Regression Quantiles and Hájek's Rank Scores

J. Jurečková[1]

[1] Charles University in Prague, Sokolovská 83, CZ-186 75 Prague 8, Czech Republic

Keywords: Regression quantile, R-estimator, Hájek's rank scores, 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 $n \times p$ matrix $\mathbf{X} = \mathbf{X}_n$ is known and $\mathbf{1}_n = (1, \ldots, 1)' \in \mathbb{R}^n$. The α-regression quantile $\left(\hat{\beta}_{n0}(\alpha), \widehat{\boldsymbol{\beta}}'_n(\alpha)\right)'$ is defined as a solution of the minimization

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - b_0 - \mathbf{x}'_i \mathbf{b}) := \min, \quad b_0 \in \mathbb{R}^1, \ \mathbf{b} \in \mathbb{R}^p \tag{2}$$

where $\rho_\alpha(x) = |x|\{\alpha I[x > 0] + (1 - \alpha)I[x < 0]\}, \ x \in \mathbb{R}^1$. Then the R-estimator $\widetilde{\boldsymbol{\beta}}_{nR}(\alpha)$ of $\boldsymbol{\beta}$, generated by the score function $\varphi_\alpha(u) = \alpha - I[u < \alpha], \ 0 < u < 1$ in the Hodges and Lehmann (1963) manner, is asymptotically equivalent to the slope component $\widehat{\boldsymbol{\beta}}_n(\alpha)$ of the regression quantile. Denoting $R_{ni}(\mathbf{b})$ the rank of $Y_i - \mathbf{x}'_i \mathbf{b}$ among $(Y_1 - \mathbf{x}'_1 \mathbf{b}, \ldots, Y_n - \mathbf{x}'_n \mathbf{b})$, $\mathbf{b} \in \mathbb{R}^p, \ i = 1, \ldots, n$, the R-estimator of $\boldsymbol{\beta}$ can be defined as $\widetilde{\boldsymbol{\beta}}_{nR}(\alpha) = \mathrm{argmin}_{\mathbf{b} \in \mathbb{R}^p} \mathcal{D}_n(\mathbf{b})$, where $\mathcal{D}_n(\mathbf{b}) = \sum_{i=1}^n (Y_i - \mathbf{x}'_i \mathbf{b}) \varphi_\alpha \left(\frac{R_{ni}(\mathbf{b})}{n+1} \right)$ is Jaeckel's measure of *rank dispersion* (Jaeckel (1972)). Due to the invariance of ranks, $\widetilde{\boldsymbol{\beta}}_{nR}(\alpha)$ estimates only the slope parameters, while $\beta_0 + F^{-1}(\alpha)$ is estimated by the $[n\alpha]$-quantile of the corresponding residuals.

Then $\widetilde{\boldsymbol{\beta}}_{nR}(\alpha)$ admits the asymptotic representation

$$f(F^{-1}(\alpha)) \Big(\sum_{i=1}^{n} (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)' \Big)^{\frac{1}{2}} \left[\widetilde{\boldsymbol{\beta}}_{nR}(\alpha) - \boldsymbol{\beta} \right] \tag{3}$$

$$= \Big(\sum_{i=1}^{n} (\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)(\mathbf{x}_{ni} - \bar{\mathbf{x}}_n)' \Big)^{-\frac{1}{2}} \sum_{j=1}^{n} (\mathbf{x}_{nj} - \bar{\mathbf{x}}_n) \left[a_n(R_j(\mathbf{0}), \alpha) - (1 - \alpha) \right] + o_p(1) \quad \text{as} \quad n \to \infty,$$

where $\bar{\mathbf{x}}_n = n^{-1} \sum_{i=1}^n \mathbf{x}_{ni}$ 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. Hájek (1965) proved, for $p = 1$, the weak convergence of the process on the right-hand side of (3) to the Brownian Bridge, and defined various rank tests in regression model as its functionals. The linear rank test statistics can be obtained by integrating the same process with respect to a suitable score function $\varphi(\alpha), \ 0 < \alpha < 1$, or by integrating $\varphi(\alpha)$ with respect to the above process, whenever it is well-defined.

If, moreover, $\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)' = \mathbf{Q}_n \to \mathbf{Q}$ as $n \to \infty$ where $\mathbf{Q}$ is a $p \times p$ positively definite matrix, then we can rewrite (3) in the form

$$\begin{aligned} & n^{\frac{1}{2}} f(F^{-1}(\alpha)) \left[\widetilde{\boldsymbol{\beta}}_{nR}(\alpha) - \boldsymbol{\beta}\right] \\ & = \Big(\sum_{i=1}^{n}(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)'\Big)^{-1} \sum_{j=1}^{n}(\mathbf{x}_{nj}-\bar{\mathbf{x}}_n)\left[a_n(R_j(\mathbf{0}),\alpha)-(1-\alpha)\right] + o_p(1); \end{aligned}$$

notice that the first term on the right-hand side equals to the least squares estimator of Hájek's scores.

The relation (3) extends up to the extreme regression quantile with $\alpha = 1 - \frac{1}{n}$, provided f belongs to the domain of attraction of the Gumbel extreme distribution and $nf(F^{-1}(1-\frac{1}{n})) \to \infty$ as $n \to \infty$. Here, too, the extreme R-estimator consistently estimates $\boldsymbol{\beta}$ and the representation (3) takes on the form

$$\begin{aligned} & nf\left(F^{-1}(1-\tfrac{1}{n})\right) \left[\widetilde{\boldsymbol{\beta}}_{nR}(1-\tfrac{1}{n}) - \boldsymbol{\beta}\right] \\ & = n\Big(\sum_{i=1}^{n}(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)(\mathbf{x}_{ni}-\bar{\mathbf{x}}_n)'\Big)^{-1} \sum_{j=1}^{n}(\mathbf{x}_{nj}-\bar{\mathbf{x}}_n)\left[a_n(R_j(\mathbf{0}),1-\tfrac{1}{n})-(1-\tfrac{1}{n})\right] + o_p(1)\Big[=O_p(1)\Big]. \end{aligned}$$

Similar representations of $\widetilde{\boldsymbol{\beta}}_{nR}(1-\frac{1}{n})$ can be also written under f with heavier tails, but then $\widetilde{\boldsymbol{\beta}}_{nR}(1-\frac{1}{n})$ is not a consistent estimator of $\boldsymbol{\beta}$.

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

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

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

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