Journal of Multivariate Analysis 102 (2011) 1280–1292

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Estimates of MM type for the multivariate linear model

Nadia L. Kudraszow [a,*], Ricardo A. Maronna [b]

[a] CONICET, University of La Plata, Argentina
[b] C.I.C.P.B.A, University of La Plata, Argentina

ARTICLE INFO

Article history:
Received 10 June 2010
Available online 5 May 2011

AMS 2000 subject classifications:
62F35
62H12
62J05

Keywords:
Robust methods
MM-estimate
Multivariate linear model

ABSTRACT

We propose a class of robust estimates for multivariate linear models. Based on the approach of MM-estimation (Yohai 1987, [24]), we estimate the regression coefficients and the covariance matrix of the errors simultaneously. These estimates have both a high breakdown point and high asymptotic efficiency under Gaussian errors. We prove consistency and asymptotic normality assuming errors with an elliptical distribution. We describe an iterative algorithm for the numerical calculation of these estimates. The advantages of the proposed estimates over their competitors are demonstrated through both simulated and real data.

1. Introduction

Consider a multivariate linear model (MLM) with random predictors, i.e., we observe n independent identically distributed (i.i.d.) $(p+q)$-dimensional vectors, $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')$ with $1 \le i \le n$, where $\mathbf{y}_i = (y_{i1}, \ldots, y_{iq})' \in \mathbb{R}^q, \mathbf{x}_i = (x_{i1}, \ldots, x_{ip})' \in \mathbb{R}^p$, and $'$ denotes the transpose. The $\mathbf{y}_i$ are the response vectors and the $\mathbf{x}_i$ are the predictors, and both satisfy the equation

$$\mathbf{y}_i = \mathbf{B}_0'\mathbf{x}_i + \mathbf{u}_i \quad 1 \le i \le n, \tag{1.1}$$

where $\mathbf{B}_0 \in \mathbb{R}^{p\times q}$ is the matrix of the regression parameters and $\mathbf{u}_i$ is a q-dimensional vector independent of $\mathbf{x}_i$. If $\mathbf{x}_{ip} = 1$ for all $1 \le i \le n$, we obtain a regression model with intercept.

We denote the distributions of $\mathbf{x}_i$ and $\mathbf{u}_i$ by G_0 and F_0, respectively, and $\mathbf{\Sigma}_0$ is the covariance matrix of the $\mathbf{u}_i$. The p-multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{\Sigma}$ is denoted by $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$.

In the case of $\mathbf{u}_i$ with distribution $N_q(\mathbf{0}, \mathbf{\Sigma}_0)$, the maximum likelihood estimate (MLE) of $\mathbf{B}_0$ is the least squares estimate (LSE), and the MLE of $\mathbf{\Sigma}_0$ is the sample covariance matrix of the residuals. It is known that these estimates are not robust: a small fraction of outliers may have a large effect on their values.

Several approaches have been proposed to deal with this problem. The first proposal of a robust estimate for the MLM was given by Koenker and Portnoy [13]. They proposed to apply a regression M-estimator, based on a convex loss function, to each coordinate of the response vector. The problems with this estimate is lack of affine equivariance and zero breakdown point. Several other estimates without these problems were defined later. Rousseeuw et al. [20] proposed estimates for the MLM based on a robust estimate of the covariance matrix of $\mathbf{z} = (\mathbf{x}', \mathbf{y}')$. Bilodeau and Duchesne [4] extended the S-estimates introduced by Davies [6] for multivariate location and scatter; then Van Aelst and Willems [23] studied the robustness of

* Corresponding author.
E-mail addresses: nkudraszow@mate.unlp.edu.ar (N.L. Kudraszow), rmaronna@retina.ar (R.A. Maronna).

0047-259X/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2011.04.011

these estimators. Agulló et al. [1] extended the minimum covariance determinant estimate introduced by Rousseeuw [19] and Roelandt et al. [18] extended the definition of GS-estimates introduced by Croux et al. [5]. These estimates have a high breakdown point but are not highly efficient when the errors are Gaussian and q is small. In order to solve this, Agulló et al. [1] improved the efficiency of their estimates, maintaining their high breakdown point, by considering one-step reweighting and one-step Newton–Raphson GM-estimates. García Ben et al. [7] extended τ-estimates for multivariate regression, obtaining a estimate with high breakdown point and a high Gaussian efficiency. Another important approach to obtain robust and efficient estimates is constrained M (CM) estimation, proposed by Mendes and Tyler in [17] for regression and by Kent and Tyler in [12] for multivariate location and scatter. The bias of CM-estimates for regression was studied by Berrendero et al. in [3]. Following this approach, Bai et al. in [2] proposed CM-estimates for the multivariate linear model.

In this paper, we propose robust estimates for the linear model based on the MM approach, first proposed by Yohai [24] for the univariate linear model, and later by Lopuhaä [15], Tatsuoka and Tyler [22], and Salibián-Barrera et al. [21] for multivariate location and scatter. We show that our estimates have both a high breakdown point and high normal efficiency.

In Section 2, we define MM-estimates for the MLM and prove some properties. In Sections 3 and 4, we study their breakdown point and influence function. In Sections 5 and 6, we study the asymptotic properties (consistency and asymptotic normality) of the MM-estimates assuming random predictors and errors with an elliptical unimodal distribution. In Section 7, we describe a computing algorithm based on an iterative weighted MLE. In Section 8, we present the results of a simulation study, and we present a real example in Section 9. All the proofs can be found in [14].

2. Definition and properties

Before defining our class of robust estimates for the MLM, we will define a robust estimate of scale.

Definition 1. Given a sample of size n, $\mathbf{v} = (v_1, \ldots, v_n)$, an *M-estimate of scale* $s(\mathbf{v})$ is defined as the value of s that is solution of

$$\frac{1}{n}\sum_{i=1}^{n} \rho_0\left(\frac{v_i}{s}\right) = b, \tag{2.1}$$

where $b \in (0, 1)$, or $s = 0$ if $\sharp(v_i = 0) \geq n(1 - b)$, where $\sharp$ is the symbol for cardinality.

In this paper, we use $b = 0.5$, which ensures the maximal asymptotic breakdown point (see [10]).

The function ρ_0 should satisfy the following definition.

Definition 2. A *ρ-function* will denote a function $\rho(u)$ which is a continuous nondecreasing function of $|u|$ such that $\rho(0) = 0$, $\sup_u \rho(u) = 1$, and $\rho(u)$ is increasing for nonnegative u such that $\rho(u) < 1$.

Note that according to the terminology of Maronna et al. [16] this would be a "bounded ρ-function". A popular ρ-function is the *bisquare function*:

$$\rho_B(u) = 1 - (1 - u^2)^3 I(|u| \leq 1), \tag{2.2}$$

where $I(\cdot)$ is the indicator function.

Definition 3. Given a vector $\mathbf{u}$ and a positive definite matrix $\mathbf{V}$, the *Mahalanobis norm of* $\mathbf{u}$ *with respect to* $\mathbf{V}$ is defined as

$$d(\mathbf{u}, \mathbf{V}) = (\mathbf{u}'\mathbf{V}^{-1}\mathbf{u})^{1/2}.$$

For particular given $\mathbf{B} \in \mathbb{R}^{p\times q}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{q\times q}$, we denote by $d_i(\mathbf{B}, \boldsymbol{\Sigma})$ $(i = 1, \ldots, n)$ the Mahalanobis norms of the residuals with respect to the matrix $\boldsymbol{\Sigma}$; that is,

$$d_i(\mathbf{B}, \boldsymbol{\Sigma}) = (\widehat{\mathbf{u}}_i(\mathbf{B})'\boldsymbol{\Sigma}^{-1}\widehat{\mathbf{u}}_i(\mathbf{B}))^{1/2},$$

with $\widehat{\mathbf{u}}_i(\mathbf{B}) = \mathbf{y}_i - \mathbf{B}'\mathbf{x}_i$.

Using the concepts defined above, we can describe an *MM-estimate for the MLM* by the following procedure.

Let $(\widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Sigma}}_n)$ be an initial estimate of $(\mathbf{B}_0, \boldsymbol{\Sigma}_0)$, with a high breakdown point and such that $|\widetilde{\boldsymbol{\Sigma}}_n| = 1$, where $|\widetilde{\boldsymbol{\Sigma}}_n|$ is the determinant of $\widetilde{\boldsymbol{\Sigma}}_n$. Compute the Mahalanobis norms of the residuals using $(\widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Sigma}}_n)$,

$$d_i(\widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Sigma}}_n) = (\widehat{\mathbf{u}}_i'(\widetilde{\mathbf{B}}_n)\widetilde{\boldsymbol{\Sigma}}_n^{-1}\widehat{\mathbf{u}}_i(\widetilde{\mathbf{B}}_n))^{1/2} \qquad 1 \leq i \leq n. \tag{2.3}$$

Then, compute the M-estimate of scale $\hat{\sigma}_n := s(\mathbf{d}(\widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Sigma}}_n))$ of the above norms, defined by (2.1), using the function ρ_0 as specified in Definition 2 and $b = 0.5$.

Let ρ_1 be another ρ-function such that

$$\rho_1 \leq \rho_0, \tag{2.4}$$

and let $\mathcal{S}_q$ be the set of all positive definite symmetric $q \times q$ matrices.

Let $(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Gamma}}_n)$ be any local minimum of

$$S(\mathbf{B}, \mathbf{\Gamma}) = \sum_{i=1}^{n} \rho_1 \left(\frac{d_i(\mathbf{B}, \mathbf{\Gamma})}{\hat{\sigma}_n} \right) \tag{2.5}$$

in $\mathbb{R}^{p\times q} \times \mathcal{S}_q$ which satisfies

$$S(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Gamma}}_n) \leq S(\widetilde{\mathbf{B}}_n, \widetilde{\mathbf{\Sigma}}_n) \tag{2.6}$$

and $|\widehat{\mathbf{\Gamma}}_n| = 1$. Then the MM-estimate of $\mathbf{B}_0$ is defined as $\widehat{\mathbf{B}}_n$, and the respective estimate of $\mathbf{\Sigma}_0$ is

$$\widehat{\mathbf{\Sigma}}_n = \hat{\sigma}_n^2 \widehat{\mathbf{\Gamma}}_n. \tag{2.7}$$

In the MM-estimates for the univariate linear model, the residuals are used as a tool of outlier detection; in the MM-estimates for the multivariate linear model, the Mahalanobis norms of the residuals play the same role. To compute the M-scale it is necessary to have an initial estimate of $\mathbf{B}_0$, to compute the residuals, and an initial estimate of the shape of $\mathbf{\Sigma}_0$, $\mathbf{\Sigma}_0/|\mathbf{\Sigma}_0|^{1/q}$, to compute the Mahalanobis norms of the residuals.

Remark 1. One method of choosing the ρ-functions ρ_0 and ρ_1 in such a way that they satisfy (2.4) is the following. Let ρ be a ρ-function and let $0 < c_0 < c_1$. We take

$$\rho_0 = \rho(u/c_0) \quad \text{and} \quad \rho_1 = \rho(u/c_1). \tag{2.8}$$

The value c_0 should be chosen such that the asymptotic value of $\hat{\sigma}_n$ is one when the errors $\mathbf{u}_i$, with $i = 1, \ldots, n$, have distribution $N_q(\mathbf{0}, \mathbf{I})$. The choice of c_1 will determine the asymptotic efficiency of the MM-estimate. For more details, see Remark 5.

The following theorem implies that the absolute minimum of $S(\mathbf{B}, \mathbf{\Gamma}/|\mathbf{\Gamma}|^{1/q})$ in $\mathbb{R}^{p\times q} \times \mathcal{S}_q$ exists. Clearly, from this absolute minimum we can obtain an MM-estimate. However, any other local minimum $(\mathbf{B}, \mathbf{\Gamma})$ which satisfies (2.6) may also be used to get an MM-estimate with a high breakdown point and with high efficiency under Gaussian errors.

Before stating the theorem, we define k_n as the maximum number of observations $(\mathbf{y}_i', \mathbf{x}_i')$ of a sample that are in a hyperplane; i.e.,

$$k_n := \max_{\|\mathbf{v}\|+\|\mathbf{w}\|>0} \#\{i : \mathbf{v}'\mathbf{x}_i + \mathbf{w}'\mathbf{y}_i = \mathbf{0}\}. \tag{2.9}$$

Theorem 1. *Let $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ be a sample of size n satisfying the MLM (1.1), where $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')$. If $k_n/n < 0.5$, then there is a pair $(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Gamma}}_n)$ that minimizes the function $S(\mathbf{B}, \mathbf{\Gamma})$, defined in (2.5), for all $(\mathbf{B}, \boldsymbol{\Gamma}) \in \mathbb{R}^{p\times q} \times \mathcal{S}_q$ such that $|\boldsymbol{\Gamma}| = 1$.*

In the following theorem, we obtain the estimating equations of MM-estimates.

Theorem 2. *Assume that ρ_1 is differentiable. Then the MM-estimates $(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n)$ satisfy the following equations:*

$$\sum_{i=1}^{n} W\left(d_i(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n)\right) \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n)\mathbf{x}_i' = \mathbf{0} \tag{2.10}$$

$$\widehat{\mathbf{\Sigma}}_n = q \frac{\sum_{i=1}^{n} W\left(d_i(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n)\right) \widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n)\widehat{\mathbf{u}}_i(\widehat{\mathbf{B}}_n)'}{\sum_{i=1}^{n} \psi_1\left(d_i(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n)\right) d_i(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n)}, \tag{2.11}$$

where $\psi_1(u) = \rho_1'(u)$ and $W(u) = \psi_1(u)/u$.

Remark 2. As we can see in Eq. (2.10), the *j*th column of $\widehat{\mathbf{B}}_n$ is the weighted LSE corresponding to the univariate regression whose dependent variable is the *j*th component of $\mathbf{y}$, the vector of independent variables is the same as that in the multivariate regression, and observation *i* receives the weight $W\left(d_i(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n)\right)$. Furthermore, by (2.11), $\widehat{\mathbf{\Sigma}}_n$ is proportional to the sample covariance matrix of the weighted residuals with the same weights. As these weights depend on the estimates $\widehat{\mathbf{B}}_n$ and $\widehat{\mathbf{\Sigma}}_n$, we cannot use the relations (2.10) and (2.11) to compute the estimates, but they will be used to formulate an iterative algorithm in Section 6.

Remark 3. If $\widetilde{\mathbf{B}}_n$ is regression, affine, and scale equivariant and $\widetilde{\mathbf{\Sigma}}_n$ is affine equivariant and regression and scale invariant, then $\widehat{\mathbf{B}}_n$ will be regression, affine, and scale equivariant and $\widehat{\mathbf{\Sigma}}_n$ will be regression and scale invariant and affine equivariant.

3. Breakdown point

Now, to investigate the robustness of the MM-estimates, we will seek a lower bound of their finite sample breakdown point. The finite sample breakdown point of the coefficient matrix estimate is the smallest fraction of outliers that make the estimator unbounded, and the finite sample breakdown point of the covariance matrix estimate is the smallest fraction of outliers that make the estimate unbounded or singular.

Let $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ be a sample of size n that satisfies the MLM (1.1), where $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')$, and let $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{\Sigma}}$ be estimates of $\mathbf{B}_0$ and $\mathbf{\Sigma}_0$, respectively.

We define

$$\mathcal{Z}_m = \{\mathbf{Z}^* = \{\mathbf{z}_1^*, \ldots, \mathbf{z}_n^*\} \text{ such that } \sharp\{i : \mathbf{z}_i = \mathbf{z}_i^*\} \geq n - m\},$$
$$S_m(\mathbf{Z}, \widehat{\mathbf{B}}) = \sup\{\|\widehat{\mathbf{B}}(\mathbf{Z}^*)\|_2 \text{ with } \mathbf{Z}^* \in \mathcal{Z}_m\},$$
$$S_m^+(\mathbf{Z}, \widehat{\mathbf{\Sigma}}) = \sup\{\lambda_1(\widehat{\mathbf{\Sigma}}(\mathbf{Z}^*)) \text{ with } \mathbf{Z}^* \in \mathcal{Z}_m\},$$

and

$$S_m^-(\mathbf{Z}, \widehat{\mathbf{\Sigma}}) = \inf\{\lambda_q(\widehat{\mathbf{\Sigma}}(\mathbf{Z}^*)) \text{ with } \mathbf{Z}^* \in \mathcal{Z}_m\},$$

where $\lambda_1(\widehat{\mathbf{\Sigma}}(\mathbf{Z}^*))$ and $\lambda_q(\widehat{\mathbf{\Sigma}}(\mathbf{Z}^*))$ are the largest and smallest eigenvalues of $\widehat{\mathbf{\Sigma}}(\mathbf{Z}^*)$, respectively.

Definition 4. The *finite sample breakdown point of* $\widehat{\mathbf{B}}$ is $\varepsilon^*(\mathbf{Z}, \widehat{\mathbf{B}}) = m^*/n$, where

$$m^* = \min\{m : S_m(\mathbf{Z}, \widehat{\mathbf{B}}) = \infty\},$$

the *finite sample breakdown point of* $\widehat{\mathbf{\Sigma}}$ is $\varepsilon^*(\mathbf{Z}, \widehat{\mathbf{\Sigma}}) = m^*/n$, where

$$m^* = \min\left\{m : \frac{1}{S_m^-(\mathbf{Z}, \widehat{\mathbf{\Sigma}})} + S_m^+(\mathbf{Z}, \widehat{\mathbf{\Sigma}}) = \infty\right\},$$

and $\varepsilon_n^*(\mathbf{Z}, \widehat{\mathbf{B}}, \widehat{\mathbf{\Sigma}}) = \min\{\varepsilon^*(\mathbf{Z}, \widehat{\mathbf{B}}), \varepsilon^*(\mathbf{Z}, \widehat{\mathbf{\Sigma}})\}$.

The following theorem gives a lower bound for the breakdown point of MM-estimates.

Theorem 3. *Let $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$, with $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')$ that satisfies the MLM (1.1), and k_n defined in (2.9). Consider ρ_0 and ρ_1 to be two ρ-functions that satisfy (2.4), and suppose that $k_n < n/2$. Then*

$$\varepsilon_n^*(\mathbf{Z}, \widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n) \geq \min\left(\varepsilon_n^*(\mathbf{Z}, \widetilde{\mathbf{B}}_n, \widetilde{\mathbf{\Sigma}}_n), \frac{[n/2] - k_n}{n}\right). \tag{3.1}$$

Since k_n is always greater than or equal to $p + q - 1$, if $\varepsilon_n^*(\mathbf{Z}, \widetilde{\mathbf{B}}_n, \widetilde{\mathbf{\Sigma}}_n)$ is close to 0.5, the maximum lower bound will be $([n/2] - (p + q - 1))/n$; i.e., when the points are in general position the finite sample breakdown point is close to 0.5 for large n.

If we did not fix $b = 0.5$ and if $k_n < n(1 - b)$, we would have the same bound as in (3.1) but with $[n(1 - b)]$ in place of $[n/2]$. In this case, the maximum finite sample breakdown point would be attained in $b = 0.5 - k_n/n$, which is very close to our choice of $b = 0.5$ when k_n/n is small.

4. Influence function

Consider an estimate $\widehat{\boldsymbol{\theta}}_n$ depending on a sample $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_n\}$ of i.i.d. variables in $\mathbb{R}^k$ with distribution $H_{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^m$. Let T be an estimating functional of $\boldsymbol{\theta}$ such that $\mathbf{T}(H_n) = \widehat{\boldsymbol{\theta}}_n$, where H_n is the corresponding empirical distribution. Suppose that $\mathbf{T}$ is Fisher consistent, i.e., $\mathbf{T}(H_{\boldsymbol{\theta}}) = \boldsymbol{\theta}$. The *influence function* of $\mathbf{T}$, introduced by Hampel [8], measures the effect on the functional of a small fraction of point mass contamination. If $\delta_{\mathbf{z}}$ denotes the probability distribution that assigns mass 1 to $\mathbf{x}$, then the influence function is defined by

$$IF(\mathbf{z}, \mathbf{T}, \boldsymbol{\theta}) = \lim_{\varepsilon \to 0} \frac{\mathbf{T}((1 - \varepsilon)H_{\boldsymbol{\theta}} + \varepsilon\delta_{\mathbf{z}}) - \mathbf{T}(H_{\boldsymbol{\theta}})}{\varepsilon} = \left.\frac{\partial \mathbf{T}((1 - \varepsilon)H_{\boldsymbol{\theta}} + \varepsilon\delta_{\mathbf{z}})}{\partial \varepsilon}\right|_{\varepsilon=0}.$$

In our case, $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$ satisfies the linear model (1.1), $\boldsymbol{\theta} = (\mathbf{B}_0, \mathbf{\Sigma}_0)$ and $H_{\boldsymbol{\theta}} = H_0$. Let $\mathbf{T}_{0,1}$, $\mathbf{T}_{0,2}$ be the functional estimates associated to the initial estimates $\widetilde{\mathbf{B}}_n$ and $\widetilde{\mathbf{\Sigma}}_n$, and $\mathbf{T}_1$, $\mathbf{T}_2$ the functional estimates corresponding to the MM-estimates $\widehat{\mathbf{B}}_n$ and $\widehat{\mathbf{\Sigma}}_n$. Then, according to (2.10) and (2.11), given a distribution function H of $(\mathbf{y}', \mathbf{x}')'$, the pair $(\mathbf{T}_1(H), \mathbf{T}_2(H))$

is the value of $(\mathbf{B}, \mathbf{\Sigma})$ satisfying

$$E_H W\left(d(\mathbf{B}, \mathbf{\Sigma})\right) \widehat{\mathbf{u}}(\mathbf{B})\mathbf{x}' = \mathbf{0},$$
$$\mathbf{\Sigma} = q\frac{E_H W\left(d(\mathbf{B}, \mathbf{\Sigma})\right) \widehat{\mathbf{u}}(\mathbf{B})\widehat{\mathbf{u}}(\mathbf{B})'}{E_{H_0}\psi_1\left(d(\mathbf{B}, \mathbf{\Sigma})\right) d(\mathbf{B}, \mathbf{\Sigma})},$$

and

$$\mathbf{\Sigma} = S(H)^2\mathbf{\Gamma}, \quad \text{with } |\mathbf{\Gamma}| = 1,$$

where $d(\mathbf{B}, \mathbf{\Sigma}) = d(\widehat{\mathbf{u}}(\mathbf{B}), \mathbf{\Sigma})$, $\widehat{\mathbf{u}}(\mathbf{B}) = \mathbf{y} - \mathbf{B}'\mathbf{x}$ and

$$E_H\rho_0\left(\frac{d(\mathbf{T}_{0,1}(H), \mathbf{T}_{0,2}(H))}{S(H)}\right) = \mathbf{0}.$$

Note that the M-estimate of scale, $\hat{\sigma}_n$, used in the definition of MM-estimates $(\widehat{\mathbf{B}}_n, \widehat{\mathbf{\Sigma}}_n)$, verifies $\hat{\sigma}_n = S(H_n)$, where H_n is the empirical distribution of $\mathbf{z}_1, \ldots, \mathbf{z}_n$.

Next we will state the influence function of MM-estimators for the case where errors in (1.1) have an elliptical distribution with unimodal density. For that, we need to make the following assumptions.

(A1) ρ_1 is strictly increasing in $[0, \kappa]$ and constant in $[\kappa, +\infty)$ for some constant $\kappa < \infty$.
(A2) $P_{G_0}(\mathbf{B}'\mathbf{x} = 0) < 0.5$ for all $\mathbf{B} \in \mathbb{R}^{p\times q}$.
(A3) The distribution F_0 of $\mathbf{u}_i$ has a density of the form

$$f_0(\mathbf{u}) = \frac{f_0^*(\mathbf{u}'\mathbf{\Sigma}_0^{-1}\mathbf{u})}{|\mathbf{\Sigma}_0|^{1/2}}, \tag{4.1}$$

where f_0^* is nonincreasing and has at least one point of decrease in the interval where ρ_1 is strictly increasing.
(A4) G_0 has second moments and $E_{G_0}(\mathbf{x}\mathbf{x}')$ is nonsingular.

Theorem 4. *Let $(\mathbf{y}_0', \mathbf{x}_0')$ be a random vector satisfying the MLM (1.1) with parameters $\mathbf{B}_0$ and $\mathbf{\Sigma}_0$. Assume that* (S1)–(S4) *hold and that the partial derivatives of $E_{H_0}W\left(d(\mathbf{B}_0, \mathbf{\Sigma}_0)/S(H_0)\right)\widehat{\mathbf{u}}(\mathbf{B}_0)\mathbf{x}'$ can be obtained by differentiating with respect to each parameter inside the expectation, where H_0 is the distribution of $(\mathbf{y}', \mathbf{x}')'$. Suppose that the functional estimates associated to the initial estimates $\widetilde{\mathbf{B}}_n$ and $\widetilde{\mathbf{\Sigma}}_n$ are affine equivariant. Then, the influence function for the functional estimator $\mathbf{T}_1$ corresponding to the MM-estimate $\widehat{\mathbf{B}}_n$ is*

$$IF(\mathbf{z}_0, \mathbf{T}_1, \mathbf{B}_0, \mathbf{\Sigma}_0) = cW\left(\frac{\left((\mathbf{y}_0 - \mathbf{B}_0'\mathbf{x}_0)'\mathbf{\Sigma}_0^{-1}(\mathbf{y}_0 - \mathbf{B}_0'\mathbf{x}_0)\right)^{1/2}}{\sigma_0}\right)\mathbf{\Sigma}_0(\mathbf{y}_0 - \mathbf{B}_0'\mathbf{x}_0)\mathbf{x}_0'E_{G_0}(\mathbf{x}\mathbf{x}')^{-1},$$

where $\sigma_0 = S(H_0)$ and

$$c = \frac{E_{F_0}W'\left((\mathbf{u}'\mathbf{\Sigma}_0\mathbf{u})^{1/2}/\sigma_0\right)(\mathbf{u}'\mathbf{\Sigma}_0\mathbf{u})^{1/2}}{\sigma_0} + E_{F_0}W\left(\frac{(\mathbf{u}'\mathbf{\Sigma}_0\mathbf{u})^{1/2}}{\sigma_0}\right).$$

As in the case of MM-estimators for univariate linear regression, the influence function of the proposed MM-estimate is unbounded.

5. Consistency

We will now show the consistency of MM-estimates for multivariate regression for the case in which errors in (1.1) have an elliptical distribution with an unimodal density. For this, we need the following additional assumptions.

Theorem 5. *Let $(\mathbf{y}_i', \mathbf{x}_i')$, $1 \le i \le n$, be a random sample of the MLM (1.1) with parameters $\mathbf{B}_0$ and $\mathbf{\Sigma}_0$. Assume that ρ_0 and ρ_1 are ρ-functions that satisfy the relation (2.4), that* (A1)–(A3) *hold, and that the initial estimator $\widetilde{\mathbf{B}}_n$ is regression and affine equivariant and $\widetilde{\mathbf{\Sigma}}_n$ is affine equivariant, and that both are consistent for $\mathbf{B}_0$ and $\mathbf{\Gamma}_0$, respectively, where $\mathbf{\Gamma}_0 = \mathbf{\Sigma}_0|\mathbf{\Sigma}_0|^{-1/q}$; then the MM-estimates $\widehat{\mathbf{B}}_n$ and $\widehat{\mathbf{\Sigma}}_n$ satisfy*

(a) $\lim_{n\to\infty}\widehat{\mathbf{B}}_n = \mathbf{B}_0$ *a.s.,*
(b) $\lim_{n\to\infty}\widehat{\mathbf{\Sigma}}_n = \sigma_0^2\mathbf{\Sigma}_0$ *a.s., with σ_0 defined by*

$$E_{F_0}\left(\rho_0\left(\frac{(\mathbf{u}'\mathbf{\Gamma}_0^{-1}\mathbf{u})^{1/2}}{\sigma_0}\right)\right) = b. \tag{5.1}$$

6. Asymptotic normality

Before obtaining the limit distribution of $\widehat{\mathbf{B}}_n$, we need to make some additional assumptions.

(A5) ρ_1 is differentiable, $\psi_1 = \rho_1'$, and $W(u) = \psi_1(u)/u$ is differentiable with bounded derivative.

(A6) $E_{G_0}\|\mathbf{x}\|^4 < \infty$, $E_{G_0}\|\mathbf{x}\|^6 < \infty$, $E_{H_0}\|\mathbf{x}\|^4\|\mathbf{y}\|^2 < \infty$ and $E_{H_0}\|\mathbf{x}\|^2\|\mathbf{y}\|^4 < \infty$, where H_0 is the distribution of $\mathbf{z} = (\mathbf{y}', \mathbf{x}')'$.

(A7) Let $\boldsymbol{\theta} = (\mathbf{B}, \boldsymbol{\Sigma})$ and

$$\phi(\mathbf{z}; \boldsymbol{\theta}) = W\left(d(\mathbf{B}, \boldsymbol{\Sigma})\right) \mathrm{vec}((\mathbf{y} - \mathbf{B}'\mathbf{x})\mathbf{x}'). \tag{6.1}$$

The function $\Phi(\boldsymbol{\theta}) = E_{H_0}\phi(\mathbf{z}; \boldsymbol{\theta})$ has a partial derivative $\partial\Phi/\partial\mathrm{vec}(\mathbf{B}')'$ which is continuous at $\boldsymbol{\theta}_0 = (\mathbf{B}_0, \sigma_0^2\boldsymbol{\Sigma}_0)$, and the matrix

$$\boldsymbol{\Lambda} = \frac{\partial\Phi(\mathbf{B}, \boldsymbol{\Sigma})}{\partial\mathrm{vec}(\mathbf{B}')'}(\mathbf{B}_0, \sigma_0^2\boldsymbol{\Sigma}_0) \tag{6.2}$$

is nonsingular.

Theorem 6. *Let $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')$, with $1 \le i \le n$, be a random sample from model (1.1) with parameters $\mathbf{B}_0$ and $\boldsymbol{\Sigma}_0$. Assume that the ρ-function ρ_1 satisfies (A1), that (A2)–(A7) hold, and that the estimates $\widetilde{\mathbf{B}}_n$ and $\widetilde{\boldsymbol{\Sigma}}_n$ are consistent for $\mathbf{B}_0$ and $\boldsymbol{\Gamma}_0 = \boldsymbol{\Sigma}_0|\boldsymbol{\Sigma}_0|^{-1/q}$, respectively; then $n^{1/2}\mathrm{vec}(\widehat{\mathbf{B}}_n' - \mathbf{B}_0') \xrightarrow{d} N_{qp}(\mathbf{0}, \mathbf{V})$, where $\xrightarrow{d}$ denotes convergence in distribution and*

$$\mathbf{V} = \boldsymbol{\Lambda}^{-1}\mathbf{M}\boldsymbol{\Lambda}^{-1'}, \tag{6.3}$$

where $\mathbf{M}$ is the covariance matrix $\phi(\mathbf{z}_1, (\mathbf{B}_0, \sigma_0^2\boldsymbol{\Sigma}_0))$, with ϕ defined in (6.1), and $\boldsymbol{\Lambda}$ is defined in (6.2).

Assumptions (A4)–(A7) are sufficient to prove Theorem 6, but we conjecture that the limit distribution of $\widehat{\mathbf{B}}_n$ can be proved under less restrictive hypotheses.

Remark 4. Note that the rate of convergence of the MM-estimates depends only on the consistency of, and not on the rate of convergence of, the initial estimates.

Under suitable differentiability conditions, we can obtain a more detailed expression of the covariance matrix $\mathbf{V}$ of Theorem 6.

Proposition 7. *If $W_1(u) = W(\sqrt{u})$ is differentiable with bounded derivative and the initial estimates $(\widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Sigma}}_n)$ are affine equivariant, then*

$$\mathbf{V} = \left[\frac{\sigma_0^2}{q}E_{F_0}\left(\psi_1\left(\frac{v}{\sigma_0}\right)\right)^2 \Big/ \left(E_{F_0}W^*\left(\frac{v}{\sigma_0}\right)\right)^2\right](E_{G_0}\mathbf{x}\mathbf{x}')^{-1} \otimes \boldsymbol{\Sigma}_0, \tag{6.4}$$

where

$$W^*\left(\frac{v}{\sigma_0}\right) = \frac{1}{q\sigma_0^2}W_1'\left(\frac{v^2}{\sigma_0^2}\right)v^2 + W\left(\frac{v}{\sigma_0}\right) \tag{6.5}$$

and

$$v = \left(\mathbf{u}'\boldsymbol{\Sigma}_0^{-1}\mathbf{u}\right)^{1/2}.$$

From the proof of Proposition 7 (see the Appendix of [14]), it is easily seen that, if $W_1(u)$ is continuously differentiable with bounded derivative, assumption (A7) holds if and only if $E_{F_0}W^*\left(\left(\mathbf{u}'\boldsymbol{\Sigma}_0^{-1}\mathbf{u}\right)^{1/2}/\sigma_0\right) \neq 0$.

Remark 5. The covariance matrix of the MLE is given by

$$\mathbf{V} = \left(E_{F_0}(v^2)/q\right)(E_{G_0}\mathbf{x}\mathbf{x}')^{-1} \otimes \boldsymbol{\Sigma}_0.$$

Then the asymptotic relative efficiency (ARE) of the MM-estimate $\widehat{\mathbf{B}}_n$ with respect to the MLE is

$$\mathrm{ARE}(\psi_1, F_0) = E_{F_0}(v^2)\frac{\left(E_{F_0}W^*\left(\frac{v}{\sigma_0}\right)\right)^2}{\sigma_0^2E_{F_0}\left(\psi_1\left(\frac{v}{\sigma_0}\right)\right)^2}. \tag{6.6}$$

As we mentioned in Remark 1, to obtain an MM-estimate which simultaneously has a high breakdown point and high efficiency under normal errors, it suffices to choose c_0 and c_1 in (2.8) appropriately. The constant c_0 can be chosen so that

$$E\left(\rho\left(\frac{(\mathbf{u}'\boldsymbol{\Gamma}_0^{-1}\mathbf{u})^{1/2}}{c_0}\right)\right) = b, \tag{6.7}$$

Table 1
Values of c_0 for the bisquare function.

q	1	2	3	4	5	10
c_0	1.56	2.66	3.45	4.10	4.65	6.77

Table 2
Values of c_1 for the bisquare function to attain given values of the asymptotic relative efficiency (ARE) under normal errors.

ARE	q					
	1	2	3	4	5	10
0.80	3.14	3.51	3.82	4.10	4.34	5.39
0.90	3.88	4.28	4.62	4.91	5.18	6.38
0.95	4.68	5.12	5.48	5.76	6.10	7.67

where $\mathbf{u}$ is $N_q(\mathbf{0}, \mathbf{\Sigma}_0)$, $\mathbf{\Sigma}_0 = |\mathbf{\Sigma}_0|^{1/q}\mathbf{\Gamma}_0$, and $b = 0.5$; this ensures a high breakdown point and that the asymptotic relative efficiency (6.6) depends only on c_1. Then, c_1 can be chosen so that the MM-estimate has the desired efficiency without affecting the breakdown point, which depends only on c_0.

Table 1 gives the values of c_0 verifying (6.7) for different values of q. Table 2 gives the values of c_1 needed to attain different levels of asymptotic efficiency. In both cases the function ρ from (2.8) is equal to the bisquare function, ρ_B, given in (2.2).

7. Computing algorithm

In this section, we propose an iterative algorithm to compute $\widehat{\mathbf{B}}_n$ and $\widehat{\mathbf{\Sigma}}_n$ based on Remark 2. Let $\mathbf{z}_i = (\mathbf{y}_i', \mathbf{x}_i')$ be a sample of size n, and assume that we have computed the initial estimates $\widetilde{\mathbf{B}}_n$ and $\widetilde{\mathbf{\Sigma}}_n$ with high breakdown point and such that $|\widetilde{\mathbf{\Sigma}}_n| = 1$.

1. Using the initial values $\widetilde{\mathbf{B}}^{(0)} = \widetilde{\mathbf{B}}_n$ and $\widetilde{\boldsymbol{\Gamma}}^{(0)} = \widetilde{\mathbf{\Sigma}}_n$, compute the M-estimate of scale $\hat{\sigma}_n := s(\mathbf{d}(\widetilde{\mathbf{B}}^{(0)}, \widetilde{\boldsymbol{\Gamma}}^{(0)}))$, defined by (2.1), using a function ρ_0 as in the definition and $b = 0.5$ and the matrix $\widetilde{\mathbf{\Sigma}}^{(0)} = \hat{\sigma}_n^2\widetilde{\boldsymbol{\Gamma}}^{(0)}$.
2. Compute the weights $\omega_{i0} = W\left(d_i(\widetilde{\mathbf{B}}^{(0)}, \widetilde{\mathbf{\Sigma}}^{(0)})\right)$ for $1 \le i \le n$. These weights are used to compute each column of $\widetilde{\mathbf{B}}^{(1)}$ separately by weighted least squares.
3. Compute the matrix
$$\widetilde{\mathbf{C}}^{(1)} = \sum_{i=1}^{n} \omega_{i0}\widehat{\mathbf{u}}_i(\widetilde{\mathbf{B}}^{(1)})\widehat{\mathbf{u}}_i'(\widetilde{\mathbf{B}}^{(1)}),$$
and with it the matrix $\widetilde{\mathbf{\Sigma}}^{(1)} = \hat{\sigma}_n^2\widetilde{\mathbf{C}}^{(1)}/|\widetilde{\mathbf{C}}^{(1)}|^{1/q}$.
4. Suppose that we have already computed $\widetilde{\mathbf{B}}^{(k-1)}$ and $\widetilde{\mathbf{\Sigma}}^{(k-1)}$. Then $\widetilde{\mathbf{B}}^{(k)}$ and $\widetilde{\mathbf{\Sigma}}^{(k)}$ are computed using steps 2 and 3, but starting from $\widetilde{\mathbf{B}}^{(k-1)}$ and $\widetilde{\mathbf{\Sigma}}^{(k-1)}$ instead of $\widetilde{\mathbf{B}}^{(0)}$ and $\widetilde{\mathbf{\Sigma}}^{(0)}$.
5. The procedure is stopped at step k if the relative absolute differences of all elements of the matrices $\widetilde{\mathbf{B}}^{(k)}$ and $\widetilde{\mathbf{B}}^{(k-1)}$ and the relative absolute differences of all the Mahalanobis norms of residuals $\widehat{\mathbf{u}}_i(\widetilde{\mathbf{B}}^{(k)})$ and $\widehat{\mathbf{u}}_i(\widetilde{\mathbf{B}}^{(k-1)})$ with respect to $\widetilde{\mathbf{\Sigma}}^{(k)}$ and $\widetilde{\mathbf{\Sigma}}^{(k-1)}$, respectively, are smaller than a given value δ.

The following theorem shows that the iterative procedure to compute MM-estimates yields the descent of the objective function.

Theorem 8. *If $W(u)$ is nonincreasing in $|u|$, then at each iteration of the algorithm the function $\sum_{i=1}^{n} \rho_1(d_i(\mathbf{B}, \mathbf{\Sigma}))$ is nonincreasing.*

8. Simulation

8.1. Simulation design

To investigate the performance of the proposed estimates we performed a simulation study.

- We consider the MLM given by (1.1) for two cases: $p = 2, q = 2$ and $p = 2, q = 5$. Due to the equivariance of the estimators we take, without loss of generality, $\mathbf{B}_0 = \mathbf{0}$ and $\mathbf{\Sigma}_0 = \mathbf{I}_q$. The errors $\mathbf{u}_i$ are generated from an $N_q(\mathbf{0}, \mathbf{I})$ distribution and the predictors $\mathbf{x}_i$ from an $N_p(\mathbf{0}, \mathbf{I})$ distribution.
- The sample size is 100 and the number of replications is 1000. We consider uncontaminated samples and samples that contain 10% of identical outliers of the form $(\mathbf{x}_0, \mathbf{y}_0)$ with $\mathbf{x}_0 = (x_0, 0, \ldots, 0)$ and $\mathbf{y}_0 = (mx_0, 0, \ldots, 0)$. The values of x_0

considered are 1 (low-leverage outliers) and 10 (high-leverage outliers). We take a grid of values of m, starting at 0. The grid was chosen in order that all robust estimates attain the maximum values of their error measure.

- Let $\widehat{\mathbf{B}}^{(k)}$ be the estimate of $\mathbf{B}_0$ obtained in the kth replication. Then, since we are taking $\mathbf{B}_0 = \mathbf{0}$, the estimate of the mean squared error (MSE) is given by

$$\mathrm{MSE} = \frac{1}{1000}\left(\sum_{k=1}^{1000}\sum_{i=1}^{p}\sum_{j=1}^{q}\left(\widetilde{\mathbf{B}}_{ij}^{(k)}\right)^2\right).$$

It must be recalled that the distributions of robust estimates under contamination are themselves heavy tailed, and it is therefore prudent to evaluate their performance through robust measures (see [10, Sec. 1.4, p. 12] and [9, p.75]). For this reason, we employed both the MSE, and the trimmed mean squared error (TMSE), which compute the 10% (upper) trimmed average of

$$\left\{\sum_{i=1}^{p}\sum_{j=1}^{q}\left(\widetilde{\mathbf{B}}_{ij}^{(k)}\right)^2\right\}_{k=1}^{1000}.$$

The results given below correspond to this MSE, although the TMSE yields qualitatively similar results (in the uncontaminated case the results are the same).

8.2. Description of the estimators

For each case, four estimates are computed: the MLE, an S-estimate, a τ-estimate, and an MM-estimate.

For the MLM, the S-estimates are defined by

$$(\widehat{\mathbf{B}}, \widehat{\boldsymbol{\Sigma}}) = \arg\min\{|\boldsymbol{\Sigma}| : (\mathbf{B}, \boldsymbol{\Sigma}) \in \mathbb{R}^{p\times q} \times \mathcal{S}_q\}$$

subject to

$$s^2(d_1(\mathbf{B}, \boldsymbol{\Sigma}), \ldots, d_n(\mathbf{B}, \boldsymbol{\Sigma})) = q,$$

where s is an M-estimate of scale.

García Ben et al. [7] extended τ-estimates to the MLM by defining

$$(\widehat{\mathbf{B}}, \widehat{\boldsymbol{\Sigma}}) = \arg\min\{|\boldsymbol{\Sigma}| : (\mathbf{B}, \boldsymbol{\Sigma}) \in \mathbb{R}^{p\times q} \times \mathcal{S}_q\}$$

subject to

$$\tau^2(d_1(\mathbf{B}, \boldsymbol{\Sigma}), \ldots, d_n(\mathbf{B}, \boldsymbol{\Sigma})) = \kappa, \tag{8.1}$$

where the τ-scale is defined by

$$\tau^2(\mathbf{v}) = (s^2(\mathbf{v})/n)\sum_{i=1}^{n}\rho_2\left(|v_i|/s(\mathbf{v})\right), \tag{8.2}$$

where $\mathbf{v} = (v_1, \ldots, v_n)$, ρ is a ρ-function, and s is an M-estimate of scale.

The robust estimates are based on bisquare ρ-functions. The M-estimate of scale used in the S-estimate is defined by $\rho_0(u) = \rho_B(u/c_0)$, and $b = 0.5$ so that the S-estimate has breakdown point 0.5 (see Table 1). The τ-estimate uses the same ρ_0 and b as the S-estimate to compute the M-scale and $\rho_2(u) = \rho_B(u/c_2)$, where c_2 is chosen together with the constant κ, from Eq. (8.1), so that the τ-estimate has an ARE equal to 0.90 when the errors are Gaussian (see Table 2 in [7] in which $\kappa = 6\kappa_2/c_2^2$). The initial estimate needed to compute the τ-estimate is computed using 2000 subsamples. The MM-estimate uses the same ρ_0 as the S-estimate to compute the M-estimate of scale and $\rho_1(u) = \rho_B(u/c_1)$, where c_1 is chosen so that the MM-estimate has an ARE equal to 0.90 when the errors are Gaussian (see Table 2). We use the S-estimates as $(\widetilde{\mathbf{B}}_n, \widetilde{\boldsymbol{\Sigma}}_n)$, and the value of δ in step 5 of the computing algorithm is taken equal to 10^{-4}.

8.3. Results

Table 3 displays the mean squared errors, the standard errors, and the relative efficiencies and asymptotic relative efficiencies with respect to the MLE for the uncontaminated case. It is seen that the relative efficiencies of all robust estimates (computed as the ratio of their respective MSEs and the MSE of the MLE) are close to their asymptotic values. The τ-estimate and MM-estimate have similar high efficiencies, and both outperform the S-estimator.

In Figs. 1–4, we show the MSEs of the different estimates under contamination.

In Fig. 1, which corresponds to $q = 2$ and $x_0 = 1$, we observe that the MM-estimate and the τ-estimate behave similarly, both having a smaller MSE than the S-estimate except when m is (approximately) between 2.8 and 4. In this case, the S-estimate has the largest maximum MSE among the robust estimates. As expected, the MSE of the MLE increases

Table 3
Simulation: mean squared error (MSE), standard error of the MSE (SE), relative efficiency (REFF), and asymptotic relative efficiency (ARE) of the estimates in the uncontaminated case for $n = 100$ and $p = 2$.

Estimate	$q = 2$				$q = 5$			
	MSE	SE	REFF	ARE	MSE	SE	REFF	ARE
MLE	0.041	0.001	1.00	1.00	0.103	0.002	1.00	1.00
S-estimate	0.074	0.002	0.55	0.58	0.125	0.002	0.83	0.85
τ-estimate	0.046	0.001	0.89	0.90	0.116	0.002	0.90	0.90
MM-estimate	0.046	0.001	0.89	0.90	0.116	0.002	0.90	0.90

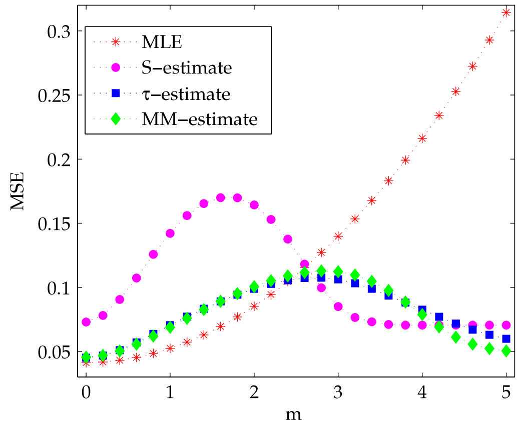

Fig. 1. Simulation: mean squared errors for $q = 2$ and $x_0 = 1$.

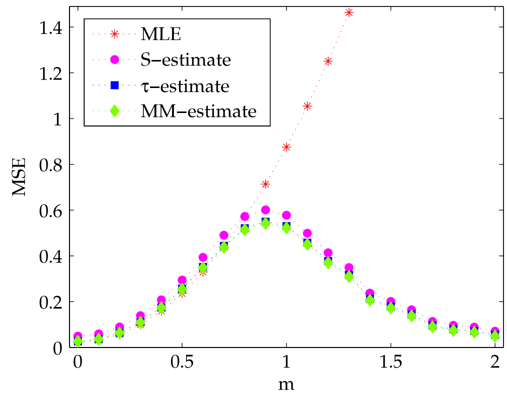

Fig. 2. Simulation: mean squared errors for $q = 2$ and $x_0 = 10$.

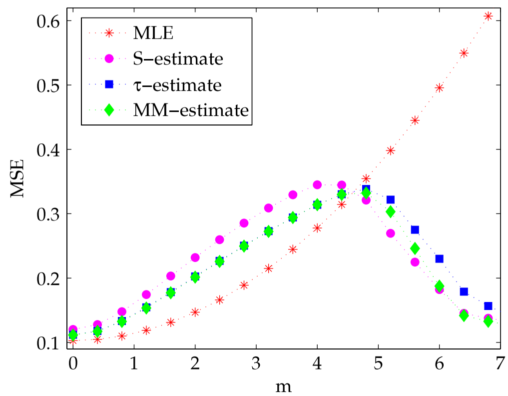

Fig. 3. Simulation: mean squared errors for $q = 5$ and $x_0 = 1$.

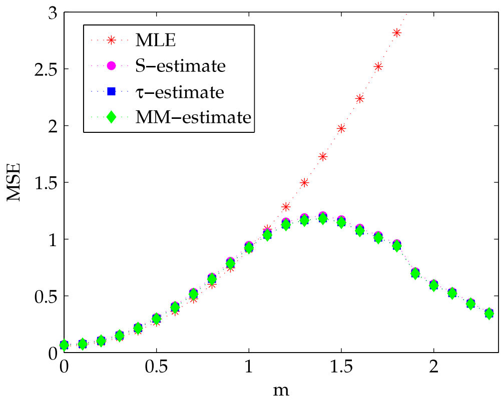

Fig. 4. Simulation: mean squared errors for $q = 5$ and $x_0 = 10$.

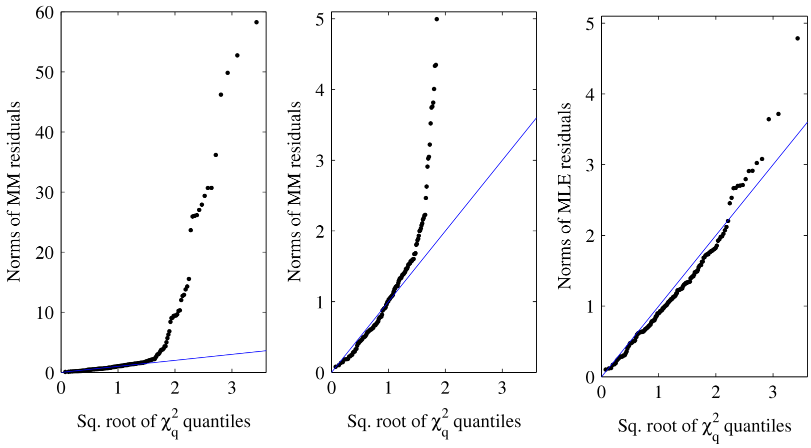

Fig. 5. QQ-plots of the Mahalanobis norms of the residuals of the MM-estimate (left), the MLE (right), and the MM-estimate in the same interval as the MLE (center).

without bound for large m. Fig. 2 shows the results for $q = 2$ and $x_0 = 10$. The S-estimate, the τ-estimate, and the MM-estimate behave similarly. In Fig. 3, which corresponds to $q = 5$ and $x_0 = 1$, the three robust estimates are seen to follow essentially the same pattern. For $m \leq 4.8$ (approximately), the τ-estimate and the MM-estimate have similar behaviors, both outperforming the S-estimate. For $m > 4.8$, the S-estimate and the MM-estimate have similar behaviors, both outperforming the τ-estimate. For $q = 5$ and $x_0 = 10$ (Fig. 4), the behavior of the robust estimates is similar to that observed for $q = 2$ and $x_0 = 10$ (Fig. 2).

In general, we see that the proposed MM-estimator competes favorably with the MLE, the τ-estimate, and the S-estimate with respect to both efficiency and robustness.

9. An example with real data

In this section, we analyze a dataset corresponding to electron-probe X-ray microanalysis of archeological glass vessels ([11]). For each of $n = 180$ vessels we have a spectrum on a set of equispaced frequencies numbered between 1 and 1920 and the contents of 13 chemical compounds; the purpose is to predict the contents on the basis of the spectra. In order to limit the size of our dataset, we considered only two compounds (responses): P_2O_5 and PbO; and we chose 12 equispaced frequencies between 100 and 400. This interval was chosen because the values of x_{ij} are almost null for frequencies below 100 and above 400. We therefore have $p = 13$ and $q = 2$.

We considered two multivariate regression estimates: the MLE and our MM-estimate. As initial estimate for the MM-estimate we use an S-estimate. The S-estimate and the MM-estimate employ bisquare ρ-functions with constants such that the MM-estimate has Gaussian ARE equal to 0.95 and the S-estimate has breakdown point 0.5. In Fig. 5, we present QQ-plots of the Mahalanobis norms of the residuals of the MLE and the MM-estimate against the root quantiles of the chi-squared distribution with q degrees of freedom. The QQ-plot of the MM-estimate shows clear outliers.

In Fig. 6 we compare the sorted absolute values of the residuals of the MLE with those corresponding to the MM-estimator for each component of the response.

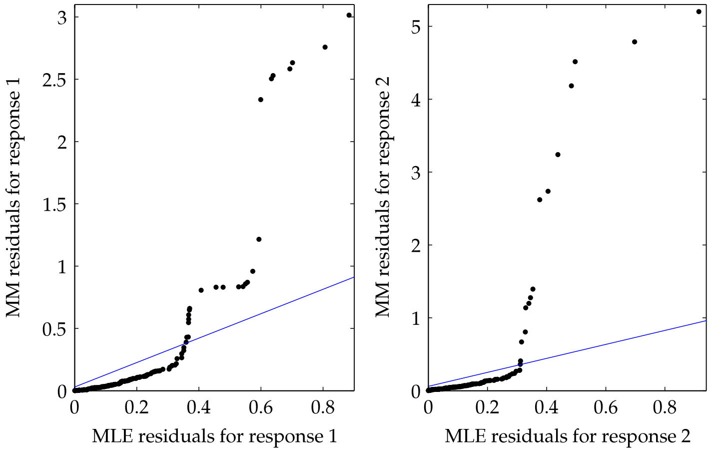

Fig. 6. QQ-plots of sorted absolute residuals of MM-estimates versus sorted absolute residuals of the MLE for each component of the response. The left plot corresponds to P_2O_5(the first component) and the right to PbO (the second component).

Table 4
MLE and MM-estimate of the covariance matrix of the errors.

MLE	MM-estimate
$\begin{pmatrix} 0.0645 & -0.0008 \\ -0.0008 & 0.0348 \end{pmatrix}$	$\begin{pmatrix} 0.0102 & -0.0014 \\ -0.0014 & 0.0084 \end{pmatrix}$

The right and left panels of Fig. 5 show respectively the QQ-plots of the Mahalanobis norms of the residuals of the MLE and the MM-estimate against the square root quantiles of the chi-squared distribution with q degrees of freedom. For ease of comparison, the center panel shows the MM-estimate's QQ-plot truncated to the size of the MLE's QQ-plot. The latter shows a very good fit of the norms to the chi-squared distribution, and therefore points out no suspect points, while the MM-estimate's QQ-plot indicates some 30 possible outliers, i.e., about 16% of the data.

The MLE's norms are in general smaller than the MM-estimate's norms, but this does not mean that the former gives a better fit, since here the residuals are normalized by the respective estimated residual dispersion matrices $\widehat{\mathbf{\Sigma}}_0$. Fig. 6 compares the sorted absolute values of the (univariate) residuals of the MLE with those of the MM-estimate for each response. We can see that the majority of the residuals corresponding to the MM-estimate are smaller than those of the MLE.

To understand why the MLE's norms are in general smaller than the MM-estimate's norms, while the respective residuals are in general smaller, we show in Table 4 the estimates given by the MLE and MM-estimate of the dispersion matrix of the errors. It is seen that the former is "much larger" than the latter, in that its two diagonal elements are respectively six and four times those of the latter.

To complete the description of the estimates' fit, Fig. 7 shows the absolute values of the coordinates of the bidimensional residual vectors; the right panel (corresponding to residuals of the MM-estimate) is truncated to the size of the left panel (corresponding to residuals of the MLE), and consequently 10% of the absolute residuals of the MM-estimate are not shown. It is seen that, while the residuals from MM-estimate have a larger range than those from the MLE, they are in general more concentrated near the origin. In general, we may conclude that the MM-estimate gives a good fit to the bulk of the data, at the expense of misfitting a reduced proportion of atypical points, while the MLE tries to fit all data points, including the atypical ones, with the cost of a poor fit to the bulk of the data.

We compared the predictive behaviors of the MLEs and the MM-estimates through five-fold cross-validation. We also included the univariate MM-estimates corresponding to each component of the response and the τ-estimate proposed by García Ben et al. [7]. As initial estimate for the univariate MM-estimates we use S-estimates. The τ-estimate, the S-estimate, and the univariate MM-estimate employ bisquare ρ-functions with constants such that the univariate MM-estimate and the τ-estimate have Gaussian AREs equal to 0.95 and the S-estimate has breakdown point 0.5. We considered two evaluation criteria: the mean squared error (MSE) and a robust criterion, namely a τ-scale (8.2) of the predictive errors, both computed separately for each component of the response. In the τ-scale, s is an M-scale with breakdown point 0.5 and ρ_2 is a bisquare ρ-function with constant such that the τ-scale has Gaussian asymptotic efficiency equal to 0.85.

Table 5 shows the results. According to the MSE, the MLE is much better than the robust estimates. However, the τ-scales yield the opposite conclusion. The reason of this fact is the MSE's sensitivity to outliers. This result shows how misleading a nonrobust criterion may be. According to the τ-scale, the predictive performance of our MM-estimate for the second component is slightly better than that of the τ-estimate, while the opposite occurs for the first component. The results obtained with the univariate MM-estimates are similar to those of the multivariate MM-estimate.

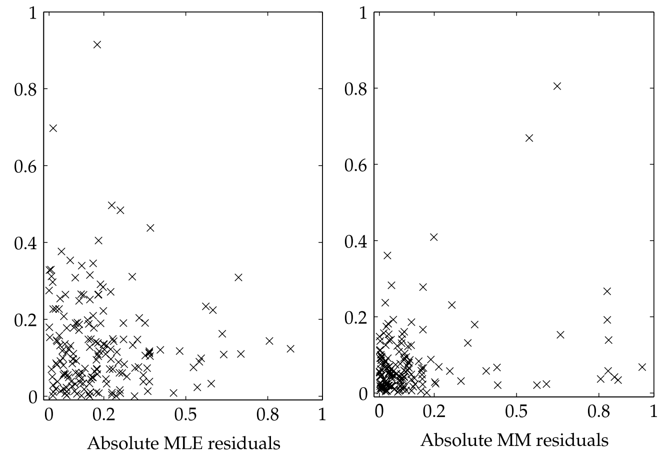

Fig. 7. Absolute values of the coordinates of the bidimensional residual vectors corresponding to the MLE (left) and to the MM-estimate (right).

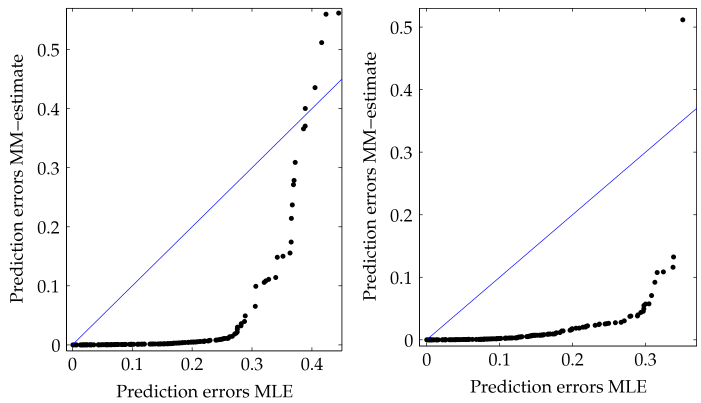

Fig. 8. QQ-plots of sorted absolute prediction errors of MM-estimates versus sorted absolute prediction errors of MLE for each component of the response, computed by cross-validation. The left plot corresponds to P_2O_5 (the first component) and the right plot to PbO (the second component).

Table 5
Mean square error (MSE) and τ-scale of the prediction errors of the MLE, multivariate MM-estimate, τ-estimate, and univariate MM-estimate for each component of the response, computed by cross-validation.

Criterion	MLE		τ-estimate		MM-estimate		MM-univariate	
Component	1	2	1	2	1	2	1	2
MSE	0.081	0.051	0.351	0.806	0.340	0.682	0.354	0.762
τ-scale	0.044	0.022	0.005	0.007	0.008	0.006	0.005	0.006

The QQ-plots in Fig. 8 compare for each response component the absolute values of the sorted cross-validation prediction errors of our MM-estimate with those of the MLE. For reasons of scale, in each QQ-plot the observations with the 12 largest absolute prediction errors were omitted. We can see that most points lie below the identity line representing the identity function, showing that the MM-estimate provides a better prediction for the bulk of the data.

10. Conclusions

In this paper, we have presented MM-estimates for the multivariate linear model and showed that they maintain the same good theoretical properties as in the univariate case, such as a high breakdown point and high Gaussian efficiency. The simulation study indicates that it has the desired high efficiency, and that its behavior is in general similar to, and in several situations superior to, that of the τ-estimate; it is also more efficient, and in most situations more robust, than the S-estimate. In the example with real data, our MM-estimate gives a good fit to the bulk of the data, pointing out the existence of atypical points, and shows a good predictive behavior.

Acknowledgments

We would like to thank the referees and the editor for their helpful comments and suggestions. This research was partially supported by grants PIP 216 from CONICET and PICT 899 from ANPCyT, Argentina.

References

[1] J. Agulló, C. Croux, S. Van Aelst, The multivariate least trimmed squares estimator, Journal of Multivariate Analysis 99 (2008) 311–338.
[2] Z. Bai, X.R. Chen, Y. Wu, On constrained M-estimation and its recursive analog in multivariate linear regression models, Statistica Sinica 18 (2008) 405–424.
[3] J.R. Berrendero, B.V.M. Mendez, D.E. Tyler, On the maximum bias functions of MM-estimates and constrained M-estimates of regression, The Annals of Statistics 35 (2007) 13–40.
[4] M. Bilodeau, P. Duchesne, Robust estimation of the SUR model, The Canadian Journal of Statistics 28 (2000) 227–288.
[5] C. Croux, P. Rousseeuw, O. Hössjer, Generalized S-estimators, Journal of the American Statistical Association 89 (1994) 1271–1281.
[6] P. Davies, Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices, The Annals of Statistics 15 (1987) 1269–1292.
[7] M.G. García Ben, E. Martinez, V.J. Yohai, Robust estimation for the multivariate linear model based on a τ-scale, Journal of Multivariate Analysis 97 (2006) 1600–1622.
[8] Frank R. Hampel, The influence curve and its role in robust estimation, Journal of the American Statistical Association. 69 (1974) 383–393.
[9] P.J. Huber, Robust Estimation of a Location Parameter, The Annals of Mathematical Statistics 35 (1) (1964) 73–101.
[10] P.J. Huber, Robust Statistics, Wiley, New York, 1981.
[11] K. Janssens, I. Deraedt, A. Freddy, J. Veekman, Composition of 15–17th century archeological glass vessels excavated in Antwerp, Belgium. Mikrochimica Acta 15 (Suppl.) (1998) 253–267.
[12] J.T. Kent, D.E. Tyler, Constrained M-estimation for multivariate location and scatter, The Annals of Statistics 24 (1997) 1346–1370.
[13] P. Koenker, S. Portnoy, M-estimation of multivariate regressions, Journal of the American Statistical Association. 85 (1990) 1060–1068.
[14] N. Kudraszow, R. Maronna, Estimates of MM type for the multivariate linear model. Technical report, 2010. Available at arxiv:1004.4883.
[15] H. Lopuhaä, Highly efficient estimators of multivariate location with high breakdown point, The Annals of Statistics 20 (1992) 398–413.
[16] R.A. Maronna, R.D. Martin, V.J. Yohai, Robust Statistics: Theory and Methods, John Wiley and Sons, New York, 2006.
[17] B. Mendes, D.E. Tyler, Constrained M estimates for regression, in: H. Rieder (Ed.), Robust Statistics: Data Analysis and Computer Intensive Methods, in: Lecture Notes in Statist, vol. 109, Springer, New York, 1996, pp. 299–320.
[18] E. Roelandt, S. Van Aelst, C. Croux, Multivariate Generalized S-estimators, Journal of Multivariate Analysis 100 (5) (2009) 876–887.
[19] P. Rousseeuw, Multivariate estimators with high breakdown point, in: W. Grossman, G. Pflug, I. Vincza, W. Wertz (Eds.), Mathematical Statistics and its Applications, vol. B, Reidel, Dordrecht, The Netherlands, 1985, pp. 283–297.
[20] P. Rousseeuw, K. Van Driessen, S. Van Aelst, J. Agulló, Robust multivariate regression, Technometrics 46 (2004) 293–305.
[21] M. Salibián-Barrera, S. Van Aelst, G. Willems, PCA based on multivariate MM-estimators with fast and robust bootstrap, Journal of the American Statistical Association 101 (2006) 1198–1211.
[22] K.S. Tatsuoka, D.E. Tyler, The uniqueness of S and M-functionals under non-elliptical distributions, The Annals of Statistics 28 (2000) 1219–1243.
[23] S. Van Aelst, G. Willems, Multivariate regression S-estimators for robust estimation and inference, Statistica Sinica 15 (2005) 981–1001.
[24] V. Yohai, High Breakdown-point and high efficiency robust estimates for regression, The Annals of Statistics 15 (2) (1987) 642–656.