Scandinavian Journal of Statistics · Theory and Applications

Scandinavian Journal of Statistics, Vol. 38: 46–62, 2011
doi: 10.1111/j.1467-9469.2010.00706.x

Efficient Estimation of an Additive Quantile Regression Model

YEBIN CHENG

Department of Statistics, Shanghai University of Finance and Economics

JAN G. DE GOOIJER

Department of Quantitative Economics, University of Amsterdam

DAWIT ZEROM

Mihaylo College of Business and Economics, California State University-Fullerton

ABSTRACT. In this paper, two non-parametric 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 more viable alternative to existing kernel-based approaches. The second estimator involves sequential fitting by univariate local polynomial quantile regressions for each additive component with the other additive components replaced by the corresponding estimates from the first estimator. The purpose of the extra local averaging is to reduce the variance of the first estimator. We show that 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.

Key words: additive models, asymptotic properties, dependent data, internalized kernel smoother, local polynomial, oracle efficiency

1. Introduction

Suppose Y denotes a response variable that depends on the vector of stochastic covariates $X=(X_1,\dots,X_d)^{\mathrm{T}}$, $d\geq 2$, 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,

$$Y=Q_\alpha(X)+\mathcal{E}_\alpha, \tag{1}$$

where $Q_\alpha(\cdot)$ is an unknown real-valued function and $\mathcal{E}_\alpha$ is an unobserved random variable that satisfies $\mathbb{P}(\mathcal{E}_\alpha\leq 0\,|\,X=x)=\alpha$ for all x where $0<\alpha<1$ is the quantile of interest. Thus, $Q_\alpha(x)$ denotes the conditional quantile of Y given $X=x$. There is a large body of literature on the estimation of $Q_\alpha(x)$ and its asymptotic properties (see e.g. Chaudhuri, 1991; Fan *et al.*, 1994). But it is well-known that for high-dimensional covariates (moderate to large value of d) non-parametric 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 example, Honda (2004), Kim (2007) and Cai & Xu (2008) consider quantile regression with varying coefficients and 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,

$$Q_\alpha(x)=c_\alpha+q_{\alpha,1}(x_1)+\cdots+q_{\alpha,d}(x_d), \tag{2}$$

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

Given observed data on X and Y, our interest is to efficiently estimate each additive component $q_{\alpha,u}(x_u)$ in (2). This non-parametric estimation problem is first considered by Fan & Gijbels (1996, pp. 296–297) where they suggest a back-fitting procedure (see also Yu & Lu, 2004). 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 & Koo (2000) introduce an easily implementable direct spline method that does not require iterations. But they do not provide asymptotic convergence results. De Gooijer & Zerom (2003) propose a simple direct kernel estimator. Horowitz & Lee (2005) suggest a hybrid stepwise approach where they use a series method in the first step and kernel smoothing in the second step. Both De Gooijer & Zerom (2003) and Horowitz & Lee (2005) provide detailed asymptotic theory, and show that their respective estimators achieve a univariate non-parametric 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 *et al.* (1999) and Manzan & Zerom (2005) to the context of conditional quantiles. We show that the proposed estimator is asymptotically normal and converges at the univariate non-parametric optimal rate. This estimator is also computationally more attractive than the average quantile estimator of De Gooijer & Zerom (2003) as it reduces the computational requirement of the latter by the order of the sample size (n), i.e. $O(n)$. 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 $(X_1,\dots,X_d)$. This is due to the need for smoothing at points that may not lie in the support of the covariate space. In contrast, our estimator is not affected by this problem.

Although the 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 additional local averaging. The local averaging involves sequential fitting by univariate local polynomial α-quantile regressions for each additive components in (2), with the other additive components replaced by the corresponding estimates from the first estimator. The second 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 the first estimator. Thus, efficiency is achieved without compromising on computational simplicity.

The asymptotic properties of our two kernel estimators are derived for dependent data. On the other hand, Horowitz & 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 $(X_1,Y_1),\dots,(X_n,Y_n)$ 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 $(Z_a,\dots,Z_b)$ with $Z_i=(X_i,Y_i)$. A process is called absolutely regular, if, as $m\to\infty$,

$$\pi(m) = \sup_{s \in \mathbb{N}} \mathbb{E}\left\{ \sup_{\mathcal{H} \in \mathcal{M}_{s+m}^{\infty}} [\mathbb{P}(\mathcal{H} \mid \mathcal{M}_{-\infty}^{s}) - \mathbb{P}(\mathcal{H})] \right\} \to 0. \tag{3}$$

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 modelling to ambulance travel times using administrative data for the city of Calgary. Section 6 provides concluding comments. Technical arguments and proofs are provided in the Appendix.

2. A modified average quantile estimator

We introduce our first kernel estimator for the additive component function $q_{\alpha,u}(x_u)$ $(u = 1, \ldots, d)$ where for identification purposes we assume that $\mathbb{E}\{q_{\alpha,u}(X_u)\} = 0$. It is also assumed that X has density denoted by $f(\cdot)$. For ease of exposition, we denote by X_u the uth element of X and W_u the set of all X variables excluding X_u, i.e. $W_u = (X_1, \ldots, X_{u-1}, X_{u+1}, \ldots, X_d)^{\mathrm{T}}$. Note that $X = (X_u, W_u)$. Let $f_u(\cdot)$ and $f_w(\cdot)$ denote the density functions of X_u and W_u, respectively. As in Manzan & Zerom (2005), 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:

$$\mathbb{E}\{\Phi(X_u, W_u) \mid X_u = x_u\} = 1 \quad \text{and} \quad \mathbb{E}\{\Phi(X_u, W_u) q_{\alpha,k}(X_k) \mid X_u = x_u\} = 0 \quad \text{for} \quad k \neq u.$$

Multiplying each side of (2) by $\Phi(\cdot,\cdot)$ and taking conditional expectations conditional on $X_u = x_u$, we obtain

$$\mathbb{E}\{\Phi(X_u, W_u)\ Q_\alpha(X) \mid X_u = x_u\} \equiv q_{\alpha,u}^{*}(x_u) = c_\alpha + q_{\alpha,u}(x_u), \quad (u = 1, \ldots, d). \tag{4}$$

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,

$$\hat{q}_{\alpha,u}(x_u) = \hat{q}_{\alpha,u}^{*}(x_u) - \hat{c}_\alpha \tag{5}$$

with the two estimators $\hat{q}_{\alpha,u}^{*}(x_u)$ and $\hat{c}_\alpha$ given in (7) and (6), respectively. Because $c_\alpha = \mathbb{E}Q_\alpha(X)$, we can estimate c_α by

$$\hat{c}_\alpha = \frac{1}{n}\sum_{i=1}^{n} \hat{Q}_\alpha(X_i), \tag{6}$$

where $\hat{Q}_\alpha(\cdot)$ is a consistent estimator of $Q_\alpha(\cdot)$ which is defined in (9). Following Jones *et al.* (1994), we compute $\hat{q}_{\alpha,u}^{*}(x_u)$ as follows,

$$\hat{q}_{\alpha,u}^{*}(x_u) = \frac{1}{nh_1}\sum_{i=1}^{n} K\left(\frac{x_u - X_{i,u}}{h_1}\right) \frac{\hat{f}_w(W_{i,u})}{\hat{f}(X_i)} \hat{Q}_\alpha(X_i), \tag{7}$$

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, i.e.

$$\frac{1}{n\hat{f}_u(x_u)h_1}\sum_{i=1}^{n}K\left(\frac{x_u - X_{i,u}}{h_1}\right)\hat{\Phi}(X_{i,u}, W_{i,u})\hat{Q}_\alpha(X_i), \tag{8}$$

(7) internalizes the normalizing density $f_u(\cdot)$ eliminating the need for explicit estimation of $f_u(x_u)$ (see the denominator in (8)). When compared to that of De Gooijer & 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 redefine (7) 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,

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

$$S = \left[\frac{1}{nh_2^d}L_2\left(\frac{X_i - X_\ell}{h_2}\right)\right]_{i,\ell},$$

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 $(\hat{q}^*_{\alpha,u}(X_{u,1}), \ldots, \hat{q}^*_{\alpha,u}(X_{u,n}))^{\mathrm{T}}$, all at once, as follows:

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

where $\odot$ and ./ denote matrix Hadamard product and division, respectively, while $e = (1, \ldots, 1)^{\mathrm{T}}$ and $\hat{Q}_\alpha = (\hat{Q}_\alpha(X_1), \ldots, \hat{Q}_\alpha(X_n))^{\mathrm{T}}$. Note also that, unlike that of De Gooijer & 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 because (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 (7) 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 neighbourhood of $x \in \mathbb{R}^d$. This will allow us to use 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 $V_1((X-x)h)$ and β_x be constructed from the elements $h^{-|\lambda|}(X-x)^\lambda$ and $h^{-|\lambda|}\partial^\lambda q(x)/x_1^{\lambda_1}\cdots x_d^{\lambda_d}$, respectively, which are arranged in natural order with respect to λ such that $|\lambda| \leq p-1$. Then, we define $\hat{Q}_\alpha(x)$ by

$$\hat{Q}_\alpha(x) = e_1^{\mathrm{T}}\hat{\beta}_x, \tag{9}$$

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

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

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 local polynomial smoothing can be implemented using the weighted linear quantile regression routine, available in the R QUANTREG library, where the weights are defined through the kernel $L(\cdot)$.

2.1. Asymptotic behaviour

Here, we derive the asymptotic behaviour of the modified average quantile estimator $\hat{q}_{\alpha,u}(x_u)$ (5) under β-mixing dependence assumption (3). 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 neighbourhood of $x_u \in \mathbb{R}$. The full-dimensional conditional quantile $Q_\alpha(x)$ is also p times continuously differentiable in the neighbourhood 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 $\bar{p} > (pd/p+1)$.

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 neighbourhood of 0.

C3 $K(\cdot)$ is a pth order kernel function that satisfies $\int K(t_1)\,\mathrm{d}t_1=1$, $\int t_1^j K(t_1)\,\mathrm{d}t_1=0$ for $j=1,\ldots,p-1$ and $\int t_1^p K(t_1)\,\mathrm{d}t_1 \neq 0$. For $i=1,2$, $L_i(\cdot)$ is a $\bar{p}$th order kernel function that satisfies $\int L_i(s)\,\mathrm{d}s=1$, $\int s^j L_i(s)\,\mathrm{d}s=0$ for $j=1,\ldots,\bar{p}-1$ and $\int s^{\bar{p}} L_i(s)\,\mathrm{d}s \neq 0$ 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 neighbour of $x_u'=x_u$.

(ii) The mixing coefficients $\pi(i)=O(i^{-\theta})$ 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 $\limsup_n nh_1^{2p+1}<\infty$.

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

(iii) For some sufficiently small constant $\epsilon>0$, $h=Cn^{-\kappa}$ with constant κ satisfying

$$\frac{1}{2p+1}<\kappa<\frac{2p+3-(2p+1)\left(\frac{1}{\theta}-2\epsilon\right)}{3d(2p+1)\left(1+\frac{2}{3\theta}-\frac{1}{3\theta^2}\right)}$$

and $h/h_1\to 0$.

(iv) It holds that $h_1^{\theta(1-\frac{2}{\delta})}h^{\frac{2}{\delta}-2}\to 0$, $nh^d\left(h_1h^d\right)^{\frac{3}{\theta}+\epsilon}\to\infty$ and $nh_1^{-1}\left(h_1h_2^{d-1}\right)^{1+\frac{3}{\theta}+\epsilon}\to\infty$ 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 $\kappa_p=\int t_1^p K(t_1)\,\mathrm{d}t_1$ and $\|K\|_2=\int K^2(t_1)\,\mathrm{d}t_1$. The following theorem summarizes the asymptotic distribution of $\hat{q}^*_{\alpha,u}(x_u)$.

Theorem 1. *When the conditions* C1–C6 *are met,*

$$\sqrt{nh_1}\left(\hat{q}^*_{\alpha,u}(x_u)-q^*_{\alpha,u}(x_u)-\frac{q^{(p)}_{\alpha,u}(x_u)\kappa_p}{p!}h_1^p\right)\to N(0,\sigma^2) \tag{10}$$

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

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

Remark 1. From theorem 1, the optimal bandwidth that minimizes the asymptotic mean squared error is given by,

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

Remark 2. Although the asymptotic variance σ^2 cannot be directly compared to the corresponding variance of the estimator of De Gooijer & 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 1. *Under the conditions of theorem* 1,

$$\hat{c}_\alpha - c_\alpha = o_{\mathbb{P}}\left(n^{-\frac{p}{2p+1}} \right). \tag{11}$$

Corollary 1. *Under the conditions of theorem* 1, *if we choose* $h_1 = h_1^{\text{opt}}$, *then it holds that*

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

in distribution.

3. Oracle efficient estimator

In section 2, we introduced a modified average quantile estimator and show that it estimates the additive components at a one-dimensional non-parametric optimal rate regardless of the size of d. However, a closer look at theorem 1 indicates 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 (1997) and Kim *et al.* (1999) to the quantile context and suggest a second estimator that involves sequential fitting of univariate locally polynomial quantile regressions for each of the additive components in (2) 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 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

$$\hat{Q}_{\alpha,-u}^*(W_u) = \hat{q}_{\alpha,1}^*(X_1) + \cdots + \hat{q}_{\alpha,u-1}^*(X_{u-1}) + \hat{q}_{\alpha,u+1}^*(X_{u+1}) + \cdots + \hat{q}_{\alpha,d}^*(X_d), \tag{12}$$

where $\hat{q}_{\alpha,j}^*(\cdot)$ $(j \neq u)$ are the additive estimates from (7). 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 (6). Also let the function $V(t)$ denote a p-dimensional vector where its jth element is given by t^{j-1}. Because $q_{\alpha,u}(x_u)$ is p times continuously differentiable in the neighbourhood of $x_u \in \mathbb{R}$, we use local polynomial smoothing to define the oracle efficient estimator as follows:

$$\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u) = e_1^{\mathrm{T}} \hat{\beta}_{x_u}, \tag{13}$$

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

$$(nh_e)^{-1} \sum_{i=1}^{n} \rho_\alpha \left(Y_i^* - \beta_{x_u}^{\mathrm{T}} V\left(\frac{x_u - X_{i,u}}{h_e} \right) \right) K_e \left(\frac{x_u - X_{i,u}}{h_e} \right), \tag{14}$$

where $K_e(\cdot)$ is a kernel function and h_e is the bandwidth.

Asymptotic behaviour

We investigate asymptotic distribution of $\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u)$ (13). 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 $g_u(t \mid x_u)$ be the conditional probability density function of $\mathcal{E}_\alpha$ given $X_u = x_u$. For any x_u in the support of X_u, it has bounded derivative with respect to the argument t in the neighbourhood of 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 is sufficiently small.

The oracle estimator. Before we provide the asymptotic distribution of $\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u)$, we first present results for an oracle estimator which we denote by $\hat{q}^{\mathrm{oracle}}_{\alpha,u}(x_u)$. Let,

$$\hat{Q}_{\alpha,-u}(W_u) = q_{\alpha,1}(X_1) + \cdots + q_{\alpha,u-1}(X_{u-1}) + q_{\alpha,u+1}(X_{u+1}) + \cdots + q_{\alpha,d}(X_d),$$

where $q_{\alpha,j}(\cdot)$ $(j \neq u)$ are the true additive components. Because $q_{\alpha,u}(x_u)$ has pth derivative and

$$\alpha = \mathbb{P}\left\{ Y_i - c_\alpha - Q_{\alpha,-u}\left(W_{i,u}\right) \leq q_{\alpha,u}(X_{i,u}) \mid X_{i,u} \right\},$$

we use local polynomial smoothing to define $\hat{q}^{\mathrm{oracle}}_{\alpha,u}(x_u)$ as,

$$\hat{q}^{\mathrm{oracle}}_{\alpha,u}(x_u) = e_1^{\mathrm{T}} \hat{\beta}^{\mathrm{oracle}}_{x_u}, \tag{15}$$

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

$$(nh_e)^{-1} \sum_{i=1}^{n} \rho_\alpha \left(Y_i - c_\alpha - Q_{\alpha,-u}\left(W_{i,u}\right) - \beta_{x_u}^{\mathrm{T}} V\left(\frac{x_u - X_{i,u}}{h_e} \right) \right) K_e \left(\frac{x_u - X_{i,u}}{h_e} \right). \tag{16}$$

Note that the oracle estimator is defined in the same way as $\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u)$ (see (13)) except that it uses true values of the other additive components and c_α.

For this oracle estimator, it can be shown that

$$\sqrt{nh_e} \left(\hat{q}^{\mathrm{oracle}}_{\alpha,u}(x_u) - q_{\alpha,u}(x_u) - h_e^p \frac{q^{(p)}_{\alpha,u}(x_u)}{p!} e_1^{\mathrm{T}} B^{-1} \kappa_e \right) \to N\left(0, \sigma_0^2\right), \tag{17}$$

where $\kappa_e = \int t_1^p V(t_1) K_e(t_1)\,\mathrm{d}t_1$, $B = \int V(t) V^{\mathrm{T}}(t) K_e(t)\,\mathrm{d}t$ and

$$\sigma_0^2 = \frac{\alpha(1-\alpha)}{g_u^2(0\,|\,x_u)f_u(x_u)} e_1^{\mathrm{T}} B^{-1} \int V(t)V^{\mathrm{T}}(t)K_e^2(t)\,\mathrm{d}t\, B^{-1} e_1. \tag{18}$$

The oracle efficient estimator. Now, we show that our estimator $\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u)$ (13) behaves analogously to the oracle estimator $\hat{q}^{\mathrm{oracle}}_{\alpha,u}(x_u)$ above. Let $r_n = n^{\frac{\varepsilon_0}{2}}/\sqrt{nh_e}$ with ε_0 being a sufficiently small positive constant. For $|t_{i,n}| \le Cr_n$, $(i=1,\dots,n)$, let $\mathbf{t}_n = (t_{1,n},\dots,t_{n,n})^{\mathrm{T}}$. 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

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

Proposition 2. *Under the conditions* C1–C9, *with probability* 1, *it holds uniformly for* $|t_{i,n}| \le Cr_n$, $i=1,2,\dots,n$, *that*

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

where $\hat{\beta}^{\mathrm{oracle}}_{x_u}$ *is as defined in* (16) *and* $B_n = \frac{1}{h_e}\mathbb{E}K_{u,i}g_u(0\,|\,X_{i,u})V_{u,i}V_{u,i}^{\mathrm{T}}$.

Theorem 2. *Under the conditions* C1–C9, *it holds that*

$$\sqrt{nh_e}(\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u) - \hat{q}^{\mathrm{oracle}}_{\alpha,u}(x_u)) = o_{\mathbb{P}}(1). \tag{19}$$

From theorem 2, we see that $\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u)$ is asymptotically normally distributed with same mean and variance as $\hat{q}^{\mathrm{oracle}}_{\alpha,u}(x_u)$. Therefore, our proposed estimator $\hat{q}^{\mathrm{e}}_{\alpha,u}(x_u)$ is oracle efficient.

4. A simulated example

In this section, we provide the finite sample performance of our oracle efficient estimator (OEE) vis-à-vis two alternative kernel estimators: the estimator of De Gooijer & Zerom (2003) (DGZ) and the back-fitting approach. We do not include the hybrid estimator of Horowitz & Lee (2005) in our comparison. But we think that the estimator of Horowitz & Lee (2005) will have a similar performance as ours at least for the independent 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,

$$Y_i = Q_\alpha(X_{i,1}, X_{i,2}) + 0.25\mathcal{E}_{\alpha,i}, \tag{20}$$

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 (20) 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

Table 1. *The average absolute deviation errors of the estimated additive components*

γ	n	$\tilde{q}_{0.5,1}(\cdot)$ OEE	DGZ	Back-fitting	$\tilde{q}_{0.5,2}(\cdot)$ 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

OEE, Oracle efficient estimator; DGZ, the estimator of De Gooijer & Zerom (2003).

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: $\hat{q}^{\mathrm{e}}_{0.5,1}(x_1)$ and $\hat{q}^{\mathrm{e}}_{0.5,2}(x_2)$ (see (13)), we need values for $\hat{Q}^*_{\alpha,-1}$ and $\hat{Q}^*_{\alpha,-2}$ (see (12)). The latter two require $\hat{q}^*_{0.5,1}(x_1)$ and $\hat{q}^*_{0.5,2}(x_2)$ (see (7)), which in turn depend on $\hat{Q}_{0.5}(x_1,x_2)$ (see (9)). 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 $h_e=h$.

We compare our median estimates $\hat{q}^{\mathrm{e}}_{0.5,1}(x_1)$ and $\hat{q}^{\mathrm{e}}_{0.5,2}(x_2)$ (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. $\mathrm{ADE}_j(k)=\mathrm{Average}\{|\tilde{q}_{0.5,k}(X_{i,k})-q_{0.5,k}(X_{i,k})|\}_i^n$ $(j=1,\dots,41; k=1,2)$ 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 $n\le 200$, 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 with existing kernel approaches, it is also robust against highly correlated covariates. For large sample sizes, the back-fitting approach tends 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 per cent of urgent urban calls in 9 minutes is a common target in North America and the National Health Service in the UK sets targets of 75 per cent in 8 minutes and 95 per cent in 14 minutes for urgent urban calls (see Budge *et al.*, 2008 and references therein). 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,

$$Y_i = \mu(X_{1,i}, X_{2,i})\,\mathrm{e}^{(\sigma \mathcal{E}_{\alpha,i})}, \quad (i = 1, \ldots, n), \tag{21}$$

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 centred t-distribution with τ degrees of freedom 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, haemorrhage and lacerations, traumatic injuries, and unknown problem. Budge *et al.* (2008) assume that the conditional median of travel time to be additive,

$$\mu(x_1, x_2) = \mu_0 + \mu_1(x_1) + \mu_2(x_2), \tag{22}$$

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 (22), the travel time distribution can be fully characterized by conditional quantiles as follows:

$$Q_\alpha(X_{1,i}, X_{2,i}) = [\mu_0 + \mu_1(X_{1,i}) + \mu_2(X_{2,i})]\,\mathrm{e}^{(\sigma Q_\alpha(\tau))}, \tag{23}$$

where $Q_\alpha(x_1, x_2)$ denotes the αth conditional quantile of Y given $(X_1, X_2) = (x_1, x_2)$ and $Q_\alpha(\tau)$ is the αth quantile of a centred t-distribution with τ degrees of freedom. Note that, under the above model set-up, the αth conditional quantile of travel time at each α is in fact additive, i.e.

$$Q_\alpha(X_{1,i}, X_{2,i}) = c_\alpha + q_{\alpha,1}(X_1) + q_{\alpha,2}(X_2), \tag{24}$$

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

Motivated by the additive conditional quantile model (24), the aim here 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 work of Budge *et al.* (2008) has a much wider scope and this section is only illustrative. 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 & 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 $\hat{q}^{*}_{\alpha,u}(\cdot)$ (7) are the same as those for estimating $\hat{q}^{\mathrm{e}}_{\alpha,u}(\cdot)$ (13). 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 & Gijbels (1996) and also adopted by Horowitz & Lee (2005). As an alternative, one may also use the data-driven bandwidth selection method by Yu & 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 Fig. 1, we plot the conditional median estimates for both our estimator and the semi-parametric approach. Those in panels (A) and (C) correspond to our median estimates corresponding to distance (X_1) and time-of-day (X_2), respectively. The confidence intervals (at the 95 per cent level) for both median estimates are based on the asymptotic variance given in (18) although we do not do any bias correction. The unknown components of the asymptotic variance are calculated using kernel estimates. On the other hand, panels (B) and (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

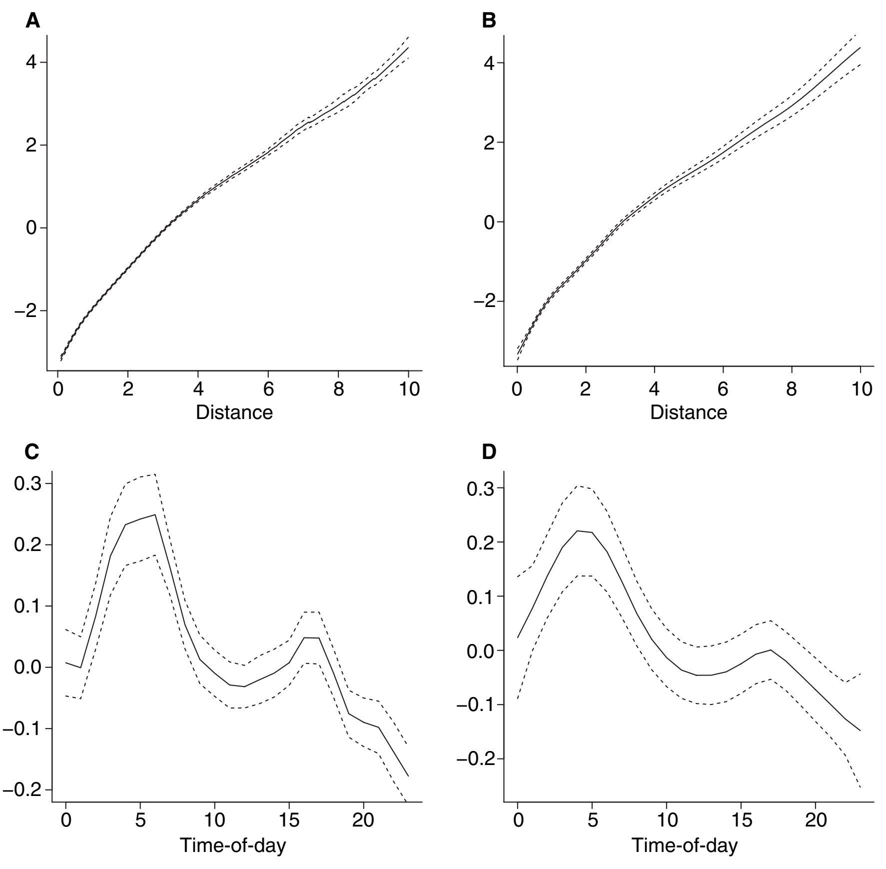

Fig. 1. (A, B) Median function with respect to distance from the non-parametric (our approach) and semi-parametric approaches, respectively. (C, D) Median function with respect to time-of-day from the non-parametric and semi-parametric approaches, respectively. For better resolution, we truncate those distances that exceed 10 km. There are very few calls that entail distances larger than 10 km.

estimates are not as smooth. In Fig. 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 occurring 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 log-travel time is leptokurtic and the t-distribution is a reasonable way to capture it.

For a complete discussion of the practical implications of the conditional quantile modelling of ambulance travel times to operational planning and related decision problems, we refer the reader to Budge *et al.* (2010).

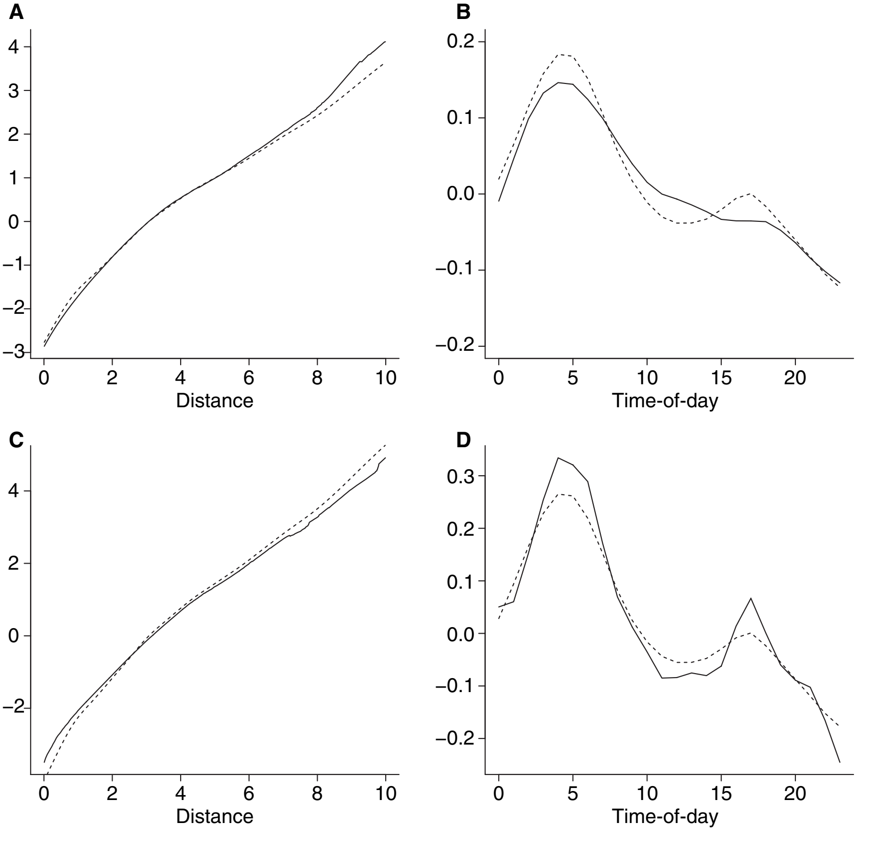

Fig. 2. (A, B) Quantile function estimates for $\alpha = 0.25$ and for distance and time-of-day, respectively. (C, D) Quantile function estimates for $\alpha = 0.75$ and for distance and time-of-day, respectively. In all panels, solid lines correspond to non-parametric estimates (our approach) while dashed lines correspond to semi-parametric estimates.

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 superior alternatives to existing kernel-based methods 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, which include independent and time series data as special cases.

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

Acknowledgements

We would like to thank the Editor, Associate Editor and the referee for their very helpful comments and suggestions. Research of Yebin Cheng is supported by National Natural Science Foundation of China (Grant No. 10871001), Project 211 Phase III provided by SUFE, and Shanghai Leading Academic Discipline Project (No. B803).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

The derivations of some intermediate results that are used in the proofs of theorem 1 and propositions 1 and 2, including a number of important lemmas.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

References

Arcones, M. A. (1998). The law of large numbers for U-statistics under absolute regular. *Electron. Commun. Probab.* **3**, 13–19.

Budge, S., Ingolfsson, A. & 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.

Budge, S., Ingolfsson, A. & Zerom, D. (2010). Empirical analysis of ambulance travel times: the case of Calgary emergency medical services. *Manag. Sci.* **56**, 715–723.

Cai, Z. & Xu, X. (2008). Nonparametric quantile estimations for dynamic smooth coefficient models, *J. Amer. Statist. Assoc.* **103**, 1595–1608.

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

De Gooijer, J. G. & Zerom, D. (2003). On additive conditional quantiles with high-dimensional covariates. *J. Amer. Statist. Assoc.* **98**, 135–146.

Doksum, K. & Koo, J. Y. (2000). On spline estimators and prediction intervals in nonparametric regression. *Comput. Statist. Data Anal.* **35**, 67–82.

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

Fan, J., Hu, T. & Truong, Y. (1994). Robust nonparametric function estimation. *Scand. J. Statist.* **21**, 433–446.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for mixing processes. *Ann. Inst. Statist. Math.* **52**, 459–470.

Honda, T. (2004). Quantile regression in varying coefficient models. *J. Statist. Plann. Inference* **121**, 113–125.

Horowitz, J. L. & Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *J. Amer. Statist. Assoc.* **100**, 1238–1249.

Jones, M. C., Davies, S. J. & Park, B. U. (1994). Versions of kernel-type regression estimators. *J. Amer. Statist. Assoc.* **89**, 825–832.

Kim, M.-O. (2007). Quantile regression with varying coefficients. *Ann. Statist.* **35**, 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. *J. Comput. Graph. Statist.* **8**, 278–297.

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

Linton, O. (1997). Efficient estimation of additive nonparametric regression models. *Biometrika* **84**, 469–473.

Manzan, S. & Zerom, D. (2005). Kernel estimation of a partially linear additive model. *Statist. Probab. Lett.* **72**, 313–322.

Rigby, R. A. & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *J. Roy. Statist. Soc. Ser. C* **54**, 507–554.

Yoshihara, K. (1978). Probability inequalities for sums of absolutely regular processes and their applications. *Z. Wahrscheinlichkeitstheorie verw Gebiete* **43**, 319–329.

Yu, K. & Lu, Z. (2004). Local linear additive quantile regression. *Scand. J. Statist.* **31**, 333–346.

Received November 2009, in final form March 2010

Dawit Zerom, Department of Information Systems and Decision Sciences, Mihaylo College of Business and Economics, California State University-Fullerton, Fullerton, CA, USA.
E-mail: dzerom@fullerton.edu

Appendix

We provide the proofs for theorems 1 and 2. In the course of the proofs, we will refer to results that can be found in Supporting Information.

Proof of theorem 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

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

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

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

$$\zeta_i = \zeta_i(x_u) = \frac{1}{h_1} K\left(\frac{x_u - X_{i,u}}{h_1}\right)\frac{f_w(W_{i,u})Q_\alpha(X_i)}{f(X_i)}. \tag{25}$$

To simplify notations, let

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

$$\phi_j \equiv \phi_j(x_u) = \mathbb{E}_j \eta_{ij}$$

and

$$\bar{\zeta}_i = h_1^{\frac{1}{2}}(\zeta_i - \mathbb{E}\zeta_i),$$

where the notations $B_{i,n}$, L_{ji} and V_{ji}, and the operator $\mathbb{E}_j$ can be referred to lemma 4 in Supporting Information. We claim that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\bar{\zeta}_i \to N(0,\sigma_2^2), \tag{26}$$

$$S_{2,n} = o_{\mathbb{P}}\left((nh_1)^{-\frac{1}{2}}\right), \tag{27}$$

$$\sqrt{nh_1}\cdot\frac{n-1}{n^2h_1h^d}\sum_{j=1}^{n}\phi_j \to N(0,\sigma_1^2), \tag{28}$$

$$S_{1,n} = \frac{n-1}{n^2h_1h^d}\sum_{j=1}^{n}\phi_j + o_{\mathbb{P}}\left(\frac{1}{\sqrt{nh_1}}\right). \tag{29}$$

Proofs of the relationships (26)–(29) are available in Supporting Information. Using (2), C1, variable substitution and Taylor expansion, we obtain

$$\begin{aligned}\mathbb{E}\zeta_i &= \int K(t_u)Q_\alpha(x_u+h_1t_u, w_u)f_w(w_u)\,\mathrm{d}t_u\mathrm{d}w_u = c_\alpha + \int K(t_u)q_{\alpha,u}(x_u+h_1t_u)\,\mathrm{d}t_u\\ &= q^*_{\alpha,u}(x_u) + \frac{q^{(p)}_{\alpha,u}(x_u)}{p!}\kappa_p h_1^p + o(h_1^p).\end{aligned} \tag{30}$$

Thus, it follows from (26) and (30) that

$$\sqrt{nh_1}\left(S_{3,n} - \frac{q^{(p)}_{\alpha,u}(x_u)\kappa_p}{p!}h_1^p\right) \to N(0,\sigma_2^2). \tag{31}$$

From the foregoing results, it can also be observed that

$$\frac{1}{nh_1^{1/2}}\sum_{i=1}^{n}\bar{\zeta}_i \quad \text{and} \quad \frac{n-1}{n^2h_1h^d}\sum_{i=1}^{n}\phi_i$$

are the two leading terms of the sum

$$S_{1,n}+S_{2,n}+S_{3,n}.$$

For any $1\le i,j\le n$, it can be noted that $\mathrm{cov}(\bar{\zeta}_i,\phi_j)=0$. Thus, the two leading terms are asymptotically uncorrelated. All other terms in the sum are asymptotically negligible and converge at the rate of $o_{\mathbb{P}}(1/\sqrt{nh_1})$.

In view of the above arguments, (27), (28), (29) and (31), the asymptotic normality result in (10) holds.

Proof of theorem 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 $j\ne u$. By following the same arguments as in the proof of theorem 1 and also applying lemma 2 in Supporting Information, we obtain three leading terms of $\hat{q}^*_{\alpha,j}(X_{i,j})$, i.e.

$$\frac{q^{(p)}_{\alpha,u}(X_{i,j})}{p!}\kappa_2 h_1^p, \quad \frac{1}{n}\sum_{1\le k\ne i\le n}\psi_k(X_{i,j}) \quad \text{and} \quad \frac{n-1}{n^2 h_1 h^d}\sum_{j=1}^{n}\phi_j(X_{i,u})$$

where $\psi_k(x_j)=\zeta_k(x_j)-\mathbb{E}\zeta_k(x_j)$. Let $\eta_{i,k}=\sum_{1\le j\ne u\le d}\psi_k(X_{i,j})$ and

$$\xi_i=\frac{\kappa_p h_1^p}{p!}\sum_{1\le j\ne u\le d} q^{(p)}_{\alpha,u}(X_{i,j})+\frac{1}{n}\sum_{1\le k\ne i\le d}\eta_{i,k}+\frac{n-1}{n^2 h_1 h^d}\sum_{j=1}^{n}\phi_j(X_{i,u})=\xi_{i1}+\xi_{i2}+\xi_{i3}. \tag{32}$$

Thus, we know, with probability 1 that

$$\hat{Q}^*_{\alpha,-u}(W_{i,u})-Q^*_{\alpha,-u}(W_{i,u})=\xi_i+O((n^{1+\varepsilon_0}h_1)^{-\frac{1}{2}}). \tag{33}$$

For $\beta\in\mathbb{R}^p$, let $f_1(\beta)$ be equal to (14) with β_{x_u} replaced by β, i.e.

$$f_1(\beta)=\frac{1}{nh_e}\sum_{i=1}^{n}K_{u,i}\rho_\alpha(Y_i+(d-2)\hat{c}_\alpha-\hat{Q}^*_{\alpha,-u}(W_{i,u})-\beta^{\mathrm{T}}V_{u,i}),$$

and

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

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

Let $\hat{\beta}_3=\arg\min_\beta f_3(\beta)$. By slight change of the proof of proposition 1 (see Supporting Information), we can obtain that $\hat{c}_\alpha-c_\alpha=o_{\mathbb{P}}(n^{-\frac{\varepsilon_0}{4}}/\sqrt{nh_e})$. Then, using (33), it can be inferred that

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

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

$$\sqrt{nh_e}(\hat{\beta}_{x_u}-\hat{\beta}_3)=O_{\mathbb{P}}(n^{-\frac{\varepsilon_0}{4}}). \tag{34}$$

Note that $\xi_{i,1}=O(h_1^p)=o(r_n)$ holds with probability 1 and uniformly for $1\le i\le n$. According to, the strong law of large numbers (SLLN) it can be inferred that $\xi_{i,2}=O(n^{\varepsilon_0/4}(nh_1)^{-\frac{1}{2}})=O(r_n)$ holds uniformly for $1\le i\le n$, and so does for ξ_{i3}. Thus, $\xi_i=O\left(r_n\right)$ holds with probability 1 and uniformly for $1\le i\le n$. Thus, using proposition 2 (see Supporting Information), we obtain that

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

Substituting (32) into the right-hand side of the above relationship, we denote the derived three terms by I_1, I_2 and I_3, respectively. Note that

$$\sqrt{nh_e}I_1=O_{\mathbb{P}}(\sqrt{nh_e}h_1^p)=O_{\mathbb{P}}(n^{-\frac{p\varepsilon_0}{2}}). \tag{35}$$

I_2 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, it follows that $\sqrt{nh_e}I_{21}=o_{\mathbb{P}}(1)$ (see (9) in Supporting Information). By the standard weak law of large numbers (WLLN), it can be obtained that $\sqrt{nh_e}I_{22}=O_{\mathbb{P}}(\sqrt{h_e})$. Therefore, we have $\sqrt{nh_e}I_2=o_{\mathbb{P}}(1)$. Analogously, it can also be obtained that $\sqrt{nh_e}I_3=o_{\mathbb{P}}(1)$. From these two relationships and (35), we have that

$$\sqrt{nh_e}(\hat{\beta}_3-\hat{\beta}_{x_u}^{\text{oracle}})=o_{\mathbb{P}}(1). \tag{36}$$

Finally, combining (36) and (34), the theorem holds.