Hacettepe Journal of Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 52 (2) (2023), 512 – 528
DOI : 10.15672/hujms.1100871

Research Article

kNN robustification equivariant nonparametric regression estimators for functional ergodic data

Somia Guenani*, Wahiba Bouabsa, Mohammed Kadi Attouch, Omar Fetitah

Lab. of Statistics and Process Stochastic, University Djillali Liabes, BP 89 Sidi Bel Abbes 22000, Algeria

Abstract

We discuss in this paper the robust equivariant nonparametric regression estimators for ergodic data with the k Nearst Neighbour (kNN) method. We consider a new robust regression estimator when the scale parameter is unknown. The principal aim is to prove the almost complete convergence (with rate) for the proposed estimator. Furthermore, a comparison study based on simulated data is also provided to illustrate the finite sample performances and the usefulness of the kNN approach and to prove the highly sensitive of the kNN approach to the presence of even a small proportion of outliers in the data.

Mathematics Subject Classification (2020). 62G05, 62G08, 62G35

Keywords. Functional data, ergodic data, kNN estimation, kernel estimate, uniform almost complete convergence rate, entropy

1. Introduction

It is very well recognized that robust regression in statistics is an attractive research method. It is used to overcome some of the weaknesses of classical regression, namely when outliers contain heteroscedastic data.

The study of the connection between a random variable W and a set of covariates Z is a common problem in statistics. In the literature, these variables are generally known as functional variables. Remember that the robustification method is an old statistical issue, This latter was investigated first by [29] who studied an estimation of alocation parameter (see also [18, 33]), for some results containing the multivariate time series case under a mixing or an ergodic condition);

The robust model is an essential alternative regression model that allows overcoming many drawbacks of the classical regression, such as the sensitivity to the outliers or the heteroscedasticity phenomena. Indeed, it was initially proposed by [8] who demonstrated the model's almost-complete convergence in the independent and identically distributed (i.i.d.) case. Several results on nonparametric robust functional regression have been obtained since this study (for example, [5–7, 12, 16, 20, 27] and references therein).

*Corresponding Author.
Email addresses: guenanisomia@gmail.com (S. Guenani), wahiba_bouab@yahoo.fr (W. Bouabsa), attou_kadi@yahoo.fr (M.K. Attouch), fetitah-omar@hotmail.com (O. Fetitah)
Received: 14.05.2022; Accepted: 03.10.2022

Furthermore, it is well known that the kNN method is better than the classical kernel method, this famous method have attracted a lot of interest in the statistical literature for evaluating multivariate data because of their flexibility and efficiency. Pushed by its attractive features, the functional kNN smoothing approach has received a growing consideration in the last years. The study of [28] is a thorough analysis of kNN estimators in the finite dimensional context. Work in this area was started by [19], and a large number of articles are now available in various estimating contexts, which including regression, discrimination, density and mode estimation, and clustering analysis, we make reference to [3, 4, 9, 15, 17, 21–23, 32, 35–37, 40, 45] and [1, 14, 30] for the most recent advances and references. Note that, such a study has a great impact on practice. However, the difficulty in the kNN smoothing is the fact that the bandwidth parameter is a random variable, unlike the classical regression in which the smoothing parameter is a deterministic scalar. So, the study of the asymptotic properties of our proposed estimator is complicated, and it requires some additional tools and techniques.

All the results involved in the functional kNN estimation above were obtained under i.i.d. case. While in many practical applications, some problems require taking into account the dependence structure that may exist within the dataset. The strong mixing dependence or α-mixing is one of the most general weak dependence modelization in the literature. The research of Nadaraya Watson (NW) kernel method for this dependent functional data analysis has been widely carried out, see, for instance [13, 25] and the bibliographical surveys by [26] and [39]. However, for the kNN approach, the papers are, as far as we know, written by [41] and [38] who studied the kNN estimator of the model under α-mixing sample.

The ergodicity hypothesis is less restrictive than the mixing condition, so we consider in this article the more global case when the scale parameter is unknown and data come from an ergodic functional time series by the kNN method. The literature on ergodic functional time series data is still restricted, with the few existing results due to [10, 27, 34] and references therein. Inspired by all the results above, the purpose of this paper show us that functional kNN approach can be used to further investigate the estimation of functional nonparametric regression opera in the case of ergodics datasets. This is motivated by the fact that the robust regression estimator has several advantages over the classical kernel regression estimator. The main profit in using a robust regression is that it allows reducing the effect of outlier data.

In NFDA, kNN robustification equivariant nonparametric regression estimators for ergodic data is new. This researches's primary goal is to provide generalizations, to the kNN case, the results obtained by [2] in ergodic dependency case with the research of [38] and [1]. More precisely, we establish the almost complete convergence with rates of the constructed estimator by combining the ideas of robustness with those of smoothed regression. We point out that the main feature of our approach is to develop an alternative prediction model to the classical regression that is not sensitive to outliers or heteroscedastic data, taking into account the local data structure. The work has not yet been addressed in the literature. We wish that this will be useful to readers who are interested in learning about and comprehending the core idea of functional kNN methods with ergodic dependence sample.

This paper's structure is as follows. In Section 2, we find some fundamental concepts and various assumptions. Then in Section 3 we give some technical tools as well as their proofs. The main result is given in Section 4, then we provides all the proofs of the main result in Section 5. Finally, simulation study is given in Section 6.

2. Principal hypotheses and basic definitions

2.1. Kolmogorov's entropy

The aim of this subsection is to emphasize the topological aspects of our study. Indeed, all asymptotic conclusions in nonparametric statistics for functional variables are intimately connected to the concentration properties of the probability measure of the functional variable Z, as Ferraty and Vieu [25] indicated. We must also consider the element of uniformity in this situation.

Definition 2.1. Let $\varepsilon > 0$ be given, and let $\mathcal{T}$ be a subset of a semi-metric space $\mathcal{F}$, a finit set of points $z_1, z_2, ..., z_n$ in $\mathcal{F}$ is called an ε-net for $\mathcal{T}$ if $\mathcal{T} \subset \bigcup_{\ell=1}^{n} B(z_\ell, \varepsilon)$. Kolmogorov's ε-entropy of the set $\mathcal{T}$ is defined as $\Psi_{\mathcal{T}_{\mathcal{F}}}(\varepsilon) = \log(N_\varepsilon(\mathcal{T}_{\mathcal{F}}))$, where $N_\varepsilon(\mathcal{T}_{\mathcal{F}})$ is the minimal number of open balls in $\mathcal{F}$ with radius ε required to cover $\mathcal{F}$.

This concept was introduced by [31] and it represents a measure of the complexity of a set, in sense that, high entropy means that much information is needed to describe an element with an accuracy ε. Therefore, the choice of the topological structure (with other words, the choice of the semi-metric) will play a crucial role when one is looking at uniform (over $\mathcal{T}$) asymptotic results. More precisely, a good semi-metric can increase the concentration of the probability measure of the functional variable Z as well as minimize the ε-entropy of the subset $\mathcal{T}_{\mathcal{F}}$.

2.2. kNN regression function model

Let $X_i = (Z_i, W_i)_{i=1,...,n}$ be n pairs independent and identically distributed (i.i.d) as (Z, W) and is defined in $\mathcal{F} \times \mathbb{R}$. We do not assume the existence of a density for the functional random variable Z since $(\mathcal{F}, d)$ is a semi-metric space, $\mathcal{F}$ is not necessarily of finite dimension. The functional nonparametric regression is defined as

$$W = r(Z) + \varepsilon \text{ with } \mathbb{E}[\varepsilon|Z] = 0.$$

The kNN kernel estimator can be written as for a fixed $z \in \mathcal{F}$

$$\widehat{r}_{kNN}(z) = \frac{\sum_{i=1}^{n} W_i L(T_{n,k}(z)^{-1} d(z, Z_i))}{\sum_{i=1}^{n} L(T_{n,k}(z)^{-1} d(z, Z_i))}, \tag{2.1}$$

where L is an asymmetrical kernel and $T_{n,k}(z)$ is defined as follows:

$$T_{n,k}(z) = \min\left\{ h \in \mathbb{R}^+ / \sum_{i=1}^{n} \mathbb{I}_{B(z,h_L)}(z_i) = k \right\}.$$

The functional version of the NW kernel type estimator of the nonparametric functional regression is as follows:

$$\widehat{r}(z) = \frac{\sum_{i=1}^{n} W_i L(h_L^{-1} d(z, Z_i))}{\sum_{i=1}^{n} L(h_L^{-1} d(z, Z_i))}, \tag{2.2}$$

where $z \in \mathcal{F}$ is fixed, and h_L denotes a non-random bandwidth.

2.3. kNN conditional cumulative distribution function

The conditional cumulative distribution function of W given $\mathrm{Z} = z$, for each $z \in \mathcal{F}$ and for any $w \in \mathbb{R}$ can be written as

$$F(w|\mathrm{Z}=z) = F_{\mathrm{Z}}^{\mathrm{W}}(z,w) = \mathbb{P}(\mathrm{W} \leq w|\mathrm{Z}=z) = \mathbb{E}[\mathbb{I}_{(-\infty,w]}(\mathrm{W})|\mathrm{Z}=z].$$

We call the following function the estimator of $F(w|\mathrm{Z}=z)$

$$\widehat{F}(w|\mathrm{Z}=z) = \sum_{i=1}^{n} L(d(z,Z_i)/T_{n,k}(z)) \left(\sum_{i=1}^{n} L(d(z,Z_i)/T_{n,k}(z)) \right)^{-1} \mathbb{I}_{(-\infty,w]}(\mathrm{W_i}). \tag{2.3}$$

Several authors have studied the estimation of the conditional cumulative distribution function in the real case (see for example [42,43]). Then, in the functional case [25] proved the almost complete convergence of a double kernel estimator of the conditional cumulative distribution function.

2.4. The kNN robust equivariant estimators and their functional relatives function

In this section we define the function of our main problem, we consider estimating a generalized regression function defined as follows:

$$\mu(z,x,\tau(z)) = \mathbb{E}\left[\Gamma_z \left(\frac{W_i - x}{\tau(z)} \right) / Z_i = z \right], \tag{2.4}$$

where Γ_z is a real-valued function, we denoted by $\vartheta(z)$ the unique solution of $\mu(z,x,t(z)) = 0$, where $t(z)$ is a robust measure of the conditional scale. The unique solution of Eq. (2.4) is the so-called robust conditional location functional, where Γ_z is a strictly increasing function (see [11]). The conditional scale measure is defined as the conditional median of the absolute deviation from the conditional median, that is

$$t(z) = med(|W - M(z)|/Z = z) = mad_c(F_W^z(.)), \tag{2.5}$$

with $M(z) = med(W/Z = z)$ is the median of the conditional distribution.
On the other hand, we note that $t(z)$ which is a robust measure of the conditional scale, always equals $\tau(z)$.
We insert an estimator of $F_W^z(z)$ into (2.3) to get $\vartheta(z)$ estimators, wich will betaken as $\widehat{F}(w|\mathrm{Z}=z)$. A robust estimator of the conditional scale is denoted by $\widehat{t}(z)$, for example, $\widehat{t}(z) = mad_c(\widehat{F}(.|Z=z)$, the scale measure given in (2.5) measured in $\widehat{F}(w|\mathrm{Z}=z)$. The solution $\widehat{\vartheta}(z)$ of $\widehat{\mu}(z,x,\widehat{t}(z)) = 0$ gives the robust nonparametric estimator of $\vartheta(z)$ in this notation, where

$$\widehat{\mu}(z,x,\widehat{t}(z)) = \frac{\sum_{i=1}^{n} L(d(z,Z_i)/h_L)\Gamma_z\left(\frac{W_i - x}{\widehat{t}(z)}\right)}{\sum_{i=1}^{n} L(d(z,Z_i)/h_L)}. \tag{2.6}$$

Hence the kNN estimator of $\mu(.)$ is written as

$$\widehat{\mu}_{kNN}(z,x,\widehat{t}(z)) = \frac{\sum_{i=1}^{n} L(d(z,Z_i)/T_{n,k}(z))\Gamma_z\left(\frac{W_i - x}{\widehat{t}(z)}\right)}{\sum_{i=1}^{n} L(d(z,Z_i)/T_{n,k}(z))}. \tag{2.7}$$

2.5. Hypotheses

In this part, we propose the following hypotheses to establish the uniform almost complete convergence of $\widehat{\mu}$ on some subset $\mathfrak{T}_{\mathcal{F}}$ of $\mathcal{F}$. To do that we denote by C and C' some real generic constants supposed strictly positive and we suppose that

(A1) The processes (Z_i, W_i) satisfies

(A1a) $\forall h_L > 0,\ \varphi_z(h_L) =: \mathbb{P}(Z \in B(z, h_L)) > 0$ where $\varphi_z(.)$ is continuous in the neighborhood of 0 and $\varphi_z(0) = 0$.

(A1b) $\forall i = \overline{1, n}$ a deterministic function exists such that

$$\forall h_L > 0,\ 0 < C\varphi_i(h_L) < \mathbb{P}(Z_i \in B(z, h_L)/\mathcal{F}_{i-1}) \leq C'\varphi_i(h_L) < \infty.$$

(A1c) For all $h_L > 0,\ \dfrac{1}{n\varphi_z(h_L)} \sum_{i=0}^{n} \mathbb{P}(Z_i \in B(z, h_L)/\mathcal{F}_{i-1}) \to 1.$ a.co.

(A2) $\exists$ function $\phi(.) \geq 0$, a bounded function $f(.) > 0$, $\alpha > 0$ and $\rho > 0$ such that

(A2a) $\phi(0) = 0$ and $\lim_{\varepsilon \to \infty} \phi(\varepsilon) = 0$,

(A2b) $\lim_{\varepsilon \to \infty} (\phi(u\varepsilon)/\phi(\varepsilon)) = u^\alpha$ with $u > 0$,

(A2c) $\sup_{z \in \mathfrak{T}_{\mathcal{F}}} |\varphi_z(\varepsilon)/\phi(\varepsilon) - f(z)| = O(\varepsilon^\rho)$, as $\varepsilon \to 0$.

(A3) The kernel $L(.)$ is defined by

(A3a) is a nonnegative function with support $[0, 1]$ such that

$$0 < C\mathbb{I}_{[0,1]}(t) < L(t) < C'\mathbb{I}_{[0,1]}(t) < +\infty.$$

(A3b) Its derivative $L'(.)$ exists on the same support and $-\infty < C < L'(t) < C' < 0$.

(A4) $\forall (x_1, x_2) \in [\vartheta(z) - \delta, \vartheta(z) + \delta] \times [\vartheta(z) - \delta, \vartheta(z) + \delta], \forall (z_1, z_2) \in \mathfrak{T}_{\mathcal{F}}$,

$$|\mu(z_1, x, t(z)) - \mu(z_2, x, t(z))| \leq Cd^\beta(z_1, z_2),\ \beta > 0.$$

(A5) For each fixed $x \in [\vartheta(z) - \delta, \vartheta(z) + \delta]$, $\forall m \geq 2$ we have that

$$\mathbb{E}[\Gamma_z\Big(\frac{W_i - x}{t(z)}\Big)/z = Z] < \delta(z) < C < \infty, \text{ with } \delta(.) \text{ continuous on } \mathfrak{T}_{\mathcal{F}}.$$

(A6) The functions φ_z and $\Psi_{\mathfrak{T}_{\mathcal{F}}}$ are such that
$\exists C > 0,\ \exists \eta_0 > 0,\ \forall \eta < \eta_0,\ \varphi_z'(\eta) < C$ and if $L(1) = 0$ we can seen that

$$\exists C > 0,\ \exists \eta_0 > 0,\ \forall 0 < \eta < \eta_0,\ \int_0^\eta \varphi(u)du > C\eta\varphi_z(\eta),$$

if in addition $k/n \to 0$ as $n \to \infty$, $\log^2 n/k < \Psi_{\mathfrak{T}_{\mathcal{F}}}(\log n/n) < k/\log n$ and $0 < C < k/\log n < C' < \infty$ for n large enough.

(A7) Kolmogorov's ε-entropy of the set $\mathfrak{T}_{\mathcal{F}}$ satisfies, for some $\varpi > 1$

$$\sum_{n=1}^{\infty} \exp\{(1 - \varpi)\Psi_{\mathfrak{T}_{\mathcal{F}}}(\log(n)/n)\} < \infty.$$

(A8) Consider that $\mathfrak{T}_{\mathcal{F}}$ is a compact set of $\mathcal{F}$ such that

(A8a) The function $F(w|Z = z)$ is uniformly continuous of z in a neighborhood of $\mathfrak{T}_{\mathcal{F}}$ for each z fixed.

(A8b) The equicontinuity condition that follows hold

$$\forall \varepsilon > 0,\ \exists \delta > 0 : |\lambda - \nu| < \delta \Longrightarrow \sup_{z \in \mathfrak{T}_{\mathcal{F}}} |F(\lambda|Z = z) - F(\nu|Z = z)| < \varepsilon.$$

Remark 2.2. Comments on the hypotheses
Our hypotheses are quite light in the context of nonparametric statistics in functional time series.

The latter is exploited together with condition (A1) which is less restrictive than the conditions imposed by [34] because the concentration function $\mathbb{P}(Z_i \in B(z, r))$ and the conditional concentration function $\mathbb{P}(Z_i \in B(z,r)/\mathcal{F}_{i-1})$ do not need to be written as products of two independent nonnegative functions of the center and radius.

Hypotheses(A3) contains two types of kernels which have been utilized in practice box and continuous kernels.

(A2), (A4) and (A5) are the usual conditions in the nonparametric setting.

About condition (A6) we can say because the derivative of φ is limited around zero, it can be considered a Lipschitzian function.

Assumption (A7) acts on Kolmogorov's ε-entropy of $\mathcal{T}_{\mathcal{F}}$.

Assumption (A8) means that there $\exists a, b \in \mathbb{R}$ such that for every $z \in \mathcal{T}_{\mathcal{F}}$, $F(b|Z = z)| > 1 - \varepsilon$ and $F(a|Z = z)| < \varepsilon$ which will be used to prove that $t(z) = mad_c(F_W^z(.))$ is bounded away from 0 for all $z \in \mathcal{T}_{\mathcal{F}}$.

3. Technical tools and their proofs

The first difficulty comes because $T_{n,k}(z)$ is random. To resolve this problem, the idea is to frame sensibly $T_{n,k}(z)$ by two non-random windows. More generally, these technical tools could be useful as long as one has to deal with random bandwidths. So we propose in this part the preliminary Lemma and their proof that is necessary to prove our main result. Following the notations in [15] or [32].

Let $(A_i, B_i)_{1\le i\le n}$ be n random pairs valued in $(\Omega\times\mathbb{R}, \mathcal{A}\times\mathcal{B}(\mathbb{R}))$, where $(\Omega, \mathcal{A})$ is a general measurable space. Let $\mathcal{T}_\Omega$ be a fixed subset of Ω, we observe that $G : \mathbb{R} \times (\mathcal{T}_\Omega \times \Omega) \to \mathbb{R}^+$ a function such that, $\forall z \in \mathcal{T}_\Omega$, $G(., (z, .))$ measurable function such that $\forall t, t' \in \mathbb{R}$, $(K_0) : t \le t' \Longrightarrow G(t, d) \le G(t', d)$ for $\forall d \in \mathcal{T}_\Omega \times \Omega$. Let $c(.) : \mathcal{T}_\Omega \to \mathbb{R}$ be a non random function such that $\sup_{z\in\mathcal{T}_\Omega} |c(z)| < \infty$. Moreover, for all $z \in \mathcal{T}_\Omega$ and $n \in \mathbb{N}^*$.

So,

$$c_{n,z}(t) = \frac{\sum_{i=1}^{n} \Gamma_z\left(\frac{B_i - x}{\widehat{t}(z)}\right) G(t, (z, A_i))}{\sum_{i=1}^{n} G(t, (z, A_i))}.$$

Lemma 3.1. *Let $\{J_n(z)\}_{n\in N^*}$ be a sequence of r.r.v. and let $(v_n)_{n\ge1}$ be a decreasing positive sequence with $\lim_{n\to\infty} v_n = 0$. If for all increasing sequence $\gamma_n \in (0, 1)$ with $\gamma_n - 1 = O(v_n)$, there exist two sequences of real random variable (r.r.v.) $(J_n^-(\gamma_n, z))_{n\in N^*}$ and $(J_n^+(\gamma_n, z))_{n\in N^*}$ such that*

(K_1) $\forall n \in N^*$, $\forall z \in \mathcal{T}_\Omega$, $J_n^-(\gamma_n, z) \le J_n^+(\gamma_n, z)$,

(K_2) $\mathbb{I}_{\left\{J_n^-(\gamma_n,z)\le J_n(z)\le J_n^+(\gamma_n,z),\ \forall z\in\mathcal{T}_\Omega\right\}} \longrightarrow 1$, *a.co. as* $n \to \infty$,

(K_3) $\sup_{z\in\mathcal{T}_\Omega}\left|\frac{\sum_{i=1}^{n} G(J_n^-(\gamma_n, z), (z, A_i))}{\sum_{i=1}^{n} G(J_n^+(\gamma_n, z), (z, A_i))} - \gamma_n\right| = O_{a.co.}(v_n)$,

(K_4) $\sup_{z\in\mathcal{T}_\Omega} |c_{n,z}(J_n^-(\gamma_n, z)) - c(z)| = O_{a.co}(v_n)$,

(K_5) $\sup_{z\in\mathcal{T}_\Omega} |c_{n,z}(J_n^+(\gamma_n, z)) - c(z)| = O_{a.co}(v_n)$.

Then,

$$\sup_{z\in\mathcal{T}_\Omega} |c_{n,z}(J_n(z)) - c(z)| = O_{a.co}(v_n). \tag{3.1}$$

Proof. The result for any real valued (r.v.) can be deduced by taking $B_i = B_i^+ - B_i^-$ where $B_i^+ = \max(B_i, 0)$ and $B_i^- = -\min(B_i, 0)$ and for i=1,...,n, we consider the quantities $\Gamma_z^i(t) = \Gamma_z\left(\frac{B_i - x}{\widehat{t(z)}}\right)$.
Under the definition of the r.v $J(n)$, we put $J_n^- \leq J_n \leq J_n^+$.

It's clear that

$$G\left(J_n^-\left(\gamma_n, z\right), \left(z, A_i\right)\right), \leqslant G\left(J_n\left(\gamma_n, z\right), \left(z, A_i\right)\right) \leqslant G\left(J_n^+\left(\gamma_n, z\right), \left(z, A_i\right)\right),$$
$$\sum_{i=1}^{n} G\left(J_n^-\left(\gamma_n, z\right), \left(z, A_i\right)\right) \leqslant \sum_{i=1}^{n} G\left(J_n\left(\gamma_n, z\right), \left(z, A_i\right)\right) \leqslant \sum_{i=1}^{n} G\left(J_n^+\left(\gamma_n, z\right), \left(z, A_i\right)\right).$$

So,

$$\frac{1}{\sum_{i=1}^{n} G\left(J_n^+\left(\gamma_n, z\right), \left(z, A_i\right)\right)} \leqslant \frac{1}{\sum_{i=1}^{n} G\left(J_n\left(\gamma_n, z\right), \left(z, A_i\right)\right)} \leqslant \frac{1}{\sum_{i=1}^{n} G\left(J_n^-\left(\gamma_n, z\right), \left(z, A_i\right)\right)}.$$

Under the hypotheses $(A1)-(A5)$, we have

$$\underbrace{\frac{\sum_{i=1}^{n} G\left(J_n^-\left(\gamma_n, z\right), \left(z, A_i\right)\right)\Gamma_z^i(t)}{\sum_{i=1}^{n} G\left(J_n^+\left(\gamma_n, z\right), \left(z, A_i\right)\right)}}_{c_{n,z}^+(\gamma_n)} \leqslant \underbrace{\frac{\sum_{i=1}^{n} G\left(J_n\left(\gamma_n, z\right), \left(z, A_i\right)\right)}{\sum_{i=1}^{n} G\left(J_n\left(\gamma_n, z\right), \left(z, A_i\right)\right)}}_{c_{n,z}(\gamma_n)} \leqslant \underbrace{\frac{\sum_{i=1}^{n} G\left(J_n^+\left(\gamma_n, z\right), \left(z, A_i\right)\right)\Gamma_z^i(t)}{\sum_{i=1}^{n} G\left(J_n^-\left(\gamma_n, z\right), \left(z, A_i\right)\right)}}_{c_{n,z}^-(\gamma_n)}.$$

In the other hand, we can express the r.r.v: $c_{n,z}^-(\gamma_n)$ and $c_{n,z}^+(\gamma_n)$, in the following way:

$$c_{n,z}^-(\gamma_n) = c_{n,z}(\gamma_n) \times \frac{\sum_{i=1}^{n} G(J_n^-(\gamma_n, z), (z, A_i))}{\sum_{i=1}^{n} G(J_n^+(\gamma_n, z), (z, A_i))}$$

and

$$c_{n,z}^+(\gamma_n) = c_{n,z}(\gamma_n) \times \frac{\sum_{i=1}^{n} G(J_n^+(\gamma_n, z), (z, A_i))}{\sum_{i=1}^{n} G(J_n^-(\gamma_n, z), (z, A_i))}.$$

So under (K_2) and (K_3), we have

$$c_{n,z}^-(\gamma_n) \xrightarrow{a.co.} \gamma_n\, c(z), \qquad \text{and} \qquad c_{n,z}^+(\gamma_n) \xrightarrow{a.co.} c(z)/\gamma_n. \tag{3.2}$$

For all sequence $\gamma_n \in (0,1)$ with $\gamma_n - 1 = O(v_n)$, (K_3), (K_4) and (K_5) give

$$\sup_{z\in\mathcal{T}_\Omega} |c_{n,z}^-(\gamma_n) - c(z)| \leq \sup_{z\in\mathcal{T}_\Omega} |c_{n,z}^-(\gamma_n) - \gamma_n c(z)| + |c(z)||\gamma_n - 1| = O_{a.co.}(v_n) \tag{3.3}$$

and

$$\sup_{z\in\mathcal{T}_\Omega} |c_{n,z}^+(\gamma_n) - c(z)| = O_{a.co.}(v_n). \tag{3.4}$$

For $\varepsilon > 0$ we note

$$T_n(\varepsilon) = \left\{\sup_{z\in\mathcal{T}_\Omega} |c_{n,z}(J_n(z)) - c(z)| \leq \varepsilon v_n\right\},$$

and for all sequence $\gamma_n \in (0,1)$ with $\gamma_n - 1 = O(v_n)$,

$$S_n^-(\varepsilon, \gamma_n) = \left\{\sup_{z\in\mathcal{T}_\Omega} |c_{n,z}^-(\gamma_n) - c(z)| \leq \varepsilon v_n\right\},$$

$$S_n^+(\varepsilon, \gamma_n) = \left\{ \sup_{z \in \mathcal{T}_\Omega} |c_{n,z}^+(\gamma_n) - c(z)| \le \varepsilon v_n \right\},$$
$$S_n(\gamma_n) = \left\{ c_{n,z}^-(\gamma_n) \le c_{n,z}(J_n(z)) \le c_{n,z}^+(\gamma_n), \forall z \in \mathcal{T}_\Omega \right\}.$$

It is evident that, for all $\gamma_n \in (0,1)$ with $\gamma_n - 1 = O(v_n)$,

$$\forall \varepsilon > 0, \ S_n^-(\varepsilon, \gamma_n) \cap S_n^+(\varepsilon, \gamma_n) \cap S_n(\gamma_n) \subset T_n(\varepsilon). \tag{3.5}$$

Let $G_n(\gamma_n) = \{J_n^-(\gamma_n, z) \le J_n(z) \le J_n^+(\gamma_n, z), \forall z \in \mathcal{T}_\Omega\}$, then (K_0) implies that $G_n(\gamma_n) \subset S_n(\gamma_n)$ and from (3.5), we obtain

$$\forall \varepsilon > 0, \ T_n(\varepsilon)^c \subset S_n^-(\gamma_n)^c \cup S_n^+(\gamma_n)^c \cup G_n(\gamma_n)^c,$$

and consequently

$$\begin{aligned}\mathbb{P}\Big(\sup_{z \in \mathcal{T}_\Omega} |c_{n,z}(J_n(z)) - c(z)| > \varepsilon v_n \Big) &\le \mathbb{P}\Big(\sup_{z \in \mathcal{T}_\Omega} |c_{n,z}^-(\gamma_n, \varepsilon) - c(z)| > \varepsilon v_n \Big) \\ &+ \mathbb{P}\Big(\sup_{z \in \mathcal{T}_\Omega} |c_{n,z}^+(\gamma_n, \varepsilon) - c(z)| > \varepsilon v_n \Big) \\ &+ \mathbb{P}\Big(\mathbb{I}_{\left\{ J_n^-(\gamma_n, z) \le J_n(z) \le J_n^+(\gamma_n, z), \ \forall z \in \mathcal{T}_\Omega \right\}} = 0 \Big).\end{aligned}$$

Then, for some $\varepsilon_0 > 0$

$$\sum_{n=1}^{\infty} \mathbb{P}\Big(\sup_{z \in \mathcal{T}_\Omega} \Big| c_{n,z}(J_n(z)) - c(z) \Big| > \varepsilon_0 v_n \Big) < \infty. \tag{3.6}$$

□

Remark 3.2. We wish to present two results, similar to Lemma 3.1, and that could be interesting for further purposes.

- **(i):** Under the same conditions, the result stated in Lemma 3.1 holds by changing all the almost complete convergence into convergence in probability.
- **(ii):** Under the same conditions, the result stated in Lemma 3.1 holds by changing all the $O_{\text{a.co.}}$ into $o_{\text{a.co.}}$.

The proof of (i) is the same as the one of Lemma 3.1, changing (3.2) into the fact that the sequence involved in () tends to zero. The proof of (ii) is similar.

4. Main result

We start by reminding the uniform asymptotic properties of $\widehat{\mu}(z, x, \widehat{t}(z))$ defined in (2.6). The Theorem 4.1 defined bellow was proved by [2] in the special case when $h_L(z) = h_L$ for all $z \in \mathcal{T}_\mathcal{F}$, but their proof can be followed line by line under (4.2)). This general condition (4.2) will be a crucial preliminary tool for us.

Theorem 4.1. *Under assumptions (A1)-(A8), if in addition, $h_L(z)$ in ((2.6)) satisfies*

$$\lim_{n \to \infty} (\varphi_z(T_{n,k}(z)) - \varphi_z(h_L(z))) = 0, \ a.co. \tag{4.1}$$

and

$$0 < C h_L \le \inf_{z \in \mathcal{T}_\mathcal{F}} h_L(z) \le \sup_{z \in \mathcal{T}_\mathcal{F}} h_L(z) \le C' h_L < \infty, \tag{4.2}$$

where $h_L \to 0$ $(n \to \infty)$ such that, for n large enough,

$$\frac{\log^2 n}{n \phi(h_L)} < \Psi_{\mathcal{T}_\mathcal{F}}\Big(\frac{\log n}{n} \Big) < \frac{n \phi(h_L)}{\log n}, \tag{4.3}$$

and

$$0 < C < \frac{n\varphi(h_L)}{\log^2 n} < C' < +\infty. \tag{4.4}$$

Then, we have

$$\sup_{z\in\mathfrak{I}_{\mathcal{F}}} \left|\widehat{\mu}(z,x,\widehat{t}(z)) - \mu(z,x,\widehat{t}(z))\right| = O(h_L^{\beta}) + O_{a.co}\left(\sqrt{\frac{\Psi_{\mathfrak{I}_{\mathcal{F}}}(\frac{\log n}{n})}{n\phi(h_L)}}\right). \tag{4.5}$$

We can now state our main result, whose proof will be presented in Section 5.

Theorem 4.2. *Under the assumptions (A1)-(A9), and for n large enough, we have*

$$\sup_{z\in\mathfrak{I}_{\mathcal{F}}} \left|\widehat{\mu}_{kNN}(z,x,\widehat{t}(z)) - \mu(z,x,\widehat{t}(z))\right| = O\left(\varphi^{-1}\left(\frac{k}{n}\right)^{\beta}\right) + O_{a.co}\left(\sqrt{\frac{\Psi_{\mathfrak{I}_{\mathcal{F}}}(\frac{\log n}{n})}{k}}\right).$$

Remark 4.3. On the rates of convergence. First of all it is worth noting that, by taking k of order $n\phi(h_L)$, the kNN estimate reaches the same rate of convergence as the kernel estimate does (see Theorem 4.1). More importantly, to attest the quality of these rates, it suffices to look at the case $\mathcal{F} = R^q$ to see that they are exactly matching the rate $(\log n/n)^{b/(2b+q)}$ which is optimal for q-dimensional functions (see [44]). Note also that, for the exponential-type processes described before the rate of convergence may look quite slow for unfamiliar people (of order $(\log n)^{-\alpha}$ for some $\alpha > 0$) but this is true only when using as "d" a standard norm; other kinds of d can be used to improve strongly these rates, as discussed in ([25], Lemma 13.6).

5. Proofs of main result

Proof. Similar to the proof of ([32], Theorem 2), we must to investigate the conditions of Lemma 3.1.
For that, we denote: $\mathfrak{I}_\Omega = \mathfrak{I}_{\mathcal{F}}$, $A_i = Z_i$, $B_i = W_i$, $G(t,(z,A_i)) = L(d(z,A_i)/t)$, $J_n(z) = T_{n,k}(z)$, $c_{n,z}(T_{n,k}(z)) = \widehat{\mu}_{kNN}(z,x,\widehat{t}(z))$ and $c(z) = \mu(z,x,\widehat{t}(z))$. We begin by recalling that the estimate

$$\widehat{\mu}_N(z) = \frac{1}{nEL\left(h_L^{-1}d\left(z,Z_1\right)\right)} \sum_{i=1}^{n} L\left(h_L^{-1}d\left(z,Z_i\right)\right),$$

satisfies under the conditions of ([2], p. 11, Lemma A.1.)

$$\sup_{z\in\mathfrak{I}_{\mathcal{F}}} |\widehat{\mu}_N^1(z) - 1| = O_{a.co}\left(\sqrt{\frac{\Psi_{\mathfrak{I}_{\mathcal{F}}}(\frac{\log n}{n})}{n\phi(h_L)}}\right). \tag{5.1}$$

Let $\gamma_n \in (0,1)$ be an increasing sequence such that $\gamma_n - 1 = O(v_n)$, where

$$v_n = \phi^{-1}\left(\frac{k}{n}\right)^{\beta} + \sqrt{\frac{\Psi_{\mathfrak{I}_{\mathcal{F}}}(\frac{\log n}{n})}{k}}.$$

Let $h_L = \phi^{-1}\left(\frac{k}{n}\right)^{\beta}$, we choose $J_n^-(\gamma_n, z)$ and $J_n^+(\gamma_n, z)$ such that

$$\varphi_z(J_n^-(\gamma_n,z)) = \varphi_z(h_L(z))\gamma_n^{1/2} \tag{5.2}$$

and

$$\varphi_z(J_n^+(\gamma_n,z)) = \varphi_z(h_L(z))\gamma_n^{-(1/2)}. \tag{5.3}$$

Checking (K_4) and (K_5): we note that the local bandwidth $J_n^-(\gamma_n, z)$ satisfies (4.1), (4.2) and (4.4), we have now

$$\sup_{z\in\mathfrak{T}_{\mathcal{F}}} |c_{n,z}(J_n^-(\gamma_n, z)) - c(z)| = O_{a.co}\left(\phi^{-1}\left(\frac{k}{n}\right)^\beta + \sqrt{\frac{\Psi_{\mathfrak{T}_{\mathcal{F}}}(\frac{\log n}{n})}{k}}\right)$$
$$= O_{a.co}(v_n).$$

Consequently, (K_4) in Lemma 3.1 is valid. We use the same steps for $J_n^+(\gamma_n, z)$, we obtain

$$\sup_{z\in\mathfrak{T}_{\mathcal{F}}} |c_{n,z}(J_n^+(\gamma_n, z)) - c(z)| = O_{a.co}(v_n).$$

Therefore (K_5) is also correct.

We check (K_1) and (K_2): with (5.2) and (5.3), we have $\varphi_z(J_n^-(\gamma_n, z)) \le \varphi_z(h_L(z)) \le \varphi_z(J_n^+(\gamma_n, z))$. Thus, with (4.1) and the property of $\varphi_z(.)$, we have as a result

$$J_n^-(\gamma_n, z) \le T_{n,k}(z) \le J_n^+(\gamma_n, z), \text{ a.co.}$$

and

$$\mathbb{I}_{\left\{J_n^-(\gamma_n,z)\le J_n(z)\le J_n^+(\gamma_n,z), \forall z\in\mathfrak{T}_\Omega\right\}} \to 1, \text{ a.co. as } n\to\infty.$$

Consequently, (K_1) and (K_2) in Lemma 3.1 are valid.

Checking (K_3): Same as [32], $\forall z \in \mathfrak{T}_{\mathcal{F}}$, indicate

$$f^*(z, h_L(z)) =: \mathbb{E}\Big[L\Big(\frac{d(z,z_1)}{h_L(z)}\Big)\Big], \quad H_1 =: \frac{f^*(z, J_n^-(\gamma_n, z))}{f^*(z, J_n^+(\gamma_n, z))},$$
$$H_2 =: \frac{\widehat{\mu}_N^1(z, J_n^-(\gamma_n, z))}{\widehat{\mu}_N^1(z, J_n^+(\gamma_n, z))} - 1, \quad H_3 =: \frac{f^*(z, J_n^-(\gamma_n, z))}{f^*(z, J_n^+(\gamma_n, z))}\gamma_n - 1.$$

After that, we come at

$$\left|\frac{\sum_{i=1}^n L\Big(\frac{d(z,z_i)}{J_n^-(\gamma_n,z)}\Big)}{\sum_{i=1}^n L\Big(\frac{d(z,z_i)}{J_n^+(\gamma_n,z)}\Big)} - \gamma_n\right| \le |H_1||H_2| + |H_1||H_3|. \tag{5.4}$$

With (A3), we get

$$\sup_{z\in\mathfrak{T}_{\mathcal{F}}} |H_1| \le C. \tag{5.5}$$

In addition, with (5.1), we have that

$$\sup_{z\in\mathfrak{T}_{\mathcal{F}}} |H_2| \le \frac{\sup_{z\in\mathfrak{T}_{\mathcal{F}}} |\widehat{\mu}_N^1(z, J_n^-(\gamma_n, z)) - 1| + \sup_{z\in\mathfrak{T}_{\mathcal{F}}} |\widehat{\mu}_N^1(z, J_n^+(\gamma_n, z)) - 1|}{\inf_{z\in\mathfrak{T}_{\mathcal{F}}} |\widehat{\mu}_N^1(z, J_n^+(\gamma_n, z))|}$$
$$= O_{a.co}\left(\sqrt{\frac{\Psi_{\mathfrak{T}_{\mathcal{F}}}(\frac{\log n}{n})}{k}}\right). \tag{5.6}$$

Then, we use ([24], Lemma 1, p. 10) with (A2), and also the fact that

$$\varphi_z(J_n^-(\gamma_n, z))/\varphi_z(J_n^+(\gamma_n, z)) = \gamma_n,$$

we get

$$\sup_{z\in\mathfrak{I}_{\mathcal{F}}}|H_3| = O(\phi(h_L)h_L^{\beta}) = O\left(\left(\sqrt{\gamma_n}\phi^{-1}\left(\frac{k}{n}\right)\right)^{\beta}\right). \tag{5.7}$$

So, (K_3) is checked because $\gamma_n \to 1$ and because (5.4)-(5.5) and (5.6) imply that

$$\sup_{z\in\mathfrak{I}_{\mathcal{F}}}\left|\frac{\sum_{i=1}^{n} L\left(\frac{d(z,z_i)}{J_n^{-}(\gamma_n,z)}\right)}{\sum_{i=1}^{n} L\left(\frac{d(z,z_i)}{J_n^{+}(\gamma_n,z)}\right)} - \gamma_n\right| = O_{a.co}(v_n).$$

Note that (K_0) is obviously satisfied because of (A3a), and that (K_1) is also easily satisfied by construction of $J_n^{-}(\gamma_n,z)$ and $J_n^{+}(\gamma_n,z)$. So, one can apply Lemma 3.1, and (3.1) is precisely the result of Theorem 4.2. □

6. Simulated data application

This section aims to show the efficiency of the proposed estimator in terms of consistency. The first direct use of the Theorem 4.2 is to predicting a functional time series processes.

Let $(X_t)_{t\in[0,b[}$ be a continuous-time real-valued random process. From the process Z_t, we may construct N functional random variables $(Z_i)_{i=1,\ldots,N}$ defined by

$$\forall t \in [0,b], \quad Z_i(t) = X_{N^{-1}((i-1)b+c)}.$$

The predictor estimator of W is defined by $\widehat{W} = \widehat{\mu}_{kNN}(z,x,\hat{t}(z))\,(Z_N)$.

To do that, we consider the following functional nonparametric model

$$W_i = r(Z_i) + \varepsilon_i \qquad for \qquad i = 1,\ldots,n,$$

where the ε_i's are generated independently according to a normal distribution with mean 0.

Let us now describe how our functional ergodic data are generated. Firstly, We use the R-routine *simul.far* of *far* package in *R* software to generate the functional explanatory variables $(Z_i)_{i=1,\ldots,n}$. This routine simulates a functional autoregressive process white Wiener noise.

For this simulation experiments, we have considered sinusoidal basis, with five functional axis, of the continuous functions from $[0,1]$ to $\mathbb{R}$. Recall that, it is shown in [34] that this kind of process satisfies the ergodicity condition. The curves Z_i's are discretized in the same grid composed by 100 points and are plotted in Figure 1.

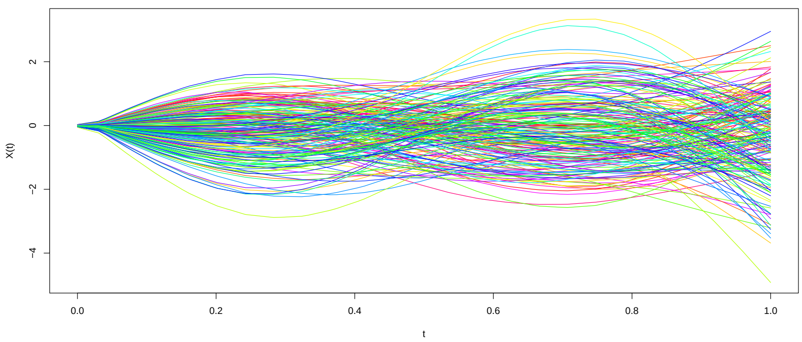

Figure 1. A sample of 100 curves, for $d_{\rho} = (0.45, 0.90, 0.34, 0.45)$.

Secondly, the scalar response W_i is computed by considering the following operator

$$r(z) = \int_0^1 \frac{10}{(1 + |z(t)|)} dt.$$

Our main goal is to compare our estimator (Robust Equivariant Estimator REE) $\widehat{\theta}(z)$ with Robust Kernel Estimator (RKE) $\tilde{\theta}(z)$ introduced by [8], and the Classical Kernel Estimator (CKE) presented by [25], where $\widehat{\theta}(z)$, $\tilde{\theta}(z)$ and $\widehat{m}(z)$ are define as following:

$\widehat{\theta}(z)$ is the zero with respect to x of $\dfrac{\sum_{i=1}^{n} L\left(\dfrac{d(z, Z_i)}{T_{n,k}(z)}\right) \Gamma_z \left(\dfrac{W_i - w}{\widehat{t}(z)}\right)}{\sum_{i=1}^{n} L\left(\dfrac{d(z, Z_i)}{T_{n,k}(z)}\right)} = 0,$

$\tilde{\theta}(z)$ is tknhe zero with respect to x of $\dfrac{\sum_{i=1}^{n} L\left(\dfrac{d(z, Z_i)}{T_{n,k}(z)}\right) \Gamma_z (W_i - w)}{\sum_{i=1}^{n} L\left(\dfrac{d(z, Z_i)}{T_{n,k}(z)}\right)} = 0,$

and $\widehat{m}(z) = \dfrac{\sum_{i=1}^{n} W_i L\left(\frac{d(z, Z_i)}{T_{n,k}(z)}\right)}{\sum_{i=1}^{n} L\left(\dfrac{d(z, Z_i)}{T_{n,k}(z)}\right)}.$

The efficiency of the predictors is evaluated by the empirical Mean Squared Error (MSE)

$$MSE_{\widehat{\theta}} = n^{-1} \sum_{i=1}^{n} (\mu(Z_i) - \widehat{\mu}(Z_i))^2, \; MSE_{\tilde{\theta}} = n^{-1} \sum_{i=1}^{n} \left(\theta(Z_i) - \tilde{\theta}(Z_i)\right)^2$$

and

$$MSE_{\widehat{m}} = n^{-1} \sum_{i=1}^{n} (\theta(Z_i) - \widehat{m}(Z_i))^2.$$

Through this simulation study, we chose the quadratic kernel L defined as $L(u) = \frac{3}{4}(1 - u^2)\, \mathbb{1}_{[0,1]}(u)$. The used semi-metric is the first derivative of sample curves, given by

$$d(Z_i, Z_j) = \sqrt{\int \left(Z_i'(t) - Z_j'(t)\right)^2 dt}.$$

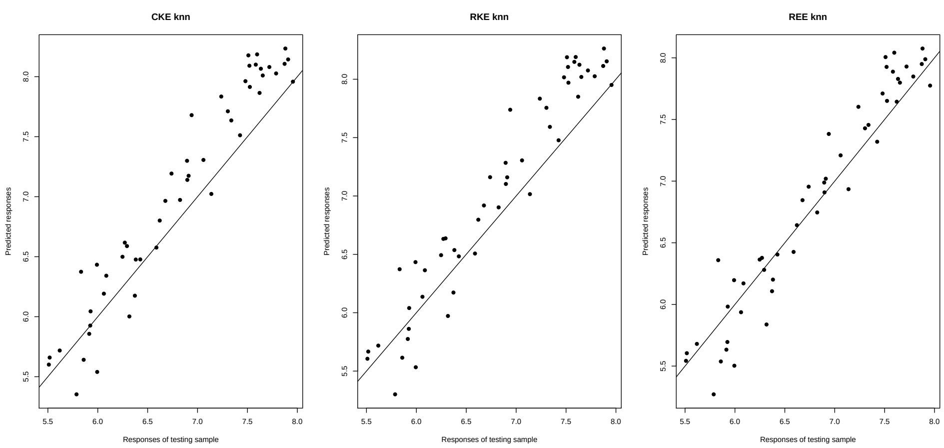

Figure 2. Predictions of the three models.

For this comparison study, we treat three estimators in the same conditions. The first illustration concerns the MSE of $\widehat{\theta}(z)$. Further we see from Figure 2 that the REE kNN estimator is much more better than the CKE kNN and the RKE estimators. Moreover, looking at both Figures, it appears clearly the MSE of REE kNN has dramatically changed compared to the classical and robust kernel estimators. So REE kNN is much more performance than the others. Furthermore, when the MSE error is taken into account, the superiority of this model becomes even more apparent.

Now, we carry out 100 independent replication with $n-$ samples($n = 200$) of the same datas for MSE and to display their distribution by means of a boxplot. Figure 3 shows the boxplots of the MSE of the prediction values.

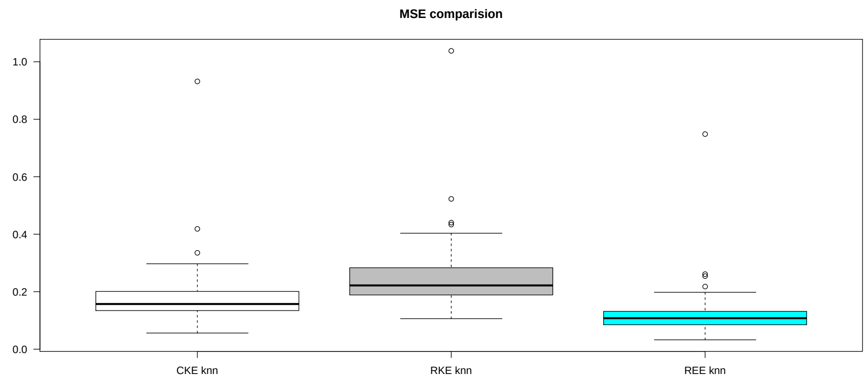

Figure 3. MSE of the three models.

Thus, the MSE comparison for the three methods illustrated in Figure 3 prove that the forecasting of REE kNN estimator is more accurate than the other methods.

Now, we will compare between kNN method and cross validation (CV) procedures in the presence of the outlier. In what follows, we randomly split the 200-sample into two parts: one is a training sample $(Z_i, W_i)_{i=1}^{100}$ which is used to model, and the other is a testing sample $(Z_j, W_j)_{j=101}^{200}$ which is used to verify the prediction effect. On the one hand, by the training sample, we can select the optimal parameter k_{opt} for kNN kernel and robust estimator, and the optimal parameter h_{opt} for CV classical kernel and robust estimator by the following CV procedures, respectively.

Concretely, we select $k_{opt} = \arg\min_k CV_1(k)$, where $CV_1(k) = \sum_{i=1}^{n}\left(W_i - \widehat{m}^{(-i)}(Z_i)\right)^2$, and

$$\hat{m}^{(-i)}(Z) = \frac{\sum_{j=1,j\neq i}^{n} W_j L\left(\frac{d\left(Z_j, Z\right)}{T_{n,k}(Z)}\right)}{\sum_{j=1,j\neq i}^{n} L\left(\frac{d(Z_j; Z)}{T_{n,k}(Z)}\right)},$$

and the robust kNN one by $k_{opt} = \arg\min_k CV_2(k)$, where $CV_2(k) = \sum_{i=1}^{n}\left(W_i - \widehat{\theta}^{(-i)}(Z_i)\right)^2$, and

$$\hat{\theta}^{(-i)}(Z) = \arg\min_t \frac{\sum_{j=1,j\neq i}^{n} \Gamma_z\left(\frac{W_i - w}{\hat{t}(z)}\right) L\left(\frac{d\left(Z_j, Z\right)}{T_{n,k}(Z)}\right)}{\sum_{j=1,j\neq i}^{n} L\left(\frac{d(Z_j, Z)}{T_{n,k}(Z)}\right)}.$$

Then the robust kernel estimator by $k_{opt} = \arg\min_k CV_3(k)$, where $CV_3(k) = \sum_{i=1}^{n} \left(W_i - \tilde{\theta}^{(-i)}(Z_i)\right)^2$, and

$$\tilde{\theta}^{(-i)}(Z) = \arg\min_t \frac{\sum_{j=1, j\neq i}^{n} \Gamma_z (W_i - x) L\left(\frac{d(Z_j, Z)}{T_{n,k}(Z)}\right)}{\sum_{j=1, j\neq i}^{n} L\left(\frac{d(Z_j, Z)}{T_{n,k}(Z)}\right)}.$$

Similarly, We adopt the selection rule, proposed by [25] where $h_{opt} = \arg\min_{h_L} CV(h_L)$, where $CV(h_L) = \sum_{i=1}^{n} \left(W_i - \widehat{m}_{(-i)}^{kernel}(Z_i)\right)^2$, and

$$\hat{m}_{(-i)}^{\text{kernel}}(Z) = \frac{\sum_{j=1, j\neq i}^{n} W_j L\left(\frac{d(Z_j, Z)}{h_L}\right)}{\sum_{j=1, j\neq i}^{n} L\left(\frac{d(Z_j, Z)}{h_L}\right)}$$

with $\hat{m}_{(-i)}^{\text{kernel}}(.)$ is the leave-one-out cross-validation (values of the estimator $\widehat{\theta}(.)$ calculate at Z_i) (see [25] for more details).

The main feature of our approach is illustrated in the second case when we perturb the data by introducing some outliers as indicated bellow. In this part, we simulated data with three values of the multiplier $MC = 0$ or 5 or 20 (MC is the number of the perturbed observations). In all six cases, we arrived at the same conclusion, usually, in the presence of outliers, the robust regression shows better behavior than that of the classical method. Even if the MSE of the both methods increases substantially with the number of perturbed points and with value of multiplicative coefficient MC, it remains very low for the robust kNN method. The results obtained in the Table 1 which shows the superiority of the kNN method to that of the CV method, and the good behavior of our functional forecasting procedure for the robust kNN method in presence of outliers.

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

MC	CKE CV	CKE kNN	RKE CV	RKE kNN	REE CV	REE kNN
0	0.3133	0.1856	0.3424	0.2585	0.1733	0.1251
5	1697.3769	2021.4049	247.2289	0.4144	147.8625	0.1350
20	20854.6054	23712.1086	775.2927	156.8678	271.5004	0.1758

7. Conclusion

The uniform kNN reliability approach is a smoothing alternative that allows for the development of an adaptive estimator for a variety of statistical problems, including bandwidth choice.

In our situation, furthermore, uniform consistency is not a straightforward extension of the pointwise approach, as it necessitates the use of additional methods and techniques. The assumption that the bandwidth parameter in the kNN method is a random variable adds to the complexity of this problem.

In the situation of equivariant robustification results, the key innovation of this approach is to estimate the regression function by mixing two essential statistical techniques: the regression estimators for ergodic data when the scale parameter is unknown with the kNN method. This strategy allowed for the development of a new estimator that combines the benefits of both methods.

Another unanswered concern is how to treat a more general case in which data are generated from a functional *alpha*-mixing dependency and the scale parameter is unknown. Precisely, we can obtain the uniform almost complete convergence of the same constructed estimator under standard conditions allowing us to explore different structural axes of the topic. We emphasize that, contrary to the usual case when the scale parameter is fixed, it must be estimated, which makes it more difficult to establish the uniform almost complete convergence of the estimator.

To summarize, the behavior of the developed estimator is unaffected by the number of outliers in the data collection. In comparison to the classical kernel method, the mixture of the kNN algorithm and the robust method allows for a reduction in the impact of outliers in results.

Acknowledgment. The authors would like to thank the referee for his constructive suggestions that have led to improve the presentation of this paper.

References

[1] I. Almanjahie, K. Aissiri, A. Laksaci and Z. Chiker el Mezouar, *The k nearest neighbors smoothing of the relative-error regression with functional regressor*, Comm. Statist. Theory Methods **51** (12), 4196-4209, 2022.

[2] I. Almanjahie, M. Attouch, Z. Kaid and H. Louab, *Robust equivariant nonparametric regression estimators for functional ergodic data*, Comm. Statist. Theory and Methods **50** (20), 3505-3521, 2020.

[3] M. Attouch and W. Bouabsa, *The k-nearest neighbors estimation of the conditional mode for functional data*, Rev. Roumaine Math. Pures Appl. **58** (4), 393-415, 2013.

[4] M. Attouch, W. Bouabsa and Z. Chiker el Mozouar, *The k-nearest neighbors estimation of the conditional mode for functional data under dependency*, International Journal of Statistics & Economics **19** (1), 48-60, 2018.

[5] M. Attouch, A. Laksaci and E. Ould Saïd, *Asymptotic distribution of robust estimator for functional non parametric models*, Comm. Statist. Theory and Methods **45** (15), 230-287, 2009.

[6] M. Attouch, A. Laksaci and E. Ould Saïd, *Asymptotic normality of a robust estimator of the regression function for functional time series data*, J. Korean Statist. Soc. **39** (22), 489-500, 2010.

[7] M. Attouch, A. Laksaci and E. Ould Saïd, *Robust regression for functional time series data*, J. Jpn. Stat. Soc. **42** (2), 125-143, 2012.

[8] N. Azzedine, A. Laksaci and E. Ould Saïd, *On the robust nonparametric regression estimation for functional regressor*, Statist. Probab. Lett. **78** (2), 3216-3221, 2008.

[9] J. Beirlant, A. Berlinet and G. Biau, *Higher order estimation at Lebesgue points*, Ann. Inst. Statist. Math. **90** (60), 651-677, 2008.

[10] F. Benziadi and A. Laksaci, *Recursive kernel estimate of the conditional quantile for functional ergodic data*, Comm. Statist. Theory and Methods **45** (12), 3097-3113, 2015.

[11] G. Boente and R. Fraiman, *Robust nonparametric regression estimation for dependent observations*, Ann. Statist. **17** (5), 1242-1256, 1989.

[12] G. Boente and A. Vahnovan, *Strong convergence of robust equivariant nonparametric functional regression estimators*, Statist. Probab. Lett. **100** (55), 1-11, 2015.

[13] D. Bosq, *Linear process in function spaces: theory and applications*, Lecture Notes in Statistics **149**, Springer-Verlag, 2000.

[14] W. Bouabsa, *Nonparametric relative error estimation via functional regressor by the k nearest neighbors smoothing under truncation random*, Appl. Appl. Math. **16** (1), 97-116, 2021.

[15] F. Burba, F. Ferraty and P. Vieu, *k-nearest neighbour method in functional nonparametric regression*, J. Nonparametr. Stat. **21** (4), 453-469, 2009.

[16] J. Cheng and L. Zhang, *Asymptotic properties of nonparametric M-estimation for mixing functional data*, J. Statist. Plann. Inference **139** (70), 533-546, 2009.

[17] G. Collomb, *Estimation non paramétrique de la régression: revue bibliographique*, Int. Stat. Rev. **49** (1), 75-93, 1981.

[18] G. Collomb and H. Härdale, *Strong uniform convergence rates in robust nonparametric time series analysis and prediction: Kernel regression estimation from dependent observations*, Stochastic Process. Appl. **23** (7), 77-89, 1979.

[19] T.M. Cover, *Estimation by the nearest neighbor rule*, IEEE Trans. Inform. Theory **7** (14), 50-55, 1968.

[20] C. Crambes and A. Laksaci, *Robust nonparametric estimation for functional data*, J. Nonparametr. Stat. **20** (5), 573-598, 2008.

[21] L. Devroye, L. Györfi, A. Krzyzak and G. Lugosi, *On the strong universal consistency of nearest neighbor regression function estimates*, Ann. Statist. **20** (22), 1371-1385, 1994.

[22] L. Devroye and T. Wagner, *The strong uniform consistency of nearest neighbor density*, Ann. Statist. **10** (5), 536-540, 1977.

[23] L. Devroye and T. Wagner, *Nearest neighbor methods in discrimination*, in: Handbook of Statistics 2: Classification, Pattern Recognition and Reduction of Dimensionality, North-Holland, Amsterdam, 1982.

[24] M. Ezzahrioui and E. Ould Saïd, *Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data*, J. Nonparametr. Stat. **20** (1), 3-18, 2008.

[25] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis: Theory and Practice*, Springer-Verlag, New York, 2006.

[26] G. Geenens, *Curse of dimensionality and related issues in nonparametric functional regression*, Stat. Surv. **5** (1), 30-43, 2011.

[27] A. Gheriballah, A. Laksaci and S. Sekkal, *Nonparametric M-regression for functional ergodic data*, Statist. Probab. Lett. **83** (20), 902-908, 2013.

[28] L. Györfi, Kohler, A. Krzyzak and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer, New York, 2002.

[29] P.J. Hurber, *Robust estimation of a location parameter*, in: S. Kotz and N.L. Johnson (ed.) Breakthroughs in Statistics, New York, Springer Science & Business Media, 1992.

[30] Z. Kara, A. Laksaci and P. Vieu, *Data-driven kNN estimation in nonparametric functional data analysis*, J. Multivariate Anal. **153** (85), 176-188, 2017.

[31] A.N. Kolmogorov and V. M. Tikhomirov, *ε-entropy and ε-capacity of sets in function spaces*, Uspekhi Mat. Nauk. **14** (2), 3-86, 1959.

[32] N.L. Kudraszow and P. Vieu, *Uniform consistency of kNN regressors for functional variables*, Statist. Probab. Lett. **83** (26), 1863-1870, 2013.

[33] N. Laïb and E. Ouled-Saïd, *A robust nonparametric estimation of the autoregression function under an ergodic hypothesis*, Canad. J. Statist. **28** (9), 817-828, 2000.

[34] N. Laïb and D. Louani, *Strong consistency of the regression function estimator for functional stationary ergodic data*, J. Statist. Plann. Inference **141** (59), 359-372, 2011.

[35] T. Laloë, *A k-nearest neighbor approach for functional regression*, Statist. Probab. Lett. **78** (10), 1189-1193, 2008.

[36] J.P. Li, *Strong convergence rates of error probability estimation in the nearest neighbor discrimination rule*, J. Math. **15** (5), 113-118, 1985.

[37] H. Lian, *Convergence of functional k-nearest neighbor regression estimate with functional responses*, Electron. J. Stat. **5** (2), 31-40, 2011.

[38] N. Ling, S. Meng and P. Vieu, *Uniform consistency rate of kNN regression estimation for functional time series data*, J. Nonparametr. Stat. **31** (5), 451-468, 2019.

[39] N. Ling and P. Vieu, *Nonparametric modelling for functional data : selected survey and tracks for future, statistics*, J. Nonparametr. Stat. **52** (4), 20-30, 2018.

[40] D. Moore and J. Yackel, *Consistency properties of nearest neighbor density function estimators*, Ann. Statist. **20** (5), 143154, 1977.

[41] S. Müller and J. Dippon, *kNN kernel estimate for nonparametric functional regression in time series analysis*, Technical Report, University of Stuttgart, Fachbereich Mathematik, 2011.

[42] G.G. Roussas, *Kernel estimates under association: strong uniform consistency*, Statist. Probab. Lett. **12** (9), 215-224, 1991.

[43] M. Samanta, *Non-parametric estimation of conditional quantiles*, Statist. Probab. Lett. **7** (5), 407-412, 1989

[44] C.J. Stone, *Optimal global rates of convergence for nonparametric regression*, Ann. Statist. **10** (4), 1040-1053, 1982.

[45] T. Tran, R. Wehrens and L. Buydens, *kNN-kernel density-based clustering for high-dimensional multivariate data*, Comput. Statist. Data Anal. **51** (50), 513-525, 2006.