Averaged extreme regression quantile

Jana Jurečková

Volume 19, Number 1, March 2016

Extremes

Statistical Theory and Applications in Science, Engineering and Economics

Springer

ISSN: 1386-1999

Springer

Springer

CrossMark

Averaged extreme regression quantile

Jana Jurečková[1]
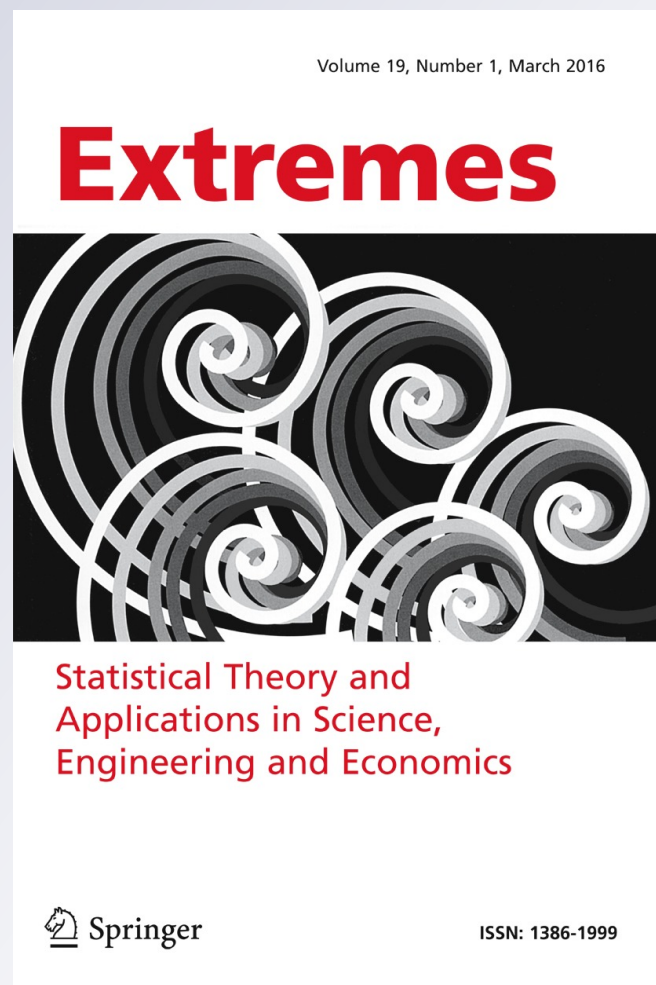
© Springer Science+Business Media New York 2015

Abstract Various events in the nature, economics and in other areas force us to combine the study of extremes with regression and other methods. A useful tool for reducing the role of nuisance regression, while we are interested in the shape or tails of the basic distribution, is provided by the *averaged regression quantile* and namely by the *average extreme regression quantile*. Both are weighted means of regression quantile components, with weights depending on the regressors. Our primary interest is the averaged extreme regression quantile (AERQ), its structure, qualities and its applications, e.g. in investigation of a conditional loss given a value exogenous economic and market variables. AERQ has several interesting equivalent forms: While it is originally defined as an optimal solution of a specific linear programming problem, hence is a weighted mean of responses corresponding to the optimal base of the pertaining linear program, we give another equivalent form as a maximum residual of responses from a specific R-estimator of the slope components of regression parameter. The latter form shows that while AERQ equals to the maximum of some residuals of the responses, it has minimal possible perturbation by the regressors. Notice that these finite-sample results are true even for non-identically distributed model errors, e.g. under heteroscedasticity. Moreover, the representations are formally true even when the errors are dependent - this all provokes a question of the right interpretation and of other possible applications.

Keywords Averaged extreme regression quantile · Extreme regression quantile ·
Regression rank scores · R-estimator

✉ Jana Jurečková
jurecko@karlin.mff.cuni.cz

1   Faculty of Mathematic and Physics, Charles University in Prague, Prague, Czech Republic

AMS 2000 Subject Classifications Primary: 60G70 · 62G30 · 62G32; Secondary: 62J05 · 90C05

1 Introduction

Consider the linear regression model

$$Y_{ni} = \beta_0 + \mathbf{x}_{ni}^\top \boldsymbol{\beta} + e_{ni}, \quad i = 1, \ldots, n \tag{1.1}$$

with observations $Y_{n1}, \ldots, Y_{nn}$, independent errors $e_{n1}, \ldots, e_{nn}$, possibly non-identically distributed with unknown distribution functions F_i, $i = 1, \ldots, n$. The covariates $\mathbf{x}_{ni} = (x_{i1}, \ldots, x_{ip})^\top$, $i = 1, \ldots, n$ are random or nonrandom, and $\boldsymbol{\beta}^* = (\beta_0, \boldsymbol{\beta}^\top)^\top = (\beta_0, \beta_1, \ldots, \beta_p)^\top \in \mathbb{R}^{p+1}$ is an unknown parameter. We shall suppress the subscript n whenever it does not cause a confusion. Occasionally we use the notation $\mathbf{x}_{ni}^* = (1, x_{i1}, \ldots, x_{ip})^\top$, $i = 1, \ldots, n$.

The regression α-quantile in model (1.1) was originated by Koenker and Bassett (1978). The extreme regression quantile (extreme RQ) was studied by Smith (1994) under *iid* model errors (under the name "nonregular regression"); he found its asymptotic distribution under heavy-tailed distribution F and under some conditions on the regressors. The extreme RQ under *iid* model errors was later studied in (Bassett 1988; Chernozhukov 2005; Jurečková 2007, 2007b; Knight 2001; Portnoy and Jurečková 1999), among others, under various distributions and under miscellaneous regularity conditions. The asymptotic distributions of the intercept and slope components of the extreme RQ were derived in (Jurečková 2007, 2007b) under *iid* errors for distributions with the tails lighter than t-distribution with 4 degrees of freedom and under some conditions on the regressors. In some cases the slope components of the extreme RQ even consistently estimate the slope parameters.

The regression α-quantile

$$\widehat{\boldsymbol{\beta}}_n^*(\alpha) = \left(\hat{\beta}_{n0}(\alpha), (\widehat{\boldsymbol{\beta}}_n(\alpha))^\top\right)^\top = \left(\hat{\beta}_{n0}(\alpha), \hat{\beta}_{n1}(\alpha), \ldots, \hat{\beta}_{np}(\alpha)\right)^\top$$

is a $(p+1)$-dimensional vector defined by means of minimization

$$\widehat{\boldsymbol{\beta}}_n^*(\alpha) = \arg \min_{\mathbf{b} \in R^{p+1}} \left\{ \sum_{i=1}^n \left[\alpha(Y_i - \mathbf{x}_i^{*\top}\mathbf{b})^+ + (1-\alpha)(Y_i - \mathbf{x}_i^{*\top}\mathbf{b})^-\right]\right\}$$

where $z^+ = \max(z, 0)$ and $z^- = \max(-z, 0)$, $z \in \mathbb{R}_1$. (1.2)

Particularly, if $\alpha = 1$, we get the extreme (maximal) regression quantile

$$\widehat{\boldsymbol{\beta}}_n^*(1) = \arg \min_{\mathbf{b} \in R^{p+1}} \sum_{i=1}^n (Y_i - \mathbf{x}_i^{*\top}\mathbf{b})^+.$$

By Eq. 1.2, $\widehat{\boldsymbol{\beta}}_n^*(1)$ can be alternatively described as any solution to the linear program:

$$\min_{\mathbf{b}^* \in R^{p+1}} \sum_{i=1}^n \mathbf{x}_i^{*\top}\mathbf{b}^* \qquad \text{s.t.} \quad Y_i \leq \mathbf{x}_i^{*\top}\mathbf{b}^*, \quad i = 1, \ldots, n. \tag{1.3}$$

If $\sum_{i=1}^{n} x_{ij} = 0$, $j = 1, \ldots, p$, then we only minimize the intercept component b_0, subject to the inequalities in Eq. 1.3. In the absence of regression, i.e. if $\mathbf{x}_i = \mathbf{0}$, $i = 1, \ldots, n$, the extreme regression quantile coincides with the extreme observation $Y_{n:n}$.

The weighted mean of components of $\widehat{\boldsymbol{\beta}}_n^*(\alpha)$, $0 \le \alpha \le 1$

$$B_n(\alpha) = \overline{\mathbf{x}}_n^{*\top} \widehat{\boldsymbol{\beta}}_n^*(\alpha) = \widehat{\beta}_{n0}(\alpha) + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{p} x_{ij} \widehat{\beta}_j(\alpha) \tag{1.4}$$

is called the averaged regression α-quantile. The behavior of $B_n(\alpha)$ with $0 < \alpha < 1$ has been illustrated in (Bassett 1988; Koenker and Bassett 1982), and summarized in (Koenker 2005). Interesting is the monotonicity of $B_n(\alpha)$ in $\alpha \in (0, 1)$.

In the present paper, we are primarily interested in the structure, properties and possible use of averaged extreme regression quantile (AERQ) $B_n(1)$, corresponding to $\alpha = 1$. It is a useful tool when we are interested in the extreme behavior of our observations while they are affected by a regression.

The quantile regression was considered by a host of authors in investigation of a conditional loss given a value of a covariate $\mathbf{x}$, involving exogenous economic and market variables (or otherwise the past observed returns). Such problems we meet in the finance, but also in the insurance and in the social statistics.

Particularly, in the context of the extreme regression quantile, we consider Y_i as the loss variable under realization $\mathbf{x}_i$, representing economic and market (exogenous) variable. Assuming that the relation of the loss and covariates is described by model (1.1), our primary goal is to find $\widehat{\boldsymbol{\beta}}^* = (\hat{\beta}_0, \widehat{\boldsymbol{\beta}})^\top$ leading to the minimum sum of positive residuals, those satisfying $Y_i > \mathbf{x}_i^{*\top} \boldsymbol{\beta}^*$. This reflects greater concern about underestimating losses Y, rather than overestimating, emphasized e.g. in (Rockafellar et al. 2014). Under continuous distribution functions F_i of e_i, $i = 1, \ldots, n$, with probability 1 we obtain

$$\overline{Y}_n - B_n(1) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i^{*\top} \widehat{\boldsymbol{\beta}}_n^*(1))$$

$$= \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i^{*\top} \widehat{\boldsymbol{\beta}}_n^*(1))^+ - \frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i^{*\top} \widehat{\boldsymbol{\beta}}_n^*(1))^- \tag{1.5}$$

$$= -\frac{1}{n} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i^{*\top} \widehat{\boldsymbol{\beta}}_n^*(1))^- \le 0.$$

This can be interpreted as the expected conditional shortfall of the loss, which is non-positive under $\boldsymbol{\beta}^* = \widehat{\boldsymbol{\beta}}_n^*(1)$. We refer to (Trindade et al. 2007; Cai and Wang 2008; Rockafellar and Uryasev 2013; Rockafellar et al. 2014) for the discussions and various solutions, and to other papers cited in. For possible applications in the insurance we refer to (Chan 2015).

The structure of $B_n(1)$ is of interest. Because $B_n(1)$ is an outcome of a linear programming, it is identical to a weighted average of $p + 1$ among observations $(Y_1, \ldots, Y_n)$, corresponding to the optimal base of the linear program (1.3). We shall give another equivalent form of $B_n(1)$, which employs a suitable R-estimator of the

slope components $\boldsymbol{\beta}$. Combining two equivalent forms indicates that $B_n(1)$ never exceeds $\max_{i \le n} Y_i$, while relative to $\max_{i \le n} Y_i$ it has minimum possible perturbations caused by the regressors, with minimum over values of $\boldsymbol{\beta}$.

2 Finite-sample form of $B_n(1)$ based on optimal base

Consider the linear program (1.3); let $\{\mathbf{x}_{i_1}, \ldots, \mathbf{x}_{i_{p+1}}\}$ be its optimal base and let $\{Y_{i_1}, \ldots, Y_{i_{p+1}}\}$ be the corresponding responses. It follows from the linear programming theory that $B_n(1)$ equals to a weighted mean of $\{Y_{i_1}, \ldots, Y_{i_{p+1}}\}$, with the weights based on the regressors. However, in this special case all weights are positive, as it is specified in the following theorem.

Theorem 1 *Assume that the regression matrix* $\mathbf{X}_n^* = \begin{bmatrix} \mathbf{x}_{n1}^{*\top} \\ \cdots \\ \mathbf{x}_{nn}^{*\top} \end{bmatrix}$ *has full rank* $p+1$ *and that the distribution functions* $F_1, \ldots, F_n$ *of model errors are continuous and increasing in* $(-\infty, \infty)$. *Then with probability 1*

$$B_n(1) = \sum_{k=1}^{p+1} w_k Y_{i_k}, \quad w_k > 0, \ k = 1, \ldots, p+1, \ \sum_{k=1}^{p+1} w_k = 1 \quad (2.1)$$

$$\overline{Y}_n \le B_n(1) < \max_{i \le n} Y_i \quad (2.2)$$

where $\overline{Y}_n = \frac{1}{n} \sum_{i=1}^{n} Y_i$ *and the vector* $\mathbf{Y}_n(1) = (Y_{i_1}, \ldots, Y_{i_{p+1}})^\top$ *corresponds to the optimal base of the linear program (1.3).*

The vector $\mathbf{w} = (w_1, \ldots, w_{p+1})^\top$ *of coefficients in Eq. 2.1 has the form*

$$\mathbf{w}^\top = n^{-1} \mathbf{1}_n^\top \mathbf{X}_n^* (\mathbf{X}_{n1}^*)^{-1}; \quad (2.3)$$

where $\mathbf{X}_{n1}^*$ *is the submatrix of* $\mathbf{X}_n^*$ *with the rows* $\mathbf{x}_{i_1}^{*\top}, \ldots, \mathbf{x}_{i_{p+1}}^{*\top}$.

Proof The dominance of $B_n(1)$ over the average $\overline{Y}_n$ follows directly from the definition of $B_n(1)$. The dominance of $\max_{i \le n} Y_i$ over $B_n(1)$ will follow from the fact proven below that the coefficients in Eq. 2.1 are all positive and sum up to 1.

To prove the identity in Eq. 2.1, consider the minimization Eq. 1.2 for $0 \le \alpha \le 1$ as a parametric linear programming. Then its dual program is a parametric linear programming of the form

$$\begin{aligned} &\text{maximize } \mathbf{Y}_n^\top \hat{\mathbf{a}}(\alpha) \\ &\text{under } \mathbf{X}_n^{*\top} \hat{\mathbf{a}}(\alpha) = (1-\alpha) \mathbf{X}_n^{*\top} \mathbf{1}_n^\top \\ &\hat{\mathbf{a}}(\alpha) \in [0, 1]^n, \ 0 \le \alpha \le 1. \end{aligned} \quad (2.4)$$

The components of the optimal solution $\hat{\mathbf{a}}(\alpha) = (\hat{a}_{n1}(\alpha), \ldots, \hat{a}_{nn}(\alpha))^\top$ of Eq. 2.4, called regression rank scores, were studied in (Gutenbrunner and Jurečková, J. 1992), where it is shown that $\hat{a}_{ni}(\alpha)$ is a continuous, piecewise linear function of $\alpha \in [0, 1]$

and $\hat{a}_{ni}(0) = 1$, $\hat{a}_{ni}(1) = 0$, $i = 1, \ldots, n$. Moreover, $\hat{\mathbf{a}}(\alpha)$ is invariant in the sense that it does not change if $\mathbf{Y}$ is replaced with $\mathbf{Y} + \mathbf{X}_n \mathbf{b}$, $\forall \mathbf{b}^* \in \mathbb{R}^{p+1}$ (see Gutenbrunner and Jurečková, J. 1992 for detail). The regression quantile $\widehat{\boldsymbol{\beta}}_n^*(\alpha)$ is a step function of $0 \leq \alpha \leq 1$. If α is a continuity point of the regression quantile trajectory, then we have the following identities, proven in (Jurečková and Picek 2014):

$$B_n(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{*\top} \widehat{\boldsymbol{\beta}}^*(\alpha) = -\frac{1}{n} \sum_{i=1}^{n} Y_i \hat{a}_{ni}'(\alpha) \tag{2.5}$$

$$\text{or} \quad B_n(\alpha) - \bar{\mathbf{x}}_n^{*\top} \boldsymbol{\beta}^* = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i^{*\top} (\widehat{\boldsymbol{\beta}}^*(\alpha) - \boldsymbol{\beta}^*) = -\frac{1}{n} \sum_{i=1}^{n} e_i \hat{a}_{ni}'(\alpha)$$

where $\hat{a}_{ni}'(\alpha)) = \frac{d}{d\alpha} \hat{a}_{ni}(\alpha)$. Moreover, Eq. 2.4 implies

$$\sum_{i=1}^{n} x_{ij}^* \hat{a}_{ni}'(\alpha) = -\sum_{i=1}^{n} x_{ij}^*, \ 1 \leq i \leq n, \ 1 \leq j \leq p, \tag{2.6}$$

and particularly,

$$\frac{1}{n} \sum_{i=1}^{n} \hat{a}_{ni}'(\alpha) = -1. \tag{2.7}$$

Notice that $\hat{a}_{ni}'(\alpha) \neq 0$ iff α is the point of continuity of $\widehat{\boldsymbol{\beta}}_n^*(\cdot)$ and $Y_i = \mathbf{x}_i^{*\top} \widehat{\boldsymbol{\beta}}_n^*(\alpha)$, $i = 1, \ldots, n$. There are exactly $p + 1$ such components pertaining to a fixed continuity point α; those for which $\mathbf{x}_i^*$ belongs to the optimal base of program (2.4). The number of points of discontinuity of $\widehat{\boldsymbol{\beta}}_n(\cdot)$ is finite under a finite n. Denote $1 - \varepsilon_1 < 1$ the largest point such that, for $1 - \varepsilon_1 < \alpha < 1$ all $\hat{a}_{ni}(\alpha)$ are non-increasing, and let $1 - \varepsilon_2 \in (1 - \varepsilon_1, 1)$ be the largest point of discontinuity of $\widehat{\boldsymbol{\beta}}_n(\cdot)$. Then either $\hat{a}_{ni}'(\alpha) = 0$ or $\frac{1}{\varepsilon_1} < -\hat{a}_{ni}'(\alpha) < \frac{1}{\varepsilon_2}$ for $1 - \varepsilon_1 < \alpha < 1$; moreover, $1 - \varepsilon_1 \leq 1 - \frac{1}{n} \leq 1 - \varepsilon_2$ by Eq. 2.7. Hence, for $1 - \varepsilon_1 < \alpha < 1$ are $\hat{a}_{ni}(\alpha)$ decreasing and $\hat{a}_{ni}'(\alpha) < 0$ for exactly $p + 1$ points $i_1, \ldots, i_{p+1}$, while $\hat{a}_{ni}'(\alpha) = 0$ otherwise.

Moreover, by Eqs. 2.6 and 2.7

$$\sum_{i=1}^{n} \hat{a}_{ni}'(1) = -n$$

$$\sum_{i=1}^{n} x_{ij} \hat{a}_{ni}'(1) = -\sum_{i=1}^{n} x_{ij}, \ j = 1, \ldots, p$$

where $\hat{a}_{ni}'(1)$ is the left-hand derivative of $\hat{a}_{ni}(\alpha)$ at $\alpha = 1$. The identity Eq. 2.5 particularly implies

$$B_n(1) = -\frac{1}{n} \sum_{i=1}^{n} Y_i \hat{a}_{ni}'(1), \tag{2.8}$$

while exactly $p + 1$ coefficients in Eq. 2.8 are different from zero, those corresponding to the base. Hence we have $B_n(1) = \sum_{k=1}^{p+1} w_k Y_{i_k}$.

Notice that the inequalities among the constraints in Eq. 1.3 reconvert in equalities for just $p + 1$ components of the optimal base $\mathbf{x}_{i_1}^*, \ldots, \mathbf{x}_{i_{p+1}}^*$. Let $\mathbf{X}_{n1}^*$ be the submatrix

of $\mathbf{X}_n^*$ with the rows $\mathbf{x}_{i_1}^{*\top}, \ldots, \mathbf{x}_{i_{p+1}}^{*\top}$ and let $(\hat{\mathbf{a}}'(1))^\top = (\hat{a}'_{i_1}(1), \ldots, \hat{a}'_{i_{p+1}}(1))$. Then $\mathbf{X}_{n1}^*$ is regular with probability 1 and

$$\mathbf{w}^\top = (-\hat{\mathbf{a}}'(1))^\top = \mathbf{1}_n^\top \mathbf{X}_n^* (\mathbf{X}_{n1}^*)^{-1}.$$

This and Eqs. 2.8 imply 2.1 and 2.3.   □

Remark 1 In the location case, where $Y_i = \beta_0 + e_i$, $i = 1, \ldots, n$, we have $\mathbf{X}_n^* = \mathbf{1}_n$ and the regression rank scores have the form (so called Hájek's rank scores, see 1965)

$$\hat{a}_{ni}(\alpha) = \begin{cases} 1 & \ldots\ 0 \leq \alpha \leq \frac{R_i - 1}{n} \\ R_i - n\alpha & \ldots\ \frac{R_i - 1}{n} < \alpha \leq \frac{R_i}{n} \\ 0 & \ldots\ \frac{R_i}{n} \leq \alpha \leq 1 \end{cases}, \qquad i = 1, \ldots, n$$

where R_i is rank of Y_i, $i = 1, \ldots, n$. Hence,

$$\hat{a}'_{ni}(\alpha) = \begin{cases} -n & \ldots\ \frac{R_i - 1}{n} < \alpha < \frac{R_i}{n} \\ 0 & \ldots\ 0 < \alpha < \frac{R_i - 1}{n} \text{ or } \frac{R_i}{n} < \alpha < 1 \end{cases}$$

and particularly $\hat{a}'_{ni}(1) = -n$ if $R_i = n$, what corresponds to the maximal Y_i, and $\hat{a}'_{ni}(1) = 0$ otherwise. Naturally, then $B_n(1) = \max_{i \leq n} Y_i$. However, it follows from Eq. 2.1 that $B_n(1) < \max_{i \leq n} Y_i$ sharply in the presence of regression. The asymptotic behavior of $\max_{i \leq n} Y_{ni} - B_n(1)$ as $n \to \infty$ in case of i.i.d. errors is of interest, and will be a subject of a further study.

3 Finite-sample expression of $B_n(1)$ using the R-estimator of slopes

The identity in Eq. 2.1 does not reveal any relation of $B_n(1)$ to the extreme $Y_{n:n}$ or to some extreme residuals, because the optimal base of the linear program can be found only numerically, not explicitly. An alternative form of the extreme regression quantile, considered by the author in (Jurečková 2007), treats the slope and intercept components in model (1.1) separately: We start with a specific R-estimate $\widetilde{\boldsymbol{\beta}}_{nR}^+$ of the slope components $\boldsymbol{\beta}$, based on the ranks of residuals $Y_i - \mathbf{x}_i^\top \mathbf{b}$, $i = 1, \ldots, n$. This R-estimate $\widetilde{\boldsymbol{\beta}}_{nR}^+$ is defined as the minimizer of the Jaeckel (1972) measure of the rank dispersion

$$\widetilde{\boldsymbol{\beta}}_{nR}^+ = \arg\min\{\mathcal{D}_n(\mathbf{b}) : \mathbf{b} \in \mathbb{R}^p\}, \tag{3.1}$$

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^n (Y_i - \mathbf{x}_i^\top \mathbf{b}) \varphi_n\left(\frac{R_{ni}(Y_i - \mathbf{x}_i^\top \mathbf{b})}{n+1}\right)$$

with the "extreme" score function

$$\varphi_n(u) = I[u \geq 1 - \tfrac{1}{n}] - \tfrac{1}{n}, \quad 0 \leq u \leq 1, \tag{3.2}$$

where $R_{ni}(Y_i - \mathbf{x}_i^\top \mathbf{b})$ is the rank of the residual $Y_i - \mathbf{x}_i^\top \mathbf{b}$ among
$Y_1 - \mathbf{x}_1^\top \mathbf{b}, \ldots, Y_n - \mathbf{x}_n^\top \mathbf{b}, \mathbf{b} \in \mathbb{R}^p, \ i = 1, \ldots, n$. The corresponding rank scores are

$$a_n(i) = \varphi_n\left(\tfrac{i}{n+1}\right) = I[i = n] - \tfrac{1}{n}, \quad i = 1, \ldots, n.$$

Hence,

$$\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^n (Y_i - \mathbf{x}_i^\top \mathbf{b}) I[R_{ni}(Y_i - \mathbf{x}_i^\top \mathbf{b}) = n] - (\bar{Y}_n - \bar{\mathbf{x}}_n^\top \mathbf{b}), \tag{3.3}$$

$$= \max_{1 \leq i \leq n} \{Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \mathbf{b}\} - \bar{Y}_n = \left(Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \mathbf{b}\right)_{n:n} - \bar{Y}_n.$$

By Jaeckel (1972), the measure $\mathcal{D}_n(\mathbf{b})$ is continuous, convex and piecewise linear function of $\mathbf{b} \in \mathbb{R}^p$. The estimate $\widetilde{\boldsymbol{\beta}}_{nR}^+$ is invariant to the shift in location, thus it estimates only $\beta_1, \ldots, \beta_p$. Define the additional intercept component $\widetilde{\beta}_{n0}^+$ as

$$\widetilde{\beta}_{n0}^+ = \max\{Y_i - \mathbf{x}_i^\top \widetilde{\boldsymbol{\beta}}_{nR}^+, \ 1 \leq i \leq n\}. \tag{3.4}$$

Denote

$$\widetilde{\boldsymbol{\beta}}_{nR}^{+*} = \begin{pmatrix} \widetilde{\beta}_{n0}^+ \\ \widetilde{\boldsymbol{\beta}}_{nR}^+ \end{pmatrix}. \tag{3.5}$$

By (Jurečková and Picek 2005), $\widetilde{\boldsymbol{\beta}}_{nR}^{+*}$ coincides with the maximal regression quantile $\widehat{\boldsymbol{\beta}}_n^*(1)$. The following theorem shows that $B_n(1)$ is identical to the maximal residual of Y_i's from the R-estimator $\widetilde{\boldsymbol{\beta}}_{nR}^+$. It further implies that $B_n(1)$ is identical to the weighted mean of components of the vector $\widetilde{\boldsymbol{\beta}}_{nR}^{+*}$ with coefficients based on the regressors.

Theorem 2 *Under the conditions of Theorem 1,*

(i)

$$B_n(1) = \max_{1 \leq i \leq n} \{Y_{ni} - (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)^\top \widetilde{\boldsymbol{\beta}}_{nR}^+\} = \max_{1 \leq i \leq n} (Y_{ni} - \widetilde{Y}_{ni}^R), \tag{3.6}$$

where $\quad \widetilde{Y}_{ni}^R = (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)^\top \widetilde{\boldsymbol{\beta}}_{nR}^+, \ i = 1, \ldots, n.$

(ii)

$$B_n(1) = \widehat{\beta}_{n0}(1) + \bar{\mathbf{x}}_n^\top \widehat{\boldsymbol{\beta}}(1) = \widetilde{\beta}_{n0}^+ + \bar{\mathbf{x}}_n^\top \widetilde{\boldsymbol{\beta}}_{nR}^+. \tag{3.7}$$

Proof Indeed, by Eq. 3.3, $\widetilde{\boldsymbol{\beta}}_{nR}^+$ minimizes the extreme residual $(Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \mathbf{b})_{n:n}$ with respect to $\mathbf{b} \in \mathbb{R}^p$. Hence, by Eqs. 3.4 and 3.5,

$$\widetilde{\beta}_{n0}^+ + \bar{\mathbf{x}}_n^\top \widetilde{\boldsymbol{\beta}}_{nR}^+ = \max_{1 \leq i \leq n} \{Y_i - (\mathbf{x}_i - \bar{\mathbf{x}}_n)^\top \widetilde{\boldsymbol{\beta}}_{nR}^+\}.$$

On the other hand, the extreme regression quantile minimizes $b_0 + \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i^\top \mathbf{b}$ among all $(b_0, \mathbf{b}^\top)^\top \in \mathbb{R}^{p+1}$ such that $b_0 + \mathbf{x}_i^\top \mathbf{b} \geq Y_i$, $i = 1, \ldots, n$. Because of Eqs. 3.3 and 3.4, if $b_0 + \mathbf{x}_i^\top \mathbf{b} \geq Y_i$ for $i = 1, \ldots, n$, then

$$b_0 + \overline{\mathbf{x}}_n^\top \mathbf{b} \geq \max_{1 \leq i \leq n}\{Y_i - b_0 - \mathbf{x}_i^\top \mathbf{b}\} + b_0 + \overline{\mathbf{x}}_n^\top \mathbf{b} = \widetilde{\beta}_{n0}^+ + \overline{\mathbf{x}}_n^\top \widetilde{\boldsymbol{\beta}}_{nR}^+$$

what confirms Eq. 3.7 and hence also (3.6).   □

Corollary 1 *Specifically, the AERQ satisfies the following inequalities:*

(i) $B_n(1) \leq \max_{i \leq n}\{Y_i - (\mathbf{x}_i - \overline{\mathbf{x}}_n)^\top \mathbf{b}\}$ $\forall \mathbf{b} \in \mathbb{R}^p$.

(ii) $B_n(1) < e_{n:n} + \beta_0 + \overline{\mathbf{x}}_n^\top \boldsymbol{\beta}$, *where* $e_{n:n} = \max_{i \leq n}\{e_1, \ldots, e_n\}$.

(iii) $B_n(1) \leq \max_{i \leq n}\{Y_i - (\mathbf{x}_i - \bar{x}_n)^\top \widehat{\boldsymbol{\beta}}_n\}$ *for any estimator of* $\boldsymbol{\beta}$.

Hence, the AERQ is equal to maximum of perturbed observations Y_i, $i = 1, \ldots, n$, *where the perturbation is minimum possible caused by the regression with regressors* $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Indeed, the inequalities follow from Eq. 3.3 for any $\mathbf{b} \in \mathbb{R}^p$, and particularly for $\mathbf{b} = \mathbf{0}$ and for any estimate $\widehat{\boldsymbol{\beta}}$. The sharp inequality in *(ii)* follows from Theorem 1. Though $B_n(1)$ is less than $\max_{i \leq n} Y_i$, the perturbations of $\max_{i \leq n} Y_i$ are minimum possible.

Corollary 2 *The expected conditional shortfall Eq. 1.5 of the loss Y under* $\boldsymbol{\beta}^* = \widehat{\boldsymbol{\beta}}_n^*(1)$ *can be alternatively written as*

$$\frac{1}{n}\sum_{i=1}^{n}\left[Y_i - \mathbf{x}_i^{*\top}\widehat{\boldsymbol{\beta}}_n^*(1)\right] = \overline{Y}_n - \max_{i \leq n}\left\{Y_i - (\mathbf{x}_i - \overline{\mathbf{x}}_n)^\top \widetilde{\boldsymbol{\beta}}_{nR}^+\right\} \leq 0$$

4 Conclusion

The extreme (maximal) regression quantile in the linear regression model is a value of the regression parameter which minimizes the sum of the positive residuals. The averaged extreme regression quantile is a useful tool in simultaneous study of extremes and regression which indicates how the regression affects the extreme observations. Two equivalent forms of the AERQ indicate that in the presence of regression it is sharply smaller than $\max_{i \leq n} Y_i$, but the perturbations are the smallest possible over possible values of $\boldsymbol{\beta}$. There are promising applications in the analysis of the risk or losses.

References

Bassett, G.W. Jr.: A property of the observations fit by the extreme regression quantiles. Comp. Stat. Data Anal. **6**, 353–359 (1988)

Cai, Z., Wang, X.: Nonparametric estimation of conditional VaR and expected shortfall. J. Econ. **147**, 120–130 (2008)

Chan, J.S.K.: Predicting loss reserves using quantile regression [Quantile regression loss reserve models]. J Data Sci **13**, 127–156 (2015)

Chernozhukov, V.: Extremal quantile regression. Ann. Stat. **30**, 806–839 (2005)

Gutenbrunner, C., Jurečková, J.: Regression rank scores and regression quantiles. Ann. Stat. **20**, 305–330 (1992)

Hájek, J.: Extension of the Kolmogorov-Smirnov test to regression alternatives. In: LeCam, L. (ed.) Proceedings of the International Research Seminar, pp. 45–60. Univ. of California Press, Berkeley (1965)

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

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

Jurečková, J.: Remark on extreme regression quantile II, Bull. of the Intern. Statist. Institute, Proceedings of the 56th Session, Section CPM026 (2007b)

Jurečková, J., Picek, J.: Two-step regression quantiles. Sankhyā **67**, 227–252 (2005)

Jurečková, J., Picek, J.: Averaged regression quantiles. In: Lahiri, S., et al. (eds.) Contemporary Developments in Statistical Theory, vol. 68, pp. 203–216. Springer Proc. in Math. & Statistics (2014)

Knight, K.: Limiting distributions for linear programming estimators. Extremes **4**, 87–103 (2001)

Koenker, R.: Quantile Regression. Cambridge University Press (2005)

Koenker, R., Bassett, G.: Regression quantiles. Econometrica **46**, 33–50 (1978)

Koenker, R., Bassett, G.: An empirical quantile function for linear models with iid errors. J. Am. Stat. Assoc. **77**, 407–415 (1982)

Portnoy, S., Jurečková, J.: On extreme regression quantiles. Extremes **2**, 227–243 (1999)

Rockafellar, R.T., Uryasev, S.: The fundamental risk quadrangle in risk management, Optimization and Statistical Estimation. Surveys in Operations Research and Management Science **18**, 33–53 (2013)

Rockafellar, R.T., Royset, J.O., Miranda, S.I.: Superquantile regression with applications to buffered reliability, uncertainty quantification, and conditional value-at-risk. Eur. J. Oper. Res. **234**, 140–154 (2014)

Smith, R.: Nonregular regression. Biometrika **81**, 173–183 (1994)

Trindade, A.A., Uryasev, S., Shapiro, A., Zrazhevsky, G.: Financial prediction with constrained tail risk. J. Bank. Financ. **31**, 3524–3538 (2007)