Ann Inst Stat Math (2015) 67:437–477
DOI 10.1007/s10463-014-0461-1

Testing for additivity in nonparametric quantile regression

Holger Dette · Matthias Guhlich · Natalie Neumeyer

Received: 13 February 2012 / Revised: 6 November 2013 / Published online: 26 April 2014

Abstract In this article, we propose a new test for additivity in nonparametric quantile regression with a high-dimensional predictor. Asymptotic normality of the corresponding test statistic (after appropriate standardization) is established under the null hypothesis, local and fixed alternatives. We also propose a bootstrap procedure which can be used to improve the approximation of the nominal level for moderate sample sizes. The methodology is also illustrated by means of a small simulation study, and a data example is analyzed.

Keywords Nonparametric regression · Quantile regression · Bootstrap · Additive estimation

1 Introduction

Quantile regression was introduced by Koenker and Bassett (1978) as a complement to least squares estimation (LSE) or maximum likelihood estimation (MLE) and leads to far-reaching extensions of "classical" regression analysis by estimating families of conditional quantile surfaces, which describe the relation between a one-dimensional response y and a high-dimensional predictor x. Since its introduction, it has found

H. Dette (✉) · M. Guhlich
Fakultät für Mathematik, Ruhr-Universität Bochum, 44780 Bochum, Germany
e-mail: holger.dette@rub.de

M. Guhlich
e-mail: matthias.guhlich@rub.de

N. Neumeyer
Fachbereich Mathematik, Universität Hamburg, 20146 Hamburg, Germany
e-mail: neumeyer@math.uni-hamburg.de

great attraction in mathematical and applied statistics because of its ease of interpretation and robustness, which yields attractive applications in such important areas as medicine, economics, engineering and environmental modeling. The interested reader is referred to the recent monograph of Koenker (2005). Many authors consider parametric quantile regression models but in the last two decades nonparametric methods for estimating conditional quantiles have also been discussed intensively. Most of the literature refers to models with a univariate predictor [see e.g., Yu and Jones (1997, 1998), Dette and Volgushev (2008) and Chernozhukov et al. (2010)]. While from a theoretical point of view, there is no difficulty to generalize this methodology to high-dimensional covariates, 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 reasonable sample sizes. A common approach in nonparametric statistics to deal with this problem is to postulate an additive nonparametric model, which allows the estimation of the regression with one-dimensional rates. In classical regression (estimating the conditional expectation of the response given in the predictor), this methodology has found considerable interest in the literature [see Linton and Nielsen (1995), Mammen et al. (1999), Carroll et al. (2002), Hengartner and Sperlich (2005), Nielsen and Sperlich (2005), among others]. In quantile regression, nonparametric models of this type have only been discussed more recently. Doksum and Koo (2000) suggest a spline estimate and Gooijer and Zerom (2003) introduce a marginal integration estimate of an additive quantile regression model. Horowitz and Lee (2005) propose a two-step procedure, which fits a parametric model in the first step (with increasing dimension) for each coordinate and smooth it in a second step by the local polynomial technique. Yu and Lu (2004) and Lee and Mammen (2010) suggest backfitting methods for additive quantile regression estimation, while Dette and Scheder (2011) combine marginal integration techniques with monotone rearrangements [see Dette et al. (2006)] for the construction of additive estimates. Although these methods estimate the unknown quantile regression with the optimal (one-dimensional) rate if the assumption of an additive model is correct, they are generally inconsistent if the quantile regression is not additive. In this case the corresponding statistics usually estimate a "best approximation" of the unknown regression by an additive quantile regression model, but the difference between the "true" curve and its best approximation can be substantial. For this reason, it is of some importance to investigate by a statistical test if the hypothesis of an additive quantile regression is satisfied. In the context of modeling the conditional expectation, this problem has found considerable interest in the literature [see for example Eubank et al. (1995), Gozalo and Linton (2001), Dette and von Lieres und Wilkau (2001), Derbort et al. (2002) or Abramovich et al. (2009), among others]. On the other hand, to the best knowledge of the authors, tests for the hypothesis of an additive quantile regression model have not been considered so far in the literature, and the purpose of the present paper is to propose and analyze such a procedure for this problem. In Sect. 2 we introduce the basic notation and an additive estimate of the conditional quantile curve. The test statistic for the problem of additive quantile regression uses the residuals from this additive fit and is introduced in Sect. 3, where we also study the main asymptotic properties. In particular, we prove weak convergence of an appropriately standardized version of the test statistic under the null hypothesis and fixed alternatives with different rates corresponding to both cases. In Sect. 4 we present a small simulation study to illustrate the finite sample properties of a bootstrap version of the proposed test. We also investigate a data example testing if the hypothesis of an additive quantile regression is satisfied. Finally, all proofs and some of the more technical details in the proofs are deferred to Appendices A, B and C.

2 Preliminaries: an additive estimator

Consider a sequence of independent, identically distributed observations $(X_1, Y_1), \dots, (X_n, Y_n)$ where $X_j = (X_{j1}, \dots, X_{jd})^T$ denotes a d—dimensional random variable with density f and f_i is the marginal density of the ith component X_{ji} of X_j $(i = 1, \dots, d)$. Throughout this paper we denote by $F(y|x)$ the conditional distribution function of Y_1 given $X_1 = x = (x_1, \dots, x_d)^T$ and by $Q(\tau|x) = F^{-1}(y|x)$ the corresponding conditional quantile function. In the following, we fix some quantile $\tau \in (0, 1)$ and are interested in the problem of testing the hypothesis of additivity

$$H_0 : Q(\tau|x) = Q(\tau|x_1, \dots, x_d) = \sum_{k=1}^{d} Q_k(\tau|x_k) + c(\tau) \tag{1}$$

for some constant $c(\tau)$ and functions $Q_k(\tau|x_k)$ $(k = 1, \dots, d)$. Note that the quantities in (1) are not uniquely determined and to make these identifiable we assume throughout this paper the conditions

$$E[Q_k(\tau|X_{jk})] = 0, \quad k = 1, \dots, d, \; j = 1, \dots, n.$$

For the construction of a test for the hypothesis (1) let $\hat{Q}_{\text{add}}$ denote an additive estimate of the quantile regression function Q (for fixed τ), which will be specified later. We propose the statistic

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} L_g\left(X_i - X_j\right) \widehat{R}_i \widehat{R}_j \pi(X_i)\pi(X_j). \tag{2}$$

Here the random variables $\widehat{R}_i$ are defined by

$$\widehat{R}_i = I\left\{Y_i \leq \widehat{Q}_{\text{add}}^{-i}(\tau|X_i)\right\} - \tau, \tag{3}$$

π is a positive weight function and the function L_g is given by

$$L_g(X_i - X_j) = \frac{1}{g^d} L\left(\frac{X_i - X_j}{g}\right), \tag{4}$$

where L denotes a d-dimensional kernel function and g is a bandwidth (note that one might use different bandwidths for each covariable, which is not reflected in our notation). Throughout this paper we use the notation $\widehat{a}$ and $\widehat{a}^{-i}$ corresponding to estimates from the full sample $\{(X_j, Y_j) | j = 1, \ldots, n\}$ and the sample without the ith observation, respectively. Thus the statistic $\widehat{Q}_{\text{add}}^{-i}(\tau|x)$ in (3) denotes the additive (nonparametric) estimate of the quantile regression from the sample without the ith observation. Similarly, $\widehat{Q}_{\text{add}}^{-i,j}$ and $\widehat{Q}_{\text{add}}^{-i,j,k}$ denote the corresponding estimators without the ith and jth and the ith, jth and kth observation, respectively. Various additive quantile regression estimates have been proposed by Gooijer and Zerom (2003),Yu and Lu (2004), Horowitz and Lee (2005), Lee and Mammen (2010) and Dette and Scheder (2011).

Note that statistics of the type (4) have been introduced by Zheng (1996) in the context of testing for a specific parametric form in nonparametric regression, and since their introduction have found considerable interest in the context of goodness-of-fit tests [see Dette and von Lieres und Wilkau (2001) or Zhang and Dette (2004) among others]. An important advantage of the statistic T_n compared to other methods is that its normalized version is asymptotically unbiased [see Dette and von Lieres und Wilkau (2001)]. In the following section, we will study the asymptotic properties of the test statistic under the null hypothesis of additivity, local alternatives and fixed alternatives. In particular, we prove weak convergence of a standardized version of the statistic T_n defined in (2) with different rates corresponding to the null hypothesis and fixed alternatives. For this discussion which is deferred to Sect. 3 we therefore recall the definition of an additive quantile regression estimate which has recently been introduced by Dette and Scheder (2011) and will be used throughout this paper for a test of an additive quantile regression. Following Dette and Scheder (2011) we denote by

$$\widehat{F}_l(y|x) = \frac{\sum_{i=1}^n K_{1,h_1}(x_l - X_{il}) K_{2,H}(x_{\underline{l}} - X_{i\underline{l}}) I\{Y_i \leq y\}}{\sum_{i=1}^n K_{1,h_1}(x_l - X_{il}) K_{2,H}(x_{\underline{l}} - X_{i\underline{l}})} \tag{5}$$

the Nadaraya Watson estimate of the conditional distribution function where for $l = 1, \ldots, d$, $x_{\underline{l}} \in \mathbb{R}^{d-1}$ denotes the vector containing the components $x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_d$ of the vector $x = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$. In (5) the functions K_1 and K_2 are one-dimensional and $(d-1)$-dimensional kernels, respectively, h_1 is a one-dimensional bandwidth and $H = \text{diag}(h_2, \ldots, h_d)$ a $(d-1)$-dimensional non-singular and diagonal (bandwidth) matrix and we use the notation

$$K_{1,h_1}(x_1) = \frac{1}{h_1} K_1(h_1^{-1} x_1),$$
$$K_{2,H}(\tilde{x}) = \frac{1}{\det(H)} K_2\big(H^{-1}\tilde{x}\big).$$

We emphasize that the statistics $\widehat{F}_l$ differ for different values of l. More precisely, the index l determines the component of the predictor x (namely x_l), which is used in the kernel K_1 while the remaining components $x_{\underline{l}} = (x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_d)$ are used in K_2. Nevertheless all statistics $\widehat{F}_l$ estimate the conditional distribution function consistently (under appropriate assumptions). Moreover, for different values of $l = 1, \ldots, d$ different bandwidths $h_1 = h_{1,l}$, $H = \text{diag}(h_{2,l}, \ldots, h_{d,l})$ will be used in the estimate $\widehat{F}_l$, although this will not be reflected in our notation. Throughout this paper we denote by $G : \mathbb{R} \to [0, 1]$ a strictly increasing given distribution function, which can be specified by the data analyst and denote by K a further positive one-dimensional kernel with compact support, say $[-1, 1]$ with corresponding bandwidth b_n. Following Dette and Volgushev (2008) we define

$$\widehat{Q}_{l,N}(\tau|x) = G^{-1}\left(\widehat{G}_{l,N}(\tau|x)\right), \tag{6}$$

where the statistic $\widehat{G}_{l,N}$ is given by

$$\widehat{G}_{l,N}(\tau|x) = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\tau} K_{b_n}\left(\widehat{F}_l\left(G^{-1}\left(\frac{i}{N}\right)\Big|x\right) - u\right)du \tag{7}$$

and we use the notation $K_{b_n}(x) = K(x/b_n)/b_n$. Note that intuitively (for example if $\widehat{F}_l(y|x)$ is uniformly consistent) we obtain for $N \to \infty$, $n \to \infty$, $b_n \to 0$ the approximation

$$\begin{aligned}\widehat{G}_{l,N}(\tau|x) \approx G_N(\tau|x) &:= \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\tau} K_{b_n}\left(F\left(G^{-1}\left(\frac{i}{N}\right)\Big|x\right) - u\right)du \\ &\approx \int I\left\{F\left(G^{-1}(s)|x\right) \le \tau\right\}ds = G(Q(\tau|x)),\end{aligned} \tag{8}$$

and therefore the statistic $\widehat{Q}_{l,N}(\tau|x)$ defined in (6) is a reasonable estimate of the conditional quantile curve $Q(\tau|x) = F^{-1}(\tau|x)$. The distribution function G is introduced to treat the case where the density of the response variable has unbounded support. 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 $\hat{Q}_{l,N}(\tau|x)$ is robust with respect to the choice of the bandwidth b_n if it is chosen sufficiently small [see Dette et al. (2006)]. The estimate (6) suffers from the curse of dimensionality if the dimension d of the predictor is large and for this reason Dette and Scheder (2011) propose to combine it with the marginal integration technique to obtain an additive estimate of the quantile regression with a one-dimensional rate of convergence. To be precise define

$$\widehat{q}_l(\tau|x_l) = \frac{1}{n}\sum_{j=1}^{n}\widehat{Q}_{l,N}\left(\tau|x_l, X_{j\underline{l}}\right), \quad l = 1, \ldots, d$$

as an estimate of the first marginal effect

$$q_l(\tau|x_l) := \int Q(\tau|x) f_{\underline{l}}(x_{\underline{l}})dx_{\underline{l}} = Q_l(\tau|x_l) + c(\tau), \tag{9}$$

where $f_{\underline{l}} : \mathbb{R}^{d-1} \to \mathbb{R}$ is the density of the random vector $X_{j\underline{l}} = (X_{j1}, \ldots, X_{jl-1}, X_{jl+1}, \ldots, X_{jd})^T$ and the second equality in (9) holds under H_0. The estimates of the marginal effects $\widehat{q}_l(\tau|x_l)$ are now used to define the final additive estimate of the conditional quantile function which is given by

$$\widehat{Q}_{\text{add}}(\tau|x) := \sum_{k=1}^{d} \widehat{q}_k(\tau|x_k) - \left(1 - \frac{1}{d}\right) \sum_{k=1}^{d} \frac{1}{n} \sum_{i=1}^{n} \widehat{q}_k(\tau|X_{ik}). \tag{10}$$

We note that this statistic is well defined even in the case when the null hypothesis (1) is not satisfied and in this case it estimates consistently (under appropriate assumptions) the function

$$Q_{\text{add}}(\tau|x) = \sum_{j=1}^{d} Q_j(\tau|x) + c(\tau),$$

where the quantities Q_j are defined as in (9). Throughout this paper we make the following assumptions regarding the kernels used in the definition of (2), (5) and (7).

Assumption 1 The one-dimensional kernel K_1 in (5) is of bounded variation and has compact support $[-1, 1]$ with existing moments of order 2 satisfying

$$\int_{-1}^{1} x K_1(x) dx = 0,$$
$$c_2(K_1) = \frac{1}{2} \int_{-1}^{1} x^2 K_1(x) dx.$$

Similarly for a multi index $\nu_{\underline{1}} = (\nu_2, \dots, \nu_d) \in \mathbb{N}^{d-1}$ we define the monomial $x_{\underline{1}}^{\nu_{\underline{1}}} = x_2^{\nu_2}, \dots, x_d^{\nu_d}$, denote by $|\nu_{\underline{1}}| := \sum_{i=2}^{d} \nu_i$ the corresponding degree.

Assumption 2 We assume that the kernel K_2 in (5) is a $(d-1)$-dimensional bounded kernel of order q with support $[-1, 1]^{d-1}$, that is

(i) K_2 is symmetric,

(ii) $\int_{[-1,1]^{d-1}} K_2(x_{\underline{1}}) dx_{\underline{1}} = 1,$

(iii) $\int_{[-1,1]^{d-1}} |x_{\underline{1}}^{\nu_{\underline{1}}}||K_2(x_{\underline{1}})| dx_{\underline{1}} < \infty \quad \text{for } |\nu_{\underline{1}}| \le q,$

(iv) $\int_{[-1,1]^{d-1}} x_{\underline{1}}^{\nu_{\underline{1}}} K_2(x_{\underline{1}}) dx_{\underline{1}} = 0 \quad \text{for } 1 \le |\nu_{\underline{1}}| \le q - 1,$

(v) $\int_{[-1,1]^{d-1}} x_{\underline{1}}^{\nu_{\underline{1}}} K_2(x_{\underline{1}}) dx_{\underline{1}} \ne 0 \quad \text{for some } |\nu_{\underline{1}}| = q,$

and fulfills the following regularity assumption [see for example Einmahl and Mason (2005)]: consider the class of functions

$$\mathcal{K} = \left\{ K_2 \left(H^{-1}(x - \cdot) \right) | H = \text{diag}(h_2, \dots, h_d), h_i > 0, x \in \mathbb{R}^{d-1} \right\}.$$

For some $C > 0$ and $V > 0$ we assume that $\mathcal{K}$ satisfies the following uniform entropy condition:

$$\sup_P \ \mathcal{N}(\varepsilon, \mathcal{K}, L_2(P)) \le \left(\frac{C}{\varepsilon}\right)^V, \quad \text{for } 0 < \varepsilon < C, \tag{11}$$

where $\mathcal{N}(\varepsilon, \mathcal{K}, L_2(P))$ denotes the minimal number of balls of $L_2(P)$-radius ε needed to cover $\mathcal{K}$.

Nolan and Pollard (1987) and van der Vaart and Wellner (1996) give some criteria under which (11) holds. For example this assumption is satisfied, if $K_2(x) = \Phi(p(x))$ where $\Phi : \mathbb{R} \to \mathbb{R}$ is a real-valued function of bounded variation and $p(x)$ is a polynomial. Similarly, the assumption is also satisfied if the $(d-1)$ dimensional kernel K_2 is a product of one-dimensional kernels, where for each factor a condition of the type (11) holds.

Assumption 3

The kernel K is Lipschitz continuous with compact support $[-1, 1]$.
The kernel L is a d-dimensional symmetric kernel of order 2 with compact support $[-1, 1]^d$ and satisfies $L(x) < \infty$, $L(x) \geq 0$ for all $x \in [-1, 1]^d$.

To motivate the use of the statistic T_n to test H_0 we introduce the "residuals"

$$R_j = I\left\{Y_j \leq Q\left(\tau|X_j\right)\right\} - \tau \tag{12}$$
$$R_j^{\text{add}} = I\left\{Y_j \leq Q_{\text{add}}\left(\tau|X_j\right)\right\} - \tau \tag{13}$$

and denote by

$$\Delta(X_j) = E[R_j - R_j^{\text{add}} \mid X_j] = -E[R_j^{\text{add}} \mid X_j] \tag{14}$$

the conditional expectation of the distance between the "unconstrained residuals" and the "restricted residuals" obtained from an additive approximation. Note that under the null hypothesis we have $\Delta(X_j) = 0$ a.s., while under the alternative it follows that $P(\Delta(X_j) = 0) < 1$. With the approximation $\widehat{R}_i \approx R_i^{\text{add}}$ we get that the expectation of T_n can be approximated by

$$\begin{aligned}
E[T_n] &= E\left[L_g(X_i - X_j)\widehat{R}_i\widehat{R}_j\pi(X_i)\pi(X_j)\right] \\
&\approx \int L_g(u-v)(F(Q_{\text{add}}(\tau|u)|u) - \tau)(F(Q_{\text{add}}(\tau|v)|v) \\
&\quad -\tau)\pi(u)\pi(v)f(u)f(v)du dv \\
&\approx \int (F(Q_{\text{add}}(\tau|u)|u) - \tau)^2\pi^2(u)f^2(u)du \\
&= E\left[\Delta^2(X_j)\pi^2(X_j)f(X_j)\right] \geq 0.
\end{aligned}$$

Here we have $E[\Delta^2(X_j)\pi^2(X_j)f(X_j)] = 0$ if and only if $P(Q_{\text{add}}(\tau|X_j) = Q(\tau|X_j)) = 1$. In fact we will see in the next section that T_n converges to $E[\Delta^2(X_j)\pi^2(X_j)f(X_j)]$ in probability. Therefore, it is reasonable to reject the null hypotheses for large T_n.

3 Asymptotic theory

In this section, we study the asymptotic properties of the statistic introduced in Sect. 2 for testing the hypothesis of an additive quantile regression. We begin with a statement regarding weak convergence under the null hypothesis. To keep the notation simple we assume that the $(d-1)$-dimensional bandwidth matrix in the definition of the estimate (5) is proportional to the identity matrix, that is $H = \operatorname{diag}(h_2, \ldots, h_2) \in \mathbb{R}^{(d-1)\times(d-1)}$, where h_2 is a one-dimensional bandwidth. We also introduce the notation $K_{2,h_2}(x)$ instead of $K_{2,H}(x)$ in this case. Moreover, to present a result regarding weak convergence under the null hypothesis we make the following basic assumptions.

Assumption 4 1. The random variables X_j have a positive density $f \in C^q([0,1]^d)$ with support $supp(f) = [0,1]^d$, where $q \geq d$ and $C^q([0,1]^d)$ denotes the set of all q times continuously differentiable functions defined on the unit cube $[0,1]^d$.
2. For any x the function $F(\cdot|x)$ is strictly increasing and continuously differentiable with uniformly bounded derivative.
3. The distribution function G is twice continuously differentiable and $(G^{-1})'$ is uniformly bounded on closed intervals $I \subset (0,1)$.
4. The positive weight function π is continuously differentiable and has compact support $S \subset (0,1)^d$.
5. For any x the function $Q(\cdot|x)$ is twice continuously differentiable in a neighborhood of τ and there exists $\varepsilon > 0$ such that

$$\sup_{x\in[0,1]^d} \sup_{|s-\tau|<\varepsilon} Q'(s|x) < \infty,$$
$$\sup_{x\in[0,1]^d} \sup_{|s-\tau|<\varepsilon} |Q''(s|x)| < \infty.$$

6. For each $l = 1, \ldots, d$, the bandwidths g, b_n, h_1, h_2 satisfy the following conditions (if $n \to \infty$)

$$\begin{aligned}
&n = O(N), \quad N \to \infty, \ b_n = o(h_1) \\
&g^d = o(h_1^2), \quad nh_1^5 = O(1) \\
&ng^d \to \infty, \quad nb_n \to \infty, \quad nh_1h_2^{d-1} \to \infty \\
&h_2^q = o(h_1^2), \quad nh_2^{2q+1} = O(1).
\end{aligned}$$

Assumption 5

For each $l = 1, \ldots, d$, the bandwidths g, b_n, h_1, h_2 satisfy the following conditions (if $n \to \infty$)

$$\frac{n^{2\alpha}}{nh_1h_2^{d-1}b_n^2} = o(1)$$
$$n^{2\alpha} g^{\frac{d}{2}} \frac{1}{h_1h_2^{d-1}} = o(1)$$

for some $\alpha > 0$.

We note that the order q of the kernel K_2 provides an upper bound for the dimension d. However, q can be chosen by the experimenter and increasing the order of the kernel

K_2 makes the test applicable to d-dimensional predictors for any $d \in \mathbb{N}$. The following result establishes weak convergence of the test statistic T_n defined in (2). Throughout this paper the symbol $\xrightarrow{\mathcal{D}}$ denotes weak convergence.

Theorem 1 *If Assumptions 1, 2, 3, 4, 5 and the null hypothesis (1) of an additive quantile regression model are satisfied, it follows that*

$$ng^{d/2}T_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{15}$$

where the asymptotic variance is given by

$$\sigma^2 = 2\tau^2(1-\tau)^2 \int L^2(u)du \int \pi^4(x) f^2(x)dx. \tag{16}$$

Remark 1 We would like to point out that a result of the form (15) is typical for the limit distribution of a statistic of the type defined in (2) [see Gozalo and Linton (2001), or Dette and von Lieres und Wilkau (2001)]. For example, recently Härdle et al. (2012) considered the problem of testing the hypothesis of causality in quantile regression, which reduces in the simplest case to the hypothesis (for a given $l \in \{1, \ldots, d\}$)

$$H_0^c : Q(\tau \mid x) = Q(\tau \mid x_l).$$

This hypothesis means that the conditional quantile given $X = x$ does not depend on the components $x_1, \ldots, x_{l-1}, x_{l+1}, \ldots, x_d$ of the vector x. Härdle et al. (2012) proposed a statistic of the form (2), where the residuals $\hat{R}_i$ are replaced by $\tilde{R}_i = I\{Y_i \leq \hat{Q}(\tau|X_{il})\}$ and $\hat{Q}(\tau|x_l)$ is an appropriate estimate of the conditional quantile function under the null hypothesis H_0^c. They claimed asymptotic normality of a normalized test statistic

$$J_n = \frac{1}{n(n-1)} \sum_{i \neq j} L_g(X_i - X_j)\tilde{R}_i \tilde{R}_j$$

with the same limit distribution as given in Theorem 5. However, it should be pointed out here that the proof in this paper is not correct. The basic argument of Härdle et al. (2012) consists in the statement that the fact

$$\sup_x | \hat{Q}(\tau|x_l) - Q(\tau|x_l)| \leq C_n$$

results in the estimate

$$J_{nU} \leq J_n \leq J_{nL} \tag{17}$$

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

$$J_{nU} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g(X_i - X_j)\varepsilon_{iU}\varepsilon_{jU},$$
$$J_{nL} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g(X_i - X_j)\varepsilon_{iL}\varepsilon_{jL},$$

and $\varepsilon_{iU} = I\{Y_i + C_n \le Q(\tau|X_{il})\} - \tau,\ \varepsilon_{iL} = I\{Y_i + C_n \le Q(\tau|X_{il})\} - \tau$ (see equation (A.11-3) in this paper). A simple calculation shows that this conclusion is not correct and in fact the inequality (17) does not hold in general. It turns out that the proof of Theorem 1 in Härdle et al. (2012) can not be corrected easily.

However, using similar arguments as given in the proof of Theorem 1, it can be shown that a similar statement of weak convergence holds for a slightly modified statistic considered in Härdle et al. (2012), that is

$$\frac{g^{d/2}}{(n-1)} \sum_{i \neq j} L_g(X_i - X_j) \left(I\left\{Y_i \le \hat{Q}^{-i}(\tau|X_{il})\right\} - \tau\right) \left(I\left\{Y_j \le \hat{Q}^{-j}(\tau|X_{jl})\right\} - \tau\right)$$
$$\pi(X_i)\pi(X_j) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$$

where $\hat{Q}^{-i}(\tau|X_{il})$ denotes the quantile regression estimate of Dette and Volgushev (2008) from the two-dimensional sample $(X_{il}, Y_i)_{i=1}^n$ and σ^2 is defined in (16) (we omit details here for the sake of brevity). A correct proof of the result claimed in Härdle et al. (2012) is still an open problem.

In the following discussion, we investigate the asymptotic properties of the statistic T_n defined in (2) under local and fixed alternatives. We first consider the properties of the test for local alternatives of the form

$$Q(\tau|x) = Q_{\text{add}}(\tau|x) + d_n l(x), \tag{18}$$

where d_n denotes a sequence satisfying $d_n = (ng^{d/2})^{-1/2} \to 0$ as $n \to \infty$ and the function $l(\cdot)$ and its first-order derivatives are bounded.

Theorem 2 *Assume that Assumptions 1, 2, 3, 4 and 5 are satisfied. Under local alternatives of the form (18) with $d_n = 1/(n^{1/2} g^{\frac{d}{4}})$ it follows that*

$$ng^{\frac{d}{2}} T_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2), \tag{19}$$

where the asymptotic variance and bias are given by (16) and

$$\mu = E\left[\left(F'\left(Q_{\text{add}}\left(\tau|X_1\right)|X_1\right)\right)^2 l^2(X_1)\pi^2(X_1) f(X_1)\right],$$

respectively.

The following result specifies the asymptotic distribution of the test statistic T_n defined in (2) under fixed alternatives. For its proof we require the following additional assumptions.

Assumption 6

1. For any $y \in \mathbb{R}$ we have $F(y|\cdot) \in C_b^q([0,1]^d)$.
2. For each $l = 1, \ldots, d$, the bandwidths b_n, h_1, h_2 satisfy the following conditions (if $n \to \infty$)

$$\frac{\log n}{n h_1^2 h_2^{2(d-1)} b_n^2} = o(1),$$
$$\frac{n^{2\alpha - 1/2}}{h_1 h_2^{d-1}} = o(1), \quad \text{for some } \alpha > 0.$$

Theorem 3 *If Assumptions 1, 2, 3, 4 and 6 are satisfied and the null hypothesis (1) does **not** hold, then we have as n → ∞*

$$n^{1/2}(T_n - E[T_n]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{20}$$

where

$$\begin{aligned} E[T_n] = E&\left[\Delta^2(X_1)\pi^2(X_1)f(X_1)\right] \\ &+2E\Big[F'(Q_{\text{add}}(\tau|X_1)|X_1)\Delta(X_1)\pi^2(X_1)f(X_1)\Big(b(X_1) \\ &-\left(1-\frac{1}{d}\right)b(X_2)\Big)\Big]h_1^2 + o\left(h_1^2\right) + O(g^2) \end{aligned}$$

with $b(x) = \sum_{\alpha=1}^d b_\alpha(x_\alpha)$ *and*

$$\begin{aligned} b_\alpha(x_\alpha) = c_2(K_1)\int\Bigg(&\frac{1}{2}\frac{\frac{\partial^2}{\partial x_\alpha^2}F(Q(\tau|x_\alpha, t_{\underline{\alpha}})|x_\alpha, t_{\underline{\alpha}})}{F'(Q(\tau|x_\alpha, t_{\underline{\alpha}})|x_\alpha, t_{\underline{\alpha}})} \\ &+\frac{\frac{\partial}{\partial x_\alpha}F(Q(\tau|x_\alpha, t_{\underline{\alpha}})|x_\alpha, t_{\underline{\alpha}})\frac{\partial}{\partial x_\alpha}f(x_\alpha, t_{\underline{\alpha}})}{F'(Q(\tau|x_\alpha, t_{\underline{\alpha}})|x_\alpha, t_{\underline{\alpha}})f(x_\alpha, t_{\underline{\alpha}})}\Bigg) f_{\underline{\alpha}}\left(t_{\underline{\alpha}}\right)dt_{\underline{\alpha}}. \end{aligned} \tag{21}$$

The asymptotic variance in (20) is given by

$$\begin{aligned} \sigma^2 = 4\mathrm{Var}\Bigg[&\Delta^2(X_1)\pi^2(X_1)f(X_1) \\ &-E\Bigg(\Delta(X_2)\pi^2(X_2)f(X_2)F'(Q_{\text{add}}(\tau|X_2)|X_2)\Bigg(\sum_{\alpha=1}^d Q\left(\tau|X_{2\alpha}, X_{1\underline{\alpha}}\right) \\ &-\left(1-\frac{1}{d}\right)\sum_{\alpha=1}^d Q\left(\tau|X_{1\alpha}, X_{3\underline{\alpha}}\right) + Q\left(\tau|X_{3\alpha}, X_{1\underline{\alpha}}\right)\Bigg)\Bigg| X_1\Bigg)\Bigg] \end{aligned}$$

$$
\begin{aligned}
&+4\tau(1-\tau)E\Bigg[\bigg(-\Delta(X_1)\pi^2(X_1)f(X_1)\\
&-\frac{1}{F'(Q(\tau|X_1)|X_1)}\Big((d-1)\int\Delta(t)\pi^2(t)f^2(t)F'(Q_{\text{add}}(\tau|t)|t)dt\\
&-\sum_{\alpha=1}^{d}\frac{f_{\underline{\alpha}}(X_{1\underline{\alpha}})}{f(X_1)}\int\Delta\left(X_{1\alpha},t_{\underline{\alpha}}\right)\pi^2\left(X_{1\alpha},t_{\underline{\alpha}}\right)f^2\left(X_{1\alpha},t_{\underline{\alpha}}\right)\\
&F'(Q_{\text{add}}(\tau|X_{1\alpha},t_{\underline{\alpha}})|X_{1\alpha},t_{\underline{\alpha}})dt_{\underline{\alpha}}\Big)\bigg)^2\Bigg].
\end{aligned}
$$

Remark 2 Note that Theorem 1 provides an asymptotic level α test for the hypothesis (1) of an additive quantile regression model by rejecting H_0, whenever

$$T_n > \hat{\sigma}_n u_{1-\alpha},$$

where $\hat{\sigma}_n^2$ is an appropriate estimate of the asymptotic variance σ^2 defined in (16). Moreover, by Theorem 3 it follows that this test is consistent, because under the alternative we have

$$T_n \xrightarrow{\mathcal{D}} E[\Delta^2(X_1)\pi^2(X_1)f(X_1)] > 0$$

from this result.

4 Finite sample properties and a data example

4.1 A small simulation study

To investigate the finite sample properties of the new test we have performed a small simulation study. To be precise, we consider the median regression model

$$Y_i = Q(0.5|X_i) + 0.25\varepsilon_i, \tag{22}$$

where ε_i are independent, standard normally distributed and independent of the four-dimensional covariates $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4}), i = 1, \ldots, n$. For the choice of the predictor we investigate the following two scenarios.

(A) X_i are uniformly distributed on the unit square $[0, 1]^4$, that is

$$X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4}) \sim \mathcal{U}\big([0, 1]^4\big), \quad i = 1, \ldots, n,$$

(B) $X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4})$ are given by

$$X_{ij} = \frac{1}{2} + \frac{1}{\pi}\arctan(N_{ij}), \quad j = 1, \ldots, 4,$$

where $N_i = (N_{i1}, N_{i2}, N_{i3}, N_{i4})$ are (independent) centered multivariate normally distributed random variables with different covariance matrices

$$V_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 1 & 0.1 & 0.2 & 0.1 \\ 0.1 & 1 & 0.1 & 0.2 \\ 0.2 & 0.1 & 1 & 0.1 \\ 0.1 & 0.2 & 0.1 & 1 \end{pmatrix}$$

$$\text{and } V_3 = \begin{pmatrix} 1 & 0.3 & 0.5 & 0.1 \\ 0.3 & 1 & 0.3 & 0.5 \\ 0.5 & 0.3 & 1 & 0.3 \\ 0.1 & 0.5 & 0.3 & 1 \end{pmatrix}.$$

Note that in Design (A) the random variables X_{i1},X_{i2},X_{i3} and X_{i4} are independent, whereas Design (B) also represents situations where X_{i1},X_{i2},X_{i3} and X_{i4} are correlated. In our simulation study, we consider six models for the conditional quantile function, that is

$$Q(0.5|x_1, x_2, x_3, x_4) = x_1 + x_2 + x_3 + x_4, \tag{23}$$

$$Q(0.5|x_1, x_2, x_3, x_4) = x_1^2 + x_2^2 + x_3^2 + x_4^2, \tag{24}$$

$$Q(0.5|x_1, x_2, x_3, x_4) = \cos(x_1) + \cos(x_2) + \cos(x_3) + \cos(x_4), \tag{25}$$

$$Q(0.5|x_1, x_2, x_3, x_4) = \cos(\pi(x_1 + x_2 + x_3 + x_4)), \tag{26}$$

$$Q(0.5|x_1, x_2, x_3, x_4) = \exp(2(x_1 + x_2 + x_3 + x_4)), \tag{27}$$

$$Q(0.5|x_1, x_2, x_3, x_4) = (x_1 + x_2 + x_3 + x_4)^3, \tag{28}$$

where the first three cases correspond to the null hypothesis of additivity and (26), (27), (28) represent three alternatives. For all kernels of order 2 in our estimators we use the Epanechnikov kernel $K(t) = \frac{3}{4}(1 - t^2)I_{[-1,1]}(t)$, and a product of kernels of this type as a multi-dimensional kernel. To construct higher order kernels we use the notation $c_i = \int u^i K(u)du$. Then

$$\tilde{K}(u) = \begin{vmatrix} c_1 & c_2 & \dots & c_l & 1 \\ c_2 & c_3 & \dots & c_{l+1} & u \\ \vdots & & & & \vdots \\ c_{l+1} & c_{l+2} & \dots & c_{2l} & u^l \end{vmatrix} K(u) \begin{vmatrix} c_1 & c_2 & \dots & c_l & c_0 \\ c_2 & c_3 & \dots & c_{l+1} & c_1 \\ \vdots & & & & \vdots \\ c_{l+1} & c_{l+2} & \dots & c_{2l} & c_l \end{vmatrix}^{-1}$$

is a kernel of order l.

In similar problems it has been observed by several authors [see Fan and Linton (2003)] that the asymptotic normal distribution under the null hypothesis does not provide a satisfactory approximation for the distribution of the statistic T_n for small sample sizes. For this reason many authors propose the application of a bootstrap in this context to calculate critical values. We follow this suggestion and use a wild bootstrap for this purpose. To be precise, in the τ-quantile model we define a bootstrap sample by

$$Y_i^* = \hat{Q}_{\mathrm{add}}\big(\tau|X_i\big) + v_i|Y_i - \hat{Q}_N\big(\tau|X_i\big)|, \quad i = 1, \ldots, n, \tag{29}$$

where $\widehat{Q}_{\mathrm{add}}$ is defined in Sect. 2 and $\widehat{Q}_N$ is the multivariate quantile regression estimator by Dette and Volgushev (2008), which is defined in the same way as (6), where the estimator of the conditional distribution function $\widehat{F}_l$ is replaced by an estimator treating all components of the predictor equally. Further v_i denote independent and identically distributed random variables satisfying $P(v_i = -1) = \tau$ and $P(v_i = 1) = 1 - \tau$, which are independent from the original sample $\mathcal{Y}_n = \{(X_j, Y_j) \mid j = 1, \ldots, n\}$. A similar bootstrap data generation was suggested by Sun (2006) and Feng et al. (2011). Note that, conditionally on the original sample, the bootstrap observations fulfill the null hypotheses of additivity and additionally fulfill a τ-quantile regression model, that is

$$P^* \left(Y_i^* \leq \hat{Q}_{\mathrm{add}}\big(\tau|X_i\big) \mid X_i\right) = P(v_i \leq 0) = \tau$$

almost surely, where P^* denotes the probability conditionally on $\mathcal{Y}_n$. Note that for the median model used in the simulations we have $\tau = \frac{1}{2}$ and v_i are Rademacher variables. Further note that by construction of the bootstrap errors we mimic the dependence of the ith error term from the ith covariate to obtain a bootstrap model approximating the unknown data generating model. This approach is similar to wild bootstrap in mean regression as introduced by Härdle and Mammen (1993). Some arguments regarding the validity of this resampling method are given in Sect. 4.3. We conjecture that the generation of bootstrap data according to the model $Y_i^* = \hat{Q}_{\mathrm{add}}(\tau|X_i) + v_i$ would also yield a valid bootstrap procedure from a theoretical point of view. However, for small sample sizes, this model is likely to yield worse approximations of the nominal level.

Now let T_n^* denote the test statistic based on the bootstrap data (X_i, Y_i^*), $i = 1, \ldots, n$. We indicate in Sect. 4.3 that both under H_0 and under fixed alternatives, conditionally on $\mathcal{Y}_n$,

$$ng^{d/2}T_n^* \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{30}$$

in probability, where σ^2 is defined in (16). The critical value for the test is then obtained from the bootstrap distribution

$$P^*(T_n^* \geq t^*_{n,(1-\alpha)}) = 1 - \alpha,$$

and the hypothesis of additivity is rejected whenever $T_n \geq t^*_{n,(1-\alpha)}$. From Theorems 1 and 3 together with (30) it follows that this hypothesis test has asymptotic level α and is consistent against fixed alternatives. For the estimation of $t^*_{n,(1-\alpha)}$ we choose the number of bootstrap replications as $B = 100$ and we have simulated the rejection probabilities of this test on the basis of 1000 replications of each experiment.

The performance of this test depends on the choice of the bandwidths and we have implemented the following data driven rules.

1. For the estimator $\widehat{Q}_N$ in (29) we select the bandwidths following Abberger (1998) by calculating

$$\underset{h_1,\dots,h_d,b_n}{\operatorname{argmin}} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \widehat{Q}_N^{-i}(X_i) \right), \tag{31}$$

where $\rho_\tau(u) = u(\tau - I\{u \le 0\})$.

2. The selection of the bandwidths for the additive estimator is more complicated. If the hypothesis of an additive quantile regression is not true, the analog of the procedure (31) might lead to overfitting. The additive estimator with the cross-validated bandwidths converges very slowly to the additive model

$$\begin{aligned} Q_{\text{add}}(\tau|\mathbf{x}) &= \sum_{j=1}^{d} Q_j(\tau|x_j) + c(\tau) \\ &= \sum_{j=1}^{d} \int Q(\tau|\mathbf{x}) f_{\underline{j}}(\mathbf{x}_{\underline{j}}) d\mathbf{x}_{\underline{j}} - \left(1 - \frac{1}{d}\right) \sum_{j=1}^{d} \int Q(\tau|\mathbf{x}) f(\mathbf{x}) d\mathbf{x}, \end{aligned}$$

and for reasonable sample sizes it still tries to interpolate the data points. To avoid this problem we introduce a theoretical additive model which we will only use for the bandwidth selection in the additive estimator, that is

$$\tilde{Y}_i = \frac{\sigma(\widehat{Q}_N^{-i})}{\sigma(\sum_{j=1}^{d} X_{ij})} \left(\sum_{j=1}^{d} X_{ij} \right) + v_i |Y_i - \widehat{Q}_N^{-i}(X_i)|. \tag{32}$$

Here $v_i, i = 1, \dots, n$ are random variables, independent from the original sample and $\sigma(\widehat{Q}_N^{-i})$ and $\sigma(\sum_{j=1}^{d} X_{ij})$ denote the standard deviations of $\widehat{Q}_N^{-1}, \dots, \widehat{Q}_N^{-n}$ and $\sum_{j=1}^{d} X_{1j}, \dots, \sum_{j=1}^{d} X_{nj}$, respectively. Now the bandwidths for the estimator $\widehat{Q}_{\text{add}}$ are chosen by calculating

$$\underset{h_1,\dots,h_d,b_n}{\operatorname{argmin}} \sum_{i=1}^{n} \left| \frac{\sigma(\widehat{Q}_N^{-i})}{\sigma(\sum_{j=1}^{d} X_{ij})} \left(\sum_{j=1}^{d} X_{ij} \right) - \tilde{Q}_{\text{add}}^{-i}(\tau|X_i) \right|. \tag{33}$$

where $\tilde{Q}_{\text{add}}^{-i}(\tau|X_i)$ is the additive estimator in the theoretical model (32). The term $\frac{\sigma(\widehat{Q}_N^{-i})}{\sigma(\sum_{j=1}^{d} X_{ij})}$ tries do imitate the scale of the true model, while $v_i|Y_i - \widehat{Q}_N^{-i}(X_i)|$ imitates the original error term.

3. Finally the bandwidths g_j in the kernel L are chosen as 0.6 times the standard deviations of $X_{1j}, \dots X_{nj}$ $(j = 1, \dots, d)$.

In Table 1 we display the results of the simulation study for model (23), (24), (25) which represent the null hypothesis where the sample size is $n = 100$. The

Table 1 Simulated level of the bootstrap test for the hypothesis of an additive quantile regression model under the null hypothesis of additivity

	Model (23)				Model (24)				Model (25)			
	(A)	(B)			(A)	(B)			(A)	(B)		
		V_1	V_2	V_3		V_1	V_2	V_3		V_1	V_2	V_3
5 %	5.1	6.4	6.8	6.0	5.5	5.6	5.9	4.2	5.9	6.3	4.5	6.7
10 %	9.8	11.3	10.6	13.6	9.8	10.5	10.7	8.5	10.9	11.2	9.5	11.6
20 %	20.7	22.4	20.2	22.4	22.5	20.5	21.3	18.2	19.7	21.8	19.7	22.8

Table 2 Simulated power of the bootstrap test for the hypothesis of an additive quantile regression model corresponding to the alternative

	Model (26)				Model (27)				Model (28)			
	(A)	(B)			(A)	(B)			(A)	(B)		
		V_1	V_2	V_3		V_1	V_2	V_3		V_1	V_2	V_3
5 %	95.0	89.6	96.0	99.9	89.8	76.3	75.2	90.3	86.5	72.7	86.4	93.2
10 %	97.3	94.4	97.2	100.0	94.6	84.7	83.3	93.9	91.2	84.0	92.8	96.3
20 %	99.2	98.8	99.2	100.0	97.4	92.9	91.3	97.3	95.5	94.0	97.6	98.7

corresponding results under the alternative defined by model (26), (27), (28) are shown in Table 2. Under the null hypothesis we observe a reasonable approximation of the nominal level under Design (A) and (B) (see Table 1). The results in Table 2 demonstrate that the bootstrap test detects alternatives with reasonable power in all cases under investigation. To investigate the properties of the test statistic for other quantiles than the median, we considered the cases $\tau = 0.25$ and $\tau = 0.75$, respectively. For the regression model (22), the conditional quantile function is given by

$$Q(\tau|\mathbf{x}) = Q(0.5|\mathbf{x}) + 0.25\Phi(\tau)$$

for all $\tau \in (0, 1)$. Here $\Phi(\tau)$ denotes the τ—quantile of the standard normal distribution. We considered one null hypothesis [model (23)] and one alternative [model (26)] for each scenario. The results can be found in Tables 3 and 4, respectively. We observe similar power properties as for the median.

Finally, to study the robustness of the procedure we investigated Cauchy distributed error variables, i.e., the ε_i are independent, standard Cauchy distributed random variables and independent from the covariates. For the sake of brevity we considered the median and one null hypothesis [model (23)] and one alternative [model (26)]. The results can be found in Table 5. The approximation of the nominal level is quite satisfactory. One the other hand for Cauchy distributed errors the procedure is less powerful.

Table 3 Simulated level and power of the bootstrap test for $\tau = 0.25$ and $n = 100$

	Model (23)				Model (26)			
	(A)	(B)			(A)	(B)		
		V_1	V_2	V_3		V_1	V_2	V_3
5 %	6.3	6.4	6.4	8.6	91.6	87.2	97.3	100.0
10 %	11.1	11.4	11.6	14.8	95.6	92.4	98.7	100.0
20 %	19.4	22.3	21.7	24.8	97.2	97.6	99.3	100.0

Table 4 Simulated level and power of the bootstrap test for $\tau = 0.75$ and $n = 100$

	Model (23)				Model (26)			
	(A)	(B)			(A)	(B)		
		V_1	V_2	V_3		V_1	V_2	V_3
5 %	5.6	6.0	6.3	4.4	90.2	86.5	84.5	92.4
10 %	10.6	11.6	10.8	9.4	94.1	91.9	90.8	95.1
20 %	18.6	23.7	19.4	19.3	97.0	95.4	93.2	98.1

Table 5 Simulated level and power of the bootstrap test for Cauchy distributed error variables

	Model (23)				Model (26)			
	(A)	(B)			(A)	(B)		
		V_1	V_2	V_3		V_1	V_2	V_3
5 %	5.6	5.1	5.4	6.8	58.7	53.1	66.3	84.5
10 %	10.7	9.9	11.3	12.5	69.9	62.3	75.8	90.9
20 %	21.9	20.8	22.9	21.9	79.8	76.4	86.9	95.6

4.2 A data example

We illustrate the test of additivity analyzing a data example from Yeh (2007), who models the slump flow of concrete. The data set contains seven input variables and three output variables. The output variables are the slump and the flow (measured in cm) of concrete, which are measures of the consistency of concrete and the 28-day compressive strength of concrete. The input variables are given by

- Cement (kg in one m^3 concrete)
- Slag (kg in one m^3 concrete)
- Fly ash (kg in one m^3 concrete)
- Water (kg in one m^3 concrete)
- SP (kg in one m^3 concrete)

- Coarse Aggr. (kg in one m^3 concrete)
- Fine Aggr. (kg in one m^3 concrete)

and the correlation matrix between these variables is estimated as

$$\begin{pmatrix} 1 & -0.24 & -0.49 & 0.22 & -0.11 & -0.31 & 0.06 \\ -0.24 & 1 & -0.32 & -0.03 & 0.31 & -0.22 & -0.18 \\ -0.49 & -0.32 & 1 & -0.24 & -0.14 & 0.17 & -0.28 \\ 0.22 & -0.03 & -0.24 & 1 & -0.16 & -0.60 & 0.11 \\ -0.11 & 0.31 & -0.14 & -0.16 & 1 & -0.10 & 0.06 \\ -0.31 & -0.22 & 0.17 & -0.60 & -0.10 & 1 & -0.49 \\ 0.06 & -0.18 & -0.28 & 0.11 & 0.06 & -0.49 & 1 \end{pmatrix}.$$

We observe that the correlations are of similar size as the correlations considered in the previous section [Design (B), correlation matrix V_3].

First, we focus on the variable slump. We want to check if the median regression function of slump given the seven covariates is additive. Therefore, we apply the bootstrap test, where we use the bandwidth selection method described before. The p value from $B = 100$ bootstrap replications is $p = 0.19$. This indicates that the hypothesis of additivity cannot be rejected at a controlled type I error of 10 %. Now we apply the test to investigate whether the median regression function of the 28-day compressive strength given the seven covariates is additive. The p value from $B = 100$ bootstrap replications is given by $p = 0.90$. This indicates that the hypothesis of additivity cannot be rejected.

Finally we apply the test to investigate whether the median regression function of the variable flow given the seven covariates is additive. The p-value from $B = 100$ bootstrap replications is $p = 0.10$. This indicates that the hypothesis of additivity can be rejected at a controlled type I error of 10 %.

4.3 Some heuristics for the bootstrap test

A rigorous proof of the conditional weak convergence (30) can be obtained by mimicking the proof of Theorem 1. Because these arguments are very lengthy we only give the main steps here. One starts with a decomposition of the bootstrap statistic, $T_n^* = T_{1n}^* + T_{2n}^* + T_{3n}^*$, analogous to (34). The proof of the statements $ng^{d/2}T_{2n}^* = o_p(1)$ and $ng^{d/2}T_{3n}^* = o_p(1)$ (under the appropriate regularity assumptions) can be conducted similarly to the proof of (39) in Appendix A (but with even more technical effort). For example for the definition of T_{3nU}^* one sets

$$R_{iU}^* = I\big\{v_i|Y_i - Q(\tau|X_i)| \leq -2C_n^* - \tilde{C}_n\big\} - \tau, \quad R_{iL}^* = I\big\{v_i|Y_i - Q(\tau|X_i)| \leq 2C_n^* + \tilde{C}_n\big\} - \tau,$$

where C_n^* and $\tilde{C}_n$ denote uniform rates of convergence of $\hat{Q}_{\text{add}}^{*,-i} - \hat{Q}_{\text{add}}$ and $\hat{Q}_N - Q$, respectively. The remaining term is

$$ng^{d/2}T_{1n}^* = \frac{g^{d/2}}{n-1}\sum_{i\neq j} H_n^*(Z_i^*, Z_j^*)\pi(X_i)\pi(X_j),$$

where

$$H_n^*(Z_i^*, Z_j^*) = L_g\left(X_i - X_j\right)\left(I\left\{Y_i^* \leq \hat{Q}_{\text{add}}(\tau|X_i)\right\} - \tau\right)$$
$$\left(I\left\{Y_j^* \leq \hat{Q}_{\text{add}}(\tau|X_j)\right\} - \tau\right)$$

and $Z_i^* = (X_i, Y_i^*)$. Note that T_{1n}^* is no U-statistic with respect to the conditional probability measure P^* because here all X_i are known. However, $ng^{d/2}T_{1n}^*$ has a structure similar to a U-statistic and the proof of conditional weak convergence in probability follows along the lines of the proofs of Theorem 1 by Hall (1984) and Corollary 3.1 by Hall and Heyde (1996), p. 58 [see Neumeyer (2009), proof of Theorem 3.4] To motivate that one obtains the same limit $\mathcal{N}(0, \sigma^2)$ as in Theorem 1, we restrict ourselves to a consideration of the conditional variance, i. e.

$$\begin{aligned}
\text{Var}^*(ng^{d/2}T_{1n}^*) &= \frac{g^d}{(n-1)^2}\sum_{i\neq j}\sum_{k\neq l} E^* \\
&\quad \left[H_n^*(Z_i^*, Z_j^*)H_n^*(Z_k^*, Z_l^*)\pi(X_i)\pi(X_j)\pi(X_k)\pi(X_l)\right] \\
&= 2\frac{g^d}{(n-1)^2}\sum_{i\neq j} L_g^2\left(X_i - X_j\right)E^*\Big[E^*\big[\left(I\{Y_i^* \leq \hat{Q}_{\text{add}}(\tau|X_i)\} - \tau\right)^2|X_i\big] \\
&\quad \times E^*\big[\left(I\{Y_j^* \leq \hat{Q}_{\text{add}}(\tau|X_j)\} - \tau\right)^2|X_j\big]\Big]\pi^2(X_i)\pi^2(X_j) \\
&= 2\tau^2(1-\tau)^2\frac{g^d}{(n-1)^2}\sum_{i\neq j} L_g^2\left(X_i - X_j\right)\pi^2(X_i)\pi^2(X_j) \;=\; \sigma_n^2
\end{aligned}$$

almost surely. Here Var^* and E^* denote variance and expectation with respect to the conditional probability measure P^* and the last equality defines σ_n^2. Now for $n \to \infty$, σ_n^2 converges in probability to the desired variance $2\tau^2(1-\tau)^2 \int L^2 \int \pi^4 f^2 = \sigma^2$.

Appendix A: proof of Theorem 1

Throughout the proofs we assume for the sake of a transparent notation $N = n$ and a uniform distribution G. The general case follows by exactly the same arguments using an additional Taylor expansion. Recall the definition of the statistic T_n in (2) and consider the decomposition

$$T_n = T_{1n} + T_{2n} + T_{3n} \tag{34}$$

where the statistics T_{jn} $(j = 1, 2, 3)$ are given by

$$T_{1n} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) R_i R_j \pi(X_i)\pi(X_j) \tag{35}$$

$$T_{2n} = \frac{2}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) R_i \big(\widehat{R}_j - R_j\big)\pi(X_i)\pi(X_j) \tag{36}$$

$$T_{3n} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) \big(\widehat{R}_i - R_i\big)\big(\widehat{R}_j - R_j\big)\pi(X_i)\pi(X_j) \tag{37}$$

and R_i and $\widehat{R}_i$ are defined in (12) and (3), respectively. The assertion follows from the following two statements, which are proved below

$$ng^{\frac{d}{2}} T_{1n} \xrightarrow{\mathcal{D}} N(0, \sigma^2) \tag{38}$$

$$ng^{\frac{d}{2}} T_{jn} = o_p(1), \quad j = 2, 3. \tag{39}$$

Proof of (38)

Defining $Z_i = (X_i, Y_i)$, $i = 1, \ldots, n$, and

$$\begin{aligned} &H_n(Z_i, Z_j) \\ &\quad = L_g\left(X_i - X_j\right) \big(I\big\{Y_i \leq Q(\tau|X_i)\big\} - \tau\big)\big(I\big\{Y_j \leq Q(\tau|X_j)\big\} - \tau\big)\pi(X_i)\pi(X_j) \end{aligned}$$

we can write the statistic $ng^{\frac{d}{2}} T_{1n}$ as

$$ng^{\frac{d}{2}} T_{1n} = \frac{g^{d/2}}{n-1} \sum_{i=1}^{n} \sum_{j \neq i} H_n(Z_i, Z_j).$$

The assertion then follows from Theorem 1 in Hall (1984) for U-statistics if the assumptions of this statement can be checked. For this purpose, note that we obtain from Assumption 5 for $i \neq j \neq k \neq i$ for some $\lambda > 0$

$$\begin{aligned} &E\big[E\big[H_n(Z_k, Z_i)H_n(Z_k, Z_j)|Z_i, Z_j\big]^2\big] \\ &\quad \leq \frac{\lambda}{g^{4d}} E\left[E\left[L\left(\frac{X_k - X_i}{g}\right) L\left(\frac{X_k - X_j}{g}\right) |X_i, X_j\right]^2\right] \\ &\quad = \frac{\lambda}{g^{4d}} \int\int\left[\int L(u)L(u+v)f(x+ug)g^d du\right]^2 \\ &\qquad f(x)f(x-vg)g^d dx dv = O\left(\frac{1}{g^d}\right) \\ &E[H_n^2(Z_i, Z_j)] = \tau^2(1-\tau)^2 \frac{1}{g^{2d}} E\left[L^2\left(\frac{X_i - X_j}{g}\right) \pi^2(X_i)\pi^2(X_j)\right] \end{aligned}$$

$$= \tau^2(1-\tau)^2 \frac{1}{g^d} \int L^2(u)du \int \pi^4(x) f^2(x)dx + o\left(\frac{1}{g^d}\right) = \frac{\sigma^2}{2g^d} + o\left(\frac{1}{g^d}\right),$$

where $\sigma^2 > 0$ is defined in (16). In a similar way one establishes the estimate $E[H_n^4(Z_i, Z_j)] = O\left(\frac{1}{g^{3d}}\right)$, which gives

$$\frac{E[E[H_n(Z_k, Z_i)H_n(Z_k, Z_j)|Z_i, Z_j]^2] + n^{-1}E[H_n^4(Z_i, Z_j)]}{(E[H_n^2(Z_i, Z_j)])^2} = O\left(g^d\right) + O\left(\frac{1}{ng^d}\right) = o(1).$$

Therefore, Theorem 1 in Hall (1984) yields $ng^{\frac{d}{2}} T_{1n} \to N(0, \sigma^2)$, where the asymptotic variance σ^2 is given by (16).

Proof of (39)

For the proof of (39) we define for $\alpha > 0$ defined in Assumption 5

$$C_n = n^\alpha \sqrt{\frac{\log n}{nh_1h_2^{d-1}}}; \quad D_n = n^\alpha \frac{1}{nh_1} \tag{40}$$

and introduce the set

$$\Omega_n = \left\{ \sup_{x\in S} |\widehat{Q}_{\text{add}}(\tau|x) - Q(\tau|x)| \le C_n, \sup_{x\in S} \max_{k=1}^{n} |\widehat{Q}_{\text{add}}^{-k}(\tau|x) - \widehat{Q}_{\text{add}}(\tau|x)| \le D_n \right\}. \tag{41}$$

First, we consider the term T_{3n} and introduce the notation

$$T_{3nU} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} L_g\left(X_i - X_j\right)(R_{iU} - R_{iL})(R_{jU} - R_{jL})\pi(X_i)\pi(X_j)$$

where

$$R_{iU} = I\left\{Y_i \le Q(\tau|X_i) + 2C_n\right\} - \tau, \quad R_{iL} = I\left\{Y_i \le Q(\tau|X_i) - 2C_n\right\} - \tau.$$

It is easy to see, that on the set Ω_n

$$I\left\{Y_i \le Q(\tau|X_i) - 2C_n\right\} \le I\left\{Y_i \le \widehat{Q}_{\text{add}}^{-i}(\tau|X_i)\right\} \le I\left\{Y_i \le Q(\tau|X_i) + 2C_n\right\}$$

which implies (note that the kernel L is non-negative) $1\{\Omega_n\}|T_{3n}| \le 1\{\Omega_n\}T_{3nU} \le T_{3nU}$. Therefore, we have

$$E[|T_{3n}|] = E[1\{\Omega_n\}|T_{3n}|] + E\left[1\left\{\Omega_n^C\right\}|T_{3n}|\right] \le E[|T_{3nU}|] + \left(E\left[|T_{3n}|^2\right]P\left(\Omega_n^C\right)\right)^{1/2}.$$

We now calculate

$$E[|T_{3nU}|] = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} E\left[L_g\left(X_i - X_j\right)(R_{iU} - R_{iL})\right.$$
$$\left.(R_{jU} - R_{jL})\pi(X_i)\pi(X_j)\right].$$

Observing that $f'(x)$ and $F'(y|x)$ are bounded we obtain by a Taylor expansion

$$\begin{aligned} &E\left[L_g\left(X_i - X_j\right)(R_{iU} - R_{iL})(R_{jU} - R_{jL})\pi(X_i)\pi(X_j)\right] \\ &= E\left[L_g\left(X_i - X_j\right)(F\left(Q(\tau|X_i) + C_n|X_i\right) - F\left(Q(\tau|X_i) - C_n|X_i\right))\right. \\ &\quad \times \left(F\left(Q(\tau|X_j) + C_n|X_j\right) - F\left(Q(\tau|X_j) - C_n|X_j\right)\right)\pi(X_i)\pi(X_j)\left.\right] = O(C_n^2). \end{aligned}$$

With Assumption 5 we have $ng^{\frac{d}{2}} T_{3nU} = O_{L_1}(ng^{\frac{d}{2}} C_n^2) = o_{L_1}(1)$ and therefore the proof of (39) in the case $j = 3$ follows from $E[T_{3n}^2] = O(1/g^{2d})$ and the following result.

Lemma 1 *For Ω_n defined in (41) we have that*

$$P(\Omega_n^C) = O\left(p(n)\exp\left(-n^{2\alpha}\right)\right) \tag{42}$$

where $p(n)$ is a polynomial in n and α is defined in Assumption 5.

Proof of Lemma 1 For a proof of (42) it suffices to show that

$$P(\sup_{x\in S}|\widehat{Q}_{\text{add}}(\tau|x) - Q_{\text{add}}(\tau|x)| > C_n) = O\left(p(n)\exp\left(-n^{\alpha}\right)\right) \tag{43}$$
$$P(\sup_{x\in S}\max_i|\widehat{Q}_{\text{add}}(\tau|x) - \widehat{Q}^{-i}_{\text{add}}(\tau|x)| > D_n) = O\left(p(n)\exp\left(-n^{\alpha}\right)\right). \tag{44}$$

At first, we consider the probability (43). We have that

$$\sup_x|\widehat{Q}_{\text{add}}(\tau|x) - Q_{\text{add}}(\tau|x)| \leq \sum_{k=1}^{d}\left\{B_{nk}^{(1)} + \left(1 - \frac{1}{d}\right)B_{nk}^{(2)}\right\} \tag{45}$$

where

$$B_{nk}^{(1)} = \sup_{x_k}|\widehat{q}_k(\tau|x_k) - q_k(\tau|x_k)|$$
$$B_{nk}^{(2)} = \frac{1}{n}\sum_{i=1}^{n}|\widehat{q}_k(\tau|X_{ik}) - c(\tau)|$$

and consider the term $B_{n1}^{(1)}$ (the other cases are treated in exactly the same way). In the following calculations, all constants are denoted by C although they might differ from line to line. With the similar arguments as in Dette et al. (2006) and the assumptions regarding the bandwidths we have $q_1(\tau|x_1) = q_{1,n}(\tau|x_1) + o(C_n)$, uniformly with respect to x_1, where

$$q_{1,n}(\tau|x_1) = \frac{1}{n}\sum_{i=1}^{n} Q_{1,n}\big(\tau|\big(x_1, X_{i\underline{1}}\big)\big) \tag{46}$$

and we introduce the notation

$$Q_{1,n}(\tau|x) = G^{-1}(G_N(\tau|x))$$

and G_N is defined in (8). Recalling the definition of $\widehat{Q}_{l,n}$ in (6) we obtain by a Taylor expansion and similar arguments as in Dette and Scheder (2011)

$$\begin{aligned}
&\widehat{q}_1(\tau|x_1) - q_{1,n}(\tau|x_1)\\
&= \frac{1}{n}\sum_{j=1}^{n}[\widehat{Q}_{1,n}(\tau|(x_1, X_{j\underline{1}})) - Q_{1,n}(\tau|(x_1, X_{j\underline{1}}))]\\
&= \frac{1}{n^2}\sum_{i,j=1}^{n}\int_{-\infty}^{\tau}\frac{1}{b_n}K'_{b_n}\Big(F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big) - u\Big)\Big(\widehat{F}_1\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big)\\
&\quad - F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big)\Big)du\\
&\quad + \frac{1}{n^2}\sum_{i,j=1}^{n}\int_{-\infty}^{\tau}\frac{1}{b_n}\Big(K'_{b_n}(\xi_i - u) - K'_{b_n}\Big(F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big) - u\Big)\Big)\\
&\quad \times\Big(\widehat{F}_1\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big) - F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big)\Big)du\\
&= \Delta_n^{(1)}(\tau|x_1) + \frac{1}{2}\Delta_n^{(2)}(\tau|x_1),
\end{aligned}$$

where the quantities $\Delta_n^{(1)}(\tau|x_1)$ and $\Delta_n^{(2)}(\tau|x_1)$ are defined by

$$\begin{aligned}
\Delta_n^{(1)}(\tau|x_1) &= -\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}K_{b_n}\Big(F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big) - \tau\Big)\Big(\widehat{F}_1\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big)\\
&\quad - F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big)\Big)\\
\Delta_n^{(2)}(\tau|x_1) &= -\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\Big(K_{b_n}\Big(\xi_i - \tau\Big) - K_{b_n}\Big(F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big) - \tau\Big)\Big)\\
&\quad \times\Big(\widehat{F}_1\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big) - F\Big(\frac{i}{n}|(x_1, X_{j\underline{1}})\Big)\Big)
\end{aligned}$$

and the random variables $\xi_i = \xi_i(\tau, x_1, X_{j\underline{1}})$ satisfy $|\xi_i - F\left(\frac{i}{n}|(x_1, X_{j\underline{1}})\right)| \leq |\widehat{F}_1\left(\frac{i}{n}|(x_1, X_{j\underline{1}})\right) - F\left(\frac{i}{n}|(x_1, X_{j\underline{1}})\right)|$ $(i = 1, \ldots, n)$. Observing the Lipschitz continuity of the kernel$K$ it follows with the notation $\mathcal{D}_n = \{\sup_{y,x} |\widehat{F}_1(y|x) - F(y|x)| \leq b_n\}$ that

$$\begin{aligned}
\sup_{x_1} |\Delta_n^{(2)}(\tau|x_1)| &\leq \sup_{x,y} 2\left|\widehat{F}_1(y|x) - F(y|x)\right|^2 \sup_{x_1} \frac{1}{n^2 b_n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} \\
&\quad I\left\{|F\left(\frac{i}{n}|(x_1, X_{j\underline{1}})\right) - \tau| \leq 2b_n\right\} \\
&\leq \sup_{x,y} 2\left|\widehat{F}_1(y|x) - F(y|x)\right|^2 \sup_{x_1} \frac{1}{nb_n^2} \sum_{j=1}^{n} \int I\left\{|F\left(u|(x_1, X_{j\underline{1}})\right) \right. \\
&\quad \left. - \tau| \leq 2b_n\right\} du (1 + o(1)) \\
&\leq \sup_{x,y} C\left|\widehat{F}_1(y|x) - F(y|x)\right|^2 \frac{1}{b_n} (1 + o(1)) \\
&\leq \sup_{x,y} C\left|\widehat{F}_1(y|x) - F(y|x)\right| (1 + o(1))
\end{aligned}$$

on the set $\mathcal{D}_n$. For the term $\Delta_n^{(1)}(\tau|x_1)$ we have

$$\begin{aligned}
\sup_{x_1} \Delta_n^{(1)}(\tau|x_1) &= -\sup_{x_1} \frac{1}{n} \sum_{j=1}^{n} \int_0^1 K_{b_n}\left(F(t|x_1, X_{j\underline{1}}) - \tau\right) \left(\widehat{F}_1\left(t|(x_1, X_{j\underline{1}})\right)\right. \\
&\quad \left. -F\left(t|(x_1, X_{j\underline{1}})\right)\right) dt (1 + o(1)) \\
&\leq \sup_{x,y} C\left|\widehat{F}_1(y|x) - F(y|x)\right| (1 + o(1)),
\end{aligned}$$

and therefore we have for sufficiently large n

$$\begin{aligned}
P\left(B_{n1}^{(1)} > C_n\right) &= P\left(B_{n1}^{(1)} > C_n \middle| \mathcal{D}_n\right) P\left(\mathcal{D}_n\right) + P\left(B_{n1}^{(1)} > C_n \middle| \mathcal{D}_n^c\right) P\left(\mathcal{D}_n^c\right) \\
&\leq P\left(\sup_{x,y} C\left|\widehat{F}_1(y|x) - F(y|x)\right| > C_n\right) + P\left(\sup_{y,x} \left|\widehat{F}_1(y|x) - F(y|x)\right| > b_n\right) \\
&\leq 2P\left(\sup_{x,y} C\left|\widehat{F}_1(y|x) - F(y|x)\right| > C_n\right) \qquad (47)
\end{aligned}$$

(note that $C_n = o(b_n)$). Introducing the following notations

$$\begin{aligned}
\widehat{h}(x, y) &= \frac{1}{n} \sum_{k=1}^{n} K_{1,h_1}(x_1 - X_{k1}) K_{2,h_2}\left(x_{\underline{1}} - X_{k\underline{1}}\right) 1\{Y_k \leq y\} \\
\widehat{f}(x) &= \frac{1}{n} \sum_{k=1}^{n} K_{1,h_1}(x_1 - X_{k1}) K_{2,h_2}\left(x_{\underline{1}} - X_{k\underline{1}}\right) \\
\widehat{h}^{-i}(x, y) &= \frac{1}{n} \sum_{k\neq i}^{n} K_{1,h_1}(x_1 - X_{k1}) K_{2,h_2}\left(x_{\underline{1}} - X_{k\underline{1}}\right) 1\{Y_k \leq y\}
\end{aligned}$$

$$\widehat{f}^{-i}(x) = \frac{1}{n}\sum_{k\neq i}^{n} K_{1,h_1}(x_1 - X_{k1})K_{2,h_2}\big(x_{\underline{1}} - X_{k\underline{1}}\big)$$
$$h(x, y) = F(y|x)f(x)$$

straightforward calculations yield

$$\left|\widehat{F}_1(y|x) - F(y|x)\right| \leq C_{n1}(x, y) + C_{n2}(x, y) \tag{48}$$

where

$$C_{n1}(x, y) = \left|\frac{(\widehat{h}(x, y) - h(x, y))}{\widehat{f}(x)}\right|$$
$$C_{n2}(x, y) = \left|\frac{h(x, y)(\widehat{f}(x) - f(x))}{\widehat{f}(x)f(x)}\right|.$$

Using the notation $\mathcal{E}_n = \{\sup_x |\widehat{f}(x) - f(x)| \leq \delta\}$ we have for the first term of the right-hand side of (48) (where $\delta > 0$ is chosen sufficiently small)

$$\begin{aligned} P(\sup_{x_1,y} C_{n1}(x_1, y) > C_n) =& P\left(\sup_{x_1,y} C_{n1}(x_1, y) > C_n\middle|\mathcal{E}_n\right)P(\mathcal{E}_n) + P\Big(\sup_{x_1,y} C_{n1}(x_1, y) \\ & > C_n\Big|\mathcal{E}_n^c\Big)P(\mathcal{E}_n^c) \\ \leq& P\left(\sup_{x,y} C\left|(\widehat{h}(x, y) - h(x, y))\right| > C_n\middle|\, \mathcal{E}_n\right)P(\mathcal{E}_n) + P(\mathcal{E}_n^c) \\ \leq& P\left(\sup_{x,y} C\left|(\widehat{h}(x, y) - h(x, y))\right| > C_n\right) + P(\mathcal{E}_n^c) \end{aligned}$$

and with similar arguments one can show

$$P\left(\sup_{x_1,y} C_{n2}(x_1, y) > C_n\right) \leq P\left(\sup_x C\left|\widehat{f}(x) - f(x)\right| > C_n\right) + P(\mathcal{E}_n^c).$$

Recalling (47) and combining these estimates we obtain

$$\begin{aligned} P\left(B_{n1}^{(1)} > C_n\right) \leq& 6P\left(C\sup_x|\widehat{f}(x) - f(x)| > C_n\right) \\ &+ 2P\left(C\sup_{x,y}|\widehat{h}(x, y) - h(x, y)| > C_n\right). \end{aligned} \tag{49}$$

For the first probability on the right-hand side of (49) we have that

$$\begin{aligned} P\left(C\sup_x\left|\widehat{f}(x) - f(x)\right| > C_n\right) \leq& P\left(2C\sup_x\left|\widehat{f}(x) - E\big[\widehat{f}(x)\big]\right| > C_n\right) \\ &+ P\left(2C\sup_x\left|E[\widehat{f}(x)] - f(x)\right| > C_n\right). \end{aligned} \tag{50}$$

The second term of the right-hand side of (50) is of order $p(n)\exp(-n^a)$ which can be shown by calculating the expectation and a Taylor expansion. For the first term, we use Lemma 22 from Nolan and Pollard (1987) and Assumption 2 and obtain that the class

$$\mathcal{G} = \left\{ K_1\left(\frac{\cdot - x_1}{a}\right) K_2\left(\frac{\cdot - x_{\underline{1}}}{b}\right) \middle| x \in [0,1]^d, a, b \in \mathbb{R} \setminus \{0\} \right\}$$

is Euclidean. Furthermore we have that $\mathcal{G}_n \subset \mathcal{G}$ for

$$\mathcal{G}_n = \left\{ K_1\left(\frac{\cdot - x_1}{h_1}\right) K_2\left(\frac{\cdot - x_{\underline{1}}}{h_2}\right) \middle| x \in [0,1]^d \right\}$$

and therefore the classes $\mathcal{G}_n$ are Euclidean with the same constants as $\mathcal{G}$. Now with

$$\sigma^2_{\mathcal{G}_n} = \|E[g - E[g]]^2\|_{\mathcal{G}_n} \le C h_1 h_2^{d-1},$$

Theorem 2.14.16 of van der Vaart and Wellner (1996) yields

$$\begin{aligned} &P\left(C \sup_x \left|\widehat{f}(x) - E\left[\widehat{f}(x)\right]\right| > C_n\right) \\ &\le O(p(n)) \exp\left(-\frac{1}{2} \frac{\tilde{K} C_n^2 n h_1 h_2^{d-1}}{K + \frac{3}{\sqrt{n h_1^2 h_2^{2(d-1)}}} + C_n}\right) = O\left(p(n) \exp\left(-n^{2\alpha}\right)\right) \end{aligned}$$

where $p(n)$ is a polynomial. The second term in (49) can be treated with the same arguments. For a proof of (43) it remains to consider the term $B_{n1}^{(2)}$ defined in (45) (the cases $k = 2, \ldots, d$ are treated in exactly the same way). We have

$$\begin{aligned} &\left| \frac{1}{n} \sum_{i=1}^n \widehat{q}_k(\tau|X_{ik}) - c(\tau) \right| \\ &= \left| \frac{1}{n} \sum_{i=1}^n \widehat{q}_k(\tau|X_{ik}) - q_k(\tau|X_{ik}) + q_k(\tau|X_{ik}) - c(\tau) \right| \\ &\le \frac{1}{n} \sum_{i=1}^n |\widehat{q}_k(\tau|X_{ik}) - q_k(\tau|X_{ik})| + \left| \frac{1}{n} \sum_{i=1}^n q_k(\tau|X_{ik}) - c(\tau) \right| \\ &\le \sup_{x_k} |\widehat{q}_k(\tau|x_k) - q_k(\tau|x_k)| + \left| \frac{1}{n} \sum_{i=1}^n q_k(\tau|X_{ik}) - E[q_k(\tau|X_{ik})] \right| \end{aligned}$$

and the assertion follows from what we have shown before and the Markov inequality. Next we consider the proof of (44). Therefore, we consider the decomposition

$$\sup_x \max_{i=1}^n \left|\widehat{Q}(\tau|x) - \widehat{Q}^{-i}(\tau|x)\right| \leq \sum_{k=1}^d \left\{ D_{nk}^{(1)} + \left(1 - \frac{1}{d}\right) D_{nk}^{(2)} \right\}$$

where

$$D_{nk}^{(1)} = \sup_{x_k} \max_{i=1}^n \left|\widehat{q}_k(\tau|x_k) - \widehat{q}_k^{-i}(\tau|x_k)\right|$$
$$D_{nk}^{(2)} = \max_{i=1}^n \left|\frac{1}{n}\sum_{j=1}^n \widehat{q}_k(\tau|X_{jk}) - \frac{1}{n-1}\sum_{j\neq i} \widehat{q}_k^{-i}(\tau|X_{jk})\right|.$$

Considering term $D_{n1}^{(1)}$ (all other terms in the first sum are treated similarly) we obtain by similar arguments for sufficiently large n

$$\begin{aligned} P\left(D_{n1}^{(1)} > C_n\right) \leq & C\Bigg(P(\mathcal{D}_n^c) + P(\mathcal{E}_n^c) + P\Bigg(\max_i \frac{1}{n}\sum_{j=1}^n \Bigg|\frac{|K_{2,h_2}(X_{\underline{j1}} - X_{i\underline{1}})|}{\int |K_2(u)|du} \\ & - f_{\underline{1}}(X_{\underline{i1}})\Bigg| > \delta\Bigg)\Bigg) \\ = & O\left(p(n)\exp(-n^\alpha)\right). \end{aligned}$$

For terms of the form $D_{nk}^{(2)}$ we use the estimate

$$\begin{aligned} & \max_{i=1}^n \left|\frac{1}{n}\sum_{j=1}^n \widehat{q}_k(\tau|X_{jk}) - \frac{1}{n-1}\sum_{j\neq i}\widehat{q}_k^{-i}(\tau|X_{jk})\right| \\ & \leq \max_{i=1}^n \left|\frac{1}{n}\sum_{j=1}^n \widehat{q}_k(\tau|X_{jk}) - \frac{1}{n-1}\sum_{j\neq i}\widehat{q}_k(\tau|X_{jk})\right| + \max_{i=1}^n \Bigg|\frac{1}{n-1}\sum_{j\neq i}(\widehat{q}_k(\tau|X_{jk}) \\ & \quad - \widehat{q}_k^{-i}(\tau|X_{jk}))\Bigg| \leq \max_{i=1}^n \Bigg|\frac{1}{n(n-1)}\sum_{j=1}^n \widehat{q}_k(\tau|X_{jk}) \\ & + \frac{1}{n-1}\widehat{q}_k(\tau|X_{ik})\Bigg| + \sup_{x_k}\max_{i=1}^n \left|\widehat{q}_k(\tau|x_k) - \widehat{q}_k^{-i}(\tau|x_k)\right| \\ & \leq \sup_{x_k} 2\left(\frac{1}{n-1}\left|\widehat{q}_k(\tau|x_k) - q_k(\tau|x_k)\right| + \frac{1}{n-1}\sup_{x_k}|q_k(\tau|x_k)|\right) + \sup_{x_k}\max_{i=1}^n |\widehat{q}_k(\tau|x_k) \\ & \quad - \widehat{q}_k^{-i}(\tau|x_k)| \end{aligned}$$

and the assertion of Lemma 1 follows by the same arguments as before. □

Now we prove assertion (39) for the term T_{2n}. Recalling its definition in (36) we have

$$T_{2n} = T_{2n}^{(1)} + T_{2n}^{(2)},$$

where the terms $T_{2n}^{(i)},\ i = 1, 2$ are given by

$$T_{2n}^{(1)} = \frac{2}{n(n-1)} \sum_i \sum_{j \neq i} L_g\left(X_i - X_j\right)\left(\widehat{R}_i - \widehat{R}_i^{-j}\right) R_j \pi(X_i)\pi(X_j)$$
$$T_{2n}^{(2)} = \frac{2}{n(n-1)} \sum_i \sum_{j \neq i} L_g\left(X_i - X_j\right)\left(\widehat{R}_i^{-j} - R_i\right) R_j \pi(X_i)\pi(X_j)$$

with $\hat{R}_i^{-j} = I\{Y_i \leq \hat{Q}_{\text{add}}^{-i,j}(\tau|X_i)\} - \tau$. Now the random variable $T_{2n}^{(1)}$ can be treated with the same arguments as the term T_{3n} and we get $ng^{d/2}T_{2n}^{(1)} = O(ng^{d/2}D_n) = o(1)$ (in L_1 and thus in probability), where the last equality follows by Assumption 5. For the second term, $T_{2n}^{(2)}$, we have that $E[T_{2n}^{(2)}] = 0$ and

$$\left(T_{2n}^{(2)}\right)^2 = U_1 + U_2$$

where

$$U_1 = \frac{4}{n^2(n-1)^2} \sum_{\substack{i_1, i_2, j \\ j \neq i_1, j \neq i_2}} L_g\left(X_{i_1} - X_j\right) L_g\left(X_{i_2} - X_j\right)\left(\widehat{R}_{i_1}^{-j} - R_{i_1}\right)\left(\widehat{R}_{i_2}^{-j} - R_{i_2}\right)$$
$$R_j^2 \pi(X_{i_1})\pi^2(X_j)\pi(X_{i_2})$$
$$U_2 = \frac{4}{n^2(n-1)^2} \sum_{\substack{i_1, i_2, j_1, j_2 \\ j_1 \neq i_1, j_2 \neq i_2, j_1 \neq j_2}} L_g\left(X_{i_1} - X_{j_1}\right) L_g\left(X_{i_2} - X_{j_2}\right)\left(\widehat{R}_{i_1}^{-j_1} - \widehat{R}_{i_1}^{-j_1, j_2}\right.$$
$$\left. + \widehat{R}_{i_1}^{-j_1, j_2} - R_{i_1}\right) R_{j_1}$$
$$\times \left(\widehat{R}_{i_2}^{-j_2} - \widehat{R}_{i_2}^{-j_2, j_2} + \widehat{R}_{i_2}^{-j_1, j_2} - R_{i_2}\right) R_{j_2} \pi(X_{i_1})\pi(X_{j_1})\pi(X_{i_2})\pi(X_{j_2})$$

with $\hat{R}_i^{-j,k} = I\{Y_i \leq \hat{Q}_{\text{add}}^{-i,j,k}(\tau|X_i)\} - \tau$. For the second term, one obtains $E[U_2] = E[\tilde{U}_2]$, where

$$\tilde{U}_2 = \frac{4}{n^2(n-1)^2} \sum_{i_1, i_2, j_1, j_2}^{\neq} L_g\left(X_{i_1} - X_{j_1}\right) L_g\left(X_{i_2} - X_{j_2}\right)\left(\widehat{R}_{i_1}^{-j_1} - \widehat{R}_{i_1}^{-j_1, j_2}\right)$$
$$R_{j_1}\left(\widehat{R}_{i_2}^{-j_2} - \widehat{R}_{i_2}^{-j_1, j_2}\right) R_{j_2} \pi(X_{i_1})\pi(X_{j_1})\pi(X_{i_2})\pi(X_{j_2})$$

$$+\frac{8}{n^2(n-1)^2}\sum_{i_1,i_2,j}^{\neq} L_g\left(X_{i_1}-X_{i_2}\right)L_g\left(X_{i_2}-X_j\right)\left(\widehat{R}_{i_1}^{-i_2}-\widehat{R}_{i_1}^{-i_2,j}\right)R_{i_2}$$
$$\left(\widehat{R}_{i_2}^{-j}-R_{i_2}\right)R_j\pi(X_{i_1})\pi^2(X_{i_2})\pi(X_j)$$

and $\sum^{\neq}$ denotes a sum where all indices are distinct. Similarly to the treatment of the term T_{3n} it can be shown that $|E[\tilde{U}_2]| \leq E[|\tilde{U}_2|] = O(D_n^2) + O(C_n D_n/n)$ applying Assumption 5. The same assumption yields analogously that $|E[U_1]| \leq E[|U_1|] = O(C_n^2/n + C_n/(n^2 g^d))$. Altogether we have $E[(T_{2n}^{(2)})^2] \leq E[|U_1|] + E[|\tilde{U}_2|] = o((ng^{d/2})^{-2})$ by the bandwidth conditions. We obtain that $ng^{d/2}T_{2n}^{(2)} = o(1)$ in L_2 and thus in probability, which completes the proof of (39).

Appendix B: Proof of Theorem 2

Recall the definition of C_n and D_n in (40) and consider the decomposition

$$T_n = (T_{1n} + 2T_{2n} + T_{3n}) + (-2T_{4n} - 2T_{5n} + T_{6n}) \tag{51}$$

where T_{1n} is defined in (35), the statistics $T_{jn}(j = 2, \ldots, 6)$ are given by

$$T_{2n} = \frac{1}{n(n-1)}\sum_{i\neq j} L_g\left(X_i - X_j\right) R_i\left(\widehat{R}_j - R_j^{\text{add}}\right)\pi(X_i)\pi(X_j)$$
$$T_{3n} = \frac{1}{n(n-1)}\sum_{i\neq j} L_g\left(X_i - X_j\right)\left(\widehat{R}_i - R_i^{\text{add}}\right)\left(\widehat{R}_j - R_j^{\text{add}}\right)\pi(X_i)\pi(X_j)$$
$$T_{4n} = \frac{1}{n(n-1)}\sum_{i\neq j} L_g\left(X_i - X_j\right) R_i\left(R_j - R_j^{\text{add}}\right)\pi(X_i)\pi(X_j)$$
$$T_{5n} = \frac{1}{n(n-1)}\sum_{i\neq j} L_g\left(X_i - X_j\right)\left(\widehat{R}_i - R_i^{\text{add}}\right)\left(R_j - R_j^{\text{add}}\right)\pi(X_i)\pi(X_j)$$
$$T_{6n} = \frac{1}{n(n-1)}\sum_{i\neq j} L_g\left(X_i - X_j\right)\left(R_i - R_i^{\text{add}}\right)\left(R_j - R_j^{\text{add}}\right)\pi(X_i)\pi(X_j)$$

and R_i, $\widehat{R}_i$ and R_i^{add} are defined in (12), (3) and (13), respectively. Observing the proofs of (38) and (39), respectively, we have that under the local alternatives of the form (18)

$$ng^{\frac{d}{2}}T_{1n} \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2); \quad T_{jn} = o\left(\frac{1}{ng^{\frac{d}{2}}}\right); \quad j = 2, 3 \tag{52}$$

in L_1, and it remains to investigate the terms T_{4n}, T_{5n} and T_{6n} in the decomposition (51). First, we study the statistic T_{4n} for which we have that $E[T_{4n}] = 0$ and

$$E\left[T_{4n}^2\right] = \frac{1}{n^2(n-1)^2}\sum_{i_1=1}^{n}\sum_{j_1\neq i_1}^{n}\sum_{i_2=1}^{n}\sum_{j_2\neq i_2}^{n} E\Big[L_g(X_{i_1}-X_{j_1})L_g(X_{i_2}-X_{j_2})R_{i_1} \\ \big(R_{j_1}-R_{j_1}^{\text{add}}\big)R_{i_2}\big(R_{j_2}-R_{j_2}^{\text{add}}\big)\pi(X_{i_1})\pi(X_{j_1})\pi(X_{i_2})\pi(X_{j_2})\Big],$$

where the expectations in this sum vanish whenever $j_2 \neq i_1 \neq i_2$ or $i_1 \neq i_2 \neq j_1$. Considering the case where $i_1 = i_2,\ j_1 \neq j_2$ we obtain by a Taylor expansion for some constant λ (conditioning on X_{i1}, X_{j1}, X_{j2} and Y_{i1})

$$\begin{aligned}
&E\Big[L_g(X_{i_1}-X_{j_1})L_g(X_{i_1}-X_{j_2})R_{i_1}^2\big(R_{j_1}-R_{j_1}^{\text{add}}\big)\big(R_{j_2}-R_{j_2}^{\text{add}}\big)\pi^2(X_{i_1})\pi(X_{j_1})\pi(X_{j_2})\Big]\\
&= E\Big[L_g(X_{i_1}-X_{j_1})L_g(X_{i_1}-X_{j_2})R_{i_1}^2 E\Big[R_{j_1}-R_{j_1}^{\text{add}}|X_{j_1}\Big]E\Big[R_{j_2}\\
&\quad -R_{j_2}^{\text{add}}|X_{j_2}\Big]\pi^2(X_{i_1})\pi(X_{j_1})\pi(X_{j_2})\Big]\\
&= E\Big[L_g(X_{i_1}-X_{j_1})L_g(X_{i_1}-X_{j_2})R_{i_1}^2(F(Q(\tau|X_{j_1})|X_{j_1})-F(Q_{\text{add}}(\tau|X_{j_1})|X_{j_1}))\\
&\quad \times F(Q(\tau|X_{j_2})|X_{j_2})-F(Q_{\text{add}}(\tau|X_{j_2})|X_{j_2})\pi^2(X_{i_1})\pi(X_{j_1})\pi(X_{j_2})\Big]\\
&\leq \lambda d_n^2 E\left[L_g(X_{i_1}-X_{j_1})L_g(X_{i_1}-X_{j_2})\right] = O\left(d_n^2\right).
\end{aligned}$$

The other cases can be treated with similar arguments and we obtain

$$\begin{aligned}
&E\left[L_g(X_{i_1}-X_{j_1})^2R_{i_1}^2\big(R_{j_1}-R_{j_1}^{\text{add}}\big)^2\right] = O\left(\frac{d_n}{g^d}\right)\\
&E\left[L_g(X_{i_1}-X_{j_1})^2R_{i_1}R_{j_1}(R_{j_1}-R_{j_1}^{\text{add}})\big(R_{i_1}-R_{i_1}^{\text{add}}\big)\right] = O\left(\frac{d_n^2}{g^d}\right).
\end{aligned}$$

Combining these estimates we have

$$ng^{\frac{d}{2}}T_{4n} = o_{L_1}(1). \tag{53}$$

The statistic T_{5n} can be treated with the same arguments as the term T_{3n} under the null hypothesis and it follows

$$ng^{\frac{d}{2}}T_{5n} = O_p\left(ng^{\frac{d}{2}}d_nC_n\right) = o_{L_1}(1). \tag{54}$$

Finally, we study the remaining term T_{6n} for which a straightforward calculation yields

$$\begin{aligned} E\big[ng^{\frac{d}{2}}T_{6n}\big] &= ng^{\frac{d}{2}}E\big[L_g(X_1-X_2)(R_1-R_1^{\text{add}})(R_2-R_2^{\text{add}})\pi(X_1)\pi(X_2)\big] \\ &= ng^{\frac{d}{2}}E\big[L_g(X_1-X_2)(F(Q_{\text{add}}(\tau|X_1)+d_n l(X_1)|X_1) \\ &\quad - F(Q_{\text{add}}(\tau|X_1)|X_1))(F(Q_{\text{add}}(\tau|X_2)+d_n l(X_2)|X_2) \\ &\quad\quad - F(Q_{\text{add}}(\tau|X_2)|X_2))\pi(X_1)\pi(X_2)\big] \\ &= E\big[(F'(Q_{\text{add}}(\tau|X_1)|X_1)l(X_1)\pi^2(X_1))f(X_1)\big]+o(1) \end{aligned} \tag{55}$$

and

$$E\left[(T_{6n}-E[T_{6n}])^2\right] = o\left(\frac{1}{ng^{\frac{d}{2}}}\right). \tag{56}$$

Thus (19) follows from (51, 52, 53, 54, 55, 56).

Appendix C: Proof of Theorem 3

For a proof of Theorem 3 we assume for a transparent notation $d=2$. The general case follows by exactly the same arguments. Recall the decomposition (51). Observing the proof of Theorem 1 we have

$$T_{jn} = o\left(\frac{1}{\sqrt{n}}\right); \quad j=1,2,3,$$

in L_1. Therefore, we obtain $E[T_n] = -2E[T_{4n}] - 2E[T_{5n}] + E[T_{6n}] + o(1/\sqrt{n})$ and

$$\begin{aligned} \sqrt{n}\Big(T_n - E[T_n]\Big) &= -2\sqrt{n}\Big(T_{4n}-E[T_{4n}]\Big) - 2\sqrt{n}(T_{5n}-E[T_{5n}]) \\ &\quad + \sqrt{n}\Big(T_{6n}-E[T_{6n}]\Big) + o_{L_1}(1) \end{aligned}$$

and it remains to investigate the statistics T_{4n}, T_{5n} and T_{6n}. We first study the term T_{4n} for which we have the stochastic expansion

$$\begin{aligned} T_{4n} &= \sum_{i=1}^n R_i E\big[T_{4n}^{(i)}|X_i\big]\pi(X_i) + o_{L_1}\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{n}\sum_{i=1}^n R_i\Delta(X_i)\pi^2(X_i)f(X_i) \\ &\quad + o_{L_1}\left(\frac{1}{\sqrt{n}}\right) \end{aligned}$$

where

$$T_{4n}^{(i)} = \frac{1}{n(n-1)}\sum_{j=1, j\neq i}^n L_g(X_i-X_j)\big(R_j-R_j^{\text{add}}\big)\pi(X_j)$$

and $\Delta(X_j)$ is defined in (14). A corresponding stochastic expansion for the term T_{5n} requires substantially more effort. More precisely, we have the following result, which is proved below.

Lemma 2 *Under the assumptions of Theorem 3 we have*

$$T_{5n} = \sum_{j=1}^{10} Z_n^{(j)} + o\left(\frac{1}{\sqrt{n}}\right) \tag{57}$$

in L_1 where the terms $Z_n^{(j)}$ in this stochastic expansion are defined by

$$\begin{aligned} Z_n^{(1)} =& E\Big[L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)K_{1,h_1}(X_{i1} - X_{l1})K_{2,h_2}(X_{k2} - X_{l2}) \\ &\times \left(\frac{F(Q(\tau|(X_{i1}, X_{k2}))|(X_{l1}, X_{l2})) - F(Q(\tau|(X_{i1}, X_{k2}))|(X_{i1}, X_{k2}))}{f(X_{i1}, X_{k2})F'(Q(\tau|(X_{i1}, X_{k2}))|(X_{i1}, X_{k2}))}\right)\pi(X_i)\pi(X_j)\Big] \end{aligned} \tag{58}$$

$$Z_n^{(2)} = -\frac{1}{n}\sum_{l=1}^{n} R_l h_2(X_l) \tag{59}$$

$$\begin{aligned} Z_n^{(3)} =& \frac{1}{n}\sum_{k=1}^{n}\Big(E[f(X)F'(Q_{\text{add}}(\tau|X)|X)\Delta(X)\pi^2(X)Q(\tau|(X_1, X_{k2}))|X_k] \\ &- E[f(X)F'(Q_{\text{add}}(\tau|X)|X)\Delta(X)\pi^2(X)Q(\tau|(X_1, X_{k2}))]\Big) \end{aligned} \tag{60}$$

$$\begin{aligned} Z_n^{(4)} =& E\Big[L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)K_{1,h_1}(X_{k1} - X_{l1})K_{2,h_2}(X_{i2} - X_{l2}) \\ &\times \left(\frac{F(Q(\tau|(X_{k1}, X_{i2}))|(X_{l1}X_{l2})) - F(Q(\tau|(X_{k1}, X_{i2}))|(X_{k1}, X_{i2}))}{f(X_{k1}, X_{i2})F'(Q(\tau|(X_{k1}, X_{i2}))|(X_{k1}, X_{i2}))}\right)\pi(X_i)\pi(X_j)\Big] \end{aligned} \tag{61}$$

$$Z_n^{(5)} = -\frac{1}{n}\sum_{l=1}^{n} R_l h_5(X_l) \tag{62}$$

$$\begin{aligned} Z_n^{(6)} =& \frac{1}{n}\sum_{k=1}^{n}\Big(E[f(X)F'(Q_{\text{add}}(\tau|X)|X)\Delta(X)\pi^2(X)Q(\tau|(X_{k1}, X_2))|X_k] \\ &- E[f_X(X)F'(Q_{\text{add}}(\tau|X)|X)\Delta(X)\pi^2(X)Q(\tau|(X_{k1}, X_2))]\Big) \end{aligned} \tag{63}$$

$$\begin{aligned} Z_n^{(7)} =& E\Big[L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)K_{1,h_1}(X_{m1} - X_{l1})K_{2,h_2}(X_{k2} - X_{l2}) \\ &\times \left(\frac{F(Q(\tau|(X_{m1}, X_{k2}))|(X_{l1}, X_{l2})) - F(Q(\tau|(X_{m1}, X_{k2}))|(X_{m1}, X_{k2}))}{f(X_{m1}, X_{k2})F'(Q(\tau|(X_{m1}, X_{k2}))|(X_{m1}, X_{k2}))}\right)\pi(X_i)\pi(X_j)\Big] \end{aligned} \tag{64}$$

$$Z_n^{(8)} = \frac{1}{n}\sum_{l=1}^{n} R_l h_8(X_l) \tag{65}$$

$$Z_n^{(9)} = \frac{-1}{n}\sum_{k=1}^{n} E\left[f(X)F'(Q_{\text{add}}(\tau|X)|X)\Delta(X)\pi^2(X)(q_1(X_{k1}) + q_2(X_{k2}) - 2c(\tau))|X_k\right] \tag{66}$$

$$Z_n^{(10)} = E\Big[L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)K_{2,h_2}(X_{m1} - X_{l1})K_{1,h_1}(X_{k2} - X_{l2}) \times \left(\frac{F(Q(\tau|(X_{m1}, X_{k2}))|(X_{l1}, X_{l2})) - F(Q(\tau|(X_{m1}, X_{k2}))|(X_{m1}, X_{k2}))}{f(X_{m1}, X_{k2})F'(Q(\tau|(X_{m1}, X_{k2}))|(X_{m1}, X_{k2}))}\right)\pi(X_i)\pi(X_j)\Big] \tag{67}$$

where (X, Y) *are independent copies of* (X_i, Y_i) *and*

$$h_2(X_l) = \frac{f_2(X_{l2}) \int \Delta(X_{l1}, t_2)\pi^2(X_{l1}, t_2) f^2(X_{l1}, t_2) F'(Q_{\text{add}}(\tau|(X_{l1}, t_2))|(X_{l1}, t_2))dt_2}{f(X_l)F'(Q(\tau|X_l)|X_l)} \tag{68}$$

$$h_5(X_l) = \frac{f_1(X_{l1}) \int \Delta(t_1, X_{l2})\pi^2(t_1, X_{l2}) f^2(t_1, X_{l2}) F'(Q_{\text{add}}(\tau|(t_1, X_{l2}))|(t_1, X_{l2}))dt_1}{f(X_l)F'(Q(\tau|X_l)|X_l)} \tag{69}$$

$$h_8(X_l) = \frac{f_1(X_{l1}) f_2(X_{l2}) \int \Delta(t)\pi^2(t) f^2(t) F'(Q_{\text{add}}(\tau|t)|t)dt}{f(X_l)F'(Q(\tau|X_l)|X_l)}. \tag{70}$$

Next, we study the term T_{6n} using Lemma 3.1 in Zheng (1996) with the kernel $H(Z_i, Z_j) = L_g(X_i - X_j)(R_i - R_i^{\text{add}})(R_j - R_j^{\text{add}})\pi(X_i)\pi(X_j)$, where $Z_i = (X_i, Y_i)$. A straightforward calculation gives $E[(H(Z_1, Z_2))^2] = o(n)$, which yields the Hoeffding decomposition

$$T_{6n} - E[T_{6n}] = \frac{2}{n}\sum_{i=1}^{n} H_1(X_i) + o_{L_1}\left(\frac{1}{\sqrt{n}}\right),$$

where $H_1(x) = E[H(Z_1, Z_2)|X_1 = x] - E[H(Z_1, Z_2)]$ and

$$E[T_{6n}] = E[\Delta^2(X_1)\pi^2(X_1) f(X_1)] + O(g^2).$$

From Lemma 2, we have for the expectation of the statistic T_{5n}

$$E[T_{5n}] = E\left[Z_n^{(1)}\right] + E\left[Z_n^{(4)}\right] + E\left[Z_n^{(7)}\right] + E\left[Z_n^{(10)}\right] + o\left(\frac{1}{\sqrt{n}}\right)$$

where

$$\begin{aligned} E\left[Z_n^{(1)}\right] &= - E\left[\Delta(X_1)\pi^2(X_1)F'(Q_{\text{add}}(\tau|X_1)|X_1) f(X_1)b_1(X_{11})\right] h_1^2 \\ &\quad + o(h_1^2) + O(h_2^q) \\ E\left[Z_n^{(4)}\right] &= - E\left[\Delta(X_1)\pi^2(X_1)F'(Q_{\text{add}}(\tau|X_1)|X_1) f(X_1)b_2(X_{12})\right] h_1^2 \\ &\quad + o(h_1^2) + O(h_2^q) \end{aligned}$$

$$
\begin{aligned}
E\left[Z_n^{(7)}\right] =& \frac{1}{2} E\left[\Delta(X_1)\pi^2(X_1)F'(Q_{\text{add}}(\tau|X_1)|X_1)f(X_1)b_1(X_{21})\right] h_1^2 \\
&+ o(h_1^2) + O(h_2^q) \\
E\left[Z_n^{(10)}\right] =& \frac{1}{2} E\left[\Delta(X_1)\pi^2(X_1)F'(Q_{\text{add}}(\tau|X_1)|X_1)f(X_1)b_2(X_{22})\right] h_1^2 \\
&+ o(h_1^2) + O(h_2^q)
\end{aligned}
$$

and the bias b_α is defined in (21). Observing (51) it therefore follows that

$$
\begin{aligned}
E[T_n] =& E\left[\Delta^2(X_1)\pi^2(X_1)f(X_1)\right] \\
&+ 2E\Big[F'(Q_{\text{add}}(\tau|X_1)|X_1)\Delta(X_1)\pi^2(X_1)f(X_1)(b(X_1) \\
&- \frac{1}{2}b(X_2))\Big]h_1^2 + o(h_1^2) + O(g^2)
\end{aligned}
$$

which is the claimed representation in Theorem 3 for the case $d = 2$. With the same argument we obtain the stochastic expansion

$$
\sqrt{n}(T_n - E[T_n]) = A_n + B_n + C_n + o_{L_1}(1),
$$

where the quantities A_n, B_n and C_n are given by

$$
\begin{aligned}
A_n =& \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \left(\Delta^2(X_i)\pi^2(X_i)f(X_i) - E[\Delta^2(X_i)\pi^2(X_i)f(X_i)]\right) \\
=& \sqrt{n}\left(T_{6n} - E[T_{6n}]\right) + o_{L_1}\left(\frac{1}{\sqrt{n}}\right), \\
B_n =& \frac{2}{\sqrt{n}} \sum_{i=1}^{n} E\Big[\Delta(X_j)\pi^2(X_j)f(X_j)F'(Q_{\text{add}}(\tau|X_j)|X_j)\Big(\frac{1}{2}(Q(\tau|X_{i1}, X_{l2}) \\
&+ Q(\tau|X_{l1}, X_{i2}) \\
&+ Q(\tau|X_{l1}, X_{i2}) + Q(\tau|X_{i1}, X_{l2})) - Q(\tau|X_{j1}, X_{i2}) - Q(\tau|X_{i1}, X_{j2})\Big)|X_i\Big] \\
&- E\Big[\Delta(X_j)\pi^2(X_j)f(X_j)F'(Q_{\text{add}}(\tau|X_j)|X_j)(2Q(\tau|X_i) - Q(\tau|X_{j1}, X_{i2}) \\
&- Q(\tau|X_{i1}, X_{j2}))\Big] \\
=& -2\sqrt{n}\left(Z_n^{(3)} + Z_n^{(6)} + Z_n^{(9)}\right), \\
C_n =& \frac{2}{\sqrt{n}} \sum_{i=1}^{n} R_i\left(-\Delta(X_i)\pi^2(X_i)f(X_i) + h_2(X_i) + h_5(X_i) - h_8(X_i)\right) \\
=& -2\sqrt{n}\left(T_{4n} + Z_n^{(2)} + Z_n^{(5)} + 2Z_n^{(8)}\right) + o_{L_1}\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

and h_2, h_5 and h_8 are defined in (68), (69) and (70), respectively. Therefore, asymptotic normality is a direct consequence of Lyapunov's central limit theorem. Finally, a straightforward calculation yields

$$\begin{aligned}
&\operatorname{Var}(A_n + B_n)\\
&\quad = 4\operatorname{Var}\Big[\Delta^2(X_1)\pi^2(X_1) f(X_1)\\
&\qquad - E\Big[\Delta(X_2)\pi^2(X_2) f(X_2) F'(Q_{\text{add}}(\tau|X_2)|X_2)\\
&\qquad \times \Big(Q(\tau|X_{11}, X_{22}) + Q(\tau|X_{21}, X_{12}) - \frac{1}{2}\big(Q(\tau|X_{11}, X_{32}) + Q(\tau|X_{31}, X_{12})\\
&\qquad + Q(\tau|X_{31}, X_{12}) + Q(\tau|X_{11}, X_{32})\big)\Big)\Big| X_1\Big]\Big],\\
&\operatorname{Var}(C_n) = 4E\Big[\tau(1-\tau)\Big(-\Delta(X_1)\pi^2(X_1) f(X_1)\\
&\quad + \frac{f_2(X_{12}) \int \Delta(X_{11}, t_2)\pi^2(X_{11}, t_2) f^2(X_{11}, t_2) F'(Q_{\text{add}}(\tau|X_{11}, t_2)|X_{11}, t_2) dt_2}{f(X_1) F'(Q(\tau|X_1)|X_1)}\\
&\quad + \frac{f_1(X_{11})\Delta(t_1, X_{12})\pi^2(t_1, X_{12}) f^2(t_1, X_{12}) F'(Q_{\text{add}}(\tau|t_1, X_{12})|t_1, X_{12}) dt_1}{f(X_1) F'(Q(\tau|X_1)|X_1)}\\
&\quad - \frac{\int \Delta(t)\pi^2(t) f^2(t) F'(Q_{\text{add}}(\tau|t)|t) dt}{F'(Q(\tau|X_1)|X_1)}\Big)^2\Big]
\end{aligned}$$

and $\operatorname{Cov}(A_n + B_n, C_n) = 0$, which completes the proof of Theorem 3.

Proof of Lemma 2 Observe the decomposition $T_{5n} = \widetilde{T}_{5n}^{(1)} + \widetilde{T}_{5n}^{(2)}$, where

$$\begin{aligned}
\widetilde{T}_{5n}^{(1)} &= \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right)\left(\widehat{R}_i - R_i^{\text{add}}\right)\left(R_j - R_j^{\text{add}} - E[R_j\right.\\
&\qquad \left. - R_j^{\text{add}}|X_j]\right)\pi(X_i)\pi(X_j)\\
\widetilde{T}_{5n}^{(2)} &= \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right)\left(\widehat{R}_i - R_i^{\text{add}}\right) E\Big[R_j - R_j^{\text{add}}|X_j\Big]\pi(X_i)\pi(X_j).
\end{aligned}$$

We calculate

$$\begin{aligned}
E\Big[\Big(\widetilde{T}_{5n}^{(1)}\Big)^2\Big] = \frac{1}{n^2(n-1)^2} \sum_{i_1 \neq j_1} \sum_{i_2 \neq j_2} & E\Big[L_g\left(X_{i_1} - X_{j_1}\right) L_g\left(X_{i_2} - X_{j_2}\right)\\
&\Big(\widehat{R}_{i_1} - R_{i_1}^{\text{add}}\Big)\Big(\widehat{R}_{i_2} - R_{i_2}^{\text{add}}\Big)\\
&\times \Big(R_{j_1} - R_{j_1}^{\text{add}} - E\Big[R_{j_1} - R_{j_1}^{\text{add}}|X_{j_1}\Big]\Big)\Big(R_{j_2} - R_{j_2}^{\text{add}}\\
&- E\Big[R_{j_2} - R_{j_2}^{\text{add}}|X_{j_2}\Big]\Big)\\
&\times \pi(X_{i_1})\pi(X_{j_1})\pi(X_{i_2})\pi(X_{j_2})\Big]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{(1+o(1))}{n^2(n-1)^2} \sum_{i_1 \neq j_1} \sum_{i_2 \neq j_2} E\Big[L_g\left(X_{i_1} - X_{j_1}\right) L_g\left(X_{i_2} - X_{j_2}\right) \\
&\quad \left(\widehat{R}_{i_1}^{-j_1} - R_{i_1}^{\text{add}}\right)\left(\widehat{R}_{i_2}^{-j_1} - R_{i_2}^{\text{add}}\right) \\
&\quad \times \left(R_{j_1} - R_{j_1}^{\text{add}} - E[R_{j_1} - R_{j_1}^{\text{add}} | X_{j_1}]\right)\Big(R_{j_2} - R_{j_2}^{\text{add}} - E[R_{j_2} \\
&\quad - R_{j_2}^{\text{add}} | X_{j_2}]\Big) \\
&\quad \times \pi(X_{i_1})\pi(X_{j_1})\pi(X_{i_2})\pi(X_{j_2})\Big] \\
&= o\left(\frac{1}{n}\right),
\end{aligned}
$$

where the last estimate follows by similar arguments as given for the term T_{3n} under the null hypothesis (see Appendix A). With similar arguments we obtain

$$
\begin{aligned}
\widetilde{T}_{5n}^{(2)} &= \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) \left(F(\widehat{Q}_{\text{add}}^{-i}(\tau|X_i)|X_i) - F(Q(\tau|X_i)|X_i)\right) \\
&\quad \times E[R_j - R_j^{\text{add}} | X_j]\pi(X_i)\pi(X_j) + o_{L_1}\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}
$$

and therefore a taylor expansion and Lemma 1 yield

$$
\begin{aligned}
T_{5n} &= \widetilde{T}_{5n}^{(2)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right) \\
&= T_{5n}^{(1)} + T_{5n}^{(2)} + T_{5n}^{(3)} + T_{5n}^{(4)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where we introduce the notation

$$
\begin{aligned}
T_{5n}^{(\ell)} &= \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j) \\
&\quad \times (\widehat{q}_\ell^{-i}(\tau|X_{i\ell}) - q_\ell(\tau|X_{i\ell})); \quad \ell = 1, 2 \\
T_{5n}^{(\ell)} &= -\frac{1}{2n^2(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) F'\left(Q_{\text{add}}(\tau|X_i)|X_i\right) \Delta(X_j)\pi(X_i)\pi(X_j) \\
&\quad \times \sum_{k=1}^{n} \left(\widehat{q}_{\ell-2}^{-i}(\tau|X_{k(\ell-2)}) - c(\tau)\right); \quad \ell = 3, 4
\end{aligned}
$$

and we treat the terms $T_{5n}^{(\ell)}$ for $\ell = 1, \ldots, 4$ separately. Recalling the notation (46) we have for the first term

$$
\begin{aligned}
&\frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) F'\Big(Q_{\text{add}}(\tau|X_i)|X_i\Big)\Delta(X_j)\pi(X_i)\pi(X_j)\Big(\widehat{q}_1^{-i}(\tau|X_{i1}) \\
&\quad - q_{1,n}(\tau|X_{i1})\Big)
\end{aligned}
$$

$$+\frac{1}{n(n-1)}\sum_{i\neq j}L_g\left(X_i-X_j\right)F'\Big(Q_{\text{add}}(\tau|X_i)|X_i\Big)\Delta(X_j)\pi(X_i)\pi(X_j)(q_{1,n}(\tau|X_{i1})$$
$$-q_1(\tau|X_{i1}))=:\ T_{5n}^{(1.1)}+T_{5n}^{(1.2)}=T_{5n}^{(1)}+o_{L_1}\left(\frac{1}{\sqrt{n}}\right), \tag{71}$$

where the first equality defines the terms $T_{5n}^{(1.1)}$ and $T_{5n}^{(1.2)}$ in an obvious manner. A straightforward but tedious calculation (using a Taylor expansion and similar arguments as in the proof of Theorem 3.1 in Dette and Scheder (2011)) yield

$$\begin{aligned}
T_{5n}^{(1.1)}=&\frac{1}{n^2(n-1)}\sum_{i\neq j}L_g\left(X_i-X_j\right)F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j)\\
&\times\sum_{k=1}^{n}\left[\widehat{Q}_{1,n}^{-i}(\tau|(X_{i1},X_{k2}))-Q_{1,n}(\tau|(X_{i1},X_{k2}))\right]\\
=&\frac{-1}{n^3(n-1)}\sum_{i\neq j}\sum_{k=1}^{n}\sum_{l=1}^{n}L_g\left(X_i-X_j\right)F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j)\\
&\times K_{b_n}\left(F\left(\frac{l}{n}|(X_{i1},X_{k2})\right)-\tau\right)\left(\widehat{F}_1^{-i}\left(\frac{l}{n}|(X_{i1},X_{k2})\right)\right.\\
&\left.-F\left(\frac{l}{n}|(X_{i1},X_{k2})\right)\right)+o_{L_1}\left(\frac{1}{\sqrt{n}}\right)\\
=&-\frac{1}{n^2(n-1)^2}\sum_{i\neq j}\sum_{k=1}^{n}\sum_{l\neq i}^{n}L_g\left(X_i-X_j\right)F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j)\\
&\times\int_0^1K_{b_n}\left(F\left(t|(X_{i1},X_{k2})\right)-\tau\right)K_{1,h_1}(X_{i1}-X_{l1})K_{2,h_2}(X_{k2}-X_{l2})\\
&\times\frac{I\{Y_l\leq t\}-F(t|(X_{i1},X_{k2}))}{f(X_{i1},X_{k2})}dt+o_{L_1}\left(\frac{1}{\sqrt{n}}\right)\\
=&Z_n^{(1)}+\tilde{Z}_n^{(2)}+o_{L_1}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

where $Z_n^{(1)}$ is defined in (58) and

$$\begin{aligned}
\tilde{Z}_n^{(2)}=&\frac{-1}{n^3(n-1)}\sum_{i\neq j}\sum_{k=1}^{n}\sum_{l\neq i}^{n}L_g\left(X_i-X_j\right)F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j)\\
&\times\int_0^1K_{b_n}\left(F\left(t|(X_{i1},X_{k2})\right)-\tau\right)K_{1,h_1}(X_{i1}-X_{l1})K_{2,h_2}(X_{k2}-X_{l2})\\
&\frac{I\{Y_l\leq t\}-F(t|(X_{l1},X_{l2}))}{f(X_{i1},X_{k2})}dt\\
=&\frac{-1}{n^3(n-1)}\sum_{i\neq j}\sum_{k=1}^{n}\sum_{l\neq i}^{n}L_g(X_i-X_j)
\end{aligned}$$

$$F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j)K_{1,h_1}(X_{i1}-X_{l1})K_{2,h_2}(X_{k2}-X_{l2})$$
$$\frac{I\{Y_l \le Q(\tau|(X_{i1},X_{k2}))\} - F(Q(\tau|(X_{i1},X_{k2}))|(X_{l1},X_{l2}))}{f(X_{i1},X_{k2})}$$
$$+ o_{L_1}\left(\frac{1}{\sqrt{n}}\right) = Z_n^{(2)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right).$$

Here, the last equality follows recalling the definition of $Z_n^{(2)}$ and some tedious but straightforward calculations. Similarly we obtain for the statistic $T_n^{(1.2)}$ defined in (71)

$$\begin{aligned} T_{5n}^{(1.2)} &= \frac{1}{n^2(n-1)}\sum_{i\neq j}^{n}\sum_{k=1}^{n} L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j) \\ &\quad \times (Q(\tau|(X_{i1},X_{k2})) - q_1(\tau|X_{i1})) + o_{L_1}\left(\frac{1}{\sqrt{n}}\right) \\ &= \frac{1}{n}\sum_{k=1}^{n} E\Big[f(X_i)F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_i)\pi^2(X_i)(Q(\tau|(X_{i1},X_{k2})) \\ &\quad - q_1(\tau|X_{i1}))|X_k\Big] + o_{L_1}\left(\frac{1}{\sqrt{n}}\right) = Z_n^{(3)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right), \end{aligned}$$

where $Z_n^{(3)}$ is defined in (60). The statistic $T_{5n}^{(2)}$ is treated similarly and we obtain the representation

$$T_{5n}^{(2)} = \sum_{j=4}^{6} Z_n^{(j)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right),$$

where $Z_n^{(4)}$, $Z_n^{(5)}$ and $Z_n^{(6)}$ are defined in (61), (62) and (63), respectively. The terms $Z_n^{(7)}, \ldots Z_n^{(12)}$ in Lemma 2 correspond to the statistics $T_{5n}^{(3)}$ and $T_{5n}^{(4)}$ in the decomposition (57) and we restrict ourselves to the calculations for the quantity $T_{5n}^{(3)}$. The corresponding representation of $T_{5n}^{(4)}$ follows exactly by the same arguments. Observing the definition of $\widehat{Q}_{l,n}$ in (6) and using a Riemann approximation and a Taylor expansion we have

$$T_{5n}^{(3)} = \left(T_{5n}^{(3.1)} + T_{5n}^{(3.2)}\right)(1+o(1)). \tag{72}$$

Here the term $T_{5n}^{(3.1)}$ is given by

$$\begin{aligned} T_{5n}^{(3.1)} &= \frac{-1}{2n^4(n-1)}\sum_{i\neq j}^{n}\sum_{k,l,m=1}^{n} L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j) \\ &\quad \times \left(\int_{-\infty}^{\tau} K_{b_n}\left(F\left(\frac{m}{n}|(X_{k1},X_{l2})\right) - u\right)du - c(\tau)\right) \end{aligned}$$

$$
\begin{aligned}
&= \frac{-1}{2n^3(n-1)} \sum_{i \neq j}^{n} \sum_{k,l=1}^{n} L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j) \\
&\quad \times (Q(\tau|(X_{k1}, X_{l2})) - c(\tau)) + o_{L_1}\left(\frac{1}{\sqrt{n}}\right) \\
&= \frac{-1}{2n^2} \sum_{k,l=1}^{n} E[F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_i)\pi^2(X_i)f(X_i)](Q(\tau|(X_{k1}, X_{l2})) \\
&\qquad - q_1(X_{k1})) \\
&\quad - \frac{1}{2n} \sum_{k}^{n} E[F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_i)\pi^2(X_i)f(X_i)](q_1(X_{k1}) - c(\tau)) \\
&\qquad + o_{L_1}\left(\frac{1}{\sqrt{n}}\right) = \frac{1}{2} Z_n^{(9)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where $Z_n^{(9)}$ is defined in (66) and the last equality follows by showing that the L^2 distance between both sides is of order $o(1/n)$. The term $T_{5n}^{(3.2)}$ in (72) is given by

$$
\begin{aligned}
&T_{5n}^{(3.2)} \\
&\quad = \frac{1}{2n^4(n-1)} \sum_{i \neq j}^{n} \sum_{k,l,m=1}^{n} L_g\left(X_i - X_j\right) F'(Q_{\text{add}}(\tau|X_i)|X_i)\Delta(X_j)\pi(X_i)\pi(X_j) \\
&\quad \times K_{b_n}\left(F\left(\frac{m}{n}|(X_{k1}, X_{l2})\right) - \tau\right)\left(\widehat{F}_1^{-i}\left(\frac{m}{n}|(X_{k1}, X_{l2})\right) - F\left(\frac{m}{n}|(X_{k1}, X_{l2})\right)\right) \\
&\quad = Z_n^{(7)} + \frac{1}{2} Z_n^{(8)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where $Z_n^{(7)}$ and $Z_n^{(8)}$ are defined in (64) and (65), respectively and the last equation follows by similar arguments as used in the treatment of the term $T_{5n}^{(1.1)}$. Finally, a similar calculation shows

$$
T_{5n}^{(4)} = Z_n^{(10)} + \frac{1}{2} Z_n^{(8)} + \frac{1}{2} Z_n^{(9)} + o_{L_1}\left(\frac{1}{\sqrt{n}}\right),
$$

where the terms $Z_n^{(j)}$ are again defined in Lemma 2. This completes the proof of the assertion.

Acknowledgments The authors thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. We are also grateful to two referees and the associate editor for their constructive comments on an earlier version of this manuscript. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilprojekt C1) of the German Research Foundation (DFG).

References

Abberger, K. (1998). Cross-validation in nonparametric quantile regression. *Allgemeines Statistisches Archiv*, *82*(2), 149–161.

Abramovich, F., De Feis, I., Sapatinas, T. (2009). Optimal testing for additivity in multiple nonparametric regression. *Annals of the Institute of Statistical Mathematics*, *61*, 691–714.

Carroll, R. J., Härdle, W., Mammen, E. (2002). Estimation in an additive model when the parameters are linked parametrically. *Econometric Theory*, *18*(4), 886–912.

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

De Gooijer, J. G., Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates. *Journal of the American Statistical Association*, *98*(461), 135–146.

Derbort, S., Dette, H., Munk, A. (2002). A test for additivity in nonparametric regression. *Annals of the Institute of Statistical Mathematics*, *54*, 60–82.

Dette, H., Scheder, R. (2011). Estimation of additive quantile regression. *Annals of the Institute of Statistical Mathematics*, *63*(2), 245–265.

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

Dette, H., von Lieres und Wilkau, C. (2001). Testing additivity by kernel-based methods—what is a reasonable test? *Bernoulli*, *7*, 669–697.

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

Doksum, K., Koo, J. Y. (2000). On spline estimators and prediction intervals in nonparametric regression. *Computational Statistics and Data Analysis*, *35*, 67–82.

Einmahl, U., Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics*, *33*(3), 1380–1403.

Eubank, R. L., Hart, J. D., Simpson, D. G., Stefanski, L. A. (1995). Testing for additivity in nonparametric regression. *Annals of Statistics*, *23*(6), 1896–1920.

Fan, Y., Linton, O. (2003). Some higher-order theory for a consistent non-parametric model specification test. *Journal of Statistical Planning and Inference*, *109*(1–2), 125–154.

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

Gozalo, P.L., Linton, O.B. (2001). Testing additivity in generalized nonparametric regression models with estimated parameters. *Journal of Econometrics*, *104*(1), 1–48.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis*, *14*, 1–16.

Hall, P., Heyde, C. (1996). *Martingale limit theory and its application. Probability and Mathematical Statistics*. New York, London: Academic Press Inc.

Härdle, W., Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, *21*(3), 1926–1947.

Härdle, W., Jeong, K., Song, R. (2012). A consistent nonparametric test for causality in quantile. *Econometric Theory*, *28*(4), 861–887.

Hengartner, N. W., Sperlich, S. (2005). Rate optimal estimation with the integration method in the presence of many covariates. *Journal of Multivariate Analysis*, *95*(2), 246–272.

Horowitz, J., Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, *100*(472), 1238–1249.

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

Koenker, R., Bassett, G. (1978). Regression quantiles. *Econometrica*, *46*(1), 33–50.

Lee, Y. K., Mammen, E., Park, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *Annals of Statistics*, *38*(5), 2857–2883.

Linton, O. B., Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, *82*(1), 93–100.

Mammen, E., Linton, O.B., Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics*, *27*(5), 1443–1490.

Neumeyer, N. (2009). Testing independence in nonparametric regression. *Journal of Multivariate Analysis*, *100*(7), 1551–1566.

Nielsen, J. P., Sperlich, S. (2005). Smooth backfitting in practice. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *67*(1), 43–61.

Nolan, D., Pollard, D. (1987). U-processes: rates of convergence. *The Annals of Statistics*, *15*(2), 780–799.
Sun, Y. (2006). A consistent nonparametric equality test of conditional quantile functions. *Econometric Theory*, *22*, 614–632.
van der Vaart, A. W., Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. New York: Springer.
Yeh, I. (2007). Modeling slump flow of concrete using second-order regressions and artificial neural networks. *Cement and Concrete Composites*, *29*(6), 474–480.
Yu, K., Jones, M.C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics and Data Analysis*, *25*(2), 159–166.
Yu, K., Jones, M.C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, *93*(441), 228–237.
Yu, K., Lu, Z. (2004). Local linear additive quantile regression. *Scandinavian Journal of Statistics*, *31*(3), 333–346.
Zhang, C., Dette, H. (2004). A power comparison between nonparametric regression tests. *Statistics and Probability Letters*, *66*, 289–301.
Zheng, J. X. (1996). A consistent test of a functional form via nonparametric estimation techniques. *Journal of Econometrics*, *75*, 263–289.