Ann. of Appl. Math.
33:4(2017), 364-378

LOCAL POLYNOMIAL DOUBLE-SMOOTHING ESTIMATION OF A CONDITIONAL DISTRIBUTION FUNCTION WITH DEPENDENT*†

Mimi Hong‡, Xianzhu Xiong
(*College of Math. and Computer Science, Fuzhou University, Fuzhou 350116, Fujian, PR China*)

Abstract

Based on the idea of local polynomial double-smoother, we propose an estimator of a conditional cumulative distribution function with dependent and left-truncated data. It is assumed that the observations form a stationary α-mixing sequence. Asymptotic normality of the estimator is established. The finite sample behavior of the estimator is investigated via simulations.

Keywords local polynomial double-smoother; conditional cumulative distribution function; left-truncated data; α-mixing; asymptotic normality

2000 Mathematics Subject Classification 62G05; 62N02

1 Introduction

Estimation of a conditional distribution function has many important applications. For example, estimation of a conditional quantile function can be obtained by inverting the estimation of the conditional distribution function. In addition, as [4] pointed out, conditional distribution estimation can be applied to construct prediction intervals for the next value in stationary time series.

The nonparametric estimation of a conditional distribution function has received much attention. For instance, in the independent and identically distributed (i.i.d.) case, [4] proposed two single-smoothing estimators (smoothing the covariates only) of the conditional distribution function, while [15,5,6] considered the local linear double-smoothing (smoothing the dependent variable and covariates) estimator of the conditional distribution function. [14] considered the local polynomial double-smoothing estimator of the conditional distribution function in time series data,

*This work was supported by National Natural Science Foundation of China (No.11301084), and Natural Science Foundation of Fujian Province (No.2014J01010).

†Manuscript received April 12, 2017; Revised June 26, 2017

‡Corresponding author. E-mail: 1114335900@qq.com

which extended the results of [15]. As [14] pointed out, compared with the single-smoothing estimator, the double-smoothing estimator not only appears closer to a distribution function, but also has more flexibility to reduce the mean-squared error when the optimal bandwidths are selected. In these papers, it is supposed that data is complete.

In some fields as reliability and survival analysis, the lifetime variables may not be completely observable, right-censored or left-truncated data are often encountered. In this paper we consider the case where the response variable is left-truncated. Left-truncated data often occurs in astronomy, epidemiology, biometry and economics, see [11,12]. For left-truncated data, [7] first proposed the kernel estimator (that is, the local constant double-smoothing estimator) of the conditional distribution function in the i.i.d. setting, and then obtained the kernel estimator of conditional quantile function by inversion. [7] also investigated asymptotic properties of two kernel estimators. [13] extended the results of [7] to dependent and left-truncated data. [8] proposed the local polynomial single-smoothing estimator of the conditional distribution function with dependent and left-truncated data and established the asymptotic normality of the estimator. Recently, in view of the advantage of local linear fitting and double-smoothing, [16] extended the local linear double-smoothing method of [15] to the left-truncated model. They proposed the local linear double-smoothing estimator of the conditional distribution function in the i.i.d. setting, and then obtained the local linear double-smoothing estimator of conditional quantile function by inversion. And they obtained the asymptotic normality of two local linear double-smoothing estimators.

Since the scenario with dependent data is an important one in lots of applications with survival data (see [1,13]), in this paper, we will consider the local polynomial double-smoothing (LPDS) estimator of the conditional distribution function in the dependent and left-truncated data, which extends the LPDS estimator of the conditional distribution function in [14] to the left-truncation model. Furthermore, the LPDS estimator here not only smooths the local polynomial single-smoothing estimator of [8], but also generalizes the local linear double-smoothing estimator of [16] in the i.i.d. setting to the LPDS estimator in the dependent data case. When left-truncated data is stationary and α-mixing, we establish the asymptotic normality of the LPDS estimator.

The rest of the paper is organized as follows. Section 2 recalls the random left-truncation model and introduces the LPDS estimator of the conditional distribution function. The asymptotic normality of the estimator is stated in Section 3, while the proof is given in Section 5. Finite-sample performance of the estimator is investigated by a simulation study in Section 4.

2 Estimator

2.1 Background for the left-truncation model

Let $\{(X_i, Y_i, T_i), 1 \le i \le N\}$ be a sequence of random vectors from (X, Y, T), where T is the truncation random variable and T is independent of (X, Y). In the random left-truncation model, for $i = 1, \cdots, N$, the observation of (X_i, Y_i) is interfered by the variable T_i such that all three quantities X_i, Y_i and T_i are observed if $Y_i \ge T_i$, and nothing is observable if $Y_i < T_i$. As a consequence of left-truncation, only a part of the original (or potential) sample $\{(X_i, Y_i, T_i), 1 \le i \le N\}$ can be observed. Such data are often encountered in many application fields including astronomy, medical studies and economics (see e.g., [11,12]). It should be noted that n, the size of the actually observed sample, is known but random, and N, the potential sample size, is unknown. Obviously, n and N satisfy that $n \le N$.

Since statistical inference is based on the observed n-sample, in what follows, the actually observed n-sample is denoted again by $\{(X_i, Y_i, T_i), 1 \le i \le n\}$, which will not lead to possible confusion. In addition, the results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the N-sample) but will be stated with respect to the conditional probability measure P (related to the actually observed n-sample). In the same manner, $\mathbb{E}$ and E denote the respective expectation operators related to $\mathbb{P}$ and P respectively. Let $\theta = \mathbb{P}\ (Y \ge T)$, then θ is the probability that Y can be observed. It is obvious that $\theta = 0$ implies that no data can be observed, thus we suppose throughout the paper that $\theta > 0$.

In the following, the actually observed sample $\{(X_i, Y_i, T_i), 1 \le i \le n\}$ is assumed to be a stationary and α-mixing sequence. Recall that a sequence $\{Z_i, i \ge 1\}$ is called to be an α-mixing (or strongly mixing) sequence, if the α-mixing coefficient

$$\alpha(n) := \sup_{k \ge 1} \sup_{A \in \mathcal{F}_1^k, B \in \mathcal{F}_{k+n}^{+\infty}} |P(AB) - P(A)P(B)|$$

converges to 0 as $n \to \infty$, where $\mathcal{F}_j^k = \sigma\{Z_i, j \le i \le k\}$ denotes the σ-algebra generated by $Z_j, Z_{j+1}, \cdots, Z_k$. There are many practical applications on the α-mixing condition, and more details can be found in [1,2].

Let $F(\cdot)$ and $G(\cdot)$ be continuous distribution functions of Y and T, respectively. For any distribution function $D(\cdot)$, we use $a_D = \inf\{y : D(y) > 0\}$ and $b_D = \sup\{y : D(y) < 1\}$ to denote its lower and upper endpoints, respectively. [11] pointed out that F and G can be completely estimated only when

$$a_G \le a_F, \quad b_G \le b_F, \quad \text{and} \quad \int_{a_F}^{\infty} \frac{\mathrm{d}F}{G} < \infty. \tag{2.1}$$

And the well-known nonparametric maximum likelihood estimators of F and G, which were proposed by [9], are given by

$$F_n(y) = 1 - \prod_{i:Y_i \le y} \Big[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)}\Big] \quad \text{and} \quad G_n(t) = \prod_{i:T_i > t} \Big[\frac{nC_n(T_i) - 1}{nC_n(T_i)}\Big], \tag{2.2}$$

respectively, where $C_n(s) = n^{-1} \sum_{i=1}^{n} I(T_i \le s \le Y_i)$.

Let $f(\cdot,\cdot)$ be the joint density function of (X, Y). In what follows, the star notation (*) relates to any characteristic function of the actually observed data. Due to the independence of T and (X, Y), [10] derived the relationship between $f(\cdot,\cdot)$ and $f^*(\cdot,\cdot)$, which is expressed as

$$f(x, y) = \theta G^{-1}(y) f^*(x, y) \quad \text{for } y > a_G. \tag{2.3}$$

2.2 Estimator

Let $F(\cdot|x)$ and $f(\cdot|x)$ be the conditional distribution and density functions of Y given $X = x$, respectively. It is of interest to estimate the conditional distribution function $F(\cdot|x)$

By conditions (A1)-(A3) in Section 3 and (2.3), we have

$$\begin{aligned}
&\theta E\Big\{K_{h_n}(X - x) G^{-1}(Y) \Big[\widetilde{\Lambda}\Big(\frac{y - Y}{b_n}\Big) - F(y|x)\Big]\Big\} \\
&= \frac{\theta}{h_n} \int_{\mathbb{R}} \int_{\mathbb{R}} K\Big(\frac{s - x}{h_n}\Big) G^{-1}(t) \Big(\widetilde{\Lambda}\Big(\frac{y - t}{b_n}\Big) - F(y|x)\Big) f^*(s, t) \mathrm{d}s\mathrm{d}t \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K(u) \Big[\widetilde{\Lambda}\Big(\frac{y - t}{b_n}\Big) - F(y|x)\Big] f(x + h_n u, t) \mathrm{d}u\mathrm{d}t \\
&= \int_{\mathbb{R}} K(u) f_X(x + h_n u) \mathrm{d}u \int_{\mathbb{R}} \Bigg(\int_{-\infty}^{\frac{y-t}{b_n}} \Lambda(v) \mathrm{d}v - F(y|x)\Bigg) f(t|x + h_n u) \mathrm{d}t \\
&= \int_{\mathbb{R}} K(u) f_X(x + h_n u) \mathrm{d}u \int_{\mathbb{R}} \Lambda(v) \big(F(y - b_n v|x + h_n u) - F(y|x)\big) \mathrm{d}v \\
&\to 0,
\end{aligned} \tag{2.4}$$

where $K(\cdot)$ is a kernel function defined on $\mathbb{R}$, and $\widetilde{\Lambda}(\cdot)$ is a distribution function defined on $\mathbb{R}$, whose density function is $\Lambda(\cdot)$, $K_{h_n}(\cdot) = K(\cdot/h_n)/h_n$, $0 < h_n \to 0$ and $0 < b_n \to 0$ as $n \to \infty$.

Assume that $F(y|s)$ (with respect to s) has $(p + 1)$th continuous derivative at point x. In a small neighborhood of x, one can approximate $F(y|s)$ locally by a pth-order polynomial

$$F(y|s) \approx F(y|x) + \cdots + F^{(p,0)}(y|x) \frac{(s - x)^p}{p!} \equiv \beta_0 + \cdots + \beta_p (s - x)^p,$$

where $F^{(i,j)}(y|x) = \frac{\partial^{(i+j)}}{\partial x^i \partial y^j} F(y|x)$. It follows from (2.4) that $F(y|x)$ can be viewed as an asymptotic nonparametric regression of $\widetilde{\Lambda}\big(\frac{y - Y_i}{b_n}\big)$ on X_i and $G^{-1}(Y_i)$. Then, based on the idea of the local polynomial fitting, the LPDS estimator of $F(y|x)$ is defined as $\widehat{F}_{LPDS}(y|x) = \widehat{\beta}_0$, where

$$\widehat{\beta}=(\widehat{\beta}_0,\cdots,\widehat{\beta}_p)^{\mathrm{T}}=\underset{\beta_j,\,j=0,\cdots,p}{\mathrm{argmin}}\sum_{i=1}^{n}\left(\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big)-\sum_{j=0}^{p}\beta_j(X_i-x)^j\right)^2 K_{h_n}(X_i-x)G_n^{-1}(Y_i). \tag{2.5}$$

As mentioned above, the estimator $\widehat{F}_{LPDS}(y|x)$ is the smoothing estimator of the local polynomial estimation of $\mathbb{E}[I(Y\le y)|\,X=x]$ in [8]. By calculation similar to that of [8] and taking the same notations to those of [8], $\widehat{\beta}$ can be expressed as

$$\widehat{\beta}=\mathbf{H}_n^{-1}\mathbf{S}_n^{-1}\mathbf{t}_n,$$

where $\mathbf{H}_n=\mathrm{diag}(1,h_n,\cdots,h_n^p)$, and

$$\mathbf{S}_n=\begin{pmatrix} s_{n,0} & \cdots & s_{n,p}\\ \vdots & \ddots & \vdots\\ s_{n,p} & \cdots & s_{n,2p}\end{pmatrix},\quad \mathbf{t}_n=\begin{pmatrix} t_{n,0}\\ \vdots\\ t_{n,p}\end{pmatrix},$$

with

$$s_{n,j}=\frac{\theta}{n}\sum_{i=1}^{n}\Big(\frac{X_i-x}{h_n}\Big)^jK_{h_n}(X_i-x)G_n^{-1}(Y_i),$$
$$t_{n,j}=\frac{\theta}{n}\sum_{i=1}^{n}\Big(\frac{X_i-x}{h_n}\Big)^jK_{h_n}(X_i-x)G_n^{-1}(Y_i)\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big).$$

Indeed, $\widehat{\beta}$ can also be expressed as

$$\widehat{\beta}=\frac{\theta}{nh_n}\sum_{i=1}^{n}\mathbf{H}_n^{-1}\mathbf{S}_n^{-1}\Big(1,\frac{X_i-x}{h_n},\cdots,\Big(\frac{X_i-x}{h_n}\Big)^p\Big)^{\tau}K\Big(\frac{X_i-x}{h_n}\Big)G_n^{-1}(Y_i)\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big), \tag{2.6}$$

then $\widehat{F}_{LPDS}(y|x)$ (that is, $\widehat{\beta}_0$) can be written as

$$\widehat{F}_{LPDS}(y|x)=\sum_{i=1}^{n}W_{pi}(x)\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big), \tag{2.7}$$

where

$$W_{pi}(x)=\frac{\theta}{nh_n}e_1^{\mathrm{T}}\mathbf{H}_n^{-1}\mathbf{S}_n^{-1}\Big(1,\frac{X_i-x}{h_n},\cdots,\Big(\frac{X_i-x}{h_n}\Big)^p\Big)^{\mathrm{T}}K\Big(\frac{X_i-x}{h_n}\Big)G_n^{-1}(Y_i),\quad 1\le i\le n, \tag{2.8}$$

and e_1 is a $(p+1)$-dimension vector $(1,0,\cdots,0)^{\mathrm{T}}$. It follows from the expression of $W_{pi}(x)$ that $W_{pi}(x)$ is independent of θ. In addition, it can be shown that

$$\sum_{i=1}^{n}W_{pi}(x)=1. \tag{2.9}$$

Hence, that is, $\widehat{F}_{LPDS}(y|x)$ is a weighted local average of the $\widetilde{\Lambda}\big(\frac{y-Y_i}{b_n}\big)$ values.

Remark 2.1 By (2.5), when $p=0$, $\widehat{F}_{LPDS}(y|x)$ is the kernel estimator proposed by [7] with

$$W_{0i}(x)=\Bigg[\sum_{j=1}^{n}G_n^{-1}(Y_j)K\Big(\frac{X_j-x}{h_n}\Big)\Bigg]^{-1}G_n^{-1}(Y_i)K\Big(\frac{X_i-x}{h_n}\Big).$$

When $p=1$, $\widehat{F}_{LPDS}(y|x)$ is the local linear double-smoothing estimator proposed by [16] with

$$W_{1i}(x)=\big[s_{n,0}s_{n,2}-s_{n,1}^2\big]^{-1}\frac{\theta}{n}G_n^{-1}(Y_i)K_{h_n}(X_i-x)\Big[s_{n,2}-\Big(\frac{X_i-x}{h_n}\Big)s_{n,1}\Big].$$

3 Assumptions and the Main Result

In what follows, we will use $U(x)$ to denote a neighborhood of x, and use C to denote some finite and positive constant, which may change from place to place. Set $\sigma^2(x,y)=\mathbb{E}[(I(Y\le y)-F(y|X))^2G^{-1}(Y)|\,X=x]$.

To establish the asymptotic result, we need the following conditions.

(A0) $a_G<a_F$ and $b_G<b_F$.

(A1) $K(\cdot)$ and $\Lambda(\cdot)$ are both bounded density functions with bounded support on $\mathbb{R}$;

(A2) The density function of X, $f_X(\cdot)$ is continuous at x with $f_X(x)>0$.

(A3) (i) For any $y\in\mathbb{R}$, $F(\cdot|\cdot)$ has continuous second partial derivatives in $U(x)\times U(y)$.

(ii) For any $y\in\mathbb{R}$, $F(y|\cdot)$ has continuous $(p+1)$th derivative in $U(x)$.

(A4) For any $y\in\mathbb{R}$, the function $\sigma^2(\cdot,y)$ is continuous at x with $\sigma^2(x,y)>0$.

(A5) For any integer $j\ge 1$, the joint density function $f_j^*(\cdot,\cdot)$ of (X_1,X_{j+1}) w.r.t P exists on $\mathbb{R}\times\mathbb{R}$ and satisfies $f_j^*(s_1,s_2)\le C$ for $(s_1,s_2)\in U(x)\times U(x)$.

(A6) Assume that $nh_n\to\infty$, and for the sequence $\alpha(n)$ satisfies, there exist positive integers q_n such that $q_n=o((nh_n)^{1/2})$ and $\lim\limits_{n\to\infty}(n(h_n)^{-1})^{1/2}\alpha(q_n)=0$.

Remark 3.1 As [16] pointed out, condition $a_G<a_F$ in (A0) implies that $G(Y)\ge G(a_F)>0$ almost surely, which ensures that $G_n(Y_i)\ne 0$ almost surely, then the resulting estimators can be well defined for enough large n. Conditions (A1)-(A3) are often used in the literature. Condition (A4) is an adaption of condition (A5) of [8] by replacing $\phi(Y)$ with $I(Y\le y)$ here. Condition (A5) is mainly technical. Condition (A6) means that the convergence rates of α-mixing coefficients relate to the selection of bandwidth. In fact, as [8] pointed out, condition (A6) can be easily satisfied. For instance, assume that $\alpha(n)=n^{-\gamma}$ for some γ, and take $h_n=Cn^{-\eta}$

for some $0 < \eta < 1$, $q_n = (nh_n/\log n)^{1/2}$, then condition (A6) holds when $\gamma > (1+\eta)/(1-\eta)$. In addition, if the α-mixing coefficients decay exponentially, then condition (A6) is redundant.

Take the same notations to those of [8] again. Set

$$u_j = \int_{\mathbb{R}} t^j K(t)\mathrm{d}t, \quad v_j = \int_{\mathbb{R}} t^j K^2(t)\mathrm{d}t, \quad \widetilde{\Lambda}_2 = \int t^2 \Lambda(t)\mathrm{d}t$$

and

$$\mathbf{S} = \begin{pmatrix} u_0 & \cdots & u_p \\ \vdots & \ddots & \vdots \\ u_p & \cdots & u_{2p} \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} v_0 & \cdots & v_p \\ \vdots & \ddots & \vdots \\ v_p & \cdots & v_{2p} \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} u_{p+1} \\ \vdots \\ u_{2p+1} \end{pmatrix}.$$

Theorem 3.1 *Let $\alpha(n) = O(n^{-\gamma})$ for some $\gamma > 3$. Suppose that conditions* (A0)-(A6) *hold, then*

$$\sqrt{nh_n}\big\{\widehat{F}_{LPDS}(y|x) - F(y|x) - B(x,y)\big\} \xrightarrow{D} N(0, V_F(x,y)), \tag{3.1}$$

where

$$B(x,y) = \frac{1}{2}\widetilde{\Lambda}_2 F^{(0,2)}(y|x)b_n^2 + o_P(b_n^2) + \frac{1}{(p+1)!}F^{(p+1,0)}(y|x)e_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{U}h_n^{p+1} + o_P(h_n^{p+1}),$$
$$V_F(x,y) = \theta f_X^{-1}(x)\sigma^2(x,y)e_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}e_1.$$

Remark 3.2 Theorem 3.1 extends Lemma 1 of [14] to the left-truncation model. If there exists no left-truncation, then $\theta = 1$ and $G(y) = 1$ for $y > a_F$. In this situation, $\sigma^2(x,y) = F(y|x)[1 - F(y|x)]$, thus Theorem 3.1 reduces to Lemma 1 of [14]. As [14] pointed out, higher-order local polynomial double-smoothing reduces the bias in the X_i direction but not the one in the Y_i direction.

Remark 3.3 When $p = 1$, the estimator $\widehat{F}_{LPDS}(y|x)$ reduces to the local linear double-smoothing (LLDS) estimator $\widehat{F}_{LLDS}(y|x)$ proposed by [16]. In this case,

$$V_F(x,y) = \theta f_X^{-1}(x)\sigma^2(x,y)\int_{\mathbb{R}} K^2(t)\mathrm{d}t$$

and

$$B(x,y) = \frac{1}{2}u_2 F^{(2,0)}(y|x)h_n^2 + \frac{1}{2}\widetilde{\Lambda}_2 F^{(0,2)}(y|x)b_n^2 + o_P(h_n^2 + b_n^2).$$

It follows from Theorem 3.1 that

$$\sqrt{nh_n}\Big\{\widehat{F}_{LLDS}(y|x) - F(y|x) - \frac{1}{2}u_2F^{(2,0)}(y|x)h_n^2 - \frac{1}{2}\widetilde{\Lambda}_2F^{(0,2)}(y|x)b_n^2 + o_P(h_n^2 + b_n^2)\Big\}$$
$$\xrightarrow{D} N\left(0, \theta f_X^{-1}(x)\sigma^2(x,y)\int_{\mathbb{R}} K^2(t)\mathrm{d}t\right), \tag{3.2}$$

which extends Theorem 3.1 of [16] to the dependent case.

4 Simulation

In this section, a simulation is carried out to investigate the finite sample performance of the LPDS estimator of the conditional distribution function. In particular, we compare the mean squared errors (MSE) of the local constant single-smoothing (LCSS) estimator, the local constant double-smoothing (LCDS) estimator, the local linear single-smoothing (LLSS) estimator and the the local linear double-smoothing (LLDS) estimator of the conditional distribution function. First we get an α-mixing observed sequence $\{(X_i, Y_i, T_i) | 1 \le i \le n\}$ after left-truncation. We generate the sequence by the same approach used by [13], and details are as follows:

(1) Generate the observed sample (X_1, Y_1, T_1).

Step 1 Simulate $e_1 \sim N(0,1)$ and take $X_1 = e_1$;

Step 2 Simulate $\epsilon_1 \sim N(0,1)$, and take $Y_1 = 2.5 + \sin(2X_1) + 2\exp\{-16X_1^2\} + 0.5\epsilon_1$;

Step 3 Simulate $T_1 \sim N(\mu, 1)$, where μ is adapted to get different values of θ. If $Y_1 < T_1$, reject the datum (X_1, Y_1, T_1) and return to Step 2, continue like this until $Y_1 \ge T_1$. Thus, we can get the observed sample (X_1, Y_1, T_1).

(2) Generate the observed sample (X_2, Y_2, T_2).

Step 4 Simulate $e_2 \sim N(0,1)$ and take $X_2 = 0.5X_1 + e_2$, which is an AR(1) model;

Step 5 Simulate $\epsilon_2 \sim N(0,1)$, and take $Y_2 = 2.5 + \sin(2X_2) + 2\exp\{-16X_2^2\} + 0.5\epsilon_2$;

Step 6 Simulate $T_2 \sim N(\mu, 1)$. If $Y_2 < T_2$, reject the datum (X_2, Y_2, T_2) and return to Step 5, continue like this until $Y_2 \ge T_2$. Thus, we can get the observed sample (X_2, Y_2, T_2).

(3) Repeating the process (2), we can generate the observed data $\{(X_i, Y_i, T_i),\ i = 1, \cdots, n\}$.

It follows from the heredity of the α-mixing property and the fact that the (X_i) is α-mixing that the (X_i, Y_i, T_i) is α-mixing. And the (X_i, Y_i, T_i) satisfies that $X_i = 0.5X_{i-1} + e_i$, $Y_i = 2.5 + \sin(2X_i) + 2\exp\{-16X_i^2\} + 0.5\epsilon_i$ and $Y_i \ge T_i$, where $e_i \sim N(0,1)$, $\epsilon_i \sim N(0,1)$, $T_i \sim N(\mu, 1)$, and μ is adapted to get different values of θ. Then the conditional density function is

$$f(y|x) = \frac{1}{0.5\sqrt{2\pi}} \exp\Big\{ - \frac{(y - 2.5 - \sin(2x) - 2\exp\{-16x^2\})^2}{2 \times 0.5^2} \Big\},$$

and the conditional distribution function is

$$F(y|x) = \int_{-\infty}^{y} f(s|x)\mathrm{d}s = \Phi\Big(\frac{y - 2.5 - \sin(2x) - 2\exp\{-16x^2\}}{0.5} \Big),$$

where $\Phi(\cdot)$ is the distribution function of $N(0,1)$.

For any estimator $\widehat{F}(\cdot|\cdot)$ among $\widehat{F}_{LCSS}(\cdot|\cdot)$, $\widehat{F}_{LCDS}(\cdot|\cdot)$, $\widehat{F}_{LLSS}(\cdot|\cdot)$ and $\widehat{F}_{LLDS}(\cdot|\cdot)$, the MSE of $\widehat{F}(\cdot|\cdot)$ is given by

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{F}(Y_i|X_i) - F(Y_i|X_i)\right]^2.$$

In the following simulation, we take $K(s) = \Lambda(s) = \frac{3}{4}(1-s^2)I(|s| \le 1)$. The bandwidths h_n and b_n will be selected by the data-driven least-square cross-validation method proposed by [17]. That is, h_n and b_n can be determined by minimizing the following criterion:

$$CV(h_n, b_n) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}\left[I(Y_i \le Y_j) - \widehat{F}_{-i}(Y_j|X_i)\right]^2,$$

where $\widehat{F}_{-i}(Y_j|X_i)$ is the leave-one-out estimator of $F(Y_j|X_i)$, which can be obtained by using the data points $\{(X_k, Y_k, T_k) : 1 \le k \le n,\ k \neq i\}$ to compute $\widehat{F}(Y_j|X_i)$.

For different values of the percentage of truncated data: $\theta \approx 30\%$, 60% and 90%, we generate the observed data $\{(X_i, Y_i, T_i),\ i = 1, \cdots, n\}$ with the sample size $n = 100$ and 200, respectively. The simulation results are reported in Table 1. The quantity in Table 1 is the average MSE based on 500 replications.

Table 1: The average MSE of $\widehat{F}_{LCSS}(\cdot|\cdot)$, $\widehat{F}_{LCDS}(\cdot|\cdot)$, $\widehat{F}_{LLSS}(\cdot|\cdot)$ and $\widehat{F}_{LLDS}(\cdot|\cdot)$.

$\theta(\%)$	n	$\widehat{F}_{LCSS}(\cdot\|\cdot)$,	$\widehat{F}_{LCDS}(\cdot\|\cdot)$	$\widehat{F}_{LLSS}(\cdot\|\cdot)$	$\widehat{F}_{LLDS}(\cdot\|\cdot)$
30	100	0.7253	0.6905	0.6832	0.6548
	200	0.6357	0.6258	0.6081	0.5802
60	100	0.6719	0.6569	0.6411	0.6204
	200	0.5721	0.5503	0.5477	0.5226
90	100	0.5814	0.5601	0.5487	0.5275
	200	0.4736	0.4511	0.4432	0.4206

Table 1 shows that: (i) In all cases, the LLDS estimator outperforms the LCSS, LCDS and LLSS estimators. In each case, the local linear estimation performs better than the local constant estimation, and the double-smoothing estimation performs better than the single-smoothing estimation. (ii) The MSE decrease as n or θ increases. It is just as we expect, since the sampling information becomes more when n or θ increases.

5 Proof of the Main Result

Lemma 5.1(Davydov's lemma) *Suppose that X and Y are random variables such that $E|X|^p < \infty$, $E|Y|^q < \infty$, where $p, q > 1$, $p^{-1} + q^{-1} < 1$. Then*

$$
\begin{aligned}
&|EXY - EXEY| \\
&\le 8(E|X|^p)^{1/p}(E|X|^q)^{1/q}\Big(\sup_{A\in\sigma(X),B\in\sigma(Y)}|P(AB)-P(A)P(B)|\Big)^{1-(p^{-1}+q^{-1})}.
\end{aligned}
$$

Lemma 5.2[8] *Suppose that* $\alpha(n) = O(n^{-\gamma})$ *for some* $\gamma > 3$. *Then under* (A0) *we have* $\sup\limits_{y\ge a_F}|G_n(y) - G_(y)| = O_p(n^{-1/2})$.

Lemma 5.3 *Suppose that conditions* (A1) *and* (A3)(i) *hold. If* $|X_i - x| \le Ch_n$, *then*

$$\mathbb{E}\Big[\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big)\Big|X_i\Big] = F(y|X_i) + \frac{1}{2}\widetilde{\Lambda}_2 F^{(0,2)}(y|X_i)b_n^2 + o(b_n^2), \tag{5.1}$$

$$\mathbb{E}\Big[\widetilde{\Lambda}^2\Big(\frac{y-Y_i}{b_n}\Big)\Big|X_i\Big] = F(y|X_i) - 2\int_{\mathbb{R}} t\widetilde{\Lambda}(t)\Lambda(t)\mathrm{d}tF^{(0,1)}(y|X_i)b_n + O(b_n^2), \tag{5.2}$$

$$\mathbb{E}\Big[\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big)I(Y_i\le y)\Big|X_i\Big] = F(y|X_i) - b_nF^{(0,1)}(y|X_i)\int_0^{+\infty} t\widetilde{\Lambda}(t)\mathrm{d}t + O(b_n^2). \tag{5.3}$$

(5.1) and (5.2) follow from Lemma A.5 of [6], and (5.3) can be easily shown by the similar proof approach used in Lemma A.5 of [6].

Lemma 5.4 *Under the conditions of Theorem* 3.1, *we have*

$$\sqrt{nh_n}\Big\{\sum_{i=1}^n W_{pi}(x)I(Y_i\le y) - F(y|x) - \frac{1}{(p+1)!}F^{(p+1,0)}(y|x)e_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{U}h_n^{p+1} + o_P(h_n^{p+1})\Big\}$$
$$\xrightarrow{D} N(0,\, V_F(x,y)).$$

By taking $\Phi(Y) = I(Y \le y)$ and comparing the conditions of Lemma 5.4 with the conditions of Theorem 3.2 of [8], Lemma 5.4 follows from Theorem 3.2 of [8].

Now consider the proof of Theorem 3.1, which is similar to that of Lemma 1 of [14].

Proof of Theorem 3.1 It follows from (2.7) that

$$
\begin{aligned}
\sqrt{nh_n}\big(\widehat{F}_{LPDS}(y|x) - F(y|x)\big) &= \sqrt{nh_n}\Big[\sum_{i=1}^n W_{pi}(x)\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big) - F(y|x)\Big] \\
&= \sqrt{nh_n}\sum_{i=1}^n W_{pi}(x)\Big[\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big) - I(Y_i\le y)\Big] \\
&\quad + \sqrt{nh_n}\Big[\sum_{i=1}^n W_{pi}(x)I(Y_i\le y) - F(y|x)\Big] \\
&:= \sqrt{nh_n}A_n + \sqrt{nh_n}B_n.
\end{aligned} \tag{5.4}
$$

From (5.4) and Lemma 5.4, Theorem 3.1 follows if we can prove that

$$A_n = \frac{1}{2}b_n^2F^{(0,2)}(y|x)\widetilde{\Lambda}_2 + o(b_n^2) + o_P\Big(\frac{1}{\sqrt{nh_n}}\Big) + O_P(n^{-1/2}). \tag{5.5}$$

It follows from Lemma 5.1 of [8] that

$$\mathbf{S}_n \xrightarrow{P} f_X(x)\mathbf{S},$$

then $W_{pi}(x)$ can be written as

$$W_{pi}(x) = [1 + o_P(1)]f_X^{-1}(x)\frac{\theta}{nh_n}G_n^{-1}(Y_i)K^*\Big(\frac{X_i - x}{h_n}\Big), \quad 1 \le i \le n, \tag{5.6}$$

where $K^*(u) = e_1^{\mathrm{T}}\mathbf{S}^{-1}(1, u, \cdots, u^p)^{\mathrm{T}}K(u)$ and the $o_P(1)$ part does not depend on u. It follows from [3, p.237-238] that the function $K^*(\cdot)$ satisfies

$$\int_{\mathbb{R}} K^*(u)\mathrm{d}u = 1 \quad \text{and} \quad \int_{\mathbb{R}} [K^*(u)]^2\mathrm{d}u = e_1^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{V}\mathbf{S}^{-1}e_1. \tag{5.7}$$

Then, A_n can be written as

$$\begin{aligned} A_n &= [1 + o_P(1)]f_X^{-1}(x)\frac{\theta}{nh_n}\sum_{i=1}^n G_n^{-1}(Y_i)K^*\Big(\frac{X_i - x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y - Y_i}{b_n}\Big) - I(Y_i \le y)\Big] \\ &:= [1 + o_p(1)]f_X^{-1}(x)Q_n. \end{aligned} \tag{5.8}$$

From (5.8), (5.5) holds if we can prove that

$$Q_n = \frac{1}{2}f_X(x)\widetilde{\Lambda}_2F^{(0,2)}(y|x)b_n^2 + o(b_n^2) + o_P\Big(\frac{1}{\sqrt{nh_n}}\Big) + O_P(n^{-1/2}). \tag{5.9}$$

Note that,

$$\begin{aligned} Q_n &= \frac{\theta}{nh_n}\sum_{i=1}^n G^{-1}(Y_i)K^*\Big(\frac{X_i - x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y - Y_i}{b_n}\Big) - I(Y_i \le y)\Big] \\ &\quad + \frac{\theta}{nh_n}\sum_{i=1}^n \big(G_n^{-1}(Y_i) - G^{-1}(Y_i)\big)K^*\Big(\frac{X_i - x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y - Y_i}{b_n}\Big) - I(Y_i \le y)\Big] \\ &:= Q_{1n} + Q_{2n}. \end{aligned} \tag{5.10}$$

First we consider the order of Q_{2n}. It follows from (A0) and Lemma 5.2 that

$$\begin{aligned} |Q_{2n}| &\le \frac{\sup_{y \ge a_F}|G_n(y) - G(y)|}{G(a_F) - \sup_{y \ge a_F}|G_n(y) - G(y)|}\frac{\theta}{nh_n}\sum_{i=1}^n G^{-1}(Y_i)K^*\Big(\frac{X_i - x}{h_n}\Big) \\ &= O_p(n^{-1/2})\frac{\theta}{nh_n}\sum_{i=1}^n G^{-1}(Y_i)K^*\Big(\frac{X_i - x}{h_n}\Big). \end{aligned} \tag{5.11}$$

By (2.3), (A1), (A2) and the expression of the function $K^*(\cdot)$,

$$E\left[\frac{\theta}{nh_n}\sum_{i=1}^{n}G^{-1}(Y_i)K^*\Big(\frac{X_i-x}{h_n}\Big)\right]=\frac{1}{h_n}\int_{\mathbb{R}}\int_{\mathbb{R}}K^*\Big(\frac{s-x}{h_n}\Big)f(s,t)\mathrm{d}s\mathrm{d}t$$
$$=\int_{\mathbb{R}}K^*(u)f_X(x+uh_n)\mathrm{d}v=O(1),$$

which implies that

$$\frac{\theta}{nh_n}\sum_{i=1}^{n}G^{-1}(Y_i)K^*\Big(\frac{X_i-x}{h_n}\Big)=O_P(1).$$

Again by (5.11), we have

$$Q_{2n}=O_P(n^{-1/2}). \tag{5.12}$$

From (5.9), (5.10) and (5.12), (5.9) holds if one can prove that

$$E[Q_{1n}]=\frac{1}{2}f_X(x)\widetilde{\Lambda}_2F^{(0,2)}(y|x)b_n^2+o(b_n^2), \tag{5.13}$$
$$Var[Q_{1n}]=o\Big(\frac{1}{nh_n}\Big). \tag{5.14}$$

Let

$$Z_i=\frac{\theta}{h_n}G^{-1}(Y_i)K^*\Big(\frac{X_i-x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y-Y_i}{b_n}\Big)-I(Y_i\le y)\Big],\quad 1\le i\le n.$$

Then

$$Q_{1n}=\frac{1}{n}\sum_{i=1}^{n}Z_i.$$

By stationarity of $\{X_i,Y_i\}$,

$$E[Q_{1n}]=E[Z_1], \tag{5.15}$$
$$Var[Q_{1n}]=\frac{1}{n}Var(Z_1)+\frac{2}{n}\sum_{k=1}^{n-1}\Big(1-\frac{k}{n}\Big)Cov(Z_1,Z_{k+1}). \tag{5.16}$$

To calculate $E[Z_1]$. It follows from (2.3), (5.1), (A2), (A3)(i) , (A0), and (5.7) that

$$\begin{aligned}E[Z_1]&=\frac{1}{h_n}\int_{\mathbb{R}}\int_{\mathbb{R}}K^*\Big(\frac{s-x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y-t}{b_n}\Big)-I(t\le y)\Big]f(s,t)\mathrm{d}s\mathrm{d}t\\
&=\mathbb{E}\Big\{\frac{1}{h_n}K^*\Big(\frac{X_1-x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y-Y_1}{b_n}\Big)-I(Y_1\le y)\Big]\Big\}\\
&=\mathbb{E}\Big\{\frac{1}{h_n}K^*\Big(\frac{X_1-x}{h_n}\Big)\mathbb{E}\Big[\widetilde{\Lambda}\Big(\frac{y-Y_1}{b_n}\Big)-I(Y_1\le y)\Big|X_1\Big]\Big\}\\
&=\mathbb{E}\Big\{\frac{1}{h_n}K^*\Big(\frac{X_1-x}{h_n}\Big)\Big[\frac{1}{2}\widetilde{\Lambda}_2F^{(0,2)}(y|X_1)b_n^2+o(b_n^2)\Big]\Big\}\\
&=\frac{1}{2}b_n^2\widetilde{\Lambda}_2\int_{\mathbb{R}}K^*(t)F^{(0,2)}(y|x+h_nt)f_X(x+h_nt)\mathrm{d}t+o(b_n^2)\int_{\mathbb{R}}K^*(t)f_X(x+h_nt)\mathrm{d}t\\
&=\frac{1}{2}f_X(x)\widetilde{\Lambda}_2F^{(0,2)}(y|x)b_n^2+o(b_n^2).\end{aligned} \tag{5.17}$$

Then (5.13) follows from (5.15) and (5.17).

To calculate $Var(Z_1)$ and $Cov(Z_1, Z_{k+1})$, from (5.17), (2.3), (A0), (5.2), (5.3), (5.7) and (A2),

$$\begin{aligned}
Var(Z_1) &= E[Z_1^2] + O(b_n^4)\\
&= \mathbb{E}\Big\{\theta G^{-1}(Y_1)\frac{1}{h_n^2}K^{*2}\Big(\frac{X_1-x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y-Y_1}{b_n}\Big) - I(Y_1\le y)\Big]^2\Big\} + O(b_n^4)\\
&\le \theta G^{-1}(a_F)\mathbb{E}\Big\{\frac{1}{h_n^2}K^{*2}\Big(\frac{X_1-x}{h_n}\Big)E\Big[\Big(\widetilde{\Lambda}\Big(\frac{y-Y_1}{b_n}\Big)-I(Y_1\le y)\Big)^2\Big|X_1\Big]\Big\}+O(b_n^4)\\
&= \theta G^{-1}(a_F)\mathbb{E}\Big[\frac{1}{h_n^2}K^{*2}\Big(\frac{X_1-x}{h_n}\Big)O(b_n)\Big] + O(b_n^4)\\
&= \frac{O(b_n)}{h_n}\int_{\mathbb{R}} K^{*2}(t)f_X(x+h_nt)\mathrm{d}t + O(b_n^4)\\
&= O\Big(\frac{b_n}{h_n}\Big). \qquad (5.18)
\end{aligned}$$

By (A0), (A2) and (A5),

$$\begin{aligned}
|Cov(Z_1, Z_{k+1})| &\le E|Z_1Z_{k+1}| + (E|Z_1|)^2\\
&= E\Big[\frac{\theta^2}{h_n^2}G^{-1}(Y_1)G^{-1}(Y_{k+1})\Big|K^*\Big(\frac{X_1-x}{h_n}\Big)\Big|\Big|K^*\Big(\frac{X_{k+1}-x}{h_n}\Big)\Big|\\
&\quad\cdot\Big|\widetilde{\Lambda}\Big(\frac{y-Y_1}{b_n}\Big)-I(Y_1\le y)\Big|\Big|\widetilde{\Lambda}\Big(\frac{y-Y_{k+1}}{b_n}\Big)-I(Y_{k+1}\le y)\Big|\Big]\\
&\quad+E^2\Big[\frac{\theta}{h_n}G^{-1}(Y_1)\Big|K^*\Big(\frac{X_1-x}{h_n}\Big)\Big|\Big|\widetilde{\Lambda}\Big(\frac{y-Y_1}{b_n}\Big)-I(Y_1\le y)\Big|\Big]\\
&\le \frac{4\theta^2}{G^2(a_F)}E\Big[\frac{1}{h_n^2}\Big|K^*\Big(\frac{X_1-x}{h_n}\Big)\Big|\Big|K^*\Big(\frac{X_{k+1}-x}{h_n}\Big)\Big|\Big]\\
&\quad+4E^2\Big[\frac{\theta}{h_n}G^{-1}(Y_1)\Big|K^*\Big(\frac{X_1-x}{h_n}\Big)\Big|\Big]\\
&= 4\theta^2G^{-2}(a_F)\int_{\mathbb{R}}\int_{\mathbb{R}}|K^*(u)||K^*(v)|f_k^*(x+h_nu, x+h_nv)\mathrm{d}u\mathrm{d}v\\
&\quad+4\Big[\int_{\mathbb{R}}|K^*(u)|f_X(x+h_nu)\mathrm{d}u\Big]^2\\
&= O(1). \qquad (5.19)
\end{aligned}$$

On the other hand, from (2.3), (A0), (A1) and (A2),

$$\begin{aligned}
E|Z_1|^{2\gamma} &= \mathbb{E}\theta^{2\gamma-1}G^{1-2\gamma}(Y_1)h_n^{-2\gamma}\Big|K^*\Big(\frac{X_1-x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y-Y_1}{b_n}\Big)-I(Y_1\le y)\Big]\Big|^{2\gamma}\\
&\le \theta^{2\gamma-1}G^{1-2\gamma}(a_F)h_n^{-2\gamma}\int_{\mathbb{R}}\Big|K^*\Big(\frac{s-x}{h_n}\Big)\Big[\widetilde{\Lambda}\Big(\frac{y-t}{b_n}\Big)-I(t\le y)\Big]\Big|^{2\gamma}f(s,t)\mathrm{d}s\mathrm{d}t\\
&\le 2^{2\gamma}\theta^{2\gamma-1}G^{1-2\gamma}(a_F)h_n^{-2\gamma}\int_{\mathbb{R}}\Big|K^*\Big(\frac{s-x}{h_n}\Big)\Big|^{2\gamma}f_X(s)\mathrm{d}s
\end{aligned}$$

$$
\begin{aligned}
&= 2^{2\gamma}\theta^{2\gamma-1}G^{1-2\gamma}(a_F)h_n^{1-2\gamma}\int_{\mathbb{R}}\Big|K^*(u)\Big|^{2\gamma}f_X(x+h_nu)\mathrm{d}u \\
&= O(h_n^{1-2\gamma}),
\end{aligned}
$$

then it follows from Lemma 5.1 and $\alpha(k) = O(k^{-\gamma})$ that

$$
|Cov(Z_1, Z_{k+1})| \le C[\alpha(k)]^{1-\frac{1}{\gamma}}(E|Z_1|^{2\gamma})^{\frac{1}{\gamma}} \le Ck^{-\gamma+1}h_n^{\frac{1}{\gamma}-2}. \tag{5.20}
$$

Let $d_n = [h_n^{-(1-1/\gamma)/\eta}]$ for $1 - \frac{1}{\gamma} < \eta < \gamma - 2$. By (5.19) and (5.20),

$$
\begin{aligned}
h_n\sum_{k=1}^{n-1}|Cov(Z_1, Z_{k+1})| &= O(1) + \left(\sum_{k=1}^{d_n} + \sum_{k=d_n+1}^{n-1}\right)\min\left\{h_n, k^{-\gamma+1}h_n^{-1+\frac{1}{\gamma}}\right\} \\
&= O(1)\left(d_nh_n + h_n^{-1+\frac{1}{\gamma}}\sum_{k=d_n+1}^{\infty}k^{-\gamma+1}\right) \\
&\le O(1)(d_nh_n + d_n^{-\gamma+2}h_n^{-1+\frac{1}{\gamma}}) \longrightarrow 0.
\end{aligned} \tag{5.21}
$$

Then (5.14) follows from (5.16), (5.18) and (5.21). The proof is completed.

References

[1] J. Cai, J. Kim, Nonparametric quantile estimation with correlated failure time data, *Lifetime Data Analysis*, **9**(2003),357-371.

[2] P. Doukhan, Mixing: Properties and Examples, Springer-Verlag, Berlin, 1994.

[3] J. Fan, Q. Yao, Nonlinear Time Series: Nonparametric and Parametric Methods, Springer-Verlag, New York, 2003.

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

[5] B.E. Hansen, Nonparametric estimation of smooth conditional distributions, Unpublished paper: Department of Economics, University of Wisconsin, 2004.

[6] Q. Li, J.S Racine, Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data, *Journal of Business & Economic Statistics*, **26**(2008),423-434.

[7] M. Lemdani, E. Ould-Said, N. Poulin, Asymptotic properties of a conditional quantile estimator with randomly truncated data, *Journal of Multivariate Analysis*, **100**(2009), 546-559.

[8] H.Y. Liang, J. de Uña-Álvarez, M.C. Iglesias-Pérez, Local polynomial estimation of a conditional mean function with dependent truncated data, *Test*, **20**(2011),653-677.

[9] D. Lynden-Bell, A method of allowing for known observational selection in small samples applied to 3CR quasars, *Monthly Notices of the Royal Astronomical Society*, **155**(1971),95-118.

[10] E. Ould-Saïd, M. Lemdani, Asymptotic properties of a nonparametric regression function estimator with randomly truncated data, *Annals of the Institute of Statistical Mathematics*, **58**(2006),357-378.

[11] M. Woodroofe, Estimating a distribution function with truncated data, *Annals of Statistics*, **13**(1985),163-177.

[12] M.C. Wang, N.P. Jewell, W.Y. Tsai, Asymptotic properties of the product limit estimate under random truncation, *Annals of Statistics*, **14**:4(1986),1597-1605.

[13] J.F. Wang, H.Y. Liang, G.L. Fan, Asymptotic properties of conditional quantile estimator under left-truncated and -mixing conditions, *Communications in Statistics-Theory and Methods*, **40**(2011),2462-2486.

[14] K.L. Xu, Nonparametric inference for conditional quantiles of time series, *Econometric Theory*, **29**(2013),673-698.

[15] K. Yu, M.C. Jones, Local linear quantile regression, *Journal of the American statistical Association*, **93**(1998),228-237.

[16] M. Yao, J.F. Wang, L. Lin, Asymptotic normality for a nonparametric estimator of conditional quantile with left-truncated data, *Communications in Statistics-Theory and Methods*, **46**(2017),6280-6292.

[17] Q. Li, J. Lin, J.S. Racine, Optimal bandwidth selection for nonparametric conditional distribution and quantile functions, *Jounnal of Business & Economic Statistics*, **31**(2013),57-65.

(*edited by Mengxin He*)