DEPARTAMENTO DE MATEMÁTICA
DOCUMENTO DE TRABAJO

"The choice of inicial Estimate for Computing MM-Estimates"

Marcela Svarc y Víctor Yohai

D.T.: N° 52 Abril 2008

+

Vito Dumas 284, (B1644BID) Victoria, Pcia. de Buenos Aires, Argentina
Tel: 4725-7053 Fax: 4725-7010
Email: economia@udesa.edu.ar

The choice of the Initial Estimate for Computing MM-Estimates

Marcela Svarc[1] and Víctor J. Yohai[2]

[1] Departamento de Matemática y Ciencias, Universidad de San Andrés, Vito Dumas 284, 1644 Victoria, Pcia. de Buenos Aires, Argentina (*msvarc@udesa.edu.ar*)

[2] Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires
Ciudad Universitaria, Pabellón 1, 1426 Buenos Aires, Argentina (*vyohai@dm.uba.ar*)

Abstract. We show, using a Monte Carlo study, that MM-estimates with projection estimates as starting point of an iterative weighted least squares algorithm, behave more robustly than MM-estimates starting at an S-estimate and similar Gaussian efficiency. Moreover the former have a robustness behavior close to the P-estimates with an additional advantage: they are asymptotically normal making statistical inference possible.

Keywords: Robust regression, S-estimates, P-estimates

1 Introduction

The most commonly used estimator for linear models is the least squares (LS) estimate. An observation is an atypical point or outlier if it does not fit the regression model which is followed by the large majority of the observations. It is well known that the LS estimate is extremely sensitive to outliers. In fact, a single outlier can have an unbounded effect on the LS estimate. Estimators which are not much influenced by outliers are called robust.

One measure of robustness of an estimate is its breakdown point. Heuristically, the breakdown point is the minimum fraction of arbitrary outliers that can take the estimate beyond any limit. Hampel (1971) introduced the breakdown point as an asymptotic concept, and Donoho and Huber (1983) gave the corresponding finite sample notion. The maximum possible asymptotic breakdown point of an equivariant regression estimate is 0.5.

Yohai (1987) introduced the class of MM-estimates which simultaneously have high breakdown point and high efficiency under normal errors. An MM-estimate requires an initial estimate with high breakdown point but not necessarily efficient. This initial estimate is used to compute an M-scale of the residuals Then using this scale and starting with the initial estimate, a redescending efficient M-estimate is computed using the iterated weighted least squares (IWLS) algorithm.

In general the MM-estimate computed with the IWLS algorithm corresponds to a local minimum of the M-estimate loss function, which is close to the initial estimate. Since, as we shall see later, this loss function may have more than one local minima, the degree of robustness of the MM-estimate is going to be related to the degree of robustness of the initial estimate used to start the IWLS algorithm. The most common implementation of the MM-estimate is to take as initial value an S-estimate with breakdown point 0.5.

Maronna and Yohai (1993) proposed the class of projection estimates (P-estimates) for linear models. They show that these estimates are highly robust. In fact, when the degree of robustness is measured by the maximum asymptotic bias (MAB), these estimates are much more robust than Least Median of Squares, Least Trimmed Squares and S-estimates. One shortcoming of P-estimates is that they are not asymptotically normal.

In this work we compare by Monte Carlo simulation the MM-estimate which uses the P-estimate as initial value with the MM-estimates that start with the S-estimate. We found that MM-estimates that use a P-estimate as initial value have better robustness properties than MM-estimates starting at an S-estimate. Moreover, the advantage of the MM-estimates starting at a P-estimate over the P-estimates is that they are asymptotically normal and efficient.

2 Robustness measures

Consider the linear model with p random regressors

$$y_i = \alpha_0 + \beta_0^{'} \mathbf{x}_i + u_i\,,\ i = 1, \ldots, n\,, \tag{1}$$

where $\beta_0 = (\beta_{01}...\beta_{0p})'$ and $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$. We will assume that $(u_1, \mathbf{x}_1), ..., (u_n, \mathbf{x}_n)$ are i.i.d. random vectors and that u_i is independent of $\mathbf{x}_i$. We denote by F_0 the distribution of the errors u_i's, G_0 the distribution of the $\mathbf{x}_i$'s and H_0 the joint distribution of $(y, \mathbf{x})$. Throughout the paper we will assume

P1. F_0 has a density f_0 which is symmetric and unimodal.

One way to measure the degree of robustness for large samples is the maximum asymptotic bias which is defined as follows.

Consider the contamination neighborhood of size ε of H_0 given by

$$V_\varepsilon(H_0) = \{H : H = (1-\varepsilon)H_0 + \varepsilon H^*, \text{ where } H^* \text{ is arbitrary.}$$

Given a sequence of estimates $\widehat{\gamma}_n = (\widehat{\alpha}_n, \widehat{\beta}_n)$ of $\gamma_0 = (\alpha_0, \beta_0)$ and a distribution function H on R^{p+1} we define

$$\widehat{\gamma}_\infty(H) = (\widehat{\alpha}_\infty(H), \widehat{\beta}_\infty(H)) = \lim_{n\to\infty} \widehat{\gamma}_n((x_1, y_1), ..., (\mathbf{x}_n, y_n))$$

where $(\mathbf{x}_1, y_1), ..., (\mathbf{x}_n,\, y_n)$ is a random sample of H. The maximum asymptotic biases (MAB) of $\widehat{\alpha}_n$ and $\widehat{\beta}_n$ are defined by

$$\mathrm{MAB}(\{\widehat{\alpha}_n\}, H_0, \varepsilon) = \max_{H \in V_\varepsilon(H_0)} |\widehat{\alpha}_\infty(H) - \alpha_0|,$$

and

$$\mathrm{MAB}(\{\widehat{\beta}_n\}, H_0, \varepsilon) = \max_{H \in V_\varepsilon(H_0)} \widehat{(\beta}_\infty(H) - \beta_0)' \Sigma_{\mathbf{x}} \widehat{(\beta}_\infty(H) - \beta_0), \qquad (2)$$

where $\Sigma_{\mathbf{x}}$ is the covariance matrix of $\mathbf{x}$. The reason why $\Sigma_{\mathbf{x}}$ is included in (2) is to make this definition affine equivariant.

For some estimates it is very complicated to compute MAB. In this cases, we can consider the pointwise MAB (PMAB)

$$\mathrm{PMAB}(\{\widehat{\beta}_n\}, H_0, \varepsilon) = \max_{H \in V_\varepsilon^*(H_0)} (\widehat{\beta}_\infty(H) - \beta_0)' \Sigma_{\mathbf{x}} (\widehat{\beta}_\infty(H) - \beta_0),$$

where

$$V_\varepsilon^*(H_0) = \{H : H = (1-\varepsilon)H_0 + \varepsilon \delta_{(y^*, \mathbf{x}^*)},\ (y^*, \mathbf{x}^*) \in R^{p+1}\},$$

and where $\delta_{(y^*, \mathbf{x}^*)}$ is the point mass distribution at $(y^*, \mathbf{x}^*)$. In a similar way we can define $\mathrm{PMAB}(\{\widehat{\alpha}_n\}, H_0, \varepsilon)$.

A measure of the robustness behavior for finite samples of an estimate $(\widehat{\alpha}_n, \widehat{\beta}_n)$, is the pointwise maximum mean square error (PMMSE) defined by

$$\mathrm{PMMSE}(\{\widehat{\alpha}_n\}, H_0, \varepsilon) = \max_{H \in V_\varepsilon^*(H_0)} E((\widehat{\alpha}_n(H) - \alpha_0)^2)$$

and

$$\mathrm{PMMSE}(\{\widehat{\beta}_n\}, H_0, \varepsilon) = \max_{H \in V_\varepsilon^*(H_0}) E((\widehat{\beta}_n(H) - \beta_0)' \mathbf{\Sigma}_{\mathbf{x}} (\widehat{\beta}_n(H) - \beta_0)).$$

3 S-estimates

Put $\gamma = (\alpha, \beta)$, $\gamma_0 = (\alpha_0, \beta_0)$. Define the residual vector $\mathbf{r}(\gamma) = (r_1(\gamma), ..., r_n(\gamma))$ by $r_i(\gamma) = y_i - \alpha - \beta' \mathbf{x}_i$. Consider a function ρ satisfying property P2 below

P2. $\rho : R \to R$ satisfies: (i) ρ is even, (ii) $\rho(0) = 0$, (iii) $0 \leq u_1 < u_2$ implies $\rho(u_1) \leq \rho(u_2)$, (iv) ρ is bounded, (v) $\sup \rho > 0$.

The S-estimates introduced by Rousseeuw and Yohai (1984) are defined by

$$\widehat{\gamma} = \arg \min_{\gamma \in R^{p+1}} S(\gamma),$$

where $S(\gamma)$ is defined as the value s solution of

$$\frac{1}{n-p-1}\sum_{i=1}^{n}\rho\left(\frac{r_i(\gamma)}{s}\right)=b,$$

where b is a given number.

Rousseeuw and Yohai (1984) and Davies (1990) proved that under general assumptions that include P1 and P2 we have

$$n^{1/2}(\widehat{\gamma}-\gamma_0)\rightarrow_D N(0,\sigma^2 c(\psi,F_0,\sigma)E(\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}')),$$

where $\widetilde{\mathbf{x}}'=(1,\mathbf{x}')$, $\rightarrow^D$ denotes convergence in distribution, $\psi=\rho'$, σ is the asymptotic value of s which is given by the solution of

$$E_{F_0}\left(\rho\left(\frac{u}{\sigma}\right)\right)=b, \tag{3}$$

and where

$$c(\psi,F,\sigma)=\frac{E_F\left(\psi^2\left(\frac{u}{\sigma}\right)\right)}{E_F^2\left(\psi'\left(\frac{u}{\sigma}\right)\right)}. \tag{4}$$

Generally ρ is calibrated so that σ coincides with the standard deviation when F_0 is normal. A necessary and sufficient condition for this is that the solution of (3) be $\sigma=1$ when F_0 is the N(0,1) distribution function.

Rousseeuw and Yohai (1984) have also shown that if $P(\mathbf{a}'\mathbf{x}=b)=0$ for all $\mathbf{a}\in R^p$ and $b\in R$, the asymptotic breakdown point of an S-estimate is given by

$$\varepsilon^*=\min\left(\frac{b}{a},1-\frac{b}{a}\right),$$

where $a=\max_u\rho(u)$. Note that if $b=a/2$ then $\varepsilon^*=0.5$, which is the highest asymptotic breakdown point that a regression equivariant estimate can have.

One family of functions satisfying P2 is the bisquare family ρ_c^B where $c>0$, given by

$$\rho_c^B(u)=\begin{cases}3\left(\frac{u}{c}\right)^2-3\left(\frac{u}{c}\right)^4+\left(\frac{u}{c}\right)^6 & \text{if } |u|\le c\\ 1 & \text{if } |u|>c.\end{cases} \tag{5}$$

If we take $c=1.56$ and $b=0.5$, the asymptotic breakdown point of the corresponding S-estimate is 0.5. Moreover with these choices, when F_0 is normal, the solution σ of (3) coincides with the standard deviation.

Hossjer (1992) showed that S-estimates can not be simultaneously highly efficient under a normal model and have a high breakdown point such as 0.5. The largest asymptotic efficiency of an S-estimate with breakdown point 0.5 is 0.33.

For S-estimates, it may be proved that the MAB and PMAB coincide, and closed expressions can be found in Martin, Yohai and Zamar (1989). For

MM-estimates there are not closed expressions for MAB and PMAB. However numerical calculations (see Chapter 5 of Maronna, Martin and Yohai (2006)) show that at least in the case that ρ_0 and ρ_1 are taken in the bisquare family, the PMAB of the MM-estimates with efficiencies 0.85 and 0.95 starting at the S-estimate coincide with the PMAB of the initial S-estimate.

4 P-estimates

Maronna an Yohai (1993) introduced the P-estimates which are defined as follows. For any $\gamma = (\alpha, \beta) \in R^{p+1}$, and $\eta = (\mu, \nu) \in R^{p+1}$ let

$$A(\gamma,\eta)=\text{median}\frac{r_i(\gamma)}{\eta'\widetilde{\mathbf{x}}_i}.$$

Note that since u_i and $\mathbf{x}_i$ are independent, under P1 we have

$$A(\gamma_0,\eta) = \text{median } \frac{u_i}{\eta'\widetilde{\mathbf{x}}_i} \to 0 \text{ a.s.. for all } \gamma \in R^{p+1}$$

Then it is natural to define the projection estimates by

$$\widehat{\gamma} = \arg \min_{\gamma \in R^{p+1}} B(\gamma),$$

where

$$B(\gamma) = \sup_{\eta \in R^{p+1}} s(\eta)|A(\gamma,\eta)|.$$

and where $s(\eta) = \text{MAD}(\eta'\widetilde{\mathbf{x}}_i)$

The main results on P-estimates that can be found in Maronna an Yohai (1993) are

- The P-estimates are regression, affine and scale equivariant.
- The rate of consistency of the P-estimates is $n^{1/2}$. However the asymptotic distribution is not normal.
- Assume that $P(\mathbf{a}'\mathbf{x} = b) = 0$ for all $\mathbf{a} \in R^p$ and $b \in R$. Then the asymptotic breakdown point of P-estimates is 0.5
- The maximum bias of the P-estimates satisfies $\text{MAB}(\widehat{\gamma}_n, \varepsilon, H) \leq 2C(\varepsilon, H) + o(\varepsilon)$, where $C(\varepsilon, H)$ is a lower bound of MAB for equivariant regression estimates

In Table 1 we compare the MAB of several estimates: the S-estimate based on the bisquare function with breakdown point 0.5, the least median of squares (LMS) and least trimmed of squares (LTS) proposed by Rousseeuw (1984) and the P-estimate. Note that the P-estimate has the smallest MAB.

Table 1. Maximum Bias of Robust Estimates

Estimate	ε			
	0.05	0.10	0.15	0.20
LMS	0.53	0.83	1.13	1.52
LTS	0.73	1.02	1.46	2.02
S	0.56	0.88	1.23	1.65
P	0.16	0.36	0.56	0.82

5 MM-estimates

The class of MM-estimates proposed by Yohai (1987) combines high breakdown point with high asymptotic efficiency under normality. To define the MM-estimates we require two functions ρ_0 and ρ_1 satisfying P2 and such that $\rho_1 \leq \rho_0$. Then the MM-estimates are defined as follows:

1-Start with a consistent estimate $\widehat{\gamma}_0$ with breakdown point 0.5. It is not necessary that this estimate be asymptotically efficient.

2- Compute an M-scale s of $\mathbf{r}(\widehat{\gamma}_0)$ with breakdown point 0.5 by

$$\frac{1}{n-p-1}\sum_{i=1}^{n}\rho_0\left(\frac{r_i(\widehat{\gamma}_0)}{s}\right)=b,$$

where $b = \max_u \rho_0(u)/2$.

3- Compute a local minimum $\widehat{\gamma}_1$ of

$$M_n(\gamma)=\frac{1}{n}\sum_{i=1}^{n}\rho_1\left(\frac{r_i(\gamma)}{s}\right)$$

such that

$$M_n(\widehat{\gamma}_1) \leq M_n(\widehat{\gamma}_0).$$

Yohai (1987) proved that, under very general assumptions, $\widehat{\beta}_1$ conserves the breakdown of $\widehat{\gamma}_0$ independently of the choice of ρ_1. Moreover, under very general assumptions,

$$n^{1/2}(\widehat{\gamma}_1-\beta)\rightarrow^{D} N(0,\sigma^2 c(\psi_1,F_0,\sigma)E(\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}'),$$

where $\psi_1 = \rho_1'$, σ is the asymptotic value of s which is given by (3) and $c(\psi, F, \sigma)$ is given by (4). When ρ_0 is chosen so that σ coincides with the standard error when u is normal, the Gaussian asymptotic efficiency of the MM-estimate with respect to the LS-estimate is

$$\text{EFF}=\frac{E_\phi^2\left(\psi_1'\left(u\right)\right)}{E_\phi\left(\psi_1^2\left(u\right)\right)},$$

where ϕ is the standard normal distribution.

Therefore, since this efficiency depends only on ρ_1, we can choose this function so that under Gaussian errors EFF be equal to any desired value. For example, we can choose ρ_0 and ρ_1 in the bisquare family; i.e., we can take $\rho_0 = \rho^B_{c_0}$ and $\rho_1(u) = \rho^B_{c_1}$. In that case it will be convenient to take $c_0 = 1.55$ so that the value of σ coincides with the standard deviation under normal errors. The value of c_1 should be chosen according to the desired Gaussian efficiency. For example for an efficiency of 0.95, $c_1 = 4.65$, and for an efficiency of 0.85 $c_1 = 3.46$.

One way to compute the MM-estimate $\widehat{\gamma}_1$ is by means of the iterative weighted least squares (IWLS) algorithm starting at $\gamma^{(1)} = \widehat{\gamma}_0$. The recursion step for this algorithm is as follows:

Given $\gamma^{(j)}$ we define the weights $w_i = w(r_i(\gamma^{(j)})/s)$, $1 \leq i \leq n$, where $w(u) = \psi(u)/u$. Then, $\gamma^{(j+1)}$ is the weighted least square estimate

$$\gamma^{(j+1)} = \arg\min_{\gamma} \sum_{i=1}^{n} w_i r_i^2(\gamma)$$

In general the function $M_n(\gamma)$ has several local minima. When $n \to \infty$, $M_n(\gamma)$ converges to

$$M_\infty(\gamma) = E_H\left(\rho_1\left(\frac{y - \gamma'\widetilde{\mathbf{x}}}{\sigma(H)}\right)\right), \tag{6}$$

where H is the joint distribution of $(y, \mathbf{x})$ and $\sigma(H)$ is the asymptotic scale defined by

$$E_H\left(\rho_0\left(\frac{y - \widehat{\alpha}_{0\infty} - \widehat{\beta}'_{0\infty}\mathbf{x}}{\sigma(H)}\right)\right) = b.$$

When the linear model is satisfied and P1 holds for F_0, the only local minima of (6) is at $\gamma = \gamma_0$.

Suppose now for simplicity that the model does not have intercept, $\beta_0 = \mathbf{0}$ and that there is a fraction $(1-\varepsilon)$ of outliers equals to $(y_0, \mathbf{x}_0)$,where $\mathbf{x}_0 = (x_{0,}\ 0, ..., 0)$. In this case M_∞ depends only on β_1, the first coordinate of β. The worst situation is when $|x_0| \to \infty$ an in this case $M_\infty(\beta_1)$ has two local minima, one at 0 and another close to the the contamination slope $m_0 = y_0/x_0$. There exists a value m^* such that that when $m_0 < m^*$ the global minimum is the local minimum closest to m_0, and when $m_0 > m^*$ the global minimum is at 0. As a consequence of this, if we choose as estimate the global minimum, the maximum asymptotic bias is m^*. We illustrate this behavior in Figure 1, where $M_\infty(\beta)$ is plotted for four values of the slope m_0. In this case we take $\mathbf{x} \sim N(0, I)$ and $u \sim N(0,1)$ so that $\beta_0 = 0$.

The local minimum $\widehat{\gamma}_1$ to which the IWLS algorithm converges, depends on the initial estimate $\widehat{\gamma}_0$. In general we can state the following rule: if we start the IWLS algorithm sufficiently close to a local minimum, it will converge to that local minimum. Therefore the degree of robustness of $\widehat{\gamma}_1$ is going to be

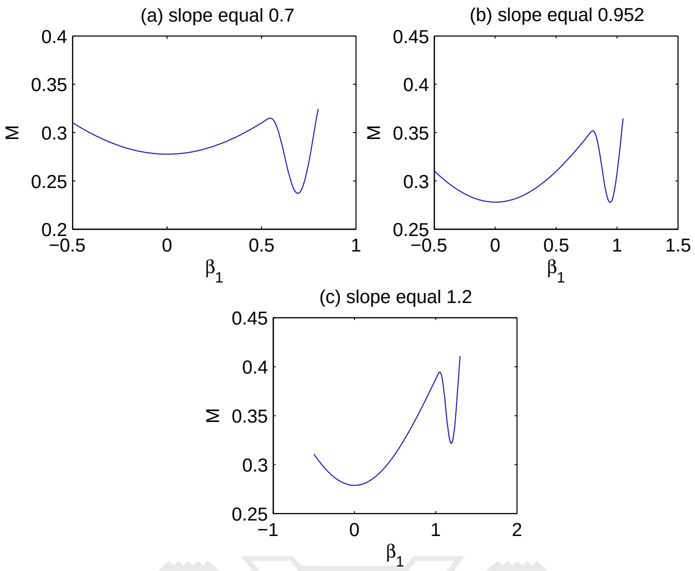

Fig. 1. (a) The global minimum is close to the contamination slope (b) There are two global minima (c) The global minimum is at 0.

related to the degree of robustness of $\widehat{\gamma}_0$. However as it is shown in Yohai (1987), the asymptotic efficiency of the MM-estimate is independent of the choice of $\widehat{\gamma}_0$.

The most popular choice of $\widehat{\gamma}_0$ (the one employed in the SPLUS, R and SAS. programs) is to start with an S-estimate based on ρ_0. However, as is shown in Table 1, the MAB of the P-estimate is smaller that the one of the S-estimate. For that reason we can expect that an MM-estimate that takes as $\widehat{\gamma}_0$ a P-estimate would be more robust than the MM-estimate that starts at an S-estimate. Simultaneously, since both MM-estimates have the same asymptotic efficiency we can expect a similar behavior under a linear model with Gaussian errors and no outliers.

6 Monte Carlo results

In this Section we report the results of a Monte Carlo study aimed to compare the performance of the MM-estimates starting at an S- and a P- estimate. The functions ρ_0 and ρ_1 were taken in the bisquare family (5). We took $\rho_0 = \rho^{(B)}_{1.55}$ and $b = 0.5$ so that (3) holds. Moreover, we chose $\rho_1 = \rho^{(B)}_{3.44}$ which corresponds to an MM-estimate having an asymptotic Gaussian efficiency of 0.85.

We consider samples of size $n = 100$, and $p = 2$, 6 and 10. In all cases a fraction $(1-\varepsilon)$ of the observations $(y_i, \mathbf{x}_i)$ were taken from a multivariate normal distribution, and the remaining observations are equal outliers $(y_0, \mathbf{x}_0)$. Because of the equivariance properties of all the estimates considered in this study, without loss of generality, the normal observations were taken with mean $\mathbf{0}$ and identity covariance matrix. This corresponds to $\beta_0 = \mathbf{0}$ and $\alpha_0 = 0$.

Because of the equivariance of the estimates considered in the study, without loss of generality, the values of $\mathbf{x}_0$ were taken of the form $(x_0, 0, .., 0)$ and $y_0 = mx_0$. We took two values of x_0: $x_0 = 1$ (low leverage outliers) and $x_0 = 10$ (high leverage outliers). We took a grid of values of m with step 0.1 and looked for the value that achieves the maximum mean square error. The number of the Monte Carlo replications was $N = 500$.

The S-estimate was computed with the fast algorithm for S-estimates proposed by Salibian-Barrera and Yohai (2006) and the P-estimate with the algorithm based on subsampling described in Maronna and Yohai (1993) taking the same set of candidates than for the S-estimate. The number of subsamples used for both estimates was 500.

In order to measure the performance of each estimate we compute the sample mean squared error (MSE) as follows. Suppose that $\gamma_n^{(1)} = (\widehat{\alpha}_n^{(1)}, \widehat{\beta}_n^{(1)})$, ..., $\widehat{\gamma}_n^{(N)} = (\widehat{\alpha}_n^{(N)}, \widehat{\beta}_n^{(N)})$ are N replications of an estimate $\widehat{\gamma}_n = (\widehat{\alpha}_n, \widehat{\beta}_n)$ of γ_0. Then we estimate the MSE of $\widehat{\alpha}_n$ and of $\widehat{\beta}_n$ by

$$\widehat{\mathrm{MSE}}\left(\widehat{\alpha}_n\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\alpha}_n^{(i)} - \alpha_0\right\|^2 . \tag{7}$$

$$\widehat{\mathrm{MSE}}\left(\widehat{\beta}_n\right) = \frac{1}{N}\sum_{i=1}^{N}\left\|\widehat{\beta}_n^{(i)} - \beta_0\right\|^2 . \tag{8}$$

We compute the following estimates: the P-estimate (P), the S-estimate (S), the MM-estimate starting from an S-estimate (SMM) and the MM-estimate starting from a P-estimate (PMM).

Table 2 show the MSE 's when there are no outliers. Tables 3 and 4 report the maximum MSE's for the case that $\varepsilon = 0.10$ for low and high leverage outliers respectively. Finally, Tables 5 and 6 give the MSE 's for the case $\varepsilon = 0.20$. From the analysis of these tables we can draw the following conclusions

- When there are not outliers both MM-estimates have a similar behavior
- When $x_0 = 1$, the MMS and MMP estimates have similar maximum MSE's for p equal 2 and 5. When $p = 10$ the MMP-estimate outperforms the MMS-estimate.
- When $x_0 = 10$, the MMP-estimate behave much better than the MMS-estimate for all values of p.

Table 2. Mean Square Errors Without Outliers

Estimates	$\widehat{\alpha}_n$			$\widehat{\beta}_n$		
	p					
	2	5	10	2	5	10
P	0.015	0.016	0.018	0.034	0.08	0.18
S	0.033	0.036	0.040	0.067	0.21	0.43
MMP	0.012	0.012	0.012	0.023	0.06	0.13
MMS	0.012	0.013	0.013	0.023	0.06	0.14

Table 3. Maximum Mean Square Errors when $\varepsilon = 0.10$ and $x_0 = 1$

Estimates	$\widehat{\alpha}_n$			$\widehat{\beta}_n$		
	p					
	2	5	10	2	5	10
P	0.040	0.044	0.054	0.099	0.19	0.33
S	0.128	0.162	0.223	0.218	0.46	0.92
MMP	0.049	0.052	0.057	0.067	0.11	0.21
MMS	0.049	0.051	0.055	0.067	0.11	0.21

Table 4. Maximum Mean Square Errors when $\varepsilon = 0.20$ and $x_0 = 1$

Estimates	$\widehat{\alpha}_n$			$\widehat{\beta}_n$		
	p					
	2	5	10	2	5	10
P	0.15	0.15	0.25	0.43.	0.84	1.92
S	0.55	0.73	1.29	1.20	2.31	4.82
MMP	0.29	0.33	0.40	0.35	0.48	0.72
MMS	0.26	0.30	0.44	0.33	0.48	1.01

Table 5. Maximum Mean Square Errors when $\varepsilon = 0.10$ and $x_0 = 10$

Estimates	$\widehat{\alpha}_n$			$\widehat{\beta}_n$		
	p					
	2	5	10	2	5	10
P	0.037	0.035	0.035	0.13	0.26	0.52
S	0.076	0.073	0.090	0.50	0.79	1.33
MMP	0.017	0.017	0.020	0.20	0.24	0.43
MMS	0.024	0.026	0.034	0.42	0.56	0.84

Table 6. Maximum Mean Square Errors when $\varepsilon = 0.20$ and $x_0 = 10$

Estimates	$\widehat{\alpha}_n$			$\widehat{\beta}_n$		
	p					
	2	5	10	2	5	10
P	0.12	0.12	0.15	0.71	1.30	2.59
S	0.21	0.27	0.31	1.90	3.09	4.90
MMP	0.03	0.05	0.07	0.65	1.12	2.17
MMS	0.07	0.11	0.17	1.75	2.48	3.84

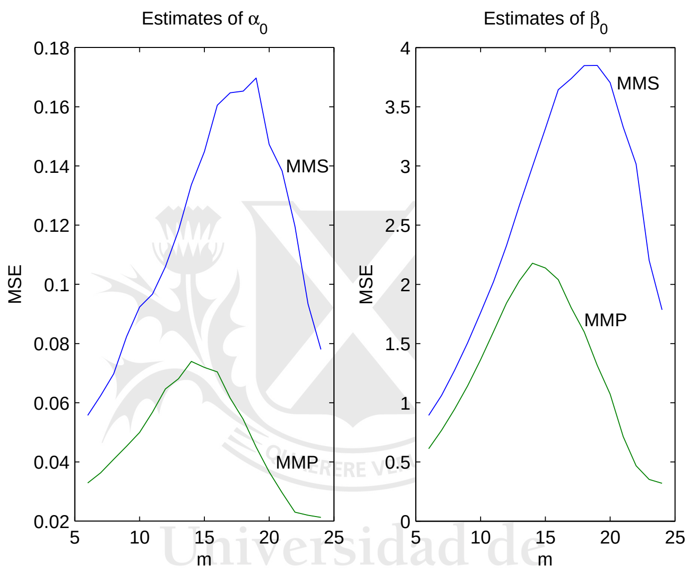

Fig. 2. MSE 's of MM-estimates for $x_0 = 10$.

Figure 2 shows the MSE as a function of the contamination slope for the two MM-estimates in the case of $x_0 = 10$, $\varepsilon = 0.10$ and $p = 10$. We observe that the MMP estimate behaves better than the MMS estimate uniformly on the contamination slope m. Similar behaviors occurs for all the other values of p and ε at $x_0 = 10$.

The computing time required to fit the PMM with 500 subsamples to a data sat of 500 observations and 10 regressors is approximately 4 seconds using a MATLAB program and a PC computer with an AMD Athlon 1.8 GHz processor

7 Concluding remarks

A Monte Carlo study has shown that MM-estimates that use a P-estimate as starting value, have a degree of robustness comparable to that of the P-estimate. and much higher than that of the MM-estimate starting at an S-estimate. On the other hand both MM-estimates have comparable Gaussian efficiencies. An additional advantage of the MM-estimate starting at a P-estimate is that, contrary to what happens with the P-estimate, it is asymptotically normal and thus allows statistical inference.

References

Davies, L. (1990). The asymptotics of S-estimators in the linear regression model. *The Annals of Statistics*, **18**, 1651-1675.

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

Hampel, F.R. (1971). A General Qualitative Definition of Robustness. *The Annals of Mathematical Statistics, 42, 1887-1896.*

Hossjer, 0, (1992). On the optimality of S-estimators, *Statistics* & *Probability Letters, 14, 413-419.*

Maronna, R. A. and Yohai, V. J. (1993). Bias-robust estimates of regression based on projections. *The Annals of Statistics, 21, 965-990.*

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

Martin, R.D., Yohai, V.J. and Zamar, R. (1989). Min-max bias robust regression. *The Annals of Statistics, 17, 1608-1630.*

Rousseeuw, P.J. (1984). Least median of squares regression, *J. Am. Stat. Assoc, 79, 871-880.*

Rousseeuw, P.J. and Yohai, V.J. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series.* (J. Franke, W. Hardle and D. Martin, eds.). Lecture Notes in Statist*istics*, **26,** 256-272. Springer-Verlag, Berlin.

Salibian-Barrera, M. and Yohai, V. J. (2006). A fast algorithm for S-regression estimates. *Journal of Computational and Graphical Statistics, 15, 414-427.*

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