Brazilian Journal of Probability and Statistics (2004), 18, pp. 183–200.

Robust M-procedures in univariate nonlinear regression models

Antonio I. Sanhueza[1] and **Pranab K. Sen**[2]
[1] *Universidad de La Frontera*
[2] *University of North Carolina*

Abstract: In univariate nonlinear regression models, estimator and test statistics based on (generalized) least squares and maximum likelihood methods are usually nonrobust; M-procedures are better in this respect. Our proposed M-estimators, and M-tests are formulated along the lines of generalized least squares procedures and their (asymptotic) properties are studied. Computational algorithms are also considered along with.

Key words: Asymptotic normality, efficiency, M-estimators, M-tests, uniform asymptotic linearity.

1 Introduction

We consider the (univariate) nonlinear regression model

$$Y_i = f(\mathbf{x}_i,\, \boldsymbol{\beta}) + a_i\, e_i,\;\; i = 1,\, ...,\, n \tag{1.1}$$

where Y_i are the observable random variables (r.v.), $\mathbf{x}_i = (x_{1i},\, x_{2i},\, ...,\, x_{mi})^t$ are known regression constants, $\boldsymbol{\beta} = (\beta_1, \beta_2,\, ...,\, \beta_p)^t$ is a vector of unknown parameters; $f(\cdot,\cdot)$ is a (nonlinear) function (of $\boldsymbol{\beta}$) of specified form, the errors e_i are assumed to be independent and identically distributed (i.i.d) r.v.'s with a distribution function (d.f.) G, defined on $\Re$, and the a_i (> 0) are known constants, possibly dependent on the $\mathbf{x}_i$. We assume that the d.f. $G(\cdot)$ is continuous and symmetric about 0 (though its functional form may be unknown). Nonlinear regression models in the homoscedastic setup can be thought of direct generalizations of linear and generalized linear models (Gallant, 1987; Seber and Wild, 1989; McCullagh and Nelder, 1989; Genning et al., 1989, and Vonesh and Chinchilli, 1997). In a more general heteroscedastic setup, they are also related to transformation and weighing in regression (Carroll and Ruppert, 1982, 1988).

Unlike linear models, in nonlinear models, even when G is assumed to be normal, exact statistical inference could be a problem mainly due to the complexities of the estimating equations, and (generalised or weighted) least squares procedures are often used for asymptotic approximations (Hartley, 1961, and Hill and Holland, 1977). Such procedures are generally not robust, and in addition, may require iterative algorithms for their solution (Bates and Watts, 1987; Dutter, 1975; Hartley, 1961; Sen, 1998, and Marquardt, 1963). For linear models, robust procedures based on M-statistics have been extensively studied in the literature (Huber, 1981; Klein and Yohai, 1981; Yohai and Maronna, 1979; Hampel et al., 1986, and Jurečková and Sen, 1996). Incorporating the classical generalized least square (GLS) procedure, we formulate suitable M-estimators of $\boldsymbol{\beta}$ and study their (asymptotic) properties, including consistency and asymptotic normality. M-tests (Sen, 1982) are also considered for testing suitable hypothesis on $\boldsymbol{\beta}$. Two needed computational algorithms for M-estimators based on Newton-Rapson and Fisher's scoring methods are presented.

In Section 2, preliminary notions and regularity conditions are presented. Section 3 deals with the asymptotic distribution theory of M-estimators of $\boldsymbol{\beta}$, and in this respect, a uniform asymptotic linearity result on M-statistics (with respect to regression parameter) is presented in detail. Section 4 is devoted to related M-tests. In Section 5, two iterative computational methods are considered.

2 Notation and regularity assumptions

We introduce the notation $\left\| \mathbf{y}\right\|^2_{\mathbf{W}} = \mathbf{y}^t\,\mathbf{W}\,\mathbf{y}$ (of a quadratic norm) with $\mathbf{W}$ denoting a positive definite (p.d.) matrix. In the current context, we let $\mathbf{W} = Diag(w_{n1},\, ...,\, w_{nn})$, and $\mathbf{h}(\mathbf{Y}-f(\mathbf{x},\,\boldsymbol{\beta})) = \big(h(Y_1-f(\mathbf{x}_1,\,\boldsymbol{\beta}),\, ...,\, h(Y_n-f(\mathbf{x}_n,\,\boldsymbol{\beta})\big)^t$, where h is a suitable function which may downweight or omit extreme values. We consider then the norm

$$\sum_{i=1}^{n} w_{ni}\,[h(Y_i - f(\mathbf{x}_i,\,\boldsymbol{\beta}))]^2 = [\mathbf{h}(\mathbf{Y} - f(\mathbf{x},\,\boldsymbol{\beta}))]^t\,\mathbf{W}\,[\mathbf{h}(\mathbf{Y} - f(\mathbf{x},\,\boldsymbol{\beta}))]$$

and define $\hat{\boldsymbol{\beta}}_n$, an M-estimator of $\boldsymbol{\beta}$, as

$$\hat{\boldsymbol{\beta}}_n = \text{Arg min}\Big\{\Big\|\mathbf{h}(\mathbf{Y} - f(\mathbf{x},\,\boldsymbol{\beta}))\Big\|^2_{\mathbf{W}} : \boldsymbol{\beta} \in \boldsymbol{\Theta} \subseteq \Re^p\Big\}, \tag{2.1}$$

We could let $h(z) = z$ in (2.1) and have the WLS estimator of $\boldsymbol{\beta}$. However, we are interested in more robust methods (Huber, 1981; Hampel et al., 1986; Jurečková and Sen, 1996, and others), so that we may use bounded and monotone functions $h(\cdot)$; the so called Hampel score function is given by

$$h(z) = \begin{cases} \frac{1}{\sqrt{2}}\,z, & \text{if } |z| \le a \\ \{a(|z| - \frac{1}{2}a)\}^{1/2}, & \text{if } -b \le z \le -a,\ a \le z \le b \\ \{a\,\frac{c|z|-\frac{1}{2}z^2}{c-b} - \frac{7a^2}{6}\}^{1/2}, & \text{if } -c \le z \le -b,\ b \le z \le c \\ \{a(b+c-a)\}^{1/2}, & \text{if } |z| > c \end{cases}$$

for suitably chosen a, b, and c $(0 < a, b, c < \infty)$. We may define $\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i, \boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})f(\mathbf{x}_i,\,\boldsymbol{\beta})$, and $\psi(z) = (\partial/\partial z)\;h^2(z)$, so that the estimating equation for the minimization in (2.1) is given by

$$\sum_{i=1}^{n}\{w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\,\hat{\boldsymbol{\beta}}_n))\,\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\,\hat{\boldsymbol{\beta}}_n)\} = 0. \tag{2.2}$$

In this context, we let

$$w_{ni} = \Big[E\big(\psi^2(Y_i - f(\mathbf{x}_i,\, \boldsymbol{\beta}))\big)\Big]^{-1} \Big/ \sum_{j=1}^{n} \Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\, \boldsymbol{\beta}))\big)\Big]^{-1},\ 1 \le i \le n.$$

We note that the equations in (2.2) need not be linear in $\hat{\boldsymbol{\beta}}_n$ or Y_i. We shall impose the following sets of regularity assumptions concerning (A) the d.f. G, (B) the score function ψ, and (C) the function $f(\cdot)$.

[A1]: G is absolutely continuous with an absolutely continuous probability density function $g(\cdot)$ having a finite Fisher information

$$\int_{-\infty}^{+\infty} [g'(z)/g(z)]^2\, dG(z) \ < \infty.$$

[B1]: ψ is nonconstant, absolutely continuous and differentiable with respect to $\boldsymbol{\beta}$.

[B2]: i) $E\,\psi^2(z) < \infty$, and $E\,\psi(z) \ = 0$.

ii) $E\big(\psi'(z)\big)^2 < \infty$, and $E\psi'(z) = \gamma\ \ (\neq 0)$.

[B3]: i) $\lim_{\delta\to 0} E\Big\{\sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big|\,\psi(Y - f(\mathbf{x},\, \boldsymbol{\beta}+\boldsymbol{\Delta})) - \psi(Y - f(\mathbf{x},\, \boldsymbol{\beta}))\Big|\Big\} = 0.$

ii) $\lim_{\delta\to 0} E\Big\{\sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big|\,\psi'(Y - f(\mathbf{x},\, \boldsymbol{\beta}+\boldsymbol{\Delta})) - \psi'(Y - f(\mathbf{x},\, \boldsymbol{\beta}))\Big|\Big\} = 0.$

[C1]: $f(\mathbf{x},\, \boldsymbol{\beta})$ is continuous and twice differentiable with respect to $\boldsymbol{\beta} \in \boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is a compact subset of $\Re^p$.

[C2]: i) $max_{1\le i\le n}\Big\{ w_{ni}\, \dot{\mathbf{f}}^t_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, (\boldsymbol{\Gamma}_n(\boldsymbol{\beta}))^{-1}\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\Big\} \longrightarrow 0$, as $n \to \infty$, where

$$\boldsymbol{\Gamma}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} w_{ni}\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i, \boldsymbol{\beta})\, \dot{\mathbf{f}}^t_{\boldsymbol{\beta}}(\mathbf{x}_i, \boldsymbol{\beta}).$$

ii) $\lim_{n\to\infty} \boldsymbol{\Gamma}_n(\boldsymbol{\beta}) = \boldsymbol{\Gamma}(\boldsymbol{\beta})$, where $\boldsymbol{\Gamma}(\boldsymbol{\beta})$ is a positive definite matrix.

[C3]: i) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big|(\partial/\partial\beta_j\,) f(\mathbf{x},\, \boldsymbol{\beta}+\boldsymbol{\Delta})\ (\partial/\partial\beta_k) f(\mathbf{x},\, \boldsymbol{\beta}+\boldsymbol{\Delta})$
$-(\partial/\partial\beta_j\,) f(\mathbf{x},\, \boldsymbol{\beta})\ (\partial/\partial\beta_k) f(\mathbf{x},\, \boldsymbol{\beta})\Big| = 0;\ \ j,\, k = 1, ..., p.$

ii) $\lim_{\delta\to 0} \sup_{\|\boldsymbol{\Delta}\|\le\delta} \Big|(\partial^2/\partial\beta_j\, \partial\beta_k) f(\mathbf{x}, \boldsymbol{\beta}+\boldsymbol{\Delta}) - (\partial^2/\partial\beta_j\, \partial\beta_k) f(\mathbf{x}, \boldsymbol{\beta})\Big|$
$= 0;\ \ j, k = 1, ..., p.$

3 Asymptotics for M-estimators

In order to prove the asymptotic normality of the M-estimator $\hat{\boldsymbol{\beta}}_n$ in (2.2), first, in Theorem 3.1, we consider the uniform asymptotic linearity result. The boundedness in probability of $n^{\frac{1}{2}} \left\| \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \right\|$ is presented in Theorem 3.2. The asymptotic distribution of the M-estimator is finally developed in Theorem 3.3.

Theorem 3.1. *Under the conditions* **[A1]**, **[B1]-[B3]**, *and* **[C1]-[C3]**,

$$\sup_{\left\|\mathbf{t}\right\| \le C} \left\| n^{\frac{1}{2}} \sum_{i=1}^{n} \{ w_{ni} \left[\lambda(Y_i, \boldsymbol{\beta} + n^{-\frac{1}{2}} \mathbf{t}) - \lambda(Y_i, \boldsymbol{\beta}) \right] \} + \gamma\, \boldsymbol{\Gamma}_n(\boldsymbol{\beta})\, \mathbf{t} \right\| = o_p(1) \qquad (3.1)$$

as $n \to \infty$, *where* $\lambda(Y_i,\, \boldsymbol{\beta}) = \psi(Y_i - f(\mathbf{x}_i,\, \boldsymbol{\beta}))\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})$.

Proof. We start the proof by denoting the j-th element of the vector $\lambda(Y_i,\, \boldsymbol{\beta})$ by $\lambda_j(Y_i,\, \boldsymbol{\beta}) = \psi(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta})$, $j = 1, ..., p$, where $f_{\beta_j}(\mathbf{x}_i,\, \boldsymbol{\beta}) = (\partial/\partial\beta_j) f(\mathbf{x}_i,\, \boldsymbol{\beta})$. By the Taylor expansion we obtain

$$\lambda_j(Y_i,\, \boldsymbol{\beta} + n^{-\frac{1}{2}}\, \mathbf{t}) - \lambda_j(Y_i,\, \boldsymbol{\beta}) = \frac{1}{\sqrt{n}} \sum_{k=1}^{p} t_k\, (\partial/\partial\beta_k) \lambda_j(Y_i,\, \boldsymbol{\beta}) +$$

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{p} t_k \left\{ (\partial/\partial\beta_k) \lambda_j(Y_i,\, \boldsymbol{\beta} + \frac{h\, \mathbf{t}}{\sqrt{n}}) - (\partial/\partial\beta_k) \lambda_j(Y_i,\, \boldsymbol{\beta}) \right\},$$

where

$$\begin{aligned} (\partial/\partial\beta_k)\lambda_j(Y_i,\, \boldsymbol{\beta}) &= \psi(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\, (\partial^2/\partial\beta_k\, \partial\beta_j) f(\mathbf{x}_i,\, \boldsymbol{\beta}) \\ &- \psi'(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\, f_{\beta_j}(\mathbf{x}_i,\, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i,\, \boldsymbol{\beta}). \end{aligned}$$

It can be seen that for each $j (= 1, ..., p)$

$$\sup_{\left\|\mathbf{t}\right\| \le C} \Big| n^{\frac{1}{2}} \sum_{i=1}^{n} \{ w_{ni} \left[\lambda_j(Y_i,\, \boldsymbol{\beta} + \frac{h\, \mathbf{t}}{\sqrt{n}}) \, - \lambda_j(Y_i,\, \boldsymbol{\beta}) \right] \} +$$

$$\gamma \sum_{i=1}^{n} \sum_{k=1}^{p} t_k\, w_{ni}\, f_{\beta_j}(\mathbf{x}_i,\, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \Big|$$

$$\le \sup_{\left\|\mathbf{t}\right\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni} \left\{ (\partial/\partial\beta_k)\lambda_j(Y_i,\, \boldsymbol{\beta} + \frac{h\, \mathbf{t}}{\sqrt{n}}) - (\partial/\partial\beta_k)\lambda_j(Y_i,\, \boldsymbol{\beta}) \right\} \Big|$$

$$+ \sup_{\left\|\mathbf{t}\right\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni}\, (\partial/\partial\beta_k) \lambda_j(Y_i,\, \boldsymbol{\beta}) +$$

$$\gamma \sum_{i=1}^{n} \sum_{k=1}^{p} t_k\, w_{ni}\, f_{\beta_j}(\mathbf{x}_i,\, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i,\, \boldsymbol{\beta}) \Big|.$$

In addition, we may have that

$$\sup_{\|\mathbf{t}\|\le C}\Big|\sum_{i=1}^{n}\sum_{k=1}^{p}t_k w_{ni}\,\{(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta})\}\Big|$$
$$\le C\sum_{i=1}^{n}\sum_{k=1}^{p}w_{ni}\sup_{\|\mathbf{t}\|\le C}\Big|(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta})\Big|,$$

and

$$\sup_{\|\mathbf{t}\|\le C}\Big|(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta})\Big|$$
$$\le\sup_{\|\mathbf{t}\|\le C}\Big\{\Big|\psi'(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}}))-\psi'(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}))\Big|\cdot$$
$$\Big|f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})\,f_{\beta_k}(\mathbf{x}_i,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})\Big|\Big\}+$$
$$\sup_{\|\mathbf{t}\|\le C}\Big\{\Big|f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta})f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta})\Big|\cdot$$
$$\Big|\psi'(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}))\Big|\Big\}+$$
$$\sup_{\|\mathbf{t}\|\le C}\Big\{\Big|\psi(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}}))-\psi(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}))\Big|$$
$$\Big|(\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})\Big|\Big\}+$$
$$\sup_{\|\mathbf{t}\|\le C}\Big\{\Big|(\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-(\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta})\Big|\cdot$$
$$\Big|\psi(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta})\Big|\Big\}.$$

Then, it can be seen that

$$
\begin{aligned}
& E\Big\{\sup_{\|\mathbf{t}\|\le C}\Big|(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta})\Big|\Big\} \\
& \le \sup_{\|\boldsymbol{\Delta}\|\le\delta}\Big|f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta}+\boldsymbol{\Delta})f_{\beta_k}(\mathbf{x}_i,\boldsymbol{\beta}+\boldsymbol{\Delta})\Big| \\
& E\Big\{\sup_{\|\boldsymbol{\Delta}\|\le\delta}\Big|\psi'(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}+\boldsymbol{\Delta}))-\psi'(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}))\Big|\Big\}+ \\
& \sup_{\|\boldsymbol{\Delta}\|\le\delta}\Big|f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta}+\boldsymbol{\Delta})f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta}+\boldsymbol{\Delta})-f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta})f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta})\Big| \\
& E\Big\{\Big|\psi'(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}))\Big|\Big\}+ \\
& \sup_{\|\boldsymbol{\Delta}\|\le\delta}\Big|(\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta}+\boldsymbol{\Delta})\Big| \\
& E\Big\{\sup_{\|\boldsymbol{\Delta}\|\le\delta}\Big|\psi(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}+\boldsymbol{\Delta}))-\psi(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}))\Big|\Big\}+ \\
& \sup_{\|\boldsymbol{\Delta}\|\le\delta}\Big|\,(\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta}+\boldsymbol{\Delta})-\,(\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta})\Big|\cdot \\
& \quad E\Big\{\Big|\psi(Y_i-f(\mathbf{x}_i,\boldsymbol{\beta}))\Big|\Big\}.
\end{aligned}
$$

Thus, applying conditions **[B3]** i)-ii) and **[C3]** i)-ii) yields

$$
E\Big\{\sup_{\|\boldsymbol{\Delta}\|\le\delta}\Big|(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta})\Big|\Big\}\longrightarrow 0,\forall i\le n,
$$

and

$$
E\Big\{\sup_{\|\mathbf{t}\|\le C}\Big|\sum_{i=1}^{n}\sum_{k=1}^{p}t_k w_{ni}\,\{(\partial/\partial\beta_k)\lambda_j(Y_i,\,\boldsymbol{\beta}+\frac{h\,\mathbf{t}}{\sqrt{n}})-(\partial/\partial\beta_k)\lambda_j(Y_i,\boldsymbol{\beta})\}\Big|\Big\}
$$

$$
\longrightarrow 0.
$$

Also,

$$
\begin{aligned}
&Var\Big\{ \sup_{\|\mathbf{t}\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni} \, \{ (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta} + \frac{h\,\mathbf{t}}{\sqrt{n}}) - \\
&\quad (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta}) \} \Big| \Big\} \\
&\le C^2 \sum_{i=1}^{n} \Big(w_{ni}^2 \, Var\{ \sum_{k=1}^{p} \sup_{\|\mathbf{t}\| \le C} \Big| (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta} + \frac{h\,\mathbf{t}}{\sqrt{n}}) - \\
&\quad (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta}) \Big| \} \Big) \\
&= \frac{C^2}{\Big(\sum_{j=1}^{n} \Big[E\big(\psi^2(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\big) \Big]^{-1} \Big)^2} \; \sum_{i=1}^{n} \Big(\Big[E\big(\psi^2(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\big) \Big]^{-2} \\
&\quad Var\{ \sum_{k=1}^{p} \sup_{\|\mathbf{t}\| \le C} \Big| (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta} + \frac{h\,\mathbf{t}}{\sqrt{n}}) - (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta}) \Big| \} \Big) \\
&\le \frac{C^2\, K}{\Big(\sum_{j=1}^{n} \Big[E\big(\psi^2(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\big) \Big]^{-1} \Big)^2} \; \sum_{i=1}^{n} \Big[E\big(\psi^2(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\big) \Big]^{-2} \; \longrightarrow 0.
\end{aligned}
$$

Hence

$$
\sup_{\|\mathbf{t}\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni} \, \{ (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta} + \frac{h\,\mathbf{t}}{\sqrt{n}}) - (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta}) \} \Big| = o_p(1). \tag{3.2}
$$

On the other hand,

$$
\begin{aligned}
&\sup_{\|\mathbf{t}\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni} \, (\partial/\partial\beta_k)\lambda_j(Y_i, \boldsymbol{\beta}) + \\
&\gamma \sum_{i=1}^{n} \sum_{k=1}^{p} t_k \, w_{ni} \, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \Big| \\
&\le \sup_{\|\mathbf{t}\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni} \, \psi(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta})) \; (\partial^2/\partial\beta_k \, \partial\beta_j) f(\mathbf{x}_i, \boldsymbol{\beta}) \Big| + \\
&\sup_{\|\mathbf{t}\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni} \, \psi'(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta})) \, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) - \\
&\gamma \sum_{i=1}^{n} \sum_{k=1}^{p} t_k \, w_{ni} \, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \Big|,
\end{aligned}
$$

where

$$\begin{aligned}
&\sup_{\|\mathbf{t}\|\le C}\Big|\sum_{i=1}^{n}\sum_{k=1}^{p} t_k w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\ (\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta})\Big| \\
&\le \sum_{k=1}^{p}\sup_{\|\mathbf{t}\|\le C}\Big|t_k\sum_{i=1}^{n} w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\ (\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta})\Big| \\
&\le C\ \sum_{k=1}^{p}\Big|\ \sum_{i=1}^{n} w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\ (\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta})\Big|.
\end{aligned}$$

Now, we may define

$$U_n = \sum_{i=1}^{n} w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\ (\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\boldsymbol{\beta}),$$

and have

$$E(U_n) = 0,$$

and

$$Var(U_n) = \frac{1}{\Big(\sum_{j=1}^{n}\Big[E\psi^2(Y_j - f(\mathbf{x}_j,\boldsymbol{\beta}))\Big]^{-1}\Big)^2}\ \sum_{i=1}^{n}\Big(\Big[E\psi^2(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\Big]^{-1}$$

$$\Big[(\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\boldsymbol{\beta})\Big]^2\Big) \longrightarrow 0.$$

Hence, we have that $U_n = o_p(1)$. Then

$$\sup_{\|\mathbf{t}\|\le C}\Big|\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{p}\{t_k w_i\,\psi(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\ (\partial^2/\partial\beta_k\,\partial\beta_j)f(\mathbf{x}_i,\,\boldsymbol{\beta})\}\Big| = o_p(1). \quad (3.3)$$

Similarly,

$$\begin{aligned}
&\sup_{\|\mathbf{t}\|\le C}\Big|\sum_{i=1}^{n}\sum_{k=1}^{p}\{t_k w_{ni}\,\psi'(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\,f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta})f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta})\} - \\
&\gamma\sum_{i=1}^{n}\sum_{k=1}^{p}\{t_k\, w_{ni}\, f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta})\, f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta})\}\Big| \\
&\le \sum_{k=1}^{p}\sup_{\|\mathbf{t}\|\le C}\Big|t_k\sum_{i=1}^{n}\{w_{ni}\Big(\psi'(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\ -\gamma\Big)\, f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta})\, f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta})\}\Big| \\
&\le C\ \sum_{k=1}^{p}\Big|\ \sum_{i=1}^{n}\{w_{ni}\Big(\psi'(Y_i - f(\mathbf{x}_i,\boldsymbol{\beta}))\ -\gamma\Big)\, f_{\beta_j}(\mathbf{x}_i,\,\boldsymbol{\beta})\, f_{\beta_k}(\mathbf{x}_i,\,\boldsymbol{\beta})\}\Big|,
\end{aligned}$$

where

$$V_n = \sum_{i=1}^{n} \Big\{ w_{ni} \left(\psi'(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta})) - \gamma \right) f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta})\, f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \Big\} = o_p(1),$$

because

$$E(V_n) = 0,$$

and

$$Var(V_n) = \frac{1}{\left(\sum_{j=1}^{n} \Big[E\psi^2(Y_j - f(\mathbf{x}_j, \boldsymbol{\beta}))\Big]^{-1}\right)^2} \sum_{i=1}^{n} \Big(\Big[E\psi^2(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\Big]^{-2}$$

$$Var(\psi'(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))) \Big[f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta})\, f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta})\Big]^2 \Big) \longrightarrow 0.$$

Then

$$\sup_{\|\mathbf{t}\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} \{ t_k\, w_{ni}\, \psi'(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta})\, f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \} -$$

$$\gamma \sum_{i=1}^{n} \sum_{k=1}^{p} \{ t_k\, w_{ni}\, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \} \Big| = o_p(1). \tag{3.4}$$

So, using (3.3) and (3.4) we have

$$\sup_{\|\mathbf{t}\| \le C} \Big| \sum_{i=1}^{n} \sum_{k=1}^{p} t_k w_{ni}\, (\partial/\partial\beta_k) \lambda_j(Y_i, \boldsymbol{\beta}) +$$

$$\gamma \sum_{i=1}^{n} \sum_{k=1}^{p} t_k\, w_{ni}\, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \Big| = o_p(1). \tag{3.5}$$

Therefore, from (3.2) and (3.5) we may conclude that

$$\sup_{\|\mathbf{t}\| \le C} \Big| n^{\frac{1}{2}} \sum_{i=1}^{n} \{ w_i\, [\lambda_j(Y_i, \boldsymbol{\beta} + \frac{h\,\mathbf{t}}{\sqrt{n}}) - \lambda_j(Y_i, \boldsymbol{\beta})] \} +$$

$$\gamma \sum_{i=1}^{n} \sum_{k=1}^{p} t_k\, w_i\, f_{\beta_j}(\mathbf{x}_i, \boldsymbol{\beta}) f_{\beta_k}(\mathbf{x}_i, \boldsymbol{\beta}) \Big| = o_p(1),\ j = 1, ..., p. \quad \bullet$$

Theorem 3.2. *Under the conditions* **[A1]**, **[B1]-[B3]**, *and* **[C1]-[C3]**, *there exists a sequence* $\hat{\boldsymbol{\beta}}_n$ *of solutions of (2.2) such that*

$$n^{\frac{1}{2}} \left\| \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} \right\| = O_p(1) \ \text{as } n \to \infty \tag{3.6}$$

and

$$\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + \frac{1}{\gamma}\,(\,\boldsymbol{\Gamma}_n(\boldsymbol{\beta}))^{-1} \sum_{i=1}^{n} \{w_{ni}\,\psi(y_i - f(\mathbf{x}_i,\,\boldsymbol{\beta}))\,\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\,\boldsymbol{\beta})\} + o_p(n^{-\frac{1}{2}}). \quad (3.7)$$

Proof. From Theorem 3.1 and following the proof of Theorem 5.2.1 of Jurečková and Sen (1996), we have that the system of equations

$$\sum_{i=1}^{n}\{w_{ni}\psi(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta} + n^{-\frac{1}{2}}\,\mathbf{t}))\ (\partial/\partial\beta_j)\,f(\mathbf{x}_i,\,\boldsymbol{\beta} + n^{-\frac{1}{2}}\,\mathbf{t})\} = 0$$

has a root $\mathbf{t}_n$ that lies in $\|\mathbf{t}\| \leq C$ with probability exceeding $1-\epsilon$ for $n \geq n_0$. Then $\hat{\boldsymbol{\beta}}_n = \boldsymbol{\beta} + n^{-\frac{1}{2}}\,\mathbf{t}_n$ is a solution of the equations in (2.2) satisfying

$$P\big(\,\|n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})\| \leq C\,\big) \geq 1-\epsilon \text{ for } n \geq n_0,$$

i.e., the expression (3.6) is proved. Now, inserting $\mathbf{t} \longrightarrow n^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ in (3.1), we have the expression in (3.7). •

Theorem 3.3. *Under the conditions* **[A1]***,* **[B1]-[B3]***, and* **[C1]-[C3]***,*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \longrightarrow \mathbf{N}_p\Big(\,\mathbf{0},\ \frac{\gamma^{-2}}{\frac{1}{n}\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\,\boldsymbol{\beta})\big)\Big]^{-1}}\ (\boldsymbol{\Gamma}(\boldsymbol{\beta}))^{-1}\Big). \quad (3.8)$$

Proof. From (3.7) we have that

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \frac{1}{\gamma}\,(\boldsymbol{\Gamma}_n(\boldsymbol{\beta}))^{-1} \sum_{i=1}^{n}\{w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\,\boldsymbol{\beta}))\,\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\,\boldsymbol{\beta})\} + o_p(n^{-\frac{1}{2}}).$$

Considering the r.v. $\mathbf{Z}_n = \sum_{i=1}^{n}\{w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\,\boldsymbol{\beta}))\,\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\,\boldsymbol{\beta})\}$, in order to prove that Z_n has asymptotically a multinormal distribution, we consider an arbitrary linear compound $Z_n^* = \lambda^t\,\mathbf{Z}_n$, $\lambda \in \Re^p$, so that

$$\begin{aligned} Z_n^* &= \lambda^t \sum_{i=1}^{n}\big\{w_{ni}\,\psi(Y_i - f(\mathbf{x}_i,\,\boldsymbol{\beta}))\,\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\,\boldsymbol{\beta})\big\} \\ &= \sum_{i=1}^{n}\Big\{\frac{1}{\sqrt{\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\,\boldsymbol{\beta})\big)\Big]^{-1}}}\,\sqrt{w_{ni}}\ \ \lambda^t\,\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\,\boldsymbol{\beta})\,Z_i(\boldsymbol{\beta})\Big\}, \end{aligned}$$

where

$$Z_i(\boldsymbol{\beta}) = \frac{1}{\sqrt{w_{ni}^{-1}\,/\,\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\,\boldsymbol{\beta})\big)\Big]^{-1}}}\ \psi(Y_i - f(\mathbf{x}_i,\,\boldsymbol{\beta})),$$

with

$$E\{Z_i(\boldsymbol{\beta})\} = 0, \text{ and } Var\{Z_i(\boldsymbol{\beta})\} = 1.$$

Thus, we may write

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

where

$$c_{ni} = \frac{1}{\sqrt{\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\, \boldsymbol{\beta})\big)\Big]^{-1}}}\, \sqrt{w_{ni}}\; \lambda^t\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta}).$$

Now, we may use the Hájek-Šidak Central Limit Theorem (Sen and Singer, 1993, p.119) to show that Z_n^* is asymptotically normally distributed; for this, we need to verify the Noether condition on the c_{ni}, namely, that $\max_{1\le i\le n} c_{ni}^2 \,/\, \sum_{i=1}^{n} c_{ni}^2 \longrightarrow 0$, as $n \to \infty$, which is equivalent to showing that as $n \to \infty$,

$$\sup_{\lambda\in\Re^p}\Big[\max_{1\le i\le n}\, \lambda^t\, w_i\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, \dot{\mathbf{f}}^t_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, \lambda \,/\, \lambda^t \boldsymbol{\Gamma}_n(\boldsymbol{\beta})\, \lambda\Big] \longrightarrow 0.$$

By using the Courant Theorem (Sen and Singer, 1993, p.28), we have that

$$\begin{aligned}
&\sup_{\lambda\in\Re^p}\Big[\lambda^t\, w_{ni}\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, \dot{\mathbf{f}}^t_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, \lambda \,/\, \lambda^t \boldsymbol{\Gamma}_n(\boldsymbol{\beta})\, \lambda\Big] \\
&= ch_1\Big\{w_{ni}\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, \dot{\mathbf{f}}^t_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, (\boldsymbol{\Gamma}_n(\boldsymbol{\beta}))^{-1}\Big\} \\
&= w_{ni}\, \dot{\mathbf{f}}^t_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\, (\boldsymbol{\Gamma}_n(\boldsymbol{\beta}))^{-1}\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta}).
\end{aligned}$$

Thus, the condition above is reduced to the Noether condition **[C2]** i). Hence, we conclude that

$$Z_n^* / [\sum_{i=1}^{n} c_{ni}^2]^{\frac{1}{2}} \longrightarrow N(0,\, 1), \text{ as } n \to \infty.$$

In addition, by using the Cramér-Wold Theorem (Sen and Singer, 1993, p.106) and condition **[C2]** ii) we may prove that

$$\begin{aligned}
&Z_n = \sum_{i=1}^{n}\{w_{ni}\, \psi(Y_i - f(\mathbf{x}_i,\, \boldsymbol{\beta}))\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\} \\
&\longrightarrow \mathbf{N}_p\Big(\mathbf{0},\, \frac{1}{\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\, \boldsymbol{\beta})\big)\Big]^{-1}}\, \boldsymbol{\Gamma}(\boldsymbol{\beta})\Big), \text{ as } n \to \infty.
\end{aligned}$$

Finally, by using the Slutsky Theorem (Sen and Singer, 1993, p.127), the asymptotic distribution of Z_n, and condition **[C2]** ii), we may conclude the result in (3.8). •

Corollary 3.4. *Under the conditions* **[A1]**, **[B1]-[B3]**, *and* **[C1]-[C3]**,

$$\Big[\frac{\hat{\gamma}^{-2}}{\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\, \hat{\boldsymbol{\beta}}_n))\Big]^{-1}}\ (\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\beta}}_n))^{-1}\Big)\Big]^{-\frac{1}{2}}\ (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \longrightarrow \mathbf{N}_p\Big(\mathbf{0},\ I_p\Big).$$

Corollary 3.5. *Under the conditions* **[A1]**, **[B1]-[B3]**, *and* **[C1]-[C3]**,

$$(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^t\ \Big[\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\, \hat{\boldsymbol{\beta}}_n))\Big]^{-1}\hat{\gamma}^2\,\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\beta}}_n)\Big]\ (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$$

$$\longrightarrow \chi^2_p.$$

4 M-tests for linear and nonlinear hypotheses

The asymptotic distribution derived in Section 3.3 allows us to construct test of hypotheses for the parameters in model in (1.1). We consider both linear and nonlinear hypothesis for $\boldsymbol{\beta}$.

4.1 Scalar hypothesis testing

From Corollary 3.4 we can write:

$$(\hat{\boldsymbol{\Sigma}})^{-\frac{1}{2}}\,(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \longrightarrow \mathbf{N}_p\Big(\mathbf{0},\ \mathbf{I}_p\Big), \tag{4.1}$$

where

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\, \hat{\boldsymbol{\beta}}_n))\Big]^{-1}}\ \hat{\gamma}^{-2}\,(\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\beta}}_n))^{-1},$$

and

$$\hat{\gamma} = \frac{1}{n}\sum_{i=1}^{n}\psi'(y_i - f(\mathbf{x}_i,\, \hat{\boldsymbol{\beta}}_n)).$$

We let β_k be the kth element of $\boldsymbol{\beta}$ ($k = 1, ..., p$), and $s.e.(\hat{\beta}_k)$ be the square root of the kth diagonal element of the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}$ in (4.1). Then, a $100(1-\alpha)\%$ confidence interval for β_k is given by:

$$\hat{\beta}_k \pm \tau_{\frac{\alpha}{2}}\ s.e.(\hat{\beta}_k),$$

where τ_α is the upper 100α-quantile of the standard normal distribution.

We also may construct test statistics for hypothesis testing using the normal distribution in (4.1).

4.2 Wald-type M-tests for linear sub-hypotheses

We are interested in a confidence region for, or test statistic about a subset of parameters; we partition $\boldsymbol{\beta}$ as $(\boldsymbol{\beta}_1^{'}, \boldsymbol{\beta}_2^{'})'$ consisting of the first r and the last $p-r$ columns, ordering the elements of $\boldsymbol{\beta}$ such that those of interest are the first r elements. We let

$$\hat{\boldsymbol{\beta}}_n = \begin{pmatrix} \hat{\boldsymbol{\beta}}_{1n} \\ \hat{\boldsymbol{\beta}}_{2n} \end{pmatrix}$$

be the M-estimator for $\boldsymbol{\beta}$. Then, from Corollary 3.4 we have that

$$[\sum_{j=1}^{n} \Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j, \hat{\boldsymbol{\beta}}_n))\big)\Big]^{-1} \hat{\gamma}^2 \,\mathbf{V}_{11}\big]^{\frac{1}{2}} \; (\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1) \longrightarrow N(\mathbf{0}, I_r)$$

where

$$(\boldsymbol{\Gamma}_n(\hat{\boldsymbol{\beta}}_n))^{-1} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}.$$

Then, by the Cochran theorem (Sen and Singer, 1993, p.137), we have

$$(\hat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_1)^t \Big(\sum_{j=1}^{n} \Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j, \hat{\boldsymbol{\beta}}_{n1}))\big)\Big]^{-1} \hat{\gamma}^2 \;\mathbf{V}_{11}\Big) (\hat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_1) \longrightarrow \chi^2_r,$$

where $\mathbf{V}_{11}$ is of rank r. Therefore, for testing the hypothesis $H_o : \boldsymbol{\beta}_1 = 0$, we use the Wald-type M-test given by

$$W = (\hat{\boldsymbol{\beta}}_{n1})^t \Big(\sum_{j=1}^{n} \Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j, \hat{\boldsymbol{\beta}}_{n1}))\big)\Big]^{-1} \hat{\gamma}^2 \;\mathbf{V}_{11}\Big) \hat{\boldsymbol{\beta}}_{n1} \qquad (4.2)$$

which, under H_o, has asymptotically the χ^2_r distribution. We can use the same result in (4.3) to construct confidence regions in the usual way.

4.3 The likelihood ratio-type M-test

We consider a partition of $\boldsymbol{\beta}$ as in (4.2) and the hypothesis

$$H_o : \boldsymbol{\beta}_1 = 0 \text{ vs } H_1 : \boldsymbol{\beta}_1 \neq 0.$$

We define $L_n(\boldsymbol{\beta})$ as

$$L_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{w_{ni}\, h^2(y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\}.$$

The "unrestricted" M-estimator for $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_n = \text{Arg} \;\; \min\{L_n(\boldsymbol{\beta}) : \boldsymbol{\beta} \in \boldsymbol{\Theta}\}.$$

Also, the "restricted" M-estimator for $\boldsymbol{\beta}$ is given by

$$\tilde{\boldsymbol{\beta}}_n = \begin{pmatrix} 0 \\ \tilde{\boldsymbol{\beta}}_2 \end{pmatrix} = \text{Arg} \;\; \min\{L_n(\boldsymbol{\beta}) : \boldsymbol{\beta}_1 = 0,\, \boldsymbol{\beta}_2 \in \boldsymbol{\Theta}_2\},$$

where $\boldsymbol{\Theta}_2 = \left\{\boldsymbol{\beta}_2 : \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \in \boldsymbol{\Theta},\, \boldsymbol{\beta}_1 = 0\right\}$.

We define

$$\begin{aligned} Q &= 2\{L_n(\hat{\boldsymbol{\beta}}_n) - L_n(\tilde{\boldsymbol{\beta}})\} \\ &= 2\{L_n(\hat{\boldsymbol{\beta}}) - L_n(\boldsymbol{\beta}) - L_n(\tilde{\boldsymbol{\beta}}_n) + L_n(\boldsymbol{\beta})\}, \end{aligned}$$

and using a Taylor approximation and Theorem 3.2 we may write

$$\begin{aligned} Q_1 &= 2\left\{L_n(\hat{\boldsymbol{\beta}}_n) - L_n(\boldsymbol{\beta})\right\} \\ &= (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})^t \left(\gamma \boldsymbol{\Gamma}(\boldsymbol{\beta})\right) (\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \\ &= \mathbf{M}_n^t(\boldsymbol{\beta})\, \frac{1}{\gamma}\, (\boldsymbol{\Gamma}(\boldsymbol{\beta}))^{-1}\, \mathbf{M}_n(\boldsymbol{\beta}) \,+ o_p(n^{-\frac{1}{2}}), \end{aligned}$$

where

$$\mathbf{M}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{w_{ni}\, \psi(Y_i - f(\mathbf{x}_i,\, \boldsymbol{\beta}))\, \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_i,\, \boldsymbol{\beta})\}.$$

Similarly, we may have

$$\begin{aligned} Q_2 &= 2\left\{L_n(\tilde{\boldsymbol{\beta}}_n) - L_n(\boldsymbol{\beta})\right\} \\ &= \mathbf{M}_n^t(\boldsymbol{\beta})\, \frac{1}{\gamma} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Gamma}_{22}(\boldsymbol{\beta}_2))^{-1} \end{pmatrix} \mathbf{M}_n(\boldsymbol{\beta}) \,+ o_p(n^{-\frac{1}{2}}), \end{aligned}$$

where,

$$\boldsymbol{\Gamma}_{22}(\boldsymbol{\beta}_2) = \lim_{n\to\infty} \sum_{i=1}^{n} w_{ni}\, \dot{\mathbf{f}}_{\boldsymbol{\beta}_2}(\mathbf{x}_i, \boldsymbol{\beta}_2)\, \dot{\mathbf{f}}^t_{\boldsymbol{\beta}_2}(\mathbf{x}_i, \boldsymbol{\beta}_2).$$

Then,

$$Q = \mathbf{M}_n^t(\boldsymbol{\beta})\, \frac{1}{\gamma} \left[(\boldsymbol{\Gamma}(\boldsymbol{\beta}))^{-1} - \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{\Gamma}_{22}(\boldsymbol{\beta}_2))^{-1} \end{pmatrix}\right] \mathbf{M}_n(\boldsymbol{\beta}) + o_p(n^{-\frac{1}{2}}).$$

Note that the asymptotic normality result in (4.1) involves the estimator $\hat{\boldsymbol{\Sigma}}$ which in turn involves the estimators $\hat{\gamma}$ and $\hat{\boldsymbol{\beta}}$, both being consistent for their respective population counterparts. As such, by using expression (4.1), the Cochran

Theorem (on the distribution of quadratic forms in normal vectors), and the Slutsky Theorem, it can be readily verified that

$$\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\,\boldsymbol{\beta}))\Big]^{-1}\gamma\, Q = \tag{4.3}$$

$$\mathbf{M}_n^t(\boldsymbol{\beta})\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\boldsymbol{\beta}))\Big]^{-1}\Big\{(\boldsymbol{\Gamma}(\boldsymbol{\beta}))^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & (\boldsymbol{\Gamma}_{22}(\boldsymbol{\beta}_2))^{-1}\end{pmatrix}\Big\}\times$$

$$\mathbf{M}_n(\boldsymbol{\beta}) + o_p(n^{-\frac{1}{2}}) \longrightarrow \chi^2_r,\ \text{as } n\to\infty.$$

In order to prove this we note that $\mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A} = \mathbf{A}$, and $tr(\boldsymbol{\Sigma}\,\mathbf{A}) = r$, where

$$\mathbf{A} = \sum_{j=1}^{n}\Big[E\big(\psi^2(Y_i - f(\mathbf{x}_i,\,\boldsymbol{\beta}))\Big]^{-1}\Big\{(\boldsymbol{\Gamma}(\boldsymbol{\beta}))^{-1} - \begin{pmatrix} 0 & 0 \\ 0 & (\boldsymbol{\Gamma}_{22}(\boldsymbol{\beta}))^{-1}\end{pmatrix}\Big\}$$

and

$$\boldsymbol{\Sigma} = \frac{1}{\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\,\boldsymbol{\beta}))\Big]^{-1}}\boldsymbol{\Gamma}(\boldsymbol{\beta}).$$

Finally, we consider the Likelihood Ratio type M-test given by:

$$F_M = 2\,\hat{\gamma}_n\sum_{j=1}^{n}\Big[E\big(\psi^2(Y_j - f(\mathbf{x}_j,\,\hat{\boldsymbol{\beta}}_n))\Big]^{-1}[L_n(\hat{\boldsymbol{\beta}}_n) - L_n(\tilde{\boldsymbol{\beta}}_n)]. \tag{4.4}$$

From expression (4.3) and the Slutsky Theorem, we may conclude that the asymptotic distribution of F_M in (4.4) is a χ^2_r distribution.

4.4 Testing nonlinear hypotheses

We consider nonlinear hypotheses of the form

$$H_o : a(\boldsymbol{\beta}) = 0 \ \text{ vs }\ H_1 : a(\boldsymbol{\beta}) \neq 0$$

where a is a real valued (nonlinear) function (of $\boldsymbol{\beta}$).

From expression (4.1) and the Delta Method, we have that

$$\{\dot{\mathbf{a}}^t_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n)\ \hat{\boldsymbol{\Sigma}}\ \dot{\mathbf{a}}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n)\}^{-\frac{1}{2}}\, n^{\frac{1}{2}}\,\big(a(\hat{\boldsymbol{\beta}}_n) - a(\boldsymbol{\beta})\big) \longrightarrow N\big(0,\,1\big),$$

where $\dot{\mathbf{a}}_{\boldsymbol{\beta}}(\boldsymbol{\beta}) = (\partial/\partial\boldsymbol{\beta})a(\boldsymbol{\beta})$.

In adition, by using the Cochran Theorem we have:

$$\big(a(\hat{\boldsymbol{\beta}}_n) - a(\boldsymbol{\beta})\big)^t\,\{\dot{\mathbf{a}}^t_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n)\ \hat{\boldsymbol{\Sigma}}\ \dot{\mathbf{a}}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n)\}^{-1}\,\big(a(\hat{\boldsymbol{\beta}}_n) - a(\boldsymbol{\beta})\big) \longrightarrow \chi^2_1.$$

Hence in order to test the null hypothesis, we may use the Wald-type M-test difined by

$$W = \big(a(\hat{\boldsymbol{\beta}}_n)\big)^t\,\{\dot{\mathbf{a}}^t_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n)\ \hat{\boldsymbol{\Sigma}}\ \dot{\mathbf{a}}_{\boldsymbol{\beta}}(\hat{\boldsymbol{\beta}}_n)\}^{-1}\,\big(a(\hat{\boldsymbol{\beta}}_n)\big),$$

which under H_o follows a χ^2_1 distribution.

5 Computational algorithm

In general, we must rely on an iterative method for solving the equations in (2.1). We propose two iterative methods based upon a Taylor expansion around some initial guess $\hat{\boldsymbol{\beta}}_n^{(0)}$. First, we define the following matrices:

$$\mathbf{A}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \big\{ w_{ni}\ \psi(Y_i - f(\mathbf{x}_i, \boldsymbol{\beta}))\ \partial^2 f(\mathbf{x}_i, \boldsymbol{\beta}) / \partial \boldsymbol{\beta}^t\, \partial \boldsymbol{\beta} \big\},$$

$$\mathbf{W} = Diag(w_{n1},\ ...,\ w_{nn}),$$

$$\mathbf{W}_1(\boldsymbol{\beta}) = Diag\Big(w_{n1}\psi'(Y_1 - f(\mathbf{x}_1, \boldsymbol{\beta})), ..., w_{nn}\psi'(Y_1 - f(\mathbf{x}_1, \boldsymbol{\beta})) \Big),$$

$$\mathbf{X}(\boldsymbol{\beta}) = \Big(\dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_1,\ \boldsymbol{\beta}),\ \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_2,\ \boldsymbol{\beta}),\ ...,\ \dot{\mathbf{f}}_{\boldsymbol{\beta}}(\mathbf{x}_n,\ \boldsymbol{\beta}) \Big)^t,$$

$$\boldsymbol{\Psi}(\mathbf{Y} - f(\mathbf{x}, \boldsymbol{\beta})) = \Big(\psi(Y_1 - f(\mathbf{x}_1, \boldsymbol{\beta})),\ ...,\ \psi(Y_n - f(\mathbf{x}_n, \boldsymbol{\beta})) \Big)^t, \text{ and}$$

$$\mathbf{U}(\boldsymbol{\beta}) = \mathbf{A}(\hat{\boldsymbol{\beta}}_n^{(l)}) - \mathbf{X}^t(\hat{\boldsymbol{\beta}}_n^{(l)})\, \mathbf{W}_1(\hat{\boldsymbol{\beta}}_n^{(l)})\, \mathbf{X}^t(\hat{\boldsymbol{\beta}}_n^{(l)}).$$

If we choose $\hat{\boldsymbol{\beta}}_n^{(0)}$ based on some consistent estimator (e.g. the GLS estimator) we may use the following algorithm

$$\begin{cases} \hat{\boldsymbol{\beta}}_n^{(0)} = \hat{\boldsymbol{\beta}}_{GLS} \\ \hat{\boldsymbol{\beta}}_n^{(l+1)} = \hat{\boldsymbol{\beta}}_n^{(l)} - \big\{ \mathbf{U}(\hat{\boldsymbol{\beta}}_n^{(l)}) \big\}^{-1} \mathbf{X}^t(\hat{\boldsymbol{\beta}}_n^{(l)})\, \mathbf{W}(\hat{\boldsymbol{\beta}}_n^{(l)})\, \boldsymbol{\Psi}(\mathbf{Y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}}_n^{(l)})), \end{cases}$$

which is similar to the Newton-Raphson method. We can also replace $\mathbf{U}(\boldsymbol{\beta})$ by its expected value:

$$E\big\{ \mathbf{U}(\boldsymbol{\beta}) \big\} = -\gamma\, \boldsymbol{\Gamma}(\boldsymbol{\beta}),$$

and propose the following algorithm:

$$\begin{cases} \hat{\boldsymbol{\beta}}_n^{(0)} = \hat{\boldsymbol{\beta}}_{GLS} \\ \hat{\boldsymbol{\beta}}_n^{(l+1)} = \hat{\boldsymbol{\beta}}_n^{(l)} + (\hat{\gamma}_n^{(l)})^{-1} \big\{ \boldsymbol{\Gamma}_n(\hat{\boldsymbol{\beta}}_n^{(l)} \big\}^{-1} \mathbf{X}^t(\hat{\boldsymbol{\beta}}_n^{(l)})\, \mathbf{W}(\hat{\boldsymbol{\beta}}_n^{(l)})\, \boldsymbol{\Psi}(\mathbf{Y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}}_n^{(l)})), \end{cases}$$

where

$$\hat{\gamma}_n^{(l)} = \frac{1}{n} \sum_{i=1}^{n} \psi'(Y_i - f(\mathbf{x}_i, \hat{\boldsymbol{\beta}}_n^{(l)})).$$

Acknowledgements

Thanks are due to the reviewers for their helpful comments on the manuscript. The first author was partially supported by Projecto PPF 110 from the Dirección de Investigación y Desarrollo de la Universidad de La Frontera.

(Received December, 1999. Accepted September, 2004.)

References

Bates, D. M. and Watts, D. G. (1987). A generalized Gauss-Newton procedure for multiresponse parameter estimation. *SIAM J. Sci. and Stat. Comput.,* **8**, 49-55.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Annals of Statistics,* **10**, 429-441.

Carroll, R. J. and Ruppert, D. (1988). Transformations and Weighting in Regression. New York: Chapman and Hall.

Dutter, R. (1975). Robust regression: different approaches to numerical solutions and algorithms. Research report 6, Fachgruppe fuer statistik, ETH, Zurich.

Gallant, A. R. (1987). Nonlinear Statistical Models. New York: Wiley.

Genning, C., Chinchilli, V. M. and Carter, W. H. (1989). Response surface analysis with correlated data: A nonlinear model approach. *Journal of the American Statistical Association,* **84**, 805-809.

Hampel, F., Rousseeuw, P., Ronchetti, E. M. and Stahel, W. A. (1986). Robust Statistics. New York: Wiley.

Hartley, H. O. (1961). The modified Gauss-Newton method for the fitting of nonlinear regression functions by least squares. *Technometrics,* **3**, 269-280.

Hill, R. W. and Holland, P. W. (1977). Two robust alternatives to least squares regression. *Journal of the American Statistical Association,* **72**, 828-833.

Huber, P. J. (1981). Robust Statistics. New York: Wiley.

Jurečková, J. and Sen, P. K. (1996). Robust Statistical Procedures, Asymptotics and Interrelations. New York: Wiley.

Klein, R. and Yohai, V. J. (1981). Asymptotic behavior of iterative M-estimators for the linear model. *Comm. Statist. A,* **10**, 2373-2388.

Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics,* **11**, 431-441.

McCullagh, P. and Nelder, J. A. (1989). Generalized Linear Models 2ed. UK: Chapman and Hall.

Seber, G. A. and Wild, C. J. (1989). Nonlinear Regression. New York: Wiley.

Sen, P. K. (1982). On M-tests in linear models. *Biometrika,* **69**, 245-248.

Sen, P. K. and Singer, J. M. (1993). Large Sample Methods in Statistics: An Introduction with Applications. New York: Chapman and Hall.

Sen, P. K. (1998). Generalized Linear and Additive Models: Robustness Perspectives. *Brazilian Journal of Probability and Statistics,* **12**, 91-112.

Singer, J. M. and Sen. P. K. (1985). M-methods in multivariate linear models. *J. Multivar. Anal.,* **17**, 168-184.

Vonesh, R. F. and Chinchilli, V. M. (1997). Linear and Nonlinear Models for the Analysis of Repeated Measurements. New York: Marcel Dekker, Inc.

Yohai, V. J. and Maronna, R. A. (1979). Asymptotic behaviour of M-estimators for the linear model. *Ann. Statist.,* **7**, 258-268.

Antonio I. Sanhueza
Departamento de Matemáticas y Estadística - CIGES
Universidad de La Frontera
Temuco, Chile

Pranab K. Sen
Departments of Biostatistics, and Statistics and Operations Research
University of North Carolina, Chapel Hill, USA
E-mail: pksen@bios.unc.edu