Robust estimation for linear regression with asymmetric errors with applications to log-gamma regression

Ana M. Bianco, Marta García Ben and Víctor J. Yohai
Universidad de Buenps Aires

October 24, 2003

Technical Report 1/2003
Instituto De Cálculo
Facultad de Ciencias Exactas y Naturales
Universidad de Buenos Aires

Abstract

In this paper we propose a class of robust estimates for regression with errors with log-gamma distribution. These estimates, which are a natural extension of the MM-estimates proposed by Yohai for ordinary regression, may combine simultaneously high asymptotically efficiency and a high breakdown point. This paper focuses on the log-gamma regression model, but the estimates can be extended to a more general class of regression models with continuous asymmetric errors.

1 Introduction

The MM-estimates for linear regression were introduced by Yohai (1987) to achieve simultaneously two properties: high breakdown point and high efficiency for normal errors. In fact, we can find MM- estimates that simultaneously have breakdown point 0.5 and asymptotic efficiency for normal errors as close to one as desired.

The MM-estimates are computed through a three step algorithm. In the first step, we compute an initial high breakdown S-estimator of the regression parameter, which may be inefficient. In the second stage, we compute a high breakdown point M-estimate of the scale-parameter. Finally, we calculate an M-estimate which is tuned, using the scale estimate of the second step, to have high efficiency. This final estimate also inherits the breakdown point of the S-estimate computed in the first step.

Other proposal that have simultaneously these two properties are the $\tau-$estimates (Zamar and Yohai (1988)) CM-estimates (Mendes and Tyler (1996)), and the adaptive estimates proposed in Gervini and Yohai (2002).

However, when the errors distribution is asymmetric all these estimates give no consistent estimates of the intercept and therefore, the corresponding prediction of a conditional mean given the regressors are also no consistent. The estimates of the slopes are consistent, but they may lose the high efficiency property.

Marazzi and Yohai (2002) introduce a class of high breakdown point highly efficient estimates when the distribution of the errors is known except by scale parameter, as is for example the case of regression with known errors distribution except from a scale parameter. This is the case, for example, when the distribution of the errors is log-Weibull. However, these estimates are not suitable for other families of asymmetric error distributions, e.g. when the errors have log-gamma distribution.

Among the robust proposal for estimating the parameters of a gamma distribution we can mention Hampel (1968) and Marazzi and Ruffieux (1996), who deal with independent identically distributed (i.i.d) observations. Stefanski, Carroll and Ruppert (1986) and Künsch, Stefanski and Carroll (1989) obtained Hampel-optimal bounded influence estimators for generalized linear models. However, the breakdown point of these estimates tends to zero when the number of regressors increases. More recently, Cantoni and Ronchetti (2001) derived a class of robust estimates based on the notion of quasi-likelihood. All these estimates require the computation of a correction term in order to make the estimates Fisher-consistent.

In this paper we generalize the MM-estimates for the case of errors with log-gamma distribution. The definition of these estimates follows closely the original proposal of Yohai (1987) with the main difference that the ordinary residuals are replaced by deviance residuals. It is shown that the MM-estimates defined in this way a have high breakdown and high efficiency under log-gamma errors.

In Section 2 we describe the log-gamma regression model and the corresponding maximum likelihood estimate. In Section 3 we introduce M-estimates for the regression model with log-gamma errors and establish their asymptotic normality. In Section 4 we define the S-estimates and study their breakdown point and consistency. In Section 5 we define the MM-estimates and show that they may simultaneously have high breakdown point and high efficiency. Finally, Section 6 is an Appendix with some proofs.

2 Log-gamma regression models

Let us consider the parametrization of the gamma distribution, denoted by $\Gamma(\alpha,\mu)$, with density function given by

$$f(y,\alpha,\mu)=\frac{\alpha^{\alpha}}{\mu^{\alpha}\,\Gamma(\alpha)}\,y^{\alpha-1}\,e^{-(\alpha/\mu)y}\quad \text{if }\ y\geq 0. \tag{2.1}$$

Then, $\mu=E(y)>0$, $\alpha>0$ determines the shape of the density function and $\mathrm{Var}(y)=\mu^2/\alpha$.

Consider now the gamma generalized linear model where the distribution of the response z given the vector of covariates $\mathbf{x}=(x_1,x_2,...x_p)'$ has $\Gamma(\alpha_0,\mu(\mathbf{x}))$ distribution and the link function is

$$\log(\mu(\mathbf{x}))=\mathbf{x}'\beta_0. \tag{2.2}$$

If the model has intercept, the last component of $x_p=1$. Observe that we are assuming that the shape parameter α does not depend on $\mathbf{x}_i$. More details of this model can be found in Chapter 8 of McCullagh and Nelder (1989).

Since $\varepsilon=z/\mu(\mathbf{x})$ has distribution $\Gamma(\alpha,1)$, putting $y=\log(z)$ and $u=\log(\epsilon)$ this model can be also written as

$$y=\mathbf{x}'\beta_0+u, \tag{2.3}$$

where the error u has the distribution of the logarithm of a $\Gamma(\alpha,1)$ variable (log $\Gamma(\alpha,1)$). Moreover, y and u are independent and the density of u is $g(u,\alpha_0)$, where

$$g(u,\alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)}\ e^{\alpha(u-e^u)}. \tag{2.4}$$

This density is asymmetric and unimodal with maximum at 0. The model given by (3) is called log-gamma regression (LGR).

Let us assume that $(\mathbf{x}_1, y_1)... (\mathbf{x}_n, y_n)$ is a random sample of the LGR model with shape parameter α_0 and regression parameter β_0.

The Maximum Likelihood Estimator (MLE) of α_0 and β_0 are the values which maximize the likelihood function $L(\alpha,\beta)$ given by

$$L(\alpha,\beta) = \prod_{i=1}^{n} g(y_i - \mathbf{x}_i'\beta, \alpha),$$

where g is given by (4). It is well known that the maximization with respect to β is equivalent to the minimization of the deviance, which compares the log likelihood under the saturated model with the log likelihood under the model we are considering

$$D(\alpha,\beta) = \sum_{i=1}^{n} 2\ \log\left(\frac{f(0,\alpha)}{f(y_i - \mathbf{x}_i'\beta,\alpha)}\right).$$

Let us call d_i the deviance component of the i–th observation which is equal to the square of the deviance residual. In the log-gamma regression model we have

$$\begin{aligned} d_i(\alpha,\beta) &= 2\ \log\left(\frac{f(0,\alpha)}{f(y_i - \mathbf{x}_i'\beta,\alpha)}\right) = 2\,\alpha\ (-(y_i - \mathbf{x}_i'\beta) + e^{y_i - \mathbf{x}_i'\beta} - 1) \\ &= 2\alpha\ d^*(y_i, \mathbf{x}_i, \beta), \end{aligned} \tag{2.5}$$

where

$$d^*(y_i, \mathbf{x}_i, \beta) = -(y_i - \mathbf{x}_i'\beta) + e^{y_i - \mathbf{x}_i'\beta} - 1. \tag{2.6}$$

The MLE of β can be defined as

$$\widehat{\beta}^{ML} = \arg\min_{\beta} \sum_{i=1}^{n} d^*(y_i, \mathbf{x}_i, \beta). \tag{2.7}$$

Let us observe that $d^*(y_i, x_i, \beta)$ does not depend on the shape parameter α, but its distribution does depend on α.

3 M-estimates

We define M-estimates of β_0 by

$$\widehat{\beta}_n^M = \arg\min_{\beta} \sum_{i=1}^{n} \rho\left(\sqrt{d^*(y_i, \mathbf{x}_i, \beta)}/\widehat{c}\right), \tag{3.1}$$

where $\widehat{c}$ is a an estimate of a tuning constant c_0 that depends on $(\mathbf{x}_i, y_i), 1 \leq i \leq n$. We will suppose that the loss function ρ satisfies the following assumptions:

A1.(i) $\rho(0) = 0$, (ii) let $a = \sup \rho(u)$, then $0 < a < \infty$, (iii) if $0 \leq u < v$, then $\rho(u) \leq \rho(v)$, (iv) ρ is differentiable, (v) if $\rho(u) < a$ and $0 \leq u < v$, then $\rho(u) < \rho(v)$.

We will prove that the M-estimator defined in (1) is Fisher-consistent, that is, for any α_0

$$\arg\min_{\beta}\ E_{\beta_0,\alpha_0}\left[\rho\left(\sqrt{d^*(y, \mathbf{x}, \beta)}/c_0\right)\right] = \beta_0. \tag{3.2}$$

The Fisher consistency will be proved for any error density f_0 satisfying:

B. The density f_0 is strictly unimodal, continuous and $f_0(x) > 0$ for all x.

Observe that the density of a $\log\Gamma(\alpha, 1)$ distribution satisfies A3

Lemma 1 below, proved in the Appendix, shows the Fisher-consistency of a M-estimate of location for any (symmetric or not) continuous and unimodal density. Lemma 2 extends this result for a M-estimate of regression. We need the following assumption on the error density

Define $y_0 = \arg\max f_0(y)$ and put

$$d(y, \mu) = 2(\log f_0(y_0) - \log f_0(y - \mu)). \tag{3.3}$$

LEMMA 1. *Let $f_\mu(y) = f_0(y - \mu)$ be a location family of density functions, where f_0 satisfies B. Assume that y is a random variable with density $f_{\mu_0}(y)$ and $\rho(u)$ is a function that satisfies condition A1. Then,*

$$\mu_0 = \arg\min_{\mu} E_{\mu_0}\left[\rho\left(\sqrt{d(y, \mu)}/c\right)\right].$$

LEMMA 2. *Let $(\mathbf{x}, y)$ be a random vector such that*

$$y = \mathbf{x}'\beta_0 + u,$$

where the error u is independent of $\mathbf{x}$ and u has density function f_0 satisfying B. Then, if $\rho(u)$ satisfies condition A1 we have

$$\beta_0 = \arg\min_{\beta} E_{\beta_0}\left[\rho\left(\frac{\sqrt{d(y, \mathbf{x}, \beta)}}{c}\right)\right].$$

Proof. This result follows from Lemma 1, taking conditional expectation on $\mathbf{x}$.

Similar Fisher-consistency results were obtained by Lenth and Green (1987) for deviance based M-estimators. The difference is that Lenth and Green results require that the estimating equation obtained differentiating (1) has a unique solution, and therefore does not include the case of ρ bounded. It is well known that M-estimates with unbounded ρ are not robust for high leverage outliers. Marazzi and Yohai (2003). prove Lemma 2 for the particular case of the truncated likelihood, which corresponds to ρ in the family $\rho_k(t) = \min(t, k)$.

The following Theorem, which follows immediately from Lemma 2 and (5), establishes the Fisher consistency of M-estimates for the LGM.

THEOREM 1. *Let* $(\mathbf{x}, y)$ *be a random vector such that*

$$y = \mathbf{x}'\beta_0 + u,$$

where the error u *is independent of* $\mathbf{x}$ *and has* $\log\Gamma(\alpha, 1)$ *distribution. Assume that* $\rho(u)$ *is a function that satisfies condition A1 and* $d^*(y, \mathbf{x}, \beta)$ *is given by (6). Then, for any* c

$$\beta_0 = \arg\min_\beta E_{\beta_0}\left[\rho\left(\sqrt{d^*(y, \mathbf{x}, \beta)}/c\right)\right].$$

3.1 Asymptotic Distribution

In this Section we establish the asymptotic normality of M-estimates.

Let $\psi(t) = \rho'(t)$ and

$$w(t) = \frac{\psi(t)}{t}$$

and define the residuals $r(y_i, x_i, \beta) = y_i - \mathbf{x}_i'\beta$. Taking derivatives in (1), we have that the M-estimate $\widehat{\beta}_n^M$ satisfies the equation

$$\sum_{i=1}^{n} \psi\left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \widehat{\beta}_n^M)}}{\widehat{c}}\right) \frac{(1 - e^{r(y_i, \mathbf{x}_i, \widehat{\beta}_n^M)})}{\sqrt{d^*(y_i, \mathbf{x}_i, \widehat{\beta}_n^M)}} \ \mathbf{x}_i = 0 \tag{3.4}$$

or equivalently

$$\sum_{i=1}^{n} w\left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \widehat{\beta}_n^M)}}{\widehat{c}}\right) (1 - e^{r(y_i, \mathbf{x}_i, \widehat{\beta}_n^M)})\ \mathbf{x}_i = 0, \tag{3.5}$$

where

$$w(t) = \frac{\psi(t)}{t}.$$

In order to derive the asymptotic distribution of $\widehat{\beta}_n^M$ we will use the following additional assumptions.

A2. ρ is twice continuously differentiable and there exists m such that $|u| \geqslant m$ implies $\rho(u) = \sup \rho$.

A3. $E(||\mathbf{x}||^2) < \infty$, where $||\ ||$ is the l_2- norm and $E(\mathbf{x}\mathbf{x}')$ is non-singular.

The following Theorem proved in the Appendix gives the asymptotic distribution of the M-estimate.

THEOREM 2. *Let $(\mathbf{x}_1, y_1), .., (\mathbf{x}_n, y_n)$ be random vectors that satisfy the LGR model with parameters α_0 and β_0. Assume that assumptions A1-A3 hold. Let $\widehat{\beta}_n$ be a sequence of estimates satisfying (5), which is strongly consistent to β_0. Assume also that $\widehat{c}$ converges to c_0 in probability. Then,*

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) \longrightarrow^D N\left(0, \frac{B(\psi, \alpha_0, c_0)}{(A(\psi, \alpha_0, c_0))^2} E(\mathbf{x}\mathbf{x}')^{-1}\right), \tag{3.6}$$

where

$$B(\psi, \alpha, c) = E_\alpha \left[w^2 \left(\frac{\sqrt{h(u)}}{c} \right) (1 - e^u)^2 \right]$$

$$A(\psi, \alpha, c) = E_\alpha \left[w' \left(\frac{\sqrt{h(u)}}{c} \right) \frac{(1 - e^u)^2}{2c\sqrt{h(u)}} + w \left(\frac{\sqrt{h(u)}}{c} \right) e^u \right],$$

where u has $\log \Gamma(\alpha, 1)$ *distribution and*

$$h(u) = -u + e^u - 1. \tag{3.7}$$

3.2 Asymptotic Efficiency

The maximum likelihood estimate $\widehat{\beta}^{ML}$ is the M-estimate with loss function $\rho_{MV}(u) = u^2$, or equivalently with score function $\psi_{MV}(u) = 2u$. It is easy to check that

$$\frac{B(\psi_{MV}, \alpha_0, c_0)}{(A(\psi_{MV}, \alpha_0, c_0))^2} = \frac{1}{\alpha_0}.$$

Therefore, the asymptotic efficiency of the M-estimate for the LGM respect to the maximum likelihood estimator is given by

$$ARE(\psi, \alpha_0, c_0) = \frac{1}{\alpha_0} \frac{(A(\psi, \alpha_0, c_0))^2}{B(\psi, \alpha_0, c_0)}.$$

Suppose that $\psi(0) = a > 0$, then it is easy to show that $\lim_{c\to\infty} w(u/c) = a$, $\lim_{c\to\infty} w'(u/c) = 0$, and then

$$\lim_{c\to\infty} ARE(\psi, \alpha_0, c) = 1 \tag{3.8}$$

Hence, we can choose c_0 to achieve any desired efficiency e. Observe that c_0 will depend also on α and then, we can define $C(e,\alpha)$

$$ARE(\psi,\alpha_0,C(\alpha_0,e))=e.$$

Then, to choose a value c_0, so that the MM-estimate defined by (1) achieves a given efficiency e, an estimate of α_0 is required. The MM-estimates defined in Section 5 includes an estimate of α_0 and then, $C(\alpha_0,e)$ can be also estimated. Therefore, they can calibrated to attain any desired efficiency.

4 S-estimators

As it was noted in the previous section, in order to calibrate the M-estimator an initial estimate of the shape parameter α is needed. For this purpose we will compute an initial estimate of β by means of a S-estimate. The S-estimates were introduced by Rousseeuw and Yohai (1984) for ordinary regression.

For each value of β let $s_n(\beta)$ be the M-scale estimate of $\sqrt{d^*(y_i,\mathbf{x}_i,\beta)}$ given by

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{\sqrt{d^*(y_i,\mathbf{x}_i,\beta)}}{s_n(\beta)}\right)=b, \tag{4.1}$$

where ρ satisfies A1. In order to obtain an estimate with high breakdown point, we choose $b=\frac{1}{2}\sup\rho$.

The S-estimate of β for the LGR model is defined by

$$\widetilde{\beta}_n=\arg\min_{\beta}\ s_n(\beta) \tag{4.2}$$

and the corresponding scale estimate by

$$\widehat{s}_n=\min_{\beta}\ s_n(\beta). \tag{4.3}$$

Next Theorem shows that the S-estimates are Fisher consistent. Define the asymptotic value of $s_n(\beta)$ as the value $s(\alpha_0,\beta,\beta_0)$ such that

$$E_{\alpha_0\beta_0}\left[\rho\left(\frac{\sqrt{d^*(y,\mathbf{x},\beta)}}{s(\alpha_0,\beta,\beta_0)}\right)\right]=b. \tag{4.4}$$

It is easy to show that $s(\alpha_0,\beta,\beta_0)=s(\alpha_0,\beta-\beta_0,0)=s^*(\alpha_0,\beta-\beta_0)$.

Theorem 3.*Let $(\mathbf{x}_1',y_1)',...,(\mathbf{x}_n',y_n)'$ be independent vectors satisfying the LGR model and assume that ρ satisfies A1. Then, the S-estimate is Fisher consistent, i.e., $\beta_0=\arg\min s^*(\alpha_0,\beta-\beta_0)$.*

Proof. From Theorem 1 we have that for any $\beta\neq\beta_0$

$$E_{\alpha_0\beta_0}\left[\rho\left(\frac{\sqrt{d^*(y,\mathbf{x},\beta_0)}}{s^*(\alpha_0,\beta-\beta_0)}\right)\right]<E_{\alpha_0\beta_0}\left[\rho\left(\frac{\sqrt{d^*(y,\mathbf{x},\beta)}}{s^*(\alpha_0,\beta-\beta_0)}\right)\right]=b.$$

From (4) and the fact that $E(\rho(u/s))$ is decreasing in s, we get

$$s(\alpha_0, \mathbf{0}) < s(\alpha_0, \beta - \beta_0),$$

and this proves the Theorem.

The following Theorem can be proved using similar arguments to those used in Theorem 4.1 of Yohai and Zamar (1988). The main change in the proof is to replace the initial estimates by the true parameters. We need the following condition:

A4. $P(\mathbf{x}'\theta = 0) < 1/2$ for all $\theta \neq 0$.

Write $S^*(\alpha) = s^*(\alpha, \mathbf{0})$,then

$$E_\alpha\left[\rho\left(\frac{\sqrt{h(u)}}{S^*(\alpha)}\right)\right] = b, \tag{4.5}$$

where $h(u)$ was defined in (7) and u has log-$\Gamma(\alpha, 1)$ distribution.

THEOREM 4. *Let $(\mathbf{x}_1', y_1).', ., (\mathbf{x}_n', y_n)'$ be independent vectors that satisfy the LGR model . Assume that conditions A1 and A4 hold. Then,*

(a) *the S-estimate $\widetilde{\beta}_n$ defined in (2) is strongly consistent to β_0,*

(b) *the scale estimate $\widehat{s}_n$ is strongly consistent to $S^*(\alpha_0)$.*

It is immediate that the S-estimate given in (2) also satisfies

$$\widetilde{\beta}_n = \arg\min_\beta \sum_{i=1}^n \rho\left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \beta)}}{\widehat{s}_n}\right)$$

and therefore is an M-estimate. Therefore, is asymptotic distribution is given by Theorem 2 with $c_0 = S^*(\alpha_0)$.

We will use part (b) of Theorem 4 to define a robust estimate of α. For this purpose we need to show that S^* is invertible. This is done in the next two Lemmas.

LEMMA 3. *Let y be a $\log \Gamma(\alpha, 1)$ distribution . Then $h(y) = -y + e^y - 1$ is strictly stochastically decreasing with α.*

Proof . We have that $h(y) = b(x) = x - \log x - 1$, where $x = e^y$ has distribution $\Gamma(\alpha, 1)$. According to (1), the density of x is

$$f(x, \alpha) = A(\alpha)e^{-\alpha x + (\alpha-1)\log x} = A(\alpha)e^{-\alpha}e^{-\alpha b(x)}x^{-1} = A^*(\alpha)e^{-\alpha b(x)}x^{-1}.$$

The function $b(x)$ is invertible in the intervals $(0, 1]$ and $[1, \infty)$. Let b_1 and b_2 the inverses of $b(x)$ in these two intervals. Then, $z = h(y) = b(x)$ takes only positive values with density

$$f^*(z, \alpha) = A^*(\alpha)e^{-\alpha z}\left(\frac{|b_1'(z)|}{b_1(z)} + \frac{|b_2'(z)|}{b_2(z)}\right) = A^*(\alpha)e^{-\alpha z}b^*(z),$$

for some function b^* and where $A^*(\alpha) = \left(\int_0^\infty e^{-\alpha w} b^*(w)dw\right)^{-1}$.

We have to show that

$$F^*(z,\alpha) = \frac{\int_0^z e^{-\alpha w} b^*(w)dw}{\int_0^\infty e^{-\alpha w} b^*(w)dw} = \frac{\int_0^z e^{-\alpha w} b^*(w)dw / \int_z^\infty e^{-\alpha w} b^*(w)dw}{\left(\int_0^z e^{-\alpha w} b^*(w)dw / \int_z^\infty e^{-\alpha w} b^*(w)dw\right) + 1}$$

is strictly increasing with α. This is equivalent to show that

$$G(z,\alpha) = \frac{\int_0^z e^{-\alpha w} b^*(w)dw}{\int_z^\infty e^{-\alpha w} b^*(w)dw}$$

is strictly decreasing with α. Take $\alpha_1 < \alpha_2$ and let $\Delta = \alpha_2 - \alpha_1$. Then

$$G(z,\alpha_2) = \frac{\int_0^z e^{-\alpha_1 w} e^{-\Delta w} b^*(w)dw}{\int_z^\infty e^{-\alpha_1 w} e^{-\Delta w} b^*(w)dw} > \frac{e^{-\Delta z} \int_0^z e^{-\alpha_1 w}\ b^*(w)dw}{e^{-\Delta z} \int_z^\infty e^{-\alpha_1 w}\ b^*(w)dw} = G(z,\alpha_1).$$

This proves the Lemma.

LEMMA 4. *The function $s^*(\alpha) : (0,\infty) \longrightarrow (0,\infty)$ defined in (5) is continuous, strictly decreasing and surjective if ρ satisfies* ***A1***.

Proof. $S^*(\alpha)$ can be defined by

$$d(\alpha, S^*(\alpha)) = b,$$

where

$$d(\alpha, s) = E_\alpha \rho\left(\frac{\sqrt{h(y)}}{s}\right),$$

where y has distribution $\log\Gamma(\alpha, 1)$.

Using the Dominated Convergence Theorem we have that $d(\alpha, s)$ is continuous. Since, by Lemma 3, the distribution of $h(y)$ is strictly stochastically decreasing with α, we have that $d(\alpha, s)$ is decreasing in α. Thus, $S^*(\alpha)$ is continuous and decreasing.

To show $\lim_{\alpha\to\infty} S^*(\alpha) = 0$, let $\alpha_n \to \infty$ and x_n a sequence of variables with distribution $\Gamma(\alpha_n, 1)$. Since $E(x_n) = 1$ and $\text{var}(x_n) = 1/\alpha_n \to 0$, we have that $y_n = \log x_n \to 0$ in probability, and therefore, $h(y_n) \to 0$ in probability too. Then, since ρ is bounded and $\rho(0) = 0$, we get that $d(\alpha_n, s) \to 0$ for any s and this implies that $S^*(\alpha_n) \to 0$.

Finally we will show that $\lim_{a\to 0} S^*(\alpha) = \infty$. We start showing that if x has distribution $\Gamma(\alpha, 1)$, then $\lim_{\alpha\to 0} E_\alpha(x^{1/2}) = 0$. We have that

$$E_\alpha(x^{1/2}) = \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha)\alpha^{1/2}}\ .$$

Since $\lim_{\alpha\to\infty}\Gamma(\alpha+1/2)=\pi^{1/2}$, it will be enough to show that $\lim_{\alpha\to 0}\Gamma(\alpha)\alpha^{1/2}=\infty$. We have that

$$\begin{aligned}\Gamma(\alpha) &= \int_0^\infty e^{-\alpha z} z^{\alpha-1} dz \\ &\geq e^{-\alpha}\int_0^1 z^{\alpha-1}dz \\ &= e^{-\alpha}/\alpha\end{aligned}$$

Therefore, $\lim_{\alpha\to 0}\Gamma(\alpha)\alpha^{1/2}\geq \lim_{\alpha\to 0} e^{-\alpha}/\alpha^{1/2}=\infty$. Consider now $\alpha_n\to 0$ and x_n a sequence of variables with distribution $\Gamma(\alpha_n,1)$. Since $E(x_n^{1/2})\to 0$, we have that $x_n\to 0$ in probability. Then, $y_n=\log(x_n)\to-\infty$ in probability and $h(y_n)\to\infty$ in probability. Hence, $d(\alpha_n,s)\to\max\rho>b$, and this implies that $S^*(\alpha_n)\to\infty$.

As a consequence of the above proposition we can compute the inverse of S^*, that we denote S^{*-1}. Since $\widehat{s}_n$ is a consistent estimate of $S^*(\alpha)$, a consistent estimate of α_0 is given by

$$\widehat{\alpha}_n = S^{*-1}(\widehat{s}_n). \tag{4.6}$$

Then, from Theorem 4 follows immediately the following result .

Theorem 5. Under the same conditions as in Theorem 4, the estimate $\widehat{\alpha}_n$ defined in (6) is strongly consistent to α_0.

Proof. Follows from Theorem 4 and the fact that S^{*-1} is continuous.

4.1 Breakdown point

One measure of the robustness of an estimate is the breakdown point. Loosely speaking, the breakdown point of an estimate is the smallest fraction of outliers than can take the estimate beyond any limit. Hampel (1971) introduced the asymptotic version of the breakdown point and Donoho and Huber (1983) defined a finite sample version.

More formally, let Z be a data set of n elements, $Z=\{\mathbf{z}_1,...,\mathbf{z}_n\}$, $\mathbf{z}_i=(\mathbf{x}_i',y_i)'$, $\mathbf{x}_i\in R^p$, $y_i\in R$. Let $\mathcal{Z}_m$ be the set of all the samples $Z^*=\{\mathbf{z}_1^*,...,\mathbf{z}_n^*\}$ such that $\#\{i:\mathbf{z}_i=\mathbf{z}_i^*\}=n-m$. Given an estimate $\widehat{\beta}$ of the regression parameter, we define

$$b_m(Z,\widehat{\beta})=\sup\{||\widehat{\beta}(Z^*)||, Z^*\in\mathcal{Z}_m\},$$

and $m^*=\inf\{m: b_m(Z,\widehat{\beta})=\infty\}$. Then the finite sample breakdown point (FSBDP) of $\widehat{\beta}$ in Z is $\varepsilon^*(Z,\beta_n)=m^*/n$.

Since the scale estimate $\widehat{s}_n$ take values between zero and infinity, we can define two FSBDP, one to zero and one to infinity as follows

$$b_m^+(Z,\widehat{s}_n)=\sup\{\widehat{s}_n(Z^*),Z^*\in\mathcal{Z}_m\}, b_m^-(Z,\widehat{s}_n)=\inf\{\widehat{s}_n(Z^*),Z^*\in\mathcal{Z}_m\},$$

and

$$m^* = \min\left\{m : b^+_m(Z, \widehat{s}_n) = \infty\right\},\ m^{**} = \min\left\{m : b^-_m(Z, \widehat{s}_n) = 0\right\}.$$

Then, the FSBDP of $\widehat{s}_n$ to infinity is defined by $\varepsilon^*(Z, \widehat{s}_n) = m^*/n$ and the FSBDP to *zero* by $\varepsilon^{**}(Z, \widehat{s}_n) = m^{**}/n$.

If the set Z has the property that the covariate vectors of any subset of p observations, are linearly independent it is said that Z is in general position. Next Theorem shows that if Z is in general position, then the S-estimates with $b = \sup\rho$ have FSBDP point close to 0.5.

THEOREM 6 . *Let* $Z = \{\mathbf{z}_1, ..., \mathbf{z}_n\}$, $\mathbf{z}_i = (\mathbf{x}'_i, y_i)'$ *and let* $c_n = \max_{a\in R^p} \#\{i : a'x_i = 0\}/n$. *Suppose that assumptions A1 and A2 hold , then*

(a) *the S-estimator* $\widetilde{\beta}_n$ *defined in (2) has FSBDP satisfying*

$$\varepsilon^*(Z, \beta_n) \geq \frac{1 - 2c_n}{2(1 - c_n)}.$$

(b) *The scale estimate* $\widehat{s}_n$ *has FSBDP satisfying*

$$\varepsilon^*(Z, \widehat{s}_n) \geq 0.5 - c_n, \varepsilon^{**}(Z, \widehat{s}_n) \geq (1 - 2c_n)/(2(1 - c_n)\ ).$$

Proof. The proof is similar to that of Lemma 3.3 in Yohai and Zamar (1986) replacing the classical residuals $y - \mathbf{x}'\beta$ by $\sqrt{d^*(y, \mathbf{x}, \beta)}$.

If the set Z is in general position $c_n = (p - 1)/n$ and $\varepsilon^*(Z, \beta_n)$ is close to 0.5.

The parameter α can take two extreme values, 0 and ∞, and therefore we can define two FSBDP of $\widehat{\alpha}_n$ similarly to those defined for $\widehat{s}_n$, one to zero $\varepsilon^*(Z, \widehat{\alpha}_n)$ and one to infinity $\varepsilon^{**}(Z, \widehat{\alpha}_n)$.

The following Theorem gives a lower bound for both FSBDP.

THEOREM 7 . *Assume that assumptions A1 and A2 hold, then*

(a) $\varepsilon^*(Z, \widehat{\alpha}_n) \geq 0.5 - c_n$.

(b) $\varepsilon^{**}(Z, \widehat{\alpha}_n) \geq (1 - 2c_n)/(2(1 - c_n)\ )$.

Proof. Follows immediately from Theorem 6 and Lemma 4.

Then if the sample is in general position, both FSBDP are close to 0.5.

5 MM-Estimates

The MM-estimator for the LGR model is a natural extension of the MM-estimate introduced by Yohai (1987) for ordinary regression. It is defined by the following three step procedure:

• **Step** 1. Compute an S-estimate $\widetilde{\beta}_n$ and the corresponding scale estimate $\widehat{s}_n$ taking $b = \frac{1}{2}\sup\rho$.

• **Step 2**. Compute $\widehat{\alpha}_n = S^{*-1}(\widehat{s}_n)$ and

$$\widehat{c}_n = \min\{c : c \geq \widehat{s}_n \text{ and } ARE(\psi, \widehat{\alpha}_n, c) \geq e\}.$$

• **Step 3**. The MM-estimate is defined as any local minimum $\widehat{\beta}_n$of $R_n(\beta)$, where

$$R_n(\beta) = \sum_{i=1}^{n} \rho\left(\frac{\sqrt{d^*(y_i, \mathbf{x}_i, \beta)}}{\widehat{c}_n}\right), \tag{5.1}$$

satisfying $R_n(\widehat{\beta}_n) \leq R_n(\widetilde{\beta}_n)$.

Since $ARE(\psi, \alpha, c)$ is continuous in α and $c,$ and $\lim_{c\rightarrow\infty} ARE(\psi, \widehat{\alpha}_n, c) = 1,$ the value $\widehat{c}_n$ is well defined. Moreover, since $s_n \rightarrow s^*(\alpha_0)$ and the efficiency of the S-estimate, $ARE(\psi, \alpha, s^*(\alpha))$, is very low, in most cases we have $s_n < C(\widehat{\alpha}_n, e)$, and therefore $\widehat{c}_n = C(\widehat{\alpha}_n, e)$.

In practice, in the first step, we compute an approximate S-estimate by subsampling of elementary sets as proposed in Rousseeuw and Leroy (1987). In the third step, we compute $\widehat{\beta}_n$ using a reweighted least square algorithm starting from the estimate $\widetilde{\beta}_n$ computed in Step 1. Details on the computer algorithm can be found in Section 6.

According to Theorems 5 and 6, the choice of $b = \sup\rho/2$ in the first step guarantees that $\widetilde{\beta}_n$ and $\widehat{\alpha}_n$ have FSBDP close to 0.5 for sample in general position. Next Theorem show that the FSBDP lower bound found for the S-estimate in Theorem 5 is also a lower bound for the FSBDP of the MM-estimate. The proof is completely similar to the one of Theorem 2.1 in Yohai (1987).

Theorem 8 . *Let $Z = \{z_1, ..., z_n\}$, $z_i = (\mathbf{x}_i', y_i)'$, and assume that* A1 *and* A2 *hold. Then $\varepsilon(Z, \widehat{\beta}_n) \geq (1 - 2c_n)/(2(1 - c_n)\ )$.*

Next theorem establishes that the efficiency of the MM estimate is at least e.

Theorem 9 . *Let $(\mathbf{x}_1', y_1)', ., (\mathbf{x}_n', y_n)'$ be random vectors that satisfy the LGR model with parameters α_0 and β_0. Suppose that assumptions A1-A3 and C hold. Let $\widehat{\beta}_n$ be a sequence of MM-estimates . Then,*

$$\sqrt{n}(\widehat{\beta}_n - \beta_o) \longrightarrow^D N\left(0, \frac{B(\psi, \alpha_0, c_0)}{(A(\psi, \alpha_0, c_0))^2} E(\mathbf{x}\mathbf{x}')^{-1}\right), \tag{5.2}$$

where $c_0 = \min\{c : c \geq s^*(\alpha_0)$ and $ARE(\psi, \alpha_0, c) \geq e\}$. Thus, the asymptotic efficiency of $\widehat{\beta}_n$ given by $ARE(\psi, \alpha_0, c_0)$ is at least e.

Proof. Since according to Theorem 4, the initial estimate $\widetilde{\beta}_n$ is strongly consistent, using arguments similar to those used in Theorem 4.1 of Yohai (1987), we can prove $\widehat{\beta}_n$ is also strongly consistent. Then, since $\lim_{n\rightarrow\infty} \widehat{c}_n = c_0$ a. s., the Theorem follows immediately from Theorem 2.

Theorems 8 and 9 show that the MM-estimate $\widehat{\beta}_n$ has simultaneously FSBDP close to 0.5 for samples in general position and the desired efficiency.

6 Appendix

Proof of Lemma 1. Since the function $\rho^*(u) = \rho(u/k)$ have the same properties than ρ,without loss of generality we can assume that $k = 1$.We begin by proving that if y has distribution $f_{\mu_0}(y)$, then

$$f_0(y - \mu_0) >_s f_0(y - \mu), \tag{6.1}$$

where $>_s$means stochastically larger. Without loss of generality it is enough to show that (1) holds for $\mu_0 = 0$.

Then, we have to show that for all $0 < z < f_0(y_0)$ and for all $\mu \neq 0$

$$P_0(f_0(y) \geq z) > P_0(f_0(y - \mu) \geq z).$$

From the assumptions on f_0 we can find A an B such that $A < y_0 < B$ and $f_0(A) = f_0(B) = z$. Then $\{f_0(y) > z\} = [A, B]$ and $\{f_0(y - \mu) > z\} = [A + \mu, B + \mu]$. Suppose that $\mu > 0$, then

$$\begin{aligned} & P_0(f_0(y) \geq z) - P_0(f_0(y - \mu) \geq z) \\ &= \int_A^{A+\mu} f_0(y)\, dy - \int_B^{B+\mu} f_0(y)\, dy \\ &= \int_A^{A+\mu} [f_0(y) - f_0(y + B - A)]\, dy. \end{aligned}$$

Then, it is enough to show that for $y \in [A, A + \mu]$ we have

$$f_0(y) > f_0(y + B - A). \tag{6.2}$$

Suppose first that $A < y < y_0$, then $f_0(y) > f_0(A) = z$ and since $y+B-A > B$, we have $f_0(y + B - A) < f_0(B) = z$. Hence, (2) follows.

Suppose now that $y_0 < y$, then since $y + B - A > y$, (2) follows from the unimodality of f_0. The case of $\mu < 0$ is similar.

Since log is a strictly increasing function, we get from (1) that $d(y, \mu_0) <_s d(y, \mu)$.Thus , if we denote $G_\mu(t)$ the distribution function of $\rho\left(\sqrt{d(y, \mu)}\right)$, A1 implies that

$$G_{\mu_0}(t) \begin{cases} = & G_\mu(0) = 0 & \text{if } t = 0 \\ > & G_\mu(t) & \text{if } 0 < t < A \\ = & G_\mu(t) = 1 & \text{if } t \geq A. \end{cases}$$

Finally,since for a non-negative random variable z with distribution function F, $E(z) = \int_o^\infty (1 - F(z))dz$, we get that

$$E_{\mu_0}\left[\rho\left(\sqrt{d(y, \mu_0)}\right)\right] < E_{\mu_0}\left[\rho\left(\sqrt{d(y, \mu)}\right)\right].$$

This proves the Lemma.

Proof of Lemma 2. Let

$$\widehat{G}(\beta) = \sum_{i=1}^{n} w\left(\frac{\sqrt{d^*(y_i,\mathbf{x}_i,\beta)}}{\widehat{c}}\right)(1 - e^{r(y_i,\mathbf{x}_i,\beta)})\,\mathbf{x}_i$$

and

$$G(\beta) = \sum_{i=1}^{n} w\left(\frac{\sqrt{d^*(y_i,\mathbf{x}_i,\beta)}}{c_0}\right)(1 - e^{r(y_i,\mathbf{x}_i,\beta)})\,\mathbf{x}_i.$$

Then, β_n satisfies $\widehat{G}(\beta_n) = 0$. Using a first order Taylor expansion we obtain

$$\widehat{G}(\widehat{\beta}_n) = \widehat{G}(\beta_0) + \widehat{G}'(\xi_n)(\widehat{\beta}_n - \beta_0) = 0,$$

where

$$\widehat{G}'(\beta) = \frac{\partial \widehat{G}(\beta)}{\partial \beta}$$

and ξ_n is an intermediate point between $\widehat{\beta}_n$ and β_0. Hence, we have that

$$\sqrt{n}(\widehat{\beta}_n - \beta_0) = -\left[\frac{1}{n}\widehat{G}'(\xi_n)\right]^{-1}\frac{1}{\sqrt{n}}\widehat{G}(\beta_0). \tag{6.3}$$

Observe that

$$\widehat{G}'(\beta) = \sum_{i=1}^{n}\left[w'\left(\frac{\sqrt{d^*(y_i,\mathbf{x}_i,\beta)}}{\widehat{c}}\right)\frac{(1-e^{r(y_i,\mathbf{x}_i,\beta)})^2}{2c\sqrt{d^*(y_i,\mathbf{x}_i,\beta)}} + w\left(\frac{\sqrt{d^*(y_i,\mathbf{x}_i,\beta)}}{\widehat{c}}\right)e^{r(y_i,\mathbf{x}_i,\beta)}\right]\mathbf{x}_i\mathbf{x}_i'.$$

From the Central Limit Theorem and the Fisher-consistency we have that

$$\frac{1}{\sqrt{n}}G(\beta_0) \longrightarrow^D N_p\left(0, E_{\alpha_0}\left[w^2\left(\frac{\sqrt{h(u)}}{c_0}\right)(1-e^u)^2\right]E(\mathbf{x}\mathbf{x}')\right). \tag{6.4}$$

Using arguments similar to those used in Lemma 5.1 in Yohai (1987) it is possible to show that

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}(\widehat{G}(\beta_0) - G(\beta_0)) = 0 \tag{6.5}$$

in probability.

Put

$$g(y,\mathbf{x},\beta,c) = \left[w'\left(\frac{\sqrt{d^*(y,\mathbf{x}_i,\beta)}}{\widehat{c}}\right)\frac{(1-e^{y-\beta'\mathbf{x}})^2}{2c\sqrt{d^*(y,\mathbf{x},\beta)}} + w\left(\frac{\sqrt{d^*(y,\mathbf{x},\beta)}}{\widehat{c}}\right)e^{y-\beta'\mathbf{x}}\right]\mathbf{x}_i\mathbf{x}_i'. \tag{6.6}$$

Then

$$\widehat{G}'(\xi_n) = \sum_{i=1}^{n} g(y_i, \mathbf{x}_i, \xi_n, \widehat{c}). \tag{6.7}$$

We also have

$$g(y_i, \mathbf{x}_i, \beta_0, c_0) = \left[w' \left(\frac{\sqrt{h(u_i)}}{c_0} \right) \frac{(1 - e^{u_i})^2}{2c_0\sqrt{h(u_i)}} + w \left(\frac{\sqrt{h(u_i)}}{c_0} \right) e^{u_i} \right] \mathbf{x}_i\mathbf{x}_i', \tag{6.8}$$

where the u_i 's have $\log\,\Gamma(\alpha, 1)$ distribution.

Therefore using Lemma 4.2 in Yohai (1987) from (6), (7), and (8) we get

$$\frac{1}{n}G'(\xi_n) \longrightarrow^p A(\alpha_0, c)E(\mathbf{x}\mathbf{x}'). \tag{6.9}$$

From (3), (4), (9) and (5) we obtain the Theorem..

References

Cantoni, E. and Ronchetti, E. (2001). Robust inference for generalized linear models. *Journal of the American Statistical Society*, **96**, 1022-1930.

Donoho, D. L. and Huber, P. C. (1983). The notion of breakdown point. In *A Festschrift for Erich L. Lehmann* (P.J. Bickel, K.L.Docksum and J.L.Hodges Jr. eds) 157–184.Wadsworth, Belmont, California.

Hampel, F. (1968). Contributions to the theory of robust estimation. Ph. D. dissertation. University of California, Berkeley.

Hampel, F. R. (1971). A general qualitative definition of robustness, *Annals of Mathematical Statistics*, **42**, 1887–1896.

Gervini, D. and Yohai, V. J. (2002). A class of robust and fully efficient regression estimators. *Annals of Statistics*, **30**, 583-616.

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.

Lenth, R. V. and Green, P. J. (1987). Consistency of deviance-based M-estimators. *Journal of the Royal Statistical Society*, Series B, **49**, 326-330.

Mc Cullagh, P. and Nelder, J.A. (1983). *Generalized linear models.* London: Chapman and Hall.

Marazzi, A. and Ruffieux, (1996). Implementing M-estimators of the gamma distribution. In *Robust Statistics, Data Analysis and Computer Intensive Methods*. In honor of Peter Huber's 60 th Birthday, H. Rieder (ed.). Lecture Notes in Statistics, Springer Verlag, **109**, 277-297.

Marazzi, A. and Yohai, V. J. (2003). Adaptively truncated maximum likelihood regression with asymmetric errors". Accepted in *Journal of Statistical Planning and Inference*.

Mendes, B. and Tyler, D.(1996). Constrained M-estimation for regression. In *Robust Statistics, Data Analysis and Computer Intensive Methods*. In honor of Peter Huber's 60 th Birthday, H. Rieder (ed.). Lecture Notes in Statistics, Springer Verlag, **109**, 299-320.

Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimates. In *Robust and nonlinear time series*, J. Franke, W. Härdle and R. Martin (eds.). Lecture Notes in Statistics, 26, 256-272.

Stefanski, L., Carroll, R. and Ruppert, D. (1986). Bounded score functions for generalized linear models. *Biometrika*, **73**, 413-424.

Yohai, V. J. (1987). High breakdown point and high efficiency robust estimates for regression. Annals of Statistics **15**, 642-656.

Yohai, V. J. and Zamar, R. (1986). High breakdown point estimates of regression by means of the minimization of an efficient scale. Technical Report No. 84, Dept. Statistics, University of Washington.

Yohai, V. and Zamar, R. (1988). High breakdown estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, **83**, 406-413.