# Regression quantile and averaged regression quantile processes

**J. Jurečková**[1*]

[1] *Charles University in Prague, Faculty of Mathematics and Physics, Czech Republic; jurecko@karlin.mff.cuni.cz*

***Keywords.*** *Regression quantile; Averaged regression quantile; Asymptotic representation; Brownian bridge.*

Consider the linear regression model $\mathbf{Y}_n = \mathbf{X}_n\beta + \mathbf{U}_n$ with observations $\mathbf{Y}_n = (Y_1, \ldots, Y_n)^\top$, i.i.d. errors $\mathbf{U}_n = (U_1, \ldots, U_n)^\top$ with an unknown distribution function $F$, and unknown parameter $\beta = (\beta_0, \beta_1, \ldots, beta_p)^\top$. The $n \times (p+1)$ matrix $\mathbf{X} = \mathbf{X}_n$ is known and $x_{i0} = 1$ for $i = 1, \ldots, n$ (i.e., $\beta_0$ is an intercept). The $\alpha$-regression quantile $\widehat{\beta}_n(\alpha)$ is a solution of the minimization $\sum_{i=1}^n \rho_\alpha(Y_i - \mathbf{x}_i^\top \mathbf{b}) := \min$ with respect to $\mathbf{b} = (b_0, \ldots, b_p)^\top \in \mathbb{R}^{p+1}$, where $\mathbf{x}_i^\top$ is the $i$-th row of $\mathbf{X}_n$, $i = 1, \ldots, n$ and $\rho_\alpha(z) = |z|\{\alpha I[z > 0] + (1-\alpha) I[z < 0]\},\ z \in \mathbb{R}^1$. The population counterpart of $\widehat{\beta}_n(\alpha)$ is the vector $\widetilde{\beta}(\alpha) = (\beta_0 + F^{-1}(\alpha), \beta_1, \ldots, \beta_p)^\top$.
**?** used a linear programming algorithm for calculation of $\widehat{\beta}_n(\alpha)$. They also used the following dual algorithm as a computational device:

$$\begin{aligned} &\mathbf{Y}_n^\top \widehat{\mathbf{a}} := \max \\ \text{under the constraint }\ &\mathbf{X}_n^\top \widehat{\mathbf{a}} = (1-\alpha)\mathbf{X}_n^\top \mathbf{1}_n,\ \ \widehat{\mathbf{a}} \in [0,1]^n,\ 0 \le \alpha \le 1. \end{aligned} \tag{1}$$

The components of the optimal solution of (**??**), $\widehat{\mathbf{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha), \ldots, \hat{a}_{nn}(\alpha))^\top$, were named the *regression rank scores* by **?**, who used them for construction of the rank tests in the linear model.
Consider the regression quantile process $\mathbf{Z}_n = \left\{\mathbf{Z}_n(\alpha) = n^{1/2}(\widehat{\beta}_n(\alpha) - \widetilde{\beta}_n(\alpha))\right\}$ and the ordinary quantile process $Z_n^{(0)} = \left\{Z_n^{(0)}(\alpha) = n^{1/2}(F_n^{-1}(\alpha) - F^{-1}(\alpha))\right\}$, $0 < \alpha < 1$ where $F_n^{-1}$ is the empirical quantile function. Under conditions on $\mathbf{X}$ and $F$, $\mathbf{Z}_n \stackrel{\mathcal{D}}{\to} \left(f \circ F^{-1}\right)^{-1} \mathbf{Q}^{-1} \mathbf{W}_{(p)}^*$ as $n \to \infty$, where $\mathbf{W}_{(p)}^*$ is a vector of $p$ independent Brownian bridges on $(0,1)$ (see **?**).

The scalar statistic

$$\bar{B}_n(\alpha) = \bar{\mathbf{x}}_n^\top \widehat{\beta}_n(\alpha), \quad \bar{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_{ni}$$

is called the *averaged regression $\alpha$-quantile*. It is scale equivariant and regression equivariant. It was shown by **?** that, for every fixed $n$, $\bar{B}_n(\alpha)$ equals to a linear combination of $p+1$ components of vector of observations $\mathbf{Y}$, corresponding to the optimal base of the linear programming (**??**). Asymptotically is $\bar{B}_n(\alpha)$ equivalent to the $[n\alpha]$-quantile of the location model, as was proved in **?**. Under mild conditions, the process $\bar{\mathcal{B}}_n$ admits the asymptotic representation

$$\bar{\mathcal{B}}_n = \frac{1}{\sqrt{n} f(F^{-1}(\alpha))} \sum_{i=1}^{n} \Big( I[U_i > F^{-1}(\alpha)] - (1-\alpha) \Big) + o_p^*(1).$$

Moreover, $\bar{\mathcal{B}}_n \xrightarrow{\mathcal{D}} (f \circ F^{-1})^{-1} W^*$ as $n \to \infty$ where $W^*$ is the Brownian bridge on (0,1); here $o_p^*(1)$ means uniform convergence on $(\varepsilon, 1-\varepsilon)$ for any $\varepsilon \in (0, 1/2)$.
The weak convergence of process $\bar{\mathcal{B}}_n$ applies also to its various functionals, what in turn leads to various useful applications under nuisance regression. Moreover, its representation coincides with the representation of the ordinary quantile process, hence $\bar{\mathcal{B}}_n$ is asymptotically equivalent to the same. This, by **?**, is in turn approximated by a sequence of Brownian bridges.

# References

Csörgő, M. & Révész, P. (1978). Strong approximation of the quantile process. *Ann. Statist.* **6**, 882–894.

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

Jurečková J. (2016). Finite sample behavior of averaged extreme regression quantile. *Extremes* **19**, 41–49.

Jurečková, J. & Picek, J. (2014). Averaged regression quantiles. In: *Contemporary Developments in Statistical Theory* (S. N. Lahiri et al. (eds.), Springer Proceedings in Mathematics & Statistics 68, Chapter 12, pp.203–216.

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