PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 14, Fasc. 2 (1993), pp. 189–206

TEST FOR DIFFERENCES BETWEEN M-ESTIMATES OF NON-LINEAR REGRESSION MODEL

BY

A. M. RUBIO (CACERES), F. QUINTANA (BADAJOZ) AND J. Á. VÍŠEK* (PRAGUE)

Abstract. For two M-estimates of the regression model (evaluated for the same data) a test of difference between them is proposed. An asymptotic representation of $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ was used as a key tool for the construction of the test. If the difference is classified as significant, it indicates that something does not correspond with the framework under which the consistency was derived. As the conditions for the consistency and the asymptotic normality of the estimates have a statistical character, it may mean that for the samples of finite sizes one ϱ-function at least was not appropriate for the given data, so that the asymptotics does not yet work. This implies that at least one of our estimates may be rather far from the "true" model.

1. Introduction. Even restricting ourselves to such ϱ-functions which generate the M-estimators with the good properties we may evaluate nearly an unlimited amount of the various estimates of the regression models for the same data. Then we (frequently) find ourselves in a situation in which the estimates of the model so evidently differ that the question of significance of the differences requires to be answered. (For numerical examples of a situation in which we obtain rather different models for real data see [13] or [14].) The question was studied for the linear models by Rubio et al. [12] and the present paper brings a very first attempt to find some results for the non-linear regression. An asymptotic representation of the estimators of the regression coefficients [7] has appeared to be a very powerful tool to study the problem for the linear regression [12]. However, for the non-linear regression no asymptotic representation has been known. The first attempt of deriving such a representation may be traced out in [10], although it was not isolated there and explicitly given. Moreover, the conditions required there (existence of the continuous second derivative of the ϱ-function) do not cover the ϱ-functions

* This paper was written while the author was visiting University of Extremadura.

frequently used. Although small modifications of these functions (in order to achieve continuity of their second derivative) would have presumably a negligible influence on the resulting values of the estimators, it may complicate their evaluation, not being a very simple task, anyway. Surely, the increase of the complexity of evaluation of the M-estimators would not be crucial, however, if we did without it, it would be preferable. Moreover, such modifications break the admissibility of estimators [5] (it is, of course, more or less an academic question). Nevertheless, it may be of interest to establish the asymptotic representation of estimators of the non-linear regression model under similarly wide conditions as in [7]; compare also with the set of conditions given for the study of the change-of-variance function in [5].

So the first task was to verify that the technique which was developed for the linear models (see, e.g., [7]) also works for non-linear ones. Since we have assumed the existence of corresponding derivatives of the non-linear surface, we may approximate this non-linear surface, in a neighbourhood of the "true" values of regression coefficients, by a plane, and hence the problem seems to be easy solvable. On the other hand, as this is a first attempt to generalize this technique to the non-linear framework, some care has been inevitable.

Moreover, there exist other problems in which the asymptotic representation of the estimators of the regression coefficients helped to find some results (see [14]), the representation for the non-linear case may be itself of interest.

Finally, since we have focussed on the problem of testing the differences between the models, we have omitted the question of the consistency of M-estimators and we have assumed simply that they are consistent. For the consistency problem consult, e.g., [10]. We have referred earlier on this paper having noted that there is an implaussible assumption of the existence of the second derivative of ϱ-function there. However, the authors needed this assumption only to prove the asymptotic normality. The consistency was established under weaker conditions which apply to the frequently employed ϱ-functions. The present paper shows that it is possible to derive an asymptotic representation (which implies here also the asymptotic normality) also under the weaker conditions.

2. Notation and conditions. Let N denote the set of all positive integers, R the real line, R^l the l-dimensional Euclidean space ($l \in N$), and $(\Omega, \mathscr{A}, P)$ a probability space. Let, moreover, for some fixed $p \in N$ and $q \in N$, $\beta^0 = (\beta_1^0, \beta_2^0, \ldots, \beta_p^0)^T$ be the vector of regression coefficients (the upper index T means transposition) and $\{X_i\}_{i=1}^{\infty}$, $X_i\colon \Omega \to R^q$, be a sequence of independent and identically distributed random variables (i.i.d.r.v.). Finally, let $\{e_i\}_{i=1}^{\infty}$, $e_i\colon \Omega \to R$, be another sequence of i.i.d.r.v., independent of the sequence $\{X_i\}_{i=1}^{\infty}$. For a function $g\colon R^{q+p} \to R$ we shall consider (for all $i \in N$) a regression model

$$Y_i = g(X_i, \beta^0) + e_i. \tag{1}$$

Let us denote by $K(x)$ the distribution function of X_1, and by $F(z)$ the distribution function of e_1, so that the joint distribution is $G(x, z) = K(x) \cdot F(z)$. The density of the distribution $F(z)$, whenever we shall assume its existence, will be denoted by $f(z)$; moreover, let S_1 denote the support of $K(x)$. We will be interested in the M-estimator of β^0 given as

$$\hat{\beta}^{(n)} = \arg\min_{\beta \in R^p} \left\{ \sum_{i=1}^{n} \varrho\left(Y_i - g(X_i, \beta)\right) \right\}, \tag{2}$$

where $\varrho\colon R \to R$ with properties specified below. Let us denote the first and the second derivatives of ϱ (at those points where they exist) by ψ and ψ', respectively. For any finite set $S = \{s_1, s_2, \ldots, s_k\} \subset R$ and positive α put

$$S(\alpha) = \bigcup_{i=1}^{k} [s_i - \alpha, s_i + \alpha].$$

The following conditions will be considered later:

CONDITIONS A. There are two sets

$$D_1 = \{d_{11}, d_{12}, \ldots, d_{1s_1}\} \quad \text{and} \quad D_2 = \{d_{21}, d_{22}, \ldots, d_{2s_2}\},$$

with s_1 and s_2 finite, such that:

(i) The derivative $\psi'(z)$ exists and is uniformly continuous on any interval (a, b) such that $(a, b) \cap \{D_1 \cup D_2\} = \emptyset$.

(ii) There is τ_0 such that $F(z)$ has a continuous density $f(z)$ on $D_1(\tau_0) \cup D_2(\tau_0)$ and $\psi(z)$ is absolutely continuous on any interval (a, b) such that $(a, b) \cap D_2 = \emptyset$.

(iii) At each d_{2i} $(i = 1, 2, \ldots, s_2)$ the limits

$$\lim_{z \searrow d_{2i}} \psi(z) = \psi(d_{2i}+) \quad \text{and} \quad \lim_{z \nearrow d_{2i}} \psi(z) = \psi(d_{2i}-)$$

exist and $\psi(d_{2i}+) \neq \psi(d_{2i}-)$. Moreover, $|\psi(d_{2i}+)| \neq \infty$ as well as $|\psi(d_{2i}-)| \neq \infty$.

(iv) Let $\mathbf{E}_F \psi(e_1) = 0$ and

$$\gamma = \mathbf{E}_F \psi'(e_1) + \sum_{k=1}^{s_1} f(d_{2k}) [\psi(d_{2k}+) - \psi(d_{2k}-)] > 0.$$

Remark 1. Observe that, due to the continuity of $f(x)$ on $\{D_1(\tau_0) \cup D_2(\tau_0)\}$, $f(x)$ is bounded there, let us say by $M < \infty$. Notice also that Condition A (i) implies that there is a finite L such that

$$\sup_z |\psi(z)| < L \quad \text{and} \quad \sup_z |\psi'(z)| < L, \quad \text{where } z \in \{D_1(\tau_0) \cup D_2(\tau_0)\} \setminus \{D_1 \cup D_2\}.$$

Remark 2. Conditions A essentially coincide with those of Hampel et al. [5] which have been used to study the change-of-variance function. The reader who is interested in a heuristic discussion of these conditions may find it in this book. The conditions cover presumably the all ϱ-functions frequently used in the present robust statistics. The corresponding ψ-functions may be written as linear combinations of three functions

$$\psi = \psi_a + \psi_c + \psi_s,$$

where ψ_a is an absolutely continuous function with uniformly continuous derivative, ψ_c is a continuous function with the derivative which is a step-function with a finite number of jumps, and ψ_s is a step-function with a finite number of jumps.

CONDITIONS B. (i) The function g is in a neighbourhood of β^0 twice differentiable in coordinates corresponding to regression coefficients, i.e., there is $\delta_0 > 0$ such that, for any $\beta \in R^p$, $\|\beta - \beta^0\| \leqslant \delta_0$, the derivatives

$$\frac{\partial}{\partial \beta_j} g(x, \beta) \ (j = 1, 2, \ldots, p) \quad \text{and} \quad \frac{\partial^2}{\partial \beta_j \partial \beta_k} g(x, \beta) \ (j, k = 1, 2, \ldots, p)$$

exist for any $x \in S_1$. Let us denote the corresponding vector and the matrix simply by $g'(x, \beta)$ and $g''(x, \beta)$, respectively, and their coordinates and elements by $g'_j(x, \beta)$ and $g''_{jk}(x, \beta)$.

(ii) There is $J < \infty$ such that

$$\max_{1 \leqslant j \leqslant p} \sup_{x \in S_1, \beta \in R^p, \|\beta - \beta^0\| < \delta_0} |g'_j(x, \beta)| < J,$$

$$\max_{1 \leqslant j,k \leqslant p} \sup_{x \in S_1, \beta \in R^p, \|\beta - \beta^0\| < \delta_0} |g''_{jk}(x, \beta)| < J.$$

(iii) The matrix $Q = \mathbf{E}_K \{g'(x, \beta^0)[g'(x, \beta^0)]^T\}$ is regular (and hence positive definite).

CONDITION C. The estimator $\hat{\beta}^{(n)}$ is consistent.

3. Preliminaries.

LEMMA 1. *Let Conditions* A *and* B *be fulfilled. Then for any* $\tau \in (0, \frac{1}{2}]$ *and any* $C > 0$ *we have for* $j = 1, 2, \ldots, p$

$$\sup_{\|t\| < C} n^{-1+\tau} \Big| \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \beta^0 + n^{-\tau} t)) g'_j(X_i, \beta^0 + n^{-\tau} t)$$

$$- \psi(Y_i - g(X_i, \beta^0)) g'_j(X_i, \beta^0) + n^{-\tau} \gamma g'_j(X_i, \beta^0) [g'(X_i, \beta^0)]^T t] \Big| \to 0$$

in probability as $n \to \infty$, *where* $\|t\|$ *denotes the Euclidean norm of* $t \in R^p$.

Proof. Fix $\varepsilon > 0$, $\tau \in (0, \frac{1}{2}]$ and $C > 0$ and define for any $t \in R^p$ and $\omega \in \Omega$

$$\xi_i(t, \omega) = \min \{e_i, e_i + g(X_i, \beta^0) - g(X_i, \beta^0 + n^{-\tau} t)\},$$

$$\zeta_i(t, \omega) = \max \{e_i, e_i + g(X_i, \beta^0) - g(X_i, \beta^0 + n^{-\tau} t)\},$$

and also for $j = 1$ and 2

$$\mathscr{H}_{n,j}(t, \omega) = \{i \in N: (\xi_i(t, \omega), \zeta_i(t, \omega)) \cap D_j \neq \emptyset\},$$

and finally

$$\mathscr{H}_{n,3}(t, \omega) = \{1, 2, \ldots, n\} \backslash \{\mathscr{H}_{n,1}(t, \omega) \cup \mathscr{H}_{n,2}(t, \omega)\}.$$

In the sequel we shall omit t and ω in $\mathscr{H}_{n,j}(t, \omega)$ provided it cannot cause a misunderstanding. Let

$$\alpha_0 = \min_{u \neq v, u, v \in \{D_1 \cup D_2\}} |u - v|.$$

Due to the assumption of the uniform (with respect to $x \in S_1$) continuity of $g(x, \beta)$ at β^0, we may find $\delta_1 > 0$, $\delta_1 < \delta_0$ (see Condition B (i)) such that for any $\beta \in R^p$, $\|\beta - \beta^0\| < \delta_1$ we have $|g(x, \beta) - g(x, \beta^0)| < \frac{1}{2}\alpha_0$ for any $x \in S_1$. Now one can find $n_1 \in N$ such that $n_1^{-\tau} C < \delta_1$. Then for any $t \in R^p$, $\|t\| < C$ and $n > n_1$ we have $\mathscr{H}_{n,1}(t, \omega) \cap \mathscr{H}_{n,2}(t, \omega) = \emptyset$ for any $\omega \in \Omega$. Then using the mean value theorem we may write for $j = 1, 2, \ldots, p$ the sum

$$n^{-1+\tau} \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \beta^0 + n^{-\tau} t)) g_j'(X_i, \beta^0 + n^{-\tau} t) - \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)]$$

as

$$(3) \qquad n^{-1+\tau} \Big\{ \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \beta^0 + n^{-\tau} t)) g_j'(X_i, \beta^0 + n^{-\tau} t) - \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)] I_{\{i \in \mathscr{H}_{n,1}\}}$$
$$+ \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \beta^0 + n^{-\tau} t)) g_j'(X_i, \beta^0 + n^{-\tau} t) - \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)] I_{\{i \in \mathscr{H}_{n,2}\}}$$
$$- n^{-\tau} \sum_{i=1}^{n} [\psi'(Y_i - g(X_i, \tilde{\beta})) g_j'(X_i, \tilde{\beta}) [g'(X_i, \tilde{\beta})]^T t - \psi(Y_i - g(X_i, \tilde{\beta})) [g''(X_i, \tilde{\beta}) t]_j] I_{\{i \in \mathscr{H}_{n,3}\}} \Big\},$$

where $\|\tilde{\beta} - \beta^0\| < n^{-\tau} \|t\|$. We shall consider the terms of (3) separately. Let us start with the first one. Now let us select $\delta_2 > 0$, $\delta_2 < \delta_1$ so that for any $\beta \in R^p$, $\|\beta - \beta^0\| < \delta_2$ we have, uniformly in $x \in S_1$, $|g(x, \beta) - g(x, \beta^0)| < \tau_0$ and select $n_2 > n_1$ so that $C n_2^{-\tau} < \delta_2$. (In what follows, at any step at which we shall look for some δ_r, we will assume that it will be chosen so that $0 < \delta_r < \delta_{r-1}$; similarly, any next n_r will be always $n_r \in N$ and larger than n_{r-1}.) Let us now put for any $t \in R^p$ and any $k = 1, 2, \ldots, p$

$$z_t^{(k)} = (t_1, t_2, \ldots, t_{k-1}, z, 0, \ldots, 0)^T.$$

Due to the absolute continuity of $\psi(z)$ on $D_1(\tau_0)$ and the existence of $g''(x, \beta)$ (for any $x \in S_1$ and $\beta \in R^p$, $\|\beta - \beta^0\| < \delta_0$) we have for any $i \in \mathscr{H}_{n,1}$

$$|\psi(Y_i - g(X_i, \beta^0 + n^{-\tau} t)) g_j'(X_i, \beta^0 + n^{-\tau} t) - \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)|$$

$$= n^{-\tau} \Big| \sum_{k=1}^{p} \int_0^{t_k} \{\bar{\psi}'(Y_i - g(X_i, \beta^0 + n^{-\tau} z_i^{(k)})) g_j'(X_i, \beta^0 + n^{-\tau} z_i^{(k)}) g_k'(X_i, \beta^0 + n^{-\tau} z_i^{(k)})$$

$$+ \psi(Y_i - g(X_i, \beta^0 + n^{-\tau} z_i^{(k)})) g_{jk}''(X_i, \beta^0 + n^{-\tau} z_i^{(k)})\} \, dz \Big|$$

$$< n^{-\tau} \sum_{k=1}^{p} L \cdot J \cdot [J+1] \, t_k \leqslant n^{-\tau} \cdot p^{1/2} \cdot L \cdot J \cdot [J+1] \, \|t\|,$$

where $\bar{\psi}'$ coincides with ψ' except for the set $D_1 \cup D_2$ and we may define $\bar{\psi}'$ at the points of $D_1 \cup D_2$ as the limit from the left of ψ'. Let us write

$$\delta_{in}(t) = g(X_i, \beta^0 + n^{-\tau} t) - g(X_i, \beta^0).$$

Then $I_{\{i \in \mathscr{H}_{n,1}\}} = 1$ implies

$$e_i \in [d_{1l} - |\delta_{in}(t)|, \; d_{1l} + |\delta_{in}(t)|]$$

(for some $l \in \{1, 2, \ldots, r\}$), and since D_1 has a finite number of points, we have

$$P_F(I_{\{i \in \mathscr{H}_{n,1}\}} = 1) \leqslant \sum_{l=1}^{s_2} \{F(d_{1l} + |\delta_{in}(t)|) - F(d_{1l} - |\delta_{in}(t)|)\}.$$

Let us assume that $\delta_{in}(t) > 0$ (for a negative value we need only to change the bounds of integration into the opposite order). Then

$$|F(d_{1l} + \delta_{in}(t)) - F(d_{1l})|$$

$$\leqslant \Big| \sum_{k=1}^{p} \int_0^{n^{-\tau} t_k} f(d_{1l} + g(X_i, \beta^0 + z_i^{(k)}) - g(X_i, \beta^0)) g_k'(X_i, \beta^0 + z_i^{(k)}) \, dz \Big|$$

(where we integrate in fact along the coordinates). But the last expression may be bounded by

$$n^{-\tau} M \cdot J \sum_{k=1}^{p} t_k \leqslant n^{-\tau} p^{1/2} M \cdot J \, \|t\|.$$

Since $G(x, z) = K(x) \cdot F(z)$, we have $\mathbf{E}_G I_{\{i \in \mathscr{H}_{n,1}\}} \leqslant n^{-\tau} p^{1/2} M \cdot J \, \|t\|$, so that

$$\mathbf{E} \sup_{\|t\| < C} \Big| \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \beta^0 + n^{-\tau} t)) g_j'(X_i, \beta^0 + n^{-\tau} t)$$

$$- \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)] \, I_{\{i \in \mathscr{H}_{n,1}\}} \Big|$$

$$\leqslant \mathbf{E} \sup_{\|t\| < C} n^{-\tau} p^{1/2} L \cdot J \cdot [J+1] \, \|t\| \sum_{i=1}^{n} I_{\{i \in \mathscr{H}_{n,1}\}} \leqslant n^{1-2\tau} p L \cdot J^2 \cdot [J+1] \cdot C^2 \cdot M.$$

Using Chebyshev's inequality for the nonnegative random variable, for any

$\varepsilon > 0$ we may find n_ε so that for any $n > n_\varepsilon$ we have

$$P_G\Big(\sup_{\|t\|<C} n^{-1+\tau}\Big|\sum_{i=1}^{n}\big[\psi\big(Y_i-g(X_i,\beta^0+n^{-\tau}t)\big)g_j'(X_i,\beta^0+n^{-\tau}t)$$
$$-\psi\big(Y_i-g(X_i,\beta^0)\big)g_j'(X_i,\beta^0)\big]I_{\{i\in\mathscr{H}_{n,1}\}}\Big|>\varepsilon\Big)<\varepsilon.$$

Let us turn our attention to the second term in (3). First of all, let us define on $D_2(\tau_0)$ the functions $\tilde\psi$ and $\tilde{\tilde\psi}$ so that

$$\tilde\psi(z)=\begin{cases}\psi(z), & z\in\bigcup_{l=1}^{s_2}[d_l-\tau_0,\,d_l),\\ \psi(d_j-), & z\in[d_j,\,d_j+\tau_0],\quad j=1,2,\ldots,s_2,\end{cases}$$

and

$$\tilde{\tilde\psi}(z)=\begin{cases}\psi(d_j+), & z\in[d_j-\tau_0,\,d_j],\quad j=1,2,\ldots,s_2,\\ \psi(z), & z\in\bigcup_{l=1}^{s_2}(d_l,\,d_l+\tau_0].\end{cases}$$

Moreover, let us put

$$\mathscr{I}_{n,1}=\Big\{i\in\mathscr{H}_{n,2}\colon\ e_i\in\bigcup_{l=1}^{s_2}[d_l-\tau_0,\,d_l]\Big\},\qquad \mathscr{I}_{n,2}=\Big\{i\in\mathscr{H}_{n,2}\colon\ e_i\in\bigcup_{l=1}^{s_2}(d_l,\,d_l+\tau_0]\Big\}.$$

As above let us keep in mind that $I_{\{i\in\mathscr{H}_{n,2}\}}=1$ implies $d_{k(i)}\in(\xi_i(t,\omega),\,\zeta_i(t,\omega))$ for some $k(i)\in\{1,2,\ldots,s\}$. Now the second term in (3) can be written as

$$(4)\qquad n^{-1+\tau}\Big\{\sum_{i\in\mathscr{I}_{n,1}}\big[\tilde{\tilde\psi}\big(Y_i-g(X_i,\beta^0+n^{-\tau}t)\big)g_j'(X_i,\beta^0+n^{-\tau}t)$$
$$-\tilde{\tilde\psi}\big(Y_i-g(X_i,\beta^0)\big)g_j'(X_i,\beta^0)$$
$$+\big[\tilde\psi\big(Y_i-g(X_i,\beta^0+n^{-\tau}t)\big)-\tilde\psi\big(Y_i-g(X_i,\beta^0)\big)\big]g_j'(X_i,\beta^0)\big]$$
$$+\sum_{i\in\mathscr{I}_{n,2}}\big[\tilde\psi\big(Y_i-g(X_i,\beta^0+n^{-\tau}t)\big)g_j'(X_i,\beta^0+n^{-\tau}t)$$
$$-\tilde\psi\big(Y_i-g(X_i,\beta^0)\big)g_j'(X_i,\beta^0)$$
$$+\big[\tilde{\tilde\psi}\big(Y_i-g(X_i,\beta^0+n^{-\tau}t)\big)-\tilde{\tilde\psi}\big(Y_i-g(X_i,\beta^0)\big)\big]g_j'(X_i,\beta^0)\big]$$
$$+\sum_{i\in\mathscr{H}_{n,2}}\operatorname{sign}(\delta_{in}(t))\,g_j'(X_i,\beta^0)\big[\psi(d_{k(i)}-)-\psi(d_{k(i)}+)\big]\Big\}.$$

Since $\tilde\psi$ and $\tilde{\tilde\psi}$ fulfill on $D_2(\tau_0)$ the same conditions as ψ on $D_1(\tau_0)$, the sum over $\{i\in\mathscr{I}_{n,1}\}$ and over $\{i\in\mathscr{I}_{n,2}\}$ may be treated in the same way as the first term in (3). In fact, the first difference in the sums over $\mathscr{I}_{n,1}$ and $\mathscr{I}_{n,2}$ are precisely of the same type as the first term in (3). Treating the second differences in these sums is simpler — we need not take care about approximating the difference of the derivative g'. Now, let us turn to the last sum in (4). First of all, let us observe that $d_{k(i)} \in (\xi_i(t, \omega), \zeta_i(t, \omega))$ means either

$$(5) \qquad d_{k(i)} \leqslant e_i \leqslant d_{k(i)} + \delta_{in}(t) \qquad (\text{when } \delta_{in}(t) \geqslant 0).$$

or

$$(6) \qquad d_{k(i)} + \delta_{in}(t) \leqslant e_i \leqslant d_{k(i)} \qquad (\text{when } \delta_{in}(t) < 0).$$

To simplify the rest of the proof of this step let us assume (without any loss of generality) that D_2 contains just one point d. We shall study the processes (for $j = 1, 2, \ldots, p$)

$$\mathscr{E}_j(t) = n^{-1+\tau}[\psi(d-) - \psi(d+)] \sum_{i=1}^{n} g_j'(X_i, \beta^0) \operatorname{sign}(\delta_{in}(t)) I_{\{i \in \mathscr{H}_{n,2}(t,\omega)\}}$$

where we have written the full notation for $\mathscr{H}_{n,2}(t, \omega)$ to indicate how the processes depend on t. First of all, let us consider $\tilde{\mathscr{E}}_j(t) = \mathscr{E}_j(t) - \mathbf{E}_F \mathscr{E}_j(t)$. Following again Jurečková [7] we have for $u \in R^p$, $u \geqslant t$ (the ordering is meant coordinatewise)

$$\begin{aligned}
&\mathbf{E}_G [\tilde{\mathscr{E}}_j(u) - \tilde{\mathscr{E}}_j(t)]^4 \\
&\quad \leqslant n^{-4+4\tau}[\psi(d-) - \psi(d+)]^4 \mathbf{E}_K \{11 \sum_{i=1}^{n} [g_j'(X_i, \beta^0)]^4 |F(d + \delta_{in}(u)) \\
&\qquad - F(d + \delta_{in}(t))| + [\sum_{i=1}^{n} [g_j'(X_i, \beta^0)]^2 |F(d + \delta_{in}(u)) - F(d + \delta_{in}(t))|]^2\} \\
&\quad \leqslant \|u - t\| \, O(n^{-3+3\tau}) + \|u - t\|^2 \, O(n^{-2+2\tau}) \text{ as } n \to \infty \\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{for } \max\{\|t\|, \|u\|\} < C.
\end{aligned}$$

So continuing to follow Jurečková [7] (see 2.12) and using the result of Jurečková and Sen [9] we obtain, for any $C > 0$, $\sup_{\|t\| < C} \|\tilde{\mathscr{E}}(t)\| = o_p(1)$ (an alternative possibility is to take the same steps as in [6] together with the result concerning multidimensional processes from [8]). An approximation to the mean values is similar: Taking into account that

$$\operatorname{sign}(\delta_{in}(t)) \cdot P_F(I_{\{i \in \mathscr{H}_{n,2}\}} = 1) = F(d + \delta_{in}(t)) - F(d)$$

we have

$$\begin{aligned}
&\left|\mathbf{E}_G [\mathscr{E}_j(t) - n^{-1}[\psi(d-) - \psi(d+)] \sum_{i=1}^{n} f(d) g_j'(X_i, \beta^0) [g'(X_i, \beta^0)]^T t]\right| \\
&\qquad = \Big|n^{-1+\tau}[\psi(d-) - \psi(d+)] \mathbf{E}_K \sum_{i=1}^{n} g_j'(X_i, \beta^0) [F(d + \delta_{in}(t)) - F(d) \\
&\qquad\qquad - f(d) [g'(X_i, \beta^0)]^T n^{-\tau} t]\Big|.
\end{aligned}$$

Now we may write similarly as above

$$F(d+\delta_{in}(t))-F(d)-f(d)[g'(X_i,\beta^0)]^T n^{-\tau} t$$

$$= \sum_{k=1}^{p} \int_{0}^{n^{-\tau}t_k} [f(d+g(X_i,\beta^0+z_t^{(k)})-g(X_i,\beta^0))g'_k(X_i,\beta^0+z_t^{(k)})$$

$$-f(d)g'_k(X_i,\beta^0)]\,dz.$$

Now from the continuity of f on $D_2(\tau_0)$ and the continuity of $g(x,\beta)$ and $g'(x,\beta)$ at β^0 (which is uniform in $x\in S_1$) we may find n_4 such that for any $n>n_4$ and any $\|t\|<C$ we have the last expression bounded by $\varepsilon\cdot J^{-1}\cdot[\psi(d-)-\psi(d+)]^{-1}n^{-\tau}\|t\|$. So we have derived that there is n_5 such that for any $n>n_5$

$$P_G\Big(\sup_{\|t\|<C} n^{-1+\tau}\Big|\sum_{i=1}^{n}[\psi(Y_i-g(X_i,\beta^0+n^{-\tau}t))g'_j(X_i,\beta^0+n^{-\tau}t)$$

$$-\psi(Y_i-g(X_i,\beta^0))g'_j(X_i,\beta^0)$$

$$+n^{-\tau}g'_j(X_i,\beta^0)[g'(X_i,\beta^0)]^T t\sum_{k=1}^{s_1} f(d_k)[\psi(d_k+)-\psi(d_k-)]]\Big|>\varepsilon\Big)<\varepsilon.$$

Finally, let us consider the last term in (3). Repeating the steps from the first part of the proof we find that

$$n^{-1}\sum_{i\notin\mathscr{H}_{n,3}}[\psi'(Y_i-g(X_i,\tilde{\beta}))g'(X_i,\tilde{\beta})[g'(X_i,\tilde{\beta})]^T$$

$$-\psi(Y_i-g(X_i,\tilde{\beta}))g''(X_i,\tilde{\beta})]\,t=o_p(1).$$

So we can treat the last term in (3) "without $I_{\{i\in\mathscr{H}_{n,3}\}}$". Now, using a standard technique of the approximation and the law of large numbers we find that

$$\Big|n^{-1}\sum_{i=1}^{n}[\psi'(Y_i-g(X_i,\tilde{\beta}))g'(X_i,\tilde{\beta})[g'(X_i,\tilde{\beta})]^T$$

$$-\psi(Y_i-g(X_i,\tilde{\beta}))g''(X_i,\tilde{\beta})]\,t-\mathbf{E}_F\psi'(e_1)g'(X_i,\beta^0)[g'(X_i,\beta^0)]^T t\Big|\to 0$$

in probability as $n\to\infty$. By this we have proved that all the three terms in (3) are of order $o_p(1)$ as $n\to\infty$, so the proof is complete. ∎

Remark 3. Let us observe that Lemma 1 is a generalization of Lemma 2.1 of Jurečková [7] for the non-linear setup. The steps of the proof were of course modified but the main ideas have been followed quite closely.

ASSERTION 1. *Let Conditions* A *and* B *be fulfilled. Then for any* $C>0$ *and* $\tau\in(0,\frac{1}{2}]$

$$n^{-1+2\tau} \sup_{\|t\|<C} \Big|\sum_{i=1}^{n} \big[\varrho\big(Y_i-g(X_i, \beta^0+n^{-\tau}t)\big)-\varrho\big(Y_i-g(X_i, \beta^0)\big)$$
$$+n^{-\tau}\psi\big(Y_i-g(X_i, \beta^0)\big)[g'(X_i, \beta^0)]^T t$$
$$-\tfrac{1}{2}n^{-2\tau}\gamma t^T g'(X_i, \beta^0)[g'(X_i, \beta^0)]^T t\big]\Big| \to 0$$

in probability as $n \to \infty$.

Proof (the schedule of the proof is analogous to the proof of Corollary 2.1 of [7]). First of all, let us define for any $i = 1, 2, \ldots, n$ and any $t \in R^p$

$$D(i, j, t) = \psi\big(Y_i-g(X_i, \beta^0+n^{-\tau}t)\big)g_j'(X_i, \beta^0+n^{-\tau}t)$$
$$-\psi\big(Y_i-g(X_i, \beta^0)\big)g_j'(X_i, \beta^0)+n^{-\tau}\gamma g_j'(X_i, \beta^0)[g'(X_i, \beta^0)]^T t.$$

By Lemma 1 we can find for any $\varepsilon > 0$ and any $C > 0$ an $n_\varepsilon \in N$ so that for $j = 1, 2, \ldots, p$ and $n > n_\varepsilon$ we have

$$P\Big\{n^{-1+\tau} \sum_{\|t\|<C} \Big|\sum_{i=1}^{n} D(i, j, t)\Big| > C^{-1}p^{-1}\varepsilon\Big\} < p^{-1}\varepsilon.$$

Let us put $B_n = \{n^{-1+\tau}\sup_{\|t\|<C}|\sum_{i=1}^n D(i, j, t)| > C^{-1}p^{-1}\varepsilon\}$. Then in the notation of $z_t^{(k)}$'s of the proof of Lemma 1 we have for any $\omega \in B_n^c$

$$n^{-1+2\tau} \sup_{\|t\|<C} \Big|\sum_{i=1}^{n} \big[\varrho\big(Y_i-g(X_i, \beta^0+n^{-\tau}t)\big)-\varrho\big(Y_i-g(X_i, \beta^0)\big)$$
$$+n^{-\tau}\psi\big(Y_i-g(X_i, \beta^0)\big)[g'(X_i, \beta^0)]^T t-\tfrac{1}{2}n^{-2\tau}\gamma t^T g'(X_i, \beta^0)[g'(X_i, \beta^0)]^T t\big]\Big|$$
$$= n^{-1+\tau} \sup_{\|t\|<C} \Big|\sum_{i=1}^{n}\sum_{k=1}^{p}\int_0^{t_k} D(i, k, z_t^{(k)})\,dz\Big|$$
$$\leqslant n^{-1+\tau}\sum_{k=1}^{p}\int_0^{t_k} \sup_{\|t\|<C}\Big|\sum_{i=1}^{n} D(i, k, z_t^{(k)})\Big|\,dz \leqslant \varepsilon$$

and the proof follows. ∎

Let us put $Z_n = n^{-1/2}\sum_{i=1}^n g'(X_i, \beta^0)\psi(e_i)$. Note that under Conditions A and B the central limit theorem holds for Z_n so that $Z_n = O_p(1)$.

ASSERTION 2. *Let Conditions* A *and* B *hold. Then for any* $\varepsilon > 0$ *there is* $C_\varepsilon > 0$ *such that for any* $C > C_\varepsilon$ *there is* $n_C \in N$ *so that for any* $n > n_C$ *we have*

$$P\big(B_n(C, \varepsilon)\big) < \varepsilon,$$

where

$$B_n(C, \varepsilon) = \Big\{\omega \in \Omega: \Big|\min_{\|t\|<C}\sum_{i=1}^{n}\big[\varrho\big(Y_i-g(X_i, \beta^0+n^{-1/2}t)\big)$$
$$-\varrho\big(Y_i-g(X_i, \beta^0)\big)\big]-\min_{\|t\|<C}\big[-t^T Z_n+\tfrac{1}{2}\gamma t^T Qt\big]\Big| > \varepsilon\Big\}.$$

Proof (again, the schedule of the proof is analogous to the proof of Lemma 3.1 of [7]). Assume that U_n is a solution of the minimization

$\min[-t^T Z_n + \frac{1}{2}\gamma t^T Qt]$. Then $U_n = \gamma^{-1} Q^{-1} Z_n$. Let us find C_ε and $n_0 \in N$ so that $P(\|U_n\| > C_\varepsilon) < \varepsilon/3$ and take any $C > C_\varepsilon$. Moreover, let us put

$$Q_n = n^{-1} \sum_{i=1}^{n} g'(X_i, \beta^0)[g'(X_i, \beta^0)]^T. \tag{7}$$

Since we assume only $t \in R^p$, $\|t\| < C$, we easily find $n_1 > n_0$ so that for any $n > n_1$

$$P_K\left(\tfrac{1}{2}\gamma |t^T(Q_n - Q)t| > \tfrac{1}{2}\varepsilon\right) < \varepsilon/3.$$

Finally, using Assertion 1 we find $n_\varepsilon > n_1$ such that putting

$$D_n = \Big\{\Big| \min_{\|t\|<C} \sum_{i=1}^{n} [\varrho(Y_i - g(X_i, \beta^0 + n^{-1/2}t)) - \varrho(Y_i - g(X_i, \beta^0))] - \min_{\|t\|<C} [-t^T Z_n + \tfrac{1}{2}\gamma t^T Qt]\Big| > \tfrac{1}{2}\varepsilon\Big\}$$

we have $P(D_n) < \varepsilon/3$ for any $n > n_\varepsilon$. Now we obtain

$$P(B_n(C, \varepsilon)) \leqslant P(D_n \cap \{\|U_n\| < C\}) + P(\{\tfrac{1}{2}\gamma |t^T(Q_n - Q)t| > \tfrac{1}{2}\varepsilon, \|U_n\| < C\}) + P(\{\|U_n\| > C\}) = \varepsilon. \blacksquare$$

Remark 4. Let D^Q be the determinant of the matrix Q and

$$\alpha = \max_{1 \leqslant i,j \leqslant p} |q_{i,j}|.$$

We have assumed that Q is regular (and hence positive definite), and hence $D^Q > 0$. Select any $\varepsilon > 0$ and $\delta > 0$, $\delta < \frac{1}{2}D^Q[p!]^{-1}\alpha^{1-p}$. Under Conditions A and B we may now find $n_\varepsilon \in N$ such that for any $n > n_\varepsilon$ we have

$$P_K\left(\max_{1 \leqslant i,j \leqslant p} |Q_{i,j} - (Q_n)_{i,j}| > \delta\right) < \varepsilon.$$

But this implies that also the determinants D^{Q_n} are positive with probability at least $1-\varepsilon$ starting with n_ε. This allows us to use Q_n^{-1} in the proof of the next lemma and formulate Corollary 1 below in a form containing also Q_n^{-1} (such a form we will need later).

LEMMA 2. *Let for the regression model*

$$Y_i = g(X_i, \beta^0) + e_i, \quad i = 1, 2, \ldots, n,$$

Conditions A, B *and* C *be fulfilled. Then the estimate* $\hat{\beta}^{(n)}$ *given in* (2) *is* $\sqrt{n}$*-consistent,*

$$\sqrt{n}(\hat{\beta}^{(n)} - \beta^0) = n^{-1/2}\gamma^{-1} Q^{-1} \sum_{i=1}^{n} g'(X_i, \beta^0)\psi(e_i) + o_p(1) \quad \text{as } n \to \infty$$

and $\sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ *has an asymptotically p-dimensional normal distribution* $\mathcal{N}_p(\mathbf{0}, \mathcal{G})$, *where* $\mathcal{G} = (\sigma_\psi^2/\gamma^2) Q^{-1}$ *and* $\sigma_\psi^2 = \mathsf{E}_F \psi^2(e_1)$.

Proof. The first step will be to prove $\sqrt{n}$-consistency of $\hat{\beta}^{(n)}$. In order to do it let us define for any $\beta \in R^p$ the residuals $r_i^{(n)}(\beta) = Y_i - g(X_i, \beta)$, $i = 1, 2, \ldots, n$, and

$$\xi_i(\beta) = \min\{e_i, r_i^{(n)}(\beta)\} \quad \text{and} \quad \zeta_i(\beta) = \max\{e_i, r_i^{(n)}(\beta)\}$$

and also for $j = 1$ and 2

$$\mathscr{K}_{n,j}(\hat{\beta}^{(n)}) = \{i \in N\colon (\xi_i(\hat{\beta}^{(n)}), \zeta_i(\hat{\beta}^{(n)})) \cap D_j \neq \emptyset\},$$

and finally

$$\mathscr{K}_{n,3}(\hat{\beta}^{(n)}) = \{1, 2, \ldots, n\} \backslash \{\mathscr{K}_{n,1}(\hat{\beta}^{(n)}) \cup \mathscr{K}_{n,2}(\hat{\beta}^{(n)})\}.$$

In what follows we shall omit $\hat{\beta}^{(n)}$ in $\mathscr{K}_{n,j}(\hat{\beta}^{(n)})$ provided it cannot cause a misunderstanding. Now we may write the expression

$$n^{-1/2} \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \hat{\beta}^{(n)})) g_j'(X_i, \hat{\beta}^{(n)}) - \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)]$$

as

$$(8) \quad n^{-1/2} \{ \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \hat{\beta}^{(n)})) g_j'(X_i, \hat{\beta}^{(n)}) - \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)] I_{\{i \in \mathscr{K}_{n,1}\}}$$
$$+ \sum_{i=1}^{n} [\psi(Y_i - g(X_i, \hat{\beta}^{(n)})) g_j'(X_i, \hat{\beta}^{(n)}) - \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0)] I_{\{i \in \mathscr{K}_{n,2}\}}$$
$$- n^{-\tau} \sum_{i=1}^{n} [\psi'(Y_i - g(X_i, \tilde{\tilde{\beta}})) g_j'(X_i, \tilde{\tilde{\beta}}) [g'(X_i, \tilde{\tilde{\beta}})]^T t - \psi(Y_i - g(X_i, \tilde{\tilde{\beta}})) [g''(X_i, \tilde{\tilde{\beta}}) t]_j] I_{\{i \in \mathscr{K}_{n,3}\}} \},$$

where $\|\tilde{\tilde{\beta}} - \beta^0\| < \|\hat{\beta}^{(n)} - \beta^0\|$ (compare (3)). Using nearly the same arguments as in the proof of Lemma 1 we find that the first term of (8) can be written as

$$o_p(1) \cdot \sqrt{n}(\hat{\beta}^{(n)} - \beta^0).$$

Also using the arguments of the proof of Lemma 1 we show that the second term is $O_p(1)$ (notice that in difference with the proof of Lemma 1 we have here a much more rough estimate of this term). Finally, we derive that the last term is asymptotically in probability equivalent to $\mathsf{E}^{-1}\psi(e_1) \cdot Q_n^{-1} \cdot \sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$ (for Q_n see (7)). So we arrive at the conclusion that

$$(9) \quad n^{-1/2} \{o_p(1) + \mathsf{E}^{-1}\psi(e_1) \cdot Q_n^{-1}\} \sqrt{n}(\hat{\beta}^{(n)} - \beta^0) = O_p(1) + n^{-1/2} \sum_{i=1}^{n} \psi(Y_i - g(X_i, \beta^0)) g_j'(X_i, \beta^0).$$

Using the Lyapunov theorem we verify that the right-hand side of (9) is $O_p(1)$. Finally, employing Lemma 4 (in the Appendix) we conclude that $\hat{\beta}^{(n)}$ is $\sqrt{n}$-consistent. The rest of the proof very closely mimicks, in the non-linear regression framework, all the steps carried out in the proof of Theorem 3.2 of [7]. In fact, once having proved $\sqrt{n}$-consistency of $\hat{\beta}^{(n)}$ the assertion of the lemma follows (directly) from Lemma 1 since we have $\|\hat{\beta}^{(n)}-\beta^0\| < n^{-1/2}C$ with probability $1-\varepsilon$, so that we may put $t = n^{1/2}(\hat{\beta}^{(n)}-\beta^0)$. ∎

COROLLARY 1. *Let the conditions of Lemma 2 hold. Then we have also*

$$\sqrt{n}(\hat{\beta}^{(n)}-\beta^0) = n^{-1/2}\gamma^{-1}Q_n^{-1}\sum_{i=1}^{n} g'(X_i, \beta^0)\psi(e_i)+o_p(1) \qquad \text{as } n\to\infty.$$

The corollary follows immediately from Lemma 2.

4. Test for differences between models. Let ϱ_j, $j = 1, 2$, be two distinct functions, both of them fulfilling Conditions A. Let us denote by $\hat{\beta}^{(n,j)}$ the M-estimate of regression coefficients generated by the function ϱ_j, i.e.,

$$\hat{\beta}^{(n,j)} = \arg\min_{\beta\in R^p}\sum_{i=1}^{n}\varrho_j\big(Y_i-g(X_i, \beta)\big).$$

Furthermore, let $r_i^{(n,j)}$ be the i-th residual with respect to the j-th estimate, i.e., for $i = 1, 2, \ldots, n$ and $j = 1, 2$

$$r_i^{(n,j)} = Y_i-g(X_i, \hat{\beta}^{(n,j)}).$$

The following lemma together with Corollary 1 will be a key tool for testing the differences between the estimates of model (1).

LEMMA 3. *Under Conditions* A, B *and* C *we have*

$$r_i^{(n,j)}-e_i = -[g'(X_i, \beta^0)]^T(\hat{\beta}^{(n,j)}-\beta^0)+O_p(n^{-1}).$$

Proof. We may write

$$\begin{aligned} r_i^{(n,j)} &= Y_i-g(X_i, \hat{\beta}^{(n,j)}) = Y_i-g(X_i, \beta^0)+g(X_i, \beta^0)-g(X_i, \hat{\beta}^{(n,j)}) \\ &= e_i-[g'(X_i, \beta^0)]^T(\hat{\beta}^{(n,j)}-\beta^0)+\tfrac{1}{2}(\hat{\beta}^{(n,j)}-\beta^0)^T g''(X_i, \tilde{\beta})(\hat{\beta}^{(n,j)}-\beta^0), \end{aligned}$$

where $\|\tilde{\beta}-\beta^0\| \leqslant \|\hat{\beta}^{(n,j)}-\beta^0\|$, and the lemma follows due to $\sqrt{n}$-consistency of $\hat{\beta}^{(n,j)}$ and the assumption B (ii). ∎

Making use of Corollary 1 and Lemma 3 we may find an asymptotic representation of $r_i^{(n,1)}-r_i^{(n,2)}$, and a test for difference between two estimates of model (1) may be based on a (normed) sum of the residuals. A derivation of corresponding results may closely follow that one in [12]. Here we will show the possibility to construct the test based on a sum of the squared differences of residuals. Using Corollary 1 and Lemma 3 we may find for any $\varepsilon > 0$ and $\delta > 0$ an $n_\varepsilon \in N$ such that for any $n > n_\varepsilon$ there is a set B_n with $P(B_n) > 1-\varepsilon$ and we have for any $\omega \in B_n$

$$n^{1/2}(r_i^{(n,1)} - r_i^{(n,2)})$$

$$= n^{-1/2} \sum_{l=1}^{p} \sum_{k=1}^{p} \sum_{j=1}^{n} [g_l'(X_i, \beta^0)]^T \tilde{q}_{lk}^{(n)} g_k'(X_i, \beta^0) \left[\frac{\psi_{(2)}(e_j)}{\gamma^{(2)}} - \frac{\psi_{(1)}(e_j)}{\gamma^{(1)}} \right] + \varkappa_{in}(\omega)$$

with $\max_{1 \leq i \leq n} \sup_{\omega \in B_n} |\varkappa_{in}(\omega)| < \delta$, where we have denoted by $\tilde{q}_{lk}^{(n)}$ the l, k-element of the matrix Q_n^{-1} (due to Remark 4 we may assume that B_n was also selected so that Q_n^{-1} exists at any $\omega \in B_n$). Putting

$$v^{(in)} = (v_{i1}, v_{i2}, \ldots, v_{in})^T \quad \text{with } v_{ij} = n^{-1/2} \sum_{l=1}^{p} \sum_{k=1}^{p} g_l'(X_i, \beta^0) \tilde{q}_{lk} g_k'(X_j, \beta^0)$$

and

$$\Psi_n(e) = \left(\frac{\psi_{(2)}(e_1)}{\gamma^{(2)}} - \frac{\psi_{(1)}(e_1)}{\gamma^{(1)}}, \frac{\psi_{(2)}(e_2)}{\gamma^{(2)}} - \frac{\psi_{(1)}(e_2)}{\gamma^{(1)}}, \ldots, \frac{\psi_{(2)}(e_n)}{\gamma^{(2)}} - \frac{\psi_{(1)}(e_n)}{\gamma^{(1)}} \right),$$

we obtain

$$n^{1/2}(r_i^{(n,1)} - r_i^{(n,2)}) = \Psi_n^T(e) \cdot v^{(in)} + \varkappa_{in},$$

and finally

$$\sum_{i=1}^{n} (r_i^{(n,1)} - r_i^{(n,2)})^2 = n^{-1} \Psi_n^T(e) \cdot v^{(in)} [v^{(in)}]^T \cdot \Psi_n(e) + \Delta_n(\omega), \tag{10}$$

where for any $\omega \in B_n$ we have $|\Delta_n(\omega)| < \delta$. Now we get

$$(11) \quad [\sum_{i=1}^{n} v^{(in)} [v^{(in)}]^T]_{jk}$$

$$= n^{-1} \sum_{i=1}^{n} \sum_{l_1=1}^{p} \sum_{l_2=1}^{p} g_{l_1}'(X_i, \beta^0) \tilde{q}_{l_1 l_2} g_{l_2}'(X_j, \beta^0) \sum_{l_3=1}^{p} \sum_{l_4=1}^{p} g_{l_3}'(X_i, \beta^0) \tilde{q}_{l_3 l_4} g_{l_4}'(X_k, \beta^0)$$

$$= \sum_{l_1=1}^{p} \sum_{l_2=1}^{p} g_{l_1}'(X_j, \beta^0) \tilde{q}_{l_1 l_2} g_{l_1}'(X_k, \beta^0).$$

We have prepared nearly everything to be able to give the promised test statistic and to derive its asymptotic distribution. Earlier however we still need to prove some other assertion.

ASSERTION 3. *Under Conditions* A *and* B *for any* $\varepsilon > 0$ *there exists* $n_\varepsilon \in N$ *so that for any* $n > n_\varepsilon$ *there is an* $(n \times p)$*-matrix* D_n *and a set* A_n *with* $P(A_n) > 1 - \varepsilon$ *such that for any* $\omega \in A_n$ *we have*

$$\sum_{i=1}^{n} v^{(in)} [v^{(in)}]^T = D_n \cdot D_n^T \quad \text{and} \quad n^{-1} D_n^T \cdot D_n = \mathscr{I}_p,$$

where $\mathscr{I}_p$ *denotes the* $(p \times p)$ *identity matrix.*

Proof. Let us find for a fixed $\varepsilon > 0$ an $n_\varepsilon \in N$ such that for any $n > n_\varepsilon$ there is a set A_n, $P(A_n) > 1-\varepsilon$, and Q_n is regular for any $\omega \in A_n$. Let us assume an arbitrary point $\omega_0 \in A_n$. Since $Q_n^{-1}(\omega_0)$ is symmetric and regular, it may be written as $C_n \cdot C_n^T$, where C_n is a regular $(p \times p)$-matrix (see, e.g., [11], 1b.1 VI). Now denote by $\mathscr{G}_n$ the matrix with $(\mathscr{G}_n)_{kj} = g_j'(X_k, \beta^0)$, $k = 1, 2, \ldots, n$, $j = 1, 2, \ldots, p$, and put $D_n = \mathscr{G}_n C_n$. Then we have

$$D_n D_n^T = \mathscr{G}_n C_n C_n^T \mathscr{G}_n^T = \mathscr{G}_n Q_n^{-1} \mathscr{G}_n^T,$$

which according to (11) gives $\sum_{i=1}^n v^{(in)} \cdot [v^{(in)}]^T$. Moreover, let us consider

$$C_n [n \mathscr{I}_p - C_n^T \mathscr{G}_n^T \mathscr{G}_n C_n] C_n^T.$$

Keeping in mind that $\mathscr{G}_n^T \mathscr{G}_n = nQ_n$ (see (7)) we find that this expression is equal to

$$nC_n C_n^T - C_n C_n^T \mathscr{G}_n^T \mathscr{G}_n Q_n^{-1} = \mathbf{0},$$

which in turn implies that $\mathscr{I}_p = n^{-1} D_n^T D_n$. ■

THEOREM 1. *Under Conditions* A, B *and* C *the sum*

$$S_n^2 = \sigma^{-2} \sum_{i=1}^{n} (r_i^{(n,1)} - r_i^{(n,2)})^2,$$

where

$$\sigma^2 = [\gamma^{(1)}]^{-2} \operatorname{var}(\psi_{(1)}(e_1)) - 2[\gamma^{(1)}\gamma^{(2)}]^{-1} \operatorname{cov}(\psi_{(1)}, \psi_{(2)}) + [\gamma^{(2)}]^{-2} \operatorname{var}(\psi_{(2)}(e_1)),$$

has an asymptotical χ^2-distribution with p degrees of freedom.

Proof. Let us put

$$T_n = \frac{1}{\sigma\sqrt{n}} \Psi_n^T(e) \cdot D_n.$$

Using the previous assertion, the Lyapunov theorem and the fact that Q_n are coordinatewise bounded in probability we find that

$$\mathscr{L}(T_n) \to \mathscr{N}_p(\mathbf{0}, \mathscr{I}_p) \quad \text{as } n \to \infty.$$

Now let $\tilde{U}$ be a random variable distributed according to $\mathscr{N}_p(\mathbf{0}, \mathscr{I}_p)$. Using Slucky's theorem we derive that $\tilde{U}^T \tilde{U} - T_n^T T_n \to 0$ in probability. On the other hand, we know that $\tilde{U}^T \tilde{U}$ has χ^2-distribution with p degrees of freedom, and the proof follows. ■

Remark 5. It is easy to see that σ^2 may be substituted with some estimate $\hat{\sigma}^2$. Some difficulties might be caused by the fact that for a non-continuous ψ-function we need to estimate

$$\sum_{l=1}^{s_2} f(d_{2l}) [\psi(d_{2l}+) - \psi(d_{2l}-)].$$

However, using some result of the strong approximation (see, e.g., [4], Theorem 6.1.1) one may cope with the problem.

Conclusion. The paper proposes a possibility how to evaluate the statistical significance of the difference between two estimates of the regression model. Since the evaluation of the required statistic is easy, the test is simple to apply.

If for some data and ψ_1 and ψ_2 the difference appears to be significant, it may indicate that for instance at least one of these ψ-functions was not appropriate for given data or that the asymptotic consistency (and asymptotic normality) does not yet work, etc. Then some further analysis is inevitable. One may for instance try to employ some graphical diagnostic tools (see [3], [16] or [2]) – probably there have been not yet available results concerning non-linear regression diagnostic of this type because, e.g., [1] is devoted mostly to other problems than those discussed in this paper. Another possibility is to find a subset of data which is the most influential for estimation (see [15]) and to look for the estimating method giving more stable estimates on the complementary subsamples.

Acknowledgement. We would like to express our gratitude to our colleague Petr Lachout for the careful reading of the manuscript and the valuable suggestions which made the proof of Lemma 4 of the Appendix trasparent.

APPENDIX

LEMMA 4. *Let for some* $p \in N$, $\{\mathscr{V}^{(n)}\}_{n=1}^{\infty}$, $\mathscr{V}^{(n)} = \{v_{ij}^{(n)}\}_{i=1,2,\ldots,p}^{j=1,2,\ldots,p}$ *be a sequence of $(p\times p)$-matrices such that for* $i = 1, 2, \ldots, p$ *and* $j = 1, 2, \ldots, p$

$$\lim_{n\to\infty} v_{ij}^{(n)} = q_{ij} \text{ in probability,}$$

where $Q = \{q_{ij}\}_{i=1,2,\ldots,p}^{j=1,2,\ldots,p}$ *is a fixed non-random regular matrix. Moreover, let* $\{\theta^{(n)}\}_{n=1}^{\infty}$ *be a sequence of p-dimensional random vectors such that*

$$\exists(\varepsilon > 0)\ \forall (K > 0)\ \limsup_{n\to\infty} P(\|\theta^{(n)}\| > K) > \varepsilon. \tag{12}$$

Then

$$\exists(k \in \{1, 2, \ldots, p\} \text{ and } \delta > 0)\ \forall (L > 0)\ \limsup_{n\to\infty} P\Big(\Big|\sum_{j=1}^{p} v_{kj}^{(n)}\theta_j^{(n)}\Big| > L\Big) > \delta.$$

Proof. Let us at first assume that for the sequence $\{\theta^{(n)}\}_{n=1}^{\infty}$ we have

$$\exists(\varepsilon > 0)\,\forall (K > 0)\ \lim_{n\to\infty} P(\|\theta^{(n)}\| > K) > p\varepsilon. \tag{13}$$

Let us fix a sequence $\{\tilde{K}_r\}_{r=1}^{\infty} \uparrow \infty$, $\tilde{K}_1 = 0$, and construct a sequence $\{K_n\}_{n=1}^{\infty}$ in the following way. For every $r \in N$ find $n_r \in N$ such that for any $n \geqslant n_r$, $n \in N$,

$$P(\|\theta^{(n)}\| > \tilde{K}_r) > p\varepsilon/2$$

and put for $l \in N$, $l \in [n_r, n_{r+1})$, $K_l = \tilde{K}_r$ (if $n_1 > 1$ put $K_l = 0$ for $l \leqslant n_1$). Write

$$B_n = \{\omega \in \Omega \colon \|\theta^{(n)}\| > K_n\}, \quad \text{i.e.,} \quad P(B_n) > p\varepsilon/2 \text{ for } n \geqslant n_1.$$

Let us consider $n > n_1$ and for any $\omega \in B_n$ let us put $\tilde{\theta}^{(n)} = \theta^{(n)} \cdot \|\theta^{(n)}\|^{-1}$. Then we have $\|\tilde{\theta}^{(n)}\| = 1$ for all $n > n_1$ and $\omega \in B_n$. Let us denote the elements of Q^{-1} by $\tilde{q}_{ij}$ and

$$\sum_{l=1}^{p} \sum_{k=1}^{p} [\tilde{q}_{lk}]^2 = (2\Delta \cdot p^{-1/2})^{-2}.$$

Now

$$[\tilde{\theta}_l^{(n)}]^2 = [(Q^{-1} Q \tilde{\theta}^{(n)})_l]^2 = \Big[\sum_{k=1}^{p} \tilde{q}_{lk} \sum_{j=1}^{p} q_{kj} \tilde{\theta}_j^{(n)}\Big]^2$$

$$\leqslant \sum_{k=1}^{p} [\tilde{q}_{lk}]^2 \sum_{i=1}^{p} \Big[\sum_{j=1}^{p} q_{ij} \tilde{\theta}_j^{(n)}\Big]^2 = \sum_{k=1}^{p} [\tilde{q}_{lk}]^2 \|Q \cdot \tilde{\theta}^{(n)}\|^2,$$

and hence

$$\|\tilde{\theta}^{(n)}\|^2 \leqslant (2\Delta \cdot p^{-1/2})^{-2} \|Q \cdot \tilde{\theta}^{(n)}\|^2,$$

so that for any $n > n_1$ and $\omega \in B_n$ we have $\|Q \cdot \tilde{\theta}^{(n)}\| \geqslant 2\Delta \cdot p^{-1/2}$. This implies that for any $n \in N$ and $\omega \in B_n$ there is $k(n) \in \{1, 2, \ldots, p\}$ so that

$$\Big|\sum_{j=1}^{p} q_{k(n)j} \tilde{\theta}_j^{(n)}\Big| \geqslant 2\Delta.$$

Putting $B_n^{k(n)} = \{\omega \in \Omega \colon |\sum_{j=1}^{p} q_{k(n)j} \tilde{\theta}_j^{(n)}| \geqslant 2\Delta\}$, we have $P(B_n^{k(n)}) > \varepsilon/2$ for $n \geqslant n_1$. Now, let us select $n_\Delta \in N$ such that, for any $n \in N$, $n \geqslant n_\Delta$,

$$P\big(\max_{i,j} |v_{ij}^{(n)} - q_{ij}| > \Delta p^{-1/2}\big) < \varepsilon/(4p^2).$$

Write $C_n = \{\omega \in \Omega \colon \max_{i,j=1,\ldots,p} |v_{ij}^{(n)} - q_{ij}| < \Delta p^{-1/2}\}$. Then for any $n > n_\Delta$ we have

$$P(C_n^c) \leqslant \sum_{i=1}^{p} \sum_{j=1}^{p} P(|v_{ij}^{(n)} - q_{ij}| \geqslant \Delta p^{-1/2}) < \frac{\varepsilon}{4p^2} p^2 = \frac{\varepsilon}{4}.$$

Since $B_n^{k(n)} \cap C_n = B_n^{k(n)} - C_n^c$, we have for any $n \in N$, $n > n_0 = \max\{n_1, n_\Delta\}$,

$$P(B_n^{k(n)} \cap C_n) \geqslant P(B_n^{k(n)}) - P(C_n^c) \geqslant \varepsilon/2 - \varepsilon/4 = \varepsilon/4.$$

For any $n > n_0$ and any $\omega \in B_n^{k(n)} \cap C_n$ we obtain

$$\Big|\sum_{j=1}^{p} v_{k(n)j}^{(n)} \tilde{\theta}_j^{(n)}\Big| \geqslant \Big|\sum_{j=1}^{p} q_{k(n)j} \tilde{\theta}_j^{(n)}\Big| - \sum_{j=1}^{p} |v_{k(n)j}^{(n)} - q_{k(n)j}| \, \tilde{\theta}_j^{(n)} \geqslant \Big|\sum_{j=1}^{p} q_{k(n)j} \tilde{\theta}_j^{(n)}\Big| - \Delta = \Delta,$$

which means that

$$\Big|\sum_{j=1}^{p} v_{k(n)j}^{(n)}\,\theta_j^{(n)}\Big| = \|\theta^{(n)}\|\,\Big|\sum_{j=1}^{p} v_{k(n)j}^{(n)}\,\tilde{\theta}_j^{(n)}\Big| \geqslant \|\theta^{(n)}\|\,\Delta$$

and the proof follows. To prove the lemma with (12) instead of (13) it is sufficient to assume that the lemma does not hold and to select a subsequence $\{\theta^{(n_l)}\}_{l=1}^{\infty}$ for which (13) holds and we get a contradiction. ∎

REFERENCES

[1] D. S. Borowiak, *Model Discrimination for Nonlinear Regression Models*, M. Dekker, New York 1989.

[2] S. Chatterjee and A. S. Hadi, *Sensitivity Analysis in Linear Regression*, Wiley, New York 1988.

[3] R. D. Cook and S. Weisberg, *Residuals and Influence in Regression*, Chapmann and Hall, New York 1982.

[4] M. Csörgő and P. Révész, *Strong Approximation in Probability and Statistics*, Akademia Kiadó, Budapest 1981.

[5] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahael, *Robust Statistics — The Approach Based on Influence Function*, Wiley, New York 1986.

[6] J. Jurečková, *Regression quantiles and trimmed least squares estimator under a general design*, Kybernetika 20 (1984), pp. 345–357.

[7] — *Consistency of M-estimators in linear model generated by non-monotone and discontinuous ψ-functions*, Probab. Math. Statist. 10 (1988), pp. 1–10.

[8] — and P. K. Sen, *On adaptive scale-equivariant M-estimators in linear models*, Statist. Decisions 2 (Suppl. Issue No 1) (1984), pp. 31–46.

[9] — *Uniform second order asymptotic linearity of M-statistics in linear models*, ibidem 7 (1989), pp. 263–276.

[10] F. Liese and I. Vajda, *M-estimators in regression with i.i.d. errors*, preprint, 1992.

[11] C. R. Rao, *Linear Statistical Inference and Its Applications*, Wiley, New York 1973.

[12] A. M. Rubio, L. Aguilar and J. Á. Víšek, *Testing difference of models*, Computational Statistics 8 (1993), pp. 57–70.

[13] D. Ruppert and R. J. Carroll, *Trimmed least squares estimation in linear model*, J. Amer. Statist. Assoc. 75 (1980), pp. 828–838.

[14] J. Á. Víšek, *Stability of regression models estimates with respect to subsamples*, Computational Statistics 7 (1992), pp. 183–203.

[15] — *Data subsamples influence in the M-estimation of the nonlinear regression models*, submitted, 1993.

[16] R. E. Welsh, *Influence function and regression diagnostic*, in: R. L. Launer and A. F. Siegel (Eds.), *Modern Data Analysis*, Academic Press, New York 1982.

A. M. Rubio, F. Quintana
Departamento de Matemáticas
Universidad de Extremadura
10071 Cáceres & Badajoz

J. Á. Víšek
Department of Stochastics Informatics
Institute of Information Theory and Automation
Academy of Sciences of Czech Republik
Pod vodárenskou věží 4, 182 08 Prague

Received on 1.4.1993