Ann. Inst. Statist. Math.
Vol. 54, No. 2, 261–290 (2002)

SENSITIVITY ANALYSIS OF M-ESTIMATES OF NONLINEAR REGRESSION MODEL: INFLUENCE OF DATA SUBSETS*

JAN ÁMOS VÍŠEK

Department of Stochastic Informatics, Institute of Information Theory and Automation, Academy of Sciences, the Czech Republic and
Department of Macroeconomics, Institute of Economic Studies, Charles University, Opletalova ulice 26, CZ-110 01 Prague 1, the Czech Republic, e-mail: visek@mbox.fsv.cuni.cz

(Received November 16, 1998; revised September 18, 2000)

Abstract. Asymptotic representations of the difference of M-estimators of the parameters of nonlinear regression model for the full data and for the subsample of data are given for the following three situation: i) fix number of points excluded from data, ii) increasing number, however asymptotically negligible part of data excluded, and finally iii) asymptotically fix portion of data excluded. Asymptotic normality of the difference of estimators (for the two latter cases) is proved.

Key words and phrases: Sensitivity analysis, asymptotic representation, asymptotic distribution of the difference of M-estimators, testing subsample stability of estimates, scale invariance, diagnostics of subsets of influential points, diversity of estimates.

1. Introduction

An analysis of the influence of individual datum or of data subsets on the results of any data-processing procedure has an eminent importance for the applications. Therefore in any theory oriented on data-processing, a part of it has been always devoted to this topic. In regression analysis one may find an amount of references to the papers treating this problem for instance in the monographs about the sensitivity analysis of estimation by Atkinson (1985), Belsley *et al.* (1980), Chatterjee and Hadi (1988) or Rousseeuw and Leroy (1987), to give at least some of them.

In the linear regression there is a well-known formula for the difference of the least squares estimators for the full data set and for a subset of data containing $n-1$ observations, namely

$$\hat{\beta}_{\mathrm{LS}}^{(n-1,\ell)} - \hat{\beta}_{\mathrm{LS}}^{(n)} = -\{[X^{(n-1,\ell)}]^{\mathrm{T}} X^{(n-1,\ell)}\}^{-1} X_\ell (Y_\ell - X_\ell^{\mathrm{T}} \hat{\beta}_{\mathrm{LS}}^{(n)}) \tag{1.1}$$

where notation is nearly selfexplaining, nevertheless, $X^{(n-1,\ell)}$ is the design matrix after deletion of the ℓ-th row from the full design matrix X and X_ℓ is the ℓ-th row (considered as a column vector) of the design matrix for the full data (see e.g. Chatterjee and Hadi (1988), Víšek (1992a) or Zvára (1989)).

The present paper derives, in a form of asymptotic representations of Bahadur type, analogical formulas for the M-estimators of nonlinear regression model for the three situations, namely when a fix number of observations is excluded from the data set, when an

*Research was supported by grant of GA UK number 255/2000/A EK /FSV.

increasing but asymptotically negligible number of observations is excluded and finally, when an asymptotically fix portion of observations is excluded. The representations for the two latter cases hint the asymptotic normality of the respective difference, and so they allow to establish a test of subsample stability of estimates in the sense of Víšek (1992a). We shall see later that it may help to select the most adequate model for given data in the case when various methods give considerably different estimates.

Let us start with notations.

Notations. Let N denote the set of all positive integers, R the real line, R^+ its positive part and $(\Omega, \mathcal{B}, P)$ a probability space. We shall consider for all $n \in N$ the nonlinear regression model

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

where $\{X_n\}_{n=1}^{\infty}$ is a fix sequence of vectors from R^q, $\beta^0 \in R^p$ and $\{e_n\}_{n=1}^{\infty}$, $e_n : \Omega \to R$ is a sequence of independent and identically distributed random variables (i.i.d.r.v.) with $\mathbb{E}e_1 = 0$ and $\mathbb{E}e_1^2 = \sigma^2 \in (0, \infty)$. $F(z)$ will denote the distribution function (d.f.) of $e_i \cdot \sigma^{-1}$, respectively. Finally, having denoted $I_k = \{i_1, i_2, \dots, i_k; 1 \le i_1 < i_2 < \cdots < i_k\} \subset N$, let us define

$$\hat{\beta}^{(n)} = \arg\min_{\beta \in R^p} \sum_{i=1}^{n} \rho([Y_i - g(X_i, \beta)]\hat{\sigma}_n^{-1}) \tag{1.2}$$

and

$$\hat{\beta}^{(n,I_k)} = \arg\min_{\beta \in R^p} \sum_{i \in \{1,2,\dots,n\} \setminus I_k} \rho([Y_i - g(X_i, \beta)]\hat{\sigma}_n^{-1}) \tag{1.3}$$

where ρ is an absolutely continuous function (with a derivative ψ) and $\hat{\sigma}_n$ is a preliminary estimator of the scale of residuals. $\hat{\sigma}_n$ is assumed to be regression-invariant and scale-equivariant in order to achieve regression- and scale-equivariance of $\hat{\beta}^{(n)}$ (see in Bickel (1975) or Jurečková and Sen (1993) and condition C.iii below). To simplify all considerations we will assume that the same estimate of the scale will be used for the full data set and for the "reduced" data. This assumption does not represent substantial restriction of generality either from the theoretical point of view or for the applications. In former case it only burden the notations by some additional items and prolongs the proofs of theorems. In the latter case it asks for the employment of robust estimators of scale and for a check whether after deletion the change of scale estimate is not dramatic. If however instability of the scale estimate occurs, we may expect the same instability of the estimate of regression model. In such a case we shall probably prefer another estimator of regression model anyway, see discussion below. Let us give now conditions under which we will derive the results.

2. Conditions

CONDITIONS A. (i) There is a positive δ_0 such that for any $\beta \in R^p$, $\|\beta - \beta^0\| < \delta_0$

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

exist for any $x \in \{X_n\}_{n=1}^{\infty}$. Let us denote the vector of the first partial derivative and the matrix of the second derivatives 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) The functions $g''_{jk}(x,\beta)(j,k = 1,2,\ldots,p)$ are uniformly in $x \in \{X_n\}_{n=1}^{\infty}$ Lipschitz of the first order in β in the δ_0-neighborhood of β^0, i.e.

$$\exists (L > 0)\ \forall (\beta \in R^p, \|\beta - \beta^0\| < \delta_0)$$
$$\max_{1\le j,k\le p} \sup_{x\in\{X_n\}_{n=1}^{\infty}} |g''_{jk}(x,\beta) - g''_{jk}(x,\beta^0)| < L \cdot \|\beta - \beta^0\|.$$

Moreover

$$\max_{1\le j,k\le p} \sup_{x\in\{X_n\}_{n=1}^{\infty}} \max\{|g(x,\beta^0)|, |g'_j(x,\beta^0)|, |g''_{jk}(x,\beta^0)|\} < \infty.$$

(iii) There is a regular matrix Q and a vector $q \in R^p$ such that

$$\frac{1}{n}\sum_{i=1}^{n} g'(X_i,\beta^0)[g'(X_i,\beta^0)]^{\mathrm{T}} = Q + \mathcal{O}(n^{-1/4})$$

and

$$\frac{1}{n}\sum_{i=1}^{n} g'(X_i,\beta^0) = q + \mathcal{O}(n^{-1/4}),$$

and denote $(Q)_{ij} = q_{ij}$.

Remark 2.1. In what follows we shall use the fact that A(ii) implies that there is $J < \infty$ such that

$$\max_{1\le j,k\le p} \sup_{x\in\{X_n\}_{n=1}^{\infty},\beta\in R^p,\|\beta-\beta^0\|<\delta_0} \max\{|g(x,\beta)|, |g'_j(x,\beta)|, |g''_{jk}(x,\beta)|\} < J. \tag{2.1}$$

Another consequence of A(iii) is that the matrix Q is positive definite.

Remark 2.2. Of course, to fulfill Conditions A we have to verify that the Lipschitz property holds on a subset of R^p, in which we assume, according to a priori knowledge, that β_0 lies. On the other hand, the most of nonlinear models $g(x,\beta)$ which can be taken into account to be able to solve efficiently the corresponding extremal problem defining the M-estimator, would be (sufficiently) smooth. Remark of the same spirit is true about uniformity in x.

Remark 2.3. From the proof of assertions given below it will be clear that the results of paper hold also for linear model and it will be also clear how conditions is to be modified (genearally weakened).

CONDITIONS B. (i) The function ψ allows decomposition in the form

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

where ψ_a has a derivative ψ'_a which is Lipschitz of the first order, ψ_c is a continuous function with derivative ψ'_c which is step- function and ψ_s is a step-function itself. Let us denote by $D_1 = \{r_{1,1}, r_{1,2}, \ldots, r_{1,s_1}\}$, ($s_1$ finite) and $D_2 = \{r_{2,1}, r_{2,2}, \ldots, r_{2,s_2}\}$, (again s_2 finite) the points of jumps of ψ_s and of ψ'_c, respectively.

(ii) $\sigma^2 = \mathrm{var}_F e_i \in (0, \infty)$ and there is a positive ϑ_0 such that $F(z)$ has a density f which is bounded on $D_2(\vartheta_0) = \cup_{i=1}^{s_2}[\sigma \cdot r_{2,i} - \vartheta_0, \sigma \cdot r_{2,i} + \vartheta_0]$ and which is Lipschitz of the first order on $D_1(\vartheta_0) = \cup_{i=1}^{s_1}[\sigma \cdot r_{1,i} - \vartheta_0, \sigma \cdot r_{1,i} + \vartheta_0]$. Let us denote by $H < \infty$ an upper bound of f on $D_2(\vartheta_0)$ as well as the corresponding Lipschitz constant of f on $D_1(\vartheta_0)$.

(iii) There is a finite Θ such that $\sup_{z \in R} |\psi(z)| < \Theta$ as well as $\sup_{z \in R \setminus \{D_1 \cup D_2\}} |\psi'(z)| < \Theta$.

(iv) $I\!E_F \psi(e_1 \cdot \sigma^{-1}) = 0$, $\mathrm{var}_F \psi(e_1 \cdot \sigma^{-1}) \in (0, \infty)$

$$\gamma = \sigma^{-1} I\!E_F \psi'(e_1 \cdot \sigma^{-1}) + \sum_{k=1}^{s_1} f(r_{1,k}\sigma)[\psi(r_{1,k}+) - \psi(r_{1,k}-)] > 0 \tag{2.2}$$

where $\psi(r_{1,k}+) = \lim_{z \searrow r_{1,k}} \psi(z)$ and $\psi(r_{1,k}-) = \lim_{z \nearrow r_{1,k}} \psi(z)$, and

$$\theta = \sigma^{-1} I\!E_F e_1 \psi'(e_1 \cdot \sigma^{-1}) + \sum_{j=1}^{s_1} r_{1,j}[\psi_s(r_{1,j}+) - \psi_s(r_{1,j}-)] f(r_{1,j}\sigma) \tag{2.3}$$

exists and is finite. Put for any positive ζ

$$\bar{\psi}''_{a,\zeta}(z) = \sup_{|v|<\zeta, |u|<\zeta} |\psi''_a(e^v(z+u))|$$

(where subindex a indicates that we take into account the second derivative of absolutely continuous part of ψ and ζ hints that the supremum is taken over the interval $(-\zeta, \zeta)$). There is ζ_0 so that for all $0 < \zeta < \zeta_0$ $I\!E_F \bar{\psi}''_{a,\zeta}(e_1) < \infty$, $I\!E_F\{e_1 \bar{\psi}''_{a,\zeta}(e_1)\} < \infty$ and $I\!E_F\{e_1^2 \bar{\psi}''_{a,\zeta}(e_1)\} < \infty$.

Remark 2.4. First of all, notice please that for $\psi = \psi_a + \psi_c$ the second term of (2.2) is equal to zero while for $\psi = \psi_s$ the first term vanishes. Similarly for (2.3).

Conditions B cover the most of the ψ-functions which are used in the robust statistics; for a discussion see Hampel *et al.* (1986). As we shall see in Remark 2.5 (below), in the case when $\psi_s \not\equiv 0$, we need to restrict the range of possible ψ-functions further by (2.4) and (2.5). On the other hand, specifying $I_{k_n} = \{\ell\}$ (for some $1 \le \ell \le n$) it is clear from Víšek (1996*a*) or (1997*b*) that for the case when $\psi_s \not\equiv 0$ the norm of $n(\hat{\beta}^{(n)} - \hat{\beta}^{(n,I_{k_n})})$, although still asymptotically bounded in probability, may be much larger than for the case $\psi_s \equiv 0$ (for numerical examples which confirm this see also Víšek (1996*a*)). It implies that in the applications we will probably prefer ψ-functions without jumps. Of course, in the following theoretical discussions we would like to treat the problem in question for the largest set of ψ-functions for which we are able to do it. Hence the case $\psi_s \not\equiv 0$ is considered, too.

Remark 2.5. Let us consider the case when $\psi_s \equiv 0$. It is clear that under given conditions the estimators $\hat{\beta}^{(n)}$ and $\hat{\beta}^{(n,I_k)}$ given in (1.2) and (1.3) fulfill the equations

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

and

$$\sum_{i \in \{1,2,\ldots,n\} \setminus I_k} \psi([Y_i - g(X_i, \hat{\beta}^{(n,I_k)})]\hat{\sigma}_n^{-1}) g'(X_i, \hat{\beta}^{(n,I_k)}) = 0.$$

When however $\psi_s \not\equiv 0$ the situation is a little more complicated. A simple consideration however shows that for some discontinuous ψ-functions we may nevertheless expect that $\hat{\beta}^{(n)}$ and $\hat{\beta}^{(n,I_k)}$ fulfill

$$\sum_{i=1}^{n} \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1}) g'(X_i, \hat{\beta}^{(n)}) = o_p(1) \tag{2.4}$$

and

$$\sum_{i \in \{1,2,\ldots,n\} \setminus I_k} \psi([Y_i - g(X_i, \hat{\beta}^{(n,I_k)})]\hat{\sigma}_n^{-1}) g'(X_i, \hat{\beta}^{(n,I_k)}) = o_p(1). \tag{2.5}$$

On the other hand, conditions under which (2.4) and (2.5) are fulfilled appear to be generally rather complicated since they have to cover all possible mutual relations of ψ and g. The discussion in Rubio and Víšek (1996) enlightens the problem and we may hope that at least for some ψ-functions one can recognize whether (2.4) and (2.5) hold. For instance, one possibility when (2.4) and (2.5) may be reached is the case when the jump in one part of ψ is "compensated" by (a large number of) small changes of values of $\psi(Y_i - g(X_i, \beta))$ in a strictly monotone part of ψ . Hence the following conditions will simply assume that (2.4) and (2.5) are fulfilled.

CONDITIONS C. Let $\{k_n\}_{n=1}^{\infty}$ be a nondecreasing sequence of positive integers and let us denote for any $b \in R^p$ by $g(X, b)$ the vector $(g(X_1, b), g(X_2, b), \ldots, g(X_n, b))^T$.

(i) There is a $\sqrt{n}$-consistent estimator $\hat{\sigma}_n = \hat{\sigma}_n(Y, X)$ of σ, i.e. $\sqrt{n}(\hat{\sigma}_n - \sigma) = \mathcal{O}_p(1)$ which is regression-invariant, sometimes this property is called affine-invariant, i.e. for any $b \in R^p$

$$\hat{\sigma}_n(Y + g(X, b), X) = \hat{\sigma}_n(Y, X)$$

and scale-equivariant, i.e. for any $c > 0$

$$\hat{\sigma}_n(cY, X) = c \cdot \hat{\sigma}_n(cY, X).$$

(ii) The estimators $\hat{\beta}^{(n)}$ and $\hat{\beta}^{(n,I_{k_n})}$ fulfill (2.4) and (2.5), respectively.

(iii) The estimator $\hat{\beta}^{(n)}$ is $\sqrt{n}$-consistent in the usual sense and $\hat{\beta}^{(n,I_{k_n})}$ is $\sqrt{n}$-consistent in the following sense

$$\forall(\varepsilon > 0)\ \exists(K > 0 \text{ and } n_\varepsilon \in N)\ \forall(n \geq n_\varepsilon \text{ and } I_{k_n})$$
$$P(\sqrt{n}\|\hat{\beta}^{(n,I_{k_n})} - \beta^0\|\hat{\sigma}_n^{-1} > K) < \varepsilon.$$

Remark 2.6. It will be clear that the form of the $\sqrt{n}$- consistency required in C(iii) is for the cases which will be studied below only slightly more demanding than the usual definition of the $\sqrt{n}$-consistency. The reader who is interested in the conditions for the consistency of M-estimators for the nonlinear models may find them in Jurečková and Procházka (1994) or Liese and Vajda (1994) where also an extensive discussion of the topic is presented. Moreover, in Rubio *et al.* (1993) it is proved that under the Conditions A and B the consistency implies the $\sqrt{n}$-consistency. Finally, in Rubio and

Víšek (1996) it is shown that under Conditions A and B for the case when $\psi_s \equiv 0$ there is a $\sqrt{n}$-consistent solution of

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

Let us recall once again that what concerns (necessity of) the scale-equivariance and regression- invariance of the scale estimate for studentization of residuals, one can find more details in Bickel (1975) or Jurečková and Sen (1993). For a proposal of scale-equivariant and regression-invariant scale estimator see Víšek (1999).

3. Preliminaries

First of all, let us denote for $t, u \in R^p$ and $\tau \in [0, \frac{1}{2}]$

$$\delta_{in}(\tau,t,u) = g(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u) - g(X_i,\beta^0). \tag{3.1}$$

In the case when $u = 0$ we shall write $\delta_{in}(\tau, t)$ instead of $\delta_{in}(\tau, t, 0)$ and when even $\tau = 0$, we shall write simply $\delta_{in}(t)$.

LEMMA 3.1. *Let Conditions* A *and* B *be fulfilled and let us put for any* $M > 0$

$$\mathcal{S}_M = \{t, u \in R^p, v \in R^+ : \max\{\|t\|, \|u\|, v\} \leq M\}.$$

Then for the case when $\psi_s \equiv 0$ *there is a sequence of random matrices* $\{\mathcal{U}_n\}_{n=1}^{\infty}$ *such that* $\mathcal{U}_n = o_p(1)$ *and we have for* $\tau \in [0, \frac{1}{2}]$

$$\sup_{\mathcal{S}_M} \left\| \sum_{i=1}^{n} \{\psi([e_i - \delta_{in}(\tau,t,u)]\sigma^{-1}e^{-n^{-1/2}v})g'(X_i, \beta^0 + n^{-1/2}t + n^{-1/2-\tau}u) \right.$$
$$- \psi([e_i - \delta_{in}(\tau,t)]\sigma^{-1}e^{-n^{-1/2}v})g'(X_i,\beta^0 + n^{-1/2}t)\}$$
$$\left. + n^{1/2-\tau}[\sigma^{-1}I\!\!E_F\psi'(e_1 \cdot \sigma^{-1})Q + \mathcal{U}_n]u \right\| = \mathcal{O}_p(n^{-\tau}) \qquad as \quad n \to \infty. \tag{3.2}$$

For the case $\psi_s \not\equiv 0$ *for any* $\tau \in [0, \frac{1}{2}]$ *there exist a family of Wiener processes* $\mathcal{W}_j = \mathcal{W}_j(y)$, *and sequences of stopping times* $\mu_{ijn}(\tau,t,u,v)$ *and of random variables* $\kappa_{jkn}(\tau)$, *and a sequence of random processes* $\mathcal{K}_{jn}(\tau,t,u,v)$ *(where* $j, k = 1,2,\ldots,p$, $y \in R^+$, $i = 1,2,\ldots,n$, $n \in N$, $t,u,v \in \mathcal{S}_M$*) so that*

$$\max_{1\leq j,k\leq p} |\kappa_{jkn}(\tau)| = o_p(1), \tag{3.3}$$

$$\max_{1\leq j\leq p} \sup_{\mathcal{S}_M} |\mathcal{K}_{jn}(\tau,t,u,v)| = \mathcal{O}_p(n^{-\tau}), \tag{3.4}$$

$$\max_{1\leq j\leq p} \sup_{\mathcal{S}_M} \|u\|^{-1} \sum_{i=1}^{n} \mu_{ijn}(\tau,t,u,v) = \mathcal{O}_p(n^{1/2-\tau}) \qquad as \quad n \to \infty \tag{3.5}$$

and for $t,u,v \in \mathcal{S}_M$ *and* $j = 1,2,\ldots,p$

$$\sum_{i=1}^{n} \{\psi([e_i - \delta_{in}(\tau,t,u)]\sigma^{-1}e^{-n^{-1/2}v})g'_j(X_i,\beta^0 + n^{-1/2}t + n^{-1/2-\tau}u) \tag{3.6}$$

$$-\psi([e_i - \delta_{in}(\tau,t)]\sigma^{-1}e^{-n^{-1/2}v})g_j'(X_i,\beta^0 + n^{-1/2}t)\}$$
$$+n^{1/2-\tau}\gamma\sum_{k=1}^{p}[q_{jk} + \kappa_{jkn}(\tau)]u_k + \mathcal{K}_{jn}(\tau,t,u,v) =_{\mathcal{D}} W_j\left(\sum_{i=1}^{n}\mu_{ijn}(\tau,t,u,v)\right)$$

where "$=_{\mathcal{D}}$" denotes the equality in distribution.

PROOF. For the proof see Lemmas 1, 2 and 3 of Víšek (1996*a*) and (1997*b*). □

Remark 3.1. Let us recall that the stopping times as well as the Wiener processes given in Lemma 1 may be defined on a probability space $(\tilde{\Omega}, \tilde{\mathcal{B}}, \tilde{P})$ different from the space $(\Omega, \mathcal{B}, P)$. On the other hand, using e.g. Csörgö and Révész (1981), Theorem 2.1.2, we can modify (3.2) and (3.6) so that they can be written simultaneously for $\psi_s \equiv 0$ and $\psi_s \not\equiv 0$. Since we shall need such a form of this assertion later, let us do it in the next lemma.

LEMMA 3.2. *Let Conditions* A *and* B *hold. Then for any $\tau \in [0, \frac{1}{2}]$ there are random variables $\kappa_{jkn}(\tau)$ and random processes $\mathcal{K}_{jn}(\tau,t,u,v)$ fulfilling* (3.3) *and* (3.4), *respectively, and random processes $\Delta^*_{jkn}(\tau,t,u,v)$ with $j = 1,2,\ldots,p$, $n \in N$, $t,u \in R^p$, such that*

$$\max_{1\le j\le p}\sup_{S_M}|\Delta^*_{jkn}(\tau,t,u,v)| = o_p(1) \quad \text{as} \quad n\to\infty \tag{3.7}$$

and

$$\max_{1\le j\le p}\sup_{S_M}\Bigg|\sum_{i=1}^{n}\{\psi([e_i - \delta_{in}(\tau,t,u)]\sigma^{-1}e^{-n^{-1/2}v})g_j'(X_i,\beta^0 + n^{-1/2}t + n^{-1/2-\tau}u) \tag{3.8}$$
$$- \psi([e_i - \delta_{in}(\tau,t)]\sigma^{-1}e^{-n^{-1/2}v})g_j'(X_i,\beta^0 + n^{-1/2}t)\}$$
$$+n^{1/2-\tau}\sum_{k=1}^{p}[\gamma(q_{jk} + \kappa_{jkn}(\tau)) + \Delta^*_{jkn}(\tau,t,u,v)]u_k + \mathcal{K}_{jn}(\tau,t,u,v)\Bigg|$$
$$= \mathcal{O}_p(1).$$

PROOF. Let us assume that $\psi_s \not\equiv 0$ (for the case $\psi_s \equiv 0$ the proof is simpler). We are going to use Lemma 3.1. Let us fix $\tau \in [0, \frac{1}{2}]$, say τ_0, and some $\Delta > 0$ and $\varepsilon > 0$. Using (3.5) let us find $K_1 < \infty$ and $n_1 \in N$ so that for the stopping times from Lemma 3.1 we have

$$\tilde{P}\left(\max_{1\le j\le p}\sup_{S_M}\|u\|^{-1}\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v) > n^{1/2-\tau_0}K_1\right) < \varepsilon.$$

Further, applying the strong law of large numbers for Wiener process (see e.g. Breiman (1968), Proposition 12.31 or Štěpán (1987), Theorem VII.1.5), let us establish $K_2 < \infty$ so that for the Wiener processes we have

$$\tilde{P}\left(\max_{1\le j\le p}\sup_{y>K_2}\frac{|W_j(y)|}{y} > \Delta\cdot K_1^{-1}\right) < \varepsilon.$$

Using e.g. the law of iterated logarithm (Csörgö and Révész (1981), Theorem 1.3.1) let us select $K_3 < \infty$ so that

$$\tilde{P}(\max_{1\le j\le p}\sup_{0\le y\le K_2}|W_j(y)| > K_3) < \varepsilon.$$

By Lemma 3.1, there exist sequences of random variables $\{\kappa_{jkn}(\tau_0)\}_{n=1}^{\infty}$ and of random processes $\{\mathcal{K}_{jn}(\tau_0,t,u,v)\}_{n=1}^{\infty}$ j, $k=1,2,\dots,p$, satisfying (3.3) and (3.4) and

$$\sum_{i=1}^{n}\{\psi([e_i-\delta_{in}(\tau_0,t,u)]\sigma^{-1}e^{-n^{-1/2}v})g_j'(X_i,\beta^0+n^{-1/2}t+n^{-1/2-\tau_0}u)$$
$$-\psi([e_i-\delta_{in}(\tau_0,t)]\sigma^{-1}e^{-n^{-1/2}v})g_j'(X_i,\beta^0+n^{-1/2}t)\}$$
$$+n^{1/2-\tau_0}\gamma\sum_{k=1}^{p}[q_{jk}+\kappa_{jkn}(\tau_0)]u_k+\mathcal{K}_{jn}(\tau_0,t,u,v)=_{\mathcal{D}}W_j\left(\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\right).$$

Now, let us denote

$$B_{1n}(\tau_0)=\left\{\tilde{\omega}:\max_{1\le j\le p}\sup_{\mathcal{S}_M}\|u\|^{-1}\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)>n^{1/2-\tau_0}K_1\right\},\tag{3.9}$$

$$B_{2n}=\left\{\tilde{\omega}:\max_{1\le j\le p}\sup_{y>K_2}\frac{|W_j(y)|}{y}>\Delta\cdot K_1^{-1}\right\}\tag{3.10}$$

and

$$B_{3n}=\left\{\tilde{\omega}:\max_{1\le j\le p}\sup_{0\le y\le K_2}|W_j(y)|>K_3\right\}.\tag{3.11}$$

Denoting $D_n=B_{1n}^c\cap B_{2n}^c\cap B_{3n}^c$ we have for any $n>n_2$

$$\tilde{P}(D_n)\ge 1-3\varepsilon.\tag{3.12}$$

Consider a fix (but arbitrary) $n>n_2$ and also a fix (but also arbitrary) $\tilde{\omega}\in D_n$, and assume at first that at this $\tilde{\omega}$ we have

$$\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)>K_2.$$

Then, taking into account at first (3.10) and then (3.9), we arrive at

$$\left|W_j\left(\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\right)\right|\le\Delta\cdot K_1^{-1}\cdot\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\le n^{1/2-\tau_0}\cdot\Delta\|u\|.\tag{3.13}$$

For the case when $\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\le K_2$ we have

$$\left|W_j\left(\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\right)\right|\le K_3\tag{3.14}$$

(see (3.11)). Since either (3.13) or (3.14) has to take place we have for our $n>n_2$ and $\tilde{\omega}\in D_n$ and for any $t,u,v\in\mathcal{S}_M$

$$\left|W_j\left(\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\right)\right|\le n^{1/2-\tau_0}\cdot\Delta\|u\|+K_3,$$

i.e. for our $n>n_2$ and $\tilde{\omega}\in D_n$ and for any $t,u,v\in\mathcal{S}_M$ we have

$$W_j\left(\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\right)-n^{1/2-\tau_0}\cdot\Delta\|u\|\le K_3\tag{3.15}$$

and also

$$W_j\left(\sum_{i=1}^{n}\mu_{ijn}(\tau_0,t,u,v)\right)+n^{1/2-\tau_0}\cdot\Delta\|u\|\geq -K_3. \tag{3.16}$$

Now, let us define for $j,k=1,2,\dots,p$, $n\in N$ and given $\tilde{\omega}\in D_n$

$$\begin{aligned}\Delta^*_{jkn}(\tau_0,t,u,v) = -\Delta\cdot\|u\|^{-1}\cdot\operatorname{sign}\Bigg\{\Bigg(&\sum_{i=1}^{n}\{\psi([e_i-\delta_{in}(\tau_0,t,u)]\sigma^{-1}e^{-n^{-1/2}v})\\ &\times g_j'(X_i,\beta^0+n^{-1/2}t+n^{-1/2-\tau_0}u)\\ &-\psi([e_i-\delta_{in}(\tau_0,t)]\sigma^{-1}e^{-n^{-1/2}v})\\ &\times g_j'(X_i,\beta^0+n^{-1/2}t)\}\\ &+n^{1/2-\tau_0}\gamma\sum_{\ell=1}^{p}[q_{j\ell}+\kappa_{j\ell n}(\tau_0)]u_\ell\\ &+\mathcal{K}_{jn}(\tau_0,t,u,v)\Bigg)\Bigg\}u_k.\end{aligned} \tag{3.17}$$

Notice that $\|\Delta^*_{jkn}(\tau_0,t,u,v)\|\leq\Delta$. Since in the considerations made a few line above the number n was an arbitrary integer larger than n_2 and the point $\tilde{\omega}$ was an arbitrary point from D_n, we have from (3.12), (3.15) and (3.16)

$$\begin{aligned}P\Bigg(\max_{1\leq j\leq p}\sup_{S_M}\Bigg|&\sum_{i=1}^{n}\{\psi([e_i-\delta_{in}(\tau_0,t,u)]\sigma^{-1}e^{-n^{-1/2}v})\\ &\times g_j'(X_i,\beta^0+n^{-1/2}t+n^{-1/2-\tau_0}u)-\psi([e_i-\delta_{in}(\tau_0,t)]\sigma^{-1}e^{-n^{-1/2}v})\\ &\times g_j'(X_i,\beta^0+n^{-1/2}t)\}+n^{1/2-\tau_0}\sum_{k=1}^{p}[\gamma(q_{jk}+\kappa_{jkn}(\tau_0))\\ &+\Delta^*_{jkn}(\tau_0,t,u,v)]u_k+\mathcal{K}_{jn}(\tau_0,t,u,v)\Bigg|>K_3\Bigg)\leq 3\varepsilon\end{aligned} \tag{3.18}$$

and (3.8) is evidently fulfilled. Since $\frac{u_k}{\|u\|}\leq 1$ we have from (3.17) $|\Delta^*_{jkn}(\tau_0,t,u,v)|\leq\Delta$. On the other hand, Δ was an arbitrary positive number (see the second line of the proof). Hence (3.7) holds, too. □

Now let us recall that we have introduced $\delta_{in}(t)=g(X_i,\beta^0+n^{-1/2}t)-g(X_i,\beta^0)$ (see (3.1) and the lines below it). In what follows we shall need also the following lemma.

LEMMA 3.3. *Let Conditions* A, B *and* C *hold. Then*

$$\begin{aligned}\sup_{S_M}\Bigg\|n^{-1/2}&\sum_{i=1}^{n}\{\psi([e_i-\delta_{in}(t)]\sigma^{-1}e^{-n^{-1/2}v})g'(X_i,\beta^0+n^{-1/2}t)\\ &-\psi(e_i\cdot\sigma^{-1})g'(X_i,\beta^0)\}+\gamma Qt+\theta qv\Bigg\|=o_p(1)\end{aligned} \tag{3.19}$$

where similarly as above

$$\mathcal{S}_M = \{t \in R^p, v \in R, \max\{\|t\|, |v|\} < M\}.$$

PROOF. Without loss of generality let us assume that $\sigma = 1$ and $M > 1$. At first we shall consider the case when $\psi = \psi_a$ and we shall show that

$$(3.20)\qquad \sup_{\mathcal{S}_M} \left\| \sum_{i=1}^{n} \{\psi_a([e_i - \delta_{in}(t)]\sigma^{-1} e^{-n^{-1/2}v}) g'(X_i, \beta^0 + n^{-1/2}t) - \psi_a(e_i \cdot \sigma^{-1}) g'(X_i, \beta^0)\} + n^{1/2}\gamma Q t + n^{1/2}\theta q v \right\| = \mathcal{O}_p(1).$$

First of all, let us observe that for $t \in \mathcal{S}_M$ $\delta_{in}(t)$ may be written as follows

$$(3.21)\qquad \begin{aligned} \delta_{in}(t) &= n^{-1/2}[g'(X_i,\beta^0)]^{\mathrm{T}} t + \frac{1}{2} n^{-1} t^{\mathrm{T}} g''(X_i, \tilde{\beta}_t) t \\ &= n^{-1/2}[g'(X_i,\beta^0)]^{\mathrm{T}} t + n^{-1} t^{\mathrm{T}} [g''(X_i, \tilde{\beta}_t) - g''(X_i,\beta^0)] t \\ &\quad + n^{-1} t^{\mathrm{T}} g''(X_i,\beta^0) t \end{aligned}$$

where $\|\tilde{\beta}_t - \beta^0\| < n^{-1/2}M$. Let us assume only $n > M^2\delta_0^{-2}$. For such n we have $n^{-1/2}M < \delta_0$ and hence $\|\tilde{\beta}_t - \beta^0\| < \delta_0$ and so, using A.ii and (2.1), we find that

$$(3.22)\qquad n^{-1/2}|[g'(X_i,\beta^0)]^{\mathrm{T}} t| \le n^{-1/2} J p^{1/2} M,$$

$$(3.23)\qquad n^{-1}|t^T g''(X_i,\beta^0) t| \le n^{-1} J M^2,$$

and

$$n^{-1}|t^T [g''(X_i,\tilde{\beta}_t) - g''(X_i,\beta^0)] t| \le n^{-3/2} L M^2.$$

It means that there is a constant $C_0 < \infty$ such that for $n > M^2\delta_0^{-2}$ we have

$$(3.24)\qquad |\delta_{in}(t)| < n^{-1/2} C_0.$$

To conclude preparatory considerations, let us verify that for any $v \in \mathcal{S}_M$ there is an $h \in [\frac{1}{2} e^{-n^{-1/2}M}, \frac{1}{2} e^{n^{-1/2}M}]$ so that

$$(3.25)\qquad e^{-n^{-1/2}v} = 1 - n^{-1/2}v + h n^{-1} v^2.$$

Now we may write for ψ_a

$$(3.26)\qquad \psi_a([e_i - \delta_{in}(t)] e^{-n^{-1/2}v}) - \psi_a(e_i) = \psi_a'(e_i) \cdot \eta_i + \frac{1}{2} \cdot \psi_a''(\xi_i) \cdot \eta_i^2$$

where

$$(3.27)\qquad \begin{aligned} \eta_i &= (e_i - \delta_{in}(t))(1 - n^{-1/2}v + h n^{-1} v^2) - e_i \\ &= -(\delta_{in}(t) + n^{-1/2} e_i v) + n^{-1/2}\{\delta_{in}(t) v + n^{-1/2} h v^2 e_i - n^{-1/2}\delta_{in}(t) \cdot h \cdot v^2\} \end{aligned}$$

and there is $n_0 \in N$ so that for all $n > n_0$ we have

$$(3.28)\qquad \begin{aligned} &(e_i - n^{-1/2} C_0) e^{-\mathrm{sign}(e_i - n^{-1/2}C_0) \cdot n^{-1/2} M} \\ &\quad < \xi_i < (e_i + n^{-1/2} C_0) e^{\mathrm{sign}(e_i + n^{-1/2}C_0) \cdot n^{-1/2} M}. \end{aligned}$$

Employing (3.26) and (3.27), we obtain

$$
\begin{aligned}
(3.29)\qquad &\psi_a([e_i-\delta_{in}(t)]e^{-n^{-1/2}v})g'(X_i,\beta^0+n^{-1/2}t)-\psi_a(e_i)g'(X_i,\beta^0)\\
&=\psi_a(e_i)[g'(X_i,\beta^0+n^{-1/2}t)-g'(X_i,\beta^0)]\\
&+g'(X_i,\beta^0+n^{-1/2}t)\{\psi_a([e_i-\delta_{in}(t)]e^{-n^{-1/2}v})-\psi_a(e_i)\}\\
&=\psi_a(e_i)[g'(X_i,\beta^0+n^{-1/2}t)-g'(X_i,\beta^0)]-g'(X_i,\beta^0+n^{-1/2}t)\\
&\cdot(\delta_{in}(t)+n^{-1/2}e_iv)\psi_a'(e_i)+n^{-1/2}g'(X_i,\beta^0+n^{-1/2}t)\\
&\cdot\{\delta_{in}(t)v+n^{-1/2}he_i-n^{-1/2}\delta_{in}(t)\cdot h\cdot v^2\}\psi_a'(e_i)\\
&+\frac{1}{2}g'(X_i,\beta^0+n^{-1/2}t)\{-(\delta_{in}(t)+n^{-1/2}e_iv)\\
&+n^{-1/2}\{\delta_{in}(t)v+n^{-1/2}he_i-n^{-1/2}\delta_{in}(t)\cdot h\cdot v^2\}\}^2\psi_a''(\xi_i).
\end{aligned}
$$

Let us consider the first term of the right hand side of (3.29). Mimicking the steps of the analysis of $\delta_{in}(t)$ (see (3.21)→(3.24)) we obtain for any $\ell=1,2,\ldots,p$

$$
\begin{aligned}
(3.30)\quad &\psi_a(e_i)[g_\ell'(X_i,\beta^0+n^{-1/2}t)-g_\ell'(X_i,\beta^0)]\\
&=n^{-1/2}\psi_a(e_i)\sum_{k=1}^{p}\{[g_{\ell,k}''(X_i,\beta^{(t,\ell)})-g_{\ell,k}''(X_i,\beta^0)]t_k+g_{\ell,k}''(X_i,\beta^0)t_k\}
\end{aligned}
$$

where $\|\beta^{(t,\ell)}-\beta^0\|<n^{-1/2}M$. Since for $n>M^2\delta_0^{-2}$

$$
\begin{aligned}
&n^{-1/2}\sup_{\mathcal{S}_M}\left|\sum_{i=1}^{n}\psi_a(e_i)\sum_{k=1}^{p}\{[g_{\ell,k}''(X_i,\beta^{(t,\ell)})-g_{\ell,k}''(X_i,\beta^0)]t_k\}\right|\\
&\le n^{-1}\cdot L\cdot p^{1/2}\cdot M^2\cdot\sum_{i=1}^{n}|\psi_a(e_i)|,
\end{aligned}
$$

an application of the Markov law of large numbers verifies that this supremum is $\mathcal{O}_p(1)$. For any fix $k,\ j\in\{1,2,\ldots,p\}$ using B.iv and the Lindeberg-Lévy central limit theorem (CLT) we can show that for any $\varepsilon>0$ there is an integer $n_1>n_0$ and a constant C_1 such that for all $n>n_1$

$$
(3.31)\qquad P\left(\left|n^{-1/2}\sum_{i=1}^{n}\psi_a(e_i)g_{kj}''(X_i,\beta^0)\right|>C_1\right)<\varepsilon.
$$

Notice please that to obtain (3.31), we had to use the assumption that $I\!E\psi_a(e_i)=0$, see B.iv. But it means that

$$
\begin{aligned}
(3.32)\quad &P\left(\sup_{\mathcal{S}_M}\left|n^{-1/2}\sum_{j=1}^{p}\sum_{i=1}^{n}\psi_a(e_i)g_{kj}''(X_i,\beta^0)t_j\right|>p\cdot C_1\cdot M\right)\\
&\le P\left(\sum_{j=1}^{p}\left|n^{-1/2}\sum_{i=1}^{n}\psi_a(e_i)g_{kj}''(X_i,\beta^0)\right|\cdot\sup_{\mathcal{S}_M}|t_j|>p\cdot C_1\cdot M\right)<\varepsilon
\end{aligned}
$$

and so the second term of the right hand side of (3.30) is also bounded in probability.

Now let us focus our attention on the second term of (3.29). We may write it as

$$(3.33)\qquad \begin{aligned}&[g'(X_i,\beta^0+n^{-1/2}t)-g'(X_i,\beta^0)](\delta_{in}(t)+n^{-1/2}e_iv)\psi_a'(e_i)\\ &+g'(X_i,\beta^0)(\delta_{in}(t)+n^{-1/2}e_iv)\psi_a'(e_i).\end{aligned}$$

Let us fix again an ℓ. Using (3.22) and (3.24) we may find $\tilde\beta^{(t,\ell)}$, $\|\tilde\beta^{(t,\ell)}-\beta^0\|<n^{-1/2}M$ so that the absolute value of the ℓ-th coordinate of the first term of (3.33) is bounded by

$$\begin{aligned}n^{-1}&\left|\sum_{k=1}^{p}[g''_{\ell,k}(X_i,\tilde\beta^{(t,\ell)})]^T\cdot t_k\right|(C_0+|e_i|M)|\psi_a'(e_i)|\\ &\le n^{-1}J\cdot p\cdot M\cdot(C_0+|e_i|M)|\psi_a'(e_i)|\end{aligned}$$

and hence the application of the Markov law gives again the expected conclusion. Now, let us consider the first part of the second term of (3.33). Making use of (3.21) we obtain

$$(3.34)\quad \begin{aligned}&g'(X_i,\beta^0)\cdot\delta_{in}(t)\cdot\psi_a'(e_i)\\ &\quad=\{n^{-1/2}g'(X_i,\beta^0)[g'(X_i,\beta^0)]^{\mathrm T}t+n^{-1}g'(X_i,\beta^0)t^{\mathrm T}g''(X_i,\tilde\beta_t)t\}\psi_a'(e_i).\end{aligned}$$

Similarly as for the second term of (3.30), employing the Lindeberg-Lévy theorem we find that for any $k=1,2,\ldots,p$

$$\sup_{\mathcal S_M}\left|n^{-1/2}\sum_{i=1}^{n}g_k'(X_i,\beta^0)\sum_{j=1}^{p}g_j'(X_i,\beta^0)t_j\psi_a'(e_i)+n^{1/2}\gamma Qt\right|$$

is bounded in probability (see (3.32)). In other words, the first term of (3.34) is bounded in probability. For the second term of (3.34) we have again for any $k=1,2,\ldots,p$

$n^{-1}\sup_{\mathcal S_M}|g_k'(X_i,\beta^0)t^{\mathrm T}g''(X_i,\tilde\beta_t)t\psi_a'(e_i)|\le n^{-1}\cdot p\cdot J^2\cdot M^2\cdot\sum_{i=1}^n|\psi_a'(e_i)|$ and we apply the Markov law once again. It remains to prove that the last term of (3.33), namely $n^{-1/2}\sup_{\mathcal S_M}\sum_{i=1}^n g'(X_i,\beta^0)\cdot e_i\cdot v\cdot\psi_a'(e_i)$ is approximately equal to $n^{1/2}\theta qv$. We have

$$\begin{aligned}n^{-1/2}\sup_{\mathcal S_M}&\left|\sum_{i=1}^n g'(X_i,\beta^0)\cdot e_i\cdot v\cdot\psi_a'(e_i)-n^{1/2}\theta qv\right|\\ &\le n^{-1/2}\left|\sum_{i=1}^n g'(X_i,\beta^0)\cdot e_i\cdot\psi_a'(e_i)-n^{1/2}\theta qv\right|\sup_{\mathcal S_M}|v|\\ &\le n^{-1/2}M\left|\sum_{i=1}^n g'(X_i,\beta^0)\cdot e_i\cdot\psi_a'(e_i)-n^{1/2}\theta qv\right|.\end{aligned}$$

Using CLT we find that this term is also bounded in probability (see again (3.32) and B.iv), so that the same is true for (3.33). Using again (3.21) we find an upper bound of the norm of the last but one term of (3.29) in a form

$$n^{-1}\cdot J\cdot\{C_2+|e_i|\cdot C_3\}\,|\psi_a'(e_i)|$$

for appropriate finite constants C_2 and C_3 and the Markov law yields again the desired result.

To conclude the proof of (3.20) we have to show that the last term of (3.29) is also $\mathcal{O}_p(1)$. This last term is sum of expressions of the type

$$\frac{1}{2}g'(X_i,\beta^0+n^{-1/2}t)w_{i1}\cdot w_{i2}\cdot\psi_a''(\xi_i)$$

where w_{ij}'s are from the set $\{\delta_{in}(t),n^{-1/2}e_iv,n^{-1/2}\delta_{in}(t)v,n^{-1}he_i,-n^{-1}\delta_{in}(t)hv^2\}$ and for ξ_i see (3.28). Let us consider at first $w_{i1}=w_{i2}=\delta_{in}(t)$, i.e. we shall analyze

$$\frac{1}{2}g'(X_i,\beta^0+n^{-1/2}t)\delta_{in}^2(t)\psi_a''(\xi_i).$$

Using (2.1), (3.24) and B.iv, we find $n_2>n_1$ and $\zeta_1<\zeta_0$ so that for all $n>n_2$ the absolute value of this term is not larger than $n^{-1}C_0^2\cdot J^2\sum_{i=1}^n|\bar{\psi}_{a,\zeta_1}''(e_i)|$ and applying (3.28), B.iv and the law of large numbers we find that this term is of order $\mathcal{O}_p(1)$. The verification of the same assertion for other expressions is very similar.

Now let us consider $\psi=\psi_c$. Recalling that we have denoted in B.i the points of jumps of ψ_c' by $r_{2,1},r_{2,2},\ldots,r_{2,s_2}$, let us put $r_{2,0}=-\infty$ and $r_{2,s_2+1}=\infty$. Let us also denote the values of the derivative of ψ_c by $\psi_c'(z)=\alpha_j$ for $z\in(r_{2,j-1},r_{2,j})$, $j=1,2,\ldots,s_2+1$. (This notations will be assumed valid only within this proof. Notice that B.iii implies that $\alpha_1=0$ as well as $\alpha_{s_2+1}=0$.) Now let $k_n=n^{-1/2}p^{1/2}JM$ and put $\mathcal{C}_n=\bigcup_{j=1}^{s_2}[r_{2,j}-k_n,r_{2,j}+k_n]$. It is clear that we can find an n_3 so that for all $n>n_3$, $k_n<\vartheta_0$ (see B.ii). In the rest of this part of proof (i.e. in the part which is devoted to ψ_c) we shall consider only $n>n_3$. Now let us define for any $n\in N$ the function $\psi_n(z)$ as follows. We shall put $\psi_n(z)=\psi_c(z)$ for $z\notin\mathcal{C}_n$ which implies that $\psi_n'(z)=\alpha_j$ for $z\in(r_{2,j-1}+k_n,r_{2,j}-k_n)$ for some $j\in\{1,2,\ldots,s_2+1\}$. For $z\in\mathcal{C}_n$ define $\psi_n(z)$ as a smooth and monotone (e.g. quadratic) function such that $\psi_c(z)=\psi_n(z)$ as well as $\psi_c'(z)=\psi_n'(z)$ at the endpoints of intervals $[r_{2,j}-k_n,r_{2,j}+k_n]$ for $j=1,2,\ldots,s_2$. One possibility is that for $j=1,2,\ldots,s_2$ the derivative will be linear from point $r_{2,j}-k_n$, at which it attains value α_j, to the point $r_{2,j}+k_n$, at which it attains of course value α_{j+1}, i.e.

$$\psi_n'(z)=\alpha_j+\frac{1}{2k_n}(\alpha_{j+1}-\alpha_j)(z-r_{2,j}+k_n)\qquad\text{for}\quad z\in[r_{2,j}-k_n,r_{2,j}+k_n].$$

We are going to show that

$$(3.35)\quad \sup_{\mathcal{S}_M}\Bigg\|n^{-1/2}\sum_{i=1}^n\{\psi_c([e_i-\delta_{in}(t)]\sigma^{-1}e^{-n^{-1/2}v})g'(X_i,\beta^0+n^{-1/2}t)$$
$$-\psi_c(e_i\cdot\sigma^{-1})g'(X_i,\beta^0)-\psi_n([e_i-\delta_{in}(t)]\sigma^{-1}e^{-n^{-1/2}v})g'(X_i,\beta^0+n^{-1/2}t)$$
$$+\psi_n(e_i\cdot\sigma^{-1})g'(X_i,\beta^0)\}\Bigg\|=\mathcal{O}_p(1).$$

Let us denote the i-th term of the sum in (3.35) as $u_{i0}^{(n)}(t)$. It can be written as

$$g'(X_i,\beta^0+n^{-1/2}t)\{\psi_c([e_i-\delta_{in}(t)]\sigma^{-1}e^{-n^{-1/2}v})-\psi_n([e_i-\delta_{in}(t)]\sigma^{-1}e^{-n^{-1/2}v})\}$$
$$+g'(X_i,\beta^0)\{\psi_n(e_i\cdot\sigma^{-1})-\psi_c(e_i\cdot\sigma^{-1})\}.$$

Due to the definition of the function $\psi_n(z)$ (please keep in mind that it is equal to $\psi_c(z)$ for $z \in \mathcal{C}_n^c$), $u_{i_0}^{(n)}(t)$ can be nonzero only in the case when for some $j_0 \in 1, 2, \ldots, s_2$ either $r_{2,j_0} - k_n < e_i < r_{2,j_0} + k_n$ or $r_{2,j_0} - k_n < [e_i - \delta_{in}(t)]e^{-n^{-1/2}v} < r_{2,j_0} + k_n$ or both. From it we obtain that $u_{i_0}^{(n)}(t)$ can be nonzero if

$$r_{2,j_0} - k_n < e_i < r_{2,j_0} + k_n$$

or

$$(r_{2,j_0} - k_n)e^{n^{-1/2}v} + \delta_{in}(t) < e_i < (r_{2,j_0} + k_n)e^{n^{-1/2}v} + \delta_{in}(t)$$

and due to the fact that $\max\{|r_{2,1}|, |r_{2,s_2}|\} < \infty$, due to the order of k_n and of $e^{n^{-1/2}M}$ and due to (3.24) there is a sequence $\{\ell_n\}_{n=1}^{\infty}$ with $\ell_n = \mathcal{O}(n^{-1/2})$ such that

$$[r_{2,j_0} - k_n, r_{2,j_0} + k_n] \subset [r_{2,j_0} - \ell_n, r_{2,j_0} + \ell_n]$$

as well as

$$[(r_{2,j_0} - k_n)e^{n^{-1/2}v} + \delta_{in}(t), (r_{2,j_0} + k_n)e^{n^{-1/2}v} + \delta_{in}(t)] \subset [r_{2,j_0} - \ell_n, r_{2,j_0} + \ell_n].$$

It implies that the probability that $u_{i_0}^{(n)}$ is nonzero is not larger than the probability of the event that $e_{i_0} \in \mathcal{C}_n^* = \bigcup_{j=1}^{s_2}[r_{2,j} - \ell_n, r_{2,j} + \ell_n]$. This probability is evidently (due to the fact that $s_2 < \infty$) $\mathcal{O}(n^{-1/2})$.

On the other hand, $\sup_{\mathcal{S}_M} \|u_{i_0}^{(n)}(t)\| \leq 2 \cdot \max_{1 \leq j \leq s_2} |\alpha_j| p^{1/2} J \cdot k_n \cdot I_{\mathcal{C}_n^*}(e_i)$, and hence

$$E_F \sup_{\mathcal{S}_M} \left\| \sum_{i=1}^n u_i^{(n)} \right\| \leq \sum_{i=1}^n E_F \sup_{\mathcal{S}_M} \|u_i^{(n)}\| \leq 2 \cdot \max_{1 \leq j \leq s_2} |\alpha_j| \cdot J \cdot k_n \cdot E_F \sum_{i=1}^n I_{\mathcal{C}_n^*}(e_i) = \mathcal{O}(1).$$

The application of the Chebyshev inequality concludes the proof of (3.35). Moreover, we find for $\nu = 1, 2, 3$

$$E_F|\psi_n'(e_1)|^\nu = E_F|\psi_c'(e_1)|^\nu + \mathcal{O}(n^{-1/2}) \quad \text{as} \quad n \to \infty \tag{3.36}$$

and

$$E_F|e_1\psi_n'(e_1)|^\nu = E_F|e_1\psi_c'(e_1)|^\nu + \mathcal{O}(n^{-1/2}) \quad \text{as} \quad n \to \infty. \tag{3.37}$$

Observe that due to the fact that $\psi_c'(z) = 0$ for $|z| > \max\{|r_{2,1}|, |r_{2,s_2}|\}$ and $\max_{1 \leq j \leq s_2+1} |\alpha_j| < \infty$ we have $E_F|e_1\psi_c'(e_1)|^\nu < \infty$ for any $\nu \in N$. But this together with (3.36) and (3.37) implies that Conditions B hold also for $\psi_n(z)$ and since $\psi_n(z)$ has derivative which is Lipschitz, we can apply on $\psi_n(z)$ the assertion of this lemma already proved for $\psi_a(z)$. In other words, (3.20) holds for $\psi_n(z)$, too. Moreover we have $\gamma_n = \sigma^{-1}\mathbb{E}_F\psi_n'(e_1 \cdot \sigma^{-1})$ and $\theta_n = \sigma^{-1}\mathbb{E}_F e_1\psi_n'(e_1 \cdot \sigma^{-1})$ (notice please that for the functions $\psi_n(z)$ the second terms in definition of γ and θ are zero, see (2.2) and (3.20)), we have $\gamma - \gamma_n = \mathcal{O}_p(n^{-1/2})$ and $\theta - \theta_n = \mathcal{O}_p(n^{-1/2})$, and (3.20) holds also for ψ_c.

For ψ_s we shall prove that

$$\sup_{\mathcal{S}_M} \Bigg\| n^{-1/2} \sum_{i=1}^n \{\psi_s([e_i - \delta_{in}(t)]\sigma^{-1}e^{-n^{-1/2}v})g'(X_i, \beta^0 + n^{-1/2}t) - \psi_s(e_i \cdot \sigma^{-1})g'(X_i, \beta^0)\} + \gamma Q t + \theta q u \Bigg\| = \mathcal{O}_p(n^{-1/4}). \tag{3.38}$$

Similarly as above we may write (let us recall that we have assumed that $\sigma = 1$ and $M > 1$)

$$\sum_{i=1}^{n}\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v})g'(X_i, \beta^0 + n^{-1/2}t) - \psi_s(e_i)g'(X_i, \beta^0)\} \tag{3.39}$$
$$= \sum_{i=1}^{n}\{g'(X_i, \beta^0 + n^{-1/2}t)\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi_s(e_i)\}$$
$$+\{g'(X_i, \beta^0 + n^{-1/2}t) - g'(X_i, \beta^0)\}\psi_s(e_i)\}.$$

Without loss of generality let us assume that $s_1 = 1$, i.e. there is only one point of discontinuity of the function $\psi_s(z)$, and let us denote it by r. Moreover, let us denote $\psi_s(z) = \tau_1$ for $z < r$ and $\psi_s(z) = \tau_2$ for $z > r$.

Let us consider at first the first sum in (3.39) and let us put

$$u_i^{(n)}(t, v) = g'(X_i, \beta^0 + n^{-1/2}t)\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi_s(e_i)\}.$$

Similarly as above $u_i^{(n)}(t, v)$ can be nonzero if either

$$e_i \le r \le [e_i - \delta_{in}(t)]e^{-n^{-1/2}v} \leftrightarrow re^{n^{-1/2}v} + \delta_{in}(t) \le e_i \le r \tag{3.40}$$

or

$$[e_i - \delta_{in}(t)]e^{-n^{-1/2}v} \le r \le e_i \leftrightarrow r \le e_i \le re^{n^{-1/2}v} + \delta_{in}(t). \tag{3.41}$$

Denote the events described in (3.40) and (3.41) successively by $B_i^{(k)}(n, t, v)$, $k = 1, 2$. First of all, please observe that (3.40) can hold when

$$re^{n^{-1/2}v} + \delta_{in}(t) \le r. \tag{3.42}$$

Similarly (3.41) can take place if

$$r \le re^{n^{-1/2}v} + \delta_{in}(t) \tag{3.43}$$

and for fix i, t, v and n only one of possibilities (3.42) and (3.43) holds. Notice please that in the case when (3.42) holds, $\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi_s(e_i) \ge 0$ and hence also

$$I\!E\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi_s(e_i)\} \ge 0$$

while for (3.43) $\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi_s(e_i) \le 0$.

Fix an $\ell \in \{1, 2, \ldots, p\}$, denote by $u_{i\ell}^{(n)}(t, v)$ the ℓ-th coordinate of $u_i^{(n)}(t, v)$ and finally denote successively by $D_i^{(j)}(n, t, v, \ell)$, $j = 1, 2, 3, 4$ the events

$$\{\omega \in \Omega : \{re^{n^{-1/2}v} + \delta_{in}(t) \le r\} \cap \{g'_\ell(X_i, \beta^0 + n^{-1/2}t) \le 0\}\},$$
$$\{\omega \in \Omega : \{re^{n^{-1/2}v} + \delta_{in}(t) \le r\} \cap \{g'_\ell(X_i, \beta^0 + n^{-1/2}t) > 0\}\},$$
$$\{\omega \in \Omega : \{re^{n^{-1/2}v} + \delta_{in}(t) > r\} \cap \{g'_\ell(X_i, \beta^0 + n^{-1/2}t) \le 0\}\},$$
$$\{\omega \in \Omega : \{re^{n^{-1/2}v} + \delta_{in}(t) > r\} \cap \{g'_\ell(X_i, \beta^0 + n^{-1/2}t) > 0\}\}.$$

Further denote for $j = 1,2,3,4$ and $k = 1,2$ by $\pi_i^{(j,k)}(n,t,v,\ell)$ the probabilities of the events $B_i^{(k)}(n,t,v) \cap D_i^{(j)}(n,t,v,\ell)$. Due to (3.24) there is $n_4 \in N$ so that for all $n > n_4$ we have $|re^{n^{-1/2}v} + \delta_{in}(t) - r| < \vartheta_0$ (see B.ii). In what follows let us assume only $n > n_4$. Let us also recall that then $f_{e_i}(u) < H$ for $e_i \in [re^{n^{-1/2}v} + \delta_{in}(t), r]$ (where the endpoints of the interval should be interchanged if $r < re^{n^{-1/2}v} + \delta_{in}(t)$). For any $n > n_4$, $j = 1,2,3,4$ and $k = 1,2$ we have

$$\begin{aligned}(3.44)\qquad \pi_i^{(j,k)}(n,t,v,\ell) &= I\!\!E I_{B_i^{(k)}(n,t,v) \cap D_i^{(j)}(n,t,v,\ell)} \\ &\leq \int_{re^{n^{-1/2}v}+\delta_{in}(t)}^{r} f_{e_i}(u)du \leq H \int_{re^{n^{-1/2}v}+\delta_{in}(t)}^{r} du\end{aligned}$$

and hence due to (2.1) there is a constant C_4 such that

$$(3.45)\qquad \pi_i^{(j,k)}(n,t,v,\ell) < n^{-1/2} C_4.$$

Of course, lower and upper bound in (3.44) should be interchanged if $r < n^{-1/2}[\delta_{in}(t) + re^{n^{-1/2}u}]$ but (3.45) holds for any combination of j and k. Now, we shall study the sum

$$\sum_{i=1}^{n} [u_{i\ell}^{(n)}(t,v) - I\!\!E u_{i\ell}^{(n)}(t,v)].$$

Since $D_i^{(j_1)}(n,t,v,\ell)$ and $D_i^{(j_2)}(n,t,v,\ell)$ are disjoint for any $j_1 \neq j_2$ and $\cup_{j=1}^4 D_i^{(j)}(n,t,v,\ell) = \Omega$, we have $u_i^{(n)}(t,v) = \sum_{j=1}^p [u_i^{(n)}(t,v) I_{D_i^{(j)}(n,t,v,\ell)}]$, and hence

$$I\!\!E u_i^{(n)}(t,v) = \sum_{j=1}^{p} I\!\!E[u_i^{(n)}(t,v) I_{D_i^{(j)}(n,t,v,\ell)}].$$

Then

$$\begin{aligned}(3.46)\qquad &\sum_{i=1}^{n} [u_{i\ell}^{(n)}(t,v) - I\!\!E u_{i\ell}^{(n)}(t,v)] \\ &= \sum_{i=1}^{n}\sum_{j=1}^{p} \{u_{i\ell}^{(n)}(t,v) I_{D_i^{(j)}(n,t,v,\ell)} - I\!\!E[u_{i\ell}^{(n)}(t,v) I_{D_i^{(j)}(n,t,v,\ell)}]\}.\end{aligned}$$

Now consider $u_{i\ell}^{(n)}(t,v) I_{D_i^{(1)}(n,t,v,\ell)} - I\!\!E[u_{i\ell}^{(n)}(t,v) I_{D_i^{(1)}(n,t,v,\ell)}]$, i.e. put $j = 1$. Let us recall that then $re^{n^{-1/2}v} + \delta_{in}(t) \leq r$ which implies that $e_i \leq r \leq [e_i - \delta_{in}(t)]e^{-n^{-1/2}v}$, see (3.40). Denoting $\varrho = \tau_2 - \tau_1 = \psi_s(r+) - \psi_s(r-)$, we easy find that

$$\begin{aligned}(3.47)\qquad &u_{i\ell}^{(n)}(t,v) I_{D_i^{(1)}(n,t,v,\ell)} - I\!\!E[u_{i\ell}^{(n)}(t,v) I_{D_i^{(1)}(n,t,v,\ell)}] \\ &= \varrho \cdot g'_\ell(X_i, \beta^0 + n^{-1/2}t)(1 - \pi_i^{(1,1)}(n,t,v,\ell)) \\ &= -\varrho \cdot |g'_\ell(X_i, \beta^0 + n^{-1/2}t)|(1 - \pi_i^{(1,1)}(n,t,v,\ell)) \\ &> -\varrho \cdot |g'_\ell(X_i, \beta^0 + n^{-1/2}t)| \quad \text{with probability} \quad \pi_i^{(1,1)}(n,t,v,\ell)\end{aligned}$$

and

$$
\begin{aligned}
(3.48) \qquad &= -\varrho \cdot g'_\ell(X_i, \beta^0 + n^{-1/2}t)\pi_i^{(1,1)}(n,t,v,\ell) \\
&= \varrho \cdot |g'_\ell(X_i, \beta^0 + n^{-1/2}t)|\pi_i^{(1,1)}(n,t,v,\ell) \\
&< n^{-1/2}\varrho \cdot C_4|g'_\ell(X_i, \beta^0 + n^{-1/2}t)| \\
&\qquad \text{with probability} \quad 1 - \pi_i^{(1,1)}(n,t,v,\ell).
\end{aligned}
$$

Taking into account the expressions staying after the first sign of equality in (3.47) and in (3.48), and corresponding probabilities, we find that

$$\mathbb{E}\{u_{i\ell}^{(n)}(t,v)I_{D_i^{(1)}(n,t,v,\ell)} - \mathbb{E}[u_{i\ell}^{(n)}(t,v)I_{D_i^{(1)}(n,t,v,\ell)}]\} = 0.$$

So, putting for any $n \in N$ and $i = 1, 2, \ldots, n$ $a_{i\ell}(n,t,v) = \varrho|g'_\ell(X_i, \beta^0 + n^{-1/2}t)|\pi_i^{(1,1)}$ (n,t,v,ℓ) and $b_{i\ell}(n,t,v) = \varrho \cdot |g'_\ell(X_i, \beta^0 + n^{-1/2}t)|(1 - \pi_i^{(1,1)}(n,t,v,\ell))$, we may utilize Lemma A.2 and define

$\mu_{i\ell}^{(1)}(n,t,v)$ the time for Wiener process to exit the interval $(-a_{i\ell}(n,t,v), b_{i\ell}(n,t,v))$

and we obtain

$$u_{i\ell}^{(n)}(t,v)I_{D_i^{(1)}(n,t,v,\ell)} - \mathbb{E}[u_{i\ell}^{(n)}(t,v)I_{D_i^{(1)}(n,t,v,\ell)}] =_{\mathcal{D}} W(\mu_{i\ell}^{(1)}(n,t,v))$$

where "$=_{\mathcal{D}}$" means equality in distribution. Similarly we find for $j = 2, 3$ and 4

$$u_{i\ell}^{(n)}(t,v)I_{D_i^{(j)}(n,t,v,\ell)} - \mathbb{E}[u_{i\ell}^{(n)}(t,v)I_{D_i^{(j)}(n,t,v,\ell)}] =_{\mathcal{D}} W(\mu_{i\ell}^{(j)}(n,t,v)).$$

Finally, putting $\mu_{i\ell}(n,t,v) = \sum_{j=1}^4 \mu_{i\ell}^{(j)}(n,t,v)$ and employing (3.46), we obtain

$$
\begin{aligned}
&n^{-1/4}\sum_{i=1}^n [u_{i\ell}^{(n)}(t,v) - \mathbb{E}u_{i\ell}^{(n)}(t,v)] \\
&\quad =_{\mathcal{D}} n^{-1/4}\sum_{i=1}^n\sum_{j=1}^4 W(\mu_{i\ell}^{(j)}(n,t,v)) =_{\mathcal{D}} n^{-1/4}\sum_{i=1}^n W(\mu_{i\ell}(n,t,v)) \\
&\quad =_{\mathcal{D}} W\left(n^{-1/2}\sum_{i=1}^n \mu_{i\ell}(n,t,v)\right).
\end{aligned}
$$

Now let us put $\tilde{g}'_\ell = \sup_{S_M}|g'_\ell(X_i, \beta^0 + n^{-1/2}t)|$. Due to (2.1) there is n_5 so that for all $n > n_5$, we have $\tilde{g}'_\ell < \infty$. Further, let us take into account inequalities which are given in (3.45), (3.47) and (3.48), and put $c_{i\ell} = \varrho \cdot \tilde{g}'_\ell$ and $d_{i\ell} = n^{-1/2}\varrho \cdot C_4 \cdot \tilde{g}'_\ell$. Defining

(3.49) $\kappa_{i\ell}^{(1)}(n,M)$ the time for Wiener process to exit the interval $(-c_{i\ell}, d_{i\ell})$,

we obtain

$$\mu_{i\ell}^{(1)}(n,t,v) \le \kappa_{i\ell}^{(1)}(n,M).$$

Having done a similar step for others j's and putting

$$\kappa_{i\ell}(n,M) = \sum_{j=1}^4 \kappa_{i\ell}^{(j)}(n,M),$$

we arrive at

$$\sup_{\mathcal{S}_M} n^{-1/4} \left|\sum_{i=1}^{n}[u_{i\ell}^{(n)}(t,v) - \mathbb{E}u_{i\ell}^{(n)}(t,v)]\right| \tag{3.50}$$
$$=_{\mathcal{D}} \sup_{\mathcal{S}_M} \left| W\left(n^{-1/2}\sum_{i=1}^{n}\mu_{i\ell}(n,t,u)\right)\right|$$
$$\leq \sup\left\{|W(s)| : 0 \leq s \leq n^{-1/2}\sum_{i=1}^{n}\kappa_{i\ell}(n,M)\right\}.$$

Moreover, see again Lemma A.2, we have from (3.49) for any $t, u \in \mathcal{S}_M$

$$\mathbb{E}\kappa_{i,\ell}(n,M) \leq 4n^{-1/2}C_4[\tilde{g}'_\ell]^2$$

for all $n \in N$, i.e.

$$n^{-1/2}\sum_{i=1}^{n}\mathbb{E}\kappa_{i,\ell}(n,M) \leq C_4[\tilde{g}'_\ell]^2.$$

It means that for any positive ε there is a constant C_5 and $n_\varepsilon > n_5$ so that for any $n > n_\varepsilon$

$$P\left\{n^{-1/2}\sum_{i=1}^{n}\kappa_i(n,M) > C_5\right\} < \frac{\varepsilon}{2} \tag{3.51}$$

and then there is also $C_6 \in (0,\infty)$ such that

$$P\{\sup\{|W(s)| : 0 \leq s \leq C_5\} > C_6\} < \frac{\varepsilon}{2}, \tag{3.52}$$

see e.g. Csörgö and Révész (1981). Taking into account (3.50), (3.51) and (3.52), we arrive at

$$P\{\sup_{\mathcal{S}_M} n^{-1/4}|u_{i\ell}^{(n)}(t,v) - \mathbb{E}u_{i\ell}^{(n)}(t,v)| > C_6\} < \varepsilon$$

and it means that also

$$\sup_{\mathcal{S}_M} n^{-1/4}\left\|\sum_{i=1}^{n}[u_{i\ell}^{(n)}(t,v) - \mathbb{E}u_{i\ell}^{(n)}(t,v)]\right\|$$

is bounded in probability. To prove that also

$$\sup_{\mathcal{S}_M} n^{-1/4}\left\|\sum_{i=1}^{n}\{g'(X_i,\beta^0 + n^{-1/2}t) - g'(X_i,\beta^0)\}\psi_s(e_i)\right\|$$

is bounded in probability is much easier. We shall give only a hint, precise arguments should be done "coordinatewise". Similarly as above we can write it as

$$\sup_{\mathcal{S}_M} n^{-3/4}\left\|\sum_{i=1}^{n} g''(X_i,\bar{\beta})t\psi_s(e_i)\right\|$$

where $\bar{\beta}_j$ is between β_j^0 and $\beta_j^0 + n^{-1/2}t_j$ for all $j = 1, 2, \ldots, p$. So we have

$$\sup_{\mathcal{S}_M} n^{-1/4} \left\| \sum_{i=1}^{n} \{g'(X_i, \beta^0 + n^{-1/2}t) - g'(X_i, \beta^0)\}\psi_s(e_i) \right\| \tag{3.53}$$
$$\leq \sup_{\mathcal{S}_M} n^{-3/4} \left\| \sum_{i=1}^{n} g''(X_i, \bar{\beta})t - g''(X_i, \beta^0)t\psi_s(e_i) \right\|$$
$$+ \sup_{\mathcal{S}_M} n^{-3/4} \left\| \sum_{i=1}^{n} g''(X_i, \beta^0)t\psi_s(e_i) \right\|.$$

Taking into account A.ii we have

$$\|g''(X_i, \bar{\beta})t - g''(X_i, \beta^0)t\| \leq n^{-1/2} \cdot p \cdot L \cdot \|\bar{\beta} - \beta^0\| M$$

and due to the fact that $\max_{z \in R} |\psi_s(z)| = \max\{|\tau_1|, |\tau_2|\}$, the first term of (3.53) is not larger than $n^{-1/4}C_7$ for some positive constant C_7. Using CLT we can verify that the second term of (3.53) is $\mathcal{O}_p(n^{-1/4})$.

We shall conclude the proof if we show that

$$\sup_{\mathcal{S}_M} \Bigg\| n^{-1/2} \sum_{i=1}^{n} \mathbb{E}\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v})g'(X_i, \beta^0 + n^{-1/2}t) \tag{3.54}$$
$$- \psi_s(e_i)g'(X_i, \beta^0)\} + \gamma Qt + \theta qu \Bigg\| = \mathcal{O}_p(n^{-1/4}).$$

Let us try to evaluate the mean value from the (3.54).

$$\sum_{i=1}^{n} \mathbb{E}\{g'(X_i, \beta^0 + n^{-1/2}t)\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - g'(X_i, \beta^0)\psi_s(e_i)\} \tag{3.55}$$

$$= \sum_{i=1}^{n} \mathbb{E}\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v})\{g'(X_i, \beta^0 + n^{-1/2}t) - g'(X_i, \beta^0)\} \tag{3.56}$$
$$+ \sum_{i=1}^{n} \mathbb{E}\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi_s(e_i)\}g'(X_i, \beta^0).$$

Now, (3.55) can be written as

$$\sum_{i=1}^{n} \mathbb{E}\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi(e_i)\} \cdot \{g'(X_i, \beta^0 + n^{-1/2}t) - g'(X_i, \beta^0)\} \tag{3.57}$$
$$+ \sum_{i=1}^{n} \mathbb{E}\psi(e_i) \cdot \{g'(X_i, \beta^0 + n^{-1/2}t) - g'(X_i, \beta^0)\}.$$

Let us denote by $\pi_i(n, t, v)$ the probability that $\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi(e_i) \neq 0$. From (2.1) it follows that $\pi_i(n, t, v) \leq n^{-1/2}C_8$ for some positive constant. Now, let us fix $\ell \in \{1, 2, \ldots, p\}$. Then similarly as above we may find $\tilde{\bar{\beta}}^{(t,\ell)}$ so that for all $j = 1, 2, \ldots, p$

$\tilde{\tilde{\beta}}_j^{(t,\ell)}$ is between β_j^0 and $\beta_j^0 + n^{-1/2} t_j$ and we have

$$(3.58)\quad |I\!E\{\psi_s([e_i - \delta_{in}(t)]e^{-n^{-1/2}v}) - \psi(e_i)\} \cdot \{g'_\ell(X_i, \beta^0 + n^{-1/2}t) - g'_\ell(X_i, \beta^0)\}|$$
$$= \left| n^{-1/2} \cdot \pi_i(n,t,v) \cdot \varrho \sum_{k=1}^{p} g''_{\ell,k}(X_i, \tilde{\tilde{\beta}}^{(t,\ell)}) t_k \right| \le n^{-1} C_8 \cdot \varrho \cdot J \cdot p \cdot M.$$

Due to the assumption that $I\!E\psi(e_i\sigma^{-1}) = 0$, the second term of (3.57) is zero for all t. Since the upper bound in (3.58) is independent of t and v whenever they are from $\mathcal{S}_M$, the supremum over $\mathcal{S}_M$ of norm of the expression in (3.55) is $\mathcal{O}(1)$.

Let us turn our attention to (3.56). It is equal to

$$\varrho \sum_{i=1}^{n} g'(X_i, \beta^0) \int_{re^{n^{-1/2}v} + \delta_{in}(t)}^{r} f_{e_i}(u) du$$
$$= \varrho \sum_{i=1}^{n} g'(X_i, \beta^0) \left\{ \int_{re^{n^{-1/2}v} + \delta_{in}(t)}^{r} f_{e_i}(r) du + \int_{re^{n^{-1/2}v} + \delta_{in}(t)}^{r} [f_{e_i}(u) - f_{e_i}(r)] du \right\}.$$

Utilizing (3.21) and (3.25) we can write

$$re^{n^{-1/2}v} + \delta_{in}(t) - r$$
$$= r(1 - n^{-1/2}v + n^{-1}hv^2) + n^{-1/2}[g'(X_i, \beta^0)]^{\mathrm{T}} t + n^{-1} t^{\mathrm{T}} g''(X_i, \tilde{\beta}_t) t$$

(where $\tilde{\beta}_t$ is given at the first line under (3.21)). Hence there is a constant C_9 so that for all $t, v \in \mathcal{S}_M$ $|re^{n^{-1/2}v} + \delta_{in}(t) - r| \le n^{-1/2} C_9$. But then we have for all $u \in [re^{n^{-1/2}v} + \delta_{in}(t), r]$ (where lower and upper bound of the interval should be changed if it is appropriate) $|f_{e_i}(u) - f_{e_i}(r)| \le n^{-1/2} C_9 H$ (see B.ii). It implies that

$$\left| \varrho \sum_{i=1}^{n} g'(X_i, \beta^0) \int_{re^{n^{-1/2}v} + \delta_{in}(t)}^{r} [f_{e_i}(u) - f_{e_i}(r)] du \right|$$
$$\le n^{-1/2} C_9 H \varrho J \sum_{i=1}^{n} \left[\int_{re^{n^{-1/2}v} + \delta_{in}(t)}^{r} du \right] \le \varrho C_9^2 H J$$

and the expression in (3.56) can be written as

$$\varrho g'(X_i, \beta^0) f_{e_i}(r) \int_{re^{n^{-1/2}v} + \delta_{in}(t)}^{r} du + \mathcal{O}(1)$$
$$= \varrho \sum_{i=1}^{n} g'(X_i, \beta^0) f_{e_i}(r)$$
$$\cdot (n^{-1/2} vr + n^{-1} hv^2 r + n^{-1/2} [g'(X_i, \beta^0)]^{\mathrm{T}} t + n^{-1} t^{\mathrm{T}} g''(X_i, \tilde{\beta}_t) t) + \mathcal{O}(1)$$

where we have used again (3.21) and (3.25) (for $\tilde{\beta}_t$ see again the first line under (3.21)). Recalling that $\varrho = \psi_s(r+) - \psi_s(r-)$, we obtain (3.56) in the following form

$$\varrho \sum_{i=1}^{n} g'(X_i, \beta^0) \int_{re^{n^{-1/2}v} + \delta_{in}(t)}^{r} f_{e_i}(u) du + \mathcal{O}(1)$$

$$= n^{-1/2}[\psi_s(r+) - \psi_s(r-)]$$
$$\cdot \sum_{i=1}^{n} g'(X_i, \beta^0)\{r f_{e_i}(r) v + [g'(X_i, \beta^0)]^{\mathrm{T}} t\} + \mathcal{O}(1).$$

Recalling that

$$\frac{1}{n}\sum_{i=1}^{n} g'(X_i, \beta^0)\,[g'(X_i, \beta^0)]^{\mathrm{T}} = Q + \mathcal{O}(n^{-1/4}) \quad \text{and}$$
$$\frac{1}{n}\sum_{i=1}^{n} g'(X_i, \beta^0) = q + \mathcal{O}(n^{-1/4}).$$

we conclude the proof. □

4. Bahadur representation

In this section we shall give the Bahadur representation for the difference of the estimators $\hat{\beta}^{(n)} - \hat{\beta}^{(n, I_{k_n})}$. We shall consider successively the following three cases:

i)

$$k_n = k \quad \text{for all} \quad n \in N, \tag{4.1}$$

ii)

$$\text{there is} \quad \nu \in (0,1) \quad \text{and} \quad \lambda \in R \ \text{so that} \ \lim_{n\to\infty} k_n \cdot n^{-\nu} = \lambda, \tag{4.2}$$

iii)

$$\text{there is} \quad \lambda \in (0, \tfrac{1}{2}] \quad \text{so that} \ \lim_{n\to\infty} k_n n^{-1} = \lambda. \tag{4.3}$$

Remark 4.1. We have assumed $\lambda \in (0, \frac{1}{2}]$ instead of a general case $\lambda \in (0,1]$. If we allowed $\lambda \in (0,1]$, at the further considerations we would have to take $\min\{k_n, n-k_n\}$ instead of k_n or $\max\{k_n, n-k_n\}$ instead of $n-k_n$. Similarly, we have assumed that $\{k_n\}_{n=1}^{\infty}$ is a nondecreasing sequences instead of a general one (see Conditions C). Not having the assumption of monotonicity but only of convergence, we would have to speak about "lim sup" or "lim inf" instead of "lim" to keep the following text fully correct.

First of all we shall study the order of convergence of $\hat{\beta}^{(n)} - \hat{\beta}^{(n, I_{k_n})}$. We obtain:

LEMMA 4.1. *Let Conditions* A, B *and* C *hold. Then we have*

$$n k_n^{-1/2}(\hat{\beta}^{(n)} - \hat{\beta}^{(n, I_{k_n})}) = \mathcal{O}_p(1). \tag{4.4}$$

PROOF. It follows from C.iii that we have

$$\sqrt{n}(\hat{\beta}^{(n)} - \hat{\beta}^{(n, I_{k_n})}) = \mathcal{O}_p(1) \quad \text{as} \quad n \to \infty$$

which is the assertion of the lemma for the case given in (4.3). So it suffices to prove (4.4) for the cases given in (4.1) and (4.2). Let us begin with (4.1). We are going to use Lemma 3.2.

Due to (3.3) and (3.7) it is clear that for the matrices $\{\mathcal{V}_n\}_{n=1}^{\infty}$

$$(\mathcal{V}_n)_{jk} = q_{jk} + \kappa_{jkn} + \frac{1}{\gamma}\Delta^*_{jkn}(\tilde{t}, \tilde{u}, \tilde{v})$$

(where $\kappa_{jkn} = \kappa_{jkn}(0)$ and $\Delta^*_{jkn}(\tilde{t}, \tilde{u}, \tilde{v}) = \Delta^*_{jkn}(0, \tilde{t}, \tilde{u}, \tilde{v})$) we can find for any $\Delta > 0$ and any $\varepsilon > 0$ an integer n_1 so that for any $n > n_1$ we have

$$P(\max_{1\le j,k\le p} |q_{jk} - (\mathcal{V}_n)_{jk}| > 2\Delta) < \varepsilon.$$

It means that the matrices $\mathcal{V}_n$ converge to Q in probability. Let us put $\tilde{t} = \sqrt{n}(\hat{\beta}^{(n)} - \beta^0)$, $\tilde{u} = \sqrt{n}(\hat{\beta}^{(n,I_{k_n})} - \hat{\beta}^{(n)})$ and $\tilde{v} = \sqrt{n}(\log\hat{\sigma}_n - \log\sigma)$. Due to Conditions C there is an integer $n_2 > n_1$ so that for any $n > n_2$ there is a set B_n with $P(B_n) > 1 - \varepsilon$ and for any $\omega \in B_n$ we have $\max\{\|\tilde{t}\|, \|\tilde{u}\| |\tilde{v}|\} < K$ for some K finite. Substituting $\tilde{t}$,$\tilde{u}$ and $\tilde{v}$ into (3.8) (for $\tau = 0$) we obtain

$$\sum_{i=1}^{n}[\psi([Y_i - g(X_i, \hat{\beta}^{(n,I_{k_n})})]\hat{\sigma}_n^{-1})g'(X_i, \hat{\beta}^{(n,I_{k_n})}) - \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1})g'(X_i, \hat{\beta}^{(n)})]$$
$$+\gamma\mathcal{V}_n n(\hat{\beta}^{(n,I_{k_n})} - \hat{\beta}^{(n)}) = \mathcal{O}_p(1) \quad \text{as} \quad n \to \infty.$$

Taking into account (2.4) and (2.5) for the case given in (4.1) we arrive at

$$\begin{aligned}(4.5)\quad & \gamma\mathcal{V}_n n(\hat{\beta}^{(n)} - \hat{\beta}^{(n,I_{k_n})}) \\ & = \sum_{\ell\in I_{k_n}} \psi([Y_\ell - g(X_\ell, \hat{\beta}^{(n,I_{k_n})})]\hat{\sigma}_n^{-1})g'(X_\ell, \hat{\beta}^{(n,I_{k_n})}) + \mathcal{O}_p(1) \quad \text{as} \quad n \to \infty.\end{aligned}$$

Since $k_n = k$ and ψ as well as g' is bounded, the right hand side of (4.5) is $\mathcal{O}_p(1)$ as $n \to \infty$. Recalling once again that $k_n = k$ and employing the Assertion A.1 of Appendix, we conclude the proof of lemma for (4.1).

It remains to prove (4.4) for (4.2). In the same way as in the previous case we obtain again (4.5) but now, due to the fact that $k_n \to \infty$ for $n \to \infty$, we cannot claim that right hand side is $\mathcal{O}_p(1)$ as $n \to \infty$. Nevertheless, dividing (4.5) by $k_n^{1/2}$ we arrive at

$$\begin{aligned}& \gamma\mathcal{V}_n n \cdot k_n^{-1/2}(\hat{\beta}^{(n)} - \hat{\beta}^{(n,I_{k_n})}) \\ & = k_n^{-1/2} \sum_{\ell\in I_{k_n}} \psi([Y_\ell - g(X_\ell, \hat{\beta}^{(n,I_{k_n})})]\hat{\sigma}_n^{-1})g'(X_\ell, \hat{\beta}^{(n,I_{k_n})}) + o_p(1) \quad \text{as} \quad n \to \infty.\end{aligned}$$

Since $\sqrt{n}(\hat{\beta}^{(n,I_{k_n})} - \beta^0) = \mathcal{O}_p(1)$ and for $\nu \in (0,1)$, $n^{-\nu}k_n \to \lambda$, we have for large n $(\frac{k_n}{n})^{1/2} < 1$ and hence $k_n^{1/2}(\hat{\beta}^{(n,I_{k_n})} - \beta^0) = \sqrt{n}(\frac{k_n}{n})^{1/2}(\hat{\beta}^{(n,I_{k_n})} - \beta^0) < \sqrt{n}(\hat{\beta}^{(n,I_{k_n})} - \beta^0)$, i.e. $k_n^{1/2}(\hat{\beta}^{(n,I_{k_n})} - \beta^0) = \mathcal{O}_p(1)$ as $n \to \infty$. Similarly $k_n^{1/2}(\log\hat{\sigma}_{n-k_n} - \log\sigma) = \mathcal{O}_p(1)$ as $n \to \infty$.

So we may substitute into (3.19) $t = \sqrt{k_n}(\hat{\beta}^{(n,I_{k_n})} - \beta^0)$ and $v = \sqrt{k_n}(\log\hat{\sigma}_{n-k_n} - \log\sigma)$, and we obtain

$$\Bigg\| k_n^{-1/2} \sum_{\ell\in I_{k_n}} \{\psi([Y_\ell - g(X_\ell, \hat{\beta}^{(n,I_{k_n})})]\hat{\sigma}_n^{-1})g'(X_\ell, \hat{\beta}^{(n,I_{k_n})}) - \psi(e_\ell \cdot \sigma^{-1})g'(X_\ell, \beta^0)\}$$
$$+ k_n^{1/2}\gamma Q(\hat{\beta}^{(n,I_{k_n})} - \beta^0) + k_n^{1/2}\theta q(\log\hat{\sigma}_{n-k_n} - \log\sigma)\Bigg\| = o_p(1) \quad \text{as} \quad n \to \infty.$$

It implies that

$$(4.6)\qquad \begin{aligned}&\gamma\mathcal{V}_n n\cdot k_n^{-1/2}(\hat\beta^{(n)}-\hat\beta^{(n,I_{k_n})})\\&=k_n^{-1/2}\sum_{\ell\in I_{k_n}}\psi(e_\ell\cdot\sigma^{-1})g'(X_\ell,\beta^0)-k_n^{1/2}\gamma Q(\hat\beta^{(n,I_{k_n})}-\beta^0)\\&\quad -k_n^{1/2}\theta q(\log\hat\sigma_{n-k_n}-\log\sigma)+o_p(1).\end{aligned}$$

Now the terms $k_n^{1/2}\gamma Q(\hat\beta^{(n,I_{k_n})}-\beta^0)$ and $k_n^{1/2}\theta q(\log\hat\sigma_{n-k_n}-\log\sigma)$ can be written as

$$\left(\frac{k_n}{n-k_n}\right)^{1/2}\gamma\sqrt{n-k_n}Q(\hat\beta^{(n,I_{k_n})}-\beta^0)\qquad\text{and}$$
$$\left(\frac{k_n}{n-k_n}\right)^{1/2}\theta q\sqrt{n-k_n}(\log\hat\sigma_{n-k_n}-\log\sigma),$$

respectively. Since $k_n\cdot(n-k_n)^{-1}=o(1)$ and $\sqrt{n-k_n}(\hat\beta^{(n,I_{k_n})}-\beta^0)=\mathcal{O}_p(1)$ as well as $\sqrt{n-k_n}(\log\hat\sigma_{n-k_n}-\log\sigma)=\mathcal{O}_p(1)$ when $n\to\infty$, using central limit theorem (to cope with the first term of right hand side of (4.6)), we obtain

$$\gamma\mathcal{V}_n n\cdot k_n^{-1/2}(\hat\beta^{(n)}-\hat\beta^{(n,I_{k_n})})=\mathcal{O}_p(1)\qquad\text{as}\quad n\to\infty.$$

Finally, utilizing Assertion A.1 of Appendix once again, we conclude the proof of lemma. □

The following theorems give the Bahadur representation of $\hat\beta^{(n)}-\hat\beta^{(n,I_{k_n})}$ for the three types of sequences $\{k_n\}_{n=1}^{\infty}$ which were indicated above.

THEOREM 4.1. *Let Conditions* A, B *and* C *be fulfilled, and let* (4.1) *hold. Then if* $\psi_s\equiv 0$

$$\begin{aligned}&n(\hat\beta^{(n,I_{k_n})}-\hat\beta^{(n)})\\&=-\gamma^{-1}Q^{-1}\sum_{i\in I_{k_n}}g'(X_i,\hat\beta^{(n,I_{k_n})})\psi([Y_i-g(X_i,\hat\beta^{(n,I_{k_n})})]\hat\sigma_n^{-1})+o_p(1)\\&\qquad\qquad\qquad\qquad\text{as}\quad n\to\infty.\end{aligned}$$

When $\psi_s\not\equiv 0$, *there exist a family of Wiener processes* $\mathcal{W}_j=W_j(y)$ *with* $y\in R^+$, *a sequence of stopping times* $\mu_{ijn}(t,u,v)$ $j=1,2,\dots,p$, $i=1,2,\dots,n$, $n\in N$, $t,u,v\in\mathcal{S}_M$, *and a sequence of random vectors* $\mathcal{R}_n\in R^p$, *such that uniformly in* $I_{k_n}\subset\{1,2,\dots,n\}$

$$\max_{1\le j\le p}\sup_{\mathcal{S}_M}\left|W_j\left(\sum_{i=1}^n\mu_{ijn}(t,u,v)\right)\right|=\mathcal{O}_p(1)\qquad\text{as}\quad n\to\infty,$$
$$(\mathcal{R}_n)_j=_{\mathcal{D}}W_j\left(\sum_{i=1}^n\mu_{ijn}(\sqrt{n}(\hat\beta^{(n)}-\beta^0),n(\hat\beta^{(n,I_{k_n})}-\hat\beta^{(n)}),\sqrt{n}(\log\hat\sigma_n-\log\sigma))\right)$$

and

$$(4.7)\qquad \begin{aligned}&n(\hat\beta^{(n,I_{k_n})}-\hat\beta^{(n)})\\&=-\gamma^{-1}Q^{-1}\left\{\sum_{i\in I_{k_n}}g'(X_i,\hat\beta^{(n,I_{k_n})})\psi([Y_i-g(X_i,\hat\beta^{(n,I_{k_n})})]\hat\sigma_n^{-1})+\mathcal{R}_n\right\}\\&\quad+o_p(1)\qquad\text{as}\quad n\to\infty.\end{aligned}$$

PROOF. We shall prove the second assertion of the theorem. It will be clear from what follows that the proof on the first one is similar but simpler, and hence it will be omitted.

From Lemma 4.1 we have $\hat{u}_n = n(\hat{\beta}^{(n,I_{k_n})} - \hat{\beta}^{(n)}) = \mathcal{O}_p(1)$ and from Conditions C $\hat{t}_n = n^{1/2}(\hat{\beta}^{(n)} - \beta^0) = \mathcal{O}_p(1)$ and $\hat{v}_n = n^{1/2}(\log \hat{\sigma}_n - \log \sigma) = \mathcal{O}_p(1)$. We are going to employ Lemma 3.1 for $\tau = \frac{1}{2}$. Let us put

$$(\mathcal{R}_n)_j = -\sum_{i=1}^{n} [\psi([Y_i - g(X_i, \hat{\beta}^{(n,I_{k_n})})]\hat{\sigma}_n^{-1}) g'_j(X_i, \hat{\beta}^{(n,I_{k_n})}) - \psi([Y_i - g(X_i, \hat{\beta}^{(n)})]\hat{\sigma}_n^{-1}) g'_j(X_i, \hat{\beta}^{(n)})] + \gamma \cdot n \cdot \sum_{\ell=1}^{p} q_{j\ell}(\hat{\beta}_\ell^{(n)} - \hat{\beta}_\ell^{(n,I_{k_n})}), \tag{4.8}$$

then due to symmetry of the distribution of Wiener process, due to the fact that

$$n \sum_{\ell=1}^{p} \kappa_{j\ell n}\left(\frac{1}{2}\right)(\hat{\beta}_\ell^{(n)} - \hat{\beta}_\ell^{(n,I_{k_n})}) = o_p(1)$$

and $\mathcal{K}_{jn}(\frac{1}{2}, \hat{t}_n, \hat{u}_n, \hat{v}_n) = o_p(1)$, we have from Lemma 3.1 for $j = 1, 2, \ldots, p$

$$(\mathcal{R}_n)_j =_{\mathcal{D}} W_j\left(\sum_{i=1}^{n} \mu_{ijn}\left(\frac{1}{2}, \hat{t}_n, \hat{u}_n, \hat{v}_n\right)\right) \tag{4.9}$$

(see (3.6)). Taking into account (2.4) and (2.5), we obtain from (4.8)

$$\sum_{i \in I_{k_n}} \psi([Y_i - g(X_i, \hat{\beta}^{(n,I_{k_n})})]\hat{\sigma}_n^{-1}) g'(X_i, \hat{\beta}^{(n,I_{k_n})}) - \gamma Q \cdot n(\hat{\beta}^{(n)} - \hat{\beta}^{(n,I_{k_n})}) + \mathcal{R}_n = o_p(1) \quad \text{as} \quad n \to \infty$$

and the proof of the theorem follows. □

Now let us consider the case described by (4.2).

THEOREM 4.2. *Let Conditions* A, B *and* C *be fulfilled, and let* (4.2) *hold. Then*

$$n k_n^{-1/2}(\hat{\beta}^{(n,I_{k_n})} - \hat{\beta}^{(n)}) = -k_n^{-1/2} \lambda \gamma^{-1} Q^{-1} \sum_{i \in I_{k_n}} g'(X_i, \beta^0)\, \psi(e_i \cdot \sigma^{-1}) + o_p(1) \qquad \textit{as} \quad n \to \infty \tag{4.10}$$

and hence $n k_n^{-1/2}(\hat{\beta}^{(n)} - \hat{\beta}^{(n,I_{k_n})})$ *is asymptotically distributed according to* $N_p(\mathbf{0}, \mathbf{\Sigma})$ *where* $\mathbf{\Sigma} = \gamma^{-2} \operatorname{var} \psi(e_1 \cdot \sigma^{-1}) \cdot Q^{-1}$.

PROOF. From Lemma 4.1 we have $\hat{u}_n = n k_n^{-1/2}(\hat{\beta}^{(n,I_{k_n})} - \hat{\beta}^{(n)}) = \mathcal{O}_p(1)$. From C.ii and C.iii it follows that $\hat{t}_n = n^{1/2}(\hat{\beta}^{(n)} - \beta^0) = \mathcal{O}_p(1)$ and $\hat{v}_n = n^{1/2}(\log \hat{\sigma}_n - \log \sigma) = \mathcal{O}_p(1)$. Notice that for $\tau_1 = \frac{1}{2}(1-\nu) > 0$, we have for the quantities from Lemma 3.2 $\kappa_{jkn}(\tau_1) = o_p(1)$, $\Delta^*_{jkn}(\tau_1, \hat{t}_n, \hat{u}_n, \hat{v}_n) = o_p(1)$ as well as $\mathcal{K}_{jn}(\tau_1, \hat{t}_n, \hat{u}_n, \hat{v}_n) = o_p(1)$.

Taking into account (2.4) and (2.5), we obtain from (3.8) for τ_1 and any $j = 1, 2, \ldots, p$

$$\sum_{i \in I_{k_n}} \psi([Y_i - g(X_i, \hat\beta^{(n,I_{k_n})})]\hat\sigma_n^{-1}) g_j'(X_i, \hat\beta^{(n,I_{k_n})})$$
$$+ n^{1/2\nu} \sum_{k=1}^{p} [\gamma(q_{jk} + \kappa_{jkn}(\tau_1)) + \Delta^*_{jkn}(\tau_1, \hat t_n, \hat u_n, \hat v_n)] n k_n^{-1/2} (\hat\beta_k^{(n,I_{k_n})} - \hat\beta_k^{(n)})$$
$$= \mathcal{O}_p(1)$$

and hence also

$$n^{-1/2\nu} \sum_{i \in I_{k_n}} \psi([Y_i - g(X_i, \hat\beta^{(n,I_{k_n})})]\hat\sigma_n^{-1}) g_j'(X_i, \hat\beta^{(n,I_{k_n})})$$
$$+ \sum_{k=1}^{p} \gamma q_{jk} n k_n^{-1/2} (\hat\beta_k^{(n,I_{k_n})} - \hat\beta_k^{(n)}) = o_p(1).$$

Taking into account (4.2)

$$n k_n^{-1/2} (\hat\beta^{(n,I_{k_n})} - \hat\beta^{(n)})$$
$$= -k_n^{-1/2} \lambda \gamma^{-1} Q^{-1} \sum_{i \in I_{k_n}} g'(X_i, \hat\beta^{(n,I_{k_n})}) \psi([Y_i - g(X_i, \hat\beta^{(n,I_{k_n})})]\hat\sigma_n^{-1})$$
$$+ o_p(1) \quad \text{as} \quad n \to \infty.$$

Utilizing now consistency of the estimator $\hat\beta^{(n,I_{k_n})}$ (see C.iii) and the assumptions about the functions $g(x, \beta)$, $\psi(z)$ and $F(z)$, we obtain (4.10). An application of the central limit theorem concludes the proof. □

It remains to treat the case when $\lim_{n\to\infty} k_n n^{-1} = \lambda \in (0, \frac{1}{2}]$. An analysis of this case may be done in a similar way as it was used in Theorems 1 and 2. However it seems simpler to use directly Lemmas 5.1, 5.2 and 7.1 of Víšek (1996*a*).

THEOREM 4.3. *Let Conditions* A, B *and* C *hold, and let* (4.3) *take place. Then*

$$n^{1/2} (\hat\beta^{(n,I_{k_n})} - \hat\beta^{(n)}) = -\gamma^{-1} Q^{-1} \lambda^{1/2} \sum_{i=1}^{n} a_{in} \psi(e_i \sigma^{-1}) g'(X_i, \beta^0) + o_p(1) \tag{4.11}$$

as $n \to \infty$ *where* $a_{in} = k_n^{-1/2}$ *for* $i \in I_{k_n}$ *and* $a_{in} = \lambda^{1/2} \cdot [(1-\lambda)(n - k_n)]^{-1/2}$ *for* $i \in \{1, 2, \ldots, n\} \setminus I_{k_n}$, *and hence* $\sqrt{n}(\hat\beta^{(n,I_{k_n})} - \hat\beta^{(n)})$ *has the asymptotic distribution* $N(\mathbf{0}, \mathbf{\Sigma}^*)$ *where*

$$\mathbf{\Sigma}^* = \frac{\lambda}{1-\lambda} \gamma^{-2} \operatorname{var} \psi\left(e_1 \sigma^{-1}\right) Q^{-1}.$$

PROOF. First of all, let us show that

$$\left\| n^{-1/2} \sum_{i=1}^{n} \psi(e_i \cdot \sigma^{-1}) g'(X_i, \beta^0) - n^{1/2} (n - k_n)^{-1} \right.$$
$$\left. \cdot \sum_{i \in \{1,2,\ldots,n\} \setminus I_{k_n}} \psi(e_i \cdot \sigma^{-1}) g'(X_i, \beta^0) - \gamma Q n^{1/2} (\hat\beta^{(n)} - \hat\beta^{(n,I_{k_n})}) \right\| = o_p(1). \tag{4.12}$$

Taking into account (2.4) and (2.5) we may bound the left hand side of (4.12) by

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

$$- \psi(e_i\cdot\sigma^{-1})g'(X_i,\beta^0)\} + \gamma Q n^{1/2}(\hat\beta^{(n)}-\beta^0) + \theta q n^{1/2}(\log\hat\sigma_n - \log\sigma)\Bigg\|$$

$$+\left[\frac{n}{n-k_n}\right]^{1/2}\Bigg\|(n-k_n)^{-1/2}\sum_{i\in\{1,2,\ldots,n\}\setminus I_{k_n}}\{\psi([Y_i - g(X_i,\hat\beta^{(n,I_{k_n})})]\hat\sigma_n^{-1})$$

$$\cdot g'(X_i,\hat\beta^{(n,I_{k_n})}) - \psi(e_i\cdot\sigma^{-1})g'(X_i,\beta^0)\} + \gamma Q(n-k_n)^{1/2}$$

$$\cdot(\hat\beta^{(n,I_{k_n})}-\beta^0) + \theta q(n-k_n)^{1/2}(\log\hat\sigma_n - \log\sigma)\Bigg\| + o_p(1),$$

and taking into account C.iii we find that the both norms are $o_p(1)$ according to Lemma 3.3. Finally, let us observe that (4.12) may be written as

$$\Bigg\|\sqrt{\frac{k_n}{n}}\cdot\frac{1}{\sqrt{k_n}}\sum_{i\in I_{k_n}}\psi(e_i\cdot\sigma^{-1})g'(X_i,\beta^0) - \sqrt{\frac{k_n}{n-k_n}}\sqrt{\frac{k_n}{n}}\cdot\frac{1}{\sqrt{n-k_n}}$$

$$\times\sum_{i\in\{1,2,\ldots,n\}\setminus I_{k_n}}\psi(e_i\cdot\sigma^{-1})g'(X_i,\beta^0) - \gamma Q n^{1/2}(\hat\beta^{(n)} - \hat\beta^{(n,I_{k_n})}))\Bigg\| = o_p(1).$$

Using central limit theorem and A.iii, we conclude the proof. □

Remark 4.2. Notice that multiplying (4.11) by $\lambda^{-1/2}$ and then assuming limit for $\lambda \to 0$, we obtain formally (4.10). It meams that for $\lambda \to 0$ Theorem 3.3 agrees formally with Theorem 3.2.

5. Conclusions

The result which we have obtained for the situation when a fix number of observations is deleted from the data, shows that the M-estimators which are generated by a discontinuous ψ-function may exhibit larger changes (after deletion of a few observations) than the estimators generated by a continuous function. Moreover, the form of the terms in corresponding relations hint that the change may be considerable and let us add that results of processing real data sets confirm it, see Víšek (1996*a*). It implies quite clear conclusion: *We should avoid M-estimators generated by discontinuous ψ-function.*

Of course, when the difference of the estimates before and after deletion of some points seems to be large, one can ask whether the change is significant or not. The results for the situations when together with increasing number of observations we allow also for an increase of number of deleted points, bring a possibility to test this significance. Naturally, when applying these results we cannot decide whether we have deleted e.g. asymptotically negligible part of data, since we have at hand a given number of observations and we delete a part of them. Nevertheless, if our analysis indicates that the subset of influential points may include more than a few observations, we can assume that the results for one of the two latter situations will work and we can probably use corresponding tests (see Benáček *et al.* (1998)).

Since all the results of paper are given for nonlinear regression, let us say now a few words about their relation to the linear regression model. Of course, the linear regression model is a special case of the nonlinear one, but.... This "but" means that the assumptions which may be acceptable for nonlinear regression, need not seem to be acceptable for the linear one. An example of such assumption may (seemingly) be A.ii. It is clear that this assumption would be fulfilled if we accept

$$\sup_{i\in N} \|x_i\| < K \tag{5.1}$$

for some $K > 0$. However, this assumption is considered by some statisticians as inadmissibly restricting while they are willing to accept the assumptions of the type

$$\sum_{i=1}^{n} \|x_i\|^k = \mathcal{O}(n) \tag{5.2}$$

Judge *et al.* (1985) or Jurečková and Sen (1993). First of all, let us look on the "difference of generality" between (5.1) and (5.2). Assertion A.2 says that if (5.2) is fulfilled for $k \geq 1$, for any $\Delta > 0$ there is a constant K_Δ so that the portion of observations, norm of which is larger than K_Δ, is asymptotically under Δ. Further, as we have seen, e.g. in (3.44), we need, at least for ψ-functions of types ψ_c and ψ_s, to evaluate probability of the type $P(u \leq e_i \leq u - n^{-1/2}x_i^{\mathrm{T}}t)$ and in fact to be able to carry out the proofs, we would like to have the upper bound for this probability given by $\|x_i\|\mathcal{O}(n^{-1/2})$. Since in the case of (5.2), the norms $\|x_i\|$, $i = 1, 2, \ldots, n$ are not uniformly bounded, we need some assumption(s) about $F(z)$ (e.g. existence of bounded density) to be fulfilled on the whole support of $F(z)$. Moreover, the proofs are rather involving while under (5.1) they are straightforward and it is evidently sufficient to require corresponding properties of $F(z)$ only in the neighborhood of u (where e.g. when the behaviour of regression quantiles is studied, $u = 0$, see Jurečková (1984)). If we take into account that under the contamination, it is already rather strong to ask for local properties of $F(z)$, it seems even more unrealistic to assume something about global properties of $F(z)$. On the other hand, although for given data we have always bounded maximum of the norms of carriers, a large ratio of

$$\sup_{i\in n} \|x_i\| / \operatorname*{med}_{i\in n} \|x_i\|$$

(where n is number of observations) may indicate that to assume that (5.1) holds, can be a little bit hazardous. Nevertheless, had we accepted (5.1), all the results of this paper are valid for linear regression, too.

So, let us summarize. Theorems 1, 2 and 3 allow us to find for the given M-estimator of the regression model the group of the most influential points. Moreover, in the case when $k_n \to \infty$ as $n \to \infty$, applying central limit theorem we may find the significance of the difference $\hat{\beta}^{(n)} - \hat{\beta}^{(n,I_{k_n})}$. We may call the estimate which is not for the full data *significantly* different from the estimate for some "reasonable" subsets of data, the *subsample stable*.

If however the difference appears to be significant, something may be wrong with our estimates. For instance, we have used ψ-function which is not adequate for the data, e.g.

the ψ-function is too far from the derivative of loglikelihood of the underlying model. So we should repeat the analysis with some other estimator. In fact, just described situation may be interpreted from a somewhat more general point of view. We shall try to explain it.

Although the robust statistics offer wide scale of estimating methods, most statisticians do not probably expect that we may obtain (considerably) different estimates of model as a result of application of different methods, of course we mean on one fix sample of data, see e.g. Ruppert and Carroll (1980). The belief that the estimates (i.e. the values of different estimators) should be similar for given data set, has probably its roots in a misinterpretation of the consistency and in a belief into an (objectively) existing mechanism somewhere behind the data which we (at least asymptotically) may discover, as Chrystofer Columbus discovered the America. But such belief may be sometimes problematic, sometimes it may be clearly demonstrated that it is an illusion (see Prigogine and Stengers (1984)). Any case, the experiences from the applications of robust procedures on the real data, especially the applications of procedures with high breakdown point, confirm the possibility of obtaining considerably different results (see e.g. Hettmansperger and Sheather (1992) or Víšek (1992*a*, 1993)). In Víšek (1994) an artificial data showed that the least median of squares and the least trimmed squares estimates may be orthogonal to each other. Due to the simplicity of these data, the reasons which caused this effect, can be traced out. Let us call this effect *diversity of estimates*. Similarly, it is not difficult to demonstrate an analogous behavior for the minimal bias estimators studied in Martin *et al.* (1989), namely surprisingly large "bias" of this estimator in some situations (converted commas hint that sometimes we are in fact sure that the estimate is considerably biased, sometimes it is far from other estimate with high breakdown point; for the discussion with more details see Víšek (1996*b*). Let us note that it is even possible to give a formalization of the *diversity of estimates* which shows that it is not contradictory to the consistency of the estimators, see Víšek (1997*a*, 2000).

It may be surprising but we may meet with *diversity of estimates* more frequently than we would expect. The fact that the most of statisticians have not yet this experience is due to the bad availability of implementations of the estimators with high breakdown point. Nevertheless when we meet with the *diversity of estimates*, the test for the *subsample stability* may help to select an appropriate estimate for given data, namely that one which exhibits the highest subsample stability (with, of course, insignificant difference $\hat{\beta}^{(n)} - \hat{\beta}^{(n,I_{k_n})}$). For a detailed discussion of the topic see Víšek (1992*b* or 1994).

Acknowledgements

We would like to express our gratitude to the anonymous referees for carefully reading the manuscript. They found at least several places where their suggestions improved comprehensibility of the text and they pointed out considerably misleading misprints. They also in fact initiated the remark on the relation of present results to the linear regression model.

Appendix

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

$$\text{(A.1)} \qquad \lim_{n\to\infty} v_{ij}^{(n)} = q_{ij} \quad \text{in probability}$$

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

$$\mathcal{V}^{(n)}\theta^{(n)} = \mathcal{O}_p(1) \quad \text{as} \quad n \to \infty.$$

Then

$$\theta^{(n)} = \mathcal{O}_p(1) \quad \text{as} \quad n \to \infty.$$

PROOF. For the proof see Víšek (1996*a*). □

LEMMA A.1. (Štěpán (1987), p. 420, VII.2.8) *Let a and b be positive numbers. Further let ξ be a random variable such that $P(\xi = -a) = \pi$ and $P(\xi = b) = 1-\pi$ (for a $\pi \in (0,1)$) and $I\!E\xi = 0$. Moreover let τ be the time for the Wiener process $W(s)$ to exit the interval $(-a, b)$. Then $\xi =_{\mathcal{D}} W(\tau)$ where "$=_{\mathcal{D}}$" denotes the equality of distributions of the corresponding random variables. Moreover, $I\!E\tau = a \cdot b = \operatorname{var}\xi$.*

Remark A.1. Since the book of Štěpán (1987) is in Czech language we refer also to Breiman (1968) where however this simple assertion is not isolated. Nevertheless, the assertion can be found directly in the first lines of the proof of Proposition 13.7 (p. 277) of Breinman's book. (See also Theorem 13.6 on the p. 276.)

Assertion A.2. Let for some positive k we have $\sum_{i=1}^{n} \|x_i\| = \mathcal{O}(n)$. Then for any $\Delta \in (0,1]$ there is a K_Δ such that denoting for any $n \in N$

$$m_n = \#\{i : 1 \le i \le n, \|x_i\| > K_\Delta\}$$

we have $m_n < \Delta \cdot n$ (where "$\#A$" denotes the number of elements of the set A).

PROOF. Due to the assumptions of lemma there is C such that for all $n \in N$ we have $\frac{1}{n}\sum_{i=1}^{n} \|x_i\| < C$. Fix $\Delta \in (0,1]$ and put $K_\Delta = \frac{C}{\Delta} + 1$. Then

$$C > \frac{1}{n}\sum_{i=1}^{n} \|x_i\| = \frac{1}{n}\left\{ \sum_{\{i:\|x_i\|\le K_\Delta\}} \|x_i\| + \sum_{\{i:\|x_i\|>K_\Delta\}} \|x_i\| \right\} > \frac{1}{n} m_n K_\Delta$$

and hence $m_n < n \cdot \frac{C}{K_\Delta} < n\Delta$.

REFERENCES

Atkinson, A. C. (1985). *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis*, Claredon Press, Oxford.

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Source of Collinearity*, Wiley, New York.

Benáček, V., Jarolím, J. and Víšek, J. Á (1998). Supply-side characteristics and the industrial structure of Czech foreign trade, *Proceedings of the Conference on Business and Economic Development in Central and Eastern Europe: Implications for Economic Intergration into Wider Europe*, 51–68, Technical University in Brno together with University of Wisconsin, Whitewater, and the Notingham Trent University, ISBN 80-214-1202-X, Brno.

Bickel, P. J. (1975). One-step Huber estimates in the linear model, *J. Amer. Statist. Assoc.*, **70**, 428–433.

Breiman, L. (1968). *Probability*, Addison-Wesley, London.

Chatterjee, S. and Hadi, A. S. (1988). *Sensitivity Analysis in Linear Regression*, Wiley, New York.

Csörgö, M. and Révész, P. (1981). *Strong Approximation in Probability and Statistics*, Akademia Kiadó, Budapest.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics–The Approach Based on Influence Functions*, Wiley, New York.

Hettmansperger, T. P. and Sheather, S. J. (1992). A Cautionary Note on the Method of Least Median Squares, *Amer. Statist.*, **46**, 79–83.

Judge, G. G., Griffiths, W. E., Hill, R. C., Liitkepohl, H. and Lee, T. C. (1985). *The Theory and Practice of Econometrics*, 2nd ed., Wiley, New York.

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

Jurečková, J. and Procházka, B. (1994). Regression quantiles and trimmed least squares estimator in nonlinear regression model, *J. Nonparametr. Statist.*, **3**, 201–222.

Jurečková, J. and Sen, P. K. (1993). Regression rank scores scale statistics and studentization in linear models, *Asymptotic Statistics, Proceedings of the Fifth Prague Symposium on Asymptotic Statistics*, 111–121, Physica, Heidelberg.

Liese, F. and Vajda, I. (1994). Consistency of M-estimators in general models, *J. Multivariate Anal.*, **50**, 93–114.

Martin, R. D., Yohai, V. J. and Zamar, R. H. (1989). Min-max bias robust regression, *Ann. Statist.*, **17**, 1608–1630.

Prigogine, I. and Stengers, I. (1984). *Out of Chaos*, William Heinemann, London.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*, Wiley, New York.

Rubio, A. M. and Víšek, J. Á. (1996). A note on asymptotic linearity of M-statistics in nonlinear models, *Kybernetika*, **32**, 353–374.

Rubio, A. M., Quintana, F. and Víšek, J. Á. (1993). Test for differences between M-estimates of nonlinear regression model, *Probab. Math. Statist.*, **14**(2), 189–206.

Ruppert, D. and Carroll, R. J. (1980). Trimmed least squares estimation in linear model, *J. Amer. Statist. Assoc.*, **75** (372), 828–838.

Štěpán, J. (1987). *Teorie pravdpodobnosti* (Theory of probability—in Czech), Academia, Praha.

Víšek, J. Á. (1992*a*). Stability of regression model estimates with respect to subsamples, *Comput. Statist.*, **2**, 183–203.

Víšek, J. Á.(1992*b*). Problems connected with selection of robust procedure, *Proceedings of PROBASTAT'91* (eds. A. Pázman and J. Volaufová), 189–203, Printing House of the Technical University of Liptovský Mikuláš.

Víšek, J. Á. (1993). On the role of contamination level and the least favorable behaviour of gross-error sensitivity, *Probab. Math. Statist.*, **14**(2), 173–187.

Víšek, J. Á. (1994). High robustness and an illusion of truth, *Transactions of ROBUST'94* (eds. J. Antoch and G. Dohnal), 172–185, JČMF (Union of Czech mathematicians), Prague.

Víšek, J. Á. (1996*a*). Sensitivity analysis of M-estimates, *Ann. Inst. Statist. Math.*, **48**, 469–495.

Víšek, J. Á.(1996*b*). On high breakdown point estimation, *Comput. Statist.*, **96**, 137– 146.

Víšek, J. Á. (1997*a*). Contamination level and sensitivity of robust tests, *Handbook of Statist.* (eds. G. S. Maddala and C. R. Rao), **15**, 633–642, Elsevier Science, Amsterdam.

Víšek, J. Á. (1997*b*). Diagnostics of regression subsample stability, *Probab. Math. Statist.*, **17**(2), 231–257.

Víšek, J. Á. (1999). Robust estimation of regression model, *Bulletin of the Czech Econometric Society*, **9**, 57–79.

Víšek, J. Á. (2000). On the diversity of estimates, *Comput. Statist. Data Anal.*, **34**, 67–89.

Zvára, K. (1989). *Regresní Analýza* (Regression Analysis), Academia, Prague (in Czech).