Significance testing in quantile regression

Stanislav Volgushev[a,d*], Melanie Birke[b], Holger Dette[a*], Natalie Neumeyer[c]
[a] Ruhr-Universität Bochum
[b] Universität Bayreuth
[c] Universität Hamburg
[d] University of Illinois at Urbana-Champaign.

Abstract

We consider the problem of testing significance of predictors in multivariate nonparametric quantile regression. A stochastic process is proposed, which is based on a comparison of the responses with a nonparametric quantile regression estimate under the null hypothesis. It is demonstrated that under the null hypothesis this process converges weakly to a centered Gaussian process and the asymptotic properties of the test under fixed and local alternatives are also discussed. In particular we show, that - in contrast to the nonparametric approach based on estimation of L^2-distances - the new test is able to detect local alternatives which converge to the null hypothesis with any rate $a_n \to 0$ such that $a_n\sqrt{n} \to \infty$ (here n denotes the sample size). We also present a small simulation study illustrating the finite sample properties of a bootstrap version of the the corresponding Kolmogorov-Smirnov test.

AMS Classification: 62G10, 62G08, 62G30
Keywords and Phrases: nonparametric quantile regression, significance testing, empirical processes, monotone rearrangement

1 Introduction

Nonparametric regression methods have become very popular in the last decades because of the fact that employing a mis-specified parametric model will typically result in inconsistent estimates and as a consequence invalid statistical inference. In recent years many authors have developed nonparametric quantile regression estimates, which provide an attractive supplement to least squares methods by focussing on the estimation of the conditional quantiles instead of

*Supported by the Sonderforschungsbereich “Statistical modelling of nonlinear dynamic processes” (SFB 823) of the Deutsche Forschungsgemeinschaft.

the mean function [see Chaudhuri (1991), Yu and Jones (1997), Yu and Jones (1998), Dette and Volgushev (2008), Chernozhukov et al. (2010) or Bondell et al. (2010) among many others]. These references mainly discuss the case of a one dimensional predictor, but from a theoretical point of view the methods can easily be generalized to multivariate predictors. On the other hand it is well known that in practical applications such nonparametric methods suffer from the curse of dimensionality and therefore do not yield precise estimates of conditional quantile surfaces for realistic sample sizes. In such cases a natural and very important question is which predictor variables are significant.

The problem of testing significance has found considerable interest in multivariate mean regression models. Gozalo (1993) considered conditional moment tests, while Yatchew (1992) constructed a test based on semi-parametric least-squares residuals. Lavergne and Vuong (1996) suggested a directional testing procedure for discriminating between two sets of regressors without specifying the functional form of the mean regression, and Racine (1997) proposed a test based on nonparametric estimates of the partial derivatives of the conditional mean of the response. Lavergne and Vuong (2000) used the kernel method to develop a test for the significance of a subset of explanatory variables and Delgado and González-Manteiga (2001) proposed a test which is based on functionals of a U-process.

Because of the well known robustness properties of the conditional quantile and the fact that conditional quantiles characterize the entire distribution it is of particular interest to develop methods for testing significance of predictors in quantile regression models. Surprisingly, in quantile regression this problem has found much less attention. Variable selection in the framework of linear quantile regression models has been recently considered by Zou and Yuan (2008), Wu and Liu (2009) and Belloni and Chernozhukov (2011) among others. Jeong et al. (2012) proposed a test for significance in a multivariate quantile regression model. The work of these authors was motivated by Granger quantile causality [Granger (1969)] and they employed an idea of Zheng (1998), who proposed to transform quantile restrictions to mean restrictions. The corresponding test is based on a U-statistic, which estimates the distance measure

$$\Delta = E[(P(Y \leq q_\tau(X)|X,Z) - \tau)^2 f_Z(Z)], \tag{1.1}$$

where Y denotes the response, (X,Z) is the predictor, f_Z the density of Z and $q_\tau(X)$ the conditional τ-quantile of Y given X. Note that the quantity Δ vanishes if and only if the conditional quantile of Y given X and Z does not depend on Z. A major drawback of this approach lies in the fact that non-parametric smoothing over both X and Z is needed for the construction of the estimate. This implies that the test is of very limited use when the dimension of (X,Z) is larger than 3. Moreover, this test can only detect local alternatives converging to the null hypothesis $H_0 : \Delta = 0$ at a rate $n^{-1/2}h^{-(d+q)/4}$, where d and q are the dimensions of the predictors X and Z, respectively, and h denotes a bandwidth converging to 0 with increasing sample size n.

The present paper is devoted to the problem of constructing a test for the hypothesis of the significance of the predictor Z, i.e. $\Delta = 0$, in the nonparametric quantile regression model, which can detect local alternatives converging to the null hypothesis at a parametric rate and at the same time does not depend on the dimension of the predictor Z, such that smoothing with respect to the covariate Z can be avoided. To be precise, the test proposed in this paper can detect alternatives converging to H_0 at any rate $a_n \to 0$ such that $a_n\sqrt{n} \to \infty$, where n denotes the sample size. Our approach is based on an empirical process, which estimates the functional

$$
\begin{aligned}
T(x,z) &= E[(P(Y \leq q_\tau(X)|X,Z)) - \tau)I\{X \leq x\}I\{Z \leq z\}] \\
&= E[(I\{Y \leq q_\tau(X)\} - \tau)I\{X \leq x\}I\{Z \leq z\}]
\end{aligned}
\tag{1.2}
$$

for all (x,z) in the support of the distribution of the predictor (X,Z), where the inequality $X \leq x$ between the vectors X and x is understood as the vector of inequalities between the corresponding coordinates and $I\{A\}$ denotes the characteristic function of the event A. The model, necessary notation and definition of this process are introduced in Section 2 and a stochastic expansion of the process $T_n(x,z)$ is established in Section 3. This result allows us to obtain the weak convergence of an appropriately scaled and centered version of $T_n(x,z)$ under the null hypothesis, fixed and local alternatives. As a result we obtain a Kolmogorov-Smirnov or a Cramer von Mises type statistic for the hypothesis of the significance of the predictor Z in the nonparametric quantile regression model. Moreover, we are also able to extend the result to the case, where the dimension q of the predictor Z is growing with the sample size, that is $q = q_n \to \infty$ as $n \to \infty$. The finite sample properties of a corresponding bootstrap test are investigated in Section 4. As a by-product of our theoretical analysis we also obtain new results on the uniform convergence of the conditional quantile estimator proposed by Dette and Volgushev (2008). Finally all proofs, which are complicated, are deferred to an Appendix in Section A.

2 Model, assumptions and test statistic

Let Y, X and Z denote one-, d and q dimensional random variables, respectively, where Y corresponds to the response and X and Z are the covariates. We assume that the random variables $\{(Y_i, X_i, Z_i)\}_{i=1,\ldots,n}$ are independent identically distributed with the same distribution as (Y,X,Z). Let $\tau \in (0,1)$ be fixed. Our aim is to test whether the predictor Z has influence on the conditional τ-quantile of Y, given (X,Z), or whether the variable Z can be omitted. Note that this problem fundamentally differs from the question whether Y is independent of Z given X. In fact, the latter is equivalent to testing whether *all* quantile curves do not depend on Z as opposed to looking at a particular quantile. Thus for fixed $\tau \in (0,1)$ we formulate the null hypothesis as

$$
H_0 : E[I\{Y \leq q_\tau(X)\} - \tau \mid X, Z] = P(Y \leq q_\tau(X) \mid X, Z) - \tau = 0 \;\; a.s., \tag{2.1}
$$

where $q_\tau(X)$ is defined as the conditional τ-quantile of Y, given X, that is

$$
P(Y \leq q_\tau(X) \mid X) = \tau. \tag{2.2}
$$

It is easy to see that the null hypothesis (2.1) is equivalent to

$$T(x,z) \equiv 0$$

for all (x,z) in the support of the random variable (X,Z), where the functional T is defined in (1.2). This functional can be be estimated by the stochastic process

$$T_n(x,z) = \frac{1}{n}\sum_{i=1}^{n}(I\{Y_i \le \hat{q}_\tau(X_i)\} - \tau)I\{X_i \le x\}I\{Z_i \le z\},, \tag{2.3}$$

where $(x,z) \in R_X \times R_Z$, R_X and R_Z denote the support of the distributions of the random variables X and Z, respectively, and $\hat{q}_\tau$ is an appropriate estimate of the conditional quantile of Y given X, which will be specified below. A test for the hypothesis of significance of the variable Z for the τ's quantile curve of Y can now easily be obtained by considering a Kolmogorov-Smirnov or Cramer von Mises type statistic based on T_n and rejecting the null hypothesis for large values of this statistic. Throughout this paper we assume that the sets R_X and R_Z are compact.
In the literature, several non-parametric quantile regression estimators have been proposed [see e.g. Yu and Jones (1997, 1998), Takeuchi et al. (2006), Chernozhukov et al. (2010) or Bondell et al. (2010) among others]. In this paper we will use an approach proposed by Dette and Volgushev (2008) who constructed non-crossing estimates of quantile curves using a simultaneous inversion and isotonization of a preliminary estimator of the conditional distribution function $F_{Y|X}$ of Y given X. For this estimator, say $\hat{F}_{Y|X}(y|x;p)$, we will use a smoothed local polynomial estimator of order p, see e.g. Fan and Gijbels (1996). Before defining this estimator, it is necessary to introduce some notation.

- For d-dimensional vectors $x = (x(1),\dots,x(d)) \in \mathbb{R}^d$ and $\mathbf{k} = (\mathbf{k}(1),\dots \mathbf{k}(d)) \in \mathbb{N}_0^d$ define

$$\begin{aligned} x^{\mathbf{k}} &:= (x(1)^{\mathbf{k}(1)},...,x(d)^{\mathbf{k}(d)}) \;, \quad \pi(x) := x(1)\cdot x(2)\cdot ... \cdot x(d) \\ \sigma(\mathbf{k}) &:= \mathbf{k}(1)+...+\mathbf{k}(d) \;, \qquad \mathbf{k}! := \mathbf{k}(1)! \cdot ... \cdot \mathbf{k}(d)! \end{aligned}$$

- For d-dimensional vectors $x \in \mathbb{R}^d$, $\mathbf{k} \in \mathbb{N}_0^d$ and a function $K : \mathbb{R} \to \mathbb{R}$ define

$$\begin{aligned} \mathbf{K}(x) &:= K(x(1))\cdots K(x(d)) \;, \qquad \mathbf{K}_{h_n,\mathbf{k}}(x) := \mathbf{K}(x/h_n)\pi((x/h_n)^{\mathbf{k}}) \\ \mathbf{K}^{(\mathbf{m})}(x) &:= K^{(\mathbf{m}(1))}(x(1))\cdots K^{(\mathbf{m}(d))}(x(d)) \;, \quad \mathbf{K}^{(\mathbf{m})}_{h_n,\mathbf{k}}(x) := \mathbf{K}^{(\mathbf{m})}_{1,\mathbf{k}}(x/h_n) \end{aligned}$$

  where $\mathbf{m} = (\mathbf{m}(1)\dots,\mathbf{m}(d))$ is a d-dimensional vector with entries from $\mathbb{N}_0$ and $\mathbf{K}^{(\ell)}$ is the ℓth derivative of a function $\mathbf{K}$.

- Define $N_j := \#\{\mathbf{k} \in \mathbb{N}_0^d | \sigma(\mathbf{k}) = j\}$ as the number of distinct d-tuples with size j, and denote the elements of this set by $\mathbf{k}_{1,m},...,\mathbf{k}_{N_m,m}$

With these notational conventions the local polynomial estimator $\hat{F}_{Y|X}(y|x;p)$ of order p can be represented as [see e.g. Fan and Gijbels (1996)]

$$\hat{F}_{Y|X}(y|x;p) := e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^t\mathbf{W}\mathbf{Y}, \tag{2.4}$$

where e_1 denotes a vector of suitable dimension with first entry one and remaining entries zero, the matrices $\mathbf{X}$, $\mathbf{W}$ and the vector $\mathbf{Y}$ are given by

$$\mathbf{X} = \begin{pmatrix} 1 & (x-X_1)^{\mathbf{k}_{1,1}} & \dots & (x-X_1)^{\mathbf{k}_{N_1,1}} & (x-X_1)^{\mathbf{k}_{1,2}} & \dots & (x-X_1)^{\mathbf{k}_{p,N_p}} \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ 1 & (x-X_n)^{\mathbf{k}_{1,1}} & \dots & (x-X_n)^{\mathbf{k}_{N_1,1}} & (x-X_n)^{\mathbf{k}_{1,2}} & \dots & (x-X_n)^{\mathbf{k}_{p,N_p}} \end{pmatrix},$$

$$\mathbf{W} = \frac{1}{nh_n^d}\mathrm{Diag}\Big(\mathbf{K}_{h_n,0}(x-X_1),\dots,\mathbf{K}_{h_n,0}(x-X_n)\Big),$$

$$\mathbf{Y} := \Big(\Omega\Big(\frac{y-Y_1}{d_n}\Big),\dots,\Omega\Big(\frac{y-Y_n}{d_n}\Big)\Big)^t, \tag{2.5}$$

and Ω denotes a smoothed version of the indicator function $I\{\cdot \le 0\}$, that is

$$\Omega(v) = \int_{-\infty}^{v} \omega(u)du \tag{2.6}$$

for a given kernel ω with support $[-1,1]$. Following Dette and Volgushev (2008) we consider a strictly increasing distribution function $G : \mathbb{R} \to (0,1)$, a nonnegative kernel κ with bandwidth b_n, and define the functional

$$H_{G,\kappa,\tau,b_n}(F) := \frac{1}{b_n}\int_0^1 \int_{-\infty}^{\tau} \kappa\Big(\frac{F(G^{-1}(u))-v}{b_n}\Big)dvdu. \tag{2.7}$$

If $\hat{F}_{Y|X}$ is the estimator of the conditional distribution function defined in (2.4), it is intuitively clear that $H_{G,\kappa,\tau,b_n}(\hat{F}_{Y|X}(\cdot|x))$ is a consistent estimate of $H_{G,\kappa,\tau,b_n}(F_{Y|X}(\cdot|x))$. If $b_n \to 0$, this quantity can be approximated as follows

$$\begin{aligned} H_{G,\kappa,\tau,b_n}(F_{Y|X}(\cdot|x)) &\approx \int_{\mathbb{R}} I\{F_{Y|X}(y|x) \le \tau\}dG(y) \\ &= \int_0^1 I\{F_{Y|X}(G^{-1}(v)|x) \le \tau\}dv \;=\; G \circ F^{-1}_{Y|X}(\tau|x), \end{aligned}$$

and as a consequence an estimate of the conditional quantile function $q_\tau(x) = F^{-1}_{Y|X}(\tau|x)$ can be defined by

$$\hat{q}_\tau(x) := G^{-1}(H_{G,\kappa,\tau,b_n}(F_{Y|X}(\cdot|x))). \tag{2.8}$$

Throughout this paper, we will assume that the kernels, the function G and the bandwidth parameters used to build the estimator satisfy the following conditions

(K1) The kernel K has support $[-1,1]$ and is $p+1 \ge d+2$ times continuously differentiable with uniformly bounded derivatives. Additionally the first $p+1$ derivatives of K vanish at the boundary points -1 and 1.

(K2) The function ω in (2.6) is a kernel of order $s \ge d+1$, has support $[-1,1]$ and is d times continuously differentiable. Additionally ω has uniformly bounded derivatives that vanish at the boundary points -1 and 1.

(K3) The kernel κ is a symmetric, positive with support $[-1,1]$ and has one Lipschitz-continuous derivative.

(K4) $G:\mathbb{R}\to[0,1]$ is a strictly increasing distribution function such that G, G^{-1} are two time continuously differentiable

(K5) $d_n^{2s}+h_n^{p+1}=o(1/\sqrt{n})$ and $\log n/(nh_n^{3d+2})+\log n/(nh_n^d d_n^{2d-1})=o(1)$

(K6) $\frac{\log n}{nh_n^d b_n^2}=o(1)$, $b_n^2+\frac{\log n}{nh_n^d b_n}+\frac{b_n\sqrt{\log n}}{\sqrt{nh_n^d}}=o(1/\sqrt{n})$

Remark 2.1 Dette and Volgushev (2008) demonstrate that the choice of the distribution function G has a negligible impact on the quality of the resulting estimate provided that an obvious centering and standardization is performed. Similarly, the estimate $\hat{q}_\tau(x)$ is robust with respect to the choice of the bandwidth b_n if it is chosen sufficiently small [see Dette et al. (2006)].

Remark 2.2 Dette and Volgushev (2008) only established point-wise weak convergence of their estimator. However, for most applications such as the construction of tests on the basis of this estimator, uniform results are needed. In the present paper, we provide general inequalities for the operator H_{G,κ,τ,b_n} defined in (2.7), see Lemma B.4 in the Appendix. In particular, these findings allow to describe uniform properties of the quantile estimator $\hat{q}_\tau$ in terms of the properties of the underlying distribution function estimator $\hat{F}_{Y|X}$. For example, in Theorem A.1 in the appendix we exploit those bounds to derive a uniform Bahadur-type representation for the estimate $\hat{q}_\tau$ defined in (2.8).

In the following discussion it turns out to be advantageous to consider a generalization of the test statistic T_n defined in (2.3), where the indicator functions $I\{X_i\le x\}$ are replaced by indicators of more general sets Θ. To be precise let Ξ denote a collection of subsets of $\mathbb{R}^d$ and define $\mathcal{D}_n:=\{x\in R_X|[x-h_n\mathbf{1},x+h_n\mathbf{1}]\subset R_X\}$ (here $\mathbf{1}$ denotes the d-dimensional vector with all entries equal to 1), then all theoretical developments will be based on the statistic

$$(2.9)\qquad T_n(\Theta,z)=\frac{1}{n}\sum_{i=1}^n(I\{Y_i\le\hat{q}_\tau(X_i)\}-\tau)I\{X_i\in\Theta\cap\mathcal{D}_n\}I\{Z_i\le z\},\quad \Theta\in\Xi, z\in R_Z.$$

The intersection of the sets $\Theta\in\Xi$ with the set $\mathcal{D}_n$ is needed in the theoretical developments to exclude "residuals" $I\{Y_i\le\hat{q}_\tau(X_i)\}-\tau$ corresponding to predictors close to the boundary of R_X. Note that if $\cup_{\Theta\in\Xi}\Theta$ has a positive distance to the boundary of R_X, the collection of sets Ξ_n will equal Ξ whenever h_n is sufficiently small. Note also that we use the same symbol T_n for the processes in (2.3) and (2.9) but the meaning is always clear from the context.
Additionally to its advantages from a theoretical point of view, the consideration of a collection of sets that are more general than sets defined by indicators of rectangles will for example allow to investigate the problem of testing the significance of the variable Z on a certain subset, say $\mathcal{D}\subset R_X$, that is

$$(2.10)\qquad H_0^{\mathcal{D}}: E[I\{Y\le q_\tau(X)\}I\{X\in\mathcal{D}\}\mid X,Z]=P(Y\le q_\tau(X)\text{ and }X\in\mathcal{D}\mid X,Z)=\tau$$

Note that $H_0^{\mathcal{D}}$ means that the conditional $\tau-$quantile of Y given (X, Z) can be represented as a function $q_\tau(X)$ for $X \in \mathcal{D} \subset R_X$. In this case a natural choice for the collection Ξ is given by $\Xi := \{\{X \leq t\} \cap \mathcal{D} | t \in \mathbb{R}^d\}$, but other choices are of course possible as well.

3 Main asymptotic results

In this section we investigate the asymptotic properties of the stochastic process defined in (2.9). For this purpose we need some additional notation and technical assumptions which are collected here for convenience and for later reference.

Define the 'error' variables as $\varepsilon = Y - q_\tau(X)$ and $\varepsilon_i = Y_i - q_\tau(X_i)$, $i = 1, \ldots, n$. We assume that the conditional distribution function $F_{\varepsilon|X}(\cdot|x)$ of ε given $X = x$ has a density, say $f_{\varepsilon|X}(y|x)$. Note that by definition we have that $F_{\varepsilon|X}(0|X) = P(\varepsilon \leq 0|X) = \tau$. In particular, this identity continues to hold even if the null hypothesis is violated. Throughout this paper we denote by $F_{Z|X,\varepsilon}(z|x, e)$ the conditional distribution function of Z given $(X, \varepsilon) = (x, e)$.

Define $\mathcal{D} := \cup_{\Theta \in \Xi} \Theta$, then we assume that the data-generating process satisfies the following conditions.

(A1) The conditional distribution function $F_{Y|X}(y|x)$ is $p + 1$ times continuously differentiable with respect to x, y and all partial derivatives are uniformly bounded on $\mathbb{R} \times R_X$. The joint density of (X, Y) is uniformly bounded on $R_X \times \mathbb{R}$. Moreover, $p \geq \max(s, d + 1)$.

(A2) The density f_X of the predictor X is $d + 1 + n_f$ times continuously differentiable with uniformly bounded partial derivatives on R_X and $n_f > d/2$. Moreover $\inf_{x \in R_X} f_X(x) > 0$.

(A3) There exist constants $a, C_1 > 0$ such that

$$\inf_{(x,y): x \in R_X, |y - q_\tau(x)| \leq a} f_{Y|X}(y|x) \geq C_1.$$

(A4) The function $(z, x) \mapsto F_{Z|X,\varepsilon}(z|x, 0)$ is Hölder-continuous of order $\gamma > 0$ with respect to z and x uniformly in $x \in \mathcal{D}$, i.e.

$$|F_{Z|X,\varepsilon}(s|x, 0) - F_{Z|X,\varepsilon}(t|\xi, 0)| \leq C\|(s, x) - (t, \xi)\|_\infty^\gamma$$

for some finite constant C.

(A5) $\sup_{x \in \mathcal{D}, y \in \mathbb{R}, z \in \mathcal{Z}} |f'_{\varepsilon|X,Z}(y \mid x, z)| < \infty$.

In conditions (A1)-(A4), R_X can be replaced by a set $\mathcal{X} \subset R_X$ provided that $\mathcal{D} \subset \mathcal{X}$. Finally, the following assumptions on the collection of sets Ξ are required.

(S1) The class of functions $\mathcal{F}_1 = \{u \mapsto I\{u \in \Theta\} | \Theta \in \Xi\}$ satisfies $N_{[]}(\mathcal{F}_1, \varepsilon, L^2(P_X)) \leq C\varepsilon^{-a}$ for any sufficiently small $\varepsilon > 0$ and a constant C, where $N_{[]}$ denotes the bracketing number [see van der Vaart and Wellner (1996)]

(S2) $\sup_{\Theta\in\Xi} P(X_i \in \Theta, \exists j : [X_i(j) - h_n, X_i(j) + h_n] \not\subset \Theta) = o(1)$ for $h_n \to 0$.

Remark 3.1 Conditions (S1) and (S2) are not strong and for example satisfied for the collection of rectangles $\Xi = \{\{s \le X \le t\} | s, t \in \mathbb{R}^d\}$ if X has a uniformly bounded density with compact support. For more details on bracketing numbers and their properties we refer to the monograph of van der Vaart and Wellner (1996).

The following result gives a stochastic expansion of the process $T_n(\Theta, z)$ under general conditions, which is crucial for deriving the asymptotic properties of the process T_n. In particular, observe that this representation continues to hold under the alternative.

Theorem 3.2 *If the assumptions (K1)-(K6), (A1) - (A5) and (S1), (S2) are satisfied, the process T_n can be represented as*

$$T_n(\Theta, z) = \frac{1}{n}\sum_{i=1}^{n}(I\{\varepsilon_i \le 0\} - \tau)I\{X_i \in \Theta_n\}(I\{Z_i \le z\} - F_{Z|X,\varepsilon}(z|X_i, 0)) + o_P(n^{-1/2}) \tag{3.1}$$

uniformly with respect to $z \in R_Z, \Theta \in \Xi$.

The proof of Theorem 3.2 is complicated and given is given in the Appendix. As an immediate consequence, we obtain that under the null hypothesis H_0 the rescaled process $\sqrt{n}T_n(\Theta, z)$ converges weakly to a centered Gaussian process.

Corollary 3.3 *If the assumptions of Theorem 3.2 and the null hypothesis H_0 in (2.1) are satisfied, the process $\sqrt{n}T_n$ converges weakly in $\ell^\infty(\Xi \times R_Z)$ to a centered Gaussian process $\mathbb{T}$ with covariance kernel*

$$\begin{aligned} k(\Theta_1, y, \Theta_2, z) = \operatorname{Cov}(\mathbb{T}(\Theta_1, y), \mathbb{T}(\Theta_2, z)) = \tau(1-\tau)\mathbb{E}\Big[I\{X \in \Theta_1 \cap \Theta_2\} \\ \times \mathbb{E}\Big[\Big(I\{Z \le y\} - F_{Z|X,\varepsilon}(y|X, 0)\Big)\Big(I\{Z \le z\} - F_{Z|X,\varepsilon}(z|X, 0)\Big)\Big|X, \varepsilon\Big]\Big]. \end{aligned} \tag{3.2}$$

As a consequence of this result we obtain the weak convergence of functionals such as the Kolmogorov-Smirnov statistic

$$K_n = \sup_{\Theta\in\Xi}\sup_{z\in R_z} |T_n(\Theta, z)|$$

by an application of the continuous mapping theorem. In general the asymptotic distribution of K_n depends on certain features of the data generating process and in the following section we will discuss bootstrap approximations for this distribution. However, in some special cases the situation simplifies substantially.

Remark 3.4 In the case where the pair (X, ε) and the covariate Z are independent it follows from (3.2) that

$$\operatorname{Cov}(\mathbb{T}(\Theta_1, y), \mathbb{T}(\Theta_2, z)) = \tau(1-\tau)P(I\{X \in \Theta_1 \cap \Theta_2\})(F_Z(y \wedge z) - F_Z(y)F_Z(z)),$$

where F_Z is the distribution function of the random variable Z and $y \wedge z$ denotes the vector of minima of the corresponding coordinates of y and z. If additionally X, Z are real-valued and $\Xi = \{(-\infty, t] | t \in \mathbb{R}\}$, the asymptotic covariance in Theorem 3.2 reduces to

$$\mathrm{Cov}(\mathbb{T}((-\infty,t],y),\mathbb{T}((-\infty,s],z)) \;=\; \tau(1-\tau)F_X(s\wedge t)(F_Z(y\wedge z)-F_Z(y)F_Z(z)).$$

Hence, for univariate independent covariates X and Z with continuous distribution functions F_X and F_Z, respectively, the Kolmogorov-Smirnov test is asymptotically distribution-free because in this case the statistic

$$\sqrt{n}\sup_{x\in R_X, z\in R_Z}|T_n(x,z)| = \sqrt{n}\sup_{s,t\in[0,1]}|T_n(F_X^{-1}(s),F_Z^{-1}(t))|$$

converges in distribution to $\sqrt{\tau(1-\tau)}\sup_{s,t\in[0,1]}|B(s,t)|$, where B is the Kiefer-Müller process on $[0,1]^2$, i.e. a centered Gaussian process with covariance kernel

$$\mathrm{Cov}(B(s_1,t_1),B(s_2,t_2)) \;=\; (s_1\wedge s_2)(t_1\wedge t_2 - t_1t_2).$$

The result obtained in Theorem 3.2 can also be used to derive the asymptotic properties of the test statistic under fixed alternatives. More precisely, the following result holds (note that under the null hypothesis, the centering term is zero, and thus this result is a generalization of Corollary 3.3).

Corollary 3.5 *Under the assumptions of Theorem 3.2 the process*

$$\sqrt{n}\Big(T_n(\Theta,z) - \int_{R_X\cap\Theta_n}\int_{R_Z}\Big(F_{Y|X,Z}(q_\tau(u)|u,v)-\tau\Big)I\{v\le z\}dF_{X,Z}(u,v)\Big)$$

converges weakly to the limiting process $\mathbb{T}$ *defined in Corollary 3.3.*

Remark 3.6 A further consequence of Corollary 3.5 is that the statistic T_n converges for all $\Theta \in \Xi$ and $z \in R_Z$ in probability to the function

$$\int_{R_X\cap\Theta}\int_{R_Z}\Big(F_{Y|X,Z}(q_\tau(u)|u,v)-\tau\Big)\Big(I\{v\le z\}-F_{Z|X,\varepsilon}(z|u,0)\Big)f_X(u)f_Z(v)dudv.$$

Consequently, if Ξ contains sufficiently many sets (for example, if $\Xi = \{(-\infty, x] \mid x \in \mathbb{R}^d\}$), the test is consistent. In order to obtain the asymptotic distribution of the test statistic under local alternatives of the form

$$F^{(n)}_{Y|X,Z}(q^{(n)}_\tau(u)|u,v) = \tau + a_n h(u,v) \tag{3.3}$$

a result on the asymptotic behavior of $T_n(\Theta, z)$ is required when the data are generated from triangular arrays. A closer look at the proofs in the appendix shows that such a result does indeed hold under suitable modifications of the conditions in Theorem 3.2. The details are omitted for the sake of brevity. In particular, a test based on the Kolmogorov-Smirnov test statistic will detect all local alternatives for which the quantity

$$K_n = \sup_{\Theta_n, z} \Big| \sqrt{n} \int_{R_X \cap \Theta_n} \int_{R_Z} \Big(F^{(n)}_{Y|X,Z}(q^{(n)}_\tau(u)|u,v) - \tau \Big) I\{v \leq z\} dF^{(n)}_{X,Z}(u,v) \Big|$$

diverges to infinity (the superscript is used to indicate that the corresponding quantities depend on n). For example $K_n \to \infty$ in probability if $\Xi = \{(-\infty, x] \mid x \in \mathbb{R}^d\}$ and $F^{(n)}_{Y|X,Z}(q^{(n)}_\tau(u)|u,v) = \tau + a_n h(u,v)$ for some function h that is not identically zero on $R_X \times R_Z$ and sequence a_n with $a_n \sqrt{n} \to \infty$. This means that the test can detect alternatives converging to the null hypothesis at rates which are "larger but arbitrarily close" to the parametric rate $n^{-1/2}$. Moreover, the test will have an asymptotically non-trivial power against many local alternatives that converge to zero at the exact parametric rate $n^{-1/2}$.

Remark 3.7 We now give a brief discussion of the properties of the proposed test statistic when alternatives of increasing dimension are considered, i.e. when the dimension of the predictor Z, say q_n, varies with n. Consider the additional assumption

(Z) The L^2 covering numbers of the classes of functions

$$\{x \mapsto F_{Z|X,\varepsilon}(z|x+s, 0) | z \in \mathcal{Z}, \|s\|_\infty \leq a\}$$

and $\{\xi \mapsto I\{\xi \leq z\} | z \in \mathcal{Z}\}$ are bounded by $C_1(C_2/\varepsilon)^{k_n}$ for some finite constants C_1, C_2.

Note that assumption (Z) holds with $k_n = q_n$ if for each n the predictor Z given (X, ε) has a conditional density $f_{Z|X,\varepsilon}$ that satisfies

$$\sup_z |f_{Z|X,\varepsilon}(z|x_1, 0) - f_{Z|X,\varepsilon}(z|x_2, 0)| \leq C\|x_1 - x_2\|$$

for a finite constant C independent of n. Under assumptions (K1)-(K6), (A1)-(A3), (Z), (A5), (S1), (S2) it is possible to prove that

$$T_n(\Theta, z) = \frac{1}{n} \sum_{i=1}^n (I\{\varepsilon_i \leq 0\} - \tau) I\{X_i \in \Theta_n\} (I\{Z_i \leq z\} - F_{Z|X,\varepsilon}(z|X_i, 0)) + o_P\Big(\frac{k_n}{n^{1/2}}\Big),$$

uniformly with respect to $z \in R_Z, \Theta \in \Xi$. In particular, this result implies

$$\sqrt{n}\Big(T_n(\Theta, z) - \int_{R_X \cap \Theta_n} \int_{R_Z} \Big(F_{Y|X,Z}(q_\tau(u)|u,v) - \tau\Big) I\{v \leq z\} dF_{X,Z}(u,v)\Big) = O_P\Big(k_n\Big).$$

Consequently, the test is able to detect local alternatives converging to the null hypothesis with any rate a_n, such that $\frac{a_n}{k_n}\sqrt{n} \to \infty$ when the sample size and dimension k_n of Z is increasing.

Remark 3.8 Jeong et al. (2012) investigated an alternative test for the hypothesis (2.1) based on ideas from Fan and Li (1996) in combination with a modification which was originally proposed by Zheng (1998). Their test is based on the statistic

$$J_n = \frac{1}{n(n-1)g_n^d} \sum_{i,j,i \neq j} L((Z_i - Z_j)/g_n)(I\{Y_i \leq \hat{Q}(\tau|X_i)\} - \tau)(I\{Y_j \leq \hat{Q}(\tau|X_j)\} - \tau)$$

where L is a kernel and g_n is a bandwidth converging to 0 with increasing sampling size. These authors claimed that a normalized version of this test statistic converges to a normal distribution. It should be pointed out here that the proof in this paper is not correct. The basic argument of Jeong et al. (2012) consists in the statement that the fact

$$\sup_x \mid \hat{Q}_\tau(x) - Q_\tau(x) \mid \leq C_n$$

results in the estimate

$$J_{nU} \leq J_n \leq J_{nL}, \tag{3.4}$$

where the statistics J_{nU} and J_{nL} are defined by

$$\begin{aligned} J_{nU} &= \frac{1}{n(n-1)g^d} \sum_{i \neq j} L((Z_i - Z_j)/g)\varepsilon_{iU}\varepsilon_{jU}, \\ J_{nL} &= \frac{1}{n(n-1)g^d} \sum_{i \neq j} L((Z_i - Z_j)/g)\varepsilon_{iL}\varepsilon_{jL}, \end{aligned}$$

and $\varepsilon_{iU} = I\{Y_i + C_n \leq Q_\tau(X_i)\} - \tau,\ \varepsilon_{iL} = I\{Y_i + C_n \leq Q_\tau(X_i)\} - \tau$ (see equation (A.11-3) in this paper). A simple calculation shows that this conclusion is not correct and in fact the inequality (3.4) does not hold. It turns out that the proof of Theorem 1 in Jeong et al. (2012) can not be corrected easily.

Even if the gap in the proof would be closed, the test of Jeong et al. (2012) still has two major drawbacks. First, it requires non-parametric smoothing with respect to the covariate Z. Second, it can only detect local alternatives converging to the null hypothesis at a rate $n^{-1/2}h^{-(d+q)/4}$ which is slower than the rate $b_n n^{-1/2}$ for any $b_n \to \infty$ detected by the test proposed in this paper and additionally depends on the dimension of the covariates.

4 Bootstrap and simulation results

In general the limit distribution derived in Theorem 3.2 depends on certain features of the data generating process which are difficult to estimate. For this reason we discuss in this section bootstrap methods that are suitable to mimic the distribution of test statistics based on T_n under the null hypothesis. To be precise, let P^* denote the conditional probability $P(\cdot \mid \mathcal{Y}_n)$, given the original sample $\mathcal{Y}_n = \{(Y_i, X_i, Z_i) \mid i = 1, \ldots, n\}$, and denote by $\mathbb{E}^*$ and Cov^* the corresponding conditional expectation and covariance. Several residual wild bootstrap approximations have been proposed in the literature for quantile regression analysis [see Sun (2006) or Feng et al. (2011)]. However, the residual wild bootstrap does not yield a valid approximation of the limiting distribution in the present context because it does not lead to an expansion of the bootstrap process analogous to the one given for T_n in Theorem 3.2.

As alternative we consider the idea of process-based wild bootstrap as considered by Delgado and González-Manteiga (2001) or He and Zhu (2003). To this end recall the definition of the "residuals" $\hat{\varepsilon}_i = Y_i - \hat{q}_\tau(X_i)$, where $\hat{q}_\tau$ denotes an estimator for the conditional τ-quantile of

Y_i, given X_i, define $\hat{\tau} = \sum_{j=1}^n I\{\hat{\varepsilon}_j \leq 0\}/n$ and introduce independent identically distributed Bernoulli random variables $B_1, \ldots, B_n$ with success probability $\hat{\tau}$, which are independent of the original data. Define the bootstrap process as

$$T_n^*(\Theta, z) = \frac{1}{n} \sum_{i=1}^n (B_i - \hat{\tau}) I\{X_i \in \Theta\} \Big(I\{Z_i \leq z\} - \hat{F}_{Z|X,\varepsilon}(z|X_i, 0) \Big),$$

where

$$\hat{F}_{Z|X,\varepsilon}(\cdot|x, y) = \frac{\sum_{j=1}^n I\{Z_j \leq \cdot\} L(\frac{X_j - x}{a}) N(\frac{\hat{\varepsilon}_j - y}{e})}{\sum_{j=1}^n L(\frac{X_j - x}{a}) N(\frac{\hat{\varepsilon}_j - y}{e})} \tag{4.1}$$

denotes a kernel estimator for the conditional distribution $F_{Z|X,\varepsilon}(\cdot|x, y)$. Here, L and N denote d- and one-dimensional kernel functions and a and e corresponding bandwidths converging to 0 with increasing sample size. For the sake of brevity we do not consider conditional weak convergence of the process T_n^* in detail, but note that $E^*[T_n^*(\Theta, z)] = 0$ and under the null hypothesis H_0 (and under suitable regularity conditions) the conditional covariance $n\text{Cov}^*(T_n^*(\Theta_1, y), T_n^*(\Theta_2, z))$ converges in probability to the covariance $\text{Cov}(T(\Theta_1, y), T(\Theta_2, z))$ as defined in Theorem 3.2.

In our numerical investigations, it turned out that the asymptotic representation (3.1) for the process defined in (2.3) is not very accurate for small sample sizes. We thus considered a slightly modified version of this process, that is

$$\tilde{T}_n(x, z) = \frac{1}{n} \sum_{i=1}^n \left(I\{Y_i \leq \hat{q}_\tau(X_i)\} - \hat{\tau} \right) I\{X_i \leq x\} (I\{Z_i \leq z\} - \hat{F}_Z(z))$$

where $\hat{F}_Z(z)$ denotes the empirical distribution function of $Z_1, ..., Z_n$, which provided much better results for moderate sample sizes. As motivation for this approach, observe that under both the null hypothesis and the alternative, we have

$$D_x := \frac{1}{n} \sum_{i=1}^n \left(I\{Y_i \leq \hat{q}_\tau(X_i)\} - \hat{\tau} \right) I\{X_i \leq x\} = o_P(n^{-1/2}), \quad \hat{\tau} = \tau + o_P(n^{-1/2})$$

uniformly with respect to x as can be seen by taking a closer look at the proofs of the main results in the Appendix. Thus the additional correction term

$$\delta_{x,z} := D_x \hat{F}_Z(z) + \frac{\hat{\tau} - \tau}{n} \sum_{i=1}^n I\{X_i \leq x\} I\{Z_i \leq z\}$$

vanishes asymptotically (uniformly with respect to x, z) under both the alternative and the null hypothesis. If, on the other hand, $\delta_{x,z}$ is relatively large because the sample size is small, the correction term $\delta_{x,z}$ induces an additional centering (the factor $\hat{F}_Z(z)$ corresponds to the amount of non-zero indicators $I\{Z_i \leq z\}$).

The simulation results described below confirm that this is a sensible approach.

For the calculation of the test statistic

$$\tilde{K}_n = \sup_x \sup_z | \tilde{T}_n(x, z) | \tag{4.2}$$

based on the process $\tilde{T}_n$, we use local polynomial estimators of order two [see (2.4)]. The bandwidth h_n of this estimator is chosen as $h_n := (\hat{\sigma}^2/2n)^{13/50}$ where $\hat{\sigma}^2$ denotes the variance estimate of Rice (1984) from the sample $\{(X_i, Y_i)|\ i = 1, \dots, n\}$ [see Yu and Jones (1997) for a related approach]. The bandwidths used in (2.5) and (4.1) are chosen as $d_n = a = e = h_n$, while the choice of b_n in (2.7) is even less critical [see also Dette and Volgushev (2008)] and we use $b_n = h_n^3$. In fact, in the simulations it turned out that the power and size properties of the test are rather insensitive with respect to the bandwidth choice, see table 3 and related discussion in the next paragraph. The function ω in (2.6) is chosen as $\omega(x) := (15/32)(3-10x^2+7x^4)I\{|x| \leq 1\}$, which is a kernel of order 2 [see Gasser et al. (1985)]. The function κ in (2.7) is defined as Epanechnikov kernel while all other kernels are Gaussian kernels. For the choice of the distribution function G in (2.7) we follow the procedure described in Dette and Volgushev (2008) who suggested a normal distribution such that the 5% and 95% quantiles coincide with the corresponding empirical quantities of the sample $Y_1, ..., Y_n$.

4.1 Simulation results

We simulate data from the location scale model

$$Y_i = q_j(X_i, Z_i) + s_k(X_i, Z_i)\varepsilon_i, \tag{4.3}$$

$j, k = 1, \dots, 4$ with the following quantile and scale functions

$$\begin{aligned} q_1(x,z) &= \exp(2x^2)\ ,\ q_2(x,z) = (x-0.5)^2 \\ q_3(x,z) &= \exp(2x^2)z^2\ ,\ q_4(x,z) = \sin(2\pi(x+z)) \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} s_1(x,z) &= 0.5(x+0.2)\ ,\ s_2(x,z) = 0.5(\sin(x)+1.2) \\ s_3(x,z) &= 0.5(z+0.2)\ ,\ s_4(x,z) = 0.5\sqrt{(x+0.2)(z+0.2)}. \end{aligned} \tag{4.5}$$

The random variables X and Z are independent and uniformly distributed on the interval $[0,1]$ while ε is standard normal. We consider the cases $\tau = 0.5$ and $\tau = 0.25$. All reported results are based on 1000 simulation runs with 300 bootstrap replications.
The bootstrap test (at level α) rejects the null hypothesis that the variable Z is not significant, whenever

$$\tilde{K}_n > K^*_{n,1-\alpha} \tag{4.6}$$

where $\tilde{K}_n$ is defined in (4.2) and $K^*_{n,1-\alpha}$ denotes the $(1-\alpha)$ bootstrap quantile of the Kolmogorov-Smirnov test statistic.
The rejection probabilities of this test under the null hypothesis are shown in Table 1 for the 50% and 25% quantile. Note that different pairs of location and scale functions in (4.4) and (4.5) correspond to the null hypothesis for $\tau = 0.5$ and $\tau = 0.25$ (more precisely the models

τ	(k,l)	$\alpha = 0.025$		$\alpha = 0.05$		$\alpha = 0.1$	
		$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
0.5	(1,1)	0.037	0.035	0.053	0.061	0.102	0.111
	(1,2)	0.026	0.025	0.044	0.048	0.090	0.101
	(1,3)	0.041	0.027	0.069	0.066	0.132	0.127
	(1,4)	0.040	0.033	0.060	0.059	0.120	0.121
	(2,1)	0.036	0.031	0.068	0.057	0.122	0.106
	(2,2)	0.024	0.028	0.051	0.046	0.092	0.085
	(2,3)	0.037	0.025	0.057	0.059	0.132	0.114
	(2,4)	0.027	0.024	0.050	0.047	0.109	0.093
0.25	(1,1)	0.024	0.019	0.044	0.035	0.089	0.082
	(1,2)	0.024	0.019	0.044	0.037	0.089	0.092
	(2,1)	0.027	0.025	0.047	0.052	0.102	0.105
	(2,2)	0.016	0.022	0.036	0.048	0.089	0.101

Table 1: *Simulated rejection probabilities of the bootstrap test (4.6) for significance of the variable z in the quantile regression model (4.3) for $\tau = 0.5$ (upper part) and $\tau = 0.25$ (lower part) under various null hypotheses. The pair (k, l) corresponds to the location function q_k and scale function s_ℓ specified in (4.4) and (4.5), respectively.*

τ	(k,l)	$\alpha = 0.025$		$\alpha = 0.05$		$\alpha = 0.1$	
		$n = 50$	$n = 100$	$n = 50$	$n = 100$	$n = 50$	$n = 100$
0.5	(3,1)	0.999	1.000	1.000	1.000	1.000	1.000
	(3,2)	0.756	0.983	0.815	0.989	0.886	0.997
	(3,3)	0.997	1.000	0.999	1.000	0.999	1.000
	(3,4)	1.000	1.000	1.000	1.000	1.000	1.000
	(4,1)	0.082	0.197	0.142	0.311	0.252	0.519
	(4,2)	0.034	0.070	0.067	0.119	0.138	0.237
	(4,3)	0.089	0.176	0.134	0.279	0.226	0.488
	(4,4)	0.070	0.203	0.123	0.321	0.218	0.508
0.25	(1,3)	0.099	0.240	0.163	0.325	0.245	0.459
	(1,4)	0.044	0.078	0.086	0.133	0.155	0.225
	(2,3)	0.139	0.295	0.204	0.405	0.332	0.540
	(2,4)	0.06	0.089	0.106	0.152	0.176	0.232
	(3,1)	0.935	1.000	0.971	1.000	0.988	1.000
	(3,2)	0.464	0.857	0.591	0.913	0.725	0.954
	(3,3)	0.792	0.990	0.873	0.996	0.934	0.999
	(3,4)	0.900	1.000	0.948	1.000	0.975	1.000
	(4,1)	0.027	0.054	0.055	0.103	0.111	0.229
	(4,2)	0.019	0.031	0.034	0.061	0.078	0.132
	(4,3)	0.022	0.051	0.043	0.091	0.104	0.176
	(4,4)	0.021	0.054	0.053	0.093	0.104	0.195

Table 2: *Simulated rejection probabilities of the bootstrap test (4.6) for significance of the variable* z *in the quantile regression model (4.3) for* $\tau = 0.5$ *(upper part) and* $\tau = 0.25$ *(lower part) under various alternatives. The pair* (k,l) *corresponds to the location function* q_k *and scale function* s_ℓ *specified in (4.4) and (4.5), respectively.*

τ	h	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.4	0.45	0.5
0.5	(1,2)	0.037	0.036	0.037	0.037	0.047	0.054	0.061	0.046	0.047	0.043
	(3,2)	0.238	0.301	0.361	0.389	0.388	0.385	0.381	0.389	0.412	0.404
0.25	(1,2)	0.017	0.031	0.037	0.033	0.031	0.048	0.042	0.049	0.041	0.053
	(3,2)	0.113	0.160	0.210	0.210	0.237	0.250	0.262	0.246	0.262	0.260

Table 3: *Simulated rejection probabilities of the bootstrap test (4.6) for various bandwidths. The sample size is* $n = 50$ *and the lower and upper part correspond to the* 50% *and* 25% *quantile, respectively. The pair* (k, l) *corresponds to the location function* q_k *and scale function* s_ℓ *specified in (4.4) and (4.5), respectively.*

α	0.025	0.050	0.100
q_1	0.026	0.042	0.096
q_2	0.998	1.000	1.000

Table 4: *Simulated rejection probabilities of the bootstrap test (4.6) for the significance of a two dimensional predictor in median regression. The models are defined in (4.7), the sample size is* $n = 50$ *and the upper (lower) row corresponds to the null hypothesis (alternative)*

defined by the pairs $(1, 3)$, $(1, 4)$, $(2, 3)$ and $(2, 4)$ correspond to the null hypothesis if $\tau = 0.5$ but to the alternative if $\tau = 0.25$). We observe from Table 1 that the level is usually approximated very well. For $\tau = 0.25$ there exist some cases where the test is slightly conservative .

The corresponding results for various alternatives are displayed in Table 2 and we observe a reasonable power for most cases. The power for $\tau = 0.25$ is always smaller than the power for $\tau = 0.5$. This corresponds to intuition because the 25%-quantile is more difficult to estimate than the median. The power of the test is smaller for alternatives corresponding to the location function $q_4(x, z) = \sin(2\pi(x + z))$ if the sample size is $n = 100$. However, if the the sample size is larger, the test also detects the alternatives with reasonable probability. For example if $n = 200$ and $\tau = 0.5$ the simulated rejection probabilities of the bootstrap test at level 5% for the alternatives $(4, 2)$, $(4, 3)$ and $(4, 4)$ are given by 0.319, 0.795 and 0.821, respectively.

Next we study the impact of the choice of the bandwidth on size and power of the bootstrap test. For this purpose we consider the sample size $n = 50$ and bandwidths 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45 and 0.50. The results for model $(1, 2)$ and $(3, 2)$ corresponding to the null hypothesis and alternative, respectively, are summarized in Table 3. We observe that the level and power are rather stable with respect to different choices of the bandwidth. Simulations for other scenarios yield similar results and are not shown for the sake of brevity.

We conclude our numerical study with a brief investigation of a two dimensional predictor, say $Z = (Z_1, Z_2)$. Because the method proposed in this paper does not require smoothing in the Z-direction, the results should not be seriously affected, if the dimension of Z is larger. To be precise we consider two different location functions

$$q_1(x, z_1, z_2) = x \ , \ q_2(x, z_1, z_2) = z_2 \cdot x + z_1^2 \tag{4.7}$$

and a constant scale function $s(x, z_1, z_2) = 0.5$ in model (4.3). Note that q_1 corresponds to the null hypothesis, while q_2 represents an alternative. The results of the bootstrap test for the median are listed in Table 4 for the sample size $n = 50$ and we observe in these examples similar satisfactory properties as in the one-dimensional setting.

Acknowledgements. The authors thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C1, C4) of the German Research Foundation (DFG).

References

Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression in high-dimensional sparse models. *Ann. Statist.*, 39(1):82–130.

Bondell, H. D., Reich, B. J., and Wang, H. (2010). Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics*, 19:760–777.

Chernozhukov, V., Fernandéz-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Delgado, M. A. and González-Manteiga (2001). Significance testing in nonparametric regression. *Annals of Statistics*, 29:1469–1507.

Dette, H., Neumeyer, N., and Pilz, K. F. (2006). A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12:469–490.

Dette, H. and Volgushev, S. (2008). Non-crossing nonparametric estimates of quantile curves. *Journal of the Royal Statistical Society, Ser. B*, 70(3):609–627.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall.

Fan, Y. and Li, Q. (1996). Consistent model specification tests: Omitted variables and semi-parametric functional forms. *Econometrica*, 64:865–890.

Feng, X., He, X., and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*, 98(4):995–999.

Gasser, T., Muller, H., and Mammitzsch, V. (1985). Kernels for nonparametric curve estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 238–252.

Gozalo, P. L. (1993). A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory*, 9:451–577.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424–438.

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

Jeong, K., Härdle, W. K., and Song, S. (2012). A consistent nonparametric test for causality in quantile. *Econometric Theory*, 3:1–27.

Lavergne, P. and Vuong, Q. H. (1996). Nonparametric selection of regressors: The nonnested case. *Econometrica*, 64:207–219.

Lavergne, P. and Vuong, Q. H. (2000). Nonparametric significance testing. *Econometric Theory*, 16:576–601.

Neumeyer, N. and Van Keilegom, I. (2010). Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, 101:1067–1078.

Racine, J. (1997). Feasible cross-validatory model selection for general stationary processes. *Journal of Applied Econometrics*, 12:169–179.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 12(4):1215–1230.

Sun, Y. (2006). A consistent nonparametric equality test of conditional quantile functions. *Econometric Theory*, 22:614–632.

Takeuchi, I., Le, Q. V., Sears, T. D., and Smola, A. J. (2006). Nonparametric quantile regression. *Journal of Machine Learning Research*, 7:1231–1264.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics.* Springer, New York.

Volgushev, S. (2006). *Nonparametric quantile regression for censored data.* PhD thesis, Fakultät für Mathematik, Ruhr-Universität Bochum, Germany.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19:801–817.

Yatchew, A. J. (1992). Nonparametric regression tests based on least squares. *Econometric Theory*, 8:435–451.

Yu, K. and Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics and Data Analysis*, 25:159–166.

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93:228–237.

Zheng, J. X. (1998). A consistent test of parametric regression models under conditional quantile restrictions. *Econometric Theory*, 14:123–138.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Annals of Statistics*, 36:1108–1126.

A Appendix: Proofs

Throughout this section, introduce the abbreviation $\Theta_n := \Theta \cap \mathcal{D}_n$ with $\mathcal{D}_n := \{x : [x - h_n, x + h_n] \subset R_X\}$.

Lemma A.1 *If assumptions (K1)-(K6) and (A1)-(A3) are satisfied, then*

$$\hat{q}_\tau(x) = q_\tau(x) - \frac{1}{f_{\varepsilon|X}(0|x)} \int_{-1}^{1} \kappa(v) \Delta_S(q_{\tau+vb_n}(x)|x) dv + o_P(n^{-1/2}) =: \hat{q}_{\tau,L}(x) + o_P(n^{-1/2})$$

uniformly in $x \in \mathcal{D}_n$ where $\Delta_S(x, y)$ is defined in Lemma B.1 and has the property

$$\sup_{v \in [-1,1], x \in \mathcal{D}_n} |\Delta_S(q_{\tau+vb_n}(x)|x)| = O_P\Big(d_n^s + \Big(\frac{\log n}{nh_n^d}\Big)^{1/2}\Big).$$

Moreover, $\hat{q}_{\tau,L}(x)$ is, with probability tending to one, $d+1$ times continuously differentiable with derivatives bounded uniformly on $\mathcal{D}_n$.

Proof. Apply part (a) of Lemma B.4 to $F_{Y|X}(\cdot|x)$ and part (c) of the same Lemma with $F_1(\cdot|x) = F_{Y|X}(\cdot|x), F_2(\cdot|x) = \hat{F}_{Y|X}(\cdot|x; p)$. Combined the results with Lemma B.1 yields the assertion. □

Lemma A.2 *If assumptions (K1) - (K6), (A1) - (A4), (S1) and (S2) are satisfied, then*

$$\int f_{\varepsilon|X}(0 \mid s)(\hat{q}_\tau(s) - q_\tau(s)) I\{s \in \Theta_n\} f_X(s) F_{Z|X,\varepsilon}(z|s, 0)\, ds$$
$$= -\frac{1}{n} \sum_{i=1}^{n} \Big(I\{\varepsilon_i \le 0\} - \tau\Big) I\{X_i \in \Theta_n\} F_{Z|X,\varepsilon}(z|X_i, 0) + o_P(\frac{1}{\sqrt{n}})$$

uniformly with respect to $\Theta \in \Xi, z \in R_Z$.

Proof. ¿¿From Lemma A.1 we obtain the representation

$$
\begin{aligned}
&-\int f_{\varepsilon|X}(0|s)(\hat{q}_\tau(s)-q_\tau(s))I\{s\in\Theta_n\}f_X(s)F_{Z|X,\varepsilon}(z|s,0)ds \\
&= \int_{-1}^{1}\kappa(v)\int \Delta_S(q_{\tau+vb_n}(s)|s)I\{s\in\Theta_n\}f_X(s)F_{Z|X,\varepsilon}(z|s,0)dsdv + o_P(n^{-1/2}) \\
&= \int_{-1}^{1}\kappa(v)\int \frac{1}{nh_n^d}\sum_i \mathbf{M}(s)\Big(\Omega\Big(\frac{q_{\tau+vb_n}(s)-Y_i}{d_n}\Big)-F_{Y|X}(q_{\tau+vb_n}(s)|X_i)\Big) \\
&\qquad \times\Big(\mathbf{K}_{h_n,0}(s-X_i),...,\mathbf{K}_{h_n,\mathbf{k}_{N_p,p}}(s-X_i)\Big)^t \\
&\qquad \times I\{s\in\Theta_n\}(I_1(X_i;\Theta_n,h_n)+I_2(X_i;\Theta_n,h_n))f_X(s)F_{Z|X,\varepsilon}(z|s,0)dsdv\ + o_P(n^{-1/2}),
\end{aligned}
$$

where

$$
\mathbf{M}(s) := e_1^t\Big(\sum_{j=0}^{n_f}(-1)^j\Big(\frac{\mathcal{M}(K)^{-1}}{f_X(x)}\sum_{1\le|\mathbf{m}|<n_f}h_n^{|\mathbf{m}|}f_X^{(\mathbf{m})}(x)M_{\mathbf{m}}\Big)^j\frac{\mathcal{M}(K)^{-1}}{f_X(x)}\Big)
$$

and

$$
\begin{aligned}
I_1(X;\Theta_n,h_n) &:= I\{\otimes_{j=1}^d[X(j)-h_n,X(j)+h_n]\subset\Theta_n\}, \\
I_2(X;\Theta_n,h_n) &:= I\{\exists j:[X(j)-h_n,X(j)+h_n]\not\subset\Theta_n,\otimes_{j=1}^d[X(j)-h_n,X(j)+h_n]\cap\Theta_n\neq\emptyset\}.
\end{aligned}
$$

We will now proceed to show that the first part in the above decomposition [i.e. the part containing I_1] determines the asymptotic expansion and establish at the end of the proof that the part corresponding to I_2 is asymptotically negligible. First, note that

$$
\begin{aligned}
&\int \frac{1}{nh_n^d}\sum_i \mathbf{M}(s)\Big(\Omega\Big(\frac{q_{\tau+vb_n}(s)-Y_i}{d_n}\Big)-F_{Y|X}(q_{\tau+vb_n}(s)|X_i)\Big) \\
&\qquad \times\Big(\mathbf{K}_{h_n,0}(s-X_i),...,\mathbf{K}_{h_n,\mathbf{k}_{N_p,p}}(s-X_i)\Big)^t I\{s\in\Theta_n\}I_1(X_i;\Theta_n,h_n)f_X(s)F_{Z|X,\varepsilon}(z|s,0)ds \\
&= \int_{[-1,1]^d}\frac{1}{n}\sum_i \mathbf{M}(X_i+h_ns)\Big(\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-Y_i}{d_n}\Big)-F_{Y|X}(q_{\tau+vb_n}(X_i+sh_n)|X_i)\Big) \\
&\qquad \times\Big(\mathbf{K}_{1,0}(s),...,\mathbf{K}_{1,\mathbf{k}_{N_p,p}}(s)\Big)^t I_1(X_i;\Theta_n,h_n)f_X(X_i+sh_n)F_{Z|X,\varepsilon}(z|X_i+sh_n,0)ds.
\end{aligned}
$$

Observe that every entry of $\mathbf{M}$ is by assumption continuously differentiable with respect to s and the derivative is uniformly bounded. The class of functions defined by

$$
\Big\{(x,y)\mapsto\Omega\Big(\frac{q_\zeta(x+a)-y}{d_n}\Big)\Big||a(j)|\le 1, j=1,...,d, |\zeta-\tau|\le\alpha\Big\}
$$

where α is a small positive number has covering numbers that satisfy the assumptions of part 1 of Lemma B.3 in Appendix B. This follows from Lemma B.2 together with the fact that under the assumptions (A1), (A3) the mapping $(\zeta,a)\mapsto q_\zeta(x+a)$ satisfies

$$
\sup_x|q_{\zeta_1}(x+a_1)-q_{\zeta_2}(x+a_2)|\le C(|\zeta_1-\zeta_2|+\|a_1-a_2\|_\infty)
$$

for some finite constant C (this inequality is a consequence of the implicit function theorem). Moreover, it follows from the smoothness assumptions on $F_{Y|X}$ and the properties of Ω that

$$\sup_{|s|\le 1,|v|\le 1}\Big|\mathbb{E}\Big[\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-Y_i}{d_n}\Big)-F_{Y|X}(q_{\tau+vb_n}(X_i+sh_n)|X_i)\Big|X_i\Big]\Big|\le R_n \quad a.s.,$$

where R_n is a nonrandom quantity of order $o(1/\sqrt{n})$. Thus the smoothness properties of $F_{Z|X,\varepsilon}, F_{Y|X}$ and $(\zeta,x)\mapsto q_\zeta(x)$ imply that by Lemma B.2 and Lemma B.3 in Appendix B we have

$$\begin{aligned}
&\frac{1}{n}\sum_i \mathbf{M}(X_i+h_n s)\Big(\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-Y_i}{d_n}\Big)-F_{Y|X}(q_{\tau+vb_n}(X_i+sh_n)|X_i)\Big)\\
&\qquad\times(\mathbf{K}_{1,0}(s),...,\mathbf{K}_{1,\mathbf{k}_{N_p,p}}(s))^t I_1(X_i;\Theta_n,h_n)f_X(X_i+sh_n)F_{Z|X,\varepsilon}(z|X_i+sh_n,0)\\
=&\ \frac{1}{n}\sum_i \mathbf{M}(X_i)\Big(\mathbf{K}_{1,0}(s),...,\mathbf{K}_{1,\mathbf{k}_{N_p,p}}(s)\Big)^t I\{X_i\in\Theta_n\}f_X(X_i)F_{Z|X,\varepsilon}(z|X_i,0)\\
&\qquad\times\Big(\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-Y_i}{d_n}\Big)-F_{Y|X}(q_{\tau+vb_n}(X_i+sh_n)|X_i)\Big)+o_P(n^{-1/2})
\end{aligned}$$

uniformly with respect to $|v|\le 1, s\in[-1,1]^d, \Theta\in\Xi$ and $z\in R_Z$. Finally, noting that

$$\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-Y_i}{d_n}\Big)=\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-q_\tau(X_i)-\varepsilon_i}{d_n}\Big)$$

yields

$$\sup_{v,s,i}\Big|\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-Y_i}{d_n}\Big)-I\{\varepsilon_i\le 0\}\Big|\le \|\Omega\|_\infty I\{|\varepsilon_i|\le R_n\}\quad a.s.,$$

where $R_n=O(h_n+b_n+d_n)$ is a non-random quantity. This, together with an application of Lemma B.3, shows that

$$\begin{aligned}
&\frac{1}{n}\sum_i \mathbf{M}(X_i)(\mathbf{K}_{1,0}(s),...,\mathbf{K}_{1,\mathbf{k}_{N_p,p}}(s))^t I\{X_i\in\Theta_n\}f_X(X_i)F_{Z|X,\varepsilon}(z|X_i,0)\\
&\qquad\times\Big(\Omega\Big(\frac{q_{\tau+vb_n}(X_i+sh_n)-Y_i}{d_n}\Big)-F_{Y|X}(q_{\tau+vb_n}(X_i+sh_n)|X_i)\Big)\\
=&\ \frac{1}{n}\sum_i \mathbf{M}(X_i)(I\{\varepsilon_i\le 0\}-F_{\varepsilon|X}(0|X_i))(\mathbf{K}_{1,0}(s),...,\mathbf{K}_{1,\mathbf{k}_{N_p,p}}(s))^t\\
&\qquad\times I\{X_i\in\Theta_n\}f_X(X_i)F_{Z|X,\varepsilon}(z|X_i,0)+o_P(n^{-1/2}).
\end{aligned}$$

In particular, noting that $F_{\varepsilon|X}(0|X_i)=\tau$, the above result implies

$$\begin{aligned}
&\int f_{\varepsilon|X}(0|s)(\hat{q}_\tau(s)-q_\tau(s))I\{s\in\Theta_n\}f_X(s)F_{Z|X,\varepsilon}(z|s,0)ds\\
=&\ \frac{1}{n}\sum_i \mathbf{M}(X_i)(I\{\varepsilon_i\le 0\}-\tau)(\mu_0(K),...,\mu_{\mathbf{k}_{N_p,p}}(K))^t I\{X_i\in\Theta_n\}f_X(X_i)F_{Z|X,\varepsilon}(z|X_i,0)\\
&+o_P(n^{-1/2}),
\end{aligned}$$

where $\mu_{\mathbf{k}}(K) := \int_{\mathbb{R}^d} \mathbf{K}_{\mathbf{1},\mathbf{k}}(u)du$. Now from the definition of $\mathbf{M}$ it is easy to see that

$$\mathbf{M}(x) = e_1^t(M_0(x)^{-1} + h_n R_M(x)) = e_1^t\Big(\frac{\mathcal{M}(K)^{-1}}{f_X(x)} + h_n R_M(x)\Big)$$

where R_M denotes a vector whose entries are uniformly bounded and Lipschitz-continuous with respect to x. Thus applying Lemma B.3 we obtain

$$\begin{aligned}&\frac{1}{n}\sum_i \mathbf{M}(X_i)(I\{\varepsilon_i \le 0\} - \tau)(\mu_0(K), ..., \mu_{\mathbf{k}_{N_p,p}}(K))^t I\{X_i \in \Theta_n\} f_X(X_i) F_{Z|X,\varepsilon}(z|X_i, 0)\\ = \;& \frac{1}{n}\sum_{i=1}^n (I\{\varepsilon_i \le 0\} - \tau) I\{X_i \in \Theta_n\} F_{Z|X,\varepsilon}(z|X_i, 0) + o_P(n^{-1/2}),\end{aligned}$$

which completes the first part of the proof.
It remains to show that

$$\begin{aligned}&\frac{1}{n}\sum_i I_2(X_i;\Theta_n, h_n)\int_{-1}^1 \kappa(v)\int \frac{1}{h_n^d}\mathbf{M}(s)\Big(\Omega\Big(\frac{q_{\tau+vb_n}(s) - Y_i}{d_n}\Big) - F_{Y|X}(q_{\tau+vb_n}(s)|X_i)\Big)\\ &\qquad \times(\mathbf{K}_{h_n,0}(s - X_i), ..., \mathbf{K}_{h_n,\mathbf{k}_{N_p,p}}(s - X_i))^t I\{s \in \Theta_n\} f_X(s) F_{Z|X,\varepsilon}(z|s, 0) ds dv = o_P(n^{-1/2})\end{aligned}$$

uniformly with respect to $\Theta \in \Xi, z \in R_Z$. To this end, consider the (n-dependent) class of functions $\mathcal{F}_n$ with elements

$$\begin{aligned}f_{z,\Theta_n,h_n,b_n}(x, y) \;=\; &\int_{-1}^1 \kappa(v)\int \frac{1}{h_n^d}\mathbf{M}(s)\Big(\Omega\Big(\frac{q_{\tau+vb_n}(s) - y}{d_n}\Big) - F_{Y|X}(q_{\tau+vb_n}(s)|x)\Big)\\ &\times\; (\mathbf{K}_{h_n,0}(s - x), ..., \mathbf{K}_{h_n,\mathbf{k}_{N_p,p}}(s - x))^t I\{s \in \Theta_n\} f_X(s) F_{Z|X,\varepsilon}(z|s, 0) ds dv\end{aligned}$$

indexed by $z \in \mathcal{Z}, \Theta \in \Xi$ contains uniformly bounded elements (the bound is also uniform with respect to n). Moreover, there exists a finite positive constant C such that

$$\text{(A.1)} \qquad N_{[]}(\mathcal{F}_n, \varepsilon, L^2(P_X)) \le \Big(N_{[]}(\mathcal{F}_{n,1}, \varepsilon/C, L^2(P_X)) N_{[]}(\mathcal{F}_{n,2}, \varepsilon/C, L^2(P_X))\Big)^2,$$

where $\mathcal{F}_{n,1} := \{s \mapsto I\{s \in \Theta_n\} | \Theta \in \Xi\}$ and $\mathcal{F}_{n,2} := \{s \mapsto F_{Z|X,\varepsilon}(z|s, \varepsilon) | z \in \mathcal{Z}\}$. To see that this holds, observe the decomposition

$$\begin{aligned}f_{z,\Theta_n,h_n,b_n}(x, y) &= f^{(1)}_{z,\Theta_n,h_n,b_n}(x, y) + f^{(2)}_{z,\Theta_n,h_n,b_n}(x, y)\\ &:= \frac{1}{h_n^d}\sum_{j=1}^2 \int\int \kappa(v) I\{\|x - s\|_\infty \le h_n\} f_X(s) g_{j,n}(x, y, s, v) I\{s \in \Theta_n\} F_{Z|X,\varepsilon}(z|s, 0) ds dv\end{aligned}$$

where $g_{1,n}$ and $g_{2,n}$ denote non-positive and non-negative, uniformly bounded functions, respectively. Moreover, $g_{j,n}$ do not depend on Θ_n or z. Obviously, it suffices to bound the bracketing number of $\mathcal{F}_{j,n} := \{(x, y) \mapsto f^{(j)}_{z,\Theta_n,h_n,b_n}(x, y)\}$ for $j = 1, 2$ separately. If we denote by $\{[b_{L,j}, b_{U,j}]\}$ a collection of $\varepsilon-$brackets (with respect to $L^2(P_X)$) for $\{s \mapsto I\{s \in \Theta_n\} F_{Z|X,\varepsilon}(z|s, 0)\}$. Then a collection of ε/C brackets for $\mathcal{F}_{n,2}$ (with respect to $L^2(P_{X,Y})$) is given by

$$B_{K,j}(x, y) := \frac{1}{h_n^d}\int\int \kappa(v) I\{\|x - s\|_\infty \le h_n\} f_X(s) g_{2,n}(x, y, s, v) b_{K,j}(s) ds dv, \quad K = U, L.$$

To see this, observe that

$$
\begin{aligned}
&\mathbb{E}[(B_{L,j}(X_1,Y_1)-B_{U,j}(X_1,Y_1))^2]\\
\le\ & \int\int\int g_{2,n}^2(x,y,s,v)\frac{1}{h_n^d}\kappa(v)I\{\|x-s\|_\infty\le h_n\}f_X(s)\kappa(v)dsdv\\
&\times\int\int\kappa(v)\frac{1}{h_n^d}I\{\|x-s\|_\infty\le h_n\}f_X(s)(b_{U,j}(s)-b_{L,j}(s))^2dsdvf_{X,Y}(x,y)dxdy\\
\le\ & C_1\int f_X(s)(b_{U,j}(s)-b_{L,j}(s))^2\int\frac{1}{h_n^d}I\{\|x-s\|_\infty\le h_n\}f_X(x)dxds
\end{aligned}
$$

for some finite constant C_1. A bound for $\mathcal{F}_{n,2}$ can be derived by similar arguments. Thus (A.1) is established. Combining the bound in (A.1) with the assumptions (S1) and (S2), the estimate $\sup_{z,\Theta}|\mathbb{E}[f_{z,\Theta_n,h_n,b_n}(X_1,Y_1)]|=o(n^{-1/2})$, and the results from Lemma B.2 and Lemma B.3 yields the assertion after noting that by assumption $\sup_{\Theta\in\Xi}\mathbb{E}I_2(X_i;\Theta_n,h_n)=o(1)$. □

Lemma A.3 *Under the assumptions of Theorem 3.2 it holds that*

$$
\begin{aligned}
T_n(\Theta_n,z) \ =\ & \frac{1}{n}\sum_{i=1}^n(I\{\varepsilon_i\le 0\}-\tau)I\{X_i\in\Theta_n\}I\{Z_i\le z\}+o_p(n^{-1/2})\\
&+\int(F_{\varepsilon|X,Z}(\hat q_{\tau,L}(s)-q_\tau(s)|s,t)-F_{\varepsilon|X,Z}(0|s,t))I\{s\in\Theta_n\}I\{t\le z\}dF_{X,Z}(s,t),
\end{aligned}
$$

uniformly with respect to $\Theta\in\Xi, z\in R_Z$, where $F_{X,Z}$ denotes the joint distribution function of X,Z.

Proof. Note that $T_n(\Theta,z)=\frac{1}{n}\sum_{i=1}^n(I\{\hat\varepsilon_i\le 0\}-\tau)I\{X_i\in\Theta\}I\{Z_i\le z\}$, and that the assertion is equivalent to

$$
\begin{aligned}
&\sup_{\Theta\in\Xi,z\in\mathcal{Z}}\Big|\frac{1}{n}\sum_{i=1}^n(I\{\hat\varepsilon_i\le 0\}-I\{\varepsilon_i\le 0\})I\{X_i\in\Theta_n\}I\{Z_i\le z\}\\
&-E\Big[(I\{\hat\varepsilon_L\le 0\}-I\{\varepsilon\le 0\})I\{X\in\Theta_n\}I\{Z\le z\}\ \Big|\ (Y_i,X_i,Z_i)_{i=1,\dots,n}\Big]\Big|=o_p(\frac{1}{\sqrt{n}}).
\end{aligned}
$$

Here we define $\hat\varepsilon_i=Y_i-\hat q_\tau(X_i)$, $\hat\varepsilon_L=Y-\hat q_{\tau,L}(X)$, where we assume that the sample (Y_i,X_i,Z_i), $i=1,\dots,n$, (used to build $\hat q_{\tau,L}$) is independent from the generic variable (Y,X,Z). The proof now proceeds in two steps. First, note that by Lemma A.1 we have $\hat q_\tau-\hat q_{\tau,L}=o_P(n^{-1/2})$ uniformly on $\mathcal{D}_n$ and thus there exists a deterministic sequence $\gamma_n=o(n^{-1/2})$ with

$$
P(\sup_{x\in\mathcal{D}_n}|\hat q_\tau(x)-\hat q_{\tau,L}(x)|\le\gamma_n)\to 1. \tag{A.2}
$$

Now on the set $\{|\hat q_\tau(x)-\hat q_{\tau,L}(x)|\le\gamma_n\}$, the probability of which tends to one, we have

$$
\sup_{\Theta\in\Xi,z\in\mathcal{Z}}\Big|\frac{1}{n}\sum_{i=1}^n(I\{\hat\varepsilon_i\le 0\}-I\{\hat\varepsilon_{i,L}\le 0\})I\{X_i\in\Theta_n\}I\{Z_i\le z\}\Big|\le\frac{1}{n}\sum_{i=1}^n I\{|\hat\varepsilon_{i,L}|\le\gamma_n\}I\{X_i\in\mathcal{D}_n\}
$$

Next, note that $I\{|\hat{\varepsilon}_{i,L}| \leq \gamma_n\} = I\{|\varepsilon_i - g(X_i)| \leq \gamma_n\}$ for $g = \hat{q}_{\tau,L} - q_\tau$. Now the assertion follows since the (n-dependent) class of functions

$$\Big\{(\epsilon, \xi) \mapsto I\{|\epsilon - g(\xi)| \leq \gamma_n\} I\{\xi \in \mathcal{D}_n\} \Big| g \in C_1^{d+1}(R_X)\Big\}$$

satisfies the assumptions of part 1 of Lemma B.3 whenever n is sufficiently large, see the proof of Lemma A.3 in Neumeyer and Van Keilegom (2010) for a similar reasoning, and $\hat{q}_{\tau,L} - q_\tau \in C_1^{d+1}(\mathcal{D}_n)$ with probability converging to one by Lemma A.1. Here $C_1^{d+1}(\mathcal{D}_n)$ is the class of $d+1$ times differentiable functions g defined on $\mathcal{D}_n$. Further, note that

$$\sup_{g \in C_1^{d+1}(\mathcal{D}_n)} \mathbb{E}\Big[I\{|\varepsilon_i - g(X_i)| \leq \gamma_n\} I\{X_i \in \mathcal{D}_n\}\Big] = o(n^{-1/2})$$

This, together with (A.2), and an application of Lemma B.3, shows that

$$\sup_{\Theta \in \Xi, z \in \mathcal{Z}} \Big|\frac{1}{n}\sum_{i=1}^n (I\{\hat{\varepsilon}_i \leq 0\} - I\{\hat{\varepsilon}_{i,L} \leq 0\}) I\{X_i \in \Theta_n\} I\{Z_i \leq z\}\Big| = o_P(n^{-1/2}).$$

Similar arguments applied to the (n-dependent) class functions

$$\Big\{(\epsilon, \xi, \zeta) \mapsto (I\{\epsilon \leq g(\xi)\} - I\{\epsilon \leq 0\}) I\{\xi \in \Theta_n\} I\{\zeta \leq z\} \Big| g \in C_1^{d+1}(R_X), \Theta \in \Xi, z \in \mathcal{Z}\Big\}$$

yield

$$\begin{aligned}\sup_{\Theta \in \Xi, z \in \mathcal{Z}} \Big|&\frac{1}{n}\sum_{i=1}^n (I\{\hat{\varepsilon}_{i,L} \leq 0\} - I\{\varepsilon_i \leq 0\}) I\{X_i \in \Theta_n\} I\{Z_i \leq z\} \\ &- E\Big[(I\{\hat{\varepsilon}_L \leq 0\} - I\{\varepsilon \leq 0\}) I\{X \in \Theta_n\} I\{Z \leq z\} \Big| (Y_i, X_i, Z_i), i = 1, \ldots, n\Big]\Big| = o_p(\frac{1}{\sqrt{n}}).\end{aligned}$$

and thus the proof is complete. □

Proof of Theorem 3.2. Starting from the stochastic expansion given in Lemma A.3 we obtain by Taylor's expansion

$$\begin{aligned}T_n(\Theta_n, z) &= \frac{1}{n}\sum_{i=1}^n (I\{\varepsilon_i \leq 0\} - \tau) I\{X_i \in \Theta_n\} I\{Z_i \leq z\} \\ &\quad + \int f_{\varepsilon|X,Z}(0|s,t)(\hat{q}_\tau(s) - q_\tau(s)) I\{s \in \Theta_n\} I\{t \leq z\} dF_{X,Z}(s,t) \\ &\quad + \int f'_{\varepsilon|X,Z}(\xi_{x,s,n}|s,t)(\hat{q}_\tau(s) - q_\tau(s))^2 I\{s \in \Theta_n\} I\{t \leq z\} dF_{X,Z}(s,t) + o_p(\frac{1}{\sqrt{n}})\end{aligned}$$

for some $\xi_{x,s,n}$ between 0 and $\hat{q}_\tau(s) - q_\tau(s)$ where the last line is of order $o_p(n^{-1/2})$ due to Lemma A.1 and the assumptions $\sup_{x \in \mathcal{D}, y \in \mathbb{R}, z \in R_Z} |f'_{\varepsilon|X,Z}(y|x,z)| < \infty$, $d_n^{2s} + \log n / nh_n^d = o(n^{-1/2})$. Note that

$$\int f_{\varepsilon|X,Z}(0|s,t)(\hat{q}_\tau(s)-q_\tau(s))I\{s\in\Theta_n\}I\{t\le z\}dF_{X,Z}(s,t)$$
$$= \int F_{Z|X,\varepsilon}(z|s,0)f_{\varepsilon|X}(0|s)f_X(s)(\hat{q}_\tau(s)-q_\tau(s))I\{s\in\Theta_n\}ds.$$

By Lemma A.2 we thus have

$$T_n(\Theta_n,z)=\frac{1}{n}\sum_{i=1}^n(I\{\varepsilon_i\le 0\}-\tau)I\{X_i\in\Theta_n\}\Big(I\{Z_i\le z\}-F_{Z|X,\varepsilon}(z|X_i,0)\Big)+o_p(\frac{1}{\sqrt{n}}).$$

This completes the proof. □

Proof of Corollary 3.3 and 3.5. Define the sequence of n-dependent classes of functions

$$\mathcal{F}_n:=\Big\{(e,\xi,\zeta)\mapsto eI\{\xi\in\Theta\cap\mathcal{D}_n\}(I\{\zeta\le z\}-F_{Z|X,\varepsilon}(z|\xi,0))\Bigm|\Theta\in\Xi,z\in R_Z\Big\}$$

and note that it is indexed by the totally bounded metric space $(\Xi\times R_Z,\rho)$ with metric

$$\rho((\Theta_1,y),(\Theta_2,z)):=(\mathbb{E}[(W_{\Theta_1,y}-W_{\Theta_2,z})^2])^{1/2}$$

where $W_{\Theta,z}:=(I\{\varepsilon_1\le 0\}-\tau)I\{X_1\in\Theta\}(I\{Z_1\le z\}-F_{Z|X,\varepsilon}(z|X_1,0))$. Moreover, it satisfies the assumptions of part 2 of Lemma B.3. A simple calculation in combination with the assumption $\sup_{\Theta\in\Xi}P(X_i\in\Theta\backslash\Theta_n)=o(1)$ shows that all the assumptions of Theorem 2.11.23 in van der Vaart and Wellner (1996) are satisfied. In particular, the covariances $\mathrm{Cov}(W_{\Theta_n,y},W_{\Theta_n',z})$ converge to $k(\Theta,y,\Theta',z)$ given in Corollary 3.3. This implies that the process

$$\sqrt{n}\Big(T_n(\Theta_n,z)-\tilde{T}_n(\Theta_n,z)\Big)$$
$$=\frac{1}{n}\sum_{i=1}^n\Big((I\{\varepsilon_i\le 0\}-\tau)I\{X_i\in\Theta_n\}(I\{Z_i\le z\}-F_{Z|X,\varepsilon}(z|X_i,0))-\tilde{T}_n(\Theta_n,z)\Big)+o_p(\frac{1}{\sqrt{n}}).$$

where $\tilde{T}_n(\Theta_n,z):=\mathbb{E}\Big[(I\{\varepsilon_i\le 0\}-\tau)I\{X_i\in\Theta_n\}(I\{Z_i\le z\}-F_{Z|X,\varepsilon}(z|X_i,0))\Big]$ converges weakly to the centered Gaussian process $T(\Theta_n,z)$ described in Corollary 3.3. Thus Corollary 3.3 and 3.5 follow after a straightforward calculation of the expectation $\tilde{T}_n(\Theta_n,z)$. Now the proof is complete. □

B Technical results

Before stating the main results of this section, we discuss some basic properties of the local polynomial estimator $\hat{F}_{Y|X}(y|x;p)$. To this end, we note that

$$\mathbf{X}^t\mathbf{W}\mathbf{Y}=(V_{n,0}(x,y),V_{n,\mathbf{k}_{1,1}}(x,y),...,V_{n,\mathbf{k}_{N_p,p}}(x,y))^t$$

with

$$V_{n,\mathbf{k}}(x,y):=\frac{h_n^{|\mathbf{k}|}}{nh_n^d}\sum_{i=1}^n\mathbf{K}_{h,\mathbf{k}}(x-X_i)\Omega\Big(\frac{y-Y_i}{d_n}\Big).$$

Lemma B.1 *Under the assumptions (K1), (K2), (K5), (A1), (A2) it holds that*

$$\begin{aligned}
&\hat{F}_{Y|X}(y|x;p) - F_{Y|X}(y|x)\\
&= e_1^t\Big(\sum_{j=0}^{n_f}\Big(-\frac{\mathcal{M}(K)^{-1}}{f_X(x)}\sum_{1\le|\mathbf{m}|<n_f} h_n^{|\mathbf{m}|} f_X^{(\mathbf{m})}(x) M_{\mathbf{m}}\Big)^j \frac{\mathcal{M}(K)^{-1}}{f_X(x)}\Big)(T_{n,0,S}(x,y),...,T_{n,\mathbf{k}_{N_p,p},S}(x,y))^t\\
&\quad + o_P(n^{-1/2})\\
&=: \Delta_S(y|x) + o_P(n^{-1/2}) = O_P(d_n^s + \Big(\frac{\log n}{nh_n^d}\Big)^{1/2})
\end{aligned}$$

uniformly with respect to $(x,y)\in\mathcal{D}_n\times\mathcal{Y}$*, where* $\mathcal{Y}$ *is any bounded subset of* $\mathbb{R}$ *and* $M_{\mathbf{k}}$ *denote some matrices with uniformly bounded entries that are independent of* x, n, y *and*

$$T_{n,\mathbf{k},S}(x,y) := \frac{1}{nh_n^d}\sum_i \mathbf{K}_{h_n,\mathbf{k}}(x-X_i)\Big(\Omega\Big(\frac{y-Y_i}{d_n}\Big) - F_{Y|X}(y|X_i)\Big).$$

Moreover, the quantity $\Delta_S(y|x)$ *is, with probability tending to one,* $d+1$ *times continuously differentiable with respect to* x *and* y *and all its partial derivatives of corresponding orders are uniformly bounded on* $\mathcal{D}_n\times\mathcal{Y}$*.*

Proof. At the end of the proof, we will establish the following two representations

$$\hat{F}_{Y|X}(y|x;p) = F_{Y|X}(y|x) + e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}(h_n^0 T_{n,0,S}(x,y),...,h_n^p T_{n,\mathbf{k}_{N_p,p},S}(x,y))^t + O_P(h_n^{p+1}), \tag{B.1}$$

$$(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1} = \mathbf{H}^{-1}\Big(\sum_{j=0}^{n_f}\Big(-\frac{\mathcal{M}(K)^{-1}}{f_X(x)}\sum_{1\le|\mathbf{l}|<n_f} h_n^{|\mathbf{l}|} M_{\mathbf{l}} f_X^{(\mathbf{l})}(x)\Big)^j\frac{\mathcal{M}(K)^{-1}}{f_X(x)} + 1_{N\times N}O_P(h_n^{n_f})\Big)\mathbf{H}^{-1}, \tag{B.2}$$

where $M_0, ..., M_{\mathbf{k}_{N_{n_f},n_f}}$ denote some matrices that do not depend on n, x, $M_0 = \mathcal{M}(K)$ is invertible, $\mathbf{H}$ is a diagonal matrix with entries $1, h_n, .., h_n, h_n^2, ..., h_n^2, ..., h_n^p, ..., h_n^p$ and the term $h_n^{|\mathbf{k}|}$ appears $N_{\mathbf{k}}$ times in this vector. By definition we have

$$\partial_y^r\partial_x^{\mathbf{m}} T_{n,\mathbf{k},S}(x,y) \;=\; \frac{1}{nh_n^{d+|\mathbf{m}|}}\sum_i \mathbf{K}_{h_n,\mathbf{k}}^{(\mathbf{m})}(x-X_i)\Big(\frac{1}{d_n^r}\omega^{(r-1)}\Big(\frac{y-Y_i}{d_n}\Big) - F_{Y|X}^{(r)}(y|X_i)\Big),$$

and tedious but straightforward calculations including integration-by parts and substitutions yield the estimates

$$\begin{aligned}
\sup_{(x,y)\in\mathcal{D}_n\times} \mathbb{E}[\partial_y^r\partial_x^{\mathbf{m}} T_{n,\mathbf{k},S}(x,y)] &\;=\; O(d_n^{s-r}),\\
\sup_{(x,y)\in\mathcal{D}_n} \mathbb{E}[(\partial_y^r\partial_x^{\mathbf{m}} T_{n,\mathbf{k},S}(x,y))^2] &\;=\; O\Big(\frac{1}{nh_n^{d+2|\mathbf{m}|}d_n^{0\vee(2r-1)}}\Big).
\end{aligned}$$

A combination of parts 1,2 and 6 of Lemma B.2 shows that, for every n, the class of functions

$$\mathcal{F}_n = \Big\{(u,v)\mapsto \mathbf{K}_{h_n,\mathbf{k}}^{(\mathbf{m})}(x-u)\Big(\frac{1}{d_n^r}\omega^{(r-1)}\Big(\frac{y-v}{d_n}\Big) - F_{Y|X}^{(r)}(y|u)\Big)\Big|x\in R_X, y\in\mathbb{R}\Big\}$$

satisfies the assumptions of part 2 of Lemma B.3 with constants not depending on n, which, together with the above estimates gives

$$\text{(B.3)} \qquad \sup_{(x,y)\in\mathcal{D}} |\partial_y^r \partial_x^{\mathbf{m}} T_{n,\mathbf{k},S}(x,y)| = O_P\Big(\frac{\log n}{n h_n^{d+2|\mathbf{m}|} d_n^{0\vee(2r-1)}}\Big)^{1/2} + O(d_n^{s-r}).$$

Combining (B.1), (B.2) and (B.3) yields

$$\begin{aligned} & e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}(h_n^0 T_{n,0,S}(x,y),...,h_n^p T_{n,\mathbf{k}_{N_p,p},S}(x,y))^t \\ = \; & e_1^t\Big(\sum_{j=0}^{n_f}\Big(\frac{\mathcal{M}(K)^{-1}}{f_X(x)}\sum_{1\le|\mathbf{l}|<n_f} h_n^{|\mathbf{l}|}M_{\mathbf{l}} f_X^{(\mathbf{l})}(x)\Big)^j \frac{\mathcal{M}(K)^{-1}}{f_X(x)}\Big)(T_{n,0,S}(x,y),...,T_{n,\mathbf{k}_{N_p,p},S}(x,y))^t + o_P(n^{-1/2}), \end{aligned}$$

and thus the proof of the first part of the Lemma is complete.

For a proof of the differentiability results, note that the $d+1-$fold differentiability of the product of every entry of a scalar product between two vectors follows from the $d+1-$fold differentiability of every entry of both vectors. This establishes that $\Delta_S(y|x)$ is $d+1$ times continuously differentiable with respect to both components and that all partial derivatives are uniformly bounded. By the results in (B.3) the proof is thus complete once we establish (B.1) and (B.2).

Proof of (B.1) A Taylor expansion of $F_{Y|X}(y|x)$ gives

$$\begin{aligned} & \frac{1}{nh_n^d}\sum_i \mathbf{K}_{h_n,\mathbf{k}}(x-X_i)F_{Y|X}(y|x) \\ = \; & \frac{1}{nh_n^d}\sum_{0\le|\mathbf{m}|\le p}\frac{\partial_x^{\mathbf{m}}F_{Y|X}(y|X_i)}{\mathbf{m}!}h_n^{|\mathbf{m}|}\sum_i \mathbf{K}_{h_n,\mathbf{k}+\mathbf{m}}(x-X_i) + O_P(h_n^{|\mathbf{m}|+p+1}). \end{aligned}$$

This fact, combined with

$$\frac{e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}}{nh_n^d}\begin{pmatrix} h_n^{|\mathbf{m}|}\sum_i K_{h_n,\mathbf{m}}(x-X_i) \\ \vdots \\ h_n^{p+|\mathbf{m}|}\sum_i K_{h_n,\mathbf{k}_{N_p,p}+\mathbf{m}}(x-X_i)\end{pmatrix} = I\{\mathbf{m}=0\},$$

yields the representation

$$\begin{aligned} F_{Y|X}(y|x) &= \frac{e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}}{nh_n^d}\begin{pmatrix} h_n^0\sum_i K_{h_n,0}(x-X_i)F_{Y|X}(y|x) \\ \vdots \\ h_n^p\sum_i K_{h_n,\mathbf{k}_{N_p,p}}(x-X_i)F_{Y|X}(y|x)\end{pmatrix} \\ &= \frac{e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}}{nh_n^d}\begin{pmatrix} h_n^0\sum_i K_{h_n,0}(x-X_i)F_{Y|X}(y|X_i) \\ \vdots \\ h_n^p\sum_i K_{h_n,\mathbf{k}_{N_p,p}}(x-X_i)F_{Y|X}(y|X_i)\end{pmatrix} + O_P(h_n^{p+1}) \end{aligned}$$

once we note that $\frac{1}{nh_n^d}\sum_i |K_{h_n,\mathbf{k}_{N_p,p}}(x-X_i)| = O_P(1)$ and $e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1} = (O_P(1),...,O_P(h_n^{-p}))$ [see the last part of the proof]. Thus

$$\hat{F}_{Y|X}(y|x) = F_{Y|X}(y|x) + e_1^t(\mathbf{X}^t\mathbf{W}\mathbf{X})^{-1}(h_n^0 T_{n,0,S}(x,y),...,h_n^p T_{n,\mathbf{k}_{N_p,p},S}(x,y))^t + O_P(h_n^{p+1}).$$

Proof of (B.2) The elements of the matrix $\mathbf{X}^t\mathbf{W}\mathbf{X}$ are of the form

$$(\mathbf{X}^t\mathbf{W}\mathbf{X})_{k,l} = \frac{1}{nh_n^d}\sum_i \mathbf{K}_{h_n,0}(x-X_i)(x-X_i)^{\mathbf{m}} = \frac{h_n^{|\mathbf{m}|}}{nh_n^d}\sum_i \mathbf{K}_{h_n,\mathbf{m}}(x-X_i)$$

where $\mathbf{m} = \mathbf{m}_1 + \mathbf{m}_2$ and $\mathbf{m}_1, \mathbf{m}_2$ denote the k'th and l'th entry in the tuple of vectors $(0, \mathbf{k}_{1,1},...,\mathbf{k}_{N_1,1},\mathbf{k}_{1,2},...,\mathbf{k}_{N_p,p})$, respectively. In particular, $d+1+n_f$-fold continuous differentiability of f_X implies that

$$\frac{1}{nh_n^d}\sum_i \mathbf{K}_{h_n,\mathbf{k}}(x-X_i) = \sum_{|\mathbf{l}|<n_f} \mu_{|\mathbf{k}|+|\mathbf{l}|}(\mathbf{K}) h_n^{|\mathbf{l}|} f_X^{(\mathbf{l})}(x) + O_P(\Big(\frac{\log n}{nh_n^d}\Big)^{1/2} + h_n^{n_f}).$$

Thus we obtain a representation of the form

$$\mathbf{X}^t\mathbf{W}\mathbf{X} = \mathbf{H}\Big(\sum_{|\mathbf{l}|<n_f} h_n^{|\mathbf{l}|} M_{\mathbf{l}} f_X^{(\mathbf{l})}(x) + 1_{N\times N} O_P(h_n^{n_f})\Big)\mathbf{H}$$

where $M_0,...,M_{\mathbf{k}_{N_{n_f},n_f}}$ denote some matrices that do not depend on n, x, $M_0 = \mathcal{M}(K)$ is invertible and $\mathbf{H}$ is a diagonal matrix with entries $1, h_n, .., h_n, h_n^2, ..., h_n^2, ..., h_n^p, ..., h_n^p$ where the term $h_n^{|\mathbf{k}|}$ appears $N_{|\mathbf{k}|}$ times in this vector [see the definition at the beginning of the section]. Thus for h_n sufficiently small an application of the Neumann series yields (B.2) with probability tending to one. □

Lemma B.2 ***Bounds on bracketing numbers***

1. *Define $\mathcal{F} + \mathcal{G} := \{f + g | f \in \mathcal{F}, g \in \mathcal{G}\}, \mathcal{F}\mathcal{G} := \{fg | f \in \mathcal{F}, g \in \mathcal{G}\}$. Then*

$$N_{[]}(\mathcal{F} + \mathcal{G}, \varepsilon, \rho) \le N_{[]}(\mathcal{F}, \varepsilon/2, \rho) N_{[]}(\mathcal{G}, \varepsilon/2, \rho)$$

   If additionally the classes $\mathcal{F}, \mathcal{G}$ are uniformly bounded by the constant C, we have

$$N_{[]}(\mathcal{F}\mathcal{G}, \varepsilon, \|.\|) \le N_{[]}^2(\mathcal{F}, \varepsilon/4C, \|.\|) N_{[]}^2(\mathcal{G}, \varepsilon/4C, \|.\|)$$

   for any seminorm $\|.\|$ with the additional property that $|f_2| \le |f_2|$ implies $\|f_1\| \le \|f_2\|$.

2. *Let $\mathcal{F}_n$ denote a class of functions f_x indexed by the bounded interval $x \in [-A, A]$ which are bounded by a given constant and have support of the form $[x-h, x+h]$. If $\sup_{f\in\mathcal{F}} |f(a) - f(b)| \le C|a-b|h^{-k}$ for some universal constant C we have $N_{[]}(\mathcal{F}_n, \varepsilon, L^2(P_X)) \le \gamma\varepsilon^{-(2k+1)}$ provided that P_X has a uniformly bounded density. Here γ denotes a constant which does not depend on n.*

3. *Consider the class of functions*

$$\mathcal{F}_n := \Big\{(x,y) \mapsto \Omega\Big(\frac{g(x)-y}{d_n}\Big)\Big| g \in \mathcal{G}\Big\},$$

where Ω is Lipschitz-continuous and there exist constants C_1, C_2 such that Ω is constant on $(-\infty, C_1]$ and $[C_2, \infty)$. Assume additionally that the distribution of (X,Y) has a uniformly bounded density, then

$$N_{[]}(\mathcal{F}_n, \varepsilon, L^2(P_{XY})) \leq C_5 N_{[]}(\mathcal{G}, C_6\varepsilon^2, \|\cdot\|_\infty)$$

for some constants C_5, C_6 independent of n.

4. *For any measure $P^{U,V}$ on the unit interval with uniformly bounded density f, the class of functions*

$$\mathcal{F} := \{u \mapsto I\{u \leq s\}|s \in [0,1]\} \cup \{u \mapsto I\{u < s\}|s \in [0,1]\}$$

can be covered by $C\varepsilon^{-2}$ brackets of $L^2(P)$ length ε.

5. *For any measure P on $\mathbb{R} \times \mathbb{R}^k$ with uniformly bounded conditional density $f_{V|U}$ the class of functions*

$$\mathcal{G} := \{(u,v) \mapsto I\{v \leq f(u)\}|f \in \mathcal{F}\}$$

satisfies $N_{[]}(\mathcal{G}, \varepsilon, \|.\|_{P,2}) \leq N_{[]}(\mathcal{F}, C\varepsilon^2, \|.\|_\infty)$ for some constant C independent of ε.

6. *Assume that $f(x;a)$ is a function indexed by the parameter $a \in A$ such that $\sup_x \|f(s;x) - f(t;x)\|_\infty \leq C\|s-t\|^\theta$ for some $\theta > 0$ and norm $\|\cdot\|$. Then the $\|\cdot\|_\infty$-bracketing numbers of the class of functions $\mathcal{F} = \{u \mapsto f(u;a)|a \in A\}$ satisfy $N_{[]}(\mathcal{F}, \varepsilon, \|\|_\infty) \leq C_1 N(A, C_2\varepsilon^{1/\theta}, \|\cdot\|)$ for some finite constants C_1, C_2.*

Proof.

Part 1 The first assertion is obvious from the definition of bracketing numbers. For the second assertion, note that $\mathcal{F}\mathcal{G} = (\mathcal{F} + C)(\mathcal{G} + C) - C\mathcal{F} - C\mathcal{G} + C^2$. Moreover, all elements of the classes $\mathcal{F} + C, \mathcal{G} + C$ are by construction non-negative and thus it also is possible to cover them with brackets consisting of non-negative functions and amounts equal to the brackets of $\mathcal{F}, \mathcal{G}$, respectively. Finally, observe that if $0 \leq f_l \leq f \leq f_u$ and $0 \leq g_l \leq g \leq g_u$, we also have $f_l g_l \leq fg \leq f_u g_u$. Moreover $\|f_l g_l - f_u g_u\| \leq C\|f_u - f_l\| + C\|g_u - g_l\|$. Thus the class $(\mathcal{F}+C)(\mathcal{G}+C)$ can be covered by at most $\leq N_{[]}(\mathcal{F}, \varepsilon, \|.\|) N_{[]}(\mathcal{G}, \varepsilon, \|.\|)$ brackets of length $2C\varepsilon$. Finding brackets for the classes $C\mathcal{F}, C\mathcal{G}$ is trivial, and applying the first assertion of the Lemma completes the proof.

Part 2 Consider two cases.

A) $\varepsilon > 4h^{1/2}$: Divide $[0,1]$ into $N := 2/\varepsilon^2$ subintervals of length $2\alpha := \varepsilon^2$ with centers $r\alpha$ for $r = 1, ..., N$ and call the intervals $I_1, ..., I_N$. Note that two adjunct intervals overlap by $\alpha > 2h$. This construction ensures that every set of the form $[x-h, x+h]$ with $x \in [h, 1-h]$ is completely contained in at least one of the intervals defined above. Then a collection of N brackets of L^2-length $D\varepsilon$ for some $D > 0$ independent of h is given by $(-CI\{u \in I_j\}, CI\{u \in I_j\})$.

B) $\varepsilon \leq 4h^{1/2}$: Observe that by assumption any element g of $\mathcal{F}$ satisfies $|g(x) - g(y)| \leq C|x-y|h^{-k}$. Consider the points $t_i := i/(N+1), i = 1, ..., N$ with $N := 2^{2k+1}C/\varepsilon^{2k+1}$. By construction, to every $x \in [h, 1-h]$ there exists $i(x)$ with $|t_{i(x)} - x| \leq \varepsilon^{2k+1}/(2^{2k+1}C)$. This implies

$$|g(x) - g(t_{i(x)})| \leq C\varepsilon^{2k+1}h^{-k}/2^{2k+1}C \leq \varepsilon/2$$

Then N $\|.\|_\infty-$brackets of length covering $\mathcal{F}$ are given by $(g(t_i) - \varepsilon/2, g(t_i) + \varepsilon/2)$, $i = 1, ..., N$. From those one can easily construct $L^2(P_X)$-brackets.

Part 3 Without loss of generality, assume that Ω equals one on $[1, \infty)$ and zero on $(-\infty, -1]$. Moreover, the assumptions on Ω imply the existence of finite constant C_l, C_u such that $C_l \leq \Omega \leq C_u$. Distinguish two cases A) $\varepsilon \leq d_n$: Starting with ε^2 supremum norm brackets for the class $\mathcal{G}$ and using the Lipschitz condition yields the desired brackets. B) $\varepsilon > d_n$: Denote by $[g_{1,l}, g_{1,u}], ..., [g_{N(\varepsilon),l}, g_{N(\varepsilon),u}]$ brackets for the class $\mathcal{G}$ of $\|\cdot\|_\infty$-size ε. For a function $g \in \mathcal{G}$, denote the bracket that contains it by $[g_{j(g),l}, g_{j(g),u}]$. Observe that

$$\Omega\Big(\frac{g(x) - y}{d_n}\Big) \begin{cases} = 0, & if \;\; y > g_{j(g),u}(x) + d_n \\ = 1, & if \;\; y < g_{j(g),l}(x) - d_n \\ \in [C_l, C_u] & else \end{cases}$$

Thus brackets of the form

$$\begin{aligned} b_{l,j}(x) &:= I\{y < g_{j,l}(x) - d_n\} + C_l I\{g_{j,l}(x) - d_n \leq y \leq g_{j,u}(x) + d_n\} \\ b_{u,j}(x) &:= I\{y < g_{j,l}(x) - d_n\} + C_u I\{g_{j,l}(x) - d_n \leq y \leq g_{j,u}(x) + d_n\} \end{aligned}$$

contain every function in $\mathcal{F}_n$. Moreover, the L^2-length of each such bracket is bounded by $(C_u - C_l)(2d_n + \varepsilon) \sup f_{X,Y}(x, y) \leq C\varepsilon$. This completes the proof.

Part 4 Follows by standard arguments.

Part 5 Follows from $|I\{v \leq g_1(u)\} - I\{v \leq g_2(u)\}| \leq I\{|v - g_1(u)| \leq 2\|g_1 - g_2\|_\infty\}$.

Part 6 Obvious □

Lemma B.3 (Basic Lemma) *Assume that the classes of functions $\mathcal{F}_n$ consist of uniformly bounded functions (by a constant not depending on n).*

1. *If for some $a < 2$ $N_{[]}(\mathcal{F}_n, \varepsilon, L^2(P)) \leq C\exp(-c\varepsilon^{-a})$ for every $\varepsilon \leq \delta_n$ with constants C, c not depending on n, then we have*

$$\sqrt{n} \sup_{f \in \mathcal{F}_n, \|f\|_{P,2} \leq \delta_n} \Big(\int f dP_n - \int f dP\Big) = o_P^*(1),$$

where the $*$ denotes outer probability, see van der Vaart and Wellner (1996) for a more detailed discussion.

2. *If $N_{[]}(\mathcal{F}_n, \varepsilon, L^2(P)) \leq C\varepsilon^{-a}$ for every $\varepsilon \leq \delta_n$, some $a > 0$ and a constant C not depending on n, then we have for any $\delta_n \sim n^{-b}$ with $b < 1/2$*

$$\sqrt{n} \sup_{f \in \mathcal{F}_n, \|f\|_{P,2} \leq \delta_n} \Big(\int f dP_n - \int f dP\Big) = O_P^*\Big(\delta_n |\log \delta_n|\Big).$$

Proof. Start by observing that the uniform boundedness of elements of $\mathcal{F}_n$ by D implies that $F \equiv D$ is a measurable envelope function with L_2-norm D. The proof of the first part follows by arguments similar to those used for the proof of the second part and is therefore omitted. For the proof of the second part, note that for η_n sufficiently small

$$\begin{aligned} a(\eta_n) &:= \eta_n D/\sqrt{1+\log N_{[]}(\eta_n D, \mathcal{F}_n, L_2(P))} \geq D\eta_n/\sqrt{1+\log C - a\log(D\eta_n)} \\ &\geq D\tilde{C}\eta_n/\sqrt{|\log \eta_n|} \end{aligned}$$

for some finite constant $\tilde{C}$ depending only on a, C, D. Thus the bound in Theorem 2.14.2 in van der Vaart, Wellner (1996) yields for δ_n sufficiently small

$$\begin{aligned} \mathbb{E}\Big[\sup_{f\in\mathcal{F}_n}\int f d\alpha_n\Big]^* &\leq DJ_{[]}(\delta_n, \mathcal{F}_n, L_2(P)) + \sqrt{n}\int F(u)I\{F(u) > \sqrt{n}a(\delta_n)\}P(du) \\ &\leq DC_1\int_0^{\delta_n}|\log\varepsilon|d\varepsilon + D\sqrt{n}I\Big\{D > \frac{D\tilde{C}\sqrt{n}\delta_n}{|\log\delta_n|}\Big\} \\ &\leq DC_2\delta_n|\log\delta_n| + D\sqrt{n}I\Big\{1 > \frac{\tilde{C}\sqrt{n}\delta_n}{|\log\delta_n|}\Big\}. \end{aligned}$$

where $\alpha_n := \sqrt{n}(P_n - P)$, P_n denotes the empirical measure, and C_1, C_2 are some finite constants. Here, the second inequality follows by a straightforward calculation and the first inequality is due to the fact that for δ_n sufficiently small by definition

$$J_{[]}(\delta_n, \mathcal{F}_n, L_2(P)) = \int_0^{\delta_n}\sqrt{1+\log N_{[]}(\varepsilon D, \mathcal{F}_n, L_2(P))}d\varepsilon \leq C_1\int_0^{\delta_n}|\log\varepsilon|d\varepsilon.$$

Now under the assumption on δ_n, the indicator in the last line will be zero for n large enough and thus the proof is complete. □

Lemma B.4 *Assume that κ is a symmetric, uniformly bounded density with support $[-1,1]$ and let $b_n = o(1)$. Introduce the notation $Q_{G,\kappa,\tau,b_n}(F) := G^{-1}(H_{G,\kappa,\tau,b_n}(F))$.*
(a) If the function $F : [0,1] \to \mathbb{R}$ is strictly increasing and F^{-1} is k times continuously differentiable in a neighborhood of the point τ, we have

$$H_{id,\kappa,\tau,b_n}(F) = F^{-1}(\tau) + \sum_{i=1}^{k}\frac{b_n^i}{i!}(F^{-1})^{(i)}(\tau)\mu_{i+1}(\kappa) + R_n(\tau)$$

with $|R_n(\tau)| \leq C_k(\kappa)b_n^k\sup_{|s-\tau|\leq b_n}|(F^{-1})^{(k)}(\tau) - (F^{-1})^{(k)}(s)|$, $\mu_i(\kappa) := \int u^i\kappa(u)du$ and a constant C_k depending only on k and κ. In particular, if $F : \mathbb{R} \to [0,1]$ is strictly increasing and F^{-1} is two times continuously differentiable in a neighborhood of τ and $G : [0,1] \to \mathbb{R}$ is two times continuously differentiable in a neighborhood of $F^{-1}(\tau)$ with $G'(F^{-1}(\tau)) > 0$, we have

$$|F^{-1}(\tau) - Q_{G,\kappa,\tau,b_n}(F)| \leq R_{n,2} := Cb_n^2\sup_{|s-G\circ F^{-1}(\tau)|\leq R_{n,1}}|(G^{-1})'(s)|\sup_{|s-\tau|\leq b_n}|(G\circ F^{-1})''(s)|$$

for some constant C that depends only on κ where $R_{n,1} := Cb_n^2 \sup_{|s-\tau|\le b_n} |(G\circ F^{-1})''(s)|$.

(b) Assume that κ is additionally differentiable with Lipschitz-continuous derivative and that the functions G, G^{-1} have derivatives that are uniformly bounded on any compact subset of $\mathbb{R}$ [the bound is allowed to depend on the interval]. Then for any increasing function F with uniformly bounded first derivative we have $|H(F_1) - H(F_2)| \le R_{n,3} + R_{n,4}$ and

$$|Q_{G,\kappa,\tau,b_n}(F_1) - Q_{G,\kappa,\tau,b_n}(F_2)| \le \sup_{u\in\mathcal{U}(H(F_1),H(F_2))} |(G^{-1})'(u)|(R_{n,3}+R_{n,4}),$$

where the constant C depends only on κ, $\mathcal{U}(a,b) := [a\wedge b, a\vee b]$, and

$$R_{n,3} := \frac{Cc_n}{b_n}\|F_1 - F_2\|_\infty \sup_{|v-\tau|\le c_n} |(G\circ F^{-1})'(v)|, \quad R_{n,4} := R_{n,3}\frac{\|F_1 - F\|_\infty + \|F_1 - F_2\|_\infty}{b_n}$$

with $c_n := b_n + 2\|F_1 - F_2\|_\infty + \|F_1 - F\|_\infty$.

(c) If additionally to the assumptions made in (b), the function F_1 is two times continuously differentiable in a neighborhood of $F^{-1}(\tau)$ with $F_1'(F_1^{-1}(\tau)) > 0$ and G is two times continuously differentiable in a neighborhood of $F_1^{-1}(\tau)$ with $G'(F^{-1}(\tau)) > 0$, we have

$$\begin{aligned} Q_{G,\kappa,\tau,b_n}(F_1) - Q_{G,\kappa,\tau,b_n}(F_2) \;=\; & -\frac{1}{F_1'(F_1^{-1}(\tau))}\int_{-1}^{1}\kappa(v)\Big(F_2(F_1^{-1}(\tau+vb_n)) - F_1(F_1^{-1}(\tau+vb_n))\Big)dv \\ & +R_n, \end{aligned}$$

where

$$|R_n| \;\le\; R_{n,5} + R_{n,6} + \frac{Cb_n \sup_{|s-\tau|\le b_n}(G\circ F^{-1})''(s)\|F_1 - F_2\|_\infty + R_{n,4}}{G'(F_1^{-1}(\tau))}$$

with a constant C depending only on κ and

$$\begin{aligned} R_{n,5} \;&:=\; \frac{1}{2}\sup_{u\in\mathcal{U}(H(F_1),H(F_2))} |(G^{-1})''(u)|(H(F_1)-H(F_2))^2 \\ R_{n,6} \;&:=\; \sup_{u\in\mathcal{U}(H(F_1),G(F_1^{-1})(\tau))} |(G^{-1})''(u)|\cdot|H(F_1) - G(F_1^{-1})(\tau)|\cdot|H(F_1)-H(F_2)|. \end{aligned}$$

Proof. The proof of the first part of (a) is essentially a Taylor expansion. Details can be found in the proof of Lemma A.4 in Volgushev (2006). For a proof of the second part of (a), observe that by definition $H_{G,\kappa,\tau,b_n}(F) = H_{id,\kappa,\tau,b_n}(F\circ G^{-1})$. Together with the first part we obtain

$$|H_{id,\kappa,\tau,b_n}(F\circ G^{-1}) - G\circ F^{-1}(\tau)| \;\le\; Cb_n^2 \sup_{|s-\tau|\le b_n} |(G\circ F^{-1})''(s)| =: R_{n,1}$$

which yields

$$\begin{aligned} |G^{-1}(H_{G,\kappa,\tau,b_n}(F)) - F^{-1}(\tau)| \;&\le\; |(G^{-1})'(\xi)|\cdot|H_{id,\kappa,\tau,b_n}(F\circ G^{-1}) - G(F^{-1}(\tau))| \\ &\le\; Cb_n^2 \sup_{|s-G\circ F^{-1}(\tau)|\le R_{n,1}} |(G^{-1})'(s)| \sup_{|s-\tau|\le b_n} |(G\circ F^{-1})''(s)| =: R_{n,2}. \end{aligned}$$

The proof of (a) is thus complete.

¿¿¿¿From now on, drop the index of H for the sake of a simpler notation. For a proof of (b), observe the decomposition

$$\begin{aligned} H(F_1) - H(F_2) &= -\frac{1}{b_n}\int_0^1 \kappa\Big(\frac{F_1(G^{-1}(u)) - \tau}{b_n}\Big)(F_1(G^{-1}(u)) - F_2(G^{-1}(u)))du \\ &\quad - \frac{1}{b_n}\int_0^1 \Big[\kappa\Big(\frac{\xi(u) - \tau}{b_n}\Big) - \kappa\Big(\frac{F_1(G^{-1}(u)) - \tau}{b_n}\Big)\Big] \\ &\quad \times (F_1(G^{-1}(u)) - F_2(G^{-1}(u)))du \end{aligned}$$

for some $|\xi(u) - F_2(G^{-1}(u))| \le |F_1(G^{-1}(u)) - F_2(G^{-1}(u))|$. This yields the bound

$$\begin{aligned} |H(F_1) - H(F_2)| \le{} & \frac{1}{b_n}\int_0^1 \kappa\Big(\frac{F_1(G^{-1}(u)) - \tau}{b_n}\Big) + \Big|\kappa\Big(\frac{\xi(u) - \tau}{b_n}\Big) - \kappa\Big(\frac{F_1(G^{-1}(u)) - \tau}{b_n}\Big)\Big|du \\ & \times\|F_1 - F_2\|_\infty \end{aligned}$$

Next, observe that by assumption κ is Lipschitz continuous and thus we have the inequality

$$\begin{aligned} & \Big|\kappa\Big(\frac{\xi(u) - \tau}{b_n}\Big) - \kappa\Big(\frac{F_1(G^{-1}(u)) - \tau}{b_n}\Big)\Big| \\ \le{} & \frac{L|\xi(u) - F_1(G^{-1}(u))|}{b_n}\big(I\{|F_1(G^{-1}(u)) - \tau| \le b_n\} + I\{|\xi(u) - \tau| \le b_n\}\big) \\ \le{} & \frac{2L\|F_1 - F_2\|_\infty}{b_n} I\{|F_1(G^{-1}(u)) - \tau| \le b_n + 2\|F_1 - F_2\|_\infty\} \\ \le{} & \frac{2L\|F_1 - F_2\|_\infty}{b_n} I\{|F(G^{-1}(u)) - \tau| \le b_n + 2\|F_1 - F_2\|_\infty + \|F_1 - F\|_\infty\}. \end{aligned}$$

Similarly

$$\begin{aligned} & \Big|\kappa\Big(\frac{F_1(G^{-1}(u)) - \tau}{b_n}\Big) - \kappa\Big(\frac{F(G^{-1}(u)) - \tau}{b_n}\Big)\Big| \\ \le{} & \frac{2L\|F_1 - F\|_\infty}{b_n} I\{|F(G^{-1}(u)) - \tau| \le b_n + \|F_1 - F\|_\infty\}, \end{aligned}$$

and moreover

$$\Big|\kappa\Big(\frac{F(G^{-1}(u)) - \tau}{b_n}\Big)\Big| \le CI\{|F(G^{-1}(u)) - \tau| \le b_n\}.$$

Define $c_n := b_n + 2\|F_1 - F_2\|_\infty + \|F_1 - F\|_\infty$. Note that the monotonicity of F, G implies

$$\{u : |F(G^{-1}(u)) - \tau| \le c_n\} \subseteq [G(F^{-1}(\tau - c_n)), G(F^{-1}(\tau + c_n))]$$

and

$$|G(F^{-1}(\tau + c_n)) - G(F^{-1}(\tau - c_n))| \le 2c_n \sup_{|v-\tau|\le c_n} |(G \circ F^{-1})'(v)|.$$

In particular, this implies the estimate

$$\int_0^1 I\{|F(G^{-1}(u)) - \tau| \leq c_n\} du \leq 2c_n \sup_{|v-\tau|\leq c_n} |(G \circ F^{-1})'(v)|.$$

Summarizing, we have obtained the bound $|H(F_1) - H(F_2)| \leq R_{n,3} + R_{n,4}$ where C denotes some constant depending only on the kernel κ. Assertion (b) follows from this estimate and a Taylor expansion of G^{-1}.

For a proof of assertion (c), note that after a substitution

$$\begin{aligned}
&\frac{1}{b_n}\int_0^1 \kappa\Big(\frac{F_1(G^{-1}(u)) - \tau}{b_n}\Big)(F_1(G^{-1}(u)) - F_2(G^{-1}(u)))du \\
&= \int_{-1}^1 (G \circ F_1^{-1})'(\tau + vb_n)\kappa(v)\Big(F_2(F_1^{-1}(\tau + vb_n)) - F_1(F_1^{-1}(\tau + vb_n))\Big)dv \\
&= (G \circ F_1^{-1})'(\tau)\int_{-1}^1 \kappa(v)\Big(F_2(F_1^{-1}(\tau + vb_n)) - F_1(F_1^{-1}(\tau + vb_n))\Big)dv + r_n
\end{aligned}$$

where

$$|r_n| \leq Cb_n \sup_{|s-\tau|\leq b_n} |(G \circ F_1^{-1})''(s)| \cdot \|F_1 - F_2\|_\infty$$

by a Taylor expansion of $(G \circ F_1^{-1})'$. A Taylor expansion of G^{-1} yields

$$\begin{aligned}
&\Big\|G^{-1}(H(F_1)) - G^{-1}(H(F_2)) - (G^{-1})'(H(F_1))(H(F_1) - H(F_2))\Big\| \\
&\leq \frac{1}{2}\sup_{u\in\mathcal{U}(H(F_1),H(F_2))} |(G^{-1})''(u)|(H(F_1) - H(F_2))^2
\end{aligned}$$

where $\mathcal{U}(a,b) := [a \wedge b, a \vee b]$. A Taylor expansion yields

$$\Big|(G^{-1})'(H(F_1)) - (G^{-1})'(G(F_1^{-1})(\tau))\Big| \leq \sup_{u\in\mathcal{U}(H(F_1),G(F_1^{-1})(\tau))} |(G^{-1})''(u)| \cdot |H(F_1) - G(F_1^{-1})(\tau)|$$

and combining this with the results obtained so far we arrive at

$$\begin{aligned}
&\Big|Q(F_1) - Q(F_2) + \frac{1}{F_1'(F_1^{-1}(\tau))}\int_{-1}^1 \kappa(v)\Big(F_2(F_1^{-1}(\tau + vb_n)) - F_1(F_1^{-1}(\tau + vb_n))\Big)dv\Big| \\
&\leq \Big\|G^{-1}(H(F_1)) - G^{-1}(H(F_2)) - (G^{-1})'(H(F_1))(H(F_1) - H(F_2))\Big\| \\
&\quad + |H(F_1) - H(F_2)| \cdot |(G^{-1})'(H(F_1)) - (G^{-1})'(G \circ F_1^{-1}(\tau))| \\
&\quad + \Big|\frac{H(F_1) - H(F_2)}{G'(F_1^{-1}(\tau))} + \frac{1}{F_1'(F_1^{-1}(\tau))}\int_{-1}^1 \kappa(v)\Big(F_2(F_1^{-1}(\tau + vb_n)) - F_1(F_1^{-1}(\tau + vb_n))\Big)dv\Big| \\
&\leq R_{n,5} + R_{n,6} + \frac{Cb_n \sup_{|s-\tau|\leq b_n}(G \circ F^{-1})''(s)\|F_1 - F_2\|_\infty + R_{n,4}}{G'(F_1^{-1}(\tau))}.
\end{aligned}$$

This completes the proof. □