IMS Collections
Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková
Vol. 7 (2010) 204–214

DOI: 10.1214/10-IMSCOLL720

Extremes of two-step regression quantiles*

Jan Picek[1] **and Jan Dienstbier**[1,2]

Technical University of Liberec and Charles University in Prague

Abstract:
The article deals with estimators of extreme value index based on two-step regression quantiles in the linear regression model. Two-step regression quantiles can be seen as a possible generalization of the quantile idea and as an alternative to regression quantiles. We derive the approximation of the tail quantile function of errors. Following Drees (1998) we consider a class of smooth functionals of the tail quantile function as a tool for the construction of estimators in the linear regression context. Pickands, maximum likelihood and probability weighted moments estimators are illustrated on simulated data.

1. Introduction

Let $E_1, \ldots, E_n$, $n \in \mathbb{N}$ be independent and identically distributed random variables with a common distribution function function F belonging to some max-domain of attraction of an extreme-value distribution G_γ for some parameter $\gamma \in \mathbb{R}$, i.e there exists a function $a(t)$ with a constant sign such that for any $x > 0$ and some $\gamma \in \mathbb{R}$

$$\lim_{t \to 0} \frac{F^{-1}(1-tx) - F^{-1}(1-t)}{a(t)} = \frac{x^{-\gamma}-1}{\gamma}. \tag{1.1}$$

The relation (1.1) is equivalent to the Fisher–Tippet result: If for some distribution function $G_\gamma(x)$ and sequences of real numbers $a(n) > 0$ and $b(n)$, $n \in \mathbb{N}$, $\lim_{n\to\infty} F^n(a(n)x + b(n)) = G_\gamma(x)$ for every continuity point x of G, then G_γ is the *extreme value distribution*, i. e. $G_\gamma(x) = \exp(-(1+\gamma x)^{-1/\gamma})$, $\gamma \neq 0$, and the case $\gamma = 0$ is interpreted as the limit $\gamma \to 0$.

The problem of estimating the so-called *extreme value index* γ, which determines the behavior of the distribution function F in its upper tail, has received much attention in the literature, see e. g. [3] and references cited there. More attention has been paid to estimators that are based on a certain number of upper order statistics. They are usually scale invariant but not invariant under a shift of the data, see [1] for some examples.

However, one of the challenging ideas of the recent advances in the field of statistical modeling of extreme events has been the development of models with time-dependent parameters or more generally models incorporating covariates.

*The research was supported by the Ministry of Education of the Czech Republic under research project LC06024.

[1]Department of Applied Mathematics, Technical University of Liberec, Studentská 2, CZ-461 17 Liberec, Czech Republic, e-mail: jan.picek@tul.cz; jan.dienstbier@tul.cz

[2]Department of Statistics, Charles University in Prague, Sokolovska 83, CZ-186 75 Prague 8, Czech Republic,

AMS 2000 subject classifications: Primary 62G30, 62G32; secondary 62J05.

Keywords and phrases: two-step regression quantile, *R*-estimator, extreme value index, tail function.

Therefore, in the present paper we aim at extending the general result given in Drees (1998) to linear regression. Consider the following linear model

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta} + \mathbf{E}, \tag{1.2}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ is a vector of observations, $\mathbf{X}$ is an $(n \times p)$ known design matrix with rows $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$, $i = 1, \ldots n$, $\mathbf{1}_n = (1, \ldots, 1)^\top \in \mathbb{R}^n$, $\mathbf{E} = (E_1, \ldots, E_n)^\top$ is a vector of i. i. d. errors with an unknown distribution function F, β_0 and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ are the unknown parameters.

The outline of this paper is as follows. Section 2 describes the construction of the two-step regression quantiles. In Section 3 the estimation of the extremes of the two-step extreme regression quantiles is given. Following Drees (1998) we establish the approximation for the tail quantile function of residual and we show the consistency and the asymptotic distribution of functionals of the tail quantile function in Section 4. The simulation study is contained in Section 5.

2. Two-step regression quantiles

Jurečková and Picek [9] proposed an alternative of the α-regression quantiles suggested by Koenker and Basset [12] in the model (1.2) as follows: Let $\hat{\boldsymbol{\beta}}_{nR}(\alpha)$ be an appropriate R-estimate of the slope parameter $\boldsymbol{\beta}$ and let $\tilde{\beta}_{n0}$ denote $[n\alpha]$-order statistic of the residuals $Y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{nR}(\alpha)$, then the vector $\tilde{\boldsymbol{\beta}}_n(\alpha) := \left(\tilde{\beta}_{n0}, \widehat{\boldsymbol{\beta}}_{nR}(\alpha)\right)^\top$ is called the two-step α-regression quantile.

The initial R-estimator of the slope parameters is constructed as an inverse of the rank test statistic calculated in the Hodges-Lehmann manner, see [11]: Denote $R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})$ the rank of $Y_i - \mathbf{x}_i^\top \mathbf{b}$ among $(Y_1 - \mathbf{x}_1^\top \mathbf{b}, \ldots, Y_n - \mathbf{x}_n^\top \mathbf{b}), \mathbf{b} \in \mathbb{R}^p, i = 1, \ldots, n$. Note that $R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})$ is also the rank of $Y_i - b_0 - \mathbf{x}_i^\top \mathbf{b}$ among $(Y_1 - b_0(\alpha) - \mathbf{x}_1^\top \mathbf{b}, \ldots, Y_n - b_0(\alpha) - \mathbf{x}_n^\top \mathbf{b})$ for any $\alpha \in (0, 1)$ because the ranks are translation invariant. Consider the vector $\mathbf{S}_n(\mathbf{b}) = (S_{n1}(\mathbf{b}), \ldots, S_{np}(\mathbf{b}))^\top$ of the linear rank statistics, where

$$S_{nj}(\mathbf{b}) = \sum_{i=1}^{n} x_{ij} \varphi_\alpha \left(\frac{R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})}{n+1} \right), \quad \mathbf{b} \in \mathbb{R}^p, \quad j = 1, \ldots, p. \tag{2.1}$$

and $\varphi_\alpha = \alpha - I[x < 0], x \in \mathbb{R}$. Then the estimator $\hat{\boldsymbol{\beta}}_{nR}$ is defined as

$$\widehat{\boldsymbol{\beta}}_{nR} = \operatorname{argmin}_{\mathbf{b} \in \mathbb{R}^p} \|\mathbf{S}_n(\mathbf{b})\|_1, \tag{2.2}$$

where $\|\mathbf{S}\|_1 = \sum_{j=1}^{p} |S_j|$ is the L_1 norm of $\mathbf{S}$, see [6]; or

$$\widehat{\boldsymbol{\beta}}_{nR} = \operatorname{argmin}_{\mathbf{b} \in \mathbb{R}^p} \mathcal{D}_n(\mathbf{b}), \tag{2.3}$$

where

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^{n} (Y_i - \mathbf{x}_i' \mathbf{b}) \varphi_\alpha \left(\frac{R_{ni}(\mathbf{Y} - \mathbf{X}\mathbf{b})}{n+1} \right) \tag{2.4}$$

is the Jaeckel's measure of rank dispersion, see [5].

$\widehat{\boldsymbol{\beta}}_{nR}$ estimates only the slope parameters and the computation is invariant of the size of the intercept.

Assume the following conditions on distribution function F of errors and on $\mathbf{X}$ in model (1.2):

(A1) F has a continuous density f that is positive on the support of F and has finite Fisher's information, i. e. $0 < \int \left(\frac{f'(x)}{f(x)}\right)^2 \mathrm{d}F(x) < \infty$.

(A2) $\lim_{n\to\infty} \max_{1\le i\le n} \mathbf{x}_i^\top \left(\sum_{k=1}^n \mathbf{x}_k\mathbf{x}_k^\top\right)^{-1} \mathbf{x}_i = 0$.

(A3) $\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathbf{x}_i^*\mathbf{x}_i^{*\top} = \mathbf{D}^*$, where $\mathbf{x}_i^* = (1, x_{i1}, \dots, \mathbf{x}_{ip})^\top$, $i = 1,\dots,n$, and $\mathbf{D}^*$ is a positively definite $(p+1)\times(p+1)$ matrix.

Under conditions (**A1**) – (**A3**), the R-estimator (2.2) and (2.3) admits the following asymptotic representation,

$$\begin{aligned} &n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{nR} - \boldsymbol{\beta}) \\ (2.5)\qquad &= n^{-\frac{1}{2}}(f(F^{-1}(\alpha))^{-1}\mathbf{D}^{-1}\sum_{i=1}^n \mathbf{x}_i\left(\alpha - I[E_i < F^{-1}(\alpha)]\right) + o_p(n^{-1/4}),\end{aligned}$$

where $\mathbf{D} = \lim_{n\to\infty}\mathbf{D}_n$, $\mathbf{D}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i^\top$, for details see [10].

The solutions of (2.2) and (2.3) are generally not unique, nevertheless the asymptotic representation (2.5) applies to any of such solution; e. g. we can take the center of gravity of the set of all solutions.

Jurečková and Picek showed in [9] that the two-step regression quantiles are asymptotically equivalent to the regression quantiles suggested by Koenker and Basset in [12]. The α-regression quantile is obtained as a solution of the minimization

$$(2.6)\qquad \hat{\boldsymbol{\beta}}_n(\alpha) := \operatorname{argmin}_{(b_0,\mathbf{b})}\left\{\sum_{i=1}^n \rho_\alpha(Y_i - b_0 - \mathbf{x}_i^\top\mathbf{b}), b_0(\alpha)\in\mathbb{R}, \mathbf{b}\in\mathbb{R}^p\right\}$$

with the loss function given by $\rho_\alpha(x) = |x|(\alpha I[x>0] + (1-\alpha)I[x<0]), x \in \mathbb{R}$. The population counterpart of the vector $\hat{\boldsymbol{\beta}}_n(\alpha)$ is the vector $\boldsymbol{\beta}(\alpha) = (\beta_0 + F^{-1}(\alpha), \beta_1, \dots, \beta_p)^\top$. The difference between empirical regression quantile and its theoretical population counterpart is $\mathcal{O}_P(n^{-3/4})$ under general conditions on $\mathbf{X}$ and F, see e. g. Theorem 7.4.1. in [10].

3. Extremes of two-step quantiles

The authors of [9] also considered the extreme two-step quantile $\hat{E}_{n:n}$, which they define as the maximum of the residuals

$$(3.1)\qquad \hat{E}_{n:n} = \max\{Y_1 - \mathbf{x}_1^\top\widehat{\boldsymbol{\beta}}_{nR}, \dots, Y_n - \mathbf{x}_n^\top\widehat{\boldsymbol{\beta}}_{nR}\}$$

calculated with respect to an appropriate R-estimate $\widehat{\boldsymbol{\beta}}_{nR}$ of $\boldsymbol{\beta}$. Under suitable conditions (see [9]) $\hat{E}_{n:n}$ is a consistent estimate of $E_{n:n} + \beta_0$ and

$$(3.2)\qquad |\hat{E}_{n:n} - E_{n:n} - \beta_0| = \mathcal{O}_p(n^{-\delta}) \quad \text{as } n\to\infty,\ 0<\delta<\tfrac{1}{2}$$

Let $\widehat{\boldsymbol{\beta}}_{nR}^{+}$ be the initial R-estimate generated by the score function $\varphi_{1-\frac{1}{n}}(u) = I[u \ge 1-\frac{1}{n}] - \frac{1}{n}$, $0<u<1$. In this case the Jaeckel measure of the rank dispersion (2.4) takes the form

$$(3.3)\qquad \max_{1\le i\le n}\{Y_i - \mathbf{x}_i^\top\mathbf{b}\} - \bar{Y}_n,$$

where $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^n Y_i$. Hence,

$$\widehat{\boldsymbol{\beta}}_{nR}^{+} = \operatorname{argmin}_{\mathbf{b}\in\mathbb{R}^p} \sum_{i=1}^{n} \left(Y_i - \mathbf{x}_i^\top \mathbf{b}\right)^{+}. \tag{3.4}$$

Then we can define the *maximal two-step regression quantile* as $\left(\hat{E}_{n:n},\, \widehat{\boldsymbol{\beta}}_{nR}^{+}\right)$. For this estimate it holds that $\hat{E}_{n:n} + \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_{nR}^{+} \geq Y_i, i = 1, \ldots, n$, while for some i_0 the inequality reduces to equality.

Jurečková [8] showed in the case $\alpha \to 0$ or $\alpha \to 1$ that the two-step regression quantile coincides exactly with the extreme regression quantile considered by Jurečková and Portnoy in [15]. Jurečková and Portnoy also derived some properties of the extreme regression quantiles. The extremes of regression quantiles have been further studied by Chernozhukov in [2]. He established the consistency of intermediate regression quantiles and simple estimators such as Pickands. Since the two-step α regression quantiles $\tilde{\boldsymbol{\beta}}_n(\alpha)$ are close to α-regression quantiles $\hat{\boldsymbol{\beta}}(\alpha)$ it should be interesting to examine the properties of $\tilde{\boldsymbol{\beta}}_n(\alpha)$ in the extreme context. This problem is closely related to the extremal properties the of high-order residuals related to the initial estimate $\widehat{\boldsymbol{\beta}}_{nR}(\alpha)$.

The methods of extreme value theory are often based not only on the maximum order statistics but also on the other higher empirical quantiles. In fact, the estimates of the extreme value index γ are calculated not only from extreme order statistics but also from the statistics of *intermediate order*, $k \to \infty$, $k/n \to 0$ as $n \to \infty$.

If we consider the regression model (1.2), then the order statistics of errors are not directly observable but the inference can be based on the estimates of errors. We shall use the residuals of a suitable R-estimate discussed above. Denote $\left\{\hat{E}_1, \ldots, \hat{E}_n\right\}$ the set of residuals $\left\{Y_1 - \mathbf{x}_1^\top(\widehat{\boldsymbol{\beta}}_{nR} - \boldsymbol{\beta}), \ldots, Y_n - \mathbf{x}_n^\top(\widehat{\boldsymbol{\beta}}_{nR} - \boldsymbol{\beta})\right\}$. The following lemma shows that k-th ordered residual $\hat{E}_{k:n}$ is an appropriate estimate of $E_{k:n}$.

Lemma 3.1. *Let $\widehat{\boldsymbol{\beta}}_{nR}$ be an R-estimate of $\boldsymbol{\beta}$, generated by a fixed nondecreasing and integrable score function $\varphi : (0,1) \mapsto \mathbb{R}$, independent of n, as in (2.1) and (2.2). Assume the conditions* **(A1)** – **(A3)** *and*

$$\max_{1\leq i\leq n} \|\mathbf{x}_i\| = \mathcal{O}\left(n^{\frac{1}{2}-\delta}\right) \quad \textit{as} \quad n \to \infty,\ 0 < \delta < \tfrac{1}{2}, \tag{3.5}$$

then

$$\sup_{1\leq k\leq n} \left|\hat{E}_{k:n} - E_{k:n} - \beta_0\right| = \mathcal{O}_P(n^{-\delta}), \qquad \textit{as } n \to \infty. \tag{3.6}$$

Proof. Let $D_1, \ldots, D_n$ denote the antiranks of $E_1, \ldots, E_n$, i. e. the indices satisfying $E_{i:n} = E_{D_i}$, $i = 1, \ldots, n$. Moreover for an R-estimate $\widehat{\boldsymbol{\beta}}_{nR}$ of the slope $\boldsymbol{\beta}$ and $n \in \mathbb{N}$

$$u_n := u_n(\widehat{\boldsymbol{\beta}}_{nR}) := \max_{i=1,\ldots,n} |\mathbf{x}_i^\top(\widehat{\boldsymbol{\beta}}_{nR} - \boldsymbol{\beta})|.$$

From the asymptotic representation of $\widehat{\boldsymbol{\beta}}_{nR}$ (2.5) and (3.5) we get $u_n = O_P(n^{-\delta})$ as $n \to \infty$.

Notice that $\hat{E}_{1:n} \leq E_{1:n}+\beta_0+u_n$, because the opposite case $\hat{E}_{1:n} > E_{1:n}+\beta_0+u_n$ implies

$$\hat{E}_{D_1} = E_{1:n} + \beta_0 + \mathbf{x}_{D_1}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{nR}) \leq E_{1:n} + \beta_0 + u_n < \hat{E}_{1:n}.$$

Hence, $\hat{E}_{1:n}$ is the smallest observation among $\left\{\hat{E}_i, i = 1, \ldots, n\right\}$, therefore it cannot be greater than $\hat{E}_{D_1}$.

Similarly, $\hat{E}_{2:n} \leq E_{2:n} + \beta_0 + u_n$ because $\hat{E}_{2:n} > E_{2:n} + \beta_0 + u_n$ leads to

$$\hat{E}_{D_2} = E_{2:n} + \beta_0 + \mathbf{x}_{D_2}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{nR}) \leq E_{2:n} + \beta_0 + u_n < \hat{E}_{2:n}$$

and

$$\hat{E}_{D_1} = E_{1:n} + \beta_0 + \mathbf{x}_{D_1}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{nR}) \leq E_{2:n} + \beta_0 + u_n < \hat{E}_{2:n}.$$

If we proceed analogously, we get

$$\hat{E}_{i,n} \leq E_{i,n} + \beta_0 + u_n, \qquad i = 1, \ldots, n. \tag{3.7}$$

On the other hand, it holds for the highest two-step ordered residual $\hat{E}_{n:n} \geq E_{n:n} + \beta_0 - u_n$, because $\hat{E}_{n:n} < E_{n:n} + \beta_0 - u_n$ implies

$$\hat{E}_{D_n} = E_{n:n} + \beta_0 + \mathbf{x}_{D_n}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_{nR}) \geq E_{n:n} + \beta_0 - u_n > \hat{E}_{n:n}.$$

We get by the similar arguments as in (3.7)

$$\hat{E}_{i,n} \geq E_{i,n} + \beta_0 - u_n, \qquad i = 1, \ldots, n. \tag{3.8}$$

Finally, $u_n = O_p(n^{-\delta})$ together with (3.8) and (3.7) imply (3.6). □

4. Estimators of extreme value index

Suppose for a while we have a simple location model, i. e. $\boldsymbol{\beta} = 0$ in (1.2). Many estimators of γ that are based on upper order statistics considered can be represented (at least approximately) as a smooth functionals $T(Q_n)$ of the *empirical tail quantile function*

$$Q_n(t) := F_n^{-1}\left(1 - \frac{k_n}{n}t\right) = X_{n-[k_n t]:n}, \quad t \in [0, 1],$$

with F_n^{-1} denoting the *empirical distribution function* and $X_{i:n}$ the ith order statistic of the i.i.d. sample. Note that Q_n depends on the $(k_n + 1)$ largest order statistics $(1 \leq k_n < n)$. Drees in [4] studied the asymptotic behaviour of such estimators.

Consider the general regression model (1.2) and the largest order statistics of the residuals. Let any $k \in \mathbb{N}$ be such that $\hat{E}_{k:n} > 0$. Then define the *tail quantile function of the residuals* as follows

$$\hat{Q}_{n,k}(t) := \hat{E}_{n-[kt]:n}.$$

Observe that $\hat{Q}_{n,k}$ is the consistent estimate of the empirical tail function of the errors $Q_{n,k}(t) = E_{n-[kt]:n}$ in the sense of Lemma 3.1. We shall provide an approximation of $\hat{Q}_{n,k}$ for the intermediate sequences of $k(n)$.

Suppose that the distribution function F in (1.2) satisfies (1.1). To obtain the approximation of $\hat{Q}_{n,k}$, however, it is useful to impose stronger condition concerning the second order approximations of the tails

$$\lim_{t\to 0} \frac{\frac{F^{-1}(1-tx)-F^{-1}(1-t)}{a(t)} - \frac{x^{-\gamma}-1}{\gamma}}{A(t)} = K(x), \tag{4.1}$$

where a is the function related to (1.1), $A(t)$ is a function of constant sign and K is some function that is not a multiple of the $(x^{-\gamma}-1)/\gamma$.
It can be shown that there is some $\rho \neq 0$ such that $K(x) = z_{\gamma-\rho} = (x^{\rho-\gamma}-1)/(\gamma-\rho)$, which for the cases $\rho = 0$ and $\gamma = 0$ is understood to be equal to the limit of $z_{\gamma-\rho}$, as $\gamma \to 0$ or $\rho \to 0$, respectively, see [3] for details.

The so-called second-order condition (4.1) naturally arises when discussing the bias of the estimators of γ, see [3] or [1]. Under second order condition (4.1) one can establish following uniform approximation of the tail quantile function.

Theorem 4.1. *Suppose that the distribution function F of errors in (1.2) satisfies (4.1) for some $\gamma \in \mathbb{R}$ and $\rho \leq 0$. Suppose that the assumptions of Lemma 3.1 are fulfilled. Then we can define a sequence of Wiener processes $\{W_n(t)\}_{t\geq 0}$ such that for suitable chosen functions A and a and each $\varepsilon > 0$,*

$$\begin{aligned}\sup_{t\in(0,1]} t^{\gamma+\frac{1}{2}+\varepsilon} \Bigg| \frac{\hat{Q}_{n,k}(t) - F^{-1}\left(1-\frac{k}{n}\right) - \beta_0}{a(k/n)} - \Big(z_\gamma(t) - k^{-\frac{1}{2}} t^{-(\gamma+1)} W_n(t) \\ + A\left(\frac{k}{n}\right) K(t)\Big)\Bigg| = o_P\left(k^{-1/2} + |A(k/n)|\right),\end{aligned} \tag{4.2}$$

$n \to \infty$, provided $k = k(n) \to \infty$, $k/n \to 0$ and $\sqrt{k}A(k/n) = O(1)$

Proof. Immediately follows from (3.6) and the approximation of the tail quantile function derived in Theorem 2.1 of [4]. □

Following [4] we consider the class of smooth statistical functionals of the estimated empirical tail quantile function $T(\hat{Q}_{n,k})$ for fixed parameter values γ. We are going to describe the properties of the functionals on space of functions that are close to the tail quantile function (or its estimate $\hat{Q}_{n,k}$). Since $F^{-1}(1-t)$ diverges as $t \to 0$ for $\gamma > 0$, we introduce weighted space $\mathcal{H}$ of real functions on the interval $[0,1]$ which are smooth and similar to the tail quantile function

$$\mathcal{H} := \left\{ h : [0,1] \to [0,\infty] \,\middle|\, h \in C[0,1], \lim_{t\downarrow 0} \frac{(\log\log(3/t))^{1/2} h(t)}{t^{1/2}}, t \in [0,1] \right\}. \tag{4.3}$$

For each $\gamma \in \mathbb{R}$ and $h \in \mathcal{H}$ we define seminorm on the space of real functions on the unit interval by $\|z\|_{\gamma,h} := t^\gamma h(t)|z(t)|$. In the view of Theorem 4.1

$$\mathcal{D}_{\gamma,h} := \left\{ z : [0,1] \to \mathbb{R} \,\middle|\, \lim_{t\downarrow 0} t^\gamma h(t) z(t) = 0, (t^\gamma h(t) z(t))_{t\in[0,1]} \in D[0,1] \right\} \tag{4.4}$$

equipped with the weighted supremum seminorm $\|z\|_{\gamma,h}$ is the suitable space in which weak convergence of $\hat{Q}_{n,k}$ can be established. Furthermore, let $C_{\gamma,h} := \left\{ z \in \mathcal{D}_{\gamma,h} | z_{|(0,1]} \in C(0,1] \right\}$ be a subset of continuous functions on $(0,1]$ of $\mathcal{D}_{\gamma,h}$. We shall formulate the key theorem showing the consistence and asymptotical normality of a broad class of functionals of $\hat{Q}_{n,k}$.

Theorem 4.2. *Suppose that for $\gamma \in \mathbb{R}$ and some $h \in \mathcal{H}$ the functional T : span$(D_{\gamma,h}, 1) \to \mathbb{R}$ satisfies*

(i) $T_{|\mathcal{D}_{\gamma,h}}$ *is* $\mathcal{B}(\mathcal{D}_{\gamma,h}, \mathcal{B}(\mathbb{R})$*-measurable (where* $\mathcal{B}$ *denotes the Borel-*σ*-field),*
(ii) $T(az+b) = T(z), \qquad$ *for all* $z \in \mathcal{D}_{\gamma,h}$, $a > 0$, $b \in \mathbb{R}$,
(iii) $T(z_\gamma) = T(1/\gamma(x^{-\gamma} - 1)) = \gamma$
(iv) $T_{|\mathcal{D}_{\gamma,h}}$ *is Hadamard differentiable tangentially to* $C_{\gamma,h} \subset D_{\gamma,h}$*, at* z_γ *with a derivative* T'_γ*, i. e. for some signed measure* $\nu_{T,\gamma}$ *it holds for all* $0 < \varepsilon_n \to 0$ *and all* $y_n \in \mathcal{D}_{\gamma,h}$ *such that* $y_n \to y \in C_{\gamma,h}$

$$\lim_{\varepsilon_n \to 0} \frac{T(z_\gamma - \varepsilon y_n) - T(z_\gamma)}{\varepsilon_n} = T'_\gamma(y) = \int_0^1 y \, \mathrm{d}\nu_{T,\gamma}. \tag{4.5}$$

Then under the assumptions of Theorem 4.1 provided that $\sqrt{k}A(k/n) \to \lambda$

(i) $T(\hat{Q}_{n,k}) \to \gamma$
(ii) $\mathcal{L}(k_n^{1/2}(T(\hat{Q}_{n,k}) - \gamma)) \to \mathcal{N}(\mu_{T,\gamma,\rho}, \sigma_{T,\gamma})$*, where*

$$\begin{aligned} \mu_{T,\gamma,\rho} &:= \int_0^1 z_{\gamma-\rho} \, \mathrm{d}\nu_{T,\gamma} \\ \sigma_{T,\gamma} &:= \mathsf{Var}\left(\int_0^1 t^{\gamma-1} W(t) \, \mathrm{d}\nu_{T,\gamma}(t)\right) \\ &= \int_0^1 \int_0^1 (st)^{\gamma-1} \min(s,t) \, \mathrm{d}\nu_{T,\gamma}(s) \, \mathrm{d}\nu_{T,\gamma}(t) \end{aligned}$$

Proof. Follows from Theorem 4.1 similarly as the proof of Theorem 3.2 in [4]. □

Theorem 4.2 assures that any location and scale invariant estimator of γ is consistent even if it is calculated from estimated residuals instead of the unobservable errors in (1.2). Moreover, as have been shown in [4] practically all location and scale invariant estimators of γ belongs to the class satisfying the assumptions of Theorem 4.2.

Example 4.1. *(i) Pickands estimator of* γ *is generated by the functional*

$$T_{Pick}(z) = \frac{1}{\log 2} \log\left(\frac{z(1/4) - z(1/2)}{z(1/2) - z(1)}\right) I[(z(1/4) - z(1/2))(z(1/2) - z) > 0]. \tag{4.6}$$

(ii) Generalized probability weighted moment can be regarded as

$$T_{PWM}(z) = \frac{\int z \, \mathrm{d}v_1}{\int z \, dv_2} I\left[\int z \, \mathrm{d}v_2 \neq 0\right] \tag{4.7}$$

for suitable finite signed Borel measures v_i *on [0,1], see [4].*

Since the larger observation approximately follow the Generalized Pareto (GP) distribution, if we apply the maximum likelihood procedure to the observations exceeding a given high threshold using GP distribution, we obtain an estimator of extreme value index (i. e. the shape parameter GP distribution). The maximum likelihood estimator is location and scale invariant, details see [3].

Note that we could also give the similar results of Theorem 4.1 and Lemma 3.1 if we would replace $\widehat{\boldsymbol{\beta}}_{nR}$ by any other suitable estimator of the slope parameter $\hat{\boldsymbol{\beta}}_n$ fulfilling

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = O_P(n^{-1/2}).$$

Nevertheless, we focus on $\widehat{\boldsymbol{\beta}}_{nR}$ because it estimates only the slope parameters in (1.2) and the computation is invariant of the size of the intercept.

But primarily we would like to stress that the nature of the two-step regression quantiles and their relation to the regression quantiles of Koenker and Basset, which makes their properties an interesting subject to study. The studied two-step regression α-quantile is asymptotically equivalent and numerically very close to the regression α-quantile and the maximal two-step regression quantile coincides with the maximal regression quantile as it was already mentioned. That is important if we have proved some results for the two-step regression quantiles only. While there were described asymptotic properties of the maximal regression quantile, see [15], [8] and others, only [2] studied the properties of the extreme and intermediate regression quantiles for different sequences of α_n but only in the pointwise sense. Theorem 4.1 gives immediately the uniform approximation of the tails of the two-step quantiles, which enables to base the tail modelling fully on the quantile function of the residuals $\hat{Q}_{n,k}(t)$.

The intrinsic connections between the regression quantiles and the two-step regression are important in the case that the assumptions are violated. There exist various interesting results showing the stability of regression quantiles even under dependency and heterogenity of the conditional distribution of the errors, for some overview see [13]. In this context, the extreme two-step regression quantiles can be observed as an interesting pattern for working with extreme regression quantiles.

On the other hand, the previously described method are directly applicable for some real case studies where the independence of the errors is assumed. We can refer e. g. the Condroz dataset presented in [1] considering calcium level and pH level of the soil in different regions. We could find other examples e. g. in the climatology, where the most widely-used method for dealing with the problem of dependency is declustering. That approach is presented in [14], where authors proposed a methodology for estimating high quantiles of distributions of daily temperature, based on the peaks-over-threshold analysis with a time-dependent threshold expressed in terms of regression quantiles.

5. Numerical Illustration

In order to check how the estimators of extreme value index perform in the linear regression model we have conducted a simulation study. We considered the model

$$Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + E_i, \quad i = 1, \ldots, n,$$

where the errors E_i, $i = 1, \ldots, n$, were simulated from the Burr, Generalized Pareto and Pareto distributions with the following parameter values: sample size $n = 400$, $\beta_0 = 2$, $\boldsymbol{\beta} = (\beta_1, \beta_2) = (-1, 2)$, $\alpha = 0.5$. Concerning the regression matrix we generated two columns $(x_{11}, \ldots, x_{n1})$ and $(x_{12}, \ldots, x_{n2})$ as two independent samples from the uniform distributions $R(0, 10)$ and $R(-5, 15)$, respectively. The R-estimator $\widehat{\boldsymbol{\beta}}_R$ was computed by minimizing Jaeckel's objective function (2.4).

10 000 replications of the model were simulated for each combination of the parameters and then the residuals based on the R-estimator $\widehat{\boldsymbol{\beta}}_R(0.5)$ were calculated. For the sake of comparison, the values of Pickands, maximum likelihood, and probability weighted moments estimator were computed for k - the varying fraction of ordered residuals.

In Figures 1 – 3 we plotted the median, the 10 %-, 25 %-, 75 %- and 90 %- quantiles of sample of 10 000 estimated values of extreme index by three considered estimators against the intermediate sequences k in the regression model. For the sake of comparison, the same procedures were performed on the (normally unobservable) errors to see how much is lost by estimating the regression coefficients. Notice that the performance of the estimators practically depends only on the distribution of errors and not on the structure of regression matrix.

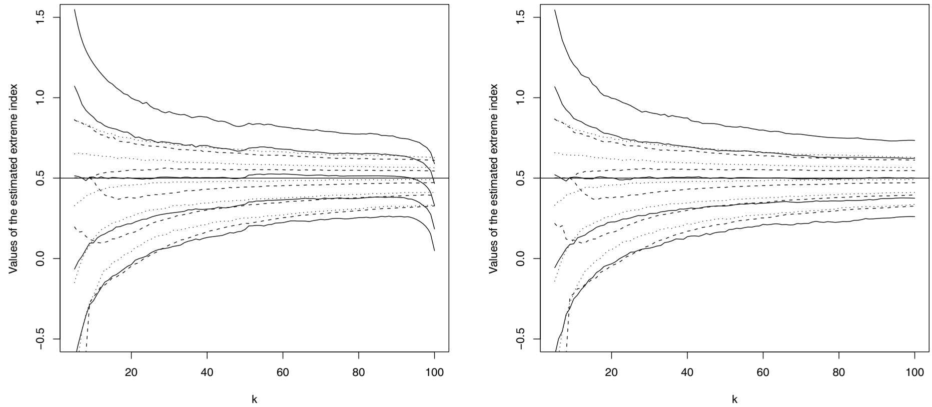

FIG 1. *The median, the 10 %-, 25 %-, 75 %- and 90 %- quantiles in the sample of 10 000 estimated values of extreme index by Pickands (solid), maximum likelihood (dotted) and probability weighted moments estimators (dashed) for Generalized Pareto distribution of errors with the shape parameter* $\gamma = 0.5$ *(denoted by the horizontal line) in the regression model (left) and in the location model with unobserved errors (right).*

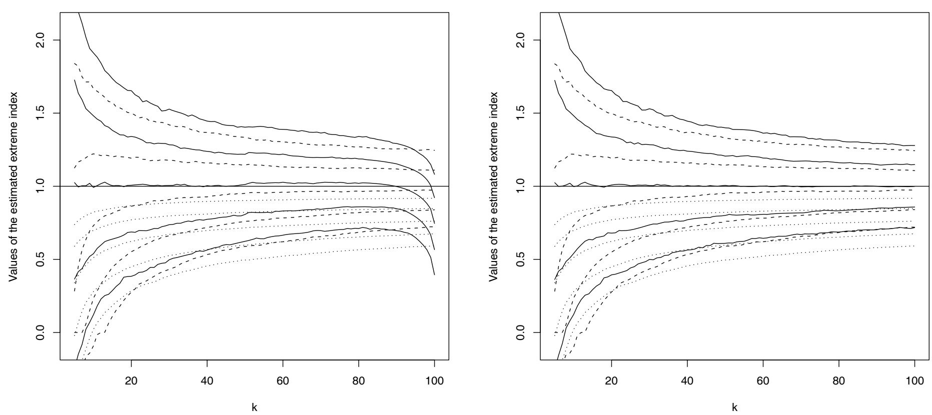

FIG 2. *The median, the 10 %-, 25 %-, 75 %- and 90 %- quantiles in the sample of 10 000 estimated values of extreme index by Pickands (solid), maximum likelihood (dotted) and probability weighted moments estimators (dashed) for Pareto distribution of errors the with shape parameter* $\gamma = 1$ *(denoted by the horizontal line) in the regression model (left) and in the location model with unobserved errors (right).*

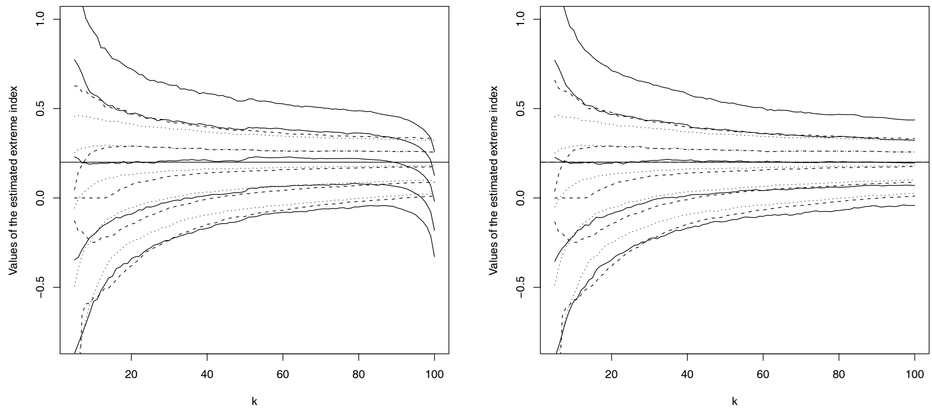

Fig 3. *The median, 10%-, 25%-, 75%- and 90%- quantiles in the sample of 10 000 estimated values of extreme index by Pickands (solid), maximum likelihood (dotted) and probability weighted moments estimators (dashed) for Burr distribution of errors with the shape parameter* $\gamma = 0.2$ *(denoted by the horizontal line) in the regression model (left) and in the location model with unobserved errors (right).*

The simulation study indicated:

(i) Results are affected by the specification of different values of k but the estimators give quite stable results for a suitable choice of fraction k. We see that the variance will be smallest for highest values of k.

(ii) The Pickands estimator, compared to the other estimators, shows a much larger variability. On the other hand, the maximum likelihood estimator is biased for the Pareto distribution. It is considered on the basis of the theoretical result that the threshold excesses have a corresponding approximate distribution within the Generalized Pareto family (see e. g. [1]). Hence, it seems that asymptotic result does not work properly in our situation.

(iii) The R-estimator is a solution of the optimization problem (2.2) in such a way it depends on initial values for the parameters to be optimized over. It seems from our simulation experiment that the resulting value of minimization does not depend (or depends very weakly) on the initial points. However, an unsuitable choice is the time expensive and it may complicate the computation considerably.

(iv) As we have verified on a considerably larger simulation experiment, the properties of the two-step regression quantiles are very weakly affected by the chosen α and by the form of the matrix.

Acknowledgments

The authors thank three referees for their careful reading and for their comments, which helped to improve the text.

References

[1] Beirlant, J., Goegebeur, Y., Teugels, J., and Segers, J. (2004). *Statistics of Extremes, Theory and Applications.* Wiley, Chichester.

[2] Chernozhukov, V. (2005). Extremal Quantile Regression. *Ann. Math. Statist.* **33 (2)** 806–839.

[3] de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory, An Introduction.* Springer, New York.

[4] Drees, H. . (1998) On Smooth Statistical Tail Functionals. *Scandinavian Journal of Statistics* **25** 187–210.

[5] Jaeckel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann. Math. Statist.* **43** 1449-1459.

[6] Jurečková, J. (1971). Nonparametric estimate of regression coefficients. *Ann. Math. Statist.* **42** 1328–1338.

[7] Jurečková, J. (1977). Asymptotic relation of M-estimates and R-estimates in the linear regression model. *Ann. Statist.* **5** 464–472.

[8] Jurečková, J. (2007). Remark on extreme regression quantile. *Sankhyā* **69** 87–100.

[9] Jurečková, J. and Picek, J. (2005). Two-step regression quantiles. *Sankhyā* 227–252.

[10] Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Inter-Relations.* J. Wiley, New York.

[11] Hodges, J. L. and Lehmann, E. L. (1963). Estimation of location based on rank tests. *Ann. Math. Statist.* **34** 598–611.

[12] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

[13] Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models. (Second Edition)* Springer, New York.

[14] Kyselý, J., Picek, J. and Beranová, R. (2010). Estimating extremes in climate change simulations using the peaks-over-threshold method with a non-stationary threshold. *Glob. Planet. Change*, doi:10.1016/j.gloplacha.2010.03.006.

[15] Portnoy, S. and Jurečková, J. (1999). On extreme regression quantiles. *Extremes* **2 (3)** 227–243.