Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Multivariate Analysis 97 (2006) 1600–1622
Journal of Multivariate Analysis
www.elsevier.com/locate/jmva

Robust estimation for the multivariate linear model based on a τ-scale

Marta García Ben[a], Elena Martínez[a], Víctor J. Yohai[b,*]

[a] *Universidad de Buenos Aires, Argentina*
[b] *CONICET, Universidad de Buenos Aires, Argentina*

Received 26 September 2004
Available online 3 October 2005

Abstract

We introduce a class of robust estimates for multivariate linear models. The regression coefficients and the covariance matrix of the errors are estimated simultaneously by minimizing the determinant of the covariance matrix estimate, subject to a constraint on a robust scale of the Mahalanobis norms of the residuals. By choosing a τ-estimate as a robust scale, the resulting estimates combine good robustness properties and asymptotic efficiency under Gaussian errors. These estimates are asymptotically normal and in the case where the errors have an elliptical distribution, their asymptotic covariance matrix differs only by a scalar factor from the one corresponding to the maximum likelihood estimate. We derive the influence curve and prove that the breakdown point is close to 0.5. A Monte Carlo study shows that our estimates compare favorably with respect to S-estimates.

AMS 1991 subject classification: 62F35; 62H12; 62J05

Keywords: Multivariate regression; Robust estimation; τ-estimates

1. Introduction

Let $y_i = (y_{i1}, \ldots, y_{iq})'$ and $x_i = (x_{i1}, \ldots, x_{ip})'$, $1 \leqslant i \leqslant n$, be the response and predictor vectors satisfying the multivariate linear model (MLM)

$$y_i = B_0' x_i + u_i, \tag{1.1}$$

* Corresponding author.
E-mail address: vyohai@uolsinectis.com.ar (V.J. Yohai).

0047-259X/$ - see front matter
doi:10.1016/j.jmva.2005.08.007

where B_0 is a $p \times q$ matrix and $u_1, \ldots, u_n$ are i.i.d. q-dimensional vectors. The x_i's may be fixed or random, and in the latter case we assume that they are i.i.d. and independent of the u_i's. If the model includes an intercept, then $x_{ip} = 1$. We denote the distributions of u_i and x_i by F_0 and by M_0, respectively.

Assuming that the errors u_i's have a multivariate normal distribution with mean 0 and covariance matrix $\boldsymbol{\Sigma}_0(N_q(0, \Sigma_0))$, the maximum likelihood estimator (MLE) of B_0 can be obtained by computing the least squares estimate (LSE) for each component of y separately, and the MLE of Σ_0 is the sample covariance matrix of the corresponding residuals. As it is well known, the LSE is extremely sensitive to outliers. In fact, just one observation may have an unbounded effect on this estimate.

The first proposal of a robust estimate for the MLM was given by Koenker and Portnoy [10]. They proposed to apply a regression M-estimator based on a convex ρ-function to each coordinate of the response vector. This proposal has two disadvantages: lack of affine equivariance and a null breakdown point. The second problem may be overcome by replacing the M-estimator by a high breakdown point estimate, but we would still lack affine equivariance.

Recently, several robust equivariant estimates for the MLM have been proposed. Rousseeuw et al. [15] proposed estimates for the MLM based on a robust estimate of the covariance matrix of $z = (x', y')'$. A different approach based on extending estimates of multivariate location and scatter was followed by Bilodeau and Duchesne [2] who extended the S-estimates introduced by Davies [4], and by Agulló et al. [1] who extended the minimum covariance determinant estimate introduced by Rousseeuw [13].

All these estimates have a high breakdown point and therefore a good robustness behavior. However, they are not highly efficient when the errors are Gaussian and q is small. 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.

In this paper we propose robust estimates for the MLM by extending the τ-estimates of multivariate location and scatter proposed by Lopuhaä [11]. We show that these estimates simultaneously have a high breakdown point and a high efficiency under Gaussian errors.

In Section 2 we define τ-estimates for MLM. In Section 3 we study their breakdown point and in Section 4 we derive the influence curve. In Section 5 we study the asymptotic properties (consistency and asymptotic normality) of the τ-estimates assuming random regressors and errors with an elliptical unimodal distribution. In Section 6 we describe a computing algorithm based on an iterative weighted MLE. In Section 7 we present the results of a Monte Carlo study and a real example in Section 8. In the Appendix we derive some mathematical results.

2. Estimates based on a robust scale for the MLM

Let (x_i, y_i), $1 \leqslant i \leqslant n$, satisfy the MLM (1.1), where the u_i are i.i.d. random vectors. Let X be the $n \times p$ matrix whose ith row is x_i' and Y the $n \times q$ matrix whose ith row is y_i'. In the rest of the paper we will assume that the rank of X is p. For any $p \times q$ matrix B define the residuals $\widehat{u}_i(B)$ as

$$\widehat{u}_i(B) = y_i - B'x_i, \quad 1 \leqslant i \leqslant n. \tag{2.1}$$

The MLE of B_0 and Σ_0 when the distribution of the u_i's is multivariate normal with mean 0 and covariance matrix Σ_0 are given by

$$\widehat{B} = (X'X)^{-1}X'Y,$$

and

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n} \widehat{u}_i(\widehat{B})\widehat{u}_i'(\widehat{B}). \tag{2.2}$$

Let $d_i(B, \Sigma)$ be the Mahalanobis norms of the residuals, given by

$$d_i(B, \Sigma) = (\widehat{u}_i'(B)\Sigma^{-1}\widehat{u}_i(B))^{1/2}, \quad 1 \leqslant i \leqslant n.$$

If the $n \times (p+q)$ matrix (X, Y) has rank $(p+q)$ then $\det(\widehat{\Sigma}) \neq 0$. In this case the MLE of B and Σ also satisfy

$$(\widehat{B}, \widehat{\Sigma}) = \arg\min_{B,\Sigma} \det(\Sigma) \tag{2.3}$$

subject to

$$\frac{1}{n}\sum_{i=1}^{n} d_i^2(B, \Sigma) = q. \tag{2.4}$$

This follows from two facts: (i) the MLE of (B_0, Σ_0) minimizes $\left(n \log|\Sigma| + \sum_{i=1}^{n} d_i^2(B, \Sigma)\right)$ and (ii) (2.2) implies that the MLE satisfies (2.4).

Given a sample $v_1, \ldots, v_n$, let $s(v_1, \ldots, v_n)$ be the scale estimate defined as the square root of the mean squared error (RMSE), that is

$$s(v_1, \ldots, v_n) = \left(\frac{1}{n}\sum_{i=1}^{n} v_i^2\right)^{1/2}. \tag{2.5}$$

Then constraint (2.4) can also be expressed as

$$s^2(d_1(B, \Sigma), \ldots, d_n(B, \Sigma)) = q. \tag{2.6}$$

One way of defining robust estimates of B_0 and of a scatter matrix Σ_0 of the u_i's is by (2.3) and (2.6), but replacing the scale s defined in (2.5) by a robust scale.

Huber [9] introduced the M-estimates of scale. For a sample $v = (v_1, \ldots, v_n)$, an M-estimate of scale $s(v_1, \ldots, v_n)$ is defined by the value s satisfying

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{|v_i|}{s}\right) = \kappa. \tag{2.7}$$

When $v_1, \ldots, v_n$ is a random sample of any arbitrary distribution H, $s(v_1, \ldots, v_n)$ converges to s_0 defined by $\kappa = E_H(\rho(|v|/s_0))$. If we wish to calibrate the M-scale so that $s_0 = 1$, we should take $\kappa = E_H(\rho(|v|))$.

The function ρ should satisfy the following properties:

(A1) $\rho(0) = 0$.

(A2) $0 \leqslant v \leqslant v^*$ implies $\rho(v) \leqslant \rho(v^*)$.
(A3) ρ is continuous.
(A4) $0 < A = \sup_v \rho(v) < \infty$.
(A5) If $\rho(u) < A$ and $0 \leqslant u < v$, then $\rho(u) < \rho(v)$.

One measure of the degree of robustness of an estimate is the breakdown point introduced by Hampel [6]. Roughly speaking, the breakdown point of an estimate is the smallest fraction of outliers that is required to take the estimate to an extreme value. In the case of scale estimates, the two possible extreme values are zero or infinity ("implosion" and "explosion" breakdown, respectively). Huber [9] proved that the breakdown point to infinity of an M-estimate of scale is $\varepsilon_\infty^* = \kappa/A$ and the breakdown point to zero is $\varepsilon_0^* = 1 - \kappa/A$. Then, the asymptotic breakdown point of this scale estimate is $\varepsilon^* = \min(\kappa/A, 1 - \kappa/A)$, and taking $\kappa/A = 0.5$ we obtain the highest possible breakdown point, $\varepsilon^* = 0.5$.

Bilodeau and Duchesne [2] proposed a class of robust estimates, called S-estimates, for the seemingly unrelated equations (SUE) model. The SUE model is more general than the MLM, since it allows different regressor vectors for each of the equations. When applied to the MLM, the S-estimates are defined by (2.3) and (2.6) where s is an M-estimate of scale.

Croux [3] proved that M-estimates of scale can combine a high breakdown point with high efficiency under normality. However, Hossjer [8] showed that estimates of regression based on an M-scale cannot combine both properties. To overcome this limitation, Yohai and Zamar [17] proposed τ-estimates of scale, defined as follows: consider two functions ρ_1 and ρ_2 satisfying A1–A5 and an arbitrary distribution H, and put

$$\kappa_i = E_H(\rho_i(|v|)), \quad i = 1, 2. \tag{2.8}$$

Let $s(v)$ be the M-estimate of scale defined in (2.7) with $\rho = \rho_1$ and $\kappa = \kappa_1$. Below we specify which H will be used to define κ_1. The τ-estimate of the scale of $v = (v_1, \ldots, v_n)$ is defined by

$$\tau^2(v) = s^2(v) \frac{1}{n} \sum_{i=1}^{n} \rho_2 \left(\frac{|v_i|}{s(v)} \right). \tag{2.9}$$

The estimate $\tau(v)$ converges to κ_2 when the v_i are independent variables with distribution H.

Yohai and Zamar [17] defined τ-estimates of univariate regression by the minimization of a τ-scale of the residuals. Put $\psi_i(v) = \rho_i'(v)$, $i = 1, 2$. To guarantee the Fisher consistency of the τ-estimates of regression, it is required that ρ_2 satisfy the following condition:

(A6) ρ_2 is continuously differentiable and $2\rho_2(v) - \psi_2(v)v > 0$ for $v > 0$.

Yohai and Zamar [17] also showed that by properly choosing ρ_1 and ρ_2, the τ-estimates of regression combine breakdown point 0.5 with a high Gaussian asymptotic efficiency.

Here we extend τ-estimates to the MLM model by defining

$$\left(\widetilde{B}, \widetilde{\Sigma}\right) = \arg \min_{B,\Sigma} \det(\Sigma) \tag{2.10}$$

subject to

$$\tau^2(d_1(B, \Sigma), \ldots, d_n(B, \Sigma)) = \kappa_2, \tag{2.11}$$

where the scale τ is defined in (2.9). The values of κ_1, κ_2 are chosen using (2.8) with $H = H_{00}$, the distribution of $(u_i' \Sigma_0^{-1} u_i)^{1/2}$ under a nominal distribution. Let H_0 be the distribution of

$(u_i'\Sigma_0^{-1}u_i)^{1/2}$. Since H_0 is unknown, H_{00} may not coincide with H_0. To prevent H_{00} from being dependent on Σ_0, the nominal model should be elliptical, for example the multivariate normal. In this case H_{00} is the distribution of $\sqrt{v}$, where v has a chi-square distribution with q degrees of freedom. We will see in Theorem 4 that if the true distribution F_0 of u_i is elliptical, the estimate $\widetilde{B}$ converges to B_0 and $\widetilde{\Sigma}$ to $\lambda\Sigma_0$, even if $H_{00} \neq H_0$. If $H_{00} = H_0$, we have $\lambda = 1$.

Let h be the maximum number of observations (x_i, y_i) lying in a hyperplane, i.e.,

$$h = \max_{\|a\|+\|b\|>0} \#\{i\colon a'x_i + b'y_i = 0\} \tag{2.12}$$

and

$$\eta = \frac{\kappa_1}{\max_u \ \rho_1(u)}. \tag{2.13}$$

It can be shown that if ρ_1 and ρ_2 satisfy A1–A5 and $h/n < 1 - \eta$, then there exists a $(\widetilde{B}, \widetilde{\Sigma})$ solution to the optimization problem defined by (2.10) and (2.11) with $\left|\widetilde{\Sigma}\right| > 0$. The proof of this is similar to the proofs of Lemmas 4 and 5 of the Appendix. As we will see in Section 3, to obtain a high breakdown point τ-estimate, η should be close to 0.5. If there are multiple solutions we take any of them.

Comparing (2.10) and (2.11) with (2.3) and (2.6), we observe that the MLE may be considered as a τ-estimate corresponding to the unbounded function $\rho_2(v) = v^2$ and $\kappa_2 = q$.

It is easy to show that when there is only one regressor variable identically equal to one, the τ-estimates $\widetilde{B}$ and $\widetilde{\Sigma}$ defined here are the estimates of multivariate location and scatter introduced by Lopuhaä [11]. When $q = 1$, they are the τ-estimates of regression proposed by Yohai and Zamar [17].

It can be verified that τ-estimates are affine and regression equivariant.

In Theorem 1 we obtain the estimating equations of τ-estimates. Define

$$d_i^*(B,\Sigma) = \frac{d_i(B,\Sigma)}{s(d_1(B,\Sigma),\ldots,d_n(B,\Sigma))}$$

and let $C_n = C_n(B,\Sigma)$, $D_n = D_n(B,\Sigma)$, $\psi_n^* = \psi_{n,B,\Sigma}^*$ and w_n^* be defined by

$$C_n = \frac{1}{n}\sum_{i=1}^n \left(2\rho_2(d_i^*(B,\Sigma)) - \psi_2(d_i^*(B,\Sigma))d_i^*(B,\Sigma)\right),$$

$$D_n = \frac{1}{n}\sum_{i=1}^n \psi_1(d_i^*(B,\Sigma))d_i^*(B,\Sigma),$$

$$\psi_n^*(v) = C_n\psi_1(v) + D_n\psi_2(v) \tag{2.14}$$

and

$$w_n^*(v) = \frac{\psi_n^*(v)}{v}. \tag{2.15}$$

Theorem 1. *Suppose that ρ_1 and ρ_2 are differentiable. Then the τ-estimates satisfy the following equations*:

$$\sum_{i=1}^n w_n^*(d_i^*(\widetilde{B},\widetilde{\Sigma}))\widehat{u}_i(\widetilde{B})x_i' = 0 \tag{2.16}$$

and

$$\widetilde{\Sigma} = \frac{q \sum_{i=1}^{n} w_n^*(d_i^*(\widetilde{B}, \widetilde{\Sigma}))\widehat{u}_i(\widetilde{B})\widehat{u}_i'(\widetilde{B})}{\widetilde{s}^2 \sum_{i=1}^{n} \psi_n^*(d_i^*(\widetilde{B}, \widetilde{\Sigma}))d_i^*(\widetilde{B}, \widetilde{\Sigma})}, \tag{2.17}$$

where $\widehat{u}_i(B)$ *is defined in* (2.1) *and* $\widetilde{s} = s(d_1(\widetilde{B}, \widetilde{\Sigma}), \ldots, d_n(\widetilde{B}, \widetilde{\Sigma}))$.

The proof of this theorem can be found in the Appendix.

Remark. Observe that according to (2.16), the *j*th column of $\widetilde{B}$ is the weighted LSE corresponding to the univariate regression whose dependent variable is the *j*th component of y, the vector of independent variables is x and the observation i receives weight $w_n^*(d_i^*(\widetilde{B}, \widetilde{\Sigma}))$. Besides, by (2.17), $\widetilde{\Sigma}$ is proportional to the weighted sample covariance of the residuals with the same weights. Since these weights depend on the estimates $\widetilde{B}$ and $\widetilde{\Sigma}$, we cannot use these relationships to compute the estimates, but they will be the base of the iterative algorithm described in Section 6.

3. Breakdown point

Donoho and Huber [5] introduced the concept of a finite sample breakdown point (FSBDP). For the case of the MLM, let $\hat{B}$ and $\hat{\Sigma}$ be estimates of B and Σ. The FSBDP of $\hat{B}$ is the smallest fraction of outliers that makes this estimate unbounded, and the FSBDP of $\hat{\Sigma}$ the smallest fraction of outliers that makes this estimate unbounded or singular. To formalize this, let Z be a data set of size n corresponding to an MLM, $Z = \{z_1, \ldots, z_n\}$, $z_i = (x_i', y_i')'$, $x_i \in R^p$, $y_i \in R^q$. Let $\mathcal{Z}_m$ be the set of all the samples $Z^* = \{z_1^*, \ldots, z_n^*\}$ such that $\#\{i : z_i = z_i^*\} \geqslant n - m$. Given estimates $\hat{B}$ and $\hat{\Sigma}$, denote by $\hat{\lambda}_1 \leqslant \cdots \leqslant \hat{\lambda}_q$ the eigenvalues of $\hat{\Sigma}$,

$$S_m(Z, \hat{B}) = \sup\{\|\widehat{B}(Z^*)\|, Z^* \in \mathcal{Z}_m\},$$

where $\| \ \|$ is the l_2 norm, and let

$$S_m^-(Z, \hat{\Sigma}) = \inf\{\widehat{\lambda}_1(\widehat{\Sigma}(Z^*)), Z^* \in \mathcal{Z}_m\}$$

and

$$S_m^+(Z, \hat{\Sigma}) = \sup\{\widehat{\lambda}_q(\widehat{\Sigma}(Z^*)), Z^* \in \mathcal{Z}_m\}.$$

Definition. The finite sample breakdown point of $\hat{B}$ is defined by $\varepsilon^*(Z, \hat{B}) = m^*/n$ where $m^* = \min\left\{m : S_m(Z, \hat{B}) = \infty\right\}$ and the breakdown point of $\hat{\Sigma}$ by $\varepsilon^*(Z, \hat{\Sigma}) = m^*/n$ where

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

The following theorem, whose proof can be found in the Appendix, gives a lower bound for the breakdown point of τ-estimates.

Theorem 2. *Let* $Z = \{z_1, z_2, \ldots, z_n\}$ *with* $z_i = (x_i', y_i')'$, *h as defined in* (2.12), *and* η *as defined in* (2.13). *Assume that* ρ_1 *and* ρ_2 *satisfy A1–A5*; *then a lower bound for* $\varepsilon^*(Z, \widetilde{\Sigma})$ *and* $\varepsilon^*(Z, \widetilde{B})$ *is given by* $\min((1 - \eta) - (h/n), \eta)$.

The maximum breakdown point is $(n-h)/(2n)$ and corresponds to $\eta=(n-h)/(2n)$. The minimum value of h is $p+q-1$. If $h=p+q-1$, we say that the points are in a general position. If $\eta=(n-h)/(2n)$ and $h=p+q-1$, we obtain that $\varepsilon^*(Z,\widetilde{\Sigma})$ and $\varepsilon^*(Z,\widetilde{B})$ are larger than or equal to $0.5-((p+q-1)/(2n))$, and then in this case the two breakdown points are close to 0.5 for large n.

4. Influence function

Consider a random variable $z\in R^k$ with distribution G_θ, where $\theta\in\Theta\subset R^m$. Let T be an estimating functional of θ, i.e., T is defined on a set of distributions G on R^k, including the empirical distributions and the contamination neighborhoods of G_θ, and takes values on Θ. Suppose that T is Fisher consistent, i.e. $T(G_\theta)=\theta$. Given a sample $z_1,z_2,\ldots,z_n$ of G_θ, the estimate $\widehat{\theta}$ associated to T is defined as $T(G_n)$, where G_n is the corresponding empirical distribution. The influence function of T measures the effect on the functional of a small fraction of point mass contamination ([7]). More precisely, let δ_z be the point mass distribution at z and $G_{\theta,\varepsilon}=(1-\varepsilon)G_\theta+\varepsilon\delta_z$, then the influence function is defined by

$$IF(z,T,\theta)=\lim_{\varepsilon\longrightarrow 0}\frac{T(G_{\theta,\varepsilon})-T(G_\theta)}{\varepsilon}=\frac{\partial}{\partial\varepsilon}T(G_{\theta,\varepsilon})\bigg|_{\varepsilon=0}.$$

In our case, $z=(x',y')'$ and $\theta=(B_0,\Sigma_0)$. The estimating functionals T_1 and T_2 corresponding to the τ-estimates $\widetilde{B}$ and $\widetilde{\Sigma}$ are given in the Appendix.

For the sake of simplicity we derive the influence function of τ-estimates assuming that the errors in (1.1) have an elliptical distribution with a unimodal density. Then, we need the following assumption:

(A7) The distribution F_0 of u_i has a density of the form

$$f_0(u)=\frac{f_0^*(u'\Sigma_0^{-1}u)}{\det(\Sigma_0)^{1/2}},\tag{4.1}$$

where f_0^* is nonincreasing and has at least one point of decrease in the interval where both functions ρ_1 and ρ_2 are strictly increasing. Besides, Σ_0 is a $q\times q$ positive-definite matrix.

An important family of unimodal elliptical distributions is the multivariate normal. In this case, $f_0^*(v)=(1/(2\pi)^{q/2})\exp(-v/2)$.

Given an arbitrary distribution H on R, define $s^*(H)$ by

$$E_H\left(\rho_1\left(\frac{v}{s^*(H)}\right)\right)=\kappa_1.\tag{4.2}$$

Observe that the M-estimate of scale $s(v_1,\ldots,v_n)=s^*(H_n)$, where H_n is the empirical distribution of $v_1,\ldots,v_n$. We denote

$$k_0=s^*(H_0),\tag{4.3}$$

where H_0 is the true distribution of $(u'\Sigma_0^{-1}u)^{1/2}$.

To define the limit of the sequence of the functions ψ_n^* and w_n^*, given in (2.14) and (2.15), put

$$C = E_{H_0}\left(2\rho_2\left(\frac{v}{k_0}\right) - \psi_2\left(\frac{v}{k_0}\right)\left(\frac{v}{k_0}\right)\right),$$

$$D = E_{H_0}\left(\psi_1\left(\frac{v}{k_0}\right)\left(\frac{v}{k_0}\right)\right).$$

Then, define $\psi^*(v) = C\psi_1(v) + D\psi_2(v)$ and $w^*(v) = \psi^*(v)/v$.

In the following we shall need some additional assumptions:

(A8) The x_i's are random, their common distribution M_0 has finite second moments and $E_{M_0}(xx')$ is nonsingular.

(A9) ρ_i is twice differentiable ($i = 1, 2$).

Theorem 3. *Assume A1–A9. Then, the influence function for the estimating functional T_1 corresponding to the τ-estimate $\widetilde{B}$ is*

$$IF(y_0, x_0, T_1, B_0, \Sigma_0) = c_0\, w^*\left(\frac{((y_0 - B_0'x_0)'\Sigma_0^{-1}(y_0 - B_0'x_0))^{1/2}}{k_0}\right) E_{M_0}(xx')^{-1}x_0(y_0 - B_0'x_0)', \tag{4.4}$$

where

$$c_0 = \frac{q}{E_{H_0}((q-1)w^*(\frac{v}{k_0}) + \psi^{*\prime}(\frac{v}{k_0}))}. \tag{4.5}$$

The proof of this theorem is given in the Appendix.

5. Consistency and asymptotic normality

In this section we state the asymptotic properties of τ-estimates for the MLM assuming that the errors have an elliptical distribution. Theorems 4 and 5 establish, respectively, the consistency and the asymptotic normality of $\widetilde{B}$.

Let $\sigma_0 = \sigma_0(\psi_1, \psi_2)$ be defined by

$$\frac{\sigma_0^2}{\kappa_2} E_{H_0}\left(\rho_2\left(\frac{v}{k_0}\right)\right) = 1, \tag{5.1}$$

where k_0 is given in (4.3). If $H_0 = H_{00}$ we have $\sigma_0 = k_0 = 1$.

Theorem 4. *Let (x_i, y_i), $1 \leqslant i \leqslant n$, be a random sample of an MLM with parameters B_0 and Σ_0, where the x_i's are random. Suppose also that A1–A7 hold; then the τ-estimates $\widetilde{B}_n$ and $\widetilde{\Sigma}_n$ satisfy*

(a) $\lim_{n\to\infty} \widetilde{B}_n = B_0$ *a.s.*.
(b) $\lim_{n\to\infty} \widetilde{\Sigma}_n = (k_0^2/\sigma_0^2)\Sigma_0$ *a.s.. In particular if $H_{00} = H_0$, then $k_0^2/\sigma_0^2 = 1$.*

To prove the asymptotic normality of $\widetilde{B}_n$ we need the following additional assumption:

(A10) There exist $m_i (i = 1, 2)$, such that $\rho_i(u)$ is constant for $|u| > m_i$.

Theorem 5. *Let* $(x_i y_i)$, $1 \leqslant i \leqslant n$, *be a random sample of an MLM with parameters* B_0 *and* Σ_0. *Suppose also that A1–A10 hold; then* $n^{1/2}\mathrm{vec}(\widetilde{B}_n - B_0) \rightarrow^D N(0, V)$, *where* $\rightarrow^D$ *denotes convergence in distribution and*

$$V = \frac{c_0^2 k_0^2}{q} E_{H_0}\left(\psi^{*2}\left(\frac{v}{k_0}\right)\right) \Sigma \otimes E_{M_0}(xx')^{-1}, \tag{5.2}$$

with k_0 *and* c_0 *given in* (4.3) *and* (4.5), *respectively*.

We do not have formal proofs of Theorems 4 and 5, but we think that they can be obtained using arguments similar to those in Yohai and Zamar [16,17], and Lopuhaä [1]. Heuristic proofs of both Theorems can be found in the Appendix, where we also give a rigorous proof of the Fisher consistency of these estimates (Lemma 12).

Remark. Estimates of V can be obtained replacing in (5.2) H_0 and M_o by the empirical distribution of the residuals and of the x's, respectively.

The MLE estimator $\widehat{B}_n$ corresponds to $\rho_2(u) = u^2$, and in this case $c_0 = 1/2$, $\psi^*(u) = 2u$ and the asymptotic covariance matrix is given by $(E_{H_0}(v^2)/q)\, \Sigma \otimes E_{M_0}(xx')^{-1}$ which differs from the one of the τ-estimate by a scalar factor. Then we obtain the following result:

Corollary. *Under the assumptions of Theorem* 5, *the asymptotic relative efficiency* (*ARE*) *of the* τ-*estimate* $\widetilde{B}$ *with respect to the MLE* $\widehat{B}$ *is*

$$ARE(\psi_1, \psi_2, H_0) = \frac{E_{H_0}(v^2)}{c_0^2\, k_0^2 E_{H_0}\left(\psi^{*2}\left(\frac{v}{k_0}\right)\right)}.$$

In order to obtain a τ-estimate which is simultaneously highly robust and highly efficient under normal errors we can choose ρ_1 so that

$$\kappa_1 / \max \rho_1 = 0.5, \tag{5.3}$$

which guarantees that the initial M-estimate of scale s has a breakdown point close to 0.5, and choose as $\rho_2(v)$ a bounded function close enough to v^2 to obtain the desired efficiency. This can be achieved, for example, by taking ρ_1 and ρ_2 in Tukey's bisquare family defined by

$$\rho_{B,c}(v) = \begin{cases} \frac{v^2}{2}\left(1 - \frac{v^2}{c^2} + \frac{v^4}{3c^4}\right) & \text{if } |v| \leqslant c \\ \frac{c^2}{6} & \text{if } |v| > c, \end{cases}$$

where c is any positive number. Observe that when c increases, $\rho_{B,c}$ approaches v^2. It is easy to verify that the functions $\rho_{B,c}(v)$ satisfy A1–A6.

Table 1 gives the values of c_1 and κ_1 such that $\rho_1 = \rho_{B,c_1}$ satisfies (5.3) and the ARE of S-estimates under Gaussian errors for different values of q. We observe that the efficiency of the S-estimate is low for small values of q, but increases with q. It may be shown that this efficiency converges to one when $q \to \infty$. Table 2 gives the values of c_2 and κ_2 to achieve different levels of asymptotic efficiency (taking $\rho_1 = \rho_{B,c_1}$ and $\rho_2 = \rho_{B,c_2}$).

Table 1
Values of c_1 and κ_1 for the bisquare function and asymptotic relative efficiency (ARE) under Gaussian errors of the S-estimate with breakdown point 0.5

q	1	2	3	4	5	10
c_1	1.55	2.66	3.45	4.10	4.65	6.77
κ_1	0.20	0.59	0.99	1.40	1.80	3.82
ARE	0.29	0.58	0.72	0.80	0.85	0.93

Table 2
Values of c_2 and κ_2 for the bisquare function to attain given values of the asymptotic relative efficiency (ARE) under Gaussian errors

ARE		q					
		1	2	3	4	5	10
0.80	c_2	3.98	3.94	4.02	4.10	4.17	4.28
	κ_2	0.42	0.77	1.10	1.40	1.67	2.56
0.90	c_2	4.97	4.97	5.10	5.25	5.39	5.98
	κ_2	0.44	0.85	1.24	1.61	1.96	3.54
0.95	c_2	6.04	6.06	6.24	6.42	6.60	7.50
	κ_2	0.46	0.90	1.32	1.73	2.13	4.03

6. Computing algorithm

Based on the remark at the end of Section 2 we propose the following iterative algorithm to compute $\widetilde{B}$ and $\widetilde{\Sigma}$.

1. Using initial values $\widetilde{B}_0$ and $\tilde{\Sigma}_0$ satisfying (2.11), compute $\widetilde{s}_0 = s(d_1(\widetilde{B}_0, \widetilde{\Sigma}_0), \ldots, d_n(\widetilde{B}_0, \widetilde{\Sigma}_0))$ and the weights $w_n^*(d_i(\widetilde{B}_0, \widetilde{\Sigma}_0)/\widetilde{s}_0)$ for $1 \leqslant i \leqslant n$. These weights are used to compute each column of $\widetilde{B}_1$ separately by WLS. Now compute $\widetilde{s}_1 = s(d_1(\widetilde{B}_1, \widetilde{\Sigma}_0), \ldots, d_n(\widetilde{B}_1, \widetilde{\Sigma}_0))$.
2. Compute the matrix
$$\widetilde{\Sigma}_1^* = \frac{q \sum_{i=1}^n w_n^* \left(\frac{d_i(\widetilde{B}_1, \widetilde{\Sigma}_0)}{\widetilde{s}_1} \right) u_i(\widetilde{B}_1) u_i'(\widetilde{B}_1)}{\widetilde{s}_1^2 \sum_{i=1}^n \psi_n^* \left(\frac{d_i(\widetilde{B}_1, \widetilde{\Sigma}_0)}{\widetilde{s}_1} \right) \frac{d_i(\widetilde{B}_1, \widetilde{\Sigma}_0)}{\widetilde{s}_1}}. \tag{6.1}$$
3. Compute $\widetilde{\tau}_1 = \tau(d_1(\widetilde{B}_1, \widetilde{\Sigma}_1^*), \ldots, d_n(\widetilde{B}_1, \widetilde{\Sigma}_1^*))$ and $\widetilde{\Sigma}_1 = (\widetilde{\tau}_1^2/\kappa_2)\widetilde{\Sigma}_1^*$. Then $(\widetilde{B}_1, \widetilde{\Sigma}_1)$ satisfy constraint (2.11).
4. Suppose now that we have already computed $(\widetilde{B}_h, \widetilde{\Sigma}_h)$ satisfying constraint (2.11). Then $(\widetilde{B}_{h+1}, \widetilde{\Sigma}_{h+1})$ is computed using steps 1–3, but starting from $(\widetilde{B}_h, \widetilde{\Sigma}_h)$ instead of $(\widetilde{B}_0, \widetilde{\Sigma}_0)$.
5. The procedure is stopped at step h if the relative absolute differences of all elements of the matrices $\widetilde{B}_h$ and $\widetilde{B}_{h-1}$ are smaller than a given value δ.

We have not proved that the reweighting step improves the value of the goal function. However in the Monte Carlo study described in Section 7 this has always occurred.

We propose to compute the initial estimates $\widetilde{B}_0$ and $\widetilde{\Sigma}_0$ by subsampling elemental sets. For this purpose we take N random subsamples of size $r = p + q$ of the original sample. For the jth subsample two values of (B, Σ) are obtained. The first $(B_j^{(1)}, \Sigma_j^{(1)})$ corresponds to the MLE of the subsample, and the second value $(B_j^{(2)}, \Sigma_j^{(2)})$ is the MLE of the $[n/2]$ observations with the smallest Mahalanobis distances $d_i(B_j^{(1)}, \Sigma_j^{(1)})$, $1 \leqslant i \leqslant n$. We now compute $d_i(B_j^{(2)}, \Sigma_j^{(2)})$, $1 \leqslant i \leqslant n$, and $s_j = s_0\left(d_1(B_j^{(2)}, \Sigma_j^{(2)}), \ldots, .d_n(B_j^{(2)}, \Sigma_j^{(2)})\right)$, where $s_0(u_1, \ldots, u_n) = \text{median}(|u|_1, \ldots, |u|_n)$ and standardize $\Sigma_j^{(2)}$ obtaining $\Sigma_j^{(3)} = s_j^2 \Sigma_j^{(2)}$. Then $s_0^2\left(d_1(B_j^{(2)}, \Sigma_j^{(3)}), \ldots, .d_n(B_j^{(2)}, \Sigma_j^{(3)})\right) = 1$.

An approximation to the estimate that minimizes $\det(\Sigma)$ subject to $s_0^2\,(d_1(B, \Sigma), \ldots, .d_n(B, \Sigma)) = 1$ is given by $(B_{j_0}^{(2)}, \Sigma_{j_0}^{(3)})$ where $j_0 = \arg\min_{1 \leqslant j \leqslant N} \det(\Sigma_j^{(3)})$.

We use the scale s_0 instead of the scale τ because it is faster to compute. This scale is inefficient, but since it is used only to compute the initial estimate, it does not affect the efficiency of the final estimate.

Finally, the initial estimate is obtained by restandardizing $\Sigma_{j_0}^{(3)}$ so that the τ-scale of the Mahalanobis distances is $\sqrt{\kappa_2}$. For this purpose we compute

$$\tau_{j_0} = \tau\left(d_1(B_{j_0}^{(2)}, \Sigma_{j_0}^{(3)}), \ldots, .d_n(B_{j_0}^{(2)}, \Sigma_{j_0}^{(3)})\right)$$

and the initial estimates are $\widetilde{B}_0 = B_{j_0}^{(2)}$ and $\widetilde{\Sigma}_0 = (\tau_{j_0}^2/\kappa_2)\Sigma_{j_0}^{(3)}$.

The reason why we compute the second value $(B_j^{(2)}, \Sigma_j^{(2)})$ is that even if the jth sample does not contain outliers, it may be badly conditioned and the corresponding fit $B_j^{(1)}$ may be very far from the true value. However, eliminating the sample half with largest $d_i(B_j^{(1)}, \Sigma_j^{(1)})$'s increases the chance of obtaining a clean sample that produces a better value, $B_j^{(2)}$. This mechanism is similar to the one proposed by Rousseeuw and Van Driessen [14].

One improvement suggested by a referee would be to proceed as in Rousseeuw and Van Driessen [14] keeping the M solutions $(B_j^{(2)}, \Sigma_j^{(3)})$ with smallest $|\Sigma_j^{(3)}|$ (for example $M = 10$) and starting the iterative process from each one of them. This gives M new values $(\widetilde{B}_h, \widetilde{\Sigma}_h)$, $1 \leqslant h \leqslant M$, and the final estimate is the one with the smallest $|\widetilde{\Sigma}_h|$.

7. Monte Carlo results

In order to assess the robustness and efficiency of the proposed estimates we performed a Monte Carlo 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, $B_0 = 0$ and $\Sigma_0 = I_q$, where I_q denotes a $q \times q$ identity matrix. The errors u_i are generated from an $\mathrm{N}_q(0,\ I_q)$ distribution and the regressors x_i from an $\mathrm{N}_p(0, I_p)$ 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 (x, y) with $x' = (x_0, 0, \ldots, 0)$ and $y' = (mx_0, 0, \ldots, 0)$. 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 last value of the grid was taken so that the maximum mean square error (MSE) of all the robust estimates is attained. Suppose that $\widehat{B}^{(k)} = \left(\widehat{B}_{ij}^{(k)}\right)$ is the estimate of B_0 obtained in the kth replication. Then, since we are taking $B_0 = 0$, the estimate of the MSE is given by

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

Table 3
Monte Carlo mean square error (MSE), standard error of the MSE (SE) and relative efficiency (REFF) of the estimates in the noncontaminated case for $n = 100$ and $p = 2$

Estimate	$q = 2$				$q = 5$			
	MSE	SE	REFF	ARE	MSE	SE	REFF	ARE
MLE	0.064	0.001	1.00	1.00	0.157	0.002	1.00	1.00
S-estimate	0.120	0.002	0.53	0.58	0.190	0.002	0.83	0.85
τ-estimate	0.072	0.001	0.89	0.90	0.176	0.002	0.89	0.90

ARE is the asymptotic relative efficiency.

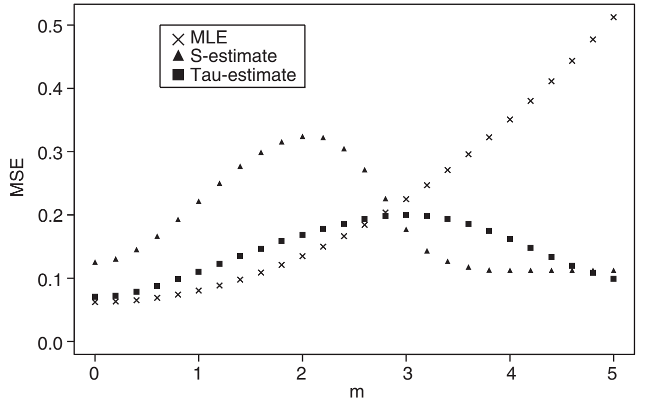

Fig. 1. Mean square error for $q = 2$ and $x_0 = 1$.

For each case, three estimates are computed: the MLE, an S-estimate and a τ-estimate. The S- and the τ-estimates are based on ρ-functions in the bisquare family. The S-estimate is based on the M-scale defined by $\rho_1 = \rho_{B,c_1}$, where c_1 and κ_1 were chosen so that this estimate has a breakdown point 0.5 (see Table 1). The τ-estimate uses the same ρ_1 and κ_1 as the S-estimate and $\rho_2 = \rho_{B,c_2}$, where c_2 and κ_2 were chosen so that the τ-estimate had an ARE equal to 0.90 when the errors are Gaussian (see Table 2). We compute the initial estimate using 2000 subsamples and the value of δ in step 5 of the computing algorithm is taken equal to 0.0001.

In Table 3 we present the MSE, its standard error (SE) and the relative efficiency (REFF) with respect to the MLE for the uncontaminated case. The efficiency of the S-estimate is low when $q = 2$ and increases for $q = 5$. We observe that the relative efficiencies of the S- and τ-estimates are close to their asymptotic values.

In Figs. 1 and 2 we show the MSE of the different estimates under contamination. In Fig. 1, which corresponds to $q = 2$ and $x_0 = 1$, we observe that the τ-estimate has a smaller MSE than the S-estimate except when m is (approximately) between 3 and 5, and the maximum MSE is smaller for the τ-estimate. As expected, the MSE of the MLE increases with m reaching very large values. Fig. 2 shows the results for $q = 2$ and $x_0 = 10$. S- and τ-estimates behave similarly, with a small advantage for the τ-estimate. Since for $q = 5$ the behavior of S- and τ-estimates is similar to the one observed for $q = 2$, we do not report the results here.

As a general conclusion we can say that when there are no outliers, τ-estimates are more efficient than S-estimates, and under outlier contamination, τ-estimates behave better than or similar to S-estimates.

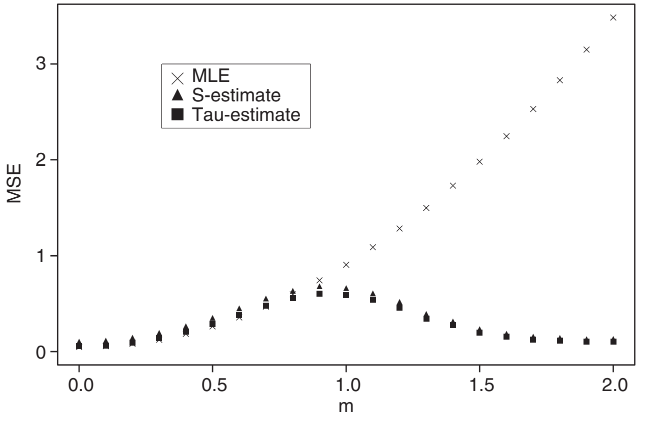

Fig. 2. Mean square error for $q = 2$ and $x_0 = 10$.

8. Example

Observational studies have suggested that low dietary intake or low plasma concentrations of retinol, beta-carotene or other carotenoids might be associated with increased risk of developing certain types of cancer. Nieremberg et al. [12] studied the determinants of plasma concentrations of these micronutrients. In an unpublished study they have collected information on 14 variables that may be the determinants of the plasma levels of beta-carotene and retinol. The number of observations was 315 and the variables considered were: Y_1 BETAPLASMA: Plasma beta-carotene (ng/ml), Y_2 RETPLASMA: Plasma Retinol (ng/ml), X_1 AGE (years), X_2 SEX ($1 =$ Male, $2 =$ Female), X_3 SMOKSTAT: Smoking status ($1 =$ Never), X_4 SMOKSTAT: Smoking status ($1 =$ Former), X_5 QUETELET (weight/(height^2)), X_6 VITUSE: Vitamin Use ($1 =$ fairly often), X_7 VITUSE: Vitamin Use ($1 =$ not often), X_8 CALORIES: Number of calories consumed per day, X_9 FAT: Grams of fat consumed per day, X_{10} FIBER: Grams of fiber consumed per day, X_{11} ALCOHOL: Number of alcoholic drinks consumed per week, X_{12} CHOLESTEROL: Cholesterol consumed (mg/day), X_{13} BETADIET: Dietary beta-carotene consumed (mcg per day), X_{14} RETDIET: Dietary retinol consumed (mcg/day). The data are available at http://lib.stat.cmu.edu/datasets/Plasma_Retinol.

We compute two estimates of the regression coefficients: the multivariate τ-estimate and the MLE. The τ-estimate uses ρ_1 and ρ_2 in the bisquare family with constants equal to those used in the Monte Carlo study of Section 7. In Fig. 3 we show, for both estimates, the box-plot of the Mahalanobis norms of the residuals $d_i = ((\mathbf{y}_i - \widehat{B}\mathbf{x}_i)'\widehat{\Sigma}^{-1}(\mathbf{y}_i - \widehat{B}\mathbf{x}_i))^{1/2},\ 1 \leqslant i \leqslant 315$. If we declare outliers those observations such that $d_i > \sqrt{\chi^2_{2,0.99}}$, the τ-estimate reveals 27 outliers while the MLE reveals only 12. In Fig. 4 we present QQ-plots of the absolute values of the residuals of the τ-estimate against the absolute value of the residuals of the MLE, after eliminating the 27 outliers detected by the τ-estimate. Fig. 4(a) shows that for the plasma beta-carotene the τ-estimate gives residuals smaller than the MLE. Instead, Fig. 4(b) shows that for the plasma retinol, the distributions of both residuals are close.

In Table 4 we show the regression coefficients and their standard errors (SE) for three estimates: the τ-estimate, the MLE and the MLE after omitting the 27 outliers. We only show these values for the variables that are statistically significant at level 0.05 for at least one estimate and one

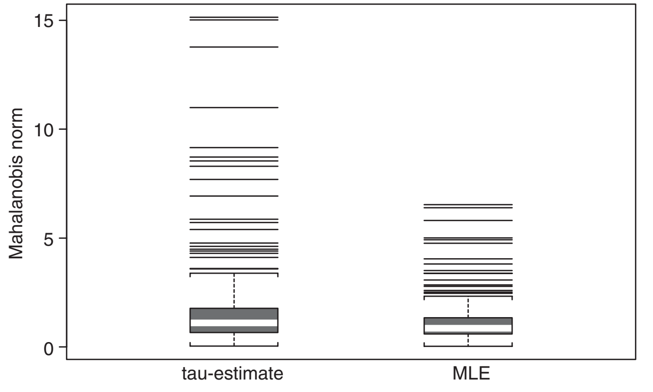

Fig. 3. Box-plot of the Mahalanobis norms of the residuals.

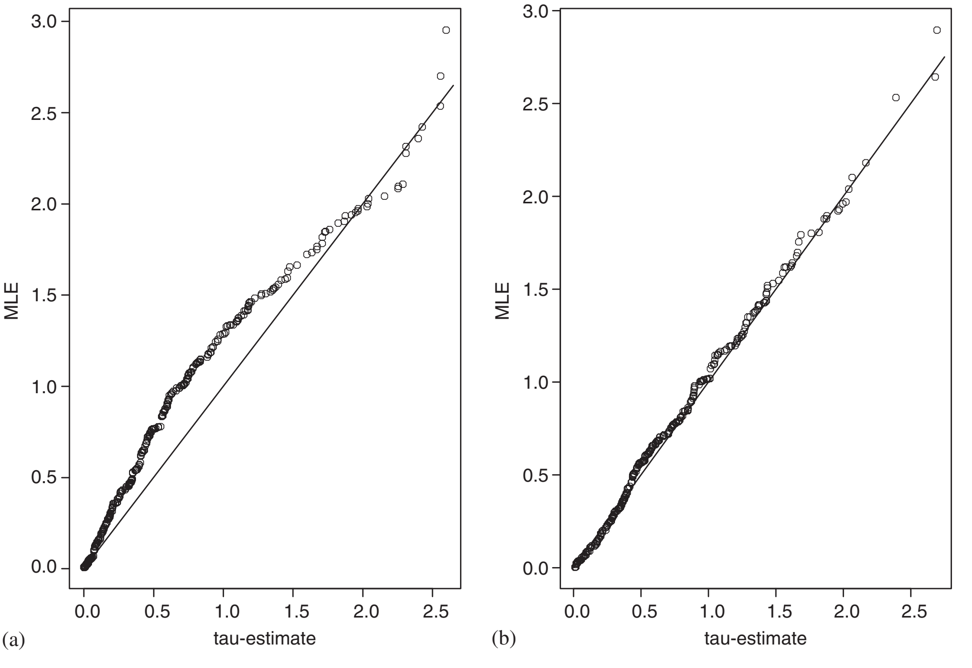

Fig. 4. QQ-plots of the absolute values of the residuals for both regressions (a) correspond to the plasma beta-carotene and (b) to the plasma retinol.

equation. We also show the error variances corresponding to each regression in this table. The estimated standard errors of the τ-estimates were calculated as proposed in the Remark after Theorem 5.

We can observe that when the MLE and τ-estimates are computed using the complete dataset, some regression coefficients and standard errors are quite different. Instead the results for the τ-estimate computed with all the observations are quite close to those of the MLE after deleting the 27 outliers. This is what a robust estimate is expected to do.

Table 4
Regression coefficients and standard errors for the τ-estimates, the MLE and the MLE after omitting 27 outliers (MLE$_{-27}$)

Equation	1			2		
Estimate	τ	MLE	MLE$_{-27}$	τ	MLE	MLE$_{-27}$
X_1	0.15	0.14	0.18	0.27	0.21	0.26
SE	0.07	0.14	0.07	0.08	0.08	0.07
X_2	−0.44	−0.37	−0.34	0.02	0.53	0.09
SE	0.19	0.35	0.18	0.21	0.20	0.18
X_5	−0.18	−0.33	−0.20	0.02	0.02	0.02
SE	0.04	0.09	0.04	0.05	0.05	0.04
X_6	0.15	0.89	0.36	0.02	0.15	0.10
SE	0.13	0.26	0.13	0.15	0.14	0.13
X_{10}	0.21	0.33	0.16	−0.13	−0.09	−0.12
SE	0.07	0.14	0.07	0.08	0.08	0.07
X_{13}	0.13	0.21	0.15	0.01	−0.01	−0.01
SE	0.05	0.10	0.05	0.06	0.05	0.05
Error variance	0.81	3.51	0.78	1.01	1.10	0.79

Acknowledgments

This work was partially supported by a grant from the Fundación Antorchas, Argentina, Grant X094 from University of Buenos Aires and Grant 03-06277 from the Agencia Nacional de Promoción Científica y Tecnológica, Argentina. We also thank the referees whose suggestions and criticisms helped to improve the paper.

Appendix

A.1. Theorem 1

Proof of Theorem 1. Put $Q(B,\Sigma) = \det(\Sigma)\ \tau^{2q}(d_1(B,\Sigma),\dots,d_n(B,\Sigma))$. It is easy to show that for any real λ, $Q(B,\lambda\Sigma) = Q(B,\Sigma)$, then the τ-estimate $(\widetilde{B},\widetilde{\Sigma})$ also minimizes $Q(B,\Sigma)$ without restrictions (observe that $(\widetilde{B},\lambda\widetilde{\Sigma})$ minimizes $Q(B,\Sigma)$ too), or equivalently $\log\left(Q(B,\Sigma)\right)$. Therefore they should satisfy the following equations:

$$\frac{\partial \log(Q(B,\Sigma))}{\partial B} = 0, \quad \frac{\partial \log(Q(B,\Sigma))}{\partial \Sigma} = 0. \tag{A.1}$$

From now on, and for the sake of simplicity, we will denote $d_i{=}d_i(B,\Sigma)$, $d_i^*{=}d_i(B,\Sigma)/s(B,\Sigma)$ and $s = s(B,\Sigma) = s(d_1(B,\Sigma),\dots,d_n(B,\Sigma))$.

Differentiating $\log(Q(B,\Sigma))$ with respect to B, after straightforward calculations, we obtain

$$\frac{\partial \log(Q(B,\Sigma))}{\partial(\mathrm{vec}(B'))'} = -2q\,\frac{\sum_{i=1}^n \psi_1\left(d_i^*\right)\frac{\widehat{u}_i'(B)}{d_i}\Sigma^{-1}\left(x_i'\otimes I_q\right)}{\sum_{i=1}^n\left(\psi_1\left(d_i^*\right)d_i\right)} - \frac{q\sum_{i=1}^n\ \psi_2\left(d_i^*\right)\frac{\widehat{u}_i'(B)}{d_i}\Sigma^{-1}\left(x_i'\otimes I_q\right)}{s\sum_{i=1}^n\ \rho_2\left(d_i^*\right)}$$

$$+\frac{q\sum_{i=1}^{n} d_i^*\psi_2\left(d_i^*\right)\sum_{j=1}^{n}\psi_1\left(d_j^*\right)\frac{\widehat{u}_j'(B)}{d_j}\Sigma^{-1}\left(x_j'\otimes I_q\right)}{\sum_{i=1}^{n}\rho_2\left(d_i^*\right)\sum_{j=1}^{n}\left(\psi_1\left(d_j^*\right)d_j\right)}.$$

Then, equating this last expression to zero, we have

$$\frac{\sum_{i=1}^{n}\psi_1\left(d_i^*\right)\frac{\widehat{u}_i'(B)}{d_i}\Sigma^{-1}\left(x_i'\otimes I_q\right)\left[-2\sum_{j=1}^{n}\rho_2\left(d_j^*\right)+\sum_{j=1}^{n}\psi_2\left(d_j^*\right)d_j^*\right]}{\sum_{i=1}^{n}\rho_2\left(d_i^*\right)\sum_{j=1}^{n}\left(\psi_1\left(d_j^*\right)d_j\right)}$$
$$-\frac{\frac{1}{s}\sum_{i=1}^{n}\psi_2\left(d_i^*\right)\frac{\widehat{u}_i'(B)}{d_i}\Sigma^{-1}\left(x_i'\otimes I_q\right)\sum_{j=1}^{n}\psi_1\left(d_j^*\right)d_j}{\sum_{i=1}^{n}\rho_2\left(d_i^*\right)\sum_{j=1}^{n}\left(\psi_1\left(d_j^*\right)d_j\right)}=0,$$

or equivalently,

$$\sum_{i=1}^{n} w_n^*\left(d_i^*\right)\widehat{u}_i'(B)\Sigma^{-1}\left(x_i'\otimes I_q\right)=0. \tag{A.2}$$

Using that $\operatorname{vec}(\Sigma^{-1}u_i x_i') = (x_i\otimes I_q)\Sigma^{-1}u_i$, we can show that expression (A.2) is equivalent to (2.16).

Differentiating $\log(Q(B,\Sigma))$ with respect to Σ, we obtain

$$\frac{\partial\log(Q(B,\Sigma))}{\partial\Sigma}=\frac{\partial\log(\det(\Sigma))}{\partial\Sigma}+\frac{2q}{s}\frac{\partial s}{\partial\Sigma}+\frac{q}{\sum_{i=1}^{n}\rho_2\left(d_i^*\right)}$$
$$\times\sum_{i=1}^{n}\frac{\partial}{\partial\Sigma}\rho_2\left(d_i^*\right). \tag{A.3}$$

It is well known that for a symmetric matrix Σ,

$$\frac{\partial\log(\det(\Sigma))}{\partial\Sigma}=\Sigma^{-1}. \tag{A.4}$$

Then, using that

$$\frac{\partial s}{\partial\Sigma}=-\frac{1}{2s}\frac{\Sigma^{-1}\left(\sum_{i=1}^{n}\psi_1(d_i^*))\frac{s}{d_i}\widehat{u}_i(B)\widehat{u}_i'(B)\right)\Sigma^{-1}}{\sum_{i=1}^{n}\psi_1(d_i^*)d_i^*}, \tag{A.5}$$

and denoting $w_i(v)=\psi_i(v)/v$ for $i=1,2$, we have

$$\frac{\partial d_i^*}{\partial\Sigma}=\frac{1}{2s}\Sigma^{-1}\left[-\frac{1}{d_i}\widehat{u}_i(B)\widehat{u}_i'(B)+\frac{d_i}{s^2}\frac{\sum_{j=1}^{n}w_1(d_j^*)\widehat{u}_j(B)\widehat{u}_j'(B)}{\sum_{j=1}^{n}\psi_1(d_j^*)d_j^*}\right]\Sigma^{-1}. \tag{A.6}$$

Replacing (A.4), (A.5) and (A.6) in (A.3), and using the fact that the τ-estimate satisfies (A.1), we obtain

$$\Sigma^{-1}-\frac{q}{s^2}\frac{\Sigma^{-1}\left(\sum_{i=1}^{n}w_1(d_i^*)\widehat{u}_i(B)\widehat{u}_i'(B)\right)\Sigma^{-1}}{\sum_{i=1}^{n}\psi_1(d_i^*)d_i^*}$$
$$-\frac{q\left[\sum_{i=1}^{n}w_2(d_i^*)\Sigma^{-1}\widehat{u}_i(B)\widehat{u}_i'(B)\Sigma^{-1}\right]}{2s^2\sum_{i=1}^{n}\rho_2(d_i^*)}$$
$$+\frac{q\sum_{i=1}^{n}\psi_2(d_i^*)d_i^*\left[\Sigma^{-1}\left(\sum_{k=1}^{n}w_1(d_k^*)\widehat{u}_k(B)\widehat{u}_k'(B)\right)\Sigma^{-1}\right]}{2s^2(\sum_{i=1}^{n}\rho_2(d_i^*))(\sum_{i=1}^{n}\psi_1(d_i^*)d_i^*)}=0.$$

Solving for Σ, we obtain

$$\Sigma = \frac{q}{s^2} \frac{\sum_{i=1}^{n} w_n^*(d_i^*)\widehat{u}_i(B)\widehat{u}_i'(B)}{\sum_{k=1}^{n} \psi_n^*(d_k^*)d_k^*},$$

and this proves (2.17). □

A.2. Theorem 2

Lemmas A.1–A.5 are required to prove Theorem 2

Lemma A.1. *Assume that ρ_1 and ρ_2 satisfy A1–A5. Given $m < \eta n$, there exists K such that for any $Z^* = \{z_1^*, \ldots, z_n^*\} \in \mathcal{Z}_m$, $z_i^* = (x_i^{*\prime}, y_i^{*\prime})'$,* we have $\tau(\|y_1^*\|, \ldots, \|y_n^*\|) < K$.

This lemma is proved in [16].

Lemma A.2. *Assume that ρ_1 and ρ_2 satisfy A1–A5, and let $\widetilde{\Sigma}$ be the τ-estimate of Σ. Given $m < \eta n$, there exist K^* such that for any $Z^* = \{z_1^*, \ldots, z_n^*\} \in \mathcal{Z}_m$, $z_i^* = (x_i^{*\prime}, y_i^{*\prime})'$* we have $\det(\widetilde{\Sigma}(Z^*)) < K^*$.

Proof. Take $B = 0$ and $\Sigma = (\tau^2(\|y_1^*\|, \ldots, \|y_n^*\|)/\kappa_2)I_q$. Observe that $d_i(B, \Sigma) = \sqrt{\kappa_2}\|y_i^*\|/\tau(\|y_1^*\|, \ldots, \|y_n^*\|)$, and therefore $\tau^2(d_1(B, \Sigma), \ldots, d_n(B, \Sigma)) = \kappa_2$. By Lemma A.1 there exists K such that

$$\det(\Sigma) = (\tau^2(\|y_1^*\|, \ldots, \|y_n^*\|)/\kappa_2)^q < (K^2/\kappa_2)^q.$$

Then using the definition of $(\widetilde{B}, \widetilde{\Sigma})$ given in (2.10) and (2.11), the lemma follows with $K^* = (K^2/\kappa_2)^q$. □

Lemma A.3. *Assume that ρ_1 and ρ_2 satisfy A1–A5. Given $K_1 > 0$ and $r > n\eta$, there exists K_2 such that for any sample $u_1, \ldots, u_n$ such that $\#\{i : |u_i| > K_2\} > r$, we have $\tau(u_1, \ldots, u_n) > K_1$.*

This lemma is proved in [16].

Lemma A.4. *Consider the same assumptions as in Theorem* 2 *and $m < \min((1-\eta)n - h, n\eta)$. Then* (i) $S_m^-(Z, \widetilde{\Sigma}) > 0$, (ii) $S_m^+(Z, \widetilde{\Sigma}) < \infty$.

Proof. Suppose that $S_m^-(Z, \widetilde{\Sigma}) = 0$. Then there exist $Z_j^* \in \mathcal{Z}_m$ such that $\lambda_1(\widetilde{\Sigma}(Z_j^*)) \to 0$. Put $\Sigma_j = \widetilde{\Sigma}(Z_j^*)$, $B_j = \widetilde{B}(Z_j^*)$ and $\lambda_{1j} = \lambda_1(\widetilde{\Sigma}(Z_j^*))$. Let U_j be an orthogonal matrix of eigenvectors of Σ_j and Λ_j the diagonal matrix with the corresponding eigenvalues. Let $Z_j^* = \{z_{j1}^*, \ldots, z_{jn}^*\}$, $z_{ji}^* = (x_{ji}^{*\prime}, y_{ji}^{*\prime})'$; then

$$\begin{aligned} d_{ji}^{*2} &= (y_{ji}^* - B_j' x_{ji}^*)'\Sigma_j^{-1}(y_{ji}^* - B_j' x_{ji}^*) \\ &= (U_j' y_{ji}^* - U_j' B_j' x_{ji}^*)'\Lambda_j^{-1}(U_j' y_{ji}^* - U_j' B_j' x_{ji}^*) \geqslant (e_j' z_{ji}^*)^2/\lambda_{1j}, \end{aligned} \tag{A.7}$$

where $e_j = (-v_{j1}, u_{j1})$, and u_{j1} and v_{j1} are, respectively, the first rows of U_j' and of $V_j = U_j' B_j'$.

Put $\delta = \inf_{\|a\|=1} \sup_{1 \leqslant j_1 < \cdots < j_{h+1} \leqslant n}\{|a' z_{j_1}|, \ldots, |a' z_{j_{h+1}}|\}$. Then by the definition of h, we have that $\delta > 0$. Since $\|e_j\| \geqslant 1$, there are at least $n - m - h > \eta n$ values d_{ji}^* larger than or equal to $\delta/(\lambda_{1_j})^{1/2}$. Therefore, from Lemma A.3 we obtain $\lim_{j\to\infty}\tau(d^*_{j1},\ldots,d^*_{jn}) = \infty$, which contradicts the definition of the τ-estimates. Then part (i) of the lemma follows.

Part (ii) follows from part (i) and Lemma A.2. □

Lemma A.5. *Consider the same assumptions as in Theorem 2. Then for any* $m < \min((1-\eta)n-h, n\eta)$ *we have* $S_m(Z,\widetilde{B}) < \infty$.

Proof. Assume that there exists $Z^*_j \in \mathcal{Z}_m$, such that $\|\widetilde{B}(Z^*_j)\| \to \infty$. Let U_j, Λ_j, V_j and δ be as in Lemma A.4. Since $\|V_j\| = \|U'_j B'_j\|$, we also have $\|V_j\| \to \infty$. Without loss of generality we can suppose that for some i_0, $\|v_{ji_0}\| \to \infty$, where v_{ji_0} is the i_0th row of V_j. According to Lemma A.4 we can assume that $\lambda_q(\widetilde{\Sigma}(Z_j)) < K$. Let $Z^*_j = \{z^*_{j1},\ldots,z^*_{jn}\}$, $z^*_{ji} = (x^{*\prime}_{ji}, y^{*\prime}_{ji})'$, then proceeding as in (A.7) we obtain $d^{*2}_{ji} \geqslant (1/K)(e'_j z^*_{ji})^2$, where $e_j = (-v_{ji_0}, u_{ji_0})$. Then there are at least $n-m-h > \eta n$ values d^*_{ji} larger than $(\delta/K^{1/2})\|v_{ji_0}\|$. Therefore by Lemma A.3, $\tau(d_{j1},\ldots,d_{jn}) \to \infty$, contradicting the definition of τ-estimates. □

Proof of Theorem 2. It follows from Lemmas A.4 and A.5.

A.3. Theorem 3

To prove Theorem 3 we need to introduce some notation and Lemmas A.6 and A.7 below.

Define $u(B) = u(B,x,y) = y - B'x$ and

$$d(B,\Sigma) = d(B,\Sigma,x,y) = (u'(B)\Sigma^{-1}u(B))^{1/2}.$$

Let (T_1,T_2) be the estimating functional corresponding to the τ-estimates $(\widetilde{B},\widetilde{\Sigma})$. Then, according to (2.16) and (2.17), given a distribution G of (x,y), $(T_1(G),T_2(G))$ are the values (B,Σ) satisfying

$$E_G\left(w^*(d^*(B,\Sigma,G))u(B)x'\right) = 0,$$

$$\Sigma = \frac{q E_G\left(w^*(d^*(B,\Sigma,G))u(B)u'(B)\right)}{(k^*(B,\Sigma,G))^2\, E_G\left(\psi^*(d^*(B,\Sigma,G))d^*(B,\Sigma,G)\right)},$$

where $d^*(B,\Sigma,G) = d(B,\Sigma)/k^*(B,\Sigma,G)$, $k^*(B,\Sigma,G) = s^*(H)$, H is the distribution of $d(B,\Sigma)$ under G and $s^*(H)$ is defined in (4.2).

Lemma A.6. *Suppose that we observe* $z \in R^m$ *with distribution function* $G_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}$, *where* $\theta_1 \in R^{k_1}$ *and* $\theta_2 \in R^{k_2}$. *Consider an M-estimating functional of* $\theta = (\theta_1,\theta_2)$, $T(G) = (T_1(G),T_2(G))$ *such that*

$$E_G(h(z,T_1(G),T_2(G),k(T_1(G),T_2(G),G))) = 0,$$

where $h: R^{m+k_1+k_2+1} \to R^{k_1}$ *is a differentiable function and* $k: R^{k_1+k_2}\times\mathcal{F} \to R$, *and* $\mathcal{F}$ *is the space of distributions on* $R^{k_1+k_2}$. *Suppose that T satisfies the following strong Fisher consistency condition*:

$$E_{G_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}}(h(z,\theta_1,\theta_2,k)) = 0 \quad \forall k, \tag{A.8}$$

and

$$E_{G_{\boldsymbol{\theta}_1,\boldsymbol{\theta}_2}}\left(h_3(z,\theta_1,\theta_2,k(\theta_1,\theta_2,G_{\theta_1,\theta_2}))\right) = 0, \tag{A.9}$$

where h_i, $1 \leqslant i \leqslant 4$, is the derivative of h with respect to the ith argument. Suppose too that $E_{G_{\theta_1,\theta_2}}\left(h(z,\theta_1,\theta_2,k(\theta_1,\theta_2,G_{\theta_1,\theta_2}))\right)$ can be differentiated inside the expectation. Then the influence function of T_1 is given by

$$IC(z_0,\mathbf{T}_1,\theta_1,\theta_2) = -\left(E_{G_{\theta_1,\theta_2}}\left(h_2(z,\theta_1,\theta_2,k(\theta_1,\theta_2,G_{\theta_1,\theta_2}))\right)\right)^{-1} \times h(z_0,\theta_1,\theta_2,k(\theta_1,\theta_2,G_{\theta_1,\theta_2})).$$

Proof. Let $G_\varepsilon = (1-\varepsilon)G_{\theta_1,\theta_2} + \varepsilon\delta_{z_0}$. Then $T(G_\varepsilon)$ satisfies

$$(1-\varepsilon)E_{G_{\theta_1,\theta_2}}(h(z,T_1(G_\varepsilon),T_2(G_\varepsilon),k\,(T_1(G_\varepsilon),T_2(G_\varepsilon),G_\varepsilon))) + \varepsilon h(z_0,T_1(G_\varepsilon),T_2(G_\varepsilon),k\,(T_1(G_\varepsilon),T_2(G_\varepsilon),G_\varepsilon)) = 0.$$

The proof of this lemma follows inmediately differentiating this expression with respect to ε at $\varepsilon = 0$ and using (A.8) and (A.9). □

Lemma A.7. *Consider assumptions A1–A9 and suppose $\Sigma_0 = I_q$. Then, if G is the distribution of (x, y), we have*

(a) $$E_G\left(\frac{\partial(\text{vec}(w^*(d(B_0,\Sigma_0)/k)u(B_0)x'))}{\partial\,(\text{vec}\,\Sigma)'}\right) = 0\ \forall k.$$

(b) $$E_G\left(\text{vec}(w^*(d(B_0,\Sigma_0)/k)u(B_0)x')\right) = 0\ \forall k.$$

Proof. (a) Straighforward computations lead to

$$E_G\left(\frac{\partial(\text{vec}(w^*(d(B_0,\Sigma_0)/k)u(B_0)x'))}{\partial\,(\text{vec}\Sigma)'}\right) = -\frac{1}{2k}E_G\left(\left(xu'\otimes I_q\right)\left(\text{vec}I_q \otimes \frac{w^{*\prime}(\|u\|/k)}{\|u\|}\text{vec}(uu'\right)'\right).$$

Since the distribution of u is assumed to be elliptical with $\Sigma_0 = I_q$, we have that for any function h, $E_G(x_j u_i u_k u_l h(\|u\|)) = 0$. Observe that all the elements of the right-hand side of the last equation are of this form, and therefore part (a) of the lemma follows immediately.

(b) This is derived from the fact that $E_G(x_j u_i h(\|u\|)) = 0$. □

Proof of Theorem 3. Suppose that $z = (x', y')'$ follows MLM (1.1) and let G_0 be its distribution function. Consider first the case $\Sigma_0 = I_q$; then $k^*(B_0, I_q, G_0) = k_0 = s^*(H_0)$, where H_0 is the distribution of $\|u\|$. Using Lemma A.6 with $\theta_1 = \text{vec}(B_0')$, $\theta_2 = \text{vec}(\Sigma_0)$, and $k\,(T_1, T_2, G_0) = k^*(T_1, T_2, G_0)$ and Lemma A.7 we obtain

$$IF(y_0,x_0,T_1,B_0,I_q) = -\left(\frac{\partial E_G\left(\text{vec}(w^*(d(B_0,I_q)/k_0)u(B_0)x')\right)}{\partial(\text{vec}(B'))'}\right)^{-1} \times\text{vec}\left(w^*\left(\frac{((y_0-B_0'x_0)'(y_0-B_0'x_0))^{1/2}}{k_0}\right)(y_0-B_0'x_0)x_0'\right).$$

As $\text{vec}(ABC) = (C'\otimes A)\text{vec}(B)$, to prove (4.4), it will be enough to show that

$$\frac{\partial E_{G_0}\left(\text{vec}(w^*(d(B_0,I_q)/k_0)u(B_0)x')\right)}{\partial(\text{vec}(B'))'} = -\frac{1}{q}E_{G_0}\left((q-1)w^*\left(\frac{d(B_0,I_q)}{k_0}\right)+\psi^{*\prime}\left(\frac{d(B_0,I_q)}{k_0}\right)\right)\left(E_{M_0}(xx')\otimes I_q\right). \quad \text{(A.10)}$$

Using that $d(B_0, I_q) = \|u\|$ and $u(B_0) = u$, we obtain

$$\begin{aligned}
&\frac{\partial E_{G_0}\left(\mathrm{vec}(w^*(d(B_0, I_q)/k_0)u(B_0)x')\right)}{\partial(\mathrm{vec}B')'} \\
&\quad = E_{G_0}\left(\frac{\partial\left(\mathrm{vec}(w^*(d(B_0, I_q)/k_0)u(B_0)x')\right)}{\partial(\mathrm{vec}B')'}\right) \\
&\quad = E_{G_0}\left[-w^*\left(\frac{\|u|}{k_0}|\right)(xx' \otimes I_q) + (xu' \otimes I_q)\right. \\
&\qquad \times\left\{\mathrm{vec}(I_q) \otimes \left[-\left(\frac{C\psi_1'(\|u\|/k_0)}{\|u\|/k_0} + \frac{D\psi_2'(\|u\|/k_0)}{\|u\|/k_0}\right.\right.\right. \\
&\qquad \left.\left.\left.\left. -\frac{\psi^*(\|u\|/k_0)}{\|u\|^2/k_0^2}\right)\frac{u'(x' \otimes I_q)}{k_0\|u\|}\right]\right\}\right].
\end{aligned} \tag{A.11}$$

Denoting

$$g(d) = -\left[\frac{C\psi_1'(d)}{d} + \frac{D\psi_2'(d)}{d} - \frac{\psi^*(d)}{d^2}\right]\frac{1}{k_0 d}$$

and using that $E_{F_0}\left[u_i u_j g(\|u\|/k_0)\right] = 0$, if $i \neq j$, we obtain

$$\begin{aligned}
&E_{G_0}\left(g\left(\frac{\|u\|}{k_0}\right)(xu' \otimes I_q)\left\{\mathrm{vec}(I_q) \otimes \left[u'(x' \otimes I_q)\right]\right\}\right) \\
&\quad = E_{M_0}(xx')E_{F_0}\left[u_i^2 g\left(\frac{\|u\|}{k_0}\right)\right] \otimes I_q \\
&\quad = E_{F_0}\left[u_i^2 g\left(\frac{\|u\|}{k_0}\right)\right]\left(E_{M_0}(xx') \otimes I_q\right).
\end{aligned} \tag{A.12}$$

We also have

$$E_{F_0}\left[u_i^2 g\left(\frac{\|u\|}{k_0}\right)\right] = \frac{1}{q}E_{F_0}\left[\|u\|^2 g\left(\frac{\|u\|}{k_0}\right)\right]. \tag{A.13}$$

Using (A.11), (A.12) and (A.13) we obtain (A.10), and then (4.4) holds for the case $\Sigma = I_q$.

Consider now the case of a general covariance matrix Σ. Take R such that $\Sigma = RR'$; then the errors u_i^* of the transformed model

$$y_i^* = R^{-1}y_i = R^{-1}B_0'x_i + R^{-1}u_i = B_0^{*\prime}x_i + u_i^*$$

have covariance matrix I_q. Then, (4.4) follows from the following relationship:

$$IF(y_0, x_0, T_1, B_0, \Sigma) = R\left[IF(R^{-1}y_0, x_0, T_1, B_0, I_q)\right]. \qquad \square$$

A.4. Theorem 4

Before proving Theorem 4 we need to introduce some notation and Lemmas A.8–A.12 below. Define the scale-estimating functional $\tau^*(H)$ by

$$\tau^{*2}(H) = s^{*2}(H)E_H\left(\rho_2\left(\frac{v}{s^*(H)}\right)\right),$$

where the functional $s^*(H)$ is defined as in (4.2). Observe that $\tau(u_1, \ldots, u_n) = \tau^*(H_n)$, where H_n is the empirical distribution of $u_1, \ldots, u_n$. Define for $a > 0$ $r_H(a) = a^2 E_H\left(\rho_2 (v/a)\right)$.

Lemma A.8. *Suppose that ρ_2 satisfies A1–A6. Then $r_H(a)$ is a nondecreasing function of a.*

Proof. Follows from $r'_H(a) = a\ E_H\left[2\rho_2\,(v/a) - \psi_2\,(v/a\,)\,v/a\right]$ and A6. □

Lemma A.9. *Suppose that ρ satisfies A1–A5 and that the distribution of u satisfies A7 with $\Sigma = \Sigma_0$. Let (μ, Σ) be such that* $\det(\Sigma) = \det(\Sigma_0)$ *and $(\mu, \Sigma) \neq (0, \Sigma_0)$. Then*

$$E(\rho((u-\mu)'\Sigma^{-1}(u-\mu))^{1/2}) > E(\rho(u'\Sigma_0^{-1}u)^{1/2}).$$

This lemma follows immediately from Theorem 1 of [4].

Lemma A.10. *Suppose that ρ satisfies A1–A5 and that the distribution of u satisfies A7 with $\Sigma = \Sigma_0$. Let (v, Σ) be such that v is a random variable independent of u,* $\det(\Sigma) = \det(\Sigma_0)$ *and either* (i) $P(v \neq 0) > 0$ *or* (ii) $\Sigma \neq \Sigma_0$. *Then*

$$E(\rho(((u-v)'\Sigma^{-1}(u-v))^{1/2}) > E(\rho(u'\Sigma_0^{-1}u)^{1/2}).$$

Proof. Suppose that (i) is true. Then, by Lemma A.9

$$\begin{aligned} E(\rho(((u-v)'\Sigma^{-1}(u-v))^{1/2}) \mid v = \mu) &= E(\rho((u-\mu)'\Sigma^{-1}(u-\mu))^{1/2}) \\ &\geqslant E(\rho(u'\Sigma_0^{-1}u)^{1/2}), \end{aligned}$$

and the inequality is strict with probability larger than 0. Then the lemma follows. The proof is similar when (ii) holds. □

Lemma A.11. *Suppose that ρ_1 and ρ_2 satisfy A1–A6 and that the distribution of u satisfies A7 with $\Sigma = \Sigma_0$. Let (v, Σ) be such that v is a random variable independent of u,* $\det(\Sigma) = \det(\lambda\Sigma_0)$ *and either* (i) $P(v \neq 0) > 0$ *or* (ii) $\Sigma \neq \lambda\Sigma_0$. *Let H^* be the distribution of $((u-v)'\Sigma^{-1}(u-v))^{1/2}$ and H the distribution of $(u'\Sigma_0^{-1}u)^{1/2}/\lambda$; then $\tau^{*2}(H^*) > \tau^{*2}(H)$.*

Proof. Since $\det(\Sigma) = \det(\lambda\Sigma_0)$, by Lemma A.10, taking as $\rho(v) = \rho_1(v/(\lambda s^*(H)))$, we obtain

$$E\left(\rho_1\left(\frac{((u-v)'(\Sigma/\lambda)^{-1}(u-v))^{1/2}}{\lambda s^*(H)}\right)\right) > E\left(\rho_1\left(\frac{(u'\Sigma_0{}^{-1}u)^{1/2}}{\lambda s^*(H)}\right)\right),$$

and therefore $s^*(H^*) > s^*(H)$.

Using Lemmas A.7 and A.9, and since $\det(\Sigma/\lambda) = \det(\Sigma_0)$, we obtain

$$\begin{aligned} \tau^{*2}(H^*) &= s^{*2}(H^*)E\left(\rho_2\left(\frac{((u-v)'(\Sigma/\lambda)^{-1}(u-v))^{1/2}}{\lambda s^*(H^*)}\right)\right) \\ &\geqslant s^{*2}(H)E\left(\rho_2\left(\frac{((u-v)'(\Sigma/\lambda)^{-1}(u-v))^{1/2}}{\lambda s^*(H)}\right)\right) \\ &> s^{*2}(H)E\left(\rho_2\left(\frac{(u'\Sigma_0^{-1}u)^{1/2}}{\lambda s^*(H)}\right)\right) = \tau^{*2}(H). \end{aligned}$$

This proves the lemma. □

Let x and y be two arbitrary random vectors of dimension p and q, respectively, and let G be any arbitrary joint distribution of (x, y). Given B and Σ, let $H_{G,B,\Sigma}$ be the distribution of $d(B, \Sigma)$ when (x, y) has distribution G and define $\tau^*_G(B, \Sigma) = \tau^*(H_{G,B,\Sigma})$. Given observations (x_i, y_i), $1 \leqslant i \leqslant n$, let G_n be their empirical distribution. Then, the definition of the τ-estimate $(\widetilde{B}, \widetilde{\Sigma})$ given in Section 2 is equivalent to

$$(\widetilde{B}, \widetilde{\Sigma}) = \arg\min \det(\Sigma) \tag{A.14}$$

subject to

$$\tau^{*2}(H_{G_n,B,\Sigma}) = \kappa_2. \tag{A.15}$$

The following lemma shows that, if we replace in (A.15) G_n for the true distribution, then the minimum of $\det(\Sigma)$ is attained at $(B_0, (k_0^2/\sigma_0^2)\Sigma_0)$.

Lemma A.12. *Suppose that the random vector* (x, y) *follows an MLM with parameters* B_0 *and* Σ_0 *where the error* u *has a distribution satisfying A7. Suppose also that* ρ_k, $k = 1, 2$ *satisfy A1–A6 and let* G_0 *be the joint distribution of* (x, y). *Then, the problem of finding* (B, Σ) *minimizing* $\det(\Sigma)$ *subject to* $\tau^{*2}(H_{G_0,B,\Sigma}) = \kappa_2$ *has* $(B_0, (k_0^2/\kappa_2)\Sigma_0)$ *as the unique solution.*

Proof. Put $\Sigma_0^* = (k_0^2/\sigma_0^2)\Sigma_0$. Since $u(B_0) = u$, then $d(B_0, \Sigma_0^*) = (\sigma_0/k_0)(u'\Sigma_0^{-1}u)^{1/2}$. Because of (4.2), we obtain $s^*(H_{G_0B_0\Sigma_0^*}) = \sigma_0$ and from (5.1), $\tau^{*2}(H_{G_0B_0\Sigma_0^*}) = \kappa_2$. Now, take $(B, \Sigma) \neq (B_0, \Sigma_0^*)$ such that $\tau^{*2}(H_{G_0B\Sigma}) = \kappa_2$. If $B \neq B_0$, then $u(B) = u - (B_0' - B')x$. Put $\lambda = (\det(\Sigma_0^*)/\det(\Sigma))^{1/q}$, and $\Sigma^* = \lambda\Sigma$. Then $\det(\Sigma^*) = \det(\Sigma_0)$, and since x is independent of u and $P((B_0' - B')x \neq 0) > 0$, by Lemma A.11 we have

$$\kappa_2 = \tau^{*2}(H_{G_0B_0\Sigma_0^*}) < \tau^{*2}(H_{G_0B\Sigma^*}) = \frac{\tau^{*2}(H_{G_0B\Sigma})}{\lambda} = \frac{\kappa_2}{\lambda}.$$

Then $\lambda < 1$ and $\det(\Sigma_0^*) < \det(\Sigma)$, proving the lemma. □

Heuristic proof of Theorem 4. The theorem follows, using standard arguments, from (A.15), (A.14), Lemma A.12 and the fact that the empirical distribution G_n converges a.s. to G_0 (see [11]).

A.5. Theorem 5

Let $T(F)$ be an estimating operator with values in R^m, and let F_n be the empirical distribution based on a random sample of size n with an underlying distribution F. Then, under suitable differentiability conditions $n^{1/2}(T(F_n)-T(F)) \to_D N(0, E_F(IF(x, T, F)IF(x, T, F)'))$, where $IF(x, T, F)$ is the influence function of T at the point x and at the distribution F.

Heuristic proof of Theorem 5. Let us first consider the case $\Sigma_0 = I_q$ and $E_{M_0}(xx') = I_p$. In this case

$$IF(y, x, T_1, B_0, I_q) = c_0 w^*_{H_0}\left(\frac{\|u\|}{k_0}\right) xu',$$

where H_0 is the distribution of $\|u\|$ and we obtain

$$
\begin{aligned}
&E(\mathrm{vec}(IF(y, x, T_1, B_0, I_q))\mathrm{vec}(IF(y, x, T_1, B_0, I_q))') \\
&= c_0^2 E_{F_0}\left(\frac{k_0^2\psi_H^{*2}\left(\frac{\|u\|}{k_0}\right)}{\|u\|^2}uu'\right) \otimes E_{M_0}(xx').
\end{aligned}
$$

As the distribution of u is assumed to be elliptical with $\Sigma_0 = I_q$, for any function h, $E_{F_0}(h(\|u\|)u_i u_j) = 0$ if $i \neq j$ and $E_{F_0}(h(\|u\|)u_i^2) = E_{F_0}(h(\|u\|)\,\|u\|^2)/q$. Then

$$
V = \frac{c_0^2 k_0^2 E_{H_0}(\psi_{H_0}^{*2}(\frac{v}{k_0}))}{q} I_q \otimes I_p.
$$

For the general case, let R and T be matrices such that $\Sigma_0 = RR'$ and $E_{M_0}(xx') = TT'$, and consider the following transformation $y^* = R^{-1}y$ and $x^* = T^{-1}x$. Then if $B_0^* = T'B_0R'^{-1}$, $y^* = B_0^{*\prime}x^* + u^*$, with $u^* = R^{-1}u$. Since the distribution of u^* is given by (4.1) with $\Sigma = I_q$ and $E(x^*x^{*\prime}) = I_q$, (5.2) follows from the equivariance of the τ-estimates and the fact that

$$
\mathrm{vec}(B_0) = \mathrm{vec}((T^{-1})'B_0^*R') = \left(R \otimes \left(T^{-1}\right)'\right)\mathrm{vec}(B_0^*).
$$

References

[1] J. Agullo, C. Croux, S. Van Aelst, The multivariate least trimmed squares estimator, Research Report 0224, K.U. Leuven D.T.E.W, 2002.

[2] M. Bilodeau, P. Duchesne, Robust estimation of the SUR model, Canad. J. Statist. 28 (2000) 277–288.

[3] C. Croux, Efficient high-breakdown M-estimators of scale, Statist. Probab. Lett. 19 (1994) 371–379.

[4] P.L. Davies, Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices, Ann. Statist. 15 (1987) 1269–1292.

[5] D.L. Donoho, P.J. Huber, The notion of breakdown point, A Festschrift for Erich L. Lehmann, 1983, pp. 157–184.

[6] F.R. Hampel, A general qualitative definition of robustness, Ann. Math. Statist. 42 (1971) 1887–1896.

[7] F.R. Hampel, The influence curve and its role in robust estimation, J. Amer. Statist. Assoc. 69 (1974) 383–393.

[8] O. Hossjer, On the optimality of S-estimators, Statist. Probab. Lett. 14 (1992) 413–419.

[9] P.J. Huber, Robust Statistics, Wiley, New York, 1981.

[10] R. Koenker, S. Portnoy, M-estimation of multivariate regressions, J. Amer. Statist. Assoc. 85 (1990) 1060–1068.

[11] H.P. Lopuhaä, Multivariate τ-estimators for location and scatter, Canad. J. Statist. 19 (1991) 307–321.

[12] D.W. Nieremberg, T.A. Stukel, J.A. Baron, B.J. Dain, E.R. Greenberg, Determinants of plasma levels of beta-carotene and retinol, Amer. J. Epidemiol. 130 (1989) 511–521.

[13] P.L. 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.

[14] P. Rousseeuw, K. Van Driessen, A fast algorithm for the minimum covariance determinant estimator, Technometrics 41 (1999) 221, 223.

[15] P.J. Rousseeuw, K. Van Driessen, S. Van Aelst, J. Agulló, Robust multivariate regression, Technometrics 46 (2004) 293–305.

[16] V.J. Yohai, R.H. Zamar, High breakdown point estimates of regression by means of the minimization of an efficient scale, Technical Report No. 84, Department of Statistics, University of Washington, 1986.

[17] V.J. Yohai, R.H. Zamar, High breakdown-point estimates of regression by means of the minimization of an efficient scale, J. Amer. Statist. Assoc. 83 (1988) 406–413.