View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Servicio de Difusión de la Creación Intelectual

Bernoulli **18**(4), 2012, 1284–1309
DOI: 10.3150/11-BEJ368

Continuity and differentiability of regression M functionals

MARÍA V. FASANO[1,*], RICARDO A. MARONNA[1,**], MARIELA SUED[2] and VÍCTOR J. YOHAI[3]

[1]*Departamento de Matemática, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calles 50 y 115, 1900 La Plata, Argentina. E-mail:* [*]*virfeather@yahoo.com.ar;*[**]*rmaronna@retina.ar*
[2]*Instituto de Cálculo, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 1, 1426 Buenos Aires, Argentina and CONICET. E-mail: msued@dm.uba.ar*
[3]*Departamento de Matemática, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, 1426 Buenos Aires, Argentina and CONICET. E-mail: vyohai@dm.uba.ar*

This paper deals with the Fisher-consistency, weak continuity and differentiability of estimating functionals corresponding to a class of both linear and nonlinear regression high breakdown M estimates, which includes S and MM estimates. A restricted type of differentiability, called weak differentiability, is defined, which suffices to prove the asymptotic normality of estimates based on the functionals. This approach allows to prove the consistency, asymptotic normality and qualitative robustness of M estimates under more general conditions than those required in standard approaches. In particular, we prove that regression MM-estimates are asymptotically normal when the observations are ϕ-mixing.

Keywords: asymptotic normality; consistency; MM estimates; nonlinear regression; S estimates

1. Introduction

We consider estimation in the regression model with random predictors

$$y_i = g(x_i, \beta_0) + u_i, \tag{1.1}$$

with data $(x_i, y_i) \in R^p \times R$, $i = 1, \ldots, n$; where $\beta_0 \in B \subseteq R^q$ is a vector of unknown parameters, $g(x, \beta)$ is a known function continuous in β, and for each i, x_i and u_i are independent. It is assumed that $\{(x_i, y_i), i \geq 1\}$ are identically distributed but not necessarily independent. The well-known fact that the least squares (LS) estimate of β_0 is sensitive to atypical observations has motivated the development of robust estimates.

An important class of robust estimators are the *M estimates*. Inside this class we can distinguish the S estimates introduced by Rousseeuw and Yohai [22] and the MM estimates proposed by Yohai [28]. For linear regression, S estimates may attain the highest possible breakdown point, and MM estimates may combine the highest possible breakdown point with a high normal efficiency; see, for example, [19], Chapter 5. In the case of nonlinear regression, MM estimates may

1350-7265 © 2012 ISI/BS

also combine high breakdown point with a high normal efficiency. In fact, the normal efficiency of these estimates can be made as close to one as desired, and Monte Carlo simulations in Fasano [10] show them to have a highly robust behavior for some nonlinear models.

In the nonlinear case, Fraiman [12] studied bounded influence estimates for nonlinear regression. Sakata and White [23] dealt with S estimates for nonlinear regression models with dependent observations; Vainer and Kukush [26] and Liese and Vajda [17,18] dealt with M estimates with a fixed scale, which therefore lack scale equivariance. The latter study the $\sqrt{n}$-consistency of M estimates in more general models, which include linear and nonlinear regression with independent observations. Stromberg [24] proved the weak consistency of the least median of squares (LMS) estimate, and Cízek [4] dealt with the consistency and the asymptotic normality of the least trimmed squares (LTS) estimate under dependency.

Three important qualitative features of an estimate are consistency, asymptotic normality and qualitative robustness. These properties have been studied in the literature through specific approaches. Yohai [28] proved these properties for MM estimates in the i.i.d. linear case, and Fasano [10] proved them in the nonlinear case, both assuming symmetrically distributed u_i's.

In this work, we propose an alternative approach, based on the representation of the estimates as *functionals* on distributions (Hampel [13]). For a large class of estimates, which includes M estimates, one can define a functional $T(G)$ on the space of data distributions, such that if G_n is the empirical distribution, then $T(G_n)$ is the estimate, and if G_0 is the underlying distribution, then $T(G_0)$ is the parameter that we want to estimate. The weak continuity of the functional T simplifies the proof of consistency of $T(G_n)$ and some suitable forms of differentiability of T, as Fréchet or Hadamard differentiability, allow simple proofs of the asymptotic normality of the estimate under very general conditions. These results hold without the requirement that G_n be the empirical distribution of a sequence of i.i.d. random variables: if we want to estimate $T(G_0)$, it suffices that G_n converges weakly to G_0 a.s. The weak continuity of M functionals at a general statistical model were studied by Clarke [5] and [6]. Fréchet differentiability was studied by Boos and Serfling [3] and Clarke [5], and Hadamard differentiability by Fernholz [11]. In all of these works, it is required that the score function used for the M estimate be bounded, and therefore their results can not be applied to regression M estimates. In this paper, we prove under very general conditions that the functionals associated to M estimates of regression are weakly continuous. Besides, since the usual forms of differentiability, like Fréchet or Hadamard differentiability, require in the case of M estimates the boundedness of the score functions, we introduce a new concept of differentiability, that we call *weak differentiability*, which is satisfied by high breakdown M estimates of regression, for example, by S and MM estimates, and which is adequate to prove the asymptotic normality of these estimates.

This work is organized as follows: In Section 2, we define the estimates to be considered and in Sections 3, 4 and 5 we shall respectively deal with the Fisher-consistency, continuity and differentiability of the functionals corresponding to the estimates defined above. These results will be shown to imply the consistency, qualitative robustness and asymptotic normality of the estimates under assumptions more general than the i.i.d. model and without the requirement of symmetric errors. In Section 6, we apply the results obtained in the former sections to MM estimates. Finally, Section 7 contains all proofs.

2. Definitions of estimates

We first define our notation. Henceforth, $\mathrm{E}_G[h(z)]$ and $\mathrm{P}_G(A)$ will respectively denote the expectation of $h(z)$ and the probability that $z \in A$, when z is distributed according to G. If z has distribution G, we write $z \sim G$ or $\mathcal{D}(z) = G$. Weak convergence of distributions, convergence in probability and convergence in distribution of random variables or vectors are denoted by $G_n \to_w G$, $z_n \to_p z$ and $z_n \to_d z$, respectively. By an abuse of notation, we will write $z_n \to_d G$ to denote $\mathcal{D}(z_n) \to_w G$. The complement and the indicator of the set A are denoted by A^c and $\mathbf{1}_A$, respectively. The scalar product of vectors a and b is denoted by $a'b$, and R_+ denotes the set of positive real numbers.

To identify β_0, without assuming that the distribution of u is symmetric around 0 or that it satisfies a centering condition (such as e.g. $\mathrm{E}_{F_0} u = 0$), we assume the following

Condition 1. *For all $\beta \neq \beta_0$ and for all α, we have*

$$\mathrm{P}\big(g(x, \beta_0) = g(x, \beta) + \alpha\big) < 1. \tag{2.1}$$

Note that when this condition is not satisfied, there exist $\beta \neq \beta_0$ and α such that (1.1) also holds with β instead of β_0 and $u_i + \alpha$ instead of u_i. Condition 1 requires that in case there is an intercept, it will be included in the error term u instead of as a parameter of the regression function $g(x, \beta)$. For linear regression, we have $g(x, \beta) = \beta' x$ and then this condition means that the vector x is not concentrated on any hyperplane.

Although model (1.1) does not contain an intercept, in order to obtain consistent estimates of β_0, our M estimates, besides an estimate $\widehat{\beta}$ of β_0, will include an additional additive term $\widehat{\alpha}$. If the model does contain an intercept, then $\widehat{\alpha}$ will be a consistent estimate of this parameter under the centering condition $\mathrm{E}\rho'(u/\sigma) = 0$, where ρ is the loss function of the M estimate and σ is the asymptotic value of the estimate of the error scale that is used to define the M estimate. If the model does not contain an intercept, then $\widehat{\alpha}$ can be ignored. Let henceforth $\xi = (\beta', \alpha)'$ with $\alpha \in R$, and define the function

$$\underline{g}(x, \xi) = g(x, \beta) + \alpha.$$

M estimates are then defined as

$$\widehat{\xi}_{\mathrm{M}} = \arg \min_{\xi \in B \times R} \sum_{i=1}^{n} \rho\left(\frac{y_i - \underline{g}(x_i, \xi)}{\widehat{\sigma}}\right), \tag{2.2}$$

where $\widehat{\sigma}$ is a robust residual scale and ρ is a loss function.

To define *S estimates*, we need an M scale $S(r)$. Given $r = (r_1, \ldots, r_n)'$, $S(r)$ is defined as the solution σ of

$$\frac{1}{n} \sum_{i=1}^{n} \rho_0\left(\frac{r_i}{\sigma}\right) = \delta, \tag{2.3}$$

where ρ_0 is another loss function and the constant δ regulates the estimate's robustness.

Then, S estimates of regression are defined by

$$\widehat{\xi}_{\mathrm{S}} = \arg \min_{\xi \in B \times R} S(r(\xi)), \tag{2.4}$$

where $r(\xi)$ is the residual vector with elements $r_i(\xi) = y_i - \underline{g}(x_i, \xi)$.

In particular, we will consider with some detail the subclass of MM estimates. These estimates are defined by (2.2) with $\widehat{\sigma}$ obtained from an S estimate, namely

$$\widehat{\sigma} = \min_{\xi \in B \times R} S(r(\xi)) \tag{2.5}$$

with $\rho \leq \rho_0$. Yohai [28] showed that in the case of linear regression the asymptotic breakdown point of MM estimates with $\delta = 0.5$ is 0.5 if $\mathrm{P}(\beta' x_i + a = 0) = 0$ for all $\beta \neq 0$, and that, simultaneously, it is possible to choose ρ so that the corresponding MM estimate yields an arbitrarily high efficiency when the errors are Gaussian.

Now in order to state our results, we must first express the already defined M and S estimates as functionals. Throughout this article, loss functions will be "bounded ρ-functions," in the following sense.

Definition 1. *A* bounded ρ-function *is a function $\rho(t)$ that is a continuous nondecreasing function of $|t|$, such that $\rho(0) = 0$, $\rho(\infty) = 1$, and $\rho(v) < 1$ implies that $\rho(u) < \rho(v)$ for $|u| < |v|$.*

Then, in the rest of the paper we will assume the following.

Condition 2. *ρ and ρ_0 are bounded ρ-functions.*

Define the residual scale functional $S^*(G, \xi)$ by

$$\mathrm{E}_G \rho_0 \left(\frac{y - \underline{g}(x, \xi)}{S^*(G, \xi)} \right) = \delta, \tag{2.6}$$

with $\delta \in (0, 1)$. Then the regression S functional T_{S} and the associated error scale M functional $S(G)$ are, respectively, defined by

$$T_{\mathrm{S}}(G) := (T_{\mathrm{S},\beta}(G), T_{\mathrm{S},\alpha}(G)) = \arg \min_{\xi \in B \times R} S^*(G, \xi) \tag{2.7}$$

and

$$S(G) = \min_{\xi \in B \times R} S^*(G, \xi). \tag{2.8}$$

We will deal with a regression M functional $T_{\mathrm{M}}(G)$ defined as

$$T_{\mathrm{M}}(G) := (T_{\mathrm{M},\beta}(G), T_{\mathrm{M},\alpha}(G)) = \arg \min_{\xi \in B \times R} M_G(\xi), \tag{2.9}$$

where the function $M_G : B \times R \to R$ is

$$M_G(\xi) = \mathrm{E}_G \rho \left(\frac{y - \underline{g}(x, \xi)}{\widetilde{S}(G)} \right) \tag{2.10}$$

and $\widetilde{S}(G)$ is an arbitrary residual scale functional, for example, the one defined in (2.8).

It is easy to show that the S regression functional defined in (2.7) is also an M functional. In fact, $T_{\mathrm{S}}(G)$ coincides with $T_{\mathrm{M}}(G)$ when in (2.10) we have $\rho = \rho_0$ and $\widetilde{S}(G) = S(G)$. We may then write

$$T_{\mathrm{S}}(G) = \arg\min_{\xi \in B \times R} \mathrm{E}_G \rho_0\left(\frac{y - \underline{g}(x, \xi)}{S(G)}\right). \tag{2.11}$$

Remark 1. In general, the minimum at (2.7) or (2.9) might be attained at more than one value of ξ. It will be henceforth assumed that the functional is well-defined by the choice of a single value. Our results will not depend on how the choice is made. However, it will be shown in Section 3 that under very general conditions, if G_0 is the distribution of (x, y) satisfying (1.1), then $T_{\mathrm{S}}(G_0)$ and $T_{\mathrm{M}}(G_0)$ are unique and $T_{\mathrm{S},\beta}(G_0) = T_{\mathrm{M},\beta}(G_0) = \beta_0$ (Fisher-consistency).

3. Fisher-consistency of M and S estimates

In this section, we give sufficient conditions to guarantee that both (2.7) and (2.9) are minimized at unique values, and that $T_{\mathrm{M},\beta}(G_0) = T_{\mathrm{S},\beta}(G_0) = \beta_0$ (Fisher consistency for β_0).

Recall that a density f is *strongly unimodal* if there exists a such that $f(t)$ is nondecreasing for $t < a$, nonincreasing for $t > a$, and f has a unique maximum at $t = a$.

We will need the following condition on ρ.

Condition 3. *The function ρ is a ρ-function such that for some $m > 0$, $\rho(u) = 1$ iff $|u| \geq m$, and* $\log(1 - \rho)$ *is concave on* $(-m, m)$.

It is easy to check that Condition 3 with $m = k$ holds in particular for the popular family of bisquare functions, defined by

$$\rho_k(u) = 1 - \left(1 - \left(\frac{u}{k}\right)^2\right)^3 I(|u| \leq k).$$

We will establish the Fisher-consistency of T_{M}. Put for brevity $\sigma = S(G_0)$ and let F_0 be the distribution of u_i in (1.1) and assume that it has a strongly unimodal density. Let Δ denote the unique minimizer of $\mathrm{E}_{F_0}\rho((u - t)/\sigma)$; note that if u_i is symmetric around μ_0, then part (b) of Theorem 3 implies that $\Delta = \mu_0$.

Theorem 1. *Let G_0 be the joint distribution of (x_i, y_i) satisfying model* (1.1), *where u_i has distribution F_0 with a strongly unimodal density. Assume that Conditions* 1 *and* 3 *hold. Then $M_{G_0}(\xi)$ is minimized at the unique point $T_{\mathrm{M}}(G_0) = (\beta_0, \Delta)$, and so T_{M} is Fisher-consistent for β_0, that is, $T_{\mathrm{M},\beta}(G_0) = \beta_0$. If we also assume that F_0 is symmetric around μ_0, we have $T_{\mathrm{M},\alpha}(G_0) = \mu_0$.*

Remark 2. Theorem 1 gives also sufficient conditions for the Fisher-consistency of the regression S functional T_{S}. In fact, according to (2.11), T_{S} is also an M functional.

4. Weak continuity of M and S regression functionals

Definition 2. *We say that a functional T is* weakly continuous *at G if $G_n \to_w G$ implies $T(G_n) \to T(G)$.*

We will show the weak continuity of the functionals defined above in two cases: nonlinear regression with a compact parameter space B, and linear regression.

Define for $G = \mathcal{D}(x, y)$

$$c(G) = \sup\{\mathrm{P}_G(\beta' x + \alpha = 0)\colon\ \beta \in R^p, \alpha \in R, \beta \neq 0\}. \tag{4.1}$$

Theorem 2. *Let $G_0 = \mathcal{D}(x, y)$ be such that* (2.9) *has a unique solution $T_{\mathrm{M}}(G_0)$. Assume that $\widetilde{S}$ is weakly continuous at G_0 and $\widetilde{S}(G_0) > 0$. Then $T_{\mathrm{M}} = (T_{\mathrm{M},\beta}, T_{\mathrm{M},\alpha})$ is weakly continuous at G_0 if either* (a) *or* (b) *holds, where*

(a) *B is compact,*
(b) *$B = R^p$, $g(x, \beta) = \beta' x$ and*

$$M_{G_0}(T_{\mathrm{M}}(G_0)) < 1 - c(G_0). \tag{4.2}$$

Theorem 3. *Let $G_0 = \mathcal{D}(x, y)$ be such that $T_{\mathrm{S}}(G_0)$ is unique and $S(G_0) > 0$. Assume that either* (a) *B is compact, or* (b) *$B = R^p$, g is linear, that is, $g(x, \beta) = \beta' x$ and $\delta < 1 - c(G_0)$ with $c(G)$ defined in* (4.1). *Then $S(G)$ and $T_{\mathrm{S}}(G) = (T_{\mathrm{S},\beta}, T_{\mathrm{S},\alpha})$ are weakly continuous at G_0.*

Let now G_0 be the distribution of (x, y) under model (1.1), and assume that T_{M} (resp., T_{S}) is Fisher-consistent for β_0, that is, $T_{\mathrm{M},\beta}(G_0) = \beta_0$ (resp., $T_{\mathrm{S},\beta}(G_0) = \beta_0$). Then the former results imply that $T_{\mathrm{M},\beta}$ (resp., $T_{\mathrm{S},\beta}$) evaluated at the empirical distribution is consistent whenever the empirical distributions converge to the underlying one. More precisely, we have the following result.

Corollary 1. *Assume the same hypotheses as in Theorem* 2 (*resp., Theorem* 3) *plus the Fisher-consistency of T_{M}* (*resp.,* T_{S})*: $T_{\mathrm{M},\beta}(G_0) = T_{\mathrm{S},\beta}(G_0) = \beta_0$. Call G_n the empirical distribution of $\{(x_i, y_i)\colon\ i = 1, \ldots, n\}$. If $G_n \to_w G_0$ a.s., then $\{T_{\mathrm{M},\beta}(G_n)\}$* (*resp.,* $\{T_{\mathrm{S},\beta}(G_n)\}$) *is strongly consistent for β_0.*

This result is immediate. The a.s. weak convergence of G_n to G_0 is well known to hold for i.i.d. (x_i, y_i). It holds also under more general assumptions on the joint distribution of $\{(x_i, y_i)\colon\ i \geq 1\}$, such as ergodicity.

We now turn to qualitative robustness. Consider a sequence of estimates $\{\widehat{\xi}_n\}$ based on a functional T, that is, $\widehat{\xi}_n = T(G_n)$ where G_n is the empirical distribution corresponding to data $(z_1, \ldots, z_n)$. Hampel [13] proved that for $\{\widehat{\xi}_n\}$ to be qualitatively robust at a distribution G_0 it suffices that T be weakly continuous at G_0 and $\widehat{\xi}_n$ be a continuous function of $(z_1, \ldots, z_n)$.

Papantoni-Kazakos and Grey [21] employ a weaker definition of robustness, which they call *asymptotic qualitative robustness*, and prove that it is equivalent to weak continuity. Therefore, Theorems 2 and 3 imply the asymptotic qualitative robustness of T_{M} and T_{S}.

5. Differentiability of estimating functionals

In this section, we shall first deal with the differentiability of general functionals and then specialize to our regression case. Let $\mathcal{G}_h$ be a set of distributions on R^h. Consider an estimating functional $T : \mathcal{G}_h \to R^k$. Hampel [14] defines the *influence function* of T at $G \in \mathcal{G}_h$ as the function $I_{T,G}(z) : R^h \to R^k$

$$I_{T,G}(z) = \left.\frac{\partial(T((1-\varepsilon)G + \varepsilon\delta_z))}{\partial\varepsilon}\right|_{\varepsilon=0}, \tag{5.1}$$

where δ_z is the point mass distribution at z. Given a distance d on $\mathcal{G}_h$ which metricizes the topology of convergence in distribution, T is *Fréchet differentiable* at G_0 under d if

$$T(G) - T(G_0) = \mathrm{E}_G I_{T,G_0}(z) + \mathrm{o}(d(G, G_0)).$$

Fréchet differentiability can be used to prove the asymptotic normality of the estimate. However, Fréchet differentiability also requires that $I_{T,G}(z)$ be bounded. Since this condition is not satisfied by regression M estimates, we are going to define a weaker type of differentiability, which suffices to prove asymptotic normality.

Definition 3. *Let T be an estimating functional that is weakly continuous at G_0. We say that T is* weakly differentiable *at a sequence $\{G_n\}$ converging weakly to G_0 if*

$$T(G_n) - T(G_0) = \mathrm{E}_{G_n} I_{T,G_0}(z) + \mathrm{o}(\|\mathrm{E}_{G_n} I_{T,G_0}(z)\|). \tag{5.2}$$

The definition of weak differentiability helps understanding the asymptotic behavior of $T(G_n) - T(G_0)$, as the next lemma shows.

Lemma 1. *Consider a random sequence of distributions $\{G_n\}$ converging weakly to G_0 a.s. Assume that T is weakly differentiable at $\{G_n\}$ a.s. and that for some sequence $\{a_n\}$ of real numbers*

$$a_n \mathrm{E}_{G_n} I_{T,G_0}(z) \to_d H.$$

Then

$$a_n\bigl(T(G_n) - T(G_0)\bigr) = a_n \mathrm{E}_{G_n} I_{T,G_0}(z) + \mathrm{o}_p(1) \tag{5.3}$$

and therefore $a_n(T(G_n) - T(G_0)) \to_d H$ too.

The proof of this lemma is immediate.

Remark. Note that if (5.3) holds for a joint functional $T = (T_1, T_2)$, it also holds for T_1, that is,

$$a_n\bigl(T_1(G_n) - T_1(G_0)\bigr) = a_n \mathrm{E}_{G_n} I_{T_1,G_0}(z) + \mathrm{o}_p(1). \tag{5.4}$$

We now deal with the differentiability of a *general M estimating functional*, that is, a functional T defined on a subset of $\mathcal{G}_p$ with values in R^q, that for some function $\Psi : R^p \times R^q \to R^q$ satisfies the equation

$$\mathrm{E}_G \Psi(z, T(G)) = 0. \tag{5.5}$$

We will assume that Ψ is continuously differentiable with respect to θ and call $\dot{\Psi}(z,\theta)$ (or alternatively $\partial\Psi(z,\theta)/\partial\theta$) the $q \times q$ differential matrix with elements $\dot{\Psi}_{jk}(z,\theta) = \partial\Psi_j(z,\theta)/\partial\theta_k$. Define

$$D(G,\theta) = \mathrm{E}_G \dot{\Psi}(z,\theta). \tag{5.6}$$

Let $\theta_0 = T(G_0)$ and assume that

$$D_0 = D(G_0, \theta_0) \tag{5.7}$$

exists. Assume that T is weakly continuous at G_0 and that the following holds.

Condition 4. *D_0 is nonsingular and there exists $\eta > 0$ such that*

$$\mathrm{E}_{G_0} \sup_{\|\theta-\theta_0\|\le\eta} \|\dot{\Psi}(z,\theta)\| < \infty, \tag{5.8}$$

where $\|\cdot\|$ denotes the l_2 norm.

Then, it is easy to show that the influence function of T at G_0 is given by

$$I_{T,G_0}(z) = -D_0^{-1}\Psi(z,\theta_0). \tag{5.9}$$

We shall now see that the following conditions are sufficient for the weak differentiability of T at $\{G_n\}$.

Condition 5. *$\{G_n\}$ is a sequence of distribution functions that converges weakly to G_0 and*

$$\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{\|\theta-\theta_0\|\le\eta} \|D(G_n,\theta) - D_0\| = 0. \tag{5.10}$$

Condition 6. *$\{G_n\}$ is a sequence of distribution functions such that, at a neighborhood of θ_0, for each n*

$$\frac{\partial}{\partial\theta}\mathrm{E}_{G_n}\Psi(z,\theta) = \mathrm{E}_{G_n}\frac{\partial}{\partial\theta}\Psi(z,\theta). \tag{5.11}$$

Condition 5 means that $D(G_n,\theta)$ approaches $D(G_0,\theta_0)$ when n is large and θ is close to θ_0. Condition 6 means that we can interchange differentiation of $\Psi(z,\theta)$ with respect to θ and expectation with respect to G_n. Theorem 4 shows that these two conditions imply weak differentiability and Theorem 5 shows that these conditions hold in very general circumstances.

Theorem 4. *Assume that T is an M functional satisfying* (5.5) *and weakly continuous at G_0, that $\dot{\Psi}(z,\theta)$ is continuous in θ, and that Condition* 4 *holds. If $\{G_n\}$ satisfies Conditions* 5 *and* 6; *then T is weakly differentiable at $\{G_n\}$.*

The following theorem gives sufficient conditions for a.s. differentiability of M functionals, at a random sequence of distributions.

Theorem 5. *Let $\{G_n\}$ be a sequence of random distributions converging weakly to G_0 and satisfying Condition* 6 *a.s. Assume also that $\dot{\Psi}(z,\theta)$ is continuous in θ and that Condition* 4 *holds. Let T be an M functional satisfying* (5.5) *and weakly continuous at G_0. Then T is weakly differentiable at $\{G_n\}$ a.s. in any of the following two cases*: (a) *for each function $d(z)$ such that $\mathrm{E}_{G_0}|d(z)|<\infty$, on a set of probability one we have that $\{\mathrm{E}_{G_n}d(z)\}$ converges to $\mathrm{E}_{G_0}d(z)$, or* (b) *$\dot{\Psi}(z,\theta)$ is bounded.*

Note case (a) contains situations where a Law of Large Numbers holds, in particular when G_n is the empirical distribution of an ergodic process.

Corollary 2. *Let $\{G_n\}$ be a sequence of empirical distributions associated to i.i.d. $\{z_i\}$ with distribution G_0. Assume that $\dot{\Psi}(z,\theta)$ is continuous in θ, that Condition* 4 *holds and that $I_{T,G_0}(z)$ has finite second moments under G_0. Let T be an M functional continuous at G_0. Then $n^{1/2}(T(G_n)-T(G_0))\rightarrow_d N(0,V)$ with*

$$V=\mathrm{E}_{G_0}I_{T,G_0}(z)I_{T,G_0}(z)'. \tag{5.12}$$

There are many examples where Fréchet differentiability does not hold and that can be dealt with using the concept of weak differentiability. One of these cases is that of MM estimates for linear an nonlinear regression which is treated in detail in the next section. Other examples where Fréchet differentiability fails are MM estimates for the multivariate linear model (see [16]) and M estimates for logistic models (see [1] and [8]). An example where the asymptotic expansions that can be obtained with weak differentiability are essential to prove asymptotic normality is the problem of robust estimation with missing data considered by Sued and Yohai [25].

6. MM estimates

In this section, we will summarize the properties derived from Theorems 1–6 for S and MM estimates of regression and location.

6.1. Regression case

Recall that MM estimates, which we denote here by $T_{\mathrm{MM}}=(T_{\mathrm{MM},\beta},T_{\mathrm{MM},\alpha})$, are defined in (2.9), where $\widetilde{S}$ is the functional S defined in (2.8). For notational convenience, we shall call ρ_1 the ρ-function employed in (2.10), and we will assume that $\rho_1\leq\rho_0$. As mentioned above, the definition of $\widehat{\xi}_{\mathrm{MM}}$ in (2.2) requires also $\widehat{\sigma}$ defined by (2.5), and hence also $\widehat{\xi}_{\mathrm{S}}$ defined in (2.4). Therefore, these three estimates must be considered simultaneously. Call

$$\widehat{\theta} = (\widehat{\xi}_{\mathrm{S}}, \widehat{\xi}_{\mathrm{MM}}, \widehat{\sigma}) \tag{6.1}$$

the joint solution of (2.2)–(2.4)–(2.5).

In the rest of this section, we assume the following properties.

Condition 7. *ρ_0 and ρ_1 are twice continuously differentiable.*

We denote by ψ_0 and ψ_1 the derivatives of ρ_0 and ρ_1, respectively. Assume also the following condition.

Condition 8. *g is twice continuously differentiable with respect to β.*

We denote by $\dot{\underline{g}}(x,\xi)$ and $\ddot{\underline{g}}(x,\xi)$ the vector of first derivatives and the matrix of second derivatives of $\underline{g}$ with respect to ξ, respectively. Analogously, we denote by $\dot{g}(x,\beta)$ and $\ddot{g}(x,\beta)$ the vector of first derivatives and the matrix of second derivatives of g with respect to β, respectively. Note that $\dot{\underline{g}}(x,\xi)$ and $\ddot{\underline{g}}(x,\xi)$ depend only on β, and for this reason we will indistinctly use also the notation $\dot{\underline{g}}(x,\beta)$ and $\ddot{\underline{g}}(x,\beta)$.

Differentiating (2.2) we have that $\widehat{\xi}_{\mathrm{MM}}$ satisfies the system

$$\frac{1}{n}\sum_{i=1}^{n}\psi_1\left(\frac{y_i - \underline{g}(x_i,\widehat{\xi}_{\mathrm{MM}})}{\widehat{\sigma}}\right)\dot{\underline{g}}(x_i,\widehat{\xi}_{\mathrm{MM}}) = 0. \tag{6.2}$$

It is immediate that $\widehat{\xi}_{\mathrm{S}}$ also satisfies

$$\widehat{\xi}_{\mathrm{S}} = \arg\min_{\xi\in B\times R}\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{y_i - \underline{g}(x_i,\xi)}{\widehat{\sigma}}\right).$$

Then, differentiating this equation we get

$$\frac{1}{n}\sum_{i=1}^{n}\psi_0\left(\frac{y_i - \underline{g}(x_i,\widehat{\xi}_{\mathrm{S}})}{\widehat{\sigma}}\right)\dot{\underline{g}}(x_i,\widehat{\xi}_{\mathrm{S}}) = 0. \tag{6.3}$$

Finally according to (2.3), $\widehat{\sigma}$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\rho_0\left(\frac{y_i - \underline{g}(x_i,\widehat{\xi}_{\mathrm{S}})}{\widehat{\sigma}}\right) - \delta = 0. \tag{6.4}$$

Then $\widehat{\theta}$ satisfies the system of $2q+3$ equations (6.2)–(6.4). Putting $z_i = (x_i, y_i)$ and denoting by G_n the empirical distribution of $\{z_1,\dots,z_n\}$, this system can be written as

$$\frac{1}{n}\sum_{i=1}^{n}\Psi(z_i,\widehat{\theta}) = \mathrm{E}_{G_n}\Psi(z,\widehat{\theta}) = 0, \tag{6.5}$$

where if $\theta = (\xi_S, \xi_{MM}, \sigma)$, $\Psi(z, \theta)$ is defined by

$$\Psi(z,\theta) = \begin{bmatrix} \psi_0\left(\dfrac{y - \underline{g}(x, \xi_S)}{\sigma}\right)\underline{\dot{g}}(x, \xi_S) \\ \psi_1\left(\dfrac{y - \underline{g}(x, \xi_{MM})}{\sigma}\right)\underline{\dot{g}}(x, \xi_{MM}) \\ \rho_0\left(\dfrac{y - \underline{g}(x, \xi_S)}{\sigma}\right) - \delta. \end{bmatrix}.$$

Let

$$T(G) = (T_S(G), T_{MM}(G), S(G)) \tag{6.6}$$

be the estimating functional associated to $\widehat{\theta}$. Then, if (5.8) holds, we can differentiate the functions to be minimized in (2.9) and (2.11) inside the expectation, obtaining that $T(G)$ satisfies the equation

$$\mathrm{E}_G \Psi(z, T(G)) = 0. \tag{6.7}$$

Note that the solution to this equation is in general not unique, and therefore, T is not defined exclusively by the equation.

To verify (5.8), in addition to Conditions 2–3–7–8 we need the following assumption:

Condition 9. *For some $\eta > 0$*

$$\mathrm{E}_{G_0} \sup_{\|\beta-\beta_0\|\le\eta} \|\dot{g}(x,\beta)\|^2 < \infty \quad \textit{and} \quad \mathrm{E}_{G_0} \sup_{\|\beta-\beta_0\|\le\eta} \|\ddot{g}(x,\beta)\| < \infty. \tag{6.8}$$

Assume that D_0 defined by (5.7) is nonsingular; then under these assumptions, we also get that $I_{T,G_0}(z)$ has finite second moments under G_0. Note that in the case of linear regression, (6.8) reduces to $\mathrm{E}_{G_0}\|x\|^2 < \infty$.

Define

$$\alpha_{0i} = \arg\min_t \mathrm{E}_{F_0}\rho_i\left(\frac{u-t}{S(G_0)}\right), \qquad i = 0, 1, \tag{6.9}$$

where F_0 is the distribution of u_i in model (1.1). We will see in Theorem 6 that under some general conditions, $T_{S,\alpha}(G_0) = \alpha_{00}$ and $T_{MM,\alpha}(G_0) = \alpha_{01}$..

Put $\theta_0 = (\beta_0, \alpha_{00}, \beta_0, \alpha_{01}, \sigma_0)$ with $\sigma_0 = S(G_0)$. The following numbers, vectors and matrices are required to derive a closed formula for the influence functions of T_{MM} and T_S. Let

$$a_{0i} = \mathrm{E}_{G_0}\psi_i'\left(\frac{y - g(x,\beta_0) - \alpha_{0i}}{\sigma_0}\right) = \mathrm{E}_{F_0}\psi_i'\left(\frac{u-\alpha_{0i}}{\sigma_0}\right), \qquad i = 0, 1,$$

$$e_{0i} = \mathrm{E}_{F_0}\left(\frac{u-\alpha_{0i}}{\sigma_0}\right)\psi_0'\left(\frac{u-\alpha_{0i}}{\sigma_0}\right), \qquad i = 0, 1,$$

$$d_0 = \mathrm{E}_{F_0}\left(\frac{u-\alpha_{00}}{\sigma_0}\right)\psi_0\left(\frac{u-\alpha_{00}}{\sigma_0}\right),$$

$$b_0 = \mathrm{E}_{G_0}\dot{g}(x,\beta_0), \qquad b_0^* = (b_0', 1)',$$

$$A_0 = \mathrm{E}_{F_0}\big(\dot{g}(x,\beta_0) - b_0\big)\big(\dot{g}(x,\beta_0) - b_0\big)'$$

and

$$C_0 = \begin{bmatrix} A_0 + b_0 b_0' & b_0 \\ b_0' & 1 \end{bmatrix}. \tag{6.10}$$

It is shown in Section 7.4 that the influence function of T_{MM} is given by

$$I_{T_{\mathrm{MM},\beta},G_0}(x,y) = \frac{\sigma_0}{a_{01}}\psi_1\left(\frac{y - \underline{g}(x,(\beta_0,\alpha_{01}))}{\sigma_0}\right)A_0^{-1}\big(\dot{g}(x,\beta_0) - b_0\big) \tag{6.11}$$

and

$$\begin{aligned} I_{T_{\mathrm{MM},\alpha},G_0}(x,y) = &-\frac{\sigma_0}{a_{01}}\psi_1\left(\frac{y - \underline{g}(x,(\beta_0,\alpha_{01}))}{\sigma_0}\right)\big[1 + b_0' A_0^{-1}\big(b_0 - \dot{g}(x,\beta_0)\big)\big] \\ &+ \frac{\sigma_0 e_{01}}{a_{01} d_0}\left(\rho_0\left(\frac{y - \underline{g}(x,(\beta_0,\alpha_{01}))}{\sigma_0}\right) - \delta\right). \end{aligned} \tag{6.12}$$

The influence functions of $T_{\mathrm{S},\beta}$ and $T_{\mathrm{S},\alpha}$ can be obtained similarly replacing α_{01}, a_{01} and e_{01} by α_{00}, a_{00} and e_{00}, respectively.

If the errors u_i have a symmetric distribution F_0, then $e_{01} = 0$ and $\alpha_{01} = \alpha_{00} = \alpha_0$, the center of symmetry of F_0. This entails a considerable simplification of the influence function $I_{T_{\mathrm{MM}}}$. In fact, in this case, we get

$$I_{T_{\mathrm{MM}},G_0}(z) = \frac{\sigma_0}{\mathrm{E}_{F_0}\psi_1'((u-\alpha_0)/\sigma_0)}\psi_1\left(\frac{y - g(x,\beta_0) - \alpha_0}{\sigma_0}\right)C_0^{-1}\underline{\dot{g}}(x,\beta_0), \tag{6.13}$$

and the asymptotic covariance matrix (5.12) is

$$V = \sigma_0^2 \frac{\mathrm{E}_{F_0}\psi_1((u-\alpha_0)/\sigma_0)^2}{(\mathrm{E}_{F_0}\psi_1'((u-\alpha_0)/\sigma_0))^2} C_0^{-1}. \tag{6.14}$$

The next theorem summarizes the properties of S and MM regression functionals.

Theorem 6. *Let $z = (x, y)$ satisfy model (1.1) where the distribution F_0 of u_i has a strongly unimodal density and Condition 1 holds. Assume that ρ_0 and ρ_1 are bounded ρ-functions that satisfy Condition 3, with $\rho_1(u) \le \rho_0(u)$. Let T be defined by (6.6) and let G_0 be the distribution of (x, y). Then:*

(i) $T_{\mathrm{S}}(G_0) = (\beta_0, \alpha_{00})$ *is the unique minimizer in (2.7). If F_0 is symmetric with respect to μ_0, then $\alpha_{00} = \mu_0$.*

(ii) $T_{\mathrm{MM}}(G_0) = (\beta_0, \alpha_{01})$ *is the unique minimizer in* (2.9). *If* F_0 *is symmetric with respect to* μ_0 *then* $\alpha_{01} = \mu_0$.

(iii) *The functional* $T = (T_{\mathrm{S}}, T_{\mathrm{MM}}, S)$ *is weakly continuous at* G_0 *if either* (a) B *is compact, or* (b) $B = R^p$, $g(x, \beta) = \beta' x$ *and* $\delta < 1 - c(G_0)$.

(iv) *Assume also that Conditions* 7, 8 *and* 9 *hold, that* $a_{00} \neq 0$, $a_{01} \neq 0$, $d_0 \neq 0$ *and that* A_0 *is invertible. Then,* $D_0 = \mathrm{E}_{G_0} \dot{\Psi}(z, T(G_0))$ *is invertible,* $I_{T_{\mathrm{MM},\beta}, G_0}(x, y)$ *and* $I_{T_{\mathrm{MM},\alpha}, G_0}(x, y)$ *are given by* (6.11) *and* (6.12), *respectively, while the influence functions* $I_{T_{\mathrm{S},\beta}, G_0}(x, y)$ *and* $I_{T_{S,\alpha}, G_0}(x, y)$ *have a similar expression replacing* α_{01}, a_{01} *and* e_{01} *by* α_{00}, a_{00} *and* e_{00}, *respectively.*

(v) *Under the same assumptions as in* (iv), *let* $\{G_n\}$ *be a sequence of random distributions converging weakly to* G_0 *and satisfying Condition* 6 *a.s. Suppose also that for each function* $d(z)$ *such that* $\mathrm{E}_{G_0}|d(z)| < \infty$, *we have that* $\{\mathrm{E}_{G_n} d(z)\}$ *converges to* $\mathrm{E}_{G_0} d(z)$ *a.s. Then, the functional* T *is weakly differentiable at* $\{G_n\}$.

(vi) *Assume the same conditions as in* (v) *and*:

$$n^{1/2} \mathrm{E}_{G_n} I_{T, G_0}(x, y) \to_d H. \tag{6.15}$$

Then

$$n^{1/2}\bigl(T(G_n) - T(G_0)\bigr) = n^{1/2} \mathrm{E}_{G_n} I_{T, G_0}(x, y) + \mathrm{o}_p(1) \tag{6.16}$$

and therefore

$$n^{1/2}\bigl(T(G_n) - T(G_0)\bigr) \to_d H. \tag{6.17}$$

(vii) *Assume that the conditions in* (iv) *hold and that* $\{(x_i, u_i)\colon i \geq 1\}$ *are i.i.d. Let* G_n *be the sequence of empirical distributions corresponding to* $\{(x_i, y_i)\colon i \geq 1\}$ *with common distribution* G_0. *Then* (6.17) *holds with* $H = N(0, V)$ *and* $V = \mathrm{E}[I_{T_{\mathrm{MM}}, G_0}(x, y) I_{T_{\mathrm{MM}}, G_0}(x, y)']$, *where* $I_{T_{\mathrm{MM}}, G_0}(x, y)$ *is defined by* (6.11) *and* (6.12).

(viii) *Assume that the conditions in* (iv) *hold, that* $\{u_i\colon i \geq 1\}$ *is stationary and ergodic and that* $\{x_i, i \geq 1\}$ *are i.i.d. and independent of* $\{u_i\colon i \geq 1\}$. *Let* G_n *be the sequence of empirical distributions corresponding to* $\{(x_i, y_i)\colon i \geq 1\}$ *with common distribution* G_0. *Then*

$$n^{1/2}\bigl(T_{\mathrm{MM},\beta}(G_n) - \beta_0\bigr) \to_d N(0, V) \tag{6.18}$$

with

$$V = \sigma_0^2 \frac{\mathrm{E}_{F_0} \psi_1^2((u - \alpha_{01})/\sigma_0)}{\mathrm{E}_{F_0}^2 \psi_1'((u - \alpha_{01})/\sigma_0)} A_0^{-1}. \tag{6.19}$$

A similar result can be obtained for $T_{\mathrm{S},\beta}$.

(ix) *Assume that the conditions in* (iv) *hold, that* $\{(u_i, x_i)\colon i \geq 1\}$ *is* ϕ*-mixing (see, e.g., Billingsley* [2] *for the definition of* ϕ*-mixing) with* $\sum_{i=1}^{\infty} \phi_n^{1/2} < \infty$, *that* u_i *have a symmetric distribution* F_0 *and that* $\{x_i, i \geq 1\}$ *and* $\{u_i\colon i \geq 1\}$ *are independent. Let* G_n *be the sequence of empirical distributions corresponding to* $\{(x_i, y_i)\colon i \geq 1\}$ *with common distribution* G_0. *Then*

$$n^{1/2}\bigl(T_{\mathrm{MM}}(G_n) - T_{\mathrm{MM}}(G_0)\bigr) \to_d N(0, V), \tag{6.20}$$

where

$$
\begin{aligned}
V &= \frac{\sigma_0^2}{\mathrm{E}_{F_0}^2 \psi_1'((u-\alpha_0)/\sigma_0)} C_0^{-1}\left(\sum_{i=-\infty}^{\infty} c_i C_i\right) C_0^{-1}, \\
c_i &= \mathrm{E}\left[\psi_1\left(\frac{u_1-\alpha_0}{\sigma_0}\right)\psi_1\left(\frac{u_{1+i}-\alpha_0}{\sigma_0}\right)\right], \qquad (6.21)\\
C_i &= \mathrm{E}\underline{\dot{g}}(x_1,\beta_0)\underline{\dot{g}}(x_{1+i},\beta_0)'
\end{aligned}
$$

and $T_{\mathrm{MM}}(G_0) = (\beta_0, a_0)$.

Remark 3. Note that (viii) implies that the asymptotic covariance matrix of $n^{1/2}(T_{\mathrm{MM},\beta}(G_n) - \beta_0)$ is the same as when the u_i are i.i.d. This result does not hold for the intercept estimate $T_{\mathrm{MM},\alpha}(G_n)$. Croux, Dhaene and Hoorelbeke [7] derived a similar result for linear regression through the origin with one covariable with mean 0.

Remark 4. The ϕ-mixing condition in (ix) can be replaced by any other type of mixing condition that guarantees the validity of the central limit theorem (see, e.g., Section 1.5.1 of Doukhan [9]). A result similar to part (ix) of Theorem 6 was stated by Croux *et al*. [7].

6.2. Location case

The location model corresponds to the case where there are no regressors: $p = q = 0$ and so $y_i = u_i$ and $\xi = \alpha$. If F_0 denotes the common distribution of the u_i, then $T(F_0) = (T_{\mathrm{S}}(F_0), T_{\mathrm{MM}}(F_0), S(F_0))$ is defined as in the regression case with $\underline{g}(x,\xi)$ replaced by α. Then, the resulting $T_{\mathrm{MM}} = T_{\mathrm{MM},\alpha}$ and $T_{\mathrm{S}} = T_{\mathrm{S},\alpha}$ are the location functionals while S is a functional estimating the error scale. In this case, I_{T_{MM},F_0} is given by

$$
\begin{aligned}
I_{T_{\mathrm{MM}},F_0}(x) = {}& \frac{\sigma_0}{a_{01}}\psi_1\left(\frac{y-\alpha_{01}}{\sigma_0}\right) \\
& - \frac{e_{01}\sigma_0}{a_{01}d_0}\left(\rho_0\left(\frac{y-\alpha_{00}}{\sigma_0}\right) - \delta\right).
\end{aligned} \qquad (6.22)
$$

The following theorem summarizes the properties of T that can be derived from the theorems in the former sections.

Theorem 7. *Assume that ρ_0 and ρ_1 are bounded ρ-functions that satisfy Condition 3, with $\rho_1 \le \rho_0$. We assume that F_0 has a strong unimodal density. Then*

(i) $T_{\mathrm{S}}(F_0) = \alpha_{00}$ *is the unique minimizer in* (2.7). *If F_0 is symmetric with respect to μ_0, we have $\alpha_{00} = \mu_0$.*
(ii) $T_{\mathrm{MM}}(F_0) = \alpha_{01}$ *is the unique minimizer in* (2.9). *If F_0 is symmetric with respect to μ_0, we have $\alpha_{01} = \mu_0$.*

(iii) *The functional* $T = (T_{\mathrm{S}}, T_{\mathrm{MM}}, S)$ *is weakly continuous at* F_0.

(iv) *Assume also that Condition* 7 *holds and that* $a_{00} \neq 0$, $a_{01} \neq 0$, $d_0 \neq 0$. *Then*, $D_0 = \mathrm{E}_{F_0} \dot{\Psi}(z, T(F_0))$ *is invertible*, $I_{T_{\mathrm{MM}}, F_0}(y)$ *is given by* (6.22). *The influence function* $I_{T_{\mathrm{S}}, F_0}(y)$ *has a similar expression replacing* α_{01}, a_{01} *and* e_{01} *by* α_{00}, a_{00} *and* e_{00}, *respectively.*

(v) *Under the same assumptions as in* (iv), *let* $\{F_n\}$ *be a sequence of random distributions converging weakly to* F_0 *and satisfying Condition* 6 *a.s. Then* T *is a.s. weakly differentiable at* $\{F_n\}$.

(vi) *Assume the same conditions as in* (v) *and*

$$n^{1/2} \mathrm{E}_{F_n} I_{T, F_0}(y) \to_d H. \tag{6.23}$$

Then

$$n^{1/2}\bigl(T(F_n) - T(F_0)\bigr) = n_{F_n}^{1/2} \mathrm{E} I_{T, F_0}(y) + \mathrm{o}_p(1), \tag{6.24}$$

and therefore

$$n^{1/2}\bigl(T(F_n) - T(F_0)\bigr) \to_d H. \tag{6.25}$$

(vii) *Assume the same conditions as in* (iv). *Let* $\{F_n\}$ *be the sequence of empirical distributions corresponding to i.i.d. observations* u_i *with common distribution* F_0. *Then* (6.23) *holds with* $H = N(0, V)$ *and* V *given by* (5.12). *If* F_0 *is symmetric, the asymptotic variance of* T_{MM} *given by* (6.14) *becomes*

$$V = \sigma_0^2 \frac{\mathrm{E}_{F_0} \psi_1 (u/\sigma_0)^2}{(\mathrm{E}_{F_0} \psi_1'(u/\sigma_0))^2}.$$

7. Proofs

7.1. Proof of Theorem 1

We shall need the following auxiliary result, which is due to Ibragimov [15].

Lemma 2. *If* f *is a strongly unimodal density and* φ *is a density such that* $\log \varphi$ *is concave on its support*, *the convolution*

$$h(t) = \int_{-\infty}^{\infty} \varphi(u - t) f(u) \, \mathrm{d}u \tag{7.1}$$

is strongly unimodal.

The following lemma is a small variation of one given by Mizera [20].

Lemma 3. *Let* ρ *satisfy Condition* 3 *and let* F *be a distribution with a strongly unimodal density* f. *Then* (a) *there exists* t_0 *such that*

$$q(t) = \mathrm{E}_F \rho(u - t) \tag{7.2}$$

has a unique minimum at t_0; (b) *if* F *is symmetric around* μ_0, *then* $t_0 = \mu_0$.

Proof. (a) Put $k = \int_{-m}^{m} \rho(x)\,\mathrm{d}x$ and $\varphi(u) = (1 - \rho(u))/k$, which vanishes for $|u| > m$. Then

$$q(t) = 1 - \mathrm{E}_F\bigl(1 - \rho(u - t)\bigr) = 1 - k\mathrm{E}_F\varphi(u - t) = 1 - kh(t),$$

where $h(t)$ is given by (7.1). Since by Lemma 2 $h(t)$ is a strongly unimodal density, part (a) of the lemma follows.

(b) It is proved in Lemma 3.1 of Yohai [27]. □

Proof of Theorem 1. Without loss of generality we may assume $\sigma = 1$. To prove the theorem, we will show that the unique minimum of $R(\beta, \alpha) = \mathrm{E}_{G_0}\rho(y - g(x, \beta) - \alpha)$ is $\beta = \beta_0, \alpha = t_0$. We will first prove that

$$R(\beta_0, t_0) < R(\beta_0, \alpha) \qquad \text{for } \alpha \neq t_0.$$

This is equivalent to

$$\mathrm{E}_{F_0}\rho(u - t_0) < \mathrm{E}_{F_0}\rho(u - \alpha) \qquad \text{for } \alpha \neq t_0,$$

which follows from Theorem 3.

Consider now (β, α) with $\beta \neq \beta_0$. Let $A = \{x\colon\ g(x, \beta_0) = g(x, \beta) + \alpha - t_0\}$ and q as in (7.2), with F replaced by F_0. Then

$$\begin{aligned} R(\beta, \alpha) &= \mathrm{E}_{G_0}\bigl\{\mathrm{E}_{G_0}\bigl[\rho\bigl(y - g(x, \beta) - \alpha\bigr)|x\bigr]\bigr\} \\ &= \mathrm{E}_{G_0}\bigl\{\mathrm{E}_{G_0}\bigl[\rho\bigl(u + g(x, \beta_0) - g(x, \beta) - \alpha\bigr)|x\bigr]\bigr\}. \end{aligned} \tag{7.3}$$

Since u and x are independent, we get

$$\mathrm{E}\bigl[\rho\bigl(u + g(x, \beta_0) - g(x, \beta) - \alpha\bigr)|x\bigr] = q\bigl(g(x, \beta) - g(x, \beta_0) + \alpha\bigr). \tag{7.4}$$

Then according to Theorem 3, the left-hand side of (7.4) is equal to $q(t_0)$ if $x \in A$ and grater than $q(t_0)$ otherwise. Condition 1 implies that $\mathrm{P}(A^c) > 0$ and from (7.3) we get that $R(\beta, \alpha) > q(t_0)$. Finally, the theorem follows from the fact that $R(\beta_0, t_0) = q(t_0)$. □

7.2. Proof of Theorems 2 and 3

Before proving Theorems 2 and 3, we need some auxiliary results.

Lemma 4. *Consider distributions* $\{G_n\}$ *and* G_0 *on* $R^p \times R$. *Let* $\{\xi_n\}$ *and* $\{\sigma_n\}$ *be sequences in* $B \times R$ *and* R_+, *respectively, such that* $\xi_n \to \xi \in B \times R$ *and* $\sigma_n \to \sigma > 0$. *Assume that* $\underline{g}(x, \xi)$ *is continuous in* ξ. *If* $G_n \to_w G_0$, *then*

$$\lim_{n\to\infty} \mathrm{E}_{G_n}\rho\left(\frac{y - \underline{g}(x, \xi_n)}{\sigma_n}\right) = \mathrm{E}_{G_0}\rho\left(\frac{y - \underline{g}(x, \xi)}{\sigma}\right).$$

Proof. Since $G_n \to_w G_0$ and ρ is continuous and bounded, we have

$$\mathrm{E}_{G_n}\rho\left(\frac{y-\underline{g}(x,\xi)}{\sigma}\right) \to \mathrm{E}_{G_0}\rho\left(\frac{y-\underline{g}(x,\xi)}{\sigma}\right),$$

and therefore it suffices to show that

$$\mathrm{E}_{G_n}\rho\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) - \mathrm{E}_{G_n}\rho\left(\frac{y-\underline{g}(x,\xi)}{\sigma}\right) \to 0.$$

Since $\{G_n\}_{n\geq 1}$ is tight, it suffices to show that if $\mathcal{P}$ is a tight set of distributions of (x, y), then

$$\sup_{F\in\mathcal{P}}\left|\mathrm{E}_F\rho\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) - \mathrm{E}_F\rho\left(\frac{y-\underline{g}(x,\xi)}{\sigma}\right)\right| \to 0.$$

To prove this, put $z = (x, y)$. Then for all $K > 0$

$$\begin{aligned}&\left|\mathrm{E}_F\rho\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) - \mathrm{E}_F\rho\left(\frac{y-\underline{g}(x,\xi)}{\sigma}\right)\right| \\ &\quad\leq 2\mathrm{E}_F\mathbf{1}_{\{\|z\|>K\}} + \mathrm{E}_F\left|\rho\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) - \rho\left(\frac{y-\underline{g}(x,\xi)}{\sigma}\right)\right|\mathbf{1}_{\{\|z\|\leq K\}}.\end{aligned} \tag{7.5}$$

If $\|z\| \leq K$, we have

$$\begin{aligned}&\left|\frac{y-\underline{g}(x,\xi_n)}{\sigma_n} - \frac{y-\underline{g}(x,\xi)}{\sigma}\right| \\ &\quad\leq \frac{1}{\sigma\sigma_n}[|\sigma_n-\sigma||y| + |\sigma_n-\sigma||\underline{g}(x,\xi)| + \sigma|\underline{g}(x,\xi_n) - \underline{g}(x,\xi)|].\end{aligned} \tag{7.6}$$

Now, given $\varepsilon > 0$, we can find K such that

$$2\sup_{F\in\mathcal{P}} \mathrm{P}_F(\|z\| > K) \leq \varepsilon/2$$

and α such that

$$|\rho(u) - \rho(v)| \leq \varepsilon/2 \qquad \text{if } |u-v| \leq \alpha.$$

Then, we can choose n_0 such that the right-hand side of (7.6) is smaller than α if $n \geq n_0$ and $\|z\| \leq K$, and so from (7.5) we obtain for all $n \geq n_0$

$$\left|\mathrm{E}_F\rho\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) - \mathrm{E}_F\rho\left(\frac{y-\underline{g}(x,\xi)}{\sigma}\right)\right| \leq \varepsilon \qquad \forall F \in \mathcal{P}.$$

□

Lemma 5. *Assume that B is closed and let G_0 be any distribution for (x, y) such that (2.9) has a unique solution $T_{\mathrm{M}}(G_0)$. Let $\{G_n\}$ be a sequence such that $G_n \to_w G_0$ and $\{T_{\mathrm{M}}(G_n)\}$ is bounded. If $\widetilde{S}(G_n) \to \widetilde{S}(G_0) > 0$, then $T_{\mathrm{M}}(G_n) \to T_{\mathrm{M}}(G_0)$.*

Proof. Put for brevity

$$\xi_n = T_{\mathrm{M}}(G_n), \qquad \xi_0 = T_{\mathrm{M}}(G_0), \qquad \sigma_n = \widetilde{S}(G_n), \qquad \sigma_0 = \widetilde{S}(G_0). \tag{7.7}$$

Since $\{\xi_n\}$ remains in a compact set, it suffices to prove that ξ_0 is the only accumulation point of $\{\xi_n\}$, that is, if a subsequence tends to some $\widehat{\xi}$, then $\widehat{\xi} = \xi_0$. Without loss of generality, assume that $\xi_n \to \widehat{\xi}$. The definition of ξ_n implies

$$\mathrm{E}_{G_n}\rho\left(\frac{y - \underline{g}(x, \xi_n)}{\sigma_n}\right) \le \mathrm{E}_{G_n}\rho\left(\frac{y - \underline{g}(x, \xi_0)}{\sigma_n}\right). \tag{7.8}$$

Using Lemma 4, we get

$$M_{G_0}(\widehat{\xi}) = \mathrm{E}_{G_0}\rho\left(\frac{y - \underline{g}(x, \widehat{\xi})}{\sigma_0}\right) \le \mathrm{E}_{G_0}\rho\left(\frac{y - \underline{g}(x, \xi_0)}{\sigma_0}\right) = M_{G_0}(\xi_0).$$

Since ξ_0 is the only minimizer of M_{G_0}, we conclude that $\widehat{\xi} = \xi_0$. □

Lemma 6. *Let $\{\xi_n\}$ and $\{\sigma_n\}$ be sequences in R^{p+1} and R_+, respectively. Assume that when $n \to \infty$, $G_n \to_w G_0$, $\|\xi_n\| \to \infty$ and $\{\sigma_n\}$ is bounded. Then*

$$\liminf_{n\to\infty} \mathrm{E}_{G_n}\rho\left(\frac{y - \xi_n'(x', 1)'}{\sigma_n}\right) \ge 1 - c_0, \tag{7.9}$$

where $c_0 = c(G_0)$ is defined in (4.1).

Proof. Assume without loss of generality that there exist $\gamma \in R^p$ and $\sigma > 0$ such that for some subsequence $\gamma_n = \xi_n / \|\xi_n\| \to \gamma$, and $\sigma_n \le \sigma$. Put $\lambda_n = \|\xi_n\|$.

For $\varepsilon > 0$ let d_ε be such that $\rho(u) \ge 1 - \varepsilon$ for $|u| \ge d_\varepsilon$. Therefore,

$$\mathrm{E}_{G_n}\rho\left(\frac{y - \xi_n'(x', 1)'}{\sigma_n}\right) \ge \mathrm{E}_{G_n}\rho\left(\frac{y - \xi_n'(x', 1)'}{\sigma}\right) \ge (1 - \varepsilon)\mathrm{P}_{G_n}\left(\frac{|y - \lambda_n \boldsymbol{\gamma}_n'(x', 1)'|}{\sigma} \ge d_\varepsilon\right).$$

Then, to prove the lemma, it suffices to show that

$$\liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \boldsymbol{\gamma}_n'(x', 1)'\right| \ge \frac{d_\varepsilon \sigma}{\lambda_n}\right) \ge 1 - c_0.$$

Let $(x_n, y_n) \sim G_n$ and $(x_0, y_0) \sim G_0$. Since $\lambda_n \to \infty$, we have $y_n/\lambda_n \to_p 0$. Then the convergence of γ_n to γ guarantees that

$$\frac{y_n}{\lambda_n} - \boldsymbol{\gamma}_n'(x_n', 1)' \to_d \gamma'(x_0', 1)'.$$

For any $\alpha > 0$ which is a point of continuity of the distribution of $|\gamma'(x_0, 1)|$, $\lambda_n \to \infty$ implies

$$\begin{aligned}\liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \boldsymbol{\gamma}_n'(x', 1)'\right| > \frac{d_\varepsilon \sigma}{\lambda_n}\right) &\ge \liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \boldsymbol{\gamma}_n'(x', 1)'\right| > \alpha\right) \\ &= \mathrm{P}_{G_0}\big(|\gamma'(x', 1)'| > \alpha\big).\end{aligned}$$

Letting $\alpha \to 0$ and recalling (4.1), we get

$$\liminf_{n\to\infty} \mathrm{P}_{G_n}\left(\left|\frac{y}{\lambda_n} - \boldsymbol{\gamma}_n'(x',1)'\right| > \frac{d_\varepsilon \sigma}{\lambda_n}\right) \geq 1 - c_0. \qquad \square$$

The proof of the following lemma is similar to that of Lemma 6.

Lemma 7. *Let $\{\xi_n\}$ be a sequence in $B \times R$, with B compact. Assume that when $n \to \infty$, $G_n \to_w G_0$, $\|\xi_n\| \to \infty$ and $\{\sigma_n\}$ is bounded. Then*

$$\liminf_{n\to\infty} \mathrm{E}_{G_n}\rho\left(\frac{y - g(x,\xi_n)}{\sigma_n}\right) = 1. \tag{7.10}$$

Finally, the following result we be used.

Lemma 8. *Let $S(G)$ be defined by (2.8) and suppose that $S(G_0) > 0$. Then, $G_n \to_w G_0$ implies that there exists n_0 such that $S(G_n) > 0$ for $n \geq n_0$.*

Proof. Suppose that the lemma is not true. Then there exists a subsequence $\{G_{n_k}\}_{k\geq 1}$ such that $S(G_{n_k}) = 0$ for all k. This means that giving $\varepsilon > 0$, there exists $(\beta_{n_k}, \alpha_{n_k})$ such that

$$\mathrm{E}_{G_{n_k}}\rho_0\left(\frac{y - g(\mathbf{x}, \beta_{n_k}) - \alpha_{n_k}}{\varepsilon}\right) < \delta \qquad \text{for any } s > 0.$$

The same arguments that we use to prove Lemma 6 let us show that $\{(\beta_{n_k}, \alpha_{n_k})\}$ is bounded and therefore (passing on to a subsequence if necessary) we can assume that $(\beta_{n_k}, \alpha_{n_k}) \to (\tilde{\beta}, \tilde{\alpha})$. Then, from Lemma 4 we get that

$$\mathrm{E}_{G_0}\rho_0\left(\frac{y - g(\mathbf{x}, \tilde{\beta}) - \tilde{\alpha}}{\varepsilon}\right) \leq \delta \qquad \text{for any } s > 0.$$

Then, $S(G_0) \leq S^*(G_0, \tilde{\beta}, \tilde{\alpha}) \leq \varepsilon$. Since this holds for any $\varepsilon > 0$, we get that $S(G_0) = 0$. This contradicts the assumption that $S(G_0) > 0$. $\square$

7.2.1. *Proof of Theorem* 2

Let $G_n \to_w G_0$. Since $\widetilde{S}$ is weakly continuous at G_0, it follows that $\widetilde{S}(G_n) \to \widetilde{S}(G_0) > 0$, by hypothesis.

Case (a): We prove first that $\{T_{\mathrm{M}}(G_n)\}$ is bounded. Suppose that it is not true; then without loss of generality we may assume that $\|T_{\mathrm{M}}(G_n)\| \to \infty$. Then Lemma 7 implies

$$1 = \liminf_{n\to\infty} M_{G_n}(T_{\mathrm{M}}(G_n)) \leq \liminf_{n\to\infty} M_{G_n}(T_{\mathrm{M}}(G_0)) = M_{G_0}(T_{\mathrm{M}}(G_0)),$$

and this implies that $M_{G_0}(\xi) = 1$ for all ξ. This contradicts the assumption that $T_{\mathrm{M}}(G_0)$ is univocally defined. Then, $\{T_{\mathrm{M}}(G_n)\}$ is bounded and from Lemma 5, we get that $T_{\mathrm{M}}(G_n) \to T_{\mathrm{M}}(G_0)$.

Case (b): Recall the notation in (7.7). Convergence of $\{\sigma_n\}$ guarantees that it is a bounded sequence. Suppose that $\{\xi_n\}$ is unbounded. Then, passing on to a subsequence if necessary, we may assume that $\|\xi_n\| \to \infty$. In this case by Lemma 6 we have

$$\liminf_{n\to\infty} M_{G_n}(\xi_n) = \liminf_{n\to\infty} \mathrm{E}_{G_n}\rho\left(\frac{y-\xi_n'(x',1)'}{\sigma_n}\right) \geq 1-c_0. \tag{7.11}$$

We also have

$$\lim_{n\to\infty} M_{G_n}(\xi_0) = \lim_{n\to\infty} \mathrm{E}_{G_n}\rho\left(\frac{y-\xi_0'(x',1)'}{\sigma_n}\right) = M_{G_0}(\xi_0) < 1-c_0. \tag{7.12}$$

Inequalities (7.11) and (7.12) imply that there exists n_0 such that for $n \geq n_0$

$$M_{G_n}(\xi_n) > M_{G_n}(\xi_0),$$

contradicting the definition of $T_{\mathrm{M}}(G_n)$. Therefore, $\{\xi_n\}$ is bounded, and then the weak continuity of T_{M} follows from Lemma 5.

7.2.2. *Proof of Theorem* 3

Let $G_n \to_w G_0$, $\xi_n = T_{\mathrm{S}}(G_n)$, $\xi_0 = T_{\mathrm{S}}(G_0)$, $\sigma_n = S(G_n)$ and $\sigma_0 = S(G_0)$. We prove first that $\{\sigma_n\}$ is bounded. Take any $\sigma_1 > \sigma_0$; then by Lemma 4

$$\mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(x,\xi_0)}{\sigma_1}\right) \to \mathrm{E}_{G_0}\rho_0\left(\frac{y-\underline{g}(x,\xi_0)}{\sigma_1}\right) < \delta,$$

and therefore there exists n_0 such that

$$S^*(\xi_0, G_n) < \sigma_1 \qquad \text{for } n \geq n_0, \tag{7.13}$$

which implies that $S^*(G_n, \xi_0)$ is bounded and therefore $\sigma_n \leq S^*(\xi_0, G_n)$ is also bounded. On the other hand, by Lemma 8, we get that $\sigma_n > 0$ for n large enough.

We now prove that $\{\xi_n\}$ is bounded. In case (a), if $\{\xi_n\}$ is unbounded, Lemma 7 implies

$$\liminf_{n\to\infty} \mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) \geq 1, \tag{7.14}$$

and this contradicts the fact that for all n

$$\mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(x,\xi_n)}{\sigma_n}\right) = \delta < 1.$$

Consider now case (b) and assume that $\{\xi_n\}$ is unbounded. Then, passing on to a subsequence if necessary, we may assume that $\|\xi_n\| \to \infty$. Then by Lemma 6

$$\liminf_{n\to\infty} \mathrm{E}_{G_n}\rho_0\left(\frac{y-\xi_n'(x',1)'}{\sigma_n}\right) \geq 1-c_0,$$

and this contradicts the fact that for all n

$$\mathrm{E}_{G_n}\rho_0\left(\frac{y-\xi_n'(x',1)'}{\sigma_n}\right)=\delta<1-c_0.$$

Then in case (b) $\{\xi_n\}$ is also bounded.

We now show that $\sigma_n \to \sigma_0$. Suppose that this is not true. By passing on to a subsequence if necessary, we may assume that $\sigma_n \to \sigma^* \neq \sigma_0$ and $\xi_n \to \xi^*$ for some ξ^* and σ^*. Since (7.13) holds for any $\sigma' > \sigma_0$ we have $\sigma^* \leq \sigma_0$ and therefore $\sigma^* < \sigma_0$. Then Lemma 4 implies

$$\delta=\lim_{n\to\infty}\mathrm{E}_{G_n}\rho_0\left(\frac{y-\underline{g}(\xi_n,x)}{\sigma_n}\right)=\mathrm{E}_{G_0}\rho_0\left(\frac{y-\underline{g}(\xi^*,x)}{\sigma^*}\right),$$

and therefore $S(G_0) \leq S^*(G_0,\xi^*) = \sigma^* < \sigma_0$. This contradicts the fact that $S(G_0) = \sigma_0$ and shows that S is weakly continuous.

Finally, the weak continuity of T_{S} follows from (2.11) and Theorem 2.

7.3. Proofs of Theorems 4 and 5

7.3.1. *Proof of Theorem* 4

Since

$$\mathrm{E}_{G_n}\Psi(z,T(G_n))=0,$$

the Mean Value theorem together with Condition 6 and the consistency of $T(G_n)$ yield

$$\mathrm{E}_{G_n}\Psi(z,T(G_0))+D(G_n,\theta_n^*)\big(T(G_n)-T(G_0)\big)=0,$$

where $\theta_n^* \to \theta_0$. Then, (5.10) implies that $D(G_n,\theta_n^*) \to D_0$ and, since for large n, $D(G_n,\theta_n^*)$ is nonsingular, we may write

$$\begin{aligned}T(G_n)-T(G_0)&=-D(G_n,\theta_n^*)^{-1}\mathrm{E}_{G_n}\Psi(z,T(G_0))\\&=\mathrm{E}_{G_n}I_{T,G_0}(z)+\big(D_0^{-1}-D(G_n,\theta_n^*)^{-1}\big)\mathrm{E}_{G_n}I_{T,G_0}(z).\end{aligned}$$

Condition 5 implies that the second term of the right–hand side divided by $\|\mathrm{E}_{G_n}I_{T,G_0}(z)\|$ tends to zero, and this proves the theorem.

7.3.2. *Proof of Theorem* 5

Under the assumptions of this theorem, we can prove that Condition 5 holds a.s. using the same arguments as in Lemma 4.2 of Yohai [27]. The only change is to replace the Law of Large Numbers for i.i.d. random variables by the assumption that $\mathrm{E}_{G_n}d(z) \to \mathrm{E}_{G_0}d(z)$ a.s. for all d such that $\mathrm{E}_{G_0}|d(z)| < \infty$ in the case (a) and for the fact that $\mathrm{E}_{G_n}d(z) \to \mathrm{E}_{G_0}d(z)$ for all function d bounded and continuous in case (b). Then, Theorem 4 implies that T is weakly differentiable at $\{G_n\}$.

7.4. Derivations of influence functions

7.4.1. *Derivation of* (6.11)–(6.12)

Put for brevity

$$t_{\mathrm{MM}} = \frac{y - \underline{g}(x, \xi_{\mathrm{MM}})}{\sigma}, \qquad t_{\mathrm{S}} = \frac{y - \underline{g}(x, \xi_{\mathrm{S}})}{\sigma}.$$

Then

$$\dot{\Psi}(z,\theta) = \begin{bmatrix} \dot{\Psi}_{11}(z,\theta) & 0 & \dot{\Psi}_{13}(z,\theta) \\ 0 & \dot{\Psi}_{22}(z,\theta) & \dot{\Psi}_{23}(z,\theta) \\ \dot{\Psi}_{31}(z,\theta) & 0 & \dot{\Psi}_{33}(z,\theta) \end{bmatrix},$$

where

$$\begin{aligned}
\dot{\Psi}_{11}(z,\theta) &= -\frac{1}{\sigma}\psi_0'(t_{\mathrm{S}})\underline{\dot{g}}(x,\xi_{\mathrm{S}})\underline{\dot{g}}(x,\xi_{\mathrm{S}})' + \psi_0(t_{\mathrm{S}})\underline{\ddot{g}}(x,\xi_{\mathrm{S}}), \\
\dot{\Psi}_{13}(z,\theta) &= -\frac{1}{\sigma}\psi_0'(t_{\mathrm{S}})t_{\mathrm{S}}\underline{\dot{g}}(x,\xi_{\mathrm{S}}), \\
\dot{\Psi}_{22}(z,\theta) &= -\frac{1}{\sigma}\psi_1'(t_{\mathrm{MM}})\underline{\dot{g}}(x,\xi_{\mathrm{MM}})\underline{\dot{g}}(x,\xi_{\mathrm{MM}})' + \psi_1(t_{\mathrm{MM}})\underline{\ddot{g}}(x,\xi_{\mathrm{MM}}), \\
\dot{\Psi}_{23}(z,\theta) &= -\frac{1}{\sigma}\psi_1'(t_{\mathrm{MM}})t_{\mathrm{MM}}\underline{\dot{g}}(x,\xi_{\mathrm{MM}}), \\
\dot{\Psi}_{31}(z,\theta) &= -\frac{1}{\sigma}\psi_0(t_{\mathrm{S}})\underline{\dot{g}}(x,\xi_{\mathrm{S}}), \\
\dot{\Psi}_{33}(z,\theta) &= -\frac{1}{\sigma}\psi_0(t_{\mathrm{S}})t_{\mathrm{S}}.
\end{aligned} \tag{7.15}$$

From (7.15) it is easy to show that

$$D_0 = \mathrm{E}_{G_0}\dot{\Psi}(z,\theta_0) = -\frac{1}{\sigma_0}\begin{bmatrix} a_{00}C_0 & 0 & e_{00}b_0^* \\ 0 & a_{01}C_0 & e_{01}b_0^* \\ 0 & 0 & d_0 \end{bmatrix}.$$

Therefore, $|D_0| = a_{00}a_{01}d_0|C_0|^2$. It follows from (6.10) $|C_0| \neq 0$ if and on only if $|A_0| \neq 0$, and that

$$C_0^{-1} = \begin{bmatrix} A_0^{-1} & -A_0^{-1}b_0 \\ -(A_0^{-1}b_0)' & 1 + b_0'A_0^{-1}b_0 \end{bmatrix},$$

Direct calculation shows that

$$D_0^{-1} = -\sigma_0\begin{bmatrix} a_{00}^{-1}C_0^{-1} & 0 & -e_{00}a_{00}^{-1}d_0^{-1}C_0^{-1}b_0^* \\ 0 & a_{01}^{-1}C_0^{-1} & -e_{01}a_{01}^{-1}d_0^{-1}C_0^{-1}b_0^* \\ 0 & 0 & d_0^{-1} \end{bmatrix},$$

and the desired results follow from (5.1).

7.4.2. *Derivation of* (6.22)

In this case from (7.15), it is easy to show that

$$D_0 = -\frac{1}{\sigma_0}\begin{bmatrix} a_{00} & 0 & e_{00} \\ 0 & a_{01} & e_{01} \\ 0 & 0 & d_0 \end{bmatrix},$$

which implies

$$D_0^{-1} = -\sigma_0\begin{bmatrix} a_{00}^{-1} & 0 & -e_{00}a_{00}^{-1}d_0^{-1} \\ 0 & a_{01}^{-1} & -e_{01}a_{01}^{-1}d_0^{-1} \\ 0 & 0 & d_0^{-1} \end{bmatrix}.$$

The rest of the derivation is straightforward.

7.5. Proof of Theorems 6 and 7

7.5.1. *Proof of Theorem* 6

Parts (i) and (ii) follow from Theorem 1 and Remark 2. To prove (iii), we need to check conditions of Theorem 2 and Theorem 3. We start showing that $S(G_0) > 0$. Let

$$h_{\beta,\alpha}(s) = \mathrm{E}\rho_0\left(\frac{y_i - g(x_i,\beta) - \alpha}{s}\right).$$

Then, we have

$$\lim_{s\to\infty} h_{\beta,\alpha}(s) = \rho_0(0) = 0 \tag{7.16}$$

and

$$\lim_{s\to 0} h_{\beta,\alpha}(s) = 1 - \mathrm{P}\big(y_i = g(x_i,\beta) + \alpha\big). \tag{7.17}$$

Since u_i has a continuous distribution and is independent of x_i, we also have

$$\begin{aligned} \mathrm{P}\big(y_i = g(x_i,\beta) + \alpha\big) &= \mathrm{P}\big(g(x_i,\beta_0) + u_i = g(x_i,\beta) + \alpha\big) \\ &= \mathrm{E}\big[\mathrm{P}\big(u_i = g(x_i,\beta) - g(x_i,\beta_0) + \alpha\big)\big] = 0. \end{aligned} \tag{7.18}$$

Equations (7.16), (7.17) and (7.19) imply that $S^*(G_0,\beta,\alpha) > 0$ for all (β,α), and so $S(G_0) = S^*(G_0,\beta_0,\alpha_{01}) > 0$.

Note that

$$\begin{aligned} M_{G_0}(T_{\mathrm{MM}}(G_0)) &= \mathrm{E}\left(\rho_1\left(\frac{y - T_{\mathrm{MM}}(G_0)}{S(G_0)}\right)\right) \\ &\le \mathrm{E}\left(\rho_1\left(\frac{y - T_{\mathrm{S}}(G_0)}{S(G_0)}\right)\right) \le \mathrm{E}\left(\rho_0\left(\frac{y - T_{\mathrm{S}}(G_0)}{S(G_0)}\right)\right) = \delta. \end{aligned}$$

Then $\delta < 1 - C(G_0)$ implies (4.2) and from Theorem 3 follows that T_{S} and S are weakly continuous. Since S is weekly continuous, Theorem 2 implies that T_{MM} is weakly continuous too, and so part (iii) follows.

Part (iv) follows from the formulas obtained in Section 7.4.

(v) follows from part (a) of Theorem 5 while part (vi) follows from Lemma 1. Part (vii) follows from (vi) as was already shown before stating the theorem.

To prove (viii) is enough to show that

$$n^{1/2}\mathrm{E}_{G_n} I_{T_{\mathrm{MM},\beta},G_0}(x, y) \to_d N(0, V), \tag{7.19}$$

where V is given by (6.19). From (6.11), is immediate that for all $\boldsymbol{\lambda} \in R^q$, $\lambda' I_{T_{\mathrm{MM},\beta},G_0}(x_i, y_i)$ is a stationary ergodic martingale difference. Then (7.19) follows from the central limit theorem for martingale differences (see, e.g., Theorem 23.1 of Billingsley [2]) and the Cramer–Wald device.

Part (ix) will follow from

$$n^{1/2}\mathrm{E}_{G_n} I_{T_{\mathrm{MM}},G_0}(x, y) \to_d N(0, V), \tag{7.20}$$

where V is given by (6.21). According to (6.13), we have that

$$I_{T_{\mathrm{MM}},G_0}(x_i, y_i) = \frac{\sigma_0}{\mathrm{E}_{F_0}\psi_1'((u-\alpha_0)/\sigma_0)}\psi_1\left(\frac{u_i-\alpha_0}{\sigma_0}\right)C_0^{-1}\underline{\dot{g}}(x, \beta_0),$$

and therefore for all $\lambda \in R^{p+1}$, $\lambda' I_{T_{\mathrm{MM}},G_0}(x_i, y_i)$ is a ϕ-mixing process with mean 0 satisfying $\sum_{i=1}^{\infty}\phi_n^{1/2} < \infty$. Then by Theorem 20.1 of Billingsley [2], we have that $n^{1/2}\lambda'\mathrm{E}_{G_n} I_{T_{\mathrm{MM}},G_0}(x, y) \to_d N(0, \lambda' V \lambda)$, where

$$V = \sum_{i=-\infty}^{\infty} \mathrm{E}[I_{T_{\mathrm{MM}},G_0}(x_1, y_1) I'_{T_{\mathrm{MM}},G_0}(x_{1+i}, y_{1+i})].$$

Finally, the proof is completed noting that

$$\mathrm{E}[I_{T_{\mathrm{MM}},G_0}(x_1, y_1) I'_{T_{\mathrm{MM}},G_0}(x_{1+i}, y_{1+i})] = \frac{\sigma_0^2 c_i}{\mathrm{E}_{F_0}^2\psi_1'((u-\alpha_0)/\sigma_0)} C_0^{-1} C_i C_0^{-1}$$

and using the Cramer–Wald device.

7.5.2. *Proof of Theorem 7*

It is completely similar to the proof of Theorem 6. The only differences are that for part (iii) we use that in the case of a location model we have $c(G_0) = 0$, and therefore condition (4.2) reduces to $M_{G_0}(T_{\mathrm{M}}(G_0)) < 1$. Note that this inequality is implied by the condition that $T_{\mathrm{M}}(G_0)$ is well defined. So, for this case, (4.2) always holds, and that for part (iv) we use part (b) of Theorem 5 instead of part (a).

Acknowledgements

This research was partially supported by Grants X–018 from University of Buenos Aires, PIP 112-200801-00216 and 0592 from CONICET and PICT 00899 and 0083 from ANPCyT. We thank the Referees and Associate Editor for their comments and suggestions which led to a much improved version of the paper.

References

[1] Bianco, A.M. and Yohai, V.J. (1996). Robust estimation in the logistic regression model. In *Robust Statistics, Data Analysis, and Computer Intensive Methods* (*Schloss Thurnau*, 1994) (H. Rieder, ed.). *Lecture Notes in Statist*. **109** 17–34. New York: Springer. MR1491394

[2] Billingsley, P. (1968). *Convergence of Probability Measures*. New York: Wiley. MR0233396

[3] Boos, D.D. and Serfling, R.J. (1980). A note on differentials and the CLT and LIL for statistical functions, with application to M-estimates. *Ann. Statist*. **8** 618–624. MR0568724

[4] Čížek, P. (2006). Least trimmed squares in nonlinear regression under dependence. *J. Statist. Plann. Inference* **136** 3967–3988. MR2299174

[5] Clarke, B.R. (1983). Uniqueness and Fréchet differentiability of functional solutions to maximum likelihood type equations. *Ann. Statist*. **11** 1196–1205. MR0720264

[6] Clarke, B.R. (2000). A remark on robustness and weak continuity of M-estimators. *J. Austral. Math. Soc. Ser. A* **68** 411–418. MR1753369

[7] Croux, C., Dhaene, G. and Hoorelbeke, D. (2003). Robust standard errors for robust estimators. Discussions Paper Series (DPS) 03.16, Center for Economic Studies, Katholieke Universiteit Leuven. Available at http://www.econ.kuleuven.be/ew/academic/econmetr/members/dhaene/papers/rsejan2004.pdf.

[8] Croux, C. and Haesbroeck, G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Comput. Statist. Data Anal*. **44** 273–295. MR2020151

[9] Doukhan, P. (1994). *Mixing: Properties and Examples*. *Lecture Notes in Statistics* **85**. New York: Springer. MR1312160

[10] Fasano, M.V. (2009). Robust estimation in nonlinear regression. Ph.D. thesis, Univ. La Plata. Available at http://www.mate.unlp.edu.ar/tesis/tesis_fasano_v.pdf.

[11] Fernholz, L.T. (1983). *Von Mises Calculus for Statistical Functionals*. *Lecture Notes in Statistics* **19**. New York: Springer. MR0713611

[12] Fraiman, R. (1983). General M-estimators and applications to bounded influence estimation for non-linear regression. *Comm. Statist. Theory Methods* **12** 2617–2631. MR0715170

[13] Hampel, F.R. (1971). A general qualitative definition of robustness. *Ann. Math. Statist*. **42** 1887–1896. MR0301858

[14] Hampel, F.R. (1974). The influence curve and its role in robust estimation. *J. Amer. Statist. Assoc*. **69** 383–393. MR0362657

[15] Ibragimov, I.A. (1956). On the composition of unimodal distributions. *Theory Probab. Appl*. **1** 255–260.

[16] Kudraszow, N.L. and Maronna, R.A. (2011). Estimates of MM type for the multivariate linear model. *J. Multivariate Anal.* **102** 1280–1292.

[17] Liese, F. and Vajda, I. (2003). A general asymptotic theory of M-estimators. I. *Math. Methods Statist*. **12** 454–477. MR2054158

[18] Liese, F. and Vajda, I. (2004). A general asymptotic theory of M-estimators. II. *Math. Methods Statist*. **13** 82–95. MR2078314

[19] Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics*: *Theory and Methods*. *Wiley Series in Probability and Statistics*. Chichester: Wiley. MR2238141

[20] Mizera, I. (1994). On consistent M-estimators: Tuning constants, unimodality and breakdown. *Kybernetika* (*Prague*) **30** 289–300. MR1291931

[21] Papantoni-Kazakos, P. and Gray, R.M. (1979). Robustness of estimators on stationary observations. *Ann*. *Probab*. **7** 989–1002. MR0548893

[22] Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis* (*Heidelberg*, 1983) (J. Franke, W. Härdle and R.D. Martin, eds.). *Lecture Notes in Statist*. **26** 256–272. New York: Springer. MR0786313

[23] Sakata, S. and White, H. (2001). S-estimation of nonlinear regression models with dependent and heterogeneous observations. *J*. *Econometrics* **103** 5–72. Studies in estimation and testing. MR1838195

[24] Stromberg, A.J. (1995). Consistency of the least median of squares estimator in nonlinear regression. *Comm*. *Statist*. *Theory Methods* **24** 1971–1984. MR1345230

[25] Sued, M. and Yohai, V.J. (2010). Robust location estimates with missing data. Available at ArXiv:1004.5418v2 [math.ST].

[26] Vaĭner, B.P. and Kukush, O.G. (1998). The consistency of M-estimators constructed from a concave weight function. *Theory Probab*. *Math*. *Stat*. **57** 11–18.

[27] Yohai, V.J. (1985). High breakdown point and high efficiency robust estimates for regression. Technical report 66, Dept. Statistics, Univ. Washington. Available at http://www.stat.washington.edu/research/reports/1985/tr066.pdf.

[28] Yohai, V.J. (1987). High breakdown-point and high efficiency robust estimates for regression. *Ann*. *Statist*. **15** 642–656. MR0888431

Received June 2010 and revised February 2011