

Marek Omelka

## Second-order linearity of Wilcoxon statistics

Received: 5 July 2005 / Revised: 9 December 2005 / Published online: 28 June 2006  
© The Institute of Statistical Mathematics, Tokyo 2006

**Abstract** The rank statistic  $S_n(\mathbf{t}) = 1/n \sum_{i=1}^n c_i R_i(\mathbf{t})$  ( $\mathbf{t} \in \mathbb{R}^p$ ), with  $R_i(\mathbf{t})$  being the rank of  $e_i - \mathbf{t}^\top \mathbf{x}_i$ ,  $i = 1, \dots, n$  and  $e_1, \dots, e_n$  being the random sample from a distribution with a cdf  $F$ , is considered as a random process with  $\mathbf{t}$  in the role of parameter. Under some assumptions on  $c_i$ ,  $\mathbf{x}_i$  and on the underlying distribution, it is proved that the process  $\{S_n(\frac{\mathbf{t}}{\sqrt{n}}) - S_n(\mathbf{0}) - \mathbf{E} S_n(\mathbf{t}), \|\mathbf{t}\|_2 \leq M\}$  converges weakly to the Gaussian process. This generalizes the existing results where the one-dimensional case  $\mathbf{t} \in \mathbb{R}$  was considered. We believe our method of the proof can be easily modified for the signed-rank statistics of Wilcoxon type. Finally, we use our results to find the second order asymptotic distribution of the  $R$ -estimator based on the Wilcoxon scores and also to investigate the length of the confidence interval for a single parameter  $\beta_l$ .

**Keywords** Rank statistics · Asymptotic linearity · Empirical processes · U-processes

### 1 Introduction

Consider the linear regression model

$$Y_i = \alpha + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i = \alpha + \boldsymbol{\beta}^\top \mathbf{x}_i + e_i, \quad i = 1, \dots, n, \quad (1.1)$$

where  $\alpha$  and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$  are unknown parameters,  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^\top$ , for  $i = 1, \dots, n$  are known constants, and  $e_1, \dots, e_n$  are independent, identically distributed random variables with a cumulative distribution function  $F$ . Let  $R_i(\mathbf{b})$  be the rank of  $Y_i - \mathbf{b}^\top \mathbf{x}_i$  among  $Y_1 - \mathbf{b}^\top \mathbf{x}_1, \dots, Y_n - \mathbf{b}^\top \mathbf{x}_n$  and  $\bar{\mathbf{x}}_n = (\bar{x}_{n1}, \dots, \bar{x}_{np})^\top$

---

M. Omelka

Department of Probability and Mathematical Statistics, Charles University,  
Sokolovská 83, 186 00 Prague 8, Czech Republic  
E-mail: omelka@karlin.mff.cuni.cz

be the vector of the means of the columns of the design matrix  $\mathbf{X}$ . Then the  $R$ -estimator  $\hat{\boldsymbol{\beta}}_R$  (based on the Wilcoxon scores) of  $\boldsymbol{\beta}$  can be defined as the solution of the following minimization

$$\sum_{j=1}^p |S_{nj}(\mathbf{b})| := \min, \quad \text{where} \quad S_{nj}(\mathbf{b}) = \frac{1}{n^{3/2}} \sum_{i=1}^n (x_{ij} - \bar{x}_{nj}) R_i(\mathbf{b}).$$

Or more informally we can say that  $\hat{\boldsymbol{\beta}}_R$  ‘almost’ solves the system of equations

$$\mathbf{S}_n(\mathbf{b}) = (S_{n1}(\mathbf{b}), \dots, S_{np}(\mathbf{b}))^\top = \frac{1}{n^{3/2}} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}}_n) R_i(\mathbf{b}) = 0.$$

The keystone for the inference about  $R$ -estimators is the uniform asymptotic linearity of the statistic  $S_n(\mathbf{b})$ . This means that the difference  $T_n(\mathbf{b}) = \mathbf{S}_n(\frac{\mathbf{b}}{\sqrt{n}} + \boldsymbol{\beta}) - \mathbf{S}_n(\boldsymbol{\beta})$  differs from a linear function by an amount which tends to zero in probability as the number of observations increases (Jurečková 1971). For the case of a one-dimensional parameter  $\beta$  and Wilcoxon scores, Jurečková (1973) showed that if the difference between the statistic  $T_n(b)$  and the linear form is multiplied by a suitable constant (usually  $\sqrt{n}$ ), one obtains stochastic process which converges weakly to a linear process. This result was further generalized for the Wilcoxon signed-rank statistics by Antille (1976) and for some other types of score functions by Hušková (1980), Puri and Wu (1985) and Kersting (1987). We will generalize the results of Jurečková (1973) for the case of a multi-dimensional parameter  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ . Our approach can be also easily modified for the Wilcoxon-signed rank statistics.

The rest of this paper is organized as follows. In Section 2 we state the general assumptions and the main theorems. Section 3 contains the proof of the asymptotic representation of the leading term of the process, while the asymptotic negligibility of the remainder term is considered in Section 4. Section 4 also contains the proof of Theorem 2.2. Section 5 uses the results of the preceding sections. In the first, part we present the second order asymptotic representation for the  $R$ -estimator  $\hat{\boldsymbol{\beta}}_R$  and in the second part we find the asymptotic distribution of the properly standardized length of the confidence interval for a single parameter.

## 2 Notations, assumptions, theorems

### 2.1 Notations

Let  $\mathbf{t} = (t_1, \dots, t_p)$  and write  $R'_i(\mathbf{t})$  for the rank of  $e_i - \mathbf{t}^\top \mathbf{x}_i / \sqrt{n}$  among  $e_1 - \mathbf{t}^\top \mathbf{x}_1 / \sqrt{n}, \dots, e_n - \mathbf{t}^\top \mathbf{x}_n / \sqrt{n}$ . In the following we will be interested in the processes

$$\tilde{S}_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n c_i R'_i(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n c_i \sum_{j=1}^n \mathbb{I}\{e_i - \frac{\mathbf{t}^\top \mathbf{x}_i}{\sqrt{n}} \geq e_j - \frac{\mathbf{t}^\top \mathbf{x}_j}{\sqrt{n}}\} \quad (2.1)$$

$$T_n(\mathbf{t}) = \tilde{S}_n(\mathbf{t}) - \tilde{S}_n(\mathbf{0}), \quad (2.2)$$

$$Y_n(\mathbf{t}) = T_n(\mathbf{t}) - \mathbf{E} T_n(\mathbf{t}), \quad (2.3)$$

with  $\mathbf{t} \in T = \{\mathbf{s} \in \mathbb{R}^p : |\mathbf{s}|_2 \leq M\}$ , where  $|\cdot|_2$  stands for the Euclidian norm, and  $M$  is an arbitrary large but fixed constant.

### 2.2 Assumptions

We will need the following assumptions on the distribution function  $F$ , the design  $\mathbf{x}_1, \dots, \mathbf{x}_n$  and the constants  $c_1, \dots, c_n$ .

**(F.1)**  $F$  has a bounded and uniformly continuous derivative  $f = F'$ .

**(F.2)**

$$\lim_{\Delta \rightarrow 0} \frac{1}{\Delta^2} \int_{-\infty}^{+\infty} \int_{-\Delta}^{+\Delta} [f(z+y) - f(y)]^2 dz dy = 0.$$

We note that the condition **(F.1)** is for our convenience to make the proofs simple and it could be weakened. A slightly weaker assumption is used in Omelka (2006). On the other hand the densities  $f$  which satisfy condition **(F.2)** but do not satisfy the condition **(F.1)** are rather curious.

According to Antille (1976) the condition **(F.2)** is satisfied in these two important cases:

(i)  $f$  is such that  $|f(x+t) - f(x)| \leq |t|^\alpha h(x)$ , with  $\alpha > \frac{1}{2}$  and  $h(x) \in L_2(-\infty, +\infty)$

(ii)  $f$  is absolutely continuous and  $f'(x) \in L_2(-\infty, +\infty)$ .

We note that the second condition is satisfied if there exists a finite Fisher information of the density  $f$ .

Let  $|\cdot|_2$  stands for the Euclidian norm.

**(X.1)**

$$\sum_{i=1}^n c_i = 0, \quad \frac{1}{n} \sum_{i=1}^n c_i^2 = 1, \quad \lim_{n \rightarrow \infty} \frac{\max_{1 \leq i \leq n} c_i^2}{\sum_{i=1}^n c_i^2} = 0,$$

**(X.1)**

$$\sum_{i=1}^n \mathbf{x}_i = \mathbf{0}, \quad \frac{1}{n} \sum_{i=1}^n |\mathbf{x}_i|_2^2 = O(1), \quad \lim_{n \rightarrow \infty} \frac{\max_{1 \leq i \leq n} |\mathbf{x}_i|_2}{\sqrt{n}} = 0$$

**(X.1)**

$$\lim_{n \rightarrow \infty} \max_{1 \leq i \leq n} \frac{|c_i| |\mathbf{x}_i|_2}{\sqrt{n}} = 0$$

**(X.1)**

$$B_n^2 = \frac{1}{n} \sum_{i=1}^n c_i^2 |\mathbf{x}_i|_2^2 = O(1)$$

The conditions **(X.1)**–**(X.3)** are analogous to the conditions in Jurečková (1973). The last condition **(X.?)** is again only for convenience. If  $B_n^2 = O(1)$  was not satisfied, we would work with the process  $\tilde{S}_n(\mathbf{t}) = (1/nB_n) \sum_{i=1}^n x_i R'_i(\mathbf{t})$  and derive analogous results.

2.3 Theorems

Put  $\gamma = \int f^2(x)dx$ .

**Theorem 2.1** *Under conditions (X.1)–(X.4) and (F.1)–(F.2) the process  $\{Y_n(\mathbf{t}), \mathbf{t} \in T\}$  satisfies uniformly in  $\mathbf{t} \in T$*

$$Y_n(\mathbf{t}) = -\frac{\mathbf{t}^\top}{\sqrt{n}} \sum_{i=1}^n \left( c_i \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^n c_j \mathbf{x}_j \right) (f(e_i) - \gamma) + o_p(1). \tag{2.4}$$

Specially, if we put

$$\mathbf{A}_n^2 = \frac{1}{n} \sum_{i=1}^n c_i^2 \mathbf{x}_i \mathbf{x}_i^\top + 3 \left( \frac{1}{n} \sum_{i=1}^n c_i \mathbf{x}_i \right) \left( \frac{1}{n} \sum_{i=1}^n c_i \mathbf{x}_i^\top \right),$$

then the process  $Y'_n(\mathbf{t}) = Y_n(\mathbf{A}_n^{-1}\mathbf{t})$  converges in distribution to the centered Gaussian process  $\{Y(\mathbf{t}), \mathbf{t} \in T\}$  with the covariance function  $\text{cov}\{Y(\mathbf{t}), Y(\mathbf{s})\} = \sigma^2 \mathbf{t}^\top \mathbf{s}$ , where  $\sigma^2 = \int f^3(x)dx - \gamma^2$ .

**Theorem 2.2** *Suppose that  $f$  satisfies the assumption (F.2). Then uniformly in  $\mathbf{t} \in T$*

$$\mathbf{E} T_n(\mathbf{t}) = -\frac{\gamma \mathbf{t}^\top}{\sqrt{n}} \sum_{i=1}^n c_i \mathbf{x}_i + o(1).$$

**Corollary 2.1** *Under conditions (X.1)–(X.4) and (F.1)–(F.2) it holds uniformly in  $\mathbf{t} \in T$*

$$\begin{aligned} \tilde{S}_n(\mathbf{t}) - \tilde{S}_n(\mathbf{0}) + \frac{\gamma \mathbf{t}^\top}{\sqrt{n}} \sum_{i=1}^n c_i \mathbf{x}_i &= -\frac{\mathbf{t}^\top}{\sqrt{n}} \sum_{i=1}^n \left( c_i \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^n c_j \mathbf{x}_j \right) \\ &\quad (f(e_i) - \gamma) + o_p(1). \end{aligned} \tag{2.5}$$

2.4 Preliminaries

In order to prove Theorem 2.1, we approximate the process  $Y_n(t)$  by its projection (see e.g. Serfling 1980)

$$Y_n^{(1)} = \hat{Y}_n = \sum_{i=1}^n \mathbf{E}[Y_n | X_i] - (n - 1)\mathbf{E} Y_n.$$

We will show that the projection  $Y_n^{(1)}$  has the asymptotic representation (2.4) and that the remainder term  $Y_n^{(2)} = Y_n - Y_n^{(1)}$  is asymptotically negligible.

### 3 The convergence of the process $Y_n^{(1)}$

Calculating the projection of the process  $Y_n$  we find out that  $Y_n^{(1)}(\mathbf{t}) = Z_n(\mathbf{t}) - \mathbf{E} Z_n(\mathbf{t})$ , where

$$Z_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (c_i - c_j) \left[ F \left( e_i - \frac{\mathbf{t}^\top (\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}} \right) - F(e_i) \right] = \sum_{i=1}^n Z_{ni}(\mathbf{t}). \tag{3.1}$$

In the following we will check the conditions of the Jain–Marcus theorem [originally in Jain and Marcus (1975), restated in van der Vaart and Wellner (1996), p. 213]. Suppose we have an index set  $\mathcal{F}$  equipped with the semimetric  $\rho$ . Then the covering number  $N(\varepsilon, \mathcal{F}, \rho)$  is the minimal number of balls of radius  $\varepsilon$  needed to cover the set  $\mathcal{F}$ .

**Theorem 3.1** *Let  $Z_{n1}, \dots, Z_{nk_n}$  be independent stochastic processes indexed by an arbitrary index set  $\mathcal{F}$  such that*

$$|Z_{ni}(f) - Z_{ni}(g)| \leq M_{ni} \rho(f, g), \quad \text{for every } f, g \in \mathcal{F}, \tag{3.2}$$

where  $M_{n1}, \dots, M_{nk_n}$  are independent random variables and  $\rho$  is a semimetric on  $\mathcal{F}$  such that

$$\int_0^\infty \sqrt{\log N(\varepsilon, \mathcal{F}, \rho)} \, d\varepsilon < \infty \tag{3.3}$$

where  $N(\varepsilon, \mathcal{F}, \rho)$  is the covering number for the index class  $\mathcal{F}$ , and

$$\sum_{i=1}^{k_n} \mathbf{E} M_{ni}^2 = O(1). \tag{3.4}$$

If the triangular array also satisfies the Lindeberg condition

$$\sum_{i=1}^{k_n} \mathbf{E} \|Z_{ni}\|_{\mathcal{F}}^2 \mathbb{I}_{\{\|Z_{ni}\|_{\mathcal{F}} > \varepsilon\}} \rightarrow 0, \quad \text{for all } \varepsilon > 0, \tag{3.5}$$

where  $\|Z_{ni}\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |Z_{ni}(f)|$ , then the sequence  $\sum_{i=1}^{k_n} (Z_{ni} - \mathbf{E} Z_{ni})$  converges in distribution in  $\ell^\infty(\mathcal{F})$  (the space of bounded functions) to a tight Gaussian process provided the sequence of covariance functions converges pointwise on  $\mathcal{F} \times \mathcal{F}$ .

In our situation the index set is quite simple  $\mathcal{F} = T = \{\mathbf{t} : |\mathbf{t}|_2 \leq M\}$  with the Euclidian metric  $\rho(\mathbf{t}, \mathbf{s}) = |\mathbf{t} - \mathbf{s}|_2 = \sqrt{\sum_{i=1}^p (t_i - s_i)^2}$ . This implies  $N(\varepsilon, \mathcal{F}, \rho) \leq \max\{(6M/\varepsilon)^p, 1\}$  [see e.g. Lemma 4.1 in Pollard (1990)] and so the condition (3.3) is satisfied.

By **(F.1)**  $K = \sup_x \{f(x)\} < \infty$ . Now calculate

$$\begin{aligned}
 |Z_{ni}(\mathbf{t}) - Z_{ni}(\mathbf{s})| &\leq \frac{1}{n} \sum_{j=1}^n |c_i - c_j| \left| F\left(e_i - \frac{\mathbf{t}^\top \mathbf{x}_i}{\sqrt{n}}\right) - F\left(e_i - \frac{\mathbf{s}^\top \mathbf{x}_i}{\sqrt{n}}\right) \right| \\
 &\leq \frac{K}{n^{3/2}} \sum_{j=1}^n |c_i - c_j| |(\mathbf{t} - \mathbf{s})^\top (\mathbf{x}_i - \mathbf{x}_j)| \\
 &\leq \frac{K}{n^{3/2}} \sum_{j=1}^n (|c_i| + |c_j|) (|\mathbf{x}_i|_2 + |\mathbf{x}_j|_2) \rho(\mathbf{t}, \mathbf{s}) \\
 &\leq \frac{K\rho(\mathbf{t}, \mathbf{s})}{n^{3/2}} \left[ n|c_i| |\mathbf{x}_i|_2 + \sum_{j=1}^n |c_j| |\mathbf{x}_i|_2 + |c_i| \sum_{j=1}^n |\mathbf{x}_j|_2 \right. \\
 &\quad \left. + \sum_{j=1}^n |c_j| |\mathbf{x}_j|_2 \right] \\
 &\leq \frac{K\rho(\mathbf{t}, \mathbf{s})}{n^{1/2}} [(|c_i| + 1) |\mathbf{x}_i|_2 + |c_i| O(1) + O(1)] \\
 &= M_{ni} \rho(\mathbf{t}, \mathbf{s}), \tag{3.6}
 \end{aligned}$$

which implies that the conditions (3.2) and (3.4) are met. Moreover, as

$$\|Z_{ni}\|_T = \sup_{\mathbf{t} \in M} |Z_{ni}(\mathbf{t})| \leq \frac{KM}{\sqrt{n}} [(|c_i| + 1) |\mathbf{x}_i|_2 + O(1)] \xrightarrow{n \rightarrow \infty} 0,$$

the Lindeberg condition (3.5) is satisfied as well. Now Theorem 3.1 implies that the process  $\bar{Z}_n = Z_n - \mathbf{E} Z_n$  is asymptotically tight. We can repeat the previous steps with the process

$$\begin{aligned}
 Z'_n(\mathbf{t}) &= \frac{1}{n} \sum_{j=1}^n (c_i - c_j) \left[ F\left(e_i - \frac{\mathbf{t}^\top (\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}\right) - F(e_i) \right] \\
 &\quad + \frac{\mathbf{t}^\top}{\sqrt{n}} \sum_{i=1}^n \left( c_i \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^n c_j \mathbf{x}_j \right) (f(e_i) - \gamma),
 \end{aligned}$$

to find out that the process  $\bar{Z}'_n = Z'_n - \mathbf{E} Z'_n$  is asymptotically tight as well. Moreover from the expansion

$$\begin{aligned}
 Z_{ni}(\mathbf{t}) &= \frac{1}{n} \sum_{j=1}^n (c_i - c_j) \left( F\left(e_i - \frac{\mathbf{t}^\top \mathbf{x}_i}{\sqrt{n}}\right) - F(e_i) \right) \\
 &= - \sum_{j=1}^n \frac{c_i - c_j}{n} \left[ f(e_i) \mathbf{t}^\top (\mathbf{x}_i - \mathbf{x}_j) + \frac{|\mathbf{x}_i|_2 + |\mathbf{x}_j|_2}{\sqrt{n}} o(1) \right]
 \end{aligned}$$

$$\begin{aligned}
 &= -\frac{f(e_i)\mathbf{t}^\top}{n^{3/2}} \sum_{j=1}^n (c_i \mathbf{x}_i - c_i \mathbf{x}_j - c_j \mathbf{x}_i + c_j \mathbf{x}_j) + o\left(\frac{1}{\sqrt{n}}\right) \\
 &= -\frac{f(e_i)\mathbf{t}^\top}{\sqrt{n}} \left( c_i \mathbf{x}_i + \frac{1}{n} \sum_{j=1}^n c_j \mathbf{x}_j \right) + o\left(\frac{1}{\sqrt{n}}\right), \tag{3.7}
 \end{aligned}$$

it follows that for every fixed  $\mathbf{t} \in T$

$$\text{var}\{\bar{Z}'_n(\mathbf{t})\} = \sum_{i=1}^n \text{var}\{Z'_{ni}(\mathbf{t})\} \xrightarrow{n \rightarrow \infty} 0, \tag{3.8}$$

which implies that the process  $\bar{Z}'_n$  converges to zero marginally. This marginal convergence and the just proved asymptotic tightness give us  $\sup_{\mathbf{t} \in T} |\bar{Z}'_n(\mathbf{t})| = o_p(1)$ , which proves the statement of the theorem.

#### 4 Asymptotic negligibility of $Y_n^{(2)}$

In the following, we will show that

$$\|Y_n^{(2)}\|_T = \sup_{\mathbf{t} \in T} |Y_n^{(2)}(\mathbf{t})| = o_p(1). \tag{4.1}$$

To prove (4.1) we will adapt the theory of U-processes introduced in Nolan and Pollard (1987). For convenience we will denote this reference as NP. Let us recall that  $Y_n^{(2)} = Y_n - Y_n^{(1)}$ . At first using (3.1) we notice that  $Y_n^{(2)}(\mathbf{t}) = U_n(\mathbf{t}) - \mathbf{E} U_n(\mathbf{t})$ , where

$$\begin{aligned}
 U_n(\mathbf{t}) &= \frac{1}{n} \sum_{i=1}^n c_i \sum_{j=1}^n \left[ \mathbb{I}\left\{e_i - \frac{\mathbf{t}^\top \mathbf{x}_i}{\sqrt{n}} \geq e_j - \frac{\mathbf{t}^\top \mathbf{x}_j}{\sqrt{n}}\right\} - \mathbb{I}\{e_i \geq e_j\} \right] \\
 &\quad - \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \left[ (c_i - c_j) \left( F\left(e_i - \frac{\mathbf{t}^\top (\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}\right) - F(e_i) \right) \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n g_{ij}\left(e_i, e_j, \frac{\mathbf{t}}{\sqrt{n}}\right),
 \end{aligned}$$

with

$$\begin{aligned}
 g_{ij}(u, v, \mathbf{w}) &= \frac{c_i}{n} \left[ \mathbb{I}\{u - \mathbf{w}^\top \mathbf{x}_i \geq v - \mathbf{w}^\top \mathbf{x}_j\} - \mathbb{I}\{u \geq v\} \right] \\
 &\quad - \frac{c_i - c_j}{n} \left[ F(u - \mathbf{w}^\top (\mathbf{x}_i - \mathbf{x}_j)) - F(u) \right].
 \end{aligned}$$

In the following, it will be more convenient to index the process  $Y_n^{(2)}$  with the class of functions  $\mathcal{G} = \{g_{\mathbf{t}}, \mathbf{t} \in T\}$ . Let us define  $R_n(g_{\mathbf{t}}) := Y_n^{(2)}(\mathbf{t}) = U_n(g_{\mathbf{t}}) - \mathbf{E} U_n(g_{\mathbf{t}})$ . The proof will be divided into several steps.

#### 4.1 Symmetrization

The first step is the symmetrization of the process  $R_n(\mathbf{g}_t)$ . Let  $e'_1, \dots, e'_n$  be independent copies of  $e_1, \dots, e_n$ . Denote

$$\begin{aligned} U'_n(\mathbf{g}_t) &= \sum_{i=1}^n \sum_{j=1}^n g_{ij}(e'_i, e_j, \frac{\mathbf{t}}{\sqrt{n}}), & R'_n(\mathbf{g}_t) &= U'_n(\mathbf{g}_t) - \mathbf{E} U'_n(\mathbf{g}_t) \\ U_n'(\mathbf{g}_t) &= \sum_{i=1}^n \sum_{j=1}^n g_{ij}(e_i, e'_j, \frac{\mathbf{t}}{\sqrt{n}}), & R_n'(\mathbf{g}_t) &= U_n'(\mathbf{g}_t) - \mathbf{E} U_n'(\mathbf{g}_t) \\ U''_n(\mathbf{g}_t) &= \sum_{i=1}^n \sum_{j=1}^n g_{ij}(e'_i, e'_j, \frac{\mathbf{t}}{\sqrt{n}}), & R''_n(\mathbf{g}_t) &= U''_n(\mathbf{g}_t) - \mathbf{E} U''_n(\mathbf{g}_t). \end{aligned}$$

With the help of these processes we define the symmetrized process

$$\begin{aligned} R_n^{\text{sym}}(\mathbf{g}_t) &= R_n(\mathbf{g}_t) - R'_n(\mathbf{g}_t) - R_n'(\mathbf{g}_t) + R''_n(\mathbf{g}_t) \\ &= \sum_{i=1}^n \sum_{j=1}^n \left[ g_{ij}(e_i, e_j, \frac{\mathbf{t}}{\sqrt{n}}) - g_{ij}(e'_i, e_j, \frac{\mathbf{t}}{\sqrt{n}}) - g_{ij}(e_i, e'_j, \frac{\mathbf{t}}{\sqrt{n}}) \right. \\ &\quad \left. + g_{ij}(e'_i, e'_j, \frac{\mathbf{t}}{\sqrt{n}}) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n g_{ij}^{\text{sym}}(e_i, e_j, e'_i, e'_j, \frac{\mathbf{t}}{\sqrt{n}}) = R_n(\mathbf{g}_t^{\text{sym}}). \end{aligned} \quad (4.2)$$

This process has the same distribution as the process

$$R_n^\sigma(\mathbf{g}_t^{\text{sym}}) = \sum_{i=1}^n \sum_{j=1}^n \sigma_i \sigma_j g_{ij}^{\text{sym}}(e_i, e_j, e'_i, e'_j, \frac{\mathbf{t}}{\sqrt{n}}),$$

where  $\sigma_1, \dots, \sigma_n$  are Rademacher random variables. Let us introduce

$$R_n^\circ(\mathbf{g}_t) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sigma_i \sigma_j g_{ij}(e_i, e_j, \frac{\mathbf{t}}{\sqrt{n}}).$$

Then it holds

$$\mathbf{E} \|R_n(\mathbf{g}_t)\|_{\mathcal{G}} \leq \mathbf{E} \|R_n(\mathbf{g}_t^{\text{sym}})\|_{\mathcal{G}} = \mathbf{E} \|R_n^\sigma(\mathbf{g}_t^{\text{sym}})\|_{\mathcal{G}} \leq 4 \mathbf{E} \|R_n^\circ(\mathbf{g}_t)\|_{\mathcal{G}},$$

where the first inequality is a complete analogy of Lemma 1 in NP (the important thing is that the process  $R_n(\mathbf{g}_t)$  is degenerated in the sense that its projection is a zero process) and the second inequality is a simple triangular inequality. Next put

$$g_{ij}^{(1)} = \frac{c_i}{n} \left[ \mathbb{I} \left\{ e_i - \frac{\mathbf{t}^\top \mathbf{x}_i}{\sqrt{n}} \geq e_j - \frac{\mathbf{t}^\top \mathbf{x}_j}{\sqrt{n}} \right\} - \mathbb{I} \{ e_i \geq e_j \} \right] \quad (4.3)$$

$$g_{ij}^{(2)} = \frac{c_i - c_j}{n} \left[ F \left( e_i - \frac{\mathbf{t}^\top (\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}} \right) - F(e_i) \right]. \quad (4.4)$$



One more application of the triangular inequality yields

$$\mathbf{E} \|R_n^\circ(g_t)\|_{\mathcal{G}} \leq \mathbf{E} \|R_n^\circ(g_t^{(1)})\|_{\mathcal{G}} + \mathbf{E} \|R_n^\circ(g_t^{(2)})\|_{\mathcal{G}}.$$

In the sequel we will show that  $\mathbf{E} \|R_n^\circ(g_t^{(1)})\|_{\mathcal{G}} = o_P(1)$ . The proof for the process  $R_n^\circ(g_t^{(2)})$  is completely analogous. For the simplicity of notation  $g_t$  will henceforward stand for  $g_t^{(1)}$ .

### 4.2 Exponential inequality

The second step is an exponential inequality. Denote  $\mathbf{E}_\sigma$  the operator of the expected value induced by the random variables  $\sigma_1, \dots, \sigma_n$  (we condition on  $e_1, \dots, e_n$ ).

As Lemma 3 in NP it is shown:

**Lemma 4.1** *Let  $\sigma_1, \dots, \sigma_n$  be independent sign variables for which  $P[\sigma = +1] = P[\sigma = -1] = \frac{1}{2}$ . Then for each real symmetric matrix  $\mathbf{A} = [a_{ij}]$  with  $\sum_{i=1}^n \sum_{j \neq i}^n a_{ij}^2 \leq \frac{1}{4\pi^2}$*

$$\mathbf{E}_\sigma \exp \left( \sum_{i=1}^n \sum_{j \neq i}^n \sigma_i \sigma_j a_{ij} \right) \leq \exp \left( \frac{\pi^2}{2} \sum_{i=1}^n \sum_{j \neq i}^n a_{ij}^2 \right).$$

But the following simple calculation shows that this Lemma is actually true for an arbitrary real square matrix  $\mathbf{A}$  (with  $\sum_{i=1}^n \sum_{j \neq i}^n a_{ij}^2 \leq \frac{1}{4\pi^2}$ ), as

$$\begin{aligned} \mathbf{E}_\sigma \exp \left( \sum_{i=1}^n \sum_{j \neq i}^n \sigma_i \sigma_j a_{ij} \right) &= \mathbf{E}_\sigma \exp \left( \sum_{i=1}^n \sum_{j \neq i}^n \sigma_i \sigma_j \frac{a_{ij} + a_{ji}}{2} \right) \\ &\stackrel{\text{Lemma 4.1}}{\leq} \exp \left( \sum_{i=1}^n \sum_{j \neq i}^n \frac{\pi^2}{2} \frac{(a_{ij} + a_{ji})^2}{4} \right) \\ &\leq \exp \left( \frac{\pi^2}{2} \sum_{i=1}^n \sum_{j \neq i}^n a_{ij}^2 \right). \end{aligned} \tag{4.5}$$

### 4.3 Chaining and the maximal inequality

In the third step we will make use of the technique known as chaining. Let  $(S, d)$  be an index class equipped with the pseudometric  $d(\cdot, \cdot)$ . Write  $N(\varepsilon, S, d)$  for the covering number of the class  $S$ . We will make use of the following Lemma 5 in NP.

**Lemma 4.2** *Let  $\Psi$  be a convex, strictly increasing function on  $[0, \infty)$  with  $0 \leq \Psi(0) \leq 1$ . Suppose that the stochastic process  $Z$  indexed by the class  $(S, d)$  satisfies:*

- (i) if  $d(s, t) = 0$ , then  $Z(s) = Z(t)$  almost surely;
- (ii) if  $d(s, t) > 0$ , then  $\mathbf{E} \Psi(Z(s) - Z(t)) / d(s, t) \leq 1$ ;
- (iii) there exists a point  $s_0 \in S$  for which  $\sup_{s \in S} d(s, s_0) < \infty$
- (iv) the sample paths of  $Z$  are continuous on  $(S, d)$ .

Then

$$\mathbf{E} \sup_{s \in S} |Z(s) - Z(s_0)| \leq 8 \int_0^\theta \Psi^{-1}(N(x, S, d)) \, dx,$$

where  $\theta$  equals one quarter of the supremum in (iii).

Define the (random) pseudometric  $d_\omega$  on the space  $T$  as

$$d_\omega(\mathbf{t}, \mathbf{s}) = \left( \sum_{i=1}^n \sum_{j=1}^n \left[ g_{ij} \left( e_i(\omega), e_j(\omega), \frac{\mathbf{t}}{\sqrt{n}} \right) - g_{ij} \left( e_i(\omega), e_j(\omega), \frac{\mathbf{s}}{\sqrt{n}} \right) \right]^2 \right)^{1/2}.$$

To make use of Lemma 4.2 we need to find an upper bound for the (random) covering numbers  $N(x, T, d_\omega)$ . For this reason we will use the technique of pseudodimension introduced in Pollard (1990) – EP. Consider

$$h_{ij}(\omega, \mathbf{s}) = \mathbb{I}\{e_i(\omega) - e_j(\omega) \geq \mathbf{s}^T(\mathbf{x}_i - \mathbf{x}_j)\} - \mathbb{I}\{e_i(\omega) - e_j(\omega) \geq 0\}.$$

Using e.g. Lemma 4.4 of EP we can deduce that the subset of the space  $\mathbb{R}^{n(n-1)}$

$$\mathcal{H}_{n\omega} = \{(h_{ij}(\omega, \mathbf{s}), 1 \leq i \neq j \leq n), \mathbf{s} \in \mathbb{R}^p\}$$

has for all  $\omega \in \Omega$  uniformly bounded pseudodimension. Now set

$$\boldsymbol{\alpha} = (\alpha_{ij}, 1 \leq i \neq j \leq n) = \left( \frac{|c_i|}{n}, 1 \leq i \neq j \leq n \right)$$

and let  $\boldsymbol{\alpha} \odot \mathbf{h}$  stand for the pointwise product in  $\mathbb{R}^{n(n-1)}$  with  $k$ th coordinate  $\alpha_k h_k$ . Then

$$N(\varepsilon, T, d_\omega) \leq N(\varepsilon, \boldsymbol{\alpha} \odot \mathcal{H}_{n\omega}, |\cdot|_2) \tag{4.6}$$

As  $|h_{ij}(\omega, \mathbf{s})| \leq 1$  we will take the vector  $\mathbf{H} = (1, 1, \dots, 1)$  as the envelope for  $\mathcal{H}_{n\omega}$ . Notice that uniformly for all  $n$

$$|\boldsymbol{\alpha} \odot \mathbf{H}|_2^2 = \sum_{i=1}^n \sum_{j \neq i}^n \frac{c_i^2}{n^2} \leq \frac{1}{n} \sum_{i=1}^n c_i^2 = 1.$$

Now Corollary 4.10 of EP guarantees the existence of universal constants  $A$  and  $W$  such that for all  $\omega$  and for all  $\varepsilon$  ( $0 < \varepsilon \leq 1$ ) and also for all  $n \in \mathbb{N}$

$$N(\varepsilon, \boldsymbol{\alpha} \odot \mathcal{H}_{n\omega}, |\cdot|_2) \leq A \left( \frac{1}{\varepsilon} \right)^W.$$

Combining this inequality with the inequality (4.6) yields  $N(\varepsilon, T, d_\omega) \leq A(\frac{1}{\varepsilon})^W$ . But this further implies that the covering integral

$$J_{n\omega}(s) = \int_0^s \log(N(x, T, d_\omega)) dx$$

is uniformly bounded for all  $\omega$  and  $J_n(s, \omega) \rightarrow 0$  for  $s \rightarrow 0$  uniformly in  $\omega \in \Omega$ .

For simplicity of notation write  $T_n$  for the measure that places mass one at each of the  $n(n - 1)$  pairs  $(e_i, e_j) (i \neq j)$  measure 1. Using the notation common in empirical processes we will write

$$T_n(f^2) = \sum_{i=1}^n \sum_{j \neq i}^n f_{ij}^2(e_i, e_j).$$

Now we are ready to formulate the analogy of Theorem 6 of NP.

**Lemma 4.3** *There exists a universal constant C such that for all  $n \in \mathbb{N}$*

$$\mathbf{E} \|R_n^\circ(g_t)\|_{\mathcal{G}} \leq C \mathbf{E} (\theta_n + J_n(\theta_n)), \tag{4.7}$$

where  $\theta_n^2 = (1/16)\|T_n(g_t^2)\|_{\mathcal{G}}$ .

*Proof* Set  $\Psi(x) = 1/2 \exp(x/2\pi - 1/8)$ . For a fixed  $\omega$  we verify that the process  $R_n^\circ(g_t)$  meets the conditions of Lemma 4.2. The only nonobvious condition is (ii). Put  $f = g_t - g_s$  and with the help of Lemma 4.1 calculate

$$\begin{aligned} \mathbf{E}_\sigma \exp\left(\frac{|R_n^\circ(g_t) - R_n^\circ(g_s)|}{2\pi d(t, s)}\right) &= \mathbf{E}_\sigma \exp\left(\frac{|R_n^\circ(f)|}{2\pi \sqrt{T_n(f^2)}}\right) \\ &\leq \mathbf{E}_\sigma \exp\left(\frac{R_n^\circ(f)}{2\pi \sqrt{T_n(f^2)}}\right) + \mathbf{E}_\sigma \exp\left(\frac{-R_n^\circ(f)}{2\pi \sqrt{T_n(f^2)}}\right) \\ &\stackrel{\text{Lemma 4.1}}{\leq} 2 \exp\left(\frac{\pi^2 \sum_{i=1}^n \sum_{j \neq i} (g_{ij}(\frac{t}{\sqrt{n}}) - g_{ij}(\frac{s}{\sqrt{n}}))^2}{4\pi^2 T_n(f^2)}\right) = 4 \exp\left(\frac{1}{8}\right). \end{aligned}$$

This gives the desired exponential inequality  $\mathbf{E}_\sigma \Psi\left(\frac{|R_n^\circ(g_t) - R_n^\circ(g_s)|}{d(t, s)}\right) \leq 1$ . The choice  $s_0 = 0$  in Lemma 4.2 yields

$$\begin{aligned} \mathbf{E}_\sigma \|R_n^\circ\| &\leq 8 \int_0^{\theta_n} \psi^{-1}(N(x, T, d_\omega)) dx \\ &\leq 8 \int_0^{\theta_n} \frac{\pi}{4} + 2\pi \log(2N(x, T, d_\omega)) dx \leq C (\theta_n + J_{n\omega}(\theta_n)). \end{aligned}$$

Averaging out over the  $\omega$  gives inequality (4.7).

By virtue of Markov's inequality and inequality (4.7) to prove  $\|R_n^\circ(g_t)\|_{\mathcal{G}} = o_p(1)$  it suffices to verify that  $\theta_n \xrightarrow[n \rightarrow \infty]{P} 0$ .

Let us denote  $\varepsilon_n = (2M/\sqrt{n}) \max_{1 \leq i \leq n} \|\mathbf{x}_i\|_2$ . The condition **(X.2)** implies  $\varepsilon_n = o(1)$ . Observe that

$$\left[ g_{ij} \left( e_i, e_j, \frac{\mathbf{t}}{\sqrt{n}} \right) \right]^2 \leq \frac{c_i^2}{n^2} \mathbb{I}\{|e_i - e_j| \leq \varepsilon_n\}.$$

This further yields

$$\mathbb{E} \theta_n^2 \leq \sum_{i=1}^n \frac{c_i^2}{n} \mathbb{E} \mathbb{I}\{|e_1 - e_2| \leq \varepsilon_n\} \xrightarrow[n \rightarrow \infty]{} 0,$$

which implies  $\theta_n \xrightarrow[n \rightarrow \infty]{P} 0$ .

The proof of  $\mathbb{E} \|R_n^\circ(g_t)\|_{\mathcal{G}} = o(1)$  for  $g_t = g_t^{(2)}$  would be completely analogous.

*Proof of Theorem 2.2.* From the proof of Theorem 2.1 we can see that

$$\mathbb{E} T_n(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty} [F(y - \frac{\mathbf{t}^\top}{\sqrt{n}}(\mathbf{x}_i - \mathbf{x}_j)) - F(y)] dF(y) + o(1)$$

Analogously as in Antille (1976) we can calculate

$$\begin{aligned} D_n(\mathbf{t}) &= \frac{1}{n} \sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty} [F(y - \frac{\mathbf{t}^\top}{\sqrt{n}}(\mathbf{x}_i - \mathbf{x}_j)) \\ &\quad - F(y)] dF(y) + \gamma \frac{\mathbf{t}^\top}{\sqrt{n}} \sum_{i=1}^n c_i \mathbf{x}_i \\ &= \frac{1}{n} \sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty} \int_0^{\frac{\mathbf{t}^\top(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}} f(y - v)f(y) - f(y)^2 dv dy \\ &= \frac{1}{2n} \sum_{i=1}^n c_i \sum_{j=1}^n \int_{-\infty}^{+\infty} \int_0^{\frac{\mathbf{t}^\top(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}} [f(y - v) - f(y)]^2 dv dy. \end{aligned} \tag{4.8}$$

With the help of the conditions **(F.2)** and **(X.1)–(X.4)** we can see that for arbitrary  $\varepsilon > 0$  for all sufficiently large  $n$  it holds

$$\begin{aligned}
 |D_n(\mathbf{t})| &\leq \frac{\varepsilon}{2n} \left| \sum_{i=1}^n |c_i| \sum_{j=1}^n \frac{|\mathbf{t}|_2^2 |\mathbf{x}_i - \mathbf{x}_j|_2^2}{n} \right| \\
 &= \frac{\varepsilon M^2}{2n} \left| \sum_{i=1}^n |c_i| \left[ |\mathbf{x}_i|_2^2 + \frac{1}{n} \sum_{j=1}^n |\mathbf{x}_j|_2^2 \right] \right| = \varepsilon O(1). \tag{4.9}
 \end{aligned}$$

### 5 Applications

#### 5.1 Second order asymptotic representation of $\hat{\beta}_R$

We will use the asymptotic expansion (2.5) to find the second order asymptotic representation for the estimator  $\hat{\beta}_R$ . Assume that there exists a positive definite matrix  $\mathbf{V}$  such that

$$\mathbf{V} = \lim_{n \rightarrow \infty} \mathbf{V}_n = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^\top.$$

Then [according to Ren (1994)] when the conditions **(F.1)** and **(X.?)** are satisfied, the estimator  $\hat{\beta}_R$  admits the following first order representation

$$\sqrt{n}(\hat{\beta}_R - \beta) = \frac{\mathbf{V}_n^{-1}}{\gamma \sqrt{n}} \sum_{i=1}^n \mathbf{x}_i F(e_i) + o_p(1). \tag{5.1}$$

Let  $\mathbf{c} = (x_{1l}, \dots, x_{nl})^\top$  be the  $l$ -th column of the matrix  $\mathbf{X}_n$  and suppose for a moment that this column is orthogonal to the other columns of the matrix  $\mathbf{X}_n$ , that is  $\sum_{i=1}^n x_{il} x_{ij} = 0$  for  $l \neq j$ . Then  $\mathbf{t}^\top / \sqrt{n} \sum_{i=1}^n c_i \mathbf{x}_i = t_l / \sqrt{n} \sum_{i=1}^n x_{il}^2$ . Further put  $T_{nl}^2 = (1/n) \sum_{i=1}^n x_{il}^2$ . Now insert  $\mathbf{t} \rightarrow \sqrt{n}(\hat{\beta}_R - \beta)$  into Eq. (2.5). After some reorganization and using the fact  $S_n(\hat{\beta}_R) = o_p(1/\sqrt{n})$  (see Jaeckel 1972) we get

$$\begin{aligned}
 &\sqrt{n}(\hat{\beta}_l - \beta_l) - \frac{1}{\gamma T_{nl}^2 \sqrt{n}} \sum_{i=1}^n x_{il} \frac{R_i(\beta)}{n} \\
 &= -(\hat{\beta}_R - \beta)^\top \frac{1}{\gamma T_{nl}^2 \sqrt{n}} \sum_{i=1}^n \left( \mathbf{x}_i x_{il} + \mathbf{e}_1 \sum_{j=1}^n \frac{x_{jl}^2}{n} \right) \\
 &\quad \times (f(e_i) - \gamma) + o_p\left(\frac{1}{\sqrt{n}}\right) \\
 &\stackrel{(5.1)}{=} -\frac{1}{\sqrt{n}} \left\{ \frac{\mathbf{V}_n^{-1}}{\gamma \sqrt{n}} \sum_{i=1}^n \mathbf{x}_i F(e_i) \right\}^\top \\
 &\quad \left\{ \frac{1}{\gamma T_{nl}^2 \sqrt{n}} \sum_{i=1}^n (\mathbf{x}_i x_{il} + \mathbf{e}_1 T_{nl}^2) \times (f(e_i) - \gamma) \right\} + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{5.2}
 \end{aligned}$$

where  $\mathbf{e}_l = (0, \dots, 0, 1, 0, \dots, 0)^\top$  is a vector of zeroes with the only one nonzero element in the  $l$ -th coordinate. Both terms in the last equations are asymptotically multivariate normal and in the case  $\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^\top \mathbf{x}_i x_{il} \rightarrow 0$  also asymptotically independent.

With the help of the Cramér–Wold device there seems to be no problem to generalize the asymptotic representation (2.5) to the vector form. Put  $\tilde{\mathbf{S}}_n(\mathbf{t}) = (1/n) \sum_{i=1}^n \mathbf{x}_i R'_i(\mathbf{t})$ .

Then it holds

$$\tilde{\mathbf{S}}_n(\mathbf{t}) - \tilde{\mathbf{S}}_n(\mathbf{0}) + \gamma \sqrt{n} \mathbf{V}_n \mathbf{t} = -\frac{1}{\sqrt{n}} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^\top + \mathbf{V}_n) (f(e_i) - \gamma) \mathbf{t} + o_p(1). \tag{5.3}$$

Inserting  $\mathbf{t} \rightarrow \sqrt{n}(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})$  and after some reorganization we get

$$\begin{aligned} & \sqrt{n}(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) - \frac{\mathbf{V}_n^{-1}}{\gamma \sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \frac{R_i(\boldsymbol{\beta})}{n} \\ &= -\frac{\mathbf{V}_n^{-1}}{\gamma \sqrt{n}} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^\top + \mathbf{V}_n) (f(e_i) - \gamma) (\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) + o_p\left(\frac{1}{\sqrt{n}}\right) \\ &\stackrel{(5.1)}{=} -\frac{1}{\sqrt{n}} \left\{ \frac{\mathbf{V}_n^{-1}}{\gamma \sqrt{n}} \sum_{i=1}^n (\mathbf{x}_i \mathbf{x}_i^\top + \mathbf{V}_n) (f(e_i) - \gamma) \right\} \\ &\quad \times \left\{ \frac{\mathbf{V}_n^{-1}}{\gamma \sqrt{n}} \sum_{i=1}^n \mathbf{x}_i F(e_i) \right\} + o_p\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{5.4}$$

The expansions (5.2) and (5.4) can be used as one of the theoretical insights to compare different estimators, especially those which are first order asymptotic equivalent. In Omelka (2005) in a simple linear regression model an  $R$ -estimator  $\hat{\boldsymbol{\beta}}_R$  (based on the Wilcoxon scores) was compared with a  $M$ -estimator  $\hat{\boldsymbol{\beta}}_M$ . It is well known that if a  $M$ -estimator is generated by the function  $\psi(x) = c(F(x) - \frac{1}{2})$ , then  $\sqrt{n}(\hat{\boldsymbol{\beta}}_R - \hat{\boldsymbol{\beta}}_M) = o_p(1)$ . In Omelka (2005) the (nontrivial) asymptotic distribution of  $n(\hat{\boldsymbol{\beta}}_R - \hat{\boldsymbol{\beta}}_M)$  was found, which implies that the  $R$ -estimator and  $M$ -estimator are not second order equivalent.

Some numerical experiments show that the remainder term ( $o_p(\frac{1}{\sqrt{n}})$ ) in representations (5.2) and (5.4) multiplied by  $\sqrt{n}$  converges to zero rather slowly. Analogous experiments for  $M$ -estimators (work is in progress) show that the second order representations can be much more accurate (provided the  $\psi$ -function and the underlying distribution of errors are smooth enough). We guess that this is caused by the fact that a  $M$ -estimator can be a much smoother functional than an  $R$ -estimator. Nevertheless the work of Lachout and Paulauskas (2000) indicates that even in the case of very smooth  $M$ -estimators we can expect only a very slow rate of convergence in the second order asymptotic distributional representations.

5.2 Length of the confidence interval for a single parameter  $\beta_l$

For  $\mathbf{b} \in \mathbb{R}^p$  denote  $\mathbf{b}(t)$  the vector  $\mathbf{b}$  with the  $l$ th coordinate replaced by  $t$ , that is  $\mathbf{b}(t) = (b_1, \dots, b_{l-1}, t, b_{l+1}, \dots, b_p)$ . Next put  $S_{nl}^\circ(t) = S_{nl}(\hat{\beta}_R(t)) = \frac{1}{n^{3/2}} \sum_{i=1}^n x_{il} R_i(\hat{\beta}_R(t))$  and so  $S_{nl}^\circ(\hat{\beta}_l) = \frac{1}{n^{3/2}} \sum_{i=1}^n x_{il} R_i(\hat{\beta}_R)$ . Let  $w_{ll}$  be the  $l$ -th diagonal element of the matrix  $\mathbf{V}_n^{-1}$  and  $z_\alpha = \Phi^{-1}(1 - \frac{\alpha}{2})$ , with  $\Phi^{-1}$  being the inverse cdf of a standard normal distribution. Then the confidence interval for the parameter  $\beta_l$  can be constructed as  $D_n^l = [b_l^-, b_l^+]$  with

$$b_l^- = \sup\{t : S_{nl}^\circ(t) > \frac{T_{nl}^2 \sqrt{w_{ll}} z_\alpha}{\sqrt{12}}\}, \quad b_l^+ = \inf\{t : S_{nl}^\circ(t) < \frac{-T_{nl}^2 \sqrt{w_{ll}} z_\alpha}{\sqrt{12}}\}. \quad (5.5)$$

For simplicity we will suppose again that the  $l$ -th column of the matrix  $\mathbf{X}_n$  is orthogonal to the other columns of this matrix and  $T_{nl}^2 = \frac{1}{n} \sum_{i=1}^n x_{il}^2 = 1$ , which further implies  $w_{ll} = 1$  [the general case can be found in Omelka (2006)].

The following theorem only restates the results of Sect. 5 of Jurečková (1973).

**Theorem 5.1** *If the conditions of Theorem 2.1 are satisfied, the  $l$ th column of the matrix  $\mathbf{X}_n$  is orthogonal to the other columns,  $T_{nl}^2 = 1$  and  $\sqrt{n}(\hat{\beta}_R - \beta) = O_p(1)$  then the confidence interval  $D_n^l$  defined by (5.5) satisfies :*

(i)

$$P(D_n^l \ni \beta_l) \xrightarrow{n \rightarrow \infty} 1 - \alpha$$

(ii)

$$L_n = \sqrt{n}(b_l^+ - b_l^-) \left( \frac{\sqrt{3}\gamma}{z_\alpha} \right) - 1 \xrightarrow{n \rightarrow \infty} 0$$

(iii) *put  $A_{nl}^2 = \frac{1}{n} \sum_{i=1}^n x_{il}^4 + 3$ , then the standardized length of the confidence interval  $\sqrt{n}L_n/A_{nl}$  is asymptotically normally distributed with the parameters  $(0, \frac{\int f^3(x)dx - \gamma^2}{\gamma^2})$ .*

*Proof* From the uniform asymptotic linearity and with the help of our assumptions about the matrix  $\mathbf{X}_n$  it follows that uniformly for  $\|\mathbf{b}\|_2 \leq M$

$$\left| S_{nl} \left( \frac{\mathbf{b}}{\sqrt{n}} + \beta \right) - S_{nl}(\beta) + \gamma b_l \right| = o_p(1). \quad (5.6)$$

As  $\mathbf{b} = \sqrt{n}(\hat{\beta}_R(\beta_l) - \beta) = O_p(1)$  and the  $l$ -th coordinate of the vector  $\mathbf{b}$  is zero, this yields  $S_{nl}(\hat{\beta}_R(\beta_l)) = S_{nl}(\beta) + o_p(1)$ , which further implies

$$\begin{aligned} P(b_l^- > \beta_l) &= P \left( S_{nl}^\circ(\beta_l) > \frac{z_\alpha}{\sqrt{12}} \right) = P \left( S_{nl}(\hat{\beta}_R(\beta_l)) > \frac{z_\alpha}{\sqrt{12}} \right) \\ &= P \left( S_{nl}(\beta) + o_p(1) > \frac{z_\alpha}{\sqrt{12}} \right) \\ &= P \left( \frac{1}{n^{3/2}} \sum_{i=1}^n x_{il} R_i(\beta) + o_p(1) > \frac{z_\alpha}{\sqrt{12}} \right) \xrightarrow{n \rightarrow \infty} \frac{\alpha}{2}. \end{aligned}$$

Analogously, we could show that  $P(b_l^+ < \beta_l) \xrightarrow[n \rightarrow \infty]{} \frac{\alpha}{2}$ , which yields the statement (i) of the theorem.

To prove (ii), at first we need to show that

$$\sqrt{n}(b_l^- - \beta_l) = O_p(1) \quad \text{and} \quad \sqrt{n}(b_l^+ - \beta_l) = O_p(1). \quad (5.7)$$

For this reason let us calculate

$$\begin{aligned} P(\sqrt{n}(b_l^- - \beta_l) > t) &= P\left(S_{nl}^o\left(\beta_l + \frac{t}{\sqrt{n}}\right) > \frac{z_\alpha}{\sqrt{12}}\right) \\ &= P\left(S_{nl}\left(\hat{\beta}_R\left(\beta_l + \frac{t}{\sqrt{n}}\right)\right) > \frac{z_\alpha}{\sqrt{12}}\right) \\ &= P\left(S_{nl}(\beta) + \gamma t + o_p(1) > \frac{z_\alpha}{\sqrt{12}}\right) \xrightarrow[n \rightarrow \infty]{} 1 - \Phi(z_\alpha + \gamma t). \end{aligned}$$

Thus  $\sqrt{n}(b_l^- - \beta_l)$  is asymptotically normal. Analogously, we can prove that  $\sqrt{n}(b_l^+ - \beta_l)$  is asymptotical normal as well, which verifies (5.7). This enables us to insert  $\mathbf{b} \rightarrow \sqrt{n}(\hat{\beta}_R(b_l^+) - \beta)$  as well as  $\mathbf{b} \rightarrow \sqrt{n}(\hat{\beta}_R(b_l^-) - \beta)$  in the asymptotic linearity result (5.6). We get

$$\begin{aligned} S_{nl}(\hat{\beta}_R(b_l^+)) - S_{nl}(\beta) + \gamma\sqrt{n}(b_l^+ - \beta_l) &= o_p(1) \\ S_{nl}(\hat{\beta}_R(b_l^-)) - S_{nl}(\beta) + \gamma\sqrt{n}(b_l^- - \beta_l) &= o_p(1). \end{aligned}$$

Combining these two equations gives

$$\gamma\sqrt{n}(b_l^+ - b_l^-) = S_{nl}(\hat{\beta}_R(b_l^-)) - S_{nl}(\hat{\beta}_R(b_l^+)) + o_p(1) = \frac{2z_\alpha}{\sqrt{12}} + o_p(1),$$

which proves the statement (ii).

To show the last statement we insert successively  $\mathbf{b} \rightarrow \sqrt{n}(\hat{\beta}_R(b_l^+) - \beta)$  and then  $\mathbf{b} \rightarrow \sqrt{n}(\hat{\beta}_R(b_l^-) - \beta)$  in Eq. (2.5). Analogously as in the proof of the statement (ii) we combine the resulting equations to get

$$\begin{aligned} \gamma n(b_l^+ - b_l^-) - \frac{2z_\alpha}{\sqrt{12}}\sqrt{n} &= -\sqrt{n}(b_l^+ - b_l^-) \frac{1}{\sqrt{n}} \sum_{i=1}^n \left( x_{il}^2 + \frac{1}{n} \sum_{j=1}^n x_{jl}^2 \right) \\ &\quad \times (f(e_i) - \gamma) + o_p(1), \end{aligned}$$

which using the statement (ii) yields after some reorganizations

$$\begin{aligned} \sqrt{n} \left[ \sqrt{n}(b_l^+ - b_l^-) \frac{\sqrt{3}\gamma}{z_\alpha} - 1 \right] &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \left( x_{il}^2 + \frac{1}{n} \sum_{j=1}^n x_{jl}^2 \right) \\ &\quad \times (f(e_i) - \gamma) + o_p(1). \end{aligned}$$

And so the statement (iii) is proved as well.



5.3 Numerical example

We illustrate the results of Theorem 5.1 on the simulated regression model. As the design matrix we used the Meyer matrix of order  $27 \times 2$  (see Stigler 1986, pp. 16–25). Further we normalize this matrix such that  $\sum_{i=1}^n x_{ij} = 0$  and  $\frac{1}{n} \sum_{i=1}^n x_{ij}^2 = 1$  for  $j = 1, 2$ . We were interested in 95% confidence intervals for the parameter  $\beta_1$  constructed by the following three methods :

- 1 (LS) the traditional least squares estimate
- 2 (R I) the conf. interval based on the first order asymptotic representation (5.1)

$$D'_n = \left[ \hat{b}_1 - \frac{z_\alpha}{\sqrt{n}} \frac{\sqrt{w_{11}}}{\hat{\gamma} \sqrt{12}}, \hat{b}_1 + \frac{z_\alpha}{\sqrt{n}} \frac{\sqrt{w_{11}}}{\hat{\gamma} \sqrt{12}} \right], \tag{5.8}$$

where  $w_{11}$  is the first element of the diagonal of the matrix  $\mathbf{V}_n^{-1}$  and  $\hat{\gamma}$  is an appropriate estimate of  $\gamma$  (to estimate  $\gamma$  we have used the R-function `wilcoxontau`, see Terpstra and McKean 2004).

- 3 (R II) The confidence interval given by (5.5)

At first we were interested in small sample coverages and mean lengths of these intervals. Some of the results, for the errors generated from standard normal distribution ( $N(0, 1)$ ),  $t$ -distribution with five degrees of freedom ( $t_5$ ) and lognormal distribution ( $LN(0, 1)$ ), are to be found in Table 1. The first row of this table gives us the estimated coverage probability, the second row presents mean length of conf. intervals and the third row the variance of the lengths. We see that for symmetric errors the method R II gives considerably smaller conf. intervals, but at the cost of a slightly smaller than nominal coverage probability. We were surprised that for asymmetric errors, R I method performs better than R II in all aspects. Table 1 also confirms the very well known fact, that LS method is tied down to normal errors.

Secondly, we wanted to assess the statements (ii) and (iii) of Theorem 5.1. We chose the sample sizes  $n = 27, 54, 108$  and  $216$  (as the design matrix we use the appropriate multiples of Meyer matrix) and estimate the mean length of conf. interval (multiplied by  $\sqrt{n}$ ) and variance of this length (multiplied by  $n^2$ ). The results for errors following exponential distribution (with density  $f(x) = e^{-x}\mathbb{I}\{x > 0\}$ )

**Table 1** Results on confidence intervals for  $\beta_1$ ,  $n = 27$  10000 random samples

	$N(0, 1)$			$t_5$			$LN(0, 1)$		
	LS	R I	R II	LS	R I	R II	LS	R I	R II
Coverage	0.952	0.954	0.944	0.947	0.950	0.940	0.956	0.959	0.937
Mean (length)	0.787	0.852	0.784	1.005	0.983	0.915	1.537	0.832	0.835
Var (length)	0.013	0.025	0.019	0.053	0.042	0.036	0.762	0.054	0.073

**Table 2** Results on confidence intervals for  $\beta_1$ , for exponential errors 10000 random samples

	$n = 27$		$n = 54$		$n = 108$		$n = 216$		$n = \infty$
	R I	R II	R I	R II	R I	R II	R I	R II	
Coverage	0.950	0.935	0.958	0.947	0.953	0.946	0.956	0.951	0.950
$\sqrt{n}$ Mean (length)	4.002	3.089	2.765	2.813	2.587	2.646	2.484	2.534	2.277
$n^2$ Var (length)	14.07	16.29	11.12	13.98	9.540	12.51	8.556	11.18	8.068

can be found in Table 2. Comparing the finite sample results with their asymptotic values (the last column of the table) we see that to approximate the mean and especially variance of length of conf. intervals with their asymptotic values is too optimistic, even in the situations with more than one hundred observations and only two explanatory variables. But to be fair we chose one of the worst cases (exponential errors). For normal errors the asymptotic approximations work for  $n > 100$  quite satisfactory.

**Acknowledgements** The author is grateful to both anonymous referees whose comments led to a considerable improvement of the text. This research was supported by the Charles University Grant GAUK 341/2005 and by the Czech Republic Grant GAČR 201/05/H007.

## References

- Antille, A. (1976). Asymptotic linearity of Wilcoxon signed-rank statistics. *Annals of Statistics*, 4, 175–186.
- Hušková, M. (1980). On bounded length sequential confidence interval for parameter in regression model based on ranks. In: *Colloquia Mathematica Societatis János Bolyai*, (pp. 435–463). Amsterdam: North-Holland.
- Jaeckel, L.A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics*, 43, 1449–1458.
- Jain, N., Marcus, M. (1975). Central limit theorems for  $C(S)$ -valued random variables. *Journal of Functional Analysis*, 19, 216–231.
- Jurečková, J. (1971). Nonparametric estimation of regression coefficients. *Annals of Mathematical Statistics*, 42, 1328–1338.
- Jurečková, J. (1973). Central limit theorem for Wilcoxon rank statistics process. *Annals of Statistics*, 1, 1046–1060.
- Kersting, G. (1987). Second-order linearity of the general signed-rank statistics. *Journal of Multivariate Analysis*, 21, 274–295.
- Lachout, P., Paulauskas, V. (2000). On the second-order asymptotic distribution of  $M$ -estimators. *Statistics and Decision*, 18, 231–257.
- Nolan, D. Pollard, D. (1987).  $U$ -processes, rates of convergence. *Annals of Statistics*, 15, 780–799.
- Omelka, M. (2005). Second order asymptotic relations of  $M$ -estimators and  $R$ -estimators in the simple linear regression. Technical Report 41, Department of Statistics, Charles University in Prague. (submitted).
- Omelka, M. (2006). *Asymptotic properties of estimators and tests*. PhD thesis, Charles University in Prague (in preparation).
- Pollard, D. (1990). *Empirical Processes: theory and applications*. Institute of Mathematical Statistics.
- Puri, M.L., Wu, T.-J. (1985). Gaussian approximation of signed rank statistics process. *Journal of Statistical Planning and Inference*, 11, 277–312.
- Ren, J.J. (1994). On Hadamard differentiability and its application to  $R$ -estimation in linear models. *Statistics and Decision*, 12, 1–12.
- Serfling, R.J. (1980). *Approximation theorems of mathematical statistics*. New York: Wiley.
- Stigler, S. (1986). *The history of statistics. The measurement of Uncertainty before 1900*. Cambridge, (MA): The Belknap Press of Harvard University. Press.
- Terpstra, J.T., McKean, J.W. (2004). Rank-based analysis of linear models using  $R$ . Technical Report.
- van der Vaart, A.W., Wellner, J.A. (1996). *Weak convergence and empirical processes with applications to statistics*. Berlin Heidelberg New York: Springer.