Austrian Journal of Statistics
2026, Volume 55, 153–186.
http://www.ajs.or.at/
doi:10.17713/ajs.v55i2.2303

Asymptotic Properties of the Location-Scale Regression Estimators for Left Truncated Data

Saloua Aliouche
University of Science and Technology
Houari Boumediene
Algeria

Ourida Sadki
University of Science and Technology
Houari Boumediene
Algeria

Hassiba Benseradj
University M'hamed Bougara of Boumerdes
Algeria

Abstract

We consider the nonparametric estimation of a heteroscedastic regression model $Y = m(X) + \sigma(X)\epsilon$, where the error term ϵ is independent of the covariate X, and the functions m and σ are unknown. The response variable Y is allowed to be interfered by a left truncation random variable T. In this work, we propose kernel-estimators for both the location and scale functions m and σ and investigate their strong uniform consistency rates as well as their asymptotic normality. The finite sample performance of the proposed estimator of the regression function $m(\cdot)$ is extensively studied via simulation and indicate that in presence of heteroscedasticity, our new estimator outperforms the classical one (when ignoring heteroscedasticity) and is less sensitive to the presence of outliers. We also investigate the performance of the scale function estimator $\sigma(\cdot)$. Furthermore, an application to real data demonstrates the effectiveness of our method in terms of prediction accuracy.

Keywords: asymptotic normality; asymptotic representation; local scale regression; random left-truncation; strong uniform consistency.

1. Introduction

Oftentimes, regression analysis is carried out on data that may exhibit high variance across different values of the independent variables. One of the effects of this type of data is heteroscedasticity which indicates variable variance around the fitted values. Heteroscedasticity is a common phenomenon encountered in many fields, including genetics (Daye, Chen, and Li (2012)), toxicology (Lim, Sen, and Peddada (2010)) and experiment designs (Box and Meyer (1985) and Box and Draper (1987)), among others. It is also frequently observed in finance and time series data, particularly in situations where the variables change drastically over time. For example Kuznietsova and Bidyuk (2020) show from experiment studies that heteroscedastic models are effective for predicting the variance of financial processes, as they account for the rapid changes and volatility inherent in such data. This prevalence and impact of heteroscedasticity motivates the study of the location scale model

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

where Y is the survival time, X is the associated covariate, ϵ is an unobservable error term that is independent of X. $m(\cdot)$ and $\sigma(\cdot)$ are unknown regression (location) and heteroscedasticity (scale) functions, respectively. This model has been extensively studied in statistics and econometrics literature, see, e.g., Härdle, Janssen, and Serfling (1988) for complete data. For censored data, Fan and Gijbels (1994) studied the estimation of $m(\cdot)$ using local linear regression techniques. Van Keilegom and Akritas (1999) consider L-functional estimators of $m(\cdot)$ and $\sigma(\cdot)$ and studied their asymptotic properties, their results were later extended by Sujika and Van Keilegom (2018) to copula dependent data. Furthermore, Heuchenne and Van Keilegom (2010) considered a more general class of L-estimators type under independent data. More recently Hössjer and Karlsson (2023) provide a survey unifying a broad class of conditional L-functionals, covering measures of location, scale, skewness, and heavy tails, within a nonparametric regression framework.

Instead of censoring, another form of incomplete data, often encountered in practice, is that data is sometimes truncated. More precisely, Y is subject to left truncation by a variable T, so that (X, Y) are observable only if $Y \geq T$.

The random left truncation (RLT) model has many applications, which originally appeared in astronomy and economics (see Woodroofe (1985)) and has since been extended to fields such as medicine, epidemiology, actuarial and beyond. Truncation is not just a theoretical problem, in practice it can cause substantial bias if ignored. An important example of such model arises in the analysis of survival data of patients infected by the AIDS virus from contaminated blood transfusions (see Chen, Chao, and Lo (1995)).

Ould-Saïd and Lemdani (2006) consider nonparametric truncated regression model with homoscedasticity, i.e $\sigma(X)$ is assumed to be constant. They proposed a kernel estimator for $m(\cdot)$ and established its uniform consistency and asymptotic normality under independent data. The same estimator is largely studied for dependent data; see, e.g., Liang, Li, and Qi (2009), Guessoum and Hamrani (2017), among others. For robust regression, notable contributions include Wang and Liang (2012) and Gheliem and Guessoum (2022). Benseradj and Guessoum (2020) considered robust regression under a right-censored and left-truncated (LTRC) model. Lemdani, Ould-Saïd, and Poulin (2009) focused on the estimation of conditional quantiles, which also provides a valuable alternative to classical regression methods.

Chen, Lu, Zhou, and Zhou (2018) consider the joint estimation of $m(\cdot)$ and $\sigma(\cdot)$ under a latent truncation heteroscedastic model. In their framework, the observed data (Y, X) arise from the conditional distribution of a latent variable Y^* given $\{Y^* > 0\}$, so that the truncation mechanism is unobservable. In contrast, the present paper considers truncation by an observed threshold, where (Y, X) are observed only when $Y > T$, with T known. Therefore, our approach for estimating $m(\cdot)$ and $\sigma(\cdot)$ using L-estimators provides a complementary contribution to the existing literature on truncated data. Specifically, it adapts and extends the results of Van Keilegom and Akritas (1999) from censored to truncated data, and generalizes the results of Ould-Saïd and Lemdani (2006) from the homoscedastic to the heteroscedastic case, and establishes uniform convergence as well as asymptotic normality. However, to the best of our knowledge, this procedure has not yet been investigated in the truncation setting.

The paper is organized as follows. In Section 2, we give some definitions and notations related to our model and define the estimators of $m(\cdot)$ and $\sigma(\cdot)$. Section 3 presents the assumptions under which we state our main results. A large simulation study is provided in Section 4, to illustrate the performance of the proposed estimators for both the regression and scale functions. Section 5 is devoted to the analysis of a real dataset, in order to validate the practical effectiveness of our estimator. Proofs and auxiliary results are relegated to Section 6. Finally, the conclusion and discussion are presented in Section 7.

2. Model and estimators

Let $\{(X_i, Y_i), 1 \le i \le N\}$ be N independent and identically distributed (iid) copies of (X, Y) defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Y denotes the lifetime under study, with a continuous distribution function (df) F. X is a one dimensional covariate with probability density v, and we denote by $F(\cdot|x)$ the conditional df of Y given $X = x$. Here, the sample size N is deterministic but unknown.

Let $(T_i)_{i=1,\dots,N}$ be a sample of truncation random variables (rv) with continuous df G. Then (Y, T) is observed only if $(Y \ge T)$; otherwise, no observation is recorded if $Y < T$. We assume that T and (X, Y) are independent. Clearly the initial sample is not completely observed, we observe only n values (among N). Obviously, in the context of truncated data, our estimators are constructed using the observed rv $(X_i, Y_i, T_i)_{i=1,..,n}$. These observed rv's remain iid (see proposition 2.1 in Lemdani and Ould-Saïd (2007)).

As a consequence of truncation, the size of the actually observed sample n is random but known. In the sequel, we denote $\alpha := \mathbb{P}(T \le Y)$ as the probability of observing Y. Clearly if $\alpha = 0$ no data can be observed. Therefore, throughout this paper, we assume that $\alpha > 0$.
In what follows, the star notation $(\star)$ refers to any characteristic of the actually observed data (conditionally on n). Moreover, throughout this study, all probability statements are to be interpreted as conditional probability statements, that is $\mathbf{P}(\cdot) = \mathbb{P}(\cdot|Y \ge T)$. Similarly $\mathbf{E}$ and $\mathbb{E}$ denote the expectation operators associated with $\mathbf{P}$ and $\mathbb{P}$, respectively. Conditional on n, all estimation results are stated by letting $n \to \infty$, which hold true with respect to the probability $\mathbb{P}$, since $n \le N$.

Following Stute (1993), the conditional joint df of an observed pair (Y, T) is given by

$$M^\star(y, t) := \mathbf{P}(Y \le y, T \le t) = \alpha^{-1} \int_0^y G(t \wedge z) dF(z),$$

where $t \wedge z := \min(t, z)$. The marginal distribution functions of Y and T are defined respectively as

$$F^\star(y) := \mathbf{P}(Y \le y) = M^\star(y, \infty) = \alpha^{-1} \int_0^y G(z) dF(z),$$
$$G^\star(t) := \mathbf{P}(T \le t) = M^\star(\infty, t) = \alpha^{-1} \int_0^{+\infty} G(t \wedge z) dF(z),$$

which are estimated respectively by

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

where $\mathbb{1}_A$ denotes the indicator function of the set A.

The well-known product-limit estimators of F and G, proposed by Lynden-Bell (1971) are

$$F_n(y) = 1 - \prod_{i: Y_i \le y} \left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)} \right],$$

and

$$G_n(t) = \prod_{i: T_i > t} \left[\frac{nC_n(T_i) - 1}{nC_n(T_i)} \right], \tag{2}$$

where

$$C_n(y) := \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{T_i \le y \le Y_i\}} = G_n^\star(y) - F_n^\star(y),$$

is the empirical estimator of

$$C(y) := \mathbf{P}(T \le y \le Y) = \frac{1}{\alpha} G(y)(1 - F(y)). \tag{3}$$

Notice that, $\hat{\alpha}_n := \frac{n}{N}$ is a consistent estimator of α, but it cannot be calculated, since N is unknown. Based on (3), He and Yang (1998) proposed the estimator

$$\alpha_n := \frac{G_n(y)(1-F_n(y))}{C_n(y)}, \tag{4}$$

for all y such that $C_n(y) \neq 0$, and they demonstrate that α_n does not depend on y.

Here, for any df W, let us define $a_W = \inf\{u : W(u) > 0\}$ and $b_W = \sup\{u : W(u) < 1\}$, as the endpoints of the W support. As pointed out in Woodroofe (1985), the df's F and G can be completely estimated only if $a_G \leq a_F$, $b_G \leq b_F$ and $\int_{a_F}^{\infty} G^{-1} dF < \infty$. Throughout this paper, we assume that $0 = a_G < a_F$ and $b_G \leq b_F$.

Define also the trivariate conditional df of (X, Y, T) as

$$H^{\star}(x,y,t) := \mathbf{P}(X \leq x, Y \leq y, T \leq t) = \alpha^{-1} \int_{a_F}^{y} \int_{-\infty}^{x} G(t \wedge w) f(z,w) dz dw,$$

where $f(\cdot,\cdot)$ denotes the joint density function of (X,Y). Taking $t = +\infty$, the observed pair (X,Y) then has the conditional df

$$F^{\star}(x,y) := H^{\star}(x,y,+\infty) = \alpha^{-1} \int_{a_F}^{y} \int_{-\infty}^{x} G(w) f(z,w) dz dw. \tag{5}$$

By differentiating (5), we obtain

$$dF^{\star}(x,y) = \alpha^{-1} G(y) f(x,y) dx dy, \quad \text{for any} \quad y > a_F. \tag{6}$$

We now return to our main problem, which is the estimation of the location and scale regression functions m and σ under RLT model. To this end, we consider the following location and scale functionals

$$m(x) = \int_0^1 F^{-1}(s|x) J(s) ds, \quad \sigma^2(x) = \int_0^1 F^{-1}(s|x)^2 J(s) ds - m^2(x), \tag{7}$$

where $F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ denotes the conditional quantile function of Y given $X = x$. The function $J(s)$ is a positive score function, satisfying $\int_0^1 J(s) ds = 1$. This class of location and scale functionals, known as L-functionals, is very flexible and covers a broad range of commonly used functionals, Serfling (1980) provides a comprehensive discussion of these functionals and their properties. For example, the choice of $J(\cdot) \equiv 1$, $m(x)$ and $\sigma^2(x)$ reduce to the conditional mean and variance, respectively. Hence, working with the definitions in (7), does not entail any loss of generality.

Note that the functionals defined in (7) depend on the conditional df $F(\cdot|x)$, which must be estimated. To this end, let $F(\cdot,\cdot)$ denote the joint distribution function of (X,Y) and suppose that it is continuously differentiable with respect to the first component. Set $F_1(x,\cdot) = \frac{\partial F(x,\cdot)}{\partial x}$, hence the conditional df of Y given $X = x$ can be written as $F(\cdot|x) = \dfrac{F_1(x,\cdot)}{v(x)}$, for $v(x) > 0$.

Let $\hat{F}_1(\cdot,\cdot)$ and $v_n(\cdot)$ denote the kernel estimators of $F_1(\cdot,\cdot)$ and $v(\cdot)$, respectively. Lemdani *et al.* (2009) defined a Nadaraya Watson kernel type estimator of $F(\cdot|x)$

$$\hat{F}(y|x) := \frac{\hat{F}_1(x,y)}{v_n(x)} \mathbb{1}\{v_n(x) \neq 0\}, \tag{8}$$

where

$$\hat{F}_1(x,y) := \frac{\alpha_n}{n h_n} \sum_{i=1}^{n} G_n^{-1}(Y_i) k\left(\frac{x - X_i}{h_n}\right) K_0\left(\frac{y - Y_i}{h_n}\right),$$

$$v_n(x) := \alpha_n n h_n \sum_{i=1}^{n} G_n^{-1}(Y_i) k\left(\frac{x - X_i}{h_n}\right),$$

for all Y_i such that $G_n(Y_i) \neq 0$. $k(\cdot)$ is a kernel function, $K_0(\cdot)$ is a df and h_n is a positive bandwidth toward 0 when $n \to \infty$. This motivates the introduction of the following estimators of $m(x)$ and $\sigma^2(x)$

$$\hat{m}(x) = \int_0^1 \hat{F}^{-1}(s|x)J(s)ds, \tag{9}$$

and

$$\hat{\sigma}^2(x) = \int_0^1 \hat{F}^{-1}(s|x)^2 J(s)ds - \hat{m}^2(x). \tag{10}$$

Note that these type of estimators are studied under censoring data by Van Keilegom and Akritas (1999), who worked under the assumption of conditional independence between Y and C given X, where C is the variable of right censoring and estimated $F(\cdot|x)$ using the estimator proposed by Beran (1981) instead of the estimator defined in (8). Sujika and Van Keilegom (2018) consider the situation where Y and C are dependent given X and estimated $F(\cdot|x)$ by the so-called conditional Copula-graphic estimator. While Chown, Heuchenne, and Van Keilegom (2020) focused on mixture cure models with censored data.

3. Assumptions and main results

Before presenting our results, we first state some regular conditions. We consider two real numbers a_x and b_x such that $a_{F(\cdot|x)} \leq a_x < b_x \leq b_{F(\cdot|x)}$. Define $\Omega_0 = \{x \in \mathbb{R}|v(x) > 0, \sigma(x) > 0\}$ and let Ω be a compact subset of Ω_0, then $\inf_{x\in\Omega} v(x) > 0$ and $\inf_{x\in\Omega} \sigma(x) > 0$. The assumptions needed for our results are listed as follows.

(A1) The bandwidth h_n satisfies $h_n \to 0$ as $n \to \infty$, and
(i) $\frac{nh_n}{\log n} \to \infty$ as $n \to \infty$.
For Asymptotic normality, we need the two following additional Assumptions.
(ii) $\frac{\sqrt{nh_n}}{\log n} \longrightarrow \infty$ as $n \to \infty$.
(iii) $nh_n^5 \longrightarrow 0$ as $n \to \infty$.

(A2) (i) k is a $\mathcal{C}^1$-probability density with compact support and satisfies $\int rk(r)dr = 0$.
(ii) K_0 admits a $\mathcal{C}^1$-probability density k_0 and compact support, k_0 satisfies $\int tk_0(t)dt = 0$.

(A3) The joint density $f(\cdot,\cdot)$ is bounded and twice continuously differentiable.

(M1) There exist $0 \leq s_0 \leq s_1 \leq 1$ such that $\sup_{x\in\Omega} F(a_x|x) \leq s_0 \leq \inf\{s \in [0,1]; J(s) \neq 0\}$, $\sup\{s \in [0,1]; J(s) \neq 0\} \leq s_1 \leq \inf_{x\in\Omega} F(b_x|x)$.

(M2) $\inf_{x\in\Omega} \inf_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x) > 0$, where s_0 and s_1 are as defined in (M1) and $f(.|x)$ is the conditional density of Y given $X = x$.

(M3) The function J is continuously differentiable on the interval (s_0, s_1), verifies $\int_0^1 J(s)ds = 1$ and $J(s) \geq 0$ for all $0 \leq s \leq 1$.

Remark 1. *1. Assumption (A1)(i) ensures the consistency of $\hat{m}$ and $\hat{\sigma}$, whereas Assumptions (A1)(ii) and (A1)(iii) are required for establishing Asymptotic normality; more precisely, they are used to make a bias term negligible. Note that Assumption (A1) is satisfied by choosing the bandwidth $h_n = O\left((logn/n)^{1/5}\right)$, or $h_n = O(n^{-\frac{\delta}{5}})$, for any $1 < \delta < 5$.*

2. *Assumptions (A2) and (A3) are the same given in Lemdani et al. (2009) and are needed to use their results.*

3. *Assumptions (M1), (M2) and (M3) are classical for L-estimators (see, e.g., Van Keilegom and Akritas (1999), Sujika and Van Keilegom (2018)), and are mainly used to ensure that the estimator is driven by a well-behaved principal part of the conditional distribution. Condition (M1) localizes the effective support of the score function, avoiding boundary effects. Condition (M2) guarantees the non-degeneracy of the conditional quantiles involved, while (M3) ensures sufficient smoothness of the score function.*

3.1. Classical examples of L-functionals

We present below several classical examples of L-functionals and illustrate how each satisfies Assumptions (M1)-(M3).

1. **Conditional mean**
   The conditional mean can be written as an L-functional of the form

$$\mathbb{E}(Y|X=x) = \int_0^1 F^{-1}(s|x)J(s)ds,$$

   where $J(s) \equiv 1$ for $s \in [0,1]$.

   - The score function satisfies $J(s) \equiv 1$ on $[0,1]$, so that its support coincides with the whole interval $[0,1]$. Consequently, Assumption (M1), which is designed to keep the integration domain away from the boundaries of the conditional support, becomes vacuous and does not impose any effective restriction in this setting.
   - Assumption (M2) reduces to the standard requirement that the conditional density $f(\cdot|x)$ be strictly positive on its support.
   - Assumption (M3) is trivial.

2. **Conditional trimmed mean**
   For $0 \le \tau < \frac{1}{2}$, the conditional trimmed mean, $TM_\tau(\cdot|x)$, can be expressed as an L-functional in the form

$$TM_\tau(Y|X=x) = \int_0^1 F^{-1}(s|x)J_\tau(s)ds,$$

   where $J_\tau(s) = \frac{1}{1-2\tau}\mathbb{1}_{[\tau,1-\tau]}(s)$.

   - Assumption (M1) is satisfied with $s_0 = \tau$ and $s_1 = 1-\tau$, and reduce (in terms of quantiles) to

$$a_x \le F^{-1}(\tau|x) \text{ and } F^{-1}(1-\tau|x) \le b_x.$$

     Hence (M1) reduces to verifying that the extreme quantile (τ and $1-\tau$) lies within the conditional support of Y, for all x.
   - Assumption (M2) requires $\inf\limits_{x\in\Omega}\inf\limits_{s\in[\tau,1-\tau]} f(F^{-1}(s|x)|x) > 0$.
   - The score function J_τ is smooth on $(\tau, 1-\tau)$ satisfies $\int_0^1 J_\tau(s)ds = 1$ and $J_\tau(s) \ge 0$, for all $s \in [0,1]$, which implies that Assumption (M3) is satisfied.

   The special case $\tau = 0$ yields $J_\tau(s) \equiv 1$ and recovers the conditional mean.

3. **Conditional median**
   The conditional median can be defined as

$$Med(Y|X=x) = \int_0^1 F^{-1}(s|x)J(s)ds,$$

where $J(s) = \delta_{\frac{1}{2}}(s)$, a Dirac mass at $\frac{1}{2}$.
It is clearly seen that the score function $\delta_{\frac{1}{2}}(s)$ is not a regular function, and therefore the conditional median is not an L-functional in the strict sense under (M3).
However, we can consider a sequence of smooth score function J_ϵ, $0 < \epsilon < \frac{1}{2}$, defined by

$$J_\epsilon(s) = \frac{1}{2\epsilon}\mathbb{1}_{[\frac{1}{2}-\epsilon,\frac{1}{2}+\epsilon]}(s),\ \ s \in [0,1].$$

The associated L-functional $Med_\epsilon(\cdot|x)$ is

$$Med_\epsilon(Y|X = x) = \frac{1}{2\epsilon}\int_{\frac{1}{2}-\epsilon}^{\frac{1}{2}+\epsilon} F^{-1}(s|x)ds,$$

which clearly converges to the conditional median $Med(Y|X = x)$ as $\epsilon \to 0$. Moreover

- Assumption (M1) is satisfied with $s_0 = \frac{1}{2} - \epsilon$, $s_1 = \frac{1}{2} + \epsilon$, $0 < s_0 < s_1 < 1$.
- Assumption (M2) requires $\inf_{x\in\Omega} \inf_{s\in[\frac{1}{2}-\epsilon,\frac{1}{2}+\epsilon]} f(F^{-1}(s|x)|x) > 0$.
  This condition means that the conditional density is strictly positive in a neighborhood of the conditional median $F^{-1}\left(\frac{1}{2}|x\right)$.
- The function J_ϵ satisfies the regularity requirements of (M3), additionally $J_\epsilon \geq 0$, for all $s \in [0,1]$, and $\int_0^1 J_\epsilon(s) = 1$.

3.2. Strong consistency and almost sure representation

In this subsection, we derive the uniform consistency rates of the location estimator $\hat{m}(x)$ and their analogues for $\hat{\sigma}(x)$ (Theorem 1). We also provide an iid asymptotic representation for $\hat{F}(y|x) - F(y|x)$, and establish the rate of convergence of the remainder term uniformly in x and y (Proposition 1). This result is instrumental for proving analogous results for the estimators of $m(x)$ and $\sigma(x)$ in Theorems 2 and 3, respectively.

Theorem 1. *Suppose that Assumptions (A1)(i), (A2), (A3) and (M1)-(M3) hold, then we have*

$$(a)\quad \sup_{x\in\Omega}|\hat{m}(x) - m(x)| = O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right),\ \mathbf{P} - a.s\ as\ n \to \infty.$$

$$(b)\quad \sup_{x\in\Omega}|\hat{\sigma}(x) - \sigma(x)| = O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right),\ \mathbf{P} - a.s\ as\ n \to \infty.$$

Remark 2. *The convergence rates of $\hat{m}(x)$ and $\hat{\sigma}(x)$ stated in Theorem 1, coincide with those obtained by Van Keilegom and Akritas (1999) (Proposition 4.5) in the context of censored data. It is worth noting that, in the complete data setting, Akritas and Van Keilegom (2001), in their Proposition 3, derived a faster convergence rates (under Assumption (A1)(i)) than in the censored case, namely of order $(nh_n)^{-1/2}(log(1/h_n))^{1/2}$. Moreover, our rate for $\hat{m}(x)$ improves upon the results of Ould-Saïd and Lemdani (2006) for truncated homoscedastic data, see Remark 4 in Section 3.4 for further details.*

To drive the results of Proposition 1, Theorems 2 and 3, we first introduce the following functions.

$$\begin{aligned}\xi(X,Y,x,y) :=& G^{-1}(Y)k\left(\frac{x-X}{h_n}\right)K_0\left(\frac{y-Y}{h_n}\right) - \mathbf{E}\left(G^{-1}(Y)k\left(\frac{x-X}{h_n}\right)K_0\left(\frac{y-Y}{h_n}\right)\right)\\ &- F(y|x)\left\{G^{-1}(Y)k\left(\frac{x-X}{h_n}\right) - \mathbf{E}\left(G^{-1}(Y)k\left(\frac{x-X}{h_n}\right)\right)\right\}\\ =:& \xi_1(X,Y,x,y) - F(y|x)\xi_2(X,Y,x).\end{aligned}$$

$$\eta(X,Y|x) := \int_0^{+\infty}\xi(X,Y,x,y)J(F(y|x))dy.$$

$$\zeta(X,Y|x) := \int_0^{+\infty}\xi(X,Y,x,y)J(F(y|x))\frac{y-m(x)}{\sigma(x)}dy.$$

Proposition 1. *Assume (A1)(i), (A2) and (A3), we have*

$$\hat{F}(y|x) - F(y|x) = (nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\xi(X_i, Y_i, x, y) + R_n(x,y),$$

for $x \in \Omega$, $y \in [a_x, b_x]$, *where*

$$\sup_{x\in\Omega}\sup_{a_x\le y\le b_x}|R_n(x,y)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right),\ \mathbf{P} - a.s\ as\ n\to\infty.$$

Theorem 2. *Under Assumptions (A1)(i), (A2), (A3) and (M1)-(M3), we have for any* $x \in \Omega$

$$\hat{m}(x) - m(x) = -(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\eta(X_i, Y_i|x) + R_n(x),$$

where

$$\sup_{x\in\Omega}|R_n(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right),\ \mathbf{P} - a.s\ as\ n\to\infty. \tag{11}$$

Theorem 3. *Under Assumptions (A1)(i), (A2), (A3) and (M1)-(M3), we have for any* $x \in \Omega$

$$\hat{\sigma}(x) - \sigma(x) = -(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\zeta(X_i, Y_i|x) + \tilde{R}_n(x),$$

where

$$\sup_{x\in\Omega}|\tilde{R}_n(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right),\ \mathbf{P} - a.s\ as\ n\to\infty.$$

Using the results of Theorem 2 and Theorem 3, we establish the asymptotic normality of the estimators $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$.

3.3. Asymptotic normality

Theorem 4 and Theorem 5 state the asymptotic normality of the estimators $\hat{m}(x)$ and $\hat{\sigma}(x)$, respectively, which will be used to construct their confidence intervals.

Theorem 4. *Suppose that Assumptions (A1)-(A3) and (M1)-(M3) hold, then we have for any* $x \in \Omega$ *and for* n *large enough*

$$\sqrt{nh_n}(\hat{m}(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Delta^2(x)\right),$$

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

$$\Delta^2(x) = \frac{\alpha}{v^2(x)}\kappa\int_{a_F}^{+\infty}\left[\int_0^t J(F(y|x))dy - \int_0^\infty \overline{F}(y|x)J(F(y|x))dy\right]^2\frac{f(x,t)}{G(t)}dt, \tag{12}$$

$\kappa = \int k^2(u)du$ *and* $\overline{F}(\cdot) = 1 - F(\cdot)$.

Theorem 5. *Suppose that Assumptions (A1)-(A3) and (M1)-(M3) hold, then we have for any* $x \in \Omega$ *and for* n *large enough*

$$\sqrt{nh_n}(\hat{\sigma}(x) - \sigma(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Delta_1^2(x)\right),$$

where

$$\Delta_1^2(x) = \frac{\alpha}{v^2(x)}\kappa\int_{a_F}^{+\infty}\left[\int_0^t J(F(y|x))\frac{y-m(x)}{\sigma(x)}dy - \int_0^{+\infty}\overline{F}(y|x)J(F(y|x))\frac{y-m(x)}{\sigma(x)}dy\right]^2\frac{f(x,t)}{G(t)}dt.$$

Remark 3. *Plug-in estimates $\hat{\Delta}^2(x)$ and $\hat{\Delta}_1^2(x)$ for the asymptotic variances $\Delta^2(x)$ and $\Delta_1^2(x)$, can easily be obtained by using $G_n(\cdot)$, α_n, $v_n(\cdot)$, $\hat{F}(\cdot|\cdot)$, $\hat{m}(\cdot)$ and $\hat{\sigma}(\cdot)$ given in (2), (4), (8), (9) and (10), respectively. Hence, we get*

$$\hat{\Delta}^2(x) = \frac{\alpha_n^2}{nh_n}\frac{\kappa}{v_n^2(x)}\sum_{i=1}^n \frac{k\left(\frac{x-X_i}{h_n}\right)}{G_n^2(Y_i)} \left[\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{[0,Y_i]}(Y_j)\frac{J(\hat{F}(Y_j|x))}{\hat{f}(Y_j)} - \frac{1}{n}\sum_{j=1}^n \hat{\overline{F}}(Y_j|x)\frac{J(\hat{F}(Y_j|x))}{\hat{f}(Y_j)}\right]^2,$$

$$\hat{\Delta}_1^2(x) = \frac{\alpha_n^2}{nh_n}\frac{\kappa}{v_n^2(x)}\sum_{i=1}^n \frac{k\left(\frac{x-X_i}{h_n}\right)}{G_n^2(Y_i)} \left[\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{[0,Y_i]}(Y_j)\frac{Y_j-\hat{m}(x)}{\hat{\sigma}(x)}\frac{J(\hat{F}(Y_j|x))}{\hat{f}(Y_j)} - \right.$$
$$\left. -\frac{1}{n}\sum_{j=1}^n \hat{\overline{F}}(Y_j|x)\frac{Y_j-\hat{m}(x)}{\hat{\sigma}(x)}\frac{J(\hat{F}(Y_j|x))}{\hat{f}(Y_j)}\right]^2,$$

where $\hat{\overline{F}}(y|x) = 1 - \hat{F}(y|x)$, and

$$\hat{f}(y) := \frac{\alpha_n}{nh_n}\sum_{i=1}^n G_n^{-1}(Y_i)k_0\left(\frac{y-Y_i}{h_n}\right).$$

As an application of the asymptotic normality, we derive the confidence interval given in the following corollary.

Corollary 3.3.1. *Under Assumptions of Theorem 4, we obtain for all $\beta \in (0,1)$, a confidence interval of asymptotic level $1-\beta$ for $m(x)$, and $\sigma(x)$, respectively.*

$$(a) \quad \left[\hat{m}(x) - \frac{u_{(1-\beta/2)}\hat{\Delta}(x)}{\sqrt{nh_n}}, \hat{m}(x) + \frac{u_{(1-\beta/2)}\hat{\Delta}(x)}{\sqrt{nh_n}}\right],$$
$$(b) \quad \left[\hat{\sigma}(x) - \frac{u_{(1-\beta/2)}\hat{\Delta}_1(x)}{\sqrt{nh_n}}, \hat{\sigma}(x) + \frac{u_{(1-\beta/2)}\hat{\Delta}_1(x)}{\sqrt{nh_n}}\right],$$

where $u_{(1-\beta/2)}$ denotes the $(1-\beta/2)$ quantile of the standard normal distribution.

3.4. Classical regression as a special case

As a direct consequence of Theorem 1 and Theorem 4, we get the strong uniform consistency rate and the asymptotic normality of the estimator of the regression curve $\mathbb{E}(Y|X=x)$, stated as particular case of the estimator defined in (9) with the score function $J(s) \equiv 1$, given by

$$m_{NW}(x) = \frac{\sum_{i=1}^n Y_i G_n^{-1}(Y_i)k\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n G_n^{-1}(Y_i)k\left(\frac{x-X_i}{h_n}\right)}. \tag{13}$$

This is the estimator proposed by Ould-Saïd and Lemdani (2006), which yield to the following results.

Corollary 3.4.1. *Under the Assumptions of Theorem 1, we have*

$$\sup_{x\in\Omega}|m_{NW}(x) - m(x)| = O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right), \mathbf{P} - a.s \text{ as } n \to \infty.$$

Remark 4. *Note that the convergence rate obtained here improves upon that established by Ould-Saïd and Lemdani (2006), which is of order $O\left(\max\left\{(n^{1-2\gamma}h_n^2)^{-1/2}(\log n)^{1/2}, h_n\right\}\right)$, $\mathbf{P}-$ a.s as $n \to \infty$, for $\gamma \in [0, \frac{3}{10})$. Indeed*

$$\frac{(nh_n)^{-1/2}(\log n)^{1/2}}{(n^{1-2\gamma}h_n^2)^{-1/2}(\log n)^{1/2}} = n^{-\gamma}h_n^{1/2}.$$

Since $\gamma > 0$ and $h_n \to 0$ as $n \to \infty$, it follows that $n^{-\gamma}h_n^{1/2} \to 0$, This shows that our rate is faster than that obtained by Ould-Saïd and Lemdani (2006).

The difference arise from the fact that the Nadaraya Watson estimator m_{NW} depends linearly on Y. Since Y is unbounded, their method requires a truncation procedure, bounding Y by a sequence n^γ. In contrast, L-estimators depend on Y only through estimator of the conditional distribution function $\widehat{F}(\cdot|x)$. Therefore, no truncation is needed, which leads to an improved convergence rate.

Corollary 3.4.2. *Under the Assumptions of Theorem 4, we have for any $x \in \Omega$,*

$$\sqrt{nh_n}(m_{NW}(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Gamma^2(x)\right),$$

where

$$\Gamma^2(x) = \frac{\alpha}{v^2(x)}\kappa \int_{a_F}^{+\infty} (t - m(x))^2 \frac{f(x,t)}{G(t)} dt.$$

Note that $\Gamma^2(\cdot)$ is the asymptotic variance which should be obtained by Ould-Saïd and Lemdani (2006) in their Theorem 4.2, but there is a typo in their $\Gamma^2(\cdot)$ as discussed in Liang *et al.* (2009) (see their Remark 2.3).

4. Simulation study

This section is divided into two parts. In the first part, we illustrate the performance of our estimator $\hat{m}(x)$ through Monte Carlo simulations and compare it with the Nadaraya-Watson type estimator $m_{NW}(x)$, defined in (13). These comparisons are conducted for various combinations of sample sizes n and no-truncation rates α, under both homoscedastic and heteroscedastic regression settings. We expect that, under the heteroscedastic model (1), the estimator $\hat{m}(x)$ outperforms $m_{NW}(x)$. Whereas under the homoscedastic model, both estimators exhibit similar performance.

The second part is devoted to evaluating the performance of the estimated scaling parameter $\hat{\sigma}(x)$ for different sample sizes and truncation rates.

4.1. Performance of the regression function estimator

1. **Homoscedastic regression case**

   We generate the observed data (X_i, Y_i, T_i), $i = 1, ..., n$ and calculate the estimators $\hat{m}(x)$ and $m_{NW}(x)$, following the procedure:

   Step 1. Generate iid rv's of the covariate X_i, $i \geq 1$ and the error ϵ_i independently from the Normal distribution $\mathcal{N}(0, 1)$.

   Step 2. Simulate the rv $Y_i, i \geq 1$, according to the homoscedastic model

   $$Y_i = m(X_i) + \sigma\epsilon_i.$$

   We take two regression functions $m(\cdot)$, linear and nonlinear, and $\sigma = 0.2$.

   Step 3. Simulate T_i, $i \geq 1$, iid rv from an exponential distribution with parameter λ, where λ is adapted in order to obtain different values of the no-truncation rate α.

   Step 4. We test

   $$\begin{cases} \text{if} \quad Y_i \geq T_i & \text{keep the observation} \quad (X_i, Y_i, T_i) \\ \text{else} & \text{reject} \quad (X_i, Y_i, T_i) \quad \text{and return to step 1.} \end{cases}$$

   Repeat this procedure until obtaining a truncated sample of size n, $(X_i, Y_i, T_i), i = 1, ..., n$. At the end of the procedure, we also obtain the value of the deterministic N, which allows us to obtain the estimated no-truncation rate $\hat{\alpha}_n = \frac{n}{N}$.

Step 5. We compute the estimator $\hat{F}(\cdot|x)$ defined in (8), using a bi-weight kernel

$$k(u) = \frac{15}{16}(1-u^2)^2\,\mathbb{1}_{\{|u|\le 1\}}(u),$$

which satisfies the required assumptions. For K_0, we take the Gaussian distribution. The bandwidth h_n is chosen according to the theoretical order $h_n = c(\log n/n)^{1/5}$, where the constant c is selected on a carefully chosen grid, in order to minimize the Global Mean Squared Error (GMSE) defined below in (14).

Step 6. Finally, we compute the Nadaraya-Watson estimator $m_{NW}(x)$ defined in (13), and our estimator $\hat{m}(x)$ defined in (9), using the score function

$$J(s) = \frac{1}{b}\mathbb{1}_{\{0\le s<b\}},$$

where $b \in (0,1)$ is chosen so that $m(x)$ can be estimated consistently (as discussed in Van Keilegom and Akritas (1999)). From the definition of $\hat{m}(x)$, it follows however that the largest possible value for b is

$$\min_{1\le i\le n}\left(\widehat{F}(+\infty|X_i)\right).$$

We then perform 500 Monte Carlo simulations. For each simulation run ℓ, $1 \le \ell \le 500$, the conditional distribution $\widehat{F}(y|X_i)$ is evaluated at the observed response Y_j, and the maximum value

$$M_i^{(\ell)} = \max_{1\le j\le n}\widehat{F}(Y_j|X_i)$$

is recorded for each X_i. The parameter $b^{(\ell)}$ for the ℓ-th iteration is then defined as

$$b^{(\ell)} = \min_{1\le i\le n} M_i^{(\ell)}.$$

Finally, averaging over all simulation runs yields the value of b, which is found to be approximately $b \approx 0.8$.

Linear model: We consider a linear regression model $Y_i = \frac{1}{3}X_i + 1 + \sigma\epsilon_i$, $i = 1, ..., n$. The means of the estimators $\hat{m}(x)$ and $m_{NW}(x)$, obtained from 150 samples for $x \in [-2, 1]$, are plotted in Figures 1 and 2.

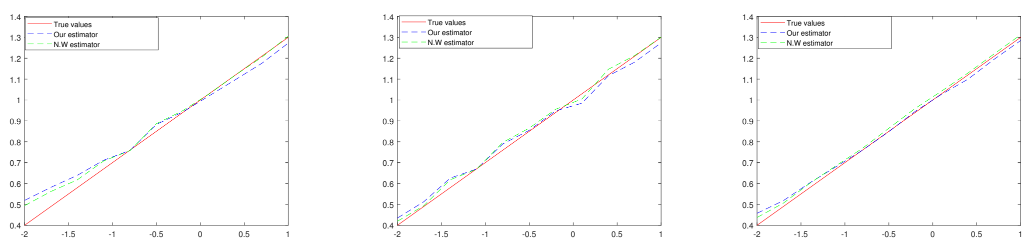

Figure 1: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $\alpha \approx 70\%$ for $n = 50, 100$ and 300 respectively

For Figure 1, we fix the no-truncation rate at $\alpha \approx 70\%$ and consider different sample sizes, $n = 50, 100$ and 300. It is evident that both estimators exhibit similar performance under the homoscedastic linear model, and the accuracy of the estimates improves as the sample size increases.

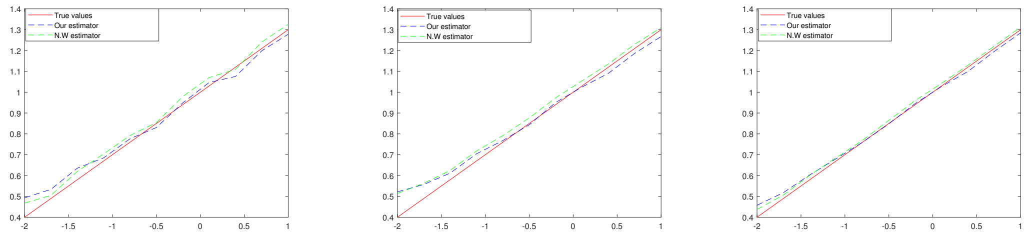

Figure 2: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $n = 300$ for $\alpha \approx 30\%, 50\%$ and 70% respectively

For Figure 2, we fix the sample size at $n = 300$ and consider different no-truncation rates $\alpha \approx 30\%, 50\%$ and 70%. We observe that the quality of fit is affected by the truncation rate, and deteriorates as the degree of truncation increases.

Nonlinear model: We consider the nonlinear regression function, $m(x) = \frac{1}{4}x^2$, and as before, the means of the estimators $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ are presented in Figures 3 and 4.

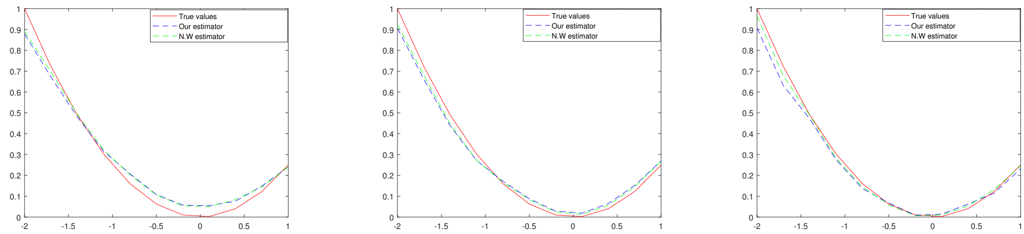

Figure 3: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $\alpha \approx 70\%$ for $n = 50, 100$ and 300 respectively

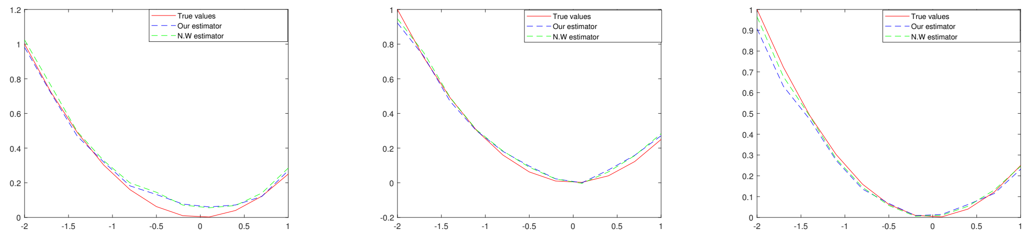

Figure 4: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $n = 300$ for $\alpha \approx 30\%, 50\%$ and 70% respectively

Again, under the nonlinear model, both estimators exhibit good and similar performance, with improved accuracy for larger sample sizes and lower truncation rates.

2. **Heteroscedastic regression case**

The model considered here assumes that the scale parameter σ depends on the covariate X. The same procedure as in the previous case is followed to evaluate the performance of the estimators. We consider the following two models

- **Model 1:** $Y_i = \frac{1}{3}X_i + 1 + \sigma(X_i)\epsilon_i$, $\sigma(X_i) = X_i^4$, $i = 1, ..., n$.
- **Model 2:** $Y_i = \frac{1}{2}X_i + \sigma(X_i)\epsilon_i$, $\sigma(X_i) = \exp(3X_i)$, $i = 1, ..., n$.

We plotted the means of the estimators $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ obtained from 150 samples, as shown in Figures 5 and 6 for Model 1, and Figures 7 and 8 for Model 2.

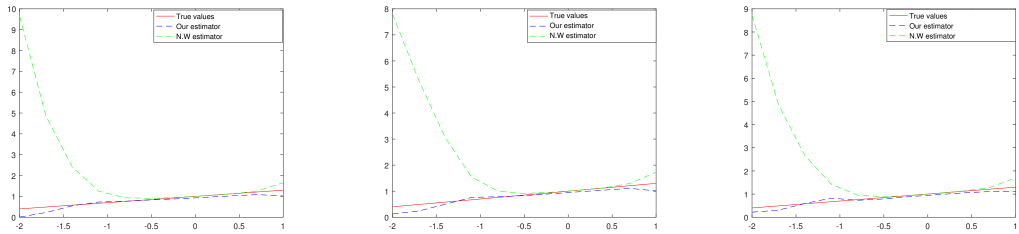

Figure 5: Model 1: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $\alpha \approx 70\%$ for $n = 50, 100$ and 300 respectively

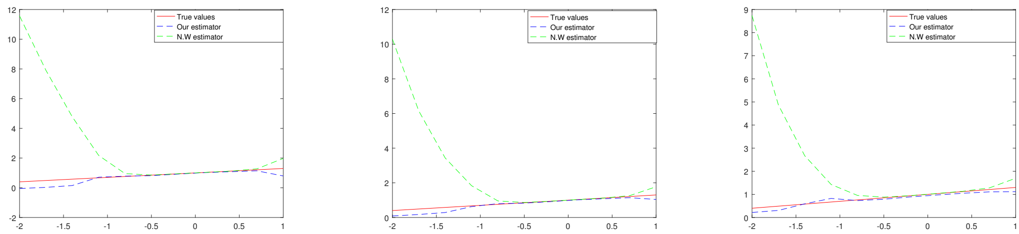

Figure 6: Model 1: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $n = 300$ for $\alpha \approx 30\%, 50\%$ and 70% respectively

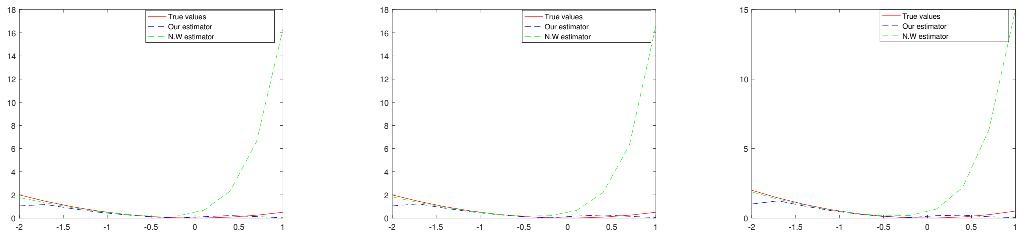

Figure 7: Model 2: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $\alpha \approx 70\%$ for $n = 50, 100$ and 300 respectively

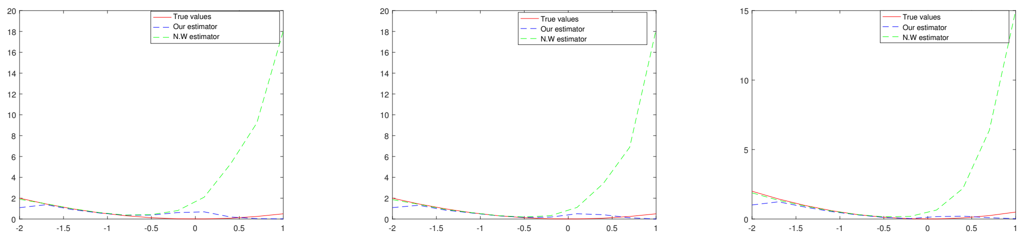

Figure 8: Model 2: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $n = 300$ for $\alpha \approx 30\%, 50\%$ and 70% respectively

From Figures 5 - 8, it is clearly seen that our estimator $\hat{m}(x)$ provides improved estimation compared to $m_{NW}(x)$. As expected, $\hat{m}(x)$ outperforms the Nadaraya-Watson estimator in the case of heteroscedastic data. These graphical findings are further confirmed by the numerical results presented in the following subsection.

3. Numerical results

Now, in order to examine and compare numerically, the performance of our estimator $\hat{m}(x)$ compared to $m_{NW}(x)$ in both homoscedastic and heteroscedastic cases, we computed the Global Mean Squared Error (GMSE) and the Global Standard Error (GSE) for both estimators. Specifically, for $\hat{m}(x)$, these quantities are defined as

$$\text{GMSE} = \frac{1}{BM}\sum_{k=1}^{B}\sum_{l=1}^{M}(\hat{m}_k(x_l) - m(x_l))^2, \tag{14}$$

and

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

where

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

$B = 150$ is the number of replications, M is the number of equidistant points x_l belonging the interval $[-2, 1]$ and $\hat{m}_k(x_l)$ is the value of $\hat{m}(x_l)$ computed at iteration k. The GMSE and the GSE are calculated for each combination of sample sizes n, and no-truncation rates α. The results are summarized in Table 1, for the following models.

- **Homoscedastic model:** $Y_i = \frac{1}{3}X_i + 1 + \sigma\epsilon_i$, $\sigma = 0.2$, $i = 1, ..., n$.
- **Heteroscedastic model:** $Y_i = \frac{1}{3}X_i + 1 + \sigma(X_i)\epsilon_i$, $\sigma(X_i) = X_i^4$, $i = 1, ..., n$.

Table 1: GMSE and GSE values of $\hat{m}$ and m_{NW} for homoscedastic and heteroscedastic cases

		GMSE			
		Homoscedastic		Heteroscedastic	
$\alpha(\%)$	n	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}
70	50	0.0169	0.0148	0.0732	16.1174
	100	0.0111	0.0091	0.0310	9.5363
	300	0.0057	0.0044	0.0293	8.9882
50	50	0.0184	0.0162	0.0927	25.5334
	100	0.0134	0.0111	0.0630	17.0770
	300	0.0060	0.0057	0.0406	15.3713
30	50	0.0226	0.0223	0.1444	28.4668
	100	0.0174	0.0133	0.1062	19.8776
	300	0.0091	0.0088	0.0680	18.3795
		GSE			
		Homoscedastic		Heteroscedastic	
$\alpha(\%)$	n	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}
70	50	0.0104	0.0098	0.0199	0.2044
	100	0.0081	0.0075	0.0094	0.1254
	300	0.0053	0.0050	0.0072	0.0767
50	50	0.0108	0.0103	0.0221	0.2492
	100	0.0092	0.0084	0.0170	0.1561
	300	0.0063	0.0059	0.0123	0.0988
30	50	0.0124	0.0117	0.0281	0.2877
	100	0.0093	0.0085	0.0222	0.1863
	300	0.0067	0.0062	0.0173	0.1322

From Table 1, the numerical results support the graphical findings and indicate that:

- The two estimators have similar performance under homoscedastic model.
- Our estimator $\hat{m}(x)$ significantly outperforms the estimator $m_{NW}(x)$ under the heteroscedastic model.
- Both estimators achieve better accuracy for lower truncation rates and larger sample sizes.
- In heteroscedastic settings, $\hat{m}(x)$ exhibits superior performance in terms of both global mean squared error and variability, highlighting its practical usefulness.

In Table 2, we examine the impact of heteroscedasticity on the two estimators by varying the scale regression function $\sigma(\cdot)$. We consider the model

$$Y_i = cos(X_i) + \sigma(X_i)\epsilon_i,\ i = 1, ..., n,$$

for different combinations of no-truncation rates α, sample sizes n and forms of the scale function $\sigma(\cdot)$. We consider $\sigma(X) = \exp(X)$, $\exp(2X)$ and $\exp(5X/2)$.

Table 2: GMSE and GSE values of $\hat{m}$ et m_{NW}

		GMSE					
		$\exp(X)$		$\exp(2X)$		$\exp(5X/2)$	
$\alpha(\%)$	n	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}
	50	0.0788	0.3455	0.0851	3.3123	0.1450	9.4887
70	100	0.0561	0.2686	0.0735	2.8707	0.0977	7.7653
	300	0.0386	0.1766	0.0484	2.6914	0.0938	7.0435
	50	0.1480	0.5898	0.1745	5.2549	0.1795	12.0482
50	100	0.0937	0.4494	0.1478	3.9874	0.1295	11.0658
	300	0.0651	0.2677	0.0546	3.1772	0.1088	9.5370
	50	0.2687	1.0657	0.2788	6.8468	0.2831	16.3236
30	100	0.1673	0.8661	0.2141	5.4003	0.2073	14.4055
	300	0.1078	0.6267	0.1143	3.3861	0.1326	11.5260
		GSE					
		$\exp(X)$		$\exp(2X)$		$\exp(5X/2)$	
$\alpha(\%)$	n	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}
	50	0.0175	0.0279	0.0113	0.0573	0.0204	0.0891
70	100	0.0133	0.0202	0.0097	0.0435	0.0108	0.0620
	300	0.0070	0.0113	0.0087	0.0258	0.0072	0.0370
	50	0.0276	0.0444	0.0240	0.1105	0.0273	0.1432
50	100	0.0208	0.0349	0.0255	0.0747	0.0217	0.1128
	300	0.0146	0.0273	0.0096	0.0345	0.0166	0.0618
	50	0.0289	0.0455	0.0296	0.1190	0.0291	0.1520
30	100	0.0219	0.0371	0.0256	0.0753	0.0247	0.1219
	300	0.0159	0.0258	0.0185	0.0448	0.0181	0.0658

From Table 2, we observe that the performance of $m_{NW}(x)$ estimator deteriorates for large values of $\sigma(\cdot)$, whereas our estimator $\hat{m}(x)$ maintains better performance under high heteroscedasticity. This is reflected in both criterions GMSE and GSE, which are consistently lower for $\hat{m}(x)$ compared to $m_{NW}(x)$. Overall, the proposed estimator clearly outperforms the Nadaraya-Watson estimator in these settings, highlighting its robustness and reliability in the presence of strong heteroscedasticity.

4. **In the presence of outliers**

To asses the robustness of our estimator compared to the classical Nadaraya-Watson estimator. We introduce artificial outliers into the data by multiplying 2% of the observed values by a multiplier factor (MF). We consider different values of multiplier $MF = 0, 10$, and 50, with a sample size $n = 300$ and a no-truncation rate $\alpha \approx 70\%$. The simulation results are displayed in Figure 9 for homoscedastic model and Figure 10 for heteroscedastic model.

- **Homoscedastic model:** $Y_i = \frac{1}{3}X_i + 1 + \sigma\epsilon_i$, $\sigma = 0.2$, $i = 1, ..., n$.

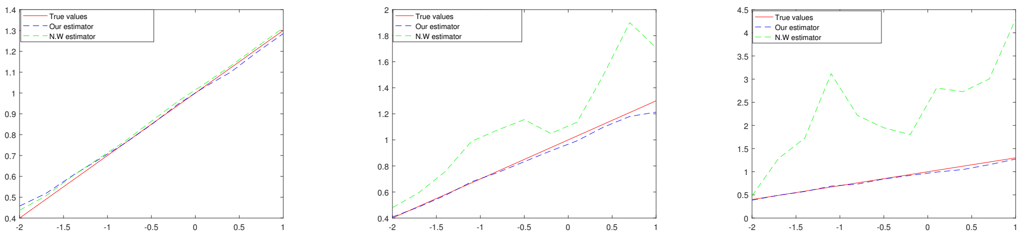

Figure 9: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $\alpha \approx 70\%$ and $n = 300$, for $MF = 0, 10$ and 50 respectively

- **Heteroscedastic model:** $Y_i = \frac{1}{3}X_i + 1 + \sigma(X_i)\epsilon_i$, $\sigma(X_i) = X_i^4$, $i = 1, ..., n$.

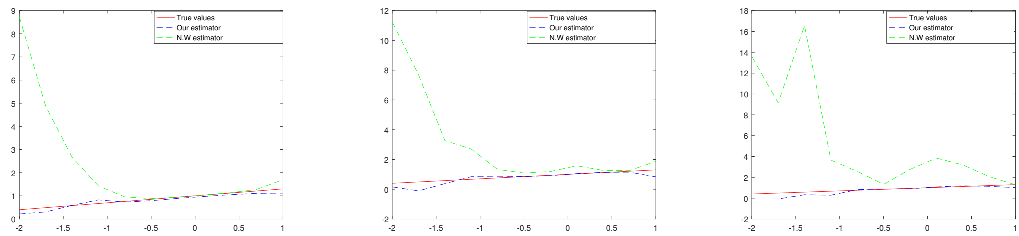

Figure 10: $m(\cdot)$, $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ with $\alpha \approx 70\%$ and $n = 300$, for $MF = 0, 10$ and 50 respectively

As expected, Figures 9 and 10 clearly show that our estimator $\hat{m}$ is more robust than the Nadaraya Watson estimator in presence of outliers. This is confirmed by Table 3 which reports the values of both the GMSE and the GSE criterions for the two estimators, $\hat{m}(x)$ and $m_{NW}(x)$, with respect to MF.

From Table 3, we observe that

- The estimator $\hat{m}$ is less influenced by the presence of outliers, than the estimator m_{NW}, so our estimator is more robust.
- The estimation quality of m_{NW} is very bad for high value of MF and small value of sample size n.

Table 3: GMSE and GSE values of $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$, for $\alpha \approx 70\%$

		Homoscedastic			
		GMSE		GSE	
n	MF	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}
100	0	0.0111	0.0091	0.0081	0.0075
	10	0.0277	0.3657	0.0104	0.0474
	50	0.0311	10.3947	0.0115	0.2524
300	0	0.0057	0.0044	0.0053	0.0050
	10	0.0196	0.1079	0.0071	0.0243
	50	0.0255	4.9365	0.0090	0.1719
		Heteroscedastic			
		GMSE		GSE	
n	MF	$\hat{m}$	m_{NW}	$\hat{m}$	m_{NW}
100	0	0.0310	9.5363	0.0094	0.1254
	10	0.3046	24.2162	0.0361	0.3054
	50	0.5212	88.2033	0.0452	0.7310
300	0	0.0293	8.9882	0.0072	0.0767
	10	0.2763	16.9654	0.0348	0.1926
	50	0.3093	62.5853	0.0411	0.6228

4.2. Performance of the scale function estimator

Building upon the previous steps, we now examine the consistency of the scale function $\sigma(\cdot)$ under the heteroscedastic model $Y_i = 2X_i + \sigma(X_i)\epsilon_i$, for $i = 1, ..., n$. We consider the exponential scale function $\sigma(x) = \sqrt{\exp(x)}$. The mean of the estimators $\hat{\sigma}^2(\cdot)$, obtained from 150 samples, are plotted for different sample sizes in Figure 11, and for different no-truncation rates in Figure 12

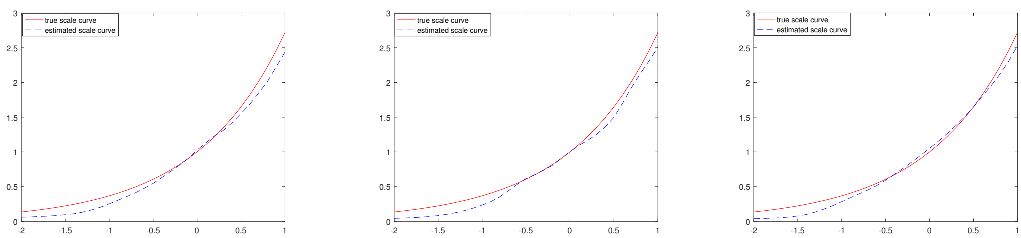

Figure 11: $\sigma^2(\cdot)$, $\hat{\sigma}^2(\cdot)$ with $\alpha \approx 70\%$ for $n = 50, 100$ and 300 respectively

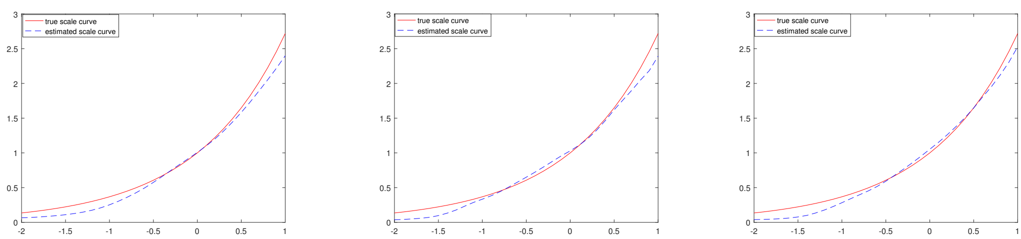

Figure 12: $\sigma^2(\cdot)$, $\hat{\sigma}^2(\cdot)$ with $n = 300$ for $\alpha \approx 30\%, 50\%$ *and* 70% respectively

Figures 11 and 12 show that the estimator accurately reproduces the true scale function and maintains good performance even for small sample sizes, although its accuracy appears to be slightly affected by the no-truncation rate. As expected, the estimation improves as the sample size increases and the no-truncation rate decreases. Table 4 presents the Global Mean Squared Error and the global standard error of the estimator $\hat{\sigma}^2(x)$ for different combinations of sample sizes n and no-truncation rates $\alpha\%$.

Table 4: GMSE and GSE values of $\hat{\sigma}^2(\cdot)$

$\alpha(\%)$	n	GMSE	GSE
	50	0.1142	0.0269
70	100	0.0706	0.0207
	300	0.0338	0.0127
	50	0.1524	0.0290
50	100	0.0925	0.0216
	300	0.0414	0.0141
	50	0.1818	0.0297
30	100	0.1142	0.0235
	300	0.0506	0.0156

The numerical results support the graphical findings. The Global Mean Squared Error indicates that the estimator achieves a good precision across different sample sizes and no-truncation rates, while the Global Standard Error reflects low variability in the estimates.

5. Application to dolphin sonar signal data

In this section, we evaluate the performance of the proposed location estimator through a real dataset, that records the sound pressure of sonar signals emitted by a dolphin at various distances from a target. The covariate X represents the distance to the dolphin (in meters), and the response variable Y corresponds to the water density. These measurements were taken off the coast of Iceland, near Keflavik in 1998. The original dataset is publicly available at `http://www.statsci.org/data/general/dolphin.html`, and were studied in the work of Zhou and Zhu (2015). A total of 1634 observations were recorded, among which at least one clear outlier was identified, where the sound pressure is significantly higher than 230. In what follows, we restrict the dataset to $N = 300$, deliberately retaining the outlier as in Figure 13 (left).

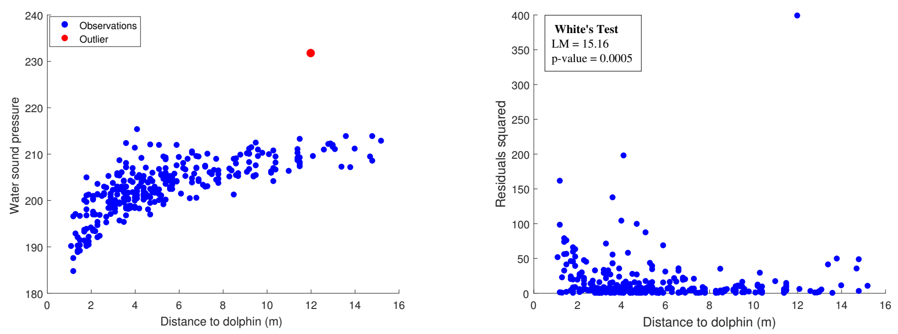

Figure 13: Dolphin sonar signal data analysis (left). Residual plot for heteroscedasticity (right)

The objective of this study is to predict the water sound pressure emitted by a dolphin based on the distance to the target, and to compare the performance of the estimators $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$. For this purpose, the full dataset is divided into two subsets: a training set of size 260, used to estimate the regression function, and a test set of size 40, used to evaluate

predictive performance.
We perform the White test for heteroscedasticity. The Lagrange Multiplier (LM) statistic, computed as $LM = n \times R^2$, where R^2 is the coefficient of determination obtained from the auxiliary regression of squared residuals on the explanatory variables. The test yields $LM = 15.16$ with a $p - value$ of 0.0005, confirming significant heteroscedasticity in the dataset. This result is also supported by the residual plot (see Figure 13, right), indicating that the variance of the residual increases with the explanatory variable.
Furthermore, a low sound pressure levels may be indistinguishable from background noise and could go undetected by standard sensors. To reflect this practical limitation, we assume that only signals exceeding a fixed threshold $T = 200$ dB (decibels) are observable. Although this threshold does not correspond to a strict physical detection limit, it reflects the practical reality that very weak signals are unlikely to be detected, thereby justifying the left truncation on the dataset. The data are therefore considered to be left-truncated, with an estimated truncation rate of $1 - \alpha \approx 25\%$.
The estimators are constructed based on the same choice of the kernel k, the distribution K_0 and the score function $J(\cdot)$ as in the previous section. The optimal bandwidths for both estimators are determined using the leave-one-out cross-validation technique, defined (for $\hat{m}(\cdot)$) by

$$h_{opt} = \arg\min_{h_n} \left\{ \frac{1}{260} \sum_{j=1}^{260} (Y_j - \hat{m}_{-j}(X_j))^2 \right\},$$

where $\hat{m}_{-j}(\cdot)$ denotes the estimator computed with the j-th observation removed. The minimization is performed over a grid of bandwidth values $h \in [0.1, 10]$ with a step size 0.1.
Figure 14, presents the fitted (left) and predicted (right) sound pressure values versus the true observed values, using both $\hat{m}$ and m_{NW} estimators.

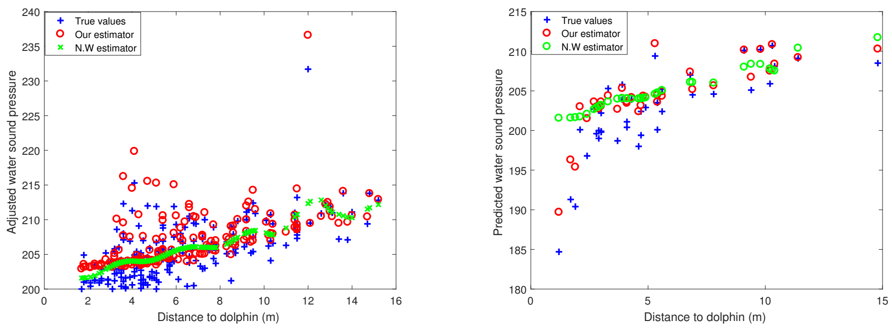

Figure 14: Adjusted water sound pressure (left) and predicted (right) values against the true observed values using $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ estimators

From the plot (Figure 14), we can observe that our estimator closely follows the true values for both the training and the predicted data, and performs better than the classical regression one, particularly in regions with higher variability.
Next, since the dataset currently includes a single outlier, we aim to introduce additional outliers in the training sample to assess the stability and robustness of our proposed estimator in comparison to the classical one. We randomly select 5 (Figure 15, left) and 10 (Figure 15, right) observations and artificially contaminate them by multiplying their values by a factor of 2.

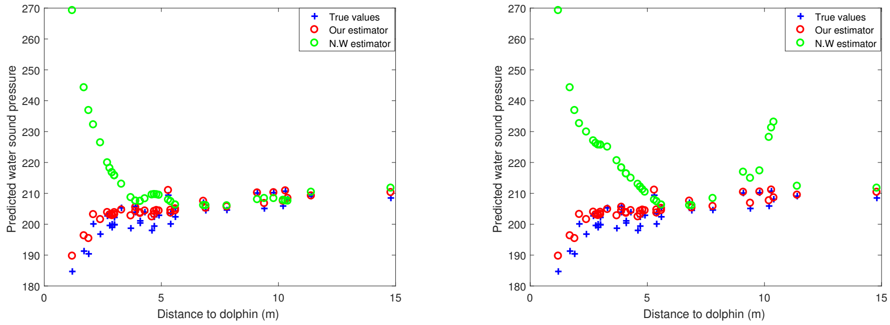

Figure 15: Effect of increased contamination with 5 and 10 contaminated values, respectively

The results in Figure 15 highlight the robustness of our estimator to perturbations compared to the classical regression estimator, our method maintains its stability and prediction accuracy.
To support the graphical findings, we compute several performance criteria for both estimators, using both the original and perturbed datasets. Specifically, we assess:
The Mean Predictive Squared Error (MPSE), which quantifies the overall predictive accuracy of the estimators

$$\mathrm{MPSE} = \frac{1}{40}\sum_{j=1}^{40}\left(Y_j - \hat{m}_{-j}(X_j)\right)^2.$$

The Mean Predictive Relative Error (MPRE), which evaluates the prediction error relative to the magnitude of the observed values.

$$\mathrm{MPRE} = \frac{1}{40}\sum_{j=1}^{40}\left|\frac{Y_j - \hat{m}_{-j}(X_j)}{Y_j}\right|.$$

The Standard Predictive Error (SPE) measures the variability of the predictions, reflecting both the dispersion of the prediction errors and the stability of the estimator.

$$\mathrm{SPE} = \frac{1}{\sqrt{40}}\sqrt{\frac{1}{39}\sum_{j=1}^{40}(e_j - \bar{e})^2},$$

where $e_j = Y_j - \hat{m}_{-j}(X_j)$ and $\bar{e} = \frac{1}{40}\sum_{j=1}^{40} e_j$.

Table 5: MPSE, MPRE (%), and SPE for $\hat{m}(\cdot)$ and $m_{NW}(\cdot)$ under original and perturbed data

Data		MPSE	MPRE (%)	SPE
Original data	$\hat{m}$	6.6293	1.01	0.2729
	m_{NW}	20.6447	1.60	0.6251
+5 outliers	$\hat{m}$	6.8538	1.04	0.2698
	m_{NW}	418.2662	6.16	2.7025
+10 outliers	$\hat{m}$	7.0611	1.07	0.2610
	m_{NW}	578.0344	8.99	2.6051

The results in Table 5 further highlight the superiority of our estimator, particularly under heteroscedasticity and in the presence of outliers. It consistently achieves lower predictive errors (MPSE and MPRE) and smaller SPE values, indicating improved accuracy and greater stability compared to the Nadaraya-Watson estimator.

6. Proofs

6.1. Strong consistency and almost sure representation

Proof of Theorem 1
(a) Using integration by parts, we get

$$\hat{m}(x) - m(x) = -\int_0^1 L(s)d\hat{F}^{-1}(s|x) + \int_0^1 L(s)dF^{-1}(s|x), \tag{15}$$

where $L(u) = \int_0^u J(s)ds$, for $0 \leq u \leq 1$. Make use of the two substitutions $y = \hat{F}^{-1}(s|x)$, and $y = F^{-1}(s|x)$ for the first and the second integral in (15), respectively, we get

$$\hat{m}(x) - m(x) = \int_0^{+\infty} [L(F(y|x)) - L(\hat{F}(y|x))]dy,$$

using again the substitution $y = F^{-1}(s|x)$, and applying the inverse function theorem, we obtain

$$\begin{aligned}\hat{m}(x) - m(x) &= \int_0^1 [L(s) - L(\hat{F}(F^{-1}(s|x)|x))]dF^{-1}(s|x) \\ &= \int_0^1 \frac{[L(s) - L(\hat{F}(F^{-1}(s|x)|x))]}{f(F^{-1}(s|x)|x)}ds.\end{aligned}$$

A Taylor expansion of order one of $L(\hat{F}(F^{-1}(s|x)|x))$ around $F(F^{-1}(s|x)|x)$, allows to write

$$\hat{m}(x) - m(x) = -\int_0^1 \left[J(\xi_{xs})(\hat{F}(F^{-1}(s|x)|x) - F(F^{-1}(s|x)|x))\right] \frac{1}{f(F^{-1}(s|x)|x)}ds,$$

where ξ_{xs} is between $\hat{F}(F^{-1}(s|x)|x))$ and $F(F^{-1}(s|x)|x)$.
By Assumption (M1), we get

$$\begin{aligned}\sup_{x\in\Omega} |\hat{m}(x) - m(x)| \leq &\left(\inf_{x\in\Omega} \inf_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x)\right)^{-1} \sup_{s_0\leq s\leq s_1} J(s) \\ &\times \sup_{x\in\Omega} \sup_{F^{-1}(s_0|x)\leq y\leq F^{-1}(s_1|x)} \left|\hat{F}(y|x) - F(y|x)\right|.\end{aligned}$$

It is worth mentioning that under Assumption (M1), for all x, $[F^{-1}(s_0|x), F^{-1}(s_1|x)] \subseteq [a_x, b_x]$, since we have supposed that $s_1 \leq \inf_{x\in\Omega} F(b_x|x)$ and $s_0 \geq \sup_{x\in\Omega} F(a_x|x)$, then from Assumption (M2), it follow that

$$\sup_{x\in\Omega} |\hat{m}(x) - m(x)| = O\left(\sup_{x\in\Omega} \sup_{a_x\leq y\leq b_x} |\hat{F}(y|x) - F(y|x)|\right), \ \mathbf{P} - a.s\ as\ n \to \infty.$$

Therefore, from Proposition 1 in Lemdani *et al.* (2009), we have for any $a_{F(\cdot)} < a < b < b_{F(\cdot)}$,

$$\sup_{x\in\Omega} \sup_{a\leq y\leq b} |\hat{F}(y|x) - F(y|x)| = O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right), \ \mathbf{P} - a.s\ as\ n \to \infty, \tag{16}$$

and as, $[a_{F(\cdot|x)}, b_{F(\cdot|x)}] \subset [a_{F(\cdot)}, b_{F(\cdot)}]$, we obtain the desired result.

Part (b) can be shown in a similar way as before. Indeed write,

$$\hat{\sigma}(x) - \sigma(x) = \frac{\hat{\sigma}^2(x) - \sigma^2(x)}{\hat{\sigma}(x) + \sigma(x)}. \tag{17}$$

For the numerator of the right hand side, we have

$$\begin{aligned}\hat{\sigma}^2(x) - \sigma^2(x) &= \int_0^1 \left[\hat{F}^{-1}(s|x)^2 - F^{-1}(s|x)^2\right] J(s)ds - (\hat{m}^2(x) - m^2(x)) \\ &=: I(x) - II(x).\end{aligned} \tag{18}$$

To treat $I(x)$ we use the same steps as in part (a), we obtain

$$\begin{aligned} I(x) =& -2\int_0^1 L(s)\hat{F}^{-1}(s|x)d\hat{F}^{-1}(s|x) + 2\int_0^1 L(s)F^{-1}(s|x)dF^{-1}(s|x) \\ =& -2\int_0^{+\infty}[L(\hat{F}(y|x)) - L(F(y|x))]ydy. \end{aligned}$$

Which gives under Assumption (M1)

$$\begin{aligned} \sup_{x\in\Omega}|I(x)| \leq& 2\left(\inf_{x\in\Omega}\inf_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x)\right)^{-1}\sup_{s_0\leq s\leq s_1} J(s)\sup_{s_0\leq s\leq s_1}|F^{-1}(s|x)| \\ &\times \sup_{x\in\Omega}\sup_{F^{-1}(s_0|x)\leq y\leq F^{-1}(s_1|x)}\left|\hat{F}(y|x) - F(y|x)\right|. \end{aligned}$$

Assumption (M2) allows to write

$$\sup_{x\in\Omega}|I(x)| = O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right), \ \mathbf{P} - a.s \ as \ n\to\infty. \tag{19}$$

On the other hand, $II(x)$, can be written as

$$II(x) = 2m(x)(\hat{m}(x) - m(x)) + (\hat{m}(x) - m(x))^2. \tag{20}$$

From the part (a) of Theorem 1, we obtain

$$\sup_{x\in\Omega}|II(x)| = O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right), \ \mathbf{P} - a.s \ as \ n\to\infty. \tag{21}$$

Combining (18), (19) and (21), we get

$$\sup_{x\in\Omega}|\hat{\sigma}^2(x) - \sigma^2(x)| = O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right), \ \mathbf{P} - a.s \ as \ n\to\infty. \tag{22}$$

Then, from (17) and (22) and the fact that $\inf_{x\in\Omega}\sigma(x) > 0$, we obtain

$$\begin{aligned} \sup_{x\in\Omega}|\hat{\sigma}(x) - \sigma(x)| \leq& \frac{\sup_{x\in\Omega}|\hat{\sigma}^2(x) - \sigma^2(x)|}{\inf_{x\in\Omega}\hat{\sigma}(x) + \inf_{x\in\Omega}\sigma(x)} \leq \frac{\sup_{x\in\Omega}|\hat{\sigma}^2(x) - \sigma^2(x)|}{\inf_{x\in\Omega}\sigma(x)} \\ =& O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right), \ \mathbf{P} - a.s \ as \ n\to\infty. \end{aligned}$$

The proof of Theorem 1 is complete. □

Proof of Proposition 1

Defining the pseudo estimators of $F_1(x, y)$ and $v(x)$, respectively by

$$\begin{aligned} \tilde{F}_1(x,y) :=& \frac{\alpha}{nh_n}\sum_{i=1}^n G^{-1}(Y_i)k\left(\frac{x - X_i}{h_n}\right)K_0\left(\frac{y - Y_i}{h_n}\right), \\ \tilde{v}(x) :=& \frac{\alpha}{nh_n}\sum_{i=1}^n G^{-1}(Y_i)k\left(\frac{x - X_i}{h_n}\right). \end{aligned}$$

A classical decomposition permits to write

$$
\begin{aligned}
\hat{F}(y|x) - F(y|x) =& \frac{\hat{F}_1(x,y)}{v_n(x)} - \frac{F_1(x,y)}{v_n(x)} + \frac{F_1(x,y)}{v_n(x)} - \frac{F_1(x,y)}{v(x)} \\
=& \frac{1}{v_n(x)} \left[\hat{F}_1(x,y) - F_1(x,y) + F(y|x)\left(v(x) - v_n(x)\right)\right] \\
=& \frac{1}{v_n(x)} \Big[\hat{F}_1(x,y) - \tilde{F}_1(x,y) + \tilde{F}_1(x,y) - \mathbf{E}(\tilde{F}_1(x,y)) \\
&+ \mathbf{E}(\tilde{F}_1(x,y)) - F_1(x,y) - F(y|x)\left(v_n(x) - \tilde{v}(x) + \tilde{v}(x) - \mathbf{E}(\tilde{v}(x))\right. \\
&+ \left.\mathbf{E}(\tilde{v}(x)) - v(x)\right)\Big] \\
=:& \frac{1}{v_n(x)} \left[\tilde{F}_1(x,y) - \mathbf{E}(\tilde{F}_1(x,y)) - F(y|x)\left(\tilde{v}(x) - \mathbf{E}(\tilde{v}(x))\right)\right] \\
&+ R_{n,1}(x,y) + R_{n,2}(x,y) - R_{n,3}(x,y) - R_{n,4}(x,y), \\
=:& I(x,y) + R_{n,1}(x,y) + R_{n,2}(x,y) - R_{n,3}(x,y) - R_{n,4}(x,y). \qquad (23)
\end{aligned}
$$

From Lemma 3 in Lemdani *et al.* (2009), under Assumptions (A1)(i), (A2) and (A3) we have, for all $x \in \Omega$,

$$
v_n(x) \longrightarrow v(x) > 0, \quad \mathbf{P} - a.s \ as \ n \to \infty. \qquad (24)
$$

We analyze the terms $R_{n,j}$, $j = 1, ..., 4$ as follows:

$$
\begin{aligned}
|R_{n,1}(x,y)| &= \frac{|\hat{F}_1(x,y) - \tilde{F}_1(x,y)|}{v_n(x)} \\
&\leq \frac{1}{v_n(x)} \left\{ \frac{|\alpha_n - \alpha|}{G_n(a_F)} + \frac{\alpha}{G_n(a_F)G(a_F)} \left(\sup_{a \leq y \leq b} |G_n(y) - G(y)| \right) \right\} \frac{1}{nh_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h_n}\right),
\end{aligned}
$$

under (A2)(ii). Hence using similar arguments as in the proof of (27) in Lemdani *et al.* (2009), we get

$$
\sup_{x \in \Omega} \sup_{a_x \leq y \leq b_x} |R_{n,1}(x,y)| = O(n^{-1/2}), \ \mathbf{P} - a.s \ as \ n \to \infty. \qquad (25)
$$

$$
\begin{aligned}
R_{n,2}(x,y) &= \frac{\mathbf{E}(\tilde{F}_1(x,y)) - F_1(x,y)}{v_n(x)} \\
&= \frac{1}{v_n(x)} \left\{ \int\int k(r)k_0(s)[F_1(x - rh_n, y - sh_n) - F_1(x,y)]drds \right\}.
\end{aligned}
$$

Under Assumptions (A2) and (A3), using a results (24) and a Taylor expansion, we obtain

$$
\sup_{x \in \Omega} \sup_{a_x \leq y \leq b_x} |R_{n,2}(x,y)| = O(h_n^2), \ as \ n \to \infty. \qquad (26)
$$

$$
\begin{aligned}
|R_{n,3}(x,y)| &= F(y|x)\frac{|v_n(x) - \tilde{v}(x)|}{v_n(x)} \\
&\leq \frac{F(y|x)}{v_n(x)} \left\{ \frac{|\alpha_n - \alpha|}{G_n(a_F)} + \frac{\alpha}{G_n(a_F)G(a_F)} \left(\sup_{a \leq y \leq b} |G_n(y) - G(y)| \right) \right\} \frac{1}{nh_n} \sum_{i=1}^{n} k\left(\frac{x - X_i}{h_n}\right).
\end{aligned}
$$

As for $R_{n,1}$, we get

$$
\sup_{x \in \Omega} \sup_{a_x \leq y \leq b_x} |R_{n,3}(x,y)| = O(n^{-1/2}), \ \mathbf{P} - a.s \ as \ n \to \infty. \qquad (27)
$$

$$\begin{aligned}R_{n,4}(x,y) &= F(y|x)\frac{\mathbf{E}(\tilde{v}(x)) - v(x)}{v_n(x)}\\ &= \frac{F(y|x)}{v_n(x)}\left\{\int k(r)[v(x-rh_n) - v(x)]dr\right\}.\end{aligned}$$

Again, using a Taylor expansion, under Assumptions (A2) and (A3), we get

$$\sup_{x\in\Omega}\sup_{a_x\le y\le b_x}|R_{n,4}(x,y)| = O(h_n^2), as\ n\to\infty. \tag{28}$$

Furthermore, for $I(x,y)$, we write

$$I(x,y) = \frac{1}{v(x)}\left[\tilde{F}_1(x,y) - \mathbf{E}(\tilde{F}_1(x,y)) - F(y|x)\left(\tilde{v}(x) - \mathbf{E}(\tilde{v}(x))\right)\right] + R_{n,5}(x,y),$$

where

$$R_{n,5}(x,y) = \left[\tilde{F}_1(x,y) - \mathbf{E}(\tilde{F}_1(x,y)) - F(y|x)\left(\tilde{v}(x) - \mathbf{E}(\tilde{v}(x))\right)\right]\left[\frac{1}{v_n(x)} - \frac{1}{v(x)}\right].$$

From the proof of (28) in Lemma 2 of Lemdani *et al.* (2009), under Assumptions (A1)(i) and (A2), we have

$$\sup_{x\in\Omega}\sup_{a_x\le y\le b_x}|\tilde{F}_1(x,y) - \mathbf{E}(\tilde{F}_1(x,y))| = O\left((nh_n)^{-1/2}(\log n)^{1/2}\right),\ \mathbf{P}-a.s\ as\ n\to\infty,$$

and, from Lemma 3 of Lemdani *et al.* (2009), under Assumptions (A1)(i) and (A2)(i), we get

$$\begin{aligned}\sup_{x\in\Omega}|\tilde{v}(x) - \mathbf{E}(\tilde{v}(x))| &= O\left((nh_n)^{-1/2}(\log n)^{1/2}\right),\ \mathbf{P}-a.s\ as\ n\to\infty,\\ \sup_{x\in\Omega}|v_n(x) - v(x)| &= O\left(\max\left\{(nh_n)^{-1/2}(\log n)^{1/2}, h_n^2\right\}\right),\ \mathbf{P}-a.s\ as\ n\to\infty.\end{aligned}$$

Which gives

$$\sup_{x\in\Omega}\sup_{a_x\le y\le b_x}|R_{n,5}(x,y)| = O\left((nh_n)^{-1}(\log n)\right),\ \mathbf{P}-a.s\ as\ n\to\infty. \tag{29}$$

Finally, using the corresponding definitions given in Section 3.2, and the results in (25)-(29), combining with (23), we get

$$\begin{aligned}\hat{F}(y|x) - F(y|x) =&(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\{\xi_1(X_i,Y_i,x,y) - F(y|x)\xi_2(X_i,Y_i,x)\} + R_n(x,y)\\ =:&(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\xi(X_i,Y_i,x,y) + R_n(x,y),\end{aligned}$$

where

$$R_n(x,y) = \sum_{i=1}^{5}R_{n,i}(x,y).$$

and

$$\sup_{x\in\Omega}\sup_{a_x\le y\le b_x}|R_n(x,y)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right),\ \mathbf{P}-a.s\ as\ n\to\infty.$$

Which complete the proof. □

Proof of Theorem 2
Using the notation $L(u) = \int_0^u J(s)ds$, under Assumption (M3), using a Taylor expansion of $L(\hat{F}(y|x))$ to order two, around $F(y|x)$, we can write

$$\begin{aligned}\hat{m}(x) - m(x) &= - \int_0^{+\infty} [L(\hat{F}(y|x)) - L(F(y|x))]dy \\ &= - \int_0^{+\infty} J(F(y|x))(\hat{F}(y|x) - F(y|x))dy \\ &\quad - \frac{1}{2}\int_0^{+\infty} J'(x_y)(\hat{F}(y|x) - F(y|x))^2 dy \\ &=: I(x) - R_{n,1}(x), \end{aligned} \tag{30}$$

where x_y lies between $F(y|x)$ and $\hat{F}(y|x)$, and $J'(\cdot)$ is the derivative of $J(\cdot)$. From (16), we get

$$\sup_{x\in\Omega} |R_{n,1}(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^4\right\}\right),\ \mathbf{P} - a.s\ as\ n \to \infty. \tag{31}$$

Make use of Proposition 1, $I(x)$ can be written as

$$\begin{aligned} I(x) =& - (nh_n)^{-1}v^{-1}(x)\alpha \sum_{i=1}^{n} \int_0^{+\infty} \xi(X_i, Y_i, x, y)J(F(y|x))dy \\ & - \int_0^{+\infty} J(F(y|x))R_n(x, y)dy \\ =&: -(nh_n)^{-1}v^{-1}(x)\alpha \sum_{i=1}^{n} \eta(X, Y|x) - R_{n,2}(x). \end{aligned} \tag{32}$$

Hence, again by Proposition 1 and Assumption (M2), we get

$$\sup_{x\in\Omega} |R_{n,2}(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right),\ \mathbf{P} - a.s\ as\ n \to \infty. \tag{33}$$

Put $R_n(x) = -(R_{n,1}(x) + R_{n,2}(x))$, then by (31) and (33), we get

$$\sup_{x\in\Omega} |R_n(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right),\ \mathbf{P} - a.s\ as\ n \to \infty.$$

So, the proof is complete. □

Proof of Theorem 3
For $x \in \Omega$, write

$$\hat{\sigma}(x) - \sigma(x) = \frac{\hat{\sigma}^2(x) - \sigma^2(x)}{2\sigma(x)} - \frac{(\hat{\sigma}(x) - \sigma(x))^2}{2\sigma(x)}. \tag{34}$$

Put

$$\tilde{R}_{n,1}(x) = \frac{(\hat{\sigma}(x) - \sigma(x))^2}{2\sigma(x)},$$

then, from Theorem 1 we get

$$\sup_{x\in\Omega} |\tilde{R}_{n,1}(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^4\right\}\right),\ \mathbf{P} - a.s\ as\ n \to \infty. \tag{35}$$

For the first term of (34), similarly as in the proof of part (b) of Theorem 1, we have

$$\begin{aligned}\hat{\sigma}^2(x) - \sigma^2(x) =& - 2\int_0^{+\infty} [L(\hat{F}(y|x)) - L(F(y|x))]ydy - (\hat{m}^2(x) - m^2(x)) \\ =&: I(x) - II(x).\end{aligned}$$

For $I(x)$, by a Taylor expansion, we have

$$I(x) = -2\int_0^{+\infty} J(F(y|x))(\hat{F}(y|x) - F(y|x))ydy - \int_0^{+\infty} J'(\tilde{x}_y)(\hat{F}(y|x) - F(y|x))^2 ydy$$
$$=: -2(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\int_0^{+\infty}\xi(X_i, Y_i, x, y)J(F(y|x))ydy - \tilde{R}_{n,2}(x) - \tilde{R}_{n,3}(x), \quad (36)$$

where $\tilde{x}_y$ is between $F(y|x)$ and $\hat{F}(y|x)$, $\tilde{R}_{n,2}(x) = 2\int_0^{+\infty} J(F(y|x))R_n(x,y)ydy$, $\tilde{R}_{n,3}(x) = \int_0^{+\infty} J'(\tilde{x}_y)(\hat{F}(y|x) - F(y|x))^2 ydy$.

Under Assumptions (M2) and (M3), the result in (16) and Proposition 1, we get

$$\sup_{x\in\Omega}|\tilde{R}_{n,2}(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right), \ \mathbf{P} - a.s \ as \ n \to \infty. \quad (37)$$
$$\sup_{x\in\Omega}|\tilde{R}_{n,3}(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^4\right\}\right), \ \mathbf{P} - a.s \ as \ n \to \infty. \quad (38)$$

For $II(x)$, using (20) and the result of Theorem 2, we obtain

$$II(x) = 2m(x)(\hat{m}(x) - m(x)) + (\hat{m}(x) - m(x))^2$$
$$=: -2m(x)(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\int_0^{+\infty}\xi(X_i, Y_i, x, y)J(F(y|x))dy$$
$$+ \tilde{R}_{n,4}(x) + \tilde{R}_{n,5}(x), \quad (39)$$

where, $\tilde{R}_{n,4}(x) = 2m(x)R_n(x)$, and $\tilde{R}_{n,5}(x) = (\hat{m}(x) - m(x))^2$.
Again, from Theorem 2, we have

$$\sup_{x\in\Omega}|\tilde{R}_{n,4}(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right), \ \mathbf{P} - a.s \ as \ n \to \infty, \quad (40)$$

and by Theorem 1, we obtain

$$\sup_{x\in\Omega}|\tilde{R}_{n,5}(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^4\right\}\right), \ \mathbf{P} - a.s \ as \ n \to \infty, \quad (41)$$

Hence, by (36) and (39), we get

$$\hat{\sigma}^2(x) - \sigma^2(x) = -2(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\left[\int_0^{+\infty}\xi(X_i, Y_i, x, y)J(F(y|x))ydy\right.$$
$$\left. -m(x)\int_0^{+\infty}\xi(X_i, Y_i, x, y)J(F(y|x))dy\right] - \tilde{R}_{2,n}(x) - \tilde{R}_{n,3}(x)$$
$$- \tilde{R}_{n,4}(x) - \tilde{R}_{n,5}(x). \quad (42)$$

Finally, combining (34) with (42), we get

$$\hat{\sigma}(x) - \sigma(x) := -(nh_n)^{-1}v^{-1}(x)\alpha\sum_{i=1}^{n}\zeta(X_i, Y_i|x) + \tilde{R}_n(x),$$

where

$$\tilde{R}_n(x) = -\sum_{i=1}^{5}\tilde{R}_{n,i}(x).$$

Then, from (35), (37), (38), (40) and (41), we obtain

$$\sup_{x\in\Omega}|\tilde{R}_n(x)| = O\left(\max\left\{(nh_n)^{-1}(\log n), h_n^2\right\}\right), \ \mathbf{P} - a.s \ as \ n \to \infty.$$

Which complete the proof. □

6.2. Asymptotic normality

Note that we prove only Theorem 4, the proof of Theorem 5 can be done following the same steps.

Proof of Theorem 4 From the expression of $(\hat{m}(x) - m(x))$ in Theorem 2, we can write

$$\begin{aligned}\sqrt{nh_n}(\hat{m}(x) - m(x)) =& \frac{-\alpha v^{-1}(x)}{\sqrt{nh_n}} \sum_{i=1}^{n} \eta(X_i, Y_i|x) + \sqrt{nh_n} R_n(x) \\ =:& \sum_{i=1}^{n} M_{ni}(x) + \sqrt{nh_n} R_n(x).\end{aligned}$$

Then to prove the asymptotic normality of $\hat{m}(\cdot)$, we first consider the negligible term $\sqrt{nh_n} R_n(x)$ (Lemma 1) and then provide the explicit expression of the variance of $\sum_{i=1}^{n} M_{ni}(x)$ (Lemma 2). Finally apply the Lindeberg's condition (Lemma 3) .

Lemma 1. *Under Assumptions of Theorem 4, for any $x \in \Omega$, we have*

$$\sqrt{nh_n}|R_n(x)| = o_p(1), \text{ as } n \to \infty.$$

Proof. The proof is a direct consequence of (11) under Assumptions A1(ii) and A1(iii). □

Lemma 2. *Under Assumptions (A2), (A3) and (M1)-(M3), for any $x \in \Omega$, we have*

$$Var\left(\sum_{i=1}^{n} M_{ni}(x)\right) \to \Delta^2(x), \text{ as } n \to \infty,$$

where $\Delta^2(x)$ is as defined in (12).

Proof.

$$\begin{aligned}Var\left(\sum_{i=1}^{n} M_{ni}(x)\right) =& \frac{\alpha^2}{h_n v^2(x)} Var\left(\eta(X_1, Y_1|x)\right) \\ =& \frac{\alpha^2}{h_n v^2(x)} Var\left(\int_0^{+\infty} \xi(X_1, Y_1, x, y) J(F(y|x)) dy\right) \\ =& \frac{\alpha^2}{h_n v^2(x)} Var\left(\int_0^{+\infty} \xi_1(X_1, Y_1, x, y) J(F(y|x)) dy\right) \\ &+ \frac{\alpha^2}{h_n v^2(x)} Var\left(\int_0^{+\infty} F(y|x) \xi_2(X_1, Y_1, x) J(F(y|x)) dy\right) \\ &- 2\frac{\alpha^2}{h_n v^2(x)} Cov\left(\int_0^{+\infty} \xi_1(X_1, Y_1, x, y) J(F(y|x)) dy,\right. \\ &\left.\int_0^{+\infty} F(y|x) \xi_2(X_1, Y_1, x) J(F(y|x)) dy\right) \\ =:& \mathscr{C} + \mathscr{L} + \mathscr{T}. \end{aligned} \tag{43}$$

We begin by analyzing term $\mathscr{C}$ as follows

$$\begin{aligned}
\mathscr{C} =&\frac{\alpha^2}{h_n v^2(x)} Var\left(\int\limits_0^{+\infty} \xi_1(X_1, Y_1, x, y) J(F(y|x))dy\right)\\
=&\frac{\alpha^2}{h_n v^2(x)} \mathbf{E}\left(G^{-1}(Y_1) k\left(\frac{x-X_1}{h_n}\right) \int\limits_0^{+\infty} K_0\left(\frac{y-Y_1}{h_n}\right) J(F(y|x))dy\right)^2\\
&-\frac{\alpha^2}{h_n v^2(x)} \mathbf{E}^2\left(G^{-1}(Y_1) k\left(\frac{x-X_1}{h_n}\right) \int\limits_0^{+\infty} K_0\left(\frac{y-Y_1}{h_n}\right) J(F(y|x))dy\right)\\
=:&\, \mathscr{C}_1 + \mathscr{C}_2.
\end{aligned} \tag{44}$$

For $\mathscr{C}_1$, using (6) and a change of variable $\left(u = \frac{x-r}{h_n}\right)$, we get

$$\begin{aligned}
\mathscr{C}_1 =&\frac{\alpha^2}{h_n v^2(x)} \int_{a_F}^{+\infty} \int_{\mathbb{R}} G^{-2}(t) k^2\left(\frac{x-r}{h_n}\right) \left(\int_0^{+\infty} K_0\left(\frac{y-t}{h_n}\right) J(F(y|x))dy\right)^2 dF^\star(r,t)\\
=&\frac{\alpha}{v^2(x)} \int_{\mathbb{R}} k^2(u) \int_{a_F}^{+\infty} G^{-1}(t) \left(\int_0^{+\infty} K_0\left(\frac{y-t}{h_n}\right) J(F(y|x))dy\right)^2 f(x-uh_n, t)dudt.
\end{aligned}$$

Under Assumption (A3), using a first-order Taylor expansion of $f(x-uh_n, y)$ around x, we obtain

$$\mathscr{C}_1 = \frac{\alpha}{v^2(x)} \int_{\mathbb{R}} k^2(u)du \int_{a_F}^{+\infty} \left(\int_0^{+\infty} K_0\left(\frac{y-t}{h_n}\right) J(F(y|x))dy\right)^2 \frac{f(x,t)}{G(t)} dt + O(h_n).$$

Note that, by Assumption A2(ii), $K_0(\cdot)$ is a continuous distribution function, and therefore,

$$K_0\left(\frac{y-t}{h_n}\right) \to \begin{cases} 1, & \text{if } \ y > t;\\ 0, & \text{if } \ y \le t. \end{cases} \quad \text{as } n \to +\infty.$$

Hence, by the Dominated Convergence Theorem, it yields

$$\begin{aligned}
\mathscr{C}_1 \to&\frac{\alpha}{v^2(x)} \int_{\mathbb{R}} k^2(u)du \int_{a_F}^{+\infty} \left(\int_t^{+\infty} J(F(y|x))dy\right)^2 \frac{f(x,t)}{G(t)} dt + o(1)\\
=&\frac{\alpha}{v^2(x)} \kappa \int_{a_F}^{+\infty} \left(\int_0^{+\infty} J(F(y|x))dy - \int_0^t J(F(y|x))dy\right)^2 \frac{f(x,t)}{G(t)} dt + o(1).
\end{aligned} \tag{45}$$

For $\mathscr{C}_2$, using the fact that $K_0(\cdot)$ is a distribution function and by the same change of variable, we get

$$\begin{aligned}
\mathscr{C}_2 =&\frac{\alpha^2}{h_n v^2(x)} \left[\alpha^{-1} \int_{a_F}^{+\infty} \int_{\mathbb{R}} k\left(\frac{x-r}{h_n}\right) \left(\int_0^{+\infty} K_0\left(\frac{y-t}{h_n}\right) J(F(y|x))dy\right) f(r,t)drdt\right]^2\\
\le&\frac{1}{h_n v^2(x)} \left[\left(\int_0^{+\infty} J(F(y|x))dy\right) \int_{\mathbb{R}} k\left(\frac{x-r}{h_n}\right) \int_{a_F}^{+\infty} f(r,t)dtdr\right]^2\\
\le&\frac{1}{h_n v^2(x)} \left[h_n \left(\int_0^{+\infty} J(F(y|x))dy\right) \int_{\mathbb{R}} k(u) \int_{a_F}^{+\infty} f(x-uh_n, t)dudt\right]^2,
\end{aligned}$$

using again a change of variable $s = F(y|x)$ and a Taylor expansion, we get

$$
\begin{aligned}
\mathscr{C}_2 \leq & \frac{h_n}{v^2(x)}\left(\left(\int_0^1 \frac{J(s)ds}{\inf\limits_{x\in\Omega}\inf\limits_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x)}\right)\int_{a_F}^{+\infty}\int_{\mathbb{R}} k(u)f(x-uh_n,t)dudt\right)^2 \\
\leq & \frac{h_n}{v^2(x)}\left(\int_0^1\left(\inf\limits_{x\in\Omega}\inf\limits_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x)\right)^{-1} J(s)ds\int_{a_F}^{+\infty} f(x,t)dt + O(h_n)\right)^2 \\
= & O(h_n),
\end{aligned} \tag{46}
$$

under Assumptions (A2)(i), (A3), (M2) and (M3).
Next, we examine term $\mathscr{L}$ as shown bellow

$$
\begin{aligned}
\mathscr{L} = & \frac{\alpha^2}{h_n v^2(x)} Var\left(\int_0^{+\infty} F(y|x)\xi_2(X_1,Y_1,x)J(F(y|x))dy\right) \\
= & \frac{\alpha^2}{h_n v^2(x)}\mathbf{E}\left(G^{-1}(Y_1)k\left(\frac{x-X_1}{h_n}\right)\int_0^{+\infty} F(y|x)J(F(y|x))dy\right)^2 \\
& - \frac{\alpha^2}{h_n v^2(x)}\mathbf{E}^2\left(G^{-1}(Y_1)k\left(\frac{x-X_1}{h_n}\right)\int_0^{+\infty} F(y|x)J(F(y|x))dy\right) \\
= & : \mathscr{L}_1 + \mathscr{L}_2.
\end{aligned} \tag{47}
$$

For $\mathscr{L}_1$, using (6) and a change of variable $\left(u = \frac{x-r}{h_n}\right)$, we get

$$
\begin{aligned}
\mathscr{L}_1 = & \frac{\alpha^2}{h_n v^2(x)}\int_{a_F}^{+\infty}\int_{\mathbb{R}} G^{-2}(t)k^2\left(\frac{x-r}{h_n}\right)\left(\int_0^{+\infty} F(y|x)J(F(y|x))dy\right)^2 dF^{\star}(r,t) \\
= & \frac{\alpha}{v^2(x)}\int_{\mathbb{R}} k^2(u)\int_{a_F}^{+\infty} G^{-1}(t)\left(\int_0^{+\infty} F(y|x)J(F(y|x))dy\right)^2 f(x-uh_n,t)dudt.
\end{aligned}
$$

Under Assumption (A3), using a first-order Taylor expansion of $f(x-uh_n,y)$ around x, we obtain

$$
\begin{aligned}
\mathscr{L}_1 = & \frac{\alpha}{v^2(x)}\int_{\mathbb{R}} k^2(u)du\int_{a_F}^{+\infty} G^{-1}(t)\left(\int_0^{+\infty} F(y|x)J(F(y|x))dy\right)^2 \frac{f(x,t)}{G(t)}dt + O(h_n) \\
= & \frac{\alpha}{v^2(x)}\int_{\mathbb{R}} k^2(u)du\left(\int_0^{+\infty} F(y|x)J(F(y|x))dy\right)^2\int_{a_F}^{+\infty}\frac{f(x,t)}{G(t)}dt + O(h_n), \\
\rightarrow & \frac{\alpha}{v^2(x)}\kappa\left(\int_0^{+\infty} F(y|x)J(F(y|x))dy\right)^2\int_{a_F}^{+\infty}\frac{f(x,t)}{G(t)}dt + o(1).
\end{aligned} \tag{48}
$$

For $\mathscr{L}_2$, since $F(y|x)$ is a distribution function, it is bounded by 1 and by the same change of variable, we get

$$
\begin{aligned}
\mathscr{L}_2 = & \frac{\alpha^2}{h_n v^2(x)}\left[\alpha^{-1}\int_{a_F}^{+\infty}\int_{\mathbb{R}} k\left(\frac{x-r}{h_n}\right)\left(\int_0^{+\infty} F(y|x)J(F(y|x))dy\right) f(r,t)drdt\right]^2 \\
\leq & \frac{1}{h_n v^2(x)}\left[\left(\int_0^{+\infty} J(F(y|x))dy\right)\int_{\mathbb{R}} k\left(\frac{x-r}{h_n}\right)\int_{a_F}^{+\infty} f(r,t)dtdr\right]^2 \\
\leq & \frac{h_n}{v^2(x)}\left[\left(\int_0^{+\infty} J(F(y|x))dy\right)\int_{\mathbb{R}} k(u)\int_{a_F}^{+\infty} f(x-uh_n,t)dudt\right]^2.
\end{aligned}
$$

Using again a change of variable $s = F(y|x)$ and a first-order Taylor expansion of $f(x-uh_n, y)$, we obtain

$$
\begin{aligned}
\mathscr{L}_2 \leq & \frac{h_n}{v^2(x)} \left(\left(\int_0^1 \frac{J(s)ds}{\inf\limits_{x\in\Omega} \inf\limits_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x)} \right) \int_{a_F}^{+\infty} \int_{\mathbb{R}} k(u) f(x-uh_n, t) du dt \right)^2 \\
\leq & \frac{h_n}{v^2(x)} \left(\int_0^1 \left(\inf_{x\in\Omega} \inf_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x) \right)^{-1} J(s)ds \int_{a_F}^{+\infty} f(x,t)dt + O(h_n) \right)^2 \\
= & O(h_n),
\end{aligned} \tag{49}
$$

under Assumptions (A2)(i), (A3), (M2) and (M3).
Finally, we analyze term $\mathscr{T}$, which can be written as

$$
\begin{aligned}
\mathscr{T} = & -2\frac{\alpha^2}{h_n v^2(x)} Cov\left(\int_0^{+\infty} \xi_1(X_1, Y_1, x, y) J(F(y|x))dy, \int_0^{+\infty} F(y|x)\xi_2(X_1, Y_1, x) J(F(y|x))dy \right) \\
= & -2\frac{\alpha^2}{h_n v^2(x)} \mathbf{E}\left(\left\{ G^{-1}(Y_1) k\left(\frac{x-X_1}{h_n}\right) \int_0^{+\infty} K_0\left(\frac{y-Y_1}{h_n}\right) J(F(y|x))\, dy \right\} \times \right. \\
& \left. \left\{ G^{-1}(Y_1) k\left(\frac{x-X_1}{h_n}\right) \int_0^{+\infty} F(y|x) J(F(y|x))\, dy \right\} \right) \\
& + 2\frac{\alpha^2}{h_n v^2(x)} \mathbf{E}\left(G^{-1}(Y_1) k\left(\frac{x-X_1}{h_n}\right) \int_0^{+\infty} K_0\left(\frac{y-Y_1}{h_n}\right) J(F(y|x))\, dy \right) \times \\
& \mathbf{E}\left(G^{-1}(Y_1) k\left(\frac{x-X_1}{h_n}\right) \int_0^{+\infty} F(y|x) J(F(y|x))\, dy \right) \\
= & :\mathscr{T}_1 + \mathscr{T}_2.
\end{aligned} \tag{50}
$$

For $\mathscr{T}_1$, by following the same steps and using the same argument as for term $\mathscr{C}_1$, we obtain

$$
\begin{aligned}
\mathscr{T}_1 = & \frac{-2\alpha}{h_n v^2(x)} \int_{a_F}^{+\infty} \int_{\mathbb{R}} G^{-1}(t) k^2\left(\frac{x-r}{h_n}\right) \left(\int_0^{+\infty} K_0\left(\frac{y-t}{h_n}\right) J(F(y|x))dy \right) \\
& \times \left(\int_0^{+\infty} F(y|x) J(F(y|x))dy \right) f(r,t) dr dt \\
\to & \frac{-2\alpha}{v^2(x)} \int_{\mathbb{R}} k^2(u)du \left(\int_0^{+\infty} F(y|x) J(F(y|x))dy \right) \int_{a_F}^{+\infty} \left(\int_t^{\infty} J(F(y|x))dy \right) \frac{f(x,t)}{G(t)} dt + o(1) \\
= & \frac{-2\alpha}{v^2(x)} \kappa \left(\int_0^{+\infty} F(y|x) J(F(y|x))dy \right) \int_{a_F}^{+\infty} \left(\int_0^{+\infty} J(F(y|x))dy - \int_0^t J(F(y|x))dy \right) \\
& \times \frac{f(x,t)}{G(t)} dt + o(1).
\end{aligned} \tag{51}
$$

For $\mathscr{T}_2$, the same steps and argument as for $\mathscr{C}_2$ and $\mathscr{L}_2$, yields

$$
\begin{aligned}
\mathscr{T}_2 \leq & h_n \frac{1}{v^2(x)} \left(\int_0^1 \left(\inf_{x\in\Omega} \inf_{s_0\leq s\leq s_1} f(F^{-1}(s|x)|x) \right)^{-1} J(s)ds \int_{a_F}^{+\infty} f(x,t)dt \right)^2 \\
= & O(h_n).
\end{aligned} \tag{52}
$$

Combining (44)-(52) with (43) we get the result. □

Now to state the asymptotic normality it remains to prove that $\sum_{i=1}^{n} M_{ni}(x) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Delta^2(x))$.
We will prove that $\sum_{i=1}^{n} M_{ni}(x)$ satisfies the Lindeberg's Theorem.

Lemma 3. *Under Assumptions (A1)-(A3) and (M1)-(M3), we have*

$$\forall \varepsilon > 0, \ \sum_{i=1}^{n} \mathbf{E}\left[M_{ni}^2(x) \mathbb{1}_{\left\{ M_{ni}^2(x) > \varepsilon^2 \mathrm{Var}\left(\sum\limits_{i=1}^{n} M_{ni}(x)\right)\right\}}\right] \longrightarrow 0, \ as \ n \ \to +\infty.$$

Proof. Clearly, we have by Lemma 2

$$Var\left(\sum_{i=1}^{n} M_{ni}(x)\right) \to \Delta^2(x), \ as \ n \to +\infty,$$

therefore for n large enough, we have

$$\left\{ M_{ni}^2(x) > \varepsilon^2 Var\left(\sum_{i=1}^{n} M_{ni}(x)\right)\right\} \subset \left\{ M_{ni}^2(x) > \varepsilon^2 \Delta^2(x)\right\}.$$

On the one hand, we have

$$\begin{aligned} M_{ni}^2(x) &= \frac{1}{nh_n}\left\{ \int\limits_0^{+\infty} \xi_1(X_i, Y_i, x, y) J(F(y|x)) dy - \int\limits_0^{+\infty} F(y|x) \xi_2(X_i, Y_i, x) J(F(y|x)) dy \right\}^2 \\ &\leq \frac{2}{nh_n}\left\{ \left(\int\limits_0^{+\infty} \xi_1(X_i, Y_i, x, y) J(F(y|x)) dy \right)^2 + \left(\int\limits_0^{+\infty} F(y|x) \xi_2(X_i, Y_i, x) J(F(y|x)) dy \right)^2 \right\} \\ &=: O\left(\frac{1}{nh_n}\right), \end{aligned}$$

from the expressions of ξ_1 and ξ_2, under Assumptions (A2), (M2), (M3) and the fact that $G(Y) > G(a_F) > 0$.
On the other hand, Assumption (A1)(i) implies that $nh_n \to \infty, \ as \ n \to \infty$, then we obtain

$$M_{ni}^2(x) \to 0, \ as \ n \to \infty.$$

Hence for n large enough, we have

$$\left\{ M_{ni}^2(x) > \varepsilon^2 Var\left(\sum_{i=1}^{n} M_{ni}(x)\right)\right\} = \emptyset,$$

and then

$$\mathbb{1}_{\left\{ M_{ni}^2(x) > \varepsilon^2 Var\left(\sum\limits_{i=1}^{n} M_{ni}(x)\right)\right\}} = 0.$$

This completes the proof of Lemma 3. □

Therefore the proof of Theorem 4 is finished. □

7. Conclusion and discussion

In this paper, we investigated a nonparametric heteroscedastic regression model in the presence of randomly left-truncated data. We proposed kernel-based estimators for both the location and scale functions, constructed using L-type functionals and relying on the conditional distribution function estimator introduced by Lemdani *et al.* (2009). We established the uniform convergence rates, asymptotic representations, and asymptotic normality of the proposed estimators.

The finite-sample performance of the regression estimator was evaluated through extensive simulation studies as well as an application to real data. The results show that, under heteroscedasticity, the proposed estimator outperforms the Nadaraya-Watson type estimator and exhibits increased robustness to outliers.
The theoretical results are derived under the assumption of independence between the truncation variable T and the pair (X,Y). This assumption may be restrictive in certain practical situations. A natural extension of this work would be to relax this condition by considering conditional independence between Y and T given X, or more generally, allowing dependence between Y and T conditional on X.
Finally, extending the proposed methodology to settings involving high-dimensional covariates may lead to a degradation in estimator performance due to the curse of dimensionality, along with practical challenges related to the selection of the bandwidth vector. Addressing these limitations constitutes an interesting direction for future research.

Acknowledgements

The authors wish to express their gratitude to the anonymous reviewer for her/his insightful comments and suggestions, which have significantly improved and enhanced an earlier version of this work.

References

Akritas MG, Van Keilegom I (2001). "Non-Parametric Estimation of the Residual Distribution." *Scandinavian Journal of Statistics*, **28**(3), 549–567. doi:10.1111/1467-9469.00254.

Benseradj H, Guessoum Z (2020). "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**(7), 2082–2115. doi:10.1080/03610926.2020.1764037.

Beran R (1981). *Nonparametric Regression with Randomly Censored Survival Data*. Berkeley: Technical Report, University of California. URL https://www.jstor.org/stable/4616062.

Box G, Meyer RD (1985). "Some New Ideas in the Analysis of Screening Designs." *Journal of Research of the National Bureau of Standards*, **90**(6), 495–500. doi:10.6028/jres.090.048.

Box GEP, Draper NR (1987). *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York. URL https://psycnet.apa.org/record/1987-97236-000.

Chen J, Chao MT, Lo SH (1995). "On Strong Uniform Consistency of the Lynden-Bell Estimator for Truncation Data." *The Annals of Statistics*, **23**(2), 440–449. doi:10.1214/aos/1176324529.

Chen S, Lu X, Zhou X, Zhou Y (2018). "Nonparametric Identification and Estimation of Truncated Regression Models with Heteroskedasticity." *Econometric Theory*, **34**(3), 543–573. doi:10.1017/S0266466617000111.

Chown J, Heuchenne C, Van Keilegom I (2020). "The Nonparametric Location-Scale Mixture Cure Model." *Test*, **29**(4), 1008–1028. doi:10.1007/s11749-019-00698-8.

Daye ZJ, Chen J, Li H (2012). "High-Dimensional Heteroscedastic Regression with an Application to eQTL Data Analysis." *Biometrics*, **68**(1), 316–326. doi:10.1111/j.1541-0420.2011.01652.x.

Fan J, Gijbels I (1994). "Censored Regression: Local Linear Approximations and Their Applications." *Journal of the American Statistical Association*, **89**(426), 560–570. doi:10.2307/2290859.

Gheliem A, Guessoum Z (2022). "Simulating the Behavior of a Kernel M-Estimator for Left-Truncated and Associated Model." *Communications in Statistics-Simulation and Computation*, **53**(5), 2366–2388. doi:10.1080/03610918.2022.2075897.

Guessoum Z, Hamrani F (2017). "Convergence Rate of the Kernel Regression Estimator for Associated and Truncated Data." *Journal of Nonparametric Statistics*, **29**(2), 425–446. doi:10.1080/10485252.2017.1303059.

Härdle W, Janssen P, Serfling R (1988). "Strong Uniform Consistency Rates for Estimators of Conditional Functionals." *The Annals of Statistics*, **16**(4), 1428–1449. doi:10.1214/aos/1176351047.

He S, Yang G (1998). "Estimation of the Truncation Probability in the Random Truncation Model." *The Annals of Statistics*, **26**(3), 1011–1027. doi:10.1214/aos/1024691086.

Heuchenne C, Van Keilegom I (2010). "Estimation in Nonparametric Location-Scale Regression Models with Censored Data." *Annals of the Institute of Statistical Mathematics*, **62**, 439–463. doi:10.1007/s10463-009-0219-3.

Hössjer O, Karlsson M (2023). "On the Use of L-Functionals Regression Models." *Open Mathematics*, **21**(1), 20220597. doi:10.1515/math-2022-0597.

Kuznietsova N, Bidyuk P (2020). "Heteroskedasticity Models for Financial Processes: Modelling and Forecasting." In *Proceedings of the IEEE Third International Conference on Data Stream Mining & Processing (DSMP)*, pp. 310–315. doi:110.1109/dsmp47368.2020.9204313.

Lemdani M, Ould-Saïd E (2007). "Asymptotic Behavior of the Hazard Rate Kernel Estimator under Truncated and Cencored Data." *Communications in Statistics-Theory and Methods*, **36**(1), 155–173. doi:10.1080/03610920600966506.

Lemdani M, Ould-Saïd E, Poulin N (2009). "Asymptotic Properties of a Conditional Quantile Estimator with Randomly Truncated Data." *Journal of Multivariate Analysis*, **100**(3), 546–559. doi:10.1016/j.jmva.2008.06.004.

Liang HY, Li D, Qi Y (2009). "Strong Convergence in Nonparametric Regression with Truncated Dependent Data." *Journal of Multivariate Analysis*, **100**(1), 162–174. doi:10.1016/j.jmva.2008.04.002.

Lim C, Sen PK, Peddada SD (2010). "Statistical Inference in Nonlinear Regression under Heteroscedasticity." *Sankhya B*, **72**, 202–218. doi:10.1007/s13571-011-0013-0.

Lynden-Bell D (1971). "A Method of Allowing for Known Observational Selection in Small Samples Applied to 3CR Quasars." *Monthly Notices Royal Astronomy Society*, **155**(1), 95–118. doi:10.1093/mnras/155.1.95.

Ould-Saïd E, Lemdani M (2006). "Asymptotic Properties of a Nonparametric Regression Function Estimator with Randomly Truncated Data." *Monthly Notices Royal Astronomy Society*, **58**, 357–378. doi:10.1007/s10463-005-0011-y.

Serfling RJ (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York. doi:10.1002/9780470316481.

Stute W (1993). "Almost Sure Representations of the Product-Limit Estimator for Truncated Data." *The Annals of Statistics*, **21**(1), 146–156. doi:10.1214/aos/1176349019.

Sujika A, Van Keilegom I (2018). "The Copula-Craphic Estimator in Censored Nonparametric Location-Scale Regression Models." *Econometrics and Statistics*, **7**, 89–114. doi:10.1016/j.ecosta.2017.07.002.

Van Keilegom I, Akritas MG (1999). "Transfer of Tail Information in Censored Regression Models." *The Annals of Statistics*, **27**(5), 1745–1784. doi:10.1214/aos/1017939150.

Wang JF, Liang HY (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. doi:10.1016/j.jkss.2011.12.001.

Woodroofe M (1985). "Estimating a Distribution Function with Truncated Data." *The Annals of Statistics*, **13**(1), 163–177. doi:10.1214/aos/1176346584.

Zhou T, Zhu L (2015). "Conditional Median Absolute Deviation." *Journal of Statistical Computation and Simulation*, **85**(10), 2101–2114. doi:10.1080/00949655.2014.922185.

Affiliation:

Saloua Aliouche
Lab. MSTD, Faculty of Mathematics
University of Science and Technology Houari Boumediene
BP 32, El-Alia 16111, Algeria
E-mail: saliouche@usthb.dz

Ourida Sadki
Lab. MSTD, Faculty of Mathematics
University of Science and Technology Houari Boumediene
BP 32, El-Alia 16111, Algeria
E-mail: osadki@usthb.dz

Hassiba Benseradj
Faculty of Sciences
University of Boumerdes UMBB
Avenue de l'indépendence, Boumerdes, Algeria
E-mail: h.benseradj@univ-boumerdes.dz

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 55 *Submitted:* 2025-08-04
2026 *Accepted:* 2026-03-19