SFB 823

Discussion Paper

Testing for additivity in nonparametric quantile regression

Holger Dette, Matthias Guhlich,
Natalie Neumeyer

Nr. 52/2011

SFB 823

Testing for additivity in nonparametric quantile regression

Holger Dette, Matthias Guhlich
Ruhr-Universität Bochum
Fakultät für Mathematik
44780 Bochum, Germany
e-mail: holger.dette@rub.de

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

December 22, 2011

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.

AMS Subject Classification: 62G05, 62G20
Keywords and Phrases: 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 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), Yu and Jones (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 De 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. Lee et al. (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 Section 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 Section 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 Section 4 we present a small simulation study in order 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 an appendix in Section 5, 6 and 7.

2 Preliminaries - an additive estimator

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

$$H_0 : Q(\tau|x) = Q(\tau|x_1, \ldots, x_d) = \sum_{k=1}^{d} Q_k(\tau|x_k) + c(\tau) \tag{2.1}$$

for some constant $c(\tau)$ and functions $Q_k(\tau|x_k)$ $(k = 1, \ldots, d)$. Note that the quantities in (2.1) are not uniquely determined and in order to make these identifiable we assume throughout this paper the conditions

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

For the construction of a test for the hypothesis (2.1) let $\hat{Q}_{add}$ denote an additive estimate of the quantile regression function Q (for fixed τ), which will be specified later. We propose the statistic

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} L_g\left(X_i - X_j\right) \widehat{R}_i \widehat{R}_j, \tag{2.2}$$

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

$$\widehat{R}_i = I\{Y_i \leq \widehat{Q}_{add}^{-i}(\tau|X_i)\} - \tau, \tag{2.3}$$

the function L denotes a d-dimensional kernel function with bandwidth g and here and throughout this paper we use the notation

$$L_g(X_i - X_j) = \frac{1}{g^d} L\Big(\frac{X_i - X_j}{g}\Big). \tag{2.4}$$

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 $\widehat{Q}_{add}^{-i}(\tau|x)$ in (2.3) denotes the additive (nonparametric) estimate of the quantile regression from the sample without the ith observation. Similarly, $\widehat{Q}_{add}^{-i,j}$ and $\widehat{Q}_{add}^{-i,j,k}$ 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 De Gooijer and Zerom (2003), Horowitz and Lee (2005), Lee et al. (2010) and Dette and Scheder (2011).

Note that statistics of the type (2.4) have been introduced by Zheng (1996) in the context of testing for a specific parametric form in nonparametric regression, and since its introduction has 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]. 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 (2.2) with different rates corresponding to the null hypothesis and fixed alternatives. For this discussion which is deferred to Section 3 we therefore recall the definition of an additive quantile regression estimate which has recently been introduced by Dette and Scheder (2010) and will be used throughout this paper for a test of an additive quantile regression. Let $F(\cdot|x)$ denote the conditional distribution function of Y_j, given $X_j = x$. Following Dette and Scheder (2011) we denote by

$$\widehat{F}_l(y|x) = \frac{\sum_{i=1}^n K_{1,h_1}(x_l - X_{il}) K_{2,H}(x_{\underline{l}} - X_{i\underline{l}}) I\{Y_i \leq y\}}{\sum_{i=1}^n K_{1,h_1}(x_l - X_{il}) K_{2,H}(x_{\underline{l}} - X_{i\underline{l}})} \tag{2.5}$$

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 (2.5) the functions K_1 and K_2 are one-dimensional and $(d-1)$-dimensional kernels, respectively, h_1 is a one-dimensional bandwidth and $H = \mathrm{diag}(h_2, \ldots, h_d)$ a $(d-1)$-dimensional non-singular and diagonal (bandwidth) matrix and we use the notation

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

We also note that the statistics $\widehat{F}_l$ differ only by the component of the predictor, which is used in the kernel K_1 but not in K_2 and that (under appropriate assumptions) all of them estimate the conditional distribution function consistently. Moreover, for different values of $l = 1, \ldots, d$ different bandwidths $h_1 = h_{1,l}, h_2 = h_{2,l}$ 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

$$\widehat{Q}_{l,N}(\tau|x) = G^{-1}(\widehat{G}_{l,N}(\tau|x)), \tag{2.6}$$

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

$$\widehat{G}_{l,N}(\tau|x) = \frac{1}{N} \sum_{i=1}^N \int_{-\infty}^{\tau} K_{b_n}\Big(\widehat{F}_l\Big(G^{-1}\Big(\frac{i}{N}\Big)\Big|x\Big) - u\Big) du \tag{2.7}$$

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

$$\begin{aligned} \widehat{G}_{l,N}(\tau|x) \approx G_N(\tau|x) &:= \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{\tau} K_{b_n}\Big(F\Big(G^{-1}\Big(\frac{i}{N}\Big)\Big|x\Big) - u\Big)du \\ &\approx \int I\{F(G^{-1}(s)|x) \le \tau\}ds = G(Q(\tau|x)), \end{aligned} \tag{2.8}$$

and therefore the statistic $\widehat{Q}_{l,N}(\tau|x)$ defined in (2.6) is a reasonable estimate of the conditional quantile curve $Q(\tau|x) = F^{-1}(\tau|x)$. 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 (2.6) 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 in order to obtain an additive estimate of the quantile regression with a one-dimensional rate of convergence. To be precise define

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

as an estimate of the first marginal effect

$$q_l(\tau|x_l) := \int Q(\tau|x) f_{\underline{l}}(x_{\underline{l}})dx_{\underline{l}} = Q_l(\tau|x_l) + c(\tau), \tag{2.9}$$

where $f_{\underline{l}} : \mathbb{R}^{d-1} \to \mathbb{R}$ is the density of the random vector $(X_{j1}, \ldots, X_{jl-1}, X_{jl+1}, \ldots, X_{jd})^T$ and the second equality in (2.9) 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

$$\widehat{Q}_{add}(\tau|x) := \sum_{k=1}^{d}\widehat{q}_k(\tau|x_k) - \Big(1 - \frac{1}{d}\Big)\sum_{k=1}^{d}\frac{1}{n}\sum_{i=1}^{n}\widehat{q}_k(\tau|X_{ik}). \tag{2.10}$$

We note that this statistic is well defined even in the case when the null hypothesis (2.1) is not satisfied and in this case it estimates consistently (under appropriate assumptions) the function

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

where the quantities Q_j are defined as in (2.9). Throughout this paper we make the following assumptions regarding the kernels used in the definition of (2.2), (2.5) and (2.7).

Assumption 2.1. *The one-dimensional kernel K_1 in (2.5) is of bounded variation and has compact support $[-1,1]$ with existing moments of order 2 satisfying*

$$\int_{-1}^{1} xK_1(x)dx = 0,$$
$$c_2(K_1) = \frac{1}{2}\int_{-1}^{1} x^2 K_1(x)dx.$$

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

Assumption 2.2. *We assume that the kernel K_2 in (2.5) is a $(d-1)$-dimensional kernel of order q with support $[-1,1]^{d-1}$, that is*

$$\begin{aligned}
&(i)\ K_2 \text{ is symmetric,}\\
&(ii)\ \int_{[-1,1]^{d-1}} K_2(x_{\underline{1}})dx_{\underline{1}} = 1,\\
&(iii)\ \int_{[-1,1]^{d-1}} |x_{\underline{1}}^{\nu_{\underline{1}}}||K_2(x_{\underline{1}})|dx_{\underline{1}} < \infty \quad \textit{for} \quad |\nu_{\underline{1}}| \leq q,\\
&(iv)\ \int_{[-1,1]^{d-1}} x_{\underline{1}}^{\nu_{\underline{1}}} K_2(x_{\underline{1}})dx_{\underline{1}} = 0 \quad \textit{for} \quad 1 \leq |\nu_{\underline{1}}| \leq q-1,\\
&(v)\ \int_{[-1,1]^{d-1}} x_{\underline{1}}^{\nu_{\underline{1}}} K_2(x_{\underline{1}})dx_{\underline{1}} \neq 0 \quad \textit{for} \quad \textit{some} \quad |\nu_{\underline{1}}| = q,
\end{aligned}$$

and is of bounded variation.

The one-dimensional kernel K and the d-dimensional kernel L satisfy

Assumption 2.3.

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 $x \in [-1,1]^d$

3 Asymptotic theory

In this section we study the asymptotic properties of the statistic introduced in Section 2 for testing the hypothesis of an additive quantile regression. We begin with a statement regarding weak convergence under the null hypothesis. In order to keep the notation simple we assume that the $(d-1)$-dimensional bandwidth matrix in the definition of the estimate (2.5) is proportional

to the identity matrix, that is $H = \mathrm{diag}(h_2, \ldots, h_2) \in \mathbb{R}^{(d-1)\times(d-1)}$, 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, in order to present a result regarding weak convergence under the null hypothesis we make the following basic assumptions.

Assumption 3.1.

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 continously differentiable and $(G^{-1})'$ is uniformly bounded on closed intervals $I \subset (0,1)$.*

4. *For any x the function $Q(\cdot|x)$ is twice continuously differentiable in a neighbourhood of τ and there exists $\varepsilon > 0$ such that*

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

5. *For each $l = 1, \ldots, d$, the bandwidths g, b_n, h_1, h_2 used in the estimate $\widehat{F}_l$ in (2.5) satisfy the following conditions (if $n \to \infty$)*

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

Assumption 3.2.

For each $l = 1, \ldots, d$, the bandwidths g, b_n, h_1, h_2 used in the estimate $\widehat{F}_l$ in (2.5) satisfy the following conditions (if $n \to \infty$)

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

for some $\alpha > 0$.

The following result establishes weak convergence of the test statistic T_n defined in (2.2) . Throughout this paper the symbol $\xrightarrow{\mathcal{D}}$ denotes weak convergence.

Theorem 3.3. *If Assumption 2.1, 2.2, 2.3, 3.1, 3.2 and the null hypothesis* (2.1) *of an additive quantile regression model are satisfied, it follows that*

$$ng^{d/2}T_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{3.1}$$

where the asymptotic variance is given by

$$\sigma^2 = 2\tau^2(1-\tau)^2 \int L^2(u)du \int f^2(x)dx. \tag{3.2}$$

Remark 3.4. We would like to point out that a result of the form (3.1) is typical for the limit distribution of a statistic of the type defined in (2.2) [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 (2.2), 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 3.2. 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

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

results in the estimate

$$J_{nU} \leq J_n \leq J_{nL} \tag{3.3}$$

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 $\varepsilon_{iU} = I\{Y_i + C_n \leq Q(\tau|X_{il})\} - \tau,\ \varepsilon_{iL} = I\{Y_i + C_n \leq Q(\tau|X_{il})\} - \tau$ (see equation (A.11-3) in this paper). A simple calculation shows that this conclusion is not correct and in fact the inequality (3.3) 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 3.2, 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

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

where $\hat{Q}^{-i}(\tau|X_{il})$ denotes the quantile regression estimate of Dette and Volgushev (2008) frome the two-dimensional sampe $(X_{il}, Y_i)_{i=1}^n$ and σ^2 is defined in (3.2) (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 (2.2) under local and fixed alternatives. For this purpose we introduce the "residuals"

$$R_j = I\{Y_j \leq Q(\tau|X_j)\} - \tau \tag{3.4}$$

and denote by

$$\Delta(X_j) = E[R_j - R_j^{add} \mid X_j] = -E[R_j^{add} \mid X_j] \tag{3.5}$$

the conditional expectation of the distance between the "residuals" defined before and the "restricted residuals"

$$R_j^{add} = I\{Y_j \leq Q_{add}(\tau|X_j)\} - \tau \tag{3.6}$$

obtained from the best additive approximation. Note that under the null hypothesis we have $\Delta(X_j) = 0$ a.s., while under the alternative it follows that $P(\Delta(X_j) = 0) < 1$. We first consider the properties of the test for local alternatives of the form

$$Q(\tau|x) = Q_{add}(\tau|x) + d_n l(x), \tag{3.7}$$

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 3.5. *Assume that Assumption 2.1, 2.2, 2.3, 3.1 and 3.2 are satisfied. Under local alternatives of the form (3.7) with* $d_n = 1/(n^{1/2} g^{\frac{d}{4}})$ *it follows that*

$$ng^{\frac{d}{2}} T_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu, \sigma^2), \tag{3.8}$$

where the asymptotic variance and bias are given by (3.2) *and*

$$\mu = E[(F'(Q(\tau|X_1)|X_1))^2 l^2(X_1) f(X_1)],$$

respectively.

The following result specifies the asymptotic distribution of the test statistic T_n defined in (2.2) under fixed alternatives. For its proof we require the following additional assumptions

Assumption 3.6.

1. *For any* $y \in \mathbb{R}$ *we have* $F(y|\cdot) \in C_b^q([0,1]^d)$.
2. *We assume the representation*

$$I\{Y_j \leq y\} - F(y|X_j) = s(y|X_j)\varepsilon_j, \ j = 1, \ldots, n, \tag{3.9}$$

   where $s^2(y|x) = F(y|x)(1 - F(y|x))$ *and the* ε_j *are independent of* Y_j*, but may depend on* X_j*. We moreover assume that* $\sup_x E[\varepsilon_j^4 | X_j = x] < \infty$.

3. *For each* $l = 1, \ldots, d$*, the bandwidths* g, b_n, h_1, h_2 *used in the estimate* $\widehat{F}_l$ *in* (2.5) *satisfy the following conditions (if* $n \to \infty$*)*

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

Note that in model (3.9) the random variables are independent identically distributed with with $E[\varepsilon_j|X_j] = 0$, $\mathrm{Var}(\varepsilon_j|X_j) = 1$. Moreover this model is a common assumption in quantile regression [see Hall et al. (1999) or Dette and Scheder (2011) among others] and observing (3.4) we have

$$s(Q(\tau|X_j)|X_j)\varepsilon_j = I\{Y_j \leq Q(\tau|X_j)\} - \tau = R_j. \tag{3.10}$$

Theorem 3.7. *If Assumption 2.1, 2.2, 2.3, 3.1 and 3.6 are satisfied and the null hypothesis* (2.8) *does* ***not*** *hold, then we have as* $n \to \infty$

$$n^{1/2}(T_n - E[T_n]) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2), \tag{3.11}$$

where

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

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

$$b_\alpha(x_\alpha) = c_2(K_1)\int\Big(\frac{1}{2}\frac{\frac{\partial^2}{\partial x_\alpha^2}F(Q(\tau|x_\alpha,t_{\underline{\alpha}})|x_\alpha,t_{\underline{\alpha}})}{F'(Q(\tau|x_\alpha,t_{\underline{\alpha}})|x_\alpha,t_{\underline{\alpha}})} + \frac{\frac{\partial}{\partial x_\alpha}F(Q(\tau|x_\alpha,t_{\underline{\alpha}})|x_\alpha,t_{\underline{\alpha}})\frac{\partial}{\partial x_\alpha}f(x_\alpha,t_{\underline{\alpha}})}{F'(Q(\tau|x_\alpha,t_{\underline{\alpha}})|x_\alpha,t_{\underline{\alpha}})f(x_\alpha,t_{\underline{\alpha}})}\Big)f_{\underline{\alpha}}(t_{\underline{\alpha}})dt_{\underline{\alpha}}. \tag{3.12}$$

The asymptotic variance in (3.11) *is given by*

$$\begin{aligned}\sigma^2 = 4\mathrm{Var}\Big[&\Delta^2(X_1)f(X_1) - E\Big(\Delta(X_2)f(X_2)F'(Q_{add}(\tau|X_2)|X_2)\Big(\sum_{\alpha=1}^d Q(\tau|X_{2\alpha},X_{1\underline{\alpha}}) \\ &- \Big(1-\frac{1}{d}\Big)\sum_{\alpha=1}^d Q(\tau|X_{1\alpha},X_{3\underline{\alpha}}) + Q(\tau|X_{3\alpha},X_{1\underline{\alpha}})\Big)\Big|X_1\Big)\Big] \\ &+ 4\tau(1-\tau)E\Big[\Big(\Delta(X_1)f(X_1) \\ &+ \sum_{\alpha=1}^d \frac{f_{\underline{\alpha}}(X_{1\underline{\alpha}})}{f(X_1)F'(Q(\tau|X_1)|X_1)}\int \Delta(X_{1\alpha},t_{\underline{\alpha}})f^2(X_{1\alpha},t_{\underline{\alpha}})F'(Q_{add}(\tau|X_{1\alpha},t_{\underline{\alpha}})|X_{1\alpha},t_{\underline{\alpha}})dt_{\underline{\alpha}} \\ &- (d-1)\frac{1}{F'(Q(\tau|X_1)|X_1)}\int \Delta(t)f^2(t)F'(Q_{add}(\tau|t)|t)dt\Big)^2\Big].\end{aligned}$$

Remark 3.8. Note that Theorem 3.3 provides an asymptotic level α test for the hypothesis (2.1) 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 (3.2). Moreover, by Theorem 3.7 it follows that this test is consistent, because under the alternative we have

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

from this result.

4 Finite sample properties and a data example

4.1 A small simulation study

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

$$Y_i = Q(0.5|X_i) + 0.25\varepsilon_i$$

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

(A) X_i are uniformly distributed on the unit square $[0,1]^2$

$$X_i = (X_{i1}, X_{i2}) \sim \mathcal{U}([0,1]^2), \quad i = 1, \ldots, n$$

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

$$X_{i1} = \frac{1}{2} + \frac{1}{\pi} \arctan(Z_{i1})$$
$$X_{i2} = \frac{1}{2} + \frac{1}{\pi} \arctan(Z_{i2})$$

where $Z_i = (Z_{i1}, Z_{i2})$ are (independent) centered normally distributed random variables with variance 1 and correlation $\rho = 0.2, \rho = 0.5, \rho = 0.8$.

Note that in Design A the random variables X_{i1} and X_{i2} are independent, whereas Design B represents a situation where X_{i1} and X_{i2} are correlated. In our simulation study we consider three models for the conditional quantile function, that is

$$Q(0.5|x_1, x_2) = x_1 + x_2 \tag{4.1}$$
$$Q(0.5|x_1, x_2) = x_1^2 + x_2^2 \tag{4.2}$$
$$Q(0.5|x_1, x_2) = \cos(c\pi(x_1 + x_2)), \ c = 0.5, 1, 2, \tag{4.3}$$

where the first two cases correspond to the null hypothesis of additivity and (4.3) represents three alternatives. For all kernels in our estimators we use the Epanechnikov kernel $K(t) = \frac{3}{4}(1-t^2)I_{[-1,1]}(t)$, and a product of two kernels of this type as a two-dimensional kernel. Following Dette and Scheder (2011) the bandwidths are chosen as

$$h_1 = 0.6n^{-\frac{1}{5}}, h_2 = 0.2n^{-\frac{1}{5}}, b_n = 0.1n^{-\frac{1}{4}}, g = 0.1n^{-\frac{1}{4}}.$$

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

$$Y_i^* = \hat{Q}_{add}(\tau|X_i) + v_i|Y_i - \hat{Q}_N(\tau|X_i)|,$$

where $\widehat{Q}_{add}$ and $\widehat{Q}_N$ are defined in Section 2 and v_i denote independent identically distributed random variables satisfying $P(v_i = -1) = \tau$ and $P(v_i = 1) = 1 - \tau$, which are independent

from the original sample. A similar bootstrap data generation was suggested by Sun (2006) The bootstrap observations, conditionally on the original sample, fulfill H_0 and additionally fulfill a τ-quantile regression model because

$$P^*(Y_i^* \leq \hat{Q}_{add}(\tau|X_i) \mid X_i = x) = P(v_i \leq 0) = \tau,$$

where P^* denotes the probability conditionally on (X_j, Y_j), $j = 1, \ldots, n$. Note also that for the median model used in the simulations we have $\tau = \frac{1}{2}$ and v_i are Rademacher variables. 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 if $T_n \geq t^*_{n,(1-\alpha)}$. 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.
In Table 1 we display the results of the simulation study for model (4.1) and (4.2) for $n = 100$ which represent the null hypothesis. The corresponding results under the alternative defined by model (4.3) are shown in Table 2. Under the null hypothesis we observe a reasonable approximation of the nominal level under Design A and B provided that the correlation between the explanatory variables in the latter case is not too large [see Table 1]. If $\rho = 0.8$ in Design B the nominal level is overestimated, which means that the critical values of the bootstrap procedure are too small. The results in Table 2 demonstrate that the bootstrap test detects alternatives with reasonable power in all cases under investigation.

	model (4.1)				model (4.2)			
	A	B			A	B		
		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
1%	2.0	1.8	2.2	3.1	1.5	2.5	1.7	3.0
5%	5.8	5.7	6.0	8.9	6.5	5.5	6.7	10.5
10%	10.1	10.6	10.0	15.6	11.5	10.5	11.4	15.8
20%	19.8	21.6	19.6	27.7	21.0	19.5	23.1	26.3

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

4.2 A data example

Recently Dette and Scheder (2011) applied additive quantile regression to investigate the Boston housing data, which has been analyzed by several authors. The dataset contains the housing values of suburbs of Boston and 13 variables, which might have an influence on the housing prices like pollution, crime and urban amenities. This dataset has been analyzed by several authors, also in the context of quantile regression. Dette and Scheder (2011) focus on the four covariates

	$c = 0.5$				$c = 1$				$c = 2$			
	A	B			A	B			A	B		
		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$		$\rho = 0.2$	$\rho = 0.5$	$\rho = 0.8$
1%	2.7	2.8	3.0	3.5	55.1	59.0	66.4	61.7	84.8	97.5	97.0	95.5
5%	7.2	6.8	6.8	9.9	73.8	77.5	83.6	76.8	92.7	99.0	99.0	98.7
10%	12.6	13.2	13.4	17.2	83.0	85.3	91.2	84.1	96.5	99.7	99.7	99.0
20%	24.9	22.4	23.2	27.3	92.0	92.9	96.0	90.1	98.8	99.7	99.7	99.8

Table 2: *Simulated power of the bootstrap test for the hypothesis of an additive quantile regression model* (4.3) *corresponding to the alternative.*

- per capita crime rate,
- average number of rooms per dwelling,
- weighted mean of distance to five Boston employment centres,
- lower status of the population

and fit their additive conditional quantile model to the data. In order to check if their assumption of an additive quantile regression model for $\tau = 0.5$ is reasonable, we apply the bootstrap test where we use the same bandwidths as Dette and Scheder (2011), i.e. $h_1 = 8.60, h_2 = 0.70, h_3 = 2.10, h_4 = 7.14, b_n = 0.02, g_1 = 8.60, g_2 = 0.70, g_3 = 2.10, g_4 = 7.14$. The value of the test statistic is given be $T_n = 0.00004057821$ and the $p-$ value from $B = 100$ Bootstrap replications is $p = 0.175$. This indicates that the hypothesis of additivity cannot be rejected at a controlled type I error of 10%.

5 Appendix: Proof of Theorem 3.2

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 (2.2) and consider the decomposition

$$T_n = T_{1n} + T_{2n} + T_{3n}$$

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

$$T_{1n} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) R_i R_j \tag{5.1}$$

$$T_{2n} = \frac{2}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) R_i (\widehat{R}_j - R_j) \tag{5.2}$$

$$T_{3n} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) (\widehat{R}_i - R_i)(\widehat{R}_j - R_j) \tag{5.3}$$

and R_i and $\widehat{R}_i$ are defined in (3.4) and (2.3), respectively. The assertion follows from the following two statements, which are proved below

$$ng^{\frac{d}{2}}T_{1n} \xrightarrow{\mathcal{D}} N(0,\sigma^2) \tag{5.4}$$

$$ng^{\frac{d}{2}}T_{jn} = o_p(1), \quad j = 2,3. \tag{5.5}$$

5.1 Proof of (5.4)

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

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

we can write the 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) if the assumptions of this statement can be checked. For this purpose note that we obtain from Assumption 3.2 for $i \neq j \neq k \neq i$ for some $\lambda > 0$

$$\begin{aligned}
&E[E[H_n(Z_k, Z_i)H_n(Z_k, Z_j)|Z_i, Z_j]^2] \\
&\leq \frac{\lambda}{g^{4d}} E\Big[E\Big[L\Big(\frac{X_k - X_i}{g}\Big)L\Big(\frac{X_k - X_j}{g}\Big)|X_i, X_j\Big]^2\Big] \\
&= \frac{\lambda}{g^{4d}} \int\int \Big[\int L(u)L(u+v)f(x+ug)g^d du\Big]^2 f(x)f(x-vg)g^d dxdv \;=\; O\Big(\frac{1}{g^d}\Big) \\
&E[H_n^2(Z_i, Z_j)] = \; \tau^2(1-\tau)^2 \frac{1}{g^{2d}} E\Big[L^2\Big(\frac{X_i - X_j}{g}\Big)\Big] \\
&= \tau^2(1-\tau)^2 \frac{1}{g^d}\int L^2(u)du \int f^2(x)dx + o\Big(\frac{1}{g^d}\Big) \;=\; \frac{\sigma^2}{2g^d} + o\Big(\frac{1}{g^d}\Big),
\end{aligned}$$

where $\sigma^2 > 0$ is defined in (3.2). In a similar way one establishes the estimate $E[H_n^4(Z_i, Z_j)] = O\Big(\frac{1}{g^{3d}}\Big)$, which gives

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

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 (3.2).

5.2 Proof of (5.5)

For the proof of (5.5) we define for $\alpha > 0$ defined in Assumption 3.2

$$C_n = n^\alpha \sqrt{\frac{\log n}{nh_1 h_2^{d-1}}};\ D_n = n^\alpha \frac{1}{nh_1} \tag{5.6}$$

and introduce the set

$$\Omega_n = \{ \sup_x |\widehat{Q}_{add}(\tau|x) - Q(\tau|x)| \le C_n, \sup_x \max_{k=1}^n |\widehat{Q}_{add}^{-k}(\tau|x) - \widehat{Q}_{add}(\tau|x)| \le D_n\}. \tag{5.7}$$

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

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

where

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

It is easy to see, that on the set Ω_n

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

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

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

We now calculate

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

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

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

With Assumption 3.2 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 (5.5) in the case $j = 3$ follows from $E[T_{3n}^2] = O(1/g^2)$ and the following result.

Lemma 5.1. *For Ω_n defined in (5.7) we have that*

$$P(\Omega_n^C) = O\left(p(n)\exp\left(-n^{2\alpha}\right)\right) \tag{5.8}$$

where $p(n)$ is a polynomial in n and α is defined in Assumption 3.2.

Proof of Lemma 5.1 For a proof of (5.8) it suffices to show that

$$P(\sup_x |\widehat{Q}_{add}(\tau|x) - Q_{add}(\tau|x)| > C_n) = O\left(p(n)\exp\left(-n^\alpha\right)\right) \tag{5.9}$$

$$P(\sup_x \max_i |\widehat{Q}_{add}(\tau|x) - \widehat{Q}^{-i}_{add}(\tau|x)| > D_n) = O\left(p(n)\exp\left(-n^\alpha\right)\right), \tag{5.10}$$

At first we consider the probability (5.9). We have that

$$\sup_x |\widehat{Q}_{add}(\tau|x) - Q_{add}(\tau|x)| \leq \sum_{k=1}^d \Big\{ B^{(1)}_{nk} + \Big(1 - \frac{1}{d}\Big) B^{(2)}_{nk} \Big\} \tag{5.11}$$

where

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

and consider the term $B^{(1)}_{n1}$ (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 $q_1(\tau|x_1) = q_{1,n}(\tau|x_1) + o\left(C_n\right)$, uniformly with respect to x_1, where

$$q_{1,n}(\tau|x_1) = \frac{1}{n}\sum_{i=1}^n Q_{1,n}(\tau|(x_1, X_{i\underline{1}})) \tag{5.12}$$

and we introduce the notation

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

and G_N is defined in (2.8). Recalling the definition of $\widehat{Q}_{l,n}$ in (2.6) we obtain by a Taylor expansion and similar arguments as in Dette and Scheder (2011)

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

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

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

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

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

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

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

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

and therefore we have for sufficiently large n

$$\begin{aligned}P\left(B_{n1}^{(1)}>C_n\right) &= P\big(B_{n1}^{(1)}>C_n\big|\mathcal{D}_n\big)P\Big(\mathcal{D}_n\Big)+P\big(B_{n1}^{(1)}>C_n\big|\mathcal{D}_n^c\big)P\left(\mathcal{D}_n^c\right)\\ &\le P\Big(\sup_{x,y}C\Big|\widehat{F}_1(y|x)-F(y|x)\Big|>C_n\Big)+P\Big(\sup_{y,x}|\widehat{F}_1(y|x)-F(y|x)|>b_n\Big)\\ &\le 2P\Big(\sup_{x,y}C|\widehat{F}_1(y|x)-F(y|x)|>C_n\Big)\end{aligned}\tag{5.13}$$

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

$$\widehat{h}(x,y) = \frac{1}{n}\sum_{k=1}^n K_{1,h_1}(x_1-X_{k1})K_{2,h_2}(x_{\underline{1}}-X_{k\underline{1}})1\{Y_k\le y\}$$

$$\widehat{f}(x) = \frac{1}{n}\sum_{k=1}^{n} K_{1,h_1}(x_1 - X_{k1})K_{2,h_2}(x_{\underline{1}} - X_{k\underline{1}})$$

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

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

$$h(x,y) = F(y|x)f(x)$$

straightforward calculations yield

$$\left|\widehat{F}_1(y|x) - F(y|x)\right| \leq \; C_{n1}(x,y) + C_{n2}(x,y) \tag{5.14}$$

where

$$\begin{aligned} C_{n1}(x,y) \; &= \; \Big|\frac{(\widehat{h}(x,y) - h(x,y))}{\widehat{f}(x)}\Big| \\ C_{n2}(x,y) \; &= \; \Big|\frac{h(x,y)(\widehat{f}(x) - f(x))}{\widehat{f}(x)f(x)}\Big|. \end{aligned}$$

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 (5.14) (where $\delta > 0$ is chosen sufficiently small)

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

and with similar arguments one can show

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

Recalling (5.13) and combining these estimates we obtain

$$P(B_{n1}^{(1)} > C_n) \leq 6P\big(C\sup_x |\widehat{f}(x) - f(x)| > C_n\big) + 2P\big(C\sup_{x,y} |\widehat{h}(x,y) - h(x,y)| > C_n\big). \tag{5.15}$$

For the first probability on the right hand side of (5.15) we have that

$$\begin{aligned} P\Big(C\sup_x |\widehat{f}(x) - f(x)| > C_n\Big) \leq& P\Big(2C\sup_x |\widehat{f}(x) - E[\widehat{f}(x)]| > C_n\Big) \\ &+ P\Big(2C\sup_x |E[\widehat{f}(x)] - f(x)| > C_n\Big). \end{aligned} \tag{5.16}$$

The second term of the right hand side of (5.16) 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 obtain that the class

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

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

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

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

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

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

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

where $p(n)$ is a polynomial. The second term in (5.15) can be treated with the same arguments. For a proof of (5.9) it remains to consider the term $B_{n1}^{(2)}$ defined in (5.11) (the cases $k = 2, \ldots, d$ are treated in exactly the same way). We have

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

and the assertion follows from what we have shown before and the Markov inequality. Next we consider the proof of (5.10).

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

where

$$D_{nk}^{(1)} = \sup_{x_k} \max_{i=1}^n |\widehat{q}_k(\tau|x_k) - \widehat{q}_k^{-i}(\tau|x_k)|$$

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

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

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

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

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

and the assertion of Lemma 5.1 follows by the same arguments as before. □

Now we prove assertion (5.5) for the term T_{2n}. Recalling its definition in (5.2) we have

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

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

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

with $\hat{R}_i^{-j} = I\{Y_i \leq \hat{Q}_{add}^{-i,j}(\tau|X_i)\} - \tau$. 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 3.2. For the second term $T_{2n}^{(2)}$ we have that $E[T_{2n}^{(2)}] = 0$ and

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

where

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

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

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

$$\tilde{U}_2 = \frac{4}{n^2(n-1)^2} \sum_{i_1,i_2,j_1,j_2}^{\neq} L_g\left(X_{i_1} - X_{j_1}\right) L_g\left(X_{i_2} - X_{j_2}\right) (\widehat{R}_{i_1}^{-j_1} - \widehat{R}_{i_1}^{-j_1,j_2}) R_{j_1} (\widehat{R}_{i_2}^{-j_2} - \widehat{R}_{i_2}^{-j_1,j_2}) R_{j_2}$$
$$+ \frac{8}{n^2(n-1)^2} \sum_{i_1,i_2,j}^{\neq} L_g\left(X_{i_1} - X_{i_2}\right) L_g\left(X_{i_2} - X_j\right) (\widehat{R}_{i_1}^{-i_2} - \widehat{R}_{i_1}^{-i_2,j}) R_{i_2} (\widehat{R}_{i_2}^{-j} - R_{i_2}) R_j$$

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 3.2. 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 (5.5).

6 Proof of Theorem 3.5

Recall the definition of C_n and D_n in (5.6) and consider the decomposition

$$T_n = (T_{1n} + 2T_{2n} + T_{3n}) + (-2T_{4n} - 2T_{5n} + T_{6n}) \tag{6.1}$$

where T_{1n} is defined in (5.1), the statistics $T_{jn} (j = 2, \ldots, 6)$ are given by

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

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

$$T_{4n} = \frac{1}{n(n-1)} \sum_{i\neq j} L_g\left(X_i - X_j\right) R_i (R_j - R_j^{add})$$

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

$$T_{6n} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right)(R_i - R_i^{add})(R_j - R_j^{add})$$

and R_i, $\widehat{R}_i$ and R_i^{add} are defined in (3.4), (2.3) and (3.6), respectively. Observing the proofs of (5.4) and (5.5), respectively, we have that under the local alternatives of the form (3.7)

$$ng^{\frac{d}{2}} T_{1n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2); \quad T_{jn} = o\Big(\frac{1}{ng^{\frac{d}{2}}}\Big); \quad j = 2, 3 \tag{6.2}$$

in L_1, and it remains to investigate the terms T_{4n}, T_{5n} and T_{6n} in the decomposition (6.1). First we study the statistic T_{4n} for which we have that $E[T_{4n}] = 0$ and

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

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})

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

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

$$E[L_g(X_{i_1} - X_{j_1})^2 R_{i_1}^2 (R_{j_1} - R_{j_1}^{add})^2] = O\Big(\frac{d_n}{g^d}\Big)$$

$$E[L_g(X_{i_1} - X_{j_1})^2 R_{i_1} R_{j_1} (R_{j_1} - R_{j_1}^{add})(R_{i_1} - R_{i_1}^{add})] = O\Big(\frac{d_n^2}{g^d}\Big)$$

Combining these estimates we have

$$ng^{\frac{d}{2}} T_{4n} = o_p(1). \tag{6.3}$$

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

$$ng^{\frac{d}{2}} T_{5n} = O_p\Big(ng^{\frac{d}{2}} d_n C_n\Big) = o_p(1). \tag{6.4}$$

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

$$E[ng^{\frac{d}{2}} T_{6n}] = \ ng^{\frac{d}{2}} E[L_g(X_1 - X_j)(R_1 - R_1^{add})(R_2 - R_2^{add})] \tag{6.5}$$

$$
\begin{aligned}
&= ng^{\frac{d}{2}} E[L_g(X_1 - X_j)(F(Q_{add}(\tau|X_1) + d_n l(X_1)|X_1) - F(Q_{add}(\tau|X_1)|X_1)) \\
&\quad \times (F(Q_{add}(\tau|X_2) + d_n l(X_2)|X_2) - F(Q_{add}(\tau|X_2)|X_2))] \\
&= E[(F'(Q_{add}(\tau|X_1)|X_1) l(X_1))^2 f(X_1)] + o(1)
\end{aligned}
$$

and

$$
E[(T_{6n} - E[T_{6n}])^2] = o\Big(\frac{1}{ng^{\frac{d}{2}}}\Big) \tag{6.6}
$$

Thus (3.8) follows from (6.1)–(6.6).

7 Proof of Theorem 3.7

For a proof of Theorem 3.7 we assume for a transparent notation $d = 2$. The general case follows by exactly the same arguments. Recall the decomposition (6.1). Observing the proof of Theorem 3.3 we have

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

in L_1. Therefore we obtain $E[T_n] = \sum_{j=4}^{6} E[T_{jn}] + o(1/\sqrt{n})$ and

$$
\sqrt{n}(T_n - E[T_n]) = \sqrt{n} \sum_{j=4}^{6} (T_{jn} - E[T_{jn}]) + o_p(1),
$$

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

$$
T_{4n} = \sum_{i=1}^{n} R_i E[T_{4n}^{(i)}|X_i] + o_p\Big(\frac{1}{\sqrt{n}}\Big) = \frac{1}{n} \sum_{i=1}^{n} R_i \Delta(X_i) f(X_i) + o_p\Big(\frac{1}{\sqrt{n}}\Big)
$$

where

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

and $\Delta(X_j)$ is defined in (3.5). A corresponding stochastic expansion for the term T_{5n} requires substantially more effort. More precisely, we have the following result, which is proved in the Appendix.

Lemma 7.1. *Under the assumptions of Theorem 3.7 we have*

$$
T_{5n} = \sum_{j=1}^{10} Z_n^{(j)} + o\Big(\frac{1}{\sqrt{n}}\Big) \tag{7.1}
$$

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

$$
Z_n^{(1)} = E\Big[L_g(X_i - X_j) F'(Q_{add}(\tau|X_i)|X_i) \Delta(X_j) K_{1,h_1}(X_{i1} - X_{l1}) K_{2,h_2}(X_{k2} - X_{l2}) \tag{7.2}
$$

$$\times \Big(\frac{F(Q(\tau|(X_{i1},X_{k2}))|(X_{l1},X_{l2})) - F(Q(\tau|(X_{i1},X_{k2}))|(X_{i1},X_{k2}))}{f(X_{i1},X_{k2})F'(Q(\tau|(X_{i1},X_{k2}))|(X_{i1},X_{k2}))}\Big)\Big]$$

$$Z_n^{(2)} = -\frac{1}{n}\sum_{l=1}^{n} R_l h_2(X_l) \tag{7.3}$$

$$\begin{aligned} Z_n^{(3)} =& \frac{1}{n}\sum_{k=1}^{n}\big(E[f(X)F'(Q_{add}(\tau|X)|X)\Delta(X)Q(\tau|(X_1,X_{k2}))|X_k] \\ &- E[f(X)F'(Q_{add}(\tau|X)|X)\Delta(X)Q(\tau|(X_1,X_{k2}))]\big) \end{aligned} \tag{7.4}$$

$$\begin{aligned} Z_n^{(4)} =& E\Big[L_g\left(X_i - X_j\right)F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)K_{1,h_1}(X_{k1}-X_{l1})K_{2,h_2}(X_{i2}-X_{l2}) \\ &\times \Big(\frac{F(Q(\tau|(X_{k1},X_{i2}))|(X_{l1}X_{l2})) - F(Q(\tau|(X_{k1},X_{i2}))|(X_{k1},X_{i2}))}{f(X_{k1},X_{i2})F'(Q(\tau|(X_{k1},X_{i2}))|(X_{k1},X_{i2}))}\Big)\Big] \end{aligned} \tag{7.5}$$

$$Z_n^{(5)} = -\frac{1}{n}\sum_{l=1}^{n} R_l h_5(X_l) \tag{7.6}$$

$$\begin{aligned} Z_n^{(6)} =& \frac{1}{n}\sum_{k=1}^{n}\big(E[f(X)F'(Q_{add}(\tau|X)|X)\Delta(X)Q(\tau|(X_{k1},X_2))|X_k] \\ &- E[f_X(X)F'(Q_{add}(\tau|X)|X)\Delta(X)Q(\tau|(X_{k1},X_2))]\big) \end{aligned} \tag{7.7}$$

$$\begin{aligned} Z_n^{(7)} =& E\Big[L_g\left(X_i - X_j\right)F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)K_{1,h_1}(X_{m1}-X_{l1})K_{2,h_2}(X_{k2}-X_{l2}) \\ &\times \Big(\frac{F(Q(\tau|(X_{m1},X_{k2}))|(X_{l1},X_{l2})) - F(Q(\tau|(X_{m1},X_{k2}))|(X_{m1},X_{k2}))}{f(X_{m1},X_{k2})F'(Q(\tau|(X_{m1},X_{k2}))|(X_{m1},X_{k2}))}\Big)\Big] \end{aligned} \tag{7.8}$$

$$Z_n^{(8)} = \frac{1}{n}\sum_{l=1}^{n} R_l h_8(X_l) \tag{7.9}$$

$$Z_n^{(9)} = \frac{-1}{n}\sum_{k=1}^{n} E[f(X)F'(Q_{add}(\tau|X)|X)\Delta(X)(q_1(X_{k1}) + q_2(X_{k2}) - 2c(\tau))|X_k] \tag{7.10}$$

$$\begin{aligned} Z_n^{(10)} =& E\Big[L_g\left(X_i - X_j\right)F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)K_{2,h_2}(X_{m1}-X_{l1})K_{1,h_1}(X_{k2}-X_{l2}) \\ &\times \Big(\frac{F(Q(\tau|(X_{m1},X_{k2}))|(X_{l1},X_{l2})) - F(Q(\tau|(X_{m1},X_{k2}))|(X_{m1},X_{k2}))}{f(X_{m1},X_{k2})F'(Q(\tau|(X_{m1},X_{k2}))|(X_{m1},X_{k2}))}\Big)\Big] \end{aligned} \tag{7.11}$$

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

$$h_2(X_l) = \frac{f_2(X_{l2})\int \Delta(X_{l1},t_2)f^2(X_{l1},t_2)F'(Q_{add}(\tau|(X_{l1},t_2))|(X_{l1},t_2))dt_2}{f(X_l)F'(Q(\tau|X_l)|X_l)} \tag{7.12}$$

$$h_5(X_l) = \frac{f_1(X_{l1})\int \Delta(t_1,X_{l2})f^2(t_1,X_{l2})F'(Q_{add}(\tau|(t_1,X_{l2}))|(t_1,X_{l2}))dt_1}{f(X_l)F'(Q(\tau|X_l)|X_l)} \tag{7.13}$$

$$h_8(X_l) = \frac{f_1(X_{l1})f_2(X_{l2})\int \Delta(t)f^2(t)F'(Q_{add}(\tau|t)|t)dt}{f(X_l)F'(Q(\tau|X_l)|X_l)}. \tag{7.14}$$

Next we study the term T_{6n} using Lemma 3.1 in Zheng (1996) with the kernel $H(Z_i,Z_j) = L_g(X_i - X_j)(R_i - R_i^{add})(R_j - R_j^{add})$, 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

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

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

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

From Lemma 7.1 we have for the expectation of the statistic T_{5n}

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

where

$$\begin{aligned}
E[Z_n^{(1)}]&=-E\left[\Delta(X_1)F'(Q_{add}(\tau|X_1)|X_1)f(X_1)b_1(X_{11})\right]h_1^2+o(h_1^2)+O(h_2^q)\\
E[Z_n^{(4)}]&=-E\left[\Delta(X_1)F'(Q_{add}(\tau|X_1)|X_1)f(X_1)b_2(X_{12})\right]h_1^2+o(h_1^2)+O(h_2^q)\\
E[Z_n^{(7)}]&=\frac{1}{2}E\left[\Delta(X_1)F'(Q_{add}(\tau|X_1)|X_1)f(X_1)b_1(X_{21})\right]h_1^2+o(h_1^2)+O(h_2^q)\\
E[Z_n^{(10)}]&=\frac{1}{2}E\left[\Delta(X_1)F'(Q_{add}(\tau|X_1)|X_1)f(X_1)b_2(X_{22})\right]h_1^2+o(h_1^2)+O(h_2^q)
\end{aligned}$$

and the bias b_α is defined in (3.12). Observing (6.1) it therefore follows that

$$E[T_n]=E[\Delta^2(X_1)f(X_1)]+2E[F'(Q_{add}(\tau|X_1)|X_1)\Delta(X_1)f(X_1)(b(X_1)-\frac{1}{2}b(X_2))]h_1^2+o(h_1^2)+O(g^2)$$

which is the claimed representation in Theorem 3.7 for the case $d=2$. With the same argument we obtain the stochastic expansion

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

where the quantities A_n, B_n and C_n are given by

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

$$=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}R_i\Big(\Delta(X_i)f(X_i)+h_2(X_i)+h_5(X_i)-h_8(X_i)\Big)$$

and h_2, h_5 and h_8 are defined in (7.12), (7.13) and (7.14), respectively. Therefore asymptotic normality is a direct consequence of Lyapunov's central limit theorem. Finally a straightforward calculation yields

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

and $\mathrm{Cov}(A_n+B_n,C_n)=0$, which completes the proof of Theorem 3.7

7.1 Proof of Lemma 7.1

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

$$\widetilde{T}_{5n}^{(1)}=\frac{1}{n(n-1)}\sum_{i\neq j}L_g\left(X_i-X_j\right)(\widehat{R}_i-R_i^{add})(R_j-R_j^{add}-E[R_j-R_j^{add}|X_j])$$

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

We calculate

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

where the last estimate follows by similar arguments as given for the term T_{3n} under the null hypothesis [see Section 5]. With similar arguments we obtain

$$\widetilde{T}_{5n}^{(2)} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right)\left(F(\widehat{Q}_{add}^{-i}(\tau|X_i)|X_i) - F(Q(\tau|X_i)|X_i)\right) E[R_j - R_j^{add}|X_j] + o\Big(\frac{1}{\sqrt{n}}\Big)$$

and therefore a taylor expansion and Lemma 5.1 yield

$$\begin{aligned} T_{5n} &= \widetilde{T}_{5n}^{(2)} + o\Big(\frac{1}{\sqrt{n}}\Big) \\ &= T_{5n}^{(1)} + T_{5n}^{(2)} + T_{5n}^{(3)} + T_{5n}^{(4)} + o\Big(\frac{1}{\sqrt{n}}\Big) \end{aligned}$$

where we introduce the notation

$$T_{5n}^{(\ell)} = \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)(\widehat{q}_\ell^{-i}(\tau|X_{i\ell}) - q_\ell(\tau|X_{i\ell})); \ \ell = 1, 2$$

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

and we treat the terms $T_{5n}^{(\ell)}$ for $\ell = 1, \ldots, 4$ separately. Recalling the notation (5.12) we have for the first term

$$\begin{aligned} &\frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)(\widehat{q}_1^{-i}(\tau|X_{i1}) - q_{1,n}(\tau|X_{i1})) \\ &+ \frac{1}{n(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)(q_{1,n}(\tau|X_{i1}) - q_1(\tau|X_{i1})) \\ &=: T_{5n}^{(1.1)} + T_{5n}^{(1.2)} = T_{5n}^{(1)} + o\Big(\frac{1}{\sqrt{n}}\Big) \end{aligned} \tag{7.15}$$

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) yields

$$\begin{aligned} T_{5n}^{(1.1)} =& \frac{1}{n^2(n-1)} \sum_{i \neq j} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j) \sum_{k=1}^{n} \Big[\widehat{Q}_{1,n}^{-i}(\tau|(X_{i1}, X_{k2})) - Q_{1,n}(\tau|(X_{i1}, X_{k2}))\Big] \\ =& \frac{-1}{n^3(n-1)} \sum_{i \neq j} \sum_{k=1}^{n} \sum_{l=1}^{n} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j) \\ &\times K_{b_n}\Big(F\Big(\frac{l}{n}|(X_{i1}, X_{k2})\Big) - \tau\Big)\Big(\widehat{F}_1^{-i}\Big(\frac{l}{n}|(X_{i1}, X_{k2})\Big) - F\Big(\frac{l}{n}|(X_{i1}, X_{k2})\Big)\Big) + o\Big(\frac{1}{\sqrt{n}}\Big) \\ =& -\frac{1}{n^2(n-1)^2} \sum_{i \neq j} \sum_{k=1}^{n} \sum_{l \neq i}^{n} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j) \end{aligned}$$

$$
\begin{aligned}
&\times \int_0^1 K_{b_n}\left(F\left(t|(X_{i1},X_{k2})\right)-\tau\right) K_{1,h_1}(X_{i1}-X_{l1})K_{2,h_2}(X_{k2}-X_{l2})\\
&\times \frac{I\{Y_l\le t\}-F(t|(X_{i1},X_{k2}))}{f(X_{i1},X_{k2})}dt+o\Big(\frac{1}{\sqrt{n}}\Big)\\
&= Z_n^{(1)}+\tilde{Z}_n^{(2)}+o\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

where $Z_n^{(1)}$ is defined in (7.2) and

$$
\begin{aligned}
\tilde{Z}_n^{(2)} =& \frac{-1}{n^3(n-1)}\sum_{i\neq j}\sum_{k=1}^n\sum_{l\neq i}^n L_g\left(X_i-X_j\right)F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)\\
&\times \int_0^1 K_{b_n}\left(F\left(t|(X_{i1},X_{k2})\right)-\tau\right) K_{1,h_1}(X_{i1}-X_{l1})K_{2,h_2}(X_{k2}-X_{l2})\frac{s(t|X_l)\varepsilon_l}{f(X_{i1},X_{k2})}dt\\
=& Z_n^{(2)}+o\Big(\frac{1}{\sqrt{n}}\Big).
\end{aligned}
$$

Here the first equality is a direct consequence of the assumption (3.9) and the last equality follows recalling the definition of $Z_n^{(2)}$, (3.10) and a standard argument. Similarly we obtain for the statistic $T_n^{(1.2)}$ defined in (7.15)

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

where $Z_n^{(3)}$ is defined in (7.4). The statistic $T_{5n}^{(2)}$ is treated similarly and we obtain the representation

$$
T_{5n}^{(2)}=\sum_{j=4}^6 Z_n^{(j)}
$$

where $Z_n^{(4)}$, $Z_n^{(5)}$ and $Z_n^{(6)}$ are defined in (7.5), (7.6) and (7.7), respectively. The terms $Z_n^{(7)},\dots Z_n^{(12)}$ in Lemma 7.1 correspond to the statistics $T_{5n}^{(3)}$ and $T_{5n}^{(4)}$ in the decomposition (7.1) 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 $\hat{F}_l$ in (2.5) and using a Riemann approximation and a Taylor expansion we have

$$
T_{5n}^{(3)}=(T_{5n}^{(3.1)}+T_{5n}^{(3.2)})(1+o(1)). \tag{7.16}
$$

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

$$
\begin{aligned}
T_{5n}^{(3.1)} =& \frac{-1}{2n^4(n-1)} \sum_{i\neq j}^{n} \sum_{k,l,m=1}^{n} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j) \\
& \times \Big(\int_{-\infty}^{\tau} K_{b_n}\left(F\left(\frac{m}{n}|(X_{k1}, X_{l2})\right) - u\right) du - c(\tau)\Big) \\
=& \frac{-1}{2n^3(n-1)} \sum_{i\neq j}^{n} \sum_{k,l=1}^{n} L_g\left(X_i - X_j\right) F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_j)(Q(\tau|(X_{k1}, X_{l2})) - c(\tau)) + o\Big(\frac{1}{\sqrt{n}}\Big) \\
=& \frac{-1}{2n^2} \sum_{k,l=1}^{n} E[F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_i)f(X_i)](Q(\tau|(X_{k1}, X_{l2})) - q_1(X_{k1})) \\
& - \frac{1}{2n} \sum_{k}^{n} E[F'(Q_{add}(\tau|X_i)|X_i)\Delta(X_i)f(X_i)](q_1(X_{k1}) - c(\tau)) + o\Big(\frac{1}{\sqrt{n}}\Big) \\
=& \frac{1}{2} Z_n^{(9)} + o\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

where $Z_n^{(9)}$ is defined in (7.10) 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 (7.16) is given by

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

where $Z^{(7)}$ and $Z^{(8)}$ are defined in (7.8) and (7.9), 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

$$
T_{5n}^{(4)} = Z_n^{(10)} + \frac{1}{2} Z_n^{(8)} + \frac{1}{2} Z_n^{(9)} + o\Big(\frac{1}{\sqrt{n}}\Big)
$$

where the terms $Z_n^{(j)}$ are again defined in Lemma 7.1. This completes the proof of the assertion.

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

References

Abramovich, F., De Feis, I., and 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., and 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., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

De Gooijer, J. G. and 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., and Munk, A. (2002). A test for additivity in nonparametric regression. *Annals of the Institute of Statistical Mathematics*, 54:60–82.

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

Dette, H. and Scheder, R. (2010). A finite sample comparison of nonparametric estimates of the effective dose in quantal bioassay. *Journal of Statistical Computation and Simulation*, 80(5):527–544.

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

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

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

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

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

Fan, Y. and 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.

Gozalo, P. L. and 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., Wolff, R. C. L., and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association*, 94(445):154–163.

Härdle, W., Jeong, K., and Song, R. (2012). A consistent nonparametric test for causality in quantile. *Econometric Theory, to appear*.

Hengartner, N. W. and 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. and 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*. Cambrige University Press, New York.

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

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

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

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

Nolan, D. and 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. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes. Springer Series in Statistics*. Springer, New York.

Yu, K. and 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. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association*, 93(441):228–237.

Zhang, C. and 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.