AVERAGED REGRESSION QUANTILES

Jana Jurečková and Jan Picek

Charles University in Prague
Technical University in Liberec

Motivation

Regression quantiles are an important tool of robust inference in linear regression and other models, because they are a straightforward extension of location quantiles. Quantile regression is intensively applied in econometric problems. Though regression quantiles are asymptotically normal, their finite sample behavior is far from normality. But it is also the case for ordinary quantiles. Hence, the finite sample behavior of regression quantiles is of interest. A good approximation of the finite-sample distribution is the saddle-point approximation (small-sample asymptotics of Field and Ronchetti and Hampel). Radka Sabolová studied the saddle-point technique during her stay in University of Geneva under supervising of E. Ronchetti and calculated the saddle-point approximations of densities of regression quantiles. They are very precise.

During our discussions, we have found the averaged regression quantile $\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^\top\widehat{\boldsymbol{\beta}}_n(\alpha)$ as a useful characteristic. Interesting is the following identity, for which I still search the right interpretation: If $\alpha\in(0,1)$ is a continuity point of $\widehat{\boldsymbol{\beta}}_n(\alpha)$, then

$$\bar{\mathbf{x}}_n^\top\widehat{\boldsymbol{\beta}}_n(\alpha) = -\frac{1}{n}\sum_{i=1}^{n} Y_i \ \frac{d}{d\alpha}\hat{a}_i(\alpha) \tag{1}$$

where $\hat{a}_{n1}(\alpha),\ldots,\hat{a}_{nn}(\alpha)$ are the so called regression rank scores of the model (dual to regression quantile, explained later). In the location model, (1) reduces to the identity $\widehat{\beta}_n(\alpha) = Y_{n:[n\alpha]}$. It leads to the conjecture that the averaged regression quantile is asymptotically equivalent to the ordinary regression quantile of the model errors.

Introduction

Consider the linear regression model $\mathbf{Y}_n = \mathbf{X}_n\boldsymbol{\beta} + \mathbf{U}_n$ (2)

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 $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^\top$. The $n \times (p+1)$ matrix $\mathbf{X} = \mathbf{X}_n$ is known and $x_{i0} = 1$ for $i = 1, \ldots, n$ (i.e., β_0 is an intercept). The α-regression quantile $\widehat{\boldsymbol{\beta}}_n(\alpha)$ of model (2) 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

$$\rho_\alpha(z) = |z|\{\alpha I[z > 0] + (1-\alpha) I[z < 0]\}, \; z \in \mathbb{R}^1.$$

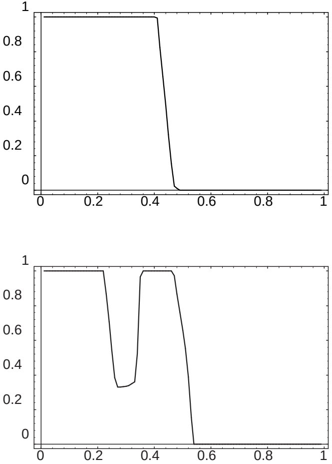
Regression rank scores

Koenker and Bassett (1978) used the following algorithm, dual to regression quantile algorithm, as a computational device:

$$\mathbf{Y}_n^\top \widehat{\mathbf{a}} := \max$$

$$\text{under the constraints } \begin{cases} (\mathbf{X}_n)^\top \widehat{\mathbf{a}} = (1-\alpha)(\mathbf{X}_n)^\top \mathbf{1}_n, \\ \widehat{\mathbf{a}} \in [0,1]^n, \ 0 \leq \alpha \leq 1. \end{cases}$$

The components of the optimal solution are called regression rank scores (RRS): $\widehat{\mathbf{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha), \ldots, \hat{a}_{nn}(\alpha))^\top, \ 0 \leq \alpha \leq 1$. They are convenient for construction of the rank tests in the linear model (see Gutenbrunner and JJur. (1992), Gutenbrunner et al. (1993)). The RRS are invariant with respect to the shift in location and scale and to changes of $\boldsymbol{\beta}$.



Typical shape of the regression rank scores

From the linear programming theory it follows that

$$\hat{a}_{ni}(\alpha) = \begin{cases} 1 & \dots \quad Y_i > \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_n(\alpha), \\ 0 & \dots \quad Y_i < \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_n(\alpha), \quad i = 1, \dots, n \end{cases}$$

and that $0 < \hat{a}_{ni}(\alpha) < 1$ if $Y_i = \mathbf{x}_i^\top \widehat{\boldsymbol{\beta}}_n(\alpha)$ (the exact fit); there are just $p + 1$ components with exact fit for each α, which correspond to the optimal base among $\mathbf{x}_1, \dots, \mathbf{x}_n$. The values of $\hat{a}_{ni}(\alpha)$ are determined by the constraints in the linear program.

The averaged regression quantile is the scalar statistic

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

$\bar{B}_n(\alpha)$ is scale equivariant and regression equivariant. Moreover, we have the following useful facts:

Lemma 1

(i) *If* $\alpha \in (0,1)$ *is a continuity point of* $\widehat{\boldsymbol{\beta}}_n(\alpha)$, *then*

$$\bar{B}_n(\alpha) = -\frac{1}{n}\sum_{i=1}^n Y_i \, \frac{d}{d\alpha}\hat{a}_i(\alpha).$$

(ii) $\bar{B}_n(\alpha)$ *and hence also* $-\frac{1}{n}\sum_{i=1}^n Y_i \, \frac{d}{d\alpha}\hat{a}_i(\alpha)$ *are nondecreasing step-functions of* $\alpha \in (0,1)$.

Moreover, the averaged regression α-quantile is asymptotically equivalent to the ordinary location α-quantile.

Assume the following conditions on matrix $\mathbf{X}_n$:

A1 $\lim_{n\to\infty} \mathbf{Q}_n = \mathbf{Q}$, where $\mathbf{Q}_n = n^{-1}\mathbf{X}_n^\top \mathbf{X}_n$ and $\mathbf{Q}$ is a positive definite matrix.

A2 $n^{-1}\sum_{i=1}^n x_{ij}^4 = \mathcal{O}(1)$, as $n \to \infty$, for $j = 1, \ldots, p$.

Theorem 1

Suppose that the distribution function F is continuous and twice differentiable in a neighborhood of $F^{-1}(\alpha)$ and that $F'(F^{-1}(\alpha)) = f(F^{-1}(\alpha)) > 0,\ 0 < \alpha < 1$. Then, under the conditions **A1** *-* **A2**,

$$n^{1/2}\left[\bar{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta}) - U_{n:[n\alpha]}\right] = \mathcal{O}_p(n^{-1/4}) \tag{3}$$

as $n \to \infty$, where $U_{n:1} \leq \ldots \leq U_{n:n}$ are the order statistics corresponding to $U_1, \ldots, U_n$.

Theorem 1 has an easy corollary, leading to the first application:

Corollary

Under the conditions of the Theorem 1,
$$n^{1/2}\left[\bar{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_n(\alpha_2)-\widehat{\boldsymbol{\beta}}_n(\alpha_1))-(U_{n:[n\alpha_2]}-U_{n:[n\alpha_1]})\right]=\mathcal{O}_p(n^{-1/4})$$
for any $0<\alpha_1\leq\alpha_2<1$.

The statistics of type $\bar{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_n(\alpha_2)-\widehat{\boldsymbol{\beta}}_n(\alpha_1))$ are invariant to the regression with design $\mathbf{X}_n$, and equivariant with respect to the scale. As such, they provide a tool for studentization of M-estimators in linear regression model, and where one needs to make a statistic scale-equivariant. For instance, we used the regression interquartile range with $\alpha_1=\frac{1}{4}$, $\alpha_2=\frac{3}{4}$ in goodness-of-fit testing with nuisance regression and scale.

The proposition of Theorem 1 remains true under a sequence of distributions, contiguous with respect to the sequence $\{\prod_{i=1}^{n} F(u_{ni})\}$. Among them, let us observe the local heteroscedasticity. The frequent heteroscedastic model has the form $Y_i = \beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta} + \sigma_i U_i, \quad i = 1, \ldots, n$ where $\mathbf{U}_n = (U_1, \ldots, U_n)^\top$ are the i.i.d. errors with the joint distribution function F and

$$\sigma_i = \exp\{\mathbf{d}_i^\top \gamma\}, \quad i = 1, \ldots, n$$

with known or observable $\mathbf{d}_i \in \mathbb{R}^q$, $1 \leq i \leq n$ and unknown parameter $\boldsymbol{\gamma} \in \mathbb{R}^q$.

We assume that

$$\sum_{i=1}^{n} d_{ij} = 0, \quad j = 1, \ldots, q, \ \max_{1 \le i \le n} \|\mathbf{d}_i\| = o(n^{1/2}) \text{ as } \quad n \to \infty,$$

$$\lim_{n\to\infty} \mathbf{D}_n = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{d}_i \mathbf{d}_i^\top = \mathbf{D},$$

$$\max_{1 \le i \le n} \left\{ \mathbf{d}_i^\top \Big(\sum_{k=1}^{n} \mathbf{d}_k \mathbf{d}_k^\top \Big)^{-1} \mathbf{d}_i \right\} \to 0 \ \text{ as } \ n \to \infty$$

where $\mathbf{D}$ is positive definite $(q \times q)$ matrix. Homoscedasticity means that $\boldsymbol{\gamma} = \mathbf{0}$; then Theorem 1 applies. The local heteroscedasticity means that
$\boldsymbol{\gamma} = \boldsymbol{\gamma}_n = n^{-\frac{1}{2}} \boldsymbol{\delta}, \quad \boldsymbol{\delta} \in \mathbb{R}^q, \ \boldsymbol{\delta} \neq \mathbf{0}, \ \|\boldsymbol{\delta}\| \le C < \infty.$

Theorem 2

(i) Under the local heteroscedasticity, Theorem 1 remains true for any fixed $\alpha \in (0,1)$.
(ii) The sequences $\{\sqrt{n}\bar{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta} - \mathbf{e}_0 F^{-1}(\alpha))\}$ *and* $\{\sqrt{n}(U_{n:[n\alpha]} - F^{-1}(\alpha))\}$ *are asymptotically normally distributed* $\mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(F^{-1}(\alpha))}\right)$, *both under homoscedasticity and under local heteroscedasticity;* $\mathbf{e}_0 = (1, 0, \ldots, 0)^\top \in \mathbb{R}_{p+1}$.

Remark

By Theorem 2, the averaged regression quantile does not register (asymptotically) the local heteroscedasticity; as such, it can be considered asymptotically ancillary with respect to the local heteroscedasticity.

Corollary

The representation

$$\sqrt{n}\bar{\mathbf{x}}_n^\top(\widehat{\boldsymbol{\beta}}_n(\alpha) - \boldsymbol{\beta} - \mathbf{e}_0 F^{-1}(\alpha)) \tag{4}$$
$$= \frac{1}{\sqrt{n} f(F^{-1}(\alpha))} \sum_{i=1}^{n} \Big(\alpha - I[U_i < F^{-1}(\alpha)]\Big) + \mathcal{O}_p(n^{-1/4})$$

holds for any fixed $0 < \alpha < 1$, *both under homoscedasticity and local heteroscedasticity. Under the homoscedasticity, it is uniform for* $0 < \alpha_0 \leq \alpha \leq 1 - \alpha_0 < 1$, *where* $\alpha_0 \in (0, \frac{1}{2})$ *is any fixed.*

From Csörgő and Révész (1978) and Theorem 1 we obtain the approximation of $B_n(\alpha)$ by Brownian bridge:

Theorem 3

Assume the conditions of Theorem 1 and that for some $\gamma > 0$

$$\sup_x \left\{ F(x)(1-F(x)) \left| \frac{f'(x)}{f^2(x)} \right| \right\} \leq \gamma.$$

Then for each n *one can define a Brownian bridge* $\{B_n(\alpha) : 0 \leq \alpha \leq 1\}$ *so that*

$$\sup_{\alpha_0 \leq \alpha \leq 1-\alpha_0} \left| \sqrt{n}\bar{\mathbf{x}}_n^\top (\widehat{\beta}_n(\alpha) - \beta - \mathbf{e}_0 F^{-1}(\alpha)) - B_n(\alpha) \right| = O_p(n^{-1/4})$$

for every fixed $\alpha_0 \in (0, \frac{1}{2})$, *as* $n \to \infty$.

Estimation of quantile density

The quantile density function $q(u) = \frac{1}{f(F^{-1}(u))}$ is often used in nonparametric statistical inference, as in the studentization, adaptive procedures, in the sequential confidence sets, in tests on β based on L_1-regression, and elsewhere. It is a scale statistic, being location invariant and scale equivariant. The sum of quantile densities is again a quantile density of some random variable. In the location model, several authors considered the histogram estimate of $q(\alpha)$; e.g. Siddiqui (1960), Bloch and Gastwirth (1968), Bofinger (1975), Lai et al. (1983), and others. The kernel-type estimators of $q(\alpha)$ was studied by Parzen (1979), Yang (1985), Falk (1986), Zelterman (1990), Soni et al. (2012) and others. Xiang (1995) studied the kernel estimator of the conditional quantile density function.

The first estimates of $q(\alpha)$ in the linear regression model were proposed by Koenker and Bassett (1982) and Welsh (1987a). Welsh (1987b) used a kernel smoothing of the empirical quantile function of residuals from some initial estimator of $\boldsymbol{\beta}$. Dodge and JJur. (1995) extended Falk's (1986) estimator to the linear model, assuming that $\bar{x}_j = \sum_{i=1}^{n} x_{ij} = 0$ for $j = 1, \ldots, p$. In the linear model, analogous estimators can be based on $\bar{B}(\alpha)$; they can be used e.g. in the autoregression and sequential models. The histogram type estimate is

$$H_n(\alpha) = \tfrac{1}{2\nu_n}[\bar{B}_n(\alpha + \nu_n) - \bar{B}_n(\alpha - \nu_n)]$$

where $\nu_n = o(n^{-1/3})$ and $n\nu_n \to \infty$ as $n \to \infty$. $H_n(\alpha)$ is consistent and asymptotically normal:

Theorem 4

Under the above assumptions,

$$H_n(\alpha) - q(\alpha) = \mathcal{O}_p(n\nu_n)^{-1/2} \quad \textit{as} \;\; n \to \infty,$$

uniformly in $\alpha \in (\varepsilon, 1-\varepsilon)$, $\forall\ \varepsilon \in (0, 1/2)$. *Moreover,* $H_n(\alpha)$ *is asymptotically normal for every fixed* $\alpha \in (\varepsilon, 1-\varepsilon)$,

$$(n\nu_n)^{1/2}(H_n(\alpha) - q(\alpha)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \tfrac{1}{2}q^2(\alpha)\right) \quad \textit{as} \;\; n \to \infty.$$

The kernel estimate of $q(\alpha)$ is defined as follows:

$$\widehat{\boldsymbol{\kappa}}_n(\alpha) = \frac{1}{\nu_n^2} \int_0^1 \bar{B}_n(u) k\Big(\frac{\alpha - u}{\nu_n}\Big) du,$$

assuming that $\nu_n \downarrow 0$, $n\nu_n^3 \downarrow 0$ and $n\nu_n^2 \to \infty$ as $n \to \infty$. The kernel function $k : \mathbb{R}^1 \mapsto \mathbb{R}^1$ is assumed to satisfy the following condition:

K1: $k(\cdot)$ is continuous on its compact support and

$$\int k(x)dx = 0, \qquad \int xk(x)dx = -1.$$

The estimator $\widehat{\boldsymbol{\kappa}}_n(\alpha)$ is consistent and asymptotically normal.

Theorem 5

Let F (distribution function of U_1) have continuous density f which is positive and finite in $\{x: \ 0 < F(x) < 1\}$. Let F^{-1} be twice differentiable with bounded second derivative in a neighborhood of α. Then, under the conditions of Theorem 2,

$$\widehat{\boldsymbol{\kappa}}_n(\alpha) - q(\alpha) = \mathcal{O}_p((n\nu_n))^{-1/2} \quad \text{as} \ \ n \to \infty.$$

Moreover, $\widehat{\boldsymbol{\kappa}}_n(\alpha)$ is asymptotically normally distributed,

$$(n\nu_n))^{1/2}(\widehat{\boldsymbol{\kappa}}_n(\alpha) - q(\alpha)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, q^2(\alpha) \int K^2(x)dx\right),$$

where $K(x) = \int_{-\infty}^{x} k(y)dy$.

Remark

Example of kernel satisfying **K1** *is the Epanechnikov (1969) kernel with*

$$k(x) = \begin{cases} -\frac{3}{2b^3} \cdot x & \text{if} \quad -b \leq x \leq b \\ 0 & \quad \text{elsewhere.} \end{cases}$$

The kernel estimate dominates the histogram for $b > \frac{6}{5}$, *when* $\int K^2(x)dx = \frac{3}{5b} < \frac{1}{2}$.

Numerical illustration

Consider the linear regression model

$$Y_i = \beta_0 + x_i\beta_1 + e_i, \quad i = 1, \ldots, n.$$

Errors e_i, $i = 1, \ldots, n$, were simulated from the normal, exponential and Cauchy distributions and $x_{1,1}, \ldots, x_{1,n}$ were generated from the uniform distribution on the interval (-5,50) for $n = 20$, 100, 500. The choice of $\boldsymbol{\beta}$ is $\beta_0 = 1$ and $\beta_1 = -2$. We compare the averaged regression α-quantiles with the location α-quantiles for $\alpha = 0.55,\ 0.95$ and 10 000 replications of models were simulated for each combination of the parameters and each α. Figures 1–3 and Tables 1–2 compare some characteristics of differences between the averaged regression α-quantile and the location α-quantile.

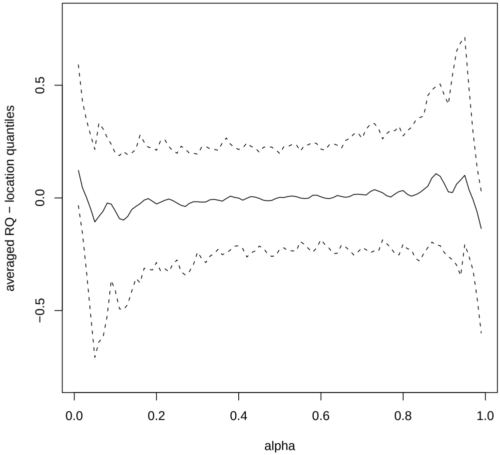


Figure 1:The median, 5%-, 95%-quantiles in the sample of 10 000 differences between averaged regression and location α-quantiles; normal distributions of errors; sample sizes n =20.

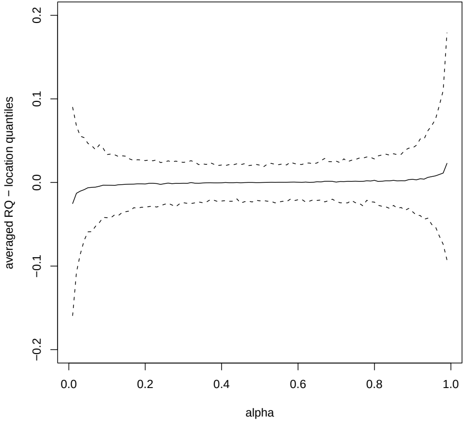


Figure 2: The median, 5%-, 95%- quantiles in the sample of 10 000 differences between averaged regression and location α-quantiles; normal distributions of errors; sample sizes n =500.

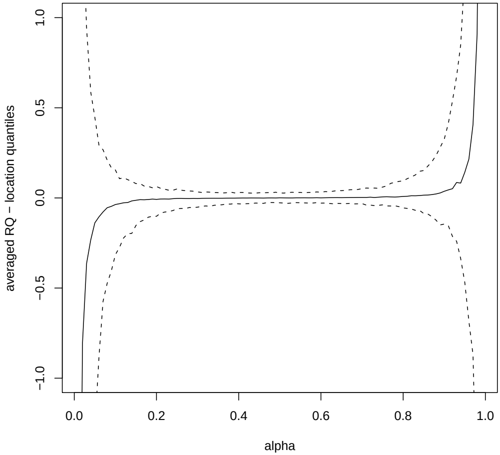


Figure 3: The median, 5%-, 95%- quantiles in the sample of 10 000 differences between averaged regression and location α-quantiles; Cauchy distributions of errors; sample sizes n =500.

n law	mean	stand. dev.	quantiles						
			0	0.05	0.25	0.5	0.75	0.95	1
20N	-0.005	0.123	-0.562	-0.220	-0.064	-0.003	0.048	0.208	0.589
20E	0.023	0.101	-0.283	-0.113	-0.028	0.004	0.060	0.211	0.681
20C	-0.029	0.216	-1.785	-0.368	-0.108	-0.009	0.065	0.287	0.825
100N	-0.002	0.042	-0.181	-0.069	-0.022	-0.001	0.018	0.065	0.170
100E	0.005	0.033	-0.135	-0.042	-0.012	0.000	0.019	0.064	0.173
100C	-0.003	0.056	-0.331	-0.099	-0.029	-0.001	0.024	0.091	0.205
500N	-0.001	0.013	-0.050	-0.025	-0.008	-0.001	0.006	0.021	0.054
500E	0.001	0.010	-0.044	-0.015	-0.004	0.000	0.006	0.018	0.038
500C	-0.001	0.017	-0.080	-0.032	-0.009	0.000	0.009	0.027	0.073

Table 1: Mean, standard deviation and quantiles of difference between averaged regression and location 0.55-quantiles; normal (N), exponential (E) and Cauchy (C) distributions of errors; sample sizes n =20, 100, 500, and 10 000 replications.

n law	mean	stand. dev.	quantiles						
			0	0.05	0.25	0.5	0.75	0.95	1
20N	0.065	0.179	-0.592	-0.215	-0.014	0.047	0.149	0.387	1.065
20E	0.123	0.269	-0.871	-0.214	-0.008	0.073	0.230	0.669	1.453
20C	0.880	2.314	-2.577	-0.378	0.024	0.260	0.875	3.899	32.680
100N	0.012	0.061	-0.281	-0.089	-0.018	0.008	0.042	0.118	0.281
100E	0.032	0.098	-0.275	-0.113	-0.019	0.021	0.078	0.196	0.579
100C	0.139	0.325	-0.629	-0.217	-0.017	0.067	0.228	0.692	3.420
500N	0.002	0.019	-0.060	-0.030	-0.008	0.003	0.013	0.034	0.085
500E	0.007	0.031	-0.123	-0.042	-0.009	0.005	0.021	0.059	0.159
500C	0.023	0.074	-0.206	-0.085	-0.015	0.013	0.057	0.152	0.401

Table 2: Mean, standard deviation and quantiles of difference between averaged regression and location 0.95-quantiles; normal (N), exponential (E) and Cauchy (C) distributions of errors; sample sizes $n = 20,\ 100,\ 500$ and 10 000 replications.

References

- Bloch, D.A. and J.L. Gastwirth (1968). On a simple estimate of the reciprocal of the density function. *Ann. Math. Statist.* 36, 457–462.
- Bofinger, E. (1975). Estimation of a density function using the order statistics, Austral. J. Statist. 17, 1–7.
- Csörgö, M. and Révész, P. (1978). Strong approximation of the quantile process. *Ann. Statist.* 6, 882–894.
- Dodge, Y. and Jurečková, J. (1995). Estimation of quantile density function based on regression quantiles. *Statistics & Probability Letters* 23, 73–78.
- Epanechnikov, V. A. (1969). Nonparametric estimation of a multivariate probability density, *Theor. Probab. Appl.* 14, 153–158.

Falk, M. (1986). On the estimation of the quantile density function. *Statist. Probab. Lett.* 4, 69–73.

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

C.Gutenbrunner, J. Jurečková, R.Koenker and S.Portnoy (1993). Tests of linear hypotheses based on regression rank scores. *Nonpar. Statist.* 2, 307-331.:

Jones, M.C. (1992). Estimating densities, quantiles, quantile densities and density quantiles. *Annals of the Institute of Statistical Mathematics* 44, 721-727.

Jurečková, J., Sen, P. K. and Picek, J. (2013). *Methodological Tools in Robust and Nonparametric Statistics*. Chapman & Hall/CRC.

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

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

Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* Lecture Notes in Statistics 166, Springer.

Koul, H. L. and Saleh, A.K.Md.E. (1995). Autoregression quantiles and related rank-scores processes. Ann. Statits. 23, 670–689.

Lai, T.L., Robbins, H. and Yu, K.F. (1983). Adaptive choice of mean or median in estimating the center of a symmetric distribution, *Proc. Nat. Acad. Sci. U.S.A.* 80, 5803–5806.

Parzen, E. (1979). Nonparametric statistical data modelling. *Journal of American Statistical Association* 74, 105-122.

Parzen, E. (2004). Quantile probability and statistical data modeling. *Statistical Science* 19, 652-662.

Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population, *J. Res. Nat. Bur. Standards* 6411, 145–150.

- Soni, P., Dewan, I. and Jain, K. (2012). Nonparametric estimation of quantile density function. CSDA 56, 3876–3886.
- Welsh, A. H. (1987) Kernel estimates of the sparsity function, in: Y. Dodge, ed. *Statist. Data Analysis Based on the L1-Norm and Rel. Methods* (Elsevier, Amsterdam) pp. 369–378.
- Xiang, X. (1995). Estimation of conditional quantile density function. *Journal of Nonparametric Statistics* 4, 309-316.
- Yang, S.S. (1985) A smooth nonparametric estimator of quantile function *J. Amer. Statist. Assoc.* 80, 1004–1011.
- Zeltermann, D. (1990). Smooth nonparametric estimation of the quantile function. *J. Statist. Plann. Inference* 26, 339–352.