Pak. J. Statist.
2010, Vol. 26(1), 1-13

M-PROCEDURES IN THE GENERAL MULTIVARIATE NONLINEAR REGRESSION MODEL

Víctor Leiva[1], **Antonio Sanhueza**[2], **Pranab K. Sen**[3], and **Nelson Araneda**[2]

[1] Departamento de Estadistica, CIMFAV, Universidad de Valparaiso, Valparaiso, Chile
Email: victor.leiva@uv.cl; victor.leiva@yahoo.com
[2] Departamento de Matematica y Estadistica, Universidad de La Frontera, Temuco, Chile
Email: asanhueza@ufro.cl, naraneda@ufro.cl
[3] Department of Biostatistics, University of North Carolina, Chapel Hill, USA
Email: pksen@bios.unc.edu

ABSTRACT

In the multivariate nonlinear regression model, parameter estimators and test statistics based on least squares and maximum likelihood methods are usually nonrobust. For this type of models, we introduce M-estimators and M-tests, which are robust to departures from normality. In addition, we study the asymptotic properties and consider a computational algorithm for these estimators.

KEYWORDS

Asymptotic normality; efficiency; M-estimators; M-tests; uniform asymptotic linearity

2000 Mathematics Subject Classification: Primary 62J02; Secondary 62G35.

1 INTRODUCTION

Modeling of data in typical biomedical studies, including bioassay, clinical research, pharmacokinetics, pharmacology, and toxicology, usually involves a characterization of the relationship between the response variable and covariables (auxiliary or explanatory variables). In many applications, the proposed systematic relationship between these variables is nonlinear (in unknown parameters of interest). In some cases, the relevant nonlinear model may be derived from physical or mechanistic grounds. In other contexts, a nonlinear model may be used simply to provide an empirical description of the data. For example, pharmacologists and toxicologists are interested in assessing antagonistic, derivative or synergistic relationships for drugs in different treatments. Response surface methods have been successfully used to describe and explore these relationships between a response variable and a set of covariables. In a typical application, various realizations of the response are statistically independent, in which case univariate techniques are employed. However, there are applications where the assumption of independence is not valid, as for example, when subjects receive all the treatments or all the treatment dose levels, so that the use of a multivariate design is the most appropriate.

In nonlinear models, most of the procedures of estimation and tests of hypotheses are basically based on least squared (LS) or maximum likelihood (ML) methods; see Galant (1987) and Seber and Wild (1981). These methods usually are sensitive to departures from normality due to the possible presence of outliers. These departures have led to various proposals for robust methods as are the M, L, and R estimators, which have been discussed by Huber (1973), Bickel (1973), and Jaeckel (1972), respectively. M-estimators (ML type) include LS and ML methods as special cases and also allow for the construction of estimators and tests that are robust against departures from normality. Maronna (1976) proposed a simultaneous M-estimation procedure for the multivariate location and scatter parameters, under the assumption of elliptical symmetry of the underlying distribution. For the multivariate linear model, M-procedures have been proposed for estimating its parameters and testing hypotheses about these parameters. Singer and Sen (1985) developed two types of M-estimators for multivariate linear models, coordinatewise M-estimators and an extension of the Maronna-type M-estimator. For more details about M procedures, see Sen and Singer (1993) and Sanhueza and Sen (2001, 2004).

The objective of this work is to develop M-procedures for the general multivariate nonlinear regression model. This study may be seen as a generalization of the robust methods for nonlinear models with repeated measurements presented by Sanhueza et al. (2009).

In Section 2, we define M-estimators for the parameters of the general multivariate nonlinear regression model. Here, we also present the notation and regularity conditions necessary to derive asymptotic results for the proposed M-procedures. In Section 3, we describe asymptotic results related to the proposed M-estimators by introducing first the uniform asymptotic linearity on these estimators. In Section 4, we develop a methodology to test hypothesis on the parameters of interest by using M-tests. Finally, in Section 5, we discussed an iterative computational method for computing the proposed M-estimators and the proofs of the theorems.

2 M-ESTIMATION

Consider the general multivariate nonlinear regression model given by

$$\mathbf{Y}_i = \mathbf{f}(\mathbf{x}_i, \beta) + \varepsilon_i,\ i = 1, \ldots, n, \tag{2.1}$$

where $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{ip})^\top$ is the response vector, $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})^\top$ is the covariable vector, $\beta = (\beta_1, \ldots, \beta_s)^\top$ is the coefficient vector, $\mathbf{f}(\mathbf{x}_i, \beta) = \big(f_1(\mathbf{x}_i, \beta), \ldots, f_p(\mathbf{x}_i, \beta)\big)^\top$ is a vector of specified functions that are nonlinear in β, and $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{ip})^\top$ is the error term vector, which has components to be independent and identically distributed with a cumulative distribution function (cdf) $G(\cdot)$ defined on $\mathbb{R}^p$, with mean vector zero and variance covariance matrix Σ. The cdf $G(\cdot)$ is assumed to be continuous and symmetric about 0, though its functional form may be unknown.

The M-estimator of β in Equation (2.1) is obtained by

$$\widehat{\beta}_n = \arg\left(\min_{\beta}\left\{\sum_{i=1}^{n}\mathbf{h}^\top\big(\varepsilon_i(\beta)\big)\,\Delta^{-1}\,\mathbf{h}\big(\varepsilon_i(\beta)\big)\right\}\right), \tag{2.2}$$

where $\arg(\min u(\beta))$ denotes that $\widehat{\beta}_n$ is the value of the parameter vector β that minimizes the function $u(\beta)$ and, for $i = 1, \ldots, n$ and $j = 1, \ldots, p$,

$$\begin{aligned}
\varepsilon_i(\beta) &= \big(\varepsilon_{i1}(\beta), \ldots, \varepsilon_{ip}(\beta)\big)^\top, \\
\varepsilon_{ij}(\beta) &= Y_{ij} - f_j(\mathbf{x}_i, \beta), \\
\mathbf{h}\big(\varepsilon_i(\beta)\big) &= \big(h_1(\varepsilon_{i1}(\beta)), \ldots, h_p(\varepsilon_{ip}(\beta))\big)^\top, \text{ and} \\
\Delta &= E\Big\{\mathbf{h}\big(\varepsilon_i(\beta)\big)\,\mathbf{h}^\top\big(\varepsilon_i(\beta)\big)\Big\} - E\Big\{\mathbf{h}\big(\varepsilon_i(\beta)\big)\Big\}\,E^\top\Big\{\mathbf{h}\big(\varepsilon_i(\beta)\big)\Big\},
\end{aligned} \tag{2.3}$$

with Δ being a variance-covariance, positive definite (p.d.) matrix.

In particular, if we let $h_j(z) = z$, for $j = 1, \ldots, p$, in Equation (2.2), we have the LS estimator of β. Incorporating the conventional setup of robust methods (see Huber, 1981, Hampel et al., 1986, and Jurečková and Sen, 1996), we mostly use bounded and monotone functions $h_j(\cdot)$ such as

$$h(z) = \begin{cases} \dfrac{z}{\sqrt{2}}, & \text{if } k_1 \le z \le k_2, \\ \sqrt{k_1\left[z - \frac{k_1}{2}\right]}, & \text{if } z \le k_1\,(<0), \\ \sqrt{k_2\left[z - \frac{k_2}{2}\right]}, & \text{if } z \ge k_2\,(>0), \end{cases}$$

where k_1 and k_2 are constants suitably chosen.

Theorem 1. *The minimization in Equation (2.2) is equivalent to solve*

$$\sum_{i=1}^{n} \lambda(\mathbf{x}_i, \mathbf{Y}_i, \widehat{\beta}_n, \Delta) = \mathbf{0}, \tag{2.4}$$

which we call robust estimating equations, where $\lambda(\mathbf{x}_i, \mathbf{Y}_i, \beta, \Delta) = \mathbf{X}_i^\top(\beta)\,\Delta^{-1}\,\Psi\big(\varepsilon_i(\beta)\big)$ *and*

$$\Psi\big(\varepsilon_i(\beta)\big) = \begin{pmatrix} \psi_1(\varepsilon_{i1}(\beta)) \\ \vdots \\ \psi_p(\varepsilon_{ip}(\beta)) \end{pmatrix},$$

with $\mathbf{X}_i(\beta) = (\partial/\partial\beta^\top)\mathbf{f}(\mathbf{x}_i, \beta)$, $\psi_j(\varepsilon_{ij}(\beta)) = 2\,h_j(\varepsilon_{ij}(\beta))\,h_j'(\varepsilon_{ij}(\beta))$, *and* $h_j'(\cdot)$ *denoting the derivative of* $h_j(\cdot)$, *for* $i = 1, \ldots, n$ *and* $j = 1, \ldots, p$.

We assume the following regularity conditions for: [RC1] the cdf $G(\cdot)$, [RC2] the score function $\psi(\cdot)$, and [RC3] the function $\mathbf{f}(\cdot)$.

[RC1]: $G(\cdot)$ is absolutely continuous with density $g(\cdot)$ having a finite Fisher information, i.e., $\int_{-\infty}^{+\infty} [g'(e)/g(e)]^2\, dG(e) < \infty$, where $g'(\cdot)$ denotes the derivative of $g(\cdot)$.

[RC21]: for $j = 1, \ldots, p$, $\psi_j(\cdot)$ is nonconstant, absolutely continuous and differentiable with respect to β_k, for $k = 1, \ldots, s$.

[RC22]: for $j, k = 1, \ldots, p$, we have that

i) $E\left\{\psi_j\big(\varepsilon(\beta)\big)\,\psi_k\big(\varepsilon(\beta)\big)\right\} = \phi_{jk} < \infty$ and $E\{\psi_j\big(\varepsilon(\beta)\big)\} = 0$.

ii) $E\left\{\big(\psi_j'(\varepsilon(\beta))\big)^2\right\} < \infty$ and $E\left\{\psi_j'\big(\varepsilon(\beta)\big)\right\} = \gamma_j \neq 0$, where $\psi'(\cdot)$ is the derivative of $\psi(\cdot)$.

[RC23]: for $j = 1, \ldots, p$ and $\|A\|$ denoting the norm of A, we have that

i) $\lim_{\delta\to 0} E\left\{\sup_{\|\Delta\|\le\delta} \big\|\,\psi_j(\varepsilon(\beta+\Delta)) - \psi_j\big(\varepsilon(\beta)\big)\big\|\right\} = 0$.

ii) $\lim_{\delta\to 0} E\left\{\sup_{\|\Delta\|\le\delta} \left\|\psi_j'(\varepsilon(\beta+\Delta)) - \psi_j'\big(\varepsilon(\beta)\big)\right\|\right\} = 0$.

[RC31]: $\mathbf{f}(\mathbf{x}, \beta)$ is continuous and twice differentiable with respect to $\beta \in \Theta$, where Θ is a compact subset of $\mathbb{R}^s$.

[RC32]: for $j, k = 1, \ldots, s$ and $l = 1, \ldots, p$, we have that

i)

$$\lim_{\delta\to 0} \sup_{\|\Delta\|\le\delta} \left\| \left(\frac{\partial f_l(\mathbf{x}, \beta+\Delta)}{\partial\beta_j}\right)\left(\frac{\partial f_l(\mathbf{x}, \beta+\Delta)}{\partial\beta_k}\right) - \left(\frac{\partial f_l(\mathbf{x}, \beta)}{\partial\beta_j}\right)\left(\frac{\partial f_l(\mathbf{x}, \beta)}{\partial\beta_k}\right)\right\| = 0.$$

ii)

$$\lim_{\delta\to 0} \sup_{\|\Delta\|\le\delta} \left\| \frac{\partial^2 f_l(\mathbf{x}, \beta+\Delta)}{\partial\beta_j\,\partial\beta_k} - \frac{\partial^2 f_l(\mathbf{x}, \beta)}{\partial\beta_j\,\partial\beta_k}\right\| = 0.$$

3 ASYMPTOTIC RESULTS

We first prove the asymptotic linearity of the proposed M-estimator, assuming that the variance-covariance matrix Δ in Equation (2.3) is known and, when it is unknown, we consider an estimate of it. Then, we prove the consistency and develop the asymptotic normality of this estimator.

Theorem 2. *Under the conditions [RC1], [RC21], [RC22], [RC23], [RC31], and [RC32], the following hold:*

i)

$$\sup_{\|\mathbf{t}\|\leq C}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta\right)\right\}+\frac{1}{n}\Gamma_n(\beta)\,\mathbf{t}\right\|=o_p(1), \tag{3.1}$$

where $\Gamma_n(\beta)=\sum_{i=1}^{n}\{\mathbf{X}_i^{\top}(\beta)\,\Delta^{-1}\,\mathbf{W}\,\mathbf{X}_i(\beta)\}$, $\mathbf{W}=diag\{\gamma_1,\ldots,\gamma_p\}$, *and* $\gamma_j=\int\psi_j'(e)\,f(e)\,de$, *for* $j=1,\ldots,p$, *as* $n\to\infty$.

ii)

$$\sup_{\substack{\|\mathbf{t}\|\leq C\\ \|vec(\mathbf{K})\|\leq D}}\left\|\frac{\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta+\tfrac{1}{\sqrt{n}}\mathbf{K}\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)\right\}}{\sqrt{n}}+\frac{1}{n}\Gamma_n\,\mathbf{t}\right\|=o_p(1), \tag{3.2}$$

where $\mathbf{K}$ *is a p.d. matrix, as* $n\to\infty$.

iii) Given a sequence $\widehat{\beta}_n$ *of solutions of Equation (2.2),*

$$\sqrt{n}\left\|\widehat{\beta}_n-\beta\right\|=O_p(1),\ as\ n\to\infty, \tag{3.3}$$

and

$$\widehat{\beta}_n=\beta+\Gamma_n^{-1}(\beta)\sum_{i=1}^{n}\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)+o_p\left(1/\sqrt{n}\right). \tag{3.4}$$

iv) As the minimum characteristic value of $\Lambda_i\,(\sum_{i=1}^{n}\Lambda_i)^{-1}$, *for* $i=1,\ldots,n$ *goes to 0 and* $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)\to\mathrm{N}_s\left(\mathbf{0},\,\Lambda\right),\quad where \tag{3.5}$$

$$\begin{aligned}
\Lambda &= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \Lambda_i \textit{ is a p.d. matrix},\\
\Lambda_i &= \mathbf{X}_i^\top(\beta)\,\Delta^{-1}\,\Phi\,\Delta^{-1}\,\mathbf{X}_i(\beta), \textit{ for } i=1,\ldots,n,\\
\Phi &= Var\left\{\Psi\big(\varepsilon(\beta)\big)\right\} \textit{ is a p.d. matrix and}\\
\phi_{kl} &= E\left\{\psi_k\big(\varepsilon_{ik}(\beta)\big)\,\psi_l\big(\varepsilon_{il}(\beta)\big)\right\}, \textit{ for } k,l=1,\ldots,p.
\end{aligned}$$

Corollary 3.1. Under the conditions of Theorem 1, the following hold:

i)

$$\sqrt{n}\left(\widehat{\beta}_n-\beta\right)\to \mathrm{N}_s\left(\mathbf{0},\,\Gamma^{-1}\,\Lambda\,\Gamma^{-1}\right), \tag{3.6}$$

where

$$\begin{aligned}
&\Gamma = \lim_{n\to\infty}\frac{\Gamma_n(\beta)}{n} \text{ is a p.d. matrix},\\
&\Gamma_n(\beta) = \sum_{i=1}^{n}\left\{\mathbf{X}_i^\top(\beta)\,\Delta^{-1}\,\mathbf{W}\,\mathbf{X}_i(\beta)\right\},\\
&\mathbf{W} = \mathrm{diag}\{\gamma_1,\ldots,\gamma_p\}, \text{ and}\\
&\gamma_j = \int \psi_j'(e)\,f(e)\,de, \text{ for } j=1,\ldots,p.
\end{aligned}$$

ii)

$$\widehat{\Omega}^{-1/2}\,\sqrt{n}\left(\widehat{\beta}_n-\beta\right)\to \mathrm{N}_s\left(\mathbf{0},\mathbf{I}_s\right), \tag{3.7}$$

where

$$\begin{aligned}
\widehat{\Omega} &= n\,\widehat{\Gamma}_n^{-1}(\widehat{\beta}_n)\,\widehat{\Lambda}_n\,\widehat{\Gamma}_n^{-1}(\widehat{\beta}_n),\\
\widehat{\Gamma}_n(\widehat{\beta}_n) &= \sum_{i=1}^{n}\left\{\mathbf{X}_i^\top(\widehat{\beta}_n)\,\widehat{\Delta}^{-1}\,\widehat{\mathbf{W}}\,\mathbf{X}_i(\widehat{\beta}_n)\right\},\\
\widehat{\Lambda}_n &= \sum_{i=1}^{n}\left\{\mathbf{X}_i^\top(\widehat{\beta}_n)\,\widehat{\Delta}^{-1}\,\widehat{\Phi}\,\widehat{\Delta}^{-1}\,\mathbf{X}_i(\widehat{\beta}_n)\right\},\\
\widehat{\Delta} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\mathbf{h}(\varepsilon_i(\widehat{\beta}_n))\;\mathbf{h}^\top(\varepsilon_i(\widehat{\beta}_n))\right\},\\
\widehat{\Phi} &= \frac{1}{n}\sum_{i=1}^{n}\left\{\Psi(\varepsilon_i(\widehat{\beta}_n))\;\Psi^\top(\varepsilon_i(\widehat{\beta}_n))\right\},\\
\widehat{\mathbf{W}} &= \mathrm{diag}\{\widehat{\gamma}_1,\ldots,\widehat{\gamma}_p\}, \text{ and}\\
\widehat{\gamma}_j &= \frac{1}{n}\sum_{i=1}^{n}\psi_j'(\varepsilon_{ij}(\widehat{\beta}_n)).
\end{aligned}$$

iii)

$$n\left(\widehat{\beta}_n-\beta\right)^\top\,\widehat{\Omega}^{-1}\,\left(\widehat{\beta}_n-\beta\right)\to\chi_s^2. \tag{3.8}$$

4 M-TESTS

We present hypothesis testing for the parameters β of the model given in Equation (2.1) based on M-tests and their asymptotic distributions. This parameter vector β can be written as

$$\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix},$$

where β_1 is an $r \times 1$ vector and β_2 is an $(s-r) \times 1$ vector, and $\widehat{\Omega}^{-1}$ given in Equation (3.7) can be written as

$$\widehat{\Omega}^{-1} = \begin{pmatrix} \widehat{\Omega}^{11} & \widehat{\Omega}^{12} \\ \widehat{\Omega}^{21} & \widehat{\Omega}^{22} \end{pmatrix}.$$

By using Cochran's theorem, we prove that

$$n\left(\widehat{\beta}_{1_n} - \beta_1\right) \widehat{\Omega}^{11} \left(\widehat{\beta}_{1_n} - \beta_1\right) \to \chi^2_r,$$

where $\widehat{\beta}_{1_n}$ is an M-estimator of β_1.

In order to test the hypotheses

$$H_o : \beta_1 = 0 \text{ vs. } H_1 : \beta_1 \neq 0,$$

we consider the Wald-type M-test defined as

$$W = n \widehat{\beta}_{1_n}^{\top} \widehat{\Omega}^{11} \widehat{\beta}_{1_n},$$

which under H_o follows asymptotically a χ^2_r distribution.

Now, we consider the nonlinear hypotheses

$$H_o : a(\beta) = 0 \text{ vs. } H_1 : a(\beta) \neq 0, \tag{4.1}$$

where $a(\cdot)$ is a real valued function that is nonlinear in β. By using Corollary 1 and the Delta method, we have that

$$\left\{A^{\top}(\widehat{\beta}_n)\ \widehat{\Omega}\ A(\widehat{\beta}_n)\right\}^{-1/2} \sqrt{n}\left(a(\widehat{\beta}_n) - a(\beta)\right) \to \mathrm{N}(0,1),$$

where $A(\widehat{\beta}) = (\partial/\partial\beta)a(\beta)$ evaluated at $\widehat{\beta}$. Once again using Cochran's theorem, we have

$$n\left(a(\widehat{\beta}_n) - a(\beta)\right)^{\top} \left\{A^{\top}(\widehat{\beta}_n)\ \widehat{\Omega}\ A(\widehat{\beta}_n)\right\}^{-1} \left(a(\widehat{\beta}_n) - a(\beta)\right) \to \chi^2_1.$$

From these last results, for testing the nonlinear hypothesis given in Equation (4.1), we can use the Wald-type M-test given by

$$W = n\,a^\top(\widehat{\beta}_n)\,\left\{A^\top(\widehat{\beta}_n)\;\widehat{\Omega}\;A(\widehat{\beta}_n)\right\}^{-1}\,\big(a(\widehat{\beta}_n)\big),$$

which under H_o follows asymptotically a χ^2_1 distribution.

Also, we can consider the q-dimensional nonlinear hypotheses

$$H_o : \mathbf{d}(\beta) = \mathbf{0} \text{ vs. } H_1 : \mathbf{d}(\beta) \neq \mathbf{0},$$

where $\mathbf{d} : \Re^s \to \Re^q$, such that $\mathbf{D}(\beta) = (\partial/\partial\beta^\top)\mathbf{d}(\beta)$ exists. By using $\widehat{\Omega}^{-1/2}\,\sqrt{n}\,(\widehat{\beta}_n - \beta)$ in Corollary 3.1(ii), we prove that

$$\left\{\mathbf{D}(\widehat{\beta}_n)\,\widehat{\Omega}\,\mathbf{D}^\top(\widehat{\beta}_n)\right\}^{-1/2}\,\sqrt{n}\,\big(\mathbf{d}(\widehat{\beta}_n) - \mathbf{d}(\beta)\big) \to \mathrm{N}_q\,(\mathbf{0}, \mathbf{I}_q)$$

and

$$n\,\big(\mathbf{d}(\widehat{\beta}_n) - \mathbf{d}(\beta)\big)^\top\,\left\{\mathbf{D}(\widehat{\beta}_n)\,\widehat{\Omega}\,\mathbf{D}^\top(\widehat{\beta}_n)\right\}^{-1}\,\big(\mathbf{d}(\widehat{\beta}_n) - \mathbf{d}(\beta)\big) \to \chi^2_q.$$

Thus, to test the nonlinear hypothesis, we use once again the Wald-type M-test given by

$$W = n\,\mathbf{d}^\top(\widehat{\beta}_n)\,\left\{\mathbf{D}(\widehat{\beta}_n)\,\widehat{\Omega}\,\mathbf{D}^\top(\widehat{\beta}_n)\right\}^{-1}\,\mathbf{d}(\widehat{\beta}_n),$$

which under H_o follows asymptotically a χ^2_q distribution. We also may define a modified Wald-type M-test given by

$$F = \frac{(N-s)\,W}{q\,\mathrm{SSE}(\widehat{\beta}_n, \widehat{\Delta})},$$

which under H_o follows asymptotically a $F(q, N-s)$ distribution, where $N = np$ is the total number of observations and

$$\mathrm{SSE}(\widehat{\beta}_n, \widehat{\Delta}) = \sum_{i=1}^{n}\left\{\mathbf{h}^\top(\varepsilon_i(\widehat{\beta}_n))\,\widehat{\Delta}^{-1}\,\mathbf{h}(\varepsilon_i(\widehat{\beta}_n))\right\}.$$

We also can define a likelihood ratio-type M-test based on

$$Q = \frac{\left\{\mathrm{SSE}(\widehat{\beta}_n^{\text{rest}}, \widehat{\Delta}) - \mathrm{SSE}(\widehat{\beta}_n^{\text{full}}, \widehat{\Delta})\right\}/q}{\mathrm{SSE}(\widehat{\beta}_n^{\text{full}}, \widehat{\Delta})/(N-s)}. \tag{4.2}$$

Here $\widehat{\beta}_n^{\text{rest}}$ is a restricted estimator obtained by minimizing Equation (2.2) subject to the constraint $\mathbf{d}(\beta) = \mathbf{0}$ and $\widehat{\beta}_n^{\text{full}}$ is the unrestricted estimator under the full model. Both estimators are obtained using the unrestricted variance-covariance estimator, $\widehat{\Delta}$, generated from the full model. Under H_o, with Δ known, we prove that Q given in Equation (4.2) follows asymptotically a $F(q, N-s)$ distribution.

APPENDIX

Computational algorithm

We present an iterative algorithm for computing the M-estimators of the parameters of the model given in Equation (2.1), which is similar to the Newton-Raphson method. Thus, in order to solve Equation (2.4), we propose the following iterative algorithm:

Step 1. Estimate β by using an initial value $\widehat{\beta}_n^{(0)}$. This value could be the weighted LS estimator of β.

Step 2. Estimate Δ by

$$\widehat{\Delta}^{(r)} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{h}\big(\varepsilon_i(\widehat{\beta}_n^{(r)})\big)\, \mathbf{h}^\top(\varepsilon_i(\widehat{\beta}_n^{(r)})).$$

Step 3. Reestimate β by using $\widehat{\Delta}^{(r)}$ obtained in Step 2 and solving

$$\widehat{\beta}_n^{(r+1)} = \widehat{\beta}_n^{(r)} + \Gamma_n^{-1}\big(\widehat{\beta}_n^{(r)}\big) \sum_{i=1}^{n} \lambda\big(\mathbf{x}_i, \mathbf{Y}_i, \widehat{\beta}^{(r)}, \widehat{\Delta}^{(r)}\big),$$

where

$$\begin{aligned}
\widehat{\Gamma}_n(\widehat{\beta}_n^{(r)}) &= \sum_{i=1}^{n}\Big\{\mathbf{X}_i^\top(\widehat{\beta}_n^{(r)})\, \big(\widehat{\Delta}^{(r)}\big)^{-1}\, \widehat{\mathbf{W}}^{(r)}\, \mathbf{X}_i(\widehat{\beta}_n^{(r)})\Big\},\\
\widehat{\mathbf{W}}^{(r)} &= \operatorname{diag}\Big\{\widehat{\gamma}_1^{(r)}, \ldots, \widehat{\gamma}_p^{(r)}\Big\},\\
\widehat{\gamma}_j^{(r)} &= \frac{1}{n}\sum_{i=1}^{n} \psi_j'\big(\varepsilon_{ij}(\widehat{\beta}_n^{(r)})\big), \text{ for } j = 1, \ldots, p, \text{ and}\\
\lambda\big(\mathbf{x}_i, \mathbf{Y}_i, \widehat{\beta}_n^{(r)}, \Delta\big) &= \mathbf{X}_i^\top(\widehat{\beta}_n^{(r)})\, \big(\widehat{\Delta}^{(r)}\big)^{-1}\, \Psi(\varepsilon_i(\widehat{\beta}_n^{(r)})).
\end{aligned}$$

Step 4. Consider as a new preliminary estimate $\widehat{\beta}_n^{(r)}$ and return to Step 2.

This scheme may be iterated a fixed number of times or until it converges, with at least one iteration recommended.

Theorem's proofs

Proof of Theorem 1. We can write $q\big(\mathbf{h}\big(\varepsilon_i(\beta)\big)\big) = \mathbf{h}^\top(\varepsilon_i)\Delta^{-1}\mathbf{h}(\varepsilon_i)$, obtaining

$$\frac{\partial q\big(\mathbf{h}\big(\varepsilon_i(\beta)\big)\big)}{\partial \beta} = -2\, \mathbf{X}_i^\top(\beta)\, \Delta^{-1}\, \Psi\big(\varepsilon_i(\beta)\big),$$

where $\mathbf{X}_i(\beta) = (\partial/\partial\beta^\top)\mathbf{f}(\mathbf{x}_i, \beta)$. ∎

Proof of Theorem 2(i). We can write

$$\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)=\mathbf{X}_i^{\top}(\beta)\,\Delta^{-1}\,\Psi\big(\varepsilon_i(\beta)\big)=\begin{pmatrix}\sum_{l=1}^{p}\sum_{m=1}^{p}(\partial/\partial\beta_1)f_m(\mathbf{x}_i,\beta)\,\delta^{ml}\,\psi_l(\varepsilon_{il}(\beta))\\ \vdots\\ \sum_{l=1}^{p}\sum_{m=1}^{p}(\partial/\partial\beta_s)f_m(\mathbf{x}_i,\beta)\,\delta^{ml}\,\psi_l(\varepsilon_{il}(\beta))\end{pmatrix},$$

where δ^{ml} is the (m,l)th element of Δ^{-1}. Then, we can consider the jth element of the vector $\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)$ denoted by

$$\lambda_j(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)=\sum_{l=1}^{p}\sum_{m=1}^{p}\Big\{(\partial/\partial\beta_j)f_m(\mathbf{x}_i,\beta)\,\delta^{ml}\ \psi_l(\varepsilon_{il}(\beta))\Big\},\ j=1,\ldots,s.$$

By using linear Taylor's expansion, we have

$$\begin{aligned}\lambda_j\left(\mathbf{x}_i,Y_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\quad=\quad&\frac{1}{\sqrt{n}}\sum_{k=1}^{s}t_k\left\{(\partial/\partial\beta_k)\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\right\}\\&+\frac{1}{\sqrt{n}}\sum_{k=1}^{s}t_k\Big\{(\partial/\partial\beta_k)\lambda_j\left(\mathbf{x}_i,Y_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)\\&\qquad-(\partial/\partial\beta_k)\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\Big\},\end{aligned}$$

where

$$\begin{aligned}(\partial/\partial\beta_k)\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\quad=\quad&\sum_{l=1}^{p}\sum_{m=1}^{p}\Big\{(\partial^2/\partial\beta_k\beta_j)f_m(\mathbf{x}_i,\beta)\,\delta^{ml}\,\psi_l(\varepsilon_{il}(\beta))\\&-(\partial/\partial\beta_j)f_m(\mathbf{x}_i,\beta)\,(\partial/\partial\beta_k)f_l(\mathbf{x}_i,\beta)\,\delta^{ml}\,\psi_l'(\varepsilon_{il}(\beta))\Big\}.\end{aligned}$$

Then, for $j=1,\ldots,s$, we have

$$\begin{aligned}&\sup_{\|\mathbf{t}\|\le C}\Bigg\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{\lambda_j\left(\mathbf{x}_i,Y_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\Big\}\\&\qquad+\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s}\sum_{l=1}^{p}\sum_{m=1}^{p}\left\{\gamma_l\,(\partial/\partial\beta_j)f_m(\mathbf{x}_i,\beta)\,\delta^{ml}\,(\partial/\partial\beta_k)f_l(\mathbf{x}_i,\beta)\,t_k\right\}\Bigg\|\\ \le\quad&\sup_{\|\mathbf{t}\|\le C}\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s}t_k\Big\{(\partial/\partial\beta_k)\lambda_j\left(\mathbf{x}_i,Y_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-(\partial/\partial\beta_k)\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\Big\}\right\|+\\&\sup_{\|\mathbf{t}\|\le C}\Bigg\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s}\left\{t_k(\partial/\partial\beta_k)\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\right\}\\&\qquad+\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s}\sum_{l=1}^{p}\sum_{m=1}^{p}\left\{\gamma_l\,(\partial/\partial\beta_j)f_m(\mathbf{x}_i,\beta)\,\delta^{ml}\,(\partial/\partial\beta_k)f_l(\mathbf{x}_i,\beta)\,t_k\right\}\Bigg\|,\end{aligned}$$

where

$$\sup_{\|\mathbf{t}\|\leq C}\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s}\{t_k(\partial/\partial\beta_k)\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\}\right.$$
$$\left.+\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s}\sum_{l=1}^{p}\sum_{m=1}^{p}\{\gamma_l\,(\partial/\partial\beta_j)\,f_m(\mathbf{x}_i,\beta)\,\delta^{ml}\,(\partial/\partial\beta_k)\,f_l(\mathbf{x}_i,\beta)\,t_k\}\right\| = o_p(1) \quad \text{and}$$

$$\sup_{\|\mathbf{t}\|\leq C}\left\|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{s}t_k\left\{(\partial/\partial\beta_k)\lambda_j\left(\mathbf{x}_i,Y_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-(\partial/\partial\beta_k)\lambda_j(\mathbf{x}_i,Y_i,\beta,\Delta)\right\}\right\| = o_p(1)$$

which proofs the theorem. ■

Proof of Theorem 2(ii). We can write

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta+\tfrac{1}{\sqrt{n}}\mathbf{K}\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)\right\}+\frac{1}{n}\Gamma_n\mathbf{t}$$
$$= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)\right.$$
$$\left.+\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta+\tfrac{1}{\sqrt{n}}\mathbf{K}\right)-\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)\right\}+\frac{1}{n}\Gamma_n\mathbf{t}.$$

Then

$$\sup_{\substack{\|\mathbf{t}\|\leq C\\ \|\mathrm{vec}(\mathbf{K})\|\leq D}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta+\tfrac{1}{\sqrt{n}}\mathbf{K}\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)\right\}+\frac{1}{n}\Gamma_n\mathbf{t}\right\|$$
$$\leq \sup_{\|\mathbf{t}\|\leq C}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)\right\}+\frac{1}{n}\Gamma_n\mathbf{t}\right\|+$$
$$\sup_{\|\mathrm{vec}(\mathbf{K})\|\leq D}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta^*,\Delta+\tfrac{1}{\sqrt{n}}\mathbf{K}\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta^*,\Delta)\right\}\right\|,$$

where $\beta^*=\beta+(1/\sqrt{n})\mathbf{t}$. Now, from Theorem 1, we have that

$$\sup_{\|\mathbf{t}\|\leq C}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta+\tfrac{1}{\sqrt{n}}\mathbf{t},\Delta\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta,\Delta)\right\}+\frac{1}{n}\Gamma_n\mathbf{t}\right\| = o_p(1).$$

Also, we prove that

$$\sup_{\|\mathrm{vec}(\mathbf{K})\|\leq D}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\lambda\left(\mathbf{x}_i,\mathbf{Y}_i,\beta^*,\Delta+\tfrac{1}{\sqrt{n}}\mathbf{K}\right)-\lambda(\mathbf{x}_i,\mathbf{Y}_i,\beta^*,\Delta)\right\}\right\| = o_p(1).$$

Consequently, we have proved the theorem. ■

Proof of Theorem 2(iii). From Theorem 1, we have that the system of equations

$$\sum_{i=1}^{n} \Lambda_j \left(\mathbf{x}_i, Y_i, \beta + \frac{1}{\sqrt{n}}\, \mathbf{t}, \Delta \right) = 0, \quad j = 1, \ldots, p,$$

has a root $\mathbf{t}_n$ that lies in $\|\mathbf{t}\| \leq C$ with a probability exceeding $1 - \eta$, for $n \geq n_0$. Then, $\widehat{\beta}_n = \beta + (1/\sqrt{n})\, \mathbf{t}_n$ is a solution of Equation (2.4) satisfying

$$P\left(\left\| \sqrt{n}\, (\widehat{\beta}_n - \beta) \right\| \leq C \right) \geq 1 - \eta, \quad n \geq n_0.$$

Now, considering $\mathbf{t} \to \sqrt{n}\, (\widehat{\beta}_n - \beta)$ in Equation (3.1), we have Equation (3.4). ∎

Proof of Theorem 2(iv). Let us consider the arbitrary linear compound

$$Z_n^* = \nu^\top \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \lambda(\mathbf{x}_i, \mathbf{Y}_i, \beta, \Delta), \quad \nu \in \Re^p,$$

so that

$$Z_n^* = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nu^\top\, \mathbf{X}_i^\top(\beta)\, \Delta^{-1}\, \Psi(\varepsilon_i(\beta)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i,$$

where $Z_i = \nu^\top\, \mathbf{X}_i^\top(\beta)\, \Delta^{-1}\, \Psi(\varepsilon_i(\beta))$ are independent random variables with

$$\begin{aligned} E\{Z_i\} &= 0, \text{ and} \\ Var\{Z_i\} &= \nu^\top\, \mathbf{X}_i^\top(\beta)\, \Delta^{-1}\, \Phi\, \Delta^{-1}\, \mathbf{X}_i(\beta)\, \nu \\ &= \sigma_i^2. \end{aligned}$$

We can write

$$Z_n^* = \sum_{i=1}^{n} c_{ni}\, Z_i^*,$$

where $c_{ni} = \sigma_i/\sqrt{n}$ and $Z_i^* = Z_i/\sigma$. Then, by using Hájek-Šidak's central limit theorem, we show that

$$\frac{\sum_{i=1}^{n} c_{ni}\, Z_i^*}{\sqrt{\sum_{i=1}^{n} c_{ni}^2}} \to \mathrm{N}(0,1)$$

resulting

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \to \mathrm{N}(0, \nu^\top\, \Lambda\, \nu).$$

∎

Acknowledgements

The authors wish to thank the editors and referees for their comments that aided in improving this article. This study was partially supported by FONDECYT 1080326 and 1090265 grants, Chile.

REFERENCES

1. Bickel, P. J. (1973). On some analogues to linear combinations of order statistics in the linear model. *Ann. Stat.*, **1**, 597-616.
2. Galant, A. (1987). *Non-linear Statistical Models.* Wiley, New York.
3. Huber, P.J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *Ann. Stat.*, **1**, 799-821.
4. Huber, P.J. (1981). *Robust Statistics.* Wiley, New York.
5. Hampel, F.R., Rousseeuw, P.J., Ronchetti, E., and Stabel, W. (1986). *Robust Statistics-the Approach Based on Influence Functions.* Wiley, New York.
6. Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of residuals. *Ann. Math. Stat.*, **42**, 366-371.
7. Jurecková, J. and Sen, P.K. (1996). *Robust Statistical Procedures, Asymptotics and Interrelations.* Wiley, New York.
8. Maronna, R.A. (1976). Robust M-estimators of multivariate location and scatter. *Ann. Stat.*, **4**, 163-169.
9. Sanhueza, A. and Sen, P.K. (2001). M-methods in Generalized Nonlinear Models. *Probability and Statistical Models with Applications, edited by Charalambides, Koutras, and Balakrishnan, Chapman-Hall / CRC, Chapter 23* pp. 359-375.
10. Sanhueza, A. and Sen, P.K. (2004). Robust M-procedures in univariate nonlinear regression Models. *Braz. J. Prob. Stat.*, **18**, 183-200.
11. Sanhueza, A., Sen, P.K., and Leiva, V. (2009). A robust procedure in nonlinear models for repeated measurements. *Comm. Stat. Theor. Meth.*, **38**, 138-155.
12. Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*. Wiley, New York.
13. Sen, P.K. and Singer, J.M. (1993). *Large Sample Methods in Statistics: An Introduction with Applications.* Chapman and Hall, New York.
14. Singer, J.M. and Sen. P.K. (1985). M-methods in multivariate linear models. *J. Mult. Anal.*, **17**, 168-184.