

A plug-in bandwidth selector for nonparametric quantile regression

Mercedes Conde-Amboage¹ ·
César Sánchez-Sellero¹

Received: 12 April 2017 / Accepted: 19 February 2018 / Published online: 3 April 2018
© Sociedad de Estadística e Investigación Operativa 2018

Abstract In the framework of quantile regression, local linear smoothing techniques have been studied by several authors, particularly by Yu and Jones (J Am Stat Assoc 93:228–237, 1998). The problem of bandwidth selection was addressed in the literature by the usual approaches, such as cross-validation or plug-in methods. Most of the plug-in methods rely on restrictive assumptions on the quantile regression model in relation to the mean regression, or on parametric assumptions. Here we present a plug-in bandwidth selector for nonparametric quantile regression that is defined from a completely nonparametric approach. To this end, the curvature of the quantile regression function and the integrated squared sparsity (inverse of the conditional density) are both nonparametrically estimated. The new bandwidth selector is shown to work well in different simulated scenarios, particularly when the conditions commonly assumed in the literature are not satisfied. A real data application is also given.

Keywords Quantile regression · Bandwidth · Nonparametric regression

Mathematics Subject Classification 62G08

✉ César Sánchez-Sellero
cesar.sanchez@usc.es

Mercedes Conde-Amboage
mercedes.amboage@usc.es

¹ Models of Optimization, Decision, Statistics and Applications Research Group (MODESTYA), Department of Statistics, Mathematical Analysis and Optimization, Universidade de Santiago de Compostela, Santiago de Compostela, Spain

1 Introduction

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.

The quantile regression model could be stated as

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

where Y is the response variable of interest, X is the covariate, q_τ is the quantile regression function of order τ and ε represents the error. Then, the conditional τ -quantile of ε given X will be zero, that is, $\mathbb{P}(\varepsilon \leq 0|X) = \tau$ almost surely.

Estimation of the quantile regression model exploits the fact that the conditional quantile, $q_\tau(x)$, is the value a that minimizes the expectation

$$\mathbb{E}[\rho_\tau(Y - a)|X = x],$$

where $\rho_\tau(u) = u(\tau - \mathbb{I}(u < 0))$ and $\mathbb{I}(\cdot)$ is the indicator function of an event. Koenker and Bassett (1978) can be considered a seminal work in estimating conditional quantiles in a parametric setup following this idea.

Along this work, we will focus on the univariate regression context, that is, the covariate X is assumed to be one-dimensional. Yu and Jones (1998) studied a local linear estimator of the quantile regression in a nonparametric framework. To this end, a random sample of independent observations $(X_1, Y_1), \dots, (X_n, Y_n)$ of the pair (X, Y) is supposed to be available. Then, the estimator will be $\widehat{q}_{\tau,h}(x) = \widehat{a}$, where \widehat{a} and \widehat{b} are the minimizers of

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

where K is a kernel function and h_τ is a bandwidth parameter.

Several authors have addressed the problem of bandwidth selection as Yu and Jones (1998), Abberger (1998), Yu and Lu (2004), Ghouch and Genton (2012) or Abberger (2002). In this work, a plug-in rule is designed to choose the bandwidth parameter, h_τ . The plug-in technique consists of minimizing the dominant terms of the mean integrated squared error (MISE) of the estimator. For the local linear quantile regression, it can be written as [see Fan et al. (1994) and Yu and Jones (1998)]

$$\begin{aligned} \text{MISE}(\widehat{q}_{\tau,h_\tau}) &= \mathbb{E} \int (\widehat{q}_{\tau,h_\tau}(x) - q_\tau(x))^2 g(x) dx \\ &\cong \frac{1}{4} h_\tau^4 \mu_2(K)^2 \int q_\tau^{(2)}(x)^2 g(x) dx + \frac{R(K)\tau(1-\tau)}{n h_\tau} \int \frac{1}{f(q_\tau(x)|x)^2} dx \quad (1) \end{aligned}$$

where g is the density of X , $f(q_\tau(x)|x)$ is the conditional density of Y at $q_\tau(x)$ given $X = x$, $q_\tau^{(2)}$ is the second derivative of q_τ , $\mu_2(K) = \int u^2 K(u) du$ and $R(K) = \int K^2(u) du$.

An asymptotically optimal bandwidth can be derived as

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

Note that $\mu_2(K)$ and $R(K)$ are obtained from the kernel function, while the two integrals in (2) are unknown and have to be estimated. Expression (2) is quite similar to the plug-in rule for mean regression. The curvature (integrated squared second derivative) is now calculated for the quantile regression function instead of the mean regression, while the integrated squared sparsity (where ‘‘sparsity’’ means the inverse of the conditional density) replaces the integrated conditional variance that appeared in mean regression. See Ruppert et al. (1995) where a plug-in rule is given for local linear mean regression.

Because of these similarities with mean regression, Yu and Jones (1998) proposed to use Ruppert et al. (1995) bandwidth with some simple transformations based on the assumptions of homoscedasticity and error normality. Homoscedasticity is useful to have the same curvature for any τ as in mean regression, while normality allows to estimate the sparsity from the conditional variance.

The purpose of this work is to provide a plug-in bandwidth for local linear quantile regression 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.

Other proposals in the literature for bandwidth selection in nonparametric quantile regression were given, based on cross-validation techniques. In particular, Abberger (1998) proposed to minimize in h the cross-validation function given by:

$$CV(h) = \sum_{i=1}^n \rho_\tau \left(Y_i - \widehat{q}_{\tau, h}^{-i}(X_i) \right) \tag{3}$$

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

In Sect. 2, a preliminary rule of thumb is obtained, and afterwards, the proposed plug-in rule is derived. In Sect. 3, a simulation study is given to explore the virtues of the new bandwidth selectors in comparison with Yu and Jones (1998) and Abberger (1998) proposals. Section 4 contains the main conclusions. In ‘‘Appendix’’, mean squared errors of curvature and sparsity estimators are derived.

2 Proposed selectors

As any plug-in rule, the crucial ingredients of our proposed selectors will be the estimators of unknown quantities, which in our case are the curvature and the sparsity.

Our first proposal will consist of a rule of thumb, where the estimators are defined on a simple partition of the sample in blocks.

2.1 Rule of thumb

Following the ideas in Ruppert et al. (1995), a rule of thumb can be constructed by doing the next steps:

1. Partition the range of X into N blocks with the same number of observations. The original sample $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ is subsequently split into the N blocks. A polynomial of order 4 is adjusted at each block, thus providing N polynomials that will be denoted by $\hat{q}_{\tau,j}$ with $j = 1, \dots, N$. The number of blocks will be chosen as \hat{N} following the Mallows's C_p criterion [see Mallows (1973)] adapted to the quantile frameworks, that is, \hat{N} will minimize

$$C_p(N) = \frac{RSQ(N)}{RSQ(N_{\max})/(n - 5N_{\max})} - (n - 10N)$$

where $RSQ(N)$ is the residual sum of quantile losses given by ρ_τ and summed over each blocked quartic fit, when the number of blocks is N , $N_{\max} = \max\{\min([n/20], N^*), 1\}$ and $N^* = 5$. Here $[\cdot]$ denotes the integer part of a number.

Remark 1 Following the ideas of Ruppert et al. (1995) in mean regression, we have considered polynomials of order 4 because it is the lowest degree that a polynomial admits estimates of the quantity $\vartheta_{24} = \int q_\tau^{(2)}(x) q_\tau^{(4)}(x) g(x) dx$ other than zero. This integral ϑ_{24} will be involved in the mean squared error of curvature estimator, see equation (4).

2. Estimate the curvature as:

$$\hat{\vartheta}_B = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\hat{N}} \hat{q}_{\tau,j}^{(2)}(X_i)^2 I(X_i \in \text{Block } j).$$

Observe that we are using the notation $\vartheta = \int q_\tau^{(2)}(x)^2 g(x) dx$.

3. Estimate the sparsity at each block j by means of

$$\hat{s}_j(\tau) = \frac{r_{[\tau+d_j]} - r_{[\tau-d_j]}}{2d_j}$$

where $r_{[\tau-d_j]}$ and $r_{[\tau+d_j]}$ are the sample quantiles of orders $(\tau - d_j)$ and $(\tau + d_j)$, respectively, of the residuals from the quartic fit at block j . This type of sparsity estimator was suggested by Siddiqui (1960) and studied by Bloch and Gastwirth (1968). For the parameter d_j , the selector proposed by Bofinger (1975) will be used here. Finally, the integrated squared sparsity will be estimated by

$$\widehat{s}_B^2 = \sum_{j=1}^{\widehat{N}} \widehat{s}_j^2 l_j$$

where l_j denotes the length of block j .

4. Finally, the selector from the rule of thumb will be obtained as:

$$\widehat{h}_{\tau,RT} = \left(\frac{R(K) \tau(1 - \tau) \widehat{s}_B^2}{n \mu_2(K)^2 \widehat{\vartheta}_B} \right)^{1/5} .$$

2.2 Plug-in rule

The plug-in rule will come from a more elaborated estimation of the curvature and the sparsity.

Curvature estimation

Now the second derivative of the regression will be nonparametrically estimated at each sample observation. In order to do this, a local polynomial of order 3 will be adjusted. Let us call $\widetilde{q}_{\tau,h_c}^{(2)}(X_i)$ to its second derivative at X_i , for $i = 1, \dots, n$. Then, we can consider the following curvature estimator:

$$\widehat{\vartheta}_{h_c} = \frac{1}{n} \sum_{i=1}^n \widetilde{q}_{\tau,h_c}^{(2)}(X_i)^2$$

At this point, a pilot bandwidth h_c for curvature estimation should be selected. The criterion for selecting h_c will be the mean squared error of the curvature estimator. As in the case of classical mean regression, see Ruppert et al. (1995), the asymptotic mean squared error coincides, up to terms not depending on the bandwidth and negligible terms, with the asymptotic squared bias, which is given by

$$\begin{aligned} \text{MSE}(\widehat{\vartheta}_{h_c}) \cong & \left[\delta_1 h_c^2 \int q_{\tau}^{(2)}(x)q_{\tau}^{(4)}(x)g(x)dx \right. \\ & \left. + \delta_2 \tau(1 - \tau) \frac{1}{nh_c^5} \int \frac{1}{f(q_{\tau}(x)|x)^2} dx \right]^2 \end{aligned} \tag{4}$$

where

$$\begin{aligned} \delta_1 &= \frac{1}{6}(\alpha_{31}\mu_4(K) + \alpha_{33}\mu_6(K)) \\ \delta_2 &= 4 \left(\alpha_{31}^2 \int K^2(v)dv + \alpha_{33}^2 \int v^4 K^2(v)dv + 2\alpha_{31}\alpha_{33} \int v^2 K^2(v)dv \right) \\ \mu_i(K) &= \int v^i K(v)dv \end{aligned}$$

$$\alpha_{31} = \frac{-\mu_2(K)^2\mu_6(K) + \mu_2(K)\mu_4(K)^2}{\mu_2(K)\mu_4(K)\mu_6(K) - \mu_4(K)^3 - \mu_2(K)^3\mu_6(K) + \mu_2(K)^2\mu_4(K)^2}$$

$$\alpha_{33} = \frac{\mu_2(K)\mu_6(K) - \mu_4(K)^2}{\mu_2(K)\mu_4(K)\mu_6(K) - \mu_4(K)^3 - \mu_2(K)^3\mu_6(K) + \mu_2(K)^2\mu_4(K)^2}$$

Minimizing the last expression, the asymptotically optimal pilot bandwidth will be

$$\tilde{h}_c = C(K) \left(\frac{\tau(1 - \tau) \int 1/f(q_\tau(x)|x)^2 dx}{|\int q_\tau^{(2)}(x)q_\tau^{(4)}(x)g(x)dx| n} \right)^{1/7}$$

where

$$C(K) = \begin{cases} C^I = \left(\frac{5\delta_2}{2\delta_1}\right)^{1/7} & \text{if } \int q_\tau^{(2)}(x)q_\tau^{(4)}(x)g(x)dx > 0 \\ C^{II} = \left(\frac{\delta_2}{\delta_1}\right)^{1/7} & \text{if } \int q_\tau^{(2)}(x)q_\tau^{(4)}(x)dx < 0 \end{cases}$$

To compute this pilot bandwidth, preliminary estimations of the integrated squared sparsity and the integral $\vartheta_{24} = \int q_\tau^{(2)}(x)q_\tau^{(4)}(x)g(x)dx$ are required. They will be obtained from blocked estimators as those considered for the rule of thumb. They can be denoted by \hat{s}_b^2 and $\hat{\vartheta}_{24,b}$. The resulting estimated pilot bandwidth will be

$$\hat{h}_c = C(K) \left(\frac{\tau(1 - \tau) \hat{s}_b^2}{|\hat{\vartheta}_{24,b}| n} \right)^{1/7} .$$

Finally, the curvature estimator will be given by

$$\hat{\vartheta}_{h_c}^\alpha = \frac{1}{n} \sum_{i=1}^n \tilde{q}_{\tau, \hat{h}_c}^{(2)}(X_i)^2 \mathbb{I}((1 - \alpha)a + \alpha b < X_i < \alpha a + (1 - \alpha)b)$$

where the sample was trimmed at each border a and b , by a small proportion $\alpha \in [0, 1]$, assuming that the covariate is supported in the interval $[a, b]$. This strategy was already used by Ruppert et al. (1995) in their estimation of similar quantities for mean regression. It is intended to prevent from the variability of local polynomial kernel estimates of high derivatives near the boundaries. Following their suggestion, we will take $\alpha = 0.05$.

Sparsity estimation

Since the sparsity, denoted by $s_\tau(x)$, is the derivative of the quantile regression function, $q_\tau(x)$, with respect to τ , we propose an estimate of this kind:

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

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 we need two pilot bandwidths, d_s and h_s . 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.

The choice of the two pilot bandwidths will be based on the asymptotic mean squared error, which comes from the asymptotic squared bias and variance:

$$\begin{aligned} \text{MSE} \left(\int \widehat{s}_{\tau, d_s, h_s}^2(x) \, dx \right) &\cong \left[\frac{1}{nd_s h_s} \int a(x) \, dx + d_s^2 \int b(x) \, dx + h_s^2 \int c(x) \, dx \right]^2 \\ &+ \frac{1}{nd_s} \int d(x) \, dx + \frac{1}{n^2 d_s^2 h_s} \int e(x) \, dx \end{aligned} \tag{5}$$

where

$$\begin{aligned} a(x) &= \frac{1}{2} \frac{R(K) s_\tau(x)^2}{g(x)} & d(x) &= 2 \frac{s_\tau(x)^4}{g(x)} \\ b(x) &= \frac{1}{3} s_\tau(x) s_\tau^{(2, \tau)}(x) & e(x) &= \left(\frac{1}{2} R(K * K) - R(K) \right) \frac{s_\tau(x)^4}{g(x)^2} \\ c(x) &= \mu_2(K) s_\tau(x) \frac{\partial q_\tau^{(2)}(x)}{\partial \tau} \end{aligned} \tag{6}$$

where $*$ represents the convolution and $s_\tau^{(2, \tau)}(x) = \frac{\partial^2}{\partial \tau^2} s_\tau(x)$.

Minimization with respect to d_s and h_s can be carried out by means of optimization algorithms as Newton–Raphson or Nelder and Mead (1965)’s method. Estimation of the six integrals in expression (5) is done by blocks. The resulting pilot bandwidths will be denoted by \widehat{d}_s and \widehat{h}_s and the estimation of the integrated squared sparsity by $\widehat{s}_{\tau, \widehat{d}_s, \widehat{h}_s}^2$. Now details are given on how to estimate the unknown integrals.

- **Estimation of $\int a(x) \, dx$** Note that $a(x) = (1/2) R(K) s_\tau(x)^2 (g(x))^{-1}$. We will make use of the sparsity estimation at each block, \widehat{s}_j , together with a simple estimation of covariate density at that block, which could be given by $n_j/(nl_j)$, where n_j is the number of observations at block j . Then, this integral can be estimated by:

$$\int \widehat{a}(x) \, dx = \frac{1}{2} R(K) \sum_{j=1}^{\widehat{N}} \widehat{s}_j^2 \left(\frac{nl_j}{n_j} \right) l_j.$$

- **Estimation of $\int b(x) \, dx$** Recall that $b(x) = (1/3) s_\tau(x) s_\tau^{(2, \tau)}(x)$, where $s_\tau^{(2, \tau)}(x)$ is the second derivative of $s_\tau(x)$ with respect to τ . The problem of estimating the second derivative of the sparsity without covariates was considered by Bofinger (1975). We apply her proposal to the residuals at each block

$$\widehat{s}_j^{(2,\tau)} = \frac{1}{2\delta^3} (r_{(\lfloor n\tau \rfloor + 2m)} - 2r_{(\lfloor n\tau \rfloor + m)} + 2r_{(\lfloor n\tau \rfloor - m)} - r_{(\lfloor n\tau \rfloor - 2m)})$$

where the value of m is taken as $m = \lfloor n\delta \rfloor = \lfloor cn^{8/9} \rfloor$ with $c = 0.25$, following Sheather and Maritz (1983) proposal. Then, the considered integral is estimated as

$$\int \widehat{b(x)} dx = \frac{1}{3} \sum_{j=1}^{\widehat{N}} \widehat{s}_j \widehat{s}_j^{(2,\tau)} l_j.$$

– **Estimation of $\int c(x) dx$** The novel ingredient in $c(x)$ is $\partial q_\tau^{(2)}(x)/\partial \tau$. Since this is a derivative with respect to τ , it can be estimated by

$$\frac{\partial \widehat{q_\tau^{(2)}(x)}}{\partial \tau} = \frac{q_{\tau+d_c}^{(2)}(x) - q_{\tau-d_c}^{(2)}(x)}{2d_c}$$

In order to choose the pilot bandwidth d_c , a location and scale model, given by $Y = q_\tau(X) + \sigma(X)\varepsilon$, is assumed. Here, ε is assumed independent of X and with a zero τ th quantile. Note that under this model, for each $\tau_1, \tau_2 \in (0, 1)$, $q_{\tau_2}(x) - q_{\tau_1}(x) = \sigma(x)(c_{\tau_2} - c_{\tau_1})$, where c_{τ_1} and c_{τ_2} are τ_1 and τ_2 quantiles of ε , respectively. Thus,

$$\frac{\partial q_\tau^{(2)}(x)}{\partial \tau} = \sigma^{(2)}(x) s_\tau(x).$$

This expression leads to consider for d_c the same selector proposed by Bofinger (1975) to estimate the sparsity without covariates. This selector will also be based on the assumption of normality for ε . Finally, we arrive at the following estimator at block j

$$\left(\frac{\partial \widehat{q_\tau^{(2)}}}{\partial \tau} \right)_j = \frac{1}{n_j} \sum_{i=1}^n \frac{\widehat{q}_{\tau+\widehat{d}_{c,j}}^{(2)}(X_i) - \widehat{q}_{\tau-\widehat{d}_{c,j}}^{(2)}(X_i)}{2\widehat{d}_c} \mathbb{I}(X_i \in \text{Block } j),$$

and the subsequent estimation of the integral

$$\int \widehat{c(x)} dx = \mu_2(K) \sum_{j=1}^{\widehat{N}} \widehat{s}_j \left(\frac{\partial \widehat{q_\tau^{(2)}}}{\partial \tau} \right)_j l_j$$

– **Estimation of $\int d(x) dx$ and $\int e(x) dx$** Note that $d(x) = 2s_\tau(x)^4 (g(x))^{-1}$ and $e(x) = (0.5R(K * K) - R(K)) s_\tau(x)^4 g(x)^{-2}$. Similarly to the previous integrals, these integrals can be estimated by

$$\int \widehat{d(x)} dx = 2 \sum_{j=1}^{\widehat{N}} \widehat{s}_j^4 \left(\frac{n l_j}{n_j} \right) l_j$$

$$\int \widehat{e(x)} dx = \left(\frac{1}{2} R(K * K) - R(K) \right) \sum_{j=1}^{\widehat{N}} \widehat{s}_j^4 \left(\frac{n l_j}{n_j} \right)^2 l_j.$$

Finally, the plug-in bandwidth selector is obtained as

$$\widehat{h}_{\tau, \text{PI}} = \left(\frac{R(K) \tau(1 - \tau) \widehat{s}_{\tau, \widehat{d}_s, \widehat{h}_s}^2}{n \mu_2(K)^2 \widehat{\vartheta}_{\widehat{h}_c}} \right)^{1/5}.$$

Remark 2 In the framework of local linear smoothing quantile regression, Yu and Jones (1998) presented a different approach based on inverting a local linear conditional distribution estimator that is called double-kernel estimator. Later, Jones and Yu (2007) proposed an improvement of their previous double-kernel estimator. Both estimators need bandwidth selectors. The choice of the main bandwidth (h_1 in their notation) could be done by the plug-in rule proposed here. A small experiment is given at the end of the simulation study to show the performance of the new plug-in rule in double-kernel estimators.

2.3 Theoretical performance

The selector from the rule of thumb includes inconsistent estimators of curvature and sparsity. Thus, consistency properties cannot be derived for this selector. Meanwhile, convergence of the plug-in bandwidth selector to the optimal bandwidth relies on the asymptotic properties of curvature and sparsity estimators, $\widehat{\vartheta}_{\widehat{h}_c}$ and $\widehat{s}_{\tau, \widehat{d}_s, \widehat{h}_s}^2$, respectively. The same arguments given in Ruppert et al. (1995) in the case of local linear mean regression can be followed here. The main difference comes from the sparsity estimator, which replaces the conditional variance in AMISE representation.

From expression (5), it can be obtained that for sequences of pilot bandwidths $d_s = D_s n^{-1/5}$ and $h_s = H_s n^{-1/5}$, where $D_s > 0$ and $H_s > 0$ are constants, $\widehat{s}_{\tau, d_s, h_s}^2 - \int s_{\tau}^2(x) dx = O_P(n^{-2/5})$. Even though this rate of convergence is slower than root-n, it is enough to achieve that the relative rate of the plug-in bandwidth selector is dominated by curvature estimation, that is,

$$(\widehat{h}_{\tau, \text{PI}} - h_{\text{MISE}}) / h_{\text{MISE}} = -\frac{1}{5} (\widehat{\vartheta}_{\widehat{h}_c} - \vartheta) / \vartheta + O_P(n^{-2/5})$$

where $\vartheta = \int q_{\tau}^{(2)}(x) g(x) dx$ is the true curvature.

Now, from expression (4), and for a sequence of pilot bandwidths $h_c = H_c n^{-1/7}$, with $H_c > 0$ a certain constant, we have that, conditionally on X_1, \dots, X_n ,

$$(\widehat{h}_{\tau, \text{PI}} - h_{\text{MISE}}) / h_{\text{MISE}} \xrightarrow{P} L \tag{7}$$

where $L = -\frac{1}{5} \vartheta^{-1} \left\{ \delta_1 \int q_\tau^{(2)}(x) q_\tau^{(4)}(x) g(x) dx H_c^2 + \delta_2 \tau(1 - \tau) \int s_\tau^2(x) dx H_c^{-5} \right\}$.

A detailed proof of (7) would follow the steps given in Sánchez-Sellero et al. (1999). Expression (7) shows that the relative rate of convergence of $\widehat{h}_{\tau, \text{pl}}$ is $O_p(n^{-2/7})$ for any choice of H_c . Observe now that the asymptotically optimal pilot bandwidth \tilde{h}_c allows to make L equal to zero, thanks to an ideal choice of H_c . This pilot bandwidth would lead to an improved $O_p(n^{-5/14})$ relative rate of convergence for the plug-in bandwidth selector. To obtain this in practice, consistent estimators of the unknown quantities in \tilde{h}_c would be needed, which could be quite complicated. Our proposed estimated pilot bandwidth, \widehat{h}_c , is based on rule-of-thumb estimates of the unknown quantities, which are simple to implement although they do not guarantee consistency, neither improved rate of convergence. The theoretical performance of the plug-in bandwidth selector is then similar to that of Ruppert et al. (1995) bandwidth selector for mean regression. The only difference was found in the sparsity estimation, which replaces the conditional variance estimation. We can conclude that the slower rate of convergence of the sparsity estimation does not have an effect on the rate of convergence of the plug-in selector.

3 Simulation study

In this section, a simulation study is presented to analyse the behaviour of the new bandwidth selectors in comparison with already existing selectors. The natural competitors would be Yu and Jones (1998)'s bandwidth and Abberger (1998)'s cross-validation bandwidth. As regards Yu and Jones (1998)'s bandwidth, some theoretical considerations are useful as an orientation to a meaningful comparison. Recall the expression given in (2) for the asymptotically optimal bandwidth

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

Observing that the same type of bandwidth for mean regression is given by

$$h_{\text{AMISE}, \text{MEAN}} = \left[\frac{R(K)}{n \mu_2(K)^2 \int m^{(2)}(x)^2 g(x) dx} \int \sigma^2(x) dx \right]^{1/5},$$

where m is the mean regression and σ^2 is the conditional variance, Yu and Jones (1998) proposed to use the following selector:

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

where \widehat{h}_{RSW} is the Ruppert et al. (1995)'s plug-in selector for local linear mean regression and the last factor is a correction for quantile regression. Yu and Jones (1998)'s selector is based on assuming that quantile and mean regression have the same cur-

vature and the error distribution is normal. The last factor then relates the conditional sparsity with the conditional variance under normality.

Since \widehat{h}_{RSW} converges to $h_{AMISE,MEAN}$, \widehat{h}_{YJ} converges to

$$h_{AMISE, YJ, \tau} = h_{AMISE, MEAN} \left[\frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2} \right]^{1/5}$$

which is generally different from the asymptotically optimal bandwidth for quantile regression, $h_{AMISE, \tau}$. Meanwhile, the proposed plug-in selector \widehat{h}_{NP} converges to $h_{AMISE, \tau}$. Then, for a sample size large enough, the new bandwidth is expected to outperform Yu and Jones (1998)'s selector, the latter selector being generally inconsistent. This simulation study will help to assess the consequences of these facts from smaller to larger sample sizes, and in models where the difference between $h_{AMISE, YJ, \tau}$ and $h_{AMISE, \tau}$ can be controlled.

In particular, for any homoscedastic quantile regression model $Y = q_{\tau}(X) + \varepsilon$, where the model error ε has τ -quantile zero and is assumed independent of X , the curvatures of mean and quantile regression coincide, and then, the quotient between $h_{AMISE, YJ, \tau}$ and $h_{AMISE, \tau}$ will be

$$\text{Ratio} = \frac{h_{AMISE, YJ, \tau}}{h_{AMISE, \tau}} = \sqrt[5]{\frac{\sigma^2 f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau))^2}{\phi(\Phi^{-1}(\tau))^2}} \tag{8}$$

where f_{ε} and F_{ε} are the density and distribution functions of ε and σ^2 denotes the variance of ε . Then, the ratio between both AMISE bandwidths only depends on the error distribution for any homoscedastic regression model. Some calculations lead to the following ratio between asymptotic mean integrated squared errors of the two bandwidths:

$$\frac{AMISE(h_{AMISE, YJ, \tau})}{AMISE(h_{AMISE, \tau})} = \frac{1}{5}\text{Ratio}^4 + \frac{4}{5}\text{Ratio}^{-1} \tag{9}$$

where Ratio is defined in (8). Note that, by construction, the ratio between AMISEs is always larger or equal to one. Part (a) of Fig. 1 shows the values taken by the ratio defined in (8) as a function of the quantile order τ and for three error distributions: exponential with expectation one, uniform on the interval (0, 1) and beta with parameters 5 and 1. Part (b) of Fig. 1 shows the values taken by the ratio defined in (9) as a function of τ and for the same three error distributions.

As shown in Fig. 1, we observe that the differences between both AMISE bandwidths will be bigger as long as the error distribution differs from the Gaussian distribution. Furthermore, if we fix an error distribution, the compared behaviour of both optimal bandwidths will depend on the quantile of interest.

Our first simulated model is given by

$$\text{Model 1: } Y = 10(X^4 + X^2 - X) + \varepsilon,$$

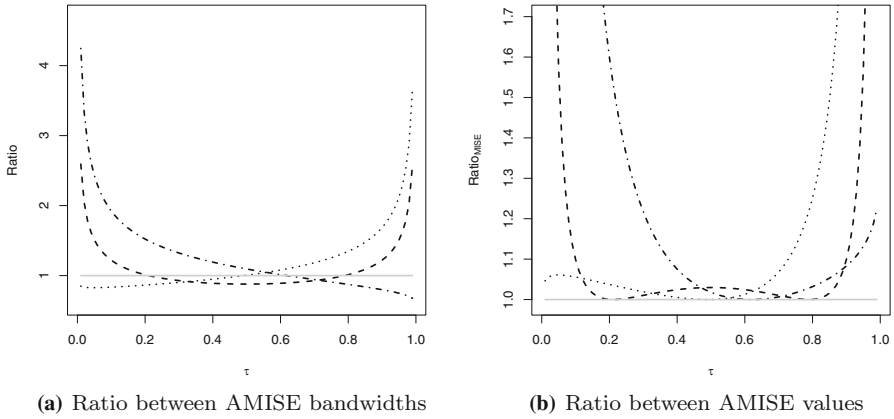


Fig. 1 Representations of the ratios between the AMISE bandwidths [detailed in (8)] and the MISE values [detailed in (9)] as a function of the quantile order τ and for three error distributions. The dashed line represents the uniform distribution, the dotted line represents the beta distribution, and the dashed and dotted line represents the exponential distribution

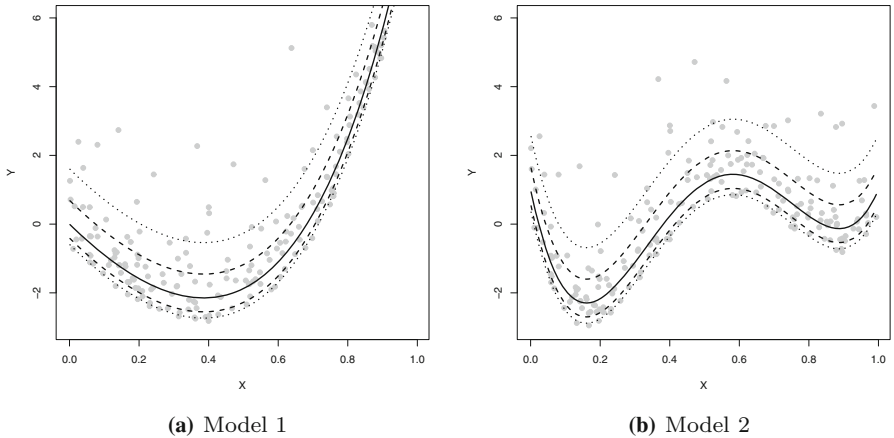


Fig. 2 Scatterplots of a sample of size 200, together with five quantile regression functions: $\tau = 0.1$ (dotted line), $\tau = 0.25$ (dashed line), $\tau = 0.5$ (solid line), $\tau = 0.75$ (dashed line) and $\tau = 0.9$ (dotted line), corresponding to Model 1 in (a) and Model 2 in (b)

where X follows a uniform distribution on the interval $(0, 1)$ and ε is the unknown error, which is drawn independently of X . Note that in this case, $q_\tau(X) = 10(X^4 + X^2 - X) + c_\tau$ where c_τ represents the τ -quantile of the error distribution. This notation is common for all the homoscedastic models that will be considered. In this model, the error follows an exponential distribution with expectation 1, which is one of the distributions represented in Fig. 1. Part (a) of Fig. 2 shows a scatterplot of one sample of size 200 drawn from this model, together with three quantile functions, for $\tau = 0.1, 0.25, 0.5$.

Figure 3 represents the boxplots corresponding to the four bandwidth selectors: the plug-in selector proposed by Yu and Jones (1998), the selector based on the new rule of thumb, the new plug-in selector and the cross-validation selector. They are denoted by

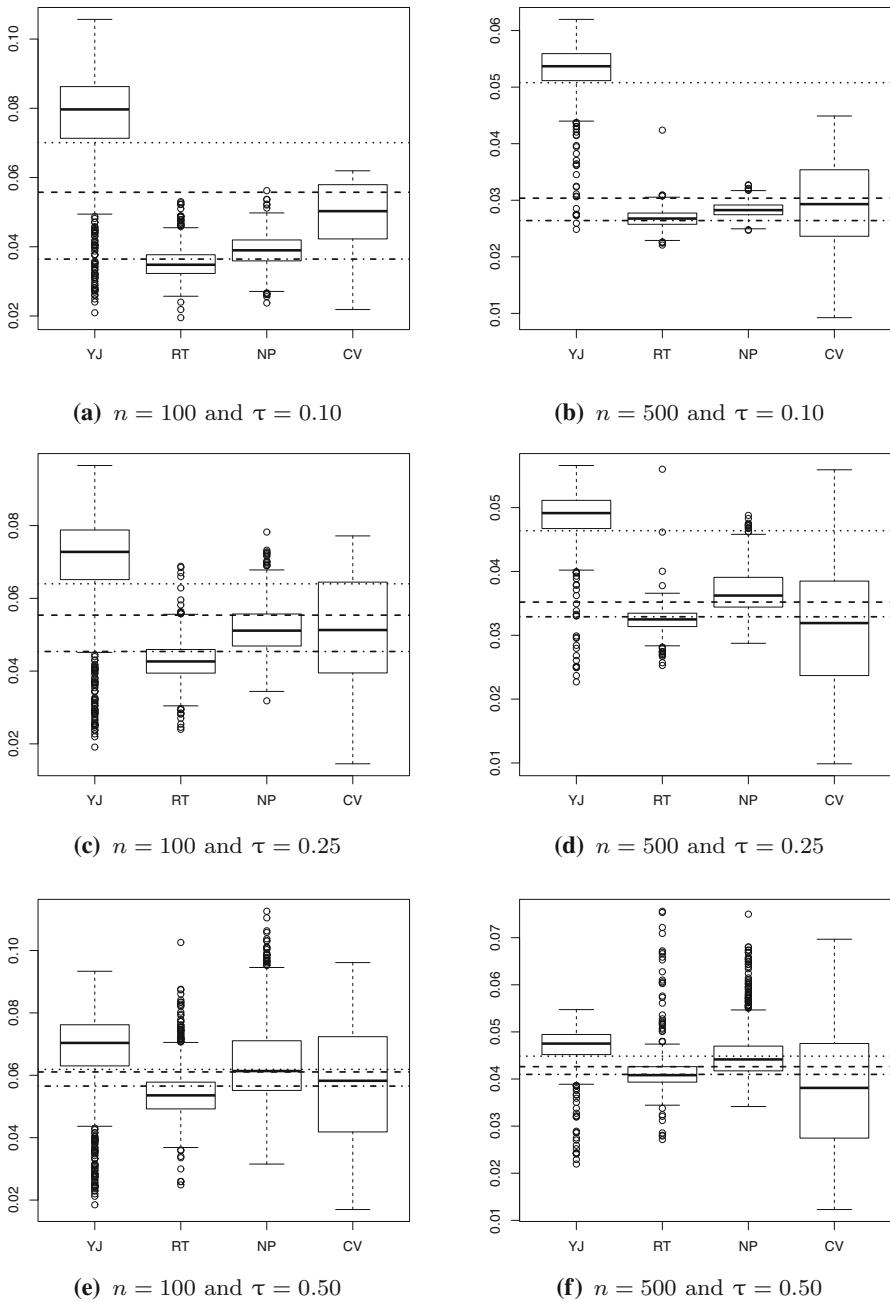


Fig. 3 Boxplot representations of Yu and Jones (1998)'s selector (YJ), the new rule of thumb (RT), the new plug-in selector (NP) and the cross-validation selector (CV), from 1000 replications of Model 1 for different values of τ and the sample size, n . The dashed line represents the MISE bandwidth, the dotted line represents the Yu and Jones (1998)'s AMISE bandwidth, and the dashed and dotted line represents the AMISE bandwidth

YJ, *RT*, *NP* and *CV*, respectively. The boxplots were obtained from 1000 replications of Model 1 for different values of τ , and sample sizes $n = 100, 500$. Three horizontal lines are added to the plots, representing the optimal bandwidths with three criteria: MISE (dashed line), Yu and Jones (1998)'s AMISE (dotted line) and AMISE (dashed and dotted line). The best of these bandwidths would be the one optimizing the MISE, so the performance of each selector is related to its approximation to this bandwidth. The AMISE bandwidth is an approximation to the MISE bandwidth. In fact, both lines approach to each other for increasing sample size. Meanwhile, Yu and Jones (1998)'s AMISE (*YJ*-AMISE) bandwidth do not approximate to MISE bandwidth even for a very large sample size. This is the cause for inconsistency of Yu and Jones (1998)'s selector. However, for a small sample size, the errors of approximation between the three bandwidths can be confounded. As regards the value of τ , Fig. 3 shows that for $\tau = 0.5$ the three bandwidths are quite similar, while for $\tau = 0.1$ they are far apart.

Yu and Jones (1998)'s selector estimates *YJ*-AMISE bandwidth, while the new selectors estimate AMISE bandwidth. For sample size $n = 500$, this leads to a clearer better performance of the new bandwidths, while for small sample size $n = 100$, the errors between optimal bandwidths are still confounded. The cross-validation bandwidth is generally centred to the MISE bandwidth, but its variability is clearly larger.

Now we are going to evaluate the performance of each selector in terms of the observed integrated squared error (OISE) in one thousand simulated samples. Following Jones (1991), for each sample the OISE will be computed for the local linear fit with the considered bandwidth selectors, that is,

$$\text{OISE}(\hat{h}_\tau) = \frac{1}{n} \sum_{i=1}^n (\hat{q}_{\tau, \hat{h}_\tau}(X_i) - q_\tau(X_i))^2,$$

where \hat{h}_τ plays the role of some bandwidth selector. Then, the sample means of the OISEs (denoted by SMISE) over the simulated samples will be employed for comparison.

To complete the presentation, a new model is included, again homoscedastic but with a larger curvature:

$$\text{Model 2: } Y = 1 - 48X + 218X^2 - 315X^3 + 145X^4 + \varepsilon,$$

where X follows a uniform distribution on the interval $(0, 1)$ and ε follows an exponential distribution with expectation 1 and is drawn independently of X . Part (b) of Fig. 2 shows a scatterplot and three quantile functions, for $\tau = 0.1, 0.25, 0.5$, corresponding to Model 2.

Table 1 contains the sample mean of the integrated squared error for the considered bandwidth selectors for several samples sizes and values of τ . We can observe that the new plug-in rule shows a better performance in terms of SMISE than the plug-in selector proposed by Yu and Jones (1998), for almost all sample sizes for $\tau = 0.10$ and 0.25 . For $\tau = 0.50$, the SMISE associated with both plug-in rules is quite similar. Note that in this case the ratio described in (8) is near to 1 as shown in Fig. 2. That is, for $\tau = 0.5$ the two AMISE bandwidths are almost equal.

Table 1 Sample mean of the integrated squared error (given values were multiplied by 10^3) associated with the considered bandwidth selectors, from 1000 replications of Model 1 and Model 2, with several sample sizes n and values of τ

τ	n	Model 1				Model 2			
		YJ	RT	NP	CV	YJ	RT	NP	CV
0.10	100	29.51	32.41	29.69	28.67	32.21	48.65	31.79	31.44
	500	6.93	4.71	4.65	5.48	7.16	5.99	5.49	6.26
	1000	4.01	2.41	2.39	2.78	4.50	3.06	2.95	3.34
0.25	100	48.81	45.82	43.49	52.64	51.66	59.66	50.58	56.84
	500	11.23	9.31	9.43	11.65	12.42	11.61	11.12	13.54
	1000	6.28	5.02	5.09	6.02	7.57	6.39	6.28	7.36
0.50	100	94.01	88.23	89.12	103.71	107.46	109.01	99.69	116.28
	500	21.97	21.61	21.98	26.62	26.23	26.35	26.36	31.56
	1000	12.04	11.80	12.03	14.23	14.98	14.86	14.99	17.41
0.75	100	220.00	243.79	226.03	241.79	264.25	286.24	248.07	264.26
	500	50.84	54.21	52.19	63.58	63.07	66.13	62.61	76.70
	1000	28.24	29.58	28.56	34.76	35.35	36.46	35.28	42.77
0.90	100	528.90	853.22	768.06	558.42	647.76	776.51	791.56	586.72
	500	132.19	227.02	168.36	153.09	161.88	235.36	180.48	181.56
	1000	71.18	114.67	83.08	85.45	89.12	120.62	91.86	103.66

On the other hand, the results associated with quantiles $\tau = 0.75$ and 0.90 are better for the selector presented by Yu and Jones (1998). These results are consequence of the proximity of the ratio (9) to 1 (see Fig. 1) and a low density of the error distribution at these high quantiles. A ratio (9) close to 1 means that inconsistency of Yu and Jones (1998)'s selector has not a severe effect for small sample sizes. A low density of the error distribution at the considered quantile makes curvature and sparsity estimation more difficult. In a sense, we are in ideal conditions for Yu and Jones (1998)'s selector versus the new plug-in selector: being curvature and sparsity similar to their analogues in mean regression, and easier to estimate in mean regression.

In any case, it should be observed that due to inconsistency of Yu and Jones (1998)'s selector, for a sample size large enough SMISE will be better for the plug-in selector proposed here. Table 2 shows this behaviour. Table 2 does not include results for the cross-validation selector because of its computational cost for large sample sizes.

It is interesting to emphasize the good behaviour of the rule of thumb, despite its simplicity. For a fair interpretation, we should note that the considered models are homoscedastic and contain polynomial quantile regression functions of order 4, these being ideal conditions for the rule of thumb. The cross-validation bandwidth shows a generally worst SMISE in the considered scenarios.

Now, we will analyse how the performance of the considered bandwidth selectors depends on the error distribution. To do this, we will generate samples from these two models:

Table 2 Sample mean of the integrated squared error (given values were multiplied by 10^4) associated with the considered bandwidth selectors, from 1000 replications of Model 1 and Model 2, with several sample sizes n and values of τ

τ	n	Model 1			Model 2		
		YJ	RT	NP	YJ	RT	NP
0.10	5000	12.79	5.56	5.57	15.57	7.19	7.13
	10000	8.07	3.17	3.18	10.10	4.05	4.06
0.25	5000	16.66	12.63	12.63	22.12	16.36	16.39
	10000	9.84	7.27	7.30	13.40	9.43	9.52
0.50	5000	30.60	30.00	30.16	39.75	38.69	38.89
	10000	17.73	17.38	17.46	23.31	22.53	22.70
0.75	5000	75.04	75.98	74.81	94.17	95.01	94.19
	10000	41.33	41.63	41.11	53.25	53.74	53.50
0.90	5000	19.00	25.92	19.54	23.79	23.07	23.78
	10000	10.75	13.23	10.44	13.51	14.85	13.13

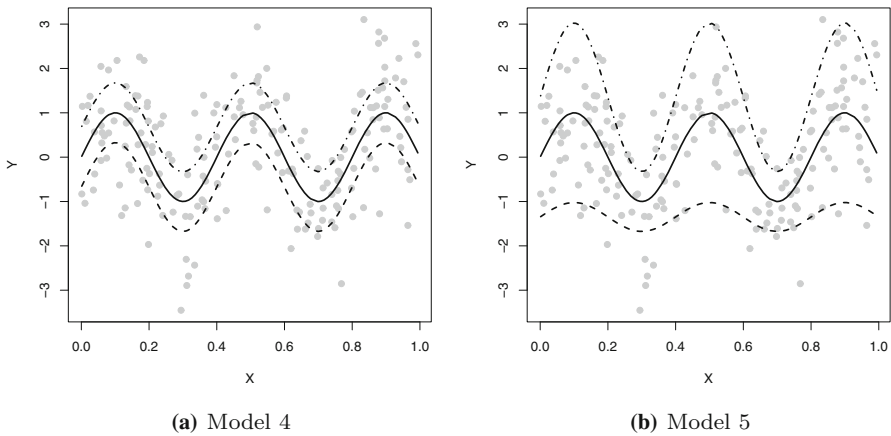


Fig. 4 Scatterplots of a sample of size 200 drawn from Model 4 and Model 5, where the error follows a standard normal distribution. The lines are quantile functions for $\tau = 0.25$ (dashed line), $\tau = 0.5$ (solid line) and $\tau = 0.75$ (dashed and dotted line)

$$\text{Model 3: } Y = 1 - 48X + 218X^2 - 315X^3 + 145X^4 + \varepsilon$$

$$\text{Model 4: } Y = \sin(5\pi X) + \varepsilon$$

where X follows a uniform distribution on the interval $(0, 1)$ and ε is independent of X and follows one of these distributions: standard normal, uniform on the interval $(-3, 3)$, Student's t with two degrees of freedom and standard log normal. Quantile function in Model 3 coincides with that of Model 2, while the error distribution now takes different shapes. Model 4 is represented in Part (a) of Fig. 4, with a standard normal distribution.

Table 3 Sample mean of the integrated squared error (given values were multiplied by 10^2) associated with the considered bandwidth selectors, from 1000 replications of Model 3 and Model 4 with $\tau = 0.5$, and several error distributions and sample sizes

ε	n	Model 3				Model 4			
		YJ	RT	NP	CV	YJ	RT	NP	CV
$N(0, 1)$	100	13.33	13.85	13.01	14.88	15.18	18.11	15.29	16.73
	500	3.47	3.54	3.46	4.25	3.92	4.25	3.95	4.74
	1000	1.98	2.02	1.99	2.37	2.23	2.35	2.24	2.59
$U(-3, 3)$	100	52.17	52.33	51.08	56.68	59.27	63.07	57.64	62.01
	500	14.61	14.95	14.26	17.36	16.31	18.48	16.24	19.43
	1000	8.49	8.57	8.25	10.04	9.36	10.39	9.31	11.19
t_2	100	23.78	20.81	19.78	21.49	29.63	31.03	26.16	24.43
	500	4.88	4.56	4.49	5.34	6.64	5.57	5.41	6.05
	1000	2.93	2.54	2.51	2.99	3.77	3.01	2.95	3.28
$\log N(0, 1)$	100	22.11	20.17	18.43	21.37	28.37	28.01	21.56	23.74
	500	4.32	4.02	3.98	4.82	4.82	4.57	4.36	5.17
	1000	2.49	2.18	2.19	2.58	2.76	2.44	2.34	2.72

Table 3 shows the sample mean of the integrated squared errors for the compared bandwidth selectors, under Model 3 and Model 4. In all cases, the quantile function is estimated for $\tau = 0.5$. The new plug-in rule outperforms the other three selectors. Yu and Jones (1998)'s selector shows a good performance for the standard normal error distribution, where its assumptions are completely satisfied. However, the new plug-in rule has similar results to Yu and Jones (1998)'s selector, even under these conditions, which shows that in this case quantile estimations of curvature and sparsity are not much less efficient than its estimations under mean regression. For distributions far from normality, as Student's t distribution or log normal, the new plug-in rule shows a clearly better behaviour. All these results are to be attributed to sparsity estimation, which is inconsistently biased in Yu and Jones (1998)'s method. Note that the simulated models are homoscedastic, and then, quantile curvature coincides with mean curvature.

The rule of thumb is slightly worse than the plug-in rule, although the difference is moderate in many cases. In particular, rule of thumb results are better under Model 3 than under Model 4, because the quantile function under Model 3 is better suited for blocked polynomial estimations carried out in the rule of thumb method. The cross-validation selector is generally worse than plug-in methods, and particularly worse than the new plug-in rule.

Moreover, we are going to consider the following heteroscedastic quantile regression model:

$$\text{Model 5: } Y = \sin(5\pi X) + (\sin(5\pi X) + 2)\varepsilon$$

where X follows a uniform distribution on the interval $(0, 1)$ and ε is independent of X . Note that in this case $q_\tau(X) = \sin(5\pi X) + (\sin(5\pi X) + 2)c_\tau$ where c_τ denotes the

Table 4 Sample mean of the integrated squared error (given values were multiplied by 10^2) associated with the compared bandwidth selectors, from 1000 replications of Model 5 and for several values of τ and the sample size n

τ	n	Standard Gaussian				Student t with 3 degrees of freedom			
		YJ	RT	NP	CV	YJ	RT	NP	CV
0.25	100	61.15	54.01	50.75	50.07	107.41	99.51	84.74	66.21
	500	14.05	12.93	11.54	13.27	19.29	16.23	15.21	16.65
	1000	7.96	7.76	6.73	7.85	10.42	8.73	8.14	9.30
0.50	100	60.04	64.32	56.88	63.08	79.28	74.23	69.68	78.93
	500	14.44	16.61	14.53	17.89	18.11	20.22	18.48	21.09
	1000	8.27	9.28	8.31	1.01	9.86	10.77	10.56	11.92
0.75	100	89.13	10.27	83.48	90.10	154.98	208.38	148.07	145.09
	500	21.16	23.88	20.06	24.89	37.37	37.25	30.64	38.01
	1000	12.40	13.42	11.57	14.47	20.11	19.21	16.46	20.71

τ -quantile of the error distribution. Firstly, ε is drawn from the standard normal distribution. Then, the main deviation of Model 5 from Yu and Jones (1998)'s assumptions is the fact that curvature depends on the quantile order, τ , and then, it is not equal to the curvature of mean regression function. Part (b) of Fig. 4 shows a representation of Model 5. A scatterplot together with three quantile functions (for $\tau = 0.25, 0.5, 0.75$) is shown. It can be seen how heteroscedasticity leads to different curvatures of the quantile regression function for different values of τ . Secondly, we will suppose that the error follows a Student t distribution with three degrees of freedom. In this second situation, neither of the assumptions considered by Yu and Jones (1998) are verified.

In Table 4, the sample mean of the integrated squared error from each of the bandwidth selectors is given for several samples sizes and values of τ . The new plug-in method provides better results than its competitors. Note that for $\tau = 0.5$ and Gaussian error distribution quantile regression coincides with mean regression, so this setup would be quite favourable for Yu and Jones (1998)'s selector. In this case, both plug-in selectors shows similar results. For quantile orders far from the median, advantages of the new plug-in rule are more noticeable. Furthermore, the differences between the sample mean of the integrated squared error associated with both plug-in methods are bigger when we suppose that the error follows a Student t distribution, as it was expected.

Now, we are going to check the robustness of the new method to deviations from some smoothness conditions assumed for the quantile regression model. In particular, we are going to generate values from a model that is not differentiable:

$$\text{Model 6: } Y = 5|X| + \sigma(X) \varepsilon$$

where X follows a uniform distribution on the interval $(-1, 1)$ and ε is independent of X . Two possible error distributions will be considered: a χ -squared distribution with two degrees of freedom and a Student's t distribution with two degrees of freedom.

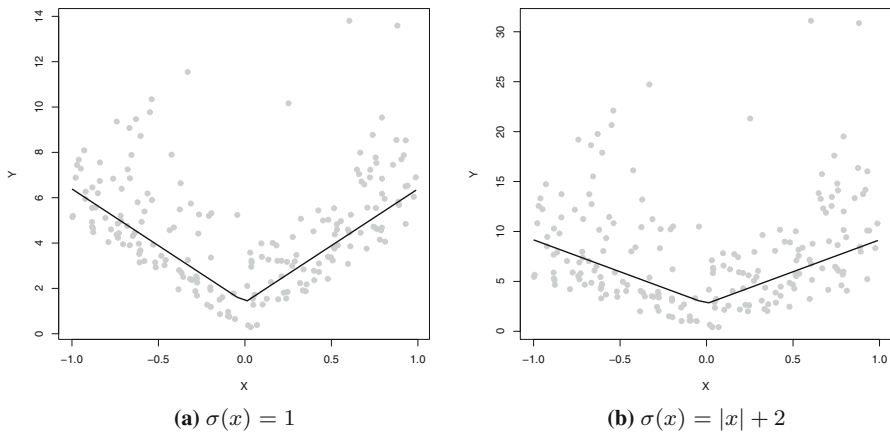


Fig. 5 Scatterplots of samples of size 200 drawn from Model 6, where the error follows a χ -squared distribution with two degrees of freedom. The lines represent the median regression function

Table 5 Sample mean of the integrated squared error (given values were multiplied by 10^2) associated with the compared bandwidth selectors, from 1000 replications of Model 6 and for several sample sizes, n

ε	n	$\sigma(x) = 1$				$\sigma(x) = x + 2$			
		YJ	RT	NP	CV	YJ	RT	NP	CV
χ_2^2	100	31.03	27.22	27.53	31.85	185.15	158.15	160.94	175.17
	500	6.95	6.95	6.88	7.84	37.04	36.63	36.24	42.42
	1000	3.85	3.76	3.88	4.24	19.86	19.50	19.44	22.37
t_2	100	20.08	15.22	16.34	17.13	120.74	84.10	84.48	86.37
	500	4.29	3.74	4.02	4.15	19.11	18.27	18.62	21.08
	1000	2.51	2.12	2.41	2.26	10.28	9.67	10.09	11.02

Note that in this case $q_\tau(X) = 5|X| + \sigma(X) c_\tau$ where c_τ denotes the τ -quantile of the error distribution. Two different options will be considered for the function $\sigma(X)$: $\sigma(X) = 1$ (homoscedastic model) and $\sigma(X) = (|X| + 2)$ (heteroscedastic model). Figure 5 shows a representation of Model 6.

Table 5 shows the sample mean of the integrated squared error when estimating the median regression with each of the bandwidth selectors. The new rule-of-thumb and plug-in methods provide better results than the other selectors in most of the considered scenarios. No relevant anomalies were observed in the performance of the selectors when smoothness conditions are not satisfied.

In a last experiment, we carried out some simulations to show the usefulness of the new plug-in selector for double-kernel methodology. In each of the two double-kernel estimators, one proposed by Yu and Jones (1998) and the other proposed by Jones and Yu (2007), two bandwidths are required. One of these bandwidths (h_1 in their notation) plays a more relevant role and behaves as a classical bandwidth for local linear quantile regression. In both works, the authors proposed to use the selector proposed by Yu and

Table 6 Sample mean of the integrated squared error (given values were multiplied by 10^2) associated with the considered estimators: ordinary local linear estimator, double-kernel estimator presented by Yu and Jones (1998) (DK YJ) and double-kernel estimator presented by Jones and Yu (2007) (DK JY), from 1000 replications of Model 3, for two error distributions, three values of τ and two sample sizes. Four bandwidth selectors were considered: the plug-in selector proposed by Yu and Jones (1998) (YJ), the selector based on the new rule of thumb (RT), the new plug-in selector (denoted by NP) and the cross-validation selector (CV)

ε	τ	n	Ordinary Local linear				DK YJ		DK JY	
			YJ	RT	NP	CV	YJ	NP	YJ	NP
t_2	0.25	100	69.80	115.17	41.77	42.24	76.89	53.35	66.26	45.92
		500	7.36	7.94	7.08	8.51	7.82	7.21	7.09	7.88
	0.50	100	25.61	24.16	23.31	25.40	26.36	24.09	23.56	20.95
		500	4.88	4.56	4.49	5.34	5.16	4.79	4.51	4.54
	0.75	100	62.75	231.53	63.83	38.18	57.95	247.97	60.78	132.50
		500	8.0	7.73	7.29	9.06	10.49	9.27	8.58	7.86
$\log N(0, 1)$	0.25	100	10.88	9.26	7.19	8.59	114.19	8.08	9.88	7.29
		500	2.11	1.60	1.53	1.84	2.08	1.63	1.96	1.61
	0.50	100	25.98	20.47	18.81	21.76	27.68	20.37	24.35	18.54
		500	4.30	3.98	3.94	4.77	5.55	4.90	4.61	4.16
	0.75	100	86.09	101.81	76.23	66.36	92.00	85.60	79.93	76.87
		500	14.13	13.29	12.13	16.30	16.16	16.05	13.16	13.07

Jones (1998) for this main bandwidth. Then, we are going to compare the plug-in rule and Yu and Jones (1998)’s rule, when applied to the selection of this bandwidth h_1 for double-kernel estimators. The second and less relevant bandwidth will be chosen following the authors’ advices. Data will be drawn from Model 3 used previously that is given by

$$\text{Model 3: } Y = 1 - 48X + 218X^2 - 315X^3 + 145X^4 + \varepsilon$$

where X follows an uniform distribution on the interval $(0, 1)$ and ε is independent of X and follows one of these two distributions: Student’s t with two degrees of freedom and standard log normal.

Table 6 contains the sample mean of the integrated squared error (SMISE) obtained from 1000 Monte Carlo replications, for different estimators: ordinary local linear estimator, double-kernel estimator proposed by Yu and Jones (1998) (denoted by DK YJ) and double-kernel estimator proposed by Jones and Yu (2007) (denoted by DK JY). Furthermore, the different bandwidth selectors used along this simulation study will be considered: YJ, RT, NP and CV. To simplify the comparison, for double-kernel estimator only YJ and NP selectors will be considered.

According to the results shown in Table 6, we can conclude that the new plug-in rule improves the performance of both double-kernel estimators. Only for a Student’s t distribution with two degrees of freedom, $\tau = 0.75$ and $n = 100$, Yu and Jones (1998)’ bandwidth leads to a better performance. Note also that ordinary local linear

estimator and the double-kernel estimator presented by Jones and Yu (2007) behave similarly when the new plug-in selector is used.

4 Real data application

The data set *Mammals*, included in the R package *quantreg*, contains 107 observations on the maximal running speed of mammal species and their body mass. Figure 6 represents the scatterplot of these two variables, together with local linear quantile fits for $\tau = 0.25, 0.5$ and 0.75 . Koenker (2005) uses this data set to illustrate how sensitive the least-squares fitting procedure is to outlying observations (see pp. 232–234). Here, we only consider local linear quantile fits, and we will compare bandwidth selectors. Note that the proposed plug-in bandwidth selector is based on quantile techniques, while Yu and Jones (1998)'s selector is based on classical estimates of curvature and conditional variance and then could be sensitive to outliers. It can also be observed that the chosen data set shows asymmetry of the response (the maximal running speed) conditionally to the explanatory variable (the body weight), with more conditional density around high quantiles and lower density around low quantiles.

Solid lines in Fig. 6 represent local linear quantile fits with the new plug-in bandwidth selector, while dotted lines are obtained with Yu and Jones (1998)'s rule. For $\tau = 0.5$, both fits are quite similar. In this case, the proposed plug-in bandwidth takes the value 1.36, while Yu and Jones (1998)'s bandwidth takes the value 1.16. As a consequence, the dotted line seems slightly more wiggly, maybe due to the effect of outliers on curvature and conditional variance estimation. For $\tau = 0.25$, the bandwidths are 1.59 for the new rule and 1.20 for Yu and Jones (1998)'s rule. Then, the dotted line is even more wiggly than the solid line, showing spurious fluctuation. Note that the density of the response around the 0.25 conditional quantile is low. This fact is taken into account by the new plug-in rule, but not by Yu and Jones (1998)'s rule. On the contrary, the density of the response around the 0.75 conditional quantile is high. The selected bandwidths are 0.97 for the new rule and 1.20 for Yu and Jones (1998)'s rule. Thus, the dotted line is over-smoothed and hides relevant features in the data. In particular, the change in slope around 1 Kg of weight is not detected by the dotted line. It can also be observed that Yu and Jones (1998)'s rule selects the same bandwidth for 0.25 and 0.75 quantiles. This is a general behaviour of this rule, which takes the same value for τ and $(1 - \tau)$ conditional quantiles, as a consequence of assuming that the conditional distribution of the response is Gaussian. Then, it does not take into account possible asymmetries in the conditional distribution, as it is the case in this real data situation.

5 Conclusions and extensions

We have proposed a new plug-in bandwidth selector for local linear quantile regression based on a nonparametric approach. This new method involves nonparametric estimation of the curvature of the quantile regression function and the integrated squared sparsity. Convergence of the new rule to the optimal bandwidth is shown, with the same rate as for mean regression selectors.

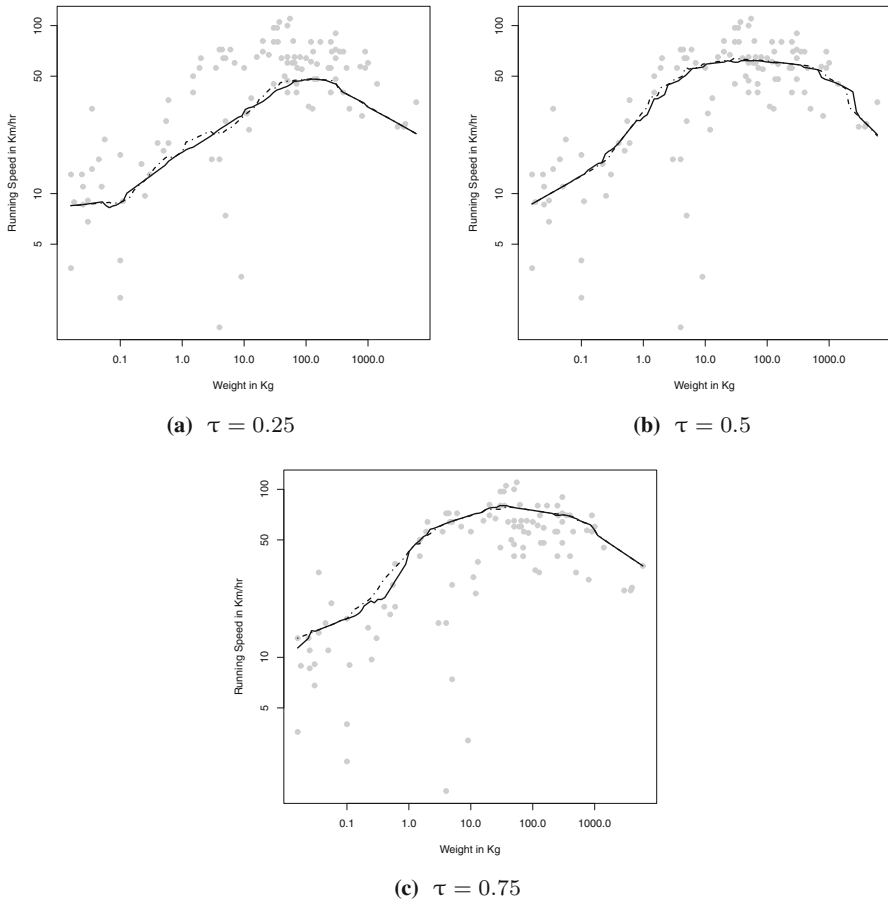


Fig. 6 Local linear quantile regression fits for Mammals data set with $\tau = 0.25, 0.5$ and 0.75 . Solid lines are obtained with bandwidths selected by the new plug-in rule. Dotted lines result from bandwidths selected by Yu and Jones (1998)'s rule

Thanks to a Monte Carlo simulation study, we have shown that the new proposal shows a good behaviour in terms of the sample mean of the integrated squared error compared with its natural competitors in both homoscedastic and heteroscedastic scenarios. Moreover, we have presented a simple rule of thumb that shows a quite good performance on a wide range of situations.

An R package called `BwQuant` has been developed to enable any user to apply the techniques proposed in this paper: rule of thumb and plug-in rule. The natural competitors, cross-validation and Yu and Jones (1998)'s bandwidths, were also implemented. Moreover, we have included a function that estimates the quantile regression function using local linear kernel regression.

The developed methodology can be used in the double-kernel estimator proposed by Yu and Jones (1998) and Jones and Yu (2007), as it was illustrated in a last experiment at the end of the simulation study. Moreover, the proposed techniques can be

extended to the case of a multidimensional covariate, particularly to nonparametric additive models in a quantile regression context as those considered by Yu and Lu (2004). Similarly to Yu and Jones (1998), Yu and Lu (2004) proposed a heuristic rule for selecting the smoothing parameter, using Opsomer and Ruppert (1998)'s bandwidth for mean regression with some transformation based on assumptions such as homoscedasticity and error normality. A plug-in rule specifically designed for additive quantile regression would be more appropriate when these assumptions are not satisfied. This plug-in rule would benefit from the ideas given in this paper.

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. Work of M. Conde-Amboage has been supported by FPU grant AP2012-5047 from the Spanish Ministry of Education. We are grateful to two anonymous referees for their constructive comments, which helped to improve the paper.

Appendix: Mean squared error of curvature and sparsity estimators

Here expressions (4) and (5) are derived. They give approximations to the mean squared error of curvature and sparsity estimators, respectively. A complete development of these expressions can be seen in Chapter 3 of Conde-Amboage (2017).

Derivation of (4)

In order to derive the asymptotic mean integrated squared error of the curvature estimator, the following assumptions will be needed:

- C1 The density function of the explanatory variable X , denoted by g , is differentiable, and its first derivative is a bounded function.
- C2 The kernel function K is symmetric and nonnegative, has a bounded support and verifies that $\int K(u) du = 1$, $\mu_6(K) = \int u^6 K(u) du < \infty$ and $\int K^2(u) du < \infty$. Moreover, it is assumed that the bandwidth parameter h_c verifies that $h_c \rightarrow 0$ and $nh_c^5 \rightarrow \infty$ when $n \rightarrow \infty$.
- C3 The conditional distribution function $F(y|X = x)$ of the response variable is three times derivable in x for each y and its first derivative verifies that $F^{(1)}(q_\tau(x)|X = x) = f(q_\tau(x)|X = x) \neq 0$. Moreover, there exist positive constants c_1 and c_2 and a positive function $\text{Bound}(y|X = x)$ such that

$$\sup_{|x_n - x| < c_1} f(y|X = x_n) \leq \text{Bound}(y|x)$$

and

$$\int |\psi_\tau(y - q_\tau(x))|^{2+\delta} \text{Bound}(y|X = x) \, dy < \infty$$

$$\int (\rho_\tau(y - t) - \rho_\tau(y) - \psi_\tau(y)t)^2 \text{Bound}(y|X = x) \, dy = o(t^2), \quad \text{as } t \rightarrow 0$$

where $\psi_\tau(r) = \tau \mathbb{I}(r > 0) + (\tau - 1)\mathbb{I}(r < 0)$.

C4 The function $q_{\tau_1}(x)$ has a continuous fourth derivative with respect to x for any τ_1 in a neighbourhood of τ . These derivatives will be denoted by $q_\tau^{(i)}$ with $i \in \{1, 2, 3, 4\}$. Moreover, all these derivatives are bounded functions in a neighbourhood of τ .

Applying the arguments of the proof of Theorem 3 in Fan et al. (1994) to a local polynomial of order 3, the estimator of the second derivative can be approximated by

$$\tilde{q}_{\tau,h_c}^{(2)}(x) \cong q_\tau^{(2)}(x) + 2h_c^{-2} \frac{1}{f(q_\tau(x)|x)g(x)} V_{n,\tau}(x)$$

where

$$V_{n,\tau}(x) = \frac{1}{nh_c} \sum_{i=1}^n \psi_\tau \left(Y_i^{(3)} \right) \left(\alpha_{31} + \alpha_{33} \left(\frac{X_i - x}{h_c} \right)^2 \right) K \left(\frac{x - X_i}{h_c} \right),$$

where $Y_i^{(3)} = Y_i - q_\tau(x) - q_\tau^{(1)}(x)(X_i - x) - (1/2)q_\tau^{(2)}(x)(X_i - x)^2 - (1/6)q_\tau^{(3)}(x)(X_i - x)^3$ and $\psi_\tau(z) = \tau - \mathbb{I}(z < 0)$. Note that assumptions C1-C4 were used here.

Now, expectation and variance of $\tilde{q}_{\tau,h_c}^{(2)}(x)$ can be obtained by some algebraic calculations:

$$\mathbb{E} \left(\tilde{q}_{\tau,h_c}^{(2)}(x) \right) \cong q_\tau^{(2)}(x) + \frac{1}{2} \delta_1 q_\tau^{(4)}(x) h_c^2$$

$$\text{Var} \left(\tilde{q}_{\tau,h_c}^{(2)}(x) \right) \cong \delta_2 \frac{1}{nh_c^5} \frac{\tau(1 - \tau)}{f(q_\tau(x)|x)^2 g(x)}$$

where δ_1 and δ_2 were defined in expression (4). Recall that the curvature estimator is given by

$$\widehat{\vartheta}_{h_c} = \frac{1}{n} \sum_{i=1}^n \tilde{q}_{\tau,h_c}^{(2)}(X_i)^2.$$

Then, combining expectation and variance of $\tilde{q}_{\tau,h_c}^{(2)}(X_i)$ conditionally to X_i , and taking expectation with respect to X_i , we obtain

$$\mathbb{E} \left(\widehat{\vartheta}_{h_c} \right) \cong \vartheta + \delta_1 h_c^2 \int q_\tau^{(2)}(x) q_\tau^{(4)}(x) g(x) \, dx$$

$$+ \delta_2 \tau(1 - \tau) \frac{1}{nh_c^5} \int \frac{1}{f(q_\tau(x)|x)^2} \, dx$$

Additional calculations, which can be found in Conde-Amboage (2017), show that the dominant terms in the variance of $\widehat{\vartheta}_{h_c}$ are of orders n^{-1} and $n^{-2} h_c^{-9}$. The term of order n^{-1} does not depend on h_c , while the term of order $n^{-2} h_c^{-9}$ is negligible with respect to the asymptotic squared bias. Because of this, the asymptotically optimal bandwidth can be obtained by minimizing the asymptotic squared bias. This fact, together with last expression for $\mathbb{E}(\widehat{\vartheta}_{h_c})$, leads to expression (4).

Derivation of (5)

The following conditions will be assumed in order to derive the asymptotic mean integrated squared error of the sparsity estimator:

- S1 The conditional density function $f(y|X = x)$ of the response variable is twice derivable in x for each y and $f^{(i)}(q_\tau(x)|X = x) \neq 0$ with $i = 0, 1, 2$. Moreover, there exist positive constants c_1 and c_2 and a positive function $\text{Bound}(y|X = x)$ such that

$$\sup_{|x_n - x| < c_1} f(y|X = x_n) \leq \text{Bound}(y|X = x)$$

and

$$\int |\psi_\tau(y - q_\tau(x))|^{2+\delta} \text{Bound}(y|X = x) dy < \infty$$

$$\int (\rho_\tau(y - t) - \rho_\tau(y) - \psi_\tau(y)t)^2 \text{Bound}(y|X = x) dy = o(t^2), \quad \text{as } t \rightarrow 0$$

where $\psi_\tau(r) = \tau \mathbb{I}(r > 0) + (\tau - 1) \mathbb{I}(r < 0)$.

- S2 The function q_{τ_1} has a continuous second derivative for any τ_1 in a neighbourhood of τ as a function of x . These derivatives will be denoted by $q_{\tau_1}^{(i)}$. Moreover, all these functions are bounded functions in a neighbourhood of τ .
- S3 The density function of the explanatory variable X , denoted by g , is differentiable, and this first derivative is a bounded function.
- S4 The kernel K is symmetric and nonnegative, has a bounded support and verifies that $\int K(u) du < \infty$, $\int K(u)^2 du < \infty$ and $\mu_2(K) < \infty$. Moreover, the bandwidth parameters verify that $d_s \rightarrow 0$, $h_s \rightarrow 0$ and $nd_s h_s \rightarrow \infty$ when $n \rightarrow \infty$.
- S5 The function q_{τ_1} has a continuous and bounded fourth derivative with respect to τ_1 for any τ_1 in a neighbourhood of τ . Moreover, $q_{\tau_1}^{(2)}$ has a continuous and bounded second derivative with respect to τ_1 for any τ_1 in a neighbourhood of τ .

Recall the definition of the proposed sparsity estimator

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

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. Applying Fan et al. (1994)'s results, we have

$$\widehat{q}_{\tau+d_s, h_s}(x) \cong q_{\tau+d_s}(x) + \frac{1}{f(q_{\tau+d_s}(x)|x)g(x)} U_{\tau+d_s, h_s}(x)$$

where

$$U_{\tau+d_s, h_s}(x) = \frac{1}{nh_s} \sum_{i=1}^n \psi_{\tau+d_s}(Y_i^{(1)}) K\left(\frac{x - X_i}{h_s}\right), \tag{10}$$

and $Y_i^{(1)} = Y_i - q_{\tau+d_s}(x) - q_{\tau+d_s}^{(1)}(x)(X_i - x)$. Analogously for $\widehat{q}_{\tau-d_s, h_s}(x)$.

Substituting these expressions in the definition of $\widehat{s}_{\tau, d_s, h_s}(x)$, we have

$$\widehat{s}_{\tau, d_s, h_s}(x) = A(x) + B(x) \tag{11}$$

with

$$A(x) = \frac{q_{\tau+d_s}(x) - q_{\tau-d_s}(x)}{2d_s}, \quad B(x) = \frac{1}{g(x)} \left(\frac{U_{\tau+d_s, h_s}(x)}{f(q_{\tau+d_s}(x)|x)} - \frac{U_{\tau-d_s, h_s}(x)}{f(q_{\tau-d_s}(x)|x)} \right).$$

Note that $A(x)$ is not random and can be approximated by a Taylor expansion as

$$A(x) \cong s_{\tau}(x) + \frac{1}{6} s_{\tau}^{(2, \tau)}(x) d_s^2$$

if S2 follows. Moreover, based on arguments developed in Lemma 2 of Fan et al. (1994), the expectation and variance of $B(x)$ can be approximated by

$$\begin{aligned} \mathbb{E}(B(x)) &\cong \frac{1}{2} \mu_2(K) \frac{\partial q_{\tau}^{(2)}(x)}{\partial \tau} h_s^2 \\ \mathbb{V}\text{ar}(B(x)) &\cong \frac{1}{2nd_s h_s} \frac{\int K^2(u) du}{f(q_{\tau}(x)|x)g(x)} \end{aligned}$$

if assumptions S1-S5 follow.

From these results, the asymptotic bias of the estimated squared sparsity is given by

$$\text{Bias} \left(\int \widehat{s}_{\tau, d_s, h_s}^2(x) dx \right) \cong \left[\frac{1}{nd_s h_s} \int a(x) dx + d_s^2 \int b(x) dx + h_s^2 \int c(x) dx \right]^2$$

where $a(x)$, $b(x)$ and $c(x)$ are given in (6).

In view of expression (11), the asymptotic variance of sparsity estimator can be decomposed as follows

$$\begin{aligned} \text{Var} \left[\int \widehat{s}_{\tau, d_s, h_s}(x)^2 dx \right] &\cong \text{Var} \left[\int \left(A(x)^2 + B(x)^2 + 2A(x)B(x) \right) dx \right] \\ &= \text{Var} \left[\int B(x)^2 dx \right] + 4 \text{Var} \left[\int A(x)B(x) dx \right] \\ &\quad + 4 \text{Cov} \left[\int B(x)^2 dx, \int A(x)B(x) dx \right]. \end{aligned}$$

Each of the previous terms can be expressed as covariances of U -expressions like that given in (10) evaluated at different points x and quantiles $\tau + d_s$ and $\tau - d_s$. These covariances can be computed (under assumptions S1, S2, S4 and S5) using similar arguments to those employed by Fan et al. (1994), adapting their φ function (given in equation (2.1) on p. 435) to each covariance. Then, the asymptotic variance of sparsity estimator can be approximated as follows

$$\text{Var} \left[\int \widehat{s}_{\tau, d_s, h_s}(x)^2 dx \right] \cong \frac{1}{nd_s} \int d(x) dx + \frac{1}{n^2 d_s^2 h_s} \int e(x) dx$$

where $d(x)$ and $e(x)$ are given in (6). Then, in view of the computed asymptotic bias and variance, expression (5) can be derived.

References

- Abberger K (1998) Cross-validation in nonparametric quantile regression. *Allg Statistisches Archiv* 82:149–161
- Abberger K (2002) Variable data driven bandwidth choice in nonparametric quantile regression. Technical Report
- Bloch DA, Gastwirth JL (1968) On a simple estimate of the reciprocal of the density function. *Ann Math Stat* 39:1083–1085
- Bofinger E (1975) Estimation of a density function using order statistics. *Aust J Stat* 17:1–7
- Conde-Amboage M (2017) Statistical inference in quantile regression models, Ph.D. Thesis. Universidade de Santiago de Compostela, Spain
- El Ghouch A, Genton MG (2012) Local polynomial quantile regression with parametric features. *J Am Stat Assoc* 104:1416–1429
- Fan J, Hu TC, Truong YK (1994) Robust nonparametric function estimation. *Scand J Stat* 21:433–446
- Hall P, Sheather SJ (1988) On the distribution of a studentized quantile. *J R Stat Soc Ser B (Methodol)* 50:381–391
- Jones MC (1991) The roles of ISE and MISE in density estimation. *Stat Probab Lett* 12:51–56
- Jones MC, Yu K (2007) Improved double kernel local linear quantile regression. *Stat Model* 7:377–389
- Koenker R (2005) *Quantile regression*. Cambridge University Press, Cambridge
- Koenker R, Bassett G (1978) Regression quantiles. *Econometrica* 46:33–50
- Mallows CL (1973) Some comments on C_p . *Technometrics* 15:661–675
- Nelder JA, Mead R (1965) A simplex algorithm for function minimization. *Comput J* 7:308–313
- Opsomer JD, Ruppert D (1998) A fully automated bandwidth selection method for fitting additive models. *J Am Stat Assoc* 93:605–618
- Ruppert D, Sheather SJ, Wand MP (1995) An effective bandwidth selector for local least squares regression. *J Am Stat Assoc* 90:1257–1270

- Sánchez-Sellero C, González-Manteiga W, Cao R (1999) Bandwidth selection in density estimation with truncated and censored data. *Ann Inst Stat Math* 51:51–70
- Sheather SJ, Maritz JS (1983) An estimate of the asymptotic standard error of the sample median. *Aust J Stat* 25:109–122
- Siddiqui MM (1960) Distribution of quantiles in samples from a bivariate population. *J Res Natl Bureau Stand* 64:145–150
- Tukey JW (1965) Which part of the sample contains the information. *Proc Natl Acad Sci* 53:127–134
- Venables WN, Ripley BD (1999) *Modern applied statistics with S-PLUS*, 3rd edn. Springer, New York
- Yu K, Jones MC (1998) Local linear quantile regression. *J Am Stat Assoc* 93:228–237
- Yu K, Lu Z (2004) Local linear additive quantile regression. *Scand J Stat* 31:333–346