UNIVERSIDADE DE SANTIAGO DE COMPOSTELA
Departamento de Estatística, Análise Matemática e Optimización

STATISTICAL INFERENCE IN QUANTILE REGRESSION MODELS

Mercedes Conde Amboage

PhD Dissertation
February 2017

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

STATISTICAL INFERENCE IN QUANTILE REGRESSION MODELS

Mercedes Conde Amboage

PhD Dissertation

February 2017

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Don Wenceslao González Manteiga, catedrático del Departamento de Estadística, Análisis Matemático y Optimización de la Universidad de Santiago de Compostela y Don César A. Sánchez Sellero, profesor titular de universidad en el Departamento de Estadística, Análisis Matemático y Optimización de la Universidad de Santiago de Compostela, informan que la memoria titulada

STATISTICAL INFERENCE IN QUANTILE REGRESSION MODELS

fue realizada bajo su dirección por Doña Mercedes Conde Amboage, estimando que la interesada se encuentra en condiciones de optar al grado de Doctor, por lo que solicitan que sea admitida a trámite para su lectura y defensa pública.

Santiago de Compostela, 22 de febrero de 2017

Fdo.: Wenceslao González Manteiga

Fdo.: César A. Sánchez Sellero

Fdo.: Mercedes Conde Amboage

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecimientos

Quiero empezar expresando mi agradecimiento a los profesores Wenceslao González Manteiga y César Sánchez Sellero por su labor de dirección en esta tesis doctoral. A Wenceslao debo agradecerle la oportunidad que me permitió dar mis primeros pasos como investigadora. Espero nunca perder el entusiasmo que me has trasmitido por la Estadística. De César debo destacar todo el tiempo que ha dedicado a este trabajo, todos tus comentarios, correcciones e ideas mejoraron notablemente este documento. Gracias por tu ayuda diaria. Espero haber estado a la altura de las expectativas que depositasteis en mi.

I would also like to express my gratitude to profesor Valentin Patilea for inviting me to visit the Centre de Reserche en Économie et Statistique (CREST) at La Grande École de la Statistique et du Traitement de L'Information (Ensai). Thanks to Valentin to allow me to live through one of the most enriching experiences of my life, not only in an academic level, but also personal.

Quisiera agradecer también a Rosa Crujeiras, Pedro Faraldo y José Manuel Prada su ayuda de cara a afrontar mis primeras experiencias como docente. Para mi ha sido un proceso de aprendizaje muy enriquecedor y agradable.

Agradezco a mis compañeros de la Facultad de Matemáticas, en especial a los del Área de Estadística e Investigación Operativa, por los buenos momentos compartidos a lo largo de estos años. Debo destacar a Maribel por convertirse en una de mis mejores amigas. Gracias, sobre todo, por entenderme.

Finalmente, no puedo dejar de acordarme de mi familia y amigos por su apoyo continuo. En especial a mis padres, que han hecho de mi la persona que hoy soy, os lo debo todo. Gracias a David por ser mejor de lo que nunca imaginé, contigo la vida es más feliz. Gracias a Fátima y a Patricia por ser las hermanas que nunca he tenido. Y también a ti abuelo, que en este momento te estarás emocionando igual que yo.

Santiago de Compostela, Febrero de 2017
Mercedes Conde Amboage

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Preface

Although mean regression achieved its greatest diffusion in the twentieth century, it is very surprising to observe that the ideas of quantile regression were 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 through concepts of quantile regression.

Later, quantile regression methods found a great development since the emergence of Robust Statistics, which reached great expansion in the 1980s. The works of Hampel et al. (1986) or Huber (1981) are good examples in this line. These concepts are still applied nowadays.

Quantile regression is employed when the aim of 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 (that is, do not require stringent assumptions, such as homoscedasticity or normality) and enjoys properties of robustness. Hereby it facilitates a more complete and robust analysis of the information. For all that, quantile regression is a very useful statistical technology for a large diversity of disciplines.

Quantile regression was introduced by Koenker and Bassett (1978) as a weighted absolute residuals fit which allows to extend some properties of classical least squares estimation to quantile regression estimates. Several classical statistical tools and procedures have been adapted to quantile regression scenario over the years. In this line, Koenker (2005) is a good review about quantile regression.

The main purpose of this dissertation is to collect different innovative statistical methods in quantile regression setup related to prediction, estimation and lack-of-fit tests. In this sense, this manuscript is organized as follows:

Chapter 1: Introduction. Along this chapter, a brief introduction to quantile regression models will be presented. We will start establishing the concept of sample quantiles and then, the idea of quantile regression introduced by Koenker and Bassett (1978) will be shown as well as how to compute the associated regression coefficients. Furthermore, the main characteristics associated with parametric quantile regression, as robustness or the importance of the sparsity function, are introduced.

Chapter 2: Predicting using quantile regression models. Quantile regression models are especially useful in applications where extremes are important, such as environmental studies where upper quantiles of pollution levels are critical from a public health perspective. The main goal of this chapter is to propose a new method in order to construct prediction intervals based on median regression and a bootstrap procedure to approximate the prediction error distribution. This new method rendered better coverage results for NO_x concentration measured in the surroundings of the power plant of As Pontes (Spain), compared to prediction intervals available in the literature. Moreover, a Monte Carlo simulation study shows the good properties of the proposed method.

This chapter is mainly based on Conde-Amboage et al. (2016).

Chapter 3: A plug-in bandwidth selector for nonparametric quantile regression. The aim of this chapter is to study the problem of bandwidth selection in local linear quantile regression. As in nonparametric least squares regression, bandwidth selection plays a very important role in local quantile estimation as well. Whereas an abundance of papers treating bandwidth choice in nonparametric mean regression may be found in the literature, this topic is less frequently discussed in local quantile estimation. Some contributions to bandwidth selection focus on plug-in methods based on several restrictive assumptions. Cross-validation techniques may also be found. Along this chapter, a new plug-in rule will be proposed based on nonparametric estimations of the sparsity and the curvature. A complete study of the mean squared error of these estimators will be presented along this chapter.

This chapter is manly based on Conde-Amboage and Sánchez-Sellero (2017).

Chapter 4: A lack-of-fit test for quantile regression models with high-dimensional covariates. Along this chapter a new lack-of-fit test for quantile regression models with multiple covariates will be presented. The test is based on the cumulative sum of residuals with respect to unidimensional linear projections of the covariates. The test is then adapting the ideas of Escanciano (2006) to cope with high-dimensional covariates, to the test proposed by He and Zhu (2003) that in turn extends the ideas of Stute (1997) to the quantile regression setup. It will be shown that the empirical process associated with the test statistic converges to a Gaussian process under the null hypothesis. Moreover, it is stated that the proposed test statistic is consistent and can detect local alternatives of order $n^{-1/2}$ from the null hypothesis. On the other hand, to approximate the critical values of the test, a wild bootstrap mechanism is used, which is similar to that proposed by Feng et al. (2011). In addition, an interesting application to real data will be presented.

This chapter is mainly based on Conde-Amboage et al. (2015).

Chapter 5: A lack-of-fit test for quantile regression models using logistic regression. The error associated with any quantile regression model verify that its conditional τ-th quantile is equal to zero. Bearing this property in mind, along this chapter a new test will be proposed in order to check the goodness-of-fit of a quantile regression model. Then, the test is based on the relation between the residuals associated with a quantile regression model and the logistic regression context, following the idea proposed by Redden et al. (2004). Furthermore, in the multivariate context, projections of the covariates of the quantile regression model have been used as predictors in the logistic model in order to avoid the well-known curse of the dimensionality. Furthermore, in order to calibrate the test, a wild bootstrap procedure is used, following the ideas developed in Chapter 4. A simulation study and an application to real data were carried out that show the good properties of the new test versus other tests available in the literature.

This chapter is mainly based on Conde Amboage et al. (2016).

We also enclose a summary of this dissertation thesis in Spanish language and a notation index.

This work has been supported by FPU grant AP2012-5047 from Spanish Ministry of Education, and graduate grant for research stays from Fundación Barrié. It is also acknowledged the support from the Spanish Ministry of Economy and Competitiveness, through grant numbers MTM2008-03010 and MTM2013-41383P, which include support from the European Regional Development Fund (ERDF). Support from the IAP network P7/06 StUDyS of the Belgian Government (Belgian Science Policy) is also acknowledged.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Contents

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 1

Introduction

Contents

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

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 chapter an introduction to quantile regression methods is developed. In this sense, in Section 1.1 the concept of sample quantile is introduced as well as the optimization problem associated with its estimation. Later, these ideas are extended to quantile regression estimator in Section 1.2 following the idea of Koenker and Bassett (1978). When dealing with the asymptotic distribution of quantile regression estimator, the so-called sparsity function comes out and will play an important role in quantile regression setup. This function will be studied in Section 1.3. Finally, in Section 1.4 robustness of quantile regression methods versus least squares regression is briefly analysed.

1.1 Sample quantiles

We start stating the definition of τ-th quantile. To this aim, it will be necessary to remember the concept of cumulative distribution function.

Definition 1.1. Given a random variable $X : \Omega \to \mathbb{R}$, defined in a sample space Ω associated with a random experiment, its distribution is characterized by the **cumulative distribution function** that is defined by

$$F_X(x) = \mathbb{P}_X\big(X \le x\big) \qquad \forall x \in \mathbb{R}.$$

Then, if X is a discrete variable, the distribution function is given by

$$F_X(x) = \mathbb{P}_X\big(X \le x\big) = \sum_{x_i \le x} \mathbb{P}_X\big(X = x_i\big)$$

where $x_i \le x$ represents all the values that could be taken by the variable X that are equal or smaller than x. On the other hand, if X is a continuous variable then

$$F_X(x) = \mathbb{P}_X\big(X \le x\big) = \int_{-\infty}^{x} f_X(z)dz$$

for each x where $f_X : \mathbb{R} \to \mathbb{R}$ is the well-known **density function**.

Now, we are able to present the idea of quantile. Its definition is given below.

Definition 1.2. 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

$$\mathbb{P}_X(X \le c_\tau) \ge \tau$$

$$\mathbb{P}_X(X \ge c_\tau) \ge 1 - \tau$$

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, that is, the greatest lower bound of the subset. The infimum is a criterion used to choose a simple quantile when the first definition provides more than one solution.

Moreover, the concept of τ-th quantile can be extended to the regression context as given below.

Definition 1.3. Given a regression setup, suppose that Y represents the response variable and X is the d-dimensional explanatory variable. Moreover, $F_Y(y|X=x) = \mathbb{P}(Y \leq y|X=x)$ denotes the conditional cumulative distribution function of Y given $X=x$. Then, the τ-th **conditional quantile** of Y given $X=x$, is defined as

$$q_\tau(Y|X=x) = \inf\{y : F_Y(y|X=x) \geq \tau\}.$$

1.1.1 Quantile loss function

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 1.1 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 a differentiable function so that standard numerical algorithms do not work. Because of this reason, most of the theory developed for mean estimation can not be applied in this context.

Note that it could seem more natural to define the sample quantiles in terms of order statistics, that is, the sample values placed in ascending order. However, the fact of using the quantile loss function introduced previously provides a natural transition to regression context as it will be shown in the following section.

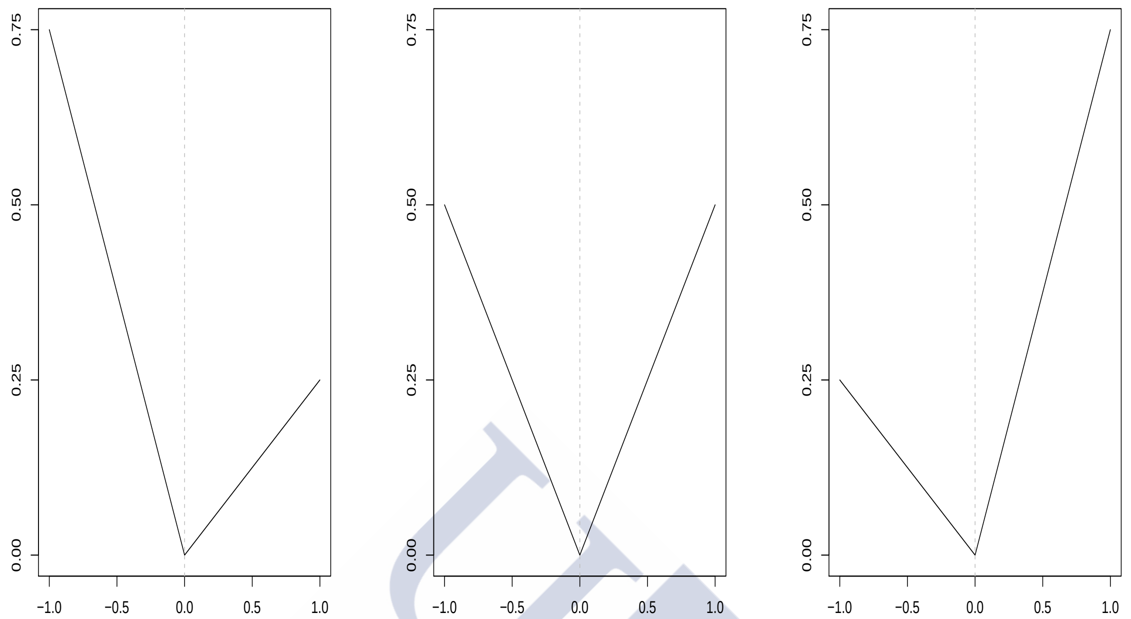
Thereupon, for each $\tau \in (0,1)$ we are going to show that the minimizer coincides with the τ-th quantile that has been denoted by c_τ. Indeed, it will be necessary to minimize in x

$$\mathbb{E}\Big[\rho_\tau(X-x)\Big] = (\tau-1)\int_{-\infty}^{x}(y-x)dF_X(y) + \tau\int_{x}^{\infty}(y-x)dF_X(y)$$

and differentiating with respect to x, it follows that

$$0 = (1-\tau)\int_{-\infty}^{x} dF_X(y) - \tau\int_{x}^{\infty} dF_X(y) = (1-\tau)\,F_X(x) - \tau\,(1-F_X(x)) = F_X(x) - \tau.$$

Since F_X is monotone, any element of $\{x : F_X(x) \geq \tau\}$ minimizes expected loss. When the solution is unique then $c_\tau = F_X^{-1}(\tau)$ and otherwise, we have an interval of τ-th quantiles from which the smallest element must be chosen.



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

Figure 1.1: 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).

1.1.2 Optimization problem

In practice, the cumulative distribution function F is replaced by the empirical distribution function that is given by

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(X_i \leq x)$$

where $\mathcal{X} = \{X_1, \cdots, X_n\}$ represents a random sample of the variable X. Then, the sample quantiles can be computed as

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

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

The problem of finding the τ-th sample quantile may be reformulated as a linear problem by introducing $2n$ artificial variables $\{u_i, v_i \text{ with } i = 1, \cdots, n\}$ representing the positive and negative parts of the vector of residuals. This yields the following optimization problem:

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

where 1_n denotes an n-dimensional vector of ones, $X = (X_1, \cdots, X_n)$, $u = (u_1, \cdots, u_n)$ and $v = (v_1, \cdots, v_n)$. Clearly, we are minimizing a linear function on a polyhedral constraint set

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

consisting of the intersection of the $(2n+1)$-dimensional hyperplane determined by the linear equality constraints and the set $\mathbb{R} \times \mathbb{R}_+^{2n}$.

A more complete explanation about the optimization problem associated with the sample quantiles can be found in Section 1.1.2 of Davino et al. (2014). 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 Hyndman and Fan (1996).

1.1.3 Exact and asymptotic distribution

Based on the fact that $\mathbb{P}(\widehat{c}_\tau > c) = \mathbb{P}(\text{Bi}(n, F_X(c)) < n\tau)$ where Bi represents a binomial distribution and using the incomplete beta function, the exact distribution of sample quantiles can be expressed as follows

$$G(c) = n\binom{n-1}{m-1}\int_0^{F_X(c)} t^{m-1}(1-t)^{n-m}dt$$

where m represents the smallest integer equal or bigger that $n\tau$ and $\binom{a}{b}$ denotes a binomial coefficient. Consequently, sample quantile density is defined as follows

$$g(c) = n\binom{n-1}{m-1}F_X(c)^{m-1}(1-F_X(c))^{n-m}f_X(c).$$

Furthermore, the asymptotic distribution of $\widehat{c}_\tau$ can be derived as a consequence of Lindeberg's central limit theorem. This result is gathered in the following theorem:

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 a distribution convergence.

The proof of Theorem 1.1 is detailed in several classical Inference Statistical works, see for instance Chatterjee (2011). Moreover, some interesting computational aspects about how to estimate asymptotic variance of sample quantiles can be shown in Section 3.5 of Wilcox (2011).

1.2 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:

$$Y_i = \theta_\tau' P_i + \varepsilon_i$$

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. This implies that the proportion of negative errors is expected to be τ, which is equivalent to the proportion of observations below the regression function being equal to τ. Note that it is analogous to assuming that $\mathbb{E}(\varepsilon_i|X = X_i) = 0$ in the classical least squares context.

Let us remember that the τ-th sample quantile of the response variable Y is given by

$$\widehat{c}_\tau = \arg\frac{1}{n}\min_{c\in\mathbb{R}}\sum_{i=1}^{n}\rho_\tau(Y_i - c).$$

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

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

where $P_i = (1, X_i)$. This idea has been introduced by Koenker and Bassett (1978) and subsequently El Bantli and Hallin (1999) demonstrated the consistency of the quantile regression estimator.

This kind of models will be applied to an environmental problem along Chapter 2. Moreover, a detail comparison between least squares regression and quantile regression will be introduced in Section 2.2.1.

1.2.1 Optimization problem

Following the ideas described in Section 1.1.2, the parameter $\widehat{\theta}_\tau$ can be obtained as the solution of the following linear optimization problem:

$$\min_{(\theta,u,v)\in\mathbb{R}^{d+1}\times\mathbb{R}^{2n}_{+}}\left\{\tau 1'_n u + (1-\tau)1'_n v \ : \mathbb{X}\theta + u - v = Y\right\} \tag{1.1}$$

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. Again, the residual vector $Y - \mathbb{X}\theta$ has been split into its positive and negative parts (u and v respectively). So we are minimizing a linear function on a polyhedral constraint set, and the solutions of this problem are the estimation of the coefficients associated with a quantile regression model that have been denoted by $\widehat{\theta}_\tau$.

The calculus of the quantile regression parameter as a linear optimization problem will be crucial because it gives place to different methods in order to compute $\widehat{\theta}_\tau$. In this line, Barrodale and Roberts (1973) 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 Barrodale and Roberts (1973)'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. Subsequently, Koenker and D'Orey (1987) extended this development to each quantile $0 < \tau < 1$.

There exist other possibilities in order to deal with the optimization problem given by (1.1) as Portnoy and Koenker (1997) that proposed a Frisch-Newton interior point method and an interior point method with preprocessing (methods recommended for larger sample sizes) or Koenker and Ng (2003) that studied a sparse regression quantile fitting (method recommended for sparse data).

1.2.2 Asymptotic distribution

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 $\psi_\tau(r) = \tau\mathbb{I}(r > 0) + (\tau - 1)\mathbb{I}(r < 0)$ denotes the derivative of the quantile loss function ρ_τ (not defined at zero), Bahadur (1966) established that

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

if the following regularity conditions are verified:

Condition C1: The distribution functions of Y_i given X_i, F_i, have continuous densities f_i that are bounded away from 0 and uniformly bounded away from ∞ in a neighbourhood of the conditional quantiles $c_i(\tau)$ with $i = 1, \ldots, n$. In addition, the first derivative of f_i is uniformly bounded in a neighbourhood of $c_i(\tau)$ with $i = 1, \ldots, n$.

Condition C2: $\max_{i=1,\ldots,n} \|X_i\| = O\left(n^{1/4}(\log n)^{-1/2}\right)$.

Condition C3: $n^{-1}\sum_{i=1}^{n} \|X_i\|^4 \leq B$ for some finite constant B.

Condition C4: There exist positive definite matrices D_0 and D_1 such that

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} P_i P_i' = D_0$$
$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_i(c_i(\tau)) P_i P_i' = D_1.$$

Differently from least squares estimator, the quantile estimator distribution is not generally known even under error normality. Koenker (2005) showed the following result about the asymptotic distribution of quantile regression estimators.

Theorem 1.2. *Let us consider a linear model*

$$Y_i = \theta_\tau' P_i + \varepsilon_i \qquad \textit{with } i = 1, \ldots, n$$

where $P_i = (1, X_i)$ *and the errors verify that* $\mathbb{P}(\varepsilon_i \leq 0 \mid X = X_i) = \tau$. *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

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

In view of Theorem 1.2, it is clear that the inverse of the conditional density of the response variable evaluated at the quantile of interest will play an important role. This function, called **sparsity function**, will be studied in Section 1.3.

1.2.3 Properties of quantile regression

Nowadays, quantile regression methods enjoy a great reputation because of their good properties. A complete description of quantile regression models can be found in Koenker (2005). Hereunder, some of these interesting properties are enumerate:

– Quantile regression allows us to study the impact of the explanatory variables on different quantiles of the response distribution, and thus provides a complete picture of the relationship between Y and X.

– Robustness to outliers in y observations (detailed in Section 1.4).

– It adapts to situations under general conditions of the error distribution, that is, quantile regression models are flexible with respect to these conditions. Theoretical results about quantile regression under heteroscedasticity and not-necessarily normal error distribution can be found in Koenker (2005), page 120.

– Some interesting properties are verified:

  1. **Equivariance property:** quantiles are equivariant to monotone transformations. That is, if h is a nondecreasing function on $\mathbb{R}$ then for any random variable Y it follows that
  $$q_\tau(h(Y)) = h(q_\tau(Y)).$$
  This property allows to obtain an interpretation of the coefficients associated with quantile regression, that is,
  $$\frac{\partial q_\tau(Y|X=x)}{\partial x_j} = \frac{\partial h^{-1}(\theta_\tau'(1,x))}{\partial x_j}.$$

2. **Interpolation:** linear quantile fit interpolates $(d+1)$ observations. The proof of this statement can be found on page 33 of Koenker (2005).

3. In a non-degenerate situation, the proportion of negative residuals is approximately τ and the proportion of positive residuals is approximately $(1-\tau)$. This property will be very useful in order to define the lack-of-fit test that is presented in Chapter 5 of this manuscript.

4. **Quantile crossing:** If we consider several quantile regression scenarios associated with different values of τ, we can find some situations in which quantile functions cross one to another, which is a really undesirable situation. Anyway, it is important to emphasize that such crossing is typically confined to outlying regions of the design space. Theorem 2.5 (page 56) of Koenker (2005) shows that at the centroid of the design (that is, the mean) the estimated conditional quantile function is monotone in τ.

Remark 1.1. All the methodology developed along this section can be extended to a nonparametric context. In this line, Chaudhuri (1991a) and Chaudhuri (1991b) can be considered as seminal works. This approach will be deeply studied in Chapter 3 where a new plug-in selector for local linear quantile regression will be presented.

1.3 The sparsity function

In view of the asymptotic distribution associated with the parametric quantile regression estimator, it will be necessary to estimate the inverse of the density function evaluated at the quantile of interest. In the parametric quantile regression model, 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 an 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. Tukey (1965) named **sparsity function** to the inverse of the density function evaluated at the quantile, that is given by

$$s(\tau) = \frac{1}{f_Y(F_Y^{-1}(\tau))}.$$

There is an extensive literature on sparsity estimation. We will review the most relevant contributions. Let us observe that the sparsity function is simply the derivative of the quantile function, that is,

$$\frac{\partial}{\partial t} F_Y^{-1}(t) = \frac{1}{f_Y(F_Y^{-1}(t))} = s(t).$$

Therefore, just as differentiating the distribution function F_Y yields the density function f_Y, differentiating the quantile function F_Y^{-1} yields the sparsity function. Given $\mathcal{Y} = \{Y_1, \cdots, Y_n\}$ a random sample of the variable Y, Siddiqui (1960) proposed to estimate the sparsity by a simple difference quotient of the empirical quantile function, that is,

$$\widehat{s}(t) = \frac{\widehat{F}_n^{-1}(t+h) - \widehat{F}_n^{-1}(t-h)}{2h} = \frac{Y_{[n(\tau+h)]} - Y_{[n(\tau-h)]}}{2h} \tag{1.2}$$

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, $[n(\tau + h)]$ and $[n(\tau \pm h)]$ are neighbouring orders to τ where $[a]$ denotes the integer part of a. Later, Bloch and Gastwirth (1968) showed that the value of the smoothing parameter that minimizes the asymptotic mean squared error of (1.2) is of order $n^{-1/5}$.

Bofinger (1975) proposed a bandwidth selector in order to compute the nonparametric estimator of the sparsity. In addition, she proved that the bandwidth

$$h_{\text{B}} = \sqrt[5]{\frac{4.5 s(\tau)^2}{s^{(2)}(\tau)^2}}\, n^{-1/5} \tag{1.3}$$

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

It is clear that in order to compute h_{B} it will be necessary to compute estimators of $s(\tau)$ and $s^{(2)}(\tau)$. Bofinger (1975) propose to estimate $s(\tau)$ in (1.3) with some non-optimal smoothing parameter. As for $s^{(2)}(\tau)$, she proposed the following estimator:

$$\widehat{s}_\tau^{(2)} = \frac{1}{2\delta^3} \left(Y_{[[n\tau]+2m]} - 2Y_{[[n\tau]+m]} + 2Y_{[[n\tau]-m]} - Y_{[[n\tau]-2m]} \right)$$

where $m = [n\delta]$ and $\delta = \Omega(n^{-\zeta})$ with $\zeta < 1/5$ where the symbol Ω represents "the same order as", that is, it will be verified that $\delta \to 0$ and $n\delta^5 \to \infty$ as $n \to \infty$. This ideas will be very useful along Chapter 3.

On the other hand, Hall and Sheather (1988) examined the effect that the selection of the smoothing parameter has on the level error of tests or confidence intervals 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 Bofinger (1975)'s proposal. Bearing this idea in mind, Hall and Sheather (1988) proposed the following smoothing parameter:

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

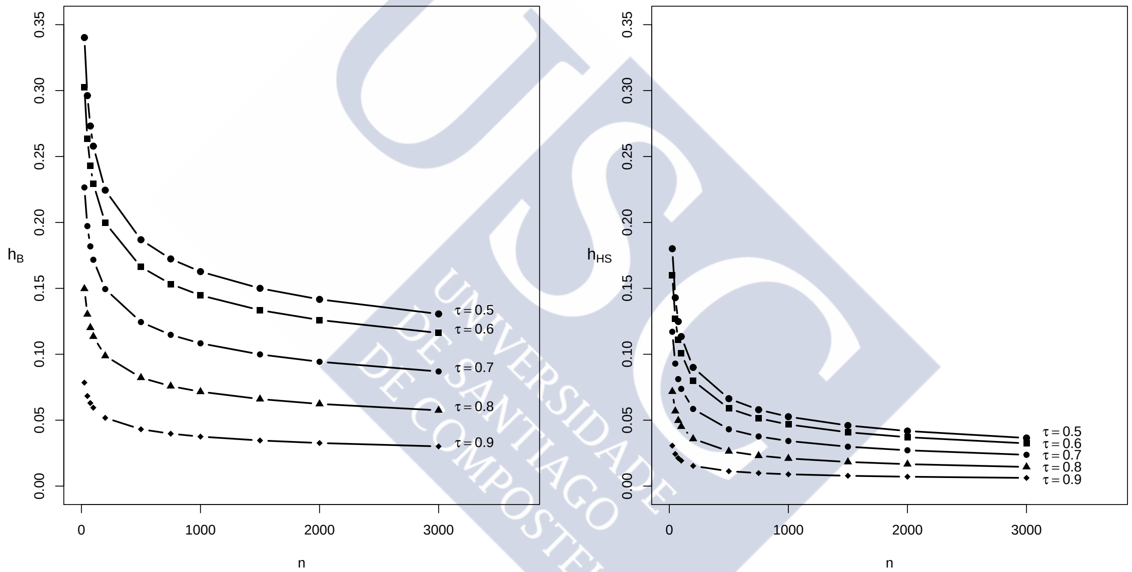
where

$$S_{d,n} = \frac{n}{2d}\left(Y_{[t+d]} - Y_{[t-d]} \right)$$

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

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

Let us consider a simple simulation scenario in order to compare the bandwidth selectors that have been proposed to build a nonparametric estimation of the sparsity function. We have simulated values from a standard Gaussian distribution associated with different sample sizes (parameter denoted by n). Figure 1.2 shows the values of the theoretical bandwidth selector proposed by Bofinger (1975) (Part a) and Hall and Sheather (1988) (Part b) associated with different values of the sample size and the τ-th quantile of interest. In view of Figure 1.2 the differences between both selectors are clear: the bandwidth selector proposed by Bofinger (1975) is bigger than Hall and Sheather (1988)'s proposal if we consider the same sample from a standard Gaussian distribution. The differences are the expected ones because of the optimal criterion is different in each case.



(a) Bofinger (1975)'s selector

(b) Hall and Sheather (1988)'s selector

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figure 1.2: Representation of theoretical bandwidth selectors proposed by Bofinger (1975) (Part (a)) and Hall and Sheather (1988) (Part (b)) in order to estimate the sparsity function following equation (1.2).

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

$$Y_i = \theta'_\tau P_i + \varepsilon_i \qquad i \in \{1, \ldots, n\}$$

where ε_i has conditional τ-th quantile equal to zero and $P_i = (1, X_i)$.

In this situation, Hendricks and Koenker (1992) 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 Hall and Sheather (1988) and $\widehat{\theta}_{\tau^+}$ and $\widehat{\theta}_{\tau^-}$ represent 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, quantiles may cross so that upper quantiles may be estimated to be smaller than lower quantiles. A modified estimator to account for this issue could be

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

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

Hendricks and Koenker (1992)'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. This method could be adapted to a local linear context, but in this case it will suffer from two biases, both controlled by the parameter h_{HS}: bias in the quantile regression estimation and bias in the sparsity estimation itself.

The study of the sparsity function in a general regression context has not been thoroughly analysed in the literature. In Chapter 3 of this manuscript this problem is addressed and the mean squared error of a newly proposed sparsity estimator is obtained.

Remark 1.2. There exist other ideas in order to estimate the asymptotic variance of the estimator $\widehat{\theta}_\tau$ as the idea of Pollard (1991) that proposed a kernel estimator of $D_1(\tau)$ that is given by

$$\widehat{D}_1(\tau) = n\hat{h}_{\text{P}} \sum_{i=1}^{n} K\left(\frac{r_i}{h_{\text{P}}}\right) P_i P_i'$$

where $r_i = Y_i - \widehat{\theta}_\tau' P_i$ and h_{P} is a bandwidth parameter satisfying $h_{\text{P}} \to 0$ and $n^{1/2} h_{\text{P}} \to \infty$ when $n \to \infty$.

1.4 Robustness

Outliers occur frequently in real data, and can make one to misinterpret patterns in plots, and may also cause that model fails to capture the important characteristics of the data. Deleting outliers from the regression model can sometimes give completely different results.

Accordingly, robust methods have been created to make outliers have much less influence on the final estimates. In this line, it is well-known the major robustness of quantile regression versus classical mean regression. We are going to focus on the influence function, introduced by Hampel (1974).

To show this, the **influence function** describes the effect of an anomalous sample point over a certain estimator. More formally, an estimator $\widehat{\gamma}$ may be seen as a functional of a distribution F, that is, $\widehat{\gamma}(F)$. We may consider contaminating F by replacing a small amount of mass t from F by an equivalent mass concentrated at y, allowing us to write the contaminated distribution function as

$$F_t = (1-t)F + t\delta_y$$

where δ_y denotes the distribution function that assigns mass 1 to the point y. Then, the influence function can be defined by

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

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

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

where sgn represents the sign function, that is given by

$$\operatorname{sgn}(u) = \begin{cases} -1 & \text{if } u < 0 \\ 0 & \text{if } u = 0 \\ 1 & \text{if } u > 0 \end{cases}$$

There is a dramatic difference between the two influence functions. In the case of the mean, the influence of contaminating F at y is simply proportional to y, that is, a point y sufficiently far from $\mu(F)$ can take the mean arbitrarily far away from its initial value at F. In contrast, the influence of contamination at y on the median is bounded by the sparsity at the median.

Figure 1.3 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. The boundedness of the quantile influence function is obviously maintained, provided that the sparsity at τ is finite.

Now, we are going to move to a regression context. Let us assume that $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ represents a random sample of two variables $(X, Y) \in \mathbb{R}^{d+1}$ whose relationship can be described by

$$Y_i = \theta' P_i + \varepsilon_i$$

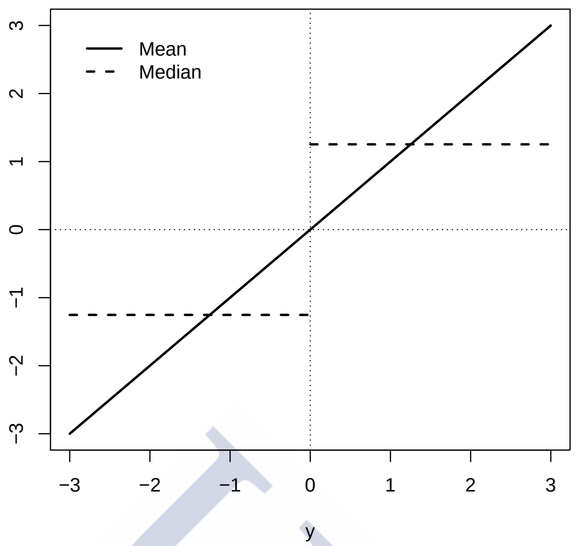


Figure 1.3: Influence function associated with mean and median estimators, where F is a standard Gaussian distribution.

where $P_i = (1, X_i) = (1, X_{i,1}, \cdots, X_{i,d})$. Moreover, the errors $\varepsilon_1, \ldots, \varepsilon_n$ are supposed to be uncorrelated and with common variance σ^2.

The least squares estimator is given by

$$\widehat{\theta}_{LS} = \arg\min_{\theta} \sum_{i=1}^{n} \left(Y_i - \theta' P_i\right)^2 . \tag{1.4}$$

Cook and Weisberg (1982) (page 106) showed that the influence function associated with the least squares estimator 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. It should be noticed that the notation $\mathbb{E}(\mathbb{X}\mathbb{X}')^{-1}$ corresponds with randomized design while for fixed design we should write $(\mathbb{X}'\mathbb{X})^{-1}$, where $\mathbb{X}$ represents the design matrix.

It is quite interesting that, 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 position 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.

Consider now the quantile regression estimator, that is given by

$$\widehat{\theta}_\tau = \arg\min_{\theta} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \theta' P_i\right) \tag{1.5}$$

where ρ_τ represents the quantile loss function associated with τ. In this case, 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).

In order to show the robustness of quantile regression against vertical deviation we are going to present a simple simulation example. Let us generate values of the following regression model:

$$\text{Model 1.1:} \quad Y = 1 + 0.5X + \varepsilon \tag{1.6}$$

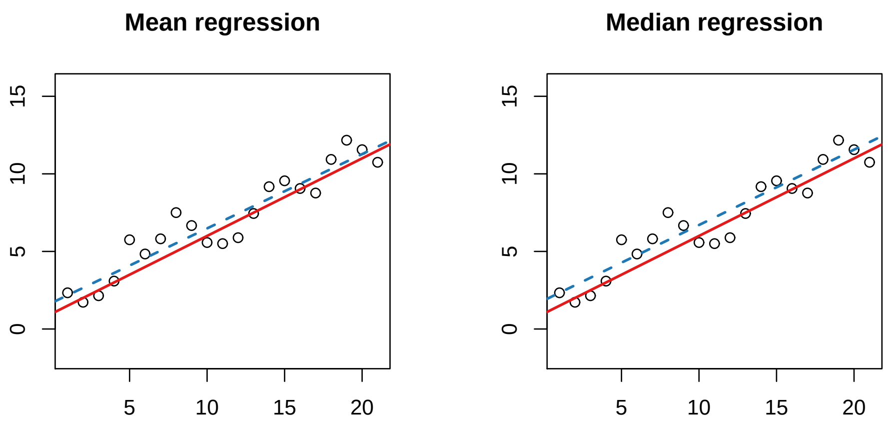
where ε follows a standard Gaussian distribution and X is a grid of equally spaced values from 1 to the sample size n (in this case $n = 21$). Figure 1.4 shows the fitted least squares and median regression models for three different scenarios:

- Part (a) shows the fitted models associated with the original sample.
- Part (b) shows the fitted models associated with the original sample plus a moderate perturbation of one possible value of the response variable represented by the solid circle.
- Part (c) shows the fitted models associated with the original sample plus a big perturbation of one possible value of the response variable represented by the solid circle.

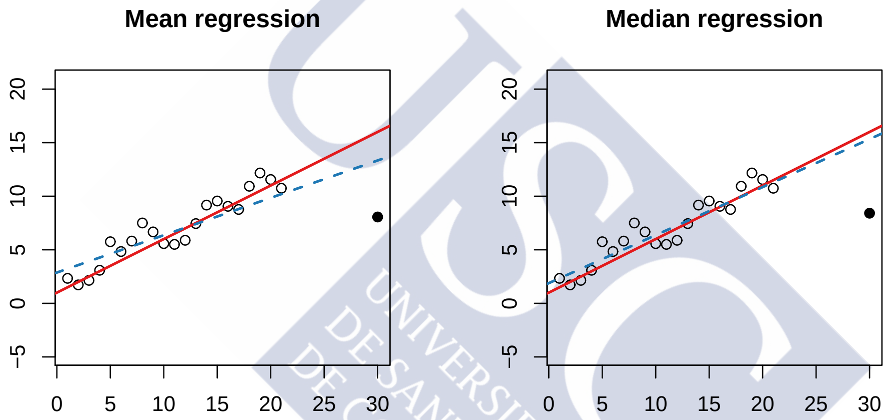
In view of Figure 1.4 we can observe the robustness of median regression against deviations related to the response variable while it is clear that the least squares estimator can completely change due to the perturbation included in the original sample.

Furthermore, in order to control both factors of the influence function, it should be necessary to introduce **generalized M-estimators** that have been studied by Maronna and Yohai (1981). Moreover, other kinds of robust estimators have been considered such as least median of squares regression proposed by Rousseeuw (1984) or regression depth proposed by Rousseeuw and Hubert (1999).

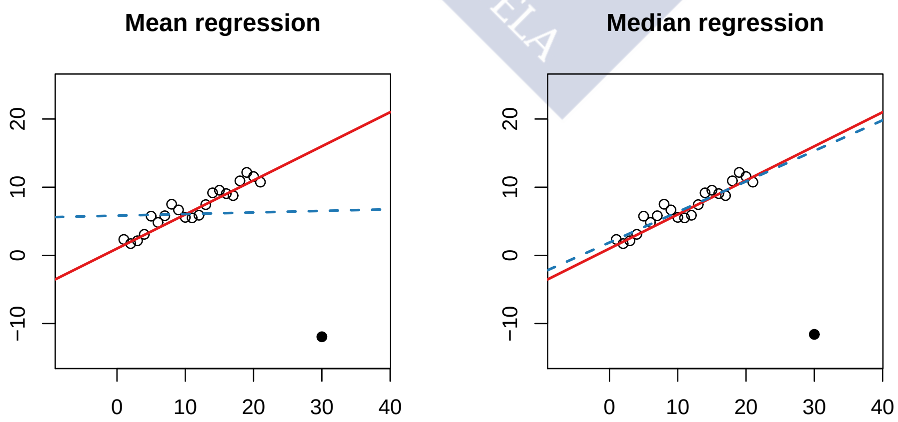
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Original data



(b) Original data plus a moderate peturbation



(c) Original data plus a big peturbation

Figure 1.4: Representation of the fitted least squares and median regression models for three different scenarios that have been considered: Original sample (Part a), Original sample plus a moderate perturbation (Part b) and Original sample plus a big perturbation (Part c). The red line represents the true model while the blue line represents the fitted model associated which each scenario.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Remark 1.3. It should be noticed that we could consider other kinds of robustness measures such as **the breakdown point** introduced by Hampel (1971). This concept is defined as the smaller proportion of atypical data in the sample needed to perturb the estimator as much as we want.

Note that the breakdown point associated with mean estimator is zero due to the fact that only one atypical point is sufficient to change it completely. On the other side, the breakdown point related to median estimator is 0.5 (the biggest possible), that is, it will be necessary to change half of the original sample.

Moreover, in view of Figure 1.4, the breakdown point associated with mean regression estimator is zero, but the one related to median regression is zero too, due to leverage problems.

Chapter 2

Predicting using quantile regression models

Contents

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Quantile regression methods are evaluated for computing predictions and prediction intervals of NO_x concentrations measured in the vicinity of the power plant in As Pontes (Spain) along this chapter. For these data, smaller prediction errors were obtained using methods based on median regression compared with mean regression. A new method to construct prediction intervals involving median regression and bootstrapping the prediction error is proposed. This new method provides better coverage for NO_x data compared with classical and bootstrap prediction intervals based on mean regression, as well as simpler prediction intervals based on quantile regression. A simulation study illustrates the features of this proposed method that leads to a better performance for obtaining prediction intervals for these particular NO_x concentration data, as well as for any other environmental dataset that does not meet assumptions of homoscedasticity and normality of the error distribution.

2.1 Introduction

As we have mentioned along Chapter 1, quantile regression models were introduced by Koenker and Bassett (1978), with the purpose of estimating certain quantiles of a response variable conditional to values of its predictors. In this way, a more complete description of the conditional distribution can be given, where the central and best known quantile is the median, but lower or upper quantiles are also taken into account. Thus, these models describe the effects of the predictors not only on the central values of the response variable, but also on its lower or upper range of values. Moreover, quantile regression is estimated in a more robust manner than common mean regression models, and does not require stringent assumptions to be satisfied, such as homoscedasticity and normality of the error distribution. For all that, quantile regression is a very useful statistical technology in diverse areas of application, like Ecology, Economy or Medicine.

For instance, quantile regression was successfully applied to environmental data by several authors in recent years. Sousa et al. (2009) made use of quantile regression to predict ozone concentrations in Oporto, Northern Portugal. Salama (2005) showed that median regression analysis is more useful for detecting relationships between environmental performance and corporate financial performance than ordinary least squares regression. A hierarchical Bayesian spatial quantile regression model was proposed by Fontanella et al. (2015) to analyze indoor radon concentrations. Cade and Noon (2003) provide a nice review of applications of quantile regression. Quantile regression has also proven to be very useful for obtaining prediction intervals. Meinshausen (2006) and Mayr et al. (2012) made use of estimated quantiles to define the endpoints of prediction intervals, while Zhou and Portnoy (1996) proposed a relatively simple correction to prediction intervals to improve their coverage.

In this chapter, quantile regression is shown to be more accurate than regression based on least squares methods for obtaining predictions and prediction intervals of NO_x concentrations measured in the vicinity of the power plant of As Pontes (Spain).

The power plant at As Pontes (A Coruña, Spain) is an important facility of Endesa Generación S.A. Figure 2.1 shows a picture of the power plant, and its geographical location within Europe. The plant comprises a thermal power station and a combined cycle power station. Its activity releases NO_x in quantities that need to be monitored for both legal and ecological reasons. European legislation imposes threshold levels on ambient NO_x

concentrations to protect human and environmental health. In addition, the location of the power plant near to natural enclaves of high ecological value requires special care to be exercised to mitigate pollution of the local environment. As a consequence, the power plant possesses several systems of pollution control. In particular, it has a 'Network of Vigilance of Atmospheric Quality', comprising seven automatic analysers for sulfur dioxide (SO_2), oxides of nitrogen (NO_x), particles in suspension, temperature, and oxygen, located in several positions around the power plant. Such measurements are used to control what happens in the neighbourhood of the plant in real time and to make modifications, if necessary, to prevent air quality level episodes that exceed the limits established by the air quality legislation. Moreover, a meteorological station also provides information to help assess and predict contamination. Predictions of 30 minutes in advance are necessary, because it takes about 30 minutes for countermeasures to be implemented at the power plant, and to arrange for other contributors to the national power grid to compensate these effects on energy production.

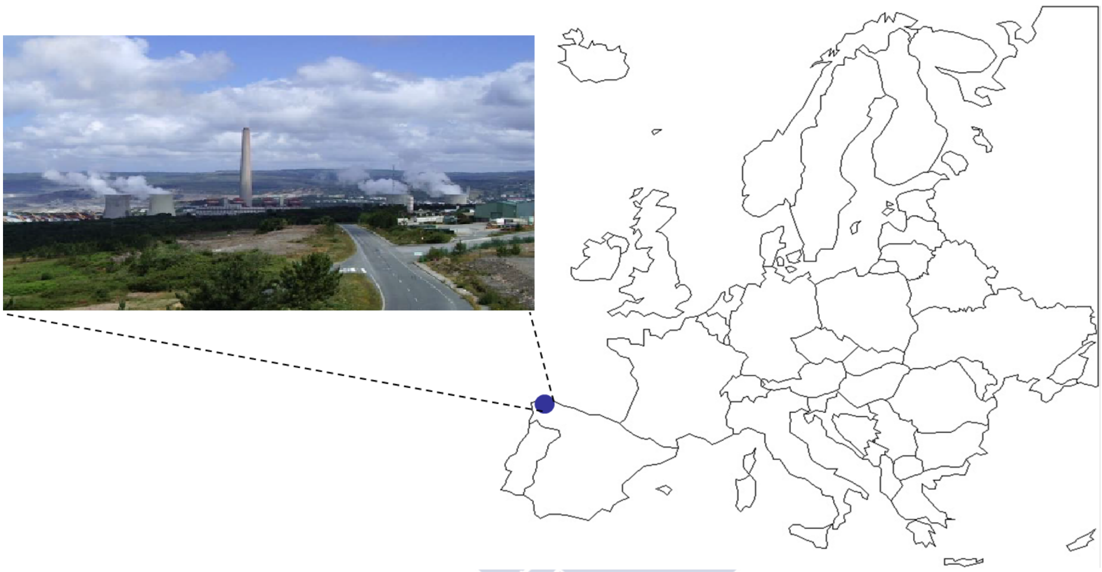


Figure 2.1: Picture of the As Pontes power plant and its geographical location within Europe.

Since 1992, the power plant in As Pontes has had an air pollution control help system, which was developed by the Department of Statistics and Operations Research of the Universidade de Santiago de Compostela in cooperation with the Environment Department of the power plant. Throughout the years, several methods have been proposed to predict future pollution in the surrounding areas of the power plant at As Pontes. García-Jurado et al. (1995) proposed a semiparametric prediction system for a time series that generalizes the Box–Jenkins model. Prada-Sánchez and Febrero-Bande (1997) introduced the concept of an historical matrix, which summarizes the information on past pollution events in a semiparametric model. Prada-Sánchez et al. (2000) considered partially linear models within an environmental context, which allowed the user to introduce additional information as meteorological variables. Fernández-Castro et al. (2003) used neural network models to predict the evolution of certain pollutant elements. Fernández-Castro et al. (2005) and Fernández-Castro and González-Manteiga (2008) employed several functional techniques for predicting sulfur dioxide levels. Roca-Pardiñas et al. (2004) and Roca-Pardiñas et al. (2005) used a generalized additive model with an unknown link function to predict the binary time series defined using a SO_2 concentration threshold. Along similar lines, a study of correlations between various contaminants around four coal-fired power plants in Greece was provided by Nanos et al. (2015) and regression modelling of atmospheric NO_x concentration in urban London can be found in Shi and Harrison (1997).

Clearly, most works rely on least squares methods and prediction of mean pollutant levels. In contrast, the purpose of this chapter is to provide prediction methods for NO_x concentration using quantile regression models. Particularly, a new method for computing prediction intervals is proposed here, based on quantile regression estimation and bootstrap approximation of the prediction error. Its performance is evaluated using real data. In addition, simulations are provided to illustrate features of this new method that make it suitable for other environmental datasets.

This chapter is organized as follows. The proposed methods are described in Section 2.2. A simulation study to compare several prediction intervals is provided in Section 2.3. In Section 2.4, each of the methods is applied to the real data example of NO_x concentration. Finally, the main conclusions are given in Section 2.5.

2.2 Prediction techniques based on quantile regression methods

Given a random sample $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ of the variables $X = (X^{(1)}, \cdots, X^{(d)}) \in \mathbb{R}^d$ and $Y \in \mathbb{R}$, let us consider a regression model

$$Y_i = \theta' P_i + \varepsilon_i \qquad (2.1)$$

where $\theta = (\theta_0, \theta_1, \cdots, \theta_d) \in \mathbb{R}^{d+1}$, $P_i = (1, X_i)$ and ε represents the unknown error. The sample $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ will be called training sample.

Along this section we are going to present different methods in order to estimate the parameter θ based on quantile or least squares methods. The performance of each method for prediction is assessed by means of an evaluation sample of new observations. Later on, we are going to present different methods available in the literature to construct prediction intervals. In addition, a new method for computing prediction intervals is proposed, based on a median regression model and a bootstrap procedure.

2.2.1 Least squares versus quantile methods

The regression model given in (2.1) can be interpreted as a mean regression model, if we assume that the error has an expectation of zero, that is, $\mathbb{E}(\varepsilon|X) = 0$. In this case, the model can be estimated by minimizing the sum of squared residuals

$$\widehat{\theta}_{\text{LS}} = \arg\min_{\theta} \sum_{t} \left(Y_i - \theta' P_i\right)^2$$

where $\theta = (\theta_0, \theta_1, \ldots, \theta_d)'$ is the vector of coefficients to be estimated, $P_i = (1, X_i)'$ is the vector of predictors, and $\widehat{\theta}_{\text{LS}}$ is the least squares estimator.

For the same expression (2.1), instead of assuming $\mathbb{E}(\varepsilon|X) = 0$, one can surmise that the error has a τ-th quantile equal to zero, that is, $\mathbb{P}(\varepsilon \leq 0|X) = \tau$ with $\tau \in (0, 1)$. This implies that the proportion of negative (non-positive) errors is expected to be τ, which is equivalent to the proportion of observations below the regression function equal to τ. In this way, the regression function is no longer the conditional expectation, but rather the conditional τ-th quantile of the response variable given the predictors.

Let us remember that to estimate the coefficients θ_τ for a certain τ, we observed that while mean regression minimizes the sum of squared residuals, the τ-quantile minimizes the sum of weighted absolute values of the residuals. Thus, the estimator $\widehat{\theta}_\tau$ is given by

$$\widehat{\theta}_\tau = \arg\min_{\theta} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \theta' P_i \right)$$

for each $\tau \in (0, 1)$, where $\rho_\tau(u) = (1 - \tau)\, u\, \mathbb{I}(u \leq 0) + \tau\, u\, \mathbb{I}(u > 0)$ is the quantile loss function. Here, ρ_τ produces a weighting effect on the residuals. Positive residuals, which correspond to observations above the regression function, are weighted by the factor τ. Negative residuals receive the weighting factor $(1 - \tau)$.

Because different regression models will be compared, validation of these models is critical. We applied well-known procedures to check whether the residuals satisfy the assumptions of homoscedasticity and normality. More details are given in Section 2.4, where this kind of models is applied to an example atmospheric data. However, validation of quantile regression models is not much addressed in the literature. One reason is that quantile regression does not require any stringent assumptions about the error distribution. The linearity of the quantile model, which is the most critical assumption, can be checked using the lack-of-fit test developed by Conde-Amboage et al. (2015) whose details are given in Chapter 4. This test evaluates the fit of a parametric quantile regression model with many predictors (in our environmental example, a linear quantile model with five predictors) versus any other possible model, that is, a nonparametric alternative. The test is based on the cumulative sum of residuals with respect to unidimensional linear projections of the covariates, and a wild bootstrap mechanism is used to approximate the critical values of the test. In Section 2.4, p-values for this test are provided to assess the linear quantile regression model associated with different values of τ in a real data application.

2.2.2 Assessment of prediction methods

The pointwise prediction for a value of Y_{i_0} at a future time i_0 using $P_{i_0} = (1, X_{i_0})'$ as predictors, can be obtained from the mean regression model by means of $\widehat{Y}_{i_0,\text{LS}} = \widehat{\theta}'_{\text{LS}} P_{i_0}$. An alternative prediction can be obtained from the median regression model in a similar way, where $\widehat{Y}_{i_0,\tau=0.5} = \widehat{\theta}'_{\tau=0.5} P_{i_0}$.

To compare the performance of these two prediction methods, we use two criteria: the mean absolute error (MAE) and the mean squared error (MSE), given by

$$MAE = n_0^{-1} \sum_{i_0} \left| Y_{i_0} - \widehat{Y}_{i_0,m} \right|$$

$$MSE = n_0^{-1} \sum_{i_0} \left(Y_{i_0} - \widehat{Y}_{i_0,m} \right)^2$$

where m represents the prediction method, either a least squares regression (LS) or median regression ($\tau = 0.5$); the indices i_0 in the summation are those of the evaluation sample; and n_0 is the evaluation sample size. Note that the estimations $\widehat{\theta}_{\text{LS}}$ and $\widehat{\theta}_{\tau=0.5}$ are calculated from the training sample.

2.2.3 Prediction intervals: conditional and unconditional coverage

Definition 2.1. A **prediction interval** for a value Y_{i_0} is an interval that is expected to contain the true value Y_{i_0} with a (presumably) high probability $(1-\alpha)$, usually called the confidence level. Let us denote a prediction interval as (L_{i_0}, U_{i_0}), where the endpoints L_{i_0} and U_{i_0} are obtained as functions of the training sample, and the values of the predictors P_{i_0} at time i_0. It would be expected that

$$P\left(Y_{i_0} \in (L_{i_0}, U_{i_0})\right) = 1 - \alpha.$$

In this expression, the probability is defined for all possible training samples and new observations. We call this *unconditional coverage.* However, because the value of the predictors for new observation, P_{i_0}, is known, it is reasonable to define the above probability as conditional to these predictors, that is,

$$P\left(Y_{i_0} \in (L_{i_0}, U_{i_0}) | P_{i_0} = p_{i_0}\right).$$

We call this probability the *conditional coverage.* The unconditional coverage can be obtained as an average of the conditional coverage, with respect to the predictors distribution. A sample analogue for the unconditional coverage would be the proportion of prediction intervals that contain the new observation in the entire evaluation sample, while the conditional coverage is the same proportion, but with evaluation samples taken at a certain value of the predictor P_{i_0}. Mayr et al. (2012) provides further explanation of these concepts.

The immediate consequence of these definitions is that: if the conditional coverage respects the nominal level $(1-\alpha)$, then the unconditional coverage will also respect it. The reverse is not necessarily true. Retaining a conditional coverage at the nominal level $(1-\alpha)$ is therefore a more stringent condition, requiring more detailed use of the information gathered by the predictor P_{i_0}.

Below, we outline a number of known methods for obtaining prediction intervals, together with our new proposed method. Each method is valid for a certain set of restrictive assumptions on the error distribution or on the conditional variability. They fail to provide unconditional or conditional coverage, when these assumptions are not satisfied. In particular, misspecification of the error distribution affects the unconditional coverage, while misspecification of the conditional variability affects the conditional coverage. The goal of our proposed method is to provide a prediction interval with appropriate conditional and unconditional coverage based on a quantile regression estimation and a bootstrapping procedure. We found that quantile methods were particularly useful because of their robustness and flexibility under more general conditions.

2.2.4 Methods for obtaining prediction intervals

In many situations, a prediction interval for a future response variable is useful. Here, we consider four published methods for obtaining prediction intervals: two based on mean regression, and two based on quantile regression. In addition, a new method is proposed here, based on median regression estimation and a bootstrapping method.

Method M1

A prediction interval for Y_{i_0} with level $(1-\alpha)$ is traditionally obtained from mean regression by

$$\left(\widehat{Y}_{i_0,\text{LS}} - t_{n-d-1,\alpha/2}\,\widehat{\sigma}\,\sqrt{1+P'_{i_0}(\mathbb{X}'\mathbb{X})^{-1}P_{i_0}}\ ,\ \widehat{Y}_{i_0,\text{LS}} + t_{n-d-1,\alpha/2}\,\widehat{\sigma}\,\sqrt{1+P'_{i_0}(\mathbb{X}'\mathbb{X})^{-1}P_{i_0}}\ \right),$$

where $t_{n-d-1,\alpha/2}$ is the $(1-\alpha/2)$ quantile of the Student's t-distribution with $(n-d-1)$ degrees of freedom; $\widehat{\sigma}^2 = (n-d-1)^{-1}\sum_i (Y_i - \widehat{\theta}'_{\text{LS}} P_i)^2$ is the error variance estimate based on the training sample; and $\mathbb{X}$ is the design matrix of the training sample.

This type of interval was used from the very beginning for estimating prediction intervals (see Seber (1977)) and it is still the most common method used to obtain prediction intervals using linear regression models (see Fahrmeir et al. (2013)). The main drawback of this method is that it heavily depends on the assumptions of homoscedasticity and error normality.

Method M2

Stine (1985) proposed a bootstrapping method to circumvent the error normality condition involved in constructing a prediction interval using mean regression. Homoscedasticity is still required for this method.

To be precise, Stine (1985) proposed computing the prediction interval as

$$\left(\widehat{Y}_{i_0,\text{LS}} + G^{\star-1}_{\text{LS}}(\alpha/2), \widehat{Y}_{i_0,\text{LS}} + G^{\star-1}_{\text{LS}}(1-\alpha/2)\right),$$

where $G^{\star-1}_{\text{LS}}$ represents the quantile function associated with the bootstrap distribution of the prediction error, denoted by $G^{\star}_{\text{LS}}$. Here, $G^{\star-1}_{\text{LS}}(\alpha/2)$ and $G^{\star-1}_{\text{LS}}(1-\alpha/2)$ denote $\alpha/2$ and $(1-\alpha/2)$ quantiles of the bootstrap distribution $G^{\star}_{\text{LS}}$, respectively. These quantities are obtained as follows:

Step 1 Bootstrap replicates of the training sample and the new observation are determined from

$$Y_i^\star = \widehat{\theta}'_{\text{LS}} P_i + \varepsilon_i^\star \qquad i \in \{1, \dots, n\}$$
$$Y_{i_0}^\star = \widehat{\theta}'_{\text{LS}} P_{i_0} + \varepsilon_{i_0}^\star$$

where $\widehat{\theta}_{\text{LS}}$ is an estimate of the mean regression coefficients obtained from the training sample. Moreover, $\varepsilon_i^\star$ and $\varepsilon_{i_0}^\star$ are drawn by sampling with replacement from the empirical distribution of the residuals, that is

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}(r_{i,\text{LS}} \leq x)$$

where $r_{i,\text{LS}} = Y_i - \widehat{\theta}'_{\text{LS}} P_i$ denote the residuals associated with a mean regression model.

Step 2 Based on the bootstrap training sample, a bootstrap replicate of the estimated coefficient can be obtained. Let us denote it by $\widehat{\theta}^\star_{\text{LS}}$. Thus, bootstrap prediction errors are computed as

$$D^\star_{\text{LS}} = Y_{i_0}^\star - \widehat{\theta}^{\star\prime}_{\text{LS}} P_{i_0}.$$

Step 3 Steps 1 and 2 are repeated B times to compute a sample of differences $D^\star_{\text{LS},1}, \dots, D^\star_{\text{LS},B}$. The empirical distribution of this sample is a Monte Carlo approximation of the distribution function $G^\star_{\text{LS}}$, from which the quantiles $G^{\star -1}_{\text{LS}}(\alpha/2)$ and $G^{\star -1}_{\text{LS}}(1-\alpha/2)$ are determined.

Although asymptotically correct as proven by Stine (1985), this kind of bootstrap intervals are liberal in small samples, especially for confidence levels near 1. Such behaviour is expected, since Efron (1983) showed that the bootstrap error rate is an underestimate of the true error rate. Thus, an interval chosen to have a given bootstrap error rate yields a liberal procedure.

A partial remedy for the lack of coverage arises from a small modification in the resampling method. The variance of the least squares residuals that are sampled to obtain $Y_i^\star$ is not σ^2, but $\sigma^2(1-h_i)$. A sample that has the correct variance is obtained by resampling from

$$\widetilde{r}_{i,\text{LS}} = \frac{r_{i,\text{LS}}}{\sqrt{1-h_i}}$$

where $h_i = X_i'(\mathbb{X}'\mathbb{X})^{-1} X_i$ and $\mathbb{X}$ represents the matrix design.

Note that this method is based on the fact that bootstrap simulates the distribution of the forecast error by sampling F_n in place of F and using $\widehat{\theta}_{\text{LS}}$ as the true mean regression coefficient vector instead of the unknown parameter θ. It is important to emphasize that homoscedasticity is still required for this method.

Method Q1

Direct use of the estimated conditional quantile function provides an intuitive approach for constructing prediction intervals and it is especially interesting because of its computational

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

efficiency. So, a prediction interval for Y_{i_0} of level $(1-\alpha)$ can be obtained from quantile regression models as

$$\left(\widehat{Y}_{i_0,\tau=\alpha/2}, \widehat{Y}_{i_0,\tau=1-\alpha/2}\right),$$

where the endpoints are estimations of the $\alpha/2$ and $(1-\alpha/2)$ quantiles of Y_{i_0} conditional to the values of the predictors P_{i_0}. The width of this prediction interval can vary greatly depending on the value of the explanatory variable for the new observation, X_{i_0}.

Intervals of this kind were used by several authors, including Meinshausen (2006) and Mayr et al. (2012), that outline the construction of these prediction intervals and their main advantages. Such intervals do not require homoscedasticity and adapt to any error distribution. Their drawback is that a parametric (commonly linear) model is assumed at extreme quantiles, which effects estimation, leading to an empirical coverage that is smaller than the nominal one. Zhou and Portnoy (1996) shows that the asymptotic coverage probability of this kind of prediction intervals is $(1-\alpha)$ with an error of the order $O(n^{-1/2})$.

Method Q2

This method is similar to the previous one, but has a small correction in order to account for its effects on estimation. This prediction interval is defined as

$$\left(\widehat{Y}_{i_0,\tau=\alpha/2-\delta}, \widehat{Y}_{i_0,\tau=1-\alpha/2+\delta}\right),$$

where $\delta = 0.5(z_{1-\alpha/2}/n)$, $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard Gaussian distribution, and n is the training sample size. The factor 0.5 comes from the expression $\sqrt{\tau(1-\tau)}$, where τ is the quantile to be estimated (here $\tau = 0.5$). This modification was proposed by Zhou and Portnoy (1996). There are other more elaborate procedures to improve the empirical coverage of this kind of prediction intervals, involving extremal quantile regression methods, such as those of Chernozhukov (2005).

Method Q3

In this case, we are going to present a new method that allows us to obtain prediction intervals based on a quantile regression model and a bootstrap procedure. In contrast to Stine (1985)'s proposal, we will consider a wild bootstrap mechanism in order to avoid homoscedasticity condition. In this sense, it will be crucial the development of Feng et al. (2011) that adapt the wild bootstrap procedure to quantile regression context.

This is a new method, proposed here, where the prediction interval is computed as

$$\left(\widehat{Y}_{i_0,\tau=0.5} + G^{\star-1}(\alpha/2)\ ,\ \widehat{Y}_{i_0,\tau=0.5} + G^{\star-1}(1-\alpha/2)\right),$$

where $G^{\star-1}(\alpha/2)$ and $G^{\star-1}(1-\alpha/2)$ denote $\alpha/2$ and $(1-\alpha/2)$ quantiles of the bootstrap distribution of the prediction error, represented by $G^{\star}$. These quantities are obtained as follows:

Step 1 Bootstrap replicates of the training sample and the new observation are determined from

$$Y_i^\star = \widehat{\theta}'_{\tau=0.5}P_i + \varepsilon_i^\star \qquad i \in \{1,\ldots,n\}$$
$$Y_{i_0}^\star = \widehat{\theta}'_{\tau=0.5}P_{i_0} + \varepsilon_{i_0}^\star$$

where $\widehat{\theta}_{\tau=0.5}$ is an estimate of the median regression coefficients obtained from the training sample.

The bootstrap errors are given by $\varepsilon_i^\star = w_i|r_i|$, where $|\cdot|$ denotes the absolute value and $r_i = Y_i - \widehat{\theta}'_{\tau=0.5}P_i$ gives the residuals in the original training sample. The use of the absolute values of the residuals in order to construct the bootstrap errors is a convenient modification of wild bootstrap for quantile regression proposed by Feng et al. (2011). The multipliers, w_i, are independently generated from a common distribution with τ-quantile equal to zero. In this case, we adopt the two-point distribution with probabilities $(1-\tau)$ and τ at $2(1-\tau)$ and -2τ, respectively, that was proposed by Feng et al. (2011) to satisfy their Conditions 3, 4 and 5. Note that other common multipliers distributions for mean regression, generally with the only condition that the variance is one and occasionally with the condition that the third moment is one (see Mammen (1993) for a two-point multipliers distribution in the mean regression), do not satisfy Conditions 4 and 5 required by Feng et al. (2011) to establish consistency of the bootstrap for quantile regression.

The bootstrap error for the new observation is given by $\varepsilon_{i_0}^\star = w_{i_0}|r_{i_0}|$, where w_{i_0} follows the same two-point distribution as w_i, while the residual r_{i_0} is drawn from the following estimate of the conditional distribution of the error for the value of the predictor P_{i_0}:

$$\widehat{F}(r|P_{i_0}) = \sum_{i=1}^{n} \mathbb{I}(r_i \le r) W_{i,P_{i_0}}$$

where $\mathbb{I}(r_i \le r)$ is the indicator function with value 1, if the condition $r_i \le r$ is satisfied, else value 0; and

$$W_{i,P_{i_0}} = \frac{K((\widehat{\theta}'_{\tau=0.5}P_i - \widehat{\theta}'_{\tau=0.5}P_{i_0})/h)}{\sum_{s=1}^{n} K((\widehat{\theta}'_{\tau=0.5}P_s - \widehat{\theta}'_{\tau=0.5}P_{i_0})/h)}$$

are nonparametric smoothing weights. The smoothing parameter was chosen as $h = cn^{-1/5}$, where c is a constant that depends on several unknown quantities, and $n^{-1/5}$ is the conventional rate for this type of Nadaraya–Watson non parametric estimator. See Hall et al. (1999) for more detail on this type of estimator and an outline of the bootstrapping method used to select the value of h. Here, we propose simpler rules to those given by Li and Racine (2007), where a rule-of-thumb is used, taking the constant c to be the standard deviation of the covariate, that is, the variable $\widehat{\theta}'_{\tau=0.5}P_i$ in our framework. In our empirical evaluation, an even simpler rule, taking $c = 1$, was used with satisfactory results.

Step 2 Based on the bootstrap training sample, a bootstrap replicate of the estimated coefficient can be obtained and it is denoted by $\widehat{\theta}^\star_{\tau=0.5}$. Then, bootstrap prediction errors are computed as

$$D^\star = Y_{i_0}^\star - \widehat{\theta}^{\star\prime}_{\tau=0.5}P_{i_0}.$$

Step 3 Steps 1 and 2 are repeated B times in order to compute a sample of differences $D_1^\star, \dots, D_B^\star$. The empirical distribution of this sample allows us to approximate $G^\star$, from which the quantiles $G^{\star-1}(\alpha/2)$ and $G^{\star-1}(1-\alpha/2)$ are determined.

2.2.5 Theoretical discussion

Here, we discuss the expected properties of these various prediction methods, with particular emphasis on the newly proposed method Q3 as it compares to published methods. The expected properties are determined based on empirical outcomes of predicting values using real or simulated data, as outlined in Sections 2.3 and 2.4. Convergence results for the new method are also provided.

Methods M1, M2, and the new method Q3 have in common that they are based on the estimation of a central quantity of the conditional distribution, that is, the conditional mean in methods M1 and M2, and the conditional median in method Q3, as well as estimation of the prediction error distribution. Estimating a mean regression, as in methods M1 and M2, is very efficient under normality, but is inefficient and lacks robustness for more general error distributions. This is one of the main reasons why we proposed a median regression estimation for our method Q3. To estimate the prediction error distribution, the classical method M1 applies a simple rule based on stringent assumptions of homoscedasticity and error normality. This is the best method under these assumptions, but it yields a poor coverage approximation, when these assumptions are not satisfied. Method M2 makes use of a bootstrapping method to estimate the error distribution, but still assumes homoscedasticity. The proposed method Q3 applies a bootstrapping method adapted for quantile regression and a heteroscedastic setup. Hence, this new method is applicable under very general conditions, and overcomes the limitations of the stringent assumptions in methods M1 and M2.

Methods Q1 and Q2 are not based on estimating any central quantity of the conditional distribution, but directly obtain the lower and upper endpoints of the prediction interval through quantile estimation. Method Q1 does not address the problem of prediction error, while method Q2 applies a simple correction for this problem. The main virtue of these two methods is that their quantile procedures of estimation are very flexible with respect to the error distribution type; it is not required to be Gaussian or similar. However, estimating non-central quantities, especially relatively extreme quantiles, has two main drawbacks: there may be few data points available for estimating these extreme quantiles, causing what is known as the problem of sparsity described in the literature dealing with quantile regression; and estimation will usually require a model assumption (most commonly, linearity) that restricts its real world applications. Heteroscedasticity can be considered for methods Q1 and Q2, but only under a specific model for the conditional variability. In other words, methods Q1 and Q2 will work well under linear heteroscedasticity, but will fail under a more general heteroscedastic pattern. This means that estimating a complex model for extreme quantiles is often infeasible in practice. Because of these restrictions, we opted to estimate a central quantile in Q3, that is, the conditional median, and to use a bootstrap approximation of the prediction error to account for general heteroscedasticity and general error distributions.

The convergence properties of the proposed method Q3 are derived using similar arguments to those given in Stine (1985). Thus, the bootstrap prediction errors can be expressed as:

$$D^\star = \varepsilon^\star_{i_0} + \left(\widehat{\theta}_{\tau=0.5} - \widehat{\theta}^\star_{\tau=0.5}\right)' P_{i_0}.$$

Given that the addends on the right are generated independently, the bootstrap distribution of the prediction error is the convolution of two distributions:

$$G^\star = \widehat{F}_{i_0} * Z^\star$$

where $\widehat{F}_{i_0}$ is the distribution of $\varepsilon^\star_{i_0}$ and $Z^\star$ is the distribution of the second addend, that is, the bootstrap approximation of the parameter estimation error multiplied by the predictors. Feng et al. (2011) obtained the consistency of $Z^\star$ under the bootstrapping mechanism proposed here. Hall and Yao (2005) provided the consistency of the estimator $\widehat{F}(r|P_{i_0})$, where smoothing is applied to projected predictors, as performed here. Since $\hat{F}_{i_0}$ is constructed from $\widehat{F}(r|P_{i_0})$ by including the bootstrap multipliers given by Feng et al. (2011), bootstrap validity depends on the consistency of $\widehat{F}(r|P_{i_0})$. Although Stine (1985) makes use of an empirical distribution function of the residuals, a locally smoothed version of the residual distribution is used here. Thus, the asymptotic coverage is attained using a smoothed version of Theorem 2 in Stine (1985).

2.3 Simulation study

A simulation study is carried out to show how deviations, such as those present in our data, from the common assumptions of the classical linear models of mean regression, lead to inadequate predictions and prediction intervals. In such situations, quantile regression is clearly a better option for prediction, while the proposed method Q3 provides a good alternative for computing prediction intervals.

Our simulated model is a linear model, with five explanatory variables, as in our case study,

$$Y = 1 + X^{(1)} + X^{(2)} + X^{(3)} + X^{(4)} + X^{(5)} + \sigma(X^{(1)}, \dots, X^{(5)})\, \varepsilon,$$

where $X^{(1)}, \dots, X^{(5)}$ are independent and have an uniform distribution on the unit interval $(0, 1)$; $\sigma(X^{(1)}, \dots, X^{(5)})$ represents the effect of the predictors on the standard deviation of the response variable; and ε is an random error variable, independent of these predictors.

Three types of conditional standard deviations are considered

Model Ho A homoscedastic model, where $\sigma(X^{(1)}, \dots, X^{(5)}) = 1$.

Model He1 A heteroscedastic model, where $\sigma(X^{(1)}, \dots, X^{(5)}) = (1 + X^{(1)} + X^{(2)} + X^{(3)} + X^{(4)} + X^{(5)})/2$. Note that in this model, the conditional standard deviation is a linear function of the predictors.

Model He2 A heteroscedastic model, where $\sigma(X^{(1)}, \dots, X^{(5)}) = 1 + (X^{(1)} + X^{(2)} + X^{(3)} + X^{(4)} + X^{(5)})^4/100$. Note that this conditional standard deviation is a non-linear function of the predictors.

These three models explore the conditions that should be verified in all datasets to achieve a good performance in estimating prediction intervals for each computation method. Given that methods M1 and M2 are designed under the assumption of homoscedasticity, we would expect a relatively poor performance for heteroscedastic models, He1 and He2. Importantly, He1 and He2 differ in whether their conditional standard deviation is a linear or non-linear function of the predictors, respectively. This creates a big difference in methods Q1 and Q2, because they are based on linear estimations of lower and upper quantiles. Such methods work well under linear heteroscedasticity, like that considered in model He1, but may be misleading under a non-linearly heteroscedastic model like He2. Finally, we would expect a good performance of method Q3 for all considered models, Ho, He1 and He2.

From all of these models, samples of independent observations were drawn of size $2n$ and these datasets will be divided into two parts: a training sample and an evaluation sample. The training set is used to fit the models while the evaluation set is used for assessment of the capacity for forecasting of the chosen model.

		M1	M2	Q1	Q2	Q3
N(0,1)	$n = 100$	90.01	88.41	84.47	85.76	88.50
	$n = 500$	89.90	89.26	88.88	89.21	89.27
	$n = 1000$	90.02	89.49	89.47	89.64	89.50
U(-1,1)	$n = 100$	93.13	88.17	84.26	85.65	88.76
	$n = 500$	94.47	88.92	88.81	89.15	89.63
	$n = 1000$	94.90	89.26	89.44	89.61	89.65
t_2	$n = 100$	93.18	89.30	83.90	85.18	88.46
	$n = 500$	95.00	89.80	88.98	89.29	89.49
	$n = 1000$	95.38	89.61	89.40	89.55	89.40
χ^2_2	$n = 100$	92.55	90.40	84.24	85.66	88.53
	$n = 500$	92.80	90.62	88.92	89.28	89.29
	$n = 1000$	92.90	90.38	89.47	89.63	89.55
Exp(1)	$n = 100$	84.17	85.51	88.73	92.62	90.39
	$n = 500$	88.86	89.17	89.32	92.85	90.56
	$n = 1000$	89.51	89.67	89.50	92.96	90.40
Cauchy(0,1)	$n = 100$	83.56	85.09	88.18	95.95	92.08
	$n = 500$	88.84	89.20	89.34	98.29	92.92
	$n = 1000$	89.43	89.60	89.46	98.76	92.98

Table 2.1: Coverage (in percentage) of prediction intervals obtained using the five methods described in Section 2.2.4, with homoscedastic model Ho. Values are for a nominal level of 90%, as well as for different error distributions and sample sizes. M1 and M2 are based on mean regression models, Q1 and Q2 are based on quantile regression, and Q3 is the median regression model proposed herein.

So, training samples of independent observations were considered of size n (different values will be considered for n) to provide estimates for both quantile and mean regression models. For each training sample, the evaluation sample was drawn of the same size to compute the empirical coverage of the prediction intervals. One thousand training samples and their corresponding evaluation samples were used to compute mean values of the prediction errors and coverage errors. Moreover, five hundred bootstrap replicates were considered. For reasons of brevity, prediction errors are omitted, and only coverages of prediction intervals are presented and discussed below.

		M1	M2	Q1	Q2	Q3
Interval I1	$n = 100$	97.13	96.17	82.82	84.26	92.62
	$n = 500$	97.29	96.90	88.51	88.87	91.83
	$n = 1000$	97.27	96.95	89.30	89.43	91.28
Interval I2	$n = 100$	93.43	91.95	84.29	85.76	90.10
	$n = 500$	93.72	93.11	89.03	89.34	90.42
	$n = 1000$	93.64	93.12	89.44	89.63	90.15
Interval I3	$n = 100$	90.42	88.61	84.80	86.05	88.54
	$n = 500$	90.74	90.06	89.27	89.61	89.69
	$n = 1000$	90.48	89.88	89.53	89.68	89.54
Interval I4	$n = 100$	87.03	85.15	84.89	86.38	87.06
	$n = 500$	87.12	86.33	89.16	89.46	88.65
	$n = 1000$	87.11	86.50	89.47	89.63	89.02
Interval I5	$n = 100$	81.79	79.52	84.86	86.43	84.55
	$n = 500$	81.81	80.98	89.04	89.36	87.65
	$n = 1000$	81.75	81.01	89.51	89.67	88.25

Table 2.2: Coverage (in percentage) of prediction intervals obtained by the five methods described in Section 2.2.4, with heteroscedastic model He1. Values are for a nominal level of 90% and two sample sizes. I1 to I5 represent five intervals of ordered expected values of the response variable. M1 and M2 are based on mean regression models, Q1 and Q2 are based on quantile regression, and Q3 is the median regression model proposed herein.

Table 2.1 contains the empirical coverages obtained for the homoscedastic model (Ho), with a nominal level of 90%, and for different error distributions and sample sizes. The error distributions investigated were standard Gaussian, uniform on the interval $(-1, 1)$, chi-square with two degrees of freedom, Student's t with two degrees of freedom, exponential distribution with mean one and a standard Cauchy distribution. The classical prediction intervals based on linear mean regression show that method M1 provides very accurate results under the standard normal error distribution, while the other three distributions provide empirical coverage that is higher than the nominal level. Method M2 based on linear mean regression with a bootstrap approximation of the prediction error provides accurate coverage for all distributions (with better accuracy for larger sample size), with the only exception being for the Cauchy distribution. The Cauchy distribution does not have a mean, which makes the classical estimator of the linear mean regression inconsistent. The two quantile-based methods, Q1 and Q2, show a coverage below the nominal level for all distributions, although

		M1	M2	Q1	Q2	Q3
Interval I1	$n = 100$	98.40	97.38	78.02	79.53	92.16
	$n = 500$	0.9868	98.05	84.41	84.76	91.03
	$n = 1000$	0.9868	98.02	85.12	85.34	90.51
Interval I2	$n = 100$	96.75	95.20	86.95	88.08	91.44
	$n = 500$	97.27	96.31	91.99	92.27	91.21
	$n = 1000$	97.22	96.26	92.38	92.52	90.73
Interval I3	$n = 100$	94.22	92.04	88.18	89.39	90.49
	$n = 500$	94.71	93.30	92.68	92.95	90.57
	$n = 1000$	94.49	93.20	92.84	92.96	90.28
Interval I4	$n = 100$	88.92	86.27	87.35	88.76	87.67
	$n = 500$	89.28	87.32	91.24	91.54	89.05
	$n = 1000$	89.24	87.47	91.60	91.75	89.34
Interval I5	$n = 100$	74.68	71.39	81.08	82.65	81.12
	$n = 500$	74.61	72.33	84.90	85.30	85.85
	$n = 1000$	74.61	72.42	85.34	85.54	86.96

Table 2.3: Coverage (in percentage) of prediction intervals obtained using the five methods described in Section 2.2.4, with heteroscedastic model He2. Values are for a nominal level of 90% and two sample sizes. I1 to I5 represent five intervals of ordered expected values of the response variable. M1 and M2 are based on mean regression models, Q1 and Q2 are based on quantile regression, and Q3 is the median regression model proposed herein.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

this under-estimation goes to zero, with increasing sample size. Method Q2 performs marginally better than Q1. The proposed method Q3 exhibits an accurate coverage for all four distributions, even for small sample sizes.

Table 2.2 shows the empirical coverages under the first heteroscedastic model (He1), for a nominal level of 90%, and different sample sizes. The error distribution was a standard normal one. In this way, we could analyze the specific effect of heteroscedasticity, without incorporating deviation from normality. Intervals I1 to I5 are defined by means of the quantiles of equal probability of the distribution of the linear function $X^{(1)} + X^{(2)} + X^{(3)} + X^{(4)} + X^{(5)}$, because this is the underlying cause of heteroscedasticity. Coverage for each interval is, in a sense, an indicator of conditional coverage. Simulations show that methods M1 and M2 provide inaccurate conditional coverage, where some intervals (I1 and I2) have under-coverage, and others (I4 and I5) have over-coverage. This effect does not diminish with increasing sample size. Observations for the first intervals have a smaller conditional standard deviation (and are more likely contained within the prediction intervals), while the last intervals have larger ones. This is a natural consequence of heteroscedasticity, since methods M1 and M2 assume homoscedasticity. Coverage derived from methods Q1 and Q2 is somewhat smaller than the nominal level for small sample size, but converges to the nominal level as sample size increases. This reflects the fact that these two methods are based on estimating extreme quantiles using a linear model, which is valid under the linear heteroscedasticity of model He1. Method Q3 provides reasonably accurate coverage at any interval, with better performance observed for larger sample sizes.

Table 2.3 presents empirical coverages for the second heteroscedastic model (He2), with a nominal level of 90% and different sample sizes. The error distribution is a standard normal one, and the five intervals I1 to I5 are constructed, as described for Table 2.2. The four methods M1, M2, Q1 and Q2 are unable to provide accurate coverage in this case. Methods M1 and M2 show the same over-coverage for the first intervals and under-coverage for the last ones, as observed under the He1 model. Methods Q1 and Q2 show an overall under-coverage up to a sample size of $n = 1000$; for larger sample sizes, the non-linearity of the heteroscedasticity produces over-coverage in some intervals, I2 to I4, and under-coverage in others, I1 and I5. Meanwhile, method Q3 is robust to heteroscedasticity of any form, and even under this non-linear model, it provides accurate coverage, with better results for larger sample sizes.

2.4 Application to environmental data

Most of the previous work on pollution around As Pontes power plant was focused on SO_2 levels, because this was the main pollutant from the power plant during its first years of operation, when combustion of local coal was the main power source. Lately, local coal has been replaced by imported coal to reduce SO_2 emissions. This change in source material, together with a new combined cycle generator, have resulted in NO_x pollution becoming more relevant. For this reason, we focus our attention on NO_x levels in this study.

The concentration of NO_x is measured every minute, and recorded by an automatic monitoring system. Simultaneously, the local meteorological station records temperature, wind speed and wind direction every minute. Our purpose is to predict the concentration of NO_x at a time $(t + 30)$ based on available information at time t, where t and $(t + 30)$ are measured in minutes. Thus, a regression model of the following type was considered:

$$Y_t = \theta' X_t + \varepsilon_t = \theta_0 + \theta_1 N_t + \theta_2 N_{t,\text{grad}} + \theta_3 Z_{1t} + \theta_4 Z_{2t} + \theta_5 Z_{3t} + \varepsilon_t \tag{2.2}$$

where N_t is the NO_x concentration at time t; $N_{t,\text{grad}} = N_t - N_{t-5}$ represents the gradient of NO_x concentration over the last 5- minute interval; Z_{1t}, Z_{2t} and Z_{3t} are the mean values of temperature, wind speed and wind direction for the interval covering the last 6 minutes (from $(t - 5)$ to t); $Y_t = N_{t+30}$ is the NO_x concentration at time $(t + 30)$, taken as the response variable; and ε_t represents the error. In this way, measurements for the latest 6-minute interval are used to predict $Y_t = X_{t+30}$. Wind direction is treated as a scalar variable because we measure the absolute value of the deviation angle from true north.

We included all five predictors in our model because they are usually considered to affect local pollution around this power plant (see Prada-Sánchez et al. (2000)). In particular, the NO_x concentration at time t, N_t, is expected to have a positive effect on the same concentration at time $t+30$, Y_t. Then, a linear effect with expected positive coefficient seems to be adequate for this predictor us, a linear effect, with a positive coefficient was selected for this predictor. The remaining four predictors, having a smaller effect on the response variable, but can also be modelled with linear terms; more complex effects are not expected to play a role. In particular, given that low ambient temperature facilitates the dispersion of pollutants, then higher temperatures would promote higher pollutant concentrations local to the power plant. Similarly, high wind speed is associated with pollutant dispersion, resulting in lower pollutant concentrations near to the power plant. Pollutant gradients and wind direction had least association with the response variable, but are included for the purposes of comparison with the literature. In Section 3, a test of linearity is applied to check the validity of our linear model.

Our model (2.2) is fitted using observations covering a period of 10 days; these are defined as the training sample. The subsequent 10-day period is used as the evaluation sample to assess the performance of our predictions and prediction intervals. Given that our model includes observations from the last 6 minutes of measurements as predictors, while the response value is scheduled for 30 minutes later, then these data are divided into blocks of 36 observations, with no predictors or response values for two of these blocks. This circumvents any possible autocorrelation issues. In Section 2.4, autocorrelation tests are applied to validate our model. Thus, after removing some missing data from the training sample, we had a sample size of 338 blocks, with corresponding observations $(N_{t-5}, \ldots, N_t, Z_{1t}, Z_{2t}, Z_{3t}, Y_t)$. Likewise, the evaluation sample comprises another 338 blocks.

Model (2.2) is adjusted for mean regression and for regression with different quantiles, using the training sample described previously. The results for the mean and the median regression are similar, in the sense that the most significant predictors are: N_t (the current concentration of NO_x); and Z_{2t} (the mean value of wind speed). Higher values of the current concentration produce higher predictions for the 30 minute future concentration, as would be expected. Higher wind speed is associated with lower future concentrations. This is consistent with the premise that wind carries NO_x away from the power plant surroundings.

Quantile regression provides a more detailed interpretation of the predictors' effects on each quantile of the future concentration. Figure 2.2 shows the estimated coefficients as a function of the quantile order, τ, together with their confidence intervals. The confidence intervals for the estimated parameters have been calculated by inverting a rank test, as described in Koenker (1994). This method involves solving a parametric linear programming problem, and for large sample sizes can be extremely slow, so by default it is not recommended when the sample size is bigger than 1000. On the other hand, we have used this method because of its several advantages: it is consistent under certain heteroscedastic conditions or ot circumvent any explicit estimation of the sparsity function.

Clearly, in view of Figure 2.2, the effect of each predictor is different at each order τ. The most significant predictor, the current NO_x concentration (N_t), has a coefficient that is positive for all quantiles, but is larger for larger τ (representing the upper range of the future NO_x concentration). The negative effect of wind speed is less dependent on a particular quantile. All coefficients show larger confidence intervals for upper quantiles, related to the higher variability in the high-end range of NO_x concentrations.

The validity of the mean regression model was determined to explore whether its assumptions were satisfied for our case study. First, a scatter plot of the residuals versus the fitted values from the model was produced (left side of Figure 2.3). This plot shows atypical observations. Clearly, more variability is found for higher-fitted values of the response, having a heteroscedastic pattern. Second, a QQ-Plot was constructed (right side of Figure 2.3) to detect deviations from normality. Deviations linked to extreme values, much larger than expected from the normal distribution, are visible. A Shapiro–Wilk test of normality showed a highly significant deviation from normality (p-value smaller than 2.2×10^{-6}). Because

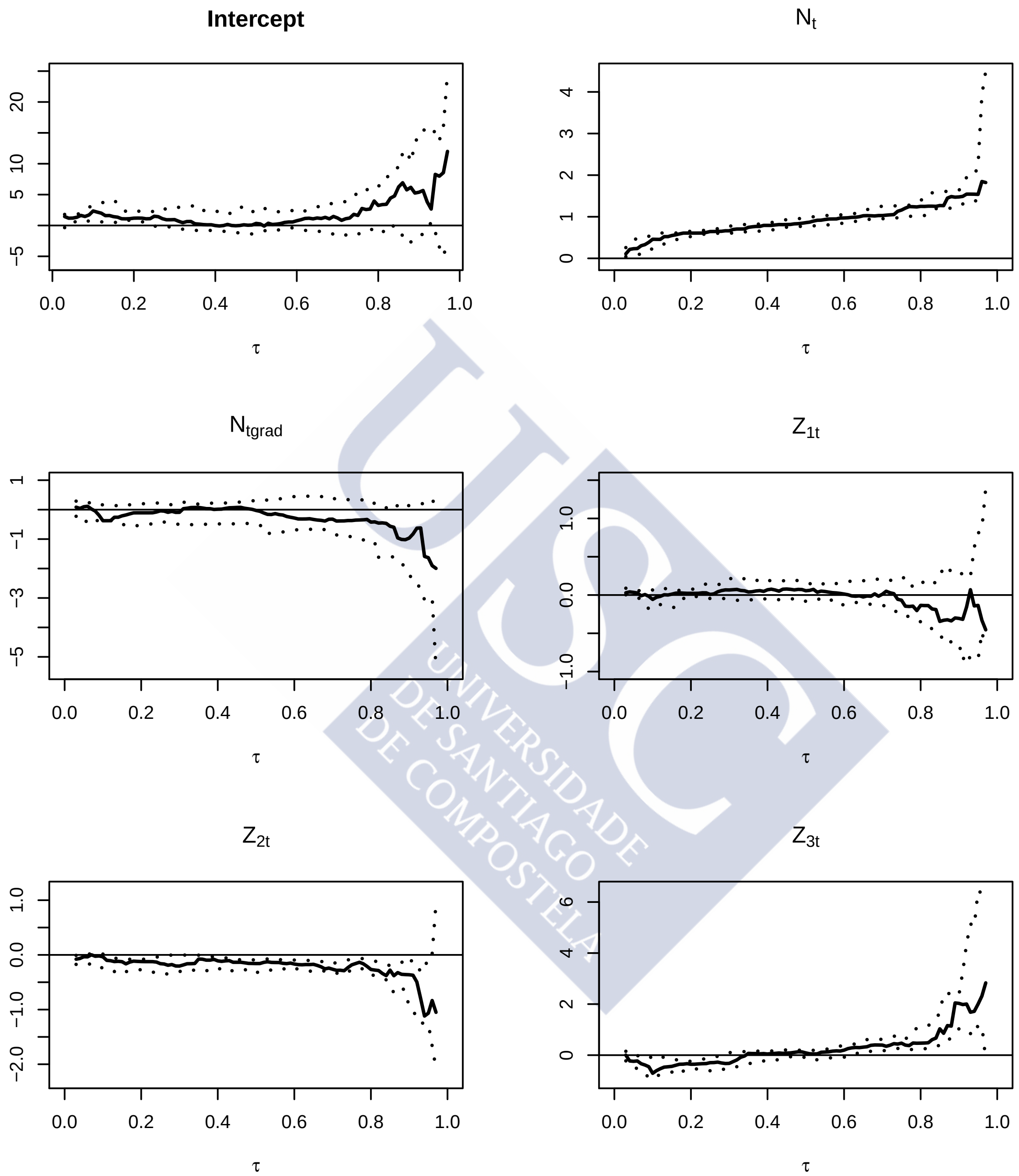


Figure 2.2: Coefficients associated with each predictor as a function of the order τ of the quantile. The solid line represents the coefficients, while the dotted lines represent the endpoints of confidence intervals for the coefficients.

the data in the training sample are obtained as a time series, autocorrelation may occur. Hence, we applied a Durbin–Watson test of one-lag autocorrelation, and a Ljung–Box test of two-lag autocorrelation. No significant autocorrelation was found in either of the tests, with a p-value=0.1656 for the Durbin–Watson test, and a p-value=0.2702 for the Ljung–Box test.

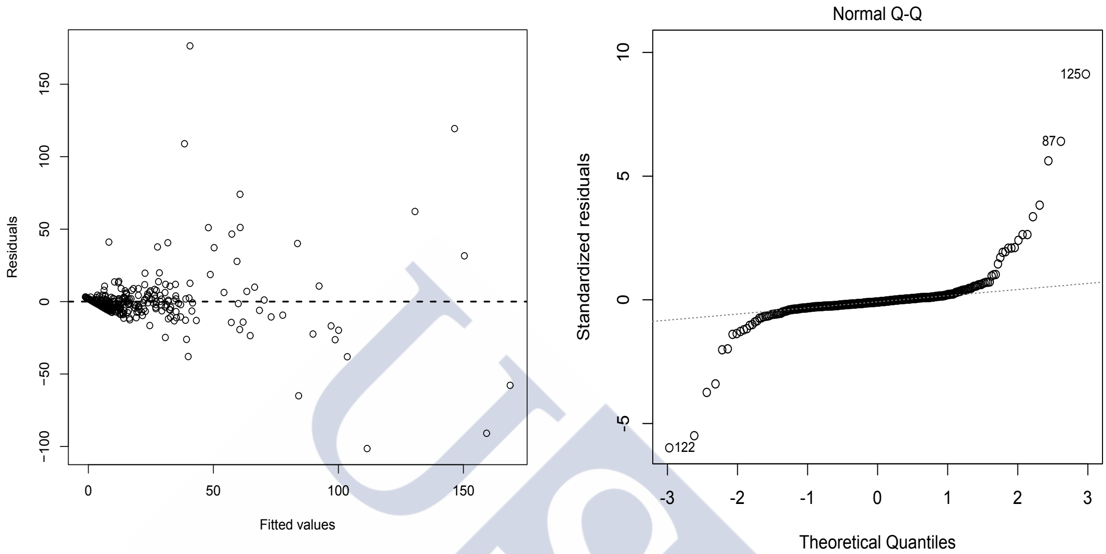


Figure 2.3: Validation of the mean regression model for our case study. Left side: Scatterplot of residuals versus fitted values. Right side: Normal QQ-Plot of standardized residuals.

The quantile regression model is flexible to more general conditions, with less constraints on the error distribution and conditional variability. Thus, it is not necessary to check homoscedasticity or error normality. The only assumption to be tested is that of linearity, that is, the assumption that the predictors effects can be explained by the linear function given by (2.2). We evaluated this assumption using the test that will be detailed along Chapter 4. Table 2.4 contains the p-values associated with this test, carried out for different quantiles. A linear model was acceptable for all quantiles evaluated.

τ	0.025	0.05	0.10	0.5	0.9	0.95	0.975
p-value	0.6702	0.2672	0.1454	0.1332	0.3596	0.3400	0.8322

Table 2.4: p-values for the linearity test carried out for different quantiles.

To evaluate different methods, we compared the mean regression models with our median regression model by means of the prediction errors obtained for the evaluation sample. Table 2.5 shows the MAE and the MSE for the median and the mean regression models for each of the 10 days in the evaluation sample. The last row of the table gives the average value. We observed that the median regression model had smaller prediction errors, both in terms of mean absolute error and mean squared error.

	Mean Absolute Error		Mean Squared Error	
	Median	Mean	Median	Mean
Day 1	4.1008	5.0523	35.925	41.570
Day 2	3.1304	4.1902	19.488	33.362
Day 3	8.3023	8.7387	246.563	248.473
Day 4	8.4403	8.8304	151.346	151.759
Day 5	5.0827	6.2487	122.750	126.733
Day 6	2.3943	3.7248	8.982	19.069
Day 7	8.2810	8.5096	487.657	478.338
Day 8	2.9789	4.0173	27.353	31.836
Day 9	2.3820	4.8768	11.525	30.377
Day 10	3.3201	5.1076	42.249	64.301
Average	4.8413	5.9296	115.384	122.582

Table 2.5: Mean absolute error (MAE) and mean squared error (MSE) associated with predictions obtained using median and mean regression models.

		M1	M2	Q1	Q2	Q3
Level = 90%	Interval I1	100.00	97.05	88.23	88.23	91.17
	Interval I2	100.00	97.05	100.00	100.00	91.17
	Interval I3	100.00	97.05	100.00	100.00	94.11
	Interval I4	100.00	97.05	97.05	97.05	88.23
	Interval I5	100.00	88.23	94.11	94.11	85.29
	Average	100.00	95.29	95.88	95.88	90.00
Level = 95%	Interval I1	100.00	97.05	88.23	88.23	91.17
	Interval I2	100.00	100.00	100.00	100.00	94.11
	Interval I3	100.00	100.00	100.00	100.00	94.11
	Interval I4	100.00	100.00	100.00	100.00	97.05
	Interval I5	100.00	100.00	97.05	100.00	91.17
	Average	100.00	100.00	98.23	99.41	94.70

Table 2.6: Coverage (in percentage) of prediction intervals obtained using the five methods described in Section 2.2.4, for nominal levels 90% and 95%. Intervals I1 to I5 are defined by splitting the ordered Y-values by their empirical quantiles. M1 and M2 are based on mean regression models, Q1 and Q2 are based on quantile regression, and Q3 is the median regression model proposed herein.

Figure 2.4 shows the fitted mean and median regression models associated with two of the days of study in order to have an idea about the motivation of the good performance of the median regression model. Note that the red line represents the median regression

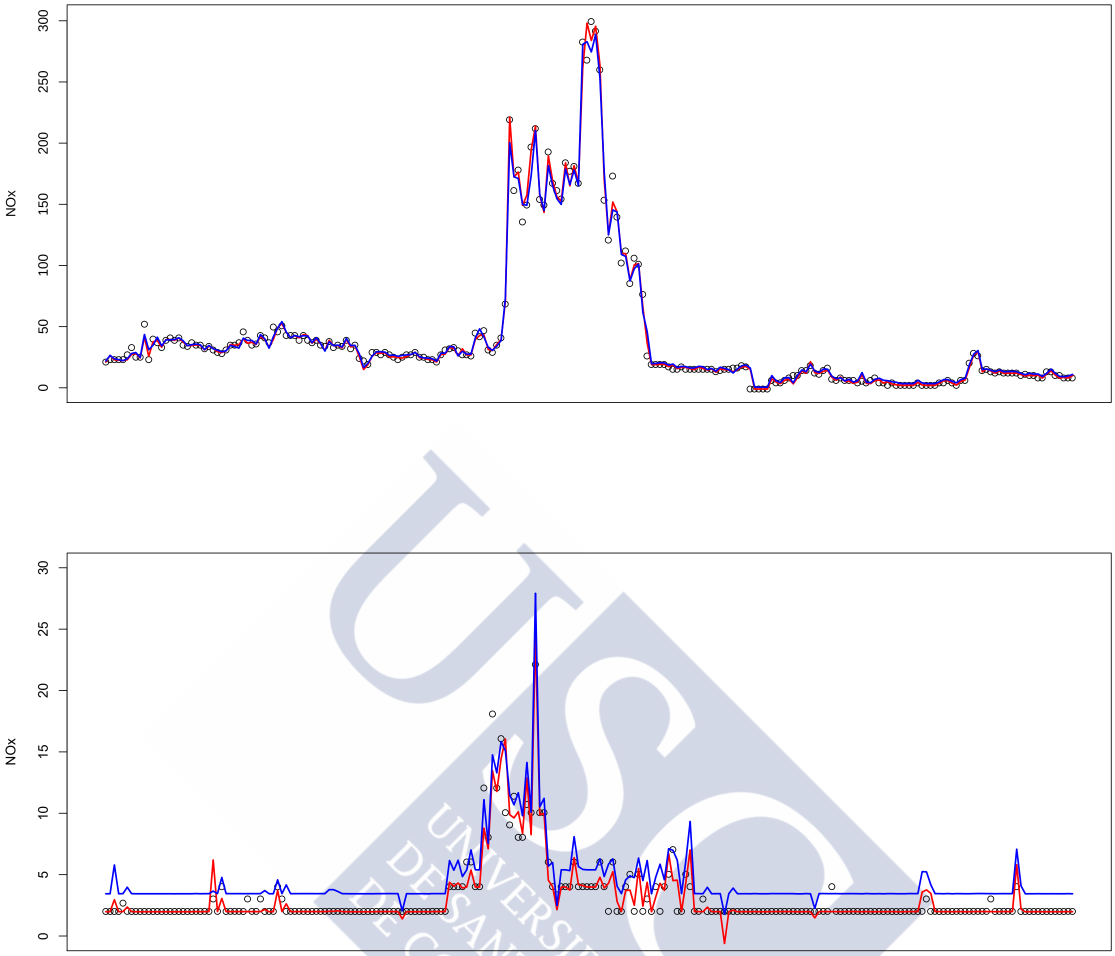


Figure 2.4: Representation of the fitted median (red line) and mean (blue line) regression models associated with a couple of the days of study.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

model while the blue line represents the mean regression model. According to the first picture of Figure 2.4, median regression seems to predict better high concentration of NO_x. In addition, the second picture highlights the leverage problems associated with mean regression. That is, some "atypical" high concentration changes too much the estimators associated with mean regression. These two effects can justify the good results that median regression shows according to Table 2.5.

We computed prediction intervals using the five methods described in Section 2.2.4. Table 2.6 shows the empirical coverage of these prediction intervals, computed as the percentage of times that the real value of NO_x concentration in the evaluation sample belonged to the prediction interval. This was done for two nominal levels: 90% and 95%. The nominal levels are compared with the actual conditional and unconditional coverages. Our validation of these regression models indicated that we are working in a heteroscedastic context. Thus, it is reasonable to expect that the variability of the response variable (NO_x concentration at time t) is not the same for high and low concentrations. In this case, heteroscedasticity is related to the value of the response. Therefore, conditional coverages were computed for five intervals, I1 to I5, each with the same number of observations, defined by evenly splitting the ordered Y-values. Because the evaluation sample size is 338, we considered these intervals to have a reasonable number of elements (around 67) in each interval. In addition, the variability of the response variable within each interval is not large. The unconditional coverage is the average of the conditional coverages in the five intervals.

Clearly, both conditional and unconditional coverages shown in Table 2.6 are much larger than the nominal level for methods M1, M2, Q1 and Q2, while the proposed method Q3 provides coverages quite close to the nominal level for each interval and the overall average. The fact that assumptions of normality and homoscedasticity are not satisfied, likely affected the behaviour of methods M1, M2, Q1 and Q2. These effects have been discussed in more detail in Section 2.3.

2.5 Conclusions

Quantile regression methods are evaluated as an alternative to mean regression for prediction and calculation of prediction intervals of NO_x concentrations around the power plant in As Pontes, Spain. We show that for these data, median regression provides smaller prediction errors than mean regression. Heteroscedasticity and a non-normal error distribution were found to characterize these data, which deviate from the assumptions for classical mean regression models and likely explain the better performance of quantile methods for these data.

Although two known methods based on quantile regression were explored for obtaining prediction intervals, because of the special features of our atmospheric data, we also proposed an additional method based on quantile regression estimation and bootstrap approximation of the prediction error. Our new method gave a markedly better performance than other methods evaluated here. In a simulation study, we showed how deviations from the assumptions of homoscedasticity and normality affected other methods for computing prediction intervals. The coverage accuracy of our new method was shown for both real and simulated scenarios.

Chapter 3

A plug-in bandwidth selector for nonparametric quantile regression

Contents

In the framework of quantile regression, local linear smoothing techniques have been studied by several authors, particularly by Yu and Jones (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. Along this chapter, 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 sparsity (inverse of the conditional density) are both nonparametrically estimated. The new bandwidth selector is shown to work well in different scenarios, particularly when the conditions commonly assumed in the literature are not satisfied.

3.1 Introduction

Along Chapter 1, parametric quantile regression models have been introduced. These models play a critical role throughout the realm of scientific data analysis. Nevertheless, there are inevitable occasions when parametric specifications fail, and data analysis must turn to be more flexible. In this context, nonparametric regression arises because it relaxes the usual assumption of linearity.

The nonparametric quantile regression model can be stated as

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

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

Along Chapter 1, it have been shown that estimation of the quantile regression model can be obtained by exploiting 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,as we have mentioned previously.

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 and h_τ represents a bandwidth parameter. This is the local linear estimator of the quantile regression function.

3.1.1 Bandwidth selectors available in the literature

As occurs in any smoothing method, bandwidth h_τ exhibits a strong influence on the resulting estimate. Too small values of the bandwidth will result in undersmoothed estimations (that is, the estimation contains too many spurious artefacts) while too big values are associated with oversmoothed estimations (that is, the estimation obscures much of the underlying structure). Several authors have addressed the problem of bandwidth selection, see Yu and Jones (1998), Abberger (1998), Yu and Lu (2004), El Ghouch and Genton (2012) or Abberger (2002).

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. Fan et al. (1994) 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

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

where g is the density of X, $f(q_\tau(x)|X = x)$ is the conditional density of Y at $q_\tau(x)$ given $X = x$, and

$$\begin{aligned} q_\tau^{(i)}(x) &= \frac{\partial^i q_\tau(x)}{\partial x^i} \\ \mu_i(K) &= \int u^i K(u)\,du \quad \text{with} \quad i = 0, 1, \dots \\ R(K) &= \int K^2(u)\,du. \end{aligned}$$

Moreover, Fan et al. (1994) obtained a similar result of the mean integrated squared error for a boundary point of the design.

Two of the major advantages of local linear fitting that apply to the quantile regression problem as much as to mean regression estimation are:

- The asymptotic bias does not depend on the design density g, and indeed it depends only on the simple quantile curvature function.

- Automatic good behaviour at boundaries, without the need for further boundary correction.

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

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

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. 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, presented in Section 1.3) 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 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).

Now, we are going to describe the plug-in selector proposed by Yu and Jones (1998). Taking into account the optimal bandwidth given in (3.2), Yu and Jones (1998) studied the relationship between optimal bandwidths for different values of the τ-th quantile of interest, that is,

$$\left(\frac{h_{\text{AMISE},\tau_1}}{h_{\text{AMISE},\tau_2}}\right)^5 = \frac{\tau_1(1-\tau_1)q_{\tau_2}^{(2)}(x)^2 f(q_{\tau_2}(x)|X=x)^2}{\tau_2(1-\tau_2)q_{\tau_1}^{(2)}(x)^2 f(q_{\tau_1}(x)|X=x)^2}.$$

Then, they simplified the previous relationship between h_{AMISE,τ_1} and h_{AMISE,τ_2} by making approximations to the unknown involved quantities, that is, the curvature and the sparsity. Firstly, Yu and Jones (1998) consider that the second derivatives could be similar for different quantiles, that is, $q_{\tau_1}^{(2)}(x) \simeq q_{\tau_2}^{(2)}(x)$. On the other hand, in order to compute the sparsity, they assume a normal error distribution. If f represents the density associated with a Gaussian distribution with mean equal to μ_x and variance equal to σ_x^2, then $f(q_\tau(x)|x) = \sigma_x^{-1}\phi(\Phi^{-1}(\tau))$ where ϕ and Φ represent the density and distribution function associated with a standard Gaussian distribution, respectively. In this case, the following equality is obtained

$$\frac{f(q_{\tau_2}(x)|x)}{f(q_{\tau_1}(x)|x)} = \frac{\phi(\Phi^{-1}(\tau_2))}{\phi(\Phi^{-1}(\tau_1))}.$$

Using these approximations, Yu and Jones (1998) concluded that

$$\left(\frac{h_{\text{AMISE},\tau_1}}{h_{\text{AMISE},\tau_2}}\right)^5 = \frac{\tau_1(1-\tau_1)\phi(\Phi^{-1}(\tau_2))^2}{\tau_2(1-\tau_2)\phi(\Phi^{-1}(\tau_1))^2}. \tag{3.3}$$

Figure 3.1 represents the quotient $\frac{h_{\text{AMISE},\tau}}{h_{\text{AMISE},0.5}}$ for different values of the τ-th quantile of interest. From Figure 3.1, the asymptotically optimal bandwidth is smallest for the median regression, this bandwidth increases symmetrically for τ above and bellow 0.5 and it goes to infinity when τ goes to zero or one. This is due to the sparsity going to infinity at the tails of the Gaussian distribution.

At this point, Yu and Jones (1998) expressed the bandwidth $h_{0.5}$ in terms of the optimal choice of the bandwidth for mean regression estimation, that has the following expression:

$$h^5_{\text{AMISE,MEAN}} = \frac{R(K)\sigma^2(x)}{n\mu_2(K)^2 m^{(2)}(x)^2 g(x)}$$

where $m(x)$ and $\sigma(x)^2$ are the conditional mean and variance. It then follows that

$$\left(\frac{h_{\text{AMISE,MEAN}}}{h_{\text{AMISE},0.5}}\right)^5 = \frac{4q_{0.5}^{(2)}(x)^2\sigma^2(x)f(q_{0.5}(x)|x)^2}{m^{(2)}(x)^2}.$$

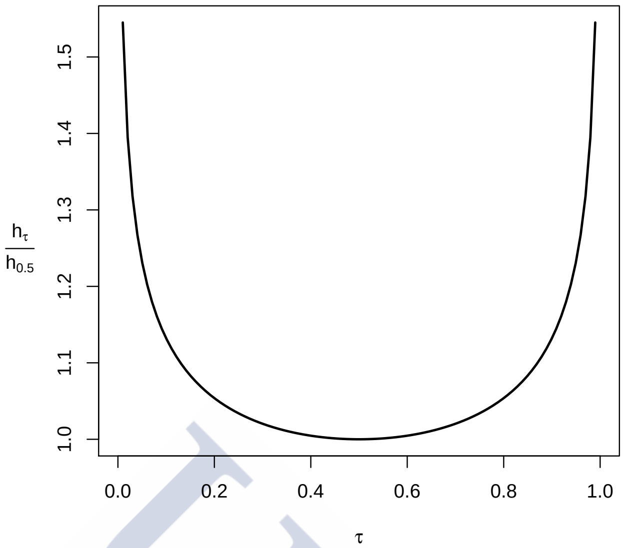


Figure 3.1: Relationship between the optimal bandwidth associated with any τ-th quantile regression and the bandwidth for median regression.

By the same arguments as those above, $q_{0.5}^{(2)}(x)$ and $m^{(2)}(x)$ may be set equal, and a normal distribution error is supposed in order to conclude that $\sigma(x)^2 f(q_\tau(x)|x)^2$ could be replaced by $\phi(\Phi^{-1}(0.5))^2 = (2\pi)^{-1}$. So, finally

$$\left(\frac{h_{\text{AMISE,MEAN}}}{h_{\text{AMISE},0.5}}\right)^5 = \frac{2}{\pi}.$$

Combining this expression with equation (3.3), Yu and Jones (1998) plug-in rule proposal is derived

$$\widehat{h}_{\tau,\text{YJ}} = \sqrt[5]{\frac{\tau(1-\tau)}{\phi(\Phi^{-1}(\tau))^2}}\,\widehat{h}_{\text{RSW}} \tag{3.4}$$

where $\widehat{h}_{\text{RSW}}$ is selected by the plug-in rule proposed by Ruppert et al. (1995).

The plug-in rule procedure proposed by Yu and Jones (1998) may lead to acceptable solutions in cases where the data structure permits all these simplifying assumptions. But in other cases these assumptions are quite restrictive and because of this reason a new plug-in rule will be presented here.

Other proposals for bandwidth selection in nonparametric quantile regression were given in literature, based on cross-validation techniques, which were very popular in early approaches of classical nonparametric regression. Cross-validation is primarily a way of measuring the predictive performance of a statistical model. Abberger (1998) 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

$$\widehat{h}_{\tau,\mathrm{CV}} = \arg\min_h \mathrm{CV}(h) = \arg\min_h \sum_{i=1}^{n} \rho_\tau\Big(Y_i - \widehat{q}^{\,-i}_{\tau,h}(X_i)\Big) \tag{3.5}$$

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. For a fixed h parameter, this method works as follows:

1.- Let us remove the i-th observation from the data set, and fit the model using the remaining data. Then, compute the residual $(r_i = Y_i - \widehat{q}_{\tau,h}(X_i))$ for the omitted observation.

2.- Repeat step 1 for $i = 1, \ldots, n$.

3.- Compute the sum of the residuals weighted by the quantile loss function ρ_τ.

It is well known that the cross-validation process has associated an accurate estimator (bias will be small) whereas the variance of the estimator will be large. The main disadvantage of cross-validation is, as in classical mean regression, its low relative convergence rate of order $n^{-1/10}$. Moreover, the required computational effort will be very large as well.

In view of the state of the art, the purpose of this chapter is to provide a plug-in bandwidth for quantile regression without imposing restrictions on the conditional variability and the error distribution. Instead, nonparametric estimations of the curvature at the given τ-th quantile will be used, as well as nonparametric estimations of the sparsity.

In Section 3.2 a preliminary rule of thumb is obtained, and afterwards the proposed plug-in rule is derived. In Sections 3.3 and 3.4 derivations of the mean integrated squared error of the curvature and sparsity estimators are given. Section 3.5 contains a simulation study in order to explore the virtues of the new bandwidth selectors in comparison with Yu and Jones (1998) and Abberger (1998) proposals. In addition, Section 3.6 describes an R package developed to implement the new bandwidth selectors. Finally, Section 3.7 contains the main conclusions of this chapter.

Remark 3.1. During this chapter, we focus on kernel smoothing techniques, although spline methods have been widely studied by several authors as Koenker et al. (1994) or Koenker and Mizera (2004). For instance, Koenker et al. (1994) proposed to estimate the function q_τ by solving the following optimization problem

$$\min\left[\sum_{i=1}^{n} \rho_\tau\Big(Y_i - q_\tau(X_i)\Big) + \lambda \mathbf{V}(\nabla q_\tau)\right] \tag{3.6}$$

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, Koenker et al. (1994) showed that the solution to (3.6) is a linear spline with nodes at the points X_i where $i = 1, \ldots, n$. Because of this reason, a quantile smoothing spline model can be fitted using l_1-type linear programming techniques. They also proposed to adapt the information criterion of Schwarz (1978) for the choice of the smoothing parameter λ involved in problem (3.6).

3.2 Newly proposed bandwidth 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. The second approach will be a plug-in rule based on nonparametric estimators of the curvature and the sparsity.

3.2.1 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 $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ is subsequently split into the N blocks. A polynomial of order four is adjusted at each block, then providing N fitted models that will be denoted by

$$\widehat{q}_{\tau,j}(x) = \widehat{\theta}_{0,j} + \widehat{\theta}_{1,j}\, x + \widehat{\theta}_{2,j}\, x^2 + \widehat{\theta}_{3,j}\, x^3 + \widehat{\theta}_{4,j}\, x^4$$

   with $j = 1, \ldots, N$. The number of blocks will be chosen as $\widehat{N}$ following the Mallows's C_p criterion (see Mallows (1973)) adapted to the quantile framework, that is, $\widehat{N}$ will minimize

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

   where $\mathrm{RSQ}(N)$ 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.

2. Estimate the curvature as

$$\widehat{\vartheta}_{\mathrm{B}} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{\widehat{N}} \widehat{q}_{\tau,j}^{(2)}(X_i)^2\, \mathbb{I}\,(X_i \in \text{Block } j),$$

   where it is clear that

$$\widehat{q}_{\tau,j}^{(2)}(x) = 2\widehat{\theta}_{2,j} + 6\widehat{\theta}_{3,j}\, x + 12\widehat{\theta}_{4,j}\, x^2.$$

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

3. Supposing an homoscedastic scenario at each of the $\widehat{N}$ blocks, then the sparsity function will be constant at each block and will be given by

$$s_j(\tau) = \frac{1}{f_{\varepsilon,j}(F_{\varepsilon,j}^{-1}(\tau))}$$

   where $f_{\varepsilon,j}$ and $F_{\varepsilon,j}$ represent the density and distribution functions associated with the error ε at each block j with $j = 1, \ldots, \widehat{N}$. So, it will be possible to estimate the sparsity at each block j by means of

$$\widehat{s}_j = \frac{r_{[\tau+d_j]} - r_{[\tau-d_j]}}{2\, d_j}$$

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), among others. For the parameter d_j, the selector proposed by Bofinger (1975) will be used here (details were given in Section 1.3). Finally, the integrated squared sparsity will be estimated by

$$\widehat{s}^2_{\mathrm{B}} = \sum_{j=1}^{\widehat{N}} \widehat{s}^2_j \, l_j \tag{3.7}$$

where l_j denotes the length of block j.

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

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

3.2.2 Plug-in rule

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

Curvature estimation

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

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

At this point, a pilot bandwidth h_c for curvature estimation should be selected. The criterion for selecting h_c will be to minimize the asymptotic mean squared error of the curvature estimator whose expression in derived in Section 3.3. As a summary, let us mention that the variance will be asymptotically negligible, so the asymptotic mean squared error coincides with the asymptotic squared bias, which is given by

$$\mathrm{MSE}\left(\widehat{\vartheta}_{h_c}\right) \cong \left[\delta_1\, h_c^2 \int q^{(2)}_{\tau}(x) q^{(4)}_{\tau}(x) g(x) dx + \delta_2\, \frac{\tau(1-\tau)}{n h_c^5} \int \frac{1}{f(q_\tau(x)|X=x)^2} dx \right]^2 \tag{3.8}$$

where

$$\delta_1 = \frac{1}{6}(\alpha_{31}\, \mu_4(K) + \alpha_{33}\, \mu_6(K))$$

$$\delta_2 = 4\left(\alpha_{31}^2 \int K^2(v)\,dv + \alpha_{33}^2 \int v^4\, K^2(v)\,dv + 2\alpha_{31}\alpha_{33} \int v^2\, K^2(v)\,dv\right)$$

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

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

Minimizing expression (3.8), the asymptotically optimal pilot bandwidth will be

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

where

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

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 needed. They will be obtained from blocked estimators as those considered for the rule of thumb and these estimators will be denoted by $\widehat{s}_{\mathrm{B}}^2$ given in expression (3.7), and

$$\widehat{\vartheta}_{24,\mathrm{B}} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{\widehat{N}} \widehat{q}_{\tau,j}^{(4)}(X_i)\, \widehat{q}_{\tau,j}^{(2)}(X_i)\, \mathbb{I}\left(X_i \in \text{Block } j\right).$$

The resulting estimated pilot bandwidth will be

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

Finally, the curvature estimator will be given by

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

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 $s_\tau(x) = 1/f(q_\tau(x)|X = x)$, results to be the derivative of the quantile regression function, $q_\tau(x)$, with respect to τ, we propose an estimate of this kind

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

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, that is obtained in Section 3.4, and is given by

$$\begin{aligned}\text{MSE}\left(\int \widehat{s}^2_{d_s,h_s}(x)\, dx\right) \cong & \left[\frac{1}{nd_sh_s}\int a(x)\, dx + d_s^2 \int b(x)\, dx + h_s^2 \int c(x)\, dx\right]^2 \\ & + \frac{1}{nd_s}\int d(x)\, dx + \frac{1}{n^2 d_s^2 h_s}\int e(x)\, dx \end{aligned} \tag{3.9}$$

where

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

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 several optimization methods such as Newton method. Estimation of the integrals in expression (3.9) is done by blocks. The resulting pilot bandwidths will be denoted by $\widehat{d}_s$ and $\widehat{h}_s$. Now details are given on how to estimate the unknown integrals.

- **Estimation of $\int \mathbf{a(x)\, dx}$.** 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, that 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:

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

- **Estimation of $\int \mathbf{b(x)}\,\mathbf{dx}$.** 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

$$\widehat{s}_j^{(2,\tau)} = \frac{1}{2\delta^3}\;\left(r_{([n\tau]+2m)} - 2\,r_{([n\tau]+m)} + 2\,r_{([n\tau]-m)} - r_{([n\tau]-2m)}\right)$$

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

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

- **Estimation of $\int \mathbf{c(x)}\,\mathbf{dx}$.** 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

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

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

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

and the subsequent estimation of the integral

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

- **Estimation of $\int \mathbf{d(x)}\,\mathbf{dx}$.** Note that $d(x) = 2\,s_\tau(x)^4\,(g(x))^{-1}$. Similarly to the previous integrals, this integral can be estimated by

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

- **Estimation of $\int \mathbf{e}(\mathbf{x})\,\mathbf{dx}$.** Note that $e(x) = (0.5R(K*K) - R(K))\; s_\tau(x)^4\, g(x)^{-2}$. 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, that 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

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

Finally, the new plug-in bandwidth selector is obtained as:

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

3.3 Derivation of the asymptotic mean integrated squared error of the curvature estimator

In this section the asymptotic mean integrated squared error of the curvature estimator will be derived. Let us recall the expression for the curvature estimator

$$\widehat{\vartheta}_{h_c} = \frac{1}{n}\sum_{i=1}^{n} \widetilde{q}^{(2)}_{\tau,h_c}(X_i)^2 \tag{3.10}$$

where $\widetilde{q}^{(2)}_{\tau,h_c}(X_i)$ represents a nonparametric estimator of the second derivative of the regression function at X_i for $i = 1,\dots,n$. In particular, $\widetilde{q}^{(2)}_{\tau,h_c}(X_i)$ is taken as the second derivative of a local polynomial of order three.

We make the following assumptions:

Conditions C

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, non negative and 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 $h_c \to 0$ and $nh_c^5 \to \infty$ when $n \to \infty$.

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 $\mathrm{Bound}(y|X=x)$ such that

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

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and

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

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

where $\psi_\tau(r) = \tau\mathbb{I}(r > 0) + (\tau - 1)\mathbb{I}(r < 0)$ is the derivative of the quantile loss function $\rho_\tau = \tau r\mathbb{I}(r > 0) + (\tau - 1)r\mathbb{I}(r < 0)$.

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.

3.3.1 Second derivative of the quantile regression function

First, we will establish the asymptotic behaviour of the nonparametric estimation of the second derivative that has been denoted by $\widetilde{q}^{(2)}_{\tau,h_c}$. The main approach to get the asymptotic representation follows the proof of Theorem 2 in Fan et al. (1994). While Theorem 2 in Fan et al. (1994) provides a representation for the nonparametric estimator of the regression function, here a representation will be obtained for the estimator of its second derivative.

Recall that $\widetilde{q}^{(2)}_\tau(x) = 2\widehat{\gamma}_2$ where $(\widehat{\gamma}_0, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}_3)$ minimizes

$$\sum_{i=1}^n \rho_\tau \left(Y_i - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3\right) K\left(\frac{X_i - x}{h_c}\right).$$

Let us introduce the following notation:

$$\bar{\theta} = \sqrt{nh_c} \begin{pmatrix} \widehat{\gamma}_0 - q_\tau(x) \\ h_c\left(\widehat{\gamma}_1 - q_\tau^{(1)}(x)\right) \\ h_c^2\left(\widehat{\gamma}_2 - \frac{1}{2}q_\tau^{(2)}(x)\right) \\ h_c^3\left(\widehat{\gamma}_3 - \frac{1}{6}q_\tau^{(3)}(x)\right) \end{pmatrix} \qquad Z_i = \begin{pmatrix} 1 \\ \frac{X_i - x}{h_c} \\ \left(\frac{X_i - x}{h_c}\right)^2 \\ \left(\frac{X_i - x}{h_c}\right)^3 \end{pmatrix}$$

$$Y_i^{(3)} = Y_i - q_\tau(x) - q_\tau^{(1)}(x)(X_i - x) - \frac{1}{2}q_\tau^{(2)}(x)(X_i - x)^2 - \frac{1}{6}q_\tau^{(3)}(x)(X_i - x)^3.$$

Then $\bar{\theta}$ minimizes the function

$$G_n(\theta) = \sum_{i=1}^n \left[\rho_\tau\left(Y_i^{(3)} - \frac{\theta' Z_i}{\sqrt{nh_c}}\right) - \rho_\tau\left(Y_i^{(3)}\right)\right] K_i$$

where $K_i = K\left(\frac{X_i - x}{h_c}\right)$. From the definition of ρ_τ it is clear that $G_n(\theta)$ is a convex function of θ. Then, it is sufficient to prove that this function converges pointwise to its conditional expectation, since it follows from the convexity lemma of Pollard (1991) that the convergence is also uniform on any compact set of θ. Let us consider the following decomposition:

$$G_n(\theta) = \mathbb{E}(G_n(\theta)|\mathcal{X}) + \frac{1}{\sqrt{nh_c}}\left(\sum_{i=1}^{n}\psi_\tau\left(Y_i^{(3)}\right)Z_iK_i - \mathbb{E}\left((\psi_\tau\left(Y_i^{(3)}\right)|X_i\right)Z_iK_i\right)'\theta + R_n(\theta) \tag{3.11}$$

where $\mathcal{X} = \{X_1, \ldots, X_n\}$ represents a random sample of the explanatory variable X.

As a consequence of a Taylor expansion of the quantile regression function it can be written that

$$\begin{aligned} q_\tau(X_i) &= q_\tau(x) + q_\tau^{(1)}(x)(X_i - x) + \frac{1}{2}q_\tau^{(2)}(x)(X_i - x)^2 + \frac{1}{6}q_\tau^{(3)}(x)(X_i - x)^3 \\ &+ \frac{1}{24}q_\tau^{(4)}(\xi_{1,i})(X_i - x)^4 \end{aligned} \tag{3.12}$$

where $\xi_{1,i}$ represents an element between X_i and x. Then, equation (3.12) can be rewritten as

$$q_\tau(X_i) = \gamma_0 + \gamma_1(X_i - x) + \gamma_2(X_i - x)^2 + \gamma_3(X_i - x)^3 + \frac{1}{24}q_\tau^{(4)}(\xi_{1,i})(X_i - x)^4$$

where we have assumed the following notation: $\gamma_0 = q_\tau(x)$, $\gamma_1 = q_\tau^{(1)}(x)$, $\gamma_2 = \frac{1}{2}q_\tau^{(2)}(x)$ and $\gamma_3 = \frac{1}{6}q_\tau^{(3)}(x)$.

Furthermore, we should introduce the auxiliary function

$$\varphi_1(t) = \mathbb{E}\left[\rho_\tau(Y - q_\tau(x) + t)|X = x\right] \tag{3.13}$$

whose first derivative is given by

$$\begin{aligned} \varphi_1^{(1)}(t) = \frac{\partial\varphi_1(t)}{\partial t} &= \lim_{z\to 0}\frac{\varphi_1(t+z) - \varphi_1(t)}{z} \\ &= \lim_{z\to 0}\frac{\mathbb{E}\left[\rho_\tau(Y - q_\tau(x) + t + z)|X = x\right] - \mathbb{E}\left[\rho_\tau(Y - q_\tau(x) + t)|X = x\right]}{z} \\ &= \mathbb{E}\left[\lim_{z\to 0}\frac{\rho_\tau(Y - q_\tau(x) + t + z) - \rho_\tau(Y - q_\tau(x) + t)}{z}\,\middle|X = x\right] \\ &= \mathbb{E}\left[\psi_\tau(Y - q_\tau(x) + t)\,|X = x\right] \\ &= \mathbb{E}\left[\tau - \mathbb{I}(Y < q_\tau(x) - t)\,|X = x\right] \\ &= \tau - \mathbb{P}(Y < q_\tau(x) - t\,|X = x) \\ &= \tau - F(q_\tau(x) - t\,|X = x) \end{aligned}$$

as a consequence of the dominated convergence theorem. So,

$$\varphi_1^{(2)}(t) = f(q_\tau(x) - t|X = x)$$

and $\varphi_1^{(2)}(0) = f(q_\tau(x)|X = x)$.

Bearing in mind the definition of function φ_1 given in (3.13), we could write

$$\mathbb{E}(G_n(\theta)|\mathcal{X}) = \sum_{i=1}\mathbb{E}\left[\rho_\tau\left(Y_i^{(3)} - \frac{\theta' Z_i}{\sqrt{nh_c}}\right) - \rho_\tau\left(Y_i^{(3)}\right)\middle|\mathcal{X}\right]K_i$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \left[\varphi_1 \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 - \frac{\theta' Z_i}{\sqrt{nh_c}} \right) \right. \\
&\left. - \varphi_1 \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 \right) \right] K_i \\
&= \sum_{i=1}^{n} \varphi_1^{(1)} \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 \right) \frac{\theta' Z_i}{\sqrt{nh_c}} \, K_i \\
&+ \frac{1}{2} \sum_{i=1}^{n} \varphi_1^{(2)} \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 \right) \frac{(\theta' Z_i)^2}{nh_c} \, K_i \\
&+ \frac{1}{6} \sum_{i=1}^{n} \varphi_1^{(3)} \left(\xi_{2,i} \right) \frac{(\theta' Z_i)^3}{(nh_c)^2} \, K_i \\
&= \sum_{i=1}^{n} C_1(X_i, x) \, \frac{\theta' Z_i}{\sqrt{nh_c}} \, K_i + \frac{1}{2} \sum_{i=1}^{n} C_2(X_i, x) \, \frac{(\theta' Z_i)^2}{nh_c} \, K_i + O\left(\frac{1}{nh_c^2} \right)
\end{aligned}
$$

where

$$
\begin{aligned}
C_1(x) &= \varphi_1^{(1)} \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 \right) \\
C_2(x) &= \varphi_1^{(2)} \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 \right)
\end{aligned}
$$

and $\xi_{2,i}$ is an element between $q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 - \frac{\theta' Z_i}{\sqrt{nh_c}}$ and $q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3$ obtained thanks to a Taylor expansion of φ_1. Then, it follows that

$$
\begin{aligned}
C_1(x) &= \varphi_1^{(1)} \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 \right) \\
&= \mathbb{E} \left[\psi_\tau \left(Y_i^{(3)} \right) \middle| X = X_i \right]. \qquad (3.14)
\end{aligned}
$$

Moreover, in view of (3.12) it follows that

$$
\begin{aligned}
C_2(x) &= \varphi_1^{(2)} \left(q_\tau(X_i) - \gamma_0 - \gamma_1(X_i - x) - \gamma_2(X_i - x)^2 - \gamma_3(X_i - x)^3 \right) \\
&= \varphi_1^{(2)} \left(\frac{1}{24} q_\tau^{(4)}(\xi_{1,i})(X_i - x)^4 \right) \\
&= \varphi_1^{(2)} (0) + \varphi_1^{(3)} \left(\xi_{3,i} \right) \left(\frac{1}{24} q_\tau^{(4)}(\xi_{1,i})(X_i - x)^4 \right) \\
&= f \left(q_\tau(X_i) | X = X_i \right) + O(h_c^4) \qquad (3.15)
\end{aligned}
$$

where $\xi_{3,i}$ is an element between $\frac{1}{24} q_\tau^{(4)}(\xi_{1,i})(X_i - x)^4$ and zero obtained thanks to a Taylor expansion of $\varphi_1^{(2)}$.

Then, in view of expressions (3.14) and (3.15) it follows that

$$
\begin{aligned}
\mathbb{E}[G_n(\theta)|\mathcal{X}] &= \frac{1}{\sqrt{nh_c}} \sum_{i=1}^{n} \mathbb{E} \left[\psi_\tau \left(Y_i^{(3)} \right) \middle| X = X_i \right] (\theta' Z_i) K_i \\
&+ \frac{1}{2nh_c} \theta' \left(\sum_{i=1}^{n} f \left(q_\tau(X_i) | X = X_i \right) Z_i Z_i' K_i \right) \theta + O(h_c^3). \qquad (3.16)
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In order to determine the behaviour of quantity $\sum_{i=1}^n f\left(q_\tau(X_i)|X=X_i\right) Z_i Z_i' K_i$ involved in equation (3.16), it will be crucial the following Lemma:

Lemma 3.1. *Under conditions C1-C4 it follows that*

$$\begin{aligned} S_j &= \frac{1}{nh_c}\sum_{i=1}^n f(q_\tau(X_i)|X=X_i)\left(\frac{X_i-x}{h_c}\right)^j K\left(\frac{X_i-x}{h_c}\right) \\ &= f(q_\tau(x)|X=x)\, g(x)\, \mu_j(K) + o_p(1) \end{aligned}$$

for each $j \in \{0,1,2,3,4,5,6\}$. *Equivalently,*

$$\frac{1}{nh_c}\sum_{i=1}^n f(q_\tau(X_i)|X=X_i)\, K\left(\frac{X_i-x}{h_c}\right) Z_i\, Z_i' = S + o_p(1)$$

where

$$S = f(q_\tau(x)|X=x)\, g(x) \begin{pmatrix} \mu_0(K) & \mu_1(K) & \mu_2(K) & \mu_3(K) \\ \mu_1(K) & \mu_2(K) & \mu_3(K) & \mu_4(K) \\ \mu_2(K) & \mu_3(K) & \mu_4(K) & \mu_5(K) \\ \mu_3(K) & \mu_4(K) & \mu_5(K) & \mu_6(K) \end{pmatrix}$$

and

$$Z_i = \begin{pmatrix} 1 \\ \frac{X_i-x}{h_c} \\ \left(\frac{X_i-x}{h_c}\right)^2 \\ \left(\frac{X_i-x}{h_c}\right)^3 \end{pmatrix}.$$

Proof. We can write

$$\begin{aligned} \mathbb{E}(S_j) &= \mathbb{E}\left[\frac{1}{nh_c}\sum_{i=1}^n f(q_\tau(X_i)|X=X_i)\left(\frac{X_i-x}{h_c}\right)^j K\left(\frac{X_i-x}{h_c}\right)\right] \\ &= \frac{1}{h_c}\,\mathbb{E}\left[f(q_\tau(X_i)|X=X_i)\left(\frac{X_i-x}{h_c}\right)^j K\left(\frac{X_i-x}{h_c}\right)\right] \\ &= \frac{1}{h_c}\int f(q_\tau(z)|X=z)\left(\frac{z-x}{h_c}\right)^j K\left(\frac{z-x}{h_c}\right) g(z)dz \\ &= \frac{1}{h_c}\int f(q_\tau(x+uh_c)|X=x+uh_c)\, u^j\, K(u)\; g(x+uh_c)\, h_c\, du \\ &= f(q_\tau(x)|X=x)\, g(x)\int u^j\, K(u)\; du + o(1) \\ &= f(q_\tau(x)|X=x)\, g(x)\, \mu_j(K) + o(1) \end{aligned}$$

and

$$\begin{aligned} \mathbb{V}\text{ar}\,(S_j) &= \mathbb{E}(S_j^2) - \mathbb{E}(S_j)^2 \le \mathbb{E}(S_j^2) \\ &= \mathbb{E}\left[\left(\frac{1}{nh_c}\sum_{i=1}^n f(q_\tau(X_i)|X=X_i)\left(\frac{X_i-x}{h_c}\right)^j K\left(\frac{X_i-x}{h_c}\right)\right)^2\right] \end{aligned}$$

$$= \frac{1}{(nh_c)^2} \sum_{i=1}^{n} \mathbb{E}\left[\left(f(q_\tau(X_i)|X = X_i) \left(\frac{X_i - x}{h_c}\right)^j K\left(\frac{X_i - x}{h_c}\right)\right)^2\right]$$
$$= \frac{1}{nh_c^2} \int f(q_\tau(z)|X = z)^2 \left(\frac{z - x}{h_c}\right)^{2j} K\left(\frac{z - x}{h_c}\right)^2 g(z)dz$$
$$= \frac{1}{nh_c^2} \int f(q_\tau(x + uh_c)|X = x + uh_c)^2\, u^{2j}\, K(u)^2\, g(x + uh_c)\, h_c\, du$$
$$= \frac{1}{nh_c} f(q_\tau(x)|X = x)^2\, g(x) \int u^{2j}\, K(u)^2\, du + o(1)$$
$$= \frac{1}{nh_c} f(q_\tau(x)|X = x)^2\, g(x)\, \phi_{2j,2}(K) + o(1) = o(1)$$

where $\phi_{i,j}(K) = \int u^i\, K(u)^j\, du$.

Thus,

$$S_j = f(q_\tau(x)|X = x)\, g(x)\, \mu_j(K) + o_p(1).$$

□

Then, equation (3.16) together with Lemma 3.1 end that

$$\mathbb{E}[G_n(\theta)|\mathcal{X}] = \frac{1}{\sqrt{nh_c}} \sum_{i=1}^{n} \mathbb{E}\left[\psi_\tau\left(Y_i^{(3)}\right)\middle| X = X_i\right](\theta' Z_i)K_i + \frac{1}{2}\theta' S\theta + o_p(1). \tag{3.17}$$

Moreover, it is important to note that $\mathbb{E}[R_n(\theta)] = 0$ as a consequence of expression (3.11) and

$$\mathbb{E}[R_n^2(\theta)] \leq n\, \mathbb{E}\left[\left(\rho_\tau\left(Y_i^{(3)} - \frac{\theta' Z_i}{\sqrt{nh_c}}\right) - \rho_\tau\left(Y_i^{(3)}\right) - \psi_\tau\left(Y_i^{(3)}\right)\frac{\theta' Z_i}{\sqrt{nh_c}}\right)^2 K_i^2\right]$$
$$\leq n \int\int \left(\rho_\tau\left(Y_i^{(3)} - \frac{\theta' Z_i}{\sqrt{nh_c}}\right) - \rho_\tau\left(Y_i^{(3)}\right) - \psi_\tau\left(Y_i^{(3)}\right)\frac{\theta' Z_i}{\sqrt{nh_c}}\right)^2$$
$$\times\, \text{Bound}(y|X = x)\, dy\, K\left(\frac{z - x}{h_c}\right)^2 g(z)\, dz$$
$$= O\left(n \int \frac{(\theta' Z_i)^2}{nh_c} K\left(\frac{z - x}{h_c}\right)^2 g(z)\, dz\right) = o(1).$$

Thus $R_n(\theta) = o_p(1)$. This, together with (3.17), leads to

$$G(\theta) = \frac{1}{2}\theta' S\theta + W'\theta + r_n(\theta) \tag{3.18}$$

where $r_n(\theta) = o_p(1)$ for each fixed θ and

$$W = \frac{1}{nh_c} \sum_{i=1}^{n} \psi_\tau\left(Y_i^{(3)}\right) Z_i K_i.$$

It is easy to see that W has a bounded second moment and hence is stochastically bounded. Since the convex function $G_n(\theta) - W'\theta$ converges in probability to the convex function $\frac{1}{2}\theta' S\theta$, it follows from the convexity lemma of Pollard (1991) that for any compact set $\mathcal{K}$

$$\sup_{\theta\in\mathcal{K}} |r_n(\theta)| = o_p(1).$$

That is, the quadratic approximation to the convex function $G_n(\theta)$ holds uniformly for θ in any compact set. Therefore, using the convexity assumption again, the minimizer $\bar{\theta}$ of $G_n(\theta)$ converges in probability to the minimizer $\widehat{\theta} = -S^{-1}W$ of the right-side of (3.18) that is

$$\bar{\theta} - \widehat{\theta} = o_p(1).$$

The third component of the vector $\bar{\theta} - \widehat{\theta}$ is

$$\sqrt{nh_c}\left(h_c^2\left(\widetilde{q}_\tau^{(2)}(x) - \frac{1}{2}q_\tau^{(2)}(x)\right) - \frac{1}{f(q_\tau(x)|X=x)g(x)}V_{\tau,h_c}(x)\right) = o_p(1)$$

where the quantities $V_{\tau,h_c}(x)$ can be simplified as a consequence of the fact that the kernel function K is symmetric, so $\mu_j(K) = 0$ if j is an odd number. As a result

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

where

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

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

Thus, we have determined the behaviour of the second derivative regression estimator, which is presented in the following theorem:

Theorem 3.2. *Under conditions C1-C4, it is verified that*

$$\sqrt{nh_c}\left(h_c^2\left(\widetilde{q}_\tau^{(2)}(x) - \frac{1}{2}q_\tau^{(2)}(x)\right) - \frac{1}{f(q_\tau(x)|X=x)g(x)}V_{\tau,h_c}(x)\right) = o_p(1)$$

where

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

Remark 3.2. Now, the goal will be to compute expectation and variance of the curvature estimator given by (3.10) based on the estimator described in Theorem 3.2. Note that the negligibility of remainders can be establish in L^p, that is, $A_n = o_{L^p}(a_n)$ can be derived from $A_n = o_p(a_n)$ with the condition that $|A_n/a_n|^p$ is uniformly integrable for $p < \infty$ (see page 221 of Gut (2005)). Note that

Definition 3.1. A sequence $A_1, A_2, \dots$ is called **uniformly integrable** if and only if

$$\mathbb{E}\left[|A_n|\,\mathbb{I}(|A_n| > a)\right] \to 0 \quad \text{as } a \to \infty \text{ uniformly in } n.$$

Another, equivalent, way to express uniform integrability is via the distribution function; $A_1, A_2, \dots$ is uniformly integrable if and only if

$$\int_{|x|>a} |x| dF_{A_n}(x) \to 0 \quad \text{as } a \to \infty \text{ uniformly in } n.$$

The assumption that $A_1, A_2, \dots$ have finite mean, implies that $\mathbb{E}\left[|A_n|\,\mathbb{I}(|A_n| > a)\right] \to 0$ as $a \to \infty$ for every n; the tails of convergent integrals converge to 0. The requirement that the sequence is uniformly integrable means that the contributions in the tails of the integrals tend to 0 uniformly for all members of the sequence.

These arguments allow us to prove that convergence established in Theorem 3.2 is also verified in terms of the mean squared convergence. Now, some auxiliary results will be needed to obtain expectation and variance of curvature estimation.

3.3.2 Auxiliary results

In order to study curvature and subsequently sparsity estimators, it will be necessary to compute different kinds of expectations related to the kernel function. These results are contained in the following lemma:

Lemma 3.3. *Under condition C1-C2 and h a bandwidth going to zero, the following statements are verified:*

$$\mathbb{E}\left[(X-x)^i\, K\left(\frac{X-x}{h}\right)^j\right] = g(x)\, h^{i+1}\, \phi_{i,j}(K) + O(h^{i+2})$$

and

$$\mathbb{E}\left[(X-x_1)^i\, K\left(\frac{X-x}{h}\right)^j\, (X-x_2)^l\, K\left(\frac{X-x_2}{h}\right)^k\right] = h^{i+1}\, g(x_1)\, (x_1-x_2)^l$$
$$\times K_i^j * K^k\left(\frac{x_2-x_1}{h_c}\right) + O(h^{i+2})$$

where i, j, k, l are integers, $K_i^j(u) = u^i K(u)^j$, $*$ represents the convolution, and

$$\phi_{i,j}(K) = \int u^i\, K(u)^j du$$

Note that $\phi_{0,2}(K) = R(K)$, $\phi_{i,1}(K) = \mu_i(K)$ and $\max\{j,k\} = 2$.

Proof. Let us consider the following Taylor expansion:

$$g(x+uh) = g(x) + g^{(1)}(\xi_3)\, uh \tag{3.19}$$

where ξ_3 represents an element between x and $x + uh$ and $g^{(1)}(x) = \frac{\partial g(x)}{\partial x}$. Then, the results of this lemma come from a change of variable, equation (3.19) and the fact that $g^{(1)}$ is a bounded function. Insightfully,

$$\begin{aligned}
\mathbb{E}\left[(X-x)^i\, K\left(\frac{X-x}{h}\right)^j\right] &= \int (z-x)^i\, K\left(\frac{z-x}{h}\right)^j\, g(z)\, dz \\
&= \int (uh)^i\, K(u)^j\, g(x+uh)\, h\, du \\
&= \int u^i\, K(u)^j\, \left(g(x) + g^{(1)}(\xi_3)\, u\, h\right)\, h^{i+1}\, du \\
&= \int u^i\, K(u)^j\, g(x)\, h^{i+1}\, du + \int u^{i+1}\, K(u)^j\, g^{(1)}(\xi_3)\, h^{i+2}\, du \\
&= g(x)\, h^{i+1} \int u^i\, K(u)^j\, du + h^{i+2} \int u^{i+1}\, K(u)^j\, g^{(1)}(\xi_3)\, du \\
&= g(x)\, h^{i+1}\, \phi_{i,j}(K) + O(h^{i+2}).
\end{aligned}$$

Analogously, it is verified that

$$\begin{aligned}
&\mathbb{E}\left[(X-x_1)^i\, K\left(\frac{X-x}{h}\right)^j\, (X-x_2)^l\, K\left(\frac{X-x_2}{h}\right)^k\right] = \\
&= \int (z-x_1)^i\, K\left(\frac{z-x_1}{h}\right)^j\, (z-x_2)^l\, K\left(\frac{z-x_2}{h}\right)^k\, g(z)\, dz \\
&= \int (uh)^i\, K(u)^j\, (hu + x_1 - x_2)^l\, K\left(u + \frac{x_1-x_2}{h}\right)^k\, g(x_1+hu)\, h\, du \\
&= h^{i+1}\, g(x_1)\, (x_1-x_2)^l \int u^i\, K(u)^j\, K\left(u + \frac{x_1-x_2}{h}\right)^k\, du + O(h^{i+2}) \\
&= h^{i+1}\, g(x_1)\, (x_1-x_2)^l\, K_i^j * K^k\left(\frac{x_2-x_1}{h}\right) + O(h^{i+2}).
\end{aligned}$$

□

On the other hand, the quantity $V_{\tau,h_c}(x)$ will play a fundamental role in order to compute the first moments of the curvature estimator. Because of this reason, expectation and variance of $V_{\tau,h_c}(x)$ will be established in the following lemma:

Lemma 3.4. *Under conditions C1-C4, it is verified that*

$$\begin{aligned}
\mathbb{E}[V_{\tau,h_c}(x)] &= \frac{1}{24}\, q_\tau^{(4)}(x)\, f\left(q_\tau(x)|X=x\right)\, h_c^4\, g(x)\, \left(\alpha_{31}\, \mu_4(K) + \alpha_{33}\, \mu_6(K)\right) + o(h_c^4) \\
\mathbb{V}ar\left[V_{\tau,h_c}(x)\right] &= \frac{\tau(1-\tau)}{nh_c}\, g(x)\, \left(\alpha_{31}^2\, R(K) + \alpha_{33}^2\, \phi_{4,2}(K) + 2\alpha_{31}\, \alpha_{33}\, \phi_{2,2}(K)\right) + o\left(\frac{1}{nh_c}\right).
\end{aligned}$$

Proof. Firstly, let us recall that the following Taylor expansion of the regression function:

$$q_\tau(X_i) = q_\tau(x) + q_\tau^{(1)}(x)(X_i - x) + \frac{1}{2} q_\tau^{(2)}(x)(X_i-x)^2 + \frac{1}{6} q_\tau^{(3)}(x)(X_i-x)^3$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$+ \frac{1}{24} q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4 \tag{3.20}$$

where $\xi_{4,i}$ is an element between X_i and x.

Expectation of $\mathbf{V}_{\tau,\mathbf{h_c}}(\mathbf{x})$

Let us remember that

$$\mathbb{E}[V_{\tau,h_c}(x)] = \mathbb{E}[\mathbb{E}(V_{\tau,h_c}(x)|\mathcal{X})]$$

where $\mathcal{X} = \{X_1, \ldots, X_n\}$ represents a random sample of the explanatory variable that we have denoted by X. Moreover,

$$\begin{aligned} \mathbb{E}[V_{\tau,h_c}(x)|\mathcal{X}] &= \mathbb{E}\left[\frac{1}{nh_c}\sum_{i=1}^{n} \psi_\tau\left(Y_i^{(3)}\right)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right) K\left(\frac{X_i - x}{h_c}\right)\middle|\mathcal{X}\right] \\ &= \frac{1}{nh_c}\sum_{i=1}^{n} \mathbb{E}\left[\psi_\tau\left(Y_i^{(3)}\right)\middle|\mathcal{X}\right]\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right) K\left(\frac{X_i - x}{h_c}\right) \end{aligned}$$

Then it follows that

$$\begin{aligned} &\mathbb{E}\left[\psi_\tau\left(Y_i^{(3)}\right)\middle|\mathcal{X}\right] = \mathbb{E}\left[\tau - \mathbb{I}\left(Y_i^{(3)} < 0\right)\middle|\mathcal{X}\right] = \tau - \mathbb{E}\left[\mathbb{I}\left(Y_i^{(3)} < 0\right)\middle|\mathcal{X}\right] \\ &= \tau - \mathbb{E}\left[\mathbb{I}\left(Y_i < q_\tau(x) + q_\tau^{(1)}(x)(X_i - x) + \frac{1}{2}q_\tau^{(2)}(x)(X_i - x)^2 + \frac{1}{6}q_\tau^{(3)}(x)(X_i - x)^3\right)\middle|\mathcal{X}\right] \\ &= \tau - F\left(q_\tau(x) + q_\tau^{(1)}(x)(X_i - x) + \frac{1}{2}q_\tau^{(2)}(x)(X_i - x)^2 + \frac{1}{6}q_\tau^{(3)}(x)(X_i - x)^3\middle|X = X_i\right) \\ &= F\left(q_\tau(X_i)|X = X_i\right) - F\left(q_\tau(X_i) - \frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right) \\ &= f\left(q_\tau(X_i)|X = X_i\right)\left(\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right) \\ &+ \frac{1}{2}f^{(1)}\left(\xi_{5,i}|X = X_i\right)\left(\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)^2 \end{aligned} \tag{3.21}$$

where $\xi_{5,i}$ is an element between $q_\tau(X_i)$ and $q_\tau(X_i) - \frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4$. Thus, we are able to compute the expectation of $V_{\tau,h_c}(x)$ as follows

$$\begin{aligned} \mathbb{E}[V_{\tau,h_c}(x)] &= \mathbb{E}[\mathbb{E}(V_{\tau,h_c}(x)|\mathcal{X})] \\ &= \mathbb{E}\left[\mathbb{E}\left(\frac{1}{nh_c}\sum_{i=1}^{n} \psi_\tau\left(Y_i^{(3)}\right)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right) K\left(\frac{X_i - x}{h_c}\right)\middle|\mathcal{X}\right)\right] \\ &= \frac{1}{h_c}\mathbb{E}\left[\mathbb{E}\left(\psi_\tau\left(Y_i^{(3)}\right)\middle|\mathcal{X}\right)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right) K\left(\frac{X_i - x}{h_c}\right)\right] \\ &= \frac{\alpha_{31}}{h_c}\, C_3(x) + \frac{\alpha_{33}}{h_c^3}\, C_4(x) \end{aligned}$$

where

$$\begin{aligned} C_3(x) &= \mathbb{E}\left[\mathbb{E}\left(\psi_\tau\left(Y_i^{(3)}\right)\middle|\mathcal{X}\right)\ K\left(\frac{X_i - x}{h_c}\right)\right] \\ C_4(x) &= \mathbb{E}\left[\mathbb{E}\left(\psi_\tau\left(Y_i^{(3)}\right)\middle|\mathcal{X}\right)\ (X_i - x)^2\, K\left(\frac{X_i - x}{h_c}\right)\right] \end{aligned}$$

Hence, we are going to study the functions C_3 and C_4 independently. On the one hand, from (3.21) and Lemma 3.3 it follows that

$$\begin{aligned}
C_3(x) &= \mathbb{E}\left[\mathbb{E}\left(\psi_\tau\left(Y_i^{(3)}\right)\Big|\mathcal{X}\right)\ K\left(\frac{X_i-x}{h_c}\right)\right] \\
&= \mathbb{E}\Bigg[\Bigg(f\left(q_\tau(X_i)|X=X_i\right)\left(\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i-x)^4\right) \\
&+ \frac{1}{2}f^{(1)}\left(\xi_{5,i}|X=X_i\right)\left(\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i-x)^4\right)\Bigg)\ K\left(\frac{X_i-x}{h_c}\right)\Bigg] \\
&= \int f\left(q_\tau(z)|X=z\right)\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(z-x)^4\ K\left(\frac{z-x}{h_c}\right)g(z)\ dz \\
&+ \int \frac{1}{2}f^{(1)}\left(\xi_{5,i}|X=z\right)\left(\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})\right)^2\ (z-x)^8\ K\left(\frac{z-x}{h_c}\right)\ g(z)\ dz \\
&= C_{3,1}(x) + O(h_c^9)
\end{aligned}$$

where

$$\begin{aligned}
C_{3,1}(x) &= \int f\left(q_\tau(z)|X=z\right)\ \frac{1}{24}\ q_\tau^{(4)}(\xi_{4,i})(z-x)^4\ K\left(\frac{z-x}{h_c}\right)g(z)\ dz \\
&= \frac{1}{24}\left(q_\tau^{(4)}(x)+O(h_c)\right)\int f\left(q_\tau(z)|X=z\right)(z-x)^4\ K\left(\frac{z-x}{h_c}\right)g(z)\ dz \\
&= \frac{1}{24}\left(q_\tau^{(4)}(x)+O(h_c)\right)\int f\left(q_\tau(x+uh_c)|X=x+uh_c\right)(uh_c)^4\ K\left(u\right)g(x+uh_c)\ h_c\ du \\
&= \frac{1}{24}\left(q_\tau^{(4)}(x)+O(h_c)\right)h_c^5\int f\left(q_\tau(x+uh_c)|X=x+uh_c\right)u^4\ K\left(u\right) \\
&\quad\times\left(g(x)+g^{(1)}(\xi_{6,i})\ uh_c\right)du \\
&= \frac{1}{24}\left(q_\tau^{(4)}(x)+O(h_c)\right)h_c^5\ g(x)\int f\left(q_\tau(x+uh_c)|X=x+uh_c\right)u^4\ K\left(u\right)\ du + O(h_c^6) \\
&= \frac{1}{24}\ q_\tau^{(4)}(x)\ g(x)\ h_c^5\ f\left(q_\tau(x)|X=x\right)\ \mu_4(K)+O(h_c^6)
\end{aligned}$$

where we have taken into account that $\xi_{4,i}=x+h_cw_i$ for a certain w_i and $q_\tau^{(4)}$ is a continuous function, so:

$$q_\tau^{(4)}(x+h_cw_i)-q_\tau^{(4)}(x)=O(h_c).$$

On the other hand, based on analogous development we can conclude that

$$\begin{aligned}
C_4(x) &= \mathbb{E}\left[\mathbb{E}\left(\psi_\tau\left(Y_i^{(3)}\right)\Big|\mathcal{X}\right)\ (X_i-x)^2\ K\left(\frac{X_i-x}{h_c}\right)\right] \\
&= \frac{1}{24}\ q_\tau^{(4)}(x)\ g(x)\ h_c^7\ f\left(q_\tau(x)|X=x\right)\ \mu_6(K)+O(h_c^8)
\end{aligned}$$

and finally

$$\mathbb{E}[V_{\tau,h_c}(x)] = \frac{1}{24}\ q_\tau^{(4)}(x)\ f\left(q_\tau(x)|X=x\right)\ h_c^4\ g(x)\ \left(\alpha_{31}\ \mu_4(K)+\alpha_{33}\ \mu_6(K)\right)+o(h_c^4).$$

Variance of $\mathbf{V_{\tau,h_c}(x)}$

The law of total variance allows to write

$$\mathbb{V}\mathrm{ar}\left[V_{\tau,h_c}(x)\right] = \mathbb{E}[\mathbb{V}\mathrm{ar}\left(V_{\tau,h_c}(x)|\mathcal{X}\right)] + \mathbb{V}\mathrm{ar}\left[\mathbb{E}(V_{\tau,h_c}(x)|\mathcal{X})\right]$$

where $\mathcal{X} = \{X_1, \ldots, X_n\}$ represents a random sample of the explanatory variable X. Firstly, we are going to focus on the conditional variance of $V_{\tau,h_c}(x)$. Let us define the following auxiliary function:

$$\begin{aligned}\varphi_2(t) &= \mathbb{V}\mathrm{ar}\left[\mathbb{I}(Y - q_\tau(x) < t)|X = x\right] \\ &= F(q_\tau(x) + t|X = x)(1 - F(q_\tau(x) + t|X = x)) \end{aligned} \tag{3.22}$$

whose first derivative is given by

$$\begin{aligned}\varphi_2^{(1)}(t) = \frac{\partial}{\partial t}\varphi_2(t) &= \left.\frac{\partial F(z|X = x)(1 - F(z|X = x))}{\partial z}\right|_{z=q_\tau(x)+t} \\ &= \Big[f(z|X = x)(1 - F(z|X = x)) - F(z|X = x)f(z|X = x)\Big]_{z=q_\tau(x)+t} \\ &= f(q_\tau(x) + t|X = x)(1 - 2F(q_\tau(x) + t|X = x)).\end{aligned}$$

So, taking into account the definition of the function φ_2 and the Taylor expansion described in (3.20), it follows that

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\mathbb{I}\left(Y_i^{(3)} < 0\right)\Big|\mathcal{X}\right] &= \mathbb{V}\mathrm{ar}\left[\mathbb{I}\left(Y_i - q_\tau(X_i) < -\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)\Big|\mathcal{X}\right] \\ &= \varphi_2\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right) \\ &= \varphi_2(0) + \varphi_2^{(1)}(\xi_{6,i})\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right) \\ &= F(q_\tau(X_i)|X = X_i)(1 - F(q_\tau(X_i)|X = X_i)) \\ &+ \varphi_2^{(1)}(\xi_{6,i})\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right) \\ &= \tau(1 - \tau) + \varphi_2^{(1)}(\xi_{6,i}|X = X_i)\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)\end{aligned} \tag{3.23}$$

where a Taylor expansion of function φ_2 has been developed and $\xi_{6,i}$ represents an element between $-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4$ and 0. So, it is verified that

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[V_{\tau,h_c}(x)|\mathcal{X}\right] &= \mathbb{V}\mathrm{ar}\left[\frac{1}{nh_c}\sum_{i=1}^n \psi_\tau\left(Y_i^{(3)}\right)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right) K\left(\frac{X_i - x}{h_c}\right)\Big|\mathcal{X}\right] \\ &= \frac{1}{(nh_c)^2}\sum_{i=1}^n \mathbb{V}\mathrm{ar}\left(\mathbb{I}\left(Y_i^{(3)} < 0\right)\Big|\mathcal{X}\right)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 K_i^2 \\ &= \frac{1}{nh_c^2}\left(\tau(1 - \tau) + \varphi_2^{(1)}(\xi_{6,i})\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)\right) \\ &\times\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 K\left(\frac{X_i - x}{h_c}\right)^2\end{aligned}$$

and

$$\mathbb{E}\left[\mathbb{V}\mathrm{ar}\left(V_{\tau,h_c}(x)|\mathcal{X}\right)\right] = \mathbb{E}\Bigg[\frac{1}{nh_c^2}\left(\tau(1 - \tau) + \varphi_2^{(1)}(\xi_{6,i})\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)\right)$$

$$\times \left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 K\left(\frac{X_i - x}{h_c}\right)^2\Bigg]$$
$$= C_5(x) + C_6(x)$$

where

$$C_5(x) = \mathbb{E}\Bigg[\frac{1}{nh_c^2}\, \tau(1-\tau)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 K\left(\frac{X_i - x}{h_c}\right)^2\Bigg]$$

$$C_6(x) = \mathbb{E}\Bigg[\frac{1}{nh_c^2}\, \varphi_2^{(1)}(\xi_{6,i})\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 \times K\left(\frac{X_i - x}{h_c}\right)^2\Bigg].$$

So, the study of the conditional variance has been reduced to computing C_5 and C_6. Firstly, the following development comes from Lemma 3.3:

$$\begin{aligned}
C_5(x) &= \mathbb{E}\Bigg[\frac{1}{nh_c^2}\, \tau(1-\tau)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 K\left(\frac{X_i - x}{h_c}\right)^2\Bigg] \\
&= \frac{1}{nh_c^2}\, \tau(1-\tau)\, \alpha_{31}^2\, \mathbb{E}\left[K\left(\frac{X_i - x}{h_c}\right)^2\right] \\
&+ \frac{1}{nh_c^6}\, \tau(1-\tau)\, \alpha_{33}^2\, \mathbb{E}\left[(X_i - x)^4\, K\left(\frac{X_i - x}{h_c}\right)^2\right] \\
&+ 2\, \frac{1}{nh_c^4}\, \tau(1-\tau)\, \alpha_{31}\, \alpha_{33}\, \mathbb{E}\left[(X_i - x)^2\, K\left(\frac{X_i - x}{h_c}\right)^2\right] \\
&= \frac{1}{nh_c}\, \tau(1-\tau)\, g(x)\left(\alpha_{31}^2\, R(K) + \alpha_{33}^2\, \phi_{4,2}(K) + 2\alpha_{31}\, \alpha_{33}\, \phi_{2,2}(K)\right) + O\left(\frac{1}{n}\right).
\end{aligned}$$

On the other hand, it follows that

$$\begin{aligned}
C_6(x) &= \mathbb{E}\Bigg[\frac{1}{nh_c^2}\, \varphi_2^{(1)}(\xi_{6,i})\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 K_i^2\Bigg] \\
&\le M_1\, M_2\, \frac{1}{nh_s^2}\mathbb{E}\left[(X_i - x)^4\left(\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right)^2 K_i^2\right] \\
&\le M_1\, M_2\, \frac{1}{nh_s^2}\Bigg[\alpha_{31}^2\mathbb{E}\left[(z - x)^4\, K\left(\frac{z - x}{h_c}\right)^2\right] \\
&+ \frac{1}{h_c^4}\, \alpha_{33}^2\, \mathbb{E}\left[(z - x)^8\, K\left(\frac{z - x}{h_c}\right)^2\right] \\
&+ \frac{2}{h_c^2}\, \alpha_{33}\, \alpha_{31}\, \mathbb{E}\left[(z - x)^6\, K\left(\frac{z - x}{h_c}\right)^2\right]\Bigg] \\
&\le M_1\, M_2\, \frac{1}{nh_s^2}\left(O(h_c^5) + O(h_c^9) + O(h_c^7)\right) = O\left(\frac{h_c^3}{n}\right)
\end{aligned}$$

where M_1 and M_2 represent upper bounds of the functions $\varphi_2^{(1)}$ and $q_\tau^{(4)}$, respectively. So, we conclude that

$$\mathbb{E}\left[\mathbb{V}\text{ar}\left(V_{\tau,h_c}(x)|\mathcal{X}\right)\right] = \frac{\tau(1-\tau)}{nh_c}\, g(x) \left(\alpha_{31}^2\, R(K) + \alpha_{33}^2\, \phi_{4,2}(K) + 2\alpha_{31}\, \alpha_{33}\, \phi_{2,2}(K)\right) + o\left(\frac{1}{nh_c}\right). \tag{3.24}$$

Finally, in order to finish the calculus of the variance of $V_{\tau,h_c}(x)$, we should introduce the following auxiliary function:

$$\begin{aligned}\varphi_3(t) &= \mathbb{E}(\psi_\tau(Y - q_\tau(x) - t)|X = x) = \tau - \mathbb{E}(\mathbb{I}(Y < q_\tau(x) + t)|X = x) \\ &= \tau - F(q_\tau(x) + t|X = x)\end{aligned}$$

and then it is verified that

$$\begin{aligned}\mathbb{V}\text{ar}\left(\mathbb{E}[V_{\tau,h_c}(x)|\mathcal{X}]\right) &= \mathbb{E}(\mathbb{E}[V_{\tau,h_c}(x)|\mathcal{X}]^2) - (\mathbb{E}[V_{\tau,h_c}(x)])^2 \leq \mathbb{E}(\mathbb{E}[V_{\tau,h_c}|\mathcal{X}]^2) \\ &= \frac{1}{(nh_c)^2}\, n\, \mathbb{E}\left(\mathbb{E}\left[\psi_\tau\left(Y_i^{(3)}\right)\middle|\mathcal{X}\right]\left[\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right] K_i\right)^2 \\ &= \frac{1}{nh_c^2}\, \mathbb{E}\left(\varphi_3\left(-\frac{1}{24} q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4\right)\left[\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right] K_i\right)^2 \\ &= \frac{1}{nh_c^2}\, \mathbb{E}\left[\varphi_3^{(1)}(\xi_{7,i})^2 \left(\frac{1}{24} q_\tau^{(4)}(\xi_{4,i})(z - x)^4\right)^2 \left[\alpha_{31} + \alpha_{33}\left(\frac{z - x}{h_c}\right)^2\right]^2 K_i^2\right] \\ &\leq M_1\, M_2\, \frac{1}{nh_c^2}\, \mathbb{E}\left[(X_i - x)^8 \left[\alpha_{31} + \alpha_{33}\left(\frac{X_i - x}{h_c}\right)^2\right]^2 K_i^2\right] \\ &= O\left(\frac{h_c^9}{nh_c^2}\right) = O\left(\frac{h_c^7}{n}\right)\end{aligned} \tag{3.25}$$

where we have considered a Taylor expansion of the function φ_3 and $\xi_{7,i}$ represents an element between $-\frac{1}{24} q_\tau^{(4)}(\xi_{4,i})(X_i - x)^4$ and 0. Note that

$$\varphi_3(0) = \tau - F(q_\tau(x)|X = x) = \tau - \tau = 0.$$

Moreover, M_3 and M_2 are upper bounds of $\varphi_3^{(1)}$ and $q_\tau^{(4)}$, respectively. So finally, in view of (3.25) and (3.24) we conclude that

$$\mathbb{V}\text{ar}\left[V_{\tau,h_c}(x)\right] = \frac{\tau(1-\tau)}{nh_c}\, g(x)\left(\alpha_{31}^2\, R(K) + \alpha_{33}^2\, \phi_{4,2}(K) + 2\alpha_{31}\, \alpha_{33}\, \phi_{2,2}(K)\right) + o\left(\frac{1}{nh_c}\right).$$

□

3.3.3 Bias and variance of the curvature estimator

In this subsection, the mean integrated squared error of the curvature estimator will be studied. Let us recall that

$$\text{MSE}\left[\widehat{\vartheta}_{h_c}\right] = \left(\mathbb{B}\text{ias}\left[\widehat{\vartheta}_{h_c}\right]\right)^2 + \mathbb{V}\text{ar}\left[\widehat{\vartheta}_{h_c}\right].$$

Firstly, the bias of the curvature estimator is presented in the following theorem:

Theorem 3.5. *Under conditions C1-C4, it follows that*

$$\mathbb{B}ias\left[\widehat{\vartheta}_{h_c}\right] \cong \frac{1}{6}\, h_c^2\, \delta_1 \int q_\tau^{(2)}(x)\, q_\tau^{(4)}(x)\, g(x)\, dx + \frac{4}{nh_c^5}\, \tau(1-\tau)\, \delta_2 \int \frac{1}{f(q_\tau(x)|X=x)^2}\, dx$$

where

$$\begin{aligned}
\delta_1 &= \alpha_{31}\, \mu_4(K) + \alpha_{33}\, \mu_6(K)\\
\delta_2 &= \alpha_{31}^2\, R(K) + \alpha_{33}^2\, \phi_{4,2}(K) + 2\alpha_{31}\, \alpha_{33}\, \phi_{2,2}(K).
\end{aligned}$$

Proof. First of all, we should remember that

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

So, we are going to start computing the expectation and the variance of $\widetilde{q}_{\tau,h_c}^{(2)}(x)$. Given x, from Lemma 3.4 and Theorem 3.2 it follows that

$$\begin{aligned}
\mathbb{E}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right] &\cong \mathbb{E}\left[q_\tau^{(2)}(x) + 2h_c^{-2}\frac{1}{f(q_\tau(x)|X=x)g(x)}V_{\tau,h_c}(x)\right]\\
&= q_\tau^{(2)}(x) + 2h_c^{-2}\frac{1}{f(q_\tau(x)|X=x)g(x)}\, \mathbb{E}\left[V_{\tau,h_c}(x)\right]\\
&= q_\tau^{(2)}(x) + \frac{1}{12}\, h_c^2\, q_\tau^{(4)}(x)\, \left(\alpha_{31}\, \mu_4(K) + \alpha_{33}\, \mu_6(K)\right) + o(h_c^2)\\
&= q_\tau^{(2)}(x) + \frac{1}{12}\, h_c^2\, q_\tau^{(4)}(x)\, \delta_1 + o(h_c^2)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right] &\cong \mathbb{V}\mathrm{ar}\left[q_\tau^{(2)}(x) + 2h_c^{-2}\frac{1}{f(q_\tau(x)|X=x)g(x)}V_{\tau,h_c}(x)\right]\\
&= 4h_c^{-4}\frac{1}{f(q_\tau(x)|X=x)^2 g(x)^2}\, \mathbb{V}\mathrm{ar}\left[V_{\tau,h_c}(x)\right]\\
&= \frac{4}{f(q_\tau(x)|X=x)^2 g(x)}\, \frac{1}{nh_c^5}\, \tau(1-\tau)\, \delta_2 + o\left(\frac{1}{nh_c^5}\right).
\end{aligned}$$

Moreover, as consequence of previous expressions, it can be established that

$$\begin{aligned}
\mathbb{E}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)^2\right] &= \left(\mathbb{E}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right]\right)^2 + \mathbb{V}\mathrm{ar}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right]\\
&= \left(q_\tau^{(2)}(x) + \mathbb{B}\mathrm{ias}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right]\right)^2 + \mathbb{V}\mathrm{ar}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)^2\right]\\
&= q_\tau^{(2)}(x)^2 + 2q_\tau^{(2)}(x)\, \mathbb{B}\mathrm{ias}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right]\\
&+ \mathbb{B}\mathrm{ias}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right]^2 + \mathbb{V}\mathrm{ar}\left[\widetilde{q}_{\tau,h_c}^{(2)}(x)\right]\\
&\cong q_\tau^{(2)}(x)^2 + 2q_\tau^{(2)}(x)\, \frac{1}{12}\, h_c^2\, q_\tau^{(4)}(x)\, \delta_1 + \left(\frac{1}{12}\, h_c^2\, q_\tau^{(4)}(x)\, \delta_1\right)^2
\end{aligned}$$

$$+ \frac{4}{f(q_\tau(x)|X=x)^2 g(x)} \frac{1}{nh_c^5} \tau(1-\tau)\delta_2 + o\left(\frac{1}{nh_c^5}\right).$$

So, it can be concluded that

$$\begin{aligned}
\mathbb{E}\left[\widehat{\vartheta}_{h_c}\right] &= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \widetilde{q}^{(2)}_{\tau,h_c}(X_i)^2\right] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[\widetilde{q}^{(2)}_{\tau,h_c}(X_i)^2\right] = \mathbb{E}\left[\widetilde{q}^{(2)}_{\tau,h_c}(X_i)^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\widetilde{q}^{(2)}_{\tau,h_c}(X_i)^2 \middle| X = X_i\right)\right] = \int \mathbb{E}\left[\widetilde{q}^{(2)}_{\tau,h_c}(z)^2\right] \, g(z)\, dz \\
&\cong \int q_\tau^{(2)}(x)^2 \, g(x)\, dx + \frac{1}{6}\, h_c^2\, \delta_1 \int q_\tau^{(2)}(x)\, q_\tau^{(4)}(x)\, g(x)\, dx \\
&+ \frac{4}{nh_c^5}\, \tau(1-\tau)\, \delta_2 \int \frac{1}{f(q_\tau(x)|X=x)^2}\, dx
\end{aligned}$$

and as a consequence

$$\mathbb{B}\text{ias}\left[\widehat{\vartheta}_{h_c}\right] \cong \frac{h_c^2}{6}\, \delta_1 \int q_\tau^{(2)}(x)\, q_\tau^{(4)}(x)\, g(x)\, dx + \frac{4\tau(1-\tau)}{nh_c^5}\, \delta_2 \int \frac{1}{f(q_\tau(x)|X=x)^2}\, dx$$

□

Now we are going to move to the calculus of the variance of the curvature estimator. Firstly, as a consequence of variance definition, we are able to obtain the following development:

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\widehat{\vartheta}_{h_c}\right] &= \mathbb{V}\text{ar}\left[\frac{1}{n}\sum_{i=1}^{n} \widetilde{q}^{(2)}_{\tau,h_c}(X_i)^2\right] = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[\widetilde{q}^{(2)}_{\tau,h_c}(X_i)^2, \widetilde{q}^{(2)}_{\tau,h_c}(X_j)^2\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[\left(q_\tau^{(2)}(X_i) + 2h_c^{-2}\frac{s_\tau(X_i)}{g(X_i)} V_{\tau,h_c}(X_i)\right)^2, \right. \\
&\left. \left(q_\tau^{(2)}(X_j) + 2h_c^{-2}\frac{s_\tau(X_j)}{g(X_j)} V_{\tau,h_c}(X_j)\right)^2\right] \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[4h_c^{-2}\, q_\tau^{(2)}(X_i)\, \frac{s_\tau(X_i)}{g(X_i)} V_{\tau,h_c}(X_i) + 4h_c^{-4}\frac{s_\tau(X_i)^2}{g(X_i)^2}\, V_{\tau,h_c}(X_i)^2, \right. \\
&\left. 4h_c^{-2}\, q_\tau^{(2)}(X_j)\, \frac{s_\tau(X_j)}{g(X_j)}\, V_{\tau,h_c}(X_i) + 4h_c^{-4}\frac{s_\tau(X_j)^2}{g(X_j)^2}\, V_{\tau,h_c}(X_j)^2\right] \\
&= C_7 + 2C_8 + C_9
\end{aligned} \qquad (3.26)$$

where

$$C_7 = \frac{16}{n^2 h_c^4}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[q_\tau^{(2)}(X_i)\, \frac{s_\tau(X_i)^2}{g(X_i)} V_{\tau,h_c}(X_i), q_\tau^{(2)}(X_j)\, \frac{s_\tau(X_j)}{g(X_j)}\, V_{\tau,h_c}(X_j)\right]$$

$$C_8 = \frac{16}{n^2 h_c^6}\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[q_\tau^{(2)}(X_i)\, \frac{s_\tau(X_i)^2}{g(X_i)} V_{\tau,h_c}(X_i), \frac{s_\tau(X_j)^2}{g(X_j)^2}\, V_{\tau,h_c}(X_j)^2\right]$$

$$C_9 = \frac{16}{n^2 h_c^8} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{C}\mathrm{ov} \left[\frac{s_\tau(X_i)^2}{g(X_i)^2} \; V_{\tau,h_c}(X_i)^2, \frac{s_\tau(X_j)^2}{g(X_j)^2} \; V_{\tau,h_c}(X_j)^2 \right].$$

In view of previous expression, it will be necessary to compute three different "kinds" of covariances between V_{τ,h_c} quantities. We are going to analyse each scenario independently. For simplicity, let us introduce the following notation:

$$\widetilde{K}_{i,j} = \left(\alpha_{31} + \alpha_{33} \left(\frac{X_i - x_j}{h_c} \right)^2 \right) K_{i,j} \quad \text{where} \quad K_{i,j} = K\left(\frac{X_i - x_j}{h_c} \right)$$

$$Y_i^{(3,j)} = Y_i - q_\tau(x_j) - q_\tau^{(1)}(x_j)(X_i - x_j) - \frac{1}{2} q_\tau^{(2)}(x_j)(X_i - x_j)^2 - \frac{1}{6} q_\tau^{(3)}(x_j)(X_i - x_j)^3.$$

We are going to start with the calculus of C_7. This result is presented in the following proposition:

Proposition 3.6. *Under conditions C1-C4 it follows that*

$$C_7 \cong \frac{4\tau(1-\tau)}{n} \int (q_\tau^{(2)} \; s_\tau)^{(2)}(x)^2 \; g(x) \; dx.$$

Proof. Let us remember that as a consequence of the law of total covariance it can be written that

$$\begin{aligned} \mathbb{C}\mathrm{ov} \left[V_{\tau,h_c}(x_1), V_{\tau,h_c}(x_2) \right] = & \; \mathbb{E} \left(\mathbb{C}\mathrm{ov} \left[V_{\tau,h_c}(x_1), V_{\tau,h_c}(x_2) \middle| \mathcal{X} \right] \right) \\ & + \mathbb{C}\mathrm{ov} \left[\mathbb{E} \left(V_{\tau,h_c}(x_1) \middle| \mathcal{X} \right), \mathbb{E} \left(V_{\tau,h_c}(x_2) \middle| \mathcal{X} \right) \right]. \end{aligned}$$

So, we are going to start computing the conditional covariance. That is,

$$\begin{aligned} & \frac{16}{n^2 h_c^4} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{C}\mathrm{ov} \left[q_\tau^{(2)}(X_i) \; \frac{s_\tau(X_i)}{g(X_i)} \; V_{\tau,h_c}(X_i), q_\tau^{(2)}(X_j) \; \frac{s_\tau(X_j)}{g(X_j)} \; V_{\tau,h_c}(X_j) \middle| \mathcal{X} \right] = \\ & = \frac{16}{n^2 h_c^4} \sum_{i=1}^{n} \sum_{j=1}^{n} q_\tau^{(2)}(X_i) \; \frac{s_\tau(X_i)}{g(X_i)} \; q_\tau^{(2)}(X_j) \; \frac{s_\tau(X_j)}{g(X_j)} \\ & \quad \times \mathbb{C}\mathrm{ov} \left[\frac{1}{n h_c} \sum_{k=1}^{n} \psi_\tau \left(Y_k^{(3,i)} \right) \; \widetilde{K}_{k,i}, \frac{1}{n h_c} \sum_{l=1}^{n} \psi_\tau \left(Y_l^{(3,j)} \right) \; \widetilde{K}_{l,j} \right] \\ & = \frac{16}{n^4 h_c^6} \sum_{i=1}^{n} \sum_{j=1}^{n} q_\tau^{(2)}(X_i) \; \frac{s_\tau(X_i)}{g(X_i)} \; q_\tau^{(2)}(X_j) \; \frac{s_\tau(X_j)}{g(X_j)} \sum_{k=1}^{n} \sum_{l=1}^{n} \\ & \quad \times \mathbb{C}\mathrm{ov} \left[\psi_\tau \left(Y_k^{(3,i)} \right), \psi_\tau \left(Y_l^{(3,j)} \right) \middle| \mathcal{X} \right] \; \widetilde{K}_{k,i} \; \widetilde{K}_{l,j} \\ & = \frac{16}{n^4 h_c^6} \sum_{i=1}^{n} \sum_{j=1}^{n} q_\tau^{(2)}(X_i) \; \frac{s_\tau(X_i)}{g(X_i)} q_\tau^{(2)}(X_j) \; \frac{s_\tau(X_j)}{g(X_j)} \sum_{l=1}^{n} \\ & \quad \times \mathbb{C}\mathrm{ov} \left[\mathbb{I} \left(Y_l^{(3,i)} < 0 \right), \mathbb{I} \left(Y_l^{(3,j)} < 0 \right) \middle| \mathcal{X} \right] \; \widetilde{K}_{l,i} \; \widetilde{K}_{l,j}. \end{aligned}$$

Then, let us define the following auxiliary function:

$$\begin{aligned}\varphi_4(t_1,t_2) &= \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y<q_\tau(x)+t_1\right),\mathbb{I}\left(Y<q_\tau(x)+t_2\right)|X=x\right]\\ &= \mathbb{E}\left[\mathbb{I}\left(Y<q_\tau(x)+t_1\right)\mathbb{I}\left(Y<q_\tau(x)+t_2\right)|X=x\right]\\ &- \mathbb{E}\left[\mathbb{I}\left(Y<q_\tau(x)+t_1\right)|X=x\right]\mathbb{E}\left[\mathbb{I}\left(Y<q_\tau(x)+t_2\right)X=x\right]\\ &= F\left(q_\tau(x)+\min\{t_1,t_2\}|X=x\right)\\ &- F\left(q_\tau(x)+t_1|X=x\right)F\left(q_\tau(x)+t_2|X=x\right)\end{aligned}$$

whose partial derivatives are given by

$$\begin{aligned}\frac{\partial\varphi_4(t_1,t_2)}{\partial t_1} &= f\left(q_\tau(x)+\min\{t_1,t_2\}|X=x\right)\mathbb{I}(t_1\le t_2)\\ &- f\left(q_\tau(x)+t_1|X=x\right)F\left(q_\tau(x)+t_2|X=x\right)\\ \frac{\partial\varphi_4(t_1,t_2)}{\partial t_2} &= f\left(q_\tau(x)+\min\{t_1,t_2\}|X=x\right)\mathbb{I}(t_2\le t_1)\\ &- f\left(q_\tau(x)+t_2|X=x\right)F\left(q_\tau(x)+t_1|X=x\right).\end{aligned}$$

Thereupon, in view of arguments developed in equation (3.20) and a Taylor expansion of function φ_4 we can determine that

$$\begin{aligned}&\mathbb{C}\mathrm{ov}\left[\left(\mathbb{I}\left(Y_l^{(3,i)}<0\right)\right),\left(\mathbb{I}\left(Y_l^{(3,j)}<0\right)\right)\Big|\mathcal{X}\right]\\ &=\varphi_4\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{9,l})(X_l-X_i)^4,-\frac{1}{24}q_\tau^{(4)}(\xi_{10,l})(X_l-X_j)^4\right)\\ &=\varphi_4(0,0)+\left(\frac{\partial\varphi_4(t_1,t_2)}{\partial t_1},\frac{\partial\varphi_4(t_1,t_2)}{\partial t_2}\right)'\Bigg|_{(t1,t2)=(\xi_{11,l},\xi_{12,l})}\\ &\times\left(-\frac{1}{24}q_\tau^{(4)}(\xi_{9,i})(X_l-X_i)^4,-\frac{1}{24}q_\tau^{(4)}(\xi_{10,i})(X_l-X_j)^4\right)\end{aligned}$$

where $\xi_{11,l}$ and $\xi_{12,l}$ are elements between $-\frac{1}{24}q_\tau^{(4)}(\xi_{9,l})(X_l-X_i)^4$ and 0 and between $-\frac{1}{24}q_\tau^{(4)}(\xi_{10,l})(X_l-X_j)^4$ and 0, respectively. Moreover,

$$\varphi_4(0,0)=F\left(q_\tau(x)|X=x\right)-F\left(q_\tau(x)|X=x\right)F\left(q_\tau(x)|X=x\right)=\tau(1-\tau).$$

So, we are able to establish that

$$\mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l^{(3,i)}<0\right),\mathbb{I}\left(Y_l^{(3,j)}<0\right)\Big|\mathcal{X}\right]=\tau(1-\tau)+O(h_c^4), \qquad (3.27)$$

and as a result

$$\begin{aligned}&\mathbb{E}\left(\frac{16}{n^2h_c^4}\sum_{i=1}^n\sum_{j=1}^n q_\tau^{(2)}(X_i)\,\frac{s_\tau(X_i)}{g(X_i)}q_\tau^{(2)}(X_j)\,\frac{s_\tau(X_j)}{g(X_j)}\mathbb{C}\mathrm{ov}\left[V_{\tau,h_c}(X_i),V_{\tau,h_c}(X_j)|\mathcal{X}\right]\right)\\ &\cong\frac{16\tau(1-\tau)}{n^4h_c^6}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\mathbb{E}\left[q_\tau^{(2)}(X_i)\,\frac{s_\tau(X_i)}{g(X_i)}\,q_\tau^{(2)}(X_j)\,\frac{s_\tau(X_j)}{g(X_j)}\widetilde{K}_{l,i}\,\widetilde{K}_{l,j}\right]\\ &\cong\frac{16\tau(1-\tau)}{nh_c^6}\mathbb{E}\left[q_\tau^{(2)}(X_i)\,\frac{s_\tau(X_i)}{g(X_i)}\,q_\tau^{(2)}(X_j)\,\frac{s_\tau(X_j)}{g(X_j)}\widetilde{K}_{l,i}\,\widetilde{K}_{l,j}\right]\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$= \frac{16\tau(1-\tau)}{nh_c^6} \int\int\int q_\tau^{(2)}(X_i)\, \frac{s_\tau(X_i)}{g(X_i)}\, q_\tau^{(2)}(X_j)\, \frac{s_\tau(X_j)}{g(X_j)} \left[\alpha_{31} + \alpha_{33}\left(\frac{X_l - X_i}{h_c}\right)^2\right]$$
$$\times K_{l,i} \left[\alpha_{31} + \alpha_{33}\left(\frac{X_l - X_j}{h_c}\right)^2\right] K_{l,j}\, g(X_i, X_j, X_l)\, dX_i\, dX_j\, dX_l$$
$$= \frac{16\tau(1-\tau)}{nh_c^6} \int\int\int q_\tau^{(2)}(X_l - uh_c)\, q_\tau^{(2)}(X_l - vh_c)\, s_\tau(X_l - uh_c)\, s_\tau(X_l - vh_c)$$
$$\times \left[\alpha_{31} + \alpha_{33}\, u^2\right]\, K(u) \left[\alpha_{31} + \alpha_{33}\, v^2\right]\, K(v)\, g(X_l)\, h_c^2\, du\, dv\, dX_l$$
$$= \frac{16\tau(1-\tau)}{nh_c^4} \int \left[\int q_\tau^{(2)}(X_l - uh_c)\, s_\tau(X_l - uh_c) \left[\alpha_{31} + \alpha_{33}\, u^2\right] K(u)\, du\right]^2 g(X_l)\, dX_l$$
$$\cong \frac{4\tau(1-\tau)}{n} \int (q_\tau^{(2)}\, s_\tau)^{(2)}(X_l)^2\, g(X_l)\, dX_l$$

where the last step of previous development comes from a Taylor expansion of the function $q_\tau^{(2)}\, s_\tau$ and the following equalities

$$\int \left[\alpha_{31} + \alpha_{33}\, u^2\right]\, K(u)\, du = \alpha_{31} + \alpha_{33}\, \mu_2(K) = 0$$
$$\int u\, \left[\alpha_{31} + \alpha_{33}\, u^2\right]\, K(u)\, du = \alpha_{31}\, \mu_1(K) + \alpha_{33}\, \mu_3(K) = 0$$
$$\int u^2\, \left[\alpha_{31} + \alpha_{33}\, u^2\right]\, K(u)\, du = \alpha_{31}\, \mu_2(K) + \alpha_{33}\, \mu_4(K) = 1. \tag{3.28}$$

Note that because of the fact that the dominant term of previous statement (that is, the first term of the Taylor expansion associated with $q_\tau^{(2)}\, s_\tau$) is zero, we could think about to consider another term in the Taylor expansion associated with the auxiliary function φ_4 defined in order to compute (3.27). This expansion will provide terms with order h_c^4 that is smaller than the term that provides the considered Taylor expansion of the function $q_\tau^{(2)}\, s_\tau$ (with order h_c^2).

On the other side, it will be necessary to compute the covariance of the conditional expectations where the calculus of $\mathbb{E}[V_{\tau,h_c}(x)|\mathcal{X}]$ has been detailed in Lemma 3.4. Then, it follows

$$\frac{16}{n^2h_c^4} \sum_{i=1}^n \sum_{j=1}^n \mathbb{C}\text{ov} \left[\mathbb{E}\left(q_\tau^{(2)}(X_i)\, \frac{s_\tau(X_i)}{g(X_i)}\, V_{\tau,h_c}(X_i)\middle|\mathcal{X}\right), \mathbb{E}\left(q_\tau^{(2)}(X_j)\, \frac{s_\tau(X_j)}{g(X_j)}\, V_{\tau,h_c}(X_j)\middle|\mathcal{X}\right)\right]$$
$$= \frac{16}{n^2h_c^4} \sum_{i=1}^n \sum_{j=1}^n \mathbb{C}\text{ov} \left[q_\tau^{(2)}(X_i)\, \frac{s_\tau(X_i)}{g(X_i)} \mathbb{E}\left(V_{\tau,h_c}(X_i)\middle|\mathcal{X}\right), q_\tau^{(2)}(X_j)\, \frac{s_\tau(X_j)}{g(X_j)} \mathbb{E}\left(V_{\tau,h_c}(X_j)\middle|\mathcal{X}\right)\right]$$
$$= \frac{16}{n^2h_c^4} \sum_{i=1}^n \sum_{j=1}^n \mathbb{C}\text{ov} \left[q_\tau^{(2)}(X_i)\, \frac{s_\tau(X_i)}{g(X_i)}\, \frac{1}{nh_c} \sum_{k=1}^n f\left(q_\tau(X_k)|X = X_k\right) \frac{1}{24} q_\tau^{(4)}(\xi_{4,k})\right.$$
$$\times (X_k - X_i)^4\, \widetilde{K}_{k,i}, q_\tau^{(2)}(X_j)\, \frac{s_\tau(X_j)}{g(X_j)}\, \frac{1}{nh_c} \sum_{l=1}^n f\left(q_\tau(X_l)|X = X_l\right)$$
$$\left.\times \frac{1}{24} q_\tau^{(4)}(\xi_{4,l}) (X_l - X_j)^4\, \widetilde{K}_{l,j}\right]$$
$$\leq \frac{16}{n^2h_c^4} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{1}{nh_c}\right)^2\, M_5^2\, M_4^2 \sum_{l=1}^n \left(\mathbb{C}\text{ov} \left[(X_l - X_i)^4 \widetilde{K}_{l,i}, (X_l - X_j)^4\, \widetilde{K}_{i,1}\right]\right. \tag{3.29}$$

where M_4 and M_5 represents bounds of the functions $q_\tau^{(4)}$ and $q_\tau^{(2)}$, respectively. In view of previous expression, it will be crucial to compute

$$\mathbb{C}\text{ov}\left[(X_l-x_1)^4\widetilde{K}_{l,1},(X_l-x_2)^4\,\widetilde{K}_{l,2}\right]=\mathbb{E}\left[(X_l-x_1)^4\,\widetilde{K}_{l,1}\,(X_l-x_2)^4\,\widetilde{K}_{l,2}\right]$$
$$-\,\mathbb{E}\left[(X_l-x_1)^4\widetilde{K}_{l,1}\right]\,\mathbb{E}\left[(X_l-x_2)^4\,\widetilde{K}_{l,2}\right].$$

Hence, bearing in mind the definition of $\widetilde{K}_{i,j}$ and Lemma 3.3 it follows that

$$\begin{aligned}\mathbb{E}\left[(X_l-x_1)^j\widetilde{K}_{l,1}\,(X_l-x_2)^l\,\widetilde{K}_{l,2}\right]&=\alpha_{31}^2\,\mathbb{E}\left[(X_l-x_1)^j\,K_{l,1}\,(X_l-x_2)^l\,K_{l,2}\right]\\&+\frac{\alpha_{31}\,\alpha_{33}}{h_c^2}\,\mathbb{E}\left[(X_l-x_1)^j\,K_{l,1}\,(X_l-x_2)^{l+2}\,K_{l,2}\right]\\&+\frac{\alpha_{31}\,\alpha_{33}}{h_c^2}\,\mathbb{E}\left[(X_l-x_1)^{j+2}\,K_{l,1}\,(X_l-x_2)^l\,K_{l,2}\right]\\&+\frac{\alpha_{33}^2}{h_c^4}\,\mathbb{E}\left[(X_l-x_1)^{j+2}\,K_{l,1}\,(X_l-x_2)^{l+2}\,K_{l,2}\right]\\&=O(h_c^{j+l+1})\end{aligned}\qquad(3.30)$$

and

$$\begin{aligned}\mathbb{E}\left[(X_l-x_1)^j\widetilde{K}_{l,1}\right]&=\alpha_{31}\,\mathbb{E}\left[(X_l-x_1)^j\,K_{l,1}\right]+\frac{\alpha_{33}}{h_c^2}\,\mathbb{E}\left[(X_l-x_1)^{j+2}\,K_{l,1}\right]\\&=O(h_c^{j+1})+\frac{1}{h_c^2}\,O(h_c^{j+3})=O(h_c^{j+1}).\end{aligned}\qquad(3.31)$$

Then, we can conclude that

$$\mathbb{C}\text{ov}\left[(X_l-x_1)^4\widetilde{K}_{l,1},(X_l-x_2)^4\,\widetilde{K}_{l,2}\right]=O(h_c^9)-O(h_c^{10})$$

and as a consequence of (3.29) it follows that

$$\begin{aligned}&\frac{16}{n^2h_c^4}\sum_{i=1}^n\sum_{j=1}^n\mathbb{C}\text{ov}\left[\mathbb{E}\left(q_\tau^{(2)}(X_i)\,\frac{s_\tau(X_i)}{g(X_i)}\,V_{\tau,h_c}(X_i)\middle|\mathcal{X}\right),\mathbb{E}\left(q_\tau^{(2)}(X_j)\,\frac{s_\tau(X_j)}{g(X_j)}\,V_{\tau,h_c}(X_j)\middle|\mathcal{X}\right)\right]\\&\le\frac{16}{nh_c^4}\,O(h_c^9)=O\left(\frac{h_c^5}{n}\right)=o\left(\frac{1}{n}\right).\end{aligned}$$

Summarizing, the following approximation is justified

$$\begin{aligned}C_7&\cong\mathbb{E}\left(\frac{16}{n^2h_c^4}\sum_{i=1}^n\sum_{j=1}^n\mathbb{C}\text{ov}\left[q_\tau^{(2)}(X_i)\,\frac{s_\tau(X_i)}{g(X_i)}\,V_{\tau,h_c}(X_i),q_\tau^{(2)}(X_j)\,\frac{s_\tau(X_j)}{g(X_j)}\,V_{\tau,h_c}(X_j)\middle|\mathcal{X}\right]\right)\\&\cong\frac{4\tau(1-\tau)}{n}\int(q_\tau^{(2)}\,s_\tau)^{(2)}(x)^2\,g(x)\,dx.\end{aligned}$$

□

Now we are going to move to the calculus associated with C_8. The following proposition shows an approximation of this quantity:

Proposition 3.7. *Under conditions C1-C4 it follows that*

$$C_8 \cong \frac{8(2\tau^2-\tau)(1-\tau)}{n^2h_c^5}\Gamma_1(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)\ s_\tau(x)^2\ dx$$
$$+\ \frac{2\tau(1-\tau)}{n}\ \Gamma_2(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)\ s_\tau(x)\ q_\tau^{(4)}(x)\ g(x)\ dx$$

where

$$\Gamma_1(K) = \int \left[\alpha_{31}+\alpha_{33}\ v^2\right]^2\ K(v)^2\,dv = \alpha_{31}^2\ R(K)+\alpha_{33}^2\ \phi_{4,2}+2\alpha_{31}\ \alpha_{33}\ \phi_{2,2}$$
$$\Gamma_2(K) = \int\int (v-w)^4\ \left[\alpha_{31}+\alpha_{33}\ (v)^2\right]\ \left[\alpha_{31}+\alpha_{33}\ (w)^2\right]\ K(v)\ K(w)\ dv\ dw.$$

Proof. Let us remember that

$$C_8 = \frac{16}{n^2h_c^6}\sum_{i=1}^n\sum_{j=1}^n \mathbb{C}\mathrm{ov}\left[q_\tau^{(2)}(X_i)\ \frac{s_\tau(X_i)}{g(X_i)}V_{\tau,h_c}(X_i),\ \frac{s_\tau(X_j)^2}{g(X_j)^2}\ V_{\tau,h_c}(X_j)^2\right].$$

Analogous to the calculus associated with C_7 (Proposition 3.6) we are going to apply the law of total covariance. Let us begin with the computation of the conditional covariance. In this case, it follows that

$$\frac{16}{n^2h_c^6}\sum_{i=1}^n\sum_{j=1}^n \mathbb{C}\mathrm{ov}\left[q_\tau^{(2)}(X_i)\ \frac{s_\tau(X_i)}{g(X_i)}\ V_{\tau,h_c}(X_i),\ \frac{s_\tau(X_j)^2}{g(X_j)^2}\ V_{\tau,h_c}(X_j)\middle|\mathcal{X}\right] =$$
$$= \frac{16}{n^2h_c^6}\sum_{i=1}^n\sum_{j=1}^n q_\tau^{(2)}(X_i)\ \frac{s_\tau(X_i)}{g(X_i)}\ \frac{s_\tau(X_j)^2}{g(X_j)^2}$$
$$\times\ \mathbb{C}\mathrm{ov}\left[\frac{1}{nh_c}\sum_{k=1}^n \psi_\tau\left(Y_k^{(3,i)}\right)\ \widetilde{K}_{k,i},\ \frac{1}{nh_c}\sum_{l=1}^n \psi_\tau\left(Y_l^{(3,j)}\right)\ \widetilde{K}_{l,j}\frac{1}{nh_c}\sum_{m=1}^n \psi_\tau\left(Y_m^{(3,j)}\right)\ \widetilde{K}_{m,j}\right]$$
$$= \frac{16}{n^5h_c^9}\sum_{i=1}^n\sum_{j=1}^n q_\tau^{(2)}(X_i)\ \frac{s_\tau(X_i)}{g(X_i)}\ \frac{s_\tau(X_j)^2}{g(X_j)^2}\sum_{k=1}^n\sum_{l=1}^n\sum_{m=1}^n$$
$$\times\ \mathbb{C}\mathrm{ov}\left[\psi_\tau\left(Y_k^{(3,i)}\right),\psi_\tau\left(Y_l^{(3,j)}\right)\psi_\tau\left(Y_m^{(3,j)}\right)\middle|\mathcal{X}\right]\ \widetilde{K}_{k,i}\ \widetilde{K}_{l,j}\ \widetilde{K}_{m,j}.$$

At this point, we have to deal with the computation of

$$C_{8,1}(X_i,X_j) = \mathbb{C}\mathrm{ov}\left[\psi_\tau\left(Y_l^{(3,i)}\right)\psi_\tau\left(Y_m^{(3,j)}\right)\ ,\ \psi_\tau\left(Y_k^{(3,i)}\right)\middle|\mathcal{X}\right]$$

for different values of the indices $\{k,l,m\}$. On the one hand, if $\mathrm{card}(\{k,l,m\}) = 3$[1] or $\mathrm{card}(\{k,l,m\}) = 2$ joint with $l=m$, then $C_{8,1}(X_i,X_j)=0$ because of the independence. As a consequence, the situations that contribute dominant terms will be $\mathrm{card}(\{k,l,m\})=2$ with $l\neq m$ and $k=l=m$. We are going to analyse each of the previous cases. First if $l\neq m=k$ it follows that

$$C_{8,1}(X_i,X_j) = \mathbb{C}\mathrm{ov}\left[\psi_\tau\left(Y_l^{(3,j)}\right)\psi_\tau\left(Y_k^{(3,j)}\right)\ ,\ \psi_\tau\left(Y_k^{(3,i)}\right)\middle|\mathcal{X}\right]$$

[1]Note that $\mathrm{card}(A)$ represents the number of different points of a subset A.

$$= \mathbb{E}\left[\psi_\tau\left(Y_l^{(3,j)}\right)\mathcal{X}\right] \mathbb{C}\mathrm{ov}\left[\psi_\tau\left(Y_k^{(3,j)}\right) , \psi_\tau\left(Y_k^{(3,i)}\right)\middle|\mathcal{X}\right]$$
$$\cong \frac{\tau(1-\tau)}{24} f(q_\tau(X_j)|X = X_j)\, q_\tau^{(4)}(X_j)(X_l - X_j)^4$$

as a result of Lemma 3.4 and equation (3.27). Secondly, if $k = l = m$ it can be written

$$\mathbb{C}\mathrm{ov}\left[\psi_\tau\left(Y_l^{(3,j)}\right)\psi_\tau\left(Y_l^{(3,j)}\right) , \psi_\tau\left(Y_l^{(3,i)}\right)\middle|\mathcal{X}\right] =$$
$$= \mathbb{C}\mathrm{ov}\left[\left(\tau - \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\right)\left(\tau - \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\right) , \tau - \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\middle|\mathcal{X}\right]$$
$$= 2\tau\, \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l^{(3,j)} < 0\right) , \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\middle|\mathcal{X}\right]$$
$$- \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_l^{(3,j)} < 0\right) , \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\middle|\mathcal{X}\right] \tag{3.32}$$

where in view of the arguments developed along Proposition 3.6 it follows that

$$\mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right), \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\middle|\mathcal{X}\right] = \tau(1-\tau) + O(h_c^4).$$

In addition, applying the same arguments developed previously in order to compute C_7, it can be established that

$$\mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_l^{(3,j)} < 0\right) , \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\middle|\mathcal{X}\right]$$
$$= \mathbb{E}\left[\mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\middle|\mathcal{X}\right]$$
$$- \mathbb{E}\left[\mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_l^{(3,j)} < 0\right)\right] \mathbb{E}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\middle|\mathcal{X}\right]$$
$$\cong \tau(1-\tau) \tag{3.33}$$

In order to prove (3.33), let us define the following auxiliary function:

$$\begin{aligned}\varphi_5(t_1,t_2,t_3) &= \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l < q_\tau(X_l) + t_1\right)\mathbb{I}\left(Y_l < q_\tau(X_l) + t_2\right), \mathbb{I}\left(Y_l < q_\tau(X_l) + t_3\right)|\mathcal{X}\right]\\ &= \mathbb{E}\left[\mathbb{I}\left(Y_l < q_\tau(X_l) + t_1\right)\ \mathbb{I}\left(Y_l < q_\tau(X_l) + t_2\right)\ \mathbb{I}\left(Y_l < q_\tau(X_l) + t_3\right)|\mathcal{X}\right]\\ &- \mathbb{E}\left[\mathbb{I}\left(Y_l < q_\tau(X_l) + t_1\right)\mathbb{I}\left(Y_l < q_\tau(X_l) + t_2\right)|\mathcal{X}\right]\\ &\times \mathbb{E}\left[\mathbb{I}\left(Y_l < q_\tau(X_l) + t_3\right)|\mathcal{X}\right]\\ &= F\left(\min\{(q_\tau(X_l) + t_1, q_\tau(X_l) + t_2, q_\tau(X_l) + t_3\}|X = X_l\right)\\ &- F\left(\min\{q_\tau(X_l) + t_1, q_\tau(X_l) + t_2\}|X = X_l\right) F\left(q_\tau(X_l) + t_3|X = X_l\right)\end{aligned}$$

where

$$\begin{aligned}\varphi_5(0,0,0) &= \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l < q_\tau(X_l)\right)\mathbb{I}\left(Y_l < q_\tau(X_l)\right) , \mathbb{I}\left(Y_l < q_\tau(X_l)\right)|\mathcal{X}\right]\\ &= F\left(q_\tau(x)|X = x\right) - F\left(q_\tau(x)|X = x\right)^2\\ &= \tau(1-\tau).\end{aligned}$$

Then, applying a Taylor expansion of the function φ_5 it follows that

$$\mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_l^{(3,j)} < 0\right), \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\middle|\mathcal{X}\right]$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$= \varphi_5 \left(-\frac{1}{24} q_\tau^{(4)}(\xi_{13,l})(X_l - X_j)^4, -\frac{1}{24} q_\tau^{(4)}(\xi_{14,l})(X_l - X_j)^4, -\frac{1}{24} q_\tau^{(4)}(\xi_{15,l})(X_l - X_i)^4 \right)$$
$$= \varphi_5(0,0,0) + \left(\frac{\partial \varphi_5(t_1,t_2,t_3)}{\partial t_1}, \frac{\partial \varphi_5(t_1,t_2,t_3)}{\partial t_2}, \frac{\partial \varphi_5(t_1,t_2,t_3)}{\partial t_3} \right)' \Bigg|_{(t_1,t_2,t_3)=(\xi_{16,l},\xi_{17,j},\xi_{18,j})}$$
$$\times \left(-\frac{1}{24} q_\tau^{(4)}(\xi_{13,l})(X_l - X_j)^4, -\frac{1}{24} q_\tau^{(4)}(\xi_{14,l})(X_l - X_j)^4, -\frac{1}{24} q_\tau^{(4)}(\xi_{15,l})(X_l - X_i)^4 \right)$$
$$\cong \tau(1-\tau)$$

where $\xi_{16,l}$, $\xi_{17,l}$ and $\xi_{18,l}$ are elements between $-\frac{1}{24} q_\tau^{(4)}(\xi_{13,l})(X_l - X_j)^4$, $-\frac{1}{24} q_\tau^{(4)}(\xi_{14,l})(X_l - X_j)^4$ and $-\frac{1}{24} q_\tau^{(4)}(\xi_{15,l})(X_l - X_i)^4$ and zero, respectively.

As a result, combining equations (3.32) and (3.33), it follows that

$$C_{8,1}(X_i, X_j) = \mathbb{C}\mathrm{ov} \left[\psi_\tau \left(Y_l^{(3,j)} \right) \psi_\tau \left(Y_k^{(3,j)} \right) \ , \ \psi_\tau \left(Y_m^{(3,i)} \right) \Big| \mathcal{X} \right]$$
$$\cong \begin{cases} \frac{\tau(1-\tau)}{24} \ f(q_\tau(X_j)|X = X_j) \ q_\tau^{(4)}(X_j)(X_l - X_j)^4 & \text{if } l \neq k = m \\ & \text{(analogous } k \neq l = m) \\ (2\tau^2 - \tau)(1-\tau) & \text{if } l = k = m \end{cases}$$

Bearing previous developments in mind, we can write the following decomposition of the expectation of the conditional covariances:

$$\mathbb{E}\left(\frac{16}{n^2 h_c^6} \sum_{i=1}^n \sum_{j=1}^n q_\tau^{(2)}(X_i) \ \frac{s_\tau(X_i)}{g(X_i)} \ \frac{s_\tau(X_j)^2}{g(X_j)^2} \mathbb{C}\mathrm{ov} \left[V_{\tau,h_c}(X_i), V_{\tau,h_c}(X_j)^2 | \mathcal{X} \right] \right) \cong C_{8,2} + C_{8,3}$$

where

$$C_{8,2} = \frac{16(2\tau^2 - \tau)(1-\tau)}{n^5 h_c^9} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \mathbb{E}\left[q_\tau^{(2)}(X_i) \ \frac{s_\tau(X_i)}{g(X_i)} \ \frac{s_\tau(X_j)^2}{g(X_j)^2} \widetilde{K}_{l,i} \ \widetilde{K}_{l,j} \ \widetilde{K}_{l,j} \right]$$

$$C_{8,3} = \frac{4\tau(1-\tau)}{3n^5 h_c^9} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{k=1}^n \mathbb{E}\left[q_\tau^{(2)}(X_i) \ \frac{s_\tau(X_i)}{g(X_i)} \ q_\tau^{(4)}(X_j)(X_k - X_j)^4 \ \frac{s_\tau(X_j)}{g(X_j)^2} \widetilde{K}_{l,j} \ \widetilde{K}_{k,j} \ \widetilde{K}_{l,i} \right].$$

On the one hand, based on arguments detailed in equation (3.28), it follows that

$$C_{8,2} = \frac{16(2\tau^2 - \tau)(1-\tau)}{n^5 h_c^9} \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \mathbb{E}\left[q_\tau^{(2)}(X_i) \ \frac{s_\tau(X_i)}{g(X_i)} \ \frac{s_\tau(X_j)^2}{g(X_j)^2} \widetilde{K}_{l,i} \ \widetilde{K}_{l,j} \ \widetilde{K}_{l,j} \right]$$
$$= \frac{16(2\tau^2 - \tau)(1-\tau)}{n^2 h_c^9} \int\int\int q_\tau^{(2)}(X_i) \ \frac{s_\tau(X_i)}{g(X_i)} \ \frac{s_\tau(X_j)^2}{g(X_j)^2} \left[\alpha_{31} + \alpha_{33} \left(\frac{X_l - X_i}{h_c} \right)^2 \right]$$
$$\times K_{l,i} \left[\alpha_{31} + \alpha_{33} \left(\frac{X_l - X_j}{h_c} \right)^2 \right]^2 \ K_{l,j}^2 \ g(X_i, X_j, X_l) \ dX_i \ dX_j \ dX_l$$
$$= \frac{16(2\tau^2 - \tau)(1-\tau)}{n^2 h_c^9} \int\int\int q_\tau^{(2)}(X_l - uh_c) \ s_\tau(X_l - uh_c) \ \frac{s_\tau(X_l - vh_c)^2}{g(X_l - vh_c)}$$

$$\times \left[\alpha_{31} + \alpha_{33}\, u^2\right]\; K(u) \left[\alpha_{31} + \alpha_{33}\, v^2\right]^2\; K(v)^2\, g(X_l)\, h_c^2\, du\, dv\, dX_l$$
$$= \frac{16(2\tau^2 - \tau)(1-\tau)}{n^2 h_c^7} \int \left[\int q_\tau^{(2)}(X_l - uh_c)\, s_\tau(X_l - uh_c) \left[\alpha_{31} + \alpha_{33}\, u^2\right]\; K(u)\; du\right]$$
$$\times \left[\int \frac{s_\tau(X_l - vh_c)^2}{g(X_l - vh_c)} \left[\alpha_{31} + \alpha_{33}\, v^2\right]^2\; K(v)^2\; dv\right] g(X_l)\, dX_l$$
$$\cong \frac{8(2\tau^2 - \tau)(1-\tau)}{n^2 h_c^5} \Gamma_1(K) \int (q_\tau^{(2)}\; s_\tau)^{(2)}(X_l)\; s_\tau(X_l)^2\, dX_l$$

where the last step is a consequence of equalities (3.28) and

$$\Gamma_1(K) = \int \left[\alpha_{31} + \alpha_{33}\, v^2\right]^2\; K(v)^2\, dv = \alpha_{31}^2\; R(K) + \alpha_{33}^2\; \phi_{4,2} + 2\alpha_{31}\; \alpha_{33}\; \phi_{2,2}.$$

On the other hand, using similar arguments the following approximation of $C_{8,3}$ can be justified

$$C_{8,3} = \frac{4\tau(1-\tau)}{3n^5 h_c^9} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{l=1}^{n} \mathbb{E}\left[q_\tau^{(2)}(X_i)\; \frac{s_\tau(X_i)}{g(X_i)}\; q_\tau^{(4)}(X_j)(X_k - X_j)^4\; \frac{s_\tau(X_j)}{g(X_j)^2} \widetilde{K}_{l,j}\; \widetilde{K}_{k,j}\; \widetilde{K}_{l,i}\right]$$
$$= \frac{32\tau(1-\tau)}{nh_c^9} \int\int\int\int q_\tau^{(2)}(X_i)\; \frac{s_\tau(X_i)}{g(X_i)}\; q_\tau^{(4)}(X_j)(X_k - X_j)^4\; \frac{s_\tau(X_j)}{g(X_j)^2}$$
$$\times\, \widetilde{K}_{l,j}\; \widetilde{K}_{k,j}\; \widetilde{K}_{l,i}\; g(X_i, X_j, X_l, X_k)\; dX_i\, dX_j\, dX_l\, dX_k$$
$$= \frac{4\tau(1-\tau)}{3nh_c^9} \int\int \left[\int q_\tau^{(2)}(X_l - uh_c)\; s_\tau(X_l - uh_c) \left[\alpha_{31} + \alpha_{33}\, u^2\right]\; K(u)\; h_c\, du\right]$$
$$\times \left[\int q_\tau^{(4)}(X_l - vh_c)\; (X_k - X_l + vh_c)^4\; \frac{s_\tau(X_l - vh)}{g(X_l - vh)}\; \left[\alpha_{31} + \alpha_{33}\, v^2\right]\; K(v)\; h_c\, dv\right]$$
$$\times \left[\alpha_{31} + \alpha_{33}\; \left(\frac{X_l - X_k}{h_c}\right)^2\right]\; K\left(\frac{X_l - X_k}{h_c}\right)\; g(X_l)\; g(X_k)\; dX_l\, dX_k$$
$$= \frac{2\tau(1-\tau)}{3nh_c^3} \int\int (q_\tau^{(2)}\; s_\tau)^{(2)}(X_l)$$
$$\left[\int q_\tau^{(4)}(X_l - vh_c)\; (-wh_c + vh_c)^4\; \frac{s_\tau(X_l - vh)}{g(X_l - vh)}\; \left[\alpha_{31} + \alpha_{33}\, v^2\right]\; K(v)\; h_c\, dv\right]$$
$$\times\, g(X_l - wh_c)\; \left[\alpha_{31} + \alpha_{33}\; (w)^2\right]\; K(w)\; h_c\, dw\, g(X_l)\, dX_l$$
$$= \frac{2\tau(1-\tau)}{n}\; \Gamma_2(K) \int (q_\tau^{(2)}\; s_\tau)^{(2)}(X_l)\; s_\tau(X_l)\; q_\tau^{(4)}(X_l)\; g(X_l)\, dX_l$$

where

$$\Gamma_2(K) = \int\int (v-w)^4\; \left[\alpha_{31} + \alpha_{33}\; (v)^2\right]\; \left[\alpha_{31} + \alpha_{33}\; (w)^2\right]\; K(v)\, K(w)\, dv\, dw.$$

Finally, it will be necessary to compute the covariance between the conditional expectations. So, it should be taken into account that

$$\mathbb{E}[V_{\tau,h_c}(x)^2|\mathcal{X}] = \mathbb{E}\left[\frac{1}{nh_s}\; \sum_{i=1}^{n} \psi_\tau\left(Y_i^{(3)}\right)\; \widetilde{K}_i\; \frac{1}{nh_s}\; \sum_{j=1}^{n} \psi_\tau\left(Y_j^{(3)}\right)\; \widetilde{K}_j \middle| \mathcal{X}\right]$$

$$
\begin{aligned}
&= \left(\frac{1}{nh_c}\right)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left[\psi_\tau\left(Y_i^{(3)}\right)\psi_\tau\left(Y_j^{(3)}\right)\Big|\mathcal{X}\right]\ \widetilde{K}_i\ \widetilde{K}_j \\
&= \left(\frac{1}{nh_c}\right)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \left(\tau^2 - 2\tau\,\mathbb{E}\left[\mathbb{I}\left(Y_j^{(3)}<0\right)\Big|\mathcal{X}\right]\right. \\
&\left. + \mathbb{E}\left[\mathbb{I}\left(Y_i^{(3)}<0\right)\ \mathbb{I}\left(Y_j^{(3)}<0\right)\Big|\mathcal{X}\right]\right)\ \widetilde{K}_i\ \widetilde{K}_j \\
&\cong \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} \tau(1-\tau)\ \widetilde{K}_i^2. \qquad (3.34)
\end{aligned}
$$

As a result,

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[\mathbb{E}\left(\frac{s_\tau(X_j)^2}{g(X_j)^2}V_{\tau,h_v}(X_j)^2\Big|\mathcal{X}\right), \mathbb{E}\left(q_\tau^{(2)}(X_i)\ \frac{s_\tau(X_i)}{g(X_i)}V_{\tau,h_c}(X_i)\Big|\mathcal{X}\right)\right] \\
&\quad \mathbb{C}\mathrm{ov}\left[\frac{s_\tau(X_j)^2}{g(X_j)^2}\ \mathbb{E}\left(V_{\tau,h_v}(X_j)^2\Big|\mathcal{X}\right), q_\tau^{(2)}(X_i)\ \frac{s_\tau(X_i)}{g(X_i)}\ \mathbb{E}\left(V_{\tau,h_c}(X_i)\Big|\mathcal{X}\right)\right] \\
&\quad \cong \mathbb{C}\mathrm{ov}\left[\frac{s_\tau(X_j)^2}{g(X_j)^2}\ \left(\frac{1}{nh_s}\right)^2 \sum_{l=1}^{n}\tau(1-\tau)\ \widetilde{K}_{l,j}^2,\right. \\
&\qquad\qquad \left. q_\tau^{(2)}(X_i)\ \frac{s_\tau(X_i)}{g(X_i)}\ \frac{1}{nh_s}\sum_{l=1}^{n} f(q_\tau(X_i)|X=X_i)\ \frac{1}{24}q_\tau^{(4)}(\xi_{4,l})(X_l-X_i)^4\ \widetilde{K}_{l,i}\right] \\
&\quad \le \left(\frac{1}{nh_s}\right)^3 \tau(1-\tau)\ h_c^4\ M_1^2\ M_2\ M_3^3\ \ M_4\ n\ \mathbb{C}\mathrm{ov}\left[\widetilde{K}_{i,1}^2, K_{i,2}\right] \\
&\quad = \left(\frac{1}{nh_s}\right)^3 \tau(1-\tau)\ h_c^4\ n\ O(h_c) = O\left(\frac{h_c^2}{n^2}\right) = o\left(\frac{1}{n^2h_c^5}\right)
\end{aligned}
$$

where M_1^{-1}, M_2, M_3^{-1} and M_4 represents bounds of the function f, $q_\tau^{(4)}$, g and $q_\tau^{(2)}$. Moreover, the last step comes from similar developments as the established in equation (3.30) related to the kernel function. That is,

$$
\mathbb{C}\mathrm{ov}\left[\widetilde{K}_{i,1}^2, \widetilde{K}_{i,2}\right] = \mathbb{E}[\widetilde{K}_{i,1}^2\ \widetilde{K}_{i,2}] - \mathbb{E}[\widetilde{K}_{i,1}^2]\ \mathbb{E}[\widetilde{K}_{i,2}] = O(h_c) - O(h_c)O(h_c) = O(h_c)
$$

Finally, it can be concluded that

$$
\begin{aligned}
C_8 \cong C_8^2 + C_8^3 \cong &\ \frac{8(2\tau^2-\tau)(1-\tau)}{n^2h_c^5}\Gamma_1(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(X_l)\ s_\tau(X_l)^2\ dX_l \\
&+ \frac{2\tau(1-\tau)}{n}\ \Gamma_2(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(X_l)\ s_\tau(X_l)\ q_\tau^{(4)}(X_l)\ g(X_l)\ dX_l
\end{aligned}
$$

□

Finally, we are going to analyse the last part of the variance of the curvature estimator that is C_9. This result is presented in the following proposition

Proposition 3.8. *Under conditions C1-C4, it follows that*

$$
C_9 \cong \frac{32}{n^2h_c^9}\ \tau^2(\tau^2-2\tau+1)\ \Gamma_3(K)^2\int s_\tau(x)^4\ dx
$$

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

where

$$\Gamma_3(K) = \int\int [\alpha_{31} + \alpha_{33}u^2]K(u)[\alpha_{31} + \alpha_{33}(u+w)^2]K(u+w)\, dw\, du.$$

Proof. In order to describe C_9 that is given by

$$C_9 = \frac{16}{n^2 h_c^8} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[\frac{s_\tau(X_i)^2}{g(X_i)^2}\, V_{\tau,h_c}(X_i)^2, \frac{s_\tau(X_j)^2}{g(X_j)^2}\, V_{\tau,h_c}(X_j)^2\right],$$

we are going to use the law of total covariance, so we start computing the conditional covariance. That is,

$$\frac{16}{n^2 h_c^8} \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\left[\frac{s_\tau(X_i)^2}{g(X_i)^2}\, V_{\tau,h_c}(X_i)^2, \frac{s_\tau(X_j)^2}{g(X_j)^2}\, V_{\tau,h_c}(X_j)^2 \middle| \mathcal{X}\right]$$
$$= \frac{16}{n^2 h_c^8} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{s_\tau(X_i)^2}{g(X_i)^2}\,\frac{s_\tau(X_j)^2}{g(X_j)^2}\, \mathbb{C}\text{ov}\left[V_{\tau,h_c}(X_i)^2, V_{\tau,h_c}(X_j)^2 \middle| \mathcal{X}\right]$$
$$= \frac{16}{n^6 h_c^{12}} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{s_\tau(X_i)^2}{g(X_i)^2}\,\frac{s_\tau(X_j)^2}{g(X_j)^2} \sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n}\sum_{r=1}^{n}$$
$$\times\, \mathbb{C}\text{ov}\left[\psi_\tau\left(Y_k^{(3,i)}\right)\, \psi_\tau\left(Y_l^{(3,i)}\right), \psi_\tau\left(Y_m^{(3,j)}\right) \psi_\tau\left(Y_r^{(3,j)}\right) \middle| \mathcal{X}\right] \widetilde{K}_{k,i}\, \widetilde{K}_{l,i}\, \widetilde{K}_{m,j}\, \widetilde{K}_{r,j}$$

So, it will be crucial to compute

$$C_{9,1}(X_i, X_j) = \mathbb{C}\text{ov}\left[\psi_\tau\left(Y_k^{(3,i)}\right)\, \psi_\tau\left(Y_l^{(3,i)}\right), \psi_\tau\left(Y_m^{(3,j)}\right) \psi_\tau\left(Y_r^{(3,j)}\right) \middle| \mathcal{X}\right]$$

for different values of the indices $\{l,k,m,r\}$. Firstly, if $\text{card}(\{l,k,m,r\}) = 4$ or $\text{card}(\{l,k,m,r\}) = 3$ and $l = k$ or $m = r$ then $C_{9,1}(X_i, X_j) = 0$ because of independence of observations $Y_1, \ldots, Y_n$. Secondly, if $\text{card}(\{l,k,m,r\}) = 3$ and $l \neq k$ or $m \neq r$ then $C_{9,1}(X_i, X_j) \neq 0$ and we have to compute it. For instance, we are going to check the scenario in which $l \neq \{k,m,r\}$ and $k = m$. In this case, as a consequence of arguments developed in Lemma 3.4 and Proposition 3.6 it follows that

$$C_{9,1}(X_i, X_j) = \mathbb{C}\text{ov}\left[\psi_\tau\left(Y_l^{(3,i)}\right) \psi_\tau\left(Y_k^{(3,i)}\right)\ ,\ \psi_\tau\left(Y_r^{(3,j)}\right)\ \psi_\tau\left(Y_k^{(3,j)}\right) \middle| \mathcal{X}\right]$$
$$= \mathbb{E}\left[\psi_\tau\left(Y_l^{(3,i)}\right) \middle| \mathcal{X}\right]\, \mathbb{E}\left[\psi_\tau\left(Y_r^{(3,j)}\right) \middle| \mathcal{X}\right] \mathbb{C}\text{ov}\left[\psi_\tau\left(Y_k^{(3,i)}\right)\ ,\ \psi_\tau\left(Y_k^{(3,j)}\right)\ \psi_\tau \mathcal{X}\right]$$
$$\cong f(q_\tau(X_i)|X = X_i)\ \frac{1}{24} q_\tau^{(4)}(X_i)\ (X_l - X_i)^4 f(q_\tau(X_j)|X = X_j)$$
$$\times\ \frac{1}{24} q_\tau^{(4)}(X_j)\ (X_r - X_i)^4 \tau(1-\tau).$$

Hence, we are going to move to the scenario in which $\text{card}(\{l,k,m,r\}) = 2$. Note that if $l = k \neq r = m$ then $C_{9,1}(X_i, X_j) = 0$ because of independence. So, it can be concluded the other addends that contribute dominated terms to the calculus of the conditional covariance will be those associated with $l = r \neq m = k$, $l = m \neq k = r$ and $l = k = r = m$. Using similar arguments to those associated with the auxiliary functions φ_4 and φ_5, we can conclude that

$$\mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \mathbb{I}\left(Y_k^{(3,i)} < 0\right)\ ,\ \mathbb{I}\left(Y_m^{(3,j)} < 0\right) \mathbb{I}\left(Y_r^{(3,j)} < 0\right) \middle| \mathcal{X}\right]$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\mathbb{I}\left(Y_k^{(3,i)} < 0\right)\mathbb{I}\left(Y_m^{(3,j)} < 0\right)\mathbb{I}\left(Y_r^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&- \mathbb{E}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\mathbb{I}\left(Y_k^{(3,i)} < 0\right)\right]\mathbb{E}\left[\mathbb{I}\left(Y_m^{(3,j)} < 0\right)\mathbb{I}\left(Y_r^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&\cong \begin{cases} \tau(1-\tau) & \text{if } l = k = m = r \\ \tau^2(1-\tau^2) & \text{if } l = m \neq k = r \text{ or } l = r \neq k = m \end{cases}
\end{aligned}
$$

Then, the case in which $l = m \neq k = r$ will be studied more insightfully. In this context, it can be written

$$
\begin{aligned}
C_{9,1}(X_i, X_j) &= \mathbb{C}\text{ov}\left[\psi_\tau\left(Y_l^{(3,i)}\right)\psi_\tau\left(Y_k^{(3,i)}\right)\,,\ \psi_\tau\left(Y_l^{(3,j)}\right)\ \psi_\tau\left(Y_k^{(3,j)}\right)\middle|\mathcal{X}\right] \\
&= \mathbb{C}\text{ov}\left[\left(\tau - \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\right)\left(\tau - \mathbb{I}\left(Y_k^{(3,i)} < 0\right)\right)\,,\right. \\
&\qquad \left.\left(\tau - \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\right)\ \left(\tau - \mathbb{I}\left(Y_r^{(3,j)} < 0\right)\right)\middle|\mathcal{X}\right] \\
&= \tau^2\ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\,,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&- \tau\ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\mathbb{I}\left(Y_k^{(3,i)} < 0\right)\,,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&+ \tau^2\ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_k^{(3,i)} < 0\right)\,,\ \mathbb{I}\left(Y_k^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&- \tau\ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\mathbb{I}\left(Y_k^{(3,i)} < 0\right)\,,\ \mathbb{I}\left(Y_k^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&- \tau\ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,\tau)} < 0\right)\,,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_k^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&- \tau\ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_k^{(3,i)} < 0\right)\,,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_k^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&+ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\mathbb{I}\left(Y_k^{(3,i)} < 0\right)\,,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\mathbb{I}\left(Y_k^{(3,j)} < 0\right)\middle|\mathcal{X}\right] \\
&= 2\tau^3(1-\tau) - 4\tau^3(1-\tau) + \tau^2(1-\tau^2) \\
&= \tau^2(1-\tau^2) - 2\tau^3(1-\tau) = \tau^2(\tau^2 - 2\tau + 1).
\end{aligned}
$$

Analogously, the case in which $l = r \neq k = m$ can be concluded. So, in order to finish the calculus of $C_{9,1}(x_1, x_2)$ it will be necessary to analyse the scenario $l = k = m = r$. In this case, it is verified that

$$
\begin{aligned}
C_{9,1}(X_i, X_j) &= \mathbb{C}\text{ov}\left[\psi_\tau\left(Y_l^{(3,i)}\right)\psi_\tau\left(Y_l^{(3,i)}\right)\,,\ \psi_\tau\left(Y_l^{(3,j)}\right)\ \psi_\tau\left(Y_l^{(3,j)}\right)\middle|\mathcal{X}\right] \\
&= \mathbb{C}\text{ov}\left[\left(\tau - \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\right)\left(\tau - \mathbb{I}\left(Y_l^{(3,i)} < 0\right)\right)\,,\right. \\
&\qquad \left.\left(\tau - \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\right)\ \left(\tau - \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\right)\middle|\mathcal{X}\right] \\
&= \tau^2\ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right)\,,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right)\middle|\mathcal{X}\right]
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$
\begin{aligned}
&+ \tau^2\, \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \ ,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \Big| \mathcal{X}\right] \\
&- \tau\, \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \mathbb{I}\left(Y_l^{(3,i)} < 0\right) \ ,\ \mathbb{I}\left(Y_m^{(3,j)} < 0\right) \Big| \mathcal{X}\right] \\
&+ \tau^2\, \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \ ,\ \mathbb{I}\left(Y_l^{(3,j)}\right) \Big| \mathcal{X}\right] \\
&+ \tau^2\, \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \ ,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \Big| \mathcal{X}\right] \\
&- \tau\, \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \mathbb{I}\left(Y_l^{(3,i)} < 0\right) \ ,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \Big| \mathcal{X}\right] \\
&- \tau\, \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,\tau)} < 0\right) \ ,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \Big| \mathcal{X}\right] \\
&- \tau\, \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \ ,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \Big| \mathcal{X}\right] \\
&+ \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_l^{(3,i)} < 0\right) \mathbb{I}\left(Y_l^{(3,i)} < 0\right) \ ,\ \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \mathbb{I}\left(Y_l^{(3,j)} < 0\right) \Big| \mathcal{X}\right] \\
&\cong 4\tau^3(1-\tau) - 4\tau^2(1-\tau) + \tau(1-\tau) = \tau(1-\tau)(4\tau^2 - 4\tau + 1).
\end{aligned}
$$

As a consequence

$$
\mathbb{E}\left(\frac{16}{n^2 h_c^8} \sum_{i=1}^n \sum_{j=1}^n \mathbb{C}\text{ov}\left[\frac{s_\tau(X_i)^2}{g(X_i)^2}\, V_{\tau,h_c}(X_i)^2, \frac{s_\tau(X_j)^2}{g(X_j)^2}\, V_{\tau,h_c}(X_j)^2 \Big| \mathcal{X} \right] \right) = C_{9,2} + 2C_{9,3} + 4C_{9,4}
$$

where

$$
C_{9,2} = \frac{16}{n^2 h_c^8}\, \tau(1-\tau)(4\tau^2 - 4\tau + 1) \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \mathbb{E}\left[\frac{s_\tau(X_i)^2}{g(X_i)^2}\, \frac{s_\tau(X_j)^2}{g(X_j)^2}\, \widetilde{K}_{l,i}\, \widetilde{K}_{l,i}\, \widetilde{K}_{l,j}\, \widetilde{K}_{l,j} \right]
$$

$$
C_{9,3} = \frac{16}{n^2 h_c^8}\, \tau^2(\tau^2 - 2\tau + 1) \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \sum_{\substack{k=1 \\ k \neq l}}^n \mathbb{E}\left[\frac{s_\tau(X_i)^2}{g(X_i)^2}\, \frac{s_\tau(X_j)^2}{g(X_j)^2}\, \widetilde{K}_{l,i}\, \widetilde{K}_{l,i}\, \widetilde{K}_{k,j}\, \widetilde{K}_{k,j} \right]
$$

$$
\begin{aligned}
C_{9,4} = \frac{1}{36 n^2 h_c^8}\, \tau(1-\tau) \sum_{i=1}^n \sum_{j=1}^n \sum_{\text{card}(\{l,k,r\})=3} &\mathbb{E}\left[\frac{s_\tau(X_i)}{g(X_i)^2}\, \frac{s_\tau(X_j)}{g(X_j)^2}\, q_\tau^{(4)}(X_i)\, (X_l - X_i)^4 \right. \\
&\times \left. q_\tau^{(4)}(X_j)\, (X_r - X_i)^4\, \widetilde{K}_{l,i}\, \widetilde{K}_{k,i}\, \widetilde{K}_{k,j}\, \widetilde{K}_{r,j} \right].
\end{aligned}
$$

Then, we are going to analyse each of the previous quantities. Firstly, it follows that

$$
C_{9,2} = \frac{16}{n^6 h_c^{12}}\, \tau(1-\tau)(4\tau^2 - 4\tau + 1) \sum_{i=1}^n \sum_{j=1}^n \sum_{l=1}^n \mathbb{E}\left[\frac{s_\tau(X_i)^2}{g(X_i)^2}\, \frac{s_\tau(X_j)^2}{g(X_j)^2}\, \widetilde{K}_{l,i}\, \widetilde{K}_{l,i}\, \widetilde{K}_{l,j}\, \widetilde{K}_{l,j} \right]
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$= \frac{16}{n^3 h_c^{12}} \tau(1-\tau)(4\tau^2 - 4\tau + 1) \int\int\int \frac{s_\tau(X_i)^2}{g(X_i)^2} \frac{s_\tau(X_j)^2}{g(X_j)^2} \widetilde{K}_{l,i} \widetilde{K}_{l,i} \widetilde{K}_{l,j} \widetilde{K}_{l,j}$$
$$\times g(X_i, X_j, X_l) \, dX_i \, dX_j \, dX_l$$
$$= \frac{16}{n^3 h_c^{10}} \tau(1-\tau)(4\tau^2 - 4\tau + 1) \int \left[\int \frac{s_\tau(X_l - uh_c)^2}{g(X_l - uh_c)} \ [\alpha_{31} + \alpha_{33} u^2]^2 K(u)^2 \right]^2$$
$$\times g(X_l) \, dv \, dX_l$$
$$\cong \frac{16}{n^3 h_c^{10}} \tau(1-\tau)(4\tau^2 - 4\tau + 1) \, \Gamma_1(K)^2 \int \frac{s_\tau(X_l)^4}{g(X_l)} \, dX_l.$$

Secondly, it is verified that

$$C_{9,3} = \frac{16}{n^6 h_c^{12}} \tau^2(\tau^2 - 2\tau + 1) \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \sum_{\substack{k=1 \\ k \neq l}}^{n} \mathbb{E}\left[\frac{s_\tau(X_i)^2}{g(X_i)^2} \frac{s_\tau(X_j)^2}{g(X_j)^2} \widetilde{K}_{l,i} \widetilde{K}_{k,i} \widetilde{K}_{l,j} \widetilde{K}_{k,j} \right]$$
$$= \frac{16}{n^2 h_c^{12}} \tau^2(\tau^2 - 2\tau + 1) \int\int\int\int \frac{s_\tau(X_i)^2}{g(X_i)^2} \frac{s_\tau(X_j)^2}{g(X_j)^2} \widetilde{K}_{l,i} \widetilde{K}_{k,i} \widetilde{K}_{l,j} \widetilde{K}_{k,j}$$
$$\times g(X_i, X_j, X_l, X_k) \, dX_i \, dX_j \, dX_l \, dX_k$$
$$= \frac{16}{n^2 h_c^{10}} \tau^2(\tau^2 - 2\tau + 1) \int\int\int\int \frac{s_\tau(X_l - uh_c)^2}{g(X_l - uh_c)} \frac{s_\tau(X_l - vh_c)^2}{g(X_l - vh_c)}$$
$$\times [\alpha_{31} + \alpha_{33} u^2] K(u) \, [\alpha_{31} + \alpha_{33} v^2] K(v) \left[\alpha_{31} + \alpha_{33} \left(u + \frac{X_k - X_l}{h_c} \right)^2 \right]$$
$$\times K \left(u + \frac{X_k - X_l}{h_c} \right) \left[\alpha_{31} + \alpha_{33} \left(v + \frac{X_k - X_l}{h_c} \right)^2 \right] K \left(v + \frac{X_k - X_l}{h_c} \right)$$
$$\times g(X_l) \, g(X_k) \, du \, dv \, dX_l \, dX_k$$
$$= \frac{16}{n^2 h_c^{9}} \tau^2(\tau^2 - 2\tau + 1) \int\int\int\int \frac{s_\tau(X_l - uh_c)^2}{g(X_l - uh_c)} \frac{s_\tau(X_l - vh_c)^2}{g(X_l - vh_c)} [\alpha_{31} + \alpha_{33} u^2] K(u)$$
$$\times [\alpha_{31} + \alpha_{33} v^2] K(v) [\alpha_{31} + \alpha_{33} \, (u + w)^2] K \, (u + w) \ [\alpha_{31} + \alpha_{33} \, (v + w)^2]$$
$$\times K \, (v + w) \ g(X_l) \, g(X_l + wh_c) \, du \, dv \, dw \, dX_l$$
$$= \frac{16}{n^2 h_c^{9}} \tau^2(\tau^2 - 2\tau + 1) \int \left[\int \frac{s_\tau(X_l - uh_c)^2}{g(X_l - uh_c)} [\alpha_{31} + \alpha_{33} u^2] K(u) \right.$$
$$\left. \times [\alpha_{31} + \alpha_{33} (u + w)^2] K(u + w) \, dw \, du \right]^2 g(X_l)^2 \, dX_l$$
$$\cong \frac{16}{n^2 h_c^{9}} \tau^2(\tau^2 - 2\tau + 1) \Gamma_3(K)^2 \int s_\tau(X_l)^4 \, dX_l$$

where

$$\Gamma_3(K) = \int\int [\alpha_{31} + \alpha_{33} u^2] K(u) [\alpha_{31} + \alpha_{33} (u + w)^2] K(u + w) \, dw \, du.$$

Finally, based on the same kind of arguments, the following approximation is justified:

$$C_{9,4} = \frac{1}{36 n^6 h_c^{12}} \tau(1-\tau) \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{\mathrm{card}(\{l,k,r\}=3)} \mathbb{E}\left[\frac{s_\tau(X_i)}{g(X_i)^2} \frac{s_\tau(X_j)}{g(X_j)^2} q_\tau^{(4)}(X_i) \, (X_l - X_i)^4 \right.$$

$$
\begin{aligned}
&\times q_\tau^{(4)}(X_j)\,(X_r-X_j)^4\,\widetilde{K}_{l,i}\,\widetilde{K}_{k,i}\,\widetilde{K}_{k,j}\,\widetilde{K}_{r,j}\Big]\\
&=\frac{1}{36n^2h_c^{12}}\,\tau(1-\tau)\int\!\!\int\!\!\int\!\!\int\!\!\int\frac{s_\tau(X_i)}{g(X_i)^2}\,\frac{s_\tau(X_j)}{g(X_j)^2}\,q_\tau^{(4)}(X_i)\,(X_l-X_i)^4\\
&\quad\times q_\tau^{(4)}(X_j)\,(X_r-X_j)^4\,\widetilde{K}_{l,i}\,\widetilde{K}_{k,i}\,\widetilde{K}_{k,j}\,\widetilde{K}_{r,j}\\
&\quad\times g(X_i,X_j,X_l,X_k,X_r)\,dX_i\,dX_j\,dX_l\,dX_k\,dX_r\\
&=\frac{1}{36n^2h_c^{2}}\,\tau(1-\tau)\int\!\!\int\!\!\int\!\!\int\!\!\int\frac{s_\tau(X_l-uh_c)}{g(X_l-uh_c)}\,\frac{s_\tau(X_r-vh_c)}{g(X_r-vh_c)}\,q_\tau^{(4)}(X_l-uh_c)\,u^4\\
&\quad\times q_\tau^{(4)}(X_r-vh_c)\,v^4\,[\alpha_{31}+\alpha_{33}u^2]K(u)\,[\alpha_{31}+\alpha_{33}v^2]K(v)\\
&\quad\times\left[\alpha_{31}+\alpha_{33}\left(\frac{X_k-X_l}{h}\right)^2\right]K\left(\frac{X_k-X_l}{h}\right)\,\left[\alpha_{31}+\alpha_{33}\left(\frac{X_k-X_l}{h}\right)^2\right]\\
&\quad\times K\left(\frac{X_k-X_l}{h}\right)\,g(X_l)\,g(X_k)\,g(X_r)\,du\,dv\,dX_l\,dX_k\,dX_r\\
&=\frac{1}{36n^2h_c}\,\tau(1-\tau)\int\!\!\int\!\!\int\!\!\int\!\!\int\frac{s_\tau(X_l-uh_c)}{g(X_l-uh_c)}\,\frac{s_\tau(X_r-vh_c)}{g(X_r-vh_c)}\,q_\tau^{(4)}(X_l-uh_c)\,u^4\\
&\quad\times q_\tau^{(4)}(X_r-vh_c)\,v^4\,[\alpha_{31}+\alpha_{33}u^2]K(u)\,[\alpha_{31}+\alpha_{33}v^2]K(v)\\
&\quad\times[\alpha_{31}+\alpha_{33}\,(w)^2]K\,(w)\,\left[\alpha_{31}+\alpha_{33}\left(\frac{X_l-X_r}{h_c}-w\right)^2\right]K\left(\frac{X_l-X-r}{h_c}-w\right)\\
&\quad\times g(X_l)\,g(Xl-wh_c)\,g(X_r)\,du\,dv\,dX_l\,dw\,dX_r\\
&=\frac{1}{36n^2}\,\tau(1-\tau)\int\!\!\int\!\!\int\!\!\int\!\!\int\frac{s_\tau(X_l-uh_c)}{g(X_l-uh_c)}\,\frac{s_\tau(X_l-zh_c-vh_c)}{g(X_l-zh_c-vh_c)}\,q_\tau^{(4)}(X_l-uh_c)\,u^4\\
&\quad\times q_\tau^{(4)}(X_l-zh_c-vh_c)\,v^4\,[\alpha_{31}+\alpha_{33}u^2]K(u)\,[\alpha_{31}+\alpha_{33}v^2]K(v)\\
&\quad\times[\alpha_{31}+\alpha_{33}\,(w)^2]K\,(w)\,\,[\alpha_{31}+\alpha_{33}\,(z-w)^2]K\,(z-w)\\
&\quad\times g(X_l)\,g(Xl-wh_c)\,g(X_l-zh_c)\,du\,dv\,dX_l\,dw\,dz\\
&\cong\frac{1}{36n^2}\,\tau(1-\tau)\,\Gamma_4(K)\int s_\tau(X_l)^2\,q_\tau^{(4)}(X_l)^2\,g(X_l)\,dX_l
\end{aligned}
$$

where

$$
\begin{aligned}
\Gamma_4(K)=&\int\!\!\int[\alpha_{31}+\alpha_{33}v^2]K(v)[\alpha_{31}+\alpha_{33}(w-v)^2]K(w-v)\,dw\,dv\\
&\times\left[\int u^4\,[\alpha_{31}+\alpha_{33}u^2]K(u)\,du\right]^2.
\end{aligned}
$$

In order to conclude the result of this proposition, it will be necessary to compute the covariance between the conditional expectations. So, it should be taken into account that

$$\mathbb{E}[V_{\tau,h_c}(x_1)^2|\mathcal{X}]\cong\left(\frac{1}{nh_c}\right)^2\sum_{i=1}^n\tau(1-\tau)\,\widetilde{K}_{i,1}^2$$

because of (3.34), and as a consequence

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[\mathbb{E}\left(V_{\tau,h_c}(x_1)^2|\mathcal{X}\right),\mathbb{E}\left(V_{\tau,h_c}(x_2)^2|\mathcal{X}\right)\right]=\\
&\qquad\mathbb{C}\mathrm{ov}\left[\left(\frac{1}{nh_c}\right)^2\sum_{i=1}^n\tau(1-\tau)\,\widetilde{K}_{i,1}^2,\left(\frac{1}{nh_s}\right)^2\sum_{j=1}^n\tau(1-\tau)\,\widetilde{K}_{j,2}^2\right]
\end{aligned}
$$

$$\begin{aligned}
&\leq \left(\frac{1}{nh_c}\right)^4 \tau^2(1-\tau)^2 \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{C}\text{ov}\,[\widetilde{K}_{i,1}^2, \widetilde{K}_{j,2}^2] \\
&= \left(\frac{1}{nh_c}\right)^4 \tau^2(1-\tau)^2\; n\, O(h_c) \\
&= O\left(\frac{1}{n^3h_c^3}\right) = o\left(\frac{1}{n^2}\right)
\end{aligned}$$

where the last step comes from the following development that comes calculus associated with the kernel function used in Proposition 3.6 and 3.7. That is,

$$\begin{aligned}
\mathbb{C}\text{ov}\,[\widetilde{K}_{i,1}^2\ ,\ \widetilde{K}_{i,2}^2] &= \mathbb{E}[\widetilde{K}_{i,1}^2\ \widetilde{K}_{i,2}^2] - \mathbb{E}[\widetilde{K}_{i,1}^2\ \mathbb{E}[\widetilde{K}_{i,2}^2] \\
&= O(h_c) - O(h_c)O(h_c) = O(h_c).
\end{aligned}$$

To sum up, it have been proved that

$$C_9 \cong \frac{32}{n^2h_c^9}\,\tau^2(\tau^2-2\tau+1)\,\Gamma_3(K)^2 \int s_\tau(X_l)^4\, dX_l$$

where

$$\Gamma_3(K) = \int\int [\alpha_{31}+\alpha_{33}u^2]K(u)[\alpha_{31}+\alpha_{33}(u+w)^2]K(u+w)\, dw\, du.$$

□

Bearing the last three proposition in mind, the variance of the curvature estimator has been computed. This results is presented in the following theorem

Theorem 3.9. *Under conditions C1-C4 the variance of the curvature estimator, given by (3.10), can be approximated as follows*

$$\begin{aligned}
\mathbb{V}ar\left[\widehat{\vartheta}_{h_c}\right] \cong \frac{4\tau(1-\tau)}{n}\Bigg[&\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)^2\, g(x)\, dx \\
&+\Gamma_2(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)\ s_\tau(x)\ q_\tau^{(4)}(x)\ g(x)\, dx\Bigg] \\
&+\frac{32}{n^2h_c^9}\,\tau^2(\tau^2-2\tau+1)\,\Gamma_3(K)^2\int s_\tau(x)^4\, dx.
\end{aligned}$$

Proof. In view of Proposition 3.6, 3.7, and 3.8 and equation (3.26) it follows that:

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\widehat{\vartheta}_{h_c}\right] \cong\ &\frac{4\tau(1-\tau)}{n}\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)^2\, g(x)\, dx \\
&\frac{16(2\tau^2-\tau)(1-\tau)}{n^2h_c^5}\Gamma_1(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)\ s_\tau(x)^2\, dx \\
&+\frac{4\tau(1-\tau)}{n}\ \Gamma_2(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)\ s_\tau(x)\ q_\tau^{(4)}(x)\ g(x)\, dx \\
&+\frac{32}{n^2h_c^9}\,\tau^2(\tau^2-2\tau+1)\Gamma_3(K)^2\int s_\tau(x)^4\, dx
\end{aligned}$$

$$\cong \frac{4\tau(1-\tau)}{n}\left[\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)^2\ g(x)\ dx\right.$$
$$\left.+\ \Gamma_2(K)\int (q_\tau^{(2)}\ s_\tau)^{(2)}(x)\ s_\tau(x)\ q_\tau^{(4)}(x)\ g(x)\ dx\right]$$
$$+\ \frac{32}{n^2 h_c^9}\ \tau^2(\tau^2-2\tau+1)\Gamma_3(K)^2\int s_\tau(x)^4\ dx.$$

□

Summarizing, if we remember the bias and the variance of the curvature estimator given in Theorems 3.5 and 3.9, it can be computed the mean squared error of the curvature estimator given by (3.10). This result is gathered in the following theorem:

Theorem 3.10. *Under assumptions C1-C4, the mean squared error of the curvature estimator, given by (3.10), can be approximated by*

$$MSE\left[\widehat{\vartheta}_{h_c}\right] \cong \left[\frac{h_c^2}{6}\ \delta_1\int q_\tau^{(2)}(x)\ q_\tau^{(4)}(x)\ g(x)\ dx + \frac{4\tau(1-\tau)}{nh_c^5}\ \delta_2\int \frac{1}{f(q_\tau(x)|X=x)^2}\ dx\right]^2$$

Proof. In view of Theorems 3.5 and 3.9, the result is derived. Note that all terms associated with the variance are negligible compared with the terms associated with the bias of the estimator. □

3.4 Derivation of the asymptotic mean squared error associated with the integrated squared sparsity estimator

Recall the definition of the proposed sparsity estimator

$$\widehat{s}_{\tau,d_s,h_s}(x) = \frac{\widehat{q}_{\tau+d_s,h_s}(x) - \widehat{q}_{\tau-d_s,h_s}(x)}{2\,d_s} \tag{3.35}$$

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 parameter. Applying Fan et al. (1994)'s results together with the argument developed in Remark 3.2, it follows that

$$\widehat{q}_{\tau+d_s,h_s}(x) = q_{\tau+d_s}(x) + \frac{s_{\tau+d_s}(x)}{g(x)}\ U_{\tau+d_s,h_s}(x) + o((nh_s)^{-1/2})$$

where

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

$\psi_\tau(r) = \tau - \mathbb{I}(r<0)$, $K_i = K\left(\frac{X_i-x}{h_s}\right)$ and $Y_i^{(1,\tau+d_s)} = Y_i - q_{\tau+d_s}(x) - q_{\tau+d_s}^{(1)}(x)(X_i-x)$. 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 can obtain the following representation of the sparsity estimator

$$\widehat{s}_{\tau,d_s,h_s}(x) = \frac{\widehat{q}_{\tau+d_s,h_s}(x) - \widehat{q}_{\tau-d_s,h_s}(x)}{2\,d_s} = A(x) + B(x) + o\left((nh_s)^{-1/2}d_s^{-1}\right)$$

where

$$A(x) = \frac{q_{\tau+d_s,h_s}(x) - q_{\tau-d_s,h_s}(x)}{2\, d_s}$$

$$B(x) = \frac{1}{2d_s}\left(\frac{s_{\tau+d_s}(x)}{g(x)}\, U_{\tau+d_s,h_s}(x) - \frac{s_{\tau-d_s}(x)}{g(x)}\, U_{\tau-d_s,h_s}(x)\right)$$

In this section, the asymptotic mean squared error of the integrated squared sparsity estimator will be established. For this, we make the following assumptions:

Conditions S

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 exists 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

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

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

where $\psi_\tau(r) = \tau\mathbb{I}(r>0) + (\tau-1)\mathbb{I}(r<0)$ is the derivative of the quantile loss function $\rho_\tau = \tau r\mathbb{I}(r>0) + (\tau-1)r\mathbb{I}(r<0)$.

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 $q_\tau^{(i)}$. 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, non negative, 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 $d_s \to 0$, $h_s \to 0$ and $nd_sh_s \to \infty$ when $n \to \infty$.

S5: The function q_{τ_1} has a continuous and bounded forth 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.

3.4.1 Auxiliary results

In order to study the properties of the sparsity estimator (3.35) it will be crucial to analyse the expectation, variance and covariance of $U_{\tau,h_s}(x)$ functions. These moments have been gathered in Lemma 3.11.

Lemma 3.11. *Given a random sample $\mathcal{X} = \{X_1, \ldots, X_n\}$ of the explanatory variable X, and under conditions S1-S4, it is verified that*

$$\mathbb{E}\left[U_{\tau_1,h_s}(x)|\mathcal{X}\right] = \frac{1}{nh_s} \sum_{i=1}^{n} f(q_{\tau_1}(X_i) + \xi_{2,i}|X = X_i)\, \frac{1}{2} q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2\, K_i$$

$$\mathbb{V}ar\left[U_{\tau_1,h_s}(x)|\mathcal{X}\right] = \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \tau_1(1-\tau_1)\; K_i^2 + \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2\right) K_i^2$$

$$\mathbb{C}ov\left[U_{\tau+d_s,h_s}(x), U_{\tau-d_s,h_s}(x)|\mathcal{X}\right] = \frac{1}{(nh_s)^2} \sum_{i=1}^{n} (\tau - d_s)(1 - \tau - d_s) K_i^2 + \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \left(\frac{\partial \varphi_6(t_1,t_2)}{\partial t_1}, \frac{\partial \varphi_6(t_1,t_2)}{\partial t_1}\right)' \Bigg|_{(\xi_{6,i},\xi_{7,i})} \times \left(-\frac{1}{2} q_{\tau+d_s}^{(2)}(\xi_{4,i})(X_i - x)^2, -\frac{1}{2} q_{\tau-ds}^{(2)}(\xi_{5,i})(X_i - x)^2\right) K_i^2$$

uniformly in τ_1 in a neighbourhood of τ.

Proof. Firstly, we should consider the following Taylor expansion

$$q_{\tau_1}(X_i) = q_{\tau_1}(x) + q_{\tau_1}^{(1)}(x)(X_i - x) + \frac{1}{2}\, q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2 \tag{3.36}$$

where $\xi_{1,i}$ is an element between X_i and x.

Conditional expectation of $\mathbf{U_{\tau_1,h_s}(x)}$

In view of (3.36), we can write

$$\begin{aligned}\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right) &= \mathbb{I}\left(Y_i - q_{\tau_1}(x) - q_{\tau_1}^{(1)}(x)(X_i - x) < 0\right) \\ &= \mathbb{I}\left(Y_i - q_{\tau_1}(X_i) < -q_{\tau_1}(X_i) + q_{\tau_1}(x) + q_{\tau_1}^{(1)}(x)(X_i - x)\right) \\ &= \mathbb{I}\left(Y_i - q_{\tau_1}(X_i) < -\frac{1}{2}\, q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2\right),\end{aligned}$$

and as a consequence

$$\begin{aligned}\mathbb{E}[U_{\tau_1,hs}(x)|\mathcal{X}] &= \mathbb{E}\left[\frac{1}{nh_s} \sum_{i=1}^{n} \psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right) K_i \middle| \mathcal{X}\right] \\ &= \frac{1}{nh_s} \sum_{i=1}^{n} \mathbb{E}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right) \middle| \mathcal{X}\right] K_i \\ &= \frac{1}{nh_s} \sum_{i=1}^{n} \left(\tau_1 - \mathbb{E}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right) \middle| \mathcal{X}\right]\right) K_i.\end{aligned} \tag{3.37}$$

Then, the following function should be defined

$$\varphi_5(t) = \mathbb{E}(\mathbb{I}(Y - q_{\tau_1}(x) < t)|X = x) = F(q_{\tau_1}(x) + t|X = x),$$

and taking into account the definition of the function φ_5 and equations (3.37) and (3.36), we can conclude that

$$\begin{aligned}\mathbb{E}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)}<0\right)\middle|\mathcal{X}\right] &= \varphi_5\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\right)\\ &= \varphi_5(0)+\varphi_5^{(1)}(\xi_{2,i})\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\right)\\ &= F(q_{\tau_1}(X_i)|X=X_i)+f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i) \qquad (3.38)\\ &\quad\times\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\right)\\ &= \tau_1+f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\right) \qquad (3.39)\end{aligned}$$

where we have developed a Taylor expansion of φ_5 and $\xi_{2,i}$ represents an element between $-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2$ and 0. Bearing in mind (3.37) and (3.39), it follows that

$$\begin{aligned}\mathbb{E}\left[U_{\tau_1,h_s}(x)|\mathcal{X}\right] &= \frac{1}{nh_s}\sum_{i=1}^{n}\left(\tau_1-\mathbb{E}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)}<0\right)\middle|\mathcal{X}\right]\right)K_i\\ &= \frac{1}{nh_s}\sum_{i=1}^{n}\left(\tau_1-\tau_1-f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\right)\right)K_i\\ &= \frac{1}{nh_s}\sum_{i=1}^{n}f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\,\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\,K_i.\end{aligned}$$

Conditional variance of $\mathbf{U_{\tau_1,h_s}(x)}$

In this case, it is verified that

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[U_{\tau_1,h_s}(x)|\mathcal{X}\right] &= \mathbb{V}\mathrm{ar}\left[\frac{1}{nh_s}\sum_{i=1}^{n}\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)K_i\middle|\mathcal{X}\right]\\ &= \frac{1}{(nh_s)^2}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\middle|\mathcal{X}\right]K_i^2\\ &= \frac{1}{(nh_s)^2}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}\left[\tau_1-\mathbb{I}\left(Y_i^{(1,\tau_1)}<0\right)\middle|\mathcal{X}\right]K_i^2\\ &= \frac{1}{(nh_s)^2}\sum_{i=1}^{n}\mathbb{V}\mathrm{ar}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)}<0\right)\middle|\mathcal{X}\right]K_i^2. \qquad (3.40)\end{aligned}$$

So, taking into account the definition of the function φ_2

$$\begin{aligned}\varphi_2(t) &= \mathbb{V}\mathrm{ar}\left(\mathbb{I}(Y-q_\tau(x)<t)|X=x\right)\\ &= F(q_\tau(x)+t|X=x)(1-F(q_\tau(x)+t|X=x))\end{aligned}$$

and equation (3.36), it is concluded that

$$\mathbb{V}\mathrm{ar}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)}<0\right)\middle|\mathcal{X}\right] = \varphi_2\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\right)$$

$$
\begin{aligned}
&= \varphi_2(0) + \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2 \right) \\
&= F(q_{\tau_1}(x)|X = x)(1 - F(q_{\tau_1}(x)|X = x)) \\
&+ \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2 \right) \\
&= \tau_1(1 - \tau_1) + \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2 \right) \qquad (3.41)
\end{aligned}
$$

where we have developed a Taylor expansion of function φ_2 and $\xi_{3,i}$ represents an element between $-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2$ and 0. So, in view of (3.40) and (3.41), it follows that

$$
\begin{aligned}
\mathbb{V}\text{ar}\left[U_{\tau_1,h_s}(x)|\mathcal{X}\right] &= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \mathbb{V}\text{ar}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right) \middle| \mathcal{X} \right] K_i^2 \\
&= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \left(\tau_1(1 - \tau_1) + \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x)^2 \right) \right) K_i^2.
\end{aligned}
$$

Conditional covariance between $\mathbf{U}_{\tau+\mathbf{d_s},\mathbf{h_s}}(\mathbf{x})$ and $\mathbf{U}_{\tau-\mathbf{d_s},\mathbf{h_s}}(\mathbf{x})$

Let us remember that

$$
\begin{aligned}
&\mathbb{C}\text{ov}\left(U_{\tau+d_s,h_s}(x), U_{\tau-d_s,h_s}(x) \middle| \mathcal{X} \right) = \\
&= \mathbb{C}\text{ov}\left[\frac{1}{nh_s} \sum_{i=1}^{n} \psi_{\tau+d_s}\left(Y_i^{(1,\tau+d_s)}\right) K_i, \frac{1}{nh_s} \sum_{j=1}^{n} \psi_{\tau-d_s}\left(Y_j^{(1,\tau-d_s)}\right) K_j \middle| \mathcal{X} \right] \\
&= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \mathbb{C}\text{ov}\left[\psi_{\tau+d_s}\left(Y_i^{(1,\tau+d_s)}\right), \psi_{\tau-d_s}\left(Y_i^{(1,\tau-d_s)}\right) \middle| \mathcal{X} \right] K_i^2 \\
&= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \mathbb{C}\text{ov}\left[\left(\tau + d_s - \mathbb{I}\left(Y_i^{(1,\tau+d_s)} < 0\right)\right), \left(\tau - d_s - \mathbb{I}\left(Y_i^{(1,\tau-d_s)} < 0\right)\right) \middle| \mathcal{X} \right] K_i^2 \\
&= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau+d_s)} < 0\right), \mathbb{I}\left(Y_i^{(1,\tau-d_s)} < 0\right) \middle| \mathcal{X} \right] K_i^2 \qquad (3.42)
\end{aligned}
$$

where we have take into account that if $i \neq j$ then $\psi_{\tau+d_s}\left(Y_i^{(1,\tau+d_s)}\right)$ and $\psi_{\tau-d_s}\left(Y_j^{(1,\tau-d_s)}\right)$ are independent as well as their covariance is zero. Furthermore, let us define the function

$$
\begin{aligned}
\varphi_6(t_1, t_2) &= \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_i < q_{\tau+d_s}(X_i) + t_1\right), \mathbb{I}\left(Y_i < q_{\tau-d_s}(X_i) + t_2\right)|\mathcal{X}\right] \\
&= \mathbb{E}\left[\mathbb{I}\left(Y_i < q_{\tau+d_s}(X_i) + t_1\right)\mathbb{I}\left(Y_i < q_{\tau-d_s}(X_i) + t_2\right)|\mathcal{X}\right] \\
&- \mathbb{E}\left[\mathbb{I}\left(Y_i < q_{\tau+d_s}(X_i) + t_1\right)|\mathcal{X}\right]\mathbb{E}\left[\mathbb{I}\left(Y_i < q_{\tau-d_s}(X_i) + t_2\right)|\mathcal{X}\right] \\
&= F\left(\min\{q_{\tau+d_s}(x) + t_1, q_{\tau-d_s}(x) + t_2\}|X = x\right) \\
&- F\left(q_{\tau+d_s}(x) + t_1|X = x\right) F\left(q_{\tau-d_s}(x) + t_2|X = x\right)
\end{aligned}
$$

whose partial derivatives are given by

$$\begin{aligned}\frac{\partial\varphi_6(t_1,t_2)}{\partial t_1} &= f\left(\min\{q_{\tau+d_s}(x)+t_1, q_{\tau-d_s}(x)+t_2\}|X=x\right)\\ &\times \mathbb{I}(q_{\tau+d_s}(x)+t_1 \leq q_{\tau-d_s}(x)+t_2)\\ &- f\left(q_{\tau+d_s}(x)+t_1|X=x\right)F\left(q_{\tau-d_s}(x)+t_2|X=x\right)\\ \frac{\partial\varphi_6(t_1,t_2)}{\partial t_2} &= f\left(\min\{q_{\tau+d_s}(x)+t_1, q_{\tau-d_s}(x)+t_2\}|X=x\right)\\ &\times \mathbb{I}(q_{\tau-d_s}(x)+t_2 \leq q_{\tau+d_s}(x)+t_1)\\ &- f\left(q_{\tau+d_s}(x)+t_1|X=x\right)F\left(q_{\tau-d_s}(x)+t_2|X=x\right).\end{aligned}$$

Then, in view of arguments developed in equation (3.36) and a Taylor expansion of function φ_6 we can determine that

$$\begin{aligned}&\mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau+d_s)}<0\right),\mathbb{I}\left(Y_i^{(1,\tau-d_s)}<0\right)\middle|\mathcal{X}\right]=\\ &=\varphi_6\left(-\frac{1}{2}q^{(2)}_{\tau+d_s}(\xi_{4,i})(X_i-x)^2,-\frac{1}{2}q^{(2)}_{\tau-d_s}(\xi_{5,i})(X_i-x)^2\right)\\ &=\varphi_6(0,0)+\left(\frac{\partial\varphi_6(t_1,t_2)}{\partial t_1},\frac{\partial\varphi_6(t_1,t_2)}{\partial t_2}\right)'\Bigg|_{(t1,t2)=(\xi_{6,i},\xi_{7,i})}\\ &\times\left(-\frac{1}{2}q^{(2)}_{\tau+d_s}(\xi_{4,i})(X_i-x)^2,-\frac{1}{2}q^{(2)}_{\tau-d_s}(\xi_{5,i})(X_i-x)^2\right)\end{aligned}\tag{3.43}$$

where $\xi_{4,i}$ and $\xi_{5,i}$ are elements between Xi and x obtained thanks to a Taylor expansion of $q_{\tau+d_s}$ and $q_{\tau-d_s}$, respectively. Moreover, $\xi_{6,i}$ is an element between $-\frac{1}{2}q^{(2)}_{\tau+d_s}(\xi_{4,i})(X_i-x)^2$ and 0, and $\xi_{7,i}$ is an element between $-\frac{1}{2}q^{(2)}_{\tau-d_s}(\xi_{5,i})(X_i-x)^2$ and 0, obtained thanks to a Taylor expansion of φ_6. In addition,

$$\begin{aligned}\varphi_6(0,0) &= F\left(\min\{q_{\tau+d_s}(x),q_{\tau-d_s}(x)\}|X=x\right)\\ &- F\left(q_{\tau+d_s}(x)|X=x\right)F\left(q_{\tau-d_s}(x)|X=x\right)\\ &= F\left(q_{\tau-d_s}(x)|X=x\right)-F\left(q_{\tau+d_s}(x)|X=x\right)F\left(q_{\tau-d_s}(x)|X=x\right)\\ &= \tau-d_s-(\tau+d_s)(\tau-d_s)=(\tau-d_s)(1-\tau-d_s).\end{aligned}$$

So, we can conclude that

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left[U_{\tau+d_s,h_s}(x),U_{\tau-d_s,h_s}(x)|\mathcal{X}\right] &= \frac{1}{(nh_s)^2}\sum_{i=1}^n(\tau-d_s)(1-\tau-d_s)K_i^2\\ &+\frac{1}{(nh_s)^2}\sum_{i=1}^n\left(\frac{\partial\varphi_6(t_1,t_2)}{\partial t_1},\frac{\partial\varphi_6(t_1,t_2)}{\partial t_2}\right)'\Bigg|_{(\xi_{6,i},\xi_{7,i})}\\ &\times\left(-\frac{1}{2}q^{(2)}_{\tau+d_s}(\xi_{4,i})(X_i-x)^2,-\frac{1}{2}q^{(2)}_{\tau-ds}(\xi_{5,i})(X_i-x)^2\right)K_i^2.\end{aligned}$$

□

3.4.2 Expectation and variance of the sparsity estimator

In this subsection, the first moments of the sparsity estimator will be studied. So we should remember the following representation of the sparsity estimator

$$\widehat{s}_{\tau,d_s,h_s}(x) = \frac{\widehat{q}_{\tau+d_s,h_s}(x) - \widehat{q}_{\tau-d_s,h_s}(x)}{2\,d_s} = A(x) + B(x) + o\left((nh_s)^{-1/2} d_s^{-1}\right)$$

where

$$A(x) = \frac{q_{\tau+d_s}(x) - q_{\tau-d_s}(x)}{2\,d_s}$$

$$B(x) = \frac{1}{2d_s}\left(\frac{s_{\tau+d_s}(x)}{g(x)}\, U_{\tau+d_s,h_s}(x) - \frac{s_{\tau-d_s}(x)}{g(x)}\, U_{\tau-d_s,h_s}(x)\right).$$

First of all, it should be noticed that $A(x)$ is not random and it can be approximated by a Taylor expansion. This result is given in the following lemma:

Lemma 3.12. *Suppose condition S5 follows, then*

$$A(x) = s_\tau(x) + \frac{1}{6}\, s_\tau^{(2,\tau)}(x)\, d_s^2 + O(d_s^3)$$

where $s_\tau^{(i,\tau)}(x) = \frac{\partial^i}{\partial \tau^i} s_\tau(x)$.

Proof. The following Taylor expansions can be considered:

$$q_{\tau+d_s}(x) = q_\tau(x) + \left.\frac{\partial q_\tau(x)}{\partial \tau}\right|_\tau d_s + \frac{1}{2}\left.\frac{\partial^2 q_\tau(x)}{\partial \tau^2}\right|_\tau d_s^2 + \frac{1}{6}\left.\frac{\partial^3 q_\tau(x)}{\partial \tau^3}\right|_\tau d_s^3 + \frac{1}{24}\left.\frac{\partial^4 q_\tau(x)}{\partial \tau^4}\right|_{\xi_8} d_s^4$$

$$q_{\tau-d_s}(x) = q_\tau(x) + \left.\frac{\partial q_\tau(x)}{\partial \tau}\right|_\tau (-d_s) + \frac{1}{2}\left.\frac{\partial^2 q_\tau(x)}{\partial \tau^2}\right|_\tau (-d_s)^2 + \frac{1}{6}\left.\frac{\partial^3 q_\tau(x)}{\partial \tau^3}\right|_\tau (-d_s)^3 + \frac{1}{24}\left.\frac{\partial^4 q_\tau(x)}{\partial \tau^4}\right|_{\xi_9} (-d_s)^4$$

where ξ_8 is an element between τ and $\tau + d_s$, and ξ_9 is an element between $\tau - d_s$ and τ. Then, we conclude that

$$\begin{aligned} A(x) &= \frac{q_{\tau+d_s}(x) - q_{\tau-d_s}(x)}{2\,d_s} \\ &= \frac{1}{2d_s}\left(2\left.\frac{\partial q_\tau(x)}{\partial \tau}\right|_\tau d_s + \frac{1}{3}\left.\frac{\partial^3 q_\tau(x)}{\partial \tau^3}\right|_\tau d_s^3 + \frac{1}{24}\left(\left.\frac{\partial^4 q_\tau(x)}{\partial \tau^4}\right|_{\xi_8} - \left.\frac{\partial^4 q_\tau(x)}{\partial \tau^4}\right|_{\xi_9}\right) d_s^4\right) \\ &= \frac{2\, s_\tau(x)\, d_s + \frac{1}{3}\, s_\tau^{(2,\tau)}\, d_s^3 + O(d_s^4)}{2\,d_s} = s_\tau(x) + \frac{1}{6}\, s_\tau^{(2,\tau)}(x)\, d_s^2 + O(d_s^3). \end{aligned}$$

□

Thus, the problem of computing bias and variance of the sparsity estimator have been reduced to the calculus of expectation and variance of $B(x)$. Firstly, Proposition 3.13 shows the expectation of $B(x)$.

Proposition 3.13. *Under assumptions S1-S5, it is verified that*

$$\mathbb{E}[B(x)] = \frac{1}{2}\, \mu_2(K)\, \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau\, h_s^2 + o(h_s^2).$$

Proof. The expectation of $B(x)$ is consequence of arguments developed in the proof of Lemma 2 in Fan et al. (1994) that computed the expectation of $U_{\tau,h_s}(x)$ that is given by:

$$\mathbb{E}[U_{\tau,h_s}(x)] = \frac{1}{2}\, q_\tau^{(2)}(x)\, g(x)\, f(q_\tau(x)|X = x)\, \mu_2(K)\, h_s^2(1 + o(1)).$$

So,

$$\begin{aligned}
\mathbb{E}[B(x)] &= \frac{1}{2d_s}\, \mathbb{E}\left[\frac{s_{\tau+d_s}(x)}{g(x)}\, U_{\tau+d_s,h_s}(x) - \frac{s_{\tau-d_s}(x)}{g(x)}\, U_{\tau-d_s,h_s}(x)\right] \\
&= \frac{1}{2d_s\, g(x)}\left(s_{\tau+d_s}(x)\, \mathbb{E}[U_{\tau+d_s,h_s}(x)] - s_{\tau-d_s}(x)\, \mathbb{E}[U_{\tau-d_s,h_s}(x)]\right) \\
&= \frac{1}{2d_s\, g(x)}\left(s_{\tau+d_s}(x)\frac{1}{2}\, q_{\tau+d_s}^{(2)}(x)\, g(x)\, f(q_{\tau+d_s}(x)|X = x)\, \mu_2(K)\, h_s^2\right. \\
&\quad \left. - s_{\tau-d_s}(x)\, \frac{1}{2}\, q_{\tau-d_s}^{(2)}(x)\, g(x)\, f(q_{\tau-d_s}(x)|X = x)\, \mu_2(K)\, h_s^2\right)(1 + o(1)) \\
&= \frac{1}{2}\, \mu_2(K)\, h_s^2\, \left(\frac{q_{\tau+d_s}^{(2)}(x) - q_{\tau-d_s}^{(2)}(x)}{2d_s}\right)(1 + o(1)) \\
&\overset{(i)}{=} \frac{1}{2}\, \mu_2(K)\, h_s^2\, \left(\frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_{\tau=\tau} + O(d_s)\right)(1 + o(1)) \\
&= \frac{1}{2}\, \mu_2(K)\, \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_{\tau=\tau}\, h_s^2 + o(h_s^2)
\end{aligned}$$

where (i) comes from the following two Taylor expansions of the regression function

$$q_{\tau+d_s}^{(2)}(x) = q_\tau^{(2)}(x) + \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_{\tau=\tau}\, d_s + \frac{1}{2}\, \frac{\partial^2 q_\tau^{(2)}(x)}{\partial \tau^2}\bigg|_{\tau=\xi_{10}}\, d_s^2$$

$$q_{\tau-d_s}^{(2)}(x) = q_\tau^{(2)}(x) + \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_{\tau=\tau}\, (-d_s) + \frac{1}{2}\, \frac{\partial^2 q_\tau^{(2)}(x)}{\partial \tau^2}\bigg|_{\tau=\xi_{11}}\, (-d_s)^2$$

where ξ_{10} is an element between τ and $\tau + d_s$, and ξ_{11} is an element between $\tau - d_s$ and τ. Hence, it follows that

$$\begin{aligned}
\frac{q_{\tau+d_s}^{(2)}(x) - q_{\tau+d_s}^{(2)}(x)}{2d_s} &= \frac{1}{2d_s}\left(\frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_{\tau=\tau}\, 2d_s + \frac{1}{2}\, \left(\frac{\partial^2 q_\tau^{(2)}(x)}{\partial \tau^2}\bigg|_{\tau=\xi_{10}} - \frac{\partial^2 q_\tau^{(2)}(x)}{\partial \tau^2}\bigg|_{\tau=\xi_{11}}\right)\, d_s^2\right) \\
&= \frac{1}{2d_s}\left(\frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_{\tau=\tau}\, 2d_s + O(d_s^2)\right) = \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_{\tau=\tau} + O(d_s).
\end{aligned}$$

□

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Additionally, in order to finish with the examination of $B(x)$'s moments, we are going to establish its variance in the following proposition:

Proposition 3.14. *Under assumptions S1-S5, it is verified that*

$$\mathbb{V}ar\,[B(x)] = \frac{1}{2nd_sh_s}\,\frac{s_\tau(x)^2}{g(x)}\,R(K) + o\left(\frac{1}{nh_sd_s}\right).$$

Proof. In this case we could obtain the following decomposition of the variance of $B(x)$:

$$\begin{aligned}\mathbb{V}\text{ar}\,[B(x)] &= \frac{1}{(2d_s)^2}\,\frac{1}{g(x)^2}\,\mathbb{V}\text{ar}\left[s_{\tau+d_s}(x)\,U_{\tau+d_s,h_s}(x) - s_{\tau-d_s}(x)\,U_{\tau-d_s,h_s}(x)\right]\\ &= \frac{1}{(2d_s)^2}\,\frac{1}{g(x)^2}\Big(s_{\tau+d_s}(x)\,B_{1,\tau+d_s}(x) + s_{\tau-d_s}(x)\,B_{1,\tau-d_s}(x)\\ &\quad - 2\,s_{\tau+d_s}(x)\,s_{\tau-d_s}(x)\,B_2(x)\Big)\end{aligned}$$

where

$$\begin{aligned}B_{1,\tau}(x) &= \mathbb{V}\text{ar}\,[U_{\tau,h_s}(x)],\\ B_2(x) &= \mathbb{C}\text{ov}\,[U_{\tau+d_s,h_s}(x), U_{\tau+d_s,h_s}(x)].\end{aligned}$$

Then, we are going to focus on studying the functions $B_{1,\tau}(x)$ and $B_2(x)$. Let us denote by $\mathcal{X} = \{X_1, \dots, X_n\}$ a random sample of the explanatory variable X. Hence,

$$\begin{aligned}B_{1,\tau}(x) = \mathbb{V}\text{ar}\,[U_{\tau,h_s}(x)] &= \mathbb{E}(\mathbb{V}\text{ar}\,[U_{\tau,h_s}(x)|\mathcal{X}]) + \mathbb{V}\text{ar}\,(\mathbb{E}[U_{\tau,h_s}(x)|\mathcal{X}])\\ &= \mathbb{E}(\mathbb{V}\text{ar}\,[U_{\tau,h_s}(x)|\mathcal{X}]) + O\left(\frac{h_s^4}{nh_s}\right)\end{aligned}$$

as consequence of the following result obtained by Fan et al. (1994) along the proof of their Lemma 2:

$$\begin{aligned}\mathbb{V}\text{ar}\,(\mathbb{E}[U_{\tau,h_s}(x)|\mathcal{X}]) &= \mathbb{E}(\mathbb{E}[U_{\tau,h_s}(x)|\mathcal{X}]) - (\mathbb{E}(\mathbb{E}[U_{\tau,h_s}(x)|\mathcal{X}]))^2\\ &= \mathbb{E}(\mathbb{E}[U_{\tau,h_s}(x)|\mathcal{X}]) - (\mathbb{E}[U_{\tau,h_s}(x)])^2 = O\left(\frac{h_s^4}{nh_s}\right).\end{aligned}$$

So, it is enough to compute $\mathbb{E}(\mathbb{V}\text{ar}\,[U_{\tau,h_s}(x)|\mathcal{X}])$ where as a consequence of Lemma 3.11 it follows that

$$\begin{aligned}\mathbb{V}\text{ar}\,[U_{\tau,h_s}(x)|\mathcal{X}] &= \frac{1}{(nh_s)^2}\sum_{i=1}^{n}\tau(1-\tau)\;K_i^2\\ &\quad + \frac{1}{(nh_s)^2}\sum_{i=1}^{n}\varphi_2^{(1)}(\xi_{3,i}|X=x)\left(-\frac{1}{2}q_\tau^{(2)}(\xi_{1,i})(X_i-x)^2\right)\;K_i^2\end{aligned}$$

and as a result

$$\mathbb{E}\,(\mathbb{V}\text{ar}\;[U_{\tau,h_s}(x)|\mathcal{X}]) = \mathbb{E}\Bigg[\frac{1}{(nh_s)^2}\sum_{i=1}^{n}\tau(1-\tau)\;K_i^2$$

$$
\begin{aligned}
&+ \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_\tau^{(2)}(\xi_{1,i})(X_i - x)^2 \right) K_i^2 \Big] \\
&= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \tau(1-\tau)\, \mathbb{E}[K_i^2] \\
&+ \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_\tau^{(2)}(\xi_{1,i}) \right) \mathbb{E}[(X_i - x)^2\, K_i^2]
\end{aligned}
$$

where $\xi_{1,i}$ is an element between X_i and x, and $\xi_{3,i}$ is an element between $-\frac{1}{2} q_\tau^{(2)}(\xi_{1,i})(X_i - x)^2$ and zero.

Bearing Lemma 3.3 in mind, it follows that

$$
\begin{aligned}
\mathbb{E}[K_i^2] &= g(x)\, h_s\, R(K) + O(h_s^2) \\
\mathbb{E}\left[(X_i - x)^2\, K_i^2\right] &= g(x)\, h_s^3\, \phi_{2,2}(K) + O(h_s^4)
\end{aligned}
$$

and finally

$$
\begin{aligned}
B_{1,\tau}(x) &= \mathbb{V}\mathrm{ar}\left[U_{\tau,h_s}(x)\right] = \mathbb{E}(\mathbb{V}\mathrm{ar}\left[U_{\tau,h_s}(x)|\mathcal{X}\right]) + O\left(\frac{h_s^4}{nh_s}\right) \\
&= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \Bigg(\tau(1-\tau)\ \left(g(x)\, h\, R(K) + O(h^2)\right) \\
&+ \varphi_2^{(1)}(\xi_{3,i}) \left(-\frac{1}{2} q_\tau^{(2)}(\xi_{1,i}) \right) \left(g(x)\, h_s^3\, \phi_{2,2}(K) + O(h_s^4)\right) \Bigg) + O\left(\frac{h_s^4}{nh_s}\right) \\
&= \frac{1}{nh_s}\, \tau(1-\tau)\, g(x)\, R(K) + O\left(\frac{h_s}{n}\right) \\
&= \frac{1}{nh_s}\, \tau(1-\tau)\, g(x)\, R(K) + o\left(\frac{1}{nh_s}\right).
\end{aligned}
$$

Thereupon, we need to compute $B_2(x)$ in order to determine the variance of $B(x)$. In this case, due to the law of total covariance it is verified that

$$
B_2(x) = \mathbb{C}\mathrm{ov}\left[U_{\tau+d_s,h_s}(x), U_{\tau-d_s,h_s}(x)\right] = \mathbb{E}\left[B_{2,1}(x)\right] + B_{2,2}(x)
$$

where

$$
\begin{aligned}
B_{2,1}(x) &= \mathbb{C}\mathrm{ov}\ \left[U_{\tau+d_s,h_s}(x), U_{\tau-d_s,h_s}(x)|\mathcal{X}\right] \\
B_{2,2}(x) &= \mathbb{C}\mathrm{ov}\ \left[\mathbb{E}\left(U_{\tau+d_s,h_s}(x)|\mathcal{X}\right), \mathbb{E}\left(U_{\tau-d_s,h_s}(x)|\mathcal{X}\right)\right].
\end{aligned}
$$

So we are going to study each of the addends. On the one hand, as a result of Lemma 3.11 and Lemma 3.3 it follows that

$$
\begin{aligned}
\mathbb{E}[B_{2,1}(x)] &= \mathbb{E}\left[\mathbb{C}\mathrm{ov}\ \left(U_{\tau+d_s,h_s}(x), U_{\tau-d_s,h_s}(x)|\mathcal{X}\right)\right] = \mathbb{E}\Bigg[\frac{1}{(nh_s)^2} \sum_{i=1}^{n} \Bigg((\tau - d_s)(1 - \tau - d_s) \\
&+ \left(\frac{\partial \varphi_6(t_1,t_2)}{\partial t_1}, \frac{\partial \varphi_6(t_1,t_2)}{\partial t_2} \right)' \Bigg|_{(\xi_{6,i},\xi_{7,i})} \\
&\times \left(-\frac{1}{2} q_{\tau+d_s}^{(2)}(\xi_{4,i})(X_i - x)^2, -\frac{1}{2} q_{\tau-ds}^{(2)}(\xi_{5,i})(X_i - x)^2 \right) \Bigg)\ K_i^2 \Bigg]
\end{aligned}
$$

$$= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \Bigg((\tau - d_s)(1-\tau-d_s)\mathbb{E}\left[K_i^2\right]$$
$$+ \left(\frac{\partial \varphi_6(t_1,t_2)}{\partial t_1}, \frac{\partial \varphi_6(t_1,t_2)}{\partial t_2} \right)' \Bigg|_{(\xi_{6,i},\xi_{7,i})}$$
$$\times \left(-\frac{1}{2} q^{(2)}_{\tau+d_s}(\xi_{4,i}), -\frac{1}{2} q^{(2)}_{\tau-d_s}(\xi_{5,i}) \right) \mathbb{E}\left[(X_i-x)^2 K_i^2\right] \Bigg)$$
$$= \frac{1}{nh_s}(\tau - d_s)(1-\tau-d_s)\; g(x)\; R(K) + O\left(\frac{h_s}{n}\right)$$
$$= \frac{1}{nh_s}(\tau - d_s)(1-\tau-d_s)\; g(x)\; R(K) + o\left(\frac{1}{nh_s}\right).$$

On the other side, it will be necessary to compute

$$B_{2,2}(x) = \mathbb{C}\text{ov}\left[\mathbb{E}\left(U_{\tau+d_s,h_s}(x)|\mathcal{X}\right), \mathbb{E}\left(U_{\tau-d_s,h_s}(x)|\mathcal{X}\right)\right]$$

in order to finish the study of $B_2(x)$. Thus, the calculus of $\mathbb{E}[U_{\tau,h_s}(x)|\mathcal{X}]$ comes from Lemma 3.11, that is

$$\mathbb{E}\left[U_{\tau,h_s}(x)|\mathcal{X}\right] = \frac{1}{nh_s} \sum_{i=1}^{n} \left(\tau - \mathbb{E}\left[\mathbb{I}\left(Y_i^{(1,\tau)} < 0\right) \Big| \mathcal{X} \right] \right) K_i$$
$$= \frac{1}{nh_s} \sum_{i=1}^{n} f(q_\tau(x) + \xi_{2,i}|X=x)\; \frac{1}{2} q_\tau^{(2)}(\xi_{1,i})(X_i - x)^2\; K_i$$

and as a result

$$B_{2,2}(x) = \mathbb{C}\text{ov}\left[\mathbb{E}\left(U_{\tau+d_s,h_s}(x)|\mathcal{X}\right), \mathbb{E}\left(U_{\tau-d_s,h_s}(x)|\mathcal{X}\right)\right]$$
$$= \mathbb{C}\text{ov}\Bigg[\frac{1}{nh_s} \sum_{i=1}^{n} f(q_{\tau+d_s}(x) + \xi_{14,i}|X=x)\; \frac{1}{2} q^{(2)}_{\tau+d_s}(\xi_{15,i})(X_i-x)^2\; K_i,$$
$$\frac{1}{nh_s} \sum_{j=1}^{n} f(q_{\tau-d_s}(x) + \xi_{16,j}|X=x)\; \frac{1}{2} q^{(2)}_{\tau-d_s}(\xi_{17,j})(X_j-x)^2\; K_i \Bigg]$$
$$= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} \mathbb{C}\text{ov}\Bigg[f(q_{\tau+d_s}(x) + \xi_{14,i}|X=x)\; \frac{1}{2} q^{(2)}_{\tau+d_s}(\xi_{15,i})(X_i-x)^2\; K_i,$$
$$f(q_{\tau-d_s}(x) + \xi_{16,j}|X=x)\; \frac{1}{2} q^{(2)}_{\tau-d_s}(\xi_{17,j})(X_i-x)^2\; K_i \Bigg]$$
$$\leq \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} M_1^2\; M_2^2 \mathbb{C}\text{ov}\left[(X_i-x)^2\; K_i, (X_j-x)^2\; K_i\right]$$
$$= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} M_1^2\; M_2^2 (\mathbb{E}[(X_i-x)^4\; K_i^2] - (\mathbb{E}[(X_i-x)^2\; K_i])^2$$
$$= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} M_1^2\; M_2^2 [O(h_s^5) - O(h_s^6)] = O\left(\frac{h_s^3}{n}\right) = o\left(\frac{1}{nh_s}\right)$$

where M_1 and M_2 represents bounds of the function f and $q_\tau^{(2)}$, respectively. Moreover, $\xi_{15,i}$ is an element between X_i and x, $\xi_{17,j}$ is an element between X_j and x, $\xi_{14,i}$ is an element between

$-\frac{1}{2}q^{(2)}_{\tau+d_s}(\xi_{15,i})(X_i-x)^2$ and zero, and $\xi_{16,j}$ is an element between $-\frac{1}{2}q^{(2)}_{\tau-d_s}(\xi_{17,j})(X_j-x)^2$ and zero.

Finally,

$$\begin{aligned} B_2(x) &= \mathbb{C}\mathrm{ov}\left[U_{\tau+d_s,h_s}(x), U_{\tau-d_s,h_s}(x)\right] = \mathbb{E}\left[B_{2,1}(x)\right] + B_{2,2}(x) \\ &= \frac{1}{nh_s}(\tau-d)(1-\tau-d)g(x)R(K) + o\left(\frac{1}{nh_s}\right) \end{aligned}$$

and it follows that

$$\begin{aligned} \mathbb{V}\mathrm{ar}\left[B(x)\right] = & \frac{1}{(2d_s)^2}\,\frac{1}{g(x)^2}\,\Big(s_{\tau+d_s}(x)^2\,B_{1,\tau+d_s}(x)\ + s_{\tau-d_s}(x)^2\,B_{1,\tau-d_s}(x) \\ & - 2\,s_{\tau+d_s}(x)\,s_{\tau-d_s}(x)\,B_2(x)\Big) \\ & = \frac{1}{(2d_s)^2}\,\frac{1}{g(x)^2}\,\Bigg[s_{\tau+d_s}(x)^2\left(\frac{1}{nh_s}\,(\tau+d_s)(1-\tau-d_s)\,g(x)\,R(K)+o\left(\frac{1}{nh_s}\right)\right) \\ & + s_{\tau-d_s}(x)^2\left(\frac{1}{nh_s}\,(\tau-d_s)(1-\tau+d_s)\,g(x)\,R(K)+o\left(\frac{1}{nh_s}\right)\right) \\ & - 2\,s_{\tau+d_s}(x)\,s_{\tau-d_s}(x)\left(\frac{1}{nh_s}(\tau-d_s)(1-\tau-d_s)\,g(x)\,R(K)+o\left(\frac{1}{nh_s}\right)\right)\Bigg] \\ & = \frac{1}{(2d_s)^2}\,\frac{1}{g(x)}\,\frac{1}{nh_s}\,R(K)\,\Big[s_{\tau+d_s}(x)^2\,(\tau+d_s)(1-\tau-d_s) \\ & + s_{\tau-d_s}(x)^2\,(\tau-d_s)(1-\tau+d_s) - 2s_{\tau+d_s}(x)\,s_{\tau-d_s}(x)\,(\tau-d_s)(1-\tau-d_s)\Big] \\ & + o\left(\frac{1}{nh_sd_s}\right). \end{aligned}$$

If we define the functions

$$\begin{aligned} &\Gamma_1(\tau) = s_\tau(x)\,\tau \quad \text{with} \quad \Gamma_1^{(1)}(\tau) = s_\tau^{(1,\tau)}(x)\tau + s_\tau(x) \\ &\Gamma_2(\tau) = s_\tau(x)\,(1-\tau) \quad \text{with} \quad \Gamma_2^{(1)}(\tau) = s_\tau^{(1,\tau)}(x)(1-\tau) - s_\tau(x) \end{aligned}$$

then we can write

$$\mathbb{V}\mathrm{ar}\left[B(x)\right] = \frac{1}{2d_s}\,\frac{1}{g(x)}\,\frac{1}{nh_s}\,R(K)\,B_3(x) + o\left(\frac{1}{nh_sd_s}\right)$$

where

$$\begin{aligned} B_3(x) &= \frac{1}{2d_s}\Big[\Gamma_1(\tau+d_s)\,\Gamma_2(\tau+d_s) + \Gamma_1(\tau-d_s)\,\Gamma_2(\tau-d_s) - 2\Gamma_1(\tau-d_s)\,\Gamma_2(\tau+d_s)\Big] \\ &= \frac{1}{2d_s}\Big[\Big(\Gamma_1(\tau+d_s)-\Gamma_1(\tau-d_s)\Big)\,\Gamma_2(\tau+d_s) + \Big(\Gamma_2(\tau-d_s)-\Gamma_2(\tau+d_s)\Big)\,\Gamma_1(\tau-d)\Big] \\ &\overset{(i)}{=} \Gamma_1^{(1)}(\tau)\,\Gamma_2(\tau+d_s) - \Gamma_2^{(1)}(\tau)\,\Gamma_1(\tau-d_s) + O(d_s) \\ &\overset{(ii)}{=} \Gamma_1^{(1)}(\tau)\,\Gamma_2(\tau) - \Gamma_2^{(1)}(\tau)\,\Gamma_1(\tau) + O(d_s) \\ &= (s_\tau^{(1,\tau)}(x)\tau + s_\tau(x))s_\tau(x)\,(1-\tau) - s_\tau(x)\tau(s_\tau^{(1,\tau)}(x)(1-\tau) - s_\tau(x)) + O(d_s) \end{aligned}$$

$$= s_\tau(x)^2 + O(d_s)$$

where (i) comes from the following Taylor expansions

$$\Gamma_1(\tau + d_s) = \Gamma_1(\tau) + \Gamma_1^{(1)}(\tau)d_s + \frac{1}{2}\Gamma_1^{(2)}(\xi_{18})d_s^2 \quad \text{where } \xi_{18} \in (\tau, \tau + d_s)$$
$$\Gamma_1(\tau - d_s) = \Gamma_1(\tau) + \Gamma_1^{(1)}(\tau)(-d_s) + \frac{1}{2}\Gamma_1^{(2)}(\xi_{19})(-d_s)^2 \quad \text{where } \xi_{19} \in (\tau - d_s, \tau)$$

and as a result

$$\Gamma_1(\tau + d_s) - \Gamma_1(\tau + d_s) = 2\Gamma_1(\tau)d + \frac{1}{2}\left(\Gamma_1(\xi) - \Gamma_1(\xi)\right) d_s^2 = 2\Gamma_1(\tau)d + O(d_s^2).$$

Analogously for Γ_2. Moreover, (ii) is consequence of the following Taylor expansions

$$\Gamma_1(\tau - d_s) = \Gamma_1(\tau) + \Gamma_1^{(1)}(\xi_{20})(-d_s) = \Gamma_1(\tau) + O(d_s) \quad \text{where } \xi_{20} \in (\tau - d_s, \tau)$$
$$\Gamma_2(\tau + d_s) = \Gamma_2(\tau) + \Gamma_2^{(1)}(\xi_{21})d_s = \Gamma_2(\tau) + O(d_s) \quad \text{where } \xi_{21} \in (\tau, \tau + d_s)$$

Finally we conclude that

$$\mathbb{V}\text{ar}\left[B(x)\right] = \frac{1}{2nd_sh_s}\,\frac{1}{g(x)}\,R(K)\,s_\tau(x)^2 + o\left(\frac{1}{nh_sd_s}\right).$$

□

Therefore, in view of all the results shown along this subsection, the bias and variance of the sparsity estimator given by (3.35) can be established. These results are gathered in the following theorem:

Theorem 3.15. *If assumptions S1-S5 follows, then*

$$\mathbb{B}ias\left[\widehat{s}_{\tau,d_s,h_s}(x)\right] \cong \frac{1}{6}\,s_\tau^{(2,\tau)}(x)\,d_s^2 + \frac{1}{2}\,\mu_2(K)\,\left.\frac{\partial q_\tau^{(2)}(x)}{\partial\tau}\right|_\tau h_s^2$$

$$\mathbb{V}ar\left[\widehat{s}_{\tau,d_s,h_s}(x)\right] \cong \frac{1}{2nd_sh_s}\,\frac{1}{g(x)}\,R(K)\,s_\tau(x)^2$$

Proof. Let us remember that

$$\widehat{s}_{\tau,d_s,h_s}(x) = \frac{\widehat{q}_{\tau+d_s,h_s}(x) - \widehat{q}_{\tau-d_s,h_s}(x)}{2\,d_s} = A(x) + B(x) + o\left((nh_s)^{-1/2}d_s^{-1}\right)$$

then as a consequence of Lemma 3.12 and Proposition 3.13 it follows that

$$\begin{aligned}\mathbb{B}\text{ias}\left[\widehat{s}_{\tau,d_s,h_s}(x)\right] &= \mathbb{E}[\widehat{s}_{\tau,d_s,h_s}(x)] - s_\tau(x) \cong \mathbb{E}[A(x) + B(x)] - s_\tau(x)\\ &\cong s_\tau(x) + \frac{1}{6}\,s_\tau^{(2,\tau)}(x)\,d_s^2 + \frac{1}{2}\,\mu_2(K)\,\left.\frac{\partial q_\tau^{(2)}(x)}{\partial\tau}\right|_\tau h_s^2 - s_\tau(x)\\ &= \frac{1}{6}\,s_\tau^{(2,\tau)}(x)\,d_s^2 + \frac{1}{2}\,\mu_2(K)\,\left.\frac{\partial q_\tau^{(2)}(x)}{\partial\tau}\right|_\tau h_s^2.\end{aligned}$$

Moreover as a result of Proposition 3.14, it is clear that

$$\mathbb{V}\text{ar}\left[\widehat{s}_{\tau,d_s,h_s}(x)\right] \cong \mathbb{V}\text{ar}\left[A(x) + B(x)\right] = \mathbb{V}\text{ar}\left[B(x)\right] \cong \frac{1}{2nd_sh_s}\,\frac{1}{g(x)}\,R(K)\,s_\tau(x)^2.$$

□

Remark 3.3. By means of Proposition 3.13, we have proved that

$$\mathbb{E}[B(x)] = \frac{1}{2}\,\mu_2(K)\,\frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau\, h_s^2 + o(h_s^2).$$

Observe that the quantity $\frac{\partial q_\tau^{(2)}(x)}{\partial\tau}$ is zero for any homoscedastic quantile regression model. That is, the dominant term of the expectation is zero and it will be necessary to consider the newt term of lower order. This new term will be associated with the new addend of the Taylor expansion detailed in (3.39). So, the following expression is justified:

$$\begin{aligned}\mathbb{E}\left[U_{\tau_1,h_s}(x)|\mathcal{X}\right] = \frac{1}{nh_s}\sum_{i=1}^{n}\Bigg[&f(q_{\tau_1}(X_i)|X=X_i)\,\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\\ &-\frac{1}{8}\,f^{(1)}(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\,q_{\tau_1}^{(2)}(\xi_{1,i})^2(X_i-x)^4\Bigg]K_i\end{aligned}$$

where $\xi_{1,i}$ is an element between X_i and x, while $\xi_{2,i}$ is an element between $-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2$ and zero. In view of previous equation and Lemma 3.3 it follows that

$$\begin{aligned}\mathbb{E}\left[U_{\tau_1,h_s}(x)\right] &= \mathbb{E}\left[\mathbb{E}\left[U_{\tau_1,h_s}(x)|\mathcal{X}\right]\right]\\ &=\frac{1}{2}\,f(q_{\tau_1}(x)|X=x)\,q_{\tau_1}^{(2)}(x)\,g(x)\,\mu_2(K)\,h_s^2\\ &-\frac{1}{8}\,f^1(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\,q_{\tau_1}^{(2)}(\xi_{1,i})^2\,g(x)\,\mu_4(K)\,h_s^4+o(h_s^4)\end{aligned}$$

and as a result

$$\begin{aligned}\mathbb{E}[B(x)] &= \frac{1}{2d_s}\,\mathbb{E}\left[\frac{s_{\tau+d_s}(x)}{g(x)}\,U_{\tau+d_s,h_s}(x)-\frac{s_{\tau-d_s}(x)}{g(x)}\,U_{\tau-d_s,h_s}(x)\right]\\ &=\frac{1}{2d_s\,g(x)}\Big(s_{\tau+d_s}(x)\,\mathbb{E}[U_{\tau+d_s,h_s}(x)]-s_{\tau-d_s}(x)\,\mathbb{E}[U_{\tau-d_s,h_s}(x)]\Big)\\ &=B_1(x)+B_2(x)+o(h_s^4)\end{aligned}$$

where

$$B_1(x)=\frac{1}{2}\,\mu_2(K)\,h_s^2\,\left(\frac{q_{\tau+d_s}^{(2)}(x)-q_{\tau-d_s}^{(2)}(x)}{2d_s}\right)=\frac{1}{2}\,\mu_2(K)\,h_s^2\,\left(\frac{\partial q_\tau^{(2)}(x)}{\partial\tau}+O(d_s)\right)$$

$$\begin{aligned}B_2(x) &= \frac{-1}{8d_s}\mu_4(K)\,h_s^4\Big(f^{(1)}(q_{\tau+ds}(x)|X=x)\,s_{\tau+d_s}(x)\,q_{\tau+ds}^{(2)}(x)^2\\ &\quad-f^{(1)}(q_{\tau-ds}(x)|X=x)\,s_{\tau-d_s}(x)\,q_{\tau-ds}^{(2)}(x)^2\Big)\\ &=\frac{-1}{4}\mu_4(K)\,h_s^4\left(\frac{\partial\left[f^{(1)}(q_{\tau+ds}(x)|X=x)\,s_\tau(x)\,q_{\tau+ds}^{(2)}(x)^2\right]}{\partial\tau}+O(d_s)\right).\end{aligned}$$

Summarizing, the quantity $B_1(x)$ match up to the result of Proposition 3.13 while $B_2(x)$ is the new term obtained in order to avoid degenerate bandwidth h_s associated with homoscedastic quantile regression models.

Computational aspects:

In order to apply the previous development in practice, note that

$$\begin{aligned}\frac{\partial}{\partial \tau}\left[s_\tau(x)\ f^{(1)}(q_\tau(x)|X=x)\ q_\tau^{(2)}(x)^2\right] &= s_\tau^{(1,\tau)}(x)\ f^{(1)}(q_\tau(x)|X=x)\ q_\tau^{(2)}(x)^2 \\ &+ s_\tau(x)\ f^{(2)}(q_\tau(x)|X=x)\ \frac{\partial}{\partial\tau}q_\tau(x)\ q_\tau^{(2)}(x)^2 \\ &+ s_\tau(x)\ f^{(1)}(q_\tau(x)|X=x)\ 2q_\tau^{(2)}(x)\frac{\partial}{\partial\tau}q_\tau^{(2)}(x) \\ &= -s_\tau(x)^3\ f^{(1)}(q_\tau(x)|X=x)^2\ q_\tau^{(2)}(x)^2 \\ &+ s_\tau(x)^2\ f^{(2)}(q_\tau(x)|X=x)\ q_\tau^{(2)}(x)^2 \\ &+ s_\tau(x)\ f^{(1)}(q_\tau(x)|X=x)\ 2q_\tau^{(2)}(x)\frac{\partial}{\partial\tau}q_\tau^{(2)}(x)\end{aligned}$$

where the quantities $s_\tau(x)$ and $q_\tau^{(2)}(x)$ can be estimated thanks to the rule of thumb, while the estimation of $\frac{\partial}{\partial\tau}q_\tau^{(2)}(x)$ has been detailed in order to compute $\int c(x)dx$ (Section 3.2.2).

Then, the novel ingredient will be to estimate the derivatives of the conditional density. Following the "blocks idea" in which the rule of thumb is based, we will estimate the conditional density at $x \in \text{Block}j$ as follows

$$\widehat{f}^{(r)}(x) = \frac{1}{n_j h_{fr}^{r+1}}\sum_{i=1}^{n_j} K^{(r)}\left(\frac{\widehat{q}_{\tau,j}(X_i)-x}{h_{fr}}\right)\ \mathbb{I}(X_i \in \text{Block}j)$$

where n_j represents the number of observations at block j and h_{fr} is a bandwidth parameter. In order to obtain an estimation of h_{fr} we will use the selector proposed by Scott (1992) (page 131-132) assuming that the error ε follows a Gaussian distribution.

3.4.3 Expectation and variance of the integrated squared sparsity estimator

In this subsection, the mean squared error of the integrated squared sparsity estimator will be studied. So, we should remember that

$$\text{MSE}\left[\int \widehat{s}_{\tau,d_s,h_s}(x)^2\ dx\right] = \left(\mathbb{B}\text{ias}\left[\int \widehat{s}_{\tau,d_s,h_s}(x)^2\ dx\right]\right)^2 + \mathbb{V}\text{ar}\left[\int \widehat{s}_{\tau,d_s,h_s}(x)^2\ dx\right]$$

We are going to start computing the bias of the integrated squared sparsity estimator that is given in the following theorem:

Theorem 3.16. *Under conditions S1-S5, it is proved that*

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

where

$$a(x) = \frac{R(K)}{2}\ \frac{s_\tau(x)^2}{g(x)}$$

$$b(x) = \frac{1}{3}\, s_\tau(x)\, s_\tau^{(2,\tau)}(x)$$
$$c(x) = \mu_2(K)\, s_\tau(x)\, \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau .$$

Proof. It follows that

$$\begin{aligned}
\mathbb{B}\text{ias}\left[\int \widehat{s}_{\tau,d_s,h_s}(x)^2\, dx\right] &= \mathbb{E}\left[\int \widehat{s}_{\tau,d_s,h_s}(x)^2\, dx\right] - \int s_\tau(x)^2\, dx \\
&= \int \mathbb{E}\left[\widehat{s}_{\tau,d_s,h_s}(x)^2\right] dx - \int s_\tau(x)^2\, dx \\
&= \int \Big(\mathbb{E}[\widehat{s}_{\tau,d_s,h_s}(x)]^2 + \mathbb{V}\text{ar}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\Big)\, dx - \int s_\tau(x)^2\, dx \\
&= \int \Big(\Big(s_\tau(x) + \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\Big)^2 + \mathbb{V}\text{ar}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\Big)\, dx \\
&- \int s_\tau(x)^2\, dx \\
&= \int s_\tau(x)^2\, dx + \int \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]^2\, dx \\
&+ 2\int s_\tau(x)\, \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\, dx \\
&+ \int \mathbb{V}\text{ar}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\, dx - \int s_\tau(x)^2\, dx \\
&= \int \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]^2\, dx + 2\int s_\tau(x)\, \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\, dx \\
&+ \int \mathbb{V}\text{ar}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\, dx.
\end{aligned}$$

Firstly, it is noticed that $\int \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]^2 dx$ is negligible compared with $\int s_\tau(x)\, \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)] dx$ as a consequence of Theorem 3.15. Indeed, let us remember that

$$\begin{aligned}
\int \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]^2\, dx &\cong \int \left(\frac{1}{6}\, s_\tau^{(2,\tau)}(x)\, d_s^2 + \frac{1}{2}\, \mu_2(K)\, \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau h_s^2\right)^2 dx \\
&= d_s^4 \int \frac{1}{36}\, s_\tau^{(2,\tau)}(x)^2\, dx + h_s^4 \int \frac{1}{4}\, \mu_2(K)^2 \left(\frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau\right)^2 dx \\
&+ 2\, d_s^2\, h_s^2 \int \frac{1}{12}\, s_\tau^{(2,\tau)}(x)\ \ \mu_2(K)\, \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau dx
\end{aligned}$$

while

$$2\int s_\tau(x)\, \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\, dx = \frac{1}{3}\, d_s^2 \int s_\tau(x)\, s_\tau^{(2,\tau)}(x)\, dx + \mu_2(K)\ \ h_s^2 \int s_\tau(x)\, \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau dx.$$

So it is concluded that

$$\mathbb{B}\text{ias}\left[\int \widehat{s}_{\tau,+d_s,h_s}(x)^2\, dx\right] \cong 2\int s_\tau(x)\, \mathbb{B}\text{ias}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\, dx + \int \mathbb{V}\text{ar}\,[\widehat{s}_{\tau,d_s,h_s}(x)]\, dx$$

$$\cong \frac{1}{3}\, d_s^2 \int s_\tau(x)\, s_\tau^{(2,\tau)}(x)\, dx + \mu_2(K)\;\; h_s^2 \int s_\tau(x)\, \frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\bigg|_\tau dx$$
$$+ \frac{1}{2nd_sh_s}\, R(K) \int \frac{s_\tau(x)^2}{g(x)}\, dx.$$

□

Now, we are going to calculate the asymptotic variance of the integrated squared sparsity estimator. Let us remember that

$$\widehat{s}_{\tau,d_s,h_s}(x) = A(x) + B(x) + o\left(\frac{1}{d_s\sqrt{nh_s}}\right).$$

Since $A(x)$ is not random, the asymptotic variance can be decomposed as

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\int \widehat{s}_{\tau,d_s,h_s}(x)^2\, dx\right] &\cong \mathbb{V}\text{ar}\left[\int \left(A(x)^2 + B(x)^2 + 2A(x)B(x)\right)\, dx\right]\\
&= \mathbb{V}\text{ar}\left[\int \left(B(x)^2 + 2A(x)B(x)\right)\, dx\right]\\
&= \mathbb{V}\text{ar}\left[\int B(x)^2\, dx\right] + 4\ \mathbb{V}\text{ar}\left[\int A(x)B(x)\, dx\right]\\
&+ 4\ \mathbb{C}\text{ov}\left[\int B(x)^2\, dx, \int A(x)B(x)\, dx\right]. \qquad (3.44)
\end{aligned}$$

To study each of the three last addends let us recall that

$$A(x) = \frac{q_{\tau+d_s}(x) - q_{\tau-d_s}(x)}{2d_s} = s_\tau(x) + O(d_s^2)$$

as a consequence of Lemma 3.12.

Now, the three addends in expression (3.44) will involve calculus about the covariance between different U_{τ,h_s} functions. These results will be established in the following Lemma:

Lemma 3.17. *Under conditions S1-S4, it follows that*

$$\mathbb{C}ov\left(U_{\tau_1,h_s}(x_1), U_{\tau_2,h_s}(x_2)\right) \cong \frac{1}{nh_s}\ \left(\min\{\tau_1,\tau_2\} - \tau_1\tau_2\right)\ K*K\left(\frac{x_2-x_1}{h_s}\right)\ g(x_1)$$

$$\begin{aligned}
\mathbb{C}ov\left(U_{\tau_1,h_s}(x_1)\ U_{\tau_2,h_s}(x_2), U_{\tau_3,h_s}(x_3)\right) &\cong \left(\frac{1}{nh_s}\right)^2 \Gamma_3(\tau_1,\tau_2,\tau_3)\ K^2*K\left(\frac{x_2-x_1}{h_s}\right)\ g(x_1)\\
&+ \frac{h_s}{2n}\, g(x_1)^2\, \mu_2(K)\ K*K\left(\frac{x_2-x_1}{h_s}\right)\\
&\times \left((\min\{\tau_2,\tau_3\} - \tau_2\tau_3) f(q_{\tau_1}(x_1)|X=x_1)\ q_{\tau_1}^{(2)}(x_1)\right.\\
&+ \left.(\min\{\tau_1,\tau_3\} - \tau_1\tau_3) f(q_{\tau_2}(x_1)|X=x_1)\ q_{\tau_2}^{(2)}(x_1)\right)
\end{aligned}$$

$$\mathbb{C}ov\left(U_{\tau_1,h_s}(x_1)U_{\tau_2,h_s}(x_1), U_{\tau_3,h_s}(x_2)U_{\tau_4,h_s}(x_2)\right) \cong$$

$$\cong \left(\frac{1}{nh_s}\right)^2 \Gamma_5(\tau_1,\tau_2,\tau_3,\tau_4)\ (K*K)^2\left(\frac{x_2-x_1}{h_s}\right)\ g(x_1)^2$$
$$+\frac{h_s^3}{4n}\ K*K\left(\frac{x_2-x_1}{h_s}\right)\ \mu_2(K)^2\ g(x_1)^2\ g(x_2)$$
$$\times\Big(\Gamma_6(\tau_1,x_1)\ \Gamma_6(\tau_3,x_2)\ \Gamma_7(\tau_2,\tau_4)+\Gamma_6(\tau_1,x_1)\ \Gamma_6(\tau_4,x_2)\ \Gamma_7(\tau_2,\tau_3)$$
$$+\Gamma_6(\tau_2,x_1)\ \Gamma_6(\tau_3,x_2)\ \Gamma_7(\tau_1,\tau_4)+\Gamma_6(\tau_2,x_1)\ \Gamma_6(\tau_4,x_2)\ \Gamma_7(\tau_1,\tau_3)\Big)$$

uniformly in τ_1, τ_2, τ_3 *and* τ_4 *where these elements are in a neighbourhood of* τ. *Here,*

$$\Gamma_4(\tau_1,\tau_2,\tau_3)=\tau_1\min\{\tau_2,\tau_3\}+\tau_2\min\{\tau_2,\tau_3\}+\tau_3\min\{\tau_1,\tau_2\}-2\tau_1\tau_2\tau_3 - \min\{\tau_1,\tau_2,\tau_3\}$$

$$\Gamma_7(\tau_i,x_j)=f(q_{\tau_i}(x_j)|X=x_j)\ q_{\tau_i}^{(2)}(x_j)$$

$$\Gamma_8(\tau_i,\tau_j)=\min\{\tau_i,\tau_j\}-\tau_i\tau_j$$

and

$$\begin{aligned}\Gamma_6(\tau_1,\tau_2,\tau_3,\tau_4)&=2\tau_1\tau_2\tau_3\tau_4-\tau_2\tau_4\min\{\tau_1,\tau_3\}-\tau_1\tau_4\min\{\tau_2,\tau_3\}\\&-\tau_2\tau_3\min\{\tau_1,\tau_4\}-\tau_1\tau_3\min\{\tau_2,\tau_4\}\\&+\min\{\tau_1,\tau_3\}\ \min\{\tau_2,\tau_4\}+\min\{\tau_1,\tau_4\}\ \min\{\tau_2,\tau_3\}.\end{aligned}$$

Proof. First of all, let us introduce the following notation:

$$K_{i,j}=K\left(\frac{X_i-x_j}{h_s}\right).$$

Covariance between $\mathbf{U_{\tau_1,h_s}(x_1)}$ and $\mathbf{U_{\tau_2,h_s}(x_2)}$

As a consequence of the law of total covariance we can write

$$\begin{aligned}\mathbb{C}\text{ov}\ (U_{\tau_1,h_s}(x_1),U_{\tau_2,h_s}(x_2))&=\mathbb{E}\left[\mathbb{C}\text{ov}\left(U_{\tau_1,h_s}(x_1),U_{\tau_2,h_s}(x_2)\middle|\mathcal{X}\right)\right]\\&+\mathbb{C}\text{ov}\left[\mathbb{E}\left(U_{\tau_1,h_s}(x_1)\middle|\mathcal{X}\right),\mathbb{E}\left(U_{\tau_2,h_s}(x_2)\middle|\mathcal{X}\right)\right]\end{aligned}$$

where $\mathcal{X}=\{X_1,\ldots,X_n\}$ represents a random sample of the explanatory variable that we have denoted by X.

On the one hand, if we denote by $B_4(x_1,x_2)=\mathbb{C}\text{ov}\ (U_{\tau_1,h_s}(x_1),U_{\tau_2,h_s}(x_2)|\mathcal{X})$, we could write

$$\begin{aligned}B_4(x_1,x_2)&=\mathbb{C}\text{ov}\left[\frac{1}{nh_s}\sum_{i=1}^n\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)K_{i,1},\frac{1}{nh_s}\sum_{j=1}^n\psi_{\tau_2}\left(Y_j^{(2,\tau_2)}\right)K_{j,2}\ \middle|\mathcal{X}\right]\\&=\frac{1}{(nh_s)^2}\sum_{i=1}^n\sum_{j=1}^n\mathbb{C}\text{ov}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right),\psi_{\tau_2}\left(Y_j^{(2,\tau_2)}\right)\middle|\mathcal{X}\right]K_{i,1}\ K_{j,2}\end{aligned}$$

$$= \frac{1}{(nh_s)^2} \sum_{i=1}^{n} \mathbb{C}\text{ov} \left[\tau_1 - \mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right), \tau_2 - \mathbb{I}\left(Y_i^{(2,\tau_2)} < 0\right) \middle| \mathcal{X} \right] K_{i,1}\, K_{i,2}$$

where we have taken into account that if $i \neq j$ then $\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)$ and $\psi_{\tau_2}\left(Y_j^{(2,\tau_2)}\right)$ are independent as well as their covariance is zero. Furthermore, let us define the function

$$\begin{aligned}
\varphi_7(t_1,t_2) &= \mathbb{C}\text{ov}\left[\tau_1 - \mathbb{I}\left(Y_i < q_{\tau_1}(X_i) + t_1\right), \tau_2 - \mathbb{I}\left(Y_i < q_{\tau_2}(X_i) + t_2\right) | \mathcal{X}\right] \\
&= \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_i < q_{\tau_1}(X_i) + t_1\right), \mathbb{I}\left(Y_i < q_{\tau_2}(X_i) + t_2\right) | \mathcal{X}\right] \\
&= \mathbb{E}\left[\mathbb{I}\left(Y_i < q_{\tau_1}(X_i) + t_1\right) \mathbb{I}\left(Y_i < q_{\tau_2}(X_i) + t_2\right) | \mathcal{X}\right] \\
&- \mathbb{E}\left[\mathbb{I}\left(Y_i < q_{\tau_1}(X_i) + t_1\right) | \mathcal{X}\right] \mathbb{E}\left[\mathbb{I}\left(Y_i < q_{\tau_2}(X_i) + t_2\right) | \mathcal{X}\right] \\
&= F\left(\min\{q_{\tau_1}(x) + t_1, q_{\tau_2}(x) + t_2\} | X = x\right) \\
&- F\left(q_{\tau_1}(x) + t_1 | X = x\right) F\left(q_{\tau_2}(x) + t_2 | X = x\right)
\end{aligned}$$

whose partial derivatives are given by

$$\begin{aligned}
\frac{\partial \varphi_7(t_1,t_2)}{\partial t_1} &= f\left(\min\{q_{\tau_1}(x) + t_1, q_{\tau_2}(x) + t_2\} | X = x\right) \mathbb{I}(q_{\tau_1}(x) + t_1 \leq q_{\tau_2}(x) + t_2) \\
&- f\left(q_{\tau_1}(x) + t_1 | X = x\right) F\left(q_{\tau_2}(x) + t_2 | X = x\right) \\
\frac{\partial \varphi_7(t_1,t_2)}{\partial t_2} &= f\left(\min\{q_{\tau_1}(x) + t_1, q_{\tau_2}(x) + t_2\} | X = x\right) \mathbb{I}(q_{\tau_2}(x) + t_2 \leq q_{\tau_1}(x) + t_1) \\
&- f\left(q_{\tau_2}(x) + t_2\right) F\left(q_{\tau_1}(X) + t_1 | X = x\right).
\end{aligned}$$

Then, in view of arguments developed in equation (3.36) and a Taylor expansion of function φ_7, it can be determined that

$$\begin{aligned}
&\mathbb{C}\text{ov}\left[\left(\tau_1 - \mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\right), \left(\tau_2 - \mathbb{I}\left(Y_i^{(2,\tau_2)} < 0\right)\right) \middle| \mathcal{X}\right] \\
&= \varphi_7\left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{22,i})(X_i - x_1)^2, -\frac{1}{2} q_{\tau_2}^{(2)}(\xi_{23,i})(X_i - x_2)^2\right) \\
&= \varphi_7(0,0) + \left(\frac{\partial \varphi_7(t_1,t_2)}{\partial t_1}, \frac{\partial \varphi_7(t_1,t_2)}{\partial t_2}\right)' \Bigg|_{(t1,t2)=(\xi_{24,i},\xi_{25,i})} \\
&\times \left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{22,i})(X_i - x_1)^2, -\frac{1}{2} q_{\tau_2}^{(2)}(\xi_{23,i})(X_i - x_2)^2\right) \qquad (3.45)
\end{aligned}$$

where $\xi_{22,i}$ and $\xi_{23,i}$ are elements between Xi and x_1 and between Xi and x_2, respectively, obtained thanks to a Taylor expansion of q_{τ_1} and q_{τ_2}, respectively. Moreover, $\xi_{24,i}$ and $\xi_{25,i}$ represent values between $-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{22,i})(X_i - x_1)^2$ and zero and $-\frac{1}{2} q_{\tau_2}^{(2)}(\xi_{23,i})(X_i - x_2)^2$ and zero, respectively. In addition,

$$\begin{aligned}
\varphi_7(0,0) &= F\left(\min\{q_{\tau_1}(x), q_{\tau_2}(x)\} | X = x\right) - F\left(q_{\tau_1}(x) | X = x\right) F\left(q_{\tau_2}(x) | X = x\right) \\
&= \min\{\tau_1, \tau_2\} - \tau_1 \tau_2.
\end{aligned}$$

As a consequence

$$\mathbb{E}\left[B_4(x_1,x_2)\right] = \mathbb{E}\Bigg[\frac{1}{(nh_s)^2} \sum_{i=1}^{n} \Bigg(\varphi_7(0,0) + \left(\frac{\partial \varphi_7(t_1,t_2)}{\partial t_1}, \frac{\partial \varphi_7(t_1,t_2)}{\partial t_1}\right)' \Bigg|_{(\xi_{24,i},\xi_{25,i})}$$

$$\times \left(-\frac{1}{2} q_{\tau_1}^{(2)}(\xi_{22,i})(X_i - x_1)^2, -\frac{1}{2} q_{\tau_2}^{(2)}(\xi_{23,i})(X_i - x_2)^2 \right) \Big) K_{i,1}\ K_{i,2} \Big]$$
$$\cong \frac{1}{nh_s^2}\ \varphi_7(0,0)\ \mathbb{E}\left[K_{i,1}\ K_{i,2}\right]$$
$$\cong \frac{1}{nh_s}\ \left(\min\{\tau_1, \tau_2\} - \tau_1\tau_2\right)\ K * K \left(\frac{x_2 - x_1}{h_s} \right)\ g(x_1) \tag{3.46}$$

where $*$ represents the convolution and the last step comes from the fact that

$$\begin{aligned}
\mathbb{E}\left[K_{i,1}\ K_{i,2}\right] &= \int K\left(\frac{z - x_1}{h_s}\right)\ K\left(\frac{z - x_2}{h_s}\right)\ g(z)\ dz \\
&= \int K(u)\ K\left(u + \frac{x_1 - x_2}{h_s}\right)\ g(x_1 + uh_s)\ h_s\ du \\
&= K * K\left(\frac{x_2 - x_1}{h_s}\right)\ g(x_1)\ h_s + O(h_s^2).
\end{aligned} \tag{3.47}$$

On the other side, we have to compute

$$\mathbb{C}\mathrm{ov}\left[\mathbb{E}\left(U_{\tau_1,h_s}(x)|\mathcal{X}\right), \mathbb{E}\left(U_{\tau_2,h_s}(x)|\mathcal{X}\right)\right]$$

where the calculus of $\mathbb{E}[U_{\tau,h_s}(x)|\mathcal{X}]$ comes from Lemma 3.11, that is

$$\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[\mathbb{E}\left(U_{\tau_1,h_s}(x)|\mathcal{X}\right), \mathbb{E}\left(U_{\tau_2,h_s}(x)|\mathcal{X}\right)\right] \\
&= \mathbb{C}\mathrm{ov}\left[\frac{1}{nh_s}\sum_{i=1}^{n} f(q_{\tau_1}(x) + \xi_{26,i}|X = x)\ \frac{1}{2} q_{\tau_1}^{(2)}(\xi_{27,i})(X_i - x)^2\ K_i, \right. \\
&\qquad \left. \frac{1}{nh_s}\sum_{j=1}^{n} f(q_{\tau_2}(x) + \xi_{28,j}|X = x)\ \frac{1}{2} q_{\tau_2}^{(2)}(\xi_{29,j})(X_j - x)^2\ K_i \right] \\
&= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} \mathbb{C}\mathrm{ov}\left[f(q_{\tau_1}(x) + \xi_{26,i}|X = x)\ \frac{1}{2} q_{\tau_1}^{(2)}(\xi_{27,i})(X_i - x)^2\ K_i, \right. \\
&\qquad \left. f(q_{\tau_2}(x) + \xi_{28,i}|X = x)\ \frac{1}{2} q_{\tau_2}^{(2)}(\xi_{29,i})(X_i - x)^2\ K_i \right] \\
&\leq \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} M_1^2\ M_2^2 \mathbb{C}\mathrm{ov}\left[(X_i - x)^2\ K_i, (X_i - x)^2\ K_i\right] \\
&= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} M_1^2\ M_2^2 (\mathbb{E}[(X_i - x)^4\ K_i^2] - (\mathbb{E}[(X_i - x)^2\ K_i])^2 \\
&= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} M_1^2\ M_2^2 [O(h_s^5) - O(h_s^6)] = O\left(\frac{h_s^3}{n}\right) = O\left(\frac{1}{nh_s}\right)
\end{aligned} \tag{3.48}$$

where M_1 and M_2 represents bounds of the functions f and $q_\tau^{(2)}$.

So, taking into account (3.48) and (3.46) it follows that

$$\mathbb{C}\mathrm{ov}\left(U_{\tau_1}(x_1), U_{\tau_2}(x_2)\right) \cong \frac{1}{nh_s}\ \left(\min\{\tau_1, \tau_2\} - \tau_1\tau_2\right)\ K * K\left(\frac{x_2 - x_1}{h_s}\right)\ g(x_1)$$

Covariance between $\mathbf{U}_{\tau_1,\mathbf{h_s}}(\mathbf{x_1})\mathbf{U}_{\tau_2,\mathbf{h_s}}(\mathbf{x_1})$ and $\mathbf{U}_{\tau_3,\mathbf{h_s}}(\mathbf{x_2})$

In this case, we can write

$$\begin{aligned}
B_5(x_1,x_2) = \mathbb{C}\mathrm{ov}\left[U_{\tau_1,h_s}(x_1)U_{\tau_2,h_s}(x_1), U_{\tau_3,h_s}(x_2)|\mathcal{X}\right] = \\
= \mathbb{C}\mathrm{ov}\left[\frac{1}{nh_s}\sum_{i=1}^{n}\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)K_{i,1}\,\frac{1}{nh_s}\sum_{j=1}^{n}\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)K_{j,2}\,,\right.\\
\left.\frac{1}{nh_s}\sum_{l=1}^{n}\psi_{\tau_3}\left(Y_l^{(1,\tau_3)}\right)K_{l,2}\middle|\mathcal{X}\right]\\
= \left(\frac{1}{nh_s}\right)^3\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}B_6(x_1,x_2)\,K_{i,1}\,K_{j,1}\,K_{l,2}
\end{aligned}$$

where

$$B_6(x_1,x_2) = \mathbb{C}\mathrm{ov}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)\,,\;\psi_{\tau_3}\left(Y_l^{(1,\tau_3)}\right)\middle|\mathcal{X}\right].$$

Then, the problem has been reduced to compute $B_6(x_1,x_2)$ for different values of the indices $\{i,j,l\}$. First, if $\mathrm{card}(\{i,j,l\}) = 3$ or $i = j \neq l$ then $B_6(x_1,x_2) = 0$ because of independence of observations $Y_1,\ldots,Y_n$. Secondly, let us consider the scenario in which $i \neq j = l$ or $j \neq i = l$. For instance, let us focus on $i \neq j = l$ where as a consequence of equation (3.45) joint withsimilar argument to those employed in Lemma 3.11, it follows that

$$\begin{aligned}
B_6(x_1,x_2) = \mathbb{E}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)\,,\;\psi_{\tau_3}\left(Y_l^{(1,\tau_3)}\right)\middle|\mathcal{X}\right]\\
- \mathbb{E}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)\middle|\mathcal{X}\right]\mathbb{E}\left[\psi_{\tau_3}\left(Y_l^{(1,\tau_3)}\right)\middle|\mathcal{X}\right]\\
= \mathbb{E}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\middle|\mathcal{X}\right]\mathbb{C}\mathrm{ov}\left[\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)\,,\;\psi_{\tau_3}\left(Y_l^{(1,\tau_3)}\right)\middle|\mathcal{X}\right]\\
\cong f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\,\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x)^2\\
\times\left(\min\{\tau_2,\tau_3\}-\tau_2\tau_3+O(h_s^2)\right)
\end{aligned}$$

where $\xi_{1,i}$ represents an element between X_i and x_1, and $\xi_{2,i}$ is an element between $-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x_1)^2$ and zero.

Similarly, we can solve the scenario in which $j \neq i = l$. Then, we are going to move to the last possible situation in which $i = j = l$. In this particular case, let us define the following function:

$$\begin{aligned}
\varphi_8(t_1,t_2,t_3) = \mathbb{C}\mathrm{ov}\left[\psi_{\tau_1}\left(Y_i < q_{\tau_1}(X_i)+t_1\right)\psi_{\tau_2}\left(Y_i < q_{\tau_2}(X_i)+t_2\right),\psi_{\tau_3}\left(Y_i < q_{\tau_3}(X_i)+t_3\right)|\mathcal{X}\right]\\
= \mathbb{E}[(\tau_1-\mathbb{I}\left(Y_i < q_{\tau_1}(X_i)+t_1\right))\;(\tau_2-\mathbb{I}\left(Y_i < q_{\tau_2}(X_i)+t_2\right))\\
\times(\tau_3-\mathbb{I}\left(Y_i < q_{\tau_3}(X_i)+t_3\right))|\mathcal{X}]\\
-\mathbb{E}\left[(\tau_1-\mathbb{I}\left(Y_i < q_{\tau_1}(X_i)+t_1\right))(\tau_2-\mathbb{I}\left(Y_i < q_{\tau_2}(X_i)+t_2\right))|\mathcal{X}\right]\\
\times\mathbb{E}\left[\tau_3-\mathbb{I}\left(Y_i < q_{\tau_3}(X_i)+t_3\right)|\mathcal{X}\right].
\end{aligned}$$

A Taylor expansion of the auxiliary function φ_8 leads to the following:

$$\begin{aligned}
B_6(x_1,x_2) &= \mathbb{C}\text{ov}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}<0\right)\psi_{\tau_2}\left(Y_i^{(1,\tau_2)}<0\right)\ ,\ \psi_{\tau_3}\left(Y_i^{(1,\tau_3)}<0\right)\Big|\mathcal{X}\right]\\
&= \varphi_8\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{26,i})(X_i-x_1)^2, -\frac{1}{2}q_{\tau_2}^{(2)}(\xi_{27,i})(X_i-x_1)^2, -\frac{1}{2}q_{\tau_3}^{(2)}(\xi_{28,i})(X_i-x_2)^2\right)\\
&= \varphi_8(0,0,0) + \left(\frac{\partial\varphi_8(t_1,t_2,t_3)}{\partial t_1}, \frac{\varphi_8(t_1,t_2,t_3)}{\partial t_2}, \frac{\varphi_8(t_1,t_2,t_3)}{\partial t_3}\right)'\Bigg|_{(\xi_{29,i},\xi_{30,i},\xi_{31,i})}\\
&\times\left(-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{26,i})(X_i-x_1)^2, -\frac{1}{2}q_{\tau_2}^{(2)}(\xi_{27,i})(X_i-x_1)^2, -\frac{1}{2}q_{\tau_3}^{(2)}(\xi_{28,i})(X_i-x_2)^2\right)\\
&= \varphi_8(0,0,0) + O(h_s^2)
\end{aligned}$$

where $\xi_{29,i}$, $\xi_{30,i}$ and $\xi_{31,i}$ represent elements between $-\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{26,i})(X_i - x_1)^2$, $-\frac{1}{2}q_{\tau_2}^{(2)}(\xi_{27,i})(X_i-x_1)^2$, $-\frac{1}{2}q_{\tau_3}^{(2)}(\xi_{28,i})(X_i-x_2)^2$ and zero, respectively. Moreover,

$$\begin{aligned}
\varphi_8(0,0,0) &= \mathbb{C}\text{ov}\left[\psi_{\tau_1}\left(Y_i<q_{\tau_1}(X_i)\right)\psi_{\tau_2}\left(Y_i<q_{\tau_2}(X_i)\right), \psi_{\tau_1}\left(Y_i<q_{\tau_3}(X_i)\right)|\mathcal{X}\right]\\
&= \mathbb{C}\text{ov}\left[\left(\tau_1-\left(Y_i<q_{\tau_1}(X_i)\right)\right)\left(\tau_2-\left(Y_i<q_{\tau_2}(X_i)\right)\right), \left(\tau_3-\left(Y_i<q_{\tau_3}(X_i)\right)\right)|\mathcal{X}\right]\\
&= \tau_2\mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_i<q_{\tau_1}(X_i)\right), \mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right]\\
&+ \tau_1\mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_i<q_{\tau_2}(X_i)\right), \mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right]\\
&- \mathbb{C}\text{ov}\left[\mathbb{I}\left(Y_i<q_{\tau_1}(X_i)\right)\ \mathbb{I}\left(Y_i<q_{\tau_2}(X_i)\right), \mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right]\\
&= \tau_2(\mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_1}(X_i)\right)\mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right] - \mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_1}(X_i)\right)\right]\mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right])\\
&+ \tau_1(\mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_2}(X_i)\right)\mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right] - \mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_2}(X_i)\right)\right]\mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right])\\
&- \mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_1}(X_i)\right)\ \mathbb{I}\left(Y_i<q_{\tau_2}(X_i)\right)\ \mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right]\\
&+ \mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_1}(X_i)\right)\ \mathbb{I}\left(Y_i<q_{\tau_2}(X_i)\right)\right]\ \mathbb{E}\left[\mathbb{I}\left(Y_i<q_{\tau_3}(X_i)\right)\right]\\
&= \tau_2(F\left(\min\{q_{\tau_1}(x), q_{\tau_3}(x)\}|X=x\right) - F\left(q_{\tau_1}(x)|X=x\right)\ F\left(q_{\tau_3}(x)|X=x\right))\\
&+ \tau_1(F\left(\min\{q_{\tau_2}(x), q_{\tau_3}(x)\}|X=x\right) - F\left(q_{\tau_2}(x)|X=x\right)\ F\left(q_{\tau_3}(x)|X=x\right))\\
&- F\left(\min\{q_{\tau_1}(x), q_{\tau_2}(x), q_{\tau_3}(x)\}|X=x\right)\\
&+ F\left(\min\{q_{\tau_1}(x), q_{\tau_2}(x)\}|X=x\right)\ F\left(q_{\tau_3}(x)|X=x\right)\\
&= \tau_2(\min\{\tau_1,\tau_3\}-\tau_1\tau_3) + \tau_1(\min\{\tau_2,\tau_3\}-\tau_2\tau_3) - \min\{\tau_1,\tau_2,\tau_3\}\\
&+ \min\{\tau_1,\tau_2\}\ \tau_3 = \Gamma_4(\tau_1,\tau_2,\tau_3).
\end{aligned}$$

As a result

$$\begin{aligned}
\mathbb{E}[B_5(x_1,x_2)] &= \left(\frac{1}{nh_s}\right)^3\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\mathbb{E}\Big[B_6(x_1,x_2)\ K_{i,1}\ K_{j,1}\ K_{l,2}\Big]\\
&= \left(\frac{1}{nh_s}\right)^3\sum_{i=1}^n\Gamma_4(\tau_1,\tau_2,\tau_3)\ \mathbb{E}\left[K_{i,1}^2\ K_{i,2}\right]\\
&+ \left(\frac{1}{nh_s}\right)^3\ (\min\{\tau_2,\tau_3\}-\tau_2\tau_3)\ \sum_{i\neq j}\mathbb{E}\Big[f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\\
&\quad\times\frac{1}{2}q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x_1)^2\ K_{i,1}\ K_{j,1}\ K_{j,2}\Big]\\
&+ \left(\frac{1}{nh_s}\right)^3\ (\min\{\tau_1,\tau_3\}-\tau_1\tau_3)\ \sum_{i\neq j}\mathbb{E}\Big[f(q_{\tau_2}(X_j)+\xi_{2,j}|X=X_j)
\end{aligned}$$

$$
\begin{aligned}
&\times \frac{1}{2} q_{\tau_2}^{(2)}(\xi_{1,j})(X_j - x_1)^2 \; K_{i,1} \; K_{j,1} \; K_{i,2}\Bigg] \\
&\cong \left(\frac{1}{nh_s}\right)^2 \Gamma_4(\tau_1, \tau_2, \tau_3) \; K^2 * K \left(\frac{x_2 - x_1}{h_s}\right) \; g(x_1) \\
&+ \frac{h_s}{2n} \; g(x_1)^2 \; \mu_2(K) \; K * K \left(\frac{x_2 - x_1}{h_s}\right) \\
&\times \Bigg((\min\{\tau_2, \tau_3\} - \tau_2\tau_3) f(q_{\tau_1}(x_1)|X = x_1) \; q_{\tau_1}^{(2)}(x_1) \\
&+ (\min\{\tau_2, \tau_3\} - \tau_1\tau_3) f(q_{\tau_2}(x_1)|X = x_1) \; q_{\tau_2}^{(2)}(x_1) \Bigg). \qquad (3.49)
\end{aligned}
$$

Note that the last step of the previous expression comes from the following developments related to the kernel function:

$$
\begin{aligned}
\mathbb{E}\left[K_{i,1}^2 \; K_{i,2}\right] &= \int K \left(\frac{X_i - x_1}{h_s}\right)^2 \; K \left(\frac{X_i - x_2}{h_s}\right) \; g(X_i) \; dX_i \\
&= \int K\,(u)^2 \;\; K \left(u + \frac{x_1 - x_2}{h_s}\right) \; g(x_1 + uh_s) \; h_s \; du \\
&= K^2 * K \left(\frac{x_2 - x_1}{h_s}\right) \; g(x_1) \; h_s + O(h_s^2)
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\Bigg[f(q_{\tau_1}(X_i) + \xi_{2,i}|X = X_i) \; q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x_1)^2 \; K_{i,1} \; K_{j,1} \; K_{j,2}\Bigg] \\
&= \int f(q_{\tau_1}(X_i) + \xi_{2,i}|X = X_i) \; q_{\tau_1}^{(2)}(\xi_{1,i})(X_i - x_1)^2 \; K \left(\frac{X_i - x_1}{h_s}\right) \; K \left(\frac{X_j - x_1}{h_s}\right) \\
&\times K \left(\frac{X_j - x_2}{h_s}\right) \; g(X_i) \; g(X_j) \; dX_i \; dX_j \\
&= \int f(q_{\tau_1}(x_1 + h_s u) + \xi_{2,i}|X = x_1 + h_s u) \; q_{\tau_1}^{(2)}(\xi_{1,i})(h_s u)^2 \; K\,(u) \;\; K\,(v) \\
&\times K \left(v + \frac{x_1 - x_2}{h_s}\right) \; g(x_1 + h_s u) \; g(x_1 + h_s v) \; h_s \; du \; h_s \; dv \\
&= f(q_{\tau_1}(x_1)|X = x_1) \; q_{\tau_1}^{(2)}(x_1) \; g(x_1)^2 \; h_s^4 \int u^2 \; K\,(u) \;\; K\,(v) \;\; K \left(v + \frac{x_1 - x_2}{h_s}\right) \; du \; dv \\
&= f(q_{\tau_1}(x_1)|X = x_1) \; q_{\tau_1}^{(2)}(x_1) \; h_s^4 \; g(x_1)^2 \int u^2 \; K\,(u) \;\; du \; \int K\,(v) \;\; K \left(v + \frac{x_1 - x_2}{h_s}\right) \; dv \\
&= f(q_{\tau_1}(x_1)|X = x_1) \; q_{\tau_1}^{(2)}(x_1) \; h_s^4 \; g(x_1)^2 \; \mu_2(K) \; K * K \left(\frac{x_2 - x_1}{h_s}\right). \qquad (3.50)
\end{aligned}
$$

Finally, it will be necessary to compute the covariance of the conditional expectations. So, it should be taken into account that

$$
\begin{aligned}
\mathbb{E}[U_{\tau_1,h_s}(x_1) \; U_{\tau_2,h_s}(x_1)|\mathcal{X}] &= \mathbb{E}\left[\frac{1}{nh_s} \; \sum_{i=1}^{n} \psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right) \; K_{i,1} \; \frac{1}{nh_s} \; \sum_{j=1}^{n} \psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right) \; K_{j,1} \middle| \mathcal{X}\right] \\
&= \left(\frac{1}{nh_s}\right)^2 \; \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right) \psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right) \middle| \mathcal{X}\right] \; K_{i,1} \; K_{j,2}
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n}\sum_{j=1}^{n}\Bigg(\tau_1\tau_2 - \tau_1\,\mathbb{E}\left[\mathbb{I}\left(Y_j^{(1,\tau_2)}<0\right)\Big|\mathcal{X}\right]\\
&- \tau_2\,\mathbb{E}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)}<0\right)\Big|\mathcal{X}\right]\\
&+ \mathbb{E}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)}<0\right)\,\mathbb{I}\left(Y_j^{(1,\tau_2)}<0\right)\Big|\mathcal{X}\right]\Bigg)\,K_{i,1}\,K_{j,1}\\
&= \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n}(\min\{\tau_1,\tau_2\}+\tau_1\tau_2)\,K_{i,1}\,K_{i,1} \qquad (3.51)
\end{aligned}
$$

because of the fact that $\mathbb{E}\left[\psi_{\tau_j}\left(Y_i^{(1,\tau_j)}\right)\right] \cong O(h_s^2)$ showed in the first part of Lemma 3.11. As a result,

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[\mathbb{E}\left(U_{\tau_1,h_s}(x_1)U_{\tau_2,h_s}(x_1)|\mathcal{X}\right),\mathbb{E}\left(U_{\tau_3,h_s}(x_2)|\mathcal{X}\right)\right]\\
&= \mathbb{C}\mathrm{ov}\Bigg[\left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n}(\min\{\tau_1,\tau_2\}+\tau_1\tau_2)\,K_{i,1}\,K_{i,1},\\
&\qquad \frac{1}{nh_s}\sum_{l=1}^{n} f(q_{\tau_3}(x_2)+\xi_{28,l}|X=x_2)\,\frac{1}{2}q_{\tau_3}^{(2)}(\xi_{29,l})(X_l-x_2)^2\,K_{l,2}\Bigg]\\
&\le \left(\frac{1}{nh_s}\right)^3 (\min\{\tau_1,\tau_2\}+\tau_1\tau_2)\,h_s^2\,M_1\,M_2\sum_{i=1}^{n}\mathbb{C}\mathrm{ov}\left[K_{i,1}\,K_{i,1},K_{i,2}\right]\\
&= \left(\frac{1}{nh_s}\right)^3 (\min\{\tau_1,\tau_2\}+\tau_1\tau_2)\,h_s^2\,n\,O(h_s) = O\left(\frac{1}{n^2}\right) = o\left(\frac{1}{n^2h_s^2}\right) \qquad (3.52)
\end{aligned}
$$

where M_1 and M_2 represents bounds of the functions f and $q_\tau^{(2)}$. Moreover, the last step comes from the following development:

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}\left[K_{i,1}\,K_{i,1},\;K_{i,2}\right] &= \mathbb{E}[K_{i,1}\,K_{i,1},\;K_{i,2}] - \mathbb{E}[K_{i,1}\,K_{i,1}]\,\mathbb{E}[K_{i,2}]\\
&= \int K\left(\frac{z-x_1}{h_s}\right)\,K\left(\frac{z-x_1}{h_s}\right)\,K\left(\frac{z-x_2}{h_s}\right)\,g(z)\,dz\\
&- \int K\left(\frac{z-x_1}{h_s}\right)\,K\left(\frac{z-x_1}{h_s}\right)\,g(z)\,dz\int K\left(\frac{z-x_2}{h_s}\right)\,g(z)\,dz\\
&= h_s\int K(u)^2\;K\left(u+\frac{x_1-x_2}{h_s}\right)\,g(x_1+uh_s)\,du\\
&- h_s^2\int K(u)^2\;g(x_1+uh_s)\,du\int K\left(\frac{z-x_2}{h_s}\right)\,g(x_2+uh_s)\,du\\
&= O(h_s).
\end{aligned}
$$

So in view of (3.49) and (3.52), it can be concluded that

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}\left(U_{\tau_1,h_s}(x_1)\,U_{\tau_2,h_s}(x_2),U_{\tau_3,h_s}(x_3)\right) &\cong \left(\frac{1}{nh_s}\right)^2\Gamma_3(\tau_1,\tau_2,\tau_3)\,K^2 * K\left(\frac{x_2-x_1}{h_s}\right)\,g(x_1)\\
&+ \frac{h_s}{2n}\,g(x_1)^2\,\mu_2(K)\,K*K\left(\frac{x_2-x_1}{h_s}\right)\\
&\times\left((\min\{\tau_2,\tau_3\}-\tau_2\tau_3)f(q_{\tau_1}(x_1)|X=x_1)\,q_{\tau_1}^{(2)}(x_1)\right.
\end{aligned}
$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

$$+ (\min\{\tau_1,\tau_3\} - \tau_1\tau_3) f(q_{\tau_2}(x_1)|X = x_1)\ q^{(2)}_{\tau_2}(x_1)).$$

Covariance between $\mathbf{U_{\tau_1,h_s}(x_1)U_{\tau_2,h_s}(x_1)}$ and $\mathbf{U_{\tau_3,h_s}(x_2)U_{\tau_4,h_s}(x_2)}$

For simplicity, let us assume the following notation

$$B_7(x_1,x_2) = \mathbb{C}\text{ov}\left[U_{\tau_1,h_s}(x_1)U_{\tau_2,h_s}(x_1), U_{\tau_3,h_s}(x_2)U_{\tau_4,h_s}(x_2)|\mathcal{X}\right].$$

Then, it can be written

$$\begin{aligned}
B_7(x_1,x_2) &= \mathbb{C}\text{ov}\Bigg[\frac{1}{nh_s}\sum_{i=1}^{n}\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)K_{i,1}\ \frac{1}{nh_s}\sum_{j=1}^{n}\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)K_{j,1}\ ,\\
&\qquad \frac{1}{nh_s}\sum_{l=1}^{n}\psi_{\tau_3}\left(Y_l^{(1,\tau_3)}\right)K_{l,2}\ \frac{1}{nh_s}\sum_{k=1}^{n}\psi_{\tau_4}\left(Y_k^{(1,\tau_4)}\right)K_{k,2}\Bigg|\mathcal{X}\Bigg]\\
&= \left(\frac{1}{nh_s}\right)^4\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{k=1}^{n}B_8(x_1,x_2)\ K_{i,1}\ K_{j,1}\ K_{l,2}\ K_{k,2}
\end{aligned}$$

where

$$B_8(x_1,x_2) = \mathbb{C}\text{ov}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)\ ,\ \psi_{\tau_3}\left(Y_l^{(1,\tau_3)}\right)\ \psi_{\tau_4}\left(Y_k^{(1,\tau_4)}\right)\Bigg|\mathcal{X}\right].$$

Now, we are going to focus on computing $B_8(x_1,x_2)$ for different values of indices $\{i,j,l,k\}$. Firstly, if $\text{card}(\{i,j,l,k\}) = 4$ or $\text{card}(\{i,j,l,k\}) = 3$ and $i = j$ or $l = k$ then $B_8(x_1,x_2) = 0$ because of independence of observations $Y_1,\ldots,Y_n$. Secondly, if $\text{card}(\{i,j,l,k\}) = 3$ and $i \neq j$ or $l \neq k$ then $B_8(x_1,x_2) \neq 0$ and we have to compute it. For instance, we are going to check the scenario in which $i \neq \{j,l,k\}$ and $j = l$. In this case, as a consequence of arguments developed in Lemma 3.11 and the first part of this Lemma it follows that

$$\begin{aligned}
B_8(x_1,x_2) &= \mathbb{C}\text{ov}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)\ ,\ \psi_{\tau_3}\left(Y_j^{(1,\tau_3)}\right)\ \psi_{\tau_4}\left(Y_k^{(1,\tau_4)}\right)\Bigg|\mathcal{X}\right]\\
&= \mathbb{E}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\Bigg|\mathcal{X}\right]\ \mathbb{E}\left[\psi_{\tau_4}\left(Y_k^{(1,\tau_4)}\right)\Bigg|\mathcal{X}\right]\mathbb{C}\text{ov}\left[\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right)\ ,\ \psi_{\tau_3}\left(Y_j^{(1,\tau_3)}\right)\Bigg|\mathcal{X}\right]\\
&\cong f(q_{\tau_1}(X_i)+\xi_{2,i}|X = X_i)\ \frac{1}{2}q^{(2)}_{\tau_1}(\xi_{1,i})\ (X_i - x_1)^2 f(q_{\tau_4}(X_k)+\xi_{2,k}|X = X_k)\\
&\times \frac{1}{2}q^{(2)}_{\tau_4}(\xi_{1,k})\ (X_k - x_2)^2(\min\{\tau_2,\tau_3\} - \tau_2\tau_3).
\end{aligned}$$

where $\xi_{1,i}$ represents an element between X_i and x_1, and $\xi_{2,i}$ is an element between $-\frac{1}{2}q^{(2)}_{\tau_1}(\xi_{1,i})(X_i - x_1)^2$ and zero, analogously for $\xi_{1,k}$ and $\xi_{2,k}$.

Analogously, we can analyse the cases in which $i \neq \{j,l,k\}$ and $j = k$, $j \neq \{i,l,k\}$ and $i = l$, and $j \neq \{i,l,k\}$ and $i = k$. For simplicity, we are going to assume the following notation:

$$\begin{aligned}
B_8^{\tau_2,\tau_4,\tau_3,\tau_4}(x_1,x_2) &= \frac{1}{4}\ f(q_{\tau_1}(X_i)+\xi_{2,i}|X = X_i)\ q^{(2)}_{\tau_1}(\xi_{1,i})\ (X_i - x_1)^2\\
&\times f(q_{\tau_4}(X_k)+\xi_{2,k}|X = X_k)\ q^{(2)}_{\tau_4}(\xi_{1,k})\ (X_k - x_2)^2
\end{aligned}$$

$$\times (\min\{\tau_2, \tau_3\} - \tau_2\tau_3).$$

Thirdly, if $\mathrm{card}(\{i, j, l, k\}) = 2$, three different situations should be considered: $i = l \neq j = k$, $i = k \neq j = l$ and $i = j \neq l = k$. If $i = j \neq l = k$ then $B_8(x_1, x_2) = 0$ as a result of the fact that $\mathcal{Y} = \{Y_1, \ldots, Y_n\}$ represents a random sample of the variable Y. In view of the previous developments, the other addends that contribute dominant terms to the calculus of

$$\mathbb{C}\mathrm{ov}\left[U_{\tau_1,h_s}(x_1)U_{\tau_2,h_s}(x_1), U_{\tau_3,h_s}(x_2)U_{\tau_4,h_s}(x_2)|\mathcal{X}\right]$$

will be those associated with $i = l \neq j = k$, $i = k \neq j = l$ and $i = j = l = k$. For this reason, it will be necessary to compute

$$\mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right), \mathbb{I}\left(Y_l^{(1,\tau_3)} < 0\right)\mathbb{I}\left(Y_k^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right]$$

in these scenarios. Using similar arguments to those associated with the auxiliary functions φ_7 and φ_8, we can conclude that

$$\mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right), \mathbb{I}\left(Y_l^{(1,\tau_3)} < 0\right)\mathbb{I}\left(Y_k^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right]$$
$$\cong \begin{cases} \min\{\tau_1,\tau_2,\tau_3,\tau_4\} - \min\{\tau_1,\tau_2\}\min\{\tau_3,\tau_4\} & \text{if } i = j = l = k \\ \min\{\tau_1,\tau_3\}\ \min\{\tau_2,\tau_4\} - \tau_1\ \tau_2\ \tau_3\tau_4 & \text{if } i = l \neq j = k \\ \min\{\tau_1,\tau_4\}\ \min\{\tau_2,\tau_3\} - \tau_1\ \tau_2\ \tau_3\ \tau_4 & \text{if } i = k \neq j = l. \end{cases}$$

Bearing previous equation in mind, it will be possible to compute $B_8(x_1, x_2)$ in the following situations: $i = l \neq j = k$, $i = k \neq j = l$ and $i = j = l = k$. For instance, we will study more insightfully the case in which $i = l \neq j = k$. In this context, it can be written

$$\begin{aligned} B_8(x_1,x_2) &= \mathbb{C}\mathrm{ov}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\psi_{\tau_2}\left(Y_j^{(1,\tau_2)}\right), \psi_{\tau_3}\left(Y_i^{(1,\tau_3)}\right)\psi_{\tau_4}\left(Y_j^{(1,\tau_4)}\right)\middle|\mathcal{X}\right] \\
&= \mathbb{C}\mathrm{ov}\left[\left(\tau_1 - \mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\right)\left(\tau_2 - \mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right)\right),\right. \\
&\qquad \left.\left(\tau_3 - \mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right)\right)\left(\tau_4 - \mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right)\right)\middle|\mathcal{X}\right] \\
&= \mathbb{C}\mathrm{ov}\left[-\tau_2\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right) - \tau_1\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right) + \mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right),\right. \\
&\qquad \left. -\tau_4\mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right) - \tau_3\mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right) + \mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right] \\
&= \tau_2\ \tau_4\ \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right), \mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right)\middle|\mathcal{X}\right] \\
&- \tau_4\ \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right), \mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right)\middle|\mathcal{X}\right] \\
&+ \tau_1\ \tau_3\ \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right), \mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right] \\
&- \tau_3\ \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right), \mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right] \\
&- \tau_2\ \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right), \mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right] \end{aligned}$$

$$
\begin{aligned}
&- \tau_1 \, \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right) \, , \, \mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right] \\
&+ \mathbb{C}\mathrm{ov}\left[\mathbb{I}\left(Y_i^{(1,\tau_1)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_2)} < 0\right) \, , \, \mathbb{I}\left(Y_i^{(1,\tau_3)} < 0\right)\mathbb{I}\left(Y_j^{(1,\tau_4)} < 0\right)\middle|\mathcal{X}\right] \\
&\cong \tau_2\tau_4(\min\{\tau_1,\tau_3\} - \tau_1\tau_3) - \tau_4(\min\{\tau_1,\tau_3\}\tau_2 - \tau_1\tau_2\tau_3) \\
&+ \tau_1 \, \tau_3 \, (\min\{\tau_2,\tau_4\} - \tau_2\tau_4) - \tau_3 \, (\min\{\tau_2,\tau_4\} \, \tau_1 - \tau_1 \, \tau_2 \, \tau_4) \\
&- \tau_2(\min\{\tau_1,\tau_3\}\tau_4 - \tau_1\tau_3\tau_4) - \tau_1(\min\{\tau_2,\tau_4\}\tau_3 - \tau_2\tau_3\tau_4) \\
&+ \min\{\tau_1,\tau_3\}\min\{\tau_2,\tau_4\} - \tau_1\tau_2\tau_3\tau_4 \\
&= \tau_1\tau_2\tau_3\tau_4 + \min\{\tau_1,\tau_3\}\min\{\tau_2,\tau_4\} - \tau_2\tau_4\min\{\tau_1,\tau_3\} - \tau_1\tau_3\min\{\tau_2,\tau_4\}.
\end{aligned}
$$

Analogously, it can be computed the conditional covariance if $i = k \neq j = l$ that will be given by

$$B_8(x_1,x_2) \cong \tau_1\tau_2\tau_3\tau_4 + \min\{\tau_1,\tau_4\}\min\{\tau_2,\tau_3\} - \tau_2\tau_3\min\{\tau_1,\tau_4\} - \tau_1\tau_4\min\{\tau_2,\tau_3\}$$

or in the case $i = j = l = k$ that can be approximated by

$$
\begin{aligned}
B_8(x_1,x_2) &= \mathbb{C}\mathrm{ov}\left[\psi_{\tau_1}\left(Y_i^{(1,\tau_1)}\right)\psi_{\tau_2}\left(Y_i^{(1,\tau_2)}\right) \, , \, \psi_{\tau_3}\left(Y_i^{(1,\tau_3)}\right) \, \psi_{\tau_4}\left(Y_i^{(1,\tau_4)}\right)\middle|\mathcal{X}\right] \\
&\cong \min\{\tau_1,\tau_2,\tau_3,\tau_4\} - \min\{\tau_1,\tau_2\}\min\{\tau_3,\tau_4\} + \tau_2\tau_4(\min\{\tau_2,\tau_3\} - \tau_1\tau_3) \\
&+ \tau_1\tau_4(\min\{\tau_2,\tau_3\} - \tau_2\tau_3) + \tau_2\tau_3(\min\{\tau_2,\tau_4\} - \tau_1\tau_4) + \tau_1\tau_3(\min\{\tau_2,\tau_4\} - \tau_3\tau_4) \\
&- \tau_4 \, (\min\{\tau_1,\tau_2,\tau_3\} - \min\{\tau_1,\tau_2\} \, \tau_3) - \tau_3 \, (\min\{\tau_1,\tau_2,\tau_4\} - \min\{\tau_1,\tau_2\} \, \tau_4) \\
&- \tau_2 \, (\min\{\tau_2,\tau_3,\tau_4\} - \min\{\tau_3,\tau_4\} \, \tau_1) - \tau_1 \, (\min\{\tau_2,\tau_3,\tau_4\} - \min\{\tau_3,\tau_4\} \, \tau_2) \\
&= \Gamma_5(\tau_1,\tau_2,\tau_3,\tau_4).
\end{aligned}
$$

As a result of previous developments, it will be possible to compute

$$
\begin{aligned}
\mathbb{E}[B_7(x_1,x_2)] = \left(\frac{1}{nh_s}\right)^4 \Bigg[&\sum_{\mathrm{card}\{i,j,l,k\}=1} \mathbb{E}[B_8(x_1,x_2) \, K_{i,1} \, K_{j,1} \, K_{l,2} \, K_{k,2}] \\
&+ \sum_{\mathrm{card}\{i,j,l,k\}=2} \mathbb{E}[B_8(x_1,x_2) \, K_{i,1} \, K_{j,1} \, K_{l,2} \, K_{k,2}] \\
&+ \sum_{\mathrm{card}\{i,j,l,k\}=3} \mathbb{E}[B_8(x_1,x_2) \, K_{i,1} \, K_{j,1} \, K_{l,2} \, K_{k,2}] \Bigg] \\
&\cong \left(\frac{1}{nh_s}\right)^4 \Gamma_5(\tau_1,\tau_2,\tau_3,\tau_4) \sum_{i=1}^{n} \mathbb{E}[K_{i,1}^2 \, K_{i,2}^2] \\
&+ \left(\frac{1}{nh_s}\right)^4 \Gamma_6(\tau_1,\tau_2,\tau_3,\tau_4) \sum_{i\neq j} \mathbb{E}[K_{i,1} \, K_{j,1} \, K_{i,2} \, K_{j,2}] \\
&+ \left(\frac{1}{nh_s}\right)^4 \sum_{i,j,l} \mathbb{E}\left[B_8^{\tau_1,\tau_3,\tau_2,\tau_4}(x_1,x_2) \, K_{i,1} \, K_{j,1} \, K_{l,2} \, K_{j,2}\right] \\
&+ \left(\frac{1}{nh_s}\right)^4 \sum_{i,j,k} \mathbb{E}\left[B_8^{\tau_1,\tau_4,\tau_2,\tau_3}(x_1,x_2) \, K_{i,1} \, K_{j,1} \, K_{j,2} \, K_{k,2}\right]
\end{aligned}
$$

$$+\left(\frac{1}{nh_s}\right)^4 \sum_{i,j,l} \mathbb{E}\left[B_8^{\tau_2,\tau_3,\tau_1,\tau_4}(x_1,x_2)\; K_{i,1}\; K_{j,1}\; K_{l,2}\; K_{i,2}\right]$$
$$+\left(\frac{1}{nh_s}\right)^4 \sum_{i,j,k} \mathbb{E}\left[B_8^{\tau_2,\tau_4,\tau_1,\tau_3}(x_1,x_2)\; K_{i,1}\; K_{j,1}\; K_{i,2}\; K_{k,2}\right] \qquad (3.53)$$

where

$$\begin{aligned}\Gamma_6(\tau_1,\tau_2,\tau_3,\tau_4) &= 2\tau_1\tau_2\tau_3\tau_4 - \tau_2\tau_4\min\{\tau_2,\tau_3\} - \tau_1\tau_4\min\{\tau_2,\tau_3\}\\ &\quad - \tau_2\tau_3\min\{\tau_2,\tau_4\} - \tau_1\tau_3\min\{\tau_2,\tau_4\}\\ &\quad + \min\{\tau_1,\tau_3\}\;\min\{\tau_2,\tau_4\} + \min\{\tau_1,\tau_4\}\;\min\{\tau_2,\tau_3\}.\end{aligned}$$

Then, we are going to compute each of the expectations involved in equation (3.53). On the one hand,

$$\begin{aligned}\mathbb{E}\left[K_{i,1}^2\; K_{i,2}^2\right] &= \int K\left(\frac{X_i-x_1}{h_s}\right)^2 K\left(\frac{X_i-x_2}{h_s}\right)^2 g(X_i)\, dX_i\\ &= \int K(u)^2\; K\left(u+\frac{x_1-x_2}{h_s}\right)^2 g(x_1+uh_s)\, h_s\, du\\ &= K^2 * K^2\left(\frac{x_2-x_1}{h_s}\right) g(x_1)\, h_s + O(h_s^2)\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left[K_{i,1}\; K_{j,1}\; K_{i,2}\; K_{j,2}\right] &= \int K\left(\frac{X_i-x_1}{h_s}\right) K\left(\frac{X_j-x_1}{h_s}\right) K\left(\frac{X_i-x_2}{h_s}\right) K\left(\frac{X_j-x_2}{h_s}\right)\\ &\quad\times g(X_i,X_j)\, dX_i\, dX_j\\ &= \int\int K(u)\; K\left(u+\frac{x_1-x_2}{h_s}\right) K(v)\; K\left(v+\frac{x_1-x_2}{h_s}\right)\\ &\quad\times g(x_1+uh_s)\, g(x_1+vh_s)\, h_s^2\, du\, dv\\ &= (K*K)^2\left(\frac{x_2-x_1}{h_s}\right) g(x_1)^2\, h_s^2 + O(h_s^3).\end{aligned}$$

Finally, it will be necessary to compute $\mathbb{E}\left[B_8^{\tau_1,\tau_3,\tau_2,\tau_4}(x_1,x_2)\; K_{i,1}\; K_{j,1}\; K_{l,2}\; K_{j,2}\right]$ that is a bit more complicated. In this case,

$$\begin{aligned}\mathbb{E}\left[B_8^{\tau_1,\tau_3,\tau_2,\tau_4}(x_1,x_2)\; K_{i,1}\; K_{j,1}\; K_{l,2}\; K_{j,2}\right] &= \frac{1}{4}(\min\{\tau_2,\tau_4\}-\tau_2\tau_4)\\ &\quad\times \mathbb{E}\Big[f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\; q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x_1)^2\; K_{i,1}\Big]\\ &\quad\times \mathbb{E}\Big[\; f(q_{\tau_3}(X_l)+\xi_{2,l}|X=X_l)\; q_{\tau_3}^{(2)}(\xi_{1,l})\;(X_l-x_2)^2\; K_{l,2}\Big]\\ &\quad\times \mathbb{E}\Big[\; K_{j,1}\; K_{j,2}\Big]\end{aligned}$$

where

$$\mathbb{E}\Big[f(q_{\tau_1}(X_i)+\xi_{2,i}|X=X_i)\; q_{\tau_1}^{(2)}(\xi_{1,i})(X_i-x_1)^2\; K_{i,1}\Big] \cong f(q_{\tau_1}(x_1)|X=x_1)\; q_{\tau_1}^{(2)}(x_1)$$

$$\times g(x_1)\, h_s^3\, \mu_2(K)$$

and

$$\mathbb{E}\Big[\, K_{j,1}\, K_{j,2} \Big] = K * K\left(\frac{x_2 - x_1}{h_s}\right)\, g(x_1)\, h_s + O(h_s^2)$$

following the arguments applied in equations (3.50) and (3.47). So, we can conclude that

$$\mathbb{E}\left[B_8^{\tau_1,\tau_3,\tau_2,\tau_4}(x_1,x_2)\, K_{i,1}\, K_{j,1}\, K_{l,2}\, K_{j,2}\right] \cong \frac{1}{4}\, K * K\left(\frac{x_2 - x_1}{h_s}\right)\, g(x_1)^2\, g(x_2)\, h_s^7\, \mu_2(K)^2 \\ \times \Gamma_7(\tau_1, x_1)\, \Gamma_7(\tau_3, x_2)\, \Gamma_8(\tau_2, \tau_4)$$

where

$$\Gamma_7(\tau_i, x_j) = f(q_{\tau_i}(x_j)|X = x_j)\, q_{\tau_i}^{(2)}(x_j) \\ \Gamma_8(\tau_i, \tau_j) = \min\{\tau_i, \tau_j\} - \tau_i\tau_j.$$

In order to finish with the calculus associated with the expectation of $B_7(x_1, x_2)$ it can be written that

$$\begin{aligned}
\mathbb{E}[B_7(x_1,x_2)] \cong& \left(\frac{1}{nh_s}\right)^4 \Gamma_5(\tau_1,\tau_2,\tau_3,\tau_4)\, n\, K^2 * K^2\left(\frac{x_2 - x_1}{h_s}\right)\, g(x_1)\, h_s \\
&+ \left(\frac{1}{nh_s}\right)^4 \Gamma_6(\tau_1,\tau_2,\tau_3,\tau_4)\, n(n-1)\, (K * K)^2\left(\frac{x_2 - x_1}{h_s}\right)\, g(x_1)^2\, h_s^2 \\
&+ \left(\frac{1}{nh_s}\right)^4\, n(n-1)(n-2)\, \frac{1}{4}\, K * K\left(\frac{x_2 - x_1}{h_s}\right)\, \mu_2(K)\, g(x_1)^2\, g(x_2)\, h_s^7 \\
&\times \Big(\Gamma_7(\tau_1,x_1)\, \Gamma_7(\tau_3,x_2)\, \Gamma_8(\tau_2,\tau_4) + \Gamma_7(\tau_1,x_1)\, \Gamma_7(\tau_4,x_2)\, \Gamma_8(\tau_2,\tau_3) \\
&+ \Gamma_7(\tau_2,x_1)\, \Gamma_7(\tau_3,x_2)\, \Gamma_8(\tau_1,\tau_4) + \Gamma_7(\tau_2,x_1)\, \Gamma_7(\tau_4,x_2)\, \Gamma_8(\tau_1,\tau_3)\Big) \\
&\cong \left(\frac{1}{nh_s}\right)^2 \Gamma_6(\tau_1,\tau_2,\tau_3,\tau_4)\, (K * K)^2\left(\frac{x_2 - x_1}{h_s}\right)\, g(x_1)^2 \\
&+ \frac{h_s^3}{4n}\, K * K\left(\frac{x_2 - x_1}{h_s}\right)\, \mu_2(K)^2\, g(x_1)^2\, g(x_2) \\
&\times \Big(\Gamma_7(\tau_1,x_1)\, \Gamma_7(\tau_3,x_2)\, \Gamma_8(\tau_2,\tau_4) + \Gamma_7(\tau_1,x_1)\, \Gamma_7(\tau_4,x_2)\, \Gamma_8(\tau_2,\tau_3) \\
&+ \Gamma_7(\tau_2,x_1)\, \Gamma_7(\tau_3,x_2)\, \Gamma_8(\tau_1,\tau_4) + \Gamma_7(\tau_2,x_1)\, \Gamma_7(\tau_4,x_2)\, \Gamma_8(\tau_1,\tau_3)\Big). \qquad (3.54)
\end{aligned}$$

Finally, it will be necessary to compute the covariance between the conditional expectations. So, it should be taken into account that

$$\mathbb{E}[U_{\tau_1,h_s}(x_1)\, U_{\tau_2,h_s}(x_1)|\mathcal{X}] \cong \left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} (\min\{\tau_1,\tau_2\} - \tau_1\tau_2)\, K_{i,1}\, K_{i,1}$$

because of (3.51), and as a consequence

$$\mathrm{Cov}\left[\mathbb{E}\left(U_{\tau_1,h_s}(x_1)\, U_{\tau_2,h_s}(x_1)|\mathcal{X}\right), \mathbb{E}\left(U_{\tau_3,h_s}(x_2)\, U_{\tau_4,h_s}(x_2)|\mathcal{X}\right)\right]$$

$$
\begin{aligned}
&= \mathbb{C}\mathrm{ov}\left[\left(\frac{1}{nh_s}\right)^2 \sum_{i=1}^{n} (\min\{\tau_1,\tau_2\} - \tau_1\tau_2)\; K_{i,1}\; K_{i,1}, \right.\\
&\qquad \left. \left(\frac{1}{nh_s}\right)^2 \sum_{l=1}^{n} (\min\{\tau_3,\tau_4\} - \tau_3\tau_4)\; K_{l,2}\; K_{l,2} \right]\\
&\leq \left(\frac{1}{nh_s}\right)^4 (\min\{\tau_1,\tau_2\} - \tau_1\tau_2)\; (\min\{\tau_3,\tau_4\} - \tau_3\tau_4)\; \mathbb{C}\mathrm{ov}\left[K_{i,1}\; K_{i,1}, K_{i,2}\; K_{i,2}\right]\\
&= \left(\frac{1}{nh_s}\right)^4 (\min\{\tau_1,\tau_2\} - \tau_1\tau_2)\;\; (\min\{\tau_3,\tau_4\} - \tau_3\tau_4)\; n\; O(h_s)\\
&= O\left(\frac{1}{n^3h_s^3}\right) = o\left(\frac{1}{n^2h_s^2}\right) \qquad (3.55)
\end{aligned}
$$

where the last step comes from the following development:

$$
\begin{aligned}
\mathbb{C}\mathrm{ov}\left[K_{i,1}\; K_{i,1}\;,\; K_{i,2}\; K_{i,2}\right] &= \mathbb{E}[K_{i,1}\; K_{i,1}\; K_{i,2}\; K_{i,2}] - \mathbb{E}[K_{i,1}\; K_{i,1}]\; \mathbb{E}[K_{i,2}\; K_{i,2}]\\
&= \int K\left(\frac{z-x_1}{h_s}\right) K\left(\frac{z-x_1}{h_s}\right) K\left(\frac{z-x_2}{h_s}\right) K\left(\frac{z-x_2}{h_s}\right)\; g(z)\; dz\\
&- \int K\left(\frac{z-x_1}{h_s}\right)\; K\left(\frac{z-x_1}{h_s}\right)\; g(z)\; dz\\
&\times \int K\left(\frac{z-x_2}{h_s}\right) K\left(\frac{z-x_2}{h_s}\right)\; g(z)\; dz\\
&= h_s \int K\left(u\right)^2 K\left(u + \frac{x1 - x_2}{h_s}\right)^2\; g(x_1 + uh_s)\; du\\
&- h_s^2 \int K\left(u\right)^2\;\; g(x_1 + uh_s)\; du \int K\left(w\right)^2\;\; g(x_2 + wh_s)\; dw\\
&= O(h_s)
\end{aligned}
$$

So in view of (3.54) and (3.55), it can be concluded that

$$
\begin{aligned}
&\mathbb{C}\mathrm{ov}\left[U_{\tau_1,h_s}(x_1)U_{\tau_2,h_s}(x_1), U_{\tau_3,h_s}(x_2)U_{\tau_4,h_s}(x_2)\right] \cong\\
&\quad\cong \left(\frac{1}{nh_s}\right)^2 \Gamma_6(\tau_1,\tau_2,\tau_3,\tau_4)\; (K*K)^2\left(\frac{x_2-x_1}{h_s}\right)\; g(x_1)^2\\
&\quad + \frac{h_s^3}{4n}\; K*K\left(\frac{x_2-x_1}{h_s}\right)\; \mu_2(K)^2\; g(x_1)^2\; g(x_2)\\
&\quad \times \Big(\Gamma_7(\tau_1,x_1)\; \Gamma_7(\tau_3,x_2)\; \Gamma_8(\tau_2,\tau_4) + \Gamma_7(\tau_1,x_1)\; \Gamma_7(\tau_4,x_2)\; \Gamma_8(\tau_2,\tau_3)\\
&\quad + \Gamma_7(\tau_2,x_1)\; \Gamma_7(\tau_3,x_2)\; \Gamma_8(\tau_1,\tau_4) + \Gamma_7(\tau_2,x_1)\; \Gamma_7(\tau_4,x_2)\; \Gamma_8(\tau_1,\tau_3)\Big).
\end{aligned}
$$

□

Now, we are able to compute all the components of the variance decomposition given in (3.44). We are going to establish the variance of $\int A(x)B(x)\; dx$ in the following proposition:

Proposition 3.18. *Under assumptions S1-S5 it is verified that*

$$\mathbb{V}ar\left[\int A(x)B(x)\,dx\right] \cong \frac{1}{2nd_s}\int \frac{s_\tau(x)^4}{g(x)}dx$$

Proof. Thereupon, applying Fubini's theorem,

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\int A(x)B(x)\,dx\right] &= \mathbb{E}\left[\left(\int A(x)B(x)\,dx\right)^2\right] - \left(\mathbb{E}\left[\int A(x)B(x)\,dx\right]\right)^2 \\
&= \mathbb{E}\left[\int\int A(x_1)B(x_1)A(x_2)B(x_2)dx_1dx_2\right] \\
&- \int\int \mathbb{E}\left[A(x_1)B(x_1)\right]\,\mathbb{E}\left[A(x_2)B(x_2)\right]\,dx_1dx_2 \\
&= \int\int\Big(\mathbb{E}\left[A(x_1)B(x_1)A(x_2)B(x_2)\right] \\
&- \mathbb{E}\left[A(x_1)B(x_1)\right]\,\mathbb{E}\left[A(x_2)B(x_2)\right]\Big)\,dx_1dx_2 \\
&= \int\int B_9(x_1,x_2)\,dx_1dx_2
\end{aligned}$$

where $B_9(x_1,x_2) = \mathbb{C}\text{ov}\left[A(x_1)B(x_1), A(x_2)B(x_2)\right]$. Moreover, the following development comes from Lemma 3.12 and the definition of $B(x)$:

$$\begin{aligned}
B_9(x_1,x_2) &= \mathbb{C}\text{ov}\left[A(x_1)B(x_1), A(x_2)B(x_2)\right] = \\
&\cong \mathbb{C}\text{ov}\Bigg[\frac{s_\tau(x_1)}{2d_s\,g(x_1)}\left(s_{\tau+d_s}(x_1)\,U_{\tau+d_s,h_s}(x_1) - s_{\tau-d_s}(x_1)\,U_{\tau-d_s,h_s}(x_1)\right)\,, \\
&\quad \frac{s_\tau(x_2)}{2d_s\,g(x_2)}\left(s_{\tau+d_s}(x_2)\,U_{\tau+d_s,h_s}(x_2) - s_{\tau-d_s}(x_2)\,U_{\tau-d_s,h_s}(x_2)\right)\Bigg] \\
&= \frac{s_\tau(x_1)}{2d_s\,g(x_1)}\,\frac{s_\tau(x_2)}{2d_s\,g(x_2)}\Big(s_{\tau+d_s}(x_1)\,s_{\tau+d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)\,,\,U_{\tau+d_s,h_s}(x_2)\right] \\
&- s_{\tau-d_s}(x_1)\,s_{\tau+d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau-d_s,h_s}(x_1)\,,\,U_{\tau+d_s,h_s}(x_2)\right] \\
&- s_{\tau+d_s}(x_1)\,s_{\tau-d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)\,,\,U_{\tau-d_s,h_s}(x_2)\right] \\
&+ s_{\tau-d_s}(x_1)\,s_{\tau-d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau-d_s,h_s}(x_1)\,,\,U_{\tau-d_s,h_s}(x_2)\right]\Big).
\end{aligned}$$

Then, the problem has been reduced to computing the covariance between $U_{\tau_1,h_s}(x_1)$ and $U_{\tau_2,h_s}(x_2)$ for certain τ_1, τ_2, x_1 and x_2, and this was obtained in Lemma 3.17. So, it follows that

$$\begin{aligned}
B_9(x_1,x_2) &\cong \frac{s_\tau(x_1)}{2d_s}\,\frac{s_\tau(x_2)}{2d_s\,g(x_2)}\,\frac{1}{nh_s} \\
&\times \Bigg[s_{\tau+d_s}(x_1)\,s_{\tau+d_s}(x_2)\,\left(\tau + d_s - (\tau+d_s)^2\right)\,K*K\left(\frac{x_2-x_1}{h_s}\right) \\
&- s_{\tau-d_s}(x_1)\,s_{\tau+d_s}(x_2)\,\left(\tau - d_s - (\tau+d_s)(\tau-d_s)\right)\,K*K\left(\frac{x_2-x_1}{h_s}\right) \\
&- s_{\tau+d_s}(x_1)\,s_{\tau-d_s}(x_2)\,\left(\tau - d_s - (\tau+d_s)(\tau-d_s)\right)\,K*K\left(\frac{x_2-x_1}{h_s}\right)
\end{aligned}$$

$$+ s_{\tau-d_s}(x_1)\, s_{\tau-d_s}(x_2)\, \left(\tau - d_s - (\tau-d_s)^2\right)\, K*K\left(\frac{x_2-x_1}{h_s}\right)\Bigg]$$

and as a consequence

$$\begin{aligned}\mathbb{V}\text{ar}\left[\int A(x)B(x)\,dx\right] &= \int\int B_9(x_1,x_2)\,dx_1dx_2\\ &\cong \frac{1}{4d_s^2}\,\frac{1}{nh_s}\int\frac{s_\tau(x)^4}{g(x)}dx\; h_s\int K*K(u)du\,(2d_s-4d_s^2)\\ &\cong \frac{1}{2nd_s}\int\frac{s_\tau(x)^4}{g(x)}dx\end{aligned}$$

where $\int K*K(u)du = \int K(u)du \int K(u)du = 1$. □

Now, we are going to move to the calculus of the covariance between $\int B(x)^2dx$ and $\int A(x)B(x)dx$. The result is available in the following proposition:

Proposition 3.19. *Under conditions S1-S5, it is verified that*

$$\begin{aligned}\mathbb{C}ov\left[\int B(x)^2\,dx, \int A(x)B(x)\,dx\right] \cong& -\frac{1}{4n^2d_s^2h_s}R(K)\int\frac{s_\tau(x)^4}{g(x)^2}dx\\ &+\frac{h_s^2}{2nd_s}\int\frac{s_\tau(x)^4}{g(x)}\frac{\partial f(q_\tau(x)|X=x)\, q_\tau^{(2)}(x)}{\partial\tau}\,dx\end{aligned}$$

Proof. Applying Fubini's theorem, as in Proposition 3.18, it follows that

$$\begin{aligned}\mathbb{C}\text{ov}\left[\int B(x)^2\,dx, \int A(x)B(x)\,dx\right] &= \int\int \mathbb{C}\text{ov}\left[B(x_1)^2, A(x_2)B(x_2)\right]\,dx_1\,dx_2\\ &= \int\int A(x_2)\,B_{10}(x_1,x_2)\,dx_1\,dx_2\end{aligned}$$

where

$$\begin{aligned}B_{10}(x_1,x_2) &= \mathbb{C}\text{ov}\left[B(x_1)^2, B(x_2)\right]\\ &= \mathbb{C}\text{ov}\Bigg[\left(\frac{1}{2d_sg(x_1)}\right)^2\left(s_{\tau+d_s}(x_1)\,U_{\tau+d_s,h_s}(x_1) - s_{\tau-d_s}(x_1)\,U_{\tau-d_s,h_s}(x_1)\right)^2,\\ &\qquad \frac{1}{2d_sg(x_2)}\left(s_{\tau+d_s}(x_2)\,U_{\tau+d_s,h_s}(x_2) - s_{\tau-d_s}(x_2)\,U_{\tau-d_s,h_s}(x_2)\right)\Bigg]\\ &= \left(\frac{1}{2d_sg(x_1)}\right)^2\frac{1}{2d_sg(x_2)}\Big(s_{\tau+d_s}(x_1)^2\, s_{\tau+d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)^2,\, U_{\tau+d_s,h_s}(x_2)\right]\\ &+ s_{\tau-d_s}(x_1)^2\, s_{\tau+d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau-d_s}(x_1)^2,\, U_{\tau+d_s,h_s}(x_2)\right]\\ &- 2\, s_{\tau+d_s}(x_1)\, s_{\tau-d_s}(x_1)\, s_{\tau+d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)\,U_{\tau-d_s,h_s}(x_1), U_{\tau+d_s,h_s}(x_2)\right]\\ &- s_{\tau+d_s}(x_1)^2\, s_{\tau-d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)^2,\, U_{\tau-d_s,h_s}(x_2)\right]\\ &- s_{\tau-d_s}(x_1)^2\, s_{\tau-d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau-d_s,h_s}(x_1)^2,\, U_{\tau-d_s,h_s}(x_2)\right]\\ &+ 2\, s_{\tau+d_s}(x_1)\, s_{\tau-d_s}(x_1)\, s_{\tau-d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)\,U_{\tau-d_s,h_s}(x_1),\, U_{\tau-d_s,h_s}(x_2)\right]\Big).\end{aligned}$$

Then, each of covariance involved in previous expression can be computed as consequence of results gathered in Lemma 3.17. After some direct calculations, the following approximation is obtained:

$$\begin{aligned}
B_{10}(x_1, x_2) &\cong \left(\frac{1}{2d_s g(x_1)}\right)^2 \frac{1}{2d_s g(x_2)} \left(\frac{1}{nh_s}\right)^2 g(x_1)\ s_\tau(x_1)^2\ s_\tau(x_2)\ (-2d_s)\ K^2 * K\left(\frac{x_2 - x_1}{h_s}\right) \\
&+ \left(\frac{1}{2d_s g(x_1)}\right)^2 \frac{1}{2d_s g(x_2)} \frac{h_s}{2n} g(x_1)^2\ s_\tau(x_1)^2\ s_\tau(x_2)\ \mu_2(K)\ K * K\left(\frac{x_2 - x_1}{h_s}\right)\ 4d_s \\
&\times \left(f(q_{\tau+d_s}(x_1)|X = x_1)\ q^{(2)}_{\tau+d_s}(x_1) - f(q_{\tau-d_s}(x_1)|X = x_1)\ q^{(2)}_{\tau-d_s}(x_1)\right) \\
&\cong -\left(\frac{1}{2d_s}\right)^2 \frac{1}{g(x_1)g(x_2)} \left(\frac{1}{nh_s}\right)^2\ s_\tau(x_1)^2\ s_\tau(x_2)\ K^2 * K\left(\frac{x_2 - x_1}{h_s}\right) \\
&+ \frac{1}{2d_s} \frac{1}{g(x_2)} \frac{h_s}{n}\ s_\tau(x_1)^2\ s_\tau(x_2)\ \frac{\partial\left[f(q_\tau(x_1)|X = x_1)\ q^{(2)}_\tau(x_1)\right]}{\partial\tau}\ \mu_2(K) \\
&K * K\left(\frac{x_2 - x_1}{h_s}\right)
\end{aligned}$$

and as a consequence

$$\begin{aligned}
&\int\int A(x_2)\ \mathbb{C}\mathrm{ov}\left[B(x_1)^2, B(x_2)\right]\ dx_1\ dx_2 = \int\int A(x_2)\ B_{10}(x_1, x_2)\ dx_1\ dx_2 \\
&\cong -\left(\frac{1}{2d_s}\right)^2 \left(\frac{1}{nh}\right)^2 \int \frac{s_\tau(x)^4}{g(x)^2}\ dx\ h_s \int K^2 * K\,(u)\ du \\
&+ \frac{h_s}{2nd_s} \int \frac{s_\tau(x)^4}{g(x)} \frac{\partial f(q_\tau(x)|X = x)\ q^{(2)}_\tau(x)}{\partial\tau}\ dx\ h_s \int K * K\,(u)\ du \\
&\cong -\frac{1}{4n^2 d_s^2 h_s}\ R(K) \int \frac{s_\tau(x)^4}{g(x)^2} dx + \frac{h_s^2}{2nd_s}\ \mu_2(K) \int \frac{s_\tau(x)^4}{g(x)} \frac{\partial\left[f(q_\tau(x)|X = x)\ q^{(2)}_\tau(x)\right]}{\partial\tau}\ dx
\end{aligned}$$

where it should be noticed that

$$\begin{aligned}
\int K^2 * K\,(u)\ du &= \int K^2\,(u)\ du \int K\,(u)\ du = R(K) \\
\int K * K\,(u)\ du &= \int K\,(u)\ du \int K\,(u)\ du = 1.
\end{aligned}$$

□

Finally, we are going to compute the variance of $\int B(x)^2 dx$. This result is gathered in the following proposition:

Proposition 3.20. *Under conditions S1-S5, it is verified that*

$$\begin{aligned}
\mathbb{V}ar\left[\int B(x)^2\ dx\right] &\cong \frac{1}{2n^2 d_s^2 h_s} \int (K * K)^2\,(u)\ du \int \frac{s_\tau(x)^4}{g(x)^2} dx \\
&+ \frac{h_s^4}{4nd_s} \int \frac{s_\tau(x)^4}{g(x)} \left[\frac{\partial f(q_\tau(x)|X = x)\ q^{(2)}_\tau(x)}{\partial\tau}\right]^2 dx.
\end{aligned}$$

Proof. In this case, as a result of Fubini's theorem, it follows that

$$\mathbb{V}\text{ar}\left[\int B(x)^2\,dx\right] = \int\int B_{11}(x_1,x_2)\,dx_1\,dx_2$$

where

$$\begin{aligned}
B_{11}(x_1,x_2) &= \mathbb{C}\text{ov}\left[B(x_1)^2, B(x_2)^2\right]\\
&= \mathbb{C}\text{ov}\left[\left(\frac{1}{2d_s g(x_1)}\right)^2 \left(s_{\tau+d_s}(x_1)\,U_{\tau+d_s,h_s}(x_1) - s_{\tau-d_s}(x_1)\,U_{\tau-d_s,h_s}(x_1)\right)^2,\right.\\
&\qquad \left.\left(\frac{1}{2dg(x_2)}\right)^2 \left(s_{\tau+d_s}(x_2)\,U_{\tau+d_s,h_s}(x_2) - s_{\tau-d_s}(x_2)\,U_{\tau-d_s,h_s}(x_2)\right)^2\right]\\
&= \left(\frac{1}{4d_s^2 g(x_1)g(x_2)}\right)^2 \left(s_{\tau+d_s}(x_1)^2\, s_{\tau+d_s}(x_2)^2\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)^2,\ U_{\tau+d_s,h_s}(x_2)^2\right]\right.\\
&+ s_{\tau-d_s}(x_1)^2\, s_{\tau+d_s}(x_2)^2\,\mathbb{C}\text{ov}\left[U_{\tau-d_s,h_s}(x_1)^2,\ U_{\tau+d_s,h_s}(x_2)^2\right]\\
&- 2\, s_{\tau+d_s}(x_1)\, s_{\tau-d_s}(x_1)\, s_{\tau+d_s}(x_2)^2\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)\,U_{\tau-d_s,h_s}(x_1), U_{\tau+d_s,h_s}(x_2)^2\right]\\
&+ s_{\tau+d_s}(x_1)^2\, s_{\tau-d_s}(x_2)^2\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)^2,\ U_{\tau-d_s,h_s}(x_2)^2\right]\\
&+ s_{\tau-d_s}(x_1)^2\, s_{\tau-d_s}(x_2)^2\,\mathbb{C}\text{ov}\left[U_{\tau-d_s,h_s}(x_1)^2,\ U_{\tau-d_s,h_s}(x_2)^2\right]\\
&- 2\, s_{\tau-d_s}(x_1)\, s_{\tau+d_s}(x_1)\, s_{\tau-d_s}(x_2)^2\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)\,U_{\tau-d_s,h_s}(x_1), U_{\tau-d_s,h_s}(x_2)^2\right]\\
&- 2\, s_{\tau+d_s}(x_1)^2\, s_{\tau+d_s}(x_2)\, s_{\tau-d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)^2, U_{\tau+d_s,h_s}(x_2)\,U_{\tau-d_s,h_s}(x_2)\right]\\
&- 2\, s_{\tau-d_s}(x_1)^2\, s_{\tau+d_s}(x_2)\, s_{\tau-d_s}(x_2)\,\mathbb{C}\text{ov}\left[U_{\tau-d_s,h_s}(x_1)^2, U_{\tau+d_s,h_s}(x_2)\,U_{\tau-d_s,h_s}(x_2)\right]\\
&+ 4\, s_{\tau+d_s}(x_1) s_{\tau-d_s}(x_1)\, s_{\tau+d_s}(x_2)\, s_{\tau-d_s}(x_2)\\
&\times \mathbb{C}\text{ov}\left[U_{\tau+d_s,h_s}(x_1)\,U_{\tau-d_s,h_s}(x_1),\ U_{\tau+d_s,h_s}(x_2)\,U_{\tau-d_s,h_s}(x_2)\right].
\end{aligned}$$

Then, in view of Lemma 3.17 we can approximate each of the covariances involve in previous expression. Then, after some direct calculations the following expression is obtained:

$$\begin{aligned}
B_{11}(x_1,x_2) &\cong \left(\frac{1}{4d_s^2 g(x_1)g(x_2)}\right)^2 \left(\frac{1}{nh_s}\right)^2 (K*K)^2\left(\frac{x_2-x_1}{h_s}\right) g(x_1)^2\, 8d_s^2\, s_\tau(x_1)^2\, s_\tau(x_2)^2\\
&+ \left(\frac{1}{4d_s^2 g(x_1)g(x_2)}\right)^2 \frac{h_s^3}{4n}\, K*K\left(\frac{x_2-x_1}{h_s}\right) \mu_2(K)^2\, g(x_1)^2\, g(x_2)\, s_\tau(x_1)^2\, s_\tau(x_2)^2\\
&\times 4d_s\Big(f(q_{\tau+d_s}(x_1)|X=x_1)\, q^{(2)}_{\tau+d_s}(x_1)\, f(q_{\tau-d_s}(x_2)|X=x_2)\, q^{(2)}_{\tau-d_s}(x_2)\\
&+ f(q_{\tau+d_s}(x_2)|X=x_2)\, q^{(2)}_{\tau+d_s}(x_2)\, f(q_{\tau-d_s}(x_1)|X=x_1)\, q^{(2)}_{\tau-d_s}(x_2)^2\\
&- f(q_{\tau+d_s}(x_1)|X=x_1)\, q^{(2)}_{\tau+d_s}(x_1)\, f(q_{\tau-d_s}(x_1)|X=x_1)\, q^{(2)}_{\tau-d_s}(x_1)\Big)\\
&- f(q_{\tau+d_s}(x_2)|X=x_2)\, q^{(2)}_{\tau+d_s}(x_2)\, f(q_{\tau-d_s}(x_2)|X=x_2)\, q^{(2)}_{\tau-d_s}(x_2)\Big)\\
&= \left(\frac{1}{4d_s^2 g(x_1)g(x_2)}\right)^2 \left(\frac{1}{nh_s}\right)^2 (K*K)^2\left(\frac{x_2-x_1}{h_s}\right) g(x_1)^2\, 8d_s^2\, s_\tau(x_1)^2\, s_\tau(x_2)^2\\
&+ \left(\frac{1}{4d_s^2 g(x_1)g(x_2)}\right)^2 \frac{h_s^3}{4n}\, K*K\left(\frac{x_2-x_1}{h_s}\right) \mu_2(K)^2\, g(x_1)^2\, g(x_2)\, s_\tau(x_1)^2\, s_\tau(x_2)^2\\
&\times 16d_s^3\, \frac{\partial[f(q_\tau(x_1)|X=x_1)\, q_\tau^{(2)}(x_1)]}{\partial\tau}\, \frac{\partial[f(q_\tau(x_2)|X=x_2)\, q_\tau^{(2)}(x_2)]}{\partial\tau}
\end{aligned}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

and as a result

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\int B(x)^2\,dx\right] &= \int\int B_{11}(x_1,x_2)\,dx_1\,dx_2\\
&\cong \left(\frac{1}{2d_s}\right)^4\left(\frac{1}{nh}\right)^2\int\frac{s_\tau(x)^4}{g(x)^2}\,dx\;\; h_s\int (K*K)^2(u)\;du\;8d_s^2\\
&+\left(\frac{1}{2d_s}\right)^4\frac{h_s^3}{4n}\,\mu_2(K)^2\,32d_s^3\,h_s\int K*K\,(u)\;du\\
&\times\int\frac{s_\tau(x)^4}{g(x)^2}\left[\frac{\partial f(q_\tau(x)|X=x)\;q_\tau^{(2)}(x)}{\partial\tau}\right]^2dx\\
&\cong \frac{1}{2n^2d_s^2h_s}R(K*K)\;\int\frac{s_\tau(x)^4}{g(x)^2}dx\\
&+\frac{h_s^4}{4nd_s}\int\frac{s_\tau(x)^4}{g(x)}\;\left[\frac{\partial f(q_\tau(x)|X=x)\;q_\tau^{(2)}(x)}{\partial\tau}\right]^2dx.
\end{aligned}$$

□

Finally, the asymptotic mean squared error of the integrated squared sparsity estimator that can be established. This result is gathered in the following theorem:

Theorem 3.21. *Under assumptions S1-S5, the mean squared error of the integrated squared sparsity estimator, given by (3.35), can be approximated by*

$$\begin{aligned}
MSE\left[\int\widehat{s}_{\tau,d_s,h_s}(x)^2\,dx\right] &\cong \left(\frac{1}{3}\,d_s^2\int s_\tau(x)\;s_\tau^{(2,\tau)}(x)\;dx+\mu_2(K)\;\;h_s^2\int s_\tau(x)\;\frac{\partial q_\tau^{(2)}(x)}{\partial\tau}\bigg|_\tau\;dx\right.\\
&\left.+\frac{1}{2nd_sh_s}\;R(K)\int\frac{1}{g(x)}\;s_\tau(x)^2\;dx\right)^2+\frac{2}{nd_s}\int\frac{s_\tau(x)^4}{g(x)}dx\\
&+\frac{1}{n^2d_s^2h_s}\int\frac{s_\tau(x)^4}{g(x)^2}dx\left(\frac{1}{2}R(K*k)-R(K)\right)
\end{aligned}$$

Proof. On the one hand, in view of Propositions 3.18, 3.19 and 3.20, and equation (3.44), the variance of the integrated sparsity estimator can de derived. That is

$$\begin{aligned}
\mathbb{V}\text{ar}\left[\int\widehat{s}_{\tau,d_s,h_s}(x)^2\,dx\right] &\cong \frac{2}{nd_s}\int\frac{s_\tau(x)^4}{g(x)}dx\\
&-\frac{1}{n^2d_s^2h_s}R(K)\int\frac{s_\tau(x)^4}{g(x)^2}dx\\
&+\frac{2h_s^2}{nd_s}\int\frac{s_\tau(x)^4}{g(x)}\frac{\partial f(q_\tau(x)|X=x)\;q_\tau^{(2)}(x)}{\partial\tau}\;dx\\
&+\frac{1}{2n^2d_s^2h_s}\int(K*K)^2(u)\;du\;\int\frac{s_\tau(x)^4}{g(x)^2}dx\\
&+\frac{h_s^4}{4nd_s}\;\int\frac{s_\tau(x)^4}{g(x)}\;\left[\frac{\partial f(q_\tau(x)|X=x)\;q_\tau^{(2)}(x)}{\partial\tau}\right]^2dx
\end{aligned}$$

where the terms of orders $\frac{h_s^2}{nd_s}$ and $\frac{h_s^4}{nd_s}$ are negligible in respect of $\frac{1}{nd_s}$ because the bandwidth h_s converges to zero.

On the other hand, as consequence of Theorem 3.16, we can write

$$\mathbb{B}\text{ias}\left[\int \widehat{s}_{\tau,+d_s,h_s}(x)^2\,dx\right] \cong \frac{1}{3}\,d_s^2 \int s_\tau(x)\,s_\tau^{(2,\tau)}(x)\,dx + \mu_2(K)\;h_s^2 \int s_\tau(x)\,\left.\frac{\partial q_\tau^{(2)}(x)}{\partial \tau}\right|_\tau dx + \frac{1}{2nd_s h_s}\,R(K) \int \frac{1}{g(x)}\,s_\tau(x)^2\,dx.$$

So, bearing in mind the previous expressions, it follows the result enunciated in this theorem. □

Remark 3.4. The therm of the mean squared error of the integrated sparsity estimator with order $\frac{1}{n^2 d_s^2 h_s}$ is negligible with respect to $\frac{1}{nd_s}$, but we have considered it in order to avoid degenerate h_s bandwidths (not convergence to zero). Remember Remark 3.3 in which other non dominant term has been included in order to avoid degenerate situations.

3.5 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 plug-in bandwidth and Abberger (1998)'s cross-validation bandwidth. As regards Yu and Jones (1998)'s proposal, some theoretical considerations are useful as an orientation to a meaningful comparison. Recall the expression given in ((3.2)) for the asymptotically optimal bandwidth

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

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

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

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

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

where $\widehat{h}_{\text{RSW}}$ 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 curvature, while the last factor relates the sparsity with the conditional variance under normality of the error distribution.

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

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

which is generally different from the asymptotically optimal bandwidth for quantile regression, $h_{\text{AMISE},\tau}$. Meanwhile, the proposed plug-in selector $\widehat{h}_{\text{NP}}$ converges to $h_{\text{AMISE},\tau}$. 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 $h_{\text{AMISE,YJ},\tau}$ and $h_{\text{AMISE},\tau}$ 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 $h_{\text{AMISE,YJ},\tau}$ and $h_{\text{AMISE},\tau}$ will be

$$\text{Ratio} = \frac{h_{\text{AMISE,YJ},\tau}}{h_{\text{AMISE},\tau}} = \sqrt[5]{\frac{\sigma^2\ f_\varepsilon(F_\varepsilon^{-1}(\tau))^2}{\phi(\Phi^{-1}(\tau))^2}} \tag{3.56}$$

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:

$$\frac{\text{AMISE}(h_{\text{AMISE, YJ},\tau})}{\text{AMISE}(h_{\text{AMISE},\tau})} = \frac{1}{5}\text{Ratio}^4 + \frac{4}{5}\text{Ratio}^{-1} \tag{3.57}$$

where Ratio is defined in (3.56). Note that, by construction, the ratio between AMISEs is always larger or equal to one. Part (a) of Figure 3.2 shows the values taken by the ratio defined in (3.56) 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 Figure 3.2 shows the values taken by the ratio defined in (3.57) as a function of τ and for the same three error distributions.

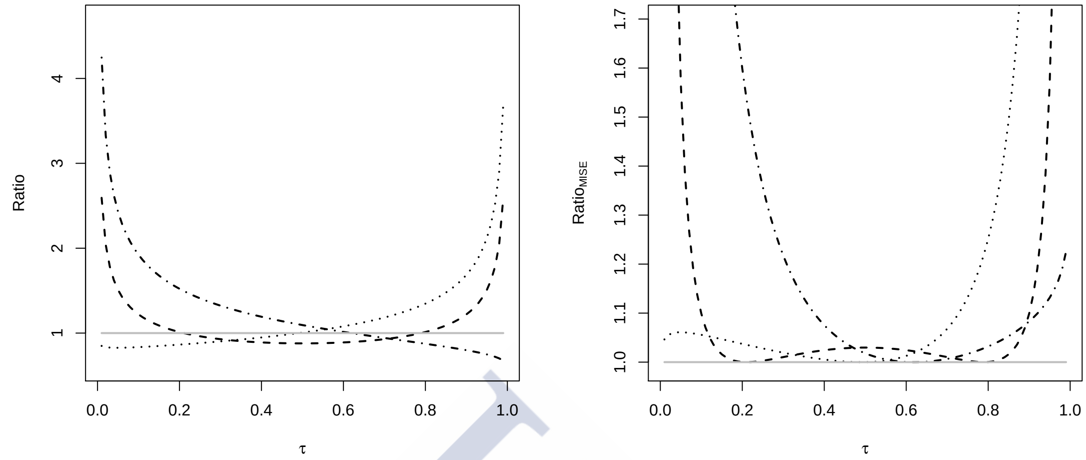
As shown in Figure 3.2, we observe that the differences between both plug-in selectors will be bigger as long as the error distribution differs from the Gaussian distribution. Furthermore, if we fix an error distribution, the compared behaviour of both plug-in selectors will depend on the quantile of interest.

Our first simulated model is given by

$$\textbf{Model 3.1:} \qquad Y = 10(X - 0.5)^4 + 3(X - 0.5)^2 + \varepsilon,$$

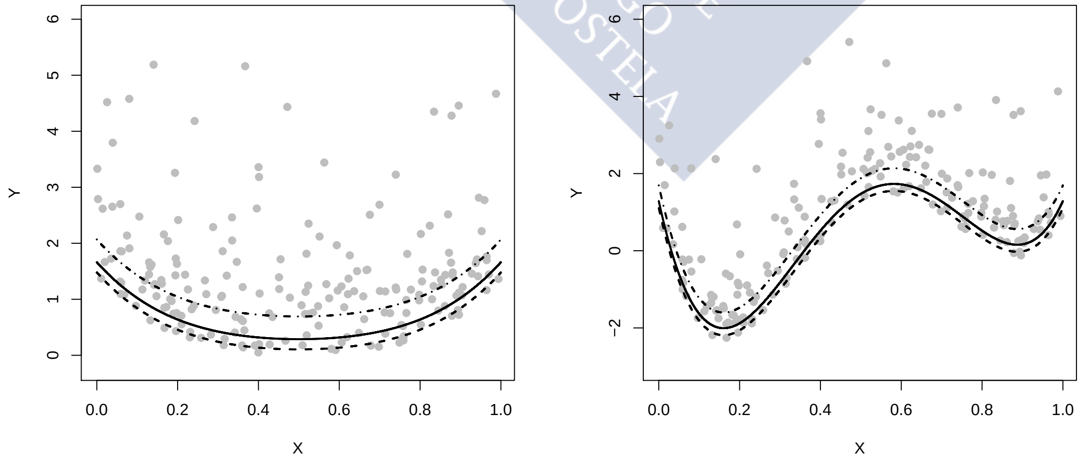
where X follows a uniform distribution on the interval $(0, 1)$ and ε is the unknown error, which is drawn independently of X. Note that in this case, $q_\tau(X) = 10(X - 0.5)^4 + 3(X - 0.5)^2 + c_\tau$ where c_τ represents the τ-quantile of the error distribution. This notation in common for all the 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 Figure 3.2. Part (a) of Figure 3.3 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 3.4 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 *YJ, RT, NP, CV*, respectively. The boxplots were obtained from 1000 replications of Model 3.1 for different values of τ, and sample sizes $n = 100, 500, 1000$. Three horizontal lines are added



(a) Ratio between AMISE bandwidths (b) Ratio between AMISE values

Figure 3.2: Representations of the ratios between the AMISE bandwidths (detailed in (3.56)) and the MISE values (detailed in (3.57)) as a function of the quantile order τ and for three error distributions. The dashed line (−−) represents the uniform distribution, the dotted line $(\cdots)$ represents the beta distribution and the dashed and dotted line $(- \cdot -)$ represents the exponential distribution.



(a) Model 3.1 (b) Model 3.2

Figure 3.3: Scatterplots of a sample of size 200, together with three quantile functions: $\tau = 0.1$ (dashed line), $\tau = 0.25$ (solid line) and $\tau = 0.5$ (dashed and dotted line), corresponding to Model 3.1 in (a) and Model 3.2 in (b).

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 approaches 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, Figure 3.4 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 MISE over 1000 samples. We expect that MISE results will be a consequence of the selectors' distances from the MISE bandwidth, observed in the boxplots. To complete the presentation, a new model is included, again homoscedastic but with a larger curvature:

$$\textbf{Model 3.2:} \qquad Y = 1 - 48X + 218X^2 - 315X^3 + 145X^4 + \varepsilon,$$

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 Figure 3.3 shows a scatterplot and three quantile functions, for $\tau = 0.1, 0.25, 0.5$, corresponding to Model 3.2.

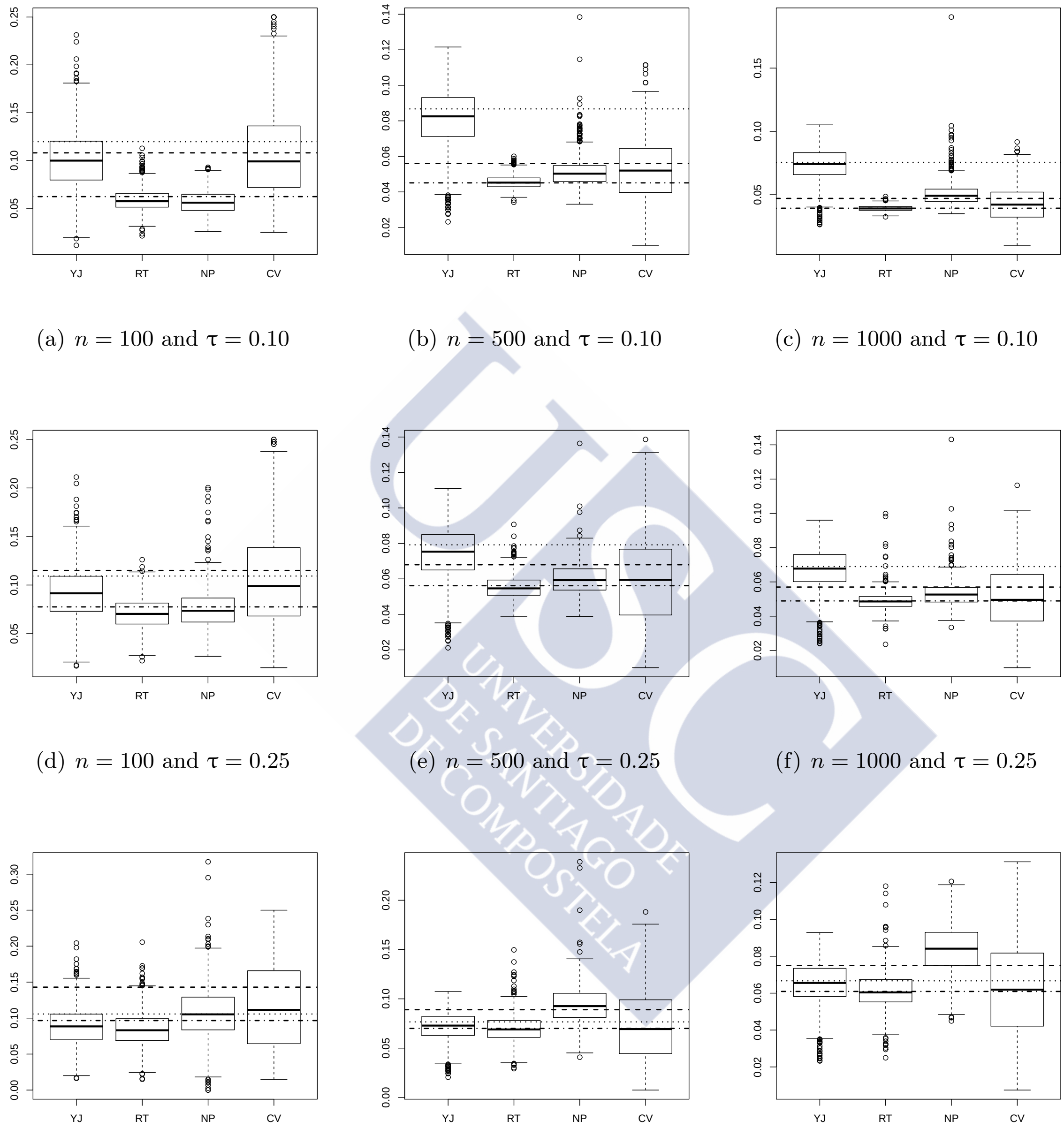
Table 3.1 contains the mean integrated squared error for the considered bandwidth selectors for several samples sizes and values of τ. We can observe that the new plug-in rule shows a better performance in terms of MISE than the plug-in selector proposed by Yu and Jones (1998), for all values of τ and 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 verify some ideal conditions for the rule of thumb. The cross-validation bandwidth shows a generally worst MISE 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:

$$\textbf{Model 3.3:} \; Y = 1 - 48X + 218X^2 - 315X^3 + 145X^4 + \varepsilon$$
$$\textbf{Model 3.4:} \; Y = \sin(5\pi X) + \varepsilon$$

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 one degree of freedom and standard log-normal. Model 3.3 has the same quantile function than Model 3.2, while the error distribution now takes different shapes. Model 3.4 is represented in Part (a) of Figure 3.5, with a standard Gaussian distribution.

Table 3.2 shows the mean integrated squared errors for the compared bandwidth selectors, under Model 3.3 and Model 3.4. In all cases, the quantile function is estimated for $\tau = 0.5$. The



(a) $n = 100$ and $\tau = 0.10$ (b) $n = 500$ and $\tau = 0.10$ (c) $n = 1000$ and $\tau = 0.10$

(d) $n = 100$ and $\tau = 0.25$ (e) $n = 500$ and $\tau = 0.25$ (f) $n = 1000$ and $\tau = 0.25$

(g) $n = 100$ and $\tau = 0.50$ (h) $n = 500$ and $\tau = 0.50$ (i) $n = 1000$ and $\tau = 0.50$

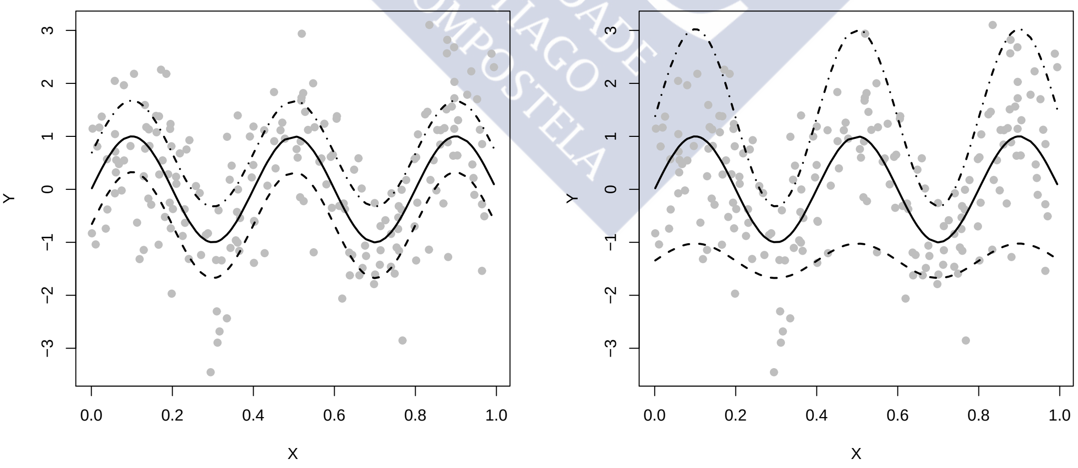
Figure 3.4: Boxplot representations of Yu and Jones (1998)'s selector (YJ), the new rule of thumb (RT), the new plug-in selector (PI) and the cross-validation selector (CV), from 1000 replications of Model 3.1 for different values of τ and the sample size, n. The dashed line $(--)$ represents the MISE bandwidth, the dotted line $(\ldots)$ represents the Yu and Jones (1998)'s AMISE bandwidth and the dashed and dotted line $(-\cdot-)$ represents the AMISE bandwidth.

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

		Model 3.1				Model 3.2			
		YJ	RT	NP	CV	YJ	RT	NP	CV
$\tau = 0.10$	$n = 100$	27.86	27.96	27.69	27.49	35.50	54.13	37.24	34.08
	$n = 500$	5.68	3.85	3.85	4.67	7.20	6.06	5.64	6.32
	$n = 1000$	3.49	2.01	2.00	2.39	4.48	3.11	2.97	3.38
$\tau = 0.25$	$n = 100$	44.08	40.49	40.05	48.84	54.41	64.91	53.27	60.80
	$n = 500$	8.91	7.69	7.73	9.67	12.50	11.59	11.37	13.36
	$n = 1000$	5.12	4.21	4.23	5.12	7.51	6.40	6.41	7.35
$\tau = 0.50$	$n = 100$	89.90	83.39	82.58	104.27	114.80	118.60	109.13	129.66
	$n = 500$	17.86	17.66	17.64	22.51	25.61	25.98	25.50	30.93
	$n = 1000$	9.85	9.76	9.74	12.41	14.91	14.80	14.77	17.78

Table 3.1: Mean integrated squared error (given values were multiplied by 10^3) associated with the considered bandwidth selectors, for Model 3.1 and Model 3.2, with several sample sizes n and values of τ.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) Model 3.4 (b) Model 3.5

Figure 3.5: Scatterplots of a sample of size 200 drawn from Model 3.4 and Model 3.5, 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).

		Model 3.3				Model 3.4			
ε	n	YJ	RT	NP	CV	YJ	RT	NP	CV
$N(0,1)$	100	13.40	13.83	12.89	15.82	15.52	18.26	15.40	17.57
	500	3.45	3.54	3.43	4.19	3.92	4.25	3.92	4.72
	1000	2.00	2.05	2.00	2.41	2.25	2.39	2.26	2.73
$U(-3,3)$	100	52.09	52.04	49.68	59.53	57.56	62.13	55.89	63.30
	500	14.71	14.83	14.21	17.80	16.41	18.33	16.33	20.85
	1000	8.57	8.67	8.29	10.54	9.25	10.15	9.18	11.24
t_1	100	3310.02	174.1	175.6	67.54	3320.08	40.81	36.50	81.01
	500	12.49	5.65	5.40	6.74	19.32	10.51	6.93	7.62
	1000	10.24	3.10	3.00	3.72	16.88	4.48	3.51	4.11
log N(0,1)	100	26.09	20.47	18.60	21.67	31.25	30.27	23.80	24.59
	500	4.30	3.98	3.88	4.77	4.75	4.55	4.18	5.12
	1000	2.54	2.21	2.19	2.65	2.78	2.50	2.33	2.84

Table 3.2: Mean integrated squared error (given values were multiplied by 10^2) associated with the considered bandwidth selectors, for Model 3.3 and Model 3.4 with $\tau = 0.5$, and several error distributions and sample sizes.

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 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 3.3 than under Model 3.4, because the quantile function under Model 3.3 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.

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

$$\textbf{Model 3.5:}\ \ Y = \sin(5\pi X) + (\sin(5\pi X) + 2)\varepsilon$$

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 τ-quantile of the error distribution. Two error distributions are considered. One is the standard Gaussiann distribution. Then, the main deviation of Model 3.5 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 Figure 3.5 shows a representation of Model 3.5. A scatterplot together with three quantile functions (for $\tau = 0.25, 0.5, 0.75$) are shown. It can be seen how heteroscedasticity leads to different curvatures of the quantile regression

		Standard Gaussian				Student's t with 2 degrees of freedom			
τ	n	YJ	RT	NP	CV	YJ	RT	NP	CV
0.25	100	59.83	53.24	50.67	51.42	120.37	97.74	84.61	69.14
	500	14.05	13.28	11.78	14.31	19.64	16.63	15.56	14.28
	1000	7.85	7.73	6.72	8.52	10.36	9.08	8.44	7.78
0.50	100	61.05	64.88	57.33	62.86	94.93	72.62	67.97	64.94
	500	14.30	16.45	14.30	18.70	18.54	20.59	17.65	17.88
	1000	8.27	9.38	8.30	10.82	10.28	10.94	9.74	10.58
0.75	100	86.53	99.14	80.57	87.86	166.40	187.47	137.90	125.29
	500	21.55	23.86	20.36	25.86	37.28	37.13	30.60	32.40
	1000	12.26	13.56	11.58	14.48	21.01	19.35	16.78	20.24

Table 3.3: Mean integrated squared error (given values were multiplied by 10^2) associated with the compared bandwidth selectors, under Model 3.5 and for several values of τ and the sample size n.

function for different values of τ. We will also suppose that the error follows a Student t distribution with two degrees of freedom. In this second situation, neither of the assumptions considered by Yu and Jones (1998) are verified.

In Table 3.3 the mean integrated squared error from each of the bandwidth selectors are 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 mean 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.

3.6 The BwQuant package

In order to facilitate the use ot the new bandwidth selector that have been proposed along this chapter, an R package has been designed. The new R package is called `BwQuant` and implements several bandwidth selectors for local linear quantile regression. In order to illustrate the usage of the different functions included in this package, we are going to employ the `mcycle` dataset used previously by several authors as Koenker (2005) or Venables and Ripley (1999). This classic nonstationary dataset consists of measurements of the acceleration of the head of a motorcycle rider as a function of time in the first moments after an impact. In addition to being nonstationary, the error associated with this data is not homoscedastic. The `mcycle` dataset is available on the R package `MASS` and it contains a series of 133 measurements of head acceleration in a simulated motorcycle accident, used to test crash helmets.

The main functions of the `BwQuant` package are:

- bwCV that implements the bandwidth selector associated with the cross-validation process proposed by Abberger (1998).

Usage: `bwCV(x,y,h,tau)`

Arguments:

x	numeric vector of explanatory variable data.
y	numeric vector of response variable data.
h	sequence of values where we want to evaluate the cross-validation function in order to obtain the bandwidth that minimizes this function.
tau	value of the τ-th quantile of interest that we want to estimate.

Example
```
library(MASS)
data(mcycle)
attach(mcycle)
hseq=seq(0.5,3,length=50)
bwCV(times,accel,hseq,0.75)
```

- bwNP that implements the new proposed plug-in selector based on a nonparametric estimations of the conditional sparsity and curvature. Note that, in this case, in order to solve the optimization problem associated with the calculus of pilot bandwidths $\widehat{h}_s$ and $\widehat{d}_s$ we will employ the optimization algorithm proposed by Nelder and Mead (1965) that is based on a Simplex idea. The selection of this method is due to necessity to impose constraints on the bandwidth $\widehat{d}_s$ that, at the end of the day, represents a quantile order. The optimization method proposed by Nelder and Mead (1965) have been implemented thanks to the R function constrOptim.

Usage:
```
bwNP(x,y,tau,blockmax=5,divisor=20,proptrun=0.05,itermax=5000,
     eps=1e-06)
```

Arguments:

x	numeric vector of explanatory variable data.
y	numeric vector of response variable data.
tau	value of the τ-th quantile of interest that we want to estimate.
blockmax	the maximum number of blocks of the data for construction of an initial parametric model estimation.
divisor	the value that the sample size is divided by to determine a lower limit on the number of blocks of the data for construction of an initial parametric model estimation.

`proptrun`	the proportion of the range of `x` at each end truncated in the curvature estimation.
`itermax`	maximum of iterations allowed in the optimization method needed to obtain the pilot bandwidths associated with the sparsity estimation.
`eps`	value considered in order to establish the converge condition related to the optimization method.

Example
```
library(MASS)
data(mcycle)
attach(mcycle)
bwNP(times,accel,0.75)
```

- `bwRT` that implements the bandwidth selector obtained as a consequence of applying the rule of thumb developed on this work.

Usage: `bwRT(x,y,tau,blockmax=5,divisor=20)`

Arguments:

`x`	numeric vector of explanatory variable data.
`y`	numeric vector of response variable data.
`tau`	value of the τ-th quantile of interest that we want to estimate.
`blockmax`	the maximum number of blocks of the data for construction of an initial parametric model estimation.
`divisor`	the value that the sample size is divided by to determine a lower limit on the number of blocks of the data for construction of an initial parametric model estimation.

Example
```
library(MASS)
data(mcycle)
attach(mcycle)
bwRT(times,accel,0.75)
```

- `bwYJ` that implements the plug-in selector proposed by Yu and Jones (1998).

Usage: `bwYJ(x,y,tau)`

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Arguments:

`x`	numeric vector of explanatory variable data.
`y`	numeric vector of response variable data.
`tau`	value of the τ-quantile of interest that we want to estimate.

Example
```
library(MASS)
data(mcycle)
attach(mcycle)
bwYJ(times,accel,0.75)
```

Moreover, we have included the `llqr` function that estimates the quantile regression function using local linear kernel regression.

Usage: `llqr(x,y,tau,h,gridsize=50)`

Arguments:

`x`	numeric vector of explanatory variable data.
`y`	numeric vector of response variable data.
`tau`	value of the τ-th quantile of interest that we want to estimate.
`h`	the bandwidth parameter.
`gridsize`	the number of points at which the local linear quantile regre-ssion is going to be estimated.

Example
```
library(MASS)
data(mcycle)
attach(mcycle)
h=bwNP(times,accel,0.75)
llqr(x,y,tau,h)
```

Thanks to the function `llqr` we can represent the fit of the local linear regression models associated with the different bandwidth selectors that we have considered. For instances, Figure 3.6 shows the smoothed 75th quantile curve related to both plug-in methods. Bearing in mind Figure 3.6 we can conclude that the estimation associated with the new plug-in method is a bite smoother than that related to the plug-in method proposed by Yu and Jones (1998).

3.7 Conclusions

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. The mean squared error of these nonparametric estimators have been obtained along this chapter.

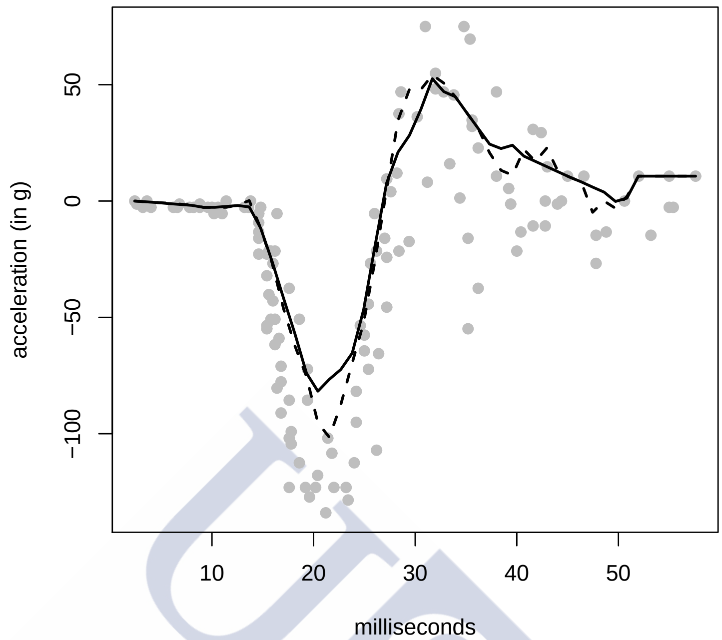


Figure 3.6: Representation of the smoothed 75th quantile regression model for the `mcycle` dataset set using the plug-in proposed by Yu and Jones (1998) (dashed line) and the new plug-in rule developed along this chapter (solid line).

Thanks to a Monte Carlo simulation study, we have shown that the new proposal performs well in terms of the mean integrated squared error compared with its natural competitors both in 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.

In addition, we have developed an R package called `BwQuant` that enables any user to apply the different techniques that have been proposed throughout this chapter. Finally an application to real data situation of the new plug-in rule will be presented in Section 5.4.

USC
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 4

A lack-of-fit test for quantile regression models with high-dimensional covariates

Contents

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

A new lack-of-fit test for quantile regression models, that is suitable even with high-dimensional covariates, is proposed along this chapter. The test is based on the cumulative sum of residuals with respect to unidimensional linear projections of the covariates. To approximate the critical values of the test, a wild bootstrap mechanism convenient for quantile regression is used. An extensive simulation study was undertaken that shows the good performance of the new test, particularly when the dimension of the covariate is high. The test can also be applied and performs well under heteroscedastic regression models. The test is illustrated with real data about the economic growth of 161 countries.

4.1 Introduction

The lack-of-fit (or in opposite terms, goodness-of-fit) of a statistical model describes how well it fits a set of observations. Measures of goodness-of-fit typically summarize the discrepancy between observed values and the values expected under the model in question. With the aim of testing if a data distribution belongs to a certain parametric family, Pearson introduced at the beginning of the twentieth century the term goodness-of-fit. Since then, there has been an enormous amount of papers on this topic. Along this section we are going to present a small introduction to lack-of-fit tests for classical mean regression (Section 4.1.1) and for quantile regression models (Section 4.1.2).

4.1.1 Lack-of-fit tests for mean regression

Let us consider a classical mean regression model denoted by

$$Y = m(X) + \varepsilon$$

where $m(x) = \mathbb{E}(Y|X = x)$ is the regression function of Y over X and the error has to verify that $\mathbb{E}(\varepsilon|X) = 0$ almost surely. Given a random sample $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ of the variables $(X, Y) \in \mathbb{R}^{d+1}$, the goal will be to test

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

Hereafter we are going to present different kinds of specification tests for mean regression models. Note that we will only mention some examples, an up-to-date review of the most important recent contributions on lack-of-fit tests for regression models is given in González-Manteiga and Crujeiras (2013).

Lack-of-fit tests based on smoothing ideas

In this first item, we are going to focus on kernel type estimators such as the Nadaraya-Watson estimator given by Nadaraya (1964) and Watson (1964). This estimator can be written as follows

$$m_{nh}(x) = \sum_{i=1}^{n} W_{ni}(x) Y_i = \sum_{i=1}^{n} \frac{K_h(x - X_i)}{\sum_{j=1}^{n} K_h(x - X_j)} Y_i$$

where $K_h(x) = h^{-1}K(x/h)$, K represents the kernel function and h is the well known bandwidth parameter. Bearing in mind the previous estimator, the following empirical process has been considered:

$$R_n^h(x) = \sqrt{nh^d} \sum_{i=1}^n \left(Y_i - m(X_i, \widehat{\theta}_{\text{LS}})\right) W_{ni}(x) = \sqrt{nh^d} \sum_{i=1}^n W_{ni}(x)\, r_i$$

where $r_i = Y_i - m(X_i, \widehat{\theta}_{\text{LS}})$ and $\widehat{\theta}_{\text{LS}}$ is a $\sqrt{n}$-consistent estimator of θ that is the value of the true parameter under H_0, computed thanks to least squares techniques.

Then, a test based on R_n^h can be devised by applying a continuous functional on the empirical process, such as

$$T_n^h = \int (R_n^h(x))^2 w(x)dx$$

where w is a weight function. Moreover, the test proposed by Härdle and Mammen (1993) can be written as

$$T_n^{HM} = \int \left(m_{nh}(x) - m_{nh}(x, \widehat{\theta}_{\text{LS}}) \right)^2 w(x)dx$$

where $m_{nh}(x, \widehat{\theta}_{\text{LS}}) = \sum_{i=1}^n W_{ni}(x) m(X_i, \widehat{\theta}_{\text{LS}})$.

Alternatively, it is possible to define other test statistics based on consistent estimators of different characteristics of the null hypothesis. A noticeable example is the test proposed by Zheng (1996) that is based on the following test statistic

$$T_n^{Zm} = \frac{1}{n(n-1)} \sum_{i \neq j} K_h(X_i - X_j)(Y_i - m(X_i, \widehat{\theta}_{\text{LS}}))(Y_j - m(X_j, \widehat{\theta}_{\text{LS}}))w(X_i)$$

or the test statistic studied by Dette (1999), given by

$$T_n^D = \frac{1}{n} \sum_{i=1}^n (Y_i - m(X_i, \widehat{\theta}_{\text{LS}}))^2 w(X_i) - \frac{1}{n} \sum_{i=1}^n (Y_i - m_{nh}(X_i))^2 w(X_i).$$

The previous lack-of-fit tests converge to a Gaussian distribution that allows us to calibrate this kind of tests. The use of the asymptotic distributions entails selecting the smoothing parameter h, a broadly studied problem in regression estimation but with serious gaps for testing problems.

Lack-of-fit tests based on the empirical regression processes

With the aim of avoiding the selection of a smoothing parameter, an alternative methodology has been developed. The resulting lack-of-fit tests are based on the integrated regression function that is given by

$$I(x) = \mathbb{E}[Y\ \mathbb{I}(X \leq x)] = \int_{-\infty}^x m(z)dF(z)$$

where $\mathbb{I}$ denotes the indicator function. Then, the integrated regression function can be estimated by

$$I_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i \leq x)\ Y_i.$$

Now, the following empirical process can be defined:

$$R_n^S(x) = n^{-1/2} \sum_{i=1}^{n} \Big(Y_i - m(X_i, \widehat{\theta}_{\mathrm{LS}})\Big) \; \mathbb{I}(X_i \leq x) = n^{-1/2} \sum_{i=1}^{n} r_i \, \mathbb{I}(X_i \leq x) \tag{4.1}$$

where $r_i = Y_i - m(X_i, \widehat{\theta}_{\mathrm{LS}})$ represent the residuals. Under the simple null hypothesis, Stute (1997) proved that the previous empirical process converges to a Gaussian process with mean equal to zero and covariance function $K(s,t) = \widetilde{K}(s \wedge t)$ where

$$\widetilde{K}(x) = \int_{\infty}^{x} \mathbb{V}\mathrm{ar}\,(Y|X = u) F(du)$$

and $s \wedge t$ represents the minimum between s and t. Moreover, the effect of θ estimation has as a consequence a complicated form of the covariance function that hinders the calibration of the test. Subsequently, Stute et al. (1998) have proposed a wild bootstrap procedure in order to calibrate the test proposed by Stute (1997).

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

We have considered test statistics constructed from the comparison of a nonparametric estimator of the regression model and an estimator under the null hypothesis, or in the comparison of the corresponding integrated regression function estimators. In both cases, the curse of dimensionality when the dimension of the explanatory variable increases, can be appreciated. Related to the first class of tests, the effect of the increasing dimension is clear when regarding the asymptotic power of order nh^d. In the other class, the power also decay for small samples as several simulation studies have shown.

The difficulties aforementioned lead to different modifications of the previous methods in order to avoid the curse of dimensionality. For the tests based on smoothing methods, the work of Lavergne and Patilea (2008) should be noticed because they propose the following modification of Zheng (1996)'s using projections (that are denoted by $\beta' X_i$)

$$T_n^{LP} = \sup_{\beta, ||\beta||=1} \sum_{i<j} K_h(\beta'(X_i - X_j))(Y_i - m(X_i, \widehat{\theta}_{\mathrm{LS}}))(Y_j - m(X_j, \widehat{\theta}_{\mathrm{LS}})).$$

Another option in this line would be to project the covariate X in a certain direction β_0 that minimizes

$$\mathbb{E}^2(\varepsilon - \mathbb{E}(\varepsilon \mid \beta' X)) = \mathbb{E}^2(\varepsilon - m_\beta(X)).$$

This idea was developed by Xia (2009).

Regarding the tests based on empirical regression processes, Stute et al. (2008) proposed replacing the empirical process R_n^S described in (4.1) by

$$R_n^g(t) = n^{-1/2} \sum_{i=1}^{n} (g(X_i) - \bar{g}) \mathbb{I}(r_i \leq t)$$

with $t \in \mathbb{R}$, where $\bar{g} = n^{-1} \sum_{i=1}^{n} g(X_i)$ and r_i denote the residuals of the model. The selection of the function g is also discussed in Stute et al. (2008) with the goal of power maximization.

In the same line, we are going to focus on Escanciano (2006) that establishes the following characterization of the null hypothesis:

$$H_0 \text{ holds} \Longleftrightarrow \mathbb{E}[\varepsilon\, \mathbb{I}(\beta' X \leq u)] = 0 \qquad \forall \beta \in \mathbb{R}^d \text{ with } \|\beta\| = 1,$$

where $\varepsilon = Y - m(X, \theta)$ and β is the direction on which the covariate X is projected. This formulation generates a new empirical process

$$R_n^E(\beta, u) = n^{-1/2} \sum_{i=1}^{n} r_i\, \mathbb{I}(\beta' X_i \leq u)$$

where $r_i = Y_i - m(X_i, \widehat{\theta}_{\text{LS}})$ denote the residuals of the model.

4.1.2 Lack-of-fit tests for quantile regression

Now we are going to move to quantile regression setup. 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 realise the following lack-of-fit test

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

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 test 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 Zheng (1998) that extends the well-known test proposed by Zheng (1996) to the quantile regression setup. In this case, the test statistic is given by

$$\begin{aligned} T_n^{Zq} = &\frac{nh^{d/2}}{\widehat{\sigma}} \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K\left(\frac{X_i - X_j}{h}\right) \left[\mathbb{I}\big(Y_i \leq q_\tau(X_i, \widehat{\theta}_\tau)\big) - \tau\right] \\ &\times \left[\mathbb{I}\big(Y_j \leq q_\tau(X_j, \widehat{\theta}_\tau)\big) - \tau\right] \end{aligned} \tag{4.2}$$

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 (4.2) 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 Zheng (1998), Dette et al. (2012) have proposed a lack-of-fit test for additive quantile models based on smoothing ideas. In this context, the following test could be raised:

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

where $X \in \mathbb{R}^d$ denotes the explanatory variable.

Given a random sample of the variables $(X, Y) \in \mathbb{R}^{d+1}$, Dette et al. (2012) proposed the following test statistic:

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

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. In spite of 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 Stute (1997) to the quantile regression setting, He and Zhu (2003) propose an omnibus lack-of-fit test for parametric quantile regression based on a cumulative sum process of the gradient vector. That is, He and Zhu (2003) based their test on the process

$$R_n^{HZ} = n^{-1/2} \sum_{i=1}^{n} \psi_\tau\left(Y_i - q_\tau(X_i, \widehat{\theta}_\tau)\right) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau)\, \mathbb{I}(X_i \leq t) \tag{4.3}$$

where $\psi_\tau(r) = \tau\mathbb{I}(r > 0) + (\tau - 1)\mathbb{I}(r < 0)$ is the derivative of the quantile loss function $\rho_\tau = \tau r\mathbb{I}(r > 0) + (\tau - 1)r\mathbb{I}(r < 0)$, $q_\tau^{(1)}(x, \theta) = \frac{\partial}{\partial\theta} q_\tau(x, \theta)$, and $\widehat{\theta}_\tau$ is an estimator of θ_τ. The test statistic proposed by He and Zhu (2003) is then defined as

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

He and Zhu (2003) proved that the empirical process (4.3) converges to a Gaussian process with mean 0 and covariance function

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

where

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

Because simulating the Gaussian process is not easy, He and Zhu (2003) proposed a multiplier bootstrap in order to calibrate their test. It is based on the following asymptotic representation of the process (4.3):

$$R_n^{HZ}(t) = n^{-1/2}\sum_{i=1}^n \psi_\tau\left(Y_i - q_\tau(X_i,\theta_\tau)\right)\left[\mathbb{I}(X_i\le t) - S(t)S^{-1}\right] q_\tau^{(1)}(X_i,\theta_\tau) + o_P(1).$$

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 can affect the performance of the specification tests. Little can be found in the literature for lack-of-fit testing adapted to multidimensional covariates in the framework of quantile regression. Wilcox (2008) used a He and Zhu type test and defined some ranks over the covariate in order to test a linear quantile regression model such as

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

where $P_i = (1, X_i)$ and ε_i represent the model error.

In order to present Wilcox (2008)'s proposal, the following notation shoul be introduced. Let us denote by $F_i = \max U_{ij}$ 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 $j = 2,\cdots,d+1$.

Then, Wilcox (2008) have considered the following empirical process:

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

where $r_i = Y_i - \widehat{\theta}_\tau' P_i$ represents the residuals. Consequently, the test statistic will be

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

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

Although T_n^W is very similar to T_n^{HZ}, both tests generally differ. It can be shown, for instance, that it is possible to have $F_k \le F_i$ yet neither $X_k \le X_i$ or $X_k > X_i$. That is, the sum when computing T_n^W contains all of the terms used to compute T_n^{HZ} plus possibly some additional terms.

[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.

The proposal of Wilcox (2008) has the virtue of simplicity but does not provide an omnibus test, i.e., it is not consistent for all alternatives.

We have mentioned some examples, but other specification tests for quantile regression models can be found in the literature as well as

- Horowitz and Spokoiny (2002) whose goal was to test if the conditional median function is linear against a nonparametric alternative with unknown smoothness.

- Whang (2006) considered an empirical likelihood method to estimate the parameters of the quantile regression models and to construct confidence regions.

- Otsu (2008) considered two empirical likelihood-based estimation, inference, and specification testing methods for quantile regression models.

- Escanciano and Velasco (2010) proposed an omnibus specification test for parametric dynamic quantile models.

- Escanciano and Goh (2014) introduces a nonparametric test for the correct specification of a linear conditional quantile function over a continuum of quantile levels.

Taking into account the state of the art, we propose and study a lack-of-fit test for parametric models of quantile regression, with good properties for multidimensional covariates and consistent for all alternatives. In Section 4.2 we present the new test based on one-dimensional projections of the covariates. Asymptotic distribution under the null hypothesis is derived and root-n consistency is established. A bootstrap method is also proposed to approximate the critical values of the test. Section 4.3 contains a simulation study where the performance of the test is studied under homo- and heteroscedastic models, with different error distributions and with increasing dimension of the covariate. We compare the proposed test with a He and Zhu test. In Section 4.4 the test is applied to real data. Some concluding remarks and extensions are provided in Section 4.5.

4.2 The proposed method

Along this section we are going to present a new lack-of-fit test for quantile regression based on the cumulative sum of residuals with respect to unidimensional linear projections of the covariates.

4.2.1 The test

As we have mentioned, the strategy to improve the performance of the test with multiple covariates consists of applying a lack-of-fit test to one-dimensional projections of the covariates. This is motivated by a fundamental result that have been brought in Lemma 4.1.

Lemma 4.1. *The null hypothesis $H_0 : q_\tau \in \mathcal{Q}_\theta$, holds if and only if, for some $\theta_\tau \in \Theta \subset \mathbb{R}^q$, and for any $\beta \in \mathbb{R}^d$ with $\|\beta\| = 1$,*

$$\mathbb{P}[Y - q_\tau(X, \theta_\tau) \leq 0 \mid \beta' X] = \tau$$

almost surely.

Proof. This result is an extension of Lemma 1 in Escanciano (2006) to the quantile regression setting. The proof if given through the following equivalences:

$$\begin{aligned} H_0: \quad q_\tau(\cdot, \theta_\tau) \in \mathcal{Q}_\theta &\Longleftrightarrow \mathbb{P}[Y - q_\tau(X, \theta_\tau) \leq 0 \mid X] = \tau \\ &\overset{(i)}{\Longleftrightarrow} \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau)) \mid X] = 0 \\ &\overset{(ii)}{\Longleftrightarrow} \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau))\, q_\tau^{(1)}(X, \theta_\tau) \mid X] = 0 \\ &\overset{(iii)}{\Longleftrightarrow} \mathbb{E}[\psi_\tau(Y - q_\tau(X, \theta_\tau))\, q_\tau^{(1)}(X, \theta_\tau) \mid \beta' X] = 0. \end{aligned}$$

Firstly, equivalence (i) is immediate from the definition of ψ_τ. Secondly, the equivalence (ii) is clear bearing in mind basic properties of the conditional expectation. Finally, in (iii) the necessity is immediate. So, we only have to prove the sufficiency. Let us denote by

$$Z = \psi_\tau(Y - q_\tau(X, \theta_\tau)) q_\tau^{(1)}(X, \theta_\tau),$$

then, for each $\beta \neq 0$, the σ-algebra generated by $\beta' X$ match with the σ-algebra generated by $\beta' X / \|\beta\|$. As a consequence of the properties of the conditional expectation, we can conclude that for any β, included $\beta = 0$,

$$0 = \mathbb{E}[e^{i\beta' X}\, \mathbb{E}[Z \mid \beta' X]] = \mathbb{E}[e^{i\beta' X} Z] = \mathbb{E}[e^{i\beta' X}\, \mathbb{E}[Z \mid X]].$$

Therefore, applying Theorem 3.1 (Page 75) of Parthasarathy (1967), we end that $\mathbb{E}\,[Z|X] = 0$ almost sure. □

If the true parameter θ_τ was known, the test could be based on the process

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

Otherwise, an estimator $\widehat{\theta}_\tau$ is substituted, yielding the process useful for lack-of-fit testing of the parametric model

$$R_n^1(\beta, u) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau\left(Y_i - q_\tau\left(X_i, \widehat{\theta}_\tau\right)\right)\; q_\tau^{(1)}\left(X_i, \widehat{\theta}_\tau\right)\; \mathbb{I}\left(\beta' X_i \leq u\right).$$

where $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ represents a random sample of the variables $(X, Y) \in \mathbb{R}^{d+1}$.

The test statistic is then defined as

$$T_n = \text{largest eigenvalue of} \quad \int_\Pi R_n^1(\beta, u)[R_n^1(\beta, u)]' F_{n,\beta}(du) d\beta, \tag{4.4}$$

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

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.2.2 Asymptotic properties

The process R_n^1 is similar to that proposed by Escanciano (2006), with two differences: the loss function is now the quantile loss function, and the gradient vector $q_\tau^{(1)}(X_i, \widehat{\theta}_\tau)$ is introduced following the suggestion of He and Zhu (2003). Theorem 4.2 shows the limit distribution of the new test under the simple null hypothesis while Theorem 4.4 shows the equivalent one under the composite null hypothesis. In order to obtain these results, it will be necessary to introduce the following conditions:

C1. The first derivative of the quantile regression function, denoted by $q_\tau^{(1)}(\cdot, \theta_\tau)$ is bounded.

C2. The errors are not assumed to be identically distributed. In particular, the conditional density $f(\cdot|X)$ of the error is bounded in a neighbourhood of 0 with $f(0|X) > 0$ and $|f(u|X) - f(0|X)| \leq c|u|^{1/2}$ for some $c < \infty$.

C3. There are functions $A(x)$, $B(x)$, and constant C such that

$$\sup_{\|\gamma - \theta_\tau\| \leq C} \|q_\tau^{(1)}(x, \gamma)\| \leq A(x)$$

$$\|q_\tau^{(1)}(x, \gamma_1) - q_\tau^{(1)}(x, \gamma_2)\| \leq B(x)\|\gamma_1 - \gamma_2\| \quad \text{for any } x,\ \gamma_1, \gamma_2$$

with $\mathbb{E}(|A(X)|^3)$, $\mathbb{E}(|h(X)A(X)|)$, $\mathbb{E}(|h(X)A(X)^3|)$ and $\mathbb{E}(|B(X)|^2) < \infty$.

Now, we are able to present the limit distribution of the empirical process R_n under the simple null hypothesis. This results is stated in the following theorem:

Theorem 4.2. *Let us consider a quantile regression model*

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

Then, under condition C1, the limit distribution of R_n under the simple null hypothesis, $H_0 : q_\tau = q_\tau(x, \theta_\tau)$ with θ_τ known, can be expressed as

$$R_n \xrightarrow{d} R_\infty,$$

where R_∞ is a Gaussian process with mean zero and covariance given by

$$K(x_1, x_2) = \tau(1-\tau)\mathbb{E}\left[q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)' \mathbb{I}(\beta_1' X \leq u_1)\mathbb{I}(\beta_2' X \leq u_2)\right],$$

where $x_1 = (\beta_1', u_1)'$ and $x_2 = (\beta_2', u_2)'$.

Proof. This result is proved by showing the convergence of finite dimensional distributions and demonstrating that the family of function is a VC-class (see Van der Vaart (2000)). To this aim, let us introduce the following notation:

$$\widetilde{x}_k = \psi_\tau(\varepsilon_i)\, q_\tau^{(1)}(X_i, \theta_\tau)\, \mathbb{I}(\beta_k' X_i \leq u_k).$$

Then, given a finite family of pairs $x_1 = (\beta_1', u_1), \ldots, x_m = (\beta_m', u_m)$, the convergence of the finite-dimensional distributions of R_n is consequence of the multivariate central limit theorem. Now we are going to check the conditions needed to apply the multivariate central limit theorem. Indeed,

$$\begin{aligned}\mathbb{E}\left(\widetilde{x}_k\right) &= \mathbb{E}\left(\psi_\tau(\varepsilon_i)\, q_\tau^{(1)}(X_i,\theta_\tau)\, \mathbb{I}(\beta_k' X_i \le u_k)\right)\\ &= \mathbb{E}\left(\mathbb{E}\left[\psi_\tau(\varepsilon_i)\, q_\tau^{(1)}(X_i,\theta_\tau)\, \mathbb{I}(\beta_k' X_i \le u_k) \;\middle|\; X = X_i\right]\right)\\ &= \mathbb{E}\left(\mathbb{E}\left[\psi_\tau(\varepsilon_i) \mid X = X_i\right]\, q_\tau^{(1)}(X_i,\theta_\tau)\, \mathbb{I}(\beta_k' X_i \le u_k)\right) \overset{(i)}{=} 0\end{aligned}$$

where in (i) we have used that $\psi_\tau(\varepsilon_i) = \tau - \mathbb{I}(\varepsilon_i < 0)$ where the conditional distribution of $\mathbb{I}(\varepsilon_i < 0)$ given X_i is a Bernoulli distribution with parameter τ. As a result $\mathbb{E}\left[\psi_\tau(\varepsilon_i) \mid X = X_i\right] = 0$. Moreover,

$$\begin{aligned}\mathbb{C}\mathrm{ov}\left(\widetilde{x}_k, \widetilde{x}_l\right) = \mathbb{E}\left(\widetilde{x}_k \widetilde{x}_l'\right) &= \mathbb{E}\left(\psi_\tau(\varepsilon_i)^2\, q_\tau^{(1)}(X_i,\theta_\tau)\, q_\tau^{(1)}(X_i,\theta_\tau)'\, \mathbb{I}(\beta_k' X_i \le u_k)\, \mathbb{I}(\beta_l' X_i \le u_l)\right)\\ &= \mathbb{E}\left(\mathbb{E}\left[\psi_\tau(\varepsilon_i)^2\, q_\tau^{(1)}(X_i,\theta_\tau)\, q_\tau^{(1)}(X_i,\theta_\tau)'\, \mathbb{I}(\beta_k' X_i \le u_k)\, \mathbb{I}(\beta_l' X_i \le u_l) \;\middle|\; X = X_i\right]\right)\\ &= \mathbb{E}\left(\mathbb{E}\left[\psi_\tau(\varepsilon_i)^2 \mid X = X_i\right]\; q_\tau^{(1)}(X_i,\theta_\tau)\, q_\tau^{(1)}(X_i,\theta_\tau)'\, \mathbb{I}(\beta_k' X_i \le u_k)\, \mathbb{I}(\beta_l' X_i \le u_l)\right)\\ &\overset{(ii)}{=} \mathbb{E}\left(\tau(1-\tau)\, q_\tau^{(1)}(X_i,\theta_\tau)\, q_\tau^{(1)}(X_i,\theta_\tau)'\, \mathbb{I}(\beta_k' X_i \le u_k)\, \mathbb{I}(\beta_l' X_i \le u_l)\right)\\ &= \tau(1-\tau)\, \mathbb{E}\left(q_\tau^{(1)}(X_i,\theta_\tau)\, q_\tau^{(1)}(X_i,\theta_\tau)'\, \mathbb{I}(\beta_k' X_i \le u_k)\, \mathbb{I}(\beta_l' X_i \le u_l)\right)\end{aligned}$$

where (ii) is consequence of the fact that $\mathbb{E}\left[\psi_\tau(\varepsilon_i)^2 \mid X = X_i\right] = \tau(1-\tau)$.

The stochastic equicontinuity (tightness) is drawn from the fact that the family of functions that we are considering is a VC-class. In order to prove this statement, we will focus on checking that the family of the indicator functions of the form $\mathbb{I}(\beta' X \le u)$ is a VC-class because the quantity $\psi_\tau(\varepsilon)\, q_\tau^{(1)}(X,\theta_\tau)$ is bounded. Then, the family of the indicator functions is a VC-class as a consequence of the two following arguments:

- Van Der Vaart and Wellner (1996) established that the set of all half spaces $\mathcal{H}_r$ in $\mathbb{R}^r$ is a VC-class (see Problem 14 of page 152) where a half space is given by

$$H(\beta, u) = \{x \in \mathbb{R}^r : \beta' x \le u\}$$

  and consequently

$$\mathcal{H}_r = \{H(\beta, u) : \beta \in \mathbb{S}_r, u \in \mathbb{R}\}$$

  where $\mathbb{S}_r = \{x \in \mathbb{R}^r : \|x\| = 1\}$. See Problem 8 of Wellner (2005) for a more complete explanation.

- Van der Vaart (2000) proved that a collection of sets is a VC-class of sets if and only if the collection of corresponding indicator functions is a CV-class of functions (see page 275).

Note that in this case, it is not necessary to lay down conditions concerning the residuals of the model, as those required by Escanciano (2006), because here the residuals are weighted by the function ψ_τ. □

Now we are going to move to the composite null hypothesis. In order to obtain the representation of the empirical process R_n^1 under this hypothesis, it will be necessary to present a Bahadur-type representation for the estimator $\widehat{\theta}_\tau$. This result will be similar to that stated by He and Zhu (2003) in Lemma A.1. Differently from their Lemma A.1, here homoscedasticity will not be assume.

Lemma 4.3 will be crucial in order to state the limit distribution of the empirical process R_n^1 under the composite null and alternative hypotheses where the true parameter θ_0 is estimated by means of $\widehat{\theta}_\tau$.

Lemma 4.3. *Let us assume that the data come from*

$$Y_i = q_\tau(X_i, \theta_\tau) + n^{-1/2} h(X_i) + \varepsilon_i \qquad i \in \{1, \dots, n\},$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent errors with conditional τ-th quantile equal to zero. Under conditions C1-C3, it follows that

$$\sqrt{n}\left(\widehat{\theta}_\tau - \theta_\tau\right) = S^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) + \mathbb{E}\left[f(0|X) h(X) q_\tau^{(1)}(X, \theta_\tau)\right]\right] + o_p(1)$$

where $S = \mathbb{E}[f(0|X) q_\tau^{(1)}(X, \theta_\tau) q_\tau^{(1)}(X, \theta_\tau)']$.

Proof. Let us remember that $\widehat{\theta}_\tau$ is defined as

$$\widehat{\theta}_\tau = \arg\min_\theta \sum_{i=1}^{n} \rho_\tau(Y_i - q_\tau(X, \theta)) = \arg\min_\theta \mathrm{LF}(\theta).$$

As He and Zhu (2003) stated, the directional derivative of $\mathrm{LF}(\theta)$ at $\widehat{\theta}_\tau$ along any direction v with $\|v\| = 1$ is nonnegative. That is,

$$\lim_{t \to 0} \frac{\mathrm{LF}(\widehat{\theta}_\tau + tv) - \mathrm{LF}(\widehat{\theta}_\tau)}{t} \geq 0.$$

Direct evaluations give

$$\begin{aligned}\sum_{Y_i \neq q_\tau(X_i, \widehat{\theta}_\tau)} & \psi_\tau\left(Y_i - q_\tau(X_i, \widehat{\theta}_\tau)\right)\ v'\ q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \\ & \leq \sum_{Y_i = q_\tau(X_i, \widehat{\theta}_\tau)} \psi_\tau\left(Y_i - v' q_\tau(X_i, \widehat{\theta}_\tau)\right)\ v'\ q_\tau^{(1)}(X_i, \widehat{\theta}_\tau)\end{aligned} \tag{4.5}$$

which is bounded by $\sum_{Y_i = q_\tau(X_i, \widehat{\theta}_\tau)} |A(X_i)|$. With probability one, there are only a finite number of points with zero residuals. The moment condition on $A(x)$ implies that $\max_{1 \leq i \leq n} |A(X_i)| = o_p(\sqrt{n})$. Thus, we can rewrite the expression (4.5) as

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau\left(Y_i - q_\tau(X_i, \widehat{\theta}_\tau)\right)\ q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) = o_p(1). \tag{4.6}$$

Let us denote by e_j a new variable that has the same distribution as ε_j and let F_j and f_j the corresponding distribution and density functions, respectively. Note that here we have introduced the main different between this Lemma 4.3 and Lemma A.2 of He and Zhu (2003) because we do not assume a common distribution for the error ε. Then, it would be verified that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \psi_\tau \left(\varepsilon_i - q_\tau(X_i, \widehat{\theta}_\tau) + q_\tau(X_i, \theta_\tau) + \frac{h(X_i)}{\sqrt{n}} \right) \right. \\
&\left. - \mathbb{E}_{e_i} \left[\psi_\tau \left(e_i - q_\tau(X_i, \widehat{\theta}_\tau) + q_\tau(X_i, \theta_\tau) + \frac{h(X_i)}{\sqrt{n}} \right) \right] \right\} q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \\
&\overset{(i)}{=} -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}_{e_i} \left[\psi_\tau \left(e_i - q_\tau(X_i, \widehat{\theta}_\tau) + q_\tau(X_i, \theta_\tau) + \frac{h(X_i)}{\sqrt{n}} \right) \right] \Big\} q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) + o_p(1) \\
&\overset{(ii)}{=} -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \tau \left[1 - F_i \left(q_\tau(X_i, \widehat{\theta}_\tau) - q_\tau(X_i, \theta_\tau) - \frac{h(X_i)}{\sqrt{n}} \right) \right] \right. \\
&\left. - (1-\tau) F_i \left(q_\tau(X_i, \widehat{\theta}_\tau) - q_\tau(X_i, \theta_\tau) - \frac{h(X_i)}{\sqrt{n}} \right) \right\} q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[F_i \left(q_\tau(X_i, \widehat{\theta}_\tau) - q_\tau(X_i, \theta_\tau) - \frac{h(X_i)}{\sqrt{n}} \right) - \tau \right] q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) + o_p(1) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[F_i \left(q_\tau(X_i, \widehat{\theta}_\tau) - q_\tau(X_i, \theta_\tau) - \frac{h(X_i)}{\sqrt{n}} \right) - F_i(0) \right] q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) + o_p(1) \\
&\overset{(iii)}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(q_\tau(X_i, \widehat{\theta}_\tau) - q_\tau(X_i, \theta_\tau) - \frac{h(X_i)}{\sqrt{n}} \right) f_i(\xi_{i,1}) \, q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) + o_p(1) \\
&\overset{(iv)}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f_i(\xi_{i,1}) \, q_\tau^{(1)}(X_i, \xi_2) \left(\widehat{\theta}_\tau - \theta_\tau \right) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau)' \\
&- \frac{1}{n} \sum_{i=1}^{n} f_i(\xi_{i,1}) \, h(X_i) \, q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) + o_p(1) \\
&= \sqrt{n} \left(\widehat{\theta}_\tau - \theta_\tau \right) S - \mathbb{E} \left(f(0|X) h(X) q_\tau^{(1)}(X_i, \theta_\tau) \right) + o_p(1) + o_p \left(\sqrt{n} \left(\widehat{\theta}_\tau - \theta_\tau \right) \right)
\end{aligned} \tag{4.7}
$$

Here, the equality (i) is a consequence of (4.6) because $q_\tau(X_i, \widehat{\theta}_\tau) + q_\tau(X_i, \theta_\tau) + \frac{h(X_i)}{\sqrt{n}} = r_i$. The equality (ii) is clear bearing in mind the definition of ψ_τ. Moreover, the equalities (iii) and (iv) are based on local expansions of F_i and q_τ where $\xi_{i,1}$ is an element between $q_\tau(X_i, \widehat{\theta}_\tau) - q_\tau(X_i, \theta_\tau) - \frac{h(X_i)}{\sqrt{n}}$ and zero while ξ_2 is an element between $\widehat{\theta}_\tau$ and θ_τ.

Following the arguments of Lemma A.1 in He and Zhu (2003), it can be shown that

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\psi_\tau \left(\varepsilon_i - q_\tau(X_i, \widehat{\theta}_\tau) + q_\tau(X_i, \theta_\tau) + \frac{h(X_i)}{\sqrt{n}} \right) \right. \\
&\left. - \mathbb{E}_{e_j} \psi_\tau \left(e_i - q_\tau(X_i, \widehat{\theta}_\tau) + q_\tau(X_i, \theta_\tau) + \frac{h(X_i)}{\sqrt{n}} \right) \right] q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) + O_p \left(\left(\|\widehat{\theta}_\tau - \theta_\tau\| + n^{-1/2} \right)^{1/2} \log(n) \right)
\end{aligned} \tag{4.8}
$$

which, together with (4.7), prove this lemma because

$$
\sqrt{n} \left(\widehat{\theta}_\tau - \theta_\tau \right) S - \mathbb{E} \left(f(0|X) h(X) q_\tau^{(1)}(X_i, \theta_\tau) \right) + o_p(1) + o_p \left(\sqrt{n}(\left(\widehat{\theta}_\tau - \theta_\tau \right) \right)
$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau \left(\varepsilon_i - q_\tau(X_i, \widehat{\theta}_\tau) + q_\tau(X_i, \theta_\tau) - \frac{h(X_i)}{\sqrt{n}} \right)$$
$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_\tau(\varepsilon_i)\; q_\tau^{(1)}(X_i, \theta_\tau) + O_p \left(\left(\|\widehat{\theta} - \theta_\tau\| + n^{-1/2} \right)^{1/2} \log(n) \right).$$

so

$$\sqrt{n} \left(\widehat{\theta}_\tau - \theta_\tau \right) = S^{-1} \left[\sum_{i=1}^{n} \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) + \mathbb{E}[f(0|X) h(X) q_\tau^{(1)}(X, \theta_\tau)] \right] + o_p(1)$$

In order to prove (4.8), we would like to apply the following result obtained by He and Shao (1996)

$$\sup_{\|\gamma - \theta_\tau\| \le \delta_n} \Bigg| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(\psi_\tau(\varepsilon_i - l(X_i, \gamma))\; q_\tau^{(1)}(X_i, \gamma) - \psi_\tau(\varepsilon_i)\; q_\tau^{(1)}(X_i, \theta_\tau)$$
$$- \mathbb{E}_{e_i} \left[\psi_\tau\left(e_i - l(X_i, \gamma)\right)\; q_\tau^{(1)}(X_i, \gamma) - \psi_\tau\left(e_i\right)\; q_\tau^{(1)}(X_i, \theta_\tau) \right] \Big) \Bigg|$$
$$= O_p \left((\delta_n + n^{-1/2})^{1/2} \log(n) \right) \tag{4.9}$$

for any $\delta_n = o(1)$ as $n \to \infty$, where $l(x, \gamma) = q_\tau(x, \gamma) - q_\tau(x, \theta_\tau) - \frac{h(x)}{\sqrt{n}}$. To apply the previous result it would only be necessary to check that the quantity

$$\mathbb{E} \left[\left(\psi_\tau\left(e - l(X, \gamma)\right) - \psi_\tau\left(e\right) \right)^2 \; \|q_\tau^{(1)}(X, \gamma)\|^2 \right]$$

in bounded. Indeed,

$$\mathbb{E} \left[\left(\psi_\tau\left(e - l(X, \gamma)\right) - \psi_\tau\left(e\right) \right)^2 \; \|q_\tau^{(1)}(X, \gamma)\|^2 \right]$$
$$= \mathbb{E} \left[\left(\mathbb{I}\left(e < l(X, \gamma)\right) - \mathbb{I}\left(e < 0\right) \right)^2 \; \|q_\tau^{(1)}(X, \gamma)\|^2 \right]$$
$$\le \mathbb{E} \left[\left| F\left(l(X, \gamma)|X\right) - F\left(0|X\right) \right| \; \|q_\tau^{(1)}(X, \gamma)\|^2 \right]$$
$$\le c\, \mathbb{E} \left[f(0|X) \left(\|\gamma - \theta_\tau\| + \frac{|h(X)|}{\sqrt{n}} \right) \|A(x)\|^3 \right]$$

as consequence of the definition of ψ_τ and by differentiability properties of $F(\cdot|X)$ and q_τ. □

Then, we are able to state the representation of the empirical process R_n^1 under the composite null hypothesis. This result is given in the following theorem:

Theorem 4.4. *Let us assume that the data come from*

$$Y_i = q_\tau(X_i, \theta_\tau) + \varepsilon_i \qquad i \in \{1, \ldots, n\},$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent errors with conditional τ-th quantile equal to zero. Under the composite null hypothesis of a parametric model, $H_0 : q_\tau \in \mathcal{Q}_\theta$, and if conditions C1-C3 are verified, the following representation can be obtained:

$$R_n^1(\beta, u) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau(\varepsilon_i) \left[\mathbb{I}(\beta' X_i \le u) - S(\beta, u) S^{-1} \right] q_\tau^{(1)}(X_i, \theta_\tau) + o_p(1)$$

uniformly in (β, u)*, where the matrices* S *and* $S(\beta, u)$ *are defined by*

$$S = \mathbb{E}[f(0|X)\, q_\tau^{(1)}(X, \theta_\tau)\, q_\tau^{(1)}(X, \theta_\tau)']$$

$$S(\beta, u) = \mathbb{E}[f(0|X)\, q_\tau^{(1)}(X, \theta_\tau)\, q_\tau^{(1)}(X, \theta_\tau)'\, \mathbb{I}(\beta' X \le u)].$$

Proof. Note that the representation itself is different from that of He and Zhu (2003), because we do not assume homoscedasticity. Moreover, from this representation, the limit distribution of the test statistic, T_n, under the null hypothesis can be derived.

The proof of this result is consequence of the following representation of the empirical process R_n^1:

$$\begin{aligned}
R_n^1(\beta, u) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(r_i) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \mathbb{I}(\beta' X_i \le u) \\
&\overset{(i)}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \mathbb{I}(\beta' X_i \le u) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big\{ \Big[\psi_\tau(r_i) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) - \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \Big) \mathbb{I}(\beta' X_i \le u) \\
&- \mathbb{E}_{\varepsilon_i} \Big[\psi_\tau(r_i) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) - \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \Big] \mathbb{I}(\beta' X_i \le u) \Big\} \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big[\mathbb{E}_{\varepsilon_i}(\psi_\tau(r_i)) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \mathbb{I}(\beta' X_i \le u) \\
&- \mathbb{E}_{\varepsilon_i}(\psi_\tau(\varepsilon_i)) q_\tau^{(1)}(X_i, \theta_\tau) \mathbb{I}(\beta' X_i \le u) \Big] \\
&= R_n(\beta, u) + R_n^{1b}(\beta, u) + R_n^{1c}(\beta, u)
\end{aligned}$$

where in the equality (i) we have only added and subtracted the following quantities:

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \mathbb{I}(\beta' X_i \le u) \\
&\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{E}_{\varepsilon_i} \Big[\psi_\tau(r_i) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) - \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \Big] \mathbb{I}(\beta' X_i \le u).
\end{aligned}$$

Then, we have to analyse the processes R_n^{1b} and R_n^{1c}. On the one hand,

$$\begin{aligned}
R_n^{1b} =& \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big\{ \Big[\psi_\tau(r_i) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) - \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \Big) \mathbb{I}(\beta' X_i \le u) \\
&- \mathbb{E}_{\varepsilon_i} \Big[\psi_\tau(r_i) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) - \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \Big] \mathbb{I}(\beta' X_i \le u) \Big\} \\
&= O_p\left(\left(\|\widehat{\theta}_\tau - \theta_\tau\| + n^{-1/2} \right)^{1/2} \log(n) \right)
\end{aligned}$$

as a consequence of Lemma 4.1 of He and Shao (1996) that is detailed in (4.9). Note that in this case we take $h(X) = 0$.

On the other hand,

$$\begin{aligned}
R_n^{1c}(\beta,u) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big[\mathbb{E}_{\varepsilon_i}(\psi_\tau(r_i))q_\tau^{(1)}(X_i,\widehat{\theta}_\tau)\mathbb{I}(\beta' X_i \le u) \\
&\quad - \mathbb{E}_{\varepsilon_i}(\psi_\tau(\varepsilon_i))q_\tau^{(1)}(X_i,\theta_\tau)\mathbb{I}(\beta' X_i \le u)\Big] \\
&\overset{(ii)}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}_{\varepsilon_i}(\psi_\tau(r_i))\, q_\tau^{(1)}(X_i,\widehat{\theta}_\tau)\,\mathbb{I}(\beta' X_i \le u) \\
&\overset{(iii)}{=} -\sqrt{n}\,(\widehat{\theta}_\tau - \theta_\tau)\Big(\frac{1}{n}\sum_{i=1}^{n} f_i(0)q_\tau^{(1)}(X_i,\theta_\tau)q_\tau^{(1)}(X_i,\theta_\tau)'\mathbb{I}(\beta' X_i \le u) \\
&\quad + o_p\left(\sqrt{n}(\widehat{\theta}_\tau - \theta_\tau)\right)\Big) \\
&\overset{(iv)}{=} -S^{-1}\, S(\beta,u)\,\frac{1}{\sqrt{n}}\,\sum_{i=1}^{n}\psi_\tau(\varepsilon_i)\, q_\tau^{(1)}(X_i,\theta_\tau) + o_p(1)
\end{aligned}$$

where $f_i(x) = f_\varepsilon(x|X = X_i)$ and the equality (ii) is obvious if we take into account the definition of the function ψ_τ, that is,

$$\mathbb{E}_{\varepsilon_i}(\psi_\tau(\varepsilon_i)) = \mathbb{E}_{\varepsilon_i}(\tau\mathbb{I}(\varepsilon_i > 0) + (\tau-1)\mathbb{I}(\varepsilon_i < 0)) = \mathbb{P}(\varepsilon_i > 0) + \tau - 1 = 0$$

Furthermore, to obtain equality (iii) we have to use the differentiability of the conditional distribution of the error and the quantile regression function. Then,

$$\begin{aligned}
\mathbb{E}_{\varepsilon_i}(\psi_\tau(r_i)) &= \mathbb{P}_{\varepsilon_i}(r_i > 0) + \tau - 1 = \tau - \mathbb{P}_{\varepsilon_i}(Y_i - q_\tau(X_i,\widehat{\theta}_\tau) \le 0) \\
&= \tau - \mathbb{P}_{\varepsilon_i}(\varepsilon_i \le q_\tau(X_i,\widehat{\theta}_\tau) - q_\tau(X_i,\theta_\tau)) \\
&= \tau - F_i(q_\tau(X_i,\widehat{\theta}_\tau) - q_\tau(X_i,\theta_\tau)) \\
&= F_i(0) - F_i(q_\tau(X_i,\widehat{\theta}_\tau) - q_\tau(X_i,\theta_\tau)) \\
&= -f_i(\xi_{i,1})\left(q_\tau(X_i,\widehat{\theta}_\tau) - q_\tau(X_i,\theta_\tau)\right) \text{ with } \xi_{i,1} \text{ between } q_\tau(X_i,\widehat{\theta}_\tau) - q_\tau(X_i,\theta_\tau) \text{ and } 0 \\
&= -f_i(\xi_{i,1})q_\tau^{(1)}(X_i,\xi_2)\left(\widehat{\theta}_\tau - \theta_\tau\right) \text{ with } \xi_2 \text{ between } \widehat{\theta}_\tau \text{ and } \theta_\tau \\
&= -f_i(0)\, q_\tau^{(1)}(X_i,\theta_\tau)\left(\widehat{\theta}_\tau - \theta_\tau\right) + o_p\left(\sqrt{n}(\widehat{\theta}_\tau - \theta_\tau)\right).
\end{aligned}$$

Finally, equality (iv) is clear in view of Lemma 4.3. So, we conclude that

$$R_n^1(\beta,u) = n^{-1/2}\sum_{i=1}^{n}\psi_\tau(\varepsilon_i)\left[\mathbb{I}(\beta' X_i \le u) - S(\beta,u)S^{-1}\right] q_\tau^{(1)}(X_i,\theta_\tau) + o_p(1).$$

□

Now, the representation under the alternative is similar to the previous case, but a new term appears which will be crucial to prove the consistency of the test. Theorem 4.5 shows that the proposed test can detect local alternatives of order $n^{-1/2}$ from the null hypothesis.

Theorem 4.5. *Let us assume that the data come from*

$$Y_i = q_\tau(X_i, \theta_\tau) + n^{-1/2} h(X_i) + \varepsilon_i \qquad i \in \{1, \dots, n\},$$

where $\varepsilon_1, \dots, \varepsilon_n$ are independent errors with conditional τ-th quantile equal to zero. Under conditions C1-C3, the process R_n^1 allows the following representation:

$$\begin{aligned} R_n^1(\beta, u) = {} & n^{-1/2} \sum_{i=1}^n \psi_\tau(\varepsilon_i) \left[\mathbb{I}(\beta' X_i \le u) - S(\beta, u) S^{-1}\right] q_\tau^{(1)}(X_i, \theta_\tau) \\ & + \mathbb{E}\left[f(0|X) h(X) q_\tau^{(1)}(X, \theta_\tau) \mathbb{I}(\beta' X \le u)\right] \\ & - S(\beta, u) S^{-1} \mathbb{E}\left[f(0|X) h(X) q_\tau^{(1)}(X, \theta_\tau)\right] + o_p(1) \end{aligned}$$

uniformly in (β, u). The second and third addends of the right-hand side are constants reflecting the deviation from the null hypothesis.

Proof. In this case, we can write R_n^1 as follows

$$\begin{aligned} R_n^1(\beta, u) &= n^{-1/2} \sum_{i=1}^n \psi_\tau(r_i)\; q_\tau^{(1)}\left(X_i, \widehat{\theta}_\tau\right)\; \mathbb{I}\left(\beta' X_i \le u\right) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\psi_\tau\left(\varepsilon_i - l(X_i, \widehat{\theta}_\tau)\right) - \mathbb{E}_{e_i}\left[\psi_\tau\left(e_i - l(X_i, \widehat{\theta}_\tau)\right)\right]\right) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau) \mathbb{I}(\beta X_i \le u) \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\mathbb{E}_{e_i}\left[\psi_\tau\left(e_i - l(X_i, \widehat{\theta}_\tau)\right)\right]\right)\; q_\tau^{(1)}(X_i, \widehat{\theta}_\tau)\; \mathbb{I}(\beta X_i \le u) \\ &= R_n^{1d}(\beta, u) + R_n^{1e}(\beta, u) \end{aligned}$$

where $l(x, \widehat{\theta}_\tau) = q_\tau(x, \widehat{\theta}_\tau) - q_\tau(x, \theta_\tau) - \frac{h(x)}{\sqrt{n}}$ and the variables e_i have the same distribution as ε_i and let F_i be its distribution function. Then,

$$R_n^{1d}(\beta, u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(\varepsilon_i)\; q_\tau^{(1)}(X_i, \theta_\tau)\; \mathbb{I}(\beta' X_i \le u) + o_p(1)$$

as a consequence of the arguments employed to prove the equation (4.8) in Lemma 4.3. Moreover, expanding R_n^{1e} as R_n^{1c} we will have

$$\begin{aligned} -R_n^{1e}(\beta, u) &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\mathbb{E}_{e_i}\left[\psi_\tau\left(e_i - l(X_i, \widehat{\theta}_\tau)\right)\right]\right)\; q_\tau^{(1)}(X_i, \widehat{\theta}_\tau)\; \mathbb{I}(\beta X_i \le u) \\ &= \sqrt{n}\; (\widehat{\theta}_\tau - \theta_\tau)\; \frac{1}{n} \sum_{i=1}^n f_i(0) q_\tau^{(1)}(X_i, \theta_\tau)\; q_\tau^{(1)}(X_i, \theta_\tau)'\; \mathbb{I}(\beta' X_i \le u) \\ &- \frac{1}{n} \sum_{i=1}^n f_i(0) h(X_i) q_\tau^{(1)}(X_i, \theta_\tau) \mathbb{I}(\beta' X_i \le u) + o_p(1) \\ &= S^{-1}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) + \mathbb{E}(f(0|X) h(X) q_\tau^{(1)}(X, \theta_\tau))\right] S(\beta, u) \\ &- \mathbb{E}\left(f(0|X) h(X) q_\tau^{(1)}(X, \theta_\tau) \mathbb{I}(\beta' X \le u)\right) + o_p(1). \end{aligned}$$

Therefore,

$$\begin{aligned}R_n^1(\beta,u) &= n^{-1/2}\sum_{i=1}^{n}\psi_\tau(\varepsilon_i)\left[\mathbb{I}(\beta' X_i \le u) - S(\beta,u)S^{-1}\right] q_\tau^{(1)}(X_i,\theta_\tau)\\ &+ \mathbb{E}\left[f(0|X)h(X)q_\tau^{(1)}(X,\theta_\tau)\mathbb{I}(\beta' X \le u)\right]\\ &- S(\beta,u)S^{-1}\mathbb{E}\left[f(0|X)h(X)q_\tau^{(1)}(X,\theta_\tau)\right] + o_p(1).\end{aligned}$$

□

Corollary 4.6. *Under conditions C1-C3, if the data come from*

$$Y_i = q_\tau(X_i,\theta_\tau) + c_n n^{-1/2}h(X_i) + \varepsilon_i \qquad i \in \{1,\dots,n\},$$

where c_n is a sequence of real numbers converging to infinity (at any rate), then the test statistic, T_n, will converge to infinity and the power of the test will converge to one. To obtain this consistency, it is assumed that the sequence $q_\tau(x,\theta_\tau) + c_n n^{-1/2}h(x)$ does not coincide with any element of the parametric model, $\mathcal{Q}_\theta = \{q_\tau(\cdot,\theta) \ : \ \theta \in \Theta \subset \mathbb{R}^q\}$, and that $\mathbb{V}ar(f(0|X)h(X)q_\tau^{(1)}(X,\theta)) > 0$ for any θ.

4.2.3 Bootstrap approximation

The approximation of critical values is a crucial issue in lack-of-fit testing. One possible solution would be to use the limit distribution. However, this would require an estimate of the limit variance which involves the estimation of complicated unknown quantities. Furthermore, the convergence to the limit distribution could be slow. Another possibility could be to use the representations as given above. Then, a bootstrap method based on multipliers can be considered (see He and Zhu (2003)). The approximation by a multipliers bootstrap is generally better than the limit distribution, but still requires estimating many unknown quantities. He and Zhu (2003) assume homoscedasticity, so the conditional density of the error at zero, $f(0|X)$, does not have to be estimated. On the other hand, Escanciano and Goh (2014) allow for heteroscedasticity and use a multipliers bootstrap, which requires an estimate of the conditional density $f(0|X)$ by a smoothing method.

On the basis of previous statements, we propose a wild bootstrap approximation. The resampling process is the following one:

Step 1: Let us consider a parametric quantile regression model given by

$$Y = q_\tau(X,\theta_\tau) + \varepsilon. \tag{4.10}$$

Given $\{(X_1,Y_1),\cdots,(X_n,Y_n)\}$ a random sample of the variable $(X,Y) \in \mathbb{R}^{p+1}$, we will fit the model (4.10) and denote by $\widehat{\theta}_\tau$ an estimation of the parameter θ_τ and $r_i = Y_i - q_\tau(X_i,\widehat{\theta}_\tau)$ are the residuals from the original sample. Then, we could compute the test statistic as

$$T_n = \text{largest eigenvalue of } \int_\Pi R_n^1(\beta,u)[R_n^1(\beta,u)]' F_{n,\beta}(du)d\beta$$

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

$$R_n^1(\beta, u) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau \left(Y_i - q_\tau \left(X_i, \widehat{\theta}_\tau\right)\right) q_\tau^{(1)} \left(X_i, \widehat{\theta}_\tau\right) \mathbb{I}\left(\beta' X_i \leq u\right).$$

Step 2: Generate the multipliers, w_i, that are independently generated from a common distribution with τ-quantile equal to zero. Following the ideas developed in Section 2.2.4, we adopt the two-point distribution with probabilities $(1-\tau)$ and τ at $2(1-\tau)$ and -2τ, respectively.

Compute $\varepsilon_i^\star = w_i |r_i|$ where $|a|$ denotes the absolute value of the element a.

Step 3: Draw new bootstrap samples, denoted by $(X_1, Y_1^\star), \ldots, (X_n, Y_n^\star)$, where

$$Y_i^\star = q_\tau(X_i, \widehat{\theta}_\tau) + \varepsilon_i^\star \qquad i = 1, \ldots, n.$$

Step 4: Given $\{(X_1, Y_1^\star), \cdots, (X_n, Y_n^\star)\}$ the bootstrap sample, fit a parametric model type (4.10) and denote by $\widehat{\theta}_\tau^\star$ an estimation of the parameter $\widehat{\theta}_\tau$. Then, we could compute the bootstrap test statistic as

$$T_{n,b}^\star = \text{largest eigenvalue of } \int_\Pi R_n^{1\star}(\beta, u)[R_n^{1\star}(\beta, u)]' F_{n,\beta}(du) d\beta$$

where

$$R_n^{1\star}(\beta, u) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau \left(Y_i^\star - q_\tau \left(X_i, \widehat{\theta}_\tau^\star\right)\right) q_\tau^{(1)} \left(X_i, \widehat{\theta}^\star\right) \mathbb{I}\left(\beta' X_i \leq u\right).$$

Note that the empirical distribution function $F_{n,\beta}(u)$ does not need to be computed for each bootstrap sample because its elements only depends on the covariate of the quantile regression model and we are not bootstrapping the explanatory variable $X \in \mathbb{R}^d$. A complete discussion about computational aspects will be presented in Section 4.2.4.

Step 5: Repeat steps 2, 3 and 4 many times.

If a number, B, of bootstrap samples are generated, then the p-value of the test may be approximated by the proportion of bootstrap values not smaller than the original test statistic, that is,

$$\frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(T_n \leq T_{n,b}^\star).$$

The validity of this bootstrap mechanism comes from the representation of the process R_n^1 under the composite null hypothesis, in terms of the true errors plus the parameters estimation,

$$R_n^1(\beta, u) = n^{-1/2} \sum_{i=1}^{n} \psi_\tau(\varepsilon_i) q_\tau^{(1)}(X_i, \theta_\tau) \mathbb{I}(\beta' X_i \leq u) - S(\beta, u)\sqrt{n}\left(\widehat{\theta}_\tau - \theta_\tau\right) + o_p(1)$$

uniformly in (β, u). A similar representation can be derived for the bootstrap process conditionally on the original sample, where the convergence of the bootstrap version of the estimation error, $\sqrt{n}(\widehat{\theta}_\tau^\star - \widehat{\theta}_\tau)$, was established in Theorem 1 of Feng et al. (2011).

The main advantage of the proposed bootstrap approximation to calibrate the lack-of-fit test, in comparison to existing methods such as those proposed by He and Zhu (2003) and Escanciano and Goh (2014), is that it allows consideration of heteroscedastic regression models of any type without needing to estimate complicated quantities in the representations, and in particular without estimating the conditional density $f(0|X)$ by smoothing methods.

4.2.4 Computational aspects

Tests that face the curse of dimension usually require additional algorithms over other more common model checks. In particular, Escanciano (2006) and Stute et al. (2008) are based on Stute (1997)'s test and require additional computations over this original method. Similarly, Lavergne and Patilea (2008) present a test for high-dimensional covariates that is based on Zheng (1996)'s test, and requires an optimization algorithm over a set of Zheng-type statistics. The proposed method here is an adaptation of He and Zhu (2003)'s test to high-dimensional covariates with a procedure similar to that given by Escanciano (2006). One important virtue of this procedure is the ease of computation and that the amount of computations does not grow dramatically with the dimension of the covariate.

To illustrate this, recall that our test statistic, T_n, was defined in (4.4) as the largest eigenvalue of a Cramer-von-Mises norm of the process R_n^1. Following Escanciano (2006), one can show that T_n represents the largest eigenvalue of a matrix MT_n that can be expressed as follows

$$\begin{aligned}
MT_n &= \int_\Pi R_n^1(\beta,u)[R_n^1(\beta,u)]' F_{n,\beta}(du)d\beta \\
&= n^{-1} \sum_{i=1}^n \sum_{j=1}^n \psi_\tau(r_i)\psi_\tau(r_j) q_\tau^{(1)}(X_i,\widehat{\theta}_\tau) q_\tau^{(1)}(X_j,\widehat{\theta}_\tau)' \\
&\quad \times \int_\Pi \mathbb{I}(\beta' X_i \le u)\mathbb{I}(\beta' X_j \le u) F_{n,\beta}(du)d\beta \\
&= n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^n \psi_\tau(r_i)\psi_\tau(r_j) q_\tau^{(1)}(X_i,\widehat{\theta}_\tau) q_\tau^{(1)}(X_j,\widehat{\theta}_\tau)' \\
&\quad \times \int_{\mathbb{S}_d} \mathbb{I}(\beta' X_i \le \beta' X_r)\mathbb{I}(\beta' X_j \le \beta' X_r) d\beta \\
&= n^{-2} \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^n \psi_\tau(r_i)\psi_\tau(r_j) q_\tau^{(1)}(X_i,\widehat{\theta}_\tau) q_\tau^{(1)}(X_j,\widehat{\theta}_\tau)' A_{ijr} \\
&= n^{-2} \sum_{i=1}^n \sum_{j=1}^n \psi_\tau(r_i)\psi_\tau(r_j) q_\tau^{(1)}(X_i,\widehat{\theta}_\tau) q_\tau^{(1)}(X_j,\widehat{\theta}_\tau)' A_\bullet[i,j]
\end{aligned}$$

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

$$A_\bullet[i,j] = \sum_{r=1}^{n} A_{ijr} = \sum_{r=1}^{n} A_{ijr}^{(0)} \frac{\pi^{d/2-1}}{\Gamma\left(\frac{d}{2}+1\right)} \quad \text{with } i,j = 1, \cdots, n,$$

where $A_{ijr}^{(0)}$ is the complementary angle between the vectors $(X_i - X_r)$ and $(X_j - X_r)$ measured in radians, Γ is the gamma function, and d is the dimension of the covariate, X. That is

$$A_{ijr}^{(0)} = \left|\pi - arcos\left(\frac{(X_i - X_r)'(X_j - X_r)}{\|X_i - X_r\| \, \|X_j - X_r\|}\right)\right| = \begin{cases} \pi & \text{if } X_i = X_j \text{ and } X_i \neq X_r \\ 2\pi & \text{if } X_i = X_j = X_r \\ \pi & \text{if } X_i \neq X_j \text{ and } X_i = X_r \text{ or } X_j = X_r \end{cases}$$

We also have a symmetric property, $A_{ijr} = A_{jir}$, which simplifies the evaluation of the test statistic T_n, and it involves that $A_\bullet$ is a symmetric matrix. This fact improves drastically the time of computation of the test statistic and allows to apply the test to larger datasets. Thus, the total number of computations required to obtain the test statistic depends on the dimension, d, only at a linear rate, which is the same rate required by He and Zhu (2003)'s test, and much less than the optimization in d dimensions required by other methods in the literature. Moreover, the symmetric property means that the memory required for storing the matrix $A_\bullet$ is substantially lower and drops to $\frac{n(n+1)}{2}$ elements, against n^2. All these computational properties are particularly useful in the case of high-dimensional or functional covariates, see García-Portugués et al. (2014) for an illustration in the mean regression functional context.

Note also that the matrix $A_\bullet$, which is the most expensive in computation time, does not need to be computed for each bootstrap sample because its elements only depends on the covariate of the quantile regression model that are not modified along the bootstrap procedure. So, the bootstrap test statistic presented in the Section 4.2.3 can be express as

$$T_{n,b}^\star = \text{largest eigenvalue of } \; n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_\tau(r_i^\star)\psi_\tau(r_j^\star) q_\tau^{(1)}(X_i, \widehat{\theta}_\tau^\star) q_\tau^{(1)}(X_j, \widehat{\theta}_\tau^\star)' A_\bullet[i,j].$$

Furthermore, in order to compute the test statistic defined in (4.4), we have used the free software R (http://cran.r-project.org/). Anyway, the calculus associated with the matrix $A_\bullet$ and the empirical process R_n^1 have been programmed in Fortran language to speed up computation time.

Table 4.1 shows the mean of the times required by 1000 original samples with $B = 500$ bootstrap replications, in units of seconds per original sample. The data are drawn from Model 4.8, whose details are given in the next section, and the sample size is $n = 100$. The dimension of the covariate is $d = t + 2$. As expected, the new test requires more computations than He and Zhu (2003)'s test, but the differences are quite small, and the amount of computations does not dramatically grow with the dimension, even for very large dimensions. The gain of power from the new test, shown in the next section, justifies the small increase in the computation time.

	$t=0$	$t=2$	$t=6$	$t=10$	$t=20$	$t=30$	$t=40$	$t=50$
Proposed test	2.76	2.84	2.85	2.91	2.91	3.10	3.20	3.38
HZ test	2.71	2.51	2.81	2.56	2.92	2.83	2.85	2.77

Table 4.1: Computational times (seconds per sample) associated with our proposed lack-of fit test (Proposed test) and with the test proposed by He and Zhu (2003) (HZ test) as a function of the dimension $(t+2)$ of the covariate.

4.3 Simulation study

We study the performance of our proposed method under the null and the alternative hypotheses using a Monte Carlo simulation study. In all experiments, the number of simulated original samples was 1000, the number of bootstrap replications $B = 500$, and the multipliers for the bootstrap approximation followed the two-point distribution given in Section 4.2.3.

We first focus on the behaviour under the null hypothesis, in order to check the adjustment of the significance level. We simulate values for the following quantile regression models with $\tau = 0.5$:

Model 4.1: $Y = 1 + X^{(1)} + X^{(2)} + \varepsilon,$

Model 4.2: $Y = 1 + X^{(1)} + X^{(2)} + X^{(3)} + X^{(4)} + X^{(5)} + \varepsilon,$

Model 4.3: $Y = 1 + X^{(1)} + X^{(2)} + f(X^{(1)})\varepsilon,$

where $X^{(i)}$ follows an uniform distribution on the interval $(0,1)$ for $i = \{1,\cdots,5\}$, and they are mutually independent; $f(x) = x + 0.5$ and ε is the unknown error that follows a standard Gaussian distribution, which is drawn independently of the covariates. In Models 4.1 and 4.3 the null hypothesis is the linear model in $X^{(1)}$ and $X^{(2)}$ versus an alternative that includes any dependence of Y on $X^{(1)}$ and $X^{(2)}$. In Model 4.2 the null hypothesis is the linear model in the five explanatory variables versus any dependence on them. Model 4.1 represents a common homoscedastic model with small dimension of the covariate. Model 4.2 is intended to show the possible effect of a larger dimension on the significant level. Model 4.3 is useful to show the possible effect of heteroscedasticity on the significant level.

Table 4.2 shows the proportions of rejections associated with different sample sizes, n, and for different nominal significance levels, α. The proposed test works well in a homoscedastic context (Models 4.1 and 4.2) as well as in a heteroscedastic context (Model 4.3) even for small sample sizes. Comparing Models 4.1 and 4.2, the increase of the dimension of the explanatory variables does not have a negative impact on the adjustment of the significance level of the test. These are important, because our bootstrap mechanism was designed to work under heteroscedastic models and the aim of the test itself was to be applied for larger dimensions of the covariate.

Table 4.3 provides the same proportions of rejections for different error distributions and quantiles, restricted to Model 4.1 and nominal level $\alpha = 0.05$. The error distributions are centred standard normal, centred log-normal, and centred exponential with expectation one. That is, $\varepsilon = Z - z_\tau$, where Z follows a standard Gaussian, standard log-normal, and exponential with expectation one, respectively, and z_τ is the τ-quantile of the Z-distribution.

		$n = 25$	$n = 50$	$n = 100$	$n = 150$	$n = 200$
Model 4.1	$\alpha = 0.10$	0.096	0.112	0.102	0.089	0.100
	$\alpha = 0.05$	0.049	0.047	0.058	0.048	0.048
	$\alpha = 0.01$	0.002	0.008	0.016	0.007	0.010
Model 4.2	$\alpha = 0.10$	0.119	0.112	0.094	0.104	0.106
	$\alpha = 0.05$	0.066	0.053	0.047	0.056	0.049
	$\alpha = 0.01$	0.017	0.014	0.011	0.014	0.010
Model 4.3	$\alpha = 0.10$	0.107	0.099	0.107	0.096	0.100
	$\alpha = 0.05$	0.061	0.045	0.049	0.055	0.054
	$\alpha = 0.01$	0.014	0.005	0.010	0.015	0.015

Table 4.2: Proportions of rejections associated with our proposed lack-of-fit test for Models 4.1, 4.2 and 4.3.

The nominal level is respected under the null hypothesis for all the error distributions considered and orders of the quantile.

We now study the performance of the new test under the alternative. To this end, the new test will be compared with that of He and Zhu (2003). Before doing so, we must remember that He and Zhu (2003) suggested a bootstrap calibration of their test based on an asymptotic representation of the empirical process in a homoscedastic scene. We will verify if this manner of calibrating the test allows a good fit to the significance level for heteroscedastic models. We simulate values of the following regression model with $\tau = 0.5$ under the null hypothesis of linearity:

$$\textbf{Model 4.4:} \quad Y = 1 + X^{(1)} + f(X^{(1)})\varepsilon,$$

where $X^{(1)}$ follows a uniform distribution on the interval $(0, 1)$, $f(x) = x + 0.5$, ε follows a standard normal, and $X^{(1)}$ and ε are independent.

The proportions of rejections associated with the test proposed by He and Zhu (2003) are shown in Table 4.4 for different sample sizes and nominal significance levels. The bootstrap method proposed by He and Zhu (2003) does not work well in a heteroscedastic context. This is due to their representation being only valid under homoscedasticity. However, the proposed bootstrap (Section 4.2.3) works well for their test also under heteroscedasticity. Therefore, with the aim to make a fair comparison between our proposal and He and Zhu (2003)'s test, subsequently we use a wild bootstrap as given in Section 4.2.3 to calibrate both lack-of-fit tests (that will be denoted by HZ test).

Once the adjustment of the level of both lack-of-fit tests has been studied, we analyse their performance under the alternative hypothesis. Consider the following regression model associated with quantiles of different orders, τ:

$$\textbf{Model 4.5:} \quad Y = 1 + \frac{1}{5}\left(X^{(1)} - X^{(2)}\right) + \varepsilon_\tau,$$

where $X^{(1)}, X^{(2)}$ follows a standard normal and they are independent, and $\varepsilon = Z - z_\tau$, where z_τ is the τ-quantile of the variable Z. Z is drawn independently of $X^{(1)}$ and $X^{(2)}$. Three possibilities are considered for the distribution of Z: standard normal, uniform on the interval $(-1, 1)$, and chi-squared with four degrees of freedom.

		$\tau = 0.10$	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$
$\varepsilon \in$ Centred Standard Normal	$n = 25$	0.043	0.053	0.049	0.051	0.070
	$n = 50$	0.048	0.061	0.047	0.063	0.043
	$n = 100$	0.049	0.060	0.058	0.047	0.049
	$n = 150$	0.052	0.053	0.048	0.057	0.051
	$n = 200$	0.049	0.056	0.048	0.045	0.058
$\varepsilon \in$ Centred Log-Normal	$n = 25$	0.043	0.057	0.067	0.057	0.072
	$n = 50$	0.051	0.047	0.057	0.058	0.053
	$n = 100$	0.053	0.052	0.052	0.037	0.041
	$n = 150$	0.059	0.057	0.063	0.048	0.058
	$n = 200$	0.050	0.059	0.057	0.044	0.057
$\varepsilon \in$ Centred Exponential	$n = 25$	0.048	0.059	0.056	0.056	0.071
	$n = 50$	0.058	0.054	0.057	0.048	0.042
	$n = 100$	0.052	0.047	0.045	0.044	0.054
	$n = 150$	0.048	0.064	0.053	0.062	0.061
	$n = 200$	0.034	0.057	0.051	0.048	0.057

Table 4.3: Proportions of rejections associated with our lack-of-fit test for Model 4.1, for different error distributions and different quantiles, with nominal level $\alpha = 0.05$.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

	Wild bootstrap of Section 4.2.3			Bootstrap proposed in He and Zhu (2003)		
	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$n = 25$	0.103	0.057	0.014	0.441	0.305	0.142
$n = 50$	0.116	0.064	0.015	0.263	0.164	0.067
$n = 100$	0.094	0.051	0.013	0.167	0.092	0.033
$n = 150$	0.104	0.051	0.010	0.155	0.085	0.025
$n = 200$	0.103	0.051	0.014	0.136	0.080	0.026

Table 4.4: Proportions of rejections associated with the test proposed by He and Zhu (2003) for the heteroscedastic Model 4.4 with two types of bootstrap approximations.

		Proposed test			HZ test		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$Z \in N(0,1)$	$\tau = 0.10$	0.346	0.229	0.092	0.183	0.094	0.023
	$\tau = 0.25$	0.498	0.362	0.180	0.210	0.121	0.030
	$\tau = 0.50$	0.575	0.444	0.231	0.208	0.110	0.032
	$\tau = 0.75$	0.487	0.377	0.200	0.191	0.096	0.016
	$\tau = 0.90$	0.357	0.245	0.102	0.128	0.052	0.007
$Z \in \text{Uniform}(-1,1)$	$\tau = 0.10$	0.930	0.885	0.707	0.524	0.335	0.112
	$\tau = 0.25$	0.866	0.789	0.593	0.397	0.242	0.066
	$\tau = 0.50$	0.809	0.691	0.475	0.325	0.186	0.056
	$\tau = 0.75$	0.877	0.795	0.587	0.381	0.229	0.054
	$\tau = 0.90$	0.945	0.872	0.693	0.382	0.193	0.027
$Z \in \chi^2_4$	$\tau = 0.10$	0.315	0.207	0.076	0.144	0.078	0.018
	$\tau = 0.25$	0.245	0.144	0.045	0.124	0.056	0.015
	$\tau = 0.50$	0.208	0.124	0.041	0.112	0.058	0.012
	$\tau = 0.75$	0.141	0.070	0.022	0.115	0.058	0.017
	$\tau = 0.90$	0.137	0.077	0.028	0.120	0.064	0.015

Table 4.5: Proportions of rejections associated with our proposed lack-of-fit test (Proposed test) and with the test proposed by He and Zhu (2003) (HZ test) for Model 4.5.

Table 4.5 shows the proportions of rejections for several quantiles and the three error distributions, when the tests are applied to check the no-effect model, i.e., to check the null hypothesis that the quantile regression function is a constant not depending on the covariates. The sample size is fixed to $n = 100$. We consider a relatively simple hypothesis and a simple deviation under the alternative, to facilitate the comparison between quantiles of different orders, and to evaluate the effect of the error distribution.

The proposed test is more powerful than He and Zhu (2003)'s test for any of the quantiles and for the three error distributions. The power of the proposed test is symmetric with respect to the order of the quantile around 0.5 for the symmetric error distributions, which are the standard normal and the uniform in Table 4.5. For the standard normal error distribution, the proposed test is more powerful for the central quantiles (around 0.5), which can be explained by the higher density at these quantiles. For the uniform error distribution, the density is constant with respect to the quantile, while the factor $\tau(1-\tau)$ appearing in the asymptotic distribution of the proposed test makes the test more powerful for the external quantiles (with orders close to 0 or 1). For the chi-squared error distribution, the proposed test is more powerful for the quantiles with smaller order, since the error distribution is asymmetric with higher density at these quantiles.

We now consider a linear model under the null hypothesis and a quadratic deviation under the alternative. The deviation is multiplied by a value $c > 0$, to evaluate the effect of the deviation on the power of the test.

$$\textbf{Model 4.6:} \quad Y = 1 + X^{(1)} + X^{(2)} + c\left((X^{(1)})^2 + (X^{(2)})^2 + X^{(1)}X^{(2)}\right) + \varepsilon_\tau,$$

where $X^{(1)}$ follows a uniform distribution on the interval $(0,1)$, $X^{(2)}$ follows a standard Gaussian distribution; and ε_τ is a log-normal distribution centred to the quantile τ, i.e., $\varepsilon_\tau = e^Z - e^{z_\tau}$, where Z denotes a standard Gaussian distribution and z_τ are the τ-quantile of the variable Z; and $X^{(1)}$, $X^{(2)}$ and ε_τ are drawn independently.

Firstly we are going to begin with the case in which $\tau = 0.5$ and $c = 1/3$ in order to check the influences of the conditional density of the error evaluated at zero on the performance of the tests. Several possibilities for the error distribution will be considered, satisfying $P(\varepsilon < 0|X) = \tau = 0.5$. All of them will have the same variance in order that the power comparison is only affected by the conditional density at zero. Results are given in Table 4.6. We conclude that the power of both tests is higher when the conditional density of the error evaluated at zero is larger (the Student's t distribution) and lower for smaller density at zero (the uniform distribution). On the other hand, we notice that our proposal is more powerful than the test described by He and Zhu (2003) regardless of the error distribution.

		Proposed test			HZ test		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$\varepsilon \in U(-\sqrt{3}, \sqrt{3})$	$n = 25$	0.192	0.124	0.029	0.170	0.091	0.027
	$n = 50$	0.459	0.344	0.165	0.340	0.225	0.090
	$n = 100$	0.774	0.675	0.472	0.564	0.450	0.240
	$n = 150$	0.930	0.884	0.712	0.765	0.667	0.433
	$n = 200$	0.978	0.950	0.851	0.870	0.776	0.562
$\varepsilon \in N(0,1)$	$n = 25$	0.214	0.122	0.037	0.185	0.108	0.028
	$n = 50$	0.586	0.464	0.228	0.401	0.281	0.100
	$n = 100$	0.891	0.810	0.597	0.709	0.592	0.334
	$n = 150$	0.971	0.945	0.842	0.878	0.788	0.568
	$n = 200$	0.996	0.990	0.949	0.955	0.924	0.775
$\varepsilon \in \frac{1}{\sqrt{2}} t_4$	$n = 25$	0.264	0.161	0.046	0.203	0.121	0.042
	$n = 50$	0.666	0.561	0.338	0.450	0.335	0.146
	$n = 100$	0.965	0.941	0.794	0.828	0.708	0.473
	$n = 150$	0.993	0.991	0.958	0.959	0.916	0.765
	$n = 200$	0.999	0.999	0.990	0.982	0.967	0.897

Table 4.6: Proportion of rejections associated with our lack-of-fit test (under the title The new test) and the test proposed by He and Zhu (2003) (under the title HZ test) for Model 4.6 with parameter $c = 1/3$.

Moreover, Figure 4.1 shows the powers of the proposed test and He and Zhu (2003)'s test as functions of the value of c, and with five orders of the quantile: 0.1, 0.25, 0.5, 0.75, and 0.9. The nominal level is $\alpha = 0.05$ and the sample size is fixed to $n = 150$. As expected, the power increases with c. The new test is more powerful than He and Zhu (2003)'s test for any value of c and for any of the considered quantiles. Both tests are more powerful for central quantiles (orders close to 0.5). Symmetry around 0.5 is not strictly satisfied, since the error distribution is not symmetric around the median, and the deviation from the null hypothesis is more complex than that given in Model 4.5.

We consider different deviations from the linear null hypothesis and error distributions, as Model 4.7.

$$\textbf{Model 4.7:} \quad Y = 1 + X^{(1)} + X^{(2)} + h(X^{(1)}, X^{(2)}) + \varepsilon,$$

where $X^{(1)}$ follows a uniform distribution on the interval $(0, 1)$, $X^{(2)}$ follows a standar Gaussian distribution; and $\varepsilon = Z - z_\tau$, with z_τ being the τ-quantile of the variable Z; and $X^{(1)}$, $X^{(2)}$, and Z are drawn independently. For the deviation $h(X)$, a quadratic function including interaction is considered, as well as a sinus, exponential, and logarithm function of the linear transformation $l(x) = 1 + x_1 + x_2$ (see Table 4.7). For the distribution of Z, the log-normal, chi-squared with two degrees of freedom, exponential with expectation one, and a mixture of normal distributions are considered. The mixture is obtained as a standard normal with probability 0.75 and a normal distribution with mean 5 and standard deviation 2 with probability 0.25.

The proposed test and He and Zhu (2003)'s test are applied to check the null hypothesis of linearity on $X^{(1)}$ and $X^{(2)}$ with nominal level $\alpha = 0.05$. Results for the proportions of rejections are given in Table 4.7. For each deviation and each error distribution, the proposed test is more powerful than He and Zhu (2003)'s.

Our main purpose in proposing a new lack-of-fit test was to overcome the curse of dimensionality. Thus, the new test should show an acceptable power for increasing dimensionality of the covariate. To check this, we simulate values of the following median regression model:

$$\textbf{Model 4.8:} \quad Y = 1 + X^{(1)} + X^{(2)} + \frac{1}{3}\Big((X^{(1)})^2 + X^{(1)}X^{(2)} + (X^{(2)})^2\Big) + \varepsilon,$$

where our goal is to realize the following lack-of-fit test:

$$\begin{cases} H_0 : Y = \theta_0 + \theta_1 X^{(1)} + \theta_2 X^{(2)} + \varepsilon \\ \\ H_a : Y = q_\tau(X^{(1)}, X^{(2)}, X^{(2+1)}, \ldots, X^{(2+t)}) + \varepsilon, \end{cases}$$

where $X^{(i)}$ follows a uniform distribution on the interval $(0, 1)$ if i is odd, and $X^{(i)}$ follows a standard Gaussian distribution if i is even; the error is drawn from the centred log-normal distribution, i.e., $\varepsilon = e^Z - 1$ where Z denotes a a standard Gaussian distribution; q_τ is any smooth (nonparametric) function of the covariates; and t represents the number of additional covariates in the alternative, and so is the additional dimension where the test is looking for deviations from the null. It would be expected that increased value of t implies decreased power of the test.

Table 4.8 shows the proportions of rejections associated with the new test and He and Zhu (2003)'s test, for different values of the additional dimension, t. Both tests suffer a loss of power due to the increase of the dimension, as expected. Nonetheless, the loss of power is more pronounced for the test proposed by He and Zhu (2003). For example, from dimension $t = 6$ the proportion of rejections associated with their test is near to the significance level, whereas our proposed test preserves noticeable power, even for very high dimensions.

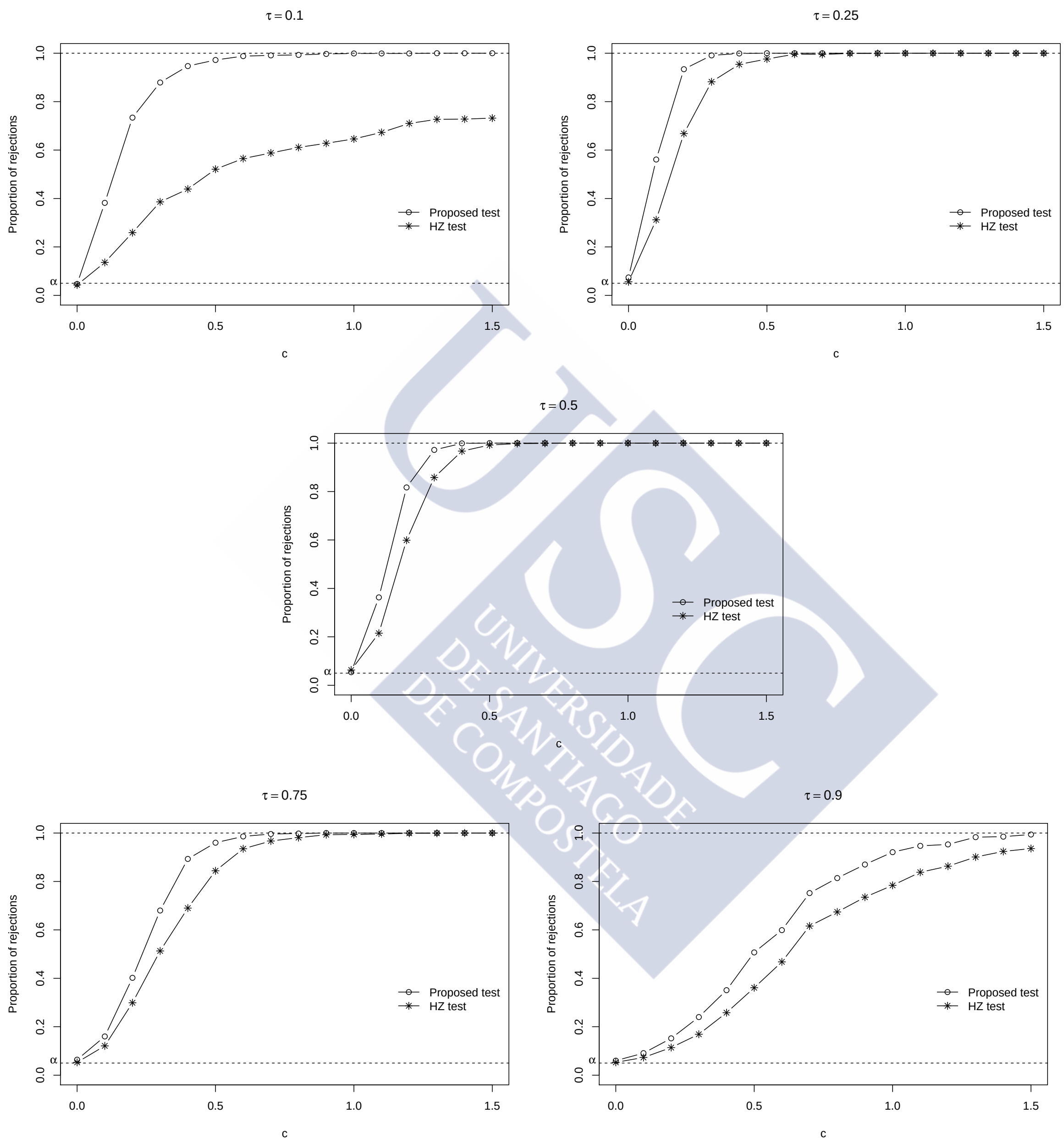


Figure 4.1: Proportion of rejections associated with our proposed lack-of-fit test (Proposed test) and the test proposed by He and Zhu (2003) (HZ test) for Model 4.6 depending on the parameter c and the τ-quantile of interest.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Note that, for very high dimensions, He and Zhu (2003)'s test statistic is almost degenerate, because for any observation of the covariate, X_i, the indicators $\mathbb{I}(X_j \leq X_i)$, involved in the computation of their test process at X_i, will be zero for most of the other observations X_j, when the dimension of the covariates X_i and X_j is large. Thus, the test is unable to make a reasonable number of evaluations to check the model, and its power is consequently destroyed, as observed in Table 4.8 for $t > 10$.

			$Z \in e^{N(0,1)}$		$Z \in \chi^2_2$		$Z \in Exp(1)$		$Z \in$ Mixture	
			Proposed	HZ	Proposed	HZ	Proposed	HZ	Proposed	HZ
$h(x) = \frac{1}{3}(x_1^2 + x_2^2 + x_1 x_2)$	$n = 50$	$\tau = 0.25$	0.373	0.162	0.199	0.097	0.448	0.184	0.135	0.083
		$\tau = 0.5$	0.577	0.364	0.345	0.208	0.696	0.435	0.287	0.175
		$\tau = 0.75$	0.309	0.217	0.200	0.150	0.490	0.365	0.074	0.068
	$n = 150$	$\tau = 0.25$	0.994	0.910	0.934	0.705	1.000	0.962	0.702	0.386
		$\tau = 0.5$	0.981	0.899	0.849	0.619	0.999	0.976	0.829	0.617
		$\tau = 0.75$	0.773	0.579	0.516	0.361	0.952	0.831	0.138	0.112
$h(x) = 5\, sin(0.6\, \pi\, l(x))$	$n = 50$	$\tau = 0.25$	0.443	0.409	0.425	0.414	0.461	0.429	0.381	0.356
		$\tau = 0.5$	0.562	0.321	0.458	0.270	0.607	0.353	0.390	0.238
		$\tau = 0.75$	0.124	0.066	0.106	0.061	0.157	0.081	0.103	0.053
	$n = 150$	$\tau = 0.25$	1.000	0.996	1.000	0.990	1.000	0.999	0.998	0.985
		$\tau = 0.5$	1.000	0.997	1.000	0.998	1.000	1.000	1.000	0.957
		$\tau = 0.75$	0.865	0.419	0.811	0.380	0.982	0.637	0.586	0.228
$h(x) = 8\, exp(-0.5\, l(x))$	$n = 50$	$\tau = 0.25$	0.190	0.113	0.154	0.112	0.169	0.135	0.133	0.109
		$\tau = 0.5$	0.411	0.251	0.254	0.167	0.483	0.268	0.225	0.161
		$\tau = 0.75$	0.244	0.145	0.164	0.097	0.382	0.251	0.102	0.089
	$n = 150$	$\tau = 0.25$	0.917	0.498	0.788	0.378	0.963	0.577	0.533	0.281
		$\tau = 0.5$	0.980	0.747	0.797	0.455	0.998	0.868	0.759	0.458
		$\tau = 0.75$	0.700	0.450	0.493	0.325	0.955	0.744	0.216	0.137
$h(x) = 6\, \log \lvert l(x) \rvert$	$n = 50$	$\tau = 0.25$	0.820	0.627	0.736	0.570	0.874	0.678	0.622	0.483
		$\tau = 0.5$	0.396	0.306	0.291	0.200	0.561	0.398	0.227	0.183
		$\tau = 0.75$	0.090	0.094	0.098	0.086	0.122	0.104	0.068	0.074
	$n = 150$	$\tau = 0.25$	1.000	0.998	0.999	0.987	1.000	1.000	0.997	0.973
		$\tau = 0.5$	0.897	0.757	0.751	0.558	0.971	0.875	0.660	0.471
		$\tau = 0.75$	0.166	0.171	0.167	0.166	0.297	0.196	0.112	0.138

Table 4.7: Proportions of rejections associated with our proposed lack-of-fit test (Proposed) and to the test proposed by He and Zhu (2003) (HZ) for Model 4.7.

On the other hand, our proposed method is able to make comparisons even for large dimensions of the covariate, because the indicators are calculated with unidimensional projections of the covariate. We conclude that the proposed method constitutes a necessary modification of He and Zhu (2003) when the dimension of the covariate is large.

		Proposed test			HZ test		
		$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
$t = 0$	$n = 25$	0.252	0.154	0.057	0.225	0.135	0.035
	$n = 50$	0.675	0.564	0.361	0.487	0.357	0.163
	$n = 100$	0.961	0.918	0.776	0.822	0.725	0.460
	$n = 150$	0.993	0.983	0.943	0.949	0.903	0.751
	$n = 200$	0.999	0.998	0.990	0.982	0.965	0.897
$t = 2$	$n = 25$	0.177	0.100	0.029	0.143	0.080	0.021
	$n = 50$	0.507	0.391	0.186	0.215	0.117	0.040
	$n = 100$	0.868	0.813	0.638	0.349	0.228	0.077
	$n = 150$	0.978	0.957	0.869	0.506	0.355	0.163
	$n = 200$	0.997	0.993	0.975	0.636	0.498	0.263
$t = 6$	$n = 25$	0.133	0.055	0.010	0.054	0.018	0.004
	$n = 50$	0.345	0.244	0.097	0.098	0.051	0.010
	$n = 100$	0.797	0.696	0.501	0.097	0.056	0.021
	$n = 150$	0.935	0.901	0.768	0.151	0.083	0.027
	$n = 200$	0.992	0.978	0.929	0.177	0.089	0.029
$t = 10$	$n = 25$	0.120	0.057	0.011	0.066	0.018	0.005
	$n = 50$	0.267	0.161	0.056	0.043	0.028	0.004
	$n = 100$	0.659	0.562	0.366	0.052	0.025	0.003
	$n = 150$	0.884	0.830	0.672	0.071	0.036	0.006
	$n = 200$	0.966	0.946	0.887	0.085	0.040	0.008
$t = 20$	$n = 25$	0.094	0.042	0.010	0.065	0.023	0.010
	$n = 50$	0.174	0.098	0.019	0.055	0.024	0.007
	$n = 100$	0.520	0.398	0.235	0.054	0.028	0.005
	$n = 150$	0.800	0.707	0.525	0.000	0.004	0.003
	$n = 200$	0.918	0.876	0.748	0.050	0.033	0.008
$t = 50$	$n = 25$	0.074	0.044	0.005	0.050	0.026	0.007
	$n = 50$	0.111	0.059	0.014	0.074	0.036	0.009
	$n = 100$	0.237	0.149	0.041	0.068	0.034	0.007
	$n = 150$	0.492	0.374	0.188	0.001	0.005	0.005
	$n = 200$	0.686	0.600	0.438	0.063	0.024	0.009

Table 4.8: Proportions of rejections associated with our proposed lack-of-fit test (Proposed test) and the test proposed by He and Zhu (2003) (HZ test) for Model 4.8.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

4.4 Application to real data

The proposed method is applied to real data from the evolution of the Gross Domestic Product (GDP) in several countries. GDP is an economic indicator that reflects the monetary value of the goods and final services produced by an economy in a certain period and it is used as a measure of the material well-being of a society. Different median regression models have been proposed to explain the annual growth rate of the Per Capita GDP in terms of a number of explanatory variables, including the initial Per Capita GDP and diverse economic and social indicators.

We focus on the model of Koenker and Machado (1999), based on the available information included in Barro and Lee (1994). A complete study of this economic model is given by Barro and Sala-i Martin (1995). The aim of Koenker and Machado (1999) was to check the combined effect of the different explanatory variables on the response in a quantile regression model. Here we test the specification of the quantile regression model itself.

The dataset **barro** that we use is available in the **R** package **quantreg**, (http://cran.at.r-project.org/web/packages/quantreg/). This data set contains measurements associated with 71 countries during the period 1965-1975 and 90 countries during the period 1975-1985, yielding a total sample size of $n = 161$ countries.

The explanatory variables used to explain the median of the annual growth of the Per Capita GDP (the response variable, Y) can be split in two groups as given below. More details about these variables and their role in the model for GDP can be found in Barro and Sala-i Martin (1995).

State variables: These variables reflect characteristics of the different countries that cannot be directly decided by political or social agents. They are measures of the steady-state position of the country, such as human capital, education or health. Koenker and Machado (1999) consider the following variables in this group:

$X^{(1)}$:= log(Initial Per Capita GDP)
$X^{(2)}$:= Male Secondary Education
$X^{(3)}$:= Female Secondary Education
$X^{(4)}$:= Female Higher Education
$X^{(5)}$:= Male Higher Education
$X^{(6)}$:= Life Expectancy
$X^{(7)}$:= Human Capital

Control and environmental variables: These variables are direct consequences of decisions made by government or private agents. The variables included in this group are

$X^{(8)}$:= Education/GDP
$X^{(9)}$:= Investment/GDP
$X^{(10)}$:= Public Consumption/GDP
$X^{(11)}$:= Black Market Premium
$X^{(12)}$:= Political Instability

$X^{(13)}$:= Growth Rate Terms Trade

We apply the AIC criterion proposed by Hurvich and Tsai (1990) to variable selection among the thirteen explanatory variables for the quantile regression model. We will consider only those variables that show as relevant for the response. Given a quantile regression model $Y = q_\tau(X, \theta_\tau) + \varepsilon$, Hurvich and Tsai (1990) define the AIC criterion as

$$AIC = n \log \left(\frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - q_\tau(X_i, \widehat{\theta}_\tau)) \right) + q$$

where q represents the number of parameter of the model and $\widehat{\theta}_\tau$ is a estimation of the parameter θ_τ provided a random sample, $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$, of the variable $(X, Y) \in \mathbb{R}^{d+1}$. Based on this criterion, we propose a model that includes the variables X_i with $i \in \mathcal{I}_1 = \{1, 2, 6, 7, 9, 10, 11, 12, 13\}$.

We apply the proposed lack-of-fit test in four different testing problems:

$$\text{Problem 1} \begin{cases} H_0 : Y = \theta_0 + \sum_{i=1}^{13} \theta_i' X^{(i)} + \varepsilon_1 \\ H_a : Y = q_\tau(X^{(1)}, X^{(2)}, \ldots, X^{(13)}) + \varepsilon_1 \end{cases}$$

$$\text{Problem 2} \begin{cases} H_0 : Y = \theta_0 + \sum_{i \in \mathcal{I}_1} \theta_i' X^{(i)} + \varepsilon_2 \\ H_a : Y = q_\tau(\{X^{(i)}, i \in \mathcal{I}_1\}) + \varepsilon_2 \end{cases}$$

$$\text{Problem 3} \begin{cases} H_0 : Y = \theta_0 + \sum_{i \in \mathcal{I}_1} \theta_i' X^{(i)} + \varepsilon_3 \\ H_a : Y = q_\tau(X^{(1)}, X^{(2)}, \ldots, X^{(13)}) + \varepsilon_3 \end{cases}$$

$$\text{Problem 4} \begin{cases} H_0 : Y = \theta_0 + \sum_{i \in \mathcal{I}_2} \theta_i' X^{(i)} + \varepsilon_4 \\ H_a : Y = q_\tau(X^{(1)}, X^{(2)}, \ldots, X^{(13)}) + \varepsilon_4 \end{cases}$$

where $\mathcal{I}_2 = \{1, 2, 3, 4, 5, 6, 7\}$ (state variables). Problem 1 is a lack-of-fit test of the linear model versus a nonparametric alternative, including all the thirteen explanatory variables under both the null and alternative hypotheses. Problem 2 is a lack-of-fit test of the linear model versus a nonparametric alternative, including only the nine variables in the set $\mathcal{I}_1$. Problem 3 is the same test as Problem 2, but with an alternative in the thirteen original variables. Problem 4 is a lack-of-fit test of a linear model that only includes the state variables.

Table 4.9 contains the p-values obtained from the application of the proposed lack-of-fit test to each of the testing problems. The number of bootstrap replications was $B = 500$. We would accept the null hypothesis in Problems 1, 2 and 3. In Problem 3, the model under the null is the simplest, while the model under the alternative is the most complex. Despite this, the p-value is quite large, so we can conclude that the simple model with the nine explanatory variables in the set $\mathcal{I}_1$ is correct, and there is no significant deviation from this model arising from any (smooth) function of the thirteen possible explanatory variables.

	Problem 1	Problem 2	Problem 3	Problem 4
p-values	0.194	0.458	0.440	0.002

Table 4.9: p-values obtained by the proposed lack-of-fit test for Problems 1, 2, 3 and 4.

On the other hand, the null hypothesis is rejected for Problem 4. Thus, a model that only includes the state variables is insufficient to explain the evolution of the GDP, that is, some of the control or environmental variables are necessary.

In summary, our proposed test confirms the validity of the model proposed by Koenker and Machado (1999). In addition, from the outcome for Problem 3, it would be sufficient to consider a model with nine explanatory variables to explain the growth rate of the Per Capita GDP.

4.5 Conclusions

We proposed a new lack-of-fit test for quantile regression models, together with a bootstrap mechanism to approximate the critical values. The bootstrap approximation does not need to estimate the conditional sparsity, and was shown to work well in homoscedastic and heteroscedastic error distributions and with high-dimensional covariates.

The proposed test is generally more powerful than its natural competitors, and particularly more powerful in the case of a high-dimensional covariate.

The proposed test was applied to a real data situation, where it was useful to validate well-known models in the economic literature, that describe the evolution of the GDP in terms of a number of explanatory variables.

The proposed method can be generalized to test models involving quantiles of different orders. The most treated model in the literature is the multiple quantile linear model, where it is assumed that the quantile regression function is linear for a subset of orders $\tau \in \mathcal{T} \subset [0,1]$,

$$q_\tau(x) = \theta'_\tau(1, x),$$

with coefficients θ_τ depending on the order, τ, of the quantile. The coefficients θ_τ allow consideration of a different effect of the covariates depending on the order of the quantile. See Escanciano and Goh (2014) for a lack-of-fit test of multiple quantile linear models, or Escanciano and Velasco (2010) for a test of multiple quantile models with time series. Our proposed method can be generalized to test multiple quantile models in a general framework of parametric (possibly nonlinear) quantile regression with heteroscedasticity and without estimating unknown quantities. To this end, one would consider a process depending on (β, u), as well as on τ. We restricted to the case of testing a single quantile to focus on the performance of the test for high-dimensional covariates and other important features of the testing problem. Extension to multiple quantile testing was left to future research.

Similarly, extensions of the proposed method to time series are possible using the results in Escanciano and Velasco (2010). These possible extensions show that the concept of projecting the covariate, given by Escanciano (2006) to overcome the curse of dimensionality, combined with the bootstrap methodology introduced by Feng et al. (2011), provide a promising strategy for checking quantile regression models.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Chapter 5

A lack-of-fit test for quantile regression models using logistic regression

Contents

A new lack-of-fit test for parametric quantile regression models is proposed along this chapter. The test is based on interpreting the residuals from the quantile regression model fit as response values of a logistic regression, the predictors of the logistic regression being functions of the covariates of the quantile model. Then a correct quantile model implies the nullity of all the coefficients but the constant in the logistic model. Given this property, we use a likelihood ratio test in the logistic regression to check the quantile regression model. In the case of a multivariate quantile regression, to avoid working in very large dimension, we use predictors obtained as functions of univariate projections of the covariates from the quantile model. Finally, we look for a "least favourable" projection for the null hypothesis of the likelihood ratio test. Our test can detect general departures from the parametric quantile model. To approximate the critical values of the test, a wild bootstrap mechanism is used, similar to that proposed by Feng et al. (2011). A simulation study and an application to real data show the good properties of the new test versus other nonparametric tests available in the literature.

5.1 Introduction

Given a pair of variables $(X, Y) \in \mathbb{R}^{d+1}$, let us consider a quantile regression model denoted by

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

where $q_\tau(\cdot)$ represents the regression function and the error ε has a conditional τ-quantile equal to zero, that is $\mathbb{P}(\varepsilon \leq 0|X = x) = \tau$ for almost all x. Along this chapter, we are going to address the same problem that have been studied in Chapter 4 but from a completely different approach. We will focus on the problem of testing a parametric quantile regression model

$$H_0 : q_\tau(\cdot) \in \mathcal{Q}_\theta = \{q_\tau(\cdot, \theta) \,:\, \theta \in \Theta \subset \mathbb{R}^q\}\,, \tag{5.1}$$

versus a nonparametric alternative $\sup_\theta \mathbb{P}(q_\tau(X) = q_\tau(X, \theta)) < 1$.

Then, the new lack-of-fit test will be based on an idea introduced by Redden et al. (2004). They have proposed a simple method using logistic regression to identify significant covariates associated with a quantile regression model. Their development is relied on the fact that observations can be classified as above or below the predicted quantile. This classification step creates a dichotomous variable that can be utilized as the response variable in a logistic regression model. If the probability of being above the predicted quantile is independent of a certain explanatory variable, then this probability will be a constant across all values of the explanatory variable indicating no association between the quantile and the explanatory variable. Otherwise, if an explanatory variable within logistic regression is statistically significant, the same variable is interpreted to be significant in the quantile regression model.

Redden et al. (2004) demonstrated that their significant test has better type I error rate control and comparable power as compared to different tests available in the literature as Koenker and Machado (1999).

Bearing Redden et al. (2004)'s idea in mind, we will try to extend this parametric significance to a nonparametric significance. In view of the previous arguments, logistic regression will play an important role along this chapter. Because of this reason, a brief introduction to logistic regression will be presented.

5.1.1 Logistic regression

In many situations, the response variable associated with a regression model is dichotomous, that is, only takes values 0 or 1. For instance, binary responses are commonly studied in medical and epidemiological research. This kind of models are known by logistic regression models. The main properties that distinguish logistic regression from mean regression are the following ones: under logistic regression the conditional mean will be bounded between zero and one and the error should follow a binomial distribution.

Logistic regression is a common technique in different applied contexts due to its case of use from a mathematical point of view and its meaningful interpretation.

Let us introduce some notation. Consider a sample $(W_1, V_1), \cdots, (W_n, V_n)$ where the response variable V only takes values 0 or 1, and W is a vector of explanatory variables with the first component equal to 1 to include an intercept. In this situation, the statistical model that is generally preferred for the analysis of binary responses is the binary logistic regression model, stated in terms of the probability that $V = 1$ given W,

$$\mathbb{P}(V = 1|W = w) = \frac{1}{1 + e^{-\varphi' w}} \tag{5.2}$$

where φ represents a vector of unknown coefficients. The binary logistic regression model was developed primarily by Cox (1958) and Walker and Duncan (1967).

Alternatively, the logistic regression (5.2) can be expressed in terms of the logistic transformation (denoted by logit) as follows:

$$\text{logit}\,(\mathbb{P}(V = 1|W = w)) = \varphi' w,$$

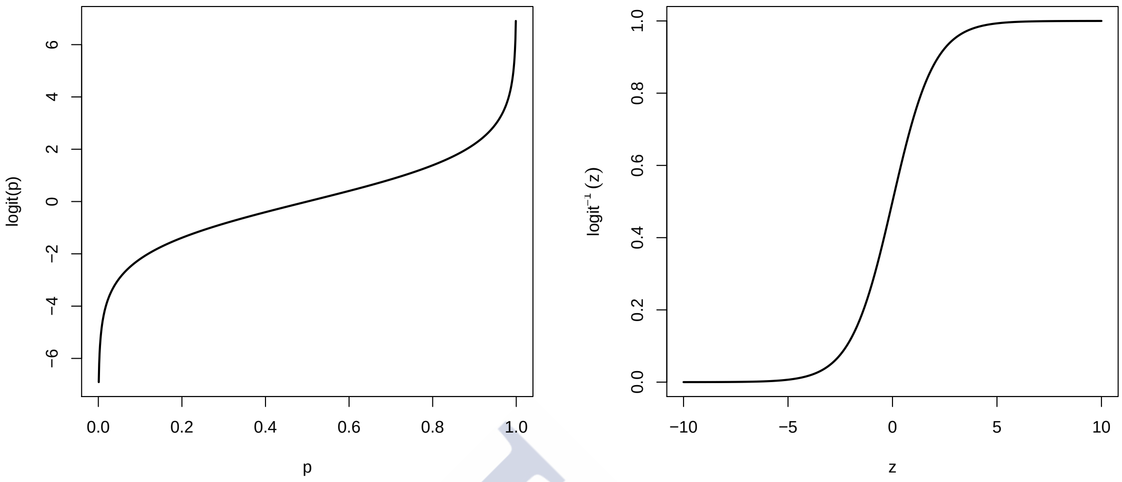
where $\text{logit}(p) = \log(p/(1-p))$. The importance of the logit transformation is that it transforms the domain $[0, 1]$, where a probability is naturally defined, in the domain $(-\infty, \infty)$, that is suitable for a linear regression. Figure 5.1 represents the logistic transformation and its inverse.

Each coefficient φ_j is the change in $\log(e^{\varphi' w})$ per unit change in w_j if w_j represents a single factor that is linear and does not interact with other factors and if all other factors are held constant.

The coefficients φ in the logistic regression model are estimated using the maximum likelihood (ML) method as

$$\begin{aligned} \widehat{\varphi} &= \arg\max_{\varphi} \left[n^{-1} L_n(\varphi, V, W) \right] \\ &= \arg\max_{\varphi} \left[\frac{1}{n} \sum_{i=1}^{n} \left(V_i \varphi' W_i - \log(1 + e^{\varphi' W_i}) \right) \right], \end{aligned} \tag{5.3}$$

where L_n denotes the likelihood function.



(a) Logistic transformation (b) Inverse of the logistic transformation

Figure 5.1: Representation of the logistic transformation and its inverse function.

Except in particularly simple cases, the ML estimates of φ cannot be written explicitly. The Newton-Raphson method or iteratively reweighted least squares method are usually used to solve iteratively for the list of values φ that maximize the likelihood. Moreover, for large enough samples, the maximum likelihood estimators are normally distributed.

From a different point of view, we can estimate the logistic regression model (5.2) via penalized maximum likelihood (PML). Then, the estimated parameter $\widehat{\varphi}$ could be computed as follows:

$$\begin{aligned} \widehat{\varphi} &= \arg\max_{\varphi} \left[n^{-1} L_n(\varphi, V, W) + \lambda \, \|\varphi\|_1 \right] \\ &= \arg\max_{\varphi} \left[\frac{1}{n} \sum_{i=1}^{n} \Big(V_i \varphi' W_i - \log(1 + e^{\varphi' W_i}) \Big) + \lambda \, \|\varphi\|_1 \right], \end{aligned} \tag{5.4}$$

where L_n denotes the likelihood function, $\| \cdot \|_1$ denotes the l_1 norm and λ is the smoothing parameter. We have considered a penalized ML estimation in order to control for large values of the coefficients that are likely to occur due to the separation problem, a well-known practical problem in logistic regression.

5.1.2 Significant tests for logistic regression models

Related to logistic regression models, let us consider the following significance test:

$$\begin{cases} H_0: & \text{logit}(\mathbb{P}(V = 1|W = w)) = \varphi' w \quad \text{where } \varphi = (\varphi_0, \cdots, \varphi_q, 0, \cdots, 0)' \in \mathbb{R}^{p+q+1}, \\ H_a: & \text{logit}(\mathbb{P}(V = 1|W = w)) = \varphi' w \quad \text{where } \varphi = (\varphi_0, \cdots, \varphi_q, \cdots, \varphi_{p+q+1})' \in \mathbb{R}^{p+q+1}. \end{cases} \tag{5.5}$$

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

In order to solve this problem, a likelihood-ratio test statistic will be considered and it adopts the form:

$$T_{\text{LR}} = 2\left(L_n(\widehat{\varphi}^{p+q}, V, W) - L_n(\widehat{\varphi}^{q}, V, W)\right),$$

where L_n denotes the likelihood function, $\widehat{\varphi}^{q}$ represents an estimator of the parameter $(\varphi_0, \cdots, \varphi_q, 0, \ldots, 0)'$ and $\widehat{\varphi}^{p+q}$ is an estimator of $(\varphi_0, \cdots, \varphi_q, \varphi_{q+1}, \ldots, \varphi_{p+q+1})'$. The test statistic T_{LR} is called the deviance difference in the logistic regression setup and it plays the same role as the residual sum of squares plays in linear regression. It is well-known that the resulting test statistic T_{LR} approximately follows a chi-square distribution, with degrees of freedom equal to the number of parameters that are constrained, that is

$$T_{\text{LR}} = 2\left(L_n(\widehat{\theta}^{p+q}, V, W) - L_n(\widehat{\theta}^{q}, V, W)\right) \longrightarrow \chi^2_{p+q-q} = \chi^2_p$$

See McCullagh and Nelder (1983), for instance.

Taking into account the state of the art, we propose and study a lack-of-fit test for parametric models of quantile regression based on logistic regression. In Section 5.2 we present the new test based on a likelihood ratio test in the logistic regression to check the quantile regression model. Moreover, a bootstrap method is also proposed to approximate the critical values of the test. Section 5.3 contains a simulation study where the performance of the test is studied under univariate and multivariate models, with different error distributions and sample sizes. We compare the proposed test with other tests available in the literature thanks to the simulation study and an application to real data (see Section 5.4). Some concluding remarks are provided in Section 5.5.

5.2 The new lack-of-fit test

As the have mentioned previously, following the idea introduced by Redden et al. (2004), the new lack-of-fit test is based on the dichotomous variable associated with the error of a parametric quantile regression model

$$Z(\theta_\tau) = \mathbb{I}(Y \leq q_\tau(X, \theta_\tau)).$$

Then, the parametric quantile regression model is correct if and only if there exists some $\theta_\tau \in \Theta$ such that the conditional probability of $Z(\theta_\tau)$ given X does not depend on X, and is equal to τ as a consequence of parametric quantile regression properties (see Section 1.2.3). In this point, in order to check the independence between a suitable $Z(\theta_\tau)$ and X, the idea is to consider a logistic regression with response $Z(\widehat{\theta}_\tau)$ and many covariates obtained as functions of the components of the vector X, and to test the nullity of all the coefficients but the constant.

5.2.1 Univariate case

Firstly, we are going to focus on the case in which the explanatory variable of the quantile regression model X is univariate. In this situation, to detect general nonparametric alternatives, the vector W used in the logistic regression should contain as many functions of the components of X, the original covariate vector in the quantile regression, as needed in order to detect all kind of alternative hypothesis. To formally describe our procedure, these function are represented by a dense basis of functions. Different basis of functions can be considered, for instance, we could use:

Hermite polynomial, that are defined as

$$H_p(x) = p! \sum_{m=0}^{[p/2]} \frac{(-1)^m}{m!(p-2m)!} \frac{x^{p-2m}}{2^m}, \qquad x \in \mathbb{R}, \qquad p \geq 0,$$

where $[a]$ denotes the integer part of a real number a.

Laguerre polynomials, that are defined as

$$L_p(x) = \sum_{k=0}^{p} \frac{(-1)^k}{k!} \binom{p}{k} x^k, \qquad x \in \mathbb{R}, \qquad p \geq 0,$$

where $\binom{p}{k}$ is a binomial coefficient.

Legendre polynomials, that are defined as

$$\mathrm{Le}_p(x) = \frac{1}{2^p} \sum_{k=0}^{p} \binom{p}{k}^2 (x+1)^{p-k}(x-1)^k, \qquad x \in [-1,1], \qquad p \geq 0.$$

Classical polynomials, that are defined as

$$C_p(x) = x^p, \qquad x \in \mathbb{R}, \qquad p \geq 0.$$

Figure 5.2 shows the first elements of these basis of polynomials. Henceforth, we are going to focus on the basis of Hermite polynomials. Let us adopt the following notation:

$$P_i = (H_0(X_i), H_1(X_i), H_2(X_i), H_3(X_i), \ldots, H_p(X_i))', \qquad 1 \leq i \leq n$$

$$= (1, H_1(X_i), H_2(X_i), H_3(X_i), \ldots, H_p(X_i))', \qquad 1 \leq i \leq n.$$

Then, the idea is to check whether, for some value θ_τ, we have $\varphi_1 = \varphi_2 = \cdots = \varphi_p = 0$ in the logistic regression model:

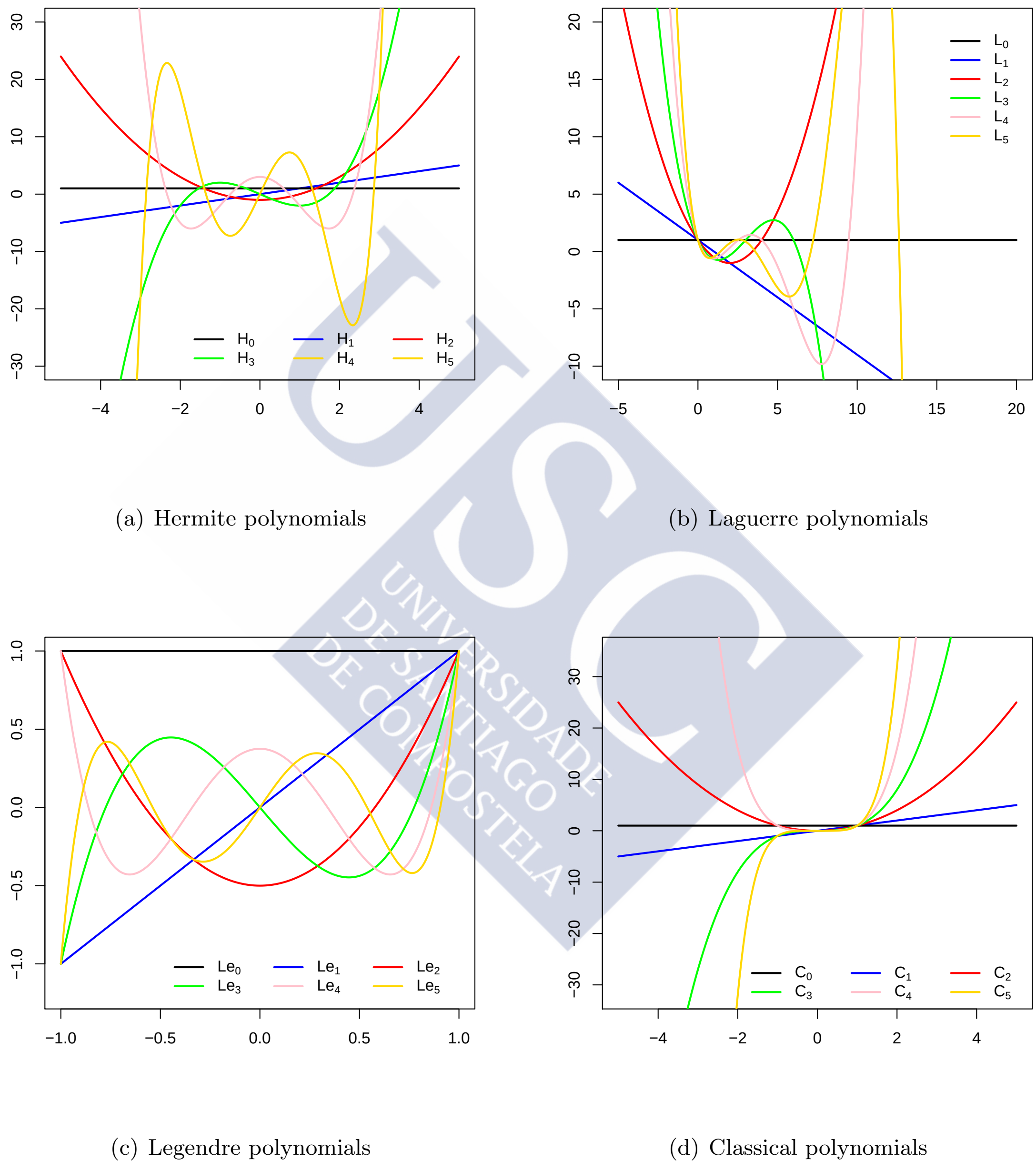
$$\mathrm{logit}\left(\mathbb{P}[Z(\theta_\tau) = 1|P]\right) = \varphi_0 + \varphi_1 H_1(X) + \varphi_2 H_2(X) + \cdots + \varphi_p H_p(X) = \varphi' P.$$

The infeasible responses $Z_i(\theta_\tau)$ are replaced by $Z_i(\widehat{\theta}_\tau)$. Meanwhile, φ is estimated via penalized maximum likelihood, that is, the procedure described in equation (5.4) that in this case is given by

$$\widehat{\varphi} = \arg\max_{\varphi} \left[\frac{1}{n} \sum_{i=1}^{n} \left(Z_i(\widehat{\theta}_\tau)\varphi' P_i - \log(1 + e^{\varphi' P_i}) \right) + \lambda \left\| \varphi \right\|_1 \right],$$

for some suitable smoothing parameter λ.

We have chosen penalized maximum likelihood in order to avoid the well known **separation problem** which is observed in the fitting process of a logistic model if the



(a) Hermite polynomials

(b) Laguerre polynomials

(c) Legendre polynomials

(d) Classical polynomials

Figure 5.2: Representation of the first polynomials associated with the different basis of functions considered.

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

likelihood converges while at least one parameter estimate diverges to $\pm\infty$. This problem has been treated by several authors like Firth (1993) or Horowitz and Spokoiny (2002).

To check the significance of the coefficients φ but φ_0, we use a likelihood ratio type statistic as described in Section 5.1.2. Gathering facts, the new lack-of-fit test for quantile regression is based on the test statistic

$$T_u = 2\left(L_n(\widehat{\varphi}, Z(\widehat{\theta}_\tau), P) - L_n(\text{logit}(\tau), Z(\widehat{\theta}_\tau), 1)\right), \tag{5.6}$$

where

$$L_n(\text{logit}(\tau), Z(\widehat{\theta}_\tau), 1) = \sum_{i=1}^{n}\left[Z_i(\widehat{\theta}_\tau)\text{logit}(\tau) - \log\left(1 + e^{\text{logit}(\tau)}\right)\right].$$

A bootstrap procedure in the quantile regression context will be proposed in order to calibrate the critical values of the test statistic (5.6). The bootstrap procedure works as follows:

1.- Let $\varepsilon_i^\star = \delta_i|r_i|$, where $r_i = Y_i - q_\tau(X_i, \widehat{\theta}_\tau)$ are the residuals from the original sample. The multipliers, δ_i, are independently generated from the two-point distribution with probabilities $(1-\tau)$ and τ at $2(1-\tau)$ and -2τ, respectively; more details have been given in Chapter 2. Compute $Y_i^\star = q_\tau(X_i, \widehat{\theta}_\tau) + \varepsilon_i^\star$ for each $i = 1, \cdots, n$.

2.- Use the bootstrap data set $\{(X_i, Y_i^\star),\ i = 1, \cdots, n\}$ to compute the bootstrap estimator $\widehat{\theta}_\tau^\star$ and the dichotomous variables $Z_i(\widehat{\theta}_\tau^\star) = \mathbb{I}(Y_i^\star \leq q_\tau(X_i, \widehat{\theta}_\tau^\star))$.

3.- Use the data set $\{(P_i, Z_i(\widehat{\theta}_\tau^\star)),\ i = 1, \cdots, n\}$ to compute the estimator $\widehat{\varphi}^\star$, following the procedure described in (5.4) with $L_n(\varphi, Z_i(\widehat{\theta}_\tau^\star), P)$, and the bootstrap test statistic

$$T_{u,b}^\star = 2\left(L_n(\widehat{\varphi}^\star, Z(\widehat{\theta}_\tau^\star), P) - L_n(\text{logit}(\tau), Z(\widehat{\theta}_\tau^\star), 1)\right).$$

4.- Repeat Steps 1, 2 and 3 B times, and estimate the α-level critical value by the $(1-\alpha)$-quantile of the resulting $T_{u,1}^\star, \ldots, T_{u,B}^\star$ values.

5.2.2 Multivariate case

Now, we are going to move to a more general scenario. In this second case, we will consider that the covariate of the quantile regression model (denoted by X) is multivariate. In this new situation, to avoid working with very large dimensions for W, we follow a projection approach. More precisely, we note that H_0 defined in (5.1) holds true if and only if, for some $\theta_\tau \in \Theta \subset \mathbb{R}^q$, and for all $\beta \in \mathbb{S}_d = \{w \in \mathbb{R}^d \ \text{ with } \ \|w\| = 1\}$,

$$\mathbb{E}\left[\mathbb{I}(Y \leq q_\tau(X, \theta_\tau)) - \tau \,\middle|\, F_\beta(\beta' X)\right] = 0, \tag{5.7}$$

where $F_\beta(t) = \mathbb{P}(\beta' X \leq t)$ represents the distribution function of the projected covariate. This property suggests that it suffices to consider the logistic regression with W a vector of univariate functions of $\beta' X$ and to check whether all the coefficients but the constant are null. Finally, it remains to search a 'least favourable' direction β for the null hypothesis (5.1), such as Conde-Amboage et al. (2015) (detailed in Chapter 4) or Patilea et al. (2016) did.

Let

$$P_i(\beta) = (1, H_1(\beta' X_i), H_2(\beta' X_i), \ldots, H_p(\beta' X_i))', \qquad 1 \leq i \leq n,$$

represent a basis of Hermite polynomial evaluated at the projections $\beta' X_1, \cdots, \beta' X_n$. Following the ideas described previously, if we consider the logistic regression model

$$\text{logit}\left(\mathbb{P}[Z(\theta_\tau) = 1 | P(\beta)]\right) = \varphi_0 + \varphi_1 H_1(\beta' X) + \cdots + \varphi_p H_p(\beta' X) = \varphi' P(\beta),$$

the idea is to check if $\varphi_2 = \cdots = \varphi_p = 0$ for some value θ_τ. Then, the new lack-of-fit test for quantile regression is based on the test statistic

$$T_m = \max_{\beta \in \mathbb{R}^d, \|\beta\| = 1} 2\left(L_n(\widehat{\varphi}, Z(\widehat{\theta}_\tau), P(\beta)) - L_n(\text{logit}(\tau), Z(\widehat{\theta}_\tau), 1)\right). \tag{5.8}$$

Finally, in order to calibrate the critical values for the test statistic (5.8), we will consider a bootstrap procedure similar to that described in Section 5.2.1. The main difference is the bootstrap test statistic that in this case will be given by

$$T_m^\star = \max_{\beta \in \mathbb{R}^d, \|\beta\| = 1} 2\left(L_n(\widehat{\varphi}^\star, Z(\widehat{\theta}_\tau^\star), P(\beta)) - L_n(\text{logit}(\tau), Z(\widehat{\theta}_\tau^\star), 1)\right).$$

Note that the covariate of the logistic model, $P_i(\beta)$, do not need to be computed for each bootstrap sample because it only depends on the covariates, and we are not bootstrapping the explanatory variable of the quantile regression model. This fact can reduce considerably the computational time associated with the new proposal.

Moreover, to compute the test statistic given by (5.8), we are going to use the sequential algorithm based on successive one dimensional-optimizations proposed by Patilea et al. (2016). In order to perform this method it will be necessary to follow these steps:

- If a bidimensional quantile regression model is considered, the different directions $\beta \in \mathbb{S}_2$ can be represented by $\beta = (\cos(\gamma), \sin(\gamma))$ with $\gamma \in [0, 2\pi)$. Then, an equally-spaced grid of values of γ in $[0, 2\pi)$ provides an equally-spaced grid of directions in $\mathbb{S}_2$.

- If $p = 3$, the first step would be to optimize with respect to the first two components as in the previous situation. Let γ_{12}^* represent an optimal direction. The next step would be to optimize in the set of directions $\cos(\gamma_3)(\cos(\gamma_{12}^*), \sin(\gamma_{12}^*), 0) + \sin(\gamma_3)(0, 0, 1)$ with $\gamma_3 \in [0, 2\pi)$. This can be solved with a grid of values in the interval $[0, 2\pi)$, that is a univariate optimization problem.

- This procedure can be applied to a possible fourth dimension, from the optimal direction obtained with the first three dimensions, and so on until the chosen number of components d is reached. Finally, this method would require $(d - 1)$ one-dimensional optimizations where d represents the dimension of the explanatory variable associated with the quantile regression model.

5.3 Simulation study

Along this section, the performance of the proposed method under the null and alternative hypotheses will be analysed using a Monte Carlo simulation study. The number of simulated original samples was 1000 and the number of bootstrap replications 500.

5.3.1 Scenario 1: Univariate case

Firstly, we will check the adjustment of the significant level associated with the proposed lack-of-fit test. In order to perform the new test, it will be necessary to select the number of Hermite polynomials (denoted by p) and the smoothing parameter related to penalized maximum likelihood estimation (denoted by λ). Now, our aim will be to study the effect of these parameters on the adjustment of the significant level. We are going to start simulating values from the following quantile regression model:

$$\textbf{Model 5.1:} \qquad Y = 1 + X^{(1)} + \varepsilon$$

where $X^{(1)}$ follows an uniform distribution on the interval $(0, 1)$ and ε represents the unknown error that follows an standard Gaussian distribution.

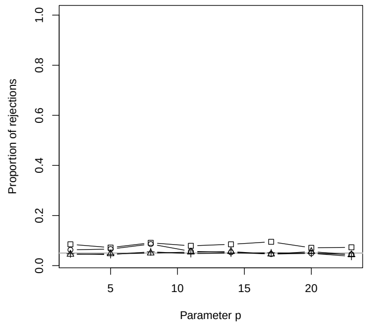
Figures 5.3 and 5.4 represents the proportion of rejections associated with our proposed lack-of-fit test for Model 5.1 depending on the parameters p, the sample size (denoted by n) and the τ-quantile of interest, for a significant nominal level $\alpha = 0.05$. Moreover, different values of the parameter λ have been considered: $\lambda = 2pn^{-1}$ represented by squares, $\lambda = \log(n)n^{-1}$ represented by circles, $\lambda = \log(p)n^{-1}$ represented by triangles and $\lambda = n^{-1}$ represented by crosses.

According to Figures 5.3 and 5.4, we can conclude that $\lambda = 2pn^{-1}$ is not a good option independently of the value of p, n or τ. Moreover $\lambda = \log(n)n^{-1}$ does not provide good results, see for instance, $\tau = 0.75$ or $\tau = 0.90$ when the sample size is 50. It seems that $\lambda = 2pn^{-1}$ or $\lambda = \log(n)n^{-1}$ are quite big penalizations and this fact might distort the shape of the original likelihood and as a result, the performance of the proposed lack-of-fit test will not be adequate. On the other hand, the differences between $\lambda = \log(p)n^{-1}$ and $\lambda = n^{-1}$ are small, but $\lambda = \log(p)n^{-1}$ seems to be a bit more appropriate, see for instance $\tau = 0.50$ and $n = 100$ or $\tau = 0.90$ and $n = 200$.

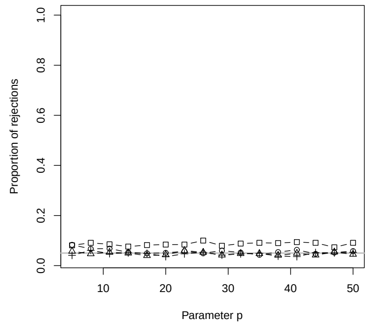
In addition, if we focus on the adjustment of the significant level associated with $\lambda = \log(p)n^{-1}$ the effect of the parameter p does not seem to be very important. However, it is expected that this parameter p will have an important effect on the power of the lack-of-fit test under the alternative. In order to answer this question, we will generate values for the following quantile regression model under the alternative:

$$\textbf{Model 5.2:} \qquad Y = 1 + X^{(1)} + cX^{(1)}X^{(1)} + \varepsilon$$

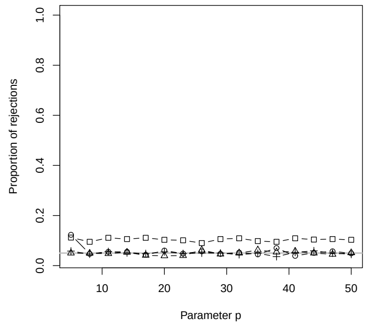
where $X^{(1)}$ follows an uniform distribution on the interval $(0, 1)$ and ε represents the error that follows an standard Gaussian distribution. Note that the parameter c represents the deviation of Model 5.2 from the null hypothesis that is the linear model in the explanatory variable $X^{(1)}$.



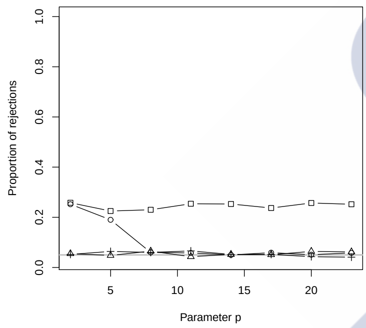
(a) $n = 50$ and $\tau = 0.10$



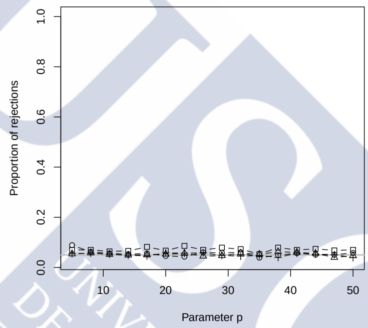
(b) $n = 100$ and $\tau = 0.10$



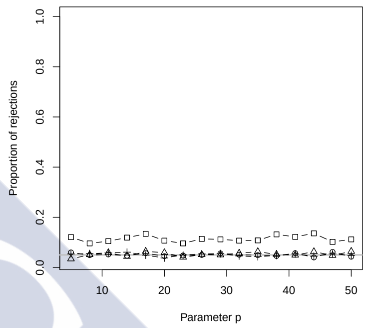
(c) $n = 200$ and $\tau = 0.10$



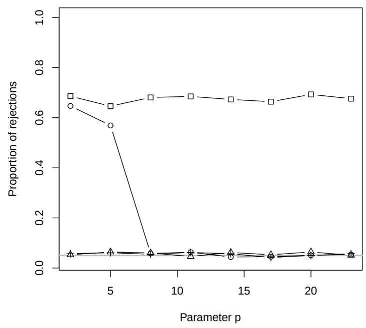
(d) $n = 50$ and $\tau = 0.25$



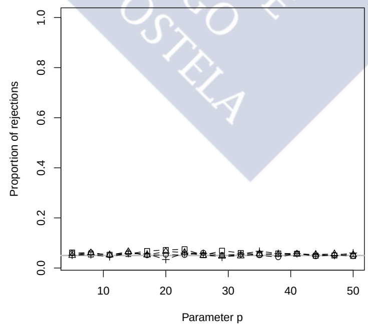
(e) $n = 100$ and $\tau = 0.25$



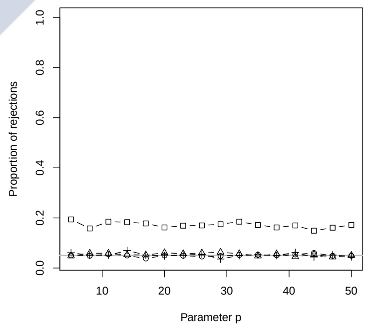
(f) $n = 200$ and $\tau = 0.25$



(g) $n = 50$ and $\tau = 0.50$

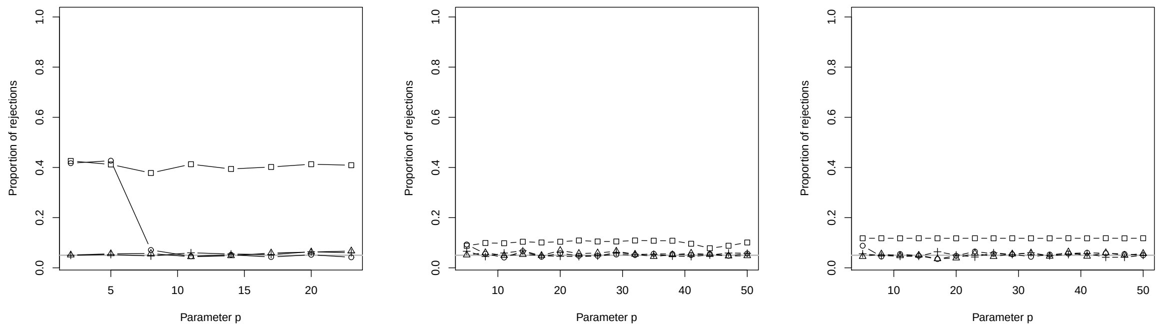


(h) $n = 100$ and $\tau = 0.50$

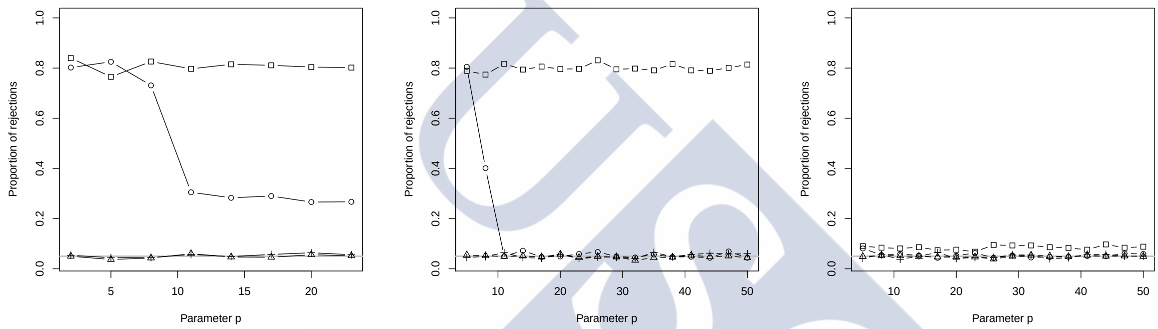


(i) $n = 200$ and $\tau = 0.50$

Figure 5.3: Proportion of rejections associated with our proposed lack-of-fit test for Model 5.1 depending on the parameters p, λ, the τ-quantile of interest and the sample size. The following values of the parameter λ have been considered: $\lambda = 2pn^{-1}$ represented by squares, $\lambda = \log(n)n^{-1}$ represented by circles, $\lambda = \log(p)n^{-1}$ represented by triangles and $\lambda = n^{-1}$ represented by crosses.



(a) $n = 50$ and $\tau = 0.75$ (b) $n = 100$ and $\tau = 0.75$ (c) $n = 200$ and $\tau = 0.75$



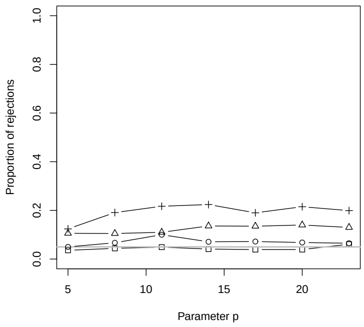
(d) $n = 50$ and $\tau = 0.90$ (e) $n = 100$ and $\tau = 0.90$ (f) $n = 200$ and $\tau = 0.90$

Figure 5.4: Proportion of rejections associated with our proposed lack-of-fit test for Model 5.1 depending on the parameters p, λ, the τ-quantile of interest and the sample size. The following values of the parameter λ have been considered: $\lambda = 2pn^{-1}$ represented by squares, $\lambda = \log(n)n^{-1}$ represented by circles, $\lambda = \log(p)n^{-1}$ represented by triangles and $\lambda = n^{-1}$ represented by crosses.

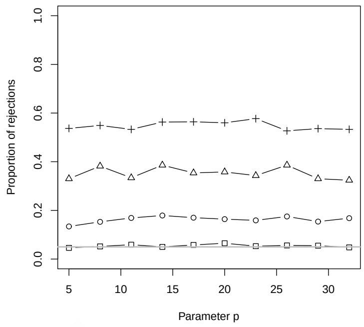
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Figures 5.5 and 5.6 represent the proportion of rejections associated with our proposed lack-of-fit test for Model 5.2 depending on the parameters p and the τ-quantile of interest for $\lambda = \log(p)n^{-1}$. Moreover, different values of the parameter c have been considered: $c = 1$ represented by squares, $c = 3$ represented by circles, $c = 5$ represented by triangles and $c = 7$ represented by crosses.

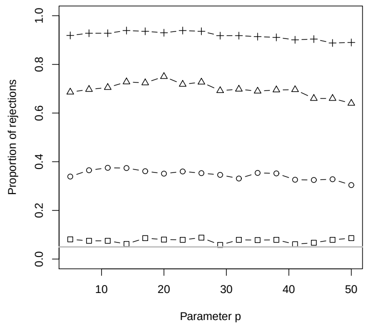
According to Figures 5.5 and 5.6, we can conclude that the parameter p, that is, the number of Hermite polynomials, does not have a clear effect on the power of the test. In order to understand this fact, we should take into account the expression of Model 5.2. In this case, the deviation from the null hypothesis only depends on the squares of the original covariate, so a small number of Hermite polynomials is enough to detect the alternative. Additionally, we can observe a relatively big and unnecessary p does not lead to a substantial lose of power. This good property can be attributed to the penalization included in the likelihood-ratio test.



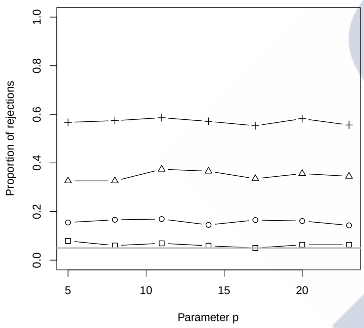
(a) $n = 50$ and $\tau = 0.10$



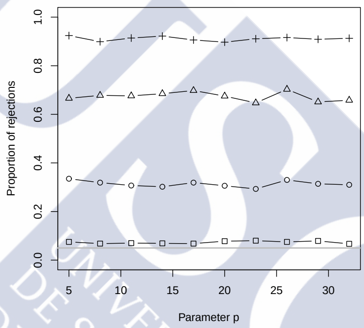
(b) $n = 100$ and $\tau = 0.10$



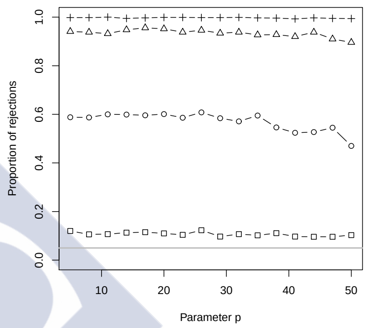
(c) $n = 200$ and $\tau = 0.10$



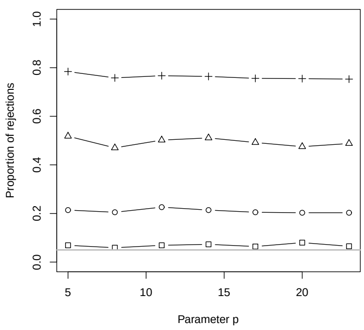
(d) $n = 50$ and $\tau = 0.25$



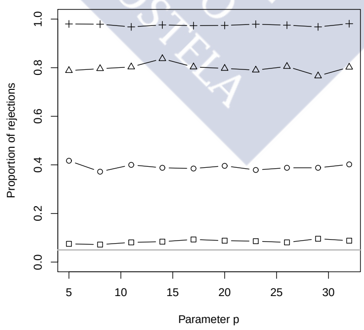
(e) $n = 100$ and $\tau = 0.25$



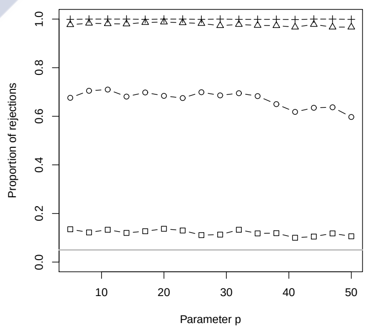
(f) $n = 200$ and $\tau = 0.25$



(g) $n = 50$ and $\tau = 0.50$

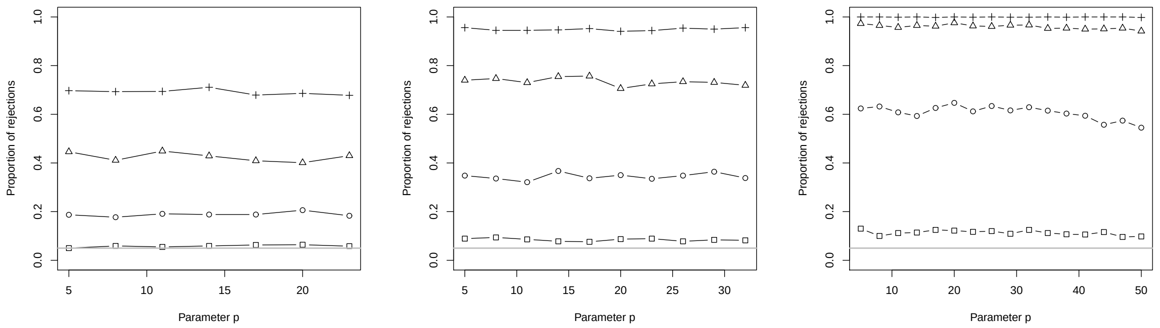


(h) $n = 100$ and $\tau = 0.50$

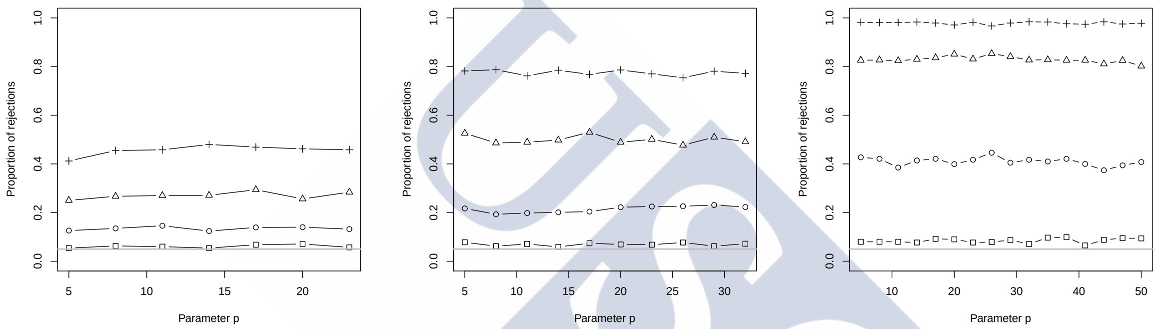


(i) $n = 200$ and $\tau = 0.50$

Figure 5.5: Proportion of rejections associated with our proposed lack-of-fit test for Model 5.2 depending on the parameters p, c, the τ-quantile of interest and the sample size. The following values of the parameter c have been considered: $c = 1$ represented by squares, $c = 3$ represented by circles, $c = 5$ represented by triangles and $c = 7$ represented by crosses.



(a) $n = 50$ and $\tau = 0.75$ (b) $n = 100$ and $\tau = 0.75$ (c) $n = 200$ and $\tau = 0.75$



(d) $n = 50$ and $\tau = 0.90$ (e) $n = 100$ and $\tau = 0.90$ (f) $n = 200$ and $\tau = 0.90$

Figure 5.6: Proportion of rejections associated with our proposed lack-of-fit test for Model 5.2 depending on the parameters p, c, the τ-quantile of interest and the sample size. The following values of the parameter c have been considered: $c = 1$ represented by squares, $c = 3$ represented by circles, $c = 5$ represented by triangles and $c = 7$ represented by crosses.

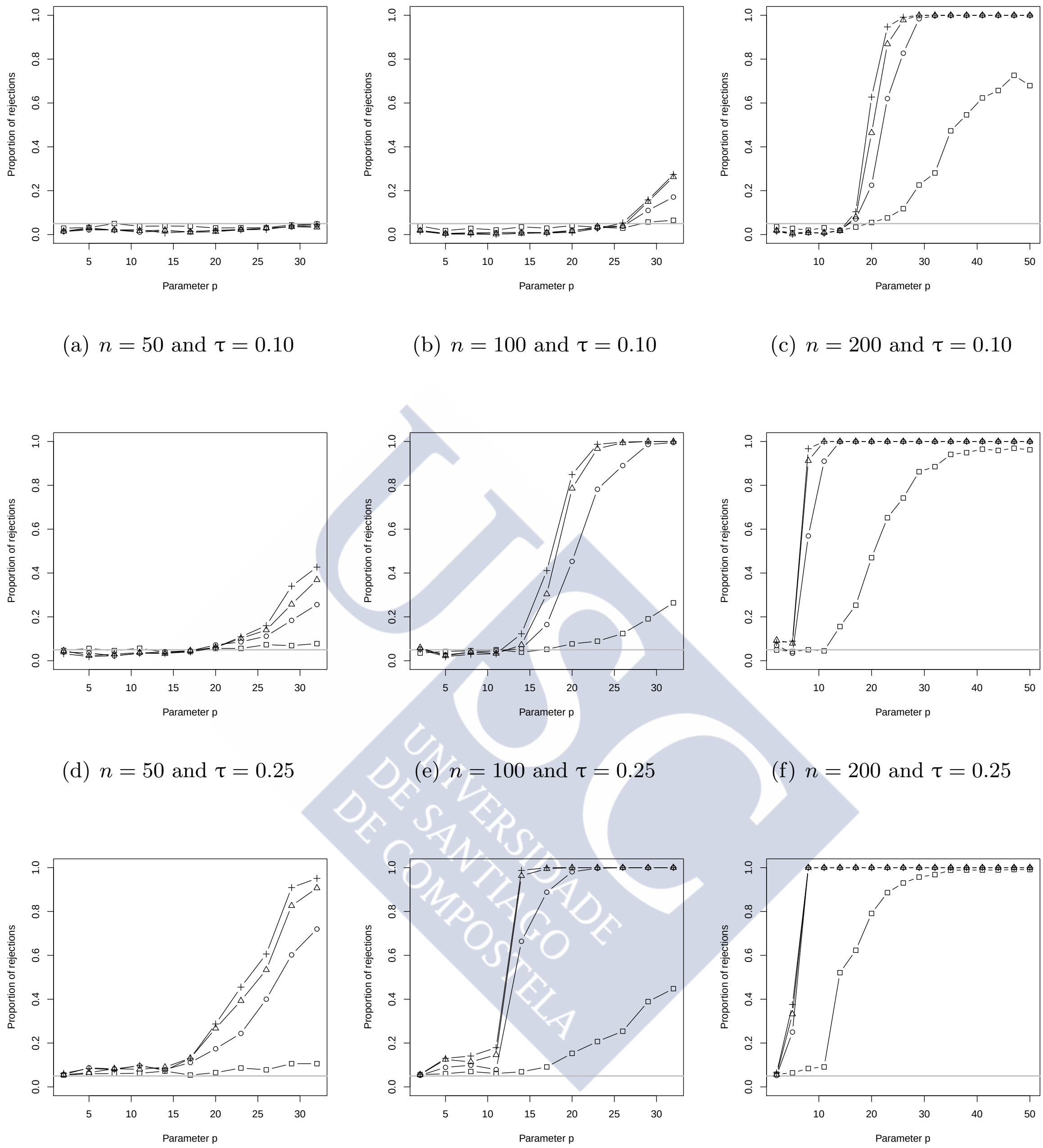
USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Now, let us consider a more complicated deviation from the null hypothesis, where a higher number of Hermite polynomials will be necessary. In particular, we are going to consider the following quantile regression model:

$$\textbf{Model 5.3:} \qquad Y = 1 + X^{(1)} + c\sin(2\pi X^{(1)}) + \varepsilon$$

where $X^{(1)}$ follows an uniform distribution on the interval $(0, 1)$ and ε represents the error that follows an standard Gaussian distribution.

Figures 5.7 and 5.8 show the proportion of rejections associated with the proposed test for Model 5.3 depending on the parameter p, the quantile of interest and different values of the deviation from the null hypothesis represented by the parameter c. The influence of the parameter p is clear in this situation. We can observe that the power of the new lack-of-fit test increases with the parameter p, specially when the sample size is big. Summarizing, the number of Hermite polynomials that should be considered in order to obtain a reasonable power for Model 5.3 is bigger than that for Model 5.2, as was expected in view of both deviations from the null hypothesis.



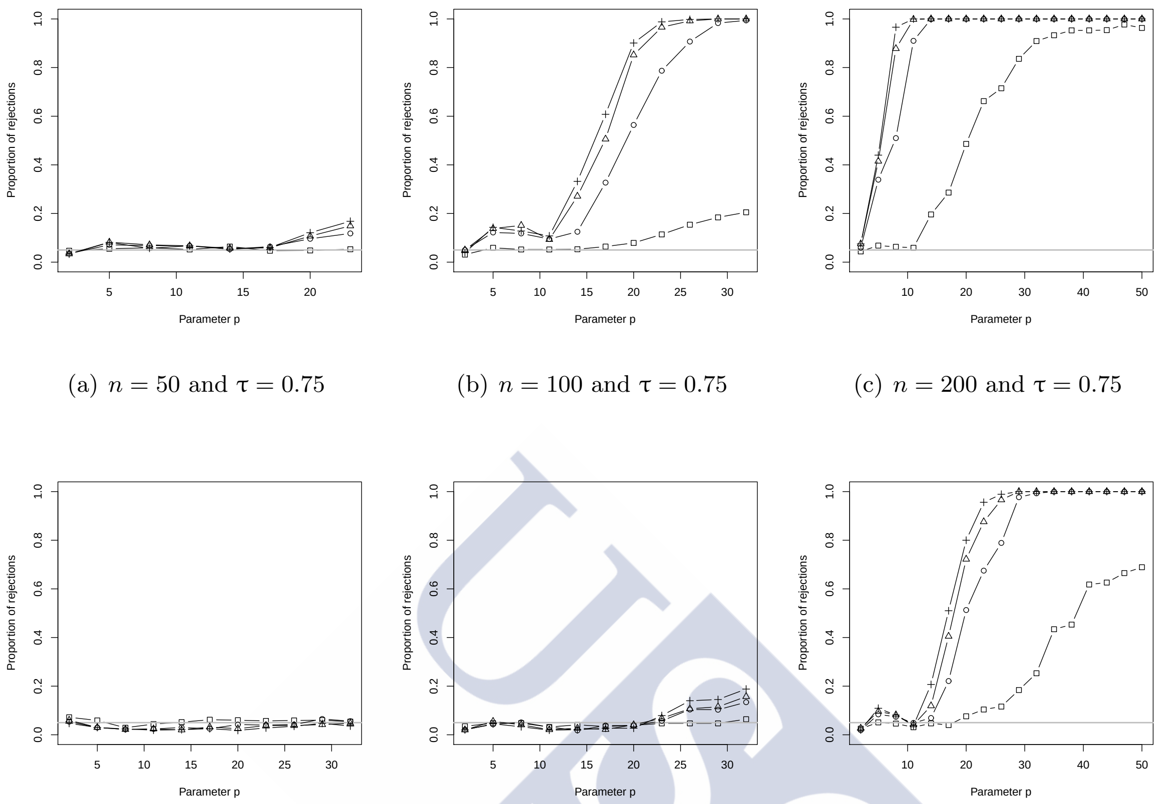
(a) $n = 50$ and $\tau = 0.10$ (b) $n = 100$ and $\tau = 0.10$ (c) $n = 200$ and $\tau = 0.10$

(d) $n = 50$ and $\tau = 0.25$ (e) $n = 100$ and $\tau = 0.25$ (f) $n = 200$ and $\tau = 0.25$

(g) $n = 50$ and $\tau = 0.50$ (h) $n = 100$ and $\tau = 0.50$ (i) $n = 200$ and $\tau = 0.50$

Figure 5.7: Proportion of rejections associated with our proposed lack-of-fit test for Model 5.3 depending on the parameters p, c, the τ-quantile of interest and the sample size. The following values of the parameter c have been considered: $c = 1$ represented by squares, $c = 3$ represented by circles, $c = 5$ represented by triangles and $c = 7$ represented by crosses.

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



(a) $n = 50$ and $\tau = 0.75$ (b) $n = 100$ and $\tau = 0.75$ (c) $n = 200$ and $\tau = 0.75$

(d) $n = 50$ and $\tau = 0.90$ (e) $n = 100$ and $\tau = 0.90$ (f) $n = 200$ and $\tau = 0.90$

Figure 5.8: Proportion of rejections associated with our proposed lack-of-fit test for Model 5.3 depending on the parameters p, c, the τ-quantile of interest and the sample size. The following values of the parameter c have been considered: $c = 1$ represented by squares, $c = 3$ represented by circles, $c = 5$ represented by triangles and $c = 7$ represented by crosses.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

5.3.2 Scenario 2: Multivariate case

Here the purpose will be to compare the proposed test with other competitors designed to deal with multiple explanatory variables. The number p of Hermite polynomials will be fixed to $p = [\sqrt{n}]$. Based on previous experiments, the smoothing parameter λ related to the penalized maximum likelihood will be set to $n^{-1}\log(p)$. Firstly, we study the behaviour under the null hypothesis, that is, the adjustment of the significant level. Data will be simulated from the following median ($\tau = 0.5$) regression model:

$$\textbf{Model 5.4:} \qquad Y = 1 + X^{(1)} + X^{(2)} + \varepsilon$$

where $X^{(i)}$ follows an uniform distribution on the interval $(0, 1)$ with $i = 1, 2$ and ε represents the unknown error. Table 5.1 presents the proportion of samples for which the null hypothesis, characterized by the linear model, was rejected, for different sample sizes n, nominal levels α and different error distributions. The new method shows a good adjustment to the nominal level, even for a small sample size and independently of the error distribution.

		$\tau=0.10$			$\tau=0.25$			$\tau=0.50$			$\tau=0.75$			$\tau=0.90$		
ε	n	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.01$
$N(0,1)$	50	0.114	0.044	0.009	0.090	0.052	0.009	0.092	0.046	0.015	0.105	0.053	0.014	0.271	0.091	0.011
	100	0.108	0.045	0.009	0.115	0.054	0.002	0.108	0.048	0.015	0.102	0.054	0.012	0.174	0.050	0.011
	200	0.118	0.057	0.011	0.105	0.046	0.016	0.093	0.052	0.016	0.115	0.061	0.013	0.105	0.058	0.017
$U[-1,1]$	50	0.097	0.052	0.009	0.093	0.049	0.007	0.114	0.062	0.011	0.099	0.057	0.009	0.314	0.079	0.010
	100	0.118	0.056	0.010	0.107	0.058	0.007	0.084	0.049	0.012	0.114	0.067	0.015	0.210	0.059	0.015
	200	0.098	0.047	0.009	0.095	0.052	0.016	0.086	0.043	0.007	0.097	0.052	0.008	0.104	0.041	0.007
χ^2_2	50	0.094	0.036	0.008	0.103	0.054	0.009	0.113	0.063	0.014	0.108	0.064	0.019	0.251	0.096	0.023
	100	0.085	0.046	0.007	0.088	0.042	0.012	0.099	0.055	0.009	0.093	0.047	0.017	0.122	0.049	0.010
	200	0.098	0.049	0.010	0.098	0.052	0.015	0.110	0.048	0.010	0.102	0.060	0.012	0.081	0.033	0.011
t_1	50	0.130	0.063	0.013	0.100	0.057	0.018	0.092	0.037	0.012	0.111	0.059	0.014	0.168	0.063	0.012
	100	0.097	0.046	0.013	0.105	0.054	0.014	0.104	0.046	0.016	0.097	0.056	0.008	0.113	0.053	0.006
	200	0.090	0.039	0.015	0.096	0.053	0.011	0.104	0.058	0.015	0.093	0.049	0.010	0.094	0.041	0.012
Exp(1)	50	0.081	0.033	0.008	0.091	0.046	0.012	0.099	0.051	0.010	0.112	0.060	0.016	0.260	0.089	0.011
	100	0.110	0.049	0.009	0.101	0.047	0.011	0.101	0.051	0.011	0.094	0.061	0.017	0.137	0.048	0.011
	200	0.103	0.058	0.013	0.093	0.046	0.012	0.116	0.060	0.014	0.110	0.049	0.015	0.088	0.047	0.020

Table 5.1: Proportions of rejections associated with the new lack-of-fit test for Model 5.4, for different nominal levels, sample sizes, τ-quantiles of interest and error distributions.

Next, the performance of the new test under different alternatives will be studied. The proposed lack-of-fit test (denoted by *NT*) will be compared with that of Maistre et al. (2014) denoted by *MLP*, the test studied along Chapter 4 denoted by *CSG*, and He and Zhu (2003) denoted by *HZ*. Our new test will be denoted by *NT*. We will consider the following median regression model:

$$\textbf{Model 5.5:} \qquad Y = 1 + X^{(1)} + X^{(2)} + h(X^{(1)}, X^{(2)}) + \varepsilon,$$

where $X^{(1)}$ follows a standard Gaussian distribution, $X^{(2)}$ follows an uniform distribution on the interval $(0, 1)$, and $\varepsilon + 1$ follows a standard log-normal distribution. The function $h(\cdot)$ represents the deviation from the null hypothesis, that is the linear model in the two variables $X^{(1)}$ and $X^{(2)}$. Two deviations will be considered:

- $h(X^{(1)}, X^{(2)}) = 5 \sin\left(2\pi(1 + X^{(1)} + X^{(2)})\right)$, that will be denoted by D1;
- $h(X^{(1)}, X^{(2)}) = 10\, (X^{(2)})^2$, that will be denoted by D2.
- $h(X^{(1)}, X^{(2)}) = \frac{1}{3}\left((X^{(1)})^2 + (X^{(2)})^2 + X^{(1)}X^{(2)}\right)$ that will be denoted by D3.

Table 5.2 shows the proportion of samples for which the null hypothesis was rejected under Model 5.5 for each of the methods (*NT*, *HZ*, *CSG* and *MLP*), for different sample sizes n and nominal levels α. According to Table 5.2, the power of the new test for deviations D1 and D2 is clearly superior, compared with the considered nonparametric competitors. On the other hand, if we focus on deviation D3 (used previously in Chapter 4, see Model 4.4), that seems "less difficult" to detect, the results of the different lack-of-fit test considered are quite similar, but *CSG* and *MLP* show a bit more power.

Summarizing, the proposed lack-of-fit test applied in a multivariate scenario shows a good adjustment of the significance level and a reasonable power compared with its natural competitors. We stand out the high power associated with the new test in situation in where the deviation from the null hypothesis is, a priori, difficult to detect as deviation D1 witch involves a sinus function.

		$\alpha = 0.10$				$\alpha = 0.05$				$\alpha = 0.01$			
		NT	HZ	CSG	MLP	NT	HZ	CSG	MLP	NT	HZ	CSG	MLP
D1	$n = 49$	0.800	0.150	0.130	0.120	0.730	0.090	0.050	0.080	0.555	0.015	0.025	0.025
	$n = 100$	0.985	0.105	0.170	0.155	0.975	0.055	0.085	0.055	0.925	0.015	0.005	0.010
D2	$n = 49$	0.775	0.260	0.075	0.145	0.695	0.160	0.035	0.075	0.485	0.070	0.010	0.025
	$n = 100$	0.995	0.740	0.185	0.145	0.980	0.570	0.055	0.075	0.940	0.220	0.010	0.010
D3	$n = 49$	0.415	0.425	0.685	0.655	0.305	0.295	0.590	0.535	0.120	0.115	0.345	0.350
	$n = 100$	0.855	0.825	0.950	0.900	0.725	0.770	0.930	0.870	0.525	0.510	0.810	0.690

Table 5.2: Proportions of rejections associated with four lack-of-fit tests (*NT*, *HZ*, *CSG* and *MLP*) for Model 5.5, for different deviations from the null hypothesis, different sample sizes n and nominal levels α.

5.4 Application to real data

Along this section, the proposed lack-of-fit test will be applied to real data. As is Chapter 2 we are going to deal with an environmental problem. In this case, we are going to consider air quality measurements for the New York Metropolitan Area that are introduced in Chambers et al. (1983). Moreover, this dataset is available in the R package `datasets` under the name **airquality**.

This data set contains 116 measurements obtained from May 1, 1973 to September 30, 1973 of the following variables:

- Mean ozone concentrations measured from 13:00 to 15:00 hours at Roosevelt Island. This variable is expressed in part per billion.

- Solar radiation measurements in the frequency band 4000-7700 Ångströms from 08:00 to 12:00 hours at Central Perk. This variable is expressed in Langleys.

- Wind speed obtained as the mean of measurements at 07:00 and 10:00 hours at La Guardia Airport This variable is expressed in miles per hour.

- Maximum daily temperature measured at La Guardia Airport. This variable is expressed in degrees Fahrenheit.

In view of this data set, we are going to focus on studying the relationship between temperature and ozone values[1]. Figure 5.9 represent the scatterplot of both variables joint with a boxplot representation of each variable in order to have a landscape about the considered scenario.

So, our goal will be to test if the relationship between temperature and ozone is linear versus a nonparametric alternative. That is, we are going to perform the following lack-of-fit test:

$$\begin{cases} H_0 : \text{Ozone} = \theta_0 + \theta_1 \text{Temperature} + \varepsilon & \text{for some } \theta_0 \text{ and } \theta_1 \\ \\ H_a : \text{Ozone} = q_\tau(\text{Temperature}) + \varepsilon & \text{for some smooth function } q_\tau \end{cases} \tag{5.9}$$

The lack-of-fit test developed along this chapter will be compared with its natural competitors thanks to this univariate real situation. Table 5.3 contains p-values from the proposed lack-of fit test (denoted by NT), the test proposed by He and Zhu (2003) (denoted by HZ) and the test proposed by Zheng (1998) (denoted by Z) for testing (5.9) and considering different values of the quantile of interest. According to the proposed test and Zheng (1998)'s test, the null hypothesis is rejected for $\tau = 0.25$ and $\tau = 0.5$, while He and Zhu (2003)'s test takes a chance on accepting the null hypothesis.

[1] In this case, we chose a lack-of-fit test problem with a one-dimensional explanatory variable to simplify the application of the new test.

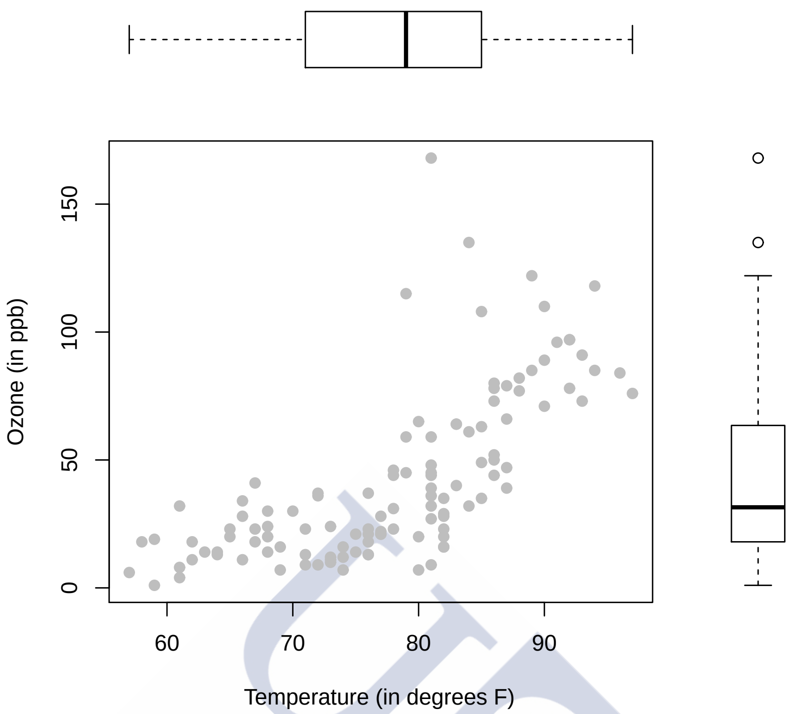


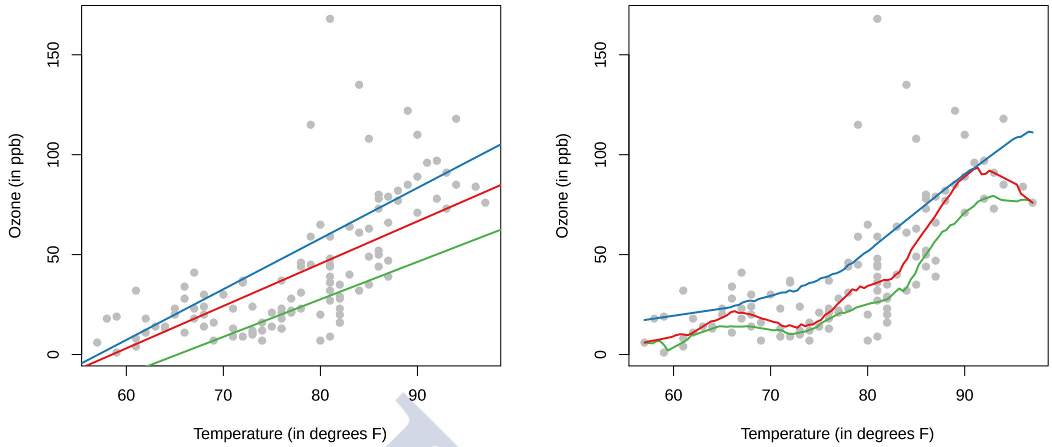
Figure 5.9: Scatterplot of the variables Ozone and Temperature joint with a boxplot representation of each variable.

In order to clarify this situation, Figure 5.10 represents the scatterplot together with fitted quantile models associated with a parametric and a nonparametric approach for different quantiles of interest: $\tau = 0.25$, $\tau = 0.5$ and $\tau = 0.75$. Part (a) of Figure 5.10 represents the fitted models under the null hypothesis, that is, linear models, whereas part (b) of Figure 5.10 represents the local linear fits. The plug-in bandwidth proposed in Chapter 3 was used here ir order to obtain the nonparametric estimation.

		$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$
Lack-of-fit tests:	NT	0.0001	0.0000	0.0821
	HZ	0.9361	0.0187	0.9177
	Z	0.0003	0.0000	0.0006

Table 5.3: p-values obtained by the proposed lack-of fit test (denoted by NT), the test proposed by He and Zhu (2003) (denoted by HZ) and the test proposed by Zheng (1998) (denoted by Z) for testing (5.9) for different values of the quantile of interest.

According to Figure 5.10, it seems clear that the quantile models associated with $\tau = 0.25$ and $\tau = 0.5$ are not linear, as the proposed test and Zheng (1998)'s test conclude. Moreover, the situation is not so clear when we talk about $\tau = 0.75$. In this case, the proposed test would accept the null hypothesis if the nominal level is $\alpha = 0.05$ but not for $\alpha = 0.10$. Figure 5.10 does not provide too much information because the parametric and nonparametric fits associated with $\tau = 0.75$ are quite similar.



(a) Linear fitted models (b) Nonparametric fitted models

Figure 5.10: Scatterplot of the variables Ozone and Temperature together with fitted quantile models associated with a parametric (Part (a)) and a nonparametric (Part (b)) approach for different quantiles of interest: $\tau = 0.25$ (green line), $\tau = 0.5$ (red line) and $\tau = 0.75$ (blue line).

5.5 Conclusions

Along this chapter we have proposed a new method for testing parametric quantile regression models versus nonparametric alternatives. The new lack-of-fit test is based on a likelihood ratio test in logistic regression context where the response variable will be determined by the sign of the quantile regression residuals. Then, a large number of functions of the quantile regression covariates are considered as explanatory variables of the logistic model. Moreover, a wild bootstrap mechanism to approximate critical values was presented.

If the covariate associated with the quantile regression model is multivariate, then we have proposed using projections in order to avoid working with very large dimension in the logistic context. This approach connects with the strategy used along Chapter 4, but now the aim is not to deal with the curse of the dimensionality.

Finally, thanks to a Monte Carlo simulation study and an application to real data, we found a promising performance of the new test in comparison with some natural competitors available in the literature. As future work, we would like to justify the good behaviour of the proposed test from a theoretical point of view. Moreover we would like a more detailed suggestion about how to select appropriates parameters p and λ.

USC

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Resumen en castellano
Summary in Spanish

Aunque la regresión en media, ajustada por el método de mínimos cuadrados, ha alcanzado la mayor difusión en la Estadística del siglo XX, resulta muy llamativo observar que las ideas de regresión cuantil fueron anteriores a los procedimientos basados en los mínimos cuadrados. Así, mientras el inicio de la regresión por mínimos cuadrados se puede datar en el año 1805 por el trabajo de Legendre, a mediados del siglo XVIII Boscovich ya ajustó datos sobre la elipticidad de la Tierra mediante procedimientos de regresión cuantil.

El método de mínimos cuadrados gozó de la ventaja que le proporcionaba la existencia de expresiones cerradas para la estimación, la sencillez de los argumentos de probabilidad y ciertos resultados de optimalidad. Aún así, siempre pesaba la duda sobre las hipótesis del modelo, y sobre la necesidad de una descripción más completa y flexible de la realidad.

Los métodos de regresión cuantil encontraron un gran desarrollo desde el surgimiento de la Estadística Robusta, que alcanzó una gran expansión a principios de los años 80. El libro de Huber (1981) o el de Hampel et al. (1986) son buenas recopilaciones de las aportaciones que hicieron sus autores a la Teoría de la Robustez, cuyos conceptos siguen siendo aplicados hoy en día a los métodos estadísticos modernos.

Los procedimientos de regresión cuantil que se tratarán en esta tesis, aunque comparten propiedades y conceptos de la Teoría de la Robustez, están basados en modelos de regresión de la función cuantil condicionada, por lo que su objetivo principal se centra en obtener una descripción más detallada de la distribución condicional. Estos modelos de regresión cuantil han sido desarrollados principalmente en los trabajos de Koenker de los años 80. Una buena recopilación de los procedimientos bajo este enfoque se encuentra en el libro de Koenker (2005).

A día de hoy la regresión cuantil es un tema de máximo interés de los investigadores en Estadística, que están adaptando gran parte de las técnicas de inferencia relacionadas con la regresión a los modelos de regresión cuantil. La razón es que los modelos de regresión cuantil permiten una descripción más detallada del comportamiento de la variable respuesta, se adaptan a situaciones bajo condiciones más generales de la distribución del error, gozan de propiedades de robustez y permiten abordar problemas de regresión con datos complejos (como por ejemplo, los datos censurados), en muchos casos en mejores condiciones que una regresión en media.

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

El objetivo principal de esta tesis doctoral es emplear los métodos no paramétricos para obtener nuevos procedimientos de inferencia en el contexto de los modelos de regresión cuantil. A continuación exponemos un breve resumen de cada un de los capítulos que constituyen esta tesis doctoral, haciendo mención de los principales avances obtenidos en cada uno de ellos.

Capítulo 1: Introducción

A lo largo del Capítulo 1 se desarrolla una pequeña introducción a los conceptos básicos asociados a la regresión cuantil. Empezamos estableciendo el concepto del cuantil asociado a un cierto orden τ.

Definición. Dada cualquier variable aleatoria X, para cada $0 < \tau < 1$ se puede definir el **cuantil de orden** τ, que denotaremos por c_τ, como el valor que verifica que:

$$\mathbb{P}(X \leq c_\tau) \geq \tau$$
$$\mathbb{P}(X \geq c_\tau) \geq 1 - \tau$$

Aparece así la **función cuantil** de una distribución de probabilidad, que se define como la inversa de la función de distribución.

Lo realmente importante es que se puede expresar el problema de la búsqueda de los cuantiles muestrales como un problema de optimización de la forma

$$\min_{c \in \mathbb{R}} \sum_{i=1}^{n} \rho_\tau(X_i - c)$$

donde $\mathcal{X} = \{X_1, \ldots, X_n\}$ representa una muestra aleatoria simple de la variable X y ρ_τ representa la **función de pérdida cuantílica** que está determinada por la siguiente función lineal definida a trozos

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

El problema anterior se puede reformular como un problema de programación lineal, es decir, se trata de minimizar una función lineal en un conjunto poliédrico de restricciones.

El razonamiento anterior se puede extender al problema de regresión. Supongamos entonces que nos interesa explicar una variable aleatoria Y escalar en función de ciertas covariables que denotaremos por $X \in \mathbb{R}^d$ de las cuales conocemos una muestra $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$. Entonces si la función cuantil condicional viene dada por $q_\tau(x) = \theta'_\tau x$, consideraremos $\widehat{\theta}_\tau$ el elemento que resuelva el siguiente problema:

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

donde $P_i = (1, X_i)$; siendo éste es el punto de partida de la idea desarrollada por Koenker and Bassett (1978). Esto nos permitiría establecer un modelo de la forma

$$Y_i = \theta'_\tau P_i + \varepsilon_i$$

donde los errores verificarían que $\mathbb{P}(\varepsilon_i \leq 0 \mid X = X_i) = \tau$, es decir, el cuantil condicional de orden τ del error es cero.

Hasta este momento, hemos conseguido expresar la búsqueda de un cuantil muestral como la solución de un problema de programación lineal y extendimos este razonamiento al contexto de la regresión cuantil. Este hecho nos permite proponer métodos para el cálculo de los estimadores de regresión cuantil.

Barrodale y Roberts (1973) proponen una simplificación de la forma estándar del método del Simplex para el resolver el problema del cálculo de los estimadores en el caso de la regresión en mediana, donde la función de pérdida sería el valor absoluto. Posteriormente, Koenker y D'Orey (1987) extendieron este razonamiento a cualquier cuantil $0 < \tau < 1$.

Debemos tener en cuenta que los estimadores asociados a la regresión cuantil no tienen expresión explícita por lo que sería necesario recurrir a expresiones asintóticas como la representación propuesta por Bahadur (1966). Además, ni siquiera la distribución de estos estimadores es conocida bajo hipótesis de normalidad como en el caso de la regresión en media estimada por mínimos cuadrados.

De todas formas si se verifican resultados sobre la distribución asintótica de los estimadores de regresión cuantil como el siguiente:

Teorema. *Consideremos un modelo lineal para explicar una variable respuesta escalar Y en función de una variable explicatica X de la forma:*

$$Y_i = \theta_\tau' P_i + \varepsilon_i \qquad \text{con } i = 1, \cdots, n$$

donde los errores verifican que $\mathbb{P}(\varepsilon_i \leq 0 \mid X = X_i) = \tau$. Supongamos además, que se verifican las siguientes condiciones:

Condición A1. *Las funciones de distribución condicionales F_i (de Y_i condicionada a X_i) son absolutamente continuas y con densidades f_i continuas y uniformemente acotadas lejos de 0 e ∞ en los cuantiles condicionales $c_i(\tau)$.*

Condición A2. *Existen matrices D_0 y $D_1(\tau)$ definidas positivas tales que*

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

se tiene que:

$$\sqrt{n}\left(\widehat{\theta}_\tau - \theta_\tau\right) \xrightarrow{d} N\left(0, \tau(1-\tau) D_1^{-1}(\tau) D_0 D_1^{-1}(\tau)\right)$$

donde $N(\mu, \sigma^2)$ representa una distribución Gaussiana con media μ y varianza σ^2.

Surge entonces la duda de cómo estimar la matriz de covarianzas asintótica de los estimadores de la regresión cuantil. La precisión de la regresión cuantil depende de la inversa de la función de densidad evaluada en el cuantil que nos interesa, a dicha función Tukey (1965) la denominó **función "sparsity"** que viene dada por:

$$s(\tau) = \frac{1}{f(F^{-1}(\tau))}.$$

Teniendo ésto en cuenta, concluimos que las estimaciones del cuantil serán más precisas cuantas más observaciones aparezcan en torno al cuantil que nos interesa. Por el contrario, si en un entorno del cuantil que estamos estudiando no existen muchas observaciones, los resultados que obtengamos no serán muy precisos.

En el caso de que los errores de la regresión sean independientes e idénticamente distribuidos, la función "sparsity" juega un papel análogo al de la desviación típica en el caso de la regresión por mínimos cuadrados en este mismo escenario independiente e idénticamente distribuido.

Si derivamos la expresión $F(F^{-1}(t)) = t$ nos damos cuenta que la función "sparsity" es la derivada de la función cuantil

$$\frac{d}{dt}F^{-1}(t) = s(t)$$

Esto nos proporcionaría un modo de estimar la función "sparsity" de la siguiente forma:

$$\widehat{s}(t) = \frac{\widehat{F}_n^{-1}(t+h) - \widehat{F}_n^{-1}(t-h)}{2h}$$

donde $\widehat{F}_n^{-1}$ estima la función F^{-1} y h es una sucesión de elementos que tienden a cero. Bofinger (1975) y Hall y Sheather (1988) propusieron diferentes sucesiones h para estimar la función "sparsity".

Ya hemos mencionado que la regresión cuantil tiene una relación muy próxima con la Estadística Robusta, pues comparte algunas propiedades y métodos con ella. En concreto, el estimador de regresión cuantil corrige los problemas de falta de robustez por atípicos en la variable respuesta, mientras que conserva los problemas relacionados con el apalancamiento, esto es, por atípicos en las variables explicativas. Para mejorar las propiedades de robustez también en este sentido, sería necesario considerar M-estimadores generalizados, regresión por mínima mediana de cuadrados o regresión profunda.

Capítulo 2: Técnicas de predicción basadas en modelos de regresión cuantil

Un gran número de trabajos estadísticos giran en torno al hecho de realizar predicciones para una determinada variable de interés. En esta línea, a lo largo del Capítulo 2 hemos propuesto un nuevo método de estimación de intervalos de predicción basado en un modelo de regresión en mediana y un plan de remuestreo bootstrap. Los principales resultados obtenidos a lo largo de este capítulo pueden verse en Conde-Amboage et al. (2016).

Dada una muestra aleatoria simple $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$, en adelante muestra de entrenamiento, de dos variables $X = (X^{(1)}, \cdots, X^{(d)}) \in \mathbb{R}^d$ e $Y \in \mathbb{R}$, consideraremos el siguiente modelo de regresión cuantil

$$Y_i = \theta' P_i + \varepsilon_i \tag{R.1}$$

donde $\theta = (\theta_0, \theta_1, \cdots, \theta_d) \in \mathbb{R}^{d+1}$, $P_i = (1, X_i)$ y ε representa el error del modelo. A la vista de este escenario, nos planteamos definir un intervalo de predicción para una nueva observación Y_{i_0} de la variable respuesta.

Definición. Un **intervalo de predicción** para un valor Y_{i_0} es un intervalo que se espera que contenga el valor verdadero Y_{i_0} con una probabilidad (presumiblemente) alta $(1-\alpha)$, conocida habitualmente como nivel de confianza. Denotaremos el intervalo de predicción por (L_{i_0}, U_{i_0}), donde los extremos L_{i_0} y U_{i_0} se obtienen como funciones de la muestra de entrenamiento y del valor del predictor P_{i_0}. Entonces se tendría que

$$P\left(Y_{i_0} \in (L_{i_0}, U_{i_0})\right) = 1 - \alpha.$$

A lo largo del Capítulo 2, hemos definido un nuevo método de estimación de intervalos de predicción de la forma:

$$\left(\widehat{Y}_{i_0,\tau=0.5} + G^{\star -1}(\alpha/2)\ ,\ \widehat{Y}_{i_0,\tau=0.5} + G^{\star -1}(1-\alpha/2)\right),$$

donde el plan de remuestreo es el siguiente:

1. Dado un modelo de regresión de la forma $Y = \theta_\tau' X + \varepsilon$ estimamos el parámetro θ_τ ajustando un modelo de regresión asociado al cuantil $\tau = 0.5$, que denotaremos por $\widehat{\theta}_{\tau=0.5}$. Supongamos que queremos calcular intervalos de predicción para una nueva observación que denotaremos por (X_{i_0}, Y_{i_0}).

2. Generar los pesos w_i procedentes de una distribución que satisfaga ciertas propiedades que proponen Feng et al. (2011). A partir de estos pesos calcular $\varepsilon_i^\star = w_i|r_i|$ donde $|\cdot|$ denota el valor absoluto y $r_i = Y_i - \widehat{\theta}'_{\tau=0.5} P_i$. Sortearemos un residuo r_{i_0} que estará asociado a la nueva observación X_{i_0} a partir de la función de distribución condicional del error que será de la forma

$$F(r|X = X_0) = \sum_{i=1}^{n} \mathbb{I}(r_i \le r) \frac{K\left(\frac{\widehat{\theta}'_{\tau=0.5}P_i - \widehat{\theta}'_{\tau=0.5}P_{i_0}}{h}\right)}{\sum_{j=1}^{n} K\left(\frac{\widehat{\theta}'_{\tau=0.5}P_j - \widehat{\theta}'_{\tau=0.5}P_{i_0}}{h}\right)}.$$

   El parámetro ventana h considerado es de la forma $cn^{-1/5}$ donde c es una contante que depende de diversas cantidades desconocidas, y $n^{-1/5}$ es la tasa tradicional asociada a estimadores no paramétricos tipo Nadaraya-Watson. Además, generamos el peso correspondiente w_{i_0} obteniendo así $\varepsilon_{i_0}^\star = w_{i_0}|r_{i_0}|$. De este modo tendríamos

$$Y_i^\star = \widehat{\theta}'_{\tau=0.5} P_i + \varepsilon_i^\star \qquad i \in \{1, \ldots, n\}$$
$$Y_{i_0}^\star = \widehat{\theta}'_{\tau=0.5} P_{i_0} + \varepsilon_{i_0}^\star$$

3. Ajustar un modelo de regresión asociado al cuantil $\tau = 0.5$ a las remuestras bootstrap y denotar por $\widehat{\theta}^\star_{\tau=0.5}$ a las estimaciones del parámetro $\widehat{\theta}_{\tau=0.5}$. Esto nos permitirá computar las diferencias

$$D^\star = Y_{i_0}^\star - \widehat{\theta}^{\star\prime}_{\tau=0.5} P_{i_0}$$

4. Repetir los pasos 2-3 B veces y calcular los cuantiles de orden $z_1 = \alpha/2$ y $z_2 = 1 - \alpha/2$ que denotamos por $G^{\star-1}(z)$ siendo $(1 - \alpha)$ en nivel de cobertura del intervalo de predicción que queremos calcular.

Las propiedades de convergencia del nuevo método propuesto son consecuencia de argumentos similares a aquellos dados en Stine (1985). Así, los errores bootstrap de predicción pueden ser expresados como

$$D^\star = \varepsilon^\star_{i_0} + \left(\widehat{\theta}_{\tau=0.5} - \widehat{\theta}^\star_{\tau=0.5}\right)' P_{i_0}.$$

Dado que el sumando de la derecha se genera de forma independiente, la distribución bootstrap del error de predicción se puede expresar como la convolución de dos distribuciones:

$$G^\star = \widehat{F}_{i_0} * Z^\star$$

donde $\widehat{F}_{i_0}$ representa la distribución de $\varepsilon^\star_{i_0}$ y $Z^\star$ denota la distribución del segundo sumando. Feng et al. (2011) obtuvo la consistencia de $Z^\star$ bajo el mecanismo bootstrap propuesto en este caso. Hall y Yao (2005) proporcionó la consistencia del estimador $\widehat{F}(r|P_{i_0})$, donde las técnicas de suavización se aplican a las variables proyectadas, como en nuestro caso. Dado que $\widehat{F}_{i_0}$ se obtiene gracias a $\widehat{F}(r|P_{i_0})$ teniendo en cuenta los multiplicadores introducidos por Feng et al. (2011), la validez del procedimiento bootstrap se deriva de la consistencia de $\widehat{F}(r|P_{i_0})$.

Hemos diseñado un completo estudio de simulación que nos permite comparar el nuevo método propuesto frente a diversos métodos disponibles en la literatura. En base a dicho estudio, podemos afirmar que los intervalos de predicción basados en cuantiles tienen asociados mejores aproximaciones del nivel de confianza y tamaños más reducidos que los clásicos intervalos de predicción en media. Además, se respalda el hecho de que cuando los residuos no verifican las hipótesis clásicas de los modelos de regresión en media (homocedasticidad y normalidad de los residuos), la regresión cuantil proporciona mejores resultados.

Finalmente, el comportamiento de los diversos métodos de estimación de intervalos de predicción, han sido aplicados a una notable base de datos medioambientales. Dichos datos están asociados a La Unidad de Producción Térmica (U.P.T.) de As Pontes, que está situada en el municipio de As Pontes de García Rodríguez (en el noreste de la provincia de A Coruña), que constituye uno de los centros productivos propiedad de Endesa Generación S.A. en la Península Ibérica. En dicho escenario, uno de los problemas que se plantea es poder predecir los niveles de óxidos de nitrógeno (NO_x), a partir de la información que se recibe en continuo de las estaciones de muestreo así como de la información pasada de dichas medidas.

El nuevo método de estimación de intervalos de predicción que hemos propuesto, proporciona mejores intervalos de predicción para las mediciones de NO_x en términos de niveles de cobertura. Estos buenos resultados están justificados por las características de estos datos, como la heterocedasticidad o la no normalidad de los residuos.

Capítulo 3: Un selector plug-in del parámetro ventana asociado a un modelo de regresión cuantil no paramétrico

Consideremos un modelo de regresión cuantil de la forma

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

donde ε representa el error del modelo cuyo cuantil condicional de orden τ es cero. Dada $\{(X_1, Y_1), \cdots, (X_n, Y_n)\}$ una muestra aleatoria simple de dos variables $(X, Y) \in \mathbb{R}^2$, se puede definir un estimador no paramétrico del cuantil condicional como $\widehat{q}_\tau(x) = \widehat{a}$ siendo $\widehat{a}$ y $\widehat{b}$ los elementos que minimizan la expresión

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

donde K representa la función núcleo y h_τ se conoce como parámetro ventana.

A la hora de abordar esta clase de modelos de regresión cuantil, será fundamental la elección del parámetro ventana h_τ. A la vista de la expresión del error cuadrático medio del estimador $\widehat{q}_\tau$ desarrollada por Fan et al. (1994), se propone la siguiente ventana óptima:

$$h^5_{\text{AMISE},\tau} = \frac{R(K)\, \tau(1-\tau)}{n\, \mu_2(K)^2 \int q_\tau^{(2)}(x)^2\, g(x)dx} \int \frac{1}{f(q_\tau(x)|x)^2} dx \tag{R.2}$$

donde g denota la densidad marginal de la variable explicativa X, n es el tamaño de muestra, $q_\tau^{(i)}$ representa la derivada de orden i de la función de regresión q_τ y

$$\mu_2(K) = \int u^2 K(u)du \qquad R(K) = \int K^2(u)du$$

son constantes que dependen de la función núcleo. Finalmente, destaca la inversa de la densidad de la variable respuesta evaluada en el cuantil de interés, es decir, la función "sparsity" que denotaremos por $s_\tau(x)$.

Existen diferentes métodos de selección del parámetro ventana h_τ en la literatura. Yu y Jones (1998) propusieron un selector plug-in de dicho parámetro basándose en un selector del parámetro ventana en el contexto de la regresión en media, como el propuesto por Ruppert et al. (1995). Por otra parte, Abberger (1998) propuso un selector basado en el método de validación cruzada adaptado al contexto de la regresión cuantil.

A lo largo del Capítulo 3, nuestro objetivo es proponer un nuevo selector plug-in del parámetro ventana basado en estimadores no paramétricos de la función "sparsity" y de la curvatura del modelo de regresión cuantil. Los resultados obtenidos a lo largo del Capítulo 3 se recogen en Conde-Amboage y Sánchez-Sellero (2017). Detallamos a continuación como construimos estimaciones no paramétricas de la "sparsity" y de la curvatura.

<u>Estimación no paramétrica de la "sparsity"</u>

A la hora de proponer un estimador para la función "sparsity" debemos recordar que la "sparsity" no es más que la derivada de la función cuantil q_τ con respecto al parámetro τ.

Teniendo esto en cuenta, proponemos el siguiente estimador:

$$\widehat{s}_{\tau,d_s,h_s}(x) = \frac{\widehat{q}_{\tau+d_s,h_s}(x) - \widehat{q}_{\tau-d_s,h_s}(x)}{2d_s} \tag{R.3}$$

siendo $\widehat{q}_{\tau+d_s,h_s}$ y $\widehat{q}_{\tau-d_s,h_s}$ estimaciones no paramétricas de las funciones de regresión en el punto x asociadas a los cuantiles $(\tau + d_s)$ y $(\tau - d_s)$ donde h_s denota el parámetro de suavización.

Para estudiar las propiedades del estimador (R.3) será fundamental un resultado probado por Fan et al. (1994) que establece la convergencia del estimador lineal local en el contexto de la regresión cuantil. Hemos obtenido el error cuadrático medio (ECM) del cuadrado del estimador (R.3) que tendría la siguiente expresión:

$$\begin{aligned}\mathrm{ECM}\left(\int \widehat{s}^2_{\tau,d_s,h_s}(x)\,dx\right) \cong & \left[\frac{1}{nd_sh_s}\int a(x)\,dx + d_s^2\int b(x)\,dx + h_s^2\int c(x)\,dx\right]^2 \\ & + \frac{1}{nd_s}\int d(x)\,dx + \frac{1}{n^2d_s^2h_s}\int e(x)\,dx\end{aligned} \tag{R.4}$$

donde

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

donde $*$ representa la convolución y $s_\tau^{(2,\tau)}(x) = \frac{\partial^2}{\partial\tau^2}s_\tau(x)$.

Por lo tanto, será necesario determinar los parámetros ventana h_s y d_s. Proponemos como selectores de dichos parámetros los elementos que hagan mínimo la expresión del error cuadrático medio dada en (R.4). Para ello utilizamos métodos numéricos de optimización.

Estimación no paramétrica de la curvatura

Dada $(X_1,Y_1),\cdots,(X_n,Y_n)$ una muestra aleatoria simple de las variables (X,Y), podemos ajustar un modelo polinómico local de la forma

$$\sum_{i=1}^{n}\rho_\tau\big(Y_i - \theta_0 - \theta_1(X_i-x) - \theta_2(X_i-x)^2 - \theta_3(X_i-x)^3\big)K\left(\frac{X_i-x}{h_c}\right)$$

para ajustar el cuantil condicional de orden τ de la variable respuesta Y dado $X = x$ que venimos denotando por $q_\tau(x)$ donde h_c denota el parámetro ventana en este caso. En este caso, se tendría que $q^{(2)}_{\tau,h_c}(X_i) = 2\theta_2$ para $i = 1,\cdots,n$. Por lo tanto, podríamos considerar el siguiente estimador para la curvatura:

$$\widehat{\vartheta}_{\widehat{h}_c} = \widehat{\int q_\tau^{(2)}(x)^2 g(x)dx} = \frac{1}{n}\sum_{i=1}^{n}\widetilde{q}^{(2)}_{\tau,\widehat{h}_c}(X_i)^2$$

A la vista de la ecuación anterior, hemos propuesto una ventana piloto para la estimación de la curvatura extendiendo el resultado de Fan et al. (1994) a orden 3. Dicha ventana piloto vendría dada por

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

donde

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

siendo

$$\delta_1 = \frac{1}{6}(\alpha_{31}\; \mu_4(K) + \alpha_{33}\; \mu_6(K))$$

$$\delta_2 = 4\left(\alpha_{31}^2 \int K^2(v)\; dv + \alpha_{33}^2 \int v^4\; K^2(v)\; dv + 2\alpha_{31}\alpha_{33} \int v^2\; K^2(v)\; dv\right)$$

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

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

$$\mu_i(K) = \int u^i K(u) du.$$

Enumeramos a continuación los pasos a seguir para obtener el nuevo selector plug-in de la ventana asociada a un modelo de regresión cuantil lineal local.

1. Obtener estimaciones piloto de la integral de la "sparsity" al cuadrado (que denotamos por $\widehat{s}^2_{\tau,B}$) y de la cantidad $\int q_\tau^{(2)}(x) q_\tau^{(4)}(x) g(x) dx$ que denotaremos por $\widehat{\vartheta}_{B,2}$.

2. Estimar la curvatura utilizando la siguiente ventana piloto:

$$\widehat{h}_c = C(K) \left(\frac{\tau(1-\tau)\widehat{s}^2_B}{|\widehat{\vartheta}_{B,2}|\; n} \right)^{1/7}$$

Consideraremos entonces el siguiente estimador de la curvatura:

$$\widehat{\vartheta}^{\alpha}_{\widehat{h}_c} = \frac{1}{n} \sum_{i=1}^{n} \widetilde{q}^{(2)}_{\tau,\widehat{h}_c}(X_i)^2\; \mathbb{I}\left((1-\alpha)a + \alpha b < X_i < \alpha a + (1-\alpha)b\right)$$

para cierto valor α pequeño, donde la variable explicativa toma valores en el intervalo $[a, b]$. El hecho de truncar la variable explicativa nos permite reducir la gran variabilidad asociada a los modelos polinómicos locales en la frontera. Consideraremos $\alpha = 0.05$.

3. Estimar la "sparsity" utilizando las ventanas piloto $\widehat{h}_s$ y $\widehat{d}_s$ obtenidas en base a la expresión del error cuadrático medio del estimador de la integral de la "sparsity" al cuadrado dado en (R.4). Denotaremos esta estimación por $\widehat{s}^2_{\tau,\widehat{h}_s,\widehat{d}_s}$.

4. Finalmente, el selector de la ventana asociada al método plug-in tendría la siguiente forma:

$$\widehat{h}_{NP} = \left(\frac{R(K)\,\tau(1-\tau)\,\widehat{s}^2_{\tau,\widehat{h}_s,\widehat{d}_s}}{\mu_2(K)^2\,\widehat{\vartheta}^{0.05}_{\widehat{h}_c}\,n} \right)^{1/5}.$$

Gracias a un estudio de simulación, se muestra que el nuevo selector del parámetro ventana muestra un menor error cuadrático medio integrado comparado con los competidores disponibles en la literatura, tanto en escenarios homocedásticos como heterocedásticos. Además, hemos desarrollado un paquete de R llamado `BwQuant` que permite acercar las nuevas herramientas estadísticas desarrolladas a la comunidad científica.

Capítulo 4: Un contrate de bondad de ajuste aplicado a covariables de alta dimensión

Dado un modelo de regresión, surge de forma natural la necesidad de verificar que dicho modelo se ajusta bien al conjunto de datos con el que estamos trabajando. Surgen de este modo los contrastes de bondad de ajuste en el contexto de la regresión que se abordan a lo largo del Capítulo 4. Los principales resultados obtenidos a lo largo de este capítulo pueden verse en Conde-Amboage et al. (2015).

Consideraremos un modelo de regresión asociado a un cuantil $\tau \in (0,1)$ de interés de la forma:

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

siendo ε el error desconocido que debe verificar que $\mathbb{P}(\varepsilon \le 0|X) = \tau$ donde $(X,Y) \in \mathbb{R}^{d+1}$. En este caso, nuestro objetivo será realizar el siguiente contraste de hipótesis:

$$\begin{cases} H_0 : q_\tau \in \mathcal{Q}_\theta = \{q_\tau(\cdot,\theta) \;:\; \theta \in \Theta \subset \mathbb{R}^q\} \\ H_a : q_\tau \notin \mathcal{Q}_\theta \end{cases} \tag{R.5}$$

Existen propuestas en la literatura que afrontar el contraste de bondad de ajuste dado en (R.5). Destacamos la propuesta de He y Zhu (2003) que extienden al contexto cuantil el conocido test basado en la función de regresión integrada propuesto por Stute (1997) en el contexto de la clásica regresión en media. Por otra parte, Zheng (1998) afronta el mismo problema utilizando métodos de suavización, extendiendo de esta forma su anterior propuesta Zheng (1996) al contexto cuantil.

Es evidente que el aumento de la dimensión de la variable explicativa $X \in \mathbb{R}^d$ afecta a los contrastes de bondad de ajuste, que es lo que se conoce como **desastre de la dimensionalidad**. A lo largo del Capítulo 4 se presenta un nuevo test que surge con el objetivo de evitar el desastre de la dimensionalidad en la línea del test propuesto por

Escanciano (2006). Es decir, vamos a proponer un contraste de bondad de ajuste para el contexto de la regresión cuantil basado en el uso de proyecciones sobre la variable explicativa X.

Si el parámetro θ_τ se supone conocido, entonces el test estará basado en el siguiente proceso empírico:

$$R_n(\beta, u) = n^{-1/2} \sum_{i=1}^{n} \psi(\varepsilon_i)\, q_\tau^{(1)}(X_i, \theta_\tau)\, \mathbb{I}(\beta' X_i \leq u)$$

donde $\psi_\tau(r) = \tau\mathbb{I}(r > 0) + (\tau - 1)\mathbb{I}(r < 0)$ representa la derivada de la función de pérdida cuantílica que viene dada por $\rho_\tau = \tau r\mathbb{I}(r > 0) + (\tau - 1)r\mathbb{I}(r < 0)$, $q_\tau^{(1)}(x, \theta) = \frac{\partial}{\partial\theta}\, q_\tau(x, \theta)$ denota la derivada de la función de regresión, $\mathbb{I}$ representa la función indicadora de un evento y $\varepsilon = Y - q_\tau(X, \theta_\tau)$.

En la práctica no conoceremos el parámetro θ_τ y será por tanto necesario estimarlo. Entonces el proceso empírico que utilizaremos en la práctica vendría dado por

$$R_n^1(\beta, u) = n^{-1/2} \sum_{i=1}^{n} \psi(r_i)\, q_\tau^{(1)}(X_i, \widehat{\theta}_\tau)\, \mathbb{I}(\beta' X_i \leq u)$$

donde $r_i = Y_i - q_\tau(X_i, \widehat{\theta}_\tau)$ representan los residuos del modelo y $\widehat{\theta}_\tau$ es una estimación de θ_τ.

Llegados a este punto debemos elegir una norma en función de la cual definir el estadístico de contraste. Consideraremos la norma de Cramér-von Mises como consecuencia de sus claras ventajas computacionales. Con lo cual, el estadístico de contraste vendría dado por:

$$T_n = \text{mayor autovalor de } \int_\Pi R_n^1(\beta, u)[R_n^1(\beta, u)]' F_{n,\beta}(du) d\beta$$

donde $F_{n,\beta}(u)$ representa la función de distribución empírica de las variables explicativas proyectadas $\{\beta' X_1, \cdots, \beta' X_n\}$ y $d\beta$ representa la densidad uniforme en la esfera unidad en el espacio $\mathbb{R}^d$.

El proceso empírico R_n^1 es similar al propuesto por Escanciano (2006), pero con dos diferencias destacables: la función de pérdida en este caso es la función de pérdida cuantílica y hemos introducido la derivada de la función de regresión siguiendo la propuesta de He y Zhu (2003). Destacar también que el proceso R_n^1 dista del considerado por He y Zhu (2003) en dos aspectos fundamentales: la consideración de proyecciones de la variable explicativa y la no suposición de homocedasticidad.

La distribución límite de R_n bajo la hipótesis nula simple $H_0 : q_\tau \in \mathcal{Q}_\theta$, viene dada por

$$R_n \xrightarrow{d} R_\infty$$

siendo R_∞ un proceso gaussiano de media cero y con matriz de covarianzas

$$K(x_1, x_2) = E\left[q_\tau^{(1)}(X, \theta_\tau)\, q_\tau^{(1)\prime}(X, \theta_\tau)\mathbb{I}(\beta_1' X \leq u_1)\, \mathbb{I}(\beta_2' X \leq u_2)\right]$$

donde $x_1 = (\beta_1', u_1)'$ y $x_2 = (\beta_2', u_2)'$.

Por otra parte, se ha probado la consistencia del test frente a alternativas que convergen a H_0 con tasa $\sqrt{n}$. Es decir, si los datos proceden de un modelo de la forma

$$Y_i = q_\tau(X_i, \theta_\tau) + n^{-1/2} h(X_i) + \varepsilon_i \qquad i \in \{1, \dots, n\}$$

donde la función h representa la desviación respecto de la hipótesis nula. Entonces, bajo ciertas condiciones de regularidad, se tiene que:

$$\begin{aligned} R_n^1(\beta, u) &= n^{-1/2} \sum_{i=1}^{n} \psi(\varepsilon_i) \left[\mathbb{I}(\beta' X_i \le u) - S(\beta, u)\, S^{-1} \right] q_\tau^{(1)}(X_i, \theta_\tau)' \\ &+ \mathbb{E}\left[f(0|X)\, h(X)\, q_\tau^{(1)}(X, \theta_\tau)'\, \mathbb{I}(\beta' X \le u) \right] \\ &- S(\beta, u)\, S^{-1} \mathbb{E}\left[f(0|X)\, h(X)\, q_\tau^{(1)}(X, \theta_\tau)' \right] + o_p(1) \end{aligned}$$

uniformemente en (β, u), siendo:

$$S = \mathbb{E}[f(0|X)\, q_\tau^{(1)}(X, \theta_\tau)\, q_\tau^{(1)}(X, \theta_\tau)']$$

$$S(\beta, u) = \mathbb{E}[f(0|X)\, q_\tau^{(1)}(X, \theta_\tau)\, q_\tau^{(1)}(X, \theta_\tau)'\, \mathbb{I}(\beta' X \le u)].$$

Una vez que hemos planteado el estadístico de contraste, se debe afrontar el problema de cómo llevar a cabo el calibrado del test. En este punto, se presentan diversas opciones. En primer lugar, podríamos utilizar la distribución límite del estadístico pero esto implicaría estimar la varianza límite que acarrea la complicada estimación de diversas cantidades desconocidas. Otra opción sería utilizar la representación del proceso empírico bajo la hipótesis nula compuesta que hemos detallado anteriormente, es decir, se plantearía un bootstrap de los multiplicadores como el considerado por He y Zhu (2003). En general, este segundo método tiene asociado mejores resultados que la aplicación directa de la distribución límite, pero todavía sería necesario estimar diversas cantidades desconocidas como la densidad condicional del error evaluada en cero (por ejemplo, Escanciano y Goh (2014) consideraron esta segunda línea).

Teniendo en cuenta todos estos argumentos, proponemos calibrar el test a través de un procedimiento wild bootstrap adaptado al contexto cuantil desarrollado por Feng et al. (2011). La mayor ventaja de dicho método es que nos permite considerar escenarios heterocedásticos sin necesidad de estimar ninguna cantidad desconocida involucrada en las representaciones del proceso empírico R_n^1.

Por otra parte, se ha presentado un extenso estudio de simulación que muestra el buen comportamiento del nuevo contraste de bondad de ajuste. Se concluye entonces que el nuevo test muestra un buen ajuste del nivel independientemente del cuantil de interés considerado. Además, es generalmente más potente que sus competidores naturales, particularmente cuando la dimensión de la covariable aumenta. Es importante destacar que la nueva propuesta también ha sido aplicada con éxito en contextos heterocedásticos.

Finalmente, el test propuesto ha sido aplicado a un conjunto de datos reales. Dicha aplicación ha servido para testear un modelo ampliamente utilizado en el contexto económico que permite describir la evolución del producto interior bruto en función de diversas variables explicativas.

Capítulo 5: Un contraste de bondad de ajuste para modelos de regresión cuantil basado en regresión logística

A lo largo del Capítulo 5 afrontamos el mismo problema que el desarrollado en el Capítulo 4 pero desde un punto de vista completamente diferente. Es decir, nuestro objetivo será llevar a cabo el contraste de hipótesis dado en (R.5). El test propuesto está basado en el hecho de considerar la función indicadora aplicada al signo de los residuos del modelo cuantil como la variable respuesta de un modelo de regresión logística. Es decir, la variable dicotómica

$$Z(\theta_\tau) = \mathbb{I}(Y \leq q_\tau(X, \theta_\tau))$$

jugará un papel fundamental a lo largo de este capítulo.

Así, siguiendo la idea desarrollada por Redden et al. (2004), si el modelo cuantil es correcto, todos los coeficientes asociados al modelo logístico serán cero salvo la constante. Nótese que dicha constantes será τ, donde τ representa el cuantil de interés en torno al cual realizamos el modelo de regresión. En base al razonamiento anterior, proponemos un test de razón de verosimilitudes sobre la significación de los coeficientes del modelo logístico, que servirá para testear la veracidad del modelo cuantil.

Consideraremos como predictores del modelo logístico funciones de proyecciones univariantes de las covariables del modelo cuantil. Para describir formalmente el nuevo contraste de bondad de ajuste, será necesario que dichas funciones representen una base densa de funciones. En particular, tomaremos polinomios de Hermite a modo de base de funciones, que denotamos por

$$P_i = (1, H_1(X_i), H_2(X_i), H_3(X_i), \ldots, H_p(X_i))', \qquad 1 \leq i \leq n$$

Así, el estadístico de contraste vendría dado por

$$T_U = 2\left(L_n(\widehat{\varphi}, Z(\widehat{\theta}_\tau), P) - L_n(\mathrm{logit}(\tau), Z(\widehat{\theta}_\tau), 1)\right),$$

donde

$$L_n(\varphi, Z(\widehat{\theta}_\tau), P) = \sum_{i=1}^{n} \left(Z_i(\widehat{\theta}_\tau)\varphi' P_i - \log(1 + e^{\varphi' P_i})\right)$$

$$L_n(\mathrm{logit}(\tau), Z(\widehat{\theta}_\tau), 1) = \sum_{i=1}^{n} \left[Z_i(\widehat{\theta}_\tau)\mathrm{logit}(\tau) - \log\left(1 + e^{\mathrm{logit}(\tau)}\right)\right].$$

Con el objetivo de estimar el parámetro φ asociado al modelo de regresión logística, hemos considerado métodos de máxima verosimilitud penalizada. Es decir, el estimador $\widehat{\varphi}$ se calcula como sigue

$$\begin{aligned}\widehat{\varphi} &= \arg\max_{\varphi} \left[n^{-1}\, L_n(\varphi, Z(\widehat{\theta}_\tau), P) + \lambda\, \|\varphi\|\right] \\ &= \arg\max_{\varphi} \left[\frac{1}{n} \sum_{i=1}^{n} \left(Z(\widehat{\theta}_\tau)_i\, \varphi' P_i - \log(1 + e^{\varphi' P_i})\right) + \lambda\, \|\varphi\|\right]\end{aligned}$$

donde λ es el conocido como parámetro de suavización. Hemos optado por métodos de máxima verosimilitud penalizada para evitar el conocido **problema de separación**, que se observa en el proceso de estimación de un modelo logístico cuando la verosimilitud converge mientras al menos uno de los parámetros estimados diverge a $\pm\infty$. Este problema ha sido tratado por varios autores como Firth (1993) y Horowitz y Spokoiny (2002).

Por otra parte, supuesto que la variable explicativa asociada al modelo de regresión cuantil sea multivariante, buscaremos la proyección "menos favorable" para la hipótesis nula del test de razón de verosimilitudes. De esta forma conseguimos que el test sea consistente para toda clase de alternativas, incluso en alta dimensión. En este segundo escenario, el estadístico de contraste viene dado por

$$T_M = \max_{\beta\in\mathbb{R}^d, \|\beta\|=1} 2\left(L_n(\widehat{\varphi}, Z(\widehat{\theta}_\tau), P(\beta)) - L_n(\text{logit}(\tau), Z(\widehat{\theta}_\tau), 1)\right).$$

donde

$$P_i(\beta) = (1, H_1(\beta' X_i), H_2(\beta' X_i), \ldots, H_p(\beta' X_i))', \qquad 1 \le i \le n.$$

Un estudio de simulación muestra las buenas propiedades del nuevo test frente a otros contrastes no paramétricos disponibles en la literatura. De igual modo, este nuevo contraste ha sido aplicado a una base de datos reales. Los principales resultados obtenidos a lo largo del Capítulo 5 pueden consultarse en Conde Amboage et al. (2016).

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Bibliography

Abberger, K. (1998). Cross-validation in nonparametric quantile regression. *Allgemeines Statistisches Archiv*, *82*, 149-161.

Abberger, K. (2002). Variable data driven bandwidth choice in nonparametric quantile regression. *Technical Report*.

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

Barro, R. J., and Lee, J. W. (1994). Data set for a panel of 138 countries, discussion paper, nber. *http://www.nber.org/pub/barro.lee*.

Barro, R. J., and Sala-i Martin, X. (1995). *Economic growth.* McGraw-Hill.

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.

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.

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

Cade, B. S., and Noon, B. R. (2003). A gentle introduction to quantile regression for ecologists. *Frontiers in Ecology and the Environment*, *1*, 412-388.

Chambers, J. M., Cleveland, B., W. S.and Kleiner, and Tukey, P. A. (1983). *Graphical methods for data analysis.* Wadsworth & Brooks. Cole Publishing Company. Advances Books & Sofware.

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

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

Chaudhuri, P. (1991b). Nonparametric estimates of regression quantiles and their local bahadur representation. *The Annals of Statistics*, *19*, 760-777.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics*, *33*, 806-839.

Conde-Amboage, M., González-Manteiga, W., and Sánchez-Sellero, C. (2016). Predicting trace gas concentrations using quantile regression models. *Stochastic Environmental Research and Risk Assessment*, *To appear*.

Conde Amboage, M., Patilea, V., and Sánchez Sellero, C. (2016). A new lack-of-fit test for quantile regression models using logistic regression. *Book of abstracts of 48émes Journées de Statistique de la Société Française de Statistique*, http://papersjds16.sfds.asso.fr/submission_180.pdf.

Conde-Amboage, M., and Sánchez-Sellero, C. (2017). A plug-in bandwidth selector for nonparametric quantile regression. *Manuscript*.

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 and Data Analysis*, *88*, 128-138.

Cook, R. D., and Weisberg, S. (1982). *Residuals and influence in regression.* Chapman and Hall.

Cox, D. R. (1958). The regression analysis of binary sequences (with discussion). *Journal of the Royal Statistical Society B*, *20*, 215-242.

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

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *The Annals of Statistics*, *27*, 1012-1040.

Dette, H., Guhlich, M., and Neumeyer, N. (2012). Testing for additivity in nonparametric quantile regression. *http://www.ruhr-uni-bochum.de/mathematik3/-research/index.html*.

Efron, B. (1983). Estimating the error rate of a prediction rule: Improvement on cross-validation. *Journal of the American Statistical Association*, *78*, 316-331.

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

El Ghouch, A., and Genton, M. G. (2012). Local polynomial quantile regression with parametric features. *Journal of the American Statistical Association*, *104*, 1416-1429.

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

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

Escanciano, J. C., and Velasco, C. (2010). Specification tests of parametric dynamic conditional quantiles. *Journal of Econometrics*, *159*, 209-221.

Fahrmeir, L., Kneib, T., Lang, S., and Marx, B. (2013). *Regression: Models, methods and applications*. Springer.

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

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, *98*, 995-999.

Fernández-Castro, B. M., and González-Manteiga, W. (2008). Boosting for real and functional samples: an application to an environmental problem. *Stochastic Environmental Research and Risk Assessment*, *22*, 27-37.

Fernández-Castro, B. M., Guillas, S., and González-Manteiga, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, *47*, 212-222.

Fernández-Castro, B. M., Prada-Sánchez, J. M., González-Manteiga, W., Febrero-Bande, B.-C. J. L., M., and Hernández-Fernádez, J. J. (2003). Prediction of so_2 levels using neural networks. *Journal of the Air and Waste Management Association*, *53*, 532-538.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*, 27-38.

Fontanella, L., Ippoliti, L., Sarra, A., Valentini, P., and Palermi, S. (2015). Hierarchical generalised latent spatial quantile regression models with applications to indoor radon concentration. *Stochastic Environmental Research and Risk Assessment*, *29*, 357-367.

García-Jurado, I., González-Manteiga, W., Prada-Sánchez, J. M., Febrero-Bande, M., and Cao, R. (1995). Predicting using box-jenkins, nonparametric and bootstrap techniques. *Technometrics*, *37*, 303-310.

García-Portugués, E., González-Manteiga, W., and Febrero-Bande, M. (2014). A goodness-of-fit test for the functional linear model with scalar response. *Journal of Computational and Graphical Statistics*, *23*, 761-778.

González-Manteiga, W., and Crujeiras, R. M. (2013). An updated review of goodness-of-fit test for regression models. *Test*, *22*, 361-411.

Gut, A. (2005). *Probability: a graduate course*. Springer.

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.

Hall, P., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimationing a conditional distribution function. *Journal of the Americam Statistical Association*, *94*, 154-163.

Hall, P., and Yao, Q. (2005). Approximating conditional distribution ffunction using dimension reduction. *The Annals of Statistics*, *33*, 1404-1421.

Hampel, F. R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics*, *42*, 1887-1896.

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

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986). *Robust statistics: The approach based on influence functions.* Wiley.

Härdle, W., and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, *21*, 1926-1497.

He, X., and Shao, Q. M. (1996). A general bahadur representation of m-estimators and its application to linear regression with nonstochastic designs. *The Annals of Statistics*, *24*, 2608-2630.

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

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.

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.

Huber, P. J. (1981). *Robust statistics.* Wiley.

Hurvich, C. M., and Tsai, C. L. (1990). Model selection for least absolute deviations regression in small samples. *Statistics & Probability Letters*, *9*, 259-265.

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

Koenker, R. (1994). Confidence intervals for regression quantiles. *Proceedings of the 5th Prague Symposium on Asymptotic Statistics, New York: Springer-Verlag*, 349-359.

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

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

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

Koenker, R., and Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, *94*, 1296-1310.

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.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Koenker, R., and Ng, P. (2003). Sparsem: a sparse matrix package for r. *Journal of Statistical Software*, *8*, 1-9.

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

Lavergne, P., and Patilea, V. (2008). Breaking the curse of dimensionality in nonparametric testing. *Journal of Econometrics*, *143*, 103-122.

Li, Q., and Racine, J. S. (2007). *Nonparametric econometrics: Theory and practise.* Princeton University Press.

Maistre, S., Lavergne, P., and Patilea, V. (2014). Powerful nonparametric checks for quantile regression. *arXiv:1404.0216 [math.ST]*.

Mallows, C. L. (1973). Some comments on c_p. *Technometrics*, *15*, 661-675.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, *21*, 255-285.

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

Mayr, A., Hothorn, T., and Fenske, N. (2012). Prediction intervals for future bmi values of individual children - a non-parametric approach by quantile boosting. *BMC Medical Research Methodology*, *12*, 6.

McCullagh, P., and Nelder, J. A. (1983). *Generalized linear models.* Chapman & Hall.

Meinshausen, N. (2006). Quantile regression forests. *Journal of Machine Learning Research*, *7*, 983-999.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and Its Applications*, *10*, 186-196.

Nanos, N., Grigoratos, T., Rodríguez-Martín, J. A., and Samara, C. (2015). Scaledependent correlations between soil heavy metals and as around four coal-fired power plants of northern greece. *Stochastic Environmental Research and Risk Assessment*, *29*, 1531-1543.

Nelder, J. A., and Mead, R. (1965). A simplex algorithm for function minimization. *Computer Journal*, *7*, 308-313.

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

Parthasarathy, K. R. (1967). *Probability measures on metric spaces.* Academic Press.

Patilea, V., Sánchez-Sellero, C., and Saumard, M. (2016). Testing the predictor effect on a functional response. *Journal of the American Statistical Association*, *11*, 1684-1695.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, *7*, 186-199.

Portnoy, S., and Koenker, R. (1997). The gaussian hare and the laplacian tortoise: computability of squared-error versus absolute-error estimators, with discusssion. *Statistical Science*, *12*, 279-300.

Prada-Sánchez, J. M., and Febrero-Bande, M. (1997). Parametric, non-parametric and mixed approaches to prediction of sparsely distributed pollution incidents: a case study. *Journal of Chemometrics*, *11*, 13-32.

Prada-Sánchez, J. M., Febrero-Bande, M., Cotos-Yáñez, T., González-Manteiga, W., Bermúdez-Cela, J. L., and Lucas-Domínguez, T. (2000). Prediction of so_2 pollution incidents near a power station using partially linear models and a historical matrix of predictor-response vectors. *Environmetrics*, *11*, 209-225.

Redden, D. T., Fernandez, J. R., and Allison, D. B. (2004). A simple significance test for quantile regression. *Statistics in Medicine*, *23*, 2587-2597.

Roca-Pardiñas, J., Cadarso-Suárez, C., and González-Manteiga, W. (2005). Testing for interactions in generalized additive models. application to so_2 pollution data. *Statistics and Computing*, *15*, 289-299.

Roca-Pardiñas, J., González-Manteiga, W., Febrero-Bande, M., Prada-Sánchez, J. M., and Cadarso-Suárez, C. (2004). Predicting binary time series of so_2 using generalized additive models with unknown link function. *Environmetrics*, *15*, 729-742.

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

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

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995). An efective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, *90*, 1257-1270.

Salama, A. (2005). A note on the impact of environmental performance on financial performance. *Structural Change and Economic Dynamics*, *16*, 413-421.

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

Scott, D. W. (1992). *Multivariate density estimation. theory, practise and visualization.* John Wiley & Sons.

Seber, G. (1977). *Linear regression analysis.* Wiley.

Sheather, S. J., and Maritz, J. S. (1983). An estimate of the asymptotic standard error of the sample median. *Australian Journal of Statistics*, *25*, 109-122.

Shi, J., and Harrison, R. (1997). Regression modelling of hourly no_x and no_2 concentrations in urban air in london. *Atmospheric Environment*, *31*, 4081-4094.

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.

Sousa, S. I. V., Pires, J. C. M., Martins, F. G., Pereira, M. C., and Alvim-Ferraz, M. C. M. (2009). Potentialities of quantile regression to predict ozone concentrations. *Environmetrics*, *20*, 147-158.

Stine, R. A. (1985). Bootstrap prediction intervals for regression. *Journal of the American Statistical Association*, *80*, 1026-1031.

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

Stute, W., González-Manteiga, W., and Presedo-Quindimil, M. (1998). Bootstrap approximations in model checks for regression. *Journal of the American Statistical Association*, *93*, 141-149.

Stute, W., Xu, W. L., and Zhu, L. X. (2008). Model diagnosis for parametric regression in high-dimensional spaces. *Biometrika*, *95*, 451-467.

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

Van der Vaart, A. W. (2000). *Asymptotic statistics.* Cambridge University Press.

Van Der Vaart, A. W., and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer New York.

Venables, W. N., and Ripley, B. D. (1999). *Modern applied statistics with s-plus. third edition.* Springer New York.

Walker, S. H., and Duncan, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, *54*, 167-178.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, *26*, 359-372.

Wellner, J. A. (2005). *Empirical processes: Theory and applications.* Notes for a course given at Delft University of Technology.

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

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

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Wilcox, R. R. (2011). *Introduction to robust estimation and hypothesis testing.* Academic Press.

Xia, Y. (2009). Model checking in regression via dimension reduction. *Biometrika*, *96*, 133-148.

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

Yu, K., and Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics*, *31*, 333-346.

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

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

Zhou, K. Q., and Portnoy, S. L. (1996). Direct use of regression quantiles to construct confidence sets in linear models. *The Annals of Statistics*, *24*, 287-306.

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Notation

a'	transpose of a vector a
α	significance level associated with a lack-of-fit test
AMISE	asymptotic mean integrated squared error
$\text{Bi}(n, p)$	binomial distribution with parameter n, p
c_τ	τ-th quantile of a random variable
card	cardinal of a subset
$*$	convolution
$\xrightarrow{d}$	distribution convergence
d	dimension of the covariate associated with a quantile regression
$\mathbb{E}$	expectation
ε	unknowm error associated with a regression model
F_X, f_X, F_X^{-1}	distribution, density and quantile function associated with a random variable X
F_n	empirical distribution function
h_τ	bandwidth parameter associated with a local linear quantile regression model
$\inf\{A\}$	infimun of a subset A
$\mathbb{I}$	indicator function
$[\cdot]$	integer part of a number
$\|\cdot\|$	l_2 norm
$\|\cdot\|_1$	l_1 norm
λ	smoothing parameter associated with penalized regression
L_n	likelihood function
m	mean regression function
$\mathcal{M}_\theta$	family of a parametric mean regression function
MAE	mean absolute error
MISE	mean integrated squared error
MSE	mean squared error
$N(\mu, \sigma^2)$	a normal distribution with mean μ and variance σ^2
n	sample size of a random sample

$\mathbb{R}^d$	d-dimensional Euclidian space
Φ, ϕ	distribution and density functions associated with a standard Gaussian distribution
ψ_τ	derivative of the quantile loss function
p	number of Hermite polynomials in the basis
q	dimension of the estimator associated with a quantile regression
q_τ	quantile regression function
$q_{\tau,h}$	local linear quantile regression with bandwidth parameter h
$q^{(i)}(X,\theta)$	$i-$th partial derivative of the quantile regression function with respect to the parameter θ, that is, $\partial^i q_\tau / \partial\theta^i$
$\mathcal{Q}_\theta$	family of a parametric quantile regression function
ρ_τ	loss quantile function
$\mathbb{S}^d$	unit sphere in $\mathbb{R}^d$
θ	estimator associted with a regression model
$\widehat{\theta}_{\text{LS}}$	least squares estimator
$\widehat{\theta}_\tau$	estimator associated with a τ-th quantile regression
T_a	test statistic associated with method ”a”
$\mathbb{V}$ar	Variance
$\mathbb{X}$	design matrix
$X = (X^{(1)}, \ldots, X^{(d)})$	covariate associated with a quantile regression model
Y	response variable associtės with a quantile regressin model
$Y^\star$	bootstrap replication associated with a variable Y

USC UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA