Ann Inst Stat Math (2011) 63:245–265
DOI 10.1007/s10463-009-0225-5

Estimation of additive quantile regression

Holger Dette · Regine Scheder

Received: 25 December 2007 / Revised: 17 November 2008 / Published online: 20 March 2009

Abstract We consider the nonparametric estimation problem of conditional regression quantiles with high-dimensional covariates. For the additive quantile regression model, we propose a new procedure such that the estimated marginal effects of additive conditional quantile curves do not cross. The method is based on a combination of the marginal integration technique and non-increasing rearrangements, which were recently introduced in the context of estimating a monotone regression function. Asymptotic normality of the estimates is established with a one-dimensional rate of convergence and the finite sample properties are studied by means of a simulation study and a data example.

Keywords Conditional quantiles · Additive models · Marginal integration · Non-increasing rearrangements

1 Introduction

Regression techniques are widely used to quantify the relation between a response and a predictor. While ordinary least squares regression refers to the conditional mean, quantile regression was introduced by Koenker and Bassett (1978) to obtain a more sophisticated picture of the relation between the response and covariates. Since the seminal paper of these authors numerous scientists have worked on methodological and practical aspects of this method and the interested reader is referred to the recent monograph of Koenker (2005). Nonparametric methods for estimating conditional

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

R. Scheder
e-mail: regine.scheder@rub.de

quantiles have lately found considerable interest in the literature (see e.g. Yu and Jones 1997, 1998). These authors concentrate on a univariate predictor, and it is well known that for high-dimensional covariates nonparametric methods suffer from the curse of dimensionality, which does not allow precise estimation of conditional quantiles with reasonable sample sizes. For this reason several authors have recommended to use additive quantile models of the form

$$Q(\alpha|\mathbf{x}) = \sum_{k=1}^{d} Q_k(\alpha|x_k) + c(\alpha) \tag{1}$$

where $\alpha \in (0, 1)$ (see Doksum and Koo 2000; De Gooijer and Zerom 2003; Horowitz and Lee 2005, among others). In (1) the quantity $c(\alpha)$ denotes a constant, $\mathbf{x} = (x_1, \ldots, x_d)^T$ are the predictors, and $Q_k(\alpha|x_k)$ are functions relating the α-quantile of the conditional distribution functions to each coordinate of the predictor [note that these have to be normalized in order to make the model (1) identifiable—see the discussion in Sect. 2]. So far several authors have proposed methods for estimating the additive components in (1). Doksum and Koo (2000) suggest a spline estimate but do not provide rates of convergence of their estimator. De Gooijer and Zerom (2003) use a marginal integration estimate, while 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 in a second step by the local polynomial technique (see Fan and Gijbels 1996).

In the present paper we propose an alternative estimate of conditional quantiles in the additive model (1) which could be considered as continuation of the work of De Gooijer and Zerom (2003). Our investigations are motivated by the observation that in the one-dimensional case many nonparametric estimates yield crossing quantile curves (see e.g. He 1997; Yu et al. 2003; Koenker 2005, Chap. 7). In the context of estimating a conditional quantile curve in the additive model (1), the situation is similar, but the focus lies on the marginal effects of the conditional quantile function. The marginal effect of the conditional quantile function is the additive component $Q_k(\alpha|x_k)$ with respect to a certain covariate x_k plus the constant term $c(\alpha)$. Throughout this paper the marginal effect will be denoted by $q_k(\alpha|x_k)$, and we note that by its definition q_k is a monotone function of α. Because Horowitz and Lee (2005) use a parametric fit based on the check function the resulting estimators of the marginal effects of the conditional quantile function are not necessarily monotone with respect to α, and a similar comment applies to the method proposed by Doksum and Koo (2000). On the other hand the procedure proposed by De Gooijer and Zerom (2003) uses a reweighted Nadaraya–Watson estimate to estimate the conditional distribution function in a first step, which is finally inverted in a second step. The estimate of the conditional distribution function is monotone if $d < 5$ and positive kernels are used. However, if $d \geq 5$ negative kernels are required in order to address bias problems in the marginal integration method. In this case the estimate of the empirical distribution function is not necessarily monotone, and the quantile curves cannot be obtained by a straightforward inversion.

It is the purpose of the present paper to study some theoretical properties of an alternative estimate of the additive conditional quantile model. The method is based on a combination of the marginal integration technique (see Linton and Nielsen 1995) with the concept of non-increasing rearrangements (see Bennett and Sharpley 1988). This methodology has been successfully applied by Dette and Volgushev (2008) and Chernozhukov et al. (2007). The last named authors use the concept of non-increasing rearrangements to isotonize parametric (possibly crossing) quantile estimates and study the weak convergence of the resulting statistics. Dette and Volgushev (2008) concentrate on the case of a one-dimensional covariate and isotonize and invert a non-parametric estimate of the conditional distribution function simultaneously in order to obtain nonparametric non-crossing estimates of quantile curves. In Sect. 2, we describe the main concept of the method for estimating the marginal effects in an additive quantile regression model. We apply a kernel distribution function estimator to a set of not necessarily monotone values obtained from an estimate of the conditional distribution function, which directly yields a monotone quantile function estimator. This method is then combined with the marginal integration technique. Our approach is applicable to any parametric or nonparametric estimate of the conditional distribution function. In Sect. 3, which contains the main contribution of the paper, we state the asymptotic distributional properties of the new statistic, if the conditional distribution function is estimated by local constant or local linear techniques. In Sect. 4 we present a small simulation study to illustrate the finite sample properties of the new method. Finally, some of the technical details of the proofs of the asymptotic results are presented in Sect. 5.

2 Monotone rearrangements and marginal integration

Let $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ denote a sample of independent and identically distributed observations, where the d-dimensional random variable $\mathbf{X}_j = (X_{j1}, \ldots, X_{jd})^{\mathrm{T}}$ has a q times continuously differentiable density, say p, with compact support $[0, 1]^d$. Following Hall et al. (1999), we introduce the random variable $Z_j = I\{Y_j \le y\}$ with

$$E[Z_j|\mathbf{X}_j = \mathbf{x}] = P(Y_j \le y|\mathbf{X}_j = \mathbf{x}) = F(y|\mathbf{x})$$

and the nonparametric regression model

$$Z_j = F(y|\mathbf{X}_j) + \sigma(y|\mathbf{X}_j)\varepsilon_j \quad j = 1, \ldots, n \tag{2}$$

where $E[\varepsilon_j] = 0$, $\mathrm{Var}(\varepsilon_j) = 1$, and $E[\varepsilon_j^4] \le c < \infty$. The variance function $\sigma(y|\mathbf{x})$ can be further specified in terms of $F(y|\mathbf{x})$, i.e.

$$\sigma^2(y|\mathbf{x}) = E[(Z_j - F(y|\mathbf{x}))^2|\mathbf{X}_j = \mathbf{x}] = F(y|\mathbf{x})(1 - F(y|\mathbf{x})).$$

We consider the model (1) and add the conditions

$$E[Q_k(\alpha|X_{jk})] = 0, \quad k = 1, \ldots, d, \ (j = 1, \ldots, n) \tag{3}$$

in order to make the components of the additive decomposition (1) identifiable. Note that this normalization yields monotone (with respect to α) marginal effects $q_k(\alpha|x_k) = Q_k(\alpha|x_k) + c(\alpha) = E[Q(\alpha|\mathbf{X_j})|X_{jk} = x_k]$.

Let $\hat{F}(y|\mathbf{x})$ denote an estimate of the conditional distribution function $F(y|\mathbf{x}) = P(Y_j \le y \mid \mathbf{X}_j = \mathbf{x})$, which will be specified below. Define $H : \mathbb{R} \to [0, 1]$ as a strictly increasing distribution function, which will be used as a transformation to the compact interval $[0, 1]$, since $F(\cdot|\mathbf{x})$ might have unbounded support. Note that $\hat{F}(y|\mathbf{x})$ is obtained by nonparametric methods and for this reason is usually not increasing which yields some difficulties in the determination of the corresponding quantiles. In the following, we solve this problem and the problem of inversion simultaneously using the concept of monotone rearrangements (see Bennett and Sharpley 1988), which was introduced in the context of estimating a monotone regression function by Dette et al. (2006). To be precise, let K_d denote a positive kernel function with compact support $[-1, 1]$ and h_d denote a bandwidth, then we define

$$\hat{H}_I(\alpha|\mathbf{x}) = \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{\alpha} K_d\left(\frac{\hat{F}(H^{-1}(\frac{i}{N})|\mathbf{x}) - u}{h_d}\right) \mathrm{d}u. \tag{4}$$

If $\hat{F}(y|\mathbf{x})$ is uniformly consistent and $N \to \infty$, $h_d \to 0$, it is intuitively clear that

$$\begin{aligned}\hat{H}_I(\alpha|\mathbf{x}) \approx H_N(\alpha|\mathbf{x}) &:= \frac{1}{Nh_d} \sum_{i=1}^{N} \int_{-\infty}^{\alpha} K_d\left(\frac{F(H^{-1}(\frac{i}{N})|\mathbf{x}) - u}{h_d}\right) \mathrm{d}u \\ &\approx \int I\{F(H^{-1}(s)|\mathbf{x}) \le \alpha\} \mathrm{d}s = H(Q(\alpha|\mathbf{x})),\end{aligned} \tag{5}$$

where $Q(\alpha|\mathbf{x}) = F^{-1}(\alpha|\mathbf{x})$. Consequently, we define

$$\hat{Q}_I(\alpha|\mathbf{x}) = H^{-1}(\hat{H}_I(\alpha|\mathbf{x})) \tag{6}$$

as the estimate of the conditional quantile $Q(y|\mathbf{x})$, and

$$Q_N(\alpha|\mathbf{x}) = H^{-1}(H_N(\alpha|\mathbf{x})) \tag{7}$$

as an approximation of the conditional quantile $Q(y|\mathbf{x})$. It will be demonstrated in the following section that the choice of the function H has no impact on the asymptotic properties of the estimate. Moreover, even for realistic sample sizes the impact of the choice of H is negligible and a practical recommendation regarding this choice will be given in Sect. 4.

Note that (4) provides a simultaneous inversion and isotonization of the not necessarily monotone estimate of the conditional distribution function (see also Dette and Volgushev 2008). Consequently, the estimate $\hat{H}_I$ is monotone with respect to α provided that the kernel K_d is positive on its support, which will be assumed throughout this paper. In the next step, we now apply the marginal integration technique (see Linton and Nielsen 1995; Chen et al. 1996; Hengartner and Sperlich 2005) to obtain an estimator in the model (1). Without loss of generality, we focus on the problem of estimating the first component $Q_1(\alpha|x_1)$ in model (1) and the marginal effect of the first covariate, respectively. We introduce the notations $X_{j\underline{1}} = (X_{j2}, \ldots, X_{jd})^T$, and $\mathbf{x} = (x_1, x_{\underline{1}})^T$. Now we define the marginal integration estimator of the first marginal effect

$$\hat{q}_1(\alpha|x_1) = \frac{1}{n}\sum_{j=1}^{n} \hat{Q}_I(\alpha|x_1, X_{j\underline{1}}), \tag{8}$$

which can be regarded as the expectation of $\hat{Q}_I(\alpha|\mathbf{X})$ with respect to the empirical distribution of $X_{\underline{1}} = (X_2, \ldots, X_d)^T$. This estimator is obviously monotone in α for fixed x_1. Note that by the strong law of large numbers and from the normalizing condition (3), we have

$$\frac{1}{n}\sum_{k=1}^{n} Q_N(\alpha|x_1, X_{k\underline{1}}) \overset{\text{a.s.}}{\longrightarrow} \int Q(\alpha|\mathbf{x})p_{\underline{1}}(x_{\underline{1}})\mathrm{d}x_{\underline{1}} = Q_1(\alpha|x_1)+c(\alpha)$$
$$=: q_1(\alpha|x_1), \tag{9}$$

where $p_{\underline{1}}$ denotes the marginal density of $X_{\underline{1}} = (X_2, \ldots, X_d)^T$. Consequently, if $\hat{Q}_I(\alpha|\mathbf{x})$ is a (uniformly) consistent estimate of $Q(\alpha|\mathbf{x})$ it follows that $\hat{q}_1(\alpha|x_1)$ is a consistent estimate of $q_1(\alpha|x_1) := Q_1(\alpha|x_1) + c(\alpha)$. Finally,

$$\hat{q}_1(\alpha|x_1) - \frac{1}{n}\sum_{i=1}^{n} \hat{q}_1(\alpha|X_{i1}) \tag{10}$$

defines a consistent estimate of $Q_1(\alpha|x_1)$ (note that $\frac{1}{n}\sum_{i=1}^{n} Q_1(\alpha|X_{i1}) \overset{\text{a.s.}}{\to} E[Q_1(\alpha|X_{i1})] = 0$). The estimates of the other components are defined in exactly the same way, and the final estimate in the additive model (1) is given by

$$\hat{Q}_{\text{add}}(\alpha|\mathbf{x}) := \sum_{k=1}^{d} \hat{q}_k(\alpha|x_k) - \left(1 - \frac{1}{d}\right)\sum_{k=1}^{d} \frac{1}{n}\sum_{i=1}^{n} \hat{q}_k(\alpha|X_{ik}). \tag{11}$$

In the following section, we study the asymptotic properties of the estimate $\hat{q}_1(\alpha|x_1)$ for the marginal effect of x_1 of the conditional quantile function. The corresponding properties of the estimate $\hat{q}_k(\alpha|x_k)$ for $k = 2, \ldots, d$ follow in a straightforward manner.

Remark 1 Note that the additive estimate $\hat{Q}_{\text{add}}$ is not necessarily monotone with respect to α. However, a monotone and asymptotically equivalent estimate could be easily obtained by applying a further monotone rearrangement as proposed by Dette et al. (2006) or Chernozhukov et al. (2007).

Remark 2 Note that the estimation $q_1(\alpha, x_1)$ requires the choice of $d + 1$ bandwidths, the bandwidth h_1 for the covariate of interest, $g_2, \ldots, g_d$ for the nuissance directions and the bandwidth h_d in definition (4). As pointed out by Dette et al. (2006)

and Dette and Volgushev (2008) the estimate $\hat{Q}_I$ is rather robust with respect to the choice h_d, as long as it is not chosen too large. On the other hand, the choice of the bandwidths in marginal integration estimation $h_1, g_2 \ldots, g_d$, is more delicate and has been discussed to some extent in the literature on nonparametric additive regression (see e.g. Neumeyer and Sperlich 2006; Hengartner and Sperlich 2005). Most authors recommend cross validation. If the covariates $X_2, \ldots, X_d$ have the same range, the choice $g_2 = \cdots = g_d$ is well justified, which simplifies the problem of choosing the bandwidth substantially.

3 Asymptotic properties

A precise statement of the main results requires the specification of an initial estimate of the conditional distribution function $F(y|\mathbf{x})$. For the sake of definiteness, we first consider a Nadaraya–Watson type estimator

$$\hat{F}(y|\mathbf{x}) = \hat{F}(y|x_1, x_{\underline{1}}) = \frac{\sum_{i=1}^{n} K_{h_1}(x_1 - X_{i1}) L_G(x_{\underline{1}} - X_{i\underline{1}}) I\{Y_i \leq y\}}{\sum_{i=1}^{n} K_{h_1}(x_1 - X_{i1}) L_G(x_{\underline{1}} - X_{i\underline{1}})}. \quad (12)$$

The kernel K in (12) is a one-dimensional kernel with compact support, say $[-1, 1]$, and existing second moments satisfying

$$\int_{-1}^{1} x K(x) \mathrm{d}x = 0, \ \frac{1}{2} \int_{-1}^{1} x^2 K(x) \mathrm{d}x = \kappa_2(K). \quad (13)$$

Let $\nu_{\underline{1}} = (\nu_2, \ldots, \nu_d)$ be a multiindex of integers with $\nu_i \geq 0$, so that $x_{\underline{1}}^{\nu_{\underline{1}}} = x_2^{\nu_2} \ldots x_d^{\nu_d}$. Moreover, define $|\nu_{\underline{1}}| = \sum_{i=2}^{d} \nu_i$. The kernel L in (12) refers to a $(d-1)$-dimensional kernel of order q supported on $[-1, 1]^{d-1}$, i.e., L satisfies the conditions

(i) $\int_{[-1,1]^{d-1}} L(x_{\underline{1}}) \mathrm{d}x_{\underline{1}} = 1$,
(ii) $\int_{[-1,1]^{d-1}} |x_{\underline{1}}^{\nu_{\underline{1}}}||L(x_{\underline{1}})| \mathrm{d}x_{\underline{1}} < \infty$ for $|\nu_{\underline{1}}| \leq q$,
(iii) $\int_{[-1,1]^{d-1}} x_{\underline{1}}^{\nu_{\underline{1}}} L(x_{\underline{1}}) \mathrm{d}x_{\underline{1}} = 0$ for $1 \leq |\nu_{\underline{1}}| \leq q - 1$,
(iv) $\int_{[-1,1]^{d-1}} x_{\underline{1}}^{\nu_{\underline{1}}} L(x_{\underline{1}}) \mathrm{d}x_{\underline{1}} \neq 0$ for some $|\nu_{\underline{1}}| = q$.

Note that the kernel L is for $q > 2$ not a probability density function any more. The bandwidth h_1 corresponds to the first covariate. Denote $K_{h_1}(\cdot) = \frac{1}{h_1} K(\cdot / h_1)$ and

$$L_G(\mathbf{x}) = \frac{1}{\det(G)} L(G^{-1}\mathbf{x})$$

for the bandwidth matrix $G = \mathrm{diag}(g_2, \ldots, g_d) \in \mathbb{R}^{(d-1)\times(d-1)}$, where g_k refers to the bandwidth of the kth coordinate $(k = 2, \ldots, d)$.

Throughout this paper we make the following basic assumptions regarding the underlying model

$$\mathbf{X}_j \quad \text{has a positive density } p \text{ with supp}\,(p) = [0,1]^d, \quad p \in C^q([0,1]^d), \tag{14}$$

$$\text{for any } y \in \mathbb{R} \quad F(y|\cdot) \in C^q([0,1]^d), \tag{15}$$

$$F(\cdot|\mathbf{x}) \in C^1([0,1]) \quad \text{and } Q'(\alpha|\mathbf{x}) > 0, \tag{16}$$

$$K_d' \text{ is Lipschitz continuous.} \tag{17}$$

In (16) the function Q' denotes the derivative of the quantile function $Q(\alpha|\mathbf{x})$ with respect to the variable α (and its existence in a neighborhood of the quantile of interest is assumed throughout this paper), while the partial derivatives with respect to the coordinates of the predictor $\mathbf{x} = (x_1, \ldots, x_d)^{\mathrm{T}}$ are denoted by $\partial^s/\partial^s x_k$ ($s = 1, \ldots, q$; $k = 1, \ldots, d$). Assumption (17) refers to the kernel used for the monotonizing inversion in (4).

In the following discussion we will investigate the asymptotic properties of the estimate $\hat{q}_1(\alpha|x_1)$ defined in (8). We focus on the marginal effect of the first component, but corresponding results for the other marginal effects can easily derived in the same way. For the sake of simplicity, we assume the same bandwidth for the remaining coordinates $x_{\underline{1}} = (x_2, \ldots, x_d)$, that is

$$g_2 = \cdots = g_d. \tag{18}$$

Regarding the bandwidths h_1, g_2 and h_d, we make the following assumptions

$$N = O(n) \tag{19}$$

$$nh_1 \to \infty, \quad ng_2^{d-1} \to \infty, \; nh_1 g_2^{d-1} \to \infty, \; nh_d \to \infty \tag{20}$$

$$nh_1^5 = O(1) \tag{21}$$

$$ng_2^{2q+1} = O(1) \tag{22}$$

$$\frac{h_d}{h_1} = o(1) \tag{23}$$

$$\frac{1}{nh_1 g_2^{2(d-1)} h_d^2} = o(1) \tag{24}$$

Our first result specifies the asymptotic properties of the estimate $\hat{q}_1(\alpha|x_1)$ defined in (8) if the Nadaraya–Watson estimator is used for estimating the conditional distribution function. The case of the local linear estimate is discussed in Theorem 2. For a precise statement of the asymptotic properties, we recall the notation $x = (x_1, x_{\underline{1}})^{\mathrm{T}}$ and obtain the following result.

Theorem 1 *If the assumptions (13)–(24) are satisfied, then we have for any* $\alpha \in (0, 1)$

$$\sqrt{nh_1}(\hat{q}_1(\alpha|x_1) - q_1(\alpha|x_1) + b_1(\alpha|x_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(\alpha|x_1)),$$

where

$$b_1(\alpha|x_1) = \kappa_2(K)h_1^2 \int \left[\frac{\partial^2}{\partial x_1^2} F(Q(\alpha|x_1, x_{\underline{1}})|x_1, x_{\underline{1}}) + 2 \frac{\frac{\partial}{\partial x_1} F(Q(\alpha|x_1, x_{\underline{1}})|x_1, x_{\underline{1}}) \frac{\partial}{\partial x_1} p(x_1, x_{\underline{1}})}{p(x_1, x_{\underline{1}})} \right] \frac{1}{F'(Q(\alpha|x_1, x_{\underline{1}})|x_1, x_{\underline{1}})} p_{\underline{1}}(x_{\underline{1}}) \mathrm{d}x_{\underline{1}},$$

$$s^2(\alpha|x_1) = \int K^2(v)\mathrm{d}v \int \frac{\alpha(1-\alpha)p_{\underline{1}}^2(x_{\underline{1}})}{(F'(Q(\alpha|x_1, x_{\underline{1}})|x_1, x_{\underline{1}}))^2 p(x_1, x_{\underline{1}})} \mathrm{d}x_{\underline{1}},$$

$$\kappa_s(K) = \frac{1}{s!} \int v^s K(v)\mathrm{d}v.$$

There are numerous alternative estimates for the conditional distribution function which could be used as initial estimate. For example, consider the case where the conditional distribution function is estimated by a local linear technique (see Masry and Fan 1997), that is

$$\tilde{F}(y|x) = e_0^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{W}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{W}I(\mathbf{Y}),$$

where $\mathbf{W} = \mathrm{diag}(K_{h_1}(x_1 - X_{i1})L_G(x_{\underline{1}} - X_{i\underline{1}}))_{i=1}^n$, $e_0 = (1, 0, \ldots, 0) \in \mathbb{R}^{d+1}$, $I(\mathbf{Y}) = (I\{Y_1 \le y\}, \ldots, I\{Y_n \le y\})^{\mathrm{T}}$ and the matrix $\mathbf{X}$ is given by

$$\mathbf{X} = \begin{pmatrix} 1 & (\mathbf{x} - \mathbf{X}_1)^{\mathrm{T}} \\ \vdots & \\ 1 & (\mathbf{x} - \mathbf{X}_n)^{\mathrm{T}} \end{pmatrix} \in \mathbb{R}^{n \times d+1}.$$

The corresponding estimate of the conditional quantile distribution and the marginal effects, say $\tilde{Q}_I(\alpha|\cdot)$ and $\tilde{q}_1(\alpha|\cdot)$, are obtained from (4), (6) and (8) where $\hat{F}$ and $\hat{Q}_I$ have to be replaced by $\tilde{F}$ and $\tilde{Q}_I$, respectively. In this case asymptotic normality of the resulting estimate is still true but the bias term has to be changed.

Theorem 2 *If the assumptions of Theorem 1 are satisfied, then*

$$\sqrt{nh_1}(\tilde{q}_1(\alpha|x_1) - q_(\alpha|x_1) + \tilde{b}_1(\alpha|x_1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(\alpha|x_1)),$$

where $s^2(\alpha|x_1)$ *is defined in Theorem 1 and*

$$\tilde{b}_1(\alpha|x_1) = \kappa_2(K)h_1^2 \int \frac{\frac{\partial^2}{\partial x_1^2} F(Q(\alpha|x_1, x_{\underline{1}})|x_1, x_{\underline{1}})}{F'(Q(\alpha|x_1, x_{\underline{1}})|x_1, x_{\underline{1}})} p_{\underline{1}}(x_{\underline{1}})\mathrm{d}x_{\underline{1}}.$$

The local linear estimator for the conditional distribution function is not necessarily monotone increasing. Using our method, this does not pose a problem for the estimation of the marginal effects of the conditional quantile $q_1(\alpha|x_1)$, since the monotonizing inversion takes care of the monotonicity of the conditional quantile function with respect to α. The proof of the last theorem follows exactly the same lines as that of Theorem 1, and is therefore omitted.

De Gooijer and Zerom (2003) applied a reweighted Nadaraya and Watson estimate for the conditional distribution function proposed by Hall et al. (1999) as preliminary estimator. This estimate comes with the superior bias of the local linear estimator and is still monotone increasing as long as positive kernels are used. Note that the bias for the additive quantile estimate derived in De Gooijer and Zerom (2003) is wrong and should be replaced by the above expression.

Remark 3 The asymptotic properties of the additive quantile function

$$\hat{Q}_{\text{add}}(\alpha|\mathbf{x}) = \sum_{k=1}^{d} \hat{q}_k(\alpha|x_k) - \left(1 - \frac{1}{d}\right) \sum_{k=1}^{d} \frac{1}{n} \sum_{j=1}^{n} \hat{q}_k(\alpha|X_{jk})$$

can be derived as well. The asymptotic bias of $\hat{Q}_{\text{add}}(\alpha|\mathbf{x})$ is

$$\sum_{k=1}^{d} b_k(\alpha|x_k) - \left(1 - \frac{1}{d}\right) \int b_k(\alpha|x_k) p_k(x_k) \mathrm{d}x_k,$$

where $b_k(\alpha|x_k)$ is the bias of $\hat{q}_k(\alpha|x_k)$. But the asymptotic variance is just the sum of the variances of $\hat{q}_k(\alpha|x_k)$, since the terms in $\hat{Q}_{\text{add}}(\alpha|\mathbf{x})$ are asymptotically uncorrelated. A similar result is obtained for the local linear estimate $\tilde{Q}_{\text{add}}$.

Remark 4 At the end of this section, some remarks about the conditions regarding the bandwidths might be appropriate. A general drawback of the marginal integration method is that for higher dimensions $d \geq 5$ the third condition in (20), namely $nh_1 g_2^{d-1} \to \infty$, is not fulfilled using the bandwidth g_2 with the rate $n^{-1/5}$. In this case, the bias-term in the directions not of interest dominates the asymptotic properties of the estimate. A way out of this problem is to take L to be a higher order kernel. For our estimators $\hat{q}_1(\alpha|x_1)$ and $\tilde{q}_1(\alpha|x_1)$ we have to deal with an additional bandwidth which yield that even for dimension $d = 2$ higher order kernel must be used or one has to weaken condition (23) and accept an extra bias-term. However, in the stated Theorems we over-smooth the variables not of interest by taking g_2 at the rate $n^{\frac{1}{2q+1}}$ for $q > 2$. This method still demonstrates simplicity and flexibility in its usage which is illustrated in more detail in the following section. Furthermore, even for higher dimensions the estimator $\hat{q}_1(\alpha|x_1)$ of the marginal effects of x_1 is monotone in α. Compared to De Gooijer and Zerom (2003), higher order kernels do not pose any problems, which means that our method can be applied for $d \geq 5$, or even for $d < 5$ with higher order kernels in order to reduce the bias.

4 Finite sample properties and data analysis

In this section, we investigate the finite sample properties of the new estimator $\hat{q}_1(\alpha|x_1)$. We also consider the procedure of De Gooijer and Zerom (2003), which is most similar in spirit with the procedure proposed in this paper and serves as a benchmark. Their estimator, which they call an average quantile estimator, is the inverse of the reweighted Nadaraya–Watson estimator for a conditional distribution function introduced by Hall et al. (1999). The reweighted estimator for the conditional distribution function has the following form

$$\breve{F}(y|\mathbf{x}) = \sum_{i=1}^{n} \frac{p_i(\mathbf{x}) K_{h_1}(x_1 - X_{i1}) L_G(x_{\underline{1}} - X_{i\underline{1}}) I\{Y_i \le y\}}{\sum_{i=1}^{n} p_i(\mathbf{x}) K_{h_1}(x_1 - X_{i1}) L_G(x_{\underline{1}} - X_{i\underline{1}})},$$

where $p_i(\mathbf{x})$ denote weights depending on the data $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with the properties that $p_i \ge 0$ for $i = 1, \ldots, n$, $\sum_{i=1}^{n} p_i = 1$, and

$$\sum_{i=1}^{n} p_i(\mathbf{x})(X_{i1} - x_1) K_{h_1}(x_1 - X_{i1}) L_G(x_{\underline{1}} - X_{i\underline{1}}) = 0.$$

The weights p_i are chosen by maximizing $\prod_{i=1}^{n} p_i$. The estimator $\breve{F}(y|\mathbf{x})$ conserves the positivity and the monotonicity of the Nadaraya–Watson estimator with positive kernel functions K and L, but provides the more attractive bias of the local linear estimator. Since $\breve{F}(y|\mathbf{x})$ is monotone increasing, the inversion is easily obtained. As pointed out in the previous discussion, this monotonicity can only be achieved if $d < 5$. Otherwise, a negative kernel (of higher order) has to be used in order to decrease the bias (see De Gooijer and Zerom 2003), which could yield some problems in the inversion of $\breve{F}(y|\mathbf{x})$. For our approach, the monotonicity is not a constraint for the estimators of conditional distribution functions. Therefore local linear methods can also be used to estimate the conditional distribution function. In the following, we will illustrate some finite sample properties of the different estimators for the marginal effect of the additive quantile regression model. Instead of using the Nadaraya and Watson estimate for the conditional distribution function, we use the local linear estimator and apply afterwards the monotonizing inversion and the marginal integration technique. We call this estimator for the marginal effect of the first variable $\tilde{q}_1(\alpha|x_1)$ and the estimate for the additive component $\tilde{Q}_1(\alpha|x_1)$. The estimator proposed by De Gooijer and Zerom (2003) is denoted by $\breve{q}_1(\alpha|x_1)$ and $\breve{Q}_1(\alpha|x_1)$, respectively. For the sake of practical convenience, we use a uniform distribution function on the interval

$$[\min(Y_i), \max(Y_i)]$$

for the function H to transform the data to [0, 1].

Example 1 We consider the two-dimensional model

$$Y = 0.75X_1 + 1.5\sin(0.5\pi X_2) + 0.25\varepsilon, \tag{25}$$

where $\varepsilon \sim \mathcal{N}(0, 1)$. We assume that the covariates $(X_1, X_2)^T$ are bivariate normal with mean 0, variance 1, and correlation ρ. For the correlation, we distinguish two cases: a weak correlation $\rho = 0.2$ and a strong correlation $\rho = 0.8$. This experiment was originally carried out by De Gooijer and Zerom (2003). The Epanechnikov kernel is used to estimate the conditional distribution function and to compute the monotonizing inversion, i.e.

$$K(x) = L(x) = K_d(x) = \frac{3}{4}(1 - x^2) I_{[-1,1]}(x).$$

We choose the bandwidths as in De Gooijer and Zerom (2003): $h_1 = 3\hat{\sigma}_1 n^{-1/5}$ for X_1 and $h_2 = \hat{\sigma}_2 n^{-1/5}$ for X_2, where $\hat{\sigma}_i$ is the standard deviation of the corresponding covariate. The bandwidth h_d was chosen as $h_d = 0.3$, but the quantile estimates are not very sensitive with respect to choice of this bandwidth if it is chosen sufficiently small (see e.g. Dette and Volgushev 2008, who considered the case of a one-dimensional covariate). The quantile estimates for $\alpha = 0.5$ and the sample sizes $n = 100, 200$, and 400 are shown in the left part of Table 1, which also contains the results of De Gooijer and Zerom (2003) as a benchmark. Instead of using the mean squared error, the mean absolute deviation error (MADE) is collected, whereas observations outside of the square $[-2, 2]^2$ are disregarded to avoid boundary effects. In order to make our results comparable to De Gooijer and Zerom (2003) 41 simulation runs have been performed in each scenario for the calculation of the new estimate. The local linear estimate has a smaller MADE than the estimate of De Gooijer and Zerom (2003), except in the case $\rho = 0.8$, $n = 100, 200$ and 400 for $Q_1(0.5|x_1) = 0.75x_1$, and $\rho = 0.2$, $n = 100$ for $Q_1(0.5|x_1) = 0.75x_1$, and $n = 400$ for $Q_2(0.5|x_2) = \sin(0.5\pi x_2)$, where the difference is rather small. It is notable that the standard deviations of the new procedure are substantially smaller than those obtained from the procedure of De Gooijer and Zerom (2003). In the right part of Table 1 we show the MADE of the estimate proposed in this paper for the marginal effects in the quantile regression corresponding to the $\alpha = 0.25$ and $\alpha = 0.75$ quantile. We observe that the more extreme quantile curves are estimated with less precision for the second component $1.5 \sin(0.5\pi x_2)$. On the other hand, the estimates for the first component $0.75x_1$ for the quantile curves have a smaller MADE than the estimate of the median curve. It is also remarkable that the standard deviation of the local linear estimates for the marginal effects of the quantile function is very similar in all cases. As pointed out by a referee, it might be of interest to investigate the performance of the cross validation procedure in this context. In Table 2 we show the corresponding results for the local linear estimate and the sample size $n = 100$. We observe that for the linear term $Q_1(0.5|x_1) = 0.75x_1$ cross validation has a worse performance than the bandwidth choice proposed by De Gooijer and Zerom (2003), while a substantial improvement can be obtained for the component $Q_2(0.5|x_2) = 1.5 \sin(0.5\pi x)$.

Example 2 To illustrate the performance on a real data set, we estimate the marginal effects for the Boston housing data which has been discussed by several authors (see e.g. Harrison and Rubinfeld 1978; Belsley et al. 1980, among others). The Boston

Table 1 The mean absolute deviation error of the different approaches (standard deviation in brackets)

ρ	n	Component	$\tilde{Q}_k(0.5\|x_k)$ (local linear)	$\breve{Q}_k(0.5\|x_k)$ (De Gooijer et. al)	$\tilde{Q}_k(0.25\|x_k)$ (local linear)	$\tilde{Q}_k(0.75\|x_k)$ (local linear)
0.2	100	$0.75x_1$	0.1176 (0.0424)	0.1374 (0.0597)	0.1019 (0.0399)	0.1105 (0.0394)
		$1.5\sin(0.5\pi x_2)$	0.2112 (0.0493)	0.1818 (0.1425)	0.2923 (0.0399)	0.2986 (0.0367)
	200	$0.75x_1$	0.0630 (0.0226)	0.1066 (0.0511)	0.0690 (0.0217)	0.0813 (0.0288)
		$1.5\sin(0.5\pi x_2)$	0.0969 (0.0376)	0.1272 (0.1120)	0.2285 (0.0267)	0.2532 (0.0243)
	400	$0.75x_1$	0.0474 (0.0211)	0.0734 (0.0431)	0.0465 (0.0207)	0.0598 (0.0124)
		$\sin(0.5\pi x_2)$	0.1169 (0.0191)	0.0936 (0.0889)	0.1862 (0.0211)	0.2161 (0.0178)
0.8	100	$0.75x_1$	0.1939 (0.0491)	0.1365 (0.1124)	0.1880 (0.0627)	0.1603 (0.0426)
		$1.5\sin(0.5\pi x_2)$	0.2801 (0.0714)	0.4865 (0.1783)	0.3200 (0.0531)	0.2851 (0.0637)
	200	$0.75x_1$	0.1882 (0.0487)	0.1093 (0.1263)	0.1780 (0.0443)	0.1615 (0.0540)
		$1.5\sin(0.5\pi x_2)$	0.2305 (0.0526)	0.4350 (0.1767)	0.3135 (0.0482)	0.2439 (0.0385)
	400	$0.75x_1$	0.1829 (0.0409)	0.0985 (0.1099)	0.1668 (0.0342)	0.1504 (0.0323)
		$\sin(0.5\pi x_2)$	0.2152 (0.0435)	0.4009 (0.1467)	0.2804 (0.0383)	0.2078 (0.0308)

Table 2 The mean absolute deviation error of the estimate $\tilde{Q}_k(0.5|x_k)$ (standard deviation in brackets), if the bandwidths are chosen by least squares cross validation

The sample size is $n = 100$

ρ	Component	$\tilde{Q}_k(0.5\|x_k)$ (local linear)
0.2	$0.75x_1$	0.1379 (0.1021)
	$1.5\sin(0.5\pi x_2)$	0.2029 (0.0891)
0.8	$0.75x_1$	0.2619 (0.0891)
	$1.5\sin(0.5\pi x_2)$	0.2071 (0.0632)

housing data contains the housing values of suburbs of Boston and 13 variables/criteria, 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. Following De Gooijer and Zerom (2003) we focus on four covariates

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

which turn out to be very significant, if a linear model is fitted to the data (see Belsley et al. 1980, p. 232). For a fit of an additive conditional quantile model we applied cross validation to determine the bandwidth for the four different variables. To simplify this problem, we set

$$h_1 = \hat{\sigma}_{\text{crime}} k, \quad h_2 = \hat{\sigma}_{\text{rooms}} k, \; h_3 = \hat{\sigma}_{\text{distance}} k, \; h_4 = \hat{\sigma}_{\text{econstatus}} k,$$

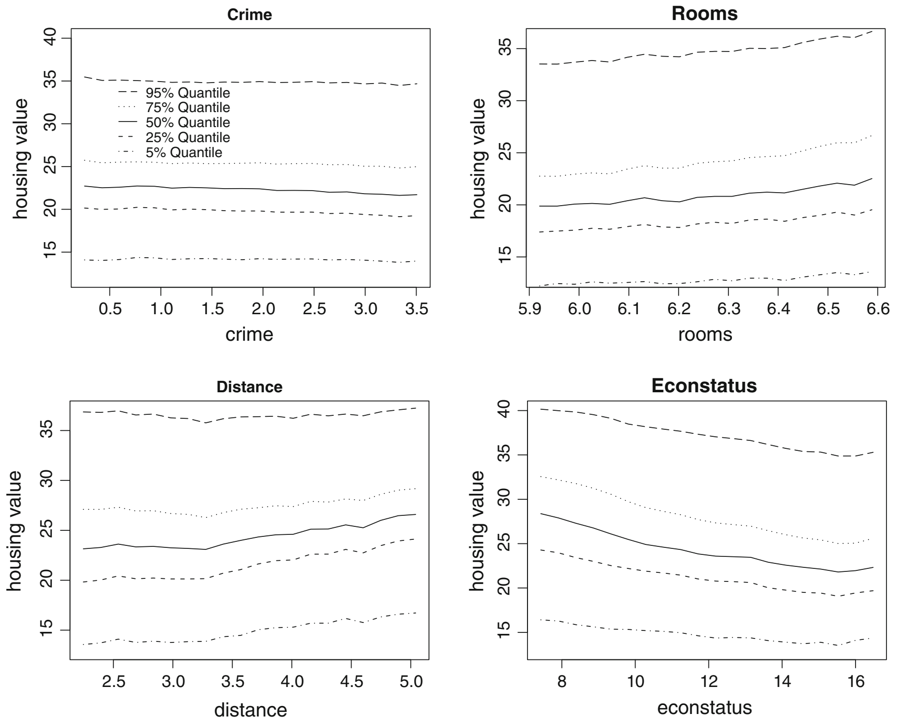

Fig. 1 Boston housing data set: the marginal effects at five different levels of α obtained by the new estimate proposed in this paper

where $\hat{\sigma}$ is the standard deviation of the corresponding variables. The cross validation criterion is minimized for $k \in [1/11, 30/11]$ for each marginal effect separately. Since the values are quite similar, we set $k = 1$ for all covariates. In Figure 1 we display five different curves of the marginal effects $Q(\alpha|X_k)$ for fixed $\alpha = 0.05, 0.25, 0.5, 0.75, 0.95$. Figure 2 shows the corresponding results for the estimate of De Gooijer and Zerom (2003). Note that the marginal effects are monotone in α for both estimates.

5 Appendix: proof of main results

Proof of Theorem 1 For the sake of simplicity, we assume that $N = n$ and that the distribution function H corresponds to the uniform distribution. Recall the definition of q_1 in (9) and of $Q_n(\alpha|\mathbf{x})$ in (7), then it is easy to see that by the law of the iterated logarithm and the bandwidth conditions (23)

$$q_1(\alpha|x_1) = q_{1,n}(\alpha|x_1) + o\left(\frac{1}{\sqrt{nh_1}}\right),$$

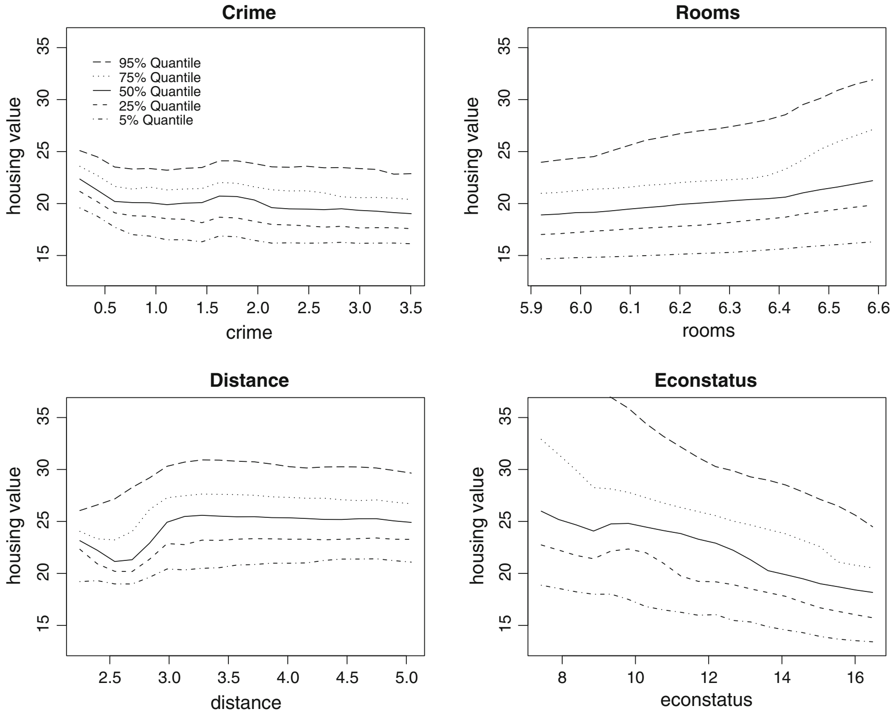

Fig. 2 Boston housing data set: The marginal effects at five different levels of α obtained by the estimate of De Gooijer and Zerom (2003)

where

$$q_{1,n}(\alpha|x_1) = \frac{1}{n}\sum_{i=1}^{n} Q_n(\alpha|x_1, X_{i\underline{1}}).$$

Now a straightforward argument shows that the assertion follows from the weak convergence

$$\sqrt{nh_1}(\hat{q}_1(\alpha|x_1) - q_{1,n}(\alpha|x_1) + b_1(\alpha|x_1)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(\alpha|x_1)), \tag{26}$$

where the bias $b_1(\alpha|x_1)$ and the variance $s^2(\alpha|x_1)$ are defined in Theorem 1. For a proof of (26) we use a Taylor expansion and obtain

$$\begin{aligned}\hat{q}_1(\alpha|x_1) - q_{1,n}(\alpha|x_1) &= \frac{1}{n}\sum_{j=1}^{n}\left[\hat{Q}_I(\alpha|x_1, X_{j\underline{1}}) - Q_n(\alpha|x_1, X_{j\underline{1}})\right] \\ &= \Delta_n^{(1)}(\alpha|x_1) + \frac{1}{2}\Delta_n^{(2)}(\alpha|x_1),\end{aligned} \tag{27}$$

where

$$\Delta_n^{(1)}(\alpha|x_1) = -\frac{1}{n^2 h_d}\sum_{j=1}^{n}\sum_{i=1}^{n} K_d\left(\frac{F(\frac{i}{n}|x_1, X_{j\underline{1}}) - \alpha}{h_d}\right)\left(\hat{F}\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right.$$
$$\left. - F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right),$$

$$\Delta_n^{(2)}(\alpha|x_1) = -\frac{1}{n^2 h_d^2}\sum_{j=1}^{n}\sum_{i=1}^{n} K_d'\left(\frac{\xi_i - \alpha}{h_d}\right)\left(\hat{F}\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right) - F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right)^2,$$

where $\xi_i = \xi_i(\alpha, x_1, X_{j\underline{1}})$ satisfies $|\xi_i - F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)| \leq |\hat{F}\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right) - F(\frac{i}{n}|x_1, X_{j\underline{1}})|$ for $i = 1, \ldots, n$. In the first step, we show that $\Delta_n^{(2)}(\alpha|x_1) = o_p\left(\frac{1}{\sqrt{nh_1}}\right)$. We obtain

$$|\Delta_n^{(2)}(\alpha|x_1)|$$
$$= \frac{1}{n^2 h_d^2}\left|\sum_{j=1}^{n}\sum_{i=1}^{n} K_d'\left(\frac{\xi_i - \alpha}{h_d}\right)\left(\hat{F}\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right) - F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right)^2\right|$$
$$= \frac{1}{n^2 h_d^2}\left|\sum_{j=1}^{n}\sum_{i=1}^{n} K_d'\left(\frac{F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right) - \alpha}{h_d}\right)\left[1 + \left(K_d'\left(\frac{F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right) - \alpha}{h_d}\right)\right)^{-1}\right.\right.$$
$$\times\left(K_d'\left(\frac{\xi_i - \alpha}{h_d}\right) - K_d'\left(\frac{F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right) - \alpha}{h_d}\right)\right)\Bigg]\left(\hat{F}\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right.$$
$$\left.\left. - F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right)^2\right|$$
$$= \frac{(1 + o_p(1))}{n^2 h_d^2}\left|\sum_{j=1}^{n}\sum_{i=1}^{n} K_d'\left(\frac{F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right) - \alpha}{h_d}\right)\left(\hat{F}\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right.\right.$$
$$\left.\left. - F\left(\frac{i}{n}|x_1, X_{j\underline{1}}\right)\right)^2\right|$$
$$= (1 + o_p(1))\Delta_n^{(2.1)}(\alpha|x_k), \tag{28}$$

where the last equation defines the quantity $\Delta_n^{(2.1)}$ in an obvious manner. In line (28), we used the Lipschitz continuity of K_d' and the uniform convergence rate of $\hat{F}(\alpha|x_1, x_{\underline{1}})$ (see Collomb and Härdle 1986), since

$$\left|K_d'\left(\frac{\xi_i-\alpha}{h_d}\right)-K_d'\left(\frac{F\left(\frac{i}{n}|x_1,X_{j\underline{1}}\right)-\alpha}{h_d}\right)\right|$$
$$\leq L\left|\frac{\xi_i-F\left(\frac{i}{n}|x_1,X_{j\underline{1}}\right)}{h_d}\right|$$
$$\leq L\left|\frac{\hat{F}\left(\frac{i}{n}|x_1,X_{j\underline{1}}\right)-F\left(\frac{i}{n}|x_1,X_{j\underline{1}}\right)}{h_d}\right|$$
$$=O_p\left(\frac{\log n}{nh_1g_2^{d-1}h_d^2}\right)^{1/2}=o_p(1). \tag{29}$$

Using the bandwidth condition (24), it follows

$$E[|\Delta_n^{(2.1)}(\alpha|x_k)|X_j]\leq\frac{1}{n^2h_d^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\left|K_d'\left(\frac{F\left(\frac{i}{n}|x_1,X_{j\underline{1}}\right)-\alpha}{h_d}\right)\right|$$
$$\times E\left[\left(\hat{F}\left(\frac{i}{n}|x_1,X_{j\underline{1}}\right)-F\left(\frac{i}{n}|x_1,X_{j\underline{1}}\right)\right)^2|X_j\right]$$
$$=O_p\left(\frac{1}{h_d}\left(\frac{1}{nh_1g_2^{d-1}}\right)\right)=o_p\left(\frac{1}{\sqrt{nh_1}}\right).$$

Now we can turn to the remaining term $\Delta_n^{(1)}(\alpha|x_1)$ which can be decomposed into bias- and variance-part. We obtain observing the representation (2)

$$\Delta_n^{(1)}(\alpha|x_1)=-\frac{(1+o_p(1))}{nh_d}\sum_{j=1}^{n}\int_0^1K_d\left(\frac{F(t|x_1,X_{j\underline{1}})-\alpha}{h_d}\right)$$
$$\times\left(\hat{F}(t|x_1,X_{j\underline{1}})-F(t|x_1,X_{j\underline{1}})\right) \tag{30}$$
$$=-\frac{(1+o_p(1))}{n^2h_d}\sum_{j=1}^{n}\sum_{k=1}^{n}\int_0^1K_d\left(\frac{F(t|x_1,X_{j\underline{1}})-\alpha}{h_d}\right)K_{h_1}(x_1-X_{k1})$$
$$\times L_G\left(X_{j\underline{1}}-X_{k\underline{1}}\right)\frac{I\{Y_k\leq t\}-F(t|x_1,X_{j\underline{1}})}{p(x_1,X_{j\underline{1}})}\mathrm{d}t$$
$$=(1+o_p(1))\left(\Delta_n^{(1.1)}(\alpha|x_1)+\Delta_n^{(1.2)}(\alpha|x_1)\right), \tag{31}$$

where

$$\Delta_n^{(1.1)}(\alpha|x_1) = -\frac{1}{n^2h_d}\sum_{j=1}^{n}\sum_{k=1}^{n}\int_0^1 K_d\left(\frac{F(t|x_1, X_{j\underline{1}}) - \alpha}{h_d}\right) K_{h_1}(x_1 - X_{k1})$$
$$\times L_G\left(X_{j\underline{1}} - X_{k\underline{1}}\right)\left(\frac{F(t|X_{k1}, X_{k\underline{1}}) - F(t|x_1, X_{j\underline{1}})}{p(x_1, X_{j\underline{1}})}\right)\mathrm{d}t \tag{32}$$

$$\Delta_n^{(1.2)}(\alpha|x_1) = -\frac{1}{n^2h_d}\sum_{j=1}^{n}\sum_{k=1}^{n}\int_0^1 K_d\left(\frac{F(t|x_1, X_{j\underline{1}}) - \alpha}{h_d}\right) K_{h_1}(x_1 - X_{k1})$$
$$\times L_G\left(X_{j\underline{1}} - X_{k\underline{1}}\right)\left(\frac{\sigma(t|X_{k1}, X_{k\underline{1}})\varepsilon_k}{p(x_1, X_{j\underline{1}})}\right)\mathrm{d}t. \tag{33}$$

The terms $\Delta_n^{(1.1)}$ and $\Delta_n^{(1.2)}$ are now investigated separately with an analysis similar as in Chen et al. (1996). First of all, $\Delta_n^{(1.1)}(\alpha|x_1)$ can be written as

$$\Delta_n^{(1.1)}(y|x_1) = -\frac{1}{n}\sum_{j=1}^{n}\eta_j(\alpha|x_1),$$

where

$$\eta_j(\alpha|x_1) = \frac{1}{n}\sum_{k=1}^{n} K_{h_1}(x_1 - X_{k1})\, L_G\left(X_{j\underline{1}} - X_{k\underline{1}}\right)$$
$$\times\frac{(F(Q(\alpha|x_1, X_{j\underline{1}})|X_{k1}, X_{k\underline{1}}) - F(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}}))}{F'(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})p(x_1, X_{j\underline{1}})}. \tag{34}$$

We break $\eta_j(\alpha|x_1)$ into two uncorrelated parts

$$\eta_j(\alpha|x_1) = E[\eta_j(\alpha|x_1)|X_j] + \left(\eta_j(\alpha|x_1) - E[\eta_j(\alpha|x_1)|X_j]\right).$$

For the conditional expectation of $\eta_j(\alpha|x_1)$, we have

$$E[\eta_j(\alpha|x_1)|X_j] = \int K_{h_1}(x_1 - u_1)\, L_G\left(X_{j\underline{1}} - u_{\underline{1}}\right)$$
$$\times\frac{(F(Q(\alpha|x_1, X_{j\underline{1}})|u_1, u_{\underline{1}}) - F(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}}))}{F'(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})p(x_1, X_{j\underline{1}})}p(u_1, u_{\underline{1}})\mathrm{d}u_1\mathrm{d}u_{\underline{1}}$$
$$= \int K(v_1)L(v_{\underline{1}})p(x_1 - h_1v_1, X_{j\underline{1}} - g_2v_{\underline{1}})$$
$$\times\frac{(F(Q(\alpha|x_1, X_{j\underline{1}})|x_1 - h_1v_1, X_{j\underline{1}} - g_2v_{\underline{1}}) - F(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}}))}{F'(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})p(x_1, X_{j\underline{1}})}\mathrm{d}v_1\mathrm{d}v_{\underline{1}}$$

$$= (1+o(1))h_1^2\kappa_2(K)\left(\frac{\frac{1}{2}\frac{\partial^2}{\partial x_1^2}F(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})}{F'(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})}\right.$$
$$\left. + \frac{\frac{\partial}{\partial x_1}F(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})}{F'(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})p(x_1, X_{j\underline{1}})}\right) + O_p(g_2^q).$$

Now it is easy to see that

$$E[\Delta_n^{(1.1)}(\alpha|x_1)] = -(1+o(1))b_1(\alpha|x_1).$$

To evaluate the variance of $\Delta_n^{(1.1)}(\alpha|x_1)$, we use the decomposition through $\eta_j(\alpha|x_1)$. Since

$$E[(\eta_j(\alpha|x_1) - E[\eta_j(\alpha|x_1)|X_j])^2|X_j] \le O_p\left(\frac{1}{nh_1g_2^{d-1}}(h_1^2+g_2^2)\right)$$

we can estimate the variance of $\Delta_n^{(1.1)}(\alpha|x_1)$ as

$$\begin{aligned}\mathrm{Var}(\Delta_n^{(1.1)}(\alpha|x_1)) &\le E\left[(\Delta_n^{(1.1)}(\alpha|x_1))^2\right] = \frac{1}{n}E\left[(E[\eta_j(\alpha|x_1)|X_j])^2\right]\\ &\quad + O\left(\frac{(h_1+g_2)^2}{n^2h_1g_2^{d-1}}\right)\\ &= O\left(\frac{(h_1^2+g_2^q)^2}{n} + \frac{(h_1+g_2)^2}{n^2h_1g_2^{d-1}}\right) = o\left(\frac{1}{nh_1}\right),\end{aligned}$$

which shows

$$\Delta_n^{(1.1)}(\alpha|x_1) + b(\alpha|x_1) = o_p\left(\frac{1}{\sqrt{nh_1}}\right). \tag{35}$$

Now we consider the term $\Delta_n^{(1.2)}$ in (33), which has expectation $E[\Delta^{(1.2)}(\alpha|x_1)] = 0$. To calculate the variance, we use a similar analysis as for $\Delta_n^{(1.1)}$

$$\Delta_n^{(1.2)}(\alpha|x_1) = -\frac{1}{nh_1}\sum_{k=1}^n K\left(\frac{x_1 - X_{k1}}{h_1}\right)\varepsilon_k\beta_k(\alpha|x_1),$$

where

$$\beta_k(\alpha|x_1) = \frac{1}{n}\sum_{j=1}^n \frac{L_G\left(X_{j\underline{1}} - X_{k\underline{1}}\right)\sigma(Q(\alpha|x_1, X_{j\underline{1}})|X_{k1}, X_{k\underline{1}})}{p(x_1, X_{j\underline{1}})F'(Q(\alpha|x_1, X_{j\underline{1}})|x_1, X_{j\underline{1}})}.$$

Now we treat $\beta_k(\alpha|x_1)$ as $\eta_k(\alpha|x_1)$ and split it up into

$$E[\beta_k(\alpha|x_1)|X_k] + (\beta_k(\alpha|x_1) - E[\beta_k(\alpha|x_1)|X_k]).$$

We calculate the conditional expectation of $\beta_k(\alpha|x_1)$ as

$$\begin{aligned} E[\beta_k(\alpha|x_1)|X_k] &= \int \frac{L_G\left(u_{\underline{1}} - X_{k\underline{1}}\right)\sigma(Q(\alpha|x_1, u_{\underline{1}})|X_{k1}, X_{k\underline{1}})p(u_1, u_{\underline{1}})}{p(x_1, u_{\underline{1}})F'(Q(\alpha|x_1, u_{\underline{1}})|x_1, u_{\underline{1}})} \mathrm{d}u_1 \mathrm{d}u_{\underline{1}} \\ &= \frac{\sigma(Q(\alpha|x_1, X_{k\underline{1}})|X_{k1}, X_{k\underline{1}})p_{\underline{1}}(X_{k\underline{1}})}{p(x_1, X_{k\underline{1}})F'(Q(\alpha|x_1, X_{k\underline{1}})|x_1, X_{k\underline{1}})} + O_p(g_2^q) \end{aligned}$$

Furthermore we have

$$\begin{aligned} &E\left[(\beta_k(\alpha|x_1) - E[\beta_k(\alpha|x_1)|X_k])^2|X_k\right] \\ &\le \frac{1}{n}\int\left(\frac{L_G\left(u_{\underline{1}} - X_{k\underline{1}}\right)\sigma(Q(\alpha|x_1, u_{\underline{1}})|X_{k1}, X_{k\underline{1}})}{p(x_1, u_{\underline{1}})F'(Q(\alpha|x_1, u_{\underline{1}})|x_1, u_{\underline{1}})}\right)^2 p(u_1, u_{\underline{1}})\mathrm{d}u_1\mathrm{d}u_{\underline{1}} \\ &= \frac{1}{ng_2^{d-1}}\left(\int L^2(v_{\underline{1}})\mathrm{d}v_{\underline{1}}\right)\frac{\sigma(Q(\alpha|x_1, X_{k\underline{1}})|X_{k1}, X_{k\underline{1}})p_{\underline{1}}(X_{k\underline{1}})}{p^2(x_1, X_{k\underline{1}})\left(F'(Q(\alpha|x_1, X_{k\underline{1}})|x_1, X_{k\underline{1}})\right)^2} \\ &= O_p\left(\frac{1}{ng_2^{d-1}}\right). \end{aligned}$$

So basically $\Delta_n^{(1.2)}(\alpha|x_1)$ is of the form

$$\begin{aligned} \Delta_n^{(1.2)}(\alpha|x_1) = &-\frac{1}{n}\sum_{k=1}^{n} K_{h_1}(x_1 - X_{k1})\frac{p_{\underline{1}}(X_{k\underline{1}})\sigma(Q(\alpha|x_1, X_{k\underline{1}})|X_{k1}, X_{k\underline{1}})\varepsilon_k}{p(x_1, X_{k\underline{1}})F'(Q(\alpha|x_1, X_{k\underline{1}})|x_1, X_{k\underline{1}})} \\ &+ o_p\left(\frac{1}{\sqrt{nh_1}}\right), \end{aligned}$$

and the variance of the dominating term on the right hand side, say $\hat{\Delta}_n^{(1.2)}(\alpha|x_1)$, can be easily calculated. i.e.

$$\begin{aligned} &\mathrm{Var}(\sqrt{nh_1}\hat{\Delta}_n^{(1.2)}(\alpha|x_1)) \\ &= \left(\int K^2(v)\mathrm{d}v\right)\int \frac{\alpha(1-\alpha)p_{\underline{1}}^2(x_{\underline{1}})\mathrm{d}x_{\underline{1}}}{p(x_1, x_{\underline{1}})(F'(Q(\alpha|x_1, x_{\underline{1}})|x_1, x_{\underline{1}})} + o(1). \end{aligned}$$

A similar calculation shows that Ljapunoff's condition is satisfied for the term $\Delta_n^{(1.2)}$, that is

$$\sum_{j=1}^{n} E\left[\frac{\sqrt{nh_1}}{nh_1}\sum_{k=1}^{n} K\left(\frac{x_1 - X_{k1}}{h_1}\right)\right.$$
$$\left.\times\frac{p_{\underline{1}}(X_{k\underline{1}})\sigma(Q(\alpha|x_1, X_{k\underline{1}})|X_{k1}, X_{k\underline{1}})\varepsilon_k}{p(x_1, X_{k\underline{1}})F'(Q(\alpha|x_1, X_{k\underline{1}})|x_1, X_{k\underline{1}}))}\right]^4$$
$$= \frac{E[\varepsilon_1^4]}{nh_1^2}\int K^4\left(\frac{x_1 - u_1}{h_1}\right)\left(\frac{\sigma(Q(\alpha|x_1, u_{\underline{1}})|u_1, u_{\underline{1}})p_{\underline{1}}(u_{\underline{1}})}{F'(Q(\alpha|x_1, u_{\underline{1}})|x_1, u_{\underline{1}})p(x_1, u_{\underline{1}})}\right)^4$$
$$\times p(u_1, u_{\underline{1}})\mathrm{d}u_1\mathrm{d}u_{\underline{1}}$$
$$= \frac{E[\varepsilon_1^4]}{nh_1}\left(\int K^4(v)\mathrm{d}v\right)\int \frac{\alpha^2(1-\alpha)^2 p_{\underline{1}}^4(u_{\underline{1}})}{(F'(Q(\alpha|x_1, u_{\underline{1}})|x_1, u_{\underline{1}}))^4 p^3(x_1, u_{\underline{1}})}\mathrm{d}u_{\underline{1}}$$
$$= O\left(\frac{1}{nh_1}\right), \tag{36}$$

which establishes the weak convergence

$$\sqrt{nh_1}\Delta_n^{(1.2)}(\alpha|x_1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2(\alpha|x_1)).$$

A combination with (35) and (30) yields (26) which proves Theorem 1. □

Acknowledgments The support of the Deutsche Forschungsgemeinschaft (SFB 475, "Komplexitätsreduktion in multivariaten Datenstrukturen"). is gratefully acknowledged. The work of H. Dette was supported in part by a NIH grant award IR01GM072876:01A1. The authors are grateful to two unknown referees for their constructive comments on an earlier version of this paper and to Martina Stein, who typed parts of the paper with considerable technical expertise.

References

Belsley, D. A., Kuh, E., Welsch, R. E. (1980). *Regression diagnostics. Identifying influential data and sources of collinearity*. New York: Wiley.

Bennett, C., Sharpley, R. (1988). *Interpolation of operators*. NY: Academic Press.

Chen, R., Härdle, W., Linton, O. B., Severence-Lossin, E. (1996). Nonparametric estimation of additive separable regression models. *Statistical theory and computational aspects of smoothing (Semmering, 1994)* (pp. 247–265).

Chernozhukov, V., Fernandez-Val, I., Galichon, A. (2007). Quantile and probability curves without crossing. arXiv:0704.3649. http://trefoil.math.ucdavis.edu/0704.3649.

Collomb, G., Härdle, W. (1986). Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations. *Stochastic Processes and their Application*, *23*(1), 77–89.

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.

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

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 & Data Analysis*, *35*, 67–82.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.

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

Harrison, D., Rubinfeld, D. L. (1978). Hedonic prices and the demand for clean air. *Journal of Environmental Economics and Management*, *5*, 81–102.

He, X. (1997). Quantile curves without crossing. *The American Statistician*, *51*(2), 186–192.

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*. London: Cambridge University Press.

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

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

Masry, E., Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics*, *24*(2), 165–179.

Neumeyer, N., Sperlich, S. (2006). Comparison of separable components in different sample. *Scandinavian Journal of Statistics*, *33*, 477–501.

Yu, K., Jones, M. C. (1997). A comparison of local constant and local linear regression quantile estimators. *Computational Statistics & 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., Stander, J. (2003). Quantile regression: applications and current research areas. *The Statistician*, *52*(3), 331–350.