



Journal of Nonparametric Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/gnst20

Nonparametric relative error estimation of the regression function for left truncated and right censored time series data

N. Bayarassou, F. Hamrani & E. Ould Saïd

To cite this article: N. Bayarassou, F. Hamrani & E. Ould Saïd (2023): Nonparametric relative error estimation of the regression function for left truncated and right censored time series data, Journal of Nonparametric Statistics, DOI: 10.1080/10485252.2023.2241572

To link to this article: https://doi.org/10.1080/10485252.2023.2241572



Published online: 02 Sep 2023.



Submit your article to this journal 🕝





View related articles



View Crossmark data 🗹



Check for updates

Nonparametric relative error estimation of the regression function for left truncated and right censored time series data

N. Bayarassou^a, F. Hamrani^a and E. Ould Saïd^b

^aLab MSTD, Faculté de Mathématiques, USTHB, Alger, Algérie; ^bLMPA, Université du Littoral Côte d'Opale, Calais, France

ABSTRACT

The paper introduces a nonparametric estimator for the regression function of left truncated and right censored data, achieved through minimising the mean squared relative error. Under α -mixing condition, strong uniform convergence of the estimator is established with a rate over a compact set. An extensive simulation study is conducted to assess the estimator's performance, comparing its efficiency to that of the classical regression estimator for finite samples across various scenarios. Moreover, a real world application is presented to demonstrate the practical utility of the proposed estimator. **ARTICLE HISTORY**

Received 8 December 2022 Accepted 20 July 2023

KEYWORDS

Kernel estimate; relative error regression; strong mixing condition; strong uniform convergence; truncated-censored data

1. Introduction

Let *Y* be a real random variable (rv) of interest and **X** a random vector of covariates taking its values in \mathbb{R}^d . The ordinary way to study the relationship between **X** and *Y* is based on the following regression model

$$Y = m(\mathbf{X}) + \epsilon,$$

where $m(\cdot)$ is an unknown regression function and ϵ is a random error variable. Usually, the regression function $m(\cdot)$ is obtained by minimising the mean squared error (MSE):

$$\mathbb{E}[(Y - m(\mathbf{X}))^2 \,|\, \mathbf{X}].$$

However, this kind of loss function is very sensitive to outliers. To avoid this problem, in this work, we use an alternative loss function called the mean squared relative error (MSRE):

$$\mathbb{E}\left[\left(\frac{Y-m(\mathbf{X})}{Y}\right)^2 \mid \mathbf{X}\right], \quad \text{for } Y > 0.$$
(1)

Park and Stefanski (1998) showed that the solution of the minimisation problem of (1) is explicitly expressed by

$$m(X) = \frac{\mathbb{E}[Y^{-1} \mid \mathbf{X}]}{\mathbb{E}[Y^{-2} \mid \mathbf{X}]},$$
(2)

CONTACT E. Ould Saïd 🖾 elias.ould-said@univ-littoral.fr 🖃 LMPA, ULCO, IUT de Calais. 19, rue Louis David. BP 699, 62228 Calais, France

^{© 2023} American Statistical Association and Taylor & Francis

2 🛞 N. BAYARASSOU ET AL.

provided that the first two conditional inverse moments of Y given **X** are finite almost surly. They also noted that this MSRE predictor is always smaller than the MSE predictor. In our study, we focus on the nonparametric approach. In this context, we recall that Jones, Park, Shin, Vines, and Jeong (2008) gave asymptotic results for bias and variance terms of an estimator minimising the MSRE by considering both estimation methods: the kernel method and the local linear approach. Attouch, Laksaci, and Messabihi (2017) established the almost complete consistency and the asymptotic normality of a kernel estimator of the relative regression function for spatial data. In the infinite dimensional Demongeot, Hamie, Laksaci, and Rachdi (2016), proved the strong and uniform consistency of a kernel estimator and its asymptotic normality.

In survival practical applications where the lifetime is the variable of interest, it is very common that the generation of data is subject to mechanisms of loss of information such as censoring and truncation. These two models are completely different from each other in the sense that a censored subject provides a partial information, while no information is available to the practitioner when a subject is truncated. The most popular types are right censoring and left truncation.

For censored data, Bouhadjera, Ould Saïd, and Remita (2019) and Khardani (2020) introduced a kernel estimator of the regression function minimising the sum of squared relative errors and they established its uniform convergence and asymptotic normality for right censored and twice censored data, respectively. Bouhadjera, Ould Saïd, and Remita (2022) proposed a local linear regression estimator for right censored data and obtained its uniform almost sure consistency with rate over a compact set.

For truncated data, in the case where the explanatory variable is of functional type, we mention both works of Altendji, Demongeot, Laksaci, and Rachdi (2018) who established the almost sure consistency and the asymptotic normality of an estimator of the relative regression operator for left truncated data, and Bouabsa (2021) who obtained a uniform consistency with convergence rate of a k Nearest Neighbors relative regression estimator for left truncated data.

In this contribution, we are interested in survival data that are subject to both left truncation and right censoring (LTRC). This kind of data often arises in medical studies, where a subject is referred to as left truncated when it is not included in the study because its lifetime origin precedes the starting time of the study, and a subject is called right censored when it is into the study but its lifetime may not be completely observed due to different causes: death for a reason unrelated to the study, leaving the study or end of the study period (for more detailed examples see Chen and Shen 2018). For this type of data, Molanes-López and Cao (2008) defined a kernel estimator of the relative density. They obtained its bias, variance and limit distribution. Benseradj and Guessoum (2022) proposed an M-estimator of regression function and established its strong uniform consistency rate under α -mixing dependence. Bey, Guessoum, and Tatachak (2022) presented a kernel estimator of the regression function and stated its almost sure uniform convergence rate under association assumptions.

Most studies addressing LTRC model consider the independent case. However, it is sometimes interesting to consider dependent samples in order to respond to practical situations where the data is not independent and identically distributed. Various types of dependency have been defined in the literature, such as α -mixing, β -mixing, ϕ -mixing, ψ -mixing and many more. Among all of those forms, α -mixing which was introduced

by Rosenblatt (1956) is the weakest and is known to be fulfilled for many stochastic process. For more properties of the different mixing processes, one can refer to Doukhan (1994) and Bradley (2005).

The paper is organised as follows: in the next section, we present the LTRC model along with relevent notations and definitions. Section 3 introduces the new estimator, while Section 4 outlines the assumptions and main result. A simulation study is carried out in Section 5, followed by a real data application in Section 6. In Section 7, we provide the concluding remarks and discuss potential avenues for future research. Finally, the proofs are given in the Appendix.

2. LTRC model and notations

Let $\{(Y_i, T_i, W_i); i = 1, ..., N\}$ be a sequence of random vectors from (Y, T, W), where Y denotes the lifetime with continuous distribution function (df) F. T and W are the variables of left truncation and right censoring times with continuous dfs L and G, respectively. In the random LTRC model, one observes (Z, T, δ) if $Z \ge T$, where $Z = Y \land W$ and $\delta = \mathbb{1}_{\{Y \le W\}}$, with \land denoting the minimum operator and $\mathbb{1}_A$ being the indicator function of the event A. When Z < T, nothing is observed. All along this paper, we suppose that Y, T and W are independent of each other, then Z has a df H = 1 - (1 - F)(1 - G). Take $\alpha := \mathbb{P}$ $(Z \ge T)$, so it is necessary to assume $\alpha > 0$ in order to have at least one observation available. As a consequence of truncation, n, the size of the actual sample, is a Bin (N, α) rv, with $n \le N$ and N fixed but not observable. Then without possible confusion, we still denote $\{(Z_i, T_i, \delta_i); i = 1, ..., n\}$ the observed sample.

Since N is unknown and n is known, our results would be stated with respect to conditional probability **P** (related to the n-sample) instead of the probability measure \mathbb{P} (related to the N-sample). Similarly, \mathbb{E} and **E** will denote the expectation operators related to \mathbb{P} and **P**, respectively. More generally, any operator or function related to the probability measure **P** will be denoted in bold.

Now, let $C(\cdot)$ be defined by

$$C(y) := \mathbf{P}(T \le y \le Z)$$

= $\mathbb{P}(T \le y \le Z \mid Z \ge T)$
= $\frac{1}{\alpha}L(y)(1 - H(y)),$ (3)

with the empirical estimator

$$C_n(y) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{T_i \le y \le Z_i\}}.$$

The nonparametric estimator of the df F is the TJW product-limit estimator F_n , defined in Tsai, Jewell, and Wang (1987) as

$$F_n(y) = 1 - \prod_{i=1}^n \left(1 - \frac{\mathbb{1}_{\{Z_i \le y\}} \delta_i}{n C_n(Z_i)} \right).$$

4 🕒 N. BAYARASSOU ET AL.

For any df Q, we define

$$a_Q = \inf\{y : Q(y) > 0\}$$
 and $b_Q = \sup\{y : Q(y) < 1\}$

as the endpoints of the Q support. Gijbels and Wang (1993) pointed out that the df F can be estimated only if

$$a_L \le a_H$$
 and $b_L \le b_H$. (4)

By the strong law of large numbers, as $N \to \infty$ we have

$$\frac{n}{N} \longrightarrow \alpha, \quad \mathbb{P}\text{- a.s.}$$

Note that, the ratio $\frac{n}{N}$ can not be used to estimate α since N is unknown. Indeed, from (3) we have

$$\alpha = \frac{L(y)(1 - F(y))(1 - G(y))}{C(y)}$$
(5)

and by following the idea of He and Yang (1998), a feasible estimator for α is given by

$$\alpha_n = \frac{L_n(y)(1 - F_n(y))(1 - G_n(y))}{C_n(y)},$$
(6)

for any *y* such that $C_n(y) \neq 0$, where G_n is the concomitant TJW estimator of the df *G*, defined as

$$G_n(y) = 1 - \prod_{i=1}^n \left(1 - \frac{\mathbb{1}_{\{Z_i \le y\}}(1 - \delta_i)}{nC_n(Z_i)} \right)$$
(7)

and L_n is the Lynden-Bell (1971) estimator of the df L, given by

$$L_n(y) = \prod_{i=1}^n \left(1 - \frac{\mathbb{1}_{\{T_i > y\}}}{nC_n(T_i)} \right).$$
(8)

Henceforth, in addition to the triplet (Z, T, δ) , assume that covariates are present and then one observes { $(X_i, Z_i, T_i, \delta_i)$; i = 1, ..., n}, with $Z_i \ge T_i$. Throughout this paper, we suppose that condition (4) is satisfied and

(T, W) are independent of (\mathbf{X}, Y) . (9)

Then, the sub-conditional df of $(\mathbf{X}, Z, \delta = 1)$ is given by

$$\begin{aligned} \mathbf{H}_{1}(\mathbf{x}, y) &:= \mathbf{P}(\mathbf{X} \leq \mathbf{x}, Z \leq y, \delta = 1) \\ &= \mathbb{P}(\mathbf{X} \leq \mathbf{x}, Z \leq y, \delta = 1 \mid Z \geq T) \\ &= \frac{1}{\alpha} \mathbb{P}(\mathbf{X} \leq \mathbf{x}, Y \leq y, Y \leq W, Y \geq T) \\ &= \frac{1}{\alpha} \int_{\mathbb{R}^{d}} \int_{a_{H}}^{y} L(t)(1 - G(t)) f_{\mathbf{X}, Y}(\mathbf{u}, t) \, \mathbf{du} \, \mathbf{dt}, \end{aligned}$$
(10)

where, $f_{\mathbf{X},Y}(\cdot, \cdot)$ is the joint density function of (\mathbf{X}, Y) .

By differentiating, (10) becomes

$$d\mathbf{H}_1(\mathbf{x}, y) = \frac{1}{\alpha} L(y)(1 - G(y)) f_{\mathbf{X}, Y}(\mathbf{x}, y).$$
(11)

In the sequel, {(X_i, Z_i, T_i, δ_i); i = 1, ..., n} is assumed to be an α -mixing sequence of random vectors. Recall that a sequence { ϑ_i ; $i \ge 1$ } is said to be α -mixing (strong mixing) if the mixing coefficient

$$\alpha(n) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^k \text{ and } B \in \mathcal{F}_{k+n}^\infty, k \in \mathbb{N}^*\}$$

converges to zero as $n \to \infty$, where \mathcal{F}_l^m denotes the σ -algebra generated by $\{\vartheta_j; l \le j \le m\}$.

3. Definition of the estimator

The main purpose of this section is to estimate the regression function expressed in (2), which can be written as

$$m(\mathbf{x}) = \frac{\int_{\mathbb{R}^d} y^{-1} f_{\mathbf{X}, Y}(\mathbf{x}, y) \, \mathrm{d}y}{\int_{\mathbb{R}^d} y^{-2} f_{\mathbf{X}, Y}(\mathbf{x}, y) \, \mathrm{d}y} =: \frac{\psi_1(\mathbf{x})}{\psi_2(\mathbf{x})},\tag{12}$$

where

$$\psi_{\ell}(\mathbf{x}) = \int_{\mathbb{R}^d} y^{-\ell} f_{\mathbf{X},Y}(\mathbf{x},y) \, \mathrm{d}y, \quad \text{for } \ell = 1, 2.$$

As already pointed out in the first section, in the case of complete data, Jones et al. (2008) constructed an estimator for (2), given by

$$\widehat{m}_N(\mathbf{x}) = \frac{\sum\limits_{i=1}^{N} Y_i^{-1} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_N}\right)}{\sum\limits_{i=1}^{N} Y_i^{-2} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_N}\right)},$$
(13)

where $K_d : \mathbb{R}^d \to \mathbb{R}$ is a kernel function and h_N is a sequence of positive real numbers that approaches zero as $N \to \infty$. Note that, the formula in (13) is the direct analogue of the estimator of Nadaraya (1964) and Watson (1964) in the MSE case. Under the random right censored model, Bouhadjera et al. (2019) used the so-called synthetic data to define a kernel estimator of (2) as

$$\widehat{m}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \frac{\delta_i Z_i^{-1}}{\overline{G}_n(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right)}{\sum_{i=1}^{n} \frac{\delta_i Z_i^{-2}}{\overline{G}_n(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right)},$$

where $\bar{G}_n = 1 - G_n$ is the estimator of the survival function of the censored rv *W*.

Now, under the random LTRC model, following the idea introduced by Carbonez, Gyorfi, and Van Der Meulen (1995), an unbiased estimate for $\mathbb{E}[Y^{-\ell} | \mathbf{X}]$, $\ell = 1, 2$ is

6 🛞 N. BAYARASSOU ET AL.

given by

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\delta_i Z_i^{-\ell}}{L(Z_i) \bar{G}(Z_i)} \mathbb{1}_{\{Z_i \ge T_i\}}.$$
(14)

Indeed, using the conditional expectation property and condition (9), we have

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{\delta_{i}Z_{i}^{-\ell}}{L(Z_{i})\bar{G}(Z_{i})}\mathbb{1}_{\{Z_{i}\geq T_{i}\}}\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\mathbb{E}\left[\frac{\delta_{i}Z_{i}^{-\ell}}{L(Z_{i})\bar{G}(Z_{i})}\mathbb{1}_{\{Z_{i}\geq T_{i}\}} \mid \mathbf{X}_{i}, Y_{i}\right]\right)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\frac{Y_{i}^{-\ell}}{L(Y_{i})\bar{G}(Y_{i})}\mathbb{E}[\mathbb{1}_{\{Y_{i}\leq W_{i}\}}\mathbb{1}_{\{Y_{i}\geq T_{i}\}} \mid \mathbf{X}_{i}, Y_{i}]\right)$$
$$= \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[Y_{i}^{-\ell} \mid \mathbf{X}_{i}]$$
$$= \mathbb{E}[Y^{-\ell} \mid \mathbf{X}].$$

Unfortunately, (14) can not be used in practice since N, L and \overline{G} are unknown. Therefore, following the same reasoning as in Ould Saïd and Lemdani (2006), we define $\widehat{m}_n(\mathbf{x})$ as a kernel estimator for (2), given for any $\mathbf{x} \in \mathbb{R}^d$ by

$$\widehat{m}_n(\mathbf{x}) =: \frac{\widehat{\psi}_1(\mathbf{x})}{\widehat{\psi}_2(\mathbf{x})},\tag{15}$$

where

$$\widehat{\psi}_{\ell}(\mathbf{x}) = \frac{\alpha_n}{nh_n^d} \sum_{i=1}^n \frac{\delta_i Z_i^{-\ell}}{L_n(Z_i) \overline{G}_n(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \quad \text{for } \ell = 1, 2$$

4. Assumptions and main result

In order to formulate the main result, we need to introduce some notations:

$$\widetilde{\psi}_n(\mathbf{x}) = \frac{\alpha}{nh_n^d} \sum_{i=1}^n \frac{\delta_i Z_i^{-\ell}}{L(Z_i) \overline{G}(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \quad \text{for } \ell = 1, 2$$

and

$$\mu_{\ell}(\mathbf{u}) = \int_{a_H}^{y} \frac{t^{-2\ell}}{L(t)\overline{G}(t)} f_{\mathbf{X},Y}(\mathbf{u},t) \,\mathrm{d}t, \quad \text{for } \ell = 1, 2.$$

Let \mathcal{C} be a compact set in \mathbb{R}^d and $\inf_{\mathbf{x}\in\mathcal{C}} \widehat{\psi}_2(\mathbf{x}) > 0$. Throughout the paper, when no confusion is possible, we denote by C any generic constant.

4.1. Assumptions

We will make use of the following assumptions gathered together for easy references.

- **K.** The kernel $K(\cdot)$ is a bounded probability density with compact support satisfies:
 - (a) $\int_{\mathbb{R}^d} K_d(\mathbf{r}) \, \mathbf{dr} = 1$, $\int_{\mathbb{R}^d} r_i K_d(\mathbf{r}) \, \mathbf{dr} = 0$ and $\int_{\mathbb{R}^d} |r_i| |r_j| K_d(\mathbf{r}) \, \mathbf{dr} < \infty$, for $i, j = 1, \ldots, d$.
 - (b) $\int_{\mathbb{R}^d} K_d^2(\mathbf{r}) \, \mathbf{dr} < \infty$ and $\int_{\mathbb{R}^d} r_i K_d^2(\mathbf{r}) \, \mathbf{dr} < \infty$, for $i = 1, \ldots, d$.
 - (c) $\forall (\mathbf{t}, \mathbf{s}) \in \mathcal{C}^2 |K_d(\mathbf{t}) K_d(\mathbf{s})| \leq ||\mathbf{t} \mathbf{s}||^{\gamma}$, for $\gamma > 0$.
- **H.** The bandwidth h_n satisfies :
 - (a) $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} nh_n^d = \infty$ and $\lim_{n\to\infty} \frac{\log n}{nh_n^d} = 0$.
 - (b) $\lim_{n \to \infty} h_n^{d(v-2)} \log n = 0.$
 - (c) $\exists \theta > 0, \exists C > 0$, such that

$$Cn^{rac{\gamma(3-\nu)}{\gamma(\nu+1)+2\gamma+1}+ heta d} \leq h_n^d$$
, for all $\nu > 3$ and $\gamma > 0$.

- **D.** (a) The function $\psi_{\ell}(\cdot)$ is twice continuously differentiable and $\sup_{\mathbf{x}\in\mathcal{C}} |\frac{\partial^2 \psi_{\ell}(\mathbf{x})}{\partial x_i \partial x_j}| < \infty$, for i, j = 1, ..., d.
 - (b) $\forall Y > 0, \exists C$, such that $Y^{-\ell} \leq C$, for $\ell = 1, 2$.
 - (c) The function $\mu_{\ell}(\cdot)$ is continuously differentiable and $\sup_{\mathbf{x}\in\mathcal{C}} |\frac{\partial \mu_{\ell}(\mathbf{x})}{\partial x_i}| < \infty$, for $i = 1, \ldots, d$.
 - (d) The joint density $\Upsilon_{i,j}(\cdot, \cdot)$ of $(\mathbf{X}_i, \mathbf{X}_j)$ exists and satisfies

$$\sup_{\mathbf{t},\mathbf{s}\in\mathcal{C}}|\Upsilon_{i,j}(\mathbf{t},\mathbf{s})-\Upsilon_i(\mathbf{t})\Upsilon_j(\mathbf{s})|\leq\infty,\quad\text{for }i,j=1,\ldots,n.$$

4.1.1. Comments on the assumptions

Assumptions K, H(a), and D(a) are commonly used in nonparametric regression estimation in both independent and dependent cases. Assumption D(b) implies that the inverse of the variable of interest Y is bounded, which is specifically useful for proving Lemmas A.2 and A.3. Assumptions H(b), H(c), D(c), and D(d) are technical and are necessary for studying the covariance term.

4.2. Main result

The following theorem presents the uniform almost sure convergence with a rate, of the relative error regression (RER) estimator defined in (15).

Theorem 4.1: Under Assumptions K, H, and D, we have

$$\sup_{\mathbf{x}\in\mathcal{C}}|\widehat{m}_n(\mathbf{x}) - m(\mathbf{x})| = O(h_n^2) + O\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right) \mathbf{P} - \text{a.s. as } n \to \infty.$$

The proof of the theorem is postponed in the Appendix.

5. Simulation study

To show how good our estimator $\widehat{m}_n(\cdot)$ is, a simulation study is performed for some particular cases of fixed size and different censoring, truncation and dependency rates, when the covariate **X** is one- and bi-dimensional rv (i.e. d = 1 and d = 2). 8 🛞 N. BAYARASSOU ET AL.

5.1. One-dimensional case

5.1.1. Algorithm

(1) Generate an α -mixing sequence { X_t ; $t \ge 1$ } by the following AR(1) model

$$X_{t+1} = \begin{cases} \rho X_t + 0.5e_{t+1}, & \text{if } \rho > 0.5\\ \rho X_t + e_{t+1}, & \text{else,} \end{cases}$$

where $0 < \rho < 1$ controls the degree of dependency, $X_1 = e_1$ and $e_t \rightsquigarrow N(0, 1)$.

- (2) Calculate $Y_t = m(X_t) + \epsilon_t$, $t \ge 1$, where the white noise $\epsilon_t \rightsquigarrow N(0, 0.1)$.
- (3) Determine $Z_t = Y_t \wedge W_t$ and $\delta_t = \mathbb{1}_{\{Y_t \leq W_t\}}, t \geq 1$, where W_t is generated according to a exponential distribution with parameter a_0 which allows obtaining different censoring percentage (CP).
- (4) Generate $T_t = \rho T_{t-1} + b_0 + \xi_t$, $t \ge 2$, where $T_1 = \xi_1, \xi_t \rightsquigarrow N(0, 0.1)$ for $t \ge 1$ and b_0 is adapted to achieve different truncation percentage (TP).
- (5) Test if $Z_t \ge T_t$, $t \ge 1$. If true, the vector $(X_t, Z_t, T_t, \delta_t)$ is included in the final sample. Otherwise, reject the observation $(X_t, Z_t, T_t, \delta_t)$ and go back to step 1.
- (6) Repeat this procedure until the final simple size is n, i.e. $(X_i, Z_i, T_i, \delta_i)$; i = 1, ..., n.
- (7) Compute $\overline{G}_n(\cdot)$ and $L_n(\cdot)$ from (7) and (8), respectively.
- (8) Calculate the estimator $\widehat{m}_n(x)$ from (15) for $x \in \mathcal{C} = [-1.5, 1.5]$. The kernel K(.) is taken as a standard normal function. The choice of the bandwidth is discussed immediately afterward.

5.1.2. Bandwidth selection

The bandwidth h_n is chosen as the minimiser of the global mean squared error (GMSE) criterion. We select this bandwidth from a grid of values denoted as \mathcal{H} . For each candidate bandwidth value $h_n \in \mathcal{H}$, we perform the following steps.

• Compute the MSE for the estimator $\widehat{m}_n(.)$ at the equidistant points $(x_i, i = 1, ..., A = 20)$ belonging to the compact set *C*. The MSE is calculated along B = 50 replications by

MSE
$$(x_i) = \frac{1}{B} \sum_{j=1}^{B} (\hat{m}_{n,j}(x_i) - m(x_i))^2, \quad i = 1, \dots, A.$$

Here $\widehat{m}_{n,j}(\cdot)$ is the value of the estimator $\widehat{m}_n(\cdot)$ at iteration *j*.

• Compute the GMSE by

$$GMSE(h_n) = \frac{1}{A} \sum_{i=1}^{A} MSE(x_i).$$
(16)

Finally, the optimal bandwidth is determined by

$$\underset{h_n \in \mathcal{H}}{\operatorname{arg\,min}} \operatorname{GMSE}(h_n).$$

5.1.3. Performance of the estimator $\widehat{m}_{n}(\cdot)$

In this part, we study the performance of our estimator when the theoretical function is of linear and nonlinear form.



Figure 1. $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with $\rho = 0.1$, CP = 20%, TP = 20%, n = 50, 100, and 300 respectively.



Figure 2. $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with n = 300, CP = 20%, TP = 20%, $\rho = 0.1, 0.5$ and, 0.8 respectively.



Figure 3. $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with $\rho = 0.1$, n = 300, TP = 20%, CP = 10%, 30%, and 60% respectively.

(1) Linear case

We consider the following linear regression function m(x) = x + 5, then $Y_i = X_i + 5 + \epsilon_i$, i = 1, ..., n.

- (a) Effect of sample size and dependency: From Figures 1 and 2 when CP and TP are fixed, we notice that the higher the sample size and smaller the rate of dependency, the better the quality of fit.
- (b) Effect of CP and TP: It is easy to see from Figure 3 that the estimator's quality is affected by CP, whereas it does not seem to be influenced by TP, as shown in Figure 4. In general, our estimator curve remains close to the theoretical curve even for a high CP.
- (c) Effect of outliers: To show the robustness of our estimator, we create artificial outliers in the data; 4% of each sample is multiplied by a multiplier factor (MF). Then, from Figure 5, it is very clear that our estimator is resistant in the presence of outliers.

(2) Nonlinear case

Now, we consider the case of nonlinear regression by choosing the following three models



Figure 4. $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with $\rho = 0.1$, n = 300, CP = 20%, TP = 10%, 30%, and 60% respectively.



Figure 5. $m(\cdot)$ and $\hat{m}_n(\cdot)$ with $\rho = 0.1$, n = 300, CP = 20%, TP = 20%, MF = 50, 100, and 150 respectively.

Model 1:

 $Y = m_1(X) + \epsilon$, with $m_1(x) = \exp(x) + 3$,

Model 2:

 $Y = m_2(X) + \epsilon$, with $m_2(x) = \sin(2x) + 4$,

Model 3:

$$Y = m_3(X) + \epsilon$$
, with $m_3(x) = x^2 + 5$.

Figure 6 shows that the quality of fit for the nonlinear model is as good as for linear model.

5.1.4. Comparison study

Here, the goal is to compare the performance of the RER estimator with the classical regression (CR) estimator studied by Bey et al. (2022) and defined as

$$\widehat{m}_{NW}(x) = \frac{\sum_{i=1}^{n} \frac{\delta_i Z_i}{L_n(Z_i)\overline{G}_n(Z_i)} K_d\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^{n} \frac{\delta_i}{L_n(Z_i)\overline{G}_n(Z_i)} K_d\left(\frac{x-X_i}{h_n}\right)},$$

in absence and presence of outliers. The performance of both estimators is evaluated via some graphic curves and the GMSE criterion under the linear regression function described in the one-dimensional case.

For the first case when there are no outliers in the observed samples, we can see from Figure 7 and Table 1 that there is no meaningful difference between the estimators $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ and both of them share the following points



Figure 6. $m(\cdot)$ and $\hat{m}_n(\cdot)$ with $\rho = 0.1$, n = 300, CP = 20% and TP = 20% for model 1, 2 and, 3 respectively.



Figure 7. $m(\cdot)$, $\hat{m}_n(\cdot)$ and $\hat{m}_{NW}(\cdot)$ with $\rho = 0.1$, CP = 20%, TP = 20%, n = 50, 100, and 300 respectively.

ρ	n		CP =	= 20%		TP = 20%				
		TP = 10%		TP = 40%		CP = 10%		CP = 40%		
		<i>m</i> _n	\widehat{m}_{NW}	m _n	\widehat{m}_{NW}	<i>m</i> _n	\widehat{m}_{NW}	<i>m</i> _n	\widehat{m}_{NW}	
0.1	100	0.00228	0.00231	0.00234	0.00237	0.00203	0.00231	0.00417	0.00437	
	300	0.00118	0.00105	0.00127	0.00116	0.00088	0.00091	0.00152	0.00150	
0.8	100	0.00347	0.00325	0.00382	0.00333	0.00289	0.00250	0.00632	0.00594	
	300	0.00181	0.00162	0.00224	0.00243	0.00095	0.00111	0.00217	0.00208	

Table 1. GMSE's values of $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$.

- The quality of fit becomes better when the sample size n increases.
- The estimator is affected by the degree of dependency and performs better for a small ρ .
- The accuracy of the estimator is influenced by CP and decreases with increasing CP, however it remains acceptable.
- The quality of estimation is slightly affected by TP for a small size of n, but this effect disappears with increasing sample size.

In the second case, we compare the estimators $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ in the presence of outliers. For this purpose, we artificially introduce outliers by multiplying 4 % of each sample by a MF. As illustrated in Figure 8 and Table 2, the RER estimator is more stable than the CR estimator. This means that even if the quality of the estimation for both estimators decreases with increasing outliers, but this decrease in the quality is still not significant in the relative error regression compared to the classical one.



Figure 8. $m(\cdot)$, $\hat{m}_n(\cdot)$ and $\hat{m}_{NW}(\cdot)$ with n = 300, $\rho = 0.1$, CP = 20%, TP = 20%, MF = 50, 100, and 150 respectively.

Table 2. GMSE's values of $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ with outliers for $\rho = 0.1$.

	CP =	: 20%			TP = 20%			
TF	TP = 10%		TP = 40%		CP = 10%		CP = 40%	
<i>m</i> _n	\widehat{m}_{NW}	\widehat{m}_n	\widehat{m}_{NW}	\widehat{m}_n	\widehat{m}_{NW}	\widehat{m}_n	\widehat{m}_{NW}	
0.00245 0.00316	3.11207×10^{2} 1.52516×10^{3}	0.00266 0.00358	4.18845×10^{2} 1.92048×10^{3}	0.00237 0.00263	2.48798×10^{2} 1.00017×10^{3}	0.00463 0.00549	5.62563×10^{2} 2.53366 $\times 10^{3}$	
0.00341 0.00107 0.00121	2.43604×10^{3} 2.30381×10^{2} 9.30793×10^{2}	0.00372 0.00107 0.00132	2.81923×10^{3} 3.05945×10^{2} 1.17058×10^{3}	0.00314 0.00092 0.00102	2.16226×10^{3} 2.26077×10^{2} 1.39562×10^{3}	0.00523 0.00164 0.00183	3.32406×10^{3} 4.51992×10^{2} 1.92666×10^{3}	
	mn 0.00245 0.00316 0.00341 0.00107 0.00121 0.00131	$\begin{tabular}{ c c c c c } \hline TP &= 10\% \\ \hline $	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	$\begin{tabular}{ c c c c c c c c c c c c c c c c c c c$	$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	

5.2. Bi-dimensional case

In this second subsection, the aim is to study the performance of our estimator in the case of a bi-dimensional covariate under the following two models Model 1:

$$Y = m_1(X_1, X_2) + \epsilon$$
, with $m_1(x_1, x_2) = x_1 + x_2 + 5$,

Model 2:

$$Y = m_2(X_1, X_2) + \epsilon$$
, with $m_2(x_1, x_2) = \cos(2x_1) + \cos(2x_2) + 4$

The data is generated using the same algorithm as for the one-dimensional case. In each model, We simulate $(X_{1,t}, X_{2,t}), t \ge 1$ as follows

$$X_{j,t+1}_{j=1,2} = \begin{cases} \rho X_{j,t} + 0.5e_{j,t+1}, & \text{if } \rho > 0.5\\ \rho X_{j,t} + e_{j,t+1}, & \text{else,} \end{cases}$$

where $X_{j,1} = e_{j,1}$ and $e_{j,t} \rightsquigarrow N(0,1)$. Then, we calculate $Y_t = m(X_{1,t}, X_{2,t}) + \epsilon_t$, $t \ge 1$. To compute the estimator $\widehat{m}_n(\cdot, \cdot)$, we use a standard multivariate normal kernel and a bandwidth that minimises the GMSE defined in (16).

The results are presented in the following figures and tables. In general, the same comments that we gave in the one-dimensional case can be given here. More clearly, in the absence of outliers, we observe from Figures 9 and 10 (Model 1), Figure 11 (Model 2) and Table 3 that the estimators $\hat{m}_n(\cdot)$ and $\hat{m}_{NW}(\cdot)$ are almost equivalent and the quality of estimation for both of them becomes better for a large sample size and a small rate of censoring,



Figure 9. True surface for Model 1 with $\rho = 0.1$, n = 300, CP = 20%, and TP = 20%.



Figure 10. RER surface for Model 1 with $\rho = 0.1$, CP = 20%, TP = 20%, n = 50, 100, and 300 respectively.

			CP =	= 20%		TP = 20%				
		TP = 10%		TP = 40%		CP = 10%		CP = 40%		
ρ	n	<i>m</i> _n	\widehat{m}_{NW}	<i>m</i> _n	\widehat{m}_{NW}	<i>m</i> _n	\widehat{m}_{NW}	- m _n	\widehat{m}_{NW}	
0.1	100 300	0.06396 0.03566	0.05726 0.03014	0.11919 0.05049	0.127636 0.05647	0.05124 0.02567	0.05513 0.02306	0.09572 0.04346	0.08368 0.04104	
0.8	100 300	0.08249 0.05667	0.09146 0.05126	0.16514 0.07682	0.15514 0.07011	0.06196 0.04590	0.07313 0.04653	0.15381 0.06998	0.14740 0.06332	

Table 3. GMSE's values of $\widehat{m}_n(\cdot, \cdot)$ and $\widehat{m}_{NW}(\cdot, \cdot)$.

truncation and dependency. However, in the presence of outliers, the RER estimator performs better than the CR estimator in all cases, as confirmed by Table 4. To conclude, the quality of fit for our estimator is good but better in the one-dimensional case.



Figure 11. True and RER surfaces for Model 2 with $\rho = 0.1$, n = 300, CP = 20%, and TP = 20%.

			CP =	: 20%		TP = 20%				
		TP	TP = 10%		TP = 40%		CP = 10%		CP = 40%	
n	MF	<i>m</i> _n	\widehat{m}_{NW}	\widehat{m}_n	\widehat{m}_{NW}	<i>m</i> _n	\widehat{m}_{NW}	<i>m</i> _n	\widehat{m}_{NW}	
	50	0.07123	7.81493×10 ²	0.11381	1.38425×10 ³	0.05571	6.68271×10 ²	0.10518	9.13577×10 ²	
100	100	0.06744	2.64428×10 ³	0.12730	4.41663×10 ³	0.06166	2.11341×10 ³	0.14459	3.78598×10 ³	
	150	0.07783	4.03922×10 ³	0.12358	5.27606×10 ³	0.06310	3.43390×10 ³	0.13849	4.54624×10 ³	
	50	0.03231	5.48369×10 ²	0.05916	8.24651×10 ²	0.02884	3.85714×10 ²	0.04664	7.21458×10 ²	
300	100	0.04076	2.02754×10 ³	0.06805	3.68605×10 ³	0.02610	1.77065×10 ³	0.04454	3.22474×10 ³	
	150	0.03820	3.88562×10^3	0.06333	4.77619×10 ³	0.02784	2.04465×10 ³	0.05414	4.03390×10 ³	

Table 4. GMSE's values of $\widehat{m}_n(\cdot, \cdot)$ and $\widehat{m}_{NW}(\cdot, \cdot)$ with outliers for $\rho = 0.1$.

6. Real data application

In this section, we present a real data application where we assess the effectiveness of the RER estimator in the context of dependent data that is left truncated and right censored. The dataset used in our analysis contains information about patients diagnosed with AIDS in Australia before 1st July 1991, and it was obtained from Dr. Patty Solomon. For more detailed information about the dataset, we recommend consulting (Venables and Ripley 2002).

The study focuses on patients who survived at least one year after diagnosis, comprising a total of 1276 individuals. The dataset includes various factors such as the dates of diagnosis, the dates of death, patients gender, age at diagnosis, survival status at the end of the study, as well as the state and transmission category they belong to.

Our main interest lies in analysing the time to diagnosis and time to death. Upon examining the relationships within the data, we found a strong correlation (0.75) between these two factors, indicating their interdependence. Furthermore, the partial correlation plot suggests that our data follows an autoregressive process of order greater than 1.

To make predictions, we divided the dataset into two parts: 80% of the data served as the learning sample to calculate the estimator, while the remaining 20% formed the test sample to evaluate the predictions quality. We utilised the standard gaussian kernel function and employed the cross-validation method to determine the optimal bandwidth. It is important



Figure 12. Scatter plot of time to diagnosis and time to death: censored vs. uncensored observations with true and RER predicted values for AIDS patients.

to note that around 43% of the data was censored. Therefore, we excluded the censored data from the predicted values since it has no meaning to predict the survival time for such observations.

Figure 12 provides visual representations of our analysis. On the left-hand side of the figure, a scatter plot of the data is shown, distinguishing between censored and uncensored observations. On the right-hand side, the true and RER predicted values are displayed. It can be observed that the majority of the RER predicted values are close to the true values, demonstrating the robustness of our method and the accuracy of our predictor.

7. Conclusion

In this paper, our focus is on estimating the relative regression function under left truncation and right censoring using a nonparametric approach. We propose a kernel type estimator that minimises the mean squared relative error and establish its strong uniform convergence, along with the corresponding rate under α -mixing condition. Furthermore, we evaluate the practical effectiveness of our estimator through both finite sample analysis and a real data application. The results demonstrated the favourable performance of the proposed methodology. As a direction for future research, it would be interesting to extend our work to the case where the covariable exhibits a functional nature.

Acknowledgments

The authors are grateful to the two anonymous referees and associate-editor whose careful reading gave us the opportunity to improve the quality of the first version of the paper and add simulation with real data.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Altendji, B., Demongeot, J., Laksaci, A., and Rachdi, M. (2018), 'Functional Data Analysis: Estimation of the Relative Error in Functional Under Left-truncated', *Journal of Nonparametric Statistics*, 30, 472–490.
- Attouch, M., Laksaci, A., and Messabihi, N. (2017), 'Nonparametric Relative Error Regression for Spatial Random Variables', *Statistical Papers*, 58, 987–1008.

- Benseradj, H., and Guessoum, Z. (2022), 'Strong Uniform Consistency Rate of An M-estimator of Regression Function for Incomplete Data Under α-mixing Condition', *Communications in Statistics Theory and Methods*, 51, 2082–2115.
- Bey, S., Guessoum, Z., and Tatachak, A. (2022), 'Kernel Regression Estimation for LTRC and Associated Data', *Communications in Statistics Theory and Methods*, 52, 1–26.
- Bouabsa, W. (2021), 'Nonparametric Relative Error Estimation Via Functional Regressor by the K Nearest Neighbors Smoothing Under Truncation Random Data', *Applications and Applied Mathematics*, 16, 97–116.
- Bouhadjera, F., Ould Saïd, E., and Remita, M.R. (2019), 'Nonparametric Relative Error Estimation of the Regression Function for Censored Data'. Available at https://hal.archives-ouvertes.fr/hal-01994512.
- Bouhadjera, F., Ould Saïd, E., and Remita, M.R. (2022), 'Strong Consistency of the Nonparametric Local Linear Regression Estimation under Censorship Model', *Communications in Statistics – Theory and Methods*, 51, 7056–7072.
- Bradley, R.C. (2005), 'Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions', *Probability Surveys*, 2, 107–144.
- Carbonez, A., Gyorfi, L., and Van Der Meulen, E.C. (1995), 'Partitioning Estimates of a Regression Function under Random Censoring', *Statistics & Risk Modeling*, 13, 21–38.
- Chen, Q., and Dai, Y. (2003), 'Kernel Estimation of Higher Derivative of Density and Hazard Rate Function for Truncated and Censored Dependent Data', *Acta Mathematica Scientia*, 23, 477–486.
- Chen, C.M., and Shen, P.S. (2018), 'Conditional Maximum Likelihood Estimation in Semi-Parametric Transformation Model with LTRC Data', *Lifetime Data Analysis*, 24, 250–272.
- Demongeot, J., Hamie, A., Laksaci, A., and Rachdi, M. (2016), 'Relative-Error Prediction in Nonparametric Functional Statistics: Theory and Practice', *Journal of Multivariate Analysis*, 146, 261–268.
- Doukhan, P. (1994), Mixing: Properties and Examples, New York: Springer-Verlag.
- Ferraty, F., and Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, New York: Springer.
- Gijbels, I., and Wang, J.L. (1993), 'Strong Representations of the Survival Function Estimator for Truncated and Censored Data with Applications', *Journal of Multivariate Analysis*, 47, 210–229.
- He, S., and Yang, G.L. (1998), 'Estimation of the Truncation Probability in the Random Truncation Model', *The Annals of Statistics*, 26, 1011–1027.
- Jones, M.C., Park, H., Shin, K.I., Vines, S.K., and Jeong, S.O. (2008), 'Relative Error Prediction Via Kernel Regression Smoothers', *Journal of Statistical Planning and Inference*, 138, 2887–2898.
- Khardani, S. (2020), 'Relative Error Prediction for Twice Censored Data', *Mathematical Methods of Statistics*, 28, 291–306.
- Lynden-Bell, D. (1971), 'A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars', *Monthly Notices of Royal Astronomical Society*, 155, 95–118.
- Masry, E. (1986), 'Recursive Probability Density Estimation for Weakly Dependent Stationary Processes', *IEEE Transactions on Information Theory*, 32, 254–267.
- Molanes-López, E.M., and Cao, R. (2008), 'Relative Density Estimation for Left Truncated and Right Censored Data', *Journal of Nonparametric Statistics*, 20, 693–720.
- Nadaraya, E.A. (1964), 'On Estimating Regression', Theory of Probability & Its Applications, 9, 141-142.
- Ould Saïd, E., and Lemdani, M. (2006), 'Asymptotic Properties of a Nonparametric Regression Function Estimator with Randomly Data', *Annals of the Institute of Statistical Mathematics*, 58, 357–378.
- Park, H., and Stefanski, L.A. (1998), 'Relative-error Prediction', *Statistics & Probability Letters*, 40, 227–236.
- Rio, E. (2000), *Théorie Asymptotique Des Processus Aléatoires Faiblement Dépendants*, Collection Mathématiques & Applications, New York: Springer.
- Rosenblatt, M. (1956), 'A Central Limit Theorem and a Strong Mixing Condition', *Proceedings of the National Academy of Science of the U.S.A.*, 42, 43–47.

- Tsai, W.Y., Jewell, N.P., and Wang, M.C. (1987), 'Anote on the Product-Limit Estimator under Right Censoring and Left Truncation', *Biometrika*, 74, 883–886.
- Venables, W.N., and Ripley, B.D. (2002), *Modern Applied Statistics with S* (4th ed.), New York: Springer.
- Watson, G.S. (1964), 'Smooth Regression Analysis', Sankhyā: The Indian Journal of Statistics, 26, 356–372.

Appendix. Proofs

From (12) and (15), we have the following classical decomposition

$$\begin{aligned} \widehat{m}_{n}(\mathbf{x}) &- m(\mathbf{x}) \\ &= \frac{\widehat{\psi}_{1}(\mathbf{x})}{\widehat{\psi}_{2}(\mathbf{x})} - \frac{\psi_{1}(\mathbf{x})}{\psi_{2}(\mathbf{x})} \\ &= \frac{1}{\widehat{\psi}_{2}(\mathbf{x})} \left\{ \left[\widehat{\psi}_{1}(\mathbf{x}) - \psi_{1}(\mathbf{x}) \right] - m(\mathbf{x}) \left[\widehat{\psi}_{2}(\mathbf{x}) - \psi_{2}(\mathbf{x}) \right] \right\} \\ &= \frac{1}{\widehat{\psi}_{2}(\mathbf{x})} \left\{ \left[(\widehat{\psi}_{1}(\mathbf{x}) - \widetilde{\psi}_{1}(\mathbf{x})) + (\widetilde{\psi}_{1}(\mathbf{x}) - \mathbf{E}[\widetilde{\psi}_{1}(\mathbf{x})]) + (\mathbf{E}[\widetilde{\psi}_{1}(\mathbf{x})] - \psi_{1}(\mathbf{x})) \right] \\ &- m(\mathbf{x}) \left[(\widehat{\psi}_{2}(\mathbf{x}) - \widetilde{\psi}_{2}(\mathbf{x})) + (\widetilde{\psi}_{2}(\mathbf{x}) - \mathbf{E}[\widetilde{\psi}_{2}(\mathbf{x})]) + (\mathbf{E}[\widetilde{\psi}_{2}(\mathbf{x})] - \psi_{2}(\mathbf{x})) \right] \right\} \\ &=: \frac{1}{\widehat{\psi}_{2}(\mathbf{x})} \left\{ \left[\Lambda_{1,1}(\mathbf{x}) + \Lambda_{2,1}(\mathbf{x}) + \Lambda_{3,1}(\mathbf{x}) \right] - m(\mathbf{x}) \left[\Lambda_{1,2}(\mathbf{x}) + \Lambda_{2,2}(\mathbf{x}) + \Lambda_{3,2}(\mathbf{x}) \right] \right\}. \end{aligned}$$
(A1)

In order to prove Theorem 4.1, some auxiliary results are needed and will be introduced in Lemmas A.1–A.3 hereafter.

Lemma A.1: Under Assumptions K(a) and D(a), we have for $\ell = 1, 2$

$$\sup_{\mathbf{x}\in\mathcal{C}}|\Lambda_{3,\ell}(\mathbf{x})|=O_{\mathrm{a.s.}}(h_n^2)\quad\text{as }n\to\infty.$$

Proof: We have

$$|\Lambda_{3,\ell}(\mathbf{x})| = \left| \mathbf{E} \left[\frac{\alpha}{nh_n^d} \sum_{i=1}^n \frac{\delta_i Z_i^{-\ell}}{L(Z_i)(1 - G(Z_i))} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \right] - \psi_\ell(\mathbf{x}) \right|$$
$$= \left| \mathbf{E} \left[\frac{\alpha}{h_n^d} \frac{\delta_1 Z_1^{-\ell}}{L(Z_1)(1 - G(Z_1))} K_d\left(\frac{\mathbf{x} - \mathbf{X}_1}{h_n}\right) \right] - \psi_\ell(\mathbf{x}) \right|.$$
(A2)

From (11) and using a change of variable, we get

$$\begin{split} \mathbf{E} & \left[\frac{\alpha}{h_n^d} \frac{\delta_1 Z_1^{-\ell}}{L(Z_1)(1 - G(Z_1))} K_d \left(\frac{\mathbf{x} - \mathbf{X}_1}{h_n} \right) \right] \\ &= \int_{\mathbb{R}^d} \int_{a_H}^y \frac{\alpha t^{-\ell}}{h_n^d L(t)(1 - G(t))} K_d \left(\frac{\mathbf{x} - \mathbf{u}}{h_n} \right) \, \mathrm{d}\mathbf{H}_1(\mathbf{u}, t) \\ &= \int_{\mathbb{R}^d} \int_{a_H}^y \frac{1}{h_n^d} t^{-\ell} K_d \left(\frac{\mathbf{x} - \mathbf{u}}{h_n} \right) f_{\mathbf{X}, \mathbf{Y}}(\mathbf{u}, t) \, \mathrm{d}\mathbf{u} \, \mathrm{d}t \\ &= \int_{\mathbb{R}^d} \frac{1}{h_n^d} K_d \left(\frac{\mathbf{x} - \mathbf{u}}{h_n} \right) \psi_\ell(\mathbf{u}) \, \mathrm{d}\mathbf{u} \\ &= \int_{\mathbb{R}^d} K_d(\mathbf{r}) \psi_\ell(\mathbf{x} - \mathbf{r}h_n) \, \mathrm{d}\mathbf{r}. \end{split}$$

18 🕒 N. BAYARASSOU ET AL.

So, (A2) becomes

$$\begin{aligned} \left| \mathbf{E} \left[\frac{\alpha}{h_n^d} \frac{\delta_1 Z_1^{-\ell}}{L(Z_1)(1 - G(Z_1))} K_d \left(\frac{\mathbf{x} - \mathbf{X}_1}{h_n} \right) \right] - \psi_\ell(\mathbf{x}) \right. \\ &= \left| \int_{\mathbb{R}^d} K_d(\mathbf{r}) \psi_\ell(\mathbf{x} - \mathbf{r}h_n) \, \mathbf{dr} - \psi_\ell(\mathbf{x}) \right| \\ &= \left| \int_{\mathbb{R}^d} K_d(\mathbf{r}) \left[\psi_\ell(\mathbf{x} - \mathbf{r}h_n) - \psi_\ell(\mathbf{x}) \right] \mathbf{dr} \right|. \end{aligned}$$

A Taylor expansion around **x** gives

$$\psi_{\ell}(\mathbf{x} - \mathbf{r}h_n) - \psi_{\ell}(\mathbf{x}) = -h_n \sum_{i=1}^d r_i \frac{\partial \psi_{\ell}(\mathbf{x})}{\partial x_i} + \frac{h_n^2}{2} \left\{ \sum_{i=1}^d r_i^2 \frac{\partial^2 \psi_{\ell}(\mathbf{x}_0)}{\partial x_i^2} + 2\sum_{i=1}^d \sum_{j=1}^d r_i r_j \frac{\partial^2 \psi_{\ell}(\mathbf{x}_0)}{\partial x_i \partial x_j} \right\}$$

when \mathbf{x}_0 is between $\mathbf{x} - \mathbf{r}h_n$ and \mathbf{x} . Then

$$\begin{split} \sup_{\mathbf{x}\in\mathcal{C}} |\Lambda_{3,\ell}(\mathbf{x})| &= \sup_{\mathbf{x}\in\mathcal{C}} \left| \int_{\mathbb{R}^d} K_d(\mathbf{r}) \left[\psi_\ell(\mathbf{x} - \mathbf{r}h_n) - \psi_\ell(\mathbf{x}) \right] d\mathbf{r} \right| \\ &= \sup_{\mathbf{x}\in\mathcal{C}} \left| \int_{\mathbb{R}^d} K_d(\mathbf{r}) \left[-h_n \sum_{i=1}^d r_i \frac{\partial \psi_\ell(\mathbf{x})}{\partial x_i} + \frac{h_n^2}{2} \left\{ \sum_{i=1}^d r_i^2 \frac{\partial^2 \psi_\ell(\mathbf{x}_0)}{\partial x_i^2} \right. \right. \\ &+ 2 \sum_{i=1}^d \sum_{j=1}^d r_i r_j \frac{\partial^2 \psi_\ell(\mathbf{x}_0)}{\partial x_i \partial x_j} \right\} \right] d\mathbf{r} \\ &\leq \frac{h_n^2}{2} \left\{ \sum_{i=1}^d \sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{\partial^2 \psi_\ell(\mathbf{x}_0)}{\partial x_i^2} \right| \int_{\mathbb{R}^d} r_i^2 K_d(\mathbf{r}) d\mathbf{r} \right. \\ &+ 2 \sum_{i=1}^d \sum_{j=1}^d \sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{\partial^2 \psi_\ell(\mathbf{x}_0)}{\partial x_i \partial x_j} \right| \int_{\mathbb{R}^d} |r_i| |r_j| K_d(\mathbf{r}) d\mathbf{r} \right\}. \end{split}$$

Assumptions K(a) and D(a) complete the proof of the lemma.

Lemma A.2: Under Assumptions K(a)-K(c), H(c), and D(a)-D(d), we have for $\ell = 1, 2$

$$\sup_{\mathbf{x}\in\mathcal{C}}|\Lambda_{2,\ell}(\mathbf{x})| = O_{\text{a.s.}}\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right) \quad as \ n \to \infty.$$

Proof: Since C is a compact set, then it can be covered by a finite number ω_n of balls $\mathcal{B}_k(\mathbf{x}_k, a_n^d)$ centred at $\mathbf{x}_k = (x_{1,k}, \ldots, x_{d,k}), 1 \le k \le \omega_n$, where ω_n and a_n^d satisfy

$$\omega_n \leq M a_n^{-d}$$
 and $a_n^d = h_n^{d(1+\frac{1}{2\gamma})} n^{-\frac{1}{2\gamma}}$,

with *M* is a positive constant and γ is the Lipschitz condition in Assumption **K**(c). Then for all $\mathbf{x} \in C$, there exists a ball B_k that contains \mathbf{x} such that

$$\|\mathbf{x} - \mathbf{x}_k\| \le a_n^d. \tag{A3}$$

For $\ell = 1, 2$, we have

$$\sup_{\mathbf{x}\in\mathcal{C}}\left|\Lambda_{2,\ell}(\mathbf{x})\right|$$

$$= \sup_{\mathbf{x}\in\mathcal{C}} \left| (\widetilde{\psi}_{\ell}(\mathbf{x}) - \widetilde{\psi}_{\ell}(\mathbf{x}_{k})) + (\mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_{k})] - \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x})]) + (\widetilde{\psi}_{\ell}(\mathbf{x}_{k}) - \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_{k})]) \right|$$

$$\leq \max_{1 \leq k \leq \omega_{n}} \sup_{\mathbf{x}\in\mathcal{C}} \left| \widetilde{\psi}_{\ell}(\mathbf{x}) - \widetilde{\psi}_{\ell}(\mathbf{x}_{k}) \right| + \max_{1 \leq k \leq \omega_{n}} \sup_{\mathbf{x}\in\mathcal{C}} \left| \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_{k})] - \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x})] \right|$$

$$+ \max_{1 \leq k \leq \omega_{n}} \left| \widetilde{\psi}_{\ell}(\mathbf{x}_{k}) - \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_{k})] \right|$$

$$=: I_{1} + I_{2} + I_{3}.$$

We start by treating the first term I_1 . Under Assumptions **D**(a) and **K**(c), we have

$$\begin{split} \left| \widetilde{\psi}_{\ell}(\mathbf{x}) - \widetilde{\psi}_{\ell}(\mathbf{x}_{k}) \right| &= \left| \frac{\alpha}{nh_{n}^{d}} \sum_{i=1}^{n} \frac{\delta_{i} Z_{i}^{-\ell}}{L(Z_{i}) G(\overline{Z}_{i})} \left[K_{d} \left(\frac{\mathbf{x} - \mathbf{X}_{i}}{h_{n}} \right) - K_{d} \left(\frac{\mathbf{x}_{k} - \mathbf{X}_{i}}{h_{n}} \right) \right] \right| \\ &= \left| \frac{\alpha}{nh_{n}^{d}} \sum_{i=1}^{n} \frac{Y_{i}^{-\ell}}{L(Y_{i}) \overline{G}(Y_{i})} \left[K_{d} \left(\frac{\mathbf{x} - \mathbf{X}_{i}}{h_{n}} \right) - K_{d} \left(\frac{\mathbf{x}_{k} - \mathbf{X}_{i}}{h_{n}} \right) \right] \right| \\ &\leq \frac{C}{nh_{n}^{d} L(a_{H}) \overline{G}(b_{H})} \sum_{i=1}^{n} \left| K_{d} \left(\frac{\mathbf{x} - \mathbf{X}_{i}}{h_{n}} \right) - K_{d} \left(\frac{\mathbf{x}_{k} - \mathbf{X}_{i}}{h_{n}} \right) \right| \\ &\leq \frac{C}{h_{n}^{d} L(a_{H}) \overline{G}(b_{H})} \left\| \left| \frac{\mathbf{x} - \mathbf{X}_{1}}{h_{n}} - \frac{\mathbf{x}_{k} - \mathbf{X}_{1}}{h_{n}} \right\| \right|^{\gamma} \\ &\leq \frac{C \parallel \mathbf{x} - \mathbf{x}_{k} \parallel^{\gamma}}{h_{n}^{d+\gamma}}. \end{split}$$

Then, from (A3), we get

$$\sup_{\mathbf{x}\in\mathcal{C}}\left|\widetilde{\psi}_{\ell}(\mathbf{x})-\widetilde{\psi}_{\ell}(\mathbf{x}_{k})\right|\leq\frac{Ca_{n}^{d\gamma}}{h_{n}^{d+\gamma}}=\frac{C}{\sqrt{nh_{n}^{d}}}h_{n}^{\gamma(d-1)}.$$

Hence,

$$I_1 = \max_{1 \le k \le \omega_n} \sup_{\mathbf{x} \in \mathcal{C}} \left| \widetilde{\psi}_{\ell}(\mathbf{x}) - \widetilde{\psi}_{\ell}(\mathbf{x}_k) \right| = O\left(\frac{1}{\sqrt{nh_n^d}}\right).$$

In the same way as for I_1 , we obtain

$$I_2 = \max_{1 \le k \le \omega_n} \sup_{\mathbf{x} \in \mathcal{C}} \left| \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x})] - \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_k)] \right| = O\left(\frac{1}{\sqrt{nh_n^d}}\right).$$

For I_3 , we use the Fuk-Nagaev exponential inequality (Ferraty and Vieu 2006), which states that if $\{U_i, i \ge 1\}$ is a sequence of rvs, with strong mixing coefficient $\alpha(n) = O(n^{-\nu})$, where $\nu > 1$, and for all $n \in \mathbb{N}$ and $i \in \mathbb{N}$, $1 \le i \le n$, $|U_i| < \infty$, then for each $\varepsilon > 0$ and q > 1, we have

$$\mathbf{P}\left\{\left|\sum_{i=1}^{n} U_{i}\right| > \varepsilon\right\} \le C\left(1 + \frac{\varepsilon^{2}}{qS_{n}^{2}}\right)^{-\frac{q}{2}} + nCq^{-1}\left(\frac{q}{\varepsilon}\right)^{\nu+1},\tag{A4}$$

where $S_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\mathbf{Cov}(U_i, U_j)|$. For that, we set

$$U_{i,\ell}(\mathbf{x}_k) = \frac{\alpha Z_i^{-\ell} \delta_i}{L(Z_i) \bar{G}(Z_i)} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n}\right) - \mathbf{E}\left[\frac{\alpha Z_1^{-\ell} \delta_1}{L(Z_1) \bar{G}(Z_1)} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_n}\right)\right], \quad \text{for } \ell = 1, 2.$$

20 🔄 N. BAYARASSOU ET AL.

Clearly, we have

$$\Lambda_{2,\ell}(\mathbf{x}_k) = \frac{1}{nh_n^d} \sum_{i=1}^n U_{\ell,i}(\mathbf{x}_k).$$

Now, we have to calculate

$$S_n^2 = \sum_{i=1}^n \sum_{j=1}^n |\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))|$$

= $n\mathbf{Var}(U_{1,\ell}(\mathbf{x}_k)) + \sum_{\substack{i=1\\i\neq j}}^n \sum_{\substack{j=1\\i\neq j}}^n |\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))|$
=: $CV_1 + CV_2$.

On the one hand, we have

$$\begin{aligned} &\mathsf{Var}(U_{1,\ell}(\mathbf{x}_k)) \\ &= \mathsf{Var}\left[\frac{\alpha_1 Z_1^{-\ell} \delta_1}{L(Z_1) \bar{G}(Z_1)} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_n}\right)\right] \\ &= \mathsf{E}\left[\frac{\alpha^2 Z_1^{-2\ell} \delta_1}{L^2(Z_1) \bar{G}^2(Z_1)} K_d^2\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_n}\right)\right] - \mathsf{E}^2\left[\frac{\alpha Z_1^{-\ell} \delta_1}{L(Z_1) \bar{G}(Z_1)} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_n}\right)\right] \\ &=: \mathcal{R}_1 - \mathcal{R}_2. \end{aligned}$$

From (11) and using a change of variable, a Taylor expansion and under Assumptions K(b) and D(c), we obtain

$$\mathcal{R}_{1} = \mathbf{E} \left[\frac{\alpha^{2} Z_{1}^{-2\ell} \delta_{1}}{L^{2}(Z_{1}) \bar{G}^{2}(Z_{1})} K_{d}^{2} \left(\frac{\mathbf{x}_{k} - \mathbf{X}_{1}}{h_{n}} \right) \right]$$

$$= \int_{\mathbb{R}^{d}} \int_{a_{H}}^{y} \frac{\alpha^{2} t^{-2\ell}}{L^{2}(t) \bar{G}^{2}(t)} K_{d}^{2} \left(\frac{\mathbf{x}_{k} - \mathbf{u}}{h_{n}} \right) d\mathbf{H}_{1}(\mathbf{u}, t)$$

$$= \int_{\mathbb{R}^{d}} \int_{a_{H}}^{y} \frac{\alpha t^{-2\ell}}{L(t) \bar{G}(t)} K_{d}^{2} \left(\frac{\mathbf{x}_{k} - \mathbf{u}}{h_{n}} \right) f_{\mathbf{X}, Y}(\mathbf{u}, t) \, \mathbf{d}\mathbf{u} \, dt$$

$$\leq \int_{\mathbb{R}^{d}} K_{d}^{2} \left(\frac{\mathbf{x}_{k} - \mathbf{u}}{h_{n}} \right) \mu_{\ell}(\mathbf{u}) \, \mathbf{d}\mathbf{u}$$

$$= h_{n}^{d} \int_{\mathbb{R}^{d}} K_{d}^{2}(\mathbf{s}) \mu_{\ell}(\mathbf{x}_{k} - \mathbf{s}h_{n}) \, \mathbf{d}\mathbf{s}$$

$$= O(h_{n}^{d}). \tag{A5}$$

For \mathcal{R}_2 , we have

$$\begin{split} \sqrt{\mathcal{R}_2} &= \mathbf{E} \left[\frac{\alpha Z_1^{-\ell} \delta_1}{L(Z_1) \bar{G}(Z_1)} K_d \left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_n} \right) \right] \\ &= \int_{\mathbb{R}^d} \int_{a_H}^y \frac{\alpha t^{-\ell}}{L(t) \bar{G}(t)} K_d \left(\frac{\mathbf{x}_k - \mathbf{u}}{h_n} \right) d\mathbf{H}_1(\mathbf{u}, t) \\ &= \int_{\mathbb{R}^d} \int_{a_H}^y t^{-\ell} K_d \left(\frac{\mathbf{x}_k - \mathbf{u}}{h_n} \right) f_{\mathbf{X}, Y}(\mathbf{u}, t) \, \mathbf{du} \, \mathbf{dt} \end{split}$$

$$= \int_{\mathbb{R}^d} K_d\left(\frac{\mathbf{x}_k - \mathbf{u}}{h_n}\right) \psi_\ell(\mathbf{u}) \, \mathbf{du}$$
$$= h_n^d \int_{\mathbb{R}^d} K_d(\mathbf{s}) \psi_\ell(\mathbf{x}_k - \mathbf{sh}_n) \, \mathbf{ds},$$

by a Taylor expansion around \mathbf{x}_k and under Assumptions $\mathbf{K}(\mathbf{a})$, $\mathbf{D}(\mathbf{a})$, we get

$$\mathcal{R}_2 = O(h_n^{2d}). \tag{A6}$$

Then, from (A5) and (A6), we obtain

$$CV_1 = n(\mathcal{R}_1 - \mathcal{R}_2) = O(nh_n^d). \tag{A7}$$

On the other hand, under Assumption $\mathbf{D}(\mathbf{b})$ and by a change of variable, we have

$$\begin{aligned} |\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| \\ &= |\mathbf{E}(U_{i,\ell}(\mathbf{x}_k)U_{j,\ell}(\mathbf{x}_k))| \\ &= \left| \mathbf{E} \left[\frac{\alpha Z_i^{-\ell} \delta_i}{L(Z_i)\bar{G}(Z_i)} K_d \left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n} \right) \frac{\alpha Z_j^{-\ell} \delta_j}{L(Z_j)\bar{G}(Z_j)} K_d \left(\frac{\mathbf{x}_k - \mathbf{X}_j}{h_n} \right) \right] \\ &- \mathbf{E} \left[\frac{\alpha Z_i^{-\ell} \delta_i}{L(Z_i)\bar{G}(Z_i)} K_d \left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n} \right) \right] \mathbf{E} \left[\frac{\alpha Z_j^{-\ell} \delta_j}{L(Z_j)\bar{G}(Z_j)} K_d \left(\frac{\mathbf{x}_k - \mathbf{X}_j}{h_n} \right) \right] \right| \\ &\leq C h_n^{2d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_d(\mathbf{s}) K_d(\mathbf{t}) |\Upsilon_{ij}(\mathbf{x}_k - \mathbf{s}h_n, \mathbf{x}_k - \mathbf{t}h_n) \\ &- \Upsilon_i(\mathbf{x}_k - \mathbf{s}h_n) \Upsilon_j(\mathbf{x}_k - \mathbf{t}h_n) | \, \mathbf{ds} \, \mathbf{dt}. \end{aligned}$$

Assumption D(d) gives

$$|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| = O(h_n^{2d}).$$
(A8)

Then, to evaluate the term CV_2 , following Masry (1986), we divide the set $\{(i, j)/1 \le |i - j| \le n\}$ into two sub-sets E_1 and E_2 by introducing a sequence of integer $\beta_n = o(n)$, such that

$$E_1 = \{(i,j)/1 \le |i-j| \le \beta_n\}$$
 and $E_2 = \{(i,j)/\beta_n + 1 \le |i-j| \le n-1\}.$

That is

$$CV_{2} = \sum_{i=1}^{n} \sum_{j \in E_{1}} |\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_{k}), U_{j,\ell}(\mathbf{x}_{k}))| + \sum_{i=1}^{n} \sum_{j \in E_{2}} |\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_{k}), U_{j,\ell}(\mathbf{x}_{k}))|$$

=: $CV_{21} + CV_{22}$.

From (A8), we get

$$CV_{21} = \sum_{i=1}^{n} \sum_{j \in E_1} |\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| = C \sum_{i=1}^{n} \sum_{1 \le |i-j| \le \beta_n} h_n^{2d} = O(nh_n^{2d}\beta_n).$$

For CV_{22} , we use the modified inequality of Davydov for mixing processes (see Rio 2000). This leads, for all $i \neq j$, to

$$|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| \le C\alpha(|i-j|).$$

22 🕒 N. BAYARASSOU ET AL.

Then, we get

$$CV_{22} = \sum_{i=1}^{n} \sum_{j \in E_2} |\operatorname{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| \le C \sum_{i=1}^{n} \sum_{\beta_n < |i-j| < n} \alpha(|i-j|)$$
$$\le Cn^2 \alpha(\beta_n)$$
$$= O(n^2 \beta_n^{-\nu}).$$

By choosing $\beta_n = [h_n^{-d}]$, we obtain

$$CV_2 = CV_{21} + CV_{22} = O(nh_n^d) + O(n^2h_n^{d\nu}).$$
 (A9)

Finally, from (A7) and (A9), we get

$$S_n^2 = CV_1 + CV_2 = O(nh_n^d) + O(n^2h_n^{dv}).$$

Now, we are ready to apply the Fuk-Nagaev exponential inequality given in (A4). For $\varepsilon > 0$, we have

$$\mathbf{P}\{|\widetilde{\psi}_{\ell}(\mathbf{x}_{k}) - \mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_{k})]| > \varepsilon\}$$

$$= \mathbf{P}\left\{\left|\sum_{i=1}^{n} U_{i,\ell}(\mathbf{x}_{k})\right| > nh_{n}^{d}\varepsilon\right\}$$

$$\leq C\left(1 + C\frac{\varepsilon^{2}nh_{n}^{d}}{q(1 + nh_{n}^{d(\nu-1)})}\right)^{-\frac{q}{2}} + nCq^{-1}\left(\frac{q}{\varepsilon nh_{n}^{d}}\right)^{\nu+1}.$$
(A10)

By taking $\varepsilon = \varepsilon_0 \left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)} \log n} \right) =: \varepsilon_n$ for all $\varepsilon_0 > 0$, (A10) becomes $\mathbf{P}\{|\widetilde{\psi}_\ell(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_\ell(\mathbf{x}_k)]| > \varepsilon_n\}$

$$\mathbf{P}\{|\psi_{\ell}(\mathbf{x}_{k}) - \mathbf{E}[\psi_{\ell}(\mathbf{x}_{k})]| > \varepsilon_{n}\}$$

$$\leq C\left\{\left(1 + C\frac{\varepsilon_{0}^{2}\log n}{q}\right)^{-\frac{q}{2}} + nq^{-1}\left(\frac{q}{\varepsilon_{0}\left(\sqrt{\frac{\log n}{nh_{n}^{d}}} + \sqrt{h_{n}^{d(\nu-2)}\log n}\right)nh_{n}^{d}}\right)^{\nu+1}\right\}$$

$$=: C(\varepsilon_{1} + \varepsilon_{2}).$$

If we replace *q* by $(\log n)^{1+b}$, with b > 0 and use a Taylor expansion of $\log(1 + x)$, we get

$$\varepsilon_{1} = (1 + C\varepsilon_{0}^{2}(\log n)^{-b})^{-\frac{(\log n)^{1+b}}{2}}$$
$$= \exp\left(-\frac{(\log n)^{1+b}}{2}\log(1 + C\varepsilon_{0}^{2}(\log n)^{-b})\right)$$
$$\simeq n^{-C\frac{\varepsilon_{0}^{2}}{2}}.$$

For the same choice of ε and q, we have

$$\varepsilon_2 \simeq n(\log n)^{\nu(1+b)} \varepsilon_0^{-(\nu+1)} (nh_n^d \log n)^{-\frac{(\nu+1)}{2}}.$$

Now, we can write

$$\mathbf{P}\left\{\max_{1\leq k\leq\omega_{n}}|\widetilde{\psi}_{\ell}(\mathbf{x}_{k})-\mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_{k})]|>\varepsilon_{n}\right\} \\
\leq \sum_{i=1}^{\omega_{n}}\mathbf{P}\{|\widetilde{\psi}_{\ell}(\mathbf{x}_{k})-\mathbf{E}[\widetilde{\psi}_{\ell}(\mathbf{x}_{k})]|>\varepsilon_{n}\}$$

$$\leq Ma_{n}^{-d} \left\{ Cn^{-C\frac{\varepsilon_{0}^{2}}{2}} + n(\log n)^{\nu(1+b)}\varepsilon_{0}^{-(\nu+1)}(nh_{n}^{d}\log n)^{-\frac{(\nu+1)}{2}} \right\}$$

$$\leq MCh_{n}^{-d(1+\frac{1}{2\gamma})}n^{\frac{1}{2\gamma}-C\frac{\varepsilon_{0}^{2}}{2}} + MC\varepsilon_{0}^{-(\nu+1)}n^{1+\frac{1}{2\gamma}}h_{n}^{-d(1+\frac{1}{2\gamma})}(\log n)^{\nu(1+b)}(nh_{n}^{d}\log n)^{-\frac{\nu+1}{2}}$$

$$=: MCA_{1} + MC\varepsilon_{0}^{-(\nu+1)}A_{2}.$$
(A11)

We have from Assumption H(c)

$$A_{2} = (\log n)^{\nu(1+b) - \frac{\nu+1}{2}} n^{1 - \frac{\nu+1}{2} + \frac{1}{2\gamma}} h_{n}^{-d\left(1 + \frac{1}{2\gamma} + \frac{\nu+1}{2}\right)}$$

$$\leq C(\log n)^{\nu(1+b) - \frac{\nu+1}{2}} n^{1 - \frac{\nu+1}{2} + \frac{1}{2\gamma}} n^{-\frac{(3-\nu)}{2} - \theta d\left(\frac{\gamma(\nu+1) + 2\gamma + 1}{2\gamma}\right)}$$

$$= C(\log n)^{\nu(1+b) - \frac{\nu+1}{2}} n^{-1 + \frac{1 - \theta d(\gamma(\nu+3) + 1)}{2\gamma}}.$$

Then, for an appropriate choice of ε_0 and θ , A_1 and A_2 are the general terms of a convergent series. Finally, applying Borel-Cantelli's lemma to (A11) gives the result.

Lemma A.3: Under Assumptions K, H(a), and D(b), we have for $\ell = 1, 2$

$$\sup_{\mathbf{x}\in\mathcal{C}} |\Lambda_{1,\ell}(\mathbf{x})| = O_{\text{a.s.}}\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{as } n \to \infty.$$

Proof: We have

$$\begin{split} \Lambda_{1,\ell}(\mathbf{x}) &= \frac{1}{nh_n^d} \sum_{i=1}^n \left(\frac{\alpha_n Z_i^{-\ell} \delta_i}{L_n(Z_i) G_n(\overline{Z}_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) - \frac{\alpha Z_i^{-\ell} \delta_i}{L(Z_i) G(\overline{Z}_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \right) \\ &= \frac{1}{nh_n^d} \sum_{i=1}^n Y_i^{-\ell} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \left(\frac{\alpha_n}{L_n(Z_i) \overline{G}_n(Z_i)} - \frac{\alpha}{L(Z_i) \overline{G}(Z_i)}\right). \end{split}$$

By replacing α and α_n , by their expressions defined in (5) and (6), respectively, we get

$$\begin{split} \sup_{\mathbf{x}\in\mathcal{C}} |\Lambda_{1,\ell}(\mathbf{x})| &\leq \sup_{a_H \leq t \leq b_H} \left| \frac{\bar{F}_n(t)}{C_n(t)} - \frac{\bar{F}(t)}{C(t)} \right| \times \sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{1}{nh_n^d} \sum_{i=1}^n Y_i^{-\ell} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) \right| \\ &\leq \left\{ \sup_{a_H \leq t \leq b_H} \left| \frac{F_n(t) - F(t)}{C_n(t)} \right| + \sup_{a_H \leq t \leq b_H} \left| \frac{\bar{F}(t)}{C_n(t)C(t)} (C_n(t) - C(t)) \right| \right\} \\ &\times \sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{1}{nh_n^d} \sum_{i=1}^n Y_i^{-\ell} K_d\left(\frac{\mathbf{x} - X_i}{h_n}\right) \right| \\ &\leq \left\{ \frac{\sup_{a_H \leq t \leq b_H} |F_n(t) - F(t)|}{\inf_{a_H \leq t \leq b_H} |C_n(t) - C(t))|} \\ &+ \frac{\sup_{a_H \leq t \leq b_H} |C(t)| - \sup_{a_H \leq t \leq b_H} |(C_n(t) - C(t))|}{\inf_{a_H \leq t \leq b_H} |C(t)| - \sup_{a_H \leq t \leq b_H} |C(t)| - \sup_{a_H \leq t \leq b_H} |(C_n(t) - C(t))|} \right\} \end{split}$$

24 🛞 N. BAYARASSOU ET AL.

$$\times \sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{1}{nh_n^d} \sum_{i=1}^n Y_i^{-\ell} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \right|.$$

From (3), we have $C(t) \ge \alpha^{-1}L(a_H)\overline{H}(b_H) > 0$ for all $a_H \le t \le b_H$, and following Chen and Dai (2003), we have

$$\sup_{a_H \le t \le b_H} |F_n(t) - F(t)| = O_{\text{a.s.}}\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{as } n \to \infty$$

and

$$\sup_{a_H \le t \le b_H} |C_n(t) - C(t)| = O_{\text{a.s.}}\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{as } n \to \infty$$

Furthermore, under Assumptions K, H(a) and D(b) we get the result.

Proof of Theorem 4.1: By triangle inequality, (A1) becomes

$$\begin{split} \sup_{\mathbf{x}\in\mathcal{C}} |\widehat{m}_n(\mathbf{x}) - m(\mathbf{x})| &\leq \frac{1}{\inf_{\mathbf{x}\in\mathcal{C}} |\widehat{\psi}_2(\mathbf{x})|} \left\{ \sup_{\mathbf{x}\in\mathcal{C}} \left[|\Lambda_{1,1}(\mathbf{x})| + |\Lambda_{2,1}(\mathbf{x})| + |\Lambda_{3,1}(\mathbf{x})| \right] \\ &+ \sup_{\mathbf{x}\in\mathcal{C}} |m(\mathbf{x})| \left[|\Lambda_{1,2}(\mathbf{x})| + |\Lambda_{2,2}(\mathbf{x})| + |\Lambda_{3,2}(\mathbf{x})| \right] \right\}. \end{split}$$

Then, the proof of Theorem 4.1 is completed by Lemmas A.1–A.3.