Pak.j.stat.oper.res. Vol.22 No.3 2026 pp 377-405 DOI: http://dx.doi.org/10.18187/pjsor.v22i3.5125

Pakistan Journal of Statistics and Operation Research

Almost sure convergence with rate of a robust regression estimator for twice-censored data

Wafaa Djelladj[1*], Hassiba Benseradj[2], Zohra Guessoum[3]

*Corresponding author

1. Department of Preparatory Classes, National Polytechnic School, ENP, Algeria, wafaa.djelladj@g.enp.edu.dz
2. Faculty of Sciences, University of Boumerdes UMBB, Algeria, h.benseradj@univ-boumerdes.dz
3. Lab. MSTD, Faculty of Mathematics, USTHB, Algeria, zguessoum@usthb.dz

Abstract

In this paper, we propose a new kernel M-estimator of the regression function using nonparametric methods for twice-censored data. The almost sure (a.s.) convergence with rate over a compact set of the estimator is established under appropriate conditions using an unbounded score function. As a by-product, we build a Nadaraya-Watson-type estimator in case of the twice-censoring and derive its strong uniform consistency with rate. Furthermore, knowing that the choice of the smoothing parameter is a difficult and important question, and based on cross-validation ideas, we construct some data-driven criterion for choosing a reasonable bandwidth. In the presence of censoring and/or outliers, a simulation study is carried out to illustrate the finite-sample behavior of the proposed estimator and to evaluate the effectiveness of the proposed cross-validation criterion through comparisons with the classical cross-validation criterion. Numerical comparisons with classical estimators are also presented. Finally, a real-data application is provided to demonstrate the practical value of the proposed estimator.

Key Words: Bandwidth selection, M-estimator, Robust regression, Strong consistency, Twice-censoring.

Mathematical Subject Classification: 62G08, 62G20.

1. Introduction

Robust regression methods have become increasingly important in asymptotic analysis and regression, particularly in the presence of outliers, aberrant observations, or heavy-tailed distributions. The M-estimators, first introduced by Huber (1964), consist of a generalization of maximum likelihood estimators (MLE) under nonstandard conditions. Assume that $(T_i)_{1\leq i\leq n}$ are random variables (rv) representing the real independent and identically distributed (iid) sequence of variables and having a common continuous distribution function (df) F^T. Huber (1964) defined the M-estimator of a location parameter as a solution of the minimization problem $\min_{\theta\in\mathbb{R}}(n^{-1}\sum_{i=1}^{n}\rho(T_i-\theta))$ that leads to resolution

$$n^{-1}\sum_{i=1}^{n}\psi(T_i-\theta)=0,$$

where $\rho(.)$ is a real continuously differentiable outlier-resistant loss function, with a strictly monotonic derivative $\psi(.):=d\rho(.)/d\theta$, chosen so that we ensure the optimality and robustness of the estimator.

Taking into account covariates $(X_i)_{1\leq i\leq n}$ valued in $\mathbb{R}^d$, the purpose of this paper is to model the relation between

the lifetime variables and the covariates as

$$T_i = m(\mathbf{X}_i) + \varepsilon_i,$$

where $m(.)$ is an unknown function to estimate and ε_i is a sequence of errors independent of $\mathbf{X}_i$. One may wish to model the relationship between T and $\mathbf{X}$ using the least-squares method. The goal is to determine $m(.)$ that minimizes the mean squared error, namely

$$m(\mathbf{x}) = \arg\min_{\theta\in\mathbb{R}}(\mathbb{E}[(T-\theta)^2|\mathbf{X}=\mathbf{x}]).$$

The solution of this minimization problem is precisely the conditional expectation of T given $\mathbf{X} = \mathbf{x}$, that is $m(\mathbf{x}) = \mathbb{E}[T|\mathbf{X} = \mathbf{x}]$. Greater robustness can be achieved by minimizing an alternative function of the errors instead of the sum of their squares. Tsybakov (1982) and Härdle (1984) adapted the M-estimator to regression curve estimation, and defined the robust regression function $m_\psi(\cdot) =: m(\cdot)$, along with the corresponding kernel M-estimator, as solutions with respect to (w.r.t) θ of the equations

$$\mathbb{E}[\psi_{\mathbf{x}}(T-\theta)|\mathbf{X}=\mathbf{x}] = 0, \tag{1}$$

and

$$(nh_n^d)^{-1}\sum_{i=1}^{n} K_d\left(\mathbf{x}-\mathbf{X}_i/h_n\right)\psi_{\mathbf{x}}(T_i-\theta) = 0, \tag{2}$$

respectively, for a kernel function $K_d(.)$ defined on $\mathbb{R}^d$ and a sequence of bandwidth h_n tending to zero along with n. Note that the objective function $\psi_{\mathbf{x}}(.)$ is indexed by $\mathbf{x}$ to include a large class of M-estimators that may depend on a scaling parameter $\sigma(\mathbf{x})$, i.e., $\psi_{\mathbf{x}}(.) := \psi(./\sigma(\mathbf{x}))$. The uniform consistency of this estimator was established by Härdle and Luckhaus (1984).

The conditions imposed on $\psi_{\mathbf{x}}$ and ρ ensure the uniqueness of $m(.)$ (see Boente and Fraiman (1990) or Koul and Stute (1998)). By properly tuning the objective function $\psi_{\mathbf{x}}(.)$, the quite general setup of solutions for equations (1) and (2) provides a wide class of regression functions and estimators. For instance, setting: $\psi_{\mathbf{x}}(\mathbf{u}) = \mathbf{u}$ leads to the classical regression function together with the Nadaraya-Watson estimator; choosing $\psi_{\mathbf{x}}(\mathbf{u}) = \mathbf{1}_{(\mathbf{u}>0)} - (1-p)$ with $p \in (0,1)$ gives the conditional p-quantile and its corresponding estimator (see Ferraty and Vieu (2006)); while taking $\psi_{\mathbf{x}}(\mathbf{u}) = \max\{-c, \min\{c, \mathbf{u}\}\}$, $c > 0$ produces the Huber-type robust regression function and its associated M-estimator. Further examples concerning the choice of $\psi_{\mathbf{x}}(.)$ are available in Hampel (1973) or Serfling (1980).

An abundant literature has been devoted to M-estimators. They were deeply studied by Huber (1964) and Huber (1967) while Huber (1981), Hampel et al. (1986) and others gave their asymptotic properties. There are also many well-developed theories dedicated to nonparametric M-estimation of the regression function when complete data are observed. We can cite Härdle (1984) who established weak and strong consistency, as well as asymptotic normality, of a robust estimator of the regression function. Later, Laïb and Ould-Saïd (2000) established the strong uniform consistency of the robust estimator under ergodic assumptions. For more recent works, we note that Attaoui et al. (2015) studied the asymptotics of the M-estimator of the regression function for the quasi-associated sequence.

On the other hand, incomplete data are frequently encountered in medical follow-up studies and in reliability, and the most important are censoring and truncation. The right-censoring phenomenon has been of great use in survival analysis: we can cite Menni and Tatachak (2016), who studied the kernel estimator of the regression function. Recently, Djelladj and Tatachak (2019) studied the strong uniform consistency with rate of the kernel conditional quantile estimator under right-censoring. Some literature dealt with robust regression under right-censored models, for example Lemdani and Ould-Saïd (2017) established asymptotic properties in the context of nonparametric robust regression estimation.

In case of left-truncated data, Wang and Liang (2012) established weak and strong consistency, as well as asymptotic normality, for an M-estimator of the regression function under α-mixing dependency, while Gheliem and Guessoum (2022) investigated asymptotic properties of the M-estimators in case of a left-truncated and associated model. In Benseradj and Guessoum (2022), the strong uniform convergence of the M-estimators is considered when data are left-truncated and right-censored under α-mixing model.

A single censoring framework is not always suitable, and more complicated censoring schemes may arise in practice, where both left- and right-censoring occur simultaneously. Two such schemes exist: twice- and double-censoring. In this paper, we focus on twice-censoring.

The doubly censored model was investigated by Turnbull (1974), who introduced the data latent model in this frame-

work using nonparametric methods. Practical examples of double-censored samples exist in the literature (see Gehan (1965) and Peto (1973)). A notable case is the study of the development of African infants by Leiderman et al. (1973), which aimed to establish baby developmental norms. To compare children in Kenya with those in the United States, and the United Kingdom, the study recorded the time taken by 65 infants to accomplish specific tasks. The time elapsed between birth and the first successful completion of a task was modeled accordingly. Ren and Gu (1997) derived the asymptotic properties of an M-estimator in a multiple linear regression framework when data are doubly censored.

The twice-censoring mechanism introduced by Patilea and Rolin (2006) (called Model I) considers the variable of interest T to be right-censored by a nonnegative rv R with continuous df F^R. Furthermore, $\min(T, R)$ is left-censored by a nonnegative rv L with continuous df F^L. To ensure identifiability conditions, it is assumed that T, R and L are independent and that

$$(T, \mathbf{X}) \quad \text{and} \quad (R, L) \ \text{ are independent.} \tag{3}$$

The observations are also independent copies of a lifetime $Y = \max(\min(T, R), L)$ which is a twice-censored observation. Assuming a sequence of covariates $(X_i)_{1\leq i\leq n}$, one then observes the triplet $(Y_i, \delta_i, \mathbf{X}_i)_{1\leq i\leq n}$, where δ_i is the censoring indicator of the i-th observation, taking values in $\{0, 1, 2\}$. Specifically, $\delta_i = 0$ reflects the real observation T_i, while $\delta_i = 1$ and $\delta_i = 2$ represent right- and left-censoring, respectively.

As a practical application, we refer the reader to Morales et al. (1991), who investigated the causes of tree mortality on a farm. In this study, trees are planted progressively, one per day, from $t = 0$ to $t = t_0$, with the objective of analyzing a specific cause of death. The lifetime T denotes the age at death due to the cause of interest, while R represents the age at death due to other causes. At the beginning of the observation period, some trees are already dead and therefore are considered as left-censored, with their survival time denoted by L. Others die during the observation period, either from the cause of interest or from another cause, corresponding to complete and right-censored observations, respectively. Trees that remain alive at the end of the study are also right-censored.

A substantial body of literature has been dedicated to nonparametric estimation within this modeling framework. For example, Messaci (2010) introduced weighted kernel regression estimators, while Kebabi and Messaci (2012) investigated their asymptotic properties, proving pointwise and uniform almost complete consistency, as well as the convergence rate of the regression function estimator. In addition, Kitouni et al. (2015) focused on the uniform almost complete convergence of the product-limit estimator of Patilea and Rolin (2006), under more relaxed conditions, while Khardani (2019) considered the problem of Relative Error Prediction, and built a new estimator of the regression function. More recently, Benrabah et al. (2021) investigated the local linear estimation method in a twice-censoring context.

Our main goal is to introduce a novel robust regression estimation procedure tailored to the twice-censoring framework. The proposed estimator provides solid theoretical guarantees, including strong uniform consistency and an explicit convergence rate, while also opening new perspectives for dealing with complex censoring structures. In this way, our work extends the results of Lemdani and Ould-Saïd (2017) to the setting of twice-censored data.

As a complementary contribution, we construct a new Nadaraya-Watson-type estimator adapted to the twice-censoring, and establish its uniform consistency along with its convergence rate. Furthermore, given that the choice of the smoothing parameter is both essential and crucial, we propose a data-driven criterion for selecting an appropriate bandwidth, motivated by cross-validation techniques.

The remainder of the paper is organized as follows: Section 2 formulates the model and presents the construction of the proposed estimator. Section 3 gathers assumptions and states the main theoretical results concerning the strong consistency of the proposed M-estimator. In Section 4, we develop a cross-validation procedure to construct a new criterion for selecting the optimal bandwidth. The performance of the estimator is assessed through a simulation study carried out in Section 5. Moreover, Section 6 illustrates a real data application that aims to predict the time to AIDS development as a function of the initial time to HIV infection. Section 7 offers concluding remarks, while the Appendix contains auxiliary results and technical proofs.

2. Model and estimator

We first recall Model I in Patilea and Rolin (2006). Let T be a rv representing the survival time, which is right-censored by another rv R with survival function $S^R = 1 - F^R$. In this setting, one observes $\min(T, R)$, which is left-censored by a rv L, say $Y = \max(\min(T, R), L)$. Consequently, the actual observation is the triplet $(Y, \delta, \mathbf{X})$, where the censoring indicator δ is defined as follows:

$$\delta = \begin{cases} 0 & \text{if} \quad L < T < R, \\ 1 & \text{if} \quad L < R < T, \\ 2 & \text{if} \quad \min(T, R) \leq L. \end{cases}$$

This framework assumes that right- and left-censoring are two phenomena which act independently but that one can censor the other. Adapting the product-limit estimator given in Patilea and Rolin (2006), S^R is estimated by S_n^R, defined as

$$S_n^R(U_j, \infty) \quad = \quad \prod_{1 \leq l \leq j} \left\{ 1 - \frac{D_{1l}}{W_{l-1} - N_{l-1}} \right\}, \tag{4}$$

with $\{U_j, 1 \leq j \leq N\}$ are the N distinct values in increasing order of $\{Y_i\ , 1 \leq i \leq n\}$, $D_{kj} = \sum_{i=1}^{n} \mathbf{1}_{\{Y_i = U_j, \delta_i = k\}}$ and $N_j = \sum_{i=1}^{n} \mathbf{1}_{\{Y_i \leq U_j\}}$, $k = 1, 2$.

We also have that $W_{j-1} = n \prod_{j \leq l \leq N} \left\{ 1 - \frac{D_{2l}}{N_l} \right\}$.

Observe that (4) reduces to the Kaplan-Meier estimator of the censored survival function S^R whenever $L = 0$.
The product-limit estimator of F^L is the reverse Kaplan-Meier estimator given by

$$F_n^L(t) \quad = \quad \prod_{l / Y_l > t} \left\{ 1 - \frac{D_{2l}}{N_l} \right\}. \tag{5}$$

From the results of Messaci and Nemouchi (2013), we obtain the strong uniform consistency of (4), viz

$$\sup_{t \in \mathbb{R}^+} \left| S_n^R(t) - S^R(t) \right| \quad = \quad O\left(\sqrt{\frac{\log \log n}{n}} \right), \ \text{a.s.} \tag{6}$$

They also adapted the law of iterated logarithm (LIL) to $F_n^L(t)$ as described in Földes and Rejtõ (1981) to establish its strong uniform consistency rate, that is

$$\sup_{t \in \mathbb{R}^+} \left| F_n^L(t) - F^L(t) \right| = O\left(\sqrt{\frac{\log \log n}{n}} \right), \ \text{a.s.} \tag{7}$$

We now return to our model. We aim to define a new M-estimator within the twice-censoring framework, specifically designed to properly account for both left- and right-censoring effects. To this end, we introduce the following synthetic ψ-function:

$$\psi_{\mathbf{x}}^*(Y, \theta) =: \frac{\mathbf{1}_{\{\delta = 0\}}}{S^R(Y) F^L(Y)} \psi_{\mathbf{x}}(Y - \theta) (= \psi_{\mathbf{x}}^*(T, \theta)). \tag{8}$$

The latter equality holds since $Y = T$ on the event $\{\delta = 0\}$, while both quantities are equal to zero on the complementary events $\{\delta = 1\}$ and $\{\delta = 2\}$. Note that the monotony of $\psi_{\mathbf{x}}$ implies that of (8). Using conditional expectation under (3), we get

$$\begin{aligned} \mathbb{E}[\psi_{\mathbf{x}}^*(Y, \theta) | \mathbf{X} = \mathbf{x}] \quad &= \quad \mathbb{E}\left[\frac{\psi_{\mathbf{x}}(T - \theta)}{S^R(T) F^L(T)} \ \mathbb{E}[\mathbf{1}_{\{\delta = 0\}} | T, \mathbf{X}] | \mathbf{X} = \mathbf{x} \right] \\ &= \quad \int \psi_{\mathbf{x}}(t - \theta) . f_{T | \mathbf{X} = \mathbf{x}}(t) dt \\ &= \quad \mathbb{E}[\psi_{\mathbf{x}}(T - \theta) | \mathbf{X} = \mathbf{x}], \end{aligned} \tag{9}$$

where $f_{T|\mathbf{X}}(\cdot)$ is the conditional probability density function of T given $\mathbf{X}$. So, we can switch from a synthetic observed function to an unobserved function. Note that the right part of (9) is defined as follows

$$\mathbb{E}[\psi_{\mathbf{x}}(T-\theta)|\mathbf{X}=\mathbf{x}] = \frac{\int \psi_{\mathbf{x}}(t-\theta) f_{\mathbf{X},T}(\mathbf{x},t)dt}{l(\mathbf{x})},$$

where $f_{\mathbf{X},T}(\cdot,\cdot)$ is the joint probability density function of $(\mathbf{X},T)$, and $l(\mathbf{x})>0$ is the marginal density function of the covariates. Consider

$$\Psi(\mathbf{x},\theta) := \mathbb{E}[\psi_{\mathbf{x}}(T-\theta)|\mathbf{X}=\mathbf{x}]l(\mathbf{x}). \tag{10}$$

Thereby, from (1), $m(\mathbf{x})$ is a solution w.r.t. θ of $\Psi(\mathbf{x},\theta)=0$. Based on (9), we define a kernel estimator of $\Psi(\mathbf{x},\theta)$ as follows:

$$\tilde{\Psi}(\mathbf{x},\theta) := \frac{1}{nh_n^d}\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\psi^*_{\mathbf{x}}(Y_i,\theta).$$

From (8), we can write

$$\tilde{\Psi}(\mathbf{x},\theta) := \frac{1}{nh_n^d}\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}\psi_{\mathbf{x}}(Y_i-\theta). \tag{11}$$

It is evident that (11) cannot be applied directly. Therefore, we substitute S^R and F^L with their estimators given in (4) and (5), respectively, to obtain a feasible kernel estimator

$$\begin{aligned}\widehat{\Psi}(\mathbf{x},\theta) &:= \frac{1}{nh_n^d}\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{\mathbf{1}_{\{\delta_i=0\}}}{S_n^R(Y_i)F_n^L(Y_i)}\psi_{\mathbf{x}}(Y_i-\theta)\\ &=: \frac{1}{nh_n^d}\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\widehat{\psi}^*_{\mathbf{x}}(Y_i,\theta),\end{aligned} \tag{12}$$

the sum in (12) is taken for the subscripts points i so that $S_n^R(Y_i)F_n^L(Y_i)\neq 0$. This gives the proposal of the current study and permits the estimation of $\widehat{m}(x)$ which is an isolated root of $\widehat{\Psi}(\mathbf{x},\theta)$ w.r.t θ.

Remark 2.1. *In special cases, the kernel estimator defined in (12) reduces to*
• *Complete data* $(L=0, R=\infty)$

$$\widehat{\Psi}(\boldsymbol{x},\theta) =: \frac{1}{nh_n^d}\sum_{i=1}^{n} K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\psi_{\boldsymbol{x}}(T_i-\theta),$$

which was given by Tsybakov (1982) and Härdle (1984).
• *Right-censored data* $(L=0)$

$$\widehat{\Psi}_n(\boldsymbol{x},\theta) := \frac{1}{nh_d}\sum_{i=1}^{n} K_d\left(\frac{\boldsymbol{x}-X_i}{h_n}\right)\frac{\mathbf{1}_{\{\delta_i=0\}}}{S_n^R(Y_i)}\psi_{\boldsymbol{x}}(Y_i-\theta).$$

where in the right-censored case, the indicator of censoring reduces to

$$\delta = \begin{cases} 0 & \textit{if} \quad T\leq R,\\ 1 & \textit{if} \quad T>R,\end{cases}$$

and $S_n^R(.)$ reduces to the well-known Kaplan-Meier estimator when dealing with right-censoring. This is the estimator proposed by Lemdani and Ould-Saïd (2017).

3. Assumptions and main results

In the sequel, for any rv Z, we define the endpoints of its support $a_Z := \inf\{y; F^Z(y) > 0\}$ and $b_Z := \sup\{y; F^Z(y) < 1\}$. We assume that

$$sup(a_L, a_R) < a_T \text{ and } b_T < b_R. \tag{13}$$

We also consider the set $\Omega_0 = \{\mathbf{x} \in \mathbb{R}^d | l(\mathbf{x}) > 0\}$ and let Ω be a compact set included in Ω_0.
Let $\Delta := [m(\mathbf{x}) - \upsilon, m(\mathbf{x}) + \upsilon]$ be a compact neighborhood of $m(\mathbf{x})$ where $\upsilon > 0$. We thereby suppose that $m(\mathbf{x}) \in \Delta \cap [a_T, b_T]$ for any $\mathbf{x} \in \Omega$. We denote by c and C generic positive constants which may vary from line to line and $\mathcal{U}(\mathbf{x})$ a neighborhood of $\mathbf{x}$. The main results will be stated using the following assumptions :

H The bandwidth $h_n \to 0$ and satisfies for $n \to +\infty$:
(i) $\dfrac{nh_n^d}{\log n} \to \infty$ and (ii) $h_n^d = O\left(\left(\dfrac{\log^7 n}{n}\right)^{1/3}\right)$.

K (i) The kernel K_d is a bounded probability density function;
(ii) K_d satisfies the Lipschitz condition such that for all $v_1, v_2 \in \mathbb{R}^d$, there exists $\sigma > 0$ with $|K_d(v_1) - K_d(v_2)| \leq \sigma\|v_1 - v_2\|$;
(iii) $\displaystyle\int_{\mathbb{R}^d} w_j K_d(\mathbf{w})\mathbf{dw} = 0$ and $\displaystyle\int_{\mathbb{R}^d} |w_i w_j| K_d(\mathbf{w})\mathbf{dw} < \infty$, for all $1 \leq i, j \leq d$, where $\mathbf{w} = (w_1, ..., w_d)^\top$;
(iv) $\displaystyle\int_{\mathbb{R}^d} K_d^2(\mathbf{w})\mathbf{dw} < \infty$.

M $\psi_{\mathbf{x}}(.)$ is a monotonic α-lipschitzian function in $\mathbb{R}$.

P (i) The function $\Psi(., \theta)$ is twice continuously differentiable w.r.t the first component and

$$\sup_{\mathbf{x}\in\Omega}\sup_{\theta\in\Delta}\left|\frac{\partial^2\Psi(\mathbf{x},\theta)}{\partial x_i \partial x_j}\right| < \infty,$$

(ii) The functions $\mathbb{E}[\psi_{\mathbf{x}}^2(T_1 - \theta)|\mathbf{X}_1 = \mathbf{s}]$ and $l(\mathbf{s})$ are uniformly bounded on $s \in \mathcal{U}(\mathbf{x})$;
(iii)The function $\Psi(\mathbf{x}, .)$ is continuously differentiable on Δ for all $\mathbf{x} \in \Omega$ and

$$\inf_{\mathbf{x}\in\Omega}\left|\frac{\partial\Psi(\mathbf{x}, m^*(\mathbf{x}))}{\partial\theta}\right| \geq c > 0,$$

with $m^*(\mathbf{x})$ lies between $m(\mathbf{x})$ and $\widehat{m}(\mathbf{x})$;
(iv) $\mathbb{E}[|\psi_{\mathbf{x}}(T_1 - \theta)|^3] < \infty$ for $(\mathbf{x}, \theta) \in \Omega \times \Delta$.

Discussion on assumptions Assumption **H(i)** ensures the uniform consistency of the estimator whereas **H(ii)** is used to establish the convergence rate. Assumptions **K(i)** and **K(ii)** are usual in kernel estimation, **K(iii)** is dedicated to the bias term whereas **K(iv)** is used to bound the variance term. The Assumptions **P** are used to establish uniform results, **P(i)** is also needed for the bias term, **P(ii)** is classical when dealing with an unbounded function $\psi(.)$ and **P(iii)** is required for the proof of Theorem 3.1. **M** is a technical assumption used to control the fluctuation term. Moreover, it ensures the existence and the uniqueness of the solution of (1).

We now state the main result related to the uniform strong consistency with the convergence rate of $\widehat{m}(x)$ given in Theorem 3.1. This result follows directly from the next proposition.

Proposition 3.1. *Suppose that Assumptions **H**, **K**, **P(i),(ii),(iv)** and **M** hold, for n large enough, we have*

$$\sup_{\boldsymbol{x}\in\Omega}\sup_{\theta\in\Delta}|\widehat{\Psi}(\boldsymbol{x},\theta) - \Psi(\boldsymbol{x},\theta)| = O\left(\sqrt{\frac{\log n}{nh_n^d}} + h_n^2\right) \text{ a.s.}$$

Theorem 3.1. *Suppose that Assumptions **H**, **K**, **P** and **M** hold and for n large enough, we have*

$$\sup_{\boldsymbol{x}\in\Omega}|\widehat{m}(\boldsymbol{x}) - m(\boldsymbol{x})| = O\left(\sqrt{\frac{\log n}{nh_n^d}} + h_n^2\right) \text{ a.s.}$$

Remark 3.1. *It should be noted that, in the case of right-censored data, the rate obtained here improves upon that established by Lemdani and Ould-Saïd (2017) which is of order $O_{a.s}\left(h_n^2 + \sqrt{\frac{\log n}{n^{1-2\xi}h_n^d}}\right)$, for $\xi \in \left(\frac{2}{\beta+1}, \frac{1}{2}\right)$, $\beta > 3$.*

3.1. Comeback to classical regression

In the special case where $\psi_{\mathbf{x}}(\mathbf{u}) = \mathbf{u}$, the estimator $\widehat{m}(\mathbf{x})$ resulting from (12) reduces to the well known Nadaraya-Watson estimator in case of the twice-censoring, namely

$$\hat{m}_{adjNW}(\mathbf{x}) = \frac{\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \mathbf{1}_{\{\delta_i=0\}} Y_i (S_n^R(Y_i) F_n^L(Y_i))^{-1}}{\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \mathbf{1}_{\{\delta_i=0\}} (S_n^R(Y_i) F_n^L(Y_i))^{-1}}, \tag{14}$$

which is different from that proposed by Kebabi and Messaci (2012) and given by

$$\hat{r}_n(\mathbf{x}) = \frac{\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \mathbf{1}_{\{\delta_i=0\}} Y_i (S_n^R(Y_i) F_n^L(Y_i))^{-1}}{\sum_{i=1}^{n} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)}. \tag{15}$$

As a direct consequence of Theorem 3.1, we get the strong consistency of the least-squares resulting estimator $\widehat{m}_{adjNW}(\mathbf{x})$ given in (14).

Theorem 3.2. *Under the Assumptions of Theorem 3.1 and for n large enough, we have*

$$\sup_{\boldsymbol{x}\in\Omega} |\widehat{m}_{adjNW}(\boldsymbol{x}) - m(\boldsymbol{x})| = O\left(\sqrt{\frac{\log n}{n h_n^d}} + h_n^2\right) \text{ a.s.}$$

Note that under complete data ($L = 0, R = \infty$), both estimators given in (14) and (15) reduce to the Nadaraya-Watson estimator.
In kernel estimation, a well-known smoothing parameter is involved in the construction of nonparametric estimators. In the forthcoming section, we propose a robust version of a cross-validation bandwidth selection, for the M-estimator $\widehat{m}(.)$ under twice-censored data.

4. The bandwidth selection rule

Following the cross-validation approach introduced by Härdle and Marron (1985), for classical regression with complete data, we develop a robust cross-validation criterion for bandwidth selection in the M-estimation framework under twice-censored data. Under Assumption P(iii), a first-order Taylor expansion of $\widetilde{\Psi}(\mathbf{x}, \widetilde{m}(\mathbf{x}))$ around $m(\mathbf{x})$, where $\widetilde{m}(\mathbf{x})$ is defined as the zero of $\widetilde{\Psi}(\mathbf{x}, \theta)$ w.r.t θ, yields

$$\begin{aligned} \widetilde{m}(\mathbf{x}) - m(\mathbf{x}) &= \frac{\Psi(\mathbf{x}, m(\mathbf{x})) - \widetilde{\Psi}(\mathbf{x}, m(\mathbf{x}))}{\frac{\partial \widetilde{\Psi}(\mathbf{x}, \widetilde{m}^*(\mathbf{x}))}{\partial \theta}} \\ &\approx \frac{\Psi(\mathbf{x}, m(\mathbf{x})) - \widetilde{\Psi}(\mathbf{x}, m(\mathbf{x}))}{\frac{\partial \Psi(\mathbf{x}, m(\mathbf{x}))}{\partial \theta}}, \end{aligned}$$

$\widetilde{m}^*(\mathbf{x})$ lies between $\widetilde{m}(\mathbf{x})$ and $m(\mathbf{x})$. The latter approximation follows from Lemmas A2 and A3, which establish that $\widetilde{\Psi}(\mathbf{x}, \theta) \to \Psi(\mathbf{x}, \theta)$ $a.s.$ Similarly, one can show that $\frac{\partial \widetilde{\Psi}(\mathbf{x}, \theta)}{\partial \theta} \to \frac{\partial \Psi(\mathbf{x}, \theta)}{\partial \theta}$ $a.s.$ and uniformly in θ as $n \to \infty$. Assumption P(iii) ensures that $\frac{\partial \Psi(\mathbf{x}, m(\mathbf{x}))}{\partial \theta}$ is bounded away from zero and finite.
The objective of the bandwidth selection procedure is to approximate the population integrated squared error of the regression estimator and to select the bandwidth that minimizes this population criterion. In view of the above expansion, this objective is asymptotically equivalent to minimizing the integrated squared error associated with the estimating equation.

Therefore, minimizing the Integrated Squared Error (ISE) of $\widetilde{m}(\mathbf{x})$

$$ISE(\widetilde{m}, m; h_n) := \int \left(\widetilde{m}(\mathbf{x}) - m(\mathbf{x})\right)^2 l(\mathbf{x}) d\mathbf{x},$$

is asymptotically equivalent to minimizing

$$ISE(\widetilde{\Psi}, \Psi; h_n) := \int \left(\widetilde{\Psi}(\mathbf{x}, m(\mathbf{x})) - \Psi(\mathbf{x}, m(\mathbf{x}))\right)^2 l(\mathbf{x}) d\mathbf{x}.$$

Since the population estimating equation (10) satisfies $\Psi(\mathbf{x}, m(\mathbf{x})) = 0$, the above quantity becomes

$$ISE(\widetilde{\Psi}, \Psi; h_n) = \int \widetilde{\Psi}^2(\mathbf{x}, m(\mathbf{x})) l(\mathbf{x}) d\mathbf{x} - 2 \int \widetilde{\Psi}(\mathbf{x}, m(\mathbf{x})) \Psi(\mathbf{x}, m(\mathbf{x})) l(\mathbf{x}) d\mathbf{x}. \tag{16}$$

Although the second term on the right-hand side of (16) is theoretically equal to zero, we retain it in the derivation for technical convenience.
From (10), since $\Psi(\mathbf{x}, m(\mathbf{x})) = 0$ and $l(\mathbf{x}) \neq 0$, identity (9) implies that

$$\mathbb{E}\left[\psi_{\mathbf{x}}^*\left(Y, m(\mathbf{x})\right) | X = \mathbf{x}\right] = 0.$$

Therefore at the point $\theta = m(\mathbf{x})$

$$\Psi(\mathbf{x}, m(\mathbf{x})) = \mathbb{E}\left[\psi_{\mathbf{x}}^*\left(Y, m(\mathbf{x})\right) | X = \mathbf{x}\right].$$

Consequently, the second term on the right-hand side of (16) can be written as

$$\begin{aligned}
\int \widetilde{\Psi}(\mathbf{x}, m(\mathbf{x})) \Psi(\mathbf{x}, m(\mathbf{x})) l(\mathbf{x}) d\mathbf{x} &= \int \widetilde{\Psi}(\mathbf{x}, m(\mathbf{x})) \mathbb{E}\left[\psi_{\mathbf{x}}^*\left(Y, m(\mathbf{x})\right) | X = \mathbf{x}\right] l(\mathbf{x}) d\mathbf{x} \\
&= \int \mathbb{E}\left[\widetilde{\Psi}(\mathbf{x}, m(\mathbf{x})) \psi_{\mathbf{x}}^*\left(Y, m(\mathbf{x})\right) | X = \mathbf{x}\right] l(\mathbf{x}) d\mathbf{x} \\
&= \mathbb{E}\left[\widetilde{\Psi}\left(\mathbf{X}, m(\mathbf{X})\right) \psi_{\mathbf{x}}^*\left(Y, m(\mathbf{X})\right)\right].
\end{aligned}$$

Replacing the population expectation by its empirical counterpart naturally leads to the estimator

$$\frac{1}{n} \sum_{i=1}^{n} \widehat{\Psi}^{(-i)}\left(X_i, m(X_i)\right) \widehat{\psi}_{\mathbf{x}}^*\left(Y_i, m(X_i)\right),$$

where $\widehat{\Psi}^{(-i)}$ is the leave-one-out estimator of $\widetilde{\Psi}$ given by

$$\widehat{\Psi}^{(-i)}\left(X_i, m(X_i)\right) := \frac{1}{(n-1)h_n^d} \sum_{j \neq i} \frac{\mathbb{I}_{\{\delta_j = 0\}} \psi_{\mathbf{x}}\left(Y_j - m(X_i)\right)}{S_n^R\left(Z_j\right) F_n^L\left(Z_j\right)} K\left(\frac{X_i - X_j}{h_n}\right),$$

and $\widehat{\psi}_{\mathbf{x}}^*$ is defined in (12). Similarly, the first term of (16) may be approximated by

$$\frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\Psi}^{(-i)}\left(X_i, m(X_i)\right)\right)^2.$$

Consequently, the quantity $ISE(\widetilde{\Psi}, \Psi; h_n)$ can be estimated by

$$\widehat{ISE}(\widetilde{\Psi}, \Psi; h_n) = \frac{1}{n} \sum_{i=1}^{n} \left(\widehat{\Psi}^{-(i)}\left(X_i, m(X_i)\right) - \widehat{\psi}_{\mathbf{x}}^*\left(Y_i, m(X_i)\right)\right)^2 - \frac{1}{n} \sum_{i=1}^{n} \widehat{\psi}_{\mathbf{x}}^{*2}\left(Y_i, m(X_i)\right).$$

Since $\widehat{\psi}^{*2}_{\mathbf{x}}(Y_i, m(X_i))$ does not depend on h_n, minimizing $\widehat{ISE}(\widetilde{\Psi}, \Psi; h_n)$ reduces to minimizing

$$\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\Psi}^{(-i)}(X_i, m(X_i)) - \widehat{\psi}^{*}_{\mathbf{x}}(Y_i, m(X_i))\right)^2. \tag{17}$$

Finally, replacing the unknown regression function by its leave-one-out estimator $\widehat{m}^{(-i)}(X_i)$, defined as solution of $\widehat{\Psi}^{(-i)}(X_i, \theta) = 0$, the quantity in (17) reduces to the following expression denoted by RCV for Robust Cross Validation criterion

$$RCV(h_n) := \frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\psi}^{*}_{\mathbf{x}}(Y_i, \widehat{m}^{(-i)}(X_i))\right)^2, \tag{18}$$

and the robust cross-validation bandwidth selection under twice-censored data consists in choosing h_n as the minimizer of RCV such that

$$h_{RCV} := \arg\min_{h_n} RCV(h_n), \tag{19}$$

which we refer to as the robust cross-validation bandwidth.

Remark 4.1. *As a special case, when taking the identity objective function $\psi_{\boldsymbol{x}}(\boldsymbol{u}) = \boldsymbol{u}$ in (18), we obtain a new criterion of the classical Least-Squares Cross-Validation $LSCV$ under twice-censored data, given by*

$$LSCV(h) := \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}_{\{\delta_i=0\}}}{(F^L_n(Y_i)S^R_n(Y_i))^2}\left(Y_i - \widehat{m}^{(i)}(X_i)\right)^2. \tag{20}$$

This criterion reduces to the definitions below in the case of
• ***Complete data**: In the absence of left- and right-censoring $(L = 0, R = \infty)$, we obtain*

$$LSCV(h) = \frac{1}{n}\sum_{i=1}^{n}\left(T_i - \widehat{m}^{(i)}(X_i)\right)^2,$$

which is the criterion defined by Härdle and Marron (1985).
• ***Censored data (RRC)**: In the absence of left-censoring (L=0), we get*

$$LSCV(h) = \frac{1}{n}\sum_{i=1}^{n}\frac{\mathbf{1}_{\{\delta_i=0\}}}{(S^R_n(Y_i))^2}\left(Y_i - \widehat{m}^{(i)}(X_i)\right)^2,$$

where $S^R_n(.)$ reduces to the Kaplan-Meier estimator, and δ_i is as defined in Remark 2.1.

5. Simulation study

In this section, we examine the performance of our estimator $\widehat{m}(x)$ using simulated data for the cases $d = 1$ and $d = 2$. We emphasize its robustness in comparison with the classical regression estimators defined earlier. The analysis is carried out with different censoring rates, sample sizes, and regression models. All numerical experiments and simulations are implemented in MATLAB(R2022b). To ensure reproducibility, the random number generator is initialized using a fixed random seed "12345" throughout all simulations.

5.1. The one-dimensional case

The algorithm proceeds as follows:
◦ The variable of interest is generated according to the model

$$T_i = m(X_i) + \epsilon_i,\ i = 1, \ldots, n,$$

for a set of regression functions $m(.)$, where X_i is normally distributed and $\epsilon_i \sim \mathcal{N}(0, 0.04)$.
◦ The right- and left-censoring variables R_i and L_i are assumed to be exponentially distributed with adjusted parameters to achieve the desired percentage of censoring (PC), including both left-censoring (LC) and right-censoring (RC). The observed data are thus given by $(Y_i = \max(\min(T_i, R_i), L_i), \delta_i, X_i)$, $i = 1, \ldots, n$, where δ_i is the cen-

soring indicator taking values in $\{0, 1, 2\}$ as specified in Section 2.
We then compute the product-limit estimators corresponding to rv's R and L, given in (4) and (5), respectively.
∘ To calculate $\widehat{m}(x)$, the solution of (12), we use the bisection algorithm with a stopping tolerance of 10^{-6} and a maximum of 500 iterations. Specifically, it consists of choosing an interval $[m(x) - c, m(x) + c]$ for a suitable constant c (ensuring the existence of the solution) and iterating the process to get $\theta_1(x)$, $\theta_2(x)$, ..., $\theta_k(x)$ until the stopping criterion $|\theta_k(x) - \theta_{k-1}(x)| < 10^{-6}$ is satisfied or the maximum of iteration is reached. At this stage, we set $\widehat{m}(x) = \theta_k(x)$. Across all simulation settings, the algorithm converged successfully in every replication, requiring on average approximately 27 iterations, with no convergence failures observed.
∘ Throughout the study, we use a Gaussian kernel and the objective function $\psi(t) = t/\sqrt{1+t^2}$ that complies with the assumptions and ensures the robustness of the estimation.
∘ For bandwidth selection, it is important to note that, in the uncontaminated setting, all competing estimators are evaluated using the same bandwidth h_{LSCV}, selected by the classical cross-validation criterion defined in (20), thereby ensuring a fair comparison among the competing methods. In the contaminated setting, each competing estimator is evaluated using both h_{LSCV} and the proposed robust bandwidth h_{RCV} defined in (19).

5.1.1. Performance of the M-estimator

This part is conducted to study the performance of our estimator when the regression function takes both linear and non-linear forms. Here, the bandwidth is selected by minimizing the classical LSCV criterion over the grid $h \in [0.01, 1]$, with an increment of 0.02.
• **Linear model**
We consider the model $T_i = 2X_i + 5 + \epsilon_i$, with $X_i \sim \mathcal{N}(2, 1)$. We plot the average of 50 realizations of $\widehat{m}(x)$ against the true curve for $x \in [0, 4]$.
∘ In Figure 1, the sample sizes n are chosen as $n = 50$, 100 and 300 with $PC \approx 60\%$ corresponding to $LC \approx 34\%$ and $RC \approx 26\%$. The results admit sufficiently good performance even for small sizes, and, as expected, the estimates improve as n increases. We observe that our estimator closely follows the theoretical curve.
∘ In Figure 2, we fix the sample size at $n = 100$ while the censoring rate varies across 20%, 40% and 60% corresponding to the left- and right-censoring $(9\%, 11\%)$, $(23\%, 17\%)$ and $(34\%, 26\%)$, respectively. Performance appears to be slightly affected by the censoring rate.

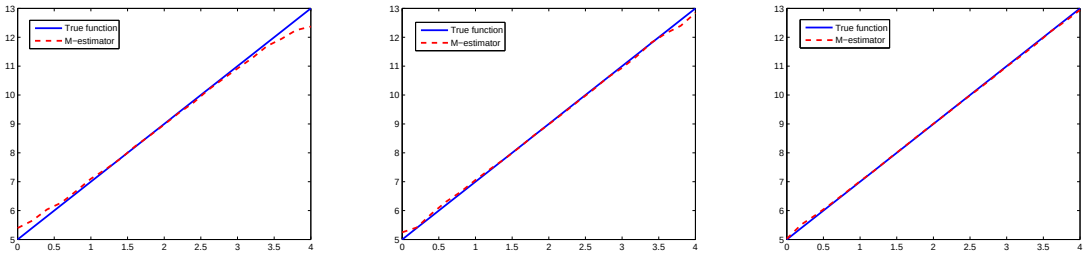

Figure 1: True function and M-estimators with $PC \approx 60\%$, for $n = 50$, 100 and 300, respectively.

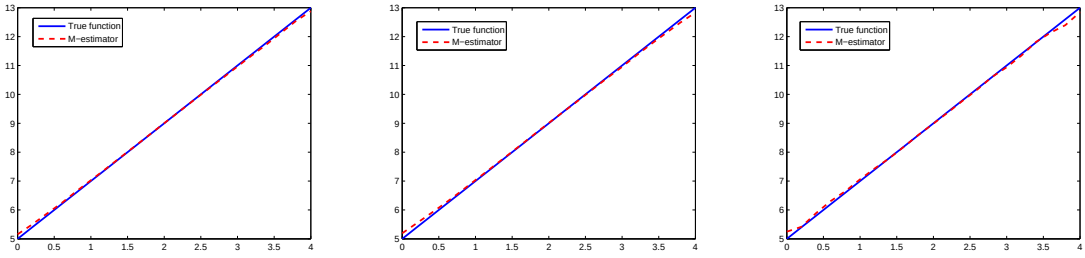

Figure 2: True function and M-estimators with $n = 100$, for $PC \approx 20\%$, 40% and 60%, respectively.

Next, we quantify both the Global Mean Square Error ($GMSE$) and the Global Standard Error (GSE) of $\widehat{m}(x)$ for $n = 50$, 100, and 300 and the corresponding cross-validated bandwidths for different censoring rates: $PC \approx$

$20\%, 40\%$ and 60%. The expression of $GMSE$ is defined by

$$GMSE = \frac{1}{MB}\sum_{\ell=1}^{M}\sum_{k=1}^{B}\left[\hat{m}_k(x_\ell) - m(x_\ell)\right]^2,$$

and that of GSE is

$$\text{GSE} = \frac{1}{\sqrt{B}}\sqrt{\frac{1}{M}\sum_{l=1}^{M}\frac{1}{B-1}\sum_{k=1}^{B}\left(\widehat{m}_k(x_l) - \overline{\widehat{m}(x_l)}\right)^2},$$

with

$$\overline{\widehat{m}(x_l)} = \frac{1}{B}\sum_{k=1}^{B}\widehat{m}_k(x_l).$$

Here, $B = 200$ is the number of replications. Note that $\hat{m}_k(x_\ell)$ is the value of $\hat{m}(x_\ell)$ in iteration k and M is the number of equidistant points x_ℓ belonging to the range $[0, 4]$. It can be seen from Table 1 that $\widehat{m}(x)$ exhibits better performance for large values of n, whereas its accuracy deteriorates as the PC rises. The optimal bandwidth values have a downward trend with the increase of n and become higher along with the PC.

Table 1: $GMSE$, GSE of $\widehat{m}$, and the optimal selector bandwidth h_{LSCV}.

	n=50			n=100			n=300		
$PC(\approx \%)$	GMSE	GSE	h_{LSCV}	GMSE	GSE	h_{LSCV}	GMSE	GSE	h_{LSCV}
20 (9,11)	0.1760	0.0289	(0.1371)	0.0298	0.0115	(0.1172)	0.0072	0.0057	(0.0990)
40 (23,17)	0.3298	0.0390	(0.1486)	0.0510	0.0150	(0.1240)	0.0104	0.0068	(0.1047)
60 (34,26)	0.8226	0.0626	(0.1671)	0.0896	0.0194	(0.1415)	0.0168	0.0086	(0.1136)

•Non-linear model

It is essential to assess the performance of our estimator under any selected model. To this end, we consider the following non-linear models :

a. Exponential model $T_i = X_i + 5 + 2\exp(-8(X_i - 3)^2) + \epsilon_i$;

b. Parabolic model $T_i = X_i^2 + 3 + \epsilon_i$;

c. Cosinusoidal model $T_i = \cos((3/2)X_i) + 5 + \epsilon_i$.

Here $X_i \sim \mathcal{N}(3, 3)$. We fix the sample size at $n = 200$, $PC \approx 60\%$ and plot the mean M-estimator versus true function for the selected models. The quality of fit seems to match that observed in the linear case, as shown in Figure 3.

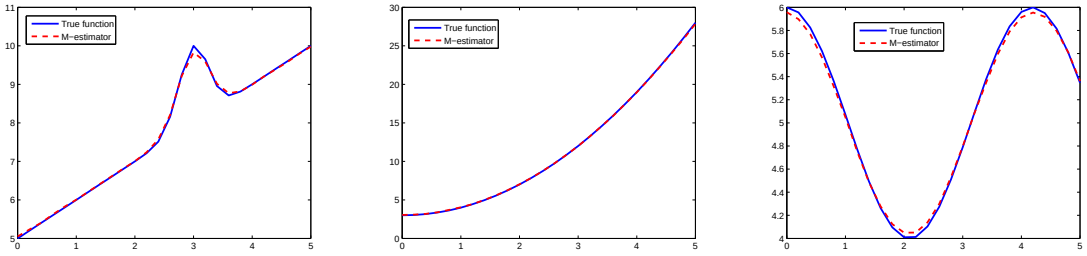

Figure 3: True functions and M-estimators with $n = 200$, $PC \approx 60\%$ for a, b and c models, respectively.

In Table 2, we report the $GMSE$ and the GSE of $\widehat{m}(x)$ based on $B = 200$ replications of our models, for $n = 200$ and $PC \approx 20\%, 40\%$ and 60%. We can see that the results summarized in this table consolidate the graphical behavior.

Table 2: $GMSE$, GSE and the optimal selector bandwidth of $\widehat{m}$, for $n = 200$.

	Exponential			Parabolic			Cosine		
$PC(\approx \%)$	GMSE	GSE	h_{LSCV}	GMSE	GSE	h_{LSCV}	GMSE	GSE	h_{LSCV}
20	0.0099	0.0067	(0.0966)	0.0449	0.0148	(0.0856)	0.0056	0.0049	(0.1727)
40	0.0132	0.0077	(0.1045)	0.1533	0.0253	(0.0816)	0.0077	0.0058	(0.1751)
60	0.0196	0.0094	(0.1105)	0.3627	0.0416	(0.1106)	0.0121	0.0074	(0.1798)

5.1.2. Performance comparisons

For the sake of comparison, the efficiency of our estimator relative to the classical least-squares estimator $\widehat{r}_n(x)$, defined in (15), is highlighted here. To ensure a fair comparison, the bandwidth h_{LSCV} for each estimator is selected by minimizing the classical cross-validation criterion over the same grid of bandwidth values $h \in [0.01, 1]$, with a step size of 0.02. Based on the linear model introduced above, we report in Table 3 the $GMSE$ and the corresponding GSE values of $\widehat{m}(x)$ and $\widehat{r}_n(x)$. The results are presented for different censoring rates and different sample sizes. The two scenarios, with $PC \approx 20\%$ and 60%, are displayed by Figures 4 and 5, respectively, for $n = 50$, 100 and 300. Since we are considering normally distributed errors, one would expect that $\widehat{r}_n(x)$ to perform similarly to, or even slightly better than, our M-estimator. However, from Table 3 and Figures 4 and 5, we can see that this is not the case.

Table 3: $GMSE$ and GSE of the proposed M-estimator and the classical estimator $\widehat{r}_n$.

	$n = 50$				$n = 100$				$n = 300$			
	$\widehat{m}$		$\widehat{r}_n$		$\widehat{m}$		$\widehat{r}_n$		$\widehat{m}$		$\widehat{r}_n$	
$PC(\approx \%)$	GMSE	GSE	GMSE	GSE	GMSE	GSE	GMSE	GSE	GMSE	GSE	GMSE	GSE
20 (9,11)	0.2708	0.0363	1.2947	0.0640	0.0301	0.0116	0.7573	0.0456	0.0072	0.0057	0.3288	0.0294
40 (23,17)	0.1388	0.0237	2.1804	0.0756	0.0509	0.0150	1.3248	0.0589	0.0104	0.0068	0.7243	0.0365
60 (34,26)	0.4442	0.0440	4.4988	0.1090	0.0900	0.0194	2.8238	0.0827	0.0166	0.0087	1.0824	0.0547

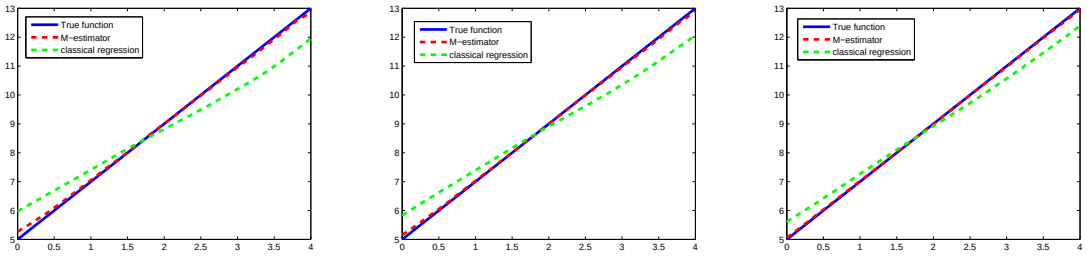

Figure 4: True function, $\widehat{m}$ and $\widehat{r}_n$ with $PC \approx 20\%$, for $n = 50$, 100 and 300, respectively.

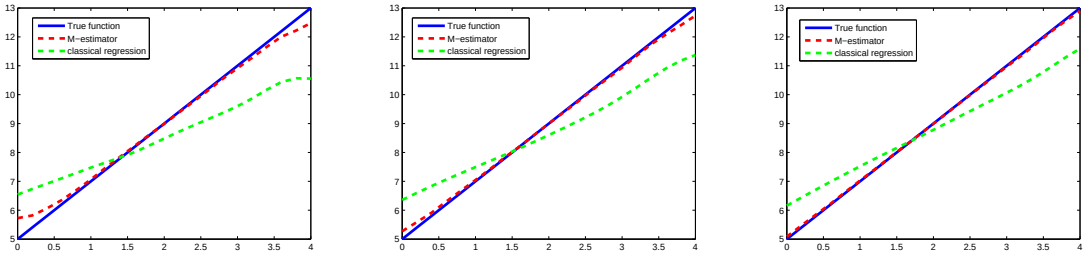

Figure 5: True function, $\widehat{m}$ and $\widehat{r}_n$ with $PC \approx 60\%$, for $n = 50$, 100 and 300, respectively.

In Kebabi's estimator $\widehat{r}_n(x)$, the effect of censoring is not reflected in the denominator which aligns with a classical approach in nonparametric estimation. This observation motivates the additional comparisons (see further compar-

isons) between $\widehat{r}_n(x)$ and the classical least-squares resulting estimator $\widehat{m}_{adjNW}(x)$ defined in (14).

5.1.3. Robustness comparisons

The robustness of $\widehat{m}$ is investigated by comparing it with that of the classical estimator $\widehat{r}_n$, under two challenging scenarios: the presence of outliers and heavy-tailed error distributions. In both settings, for each estimator, the bandwidth is selected over the same grid $h \in [0.01, 1]$ enlarged when necessary, using either the classical least-squares cross-validation criterion (h_{LSCV}) or the proposed robust cross-validation criterion (h_{RCV}). This allows us to quantify not only the robustness of the estimators, but also the benefit of the proposed robust bandwidth selection procedure for M-estimation.

• Sensitivity to outliers

To assess the sensitivity of the estimators to outliers, we artificially contaminate 2% of the response data by applying a multiplicative factor (MF) taking the values of 5, 20 and 50. The corresponding $GMSE$ values are first reported in Table 4, for different sample sizes and censoring proportions, to quantify the effect of contamination and evaluate the impact of the bandwidth selection procedure. (To keep the table concise and improve readability, only the $GMSE$ values are reported in the outlier scenarios. Since the primary objective of this experiment is to illustrate the robustness of the competing estimators under contamination rather than to assess the variability of the simulation results, the corresponding GSE values are omitted).

Table 4: $GMSE$ **of** $\widehat{m}$ **and** $\widehat{r}_n$ **under different levels of contamination.**

MF	n	$PC \approx 20\%$				$PC \approx 40\%$			
		$\widehat{m}$		$\widehat{r}_n$		$\widehat{m}$		$\widehat{r}_n$	
		$\widehat{m}(LSCV)$	$\widehat{m}(RCV)$	$\widehat{r}_n(LSCV)$	$\widehat{r}_n(RCV)$	$\widehat{m}(LSCV)$	$\widehat{m}(RCV)$	$\widehat{r}_n(LSCV)$	$\widehat{r}_n(RCV)$
5	100	0.9180	1.5151	3.7557	3.0812	1.7359	2.2033	3.4245	5.0129
	300	0.2761	0.2446	1.8411	1.9833	0.964	0.7681	2.5664	3.6719
20	100	2.1569	1.4305	58.0126	172.759	2.7917	2.2137	83.0999	183.5154
	300	0.4762	0.2642	47.6135	186.9012	1.1816	0.7791	67.8389	297.9003
50	100	2.7284	1.4305	375.0356	1.3275×10^3	2.9434	2.2137	579.0723	1.3864×10^3
	300	0.6125	0.2642	308.8019	1.4681×10^3	1.1374	0.7791	465.5143	2.1712×10^3

The numerical results clearly show the superior robustness of the proposed M-estimator. As the contamination level increases (i.e., as MF increases), the GMSE associated with the classical estimator increases dramatically, whereas that of the proposed M-estimator remains remarkably stable, illustrating its resistance to outlying observations.
• Regarding bandwidth selection, the proposed RCV criterion generally improves the estimation accuracy of the M-estimator under moderate to severe contamination. Although the classical LSCV criterion may yield slightly smaller errors for mild contamination (e.g., MF=5 and n=100), the advantage of RCV becomes clear as the contamination level or the sample size increases. These results indicate that the proposed RCV criterion provides a more reliable bandwidth selection for the proposed M-estimator in the presence of substantial outliers.
• A different behavior is observed for the classical estimator. When combined with the robust bandwidth h_{RCV}, its estimation error is even larger than that obtained with the classical bandwidth h_{LSCV}. This behavior may be explained by the fact that the proposed RCV criterion is derived from the robust estimation framework underlying the proposed M-estimator. Consequently, it is primarily intended for estimators based on this framework rather than for the classical least-squares estimator.

In view of the above numerical results, Figure 6 illustrates the robustness of the proposed M-estimator relative to the classical least-squares estimator under different levels of contamination, using the bandwidth selection criterion best suited to their estimation procedure, namely h_{RCV} for the proposed M-estimator and h_{LSCV} for the classical estimator.

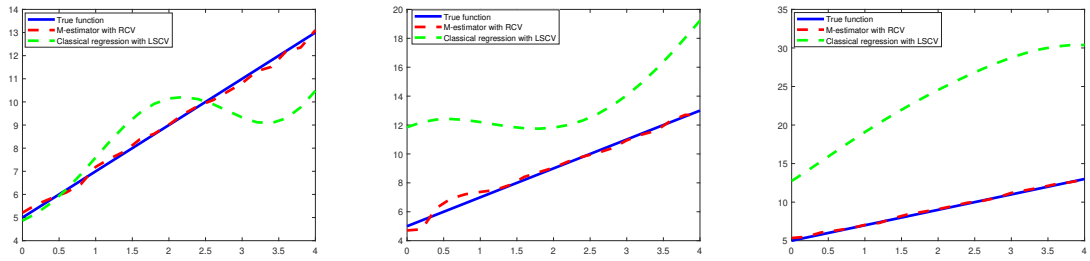

Figure 6: True function, $\widehat{m}$ and $\widehat{r}_n$ with $n = 100$, $PC \approx 40\%$ for $MF = 5$, 20 and 50, respectively.

• Heavy-tailed error distributions

To further assess the robustness of the estimators, we consider heavy-tailed errors generated from a Student's t-distribution with five degrees of freedom. As in the previous experiment, both the classical bandwidth h_{LSCV} and the proposed robust bandwidth h_{RCV} are considered. Table 5 summarizes the corresponding $GMSE$ and GSE values for different sample sizes with $PC = 20\%$, allowing us to evaluate the effect of heavy-tailed errors as well as the influence of the bandwidth selection procedure.

Table 5: $GMSE$ **(with** GSE **in parentheses) of** $\widehat{m}$ **and** $\widehat{r}_n$ **under Student's** t**-distributed errors (**$PC \approx 20\%$**).**

n	$\widehat{m}(LSCV)$	$\widehat{m}(RCV)$	$\widehat{r}_n(LSCV)$	$\widehat{r}_n(RCV)$
50	0.4537 (0.0447)	0.5224 (0.0537)	1.3480 (0.0746)	1.6762 (0.0889)
100	0.2297 (0.0319)	0.2639 (0.0381)	0.9524 (0.0525)	1.0238 (0.0652)
300	0.0879 (0.0193)	0.0966 (0.0228)	0.4219 (0.0345)	0.3979 (0.0422)

It can be seen that the proposed M-estimator performs better than the classical estimator under heavy-tailed errors. Interestingly, the classical bandwidth h_{LSCV} yields slightly smaller $GMSE$ and GSE values than the proposed robust bandwidth h_{RCV} for the M-estimator. This behavior is consistent with the results obtained under mild contamination, where the advantage of the proposed RCV criterion is less pronounced. A possible explanation is that, unlike contamination by isolated outliers, heavy-tailed errors arise from a homogeneous distribution. As a result, observations with large errors are not necessarily atypical, and reducing their influence through a robust bandwidth selection criterion is therefore less beneficial than in the presence of isolated outliers. Figure 7 displays the resulting regression curves for different sample sizes with $PC = 20\%$, where both the proposed M-estimator and the classical estimator are computed using the classical bandwidth selection criterion h_{LSCV}.

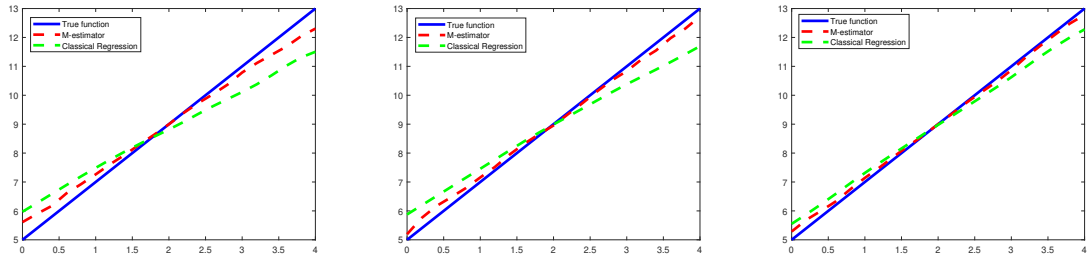

Figure 7: True function, $\widehat{m}$ and $\widehat{r}_n$ with $PC \approx 20\%$, for $n = 50$, 100 and 300, respectively.

5.1.4. Further comparisons

In this Subsection, we make some comparisons between the resulting classical estimator $\widehat{m}_{adjNW}(x)$ and the estimator $\widehat{r}_n(x)$ defined in (14) and (15), respectively.
We choose three different theoretical models:

1. Linear model $T_i = 2X_i + 5 + \epsilon_i$;
2. Sinusoidal model $T_i = \sin(2X_i) + 5 + \epsilon_i$;
3. Parabolic model $T_i = X_i^2 + 3 + \epsilon_i$;

with $X_i \sim \mathcal{N}(3,3)$. In Table 6, we report the $GMSE$ and the GSE for both estimators $\widehat{r}_n(x)$ and $\widehat{m}_{adjNW}(x)$, based on B=200 replications, for the three models, with n=100, 300, and different censoring rates. Here, the classical least-squares cross-validation criterion is applied over the same grid of bandwidth values, $h \in [0.01, 1]$ with a step size of 0.02, for both estimators. The results clearly show that, in all cases, $\widehat{r}_n(x)$ performs worse than the proposed estimator $\widehat{m}_{adjNW}(x)$, so the latter shows greater resistance to the presence of censored data.

Table 6: $GMSE$ **and** GSE **of** $\widehat{r}_n$ **and** $\widehat{m}_{adjNW}$**.**

PC	n	Linear				Sinusoidal				Parabolic			
		$\widehat{r}_n$		$\widehat{m}_{adjNW}$		$\widehat{r}_n$		$\widehat{m}_{adjNW}$		$\widehat{r}_n$		$\widehat{m}_{adjNW}$	
		GMSE	GSE	GMSE	GSE	GMSE	GSE	GMSE	GSE	GMSE	GSE	GMSE	GSE
20	100	0.7570	0.0455	0.0258	0.0111	0.3450	0.0300	0.0192	0.0091	1.6302	0.0653	0.0560	0.0159
	300	0.3291	0.0293	0.0070	0.0056	0.1685	0.0228	0.0052	0.0048	0.7422	0.0398	0.0155	0.0085
40	100	1.3270	0.0589	0.0491	0.0146	0.5174	0.0260	0.0261	0.0107	3.1294	0.0919	0.0980	0.0210
	300	0.7250	0.0424	0.0099	0.0066	0.3048	0.0274	0.0070	0.0056	1.6668	0.0642	0.0216	0.0101
60	100	2.8349	0.0829	0.0898	0.0191	0.6749	0.0283	0.0456	0.0137	6.0724	0.1331	0.1425	0.0260
	300	1.3427	0.0565	0.0168	0.0085	0.4791	0.0279	0.0115	0.0071	4.3534	0.1050	0.0396	0.0136

Some of the simulation results are displayed in Figure 8, which illustrates the improved performance of the estimator $\widehat{m}_{adjNW}$.

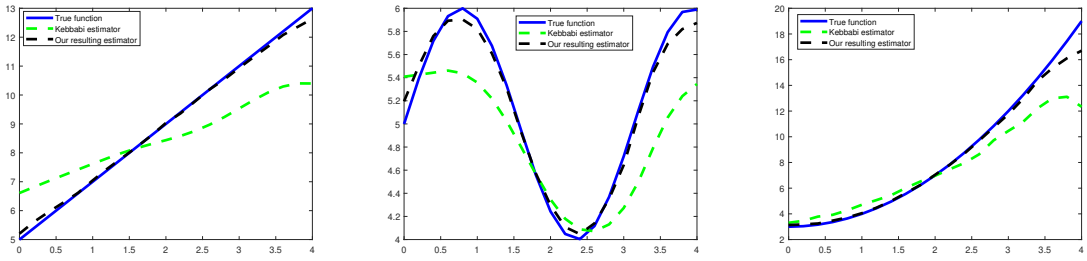

Figure 8: True function, $\widehat{r}_n$ and $\widehat{m}_{adjNW}$ with $n = 100$, $PC \approx 40\%$ for models 1,2 and 3, respectively.

Remark 5.1. *Benseradj and Guessoum (2023) showed (see their proposition 3.3) that, under the right random censoring model, the resulting estimator of the regression function defined by*

$$m_{adjNW}(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right) \mathbf{1}_{\{\delta_i=0\}} Y_i G_n^{-1}(Y_i)}{\sum_{i=1}^{n} K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right) \mathbf{1}_{\{\delta_i=0\}} G_n^{-1}(Y_i)} \tag{21}$$

(with $G_n(.)$ is the Kaplan-Meier estimator and δ_i is as defined in Remark 2.1) has a smaller asymptotic variance than

that of Carbonez et al. (1995) given by

$$m_n(\boldsymbol{x}) = \frac{\sum_{i=1}^{n} K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right) \mathbf{1}_{\{\delta_i=0\}} Y_i G_n^{-1}(Y_i)}{\sum_{i=1}^{n} K_d\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)}.$$

This result confirms the good performance of the estimator (14) which is similar to that of (21) when $L = 0$ (without left-censoring).

5.2. The two-dimensional case

5.2.1. Performance of the M-estimator

In this part, the X's represent a random vector such that $X = (X^1, X^2)$. We consequently have $T_i = m(X_i^1, X_i^2) + \epsilon_i$, $i = 1, ..., n$, where $X_i^j \sim \mathcal{N}(2, 1)$ with $j = 1, 2$ and $\epsilon_i \sim \mathcal{N}(0, 0.02^2)$. Following the same steps as in the one-dimensional case, we generate the data from the model

$$m(x_1, x_2) = 2(x_1 + x_2) + 1. \tag{22}$$

The bivariate version of the Gaussian kernel is given by : $K(x_1, x_2) = K(x_1)K(x_2)$. Moreover, we calculate $\hat{m}(x_1, x_2)$ over the range $[0, 3] \times [0, 3]$. Here, the bandwidth is fixed at $h = 0.3$. In Figure 10, we consider sample sizes $n = 50, 300$ and 500 with $PC \approx 40\%$. The corresponding M-estimators are plotted, showing the same trend as in the one-dimensional case, with an improved quality of fit as n increases.

Besides, we consider the following non-linear models for $d = 2$ to emphasize the high quality of our estimator :

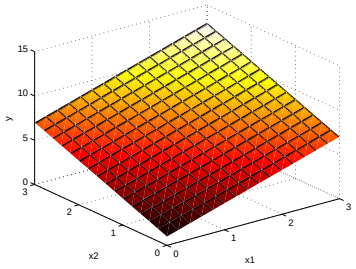

Figure 9: True function $m(x_1, x_2)$.

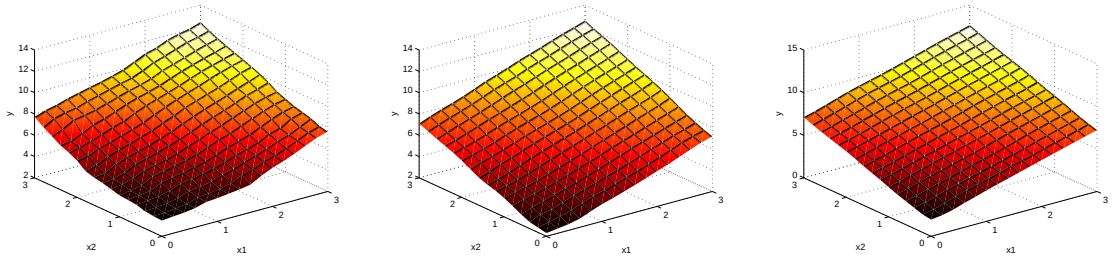

Figure 10: M-estimators $\widehat{m}$ with $PC \approx 40\%$, for $n = 50$, 300 and 500, respectively.

a. Sinusoidal model $T_i = \sin(3X_i^1) + \sin(3X_i^2) + \epsilon_i$;

b. Exponential model $T_i = \exp(-4X_i^1) + 1/2\exp(-4X_i^2) + \epsilon_i$.

The process is conducted for $X_i^j \sim \mathcal{N}(3, 3)$, $j = 1, 2$ with $n = 500$ and the percentage of censoring is $PC \approx 40\%$. Here, we take $(x_1, x_2) \in [1, 3] \times [1, 3]$. The good performance of $\widehat{m}$ appears in figures 11 and 12.

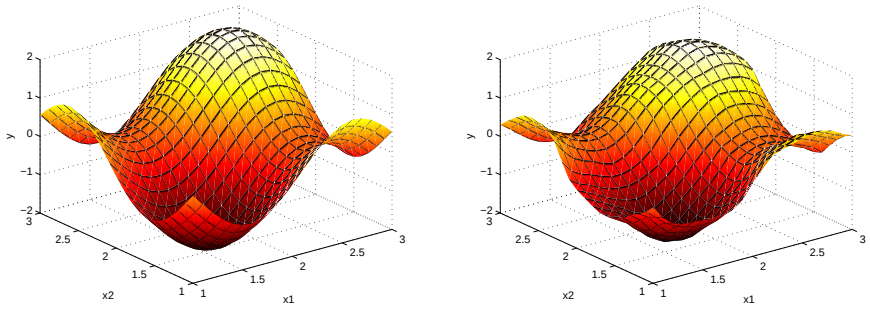

Figure 11: True function and M-estimator surface, respectively, for model (a) with $PC \approx 40\%$, for $n = 500$.

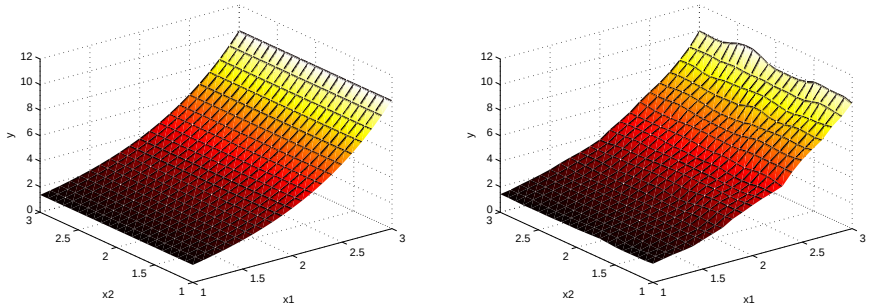

Figure 12: True function and M-estimator surface, respectively, for model (b) with $PC \approx 40\%$, for $n = 500$.

5.2.2. Performance comparisons

Here, we investigate the performance of the M-estimators $\widehat{m}(\mathbf{x})$, $\widehat{m}_{adjNW}(\mathbf{x})$ and $\widehat{r}_n(\mathbf{x})$ based on the model (22), for $d = 2$ and $n = 300$. From Figure 13, and compared with $m(x_1, x_2)$ reported in Figure 9, we observe no substantial differences in the performance of the estimates $\widehat{m}(\mathbf{x})$ and $\widehat{m}_{adjNW}(\mathbf{x})$. However, the classical estimator $\widehat{r}_n(\mathbf{x})$ appears to be less efficient.

We report in Table 7 the $GMSE$ of the estimators $\widehat{m}(\mathbf{x})$, $\widehat{m}_{adjNW}(\mathbf{x})$, and $\widehat{r}_n(\mathbf{x})$ for different values of PC. Based on the results, we confirm that our estimator $\widehat{m}(\mathbf{x})$ and the adjusted classical estimator $\widehat{m}_{adjNW}(\mathbf{x})$ perform reasonably well comparing to the classical one, particularly for small sample sizes.

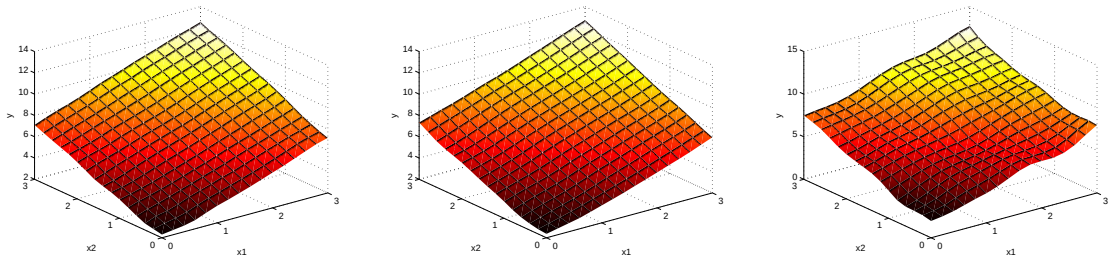

Figure 13: $\widehat{m}$, $\widehat{m}_{adjNW}$ and $\widehat{r}_n$, respectively for $n = 300$ with $PC \approx 40\%$.

Table 7: $GMSE$ **of** $\widehat{m}$**,** $\widehat{m}_{adjNW}$ **and** $\widehat{r}_n$**.**

	$PC \approx 20\%$			$PC \approx 40\%$		
n	$\widehat{m}$	$\widehat{m}_{adjNW}$	$\widehat{r}_n$	$\widehat{m}$	$\widehat{m}_{adjNW}$	$\widehat{r}_n$
50	0.5609	0.5135	3.5806	0.7722	0.7048	9.2680
300	0.1468	0.1384	0.7068	0.1861	0.1744	1.6804
500	0.1349	0.1314	0.5570	0.1296	0.1228	1.0360

6. Application to real data

We present a real data application to emphasize the effectiveness of the M-estimator in case of the twice-censoring. Furthermore, we compare the behavior of our estimator with that of the classical ones defined in (14), and (15). The data consist of measurements taken from patients infected with HIV who were following a detoxification program at the German Trias i Pujol Hospital, located in Badalona (Spain). The dataset records several events, such as the day patients started injecting drugs, the day they were diagnosed with AIDS (if available), the time of death, and whether they had AIDS at the time of death. This dataset is available in the R language database under the name `IVaids`. The most important time variable considered is the time elapsed since the first exposure to HIV through drug injection. Specifically, we explore the time to AIDS among 232 patients infected with HIV. Among these patients, 82 developed AIDS while they were following the program and are considered uncensored; 136 reached the end of the study without developing AIDS and are treated as right-censored observations. The remaining 14 patients were lost to follow-up and were subsequently found to have died with AIDS without a recorded date of AIDS diagnosis; they are therefore considered left-censored. We are thus interested in evaluating the relationship between the elapsed time from IV initiation to AIDS diagnosis (the response variable) and the date of IV initiation (the covariate).

As described by Julià and Gómez (2011), the left- and right-censoring mechanisms arise from the follow-up and observation process rather than from the biological progression of the disease itself. Consequently, the study design does not suggest that the censoring variables are related to either the response variable or the covariate, making the independence assumption $(T, X) \perp (L, R)$ a reasonable working assumption for this application. As in most observational survival studies, however, this assumption cannot be verified directly from the observed data.

To compare predictive performance, we perform B=30 random train–test splits. For each split, 80% of the data are randomly selected as the training sample to compute the estimators, while the remaining 20% constitute the test sample. Since the response variable is subject to censoring, the predictive performance is evaluated only on the uncensored observations of the test sample. It is worth noting that the dataset is heavily censored, with approximately 64% of the observations censored (58% right-censored and 6% left-censored), so that only about 36% of the response values are fully observed.

The choice of the objective function and the kernel is the same as for the simulated data. We compare the predictive performance of the proposed M-estimator, the adjusted Nadaraya–Watson-type estimator $\widehat{m}_{adjNW}(x)$, and the classical least-squares regression estimator $\widehat{r}_n(x)$. The bandwidth is selected over the grid $h \in [0.01, 1]$ with a step size of 0.01. For each estimator, the classical least-squares cross-validation (LSCV) criterion is used to determine the bandwidth .

As in the simulation study, the proposed iterative algorithm exhibits stable convergence for all random train–test splits. The average number of iterations required to reach the convergence criterion is approximately 28.

To evaluate the predictive performance, we compute the mean absolute error (MAE), defined by

$$\text{MAE} = \frac{1}{n_b} \sum_{i=1}^{n_b} \left| \hat{Y}_i - Y_i \right|,$$

where n_b denotes the number of uncensored observations in the b-th random test sample, $b = 1..., 30$. $\hat{Y}_i$ and Y_i denote the predicted and true values of the i-th patient. The prediction errors obtained from the first 10 random splits are reported in Table 8. In addition, the average MAE over 30 random train-test splits is also reported to provide an overall assessment of the predictive performance of the competing estimators.

The M-estimator and the adjusted Nadaraya–Watson estimator $\widehat{m}_{\text{adjNW}}(x)$ consistently yield similar prediction errors across the different splits. In all cases, the classical estimator $\widehat{r}_n(x)$ produces larger prediction errors. Overall, these results indicate that the proposed M-estimator and our resulting classical estimator $\widehat{m}_{\text{adjNW}}(x)$ provide better predictive performance and are consistent with the conclusions drawn in the simulation study (see Section 5). For

Table 8: MAE obtained over 30 random train-test splits.

Split	$\widehat{m}$	$\widehat{m}_{\text{adjNW}}$	$\widehat{r}_n$
1	2.4904	2.5594	4.2102
2	1.9707	2.2479	4.8553
3	1.6800	1.7973	4.1589
4	2.0984	2.4903	4.0257
5	2.7152	2.8845	5.0626
6	1.9875	1.7482	3.6910
7	2.1264	2.7185	3.1052
8	2.2217	2.6649	3.6540
9	1.7903	2.1110	4.5262
10	1.7527	1.8176	4.2419
Average over 30 random splits	2.3759	2.2078	3.9848

visual illustration, the prediction results corresponding to a few representative random train-test splits are displayed in Figure 14.

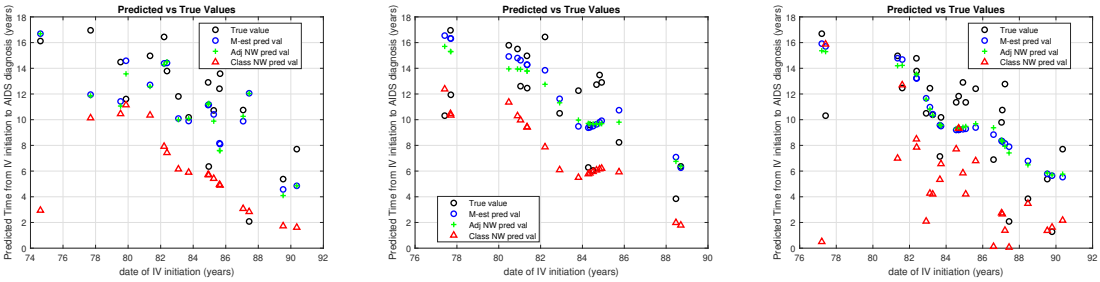

Figure 14: True and predicted AIDS diagnosis times for three random train-test splits.

Next, to assess the robustness of the competing estimators, we consider 30 random train–test splits. For each split, the training sample is contaminated by randomly selecting 2% of the observations and multiplying their response values by $MF = 5$, 10, and 20. The predictive performance is evaluated using the MAE, where the bandwidth for each estimator is selected separately using the LSCV and RCV criteria. Table 9 reports the average MAE over the 30 random train–test splits for each contamination level and reveals three main observations.

Table 9: Average MAE obtained over 30 random train-test splits for different contamination levels.

MF	LSCV			RCV		
	$\widehat{m}$	$\widehat{m}_{\text{adjNW}}$	$\widehat{r}_n$	$\widehat{m}$	$\widehat{m}_{\text{adjNW}}$	$\widehat{r}_n$
5	9.4713	20.4863	23.4460	8.4786	9.5230	28.0083
10	11.7365	42.6915	49.0172	10.4904	15.7511	49.7511
20	12.8610	81.4960	93.4866	11.4410	35.5541	98.5598

First, as expected, the MAE values increase significantly for the classical regression estimators, whereas the proposed M-estimator maintains relatively low and stable prediction errors, further confirming its robustness to outliers compared to the classical regression estimators.

Second, on average over the 30 random train–test splits, the RCV criterion yields lower prediction errors than the LSCV criterion for both the proposed M-estimator and the adjusted Nadaraya–Watson estimator. The improvement observed for the adjusted Nadaraya–Watson estimator is particularly noteworthy, as this aspect was not investigated in the simulation study. This is not unexpected since the adjusted estimator is obtained from the proposed M-estimator by replacing the robust score function ψ with the identity function. Therefore, it retains the same estimation framework, which explains why the RCV criterion remains beneficial for its bandwidth selection, unlike the classical estimator.

Finally, the classical estimator $\widehat{r}_n$ remains highly sensitive to contamination, and the use of the RCV criterion provides virtually no improvement over LSCV.

Overall, these findings support the conclusions drawn from the simulation study, while indicating that the advantage of the proposed RCV criterion becomes apparent on average over repeated random train–test splits, even under mild

contamination.
Figure 15 further illustrates this behavior by presenting the prediction results obtained under different contamination levels using the RCV bandwidth selection criterion, highlighting the stability of the proposed M-estimator compared with the classical regression estimators.

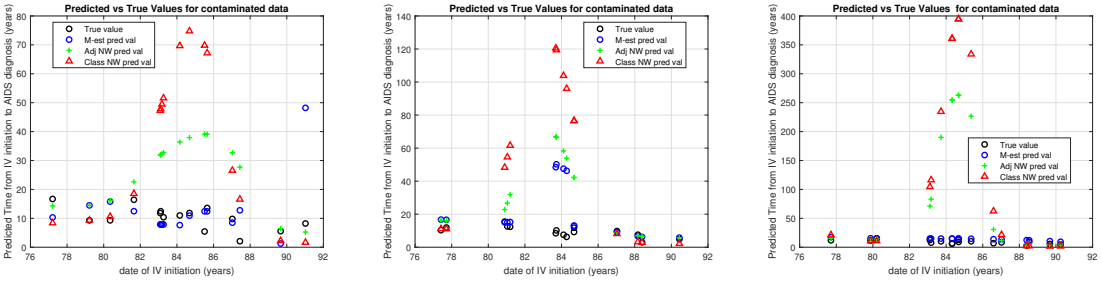

Figure 15: True and predicted AIDS diagnosis times under contamination for MF=5,10 and 20, respectively.

7. Conclusion

The aim of this paper was to establish the strong uniform consistency, together with the corresponding rate of convergence, for a new robust regression estimator when data are twice-censored, which represents a particularly complex form of incompleteness. Bandwidth selection plays a crucial role in nonparametric methods, and it is therefore essential to rely on data-driven criteria for the smoothing parameter. To this end, we propose a new criterion to compute the optimal bandwidth.
In conclusion, both the simulation study and the real-data application to AIDS survival analysis demonstrate that the proposed M-estimator provides superior robustness while maintaining excellent estimation accuracy compared with the existing classical estimators. The results also show that the proposed robust cross-validation criterion provides an effective data-driven bandwidth selection procedure for the proposed M-estimator, particularly in the presence of outliers. Furthermore, the resulting adjusted Nadaraya–Watson-type estimator substantially improves upon the existing classical estimator, exhibiting markedly better estimation accuracy while being considerably less sensitive to the presence of censoring.

Appendix. Proofs

The proof of Proposition 3.1 is based upon the following decomposition:

$$\begin{aligned}\widehat{\Psi}(\mathbf{x},\theta)-\Psi(\mathbf{x},\theta) &= \left[\widehat{\Psi}(\mathbf{x},\theta)-\tilde{\Psi}(\mathbf{x},\theta)\right]+\left[\tilde{\Psi}(\mathbf{x},\theta)-\mathbb{E}[\tilde{\Psi}(\mathbf{x},\theta)]\right] \\ &+ \left[\mathbb{E}[\tilde{\Psi}(\mathbf{x},\theta)]-\Psi(\mathbf{x},\theta)\right] \\ &=: \Xi_1+\Xi_2+\Xi_3.\end{aligned} \tag{A1}$$

Thanks to the following lemmas, the right terms in (A1) will be studied. Ξ_3 denotes the bias term (Lemma A1), whereas Ξ_2 denotes the fluctuation term (Lemma A2) and Ξ_1 treats the negligible term (Lemma A3).

Lemma A1. *Under Assumptions **K(i)**, **K(iii)** and **P(i)**, for n large enough we have*

$$\sup_{\boldsymbol{x}\in\Omega}\sup_{\theta\in\Delta}\left|\mathbb{E}[\tilde{\Psi}(\boldsymbol{x},\theta)]-\Psi(\boldsymbol{x},\theta)\right|=O(h_n^2),\quad \text{a.s.}$$

Proof. From (11), we get

$$\begin{aligned}
\mathbb{E}[\tilde{\Psi}(\mathbf{x},\theta)] &= \frac{1}{h_n^d}\mathbb{E}\left[K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}\psi_{\mathbf{x}}(Y_i-\theta)\right]\\
&= \frac{1}{h_n^d}\mathbb{E}\left[K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\mathbb{E}\left[\frac{\mathbf{1}_{\{L_1\le T_1\le R_1\}}}{S^R(T_1)F^L(T_1)}\psi_{\mathbf{x}}(T_1-\theta)|\mathbf{X}_1\right]\right]\\
&= \frac{1}{h_n^d}\int_{\mathbb{R}^d}K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)\mathbb{E}\left[\psi_{\mathbf{x}}(T_1-\theta)|\mathbf{X}_1=\mathbf{u}\right]l(\mathbf{u})d\mathbf{u}\\
&= \int_{\mathbb{R}^d}K_d(\mathbf{z})\,\mathbb{E}\left[\psi_{\mathbf{x}}(T_1-\theta)|\mathbf{X}_1=\mathbf{x}-\mathbf{z}h_n\right]l(\mathbf{x}-\mathbf{z}h_n)d\mathbf{z}. \qquad \text{(A2)}
\end{aligned}$$

Using (10) and Assumption **K(i)**

$$\Xi_3=\int_{\mathbb{R}^d}K_d(\mathbf{z})[\Psi(\mathbf{x}-\mathbf{z}h_n,\theta)-\Psi(\mathbf{x},\theta)]d\mathbf{z}.$$

Expanding $\Psi(\mathbf{x}-\mathbf{z}h_n,\theta)$ up to order 2 around $\mathbf{x}$, we find

$$\Psi(\mathbf{x}-\mathbf{z}h_n,\theta)=\Psi(\mathbf{x},\theta)-h_n\left[z_1\frac{\partial\Psi(\mathbf{x},\theta)}{\partial x_1}+\cdots+z_d\frac{\partial\Psi(\mathbf{x},\theta)}{\partial x_d}\right]+\frac{h_n^2}{2}\sum_{1\le i\le d}\sum_{1\le j\le d}z_iz_j\frac{\partial^2\Psi(\mathbf{x}^*,\theta)}{\partial x_i\partial x_j},$$

where $(\mathbf{x}^*,\theta)$ lies between $(\mathbf{x}-\mathbf{z}h_n,\theta)$ and $(\mathbf{x},\theta)$. Under Assumption **K(iii)**, we can restrict the expression Ξ_3 to

$$\Xi_3=\frac{h_n^2}{2}\sum_{1\le i\le d}\sum_{1\le j\le d}\int_{\mathbb{R}^d}K_d(z).z_iz_j\frac{\partial^2\Psi(\mathbf{x}^*,\theta)}{\partial x_i\partial x_j}dz.$$

Again, thanks to both Assumptions **K(iii)** and **P(i)**, we can set

$$\begin{aligned}
\sup_{\mathbf{x}\in\Omega}\sup_{\theta\in\Delta}|\Xi_3| &\le c\frac{h_n^2}{2}\sum_{1\le i\le d}\sum_{1\le j\le d}\int_{\mathbb{R}^d}|z_iz_j|K_d(\mathbf{z})d\mathbf{z}\\
&= O(h_n^2),
\end{aligned}$$

which concludes the proof. □

Lemma A2. *Using Assumptions **H**, **K**, **M** and **P**, for n large enough we have*

$$\sup_{\boldsymbol{x}\in\Omega}\sup_{\theta\in\Delta}\left|\tilde{\Psi}(\boldsymbol{x},\theta)-\mathbb{E}[\tilde{\Psi}(\boldsymbol{x},\theta)]\right|=O\left(h_n^2+\sqrt{\frac{\log n}{nh_n^d}}\right),\quad \text{a.s.}$$

Proof. To establish the strong consistency of the fluctuation term Ξ_2, we apply Bernstein's inequality and compact covering techniques. The compactness of Δ allows us to divide the latter into a finite number of sub-intervals $J_1,\ldots,J_{\wp_n}$, centered at θ_k such that $k\in\{1,\ldots,\wp_n\}$, then we have $|\theta-\theta_k|\le h_n^{d+2}$ and $\wp_nh_n^{d+2}\le c$.
Note that Ξ_2 is upper bounded as follows

$$\begin{aligned}
\sup_{\mathbf{x}\in\Omega}\sup_{\theta\in\Delta}\left|\tilde{\Psi}(\mathbf{x},\theta)-\mathbb{E}\left[\tilde{\Psi}(\mathbf{x},\theta)\right]\right| &\le \max_{1\le k\le\wp_n}\sup_{\mathbf{x}\in\Omega}\sup_{\theta\in\Delta}\left|\tilde{\Psi}(\mathbf{x},\theta)-\tilde{\Psi}(\mathbf{x},\theta_k)\right|\\
&+ \max_{1\le k\le\wp_n}\sup_{\mathbf{x}\in\Omega}\sup_{\theta\in\Delta}\left|\mathbb{E}\left[\tilde{\Psi}(\mathbf{x},\theta_k)\right]-\mathbb{E}\left[\tilde{\Psi}(\mathbf{x},\theta)\right]\right|\\
&+ \max_{1\le k\le\wp_n}\sup_{\mathbf{x}\in\Omega}\left|\tilde{\Psi}(\mathbf{x},\theta_k)-\mathbb{E}\left[\tilde{\Psi}(\mathbf{x},\theta_k)\right]\right|\\
&=: I_{1n}+I_{2n}+I_{3n}.
\end{aligned}$$

By simple calculation, we deal with I_{1n}. Under Assumptions **K(i)** and **M**, we have

$$\begin{aligned}\left|\tilde{\Psi}(\mathbf{x},\theta)-\tilde{\Psi}(\mathbf{x},\theta_k)\right| &\leq \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)|\psi_{\mathbf{x}}(Y_i-\theta)-\psi_{\mathbf{x}}(Y_i-\theta_k)|\\ &\leq \frac{\|K_d\|_\infty\,\alpha|\theta-\theta_k|}{h_n^d S^R(b_T)F^L(a_T)}\\ &\leq \frac{\|K_d\|_\infty\,\alpha h_n^2}{S^R(b_T)F^L(a_T)}\\ &= O(h_n^2),\end{aligned}$$

in view of condition (13), $S^R(b_T)>0$ and $F^L(a_T)>0$. Then, $I_{1n}=O(h_n^2)$ a.s.
As for I_{2n}, it holds straightforwardly from I_{1n} using the fact that $|\mathbb{E}(I)|\leq\mathbb{E}|I|$ for any rv I. Hence, $I_{2n}=O(h_n^2)$.
It remains to give the upper bound of I_{3n} using Bernstein's inequality. Recall that $\psi_{\mathbf{x}}(.)$ is not necessarily bounded, we use for this purpose and for all $\mathbf{x}\in\Omega$ the following expression

$$\tilde{\Psi}(\mathbf{x},\theta)=\frac{1}{nh_n^d}\sum_{i=1}^{n}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\psi_{\mathbf{x}}^*(Y_i,\theta)\mathbf{1}_{\{|\psi_{\mathbf{x}}(Y_i-\theta)|\leq\beta_n\}}+\frac{1}{nh_n^d}\sum_{i=1}^{n}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\psi_{\mathbf{x}}^*(Y_i,\theta)\mathbf{1}_{\{|\psi_x(Y_i-\theta)|>\beta_n\}}.$$

Then, it can be written as follows

$$\tilde{\Psi}(\mathbf{x},\theta)=:\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta)+\Re_n(\mathbf{x},\theta), \tag{A3}$$

and the purpose is to prove that $\Re_n(\mathbf{x},\theta_k)$ is a negligible term.
We point out that β_n is infinite and make choice that $\beta_n=\left(n\log^2 n\right)^{1/3}$. Therefore, we write

$$\begin{aligned}\max_{1\leq k\leq\wp_n}\sup_{\mathbf{x}\in\Omega}\left|\tilde{\Psi}(\mathbf{x},\theta_k)-\mathbb{E}\left[\tilde{\Psi}(\mathbf{x},\theta_k)\right]\right| &\leq \max_{1\leq k\leq\wp_n}\sup_{\mathbf{x}\in\Omega}\left|\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta_k)-\mathbb{E}\left[\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta_k)\right]\right|\\ &+ \max_{1\leq k\leq\wp_n}\sup_{\mathbf{x}\in\Omega}\left|\Re_n(\mathbf{x},\theta_k)-\mathbb{E}\left[\Re_n(\mathbf{x},\theta_k)\right]\right|.\end{aligned}$$

Observe that

$$\begin{aligned}\left|\mathbb{E}\left[\Re_n(\mathbf{x},\theta_k)\right]\right| &\leq \mathbb{E}\left|\frac{1}{nh_n^d}\sum_{i=1}^{n}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\psi_{\mathbf{x}}^*(Y_i,\theta_k)\mathbf{1}_{\{|\psi_{\mathbf{x}}(Y_i-\theta_k)|>\beta_n\}}\right|\\ &\leq \mathbb{E}\left[\frac{1}{nh_n^d}\sum_{i=1}^{n}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}|\psi_{\mathbf{x}}(Y_i-\theta_k)|\mathbf{1}_{\{|\psi_{\mathbf{x}}(Y_i-\theta_k)|>\beta_n\}}\right].\end{aligned}$$

Proceeding as in the proof of Lemma A1, and using Assumptions **K(i)** and **P(ii)**, we get

$$\begin{aligned}\left|\mathbb{E}\left[\Re_n(\mathbf{x},\theta_k)\right]\right| &\leq \int_{\mathbb{R}^d}K_d(\mathbf{z})\,\mathbb{E}\left[|\psi_{\mathbf{x}}(T_1-\theta_k)|\mathbf{1}_{\{|\psi_{\mathbf{x}}(T_1-\theta_k)|>\beta_n\}}|\mathbf{X}_1=\mathbf{x}-\mathbf{z}h_n\right]l(\mathbf{x}-\mathbf{z}h_n)d\mathbf{z}\\ &\leq \frac{1}{\beta_n}\int_{\mathbb{R}^d}K_d(\mathbf{z})\,\mathbb{E}\left[\psi_{\mathbf{x}}^2(T_1-\theta_k)|\mathbf{X}_1=\mathbf{x}-\mathbf{z}h_n\right]l(\mathbf{x}-\mathbf{z}h_n)d\mathbf{z}\\ &\leq \frac{c}{\beta_n}.\end{aligned} \tag{A4}$$

On the one hand, using Markov's inequality and Assumption **P(iv)** we get

$$\begin{aligned}\mathbb{P}\left(|\psi_{\mathbf{x}}(T_n-\theta_k)|>\beta_n\right) &\leq \frac{\mathbb{E}|\psi_{\mathbf{x}}(T_n-\theta_k)|^3}{\beta_n^3}\\ &\leq \frac{c}{\beta_n^3}.\end{aligned}$$

Since

$$\begin{aligned}\sum_{n=1}^{\infty}\mathbb{P}\left(|\psi_{\mathbf{x}}(T_n-\theta_k)|>\beta_n\right) &\leq c\sum_{n=1}^{\infty}\frac{1}{\beta_n^3}\\ &=: c\sum_{n=1}^{\infty}\frac{1}{n\log^2 n}<\infty,\end{aligned}$$

the Borel-Cantelli's Lemma yields

$$\mathbb{P}\left(|\psi_{\mathbf{x}}(T_n-\theta_k)|>\beta_n\right)=0, \text{ infinitely often.}$$

Hence, there exists a set of event Ω_0 such that $\mathbb{P}(\Omega_0)=1$ and, for every $\omega\in\Omega_0$, there exists a finite integer $N_0(\omega)$ satisfying

$$|\psi_{\mathbf{x}}(T_n(\omega)-\theta_k)|\leq\beta_n,\quad \forall n\geq N_0(\omega).$$

Since (β_n) is increasing, we also have

$$|\psi_{\mathbf{x}}(T_i(\omega)-\theta_k)|\leq\beta_n,\quad N_0(\omega)\leq i\leq n,\quad \text{a.s.}$$

Fix now an arbitrary $\omega\in\Omega_0$. In the remainder of the proof, we work pathwise and omit the dependence on ω for notational simplicity. Consequently, $N_0:=N_0(\omega)<\infty$, while X_i, Y_i (eventually T_i) and δ_i denote the corresponding realizations. Therefore, for sufficiently large n, $\Re_n(\mathbf{x},\theta_k)$ can be decomposed as

$$\begin{aligned}\Re_n(\mathbf{x},\theta_k)=&\frac{1}{nh_n^d}\sum_{i=1}^{N_0-1}K_d\left(\frac{\mathbf{x}-X_i}{h_n}\right)\frac{\mathbb{I}_{\{\delta_i=0\}}}{S^R(T_i)F^L(T_i)}\psi_{\mathbf{x}}(T_i-\theta_k)\mathbb{I}_{\{|\psi_{\mathbf{x}}(T_i-\theta_k)|>\beta_n\}}\\ &+\frac{1}{nh_n^d}\sum_{i=N_0}^{n}K_d\left(\frac{\mathbf{x}-X_i}{h_n}\right)\frac{\mathbb{I}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}\psi_{\mathbf{x}}(T_i-\theta_k)\mathbb{I}_{\{|\psi_{\mathbf{x}}(T_i-\theta_k)|>\beta_n\}}.\end{aligned}$$

For every $i\geq N_0$, $\mathbb{I}_{\{|\psi_{\mathbf{x}}(T_i-\theta_k)|>\beta_n\}}=0$. Hence, the second sum vanishes.
On the other hand, the first sum contains only the finite number of indices $i=1,...,N_0-1$. Using Assumption **K(i)**, the boundedness of the kernel K_d, together with the fact that $S^R(\cdot)$ and $F^L(\cdot)$ are bounded away from zero on the support of T, we obtain

$$\sum_{i=1}^{N_0-1}K_d\left(\frac{\mathbf{x}-X_i}{h_n}\right)\frac{\mathbb{I}_{\{\delta_i=0\}}}{S^R(T_i)F^L(T_i)}\psi_{\mathbf{x}}(T_i-\theta_k)\mathbb{I}_{\{\psi_{\mathbf{x}}(T_i-\theta_k)>\beta_n\}}\leq\frac{(N_0-1)\|K_d\|_\infty}{S^R(b_T)F^L(a_T)}\max_{1\leq i\leq N_0-1}\left\{\psi_{\mathbf{x}}(T_i-\theta_k)\right\}. \tag{A5}$$

Now, for every fixed $i\leq N_0-1$,

$$\max_{1\leq k\leq\wp_n}\left\{\psi_{\mathbf{x}}(T_i-\theta_k)\right\}\leq\sup_{\theta\in\Delta}\left\{\psi_{\mathbf{x}}((T_i-\theta)\right\}<\infty,$$

since, $\{\theta_k\}_{k=1}^{\wp_n}\subset\Delta$. Moreover, by Assumption **M**, $\theta\to\psi_{\mathbf{x}}(T_i-\theta)$, is continuous on the compact set Δ.
Therefore, by the Weierstrass theorem,

$$\sup_{\theta\in\Delta}\left\{\psi_{\mathbf{x}}(T_i-\theta)\right\}<\infty,\quad \text{for every fixed } i\leq N_0-1.$$

Since only the finite indices $i=1,...,N_0-1$ are involved, we obtain

$$\max_{1\leq i\leq N_0-1}\sup_{\theta\in\Delta}\left\{\psi_{\mathbf{x}}(Y_i-\theta)\right\}=:C_\omega<\infty.$$

Therefore, it follows from (A5)

$$\max_{1\leq k\leq\wp_n}\sup_{\mathbf{x}\in\Omega}\Re_n(\mathbf{x},\theta_k)\leq\frac{C_\omega}{nh_n^d}. \tag{A6}$$

Finally, since the above argument holds for an arbitrary $\omega\in\Omega_0$, where $P(\Omega_0)=1$, it follows from (A4), (A5) and (A6) that

$$\max_{1\leq k\leq\wp_n}\sup_{\mathbf{x}\in\Omega}|\Re_n(\mathbf{x},\theta_k)-\mathbb{E}\left[\Re_n(\mathbf{x},\theta_k)\right]|=O\left(\frac{1}{nh_n^d}\right)+O\left(\frac{1}{(n\log n^2)^{1/3}}\right)\qquad\text{a.s.}$$

We now turn to the other term of (A3). To this end, we again use the covering techniques. So, Ω can be covered by a finite number $b_{x,n}$ of balls $B_j(\mathbf{x}_j,\omega_n)$ centred at $\mathbf{x}_j=(x_{j,1},...,x_{j,d})$ such that for all $\mathbf{x}\in\Omega$, there exist integers $j\in\{1,...,b_{x,n}\}$ so that $\|\mathbf{x}-\mathbf{x}_j\|\leq\omega_n$ with $\omega_n=\frac{h_n^{d+3}}{\beta_n}$. Then, as Ω is bounded, let c_2 be a positive generic constant

satisfying $b_{x,n}\omega_n \leq c_2$.

$$\begin{aligned}\max_{1\leq k\leq \wp_n}\sup_{\mathbf{x}\in\Omega}\left|\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta_k)-\mathbb{E}\left[\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta_k)\right]\right| &\leq \max_{1\leq k\leq \wp_n}\max_{1\leq j\leq b_{x,n}}\sup_{\mathbf{x}\in B_j}\left|\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta_k)-\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)\right| \\ &+ \max_{1\leq k\leq \wp_n}\max_{1\leq j\leq b_{x,n}}\sup_{\mathbf{x}\in B_j}\left|\mathbb{E}\left[\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)\right]-\mathbb{E}\left[\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta_k)\right]\right| \\ &+ \max_{1\leq k\leq \wp_n}\max_{1\leq j\leq b_{x,n}}\left|\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)-\mathbb{E}\left[\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)\right]\right| \\ &=: I'_{1n}+I'_{2n}+I'_{3n}.\end{aligned}$$

For I'_{1n}, by virtue of Assumption **K(ii)** one may observe that

$$\left|K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)-K_d\left(\frac{\mathbf{x}_j-\mathbf{X}_i}{h_n}\right)\right|\leq \sigma\left\|\frac{\mathbf{x}-\mathbf{x}_j}{h_n}\right\|.$$

Since, $\forall \mathbf{x}\in B_j$, we write

$$\begin{aligned}\sup_{\mathbf{x}\in B_j}\left|\tilde{\Psi}_{\beta_n}(\mathbf{x},\theta_k)-\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)\right| &\leq \frac{\sigma}{nh_n^d}\sum_{i=1}^{n}\frac{\|\mathbf{x}-\mathbf{x}_j\|}{h_n}\frac{\mathbf{1}_{\{\delta_i=0\}}|\psi_{\mathbf{x}}(Y_i-\theta_k)|}{S^R(Y_i)F^L(Y_i)}\mathbf{1}_{\{|\psi_{\mathbf{x}}(Y_i-\theta_k)|\leq\beta_n\}} \\ &\leq \frac{\sigma\omega_n\beta_n}{h_n^{d+1}S^R(b_T)F^L(a_T)}.\end{aligned}$$

Therefore $I'_{1n}=O(h_n^2)$ a.s. as $n\to\infty$. Similarly, we get $I'_{2n}=O(h_n^2)$.

The last step is dedicated to I'_{3n}, set

$$\Lambda_i(\mathbf{x}_j,\theta_k):=\frac{1}{\beta_n}\left\{K_d\left(\frac{\mathbf{x}_j-\mathbf{X}_i}{h_n}\right)\psi^*_{\mathbf{x}}(Y_i,\theta_k)\mathbf{1}_{\{|\psi_{\mathbf{x}}(Y_i-\theta_k)|\leq\beta_n\}}-\mathbb{E}\left[K_d\left(\frac{\mathbf{x}_j-\mathbf{X}_i}{h_n}\right)\psi^*_{\mathbf{x}}(Y_i,\theta_k)\mathbf{1}_{\{|\psi_{\mathbf{x}}(Y_i-\theta_k)|\leq\beta_n\}}\right]\right\}.$$

Consequently,

$$\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)-\mathbb{E}\left[\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)\right]=\frac{\beta_n}{nh_n^d}\sum_{i=1}^{n}\Lambda_i(\mathbf{x}_j,\theta_k)$$

for all $i=1,\ldots,n,\ j=1,\ldots,b_{x,n}$ and $k=1,\ldots,\wp_n$.

Observe that $\mathbb{E}[\Lambda_i]=0$ and $|\Lambda_i|\leq C<\infty$ for $C=\dfrac{2\|K_d\|_\infty}{S^R(b_T)F^L(a_T)}$. These properties are sufficient to apply Bernstein's inequality, viz

$$\begin{aligned}\mathbb{P}\left(\left|\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)-\mathbb{E}[\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)]\right|>\varepsilon\right) &= \mathbb{P}\left(\left|\sum_{i=1}^{n}\Lambda_i(\mathbf{x}_j,\theta_k)\right|>\frac{nh_n^d\varepsilon}{\beta_n}\right) \\ &\leq 2\exp\left(-\frac{nh_n^{2d}\varepsilon^2}{2\beta_n^2(S^2+\varepsilon\frac{h_n^d}{\beta_n}C)}\right), \qquad \text{(A7)}\end{aligned}$$

where $S^2=\mathrm{var}(\Lambda_i(\mathbf{x}_j,\theta_k))$. Using the classical expectation techniques and Assumptions **K(iv)**-**P(ii)**, we have

$$\begin{aligned}S^2 &\leq \mathbb{E}\left[\frac{1}{\beta_n^2}K_d^2\left(\frac{\mathbf{x}_j-\mathbf{X}_i}{h_n}\right)\psi^{*2}_{\mathbf{x}}(Y_i,\theta_k)\mathbf{1}_{\{|\psi_{\mathbf{x}}(Y_i-\theta_k)|\leq\beta_n\}}\right] \\ &\leq \frac{1}{\beta_n^2}\mathbb{E}\left[K_d^2\left(\frac{\mathbf{x}_j-\mathbf{X}_i}{h_n}\right)\mathbb{E}\left[\frac{\psi^2_{\mathbf{x}}(T_1-\theta_k)}{S^R(T_1)F^L(T_1)}|\mathbf{X}_1\right]\right] \\ &\leq \frac{1}{\beta_n^2S^R(b_T)F^L(a_T)}\int_{\mathbb{R}^d}K_d^2\left(\frac{\mathbf{x}_j-u}{h_n}\right)\mathbb{E}\left[\psi^2_{\mathbf{x}}(T_1-\theta_k)|X_1=\mathbf{u}\right]l(\mathbf{u})d\mathbf{u} \\ &= \frac{h_n^d}{\beta_n^2S^R(b_T)F^L(a_T)}\int_{\mathbb{R}^d}K_d^2(\mathbf{z})\mathbb{E}\left[\psi^2_{\mathbf{x}}(T_1-\theta_k)|\mathbf{X}_1=\mathbf{x}-h_n\mathbf{z}\right]l(\mathbf{x}-h_n\mathbf{z})d\mathbf{z} \\ &= O\left(\frac{h_n^d}{\beta_n^2}\right).\end{aligned}$$

Consequently,

$$S^2=O\left(\frac{h_n^d}{\beta_n^2}\right).$$

It follows from (A7) and letting $\varepsilon = \varepsilon_0\sqrt{\frac{\log n}{nh_n^d}}$ that

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq k\leq \wp_n}\max_{1\leq j\leq b_{x,n}}\left|\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)-\mathbb{E}[\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)]\right|>\varepsilon\right) &= \mathbb{P}\left(\max_{1\leq k\leq \wp_n}\max_{1\leq j\leq b_{x,n}}\left|\sum_{i=1}^{n}\Lambda_i(\mathbf{x}_j,\theta_k)\right|>\frac{nh_n^d\varepsilon}{\beta_n}\right)\\
&\leq \sum_{k=1}^{\wp_n}\sum_{j=1}^{b_{x,n}}\mathbb{P}\left(\left|\sum_{i=1}^{n}\Lambda_i(\mathbf{x}_j,\theta_k)\right|>\frac{nh_n^d\varepsilon}{\beta_n}\right)\\
&\leq 2\wp_n b_{x,n}\exp\left(\frac{-nh_n^{2d}\varepsilon^2}{2\beta_n^2\left(\frac{ch_n^d}{\beta_n^2}+\varepsilon C\frac{h_n^d}{\beta_n}\right)}\right)\\
&\leq \frac{c\beta_n}{h_n^{2d+5}}\exp\left(\frac{-\frac{\varepsilon_0^2}{2}\log n}{\left(c+C\varepsilon_0\left(n\log^2 n\right)^{1/3}\left(\frac{\log n}{nh_n^d}\right)^{1/2}\right.}\right)\\
&= \frac{c\beta_n}{h_n^{2d+5}}\exp\left(\frac{-\frac{\varepsilon_0^2}{2}\log n}{c+C\varepsilon_0\left(\frac{(\log n)^{\frac{7}{3}}}{n^{\frac{1}{3}}h_n^d}\right)^{1/2}}\right)\\
&\leq c\frac{\beta_n}{\left(\sqrt{nh_n^d}\right)^{2\left(\frac{2d+5}{d}\right)}}n^{-c\varepsilon_0^2+\frac{2d+5}{d}}. \qquad \text{(A8)}
\end{aligned}$$

From Assumptions **H(i)-(ii)**, we have $\frac{\beta_n}{\left(\sqrt{nh_n^d}\right)^{2\left(\frac{2d+5}{d}\right)}}=o(1)$. Hence, taking ε_0 sufficiently large, the second term in (A8) is the general term of a convergent series. Then, we get

$$\sum_{n\geq 1}\mathbb{P}\left(\max_{1\leq k\leq \wp_n}\max_{1\leq j\leq b_{x,n}}\left|\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)-\mathbb{E}[\tilde{\Psi}_{\beta_n}(\mathbf{x}_j,\theta_k)]\right|\geq\varepsilon_0\sqrt{\frac{\log n}{nh_n^d}}\right)<\infty$$

By the use of Borel-Cantelli's Lemma, we find $I'_{3n}=O\left(\sqrt{\frac{\log n}{nh_n^d}}\right)$ a.s. as $n\to\infty$.

Finally, $I_{3n}=O\left(h_n^2+\sqrt{\frac{\log n}{nh_n^d}}\right)$ a.s. as $n\to\infty$. This makes the proof of Lemma A2 achieved. □

Lemma A3. *Under the same assumptions as in Lemma A2, for n large enough we have*

$$\sup_{\boldsymbol{x}\in\Omega}\sup_{\theta\in\Delta}\left|\widehat{\Psi}(\boldsymbol{x},\theta)-\tilde{\Psi}(\boldsymbol{x},\theta)\right|=O\left(\sqrt{\frac{\log\log n}{n}}\right),\quad \text{a.s. as } n\to\infty.$$

Proof.

$$\begin{aligned}
|\Xi_1| &= \frac{1}{nh_n^d}\left|\sum_{i=1}^{n}\mathbf{1}_{\{\delta_i=0\}}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\left(\frac{1}{S_n^R(Y_i)F_n^L(Y_i)}-\frac{1}{S^R(Y_i)F^L(Y_i)}\right)\psi_{\mathbf{x}}(Y_i-\theta)\right|\\
&= \frac{1}{nh_n^d}\left|\sum_{i=1}^{n}\mathbf{1}_{\{\delta_i=0\}}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\psi_{\mathbf{x}}(Y_i-\theta)\right.\\
&\quad\left.\times\left\{\frac{1}{F_n^L(Y_i)}\left[\frac{1}{S_n^R(Y_i)}-\frac{1}{S^R(Y_i)}\right]+\frac{1}{S^R(Y_i)}\left[\frac{1}{F_n^L(Y_i)}-\frac{1}{F^L(Y_i)}\right]\right\}\right|\\
&\leq \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)|\psi_{\mathbf{x}}(Y_i-\theta)|\left\{F^L(Y_i)\frac{|S^R(Y_i)-S_n^R(Y_i)|}{F_n^L(Y_i)S_n^R(Y_i)}\right.\\
&\quad\left.+\frac{|F_n^L(Y_i)-F^L(Y_i)|}{F_n^L(Y_i)}\right\}.
\end{aligned}$$

Hence, we have

$$\begin{aligned}\sup_{\mathbf{x}\in\Omega}\sup_{\theta\in\Delta}|\Xi_1| &\leq \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)|\psi_{\mathbf{x}}(Y_i-\theta)| \\ &\times \left\{\frac{\sup\limits_{a_T\leq t\leq b_T}|S_n^R(t)-S^R(t)|}{(F^L(a_T)-\sup\limits_{a_T\leq t\leq b_T}|F_n^L(t)-F^L(t)|)(S^R(b_T)-\sup\limits_{a_T\leq t\leq b_T}|S_n^R(t)-S^R(t)|)}\right. \\ &\left.+\frac{\sup\limits_{a_T\leq t\leq b_T}|F_n^L(t)-F^L(t)|}{F^L(a_T)-\sup\limits_{a_T\leq t\leq b_T}|F_n^L(t)-F^L(t)|}\right\} \\ &=: \Upsilon_1\times\Upsilon_2.\end{aligned}$$

In view of (6), (7) and condition (13) we get

$$\Upsilon_2 = O\left(\sqrt{\frac{\log\log n}{n}}\right), \text{ a.s.} \tag{A9}$$

Now, to deal with Υ_1 we use the following decomposition

$$\begin{aligned}\Upsilon_1 &= \frac{1}{nh_n^d}\sum_{i=1}^{n}\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)|\psi_{\mathbf{x}}(Y_i-\theta)| \\ &- \frac{1}{h_n^d}\mathbb{E}\left[\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)|\psi_{\mathbf{x}}(Y_i-\theta)|\right] \\ &+ \frac{1}{h_n^d}\mathbb{E}\left[\frac{\mathbf{1}_{\{\delta_i=0\}}}{S^R(Y_i)F^L(Y_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)|\psi_{\mathbf{x}}(Y_i-\theta)|\right] \\ &=: \Upsilon_{11}+\Upsilon_{12}.\end{aligned} \tag{A10}$$

Similar to the proof as in Lemma A2, we treat Υ_{11} and obtain

$$\sup_{\mathbf{x}\in\Omega}\sup_{\theta\in\Delta}\Upsilon_{11} = O\left(h_n^2+\sqrt{\frac{\log n}{nh_n^d}}\right) \text{ a.s.} \tag{A11}$$

On the other hand, using the same method as in (A2) under Assumptions **K(i)** and **P(ii)**, we get

$$\begin{aligned}\Upsilon_{12} &= \frac{1}{h_n^d}\int_{\mathbb{R}^d}K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)\mathbb{E}\left[|\psi_{\mathbf{x}}(T_1-\theta)||\mathbf{X}_1=\mathbf{u}\right].l(\mathbf{u})d\mathbf{u} \\ &\leq \int_{\mathbb{R}^d}K_d(\mathbf{z})\,\mathbb{E}\left[\psi_{\mathbf{x}}^2(T_1-\theta)|\mathbf{X}_1=\mathbf{x}-\mathbf{z}h_n\right]^{1/2}l(\mathbf{x}-\mathbf{z}h_n)d\mathbf{z} \\ &\leq c\int_{\mathbb{R}^d}K_d(\mathbf{z})d\mathbf{z} \\ &= O(1).\end{aligned} \tag{A12}$$

From (A10)-(A12), we conclude that $\sup\limits_{\mathbf{x}\in\Omega}\sup\limits_{\theta\in\Delta}\Upsilon_1 = O(1)$ a.s. which combined with (A9) achieves the proof of Lemma A3. □

***Proof of Proposition* 3.1** . The result is a straightforward consequence of (A1), Lemma A1, Lemma A2 and Lemma A3. □

***Proof of Theorem* 3.1** . Without loss of generality, we assume that $\psi_{\mathbf{x}}$ is an increasing function. This allows us to

assert that for all $\varsigma > 0$ and any $\mathbf{x} \in \Omega$, we have:

$$\Psi(\mathbf{x}, m(\mathbf{x}) + \varsigma) < 0 < \Psi(\mathbf{x}, m(\mathbf{x}) - \varsigma). \tag{A13}$$

Note that (A13) is obtained from $\Psi(\mathbf{x}, m(\mathbf{x})) = 0$. As $\widehat{\Psi}(\mathbf{x}, \widehat{m}(\mathbf{x})) = 0$ and $\widehat{\Psi}(\mathbf{x}, .)$ is continuous on Δ, it makes sense to write that

$$m(\mathbf{x}) - \varsigma < \widehat{m}(\mathbf{x}) < m(\mathbf{x}) + \varsigma \quad a.s. \ \ as \ \ n \to \infty.$$

On the other hand, a Taylor expansion of $\Psi(\mathbf{x}, .)$ in the neighborhood of $\widehat{m}(\mathbf{x})$ and the first part of Assumption **P(iii)** allows us to obtain

$$\widehat{\Psi}(\mathbf{x}, \widehat{m}(\mathbf{x})) - \Psi(\mathbf{x}, \widehat{m}(\mathbf{x})) = (m(\mathbf{x}) - \widehat{m}(\mathbf{x})) \times \frac{\partial \Psi(\mathbf{x}, m^*(\mathbf{x}))}{\partial \theta},$$

where $m^*(\mathbf{x})$ is between $m(\mathbf{x})$ and $\widehat{m}(\mathbf{x})$.
The purpose is to establish the strong consistency of $m(\mathbf{x}) - \widehat{m}(\mathbf{x})$ which comes straightforwardly from

$$\begin{aligned} \sup_{\mathbf{x}\in\Omega} |\widehat{m}(\mathbf{x}) - m(\mathbf{x})| &\leq \frac{1}{\inf_{\mathbf{x}\in\Omega} \left| \frac{\partial}{\partial \theta} \Psi(\mathbf{x}, m^*(\mathbf{x})) \right|} \sup_{\mathbf{x}\in\Omega} |\widehat{\Psi}(\mathbf{x}, \widehat{m}(\mathbf{x})) - \Psi(\mathbf{x}, \widehat{m}(\mathbf{x}))| \\ &\leq \frac{1}{c} \sup_{\mathbf{x}\in\Omega} \sup_{\theta\in\Delta} |\widehat{\Psi}(\mathbf{x}, \theta) - \Psi(\mathbf{x}, \theta)|, \end{aligned}$$

by the second part of Assumption **P(iii)**. Proposition 3.1 ends the proof.

□

References

Attaoui, S., Laksaci, A., and Ould-Saïd, E. (2015). Asymptotic results for an M-estimator of the regression function for quasi-associated processes. *Statistical Papers*, pages 3–28.

Benrabah, O., Bouhadjera, F., and Ould-Saïd, E. (2021). Local linear estimation of the regression function for twice-censored data. *Statistical Papers*, 63:489–514.

Benseradj, H. and Guessoum, Z. (2022). Strong uniform consistency rate of an M-estimator of regression function for incomplete data under α-mixing condition. *Communications in Statistics – Theory and Methods*, 51:2082–2115.

Benseradj, H. and Guessoum, Z. (2023). Asymptotic normality of an M-estimator of regression function for truncated-censored data under α-mixing condition. *Statistical Papers*.

Boente, G. and Fraiman, R. (1990). Asymptotic distribution of robust estimators for nonparametric models from mixing processes. *The Annals of Statistics*, 18:891–906.

Carbonez, A., Györfi, L., and van der Meulen, E. C. (1995). Partitioning-estimates of a regression function under random censoring. *Statistics & decisions*, 13:21–37.

Djelladj, W. and Tatachak, A. (2019). Strong uniform consistency of a kernel conditional quantile estimator for censored and associated data. *Hacettepe Journal of Mathematics and Statistics*, 48:290–311.

Ferraty, F. and Vieu, P. (2006). Nonparametric functional data analysis: Theory and practice. *Springer*.

Földes, A. and Rejtõ, L. (1981). A LIL type result for the product-limit estimator. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 56:75–86.

Gehan, E. A. (1965). A generalized two-sample wilcoxon test for doubly-censored data. *Biometrika*, 52:650–653.

Gheliem, A. and Guessoum, Z. (2022). Simulating the behavior of a kernel M-estimator for left-truncated and associated model. *Communications in Statistics - Simulation and Computation*, 53:2366–2388.

Hampel, F. R. (1973). Robust estimation: A condensed partial survey. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 27:87–104.

Hampel, F. R., Ronchetti, E., Rousseeuw, P. J., and Stahel, W. A. (1986). Robust statistics: The approach based on influence functions. *Wiley, New York*.

Härdle, W. K. (1984). Robust regression function estimation. *Journal of Multivariate Analysis*, 14:169–180.

Härdle, W. K. and Luckhaus, S. (1984). Uniform consistency of a class of regression function estimators. *Annals of Statistics*, 12:612–623.

Härdle, W. K. and Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Annals of Statistics*, 13:1465–1481.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:492–518.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Berkeley Symposium*, 5.1:221–233.

Huber, P. J. (1981). Robust statistics. *Wiley Series in Probability and Statistics*, pages 1248–1251.

Julià, O. and Gómez, G. (2011). Simultaneous marginal survival estimators when doubly censored data is present. *Lifetime Data Analysis*, 17:347–372.

Kebabi, K. and Messaci, F. (2012). Rate of the almost complete convergence of a kernel regression estimate with twice-censored data. *Statistics & Probability Letters*, 82:1908–1913.

Khardani, S. (2019). Relative error prediction for twice-censored data. *Mathematical Methods of Statistics*, 28:291–306.

Kitouni, A., Boukeloua, M., and Messaci, F. (2015). Rate of strong consistency for nonparametric estimators based on twice-censored data. *Statistics & Probability Letters*, 96:255–261.

Koul, H. L. and Stute, W. (1998). Regression model fitting with long memory errors. *Journal of Statistical Planning and Inference*, 71:35–56.

Laïb, N. and Ould-Saïd, E. (2000). A robust nonparametric estimation of the autoregression function under an ergodic hypothesis. *Canadian Journal of Statistics*, 28:817–828.

Leiderman, P. H., Babu, B. A. P., Kagia, J., Kraemer, H. C., and Leiderman, G. F. (1973). African infant precocity and some social influences during the first year. *Nature*, 242:247–249.

Lemdani, M. and Ould-Saïd, E. (2017). Nonparametric robust regression estimation for censored-data. *Statistical Papers*, 58:505–525.

Menni, N. and Tatachak, A. (2016). A note on estimating the conditional expectation under censoring and association: Strong uniform consistency. *Statistical Papers*, 59:1009–1030.

Messaci, F. (2010). Local averaging estimates of the regression function with twice-censored-data. *Statistics & Probability Letters*, 80:1508–1511.

Messaci, F. and Nemouchi, N. (2013). Erratum to a law of the iterated logarithm for the product-limit estimator with doubly censored data. *Statistics & Probability Letters*, 83:2142.

Morales, D., Pardo, L., and Quesada, V. (1991). Bayesian survival estimation for incomplete data when the life distribution is proportionally related to the censoring time. *Communications in Statistics – Theory and Methods*, 20:831–850.

Patilea, V. and Rolin, J. M. (2006). Product-limit estimators of the survival function with twice-censored data. *Annals of Statistics*, 34:925–938.

Peto, R. (1973). Experimental survival curves for interval-censored data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 22:86–91.

Ren, J. J. and Gu, M. (1997). Regression M-estimators with doubly censored data. *Annals of Statistics*, 25:2638–2664.

Serfling, R. (1980). Approximation theorems of mathematical statistics. *Wiley, New York*.

Tsybakov, A. B. (1982). Robust estimates of a function. *Probability Peredachi Inference*, 18:39–52.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship sunction with doubly censored data. *Journal of the American Statistical Association*, 69:169–173.

Wang, J. F. and Liang, H. Y. (2012). Asymptotic properties for an M-estimator of the regression function with truncation and dependent data. *Journal of The Korean Statistical Society*, 41:351–367.