Journal of Nonparametric Statistics

ISSN: 1048-5252 (Print) 1029-0311 (Online) Journal homepage: http://www.tandfonline.com/loi/gnst20

Convergence rate of the kernel regression estimator for associated and truncated data

Z. Guessoum & F. Hamrani

To cite this article: Z. Guessoum & F. Hamrani (2017) Convergence rate of the kernel regression estimator for associated and truncated data, Journal of Nonparametric Statistics, 29:2, 425-446, DOI: 10.1080/10485252.2017.1303059

To link to this article: http://dx.doi.org/10.1080/10485252.2017.1303059

Published online: 20 Mar 2017.

JOURNAL OF NONPARAMETRIC STATISTICS, 2017
VOL. 29, NO. 2, 425–446
http://dx.doi.org/10.1080/10485252.2017.1303059

Taylor & Francis
Taylor & Francis Group

REVIEW

Convergence rate of the kernel regression estimator for associated and truncated data

Z. Guessoum[a] and F. Hamrani[b]

[a]Lab. M.S.T.D., Faculté de Math., U.S.T.H.B., Algiers, Algeria; [b]Faculté de Math., U.S.T.H.B., Algiers, Algeria

ABSTRACT
This paper studies the behaviour of the kernel estimator of the regression function for associated data in the random left truncated model. The uniform strong consistency rate over a real compact set of the estimate is established. The finite sample performance of the estimator is investigated through extensive simulation studies.

ARTICLE HISTORY
Received 9 January 2015
Accepted 17 December 2016

KEYWORDS
Association; kernel estimator; nonparametric regression; random left-truncation (RLT) model; rate of convergence; strong consistency

1. Introduction

Let Y be a real random variable (rv) of interest with a continuous distribution function (df) F, and X an $\mathbb{R}^d$-valued random vector of covariates with a joint df V and a joint density v. We wish to estimate Y given X using a regression-based approach. This means looking for a function which realises the minimum of the mean squared error (MSE). The regression function that achieves this minimum is defined on $\mathbb{R}^d$ by the conditional expectation of Y given $X = x$ that is

$$m(x) := \mathbb{E}(Y \mid X = x).$$

In the case of complete data there is a vast literature devoted to the study of the non-parametric kernel estimator of $m(.)$. Far from being exhaustive, we can quote Walk (2005) and the references therein.

Nevertheless, in many survival practical applications, it happens that one is not able to observe a subject's entire lifetime. The subject may leave the study, may survive to the closing date, or may enter the study at some time after its lifetime has started. The most current forms of such incomplete data are censorship and truncation. The model studied here is based on the random left truncated (RLT) data, where the observation (X, Y) is interfered by another independent rv T such that the random quantities Y, X and T are observable only if $Y \geq T$.

CONTACT Z. Guessoum zguessoum@usthb.dz, z0guessoum@hotmail.com Lab. M.S.T.D., Faculté de Math., U.S.T.H.B., BP 32, El Alia, 16111 Algiers, Algeria

The RLT model is originally appeared in astronomy and economics (Woodroofe 1985; Chen, Chao, and Lo 1995), then extended to several domains as epidemiology, demographics, reliability testing and actuarial (Wang, Jewell, and Tsai 1986; Tsai, Jewell, and Wang 1987). For example, in an AIDS study (Kalbfleisch and Lawless 1989), let X be the infection time where 1 represents January 1978 and let T be the incubation time in months for people who were infected by contaminated blood transfusions and developed AIDS by 1 July 1986. Since the total study period is 102 months only individuals with $X+T<102$ were included in the sample. Then, letting $Y = 102-X$ yields the model described: (Y, T) is observed only if $T < Y$. Another example is that of a retirement centre (Klein and Moeschberger 2003), where subjects are observed only if they live long enough to enter the centre. The lifetime Y is then left truncated by the retirement house entry age, T. People who enter the centre earlier may get better medical attention and therefore live longer. On the other hand, people with poor health and shorter expected lifetime may retire earlier.

In the i.i.d. case, Ould Saïd and Lemdani (2006) constructed a nonparametric kernel estimator of the regression function $m(.)$ under RLT model. They established its strong uniform consistency as well as its asymptotic properties. The aim of this paper is to extend some of their results to the case of dependent data. Two types of dependency are widely used in literature: mixing and association.

The α-mixing condition, also called strong mixing, is the weakest among mixing conditions known in the literature. Many stochastic processes satisfy the α-mixing condition, see, for example, Doukhan (1994) and Carrasco, Chernov, Florens, and Ghysels (2007).

For truncated data, under α-mixing condition, the strong convergence of the estimator of the regression function defined by Ould Saïd and Lemdani (2006) is treated in Liang, Li, and Qi (2009) and its asymptotic normality is established later in Liang (2011).

In this paper, we focus on the concept of association which has been introduced and defined by Esary, Proschan, and Walkup (1967). A set of finite family of rv's $Y = (Y_1 \ldots, Y_N)$ is said associated if

$$\mathrm{cov}(f(Y), g(Y)) \geq 0$$

for all nondecreasing functions f and g from $\mathbb{R}^N$ to $\mathbb{R}$ for which this covariance exists. An infinite family is said to be associated if every finite subfamily is associated.

It is of interest to note that association and mixing define two distinct but not disjoint classes of processes (see Doukhan and Louhichi 1999). In the linear time series framework, the sequence X_k defined by $X_k = \sum_{j=0}^{\infty} a_j v_{k-j}$ (where $(v_i)_{-\infty<i<+\infty}$ is a sequence of i.i.d. rvs with mean zero and variance σ^2) is associated if $a_j \geq 0$. On the other hand Pham and Tran (1985) showed that $(X_k)_{k\geq 0}$ is α-mixing under suitable conditions on a_j. In particular, Andrew (1984) showed that when $(v_i)_{-\infty<i<+\infty}$ is a sequence of i.i.d. Bernoulli rvs and $a_j = \varsigma$ with $0 < \varsigma < \frac{1}{2}$, the sequence $(X_k)_{k\geq 0}$ is not α-mixing, whereas it is still associated.

For more details on the concept of association, we refer the reader to Bulinski and Shashkin (2007). In that book, the reader can find some results and examples related to associated random sequences and random fields.

In the complete associated data case, there is a vast literature devoted to the study of the nonparametric kernel estimation and many papers deal with density estimation. We cite only a few of them as Bagai and Prakasa Rao (1995) who have obtained the strong uniform consistency of the kernel density estimator. They also proposed an estimator of the survival function and established its consistency. Roussas (2000) has established the asymptotic normality of the usual kernel estimate of the marginal probability density function. Douge (2007) has stated a new exponential inequality and has derived a uniform almost sure rate of convergence over compact sets for the kernel density estimator.

In the incomplete data case, for associated rv there are no much works done for this kind of model. Under random right censoring one, we can cite Cai and Roussas (1998) who have established uniform strong consistency and asymptotic normality of the Kaplan–Meier estimator. Ferrani, Ould Saïd, and Tatachak (2016) established the strong uniform consistency of the kernel estimator of the underlying density function and the almost sure convergence of a smooth kernel mode estimator under right censored model. Under RLT model, Guessoum, Sadki, and Tatachak (2012) established the strong uniform convergence with a rate of the Lynden-Bell estimator.

To the best of our knowledge, the problem of estimating the regression function under association and truncation has not been addressed in the literature. The goal of this paper is to establish the strong uniform convergence with a rate for the kernel regression estimate, under RLT model when the variable of interest Y and the multivariate covariates X are associated.

This paper is organised as follows. In Section 2, we recall some results stated in RLT model with the estimators studied in the current work. In Section 3, we list the assumptions and give our main results. In Section 4, some simulations are given. The proofs of the main results are detailed in the appendix with some preliminary lemmas.

2. Model and estimators

Let Y be a bounded real rv defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with a continuous df F and T a real rv independent from Y , defined on the same probability space with a Lipschitz df G.

In the following, $\{Y_i;\ i = 1, \dots, N\}$ and $\{T_i;\ i = 1, \dots, N\}$ denote, respectively, a strictly stationary associated sequence and an i.i.d. sequence of N copies of Y and T, where the sample size N is fixed but unknown.

In the RLT model, as mentioned above, the rv of interest Y and the truncated rv T are observable only when $Y \geq T$, whereas nothing is observed if $Y < T$. Then without possible confusion, we still denote $\{(Y_i, T_i);\ i = 1, \dots, n\}$ $(n \leq N)$ the actually observed sample. Note that as the original sequence of interest is associated, the observed one is associated (by property *(P1)* in Esary et al. 1967) and the observed sequence of truncation is also i.i.d. (by Proposition 2.1 in Lemdani 2007).

Let $\alpha := \mathbb{P}(Y \geq T)$ be the probability to observe at least one pair from (Y, T). As a consequence of truncation, the size n of the actually observed sample is a $\mathrm{Bin}(N, \alpha)$ rv. We suppose hereafter that $\alpha > 0$ otherwise no data can be observed. Since N is unknown and n known (although random), our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the N-sample) but will involve the probability $\mathbf{P}$ (related to the n-sample) defined as $\mathbf{P}(.) = \mathbb{P}(.|Y \geq T)$. In the same way $\mathbb{E}$ and $\mathbf{E}$ will denote the expectation operators related to $\mathbb{P}$ and $\mathbf{P}$, respectively.

Following Stute (1993), the joint **P**-distribution of an observed (Y, T) is given by

$$\begin{aligned} H^*(y,t) &:= \mathbf{P}\{Y \leq y, T \leq t\} \\ &= \mathbb{P}\{Y \leq y, T \leq t | Y \geq T\} \\ &= \frac{1}{\alpha}\int_{-\infty}^{y} G(t \wedge u)\, \mathrm{d}F(u), \end{aligned}$$

where $t \wedge u := \min(t, u)$. The marginal distributions of Y and T, respectively, are defined by

$$F^*(y) := H^*(y,\infty) = \frac{1}{\alpha}\int_{-\infty}^{y} G(u)\, \mathrm{d}F(u) \quad \text{and}$$
$$G^*(t) := H^*(\infty,t) = \frac{1}{\alpha}\int_{-\infty}^{\infty} G(t \wedge u)\, \mathrm{d}F(u),$$

which can be estimated by

$$F_n^*(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{Y_i \leq y\}} \quad \text{and} \quad G_n^*(t) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{T_i \leq t\}},$$

respectively, where $\mathbb{I}_A$ denotes the indicator function of the set A. Let $C(.)$ be defined by

$$\begin{aligned} C(y) &:= \mathbf{P}\{T \leq y \leq Y\} = \mathbb{P}\{T \leq y \leq Y\,|Y \geq T\} \\ &= \frac{1}{\alpha}G(y)(1 - F(y)) = G^*(y) - F^*(y), \end{aligned}$$

with empirical estimator

$$C_n(y) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{T_i \leq y \leq Y_i\}} = G_n^*(y) - F_n^*(y).$$

The well-known nonparametric estimators of F and G in RLT model, proposed by Lynden-Bell (1971) are

$$\begin{aligned} F_n(y) &:= 1 - \prod_{i:Y_i \leq y}\left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)}\right], \\ G_n(t) &:= \prod_{i:T_i > t}\left[\frac{nC_n(T_i) - 1}{nC_n(T_i)}\right]. \end{aligned} \tag{1}$$

Here, for any df W we define

$$a_W := \inf\{u : W(u) > 0\} \quad \text{and} \quad b_W := \sup\{u : W(u) < 1\} \tag{2}$$

as the endpoints of the W support. Woodroofe (1985) pointed out that F and G can be completely estimated only if

$$a_G \leq a_F, \quad b_G \leq b_F \quad \text{and} \quad \int_{a_F}^{\infty} \frac{\mathrm{d}F}{G} < \infty.$$

For the unknown probability $\alpha := \mathbb{P}(Y \geq T)$, which cannot be classically estimated by $\hat{\alpha}_n := n/N$ since N is unknown, He and Yang (1998) proposed the estimator

$$\alpha_n(y) := \frac{G_n(y)(1 - F_n(y))}{C_n(y)} =: \alpha_n. \tag{3}$$

The authors proved that α_n does not depend on y and its value can be obtained for any y such that $C_n(y) \neq 0$.

Now, let $\{(X_i, Y_i, T_i);\ i = 1, \ldots, n\}$ be the n triplets observed among the N ones such that $Y_i \geq T_i$ and T independent of (X, Y) where $\{(X_i, Y_i);\ i = 1, \ldots, n\}$ are supposed strictly stationary and associated. Then we first consider the joint $\mathbf{P}$-distribution of (X, Y, T)

$$\begin{aligned} L^*(x, y, t) &:= \mathbf{P}\{X \leq x, Y \leq y, T \leq t\} \\ &= \mathbb{P}\{X \leq x, Y \leq y, T \leq t | Y \geq T\} \\ &= \frac{1}{\alpha} \int_{u \leq x} \int_{a_G \leq w \leq y} G(w \wedge t) F(\mathrm{d}u, \mathrm{d}w), \end{aligned}$$

the joint $\mathbf{P}$-df $F^*_{X,Y}(.,.)$ of an observed (X, Y) is given by

$$F^*_{X,Y}(x, y) := L^*(x, y, \infty) = \frac{1}{\alpha} \int_{u \leq x} \int_{a_G \leq w \leq y} G(w) F_{X,Y}(\mathrm{d}u, \mathrm{d}w),$$

which gives

$$F_{X,Y}(\mathrm{d}x, \mathrm{d}y) = \frac{\alpha}{G(y)} F^*_{X,Y}(\mathrm{d}x, \mathrm{d}y) \quad \text{for } y \geq a_G. \tag{4}$$

By integrating over y, we get the df of X

$$V(x) = \alpha \int_{u \leq x} \int_{y \geq a_G} \frac{1}{G(y)} F^*(\mathrm{d}u, \mathrm{d}y).$$

Following Ould Saïd and Lemdani (2006), an estimator of $V(x)$ is given by

$$\tilde{V}_n(x) := \frac{\alpha}{n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} \mathbb{I}_{\{X_i \leq x\}}. \tag{5}$$

Note that in Equation (5) and in the sequel, the sum is taken only over the i's such that $G(Y_i) \neq 0$. Thus, Equation (5) yields the kernel density estimator

$$\tilde{v}_n(x) := \frac{\alpha}{nh_n^d} \sum_{i=1}^{n} \frac{1}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right), \tag{6}$$

where $K_d : \mathbb{R}^d \to \mathbb{R}$ is a smooth kernel function and h_n is a positive bandwidth sequence that tends to zero as $n \to \infty$. Observe that the regression function $m(x)$ can be written as

$$m(x) = \frac{\psi(x)}{v(x)}, \tag{7}$$

where $\psi(x) = \int_{\mathbb{R}} y f_{X,Y}(x, y)\, \mathrm{d}y$ and $f_{X,Y}(.,.)$ is the joint density of (X, Y). Then, assuming that $\tilde{v}_n(x) > 0$ for all $x \in \mathbb{R}^d$, it is well known that the kernel estimator of the regression function $m(x)$ under RLT model is given by

$$\tilde{m}_n(x) := \frac{\tilde{\psi}_n(x)}{\tilde{v}_n(x)},$$

where

$$\tilde{\psi}_n(x) := \frac{\alpha}{nh_n^d} \sum_{i=1}^{n} \frac{Y_i}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right). \tag{8}$$

As in practice α and G are usually unknown, we replace them by their consistent estimators G_n and α_n defined in Equations (1) and (3), respectively. Thus by plug-in method one can define the feasible kernel estimate for $m(.)$ by

$$\hat{m}_n(x) := \frac{\hat{\psi}_n(x)}{\hat{v}_n(x)}, \tag{9}$$

where

$$\hat{\psi}_n(x) := \frac{\alpha_n}{nh_n^d} \sum_{i=1}^{n} \frac{Y_i}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right),$$

$$\hat{v}_n(x) := \frac{\alpha_n}{nh_n^d} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right).$$

3. Assumptions and main results

Throughout this paper we assume that $a_G < a_F$ and $b_G \leq b_F$. Let D be a compact set which is included in $\Xi = \{x \in \mathbb{R}^d / v(x) > \delta > 0\}$ for a real $\delta > 0$, and let us define

$$\theta_{i,j} := \sum_{k=1}^{d} \sum_{l=1}^{d} \operatorname{cov}(X_{i,k}, X_{j,l}) + 2 \sum_{k=1}^{d} \operatorname{cov}(X_{i,k}, Y_j) + \operatorname{cov}(Y_i, Y_j), \tag{10}$$

where $X_{i,k}$ is the k-th component of X_i. We will make use of the following assumptions gathered here for easy reference.

- **(H)** The bandwidth h_n satisfies: $h_n \to 0$, $nh_n^d \to +\infty$ and $\log^5 n / nh_n^d \to 0$ as $n \to +\infty$,
- **(A)** $\int_{\mathbb{R}} (\mathrm{d}F(z)/G^2(z)) < +\infty$,
- **(K1)** The kernel K_d is a bounded probability density with compact support,
- **(K2)** $\int_{\mathbb{R}^d} z_i K_d(z)\,\mathrm{d}z = 0$ for all $i = 1, \ldots, d$ and $\int_{\mathbb{R}^d} |z_1^{i_1} \cdots z_d^{i_d}| K_d(z)\,\mathrm{d}z < +\infty$ for $i_1 + \cdots + i_d = 2$,
- **(K3)** $\int_{\mathbb{R}^d} |z_1 + \cdots + z_d| K_d^2(z)\,\mathrm{d}z < +\infty$ and $\int_{R^d} K_d^2(z)\,\mathrm{d}z < +\infty$,
- **(K4)** K_d is Hölder continuous with exponent $\beta > 0$,
- **(R)** The covariance term defined by $\rho(s) := \sup_{|i-j| \geq s} \theta_{i,j}$ for $s > 0$ satisfies $\rho(s) \leq \gamma_0 e^{-\gamma s}$ for some positive constants γ_0 and γ,
- **(D1)** The function $\psi(.)$ is bounded, twice differentiable with $\sup_{x \in D} |(\partial^k \psi / \partial x_i \partial x_j^{k-1})(x)| < +\infty$ for $i, j = 1, \ldots, d$ and $k = 1{,}2$,

(D2) The function $\psi_1(x) := \int_{\mathbb{R}} (y^2/G(y)) F(x, \mathrm{d}y)$ is bounded and continuously differentiable with $\sup_{x\in D} |(\partial\psi_1/\partial x_i)(x)| < +\infty$ for $i = 1, \ldots, d$,

(D3) The probability density $v(.)$ is bounded and twice differentiable with $\sup_{x\in D} |(\partial^k v/\partial x_i \partial x_j^{k-1})(x)| < +\infty$ for $i = 1, \ldots, d$ and $k = 1,2$,

(D4) The conditional joint density $v^*_{i,j}$ of (X_i, X_j) exists and $\sup_{x_1,x_2\in D} |v^*_{i,j}(x_1, x_2)| < +\infty$.

Remark 3.1: Assumptions **H, K1, K4** and **D1** are very common in functional estimation both in independent and dependent cases. Assumptions **A** and **R** imply assumptions of Guessoum et al. (2012) and are needed, among other, to use their results. Furthermore, assumption **R** quantifies a geometric decay of the covariance terms needed to establish a uniform almost sure convergence type. This condition is similar to the one used in Bulinski (1996), Douge (2007) and Doukhan and Neumann (2007) (without covariate). Assumptions **D3** and **D4** are technical and are used to compute covariances.

The following two theorems give the uniform asymptotic expression of the fluctuation terms for the two estimators $\tilde{\psi}_n(x)$ and $\tilde{v}_n(x)$ defined respectively in Equations (8) and (6).

Theorem 3.1: *Under assumptions **K1–K4, D1–D2, D4, R** and **H** we have*

$$\sup_{x\in D} |\tilde{\psi}_n(x) - \mathbf{E}(\tilde{\psi}_n(x))| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad \boldsymbol{P}\text{ - a.s, as } n \to +\infty.$$

Theorem 3.2: *Under assumptions **K1–K4, D3–D4, R** and **H** we have*

$$\sup_{x\in D} |\tilde{v}_n(x) - \mathbf{E}(\tilde{v}_n(x))| = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right) \qquad \boldsymbol{P}\text{ - a.s, as } n \to +\infty.$$

The proofs of Theorem 3.1 and Theorem 3.2 are mainly based on a Bernstein-type inequality due to Doukhan and Neumann (2007) recalled in Lemma A.1 in the appendix.

Now, to state our main result observe that from Equations (7) and (9) we have

$$\begin{aligned}\hat{m}_n(x) - m(x) &= \frac{\hat{\psi}_n(x)}{\hat{v}_n(x)} - \frac{\psi(x)}{v(x)} \\ &= \left(\frac{\hat{\psi}_n(x)}{\hat{v}_n(x)} - \frac{\tilde{\psi}_n(x)}{\hat{v}_n(x)}\right) + \left(\frac{\tilde{\psi}_n(x)}{\hat{v}_n(x)} - \frac{\mathbf{E}(\tilde{\psi}_n(x))}{\hat{v}_n(x)}\right) \\ &\quad + \left(\frac{\mathbf{E}(\tilde{\psi}_n(x))}{\hat{v}_n(x)} - \frac{\psi(x)}{\hat{v}_n(x)}\right) + \psi(x)\frac{v(x) - \hat{v}_n(x)}{\hat{v}_n(x)v(x)},\end{aligned} \tag{11}$$

then application of Theorem 3.1 and Theorem 3.2 leads to

Theorem 3.3: *Under assumptions **A, K1–K4, D1–D4, R** and **H** we have*

$$\sup_{x\in D} |\hat{m}_n(x) - m(x)| = O\left\{\sqrt{\frac{\log n}{nh_n^d}} \vee \left(\frac{\log\log n}{n}\right)^{\theta} \vee h_n^2\right\} \qquad \boldsymbol{P}\text{ - a.s, as } n \to \infty,$$

where $0 < \theta < (\gamma/(2\gamma + 6 + 3\kappa/2))$ *for any real* $\kappa > 0$.

Remark 3.2: The rate obtained by Ould Saïd and Lemdani (2006) in the independent case for $d=1$ (their Theorem 4.1) is $O(\sqrt{\log n/n^{1-2\lambda}h_n^2} \vee h_n)$ where λ is defined in their assumptions **A1** and **A2**. If we compare it with our result, we find an added term depending on θ, which is due to the association effect. Note that our rate is slightly better than theirs due to the symmetric condition on the kernel K in our assumption **K2.** In α-mixing case, Liang et al. (2009) obtained the rate $O\{\sqrt{\log n/nh_n^d} \vee \hat{w}_n \vee h_n^p\}$ where p denotes the order of the kernel K and $\hat{w}_n$ quantifies the α-mixing effect. Then if we take $p=2$ (symmetric kernel), their rate becomes similar to ours.

4. Simulations study

The main purpose of this section is to investigate the finite sample performance of the estimator $\hat{m}_n(x)$ in the case $d = 1$, for some particular regression functions $m(x)$. For that, we generate data as follows

(1) The covariate X:
- Generate $N+1$ i.i.d. $\mathcal{N}(0,1)$ rv's $\{W_t;\ t = -1, 0, \dots, N-1\}$.
- Given W_t, generate the associated sequence $\{X_t,\ \ t = 1, \dots, N\}$ by $X_t = \exp[\frac{1}{2}(W_{t-1} + W_{t-2})]$. This model comes from Chaubey, Dewan, and Li (2011) where it is shown that the autoregressif model of order 2 defined by $S_t = W_{t-1}/2 + W_{t-2}/2$ is associated. Furthermore, as the exponential is a nondecreasing function so, from property (P4) of Esary et al. (1967), X_t are associated variables.

(2) The interest variable Y :
- Generate N i.i.d. rv's $\{\varepsilon_t;\ t = 1, \dots, N\}$ with distribution specified below.
- Set $Y_t = m(X_t) + \varepsilon_t$ with $m(.)$ the regression function.

(3) The truncated variable T:
- Generate independently the i.i.d. rv's $\{T_t\}$ with distribution $\mathcal{N}(\mu, 1)$ (μ is adapted in order to obtain different values of truncation).

(4) The observed data:
- We keep the n observations $\{(X_i, Y_i, T_i),\ i = 1, \dots, n\}$ of the triplet of rv's (X, Y, T) satisfying the condition $Y_i \geq T_i$.

The supremum is taken over a compact set $D = [a, b]$ for which we consider a subdivision ϖ defined by

$$\varpi = \{a_0 = a, x_1, \dots, a_J = b\}, \quad J \geq 1,$$

and we calculate the values of the estimator $\hat{m}_n(x)$, at each point a_j of ϖ by choosing a Gaussian kernel K and a bandwidth $h_n = O((\log n/n)^{1/3})$.

This section is divided in three parts. In the first one, we focus on the performance of the estimator when the model $m(x)$ is linear, for different values of α (the rate related to the (no) truncation) and n (the size of the observed sample). In the second part, we compare the performance of the estimator by considering two different distributions of the errors ε_i and thereafter we propose to compare $\hat{m}_n(x)$ to the restricted conditional mean survival time (RCMST) estimate. Finally some nonlinear regression functions are chosen for highlighting the robustness of our estimator.

4.1. Performances of $\hat{m}_n(.)$ under linear model

We generate the data with observed sample sizes $n = 50$, 100 and 300, respectively, from the linear model given by

$$Y_i = 2X_i + 1 + \varepsilon_i, \quad i = 1, \dots, N,$$

where ε_i comes from $\mathcal{N}(0, 0.2)$. Hence, the true underlying regression function is $m(x) = 2x + 1$. In Tables 1 and 2, we take $\alpha \approx 60\%$ and 80% (obtained for $\mu = 2.5$ and $\mu = 2$, respectively) and we report the bias, variance and MSE of the estimator $\hat{m}_n$ at the bounds $x = 0$ and $x = 2$ and at a middle point $x = 1$, based on $B = 1000$ replications.

From Tables 1 and 2 it can be seen that:

- The estimator does not have the same behaviour at the bounds of the support and inside this one. In particular in the left we notice, as it is well known, that the estimator has boundary effects which may be attributed to the fact that the support of the kernel exceeds the available range of data.
- Each of the MSE, bias and variance decrease when n increases, so the quality of the estimator is better for high observed sample sizes which is confirmed by Figure 1 , where we plot $m(.)$ and its estimator $\hat{m}_n(.)$ with $\alpha \approx 80\%$ for $n = 50$, 100 and 300.

Hereafter, to take into account all points in D, we calculate the MSE along the interval $D = [0, 2]$ by taking the median over $x \in [0, 2]$. The results are given in Table 3 for $\alpha \approx 60\%, 70\%$, and 80% (in this case the different values of α are obtained for the same parameter $\mu = 2.5$, in order to make some comparisons using the same law for the simulated variables) for $n = 50$, 100 and 300.

As the MSE gives a pointwise error, we extend the study to the global behaviour of the estimator by using the mean integrated squared error (MISE). The results are given in Table 4 with $\alpha \approx 60\%, 70\%$ and 80% (obtained for $\mu = 2.5$) for $n = 50$, 100 and 300.

Tables 3 and 4 insure that the performance of the estimator is better for high sample size and great value of α. This is also confirmed by Figure 2, where we plot $m(.)$ and its estimator $\hat{m}_n(.)$ with $n = 300$ for $\alpha \approx 60\%, 70\%$ and 80%.

Table 1. MSE, bias and variance of $\hat{m}_n$ for $\alpha \approx 60\%$.

	$n = 50$			$n = 100$			$n = 300$		
	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var
$x = 0$	0.3955	0.6028	0.0321	0.2754	0.5053	0.0201	0.1547	0.3787	0.0114
$x = 1$	0.0023	−0.0171	0.0020	0.0008	−0.0117	0.0007	0.0002	−0.0059	0.0002
$x = 2$	0.0293	−0.0998	0.0193	0.0084	−0.0191	0.0080	0.0027	−0.0113	0.0026

Table 2. MSE, bias and variance of $\hat{m}_n$ for $\alpha \approx 80\%$.

	$n = 50$			$n = 100$			$n = 300$		
	MSE	Bias	Var	MSE	Bias	Var	MSE	Bias	Var
$x = 0$	0.2073	0.4212	0.0299	0.1182	0.3250	0.0126	0.0901	0.2971	0.0018
$x = 1$	0.0021	−0.0121	0.0019	0.0009	−0.0100	0.0008	0.0002	−0.0046	0.0002
$x = 2$	0.0102	−0.0678	0.0056	0.0044	−0.0324	0.0034	0.0019	−0.0223	0.0014

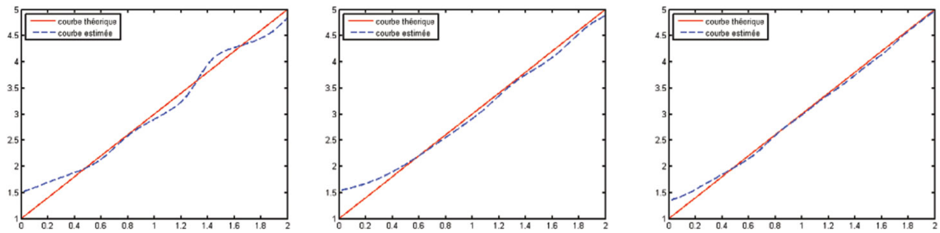

Figure 1. $m(.)$ and $\hat{m}_n(.)$ with $\alpha \approx 80\%$ and $n = 50, 100$ and 300, respectively.

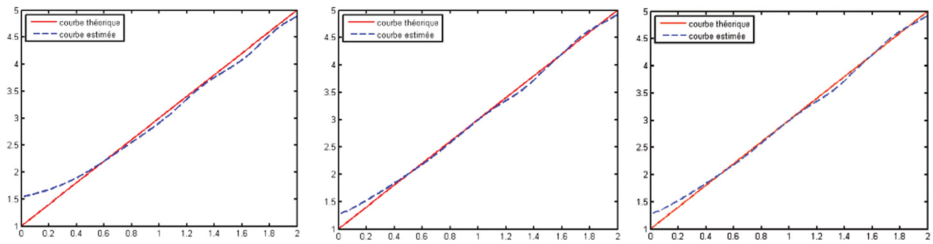

Figure 2. $m(.)$ and $\hat{m}_n(.)$ with $n = 300$ and $\alpha \approx 60\%, 70\%$ and 80%, respectively.

Table 3. MSE's median of $\hat{m}_n$.

α (%)	$n = 50$	$n = 100$	$n = 300$
60	0.0033	0.0012	3.0539×10^{-4}
70	0.0028	0.0009	3.0283×10^{-4}
80	0.0027	0.0006	2.7858×10^{-4}

Table 4. The MISE of $\hat{m}_n$.

α (%)	$n = 50$	$n = 100$	$n = 300$
60	0.0539	0.0305	0.0228
70	0.0341	0.0284	0.0128
80	0.0298	0.0198	0.0127

4.2. *Some comparisons*

- We compare the performances of $\hat{m}_n(x)$, via the MISE, based on $B = 1000$ replications, for two different distributions of the error ε_i by generating the data with observed sample sizes $n = 100$ and 300, respectively, from the linear model given by

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \ldots, N,$$

where (a) ε_i comes from $\mathcal{N}(0, 1)$ and (b) ε_i comes from $\mathcal{S}t(2)$.

In Table 5, it can be seen that the estimation is most robust when the error distribution is normal. One of the reasons may be that in distribution (b) the errors ε_i are heavy-tailed.

- Using the same model with normal error, we compare $\hat{m}_n(x)$ to the RCMST estimate defined below. Recall that in the case of complete and positive data the regression

Table 5. The MISE of $\hat{m}_n$.

α (%)	n	Distribution (a)	Distribution (b)
60	100	0.0591	0.9509
	300	0.0281	0.8092
70	100	0.0398	0.9291
	300	0.0194	0.7523
80	100	0.0368	0.9022
	300	0.0131	0.6993

function can be written as a conditional mean survival time that is

$$\mathbb{E}(Y \mid X = x) = \int_0^{+\infty} S(y|x)\,\mathrm{d}y,$$

where $S(y|x)$ is the conditional survival function of Y given $X = x$. Under the RLT model, we define the RCMST as

$$\text{RCMST} = \int_{a_F}^{b_F} S(y|x)\,\mathrm{d}y = \int_{a_F}^{b_F} [1 - F(y|x)]\,\mathrm{d}y,$$

where $F(y|x)$ is the conditional df of Y given $X = x$ and a_F, b_F are defined in Equation (2). Then we propose to estimate RCMST by

$$\bar{m}_n(x) = \int_{a_F}^{b_F} [1 - F_n(y|x)]\,\mathrm{d}y,$$

where $F_n(y|x)$ is the estimator of $F(y|x)$ given in Lemdani, Ould Saïd, and Poulin (2009) by

$$F_n(y|x) = \frac{\sum_{i=1}^n G_n^{-1}(Y_i) K_d\left(\frac{x-X_i}{h_n}\right) K_0\left(\frac{y-Y_i}{h_n}\right)}{\sum_{i=1}^n G_n^{-1}(Y_i) K_d\left(\frac{x-X_i}{h_n}\right)},$$

with K_0 a smooth df defined on $\mathbb{R}$.

In order to compare the performance of the two estimators $\hat{m}_n(x)$ and $\bar{m}_n(x)$ through their MISE, we generate the data with observed sample sizes 100 and 300, respectively, from the model given by

$$Y_i = X_i + \varepsilon_i, \quad i = 1, \ldots, N,$$

where $\varepsilon_i \sim \mathcal{N}(0, 1)$.

In Table 6, we take $\alpha \approx 60\%$ and 80% (obtained for $\mu = 1$) and we report the MISE of $\hat{m}_n(x)$ and $\bar{m}_n(x)$ based on $B = 1000$ replications. We notice that the two estimators have similar performance which is confirmed by Figure 3 where we plot the true regression function $m(x) = x$ together with the estimators $\hat{m}_n(x)$ and $\bar{m}_n(x)$ for $\alpha \approx 80\%$, $n = 100$ and 300.

Table 6. The MISE of $\hat{m}_n$ and $\bar{m}_n(x)$.

α (%)	n	$\hat{m}_n(x)$	$\bar{m}_n(x)$
60	100	0.0585	0.0591
	300	0.0279	0.0281
80	100	0.0368	0.0365
	300	0.0131	0.0132

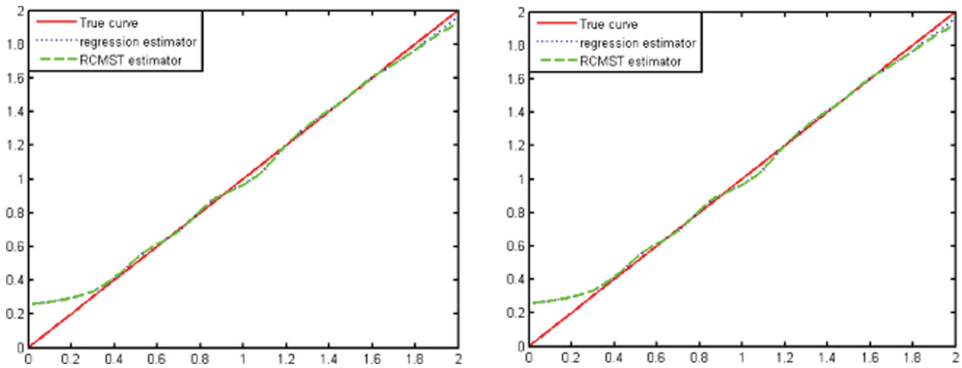

Figure 3. $m(.)$, $\hat{m}_n(.)$ and $\bar{m}_n(.)$ with $\alpha \approx 80\%$ and $n = 100, 300$, respectively.

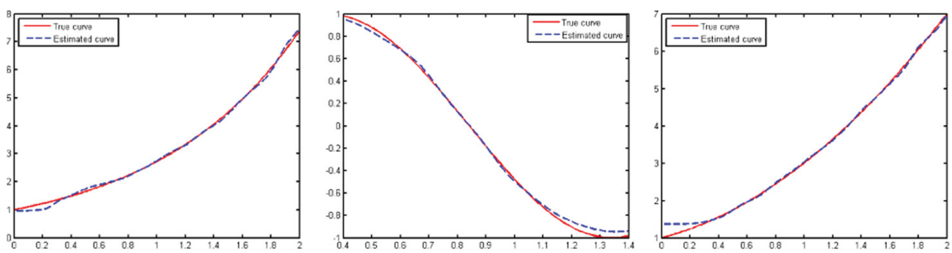

Figure 4. Exponential, sinus, and parabolic with $n = 300$ and $\alpha \approx 80\%$.

4.3. Nonlinear model

We consider the case of nonlinear regression by choosing the following three models:

$$m(x) = \exp(x) + \varepsilon_i, \text{ exponential,}$$

$$m(x) = \sin\left(\pi x + \frac{1}{2}\right) + \varepsilon_i, \text{ sinus,}$$

$$m(x) = x^2 + x + 1 + \varepsilon_i, \text{ parabolic,}$$

with $D = [0, 2]$ for exponential and parabolic cases and $D = [0.4, 1.4]$ for sinus case. The error ε_i is taken normal, the size n equal to 300 and α equal to 80% (Figure 4).

Here again, we can see the good performance of our estimator for nonlinear regression functions.

5. Conclusion

This paper has established the uniform strong consistency along with a rate of the kernel regression estimator over a real compact, when the variable of interest is subject to left truncation under association hypothesis. In doing so, a Bernstein-type inequality due to Doukhan and Neumann (2007) has been used. A large simulation study was conducted through which our estimator performance was highlighted in spite of well-known boundary effects of kernel estimation. Alternative methods such as local linear and k-nearest neighbour will be the subject of future research. These methods could improve the bias and the boundary effects. Comparisons of the studied estimator with the RCMST estimate have been made. Regarding this latter estimator, another possible estimation procedure is to consider the conditional product limit estimator defined in Akritas and LaValley (2005), under the assumption that Y is conditionally independent of T given X, for which theoretical results do not exist for associated data. Many other results remain to be established as asymptotic normality and more generally asymptotic behaviour of nonparametric functional estimator.

Acknowledgements

The authors are grateful to the Chief Editor, the anonymous referees and the Associated Editor for their very careful reading of the original version of this paper. Their pertinent and helpful comments have led to substantial improvement in the quality of the paper.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akritas, M.G., and LaValley, M.P. (2005), 'A Generalized Product-limit Estimator for Truncated Data', *Journal of Nonparametric Statistics*, 17, 643–663.

Andrew, D.W.K. (1984), 'Non-strong Mixing Autoregressive Processes', *Journal of Applied Probability*, 21, 930–934.

Bagai, I., and Prakasa Rao, B.L.S. (1995), 'Kernel-type Density and Failure Rate Estimation for Associated Sequences', *Annals of the Institute of Statistical Mathematics*, 47, 253–266.

Bulinski, A. (1996), 'On the Convergence Rates in the CLT for Positively and Negatively Dependent Random Fields', in *Probability Theory and Mathematical Statistics*, eds. I.A. Ibragimov and A.Yu. Zaitsev, Netherlands: Gordon and Breach, pp. 3–14.

Bulinski, A., and Shashkin, A. (2007), *Limit Theorems for Associated Random Fields and Related Systems*, Advanced Series on Statistical Science & Applied Probability (Vol. 10), Hackensack, NJ: World Scientific Publishing.

Cai, Z., and Roussas, G.G. (1998), 'Kaplan–Meier Estimator Under Association', *Journal of Multivariate Analysis*, 67, 318–348.

Carrasco, M., Chernov, M., Florens, J., and Ghysels, E. (2007), 'Efficient Estimation of General Dynamic Models with a Continuum of Moment Conditions', *Journal of Econometrics*, 140, 529–573.

Chaubey, Y.P., Dewan, I., and Li, J. (2011), 'Smooth Estimation of Survival and Density Functions for a Stationary Associated Process Using Poisson Weights', *Statistics and Probability Letters*, 81, 267–276.

Chen, K., Chao, M.T., and Lo, S.W. (1995), 'On Strong Uniform Consistency of the Lynden-Bell Estimator for Truncated Data', *The Annals of Statistics*, 23, 440–449.

Douge, L. (2007), 'Vitesses de convergence dans la loi forte des grands nombres et dans l'estimation de la densité pour des variables alé atoires associées', *Comptes Rendus de l'Académie des Sciences Paris, Series I*, 344, 515–518.

Doukhan, P. (1994), *Mixing: Properties and Examples (Vol. 85)*, New York: Springer-Verlag.

Doukhan, P., and Louhichi, S. (1999), 'A New Weak Dependence Condition and Applications to Moment Inequalities', *Stochastic Processes and Their Applications*, 84, 313–342.

Doukhan, P., and Neumann, M. (2007), 'Probability and Moment Inequalities for Sums of Weakly Dependent Random Variables, with Applications', *Stochastic Processes and Their Applications*, 117, 878–903.

Esary, J., Proschan, F., and Walkup, D. (1967), 'Association of Random Variables with Applications', *The Annals of Mathematical Statistics*, 38, 1466–1476.

Ferrani, Y., Ould Saïd, E., and Tatachak, A. (2016), 'On Kernel Density and Mode Estimates for Associated and Censored Data', *Communications in Statistics – Theory and Methods*, 45, 1853–1862.

Guessoum, Z., Sadki, O., and Tatachak, A. (2012), 'A Note on the Lynden-Bell Estimator under Association', *Statistics & Probability Letters*, 82, 1994–2000.

He, S., and Yang, G. (1998), 'Estimation of the Truncation Probability in the Random Truncation Model', *The Annals of Statistics*, 26, 1011–1027.

Kalbfleisch, J.D., and Lawless, J.F. (1989), 'Inference Based on Retrospective Ascertainment: An Analysis of the Data on Transfusionrelated AIDS', *Journal of the American Statistical Association*, 84, 360–372.

Klein, J.P., and Moeschberger, M.L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data (2nd ed.)*, New York: Springer.

Lemdani, M. (2007), 'Asymptotic Behavior of the Hazard Rate Kernel Estimator Under Truncated and Censored Data,', *Communications in Statistics – Theory and Methods*, 36(1):155–173.

Lemdani, M., Ould Saïd, E., and Poulin, N. (2009), 'Asymptotic Properties of a Conditional Quantile Estimator with Randomly Truncated Data', *Journal of Multivariate Analysis*, 100, 546–559.

Liang, H.Y. (2011), 'Asymptotic Normality for Regression Function Estimate Under Truncation and α-Mixing Conditions', *Communications in Statistics – Theory and Methods*, 40, 1999–2021.

Liang, H.Y., Li, D.L., and Qi, Y.C. (2009), 'Strong Convergence in Nonparametric Regression with Truncated Dependent Data', *Journal of Multivariate Analysis*, 100, 162–174.

Lynden-Bell, D. (1971), 'A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars', *Monthly Notices Royal Astronomy Society*, 155, 95–118.

Masry, E. (1986), 'Recursive Probability Density Estimation for Weakly Dependent Processes', *IEEE Transactions on Information Theory*, 32, 254–267.

Ould Saïd, E., and 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., and 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.

Pham, T.D., and Tran, L.T. (1985), 'Some Mixing Properties of Time Series Models', *Stochastic Processes and Their Applications*, 19, 297–303.

Roussas, G.G. (2000), 'Asymptotic Normality of the Kernel Estimate of a Probability Density Function under Association', *Statistics and Probability Letters*, 50, 1–12.

Stute, W. (1993), 'Almost Sure Representation of the Product-limit Estimator for Truncated Data', *Annals of Statistics*, 21, 146–156.

Tsai, W.Y., Jewell, N.P., and Wang, M.C. (1987), 'A Note on the Product-limit Estimator Under Right Censoring and Left Truncation', *Biometrika*, 74, 883–886.

Walk, H. (2005), 'Almost Sure Convergence Properties of Nadaraya-Watson Regression Estimates', in *Modeling Uncertainty*, eds M. Dror, P. l'Ecuyer and F. Szidarovsky, International Series in Operations Research & Management Science (Vol. 46), Springer, pp. 201–223.

Wang, M.C., Jewell, N.P., and Tsai, W.Y. (1986), 'Asymptotic Properties of the Product-limit Estimate under Random Truncation', *Annals of Statistics*, 14, 1597–1605.
Woodroofe, M. (1985), 'Estimating a Distribution Function with Truncated Data', *Annals of Statistics*, 13, 163–177.

Appendix 1. Proofs

First, let

$$Z_i(x) := Z(X_i, Y_i; x) = \frac{\alpha Y_i}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right) - \mathbf{E}\left(\frac{\alpha Y_i}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right)\right). \tag{A1}$$

We recall in the following lemma an exponential inequality stated in Doukhan and Neumann (2007), used in the proofs of Theorems 3.1 and 3.2.

Lemma A.1 (Doukhan and Neumann 2007, Theorem 1, p. 880): *Suppose that $X_1, \ldots, X_n$ are real-valued rvs with zero mean, defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We assume that there exist constants $K, M, L_1, L_2 < +\infty, \mu, \nu \geq 0$ and a non-increasing sequence of real coefficients $(\delta(n))_{n\geq 0}$ such that, for all u-tuples $(s_1, \ldots, s_u)$ and all v-tuples $(t_1, \ldots, t_v)$ with $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v \leq n$, the following inequalities are fulfilled:*

(a) $|\mathrm{cov}(X_{s_1} \cdots X_{s_u}, X_{t_1} \cdots X_{t_v})| \leq K^2 M^{u+v-2}((u+v)!)^{\nu} uv\delta(t_1 - s_u)$,
(b) $\sum_{s=0}^{\infty}(s+1)^k(\delta(s)) \leq L_1 L_2^k (k!)^{\mu}\ \forall k \geq 0$,
(c) $\mathbb{E}(|X_t|^k) \leq (k!)^{\nu} M^k\ \forall k \geq 0$.

Then, for all $t \geq 0$

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\right) \leq \exp\left(-\frac{t^2/2}{A_n + B_n^{1/(\mu+\nu+2)} t^{(2\mu+2\nu+3)/(\mu+\nu+2)}}\right), \tag{A2}$$

where A_n can be chosen as any number greater than or equal to $\sigma_n^2 := \mathbb{V}ar(\sum_{i=1}^{n} X_i)$ and $B_n = 2(K \vee M)L_2(2^{4+\mu+\nu} nK^2 L_1/A_n \vee 1)$.

Throughout the proofs, we denote by c (different) constants whose values are allowed to change. The following two lemmas are used to show that the process $Z_i(x)$ satisfies the conditions of Lemma A.1, that will allow us to use the exponential inequality (A2). In the first lemma, to lighten the notations, we note $Z_i(x)$ by Z_i.

Lemma A.2: *Under assumption **R**, there exist constants $K, M, L_1, L_2 < +\infty, \mu, \lambda \geq 0$ such that for all $(s_1, \ldots, s_u) \in \mathbb{N}^u$ and all $(t_1, \ldots, t_v) \in \mathbb{N}^v$ with $1 \leq s_1 \leq \cdots \leq s_u \leq t_1 \leq \cdots \leq t_v \leq n$, we have*

(a) $\mathrm{cov}(Z_{s_1} \cdots Z_{s_u}, Z_{t_1} \cdots Z_{t_v}) \leq K^2 M^{u+v-2}((u+v)!)^{\lambda} uv(\rho(t_1 - s_u))^{d/(2d+2)}$,
(b) $\sum_{s=0}^{\infty}(s+1)^k(\rho(s))^{d/(2d+2)} \leq L_1 L_2^k (k!)^{\mu}\ \forall k \geq 0$,
(c) $\mathbf{E}(|Z_i|^k) \leq (k!)^{\lambda} M^k\ \forall k \geq 0$.

Proof: Following Bulinski and Shashkin (2007), let us define for a function $\Phi_m : \mathbb{R}^{m(d+1)} \to \mathbb{R}$, the partial Lipschitz constants of Φ_m, that is

$$\mathrm{Lip}_i(\Phi_m) := \sup_{z_i \neq z_i'} \frac{|\Phi_m(z_1, \ldots, z_{i-1}, z_i, z_{i+1}, \ldots, z_{m(d+1)}) - \Phi_m(z_1, \ldots, z_{i-1}, z_i', z_{i+1}, \ldots, z_{m(d+1)})|}{|z_i - z_i'|},$$

where $z_j \in \mathbb{R}$. For a function $\Phi : \mathbb{R}^m \to \mathbb{R}$, let $\mathrm{Lip}(\Phi)$ denote the Lipschitz modulus of continuity of Φ, that is

$$\mathrm{Lip}(\Phi) = \sup_{x \neq y} \frac{|\Phi(x) - \Phi(y)|}{\|x - y\|_1},$$

where $\|x\|_1 = |x_1| + \cdots + |x_m|$.

Set $\Phi_u = \prod_{i=s_1}^{s_u} Z_i$ and $\Phi_v = \prod_{j=t_1}^{t_v} Z_j$. Then from one hand, by using Theorem 5.3 in Bulinski and Shashkin (2007) under the Lipschitz condition on K_d and G and from Equation (10), we have

$$\mathrm{cov}(Z_{s_1} \cdots Z_{s_u}, Z_{t_1} \cdots Z_{t_v}) \leq \sum_{i=s_1}^{s_u} \sum_{j=w_1}^{w_v} \mathrm{Lip}_i(\Phi_u)\, \mathrm{Lip}_j(\Phi_v) \theta_{i,j}.$$

As Y is bounded, and from the definition of $\mathrm{Lip}_i(.)$ given above, we get

$$\mathrm{Lip}_i(\Phi_u) \leq \frac{C_1}{h_n} \left(\frac{c}{G(a_F)} \right)^{u-1} \|K_d\|_\infty^{u-1},$$
$$\mathrm{Lip}_i(\Phi_v) \leq \frac{C_1}{h_n} \left(\frac{c}{G(a_F)} \right)^{v-1} \|K_d\|_\infty^{v-1},$$

where $C_1 = \mathrm{Max}\{\mathrm{Lip}(K_d)(c/G(a_F)), \|K_d\|_\infty (1/G(a_F) + \mathrm{Lip}(G)(c/G^2(a_F)))\}$. Therefore, the stationarity and assumption **R** give

$$\begin{aligned}\mathrm{cov}(Z_{s_1} \cdots Z_{s_u}, Z_{t_1} \cdots Z_{t_v}) &\leq \frac{C_1^2}{h_n^2} \left(\frac{c}{G(a_F)} \right)^{u+v-2} \|K_d\|_\infty^{u+v-2} uv\rho(t_1 - s_u) \\ &\leq C_1^2 \left(\frac{c\|K_d\|_\infty}{G(a_F)} \right)^{u+v-2} \frac{1}{h_n^2} uv\rho(t_1 - s_u). \end{aligned} \tag{A3}$$

On the other hand, under assumptions **K1** and **D4** we have

$$\mathbf{E}\left(\frac{\alpha^2 Y_i Y_j}{G(Y_i)G(Y_j)} K_d \left(\frac{x - X_i}{h_n} \right) K_d \left(\frac{x - X_j}{h_n} \right) \right) \leq c \left(\frac{1}{G(a_F)} \right)^2 h_n^{2d} \tag{A4}$$

and

$$\mathbf{E}\left(\frac{\alpha Y_i}{G(Y_i)} K_d \left(\frac{x - X_i}{h_n} \right) \right) \leq \frac{c}{G(a_F)} h_n^d. \tag{A5}$$

Then from Equations (A4) and (A5) we deduce

$$\begin{aligned}\mathrm{cov}(Z_{s_1} \cdots Z_{s_u}, Z_{t_1} \cdots Z_{t_v}) &= \mathbf{E}(Z_{s_1} \cdots Z_{s_u} Z_{t_1} \cdots Z_{t_v}) - \mathbf{E}(Z_{s_1} \cdots Z_{s_u})\mathbf{E}(Z_{t_1} \cdots Z_{t_v}) \\ &\leq c \left(\frac{1}{G(a_F)} \right)^{u+v-2} h_n^{2d}. \end{aligned} \tag{A6}$$

By combining Equations (A3) and (A6), we get

$$\mathrm{cov}(Z_{s_1} \cdots Z_{s_u}, Z_{t_1} \cdots Z_{t_v}) \leq \left(\frac{c\|K_d\|_\infty}{G(a_F)} \right)^{u+v} h_n^d uv(\rho(t_1 - s_u))^{d/(2d+2)}. \tag{A7}$$

Choosing $K = c\|K_d\|_\infty / G(a_F)\sqrt{h_n^d}$, $M = c\|K_d\|_\infty / G(a_F)$ and $\lambda = 0$ we get the result in item (a). To prove the item (b), we follow the same steps as in the proof of Proposition 8 of Doukhan and

Neumann (2007). For all $k \geq 0$, under Assumption **R**, we have

$$\sum_{s=0}^{\infty}(s+1)^k(\rho(s))^{d/(2d+2)} \leq \sum_{s=0}^{\infty}(s+1)\cdots(s+k)e^{-\gamma sd/(2d+2)}$$
$$= \frac{d^k}{dt^k}\left(\frac{1}{1-t}\right)\Bigg|_{t=e^{-\gamma d/(2d+2)}}$$
$$= (k)!\left(\frac{1}{1-e^{-\gamma d/(2d+1)}}\right)^{k+1}.$$

Thus by choosing $\mu = 1$ and $L_1 = L_2 = 1/(1-e^{-\gamma d/(2d+2)})$ we get the result in item (b). For the proof of item (c), we have for all $k \geq 0$

$$\mathbf{E}(|Z_i|^k) \leq \left(\frac{c\|K_d\|_\infty}{G(a_F)}\right)^k$$
$$= (k!)^\lambda M^k,$$

where λ and M have the same values chosen to get result (a). ■

Lemma A.3: *Under assumptions **H, K1–K3, R** and **D1–D2**, we have*

$$\sigma_n^2 := \mathrm{Var}\left(\sum_{i=1}^n Z_i(x)\right) = O(nh_n^d).$$

Proof: We have

$$\sigma_n^2 = \mathrm{Var}\left(\sum_{i=1}^n Z_i(x)\right) = (nh_n^d)^2\mathrm{Var}(\tilde{\psi}_n(x) - \mathbf{E}(\tilde{\psi}_n(x))) = n^2h_n^{2d}\,\mathrm{Var}(\tilde{\psi}_n(x))$$
$$= n^2h_n^{2d}\,\mathrm{Var}\left(\frac{1}{nh_n^d}\sum_{i=1}^n \frac{\alpha Y_i}{G(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right)\right)$$
$$= n\mathrm{Var}\left(\frac{\alpha Y_1}{G(Y_1)}K_d\left(\frac{x-X_1}{h_n}\right)\right) + \sum_{i=1}^n\sum_{\substack{j=1\\ j\neq i}}^n \mathrm{Cov}\left(\frac{\alpha Y_i}{G(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right), \frac{\alpha Y_j}{G(Y_j)}K_d\left(\frac{x-X_j}{h_n}\right)\right)$$
$$=: V + CV.$$

On the one hand, we have

$$V = n\left[\mathbf{E}\left(\frac{\alpha^2Y_1^2}{G^2(Y_1)}K_d^2\left(\frac{x-X_1}{h_n}\right)\right) - \mathbf{E}^2\left(\frac{\alpha Y_1}{G(Y_1)}K_d\left(\frac{x-X_1}{h_n}\right)\right)\right]$$
$$=: n(V_1 - V_2).$$

Using Equation (4) and a change of variable, we get

$$V_1 = \int_{\mathbb{R}^d}\int_{\mathbb{R}} \frac{\alpha^2y^2}{G^2(y)}K_d^2\left(\frac{x-u}{h_n}\right)F^*(\mathrm{d}u,\mathrm{d}y)$$
$$= \int_{\mathbb{R}^d}\int_{\mathbb{R}} \frac{\alpha y^2}{G(y)}K_d^2\left(\frac{x-u}{h_n}\right)F(\mathrm{d}u,\mathrm{d}y)$$
$$\leq h_n^d\int_{\mathbb{R}^d} K_d^2(z)\psi_1(x-zh_n)\,\mathrm{d}z.$$

A Taylor expansion gives

$$\psi_1(x - zh_n) = \psi_1(x) - h_n\left(z_1\frac{\partial\psi_1}{\partial z_1}(x^*) + \cdots + z_d\frac{\partial\psi_1}{\partial z_d}(x^*)\right),$$

where x^* is between $x - h_n z$ and x. Then under assumptions **H, K3** and **D2** we have $V_1 = O(h_n^d)$. In the same way, we get

$$\begin{aligned}
V_2 &= \left[\int_{\mathbb{R}^d}\int_{\mathbb{R}} \frac{\alpha y}{G(y)} K_d\left(\frac{x-u}{h_n}\right) F^*(\mathrm{d}u, \mathrm{d}y)\right]^2 \\
&= \left[\int_{\mathbb{R}^d}\int_{\mathbb{R}} y K_d\left(\frac{x-u}{h_n}\right) F(\mathrm{d}u, \mathrm{d}y)\right]^2 \\
&= \left[h_n^d \int_{\mathbb{R}^d} K_d(z)\psi(x - zh_n)\,\mathrm{d}z\right]^2.
\end{aligned}$$

Therefore under Assumptions **H, K1–K2** and **D1**, a Taylor expansion gives $V_2 = O(h_n^{2d})$. Thus $V = O(nh_n^d)$. On the other hand, from Equation (A6) we can write

$$\mathrm{Cov}\left(\frac{\alpha Y_i}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right), \frac{\alpha Y_j}{G(Y_j)} K_d\left(\frac{x - X_j}{h_n}\right)\right) = O(h_n^{2d}). \tag{A8}$$

To evaluate CV, we use a technique developed in Masry (1986). Let us define

$$\begin{aligned}
\mathcal{B}_1 &= \{(i,j); 1 \le |i-j| \le \vartheta_n\}, \\
\mathcal{B}_2 &= \{(i,j); \vartheta_n + 1 \le |i-j| \le n-1\},
\end{aligned}$$

where $\vartheta_n = o(n)$. Then

$$\begin{aligned}
CV &= \sum_{i=1}^n \sum_{j\in\mathcal{B}_1} \mathrm{Cov}\left(\frac{\alpha Y_i}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right), \frac{\alpha Y_j}{G(Y_j)} K_d\left(\frac{x - X_j}{h_n}\right)\right) \\
&\quad + \sum_{i=1}^n \sum_{j\in\mathcal{B}_2} \mathrm{Cov}\left(\frac{\alpha Y_i}{G(Y_i)} K_d\left(\frac{x - X_i}{h_n}\right), \frac{\alpha Y_j}{G(Y_j)} K_d\left(\frac{x - X_j}{h_n}\right)\right) \\
&=: CV_1 + CV_2.
\end{aligned}$$

From Equation (A8) we get

$$CV_1 = O(\vartheta_n n h_n^{2d}). \tag{A9}$$

By Assumption **R** and Equation (A7) we obtain

$$\begin{aligned}
CV_2 &\le cnh_n^d \sum_{j\in\mathcal{B}_2} \mathrm{e}^{-\gamma d/(2d+2)|i-j|} \le cnh_n^d \int_{\vartheta_n}^n \mathrm{e}^{-(\gamma d/(2d+2))u}\,\mathrm{d}u \\
&= O(nh_n^d \mathrm{e}^{-\gamma\vartheta_n d/(2d+2)}).
\end{aligned} \tag{A10}$$

Now choosing $\vartheta_n = O(h_n^{\nu-d})$ with $0 < \nu < d$, Equations (A9) and (A10) become

$$\begin{aligned}
CV_1 &= O(nh_n^d h^\nu) = o(nh_n^d), \\
CV_2 &= O(nh_n^d e^{-(\gamma d/(2d+2))Ch_n^{\nu-d}}) = o(nh_n^d).
\end{aligned}$$

Finally $CV = o(nh_n^d)$ and $\sigma_n^2 = O(nh_n^d)$. ■

Proof of Theorem 3.1: Note that

$$\tilde{\psi}_n(x) - \mathbf{E}(\tilde{\psi}_n(x)) = \frac{1}{nh_n^d}\sum_{i=1}^{n} Z_i(x). \tag{A11}$$

We use a classical technique which consists in covering the compact D by a finite number p_n of balls $B_k(x_k, a_n^d)$ centred at $x_k = (x_{k,1}, \dots, x_{k,d})$, for $k \in 1, \dots, p_n$ and where $a_n^d = n^{-1/2\beta} h^{1+d/2\beta}$.

Then for all $x \in D$, there exists $k \in 1, \dots, p_n$ such that $\|x - x_k\| \leq a_n^d$. As D is bounded, there exists a constant $M' > 0$ such that

$$p_n a_n^d \leq M' \Rightarrow p_n \leq \frac{M'}{a_n^d} \Rightarrow p_n = O((a_n^d)^{-1}).$$

Hence we consider the following decomposition:

$$\begin{aligned}
\sup_{x\in D} |\tilde{\psi}_n(x) - \mathbf{E}(\tilde{\psi}_n(x))| &= \sup_{x\in D}\left|\frac{1}{nh_n^d}\sum_{i=1}^{n} Z_i(x) - \frac{1}{nh_n^d}\sum_{i=1}^{n} Z_i(x_k) + \frac{1}{nh_n^d}\sum_{i=1}^{n} Z_i(x_k)\right| \\
&\leq \max_{1\leq k\leq p_n}\sup_{x\in B_k}\frac{1}{nh_n^d}\sum_{i=1}^{n}|Z_i(x) - Z_i(x_k)| + \max_{1\leq k\leq p_n}\frac{1}{nh_n^d}\left|\sum_{i=1}^{n} Z_i(x_k)\right| \\
&=: S_1 + S_2.
\end{aligned} \tag{A12}$$

First, we have under assumption **K4** and **H**,

$$\begin{aligned}
\frac{1}{nh_n^d}\sum_{i=1}^{n}|Z_i(x) - Z_i(x_k)| &\leq \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\alpha|Y_i|}{G(Y_i)}\left|K_d\left(\frac{x - X_i}{h_n}\right) - K_d\left(\frac{x_k - X_i}{h_n}\right)\right| \\
&\quad + \frac{1}{h_n^d}\mathbf{E}\left(\frac{\alpha|Y_i|}{G(Y_i)}\left|K_d\left(\frac{x - X_i}{h_n}\right) - K_d\left(\frac{x_k - X_i}{h_n}\right)\right|\right) \\
&\leq \frac{2\alpha c}{nh_n^d G(a_F)} n \left\|\frac{x - X_1}{h_n} - \frac{x_k - X_1}{h_n}\right\|^{\beta} \\
&= \frac{c}{h_n^d}.\frac{\|x - x_k\|^{\beta}}{h_n^{\beta}} \\
&\leq \frac{c(a_n^d)^{\beta}}{h_n^{d+\beta}} = O\left(\frac{1}{\sqrt{nh_n^d}}\right).
\end{aligned}$$

So, we get

$$S_1 = O\left(\frac{1}{\sqrt{nh_n^d}}\right). \tag{A13}$$

We now turn to the term S_2 in Equation (A12). The use of Lemma A.2 shows that $Z_i(x)$ defined in Equation (A1) satisfies the conditions of Lemma A.1 with $\delta(s) = (\rho(s))^{d/(2d+2)}$ and we have

$$\mathbf{P}\left(\sum_{i=1}^{n} Z_i(x_k) \geq \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2/2}{A_n + B_n^{1/(\mu+\lambda+2)}\varepsilon^{(2\mu+2\lambda+3)/(\mu+\lambda+2)}}\right), \tag{A14}$$

where A_n can be chosen such that $A_n \leq \sigma_n^2$ with $\sigma_n^2 := \mathrm{Var}(\sum_{i=1}^{n} Z_i(x))$ and $B_n = 2cL_2(2^{4+\mu+\lambda} nch_n^d L_1/A_n \vee 1)$. From Lemma A.3 we have $A_n = O(nh_n^d)$. Furthermore, following the proof of Lemma A.2, we have $\mu = 1$, $\lambda = 0$, $L_1 = L_2 = 1/(1 - e^{-\gamma d/(2d+2)})$ and $B_n = O(1)$, then in

Equation (A14) we get

$$\mathbf{P}\left(\sum_{i=1}^{n} Z_i(x_k) \geq \varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2/2}{cnh_n^d + \varepsilon^{5/3}}\right). \tag{A15}$$

Next, if we choose $\varepsilon = \varepsilon_0\sqrt{\log n/nh^d}$ for all $\varepsilon_0 > 0$ then from Equation (A15) we have

$$\begin{aligned}
\mathbf{P}\left(\max_{1\leq k\leq p_n}\left|\sum_{i=1}^{n} Z_i(x_k)\right| > nh_n^d\varepsilon_0\sqrt{\frac{\log n}{nh^d}}\right) &\leq \sum_{k=1}^{p_n}\mathbf{P}\left(\left|\sum_{i=1}^{n} Z_i(x_k)\right| > nh_n^d\varepsilon_0\sqrt{\frac{\log n}{nh^d}}\right)\\
&\leq 2p_n \exp\left(-\frac{\frac{\varepsilon_0^2}{2}n^2h_n^{2d}\frac{\log n}{nh^d}}{cnh_n^d + \varepsilon_0^{\frac{5}{3}}(nh_n^d)^{5/3}\left(\frac{\log n}{nh_n^d}\right)^{5/6}}\right)\\
&\leq 2M'(a_n^d)^{-1}\exp\left(-\frac{\frac{\varepsilon_0^2}{2}\log n}{c+\varepsilon_0^{5/3}\left(\frac{\log^5 n}{nh_n^d}\right)^{1/6}}\right)\\
&= 2M'(a_n^d)^{-1}n^{-\varepsilon_0^2/2/c+\varepsilon_0^{5/3}(\log n^5/nh_n^d)^{1/6}}\\
&= 2M'n^{1/2\beta}h^{-1-d/2\beta}n^{-c\varepsilon_0^2}\\
&= 2M'(nh)^{-1-d/2\beta}n^{-c\varepsilon_0^2+1+(d+1)/2\beta}\\
&= 2M'\frac{1}{\sqrt{(nh)^{(2\beta+d)/\beta}}}n^{-c\varepsilon_0^2+1+(d+1)/2\beta}.
\end{aligned} \tag{A16}$$

By assumption **H** and for a suitable choice of ε_0 (i.e. $\varepsilon_0^2 > (1/C)(2 + (d + 1)/\beta)$) the last term in Equation (A16) is the general term of a convergent series. Finally, applying Borel–Cantelli 's lemma to Equation (A16) gives

$$S_2 = O\left(\sqrt{\frac{\log n}{nh_n^d}}\right). \tag{A17}$$

Thus combining Equations (A13), (A17) and (A12) ends the proof. ■

Proof of Theorem 3.2: We follow step by step the proof of Theorem 3.1 with

$$Z_i(x) = \frac{\alpha}{G(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right) - \mathbf{E}\left(\frac{\alpha}{G(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right)\right),$$

we get the result. ■

Lemma A.4: *Under assumptions **A, R** and **K1** we have*

$$\sup_{x\in D}|\hat{\psi}_n(x) - \tilde{\psi}_n(x)| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right] \quad \boldsymbol{P}\text{ - a.s, as } n \to +\infty.$$

Proof:

$$
\begin{aligned}
|\hat{\psi}_n(x) - \tilde{\psi}_n(x)| &= \left| \frac{1}{nh_n^d} \sum_{i=1}^{n} \left[\frac{\alpha_n}{G_n(Y_i)} - \frac{\alpha}{G(Y_i)} \right] Y_i K_d \left(\frac{x - X_i}{h_n} \right) \right| \\
&\leq \frac{1}{nh_n^d} \sum_{i=1}^{n} \left| \frac{\alpha_n - \alpha}{G_n(Y_i)} - \frac{\alpha}{G_n(Y_i)G(Y_i)} (G_n(Y_i) - G(Y_i)) \right| |Y_i| K_d \left(\frac{x - X_i}{h_n} \right) \\
&\leq \left\{ \frac{|\alpha_n - \alpha|}{G_n(a_F)} + \frac{\alpha}{G_n(a_F)G(a_F)} \sup_{y \geq a_F} |G_n(y) - G(y)| \right\} \frac{1}{nh_n^d} \sum_{i=1}^{n} |Y_i| K_d \left(\frac{x - X_i}{h_n} \right).
\end{aligned}
$$

By using Markov's inequality, assumption **K1** and for $\varepsilon > 0$ we get

$$
\begin{aligned}
\mathbf{P}\left(\frac{1}{nh_n^d} \sum_{i=1}^{n} |Y_i| K_d \left(\frac{x - X_i}{h_n} \right) \geq \varepsilon \right) &\leq \frac{\mathbf{E}\left(\frac{1}{nh_n^d} \sum_{i=1}^{n} |Y_i| K_d (\frac{x - X_i}{h_n}) \right)}{\varepsilon} \\
&\leq \frac{c}{\varepsilon h_n^d} \mathbf{E}\left(K_d \left(\frac{x - X_1}{h_n} \right) \right) \\
&= \frac{c}{\varepsilon h_n^d} \int_{R^d} K_d \left(\frac{x - u}{h_n} \right) v^*(u)\, \mathrm{d}u \\
&= c\varepsilon^{-1} \int_{R^d} K_d(z) v^*(x - h_n z)\, \mathrm{d}z \\
&= O(1).
\end{aligned}
$$

Now, by following Ould Saïd and Tatachak (2009), Guessoum et al. (2012) and using assumptions **A** and **R** we get

$$
\sup_{y \geq a_F} |G_n(y) - G(y)| = O\left[\left(\frac{\log \log n}{n} \right)^{\theta} \right]
$$

and

$$
|\alpha_n - \alpha| = O\left[\left(\frac{\log \log n}{n} \right)^{\theta} \right].
$$

Finally, since $G_n(a_F) \xrightarrow{\mathbf{P}-a.s.} G(a_F)$ and $(1/nh_n^d) \sum_{i=1}^{n} |Y_i| K_d((x - X_i)/h_n) = O(1)$, we deduce the result. ∎

Lemma A.5: *Under assumptions **K1–K2**, **D1** and **H** we have*

$$
\sup_{x \in D} |\mathbf{E}(\tilde{\psi}_n(x)) - \psi(x)| = O(h_n^2) \text{ a.s, as } n \to \infty.
$$

Proof:

$$
\begin{aligned}
\mathbf{E}(\tilde{\psi}_n(x)) - \psi(x) &= \mathbf{E}\left(\frac{1}{nh_n^d} \sum_{i=1}^{n} \frac{\alpha Y_i}{G(Y_i)} K_d \left(\frac{x - X_i}{h_n} \right) \right) - \psi(x) \\
&= \frac{1}{h_n^d} \int_{\mathbb{R}^d} \int_{R} \frac{\alpha y}{G(y)} K_d \left(\frac{x - u}{h_n} \right) F^*(\mathrm{d}u, \mathrm{d}y) - \psi(x) \\
&= \frac{1}{h_n^d} \int_{\mathbb{R}^d} \int_{\mathbb{R}} y K_d \left(\frac{x - u}{h_n} \right) F(\mathrm{d}u, \mathrm{d}y) - \psi(x) \\
&= \int_{\mathbb{R}^d} K_d(z) \left(\psi(x - zh_n) - \psi(x) \right) \mathrm{d}z.
\end{aligned}
$$

Under assumptions **K1–K2**, and **D1**, a Taylor expansion finishes the proof. ■

Lemma A.6: *Under assumptions **K1–K4**, **A**, **R**, **D3–D4** and **H***

$$\sup_{x\in D}|\hat{v}_n(x)-v(x)| = O\left\{\sqrt{\frac{\log n}{nh_n^d}} \vee \left(\frac{\log\log n}{n}\right)^{\theta} \vee h_n^2\right\} \quad \boldsymbol{P}\text{-}a.s,\ as\ n\to\infty,$$

where $0<\theta<\gamma/(2\gamma+6+3/2\kappa)$ *for any real* $\kappa>0$.

Proof: We have

$$\sup_{x\in D}|\hat{v}_n(x)-v(x)| \le \sup_{x\in D}|\hat{v}_n(x)-\tilde{v}_n(x)| + \sup_{x\in D}|\tilde{v}_n(x)-\mathbf{E}(\tilde{v}_n(x))| + \sup_{x\in D}|\mathbf{E}(\tilde{v}_n(x))-v(x)|.$$

Using the same steps and arguments as in the proof of Lemma A.4 we get under assumptions **A** and **K1**

$$\sup_{x\in D}|\hat{v}_n(x)-\tilde{v}_n(x)| = O\left[\left(\frac{\log\log n}{n}\right)^{\theta}\right]. \tag{A18}$$

Furthermore, under assumptions **K1–K2, D3** and **H** and using a Taylor expansion, we get

$$\sup_{x\in D}|\mathbf{E}(\tilde{v}_n(x))-v(x)| = O(h_n^2). \tag{A19}$$

By combining Equations (A18), (A19) and Theorem 3.2 we end the proof. ■

Proof of Theorem 3.3: As already mentioned in Section 3, let $\delta>0$ such that $\inf_{x\in D}|v(x)|>\delta$, hence we have from Equation (11)

$$\begin{aligned}
\sup_{x\in D}|\hat{m}_n(x)-m(x)| &\le \frac{1}{\inf\limits_{x\in D}(\hat{v}_n(x))}\left\{\sup_{x\in D}|\hat{\psi}_n(x)-\tilde{\psi}_n(x)| + \sup_{x\in D}|\tilde{\psi}_n(x)-\mathbf{E}(\tilde{\psi}_n(x))|\right.\\
&\quad\left. + \sup_{x\in D}|\mathbf{E}(\tilde{\psi}_n(x))-\psi(x)| + \frac{\sup_{x\in D}|\psi(x)|}{\inf_{x\in D}v(x)}\sup_{x\in D}|\hat{v}_n(x)-v(x)|\right\}\\
&\le \frac{1}{\delta-\sup_{x\in D}|\hat{v}_n(x)-v(x)|}\left\{\sup_{x\in D}|\hat{\psi}_n(x)-\tilde{\psi}_n(x)| + \sup_{x\in D}|\tilde{\psi}_n(x)\right.\\
&\quad - \mathbf{E}(\tilde{\psi}_n(x))| + \sup_{x\in D}|\mathbf{E}(\tilde{\psi}_n(x))-\psi(x)|\\
&\quad\left. + \delta^{-1}\sup_{x\in D}|\psi(x)|\sup_{x\in D}|\hat{v}_n(x)-v(x)|\right\}.
\end{aligned}$$

Then by using Theorem 3.1, Lemma A.4, Lemma A.5 and Lemma A.6 we get the result. ■