KYBERNETIKA — VOLUME 32 (1996), NUMBER 4, PAGES 353–374

A NOTE ON ASYMPTOTIC LINEARITY OF M–STATISTICS IN NONLINEAR MODELS

ASUNCIÓN MÁRIA RUBIO AND JAN ÁMOS VÍŠEK

For a smooth nonlinear regression model the conditions for the uniform second order asymptotic linearity of the M-statistics in the regression parameters are given. The existence of the $\sqrt{n}$-consistent estimator of the regression parameters and the role of the rescaling residuals in the M-estimation are briefly discussed.

1. INTRODUCTION

Recently more and more attention has been paid to nonlinear models. Some results, as e. g. testing the differences between models or the study of the subsample stability of models (see [16], [18] and [19]), were established for the linear models using as a key tool the Bahadur representation of the estimators. For the nonlinear models this representation has been derived in [17] and used for constructing a test of the differences of estimates. Due to the importance of the rescaling residuals in the statistical inference, earlier or later there will be a need of a version of this representation with rescaled residuals.

This note derives it by generalizing the results of Jurečková and Sen [9] for linear models. To facilitate reading for a reader who is familiar with the paper [9] we have preserved the structure of it so far as possible (moreover, the generalization follows very closely all steps from [9] and the whole matter is mainly a technical one). So, it seems that more important are some related problems which were raised by Jana Jurečková. First of all, in the nonlinear setup we know much less about consistency of M-estimators than in the linear setup. Although there are already some results (see [11]), they were not established for the case when the residuals are assumed to be rescaled. Moreover we need even $\sqrt{n}$-consistency. Similar situation is with the rescaling of residuals. That is why we shall at first consider these questions.

So the plan of the present paper is as follows. At first we shall give a basic notation and conditions on the regression model. Secondly, we will briefly discuss some problems which were mentioned a few lines above. We shall give ideas how to cope with them. Without this discussion the next generalization would be only a theoretical game, may be without any consequences from the practical point of view. Then we shall present the promised asymptotic linearity of M-statistics in nonlinear setup, and finally, we shall pay attention to the discontinuous ψ-functions.

2. NOTATION AND CONDITIONS ON REGRESSION MODEL

Let $(\Omega, \mathcal{B}, P)$ be a probability space. (In what follows the all "$o_p(\cdot)$" as well as "$O_p(\cdot)$" are understood with respect to this P.) We shall consider the nonlinear regression model

$$Y_i(\omega) = g(X_i(\omega), \beta^0) + e_i(\omega), \quad i = 1, 2, \dots \tag{1}$$

where the sequence $\{Y_i(\omega)\}_{i=1}^{\infty}$, $Y_i(\omega) : \Omega \to R$ represents responses of the model, $\{X_i(\omega)\}_{i=1}^{\infty}$, $X_i(\omega) \to R^p$, the carriers of model, are assumed to be a sequence of independent and identically disributed random variables (i.i.d.r.v.) and $\beta^0 = (\beta_1^0, \beta_2^0, \dots, \beta_p^0)^{\mathrm{T}}$ is the vector of the regression parameters (coefficients) ("T" indicates the transposition). Further, $\{e_i(\omega)\}_{i=1}^{\infty}$, $e_i(\omega) : \Omega \to R$ is another sequence of i.i.d.r.v., independent from $\{X_i(\omega)\}_{i=1}^{\infty}$. We shall also assume that $\mathrm{var}_P(e_i) \in (0, \infty)$. Finally, let the two times differentiable function g fulfill the following conditions:

CONDITIONS A

i) $\exists\, (\kappa > 0) \quad \forall\, (\|\beta - \beta^0\| < \kappa, \; x \in R \text{ and } j,\, k = 1, 2, \dots, p)$
$\exists (g_j'(x, \beta) = \frac{\partial}{\partial \beta_j} g(x, \beta) \text{ and } g_{jk}'(x, \beta), g_{jk}''(x, \beta) = \frac{\partial^2}{\partial \beta_j \partial \beta_k} g(x, \beta))$.

ii) $\exists\, (J < \infty)$

$$\max_{1 \le j,k \le p} \; \sup_{x \in R, \|\beta - \beta^0\| < \kappa} \; \max \left\{ |g(x, \beta)|, |g_j'(x, \beta)|, |g_{jk}''(x, \beta)| \right\} < J.$$

iii) $\exists\, (L > 0) \qquad \forall (\beta \in R^p, \|\beta - \beta^0\| < \kappa)$

$$\max_{1 \le j,k \le p} \sup_{x \in S} |g_{jk}''(x, \beta) - g_{jk}''(x, \beta^0)| < L \cdot \|\beta - \beta^0\|.$$

Recalling that $e_i(\omega) = Y_i(\omega) - g(X_i(\omega), \beta^0)$, let us put $\delta_{in}(t) = g\left(X_i, \beta^0 + n^{-\frac{1}{2}} t\right) - g(X_i, \beta^0)$. Further, denote $q = \mathsf{E}_P g''(x, \beta^0)$, $Q = \mathsf{E}_P \left\{ g'(x, \beta^0) \left[g'(x, \beta^0) \right]^{\mathrm{T}} \right\}$ and for any finite set $A = \{a_1, \dots, a_s\}$ and $\nu > 0$ put $A(\nu) = \bigcup_{i=1}^{s} [a_i - \nu, a_i + \nu]$. Moreover, let $F(z)$ and $G(x, z)$ denote the distribution function of $e_i \sigma^{-1}$ and of $(X_i^{\mathrm{T}}, e_i \sigma^{-1})^{\mathrm{T}}$, respectively.

The behaviour of the sum

$$S_n(t, u) = \sum_{i=1}^{n} \left[\psi\left([e_i - \delta_{in}(t)]\, \sigma^{-1} e^{-n^{-\frac{1}{2}} u} \right) g'(X_i, \beta^0 + n^{-\frac{1}{2}} t) - \psi(e_i \sigma^{-1})\, g'(X_i, \beta^0) \right] \tag{2}$$

for $\max\{\|t\|, |u|\} < C$ will be studied under various conditions on the ψ-function.

Finally, let $S_{n1}(t, u)$ denote the first coordinate of the sum $S_n(t, u)$ and I_B the indicator of a set B. We shall assume that $\mathsf{E}_P \psi(e_1 \sigma^{-1}) < \infty$. Then, taking $\tilde{\psi}(t) = \psi(t) - \mathsf{E}_P \psi(e_1 \sigma^{-1})$, if necessary, we have $\mathsf{E}_P \tilde{\psi}(e_1 \sigma^{-1}) = 0$. Hence if it will not be said something else, we shall assume that $\mathsf{E}_P \psi(e_1 \sigma^{-1}) = 0$.

3. CONSISTENCY AND THE RESCALING OF RESIDUALS

As it was already mentioned in Introduction the applicability of the results which will be established requires a discussion of several questions, namely:

- Is there, under the Conditions A (and possibly some additional ones, e. g. on ψ-function), any $\sqrt{n}$-consistent M-estimator on which we can then apply our results?
- What is the role of rescaling in the nonlinear regression where the scale-invariance of the M-estimators has no (or at least considerably modified) sense?
- Can we hope that the equation

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - g(X_i, \beta)}{\hat{\sigma}_n}\right) g'(X_i, \beta) = 0 \tag{3}$$

can be (approximately) fulfilled also for noncontinuous ψ-functions (as e. g. in [15])?

We shall briefly discuss now the first two problems and we shall leave the last one to the end of paper.

3.1. Consistency of the M-estimators

In this subsection we shall assume that the ψ-function can be decomposed as

$$\psi(z) = \psi_a(z) + \psi_c(z) \tag{4}$$

where $\psi_a(z)$ is absolutely continuous with the absolutely continuous derivative and $\psi_c(z)$ is continuous with the derivative which is a step function (with finite number of jumps, say at the points $r_1, r_2, \ldots, r_k$) and let us define γ_1 and γ_2 by

$$\gamma_1 = \mathsf{E}_P\left\{\sigma^{-1}\psi'(e_1\sigma^{-1})\right\} \qquad \text{and} \qquad \gamma_2 = \mathsf{E}_P\left\{(e_1\sigma^{-1})\,\psi'(e_1\sigma^{-1})\right\}.$$

Moreover, let Conditions A hold and let F have a bounded derivative f in neighborhoods of the points $r_1, r_2, \ldots, r_k$. Finally, putting for any $\delta > 0$

$$\psi''_\delta(y) = \sup\left\{|\psi''(y+z)| : |z| \le \delta\right\}$$

and

$$\overline{\psi}''_\delta(y) = \sup\left\{|\psi''(\exp(w)\,(y+z))| : \{|z|, |w|\} \le \delta\right\},$$

let for some $\delta_0 > 0$ and $\nu > 1$

$$\mathsf{E}_P\left\{|t\overline{\psi}''_{\delta_0}(t)|^\nu\right\} < \infty \qquad \text{and} \qquad \mathsf{E}_P\left\{|t^2\psi''_{\delta_0}(t)|^\nu\right\} < \infty$$

for all $\delta \in (0, \delta_0]$ and γ_1 as well as γ_2 are finite.

Then we shall show that under these conditions for any $\varepsilon > 0$ there is $K_1 > 0$, $K_2 > 0$ and $n_0 \in N$ such that for any $n > n_0$ there is a set B_n such that $P(B_n) > 1 - \varepsilon$

and for any $\omega \in B_n$ and any $u \in R^+$, $|u| < K_1$ there is a solution $\hat{\beta}^{(n)}(u,\omega)$ of the equation

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - g(X_i,\beta)}{u}\right) g'(X_i,\beta) = 0 \tag{5}$$

such that we have $\sqrt{n}\,\|\hat{\beta}^{(n)}(u,\omega) - \beta^0\| < K_2$. The basic step of the proof will be the utilization of the fix-point theorem (similarly as in [8]) in a form which we shall now recall.

Assertion 1. Let U be an open, bounded set in R^p and assume that $Q(z) : \overline{U} \subset R^p \to R^p$ ($\overline{U}$ is the closure of U) is continuous and satisfies $(z - z_0)^{\mathrm{T}} Q(z) \geq 0$ for some $z_0 \in U$ and all $z \in \overline{U} \setminus U$. Then the equation $Q(z) = 0$ has a solution in $\overline{U}$.

For the proof see [13], Assertion 6.3.4 on the page 163.

Now using (18) and (26) we arrive at

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi\left(\left[e_i - \delta_{in}(t)\right] \sigma^{-1} e^{-n^{-\frac{1}{2}}u}\right) g'\left(X_i, \beta^0 + n^{-\frac{1}{2}}t\right)$$

$$= n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi(e_i\sigma^{-1})\, g'(X_i,\beta^0) - \gamma_1 Q t - \gamma_2 q u + O_p\left(n^{-\frac{1}{2}}\right). \tag{6}$$

Due to Conditions A and the assumptions on the functions ψ_a and ψ_c it is possible to verify that the assumptions of Feller–Lindeberg theorem are fulfilled for the sequence of random variables

$$\left\{\psi(e_i\sigma^{-1})\, g'(X_i,\beta^0)\right\}_{i=1}^{\infty}$$

and due to the fact that we have assumed that $\mathsf{E}_P \psi(e_1\sigma^{-1}) = 0$,

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi(e_i\sigma^{-1})\, g'(X_i,\beta^0)$$

is bounded in probability (independently on t and u). It means that for any $\varepsilon > 0$ there is a constant $K_3 > 0$ and $n_0 \in N$ so that for any $n > n_0$ we have for

$$B_n = \left\{\omega \in \Omega \;:\; \left\| n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi(e_i\sigma^{-1})\, g'(X_i,\beta^0) \right\| < K_3 \right\}$$

$P(B_n) > 1 - \varepsilon$.

However it implies that for any $K_1 > 0$ and any $u \in (0, K_1)$, due to the linearity in t of

$$t^{\mathrm{T}} n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi(e_i\sigma^{-1})\, g'(X_i,\beta^0)$$

and of

$$t^{\mathrm{T}} \gamma_2 q u,$$

there is $K_2 > 0$ so that for any $n > n_0$ and $\omega \in B_n$ we have for any $t \in R^p$ such that $\|t\| = K_2$

$$-t^{\mathrm{T}} n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi\left(\left[e_i - \delta_{in}(t)\right] \sigma^{-1} e^{-n^{-\frac{1}{2}} u}\right) g'\left(X_i, \beta^0 + n^{-\frac{1}{2}} t\right)$$

$$= \quad -t^{\mathrm{T}} n^{-\frac{1}{2}} \sum_{i=1}^{n} \psi(e_i \sigma^{-1})\, g'(X_i, \beta^0) + t^{\mathrm{T}} \gamma_1 Q t + t^{\mathrm{T}} \gamma_2 q u + O_p\left(n^{-\frac{1}{2}}\right) \geq 0.$$

Applying Assertion 1 we find that there is $t \in R^p$ such that $\|t\| \leq K_2,\ t = t(u, \omega)$ which solves

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - g(X_i, \beta^0 + n^{-\frac{1}{2}} t)}{u}\right) g'\left(X_i, \beta^0 + n^{-\frac{1}{2}} t\right) = 0.$$

Writing

$$t(u) = \sqrt{n}\left(\hat{\beta}^{(n)}(u, \omega) - \beta^0\right),$$

we conclude the proof of the promised assertion.

3.2. The role of rescaling residuals in the regression analysis

Now let us discuss the problem for what is useful the rescaling of residuals in the nonlinear regression setup. One may find that the residuals were studentized in linear regression to make the M-estimators scale invariant. But in nonlinear regression this reason seems to be problematic (anyway, the group of functions for which we would like to have the invariance, would be surely different from the group of linear functions). But let us look on the situation more carefully.

The rescaling has been used generally to avoid difficulties with unknown scale parameter. Do we need it also in the regression analysis? Let us return to the history of building up the robust methods to clarify the question.

On the base of theoretical results (see [5], [6]) we use for M-estimation the families of optimal ψ-functions. E.g. when the central model (i.e. the distribution of the bulk of residuals) is assumed to be standard normal one, then we use $\{\psi_k(z)\}_{k>0}$ where $\psi_k(z) = \operatorname{sign}(z) \cdot \min\{|z|, k\}$. In both approaches, presented in [5] and [6], the optimality is reached when the underlying model which generated bulk of data is the same as the model which was used to determine ψ-function, and when the "tuning" constant was properly selected. Let us discuss at first the selection of the "tuning" constant.

According to the first approach (see [5]) the "tuning" constant k should be selected so that the estimator attains a required gross error sensitivity. Although the value which we assign to the gross error sensitivity seems to be rather arbitrary (depending only on our taste how much we admit that the estimator may react on the gross errors), implicitly it is related to the contamination level of data and to the variance of data, see [20].

On the other hand, according to the second approach, as we may see from the pioneering paper of Huber [6], the "tuning" constant should be found so that the asymptotic variance of the corresponding estimator is minimal (under an assumed contamination level). The contamintion level of data is of course unknown but it does not mean that we should not try to adapt the estimator to this unknown level. As it was shown in [20] the effect of selection of $k(\varepsilon)$ on the asymptotic variance (i.e. on efficiency) of the M-estimator is small (or at least very "smooth"). However the effect in the sense whether we obtain the estimate (i.e. numerical values which we obtain when applying the estimator on given data) near to the "true" model is unfortunately considerable (because one may easy find examples of data for which, with varying tuning constant, the estimate of regression parameters varies much more than we would expect and than it is acceptable for applications, see e.g. [21]). So the selection of the tuning constant which is appropriately adapted to the contamination level of given data and their variance is crutial.

Of course, we reach the full optimality only when for former approach the model which generated bulk of data is the same as the model which generated ψ-function, and for the latter when the ψ-function is the derivative of the logarithm of the density of data-generating model (which is nearly the same).

Anyway, in both [5] and [6] (and also in others, e.g. [4]) we assume that the variance of the residuals is not very far from the variance given by the central model. Naturally, instead of rescaling the residuals we may use e.g. the family $\left\{\psi^{(\sigma)}(z)\right\}_{k>0}$ where $\psi_k^{(\sigma)}(z) = \text{sign}(z)\cdot\sigma^{-1}\cdot\min\{|z|, k(\sigma)\}$. However the employment of the latter possibility may lead to some numerical difficulties, and hence the practitioners are used to utilize the families of ψ-functions which assume a fix variance of residuals (see [12] and the library ROBETH; but a rescaling of data before evaluating estimates (and not only the estimates of regression parameters) is performed practically by any software).

So, the rescaling of residuals (both in linear and nonlinear models but of course anywhere else) allows us to rid of dependence of the procedure on the scale parameter, i.e. it allows us to use standartized families of the criterial functions (consequence of which is that we avoid computational difficulties and in the theoretical reflection we may use one fix criterial function instead of a sequence of them). In M-estimation, it means some standartized families of ψ-functions. It simplifies selection of the proper ψ-function, sometimes reducing it on selection of a proper tuning constant.

4. SECOND ORDER ASYMPTOTIC LINEARITY

4.1. Step-function ψ

Let $\psi(x) = \alpha_j$, for $x \in (r_j, r_{j+1}]$, $j = 0, 1, \ldots, k$ (please, read $(r_k, \infty]$ as (r_k, ∞)) where $\alpha_0, \ldots, \alpha_k$ are real distinct numbers and $-\infty = r_0 < r_1 < \ldots < r_k < r_{k+1} = \infty$, k being a positive integer. Put $\gamma_1 = \sum_{j=1}^k (\alpha_j - \alpha_{j-1}) f(\sigma r_j)$ and $\gamma_2 = \sum_{j=1}^k r_j(\alpha_j - \alpha_{j-1}) f(\sigma r_j)$.

Theorem 1. Let Conditions A hold. Moreover, let F has in neighborhoods of the points $r_1, r_2, \ldots, r_k$ bounded derivatives f and f'. Then for any $C > 0$

$$\sup\left\{\left\|n^{-\frac{1}{2}}S_n(t,u)+\gamma_1 Qt+\gamma_2 qu\right\| : \max\{\|t\|,|u|\}<C\right\}=O_p\left(n^{-\frac{1}{4}}\right).$$

Proof. (The proof mimics the steps of the proof of Theorem 2.2.in [9].) Without loss of generality we may assume $\sigma = 1$ and $k = 1$, and write r instead of r_1. Let n_0 be the smallest integer such that $C^2 < \kappa^2 n_0$ (see $A.i$), and let us consider throughout the proof only $n \geq n_0$. Denote

$$A_{in}(r)=\left\{\tilde{t}\in R^p, \tilde{u}\in R : \delta_{in}(\tilde{t})+re^{n^{-\frac{1}{2}}\tilde{u}}\geq r\right\}$$

and

$$\Delta(i,n,t)=g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-g_1'(X_i,\beta^0).$$

Then we have

$$\begin{aligned}
& S_{n1}(t,u)-\mathsf{E}_P S_{n1}(t,u) \qquad (7)\\
= & \sum_{i=1}^{n}\Bigg\{\alpha_1\Delta(i,n,t)\left[I_{\left\{\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}<e_i\right\}}-1+F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right]I_{A_{in}(r)}(t,u)\\
+ & \alpha_1\Delta(i,n,t)\left[I_{\{r<e_i\}}-1+F(r)\right]I_{A^c_{in}(r)}(t,u)\\
+ & \left[\alpha_0\, g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_1 g_1'(X_i,\beta^0)\right]\\
\times & \left[I_{\left\{r\leq e_i\leq\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right\}}(t,u)-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)+F(r)\right]I_{A_{in}(r)}(t,u)\\
+ & \left[\alpha_1 g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\, g_1'(X_i,\beta^0)\right]\\
\times & \left[I_{\left\{\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\leq e_i\leq r\right\}}(t,u)-F(r)+F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right]I_{A^c_{in}(r)}(t,u)\\
+ & \alpha_0\,\Delta(i,n,t)\left[I_{\{e_i<r\}}-F(r)\right]I_{A_{in}(r)}(t,u)\\
+ & \alpha_0\,\Delta(i,n,t)\left[I_{\left\{e_i<\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right\}}-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right]I_{A^c_{in}(r)}(t,u)\Bigg\}.
\end{aligned}$$

Similarly as Jurečková, Sen [9] let us consider first of all the sum

$$\begin{aligned}
S_{n1}^{(1)}(t,u)=\sum_{i=1}^{n}\Bigg\{&\left[\alpha_0\, g_1'(X_i,\beta^0+n^{-\frac{1}{2}}t)-\alpha_1 g_1'(X_i,\beta^0)\right] \qquad (8)\\
\times&\left[I_{\left\{r\leq e_i\leq\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right\}}(t,u)-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)+F(r)\right]I_{A_{in}(r)}(t,u)
\end{aligned}$$

$$+ \quad \left[\alpha_1 g_1'(X_i,\beta^0+n^{-\frac{1}{2}}t)-\alpha_0\, g_1'(X_i,\beta^0)\right]$$

$$\times \quad \left[I_{\left\{\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\le e_i\le r\right\}}(t,u)-F(r)+F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right] I_{A^c_{in}(r)}(t,u)\Bigg\}.$$

Following [9] and [14] let us denote $\mathcal{W}=\{W(s),\ s\in R\}$ a Wiener process, and define

$$\tau_i(t,u)=\text{time for } W(s) \text{ to exit the interval } \left(c_i^{(1)},d_i^{(1)}\right),\quad i\ge 1$$

where

$$c_i^{(1)}=\min\Bigg\{\left[\alpha_0\, g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_1 g_1'(X_i,\beta^0)\right]\left[1+F(r)-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right],$$

$$\left[\alpha_0\, g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_1 g_1'(X_i,\beta^0)\right]\left[F(r)-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right]\Bigg\} I_{A_{in}(r)}(t,u)$$

$$+\min\Bigg\{\left[\alpha_1 g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\, g_1'(X_i,\beta^0)\right]\left[1+F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right],$$

$$\left[\alpha_1 g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\, g_1'(X_i,\beta^0)\right]\left[F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right]\Bigg\} I_{A^c_{in}(r)}(t,u)$$

and

$$d_i^{(1)}=\max\Bigg\{\left[\alpha_0\, g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_1 g_1'(X_i,\beta^0)\right]\left[1+F(r)-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right],$$

$$\left[\alpha_0\, g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_1 g_1'(X_i,\beta^0)\right]\left[F(r)-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right]\Bigg\} I_{A_{in}(r)}(t,u)$$

$$+\max\Bigg\{\left[\alpha_1\, g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\, g_1'(X_i,\beta^0)\right]\left[1+F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right],$$

$$\left[\alpha_1 g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\, g_1'(X_i,\beta^0)\right]\left[F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right]\Bigg\} I_{A^c_{in}(r)}(t,u).$$

Using the Skorokhod embedding of the Wiener process, we have

$$n^{-\frac{1}{4}}S_{n1}^{(1)}(t,u)=_{\mathcal{D}}\ n^{-\frac{1}{4}}W\left(\sum_{i=1}^{n}\tau_i(t,u)\right)=_{\mathcal{D}}\ W\left(n^{-\frac{1}{2}}\sum_{i=1}^{n}\tau_i(t,u)\right),\quad \forall(t,u),$$

where " $=_{\mathcal{D}}$ " denotes the equality in distribution. Since the embedding is in fact constructed on the space $(\Omega,\mathcal{B},P)$ (see e.g. [3]) it is undertood with respect to $G(x,z)$. For $\max\{\|t\|,|u|\}<C$

$$\max\left\{|\alpha_1 g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\, g_1'(X_i,\beta^0)|,\ |\alpha_0\, g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_1 g_1'(X_i,\beta^0)|\right\}$$

$$\times\left|F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right|$$

$$\le 2|\alpha_0-\alpha_1|J\cdot\left\{\left|F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F\left(re^{n^{-\frac{1}{2}}u}\right)\right|+\left|F\left(re^{n^{-\frac{1}{2}}u}\right)-F(r)\right|\right\}$$

$$\le 2n^{-\frac{1}{2}}|\alpha_0-\alpha_1|\cdot J\cdot K_1(\|t\|+|u|)\le 4n^{-\frac{1}{2}}|\alpha_0-\alpha_1|\cdot J\cdot K_1\cdot C$$

where K_1 is a positive constant. So denoting for $j = 0, 1$

$$V_{ji}(C) = \text{time for } W(s) \text{ to exit the interval } (a_{ji}, b_{ji}), \quad j = 0, 1 \tag{9}$$

with

$$a_{ji} = \min\left\{(-1)^{j+1} 4n^{-\frac{1}{2}}|\alpha_0 - \alpha_1| \cdot J \cdot K_1 \cdot C, (-1)^j 2|\alpha_0 - \alpha_1| \cdot J\right\}$$

and

$$b_{ji} = \max\left\{(-1)^{j+1} 4n^{-\frac{1}{2}}|\alpha_0 - \alpha_1| \cdot J \cdot K_1 \cdot C, (-1)^j 2|\alpha_0 - \alpha_1| \cdot J\right\},$$

we have $\tau_i(t,u) \le V_{0i}(C) + V_{1i}(C) \; \forall (i = 1, \ldots, n)$. Hence

$$\begin{aligned} & \sup\left\{\left|W\left(n^{-\frac{1}{2}}\sum_{i=1}^{n}\tau_i(t,u)\right)\right| : \max\{\|t\|, |u|\} < C\right\} \\ \le \; & \sup\left\{|W(s)| : 0 \le s \le n^{-\frac{1}{2}}\sum_{i=1}^{n}(V_{0i}(C) + V_{1i}(C))\right\}. \end{aligned}$$

Notice that while $\tau_i(t,u)$ still depends on $X_i(\omega)$, V_{0i} and V_{1i} already do not depend on it, and they are the same for all i. Using (9) we find $\mathsf{E}_P\left[n^{-\frac{1}{2}}\sum_{i=1}^{n}(V_{0i}(C)+V_{1i}(C))\right]$ $\le 8|\alpha_0 - \alpha_1| \cdot J \cdot K_1 \cdot C < K_2$ for all n starting with some n_1, where K_2 is a finite (positive) constant. Hence for a given $\varepsilon > 0$ there is a constant $T > 0$ such that (for $n \ge n_2$ and $j = 0, 1$)

$$P\left\{n^{-\frac{1}{2}}\sum_{i=1}^{n}V_{ji}(C) > T\right\} < \frac{\varepsilon}{2}.$$

Moreover, for these $\varepsilon > 0$ and $T > 0$ there is a positive constant K_3 such that

$$P\left\{\sup\{|W(s)|, \; 0 \le s \le T\} > K_3\right\} < \frac{\varepsilon}{2}$$

and hence

$$\sup\left\{n^{-\frac{1}{4}}\left|S_{n1}^{(1)}(t,u)\right| : \max\{\|t\|, |u|\} < C\right\} = O_p(1). \tag{10}$$

Now, recalling that

$$\Delta(i,n,t) = g_1'\left(X_i, \beta^0 + n^{-\frac{1}{2}}t\right) - g_1'(X_i, \beta^0)$$

and keeping in mind *A.iii*, let us write

$$\begin{aligned} |\Delta(i,n,t)| &= n^{-\frac{1}{2}}\left|\sum_{j=1}^{p} g_{1j}''(X_i, \tilde{\beta}^{(j)})t_j\right| \\ &\le n^{-\frac{1}{2}}\left\{\sum_{j=1}^{p}[g_{1j}''(X_i, \tilde{\beta}^{(j)})]^2\right\}^{\frac{1}{2}} \|t\| \le n^{-\frac{1}{2}} \cdot p^{\frac{1}{2}} \cdot J \cdot C \end{aligned} \tag{11}$$

where $\tilde{\beta}^{(j)}$ are appropriate points from the neighborhood of β^0 such that $||\tilde{\beta}^{(j)} - \beta^0|| < \kappa$. Considering now

$$S_{n1}^{(2)}(t,u) = \sum_{i=1}^{n} \alpha_1 \Delta(i,n,t) \left\{ \left[I_{\left\{\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}<e_i\right\}} - 1 \right.\right.$$
$$\left.\left. + F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right] I_{A_{in}(r)}(t,u) + \left[I_{\{r<e_i\}} - 1 + F(r)\right] I_{A_{in}^c(r)}(t,u)\right\}$$

let us put similarly as above

$$\begin{aligned} c_i^{(2)} &= \min\left\{\alpha_1\Delta(i,n,t)\left[F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-1\right],\right. \\ &\qquad \left.\alpha_1\Delta(i,n,t)F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right\} I_{A_{in}(r)}(t,u) \\ &+ \min\left\{\alpha_1\Delta(i,n,t)F(r), \alpha_1\Delta(i,n,t)\left[F(r)-1\right]\right\} I_{A_{in}^c(r)}(t,u) \end{aligned}$$

and

$$\begin{aligned} d_i^{(2)} &= \max\left\{\alpha_1\Delta(i,n,t)\left[F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-1\right],\right. \\ &\qquad \left.\alpha_1\Delta(i,n,t)F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right\} I_{A_{in}(r)}(t,u) \\ &+ \max\left\{\alpha_1\Delta(i,n,t)F(r), \alpha_1\Delta(i,n,t)\left[F(r)-1\right]\right\} I_{A_{in}^c(r)}(t,u). \end{aligned}$$

Repeating the steps from the previous part of proof and making use of (11) we obtain

$$\sup\left\{n^{-\frac{1}{4}}\left|S_{n1}^{(2)}(t,u)\right| : \max\{||t||,|u|\} < C\right\} = O_p(1). \tag{12}$$

Modifying slightly the previous lines we may also find that

$$\sup\left\{n^{-\frac{1}{4}}\left|S_{n1}^{(3)}(t,u)\right| : \max\{||t||,|u|\} < C\right\} = O_p(1) \tag{13}$$

where

$$\begin{aligned} S_{n1}^{(3)}(t,u) &= \sum_{i=1}^{n} \alpha_0\Delta(i,n,t)\left\{ \left[I_{\{e_i<r\}} - F(r)\right] I_{A_{in}(r)}(t,u) \right. \\ &+ \left.\left[I_{\left\{e_i<\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right\}} - F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right] I_{A_{in}^c(r)}(t,u)\right\}, \end{aligned}$$

and hence

$$\sup\left\{n^{-\frac{1}{4}}||S_{n1}(t,u) - \mathsf{E}_P S_{n1}(t,u)|| : \max\{||t||,|u|\} < C\right\} = O_p(1). \tag{14}$$

Now, let us estimate $n^{-\frac{1}{4}}\mathsf{E}_P S_{n1}(t,u)$. At first, let us consider

$$n^{-\frac{1}{4}}\sum_{i=1}^{n}\Delta(i,n,t)\left\{\alpha_1\left[\left[1-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right]I_{A_{in}(r)}(t,u)+[1-F(r)]I_{A^c_{in}(r)}(t,u)\right]\right.$$
$$\left.+\ \alpha_0\left[F(r)I_{A_{in}(r)}(t,u)+F(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u})I_{A^c_{in}(r)}(t,u)\right]\right\}. \quad (15)$$

which may be rewritten as

$$n^{-\frac{1}{4}}\sum_{i=1}^{n}\Delta(i,n,t)\left\{\alpha_1\left[F(r)-F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)\right]I_{A_{in}(r)}(t,u)\right.$$
$$\left.+\alpha_1(1-F(r))+\alpha_0F(r)+\alpha_0\left[F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right]I_{A^c_{in}(r)}(t,u)\right\}. \quad (16)$$

Recalling that we have assumed $\mathsf{E}_P\psi(e_i)=0$, we have

$$\mathsf{E}_P\psi(e_i)=\alpha_1\ (1-F(r))+\alpha_0\ F(r)=0. \quad (17)$$

Taking into account the assumption that the density f is bounded we easy find a positive constant K_4 such that

$$\left|F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right|\le n^{-\frac{1}{2}}\cdot K_4\cdot C,$$

which together with (17) implies that (16) (and hence also (15)) is of order $O\left(n^{-\frac{1}{4}}\right)$.

Finally, we may write

$$n^{-\frac{1}{4}}\left|\sum_{i=1}^{n}\left\{\left[\left[\alpha_1g_1'(X_i,\beta^0)-\alpha_0\ g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)\right]I_{A_{in}(r)}(t,u)\right.\right.\right.$$
$$+\quad\left.\left[\alpha_1g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\ g_1'(X_i,\beta^0)\right]I_{A^c_{in}(r)}(t,u)\right]$$
$$\times\quad\left.\left[F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F(r)\right]\right\}$$
$$-\quad\left.n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{\gamma_1g_1'(X_i,\beta^0)\cdot\left[g'(X_i,\beta^0)\right]^{\mathrm{T}}t+\gamma_2g_1'(X_i,\beta^0)\,u\right\}\right|$$
$$=\quad n^{-\frac{1}{4}}\left|\sum_{i=1}^{n}\left\{\left[\left[\alpha_1g_1'(X_i,\beta^0)-\alpha_0\ g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)\right]I_{A_{in}(r)}(t,u)\right.\right.\right.$$
$$+\quad\left.\left[\alpha_1g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)-\alpha_0\ g_1'(X_i,\beta^0)\right]I_{A^c_{in}(r)}(t,u)\right]$$
$$\times\quad\left.\left[F\left(\delta_{in}(t)+re^{n^{-\frac{1}{2}}u}\right)-F\left(re^{n^{-\frac{1}{2}}u}\right)-n^{-\frac{1}{2}}\left[g'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)\right]^{\mathrm{T}}tf\left(re^{n^{-\frac{1}{2}}u}\right)\right]\right\}$$

$$+ \; n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{ \left[\left[\alpha_1 g_1'(X_i, \beta^0) - \alpha_0 \; g_1' \left(X_i, \beta^0 + n^{-\frac{1}{2}} t \right) \right] I_{A_{in}(r)}(t,u) \right. \right.$$

$$\left. + \; \left[\alpha_1 g_1' \left(X_i, \beta^0 + n^{-\frac{1}{2}} t \right) - \alpha_0 \; g_1'(X_i, \beta^0) \right] I_{A^c_{in}(r)}(t,u) \right] \left[g'(X_i, \beta^0 + n^{-\frac{1}{2}} t) \right]^{\mathrm{T}} t$$

$$\left. - \; (\alpha_1 - \alpha_0) \, g_1'(X_i, \beta^0) \, [g_1'(X_i, \beta^0)]^{\mathrm{T}} t \right\} \cdot f \left(r e^{n^{-\frac{1}{2}} u} \right)$$

$$+ \; n^{-\frac{1}{2}} (\alpha_1 - \alpha_0) \sum_{i=1}^{n} g_1'(X_i, \beta^0) \, [g'(X_i, \beta^0)]^{\mathrm{T}} t \left\{ f \left(r e^{n^{-\frac{1}{2}} u} \right) - f(r) \right\}$$

$$+ \; \sum_{i=1}^{n} (\alpha_0 + \alpha_1) \left\{ g_1' \left(X_i, \beta^0 + n^{-\frac{1}{2}} t \right) - g_1'(X_i, \beta^0) \right\} I_{A^c_{in(r)}}(t,u) \left[F \left(r e^{n^{-\frac{1}{2}} u} \right) - F(r) \right]$$

$$+ \; \alpha_1 \sum_{i=1}^{n} \left[g_1'(X_i, \beta^0) - g_1' \left(X_i, \beta^0 + n^{-\frac{1}{2}} t \right) \right] \left\{ F(r e^{n^{-\frac{1}{2}} u}) - F(r) - n^{-\frac{1}{2}} r u f(r) \right\}$$

$$\left. + \; (\alpha_1 - \alpha_0) \sum_{i=1}^{n} g_1'(X_i, \beta^0) \left\{ F \left(r e^{n^{-\frac{1}{2}} u} \right) - F(r) - n^{-\frac{1}{2}} r u f(r) \right\} \right|$$

$$= \sum_{j=1}^{6} A_j .$$

Since

$$F(a + n^{-\frac{1}{2}} b) - F(a) - n^{-\frac{1}{2}} b f(a) = \int_0^{n^{-\frac{1}{2}} b} [f(a+t) - f(a)] \, \mathrm{d}t,$$

using *A.i* and *A.ii*, and the assumption that f as well as f' are bounded, we find that uniformly in $\max\{\|t\|, |u|\} < C$ for $j = 1, \ldots, 6$ we have $|A_j| < K_5 n^{-\frac{1}{4}}$ where K_5 does not depend on X_i's. Finally, keeping in mind that $\|g'(X_i, \beta)\| < p^{\frac{1}{2}} J$, which implies that the variances $\mathrm{var}_P g_i'(X_i, \beta)$ and $\mathrm{var}_P [g_i'(X_i, \beta) \, g_j'(X_i, \beta)]$ are finite and using the Lindeberg-Feller central limit theorem we find that

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{ g'(X_i, \beta^0) \cdot \left[g'(X_i, \beta^0) \right]^{\mathrm{T}} - Q \right\} = O_p(1)$$

as well as

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} \{ g'(X_i, \beta^0) - q \} = O_p(1)$$

which concludes the proof. □

4.2. Absolutely continuous ψ-function with ψ' step-function

Let $\psi'(z) = \alpha_j$ for $z \in (r_j, r_{j+1}]$, $j = 0, 1, \ldots, k-1$, and $\psi'(z) = \alpha_k$ for $z \in (r_k, \infty)$ where $\alpha_0 = -\infty < \alpha_1 < \ldots < \alpha_k$. Following Jurečková and Sen let us change from here the meaning of γ_1 and γ_2 so that

$$\gamma_1 = \mathsf{E}_P \left\{ \sigma^{-1} \psi'(e_1 \sigma^{-1}) \right\} \qquad \text{and} \qquad \gamma_2 = \mathsf{E}_P \left\{ (e_1 \sigma^{-1}) \, \psi'(e_1 \sigma^{-1}) \right\}.$$

Theorem 2. Let Conditions A hold. Moreover, let F have a bounded derivative f in neighborhoods of the points $r_1, r_2, \ldots, r_k$. Then for any $C > 0$

$$\sup\left\{\left\|S_n(t,u) + n^{\frac{1}{2}}\left[\gamma_1 Qt + \gamma_2 qu\right]\right\| : \max\{\|t\|, |u|\} < C\right\} = O_p(1). \qquad (18)$$

Proof. (The proof again mimics the steps of the proof of Theorem 3.2 of [9].) Without any loss of generality let us assume that there are just two steps of $\psi'(t)$ (due to fact that we have assumed that ψ is bounded we cannot assume only one step), let us say at the points r_1 and r_2. Let n_0 be the smallest integer such that $C < n_0 \cdot |r_1 - r_0|$, and hereafter consider only $n > n_0$. First of all observe that

$$\psi\left(\left[e_i - \delta_{in}(t)\right] e^{-n^{-\frac{1}{2}}u}\right) = \psi\left(\left(1 - n^{-\frac{1}{2}}u\right) e_i - \delta_{in}(t)\right)$$
$$+ \left\{e_i\left(e^{-n^{-\frac{1}{2}}u} - 1 + n^{-\frac{1}{2}}u\right) + \delta_{in}(t)\left(1 - e^{-n^{-\frac{1}{2}}u}\right)\right\}\xi_i \qquad (19)$$

where ξ_i lies between $\psi'\left(\left[e_i - \delta_{in}(t)\right] e^{-n^{-\frac{1}{2}}u}\right)$ and $\psi'\left(\left[1 - n^{-\frac{1}{2}}u\right]e_i - \delta_{in}(t)\right)$. Since $\left|e^{-n^{-\frac{1}{2}}u} - 1 + n^{-\frac{1}{2}}u\right| \le 2n^{-1}C^2$ (for $|u| < C$) and $\left|\delta_{in}(t)\left(1 - e^{-n^{-\frac{1}{2}}u}\right)\right| \le K_6 \cdot n^{-1}C$ for some positive constant K_6 and $\|t\| < C$, there is a positive constant K_7 such that

$$\sup\left\{\left|\sum_{i=1}^n g_1'\left(X_i, \beta^0 + n^{-\frac{1}{2}}t\right)\left[\psi\left(\left[e_i - \delta_{in}(t)\right]e^{-n^{-\frac{1}{2}}u}\right) - \psi\left(\left[1 - n^{-\frac{1}{2}}u\right]e_i - \delta_{in}(t)\right)\right]\right| :\right.$$
$$\left.\max\{\|t\|, |u|\} < C\right\} \le K_7 n^{-1}\sum_{i=1}^n \{|e_i| + 1\}, \qquad (20)$$

and hence using the Markov law of large numbers we find that (20) is of order $O_p(1)$. So, let us consider instead of sum (2) the sum

$$S_n^M(t,u) = \sum_{i=1}^n \left[\psi\left(\left[1 - n^{-\frac{1}{2}}u\right] e_i - \delta_{in}(t)\right) g'\left(X_i, \beta^0 + n^{-\frac{1}{2}}t\right) - \psi(e_i) g'(X_i, \beta^0)\right]$$

and let us denote

$$\mathcal{H}_{n1}(t,u) = \left\{i \in N : \left[\min\left\{(1 - n^{-\frac{1}{2}}u)e_i - \delta_{in}(t), (1 - n^{-\frac{1}{2}}u)e_i, e_i\right\},\right.\right.$$
$$\left.\left.\max\left\{(1 - n^{-\frac{1}{2}}u)e_i - \delta_{in}(t), (1 - n^{-\frac{1}{2}}u)e_i, e_i\right\}\right] \cap \{r_1, r_2\} \ne \emptyset\right\},$$
$$\mathcal{H}_{n2}(t,u) = \{1, 2, \ldots, n\} \setminus \mathcal{H}_{n1}(t,u).$$

Moreover denote the ith element of the sum $S_n^M(t,u)$ by $s_{in}(t,u)$ and its kth coordinate by $s_{in(k)}(t,u)$. Then we may write

$$S_n^M(t,u) = \sum_{i=1}^n s_{in}(t,u)\, I_{\{i \in \mathcal{H}_{n1}\}} + \sum_{i=1}^n s_{in}(t,u)\, I_{\{i \in \mathcal{H}_{n2}\}}. \qquad (21)$$

Denote moreover for any $t \in R^p$ $z_t^{(\ell)} = (t_1, t_2, \ldots, t_{\ell-1}, z, 0, \ldots, 0)^{\mathrm{T}}$. Then due to the fact that

$$s_{in(k)}(t,u) = -n^{-\frac{1}{2}} \left\{ \int_0^u \psi'([1-n^{-\frac{1}{2}}v]e_i) \cdot g_k'(X_i, \beta^0)\, \mathrm{d}v \right.$$
$$+ \sum_{\ell=1}^{p} \int_0^{t_\ell} \left\{ \psi'\left(\left[1-n^{-\frac{1}{2}}u\right] e_i - \delta_{in}(z_t^{(\ell)})\right) g_k'\left(X_i, \beta^0 + n^{-\frac{1}{2}} z_t^{(\ell)}\right) g_\ell'\left(X_i, \beta^0 + n^{-\frac{1}{2}} z_t^{(\ell)}\right) \right.$$
$$\left. \left. - \psi\left(\left[1-n^{-\frac{1}{2}}u\right] e_i - \delta_{in}(z_t^{(\ell)})\right) g_{k\ell}''\left(X_i, \beta^0 + n^{-\frac{1}{2}} z_t^{(\ell)}\right) \right\} \mathrm{d}z \right\}, \tag{22}$$

we find that for $1 \le i \le n$ $|s_{in(k)}| \le n^{-\frac{1}{2}} K_8(|e_i| + 1)$, where $K_8 < \infty$ depends of course on C. In a similar way we find that, starting with some n_3, there is $K_9 < \infty$ such that $\{i \in \mathcal{H}_{n1}(t,u)\} \subset D_{in}$, where $D_{in} = \left\{ \left\{ r_1 - n^{-\frac{1}{2}} K_9 \le e_i \le r_1 + n^{-\frac{1}{2}} K_9 \right\} \right.$ $\left. \cup \left\{ r_2 - n^{-\frac{1}{2}} K_9 \le e_i \le r_2 + n^{-\frac{1}{2}} K_9 \right\} \right\}$ and $P(D_{in}) = O\left(n^{-\frac{1}{2}}\right)$. It implies that

$$\mathsf{E}_P \sup \left\{ \left| \sum_{i=1}^n s_{in(k)}(t,u) I_{\{i \in \mathcal{H}_{n1}\}} \right| : \max\{\|t\|, |u|\} < C \right\}$$
$$\le \mathsf{E}_P \left\{ \mathsf{E}_P \sup \sum_{i=1}^n \left\{ |s_{in(k)}(t,u)\, I_{\{i \in \mathcal{H}_{n1}\}}| : \max\{\|t\|, |u|\} < C \;\middle|\; X_i = x_i \right\} \right\}$$
$$\le n^{-\frac{1}{2}} \mathsf{E}_P \left\{ \mathsf{E}_P \left\{ \sum_{i=1}^n K_8(|e_i| + 1) I_{D_{in}} \right\} \right\} = O(1),$$

and the Chebyshev inequality (for nonnegative random variable) applied on $n^{-\frac{1}{2}} \sum_{i=1}^n K_8(|e_i|+1)\, I_{D_{in}}$ gives that the first sum of the right-hand-side of (21) is bounded in probability. Now, denoting the first coordinate of the second sum of the right-hand-side of (21) by S_{n1}^{M2}, we have for any pair (t_1, u_1) and (t_2, u_2) of distinct points

$$\mathrm{var}_G \left(S_{n1}^{M2}(t_1,u_1) - S_{n1}^{M2}(t_2,u_2)\right) \tag{23}$$
$$\le 2 \sum_{i=1}^n \mathsf{E}_P \left\{ \psi\left(\left[1 - n^{-\frac{1}{2}} u_1\right] e_i - \delta_{in}(t_1)\right) g_k'\left(X_i, \beta^0 + n^{-\frac{1}{2}} t_1\right) \right.$$
$$\left. - \psi\left(\left[1 - n^{-\frac{1}{2}} u_2\right] e_i - \delta_{in}(t_2)\right) g_1'\left(X_i, \beta^0 + n^{-\frac{1}{2}} t_2\right) \right\}^2 I_{\{i \in \mathcal{H}_{n2}\}}.$$

Let us denote for $\ell = 1, 2, \ldots, p$ $z_{t_1,t_2}^{(\ell)} = (t_{21}, t_{22}, \ldots t_{2\,\ell-1}, z, t_{1\,\ell+1}, \ldots, t_{1\,p})^{\mathrm{T}}$. Again due to the absolute continuity of ψ we may write

$$\psi\left(\left[1 - n^{-\frac{1}{2}} u_1\right] e_i - \delta_{in}(t_1)\right) - \psi\left(\left[1 - n^{-\frac{1}{2}} u_2\right] e_i - \delta_{in}(t_1)\right)$$
$$= e_i n^{-\frac{1}{2}} \int_{u_1}^{u_2} \psi'\left(\left[1 - n^{-\frac{1}{2}} u\right] e_i - \delta_{in}(t_1)\right) \mathrm{d}u$$

and

$$\psi\left(\left[1-n^{-\frac{1}{2}}u_2\right]e_i-\delta_{in}(t_1)\right)g_k'\left(X_i,\beta^0+n^{-\frac{1}{2}}t_1\right)$$
$$-\psi\left(\left[1-n^{-\frac{1}{2}}u_2\right]e_i-\delta_{in}(t_2)\right)g_k'\left(X_i,\beta^0+n^{-\frac{1}{2}}t_2\right)$$
$$=-n^{-\frac{1}{2}}\sum_{\ell=1}^{p}\int_{t_{1\ell}}^{t_{2\ell}}\left\{\psi'\left(\left[1-n^{-\frac{1}{2}}u_2\right]e_i-\delta_{in}(z_{t_1,t_2}^{(\ell)})\right)\right.$$
$$\times g_k'\left(X_i,\beta^0+n^{-\frac{1}{2}}z_{t_1,t_2}^{(\ell)}\right)g_\ell'\left(X_i,\beta^0+n^{-\frac{1}{2}}z_{t_1,t_2}^{(\ell)}\right)$$
$$\left.-\psi\left(\left[1-n^{-\frac{1}{2}}u_2\right]e_i-\delta_{in}(z_{t_1,t_2}^{(\ell)})\right)g_{k\ell}''\left(X_i,\beta^0+n^{-\frac{1}{2}}z_{t_1,t_2}^{(\ell)}\right)\right\}\mathrm{d}z$$

and hence we may bound (23) by

$$K_{10}\left\{(u_1-u_2)^2+\|t_1-t_2\|^2\right\},\quad 0<K_{10}<\infty,$$

uniformly in

$$\max\left\{\|t_1\|,\|t_2\|,|u_1|,|u_2|\right\}<C.$$

Using analogous steps as above we derive that also uniformly in $\max\{\|t_1\|,\|t_2\|,|u_1|,|u_2|\}<C$

$$\left|\mathsf{E}_P(S_{n1}^{M2}(t_1,u_1)-S_{n1}^{M2}(t_2,u_2))+n^{-\frac{1}{2}}\sum_{i=1}^{n}\{\gamma_1g_1'(X_i,\beta^0)[g'(X_i,\beta^0)]^{\mathrm{T}}(t_1-t_2)\right.$$
$$\left.+\quad\gamma_2g_1'(X_i,\beta^0)(u_1-u_2)\right\}I_{\{i\in\mathcal{H}_{n2}\}}\Bigg|\le K_{11}\{\|t_1-t_2\|+|u_1-u_2|\},\ 0<K_{11}<\infty.$$

Earlier than proceeding further let us make following linearization (let $g_1''(X_i,\beta)$ denotes the first column of $g''(X_i,\beta)$):

$$\psi\left(\left[1-n^{-\frac{1}{2}}u\right]e_i-\delta_{in}(t)\right)g_1'\left(X_i,\beta^0+n^{-\frac{1}{2}}t\right)I_{\{i\in\mathcal{H}_{n2}\}}$$
$$=\quad\left\{\psi\left(\left[1-n^{-\frac{1}{2}}u\right]e_i\right)g_1'(X_i,\beta^0)\right.$$
$$-\quad n^{-\frac{1}{2}}\left\{\psi'\left(\left[1-n^{-\frac{1}{2}}u\right]e_i\right)g_1'(X_i,\beta^0)\left[g'(X_i,\beta^0)\right]^{\mathrm{T}}\right.$$
$$-\quad\left.\psi\left(\left[1-n^{-\frac{1}{2}}u\right]e_i\right)\left[g_1''(X_i,\beta^0)\right]^{\mathrm{T}}\right\}t+n^{-1}\left\{\psi''(\tilde{e}_i)\left[\left[g'(X_i,\tilde{\beta})\right]^{\mathrm{T}}t\right]^2g_1'(X_i,\tilde{\beta})\right.$$
$$-\quad\left.\psi'(\tilde{e}_i)\,t^{\mathrm{T}}g_1''(X_i,\beta^0)\left[g'(X_i,\tilde{\beta})\right]^{\mathrm{T}}t+\psi'(\tilde{e}_i)t^{\mathrm{T}}g_1''(X_i,\tilde{\beta})\left[g'(X_i,\tilde{\beta})\right]^{\mathrm{T}}t\right\}$$
$$+\quad\left.n^{-\frac{1}{2}}\psi\left(\left[1-n^{-\frac{1}{2}}u\right]e_i\right)\left\{g_1''(X_i,\tilde{\tilde{\beta}})-g_1''(X_i,\beta^0)\right\}^{\mathrm{T}}t\right\}I_{\{i\in\mathcal{H}_{n2}\}}\qquad(24)$$

where $\tilde{e}_i$ lies between $\left(1-n^{-\frac{1}{2}}u\right)e_i-\delta_{in}(t)$ and $(1-n^{-\frac{1}{2}}u)e_i$ (remember that for $i\in\mathcal{H}_{n2}$ between $\left(1-n^{-\frac{1}{2}}u\right)e_i-\delta_{in}(t)$ and $\left(1-n^{-\frac{1}{2}}u\right)e_i$ we have continuous – in fact constant – derivative of ψ) and $\|\tilde{\beta} - \beta^0\| < n^{-\frac{1}{2}}t$ as well as $\|\tilde{\tilde{\beta}} - \beta^0\| < n^{-\frac{1}{2}}t$. Due to the fact that $\psi'' \equiv 0$ and all other derivatives are assumed bounded and g'' Lipschitz, the supremum (over $\max\{\|t\|, |u|\} < C$) of the sum (over $1 \le i \le n$) of the last terms of (24) is $O_p(1)$. Similarly, the sum of the elements of order n^{-1} is of course also $O_p(1)$. Hence it remains to cope with the first three terms of the right-hand side of the last expression. Taking into account that $\mathsf{E}_P\psi(e_1) = 0$, we may write them as

$$\left\{\psi(e_i)\, g_1'(X_i,\beta^0) - n^{-\frac{1}{2}}\psi'(e_i)\, g_1'(X_i,\beta^0)\left\{e_i u + \left[g'(X_i,\beta^0)\right]^{\mathrm{T}} t\right\} + R_{in}\right\} I_{\{i\in\mathcal{H}_{n2}\}}$$

where $\sup\left\{\left|\sum_{i=1}^n R_{in} I_{\{i\in\mathcal{H}_{n2}\}}\right| : \max\{\|t\|, |u|\} < C\right\} = O_p(1)$. So we may substitute the process $S_n^M(t,u)$ by the process

$$\tilde{S}_n(t,u) = -n^{-\frac{1}{2}}\sum_{i=1}^n \left\{\psi'(e_i)\left\{e_i u + \left[g'(X_i,\beta^0)\right]^{\mathrm{T}} t\right\} g'(X_i,\beta^0)\right\} I_{\{i\in\mathcal{H}_{n2}\}}.$$

and we have $\sup\left\{\|S_n^M(t,u) - \tilde{S}_n(t,u)\| : \max\{\|t\|, |u|\} < C\right\} = O_p(1)$. By this step we have modified the original processes $S_n(t,u)$ so that the new processes $\tilde{S}_n(t,u)$ is equivalent to processes treated for the linear model (see [9]; the only difference is that instead of the ith row of the design matrix, say X_i^{T}, we have $[g'(X_i,\beta^0)]^{\mathrm{T}}$). Now, the problem is that the processes $\tilde{S}_n(t,u)$ may not vanish along the lower boundary and hence, to be able to use e.g. result of Bickel and Wichura [2], the following reparametrization is necessary. Let us denote by

$$\tilde{\tilde{S}}_n(t,u) = \tilde{S}_n(t,u) + n^{-\frac{1}{2}}\sum_{i=1}^n g_1'(X_i,\beta^0)\left\{\gamma_1\left[g'(X_i,\beta^0\right]^{\mathrm{T}} t + \gamma_2 u\right\}$$

and $\mathcal{D} = \{E = \mathrm{Diag}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{p+1}) : (\varepsilon_k \in \{0,1\}, k = 1, 2, \ldots, p+1)\}$. For any $E \in \mathcal{D}$ let $E^{(p)}$ be the main submatrix of dimension p and put

$$S_n^*(t,u) = \sum_{E\in\mathcal{D}} (-1)^{\mathrm{Tr}(E)}\quad \tilde{\tilde{S}}_n((\mathcal{I} - E^{(p)})\, t, (1-\varepsilon_{p+1})\, u)$$

where $\mathcal{I}$ denotes the identity matrix. At this point we have reached the full coincidence (even in the notation) with the proof of Theorem 3.2 of [9] (see (3.12)) and the rest of the proof coincides with a part of the proof of Theorem 3.2 of [9]. Hence it will be omitted. □

4.3. Absolutely continuous ψ-function with absolutely continuous derivative

Put for any $\delta > 0$

$$\psi_\delta''(y) = \sup\{|\psi''(y+z)| : |z| \le \delta\}$$

and

$$\overline{\psi}_\delta''(y) = \sup\{|\psi''(\exp(w)(y+z))| : \{|z|, |w|\} \le \delta\}.$$

Theorem 3. Let Conditions A hold. Moreover, let for some $\delta_0 > 0$ and $\nu > 1$

$$\mathsf{E}_P\left\{|t\overline{\psi}''_{\delta_0}(t)|^\nu\right\} < \infty \qquad \text{and} \qquad \mathsf{E}_P\left\{|t^2\psi''_{\delta_0}(t)|^\nu\right\} < \infty \tag{25}$$

for all $\delta \in (0, \delta_0]$ and γ_1 as well as γ_2 are finite. Then

$$\sup\left\{\left\|S_n(t,u) + n^{\frac{1}{2}}\left[\gamma_1 Qt + \gamma_2 qu\right]\right\| : \max\left\{\|t\|, |u|\right\} < C\right\} = O_p(1). \tag{26}$$

The proof can be again carry out by mimicing the steps of the proof of Theorem 4.2 of [9]. We hope that the proofs of the two previous theorems have demonstrated that modifications of corresponding technique from the linear framework to the nonlinear one is straightforward. That is why in this case, where due to existence of all derivatives the modifications are simpler than above, we omit the proof.

5. UTILIZATION OF THE ASYMPTOTIC LINEARITY FOR THE DISCONTINUOUS ψ FUNCTIONS

As it was already indicated above, there is still some other problem. To be able to apply the asymptotic linearity of the M-statistics on the M-estimator $\hat\beta^{(\psi,n)}$, we need to know something about the behavior of the sum

$$\sum_{i=1}^n \psi\left(\frac{Y_i - g(X_i, \hat\beta^{(\psi,n)})}{\hat\sigma^{(n)}}\right) g'(X_i, \hat\beta^{(\psi,n)}).$$

It is clear that when the derivative of the function ρ exist everywhere and the M-estimator is defined by (3) we have

$$\sum_{i=1}^n \psi\left(\frac{Y_i - g(X_i, \hat\beta^{(\psi,n)})}{\hat\sigma^{(n)}}\right) g'(X_i, \hat\beta^{(\psi,n)}) = 0. \tag{27}$$

Generally it does not hold for non-smooth ρ-functions derivative of which is discontinuous. A possible solution of the problem may be as follows. Rao and Zhao [15] proved consistency of the estimator given by equation

$$S_n^{-1}\sum_{i=1}^n \psi\left(Y_i - X_i^{\mathrm{T}}\beta^{(n,\sigma)}\right) X_i = o_p(1) \tag{28}$$

(where S_n is an estimator of the scatter matrix) for nondecreasing ψ (without the assumption of continuity). Nevertheles a modification for the standardized version would be necessary (and then of course we would have again to utilize considerations which we employed above, applying the lemma of the Appendix).

Moreover Jurečková and Welsh [10] proved for increasing ψ-step-function the $\sqrt{n}$-consistency of the M-estimator defined through the equation

$$n^{-\frac{1}{2}}\sum_{i=1}^n \psi\left(\frac{Y_i - X_i^{\mathrm{T}}\beta}{\hat\sigma^{(n)}}\right) X_i = O_p\left(n^{-\frac{1}{4}}\right) \tag{29}$$

(see Theorem 4.2 of [9]; of course (29) is somewhat stronger than (28) but the result is also stronger).

On the other hand to be able to apply the asymptotic linearity of the M-statistics on the M-estimator of the parameters of the nonlinear models we do not need (27) but it is sufficient to know that

$$\sum_{i=1}^{n} \psi\left(\frac{Y_i - g(X_i, \hat{\beta}^{(\psi,n)})}{\hat{\sigma}^{(n)}}\right) g'(X_i, \hat{\beta}^{(\psi,n)}) = o_p\left(n^{\frac{1}{2}}\right). \tag{30}$$

which is a standardized nonlinear version of (28).

But let us try to carry out directly some very first consideration clarifying this problem. For the ψ-function which is not continuous, general conditions under which (30) is fulfilled are not known, although for some discontinuos ψ-functions, e.g. for ψ_{med} (given by

$$\psi_{\text{med}} = \begin{array}{lll} -1 & \text{if} & x < 0, \\ 0 & \text{if} & x = 0, \\ 1 & \text{if} & x > 0) \end{array}$$

we may reach again even precise equality in (30) – under some conditions for symmetry of $g'(X_i, \beta)$ without which it seems questionable to use ψ_{med}.

To create an idea about the problem let us look at first on the much simpler case of estimating location parameter in the case when the central model is assumed to be the standard normal one. After all, in other cases, under assumptions which was used in Huber's paper [6], namely that $-\log\frac{f'(x)}{f(x)}$ is strictly convex, we may for theoretical considerations assume that we transform random variables to the normal ones. Let us assume that we shall use skipped Huber's ψ-function $\psi_H(x)$, i.e. $\psi_H(x) = -\psi_H(-x)$ and

$$\psi_H(x) = \begin{array}{lll} x & \text{if} & x \in [0, a], \\ a & \text{if} & x \in (a, b], \\ 0 & \text{if} & x > b \end{array}$$

for some $0 < a < b < \infty$. Let $Y_{(1)} \le Y_{(2)} \le \cdots \le Y_{(n)}$ be our observation (in fact we may assume $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ because if any sharp inequality is distorted the (absolute) continuity, is questionable; from the similar reasons we have also $Y_{(i)} - Y_{(j)} \ne 2b$ for $i, j = 1, 2, \ldots, n$ a.e. for any $n \in N$). Now, let us observe that for $t \in (-\infty, Y_{(1)} - b) \cup (Y_{(n)} + b, \infty)$ we have $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t) = 0$.

Since for any $n \in N$, any $\omega \in \Omega$ and $t \in (-\infty, Y_{(1)}(\omega)) \cup (Y_{(n)}(\omega), \infty)$ (31) holds, it is clear that we may obtain inconsistent solution of (31) (below). In other words, for strongly redescending ψ-function (regardless continuous or discontinuous) among the estimators given by (26) is at least one inconsistent. Nevertheless for $t = Y_{(1)} - b$ we obtain $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t) = a$ and for $t = Y_{(n)} + b$ we finally get $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t) = -a$. Moreover, $\sum_{i=1}^{n} \psi_H(Y_{(i)} - t)$ is continuous (and nonincreasing) in t except for a finite number of discontinuities, at which it has the positive jumps equal to a. It implies that there is at least one point $\hat{t}^{(n)} \in (Y_{(1)} - b, Y_{(n)} + b)$ such that

$$\sum_{i=1}^{n} \psi_H\left(Y_{(i)} - \hat{t}^{(n)}\right) = 0. \tag{31}$$

We may observe that the reason why for ψ_H we are able to fulfil (31) is a "compensation" of the jump(s) by a decrease of the value of the terms which have argument in the linear part of the ψ-function.

It is easy to see that the point(s) which solves (31) is a (local) minimum of the function $\sum_{i=1}^n \rho(Y_{(i)} - t)$ because $-\psi_H(y-t)$ is increasing in t. Moreover at any point t^* of jump of $\frac{\partial}{\partial t}\sum_{i=1}^n \rho(Y_{(i)} - t)$ we have

$$\lim_{t\to t^*_-} \frac{\partial}{\partial t}\sum_{i=1}^n \rho(Y_{(i)} - t) > \lim_{t\to t^*_+} \frac{\partial}{\partial t}\sum_{i=1}^n \rho(Y_{(i)} - t)$$

so that the function $\sum_{i=1}^n \rho(Y_{(i)} - t)$ either increases when $t \to t^*_-$, and then for $t > t^*$ either decreases or increases less steeply, or decreases for $t \to t^*_-$, and then for $t > t^*$ it decreases more steeply. Anyway, the function $\sum_{i=1}^n \rho(Y_{(i)} - t)$ cannot have at t^* minimum. So we may concludes that the global minimum is among the points for which (31) holds.

More detailed analysis would reveal that a similar situation holds for many ψ-functions, namely that we may hope to fulfill

$$\sum_{i=1}^n \psi(Y_i - t) = o_p\left(n^{\frac{1}{2}}\right)$$

for rather large family of ψ-functions.

Some difficulties may appear e. g. for skipped median (or for some other estimators with both types of jumps).

Let us now consider the linear regression. We would want again to show that there is a point $\hat\beta^{(n)}$ such that

$$\sum_{i=1}^n \psi_H\left(Y_{(i)} - X_i^{\mathrm{T}}\hat\beta^{(n)}\right) X_i = 0.$$

Let us consider at first $\sum_{i=1}^n \rho_H(Y_i - L\cdot X_i^{\mathrm{T}}\gamma)$ for $\|\gamma\| = 1$. We easy verify that for any γ

$$-\frac{\partial}{\partial L}\sum_{i=1}^n \rho_H\left(Y_i - L\cdot X_i^{\mathrm{T}}\gamma\right) = \sum_{i=1}^n \psi_H\left(Y_i - L\cdot X_i^{\mathrm{T}}\gamma\right) X_i^{\mathrm{T}}\gamma$$

is nonincreasing in L (except of finite number ℓ ($\ell \le 2n$) of possitive jumps), and along similar lines as above we again find that there is $L_\gamma^{(1)} < 0$ such that for $L < L_\gamma^{(1)}$ we have $\sum_{i=1}^n \psi_H(Y_i - L\cdot X_i^{\mathrm{T}}\gamma)\, X_i^{\mathrm{T}}\gamma = 0$ and

$$\sum_{i=1}^n \psi_H\left(Y_i - L_\gamma^{(1)}\cdot X_i^{\mathrm{T}}\gamma\right) X_i^{\mathrm{T}}\gamma = \sum_{i\in\mathcal{I}_\gamma^{(1)}} |X_i^{\mathrm{T}}\gamma|\cdot a$$

where $\mathcal{I}_\gamma^{(1)} = \left\{i \in N : \mathrm{sign}(X_i^{\mathrm{T}}\gamma)\,\psi_H(Y_i - L_\gamma^{(1)}\cdot X_i^{\mathrm{T}}\gamma) = a\right\}$. Similarly we may find an upper "bound" L_2. Then there is again at least one $L_\gamma^* \in (L_\gamma^{(1)}, L_\gamma^{(2)})$ such that

$$\sum_{i=1}^n \psi_H\left(Y_i - L_\gamma^*\cdot X_i^{\mathrm{T}}\gamma\right) X_i^{\mathrm{T}}\gamma = 0. \tag{32}$$

Due to similar arguments as above we find that at one of these points (if they are multiple) the function $\sum_{i=1}^{n} \rho_H(Y_i - L \cdot X_i^{\mathrm{T}}\gamma)$, as the function of L, attains its minimum, and that the points $Y_i - L_\gamma^* \cdot X_i^{\mathrm{T}}\gamma$, $i = 1, 2, \ldots, n$ are not points of discontinuity of the function ψ_H. Let $\rho_0 = \inf_{\|\gamma\|=1} \sum_{i=1}^{n} \rho_H(Y_i - L_\gamma^* \cdot X_i^{\mathrm{T}}\gamma)$. Taking into account the compactness of the surface of unit ball we find that there is a $\gamma_0, \|\gamma_0\| = 1$ such that

$$\rho_0 = \sum_{i=1}^{n} \rho_H \left(Y_i - L_{\gamma_0}^* \cdot X_i^{\mathrm{T}} \gamma_0 \right).$$

Let us recall that the points $Y_i - L_{\gamma_0}^* \cdot X_i^{\mathrm{T}}\gamma_0$, $i = 1, 2, \ldots, n$ are not the points of discontinuity of the function ψ_H, i.e. in the neighborhood of the point $\hat{\beta}^{(n)} = L_{\gamma_0}^* \gamma_0$ the function $\sum_{i=1}^{n} \rho_H(Y_i - X_i^{\mathrm{T}}\beta)$ has (continuous) partial derivatives, and hence

$$\sum_{i=1}^{n} \psi_H \left(Y_i - X_i^{\mathrm{T}} \hat{\beta}^{(n)} \right) X_i^{\mathrm{T}} = 0.$$

Of course, for the nonlinear setup it is more complicated to describe the situation because it depends on mutual relations of ψ and g. E.g. considering again $\psi_H(z)$ we may find that for the function g which is for any fix X coordinatewise increasing and convex or coordinatewise decreasing and concave in β we have again

$$\sum_{i=1}^{n} \psi_H \left(Y_i - g(X_i, L\gamma) \right) \sum_{\ell=1}^{p} g_\ell'(X_i, L\gamma)^{\mathrm{T}} \gamma = 0$$

nonincreasing in L and the considerations which we made above might be repeated. But in such a case we may probably cope with the problem even without the convexity (or concavity) of function, just reparametrizing the problem to the linear one (due to the monotonicity).

Nevertheless, the set of conditions covering all possibilities of the "compensation" for the regression setup would be rather complicated. So that, one may only hope that for some ψ-function one can recognize whether the "compensation" which was described above is possible or not. On the other hand, the discontinuous ψ-functions may not only imply infite local shift sensitivity but also they may have infinite change-of-variance sensitivity (see [5]; consult also [18]). Some recent results moreover indicate that the change of (the norm of) estimate when excluding some observations may be rather large for them (although asymptotically bounded in probability), while for continuous function it is proportional to the gross error sensitivity ([22]). It implies that the importance of the discontinuous ψ-functions for robust estimation is limited.

APPENDIX

Lemma 7. Let for some $p \in N$, $\{\mathcal{V}^{(n)}\}_{n=1}^{\infty}$, $\mathcal{V}^{(n)} = \left\{v_{ij}^{(n)}\right\}_{i=1,2,\dots,p}^{j=1,2,\dots,p}$ be a sequence of $(p \times p)$ matrixes such that for $i = 1, 2, \dots, p$ and $j = 1, 2, \dots, p$

$$\lim_{n\to\infty} v_{ij}^{(n)} = q_{ij} \qquad \text{in probability} \tag{33}$$

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

$$\exists\,(\varepsilon > 0)\ \forall\ (K > 0)\ \limsup_{n\to\infty} P\left(\|\theta^{(n)}\| > K\right) > \varepsilon. \tag{34}$$

Then so that

$$\exists\,(\delta > 0) \quad \forall\,(L > 0)$$

$$\limsup_{n\to\infty}\ P\left(\left\|\mathcal{V}^{(n)}\theta^{(n)}\right\| > L\right) > \delta.$$

Proof. Due to (33) the matrix $\mathcal{V}^{(n)}$ is regular in probability. Let then $0 < \lambda_{1n} < \lambda_{2n} < \dots < \lambda_{pn}$ and $z_{1n}, z_{2n}, \dots, z_{pn}$ be eigenvalues and corresponding eigenvectors (selected to be mutually orthogonal) of the matrix $[\mathcal{V}^{(n)}]^{\mathrm{T}}\mathcal{V}^{(n)}$. Let us write $\theta^{(n)} = \sum_{j=1}^{p} a_{jn} z_{jn}$ (for an appropriate vector $a_n = (a_{1n}, a_{1n}, \dots, a_{pn})^{\mathrm{T}}$). Then we have

$$\left\|\mathcal{V}^{(n)}\theta^{(n)}\right\|^2 = \sum_{j=1}^{p}[a_{jn}]^2\lambda_{jn}\|z_{jn}\|^2 \le \lambda_{1n}\|\theta^{(n)}\|. \tag{35}$$

Moreover, denoting λ_1 the smallest eigenvalue of the matrix $Q^{\mathrm{T}}Q$, we have $\lambda_{1n} \to \lambda_1$ in probability as $n \to \infty$. The assertion of the lemma then follows from (35). □

ACKNOWLEDGEMENT

We would like to express our gratitude to our colleagues Jana Jurečková and Pavel Charamza for the carefull reading of the manuscript and the valuable suggestions which helped to remove a gap in the proof of Theorem 1 and also led to the general improvement of the paper. Especially, those parts which treat problems of $\sqrt{n}$-consistency, standartization and the applicability of the asymptotic linearity for the discontinuous ψ-functions were included as a result of the discussion with Jana Jurečková.

(Received October 21, 1993.)

REFERENCES

[1] D. M. Bates and D. G. Watts: Nonlinear Regression Analysis and Its Applications. J. Wiley & Sons, New York 1988.

[2] P J. Bickel and M. J. Wichura: Convergence criteria for multiparameter stochastic processes and some applications. Ann. Math. Statist. *42* (1971), 1656–1670.

[3] M. Csörgő and P. Révész: Strong Approximation in Probability and Statistics. Akademia Kiadó, Budapest 1981.
[4] P. L. Davis: Aspects of robust linear regression. Ann. Statist. *21* (1993), 1843–1899.
[5] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw and W. A. Stahel: Robust Statistics – The Approach Based on Influence Functions. J. Wiley & Sons, New York 1986.
[6] P. J. Huber: Robust estimation of a location parameter. Ann. Math. Statist. *35* (1964), 73–101.
[7] J. Jurečková: Consistency of M-estimators in linear model generated by non-monotone and discontinuous ψ-functions. Probab. Math. Statist. *10* (1988), 1–10.
[8] J. Jurečková and B. Procházka: Regression quantiles and trimmed least squares estimator in nonlinear regression model. Nonparametric Statist. *3* (1994), 201–222.
[9] J. Jurečková and P. K. Sen: Uniform second order asymptotic linearity of M-statistics in linear models. Statist. Decisions 7 (1989), 263–276.
[10] J. Jurečková and A. H. Welsh: Asymptotic relations between L- and M-estimators in the linear model. Ann. Inst. Statist. Math. *42* (1990), 671–698.
[11] F. Liese and I. Vajda: Consistency of M-estimators in general models. J. Multivariate Anal. *50* (1994), 93–114.
[12] A. Marazzi: Algorithms, Routines and S Functions for Robust Statistics. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California 1992.
[13] J.M. Ortega and W.C. Rheinboldt: Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, New York and London 1970.
[14] S. Portnoy: Tightness of the sequence of empiric c.d.f. processes defined from regression fractiles. In: Robust and Nonlinear Time-Series Analysis (J. Franke, W. Härdle, D. Martin, eds.), Springer–Verlag, New York, 1983, pp. 231–246.
[15] C.R. Rao and L. C. Zhao: On the consistency of M-estimate in linear model obtained through an estimating equation. Statist. Probab. Lett. *14* (1992), 79–84.
[16] A. Rubio, L. Aguilar, and J. Á. Víšek: Testing for difference between models. Comput. Statist. *8* (1992), 57–70.
[17] A. Rubio, F. Quintana and J. Á. Víšek: Test for differences of M-estimates between nonlinear regression models. Probab. Math. Statist. *14* (1993), 2, 189–206.
[18] J. Á. Víšek: Stability of regression models estimates with respect to subsamples. Computat. Statist. 7 (1992), 183–203.
[19] J. Á. Víšek: Problems connected with selection of robust procedure. In: Proceedings of PROBASTAT'91 (A. Pázman and J. Volaufová, eds.), Printing House of the Technical University of Liptovský Mikuláš 1992, pp. 189–203.
[20] J. Á. Víšek: On the role of contamination level and the least favorable behaviour of gross-error sensitivity. Probab. Math. Statist. *14* (1993), 2, 173–187.
[21] J. Á. Víšek: A cautionary note on the method of Least Median of Squares reconsidered. In: Transactions of the Twelfth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes (J. Á. Víšek and P. Lachout, eds.), Prague 1994, pp. 254–259.
[22] J. Á. Víšek: Sensitivity analysis of M-estimates. Ann. Inst. Statist. Math., to appear.

Dr. Asunción Mária Rubio, Departamento de Matemáticas, Universidad de Extremadura, 10071 Cáceres. Spain.

RNDr. Jan Ámos Víšek, CSc., Ústav teorie informace a automatizace AV ČR (Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic), Pod vodárenskou věží 4, 182 08 Praha 8 & Katedra makroekonomie a financí, Institut ekonomických studií fakulty sociálních věd, Karlova universita (Department of Macroeconomics and Finance, Institute of Economic Studies, Faculty of Social Sciences, Charles University), Smetanovo nábřeží 6, 110 01 Praha 1. Czech Republic.