Ann Inst Stat Math (2018) 70:155–189
https://doi.org/10.1007/s10463-016-0587-4

A weighted estimator of conditional hazard rate with left-truncated and dependent data

Han-Ying Liang[1] · **Elias Ould Saïd**[2]

Received: 26 January 2015 / Revised: 17 September 2016 / Published online: 31 October 2016

Abstract Based on empirical likelihood method, we construct new weighted estimators of conditional density and conditional survival functions when the interest random variable is subject to random left-truncation; further, we define a plug-in weighted estimator of the conditional hazard rate. Under strong mixing assumptions, we derive asymptotic normality of the proposed estimators which permit to built a confidence interval for the conditional hazard rate. The finite sample behavior of the estimators is investigated via simulations too.

Keywords Asymptotic normality · Conditional hazard rate · Strong mixing · Truncated data · Weighted estimator

1 Introduction

In medical follow-up or in engineering life-test study, the lifetime variables may not be completely observable, right-censored or left-truncated data are often encountered. In this paper we consider the case where the response variable is left-truncated, construct and study a weighted estimation of the conditional hazard function.

H.-Y. Liang was supported by the National Natural Science Foundation of China (11271286, 11671299).

✉ Han-Ying Liang
hyliang@tongji.edu.cn

Elias Ould Saïd
ouldsaid@lmpa.univ-littoral.fr; Elias.Ould-Said@univ-littoral.fr

1 School of Mathematical Science, Tongji University, Shanghai 200092, People's Republic of China

2 Univ. Littoral Côte d'Opale, LMPA, CS: 80699, Calais, France

Let $(\mathscr{Y}_i)_{i=1,\ldots,N}$ be a sample of real random variables (rv) with common unknown distribution function (df) F. Here the integer N is unknown but deterministic. Let $(\mathscr{X}_i)_{i=1,\ldots,N}$ a corresponding sample of real random covariate vectors with density function $\ell(\cdot)$. The $\mathscr{Y}_i$s are regarded as the lifetimes of the items under study and are supposed to be subject to left-truncation which may occur if the time origin of the lifetime precedes the time origin of the study. Only subjects that fail after the start of the study are being followed, the others being truncated. We denote by $(\mathscr{T}_i)_{i=1,\ldots,N}$ the sample of truncation rv with d.f. G. The $\mathscr{T}_i$s are assumed to be independent of the $\mathscr{Y}_i$s. Then $(\mathscr{Y}_i, \mathscr{T}_i)$ are observed only when $(\mathscr{Y}_i \geq \mathscr{T}_i)$. Clearly the initial sample is not completely observed and only n observations (among N) are obtained. We point out that n is random but known. Such data, arise in many fields such as astronomy, economics and medical studies (see e.g., Woodroofe 1985). In what follows, we denote again (if there is no-confusion for the index) $\{(Y_i, T_i, X_i), 1 \leq i \leq n\}, (n \in \mathbb{N})$, the observed sample from the original N-sample. As a consequence of truncation, the size of the actually observed sample, n is a $Bin(N, \mu)$ random variable, with $\theta := \mathbb{P}(\mathscr{Y} \geq \mathscr{T})$. It is clear that if $\theta = 0$, no data can be observed and therefore, we suppose throughout this paper that $\theta > 0$. By the strong law of large numbers (SLLN) we have, as $N \to \infty$

$$\widehat{\theta}_n := \frac{n}{N} \longrightarrow \theta, \quad \mathbb{P} - a.s. \tag{1}$$

Since the N is unknown and the n is known (although random), our results would not be stated with respect to the probability measure $\mathbb{P}$ (related to the N-sample) but will involve the conditional probability $\mathbf{P}(\cdot) = \mathbb{P}(\cdot|\mathscr{Y} \geq \mathscr{T})$ with respect to the actually observed n-sample. Also $\mathbf{E}$ and $\mathbb{E}$ will denote the expectation operators under $\mathbf{P}$ and $\mathbb{P}$, respectively.

It is well known that a useful tool in survival analysis (complete or incomplete data) is the hazard function, which reflects the instantaneous probability that a duration will end within the next time instant. An increasing hazard rate indicates positive duration dependence; that is the probability that a spell is completed increases with the duration between the events. Similarly, a decreasing hazard rate reflects negative duration dependence. However, in practice the hazard function depends on covariates, such age, rate of cholesterol.

In the complete data case, unconditional or conditional hazard rates have been widely studied by many authors. To quote only a few, we cite Collomb et al. (1985), Sarda and Vieu (1991) and Quintela-del-Río (2008). When the covariates take their values in infinite dimensional spaces, Berlinet et al. (2011) studied a non-linear regression model with functional data as inputs and scalar response, they get a pointwise estimate of the regression function that maps a Hilbert space onto the real line by a local linear method and derive its asymptotic mean square error.

For censored data and unconditional case, Lecoutre and Ould Saïd (1995) studied the strong consistency of a kernel estimate when the data exhibit a strong mixing condition. Bagkavos (2011a, b) defined a new kernel based local linear estimator of the hazard rate, he studied its finite sample and asymptotic properties as well as proved its asymptotic normality. For the conditional case, Lecoutre and Ould Saïd (1992) established a pointwise and uniform almost complete convergence of a kernel estimators of density and hazard functions under strong mixing condition. Spierdijk (2008) defined a new estimator of the hazard rate by using a ratio of local linear estimators for the conditional density and survivor function. The resulting hazard rate estimator has been shown to be pointwise consistent and asymptotically normally distributed under appropriate conditions. Van Keilegom and Veraverbeke (2001) studied hazard rate estimation using a nonparametric regression. Recently, Kim et al. (2010) proposed a new local linear estimator of the hazard rate which is motivated by the ideas of Fan et al. (1996) and Kim et al. (2005). The asymptotic distribution of the estimator is derived, and some numerical results have been also given.

For the left-truncated data and parameter case, Zhou (2011) developed a weighted quantile regression approach. The method leads to a simple algorithm that has been conveniently implemented with R software. Furthermore, he showed that the proposed estimator is strongly consistent and asymptotically normal under appropriate conditions and independent and identically distributed (i.i.d.) case. In the nonparametric estimation setting and as far as we know, the previous paper is of Ould Saïd and Lemdani (2006), where they built a new Nadaraya–Watson (NW) estimator of $m(x) = E(Y|X = x)$ and studied its asymptotic properties in the i.i.d. case. Further Liang et al. (2009) extended the results of Ould Saïd and Lemdani (2006) to the dependent data. For local linear estimators of the regression function and conditional density function, Liang et al. (2011) and Liang and Baek (2016), respectively, studied the asymptotic properties of the proposed estimator when the data satisfy the strong mixing conditions. We point out, that in this case, even we study the density and survival distribution, it is not possible to deduce the same results for conditional hazard, which is not the case for Nadaraya–Watson and weighted estimation approaches. Using a weighted estimation, Liang (2012) established the asymptotic normality and weak consistency of the estimator of the regression function. Furthermore, he made a comparison study between Nadaraya–Watson, local linear and weighted estimators by the finite sample performance of the proposed estimators by studying their mean square errors.

As pointed by Stute (1993), the purely truncated data is completely different to censored data and then we cannot deduce the results for truncated data from those obtained in the censored case. As far as we know, estimation of the conditional hazard function for truncated data based on weighted method has not been studied yet in the literature. This is the goal of this paper.

The rest of the paper is organized as follows: in Sect. 2 we recall the truncation framework; the different notations and defining the estimator of the conditional hazard function are presented in Sect. 3. The assumptions and main results are detailed in Sect. 4. A simulation study is presented in Sect. 5. Sect. 6 is devoted to the proofs of the main results.

2 Background for truncation models

In this section, we give the main definitions and results related to the truncation model. Throughout the paper, the star notation (*) relates to any characteristic function of the actually observed data. Recall that under random left-truncation, the conditional joint distribution of an observed (Y, T) (see e.g., Stute 1993), is given by

$$V^*(y,t) = \mathbf{P}(Y \le y, T \le t) = \mathbb{P}(Y \le y, T \le t | Y \ge T)$$
$$= \theta^{-1} \int_{-\infty}^{y} G(t \wedge u) dF(u),$$

where $t \wedge u := \min(t, u)$. Then the marginal laws are defined by

$$F^*(y) = \theta^{-1} \int_{-\infty}^{y} G(u) dF(u) \text{ and } G^*(t) = \theta^{-1} \int_{-\infty}^{\infty} G(t \wedge u) dF(u),$$

which are estimated by $F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{Y_i \le y\}}$ and $G_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{T_i \le t\}}$, respectively, where $\mathbf{1}_A$ denotes the indicator function of the set A.

For any d.f. W, define $a_W = \inf\{z : W(z) > 0\}$ and $b_W = \sup\{z : W(z) < 1\}$, as the endpoints of the W support. As pointed out by Woodroofe (1985), the d.f.s F and G can be completely estimated only under the conditions

$$a_G \le a_F, \quad b_G \le b_F \text{ and } \int_{a_F}^{\infty} \frac{dF}{G} < \infty.$$

Now, let $R(\cdot)$ be a function defined by $R(y) = G^*(y) - F^*(y) = \theta^{-1} G(y)[1 - F(y)]$ with empirical estimator

$$R_n(y) = G_n^*(y) - F_n^*(y^-) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{T_i \le y \le Y_i\}} \text{ for } y \ge a_F.$$

Then, the well-known nonparametric estimators of F and G, originally proposed by Lynden-Bell (1971), are given by

$$F_n(y) = 1 - \prod_{i:Y_i \le y} \left[\frac{nR_n(Y_i) - 1}{nR_n(Y_i)}\right] \text{ and } G_n(t) = \prod_{i:T_i > t} \left[\frac{nR_n(T_i) - 1}{nR_n(T_i)}\right],$$

respectively, assuming no ties among the Ys. Note that Stute and Wang (2008) showed how to break ties without destroying the product limit structure. Therefore, throughout we shall assume without loss of generality that there are no ties among the Ys.

Another important quantity is the unknown probability θ, which estimated by (1) but cannot be calculated since N is unknown. Note that

$$\theta(t) := \frac{G(t)(1 - F(t))}{R(t)} =: \theta$$

for all $a_F < t < b_F$. This observation led He and Yang (1998) to propose

$$\theta_n(t) = \frac{G_n(t)(1 - F_n(t^-))}{R_n(t)} =: \theta_n$$

as an estimator of θ. Moreover, they showed (see their Corollary 2.5) its **P**-a.s. consistency in i.i.d. case, that is $\theta_n \to \theta$, **P**-a.s. as $n \to \infty$, while under an α-mixing hypothesis, Ould Saïd and Tatachak (2009) stated that (see their Lemma 5.2),

$$\theta_n - \theta = O_{\mathbf{P}}\left(\sqrt{\frac{\ln\ln(n)}{n}}\right).$$

This result was established by Liang et al. (2009) in almost surely case.

In order to discuss the estimation of the conditional hazard function for truncated data based on the weighted method in more general situation, the observed sample $\{(Y_i, T_i, X_i), 1 \le i \le n\}$, in the sequel, is assumed to be a stationary α-mixing sequence. We recall definition of the α-mixing sequence.

Definition 1 Let $\{Z_i,\ i \ge 1\}$ denotes a sequence of r.v.s. Given an integer $n \ge 1$, set

$$\alpha(n) = \sup\left\{|\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)| : A \in \mathscr{F}_1^k(Z),\ B \in \mathscr{F}_{k+n}^{\infty}(Z),\ k \in \mathbb{N}\right\},$$

where $\mathscr{F}_i^k(Z)$ denotes the σ-field of events generated by $\left\{Z_j,\ i \le j \le k\right\}$. The sequence is said to be α-mixing (or strongly mixing) if the mixing coefficient $\alpha(n) \to 0$ as $n \to \infty$.

The α-mixing condition has many practical applications (see e.g., Bradley (2007) and Dedecker et al. (2007) for more details). The α-mixing has been used in applications with clustered survival data, see, for instance, Cai and Kim (2003).

3 Estimators

Denote by $S(\cdot \mid x)$ and $f(\cdot \mid x)$ the conditional survival function and conditional density function of Y given $X = x$, respectively. It is well known that the conditional hazard function is defined by the ratio between $f(y|x)$ and $S(y|x)$ that is

$$\lambda(y|x) := \frac{f(y|x)}{S(y|x)}$$

such that $S(y|x) > 0$. Let $H(\cdot)$ be a probability density function (called kernel) and $0 < h_{n,H} =: h_H \to 0$ as $n \to +\infty$. In what follows, we denote by $H_{h_H}(\cdot) := \dfrac{H(\cdot/h_H)}{h_H}$.

Let $f(\cdot,\cdot)$ stand for joint density function of (X, Y), then from formula (12) in Ould Saïd and Lemdani (2006) we have

$$f(x, y) = [\theta^{-1}G(y)]^{-1} f^*(x, y) \quad \text{for} \quad y > a_G. \tag{2}$$

Hence, under assumptions (A1) and (A2) in Sect. 4 below, it follows that

$$\begin{aligned}
&\theta \mathbf{E}\Big\{K_{h_K}(X-x)G^{-1}(Y)[H_{h_H}(Y-y)-f(y|x)]\Big\}\\
&=\frac{\theta}{h_K}\int_{\mathbb{R}}\int_{\mathbb{R}}K\Big(\frac{s-x}{h_K}\Big)G^{-1}(t)\Big[h_H^{-1}H\Big(\frac{t-y}{h_H}\Big)-f(y|x)\Big]f^*(s,t)dsdt\\
&=\frac{1}{h_Kh_H}\int_{\mathbb{R}}\int_{\mathbb{R}}K\Big(\frac{s-x}{h_K}\Big)H\Big(\frac{t-y}{h_H}\Big)f(s,t)dsdt-\frac{f(y|x)}{h_K}\int_{\mathbb{R}}K\Big(\frac{s-x}{h_K}\Big)\ell(s)ds\\
&=\int_{\mathbb{R}}K(s)\ell(x+h_Ks)\Big(\int_{\mathbb{R}}H(t)[f(y+h_Ht|x+h_Ks)-f(y|x)]dt\Big)ds\to 0,
\end{aligned} \tag{3}$$

where $K(\cdot)$ is a kernel function and $h_K := h_{K,n}$ is a sequence of bandwidth which goes to zero as n tends to infinity. Define $\mathbb{H}(y)=\int_{-\infty}^{y}H(u)du$, analogously as before, one has

$$\theta \mathbf{E}\Big\{K_{h_K}(X-x)G^{-1}(Y)\Big[\mathbb{H}\Big(\frac{Y-y}{h_H}\Big)-S(y|x)\Big]\Big\}\to 0. \tag{4}$$

Put $f^{(i,j)}(y|x)=\partial^{i+j}f(y|x)/\partial x^i\partial y^j$. If $f(y|x)$ is assumed to have $(p+1)$th continuous derivative respect to x, then, in a small neighborhood of x, it can be approximated by a polynomial function as

$$f(y|z)\approx f(y|x)+\cdots+f^{(p,0)}(y|x)(z-x)^p/p!\equiv\beta_0+\cdots+\beta_p(z-x)^p.$$

Based on the idea of the local polynomial smoother, from (3) the estimator of $(f(y|x),\ldots,f^{(p,0)}(y|x)/p!)^\tau$, is defined as $(\widehat{\beta_0},\ldots,\widehat{\beta_p})^\tau$, which minimizes

$$\sum_{i=1}^{n}\Big(H_{h_H}(Y_i-y)-\sum_{j=0}^{p}\beta_j(X_i-x)^j\Big)^2K_{h_K}(X_i-x)G_n^{-1}(Y_i). \tag{5}$$

When $p=0$ in (5), the estimator of $f(y|x)$ is the well-known NW type estimator, defined by

$$\widehat{f}_{\mathrm{NW}}(y|x)=\sum_{i=1}^{n}w_i^{\mathrm{NW}}(x)H_{h_H}(Y_i-y)\text{ with }w_i^{\mathrm{NW}}(x)=\frac{G_n^{-1}(Y_i)K_{h_K}(X_i-x)}{\sum_{j=1}^{n}G_n^{-1}(Y_j)K_{h_K}(X_j-x)}.$$

When $p=1$ in (5), the local linear (LL) estimator of $f(y|x)$ is

$$\widehat{f}_{LL}(y|x)=\sum_{i=1}^{n}w_i^{\mathrm{LL}}(x)H_{h_H}(Y_i-y),$$

here $w_i^{\mathrm{LL}}(x)=\frac{K_{h_K}(X_i-x)G_n^{-1}(Y_i)\{s_{n,2}-(X_i-x)s_{n,1}\}}{s_{n,0}s_{n,2}-s_{n,1}^2}$, $s_{n,j}=\sum_{i=1}^{n}(X_i-x)^jK_{h_K}(X_i-x)G_n^{-1}(Y_i)$.

Remark 1 In view of $f(y|x) = f(x, y)/\ell(x)$, Ould Saïd and Tatachak (2007) defined for the first time the plug-in estimator $\widehat{f}_{NW}(y|x)$ by using the estimators of $f(x, y)$ and $\ell(x)$; based on the idea of the local polynomial smoother, Liang and Baek (2016) constructed the estimators $\widehat{f}_{\mathrm{NW}}(y|x)$ and $\widehat{f}_{\mathrm{LL}}(y|x)$ of $f(y|x)$.

Similarly, from (4) the NW type and LL estimators of $S(y|x)$ are defined, respectively, by

$$\widehat{S}_{\mathrm{NW}}(y|x) = \sum_{i=1}^{n} w_i^{\mathrm{NW}}(x)\mathbb{H}\left(\frac{Y_i - y}{h_H}\right) \quad \text{and} \quad \widehat{S}_{\mathrm{LL}}(y|x) = \sum_{i=1}^{n} w_i^{\mathrm{LL}}(x)\mathbb{H}\left(\frac{Y_i - y}{h_H}\right).$$

From least squares theory, it is easy to see that the LL weights $w_i^{\mathrm{LL}}(x)$ satisfy:

$$\sum_{i=1}^{n} w_i^{\mathrm{LL}}(x) = 1 \quad \text{and} \quad \sum_{i=1}^{n}(X_i - x)w_i^{\mathrm{LL}}(x) = 0. \tag{6}$$

Motivated by the information (6), we use the empirical likelihood method to define the new weighted NW type estimator of $f(y|x)$ for the left-truncation model as follows. This approach was proposed first by Hall and Presnell (1999) for estimating regression function under the independent samples and it was used by Hall et al. (1999) for estimating conditional distribution. We define the empirical likelihood function $H = \prod_{i=1}^{n} p_i(x)$, where $p_1(x), \ldots, p_n(x)$ are subject to the restrictions:

$$p_i(x) \geq 0, \ \sum_{i=1}^{n} p_i(x) = 1 \quad \text{and} \quad \sum_{i=1}^{n}(X_i - x)p_i(x)K_{h_K}(X_i - x) = 0. \tag{7}$$

Using Lagrange multipliers, we get $H_{\max} = \prod_{i=1}^{n} \widehat{p}_i(x)$, where

$$\widehat{p}_i(x) = \frac{1}{n} \cdot \frac{1}{1 + \eta(X_i - x)K_{h_K}(X_i - x)}, \quad i = 1, \ldots, n,$$

and η is the solution of the following equation:

$$\sum_{i=1}^{n} \frac{(X_i - x)K_{h_K}(X_i - x)}{1 + \eta(X_i - x)K_{h_K}(X_i - x)} = 0. \tag{8}$$

The proposed weighted NW type estimators of $f(y|x)$ and $S(y|x)$, respectively, are

$$\widehat{f}_n(y|x) = \sum_{i=1}^{n} \widehat{w}_i(x)H_{h_H}(Y_i - y) \quad \text{and} \quad \widehat{S}_n(y|x) = \sum_{i=1}^{n} \widehat{w}_i(x)\mathbb{H}\left(\frac{Y_i - y}{h_H}\right)$$

with $\widehat{w}_i(x) = \frac{\widehat{p}_i(x)K_{h_K}(X_i - x)G_n^{-1}(Y_i)}{\sum_{j=1}^{n} \widehat{p}_j(x)K_{h_K}(X_j - x)G_n^{-1}(Y_j)}$.

Therefore, we define the following plug-in NW, LL and weighted estimators of $f(y|x)$ and $S(y|x)$, respectively

$$\widehat{\lambda}_{\mathrm{NW}}(y|x)=\frac{\widehat{f}_{\mathrm{NW}}(y|x)}{\widehat{S}_{\mathrm{NW}}(y|x)},\quad \widehat{\lambda}_{\mathrm{LL}}(y|x)=\frac{\widehat{f}_{\mathrm{LL}}(y|x)}{\widehat{S}_{\mathrm{LL}}(y|x)},\quad \widehat{\lambda}_{n}(y|x)=\frac{\widehat{f}_{n}(y|x)}{\widehat{S}_{n}(y|x)}.$$

Remark 2 Although the estimators $\widehat{\lambda}_{\mathrm{NW}}(y|x)$ and $\widehat{\lambda}_n(y|x)$ are considered only for covariate X in univariate case, it is worthy of pointing out that the basic ideas of our methodology hold for multivariate situations.

4 Assumptions and main results

In the sequel, let C, c_0 and c denote generic finite positive constants, whose values are unimportant and may change from line to line, and let $U(x)$ represent a neighborhood of x. $A_n = O(B_n)$ means $|A_n| \le C|B_n|$. In order to formulate the main results, we need the following assumptions.

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

(A1) Both $K(\cdot)$ and $H(\cdot)$ are symmetric and bounded density functions with compact support on $\mathbb{R}$, respectively.

(A2) (i) The second derivative of $\ell(\cdot)$ is continuous in $U(x)$ and $\ell(x) > 0$;
(ii) The second partial derivatives of $f(\cdot|\cdot)$ are continuous in $U(x) \times U(y)$ and $f(y|x) > 0$.

(A3) (i) For all integers $j \ge 1$, the joint conditional density $l_j^*(\cdot,\cdot)$ of (X_1, X_{j+1}) exists on $\mathbb{R}\times\mathbb{R}$ and satisfies $l_j^*(s_1, s_2) \le C$ for $(s_1, s_2) \in U(x) \times U(x)$;
(ii) For all integers $j \ge 1$, the joint conditional density $l_j^*(\cdot,\cdot,\cdot)$ of (X_1, X_{j+1}, Y_1) exists on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}$ and satisfies $l_j^*(s_1, s_2, t_1) \le C$ for $(s_1, s_2, t_1) \in U(x) \times U(x) \times U(y)$;
(iii) For all integers $j \ge 1$, the joint conditional density $l_j^*(\cdot,\cdot,\cdot)$ of (X_1, X_{j+1}, Y_{j+1}) exists on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}$ and satisfies $l_j^*(s_1, s_2, t_2) \le C$ for $(s_1, s_2, t_2) \in U(x) \times U(x) \times U(y)$;
(iv) For all integers $j \ge 1$, the joint conditional density $l_j^*(\cdot,\cdot,\cdot,\cdot)$ of $(X_1, X_{j+1}, Y_1, Y_{j+1})$ exists on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$ and satisfies $l_j^*(s_1, s_2, t_1, t_2) \le C$ for $(s_1, s_2, t_1, t_2) \in U(x) \times U(x) \times U(y) \times U(y)$.

(A4) Assume that $nh_K h_H \to \infty$, and that the sequence $\alpha(n)$ satisfies that for positive integers q_n such that $q_n = o((nh_K h_H)^{1/2})$ and $\lim_{n\to\infty}(n(h_K h_H)^{-1})^{1/2}\alpha(q_n) = 0$.

(A5) Assume that $nh_K \to \infty$, and that the sequence $\alpha(n)$ satisfies that for positive integers u_n such that $u_n = o((nh_K)^{1/2})$ and $\lim_{n\to\infty}(n(h_K)^{-1})^{1/2}\alpha(u_n) = 0$.

(A6) The second partial derivatives of $S(\cdot|\cdot)$ are continuous in $U(x) \times U(y)$ and $S(y|x) > 0$.

Remark 3 The condition $a_G < a_F$ will be needed if we state uniform result which imply a sufficient rate of convergence of G_n (see Lemma 6.3 in Woodroofe (1985)). Then we have to consider a set of values of Y_i which do not include a_G (a uniform rate for G_n is given in Woodroofe (1985) on $[a, b_G]$) with $a > a_G$) that is $a_F > a_G$, thus

$G(Y) \geq G(a_F) > 0$, which ensures $G_n(Y_i) \neq 0$ eventually, so the given estimators are well defined for large n. Assumptions (A1) and (A2) are used commonly in the literature. Assumption (A3) is mainly technical, which is employed to simplify the calculations of covariances in the proof, the assumption is redundant for the independent setting. Assumption (A4) implies restrictions when choosing the bandwidth, the conditions in (A4) can be satisfied easily, for example, choose $h_K = h_H = cn^{-\eta}$ for some $0 < \eta < 1/2$, $q_n = (nh_K^2/\log n)^{1/2}$ and assume $\alpha(k) = O(k^{-\gamma})$ for some $\gamma > 0$, then (A4) automatically holds if γ is large enough, specifically $\gamma > 2(1+2\eta)/(1-2\eta)$ (note that γ can be arbitrarily large if $\alpha(k) = O(\rho^k)$ for some $0 < \rho < 1$).

Put $\Delta_{ij} = \int_{\mathbb{R}} s^i K^j(s)ds$, $\Lambda_{ij} = \int_{\mathbb{R}} v^i H^j(v)dv$, $\sigma^2(y|x) = \frac{f(y|x)}{\ell(x)G(y)} \int_{\mathbb{R}^2} K^2(u)H^2(v)dudv$.

Proposition 1 *Let $\alpha(n) = O(n^{-\gamma})$ for some $\gamma > 3$, and let $x \in \mathbb{R}$ and $y > a_F$. Suppose that (A0)–(A4) are satisfied. If $nh_K^{1+r} = O(1)$ for some $r \geq 2$, then*

$$\sqrt{nh_K h_H}\left\{\widehat{f}_n(y|x) - f(y|x) - \frac{h_K^2}{2}\Delta_{21} f^{(2,0)}(y|x) - \frac{h_H^2}{2}\Lambda_{21} f^{(0,2)}(y|x) + o_{\mathbf{P}}(h_K^2 + h_H^2) + O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}} + \frac{1}{nh_K\sqrt{h_H}} + \sqrt{\frac{h_K}{nh_H}}\right)\right\} \xrightarrow{\mathscr{D}} N\big(0, \theta\sigma^2(y|x)\big).$$

Proposition 2 *Let $\alpha(n) = O(n^{-\gamma})$ for some $\gamma > 3$, and let $x \in \mathbb{R}$ and $y > a_F$. Suppose that (A0)–(A3) and (A5)–(A6) hold. Set*

$$\Delta_n(y|x) = \theta\ell(x)\mathbb{E}\left(\left[\mathbb{H}\left(\frac{Y-y}{h_H}\right) - S(y|x)\right]^2 G^{-1}(Y)\Big|X = x\right)\int_{\mathbb{R}} K^2(u)du.$$

Suppose that there exists $\varepsilon(y|x) > 0$ such that $\Delta_n(y|x) \geq \varepsilon(y|x)$. If $nh_K \to \infty$ and $nh_K^{1+r} = O(1)$ for some $r \geq 2$, then

$$\sqrt{\frac{nh_K\ell^2(x)}{\Delta_n(y|x)}}\left\{\widehat{S}_n(y|x) - S(y|x) - \frac{h_K^2}{2}\Delta_{21} S^{(2,0)}(y|x) - \frac{h_H^2}{2}\Lambda_{21} S^{(0,2)}(y|x) + o_{\mathbf{P}}(h_K^2 + h_H^2) + O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}} + \frac{1}{nh_K} + \sqrt{\frac{h_K}{n}}\right)\right\} \xrightarrow{\mathscr{D}} N\big(0, 1\big).$$

Remark 4 (a) In Proposition 2, though $\Delta_n(y|x)$ depends on n, it is bounded. (b) Under complete data, $\theta = 1$ and $G(\cdot) = 1$, in this case, one can verify that $\Delta_n(y|x) = \ell(x)S(y|x)[1 - S(y|x)]\int_{\mathbb{R}} K^2(u)du$, which is a constant.

Proposition 3 *Under the assumptions of Proposition 1, if (A6) holds, then $\widehat{S}_n(y|x) \to S(y|x)$ in Probability.*

Remark 5 In Proposition 3, the conditions in (A2) and (A6) can be weakened, i.e., the functions $\ell(\cdot)$, $f^{(2,0)}(\cdot|\cdot)$ and $S^{(2,0)}(\cdot|\cdot)$ are continuous in $U(x)$ and $U(x) \times U(y)$, respectively, may replace the conditions (A2) and (A6) in Proposition 3.

Theorem 1 *Under the assumption of Proposition 1, if (A6) holds, then*

$$\sqrt{nh_K h_H}\Big\{\widehat{\lambda}_n(y|x) - \lambda(y|x) - Asymp.bias + o_{\mathbf{P}}(h_K^2 + h_H^2) + O_{\mathbf{P}}\big((nh_K)^{-1/2} + (h_K/(nh_H))^{1/2}\big)\Big\} \xrightarrow{\mathscr{D}} N\left(0, \frac{\theta\sigma^2(y|x)}{S^2(y|x)}\right),$$

where

$$Asymp.bias = \frac{1}{S(y|x)}\Big\{\frac{h_K^2}{2}\Lambda_{21} f^{(2,0)}(y|x) + \frac{h_H^2}{2}\Lambda_{21} f^{(0,2)}(y|x) - \lambda(y|x)\Big[\frac{h_K^2}{2}\Lambda_{21} S^{(2,0)}(y|x) + \frac{h_H^2}{2}\Lambda_{21} S^{(0,2)}(y|x)\Big]\Big\}.$$

Remark 6 (a) Under suitable conditions, Liang and Baek (2016) established the following results:

$$\sqrt{nh_K h_H}\Big\{\widehat{f}_{NW}(y|x) - f(y|x) - \frac{h_K^2}{2}\Lambda_{21}\Big[f^{(2,0)}(y|x) + \frac{2\ell'(x)}{\ell(x)} f^{(1,0)}(y|x)\Big] - \frac{h_H^2}{2}\Lambda_{21} f^{(0,2)}(y|x) \quad + o_{\mathbf{P}}(h_K^2 + h_H^2)\Big\} \xrightarrow{\mathscr{D}} N\big(0, \theta\sigma^2(y|x)\big);$$

$$\sqrt{nh_K h_H}\Big\{\widehat{f}_{LL}(y|x) - f(y|x) - \frac{h_K^2}{2}\Lambda_{21} f^{(2,0)}(y|x) - \frac{h_H^2}{2}\Lambda_{21} f^{(0,2)}(y|x) + o_{\mathbf{P}}(h_K^2 + h_H^2)\Big\} \xrightarrow{\mathscr{D}} N\big(0, \theta\sigma^2(y|x)\big).$$

Based on these results, once the asymptotic bias of $\widehat{S}_{\rm NW}(y|x)$ and $\widehat{S}_{\rm LL}(y|x)$ are deduced, respectively, as in Proposition 2, one can give asymptotic normality of $\widehat{\lambda}_{\rm NW}(y|x)$ and $\widehat{\lambda}_{\rm LL}(y|x)$. Though the asymptotic normality of $\widehat{\lambda}_{\rm NW}(y|x)$ and $\widehat{\lambda}_{\rm LL}(y|x)$ has not been formulated in this paper, from the results above, we can guess that the estimators $\widehat{\lambda}_n(y|x)$ and $\widehat{\lambda}_{\rm LL}(y|x)$ have same asymptotic normality, and both of them are superior to that of the estimator of $\widehat{S}_{\rm NW}(y|x)$ if comparing their asymptotic mean squared errors.

(b) Put $\Delta^2(x, y) = \frac{\theta\ell(x)f(y|x)}{G(y)}\int_{\mathbb{R}} K^2(u)du \int_{\mathbb{R}} H^2(v)dv$. Define

$$\hat{\ell}_n(x) = \frac{\theta}{nh_K}\sum_{i=1}^n \frac{1}{G_n(Y_i)} K\left(\frac{X_i - x}{h_K}\right),$$

$$\hat{\Delta}_n^2(x, y) = \frac{\theta^2}{nh_K h_H}\sum_{i=1}^n \frac{1}{G_n^2(Y_i)} K^2\Big(\frac{X_i - x}{h_K}\Big) H^2\Big(\frac{Y_i - y}{h_H}\Big).$$

Hence, we obtain a plug-in estimator $\hat{\sigma}_n^2(y|x) = \hat{\Delta}_n^2(x, y)\hat{\ell}_n^{-2}(x)$ of $\theta\sigma^2(y|x)$.

It is well known that the main feature of the asymptotic normality is to get a confidence intervals, which are given below.

Corollary 1 *Under the assumptions of Theorem 1, based on the estimators* $\hat{\sigma}_n^2(y|x)$ *and* $\widehat{S}_n(y|x)$ *of* $\theta\sigma^2(y|x)$ *and* $S(y|x)$*, respectively, we get a confidence interval of asymptotic level* $1-\alpha$ *for* $\lambda(y|x)$

$$\left[\widehat{\lambda}_n(y|x) - \frac{u_{1-\alpha/2}\widehat{\sigma}_n(y|x)}{\sqrt{nh_K h_H}\widehat{S}_n(y|x)}, \widehat{\lambda}_n(y|x) + \frac{u_{1-\alpha/2}\widehat{\sigma}_n(y|x)}{\sqrt{nh_K h_H}\widehat{S}_n(y|x)}\right],$$

where $u_{1-\alpha/2}$ *denotes the* $(1-\alpha/2)$*-quantile of the standard normal distribution.*

5 Simulation study

In this section, we present the results of a simulation study, in which the finite sample performance of the proposed estimators $\widehat{\lambda}_{NW}(y|x)$, $\widehat{\lambda}_{LL}(y|x)$ and $\widehat{\lambda}_n(y|x)$ of $\lambda(y|x)$ at $x = 0.5$ is investigated. In order to obtain an α-mixing observed sequence $\{X_i, Y_i, T_i\}$ after truncation, we generate the observed data as follow.

(1) We generate the triplet (X_1, Y_1, T_1) as follows:
Step 1. We simulate $e_1 \sim N(0, 0.5^2)$ and take $X_1 = e_1$;
Step 2. Let Y_1 follows from the model $Y_1 = \sin(\pi X_1)+\varepsilon_1$, where $\varepsilon_1 \sim N(0, 0.5^2)$;
Step 3. We simulate $T_1 \sim N(\mu, 1)$, where μ is adapted in order to get different values of θ. If $Y_1 < T_1$, we reject the datum and go back to *Step 2*, do this until $Y_1 \geq T_1$. Thus we obtain the observed sample (X_1, Y_1, T_1).
(2) Then we generate the triplet (X_2, Y_2, T_2) as follows:
Step 4. X_2 is generated by an AR(1) model $X_2 = \rho X_1 + e_2$, where $e_2 \sim N(0, 0.5^2)$;
Step 5. Y_2 follows from the model $Y_2 = \sin(\pi X_2) + \varepsilon_2$, where $\varepsilon_2 \sim N(0, 0.5^2)$;
Step 6. We simulate $T_2 \sim N(\mu, 1)$. If $Y_2 < T_2$, we reject the datum and go back to
Step 5, do this until $Y_2 \geq T_2$. Thus we obtain the observed sample (X_2, Y_2, T_2).

By replicating the process (2) above, we generate the observed data (X_i, Y_i, T_i), $i = 1, \ldots, n$. The generating process shows that $X_i = \rho X_{i-1}+e_i$, $Y_i = \sin(\pi X_i)+\varepsilon_i$ and $Y_i \geq T_i$, where $e_i \sim N(0, 0.5^2)$, $\varepsilon_i \sim N(0, 0.5^2)$ and $T_i \sim N(\mu, 1)$, here μ is adapted in order to get different values of θ. Hence the conditional density function $f(y|x) = \frac{1}{0.5\sqrt{2\pi}}\exp\{-\frac{(y-\sin(\pi x))^2}{2\times 0.5^2}\}$, the conditional survival function $S(y|x) = 1 - \Phi(\frac{y-\sin\pi x}{0.5})$, where $\Phi(u)$ stands for the standard normal distribution function, hence $\lambda(y|x) = \frac{f(y|x)}{S(y|x)}$. Note that the α-mixing property (see Doukhan (1994)) of the observable X_i is immediately transferred to the (X_i, Y_i, T_i). For the estimators, we employ the kernel $K(x) = H(x) = \frac{15}{16}(1-x^2)^2 I(|x| \leq 1)$ based on $M = 500$ replications.

5.1 Comparison among $\widehat{\lambda}_{NW}(y|x)$, $\widehat{\lambda}_{LL}(y|x)$ and $\widehat{\lambda}_n(y|x)$

Now we draw random sample with sample size n from the above model. For comparing the estimators, we compute for each estimator $\widehat{\lambda}_{a,b}(\cdot|x)$ of $\lambda(\cdot|x)$ the global mean

Table 1 Minimum GMSEs of $\widehat{\lambda}_{NW}(y|x)$, $\widehat{\lambda}_{LL}(y|x)$ and $\widehat{\lambda}_n(y|x)$ and corresponding optimal bandwidths for several sample sizes and truncation rates

ρ	$\theta(\%)$	n	h_K	h_H	$\widehat{\lambda}_{\text{NW}}$	h_K	h_H	$\widehat{\lambda}_{\text{LL}}$	h_K	h_H	$\widehat{\lambda}_n$
0.1	30	100	0.63	0.69	0.207	0.61	0.79	0.150	0.46	0.66	0.130
		500	0.49	1.12	0.145	0.58	0.73	0.124	0.64	1.11	0.118
	60	100	0.60	1.01	0.173	0.53	0.91	0.134	0.72	1.20	0.130
		500	0.57	0.98	0.128	0.69	0.84	0.108	0.45	1.02	0.106
	90	100	0.50	1.16	0.160	0.74	0.68	0.114	0.51	0.86	0.102
		500	0.71	1.14	0.115	0.52	0.96	0.091	0.52	0.81	0.086
0.9	30	100	0.40	0.69	0.403	0.57	0.94	0.268	0.65	0.75	0.252
		500	0.55	0.82	0.284	0.64	0.80	0.201	0.64	0.78	0.198
	60	100	0.62	0.78	0.373	0.55	1.26	0.250	0.53	0.59	0.249
		500	0.53	0.72	0.239	0.71	1.01	0.174	0.44	0.61	0.164
	90	100	0.55	1.04	0.336	0.47	0.97	0.233	0.49	0.84	0.214
		500	0.50	0.93	0.219	0.57	0.79	0.164	0.53	0.87	0.163

squared errors (GMSE) at $x = 0.5$ and a grid of bandwidths $a := h_K$ and $b := h_H$; the GMSE are defined as

$$\text{GMSE}(a, b) = \frac{1}{Mn}\sum_{l=1}^{M}\sum_{k=1}^{n}[\widehat{\lambda}_{a,b}(Y_k, l|x) - \lambda(Y_k, l|x)]^2.$$

The minimal values of GMSE(a, b) along the grid, and the corresponding optimal bandwidths minimizing the errors, are reported in Table 1.

From Table 1, it can be seen that the minimum GMSE of the estimators decrease as the sample size, this was expected, since we are moving to situations with more sampling information; the values of the minimum GMSE of the estimators seems to be less affected by the truncation proportion for the same sample size. More interestingly, we can appreciate how the estimators $\widehat{\lambda}_{\text{LL}}(y|x)$ and $\widehat{\lambda}_n(y|x)$, respectively, based on plug-in LL and weighted methods outperform the estimator $\widehat{\lambda}_{\text{NW}}(y|x)$ based on NW method in all the considered situations; the estimators $\widehat{\lambda}_{\text{LL}}(y|x)$ and $\widehat{\lambda}_n(y|x)$ have similar performance or the estimator $\hat{\lambda}_n(y|x)$ performs better slightly than the estimator $\hat{\lambda}_{\text{LL}}(y|x)$; In addition, Table 1 shows also that as the dependence of the observations increases, that is, the value of ρ increases, the minimum GMSE of the estimators increase for each considered cases.

5.2 Asymptotic normality

In this subsection, we examine how good the asymptotic normality of the estimator $\widehat{\lambda}_n(y|x)$ is by the normal-probability plots against the normal distribution at $(x, y) = (0.5, 0.5)$. For the estimator, we choose the bandwidths $h_K = h_H = n^{-1/4}$. In Figs. 1, 2, 3 and 4, we plot the normal-probability plots with different θ, ρ and sample sizes, respectively.

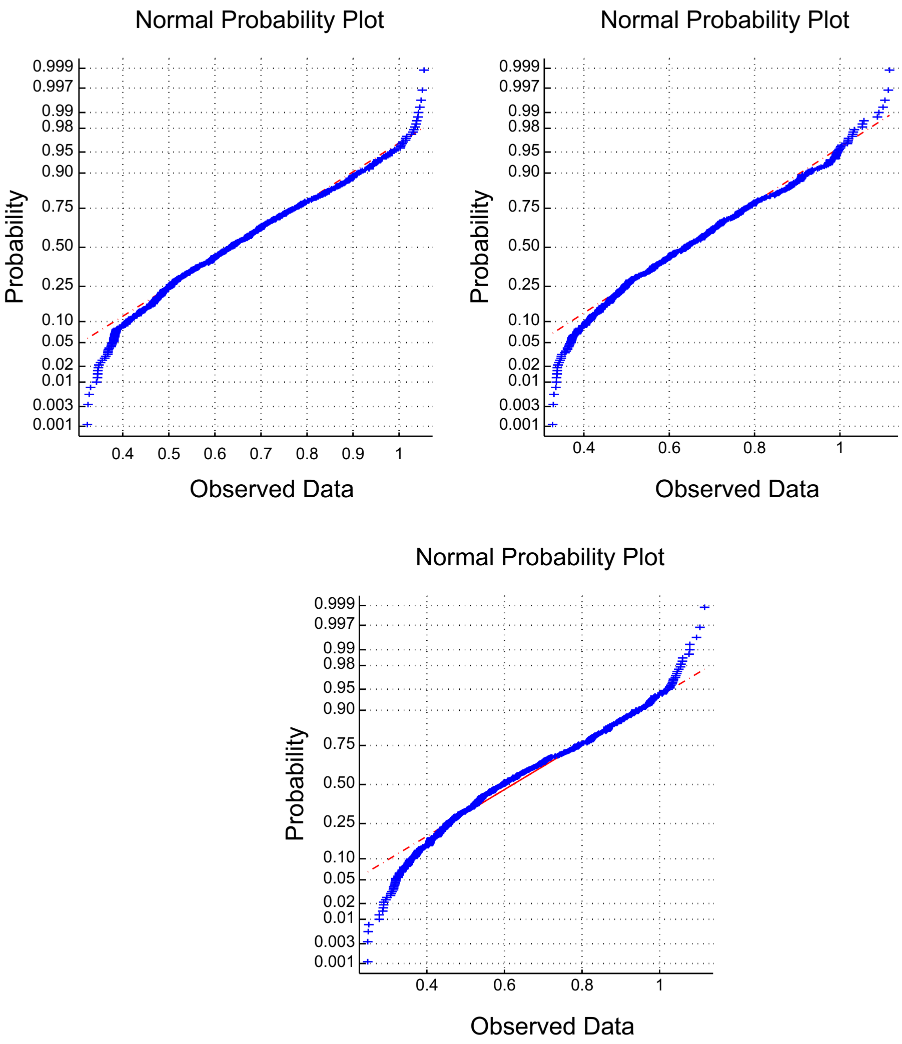

Fig. 1 The normal-probability plots of $\widehat{\lambda}_n(y|x)$ with $\theta \approx 90$ % and $n = 300$. From *left to right*, $\rho = 0.1, 0.5, 0.9$, respectively

From Figs. 1 and 2, it is seen that the sampling distribution of the estimator fits reasonably the normal, this fit being better when increasing the sample size otherwise, and that as the dependence of the observations increases, the quality of fit decreases.

Figures 3 and 4 show again that the normality in the distribution of the estimators increases as increasing of the sample size n, and that the estimator's quality seems to be less affected by the truncation proportion.

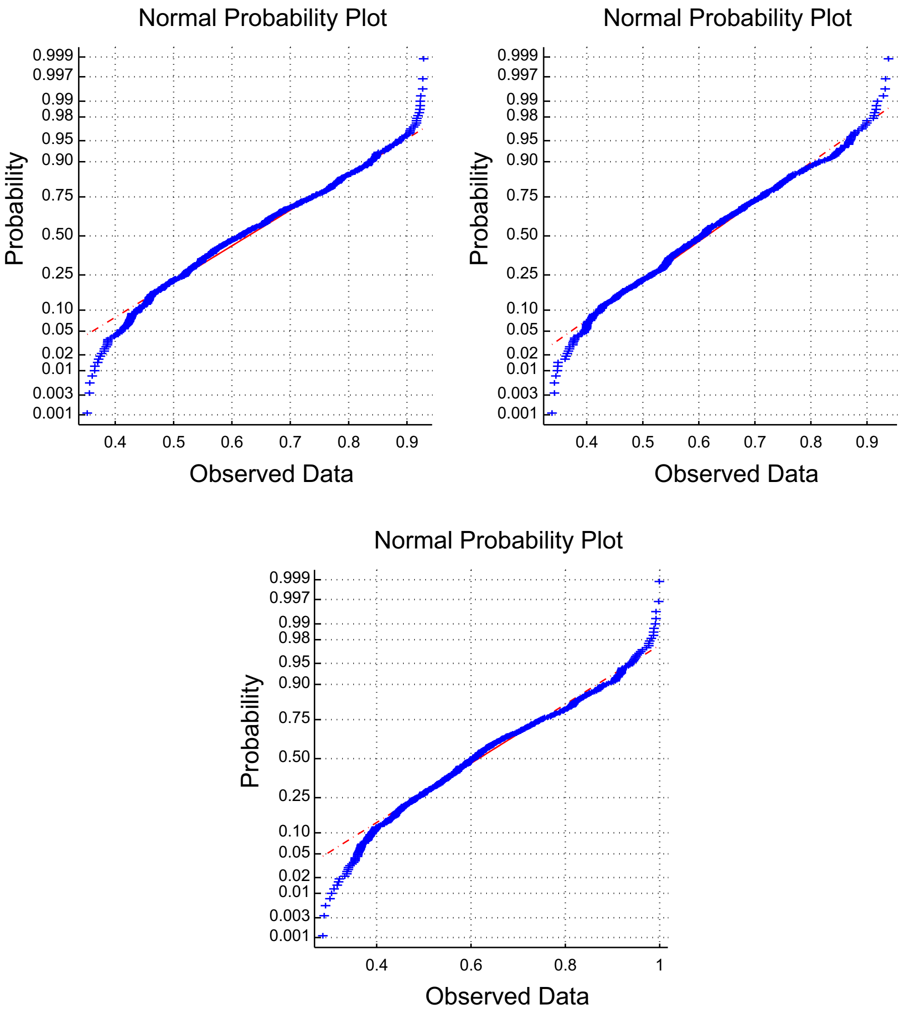

Fig. 2 The normal-probability plots of $\widehat{\lambda}_n(y|x)$ with $\theta \approx 90\ \%$ and $n = 600$. From *left to right*, $\rho = 0.1, 0.5, 0.9$, respectively

5.3 Confidence intervals

In this subsection, we generate the observed data with $\rho = 0.1$ and 0.9, and sample sizes $n = 100$ and 500, respectively, from the model above. In Table 2, we report the coverage probabilities (CP) and average lengths (AL) of 95 % confidence intervals of $\lambda(y|x)$ based on the estimator $\hat{\lambda}_n(y|x)$ for $h_K = h_H = n^{-1/5}$ at $(x, y) = (0.5, 0.5)$.

Table 2 shows that the coverage probabilities of the confidence intervals tend to increase as the sample size n becomes larger, and the average lengths decrease as the sample size or the no truncation proportion increases.

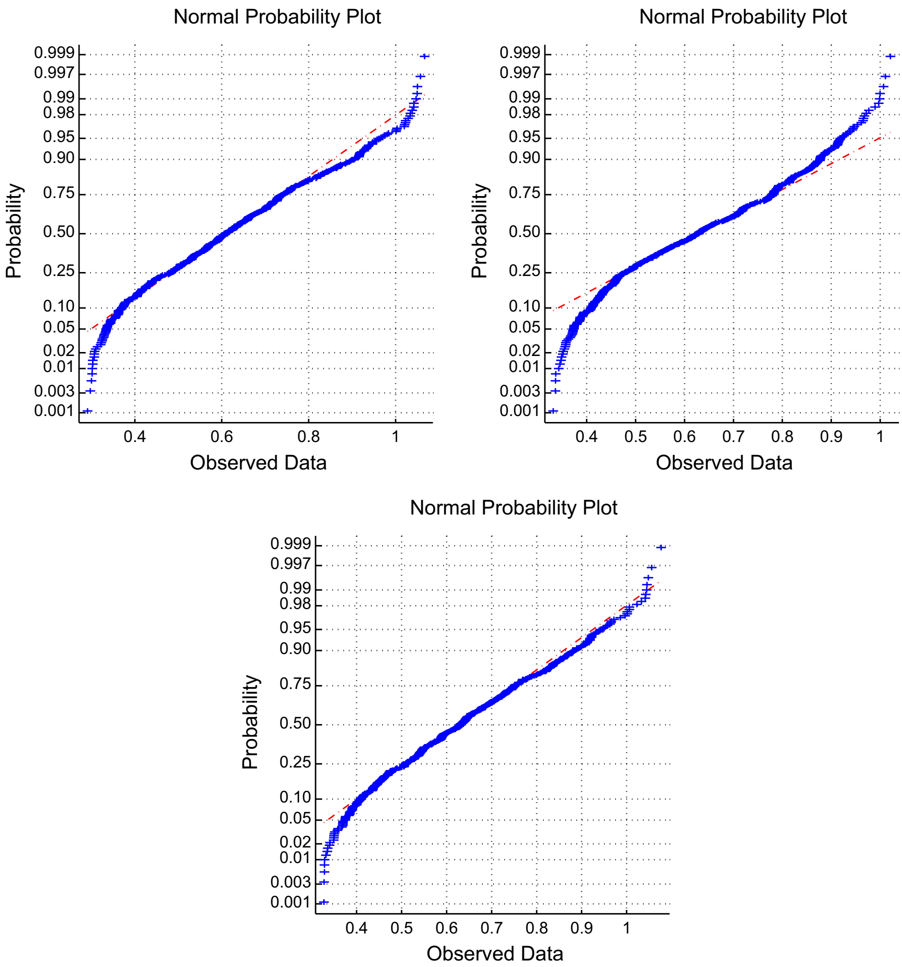

Fig. 3 The normal-probability plots of $\widehat{\lambda}_n(y|x)$ with $\rho = 0.1$ and $n = 300$. From *left to right*, $\theta \approx 30$, 60, 90 %, respectively

In Fig. 5, we plot the confidence bands of 95 % confidence intervals of $\lambda(y|x)$ based on the estimator $\hat{\lambda}_n(y|x)$ with $\rho = 0.1$ for $h_K = h_H = n^{-1/5}$ and $n = 500$ at $x = 0.5$.

From Fig. 5, it can be seen that the confidence bands become slightly narrow as no truncation proportion increases, this can be understood since more sampling information is observed as the values of θ increase.

In order to further show the global performance of the confidence bands of $\widehat{\lambda}_n(y|x)$, we, in Fig. 6, plot the confidence bands of $\widehat{\lambda}_n(y|x)$ for x and y from 0 to 1 with $n = 500$, $\rho = 0.1$ and $\theta \approx 30$, 90 %, respectively. Clearly, Fig. 6 gives similar performance of the confidence bands as that in Fig. 5.

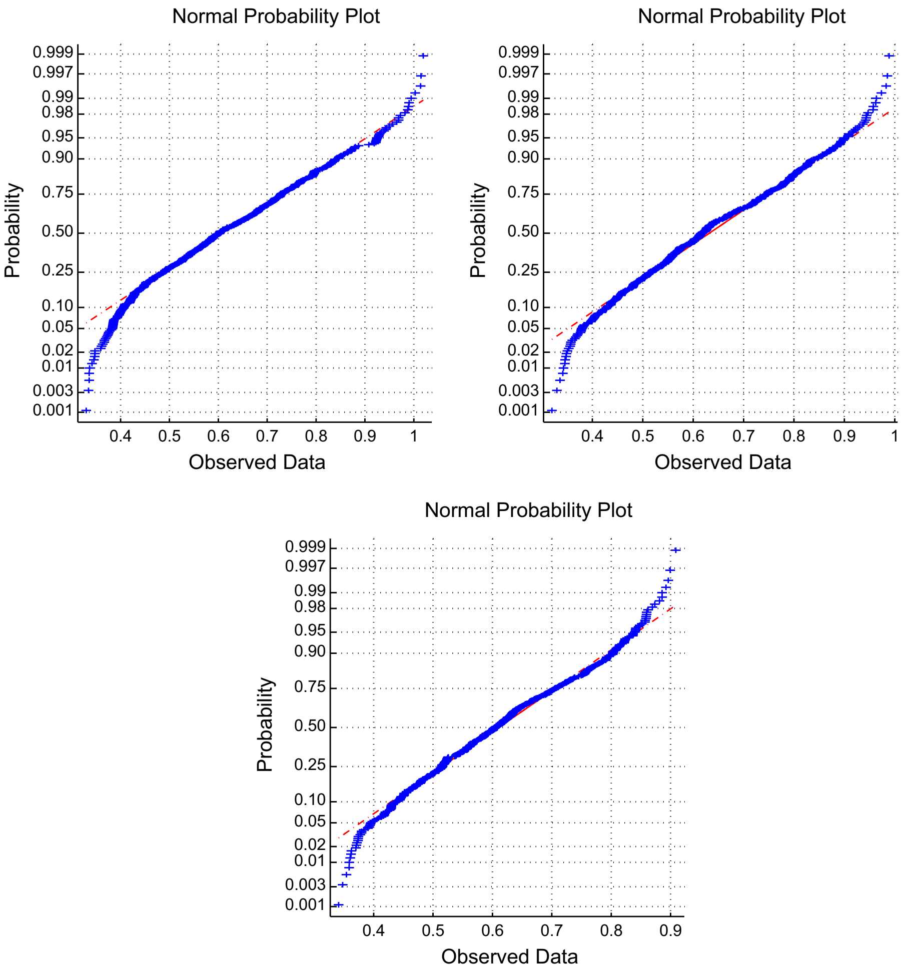

Fig. 4 The normal-probability plots of $\widehat{\lambda}_n(y|x)$ with $\rho = 0.1$ and $n = 600$. From left to right, $\theta \approx 30$, 60, 90 %, respectively

6 Proof of main results

Put $\Theta_j = \frac{\theta}{n}\sum_{i=1}^{n} w_i^j(x)$ for $j = 1,\ 2$ with $w_i(x) = (X_i - x)K_{h_K}(X_i - x)$.

Lemma 1 *Let* $\alpha(n) = O(n^{-\gamma})$ *for some* $\gamma > 3$. *Suppose that (A1) for* K, *(A2) and (A3)(i) are satisfied. If* $nh_K \to \infty$ *and* $nh_K^{1+r} = O(1)$ *for some constant* $r \geq 2$, *then* $\eta = O_{\mathbf{P}}((nh_K)^{-1/2} + h_K)$ *and* $\max_{1\leq i\leq n} |\eta w_i(x)| = o_{\mathbf{P}}(1)$.

Table 2 The coverage probabilities (CP) and average lengths (AL) of 95 % confidence intervals of $\lambda(y|x)$ based on $\hat{\lambda}_n(y|x)$ at $(x, y) = (0.5, 0.5)$ for several sample sizes and truncation rates

ρ	θ (%)	n	CP	AL
0.1	30	100	0.866	0.473
		500	0.910	0.290
	60	100	0.882	0.426
		500	0.912	0.262
	90	100	0.896	0.401
		500	0.922	0.228
0.9	30	100	0.754	0.566
		500	0.852	0.342
	60	100	0.822	0.535
		500	0.874	0.290
	90	100	0.848	0.479
		500	0.904	0.279

Proof From (8) it is easy to see that

$$\begin{aligned} 0 &= \left|\frac{\theta}{n}\sum_{i=1}^{n}\frac{w_i(x)}{1+\eta w_i(x)}\right| = \left|\frac{\theta}{n}\sum_{i=1}^{n}w_i(x) - \frac{\theta\eta}{n}\sum_{i=1}^{n}\frac{w_i^2(x)}{1+\eta w_i(x)}\right| \\ &\geq \frac{|\eta|\Theta_2}{1+\max_{1\leq i\leq n}|\eta w_i(x)|} - |\Theta_1|. \end{aligned}$$

If we can prove $\Theta_1 = O_{\mathbf{P}}(h_K^2 + (h_K/n)^{1/2})$ and

$$\Theta_2 = h_K \ell(x)\mathbb{E}\{G(Y)|X=x\}\Delta_{22} + O_{\mathbf{P}}((h_K/n)^{1/2}) + O(h_K^3), \tag{9}$$

then from $nh_K \to \infty$ we have

$$\frac{|\eta|}{1+\max_{1\leq i\leq n}|\eta w_i(x)|} \leq \frac{|\Theta_1|}{\Theta_2} = O_{\mathbf{P}}(h_K + (nh_K)^{-1/2}). \tag{10}$$

Since $\mathbf{E}|h_K^{-1/r}w_i(x)|^r < \infty$ from (A1) and (A2), $\max_{1\leq i\leq n}|w_i(x)| = o((nh_K)^{1/r})$ $a.s.$ from the proof of Lemma 3 in Owen (1990). Therefore, from $nh_K^{1+r} = O(1)$ for some $r \geq 2$ and (10) it follows that

$$\eta = O_{\mathbf{P}}(h_K + (nh_K)^{-1/2})$$

and $\max_{1\leq i\leq n}|\eta w_i(x)| = O_{\mathbf{P}}(h_K + (nh_K)^{-1/2}) \cdot o_{\mathbf{P}}((nh_K)^{1/r}) = o_{\mathbf{P}}(1)$.

Next we verify $\Theta_1 = O_{\mathbf{P}}(h_K^2 + (h_K/n)^{1/2})$ and (9). We prove only (9), the evaluation related to Θ_1 is similar. Note that $\Theta_2 = \mathbf{E}\Theta_2 + O_{\mathbf{P}}(\sqrt{\mathrm{Var}(\Theta_2)})$. So, we need to evaluate $\mathbf{E}\Theta_2$ and $\mathrm{Var}(\Theta_2)$. Note that the second derivatives of $\mathbb{E}(G(Y)|X =$

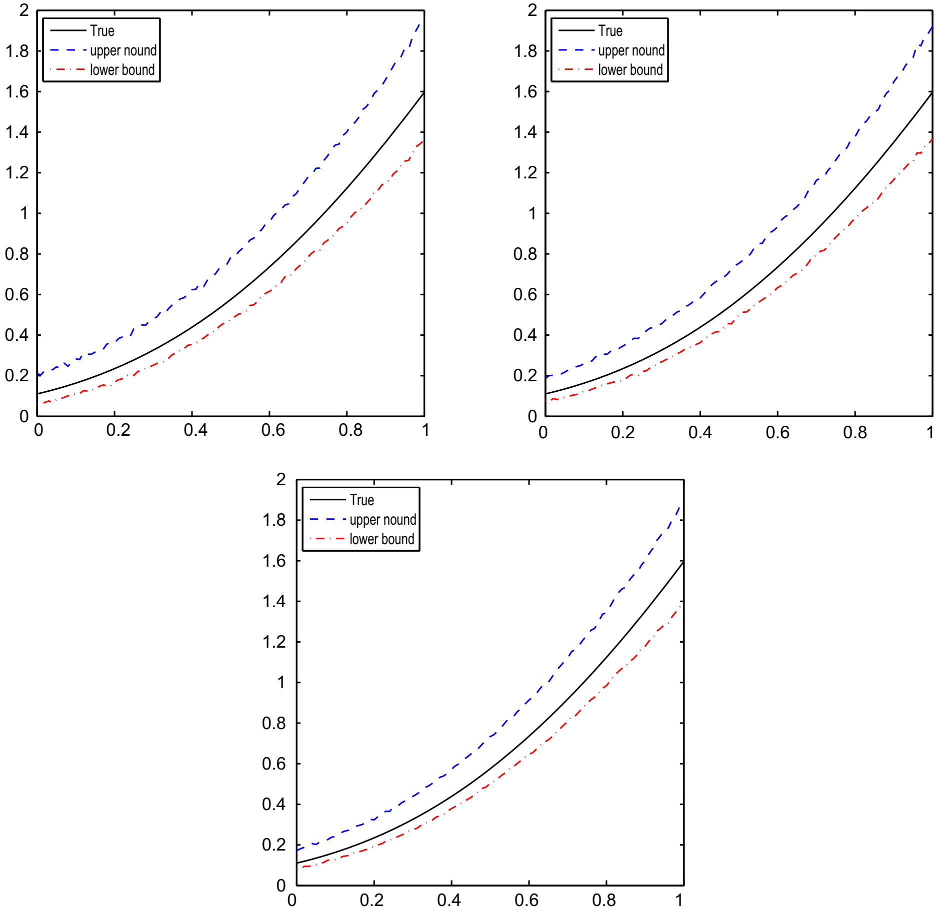

Fig. 5 Confidence bands of $\lambda(y|x)$ based on $\hat{\lambda}_n(y|x)$ with $\rho = 0.1$ and $n = 500$ at $x = 0.5$. From *left to right*, $\theta \approx 30, 60, 90$ %, respectively

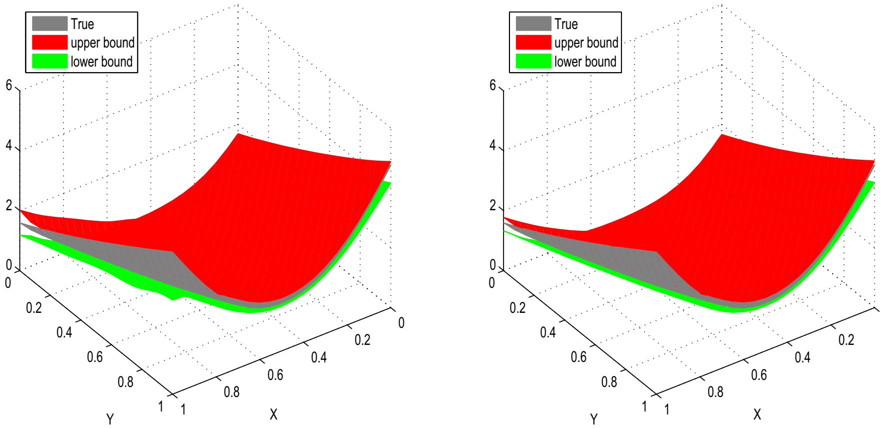

Fig. 6 Confidence band of $\widehat{\lambda}_n(y|x)$ with $\rho = 0.1$. From *left to right*, $\theta \approx 30, 90$ %

$x) = \int G(y)f(y|x)dy$ are continuous with respect to x from (A2). Then, from $f^*(x,y) = \theta^{-1}G(y)f(x,y)$ (see (2)) and according to (A1) and (A2) we have

$$\begin{aligned}\mathbf{E}\Theta_2 &= \frac{\theta}{nh_K^2}\sum_{i=1}^{n}\mathbf{E}\Big(K^2\Big(\frac{X_i-x}{h_K}\Big)(X_i-x)^2\Big)\\ &= h_K\int_{\mathbb{R}} s^2K^2(s)\ell(x+h_Ks)\mathbb{E}(G(Y)|X=x+h_Ks)ds\\ &= h_K\ell(x)\mathbb{E}\{G(Y)|X=x\}\Delta_{22}+O(h_K^3).\end{aligned} \tag{11}$$

Write

$$\begin{aligned}\mathrm{Var}(\Theta_2) &= \Big(\frac{\theta}{nh_K^2}\Big)^2\Big\{\sum_{i=1}^{n}\mathrm{Var}\Big(K^2\Big(\frac{X_i-x}{h_K}\Big)(X_i-x)^2\Big)\\ &\quad+\sum_{i\neq j}\mathrm{Cov}\Big(K^2\Big(\frac{X_i-x}{h_K}\Big)(X_i-x)^2, K^2\Big(\frac{X_j-x}{h_K}\Big)(X_j-x)^2\Big)\Big\}\\ &:= \Theta_{21}+\Theta_{22}.\end{aligned}$$

In view of (2), from (A1) and (A2) we obtain that

$$\begin{aligned}\Theta_{21} &= \frac{\theta^2}{nh_K^4}\Big\{\mathbf{E}\Big[K^4\Big(\frac{X_i-x}{h_K}\Big)(X_i-x)^4\Big]-\Big[\mathbf{E}\Big(K^2\Big(\frac{X_i-x}{h_K}\Big)(X_i-x)^2\Big)\Big]^2\Big\}\\ &= \frac{\theta^2}{nh_K^4}\Big\{\frac{h_K^5}{\theta}\int_{\mathbb{R}} s^4K^4(s)\mathbb{E}\big(G(Y)|X=x+h_Ks\big)\ell(x+h_Ks)ds\\ &\quad-\frac{h_K^6}{\theta^2}\Big(\int_{\mathbb{R}} s^2K^2(s)\mathbb{E}(G(Y)|X=x+h_Ks)\ell(x+h_Ks)ds\Big)^2\Big\} = O(h_K/n).\end{aligned} \tag{12}$$

Let $\xi_i = \frac{\theta}{h_K^2}K^2(\frac{X_i-x}{h_K})(X_i-x)^2$. Then

$$\Theta_{22} = \frac{1}{n^2}\left\{\sum_{|i-j|\le[h_K^{-1}]}+\sum_{|i-j|>[h_K^{-1}]}\right\}\mathrm{Cov}(\xi_i,\xi_j). \tag{13}$$

For $i<j$, applying (A1)–(A3) we have

$$\begin{aligned}&|\mathrm{Cov}(\xi_i,\xi_j)|\\ &= \frac{\theta^2}{h_K^4}\Big|\mathbf{E}\Big(K^2\Big(\frac{X_i-x}{h_K}\Big)K^2\Big(\frac{X_j-x}{h_K}\Big)(X_i-x)^2(X_j-x)^2\Big)\\ &\quad-\Big[\mathbf{E}\Big(K^2\Big(\frac{X_i-x}{h_K}\Big)(X_i-x)^2\Big)\Big]^2\Big|\\ &\le \frac{\theta^2h_K^6}{h_K^4}\int_{\mathbb{R}}\int_{\mathbb{R}}(stK(s)K(t))^2l_{j-i}^*(x+h_Ks,x+h_Kt)dsdt+O(h_K^2) = O(h_K^2).\end{aligned} \tag{14}$$

On the other hand, from Lemma 3 it follows $|\mathrm{Cov}(\xi_i,\xi_j)| \le C[\alpha(j-i)]^{1-1/\gamma}(E|\xi_i|^{2\gamma})^{1/\gamma}$ and

$$\begin{aligned}\mathbf{E}|\xi_i|^{2\gamma} &= \frac{\theta^{2\gamma}}{h_K^{4\gamma}}\mathbf{E}\Big|K\Big(\frac{X_i-x}{h_K}\Big)(X_i-x)\Big|^{4\gamma}\\ &\le \theta^{2\gamma-1}h_K\int_{\mathbb{R}}|sK(s)|^{4\gamma}\ell(x+h_Ks)ds = O(h_K),\end{aligned}$$

which, together with (14), gives $|\mathrm{Cov}(\xi_i,\xi_j)| \le C\min\{h_K^2, [\alpha(j-i)]^{1-1/\gamma}h_K^{1/\gamma}\}$. Thus from (13) we have $\Theta_{22} = O(h_K/n) + O\big(n^{-1}h_K^{1/\gamma}\sum_{l=[h_K^{-1}]}^{\infty}[\alpha(l)]^{1-1/\gamma}\big) = O(h_K/n)$. Then $\mathrm{Var}(\Theta_2) = O(h_K/n)$. Therefore, $\Theta_2 = h_K\ell(x)\mathbb{E}\{G(Y)|X=x\}\Delta_{22} + O_{\mathbf{P}}((h_K/n)^{1/2}) + O(h_K^3)$. □

Proof of Proposition 1 Put $\ell_n(x) = \theta\sum_{j=1}^n \widehat{p}_j(x)K_{h_K}(X_j-x)G_n^{-1}(Y_j)$. Write

$$\begin{aligned}\widehat{f}_n(y|x) - f(y|x) &= \ell_n^{-1}(x)\cdot\theta\sum_{i=1}^n \widehat{p}_i(x)G_n^{-1}(Y_i)K_{h_K}(X_i-x)\big\{[H_{h_H}(Y_i-y) - f(y|X_i)]\\ &\quad + [f(y|X_i) - f(y|x)]\big\} := \ell_n^{-1}(x)[U_n(y|x) + V_n(y|x)].\end{aligned}$$

Note that $\widehat{p}_i(x) = \frac{1}{n}\cdot\frac{1}{1+\eta w_i(x)} = \frac{1}{n}\big\{\sum_{j=1}^k(-\eta w_i(x))^{j-1} + \frac{(-\eta w_i(x))^k}{1+\eta w_i(x)}\big\}$ for positive integer k. Write

$$\begin{aligned}&U_n(y|x)\\ &\quad= \frac{\theta}{nh_K}\Big\{\sum_{i=1}^n\Big(\frac{1}{G_n(Y_i)} - \frac{1}{G(Y_i)}\Big)\frac{H_{h_H}(Y_i-y) - f(y|X_i)}{1+\eta w_i(x)}K\Big(\frac{X_i-x}{h_K}\Big)\\ &\qquad + \sum_{i=1}^n\Big[\frac{H_{h_H}(Y_i-y) - f(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\\ &\qquad - \mathbf{E}\Big(\frac{H_{h_H}(Y_i-y) - f(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big]\\ &\qquad + \sum_{i=1}^n\mathbf{E}\Big(\frac{H_{h_H}(Y_i-y) - f(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\\ &\qquad - \eta\sum_{i=1}^n\Big[\frac{[H_{h_H}(Y_i-y) - f(y|X_i)]w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\\ &\qquad - \mathbf{E}\Big(\frac{[H_{h_H}(Y_i-y) - f(y|X_i)]w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big]\\ &\qquad - \eta\sum_{i=1}^n\mathbf{E}\Big(\frac{[H_{h_H}(Y_i-y) - f(y|X_i)]w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\\ &\qquad + \eta^2\sum_{i=1}^n\Big[\frac{[H_{h_H}(Y_i-y) - f(y|X_i)]w_i^2(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\end{aligned}$$

$$
\begin{aligned}
&- \mathbf{E}\Big(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)]w_i^2(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big]\\
&+\eta^2\sum_{i=1}^n \mathbf{E}\Big(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)]w_i^2(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\\
&-\eta^3\sum_{i=1}^n \frac{[H_{h_H}(Y_i-y)-f(y|X_i)]w_i^3(x)}{G(Y_i)(1+\eta w_i(x))}K\Big(\frac{X_i-x}{h_K}\Big)\Big\}\\
:=\ & \frac{\theta}{nh_K}\big\{U_{n1}(y|x)+U_{n2}(y|x)+U_{n3}(y|x)-U_{n4}(y|x)-U_{n5}(y|x)+U_{n6}(y|x)\\
&+U_{n7}(y|x)-U_{n8}(y|x)\big\}.
\end{aligned}
$$

In view of (7) we have

$$
V_n(y|x)=(\theta/2)\sum_{i=1}^n (X_i-x)^2 f^{(2,0)}(y|X_i^*)\widehat{p}_i(x)K_{h_K}(X_i-x)G_n^{-1}(Y_i),
$$

where X_i^* is between X_i and x. Then

$$
\begin{aligned}
&V_n(y|x)\\
&=\frac{\theta}{2nh_K}\Big\{\sum_{i=1}^n\Big(\frac{1}{G_n(Y_i)}-\frac{1}{G(Y_i)}\Big)\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)}{1+\eta w_i(x)}K\Big(\frac{X_i-x}{h_K}\Big)\\
&\quad+\sum_{i=1}^n\Big[\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\\
&\quad-\mathbf{E}\Big(\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big]\\
&\quad+\sum_{i=1}^n\mathbf{E}\Big(\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\\
&\quad-\eta\sum_{i=1}^n\Big[\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\\
&\quad-\mathbf{E}\Big(\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big]\\
&\quad-\eta\sum_{i=1}^n\mathbf{E}\Big(\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\\
&\quad+\eta^2\sum_{i=1}^n\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)w_i^2(x)}{G(Y_i)[1+\eta w_i(x)]}K\Big(\frac{X_i-x}{h_K}\Big)\Big\}\\
&:=\frac{\theta}{2nh_K}\big\{V_{n1}(y|x)+V_{n2}(y|x)+V_{n3}(y|x)-V_{n4}(y|x)-V_{n5}(y|x)+V_{n6}(y|x)\big\}.
\end{aligned}
$$

Following the line in Step 1 in the proof of Liang (2012), it is easy to verify that $\ell_n(x) = \ell(x) + o_{\mathbf{P}}(1)$. It suffices to show that

$$
\begin{aligned}
&\frac{\theta}{nh_K} U_{n1}(y|x) = O_{\mathbf{P}}(n^{-1/2}),\\
&\sqrt{nh_K h_H} \cdot \frac{\theta}{nh_K} U_{n2}(y|x) \xrightarrow{\mathscr{D}} N\left(0, \frac{\theta \ell(x) f(y|x)}{G(y)} \int_{\mathbb{R}^2} K^2(u) H^2(v) du dv\right),\\
&\frac{\theta}{nh_K} U_{n3}(y|x) = \frac{h_H^2}{2} \ell(x) f^{(0,2)}(y|x) \int v^2 H(v) dv + o(h_H^2),\\
&\frac{\theta}{nh_K} U_{n4}(y|x) = O_{\mathbf{P}}\left(\frac{1}{nh_K \sqrt{h_H}} + \frac{h_K}{\sqrt{nh_K h_H}}\right),\\
&\frac{\theta}{nh_K} U_{n5}(y|x) = O_{\mathbf{P}}(h_K^2 h_H^2) + O_{\mathbf{P}}\left(\frac{h_K h_H^2}{\sqrt{nh_K}}\right),\\
&\frac{\theta}{nh_K} U_{n6}(y|x) = O_{\mathbf{P}}\left(\frac{1}{nh_K \sqrt{nh_K h_H}} + \frac{h_K^2}{\sqrt{nh_K h_H}}\right),\\
&\frac{\theta}{nh_K} U_{n7}(y|x) = O(h_K^2 h_H^2) + O_{\mathbf{P}}\left(\frac{h_K h_H^2}{\sqrt{nh_K}}\right) + O_{\mathbf{P}}\left(\frac{h_H^2}{nh_K}\right),\\
&\frac{\theta}{nh_K} U_{n8}(y|x) = O_{\mathbf{P}}((nh_K)^{-3/2} + h_K^3),\\
&\frac{\theta}{nh_K} V_{n1}(y|x) = O_{\mathbf{P}}\left(\frac{h_K^2}{\sqrt{n}}\right), \ \frac{\theta}{2nh_K} V_{n2}(y|x) = O_{\mathbf{P}}\left(\sqrt{\frac{h_K^3}{n}}\right),\\
&\frac{\theta}{2nh_K} V_{n3}(y|x) = \frac{h_K^2}{2} \ell(x) f^{(2,0)}(y|x) \int u^2 K(u) du + o(h_K^2),\\
&\frac{\theta}{2nh_K} V_{n4}(y|x) = O_{\mathbf{P}}\left(\frac{h_K}{n} + h_K^2 \sqrt{\frac{h_K}{n}}\right), \ \frac{\theta}{2nh_K} V_{n5}(y|x) = O_{\mathbf{P}}\left(h_K^3 + \frac{h_K^2}{\sqrt{nh_K}}\right),\\
&\frac{\theta}{2nh_K} V_{n6}(y|x) = O_{\mathbf{P}}\left(\frac{h_K}{n} + h_K^4\right).
\end{aligned}
$$

Step 1. We prove $\sqrt{nh_K h_H} \cdot \frac{\theta}{nh_K} U_{n2}(y|x) \xrightarrow{\mathscr{D}} N\big(0, \frac{\theta \ell(x) f(y|x)}{G(y)} \int_{\mathbb{R}^2} K^2(u) H^2(v) du dv\big)$.

From (A4), it follows that there exists a sequence of positive integers $\delta_n \to \infty$ such that $\delta_n q_n = o((nh_K h_H)^{1/2})$ and $\delta_n (n(h_K h_H)^{-1})^{1/2} \alpha(q_n) \to 0$. Let $w_n = [\frac{n}{p_n + q_n}]$ and $p_n = [(nh_K h_H)^{1/2}/\delta_n]$. Then

$$
q_n/p_n \to 0, \ w_n \alpha(q_n) \to 0, \ w_n q_n/n \to 0, \ p_n/n \to 0, \ p_n/(nh_K h_H)^{1/2} \to 0. \tag{15}
$$

Define $\chi_{mn} = \sum_{i=k_m}^{k_m+p_n-1} \eta_i$, $\chi'_{mn} = \sum_{i=l_m}^{l_m+q_n-1} \eta_i$, $\chi''_{w_n n} = \sum_{i=w_n(p_n+q_n)+1}^{n} \eta_i$ with

$$\eta_i = \theta\sqrt{\frac{h_H}{h_K}}\Big[\frac{H_{h_H}(Y_i-y)-f(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\\ -\mathbf{E}\Big(\frac{H_{h_H}(Y_i-y)-f(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big],$$

where $k_m = (m-1)(p_n+q_n)+1, l_m = (m-1)(p_n+q_n)+p_n+1, m = 1, \ldots, w_n$. Then

$$\sqrt{nh_Kh_H}\cdot\frac{\theta}{nh_K}U_{n2}(y|x) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_i = \frac{1}{\sqrt{n}}\left\{\sum_{m=1}^{w_n}\chi_{mn}+\sum_{m=1}^{w_n}\chi'_{mn}+\chi''_{w_n n}\right\}\\ := \frac{1}{\sqrt{n}}\{S'_n+S''_n+S'''_n\}.$$

Then it is sufficient to show that

$$n^{-1}\mathbf{E}(S''_n)^2 \to 0,\ n^{-1}\mathbf{E}(S'''_n)^2 \to 0, \tag{16}$$

$$n^{-1}\text{Var}(S'_n) \to \frac{\theta f(x,y)}{G(y)}\int_{\mathbb{R}^2}K^2(u)H^2(v)dudv, \tag{17}$$

$$\Big|\mathbf{E}\exp\Big(it\sum_{m=1}^{w_n}n^{-1/2}\chi_{mn}\Big)-\prod_{m=1}^{w_n}\mathbf{E}\exp\big(itn^{-1/2}\chi_{mn}\big)\Big| \to 0, \tag{18}$$

$$A_n(\varepsilon) = \frac{1}{n}\sum_{m=1}^{w_n}\mathbf{E}\chi_{mn}^2 I(|\chi_{mn}|>\varepsilon\sqrt{n}) \to 0\ \ \forall\varepsilon>0. \tag{19}$$

It is easy to see that

$$\text{Var}(\eta_i) = \frac{h_H\theta^2}{h_K}\Big\{\mathbf{E}\Big(\frac{K^2(\frac{X_i-x}{h_K})}{G^2(Y_i)}\Big[\frac{1}{h_H^2}H^2\Big(\frac{Y_i-y}{h_H}\Big)-\frac{2f(y|X_i)}{h_H}H\Big(\frac{Y_i-y}{h_H}\Big)\\ +f^2(y|X_i)\Big]-\Big[\mathbf{E}\Big(\frac{H_{h_H}(Y_i-y)-f(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big]^2\Big\}\\ = \theta\int_{\mathbb{R}^2}\frac{K^2(u)H^2(v)f(x+h_Ku,y+h_Hv)dudv}{G(y+h_Hv)}\\ -2\theta h_H\int_{\mathbb{R}^2}\frac{K^2(u)H(v)f(y|x+h_Ku)f(x+h_Ku,y+h_Hv)dudv}{G(y+h_Hv)}\\ +\theta h_H\int_{\mathbb{R}^2}\frac{K^2(u)f^2(y|x+h_Ku)f(x+h_Ku,v)dudv}{G(v)}\\ -h_Kh_H\Big\{h_H\int_{\mathbb{R}^2}K(u)H(v)f(x+h_Ku,y+h_Hv)dudv$$

$$
\begin{aligned}
&- \int_{\mathbb{R}^2} K(u) f(y|x + h_K u) f(x + h_K u, v) du dv \Big\}^2 \\
&\to \frac{\theta f(x, y)}{G(y)} \int_{\mathbb{R}^2} K^2(u) H^2(v) du dv. \qquad (20)
\end{aligned}
$$

From (A3), for $i < j$ we have

$$
\begin{aligned}
&|\mathrm{Cov}(\eta_i, \eta_j)| \le |\mathbf{E}(\eta_i, \eta_j)| + |\mathbf{E}\eta_i \cdot \mathbf{E}\eta_j| \\
&\le \theta^2 \frac{h_H}{h_K} \Big| \mathbf{E} \Big\{ \frac{K(\frac{X_i - x}{h_K}) K(\frac{X_j - x}{h_K})}{G(Y_i) G(Y_j)} \Big[\frac{1}{H_H^2} H\Big(\frac{Y_i - y}{h_H}\Big) H\Big(\frac{Y_j - y}{h_H}\Big) \\
&- \frac{1}{H_H} f(y|X_i) H\Big(\frac{Y_j - y}{h_H}\Big) \\
&- \frac{1}{H_H} f(y|X_j) H\Big(\frac{Y_i - y}{h_H}\Big) + f(y|X_i) f(y|X_j) \Big] \Big\} \Big| \\
&+ \theta^2 \frac{h_H}{h_K} \Big[\mathbf{E}\Big(\frac{H_{h_H}(Y_i - y) - f(y|X_i)}{G(Y_i)} K\Big(\frac{X_i - x}{h_K}\Big)\Big)\Big]^2 = O(h_H h_K).
\end{aligned}
$$

In addition, applying Lemma 3 (take $p = q = 20\gamma$), one can prove that for $i < j$, $|\mathrm{Cov}(\eta_i, \eta_j)| \le C[\alpha(j - i)]^{1-1/(10\gamma)} (\mathbf{E}|\eta_i|^{20\gamma})^{1/(10\gamma)}$ and $\mathbf{E}|\eta_i|^{20\gamma} = O((h_H h_K)^{-(10\gamma - 1)})$. Thus

$$
\begin{aligned}
&\frac{1}{n} \sum_{1 \le i < j \le n} |\mathrm{Cov}(\eta_i, \eta_j)| \\
&= \frac{O(1)}{n} \Bigg(\sum_{1 \le j - i \le c_n} + \sum_{j - i \ge c_n + 1} \Bigg) \min\{h_H h_K, [\alpha(j - i)]^{1 - 1/(10\gamma)} \\
&\quad \times (h_H h_K)^{-(1 - 1/(10\gamma))}\} \\
&= O(1)\big[c_n h_H h_K + (h_H h_K)^{-(1 - 1/(10\gamma))} c_n^{-(\gamma - 11/10)}\big] \to 0, \qquad (21)
\end{aligned}
$$

where $c_n = [(h_H h_K)^{-(1-1/(10\gamma))/\phi}]$ for some $1 - 1/(10\gamma) < \phi < \gamma - 11/10$.

From (20), (21), according to (15) we have

$$
\begin{aligned}
\frac{1}{n}\mathbf{E}(S_n'')^2 &= \frac{1}{n} \sum_{m=1}^{w_n} \sum_{i=l_m}^{l_m + q_n - 1} \mathrm{Var}(\eta_i) + \frac{2}{n} \sum_{m=1}^{w_n} \sum_{l_m \le i < j \le l_m + q_n - 1} \mathrm{Cov}(\eta_i, \eta_j) \\
&\quad + \frac{2}{n} \sum_{1 \le i < j \le w_n} \mathrm{Cov}(\chi'_{in}, \chi'_{jn}) \\
&= O(w_n q_n / n) + \frac{O(1)}{n} \sum_{1 \le i < j \le n} |\mathrm{Cov}(\eta_i, \eta_j)| \to 0, \\
\frac{1}{n}\mathbf{E}(S_n''')^2 &= \frac{1}{n} \sum_{i = w_n(p_n + q_n) + 1}^{n} \mathrm{Var}(\eta_i) + \frac{2}{n} \sum_{w_n(p_n + q_n) + 1 \le i < j \le n} \mathrm{Cov}(\eta_i, \eta_j) \to 0,
\end{aligned}
$$

and $n^{-1}\mathrm{Var}(S_n') \to \frac{\theta f(x,y)}{G(y)} \int_{\mathbb{R}^2} K^2(u)H^2(v)dudv$ by $w_n p_n/n \to 1$. Thus (16) and (17) are proved.

As for (18), according to Lemma 2, from (15) we have

$$\left|\mathbf{E}\exp\left(it\sum_{m=1}^{w_n} n^{-1/2}\chi_{mn}\right) - \prod_{m=1}^{w_n}\mathbf{E}\exp\left(itn^{-1/2}\chi_{mn}\right)\right| \le 16w\alpha(q_n+1) \to 0.$$

Finally, we establish (19). Note that $\max_{1\le m\le w}|\chi_{mn}| = O\big(p_n/\sqrt{h_K h_H}\big)$, which leads that for large n, $I(|\chi_{mn}| > \varepsilon\sqrt{n}) = 0$ by $p_n/\sqrt{nh_K h_H} \to 0$. Therefore $A_n(\varepsilon) \to 0$.

Step 2. Observe that

$$\begin{aligned}\frac{\theta}{nh_K}|U_{n1}(y|x)| \le & \frac{\sup_{y\ge a_F}|G_n(y)-G(y)|}{[1-\max_{1\le i\le n}|\eta w_i(x)|][G(a_F)-\sup_{y\ge a_F}|G_n(y)-G(y)|]} \\ & \cdot \frac{\theta}{nh_K}\sum_{i=1}^{n}\frac{|H_{h_H}(Y_i-y)-f(y|X_i)|}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big), \\ \frac{\theta}{2nh_K}|V_{n1}(y|x)| \le & \frac{\sup_{y\ge a_F}|G_n(y)-G(y)|}{[1-\max_{1\le i\le n}|\eta w_i(x)|][G(a_F)-\sup_{y\ge a_F}|G_n(y)-G(y)|]} \\ & \cdot \frac{\theta}{2nh_K}\sum_{i=1}^{n}\frac{(X_i-x)^2|f^{(2,0)}(y|X_i^*)|}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big).\end{aligned}$$

From (A1) and (A2), it is easy to verity that $\frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\big\{\frac{|H_{h_H}(Y_i-y)-f(y|X_i)|}{G(Y_i)}K(\frac{X_i-x}{h_K})\big\}$ $= O(1)$ and $\frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\big\{\frac{(X_i-x)^2|f^{(2,0)}(y|X_i^*)|}{G(Y_i)}K(\frac{X_i-x}{h_K})\big\} = O(h_K^2)$, which follow that

$$\frac{\theta}{nh_K}\sum_{i=1}^{n}\frac{|H_{h_H}(Y_i-y)-f(y|X_i)|}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big) = O_{\mathbf{P}}(1)$$

and $\frac{\theta}{nh_K}\sum_{i=1}^n \frac{(X_i-x)^2|f^{(2,0)}(y|X_i^*)|}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big) = O_{\mathbf{P}}(h_K^2)$. Therefore, in view of Lemmas 1 and 4 we have $\frac{\theta}{nh_K}U_{n1}(y|x) = O_{\mathbf{P}}(n^{-1/2})$ and $\frac{\theta}{nh_K}V_{n1}(y|x) = O_{\mathbf{P}}(n^{-1/2}h_K^2)$.

Similarly

$$\begin{aligned}\frac{\theta}{nh_K}|U_{n8}(y|x)| \le & \frac{|\eta|^3}{[1-\max_{1\le i\le n}|\eta w_i(x)|} \cdot \frac{\theta}{nh_K^4} \\ & \times \sum_{i=1}^{n}\frac{|[H_{h_H}(Y_i-y)-f(y|X_i)](X_i-x)^3|K^4\big(\frac{X_i-x}{h_K}\big)}{G(Y_i)} \\ = & O_{\mathbf{P}}((nh_K)^{-3/2}+h_K^3),\end{aligned}$$

$$\frac{\theta}{2nh_K}|V_{n6}(y|x)| \le \frac{\eta^2}{[1-\max_{1\le i\le n}|\eta w_i(x)|} \cdot \frac{\theta}{2nh_K^3}\sum_{i=1}^n \frac{(X_i-x)^4|f^{(2,0)}(y|X_i^*)|}{G(Y_i)}$$
$$\times K^3\Big(\frac{X_i-x}{h_K}\Big)$$
$$= O_{\mathbf{P}}((nh_K)^{-1}+h_K^2)h_K^2 = O_{\mathbf{P}}(n^{-1}h_K+h_K^4).$$

Step 3. From (A1) and (A2) we have

$$\frac{\theta}{nh_K}U_{n3}(y|x) = \frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{H_{h_H}(Y_i-y)-f(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)$$
$$= \int\int K(u)H(v)\ell(x+h_K u)[f(y+h_H v|x+h_K u)$$
$$- f(y|x+h_K u)]dudv$$
$$= \frac{h_H^2}{2}\ell(x)f^{(0,2)}(y|x)\int v^2 H(v)dv + o(h_H^2),$$
$$\frac{\theta}{2nh_K}V_{n3}(y|x) = \frac{\theta}{2nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{(X_i-x)^2 f^{(2,0)}(y|X_i^*)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)$$
$$= \frac{h_K^2}{2}\ell(x)f^{(2,0)}(y|x)\int u^2 K(u)du + o(h_K^2).$$

Similarly

$$\frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)]w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)$$
$$= \frac{\theta}{h_K^2}\mathbf{E}\Big\{\frac{(X-x)K^2(\frac{X-x}{h_K})}{G(Y_i)}\Big[\frac{1}{h_H}H\Big(\frac{Y_i-y}{h_H}\Big)-f(y|X_i)\Big]\Big\}$$
$$= \int\int uK^2(u)H(v)\ell(x+h_K u)[f(y+h_H v|x+h_K u)-f(y|x+h_K u)]dudv$$
$$= \frac{h_K h_H^2}{2}l'(x)f^{(0,2)}(y|x)\int u^2K^2(u)du\int v^2 H(v)dv + o(h_K h_H^2),$$
$$\times\frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)]w_i^2(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)$$
$$= \frac{\theta}{h_K^3}\mathbf{E}\Big\{\frac{(X-x)^2K^3(\frac{X-x}{h_K})}{G(Y_i)}\Big[\frac{1}{h_H}H\Big(\frac{Y_i-y}{h_H}\Big)-f(y|X_i)\Big]\Big\}$$
$$= \int\int u^2K^2(u)H(v)\ell(x+h_K u)[f(y+h_H v|x+h_K u)-f(y|x+h_K u)]dudv$$

$$= \frac{h_H^2}{2}\ell(x)f^{(0,2)}(y|x)\int u^2K^2(u)du\int v^2H(v)dv + o(h_H^2),$$

$$\frac{\theta}{nh_K}\sum_{i=1}^{n}\mathbf{E}\left(\frac{(X_i-x)^2f^{(2,0)}(y|X_i^*)w_i(x)}{G(Y_i)}K\left(\frac{X_i-x}{h_K}\right)\right) = O(h_K^2).$$

Therefore, from Lemma 1 we obtain that

$$\frac{\theta}{nh_K}U_{n5}(y|x) = \frac{h_K^2h_H^2}{2}\frac{(l'(x))^2f^{(0,2)}(y|x)\Delta_{21}\Lambda_{21}}{\ell(x)\mu(x)} + O_{\mathbf{P}}\left(\frac{h_Kh_H^2}{\sqrt{nh_K}}\right) + o_{\mathbf{P}}(h_K^2h_H^2),$$

$$\frac{\theta}{nh_K}U_{n7}(y|x) = \frac{h_K^2h_H^2}{2}\frac{(l'(x))^2f^{(0,2)}(y|x)\Delta_{21}^2\Lambda_{21}}{\ell(x)\mu^2(x)\Delta_{22}^2} + O_{\mathbf{P}}\left(\frac{h_Kh_H^2}{\sqrt{nh_K}}\right) + o_{\mathbf{P}}(h_K^2h_H^2) + O_{\mathbf{P}}\left(\frac{h_H^2}{nh_K}\right),$$

$$\frac{\theta}{2nh_K}V_{n5}(y|x) = O_{\mathbf{P}}\left(h_K^3 + \frac{h_K^2}{\sqrt{nh_K}}\right).$$

Step 4. Note that $\mathrm{Var}\left(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)](X_i-x)}{G(Y_i)}K^2(\frac{X_i-x}{h_K})\right) = O\left(\frac{h_K^3}{h_H}\right)$ and for $i<j$,

$$\mathrm{Cov}\left(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)](X_i-x)}{G(Y_i)}K^2\left(\frac{X_i-x}{h_K}\right), \frac{[H_{h_H}(Y_j-y)-f(y|X_j)](X_j-x)}{G(Y_j)}K^2\left(\frac{X_j-x}{h_K}\right)\right)$$
$$\leq C\min\left\{h_k^4, [\alpha(j-i)]^{1-1/(10\gamma)}(h_Kh_H)^{1/(10\gamma)}(h_K/h_H)^2\right\}.$$

Then, similarly to the argument as in (21) we have

$$\mathrm{Var}\left(\frac{\theta}{nh_K}\sum_{i=1}^{n}\frac{[H_{h_H}(Y_i-y)-f(y|X_i)]w_i(x)}{G(Y_i)}K\left(\frac{X_i-x}{h_K}\right)\right)$$
$$= \frac{\theta^2}{n^2h_K^4}\left[\sum_{i=1}^{n}\mathrm{Var}\left(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)](X_i-x)}{G(Y_i)}K^2\left(\frac{X_i-x}{h_K}\right)\right)\right.$$
$$+\sum_{i\neq j}\mathrm{Cov}\left(\frac{[H_{h_H}(Y_i-y)-f(y|X_i)](X_i-x)}{G(Y_i)}K^2\left(\frac{X_i-x}{h_K}\right),\right.$$
$$\left.\left.\frac{[H_{h_H}(Y_j-y)-f(y|X_j)](X_j-x)}{G(Y_j)}K^2\left(\frac{X_j-x}{h_K}\right)\right)\right] = O\left(\frac{1}{nh_Kh_H}\right).$$

Similarly

$$\mathrm{Var}\left(\frac{\theta}{nh_K}\sum_{i=1}^{n}\frac{[H_{h_H}(Y_i-y)-f(y|X_i)]w_i^2(x)}{G(Y_i)}K\left(\frac{X_i-x}{h_K}\right)\right)=O\left(\frac{1}{nh_Kh_H}\right),$$
$$\mathrm{Var}\left(\frac{\theta}{2nh_K}\sum_{i=1}^{n}\frac{(X_i-x)^2f^{(2,0)}(y|X_i^*)}{G(Y_i)}K\left(\frac{X_i-x}{h_K}\right)\right)=O\left(\frac{h_K^3}{n}\right),$$
$$\mathrm{Var}\left(\frac{\theta}{2nh_K}\sum_{i=1}^{n}\frac{(X_i-x)^2f^{(2,0)}(y|X_i^*)w_i(x)}{G(Y_i)}K\left(\frac{X_i-x}{h_K}\right)\right)=O\left(\frac{h_K^3}{n}\right).$$

Therefore, from Lemma 1 we have

$$\frac{\theta}{nh_K}U_{n4}(y|x)=O_{\mathbf{P}}\left(\frac{1}{nh_K\sqrt{h_H}}+\frac{h_K}{\sqrt{nh_Kh_H}}\right),\ \frac{\theta}{2nh_K}V_{n2}(y|x)=O_{\mathbf{P}}\left(\sqrt{\frac{h_K^3}{n}}\right),$$
$$\frac{\theta}{nh_K}U_{n6}(y|x)=O_{\mathbf{P}}\left(\frac{1}{nh_K\sqrt{nh_Kh_H}}+\frac{h_K^2}{\sqrt{nh_Kh_H}}\right),$$
$$\frac{\theta}{2nh_K}V_{n4}(y|x)=O_{\mathbf{P}}\left(\frac{h_K}{n}+h_K^2\sqrt{\frac{h_K}{n}}\right).$$

□

Proof of Proposition 2. Let the definition of $\ell_n(x)$ be same as that in the proof of Proposition 1, and we define $\mathscr{U}_{nj}(\cdot)$ and $\mathscr{V}_{nl}(\cdot)$ as in $U_{nj}(\cdot)$ and $V_{nl}(\cdot)$ by replacing $H_{h_H}(Y_i-y)$ and $f(\cdot)$ with $\mathbb{H}(\frac{Y_i-y}{h_H})$ and $S(\cdot)$, respectively. Thus we write

$$\begin{aligned}
&\widehat{S}_n(y|x)-S(y|x)\\
&\quad=\frac{1}{\ell_n(x)}\cdot\frac{\theta}{nh_K}\big\{\mathscr{U}_{n1}(y|x)+\mathscr{U}_{n2}(y|x)+\mathscr{U}_{n3}(y|x)-\mathscr{U}_{n4}(y|x)-\mathscr{U}_{n5}(y|x)\\
&\qquad+\mathscr{U}n6(y|x)+\mathscr{U}_{n7}(y|x)-\mathscr{U}_{n8}(y|x)\big\}+\frac{1}{\ell_n(x)}\cdot\frac{\theta}{2nh_K}\big\{\mathscr{V}_{n1}(y|x)+\mathscr{V}_{n2}(y|x)\\
&\qquad+\mathscr{V}_{n3}(y|x)-\mathscr{V}_{n4}(y|x)-\mathscr{V}_{n5}(y|x)+\mathscr{V}_{n6}(y|x)\big\}.
\end{aligned}$$

Following the line in the proof of Proposition 1, it is easy to verify the following facts

$$\frac{\theta}{nh_K}\mathscr{U}_{n1}(y|x)=O_{\mathbf{P}}(n^{-1/2}),\ \frac{\theta}{nh_K}\mathscr{U}_{n8}(y|x)=O_{\mathbf{P}}((nh_K)^{-3/2}+h_K^3),$$
$$\frac{\theta}{nh_K}\mathscr{V}_{n1}(y|x)=O_{\mathbf{P}}\left(\frac{h_K^2}{\sqrt{n}}\right),\ \frac{\theta}{2nh_K}\mathscr{V}_{n2}(y|x)=O_{\mathbf{P}}\left(\sqrt{\frac{h_K^3}{n}}\right),$$
$$\frac{\theta}{2nh_K}\mathscr{V}_{n4}(y|x)=O_{\mathbf{P}}\left(\frac{h_K}{n}+h_K^2\sqrt{\frac{h_K}{n}}\right),$$

$\frac{\theta}{2nh_K}\mathscr{V}_{n5}(y|x) = O_{\mathbf{P}}\left(h_K^3 + \frac{h_K^2}{\sqrt{nh_K}}\right)$ and $\frac{\theta}{2nh_K}\mathscr{V}_{n6}(y|x) = O_{\mathbf{P}}\left(\frac{h_K}{n} + h_K^4\right)$.

Next we need only to prove that

$$\frac{\theta}{\sqrt{nh_K\Delta_n(y|x)}}\mathscr{U}_{n2}(y|x) \xrightarrow{\mathscr{D}} N\big(0,1\big),$$
$$\frac{\theta}{nh_K}\mathscr{U}_{n3}(y|x) = \frac{h_H^2}{2}\ell(x)S^{(0,2)}(y|x)\int v^2H(v)dv + o(h_H^2),$$
$$\frac{\theta}{2nh_K}\mathscr{V}_{n3}(y|x) = \frac{h_K^2}{2}\ell(x)S^{(2,0)}(y|x)\int u^2K(u)du + o(h_K^2),$$
$$\frac{\theta}{nh_K}\mathscr{U}_{n4}(y|x) = O_{\mathbf{P}}\Big(\frac{1}{nh_K} + \frac{h_K}{\sqrt{nh_K}}\Big),\ \frac{\theta}{nh_K}\mathscr{U}_{n5}(y|x) = O_{\mathbf{P}}\Big(\frac{h_Kh_H^2}{\sqrt{nh_K}} + h_K^2h_H^2\Big),$$
$$\frac{\theta}{nh_K}\mathscr{U}_{n6}(y|x) = O_{\mathbf{P}}\Big(\frac{1}{nh_K\sqrt{nh_K}} + \frac{h_K^2}{\sqrt{nh_K}}\Big),$$
$$\frac{\theta}{nh_K}\mathscr{U}_{n7}(y|x) = O_{\mathbf{P}}\Big(h_K^2h_H^2 + \frac{h_H^2}{nh_K}\Big).$$

Step 5. We verify $\frac{\theta}{\sqrt{nh_K\Delta_n(y|x)}}\mathscr{U}_{n2}(y|x) \xrightarrow{\mathscr{D}} N\big(0,1\big)$. From (A5), it follows that there exists a sequence of positive integers $\zeta_n \to \infty$ such that $\zeta_n u_n = o((nh_K)^{1/2})$ and $\zeta_n(n(h_K)^{-1})^{1/2}\alpha(u_n) \to 0$. Let $\pi_n = [\frac{n}{u_n+v_n}]$ and $v_n = [(nh_K)^{1/2}/\zeta_n]$. Then

$$u_n/v_n \to 0,\ \pi_n\alpha(u_n) \to 0,\ \pi_n u_n/n \to 0,\ v_n/n \to 0,\ v_n/(nh_K)^{1/2} \to 0.$$

Define $\xi_{mn} = \sum_{i=k_m}^{k_m+v_n-1}\varpi_i$, $\xi'_{mn} = \sum_{i=l_m}^{l_m+u_n-1}\varpi_i$, $\xi''_{\pi_n n} = \sum_{i=\pi_n(u_n+v_n)+1}^{n}\varpi_i$ with

$$\varpi_i = \theta\sqrt{\frac{1}{h_K\Delta_n(y|x)}}\Big[\frac{\mathbb{H}(\frac{Y_i-y}{h_H}) - S(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)$$
$$- \mathbf{E}\Big(\frac{\mathbb{H}(\frac{Y_i-y}{h_H}) - S(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\Big],$$

where $k_m = (m-1)(u_n+v_n)+1$, $l_m = (m-1)(u_n+v_n)+v_n+1$, $m = 1,\ldots,\pi_n$. Then

$$\sqrt{nh_K}I_{n2}(y|x) = \frac{1}{\sqrt{n}}\Big\{\sum_{m=1}^{\pi_n}\xi_{mn} + \sum_{m=1}^{\pi_n}\xi'_{mn} + \xi''_{\pi_n n}\Big\} := \frac{1}{\sqrt{n}}\big\{T'_n + T''_n + T'''_n\big\}.$$

Since $0 < \varepsilon(y|x) \le \Delta_n(y|x) \le C$, from (A1) and (A2) we have

$$\mathrm{Var}(\varpi_i) = \frac{\theta^2}{h_K\Delta_n(y|x)}\Big\{\mathbf{E}\Big(\frac{K^2(\frac{X_i-x}{h_K})}{G^2(Y_i)}\Big[\mathbb{H}\Big(\frac{Y_i-y}{h_H}\Big) - S(y|X_i)\Big]^2$$

$$
\begin{aligned}
&\quad - \Big[\mathbf{E}\Big(\frac{\mathbb{H}(\frac{Y_i-y}{h_H}) - S(y|X_i)}{G(Y_i)} K\Big(\frac{X_i - x}{h_K}\Big)\Big)\Big]^2\Big\} \\
&= \frac{\theta}{\Delta_n(y|x)} \int_{\mathbb{R}^2} \frac{K^2(u)}{G(t)} \Big[\mathbb{H}\Big(\frac{t-y}{h_H}\Big) - S(y|x + h_K u)\Big]^2 f(x + h_K u, t) du dt \\
&\quad + O(h_K) \\
&= \frac{\theta \ell(x)}{\Delta_n(y|x)} \mathbb{E}\Big(\Big[\mathbb{H}\Big(\frac{Y-y}{h_H}\Big) - S(y|x)\Big]^2 G^{-1}(Y)\Big| X = x\Big) \int_{\mathbb{R}} K^2(u) du \\
&\quad + o(1) \to 1.
\end{aligned}
$$

Then, following similar line as *Step 1* in the proof of Proposition 1, one can prove that $n^{-1}\mathbf{E}(T_n'')^2 \to 0, n^{-1}\mathbf{E}(T_n''')^2 \to 0$, $A_n(\varepsilon) = \frac{1}{n}\sum_{m=1}^{\pi_n} \mathbf{E}\xi_{mn}^2 I(|\xi_{mn}| > \varepsilon\sqrt{n}) \to 0 \ \forall \varepsilon > 0, n^{-1}\mathrm{Var}(T_n') \to 1$ and

$$
\Big|\mathbf{E}\exp\Big(it\sum_{m=1}^{\pi_n} n^{-1/2}\xi_{mn}\Big) - \prod_{m=1}^{\pi_n}\mathbf{E}\exp\big(itn^{-1/2}\xi_{mn}\big)\Big| \to 0.
$$

Step 6. From (A1), (A2) and (A6) we have

$$
\begin{aligned}
\frac{\theta}{nh_K}\mathscr{U}_{n3}(y|x) &= \frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{\mathbb{H}(\frac{Y_i-y}{h_H}) - S(y|X_i)}{G(Y_i)} K\Big(\frac{X_i - x}{h_K}\Big)\Big) \\
&= \int K(u)\ell(x + h_K u) du \int \Big[\mathbb{H}\Big(\frac{t-y}{h_H}\Big) - S(y|x + h_K u)\Big] \\
&\quad \times f(t|x + h_K u) dt \\
&= \int K(u)\ell(x + h_K u)\Big(\int H(v)[S(y + h_H v|x + h_K u) \\
&\quad - S(y|x + h_K u)] dv\Big) du \\
&= \frac{h_H^2}{2}\ell(x) S^{(0,2)}(y|x) \int v^2 H(v) dv + o(h_H^2), \\
\frac{\theta}{2nh_K}\mathscr{V}_{n3}(y|x) &= \frac{\theta}{2nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{(X_i - x)^2 S^{(2,0)}(y|X_i^*)}{G(Y_i)} K\Big(\frac{X_i - x}{h_K}\Big)\Big) \\
&= \frac{h_K^2}{2}\ell(x) S^{(2,0)}(y|x) \int u^2 K(u) du + o(h_K^2).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
&\frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{[\mathbb{H}(\frac{Y_i-y}{h_H}) - S(y|X_i)] w_i(x)}{G(Y_i)} K\Big(\frac{X_i - x}{h_K}\Big)\Big) \\
&\quad = \frac{\theta}{h_K^2}\mathbf{E}\Big\{\frac{(X - x)K^2(\frac{X-x}{h_K})}{G(Y_i)}\Big[\mathbb{H}\Big(\frac{Y_i - y}{h_H}\Big) - S(y|X_i)\Big]\Big\}
\end{aligned}
$$

$$= \int uK^2(u)\ell(x+h_Ku)\Big(\int H(v)[S(y+h_Hv|x+h_Ku)-S(y|x+h_Ku)]dv\Big)du$$

$$= \frac{h_Kh_H^2}{2}l'(x)S^{(0,2)}(y|x)\int u^2K^2(u)du\int v^2H(v)dv+o(h_Kh_H^2),$$

$$\frac{\theta}{nh_K}\sum_{i=1}^n \mathbf{E}\Big(\frac{[\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)]w_i^2(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)$$

$$= \frac{\theta}{h_K^3}\mathbf{E}\Big\{\frac{(X-x)^2K^3(\frac{X-x}{h_K})}{G(Y_i)}\Big[\mathbb{H}\Big(\frac{Y_i-y}{h_H}\Big)-S(y|X_i)\Big]\Big\}$$

$$= \int u^2K^2(u)\ell(x+h_Ku)\Big(\int H(v)[S(y+h_Hv|x+h_Ku)-S(y|x+h_Ku)]dv\Big)du$$

$$= \frac{h_H^2}{2}\ell(x)S^{(0,2)}(y|x)\int u^2K^2(u)du\int v^2H(v)dv+o(h_H^2).$$

Therefore, from Lemma 1 we obtain that

$$\frac{\theta}{nh_K}\mathscr{U}_{n5}(y|x)=O_{\mathbf{P}}\left(\frac{h_Kh_H^2}{\sqrt{nh_K}}+h_K^2h_H^2\right),\ \frac{\theta}{nh_K}\mathscr{U}_{n7}(y|x)=O_{\mathbf{P}}\left(h_K^2h_H^2+\frac{h_H^2}{nh_K}\right).$$

Step 7. Following the argument as for $\mathrm{Var}(\varpi_i)$ in *Step 5*, for positive integers l_1 and l_2 we have

$$\frac{\theta}{h_K^{1+2l_1}}\mathrm{Var}\left(\frac{[\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)](X_i-x)^{l_1}}{G(Y_i)}K^{l_2}\left(\frac{X_i-x}{h_K}\right)\right)\le C, \tag{22}$$

and applying Lemma 3, for $i<j$ we have

$$\mathrm{Cov}\left(\frac{[\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)](X_i-x)^{l_1}K^{l_2}(\frac{X_i-x}{h_K})}{G(Y_i)},\right.$$
$$\left.\frac{[\mathbb{H}(\frac{Y_j-y}{h_H})-S(y|X_j)](X_j-x)^{l_1}K^{l_2}(\frac{X_j-x}{h_K})}{G(Y_j)}\right)$$
$$\le C\min\left\{h_K^{2l_1+2},[\alpha(j-i)]^{1-1/(10\gamma)}h_K^{2l_1+1/(10\gamma)}\right\}. \tag{23}$$

Then

$$\mathrm{Var}\Big(\frac{\theta}{nh_K}\sum_{i=1}^n\frac{[\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)]w_i(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)$$
$$=\frac{\theta^2}{n^2h_K^4}\Big[\sum_{i=1}^n\mathrm{Var}\Big(\frac{[\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)](X_i-x)}{G(Y_i)}K^2\Big(\frac{X_i-x}{h_K}\Big)\Big)$$

$$+\sum_{i\neq j}\mathrm{Cov}\Big(\frac{[\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)](X_i-x)}{G(Y_i)}K^2\Big(\frac{X_i-x}{h_K}\Big),$$
$$\frac{[\mathbb{H}(\frac{Y_j-y}{h_H})-S(y|X_j)](X_j-x)}{G(Y_j)}K^2\Big(\frac{X_j-x}{h_K}\Big)\Big)\Big]=O\Big(\frac{1}{nh_K}\Big),$$

which gives $\mathrm{Var}\Big(\frac{\theta}{nh_K}\sum_{i=1}^{n}\frac{[\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)]w_i^2(x)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)=O\Big(\frac{1}{nh_K}\Big)$.

Therefore, from Lemma 1 we have

$$\frac{\theta}{nh_K}\mathscr{U}_{n4}(y|x)=O_{\mathbf{P}}\Big(\frac{1}{nh_K}+\frac{h_K}{\sqrt{nh_K}}\Big),$$
$$\frac{\theta}{nh_K}\mathscr{U}_{n6}(y|x)=O_{\mathbf{P}}\Big(\frac{1}{nh_K\sqrt{nh_K}}+\frac{h_K^2}{\sqrt{nh_K}}\Big).$$

□

Proof of Proposition 3. From the proof of Proposition 2 we write

$$\begin{aligned}&\widehat{S}_n(y|x)-S(y|x)\\&=\frac{1}{\ell_n(x)}\cdot\frac{\theta}{nh_K}\big\{\mathscr{U}_{n1}(y|x)+\mathscr{U}_{n2}(y|x)+\mathscr{U}_{n3}(y|x)-\mathscr{U}_{n4}(y|x)-\mathscr{U}_{n5}(y|x)\\&\quad+\mathscr{U}_{n6}(y|x)+\mathscr{U}_{n7}(y|x)-\mathscr{U}_{n8}(y|x)\big\}+\frac{1}{\ell_n(x)}\cdot\frac{\theta}{2nh_K}\big\{\mathscr{V}_{n1}(y|x)+\mathscr{V}_{n2}(y|x)\\&\quad+\mathscr{V}_{n3}(y|x)-\mathscr{V}_{n4}(y|x)-\mathscr{V}_{n5}(y|x)+\mathscr{V}_{n6}(y|x)\big\}.\end{aligned}$$

The proof of Proposition 2 shows that

$$\frac{\theta}{nh_K}\mathscr{U}_{nj}(y|x)=o_{\mathbf{P}}(1)\ \text{ for } j=1,3,,4,\ldots,8;\ \frac{\theta}{nh_K}\mathscr{V}_{nl}(y|x)=o_{\mathbf{P}}(1)\quad\text{for } l=1,2,\ldots,6.$$

Therefore, it suffices to prove that $\frac{\theta}{nh_K}\mathscr{U}_{n2}(y|x)\to 0$ in Probability.

From (22) and (23) it follows that

$$\begin{aligned}&\mathrm{Var}\Big(\frac{\theta}{nh_K}\mathscr{U}_{n2}(y|x)\Big)\\&=\Big(\frac{\theta}{nh_K^2}\Big)^2\Big\{\sum_{i=1}^{n}\mathrm{Var}\Big(\frac{\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big)\end{aligned}$$

$$\sum_{i\neq j}\text{Cov}\Big(\frac{\mathbb{H}(\frac{Y_i-y}{h_H})-S(y|X_i)}{G(Y_i)}K\Big(\frac{X_i-x}{h_K}\Big)\Big),\frac{\mathbb{H}(\frac{Y_j-y}{h_H})-S(y|X_j)}{G(Y_j)}K\Big(\frac{X_j-x}{h_K}\Big)\Big)\Big\}$$
$$=O\Big(\frac{1}{nh_K}\Big)+2\Big(\frac{\theta}{nh_K^2}\Big)^2\sum_{1\le i<j\le n}\min\Big\{h_K^2,[\alpha(j-i)]^{1-1/(10\gamma)}h_K^{1/(10\gamma)}\Big\}\to 0. \tag{24}$$

Therefore $\frac{\theta}{nh_K}\mathscr{U}_{n2}(y|x)\to 0$ in Probability. □

Proof of Theorem 1. It is easy to see that

$$\widehat{\lambda}_n(y|x)-\lambda(y|x)=\widehat{S}_n^{-1}(y|x)\big\{\widehat{f}_n(y|x)-f(y|x)-\lambda(y|x)[\widehat{S}_n(y|x)-S(y|x)]\big\}.$$

From the proof of Propositions 2 and 3 it follows that

$$\widehat{S}_n(y|x)-S(y|x)=l^{-1}(x)\Big[\frac{h_K^2}{2}\Delta_{21}S^{(2,0)}(y|x)+\frac{h_H^2}{2}\Lambda_{21}S^{(0,2)}(y|x)$$
$$+\frac{\theta}{nh_K}\mathscr{U}_{n2}(y|x)\Big]+o_{\mathbf{P}}(h_K^2+h_H^2)+O_{\mathbf{P}}\Big(\frac{1}{\sqrt{n}}+\frac{1}{nh_K}+\sqrt{\frac{h_K}{n}}\Big).$$

(24) in the proof of Proposition 3 gives that $\frac{\tau}{nh_K}\mathscr{U}_{n2}(y|x)=O_{\mathbf{P}}\big((nh_K)^{-1/2}\big)$. Therefore

$$\widehat{S}_n(y|x)-S(y|x)=l^{-1}(x)\Big[\frac{h_K^2}{2}\Delta_{21}S^{(2,0)}(y|x)+\frac{h_H^2}{2}\Lambda_{21}S^{(0,2)}(y|x)\Big]$$
$$+o_{\mathbf{P}}(h_K^2+h_H^2)+O_{\mathbf{P}}\big((nh_K)^{-1/2}\big).$$

Proposition 1 shows that

$$\sqrt{nh_Kh_H}\left\{\widehat{f}_n(y|x)-f(y|x)-\frac{h_K^2}{2}\Delta_{21}f^{(2,0)}(y|x)-\frac{h_H^2}{2}\Lambda_{21}f^{(0,2)}(y|x)\right.$$
$$\left.+o_{\mathbf{P}}(h_K^2+h_H^2)+O_{\mathbf{P}}\left(\frac{1}{\sqrt{n}}+\frac{1}{nh_K\sqrt{h_H}}+\sqrt{\frac{h_K}{nh_H}}\right)\right\}\xrightarrow{\mathscr{D}}N\left(0,\theta\sigma^2(y|x)\right).$$

Therefore

$$\sqrt{nh_Kh_H}\Big\{\widehat{\lambda}_n(y|x)-\lambda(y|x)-Asymp.bias+o_{\mathbf{P}}(h_K^2+h_H^2)+O_{\mathbf{P}}\big((nh_K)^{-1/2}$$
$$+\ (h_K/(nh_H))^{1/2}\big)\Big\}\xrightarrow{\mathscr{D}}N\Big(0,\frac{\theta\sigma^2(y|x)}{S^2(y|x)}\Big).$$

□

7 Appendix

Lemma 2 (Fan and Yao (2003), *Proposition 2.6, p. 72) Let* $V_1, \ldots, V_m$ *be* α*-mixing and complex-valued random variables measurable with respect to the* σ*-algebra* $\mathscr{F}_{i_1}^{j_1}, \ldots, \mathscr{F}_{i_m}^{j_m}$*, respectively, with* $1 \le i_1 < j_1 < \cdots < j_m \le n$, $i_{l+1} - j_l \ge w \ge 1$ *and* $P(|V_j| \le 1) = 1$ *for* $l, j = 1, 2, \ldots, m$*. Then* $|E(\prod_{j=1}^m V_j) - \prod_{j=1}^m EV_j| \le 16(m-1)\alpha(w)$*, where* $\mathscr{F}_a^b = \sigma\{V_i, a \le i \le b\}$ *and* $\alpha(w)$ *is the mixing coefficient.*

Lemma 3 (Hall and Heyde (1980), *Corollary A.2, p. 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*

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

Lemma 4 (Liang et al. (2011), *Lemma 5.4) Suppose that* $\alpha(k) = O(k^{-\gamma})$ *for some* $\gamma > 3$*, and that (A0) holds. Then* $\sup_{y\ge a_F} |G_n(y) - G(y)| = O_{\mathbf{P}}(n^{-1/2})$.

References

Bagkavos, D. (2011a). Local linear hazard rate estimation and bandwidth selection. *Journal of Statistical Computation and Simulation*, *81*, 1517–1531.

Bagkavos, D. (2011b). Local linear hazard rate estimation and bandwidth selection. *Annals of the Institute of Statistical Mathematics*, *63*, 1019–1046.

Berlinet, A., Elamine, A., Mas, A. (2011). Local linear regression for functional data. *Annals of the Institute of Statistical Mathematics*, *63*, 1047–1075.

Bradley, R. C. (2007). *Introduction to strong mixing conditions* (Vol. I–III). Utah: Kendrick Press.

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

Collomb, G., Hassani, S., Sarda, P., Vieu, P. (1985). Convergence uniforme d'estimateurs de la fonction de hasard pour des observations dépendantes: méthodes du noyau et des k-points les plus proches. *Comptes Rendus de l'Académie des Sciences. Series I, Mathematics*, *301*, 653–656.

Dedecker, J., Doukhan, P., Lang, G., León, R., Louhichi, S., Prieur, C. (2007). *Weak dependence: with examples and applications.* Lecture notes in statistics, 190. New York: Springer.

Fan, J., Yao, Q. (2003). *Nonlinear time series*. New York: Springer.

Fan, J., Yao, Q., Tong, H. (1996). Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems. *Biometrika*, *83*, 189–206.

Hall, P., Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic Press.

Hall, P., Presnell, B. (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society, Series B*, *61*, 143–158.

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

He, S., Yang, G. (1998). Estimation of the truncation probability in the random truncation model. *The Annals of Statistics*, *26*, 1011–1027.

Kim, C., Bae, W., Choi, H., Park, B. U. (2005). Non-parametric hazard function estimation using the Kaplan–Meier estimator. *Journal of Nonparametric Statistics*, *17*, 937–948.

Kim, C., Oh, M., Yang, S. J., Choi, H. (2010). A local linear estimation of conditional hazard function in censored data. *Journal of the Korean Statistical Society*, *39*, 347–355.

Lecoutre, J. P., Ould Saïd, E. (1992). Estimation de la densité conditionnelle et de la fonction de hasard conditionnelle pour un processus fortement mélangeant avec censure. *Comptes Rendus de l'Académie des Sciences. Series I, Mathematics*, *314*, 295–300.

Lecoutre, J. P., Ould Saïd, E. (1995). Hazard rate estimation for strong-mixing and censored processes. *Journal of Nonparametric Statistics*, *5*, 83–89.

Liang, H. Y. (2012). Weighted nonparametric regression estimation with truncated and dependent data. *Journal of Nonparametric Statistics*, *24*(4), 1051–1073.

Liang, H. Y., Baek, J. I. (2016). Asymptotic normality of conditional density estimation with left-truncated and dependent data. *Statistical Papers*, *57*, 1–20.

Liang, H. Y., Fan, G. L. (2009). Berry–Esseen type bounds of estimators in a semiparametric model with linear process errors. *Journal of Multivariate Analysis*, *100*, 1–15.

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*, 653–677.

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

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

Ould Saïd, E., Tatachak, A. (2007). Asymptotic properties of the kernel estimator of the conditional mode for the left truncated model. *Comptes Rendus de l'Académie des Sciences. Series I, Mathematics*, *344*, 651–656.

Ould Saïd, E., Tatachak, A. (2009). Strong consistency rate for the kernel mode under strong mixing hypothesis and left truncation. *Communications in Statistics: Theory and Methods*, *38*, 1154–1169.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, *18*, 90–120.

Quintela-del-Río, A. (2008). Hazard function given a functional variable: non-parametric estimation under strong mixing conditions. *Journal of Nonparametric Statistics*, *20*, 413–430.

Sarda, P., Vieu, P. (1991). Smoothing parameter selection in hazard estimation. *Statistics and Probability Letters*, *11*, 429–434.

Spierdijk, L. (2008). Nonparametric conditional hazard rate estimation: A local linear approach. *Computational Statistics and Data Analysis*, *52*, 2419–2434.

Stute, W. (1993). Almost sure representation of the product-limit estimator for truncated data. *The Annals of Statistics*, *21*, 146–156.

Stute, W., Wang, J. L. (2008). The central limit theorem under random truncation. *Bernoulli*, *14*, 604–622.

Van Keilegom, I., Veraverbeke, N. (2001). Hazard rate estimation in nonparametric regression with censored data. *Annals of the Institute of Statistical Mathematics*, *53*, 730–745.

Woodroofe, M. (1985). Estimating a distribution function with truncated data. *The Annals of Statistics*, *13*, 163–177.

Zhou, W. (2011). A weighted quantile regression for randomly truncated data. *Computational Statistics and Data Analysis*, *55*, 554–566.