J. Korean Math. Soc. **44** (2007), No. 3, pp. 525–539

ASYMPTOTIC NORMALITY OF ESTIMATOR IN NON-PARAMETRIC MODEL UNDER CENSORED SAMPLES

SI-LI NIU AND QIAN-RU LI

ABSTRACT. Consider the regression model $Y_i = g(x_i)+e_i$ for $i = 1, 2, \dots$, n, where: (1) x_i are fixed design points, (2) e_i are independent random errors with mean zero, (3) $g(\cdot)$ is unknown regression function defined on $[0, 1]$. Under Y_i are censored randomly, we discuss the asymptotic normality of the weighted kernel estimators of g when the censored distribution function is known or unknown.

1. Introduction

Consider the fixed design regression model

$$Y_i = g(x_i) + e_i, \quad i = 1, 2, \dots, n, \tag{1.1}$$

where $Y_1, Y_2, \dots, Y_n$ are the observations on the fixed design points $x_1, x_2, \dots$, x_n; $e_1, e_2, \dots, e_n$ are independent random errors with mean zero; $g(\cdot)$ is the unknown regression function defined on $[0, 1]$. Without loss of generality, we assume that $0 = x_0 < x_1 < \cdots < x_n = 1$.

Under independent samples, many researchers have studied the large sample properties of estimator of $g(\cdot)$. For instance, consistency and asymptotic normality have been studied by Priestly and Chao [9], Benedetti [1], Georgiev and Greblicki [5] and Georgiev [4] among others. Under various dependent samples, the estimators of $g(\cdot)$ also have been widely studied, such as Fan [2], Roussas [10], Roussas et al. [11] and Tran et al. [12]. In particular, Prieshey and Chao [9] proposed the following weighted kernel estimator of $g(x)$

$$g_n(x) = \sum_{i=1}^{n} \frac{x_i - x_{i-1}}{h_n} K(\frac{x - x_i}{h_n}) Y_i, \ x \in [0, 1], \tag{1.2}$$

where $K(\cdot)$ is a kernel function, h_n is a sequence of positive constants tending to 0.

Received August 22, 2005.
2000 *Mathematics Subject Classification.* 62G05.
Key words and phrases. censored sample, non-parametric regression model, weighted kernel estimator, asymptotic normality.

In this paper, we consider the random censorship case for $Y_i, 1 \le i \le n$, i.e. when we observe $Y_i,\ 1 \le i \le n$, we observe only the pairs $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$:

$$Z_i = \min\{Y_i, T_i\}, \quad \delta_i = I[Y_i \le T_i], \quad i = 1, 2, \ldots, n,$$

where $\{T_i,\ 1 \le i \le n\}$ are censored variables, $I[\cdot]$ denotes the indicator function. The estimator of g will be constructed by the censored data $\{(Z_i,\ \delta_i),\ i = 1, 2, \ldots, n\}$.

Considering the practical background of the random censorship model, we assume that both $Y_i,\ 1 \le i \le n$ and $T_i,\ 1 \le i \le n$ are nonnegative and independent random variables; moreover every T_i is independent of every Y_i with distribution function $F_i,\ 1 \le i \le n$. The censored variables $T_1,\ T_2, \ldots$ are assumed to be independent, identically distributed with distribution function G. Then the distribution function of Z_i is

$$H_i(x) = P(Z_i \le x) = 1 - (1 - F_i(x))(1 - G(x)) =: 1 - F_{is}(x)G_s(x),$$

where $F_{is} = 1 - F_i,\ 1 \le i \le n,\ G_s = 1 - G$. Set

$$\bar{F}_s = \frac{1}{n}\sum_{i=1}^{n} F_{is},\ H_{is} = 1 - H_i = F_{is}G_s,\ \bar{H}_s = \frac{1}{n}\sum_{i=1}^{n} H_{is}.$$

For any distribution function F, define

$$\tau_F = \inf\{t : F(t) = 1\}, \quad F^{-p}(t) = \left(\frac{1}{F(t)}\right)^p \quad \text{for all } p > 0.$$

Throughout this paper, we assume $\tau_{F_i} \le \tau_G,\ i = 1, \ldots, n$. Note that

$$E\delta_i Z_i G_s^{-1}(Z_i) = \int_o^{\tau_{F_i}} dF_i(y) \int_y^{\tau_G} yG_s^{-1}(y)dG(t) = EY_i = g(x_i). \tag{1.3}$$

Therefore, we may assume that $\{\delta_i Z_i G_s^{-1}(Z_i),\ 1 \le i \le n\}$ obey the following model:

$$\delta_i Z_i G_s^{-1}(Z_i) = g(x_i) + e_i^{'}, \quad i = 1, 2, \ldots, n, \tag{1.4}$$

where $\{e_i^{'},\ 1 \le i \le n\}$ are independent random errors with mean zero. Hence, when G is known, taking advantage of model (1.4) and comparing with (1.2), the weighted kernel estimator of $g(x)$ is defined as follows:

$$g_n^{(1)}(x) = \sum_{i=1}^{n} \frac{x_i - x_{i-1}}{h_n} K(\frac{x - x_i}{h_n})\delta_i Z_i G_s^{-1}(Z_i), \quad x \in [0, 1].$$

Also, we assume that $\tau_{(n)} =: \max\{\tau_{F_i} : 1 \le i \le n\} \le \tau_G \le +\infty$ for all $n \ge 1$, and that $\tau_0 =: \lim_n \tau_{(n)}$ exists.

When G is unknown, the estimate of $g(x)$ is defined by

$$g_n^{(2)}(x) = \sum_{i=1}^{n} \frac{x_i - x_{i-1}}{h_n} K(\frac{x - x_i}{h_n})\delta_i Z_i \hat{G}_{ns}^{-1}(Z_i) I[Z_i \le M_n], \quad x \in [0, 1],$$

where M_n is a sequence of constants, which are less than $\tau_{(n)}$ and monotonously increase to τ_0, $\hat{G}_{ns}$ denotes the Kaplan-Meier estimator of the distribution function G, i.e.,

$$\hat{G}_{ns}(t) = \begin{cases} \prod_{j=1}^{n} \left(\frac{1+N^+(Z_j)}{2+N^+(Z_j)}\right)^{I[\delta_j=0,\ Z_j\le t]}, & t \le Z_{(n)}, \\ 0, & t > Z_{(n)} \end{cases}$$

(cf. Kaplan and Meier [6]), $Z_{(n)} = \max\{Z_1, Z_2, \dots, Z_n\}$, $N^+(Z_j) = \sum_{i=1}^{n} I[Z_i > Z_j]$, $1 \le j \le n$.

In the sequel, let C, C_1 and $C_2, \dots$ will represent positive constants whose value may change from one place to another.

For model (1.1), the consistent rates of estimators of g have also been studied by some authors under censored assumptions, see Xue [14] and Wang [13]. In particular, Wang [13] discussed the weak consistency and consistent rates of the estimators $g_n^{(1)}$ and $g_n^{(2)}$ of g. His results are as follows.

Theorem 1.1. *Let $p > \sqrt{2}$, $0 < \alpha < 1$, and let $K(u)$ be a continuous probability density kernel. Set $\hat{K}(u) = |u|^\alpha K(u)$ with $\int_{-\infty}^{\infty} \hat{K}(u)du < \infty$. Suppose that*

(A.1) *there exists some $M_0 \ge 0$ such that $\hat{K}(u)$ is non-increasing when $u > M_0$, and nondecreasing when $u < -M_0$,*
(A.2) $\max_{1\le i\le n}(x_i - x_{i-1}) \le C/n$ *for all $n \ge 1$,*
(A.3) *$g(x)$ satisfies Lipschitz condition of order α on $[0, 1]$.*

If $\max_{1\le i\le n} E[Y_i^p G_s^{1-p}(Y_i)] \le C$, and one of the following conditions holds.

$$(i)\ \sqrt{2} < p \le 2,\ n^{2/p-p}h_n^{1-p} \to 0; \quad (ii)\ p > 2,\ n^{-p/2}h_n^{1-p} \to 0.$$

Then $g_n^{(1)}(x) \xrightarrow{P} g(x)$ *for all* $x \in [0, 1]$.

Theorem 1.2. *Let $p > \frac{1-\beta+\sqrt{(1-\beta)^2+8}}{2}$ for some $0 \le \beta \le 1$. Suppose that (A.1)-(A.3) in Theorem 1.1 are satisfied. If $\max_{1\le i\le n} E[Y_i^p G_s^{1-p}(Y_i)] \le C$, and one of the following conditions holds.*

$$(a) \quad 0 < \beta \le 1,\ \frac{1-\beta+\sqrt{(1-\beta)^2+8}}{2} < p \le 2,\ \sum_{n=1}^{\infty} n^{2/p-p-\beta}h_n^{1-p} < \infty;$$

$$(b) \quad 0 \le \beta \le 1,\ p > 2,\ \sum_{n=1}^{\infty} n^{-2/p-\beta}h_n^{1-p} < \infty.$$

Then, for any fixed $x \in [0, 1]$, we have

$$\sum_{n=1}^{\infty} n^{-\beta} P(|g_n^{(1)}(x) - g(x)| > \epsilon) < \infty.$$

Theorem 1.3. *Suppose that* (A.1)-(A.3) *in Theorem 1.1 are satisfied, and that*

(A.4) *both F_i $(1 \le i \le n)$, and G are continuous distribution functions,* (A.5) $\tau_0 < \tau_G$,

(A.6) $\liminf_{n\to\infty} \overline{F}_s(M_n)\log n \geq b$ *for some* $b > 0$.

If $\{F_i, i \geq 1\}$ *are equi-continuous at* τ_0 *and* $\sum_{n=1}^{\infty} n^{-p/2} h_n^{1-p} < \infty$, *then*

$$\sum_{n=1}^{\infty} P(|g_n^{(2)}(x) - g(x)| > \epsilon) < \infty \text{ for all } \epsilon > 0 \text{ and } x \in [0,1].$$

In this paper, we will establish the asymptotic normality of the estimators $g_n^{(1)}$ and $g_n^{(2)}$ of g under some suitable conditions.

The paper is organized as follows. In Section 2, we introduce the assumptions used throughout the paper and give main results. Proofs will be provided in Sections 3.

2. Main results

In order to state the main results, we first list the following some assumptions.

(B.1) Let $K(u)$ be a continuous probability density kernel with $0 < \int_{-\infty}^{+\infty} K^2(u)du < \infty$; there exists some $M_0 \geq 0$ such that $K(u)$ is non-increasing when $u > M_0$ and nondecreasing when $u < -M_0$.

(B.2) For all $n \geq 1$, $\frac{C_1}{n} \leq x_i - x_{i-1} \leq \frac{C_2}{n}$, $1 \leq i \leq n$.

(B.3) For $0 < \theta < 1$, $\liminf_{n\to\infty} n^{\theta}\bar{F}_s(M_n) \geq b$ for some $b > 0$; $\lim_{n\to\infty} h_n^{\frac{p}{2}} \bar{F}_s^{1-2p}(M_n) = 0$ for some $p \geq 2$.

Theorem 2.1. *Suppose that* (B.1) *and* (B.2) *are satisfied and* $nh_n \to \infty$. *Set* $Z_{iG} = \delta_i Z_i G_s^{-1}(Z_i)$. *If* $\lim_{C\to\infty}\sup_i EZ_{iG}^2 I(Z_{iG} > C) = 0$, *and* $\sigma_i^2 =: \mathrm{Var}(Z_{iG}) \geq \delta^2$, $1 \leq i \leq n$ *for some* $\delta > 0$, *then*

$$\frac{g_n^{(1)}(x) - Eg_n^{(1)}(x)}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}} \to_d N(0,1), \quad x \in [0,1].$$

Theorem 2.2. *Suppose that* (A.4)-(A.5) *and* (B.1)-(B.3) *are satisfied. If* $\sigma_i^2 =: \mathrm{Var}(Z_{iG}) \geq \delta^2$, $1 \leq i \leq n$ *for some* $\delta > 0$ *and* $\sqrt{nh_n}\sup_i[F_i(\tau_0) - F_i(M_n)] \to 0$, *then*

$$\frac{g_n^{(2)}(x) - Eg_n^{(1)}(x)}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}} \to_d N(0,1), \quad x \in [0,1].$$

3. Proofs of main results

Lemma 3.1. ([13, Lemma 3.1]) *Suppose that* $\bar{K}(\cdot)$ *is a continuous function with* $\int_{-\infty}^{\infty} \bar{K}(u)du < +\infty$, *and there exists some* $M_0 \geq 0$ *such that* $\bar{K}(u)$ *is non-increasing when* $u > M_0$ *and non-decreasing when* $u < -M_0$. *If* $\max_{1\leq i\leq n}(x_i - x_{i-1}) \leq C/n$ *for all* n *and* $\lim_{n\to\infty} nh_n = \infty$, *then*

$$h_n^{-1}\sum_{i=1}^{n}(x_i - x_{i-1})\bar{K}\left(\frac{x - x_i}{h_n}\right) \to \int_{-\infty}^{+\infty} \bar{K}(u)du.$$

Lemma 3.2. ([7]) *For any* $r > 0$,

$$E[(1+N^+(Z_i))^{-r}|(Z_i,\delta_i)] \le n^{-r}[\bar{H}_s(Z_i) - n^{-1}]^{-r}.$$

Lemma 3.3. *Suppose that* (A.4)-(A.5) *and* (B.1)-(B.3) *are satisfied. If* $\sigma_i^2 =: \mathrm{Var}(Z_{iG}) \ge \delta^2$, $1 \le i \le n$ *for some* $\delta > 0$, *then*

$$\frac{\sup_{0\le t\le M_n} |\log \hat{G}_{ns}(t) - \log G_s(t)|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}} \to_p 0.$$

Proof. For every $n \ge 1$ and $t \in [0, M_n]$, define $\beta_j(t) = I[\delta_j = 0, Z_j \le t]$, then by the Taylor expansion we have

$$\begin{aligned}
\log \hat{G}_{ns}(t) &= \sum_{j=1}^n I[\delta_j = 0, Z_j \le t] \log(\frac{1+N^+(Z_j)}{2+N^+(Z_j)}) \\
&= \sum_{j=1}^n \beta_j(t) \log(1 - \frac{1}{2+N^+(Z_j)}) \\
&= -\sum_{j=1}^n \beta_j(t) \sum_{l=1}^\infty \frac{1}{l}(2+N^+(Z_j))^{-l}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\log \hat{G}_{ns}(t) - \log G_s(t) \\
= \;& [-\frac{1}{n}\sum_{j=1}^n \beta_j(t)\bar{H}_s^{-1}(Z_j) - \log G_s(t)] - \sum_{j=1}^n \beta_j(t) \sum_{l=2}^\infty \frac{1}{l}(2+N^+(Z_j))^{-l} \\
&-\frac{1}{n}\sum_{j=1}^n \beta_j(t)[n(2+N^+(Z_j))^{-1} - \bar{H}_s^{-1}(Z_j)] \\
=: \;& L_{n1} + L_{n2} + L_{n3}.
\end{aligned}$$

Since $\{\delta_i Z_i G_s^{-1}(Z_i), 1 \le i \le n\}$ are independent random variables, by assumption (B.2) and Lemma 3.1, we have

$$\begin{aligned}
\mathrm{Var}(g_n^{(1)}(x)) &= \mathrm{Var}[\sum_{i=1}^n (\frac{x_i - x_{i-1}}{h_n}) K(\frac{x - x_i}{h_n}) \delta_i Z_i G_s^{-1}(Z_i)] \\
&= \sum_{i=1}^n (\frac{x_i - x_{i-1}}{h_n})^2 K^2(\frac{x - x_i}{h_n}) \sigma_i^2 \\
&\ge \frac{C\delta^2}{nh_n} \sum_{i=1}^n \frac{x_i - x_{i-1}}{h_n} K^2(\frac{x - x_i}{h_n}) \\
&\ge \frac{C}{nh_n}. \qquad (3.5)
\end{aligned}$$

Hence, for any $\epsilon > 0$ and $t \in [0, M_n]$,

$$\begin{aligned}
& P(\frac{|\log \hat{G}_{ns}(t) - \log G_s(t)|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}} > \epsilon) \\
\leq\ & P(|\log \hat{G}_{ns}(t) - \log G_s(t)| > \frac{C\epsilon}{\sqrt{nh_n}}) \\
\leq\ & P(|L_{n1}| > \frac{C\epsilon}{3\sqrt{nh_n}}) + P(|L_{n2}| > \frac{C\epsilon}{3\sqrt{nh_n}}) \\
& +P(|L_{n3}| > \frac{C\epsilon}{3\sqrt{nh_n}}).
\end{aligned} \tag{3.6}$$

Next, we divide three steps to estimate the three terms in (3.6) respectively. *Step 1.* Note that

$$\begin{aligned}
& -E\beta_j(t)\bar{H}_s^{-1}(Z_j) \\
=\ & -\int_o^t (\int_u^{\tau_{F_j}} \frac{1}{\bar{H}_s(u)} dF_j(y))dG(u) \\
=\ & -\int_0^t \frac{1-F_j(u)}{\frac{1}{n}\sum_{i=1}^n F_{is}(u)G_s(u)} dG(u) \\
=\ & -n\int_0^t \frac{F_{js}(u)}{G_s(u)\sum_{i=1}^n F_{is}(u)} dG(u),
\end{aligned}$$

therefore,

$$-\frac{1}{n}\sum_{j=1}^n E[\beta_j(t)\bar{H}_s^{-1}(Z_j)] = -\int_0^t \frac{1}{G_s(u)} dG(u) = \log G_s(t).$$

Thus, $L_{n1}(t) = \frac{1}{n}\sum_{j=1}^n[-\beta_j(t)\bar{H}_s^{-1}(Z_j) + E\beta_j(t)\bar{H}_s^{-1}(Z_j)]$ and we obtain from (B.3) and the Dharmadhikar-Jogdeo inequality (Praksas Rao [8]) that for $p \geq 2$

$$\begin{aligned}
& P(|L_{n1}| > \frac{C\epsilon}{3\sqrt{nh_n}}) \\
\leq\ & Cn^{-\frac{p}{2}} h_n^{\frac{p}{2}} E|\sum_{j=1}^n[-\beta_j(t)\bar{H}_s^{-1}(Z_j) + E\beta_j(t)\bar{H}_s^{-1}(Z_j)]|^p \\
\leq\ & Cn^{-\frac{p}{2}} h_n^{\frac{p}{2}} n^{\frac{p}{2}-1} \sum_{j=1}^n E|\beta_j(t)\bar{H}_s^{-1}(Z_j) - E\beta_j(t)\bar{H}_s^{-1}(Z_j)|^p \\
\leq\ & Cn^{-\frac{p}{2}} h_n^{\frac{p}{2}} n^{\frac{p}{2}-1} \sum_{j=1}^n [E|\beta_j(t)\bar{H}_s^{-1}(Z_j)|^p + |E\beta_j(t)\bar{H}_s^{-1}(Z_j)|^p]
\end{aligned}$$

$$
\begin{aligned}
&\leq Cn^{-1}h_n^{\frac{p}{2}}\sum_{j=1}^{n}E\beta_j(t)\bar{H}_s^{-p}(Z_j)\\
&= Cn^{-1}h_n^{\frac{p}{2}}\cdot n\int_0^t \frac{\frac{1}{n}\sum_{j=1}^{n}F_{js}(u)}{G_s^p(u)(\frac{1}{n}\sum_{i=1}^{n}F_{is}(u))^p}dG(u)\\
&\leq Ch_n^{\frac{p}{2}}\bar{F}_s^{1-p}(M_n)\to 0.
\end{aligned}
$$

Step 2. We observe that

$$
\begin{aligned}
|L_{n2}| &\leq \sum_{j=1}^{n}\beta_j(t)\sum_{l=2}^{\infty}(2+N^+(Z_j))^{-l}\\
&= \sum_{j=1}^{n}\beta_j(t)\frac{\frac{1}{(2+N^+(Z_j))^2}}{1-\frac{1}{2+N^+(Z_j)}}\\
&\leq \sum_{j=1}^{n}\beta_j(t)\frac{1}{(1+N^+(Z_j))^2}.
\end{aligned}
$$

Hence, by applying the Minkowski inequality we have

$$
\begin{aligned}
&P(|L_{n2}|\geq \frac{C\epsilon}{3\sqrt{nh_n}})\\
&\leq P(|\sum_{j=1}^{n}\beta_j(t)\frac{1}{(1+N^+(Z_j))^2}|>\frac{C\epsilon}{3\sqrt{nh_n}})\\
&\leq Cn^{\frac{p}{2}}h_n^{\frac{p}{2}}E|\sum_{j=1}^{n}\beta_j(t)\frac{1}{(1+N^+(Z_j))^2}|^p\\
&\leq Cn^{p-1}n^{\frac{p}{2}}h_n^{\frac{p}{2}}\sum_{j=1}^{n}E\beta_j(t)(1+N^+(Z_j))^{-2p}\\
&= Cn^{\frac{3p}{2}-1}h_n^{\frac{p}{2}}\sum_{j=1}^{n}E\{\beta_j(t)E[(1+N^+(Z_j))^{-2p}|(Z_j,\delta_j)]\}.
\end{aligned}
\tag{3.7}
$$

Lemma 3.2 yields

$$
E[(1+N^+(Z_j))^{-2p}|(Z_j,\delta_j)]\leq n^{-2p}[\bar{H}_s(Z_j)-n^{-1}]^{-2p}.
\tag{3.8}
$$

The assumptions (A.5) and (B.3) imply that for sufficiently large n and all $t\in[0,M_n]$,

$$
\begin{aligned}
&n\bar{H}_s(t)-1\geq n\bar{H}_s(M_n)-1\geq G_s(\tau_0)n\bar{F}_s(M_n)-1\\
&= G_s(\tau_0)n^{1-\theta}n^{\theta}\bar{F}_s(M_n)-1\geq bG_s(\tau_0)n^{1-\theta}-1\to\infty,
\end{aligned}
$$

further we have

$$
[\bar{H}_s(t)-n^{-1}]^{-1}=\frac{1}{\bar{H}_s(t)}[1+\frac{1}{n\bar{H}_s(t)-1}]\leq C\bar{H}_s^{-1}(t).
\tag{3.9}
$$

From (3.7)-(3.9) and (B.3), we find

$$\begin{aligned}
&P(|L_{n2}| \geq \frac{C\epsilon}{3\sqrt{nh_n}}) \\
\leq\ & Cn^{\frac{3p}{2}-1}h_n^{\frac{p}{2}}n^{-2p}\sum_{j=1}^{n} E\beta_j(t)\bar{H}_s^{-2p}(Z_j) \\
=\ & Cn^{-\frac{p}{2}-1}h_n^{\frac{p}{2}}\sum_{j=1}^{n} EI[Y_j > T_j, Y_j \wedge T_j \leq t]\bar{H}_s^{-2p}(Y_j \wedge T_j) \\
\leq\ & Cn^{-\frac{p}{2}-1}h_n^{\frac{p}{2}}\sum_{j=1}^{n}\int_o^t\int_u^{\tau_{F_j}} \frac{1}{\bar{H}_s^{2p}(u)}dF_j(y)dG(u) \\
\leq\ & Cn^{-\frac{p}{2}-1}h_n^{\frac{p}{2}}\int_0^t \frac{\sum_{j=1}^n F_{js}(u)}{G_s^{2p}(u)(\frac{1}{n}\sum_{i=1}^n F_{is}(u))^{2p}}dG(u) \\
\leq\ & Cn^{-\frac{p}{2}}h_n^{\frac{p}{2}}\bar{F}_s^{1-2p}(M_n) \to 0.
\end{aligned}$$

Step 3. Note that $\frac{-1}{2+N^+(Z_j)} \leq \frac{1}{(1+N^+(Z_j))^2} - \frac{1}{1+N^+(Z_j)}$, so

$$\begin{aligned}
L_{n3} \ &\leq\ \frac{1}{n}\sum_{j=1}^{n}\beta_j(t)[\frac{n}{(1+N^+(Z_j))^2} - \frac{n}{1+N^+(Z_j)}] + \frac{1}{n}\sum_{j=1}^{n}\beta_j(t)\bar{H}_s^{-1}(Z_j) \\
&=\ \sum_{j=1}^{n}\frac{\beta_j(t)}{(1+N^+(Z_j))^2} - \frac{1}{n}\sum_{j=1}^{n}\beta_j(t)(\frac{n}{1+N^+(Z_j)} - \bar{H}_s^{-1}(Z_j)) \\
&=:\ L_{n31} - L_{n32}.
\end{aligned} \tag{3.10}$$

From *Step 2* we have

$$P\left(|L_{n31}| \geq \frac{C\epsilon}{6\sqrt{nh_n}}\right) \to 0. \tag{3.11}$$

As to L_{n32} we have

$$\begin{aligned}
&P\left(|L_{n32}| \geq \frac{C\epsilon}{6\sqrt{nh_n}}\right) \\
=\ & P(|\frac{1}{n}\sum_{j=1}^{n}\beta_j(t)(\frac{n}{1+N^+(Z_j)} - \frac{1}{\bar{H}_s(Z_j)})| \geq \frac{C\epsilon}{6\sqrt{nh_n}}) \\
\leq\ & Cn^{\frac{p}{2}}h_n^{\frac{p}{2}}n^{-p}E|\sum_{j=1}^{n}\beta_j(t)(\frac{n}{1+N^+(Z_j)} - \frac{1}{\bar{H}_s(Z_j)})|^p \\
\leq\ & Cn^{-\frac{p}{2}}h_n^{\frac{p}{2}}n^{p-1}\sum_{j=1}^{n}E\beta_j(t)|\frac{n}{1+N^+(Z_j)} - \frac{1}{\bar{H}_s(Z_j)}|^p \\
=\ & Cn^{\frac{p}{2}-1}h_n^{\frac{p}{2}}\sum_{j=1}^{n}E\{\beta_j(t)E[|\frac{n}{1+N^+(Z_j)} - \frac{1}{\bar{H}_s(Z_j)}|^p|(Z_j,\delta_j)]\}
\end{aligned} \tag{3.12}$$

and

$$
\begin{aligned}
& E\left[|\frac{n}{1+N^+(Z_j)} - \frac{1}{\bar{H}_s(Z_j)}|^p|(Z_j,\delta_j)\right] \\
= \;& E\left[|\frac{1+N^+(Z_j)-n\bar{H}_s(Z_j)}{\bar{H}_s(Z_j)(1+N^+(Z_j))}|^p|(Z_j,\delta_j)\right] \\
\leq \;& \bar{H}_s^{-p}(Z_j)E^{\frac{1}{2}}[(1+N^+(Z_j))^{-2p}|(Z_j,\delta_j)] \\
& \cdot E^{\frac{1}{2}}[|1+N^+(Z_j)-n\bar{H}_s(Z_j)|^{2p}|(Z_j,\delta_j)]. \qquad (3.13)
\end{aligned}
$$

By Lemma 3.2, for $1 \leq j \leq n$,

$$
\begin{aligned}
E[(1+N^+(Z_j))^{-2p}|(Z_j,\delta_j)] &\leq n^{-2p}[\bar{H}_s(Z_j)-n^{-1}]^{-2p} \\
&\leq Cn^{-2p}\bar{H}_s^{-2p}(Z_j). \qquad (3.14)
\end{aligned}
$$

On applying the Dharmadhikar-Jogdeo inequality (Praksas Rao [8]) we have

$$
\begin{aligned}
& E[|1+N^+(Z_j)-n\bar{H}_s(Z_j)|^{2p}|(Z_j,\delta_j)] \\
= \;& E[|1+\sum_{l=1}^{n}(I(Z_l>Z_j)-H_{ls}(Z_j))|^{2p}|(Z_j,\delta_j)] \\
\leq \;& C\{1+n^{p-1}\sum_{l=1}^{n}E[|I(Z_l>Z_j)-H_{ls}(Z_j)|^{2p}|(Z_j,\delta_j)]\} \\
\leq \;& Cn^p. \qquad (3.15)
\end{aligned}
$$

Hence, (3.12)-(3.15) and (B.3) yield that

$$
\begin{aligned}
P(|L_{n32}| \geq \frac{C\epsilon}{6\sqrt{nh_n}}) &\leq Cn^{-1}h_n^{\frac{p}{2}}\sum_{j=1}^{n}E\beta_j(t)\bar{H}_s^{-2p}(Z_j) \\
&= Cn^{-1}h_n^{\frac{p}{2}}\sum_{j=1}^{n}\int_o^t\int_u^{\tau_{F_j}}\frac{1}{\bar{H}_s^{2p}(u)}dF_j(y)dG(u) \\
&\leq Cn^{-1}h_n^{\frac{p}{2}}\int_0^t\frac{\sum_{j=1}^{n}F_{js}(u)}{G_s^{2p}(u)(\frac{1}{n}\sum_{i=1}^{n}F_{is}(u))^{2p}}dG(u) \\
&\leq Ch_n^{\frac{p}{2}}\bar{F}_s^{1-2p}(M_n) \to 0. \qquad (3.16)
\end{aligned}
$$

Then, (3.10), (3.11) and (3.16) follow that $P(|L_{n3}| \geq \frac{C\epsilon}{3\sqrt{nh_n}}) \to 0$.

(3.6) and *Steps 1-3* yield that for all $t \in [0, M_n]$,

$$
\frac{|\log \hat{G}_{ns}(t) - \log G_s(t)|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}} \to_p 0. \qquad (3.17)
$$

Finally, by the continuity of G and (3.17), utilizing the method used in Földes and Rejtö [3] (or see Wang [13]), we can obtain that

$$\frac{\sup_{0\le t\le M_n}|\log \hat{G}_{ns}(t)-\log G_s(t)|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}}\to_p 0$$

for sufficiently large n. □

Proof of Theorem 2.1. Set $\xi_{ni}=\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})Z_{iG}$. Then $g_n^{(1)}(x)=\sum_{i=1}^n \xi_{ni}$. From the assumption of independence on Y_i and T_i, it suffices to verify Linderberg condition: for any selected $x\in[0,1]$ and any $\epsilon>0$,

$$A_n(\varepsilon)=\frac{1}{s_n^2}\sum_{i=1}^n E|\xi_{ni}-E\xi_{ni}|^2 I(|\xi_{ni}-E\xi_{ni}|\ge \varepsilon s_n)\to 0,$$

where $s_n^2=\mathrm{Var}(g_n^{(1)}(x))$. Note that $\lim_{C\to\infty}\sup_i EZ_{iG}^2 I(|Z_{iG}|>C)=0$, namely

$$\lim_{C\to\infty}\sup_i\int_{|\delta_i Z_i G_s^{-1}(Z_i)|>C}|\delta_i Z_i G_s^{-1}(Z_i)|^2 dP=0,$$

which follows that

$$\begin{aligned}&\sup_i E|\delta_i Z_i G_s^{-1}(Z_i)|^2<\infty \text{ and}\\ &\sup_i |g(x_i)|\le \sup_i E|\delta_i Z_i G_s^{-1}(Z_i)|<\infty.\end{aligned}\tag{3.18}$$

Therefore, from (B.1), (B.2) and (3.5) we have

$$\begin{aligned}A_n(\varepsilon)&=\frac{1}{s_n^2}\sum_{i=1}^n E|\xi_{ni}-E\xi_{ni}|^2 I(|\xi_{ni}-E\xi_{ni}|\ge\varepsilon s_n)\\
&=\frac{1}{s_n^2}\sum_{i=1}^n E[\delta_i Z_i G_s^{-1}(Z_i)-g(x_i)]^2[\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})]^2\\
&\quad\cdot I(|(\delta_i Z_i G_s^{-1}(Z_i)-g(x_i))\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})|\ge\varepsilon s_n)\\
&\le C\sum_{i=1}^n E(\delta_i Z_i G_s^{-1}(Z_i)-g(x_i))^2[\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})]\\
&\quad\cdot I(|\delta_i Z_i G_s^{-1}(Z_i)-g(x_i)|\ge\frac{\varepsilon s_n}{|\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})|})\\
&\le C\sum_{i=1}^n E(\delta_i Z_i G_s^{-1}(Z_i)-g(x_i))^2[\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})]\\
&\quad\cdot I(|\delta_i Z_i G_s^{-1}(Z_i)-g(x_i)|\ge C\epsilon\sqrt{nh_n}).\end{aligned}\tag{3.19}$$

From Lemma 3.1

$$\sum_{i=1}^{n} \frac{x_i - x_{i-1}}{h_n} K(\frac{x - x_i}{h_n}) \to \int_{-\infty}^{+\infty} K(u)du = 1. \tag{3.20}$$

We observe that

$$\begin{aligned} & E(\delta_i Z_i G_s^{-1}(Z_i) - g(x_i))^2 I(|\delta_i Z_i G_s^{-1}(Z_i) - g(x_i)| \geq C\epsilon\sqrt{nh_n}) \\ \leq \ & C\{E(\delta_i Z_i G_s^{-1}(Z_i))^2 I(|\delta_i Z_i G_s^{-1}(Z_i) - g(x_i)| \geq C\epsilon\sqrt{nh_n}) \\ & + g^2(x_i) EI(|\delta_i Z_i G_s^{-1}(Z_i) - g(x_i)| \geq C\epsilon\sqrt{nh_n})\} \\ =: \ & E_1 + E_2. \end{aligned} \tag{3.21}$$

(3.18) and $nh_n \to \infty$ yield that

$$\begin{aligned} E_1 & \leq CE(\delta_i Z_i G_s^{-1}(Z_i))^2 I(|\delta_i Z_i G_s^{-1}(Z_i)| \geq C\epsilon\sqrt{nh_n} - |g(x_i)|) \\ & \to 0. \end{aligned} \tag{3.22}$$

On applying the Chebyshev inequality, from (3.18) and $nh_n \to \infty$ we have

$$\begin{aligned} E_2 & \leq \frac{g^2(x_i) E(\delta_i Z_i G_s^{-1}(Z_i) - g(x_i))^2}{C^2 \epsilon^2 nh_n} \\ & \leq \frac{g^2(x_i)[E(\delta_i Z_i G_s^{-1}(Z_i))^2 + g^2(x_i)]}{Cnh_n} \to 0. \end{aligned} \tag{3.23}$$

Hence, we complete the proof of Theorem 2.1 by (3.19)-(3.23). □

Proof of Theorem 2.2. Put $a_{ni}(x) = \frac{x_i - x_{i-1}}{h_n} K(\frac{x - x_i}{h_n})$. For any selected $x \in [0, 1]$, we have

$$\begin{aligned} & g_n^{(2)}(x) - Eg_n^{(1)}(x) \\ = \ & [\sum_{i=1}^{n} a_{ni} \delta_i Z_i \hat{G}_{ns}^{-1}(Z_i) I(Z_i \leq M_n) - \sum_{i=1}^{n} a_{ni} Z_{iG} I(Z_i \leq M_n)] \\ & + [\sum_{i=1}^{n} a_{ni} Z_{iG} - Eg_n^{(1)}(x)] - E\sum_{i=1}^{n} a_{ni} Z_{iG} I(Z_i > M_n) \\ & - [\sum_{i=1}^{n} a_{ni} Z_{iG} I(Z_i > M_n) - E\sum_{i=1}^{n} a_{ni} Z_{iG} I(Z_i > M_n)] \\ =: \ & B_1 + B_2 + B_3 + B_4. \end{aligned} \tag{3.24}$$

Note that, from (A.4) and (A.5) we have

$$\begin{aligned} & \sup_i EZ_{iG}^2 I(Z_i \geq M_n) \\ = & \sup_i \int_{M_n}^{\tau_{F_i}} G_s^{-2}(y) y^2 dF_i(y) \int_y^{\tau_G} dG(t) \\ = & \sup_i \int_{M_n}^{\tau_{F_i}} \frac{y^2}{1-G(y)} dF_i(y) \\ \leq & \frac{\tau_0^2}{1-G(\tau_0)} \sup_i [F_i(\tau_0) - F_i(M_n)] \to 0 \text{ as } n \to \infty. \end{aligned} \tag{3.25}$$

Therefore

$$\lim_{C\to\infty} \sup_i EZ_{iG}^2 I(Z_{iG} > C) = 0.$$

Next, we will analyse the four parts in (3.24) respectively as follows.

(i) Obviously, $B_2 = g_n^{(1)}(x) - Eg_n^{(1)}(x)$. Hence, from Theorem 2.1 we have

$$\frac{B_2}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}} \to_d N(0,1).$$

(ii) Note that

$$\begin{aligned} & \sqrt{nh_n} \max_{1\leq i\leq n} EZ_{iG} I(Z_i > M_n) \\ = & \sqrt{nh_n} \max_{1\leq i\leq n} EY_i G_s^{-1}(Y_i) I(Y_i > M_n) \\ = & \sqrt{nh_n} \max_{1\leq i\leq n} \int_{M_n}^{\tau_{F_i}} G_s^{-1}(y) y dF_i(y) \int_y^{\tau_G} dG(t) \\ = & \sqrt{nh_n} \max_{1\leq i\leq n} \int_{M_n}^{\tau_{F_i}} \frac{y(1-G(y))}{G_s(y)} dF_i(y) \\ \leq & \tau_0 \sqrt{nh_n} \max_{1\leq i\leq n} [F_i(\tau_0) - F_i(M_n)] \to 0, \end{aligned}$$

which, together with Lemma 3.1, follows that

$$\begin{aligned} & \frac{|B_3|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}} \\ \leq & C\sqrt{nh_n} \sum_{i=1}^n (\frac{x_i - x_{i-1}}{h_n}) K(\frac{x - x_i}{h_n}) EZ_{iG} I(Z_i > M_n) \to 0. \end{aligned}$$

(iii) We observe that

$$P\Big(|B_4|\Big/\sqrt{\mathrm{Var}(g_n^{(1)}(x))}\geq\varepsilon\Big)\leq P(|B_4|\geq\frac{C\varepsilon}{\sqrt{nh_n}})$$
$$\leq\ Cnh_nE[\sum_{i=1}^{n}(a_{ni}Z_{iG}I(Z_i>M_n)-Ea_{ni}Z_{iG}I(Z_i>M_n))]^2$$
$$\leq\ Cnh_n\sum_{i=1}^{n}E|a_{ni}Z_{iG}I(Z_i>M_n)-Ea_{ni}Z_{iG}I(Z_i>M_n)|^2$$
$$\leq\ C\sum_{i=1}^{n}\frac{x_i-x_{i-1}}{h_n}K^2(\frac{x-x_i}{h_n})EZ_{iG}^2I(Z_i>M_n). \tag{3.26}$$

By Lemma 3.1 we have

$$\sum_{i=1}^{n}\frac{x_i-x_{i-1}}{h_n}K^2(\frac{x-x_i}{h_n})\to\int K^2(u)du<\infty. \tag{3.27}$$

Hence, (3.25)-(3.27) follow that $P\Big(|B_4|\Big/\sqrt{\mathrm{Var}(g_n^{(1)}(x))}\geq\varepsilon\Big)\to 0$.

(iv) Note that $M_n<\tau_0<\tau_G$. By Lemma 3.1, for sufficiently large n we have

$$P\Big(\frac{|B_1|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}}\geq\varepsilon\Big)$$
$$\leq\ P\Big(\frac{1}{s_n}\sum_{i=1}^{n}\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})|Z_i\delta_i(\hat{G}_{ns}^{-1}(Z_i)-G_s^{-1}(Z_i))|I(Z_j\leq M_n)>\varepsilon\Big)$$
$$\leq\ P\Big(\frac{\tau_0}{s_n}\sup_{0\leq t\leq M_n}|\hat{G}_{ns}^{-1}(t)-G_s^{-1}(t)|\sum_{i=1}^{n}\frac{x_i-x_{i-1}}{h_n}K(\frac{x-x_i}{h_n})>\varepsilon\Big)$$
$$\leq\ P\Big(\frac{1}{s_n}\sup_{0\leq t\leq M_n}|\exp\{\log\hat{G}_{ns}^{-1}(t)\}-\exp\{\log G_s^{-1}(t)\}|>\frac{\varepsilon}{2\tau_0}\Big)$$
$$=\ P\Big(\frac{1}{s_n}\sup_{0\leq t\leq M_n}|\exp\{\theta(\log\hat{G}_{ns}^{-1}(t)-\log G_s^{-1}(t))\}\cdot(\log\hat{G}_{ns}^{-1}(t)-\log G_s^{-1}(t))|>\frac{\varepsilon}{2\tau_0}\Big)$$
$$\leq\ P\Big(\frac{1}{s_n}\exp\{\sup_{0\leq t\leq M_n}|\log\hat{G}_{ns}(t)-\log G_s(t)|\}\cdot\sup_{0\leq t\leq M_n}|\log\hat{G}_{ns}(t)-\log G_s(t)|>\frac{\varepsilon}{2\tau_0}\Big)$$

$$\begin{aligned}
&\leq P\Big(\exp\{\frac{\sup_{0\leq t\leq M_n}|\log \hat{G}_{ns}(t)-\log G_s(t)|}{s_n}\} \\
&\qquad\qquad \cdot\frac{1}{s_n}\sup_{0\leq t\leq M_n}|\log \hat{G}_{ns}(t)-\log G_s(t)|>\frac{\varepsilon}{2\tau_0}\Big) \\
&= P\Big(e^{\Delta_n}\cdot\Delta_n>\frac{\varepsilon}{2\tau_0},\Delta_n>1\Big)+P\Big(e^{\Delta_n}\cdot\Delta_n>\frac{\varepsilon}{2\tau_0},\Delta_n\leq 1\Big) \\
&\leq P(\Delta_n>1)+P(\Delta_n>\varepsilon/2\tau_0 e),
\end{aligned}$$

where $0<\theta<1$, $\Delta_n=\frac{\sup_{0\leq t\leq M_n}|\log \hat{G}_{ns}(t)-\log G_s(t)|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}}$. Therefore, according to Lemma 3.3 we have

$$P\Big(\frac{|B_1|}{\sqrt{\mathrm{Var}(g_n^{(1)}(x))}}\geq\varepsilon\Big)\to 0.$$

On combining with (3.24) and (i)-(iv), Theorem 2.2 is proved. □

Acknowledgements. This research was supported by the National Natural Science Foundation of China (No. 10571136).

References

[1] J. K. Benedetti, *On the nonparametric estimation of regression functions*, J. Roy. Statist. Soc. Ser. B **39** (1977), no. 2, 248–253.

[2] Y. Fan, *Consistent nonparametric multiple regression for dependent heterogeneous processes: the fixed design case*, J. Multivariate Anal. **33** (1990), no. 1, 72–88.

[3] A. Földes and L. Rejtö, *Strong uniform consistency for nonparametric survival curve estimators from randomly censored data*, Ann. Statist. **9** (1981), no. 1, 122–129.

[4] A. A. Georgiev, *Consistent nonparametric multiple regression: the fixed design case*, J. Multivariate Anal. **25** (1988), no. 1, 100–110.

[5] A. A. Georgiev and W. Greblicki, *Nonparametric function recovering from noisy observations*, J. Statist. Plann. Inference **13** (1986), no. 1, 1–14.

[6] E. L. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations*, J. Amer. Statist. Assoc. **53** (1958), 457–481.

[7] H. Koul, V. Susarla and J. Van Ryzin, *Regression analysis with randomly right-censored data*, Ann. Statist. **9** (1981), no. 6, 1276–1288.

[8] B. L. S. Prakasa Rao, *Asymptotic theory of statistical inference*, Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1987.

[9] M. B. Priestley and M. T. Chao, *Non-parametric function fitting*, J. Roy. Statist. Soc. Ser. B **34** (1972), 385–392.

[10] G. G. Roussas, *Consistent regression estimation with fixed design points under dependence conditions*, Statist. Probab. Lett. **8** (1989), no. 1, 41–50.

[11] G. G. Roussas, L. T. Tran, and D. A. Ioannides, *Fixed design regression for time series: asymptotic normality*, J. Multivariate Anal. **40** (1992), no. 2, 262–291.

[12] L. Tran, G. Roussas, S. Yakowitz, and B. T. Van, *Fixed-design regression for linear time series*, Ann. Statist. **24** (1996), no. 3, 975–991.

[13] Q. H. Wang, *Some convergence properties of weighted kernel estimators of regression functions under random censorship*, Acta Math. Appl. Sin. **19** (1996), no. 3, 338–350.

[14] L. G. Xue, *Strong uniform convergence rates of wavelet estimates of regression function under complete and censored data*, Acta Math. Appl. Sin. **25** (2002), no. 3, 430–438.

Si-Li Niu
Department of Applied Mathematics
Tongji University
Shanghai 200092, P. R. China
E-mail address: siliniu05@yahoo.com.cn

Qian-Ru Li
Department of Applied Mathematics
Tongji University
Shanghai 200092, P. R. China
E-mail address: leeqianru@163.com