ISSN 1066-5307, Mathematical Methods of Statistics, 2019, Vol. 28, No. 1, pp. 39–56. © Allerton Press, Inc., 2019.

A Semi-Parametric Mode Regression with Censored Data

S. Khardani[1*]

[1] *Laborat. de Phys. Math., Fonctions spéciales et Appl., Hammam-Sousse, Tunisia*

Received June 1, 2018; in final form, December 3, 2018

Abstract—In this work we suppose that the random vector (X, Y) satisfies the regression model $Y = m(X) + \epsilon$, where $m(\cdot)$ belongs to some parametric class $\{m_\beta(\cdot)\colon \beta \in \mathbb{K}\}$ and the error ϵ is independent of the covariate X. The response Y is subject to random right censoring. Using a nonlinear mode regression, a new estimation procedure for the true unknown parameter vector β_0 is proposed that extends the classical least squares procedure for nonlinear regression. We also establish asymptotic properties for the proposed estimator under assumptions of the error density. We investigate the performance through a simulation study.

Keywords: asymptotic normality, censored data, nonlinear model regression, survival data, strong consistency, kernel smoothing, mode estimation, semi-parametric regression..

AMS 2010 Subject Classification: 62G20, 62G07, 62E20, 62N01.

DOI: 10.3103/S1066530719010034

1. INTRODUCTION

Nonlinear models are often used when analyzing a possibly censored survival time depending on a covariate. For example, in medical surveys, the relationship between the survival time and the age of a patient who has received a given treatment is often nonlinear and the survival time is subject to right censoring since the patient may decide to leave the study, die due to another cause than the disease from which he suffers or the study itself can be stopped. To extend the above relation to regression setting, let (X, Y) be a random vector, where X is a d-dimensional covariate and $Y (\in \mathbb{R})$ represents the response.

We suppose that Y is subject to random right censoring, i.e., instead of observing Y we only observe (Z, δ), where $Z = \min(Y, C)$, $\delta = \mathbf{1}_{\{Y \leqslant C\}}$ and C represents the censoring time, which is supposed to be independent of Y conditionally on X. Let (Y_i, C_i, X_i, Z_i) $(i = 1, \ldots, n)$ be n independent copies of (Y, C, X, Z). We assume that the relation between X and Y is given by:

$$Y = m(X, \beta_0) + \epsilon, \tag{1}$$

where $m(\cdot)$ is a known function, $\beta_0 = (\beta_1, \cdots, \beta_p)^T \in \mathbb{K}$ is an unknown $p \times 1$ parameter to be estimated, the error term ϵ is independent of the covariate X and $\mathbb{K}$ is a compact subset of $\mathbb{R}^p$.

This formulation includes both the conditional mean and conditional median (or more general quantile) regression models. In many cases, economic theory implies a particular functional form for an empirical model specification. An incorrect parametrization of the regression equation might result in inconsistent estimates. Sometimes the researcher might feel more confident about the functional form of some parts of the regression equation but be less confident about the form of the other parts. Combining the parametric and nonparametric techniques to yield the semi-parametric regression model could then help obtain consistent estimates of the parameters of interest.

In this paper, a new estimation procedure for the true unknown parameter vector β_0 is proposed that extends the classical least squares method (LSM) for nonlinear regression to the case where the response is subject to censoring. For that, we propose a semi-parametric modal regression estimator for the case in which the dependent variable has a continuous conditional density with a well-defined global mode. The estimator is semi-parametric in that the conditional mode is specified as a parametric

[*] E-mail: khardani_salah@yahoo.fr

function, but only mild assumptions are made about the nature of the conditional density of interest. We show that the proposed estimator is consistent and has a tractable asymptotic distribution.

Classically, the conditional mean $r(x) = \mathbb{E}(Y \mid X = x) = m(x, \beta_0)$ or median estimation are used to model the link function and the parameter β_0 is estimated by the least-squares method or any of its robust or weighted version. But it is not uncommon in many fields to encounter data distributions that are highly skewed (e.g., wages, prices, energy intake) with several peaks or contain outliers. Then other alternative statistics are necessary to model the link function. Among them, the mode of the condition distribution of $Y \mid X = x$ which leads to less bias (than r). This parameter will be called *mode regression* (MR). It can bring some helpful information to understand relationships between the covariate X and a response variable Y. MR may provide shorter prediction intervals than other regression approaches for a nominal confidence level. It is robust to outliers and is very justified in situations where conditional distributions are highly skewed. Mode regression is potentially a very useful addition to current data analysis tools. However, estimation of modal regression coefficients is not trivial. In this work we propose an expectation-maximization (EM) algorithm that minimizes a kernel-based objective function for estimating mode regression coefficients.

In the literature, model (1) has been thoroughly investigated for parametric/semi-parametric mean regression, where $m(\cdot)$ is characterized by a finite-dimensional parameter and $\mathbb{E}(\epsilon \mid X) = 0$. In this context and using the ordinary least squares (OLS) method, Yatchew [32] estimates the relationship between variable costs of distributing electricity per customer as a nonlinear function of the scale of operation as measured by the number of customers.

Semi-parametric regression models are less studied but are extremely useful due to their flexibility to accommodate non-linearity and to circumvent curse of dimensionality [9, 24, 32, 15]. In particular, we consider the general setup with $m(x) = m(x, \beta_0)$, where β_0 is a finite-dimensional parameter. The main interest is often in making inference about β_0.

In partially linear model, Severini and Staniswalis [25] outlined a method for estimating the parameter β_0 of this type of semi-parametric model using a quasi-likelihood function. Algorithms for computing the estimates are given and the asymptotic distribution theory for the estimators is developed.

Liang and Härdle [20] considered a simple modification of the last estimator and derived its asymptotic distribution theory. For nonrandom design Jennrich [10] proved strong consistency of the least-squares estimator and derived its limit distribution. Liang, Härdle and Carroll [21] have studied heteroscedastic partially linear mean regression models using a quasi-likelihood function. Lee [17] introduced a semi-parametric method and used a uniform kernel to estimate mode regression coefficients based on a loss function.

In fact, in complete data, nonparametric estimation of mode had been discussed in decades (see [23, 6, 7]. Shoung and Zhang [26] studied the least squares estimators of the mode and Ziegler [35] proved its asymptotic normality).

In incomplete data, a number of extensions to censored data of the least squares procedure for estimating $\beta_1, \ldots, \beta_p$ have been studied in the literature. The list of first-generation estimators includes, e.g., Müller [22] who studied a least squares regression, Buckley and James [3] who gave a definition of the β estimator using a mean regression and studied its asymptotic properties, Koul, Susarla, Van Ryzin [16] and Leurgans [19] who studied a linear model using a synthetic data, while more recent contributions have been made in [34, 27, 1, 2, 29]. Recently, Khardani et al. [13–15] established strong uniform convergence with a rate for the kernel estimator under random censorship and stated its asymptotic normality.

In this paper, using a synthetic data, we study the estimation of the parameter β_0 (based on both the MR) using a kernel smoother. This paper is organized as follows. In Section 2, we introduce some notation and describe the estimation procedure in detail. In Section 3 we state asymptotic normality of the regression parameter estimators and the weak convergence results. In Section 4 we analyze the finite-sample performance of the proposed estimator via a simulation study, while the Appendix contains the proofs of the results of Section 5.

2. DESCRIPTION OF THE MODEL AND ESTIMATOR

Consider a randomly right-censored model given by two nonnegative stationary random sequences $Y_1, \dots, Y_n$ (survival times) and $C_1, \dots, C_n$ censoring times. Assume that the latter are i.i.d. and independent of the survival times $(Y_i)_{1 \le i \le n}$. For any distribution function (df) L, let $\tau_L = \sup\{t, L(t) < 1\}$ be the right endpoint of its support.

Further, we will denote by $H(\cdot)$ (resp. $G(\cdot)$) the df of Y (resp. of C) and by τ_H (resp. τ_G) the upper endpoints of the survival function $\overline{H}$ (resp. of $\overline{G}$). In the following we assume that $\tau_H < \infty$, $\overline{G}(\tau_H) > 0$ and let $\tau < \min(\tau_H, \tau_G)$.

In this kind of model, it is well known that the empirical distribution is not a consistent estimator for the distribution function G. Therefore Kaplan and Meier [11] proposed a consistent estimator for the survival function $\bar{G} = 1 - G$ which is defined as

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^{n} \left(1 - \frac{1-\Delta_{(i)}}{n-i+1}\right)^{\mathbf{1}_{\{Z_{(i)} \le t\}}} & \text{if} \quad t < Z_{(n)}, \\ 0 & \text{otherwise,} \end{cases}$$

where $Z_{(1)} < Z_{(2)} < \dots < Z_{(n)}$ are the order statistics of $(Z_i)_{1 \le i \le n}$ and $\Delta_{(i)}$ is the concomitant of $Z_{(i)}$.

The purpose of this paper is to present a way to overcome this problem by imposing the following weak model assumption: we assume that the relation between Y_i and X_i is given by

$$Y_i = m(X_i, \beta_0) + \epsilon_i, \qquad (i = 1, \dots, n), \tag{2}$$

where $m \colon \mathbb{R}^d \times \mathbb{K} \longrightarrow \mathbb{R}$ is a known function measurable on $\mathbb{R}^d$ for each $\beta \in \mathbb{K}$ and continuous on $\mathbb{K}$ (a compact subset of $\mathbb{R}^p$, $p \leqslant d$); β_0 is an unknown $p \times 1$ vector to be estimated.

Now, there exists a loss function whose expectation is minimized at the conditional mode of Y given $X = x$. In model (2), we assume that: mode $(Y \mid X = x) = m(x, \beta_0) \Leftrightarrow$ mode $(\epsilon \mid X = x) = 0$.

Second, we recall that the model (2) suffers from censorship data. For that, we use the so-called "synthetic data" which allow us to take into account the censoring effect on the lifetime distribution.

For this purpose, we let

$$\varphi(Y_i^*) = \frac{\delta_i \varphi(Z_i)}{\overline{G}(Z_i)}, \quad 1 \le i \le n, \tag{3}$$

for any measurable function φ, where $\overline{G} = 1 - G$, $Z_i = \min(Y_i, C_i)$ and $\delta_i = \mathbf{1}_{\{Y_i < C_i\}}$.

Assuming a sequence of covariates is given, we observe the triplets $(Z_i, \delta_i, X_i)_{1 \leqslant i \leqslant n}$. All along this paper we suppose that

$$(Y_i, X_i)_i \quad \text{and} \quad (C_i)_i \quad \text{are independent.} \tag{4}$$

Then from (3) and (4) we get

$$\begin{aligned} \mathbb{E}\big[\varphi(Y_1^*) \mid X_1\big] &= \mathbb{E}\left[\frac{\delta_1 \varphi(Z_1)}{\overline{G}(Z_1)} \mid X_1\right] = \mathbb{E}\left\{\mathbb{E}\left[\frac{\delta_1 \varphi(Z_1)}{\overline{G}(Z_1)} \mid Y_1, X_1\right] \mid X_1\right\} \\ &= \mathbb{E}\left\{\mathbb{E}\left[\frac{\delta_1 \varphi(Y_1)}{\overline{G}(Y_1)} \mid Y_1, X_1\right] \mid X_1\right\} = \mathbb{E}\left\{\frac{\varphi(Y_1)}{\overline{G}(Y_1)} \mathbb{E}[\mathbf{1}_{\{Y_1 \le C_1\}} \mid Y_1] \mid X_1\right\} \\ &= \mathbb{E}(\varphi(Y_1) \mid X_1). \end{aligned} \tag{5}$$

In order to take in account the censoring phenomenon, the idea in Lee [17, 18] for complete data is adapted. Using (4) and (5) we have

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{Y_1 \le C_1\}} K_0\left(\frac{Z_1 - m(X_1, \beta)}{h_n}\right)}{\overline{G}(Z_1)} \mid X_1\right] = \mathbb{E}\left[K_0\left(\frac{Y_1 - m(X_1, \beta)}{h_n}\right) \mid X_1\right], \tag{6}$$

where $K_0(\cdot)$ denotes a smooth kernel function and h_n a bandwidth.

From (6), using a mode regression to estimate the parameter β_0, we propose maximizing the kernel based objective function

$$\bar{S}_n(\beta) := \frac{1}{nh_n}\sum_{i=1}^{n} \mathbf{1}_{\{T_i \le C_i\}} \bar{G}^{-1}(Z_i) K_0\biggl(\frac{Z_i - m(X_i,\beta)}{h_n}\biggr). \tag{7}$$

In practice $\bar{G}(\cdot)$ is unknown, hence it is replaced by its Kaplan–Meier estimate $\bar{G}_n(\cdot)$. Therefore the feasible estimator of $\bar{S}_n$ is given by

$$\hat{S}_n(\beta) := \frac{1}{nh_n}\sum_{i=1}^{n} \delta_i \bar{G}_n^{-1}(Z_i) K\biggl(\frac{Z_i - m(X_i,\beta)}{h_n}\biggr). \tag{8}$$

Then a natural estimator of β_0 is

$$\hat{\beta}_n = \arg\max_{\beta} \hat{S}_n(\beta). \tag{9}$$

3. ASSUMPTIONS AND MAIN RESULTS

Throughout the paper, when no confusion is possible, we denote by M and/or C any generic positive constant and by

$$\|K_0\|_\infty = \sup_{t\in\mathbb{R}} K_0(t), \quad \|K_0\|_2^2 = \int_{-\infty}^{\infty} K_0^2(t)\,dt, \quad K_0'(t) = \frac{\partial K_0}{\partial t}(t), \quad \hat{S}_n^{(j)}(\beta) = \frac{\partial^j \hat{S}_n}{\partial^j \beta}(\beta),$$

$$\dot{m}(x,\beta) = \frac{\partial m}{\partial \beta}(x,\beta), \quad \ddot{m}(x,\beta) = \frac{\partial^2 m}{\partial^2 \beta}(x,\beta), \quad g_{T|X}^{(i)}(t \mid x) = \frac{\partial^i g_{T|X}}{\partial^i t}(t \mid x).$$

Let $(X_i, \epsilon_i)_{i\le 1}$ be a sequence of i.i.d. random variables. For any x denote by $g_{\epsilon|X}(\cdot \mid x)$ the conditional probability density function of ϵ_i given $X_i = x$ and assume that the covariate X has df $F(\cdot)$.

Now we give the assumptions needed to get our results.

For any sequences (u_n) and (v_n) we put $u_n = O(v_n)$ if $|u_n| \leqslant C|v_n|$ for all n (if the property holds in probability we use the symbol $O_{\mathbf{P}}$).

(A1): $\forall(\epsilon_1,\epsilon_2) \in \mathbb{R}\times\mathbb{R}, \forall(x_1,x_2) \in \mathbb{R}^d \times \mathbb{R}^d$,

$$|g_{\epsilon|X}(\epsilon_1 \mid x_1) - g_{\epsilon|X}(\epsilon_2 \mid x_2)| \le C\bigl(\|x_1 - x_2\| + |\epsilon_1 - \epsilon_2|\bigr).$$

(A2): The conditional density $g_{\epsilon|X}$ is differentiable up to order 3,

- (i) $\sup_{x,\epsilon} |g_{\epsilon|X}^{(j)}(\epsilon \mid x)| < \infty$ for $0 \le j \le 3$,
- (ii) $g_{\epsilon|X}(\epsilon \mid x) < g_{\epsilon|X}(0 \mid x)$ for all $\epsilon \ne 0$ and x.

(A3): The kernel K_0 is differentiable up to order 3 and:

- (i) $\displaystyle\int_{\mathbf{R}} tK_0(t)\,dt = 0$, (ii) $\displaystyle\int_{\mathbf{R}} t^2 K_0(t)\,dt < \infty$, (iii) $\displaystyle\lim_{|t|\to\infty} K_0(t) = 0$,
- (iv) $\displaystyle\sup_t |K_0^{(j)}(t)| < \infty$ for $0 \le j \le 3$,
- (v) $\displaystyle\int_{\mathbf{R}} K_0^{(i)}(t)\,dt < \infty$ for $0 \le i \le 2$.

(A4): The set

$$\mathcal{F}_0 = \Big\{ K_0\Big(\frac{t-\cdot}{h}\Big),\ t \in \mathbb{R}, h \in \mathbb{R}^* \Big\}$$

is a VC-class of measurable functions.

(A5): $\forall x \in \mathbb{R}^d$, $\forall (\beta_1, \beta_2) \in \mathbb{K}^2$, $|m(x,\beta_1) - m(x,\beta_2)| \leq f(x)\|\beta_1 - \beta_2\|$ for some integrable positive function f.

(A6): All partial derivatives of $m(x,\beta)$ with respect to x and the components of β of order 0, 1 or 2 exist and are continuous in (x,β) for all (x,β).

(A7): (i) $\mathbb{E}\big[\sup_{\beta\in\mathbb{K}}(\dot{m}(X_i,\beta))^{5+s}\big] < \infty, \quad s > 0,$

(ii) $\mathbb{E}\big[\sup_{\beta\in\mathbb{K}}|\ddot{m}(X_i,\beta)|\big] < \infty,$

(iii) For all $\epsilon > 0$, $\inf_{\|\beta-\beta_0\|>\epsilon} \mathbb{E}\big[\big(m(X_i,\beta) - m(X_i,\beta_0)\big)^2\big] > 0,$

(iv) $\mathbb{E}\big[\big(\dot{m}(X_i,\beta_0)\big)^2 \times g^{(2)}_{\epsilon|X}(0 \mid X_i)\big]$ is nonsingular.

(A8): (i) $\dfrac{nh_n^5}{\log n} \longrightarrow \infty,$

(ii) $nh_n^7 \longrightarrow 0 \quad \text{as} \quad n \to +\infty.$

Discussion of the Assumptions and Examples

- The independence assumption between $(C_n)_n$ and $(X_n, Y_n)_n$ in (4) may seem to be strong and one can think of replacing it by a classical conditional independence assumption between $(C_n)_n$ and $(Y_n)_n$ given $(X_n)_n$.
- Assumptions (A1), (A2 (i)), and (A3) are classical in nonparametric estimation.
- Assumption (A2 (ii)) is specific to the context of mode regression and requires that the conditional density of ϵ has a mode at 0. An example that satisfies condition (A2 (ii)) is the following (mixture of two Gaussian densities):

$$g_{\epsilon|X}(\epsilon \mid x) = \frac{1}{2\sqrt{2\pi}}\Big\{ \exp -\frac{1}{2}(\epsilon - x)^2 + \exp -\frac{1}{2}(\epsilon + x)^2 \Big\}.$$

  Here $g_{\epsilon|X}(\cdot \mid x)$ is symmetric about 0 and $\mathbb{E}(\epsilon \mid X = x) = \text{Mode}\,(\epsilon \mid X = x) = x + (-x) = 0$.
- Assumption (A4) is a consequence of Theorems 4.2.1 and 4.2.4 in [5]. This assumption is needed in order to use Talagrand's inequality.
- Assumptions (A5), (A6) and (A7) specify the model. Example: Let (X, Y) satisfy the model assumptions (A5)–(A7),

$$Y = m(X,\beta) + \epsilon,$$

  where $m(X,\beta) = \beta_0 + \beta_1 X + X^2$ and ϵ (with conditional density $g_{\epsilon|X}$) is independent of X.
- Assumption (A8) gives conditions for the bandwidth which allow to get the rate of convergence.

Theorem 3.1. *Under Assumptions* (A1)–(A7)*, we have*

$$\hat{\beta}_n^T \xrightarrow{\mathbf{P}} \beta_0^T \qquad as \quad n \longrightarrow \infty.$$

Corollary 3.2. *Under Assumptions* (A1)–(A7) *and* (A8 (i))*, we have*

$$\|\hat{\beta}_n - \beta_0\| = O_{\mathbf{P}}\left(\left(\frac{\log n}{nh_n}\right)^{1/4}\right) \qquad as \quad n \longrightarrow \infty, \tag{10}$$

where $\xrightarrow{\mathbf{P}}$ *denotes convergence in probability.*

Theorem 3.3. *Assume that* (A1)–(A8) *hold. We have*

$$\sqrt{nh_n^3}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega^2),$$

where $\xrightarrow{\mathcal{D}}$ *denotes convergence in distribution, where*

$$\Omega^2 = \Omega_1^{-1}\Omega_0\Omega_1^{-1}, \tag{11}$$

$$\Omega_0 = \|K_0'\|_2^2\mathbb{E}\left[\frac{\dot{m}^2(X_i,\beta_0)g_{\epsilon|X}(0 \mid X_i)}{\bar{G}(m(X_i,\beta_0))}\right], \tag{12}$$

$$\Omega_1 = \mathbb{E}\big[(\dot{m}(X_i,\beta_0))^2 \times g^{(2)}_{\epsilon|X}(0 \mid X_i)\big]. \tag{13}$$

Remark 3.4. In complete data, Parzen [23] and Eddy [6] and Khardani et al. [12] in censored data have proven similar asymptotic results for kernel estimators of the mode of the distribution of a variable response Y without conditioning on X. Therefore the results of [23, 6] can be considered as special cases of Theorem 3.3 when there is no predictor involved. By Theorem 3.3, the asymptotic variance is $\Omega^2/(nh_n^3)$.

Theorem 3.3 generalizes the result of [31] in the linear case and when $\bar{G} = 1$. We give the same rate of convergence of $\hat{\beta}_n$, which is $\sqrt{nh_n^3}$, except that the effect of censoring is present in our case in the variance by $1/\bar{G}$: which increases this quantity. A theoretic optimal bandwidth h_n for estimating can be obtained by minimizing the asymptotic weighted mean squared errors.

Theorem 3.5. *Assume that* (A1)–(A8) *hold. We have*

$$\Omega_n^2 \xrightarrow{\mathbf{P}} \Omega^2,$$

where

$$\Omega_n^2 = \hat{\Omega}_{1,n}^{-1}\hat{\Omega}_{0,n}\hat{\Omega}_{1,n}^{-1},$$

$$\hat{\Omega}_{1,n} = \frac{1}{nh_n^3}\sum_{i=1}^{n}\frac{\delta_i}{\bar{G}(Z_i)}\left[K_0''\left(\frac{Z_i - m(X_i,\hat{\beta}_n)}{h_n}\right)\dot{m}^2(X_i,\hat{\beta}_n) - h_nK_0'\left(\frac{Z_i - m(X_i,\hat{\beta}_n)}{h_n}\right)\ddot{m}(X_i,\hat{\beta}_n)\right],$$

$$\hat{\Omega}_{0,n} = \frac{1}{nh_n}\sum_{i=1}^{n}\left[\left(\frac{\delta_i}{\bar{G}(Z_i)}K_0'\left(\frac{Z_i - m(X_i,\hat{\beta}_n)}{h_n}\right)\dot{m}(X_i,\hat{\beta}_n)\right)^2\right].$$

Corollary 3.6. *Based on* $\hat{\Omega}_{0,n}$ *and* $\hat{\Omega}_{1,n}$ *we easily get a plug-in estimator* $\hat{\Omega}_n^2$ *for* Ω^2 *which, under the assumptions of Theorem* 3.5*, gives a confidence interval of asymptotic level* $1-\alpha$ *for* β_0

$$\beta_0^i \in \left[\hat{\beta}_n^i - \frac{\hat{\Omega}_n^i}{\sqrt{nh_n^3}} \times \eta_{1-\alpha/2}, \quad \hat{\beta}_n^i + \frac{\hat{\Omega}_n^i}{\sqrt{nh_n^3}} \times \eta_{1-\alpha/2}\right],$$

where $\hat{\Omega}_n^i = \mathrm{Var}(\hat{\beta}_n^i)$ *for* $i = 1,\dots,p$ *and* $\eta_{1-\alpha/2}$ *denotes the* $(1-\alpha/2)$*-quantile of the standard normal distribution.*

4. SIMULATION STUDY

In this section, we discuss the feasibility and the performance of our estimates $\hat{\beta}_n = (\hat{\beta}_{0n}, \hat{\beta}_{1n})^T$. We are interested in the behavior of the bias, variance and MSE of the two estimators. In the first setting, we generate independently and identically distributed (i.i.d.) sample $(X_i, Y_i)_{1\leq i\leq n}$ from the following model

$$Y = \frac{4}{3}\exp(\beta_0 X + \beta_1 X^2) + \sigma\epsilon, \tag{14}$$

where $\beta_0 = \frac{4}{5}$ and $\beta_1 = 1$. The sample $(X_i, Y_i)_{1\leq i\leq n}$ was generated as follows: X_i with a uniform distribution on $[-2, 2]$ and the error term ϵ_i a standard normal random variable. The censoring variable C satisfies $C = \theta_0 \exp(\theta_1 X + \theta_2 X^2) + \sigma\eta$ for certain choices of θ_0, θ_1, θ_2 and σ, where η has a standard normal distribution. We assume that ϵ and η are independent of X and that ϵ is independent of η.

Our simulation scheme is as follows: N independent samples of size n were generated. We worked with the standard normal density $K_0(x) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{1}{2}x^2\}$. Estimation of $\beta = (\beta_0, \beta_1))^T$ defined by (9) can be seen as solving the set of moment conditions

$$\mathbb{E}\left[\exp\left(\frac{(Y_i - m(X_i,\beta))^2}{2h_n^2}\big(Y_i - m(X_i,\beta)\dot{m}(X_i,\beta)\big)\right)\right] = 0. \tag{15}$$

A Newton-type algorithm is used. In nonparametric estimation, it is well known that optimality (in the MSE sense) is not seriously affected by the choice of kernel K_0 but can be swayed by that of the bandwidth h_n. In the censored model, the estimator depends on the choice of many parameters: the bandwidth h_n, the sample size n, the percentage of censoring CP (controlled by σ). Now, for this empirical study we use the Gaussian kernel and we consider the well-known smoothing parameter defined by $h_n = \sigma_n^2 n^{-0.143}$ (to satisfy (A8)), where

$$\sigma_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{and} \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

We performed simulation studies based on data that were obtained in the following manner. We consider the values for the response and explanatory random variables X and Y given by the models (14) and (16) below, but we keep in mind that the 'observable' variables are (X, Z, δ) defined in Section 1. Recall that the proportion of censoring (in % and denoted in the tables) is computed as the average of $\mathbb{P}(\delta = 0 \mid x)$ for an equispaced grid of values of x. To evaluate the finite sample performance of our estimator at each scenario, a sample of size $n = 100$ and $N = 500$ replications were used. The distance measure that was approximated is the mean squared error total (MSE). Tables 1, 2 summarize the results of this simulation study: we show the effect of a variation of the constant σ (obviously its effect on the term of variance), the size of the sample and the effect of the choice of the bandwidth. These tables include variation in the three parameters (bias, variance, MSE) depending on $(n, h, \theta_0, \theta_1, \sigma, CP)$.

In the second setting, we generate i.i.d. data from the regression model

$$Y = \frac{\beta_0}{11}\cos(\beta_1 X^2) + \sigma\epsilon, \tag{16}$$

where $\beta_0 = \beta_1 = 11$, $\sigma^2 = 0.4$ or 1. The sample $(X_i, Y_i)_{1\leq i\leq n}$ was generated as follows: X_i with a uniform distribution on $[-2, 2]$ and the error term ϵ_i a standard normal random variable. The censoring variable C satisfies $C = \frac{\theta_0}{11}\cos(\theta_1 X^2) + \sigma\epsilon^*$ for certain choices of (θ_0, θ_1), where ϵ^* has a standard normal distribution. By a simple calculus, under this model

$$\mathbb{P}(\delta = 0 \mid X = x) = 1 - \Phi\left(\frac{\theta_0/11\cos(\theta_1 x^2) - \beta_0/11\cos(\beta_1 x^2)}{\sqrt{2}\sigma}\right).$$

Table 1. Average values of $\hat{\beta}_{0n}$ and $\hat{\beta}_{1n}$ of the model (14)

θ_0 / σ^2	θ_1 / C.P	θ_2	$\hat{\beta}_{0n}$ Bias	$\hat{\beta}_{0n}$ Var	$\hat{\beta}_{0n}$ MSE	$\hat{\beta}_{1n}$ Bias	$\hat{\beta}_{1n}$ Var	$\hat{\beta}_{1n}$ MSE
1.3 / 1	1.2 / 35	1	0.002	0.074	0.076	−0.01	0.1	0.1
1.3 / 1	0.9 / 46	1	0.03	0.086	0.087	−0.08	0.15	0.1564
1.3 / 0.75	0.5 / 57	1.5	0.046	0.26	0.28	−0.1	0.2	0.21
1.3 / 0.5	0.8 / 38.5	0.95	0.03	0.045	0.05	0.06	0.07	0.08

Table 2. Average values of $\hat{\beta}_{0n}$ and $\hat{\beta}_{1n}$ of the model (16)

θ_0 / σ^2	θ_1 / C.P	$\hat{\beta}_{0n}$ Bias	$\hat{\beta}_{0n}$ Var	$\hat{\beta}_{0n}$ MSE	$\hat{\beta}_{1n}$ Bias	$\hat{\beta}_{1n}$ Var	$\hat{\beta}_{1n}$ MSE
11 / 0.5	12 / 53	0.8	2.4	3.04	0.4	0.18	0.34
11 / 1	11 / 46	−0.56	2.01	2.33	0.15	0.077	0.097
25 / 0.5	1.5 / 37	−0.30	1.85	2	−0.07	0.063	0.07
24 / 0.5	1.6 / 35	−0.5	2.9	3.03	0.039	0.162	0.18

5. APPENDIX: AUXILIARY RESULTS AND PROOFS

Proof of Theorem 3.1. There are two parts of the proof of this theorem. First, in Lemma 5.1 below we establish that $\bar{S}(\beta) = \lim_{n\to\infty} \mathbb{E}(\hat{S}_n(\beta))$ exists and is continuous in β with a unique global maximum at $\beta = \beta_0$. Second, in Lemma 5.2 below we establish the almost sure uniform convergence of $\hat{S}_n(\beta)$ to $\bar{S}(\beta)$.

Lemma 5.1. *Under Assumptions* (A1)–(A7) *we have that*

$$\Psi(\beta, h) = \int_{\mathbb{R}} \int_{\mathbb{R}^d} K_0(s) g_{\epsilon|X}(m(x,\beta) - m(x,\beta_0) + hs \mid x)\, ds\, dF_X(x)$$

exists and is continuous for all (β, h). *In addition,* $\lim_{n\to\infty} \mathbb{E}(\hat{S}_n(\beta))$ *is equal to* $\Psi(\beta, 0))$ *and has a unique global maximum over a compact set around* $\beta = \beta_0$.

Proof. First, observe that

$$\mathbf{1}_{\{Y_1 \le C_1\}} \varphi(Z_1) = \mathbf{1}_{\{Y_1 \le C_1\}} \varphi(Y_1), \tag{17}$$

for any measurable function φ. Using the fact that $\sup_{t<\tau}|\bar{G}_n(t)-\bar{G}(t)| = O\big(\sqrt{\frac{\log\log n}{n}}\,\big) = o(1)$ as $n\to\infty$ (see formula (4.28) in [4]) and by the conditional expectation properties, we have

$$\begin{aligned}
\mathbb{E}[\hat{S}_n(\beta)] &= \frac{1}{nh_n}\sum_{i=1}^{n}\mathbb{E}\Big[\delta_i\bar{G}^{-1}(Z_i)K_0\Big(\frac{Z_i-m(X_i,\beta)}{h_n}\Big)\Big]\\
&= \frac{1}{h_n}\mathbb{E}\Big[K_0\Big(\frac{Y_i-m(X_i,\beta)}{h_n}\Big)\mathbb{E}\big[\delta_i\bar{G}^{-1}(Z_i)\mid (X_1,Y_1)\big]\Big]\\
&= \frac{1}{h_n}\int_{\mathbb{R}}\int_{\mathbb{R}^d}K_0\Big(\frac{\epsilon+m(x,\beta_0)-m(x,\beta)}{h_n}\Big)g_{\epsilon|X}(\epsilon\mid x)\,d\epsilon\,dF_X(x)\\
&= \int_{\mathbb{R}}\int_{\mathbb{R}^d}K_0(w)g_{\epsilon|X}\big(m(x,\beta)-m(x,\beta_0)+h_nw\mid x\big)\,dw\,dF_X(x).
\end{aligned}$$

By Assumptions (A1), (A2(i)) and (A6), and by dominated convergence we obtain that $\Psi(\beta,h_n)$ exists and is continuous for every (β,h). We have $\mathbb{E}[\hat{S}_n(\beta)] = \Psi(\beta,h_n)$ and the continuity of $\Psi(\beta,h_n)$ implies

$$\lim_{n\to+\infty}\mathbb{E}[\hat{S}_n(\beta)] = \mathbb{E}\big[g_{\epsilon|X}((m(X_i,\beta)-m(X_i,\beta_0))\mid X_i)\big] = \Psi(\beta,0) = \bar{S}(\beta).$$

Secondly, by (A2(ii)) $g_{\epsilon|X}(\epsilon\mid x)$ achieves a strict global maximum at $\epsilon=0$ for every x and by (A7(iii)) it follows that $\bar{S}(\beta) = \lim_{n\to+\infty}\mathbb{E}[\hat{S}_n(\beta)]$ achieves a strict global maximum at $\beta=\beta_0$. □

Lemma 5.2. *Under Assumptions* (A1)–(A8(i)) *we have*

$$\sup_{\beta\in\mathbb{K}}|\hat{S}_n(\beta)-\bar{S}(\beta)| = O\bigg(\sqrt{\frac{\log n}{nh_n}}\,\bigg)\ a.s.$$

Proof. $\mathbb{K}$ is a compact set, hence it admits a covering by a finite number l_n of balls $\mathcal{B}_k(\beta_k^*,r_n)$ centered at $\beta_k^*, 1\le k\le l_n$,

$$\mathbb{K}\subset\bigcup_{k=1}^{l_n}B(\beta_k^*,r_n),$$

where

$$r_n = n^{-1/2}h_n^{3/2}. \tag{18}$$

Since $\mathbb{K}$ is bounded, there exists a constant $\kappa>0$ such that $l_n\le\kappa r_n^{-p}$. For any $\beta\in\mathbb{K}$, there exists k such that

$$\|\beta-\beta_k^*\|\le r_n. \tag{19}$$

We write

$$\begin{aligned}
\sup_{\beta\in\mathbb{K}}|\hat{S}_n(\beta)-\bar{S}(\beta)| &\le \sup_{\beta\in\mathbb{K}}|\hat{S}_n(\beta)-\hat{S}_n({\beta^*}_k)| + \sup_{\beta\in\mathbb{K}}|\hat{S}_n(\beta_k^*)-\mathbb{E}\hat{S}_n(\beta_k^*)| + \sup_{\beta\in\mathbb{K}}|\mathbb{E}\hat{S}_n(\beta_k^*)-\bar{S}(\beta)|\\
&:= \Sigma_{1n}+\Sigma_{2n}+\Sigma_{3n}.
\end{aligned}$$

We have

$$\Sigma_{1n} = O\Big(\frac{1}{\sqrt{nh_n}}\Big)\ a.s.;\qquad \Sigma_{3n} = O\Big(\frac{1}{\sqrt{nh_n}}\Big)\ a.s. \tag{20}$$

and

$$\Sigma_{2n} = O\bigg(\sqrt{\frac{\log n}{nh_n}}\,\bigg)\ a.s. \tag{21}$$

Proof of (20). From (18), we have

$$\begin{aligned}\Sigma_{1n} &= \sup_{\beta\in\mathbb{K}} |\hat{S}_n(\beta) - \hat{S}_n(\beta_k^*)| \\ &= \sup_{\beta\in\mathbb{K}} \left| \frac{1}{nh_n} \sum_{i=1}^{n} \delta_i \bar{G}_n^{-1}(Z_i) K_0\left(\frac{Z_i - m(X_i,\beta)}{h_n}\right) - \frac{1}{nh_n} \sum_{i=1}^{n} \delta_i \bar{G}_n^{-1}(Z_i) K_0\left(\frac{Z_i - m(X_i,\beta_k^*)}{h_n}\right)\right| \\ &\le \frac{C\mathbb{E}(f(X_i))}{\bar{G}(\tau)h_n^2} \sup_{\beta\in\mathbb{K}} |\beta - \beta_k^*| = O\left(\frac{1}{\sqrt{nh_n}}\right).\end{aligned}$$

For Σ_{3n} we clearly have $\Sigma_{3n} \le \mathbb{E}[\Sigma_{1n}] \le C\left(\frac{1}{\sqrt{nh_n}}\right)$. □

Proof of (21). In order to study Σ_{2n}, we use an exponential inequality which was first derived in [28]. For that we consider the family of functions defined on $\mathbb{R}\times\mathbb{R}\times\mathbb{R}^p$ by

$$\psi_\beta(u,v,x) = \frac{1}{nh}\mathbf{1}_{\{u<v\}} \bar{G}^{-1}(u) K_0\left(\frac{u - m(x,\beta)}{h}\right), \quad x\in\mathbb{R}^p, \quad u\in\mathbb{R}, \quad v\in\mathbb{R}.$$

Under (A3 (iv)) and (A4) and using Lemma (2.6.20) in [30, p. 148], the set

$$\mathcal{F}_1 := \left\{\psi_\beta(u,v,x) = \frac{1}{nh}\mathbf{1}_{\{u<v\}} \bar{G}^{-1}(u) K_0\left(\frac{u - m\,(x,\beta)}{h_n}\right), x\in\mathbb{R}^p, y\in\mathbb{R}\right\}$$

is a VC-class of measurable functions.

Now we write

$$\Sigma_{2n} = \sum_{i=1}^{n} \Big\{[\psi_\beta(T_i,C_i,X_i)] - \mathbb{E}[\psi_\beta(T_i,C_i,X_i)]\Big\}. \tag{22}$$

To deal with (22), we first note that the envelope of $\mathcal{F}_1$ is $U_n := \frac{\|K_0\|_\infty}{\bar{G}(\tau)}\frac{1}{nh_n}$. Moreover, proceeding as in (5) and using (A3 (iv)), we get

$$\sup_{x,t\leqslant\tau} \mathbb{E}\big[\psi_\beta^2(T_i,C_i,X_i)\big] \le \frac{\|K_0\|_\infty^2 \|g_{\epsilon|X}\|_\infty \times M}{n^2h_n} := \frac{M2}{n^2h_n} = \sigma_n^2,$$

with $\sigma_n \le U_n$ for n large enough.

Now applying Talagrand's inequality (see Proposition **A** in [8]), there exist two positive constants m_1 and m_2 such that

$$\begin{aligned}&\mathbb{P}\Big\{\sup_{\psi_\beta(u,v,x)\in\mathcal{F}_1} \Big|\sum_{i=1}^{n}\{\,\psi_\beta(T_i,C_i,X_i) - \mathbb{E}[\psi_\beta(T_i,C_i,X_i)]\}\Big| > t\Big\} \\ &\qquad\le m_1 \exp\left\{-\frac{t}{m_1U}\log\left[1 + \frac{tU}{m_1\big(\sigma_n\sqrt{n} + U\sqrt{\log(m_2U/\sigma_n)}\big)^2}\right]\right\}.\end{aligned}$$

Then, under (A8), simple algebraic calculations show that $\sigma_n\sqrt{n} \geqslant U\sqrt{\log(m_2U/\sigma_n)}$ for n large enough, which gives

$$\mathbb{P}\big[\sup_{\psi_\beta(x,y)\in\mathcal{F}_1} |\Sigma_{2,n}| > t\big] \le m_1 \exp\left\{-\frac{t}{m_1U}\log\left[1 + \frac{tU}{4m_1n\sigma_n^2}\right]\right\}. \tag{23}$$

Taking $t = B_3\sqrt{\frac{\log n}{nh_n}}$, where B_3 is a positive constant, a Taylor expansion using $\log(1+w)\sim w$ (for $w\to 0$) shows that the right-hand side in (23) is of order

$$m_1 \exp\left\{-\frac{B_3^2}{4m_1^2M_2}\log n\right\}$$

which, for $B_3 > 2m_1\sqrt{M2}$, is the general term of a convergent Riemann series which in turn, by Borel–Cantelli's lemma proves that

$$\Sigma_{2n} = O\biggl(\sqrt{\frac{\log n}{nh_n}}\biggr) \ a.s.$$

Finally, Lemma 5.1 and Lemma 5.2 end the proof of Theorem 3.1. □

Proof of Corollary 3.2. Standard argument gives us

$$\begin{aligned}
|\bar{S}(\hat{\beta}_n) - \bar{S}(\beta_0)| &\le |\bar{S}(\hat{\beta}_n) - \hat{S}_n(\hat{\beta}_n)| + |\hat{S}_n(\hat{\beta}_n) - \bar{S}(\beta_0)| \\
&\le \sup_{\beta\in\mathbb{K}} |\hat{S}_n(\beta) - \bar{S}(\beta)| + \bigl|\sup_{\beta\in\mathbb{K}} \hat{S}_n(\beta) - \sup_{\beta\in\mathbb{K}} \bar{S}(\beta)\bigr| \\
&\le 2\sup_{\beta\in\mathbb{K}} |\hat{S}_n(\beta) - \bar{S}(\beta)|. \qquad (24)
\end{aligned}$$

The a.s. consistency of $\hat{\beta}_n$ follows then immediately from Theorem 3.1. Now a Taylor expansion gives

$$\bar{S}(\hat{\beta}_n) - \bar{S}(\beta_0) = \frac{1}{2}(\hat{\beta}_n - \beta_0)^2 \bar{S}^{(2)}(\bar{\beta}_n^\star), \qquad (25)$$

where $\bar{\beta}_n^*$ is between β_0 and $\hat{\beta}_n$. Then by (24) and (25) we have

$$|\hat{\beta}_n - \beta_0| \le 2\sqrt{\frac{\sup_{\beta\in\mathbb{K}} |\hat{S}_n(\beta) - \bar{S}(\beta)|}{\bigl|\bar{S}^{(2)}(\bar{\beta}_n^*)\bigr|}},$$

and Lemma 5.2 completes the proof. □

Proof of Theorem 3.3. Denote

$$\nabla_n^{\beta_0} = \frac{\partial \hat{S}_n(\beta)_{|\beta_0}}{\partial\beta} = -\frac{1}{nh_n^2}\sum_{i=1}^n \delta_i \bar{G}_n^{-1}(Z_i) K_0'\biggl(\frac{\epsilon_i}{h_n}\biggr) \times \dot{m}(X_i, \beta_0). \qquad (26)$$

By the definition in (9), we have $\hat{S}_n^{(1)}(\hat{\beta}_n) = 0$. Now using a Taylor expansion of $\hat{S}_n^{(1)}(\cdot)$ in the neighborhood of β_0, we get

$$\hat{\beta}_n - \beta_0 = -\frac{\hat{S}_n^{(1)}(\beta_0)}{\hat{S}_n^{(2)}(\bar{\beta}_n)}, \qquad (27)$$

where $\bar{\beta}_n$ belongs to segment $[\hat{\beta}_n\ ,\ \beta_0]$ if the denominator does not vanish.

From (27), we have

$$\begin{aligned}
\sqrt{nh_n^3}(\hat{\beta}_n - \beta_0) &= \sqrt{nh_n^3}\frac{\hat{S}_n^{(1)}(\beta_0)}{\hat{S}_n^{(2)}(\bar{\beta}_n)} = \sqrt{nh_n^3}\frac{\nabla_n^{\beta_0} - \mathbb{E}(\nabla_n^{\beta_0})}{S_n^{(2)}(\bar{\beta}_n)} + \sqrt{nh_n^3}\frac{\mathbb{E}(\nabla_n^{\beta_0})}{S_n^{(2)}(\bar{\beta}_n)} \\
&= \frac{\mathcal{I}_{1n}}{S_n^{(2)}(\bar{\beta}_n)} + \frac{\mathcal{I}_{2n}}{S_n^{(2)}(\bar{\beta}_n)}. \qquad (28)
\end{aligned}$$

We complete the proof by showing that $\mathcal{I}_{2n}$ is negligible (see Lemma 5.3), whereas $\mathcal{I}_{1n}$ is asymptotically normal (see Lemma 5.4) and the denominator converges in probability to $\bar{S}^{(2)}(\beta)$ (see Lemma 5.5).

Lemma 5.3. *Assume* (A1)–(A8). *Then we have*

$$\mathcal{I}_{2n} \longrightarrow 0 \ \ a.s. \quad as \quad n \longrightarrow \infty.$$

Proof. Using (17) and integrating by parts we obtain

$$-\mathcal{I}_{2n} = \sqrt{nh_n^3}\mathbb{E}(\nabla_n^{\beta_0}) = \sqrt{nh_n^3}\mathbb{E}\biggl[\frac{1}{nh_n^2}\sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Z_i)} K_0'\biggl(\frac{\epsilon_i}{h_n}\biggr)\dot{m}(X_i, \beta_0)\biggr]$$

$$
\begin{aligned}
&= \sqrt{nh_n^3}\mathbb{E}\Big[\frac{1}{h_n^2}K_0'\Big(\frac{\epsilon_i}{h_n}\Big)\dot{m}(X_i,\beta_0)\mathbb{E}[\mathbf{1}_{\{Y_1\le C_1\}}\bar{G}^{-1}(Z_1)\mid(X_1,Y_1)]\Big]\\
&= n^{\frac{1}{2}}h_n^{-\frac{1}{2}}\int\int K_0'\Big(\frac{\epsilon}{h_n}\Big)\dot{m}(x,\beta_0)g_{\epsilon|X}(\epsilon\mid x)\,d\epsilon\,dF_X(x)\\
&= \int n^{\frac{1}{2}}h_n^{\frac{1}{2}}\Big[K_0\Big(\frac{\epsilon}{h_n}\Big)g_{\epsilon|X}(\epsilon\mid x)\Big]_{-\infty}^{+\infty}\dot{m}(x,\beta_0)\,dF_X(x)\\
&\qquad - n^{\frac{1}{2}}h_n^{\frac{1}{2}}\int\int K_0\Big(\frac{\epsilon}{h_n}\Big)\dot{m}(x,\beta_0)g^{(1)}_{\epsilon|X}(\epsilon\mid x)\,d\epsilon\,dF_X(x)\\
&= \mathcal{S}_1^n+\mathcal{S}_2^n.
\end{aligned}
$$

Under Assumptions (A3 (iii)) and (A7 (i))

$$
\mathcal{S}_1^n = o(1). \tag{29}
$$

On the other hand, using a change of variable we can write

$$
\begin{aligned}
\mathcal{S}_2^n &= n^{\frac{1}{2}}h_n^{\frac{1}{2}}\int\int K_0\Big(\frac{\epsilon}{h_n}\Big)\dot{m}(x,\beta_0)g^{(1)}_{\epsilon|X}(\epsilon\mid x)\,d\epsilon\,dF_X(x)\\
&= n^{\frac{1}{2}}h_n^{\frac{3}{2}}\int\int K_0(w)\dot{m}(x,\beta_0)g^{(1)}_{\epsilon|X}(h_nw\mid x)\,dw\,dF_X(x).
\end{aligned}
$$

Using (A2 (i)), we obtain by a Taylor expansion

$$
g^{(1)}_{\epsilon|X}(h_nw\mid x) = g^{(1)}_{\epsilon|X}(0\mid x)+h_nwg^{(2)}_{\epsilon|X}(0\mid x)+\frac{h_n^2w^2}{2}g^{(3)}_{\epsilon|X}(t^*\mid x),\quad t^*\in[0,h_nw].
$$

Integrating by parts, we have by (A3 (ii))

$$
\begin{aligned}
\mathcal{S}_2^n &= n^{1/2}h_n^{3/2}\int\int K_0(w)\Big[g^{(1)}_{\epsilon|X}(0\mid x)+h_nwg^{(2)}_{\epsilon|X}(0\mid x)+\frac{h_n^2w^2}{2}g^{(3)}_{\epsilon|X}(t^*\mid x)\Big]\dot{m}(x,\beta_0)\,d\epsilon\,dF_X(x)\\
&= n^{1/2}h^{5/2}\int\int wK_0(w)g^{(2)}_{\epsilon|X}(0\mid x)\dot{m}(x,\beta_0)\,dw\,dF_X(x)\\
&\qquad + n^{1/2}h^{7/2}\int\int w^2K_0(w)g^{(3)}_{\epsilon|X}(t^*\mid x)\dot{m}(x,\beta_0)\,dw\,dF_X(x)\\
&= A_{1n}+A_{2n}.
\end{aligned}
$$

Moreover, under Assumptions (A2 (i)), (A3 (i)) and (A7 (i))

$$
A_{1n} = n^{1/2}h_n^{5/2}\int wK_0(w)\,dw\int g^{(2)}_{\epsilon|X}(0\mid x)\dot{m}(x,\beta_0)\,dF_X(x) = o(1).
$$

In addition,

$$
\begin{aligned}
A_{2n} &= n^{1/2}h_n^{7/2}\int\int w^2K_0(w)g^{(3)}_{\epsilon|X}(t^*\mid x)\dot{m}(x,\beta_0)\,dw\,dF_X(x)\\
&= n^{1/2}h_n^{7/2}\underbrace{\int w^2K_0(w)\,dw}_{<\infty}\underbrace{\int g^{(3)}_{\epsilon|X}(t^*|x)\dot{m}(x,\beta_0)\,dF_X(x)}_{<\infty} = o(1).
\end{aligned}
$$

Under (A2 (i)), (A3 (ii)), (A7 (i)), and (A8 (ii)), we conclude

$$
\mathcal{S}_2^n = o(1). \tag{30}
$$

Finally (29) and (30) finish the proof of Lemma 5.3. □

Lemma 5.4. *Under* (A1)–(A8) *we have*

$$
\mathrm{Var}\left[\mathcal{I}_{1n}\right] \longrightarrow \Omega_0 \;\; a.s. \quad as \quad n\longrightarrow +\infty,
$$

where

$$\Omega_0 = \|K_0'\|_2^2 \mathbb{E}\Big[\frac{\dot m^2(X_i,\beta_0) g_{\epsilon|X}(0 \mid X_i)}{\bar G(m(X_i,\beta_0))}\Big].$$

Proof. We have

$$\begin{aligned}
\mathrm{Var}(\mathcal{I}_{1n}) &= \mathrm{Var}\left[\sqrt{nh_n^3}(\nabla_n^{\beta_0} - \mathbb{E}[\nabla_n^{\beta_0}])\ \right] = nh_n^3\,\mathrm{Var}[\nabla_n^{\beta_0}]\\
&= nh_n^3\,\mathrm{Var}\left[\frac{1}{nh_n^2}\sum_{i=1}^n \frac{\delta_i}{\bar G_n(Z_i)} K_0'\Big(\frac{\epsilon_i}{h_n}\Big)\dot m(X_i,\beta_0)\right]\\
&= \frac{1}{h_n}\,\mathrm{Var}\left[\frac{\delta_1}{\bar G_n(Y_1)} K_0'\Big(\frac{\epsilon_1}{h_n}\Big)\dot m(X_1,\beta_0)\right]\\
&= \frac{1}{h_n}\mathbb{E}\Big[\bar G^{-2}(Y_1)\Big(K_0'\Big(\frac{\epsilon_1}{h_n}\Big)\dot m(X_i,\beta_0)\Big)^2 \mathbb{E}[\mathbf{1}_{\{Y_1\le C_1\}}|X_1,Y_1]\Big]\\
&\quad - \frac{1}{h_n}\mathbb{E}\Big[\bar G^{-1}(Y_1)\Big(K_0'\Big(\frac{\epsilon_1}{h_n}\Big)\dot m(X_i,\beta_0)\Big)\mathbb{E}[\mathbf{1}_{\{Y_1\le C_1\}}|X_1,Y_1]\Big]^2\\
&=: \Psi_{1n} + \Psi_{2n}.
\end{aligned}$$

On the one hand, by Lemma 5.3

$$\Psi_{2n} = n^{-1}\mathcal{I}_{2n}^2 \longrightarrow 0 \qquad \text{as} \quad n\longrightarrow\infty.$$

We have

$$\Psi_{1n} = \frac{1}{h_n}\mathbb{E}\Big[\bar G^{-2}(Y_1)\Big(K_0'\Big(\frac{\epsilon_1}{h_n}\Big)\dot m(X_1,\beta_0)\Big)^2\mathbb{E}[\mathbf{1}_{\{Y_1\le C_1\}} \mid X_1,Y_1]\Big].$$

Then, since $G_n(\cdot)$ is continuous and consistent, we have

$$\begin{aligned}
\lim_{n\to\infty}\Psi_{1n} &= \int K_0'^2(w)\,dw \int \frac{\dot m^2(x,\beta_0) g_{\epsilon|X}(0\mid x)\,dF_X(x)}{\bar G(m(x,\beta_0))}\\
&= \|K_0'\|_2^2\,\mathbb{E}\Big[\frac{\dot m^2(X_i,\beta_0) g_{\epsilon|X}(0\mid X_i)}{\bar G(m(X_i,\beta_0))}\Big],
\end{aligned}$$

which gives the result. $\square$

Now, to complete the proof of Theorem 3.3, it suffices to prove that

$$\hat S_n^{(2)}(\bar\beta_n) \xrightarrow{\mathbb{P}} \bar S^{(2)}(\beta_0).$$

Lemma 5.5. *Under Assumptions* (A1)–(A8) *we have*

$$\sup_{\beta\in\mathbb{K}}\big|\hat S_n^{(2)}(\beta) - \bar S^{(2)}(\beta)\big| \xrightarrow{\mathbb{P}} 0 \qquad as \quad n\longrightarrow\infty.$$

Proof. Using the triangle inequality we have

$$\begin{aligned}
\sup_{\beta\in\mathbb{K}}\big|\hat S_n^{(2)}(\beta) - \bar S^{(2)}(\beta)\big| &\le \sup_{\beta\in\mathbb{K}}\big|\hat S_n^{(2)}(\beta) - \hat S_n^{(2)}(\beta_k)\big| + \sup_{\beta\in\mathbb{K}}\big|\hat S_n^{(2)}(\beta_k) - \mathbb{E}[\hat S_n^{(2)}(\beta_k)]\big|\\
&\quad + \sup_{\beta\in\mathbb{K}}\big|\mathbb{E}[\hat S_n^{(2)}(\beta_k)] - \mathbb{E}[\hat S_n^{(2)}(\beta)]\big| + \sup_{\beta\in\mathbb{K}}\big|\mathbb{E}\big[\hat S_n^{(2)}(\beta)\big] - \bar S^{(2)}(\beta)\big|\\
&= \gamma_{1n}(\beta) + \gamma_{2n}(\beta) + \gamma_{3n}(\beta) + \gamma_{4n}(\beta).
\end{aligned}$$

We have by (8)

$$\hat S_n^{(2)}(\beta) = \frac{1}{nh_n^3}\sum_{i=1}^n \delta_i \bar G_n^{-1}(Z_i)\big(\dot m(X_i,\beta)\big)^2 K_0''\Big(\frac{Y_i - m(X_i,\beta)}{h_n}\Big)$$

$$-\frac{1}{nh_n^2}\sum_{i=1}^{n}\delta_i\bar{G}_n^{-1}(Z_i)\ddot{m}(X_i,\beta)K_0'\left(\frac{Y_i-m(X_i,\beta)}{h_n}\right)$$
$$=V_{1,n}(\beta)+V_{2,n}(\beta). \tag{31}$$

The result is then a consequence of the following Lemmas 5.6 and 5.7 below. □

Lemma 5.6. *Assume* (A3), (A5)–(A6) *and* (A7 (i)–(ii)). *Then*

$$\gamma_{1n}(\beta)\longrightarrow 0\ \ a.s.,\qquad \gamma_{3n}(\beta)\longrightarrow 0\ ;a.s.$$

Proof. Using the same idea as in the proof of Lemma 5.2 by taking

$$r_n'=h_n^{\xi},\quad \xi>4. \tag{32}$$

□

Lemma 5.7. *Assume* (A1)–(A8). *Then we have*

$$\gamma_{2n}(\beta)\longrightarrow 0\ \ a.s.,\quad \gamma_{4n}(\beta)\longrightarrow 0\ \ a.s.$$

Proof. Under equation (31), we can write

$$\gamma_{2n}(\beta)=\gamma_{2n}^1(\beta)+\gamma_{2n}^2(\beta),$$

where $\gamma_{2n}^j(\beta)=V_{j,n}(\beta)-\mathbb{E}(V_{j,n}(\beta))$ for $j=1,2$.

We will now prove the convergence to zero of the first quantity. The second is proved in the same way.

For $\gamma_{1n}^j(\beta)$, we set

$$\zeta_{in,1}(\beta)=\frac{1}{h_n^3}(\dot{m}(x,\beta))^2\delta_i\bar{G}_n^{-1}(Z_i)K_0''\left(\frac{Z_i-m(X_i,\beta)}{h_n}\right)\mathbf{1}_{[(\dot{m}(x,\beta))^2\le h_n^{-2}]},$$
$$\zeta_{in,2}(\beta)=\frac{1}{h_n^3}\dot{m}(x,\beta))^2\delta_i\bar{G}_n^{-1}(Z_i)K_0''\left(\frac{Z_i-m(X_i,\beta)}{h_n}\right)\mathbf{1}_{[(\dot{m}(x,\beta))^2> h_n^{-2}]},$$
$$\mathcal{K}_{in,k}(\beta)=\frac{1}{n}\sum_{i=1}^{n}\zeta_{in,k}(\beta)-\mathbb{E}(\zeta_{in,k}(\beta))\quad\text{for}\quad k=1,2.$$

We get $|\zeta_{in,1}(\beta)-\mathbb{E}(\zeta_{in,1}(\beta))|\le 2h_n^{-5}\bar{G}^{-1}(\tau)\|K_0''\|_\infty$ and by Assumptions (A2), (A3) and (A7 (i,iv))

$$\begin{aligned}
\operatorname{Var}\big(\zeta_{in,1}(\beta)-\mathbb{E}(\zeta_{in,1}(\beta))\big)&\le\mathbb{E}\big[\mathbb{E}(\zeta_{in,1}^2(\beta)\mid(X_1,Y_1))\big]\\
&=\mathbb{E}\left[\left(h_n^{-3}\big(\dot{m}(x,\beta)\big)^2K_0''\left(\frac{Y_i-m(x,\beta)}{h_n}\right)\right)^2\mathbb{E}[\delta_1\bar{G}^{-2}(Z_1)\mid(X_1,Y_1)]\right]\\
&\le\frac{h_n^{-6}}{\bar{G}(\tau)}\int\left[K_0''\left(\frac{\epsilon+m(x,\beta_0)-m(x,\beta)}{h_n}\right)\right]^2\big(\dot{m}(x,\beta)\big)^4g_{\epsilon|X}(\epsilon x)\,d\epsilon\,dF_X(x)\\
&\le Mh_n^{-5}\int[K_0''(w)]^2dw\,\mathbb{E}\big[(\dot{m}(X_i,\beta))^4\big]=c_0h_n^{-5},
\end{aligned}$$

where c_0 is a finite positive constant as a consequence of Assumptions (A2 (i)), (A3), and (A7). Then, from Bernstein's inequality, we have

$$\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_{in,1}(\beta)-\mathbb{E}(\zeta_{in,1}(\beta))\right|>\epsilon\right\}\le 2\exp\left\{-\frac{3nh_n^5\epsilon^2}{6c_0+4\epsilon\|K_0''\|_\infty}\right\}.$$

Consequently,

$$\mathbb{P}\left\{\sup_{\beta\in\mathbb{K}}\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_{in,1}(\beta_k)-\mathbb{E}(\zeta_{in,1}(\beta_k))\right|>\epsilon\right\}\le\sum_{k=1}^{\lambda_n}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_{in,1}(\beta_k)-\mathbb{E}(\zeta_{in,1}(\beta_k))\right|>\epsilon\right\}$$

$$\leq 2(r_n')^{-1}\exp\Big\{-\frac{3nh_n^5\epsilon^2}{6c_0+4\|K_0''\|_\infty\epsilon}\Big\}.$$

Since $(r_n')^{-1}\sim h_n^{-\xi}$, it follows from Assumption (A8(i)) that $\mathbb{P}\big\{\sup_{\beta\in\mathbb{K}}|\mathcal{K}_{in,1}(\beta)|>\epsilon\big\}$ tends to zero as $n\to\infty$ for any fixed value of ϵ, and thus

$$\sup_{\beta\in\mathbb{K}}|\mathcal{K}_{in,1}(\beta)|=o_{\mathbb{P}}(1). \tag{33}$$

Secondly, by Assumption **A3(iv)**,

$$\sup_{\beta\in\mathbb{K}}|\mathcal{K}_{in,2}(\beta)|\leq\sup_{\beta\in\mathbb{K}}|\mathbb{E}(\zeta_{in,2}(\beta))|+\frac{1}{h_n^3}\|K_0''\|_\infty\frac{1}{n}\sum_{i=1}^{n}\big((\dot m(X_i,\beta))^2\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}\big).$$

Now, it is clear that

$$\mathbb{E}\big\{\sup_{\beta\in\mathbb{K}}|\mathcal{K}_{in,2}(\beta)|\big\}\leq 2\frac{1}{h_n^3}\|K_0''\|_\infty\mathbb{E}\big\{\sup_{\beta\in\mathbb{K}}(\dot m(X_i,\beta))^2\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}\big\}.$$

Take $p>1$ such that $\mathbb{E}|\sup_{\beta\in\mathbb{K}}|\dot m(X_i,\beta)|^{2p}|<\infty$, then by Hölder's inequality we have

$$\mathbb{E}\big\{\sup_{\beta\in\mathbb{K}}(\dot m(X_i,\beta))^2\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}\big\}\leq\big[\mathbb{E}|\sup_{\beta\in\mathbb{K}}\dot m(X_i,\beta)|^{2p}\big]^{1/p}\big[\mathbb{E}|\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}|^q\big]^{1/q},$$

where $1/p+1/q=1$. Since $s<0$, we have by Markov's inequality

$$\begin{aligned}\big[\mathbb{E}|\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}|^q\big]&=\mathbb{E}\big[\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}\big]=\mathbb{P}\big[[(\dot m(X_i,\beta))^2>h_n^{-2}]\big]\\&=\mathbb{P}\big[[(\dot m(X_i,\beta))^{2p}>h_n^{-2p}]\big]\leq h_n^{2p}\mathbb{E}\big[\big(\sup_{\beta\in\mathbb{K}}\dot m(X_i,\beta)\big)^{2p}\big].\end{aligned}$$

Hence

$$\begin{aligned}\mathbb{E}\big\{\sup_{\beta\in\mathbb{K}}(\dot m(X_i,\beta))^2\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}\big\}&\leq\big[\mathbb{E}|\sup_{\beta\in\mathbb{K}}\dot m(X_i,\beta)|^{2p}\big]^{1/p}\left[\frac{\mathbb{E}\big[[(\dot m(X_i,\beta))^{2p}]\big]}{h_n^{-2p}}\right]^{1/q}\\&=\mathbb{E}\big[[(\dot m(X_i,\beta))^{2p}]\big]h_n^{2p/q}.\end{aligned}$$

Then, since $2p/q=2(p-1)$, we have

$$\mathbb{E}\big\{\sup_{\beta\in\mathbb{K}}(\dot m(X_i,\beta))^2\mathbf{1}_{[(\dot m(X_i,\beta))^2>h_n^{-2}]}\big\}\leq\mathbb{E}\big[[(\dot m(X_i,\beta))^{2p}]\big]h_n^{2(p-1)},$$

and we conclude that

$$\mathbb{E}\big\{\sup_{\beta\in\mathbb{K}}|\mathcal{K}_{in,2}(\beta)|\big\}=O(h_n^{2p-5}).$$

Taking $p=(5+\xi)/2$, we get $\mathbb{E}\big\{\sup_{\beta\in\mathbb{K}}|\mathcal{K}_{in,2}(\beta)|\big\}=O(h_n^{\xi})=o_{\mathbb{P}}(1)$. Hence

$$\sup_{\beta\in\mathbb{K}}|\mathcal{K}_{in,2}(\beta)|=O(h_n^{\xi})=o_{\mathbb{P}}(1). \tag{34}$$

We conclude by (33) and (34) that $\gamma_{2n}^1(\beta)=o_{\mathbb{P}}(1)$ and in the same way we prove that $\gamma_{2n}^2(\beta)=o_{\mathbb{P}}(1)$.

Using integration by parts and a change of variable, under (A1)–(A3) and (A6)–(A7), we have $\gamma_{4n}(\beta)=o_{\mathbb{P}}(1)$. Then

$$\begin{aligned}\lim_n\mathbb{E}\hat S_n^{(2)}(\beta)&=\lim_n\mathbb{E}(V_{1,n}(\beta))+\lim_n\mathbb{E}(V_{2,n}(\beta))\\&=\int\int K_0(w)(\dot m(x,\beta))^2g_{\epsilon|X}^{(2)}\big(m(x,\beta)-m(x,\beta_0)\mid x\big)\,dw\,dF_X(x)\\&\quad-\int\int K_0(w)\ddot m(x,\beta)g_{\epsilon|X}^{(1)}\big(m(x,\beta)-m(x,\beta_0)\mid x\big)\,dw\,dF_X(x)\end{aligned}$$

$$= \frac{\partial^2}{\partial\beta\partial\beta'}\bar{S}(\beta) = \bar{S}^{(2)}(\beta).$$

It follows that

$$\sup_{\beta\in\mathbb{K}}\left|\frac{\partial^2\hat{S}_n(\beta)}{\partial\beta\partial\beta'} - \frac{\partial^2\bar{S}(\beta)}{\partial\beta\partial\beta'}\right| = o_{\mathbb{P}}(1).$$

Now the final step in the proof of Theorem 3.3 is to show the Lindberg condition for $\mathcal{I}_1$. For that, in view of (28), put

$$\mathcal{I}_1 =: \sum_{i=1}^n \Delta_{i,n}(x,\epsilon),$$

where

$$\Delta_{i,n}(x,\epsilon) = (nh_n)^{-1/2}\Big\{\delta_i\bar{G}_n^{-1}(Z_i)K_0\Big(\frac{\epsilon_i}{h_n}\Big)\dot{m}(X_i,\beta_0) - \mathbb{E}\Big[\delta_i\bar{G}_n^{-1}(Z_i)K_0\Big(\frac{\epsilon_i}{h_n}\Big)\dot{m}(X_i,\beta_0)\Big]\Big\}.$$

Then we have from Lemma (5.4)

$$\operatorname{Var}\Big(\sum_{i=1}^n \Delta_{i,n}(x,\epsilon)\Big) = nh_n^3\operatorname{Var}(\nabla_n^{\beta_0}) \longrightarrow \|K_0'\|_2^2\mathbb{E}\Big[\frac{(\dot{m}(X_i,\beta_0))^2\times g_{\epsilon|X}(0\mid X_i)}{\bar{G}(m(X_i,\beta_0))}\Big] := \omega^2. \tag{35}$$

Lemma 5.8. *Under Assumptions* (A1)–(A8) *we have*

$$\forall\eta>0,\ \sum_{i=1}^n\int_{\{\Delta_{i,n}^2(x,\epsilon)>\eta^2\operatorname{Var}(\Delta_{i,n}(x,\epsilon))\}}\Delta_{i,n}^2(x)\,d\mathbb{P}(x,\epsilon)\longrightarrow 0\quad as\quad n\to\infty.$$

Proof. On the one hand, we have

$$\Delta_{i,n}^2(x,\epsilon) \le \frac{2}{nh_n\bar{G}^2(\tau)}K_0^2\Big(\frac{\epsilon_i}{h_n}\Big)(\dot{m}(X_i,\beta_0))^2 + \frac{2}{nh_n\bar{G}^2(\tau)}\mathbb{E}^2\Big[K_0\Big(\frac{\epsilon_i}{h_n}\Big)\dot{m}(X_i,\beta_0)\Big]. \tag{36}$$

Note that by Lemma 5.3 the second term in the right-hand side of (36) goes to zero as $n\to\infty$:

$$\frac{2}{nh_n}\mathbb{E}^2\Big[K_0\Big(\frac{\epsilon_i}{h_n}\Big)\dot{m}(X_i,\beta_0)\Big]\longrightarrow 0 \qquad \text{as}\quad n\longrightarrow\infty. \tag{37}$$

On the other hand, taking $\eta = \frac{\omega^2}{2}$ we have by (35) that $\exists n_0\in\mathbb{N}^*$ such that $\forall n>n_0$

$$\operatorname{Var}\Big(\sum_{i=1}^n \Delta_{i,n}(x,\epsilon)\Big) \ge \|K_0'\|_2^2\mathbb{E}\Big[\frac{(\dot{m}(X_i,\beta_0))^2\times g_{\epsilon|X}(0\mid X_i)}{\bar{G}(m(X_i,\beta_0))}\Big]. \tag{38}$$

Now, set

$$W(x,\epsilon) = \frac{1}{\bar{G}^2(\tau)}K_0^2\Big(\frac{\epsilon_i}{h_n}\Big)(\dot{m}(X_i,\beta_0))^2 + \mathbb{E}^2\Big[\frac{1}{\bar{G}^2(\tau)}K_0\Big(\frac{\epsilon_i}{h_n}\Big)\dot{m}(X_i,\beta_0)\Big].$$

We clearly have from (36) that

$$\Delta_{i,n}^2(x,\epsilon) \le \frac{2W_{i,n}(x,\epsilon)}{nh_n}.$$

Now, set $\eta' = \frac{\eta^2\omega^2}{4}$, then using (38) we have for $n\ge n_0$

$$\Big\{\Delta_{i,n}^2(x,\epsilon) > \eta^2\operatorname{Var}\Big(\sum_{i=1}^n\Delta_{i,n}(x,\epsilon)\Big)\Big\} \subset \{\Delta_{i,n}^2(x,\epsilon)>\eta^2\omega^2\} = \{\Delta_{i,n}^2(x,\epsilon)>2\eta'\}$$

$$:= \{W_{i,n}(x,\epsilon)>\eta' nh_n\} \subset \Big\{K_0^2\Big(\frac{\epsilon_i}{h_n}\Big)(\dot{m}(X_i,\beta_0))^2 > \frac{\eta' nh_n}{2}\Big\}$$

$$\cup \left\{ \mathbb{E}^2 \Big[K_0\Big(\frac{\epsilon_i}{h_n}\Big) \dot{m}(X_i, \beta_0) \Big] > \frac{\eta' n h_n}{2} \right\} =: \Sigma_{1,n} \cup \Sigma_{2,n}.$$

By (37), for n large enough, we have

$$\Sigma_{2,n} = \left\{ \mathbb{E}^2 \Big[K_0\Big(\frac{\epsilon_i}{h_n}\Big) \dot{m}(X_i, \beta_0) \Big] > \frac{\eta' n h_n}{2} \right\} = \emptyset. \tag{39}$$

In the same way, by (A2)–(A3) and (A7)–(A8) we have for n large enough that $\Sigma_{1,n}$ is empty.

Therefore $\{\Delta_{i,n}^2(x,\epsilon) > \eta^2 \operatorname{Var}\big(\sum_{i=1}^n \Delta_{i,n}(x,\epsilon)\big)\}$ is empty for n large enough, which completes the proof. □

Proof of Theorem 3.5. It is sufficient to establish convergences in probability $\hat{\Omega}_{0,n}$ to Ω_0 and $\hat{\Omega}_{1,n}$ to Ω_1. By consistency of $\hat{\beta}_n$ to β_0 it follows from Theorem 3.1 and Lemma 5.5 that $\hat{\Omega}_{1,n} \xrightarrow{\mathbf{P}} \Omega_1$. Secondly, to establish that $\hat{\Omega}_{0,n} \xrightarrow{\mathbf{P}} \Omega_0$, we use an approach similar to that used in the proofs of Lemma 5.2 and Lemma 5.5 above. □

ACKNOWLEDGMENTS

The author would like to thank the editor, the associate editor and the anonymous referees for their valuable comments and suggestions, which led to substantial improvements in the manuscript.

REFERENCES

1. M. G. Akritas, “Nearest Neighbor Estimation of a Bivariate Distribution under Random Censoring”, Ann. Statist. **22**, 1299–1327 (1994).
2. M. G. Akritas, “On the Use of Nonparametric Regression Techniques for Fitting Parametric Regression Models”, Biometrics **52**, 1342–1362 (1996).
3. J. Buckley and I. James, “Linear Regression with Censored Data”, Biometrika **66**, 429–436 (1979).
4. P. Deheuvels and J. H. J. Einmahl, “Functional Limit Laws for the Increments of Kaplan–Meier Product-Limit Processes and Applications”, Ann. Probab. **28**, 1301–1335 (2000).
5. R. M. Dudley, *Uniform Central Limit Theorems* (Cambridge Univ. Press, Cambridge, UK, 1999).
6. W. F. Eddy, “Optimum Kernel Estimators of the Mode”, Ann. Statist. **8**, 870–882 (1980).
7. T. Gasser and H. G. Müller, “Estimating Regression Functions and Their Derivatives by the Kernel Method”, Scand. J. Statist. **11**, 171–185 (1984).
8. E. Giné and A. Guillou, “Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators”, Ann. Inst. Henri Poincaré **38**, 907–921 (2002).
9. W. Härdle, H. Liang, and J. Gao, *Partially Linear Models* (Physica-Verlag, Heidelberg, 2000).
10. R. Jennrich, “Asymptotic Properties of Nonlinear Least-Squares Estimator”, Ann. Math. Statist. **40**, 633–643 (1969).
11. E. L. Kaplan and P. Meier, “Nonparametric Estimation from Incomplete Observations”, J. Amer. Statist. Assoc. **53**, 457–481 (1958).
12. S. Khardani, M. Lemdani, and E. Ould Saïd, “Some Asymptotic Properties for a Smooth Kernel Estimator of the Conditional Mode under Random Censorship”, J. Korean Statist. Soc. **39**, 455–469 (2010).
13. S. Khardani, M. Lemdani, and E. Ould Saïd, “Uniform Rate of Strong Consistency for a Smooth Kernel Estimator of the Conditional Mode for Censored Time Series”, J. Statist. Plann. Inference **141**, 3426–3436 (2011).
14. S. Khardani, M. Lemdani, and E. Ould Saïd, “On the Strong Uniform Consistency of the Mode Estimator for Censored Time Series”, Metrika **135**, 1–13 (2012).
15. S. Khardani and A. F. Yao, “Nonlinear Parametric Mode Regression”, Commun. Statist. Theory and Methods **46**, 3006–3024 (2017).
16. H. Koul, V. Susarla, and J. Van Ryzin, “Regression Analysis with Randomly Right-Censored Data”, Ann. Statist. **9**, 1276–1288 (1981).
17. M. J. Lee, “Mode Regression”, J. Econometric **42**, 337–349 (1989).
18. M. J. Lee, “Quadratic Mode Regression”, J. Econometric **57**, 1–19 (1993).
19. S. Leurgans, “Linear Models, Random Censoring and Synthetic Data”, Biometrika **74**, 301–309 (1987).
20. H. Liang and W. Härdle, *Asymptotic normality of parametric part in partially linear heteroscedastic regression models* (DP 33, SFB 373, Humboldt Univ. Berlin).
21. H. Liang, W. Härdle, and R. Carroll, “Estimation in a Semi-Parametric Partially Linear Errors in Variables Model”, Ann. Statist. **27**, 1519–1535 (1999).

22. R. G. Müller, “Least Squares Regression with Censored Data”, Biometrika **63**, 449–464 (1976).
23. E. Parzen, “On Estimation of a Probability Density Function and Mode”, Ann. Math. Statist. **33**, 1065–1076 (1962).
24. D. Ruppert, M. P. Wand, and R. J. Carroll, *Semi-Parametric Regression* (Cambridge Univ. Press, Cambridge, 2003).
25. T. A. Severini and J. G. Staniswalis, “Quasi-Likelihood Estimation in Semi-Parametric Models”, J. Amer. Statist. Assoc. **89**, 501–511 (1994).
26. J. Shoung and C. Zhang, “Least Squares Estimators of the Mode of a Unimodal Regression Function”, Ann. Statist. **29**, 648–665 (2001).
27. W. Stute, “Consistent Estimation under Random Censorship when Covariates are Present”, J. Multivar. Anal. **45**, 89–103 (1993).
28. M. Talagrand, “New Concentration Inequalities in Product Spaces”, Invent. Math. **126**, 505–563 (1996).
29. I. Van Keilegom and M. G. Akritas, “The Least Squares Method in Heteroscedastic Censored Regression Models”, in: *Asymptotics in Statistics and Probability*, Ed. by M. L. Puri (2000), pp. 379–391.
30. A. W. Van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes with Applications to Statistics* (Springer-Verlag, New York, 1996).
31. W. Yao and L. Li, “A New Regression Model: Modal Linear Regression”, Scand. J. Statist. **41**, 656–671 (2014).
32. A. Yatchew, *Semi-Parametric Regression for the Applied Econometric* (Cambridge Univ. Press, Cambridge, 2003).
33. A. Yatchew, “An Elementary Estimator of the Partial Linear Model”, Economics Lett. **57**, 135–143 (1997).
34. M. Zhou, “Asymptotic Normality of the Synthetic Data Regression Estimator for Censored Survival Data”, Ann. Statist. **20**, 1002–1021 (1992).
35. K. Ziegler, “On the Asymptotic Normality of Kernel Regression Estimators of the Mode in the Random Design Model”, J. Statist. Plann. Inference **115**, 123–144 (2003).