Weighted estimation of conditional mean function with truncated, censored and dependent data

Han-Ying Liang[a], María del Carmen Iglesias-Pérez[b]
[a]School of Mathematical Science, Tongji University, Shanghai 200092, P. R. China
E-mail: hyliang@tongji.edu.cn
[b]Department of Statistics and OR, Escuela de Ingeniería Forestal Organization, Universidad de Vigo,
Campus A Xunqueira, 36005, Pontevedra, Spain. *E-mail: mcigles@uvigo.es*

Abstract. By applying the empirical likelihood method, we construct a new weighted estimator of the conditional mean function for a left-truncated and right-censored model. Assuming that the observations form a stationary α-mixing sequence, we derive weak convergence with a certain rate and prove asymptotic normality of the weighted estimator. The asymptotic normality shows that the weighted estimator preserves the bias, variance, and, more importantly, automatic good boundary behavior of a local linear estimator of the conditional mean function. Also, a Berry-Esseen type bound for the weighted estimator is established. A simulation study is conducted to study the finite sample behavior of the new estimator and a real data application is provided.

Key words and phrases: Asymptotic normality; Berry-Esseen type bound; conditional mean function; truncated and censored data; weighted estimator; α-mixing
AMS 2000 subject classifications: 62N01; 62E20

1 Introduction

Regression techniques are commonly used for exploring the relationship between response Y and covariate X. This can be done via estimating the function $m(x) = E(\phi(Y)|X = x)$, where the (known) transformation $\phi(\cdot)$ is introduced to include various targets of interest. For example, taking $\phi(y) = y^r$ we get the rth conditional moment (when $r = 1$, the 1th conditional moment is the regression function), and if we take $\phi(y) = I(y \le t)$ then $m(x)$ becomes the conditional distribution function of Y given $X = x$ at t. There is a large literature about the Nadaraya-Watson (NW) estimator of the conditional mean under complete data and different kinds of associations, like mixing processes and Markovian chains, for example, Bercu et al. (2015) for the recursive NW estimator associated with the recursive sliced inverse regression method, Huang and Su (2015) for characterizations based on regression assumptions of order statistics, Yanev and Ahsanullah (2012) for characterizations of Student's t-distribution via regressions of order statistics, Ling and Wu (2012) for modified kernel regression estimation for functional data, Goegebeur et al. (2014) for nonparametric regression estimation of conditional tails. For more details see Györfi et al. (1989) and Bosq (1998) and the references therein. Masry and Fan (1997) considered estimating $E(Y|X = x)$ for mixing sequences using the local polynomial estimator. For more references about non-parametric regression techniques with dependent data see, for example, the bibliographical notes given in Fan and Yao (2003).

In some fields as reliability, survival analysis or economics, right-censored or/and left-truncated data are often encountered. The random variable Y can be regarded as the lifetime of a device, the survival time of a patient or the duration time of some economic variable. El Ghouch and Van Keilegom (2008) discussed for the first time asymptotic normality for an estimator of $m(x)$ based on a transformed variable of Y, for the right-censored model with dependent data; Liang et al. (2011) for local polynomial estimation of the conditional mean function with dependent truncated data; Wang et al. (2013) for local polynomial quasi-likelihood regression with truncated and dependent data; de Uña-Álvarez et al. (2010) for wavelet regression with left-truncated dependent data, which is useful in the presence of discontinuities; Zou and Liang (2017) for wavelet estimation of density for censored data with censoring indicator missing at random. Recently, Liang et al. (2015) constructed the NW and the local linear (LL) estimators of $m(x)$ and its derivatives for the left truncated and right censored (LTRC) model, and established asymptotic normality for the NW and LL estimators of $m(x)$ under dependent observations. It can be seen easily that the implementation of the NW estimate of the regression function is much easier than the LL method, and the estimated values of the regression function are always within the range of the response variable. However, it is well-known that the NW method is inferior to the LL approach due to the limitations such as larger bias, non-adaptation and boundary effects (see, e.g., Fan and Gijbels (1996)). To take the advantages of both the NW method and LL estimate, Hall and Presnell (1999) proposed for the first time a weighted NW estimator of the regression function under the independent and completed samples. Liang (2012) used this method to define a weighted NW estimator for the nonparametric regression with truncated data.

We, in this paper, use empirical likelihood method to construct a new weighted estimator of $m(x)$ for the LTRC model, and investigate weak consistency and asymptotic normality of the weighted estimator under α-mixing observations. Also, a Berry-Esseen type bound for the weighted estimator is established. These results are new, even in independent cases. Theoretical results in Section 3 and simulation study in Section 4 show that the weighted estimator shares the advantages of both the NW method and LL fitting. Specifically, the new estimator reduces the bias of the NW estimator (matching that of the LL estimator).

Let (X, Y, T, W) be a random vector, where Y is the lifetime with distribution function (df) F, T is random left truncation time with the df L and W denotes random right censoring time with the df G. Assume that X admits df $V(\cdot)$ and density $v(\cdot)$. In the LTRC model one observes (X, Z, T, δ) if $Z \geq T$, where $Z = \min(Y, W)$ and $\delta = I(Y \leq W)$. When $Z < T$ nothing is observed. Clearly, if Y is independent of W, then Z has the df $H = 1 - (1 - F)(1 - G)$. Take $\theta = P(T \leq Z)$, then necessarily, we assume $\theta > 0$. Let $(X_i, Z_i, T_i, \delta_i)$, for $i = 1, 2, \cdots, n$, be a stationary random sample from (X, Z, T, δ) which one observes then $(T_i \leq Z_i, \forall i)$. We assume that Y, T and W are nonnegative random variables, as usual in survival analysis. Iglesias-Pérez and González-Manteiga (1999) defined a generalized product-limit estimator of conditional distribution function $F(y|x)$ of Y given $X = x$ for the LTRC data by

$$\widehat{F}_n(y|x) = 1 - \prod_{i=1}^{n}\Big(1 - \frac{I(Z_i \le y)\delta_i B_{ni}(x)}{\sum_{j=1}^{n} I(T_j \le Z_i \le Z_j)B_{nj}(x)}\Big), \tag{1.1}$$

where $B_{ni}(x) = K(\frac{x-X_i}{h_n})/\sum_{j=1}^{n} K(\frac{x-X_j}{h_n})$, K denotes a kernel function, and $0 < h_n \to 0$ is a bandwidth parameter. Under the i.i.d. setting, Iglesias-Pérez and González-Manteiga (1999) obtained for the first time an almost sure representation and asymptotic normality of $\widehat{F}_n(\cdot)$. Liang et al. (2012) investigated asymptotic properties of $\widehat{F}_n(\cdot)$ with α-mixing data.

In the sequel, $\{(X_i, Z_i, T_i, \delta_i) =: \zeta_i,\ 1 \le i \le n\}$ is assumed to be a stationary α-mixing sequence of random vectors. Recall that the sequence $\{\zeta_k, k \ge 1\}$ is said to be α-mixing if the α-mixing coefficient

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

converges to zero as $n \to \infty$, where $\mathcal{F}_l^m = \sigma\{\zeta_l, \zeta_{l+1}, \ldots, \zeta_m\}$ denotes the σ-algebra generated by $\zeta_l, \zeta_{l+1}, \ldots, \zeta_m$ with $l \le m$. Among various mixing conditions used in the literature, the α-mixing is reasonably weak, and has many practical applications. Many stochastic processes and time series are known to be α-mixing. Withers (1981) obtained various conditions for a linear process to be α-mixing. Also see, e.g., Doukhan (1994), page 99, for more details; and Cai and Kim (2003) for motivation in the scope of survival analysis. In fact, under very mild assumptions linear autoregressive and more generally bilinear time series models are α-mixing with mixing coefficients decaying exponentially, i.e., $\alpha(n) = O(\rho^n)$ for some $0 < \rho < 1$.

The rest of the paper is organized as follows. In the next section we introduce the estimators of $m(x)$. Main results are formulated in Section 3. A simulation study is presented in Section 4, while a real data illustration is given in Section 5. Section 6 gives the proofs of the main results. Some preliminary lemmas, which are used in the proofs of the main results, are collected in Appendix (Section 7).

2 Estimators

In the sequel, for any df $Q(y) = P(\eta \le y)$, its density function is denoted by $q(y)$, we denote the left and right support endpoints by $a_Q = \inf\{y : Q(y) > 0\}$ and $b_Q = \sup\{y : Q(y) < 1\}$, respectively.

For any given $x \in \mathbb{R}$, define $Q(y|x) = P(\eta \le y|X = x)$ and $Q^*(y) = P(\eta \le y|T \le Z)$, their density functions are denoted by $q(y|x)$ and $q^*(y)$, respectively. Set $\theta(x) = P(T \le Z|X = x)$,

$$C(y|x) = P(T \le y \le Z|X = x, T \le Z) \quad\text{and}\quad H_1^*(y|x) = P(Z \le y, \delta = 1|X = x, T \le Z).$$

We observe that $V^*(x) = P(X \le x | T \le Z) = \theta^{-1} \int_{-\infty}^{x} \theta(t) v(t) dt$, which gives $v^*(x) = \theta^{-1}\theta(x)v(x)$. In addition, assume that Y, T and W are conditionally independent at $X = x$, and $F(y|x)$ and $G(y|x)$ are continuous with respect to y then it is easy to verify that

$$C(y|x) = \theta^{-1}(x)L(y|x)(1 - G(y|x))(1 - F(y|x)) = \theta^{-1}(x)L(y|x)(1 - H(y|x))$$

and $H_1^*(y|x) = \theta^{-1}(x) \int_0^y L(t|x)(1-G(t|x))f(t|x)dt$, which gives $h_1^*(y|x) = \theta^{-1}(x)L(y|x)(1-G(y|x))f(y|x)$. The estimator of $C(y|x)$ is defined by $\widehat{C}_n(y|x) = \sum_{i=1}^n I(T_i \le y \le Z_i)B_{ni}(x)$.

From now on we only consider classes of the functions ϕ that satisfy the following conditions:

($\mathcal{H}$):($\mathcal{H}1$) ϕ vanishes outside the interval $[\tau_1, \tau_2]$ for some $a_{L(\cdot|x)} < \tau_1 \le \tau_2 < b_{H(\cdot|x)}$;

($\mathcal{H}2$) ϕ is a bounded function on $[\tau_1, \tau_2]$.

Remark 2.1 *If the lifetime Y has no left truncation (i.e., $T = 0$), then one can choose $\tau_1 = 0$ in the condition ($\mathcal{H}$), in this case, ($\mathcal{H}$) reduces to assumption ($\mathcal{H}$) of El Ghouch and Van Keilegom (2008).*

Under condition ($\mathcal{H}$), if $\Lambda(\cdot)$ is a bounded function with bounded support, and the functions $v^*(t)$ and $f(y|t)$ are continuous at $t = x$ we find

$$E\big\{\Lambda_{a_n}(X - x)[\delta\phi(Z)C^{-1}(Z|X)(1 - F(Z|X)) - m(x)]\big\} \to 0, \tag{2.1}$$

where $\Lambda_{a_n}(\cdot) = \Lambda(\cdot/a_n)/a_n$ and $0 < a_n \to 0$.

If $m(z)$ is assumed to have $(p+1)$th continuous derivative in a small neighbourhood of x, it can be approximated by a polynomial function as

$$m(z) \approx m(x) + \cdots + m^{(p)}(x)(z - x)^p/p! := \beta_0 + \cdots + \beta_p(z - x)^p.$$

Note that (2.1) suggests that $m(x)$ can be viewed as a nonparametric regression of $Y_i^* = \delta_i\phi(Z_i)C^{-1}(Z_i|X_i)(1-F(Z_i|X_i))$ on $X_i = x$. Therefore, based on the idea of the local polynomial smoother, the estimator of $(m(x), \cdots, m^{(p)}(x)/p!)^\tau$ is defined as $(\widehat{\beta}_0, \cdots, \widehat{\beta}_p)^\tau$, which minimizes

$$\sum_{i=1}^n \Big(\widehat{Y}_i^* - \sum_{j=0}^p \beta_j(X_i - x)^j\Big)^2 \Lambda_{a_n}(X_i - x), \tag{2.2}$$

where $\widehat{Y}_i^* = \delta_i\phi(Z_i)\widehat{C}_n^{-1}(Z_i|X_i)(1 - \widehat{F}_n(Z_i|X_i))$.

From (2.2), when $p = 0$, we obtain the following NW type estimator of $m(x)$

$$\widehat{m}_{NW}(x) = \sum_{i=1}^n \widehat{Y}_i^* w_i^{NW}(x) \quad \text{with} \quad w_i^{NW}(x) = \frac{\Lambda_{a_n}(X_i - x)}{\sum_{j=1}^n \Lambda_{a_n}(X_j - x)},$$

when $p = 1$, the LL estimator of $m(x)$ is $\widehat{m}_{LL}(x) = \sum_{i=1}^n \widehat{Y}_i^* w_i^{LL}(x)$, where

$$w_i^{LL}(x) = \frac{\Lambda_{a_n}(X_i - x)\{s_{n2} - (X_i - x)s_{n1}\}}{s_{n0}s_{n2} - s_{n1}^2} \quad \text{with} \quad s_{nj} = \sum_{i=1}^n (X_i - x)^j \Lambda_{a_n}(X_i - x).$$

Remark 2.2 *The estimators $\widehat{m}_{NW}(x)$ and $\widehat{m}_{LL}(x)$ are constructed for the first time in Liang et al. (2015), who investigated the asymptotic normality of $\widehat{m}_{NW}(x)$ and $\widehat{m}_{LL}(x)$ under dependent assumptions.*

It is easy to see that the LL weights $w_i^{LL}(x)$ satisfy:

$$\sum_{i=1}^{n} w_i^{LL}(x) = 1 \quad \text{and} \quad \sum_{i=1}^{n} (X_i - x) w_i^{LL}(x) = 0. \tag{2.3}$$

But for the NW type weights $w_i^{NW}(x)$ this moment condition is not fulfilled.

In view of the information (2.3), we use the empirical likelihood method to define the new weighted estimator of $m(x)$ for the LTRC model as follows. We first introduce the empirical likelihood function $\mathcal{L} = \prod_{i=1}^{n} p_i(x)$, where $p_1(x), \cdots, p_n(x)$ are subject to the restrictions:

$$p_i(x) \geq 0, \ \sum_{i=1}^{n} p_i(x) = 1 \ \text{ and } \ \sum_{i=1}^{n} p_i(x)(X_i - x)\Lambda_{a_n}(X_i - x) = 0. \tag{2.4}$$

The maximum of $\mathcal{L}$ can be found via Lagrange multipliers. It may be shown that $\mathcal{L}_{\max} = \prod_{i=1}^{n} \widehat{p}_i(x)$, where $\widehat{p}_i(x) = \frac{1}{n} \cdot \frac{1}{1+\eta(X_i - x)\Lambda_{a_n}(X_i - x)}$, $i = 1, \cdots, n$, and η is the solution of the following equation:

$$\sum_{i=1}^{n} \frac{(X_i - x)\Lambda_{a_n}(X_i - x)}{1 + \eta(X_i - x)\Lambda_{a_n}(X_i - x)} = 0. \tag{2.5}$$

Equivalently, η is chosen to minimize $L_n(\eta) = \sum_{i=1}^{n} \log\{1 + \eta(X_i - x)\Lambda_{a_n}(X_i - x)\}$, or to find the root of the equation $L_n'(\eta) = 0$ by using the New-Raphson iteration scheme.

The proposed weighted estimator of $m(x)$ is

$$\widehat{m}_n(x) = \sum_{i=1}^{n} \widehat{Y}_i^* w_{ni}(x) \ \text{ with } \ w_{ni}(x) = \frac{\widehat{p}_i(x)\Lambda_{a_n}(X_i - x)}{\sum_{j=1}^{n} \widehat{p}_j(x)\Lambda_{a_n}(X_j - x)}.$$

Remark 2.3 *Although the estimator $\widehat{m}_n(x)$ is considered only for covariate X in univariate case, we would like to mention that the basic ideas of our methodology hold for multivariate situations.*

3 Main Results

Throughout the paper, let $C, C_1, \cdots$ and $c_0, c, c_1, \cdots$ denote generic finite positive constants, whose values may change from line to line, and let $\Phi(u)$ stand for the standard normal distribution function and $[t]$ be the integer part of t. $A_n = O(B_n)$ means $|A_n| \leq C|B_n|$ and $f^{(i,j)}(x, y) := \partial^{i+j} f(x, y)/\partial x^i \partial y^j$. Let $I = [x_1, x_2]$ be an interval contained in the support of $m^*(\cdot)$ such that $\inf_{x \in I_\epsilon} m^*(x) \geq \epsilon_0 > 0$ for $I_\epsilon = [x_1 - \epsilon, x_2 + \epsilon]$ with chosen small $\epsilon > 0$.

Throughout the paper we assume that $\alpha(n) = O(n^{-\lambda})$ for some positive constant λ. In order to formulate the main results, we need the following assumptions.

(A0) $m(x)$ has 2th continuous derivative for $x \in I$.

(A1) $K(\cdot)$ is a Lipschitz-continuous density function with bounded support and $\int_{\mathbb{R}} t^j K(t)dt = 0$ for $j = 1, 2, \cdots k_0 - 1$ and $k_0 > 2$.

(A2) (i) For $s \in I_\epsilon$, Y, T and W are conditionally independent at $X = s$;

(ii) τ_1 and τ_2 are two real numbers such that $a_{L(\cdot|x)} < \tau_1 \leq \tau_2 < b_{H(\cdot|x)}$ and $a_{L(\cdot|x)} < a_{H(\cdot|x)}$ for $x \in I_\epsilon$.

(A3) Let $k_0 > 2$. The first k_0 derivatives of functions $\theta(s)$ and $v(s)$ are bounded for $s \in I_\epsilon$, and the first k_0 partial derivatives with respect to s of functions $L(y|s)$, $G(y|s)$, $F(y|s)$, $l(y|s)$, $g(y|s)$ and $f(y|s)$ are bounded for $(s, y) \in I_\epsilon \times \mathbb{R}$.

(A4) For all integers $j \geq 1$, the joint conditional density $v_j^*(\cdot, \cdot)$ of X_1 and X_{j+1} exists on $\mathbb{R} \times \mathbb{R}$ and satisfies $v_j^*(s_1, s_2) \leq C_1$ for $(s_1, s_2) \in I_\epsilon \times I_\epsilon$.

(A5) (i) $h_n^{-2}(nh_n/\ln(n))^{-(\lambda-3)/2} = O(1)$; (ii) $\sum_{n=1}^{\infty} h_n^{-2}(nh_n/\ln(n))^{-(\lambda-3)/2} < \infty$.

(A6) $\Lambda(\cdot)$ is a symmetric and bounded density function with a bounded support $[-1, 1]$.

(A7) Function $m(x)$ has continuous second order derivatives at $x \in I$.

(A8) Assume that $na_n \to \infty$, and that the sequence $\alpha(n)$ satisfies for positive integers $q := q_n$ that $q = o((na_n)^{1/2})$ and $\lim_{n\to\infty}(na_n^{-1})^{1/2}\alpha(q) = 0$.

Remark 3.1 *Conditions (A1)-(A3) and (A6)-(A7) are standard regularity conditions and used commonly in the literature, see, e.g., Iglesias-Pérez and González-Manteiga (1999) for conditions (A1)-(A3), El Ghouch and Van Keilegom (2008) for conditions (A6)-(A7); Condition (A4) is mainly technical, which is employed to simplify the calculations of covariances in the proof, these assumptions are redundant for the independent setting; the role of condition (A5) is to obtain the rate of convergence of the estimator $\widehat{F}_n(y|x)$. Condition (A8) is used to prove asymptotic normality for an α-mixing sequence.*

Set $\Delta_{ij} = \int_{\mathbb{R}} s^i \Lambda^j(s)ds$, $\sigma^2(x) = \theta(x)\int_{\mathbb{R}} \frac{\phi^2(y)f(y|x)dy}{L(y|x)(1-G(y|x))} - m^2(x)$ and $\Gamma^2(x) = \frac{\theta\Delta_{02}\sigma^2(x)}{\theta(x)v(x)}$.

Theorem 3.1 *Let $x \in I$ and $\alpha(n) = O(n^{-\lambda})$ for some $\lambda \geq 2.1$. Suppose that conditions ($\mathcal{H}$), (A0)-(A4), (A5)(i) and (A6)-(A7) are satisfied. If $na_n^{1+r_0} = O(1)$ for some constant $r_0 \geq 2$, then*

$$\widehat{m}_n(x) - m(x) = \frac{a_n^2}{2}m''(x)\Delta_{21} + O_p((na_n)^{-1/2}) + o_p(a_n^2) + O_p\Big(\Big(\frac{\ln(n)}{nh_n}\Big)^{1/2} + h_n^{k_0}\Big).$$

Theorem 3.2 *Under the assumptions of Theorem 3.1 and $\Gamma^2(x) > 0$, if (A8) holds, then*

$$\sqrt{na_n}\Big\{\widehat{m}_n(x) - m(x) - \frac{a_n^2}{2}m''(x)\Delta_{21} + o_p(a_n^2) + O_p\Big(\Big(\frac{\ln(n)}{nh_n}\Big)^{1/2} + h_n^{k_0}\Big)\Big\} \xrightarrow{\mathcal{D}} N\big(0, \Gamma^2(x)\big).$$

Remark 3.2 *(a) It may be seen from Theorem 3.1 that the weighted estimator $\widehat{m}_n(x) \to m(x)$ in probability with a rate, which implies that $\widehat{m}_n(x)$ is consistent.*

(b) From Theorem 3.2, if $a_n \ln(n)/h_n \to 0$ and $na_nh_n^{2k_0} \to 0$, it is easy to see that the asymptotic mean-squared errors (AMSE) of $\widehat{m}_n(x)$ is $AMSE(\widehat{m}_n) = a_n^4 B + (na_n)^{-1}\theta\Delta_{02}(\theta(x)v(x))^{-1}\sigma^2(x)$ with $B = \Delta_{21}^2[2^{-1}m''(x)]^2$, thus we get the aymptotic optimal bandwidth for the bandwidth a_n, $a_n^{opt} = (\frac{\theta\Delta_{02}\sigma^2(x)}{4B\theta(x)v(x)})^{1/5}n^{-1/5}$ by minimizing the AMSE of $\widehat{m}_n(x)$.

(c) Note that $\Gamma^2(x) = \Delta_{02}(v^*(x))^{-1}(\sigma_1^2(x) - m^2(x))$, where

$$\sigma_1^2(x) = \int_{\mathbb{R}} \phi^2(y)C^{-2}(y|x)(1 - F(y|x))^2 dH_1^*(y|x)$$

and $m(x) = \int_{\mathbb{R}} \phi(y)C^{-1}(y|x)(1-F(y|x))dH_1^*(y|x)$. Define the estimators $\widehat{v}_n^*(x) = \frac{1}{nh_n}\sum_{i=1}^n K\big(\frac{x-X_i}{h_n}\big)$ and $\widehat{H}_{1n}^*(y|x) = \sum_{i=1}^n I(Z_i \le y, \delta_i = 1)B_{ni}(x)$ of $v^*(x)$ and $H_1^*(y|x)$, respectively. Hence, one can get plug-in estimators $\widehat{\sigma}_{1n}^2(x) = \sum_{i=1}^n \frac{\phi^2(Z_i)\delta_i(1-\widehat{F}_n(Z_i|x))^2 B_{ni}(x)}{\widehat{C}_n^2(Z_i|x)}$ and $\widehat{m}(x) = \sum_{i=1}^n \frac{\phi(Z_i)\delta_i(1-\widehat{F}_n(Z_i|x))B_{ni}(x)}{\widehat{C}_n(Z_i|x)}$ of $\sigma_1^2(x)$ and $m(x)$, respectively. Hence, a plug-in estimator of $\Gamma^2(x)$ is $\widehat{\Gamma}_n^2(x) = \frac{\widehat{\sigma}_{1n}^2(x)-\widehat{m}^2(x)}{\widehat{v}_n^*(x)}\int_{\mathbb{R}}\Lambda^2(t)dt$.

(d) For the LL estimator $\widehat{m}_{LL}(x)$ of $m(x)$, Liang et al. (2015) proved that

$$\sqrt{na_n}\Big\{\widehat{m}_{LL}(x) - m(x) - \frac{a_n^2}{2}m''(x)\Delta_{21} + o_p(a_n^2) + O_p\Big(\Big(\frac{\ln(n)}{nh_n}\Big)^{1/2} + h_n^{k_0}\Big)\Big\} \xrightarrow{\mathcal{D}} N\big(0, \Gamma^2(x)\big).$$

Obviously, the estimators $\widehat{m}_{LL}(x)$ and $\widehat{m}_n(x)$ of $m(x)$ have same asymptotic normality, i.e., $\widehat{m}_n(x)$ matches both the asymptotic bias and the asymptotic variance of $\widehat{m}_{LL}(x)$.

In order to give a Berry-Esseen type bound for $\widehat{m}_n(x)$ to assess the quality of the normal approximation in Theorem 3.2, we need the following additional assumption.

(B) $\mu := \mu_n$ and $\nu := \nu_n$ are positive integers such that $\mu + \nu \le n$, $\mu/n \to 0$ and $\nu\mu^{-1} \to 0$.

Put $\gamma_{1n} = (na_n^5)^{1/2}$, $\gamma_{2n} = (a_n \ln(n)/h_n)^{1/4} + (na_nh_n^{2k_0})^{1/4}$, $\gamma_{3n} = a_n^{-1}(\ln(n)/(na_n))^{(\lambda-1)/2}$, $\gamma_{4n} = \nu\mu^{-1}a_n^{-\varsigma/(2+\varsigma)}u(\mu)$, $\gamma_{5n} = (\mu/n)^\beta a_n^{-\varsigma(1+\beta)/(2+\varsigma)}$, $\gamma_{6n} = n\mu^{-1}\alpha(\nu)$ and $u(\mu) = \sum_{i=\mu}^\infty \alpha^{\varsigma/(2+\varsigma)}(i)$.

Theorem 3.3 *Let $x \in I$, $\alpha(n) = O(n^{-\lambda})$ and $\sum_{n=1}^\infty \frac{1}{a_n}(\frac{\ln(n)}{na_n})^{(\lambda-1)/2} < \infty$ for some $\lambda > (2+\varsigma)/\varsigma$, where $\varsigma > 0$. Suppose that conditions $(\mathcal{H})$, (A0)-(A4), (A5)(ii), (A6)-(A7) and (B) are satisfied, and that $\Gamma^2(x) > 0$. Let $\gamma_{in} \to 0$ $(i = 1, \cdots, 6)$ for $0 < 2\beta < \varsigma$ and $\beta \le [\varsigma\lambda - (2+\varsigma)]/[2\lambda + (2+\varsigma)]$. If $na_n^{1+r_0} = O(1)$ and $\ln(n)/(na_n)^{1-2/r_0} = O(1)$ for some constant $r_0 > 2$, then for $(20\lambda - 1)/[\lambda(10\lambda - 1)] \le \rho < 1$ we have*

$$\sup_u \Big|P\Big(\frac{(na_n)^{1/2}(\widehat{m}_n(x) - m(x))}{\Gamma(x)} \le u\Big) - \Phi(u)\Big| = O\big((\mu^{-1}\nu)^{1/3} + (n^{-1}\mu)^{1/3} + a_n^{(1-\rho)/3} + (\ln(n)/(na_n))^{1/4} + \gamma_{1n} + \gamma_{2n} + \gamma_{3n} + \gamma_{4n}^{1/3} + \gamma_{5n} + \gamma_{6n}^{1/4}\big).$$

Remark 3.3 *The assumptions* $\gamma_{in} \to 0$ $(i = 1, \cdots, 6)$ *in Theorem 3.3 can be satisfied by appropriate choosing for* a_n*,* h_n*,* p *and* q *when* λ *is large enough (note that, if we replace* $\alpha(n) = O(n^{-\lambda})$ *by the exponential decay* $\alpha(n) = O(\rho^n)$ *for some* $0 < \rho < 1$*, which has been used by some authors (see Doukhan (1994)), then* λ *can be arbitrarily large).*

4 Simulation Study

In this section, we conduce a simulated study to investigate the finite sample performance of the new estimator $\widehat{m}_n(x)$ (denoted by NWP) in the case $\phi(y) = y$, and compare it with that of $\widehat{m}_{NW}(x)$ and $\widehat{m}_{LL}(x)$ (results by Liang et al (2015)). In particular, (i) we compare the global mean squared errors of estimators $\widehat{m}_n(x)$, $\widehat{m}_{NW}(x)$ and $\widehat{m}_{LL}(x)$ for different sample sizes, truncation rates, percentage of censoring and bandwidth choices, and we explore their graphical fit to the true underlying curve; (ii) we examine how good the asymptotic normality of the new estimator is by Normal-Probability-plots of $\widehat{m}_n(x)$ at the specific point $x = 1$.

In order to obtain an α-mixing observed sequence $\{X_i, Z_i, T_i, \delta_i\}$, we generate the data as in the simulated study by Liang et al. (2015), which is as follows.

(1) Drawing of the first observation $(X_1, Z_1, T_1, \delta_1)$ in the final sample.

Step 1. Draw $e_1 \sim N(0, 1)$, take $X_1 = 0.5e_1$;

Step 2. Compute Y_1 and W_1, respectively, from the model $Y_1 = \sin(\pi X_1)+\phi_1(1+0.3\cos(\pi X_1))\epsilon_1$, $W_1 = \sin(\pi X_1) + 0.5\phi_2(1 + 0.3\cos(\pi X_1)) + \phi_3(1 + 0.3\cos(\pi X_1))\tilde{\epsilon}_1$, where both ϵ_1 and $\tilde{\epsilon}_1$ are $N(0, 1)$ random variables, ϵ_1, $\tilde{\epsilon}_1$ and X_1 are mutually independent, and ϕ_i $(i = 1, 2, 3)$ are chosen (see below) to control the percentage of censoring. Take $Z_1 = \min(Y_1, W_1)$, $\delta_1 = I(Y_1 \leq W_1)$;

Step 3. Draw independently $T_1 \sim N(\mu, 1)$, where μ is adapted in order to get different values of θ. If $Z_1 < T_1$, reject the datum $(X_1, Z_1, T_1, \delta_1)$ and go back to *Step 2*; do this until $Z_1 \geq T_1$.

(2) Drawing of the second observation $(X_2, Z_2, T_2, \delta_2)$ in the final sample.

Step 4. Draw X_2 according to the AR(1) model $X_2 = \rho X_1 + 0.5e_2$, where $e_2 \sim N(0, 1)$ is independent of X_1, and $|\rho| < 1$ is some constant, which is chosen to control the dependence of the observations;

Step 5. Compute Y_2 and W_2, respectively, from the model $Y_2 = \sin(\pi X_2)+\phi_1(1+0.3\cos(\pi X_2))\epsilon_2$, $W_2 = \sin(\pi X_2) + 0.5\phi_2(1 + 0.3\cos(\pi X_2)) + \phi_3(1 + 0.3\cos(\pi X_2))\tilde{\epsilon}_2$, where both ϵ_2 and $\tilde{\epsilon}_2$ are $N(0, 1)$ random variables, and ϵ_2, $\tilde{\epsilon}_2$ and X_2 are mutually independent. Take $Z_2 = \min(Y_2, W_2)$, $\delta_2 = I(Y_2 \leq W_2)$;

Step 6. Draw independently $T_2 \sim N(\mu, 1)$. If $Z_2 < T_2$, reject the datum $(X_2, Z_2, T_2, \delta_2)$ and go back to *Step 5*; do this until $Z_2 \geq T_2$.

By replicating the process (2) above, we generate the observed data $(X_i, Z_i, T_i, \delta_i)$, $i = 1, \cdots, n$. The generating process shows that $X_i = \rho X_{i-1} + 0.5 e_i$, $Y_i = \sin(\pi X_i) + \phi_1(1 + 0.3\cos(\pi X_i))\epsilon_i$, $W_i = \sin(\pi X_i) + 0.5\phi_2(1 + 0.3\cos(\pi X_i)) + \phi_3(1 + 0.3\cos(\pi X_i))\tilde{\epsilon}_i$, $Z_i = \min(Y_i, W_i)$, $\delta_i = I(Y_i \leq W_i)$, where $e_i \sim N(0,1)$, $\epsilon_i \sim N(0,1)$, $\tilde{\epsilon}_i \sim N(0,1)$ and $T_i \sim N(\mu, 1)$, and everything is distributed conditionally on $Z_i \geq T_i$. Note that the α-mixing property of the observable X_i is immediately transferred to the $(X_i, Z_i, T_i, \delta_i)$. Hence, the true underlying regression function is $m(x) = \mathbb{E}(Y|X = x) = \sin(\pi x)$.

For the proposed estimators, we employ the kernel $K(x) = \Lambda(x) = \frac{15}{16}(1 - x^2)^2 I(|x| \leq 1)$. In addition, the parameters ϕ_i $(i = 1, 2, 3)$ allow to control the percentage of censoring (PC) which is given by

$$\text{PC} = P(Y_i > W_i | X_i = x) \overset{\phi_1=\phi_3=0.7}{=} 1 - \Phi\Big(\frac{5\sqrt{2}}{14}\phi_2\Big) = \begin{cases} 10\%, & \text{when } \phi_2 = 2.537, \\ 30\%, & \text{when } \phi_2 = 1.038, \\ 50\%, & \text{when } \phi_2 = 0. \end{cases}$$

4.1 *Comparison of consistency among the estimators of $m(x)$*

To compare the global performance of the estimators $\widehat{m}_n(x)$, $\widehat{m}_{NW}(x)$ and $\widehat{m}_{LL}(x)$, we compute for the estimator $\widehat{m}_{h,a}$ of m the global mean square error (GMSE) along $M = 500$ Monte Carlo trials and a grid of bandwidths $h := h_n$ and $a := a_n$; the GMSE is defined as

$$GMSE(h, a) = \frac{1}{Mn} \sum_{l=1}^{M} \sum_{k=1}^{n} [\widehat{m}_{h,a}(X_{k,l}) - m(X_{k,l})]^2,$$

where $X_{k,l}$ is the k-th datum in the l-th trial. The minimal values of $GMSE(h, a)$ along the grid, and the corresponding bandwidths minimizing the error, are reported in Table 1. Specifically, for $n = 100$, h_n and a_n range from 0.05 to 0.80 with a step 0.01; for $n = 300$, h_n ranges from 0.05 to 0.80 with step 0.01 and a_n ranges from 0.05 to 0.60 with step 0.01. When $h_n = 0.05$ was obtained, then the grid was expanded.

In Table 1 we see, for the new estimator $\widehat{m}_n$, that the minimum GMSE decreases as the sample size n or the no truncation proportion θ increase, and that it increases as the dependence of the observations increases (ρ increases) or as the percentage of censoring PC increases. All these features were expected, and they were also observed for NW and LL estimators by Liang et al. (2015). More interestingly, we can appreciate how the new estimator $\widehat{m}_n$ outperforms the NW estimator in most of the cases and how the local linear smoother outperforms both of them. It is also seen in Table 1 that the bandwidth a_n decreases as the sample size increases or the percentage of censoring decreases, and that, in many cases, a_n for the new estimator is between the a_n for the NW estimator and the a_n for the LL estimator. The bandwidth h_n is very similar for the three estimators and it increases as the percentage of censoring increases.

Table 1: Minimum GMSE's of $\widehat{m}_n$, $\widehat{m}_{NW}$ and $\widehat{m}_{LL}$ along $M = 500$ Monte Carlo trials, and corresponding optimal bandwidths, for several truncation rates and percentage of censoring (PC).

ρ	θ	PC	n	$\widehat{m}_n$	a_n	h_n	$\widehat{m}_{NW}$	a_n	h_n	$\widehat{m}_{LL}$	a_n	h_n
0.1	60%	10%	100	5.9923×10^{-2}	0.36	0.04	6.2453×10^{-2}	0.32	0.03	6.0215×10^{-2}	0.41	0.04
			300	3.0086×10^{-2}	0.33	0.07	3.1718×10^{-2}	0.29	0.06	2.9861×10^{-2}	0.37	0.10
		50%	100	9.7991×10^{-2}	0.45	0.48	9.3781×10^{-2}	0.43	0.49	9.4899×10^{-2}	0.52	0.48
			300	5.4613×10^{-2}	0.37	0.41	5.2431×10^{-2}	0.41	0.68	5.2705×10^{-2}	0.44	0.45
	90%	10%	100	4.3330×10^{-2}	0.40	0.22	4.4696×10^{-2}	0.36	0.34	4.1301×10^{-2}	0.46	0.26
			300	2.0164×10^{-2}	0.32	0.24	2.0731×10^{-2}	0.31	0.45	1.9873×10^{-2}	0.36	0.23
		50%	100	8.5158×10^{-2}	0.46	0.72	8.6258×10^{-2}	0.42	0.67	8.3940×10^{-2}	0.51	0.71
			300	4.5092×10^{-2}	0.38	0.53	4.3951×10^{-2}	0.39	0.68	4.3736×10^{-2}	0.47	0.66
0.5	60%	10%	100	6.0644×10^{-2}	0.39	0.05	6.2637×10^{-2}	0.36	0.05	5.9977×10^{-2}	0.44	0.07
			300	3.1768×10^{-2}	0.34	0.09	3.2710×10^{-2}	0.31	0.09	3.1498×10^{-2}	0.35	0.09

In Figure 1, we plot the theoretical curve $m(x) = \sin(\pi x)$ together with the estimators $\widehat{m}_n(x)$ (denoted by m_{NWP}), $\widehat{m}_{NW}(x)$ and $\widehat{m}_{LL}(x)$, respectively, averaged along the 500 Monte Carlo trials for the two sample sizes, $\rho = 0.1$, $PC = 10\%$, $\theta = 60\%$, a_n and h_n from Table 1. From this Figure it is seen that all estimators perform better with a larger n, and that the new NWP estimator and the LL estimator fit better the theoretical curve specially at the boundaries.

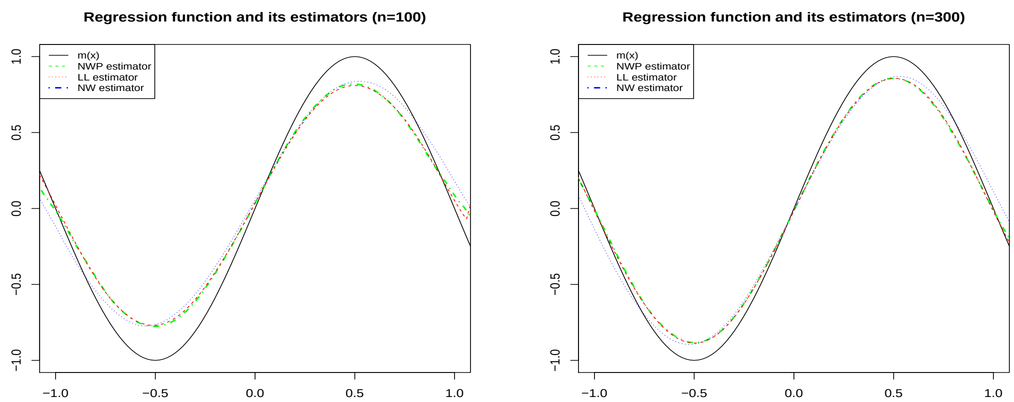

Figure 1. Regression function $m(x) = \sin(\pi x)$ and its estimators averaged along 500 Monte Carlo trials with $\rho = 0.1$, $PC = 10\%$, $\theta = 60\%$, a_n and h_n from Table 1, n =100 (left) and 300 (right).

4.2 *Asymptotic normality*

We explore how good is the asymptotic normal approximation of the estimator $\widehat{m}_n(x)$ at $x = 1$. Observe that the regression function in $x = 1$ is $sin(\pi) = 0$. In Figure 2, we plot the histograms and the Q-Q plots of $\widehat{m}_n(x)$ for $\rho = 0.1$, $PC = 10\%$, $\theta = 60\%$, a_n and h_n from Table 1, based on $M = 250$ replications with sample size $n = 100$ and 300, respectively. From Figure 2, it is seen that the normality in the distribution of the estimators is acceptable for the two sample sizes n, and it is better when n increases.

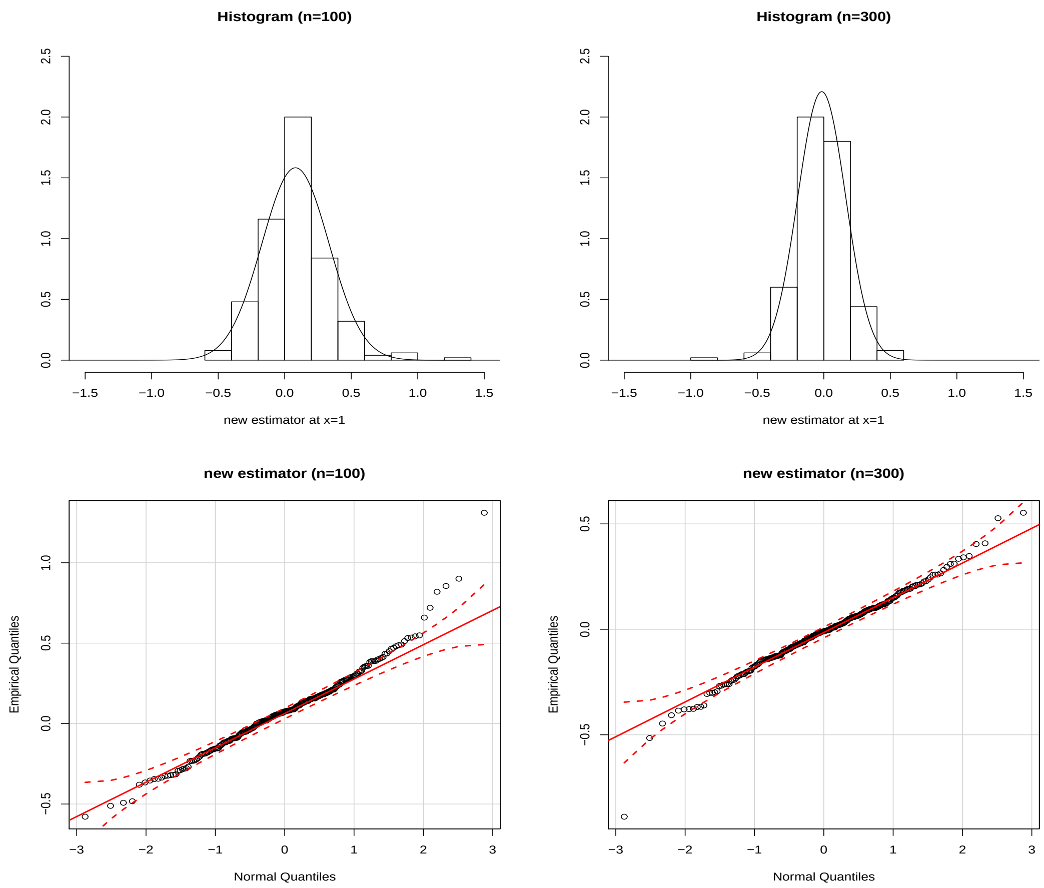

Figure 2. The histograms of $\widehat{m}_n(x)$ at $x = 1$, averaged along 250 Monte Carlo trials with $\rho = 0.1$, $PC = 10\%$, $\theta = 60\%$, a_n and h_n from Table 1, n =100 and 300, respectively.

5 Real data application

In this section, we apply the new NWP regression estimator to the analysis of Spanish unemployment data, and we compare its performance with the performance of NW and LL estimators which were studied in Liang et al. (2015).

The Spanish unemployment data set comprises 1,009 unemployment times (response variable) of married women living in Galicia (NW of Spain), recruited by means of quarterly inquires at the individuals' homes. Several covariates such as age or education are present too. The data are left-truncated because the sampled information corresponds to those women unemployed by the inquiry date. Moreover, limitation in the following-up period results in right-censoring; specifically, 56% of the unemployment times are right-censored. Additionally, the survey is conducted for periods, so that the women are grouped according to the inquiry date (38 clusters). Some dependence among the unemployment times within each cluster is expected because the specific economic situation in the corresponding period. This dependence can be modeled using the α-mixing condition assumed in this work. See de Uña-Álvarez and Iglesias-Pérez (2010) and Liang et al. (2015) for further description of this data set.

A regression analysis on the covariate 'age (in years) when entering the unemployed stock' (X) was performed. NW and LL estimators of the conditional mean function $m(x)$ were computed by Liang et al. (2015) using the biweight kernel and their bandwidths being the minimizers of the cross-validation (CV) criterion introduced in their Section 3 (with $r = 3$). Specifically, $h_n = 50$ for both estimators, and a_n took values 10.8 for the NW estimator and 10.1 for the LL estimator. Here, the new NWP estimator was computed with the same kernel, and bandwidths $h_n = 50$ and $a_n = 10.5$, for comparative purposes. An outlier corresponding to $X = 15$ was removed for the three estimations. The estimators are displayed in Figure 3. The new NWP estimator is closer to the LL estimator, taking values higher than the NW estimator at the left boundary. The three estimators provide roughly the same regression curve for X between 23 and 53 years, showing a decreasing trend of the unemployment time from 20 to 30 years, and an increasing trend between 30 and 50 years. This indicates that women with intermediate ages (around 30) at the beginning of unemployment are the ones with the shortest expected unemployment times. This is in agreement with previous results (de Uña-Álvarez and Iglesias-Pérez (2010); Liang et al. (2015)).

6 Proof of Main Results

Set $\Theta_j = \frac{1}{n}\sum_{i=1}^n \chi_{ni}^j(x)$ and $\chi_{ni}(x) = (X_i - x)\Lambda_{a_n}(X_i - x)$.

Lemma 6.1 *Let $\alpha(n) = O(n^{-\lambda})$ for some $\lambda > 0$. Suppose that (A3), (A4) and (A6) are satisfied.*

(i) Let $na_n \to \infty$ and $\lambda \geq 2.1$. If $na_n^{1+r_0} = O(1)$ for some constant $r_0 \geq 2$, then $\Theta_1 = a_n^2(v^*(x))'\Delta_{21} + O_p((\frac{a_n}{n})^{1/2}) + O(a_n^3)$, $\Theta_2 = a_n v^*(x)\Delta_{22} + O_p((\frac{a_n}{n})^{1/2}) + O(a_n^3)$, $\eta = O_p((na_n)^{-1/2} + a_n)$ and $\max_{1\leq i\leq n} |\eta\chi_{ni}(x)| = o_p(1)$;

(ii) Let $\sum_{n=1}^\infty \frac{1}{a_n}(\frac{\ln(n)}{na_n})^{(\lambda-1)/2} < \infty$ for some $\lambda > 3$. If $na_n^{1+r_0} = O(1)$ and $\ln(n)/(na_n)^{1-2/r_0} = O(1)$ for some constant $r_0 > 2$, then $\Theta_1 = a_n^2(v^*(x))'\Delta_{21} + O((\frac{a_n\ln(n)}{n})^{1/2}) + O(a_n^3)$ a.s., $\Theta_2 = a_n v^*(x)\Delta_{22} + O((\frac{a_n\ln(n)}{n})^{1/2}) + O(a_n^3)$ a.s., $\eta = O((\frac{\ln(n)}{na_n})^{1/2} + a_n)$ a.s. and $\max_{1\leq i\leq n} |\eta\chi_{ni}(x)| = o(1)$ a.s.

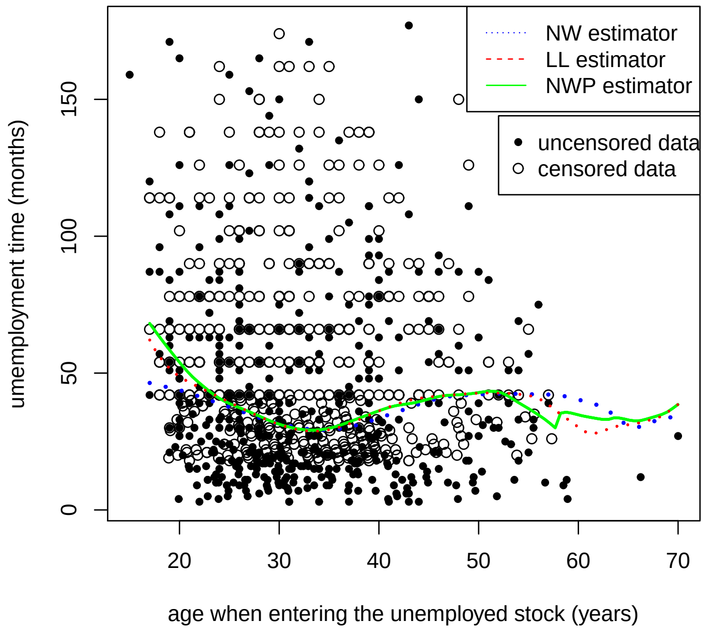

Figure 3. Estimators LL (dashed line) and NW (dotted line) based on CV bandwidths and NWP estimator (solid line) with $h_n = 50$ and $a_n = 10.5$. Uncensored and censored data are showed (black and white circles respectively).

Proof. (i) We prove only $\Theta_1 = a_n^2(v^*(x))'\Delta_{21} + O_p((a_n/n)^{1/2}) + O(a_n^3)$, the proof related to Θ_2 is similar. Note that $\Theta_1 = E\Theta_1 + O_p(\sqrt{\mathrm{Var}(\Theta_1)})$. So, we need to evaluate that $E\Theta_1$ and $\mathrm{Var}(\Theta_1)$.

In view of (A3) and (A6), we have

$$E\Theta_1 = \frac{1}{a_n}E(X_i - x)\Lambda\Big(\frac{X_i - x}{a_n}\Big) = a_n^2(v^*(x))'\Delta_{21} + O(a_n^3). \tag{6.1}$$

Write $\mathrm{Var}(\Theta_1) = \frac{1}{n}\mathrm{Var}(\chi_{n1}(x)) + \frac{2}{n^2}\sum_{1\le i<j\le n}\mathrm{Cov}(\chi_{ni}(x), \chi_{nj}(x)) := \Theta_{11} + 2\Theta_{12}$. From (A3) and (A6), it is easy to verify that $\Theta_{11} = O(a_n/n)$.

For $i < j$, applying (A3), (A4) and (A6) we have $|\mathrm{Cov}(\chi_{ni}(x), \chi_{nj}(x))| = O(a_n^2)$. On the other hand, from Lemma 7.1 (take $p = q = 20\lambda$) it follows that $|\mathrm{Cov}(\chi_{ni}(x), \chi_{nj}(x))| \le C[\alpha(j-i)]^{1-1/(10\lambda)}a_n^{1/(10\lambda)}$. Then $\Theta_{12} = O(a_n/n)$. Therefore $\mathrm{Var}(\Theta_1) = O(a_n/n)$, which, together with (6.1), implies that $\Theta_1 = a_n^2(v^*(x))'\Delta_{21} + O_p((a_n/n)^{1/2}) + O(a_n^3)$.

From (2.5), it is easy to see that

$$0 = \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\chi_{ni}(x)}{1+\eta\chi_{ni}(x)}\Big| = \Big|\frac{1}{n}\sum_{i=1}^{n}\chi_{ni}(x)\Big(1 - \frac{\eta\chi_{ni}(x)}{1+\eta\chi_{ni}(x)}\Big)\Big| \ge \frac{|\eta|\Theta_2}{1+\max_{1\le i\le n}|\eta\chi_{ni}(x)|} - |\Theta_1|. \tag{6.2}$$

Hence, applying the proof of Lemma 3 in Owen (1990), it follows that $\eta = O_p(a_n + (na_n)^{-1/2})$ and $\max_{1\le i\le n}|\eta\chi_{ni}(x)| = o_p(1)$.

(ii) From (6.1) we have

$$\Theta_1 = \frac{1}{na_n}\sum_{i=1}^{n}\Big[(X_i-x)\Lambda\Big(\frac{X_i-x}{a_n}\Big) - E\Big((X_i-x)\Lambda\Big(\frac{X_i-x}{a_n}\Big)\Big)\Big] + a_n^2(v^*(x))'\Delta_{21} + O(a_n^3).$$

Using Lemmas 7.5 and 7.6, one can prove that

$$\frac{1}{na_n}\sum_{i=1}^{n}\Big[(X_i-x)\Lambda\Big(\frac{X_i-x}{a_n}\Big) - E\Big((X_i-x)\Lambda\Big(\frac{X_i-x}{a_n}\Big)\Big)\Big] = O\Big(\Big(\frac{a_n\ln(n)}{n}\Big)^{1/2}\Big)\ a.s.$$

Similarly, one can verify that

$$\Theta_2 = \frac{1}{na_n^2}\sum_{i=1}^{n}(X_i-x)^2\Lambda^2\Big(\frac{X_i-x}{a_n}\Big) = a_n v^*(x)\Delta_{22} + O\Big(\Big(\frac{a_n\ln(n)}{n}\Big)^{1/2}\Big) + O(a_n^3)\ \ a.s.$$

Hence, from $\frac{\ln(n)}{na_n}\to 0$ and (6.2) we have $\frac{|\eta|}{1+\max_{1\le i\le n}|\eta\chi_{ni}(x)|} = O(a_n + (\frac{\ln(n)}{na_n})^{1/2})$ *a.s.* and $\eta = O(a_n + (\frac{\ln(n)}{na_n})^{1/2})$ *a.s.*, further we have $\max_{1\le i\le n}|\eta\chi_{ni}(x)| = o(1)$ *a.s.* ■

Proof of Theorem 3.1. Let $v_n^*(x) = \sum_{i=1}^n \widehat{p}_i(x)\Lambda_{a_n}(X_i-x)$ and write

$$\begin{aligned}\widehat{m}_n(x) - m(x) =& \frac{1}{v_n^*(x)}\Big\{\sum_{i=1}^{n}(\widehat{Y}_i^* - Y_i^*)\widehat{p}_i(x)\Lambda_{a_n}(X_i-x) + \sum_{i=1}^{n}(Y_i^* - m(X_i))\widehat{p}_i(x)\Lambda_{a_n}(X_i-x)\\ &+\sum_{i=1}^{n}(m(X_i)-m(x))\widehat{p}_i(x)\Lambda_{a_n}(X_i-x)\Big\}\\ :=&(v_n^*(x))^{-1}[R_{1n}(x)+R_{2n}(x)+R_{3n}(x)].\end{aligned}$$

It suffices to show that $v_n^*(x)\xrightarrow{P} v^*(x) = \theta^{-1}\theta(x)v(x)$, $R_{1n}(x) = O_p\big((\ln(n)/(nh_n))^{1/2} + h_n^{k_0}\big)$, $R_{2n}(x) = O_p((na_n)^{-1/2} + a_n^3)$ and $R_{3n}(x) = \frac{1}{2\theta}a_n^2\theta(x)v(x)m''(x)\Delta_{21} + o_p(a_n^2)$.

(I) We prove $v_n^*(x) = v^*(x) + o_p(1) = \theta^{-1}\theta(x)v(x) + o_p(1)$.

From $\widehat{p}_i(x) = \frac{1}{n}\cdot\frac{1}{1+\eta\chi_{ni}(x)} = \frac{1}{n}(1-\frac{\eta\chi_{ni}(x)}{1+\eta\chi_{ni}(x)})$ we have

$$v_n^*(x) = \frac{1}{na_n}\sum_{i=1}^{n}\Lambda\Big(\frac{X_i-x}{a_n}\Big) - \frac{\eta}{na_n}\sum_{i=1}^{n}\frac{\chi_{ni}(x)}{1+\eta\chi_{ni}(x)}\Lambda\Big(\frac{X_i-x}{a_n}\Big) := D_{1n}(x) - D_{2n}(x).$$

From (A3) and (A6) it follows that $ED_{1n}(x) = \int\Lambda(u)v^*(x+a_nu)du = v^*(x)+o(1)$. Similarly to the arguments as in (i) of the proof of Lemma 6.1 it is easy to verify that $\mathrm{Var}(D_{1n}(x))\to 0$. Therefore, $D_{1n}(x) = v^*(x) + o_p(1)$. Note that

$$|D_{2n}(x)| \le \frac{|\eta|}{1-\max_{1\le i\le n}|\eta\chi_{ni}(x)|}\cdot\frac{1}{na_n^2}\sum_{i=1}^{n}|X_i-x|\Lambda^2\Big(\frac{X_i-x}{a_n}\Big)$$

and $E\{\frac{1}{na_n^2}\sum_{i=1}^n|X_i-x|\Lambda^2(\frac{X_i-x}{a_n})\} = \int|u|\Lambda^2(u)v^*(x+a_nu)du = O(1)$. Then $D_{2n}(x) = O_p((na_n)^{-1/2} + a_n) = o_p(1)$ by Lemma 6.1(i).

(II) We prove $R_{1n}(x) = O_p\big((\ln(n)/(nh_n))^{1/2} + h_n^{k_0}\big)$.

Since $x \in I$, $[x - Ca_n, x + Ca_n] \in I_\epsilon$ for large n and $0 < C_* \le C(y|x) \le C^* < \infty$ for $(x, y) \in I_\epsilon \times [\tau_1, \tau_2]$ by $C(y|x) = \theta^{-1}(x)L(y|x)(1 - H(y|x))$ and (A2). Then from (A6) we have

$$\begin{aligned}
&|R_{1n}(x)| \\
\le& \frac{1}{1 - \max_{1\le i\le n} |\eta\chi_{ni}(x)|} \cdot \frac{1}{na_n}\Big\{ \sup_{x\in I_\epsilon} \sup_{\tau_1\le y\le\tau_2} |\widehat{F}_n(y|x) - F(y|x)| \sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})}{C(Z_i|X_i)} \\
&+ \frac{\sup_{x\in I_\epsilon} \sup_{\tau_1\le y\le\tau_2} |\widehat{C}_n(y|x) - C(y|x)|}{C_* - \sup_{x\in I_\epsilon} \sup_{\tau_1\le y\le\tau_2} |\widehat{C}_n(y|x) - C(y|x)|} \sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})(1 - F(Z_i|X_i))}{C(Z_i|X_i)} \\
&+ \frac{\sup_{x\in I_\epsilon} \sup_{\tau_1\le y\le\tau_2} |\widehat{C}_n(y|x) - C(y|x)||\widehat{F}_n(y|x) - F(y|x)|}{C_* - \sup_{x\in I_\epsilon} \sup_{\tau_1\le y\le\tau_2} |\widehat{C}_n(y|x) - C(y|x)|} \sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})}{C(Z_i|X_i)}\Big\}.
\end{aligned}$$

Note that $E\big(\frac{1}{na_n}\sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})}{C(Z_i|X_i)}\big) = O(1)$, $E\big(\frac{1}{na_n}\sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})(1-F(Z_i|X_i))}{C(Z_i|X_i)}\big) = O(1)$. Therefore, using Lemma 6.1(i) and Lemma 7.3, it follows that $R_{1n}(x) = O_p\big((\ln(n)/(nh_n))^{1/2} + h_n^{k_0}\big)$.

(III) We prove $R_{3n}(x) = \frac{1}{2\theta}a_n^2\theta(x)v(x)m''(x)\Delta_{21} + o_p(a_n^2)$.

In view of (2.4) we have $R_{3n}(x) = \frac{1}{2}\sum_{i=1}^n m''(X_i^*)(X_i - x)^2\widehat{p}_i(x)\Lambda_{a_n}(X_i - x)$, where X_i^* is between X_i and x, and from $\widehat{p}_i(x) = \frac{1}{n}(1 - \frac{\eta\chi_{ni}(x)}{1+\eta\chi_{ni}(x)})$ we write

$$R_{3n}(x) = \frac{1}{2na_n}\sum_{i=1}^n m''(X_i^*)(X_i - x)^2\Lambda\Big(\frac{X_i - x}{a_n}\Big)\Big[1 - \frac{\eta\chi_{ni}(x)}{1 + \eta\chi_{ni}(x)}\Big] := R_{31n}(x) - R_{32n}(x). \quad (6.3)$$

We observe that

$$E\Big(\frac{1}{2na_n}\sum_{i=1}^n m''(X_i)(X_i - x)^2\Lambda\Big(\frac{X_i - x}{a_n}\Big)\Big) = \frac{1}{2\theta}a_n^2\theta(x)v(x)m''(x)\int_{\mathbb{R}} s^2\Lambda(s)ds + o(a_n^2).$$

Using the similar arguments as those employed in the proof of Lemma 6.1(i) one can verify that

$$\mathrm{Var}\Big(\frac{1}{2na_n}\sum_{i=1}^n m''(X_i)(X_i - x)^2\Lambda\Big(\frac{X_i - x}{a_n}\Big)\Big) = O(a_n^3/n).$$

Hence, from (A0) and $na_n \to \infty$ we have $R_{31n}(x) = \frac{1}{2\theta}a_n^2\theta(x)v(x)m''(x)\Delta_{21} + o_p(a_n^2)$.

From $\frac{1}{2na_n}\sum_{i=1}^n |m''(X_i^*)|(X_i - x)^2\Lambda(\frac{X_i-x}{a_n}) = O_p(a_n^2)$, using Lemma 6.1 we have $|R_{32n}(x)| = o_p(a_n^2)$. Therefore, from (6.3) we obtain that $R_{3n}(x) = \frac{1}{2\theta}a_n^2\theta(x)v(x)m''(x)\Delta_{21} + o_p(a_n^2)$.

(IV) We prove $R_{2n}(x) = O_p((na_n)^{-1/2} + a_n^3)$.

From $\widehat{p}_i(x) = \frac{1}{n}\{1 - \eta\chi_{ni}(x) + \eta^2\chi_{ni}^2(x) - \frac{\eta^3\chi_{ni}^3(x)}{1+\eta\chi_{ni}(x)}\}$, we write

$$\begin{aligned}
R_{2n}(x) =& \frac{1}{na_n}\sum_{i=1}^n (Y_i^* - m(X_i))\Lambda\Big(\frac{X_i - x}{a_n}\Big)\Big[1 - \eta\chi_{ni}(x) + \eta^2\chi_{ni}^2(x) - \frac{\eta^3\chi_{ni}^3(x)}{1 + \eta\chi_{ni}(x)}\Big] \\
:=& R_{21n}(x) - R_{22n}(x) + R_{23n}(x) - R_{24n}(x).
\end{aligned}$$

It is easy to verify that $\frac{1}{na_n}\sum_{i=1}^n E\{(Y_i^* - m(X_i))\Lambda(\frac{X_i-x}{a_n})\chi_{ni}^l(x)\} = 0$ for $l = 0, 1, 2$. Note that

$$\mathrm{Var}(R_{21n}(x)) = \frac{1}{na_n^2}E\Big[(Y_i^* - m(X_i))\Lambda\Big(\frac{X_i - x}{a_n}\Big)\Big]^2$$

$$+\frac{2}{n^2a_n^2}\sum_{1\le i<j\le n}\mathrm{Cov}\Big((Y_i^*-m(X_i))\Lambda\Big(\frac{X_i-x}{a_n}\Big),(Y_j^*-m(X_j))\Lambda\Big(\frac{X_j-x}{a_n}\Big)\Big). \quad (6.4)$$

Since Y_i^* and $m(\cdot)$ are bounded by $(\mathcal{H})$, from $(\mathcal{H})$, (A3), (A4) and (A6) it follows that $\frac{1}{na_n^2}E\big[(Y_i^*-m(X_i))\Lambda(\frac{X_i-x}{a_n})\big]^2=O((na_n)^{-1})$ and similarly to the arguments as those employed in the proof of Lemma 6.1(i) we have

$$\frac{2}{n^2a_n^2}\sum_{1\le i<j\le n}\Big|\mathrm{Cov}\Big((Y_i^*-m(X_i))\Lambda\Big(\frac{X_i-x}{a_n}\Big),(Y_j^*-m(X_j))\Lambda\Big(\frac{X_j-x}{a_n}\Big)\Big)\Big|=o((na_n)^{-1}).$$

Therefore $\mathrm{Var}(R_{21n}(x))=O((na_n)^{-1})$, which gives that $R_{21n}(x)=O_p((na_n)^{-1/2})$.

Similarly to the arguments as for (6.4) one can verify that $\mathrm{Var}\{\frac{1}{na_n}\sum_{i=1}^n(Y_i^*-m(X_i))\Lambda(\frac{X_i-x}{a_n})\chi_{ni}^l(x)\}=O((na_n)^{-1})$ for $l=1,2$. Then from Lemma 6.1 we have

$$R_{22n}(x)=O_p((a_n/n)^{1/2}+(na_n)^{-1}),\ R_{23n}(x)=O_p((a_n^3/n)^{1/2}+(na_n)^{-3/2}).$$

From $(\mathcal{H})$, (A3) and (A6), it is easy to verify that $\frac{1}{na_n^4}\sum_{i=1}^n|(Y_i^*-m(X_i))(X_i-x)^3|\Lambda^4(\frac{X_i-x}{a_n})\}=O_p(1)$. Then, in view of Lemma 6.1 it follows that $|R_{24n}(x)|=O_p(a_n^3+(na_n)^{-3/2})$. ■

Proof of Theorem 3.2. From the proof of Theorem 3.1 it suffices to show that

$$\sqrt{na_n}R_{21n}(x)\xrightarrow{\mathcal{D}}N\big(0,\Delta_{02}\theta^{-1}\theta(x)v(x)\sigma^2(x)\big).$$

In fact, set $\Xi_i(x)=a_n^{-1/2}(Y_i^*-m(X_i))\Lambda(\frac{X_i-x}{a_n})$. Then $E\Xi_i(x)=0$ and $\sqrt{na_n}R_{21n}(x)=n^{-1/2}\sum_{i=1}^n\Xi_i(x)$.

Note that (A8) implies that there exists a sequence of positive integers $\delta_n\to\infty$ such that $\delta_nq=o((na_n)^{1/2})$, $\delta_n(na_n^{-1})^{1/2}\alpha(q)\to 0$. Let $\pi:=\pi_n=[\frac{n}{p+q}]$ and $p:=p_n=[(na_n)^{1/2}/\delta_n]$. Then

$$q/p\to 0,\quad \pi\alpha(q)\to 0,\quad \pi q/n\to 0,\quad p/n\to 0,\quad p/(na_n)^{1/2}\to 0.$$

Applying Bernstein's big-block and small-block procedure, Following the proof line as in Liang et al. (2015), one can prove the conclusion. ■

Proof of Theorem 3.3. From the proof of Theorem 3.1 one can write

$$\begin{aligned}\frac{(na_n)^{1/2}(\widehat{m}_n(x)-m(x))}{\Gamma(x)}&=\frac{v^*(x)}{v_n^*(x)}\cdot\frac{(na_n)^{1/2}}{v^*(x)\Gamma(x)}\cdot\frac{1}{n}\Big\{\sum_{i=1}^n(Y_i^*-m(X_i))\Lambda_{a_n}(X_i-x)\\&+\sum_{i=1}^n\frac{(\widehat{Y}_i^*-Y_i^*)\Lambda_{a_n}(X_i-x)}{1+\eta\chi_{ni}(x)}-\eta\sum_{i=1}^n\frac{(Y_i^*-m(X_i))\Lambda_{a_n}(X_i-x)\chi_{ni}(x)}{1+\eta\chi_{ni}(x)}\\&+\frac{1}{2}\sum_{i=1}^n[m''(X_i^*)(X_i-x)^2\Lambda_{a_n}(X_i-x)-E(m''(X_i^*)(X_i-x)^2\Lambda_{a_n}(X_i-x))]\\&-\frac{\eta}{2}\sum_{i=1}^n\frac{m''(X_i^*)(X_i-x)^2\Lambda_{a_n}(X_i-x)\chi_{ni}(x)}{1+\eta\chi_{ni}(x)}+\frac{1}{2}\sum_{i=1}^nE(m''(X_i^*)(X_i-x)^2\Lambda_{a_n}(X_i-x))\Big\}\\&:=\frac{v^*(x)}{v_n^*(x)}[I_{1n}(x)+I_{2n}(x)-I_{3n}(x)+I_{4n}(x)-I_{5n}(x)+I_{6n}(x)],\end{aligned}$$

where X_i^* is between X_i and x.

Let $\gamma_{7n} = (\ln(n)/(na_n))^{1/4} + a_n^{1/2}$, $\gamma_{8n} = (\ln(n)/(na_n))^{1/3} + (na_n^7)^{1/4} + a_n^{2/3}$, $\gamma_{9n} = a_n^{4/3}$ and $\gamma_{10n} = (a_n^4 \ln(n))^{1/4} + (na_n^7)^{1/4}$. Using Lemma 7.7 we have

$$\begin{aligned}
&\sup_u \Big| P\Big(\frac{(na_n)^{1/2}(\widehat{m}_n(x) - m(x))}{\Gamma(x)} \le u\Big) - \Phi(u)\Big| \\
\le& \sup_u |P(I_{1n}(x) \le u) - \Phi(u)| + C\Big(\gamma_{2n} + \sum_{k=7}^{10} \gamma_{kn} + |I_{6n}(x)|\Big) + P(|I_{2n}(x)| > c\gamma_{2n}) \\
&+ P\Big(\Big|\frac{v_n^*(x)}{v^*(x)} - 1\Big| > c\gamma_{7n}\Big) + \sum_{k=3}^{5} P(|I_{kn}(x)| > c\gamma_{(k+5)n}).
\end{aligned}$$

Then, it suffices to show that

$$P\Big(\Big|\frac{v_n^*(x)}{v^*(x)} - 1\Big| > c\gamma_{7n}\Big) = O(\gamma_{7n} + \gamma_{3n}),\ P(|I_{2n}(x)| > c\gamma_{2n}) = O(\gamma_{2n}),\ P(|I_{3n}(x)| > c\gamma_{8n}) = O(\gamma_{8n}),$$
$$P(|I_{4n}(x)| > c\gamma_{9n}) = O(\gamma_{9n}),\ P(|I_{5n}(x)| > c\gamma_{10n}) = O(\gamma_{10n}),\ I_{6n}(x) = O((na_n^5)^{1/2})$$

and $\sup_u |P(I_{1n}(x) \le u) - \Phi(u)| = O\big((\mu^{-1}\nu)^{1/3} + (n^{-1}\mu)^{1/3} + a_n^{(1-\rho)/3} + \gamma_{4n}^{1/3} + \gamma_{5n} + \gamma_{6n}^{1/4}\big)$.

(V) We prove $P(|\frac{v_n^*(x)}{v^*(x)} - 1| > c\gamma_{7n}) = O(\gamma_{7n} + \gamma_{3n})$. Write

$$\begin{aligned}
v_n^*(x) - v^*(x) =& \frac{1}{na_n}\sum_{i=1}^n \Big[\Lambda\Big(\frac{X_i - x}{a_n}\Big) - E\Big(\Lambda\Big(\frac{X_i - x}{a_n}\Big)\Big)\Big] + \Big[\frac{1}{a_n}E\Big(\Lambda\Big(\frac{X_1 - x}{a_n}\Big)\Big) - v^*(x)\Big] \\
&- \frac{\eta}{na_n}\sum_{i=1}^n \frac{\chi_{ni}(x)}{1 + \eta\chi_{ni}(x)}\Lambda\Big(\frac{X_i - x}{a_n}\Big).
\end{aligned}$$

Note that (A3) and (A6) imply that $|\frac{1}{a_n}E(\Lambda(\frac{X_1-x}{a_n})) - v^*(x)| \le Ca_n^2$, and from Lemma 6.1(ii) we have $E\big|\frac{\eta}{na_n}\sum_{i=1}^n \frac{\chi_{ni}(x)}{1+\eta\chi_{ni}(x)}\Lambda\big(\frac{X_i-x}{a_n}\big)\big| \le C\big((\ln(n)/(na_n))^{1/2} + a_n\big)$. Following the proof line as in Lemma 6.1(ii), applying Lemmas 7.5-7.6 we have

$$P\Big(\frac{1}{na_n}\Big|\sum_{i=1}^n \Big[\Lambda\Big(\frac{X_i - x}{a_n}\Big) - E\Big(\Lambda\Big(\frac{X_i - x}{a_n}\Big)\Big)\Big]\Big| > c_0\Big(\frac{\ln(n)}{na_n}\Big)^{1/4}\Big) \le \frac{C}{a_n}\Big(\frac{\ln(n)}{na_n}\Big)^{(\lambda-1)/2} = O(\gamma_{3n}).$$

(VI) We prove $P(|I_{2n}(x)| > c\gamma_{2n}) = O(\gamma_{2n})$. From the proof in (II) above we have

$$\begin{aligned}
&|I_{2n}(x)| = \frac{(na_n)^{1/2}}{v^*(x)\Gamma(x)}|R_{1n}(x)| \\
\le& \frac{(na_n)^{1/2}}{v^*(x)\Gamma(x)(1 - \max_{1\le i\le n}|\eta\chi_{ni}(x)|)} \cdot \frac{1}{na_n}\Big\{\sup_{x\in I_\epsilon}\sup_{\tau_1\le y\le\tau_2}|\widehat{F}_n(y|x) - F(y|x)|\sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})}{C(Z_i|X_i)} \\
&+ \frac{\sup_{x\in I_\epsilon}\sup_{\tau_1\le y\le\tau_2}|\widehat{C}_n(y|x) - C(y|x)|}{C_* - \sup_{x\in I_\epsilon}\sup_{\tau_1\le y\le\tau_2}|\widehat{C}_n(y|x) - C(y|x)|}\sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})(1 - F(Z_i|X_i))}{C(Z_i|X_i)} \\
&+ \frac{\sup_{x\in I_\epsilon}\sup_{\tau_1\le y\le\tau_2}|\widehat{C}_n(y|x) - C(y|x)||\widehat{F}_n(y|x) - F(y|x)|}{C_* - \sup_{x\in I_\epsilon}\sup_{\tau_1\le y\le\tau_2}|\widehat{C}_n(y|x) - C(y|x)|}\sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})}{C(Z_i|X_i)}\Big\}.
\end{aligned}$$

Since $E(\frac{1}{na_n}\sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})}{C(Z_i|X_i)}) = O(1)$ and $E(\frac{1}{na_n}\sum_{i=1}^n \frac{\delta_i|\phi(Z_i)|\Lambda(\frac{X_i-x}{a_n})(1-F(Z_i|X_i))}{C(Z_i|X_i)}) = O(1)$, from Lemma 6.1(ii) and Lemma 7.3 we obtain that $P(|I_{2n}(x)| > c\gamma_{2n}) = O(\gamma_{2n})$ with $\gamma_{2n} = (a_n \ln(n)/h_n)^{1/4} + (na_n h_n^{2k_0})^{1/4}$.

(VII) We prove $P(|I_{3n}(x)| > c\gamma_{8n}) = O(\gamma_{8n})$. Write

$$\begin{aligned} I_{3n}(x) =& \frac{(na_n)^{1/2}}{v^*(x)\Gamma(x)}\Big[\frac{\eta}{na_n}\sum_{i=1}^n (Y_i^* - m(X_i))\Lambda\Big(\frac{X_i-x}{a_n}\Big)\chi_{ni}(x) \\ &- \frac{\eta^2}{na_n}\sum_{i=1}^n (Y_i^* - m(X_i))\Lambda\Big(\frac{X_i-x}{a_n}\Big)\chi_{ni}^2(x) + \frac{\eta^3}{na_n}\sum_{i=1}^n \frac{(Y_i^* - m(X_i))\Lambda(\frac{X_i-x}{a_n})\chi_{ni}^3(x)}{1+\eta\chi_{ni}(x)}\Big] \\ :=& I_{31n}(x) - I_{32n}(x) + I_{33n}(x). \end{aligned} \tag{6.5}$$

From (IV) above, we have $E\{\frac{1}{na_n}\sum_{i=1}^n (Y_i^* - m(X_i))\Lambda(\frac{X_i-x}{a_n})\chi_{ni}^l(x)\}^2 = O((na_n)^{-1})$ for $l = 1, 2$. Then, using Lemma 6.1(ii), it follows that $EI_{31n}^2(x) = O(\frac{\ln(n)}{na_n} + a_n^2)$ and $EI_{32n}^2(x) = O((\frac{\ln(n)}{na_n})^2 + a_n^4)$, which imply that

$$P(|I_{31n}(x) - I_{32n}(x)| > c((\ln(n)/(na_n))^{1/3} + a_n^{2/3})) = O((\ln(n)/(na_n))^{1/3} + a_n^{2/3})). \tag{6.6}$$

Note that $E|I_{33n}(x)| \le E\{\frac{(na_n)^{1/2}|\eta|^3}{v^*(x)\Gamma(x)(1-\max_{1\le i\le n}|\eta\chi_{ni}(x)|)} \cdot \frac{1}{na_n^4}\sum_{i=1}^n |(Y_i^* - m(X_i))(X_i - x)^3|\Lambda^4(\frac{X_i-x}{a_n})\} = O(\ln^{3/2}(n)/(na_n)) + (na_n^7)^{1/2})$, which gives that

$$P(|I_{33n}(x)| > c((\ln^{3/2}(n)/(na_n))^{1/2} + (na_n^7)^{1/4})) = O((\ln^{3/2}(n)/(na_n))^{1/2} + (na_n^7)^{1/4}). \tag{6.7}$$

From (6.5)-(6.7) and $(\ln(n)/(na_n))^{1/3} > (\ln^{3/2}(n)/(na_n))^{1/2}$ we find

$$P(|I_{3n}(x)| > c((\ln(n)/(na_n))^{1/3} + (na_n^7)^{1/4} + a_n^{2/3})) = O((\ln(n)/(na_n))^{1/3} + (na_n^7)^{1/4} + a_n^{2/3}).$$

(VIII) We evaluate $|I_{6n}(x)|$, and prove that $P(|I_{4n}(x)| > c\gamma_{9n}) = O(\gamma_{9n})$ and $P(|I_{5n}(x)| > c\gamma_{10n}) = O(\gamma_{10n})$. Write

$$\begin{aligned} EI_{4n}^2(x) =& \frac{1}{4(v^*(x)\Gamma(x))^2}\Big\{\frac{1}{a_n}\mathrm{Var}\Big(m''(X_i^*)(X_i - x)^2\Lambda\Big(\frac{X_i-x}{a_n}\Big)\Big) \\ &+ \frac{2}{na_n}\sum_{1\le i<j\le n}\mathrm{Cov}\Big(m''(X_i^*)(X_i - x)^2\Lambda\Big(\frac{X_i-x}{a_n}\Big), m''(X_j^*)(X_j - x)^2\Lambda\Big(\frac{X_j-x}{a_n}\Big)\Big)\Big\}. \end{aligned}$$

The proof in (III) shows that $I_{6n}(x) = O((na_n^5)^{1/2})$, and $EI_{4n}^2(x) = O(a_n^4)$, which gives that $P(|I_{4n}(x)| > ca_n^{4/3}) = O(a_n^{4/3})$. Note that using Lemma 6.1(ii) we obtain that $E|I_{5n}(x)| = O((a_n^4\ln(n))^{1/2} + (na_n^7)^{1/2})$, which gives that $P(|I_{5n}(x)| > c((a_n^4\ln(n))^{1/4} + (na_n^7)^{1/4})) = O((a_n^4\ln(n))^{1/4} + (na_n^7)^{1/4})$.

(IX) We verify $\sup_u|P(I_{1n}(x) \le u) - \Phi(u)| = O((\mu^{-1}\nu)^{1/3} + (n^{-1}\mu)^{1/3} + a_n^{(1-\rho)/3} + \gamma_{4n}^{1/3} + \gamma_{5n} + \gamma_{6n}^{1/4})$. Let $\xi_i(x) = (v^*(x)\Gamma(x))^{-1}a_n^{-1/2}(Y_i^* - m(X_i))\Lambda(\frac{X_i-x}{a_n})$. Then $E\xi_i(x) = 0$ and $I_{1n}(x) = n^{-1/2}\sum_{i=1}^n \xi_i(x)$. Put $w := w_n = [\frac{n}{\mu+\nu}]$ and define

$$d_{mn}(x) = \sum_{i=k_m}^{k_m+\mu-1}\xi_i(x),\ d'_{mn}(x) = \sum_{i=l_m}^{l_m+\nu-1}\xi_i(x),\ d''_{wn}(x) = \sum_{i=w(\mu+\nu)+1}^{n}\xi_i(x),$$

where $k_m = (m-1)(\mu+\nu)+1$, $l_m = (m-1)(\mu+\nu)+\mu+1$, $m = 1, \cdots, w$. Then

$$I_{1n}(x) = n^{-1/2}\Big\{\sum_{m=1}^{w} d_{mn}(x) + \sum_{m=1}^{w} d'_{mn}(x) + d''_{wn}(x)\Big\} := n^{-1/2}\big\{\Omega'_n(x) + \Omega''_n(x) + \Omega'''_n(x)\big\}.$$

Hence, applying Lemma 7.7, it follows that

$$\sup_u |P(I_{1n}(x) \le u) - \Phi(u)| \le \sup_u |P(n^{-1/2}\Omega'_n(x) \le u) - \Phi(u)| + P(n^{-1/2}|\Omega''_n(x)| > \tau_{1n}^{1/3})$$
$$+P(n^{-1/2}|\Omega'''_n(x)| > \tau_{2n}^{1/3}) + (2\pi)^{-1/2}(\tau_{1n}^{1/3} + \tau_{2n}^{1/3}),$$

where $\tau_{1n} = \nu\mu^{-1} + a_n^{1-\rho} + \gamma_{4n}$, $\tau_{2n} = \mu n^{-1} + a_n^{1-\rho}$. Next we need only to prove that

$$n^{-1}E(\Omega''_n(x))^2 = O(\tau_{1n}),\ n^{-1}E(\Omega'''_n(x))^2 = O(\tau_{2n}), \tag{6.8}$$
$$\sup_u |P(n^{-1/2}\Omega'_n(x) \le u) - \Phi(u)| = O(\mu^{-1}\nu + n^{-1}\mu + a_n^{1-\rho} + \gamma_{5n} + \gamma_{6n}^{1/4}). \tag{6.9}$$

(i) We verify (6.8). Note that

$$\frac{1}{n}E(\Omega''_n(x))^2 = \frac{1}{n}\sum_{m=1}^{w}\sum_{i=l_m}^{l_m+\nu-1}\mathrm{Var}(\xi_i(x)) + \frac{2}{n}\sum_{m=1}^{w}\sum_{l_m\le i<j\le l_m+\nu-1}\mathrm{Cov}(\xi_i(x),\xi_j(x))$$
$$+\frac{2}{n}\sum_{1\le i<j\le w}\mathrm{Cov}(d'_{in}(x), d'_{jn}(x)). \tag{6.10}$$

The proof of of Theorem 3.2 shows that $\mathrm{Var}(\xi_i(x)) = (v^*(x)\Gamma(x))^{-2}\mathrm{Var}(\Xi(x)) = 1$, and $|\mathrm{Cov}(\xi_i(x),\xi_j(x))| \le C\min\big\{a_n, [\alpha(j-i)]^{1-1/(10\lambda)}a_n^{-(1-1/(10\lambda))}\big\}$ for $i<j$. Let $c_n = a_n^{-\rho}$ for $(20\lambda-1)/[\lambda(10\lambda-1)] \le \rho < 1$, then

$$\frac{1}{n}\sum_{1\le i<j\le n}|\mathrm{Cov}(\xi_i(x),\xi_j(x))| \le C\big\{c_n a_n + a_n^{-(1-1/(10\lambda))}c_n^{-(\lambda-11/10)}\big\} = O(a_n^{1-\rho}). \tag{6.11}$$

Using Lemma 7.1 again we have

$$\frac{1}{n}\Big|\sum_{1\le i<j\le w}\mathrm{Cov}(d'_{in}(x), d'_{jn}(x))\Big| \le C\nu\mu^{-1}a_n^{-\varsigma/(2+\varsigma)}u(\mu) = O(\gamma_{4n}). \tag{6.12}$$

From (6.10)-(6.12) it follows that $n^{-1}E(\Omega''_n(x))^2 = O(\nu\mu^{-1} + a_n^{1-\rho} + \gamma_{4n}) = O(\tau_{1n})$ and

$$\frac{1}{n}E(\Omega'''_n(x))^2 = \frac{1}{n}\sum_{i=w(\mu+\nu)+1}^{n}E\xi_i^2(x) + \frac{2}{n}\sum_{w(\mu+\nu)+1\le i<j\le n}\mathrm{Cov}(\xi_i(x),\xi_j(x)) = O(\tau_{2n}).$$

(ii) We prove (6.9). Let $e_{mn}(x), m = 1, 2, \cdots, w$ be independent random variables where the distribution of $e_{mn}(x)$ is the same as that of $d_{mn}(x)$ for $m = 1, 2, \cdots, w$. Put $U_n(x) = n^{-1/2}\sum_{m=1}^{w} e_{mn}(x)$ and $s_n^2 = n^{-1}\sum_{m=1}^{w} Ed_{mn}^2(x)$. Then

$$\sup_u |P(n^{-1/2}\Omega'_n(x) \le u) - \Phi(u)| \le \sup_u |P(n^{-1/2}\Omega'_n(x) \le u) - P(n^{-1/2}U_n(x) \le u)|$$
$$+\sup_u |P(n^{-1/2}U_n(x) \le u) - \Phi(u/s_n)| + \sup_u |\Phi(u/s_n) - \Phi(u)|. \tag{6.13}$$

From $E\xi_i^2(x)=1$ and (6.11) it follows that $s_n^2=1+O(\mu^{-1}\nu+n^{-1}\mu+a_n^{1-\rho})$, which implies that $s_n^2\to 1$ and

$$\sup_u|\Phi(u/s_n)-\Phi(u)|=O(|s_n^2-1|)=O(\mu^{-1}\nu+n^{-1}\mu+a_n^{1-\rho}). \tag{6.14}$$

By Berry-Esseen inequality (see Petrov (1995), page 154, Theorem 5.7), for $l>2$ there exists some constant $C>0$ such that

$$\sup_u|P(n^{-1/2}U_n(x)\le u)-\Phi(u/s_n)|\le\frac{C}{n^{l/2}s_n^l}\sum_{m=1}^{w}E|e_{mn}(x)|^l. \tag{6.15}$$

Taking $l=2(1+\beta)$, $\mu=\varsigma-2\beta$, then $l+\mu=2+\varsigma$. Note that $\beta\le[\varsigma\lambda-(2+\varsigma)]/[2\lambda+(2+\varsigma)]$ implies that $\lambda\ge(1+\beta)(2+\varsigma)/(\varsigma-2\beta)=l(l+\mu)/2\mu$. Then, using Lemma 7.4 (take $p=l$ and $q=l+\mu$) and $E|\xi_1(x)|^{2+\varsigma}\le Ca_n^{-\varsigma/2}$, we have $\sum_{m=1}^{w}E|e_{mn}(x)|^l=O\big(n\mu^\beta a_n^{-\varsigma(1+\beta)/(2+\varsigma)}\big)$, which, together with (6.15), yields that

$$\sup_u|P(n^{-1/2}U_n(x)\le u)-\Phi(u/s_n)|=O\big(n^{-(1+\beta)}n\mu^\beta a_n^{-\varsigma(1+\beta)/(2+\varsigma)}\big)=O(\gamma_{5n}). \tag{6.16}$$

Assume that $\varphi(t)$ and $\psi(t)$ are the characteristic functions of $n^{-1/2}\Omega_n'(x)$ and $n^{-1/2}U_n(x)$, respectively. By Esseen inequality (see Petrov (1995), page 146, Theorem 5.3), for any $\Upsilon>0$ we have

$$\begin{aligned}&\sup_u|P(n^{-1/2}\Omega_n'(x)\le u)-P(n^{-1/2}U_n(x)\le u)|\\ \le&\int_{-\Upsilon}^{\Upsilon}\Big|\frac{\varphi(t)-\psi(t)}{t}\Big|dt+\Gamma\sup_u\int_{|v|\le\frac{C}{\Upsilon}}|P(n^{-1/2}U_n(x)\le u+v)-P(n^{-1/2}U_n(x)\le u)|dv\\ :=&J_{1n}+J_{2n}.\end{aligned} \tag{6.17}$$

Using Lemma 7.8, we have $|\varphi(t)-\psi(t)|\le C|t|\alpha^{1/2}(\nu)n^{-1/2}\sum_{m=1}^{w}\big\{E\big|\sum_{i=k_m}^{k_m+\mu-1}\xi_i(x)\big|^2\big\}^{1/2}$. From $E\xi_i^2(x)=1$ and $|\mathrm{Cov}(\xi_i(x),\xi_j(x))|\le C\min\big\{a_n,[\alpha(j-i)]^{1-1/(10\lambda)}a_n^{-(1-1/(10\lambda))}\big\}$ for $i<j$ we have $E\big|\sum_{i=k_m}^{k_m+\mu-1}\xi_i(x)\big|^2=O(\mu)$. Thus $J_{1n}=O(\Upsilon(w^2n^{-1}\mu\alpha(\nu))^{1/2})=O(\Upsilon(n\mu^{-1}\alpha(\nu))^{1/2})=O(\Upsilon\gamma_{6n}^{1/2})$. From (6.16) we have

$$\begin{aligned}&\sup_u|P(n^{-1/2}U_n(x)\le u+v)-P(n^{-1/2}U_n(x)\le u)|\\ \le&\sup_u|P(n^{-1/2}U_n(x)\le u+v)-\Phi((u+v)/s_n)|+\sup_u|P(n^{-1/2}U_n(x)\le u)-\Phi(u/s_n)|\\ &+\sup_u|\Phi((u+v)/s_n)-\Phi(u/s_n)|=O(\gamma_{5n})+O(|v|/s_n),\end{aligned}$$

which yields that $J_{2n}=O(\gamma_{5n}+1/\Upsilon)$. Choose $\Upsilon=\gamma_{6n}^{-1/4}$. Then from (6.17) we have

$$\sup_u|P(n^{-1/2}\Omega_n'(x)\le u)-P(n^{-1/2}U_n(x)\le u)|=O(\gamma_{5n}+\gamma_{6n}^{1/4}). \tag{6.18}$$

Therefore, from (6.13), (6.14), (6.16) and (6.18) we have

$$\sup_u|P(n^{-1/2}\Omega_n'(x)\le u)-\Phi(u)|=O(\mu^{-1}\nu+n^{-1}\mu+a_n^{1-\rho}+\gamma_{5n}+\gamma_{6n}^{1/4}).$$

■

7 Appendix

In this section, we give some preliminary Lemmas, which have been used in Section 6. Let $\{Z_i, i \geq 1\}$ be a sequence of α-mixing real random variables with the mixing coefficients $\{\alpha(k)\}$.

Lemma 7.1 (Hall and Heyde (1980), Corollary A.2, page 278) *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*

$$|EXY - EXEY| \leq 8\|X\|_p\|Y\|_q\Big\{\sup_{A\in\sigma(X),B\in\sigma(Y)}|P(AB) - P(A)P(B)|\Big\}^{1-p^{-1}-q^{-1}}.$$

Lemma 7.2 (Volkonskii and Rozanov (1959)) *Let $V_1, \cdots, V_m$ be α-mixing random variables measurable with respect to the σ-algebra $\mathcal{F}_{i_1}^{j_1}, \cdots, \mathcal{F}_{i_m}^{j_m}$, respectively, with $1 \leq i_1 < j_1 < \cdots < j_m \leq n$, $i_{l+1} - j_l \geq w \geq 1$ and $|V_j| \leq 1$ for $l, j = 1, 2, \cdots, m$. Then $|E(\prod_{j=1}^m V_j) - \prod_{j=1}^m EV_j| \leq 16(m-1)\alpha(w)$, where $\mathcal{F}_a^b = \sigma\{V_i, a \leq i \leq b\}$ and $\alpha(w)$ is the mixing coefficient.*

Lemma 7.3 *Let $\alpha(n) = O(n^{-\lambda})$ for some $\lambda > 2$. Suppose that conditions (A1)-(A4) are satisfied.*

(a) If (A5)(i) holds, then $\sup_{x\in I_\epsilon}\sup_{\tau_1\leq y\leq\tau_2}|\widehat{C}_n(y|x) - C(y|x)| = O_p(\max\{(\ln(n)/(nh_n))^{1/2}, h_n^{k_0}\})$ and $\sup_{x\in I_\epsilon}\sup_{\tau_1\leq y\leq\tau_2}|\widehat{F}_n(y|x) - F(y|x)| = O_p\big(\max\big\{(\ln(n)/(nh_n))^{1/2}, h_n^{k_0}\big\}\big)$.

(b) If (A5)(ii) holds, then $\sup_{x\in I_\epsilon}\sup_{\tau_1\leq y\leq\tau_2}|\widehat{C}_n(y|x)-C(y|x)| = O(\max\{(\ln(n)/(nh_n))^{1/2}, h_n^{k_0}\})$ a.s. and $\sup_{x\in I_\epsilon}\sup_{\tau_1\leq y\leq\tau_2}|\widehat{F}_n(y|x) - F(y|x)| = O\big(\max\big\{(\ln(n)/(nh_n))^{1/2}, h_n^{k_0}\big\}\big)$ a.s.

Proof. The proof of Lemma 7.3 comes from Theorem 2.1 of Liang et al. (2012).

Lemma 7.4 (Shao and Yu (1996), Theorem 4.1) *Let $2 < p < q \leq \infty$. Assume that $EZ_i = 0$ and $\alpha(n) = O(n^{-\gamma})$ for $\gamma > 0$. Then there exists $Q = Q(p, q, \gamma) < \infty$ such that $E|\sum_{i=1}^n Z_i|^p \leq Qn^{p/2}\max_{1\leq i\leq n}\|Z_i\|_q^p$ if $\gamma \geq pq/[2(q-p)]$.*

Lemma 7.5 (Liebscher (2001), Proposition 5.1) *Assume that $EZ_i = 0$ and $|Z_i| \leq S < \infty$ a.s. ($i = 1, 2, \cdots, n$). Set $D_N = \max_{1\leq j\leq 2N}\mathrm{Var}(\sum_{i=1}^j Z_i)$. Then, for n, $N \in \mathbb{N}$, $0 < N \leq n/2$, $\varepsilon > 0$, $P\big(\big|\sum_{i=1}^n Z_i\big| > \varepsilon\big) \leq 4\exp\big\{-\frac{\varepsilon^2}{16}\big(nN^{-1}D_N + \frac{1}{3}\varepsilon SN\big)^{-1}\big\} + 32\frac{S}{\varepsilon}n\alpha(N)$.*

Lemma 7.6 (Liebscher (1996), Lemma 2.3) *Assume $\alpha(n) \leq C_1n^{-q}$ for some $q > 1$, $C_1 > 0$. Let $\sup_{1\leq i,j\leq n,i\neq j}|\mathrm{Cov}(Z_i, Z_j)| := R^*(n) < \infty$ be satisfied. Moreover, let $R_m(n) < \infty$ for some m, $2q/(q-1) < m \leq \infty$, where $R_m(n) = \sup_{1\leq i\leq n}(E|Z_i|^m)^{1/m}$, for $1 \leq m < \infty$, and $R_\infty(n) = \sup_{1\leq i\leq n}\operatorname{ess\,sup}_{w\in\Omega}|Z_i|$. Then $Var\big(\sum_{i=1}^n Z_i\big) \leq n\big\{C_2(q, m)(R_m(n))^{2m/(q(m-2))}(R^*(n))^{1-m/(q(m-2))} + R_2^2(n)\big\}$ holds with $C_2(\gamma, m) := \frac{20q-40q/m}{q-1-2q/m}C_1^{1/q}$.*

Lemma 7.7 *Let X, V and $Y_1, \cdots, Y_m$ be random variables, then for positive numbers a, $w_1, \cdots, w_m$ we have $\sup_u|P(X \leq uV) - \Phi(u)| \leq \sup_u|P(X \leq u) - \Phi(u)| + P(|V-1| > a) + a$ and*

$$\sup_u\Big|P\Big(X + \sum_{i=1}^m Y_i \leq u\Big) - \Phi(u)\Big| \leq \sup_u|P(X \leq u) - \Phi(u)| + \sum_{i=1}^m\frac{w_i}{\sqrt{2\pi}} + \sum_{i=1}^m P(|Y_i| > w_i).$$

Proof. The first inequality is a consequence of Michel and Pfanzagl (1971) and the second one follows from Lemma 3.1 of Liang and Fan (2009). ■

Lemma 7.8 (Yang and Li (2006)) *Let p and q be positive integers. Set $\eta_r = \sum_{j=(r-1)(p+q)+1}^{(r-1)(p+q)+p} Z_j$ for $1 \leq r \leq w$. If $s > 0$, $r > 0$ with $1/s + 1/r = 1$, then there exists constant $C > 0$ such that $|E\exp(it\sum_{r=1}^{w}\eta_r) - \prod_{r=1}^{w} E\exp(it\eta_r)| \leq C|t|\alpha^{1/s}(q)\sum_{r=1}^{w}\|\eta_r\|_r$.*

Acknowledgments. The authors are grateful to the Editor, an associate Editor, and two anonymous referees for providing a detailed list of comments which greatly improved the presentation of the paper. The first author was supported by the National Natural Science Foundation of China (11671299). The second author was supported by the Grant MTM2014-55966-P of the Spanish Ministry of Economy and Competitiveness.

References

[1] Bercu, B., Nguyen, T. M. N., Saracco, J. (2015). On the asymptotic behaviour of the recursive Nadaraya-Watson estimator associated with the recursive sliced inverse regression method. *Statistics* **49(3)**, 660-679.

[2] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes.* Springer, New York.

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

[4] de Uña-Álvarez, J., Iglesias-Pérez, M. C. (2010). Nonparametric estimation of a conditional distribution from length-biased data. *Ann. Inst. Statist. Math.* **62**, 323-341.

[5] de Uña-Álvarez, J., Liang, H. Y., Rodríguez-Casal, A. (2010). Nonlinear wavelet estimator of the regression function under left truncated dependent data. *J. Nonparametric Statist.* **22(3)**, 319-344.

[6] Doukhan. P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics, vol. 85. Springer, Berlin.

[7] El Ghouch, A., Van Keilegom, I. (2008). Non-parametric regression with dependent censored data. *Scand. J. Statist.* **35(2)**, 228-247.

[8] Fan, J., Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman & Hall, London.

[9] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series.* Springer, New York.

[10] Goegebeur, Y., Guillou, A., Schorgen, A. (2014). Nonparametric regression estimation of conditional tails: the random covariate case. *Statistics* **48(4)**, 732-755.

[11] Györfi, L., Härdle, W., Sarda, P. and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series.* Lecture Notes in Statistics, Vol. 60. Springer, New York.

[12] Hall, P., Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* Academic Press, New York.

[13] Hall, P., Presnell, B. (1999). Intentionally biased bootstrap methods. *J. Roy. Statist. Soc. Ser. B* **61**, 143-158.

[14] Huang, W. J., Su, N. C. (2015). Characterizations based on regression assumptions of order statistics. *Statistics* **49(3)**, 578-587.

[15] Iglesias-Pérez, M. C., González-Manteiga, W. (1999). Strong representation of a generalized product-limit estimator for truncated and censored data with some applications. *J. Nonparametric Statist.* **10**, 213-244.

[16] Liang, H. Y., Fan, G. L. (2009). Berry-Esseen type bounds of estimators in a semiparametric model with linear process errors. *J. Multivar. Analysis* **100**, 1-15.

[17] Liang, H. Y. (2012). Weighted nonparametric regression estimation with truncated and dependent data. *J. Nonparametric Statist.* **24(4)**, 1051-1073.

[18] Liang, H. Y., de Uña-Álvarez, J., Iglesias-Pérez, M. C. (2011). Local polynomial estimation of a conditional mean function with dependent truncated data. *Test* **20(3)**, 653-677.

[19] Liang, H. Y., de Uña-Álvarez, J., Iglesias-Pérez, M. C. (2012). Asymptotic properties of conditional distribution estimator with truncated, censored and dependent data. *Test* **21(4)**, 790-810.

[20] Liang, H. Y., de Uña-Álvarez, J., Iglesias-Pérez, M. C. (2015). A CLT in nonparametric regression with truncated, censored and dependent data. *Scandinavian Journal of Statistics* **42**, 256-269.

[21] Liebscher, E. (1996). Strong convergence of sums of α-mixing random variables with applications to density estimation. *Stochastic Processes Appl.* **65(1)**, 69-80.

[22] Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions. *Statist. Decisions* **19(1)**, 9-26.

[23] Ling, N., Wu, Y. (2012). Consistency of modified kernel regression estimation for functional data. *Statistics* **46(2)**, 149-158.

[24] Masry, E., Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scand. J. Statist.* **24**, 165-179.

[25] Michel, R., Pfanzagl, J. (1971). The accuracy of the normal approximation for minimum constrast estimates. *Z. Wahrsch. Verw. Gebiete.* **18**, 73-84.

[26] Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18**, 90-120.

[27] Petrov, V. V. (1995). *Limit Theorems of Probability Theory.* Oxford Univ. Press Inc., New York.

[28] Shao, Q., Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Ann. Probab.* **24**, 2098-2127.

[29] Volkonskii, V. A. and Rozanov, Y. A. (1959). Some limit theorems for random functions. *Theory Probab. Appl.* **4**, 178-197.

[30] Wang, J. F., Liang, H. Y., Fan, G. L. (2013). Local polynomial quasi-likelihood regression with truncated and dependent data. *Statistics* **47(4)**, 744-761.

[31] Withers, C. S. (1981). Conditions for linear processes to be strong mixing. *Z. Wahrsch. verw. Gebiete* **57**, 477-480.

[32] Yanev, G. P., Ahsanullah, M. (2012). Characterizations of Student's t-distribution via regressions of order statistics. *Statistics* **46(4)**, 429-435.

[33] Yang, S. C., Li, Y. M. (2006). Uniformly asymptotic normality of the regression weighted estimator for strong mixing samples. *Acta. Math. Sinica* **49(5)**, 1163-1170.

[34] Zou, Y. Y., Liang, H. Y. (2017) Wavelet estimation of density for censored data with censoring indicator missing at random. *Statistics* **51(6)**, 1214-1237.