Electronic Journal of Statistics
Vol. 4 (2010) 117–132
ISSN: 1935-7524
DOI: 10.1214/08-EJS195

Kernel regression uniform rate estimation for censored data under α-mixing condition

Zohra Guessoum

Lab. M.S.T.D., Faculté de Math., Univ. Sci. Tech. Houari Boumédiène, BP 32, El Alia, 16111, Algeria
e-mail: z0guessoum@hotmail.com

Elias Ould Saïd*

Univ. Lille Nord de France, F-59000, France
ULCO, LMPA, F-62000, Calais, France
e-mail: ouldsaid@lmpa.univ-littoral.fr

Abstract: In this paper, we study the behavior of a kernel estimator of the regression function in the right censored model with α-mixing data. The uniform strong consistency over a real compact set of the estimate is established along with a rate of convergence. Some simulations are carried out to illustrate the behavior of the estimate with different examples for finite sample sizes.

AMS 2000 subject classifications: Primary 62G05, 62G20.
Keywords and phrases: Censored data, Kernel estimator, nonparametric regression, rate of convergence, strong consistency, strong mixing.

Received February 2008.

Contents

1. Introduction

In many survival practical applications, censored dependent data appear. For example, in the clinical trials domain, it is frequently happens that the patients

*Corresponding author.

from the same hospital have correlated survival times due to unmeasured variables like the quality of the hospital equipment. An example of such data can be found in Lipsitz and Ibrahim (2000). In economic duration data, event times are often correlated and the observation of the event may be prevented by the occurrence of an earlier competing event called censoring. Another example is the observations on duration of unemployment, which may be right censored and are typically correlated. For real data, the reader can refer to Wei and Lin (1989), Cai and Prentice (1995). However few papers deal with the regression function under censoring in the dependent case, we can cite the recent paper of El-Ghouch and Van Keilegom (2008) who estimate the regression function by applying polynomial local linear regression techniques using Beran's estimator.

Having in mind such kind of applications, consider a real random variable (rv) Y and a sequence of strictly stationary rv's $(Y_i)_{i\geq 1}$ with common unknown absolutely continuous disribution function (df) F and let $(C_i)_{i\geq 1}$, be a sequence of censoring rv's with common unknown df G. In contrast to statistics for complete data studies, censored model involves pairs $(T_i, \delta_i)_{i=1,\dots,n}$ where only $T_i = Y_i \wedge C_i$ and $\delta_i = \mathbb{I}_{\{Y_i \leq C_i\}}$ are observed.

Let X be an $\mathbb{R}^d$-valued random vector. Let $(X_i)_{i\geq 1}$ be a sequence of copies of the random vector X and denote by $X_{i,1}, \dots, X_{i,d}$ the coordinates of X_i. The study we perform below is then on the set of observations $(T_i, \delta_i, X_i)_{i\geq 1}$. In regression analysis one expectes to identify, if any, the relationship between the Y_i's and X_i's. This means looking for a function $m^*(X)$ describing this relationship that realizes the minimum of the mean squared error criterion. It is well known that this minimum is achieved by the regression function $m(x)$ defined on $\mathbb{R}^d$ by

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

There is a wide range of literature on nonparametric estimation of the regression function and many nonlinear smoothers including kernel, spline, local polynomial, orthogonal methods and so on. For an overview on methods and results for both theoretical and application points of view considering independent or dependent case, we refer the reader to Collomb (1981), Silverman (1986), Härdle (1990), Wahba (1990), Wand and Jones (1995), Masry and Fan (1997), Cai (2003) and Cai and Ould Saïd (2003).

In the uncensored case, the behavior of nonparametric estimators built upon mixing sequences is extensively studied. The consistency has been investigated by many authors. We only cite the recent work of Gonzalez-Manteiga et al. (2002) where they developed a nonparametric test, based on kernel smoothers, to decide whether some covariates could be suppressed in a multidimensional nonparametric regression study. Under the α-mixing condition, the uniform strong convergence of the Nadaraya-Watson estimator is treated in Doukhan (1994), Bosq (1998), Liebscher (2001) and the references therein.

Our goal is to establish the strong uniform convergence with rate for the kernel regression estimate under α-mixing condition in random censorship models where the independence condition is relaxed. For this kind of model, Cai (2001) established the asymptotic properties of the Kaplan-Meier estimator. Liebscher

(2002) derive a rate uniform for the strong convergence of kernel density and hazard rate estimators. His result represents an improvement of that given in Cai (1998b).

To this end, we were interested in extending the result of Guessoum and Ould Saïd (2008) from the iid to the dependent case. The paper is organized as follows: In Section 2 we give some definitions and notations under the censorship model of the regression function and strong-mixing process. Section 3 is devoted to the assumptions and main result. In Section 4, some simulations are drawn to lend further support to our theoretical results. Proof with auxiliary results are relegated to Section 5.

2. Definition of estimators

Suppose that $\{Y_i, i \geq 1\}$ and $\{C_i, i \geq 1\}$ are two independent sequences of stationary random variables. We want to estimate $m(x) = \mathbb{E}(Y \mid X = x)$ which can be written as $m(x) = \frac{r_1(x)}{\ell(x)}$ where

$$r_1(x) = \int_{\mathbb{R}} y f_{X,Y}(x, y) dy \tag{1}$$

with $f_{\cdot,\cdot}(x, y)$ being the joint density of (X, Y) and $\ell(\cdot)$ the density function of the covariates.

Now, it is well known that the kernel estimator of the regression function $m(\cdot)$ under censorship model (see, eg Carbonez et al. (1995)) is given by

$$\tilde{m}_n(x) = \sum_{i=1}^{n} W_{in}(x) \frac{\delta_i T_i}{\bar{G}(T_i)} \tag{2}$$

where $\bar{G}$ is the survival function of the rv C and

$$W_{in}(x) = \frac{K_d\left(\frac{x - X_i}{h_n}\right)}{\sum_{j=1}^{n} K_d\left(\frac{x - X_j}{h_n}\right)}$$

are the Watson-Nadaraya weights, K_d is a probability density function (pdf) defined on $\mathbb{R}^d$ and h_n a sequence of positive numbers converging to 0 as n goes to infinity. Then (2) can be written

$$\tilde{m}_n(x) =: \frac{\tilde{r}_{1,n}(x)}{\ell_n(x)}$$

where

$$\tilde{r}_{1,n}(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} \frac{\delta_i T_i}{\bar{G}(T_i)} K_d\left(\frac{x - X_i}{h_n}\right) \quad \text{and} \quad \ell_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K_d\left(\frac{x - X_i}{h_n}\right). \tag{3}$$

In practice, G is usually unknown, we replace it by the corresponding Kaplan–Meier (1958) estimator (KME) G_n defined by

$$\overline{G}_n(t) = \begin{cases} \prod_{i=1}^{n}\left(1 - \frac{1-\delta_i}{n-i+1}\right)^{\mathbb{1}_{\{Y_i \le t\}}}, & \text{if } t < Y_{(n)}, \\ 0, & \text{if } t \ge Y_{(n)}. \end{cases}$$

The properties of the KME for dependent variables can be found in Cai (1998a, 2001). Then a feasible estimator of $m(x)$ is given by:

$$m_n(x) = \frac{r_{1,n}(x)}{\ell_n(x)}$$

where

$$r_{1,n}(x) = \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\delta_i T_i}{\bar{G}_n(\,T_i)}K_d\left(\frac{x - X_i}{h_n}\right) \tag{4}$$

is an estimator of $r_1(x)$ and $\ell_n(x)$ (defined in (3)) an estimator of $\ell(x)$.

In what follows, we define the endpoints of F and G by $\tau_F = \sup\left\{y, \bar{F}(y) > 0\right\}$, $\tau_G = \sup\left\{y, \bar{G}(y) > 0\right\}$ and we assume that $\tau_F < \infty$ and $\bar{G}(\tau_F) > 0$ (this implies $\tau_F < \tau_G$).

For technical reasons (see Lemma 1), we assume that $\{C_i, i \ge 1\}$ and $\{(X_i, Y_i),\ i \ge 1\}$ are independent; furthermore this condition is plausible whenever the censoring is independent of the characteristics of the patient under study. We point out that since Y can be a lifetime we can suppose it bounded. We put $\|t\| = \sum_{j=1}^{d}|t_j|$ for $t \in \mathbb{R}^d$.

In order to define the α-mixing property, we introduce the following notations. Denote by $\mathcal{F}_i^k(Z)$ the $\sigma-$algebra generated by $\{Z_j, i \le j \le k\}$.

Definition *Let $\{Z_i, i = 1, 2, \ldots\}$ denote a sequence of rv's. Given a positive integer n, set*

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

The sequence is said to be α-mixing (strong mixing) if the mixing coefficient $\alpha(n) \to 0$ as $n \to \infty$.

Many processes do fulfill the strong mixing property. We quote, here, the usual ARMA processes which are geometrically strongly mixing, *i.e.*, there exist $\rho \in (0,1)$ and $a > 0$ such that, for any $n \ge 1$, $\alpha(n) \le a\rho^n$ (see, *e.g.*, Jones (1978)). The threshold models, the EXPAR models (see, Ozaki (1979)), the simple ARCH models (see Engle (1982)), their GARCH extension (see Bollerslev (1986)) and the bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions. We refer the reader to the recent Bradley's monograph.

We suppose that the sequences $\{Y_i, i \ge 1\}$ and $\{C_i, i \ge 1\}$ are α-mixing with coefficients $\alpha_1(n)$ and $\alpha_2(n)$, respectively. Cai (2001, Lemma 2) showed that $\{T_i, i \ge 1\}$ is then strongly mixing, with coefficient

$$\alpha(n) = 4\ \max(\alpha_1(n), \alpha_2(n)).$$

From now on, we suppose that $\{(T_i, \delta_i, X_i) \quad i = 1, \ldots, n\}$ is strongly mixing with mixing's coefficient $\alpha(n)$ such that $\alpha(n) = O(n^{-\nu})$ for some $\nu > 3$.

Now we are in position to give our assumptions and main result.

3. Assumptions and main result

Let $\mathcal{C}$ be a compact set of $\mathbb{R}^d$ which is included in $\mathcal{C}_0 = \left\{x \in \mathbb{R}^d \;/\; \ell(x) > 0\right\}$. We will make use of the following assumptions gathered here for easy reference:

A1) The bandwidth h_n satisfies:

i) $\lim_{n\to+\infty} \frac{nh_n^d}{\log n} = +\infty$,

ii) $\sqrt{\frac{\log\log n}{n}} = o(h_n)$,

iii) $\lim_{n\to+\infty} h_n^{d(\nu-2)} \log n = 0$.

A2) The kernel K_d is bounded and satisfies:

i) $\int_{\mathbb{R}^d} \|t\| \, K_d(t)dt < +\infty$,

ii) $\int_{\mathbb{R}^d} (t_1 + t_2 + \cdots + t_d) K_d^2(t)dt < +\infty$ and $\int_{\mathbb{R}^d} K_d^2(t)dt < +\infty$,

iii) $\forall (t, s) \in \mathcal{C}^2 \quad |K_d(t) - K_d(s)| \le \|t - s\|^\gamma$ for $\gamma > 0$.

A3) The function $r_1(\cdot)$ defined in (1) is continuously differentiable and $\sup_{x\in\mathcal{C}} \left|\frac{\partial r_1}{\partial x_i}(x)\right| < +\infty$ for $i = 1, \ldots, d$.

A4) The function $r_2(x) := \int_{\mathbb{R}^d} \frac{y^2}{G(y)} \, f_{X,Y}(x, y)dy$ is continuously differentiable and $\sup_{x\in\mathcal{C}} \left|\frac{\partial r_2}{\partial x_i}(x)\right| < +\infty$ for $i = 1, \ldots, d$.

A5) $\exists D_1 > 0$ and $\exists D_2 > 0$ such that $\sup_{u,v\in\mathcal{C}} |\ell_{ij}(u, v)| < D_1$ and $\sup_{u\in\mathcal{C}} |\ell(u)| < D_2$, where ℓ_{ij} is the joint distribution of (X_i, X_j).

A6) $\exists \theta > 0, \exists \; c_1 > 0$, such that

$$c_1 n^{\frac{\gamma(3-\nu)}{d[\gamma(\nu+1)+2\gamma+1]} + \theta d} \le h_n^d.$$

A7) The marginal density $\ell(.)$ is continuously differentiable and $\sup_{x\in\mathcal{C}} \left|\frac{\partial \ell}{\partial x_i}(x)\right| < +\infty$ for $i = 1, \ldots, d$ and there exists $\xi > 0$ such that $\ell(x) > \xi \quad \forall x \in \mathcal{C}$.

Remark 3.1. Assumption **A1** is very common in functional estimation both in independent and dependent cases. However, it must be reinforced by Assumption **A6** which ensure the convergence of the series which appear in proof of Lemma 3. We point out that **A1** *i)* implies **A1** *ii)* for $d \ge 2$. Assumptions **A2**, **A3**, **A4** and **A6** are needed in the study of the bias term of $r_{1,n}(x)$ which is the kernel estimator of $r_1(x)$. We point out that we do not require for K_d to be symmetric as in Guessoum and Ould Saïd (2008). Assumption **A7** intervenes in the convergence of the kernel density.

In the sequel the letter C denotes any generic constant.

Our main result is given in the following theorem which concerns the rate of the almost sure uniform convergence of the regression function.

Theorem 3.1. *Under Assumptions **A1-A7**, we have*

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

Remark 3.2. The rate obtained here is slightly different from that obtained by Guessoum and Ould Saïd (2008) in the independent case, which is $O\Big(\max\big\{\sqrt{\frac{\log n}{nh_n}},h_n^2\big\}\Big)$ for $d=1$. Their result can easily be generalized for higher dimensional covariate, ie $X\in\mathbb{R}^d$, by adapting their Assumptions **A2** and **A4** to obtain the rate $O\Big(\max\big\{\sqrt{\frac{\log n}{nh_n^d}},h_n^2\big\}\Big)$.

Remark 3.3. If we choose $h_n = O\Big(\big(\frac{\log n}{n}\big)^{1/d+2}\Big)$ then Theorem 3.1 becomes:

$$\sup_{x\in\mathcal{C}}|m_n(x)-m(x)| = O\left(\left(\frac{\log n}{n}\right)^{\frac{1}{d+2}}\right) \quad a.s.$$

This is the optimal rate obtained by Liebscher (2001) in the uncensored case.

4. Simulations study

First, we consider the strong mixing bidimentionnal process generated by:

$$\begin{aligned} X_i &= \rho X_{i-1}+\sqrt{1-\rho^2}\epsilon_i,\\ Y_i &= X_{i+1}\quad i=1,2,\dots,n, \end{aligned}$$

where $0<\rho<1$, $(\epsilon_i)_i$ is a white noise with standard Gaussian distribution and X_0 is a standard Gaussian rv independent of $(\epsilon_i)_i$. We also simulate n iid rv C_i exponentially distributed with parameter $\lambda = 1.5$. It is clear that the process (X_n,Y_n,C_n) is stationary and strongly mixing, in fact the process (X_n) is an AR(1) and given $X_1 = x$, we have $Y_1 = \rho x+\sqrt{1-\rho^2}\epsilon_2$, then, $Y_1 \rightsquigarrow N(\rho x, 1-\rho^2)$. We calculate our estimator based on the observed data $(X_i, T_i\ , \delta_i,)\ i=1,\dots,n$, by choosing a Gaussian kernel K. In this case, we have $m(x) = \mathbb{E}\left(Y_1\ |X_1 = x\right) = \rho x$. In all cases we took h_n satisfying **A6**, that is $h_n = O\big((\log n/n)^{1/3}\big)$. The following graphs show the behavior of our estimator when the percentage of censoring and ρ increase, for n rather large (here $n=300$). One notices without surprise that the best behavior is obtained for a rate of censoring equal to 0 (uncensored case) and an index of weak dependence $(\rho = 0.1)$.

We see that the asymptotic behavior is better for a small ρ and a weak rate of censoring which is conforted by the following table, where we show how the quality of the estimation is influenced by the percentage of the censoring which appear clearly when we have high censoring, for various values of ρ and n. The quality becomes slightly worse when we have high percentage of censoring, however it remains acceptable.

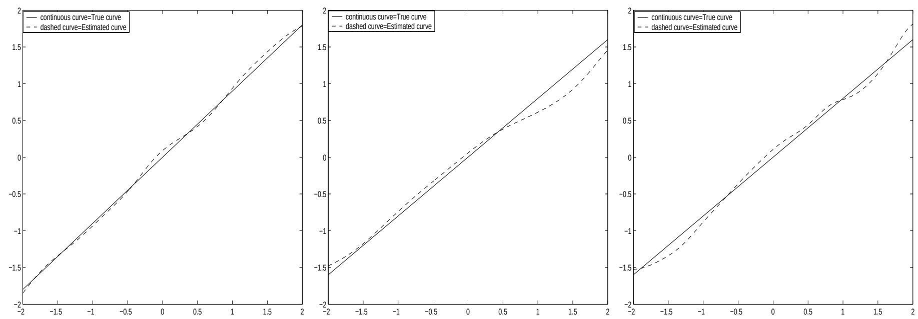

FIG 1. *$m(x) = \rho x$, with $\rho = 0.1, n = 300$ and a percentage of censoring* $= 0\%,\ 20\%$ *and* 40%, *respectively.*

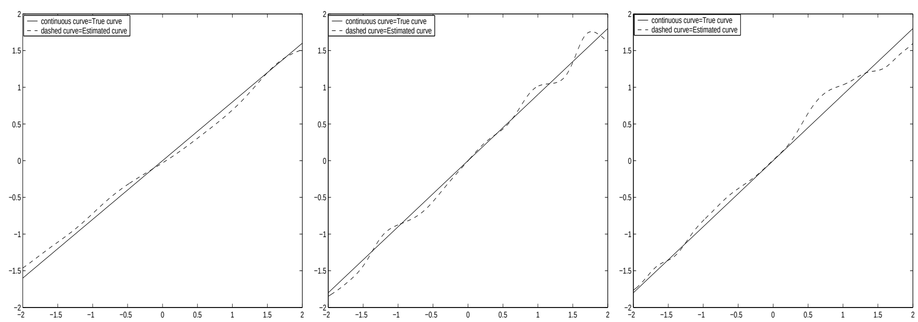

FIG 2. *$m(x) = \rho x$, with $\rho = 0.9, n = 300$ and a percentage of censoring* $= 0\%,\ 20\%$ *and* 40%, *respectively.*

Here we use the mean squarre error of the estimator with respect to the theoretical value. For each values n and ρ we replicate $B = 100$ times and taking the median over $x \in [-2,\ 2]$, that is

$$\underset{x\in[-2,\ 2]}{med}\left[\frac{1}{100}\sum_{i=1}^{100}\mathbb{E}\left[m_n(x) - m(x)\right]^2\right],$$

which are consigned in the following table.

It is clear that the high censoring and the dependency (for high values of ρ) affect slitghly the accuracy which is predictable but it remains all in all satisfactory.

We also consider two nonlinear cases with a percentage of censoring $= 15\%$ and $\rho = 0.9$

$$Y_i \quad = \quad \sin\left(\frac{\pi}{2}X_i\right), \qquad \text{sinus case,} \tag{5}$$

$$Y_i \quad = \quad \frac{5}{12}X_{i+1}^2 - 0.15, \qquad \text{parabolic case.} \tag{6}$$

percentage of censoring	$n = 50$			$n = 100$		
	$\rho = 0.9$	$\rho = 0.3$	$\rho = 0.1$	$\rho = 0.9$	$\rho = 0.3$	$\rho = 0.1$
0%	0.180	0.070	0.059	0.099	0.045	0.041
20%	0.484	0.086	0.064	0.365	0.073	0.035
40%	0.648	0.094	0.063	0.577	0.088	0.053

percentage of censoring	$n = 300$		
	$\rho = 0.9$	$\rho = 0.3$	$\rho = 0.1$
0%	0.043	0.016	0.010
20%	0.305	0.049	0.021
40%	0.521	0.089	0.018

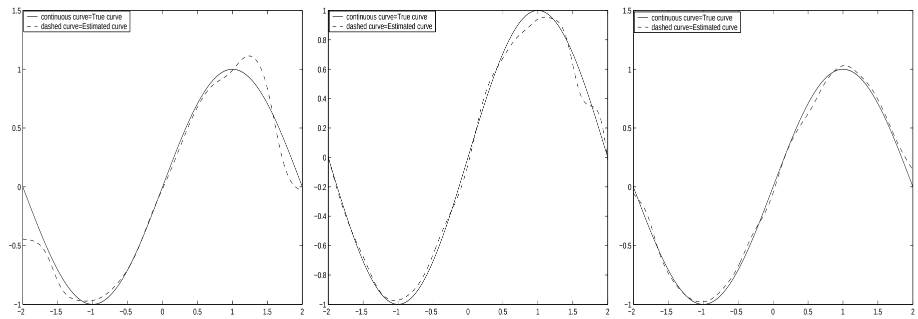

FIG 3. $m(x) = \sin \frac{\pi}{2} x$, *with* $n = 50$, 100 *and* 300, *respectively.*

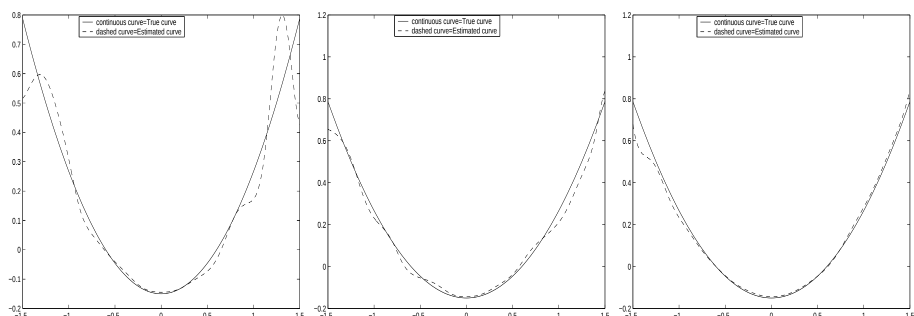

FIG 4. $m(x) = \frac{5}{12}\rho^2 x^2 + \frac{5}{12}(1-\rho^2) - 0.15$, *with* $n = 50$, 100 *and* 300, *respectively.*

Then we have $m(x) = \sin(\frac{\pi}{2}x)$ for (5) and $m(x) = \frac{5}{12}\rho^2 x^2 + \frac{5}{12}(1-\rho^2) - 0.15$ for (6).

Figures 3 and 4 show that the quality of fit for the non linear model is as good as in the linear model. Furthermore, we see that the quality is better when n increase.

5. Proofs

We split the proof of the Theorem 3.1 in the following Lemmata.

Lemma 5.1. *Under Assumptions **A1 A2** i) and **A3**, for n large enough:*

$$\sup_{x\in\mathcal{C}} |\mathbb{E}(\tilde{r}_{1,n}(x)) - r_1(x)| = O(h_n) \qquad a.s.$$

Proof of Lemma 5.1. Observe that

$$\begin{aligned}
\mathbb{E}\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}|X_1 = u\right) &= \mathbb{E}\left[\mathbb{E}\left[\frac{\mathbb{1}_{\{Y_1\le C_1\}}Y_1}{\bar{G}(Y_1)}|Y_1\right] \mid X_1 = u\right] \\
&= \mathbb{E}\left[\frac{Y_1}{\bar{G}(Y_1)}\mathbb{E}\left[\mathbb{1}_{\{Y_1\le C_1\}}|Y_1\right] \mid X_1 = u\right] \\
&= \mathbb{E}\left[Y_1|X_1 = u\right] \\
&= m(u).
\end{aligned}$$

Then, we have from (3)

$$\begin{aligned}
\mathbb{E}(\tilde{r}_{1,n}(x)) - r_1(x) &= \mathbb{E}\left(\frac{1}{h_n^d}\frac{\delta_1 T_1}{\bar{G}(T_1)}K_d\left(\frac{x - X_1}{h_n}\right)\right) - r_1(x) \\
&= \mathbb{E}\left(\frac{1}{h_n^d}K_d\left(\frac{x - X_1}{h_n}\right)\mathbb{E}\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}|X_1\right)\right) - r_1(x) \\
&= \int_{\mathbb{R}^d}\frac{1}{h_n^d}K_d\left(\frac{x-u}{h_n}\right)m(u)\ell(u)du - r_1(x) \\
&= \int_{\mathbb{R}^d}K_d(t)\left[r_1(x - h_n t) - r_1(x)\right]dt
\end{aligned}$$

since $r_1 = m\ell$.

A Taylor expansion gives

$$r_1(x - h_n t) - r_1(x) = -h_n\big(t_1\frac{\partial r_1}{\partial x_1}(x') + \cdots + t_d\frac{\partial r_1}{\partial x_d}(x')\big)$$

where x' is between $x - h_n t$ and x. Then

$$\begin{aligned}
\sup_{x\in\mathcal{C}}|\mathbb{E}(\tilde{r}_{1,n}(x)) - r_1(x)| &= \sup_{x\in\mathcal{C}}\left|\int_{\mathbb{R}^d}K_d(t)\left[r_1(x - h_n t) - r_1(x)\right]dt\right| \\
&\le h_n\sup_{x\in\mathcal{C}}\int_{\mathbb{R}^d}\left|K_d(t)\big(t_1\frac{\partial r_1}{\partial x_1}(x') + \cdots + t_d\frac{\partial r_1}{\partial x_d}(x')\big)dt\right|.
\end{aligned}$$

Then Assumptions **A1, A2** $i)$ and **A3**, give the result. □

Now, we introduce the following lemma (Ferraty and Vieu (2006) Proposition A.11 ii), p. 237).

Lemma 5.2. *Let $\{U_i, i \in \mathbb{N}\}$ be a sequence of real random variables, with strong mixing coefficient $\alpha(n) = O(n^{-\nu})$, $\nu > 1$, such that $\forall n \in \mathbb{N}, \forall i \in \mathbb{N}, 1 \leq i \leq n \quad |U_i| < +\infty$. Then for each $\varepsilon > 0$ and for each $q > 1$:*

$$\mathbb{P}\left\{ \left|\sum_{i=1}^{n} U_i\right| > \varepsilon \right\} \leq C\left(1 + \frac{\varepsilon^2}{qS_n^2}\right)^{-\frac{q}{2}} + nCq^{-1}\left(\frac{2q}{\varepsilon}\right)^{\nu+1}$$

where $S_n^2 = \sum_i \sum_j |cov(U_i, U_j)|$.

Lemma 5.3. *Under Assumptions **A1-A6**, we have*

$$\sup_{x \in \mathcal{C}} |\tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x))| = O\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)} \log n}\right) \quad a.s \quad as \ n \to \infty.$$

Proof of Lemma 5.3. $\mathcal{C}$ is a compact set, then it admits a covering $\mathfrak{S}$ by a finite number $\mathfrak{s}_n$ of balls $\mathcal{B}_k(x_k^*, a_n^d)$ centered at $x_k^* = (x_{1,k}^*, \dots x_{d,k}^*)$, $k \in \{1, \dots, \mathfrak{s}_n\}$. Then for all $x \in \mathcal{C}$ there exists $k \in \{1, \dots, \mathfrak{s}_n\}$ such that $\|x - x_k^*\| \leq a_n^d$, where a_n verifies $a_n^{d\gamma} = h_n^{d(\gamma+\frac{1}{2})} n^{-\frac{d}{2}}$, ($\gamma$ is the same as in Assumption **A2** *iii*)). Since $\mathcal{C}$ is bounded there exists a constant $M > 0$ such that $\mathfrak{s}_n \leq \frac{M}{a_n^d}$.

Now we set, for $x \in \mathcal{C}$:

$$\Delta_i(x) = \frac{1}{nh_n^d} \frac{\delta_i T_i}{\bar{G}(T_i)} K_d\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}\left(\frac{1}{nh_n^d} \frac{\delta_1 T_1}{\bar{G}(T_1)} K_d\left(\frac{x - X_1}{h_n}\right)\right).$$

It is obvious that

$$\sum_{i=1}^{n} \Delta_i(x) = \tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x)).$$

Writing $\Delta_i(x) - \Delta_i(x_k^*) = \tilde{\Delta}_i(x)$, we have clearly $|\Delta_i(x)| \leq \left|\tilde{\Delta}_i(x)\right| + |\Delta_i(x_k^*)|$. Then,

$$\begin{aligned}
\sup_{x \in \mathcal{C}} \left|\sum_{i=1}^{n} \tilde{\Delta}_i(x)\right| \leq\ & \sup_{x \in \mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i |T_i|}{\bar{G}(T_i)} \frac{1}{h_n^d} \left| K_d\left(\frac{x - X_i}{h_n}\right) - K_d\left(\frac{x_k^* - X_i}{h_n}\right)\right| \right\} \\
& + \sup_{x \in \mathcal{C}} \left\{ \mathbb{E}\left(\frac{\delta_1 |T_1|}{\bar{G}(T_1)} \frac{1}{h_n^d} \left| K_d\left(\frac{x - X_1}{h_n}\right) - K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right|\right) \right\}.
\end{aligned}$$

From Assumption **A2** *iii*)

$$\begin{aligned}
\sup_{x \in \mathcal{C}} \left|\sum_{i=1}^{n} \tilde{\Delta}_i(x)\right| \leq\ & \sup_{x \in \mathcal{C}} \left(\frac{2\mathbb{E}(|Y_1|)}{\bar{G}(\tau_F)} \frac{1}{h_n^d} \left\|\frac{x - x_k^*}{h_n}\right\|^{\gamma}\right) \leq \frac{\mathbb{E}(|Y_1|)}{\bar{G}(\tau_F)} \frac{a_n^{d\gamma}}{h_n^{\gamma+d}} \\
\leq\ & \frac{\mathbb{E}(|Y_1|)}{\bar{G}(\tau_F)} \frac{h_n^{d(\gamma+\frac{1}{2})} n^{-\frac{d}{2}}}{h_n^{\gamma+d}} \leq \frac{C}{\sqrt{nh_n^d}} h_n^{\gamma(d-1)}.
\end{aligned}$$

Assumption **A1** *i*) implies that $\sup_{x \in \mathcal{C}} \left|\sum_{i=1}^{n} \tilde{\Delta}_i(x)\right| = O\left(\frac{1}{\sqrt{nh_n^d}}\right) \quad a.s.$

On the other hand, let $U_i = nh_n^d \Delta_i(x_k^*)$. In order to apply Lemma 5.2, we have to calculate S_n^2. It is clear that

$$S_n^2 = \sum_{\substack{i \\ i \neq j}} \sum_j |cov(U_i, U_j)| + nVar(U_1).$$

We have

$$\begin{aligned} Var(U_1) &= \mathbb{E}\left[\frac{\delta_1^2 T_1^2}{\bar{G}^2(T_1)} K_d^2\left(\frac{x_k^* - X_1}{h_n}\right)\right] - \mathbb{E}^2\left[\frac{\delta_1 T_1}{\bar{G}(T_1)} K_d\left(\frac{x_k^* - X_1}{h_n}\right)\right] \\ &=: \mathcal{I}_1 - \mathcal{I}_2. \end{aligned}$$

Using the conditional expectation properties and a change of variables, we get

$$\begin{aligned} \mathcal{I}_1 &= \mathbb{E}\left[\frac{\delta_1^2 T_1^2}{G^2(T_1)} K_d^2\left(\frac{x_k^* - X_1}{h_n}\right)\right] \\ &= \mathbb{E}\left[K_d^2\left(\frac{x_k^* - X_1}{h_n}\right) \mathbb{E}\left(\frac{\delta_1^2 T_1^2}{G^2(T_1)} | X_1\right)\right] \\ &\leq \frac{h_n^d}{\bar{G}(\tau_F)} \int_{\mathbb{R}^d} K_d^2(t) r_2(x_k^* - h_n t) dt. \end{aligned}$$

By a Taylor expansion around x_k^*, under Assumptions **A2** *ii*) and **A4**, we obtain

$$\mathcal{I}_1 = O(h_n^d).$$

From Assumption **A3**,

$$\begin{aligned} \mathcal{I}_2 &= \mathbb{E}^2\left[K_d\left(\frac{x_k^* - X_1}{h_n}\right) \mathbb{E}\left[\frac{\delta_1 T_1}{\bar{G}(T_1)} | X_1\right]\right] \\ &= \left[\int_{\mathbb{R}^d} K_d\left(\frac{x_k^* - u}{h_n}\right) r_1(u) dt\right]^2 \\ &= O(h_n^{2d}). \end{aligned}$$

Finally $Var(U_1) = O(h_n^d)$.

Now let $S_n^{2*} = \sum_{\substack{i \\ i \neq j}} \sum_j |cov(U_i, U_j)|$, a direct calculus of $|cov(U_i, U_j)|$ gives

$$\begin{aligned} |cov(U_i, U_j)| &= |\mathbb{E}(U_i U_j)| \\ &= \left|\mathbb{E}\left\{\left[\frac{\delta_i T_i}{\bar{G}(T_i)} K_d\left(\frac{x_k^* - X_i}{h_n}\right) - \mathbb{E}\left(\frac{\delta_i T_i}{\bar{G}(T_i)} K_d\left(\frac{x_k^* - X_i}{h_n}\right)\right)\right]\right.\right. \\ &\quad \left.\left. \times \left[\frac{\delta_j T_j}{\bar{G}(T_j)} K_d\left(\frac{x_k^* - X_j}{h_n}\right) - \mathbb{E}\left(\frac{\delta_j T_j}{\bar{G}(T_j)} K_d\left(\frac{x_k^* - X_j}{h_n}\right)\right)\right]\right\}\right| \\ &\leq \left|\mathbb{E}\left(\frac{Y_i}{\bar{G}(Y_i)} K_d\left(\frac{x_k^* - X_i}{h_n}\right) \frac{Y_j}{\bar{G}(Y_j)} K_d\left(\frac{x_k^* - X_j}{h_n}\right)\right)\right| \\ &\quad + \left|\mathbb{E}\left(\frac{Y_i}{\bar{G}(Y_i)} K_d\left(\frac{x_k^* - X_i}{h_n}\right)\right) \mathbb{E}\left(\frac{Y_j}{\bar{G}(Y_j)} K_d\left(\frac{x_k^* - X_j}{h_n}\right)\right)\right| \\ &\leq C h_n^{2d} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_d(z) K_d(t) \\ &\quad \times \left[\ell_{ij}\left(x_k^* - z h_n, x_k^* - t h_n\right) + \ell_i(x_k^* - z h_n) \ell_j(x_k^* - t h_n)\right] dz dt. \end{aligned}$$

Assumption **A5** gives

$$|cov(U_i, U_j)| = O(h_n^{2d}). \tag{7}$$

On the other hand, from a result in Bosq (1998, p. 22), we have

$$|cov(U_i, U_j)| \leq C\alpha\left(|i-j|\right). \tag{8}$$

Then to evaluate S_n^{2*} the idea is to introduce a sequence of integers w_n which we will precise below. Then we use (7) for the close i and j and (8) otherwise. That is

$$\begin{aligned} S_n^{2*} &= \sum\sum_{\{0<|i-j|\leq w_n\}} |cov(U_i, U_j)| + \sum\sum_{\{|i-j|>w_n\}} |cov(U_i, U_j)| \\ &\leq C \sum\sum_{\{0<|i-j|\leq w_n\}} h_n^{2d} + C\sum\sum_{\{|i-j|>w_n\}} \alpha\left(|i-j|\right) \\ &\leq C\left(nh_n^{2d}w_n\right) + Cn^2\alpha\left(w_n\right). \end{aligned}$$

Now choosing $w_n = \left[\frac{1}{h_n^d}\right] + 1$, we have $S_n^{2*} \leq O(nh_n^d) + Cn^2\alpha\left(\frac{1}{h_n^d}\right)$. The mixing coefficient yields $n^2\alpha\left(\frac{1}{h_n^d}\right) = O(n^2h_n^{d\nu})$.

So

$$S_n^{2*} = O(nh_n^d) + O(n^2h_n^{d\nu}).$$

Finally, we have

$$S_n^2 = S_n^{2*} + nVar(U_1) = O(nh_n^d) + O(n^2h_n^{d\nu}).$$

Then, for $\varepsilon > 0$, applying Lemma 5.2, we have

$$\begin{aligned} \mathbb{P}\left\{\left|\sum_{i=1}^n \Delta_i(x^{*k})\right| > \varepsilon\right\} &= \mathbb{P}\left\{\left|\sum_{i=1}^n U_i\right| > nh_n^d\varepsilon\right\} \\ &\leq C\left(1 + C\frac{\varepsilon^2 nh_n^d}{q(1+nh_n^{d(\nu-1)})}\right)^{-\frac{q}{2}} + nCq^{-1}\left(\frac{q}{\varepsilon nh_n^d}\right)^{\nu+1}. \end{aligned} \tag{9}$$

If we replace ε by $\varepsilon_0\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right) =: \varepsilon_n$ for all $\varepsilon_0 > 0$ in (9), we get

$$\begin{aligned} &\mathbb{P}\left\{\left|\sum_{i=1}^n \Delta_i(x^{*k})\right| > \varepsilon_n\right\} \\ &\leq C\left(1 + C\frac{\varepsilon_0^2\log n}{q}\right)^{-\frac{q}{2}} + nCq^{-1}\left(\frac{q}{\varepsilon_0\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right)}\right)^{\nu+1}. \end{aligned} \tag{10}$$

By choosing $q = (\log n)^{1+b} \quad (b > 0)$, (10) becomes

$$\begin{aligned}
\mathbb{P}\left\{\left|\sum_{i=1}^{n}\Delta_i(x^{*k})\right| > \varepsilon_n\right\} &\leq Cn^{-C\varepsilon_0^2} + nCq^{-1}\left(\frac{q}{\varepsilon_0}\right)^{\nu+1}\left(nh_n^d\log n\right)^{-\frac{\nu+1}{2}} \\
&\leq Cn^{-C\varepsilon_0^2} + C\varepsilon_0^{-(\nu+1)}\left(\log n\right)^{\nu(1+b)} n^{1-\frac{\nu+1}{2}} h_n^{-\frac{d(\nu+1)}{2}}.
\end{aligned}$$

Now we can write

$$\begin{aligned}
&\mathbb{P}\left\{\max_{k=1,\ldots,\mathfrak{s}_n}\left|\sum_{i=1}^{n}\Delta_i(x^{*k})\right| > \varepsilon_0\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right)\right\} \\
&\leq \sum_{i=1}^{\mathfrak{s}_n}\mathbb{P}\left\{\left|\sum_{i=1}^{n}\Delta_i(x^{*k})\right| > \varepsilon_n\right\} \\
&\leq Ma_n^{-d}\left(Cn^{-C\varepsilon_0^2} + C\varepsilon_0^{-(\nu+1)}\left(\log n\right)^{\nu(1+b)} n^{1-\frac{\nu+1}{2}} h_n^{-\frac{d(\nu+1)}{2}}\right) \\
&\leq CMh_n^{-d(1+\frac{1}{2\gamma})} n^{\frac{d}{2\gamma}-C\varepsilon_0^2} \\
&+ MC\varepsilon_0^{-(\nu+1)}\left(\log n\right)^{\nu(1+b)} n^{1-\frac{\nu+1}{2}+\frac{d}{2\gamma}} h_n^{-\frac{d(\nu+1)}{2}-\frac{d}{2\gamma}-d} \\
&=: CMJ_1 + MC\varepsilon_0^{-(\nu+1)}J_2. \qquad (11)
\end{aligned}$$

We have from Assumption **A6**

$$\begin{aligned}
J_2 &\leq C\left(\log n\right)^{\nu(1+b)} n^{1-\frac{\nu+1}{2}+\frac{d}{2\gamma}} n^{-\frac{(3-\nu)}{2}-\theta d\left(\frac{\gamma(\nu+1)+2\gamma+1}{2\gamma}\right)} \\
&\leq C\left(\log n\right)^{\nu(1+b)} n^{-1-\theta d\left(\frac{\gamma(\nu+1)+2\gamma+1-\frac{1}{\theta}}{2\gamma}\right)}.
\end{aligned}$$

Then, for an appropriate choice of θ, J_2 is the general term of a convergent series. In the same way, $J_1 \leq n^{\varsigma - C\varepsilon_0^2}$ and we can choose ε_0 such that J_1 is the general term of convergent series. Finally, applying Borel-Cantelli lemma, to (11) gives the result. □

Remark 5.1. We point out that the parameter θ of Assumption **A6** can be chosen such as:

$$\theta > \frac{1}{\gamma(\nu+1)+2\gamma+1}.$$

This condition ensures the convergence of the series of Lemma 5.3.

Lemma 5.4. *Under Assumptions **A1-A2** and **A5-A7**,*

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

Proof of Lemma 5.4. We have

$$\sup_{x\in\mathcal{C}}\left|\ell(x) - \ell_n(x)\right| \leq \sup_{x\in\mathcal{C}}\left|\ell_n(x) - \mathbb{E}\left(\ell_n(x)\right)\right| + \sup_{x\in\mathcal{C}}\left|\mathbb{E}\left(\ell_n(x)\right) - \ell(x)\right|.$$

By Assumptions **A1**–**A2**, **A5**–**A7**, by an analogous proof to that of Lemma 5.3 without censoring (that is $\bar{G}(T_i) = 1$, $\delta_i = 1$ and $Y_i = 1$) and putting $\varepsilon = \varepsilon_0\big(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\big)$ we get

$$\sup_{x\in\mathcal{C}} |\ell_n(x) - \mathbb{E}(\ell_n(x))| = O\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(\nu-2)}\log n}\right). \tag{12}$$

Furthermore, under **A2** *i*) and **A7** and using a Taylor expansion, we get

$$\sup_{x\in\mathcal{C}} |\mathbb{E}(\ell_n(x)) - \ell(x)| = O(h_n)$$

which permit us to conclude. □

Lemma 5.5. *Under Assumptions **A1**–**A2**, **A5**–**A7**, we have*

$$\sup_{x\in\mathcal{C}} |r_{1,n}(x) - \tilde{r}_{1,n}(x)| = o(h_n) \quad a.s \qquad as\ \ n \to \infty.$$

Proof of Lemma 5.5. We have from (3) and (4)

$$\begin{aligned}
&|r_{1,n}(x) - \tilde{r}_{1,n}(x)| \\
&\quad = \frac{1}{nh_n^d}\left|\sum_{i=1}^n \frac{\mathbb{1}_{\{Y_i<C_i\}}Y_i}{\bar{G}_n(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right) - \frac{\mathbb{1}_{\{Y_1<C_1\}}Y_i}{\bar{G}(Y_i)}K_d\left(\frac{x-X_i}{h_n}\right)\right| \\
&\quad \le \frac{1}{nh_n^d}\left|\sum_{i=1}^n Y_iK_d\left(\frac{x-X_i}{h_n}\right)\frac{\bar{G}(Y_i)-\bar{G}_n(Y_i)}{\bar{G}_n(Y_i)\bar{G}(Y_i)}\right| \\
&\quad \le \frac{1}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)}\sup_{t\le\tau_F}\left(\left|\bar{G}_n(t)-\bar{G}(t)\right|\right)\frac{1}{nh_n^d}\sum_{i=1}^n |Y_i|\,K_d\left(\frac{x-X_i}{h_n}\right).
\end{aligned}$$

In the same way as for Theorem 2 of Cai (2001), it can be shown under **A1** that

$$\sup_{t\le\tau_F}\left(\left|\bar{G}_n(t)-\bar{G}(t)\right|\right) = O\left(\sqrt{\frac{\log\log n}{n}}\right) \quad a.s.$$

Furthermore, from the definition of $\ell_n(x)$, Lemma 5.4, Assumptions **A1, A2** and **A7**, and the fact that Y is bounded we get the result. □

Proof of Theorem 3.1. We have

$$\begin{aligned}
&\sup_{x\in\mathcal{C}}|m_n(x) - m(x)| \\
&\quad \le \sup_{x\in\mathcal{C}}\left|\frac{r_{1,n}(x)}{\ell_n(x)} - \frac{\tilde{r}_{1,n}(x)}{\ell_n(x)}\right| + \left|\frac{\tilde{r}_{1,n}(x)}{\ell_n(x)} - \frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)}\right| \\
&\qquad + \left|\frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{\ell_n(x)} - \frac{r_1(x)}{\ell_n(x)}\right| + \left|\frac{r_1(x)}{\ell_n(x)} - \frac{r_1(x)}{\ell(x)}\right|
\end{aligned}$$

$$\leq \frac{1}{\inf \ell_n(x)} \Big\{ \sup_{x\in\mathcal{C}} |r_{1,n}(x) - \tilde{r}_{1,n}(x)| + \sup_{x\in\mathcal{C}} |\tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x))| + \sup_{x\in\mathcal{C}} |\mathbb{E}(\tilde{r}_{1,n}(x) - r_1(x))| + \sup_{x\in\mathcal{C}} \left(|r_1(x)|\, \xi^{-1}\right) \sup_{x\in\mathcal{C}} |\ell(x) - \ell_n(x)| \Big\} \quad (13)$$

The kernel estimator $\ell_n(x)$ is almost surely bounded away from 0 because of Lemma 5.4 and Assumption **A7**.

Then (13) in conjunction with Lemmas 5.1, 5.3, 5.4 and 5.5 we conclude the proof. □

Acknowledgments

The authors are grateful to the associate editor and the two anonymous referees for their careful reading and appropriate remarks which gave them the opportunity to improve the rate of convergence, the simulation part and the quality of the paper.

References

Bollerslev, T. (1986). General autoregressive conditional heteroskedasticity. *J. Economt.* **31**:307–327. MR0853051

Bosq, D. (1998). Nonparametric statistics for stochastic processes. Estimation and Prediction. *Lecture Notes in Statistics*, **110**, Springer-Verlag, New York. MR1640691

Bradley, R.D. (2007). *Introduction to strong mixing conditions.* Vol I-III, Kendrick Press, Utah.

Cai, J., Prentice, R.L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. MR1332846

Cai, Z. (1998a). Asymptotic properties of Kaplan-Meier estimator for censored dependent data. *Statist. & Probab. Lett.* **37**:381–389. *Biometrika.* **82**:151–164. MR1624415

Cai, Z. (1998b). Kernel density and hazard rate estimation for censored dependent data. *J. Multivariate Anal.* **67**:23–34. MR1659104

Cai, Z. (2001). Estimating a distribution function for censored time series data. *J. Multivariate Anal.* **78**:299–318. MR1859760

Cai, Z. (2003). Weighted local linear approach to censored nonparametric regression. In *Recent advances and trends in nonparametric staistics* (eds M. G. Atrikas & D.M. Politis), Elsevier, Amesterdam, the Netherlands; San Diego, CA, 217–231.

Cai, Z., Ould Saïd, E. (2003). Local M-estimator for nonparametric time series. *Statist & Probab. Lett.* **65**:433–449. MR2039887

Carbonez, A., Györfi, L., Van der Meulin, E.C. (1995). Partition-estimate of a regression function under random censoring. *Statist. & Decisions*, **13**:21–37. MR1334076

Collomb, G. (1981). Estimation non-paramétrique de la régression: Revue Bibliographique. *Internat. Statist. Rev.* **49**:75–93. MR0623011

Doukhan, P. (1994). *Mixing: Properties and examples. Lecture Notes in Statistics*, **85**, Springer-Verlag, New York. MR1312160

El-Ghouch, A., Van Keilegom, I. (2008). Non-parametric regression with dependent censored data. *Scand. J. Statist.* **35**:228–247. MR2418738

Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* **50**:987–1007. MR0666121

Ferraty, F., Vieu, P. (2006). *Nonparametric Functionnal Data Analysis. Theory and Practice.* Springer-Verlag, New York. MR2229687

Gonzalez-Manteiga, W., Quintela-del-Rio, A., Vieu, P. (2002). A note on variable selection in nonprametric regression with dependent data. *Statist. & Probab. Lett.* **57**:259–268. MR1912084

Guessoum, Z., Ould Saïd, E. (2008). On the nonpararametric estimation of the regression function under censorship model. *Statist. & Decisions* **26**:159–177. MR2512266

Härdle, W. (1990). *Applied nonparametric regression.* Cambridge University Press. MR1161622

Jones, D.A. (1978). Nonlinear autoregressive processes. *Proc. Roy. Soc. London A* **360**, 71–95. MR0501672

Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**:457–481. MR0093867

Liebscher, E. (2001). Estimation of the density and the regression function under mixing condition. *Statist. & Decisions* **19**:9–26. MR1817218

Liebscher, E. (2002). Kernel density and hazard rate estimation for censored data under $\alpha-$mixing condition. *Ann. Inst. Statist. Math.* **34**:19–28. MR1893539

Lipsitz, S.R., Ibrahim, J.G. (2000). Estimation with correlated censored survival data with missing covariates. *Biostatistics.* **1**: 315–327.

Masry, E., Fan, J. (1997). Local polynomial estimation of regression function for mixing processes. *Scand. J. Statist.* **24**:165–179. MR1455865

Ozaki, T. (1979). Nonlinear time series models for nonlinear random vibrations. Technical report. Univ. of Manchester. MR0557437

Robinson, P.M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4**:185–297. MR0732897

Silverman, B.W. (1986). *Estimation for statistics and data analysis. Monographs on Statist. and Appl. Probab.* Chapman & Hall. London. MR0848134

Tae, Y.K., Cox, D.D. (1996). Uniform strong consistency of kernel density estimator under dependence. *Statist. & Probab. Lett.* **26**:179–185. MR1381469

Wand, M.P., Jones, M.C. (1995). *Kernel smoothing.* Chapman & Hall. London. MR1319818

Wahba, G. (1990). *Spline models for observational data.* SIAM, Philadelphia. MR1045442

Wei, L.J., Lin, D.Y. (1989). Regression analysis of multivariate incomplete failure times by modelling marginal distribution. *J. Amer. Statist. Assoc.* **84**: 1064–1073.