Mathematics

http://www.aimspress.com/journal/Math

AIMS Mathematics, 8(6): 13000–13023.
DOI: 10.3934/math.2023655
Received: 09 January 2023
Revised: 17 February 2023
Accepted: 22 February 2023
Published: 03 April 2023

Research article

Robust kernel regression function with uncertain scale parameter for high dimensional ergodic data using k-nearest neighbor estimation

Fatimah Alshahrani[1], **Wahiba Bouabsa**[2], **Ibrahim M. Almanjahie** [3,*] **and Mohammed Kadi Attouch**[2]

[1] Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

[2] Laboratory of Statistics and Stochastic Processes, University of Djillali Liabes BP 89, Sidi Bel Abbes 22000, Algeria

[3] Department of Mathematics, College of Science, King Khalid University, Abha 62223, Saudi Arabia

* **Correspondence:** Email: imalmanjahi@kku.edu.sa.

Abstract: In this paper, we consider a new method dealing with the problem of estimating the scoring function γ_a, with a constant a, in functional space and an unknown scale parameter under a nonparametric robust regression model. Based on the k Nearest Neighbors (kNN) method, the primary objective is to prove the asymptotic normality aspect in the case of a stationary ergodic process of this estimator. We begin by establishing the almost certain convergence of a conditional distribution estimator. Then, we derive the almost certain convergence (with rate) of the conditional median (scale parameter estimator) and the asymptotic normality of the robust regression function, even when the scale parameter is unknown. Finally, the simulation and real-world data results reveal the consistency and superiority of our theoretical analysis in which the performance of the kNN estimator is comparable to that of the well-known kernel estimator, and it outperforms a nonparametric series (spline) estimator when there are irrelevant regressors.

Keywords: asymptotic normality; functional data analysis; robust regression estimation; k-NN method; financial data; kernel method
Mathematics Subject Classification: 62H12, 62G07, 62G35, 62G20

1. Introduction

Nonparametric regression using kernel methods is a well-known technique for examining the underlying relationship between response variables and covariates. In research including functional data, estimators based on this technique are provided by [1, 2]. Similar to the estimator utilized in the parametric method, the kernel estimator might be affected by outliers; hence, robustness is necessary.

Let $(A_i, B_i)_{i=1,\dots,n}$ represent a set of strictly stationary dependent random variables generated by (A, B). The latter is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ represents a semi-metric space and d signifies a semi-metric, we denote by $\Upsilon(a)$ the unique solution of $\Gamma(a, x, \delta)$. In this study, we aim to investigate the nonparametric estimate of the robust regression $\Upsilon(a)$ when the scale parameter is unidentified and strong dependencies occur (ergodicity). In the following equation,

$$\Gamma(a, x, \delta) = \mathbb{E}\left[\gamma_a\left(\frac{B-x}{\delta}\right) \mid A = a\right] = 0, \tag{1.1}$$

$\gamma(a)$ is defined as 0 with regard to the parameter x for each $a \in \mathcal{F}$. The δ is a robust measure of conditional scale, and γ_a is a real-valued function that satisfies a number of regularity constraints, as detailed below. In what follows, we assume that the robust regression Γ exists and is unique for all $a \in \mathcal{F}$ (see, for example, [3]).

It is essential to keep in mind that robust regression modeling is an age-old statistical problem. In his 1964 debut publication, [4] investigated the estimation of a location parameter. There are a number of conclusions concerning multivariate time series that are cited in the works of [5–7] (ergodicity, mixing conditions, or Bayesian robustness). In functional nonparametric statistics, robust regression is intensively explored. In fact, [8] introduced it for the first time in 2008. In the case of independent and identically distributed variables, they established this model's complete and absolute convergence. Other robust nonparametric functional regression studies have been conducted since their work. The work of [9] is important, and the references therein on this topic. It is important to note that all these results were achieved with the scale parameter set to a fixed value. In this study, we look at a more general scenario in which the scale is unknown, and the data are dependent (for a state-of-the-art and some discussions on this topic, see, for example, [10] and references therein).

In contrast to the mixing requirement typically used in functional time series research, the ergodicity assumption is more flexible. Specifically, it is present in the majority of mixing circumstances. Few studies have been conducted on time series data with an ergodic functional. Among the limited results are those of ([11–15]).

We simply mention the overviews of parametric models presented by [16, 17], as well as the monographs of [2, 18] as important contributions to the non-parametric model, among the literature review on functional data analysis.

In this paper, as mentioned early, we consider the case in which the scale is unknown. In this sense, we extend the result found by [10] from the independent case to the ergodic case, using the k Nearest Neighbors (kNN) approach and rely mainly on [19] study and also the studies of [20, 21]. This study was carried out under standard conditions, allowing us to examine the subject's multiple structural axes, including the robustness of the regression function and the correlation between the observations. In this case, it is important to note that it is necessary to estimate the scale parameter, which makes it harder to find asymptotic properties compared to the case of a fixed scale. Then, the objective of this contribution is to estimate this robust regression model using the kNN-method. It is an alternative smoothing technique that permits an estimate of the bandwidth parameter of the robust regression operator based on data.

The distance between the functional random variables precisely determines the bandwidth parameter. The kNN algorithm permits the exploration of the data's topological and spectral components. In recent years, the functional kNN smoothing method has garnered great interest due to its sophisticated bandwidth selection mechanism. Burba et al. [22] developed the first results in this subject; they offered a convergence rate that was nearly consistent. Attouch and Bouabca [23, 24] used the same methods to establish the almost complete consistency of the conditional mode estimator, and [25] deduced the conditional hazard function. Recent developments in the field are listed by [26]; we refer to [20, 21] for the most up-to-date developments and references.

The study is structured as follows: Section 2 presents the robust estimator with an unknown scale parameter. In Section 3, we list the necessary assumptions and notations. In Section 4, we give the results with their proofs. Sections 5 and 6 are devoted to using simulations and real data to prove the efficiency of the estimators.

2. The equivariant robust estimators with the k Nearest Neighbors estimation

Consider the ergodic stationary functional process $M_i = (A_i, B_i)_{i=1,\dots,n}$ (for various definitions and examples, see [12]). The robust estimator can be built using the two steps if the scale parameter is unknown. The scale parameter δ is initially estimated via the local median of the absolute deviation from the conditional median (MED), $\widehat{m}_{MED}(a)$ of the conditional distribution of B given $A = a$, denoted $F(b|A = a) = \mathbb{E}\left(\mathbb{1}_{(-\infty,b]}(B)|A = a\right)$, for any $b \in \mathbb{R}$, where $\mathbb{1}_D$ represents the indicator function on the set D. Thus, the kernel estimator $\hat{t}$ of $t(a)$ is the zero of the equation that follows for $a \in \mathcal{F}$:

$$\widehat{F}(t|A = a) = \frac{1}{2},$$

where, $\widehat{F}(b|A = a)$ is given by

$$\widehat{F}(b|A = a) = \frac{\sum_{i=1}^{n} Z\left(h^{-1}d(a, A_i)\right) \mathbb{1}_{(-\infty,b]}(B_i)}{\sum_{i=1}^{n} Z\left(h^{-1}d(a, A_i)\right)}, \tag{2.1}$$

where Z is a kernel function and $h = h_n$ is a series of positive real numbers that tend to zero when n tends to infinity. Then, the classical estimator $\check{\Upsilon}(a)$ of $\Upsilon(a)$ is the zero, with respect to l, of the equation

$$\check{\Gamma}(a, l, \widehat{t}) = 0 \quad \text{with} \quad \check{\Gamma}(a, l, \widehat{t}) = \frac{\sum_{i=1}^{n} Z\left(h^{-1}d(a, A_i)\right) \gamma_a\left(\frac{B_i - l}{\widehat{t}}\right)}{\sum_{i=1}^{n} Z\left(h^{-1}d(a, A_i)\right)}. \tag{2.2}$$

This estimator's asymptotic properties have been examined by [27]. Alternatively, in this paper, we focus on the asymptotic properties of the kNN estimator of the robust kernel regression function with uncertain scale parameter for which the scalar bandwidth parameter h is replaced by a random sequence of positive real integers defined by $H_{n,k}(a) = \min\left\{u_n \in \mathbb{R}^+ : \sum_{i=1}^{n} \mathbb{1}_{B(a,u_n)}(a_i) = k\right\}$. So, our main estimator $\hat{\Upsilon}(a)$ of the robust estimator $\Upsilon(a)$ is the zero, with respect to l, of the following equation

$$\widehat{\Gamma}(a, l, \widehat{t}) = 0 \quad \text{with} \quad \widehat{\Gamma}(a, l, \widehat{t}) = \frac{\sum_{i=1}^{n} Z\left(H_{n,k}(a)^{-1}d(a, A_i)\right) \gamma_a\left(\frac{B_i - l}{\widehat{t}}\right)}{\sum_{i=1}^{n} Z\left(H_{n,k}(a)^{-1}d(a, A_i)\right)}. \tag{2.3}$$

3. Notations, hypotheses, and comments

In this work, some strictly positive generic constants will be represented by C and C' where there is no possibility of error. In $\mathcal{F}$, a is a fixed point, and N_a designates the fixed area around a. Let consider $\mathcal{B}(a,s) := \{a' \in \mathcal{F}/d(a',a) < s\}$, for $s > 0$. In addition, we propose that the σ-field produced by $((A_1, B_1), \ldots, (A_{\mathcal{K}}, B_{\mathcal{K}}), A_{\mathcal{K}+1})$ is known as $\mathfrak{B}_{\mathcal{K}}$. We take the following assumptions into account:

(AS$_1$): γ_a's function is monotone and continuous with respect to the second component.

(AS$_2$): The processes $(A_i, B_i)_{i\in N}$ satisfies:

i) $\sigma(a,s) = \mathbb{P}(A \in B(a,s)) > 0$ and
$\sigma_i(a,s) = \mathbb{P}(A_i \in B(a,s)|\mathcal{F}_{i-1}) > 0 \; \forall s > 0$.

ii) For all $s > 0$, $\dfrac{1}{n\sigma(a,s)} \sum_{i=1}^{n} \sigma_i(a,s) \to^p 1$ *and*
$n\sigma(a,s) \to \infty$ *as* $h \to 0$.

(AS$_3$): The function Γ is as follows:

i) $\Gamma(a, ., \delta)$ is of class C^1 in $\mathcal{N}_a$, a fixed neighborhood of $\Upsilon(a)$.

ii) For any fixed r in $\mathcal{N}_a$ the functions $\Gamma(., r.\delta)$ and

$$\Lambda_2(., r, \delta) = \mathbb{E}[\gamma_a^2(\frac{B-r}{\delta})|A = .],$$

are continuous at a.

iii) The derivative of the real function

$$\Theta(a,u,\delta) = \mathbb{E}[\Gamma(A_1,u,\delta) - \Gamma(a,u,\delta)|d(a,A_1) = t],$$

exists at $t = 0$ and is continuous in $\mathcal{N}_a$ second component.

(AS$_4$): For any fixed r in the neighborhood of $\Upsilon(a)$ and $\forall j \geq 2$,

$$\mathbb{E}\left[\gamma_a^j\left(\frac{B-r}{\delta}\right)|\mathfrak{B}_{i-1}\right] = \mathbb{E}\left[\gamma_a^j\left(\frac{B-r}{\delta}\right)|A_i\right] < c < \infty, \; a.s.$$

(AS$_5$): The kernel $Z(\cdot)$ is a positive function supported on $[0,1]$ and is a differentiable function on $]0,1[$ with derivative $Z'(\cdot)$ such that

$$-\infty < C < Z'(\cdot) < C' < 0.$$

(AS$_6$): There exists a function $\vartheta_a(.)$ such that

$$\forall r \in [0,1] \; \lim_{h\to 0} \frac{\sigma(a,rh)}{\sigma(a,h)} = \vartheta_a(r),$$

$$Z^2(1) - \int_0^1 (Z^2(v))' \, \vartheta_a(v)dv > 0,$$

and

$$Z(1) - \int_0^1 Z'(v) \, \vartheta_a(v)dv \neq 0.$$

(AS$_7$):

i) In the neighborhood of $\mathcal{S}$, $F(b|A = a)$ is a continuous function of a. It also fulfills the following equicontinuity condition:

$$\forall\, \epsilon > 0, \exists \mu > 0 : |v - w| < \mu \Longrightarrow |F(v|A = a) - F(w|A = a)| < \epsilon.$$

ii) $F(b|A = a)$ is symmetric around $\Upsilon(a)$ and a continuous function of b for each fixed a.

Remark Convexity and boundedness are well-known properties of robust functions and are present in the scoring function. The existence and uniqueness of the estimate are dependent on the primer, as is its usefulness in minimizing the effects of atypical data. This study employs the monotonicity criterion **(AS$_1$)** to control convexity. For example, the classical regression has been read in the ergodic process (see [12]), hence we adopted a contribution without the boundedness condition.

Assumptions **(AS$_2$)** and **(AS$_3$)** are the same conditions utilized in [13]. Additionally, the conditions used by [28] are very similar to conditions **(AS$_4$)** and **(AS$_6$)**.

4. Main result

For either kernel or nearest neighbor with kernel estimates, the result in Proposition 4.1 ensures the consistency on a set $\mathcal{S} \in \mathcal{F}$. Following this, the asymptotic normality of the suggested estimator is handled by the Theorem 1.

Proposition 4.1. *Suppose that* ***(AS$_2$)***, ***(AS$_4$)*** *and* ***(AS$_5$)***, *hold. Then, for any set* $\mathcal{S}$,

1) Under assumptions ***(AS$_7$(i))***, *we get*

$$\left|\hat{\Upsilon}(a) - \Upsilon(a)\right| \xrightarrow{a.s} 0.$$

2) Additionally, if, $F(b|A = a)$ *has a unique median at* $\Upsilon(a)$, *then*

$$|\hat{m}_{MED}(a) - \Upsilon(a)| \xrightarrow{a.s} 0.$$

For the sake of shortness, the Proof of this proposition is omitted. It is obtained by combining the classical techniques of [22] to those used by [19].

Theorem 1. *Suppose that* ***(AS$_1$)***–***(AS$_6$)***, *with* ***(AS$_7$.ii)*** *hold. Then, if* $k\sigma^{-1}(a, k/n) \to 0$ *we have*

$$\left(\frac{k}{\delta^2(a, \Upsilon(a))}\right)^{1/2} \left(\hat{\Upsilon}(a) - \Upsilon(a)\right) \xrightarrow{D} \mathcal{N}(0, 1) \;\; as \;\; n \to \infty,$$

where

$$\delta^2(a, \Upsilon(a) = \frac{\varrho_2 \Lambda_2(a, \Upsilon(a), \delta)}{\varrho_1^2(\phi_1(a, \Upsilon(a), \delta))^2},$$

with

$$\varrho_0 = \int_1^0 (tZ(t))' \varrho_a(t) dt,$$

$$\varrho_j = -\int_1^0 \left(Z^j\right)'(t)\varrho_a(t)dt,\ j = 1, 2,$$
$$\phi_1(a, \Upsilon(a), \delta) = \frac{\partial\Gamma(a, r, \delta)}{\partial r}\Big|_{r=\Upsilon(a)},$$
$$\mathcal{M} = \{w \in \mathcal{F}, \Lambda_2(w, \Upsilon(w), \delta)\phi_1(w, \Upsilon(w), \delta)\} \neq 0,$$

signifies the convergence of distributions.

Proof. We prove the increasing case γ_a and then the decreasing case is getting by looking at $-\gamma_a$. To do that, for ζ_n such that $\zeta_n - 1 = o(1)$, we define $x_n := \sigma^{-1}(a, \zeta_n \frac{k}{n})$ and $y_n := \sigma^{-1}(a, \frac{k}{n\zeta_n})$ and

$$\forall v \in \mathbb{R}, \quad u = \Upsilon(a) + v[n\sigma(a, h)]^{1/2}\delta(a, \Upsilon(a)).$$

It is clear that, $\mathbb{1}_{x_n \le h \le y_n} \overset{a.co.}{\longrightarrow} 1$ when $\frac{k}{n} \longrightarrow 0$, so

$$\mathbb{P}\left\{\left(\frac{n\sigma(a, h)}{\delta^2(a, \Upsilon(a))}\right)^{1/2}\left(\hat{\Upsilon}(a) - \Upsilon(a)\right) < v\right\}$$
$$= \mathbb{P}\left\{\hat{\Upsilon}(a) < \Upsilon(a) - +v[n\sigma(a, h)]^{1/2}\delta(a, \Upsilon(a))\right\} = \mathbb{P}\left\{0 < \hat{\Gamma}(a, u, \hat{t})\right\}.$$

We can write

$$\hat{\Gamma}(a, r, \hat{t}) = \mathfrak{B}_n(h, a, r, \hat{t}) + \frac{\mathfrak{R}_n(h, a, r, \hat{t})}{\hat{\Gamma}_D(h, a)} + \frac{\mathfrak{Q}_n(h, a, r, \hat{t})}{\hat{\Gamma}_D(h, a)},$$

with

$$\mathfrak{Q}_n(h, a, r, \hat{t}) = \left(\hat{\Gamma}_N(h, a, r, \hat{t}) - \bar{\Gamma}_N(h, a, r, \hat{t})\right) - \Gamma_N(h, a, r, \hat{t})\left(\hat{\Gamma}_D(h, a) - \bar{\Gamma}_D(h, a)\right),$$
$$\mathfrak{R}_n(h, a, r, \hat{t}) = -\left(\frac{\bar{\Gamma}_N(h, a, r, \hat{t})}{\bar{\Gamma}_D(h, a)} - \Gamma(h, a, r, \hat{t})\right)\left(\hat{\Gamma}_N(h, a, r, \hat{t}) - \bar{\Gamma}_N(h, a, r, \hat{t})\right),$$
$$\mathfrak{B}_n(h, a, r, \hat{t}) = \frac{\bar{\Gamma}_N(h, a, r, \hat{t})}{\bar{\Gamma}_D(h, a)},$$

where

$$\hat{\Gamma}_N(h, a, r, \hat{t}) = \frac{1}{n\mathbb{E}[Z(y_n^{-1}d(a, A_1))]}\sum_{i=1}^{n} Z\left(h^{-1}d(a, A_i)\right)\gamma_a\left(\frac{B_i - r}{\hat{t}}\right),$$
$$\bar{\Gamma}_N(h, a, r, \hat{t}) = \frac{1}{n\mathbb{E}[Z(y_n^{-1}d(a, A_1))]}\sum_{i=1}^{n}\mathbb{E}\left[Z\left(h^{-1}d(a, A_i)\right)\gamma_a\left(\frac{B_i - r}{\hat{t}}\right)/\mathcal{F}_{i-1}\right],$$
$$\hat{\Gamma}_D(h, a) = \frac{1}{n\mathbb{E}[Z(y_n^{-1}d(a, A_1))]}\sum_{i=1}^{n} Z\left(h^{-1}d(a, A_i)\right),$$
$$\bar{\Gamma}_D(h, a) = \frac{1}{n\mathbb{E}[Z(y_n^{-1}d(a, A_1))]}\sum_{i=1}^{n}\mathbb{E}\left[Z\left(h^{-1}d(a, A_i)\right)/\mathcal{F}_{i-1}\right].$$

Consequently,

$$\mathbb{P}\left\{\left(\frac{k\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}\left(\hat{\Upsilon}(a)-\Upsilon(a)\right)<v\right\}=$$
$$\mathbb{P}\left\{-\hat{\Gamma}_D(h,a)\mathfrak{B}_n(h,a,u,\hat{t})-\mathfrak{R}_n(h,a,u,\hat{t})<\mathfrak{Q}_n(h,a,u,\hat{t})\right\}.$$

The following intermediate results lead to our main consequence. □

Lemma 1. *Assume that the assumptions of Theorem 1 holds, then, we obtain, for any $a\in\mathcal{M}$,*

$$\left(\frac{k\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}Q_n(h,a,u,\hat{t})\xrightarrow{D}\mathcal{N}(0,1)\ \ as\ \ n\to\infty.$$

Proof. We write

$$Q_n(h,a,u,\hat{t})=Q_n(y_n,a,u,\hat{t})+(Q_n(h,a,u,\hat{t})-Q_n(y_n,a,u,\hat{t})).$$

As

$$Q_n(y_n,a,u,\hat{t})\le Q_n(h,a,u,\hat{t})\le Q_n(x_n,a,u,\hat{t}),\tag{4.1}$$

then

$$|Q_n(h,a,u,\hat{t})-Q_n(y_n,a,u,\hat{t})|\le|Q_n(x_n,a,u,\hat{t})-Q_n(y_n,a,u,\hat{t})|.$$

So, all it remains is to prove that

$$\left(\frac{k\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}Q_n(y_n,a,u,t)\xrightarrow{D}\mathcal{N}(0,1)\ \ as\ \ n\to\infty.\tag{4.2}$$

$$\left(\frac{k\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}Q_n(y_n,a,u,\hat{t})-Q_n(y_n,a,u,\hat{t})=o_p(1).\tag{4.3}$$

and

$$\left(\frac{k\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}\||Q_n(x_n,a,u,\hat{t})-Q_n(y_n,a,u,\hat{t})|=o_p(1).\tag{4.4}$$

The required result in (4.4) can be deduced from the standard consistency in [13], while (4.3) is similar to the Eq (A. 25) in [27]. Hence, we focus now on (4.2). For this purpose, we put $Z_i(a)=Z\left(h^{-1}d(a,A_i)\right)$, for all $i=1,...,n$, and

$$\tau_{ni}=\left(\frac{\sigma(a,h)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}\left(\gamma_a\left(\frac{B_i-u}{t}\right)-\Gamma(a,u,t)\right)\frac{Z_i(a)}{\mathbb{E}[Z_1(a)]},$$

which determines $\xi_{ni}=\tau_{ni}-\mathbb{E}\left[\tau_{ni}|\mathcal{F}_{i-1}\right]$. Then, we get

$$\left(\frac{n\sigma(a,x_n)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}Q\left(a,u,t\right)=\frac{1}{\sqrt{n}}\sum_{i=1}^{\mathrm{n}}\xi_{ni}.$$

Since ξ_{ni} is a triangular array of martingale differences (according to the $\delta-$ field $\mathcal{F}_{i-1}$) and based on the unconditional Lindeberg condition, we may apply the central limit theorem (see [29]). Specifically, we should investigate the following conditions:

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\xi_{ni}^2|\mathcal{F}_{i-1}] \rightarrow 1 \;\; in \;\; probability, \tag{4.5}$$

and

$$\forall\varepsilon > 0, \;\; \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\xi_{ni}^2 I_{\xi_{ni}^2>\epsilon n}.] \rightarrow 0. \tag{4.6}$$

We start with the proof of (4.5). To do this, we write

$$\mathbb{E}[\xi_{ni}^2|\mathcal{F}_{i-1}] = \mathbb{E}[\tau_{ni}^2|\mathcal{F}_{i-1}] - \mathbb{E}^2[\tau_{ni}|\mathcal{F}_{i-1}].$$

So, it is enough to show that

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}^2[\xi_{ni}|\mathcal{F}_{i-1}] \xrightarrow{P} 0, \tag{4.7}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\xi_{ni}^2|\mathcal{F}_{i-1}] \xrightarrow{P} 1. \tag{4.8}$$

For the first convergence, we have

$$\begin{aligned}|\mathbb{E}[\xi_{ni}|\mathcal{F}_{i-1}]| &= \frac{1}{\mathbb{E}Z_1(a)}\left(\frac{\sigma(a,h)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}|\mathbb{E}[(\Gamma(A_i,r,t)-\Gamma(a,r,t)Z_i(a))|\mathcal{F}_{i-1}]|,\\ &\leq \frac{1}{\mathbb{E}[Z_1(a)]}\left(\frac{\sigma(a,h)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}\sup_{v\in B(a,y_n)}|\Gamma(v,r,t)-\Gamma(a,r,t)|\mathbb{E}[Z_i(a)|\mathcal{F}_{i-1}].\end{aligned}$$

So, under the **($\mathbf{AS}_2$)** and **($\mathbf{AS}_3$.ii))** means that

$$\sup_{v\in B(a,y_n)}|\Gamma(v,r,t)-\Gamma(a,r,t)| = o(1).$$

By combining the last three results, we get

$$\begin{aligned}(|\mathbb{E}[\xi_{ni}|\mathcal{F}_{i-1}]|)^2 &\leq \left|\Gamma(v,r,t)-\Gamma(a,r,t)\left(\frac{\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)\right|\frac{1}{\sigma(a,y_n)}\sigma_i^2(a,y_n),\\ &\leq \left|\Gamma(v,r,t)-\Gamma(a,r,t)\left(\frac{\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)\right|\frac{1}{\sigma(a,y_n)}\sigma_i(a,y_n).\end{aligned}$$

Finally, using the fact that

$$\frac{1}{n\sigma(a,y_n)}\sum_{i=1}^{\mathrm{n}}\sigma_i(a,y_n) \xrightarrow{P} 1,$$

we get

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\xi_{ni}|\mathcal{F}_{i-1}])^2 = \sup_{v\in B(a,h)}|\Gamma(v,r,t)-\Gamma(a,r,t)|\,,$$

$$\left(\frac{\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)\left(\frac{1}{n\sigma(a,y_n)}\sum_{i=1}^{\mathrm{n}}\sigma_i(a,y_n)\right) = o_p(1).$$

Now, we look at convergence in (4.7). We compose

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\xi_{ni}^2|\mathcal{F}_{i-1}]\\
&= \frac{1}{n(\mathbb{E}Z_1(a))^2}\left(\frac{\sigma(a,h)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)\sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[\left(\gamma_a\left(\frac{B_i-u}{t}\right)-\Gamma(a,u,t)\right)^2 Z_i^2(a)|\mathcal{F}_{i-1}\right],\\
&= \frac{1}{n(\mathbb{E}Z_1(a))^2}\left(\frac{\sigma(a,h)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)\left(\sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[\gamma_a^2\left(\frac{B_i-u}{t}\right)\Phi_i^2(a)|\mathcal{F}_{i-1}\right]\right)\\
&- \frac{2\Gamma(a,u,t)}{n(\mathbb{E}Z_1(a))^2}\left(\frac{\sigma(a,h)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)\sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[\gamma_a\left(\frac{B_i-u}{t}\right)\Phi_i^2(a)|\mathcal{F}_{i-1}\right]\\
&+ \frac{1}{n(\mathbb{E}Z_1(a))^2}\left(\frac{\sigma(a,h)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)\Gamma^2(a,u,t)\sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[\Phi_i^2(a)|\mathcal{F}_{i-1}\right].
\end{aligned}$$

Next, set the following variables:

$$\mathcal{D}_1 = \sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[\gamma_a^2\left(\frac{B_i-u}{t}\right)\Phi_i^2(a)|\mathcal{F}_{i-1}\right], \mathcal{D}_2 = \sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[\gamma_a\left(\frac{B_i-u}{t}\right)\Phi_i^2(a)|\mathcal{F}_{i-1}\right],$$

and

$$\mathcal{D}_3 = \sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[\Phi_i^2(a)|\mathcal{F}_{i-1}\right].$$

Then, use mathematical manipulation to write $\mathcal{D}_1$ as:

$$\begin{aligned}
\mathcal{D}_1 =&\Lambda_2(a,u,t)\sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right] + \sum_{i=1}^{\mathrm{n}}\left[\mathbb{E}\left[\gamma_a^2\left(\frac{B_i-u}{t}\right)Z_i^2(a)\right]|\mathcal{F}_{i-1}\right]\\
&- \sum_{i=1}^{\mathrm{n}}\Lambda_2(a,u,t)\mathbb{E}\left[Z_i^2|\mathcal{F}_{i-1}\right],\\
=&\Lambda_2(a,u,t)\sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right] + \sum_{i=1}^{\mathrm{n}}\left[\mathbb{E}\left[Z_i^2(a)\mathbb{E}\left[\gamma_a^2\left(\frac{B_i-u}{t}\right)|\mathfrak{B}_{i-1}\right]|\mathcal{F}_{i-1}\right]\right]\\
&- \sum_{i=1}^{\mathrm{n}}\Lambda_2(a,u,t)\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right],\\
=&\Lambda_2(a,u,t)\sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right] + \sum_{i=1}^{\mathrm{n}}\mathbb{E}\left[Z_i^2(a)\mathbb{E}\left[\gamma_a^2\left(\frac{B_i-u}{t}\right)|A_i\right]|\mathcal{F}_{i-1}\right]
\end{aligned}$$

$$-\sum_{i=1}^{n}\Lambda_2(a,u,t)\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right].$$

We may determine the second term by using

$$\frac{1}{n\mathbb{E}\left[Z_1(a)\right]}\sum_{i=1}^{n}\left[\mathbb{E}\left[Z_i^2(a)\mathbb{E}\left[\gamma_a^2\left(\frac{B_i-u}{t}\right)|A_i\right]|\mathcal{F}_{i-1}\right]-\Lambda_2(a,u,t)\mathbb{E}[Z_i^2(a)|\mathcal{F}_{i-1}]\right]$$
$$\leq \sup_{v\in B(a,h)}|\Lambda_2(a,v,t)-\Lambda_2(a,u,t)|\left(\frac{1}{n\sigma(a,h)}\sum_{i=1}^{n}\mathbb{P}(\sigma(a,h)|\mathcal{F}_{i-1})\right).$$

Next, employ the continuity of Λ_2 to get

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\tau_{ni}^2|\mathcal{F}_{i-1}]=\frac{1}{(n\mathbb{E}\left[Z_1(a)\right])^2}\left(\frac{\sigma(a,y_n)\varrho_1^2}{\varrho_2}\right)\sum_{i=1}^{n}\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right]+o(1).$$

Here, we apply the identical methods described by [30] to

$$\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right]=Z^2(1)\sigma_i(a,y_n)-\int_0^1(Z^2(v))^{'}\sigma_i(a,vy_n)dv$$

and

$$\mathbb{E}\left[Z_1(a)\right]=Z(1)\sigma(a,y_n)-\int_0^1(Z(v))^{'}\sigma_i(a,vy_n)dv.$$

It follows that

$$\frac{1}{n\sigma(a,y_n)}\sum_{i=1}^{n}\mathbb{E}\left[Z_i^2(a)|\mathcal{F}_{i-1}\right]=\frac{Z^2(1)}{n\sigma(a,y_n)}\sum_{i=1}^{n}\sigma_i(a,y_n)$$
$$-\int_0^1(Z^2(v))^{'}\frac{\sigma_i(a,vy_n)}{n\sigma(a,y_n)\sigma(a,vy_n)}\sum_{i=1}^{n}\sigma_i(a,vy_n)dv$$
$$=Z^2(1)-\int_0^1(Z^2(v))^{'}\eta_a(v)dv+o_p(1)=\varrho_2+o_p(1),$$

and

$$\frac{1}{n\sigma(a,y_n)}\mathbb{E}\left[Z_1(a)\right]=\varrho_2+o_p(1).$$

We conclude that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[\tau_{ni}^2|\mathcal{F}_{i-1}]=1,$$

which completes the proof of (4.5). In regards to (4.6), we write

$$\xi_{ni}^2 I_{\xi_{ni}^2>\epsilon n}\leq\frac{|\xi_{ni}|^{2+\mu}}{\sqrt{(\epsilon n)^{\mu}}},\quad\forall\ \mu>0.$$

Note that

$$\mathbb{E}[\xi_{ni}^{2+\mu}]=\mathbb{E}\left[|\tau_{ni}(a)-\mathbb{E}[\tau_{ni}(a)|\mathcal{F}_{i-1}]|^{2+\mu}\right]$$
$$\leq 2^{1+\mu}\mathbb{E}\left[|\tau_{ni}(a)|^{2+\mu}\right]+2^{1+\mu}\left|\mathbb{E}\left[\mathbb{E}[\tau_{ni}(a)|\mathcal{F}_{i-1}]^{2+\mu}\right]\right|.$$

Then, use Jensen's inequality to get

$$\mathbb{E}[\xi_{ni}^{2+\mu}] \le C\mathbb{E}\left[|\tau_{ni}(a)|^{2+\mu}\right].$$

Consequently, it is necessary to evaluate $\mathbb{E}\left[|\tau_{ni}(a)|^{2+\mu}\right]$. To do this, we again employ the C_r-inequality to obtain

$$\begin{aligned}\mathbb{E}\left[|\tau_{ni}(a)|^{2+k}\right] &\le C\left(\frac{\sigma(a,x_n)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\mathbb{E}^2[Z_1(a)]\right)^{1+\mu/2}\mathbb{E}\left[Z_i^{2+\mu}(a)\gamma_a^{2+\mu}\left(\frac{B_i-r}{t}\right)\right]\\ &+\Gamma^{2+\mu}(a,u,t)\mathbb{E}\left[Z_i^{2+\mu}(a)\right].\end{aligned}$$

Now conditioning on A_i and using the fact that

$$\mathbb{E}\left[\gamma_a^{2+\mu}\left(\tfrac{B_i-r}{t}\right)|A_i\right] < \infty,$$

we obtain, with $h \in (x_n, y_n)$, that

$$\mathbb{E}\left[|\tau_{ni}(a)|^{2+\mu}\right] \le C\left(\frac{1}{\sigma(a,x_n)}\right)^{2+\mu/2}\mathbb{E}\left([Z_i(a)]^{2+\mu}\right) \le C\left(\frac{1}{\sigma(a,x_n)}\right)^{\mu/2}.$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[\xi_{ni}^2 I_{\xi_{ni}^2>\epsilon n}\right] \le C\left(\frac{1}{n\sigma(a,x_n)}\right)^{\mu/2} \to 0,$$

completes the proof.

□

Lemma 2. *Assume that the assumptions* **(AS**$_1$**)**, **(AS**$_2$**)** *and* **(AS**$_7$**)** *holds. Then, we get that*

$$\hat{\Gamma}_D(h,a) - 1 = o_p(1).$$

Proof. Observe that $\hat{\Gamma}_D(h,a) - 1 = \mathcal{R}_{1,D}(a) - \mathcal{R}_{2,D}(a)$, where

$$\begin{aligned}\mathcal{R}_{1,D}(h,a) &:= \hat{\Gamma}_D(h,a) - \bar{\Gamma}_D(h,a)\\ &:= \frac{1}{n\mathbb{E}[Z(h^{-1}d(a,A_1))]}\sum_{i=1}^{\mathrm{n}}\left(Z\left(h^{-1}d(a,A_i)\right) - \mathbb{E}[Z(h^{-1}d(a,A_i))|\mathcal{F}_{i-1}]\right),\\ \mathcal{R}_{2,D}(h,a) &:= \hat{\Gamma}_D(h,a) - \mathbb{E}\left(\bar{\Gamma}_D(h,a)\right)\\ &:= \frac{1}{n\mathbb{E}[Z(h^{-1}d(a,A_1))]}\sum_{i=1}^{\mathrm{n}}\left(\mathbb{E}[Z(h^{-1}d(a,A_i))|\mathcal{F}_{i-1}] - \mathbb{E}Z(h^{-1}d(a,A_1))\right),\\ &= \frac{1}{n\mathbb{E}[Z(h^{-1}d(a,A_1))]}\sum_{i=1}^{\mathrm{n}}\mathbb{E}[Z(h^{-1}d(a,A_i))|\mathcal{F}_{i-1}] - 1.\end{aligned}$$

So,

$$\lim_{x_n\le h\le y_n}\mathcal{R}_{2,D}(a) = o_{a.s.}(1).$$

To address the first term $\mathcal{R}_{1,D}(a) = \sum_{i=1}^{\mathrm{n}} L_{ni}(a)$, observe that $L_{ni}(a)$ is a triangular array of martingale differences relative to the δ-field $\mathcal{F}_{i-1}$. Combining the Burkholder ([31], p. 23) and Jensen inequalities yield the following result:

$$\forall\, \varepsilon > 0, \text{ there exists a constant } C_0 > 0 \text{ such that}$$
$$\mathbb{P}(|\mathcal{R}_{1,D}(a)| > \varepsilon) \leq \quad C_0 \frac{\mathbb{E}\left(Z(h^{-1}d(a,A_1))\right)^2}{\varepsilon^2 n\left(\mathbb{E}Z(h^{-1}d(a,A_1))\right)^2} = O\left(\frac{1}{\varepsilon^2 n\sigma(a,x_n)} + o(1)\right).$$

Hence, we conclude that

$$\mathcal{R}_{1,D}(a) = o_{\mathbb{P}}(1) \;\; as \;\; n \to \infty.$$

□

Lemma 3. *Assume that the assumptions* **(AS$_1$), (AS$_2$), (AS$_4$)** *and* **(AS$_7$)** *holds. Then, we get that*

$$\left(\frac{n\sigma(a,x_n)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2} \mathfrak{B}_n(a,u,\hat{t}) = v + o(1) \;\; as \;\; n \to \infty.$$

Proof. We get the following results from a simple manipulation:

$$\begin{aligned}
&\frac{\bar{\Gamma}_N(h,a,u,\hat{t})}{\bar{\Gamma}_D(h,a)} \\
&= \frac{1}{\sum_{i=1}^{n}\mathbb{E}\left[Z_i(a)|\mathcal{F}_{i-1}\right]}\sum_{i=1}^{n}\mathbb{E}\left[Z_i\left[\mathbb{E}\left[\gamma_a\left(\frac{B-u}{\hat{t}}\right)|A_1\right] - \mathbb{E}\left[\gamma_a\left(\frac{B-u}{\hat{t}}\right)|A=a\right]\right]|\mathcal{F}_{i-1}\right] \\
&+ \mathbb{E}\left[\gamma_a\left(\frac{B-u}{\hat{t}}\right)|A=a\right] - \mathbb{E}\left[\gamma_a\left(\frac{B-\Upsilon(a)}{\hat{t}}\right)|A=a\right], \\
&= D_1(a) + D_2(a).
\end{aligned} \tag{4.9}$$

The key idea of the proof for $D_1(a)$ is to employ the same method as in [28]. We obtain the following under (**AS$_3$.iii**):

$$\begin{aligned}
\mathcal{M}_i &= \mathbb{E}\left[Z_i\left[\mathbb{E}\left[\gamma_a\left(\frac{B-u}{\hat{t}}\right)|A_i\right] - \mathbb{E}\left[\gamma_a\left(\frac{B-u}{\hat{t}}\right)|A=a\right]\right]|\mathcal{F}_{i-1}\right], \\
&= \mathbb{E}\left[Z_i\left[\mathbb{E}\left[\Gamma(A_i,u,\hat{t}) - \Gamma(a,u,\hat{t})|d(a,A_i)|\mathcal{F}_{i-1}\right]\right]\right], \\
&= \mathbb{E}\left[Z_i\Theta(d(a,A_i),u)|\mathcal{F}_{i-1}\right], \\
&= \int \Theta(rx_n,u)Z(r)d\mathbb{P}^{\mathcal{F}_{i-1}}(rh), \\
&= h\Theta'(0,u)\int rZ(r)d\mathbb{P}^{\mathcal{F}_{i-1}}(rh).
\end{aligned}$$

So, with $h \in (x_n, y_n)$, we employ continuity and the fact that

$$\int rZ(r)d\mathbb{P}^{\mathcal{F}_{i-1}}(rh) = Z(1)\sigma_i(a,h) - \int_0^1 (tZ(t))'\sigma_i(a,th)dt,$$

to get

$$\frac{1}{n}\sum_{i=1}^{n}\mathcal{M}_i = x_n\Theta'(0,\Upsilon(a))\left(Z(1) - \int_0^1 (tZ(t))'\eta_a(t)dt\right) + o_p(y_n).$$

Similarly, we obtain

$$\tfrac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[Z_i(a)|\mathcal{F}_{i-1}\right] = \left(Z(1) - \int_0^1 Z^{'}(t)\eta_a(t)dt\right) + o_p(1).$$

Lastly,

$$D_1 = o(y_n).$$

In the case of D_2, we can use the Taylor expansion to lead, under (AS_3),

$$D_2 = +v[n\sigma(a,x_n)]^{-1/2}\delta(a,\Upsilon(a))\frac{\partial}{\partial r}\Gamma(h,a,\Upsilon(a),\hat{t}) + o\left([n\sigma(a,x_n)]^{-1/2}\right).$$

As a result of the decomposition in (4.9), the outcome is obtained. □

Lemma 4. *Assume that the assumptions* **(AS_1), (AS_2), (AS_4)** *and* **(AS_7)** *holds. Then, we get that*

$$\left(\frac{n\sigma(a,x_n)\varrho_1^2}{\varrho_2\Lambda_2(a,\Upsilon(a),t)}\right)^{1/2}\mathfrak{R}_n(a,u,\hat{t}) = o(1), \quad a.co.$$

Proof. It suffices to demonstrate that

$$\frac{\bar{\Gamma}_N(h,a,r,\hat{t}) - \Gamma(h,a,r,\hat{t})\bar{\Gamma}_D(h,a)}{\bar{\Gamma}_D(h,a)} = o_p(1),$$

and

$$\left|\hat{\Gamma}_N(h,a,r,\hat{t}) - \bar{\Gamma}_N(h,a,r,\hat{t})\right| = o_p(1).$$

On other hand, there is

$$\begin{aligned}
&\frac{\bar{\Gamma}_N(h,a,r,\hat{t}) - \Gamma(h,a,r,\hat{t})\bar{\Gamma}_D(h,a)}{\bar{\Gamma}_D(h,a)} = \\
&\frac{1}{n\mathbb{E}[Z_1(a)]\bar{\Gamma}_D(h,a)}\sum_{i=1}^{n}\mathbb{E}\left[Z_i(a)\mathbb{E}\left[\gamma_a\left(\frac{B_i-r}{\hat{t}}\right)|\mathfrak{B}_{i-1}\right]|\mathcal{F}_{i-1}\right] - \Gamma(a,r,\hat{t})\mathbb{E}\left[Z_i(a)|\mathcal{F}_{i-1}\right],\\
&= \frac{1}{n\mathbb{E}[Z_1(a)]\bar{\Gamma}_D(h,a)}\sum_{i=1}^{n}\mathbb{E}\left[Z_i(a)\mathbb{E}\left[\gamma_a\left(\frac{B_i-r}{\hat{t}}\right)|A_i\right]|\mathcal{F}_{i-1}\right] - \Gamma(a,r,\hat{t})\mathbb{E}\left[Z_i(a)|\mathcal{F}_{i-1}\right],\\
&\leq \frac{1}{n\mathbb{E}[Z_1(a)]\bar{\Gamma}_D(h,a)}\sum_{i=1}^{n}\mathbb{E}\left[Z_i(a)|\Gamma(A_i,r,\hat{t}) - \Gamma(a,r,\hat{t})||\mathcal{F}_{i-1}\right].
\end{aligned}$$

We can conclude by using **(AS_2.ii))** that

$$\frac{\bar{\Gamma}_N(h,a,r,\hat{t}) - \Gamma(a,r,\hat{t})\bar{\Gamma}_D(h,a)}{\bar{\Gamma}_D(h,a)} \leq \sup_{a^{'}\in B(a,h)}\left|\Gamma(a^{'},r,\hat{t}) - \Gamma(a,r,\hat{t})\right| \to 0.$$

However, when viewed from the other side,

$$\hat{\Gamma}_N(h,a,u,\hat{t}) - \bar{\Gamma}_N(h,a,u,\hat{t}) = o_p(1).$$

Our next objective is to present the following two results:

$$\mathbb{E}\left[\hat{\Gamma}_N(h,a,u,\hat{t}) - \bar{\Gamma}_N(a,u,\hat{t})\right] \to 0,$$

and

$$Var\left[\hat{\Gamma}_N(a,u,\hat{t}) - \bar{\Gamma}_N(a,u,\hat{t})\right] \to 0.$$

The first comes as a result of the $\hat{\Gamma}_N(a,u,\hat{t})$ and $\bar{\Gamma}_N(a,u,\hat{t})$ definitions. For the second one, we have

$$\hat{\Gamma}_N(a,u,\hat{t}) - \bar{\Gamma}_N(a,u,\hat{t}) = \sum_{i=1}^{\mathrm{n}} \mu_i(a,u,\hat{t}),$$

where,

$$\mu_i(a,u,\hat{t}) = \frac{1}{n\mathbb{E}[Z_1]} Z_i \gamma_a\left(\frac{B_i - u}{\hat{t}}\right) - \mathbb{E}\left[Z_i \gamma_a\left(\frac{B_i - u}{\hat{t}}\right) | \mathcal{F}_{i-1}\right].$$

Using the Burkholder's inequality, we have

$$\mathbb{E}\left[\sum_{i=1}^{\mathrm{n}} \mu_i(a,u,\hat{t})\right]^2 \leq \sum_{i=1}^{\mathrm{n}} \mathbb{E}\left[\mu_i(a,u,\hat{t})\right]^2.$$

Furthermore, using Jensen's inequality, we can establish that

$$\begin{aligned}\mathbb{E}^2\left[\mu_i(a,u,\hat{t})\right] &\leq \frac{1}{n^2\mathbb{E}^2[Z_1]} \mathbb{E}\left[Z_i^2 \gamma_a^2\left(\frac{B_i - u}{\hat{t}}\right)\right] \leq \frac{1}{n^2\mathbb{E}^2[Z_1]} \mathbb{E}\left[Z_i^2\right] \\ &\leq \frac{1}{n\sigma^2(a,x_n)} \sigma_i(a,x_n).\end{aligned}$$

(AS$_2$) now produces

$$Var\left[\hat{\Gamma}_N(a,u,\hat{t}) - \bar{\Gamma}_N(a,u,\hat{t})\right] \to 0.$$

□

Lemma 5. *Assume that the assumptions* ***(AS$_1$)****,* ***(AS$_2$)*** *and* ***(AS$_7$)*** *hold. For all n that is large enough, $\hat{\Upsilon}(a)$ exists a.s.*

Proof. From the monotonicity of $\gamma_a\left(\frac{B-.}{\hat{t}}\right)$, for all $\epsilon > 0$,

$$\Gamma(a, \Upsilon(a) - \epsilon, \hat{t}) \leq \Gamma(a, \Upsilon(a), \hat{t}) \leq \Gamma(a, \Upsilon(a) + \epsilon, \hat{t}).$$

It is shown to us by employing an argument similar to that used in previous Lemmas that

$$\hat{\Gamma}(a,r,\hat{t}) \to \Gamma(a,r,\hat{t}) \text{ in probability},$$

for all real fixed $r \in \mathcal{N}_a$. So, for n large enough and, $\forall\epsilon$ small enough

$$\hat{\Gamma}(a, \Upsilon(a) - \epsilon, \hat{t}) \leq 0 \leq \hat{\Gamma}(a, \Upsilon(a) + \epsilon, \hat{t})$$

is verified with a probability approaching 1.

As γ_a is a continuous function, this implies that $\hat{\Gamma}(a,r,\hat{t})$ is a continuous function of r, and there exists $\hat{\Upsilon}(a) \in [\Upsilon(a) - \epsilon, \Upsilon(a) + \epsilon]$ such that $\hat{\Gamma}(a, \hat{\Upsilon}(a), \hat{t}) = 0$.

The uniqueness of $\hat{\Upsilon}(a)$ is a direct result of the strict monotonicity of γ_a in the second component and the fact that

$$\mathbb{P}\left(\sum_{i=1}^{n} Z_i = 0\right) = \mathbb{P}\left(\hat{\Gamma}_D(a) = 0\right) \to 0 \text{ as } n \to 0,$$

which indicates $\left(\sum_{i=1}^{n} Z_i \neq 0\right)$ with a probability tends to 1. Furthermore, since $\hat{\Upsilon}(a) \in [\Upsilon(a) - \epsilon, \Upsilon(a) + \epsilon]$ in probability, accordingly, it produces

$$\hat{\Upsilon}(a) \to \Upsilon(a) \text{ in probability as } n \to \infty.$$

□

5. Simulation studies

In this section, we use simulation to evaluate the proposed estimate's finite sample performance compared to the other classical estimators.

Applying the following regression model:

$$B = r(A) + \epsilon,$$

where B is the scalar response, $r(A)$ is the functional variables, and ϵ is a normally distributed random variable with a variance of 0.075. The functional variables used for explanation are built by:

$$A_i(r) = 2\omega_i r^2 + \frac{1}{2}\cos(\pi u_i r) \quad i = 1, \ldots 200, r \in [0, 1],$$

where ω_i are n independent real random variables, and follow Unif(0,1). Here ς_i are i.i.d. realizations of $N(0, 1)$ and are independent from ω_i and u_i, which are generated independently by $u_0 \sim N(0, 1)$. All the curves A_i's were discretized on the same grid generated from 200 measurements with equal spacing in the interval $(0, 1)$. The curves, A_i's, are plotted in Figure 1.

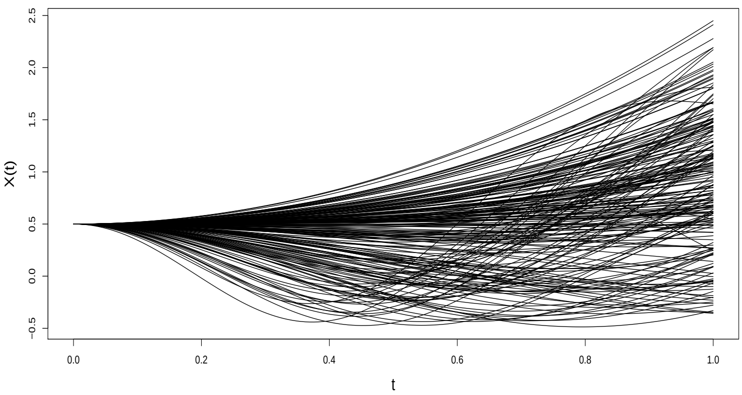

Figure 1. The curves A_i.

The second step in calculating the scalar response B_i is to consider the following operator:

$$r(a) = \int_0^1 \frac{10}{(1 + |a(r)|)} dr.$$

Our main goal is to compare the following three methods.

Method 1. Our estimator kNN Robust Equivariant Estimator (kNN REE) $\widehat{\Upsilon}(a)$ is the zero with respect to x of

$$\frac{\sum_{i=1}^{n} Z\left(\frac{d(a,A_i)}{H_{n,k}(a)}\right)\gamma_a\left(\frac{B_i - x}{\widehat{t}(a)}\right)}{\sum_{i=1}^{n} Z\left(\frac{d(a,A_i)}{H_{n,k}(a)}\right)} = 0.$$

Method 2. kNN robust Kernel Estimator (kNN RKE), $\tilde{\Upsilon}(a)$ introduced by [32] $\widetilde{\Upsilon}(a)$ is the zero with respect to x of

$$\frac{\sum_{i=1}^{n} Z\left(\frac{d(a,A_i)}{H_{n,k}(a)}\right)\gamma_a(B_i - x)}{\sum_{i=1}^{n} Z\left(\frac{d(a,A_i)}{H_{n,k}(a)}\right)} = 0.$$

Method 3. The Classical kNN Kernel Estimator (kNN CKE) $\widehat{m}(a)$ (see [22]), is defined as

$$\widehat{m}(a) = \frac{\sum_{i=1}^{n} B_i Z\left(\frac{d(a,A_i)}{H_{n,k}(a)}\right)}{\sum_{i=1}^{n} Z\left(\frac{d(a,A_i)}{H_{n,k}(a)}\right)}.$$

Next, the 200-sample is divided randomly into two parts: a training sample $(A_i, B_i)_{i=1}^{150}$ used for modeling, and a testing sample $(A_i, B_i)_{i=151}^{200}$ used to validate the prediction effect. Using the training sample, we can choose the optimal parameter k_{opt} for kNN robust equivariant and robust estimators, and the optimal parameter h_{opt} for NW robust estimator by the following cross-validation procedures, respectively. Specifically, for the robust equivariant kNN, we choose $k_{opt} = \arg\min_k CV_1(k)$, where $CV_1(k) = \sum_{i=1}^{n}\left(Y_i - \widehat{\Upsilon}_{(-i)}^{kNN}(a)\right)^2$ and $\widehat{\Upsilon}_{(-i)}^{kNN}(a)$ is the zero with respect to x of

$$\frac{\sum_{j=1,j\neq i}^{n} Z\left(\frac{d\left(a,A_j\right)}{H_{n,k}(a)}\right)\gamma_a\left(\frac{B_j - x}{\widehat{t}(a)}\right)}{\sum_{j=1,j\neq i}^{n} Z\left(\frac{d\left(a,A_j\right)}{H_{n,k}(a)}\right)}.$$

And the robust kNN one by $k_{opt} = \arg\min_k CV_2(k)$, where $CV_2(k) = \sum_{i=1}^{n}\left(Y_i - \widetilde{\Upsilon}_{(-i)}^{kNN}(a)\right)^2$ and $\widetilde{\Upsilon}_{(-i)}^{kNN}(a)$ is the zero with respect to x of

$$\frac{\sum_{j=1,j\neq i}^{n} Z\left(\frac{d\left(a,A_j\right)}{H_{n,k}(a)}\right)\gamma_a\left(B_j - x\right)}{\sum_{j=1,j\neq i}^{n} Z\left(\frac{d\left(a,A_j\right)}{H_{n,k}(a)}\right)}.$$

Similarly, we choose $h_{opt} = \arg\min_h CV_3(h)$ for the NW-kernel regression method, where $CV_3(h) = \sum_{i=1}^{n}\left(Y_i - \widehat{m}_{(-i)}^{kernel}(a)\right)^2$ and

$$\widehat{m}_{(-i)}^{kernel}(a) = \frac{\sum_{j=1, j\neq i}^{n} B_i Z\left(\frac{d\left(a, A_j\right)}{H_{n,k}(a)}\right)}{\sum_{j=1, j\neq i}^{n} Z\left(\frac{d\left(a, A_j\right)}{H_{n,k}(a)}\right)}.$$

Through this simulation study, we use the quadratic kernel Z, defined as $Z(v) = \frac{3}{4}\left(1 - v^2\right)\mathbb{1}_{[0,1]}(v)$. The semi-metric employed here is the first derivative of the sample curves provided by:

$$d\left(A_i, A_j\right) = \sqrt{\int \left(A_i'(r) - A_j'(r)\right)^2 dr}.$$

We worked with several functions ($L_1 - L_2$, Andrews, Tuckey, Cauchy ...), but we found that the best results are obtained when the $L_1 - L_2$ function $\left(\gamma_a(t) = \frac{t}{\sqrt{1 + t^2/2}}\right)$ is used. The predictions of the three models are displayed in Figure 2.

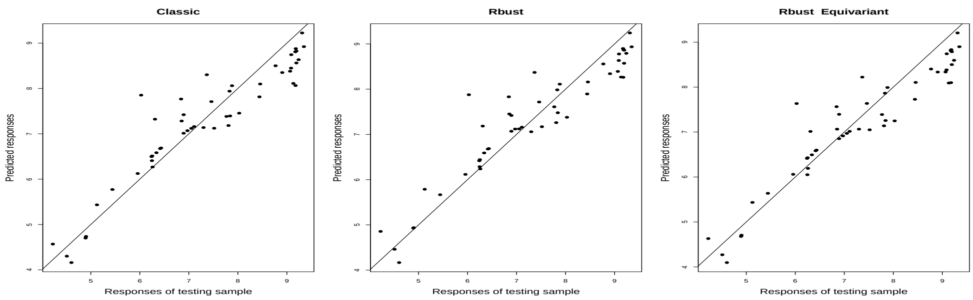

Figure 2. Predictions of the three models.

The empirical mean square error (MSE) is used to evaluate the predictors' effectiveness, where,

$$MSE_{\widehat{\Upsilon}} = n^{-1}\sum_{i=1}^{n}\left(\Upsilon(A_i) - \widehat{\Upsilon}_{(-i)}^{kNN}(A_i)\right)^2, \; MSE_{\widetilde{\Upsilon}} = n^{-1}\sum_{i=1}^{n}\left(\Upsilon(A_i) - \widetilde{\Upsilon}_{(-i)}^{kNN}(A_i)\right)^2,$$

and

$$MSE_{\widehat{m}} = n^{-1}\sum_{i=1}^{n}\left(\Upsilon(A_i) - \widehat{m}_{(-i)}^{kernel}(A_i)\right)^2.$$

The MSEs of the three models are displayed in Figure 3. In this comparative study, identical conditions are applied to the three estimators. The first illustration relates to the MSE of each estimator. Then, 100 independent replications of the same data are generated using $n-$samples$(n = 200)$.

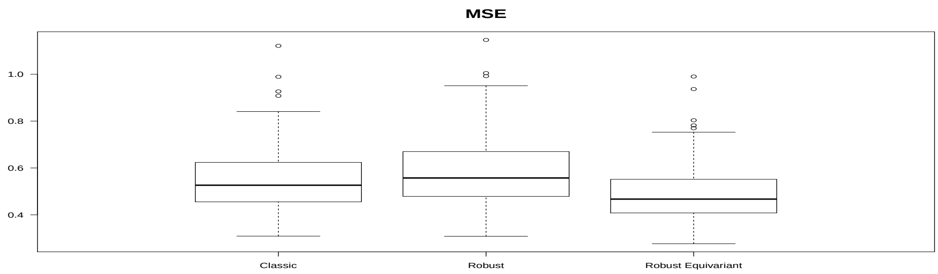

Figure 3. MSE of the three models.

For the second scenario, we introduce some outliers to the data to highlight the major aspect of our approach. In this part, we simulated data with $MC = 0$, 5, and 10 for the multiplier (MC is the number of perturbed observations). In all six instances, we reached the same conclusion: typically, when outliers are presented, the robust regression method performs better than the traditional one. Even if the MSE of both approaches increases significantly with the number of perturbed points and the multiplicative coefficient MC value, the MSE of the robust kNN method remains extremely low. The results presented in Table 1 demonstrate that the robust kNN technique is superior to the other methods and that our functional forecasting procedure for the robust method behaves well in the presence of outliers.

Table 1. Comparison between the three methods in the presence of outliers.

MC	kNN CKE	kNN RKE	kNN REE
0	0.5507103	0.5919875	0.5519775
5	944.1715413	0.9448654	0.5738740
10	2819.2037625	1.6516898	0.5877922

The primary use of Theorem 1 is to construct a confidence interval for the actual value of r(a) given the curve $A = a$. Figures 4 and 5 clearly demonstrate the superior performance of our estimator compared to the standard regression, both in the absence and presence of outliers.

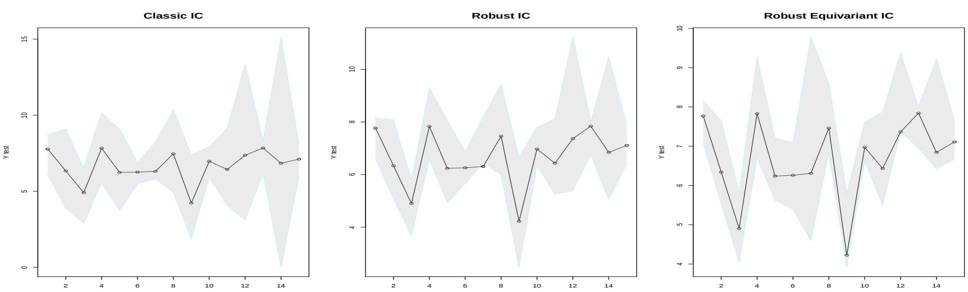

Figure 4. Extremes of real values and confidence intervals (simulation data in the presence of 10 outliers). The true values are connected by the solid black curve. The dashed Blue curves connect the expected minimum and maximum values.

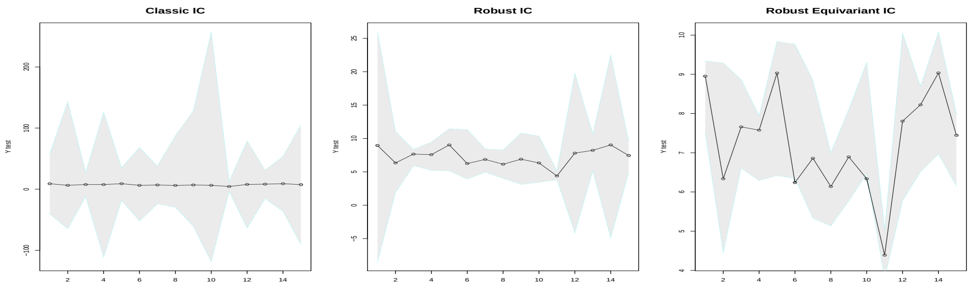

Figure 5. Extremes of real values and confidence intervals (simulation data in the presence of 10 outliers). The true values are connected by the solid black curve. The dashed Blue curves connect the expected minimum and maximum values.

6. Real data application

We will also look at how straightforward it is to apply the k-NN robust regression with an unknown scale parameter in practice, in addition to discussing its advantages over competing models (regression and robust regression). In this paper, we compare the k-NN robust regression with a scale parameter to other regressions in the risk analysis field (robust also). Therefore, we perform an empirical study based on the daily returns of $r(t)$ of 4 worldwide financial stock indices to quantify this gain in the real world.

The link to the data used in this study is mentioned in the "Data Availability Statement" Section, which has all the information you need. It covers the time span from January 1, 2017 to December 31, 2020. We use this information to analyze 192 random months from various sources. All primary characteristics are still present in the evaluated functional data (see Figure 6).

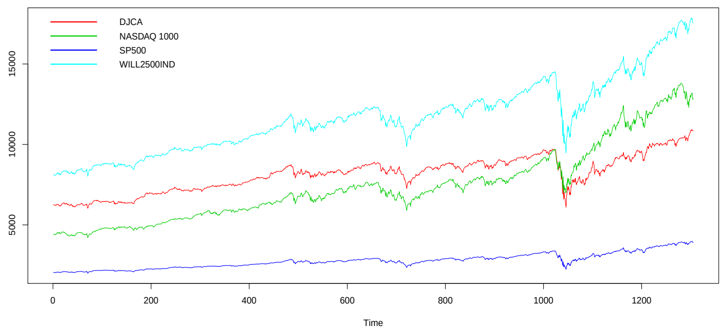

Figure 6. The daily values of $Z(t) = -100\log\left(\frac{r(t)}{r(t-1)}\right)$ for the 4 stocks index.

We assume that the curve $X(.)$ represents the monthly curve of the time series. $Z(t) = -100log(\frac{r(t)}{r(t-1)})$, and $Y = Z(t_l + 1)$ indicates the true response variable, where t_l her means the month's last day. The efficiency of these approaches, like that of all statistical modeling, is firmly connected to the selection of the various parameters included in the estimators' definition. To perform a fair comparison of the two mean squared errors, we follow the same procedure when selecting the fundamental parameters in the estimators. Specifically, we employ this procedure with a quadratic kernel on the interval (0, 1) and the PCA metric. Note that the ideal bandwidth was determined by dividing the data into subsets (Learning and testing samples (150+42)) and using the Cross-Validation rule to the 150-learning observations to determine the best smoothing parameter. We have only studied the most prevalent approach, cross-validated selected bandwidth, for the purpose of brevity. Lastly, the effectiveness of the proposed estimator's strategies is computed using the empirical mean square error (MSE) criterion (defined in the simulation section).

We test the proposed criteria 100 times, and in each case, we switch the observations from the learning part to the testing part. These box-plot figures show the errors that were found. It is clear that the k-NN robust equivariant estimation is slightly more accurate than the other estimation methods (see Figures 7–10).

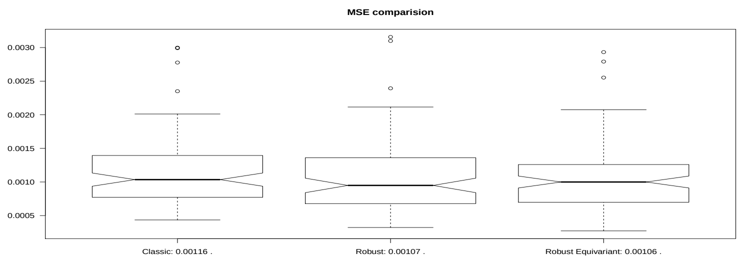

Figure 7. Comparison of the MSE between the kNN CKE, kNN RKE and kNN REE approach for DJA Index.

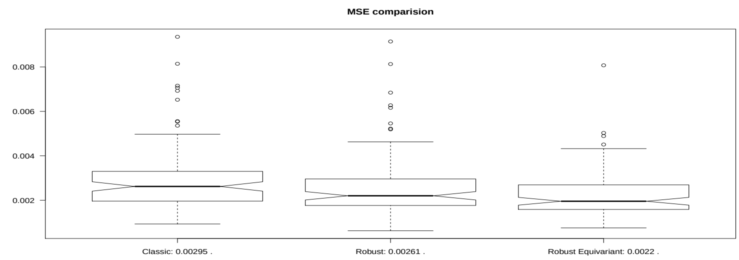

Figure 8. Comparison of the MSE between the kNN CKE, kNN RKE and kNN REE approach for NASDAQ Index.

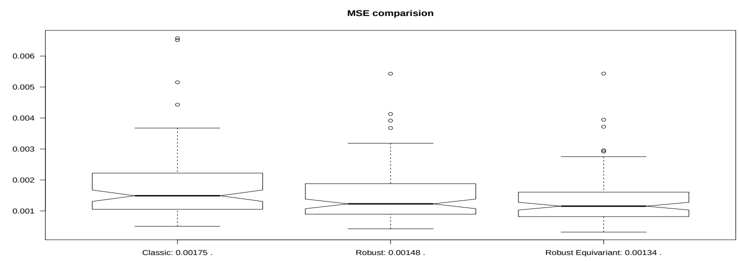

Figure 9. Comparison of the MSE between the kNN CKE, kNN RKE and kNN REE approach for SP500 Index.

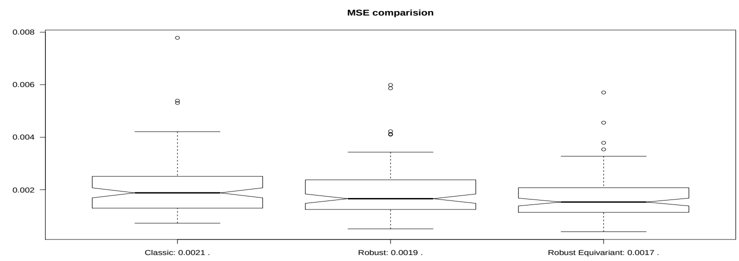

Figure 10. Comparison of the MSE between the kNN CKE, kNN RKE and kNN REE approach for WILL2500 Index.

Figures 11–14 show the observed responses in terms of their predictions of the four Index markets for the various proposed estimators. The robust equivariant k-NN method appears to have a slight precision advantage over the other presented models.

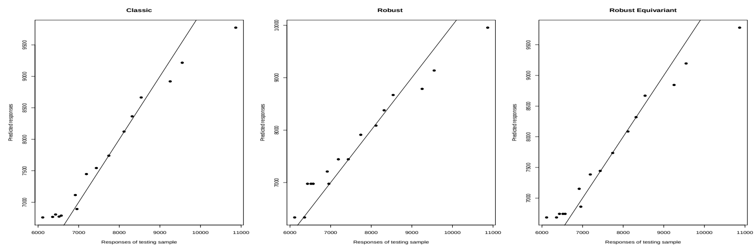

Figure 11. Y_i's predictions versus their Classical (left panel), Robust (middle panel), and Equivariant models (right panel) for DJA Index.

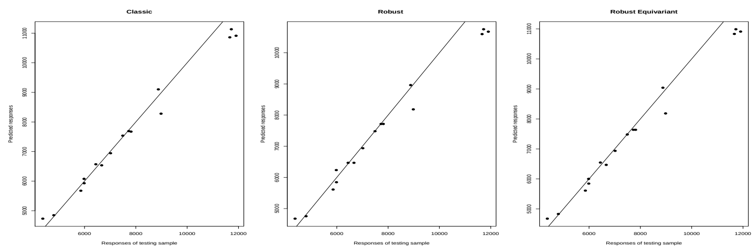

Figure 12. Y_i's predictions versus their Classical (left panel), Robust (middle panel), and Equivariant models (right panel) for NASDAQ Index.

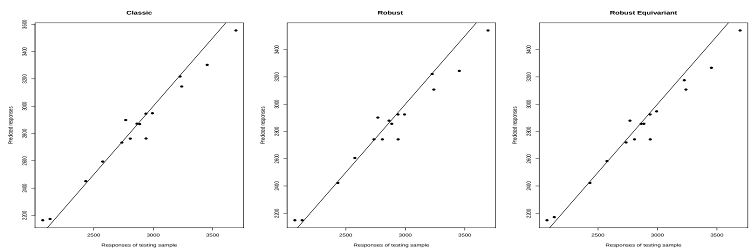

Figure 13. Y_i's predictions versus their Classical (left panel), Robust (middle panel), and Equivariant models (right panel) for SP500 Index.

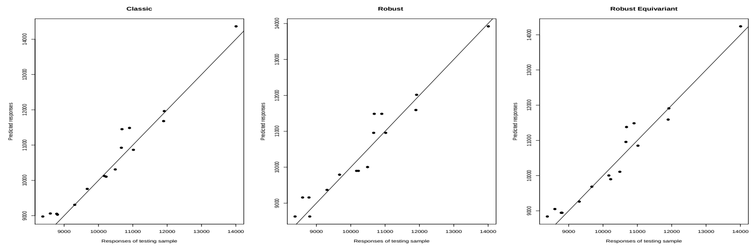

Figure 14. Y_i's predictions versus their Classical (left panel), Robust (middle panel), and Equivariant models (right panel) for WILL2500 Index.

7. Conclusions

In this paper, we estimate the functional ergodic data with the k-NN method and show their almost certainly convergence. The asymptotic normality aspect in the case of the robust regression function holds even when the scale parameter is unknown. The results were established under enough standard conditions that make it possible to look at different structural axes, such as the functional naturalness of the model and the data, the regression function's robustness, and the observations' correlation. The simulation study results and data application reveal the consistency and superiority of our theoretical analysis since our k-NN estimator's performance outperforms a nonparametric series estimator.

For future work, most techniques that use nonparametric functional kernel smoothers could be effectively extended. Among other things, the challenge here would be to extend the concepts to other nonparametric predictors, such as functional local linear, functional kNN, etc. Further applications to other types of prediction models (such as functional single index models, partial linear models, ..., etc.) in which an initial kernel stage plays a crucial role are also possible. Also, our asymptotic result could be applied to other types of dependency data, especially those associated positively (see [33]).

Acknowledgments

The authors thank the reviewers for their valuable comments and extend their appreciation to the funder of this work. This research project was funded by the Deanship of Scientific Research, Princess Nourah bint Abdulrahman University, through the Program of Research Project Funding After Publication, grant No (43- PRFA-P-23).

Conflict of interest

The authors declare no conflicts of interest.

References

1. F. Ferraty, P. Vieu, Nonparametric models for functional data, with applications in regression, time series prediction and curves discrimination, *J. Nonparametr. Stat.*, **16** (2004), 111–125. https://doi.org/10.1080/10485250310001622686
2. F. Ferraty, P. Vieu, *Nonparametric functional data analysis theory and practice*, New York: Springer, 2006.

3. G. Boente, R. Fraiman, Robust nonparametric regression estimation, *J. Multivariate Anal.*, **29** (1989), 180–198. https://doi.org/10.1016/0047-259X(89)90023-7
4. P. J. Huber, Robust estimation of a location parameter: annals mathematics statistics, *IEEE T. Signal Proces.*, **56** (1964), 2356–2356.
5. G. Collomb, W. Hardle, Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations, *Stoch. Proc. Appl.*, **23** (1986), 77–89. https://doi.org/10.1016/0304-4149(86)90017-7
6. N. Laïb, E. Ould-Said, A robust nonparametric estimation of the autoregression function under ergodic hypothesis, *Canad. J. Stat.*, **28** (2000), 817–828. https://doi.org/10.2307/3315918
7. F. Ruggeri, Nonparametric bayesian robustness, *Chil. J. Stat.*, **1** (2010), 51–68.
8. N. Azzedine, A. Laksaci, E. Ould-Said, On the robust nonparametric regression estimation for functional regressor, *Stat. Probabil. Lett.*, **78** (2008), 3216–3221. https://doi.org/10.1016/j.spl.2008.06.018
9. C. Crambes, L. Delsol, A. Laksaci, Robust nonparametric estimation for functional data, *J. Nonparametr. Stat.*, **20** (2008), 573–598. https://doi.org/10.1080/10485250802331524
10. G. Boente, A. Vahnovan, Strong convergence of robust equivariant nonparametric functional regression estimators, *Stat. Probabil. Lett.*, **100** (2015), 1–11. https://doi.org/10.1016/j.spl.2015.01.028
11. N. Laïb, D. Louani, Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties, *J. Multivariate Anal.*, **101** (2010), 2266–2281. https://doi.org/10.1016/j.jmva.2010.05.010
12. N. Laïb, D. Louani, Rates of strong consistencies of the regression function estimator for functional stationary ergodic data, *J. Stat. Plann. Infer.*, **141** (2011), 359–372. https://doi.org/10.1016/j.jspi.2010.06.009
13. A. Gheriballah, A. Laksaci, S. Sekkal, Nonparametric Mregression for functional ergodic data, *Stat. Probabil. Lett.*, **83** (2013), 902–908. https://doi.org/10.1016/j.spl.2012.12.004
14. F. Benziadi, A. Gheriballah, A. Laksaci, Asymptotic normality of kernel estimator of ψ-regression functional ergodic data, *New Trends Math. Sci.*, **1** (2016), 268–282.
15. F. Benziadi, A. Laksaci, F. Tebboune, Recursive kernel estimate of the conditional quantile for functional ergodic data, *Commun. Stat. Theor. M.*, **45** (2016), 3097–3113. https://doi.org/10.1080/03610926.2014.901364
16. D. Bosq, *Linear processes in function spaces: theory and applications*, Berlin: Springer, 2000.
17. J. O. Ramsay, B. W. Silverman, *Applied functional data analysis: methods and case studies*, New York: Springer, 2002.
18. G. Geenens, Curse of dimensionality and related issues in nonparametric functional regression, *Stat. Surv.*, **5** (2011), 30–43. https://doi.org/10.1214/09-SS049
19. I. M. Almanjahie, M. K. Attouch, O. Fetitah, H. Louhab, Robust kernel regression estimator of the scale parameter for functional ergodic data with applications, *Chil. J. Stat.*, **11** (2020), 73–93.

20. I. M. Almanjahie, K. Aissiri, A. Laksaci, Z. Chiker Elmezouar, The k nearest neighbors smoothing of the relative-error regression with functional regressor, *Commun. Stat. Theor. M.*, **51** (2020), 4196–4209. https://doi.org/10.1080/03610926.2020.1811870

21. W. Bouabsa, Nonparametric relative error estimation via functional regressor by the k Nearest Neighbors smoothing under truncation random data, *AAM*, **16** (2021), 97–116.

22. F. Burba, F. Ferraty, P. Vieu, k-Nearest neighbor method in functional nonparametric regression, *J. Nonparametr. Stat.*, **21** (2009), 453–469. https://doi.org/10.1080/10485250802668909

23. M. Attouch, W. Bouabsa, The k-nearest neighbors estimation of the conditional mode for functional data, *Rev. Roumaine Math. Pures Appl.*, **58** (2013), 393–415.

24. M. Attouch, W. Bouabsa, Z. Chiker el mozoaur, The k-nearest neighbors estimation of the conditional mode for functional data under dependency, *Int. J. Stat. Econ.*, **19** (2018), 48–60.

25. M. Attouch, F. Belabed, The k nearest neighbors estimation of the conditional hazard function for functional data, *REVSTAT Stat. J.*, **12** (2014), 273–297. https://doi.org/10.57805/revstat.v12i3.154

26. L. Z. Kara, A. Laksaci, M. Rachdi, P. Vieu, Data-driven kNN estimation in nonparametric functional data analysis, *J. Multivariate Anal.*, **153** (2017), 176–188. https://doi.org/10.1016/j.jmva.2016.09.016

27. I. M. Almanjahie, O. Fetitah, M. Attouch, H. Louhab, Asymptotic normality of the robust equivariant estimator for functional nonparametric models, *Math. Probl. Eng.*, **2022** (2022), 8989037. https://doi.org/10.1155/2022/8989037

28. F. Ferraty, A. Mas, P. Vieu, Nonparametric regression on functional data: inference and practical aspect, *Aust. New Zeal. J. Stat.*, **49** (2007), 267–286. https://doi.org/10.1111/j.1467-842X.2007.00480.x

29. P. Gaenssler, J. Strobel, W. Stute, On central limit theorems for martingale triangular arrays, *Acta Math. Acad. Sci. H.*, **31** (1978), 205–216.

30. F. Ferraty, P. Vieu, Additive prediction and boosting for functional data, *Comput. Stat. Data Anal.*, **53** (2009), 1400–1413. https://doi.org/10.1016/j.csda.2008.11.023

31. P. Hall, C. Heyde, *Martingale limit theory and its application*, New York: Academic Press, 1980.

32. M. Attouch, T. Benchikh, Asymptotic distribution of robust k-nearest neighbour estimator for functional nonparametric models, *Mat. Vestn.*, **64** (2012), 275–285.

33. C. Azevedo, P. E. Oliveira, On the kernel estimation of a multivariate distribution function under positive dependence, *Chil. J. Stat.*, **2** (2011), 99–113.

AIMS Press