Averaged Extreme Regression Quantile

Jana Jurečková[1]
[1] Charles University in Prague, Czech Republic

Consider the linear regression model

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

with observations $Y_{n1}, \dots, Y_{nn}$, independent errors $e_{n1}, \dots, e_{nn}$, identically distributed according to an unknown distribution function F;
$\mathbf{x}_{ni} = (x_{i1}, \dots, x_{ip})^\top$ with $x_{i1} = 1,\ i = 1, \dots, n$ is the vector of covariates, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^\top$ is an unknown parameter. The *maximal regression quantile* $\widehat{\boldsymbol{\beta}}_n(1)$ is a solution of the minimization problem:

$$\min_{\mathbf{b} \in \boldsymbol{R}^p} \sum_{i=1}^{n} (Y_i - \mathbf{x}_i^\top \mathbf{b})^+,$$

which can be alternatively described as any solution to the linear program:

$$\min_{\mathbf{b} \in \boldsymbol{R}^p} \sum_{i=1}^{n} \mathbf{x}_i^\top \mathbf{b} \qquad \text{s.t.} \quad Y_i \le \mathbf{x}_i^\top \mathbf{b}, \quad i = 1, \dots, n.$$

We are interested in the averaged extreme regression quantile $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_{ni}^\top \widehat{\boldsymbol{\beta}}_n(1)$ and in the relation of $\frac{1}{n}\sum_{i=1}^n \mathbf{x}_{ni}^\top (\widehat{\boldsymbol{\beta}}_n(1) - \boldsymbol{\beta})$ to the extreme quantiles of model errors $e_{n1}, \dots, e_{nn}$. Jurečková and Picek (2014) studied the averaged regression α-quantile $\widehat{\boldsymbol{\beta}}_n(\alpha)$ for $0 < \alpha < 1$, and proved that

$$n^{1/2}[\overline{\mathbf{x}}_n^\top (\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}) - e_{n:[n\alpha]}] = \mathcal{O}_p(n^{-1/4}) \ \text{ as } \ n \to \infty.$$

The situation is different with the averaged extreme regression quantile. Rewrite the regression model in the form $\mathbf{Y}_n = \mathbf{X}_n \boldsymbol{\beta} + \mathbf{e}_n$, and consider the following linear programming problem, connected with $\widehat{\boldsymbol{\beta}}_n(1)$:

$$\begin{aligned} \mathbf{Y}_n^T \mathbf{A} &= \max \\ \mathbf{X}_n^T \mathbf{A} &= \mathbf{X}_n^T \mathbf{1}_n \\ A_i \ge 0, & \quad i = 1, \dots, n \end{aligned}$$

with the optimal base $\mathbf{X}_1$, a submatrix of $\mathbf{X}_n$, of order $p \times p$. Then

$$\overline{\mathbf{x}}_n^\top (\widehat{\boldsymbol{\beta}}_n(1) - \boldsymbol{\beta}) = \sum_{i=1}^{p} \lambda_i \, e_{n:\, n-i+1} \ \text{ for any finite } \quad n > p,$$

where $e_{n:1} \le \ldots \le e_{n:n}$ are order statistics corresponding to $e_1, \dots, e_n$ and

$$\boldsymbol{\lambda}^\top = (\lambda_1, \dots, \lambda_p) = \frac{1}{n} \mathbf{1}_n^\top \mathbf{X}_n \mathbf{X}_1^{-1}.$$

The coefficients λ_i are obviously random and satisfy $\lambda_i > 0,\ i = 1, \ldots, p$ and $\sum_{i=1}^{p} \lambda_i = 1$.

Keywords: α-regression quantile, Extreme regression quantile, Averaged extreme regression quantile

References:

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

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

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

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

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

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