Efficient Estimation of an Additive Quantile Regression Model

Cheng, Yebin; De Gooijer, Jan and Zerom, Dawit
California State University Fullerton

14. March 2009

Efficient Estimation of an Additive Quantile Regression Model

Yebin Cheng[a], Jan G. De Gooijer[b] and Dawit Zerom[c*]

[a] *Shanghai University of Finance and Economics, China*

[b] *University of Amsterdam, The Netherlands*

[c] *California State University - Fullerton, USA*

Abstract

In this paper two kernel-based nonparametric estimators are proposed for estimating the components of an additive quantile regression model. The first estimator is a computationally convenient approach which can be viewed as a viable alternative to the method of De Gooijer and Zerom (2003). With the aim to reduce variance of the first estimator, a second estimator is defined via sequential fitting of univariate local polynomial quantile smoothing for each additive component with the other additive components replaced by the corresponding estimates from the first estimator. The second estimator achieves oracle efficiency in the sense that each estimated additive component has the same variance as in the case when all other additive components were known. Asymptotic properties are derived for both estimators under dependent processes that are strictly stationary and absolutely regular. We also provide a demonstrative empirical application of additive quantile models to ambulance travel times.

Keywords: Additive models; Asymptotic properties; Dependent data; Internalized kernel smoothing; Local polynomial; Oracle efficiency.

*Corresponding author: Mihaylo College of Business and Economics, California State University, Fullerton, CA, 92834-6848, (714) 278-3635, dzerom@fullerton.edu.

1 Introduction

Suppose Y denotes a response variable that depends on the vector of stochastic covariates $X = (X_1, \ldots, X_d)^T$, $d \geq 2$, where T denotes the transpose of a matrix or a vector. We consider the case where the relationship between Y and X follows a quantile regression set-up,

$$Y_i = Q_\alpha(X_i) + \mathcal{E}_{\alpha,i}, \quad i = 1, \ldots, n \tag{1.1}$$

where $Q_\alpha(\cdot)$ is an unknown real-valued function and $\mathcal{E}_\alpha$ is an unobserved random variable that satisfies $I\!P(\mathcal{E}_{\alpha,i} \leq 0 | X = x) = \alpha$ for all x where $0 < \alpha < 1$ is the quantile of interest. In this way, $Q_\alpha(x)$ denotes the conditional quantile of Y_i given $X_i = x$. Indeed, there is a large body of literature on the estimation of $Q_\alpha(x)$ and its asymptotic properties (see, e.g., Chaudhuri 1991; Fan, Hu, and Troung 1994). But it is well-known that for high-dimensional covariates (moderate to large value of d) 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 proposed dimension reduction techniques. For instance, Honda (2004), Kim (2007) and Cai and Xu (2008) consider quantile regression with varying coefficients. Alternatively, Lee (2003) studies conditional quantiles using a partially linear regression model. In this paper, we assume $Q_\alpha(\cdot)$ to be additive of the following form,

$$Q_\alpha(x) = c_\alpha + q_{\alpha,1}(x_1) + \ldots + q_{\alpha,d}(x_d), \tag{1.2}$$

where $x = (x_1, \ldots, x_d)^T$, c_α is a constant, and $q_{\alpha,u}(x_u)$ $(u = 1, \ldots, d)$ are smooth nonparametric functions representing the α-th quantile function of Y related only to X_u. Additive models are simple, easily interpretable, and sufficiently flexible for many practical applications.

Given observe data $(X_1, Y_1), \ldots, (X_n, Y_n)$, our interest is to efficiently estimate each additive components $q_{\alpha,u}(x_u)$ in (1.2). This nonparametric estimation problem is first considered by Fan and Gijbels (1996, pp. 296-297) where they suggest a back-fitting procedure for estimating the additive components. Yu and Lu (2004) later re-consider the back-fitting procedure. Although the back-fitting algorithm is easy to implement, there is no guarantee for convergence and its iterative structure makes it difficult to establish asymptotic results. Doksum and Koo (2000) introduce an easily implementable direct spline method that does not require iterations. But they do not provide asymptotic convergence results. De Gooijer and Zerom (2003) propose a simple direct kernel estimator. Horowitz and Lee (2005) suggest a hybrid step-wise approach where they use a series method in the first step and kernel smoothing in the second step. Both

De Gooijer and Zerom (2003), and Horowitz and Lee (2005) provide detailed asymptotic theory, and show that their respective estimators achieve a univariate nonparametric rate of convergence regardless of the dimension of X.

In this paper, we propose two kernel-based estimators for estimating the additive component functions. Our first estimator extends the works of Kim, Linton and Hengartner (1999) and Manzan and Zerom (2005) to the context of conditional quantiles. We show that the proposed estimator is asymptotically normal and converges at the univariate nonparametric optimal rate. This estimator is computationally more attractive than the average quantile estimator of De Gooijer and Zerom (2003) as it reduces the computational requirement of the latter by the order of the sample size $O(n)$. In applications, this computational advantage can be very significant when n is large and/or when implementing computer-intensive methods such as bootstrap or cross-validation. For example, in the empirical analysis of ambulance travel times (see Section 5), we have over 7000 observations. For n of this size, implementing the average quantile estimator requires excessively large computational time. In addition to its computational inconvenience, the average quantile estimator is also not robust to correlated covariates in the sense that its efficiency deteriorates with an increase in the correlation among the covariates $(X_1, \ldots, X_d)$. This is the result of the need to smooth at points that may not lie in the support of the covariate space. On the other hand, our estimator is not affected by this problem.

Although our first estimator is practically appealing, its asymptotic variance has an undesirable additional term. To mitigate this efficiency problem, we propose a second estimator that uses further local averaging. The local averaging involves sequential fitting of univariate local polynomial quantile smoothing for each additive components with the other additive components replaced by the corresponding estimates from the first estimator. The second proposed estimator is also shown to be asymptotically normal and converges at the univariate nonparametric optimal rate. Further, we show that it achieves oracle efficiency where each estimated additive component has the same variance as in the case when all other additive components were known. In terms of computer implementation, this efficient estimator only takes twice as many computational operations as our estimator. Thus, efficiency is achieved without compromising on computational simplicity. The estimator of Horowitz and Lee (2005) also shares the oracle property.

The asymptotic properties of our two kernel estimators are derived for dependent data. On the other hand, Horowitz and Lee (2005) establish the asymptotic properties of their estimator only for the case of independent data. Thus, our theoretical results are more general. We assume that the sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is a strictly stationary weakly dependent data from the population $\{X, Y\}$. We focus on absolutely regular (or β-mixing) processes. For any $a < b$, let $\mathcal{M}_a^b$ denote the sigma algebra generated by $(Z_a, \ldots, Z_b)$ with $Z_i = (X_i, Y_i)$. A process is called absolutely regular, if, as $m \to \infty$,

$$\pi(m) = \sup_{s \in \mathbb{N}} I\!E \left\{ \sup_{\mathcal{H} \in \mathscr{M}_{s+m}^{\infty}} [I\!P(\mathcal{H}|\mathcal{M}_{-\infty}^{s}) - I\!P(\mathcal{H})] \right\} \to 0.$$

For more details on β-mixing processes, see, for example, Yoshihara (1978) and Arcones (1998).

The paper is organized as follows. In Section 2 we provide a description of a modified average quantile estimator together with its asymptotic properties. In Section 3, an oracle efficient estimator is introduced and its asymptotic properties are also established. In Section 4, we illustrate the numerical performance of the proposed estimators using simulated data. In Section 5, we provide a demonstrative empirical application of additive quantile modeling to ambulance travel times using administrative data for the city of Calgary. Section 6 provides concluding comments. Technical arguments and proofs are provided in two Appendices.

2 A modified average quantile estimator

Here, we introduce our first kernel estimator for the additive component function $q_{\alpha,u}(x_u)$ for $u = 1, \ldots, d$. In order to make the u-th component $q_{\alpha,u}(x_u)$ identifiable, it is assumed that $I\!E\{q_{\alpha,u}(X_u)\} = 0$ for $u = 1, \ldots, d$. For ease of exposition, we denote by X_u the u-th element of X and W_u the set of all X variables excluding X_u, i.e. $W_u = (X_1, \ldots, X_{u-1}, X_{u+1}, \ldots, X_d)^T$. Note that $X = (X_u, W_u)$. Also let $f_u(\cdot)$, $f_w(\cdot)$ and $f(\cdot)$ denote the density functions of X_u, W_u and X, respectively. Following Kim, Linton and Hengartner (1999), we define the function

$$\phi(x_u, w_u) = \frac{f_u(x_u) f_w(w_u)}{f(x_u, w_u)}.$$

It is easy to show that this function has two desirable properties:

$$I\!E\{\phi(X_u, W_u)\ |X_u = x_u\} = 1 \quad \text{and} \quad I\!E\{\phi(X_u, W_u) q_{\alpha,k}(X_k)|X_u = x_u\} = 0 \quad \text{for} \quad k \neq u.$$

Multiplying each side of equation (1.2) by $\phi(\cdot, \cdot)$ and taking conditional expectations conditional on $X_u{=}x_u$, we obtain

$$I\!E\{\phi(X_u, W_u)\ Q_\alpha(X)|X_u = x_u\} \equiv q^*_{\alpha,u}(x_u) = c_\alpha + q_{\alpha,u}(x_u), \quad (u = 1, \ldots, d).$$

Therefore, $q^*_{\alpha,u}(x_u)$ coincides, up to a constant, with the component $q_{\alpha,u}(x_u)$ of the additive quantile model. Thus, we can estimate $q_{\alpha,u}(x_u)$ by the following estimator which we call the modified average quantile estimator,

$$\hat{q}_{\alpha,u}(x_u) = \hat{q}^*_{\alpha,u}(x_u) - \hat{c}_\alpha \tag{2.1}$$

with the two estimators $\hat{q}^*_{\alpha,u}(x_u)$ and $\hat{c}_\alpha$ given in (2.3) and (2.2), respectively. Because $c_\alpha = I\!\!EQ_\alpha(X)$, we can estimate c_α by

$$\hat{c}_\alpha = \frac{1}{n}\sum_{i=1}^{n} \hat{Q}_\alpha(X_i), \tag{2.2}$$

where $\hat{Q}_\alpha(\cdot)$ is a consistent estimator of $Q_\alpha(\cdot)$ which is defined in (2.4). To compute $\hat{q}^*_{\alpha,u}(x_u)$, we use an internalized kernel smoothing as follows,

$$\hat{q}^*_{\alpha,u}(x_u) = \frac{1}{nh_1}\sum_{i=1}^{n} K\left(\frac{x_u - X_{i,u}}{h_1}\right)\frac{\hat{f}_w(W_{i,u})}{\hat{f}(X_i)}\hat{Q}_\alpha(X_i), \tag{2.3}$$

where $K(\cdot)$ is a kernel function, h_1 is a bandwidth (or smoothing parameter) and $\hat{f}_w(\cdot)$ and $\hat{f}(\cdot)$ are kernel smoothers of the corresponding densities. Note that, unlike the usual kernel-based conditional expectation smoothers, (2.3) eliminates explicit estimation of the density $f_u(x_u)$ in the denominator and hence named an internalized smoother; see Jones, Davies and Park (1994) for details on internalized smoothing. When compared to that of De Gooijer and Zerom (2003), this internalization offers a significant practical advantage by reducing computational cost by the order n (i.e., $O(n)$). To better see this advantage, we can re-define (2.3) in a more computationally convenient way as follows. Say, the aim is to estimate $\hat{q}^*_{\alpha,u}(\cdot)$ at all observation points $X_{u,i}$ for $i = 1, \ldots, n$. First, define the following $n \times n$ smoother matrices,

$$\begin{aligned} S_u^x &= \left[\frac{1}{nh_1}K\left(\frac{X_{i,u} - X_{\ell,u}}{h_1}\right)\right]_{i,\ell}, \qquad S_u^w = \left[\frac{1}{nh_2^{d-1}}L_1\left(\frac{W_{i,u} - W_{\ell,u}}{h_2}\right)\right]_{i,\ell}, \\ S &= \left[\frac{1}{nh_2^{d}}L_2\left(\frac{X_i - X_\ell}{h_2}\right)\right]_{i,\ell}, \end{aligned}$$

where $L_1(\cdot)$ and $L_2(\cdot)$ are two kernel functions, and h_2 is the bandwidth. Then, we can estimate the $n \times 1$ vector of estimates $(\hat{q}^*_{\alpha,u}(X_{u,1}), \ldots, \hat{q}^*_{\alpha,u}(X_{u,n}))^T$, all at once, as follows

$$(\hat{q}^*_{\alpha,u}(X_{u,1}), \ldots, \hat{q}^*_{\alpha,u}(X_{u,n}))^T = S_u^x\{\hat{Q}_\alpha \odot (S_u^w\ e)./(S\ e)\},$$

where $\odot$ and ./ denote matrix Hadamard product and division, respectively, while $e = (1, \ldots, 1)^T$ and $\hat{Q}_\alpha$=$(\hat{Q}_\alpha(X_1), \ldots, \hat{Q}_\alpha(X_n))^T$. Further, unlike that of De Gooijer and Zerom (2003), the computation of $\hat{q}^*_{\alpha,u}(x_u)$ does not require smoothing at pairs (x_u, W_u). This feature is important as (x_u, W_u) may not lie in the support of (X_u, W_u). Unless the product of the marginal supports is equal to the joint support, we may be estimating at points where the joint density is zero. Many data sets have highly correlated design, which causes the finite support to violate the above requirement. The estimator in (2.3) does not face this problem and hence is robust against correlated design.

Now we define an estimator for $Q_\alpha(x)$. We assume that $Q_\alpha(x)$ is p-times ($p \geq 2$) continuously differentiable in the neighborhood of $x \in \mathbb{R}^d$. This will allow us to carry the well-known local polynomial quantile smoothing; see Honda (2000). For non-negative integer vector $\lambda = (\lambda_1, \ldots, \lambda_d)$, let $|\lambda| = \sum_i \lambda_i$ and $x^\lambda = \Pi x_i^{\lambda_i}$. Also let the vectors $V_1(\frac{X-x}{h})$ and β_x be constructed from the elements $h^{-|\lambda|}(X-x)^\lambda$ and $\frac{h^{-|\lambda|}\partial^\lambda q(x)}{x_1^{\lambda_1}\cdots x_d^{\lambda_d}}$, respectively, which are arranged in natural order with respect to λ such that $|\lambda| \leq p-1$. As usual, we define $\hat{Q}_\alpha(x)$ by

$$\hat{Q}_\alpha(x) = e_1^T \hat{\beta}_x, \tag{2.4}$$

where e_1 is an p-dimensional unit vector with the first element 1 and all other elements 0 and the vector $\hat{\beta}_x$ minimizes

$$(nh^d)^{-1} \sum_{i=1}^n \rho_\alpha\left(Y_i - \beta_x^T V_1\left(\frac{X_i - x}{h}\right)\right) L\left(\frac{x - X_i}{h}\right),$$

where $\rho_\alpha(\cdot)$ is a check function that is defined as $\rho_\alpha(s) = |s| + (2\alpha - 1)s$ for $0 < \alpha < 1$ and $L(\cdot)$ is a kernel function and h is the bandwidth. The above polynomial smoothing is easy to implement in the major statistical softwares using a weighted linear quantile regression routine where the weights are defined through the kernel $L(\cdot)$.

2.1 Asymptotic behavior

Here, We derive the asymptotic behavior of the modified average quantile estimator $\hat{q}_{\alpha,u}(x_u)$ (2.1) under β-mixing. In this paper, $C < \infty$ denotes a positive generic constant. We use the following regularity conditions to derive the asymptotic properties.

C1. The additive function $q_{\alpha,u}(x_u)$ is p-times continuously differentiable in the neighborhood of $x_u \in \mathbb{R}$. The full-dimensional conditional quantile $Q_\alpha(x)$ is p-times continuously differentiable in the neighborhood of $x \in \mathbb{R}^d$. The probability density function $f(x)$ of X is bounded from above and has $\bar{p}$th derivatives on their support set, where $\bar{p} > \frac{pd}{p+1}$.

C2. Let $g(y|x)$ be the conditional probability density function of $\mathcal{E}_\alpha$ given $X = x$. For any x in the support set of X, it has the first continuous derivative with respect to the argument y in the neighborhood of 0.

C3. $K(\cdot)$ is a p-th order kernel function that satisfies $\int K(t_1)dt_1 = 1$, $\int t_1^j K(t_1)dt_1 = 0$ for $j = 1, \ldots, p-1$ and $\int t_1^p K(t_1)dt_1 \neq 0$. For $i = 1, 2$, $L_i(\cdot)$ is a $\bar{p}$-th order kernel function that satisfies $\int L_i(s)ds = 1$, $\int s^j L_i(s)ds = 0$ for $j = 1, \ldots, \bar{p}-1$ and $\int s^{\bar{p}} L_i(s)ds \neq 0$ with s in $d-1$ or d dimensional spaces according to $L_i(\cdot)$. $L(t)$ is a second-order kernel which has bounded and continuous partial derivatives of order 1.

C4. i). There exist two constants $\delta > 2$ and $\gamma > 0$ such that $\delta > 2 + \gamma$ and the function

$$\mathbb{E}\left\{ \left| \frac{f_w(W_u)}{f(X)} Q_\alpha(X) \right|^\delta \middle| X_u = x_u' \right\}$$

is bounded in the neighbor of $x_u' = x_u$.

ii). The mixing coefficients $\pi(i) = O\left(i^{-\theta}\right)$ with $\theta \geq \max\left\{p + \frac{4}{p} + 6, \frac{2(p+1)\delta}{\delta-2} + 1\right\}$.

C5. i). It holds that $n^{-\gamma/4} h_1^{(2+\gamma)/\delta - 1 - \gamma/4} = O(1)$ and $\limsup_n n h_1^{2p+1} < \infty$.

ii). Assume that there exists a sequence of positive integers s_n such that $s_n \to \infty$, $s_n = o\left((nh_1)^{1/2}\right)$, and $(n/h_1)^{1/2}\pi(s_n) \to 0$, as $n \to \infty$.

iii). $h = Cn^{-\kappa}$ with constant κ satisfying $\frac{1}{2p+1} < \kappa < \frac{2p+3}{3d(2p+1)}$ and $h/h_1 \to 0$.

iv). For some sufficiently small constant $\epsilon > 0$, it holds that $h_1^{\theta(1-\frac{2}{\delta})} h^{\frac{2}{\delta}-2} \to 0$, $nh^d \left(h_1 h^d\right)^{\frac{3}{\theta}+\epsilon} \to \infty$ and $nh_1^{-1}\left(h_1 h_2^{d-1}\right)^{1+\frac{3}{\theta}+\epsilon} \to \infty$ with $h_2 = Cn^{-\frac{1}{d+\bar{p}}}$.

C6. For any $j \geq 1$, the joint density functions (X_1, X_{j+1}) are bounded from above.

Let $\kappa_p = \int t_1^p K(t_1)dt_1$ and $\|K\|_2 = \int K^2(t_1)dt_1$. The following theorem summarizes the asymptotic distribution of $\hat{q}^*_{\alpha,u}(x_u)$.

Theorem 2.1. *When the conditions C1 to C6 are met,*

$$\sqrt{nh_1}\left(\hat{q}^*_{\alpha,u}(x_u) - q^*_{\alpha,u}(x_u) - \frac{q^{(p)}_{\alpha,u}(x_u)\kappa_p}{p!} h_1^p\right) \to N\left(0, \sigma^2\right) \tag{2.5}$$

in distribution with $\sigma^2 = \sigma_1^2 + \sigma_2^2$ *where*

$$\sigma_1^2 = \frac{\alpha(1-\alpha)\|K\|_2}{f_u(x_u)} \mathbb{E}\left(\frac{\phi^2(X)}{g^2(0|X)} \middle| X_u = x_u \right) \quad \& \quad \sigma_2^2 = \frac{\|K\|_2}{f_u(x_u)} \mathbb{E}\left[\phi^2(X) Q_\alpha^2(X) \middle| X_u = x_u\right].$$

Remark 1. To simplify our presentation, we assume that smoothness of $Q_\alpha(x)$ and its uth additive component is of the same order p. But, it is possible that the smoothness of these functions can be different. For example, when $Q_\alpha(x_1, x_2) = c_\alpha + x_1^2 + \sin(x_2)$, $Q_\alpha(x_1, x_2)$ has derivatives of any order but x_1^2 only has the second-order differentiability, i.e., $p = \infty$ and $p_1 = 2$. Following the same lines of the proofs and using $\limsup_n nh_1^{2p_1+1} < \infty$, Theorem 2.1 will still hold where p is replaced by p_1 in the asymptotic distribution expression.

Remark 2. From Theorem 2.1, the optimal bandwidth that minimizes the asymptotic mean squared error (AMSE) is given by,

$$h_1^{opt} = \left(\frac{p!\sigma}{q_{\alpha,u}^{(p)}(x_u)\kappa_p} \right)^{\frac{2}{2p+1}} n^{-\frac{1}{2p+1}}.$$

Remark 3. Although the asymptotic variance σ^2 can not be directly compared to the corresponding variance of the estimator of De Gooijer and Zerom (2003), there is a visible additional term (σ_2^2) in the case of our estimator. A similar problem has also been shown by Kim, *et al* (1999) for the conditional mean case. This motivates us to introduce our second estimator (see Section 3) whose goal is to mitigate this efficiency problem without compromising on bias.

Proposition 2.2. *Under the conditions of Theorem 2.1,*

$$\hat{c}_\alpha - c_\alpha = o_{I\!P}\left(n^{-\frac{p}{2p+1}}\right). \tag{2.6}$$

Corollary 2.3. *Under the conditions of Theorem 2.1, if we choose* $h_1 = h_1^{opt}$*, then it holds that*

$$\left(\frac{p!\sigma}{\left|q_{\alpha,u}^{(p)}(x_u)\right| \kappa_p} \right)^{\frac{1}{2p+1}} n^{\frac{p}{2p+1}} \left(\hat{q}_{\alpha,u}(x_u) - q_{\alpha,u}(x_u) - \frac{q_{\alpha,u}^{(p)}(x_u)\kappa_p}{p!} n^{-\frac{p}{2p+1}} \right) \to N\left(0, \sigma^2\right)$$

in distribution.

3 Oracle efficient estimator

In Section 2 we introduce a modified average quantile estimator and show that it estimates the additive components at a one-dimensional nonparametric optimal rate regardless of the size of d. However, a closer look at Theorem 2.1 indicate that the asymptotic variance includes a second term (σ_2^2) which inflates the value of the variance. To deal with this inefficiency, we extend the idea of Linton (1996) and Kim *et al* (1999) to the quantile context and suggest a second estimator that involves sequential fitting of univariate local polynomial quantile smoothing for each additive components with the other additive components replaced by the corresponding estimates from the average quantile estimator. In fact, we will show in Section 3.1) that the proposed estimator is oracle efficient in the sense that it is asymptotically distributed with same mean and variance as it would have if the other additive components were known. Importantly, this efficient estimator only takes twice as many computational operations as the modified average quantile estimator. Thus, efficiency is achieved without compromising on computational simplicity.

We construct this estimator as follows. First, define

$$\hat{Q}^*_{\alpha,-u}(W_u) = \hat{q}^*_{\alpha,1}(X_1) + \cdots + \hat{q}^*_{\alpha,u-1}(X_{u-1}) + \hat{q}^*_{\alpha,u+1}(X_{u+1}) + \cdots + \hat{q}^*_{\alpha,d}(X_d), \tag{3.1}$$

where $\hat{q}^*_{\alpha,j}(\cdot)$ $(j \neq u)$ are the additive estimates from (2.3). For technical convenience, we consider the one-leave-out versions of these first-stage estimates. Let

$$Y_i^* = Y_i + (d-2)\hat{c}_\alpha - \hat{Q}^*_{\alpha,-u}(W_{i,u})$$

where $\hat{c}_\alpha$ is given by (2.2). Let the function $V(t)$ denotes a p-dimensional vector where its jth element given by t^{j-1}. Then, using the local polynomial smoothing, we define the oracle efficient estimator by

$$\hat{q}^e_{\alpha,u}(x_u) = e_1^T \hat{\beta}_{x_u}, \tag{3.2}$$

where e_1 is a p-dimensional unit vector with the first element 1 and all other elements 0 and the vector $\hat{\beta}_{x_u}$ minimizes

$$(nh_e)^{-1} \sum_{i=1}^{n} \rho_\alpha \left(Y_i^* - \beta_{x_u}^T V\left(\frac{x_u - X_{i,u}}{h_e} \right) \right) K_e \left(\frac{x_u - X_{i,u}}{h_e} \right), \tag{3.3}$$

where $K_e(\cdot)$ is a kernel function and h_e is the bandwidth. The computer implementation of this estimator is similar to that used to compute $\hat{Q}_\alpha(x)$ in Section 2.

3.1 Asymptotic behavior

We investigate asymptotic distribution of $\hat{q}^e_{\alpha,u}(x_u)$ (3.2). To derive our results, we use the following extra regularity conditions.

C7. $K_e(t_1)$ is a second-order kernel which has bounded and continuous first order derivative.

C8. Let $g_u(t|x_u)$ be the conditional probability density function of $\mathcal{E}_\alpha$ given $X_u = x_u$ and $g_u\,(t|x_u)$ has bounded derivative in the neighborhood of $t = 0$.

C9. It holds that $h_e = Cn^{-\frac{1}{2p+1}}$ and the bandwidth of the modified average quantile estimator satisfies that $h_1 = h_e n^{-\frac{\varepsilon_0}{2}}$ with some constant $\varepsilon_0 > 0$ which sufficiently small.

The oracle estimator

Before we provide the asymptotic distribution of $\hat{q}^e_{\alpha,u}(x_u)$, we first present results for an oracle estimator which we denote by $\hat{q}^{oracle}_{\alpha,u}(x_u)$. We define $\hat{q}^{oracle}_{\alpha,u}(x_u)$ in the same way as $\hat{q}^e_{\alpha,u}(x_u)$ except that the oracle estimator is based on true values of the other additive components. Thus, $\hat{q}^{oracle}_{\alpha,u}(x_u)$ is some desirable estimator while being infeasible in practice. Let

$$Q_{\alpha,-u}\,(w_u) = \sum_{1\le j\ne u\le d} q_{\alpha,j}(x_j).$$

Suppose we know $\{c_\alpha, q_{\alpha,j}(x_j),\ 1 \le j \ne u \le d\}$. But we do not know $q_{\alpha,u}(x_u)$. Note that $\alpha = I\!P\left\{Y_i - c_\alpha - Q_{\alpha,-u}\,(W_{i,u}) \le q_{\alpha,u}(X_{i,u})|X_{i,u}\right\}$ and $q_{\alpha,u}(x_u)$ has pth derivative. Then, using the local polynomial smoothing, we define $\hat{q}^{oracle}_{\alpha,u}(x_u)$ by,

$$\hat{q}^{oracle}_{\alpha,u}(x_u) = e_1^T \hat{\beta}^{oracle}_{x_u}, \tag{3.4}$$

where the vector $\hat{\beta}_{x_u}$ minimizes

$$(nh_e)^{-1}\sum_{i=1}^{n}\rho_\alpha\left(Y_i - c_\alpha - Q_{\alpha,-u}\,(W_{i,u}) - \beta^T_{x_u}V\left(\frac{x_u - X_{i,u}}{h_e}\right)\right)K_e\left(\frac{x_u - X_{i,u}}{h_e}\right) \tag{3.5}$$

Using (A.18) and by a similar methods to the proofs of (A.5) and (A.2) (see Appendix A), it can be obtained that

$$\sqrt{nh_e}\left(\hat{q}^{oracle}_{\alpha,u}(x_u) - q_{\alpha,u}(x_u) - h_e^p\frac{q^{(p)}_{\alpha,u}(x_u)}{p!}e_1^TB^{-1}\kappa_e\right) \to N\left(0, \sigma_0^2\right), \tag{3.6}$$

where $\kappa_e = \int t_1^p V(t_1)K_e(t_1)dt_1$, $B = \int V(t)V^T(t)K_e(t)dt$ and

$$\sigma_0^2 = \frac{\alpha(1-\alpha)}{g_u^2(0|x_u)f_u(x_u)}e_1^TB^{-1}\int V(t)V^T(t)K_e^2(t)dtB^{-1}e_1. \tag{3.7}$$

For more details on the local polynomial estimator for one dimensional conditional quantiles refer to Chaudhuri (1991) and Honda (2000).

The oracle efficient estimator

Now, we show that our estimator $\hat{q}^{e}_{\alpha,u}(x_u)$ (3.2) behaves analogously to the oracle estimator $\hat{q}^{oracle}_{\alpha,u}(x_u)$ above. Let $r_n = n^{\frac{\varepsilon_0}{2}}/\sqrt{nh_e}$ with ε_0 being a sufficiently small positive constant. For $|t_{i,n}| \leq Cr_n$, $(i = 1, \ldots, n)$, let $\mathbf{t}_n = (t_{1,n}, \ldots, t_{n,n})^T$. Denote by $V_{u,i} = V\left(\frac{X_{i,u}-x_u}{h_e}\right)$, $K_{u,i} = K_e\left(\frac{x_u - X_{i,u}}{h_e}\right)$ and

$$\hat{\beta}_{\mathbf{t}_n} = \arg\min_{a} \frac{1}{nh_e} \sum_{i=1}^{n} K_{u,i} \left| Y_i - c_\alpha - Q_{\alpha,-u}\left(W_{i,u}\right) - a^T V_{u,i} - t_{i,n} \right|.$$

Proposition 3.1. *Under the conditions C1 to C9, with probability one, it holds uniformly for* $|t_{i,n}| \leq Cr_n$, $i = 1, 2, \ldots, n$, *that*

$$\hat{\beta}_{\mathbf{t}_n} - \hat{\beta}^{oracle}_{x_u} = \frac{B_n^{-1} I\!E\left(K_{u,i} V_{u,i} g_u\left(0|X_{i,u}\right)\right)}{nh_e} \sum_{i=1}^{n} t_{i,n} + O\left(\frac{n^{-\varepsilon_0}}{\sqrt{nh_e}}\right),$$

where $\hat{\beta}^{oracle}_{x_u}$ *is as defined in (3.5) and* $B_n = \frac{1}{h_e} I\!E K_{u,i} g_u\left(0|X_{i,u}\right) V_{u,i} V_{u,i}^T$.

Theorem 3.2. *Under the conditions C1 to C9, it holds that*

$$\sqrt{nh_e}\left(\hat{q}^{e}_{\alpha,u}(x_u) - \hat{q}^{oracle}_{\alpha,u}(x_u)\right) = o_{I\!P}(1). \tag{3.8}$$

From Theorem 3.2, we see that $\hat{q}^{e}_{\alpha,u}(x_u)$ is asymptotically normally distributed with same mean and variance as $\hat{q}^{oracle}_{\alpha,u}(x_u)$. Therefore, our proposed estimator $\hat{q}^{e}_{\alpha,u}(x_u)$ is oracle efficient.

4 A simulated example

In this section, we provide the finite sample performance of our oracle efficient estimator (denoted in this section by OEE) vis-à-vis two alternative kernel estimators: the estimator of De Gooijer and Zerom(2003) (denoted as DGZ) and the back-fitting approach. We do not include the hybrid estimator of Horowitz and Lee (2005) in our comparison. But we think that the estimator of Horowitz and Lee (2005) will have a similar performance as ours at least for the i.i.d. data case. We use the standard normal density for all kernel functions: $K_1(\cdot)$, $K_2(\cdot)$, $K(\cdot)$, and $K_e(\cdot)$. These choices are consistent with the assumptions used to derive the asymptotic properties. As in DGZ, we assume the following data generating process,

$$Y_i = Q_\alpha(X_{i,1}, X_{i,2}) + 0.25\mathcal{E}_{\alpha,i}, \tag{4.1}$$

where the errors $\mathcal{E}_{\alpha,i}$ are *i.i.d.* $N(0,1)$ and the covariates X_1 and X_2 are bivariate normal with zero mean, unit variance, and correlation γ. We consider $\alpha = 0.5$ (the case of conditional median), correlations $\gamma = 0.2$ (low correlation between covariates), 0.8 (high correlation) and sample sizes $n = 100$, 200, 400 and 800. The conditional median of Y is assumed to be additive,

$$\begin{aligned} Q_{0.5}(x_1, x_2) &= q_{0.5,1}(x_1) + q_{0.5,2}(x_2), \\ &= 0.75x_1 + 1.5\sin(0.5\pi x_2). \end{aligned}$$

We simulate model (4.1) 41 times and in each simulation the three approaches are used to compute the additive median functions $q_{0.5,1}(\cdot)$ and $q_{0.5,2}(\cdot)$. To avoid the sensitivity of the performance of the compared approaches on bandwidth selection, we use the bandwidth values used in DGZ, although these values may not be optimal. To compute the oracle efficient median estimates: $\hat{q}^e_{0.5,1}(x_1)$ and $\hat{q}^e_{0.5,2}(x_2)$ (see (3.2)), we need values for $\hat{Q}^*_{\alpha,-1}$ and $\hat{Q}^*_{\alpha,-2}$ (see 3.1). The latter two in require $\hat{q}^*_{0.5,1}(x_1)$ and $\hat{q}^*_{0.5,2}(x_2)$ (see (2.3)), which in turn depend on $\hat{Q}_{0.5}(x_1, x_2)$ (see (2.4)). Thus, we need different bandwidth values at various stages. Instead of a single value, we let h (used for $\hat{Q}_{0.5}(x_1, x_2)$) vary with the variability of the covariates in the following way. For smoothing in the direction of X_1, $h = 3s_1n^{-1/5}$ and for smoothing in the direction of X_2, $h = s_2n^{-1/5}$ where s_k is the sample standard deviation of X_k $(k = 1, 2)$. We also need to choose h_1 and h_2. We use $\{h_1 = 3s_1n^{-1/5}, h_2 = s_2n^{-1/5}\}$ for $\hat{q}^*_{0.5,1}(x_1)$ and $\{h_1 = s_2n^{-1/5}, h_2 = 3s_1n^{-1/5}\}$ for $\hat{q}^*_{0.5,2}(x_2)$. Finally, we take $h_e = h$.

Table 1: The average absolute deviation errors (AADE) of the estimated additive components.

γ	n	$\tilde{q}_{0.5,1}(\cdot)$			$\tilde{q}_{0.5,2}(\cdot)$		
		OEE	DGZ	Back-fitting	OEE	DGZ	Back-fitting
0.2	100	0.0383	0.1374	0.0597	0.1124	0.1818	0.1425
	200	0.0324	0.1066	0.0511	0.0883	0.1272	0.1120
	400	0.0214	0.0734	0.0431	0.0678	0.0936	0.0889
	800	0.0143	0.0625	0.0264	0.0546	0.0703	0.0704
0.8	100	0.0522	0.1365	0.1124	0.1491	0.4865	0.1783
	200	0.0505	0.1093	0.1263	0.1232	0.4350	0.1767
	400	0.0526	0.0985	0.0780	0.1027	0.4009	0.1467
	800	0.0526	0.0882	0.0630	0.0928	0.3690	0.1124

We compare our median estimates $\hat{q}^e_{0.5,1}(x_1)$ and $\hat{q}^e_{0.5,2}(x_2)$ (OEE) with (DGZ) and the back-fitting approach. The three approaches are compared based on the average absolute deviation error (AADE). First, the absolute deviation error (ADE) for each estimated function $\tilde{q}_{0.5,k}(\cdot)$, $k = 1, 2$ is computed at each replication j, i.e. $ADE_j(k) = \text{Average}\{|\tilde{q}_{0.5,k}(X_{i,k}) - q_{0.5,k}(X_{i,k})|\}_i^n$ $(j = 1, \ldots, 41; k = 1, 2)$ where the average is only taken for $X_k \in [-2, 2]$, to avoid data sparsity. Then, the AADE is defined as the average of the ADE over the 41 replications. In Table 1, we report the AADE values by changing γ and/or n.

When $\gamma = 0.2$ and $n \leq 200$, the OEE is significantly more accurate than DGZ. While the performance of the three estimators improves with increasing sample size, the OEE maintains its superiority at all sample sizes. For $\gamma = 0.8$, the performance of the three estimators decreases although the OEE still achieves a decent accuracy at all sample sizes especially for the estimation of $q_{0.5,1}(\cdot)$. The DGZ is highly inaccurate even at sample sizes as large as $n = 800$. Although the back-fitting approach tends to converge a lot faster than DGZ, its accuracy is still worse than OEE. From the above simulation experiment, we observe that the OEE is not only a superior approach when compared to existing kernel approaches, it is also robust against highly correlated covariates. For large sample sizes, the back-fitting approach tend to be competitive against OEE. One advantage of the OEE is that it is computed in two easy and fast steps with guaranteed convergence while the back-fitting is iterative and convergence is not assured.

5 Additive models for ambulance travel times

The most common performance measure of emergency medical service (EMS) operations is the fraction of calls with a *response time* below one or more thresholds. For instance, reaching 90% of urgent urban calls in 9 minutes is a common target in North America and the National Health Service in the U.K. sets targets of 75% in 8 minutes and 95% in 14 minutes for urgent urban calls (Budge, Ingolfsson and Zerom, 2008). Note that these performance targets correspond to quantiles of the response time distribution. Budge *et al.* (2008) introduce the following semi-parametric model to predict the *travel time* (travel time of an ambulance to the scene of an emergency is typically the largest component of response time) distribution of high-priority calls for the city of Calgary, Canada,

$$Y_i = \mu(X_{1,i}, X_{2,i}) e^{(\sigma\ \mathcal{E}_{\alpha,i})}, \quad (i = 1, \ldots, n), \tag{5.1}$$

where i denotes a 911 call, Y denotes travel time and the two predictors X_1 and X_2 are network distance and time-of-day, respectively. The error $\mathcal{E}_{\alpha,i}$ follows a standard t-distribution with τ degrees of freedom, i.e. $\mathcal{E}_{\alpha,i} \sim t_\tau(0, 1)$ and σ is a scaling parameter. Under this set-up, the function $\mu(x_1, x_2)$ represents the conditional median of Y given $(X_1, X_2) = (x_1, x_2)$. In 2003, Calgary EMS responded to $n = 7457$ high priority calls that involves heart problems, breathing problems, traffic accident, building fire, unconsciousness, house fire, fall, convulsions and seizures, hemorrhage and lacerations, traumatic injuries, and unknown problem.

Budge *et al.* (2008) assume that the conditional median of travel time to be additive,

$$\mu(x_1, x_2) = \mu_0 + \mu_1(x_1) + \mu_2(x_2), \tag{5.2}$$

where μ_0 is a constant and no parametric form is imposed on the functions $\mu_1(x_1)$ and $\mu_2(x_2)$ except that they should be arbitrary twice continuously-differentiable. With (5.2), the travel time distribution can be fully characterized by conditional quantiles as follows,

$$Q_\alpha(X_{1,i}, X_{2,i}) = [\mu_0 + \mu_1(X_{1,i}) + \mu_2(X_{2,i})]e^{(\sigma\ Q_\alpha(\tau))}, \tag{5.3}$$

where $Q_\alpha(x_1, x_2)$ denotes the α-th conditional quantile of Y given $(X_1, X_2) = (x_1, x_2)$ and $Q_\alpha(\tau)$ is the α-th quantile of a $t_\tau(0, 1)$-distribution. Note that, under the above model set-up, the α-th conditional quantile of travel time at all α is in fact additive, i.e.,

$$Q_\alpha(X_{1,i}, X_{2,i}) = c_\alpha + q_{\alpha,1}(X_1) + q_{\alpha,2}(X_2), \tag{5.4}$$

where $c_\alpha = \mu_0 e^{(\sigma\ Q_\alpha(\tau))}$, $q_{\alpha,1}(x_1) = \mu_1(x_1)\ e^{(\sigma\ Q_\alpha(\tau))}$ and $q_{\alpha,2}(x_2) = \mu_2(x_2)e^{(\sigma\ Q_\alpha(\tau))}$.

Motivated by the additive conditional quantile set-up (5.4), our aim is to compare our oracle efficient estimates of the additive quantiles and the corresponding estimates from the semi-parametric approach. It should be noted that the paper of Budge *et al.* (2008) has a much wider scope and our aim here is only illustrative. Although limited in scope, this example serves two purposes. First, we illustrate how to implement our estimator in practice with a novel data set. Second, we use our estimates to validate, albeit indirectly, the distributional assumption of the semi-parametric model. Although both the semi-parametric approach and the non-parametric approach rely on an underlying additive structure, the non-parametric estimator does not impose an assumption on the distribution of the travel time and hence is more general. We consider three quantile levels $\alpha = 0.25$, 0.5 and 0.75. In the estimation of the additive median components ($\mu_1(\cdot)$ and $\mu_2(\cdot)$) for the semi-parametric model, we use cubic smoothing splines with degrees of freedom chosen via minimization of Akaike's information (AIC) criterion. All unknown components of the semi-parametric model are estimated using the penalized maximum likelihood algorithm of Rigby and Stasinopoulos (2005) which is readily available in the `R` library GAMLSS.

To implement our oracle efficient estimator, we need to select bandwidth values. As in Section 4, we assume that the bandwidth values used to estimate $\hat{q}^*_{\alpha,u}$ are the same as those for estimating $\hat{q}^e_{\alpha,u}$ But, at the same time, to allow varying level of smoothness for the two additive quantile functions (corresponding to distance and time-of-day), we adopt separate bandwidth values. So, for each quantile level α, we select two bandwidth values using a rule-of-thumb suggested by Fan and Gijbels (1996) and also adopted by Horowitz and Lee (2005). As an alternative one may also use the data-driven bandwidth selection method by Yu and Lu (2004). We obtain the following bandwidth values for smoothing in the direction of X_1 (distance): 0.58 ($\alpha = 0.5$), 0.62 ($\alpha = 0.25$), and 0.65 ($\alpha = 0.75$). Similarly, for smoothing in the direction of X_2 (time-of-day), the selected bandwidth values are 1.13 ($\alpha = 0.5$), 1.29 ($\alpha = 0.25$), and 1.06 ($\alpha = 0.75$). We use the standard normal density for all kernel functions: $K_1(\cdot)$, $K_2(\cdot)$, $K(\cdot)$, and $K_e(\cdot)$.

In Figure 1 we plot the conditional median estimates for both our estimator and the semi-parametric approach. Those in panel (a) and panel (c) correspond to our median estimates corresponding to distance (X_1) and time-of-day (X_2), respectively. The confidence intervals (at the 95% level) for both median estimates are based on the asymptotic variance given in equation 3.7 although we do not do any bias correction. The unknown components of the asymptotic variance are calculated using kernel estimates. On the other hand, panel (b) and panel (d) show the estimated median functions $\hat{\mu}_1(x_1)$ and $\hat{\mu}_2(x_2)$ from the semi-parametric method. Comparing the corresponding median estimates from the two approaches, it is interesting to see that both produce closely similar estimated functions. The only difference is that our estimates are not as smooth. In Figure 2 we plot the estimated additive conditional quantile functions for $\alpha = 0.25$ (panels (a) and (b)) and for $\alpha = 0.75$ (panels (c) and (d)). Solid lines correspond to our estimates and dashed lines to the semi-parametric approach. Note that the general shape of both quantile functions is similar to those of the median for both distance and time-of-day. As in the case of median, the estimates from the proposed approach are less smooth. It is also interesting to see that the estimated quantiles from both approaches are very close although they seem to differ slightly in their estimated peaks. It should be noted that quantile estimates of the semi-parametric approach are functions of the estimated medians $\hat{\mu}_1(\cdot)$ and $\hat{\mu}_2(\cdot)$ as well as $\hat{\sigma}$ and $\hat{\tau}$. We find that $\hat{\sigma} = 0.24$ and kurtosis $\hat{\tau} = 3.35$ where the latter estimate indicates leptokurtosis in travel times due to infrequently occuring large travel times. Given that the semi-parametric conditional quantile estimates mimics the distribution-free conditional quantile estimates (based on our approach), we may conclude that the conditional distribution of travel time is leptokurtic and the student t-distribution is a reasonable way to capture it.

For a complete discussion of the practical implications of the conditional quantile modeling of ambulance travel times to operational planning and related decision problems, we refer the reader to Budge *et al.* (2008). These authors also discuss the additive median function estimates in the context of existing operations research models.

6 Concluding remarks

We have introduced two simple kernel estimators for estimating additive components of an additive quantile regression model. Taken together, these estimators are are offered as a better alternative to existing kernel-based methods (De Gooijer and Zerom, 2003 and Yu and Lu, 2004) due to better efficiency and computational convenience. We provide asymptotic properties for both estimators. The validity of the asymptotic properties is established for dependent data and in particular for β-mixing processes, that include independent and time series data as special cases. On the other hand, the asymptotic validity of Horowitz and Lee (2005) is proved only for independent data.

It is well known that proper choice of the bandwidth is critical for the accuracy of any nonparametric function. This paper does not address this issue for the proposed nonparametric estimators. In practice, it is desirable to have a feasible data-driven method of choosing bandwidth values. For example, Yu and Lu (2004) suggest a simple practical bandwidth selection rule for their back-fitting approach. We defer this important topic for future research.

Appendix: Proofs

In this Appendix, we provide proofs of theoretical results. For better exposition of the derivations, we divide this section into two appendices. In Appendix A, we provide proofs of the main results, i.e. theorems and propositions. In Appendix B, proofs are provided for intermediate lemmas that are used in the derivations of the theorems and propositions.

Appendix A

Proof of Theorem 2.1

We note that $\hat{q}^*_{\alpha,u}(x_u) - q^*_{\alpha,u}(x_u) = S_{1,n} + S_{2,n} + S_{3,n}$, where

$$S_{1,n} = \frac{1}{nh_1}\sum_{i=1}^{n} K\left(\frac{x_u - X_{i,u}}{h_1}\right)\frac{\hat{f}_w(W_{i,u})}{\hat{f}(X_i)}\left(\hat{Q}_\alpha(X_i) - Q_\alpha(X_i)\right),$$

$$S_{2,n} = \frac{1}{nh_1}\sum_{i=1}^{n} K\left(\frac{x_u - X_{i,u}}{h_1}\right)\left(\frac{\hat{f}_w(W_{i,u})}{\hat{f}(X_i)} - \frac{f_w(W_{i,u})}{f(X_i)}\right)Q_\alpha(X_i)$$

and $S_{3,n} = \frac{1}{n}\sum_{i=1}^{n}\zeta_i - q^*_{\alpha,u}(x_u)$ with

$$\zeta_i = \zeta_i(x_u) = \frac{1}{h_1}K\left(\frac{x_u - X_{i,u}}{h_1}\right)\frac{f_w(W_{i,u})\,Q_\alpha(X_i)}{f(X_i)}. \tag{A.1}$$

Hereby, we investigate $S_{1,n}$, $S_{2,n}$ and $S_{3,n}$ in the following three steps, in a reverse order.

We first consider $S_{3,n}$. Using equation (1.2), C1, variable substitution, Taylor expansion, we have

$$\begin{aligned} I\!E\zeta_i &= \int K(t_u)\,Q_\alpha(x_u + h_1 t_u, w_u) f_w(w_u)\,dt_u dw_u = c_\alpha + \int K(t_u)\,q_{\alpha,u}(x_u + h_1 t_u) dt_u \\ &= q^*_{\alpha,u}(x_u) + \frac{q^{(p)}_{\alpha,u}(x_u)}{p!}\kappa_p h_1^p + o\left(h_1^p\right). \end{aligned} \tag{A.2}$$

Let $\bar{\zeta}_i = h_1^{\frac{1}{2}}(\zeta_i - I\!E\zeta_i)$. By variable substitution, Taylor expansion and i) of C4, it can be inferred that

$$I\!E\bar{\zeta}_i^2 = h_1^2\int\left(\frac{K(t_u)}{h_1}\right)^2\frac{Q_\alpha^2(x_u + h_1 t_u, w_u) f_w^2(w_u)}{f(x_u + h_1 t_u, w_u)}dt_u dw_u(1 + o(1)) = \sigma_2^2(1 + o(1)) \tag{A.3}$$

where σ_2^2 is as defined in Theorem 2.1. Using C6 and variable substitution, it follows that $\mathrm{cov}(\bar{\zeta}_1, \bar{\zeta}_{i+1}) = O(h_1)$. Let m_n be a sufficiently large integer with the restriction $m_n h_1 \to 0$. For $\delta > 2$ introduced in C4, by Davydov (1968) inequality, C4 and the fact that $\left(I\!E|\bar{\zeta}_1|^\delta\right)^{\frac{2}{\delta}} = O\left(h_1^{\frac{2}{\delta}-1}\right)$, we obtain

$$\sum_{i=m_n}^{n}(n-i)|\mathrm{cov}(\bar{\zeta}_1, \bar{\zeta}_{i+1})| \le Cn\sum_{i=m_n}^{n}\pi_i^{1-\frac{2}{\delta}}\left(I\!E|\bar{\zeta}_1|^\delta\right)^{\frac{2}{\delta}} = o(1).$$

Hereby, we know that

$$I\!D\left(\sum_{i=1}^{n}\frac{\bar{\zeta}_i}{\sqrt{n}}\right) = I\!E\bar{\zeta}_1^2 + \left(\sum_{i=1}^{m_n} + \sum_{i=m_n+1}^{n}\right)\frac{2(n-i)}{n}\mathrm{cov}(\bar{\zeta}_1, \bar{\zeta}_{i+1}) = \sigma_2^2(1 + o(1)) \tag{A.4}$$

where the operator $I\!\!D$ is defined in Lemma 2. Then, in view of (A.4), C4, C5, and following the same line as the proofs of (32)-(35) in Cai and Ould-Saïd (2003), we infer that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\zeta}_i \to N(0,\sigma_2^2). \tag{A.5}$$

Thus, it follows from (A.5) and (A.2) that

$$\sqrt{nh_1}\left(S_{3,n}-\frac{q^{(p)}_{\alpha,u}(x_u)\kappa_p}{p!}h_1^p\right)\to N\left(0,\sigma_2^2\right). \tag{A.6}$$

Moving to $S_{2,n}$, we first write it as $S_{2,n}=I_1-I_2+I_3+I_4+I_5$, where

$$I_1=\frac{1}{nh_1}\sum_{i=1}^{n}K\left(\frac{x_u-X_{i,u}}{h_1}\right)\left(\hat{f}_w\left(W_{i,u}\right)-I\!\!E_i\hat{f}_w\left(W_{i,u}\right)\right)\left(\frac{1}{\hat{f}(X_i)}-\frac{1}{f(X_i)}\right)Q_\alpha(X_i),$$

$$I_2=\frac{1}{nh_1}\sum_{i=1}^{n}K\left(\frac{x_u-X_{i,u}}{h_1}\right)\frac{\hat{f}(X_i)-f(X_i)}{f^2(X_i)}Q_\alpha(X_i)I\!\!E_i\hat{f}_w\left(W_{i,u}\right),$$

$$I_3=\frac{1}{nh_1}\sum_{i=1}^{n}K\left(\frac{x_u-X_{i,u}}{h_1}\right)\frac{(\hat{f}(X_i)-f(X_i))^2}{\hat{f}(X_i)f^2(X_i)}Q_\alpha(X_i)I\!\!E_i\hat{f}_w\left(W_{i,u}\right),$$

$$I_4=\frac{1}{nh_1}\sum_{i=1}^{n}K\left(\frac{x_u-X_{i,u}}{h_1}\right)\frac{Q_\alpha(X_i)}{f(X_i)}\left(\hat{f}_w\left(W_{i,u}\right)-I\!\!E_i\hat{f}_w\left(W_{i,u}\right)\right),$$

$$I_5=\frac{1}{nh_1}\sum_{i=1}^{n}K\left(\frac{x_u-X_{i,u}}{h_1}\right)\frac{Q_\alpha(X_i)}{f(X_i)}\left(I\!\!E_i\hat{f}_w\left(W_{i,u}\right)-f_w\left(W_{i,u}\right)\right)$$

and the operator $I\!\!E_i$ is as defined in Lemma 4. By C1 and the dominated convergence theorem, it can be inferred that $f_w(w_u)$ has the partial derivatives up to order $\bar{p}$. Thus, it holds that

$$I\!\!E_i\hat{f}_w\left(W_{i,u}\right)-f_w\left(W_{i,u}\right)=\frac{1}{h_2^{d-1}}I\!\!E_iL_1\left(\frac{W_{i,u}-W_{j,u}}{h_2}\right)-f_w(W_{i,u})=O\left(h_2^{\bar{p}}\right).$$

From this, WLLN and $\bar{p}>\frac{pd}{p+1}$, we know that $I_5=O_{I\!\!P}\left(h_2^{\bar{p}}\right)=o_{I\!\!P}\left((nh_1)^{-\frac{1}{2}}\right)$. By virtue of the uniform weak law of large number, it can be inferred that

$$\hat{f}_w\left(W_{i,u}\right)-I\!\!E_i\hat{f}_w\left(W_{i,u}\right)=O_{I\!\!P}\left(\left(nh_2^{d-1}\right)^{-\frac{1}{2}}\right). \tag{A.7}$$

Let

$$\xi_{ij}=K\left(\frac{x_u-X_{i,u}}{h_1}\right)\left(L_1\left(\frac{W_{i,u}-W_{j,u}}{h_2}\right)-I\!\!E_iL_1\left(\frac{W_{i,u}-W_{j,u}}{h_2}\right)\right)\frac{Q_\alpha(X_i)}{f(X_i)},$$

$$J_1=\frac{1}{n^2h_1h_2^{d-1}}\sum_{i=1}^{n}\sum_{1\le j\ne i\le n}(\xi_{ij}-I\!\!E_j\xi_{ij})\quad\text{and}\quad J_2=\frac{2(n-1)}{n^2h_1h_2^{d-1}}\sum_{j=1}^{n}I\!\!E_j\xi_{ij}.$$

Thus, $I_4=J_1+J_2$. Noting that $I\!\!E_i\xi_{ij}=0$, J_1 is a degenerated U-statistic. By C4 and C6, it can be inferred that $\sup_{i\ne j}I\!\!E|\xi_{ij}|^k=O\left(h_1h_2^{d-1}\right)$. Hence, applying Lemma 3 and vi) of C5, we obtain $J_1 = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right)$. To arrive at this result, we use the same techniques as that of (A.13) - see below. To deal with J_2, let

$$\varsigma_j = I\!E_j\left\{K\left(\frac{x_u - X_{i,u}}{h_1}\right) L_1\left(\frac{W_{i,u} - W_{j,u}}{h_2}\right)\frac{Q_\alpha(X_i)}{f(X_i)}\right\}.$$

As in (A.4) and taking $m_n h_1 = o(1)$, it follows that $I\!E(\varsigma_1^2) = O\left(\left(h_1 h_2^{d-1}\right)^2\right)$,

$$\sum_{j=1}^{m_n}(n-j)|\mathrm{cov}(\varsigma_1, \varsigma_{j+1})| = O\left(nm_n\left(h_1 h_2^{d-1}\right)^2\right),$$

and

$$\sum_{j=m_n+1}^{n}(n-j)|\mathrm{cov}(\varsigma_1, \varsigma_{j+1})| \le Cn\sum_{j=m_n+1}^{n}\beta_j^{1-\frac{2}{\delta}}\left(\left(h_1 h_2^{d-1}\right)^{\delta}\right)^{\frac{2}{\delta}} = Cn\sum_{j=m_n+1}^{n}\beta_j^{1-\frac{2}{\delta}}\left(h_1 h_2^{d-1}\right)^2.$$

From the above three equations, we see that $I\!D\left(J_2\right) = o\left(\frac{1}{nh_1}\right)$. Thus, $J_2 = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right)$ and so does for I_4. We begin to handle I_2. Note that

$$I\!E_i\hat{f}\left(X_i\right) - f\left(X_i\right) = O\left(n^{-\bar{p}/(d+\bar{p})}\right) = o\left((nh_1)^{-\frac{1}{2}}\right), \tag{A.8}$$

where the last equation follows from $\bar{p} > \frac{pd}{p+1}$ and $h_1 \le Cn^{-\frac{1}{2p+1}}$. Similar to the proof of I_4, it can be obtained that

$$\frac{1}{nh_1}\sum_{i=1}^{n}K\left(\frac{x_u - X_{i,u}}{h_1}\right)\frac{\hat{f}(X_i) - I\!E_i\hat{f}(X_i)}{f^2(X_i)}Q_\alpha(X_i)I\!E_i\hat{f}_w\left(W_{i,u}\right) = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right). \tag{A.9}$$

Hereby, from (A.8), (A.9) and WLLN, it can be obtained that $I_2 = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right)$. Moving to I_3, using WLLN and (A.8), we know that

$$|\hat{f}(X_i) - f(X_i)| \le |\hat{f}(X_i) - I\!E_i\hat{f}(X_i)| + |I\!E_i\hat{f}(X_i) - f(X_i)| = O_{I\!P}\left(n^{-\frac{\bar{p}}{2(d+\bar{p})}}\right). \tag{A.10}$$

Combination of this, $\bar{p} > \frac{pd}{p+1}$ and WLLN implies that $I_3 = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right)$. Finally, from (A.10), (A.7), WLLN and $\bar{p} > \frac{pd}{p+1}$, it can be inferred that

$$I_1 = O_{I\!P}\left(n^{-\frac{1}{2}\left(1-\frac{d-1}{d+\bar{p}}\right)}\cdot n^{-\frac{\bar{p}}{2(d+\bar{p})}}\right) = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right).$$

We now consider $S_{1,n}$. By the same method as that of $S_{2,n}$, it can be proved that $S'_{1,n} = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right)$, where $S'_{1,n} = S_{1,n} - S''_{1,n}$ and

$$S''_{1,n} = \frac{1}{nh_1}\sum_{i=1}^{n}K\left(\frac{x_u - X_{i,u}}{h_1}\right)\frac{f_w\left(W_{i,u}\right)}{f(X_i)}\left(\hat{Q}_\alpha(X_i) - Q_\alpha(X_i)\right). \tag{A.11}$$

Inserting (B.1) into (A.11), we obtain two terms (denoted by $S_{4,n}$ and $S_{5,n}$) where the second (which is $S_{5,n}$) is the remainder term that is of order $o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right)$ because $\kappa < \frac{2p+3}{3d(2p+1)}$. Let $\varphi_{ij} = e_1^T B_{i,n}^{-1} K_{ji} V_{ji} \left(I\!I\left(Y_j \le V_{ij}\beta_{X_j}\right) - I\!I\left(\mathcal{E}_{\alpha,j} \le 0\right)\right)$. By SLLN, $\kappa > \frac{1}{2p+1}$ and Taylor expansion for $Q_\alpha(x)$, it could be proved that

$$\frac{1}{nh^d} \sum_{1\le j\ne i\le n} \varphi_{ij} = O\left(h^p\right) = o\left(\frac{1}{\sqrt{nh_1}}\right) \tag{A.12}$$

holds almost surely and uniformly for $i = 1, \ldots, n$. Thus, by WLLN, we obtain that

$$\frac{1}{n^2 h_1 h^d} \sum_{i=1}^{n} \sum_{1\le j\ne i\le n} K\left(\frac{x_u - X_{i,u}}{h_1}\right) \frac{f_w\left(W_{i,u}\right)}{f(X_i)} \varphi_{ij} = o_{I\!P}\left(\frac{1}{\sqrt{nh_1}}\right).$$

Therefore,

$$S_{4,n} = \frac{1}{n^2 h_1 h^d} \sum_{i=1}^{n} \sum_{1\le j\ne i\le n} \eta_{ij} + o_{I\!P}\left(\frac{1}{\sqrt{nh_1}}\right),$$

where

$$\eta_{ij} = K_1\left(\frac{x_u - X_{i,u}}{h_1}\right) \frac{f_w\left(W_{i,u}\right)}{f(X_i)}\; e_1^T B_{i,n}^{-1} L_{ji} A_{ji} \left(\alpha - I\!I\left(\mathcal{E}_{\alpha,j} \le 0\right)\right).$$

Note that $I\!E_i \eta_{ij} = 0$. According to H-decomposition of U-statistic,

$$S_{4,n} = \frac{1}{n^2 h_1 h^d} \sum_{i=1}^{n} \sum_{1\le j\ne i\le n} \left(\eta_{ij} - I\!E_j \eta_{ij}\right) + \frac{n-1}{n^2 h_1 h^d} \sum_{j=1}^{n} I\!E_j \eta_{ij} + o_{I\!P}\left(\frac{1}{\sqrt{nh_1}}\right).$$

Denote the first two terms of $S_{4,n}$ by $S'_{4,n}$ and $S''_{4,n}$, respectively. Trivially, $S'_{4,n}$ is a degenerated U-statistic. Therefore, by applying iv) of C5 and Lemma 3 and taking $\frac{1}{s} = 1 - \frac{k+1}{\theta} - \epsilon$ and $k = 2$, we know that

$$\begin{aligned} (nh_1)^{\frac{k}{2}} I\!E\left|S'_{4,n}\right|^k &\le \frac{(nh_1)^{\frac{k}{2}} n^k}{(n^2 h_1 h^d)^k} \left(1 + \sum_{i=1}^{n} i^k \beta_i^{1-\frac{1}{s}}\right) \sup_{1\le i\ne j\le n} \left(I\!E|\eta_{ij}|^{sk}\right)^{\frac{1}{s}} \\ &= O\left(\left(nh^d \left(h_1 h^d\right)^{\frac{3}{\theta}+\epsilon}\right)^{-1}\right) = o(1). \end{aligned} \tag{A.13}$$

Hereby, from this and Markov inequality, it can be inferred that $S'_{4,n} = o_{I\!P}\left((nh_1)^{-\frac{1}{2}}\right)$. To simplify notations, let $\phi_j = \phi_j(x_u) = I\!E_j \eta_{ij}$. By variable substitution and $h/h_1 \to 0$, it can be inferred that

$$\begin{aligned} I\!E(\phi_1^2) &= \alpha(1-\alpha) I\!E\left(I\!E_j K\left(\frac{x_u - X_{i,u}}{h_1}\right) \frac{f_w\left(W_{i,u}\right)}{f(X_i)} e_1^T B_{i,n}^{-1} L_{ji} V_{ji}\right)^2 \\ \sim\; & \alpha(1-\alpha) h^{2d} I\!E\left(\int K\left(\frac{x_u - hs_u - X_{j,u}}{h_1}\right) \frac{e_1^T B_2^{-1} L\left(s\right) V_1\left(s\right)}{g(0|X_j + hs) f(X_j + hs)} f_w\left(W_{j,u} + hs^{-u}\right) ds\right)^2 \\ \sim\; & C h_1 h^{2d} \sigma_1^2, \end{aligned} \tag{A.14}$$

where $B_2 = \int V_1(s)V_1^T(s)L(s)ds$, and the fact $e_1^T B_2^{-1} \int V_1(s)L(s)ds = 1$, which follows from properties of the inverse matrix and the adjoin matrix, is used in the last step. Taking $m_n h_1 = o(1)$ and using the same arguments used earlier, we can obtain that

$$\sum_{j=1}^{m_n}(n-j)|\text{cov}(\phi_1, \phi_{j+1})| = O\left(nm_n h^{2d} h_1^2\right).$$

Further, according to Davydov's inequality, $\delta > 2$ in A4 and iv) of A5, it can be shown that

$$\sum_{j=m_n+1}^{n}(n-j)|\text{cov}(\phi_1, \phi_{j+1})| \le Cn \sum_{j=m_n+1}^{n} \pi_j^{1-\frac{2}{\delta}} \left(h^{d\delta} h_1\right)^{\frac{2}{\delta}} = Cnm_n^{1-\theta(1-\frac{2}{\delta})} h^{2d} h_1^{\frac{2}{\delta}}.$$

Thus using (A.14) and the above two equations, we see that $I\!D\left(S''_{4,n}\right) = \frac{\sigma_2^2}{nh_1}$. Further, following the same line of proofs as(A.5), it can be obtained that

$$\sqrt{nh_1} \cdot \frac{n-1}{n^2 h_1 h^d} \sum_{j=1}^{n} \phi_j \to N\left(0, \sigma_1^2\right). \tag{A.15}$$

From the foregoing proofs, it can be observed that $\frac{1}{nh_1^{1/2}} \sum_{i=1}^n \bar{\zeta}_i$ and $\frac{n-1}{n^2 h_1 h^d} \sum_{i=1}^n \phi_i$ are the two leading terms for the sum $S_{1,n} + S_{2,n} + S_{3,n}$. On the other hand, all other terms are asymptotically negligible and convergence at the rate of $o_{I\!P}\left(\frac{1}{\sqrt{nh_1}}\right)$. For any $1 \le i, j \le n$, we note from the conditional expectation that $\text{cov}\left(\bar{\zeta}_i, \phi_j\right) = 0$. Therefore, the two leading terms are asymptotically uncorrelated.

In view of the above arguments, (A.15) and (A.6), the asymptotic normal relationship (2.5) can be inferred directly.

Proof of Proposition 2.2

Note that $\hat{c}_\alpha = \frac{1}{n} \sum_{i=1}^n \hat{Q}_\alpha(X_i)$. Let

$$\hat{c}_1 = \frac{1}{n} \sum_{i=1}^{n} \left(\hat{Q}_\alpha(X_i) - Q_\alpha(X_i)\right) \quad \text{and} \quad \hat{c}_2 = \frac{1}{n} \sum_{i=1}^{n} \left(Q_\alpha(X_i) - I\!E Q_\alpha(X_i)\right).$$

Then, $\hat{c}_\alpha - c_\alpha = \hat{c}_1 + \hat{c}_2$. Similar to the proof of (A.12), it can be obtained that

$$\hat{c}_1 = \frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{1 \le j \ne i \le n} e_1^T B_{i,n}^{-1} L_{ji} V_{ji} \left[\alpha - I\!I\left(\mathcal{E}_{\alpha,i} \le 0\right)\right] + O_{I\!P}\left(h^p + \left(\frac{\log n}{nh^d}\right)^{3/4}\right).$$

According to iii) of C5, the remainder term is of order $o_{I\!P}\left(n^{-\frac{p}{2p+1}}\right)$. Denote the first term on the right hand side of the relationship above by $\hat{c}'_1$. And let $\eta_{i,j} = e_1^T B_{i,n}^{-1} L_{ji} V_{ji} \left[\alpha - I\!I\left(\mathcal{E}_{\alpha,i} \le 0\right)\right]$. Since $I\!E_j \eta_{i,j} = 0$, by H-decomposition of U-statistic, we have

$$\hat{c}'_1 = \frac{1}{n^2 h^d} \sum_{i=1}^{n} \sum_{1 \le j \ne i \le n} \left(\eta_{i,j} - I\!E_i \eta_{i,j}\right) + \frac{n-1}{n^2 h^d} \sum_{i=1}^{n} I\!E_i \eta_{i,j} = A_1 + A_2.$$

Since A_1 is a degenerated U-statistic, by Lemma 3, iv) of C5, and taking $\frac{1}{s} = 1 - \frac{k+1}{\theta} - \varepsilon$ and $k = 2$, we have

$$(nh_1)^{k/2}\mathbb{E}|A_1|^k \leq C(nh_1)^{k/2}\frac{n^k(\sup_{i\neq j}\mathbb{E}|\eta_{i,j}|^{sk})^{1/s}}{(n^2h^d)^k} \leq C(nh_1)^{k/2}\frac{n^kh^{d/s}}{(n^2h^d)^k} \leq \frac{Ch_1}{nh^{d+\frac{3}{\theta}+\varepsilon}} \to 0.$$

Thus, it follows that $A_1 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_1}}\right)$. By WLLN, it can be derived that $A_2 = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$. From the two relationships above and $\hat{c}_2 = o_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_1}}\right)$, Proposition 2.2 holds.

Proof of Proposition 3.1

For $a \in \mathbb{R}^p$, denote by $\Delta_n(a, \mathbf{t}_n) = \Delta_{n,1}(a, \mathbf{t}_n) - \Delta_{n,1}(\beta_{x_u}, \mathbf{0})$, $\Delta_{n,1}(a, \mathbf{t}_n) = \sum_{i=1}^n \Delta_{i,1}(a, t_{i,n})$, $\Delta_{i,2}(a, t_{i,n}) = \Delta_{i,1}(a, t_{i,n}) - \Delta_{i,1}(a, 0)$,

$$\Delta_{i,1}(a, t_{i,n}) = \sum_{i=1}^n K_{u,i}V_{u,i}\left[\alpha - \mathbb{I}\left(\mathcal{E}_{\alpha,i} \leq a^TV_{u,i} + t_{i,n} - q_{\alpha,u}(X_{i,u})\right)\right].$$

Let $G(y|x_u)$ be the conditional distribution function of $\mathcal{E}_{\alpha,i}$ given that $X_{i,u} = x_u$. Then, by Taylor expansion and condition C8, it can be inferred that

$$\begin{aligned}\frac{\mathbb{E}\Delta_n(a, \mathbf{t}_n)}{nh_e} &= \sum_{i=1}^n \mathbb{E}\frac{K_{u,i}V_{u,i}}{nh_e}\left[G\left(a^TV_{u,i} + t_{i,n} - q_{\alpha,u}(X_{i,u})|X_{i,u}\right) - G\left(\beta_u^TV_{u,i} - q_{\alpha,u}(X_{i,u})|X_{i,u}\right)\right] \\ &= B_n(a - \beta_u) + \frac{\mathbb{E}\left(K_{u,i}V_{u,i}g_u(0|X_{i,u})\right)}{nh_e}\sum_{i=1}^n t_{i,n} + O\left(r_n^2\right). \qquad \text{(A.16)}\end{aligned}$$

From the definition of $\hat{\beta}_{x_u}$, we know that $\Delta_{n,1}\left(\hat{\beta}_{\mathbf{t}_n}, \mathbf{t}_n\right) = O(1)$. From this, (A.16), Lemma 12 and the following relationship

$$-\Delta_{n,1}(\beta_{x_u}, \mathbf{0}) = [\Delta_n(a, \mathbf{t}_n) - \mathbb{E}\Delta_n(a, \mathbf{t}_n)] + \mathbb{E}\Delta_n(a, \mathbf{t}_n) - \Delta_{n,1}(a, \mathbf{t}_n),$$

it can be inferred that

$$\begin{aligned}\hat{\beta}_{\mathbf{t}_n} - \beta_{x_u} &= -\frac{B_n^{-1}\mathbb{E}\left(K_{u,i}g_u(0|X_{i,u})V_{u,i}\right)}{nh_e}\sum_{i=1}^n t_{i,n} \\ &- \frac{B_n^{-1}}{nh_e}\sum_{i=1}^n K_{u,i}V_{u,i}\left[\alpha - \mathbb{I}\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - \beta_{x_u}^TV_{u,i} \leq 0\right)\right] + O\left(\frac{n^{-\varepsilon_0}}{\sqrt{nh_e}}\right). \qquad \text{(A.17)}\end{aligned}$$

In the relationship above, if we set $\mathbf{t}_n = \mathbf{0}$, then we can derive that

$$\begin{aligned}\hat{\beta}_{x_u}^{oracle} - \beta_{x_u} &= -\frac{B_n^{-1}}{nh_e}\sum_{i=1}^n K_{u,i}V_{u,i}\left[\alpha - \mathbb{I}\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - \beta_{x_u}^TV_{u,i} \leq 0\right)\right] \\ &+ O\left(\frac{n^{-\varepsilon_0}}{\sqrt{nh_e}}\right). \qquad \text{(A.18)}\end{aligned}$$

In view of Equations (A.17) and (A.18), it can be inferred that Proposition 3.1 holds.

Proof of Theorem 3.2

Note that

$$\hat{Q}^*_{\alpha,-u}(W_{i,u}) - Q^*_{\alpha,-u}(W_{i,u}) = \sum_{1\le j\ne u\le d} \left[\hat{q}^*_{\alpha,j}(X_{i,j}) - q^*_{\alpha,j}(X_{i,j})\right].$$

We now consider the asymptotic representation of $\hat{q}^*_{\alpha,j}(X_{i,j}) - q^*_{\alpha,j}(X_{i,j})$ for $j \ne u$. From the proof below, we are mainly interested in its leading term. While, all other left terms are controlled at the rate $O\left((n^{1+\varepsilon_0}h_1)^{-\frac{1}{2}}\right)$ and kept in the remainder term. According to the proof of Theorem 2.1, $\hat{q}^*_{\alpha,j}(x_j) - q^*_{\alpha,j}(x_j)$ includes three leading terms $\frac{\kappa_p q^{(p)}_{\alpha,u}(x_j)}{p!}h_1^p$, $\frac{1}{n}\sum_{k=1}^n(\zeta_k(x_j) - I\!\!E\zeta_k(x_j))$ and $\frac{n-1}{n^2h_1h^d}\sum_{j=1}^n \phi_j(x_u)$, see (A.2), (A.5) and (A.15). In fact, all other left terms can be controlled at the rate $O\left((n^{1+\varepsilon_0}h_1)^{-\frac{1}{2}}\right)$ with probability one by slight change of the proof of Theorem 2.1. For example, when dealing with the appeared degenerated U-statistics, we could choose a suitable large k for the purpose of using Borel-Cantelli lemma. By following the same line as that of Theorem 2.1 and applying Lemma 2, we could obtain the same two leading terms of $\hat{q}^*_{\alpha,j}(X_{i,j})$ as that of $\hat{q}^*_{\alpha,j}(x_j)$ with x_j replaced by $X_{i,j}$, which are $\frac{q^{(p)}_{\alpha,u}(X_{i,j})}{p!}\kappa_2 h_1^p$, $\frac{1}{n}\sum_{1\le k\ne i\le n}\psi_k(X_{i,j})$ and $\frac{n-1}{n^2h_1h^d}\sum_{j=1}^n \phi_j(X_{i,u})$, respectively, where the function $\psi_k(x_j) = \zeta_k(x_j) - I\!\!E\zeta_k(x_j)$ (see (A.1) for the notation $\zeta_k(x_j)$). Let $\eta_{i,k} = \sum_{1\le j\ne u\le d}\psi_k(X_{i,j})$ and

$$\xi_i = \frac{\kappa_p h_1^p}{p!}\sum_{1\le j\ne u\le d} q^{(p)}_{\alpha,u}(X_{i,j}) + \frac{1}{n}\sum_{1\le k\ne i\le d}\eta_{i,k} + \frac{n-1}{n^2h_1h^d}\sum_{j=1}^n \phi_j(X_{i,u}) = \xi_{i1} + \xi_{i2} + \xi_{i3}. \quad \text{(A.19)}$$

By virtue of the argument above, we know that, with probability one,

$$\hat{Q}^*_{\alpha,-u}(W_{i,u}) - Q^*_{\alpha,-u}(W_{i,u}) = \xi_i + O\left((n^{1+\varepsilon_0}h_1)^{-\frac{1}{2}}\right). \quad \text{(A.20)}$$

For $\beta \in \mathbb{R}^p$, let $f_1(\beta)$ be equal to (3.3) with β_{x_u} replaced by β,

$$f_2(\beta) = \frac{1}{nh_e}\sum_{i=1}^n K_{u,i}\rho_\alpha\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - \beta^T V_{u,i}\right),$$

$$f_3(\beta) = \frac{1}{nh_e}\sum_{i=1}^n K_{u,i}\rho_\alpha\left\{Y_i + (d-2)c_\alpha - Q^*_{\alpha,-u}(W_{i,u}) - \xi_i - \beta^T V_{u,i}\right\}$$

and $\hat{\beta}_3 = \arg\min_\beta f_3(\beta)$. By slight change of the proof of Proposition 2.2, we know that $\hat{c}_\alpha - c_\alpha = o_{I\!\!P}\left(\frac{n^{-\frac{\varepsilon_0}{4}}}{\sqrt{nh_e}}\right)$. Hereby, from this and (A.20), it can be inferred directly that

$$\sup_\beta |f_1(\beta) - f_3(\beta)| = O_{I\!\!P}\left(\frac{n^{-\frac{\varepsilon_0}{4}}}{\sqrt{nh_e}}\right).$$

Combination of this and the fact that both $f_1(\beta)$ and $f_3(\beta)$ are linear functions leads to

$$\sqrt{nh_e}\left(\hat{\beta}_{x_u} - \hat{\beta}_3\right) = O_{I\!\!P}\left(n^{-\frac{\varepsilon_0}{4}}\right). \quad \text{(A.21)}$$

Note that $\xi_{i,1} = O(h_1^p) = o(r_n)$ holds with probability one and uniformly for $1 \le i \le n$. According to SLLN, it can be inferred that $\xi_{i,2} = O\left(n^{\varepsilon_0/4}(nh_1)^{-\frac{1}{2}}\right) = O(r_n)$ holds uniformly for $1 \le i \le n$, and so does for ξ_{i3}. Thus, $\xi_i = O(r_n)$ holds with probability one and uniformly for $1 \le i \le n$. From this and Proposition 3.1, we obtain that

$$\hat{\beta}_3 - \hat{\beta}_{x_u}^{oracle} = \frac{B_n^{-1}\mathbb{E}\left(K_{2,i}V_{u,i}g_u\left(0|X_{i,u}\right)\right)}{nh_e}\sum_{i=1}^{n}\xi_i + O_{\mathbb{P}}\left((n^{1+2\varepsilon_0}h_e)^{-\frac{1}{2}}\right).$$

Substituting (A.19) into the right hand side of the relationship above, we denote the derived three terms by I_1, I_2 and I_3, respectively. Clearly,

$$\sqrt{nh_e}I_1 = O_{\mathbb{P}}\left(\sqrt{nh_e}h_1^p\right) = O_{\mathbb{P}}\left(n^{-\frac{p\varepsilon_0}{2}}\right). \tag{A.22}$$

As for I_2, it is of the same order as that of

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{1\le j\ne i\le n}\eta_{i,j} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{1\le j\ne i\le n}(\eta_{i,j} - \mathbb{E}_i\eta_{i,j}) + \frac{n-1}{n^2}\sum_{i=1}^{n}\mathbb{E}_i\eta_{i,j} = I_{21} + I_{22}.$$

Since I_{21} is a degenerated U-statistic, by following the same line as that of (A.13), it follows that $\sqrt{nh_e}I_{21} = o_{\mathbb{P}}(1)$. By the standard WLLN, it can be obtained that $\sqrt{nh_e}I_{22} = O_{\mathbb{P}}\left(\sqrt{h_e}\right)$. Therefore, we have that $\sqrt{nh_e}I_2 = o_{\mathbb{P}}(1)$. Analogously, it can also obtained that $\sqrt{nh_e}I_3 = o_{\mathbb{P}}(1)$. From these two relationships and (A.22), we have that $\sqrt{nh_e}(\hat{\beta}_3 - \hat{\beta}_{x_u}^{oracle}) = o_{\mathbb{P}}(1)$. Whence, by following this and (A.21), this theorem holds.

Appendix B

Lemmas for Theorems 2.1 and Proposition 2.2

Lemmas 1, 2 and 3 are on absolutely regular processes. For the proof of Lemma 1, we refer to Yoshihara (1978). The proofs of Lemmas 2, 3 and 4 are given by Cheng and De Gooijer (2008).

Lemma 1. *Let F_1 and F_2 be the two distribution functions of the random vectors $\xi_{i_1}, \ldots, \xi_{i_j}$ and $\xi_{i_{j+1}}, \ldots, \xi_{i_k}$. $h(x_1, \ldots, x_k)$ is a Borel measurable function with the bound $M > 0$. Then*

$$\left|\mathbb{E}h(\xi_{i_1}, \ldots, \xi_{i_k}) - \int\cdots\int h\left(x_{i_1}, \ldots, x_{i_j}, x_{i_{j+1}}, \ldots, x_{i_k}\right) dF^{(1)}\left(x_{i_1}, \ldots, x_{i_j}\right) dF^{(2)}\left(x_{i_{j+1}}, \ldots, x_{i_k}\right)\right| \le 2M\pi(i_{j+1} - i_j).$$

Lemma 2. *Suppose that $g(\cdot,\cdot)$ is a Borel measurable function with the bound $M > 0$. Let $g_2(\cdot) = \mathbb{E}g(\xi,\cdot)$, $\sigma(\cdot) = \mathbb{D}(\sum_{i=1}^{q} g(\xi_i,\cdot))$ and $U_{ij_0} = g(\xi_i, \xi_{j_0}) - g_2(\xi_{j_0})$, where $1 \le j_0 \le n$ is* *fixed. Then, for any $x > 0$, $r_1 > 1$ and positive integer $q \leq \frac{n}{4}$ it holds that*

$$I\!P\left\{\left|\sum_{1\leq i\leq n, i\neq j_0} U_{ij_0}\right| \geq x\right\} \leq 2I\!E\exp\left\{\frac{-\left(\frac{x}{4}\right)^2}{\frac{n}{2q}\sigma(\xi_0)+\frac{2}{3}qM\frac{x}{4}}\right\}+\frac{n}{q}\beta(q)+\frac{2^{r_1}q^{r_1-1}}{x^{r_1}}\sum_{|i-j_0|<2q} I\!E|U_{ij_0}|^{r_1}.$$

Lemma 3. *Let $U_n = \sum_{1\leq i<j\leq n} h_n(\xi_i,\xi_j)$ be a degenerated U-statistic with the symmetric kernel $h_n(\cdot,\cdot)$, i.e., for any $t \in \mathbb{R}$, $I\!Eh_n(\xi_i,\cdot) = 0$. Then for $k \in \mathbb{N}$, there exists a universal constant $C > 0$ such that*

$$I\!EU_n^k \leq Cn^k\left(1+\sum_{i=1}^{n-1} i^k\beta_i^{1-\frac{1}{s}}\right)M_{sk}^k,$$

where $s > 1$ and

$$M_{sk} = \sup_{(i_1,i_2),I\!P}\left(\int |h_n(\xi_{i_1},\xi_{i_2})|^{sk}dI\!P\right)^{\frac{1}{sk}}$$

with $I\!P$ being either the probability measure $I\!P_{(\xi_{i_1},\xi_{i_2})}$ or $I\!P_{\xi_{i_1}}\otimes I\!P_{\xi_{i_2}}$.

Lemma 4. *Define an operator $I\!E_i$ as $I\!E_i g(\xi_j,\xi_i) = g_2(\xi_i)$ for any $i \neq j$. Then, under the conditions C1 to C6, the following Bahadur representation for conditional quantiles holds almost surely and uniformly for $1 \leq i \leq n$ that*

$$\hat{Q}_\alpha(X_i) - Q_\alpha(X_i) = \frac{e_1^T B_{i,n}^{-1}}{nh^d}\sum_{1\leq j\neq i\leq n} L_{ji}V_{ji}\left(\alpha - I\!I\left(Y_j \leq V_{ij}^T\beta_{X_i}\right)\right) + O\left(\left(\frac{\log n}{nh^d}\right)^{\frac{3}{4}}\right), \tag{B.1}$$

where $L_{ji} = L\left(\frac{X_i-X_j}{h}\right)$, $V_{ij} = V_1\left(\frac{X_i-X_j}{h}\right)$ and $B_{i,n} = \frac{1}{h^d}I\!E_i\left(L_{ji}V_{ji}V_{ji}^T g(0|X_j)\right)$.

Lemmas for Proposition 3.1

We show a sequence of lemmas with proofs that help to show the result in Proposition 3.1. Without loss of generality, in the proofs of Lemma 5 to Lemma 12, α is taken to be $\frac{1}{2}$, condensing presentation of the proofs.

For $a \in \mathbb{R}^p$, denote by $\Lambda(a,\mathbf{t}_n) = \frac{1}{nh_e}\sum_{i=1}^n \Lambda_i$ and

$$\begin{aligned}\Lambda_i = \Lambda_i(a,t_{i,n}) = K_{u,i}&\left(\left|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - a^T V_{u,i} - t_{i,n}\right|\right.\\ &\left.-\left|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - \beta_{x_u}^T V_{u,i}\right|\right).\end{aligned}$$

Lemma 5. *There exists a constant $M_1 > 0$ such that, with probability one,*

$$|\Lambda(a,\mathbf{t}_n) - I\!E\Lambda(a,\mathbf{t}_n)| \leq \frac{M_1 r_n}{\sqrt{\log n}} \tag{B.2}$$

holds uniformly for $\|a-\beta_{x_u}\| = Cr_n^{\frac{1}{2}}$ and $|t_{i,n}| \leq Cr_n$, $i = 1,2,\ldots,n$.

Proof. Divide the interval $|t| \leq Cr_n$ into a sequence of subintervals with equidistance $l_n = r_n/\sqrt{\log n}$. Let $\{v_{i,n}\}$ be the set of all the grid points of the number $N_1 = O\left(\sqrt{\log n}\right)$. For any $|s_{i,n}| \leq Cr_n$, $i = 1, 2, \ldots, n$, let $t_{i,n}$ be the left abscissa of the corresponding subinterval. Denote by $\mathbf{s}_n = (s_{1,n}, \ldots, s_{n,n})$ and $\mathbf{t}_n = (t_{1,n}, \ldots, t_{n,n})$. Next, we divide the sphere $\|a - \beta_{x_u}\| = Cr_n^{\frac{1}{2}}$ into a sequence of smaller areas with the length of the sides l_n. The number of such kinds of smaller areas is of order $N_2 = \left(r_n^{-1}\log n\right)^{\frac{p-1}{2}}$. For any b in the mentioned sphere, let a be the nearest grid point to b. Then, we have that

$$|\Lambda(b, \mathbf{s}_n) - \Lambda(a, \mathbf{t}_n) - \mathbb{E}\left[\Lambda(b, \mathbf{s}_n) - \Lambda(a, \mathbf{t}_n)\right]| \leq \frac{l_n}{nh_e}\sum_{i=1}^{n}\left(K_{u,i} + \mathbb{E}K_{u,i}\right). \tag{B.3}$$

Below, we will use the fact that $\mathbb{E}\left|\Lambda_i(a, t_{j,n})\right|^s \leq Ch_e r_n^{\frac{s}{2}}$ for any $s > 0$ is used. By using Theorem 3 of Yoshihara (1978), Corollary 2.1 of Hall and Heyde (1980) and Rosenthal inequality, and taking $m = (nh_e)^{\frac{1}{4}}$ and constant r sufficiently large, we know that

$$\begin{aligned}
& N_2 \mathbb{P}\left\{\cup_{\mathbf{t}_n}\left\{|\Lambda(a, \mathbf{t}_n) - \mathbb{E}\Lambda(a, \mathbf{t}_n)| \geq \frac{M_1 r_n}{\sqrt{\log n}}\right\}\right\} \\
\leq\; & N_2 \sum_{j=1}^{N_1} \mathbb{P}\left\{\max_{1\leq k\leq n}\left|\sum_{i=1}^{k}\left(\Lambda_i(a, v_{j,n}) - \mathbb{E}\Lambda_i(a, v_{j,n})\right)\right| \geq \frac{nh_2 r_n}{N_1\sqrt{\log n}}\right\} \\
\leq\; & N_2 \sum_{j=1}^{C\log n}\left\{C(r)\sum_{l=1}^{m}\frac{n\mathbb{E}\left|\Lambda_i(a, v_{j,n})\right|^r + \left[\sum_{i=1}^{n} E(\Lambda_i^2(a, v_{j,n}))\right]^{\frac{r}{2}}}{\left(m^{-1}r_n^{-1}\right)^r} + n\pi(m)\right\} \\
\leq\; & CN_1N_2\left\{m^{r+1}\left(\frac{(\log n)^2}{nh_e r_n}\right)^{\frac{r}{2}} + n\pi(m)\right\} \leq \frac{C}{n(\log n)^2},
\end{aligned} \tag{B.4}$$

where $C(r)$ is a constant only related to r, and the condition C4 on θ is used in the last inequality. Next, we will verify (B.6). Let $L_1 = \sum_{i=1}^{u} K_{u,i}$. According to Lemma 1 and the two facts $\theta \geq p + 9$ and $m_n = O(h_2^{-1})$, it can be inferred that

$$\begin{aligned}
\mathbb{D}(L_1) \;\leq\; & u\mathbb{E}K_{u,1}^2 + u\sum_{i=1}^{u-1}\operatorname{cov}(K_{u,1}, K_{u,i+1}) \leq uh_e + u\sum_{i=1}^{m_n}\operatorname{cov}(K_{u,1}, K_{u,i+1}) \\
& + Cu\sum_{i=m_n}^{\infty}\pi_i \leq Cuh_e + um_n h_e^2 + Cum_n^{-(\theta-1)} \leq Cuh_e.
\end{aligned} \tag{B.5}$$

For suitable large constant M_1, we have that $\frac{l_n}{nh_e}n\mathbb{E}K_{u,i} \leq M_1 r_n/\sqrt{\log n}$. Thus, by letting

$u = \frac{nh_e}{\log n}$, it can be obtained subsequently that

$$\begin{aligned} & N_2 I\!P \left\{ \frac{l_n}{nh_e} \sum_{i=1}^{n} (K_{u,i} + I\!E K_{u,i}) \geq M_1 r_n / \sqrt{\log n} \right\} \\ \leq\ & N_2 I\!P \left\{ \sum_{i=1}^{n} (K_{u,i} - I\!E K_{u,i}) \geq C M_1 n h_e \right\} \\ \leq\ & N_2 \left\{ \exp\left\{ \frac{(C M_1 n h_e)^2}{2v I\!D(L_1) + 2u C M_1 n h_e} \right\} + \frac{n}{u} \pi(u) \right\} \leq \frac{C}{n (\log n)^2}. \end{aligned} \tag{B.6}$$

By virtue of Borel-Cantelli Lemma, (B.4) and (B.6), we can get that (B.2) holds. □

Lemma 6. *It holds uniformly for $|t_{i,n}| \leq C r_n$, $i = 1, 2, \ldots, n$, with probability one that $\|\hat{\beta}_{\mathbf{t}_n} - \beta_{x_u}\| \leq M_1 r_n^{\frac{1}{2}}$.*

The proof of the lemma above is similar to Lemma 3.2 of Honda (2000).

Denote by

$$\xi_{n,i}(a, t_{i,n}) = -K_{u,i} V_{u,i} \left[2 I\!I \left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - a^T V_{u,i} - t_{i,n} \geq 0 \right) - 1 \right].$$

Lemma 7. *There exists a constant $M_2 > 0$ such that, with probability one,*

$$\left| \sum_{i=1}^{n} \xi_{n,i}(a, t_{i,n}) - 2nh_e g_u(0|x_u) f_u(x_u) \int K_e(s) V(s) V^T(s) ds (a - \beta_{x_u}) \right| \leq M_2 \sqrt{n^{1+\varepsilon_0} h_e}$$

holds uniformly for $\|a - \beta_{x_u}\| \leq M_1 r_n^{\frac{1}{2}}$ and $|t_{i,n}| \leq C r_n$ with $1 \leq i \leq n$.

Proof. First note that, when $|x_{i,u} - x_u| \leq h_e$, it then holds that

$$-q_{\alpha,u}(x_{i,u}) + a^T V_{u,i} = -(q_{\alpha,u}(x_{i,u}) - \beta_{x_u}^T V_{u,i}) + (a - \beta_{x_u})^T V_{u,i} = O\left(h_e^p\right) + (a - \beta_{x_u})^T V_{u,i}.$$

Thus, by using variable substitution, the mean value of the integration and Taylor expansion, we can obtain that

$$\begin{aligned} \sum_{i=1}^{n} I\!E \xi_{n,i}(a, 0) &= 2 \sum_{i=1}^{n} \int K_{u,i} V_{u,i} \int_0^{-q_{\alpha,u}(x_{i,u}) + a^T V_{u,i}} g_u(s|x_{i,u}) ds f_u(x_{i,u}) dx_{i,u} \\ &= 2nh_e g_u(0|x_u) f_u(x_u) \int K_e(s) V(s) V^T(s) ds (a - \beta_{x_u}) + O\left(nh_e |a - \beta_{x_u}|^2 + nh_e^{2p+1} \right). \end{aligned}$$

Divide the ball $\|a - \beta_{x_u}\| \leq M_1 r_n^{\frac{1}{2}}$ into a sequence of smaller cubics with the length of the side $l_n = r_n$. Thus, the total number of different smaller cubics is of order $r_n^{-\frac{p}{2}}$. Denote by

$\eta_i = K_{u,i} I\!I\left(\left|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - a^T V_{u,i}\right| \le C r_n\right)$. For any b in the mentioned ball, a is the corresponding nearest grid point. Clearly, we have that

$$\left|\sum_{i=1}^{n}(\xi_{n,i}(b,t_{i,n}) - I\!E\xi_{n,i}(b,t_{i,n})) - \sum_{i=1}^{n}(\xi_{n,i}(a,0) - I\!E\xi_{n,i}(a,0))\right| \le 2\sum_{i=1}^{n}(\eta_i + I\!E\eta_i).$$

It can be calculated that $nI\!E\eta_i \le C\sqrt{n^{1+\varepsilon_0}h_e}$. Thus, by the same method as that of (B.6) and taking $u = \sqrt{n^{1+\varepsilon_0}h_e}/\log n$ and $m_n = h_e r_n^2$, we obtain that

$$r_n^{-\frac{p}{2}} I\!P\left\{\sum_{i=1}^{n}\eta_i \ge M_2\sqrt{n^{1+\varepsilon_0}h_e}\right\} \le \frac{1}{n(\log n)^2}.$$

Analogously, by letting $u = \sqrt{n^{1+\varepsilon_0}h_e}/\log n$ and $m_n = Ch_e$, it holds that

$$r_n^{-\frac{p}{2}} I\!P\left\{\left|\sum_{i=1}^{n}(\xi_{n,i}(a,0) - I\!E\xi_{n,i}(a,0))\right| \ge M_2\sqrt{n^{1+\varepsilon_0}h_e}\right\} \le \frac{1}{n(\log n)^2}.$$

Then, by Borel-Cantelli Lemma, we can obtain Lemma 7. □

We now divide the ball $\|a - \beta_{x_u}\| \le M_1 r_n^{\frac{1}{2}}$ into a sequence of smaller cubics with the length of the sides $(n^{1-\varepsilon_0}h_e)^{-1}$. Denote $\mathcal{B}_n$ by the set of all the grid points, the number of which is of order $\left(n^{1-\varepsilon_0}h_e\right)^{\frac{p}{2}}$. Let $\beta^*_{x_u}$ be nearest grid point to the point $\hat\beta_{\mathbf{t}_n}$. Then, we can see that $\|\beta^*_{x_u} - \hat\beta_{\mathbf{t}_n}\| \le (n^{1-\varepsilon_0}h_e)^{-1}$.

Lemma 8. *There exists a constant $M_3 > 0$ such that, with probability one,*

$$\max_{a\in\mathcal{B}_n}\sum_{i=1}^{n}\left|K_{u,i}V_{u,i} I\!I\left(|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - a^T V_{u,i} - t_{i,n}| \le Cr_n\right)\right| \le M_3\sqrt{n^{1+\varepsilon_0}h_e}$$

holds uniformly for $|t_{i,n}| \le C_1 r_n$, $1 \le i \le n$.

Proof. Let

$$\zeta_i(a, t_{i,n}, C_1) = K_{u,i}V_{u,i} I\!I\left(|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - a^T V_{u,i} - t_{i,n}| \le C_1 r_n\right).$$

Then, there exists a constant $C_2 > 0$ such that

$$\left|\sum_{i=1}^{n}[\zeta_i(a,t_{i,n},C_1) - \zeta_i(a,0,C_1)]\right| \le \sum_{i=1}^{n}|\zeta_i(a,0,C_2)|.$$

Analogous to the proofs of (B.6) and (B.5), through letting $u = \frac{1}{\log n}\sqrt{n^{1+\varepsilon_0}h_e}$ and $m_n = h_e^{-1}r_n^2$, we know that

$$\left(n^{1-\varepsilon_0}h_e\right)^{\frac{p}{2}} I\!P\left\{\left|\sum_{i=1}^{n}\zeta_i(a,0,C_2)\right| \ge M_3\sqrt{n^{1+\varepsilon_0}h_e}\right\} \le \frac{C}{n(\log n)^2}.$$

Thus, Borel-Cantelli lemma leads to Lemma 8. □

Lemma 9. *There exists a constant $C > 0$ such that, with probability one,*

$$\left|\sum_{i=1}^{n} K_{u,i}V_{u,i}\left[2I\!I\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - \hat{\beta}_{\mathbf{t}_n}^T V_{u,i} - t_{i,n} \geq 0\right) - 1\right]\right| \leq C$$

holds uniformly for $|t_{i,n}| \leq C_1 r_n$, $1 \leq i \leq n$.

Proof. This can be obtained by the definition of $\hat{\beta}_{\mathbf{t}_n}$, see, for example, Lemma 3.3 of Honda (2000). □

Lemma 10. *There exists a constant $M_4 > 0$ such that, with probability one,*

$$\left|\sum_{i=1}^{n} K_{u,i}V_{u,i}\left[2I\!I\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - (\beta_{x_u}^*)^T V_{u,i} - t_{i,n} \geq 0\right) - 1\right]\right| \leq M_4\sqrt{n^{1+\varepsilon_0}h_e}$$

holds uniformly for $|t_{i,n}| \leq C_1 r_n$, $1 \leq i \leq n$.

Proof. The left hand side of the relationship above is bounded by $I_1 + I_2$ with

$$\begin{aligned} I_1 &= 2\left|\sum_{i=1}^{n} K_{u,i}V_{u,i}\left[I\!I\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - (\beta_{x_u}^*)^T V_{u,i} - t_{i,n} \geq 0\right)\right.\right. \\ &\quad \left.\left. - I\!I\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - \hat{\beta}_{\mathbf{t}_n}^T V_{u,i} - t_{i,n} \geq 0\right)\right]\right| \end{aligned}$$

and

$$I_2 = \left|\sum_{i=1}^{n} K_{u,i}V_{u,i}\left[2I\!I\left(Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - \hat{\beta}_{\mathbf{t}_n}^T V_{u,i} - t_{i,n} \geq 0\right) - 1\right]\right|,$$

respectively. When $\left|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - (\beta_{x_u}^*)^T V_{u,i} - t_{i,n}\right| \leq r_n$, by SLLN, it follows that

$$I_1 \leq 2\sum_{i=1}^{n} K_{u,i}V_{u,i}I\!I\left(\left|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - (\beta_{x_u}^*)^T V_{u,i} - t_{i,n}\right| \leq r_n\right) \leq C\sqrt{n^{1+\varepsilon_0}h_e}$$

holds with probability one. For any two reals a_1 and a_2, define $U(a_1) = \frac{a_1}{|a_1|}$ for $a_1 \neq 0$ and $U(a_1) = 1$ if $a_1 = 0$. Note that $2I\!I(a_1 \geq 0) - 1 = U(a_1)$ and $|U(a_1) - U(a_2)| \leq \frac{2|a_1 - a_2|}{|a_1|}$. Then, if $\left|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - (\beta_{x_u}^*)^T V_{u,i} - t_{i,n}\right| \geq r_n$, from the known result $\left|\beta_{x_u}^* - \hat{\beta}_{\mathbf{t}_n}\right| \leq (n^{1-\varepsilon_0}h_e)^{-1}$ and SLLN, we get to know that, with probability one,

$$I_1 \leq 2\sum_{i=1}^{n} \frac{K_{u,i}\left|\beta_{x_u}^* - \hat{\beta}_{\mathbf{t}_n}\right|}{\left|Y_i - c_\alpha - Q_{\alpha,-u}(W_{i,u}) - (\beta_{x_u}^*)^T V_{u,i} - t_{i,n}\right|} \leq C\sqrt{n^{1+\varepsilon_0}h_e}.$$

In view of Lemma 9, $I_2 \leq C$. Therefore, this lemma holds. □

Lemma 11. *It holds uniformly for $|t_{i,n}| \leq Cr_n$, $i = 1, \ldots, n$, with probability one that $\left\|\hat{\beta}_{\mathbf{t}_n} - \beta_{x_u}\right\| \leq Cr_n$.*

Proof. For a satisfying that $M_1 r_n^{\frac{1}{2}} \geq \|a - \beta_{x_u}\| > M_5 r_n$, it follows from Lemma 7 that

$$\begin{aligned}
& \left|K_{u,i} V_{u,i} \left[2I\!I\left(Y_i - c_\alpha - Q_{\alpha,-u}(X_{i,-u}) - a^T V_{u,i} - t_{i,n} \geq 0\right) - 1\right]\right| \\
\geq \; & -M_2 \sqrt{n^{1+\varepsilon_0} h_e} + \left\| 2nh_e g_u(0|x_u) f(x_u) \int K_e(s) V(s) V^T(s) ds (a - \beta_\tau) \right\| \\
\geq \; & (CM_5 - M_2)\sqrt{n^{1+\varepsilon_0} h_e}.
\end{aligned}$$

We could choose sufficiently large $M_5 > 0$ such that $CM_5 - M_2 > M_4$. Hence, by Lemma 10, we have that $\|\beta^*_{x_u} - \beta_{x_u}\| \leq M_5 r_n$. Therefore, in view that $\|\beta^*_{x_u} - \hat{\beta}_{x_u}\| \leq (n^{1-\varepsilon_0} h_e)^{-1}$, it holds with probability one that $\|\hat{\beta}_{\mathbf{t}_n} - \beta_{x_u}\| \leq (M_5 + 1) r_n$. □

Lemma 12. *With probability one, it holds uniformly for $\|a - \beta_{x_u}\| \leq C r_n$ and $|t_{i,n}| \leq C r_n$, $i = 1, 2, \ldots, n$, that*

$$\Delta_n(a, \mathbf{t}_n) - I\!E \Delta_n(a, \mathbf{t}_n) = O\left((nh_e)^{\frac{1}{2}} n^{-\varepsilon_0}\right). \quad \text{(B.7)}$$

Proof. Divide the ball $\|a - \beta_u\| \leq C r_n$ into a sequence of cubics with the length of the sides $l_n = \frac{n^{-\varepsilon_0}}{\sqrt{nh_e}}$. As before, the total number of different cubics N_2 is of order $n^{\frac{3}{2} p \varepsilon_0}$. For any b in the mentioned ball, let a be the center of the cubic in which b locates. Then, we divide the interval $|t| \leq C r_n$ analogously to Lemma 5 but with equidistance l_n. The same notations related to this division is adopted here. Then, $N_1 = C n^{\frac{3}{2}\varepsilon_0}$. Clearly, we have that

$$|(\Delta_n(b, \mathbf{s}_n) - \Delta_n(a, \mathbf{t}_n)) - I\!E(\Delta_n(b, \mathbf{s}_n) - \Delta_n(a, \mathbf{t}_n))| \leq \sum_{i=1}^n (\xi_i + I\!E \xi_i),$$

where

$$\xi_i = K_{u,i} A_{u,i} I\!I\left(\left|\mathcal{E}_{\alpha,i} - \left(a^T V_{u,i} + t_{i,n} - q_{\alpha,u}(X_{i,u})\right)\right| \leq l_n\right).$$

Note that, for any $s > 0$, it holds that $I\!E|\Delta_{i,2}(a, v_{j,n})|^s \leq C h_e r_n$. Then, by using Theorem 3 of Yoshihara (1978), Corollary 2.1 of Hall and Heyde (1980) and Rosenthal's inequality, and taking $m = n^{-6\varepsilon_0}(nh_e)^{\frac{1}{4}}$, and letting constant $r > 0$ sufficiently large, we know that

$$\begin{aligned}
& N_2 I\!P\left\{\cup_{\mathbf{t}_n}\left\{|\Delta_n(a, \mathbf{t}_n) - I\!E \Delta_n(a, \mathbf{t}_n)| \geq C\sqrt{nh_e} n^{-\varepsilon_0}\right\}\right\} \\
\leq \; & N_2 \sum_{j=1}^{N_1} I\!P\left\{\max_{1 \leq k \leq n}\left|\sum_{i=1}^k (\Delta_i(a, v_{j,n}) - I\!E \Delta_i(a, v_{j,n}))\right| \geq \frac{C\sqrt{nh_e} n^{-\varepsilon_0}}{N_1}\right\} \\
\leq \; & N_2 \sum_{j=1}^{N_1}\left\{C(r) \sum_{l=1}^m \frac{n I\!E|\Delta_i(a, v_{j,n})|^r + \left[\sum_{i=1}^n I\!E(\Delta_i^2(a, v_{j,n}))\right]^{\frac{r}{2}}}{\left(\frac{C\sqrt{nh_e} n^{-\varepsilon_0}}{mN_1}\right)^r} + n\pi_m\right\} \\
\leq \; & C N_1 N_2 \left\{m\left(\frac{m^2 n^{11\varepsilon_0/2}}{C^2\sqrt{nh_e}}\right)^{\frac{r}{2}} + n\pi_m\right\} \leq \frac{C}{n(\log n)^2}. \quad \text{(B.8)}
\end{aligned}$$

Note that, for any $s > 0$, it holds that $I\!\!E|\xi_i|^s \leq C h_e l_n$. On the other hand, by the same strategy and taking $m = n^{-3\varepsilon_0}(nh_e)^{\frac{1}{4}}$, it can be derived subsequently that

$$
\begin{aligned}
& N_2 I\!\!P\left\{\cup_{\mathbf{t}}\left\{\left|\sum_{i=1}^{n}(\xi_i + I\!\!E\xi_i)\right| \geq C\sqrt{nh_e}n^{-\varepsilon_0}\right\}\right\} \\
\leq\ & N_2 I\!\!P\left\{\cup_{\mathbf{t}}\left\{\left|\sum_{i=1}^{n}(\xi_i - I\!\!E\xi_i)\right| \geq \frac{C}{2}\sqrt{nh_e}n^{-\varepsilon_0}\right\}\right\} \\
\leq\ & N_2 \sum_{j=1}^{N_1}\left\{C(r)\sum_{l=1}^{m}\frac{nI\!\!E|\xi_i|^r + \left[\sum_{i=1}^{n} I\!\!E(\xi_i^2)\right]^{\frac{r}{2}}}{\left(\frac{C\sqrt{nh_e}n^{-\varepsilon_0}}{mN_1}\right)^r} + n\pi_m\right\} \\
\leq\ & CN_1N_2\left\{m\left(\frac{m^2 n^{4\varepsilon_0}}{C\sqrt{nh_e}}\right)^{\frac{r}{2}} + n\pi_m\right\} \leq \frac{C}{n(\log n)^2}.
\end{aligned}
\tag{B.9}
$$

Then, from Borel-Cantelli Lemma, (B.8) and (B.9), we know that (B.7) holds. □

Acknowledgement. Research of Yebin Cheng is supported by National Natural Science Foundation of China (Grant No. 10871001), Project 211 Phrase Ⅲ provided by SUFE, and Shanghai Leading Academic Discipline Project (No. B803). Dawit Zerom acknowledges financial support of Mihaylo College of Business and Economics Dean's research grant.

References

Arcones, M.A. (1998), "The law of large numbers for U-statistcs under absolute regular", *Elect. Communications in Probability*, **3**, 13–19.

Budge, S., Ingolfsson, A. and Zerom, D. (2008), "Empirical analysis of ambulance travel times: the case of Calgary emergency medical services", Working paper, School of Business, University of Alberta, Canada.

Cai, Z. and Ould-Saïd, E. (2003), "Local M-estimator for nonparametric time series", *Statistics & Probability Letters*, **65(4)**, 433–449.

Cai, Z. and Xu, X. (2008), "Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models", *Journal of the American Statistical Association*, **103**, 1595–1608.

Chaudhuri, P. (1991), "Nonparametric estimates of regression quantiles and their local Bahadur representation", *Annals of Statistics*, **19(2)**, 760–777.

Chaudhuri, P. (1997), "On average derivative quantile regression", *Annals of Statistics*, **25(2)**, 715–744.

Cheng, Y. and De Gooijer, J.G. (2008), "Estimation of generalized additive conditional quantiles", Working paper, Department of Quantitative Economics, University of Amsterdam, The Netherlands.

Davydov, Y.A. (1968), "Convergence of distributions generalized by stationary stochastic processes", *Probability Theory and Applied Statistics*, **13**, 691–696.

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

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

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

Hall, P. and Heyde, C.C. (1980), *Martingale Limit Theory and its Application*, Academic Press: New York.

Honda, T. (2000), "Nonparametric estimation of a conditional quantile for -mixing processes", *Annals Institute of Statistical Mathematics*, **52**, 459–470.

Honda, T. (2004), "Quantile regression in varying coefficient models", *Journal of Statistical Planning and Inference*, **121**, 113–125.

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

Jones, M.C., Davies, S.J. and Park, B.U. (1994), "Versions of kernel-type regression estimators", *Journal of the American Statistical Association*, **89**, 825–832.

Kim, M-O. (2007), "Quantile regression with varying coefficients", *Annals of Statistics*, **35(1)**, 92–108.

Kim, W., Linton, O.B. and Hengartner, N.W. (1999), "A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals", *Journal of Computational and Graphical Statistics*, **8(2)**, 278–297.

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

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

Lee, S. (2003), "Efficient semiparametric estimation of a partially linear quantile regression model", *Econometric Theory*, **19**, 1–31.

Manzan, S. and Zerom, D. (2005), "Kernel estimation of a partially linear additive model", *Statistics & Probability Letters*, **72**, 313–322.

Rigby, R.A. and Stasinopoulos, D.M. (2005), "Generalized additive models for location, scale and shape", *Journal of the Royal Statistical Society: Series C*, **54**, 507–554.

Yoshihara, K. (1978), "Probability inequalities for sums of absolutely regular processes and their applications", *Zeitschrift für Wahrscheinlichkeit und Verwante Gebiete*, **43(4)**, 319–329.

Yu, K. and Lu, Z. (2004), "Local linear additive quantile regression" , *Scandinavian Journal of Statistics*, **31(3)**, 333–346.

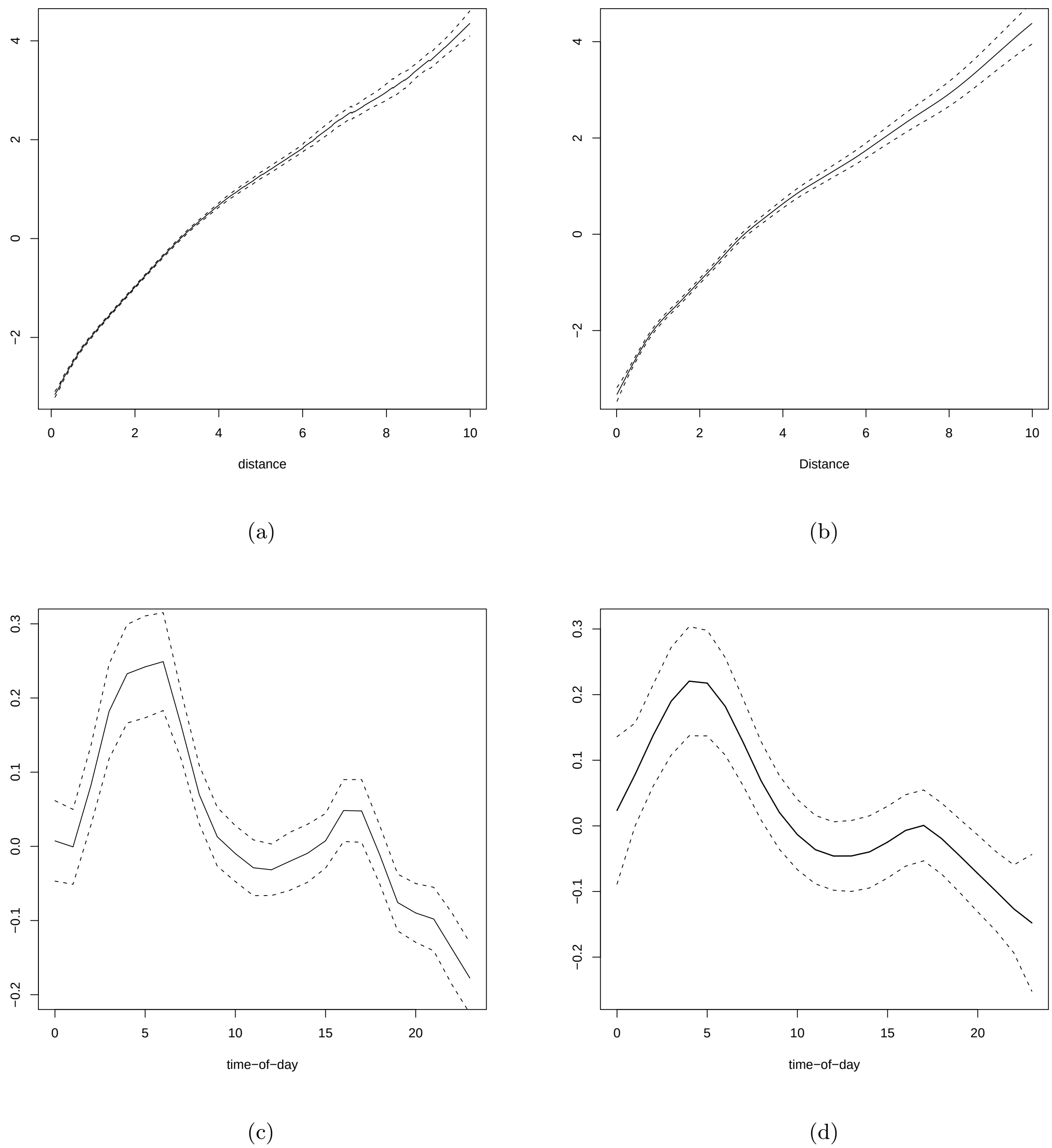

Figure 1: Panel (a) and panel (b) plot the median function with respect to distance from the non-parametric (our approach) and semi-parametric approaches, respectively. Panel (c) and panel (d) give the median function with respect to time-of-day from the non-parametric and semi-parametric approaches, respectively. For better resolution, we truncate those distances that exceed 10 kms. There are very few calls that entail distances larger than 10 kms.

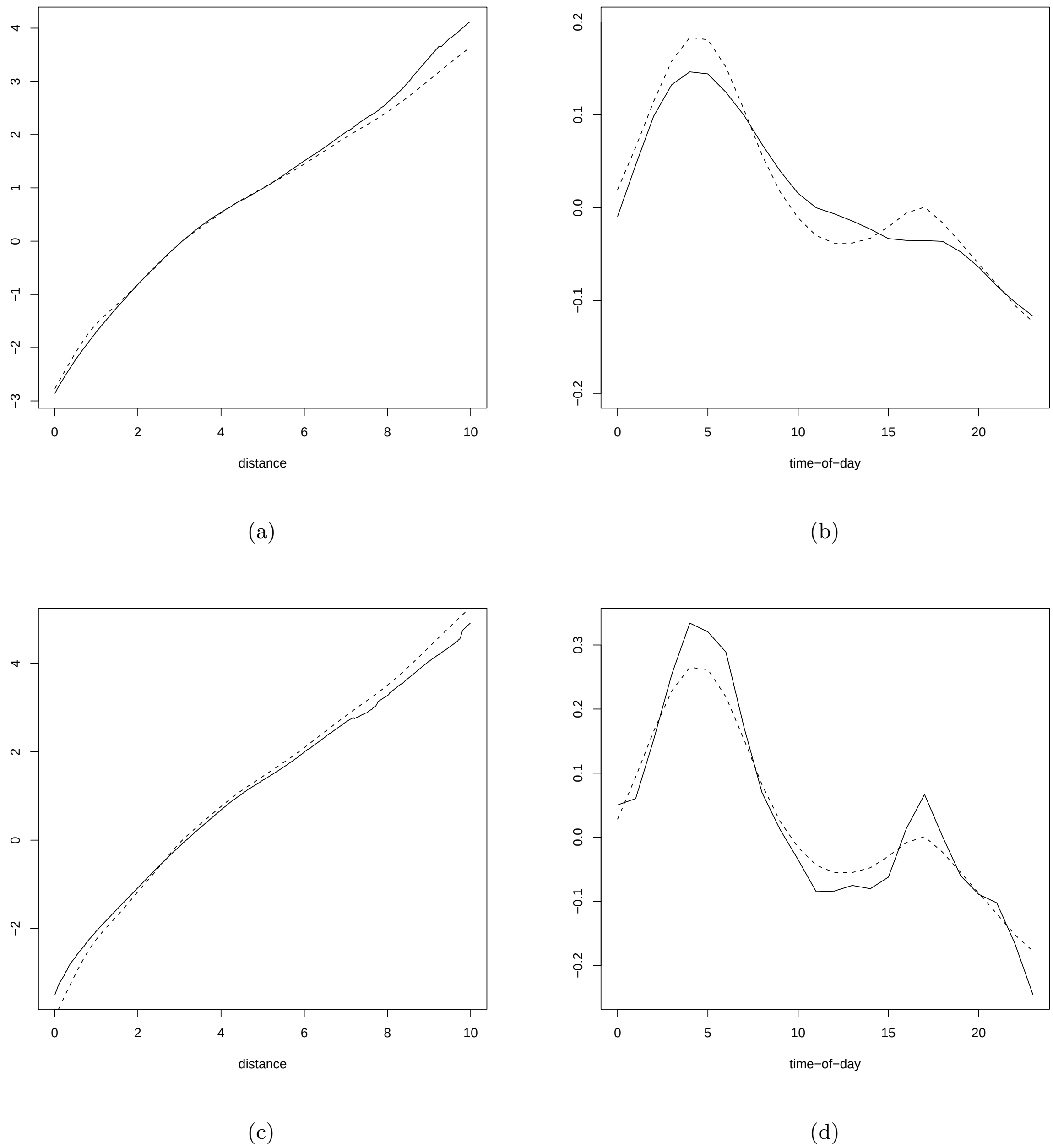

Figure 2: Panel (a) and panel (b) give the quantile function estimates for α=0.25 and for distance and time-of-day, respectively. Panel (c) and panel (d) plot the quantile function estimates for α=0.75 and for distance and time-of-day, respectively. In all panels, solid lines correspond to non-parametric estimates while dashed lines correspond to semi-parametric estimates.