Asymptotics for Extreme Regression Quantiles

J. Jurečková [1]

[1] Department of Probability and Statistics, Charles University in Prague, Prague, Czech Republic

Keywords: Regression Quantile, R-estimator, Extreme Regression Quantile.

Abstract

Consider the linear regression model

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

with observations $\mathbf{Y} = (Y_1, \ldots, Y_n)'$, i.i.d. errors $\mathbf{E} = (E_1, \ldots, E_n)'$ with an unknown distribution function F, increasing on the set $\{x : 0 < F(x) < 1\}$, and unknown parameter $\boldsymbol{\beta}^* = (\beta_0, \beta_1, \ldots, \beta_p)'$. The extreme (maximal) regression quantile is defined as a solution of the linear program $\sum_{i=1}^n (b_0 + \mathbf{x}_i'\mathbf{b}) =:$ min under the restrictions $b_0 + \mathbf{x}_i'\mathbf{b} \geq Y_i,\ i = 1, \ldots, n,\ b_0 \in \mathbf{R},\ \mathbf{b} \in \mathbf{R}^p$. Jurečková and Picek (2005) showed that the extreme regression quantile can be equivalently written in a two step version, starting with an R-estimator $\widetilde{\boldsymbol{\beta}}_{nR}$ of the slope parameters, generated by the score function $\varphi(u) = I[u \geq 1 - \frac{1}{n}] - \frac{1}{n},\ 0 \leq u \leq 1$, and then ordering the residuals with respect to $\widetilde{\boldsymbol{\beta}}_{nR}$. Jurečková (2005) showed that, provided the density f of the E_i belongs to the domain of attraction of the Gumbel extreme distribution and $nf(F^{-1}(1 - \frac{1}{n})) \to \infty$ as $n \to \infty$, the slope component $\widetilde{\boldsymbol{\beta}}_{nR}$ of the extreme regression quantile consistently estimates $\boldsymbol{\beta}$ and admits the asymptotic representation

$$nf\left(F^{-1}(1 - \tfrac{1}{n})\right)\left[\widetilde{\boldsymbol{\beta}}_{nR}(1 - \tfrac{1}{n}) - \boldsymbol{\beta}\right] \tag{2}$$

$$= n\Big(\sum_{i=1}^n (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)'\Big)^{-1} \sum_{j=1}^n (\mathbf{x}_{nj} - \bar{\mathbf{x}}_n)\left[a_n(R_j(\mathbf{0}), 1 - \tfrac{1}{n}) - (1 - \tfrac{1}{n})\right] + o_p(1)\Big[= O_p(1)\Big]$$

where $R_j(\mathbf{0})$ is the rank of Y_i among $Y_1, \ldots, Y_n$ under $\boldsymbol{\beta} = \mathbf{0}$, $\bar{\mathbf{x}}_n = n^{-1}\sum_{i=1}^n \mathbf{x}_{ni}$ and

$$a_n(j, \alpha) = \begin{cases} 0, & j \leq n\alpha, \\ j - n\alpha, & n\alpha \leq j \leq n\alpha + 1, \\ 1, & n\alpha + 1 \leq j, \quad j = 1, \ldots, n. \end{cases}$$

are Hájek's rank scores. If $\mathbf{x}_{n1}, \ldots, \mathbf{x}_{nn}$ are random, independent of $E_1, \ldots, E_n$, and create a random sample from a p-variate distribution function H with expectation $\mathbf{0}$ and satisfying $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_{ni}\mathbf{x}_{ni}' \xrightarrow{p} \mathbf{Q}$ as $n \to \infty$, with a positively definite matrix $\mathbf{Q}$ of order $p \times p$, then the representation (2) changes to the form

$$nf\left(F^{-1}(1 - \tfrac{1}{n})\right)\left[\widetilde{\boldsymbol{\beta}}_{nR}(1 - \tfrac{1}{n}) - \boldsymbol{\beta}\right] \tag{3}$$

$$= \mathbf{Q}^{-1}\sum_{j=1}^n \mathbf{x}_{nj}\left[a_n(R_j(\mathbf{0}), 1 - \tfrac{1}{n}) - (1 - \tfrac{1}{n})\right] + o_p(1)\Big[= O_p(1)\Big].$$

The representations (2) and (3) enable to derive the asymptotic distributions of
$\Big\{nf\left(F^{-1}(1 - \tfrac{1}{n})\right)\left[\widetilde{\boldsymbol{\beta}}_{nR}(1 - \tfrac{1}{n}) - \boldsymbol{\beta}\right]\Big\}_{n=1}^{\infty}$ both for the random and nonrandom $\mathbf{x}_{ni}$.

Acknowledgement

The research was partially supported by the Czech Republic Grant 201/05/2340 and by the Research Project MSM 0021620839.

References

J. Hájek (1965). Extension of the Kolmogorov-Smirnov test to regression alternatives. *Proc. of Bernoulli-Bayes-Laplace Seminar* (L. LeCam, ed.), pp. 45–60. Univ. of California Press.

J. Jurečková (2005). Regression Quantiles and Hájek's Rank Scores. *ICORS'2005* (abstract).

J. Jurečková, J. Picek (2005). Two-step regression quantiles. Submitted.

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