Ann. Inst. Statist. Math.
Vol. 52, No. 3, 459–470 (2000)

NONPARAMETRIC ESTIMATION OF A CONDITIONAL QUANTILE FOR α-MIXING PROCESSES

TOSHIO HONDA

Institute of Social Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan

(Received August 5, 1997; revised March 23, 1999)

Abstract. Let $(X_i', Y_i)'$ be a set of observations form a stationary α-mixing process and $\theta(x)$ be the conditional α-th quantile of Y given $X = x$. Several authors considered nonparametric estimation of $\theta(x)$ in the i.i.d. setting. Assuming the smoothness of $\theta(x)$, we estimate it by local polynomial fitting and prove the asymptotic normality and the uniform convergence.

Key words and phrases: Conditional quantile, local polynomial fitting, α-mixing process, exponential inequality, Bahadur representation, uniform convergence.

1. Introduction

Let $(X_1', Y_1)', (X_2', Y_2)', \dots, (X_n', Y_n)'$ be a set of observations from a stationary α-mixing process, where Y_i are real valued and X_i are d-dimensional. For a fixed $0 < \alpha < 1$, let $\theta(x)$ denote the conditional α-th quantile of Y_i given $X_i = x$. We deal with the nonparametric estimation of $\theta(x)$ based on local polynomial fitting. Several authors considered the same problem in the i.i.d. setting, for example, Bhattacharya and Gangopadhyay (1990), Jones and Hall (1990), Mehra *et al.* (1991), Chaudhuri (1991*a*, 1991*b*), Fan *et al.* (1994), Welsh (1996), Xiang (1996). See also the references therein. Local polynomial estimators are examined in Chaudhuri (1991*a*, 1991*b*), Fan *et al.* (1994), and Welsh (1996). It is well known that local polynomial fitting by the method of weighted least squares gives estimators with some desirable properties. See Fan and Gijbels (1996) about this.

For stationary α-mixing processes, Truong and Stone (1992) considered estimation of the conditional median function by local median, which corresponds to $p = 1$ and $K(z) = I\{|z| \le 1\}$ here. They derived the rates of convergence. We define the estimators of $\theta(x)$ and the derivatives based on local polynomial fitting of any order following Jones and Hall (1990) and Chaudhuri (1991*a*). Chaudhuri (1991*a*) defined the estimators by local polynomial fitting under the loss function

$$H_\alpha(t) = |t| + (2\alpha - 1)t. \tag{1.1}$$

Then we investigate the asymptotic properties by the method of Babu (1989), derive the Bahadur representations, show the asymptotic normality, and evaluate the remainder terms closely.

Estimators based on the method of least squares have many optimal properties and it is easy to study the asymptotic properties. However, they do not perform well when the error distribution is heavy-tailed. They are also sensitive to outliers. On the other hand the method of least absolute deviations produces estimators which are robust to heavy-tailed errors and outliers. In addition regression quantiles give a good description of data. For example, see Chaudhuri *et al.* (1997).

For dependent observations, a lot of authors considered nonparametric estimation of the conditional mean function. Particularly Masry (1996a, 1996b) and Masry and Fan (1997) dealt with α-mixing processes by local polynomial fitting and proved the consistency and asymptotic normality. Although estimators of the conditional mean function are useful for data description, prediction, and examination of the effect of eplanatory variables, they might have the drawbacks that we referred to above. Our estimators of the onditional quantile functions, especially the conditional median function, are useful for the same purposes and robust to heavy-tailed distributions and outliers. In spite of the usefulness and robustness, the rates of convergnce are derived only for $p = 1$. Therefore it is improtant to investigate the properties from a theoretical point of view. Note that the exponential inequality for α-mixing processes of Rio (1995) and the related results of Liebscher (1996) are useful in this article.

In Section 2 we introduce several notations to define the estimator, and then state the theorems on the local Bahadur representation and the asymptotic normality. Those theorems are proved in Section 3. Then we present the theorems on the global properties with respect to x and outline the proofs in Section 4.

2. Local Bahadur representation and asymptotic normality

Let $\{(X_i', Y_i)'\}$ be a stationary α-mixing process with mixing coefficients α_k. See Hall and Heyde (1980) for the definition and the basic properties. At first we introduce several notations and assumptions. C_i are generic positive constants throughout this paper and the subscripts have no specific meaning. We define the sign function by

$$\mathrm{sign}(v) = 1(v > 0), 0(v = 0), \quad \text{and} \quad -1(v < 0).$$

$\mathcal{S}_i$ denotes the σ-field which is generated by $\{X_i, X_{i-1}, \ldots, Y_{i-1}, Y_{i-2}, \ldots\}$. The conditional distribution functions are defined as follows:

$$F(x, y) = P(Y_i \leq y \mid X_i = x) \quad \text{and} \quad G(x, u) = P(\epsilon_i \leq u \mid X_i = x) = F(x, u + \theta(x)),$$

where $\theta(x)$ is the conditional α-th quantile and $\epsilon_i = Y_i - \theta(X_i)$.

Here we estimate $\theta(x)$ for a fixed x. In Section 4 we consider the global properties of the estimator defined below.

ASSUMPTION 1. $\theta(x)$ is p times continuously differentiable in some neighborhood of x. The Taylor expansion up to the $(p-1)$-th order is written as

$$\theta(v) = \sum_{\lambda \in \Lambda} \beta_\lambda^x h^{-|\lambda|}(v - x)^\lambda - r_x(v) = P_h(\beta^x, v - x) - r_x(v), \tag{2.1}$$

where $\Lambda = \{(\lambda_1, \ldots, \lambda_d) \mid \lambda_i$ are nonnegative integers and $\sum \lambda_i < p\}$, $|\lambda| = \sum \lambda_i$, $v^\lambda = \prod v_i^{\lambda_i}$ ($\lambda \in \Lambda$ and $v \in R^d$). Let $|\Lambda|$ be the cardinal number of Λ. We arrange $h^{-|\lambda|}(v-x)^\lambda$ and β_λ^x in the ascending order with respect to λ and denote them by $\overline{v-x} \in R^{|\Lambda|}$ and $\beta^x \in R^{|\Lambda|}$, respectively. Note that both β^x and $\overline{v-x}$ depend on h. Considering the Taylor expansion up to the p-th order, we can define $(\overline{v-x}', \widetilde{v-x}')'$ and $(\beta^{x\prime}, \gamma^{x\prime})'$ in the same way as $\overline{v-x}$ and β^x. Then the Taylor expansion can be written as

$$\theta(v) = \overline{v-x}'\beta^x + \widetilde{v-x}'\gamma^x + o(|v-x|^p).$$

ASSUMPTION 2. For any $i_1 < \cdots < i_{|\Lambda|}$,

$$P(\{\overline{X_{i_1} - x}, \ldots, \overline{X_{i_{|\Lambda|}} - x}\} \text{ is linearly independent.}) = 1.$$

It is easily shown that Assumption 2 does not depend on h. When we take $d = 2$ and $p = 2$ with $i_1 = 1$, $i_2 = 2$, $i_3 = 3$, and $X_i = (X_{i1}, X_{i2})'$, the linear independence holds if and only if

$$h^2 \begin{vmatrix} 1 & 1 & 1 \\ \frac{X_{11}-x_1}{h} & \frac{X_{21}-x_1}{h} & \frac{X_{31}-x_1}{h} \\ \frac{X_{12}-x_2}{h} & \frac{X_{22}-x_2}{h} & \frac{X_{32}-x_2}{h} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ X_{11}-x_1 & X_{21}-x_1 & X_{31}-x_1 \\ X_{12}-x_2 & X_{22}-x_2 & X_{32}-x_2 \end{vmatrix} \neq 0.$$

Since all the first order derivatives of the determinant with respect to x are zero, it does not depend on x. This argument easily extends to the general cases. Therefore Assumption 2 is equivalent to

$$P(\{\overline{X_{i_1} - x}, \ldots, \overline{X_{i_{|\Lambda|}} - x}\} \text{ is linearly independent for any } x.) = 1$$

for any $i_1 < \cdots < i_{|\Lambda|}$.

ASSUMPTION 3. $G(u, v)$ has the density $g(u, v)$ in some neighborhood of $(x, 0)$ and $g(u, v)$ satisfies in the neighborhood

$$|g(u,v) - g(u,0)| \leq C_1|v|^{C_2}(C_2 \geq 1/2) \quad \text{and} \quad C_3 < g(u,0) < C_4.$$

ASSUMPTION 4. $K(v)$ is a bounded nonnegative kernel function with the compact support $\{|v| \leq 1\} \subset R^d$. The bandwidth h is put to $C_1 n^{-1/(2p+d)}$ and the dependence on n is suppressed.

$K(\cdot/h)$ is denoted by $K_h(\cdot)$.

ASSUMPTION 5.

$$C_1 h^d I_{|\Lambda|} < E\{K_h(X_1 - x)\overline{X_1 - x}\,\overline{X_1 - x}' g(X_1, 0)\} < C_2 h^d I_{|\Lambda|}$$
$$E\{K_h(X_i - x)K_h(X_j - x)\} < C_3 h^{d+1} \quad \text{for} \quad i \neq j.$$

ASSUMPTION 6. The conditional distribution of Y_i given $\mathcal{S}_i$ has no atom with probability 1.

Remark 2.1. Assumption 6 holds when

$$Y_i = g(\mathcal{S}_i) + \sigma(\mathcal{S}_i)\eta_i, \tag{2.2}$$

where η_i and $\mathcal{S}_i$ are independent and η_i has the continuous distribution. Assumption 6 can be replaced with the primitive one: (3.5) $\neq 0$ a.s. for any $i_1 < i_2 < \cdots < i_{|\Delta|+1}$.

ASSUMPTION 7.

$$C_1 h^d I_{|\Lambda|} < E\{K_h^2(X_1 - x)\overline{X_1 - x}\,\overline{X_1 - x}'\} < C_2 h^d I_{|\Lambda|}.$$

ASSUMPTION 8. $P(Y_i \le y \mid X_i) = P(Y_i \le y \mid \mathcal{S}_i)$ for any y and i with probability 1.

Remark 2.2. Assumption 8 holds if $\mathcal{S}_i$ in (2.2) are replaced with X_i.

We give an estimator of $\theta(x)$ by estimating β^x. Consider the minimization problem

$$\sum_{i=1}^{n} K_h(X_i - x) H_\alpha(Y_i - P_h(\beta, X_i - x)) \to \min .$$

Take one of the solutions as $\hat{\beta}^x$. Then $\theta(x)$ is estimated by the first element of $\hat{\beta}^x$. The derivatives of $\theta(x)$ up to the $(p-1)$-th order can be estimated by multiplying the corresponding elements of $\hat{\beta}^x$ by $C_\lambda h^{-|\lambda|}$, where C_λ depends on λ.

When Assumption 8 holds, we have the same results as in the i.i.d. setting.

THEOREM 2.1. *Suppose Assumptions* 1–6 *and* 8 *hold and that* $\alpha_k \le C_1 k^{-r}$ *and* $r > d + 7 + 4d/p$. *Then*

$$(2.3) \quad \hat{\beta}^x - \beta^x = \left(2E\left\{ \sum_{i=1}^{n} K_h(X_i - x)\overline{X_i - x}\,\overline{X_i - x}' g(X_i, 0) \right\} \right)^{-1} \times \sum_{i=1}^{n} K_h(X_i - x)\overline{X_i - x}\{(\mathrm{sign}(\epsilon_i) + 2\alpha - 1) + 2(\alpha - G(X_i, r_x(X_i)))\} + O(h^{3p/2}(\log n)^{3/4}) \quad a.s.$$

Without Assumption 8, the result deteriorates to

THEOREM 2.2. *Suppose Assumptions* 1–6 *hold and that* $\alpha_k \le C_1 k^{-r}$ *and for some* $p/2 < \eta < p$, $r > (d+p)/(2\eta - p) \vee (d + 2 + 2d/p) \vee [\{d(p-\eta) + 2(2p+d)\}/\eta - 1]$. *Then*

$$(2.4) \quad \hat{\beta}^x - \beta^x = \left(2E\left\{ \sum_{i=1}^{n} K_h(X_i - x)\overline{X_i - x}\,\overline{X_i - x}' g(X_i, 0) \right\} \right)^{-1} \times \sum_{i=1}^{n} K_h(X_i - x)\overline{X_i - x}\{(\mathrm{sign}(\epsilon_i) + 2\alpha - 1) + 2(\alpha - G(X_i, r_x(X_i)))\} + O(h^{2p-\eta}) \quad a.s.$$

We defer the proofs to the next section and state the uniform versions of the theorems in Section 4. As to the asymptotic distribution of $\hat{\beta}^x - \beta^x$, we have the following corollary to Theorems 2.1 and 2.2.

COROLLARY 2.1. *Suppose Assumptions* 1–7 *hold and that* $\alpha_k \le C_1 k^{-r}$ *and* $r > d + 2 + 2d/p$. *Then*

$$(A_n^{-1} C_n A_n^{-1})^{-1/2}\{h^{-p}(\hat{\beta}^x - \beta^x) - A_n^{-1} B_n\} \xrightarrow{D} N_{|\Lambda|}(0, I)$$

where

$$A_n = \frac{1}{nh^d} E\left\{ \sum_{i=1}^n K_h(X_i - x)\overline{X_i - x}\,\overline{X_i - x}' g(X_i, 0) \right\},$$

$$B_n = \frac{1}{nh^d} E\left\{ \sum_{i=1}^n K_h(X_i - x)\overline{X_i - x}\widetilde{X_i - x}' g(X_i, 0) \right\} h^{-p}\gamma^x,$$

$$C_n = \frac{\alpha(1-\alpha)}{nh^d} E\left\{ \sum_{i=1}^n K_h^2(X_i - x)\overline{X_i - x}\,\overline{X_i - x}' \right\}, \quad \textit{and}$$

$(A_n^{-1}C_nA_n^{-1})^{1/2}$ *is the symmetric root matrix of* $A_n^{-1}C_nA_n^{-1}$.

$h^{-p}\gamma^x$ consists of constant multiples of the p-th order derivatives of $\theta(x)$.

Remark 2.3. With further assumptions, for example, the smoothness of the density of X_i, A_n, B_n, and C_n converge to some constants.

3. Proofs

We almost follow Babu (1989). In this section we can take $\alpha = 1/2$ and $x = 0$ without loss of generality. We sometimes write K_{hi}, $G_i(v)$, and r_i for $K_h(X_i)$, $G(X_i, v)$, and $r_0(X_i)$, respectively for notational convenience.

LEMMA 3.1. *Suppose that* $\alpha_k \le C_1 k^{-r} (r > d + 2 + 2d/p)$ *and that Assumptions* 1, 4, *and* 5 *hold. Then for any positive* a, *if* B *is sufficiently large,*

$$(3.1) \quad \left| \sum_{i=1}^n K_{hi}(|Y_i - P_h(\beta^0, X_i)| - |Y_i - P_h(\beta, X_i)|) - E\{K_{hi}(|Y_i - P_h(\beta^0, X_i)| - |Y_i - P_h(\beta, X_i)|)\} \right|$$
$$\le aB^2 \log n, \quad \textit{uniformly on} \quad \{\beta \mid |\beta - \beta^0| = Bh^p(\log n)^{1/2}\} \ \textit{a.s.}$$

PROOF. We write $P_{\beta,i}$ for $P_h(\beta, X_i)$. Noting

$$Y_i - P_{\beta,i} = \epsilon_i - P_{\beta-\beta^0,i} - r_i \quad \text{and} \quad Y_i - P_{\beta^0,i} = \epsilon_i - r_i,$$

we get

$$|Y_i - P_{\beta^0,i}| - |Y_i - P_{\beta,i}| = \int_{r_i}^{P_{\beta-\beta^0,i}+r_i} (1 - 2I\{\epsilon_i \le v\})dv.$$

Then we can represent the summand of (3.1) as

$$Z_i = K_{hi} \int_{r_i}^{P_{\beta-\beta^0,i}+r_i} (1 - 2I\{\epsilon_i \le v\})dv - E\left\{ K_{hi} \int_{r_i}^{P_{\beta-\beta^0,i}+r_i} (1 - 2I\{\epsilon_i \le v\})dv \right\}.$$

We prove the lemma using Theorem 2.1 and Lemma 2.3 of Liebscher (1996). From Assumption 5 we can evaluate $R^*(n)$, R_∞, and $R_2^2(n)$ of Liebscher (1996) as

$$R^*(n) < C_1 B^2 n^{-1} h \log n, \quad R_\infty < C_2 Bh^p (\log n)^{1/2}, \quad \text{and} \quad R_2^2(n) < C_3 B^2 n^{-1} \log n.$$

From Lemma 2.3 there, we can choose the parameters of the theorem as

$$D(n,m) < C_4 \frac{B^2 m}{n} \log n, \qquad S(n) < C_5 B h^p (\log n)^{1/2},$$
$$\epsilon = B^2 a \log n, \qquad \text{and} \qquad N = n^{C_6} \quad \text{for some } C_6 < p/(2p+d).$$

Therefore we can prove the lemma by dividing $\{\beta \mid |\beta - \beta^0| = Bh^p(\log n)^{1/2}\}$ into sufficiently small cells. The number of the cells is $O((\frac{h^p(\log n)^{1/2}}{(nh^d)^{-1}\log n})^{d-1})$. □

In Lemma 3.2 we prove the consistency of $\hat{\beta}^x$. The convexity of $K_h(X_i - x)H_\alpha(\cdot)$ plays an important role in the proof.

LEMMA 3.2 *Suppose that Assumption* 3 *holds in addition to the assumptions of Lemma* 3.1. *Then*

$$|\hat{\beta}^0 - \beta^0| = O(h^p(\log n)^{1/2}) \quad a.s.$$

PROOF. From the convexity of the objective function and Lemma 3.1, all we have to do is to show that for some $a > 0$, there exists a sufficiently large $B > 0$ such that

$$\sum_{i=1}^{n} E\{K_{hi}(|Y_i - P_h(\beta^0, X_i)| - |Y_i - P_h(\beta, X_i)|)\} \le -aB^2 \log n \tag{3.2}$$

on $\{\beta \mid |\beta - \beta^0| = Bh^p(\log n)^{1/2}\}$. The left hand side of (3.2) is rewritten as

$$\begin{aligned}
&\sum_{i=1}^{n} E\left\{K_{hi}\int_{r_i}^{P_{\beta-\beta^0}+r_i}(1-2G_i(v))dv\right\}\\
&\quad = -2\sum_{i=1}^{n} E\left\{K_{hi}\int_{r_i}^{P_{\beta-\beta^0}+r_i} g(X_i,\theta_i v)v dv\right\}\\
&\quad < -C_1\sum_{i=1}^{n} E\{K_{hi}(\beta-\beta^0)'\overline{X}_i\overline{X}_i'(\beta-\beta^0)g(X_i,0)\} + O(B(\log n)^{1/2})\\
&\quad < -C_2 B^2 \log n + O(B(\log n)^{1/2}).
\end{aligned}$$

Then the proof is complete. □

For the results in Section 4, we present the following lemma in the general form.

LEMMA 3.3. *Suppose that Assumptions* 2 *and* 6 *hold. Then*

$$\left|\sum_{i=1}^{n} K_h(X_i - x)h^{-|h|}(X_i - x)^\lambda \operatorname{sign}(Y_i - P_h(\hat{\beta}^x, X_i - x))\right| \le C_1 \quad \textit{for any } x \in S \ \ a.s.,$$

where S is any compact subset on which Assumption 5 *holds uniformly.*

PROOF. All the directional derivatives of the objective function at $\beta = \hat{\beta}^x$ must be non-negative. This implies that

$$\begin{aligned}
&\left|\sum_{i=1}^{n} K_{hi}h^{-|\lambda|}(X_i - x)^\lambda \operatorname{sign}(Y_i - P_h(\hat{\beta}^x, X_i - x))\right|\\
&\qquad \le \sum_{i=1}^{n} K_{hi} I\{Y_i = P_h(\hat{\beta}^x, X_i - x)\}|h^{-|\lambda|}(X_i - x)^\lambda|.
\end{aligned} \tag{3.3}$$

We show the right hand side of (3.3) is bounded by a constant. We have only to prove the event

$$\begin{pmatrix} Y_{i_1} \\ \vdots \\ Y_{i_{|\Lambda|+1}} \end{pmatrix} = \begin{pmatrix} \overline{X_{i_1} - x}' \\ \vdots \\ \overline{X_{i_{|\Lambda|+1}} - x}' \end{pmatrix} \hat{\beta}^x \tag{3.4}$$

for some x happens with probability 0. The determinant of the $(|\Lambda|+1) \times (|\Lambda|+1)$ matrix

$$\begin{pmatrix} \dfrac{Y_{i_1}}{\overline{X_{i_1} - x}} & \cdots & \dfrac{Y_{i_{|\Lambda|+1}}}{\overline{X_{i_{|\Lambda|+1}} - x}} \end{pmatrix}$$

can be written as

$$(-1)^{|\Lambda|} h_1(X_{i_1}, \ldots, X_{i_{|\Lambda|}}) Y_{i_{|\Lambda|+1}} + h_2(X_{i_1}, \ldots, X_{i_{|\Lambda|+1}}, Y_{i_1}, \ldots, Y_{i_{|\Lambda|}}), \tag{3.5}$$

where $h_1(\cdot)$ is the determinant of Assumption 2 and $h_2(\cdot)$ does not depend on x. Then the lemma follows from Assumption 6. □

In the next two lemmas we evaluate

$$\begin{aligned} &\sum_{i=1}^{n} K_{hi} h^{-|\lambda|} X_i^{\lambda} \operatorname{sign}(Y_i - P_h(\beta, X_i)) \\ &\qquad = \sum_{i=1}^{n} K_{hi} h^{-|\lambda|} X_i^{\lambda} \operatorname{sign}(\epsilon_i - P_h(\beta - \beta^0, X_i) - r_0(X_i)) \end{aligned}$$

on $\{\beta \mid |\beta - \beta^0| \le B h^p (\log n)^{1/2}\}$. Define $\Delta^0_{\lambda,i}(\beta)$ by

$$\begin{aligned} \Delta^0_{\lambda,i}(\beta) = K_{hi} h^{-|\lambda|} X_i^{\lambda} \{ &\operatorname{sign}(\epsilon_i - P_h(\beta - \beta^0, X_i) - r_0(X_i)) - \operatorname{sign}(\epsilon_i - r_0(X_i)) \\ &+ 2(G(X_i, P_h(\beta - \beta^0, X_i) + r_0(X_i)) - G(X_i, r_0(X_i)))\}. \end{aligned}$$

LEMMA 3.4. *Let α_k be bounded by $C_1 k^{-r}$ and Assumptions* 1, 3, 4, *and* 5. *For $p/2 \le \eta < p$, we put $q = 2(p+d)/(2p - 2\eta + d)$. Suppose that $\sum_{k=1}^{\infty} \alpha_k^{1-2/q} < \infty$ when $q > 2$ and Assumption* 8 *holds when $q = 2$. Then for any $\gamma_1 > 0$ and $0 < \gamma_2 < \eta/(2p+d)$, if B is sufficiently large,*

$$P\left(\left| \sum_{i=1}^{n} \Delta^0_{\lambda,i}(\beta) \right| \ge B h^{-\eta} (\log n)^{(q+1)/(2q)} \right) \le C_2 (n^{-\gamma_1} + n^{1-(r+1)\gamma_2})$$

uniformly on $\{\beta \mid |\beta - \beta^0| \le B h^p (\log n)^{1/2}\}$

PROOF. We choose the parameters of Theorem 2.1 of Liebscher as

$$\begin{aligned} &D(n,m) < C_3 B^{2/q} m h^{2(p+d)/q} (\log n)^{1/q}, \qquad S(n) < C_4, \\ &\epsilon = B h^{-\eta} (\log n)^{(q+1)/(2q)}, \qquad \text{and} \qquad N = n^{\gamma_2}. \end{aligned}$$

Lemma 2.2 there is employed for $D(n,m)$. The details are omitted. □

LEMMA 3.5. *Besides the assumptions of lemma 3.4 we assume that* $\{d(p-\eta)+2(2p+d)\}\eta^{-1}-1<r$. *Then if* B *is sufficiently large,*

$$\left|\sum_{i=1}^{n}\Delta_{\lambda,i}^{0}(\beta)\right|\le Bh^{-\eta}(\log n)^{(q+1)/(2q)}$$

uniformly on $\{\beta\mid|\beta-\beta^0|\le Bh^p(\log n)^{1/2}\}$ *a.s.*

PROOF. We put $v_n=h^{2p-\eta}(\log n)^{(q+1)/(2q)}$. When $|\alpha-\beta|\le v_n$, we can choose C_1 and C_2 such that

$$|\Delta_{\lambda,i}^{0}(\beta)-\Delta_{\lambda,i}^{0}(\alpha)|\le C_1K_{hi}\{I\{A_{\beta,i}\}-P(A_{\beta,i}\mid X_i)+2P(A_{\beta,i}\mid X_i)\}$$

where $A_{\beta,i}=\{|\epsilon_i-P_h(\beta-\beta^0,X_i)-r_0(X_i)|\le C_2v_n\}$. We have to evaluate

$$\sum_{i=1}^{n}K_{hi}P(A_{\beta,i}\mid X_i)\quad\text{and}\tag{3.6}$$

$$\sum_{i=1}^{n}K_{hi}\{I\{A_{\beta,i}\}-P(A_{\beta,i}\mid X_i)\}.\tag{3.7}$$

We have from Assumption 3

$$(3.6)\le C_3v_nnh^d=C_3h^{-\eta}(\log n)^{(q+1)/(2q)}.$$

We evaluate (3.7) by applying Theorem 2.1 of Liebscher with

$$D(n,m)<C_4m(h^dv_n)^{2/q},\qquad S(n)<C_5,$$
$$\epsilon=Bh^{-\eta}(\log n)^{(q+1)/(2q)},\qquad\text{and}\qquad N=n^{C_6}\qquad\text{for some}\quad C_6<\eta/(2p+d).$$

Then the lemma is verified by using Lemma 3.4 and dividing $\{\beta\mid|\beta-\beta^0|\le Bh^p(\log n)^{1/2}\}$ into sufficiently small cells whose number is $O(h^{d(\eta-p)})$. □

PROOF OF THEOREM 2.1 AND 2.2. We can take an η' which satisfies the assumptions and is smaller than η when $p/2<\eta<p$. Hereafter we write η for η'.

From Lemmas 3.3–3.5,

$$\begin{aligned}\sum_{i=1}^{n}&K_{hi}\overline{X}_i\,\mathrm{sign}(\epsilon_i-r_0(X_i))\\&=2\sum_{i=1}^{n}K_{hi}\overline{X}_i(G(X_i,P_h(\hat\beta^0-\beta)+r_0(X_i))-G(X_i,r_0(X_i)))\\&\quad+O(h^{-\eta}(\log n)^{(q+1)/(2q)}).\end{aligned}\tag{3.8}$$

By the Taylor expansion of the first order and the argument similar to the density estimation, we have

$$\begin{aligned}\sum_{i=1}^{n}&K_{hi}\overline{X}_i(G(X_i,P_h(\hat\beta^0-\beta)+r_0(X_i))-G(X_i,r_0(X_i)))\\&=\sum_{i=1}^{n}K_{hi}\overline{X}_i\overline{X}_i'g(X_i,0)(\hat\beta^0-\beta)+O(h^{-\eta}(\log n)^{(q+1)/(2q)})\\&=E\left\{\sum_{i=1}^{n}K_{hi}\overline{X}_i\overline{X}_i'g(X_i,0)\right\}(\hat\beta^0-\beta)+O(h^{-\eta}(\log n)^{(q+1)/(2q)}).\end{aligned}\tag{3.9}$$

We can verify in the same way as Lemmas 3.4 and 3.5 that

$$(3.10)\qquad \sum_{i=1}^{n} K_{hi}\overline{X}_i \operatorname{sign}(\epsilon_i - r_0(X_i)) = \sum_{i=1}^{n} K_{hi}\overline{X}_i\{\operatorname{sign}(\epsilon_i) + (1 - 2G(X_i, r_0(X_i)))\} + O(h^{-\eta}(\log n)^{(q+1)/(2q)}).$$

Then the proof is completed by combining (3.8)–(3.10).

PROOF OF COROLLARY 2.1. It is similar to that of Theorem 3 of Masry and Fan (1997). The details are omitted. □

4. Uniform convergence and uniform bahadur representation

We consider the properties of $\hat{\beta}^x$ on some fixed compact set S. Assumptions 1, 3, 4 and 5 are strengthened and Assumption 9 is added for this purpose. Then we have the same results as in Section 2. Strengthened Assumption 1 is called Assumption 1' and so on.

Assumptions 1' and 5' will be clear. $(x, 0)$ is replaced with $\{(u, 0) \mid u \in S\}$ in Assumption 3'.

ASSUMPTION 4'. $K(v) = W(v)I\{|v| \le 1\}$ and $W(v)$ is bounded, nonnegative, and Lipschitz continuous. $h = C_1 n^{-1/(2p+d)}$.

ASSUMPTION 9. X_i has a bounded density.

Assumption 9 is added to evaluate $I\{h - \delta_n \le |X_i - x| \le h + \delta_n\}$ in Lemmas 4.1 and 4.2.

THEOREM 4.1. *Suppose that Assumptions* 1', 2, 3'–5', 6, 8, *and* 9 *hold and that* $\alpha_k \le C_1 k^{-r}(r > (5dp + 7p + 6d)/p)$. *Then* (2.3) *with* $E\{\}$ *removed holds uniformly on* S.

THEOREM 4.2. *Suppose that Assumptions* 1', 2, 3'–5', 6, *and* 9 *hold and that* $\alpha_k \le C_1 k^{-r}(r > (d + p)/(2\eta - p) \vee ((3dp + 4p + 3d - d\eta)/\eta - 1) \vee (3dp + 2p + 3d)/p)$ *for* $p/2 < \eta < p$. *Then* (2.4) *with* $E\{\}$ *removed holds uniformly on* S.

The following theorem corresponds to Lemma 3.2.

THEOREM 4.3. *Suppose that Assumptions* 1', 2, 4', 5', *and* 9 *hold and that* $\alpha_k \le C_1 k^{-r}(r > (3dp + 2p + 3d)/p)$. *Then*

$$|\hat{\beta}^x - \beta^x| = O(h^p(\log n)^{1/2}) \qquad uniformly\ on \quad S\ a.s.$$

We only outline the proofs of the theorems because they are similar to those in the previous section. Lemma 4.1 corresponds to Lemma 3.1.

LEMMA 4.1. *Suppose that Assumptions 1', 4', 5' and 9 hold and that $\alpha_k \le C_1 k^{-r}$ $(r > (3dp + 2p + 3d)/p)$. Then for any positive a, if B is sufficiently large,*

$$\begin{aligned}(4.1)\qquad &\Big|\sum_{i=1}^{n} K_h(X_i - x)(|Y_i - P_h(\beta^x, X_i - x)| - |Y_i - P_h(\beta, X_i - x)|) \\ &\quad - E\{K_h(X_i - x)(|Y_i - P_h(\beta^x, X_i - x)| - |Y_i - P_h(\beta, X_i - x)|)\}\Big| \\ &\quad \le aB^2 \log n,\end{aligned}$$

uniformly in $x \in S$ and $\beta \in \{\beta \mid |\beta - \beta^x| = Bh^p(\log n)^{1/2}\}$ a.s.

PROOF. We can verify the lemma by the standard argument for proving the uniform convergence of the kernel density estimator or regression estimator. We only outline the proof. For a fixed $x \in S$, the evaluation of (4.1) is done in Lemma 3.1. Next we deal with

$$\begin{aligned}(4.2)\qquad &K_h(X_i - x)(|Y_i - P_h(\beta^x, X_i - x)| - |Y_i - P_h(\beta, X_i - x)|) \\ &\quad - K_h(X_i - y)(|Y_i - P_h(\beta^y, X_i - y)| - |Y_i - P_h(\alpha, X_i - y)|)\end{aligned}$$

when $|x - y| \le \delta_n$, $|\beta - \beta^x| = Bh^p(\log n)^{1/2}$, and $|\alpha - \beta^y| = Bh^p(\log n)^{1/2}$. Note that $|\beta^x - \beta^y| \le C_1\delta_n$ and there exists a β s.t. $|\beta - \alpha| \le C_1\delta_n$ for any α.

We have only to consider two cases.

$|X_i - x| \le h$ and $|X_i - y| \le h$:

Since

$$\begin{aligned}&|(K_h(X_i - x) - K_h(X_i - y))(|Y_i - P_h(\beta^x, X_i - x)| - |Y_i - P_h(\beta, X_i - x)|)| \\ &\quad \le C_2\delta_n B h^{p-1}(\log n)^{1/2} I\{|X_i - x| \le h\}\end{aligned}$$

and

$$\begin{aligned}&|K_h(X_i - y)\{(|Y_i - P_h(\beta^x, X_i - x)| - |Y_i - P_h(\beta, X_i - x)|) \\ &\qquad - (|Y_i - P_h(\beta^y, X_i - y)| - |Y_i - P_h(\alpha, X_i - y)|)\}| \\ &\quad \le C_3\delta_n h^{-1} I\{|X_i - x| \le h\}, \\ &|(4.2)| \le C_4 B\delta_n h^{-1} I\{|X_i - x| \le h\}.\end{aligned}$$

$|X_i - x| \le h$ and $|X_i - y| > h$, or $|X_i - x| > h$ and $|X_i - y| \le h$:

$$|(4.2)| \le C_5 Bh^p(\log n)^{1/2} I\{h - \delta_n \le |X_i - x| \le h + \delta_n\}.$$

Therefore we can prove the uniformity by dividing S into small cells with the diameter $\delta_n = O((nh^{d-1})^{-1})$ and using the argument analogous to Lemmas 3.1, 3.2 and 3.5. Then we must have

$$\sum_{n=1}^{\infty} n^{1-\lambda(1+r)} n^{p(d-1)/(2p+d)} n^{d(2p+1)/(2p+d)} < \infty,$$

where $n^{1-\lambda(1+r)}$ $(\lambda < p/(2p+d))$ is from the exponential inequality and $n^{p(d-1)/(2p+d)}$ and $n^{d(2p+1)/(2p+d)}$ are from the division of $\{|\beta - \beta^x| = Bh^p(\log n)^{1/2}\}$ and S, repectively. The assumption on α^k follows from this. □

Theorem 4.3 can be shown in the same way as Lemma 3.2. So omitted. Lemma 4.2 corresponds to Lemma 3.5.

LEMMA 4.2. *For $p/2 \le \eta < p$, we put $q = 2(p+d)/(2p - 2\eta + d)$. Suppose that Assumptions 1', 3', 4', 5', and 9 hold and that $\alpha_k \le C_1 k^{-r} (r > (3dp+4p+3d-d\eta)/\eta - 1)$. If $q = 2$, we have to impose Assumption 8. Otherwise $r > (d+p)/(2\eta - p)$. Then if B is sufficiently large,*

$$\left| \sum_{i=1}^{n} \Delta_{\lambda,i}^{x}(\beta) \right| \le Bh^{-\eta} (\log n)^{(q+1)/(2q)}$$

uniformly in $x \in S$ and $\beta \in \{\beta \mid |\beta - \beta^x| \le Bh^p (\log n)^{1/2}\}$ a.s.

PROOF. The probability of the event $\{|\sum_{i=1}^n \Delta_{\lambda,i}^x(\beta)| > Bh^{-\eta}(\log n)^{(q+1)/(2q)}\}$ is evaluated in Lemma 3.4. We consider

$$\sum_{i=1}^{n} |\Delta_{\lambda,i}^{x}(\beta) - \Delta_{\lambda,i}^{y}(\alpha)| \tag{4.3}$$

when $|x - y| \le \delta_n$, $|\beta - \beta^x| \le Bh^p(\log n)^{1/2}$, and $|\alpha - \beta^y| \le Bh^p(\log n)^{1/2}$. For some C_1, $|\beta^x - \beta^y| \le C_1 \delta_n$ and there exists a β s.t. $|\beta - \alpha| \le C_1 \delta_n$ for any α. Then we have

$$\begin{aligned} &|P_h(\alpha, X_i - y) - P_h(\beta, X_i - x)| \le C_2 \frac{\delta_n}{h}, \\ &|P_h(\beta^y, X_i - y) - P_h(\beta^x, X_i - x)| \le C_2 \frac{\delta_n}{h}. \end{aligned} \tag{4.4}$$

By (4.4) and Assumption 4' we have

$$\begin{aligned} &|\Delta_{\lambda,i}^{x}(\beta) - \Delta_{\lambda,i}^{y}(\alpha)| \\ &\quad \le C_3 I\{|X_i - x| \le h\} \left[\frac{\delta_n}{h} + I\left\{ |\epsilon_i - P_h(\beta - \beta^x, X_i - x) - r_x(X_i)| \le C_4 \frac{\delta_n}{h} \right\} \right. \\ &\qquad \left. + I\left\{ |\epsilon_i - r_x(X_i)| \le C_5 \frac{\delta_n}{h} \right\} \right] + C_6 I\{h - \delta_n \le |X_i - x| \le h + \delta_n\}. \end{aligned} \tag{4.5}$$

Putting $\delta_n = O((nh^{d-1})^{-1})$, we can prove the lemma by dividing S and $\{\beta \mid |\beta - \beta^x| \le Bh^p(\log n)^{1/2}\}$ into small cells and applying the argument analogous to Lemmas 3.1, 3.2 and 3.5. Then we must have

$$\sum_{n=1}^{\infty} n^{1-\lambda(1+r)} n^{d(p-\eta)/(2p+d)} n^{d(2p+1)/(2p+d)} < \infty,$$

where $n^{1-\lambda(1+r)} (\lambda < \eta/(2p+d))$ is from the exponential inequality and $n^{p(d-1)/(2p+d)}$ and $n^{d(2p+1)/(2p+d)}$ are from the division of $\{|\beta - \beta^x| \le Bh^p(\log n)^{1/2}\}$ and S, repectively. One of the assumptions on α^k follows from this. □

Theorems 4.1 and 4.2 can be proved in the same way as Theorems 2.1 and 2.2. The details are omitted.

Acknowledgements

The author is very grateful to the two referees for their valuable comments.

References

Babu, G. J. (1989). Strong Representation for LAD estimators in linear models, *Probab. Theory. Related Fields*, **83**, 547–558.

Bhattacharya, P. K. and Gangopadhyay, A. (1990). Kernel and nearest neighbor estiamtion of a conditional quantile, *Ann. Statist.*, **18**, 1400–1415.

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

Chaudhuri, P. (1991*b*). Global nonparametirc estimation of conditional quantile functions and their derivatives, *J. Multivariate Anal.*, **39**, 246–269.

Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression, *Ann. Statist.*, **25**, 715–744.

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

Fan, J., Hu, T.-C. and Truong, Y. K. (1994). Robust nonparametric function stimation, *Scand. J. Statist.*, **21**, 433–446.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*, Academic Press, San Diego.

Jones, M. C. and Hall, P. (1990). Mean squared error properties of kernel estimates of regression quantiles, *Statist. Probab. Lett.*, **10**, 283–289.

Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation, *Stochastic Process. Appl.*, **65**, 69–80.

Masry, E. (1996*a*). Multivariate local polynomial regression for time series: uniform strong consisitency and rates, *J. Time Ser. Anal.*, **17**, 571–599.

Masry, E. (1996*b*). Multivariate regression estimation: local polynomial fitting for time series, *Stochastic Processes Appl.*, **65**, 81–101.

Masry, E. and Fan, J. (1997). Local plynomial estimation of regression functions for mixing processes, *Scand. J. Statist.*, **24**, 166–179.

Mehra, K. L., Rao, M. S. and Upadrasta, S. P. (1991). A smooth conditional quantile estimator and related applications of conditional empirical processes, *J. Multivariate Anal.*, **37**, 151–179.

Rio, E. (1995). The functional law of the iterated logarithm for strongly mixing sequences, *Ann. Probab.*, **23**, 1188–1203.

Truong, Y. K. and Stone, C. J. (1992). Nonparametric function estimation involving time series, *Ann. Statist.*, **20**, 77–97.

Welsh A. H. (1996). Robust estimation of smooth regression and spread functions and their derivatives, *Statist. Sinica*, **6**, 347–366.

Xiang, X. (1996). A kernel estimator of a conditional quantile, *J. Multivariate Anal.*, **59**, 206–216.