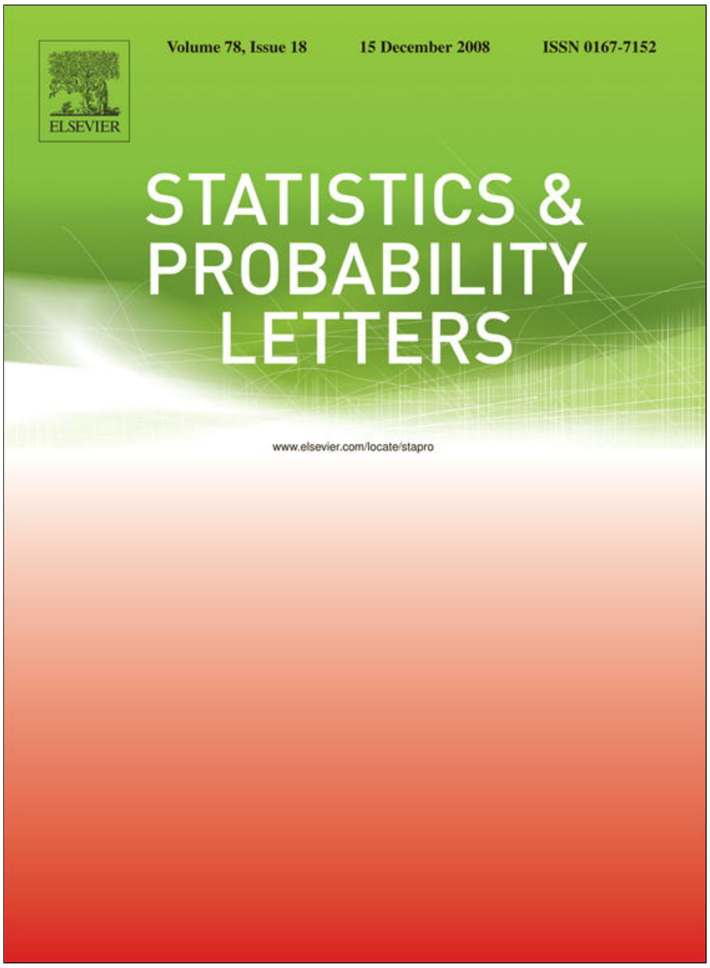
Volume 78, Issue 18
15 December 2008
ISSN 0167-7152
ELSEVIER
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Statistics and Probability Letters 78 (2008) 3366–3372

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Comparison of two types of confidence intervals based on Wilcoxon-type R-estimators

Marek Omelka

Jaroslav Hájek Centre for Theoretical and Applied Statistics, Charles University in Prague, Czech Republic

ARTICLE INFO

Article history:
Received 1 December 2007
Received in revised form 2 February 2008
Accepted 9 July 2008
Available online 15 July 2008

ABSTRACT

Omelka [Omelka, M., 2007. Second-order linearity of Wilcoxon rank statistics. Ann. Inst. Statist. Math. 59, 385–402] proposed an alternative way of constructing a confidence interval based on R-estimators for single parameters in linear models. We will compare this confidence interval with a traditional (Wald type) confidence interval theoretically as well as by means of a Monte Carlo experiment. As a by-product we will show the asymptotic normality of the estimator of $\int f^2$ proposed in [Koul, H.L., Sievers, G.L., McKean, J., 1987. An estimator of the scale parameter for the rank analysis of linear models under general score functions. Scand. J. Statist. 14, 131–141] under somewhat different assumptions from those in [Thewarepperuma, P.S., 1987. On estimation of some density functionals under regression and one sample models. Ph.D. Thesis. Michigan State University, Department of Statistics and Probability].

1. Introduction

Consider the linear regression model

$$Y_i = \alpha + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i = \alpha + \boldsymbol{\beta}^{\mathsf{T}}\mathbf{x}_i + e_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where α and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathsf{T}}$ are unknown parameters, $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathsf{T}}$, for $i = 1, \ldots, n$ are known constants, and $e_1, \ldots, e_n$ are independent, identically distributed random variables with a cumulative distribution function F with a density f. Let $R_i(\mathbf{b})$ be the rank of $Y_i - \mathbf{b}^{\mathsf{T}}\mathbf{x}_i$ among $Y_1 - \mathbf{b}^{\mathsf{T}}\mathbf{x}_1, \ldots, Y_n - \mathbf{b}^{\mathsf{T}}\mathbf{x}_n$ and $\bar{\mathbf{x}}_n = (\bar{x}_{n1}, \ldots, \bar{x}_{np})^{\mathsf{T}}$ be the vector of the means of the columns of the design matrix $\mathbf{X}_n$. In the following we will suppose that the columns of the matrix $\mathbf{X}_n$ are centered, that is $\bar{\mathbf{x}}_n = 0$. The R-estimator $\hat{\boldsymbol{\beta}}_n$ (based on the Wilcoxon scores) of $\boldsymbol{\beta}$ can be defined as the solution of the following minimization of 'Jaeckel' dispersion function (see Jaeckel (1972))

$$\hat{\boldsymbol{\beta}}_n = \arg\min_{\mathbf{b}\in\mathbb{R}_p} D_n(\mathbf{b}), \quad \text{where } D_n(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - \mathbf{b}^{\mathsf{T}}\mathbf{x}_i)\left(\frac{R_i(\mathbf{b})}{n+1} - \frac{1}{2}\right). \tag{1.2}$$

As this estimator is relatively easy to compute as well as to explain to non-statisticians, it belongs to the most popular R-estimators in particular in applications, see e.g. Hettmansperger and McKean (1977), Sievers and McKean (1986), Abebe et al. (2001), and McKean (2004).

A standard (Wald type) approach of constructing a confidence interval (CI) for β_l ($1 \le l \le p$) is based on the asymptotic normality of $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$. This asymptotic normality follows from the following first order asymptotic representation

E-mail address: omelka@karlin.mff.cuni.cz.

0167-7152/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2008.07.006

(e.g. Jurečková and Sen (1996))

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = \frac{\mathbf{V}_n^{-1}}{\gamma\sqrt{n}} \sum_{i=1}^{n} \mathbf{x}_i \left[F(e_i) - \frac{1}{2}\right] + o_p(1), \tag{1.3}$$

where $\mathbf{V}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i^{\mathsf{T}}$ and $\gamma = \mathsf{E}f(e_1) = \int f^2(x)\mathrm{d}x$. Thus the confidence interval for the parameter β_l is

$$D_n^{\mathrm{I}} = \left[\hat{\beta}_l - \frac{t_{n-p}(\alpha)}{\sqrt{n}} \frac{\sqrt{\omega_{ll}^n}}{\hat{\gamma}_n\sqrt{12}}, \hat{\beta}_l + \frac{t_{n-p}(\alpha)}{\sqrt{n}} \frac{\sqrt{\omega_{ll}^n}}{\hat{\gamma}_n\sqrt{12}}\right], \tag{1.4}$$

where ω_{ll}^n is the lth element of the diagonal of the matrix $\mathbf{V}_n^{-1}$, $\hat{\gamma}_n$ is an appropriate estimate of γ and $t_{n-p}(\alpha)$ is a $(1-\alpha/2)$ quantile of t-distribution with $n-p$ degrees of freedom. The advantage of using quantiles of t-distribution over quantiles of normal distribution was shown in Hettmansperger and McKean (1977). The mostly used estimate of γ was suggested in Koul et al. (1987).

An alternative confidence interval was originally proposed and studied in Jurečková (1973) for a simple linear model (one explanatory variable). Omelka (2007) generalized this type of confidence interval to a multiple linear model (more explanatory variables) in the following way. Let $r_1, \ldots, r_n$ be the residuals, that is $r_i = Y_i - \hat{\boldsymbol{\beta}}_n^{\mathsf{T}}\mathbf{x}_i$ for $i = 1, \ldots, n$. For $l = 1, \ldots, p$ define $S_{nl}(t) = \frac{1}{n^{3/2}}\sum_{i=1}^{n} x_{il} R_i(t)$, where $R_i(t)$ is the rank of the random variable $r_i - t\,x_{il}$ among $r_1 - t\,x_{1l}, \ldots, r_n - t\,x_{nl}$. Finally put

$$c_n = \frac{T_{nl}^2\sqrt{\omega_{ll}^n}}{\sqrt{12}}\sqrt{\frac{n}{n-p-1}}, \quad \text{where } T_{nl}^2 = \frac{1}{n}\sum_{i=1}^{n} x_{il}^2.$$

Then the (type II) confidence interval for the parameter β_l can be constructed as $D_n^{\mathrm{II}} = [\hat{b}_l^-, \hat{b}_l^+] = [\hat{\beta}_l + \delta_n^-, \hat{\beta}_l + \delta_n^+]$, where

$$\delta_n^- = \sup\left\{t < 0 : S_{nl}(t) \geq c_n\, t_{n-p}(\alpha)\right\}, \qquad \delta_n^+ = \inf\left\{t > 0 : S_{nl}(t) \leq -c_n\, t_{n-p}(\alpha)\right\}. \tag{1.5}$$

Notice that we do not need to estimate any unknown parameters.

By means of simulations Omelka (2007) showed that D_n^{II} is an interesting alternative to D_n^{I}. We believe that the comparison of these two types of confidence intervals deserves a further study. But to do that we need to decide on the estimator of $\hat{\gamma}_n$. The properties of various estimators of $\gamma = \int f^2(x)\,\mathrm{d}x$ have been studied in Aubuchon and Hettmansperger (1984), Sievers and McKean (1986), Aubuchon and Hettmansperger (1989), and Brown and Hettmansperger (2002). We chose the estimator which is mostly recommended in the textbooks as well as software manuals, and which was suggested in Koul et al. (1987). The precise definition of this estimator is given in Section 2. As a by-product of comparison of confidence intervals, we will find the Bahadur–Kiefer representation of this estimator, which further implies asymptotic normality. As pointed out by H. L. Koul in a personal communication, the asymptotic normality was already proved in Thewarepperuma (1987). As we were not aware of this result, we made an independent proof based on the recent results for the asymptotic expansion of Wilcoxon rank statistics (Omelka, 2007). Although our assumption (see W) is not generally comparable to the boundedness of the second derivative of density required in Thewarepperuma (1987), it is usually less restrictive, see Remarks 1 and 3.

As the only requirement on the estimator $\hat{\boldsymbol{\beta}}_n$ is its $\sqrt{n}$-consistency, that is $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = O_p(1)$, our result can be used to analyze the inference based on high-breakdown point modifications of Wilcoxon-type R-estimators as well, see e.g. Naranjo and Hettmansperger (1994) and Chang et al. (1999).

The rest of the paper is organized as follows. In Section 2 we formulate the main results concerning the asymptotic behavior of the confidence intervals D_n^{I} and D_n^{II}. Section 3 illustrates the finite sample properties by means of Monte Carlo experiments. Finally, the Appendix contains the proof.

2. Main results

To compare the confidence intervals we will work with the estimate of γ proposed in Koul et al. (1987). The estimate is constructed in the following way:

1. Denote

$$H_n(t) = \frac{1}{\binom{n}{2}} \sum_{i<j} \mathbb{I}\{|r_i - r_j| \leq t\}$$

the distribution function of the pairwise differences of the residuals $r_i = Y_i - \hat{\boldsymbol{\beta}}_n^{\mathsf{T}}\mathbf{x}_i$.

2. Put $\tau_n = \frac{H_n^{-1}(0.8)}{\sqrt{n}}$.

3. Finally estimate γ by

$$\hat{\gamma}_n = \frac{H_n(\tau_n)}{2\,\tau_n}\sqrt{\frac{n-p-1}{n}}.$$

Before we formulate the main theorem we need to make some assumptions. As it is convenient to have $\bar{\mathbf{x}}_n = 0$, it is necessary to formulate the conditions about the rows of the design matrix in the form of a triangular array $\{\mathbf{x}_{1n}, \ldots, \mathbf{x}_{nn}\}$.

X.1.

$$\frac{1}{n}\sum_{i=1}^{n}|\mathbf{x}_{in}|_2^4 = O(1), \qquad \lim_{n\to\infty}\frac{\max_{1\le i\le n}|\mathbf{x}_{in}|_2}{\sqrt{n}} = 0.$$

X.2. There exists a limit $(p \times p)$ matrix $\mathbf{V}$

$$\mathbf{V} = \lim_{n\to\infty}\mathbf{V}_n = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_{in}\,\mathbf{x}_{in}^{\mathsf{T}},$$

which is positive definite.

But for the simplicity of notation we will drop the index n in the following formulas and write shortly $\mathbf{x}_i$ instead of $\mathbf{x}_{in}$.

We need the distribution of the errors to satisfy:

W. F is absolutely continuous with a derivative f such that

$$\lim_{\Delta\to 0}\frac{1}{\Delta^2}\int_{-\infty}^{+\infty}\int_{-\Delta}^{+\Delta}[f(z+y)-f(y)]^2\mathrm{d}z\,\mathrm{d}y = 0.$$

Remark 1. According to Antille (1976) the condition W is satisfied in these two important situations:

(i) f is such that $|f(x+t)-f(x)| \le |t|^{\alpha}h(x)$, with $\alpha > \frac{1}{2}$ and $h(x) \in L_2(-\infty,+\infty)$;
(ii) f is absolutely continuous and $f'(x) \in L_2(-\infty,+\infty)$.

Notice that the second condition is satisfied if there exists a Fisher information for location parameter at the error density f (see Corollary 3.2.1 of Koul (2002)).

For the simplicity of notation put

$$a_F^n = \frac{t_{n-p}(\alpha)\sqrt{\omega_{ll}^n}}{\gamma\sqrt{12}} \quad\text{and}\quad a_F = \lim_{n\to\infty}a_F^n \overset{\mathbf{X.2}}{=} \frac{z_\alpha\sqrt{\omega_{ll}}}{\gamma\sqrt{12}},$$

where ω_{ll} is the lth diagonal element of the inverse of the matrix $\mathbf{V}$, z_α is $(1-\alpha/2)$ quantile of standardized normal distribution and $\gamma = \mathsf{E}f(e_1) = \int f^2(x)\mathrm{d}x$. Further let ℓ_n^{I} and ℓ_n^{II} be the lengths of the confidence intervals D_n^{I} and D_n^{II} and denote

$$L_n^{\mathrm{I}} = \frac{\sqrt{n}[\sqrt{n}\,\ell_n^{\mathrm{I}} - 2\,a_F^n]}{2\,a_F}, \qquad L_n^{\mathrm{II}} = \frac{\sqrt{n}[\sqrt{n}\,\ell_n^{\mathrm{II}} - 2\,a_F^n]}{2\,a_F}.$$

Theorem 1. *If the conditions* X.1–2, W *are satisfied, and the representation* (1.3) *holds, then the confidence intervals* D_n^{I}, D_n^{II} *satisfy:*

(1)

$$P(D_n^{\mathrm{I}} \ni \beta_l) \xrightarrow[n\to\infty]{} 1-\alpha, \qquad P(D_n^{\mathrm{II}} \ni \beta_l) \xrightarrow[n\to\infty]{} 1-\alpha.$$

(2)

$$\sqrt{n}\,\ell_n^{\mathrm{I}} = a_F + o_p(1) = \sqrt{n}\,\ell_n^{\mathrm{II}}.$$

(3) *The random variable* L_n^{I}, L_n^{II} *are asymptotically normal and admit the first order asymptotic representations*

$$L_n^{\mathrm{I}} = -\frac{2}{\gamma\sqrt{n}}\sum_{i=1}^{n}[f(e_i)-\gamma] + o_p(1), \tag{2.1}$$

$$L_n^{\mathrm{II}} = -\frac{1}{\gamma\sqrt{n}}\sum_{i=1}^{n}\left[\frac{x_{il}^2}{T_{nl}^2}+1\right][f(e_i)-\gamma] + o_p(1). \tag{2.2}$$

Remark 2. Notice, that for $p = 1$ the conclusions of Theorem 1 about D_n^{II} are in agreement with Jurečková (1973).

Theorem 1 implies that both confidence intervals are asymptotically correct (1) and their lengths are asymptotically equivalent (2).

Table 1
The ratio of the asymptotic variances of the lengths of confidence intervals

	$\mathcal{N}(0,1)$	$U(0,1)$	t_5	t_{10}	Logistic	Exponential
κ	3.00	1.80	9.00	4.00	4.20	9.00
$\frac{\sigma_2^2}{\sigma_1^2}$	1.50	1.20	3.00	1.75	1.80	3.00

But the main contribution of Theorem 1 is in the point (3). Let us denote by σ_{n1}^2 and σ_{n2}^2 the asymptotic variances of the properly standardized lengths of the confidence intervals D_n^{I} and D_n^{II}, i.e. asymptotic variances of L_n^{I} and L_n^{II}. Then by Theorem 1

$$\frac{\sigma_{n2}^2}{\sigma_{n1}^2} = \frac{\kappa+3}{4}, \quad \text{where } \kappa = \frac{1}{n}\sum_{i=1}^{n}\frac{x_{il}^4}{(T_{nl}^2)^2}. \tag{2.3}$$

As the Cauchy–Schwarz inequality implies

$$(T_{nl}^2)^2 \leq \frac{1}{n}\sum_{i=1}^{n} x_{il}^4,$$

we see that generally $\kappa \geq 1$ implying $\sigma_{n2}^2 \geq \sigma_{n1}^2$. Thus we can conclude that Wald-type confidence interval D_n^{I} is more stable in the sense, that its length has a smaller asymptotic variance. Further the variances σ_{n1}^2 and σ_{n2}^2 are equal if and only if $x_{il}^2 = const$ for all $i = 1, \ldots, n$, i.e. the values of the lth explanatory variable $\mathbf{x}_{\cdot 1}$ differ only in sign. As we suppose the explanatory variables to be centered, this corresponds to $\mathbf{x}_{\cdot l}$ being a dichotomous random variable with the same number of observations for both values.

Omelka (2006a) considered a similar comparison of confidence intervals based on M-estimators. He found out that the ratio of the asymptotic variances of (properly standardized) lengths of confidence intervals is directly κ. By (2.3) we see that for R-estimators the effect of the fourth moment of the explanatory variable $\mathbf{x}_{\cdot l}$ is weaker in comparison to an analogous situation for M-estimators. Table 1 summarizes the ratio of asymptotic variances (2.3) if the explanatory variable $\mathbf{x}_{\cdot l}$ is generated from some of the common distributions.

As by a simple algebra we can find out that

$$L_n^{\mathrm{I}} = \frac{1}{\hat{\gamma}_n}\sqrt{n}(\gamma - \hat{\gamma}_n),$$

Theorem 1 immediately implies

$$\sqrt{n}(\hat{\gamma}_n - \gamma) = \frac{2}{\sqrt{n}}\sum_{i=1}^{n}[f(e_i) - \gamma] + o_p(1). \tag{2.4}$$

Thus the estimator $\hat{\gamma}_n$ is asymptotically normal with the mean γ and the variance σ^2/n, where σ^2 is given by the formula

$$\sigma^2 = 4\left(\int f^3(x)\mathrm{d}x - \gamma^2\right).$$

By a closer inspection of the proof of Theorem 1 we can notice that to prove only the expansion (2.4), we are allowed to drop the condition on the fourth moment of $\mathbf{x}_i$ from the assumption X.1, which is needed to prove (2.2). It is also sufficient to assume only $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = O_p(1)$ instead of (1.3). Furthermore, (2.4) holds true even if we replace condition X.2 with a more general condition that the matrix $\mathbf{V}_n$ is regular for all sufficiently large n and it satisfies

$$\max_{1\leq i\leq n}\frac{\mathbf{x}_i^{\mathrm{T}}\mathbf{V}_n^{-1}\mathbf{x}_i}{n} \to 0, \quad \text{and} \quad \sqrt{n}\,\mathbf{V}_n^{1/2}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) = O_p(1).$$

The trick would be to replace $\mathbf{x}_i$ with $\mathbf{x}_i^* = \mathbf{V}_n^{-1/2}\mathbf{x}_i$ in the proof.

Remark 3. As we have mentioned in the Introduction, the asymptotic normality of estimator $\hat{\gamma}_n$ was already proved by Thewarepperuma (1987), who required the second derivative of the density of the errors to be bounded. Although this condition is generally not comparable with our condition W, our condition W is usually less restrictive for standard families of distributions (see Table 2). In fact, it is rather tricky to construct a density whose second derivative exists and is bounded and which does not meet the condition W.

3. Numerical results

Several numerical experiments have been conducted (for some of them see Omelka (2006b)). From these results we can conclude that:

Table 2
Comparisons of regularity conditions for some standard families of distributions

Distribution	Is f'' bounded?	Is W satisfied?
$\mathcal{N}(0, 1)$	Yes	Yes
Cauchy	Yes	Yes
Logistic	Yes	Yes
Lognormal	Yes	Yes
Laplace	No	Yes
Exponential	No	No
$U(0, 1)$	No	No
Convolution of two $U(0, 1)$	No	Yes
Gamma $f(x) = x^{k-1}\exp\{-\alpha x\}$	$k \geq 3$	$k > 1.5$
Beta $f(x) = x^{\alpha-1}(1-x)^{\beta-1}$	$\min(\alpha, \beta) \geq 3$	$\min(\alpha, \beta) > 1.5$

Table 3
Actual coverage probabilities of the true value of the parameter β_1 for the sample sizes $n = 34$, $n = 68$, and $n = 136$

n		$N(0, 1)$		Logistic		Cauchy		Lnorm	
		R I	R II	R I	R II	R I	R II	R I	R II
34	Coverage	0.953	0.948	0.953	0.947	0.963	0.940	0.961	0.946
	$\mathsf{E}(\sqrt{n}\,\ell_n)$	4.380	4.159	7.497	7.174	9.750	9.725	4.254	4.350
	$\mathrm{var}(L_n)$	20.605	15.969	69.721	57.705	384.939	402.157	40.033	51.569
68	Coverage	0.929	0.948	0.936	0.948	0.957	0.945	0.945	0.947
	$\mathsf{E}(\sqrt{n}\,\ell_n)$	3.775	4.061	6.451	6.938	7.830	8.190	3.491	3.766
	$\mathrm{var}(L_n)$	15.405	14.510	53.199	51.569	207.571	203.546	28.385	34.459
136	Coverage	0.935	0.949	0.940	0.949	0.955	0.946	0.946	0.949
	$\mathsf{E}(\sqrt{n}\,\ell_n)$	3.793	4.035	6.457	6.858	7.356	7.634	3.294	3.484
	$\mathrm{var}(L_n)$	14.152	13.724	49.497	49.208	151.988	156.014	23.281	27.202

- both D_n^{I} and D_n^{II} have good coverage properties;
- contrary to our expectations, D_n^{II} works very well for symmetric errors and outperforms D_n^{I} if the errors are normally distributed;
- when the errors are not symmetric then D_n^{I} is usually shorter and its length is less variable, but the one-sided confidence intervals may not be appropriate;
- it is usually not recommended to use D_n^{II} unless $\kappa \leq 3$.

These results suggest that to show the possible advantages of D_n^{II}, we need to consider designs with κ close to 1. As Omelka (2006b) considered only $p = 2$, we decided to increase the number of explanatory variables and to consider a linear model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + e_i.$$

For $n = 34$ we took the design matrix from the potency data example (see Table 6.6.2 of Hettmansperger and McKean (1998)) and for $n = 68$ and $n = 136$ we took appropriate multiples of this matrix. We were interested in the confidence intervals for the first variable (named SAE). The constant κ for this explanatory variable is 1.7 implying the ratio of variances (2.3) to be 1.175. The errors e_i were generated from a normal, logistic, Cauchy and lognormal distribution. For each of the situations the number of random samples was at least 100 000. The nominal coverage probability was set to 0.95.

The results can be found in Table 3. For simplicity of notation, the symbols D_n^{I} and D_n^{II} were replaced by R I and R II. For each sample size n the first row stands for the estimate of the true coverage, the second row for the estimate of the mean of $\sqrt{n}\,\ell_n$ and the third row the estimate of variance of L_n.

From Table 3 we can conclude that even when the ratio $\frac{n}{p}$ is small to moderate, D_n^{II} works better for symmetric errors unless they are heavy tailed (Cauchy). The main problem of D_n^{II} is possible undercoverage. On the contrary for asymmetric errors D_n^{I} is preferable. Although Table 3 does not include results for one-sided intervals, our numerical experience is that for such nice symmetric designs there are no problems with D_n^{I}-type one-sided intervals as well. We can notice that for moderate $\frac{n}{p}$, the ratio of asymptotic variances (2.3) is not very informative and finite sample properties are dominant. Some further results suggest that $\frac{n}{p}$ should be larger than 100 such that the asymptotic ratio (2.3) comes into play.

Acknowledgements

The author is grateful to H. L. Koul for pointing out to him the work of Thewarepperuma (1987), as well as to an anonymous referee whose comments led to a considerable improvement of the text. The author's research was supported by Research Project LC06024.

Appendix. Proof of Theorem 1

The proofs of the statements concerning confidence interval D_n^{II} can be found for a special case of the orthogonal columns of the design matrix $\mathbf{X}_n$ in Omelka (2007). In Omelka (2006b) it is indicated how to treat the general case. Thus we can focus on the confidence interval D_n^{I}.

To prove the Bahadur–Kiefer representation (1.3), we can argue similarly as in Jurečková and Sen (1996), pp. 272–273. The only difference is that to avoid the assumption of finite Fisher information for the density of the errors, we can use Corollary 4.4. of Omelka (2006b) to deduce the asymptotic linearity result (6.6.29). Once expansion (1.3) is proved, the statements (1) and (2) follow immediately from the consistency of $\hat{\gamma}_n$, which was proved in Koul et al. (1987). Thus it only remains to show (3).

As $L_n^{\mathrm{I}} = \frac{1}{\hat{\gamma}_n}\sqrt{n}(\gamma - \hat{\gamma}_n)$, it is sufficient to find the asymptotic expansion for the difference $\sqrt{n}(\hat{\gamma}_n - \gamma)$.

Let us define the process

$$M_n(\mathbf{t}, s) = \frac{2}{n}\sum_{i<j}^{n} \mathbb{I}\left\{\left|e_i - e_j - \frac{\mathbf{t}^{\mathsf{T}}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}\right| \le \frac{s}{\sqrt{n}}\right\}$$

where the index set is given by $T = \{(\mathbf{t}, s) : \max\{|\mathbf{t}|_2, s\} \le C \,\&\, s \ge 0\}$ and C is arbitrarily large but fixed. Notice that

$$\hat{\gamma}_n = \frac{\sqrt{n}\, M_n(\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}),\, H_n^{-1}(0.8))}{2\,(n-1)\, H_n^{-1}(0.8)}\sqrt{\frac{n-p-1}{n}} = \frac{M_n(\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}),\, H_n^{-1}(0.8))}{2\sqrt{n}\, H_n^{-1}(0.8)}\left(1 + O\left(\frac{1}{n}\right)\right). \tag{A.1}$$

and $M_n(\mathbf{0}, 0) = 0$ almost surely.

It will be more convenient to rewrite the process $M_n(\mathbf{t}, s)$ as

$$M_n(\mathbf{t}, s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{I}\left\{\left|e_i - e_j - \frac{\mathbf{t}^{\mathsf{T}}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}\right| \le \frac{s}{\sqrt{n}}\right\} - \frac{1}{n} = M'_{n1}(\mathbf{t}, s) - M'_{n1}(\mathbf{t}, -s) - \frac{1}{n}, \tag{A.2}$$

where

$$M'_{n1}(\mathbf{t}, s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{I}\left\{e_i - e_j \le \frac{s}{\sqrt{n}} + \frac{\mathbf{t}^{\mathsf{T}}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}\right\}.$$

The technique of how to deal with such processes was presented in Omelka (2007) and it can be found as well in Omelka (2006b), where a similar process

$$\tilde{S}_n(\mathbf{t}, s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} c_{in}\, \mathbb{I}\left\{e_i - e_j \le \frac{\mathbf{t}^{\mathsf{T}}(\mathbf{x}_i - \mathbf{x}_j)}{\sqrt{n}}\right\},$$

with $\{c_{in}\}$ being a triangular array of constants, was considered. The crucial step is to decompose the process

$$T_n(\mathbf{t}, s) = M'_{n1}(\mathbf{t}, s) - M'_{n1}(\mathbf{0}, 0) = P_n(\mathbf{t}, s) + R_n(\mathbf{t}, s)$$

by means of 'Hájek projection' (see Hájek (1968) or Serfling (1980)), where

$$P_n = \sum_{i=1}^{n} \mathsf{E}\,[T_n |\, e_i] - (n-1)\mathsf{E}\, T_n = \sum_{i=1}^{n} \mathsf{E}\,[T_n |\, e_i]\,.$$

Then by a minor modification of the proof of Corollary 3.3 of Omelka (2006b) we can find that the asymptotic distribution of the process T_n is given by the leading term P_n and the remainder term R_n is negligible in probability. As the assumption W implies W.1–3 of Omelka (2006b), we arrive at

$$\sup_{(\mathbf{t},s)\in T}\left|M'_{n1}(\mathbf{t}, s) - M'_{n1}(\mathbf{0}, 0) - \sqrt{n}\,\gamma\, s - \frac{2\, s}{\sqrt{n}}\sum_{i=1}^{n}(f(e_i) - \gamma)\right| = o_p(1), \tag{A.3}$$

which together with (A.2) gives us

$$\sup_{(\mathbf{t},s)\in T}\left|M_n(\mathbf{t}, s) - M_n(\mathbf{0}, 0) - 2\sqrt{n}\,\gamma\, s - \frac{4\, s}{\sqrt{n}}\sum_{i=1}^{n}(f(e_i) - \gamma)\right| = o_p(1). \tag{A.4}$$

By Lemma 6 of Koul et al. (1987) we know that $H_n^{-1}(0.8) = H(0.8) + o_p(1)$ as $n \to \infty$. Thus if we restrict the set T to $T' = \{(\mathbf{t}, s) : |\mathbf{t}|_2 \le C, |\frac{s - H^{-1}(0.8)}{H^{-1}(0.8)}| \le \frac{1}{2}\}$, then (A.4) implies

$$\sup_{(\mathbf{t},s)\in T'}\left|\frac{M_n(\mathbf{t}, s)}{s} - 2\sqrt{n}\,\gamma - \frac{4}{\sqrt{n}}\sum_{i=1}^{n}(f(e_i) - \gamma)\right| = o_p(1). \tag{A.5}$$

Now we can substitute $\sqrt{n}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta})$ for $\mathbf{t}$ and $H_n^{-1}(0.8)$ for s into (A.5) and with the help of (A.1) and some algebra get

$$\sqrt{n}(\hat{\gamma}_n - \gamma) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} (f(e_i) - \gamma) + o_p(1),$$

which completes the proof. □

References

Abebe, A., Crimin, K., McKean, J.W., Haas, J.V., Vidmar, J., 2001. Rank-based procedures for linear models: Applications to pharmaceutical science data. Drug Information Journal 35, 947–971.

Antille, A., 1976. Asymptotic linearity of Wilcoxon signed-rank statistics. Ann. Statist. 4, 175–186.

Aubuchon, J.C., Hettmansperger, T.P., 1984. A note on the estimation of the integral of $f^2(x)$. J. Statist. Plann. Inference 9, 321–331.

Aubuchon, J.C., Hettmansperger, T.P., 1989. Rank-based inference for linear models: Asymmetric errors. Statist. Probab. Lett. 8, 97–107.

Brown, B.M., Hettmansperger, T.P., 2002. On the estimation of integral f-squared. J. Statist. Plann. Inference 107, 365–375.

Chang, W.H., McKean, J.W., Naranjo, J.D., Sheather, S.J., 1999. High-Breakdown rank regression. J. Amer. Statist. Assoc. 94, 205–219.

Hájek, J., 1968. Asymptotic normality of simple linear rank statistics under alternatives. Ann. Math. Statist. 39, 325–346.

Hettmansperger, T.P., McKean, J.W., 1977. A robust alternative based on ranks to least squares in analyzing general linear models. Technometrics 19, 275–284.

Hettmansperger, T.P., McKean, J.W., 1998. Robust Nonparametric Statistical Methods. Arnold, London.

Jaeckel, L.A., 1972. Estimating regression coefficients by minimizing the dispersion of the residuals. Ann. Math. Statist. 43, 1449–1458.

Jurečková, J., 1973. Central limit theorem for Wilcoxon rank statistics process. Ann. Statist. 1, 1046–1060.

Jurečková, J., Sen, P.K., 1996. Robust Statistical Procedures: Asymptotics and Interrelations. Wiley, New York.

Koul, H.L., 2002. Weighted Empirical Processes in Dynamic Nonlinear Models, second edition. Springer, New York.

Koul, H.L., Sievers, G.L., McKean, J., 1987. An estimator of the scale parameter for the rank analysis of linear models under general score functions. Scand. J. Statist. 14, 131–141.

McKean, J.W., 2004. Robust analysis of linear models. Statist. Sci. 19, 562–570.

Naranjo, J.D., Hettmansperger, T.P., 1994. Bounded influence rank regression. J. Roy. Statist. Soc. Ser. B. 56, 209–220.

Omelka, M., 2006a An alternative method for constructing confidence intervals from M-estimators in linear models. In: Proceedings of Prague Stochastics 2006, pp. 568–578.

Omelka, M., 2006b. Second order properties of some M-estimators and R-estimators. Ph.D. Thesis. Charles University in Prague. Available at http://www.karlin.mff.cuni.cz~omelka/Soubory/omelka_thesis.pdf.

Omelka, M., 2007. Second-order linearity of Wilcoxon rank statistics. Ann. Inst. Statist. Math. 59, 385–402.

Serfling, R.J., 1980. Approximation Theorems of Mathematical Statistics. Wiley, New York.

Sievers, G.L., McKean, J.W., 1986. On the robust rank analysis of linear models with nonsymmetric error distributions. J. Statist. Plann. Inference 13, 215–230.

Thewarepperuma, P.S., 1987. On estimation of some density functionals under regression and one sample models. Ph.D. Thesis. Michigan State University, Department of Statistics and Probability.