Strong Convergence Rates of Wavelet Estimators in Semiparametric Regression Models with Censored Data*

Hongchang Hu*

School of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, China

Abstract: The paper studies a semiparametric regression model

$$Y_i = X_i\beta + g(T_i) + e_i, i = 1, 2, \ldots, n.$$

where Y_i is censored on the right by another random variable C_i with known or unknown distribution G. Firstly, the wavelet estimators of parameter β and nonparameter $g(t)$ are given by wavelet smoothing and the synthetic data methods. Secondly, under general conditions, the laws of the iterated logarithm for the wavelet estimators of parameter and strong uniform convergence rates for the wavelet estimators of nonparameter are investigated. Lastly, the validity of method is illuminated by the simulation example.

Keywords: Semiparametric regression model, Wavelet estimate, Censored data, Law of the iterated logarithm, Strong uniform convergence rate.

1. INTRODUCTION

Consider the semiparametric regression model

$$Y_i = X_i\beta + g(T_i) + e_i, \quad i = 1, 2, \ldots, n. \tag{1}$$

where Y_i's are scalar response variables, X_i's and T_i's are explanatory variables, β is a one-dimensional unknown parameter, $g(\cdot)$ is an unknown regression function on $[0,1]$, and e_i's are independent and identically distributed random errors with zero mean and finite variance σ^2.

Following Speckman (See [1]), denote

$$X_i = f(T_i) + \eta_i, 1 \le i \le n. \tag{2}$$

where $f(\cdot)$ is some unknown smooth function on $[0,1]$, η_i's are independent and identically distributed random errors with zero mean and finite variance σ_η^2 and independent of Y_i's and T_i's.

Since the introductory work by Engle *et al.* (See [2]), the model (1) has been widely studied (See [1,3-6] and references therein) and put into use in many fields of applied statistics. A partial list of estimation methods for β and $g(\cdot)$ includes penalized least squares method (See [7]), smoothing splines method (See [8]), piecewise polynomial method (See [3]), near neighbor method (See [9]) and wavelet method (See [10-13]), etc.

In practice, particularly in medical studies, Y_i may be censored randomly on the right by some censoring variable $C_i, i = 1, 2, \ldots, n$, and hence cannot be observed completely. One only observes $\{X_i, T_i, Z_i, \delta_i\}, i = 1, 2, \ldots, n$, where

$$Z_i = \min(Y_i, C_i), \ \delta_i = I(Y_i \le C_i),$$

C_i are assumed to be independent of Y_i's and η_i's. For right censored data, the model (1) has been studied by Wang and Li [14] and Wang and Zheng [15], and the model (1)-(2) has been investigated by Qin and Cai [16], Qin and Jing [17], Pan and Fu [18] and Liang and Zhou [19].

With wavelet method, the paper discusses the model (1)-(2). The organization of the paper is as follows. In section 2 the wavelet estimators of parameter β and nonparameter $g(t)$ are given by wavelet smoothing and the synthetic data methods. Under general conditions, the laws of the iterated logarithm for the wavelet estimators of parameter and strong uniform convergence rates for the wavelet estimators of nonparameter are investigated in section 3. The main proofs are presented in section 4, with the simulation example in section 5.

2. ESTIMATION METHOD

Let $F(\cdot)$ and $G(\cdot)$ denote the distributions for Y_1 and C_1, respectively. Assume that $G(\cdot)$ is continuous and known at the moment. Using the synthetic data method (See [16,17,20], and references therein), we transform the censored data $\{(Z_i, \delta_i), i = 1, 2, \ldots, n\}$ into the following synthetic data

$$Y_i^* = \phi_1(Z_i; G)\delta_i + \phi_2(Z_i; G)(1 - \delta_i) \tag{3}$$

where (ϕ_1, ϕ_2) are continuous functions which satisfy

(i) $\quad (1 - G(Y))\phi_1(Y; G) + \int_{-\infty}^{Y} \phi_2(t; G)\, dG(t) = Y,$

(ii) $\quad \phi_1$ and ϕ_2 do not depend on F.

The class of all pairs (ϕ_1, ϕ_2) of such functions will be denoted by K. It can be easily showed that, if $(\phi_1, \phi_2) \in K$, then

*Address correspondence to this author at the School of Mathematics and Statistics, Hubei Normal University, Huangshi 435002, China; E-mail: retutome@163.com

1874-1142/08 2008 Bentham Science Publishers Ltd.

$$E\left(Y_i^*\middle|\mathbf{X},\mathbf{T}\right)=E\left(Y_i\middle|\mathbf{X},\mathbf{T}\right)\quad\text{and}\quad E\left(Y_i^*\right)=E\left(Y_i\right), \tag{4}$$

where $\mathbf{X}=\{X_1,\cdots,X_n\}$, $\mathbf{T}=\{T_1,\cdots,T_n\}$. The class K includes many interesting cases (See [19-20]).

If $G(\cdot)$ is unknown, we use a modified version $\tilde G_n(t)$ of the Kaplan-Meier estimator of $G(t)$ defined by

$$\tilde G_n(t)=\begin{cases}\hat G_n(t),\text{if} & t\le Z_{(n)}\quad\text{or}\quad\delta_{(n)}=0\\ \hat G_n\left(Z_{(n)}\right),\text{if} & t>Z_{(n)}\quad\text{and}\quad\delta_{(n)}=1\end{cases}$$

Where

$$\hat G_n(t)=1-\prod_{j=1}^n\left(N^+\left(Z_j\right)\Big/\left[1+N^+\left(Z_j\right)\right]\right)^{I\left(Z_j\le t,\delta_j=0\right)},$$

$$N^+(t)=\sum_{i=1}^n I\left(Z_i>t\right),$$

$Z_{(n)}=\max_{1\le i\le n}\{Z_i\}$ and $\delta_{(n)}$ is the corresponding δ_i. It is easily showed that

$$\tilde G_n(t)\le 1-(n+1)^{-1}<1$$

for all t. Substituting $\tilde G_n(\cdot)$ for $G(\cdot)$ in (3), we obtain the following synthetic data:

$$\hat Y_i^*=\phi_1\left(Z_i;\tilde G_n\right)\delta_i+\phi_2\left(Z_i;\tilde G_n\right)\left(1-\delta_i\right) \tag{5}$$

Suppose that there exists a scaling function $\phi(x)$ in the Schwartz space S_l and a multiresolution analysis $\{V_m\}$ in the concomitant Hilbert space $L^2(R)$, with its reproducing kernel $E_m(t,s)$ given by

$$E_m(t,s)=2^m E_0\left(2^m t,2^m s\right)=2^m\sum_{k\in Z}\phi\left(2^m t-k\right)\phi\left(2^m s-k\right)$$

where Z denotes the collection of integers. Let $A_i=[s_{i-1},s_i]$ be intervals that partition [0,1] with $t_i\in A_i$ and $1\le i\le n$. The estimate method will be introduced as following:

Firstly, suppose that β is known, we define estimator of $g(\cdot)$ by

$$g_n(t,\beta)=\sum_{i=1}^n\left(Y_i^*-X_i\beta\right)\int_{A_i}E_m(t,s)ds\cdot$$

In succession, we define wavelet estimator β_n^* by minimizing

$$\beta_n^*=\arg\min_\beta\left\{\sum_{i=1}^n\left[Y_i^*-X_i\beta-g_n\left(t_i,\beta\right)\right]^2\right\}=\sum_{i=1}^n\tilde X_i\tilde Y_i^*\Big/\sum_{i=1}^n\tilde X_i^2\cdot$$

where

$$\tilde X_i=X_i-\sum_{j=1}^n\int_{A_j}E_m\left(t_i,s\right)dsX_j,\ \tilde Y_i^*=Y_i^*-\sum_{j=1}^n\int_{A_j}E_m\left(t_i,s\right)dsY_i^*\cdot$$

Finally, we define linear wavelet estimator of $g(\cdot)$ by

$$g_n^*(t)=g_n\left(t,\beta_n^*\right)=\sum_{i=1}^n\left(Y_i^*-X_i\beta_n^*\right)\int_{A_i}E_m(t,s)ds$$

Using the similar procedure, if $G(\cdot)$ is unknown, then we define estimators of β and $g(\cdot)$ by

$$\hat\beta_n^*=\sum_{i=1}^n\tilde X_i\hat{\tilde Y}_i^*\Big/\sum_{i=1}^n\tilde X_i^2,\quad \hat g_n^*(t)=\sum_{i=1}^n\left(\hat Y_i^*-X_i\hat\beta_n^*\right)\int_{A_i}E_m(t,s)ds$$

where $\hat{\tilde Y}_i^*=\hat Y_i^*-\sum_{j=1}^n\hat Y_j^*\int_{A_j}E_m\left(t_i,s\right)ds\cdot$

To obtain our results, the following conditions are sufficient.

(A$_1$) $g(\cdot)$ and $f(\cdot)$ belong to the Sobolev space with order $\alpha>1/2$.

(A$_2$) $g(\cdot)$ and $f(\cdot)$ satisfy the Lipschitz condition with order $\gamma>0$.

(A$_3$) $\phi(\cdot)$ is in the Schwartz space with order $l\ge\alpha$, satisfies the Lipschitz condition with order 1 and has a compact support. Furthemore, $\left|\hat\phi(\xi)-1\right|=O(\xi)$ as $\xi\to 0$, where $\hat\phi$ is the Fourier transform of ϕ.

(A$_4$) $\max_{1\le i\le n}\left(s_i-s_j\right)=O\left(n^{-1}\right)$ and $2^m=O\left(n^{1/3}\right)$.

(A$_5$) $F(\cdot)$ and $G(\cdot)$ are continuous functions. In the case $G\left(\tau_F\right)<1$, assume that

$$\int_{-\infty}^{\tau_F}\left(1-F(s-)\right)^{-1}dG(s)<\infty,$$

where $\tau_F=\inf\{t:F(t)=1\}$. In the case $G\left(\tau_F\right)=1$, assume that

$$\left(1-G(u)\right)^q=O\left(1-F(u-)\right)\text{ as } u\uparrow\tau_F,$$

for some $q\in(0,1)$.

In the following discussion, we shall confine ourselves to some more restricted classes of $\left(\phi_1,\phi_2\right)$ than the class K. In particular, we define

$K^*=\{\left(\phi_1,\phi_2\right)\in K:$ for each $s<\tau_G$, there exists a constant $0<C^*<\infty$ such that

$$\sup_{t\le s}\left|\phi_j(t;G)\right|\le C^*,\ j=1,2\},$$

and

$\tilde K^*=\{\left(\phi_1,\phi_2\right)\in K^*:$ there exist constants $0<L=L(s)<\infty$ and $\delta>0$ such that

$$\sup_{t\le s}\left|\phi_j\left(t;\tilde G\right)-\phi_j(t;G)\right|\le L\sup_{t\le s}\left|\tilde G(t)-G(t)\right|,\ j=1,2,$$

for all d.f. $\tilde G$ with $\sup_{t\le s}\left|\tilde G(t)-G(t)\right|<\delta\}$.

For ease of exposition, we shall introduce the following notations which will be used later in the paper. Define

$$E_n^* = n^{-1}\sum_{i=1}^n \tilde{X}_i^2\left[Y_i^* - X_i\beta_n^* - g_n^*(T_i)\right]^2, \qquad \hat{E}_n^* = n^{-1}\sum_{i=1}^n \tilde{X}_i^2\left[\hat{Y}_i^* - X_i\hat{\beta}_n^* - \hat{g}_n^*(T_i)\right]^2$$

$$e_i^* = Y_i^* - E\left(Y_i^* \middle| \mathbf{X},\mathbf{T}\right) = Y_i^* - X_i\beta - g(T_i), \qquad \tilde{f}(T_i) = f(T_i) - \sum_{j=1}^n f(T_j)\int_{A_j} E_m(t_i,s)ds,$$

$$\tilde{g}(T_i) = g(T_i) - \sum_{j=1}^n g(T_j)\int_{A_j} E_m(t_i,s)ds, \qquad \tilde{e}_i^* = e_i^* - \sum_{j=1}^n e_j^*\int_{A_j} E_m(t_i,s)ds,$$

$$\bar{\eta}_i = \sum_{j=1}^n \eta_j\int_{A_j} E_m(t_i,s)ds.$$

3. STATEMENT OF THE RESULTS

Theorem 3.1. Suppose that conditions (A$_1$) - (A$_5$) hold. $E|e_1|^3<\infty, E|\eta_1|^3<\infty$ and $(\phi_1,\phi_2)\in K^*$. Then for every $\gamma \ge 1/3$ $\alpha > 3/2$,

$$\limsup_{n\to\infty}\left[n(2\log\log n)^{-1}\right]^{1/2}\left|\beta_n^*-\beta\right| = \sigma_\eta^{-2}\left(E\left(\eta_i e_i^*\right)^2\right)^{1/2} \quad a.s. \tag{6}$$

$$\sup_{0\le t\le 1}\left|g_n^*(t) - g(t)\right| = O\left(n^{-1/3}\log n\right) \quad a.s. \tag{7}$$

Theorem 3.2. Suppose that conditions (A$_1$) - (A$_5$) hold. $E|e_1|^3<\infty, E|\eta_1|^3<\infty$ and $(\phi_1,\phi_2)\in \tilde{K}^*$. Then for every $\gamma \ge 1/3$ $\alpha > 3/2$,

$$\limsup_{n\to\infty}\left[n(2\log\log n)^{-1}\right]^{1/2}\left|\hat{\beta}_n^*-\beta\right| = \sigma_\eta^{-2}\left(E\left(\eta_1 e_1^*\right)^2\right)^{1/2} \quad a.s. \tag{8}$$

$$\sup_{0\le t\le 1}\left|\hat{g}_n^*(t) - g(t)\right| = O\left(n^{-1/3}\log n\right) \quad a.s. \tag{9}$$

4. PROOFS OF THEOREMS

Throughout this paper, let C denote a generic positive constant which could take different value at each occurrence. Before the proofs of the theorems, we introduce some preliminary results.

Lemma 4.1. (Antoniads *et al*. [21]). If condition (A$_3$) holds, then

(I) $|E_0(t,s)| \le C_k/(1+|t-s|)^k$ and $|E_m(t,s)| \le 2^m C_k/(1+2^m|t-s|)^k$ for $k\in N$, where C_k is a real constant depending on k only.

(II) $\sup_{0\le s\le 1}|E_m(t,s)| = O(2^m)$.

(III) $\sup_t \int_0^1 |E_m(t,s)|ds \le C$.

Lemma 4.2. (Hu and Hu [22]). Let $\tau_m = 2^{-m(\alpha-1/2)}$ when $1/2<\alpha<3/2$, $\tau_m = \sqrt{m}\cdot 2^{-m}$ when $\alpha = 3/2$, $\tau_m = 2^{-m}$ when $\alpha > 3/2$. If conditions (A$_1$)—(A$_4$) hold, then

$$\sup_t\left|f(t) - \sum_{k=1}^n \left(\int_{A_k} E_m(t,s)ds\right) f(T_k)\right| = O\left(n^{-\gamma}\right) + O(\tau_m)$$

$$\sup_t\left|g(t) - \sum_{k=1}^n \left(\int_{A_k} E_m(t,s)ds\right) g(T_k)\right| = O\left(n^{-\gamma}\right) + O(\tau_m)$$

Lemma 4.3. (Qian and Cai [13]). If conditions (A$_1$)—(A$_4$) hold, and $E|e_1|^{3/2+\delta}<\infty$ for some $\delta>0$, then

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \tilde{X}_i^2 = \sigma_\eta^2 \quad a.s.$$

Lemma 4.4. (Qian and Cai [13]). If conditions (A$_1$)—(A$_4$) hold, and $E\eta_1^2<\infty$, then

$$\sup_t\left|\sum_{j=1}^n \eta_j\int_{A_j} E_m(t,s)ds\right| = O\left(n^{-1/6}\log^{1/2} n\right) \quad a.s.$$

Lemma 4.5. (Xue [11]). If conditions (A$_1$)—(A$_4$) hold, then

$$\max_j\left|\sum_{i=1}^n \int_{A_j} E_m(t_i,s)ds\right| = O(1) \quad a.s.$$

Lemma 4.6. (Gu and Lai [23]). Assume condition (A$_5$) holds. If $G(\tau_F)<1$, let

$$\tau_n = \sup\left\{t: 1-F(t)\ge n^{-(1-\kappa)}\right\},$$

where $1/3<\kappa<1/2$, then

$$\limsup_{n\to\infty}\left[n(2\log\log n)^{-1}\right]^{1/2}\sup_{t\le\tau_n}\left|\hat{G}_n(t) - G(t)\right| = \sup_{t\le\tau_F}\left[S(t)\sigma^{-1}(t)\right] \quad a.s.$$

If $G(\tau_F)=1$, then

$$\limsup_{n\to\infty}\left[n(2\log\log n)^{-1}\right]^{1/2}\sup_{t\le\tau_G}\left|\hat{G}_n(t) - G(t)\right| = \sup_{t\le\tau_G}\left[S(t)\sigma^{-1/2}(t)\right] \quad a.s.$$

where

$$S(t) = 1-G(t), \sigma^2(t) = \int_{-\infty}^t (1-G(s))^{-2}(1-F(s))^{-1}dG(s).$$

Lemma 4.7. Under the conditions of theorem 1, we have

$$n^{1/2}\left(\hat{\beta}_n^* - \beta_n^*\right) = o(1) \quad a.s.$$

The proof is similar to that of lemma 1 in [17] and hence omitted here.

Lemma 4.8. (Hardle, Liang and Gao [24]) Let $\{\xi_i, i=1,2,\ldots,n\}$ be independent random variables with $E\xi_i = 0$ and finite variances, and $\sup_{1\le i\le n} E|\xi_i|^r \le C<\infty$ $(r\ge 2)$. Assume that $\{a_{ij}, i,j=1,2,\ldots,n\}$ is a sequence of real numbers such that $\sup_{1\le i,j\le n}|a_{ij}| = O(n^{-p_1})$ for some $0<p_1<1$ and $\sum_{i=1}^n a_{ij} = O(n^{p_2})$ for $p_2 \ge \max\{0, 2/r - p_1\}$. Then

$$\max_{1\le j\le n}\left|\sum_{i=1}^{n} a_{ij}\xi_i\right| = O\left(n^{-s}\log n\right) \text{ for } s=\left(p_1-p_2\right)/2 \quad a.s.$$

Lemma 4.9. Let $\{\xi_i, i=1,2,\ldots,n\}$ be independent random variables with $E\xi_i=0$ and finite variances, and $\sup_{1\le i\le n}E\left|\xi_i\right|^r\le C<\infty$ $(r\ge 3)$. Assume that conditions (A_1)—(A_4) hold. Then

$$\max_{1\le j\le n}\left|\sum_{i=1}^{n}\int_{A_i}E_m\left(t_j,s\right)ds\,\xi_i\right| = O\left(n^{-1/3}\log n\right) \quad a.s.$$

Proof Let $r=3,\ p_1=2/3,\ p_2=0$ and $a_{ij}=\int_{A_i}E_m\left(t_j,s\right)ds$. The lemma follows from lemma 4.1, lemma 4.5 and lemma 4.8.

Proof of Theorem 3.1 Note that

$$\beta_n^*-\beta=\left(\sum_{i=1}^{n}\tilde{X}_i^2\right)^{-1}\left(I_1+I_2\right) \tag{10}$$

where

$$\begin{aligned}
I_1 &= \sum_{i=1}^{n}\tilde{X}_i\left(\tilde{Y}_i^*-E\left(\tilde{Y}_i^*|\mathbf{X,T}\right)\right)\\
&=\sum_{i=1}^{n}\eta_i e_i^*-\sum_{i=1}^{n}\left(\bar{\eta}_i-\sum_{j=1}^{n}\bar{\eta}_j\int_{A_i}E_m\left(t_j,s\right)ds\right)e_i^*\\
&+\sum_{i=1}^{n}\left(\tilde{f}\left(T_i\right)-\sum_{j=1}^{n}\tilde{f}\left(T_j\right)\int_{A_j}E_m\left(t_j,s\right)ds\right)e_i^*-\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\eta_i\int_{A_j}E_m\left(t_i,s\right)ds\right)e_j^*\\
&=I_{11}+I_{12}+I_{13}+I_{14}
\end{aligned} \tag{11}$$

$$\begin{aligned}
I_2 &= \sum_{i=1}^{n}\tilde{X}_i\tilde{g}\left(T_i\right)=\sum_{i=1}^{n}\tilde{f}\left(T_i\right)\tilde{g}\left(T_i\right)+\sum_{i=1}^{n}\left(\tilde{g}\left(T_i\right)-\sum_{j=1}^{n}\tilde{g}\left(T_j\right)\int_{A_i}E_m\left(t_j,s\right)ds\right)\eta_i\\
&=I_{21}+I_{22}
\end{aligned} \tag{12}$$

Let $\hat{e}_i^*=e_i^*I\left(\left|e_i^*\right|<\varepsilon^2 i^{1/2}\right)$, $\tilde{e}_i^*=e_i^*-\hat{e}_i^*$, $\hat{h}_i=E\left(\hat{e}_i^*|\mathbf{X,T}\right)$, $\tilde{h}_i=E\left(\tilde{e}_i^*|\mathbf{X,T}\right)=-\hat{h}_i$, $V_i=a_{ni}\left(\hat{e}_i^*-\hat{h}_i\right)$ and $a_{ni}=\tilde{f}\left(T_i\right)-\sum_{j=1}^{n}\tilde{f}\left(T_j\right)\int_{A_j}E_m\left(t_j,s\right)ds$. Then for any $\varepsilon>0$,

$$I_{13}=\sum_{i=1}^{n}V_i+\sum_{i=1}^{n}a_{ni}\left(\tilde{e}_i^*-\tilde{h}_i\right)=L_1+L_2 \tag{13}$$

By $E\left(e_i^*\right)^2<\infty$ (See lemma 2 in [16]) and three series theorem, we can easily obtain that $\sum_{i=1}^{n}\left|\tilde{e}_i^*\right|<\infty$, *a.s.* By Bernstein's inequality and Bore-Cantelli lemma,

$$\sum_{i=1}^{n}\left|\tilde{h}_i\right|\le\varepsilon n^{1/2}\log n,\ a.s.$$

Hence from lemma 4.2 and 4.5, we obtain that

$$\begin{aligned}
\max_{1\le i\le n}\left|a_{ni}\right| &= \sup_t\left|\tilde{f}\left(t\right)\right|\max_{1\le i\le n}\left(1+\sum_{j=1}^{n}\int_{A_j}E_m\left(t_j,s\right)ds\right)\\
&\le C\sup_t\left|\tilde{f}\left(t\right)\right|=O\left(n^{-\gamma}\right)+O\left(\tau_m\right)
\end{aligned} \tag{14}$$

$$\begin{aligned}
L_2 &\le \sum_{i=1}^{n}\left|a_{ni}\right|\left(\left|\tilde{e}_i^*\right|+\left|\tilde{h}_i\right|\right)\le\max_{1\le i\le n}\left|a_{ni}\right|\left(\sum_{i=1}^{n}\left|\tilde{e}_i^*\right|+\sum_{i=1}^{n}\left|\tilde{h}_i\right|\right)\\
&=\left(O\left(n^{-\gamma}\right)+O\left(\tau_m\right)\right)O\left(\varepsilon n^{1/2}\log n\right)=o\left(n^{1/2}\right)\quad a.s.
\end{aligned} \tag{15}$$

By Bernstein's inequality and Bore-Cantelli lemma,

$$\sum_{i=1}^{n}\left|V_i\right|\le\varepsilon n^{1/2},\ a.s. \tag{16}$$

From (13) (15) and (16), we have that

$$I_{13}=o\left(n^{1/2}\right),\ a.s. \tag{17}$$

Using the similar argument as above, by lemma 4.2, 4.4 and 4.5, we obtain that

$$I_{12}=o\left(n^{1/2}\right),\ a.s.,\ I_{14}=o\left(n^{1/2}\right),\ a.s.,\ I_{22}=o\left(n^{1/2}\right),\ a.s. \tag{18}$$

From lemma 4.2, it follows that

$$I_{21}\le n\sup_t\left|\tilde{f}\left(t\right)\right|\left|\tilde{g}\left(t\right)\right|=O\left(n^{1-2\gamma}\right)+O\left(n\tau_m^2\right)=o\left(n^{1/2}\right) \tag{19}$$

From (10)-(12) and (17)-(19),

$$n^{1/2}\left(\beta_n^*-\beta\right)=n\left(\sum_{i=1}^{n}\tilde{X}_i^2\right)^{-1}\left(n^{-1/2}\sum_{i=1}^{n}\eta_i e_i^*+o\left(1\right)\right)\ a.s. \tag{20}$$

Hence, the (6) follows immediately from lemma 4.3 and LIL of Hartman-Winter.

Note that

$$\begin{aligned}
\sup_{0\le t\le 1}\left|g_n^*\left(t\right)-g\left(t\right)\right| &\le \sup_{0\le t\le 1}\left|\left(\beta-\beta_n^*\right)\sum_{i=1}^{n}\tilde{X}_i\int_{A_i}E_m\left(t,s\right)ds\right|\\
&+\sup_{0\le t\le 1}\left|\sum_{i=1}^{n}g\left(T_i\right)\int_{A_i}E_m\left(t,s\right)ds-g\left(t\right)\right|+\sup_{0\le t\le 1}\left|\sum_{i=1}^{n}e_i^*\int_{A_i}E_m\left(t,s\right)ds\right|\\
&\triangleq n^{1/2}\left|\beta-\beta_n^*\right|I_1+I_2+I_3
\end{aligned} \tag{21}$$

By lemma 4.1, lemma 4.2 and lemma 4.4, we have that

$$\begin{aligned}
I_1 &\le Cn^{-1/2}\left(\log\log n\right)^{1/2}\sup_{0\le t\le 1}\left|\sum_{i=1}^{n}\tilde{f}\left(T_i\right)\int_{A_i}E_m\left(t,s\right)ds\right|\\
&+n^{-1/2}\left(\log\log n\right)^{1/2}\sup_{0\le t\le 1}\left|\sum_{i=1}^{n}\tilde{\eta}_i\int_{A_i}E_m\left(t,s\right)ds\right|\\
&\le Cn^{-1/2}\left(\log\log n\right)^{1/2}\max_{1\le i\le n}\left|\tilde{f}\left(T_i\right)\right|\sup_{0\le t\le 1}\int_0^1\left|E_m\left(t,s\right)\right|ds\\
&+Cn^{-1/2}\left(\log\log n\right)^{1/2}\sup_{0\le t\le 1}\left|\sum_{i=1}^{n}\eta_i\int_{A_i}E_m\left(t,s\right)\right|\\
&+Cn^{-1/2}\left(\log\log n\right)^{1/2}\sup_{0\le t\le 1}\left|\sum_{j=1}^{n}\left(\max_i\left(\int_{A_j}E_m\left(t_i,s\right)ds\right)\left(\sum_{i=1}^{n}\int_{A_i}E_m\left(t,s\right)ds\right)\right)\eta_j\right|\\
&=O\left(n^{-1/2-\gamma}\left(\log\log n\right)^{1/2}\right)+O\left(n^{-1/2}\tau_m\left(\log\log n\right)^{1/2}\right)\\
&+O\left(n^{-2/3}\log^{1/2}n\left(\log\log n\right)^{1/2}\right)=O\left(n^{-2/3}\log^{1/2}n\left(\log\log n\right)^{1/2}\right)
\end{aligned} \tag{22}$$

By lemma 4.2, we have that

$$I_2=O\left(n^{-\gamma}\right)+O\left(\tau_m\right) \tag{23}$$

From $E|e_1|^3 < \infty$ and lemma 2 in [16], we have that $E|e_i^*|^3 < \infty$. Hence, by lemma 4.9,

$$I_3 = O\left(n^{-1/3}\log n\right) \ a.s. \tag{24}$$

Thus (7) follows from (21)-(24). We therefore complete the proof of Theorem 3.1.

Proof of Theorem 3.2 Note that

$$n^{1/2}\left(\hat{\beta}_n^* - \beta\right) = n^{1/2}\left(\hat{\beta}_n^* - \beta_n^*\right) + n^{1/2}\left(\beta_n^* - \beta\right)$$

The (8) follows immediately from lemma 4.7 and theorem 3.1.

Write

$$\hat{g}_n^*(t) - g(t) = \left(\hat{g}_n^*(t) - g_n^*(t)\right) + \left(g_n^*(t) - g(t)\right) \hat{=} I_1 + I_2 \tag{25}$$

$$I_1 = \sum_{i=1}^n \int_{A_i} E_m(t,s)ds\left(\hat{Y}_i^* - Y_i^*\right) - \left(\hat{\beta}_n^* - \beta_n^*\right)\sum_{i=1}^n X_i \int_{A_i} E_m(t,s)ds \hat{=} I_{11} - I_{12} \tag{26}$$

$$\begin{aligned} I_{11} &= \sum_{i=1}^n \int_{A_i} E_m(t,s)ds\left(\hat{Y}_i^* - Y_i^*\right) I\left(Z_i \le \tau_n\right) \\ &+ \sum_{i=1}^n \int_{A_i} E_m(t,s)ds\left(\hat{Y}_i^* - Y_i^*\right) I\left(Z_i > \tau_n\right) \\ &\hat{=} R_1 + R_2 \end{aligned} \tag{27}$$

By lemma 4.1 and 4.6, we have that

$$\begin{aligned} |R_1| &\le \int_0^1 \left|E_m(t,s)\right| ds \sup_{t \le \tau_n}\left|\hat{G}_n(t) - G(t)\right| \\ &\le C\left(n^{-1}\log\log n\right)^{1/2} = o\left(n^{-1/3}\log n\right) \ a.s. \end{aligned} \tag{28}$$

In the case $G(\tau_F) = 1, \max_i Z_i \le \tau_G$ *a.s.*. By lemma 4.1 and 4.6,

$$\begin{aligned} \sup_t |R_2| &\le \sup_t \left|\sum_{i=1}^n \int_{A_i} E_m(t,s)ds\right| \max_{j=1,2}\left|\phi_j\left(Z_i, \hat{G}_n\right) - \phi_j\left(Z_i, G\right)\right| \\ &\le \sup_t \int_0^1 \left|E_m(t,s)\right| ds \sup_{t \le \tau_G}\left|\hat{G}_n(t) - G(t)\right| \\ &\le C\left(n^{-1}\log\log n\right)^{1/2} = o\left(n^{-1/3}\log n\right) \quad a.s. \end{aligned} \tag{29}$$

In the case $G(\tau_F) < 1$. Let $\xi_i = I(Z_i > \tau_n) - E(I(Z_i > \tau_n))$, ξ_i's are independent random variables with zero mean. Since $(\phi_1, \phi_2) \in \tilde{K}^*$, (3) and (5), $\max_i \left|\hat{Y}_i^* - Y_i^*\right| < C$ *a.s.*. Hence from lemma 4.9, we have that

$$\begin{aligned} &\sup_t |R_2| \le \max_i \left|\hat{Y}_i^* - Y_i^*\right| \\ &\sup_t \left|\sum_{i=1}^n \int_{A_i} E_m(t,s)ds\left(I\left(Z_i > \tau_n\right) - P\left(Z_i > \tau_n\right) + P\left(Z_i > \tau_n\right)\right)\right| \\ &\le C\sup_t \left|\sum_{i=1}^n \xi_i \int_{A_i} E_m(t,s)ds\right| + C\sup_t\left|\sum_{i=1}^n \int_{A_i} E_m(t,s)dsP\left(Z_i > \tau_n\right)\right| \\ &\le C\left(n^{-1/3}\log n + n^{-(1-\kappa)}\right) = O\left(n^{-1/3}\log n\right) \quad a.s. \end{aligned} \tag{30}$$

Therefore the desired conclusion (9) follows from (25)-(30) and theorem 3.1.

5. NUMERICAL EXAMPLE

We will simulate a simple semiparametric regression model

$$Y_i = X_i\beta + \cos\left(2\pi T_i\right) + e_i, i = 1, 2, \cdots, 64$$

where $\beta = 1$, $T_i = i/64$, $X_i = 5T_i^2 + \eta_i$, $\eta_i \sim N(0, 0.25)$, $e_i \sim N(0,1)$. The right-censored random variables C_i are independent and identically distribution function $G(u) = 1 - \exp(-2.4u)$ ($u \ge 0$).

Choose $\phi_1(u;G) = u, \phi_2(u;G) = u + G(u)/G'(u)$ and Daubechies scaling `function ${}_2\phi(t)$. By calculation, we have $\beta_n^* = 1.0158$. It closely approximates the true value of parameter β. It can be seen that our method is successful, especial in estimating the parameter. However, a further discussion of the choices the scaling function and (ϕ_1, ϕ_2) is needed so that we can find a good method to use in practical applications.

ACKNOWLEDGMENTS

***Supported** by Scientific Reasearch Item of Education Office, Hubei (No.Q200622001).

REFERENCES

[1] Speckman P. Kernel smoothing in partial linear models. J R Statist Soc 1988; 50: 413-36.

[2] Engle RF, Granger WJ, Rice J, Weiss A. Semiparametric estimates of the relation between weather and electricity sales. J Am Stat Assoc 1986; 80: 310-9.

[3] Chen H. Convergence rates for parametric components in a partly linear model. Ann Stat 1988; 16(1): 136-46.

[4] Bianco A, Boente G. Robust estimators in semiparamtric partly linear regression models. J Stat Plan Inf 2004; 122: 229-52.

[5] Ivet P, King ML, Zhang X. Smoothing spline based tests for nonlinearity in a partially linear model. J Stat Plan Inf 2006; 136: 2446-69.

[6] Hu H. Ridge estimation of a semiparametric regression model. J. Comput Appl Math 2005; 176: 215-22.

[7] Fischer B, Hegland M. Collocation, filtering and nonparametric regression, part 1. ZfV 1999; 2: 46-52.

[8] Green PJ, Silverman BW. Nonparametric regression and generalized linear models. London: Chapman & Hall 1994.

[9] Hong SY. The estimate theory of a semiparametric regression model. Sci China (Ser A) 1991; 21: 1258-69.

[10] Chang XW, Qu LM. Wavelet estimation of partially linear models. Comput Stat Data Anal 2004; 47: 31-48.

[11] Xue LG. Rates of random weighting approximation of wavelet estimates in semiparametric regression model. Chin Acta Math Appl Sinica 2003; 26(1): 11-25.

[12] Sun H, Zhao XM. Convergence rates of estimators for parametric components and nonparametric components in partial linear model. J Eng Math 1999; 16(3): 11-18.

[13] Qian WM, Cai GX. Strong approximability of wavelet estimate in semiparamtric regression model. Sci China (Ser A) 1999; 29: 233-40.

[14] Wang QH, Li G. Empirical likelihood semiparametric regression analysis under random censorship. J Multivar Anal 2002; 83: 469-86.

[15] Wang QH, Zheng ZG. Asymptotic properties for the semiparametric regression model with randomly censored data. Sci China (Ser A) 1999; 40: 945-57.

[16] Qin GS, Cai L. Estimation for the asymptotic variance of parametric estimates in partial linear model with censored data. Acta Math Sci 1996; 16(2): 192-208.

[17] Qin GS, Jing BY. Asymptotic properties for estimation of partial linear models with censored data. J Stat Plan Inf 2000; 84: 95-110.

[18] Pan X, Fu ZT. The asymptotic properties of wavelet estimation in a semiparametric regression model under random censorship. Acta Math Appl Sinica 2006; 29(1): 68-80.
[19] Liang H, Zhou Y. Asymptotic normality in a semiparametric partially linear model with right-censored data. Commun Statist Theory Meth 1998; 27(12): 2895-907.
[20] Zheng ZK. A class of estimators of the parameters in linear regression with censored data. Acta Math Appl Sinica 1987; 3(3): 231-41.
[21] Antoniads A, Gregoire G, Mckeague IW. Wavelet methods for curve estimation. J. Am Stat Assoc 1994; 89: 1340-53.
[22] Hu HC, Hu DH. Strong consistency of wavelet estimation in semiparametric regression models. Acta Math Sinica (Chinese Series) 2006; 49(6): 1417-24.
[23] Gu MG, Lai TL. Functional law of the iterated logarithm for the product-limit estimator of a function under random censorship or truncation. Ann Probab 1990; 18: 160-89.
[24] Hardle WG, Liang H, Gao JT. Partially linear models. Heidelberg: Physica-Verlag 2000.

Received: November 23, 2007 Revised: December 14, 2007 Accepted: March 06, 2008