Boletín de Estadística e Investigación Operativa
Vol. 34, No. 2, Julio 2018, pp. 97-116

Estadística

Quantile regression: estimation and lack-of-fit tests

Mercedes Conde-Amboage, Wenceslao González-Manteiga and César Sánchez-Sellero

Department of Statistics, Mat. Analysis and Optimization
Universidade de Santiago de Compostela

✉ mercedes.amboage@usc.es, ✉ wenceslao.gonzalez@usc.es,
✉ cesar.sanchez@usc.es

Abstract

Although mean regression achieved its greatest diffusion in the twentieth century, it is very surprising to observe that the ideas of quantile regression appeared earlier. While the beginning of the least-squares regression can be dated in the year 1805 by the work of Legendre, in the mid-eighteenth century Boscovich already adjusted data on the ellipticity of the Earth using concepts of quantile regression.

Quantile regression is employed when the aim of the study is centred on the estimation of the different positions (quantiles). This kind of regression allows a more detailed description of the behaviour of the response variable, adapts to situations under more general conditions of the error distribution and enjoys robustness properties. For all that, quantile regression is a very useful statistical technology for a large diversity of disciplines. In this paper a review on quantile regression methods will be presented.

Keywords: Quantile regression, Estimation, Lack-of-fit tests, Robustness, Sparsity.
AMS Subject classifications: 62J05, 62G08, 62F35, 62F03.

1. Introduction

Given a random variable X, for each $0 < \tau < 1$ its **τ-th quantile**, that will be denoted by c_τ, is defined as the value that verifies

$$\mathbb{P}_X(X \leq c_\tau) \geq \tau \qquad \text{and} \qquad \mathbb{P}_X(X \geq c_\tau) \geq 1 - \tau.$$

Then, the **quantile function** of a probability distribution is given by the inverse of the cumulative distribution function. More formally, the quantile function is defined as follows

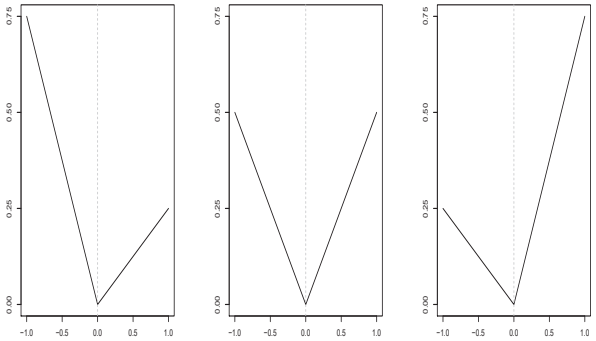
$$F_X^{-1}(\tau) = \inf\{x \in \mathbb{R} : \tau \leq F_X(x)\},$$

where $\inf\{A\}$ represents the infimum of a subset A. The infimum is a criterion used to choose a simple quantile when the definition in terms of the probability function provides more than one solution.

Quantiles can be computed as the result of an optimization problem. First, let us call **quantile loss function** to the following piecewise linear function:

$$\rho_\tau(u) = u\big(\tau - \mathbb{I}(u < 0)\big) = \begin{cases} u\,\tau & \text{if } u \geq 0, \\ u\,(\tau - 1) & \text{if } u < 0, \end{cases}$$

where $\mathbb{I}$ represents the indicator function of an event. Figure 1 shows the representation of the quantile loss function for different values of the τ-th quantile of interest. Note that the quantile loss function is not differentiable so that standard numerical algorithms cannot be directly applied. Because of this reason, most of the theory developed for mean estimation can not be applied in this context.



(a) Quantile $\tau = 0.25$ (b) Quantile $\tau = 0.50$ (c) Quantile $\tau = 0.75$

Figure 1: Representation of the quantile loss function for three different values of the τ-th quantile of interest: $\tau = 0.25$ (Part a), $\tau = 0.50$ (Part b) and $\tau = 0.75$ (Part c).

Thereupon, for each $\tau \in (0,1)$, the τ-th quantile that has been denoted by c_τ can be written as

$$c_\tau = \arg\min_x \mathbb{E}\Big[\rho_\tau(X-x)\Big].$$

In practice, the cumulative distribution function F is replaced by the empirical distribution function. So, given $\{X_1,\ldots,X_n\}$ a random sample of the variable X, the **sample quantiles** can be computed as

$$\widehat{c}_\tau = \arg\min_c \int \rho_\tau(x-c)\, dF_n(x) = \arg\min_c \frac{1}{n}\sum_{i=1}^{n}\rho_\tau(X_i-c) \tag{1.1}$$

for each $\tau \in (0,1)$.

The problem of finding the τ-th sample quantile may be reformulated as a linear problem. A more complete explanation about this optimization problem can be found in Section 1.1.2 of [12]. In practice, there exists several methods in order to compute sample quantiles, and a clear review about these possibilities in R language is detailed in [24].

The asymptotic distribution of $\widehat{c}_\tau$ can be derived as a consequence of Lindeberg's central limit theorem. This result is gathered in Theorem 1.1 and its proof is detailed in several classical works on Statistical Inference, see for instance [6].

Theorem 1.1. *Given a random variable X with associated cumulative distribution function F_X that is absolutely continuous in a neighbourhood of the τ-th quantile of interest, c_τ, with $f_X(c_\tau) > 0$. Then, the asymptotic distribution of the sample quantile, $\widehat{c}_\tau$, is given by*

$$\sqrt{n}\,(\widehat{c}_\tau - c_\tau) \xrightarrow{d} N(0,\omega^2),$$

where $\omega^2 = \tau(1-\tau)/f_X^2(c_\tau)$, $N(0,\omega^2)$ represents the Gaussian distribution with zero mean and variance ω^2, and $\xrightarrow{d}$ denotes convergence in distribution.

According to the asymptotic distribution of $\widehat{c}_\tau$, the inverse of the density evaluated in the quantile, that is known in this context as sparsity, will play a crucial role in this context. A complete description of the sparsity function will be presented in Section 4.

It is well-known the major robustness of quantile methods versus classical least squares estimation. To show that, we are going to focus on the influence function, introduced by [20]. The **influence function** describes the effect of an anomalous sample point over a certain estimator. More formally, the influence function can be defined by

$$IF(y,\widehat{\gamma},F) = \lim_{t\to 0}\frac{\widehat{\gamma}(F_t)-\widehat{\gamma}(F)}{t},$$

where $\widehat{\gamma}(F)$ represents an estimator that depends on a distribution F and $F_t = (1-t)F + t\delta_y$ where δ_y denotes the distribution function that assigns mass 1 to the contaminated point y.

So, the influence function associated with mean estimator (denoted by $\widehat{\mu}$) will be given by

$$IF(y, \widehat{\mu}, F) = \lim_{t \to 0} \frac{\widehat{\mu}(F_t) - \widehat{\mu}(F)}{t} = y - \widehat{\mu}(F),$$

while the influence function of median estimator (denoted by $\widehat{c}_{0.5}$) will be given by

$$IF(y, \widehat{c}_{0.5}, F) = \lim_{t \to 0} \frac{\widehat{c}_{0.5}(F_t) - \widehat{c}_{0.5}(F)}{t} = \frac{0.5 \ \mathrm{sgn}(y - \widehat{c}_{0.5}(F))}{f(\widehat{c}_{0.5}(F))},$$

where sgn represents the sign function.

There is a fundamental difference between the two influence functions. While the influence function of the mean, is simply proportional to y, the influence of contamination at y on the median is bounded by the sparsity at the median. Figure 2 shows the comparison of the influence functions of mean and median estimators associated with a standard Gaussian distribution F. Let us observe the fragility of the mean and the robustness of the median in withstanding the contamination of outlying observations. Much of what has already been said extends immediately to the quantiles generally for any τ, and from them to quantile regression.

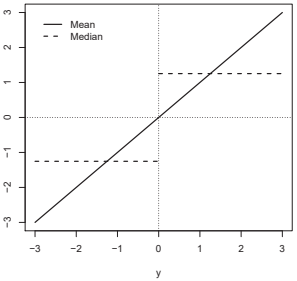


Figure 2: Influence function associated with mean and median estimators, where F is a standard Gaussian distribution.

Taking into account the good properties of sample quantiles, we are going to extend these ideas to a regression context with a parametric (see Section 2) and a nonparametric (see Section 3) perspective. In Section 4 we have estab-

lished different sparsity estimators because of its fundamental role in quantile regression context. In Section 5 an introduction to lack-of-fit tests for quantile regression is presented. Finally, in Section 6 some conclusions are presented.

2. Parametric quantile regression

Now, our main goal will be to extend the theory developed in the previous section to the regression context. Then, for simplicity, let us consider the following linear regression model:

$$Y_i = \theta'_\tau P_i + \varepsilon_i, \tag{2.1}$$

where $P_i = (1, X_i)$ and $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ represents a random sample of the response variable (denoted by $Y \in \mathbb{R}$) and the explanatory variable (denoted by $X \in \mathbb{R}^d$). Moreover, the errors ε_i should verify that $\mathbb{P}(\varepsilon_i \leq 0 \mid X = X_i) = \tau$, that is, its conditional τ-th quantile is zero. Note that it is analogous to assume that $\mathbb{E}(\varepsilon_i | X = X_i) = 0$ in the classical least squares context.

If the conditional quantile function is defined by $q_\tau(x) = \theta'_\tau(1, x)$, in view of (1.1), it is reasonable to consider the estimator $\widehat{\theta}_\tau$ obtained as the solution of the following optimization problem:

$$\widehat{\theta}_\tau = \arg \min_{\theta \in \mathbb{R}^{d+1}} \sum_{i=1}^{n} \rho_\tau(Y_i - \theta' P_i). \tag{2.2}$$

This idea has been introduced by [26] and subsequently [14] demonstrated the consistency of the quantile regression estimator.

Following the ideas described in Section 1, the parameter $\widehat{\theta}_\tau$ can be obtained as the solution of the following linear optimization problem:

$$\min_{(\theta, u, v) \in \mathbb{R}^{d+1} \times \mathbb{R}^{2n}_+} \left\{ \tau 1'_n u + (1 - \tau) 1'_n v \ : \mathbb{X}\theta + u - v = Y \right\}, \tag{2.3}$$

where $\mathbb{X}$ denotes the regression design matrix that is a $n \times (d+1)$ matrix whose $j-$th row is given by $(1, X_j)'$ and 1_n represents a n-dimensional vector of ones. The residual vector $Y - \mathbb{X}\theta$ has been split into its positive and negative parts (u and v respectively).

The calculus of the quantile regression parameter as a linear optimization problem is crucial because it gives place to different methods in order to compute $\widehat{\theta}_\tau$. In this line, [3] proposed a modified version of the Simplex method in order to solve the optimization problem associated with $\tau = 0.5$ in which case the quantile loss function is the absolute value. It is important to emphasize that [3]'s proposal manages to reduce substantially the computational time needed to compute the estimator $\widehat{\theta}_\tau$ for $\tau = 0.5$ compared with the original Simplex

algorithm. Later, [27] extended this development to each quantile $0 < \tau < 1$.

Since quantile regression estimators do not have explicit expression, it would be necessary to resort to asymptotic expressions such as Bahadur's representation. If we assume that $\psi_\tau(r) = \tau\mathbb{I}(r > 0) + (\tau - 1)\mathbb{I}(r < 0)$, [2] established that

$$\sqrt{n}\ \left(\widehat{\theta}_\tau - \theta_\tau\right) = D_1^{-1}\ n^{-1/2}\ \sum_{i=1}^{n} P_i\ \psi_\tau(Y_i - \theta'_\tau P_i) + O_p\left(n^{-1/4}\sqrt{\log n}\right),$$

under certain regularity conditions.

Differently from least squares estimator, the quantile estimator distribution is not generally known even under error normality. [25] showed the following result about the asymptotic distribution of quantile regression estimators.

Theorem 2.1. *Let us consider a linear model as given in* (2.1). *Under the following conditions:*

Condition A1. *The conditional distribution functions F_i (Y_i conditioned to X_i) are absolutely continuous with continuous density functions f_i uniformly bounded away from 0 and ∞ at the conditional quantiles $c_i(\tau)$.*

Condition A2. *There exist positive definite matrices D_0 and $D_1(\tau)$ such that*

1. $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} P_i\ P_i' = D_0$,
2. $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_i(c_i(\tau)) P_i\ P_i' = D_1(\tau)$,
3. $\max_{i=1,\ldots,n} \|X_i\|/\sqrt{n} \to 0$,

it follows that

$$\sqrt{n}\left(\widehat{\theta}_\tau - \theta_\tau\right) \xrightarrow{d} N\Big(0, \tau(1-\tau)D_1(\tau)^{-1}D_0D_1(\tau)^{-1}\Big).$$

Again, in view of Theorem 2.1, it is clear that the sparsity function will play an important role. Furthermore, in a regression context, the quantile methods still enjoys properties of robustness. [11] (page 106) showed that the influence function associated with the least squares estimator (denoted by $\widehat{\theta}_{LS}$) is given by

$$IF((x,y), \widehat{\theta}_{LS}, F) = \mathbb{E}(\mathbb{X}\mathbb{X}')^{-1}(1,x)(y - \widehat{\theta}_{LS}(F)'(1,x)),$$

where F represents the distribution function of the random vector (X, Y) and the pair (x, y) denotes a new observation. In this case, the influence function can be split into two factors

$$IP(x, \widehat{\theta}_{LS}, F_X) = \mathbb{E}(\mathbb{X}\mathbb{X}')^{-1}(1,x),$$
$$IR(r, \widehat{\theta}_{LS}, F_\varepsilon) = r = y - \widehat{\theta}_{LS}(F)'(1,x),$$

where F_X represents the marginal distribution of the explanatory variable, F_ε denotes the error distribution and $r = y - \widehat{\theta}_{LS}(F)'(1,x)$ represents the residual associated with a pair (x,y).

In this sense, the factor IP represents the influence of the new observation x. This is closely related to the well-known leverage problem in the regression context. In addition, the factor IR contains the influence of the residual, that is, the effect of a deviation of the response variable y.

Considering now the quantile regression estimator (see equation (2.2)), the influence function can be split into the following two parts:

$$IP(x, \widehat{\theta}_\tau, F_X) = \mathbb{E}(\mathbb{X}\mathbb{X}')^{-1}(1,x),$$
$$IR(r, \widehat{\theta}_\tau, F_\varepsilon) = \operatorname{sgn}(r) = \operatorname{sgn}\left(y - \widehat{\theta}_\tau(F)'(1,x)\right).$$

Then, the influence due to the new observation x matches with the least squared estimator while the influence due to the residual coincides with the influence of the quantile estimator without covariates.

It can then be established that quantile regression can correct robustness problems due to vertical deviations (that is, related to the response variable), but not those caused by horizontal deviations (that is, related to the explanatory variables). Furthermore, in order to control both factors of the influence function, it should be necessary to introduce **generalized M-estimators** that were studied by [31]. Moreover, other kinds of robust estimators have been considered such as least median of squares regression proposed by [34] or regression depth proposed by [35].

We have focused on linear quantile regression but all the ideas presented in this section can be extended to non linear context. Let us consider the following regression scenario:

$$Y_i = q_\tau(X_i, \theta_\tau) + \varepsilon_i,$$

where the function q_τ is known apart from the parameter θ_τ and $\{(X_1,Y_1),\ldots,(X_n,Y_n)\}$ represents a random sample of the variables $(X,Y) \in \mathbb{R}^{d+1}$. Moreover, the conditional τ-quantile of the errors is zero. In this context, we can considerer the following estimator

$$\widehat{\theta}_\tau = \arg\min_{\theta \in \mathbb{R}^q} \sum_{i=1}^{n} \rho_\tau(Y_i - q_\tau(X_i, \theta)). \tag{2.4}$$

In Section 4.5 of [25], the asymptotic behaviour of estimator (2.4) is presented. This result is an extension of Theorem 2.1. Moreover, in some situations, in order to get more flexible approaches, it will be necessary to introduce nonparametric techniques that will be introduced in the Section 3.

3. Nonparametric quantile regression

All the methodology developed along the previous section can be extended to a nonparametric context. In this line, [7] and [8] can be considered as seminal works. In this section we are going to focus on local linear smoothing techniques. Let us consider a regression scenario as

$$Y = q_\tau(X) + \varepsilon,$$

where the conditional τ-quantile of the error given the covariate is zero. Given a random sample of independent observations $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ of the pair $(X, Y) \in \mathbb{R}^2$, a nonparametric estimator of the conditional quantile can be defined as $\widehat{q}_{\tau,h_\tau}(x) = \widehat{a}$, where $\widehat{a}$ and $\widehat{b}$ are the minimizers of

$$\sum_{i=1}^{n} \rho_\tau \left(Y_i - a - b(X_i - x)\right) K\left(\frac{X_i - x}{h_\tau}\right),$$

where K is a kernel function (usually a symmetric density) and h_τ represents a bandwidth parameter. This is the local linear estimator of the quantile regression function.

As it happens for any smoothing method, bandwidth h_τ exhibits a strong influence on the resulting estimate. Several authors have addressed the problem of bandwidth selection, see [43], [1], [44] or [18].

One of the main approaches to bandwidth selection is the plug-in technique which consists of minimizing the dominant terms of the mean integrated squared error (MISE) of the estimator. [17] established the asymptotic MISE for the local linear quantile regression when $n \to \infty$, $h_\tau = h_\tau(n) \to 0$ and $nh_\tau \to \infty$, that is given by

$$\begin{aligned} \text{MISE}\left(\widehat{q}_{\tau,h_\tau}\right) \cong{}& \frac{1}{4} h_\tau^4 \mu_2(K)^2 \int q_\tau^{(2)}(x)^2 g(x)\, dx \\ &+ \frac{R(K)\tau(1-\tau)}{nh_\tau} \int \frac{1}{f(q_\tau(x)|X=x)^2}\, dx, \end{aligned} \tag{3.1}$$

where g is the density of X, $f(q_\tau(x)|X = x)$ is the conditional density of Y at $q_\tau(x)$ given $X = x$, $q_\tau^{(i)}(x) = \partial^i q_\tau(x)/\partial x^i$, $\mu_i(K) = \int u^i K(u)\, du$ and $R(K) = \int K^2(u)\, du$.

Moreover, in view of (3.1), an asymptotically optimal bandwidth can be derived as

$$h_{\text{AMISE},\tau} = \left[\frac{R(K)\ \tau(1-\tau)}{n\ \mu_2(K)^2\ \int q_\tau^{(2)}(x)^2\ g(x)\ dx} \int \frac{1}{f(q_\tau(x)|X=x)^2}\, dx \right]^{1/5}. \tag{3.2}$$

Note that $\mu_2(K)$ and $R(K)$ are obtained from the kernel function, while the two integrals in (3.2) are unknown and have to be estimated. Expression (3.2) is quite similar to the plug-in rule for mean regression but again the sparsity function will play an important role. Because of these similarities with mean regression, [43] proposed to use [36] bandwidth selector with some simple transformations based on the assumptions of homoscedasticity (it is useful to have the same curvature for any τ as in mean regression) and error normality (it allows to estimate the sparsity from the conditional variance). As a result, Yu and Jones (1998) plug-in rule proposal is derived

$$\widehat{h}_{\tau,\text{YJ}} = \sqrt[5]{\frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2}}\, \widehat{h}_{\text{RSW}}, \tag{3.3}$$

where $\widehat{h}_{\text{RSW}}$ is selected by the plug-in rule proposed by [36].

On the other hand, [1] suggested a modification of classical cross-validation function that consisted of replacing the squared loss criterion by the quantile loss function. Bearing this idea in mind, a cross-validation procedure can be applied to select the bandwidth parameter associated with a kernel quantile regression, as follows

$$\widehat{h}_{\tau,\text{CV}} = \arg\min_h \text{CV}(h) = \arg\min_h \sum_{i=1}^{n} \rho_\tau\left(Y_i - \widehat{q}^{-i}_{\tau,h}(X_i)\right),$$

where $\widehat{q}^{-i}_{\tau,h}(X_i)$ is the estimator of the τ-th quantile function obtained from a sample without the i-th individual, that is, the classical leave-one-out estimator, evaluated with bandwidth h.

More recently, [9] provided a plug-in bandwidth for local linear quantile regression based on expression (3.2) without imposing restrictions on the conditional variability and the error distribution. Instead, nonparametric estimations of the curvature at the given quantile τ will be used, as well as nonparametric estimations of the sparsity. Moreover, they prove the convergence of their plug-in estimator to the optimal bandwidth and the convergence rate is the same that in the classical mean regression context.

The aforementioned methods can be extended to the case of a multi-dimensional covariate. For instance, [44] extends the ideas of [43] to nonparametric additive models. Again, the goal is to reduce the problem to a mean regression context under assumptions of homoscedasticity and error normality and then use the selector presented by [32].

Finally, during this section, we focus on kernel smoothing techniques, although spline methods have been widely studied by several authors as [29] or [28]. For instance, [29] proposed to estimate the function q_τ by solving the

following optimization problem

$$\min \left[\sum_{i=1}^{n} \rho_{\tau} \Big(Y_i - q_{\tau}(X_i) \Big) + \lambda \mathbf{V}(\nabla q_{\tau}) \right], \tag{3.4}$$

where $\mathbf{V}(\nabla q_{\tau})$ denotes the total variation of the derivative of q_{τ} and λ represents the well-know smoothing parameter in this context. Moreover, [29] showed that the solution to (3.4) is a linear spline with nodes at the points X_i where $i = 1, \ldots, n$. Hence, a quantile smoothing spline model can be fitted using l_1-type linear programming techniques. They also proposed to adapt the information criterion of [37] for the choice of the smoothing parameter λ involved in problem (3.4).

4. The sparsity function

In view of the asymptotic behaviour of the univariate, parametric and non-parametric quantile regression estimators, it will be necessary to estimate the inverse of the density function evaluated at the quantile of interest. In the regression setup, this function plays an analogous role to the standard deviation of the errors in least squares estimation of the mean regression model.

It is perfectly natural that the precision of quantile estimates should depend on the inverse of the density because it reflects the density of observations near the quantile of interest. If the data are very sparse at the quantile of interest, this quantile will be difficult to estimate. On the other hand, when the sparsity is low and the density is high, the quantile is more precisely estimated.

We are going to start studying the sparsity function associated with a univariate variable, without considering covariates or a regression scenario. Let us consider a random variable Y with associated distribution and density function denoted by $F_{_Y}$ and $f_{_Y}$, respectively. [40] named **sparsity function** to the inverse of the density function evaluated at the quantile, that is given by

$$s(\tau) = \frac{1}{f_{_Y}(F_{_Y}^{-1}(\tau))}.$$

Let us observe that the sparsity function is simply the derivative of the quantile function, that is,

$$\frac{\partial}{\partial t} F_{_Y}^{-1}(t) = \frac{1}{f_{_Y}(F_{_Y}^{-1}(t))} = s(t).$$

Given $\mathcal{Y} = \{Y_1, \cdots, Y_n\}$ a random sample of the variable Y, [38] proposed to estimate the sparsity by a simple difference quotient of the empirical quantile

function, that is,

$$\widehat{s}(t) = \frac{\widehat{F}_n^{-1}(t+h) - \widehat{F}_n^{-1}(t-h)}{2h} = \frac{Y_{[n(\tau+h)]} - Y_{[n(\tau-h)]}}{2h}, \tag{4.1}$$

where $\widehat{F}_n^{-1}$ is the empirical quantile function and h is a bandwidth that tends to zero as the sample size tends to infinity, as well, $Y_{[z]}$ are order statistics. Moreover, $[n(\tau \pm h)]$ are neighbouring orders to τ where $[a]$ denotes the integer part of a. Later, [4] showed that the value of the smoothing parameter that minimizes the asymptotic mean squared error of (4.1) is of order $n^{-1/5}$.

[5] proposed a bandwidth selector in order to compute the nonparametric estimator of the sparsity. In addition, the author proved that the bandwidth

$$h_{\text{B}} = \sqrt[5]{\frac{4.5s(\tau)^2}{s^{(2)}(\tau)^2}}\, n^{-1/5}$$

is optimal from the standpoint of minimizing the mean squared error, where $s^{(2)}(\tau) = \frac{\partial^2}{\partial \tau^2} s(\tau)$.

On the other hand, [19] examined the effect that the selection of the smoothing parameter has on the empirical level of tests or confidence intervals coverage based on Studentized quantiles. In this line, they showed that if we would like to minimize this error, the bandwidth should be of smaller order than that required by squared error theory, such as [5]'s proposal. Bearing this idea in mind, [19] proposed the following smoothing parameter:

$$h_{\text{HS}} = z_{\alpha/2}^{2/3} \sqrt[3]{\frac{1.5S_{d,n}}{|V_{h,n}|}}\, n^{-1/3},$$

where

$$S_{d,n} = \frac{n}{2d}\Big(Y_{[t+d]} - Y_{[t-d]}\Big),$$
$$V_{h,n} = 0.5\left(\frac{n}{h}\right)^3 (Y_{[r+2h]} - 2Y_{[r+h]} + 2Y_{[r-h]} - Y_{[r-2h]}),$$

$t = [n\tau] + 1$, $d = 0.5n^{4/5}$, $r = [0.5n] + 1$, $h = 0.25n^{8/9}$ and $z_{\alpha/2}$ satisfies that $\Phi(z_{\alpha/2}) = 1 - \alpha/2$ with $\alpha = 0.05$ where Φ represents the standard Gaussian distribution.

Now, we are going to move to a regression scenario. Let us consider (X_1, Y_1), $\cdots, (X_n, Y_n)$ a random sample of two variables $(X, Y) \in \mathbb{R}^{d+1}$ drawn from a linear quantile regression model such as (2.1). In this situation, [22] proposed

to estimate the density of the response variable Y given $X = X_i$ as follows

$$\widehat{f}_i = \frac{2h_{\text{HS}}}{(\widehat{\theta}_{\tau^+} - \widehat{\theta}_{\tau^-})' P_i},$$

where h_{HS} represents a smoothing parameter associated with sparsity estimation for Y (without regression) as that given by [19] and $\widehat{\theta}_{\tau^+}$ and $\widehat{\theta}_{\tau^-}$ represent the estimated coefficients of the linear model for neighbouring quantiles

$$\tau^+ = \frac{[n\tau] + nh_{\text{HS}} + 1}{n} \qquad \text{and} \qquad \tau^- = \frac{[n\tau] - nh_{\text{HS}} + 1}{n}.$$

In finite samples, [22] proposed the following modified estimator to combat possible crossing quantiles estimations:

$$\widehat{f}_i = \max\left\{0, \frac{2h_{\text{HS}}}{\left(\widehat{\theta}_{\tau^+} - \widehat{\theta}_{\tau^-}\right)' P_i - \delta}\right\},$$

where δ is a small positive constant included in order to avoid zero denominator.

[22]'s proposal is based on supposing a global linear model, and intended to make inference about its coefficients. To this end the sparsity was estimated by $\frac{1}{\widehat{f}_i}$ using information of neighbouring quantiles. This procedure will properly work only when the relation between X and Y could be fitted by a linear model for different values of the τ.

The study of the sparsity function in a general regression context has not been thoroughly analysed in the literature. [9] presented the first nonparametric sparsity estimator for regression context. Since the sparsity results to be the derivative of the quantile regression function, $q_\tau(x)$, with respect to τ, they propose an estimate of this kind

$$\widehat{s}_{\tau,d_s,h_s}(x) = \frac{\widehat{q}_{\tau+d_s,h_s}(x) - \widehat{q}_{\tau-d_s,h_s}(x)}{2\,d_s}, \tag{4.2}$$

where $\widehat{q}_{\tau+d_s,h_s}$ and $\widehat{q}_{\tau-d_s,h_s}$ are local linear quantile regression estimates at the quantile orders $(\tau + d_s)$ and $(\tau - d_s)$, respectively, and h_s denotes their bandwidth.

Note that two pilot bandwidths, d_s and h_s, are needed to use estimator (4.2). The bandwidth d_s is placed in the Y-axis and plays a similar role to that of the bandwidth d_j in the rule of thumb. The bandwidth h_s is necessary to compute the nonparametric estimations of the regression functions. In order to select these smoothing parameters, it can be use the plug-in technique which consists of minimizing the dominant terms of the mean integrated squared error (MISE) of the estimator given in (4.2). [9] presented the mean squared error of

this sparsity estimator to obtain optimal bandwidths d_s and h_s.

5. Lack-of-fit tests for quantile regression

The lack-of-fit (or in opposite terms, goodness-of-fit) of a statistical model describes how well it fits a set of observations. At the beginning of the twentieth century, Pearson introduced the term goodness-of-fit which main goal is to measure the discrepancy between observed values and the values expected under a specific model. Along this section we are going to present a brief introduction to lack-of-fit tests for quantile regression models.

Let us consider a regression model associated with a quantile of interest $\tau \in (0,1)$,

$$Y = q_\tau(X) + \varepsilon,$$

where ε is the unknown model error of the model that should verify that $\mathbb{P}(\varepsilon \leq 0|X) = \tau$. In this new scenario, the main goal will be to carry out the following lack-of-fit test:

$$\begin{cases} H_0 : q_\tau \in \mathcal{Q}_\theta = \left\{q_\tau(\cdot,\theta) \ : \ \theta \in \Theta \subset \mathbb{R}^q\right\} & \text{Null hypothesis} \\ H_a : q_\tau \notin \mathcal{Q}_\theta & \text{Alternative hypothesis} \end{cases}$$

that is equivalent to

$$H_0 \ : \ \mathbb{E}\left[\mathbb{I}(Y \leq q_\tau(X,\theta_\tau)) \mid X\right] = \tau,$$

for some $\theta_\tau \in \Theta \subset \mathbb{R}^q$.

Then, given $\{(X_1,Y_1),\cdots,(X_n,Y_n)\}$ a random sample of the variables $(X,Y) \in \mathbb{R}^{d+1}$, we are going to review different goodness-of-fit tests in the quantile regression context available from the literature.

Lack-of-fit tests based on smoothing ideas

Regarding the lack-of-fit tests for quantile regression based on smoothing ideas, we should highlight the work developed by [46] that extends the well-known test proposed by [45] to the quantile regression setup. In this case, the test statistic is given by

$$\begin{aligned} T_n^z =& \frac{nh^{d/2}}{\widehat{\sigma}} \ \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \left[\mathbb{I}\big(Y_i \leq q_\tau(X_i,\widehat{\theta}_\tau)\big) - \tau\right] \\ & \times \left[\mathbb{I}\big(Y_j \leq q_\tau(X_j,\widehat{\theta}_\tau)\big) - \tau\right], \end{aligned} \tag{5.1}$$

where K is the kernel function, h is the smoothing parameter and

$$\widehat{\sigma}^2 = 2\tau^2(1-\tau)^2 \, \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} \, K^2\left(\frac{X_i - X_j}{h}\right).$$

The statistic (5.1) converges to a Gaussian distribution. It should be noted the well-known problem associated with the selection of the smoothing parameter, h.

Following the idea of [46], [13] proposed a lack-of-fit test for additive quantile models based on smoothing ideas. In this context, the following test could be raised:

$$H_0 : q_\tau(X) = q_\tau(X^{(1)}, \cdots, X^{(d)}) = \sum_{i=1}^{d} q_{\tau,i}\left(X^{(i)}\right) + c(\tau),$$

where $X = (X^{(1)}, \cdots, X^{(d)}) \in \mathbb{R}^d$ denotes the explanatory variable.

Given a random sample of the variables $(X, Y) \in \mathbb{R}^{d+1}$, [13] proposed the following test statistic:

$$T_n^{\text{DGN}} = \frac{1}{n(n-1)h^d} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K\left(\frac{X_i - X_j}{h}\right) \widehat{R}_i \, \widehat{R}_j,$$

where K represents the kernel function, h is the smoothing parameter and

$$\widehat{R}_i = \mathbb{I}(Y_i \leq \widehat{q_\tau}^{-i}(X_i)) - \tau,$$

where $\widehat{q_\tau}^{-i}(X_i)$ denotes an additive estimation of the quantile regression function without considering the i-th observation. Despite having obtained the asymptotic convergence to a Gaussian distribution, it is more recommended to use a bootstrap procedure in order to calibrate this test.

Lack-of-fit tests based on empirical regression processes

Extending the work developed by [39] to the quantile regression setting, [21] proposed an omnibus lack-of-fit test for parametric quantile regression based on a cumulative sum process of the gradient vector. That is, [21] based their test on the process

$$R_n^{\text{HZ}} = n^{-1/2} \sum_{i=1}^{n} \psi_\tau\left(Y_i - q_\tau(X_i, \widehat{\theta}_\tau)\right) \, q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \, \mathbb{I}(X_i \leq t), \qquad (5.2)$$

where $\psi_\tau(r) = \tau \mathbb{I}(r > 0) + (\tau - 1)\mathbb{I}(r < 0)$, $q_\tau^{(1)}(x, \theta) = \frac{\partial}{\partial \theta}\, q_\tau(x, \theta)$, and $\widehat{\theta}_\tau$ is an estimator of θ_τ. The test statistic proposed by [21] is then defined as

$$T_n^{\text{HZ}} = \text{largest eigenvalue of } \; n^{-1} \sum_{i=1}^{n} R_n^{\text{HZ}}(X_i)\, R_n^{\text{HZ}}(X_i)'.$$

[21] proved that the empirical process (5.2) converges to a Gaussian process with mean 0 and covariance function

$$W(t_1, t_2) = \tau(1-\tau)\, \mathbb{E}\Big(q_\tau^{(1)}(X, \theta_\tau)\, q_\tau^{(1)}(X, \theta_\tau)'\, \mathbb{I}(X \leq \min(t_1, t_2)) - S(t_1) S^{-1} S(t_2) \Big),$$

where

$$S = \mathbb{E}\left[q_\tau^{(1)}(X, \widehat{\theta}_\tau)\, q_\tau^{(1)}(X, \widehat{\theta}_\tau)' \right],$$
$$S(t) = \mathbb{E}\left[q_\tau^{(1)}(X, \widehat{\theta}_\tau)\, q_\tau^{(1)}(X, \widehat{\theta}_\tau)'\, \mathbb{I}(X \leq t) \right].$$

Given that simulating the Gaussian process is not easy, [21] proposed a multiplier bootstrap in order to calibrate their test.

Lack-of-fit tests design for avoiding the curse of dimensionality

It is well-known that a high (or even moderate) dimension of the covariate may affect the performance of the specification tests. In this line, [42] used a He and Zhu type test and defined some ranks over the covariate in order to test a linear quantile regression model. He considered the following empirical process:

$$R_n^{\text{W}}(t) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau(r_i)\, P_i\, \mathbb{I}(F_k \leq t),$$

where $r_i = Y_i - \widehat{\theta}_\tau' P_i$ represents the residuals and $F_i = \max U_{ij}$ where U_{ij} represents the ranks of the n values of the $j-$th column of the design matrix, represented by $\mathbb{X}$[1] for each $j = 2, \cdots, d+1$. Consequently, the test statistic will be

$$T_n^{\text{W}} = \text{largest eigenvalue of } \int R_n^{\text{W}}(t)[R_n^{\text{W}}(t)]' dF_{n,W}(t),$$

where $F_{n,W}$ is the empirical distribution function of the variables F_i.

The proposal of [42] has the virtue of simplicity but does not provide an omnibus test, i.e., it is not consistent for all alternatives. To solve this problem [10] presented an omnibus lack-of-fit test for quantile regression models, that is suitable even with high-dimensional covariates. This test is based on the cumulative sum of residuals with respect to unidimensional linear projections

[1]The design matrix is a $n \times (d+1)$ matrix which $j-$th row is given by $(1, X_j)'$ where $\{X_1, \cdots, X_n\}$ is a random sample of the explanatory variable X.

of the covariates following the ideas of [15] for mean regression context. Their test statistic is defined as

$$T_n^{\text{CSG}} = \text{largest eigenvalue of } \int_{\Pi} R_n^{\text{CSG}}(\beta,u)[R_n^{\text{CSG}}(\beta,u)]' F_{n,\beta}(du)d\beta,$$

where

$$R_n^{\text{CSG}}(\beta,u) = n^{-1/2}\sum_{i=1}^{n}\psi_\tau\left(Y_i - q_\tau(X_i,\theta_\tau)\right)\ q_\tau^{(1)}\left(X_i,\theta_\tau\right)\ \mathbb{I}\left(\beta' X_i \le u\right),$$

$\Pi = \mathbb{S}_d \times [-\infty,+\infty]$, $\mathbb{S}_d$ is the unit sphere on $\mathbb{R}^d$, and $F_{n,\beta}$ is the empirical distribution of the projected covariates $\beta' X_1,\ldots,\beta' X_n$.

On the other hand, [30] adapted the ideas of [46] to a multivariate scenario. The main difference between both tests is that [30]'s work only involves unidimensional kernel smoothing, so that the rate at which it detects local alternatives does not depend on the dimension of covariate. This lack-of-fit test is based on the following test statistic:

$$\begin{aligned} T_n^{\text{MLP}} =& \frac{nh^{1/2}}{\widehat{\sigma}}\ \frac{1}{n(n-1)}\ \sum_{i\neq j}\ \frac{1}{h}\ K\left(\frac{W_i - W_j}{h}\right)\ \psi(Z_i - Z_j) \\ & \times \left[\mathbb{I}\big(Y_i \le q_\tau(X_i,\widehat{\theta}_\tau)\big) - \tau\right]\left[\mathbb{I}\big(Y_j \le q_\tau(X_j,\widehat{\theta}_\tau)\big) - \tau\right], \end{aligned}$$

where

$$\widehat{\sigma}^2 = \frac{2\tau^2(1-\tau)^2}{n(n-1)}\sum_{i\neq j}\frac{1}{h}\ K\left(\frac{W_i - W_j}{h}\right)^2\ \psi(Z_i - Z_j)^2,$$

K and ψ are bounded, even, integrable functions with (almost everywhere) positive, and h represent the univariate smoothing parameter. Note that they assumed that the covariate can be written as $X = (W,Z) \in \mathbb{R}^d$ where W is a unidimensional continuous random variable while Z may include both continuous and discrete variables.

We have mentioned some examples, but other specification tests for quantile regression models can be found in the literature as well as [23] whose goal was to test if the conditional median function is linear against a nonparametric alternative with unknown smoothness; [41] considered an empirical likelihood method to estimate the parameters of the quantile regression models and to construct confidence regions; [33] considered two empirical likelihood-based estimation, inference, and specification testing methods for quantile regression models; or [16] introduced a nonparametric test for the correct specification of a linear conditional quantile function over a continuum of quantile levels.

6. Conclusions

Although mean regression is still a traditional benchmark in regression studies, the quantile approach is receiving increasing attention, because it allows a more complete description of the conditional distribution of the response given the covariate, and it is more robust to deviations from error normality. That is, while classical regression gives only information on the conditional expectation, quantile regression extends the viewpoint on the whole conditional distribution of the response variable.

Along this work an introduction to quantile regression methods is presented. Parametric and nonparametric methods have been introduced and the main advantages of these procedures were mentioned. Finally, some lack-of-fit tests for quantile regression have been shown.

Acknowledgements. The authors gratefully acknowledge the support of Projects MTM2013–41383–P (Spanish Ministry of Economy, Industry and Competitiveness) and MTM2016–76969–P (Spanish State Research Agency, AEI), both co–funded by the European Regional Development Fund (ERDF). Support from the IAP network StUDyS, from Belgian Science Policy, is also acknowledged.

References

[1] Abberger K. 1998. Cross-validation in nonparametric quantile regression. Allgemeines Statistisches Archiv, 82: 149-161.

[2] Bahadur, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics*, **37**, 577-580.

[3] Barrodale, I. and Roberts, F. D. K. (1973). An improved algorithm for discrete L_1 linear approximation. *SIAM Journal on Numerical Analysis*, **10**, 839-848.

[4] Bloch, D. A. and Gastwirth, J. L. (1968). On a simple estimate of the reciprocal of the density function. *The Annals of the Mathematical Statistics*, **39**, 1083-1085.

[5] Bofinger, E. (1975). Estimation of a density function using order statistics. *Australian Journal of Statistics*, **17**, 1-7.

[6] Chatterjee, A. (2011). Asymptotic properties of sample quantiles from a finite population. *Annals of the Institute of Statistical Mathematics*, **63**, 157 - 179.

[7] Chaudhuri, P. (1991a). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of Statistics*, **19**, 760-777.

[8] Chaudhuri, P. (1991b). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis*, **39**, 246-269.

[9] Conde-Amboage, M. and Sánchez-Sellero, C. (2018). A plug-in bandwidth selector for nonparametric quantile regression. *TEST*. `https://doi.org/10.1007/s11749-018-0582-6`.

[10] Conde-Amboage, M., Sánchez-Sellero, C. and González-Manteiga, W. (2015). A lack-of-fit test for quantile regression models with high-dimensional covariates. *Computational Statistics & Data Analysis*, **88**, 128 - 138.

[11] Cook, R. D. and Weisberg, S. (1982). Residuals and Influence in Regression. *Chapman and Hall*.

[12] Davino, C., Furno, M. and Vistocco, D. (2014). *Quantile regression: theory and applications*. John Wiley & Sons.

[13] Dette, H., Guhlich, M., and Neumeyer, N. (2015). Testing for additivity in nonparametric quantile regression. *Annals of the Institute of Statistical Mathematics*, **67**, 437-477.

[14] El Bantli, F. and Hallin, M. (1999). L_1-estimation in linear models with heterogeneous white noise. *Statistics & Probability Letters*, **45**, 305-315.

[15] Escanciano, J.C. (2006). A consistent diagnostic test for regression models using projections. Econometric Theory, 22, 1030-1051.

[16] Escanciano, J.C. and Goh, S.C. (2014). Specification analysis of linear quantile models. Journal of Econometrics, 178, 495-507.

[17] Fan, J., Hu, T. C. and Truong, Y. K. (1994). Robust nonparametric function estimation. *Scandinavian Journal of Statistics*, **21**, 433-446.

[18] El Ghouch A and Genton MG. 2012. Local polynomial quantile regression with parametric features. Journal of the American Statistical Association, 104: 1416-1429.

[19] Hall, P. and Sheather, S. J. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society. Series B (Methodological)*, **50**, 381-391.

[20] Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, **69**, 383-393.

[21] He, X. and Zhu, L.-X. (2003). A lack-of-fit test for quantile regression. *Journal of the American Statistical Association*, **98**, 1013-1022.

[22] Hendricks, W. and Koenker, R. (1992). Hierarchical spline models for conditional quantiles and the demand for electricity. *Journal of the American Statistical Association*, **87**, 58-68.

[23] Horowitz, J.L. and Spokoiny, V.G. (2002). An adaptive, rate-optimal test of linearity for median regression models. *Journal of the American Statistical Association*, **97**, 822-835.

[24] Hyndman, R. J. and Fan, Y. (1996). Sample quantiles in statistical packages. *The American Statistician*, **50**, 361 - 365.

[25] Koenker, R. (2005). Quantile Regression. *Cambridge: Cambridge University Press.*

[26] Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.

[27] Koenker, R. and D'Orey, V. (1987). Computing regression quantiles. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **36**, 383-393.

[28] Koenker, R. and Mizera, I. (2004). Penalized triograms: total variation regularization for bivariate smoothing. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, **66**, 145-163.

[29] Koenker, R., Ng, P. and Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81**, 673-680.

[30] Maistre, S., Lavergne, P. and Patilea, V. (2017). Powerful nonparametric checks for quantile regression. *Journal of Statistical Planning and Inference*, **180**, 13 - 29.

[31] Maronna, R.A. and Yohai, V.J. (1981). Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Probability Theory and Related Fields*, **58**, 7-20.

[32] Opsomer, J. D. and Ruppert, D. (1998). A fully automated bandwidth selection method for fitting additive models. *Journal of the American Statistical Association*, **93**, 605-618.

[33] Otsu, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression models. Journal of Econometrics, 142, 508-538.

[34] Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, **79**, 871-880.

[35] Rousseeuw, P. J. and Hubert, M. (1999). Regression depth. *Journal of the American Statistical Association*, **94**, 388-402.

[36] Ruppert D, Sheather SJ and Wand MP. 1995. An efective bandwidth selector for local least squares regression. Journal of the American Statistical Association. 90: 1257-1270.

[37] Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, **6**, 461-464.

[38] Siddiqui, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *Journal of Research of the National Bureau of Standards*, **64B**, 145-150.

[39] Stute, W. (1997). Nonparametric model checks for regression. The Annals of Statistics, 25, 613-641.

[40] Tukey, J. W. (1965). Which part of the sample contains the information, *Proceedings of the National Academy of Sciences*, **53**, 127-134.

[41] Whang, Y.-J. (2006). Smoothed empirical likelihood methods for quantile regression models. Econometric Theory, 22, 173-205.

[42] Wilcox, R. R. (2008). Quantile regression: A simplified approach to a goodness-of-fit test. Journal of Data Science, 6, 547-556.

[43] Yu, K., and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American statistical Association*, **93**, 228-237.

[44] Yu K and Lu Z. 2004. Local linear additive quantile regression. Scandinavian Journal of Statistics, 31, 333-346.

[45] Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. Journal of Econometrics, 75, 263-289.

[46] Zheng, J. X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restrictions. Econometric Theory, 14, 123-138.