Computers, Materials & Continua CMC, vol.65, no.3, pp.2049-2064, 2020

The k Nearest Neighbors Estimator of the M-Regression in Functional Statistics

Ahmed Bachir[1, *], Ibrahim Mufrah Almanjahie[1, 2] and Mohammed Kadi Attouch[3]

Abstract: It is well known that the nonparametric estimation of the regression function is highly sensitive to the presence of even a small proportion of outliers in the data. To solve the problem of typical observations when the covariates of the nonparametric component are functional, the robust estimates for the regression parameter and regression operator are introduced. The main propose of the paper is to consider data-driven methods of selecting the number of neighbors in order to make the proposed processes fully automatic. We use the k Nearest Neighbors procedure (kNN) to construct the kernel estimator of the proposed robust model. Under some regularity conditions, we state consistency results for kNN functional estimators, which are uniform in the number of neighbors (UINN). Furthermore, a simulation study and an empirical application to a real data analysis of octane gasoline predictions are carried out to illustrate the higher predictive performances and the usefulness of the kNN approach.

Keywords: Functional data analysis, quantile regression, kNN method, uniform nearest neighbor (UNN) consistency, functional nonparametric statistics, almost complete convergence rate.

1 Introduction

Let us introduce n pairs of random variables (X_i, Y_i), for $i = 1, \ldots, n$, that we assume drawn from the pair (X, Y), which is valued in $\mathcal{F} \times \mathbb{R}$, where $\mathcal{F}$ is a semi-metric space equipped with a semi-metric d. Our purpose is to evaluate the impact of the functional variable X on the real variable Y using the robust estimation of the regression function. Recall that, the relationship between X and Y is usually modeled by the classical regression which is given by $Y = r(X) + \epsilon$, where ϵ represents an independent random variable of X with a symmetric distribution. However, in the alternative case, when the distribution is not symmetric, the classical regression is not adequate. In this case, robust regression is more

[1] Department of Mathematics, College of Science, King Khalid University, Abha, 62529, Saudi Arabia.

[2] Statistical Research and Studies Support Unit, King Khalid University, Abha, 62529, Saudi Arabia.

[3] Djillali Liabes University, Sidi Belabbes, 22000, Algeria.

[*] Corresponding Author: Ahmed Bachir. Email: bachir_ahmed@hotmail.com; abishr@kku.edu.sa.

Received: 11 May 2020; Accepted: 19 August 2020.

appropriate. Moreover, this last is more often used than the classical regression because it is insensitive to outliers presented in a given data and resisting heteroscedasticity phenomena.

Historically, the robust method was introduced by Maronna et al. [Maronna, Douglas Martin, Yohai et al. (2019)] and defined, for any loss function $\rho(.,.)$ on IR, as the unique minimizer with respect to (w.r.t.) the component t in the model $\Gamma_x(t) = \mathbb{E}[\rho(Y,t)|X = x]$. The theoretical estimator of this model is defined by

$$\theta_x = \underset{t\in\mathbb{R}}{\operatorname{argmin}}\, \Gamma_x(t). \tag{1}$$

Furthermore, the statistical model, $\Gamma_x(t)$, deals with a large class of M-regression models based on the types of $\rho(y,t)$. For examples, $\rho(y,t) = (y-t)^2$ yields the classical regression, $\rho(y,t) = |y-t|$ gives the conditional median function $m(x) = med(Y|X = x)$, and $\rho(y,t) = |y-t| + (2\alpha-1)(y-t)$ leads to the α^{th}-conditional quantile; for further examples of the function ρ[Stone (2000)].

According to Eq. (1), the best approximation of Y given X is based on the estimation of θ_x denoted by $\hat{\theta_x}$, given by $\hat{\theta_x} = \underset{t\in\mathbb{R}}{\operatorname{argmin}}\, \hat{\Gamma}_x(t)$ where

$$\hat{\Gamma}_x(t) = \frac{\sum_{i=1}^{n} K\left(h_k^{-1} d(x,X_i)\right)\rho(Y_i,t)}{\sum_{i=1}^{n} K\left(h_k^{-1} d(x,X_i)\right)}, \tag{2}$$

with K is a kernel function and $h_k = \min\left\{h \in \mathbb{R}^+ \; such\; that \; \sum_{i=1}^{n} \mathbb{1}_{B(x,h)}(X_i) = k\right\}$ with k is given as a sequence of integers.

It should be noted that, in recent years, the field of functional statistical analysis has become an essential topic of research. This field of statistics concerns the modeling of the random variable takes values in a functional space. Data with functional structure arises in many applied sciences fields such as soil science, economics, epidemiology, and others. Key references in this topic are Bosq [Bosq (2000)] and Belarbi et al. [Belarbi, Chemikh and Laksaci (2018)] in the parametric model and the monograph of Ferraty et al. [Ferraty and Vieu (2006)] in the nonparametric case. The main purpose of this contribution is to construct an estimator of a regression model using the k Nearest Neighbors procedure (kNN). Unlike the standard kernel method, the kNN smoothing allows estimating the M-regression, θ_x, with varied bandwidth parameter strongly depends on the data. Precisely, the bandwidth parameter is priory defined according to the distance between the functional random variable. Such consideration allows for exploring the topological as well as the specter component of the data.

The nonparametric robust regression estimation was firstly introduced by Azzedine et al. [Azzedine, Laksaci and Ould Said (2008)]. They obtained the almost complete convergence with rates in the independent and identically distributed (i.i.d.) case. The asymptotic normality of their model has been established by Attouch et al. [Attouch, Laksaci and Ould Said (2010, 2012)] in both dependent and independent cases. However, all these results are obtained in the complete data. In the incomplete data cases, we can refer to Derrar et al. [Derrar, Laksaci and Ould Said (2020)] and the references therein.

Pushed by its attractive features, the functional kNN smoothing approach has received a growing consideration in the last years. The first study in this topic was obtained by Burba et al. [Burba, Ferraty and Vieu (2009)]. They proved the almost complete consistency of the kNN estimator of the regression operator of functional regressors. Some recent

advances in the functional kNN method can be found in Almanjahie et al. [Almanjahie, Chikr, Laksaci et al. (2018); Kara-Zaitri, Laksaci, Rachdi et al. (2017); Almanjahie, Chikr and Laksaci (2020)].

We aim to construct a new estimator of the regression function. Our estimator is built by combining the robust method estimation with (weighted by) the kernel nearest-neighbor (kNN) procedure uniformly on the number of neighbors. 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.

Note that, such a study has a great impact on practice. On the one hand, the robust regression 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. Furthermore, it is well known that the kNN method is better than the classical kernel method. 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.

On the other hand, the UNN consistency ensures the convergence of the estimator even if the number of neighbors is variable. Of course, this case incorporates the cross-validation criterion, which provides a varied random bandwidth parameter. So, the UNN consistency offers mathematical support for this kind of data-driven procedure. For other studies related to nonparametric estimates of the robust regression, we refer to Attouch et al. [Attouch, Kaid and Louhab (2019); Attouch, Laksaci and Ould Said (2010, 2012); Azzedine, Laksaci and Ould Said (2008); Gheriballah, Laksaci and Sekkal (2013); Gheriballah, Laksaci and Rouane (2010); Belarbi, Chemikh and Laksaci (2018)].

The main objective of this paper is to generalize, to the k nearest neighbor case, the results obtained by Azzedine et al. [Azzedine, Laksaci and Ould Said (2008)] in the i.i.d case. More precisely, we establish the almost complete convergence uniformly on the number of neighbors with rates of an estimator constructed 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 paper is organized as follows. Section 2 is dedicated to fixing notations and conditions, and to establishing the development of our main results. Also, we present simulation studies and real data applications to investigate the efficiency of the proposed robust estimator. The conclusion is presented in Section 3.

2 Main results

To establish the almost complete convergence of $\hat{\theta_x}$ uniformly in the numbers of neighbors $k \in (k_{1,n}, k_{2,n})$, we need the following conditions and notations:

(A1) For all $r > 0$, $\mathbb{P}(X \in B(x,r)) = \phi_x(r) > 0$ such that, for all $s \in (0,1)$,
$\lim_{r\to 0} \frac{\phi_x(sr)}{\phi_x(r)} = \tau_x(s) < \infty$.

(A2) The function Γ is such that:

(i) The function $\Gamma_x(\cdot)$ is of class $\mathcal{C}^2$ on $[\theta_x - \delta, \theta_x + \delta]$, $\delta > 0$.

(ii) $\forall t \in [\theta_x - \delta, \theta_x + \delta]$, $\forall (x_1, x_2) \in \mathcal{N}_x \times \mathcal{N}_x$, $|\Gamma_{x_1}(t) - \Gamma_{x_2}(t)| \leq C d^b(x_1, x_2)$, where $\mathcal{N}_x$ is a fixed neighborhood of x.

(iii) For each fixed $t \in [\theta_x - \delta, \theta_x + \delta]$, the function $\Gamma.(t)$ is continuous at x.

(A3) The function ρ is a strictly convex, continuous and differentiable w.r.t. the variable t, and its derivative, $\psi(y,t) = \frac{\partial \rho(y,t)}{\partial t}$, fulfills $\mathbb{E}[|\psi(Y,t)|^2|X] < C < \infty$ and $\mathbb{E}[|\psi(Y,t)|^p] < C < \infty, p > 1$.

(A4) The kernel function K is supported within $(0,1/2)$ and the derivative function of K is continuous on $(0,1/2)$ such that

$$0 < C\mathbb{1}_{\left(\frac{0,1}{2}\right)}(\cdot) \leq K(\cdot) \leq C'^{\mathbb{1}_{\left(\frac{0,1}{2}\right)}}(\cdot) \text{ and } K(1/2) - \int_0^{\frac{1}{2}} K'(s)\tau_x(s)ds > 0,$$

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

(A5) Let define the class $\mathcal{K}$ of functions by $\mathcal{K} = \{\cdot \mapsto K(\gamma^{-1}d(x,\cdot)),\ \gamma > 0\}$ which is a pointwise measurable class[4] such that $\sup_Q \int_0^1 \sqrt{1 + \log \mathcal{N}(\epsilon \parallel F \parallel_{Q,2}, \mathcal{K}, d_Q)}\, d\epsilon < \infty$, where the supremum is taken over all probability measures Q on the space $\mathcal{F}$ with $Q(F^2) < \infty$ and $\mathcal{F}$ is the envelope function[5] of the set $\mathcal{K}$. Here, d_Q is the $L_2(Q)$-metric and $\mathcal{N}(\epsilon, \mathcal{K}, d_Q)$ is the minimal number of open balls (for the $L_2(Q)$-metric) with radius ϵ which are needed to cover the function class $\mathcal{K}$. We will denote by $\|\cdot\|_{Q,2}$ the $L_2(Q)$-norm.

(A6) The sequence of numbers $(k_{1,n})$ verifies

$$\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right) \to 0 \qquad and \qquad \frac{\log n}{\min(n\phi_x^{-1}\left(\frac{k_{1,n}}{n}\right), k_{1,n})} \to 0.$$

C or/and C' denotes a generic positive constant.

In the following theorem, we present the consistency result.

Theorem: Assume that conditions (A1) -(A6) are satisfied, then $\hat{\theta_x}$ exists and is unique almost surely for all larger value of n. Furthermore, if $\Gamma_x^{''}(\theta_x) \neq 0$, we have

$$\sup_{k_{1,n} \leq k \leq k_{2,n}} \left|\hat{\theta_x} - \theta_x\right| = O\left(\phi_x^{-1}\left(\frac{k_{2,n}}{n}\right)^{\min(k_1,k_2)}\right) + O_{a.co.}\left(\sqrt{\frac{\log(n)}{k_{1,n}}}\right). \tag{3}$$

Proof of the Theorem

---

[4]A class $\mathcal{C}$ of functions is called a pointwise measurable class, if $\exists \mathcal{C}_0$ a countable subclass, such that, $\forall g \in \mathcal{C}$ there exists a sequence of functions $(g_m)_{m\in\mathbb{N}} \in \mathcal{C}_0$ satisfying $|g_m(z) - g(z)| = o(1)$.

[5]An envelope function G for a class of functions $\mathcal{C}$ is any measurable function such that: $\sup_{g\in\mathcal{C}}|g(z)| \leq G(z)$, for all$z$.

To prove the above theorem, we use the strict convexity of ρ, which implies that the function ψ is strictly monotone concerning the variable t. Note that, for the shortness of the prove, we determine the result for the increasing case, and the decreasing case can be obtained by $-\psi(Y,\cdot)$. Then, in the increasing case of $\psi(Y,\cdot)$, we can get

$\forall\epsilon > 0 \;\;,\;\; \Gamma_x'(\theta_x - \epsilon) \leq \Gamma_x'(\theta_x) = 0 \leq \Gamma_x'(\theta_x + \epsilon)$ and $\hat{\Gamma}'_x(\hat{\theta}_x - \epsilon) \leq \hat{\Gamma}'_x(\hat{\theta}_x) = 0 \leq \hat{\Gamma}'_x(\hat{\theta}_x + \epsilon)$.

So, by a simple analytical argument, we prove that, for all $\epsilon > 0$,

$$\mathbb{P}\left(\sup_{k_{1,n}\leq k\leq k_{2,n}} |\hat{\theta}_x - \theta_x| \geq \epsilon\right)$$
$$\leq \mathbb{P}\left(\sup_{k_{1,n}\leq k\leq k_{2,n}} |\hat{\Gamma}'_x(\theta_x + \epsilon) - \Gamma_x'(\theta_x + \epsilon)| \geq \Gamma_x'(\theta_x + \epsilon)\right)$$
$$+\mathbb{P}\left(\sup_{k_{1,n}\leq k\leq k_{2,n}} |\hat{\Gamma}'_x(\theta_x - \epsilon) - \Gamma_x'(\theta_x - \epsilon)| \geq -\Gamma_x'(\theta_x - \epsilon)\right).$$

Furthermore, by ((A2) (i)), we write

$$\hat{\theta}_x - \theta_x = \frac{\Gamma_x'(\hat{\theta}_x) - \hat{\Gamma}'_x(\hat{\theta}_x)}{\Gamma_x''(\xi_n)},$$

where ξ_n is between $\hat{\theta}_x$ and θ_x. This result allows us to demonstrate that

$$\exists\tau > 0,\ \sum_{n=1}^{\infty} \mathbb{P}\left(\inf_{k_{1,n}\leq k\leq k_{2,n}} \Gamma_x''(\xi_n) < \tau\right) < \infty. \tag{4}$$

We would have

$$\sup_{k_{1,n}\leq k\leq k_{2,n}} \left|\hat{\theta}_x - \theta_x\right| = O_{a.co.}\left(\sup_{t\in[\theta_x-\delta,\ \theta_x+\delta]} \sup_{k_{1,n}\leq k\leq k_{2,n}} |\Gamma_x'(t) - \hat{\Gamma}'_x(t)|\right).$$

Therefore, we still have to prove that

$$\sup_{t\in[\theta_x-\delta,\ \theta_x+\delta]} \sup_{k_{1,n}\leq k\leq k_{2,n}} \left|\Gamma_x'(t) - \hat{\Gamma}'_x(t)\right|,$$

and for this sequence, we use the decomposition

$$\Gamma_x'(t) - \hat{\Gamma}'_x(t) = \tilde{B}_x(t) + \frac{\tilde{D}_x(t)}{\hat{\Psi}_D(x)} + \frac{\tilde{Q}_x(t)}{\hat{\Psi}_D(x)},$$

where

$$\tilde{Q}_x(t) = (\hat{\Psi}_N(x,t) - \mathbb{E}[\hat{\Psi}_N(x,t)]) - \Gamma_x'(t)(\hat{\Psi}_D(x) - \mathbb{E}[\hat{\Psi}_D(x)),$$

$$\tilde{B}_x(t) = \frac{\mathbb{E}[\hat{\Psi}_N(x,t)]}{\mathbb{E}[\hat{\Psi}_D(x)} - \Gamma_x'(t)$$

and

$$\tilde{D}_x(t) = -\tilde{B}_x(t)(\hat{\Psi}_D(x) - \mathbb{E}[\hat{\Psi}_D(x)),$$

with

$$\hat{\Psi}_N(x,t) = 1/(n\mathbb{E}[K(h^{-1}d(x,X_1))])\sum_{i=1}^{n} K(h^{-1}d(x,X_i))\psi(Y_i,t),$$

$$\hat{\Psi}_D(x) = 1/(n\mathbb{E}[K(h^{-1}d(x,X_1))])\sum_{i=1}^{n} K(h^{-1}d(x,X_i)).$$

Hence, the theorem is obtained by the following Lemmas.

Lemma 1: Under the conditions (A1) and (A3)–(A6), we can obtain that

$$\sup_{a_n\le h\le b_n}\left|\hat{\Psi}_D(x) - \mathbb{E}[\hat{\Psi}_D(x)]\right| = O_{a.co.}\left(\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right).$$

Furthermore, $\exists C > 0$ a real number, such that

$$\sum_{n=1}^{\infty}\mathbb{P}\left(\inf_{a_n\le h\le b_n}\hat{\Psi}_D(x) < C\right) < \infty,$$

where $a_n = \phi_x^{-1}\left(\frac{\alpha k_{1,n}}{n}\right)$ and $b_n = \phi_x^{-1}\left(\frac{k_{2,n}}{n\alpha}\right)$.

Lemma 2: Under the conditions (A1) -(A6), we have

$$\sup_{a_n\le h\le b_n}\sup_{t\in[t_p(x)-\delta,\ t_p(x)+\delta]}\left|\hat{\Psi}_N(x,t) - \mathbb{E}[\hat{\Psi}_N(x,t)]\right| = O_{a.co.}\left(\sqrt{\frac{\log(n)}{n\phi_x(a_n)}}\right).$$

Lemma 3: Under the conditions (A2), (A3) and (A4), we have

$$\sup_{a_n\le h\le b_n}\sup_{t\in[t_p(x)-\delta,\ t_p(x)+\delta]}\left|\tilde{B}(x,t)\right| = O(b_n^{\min(k_1,k_2)}).$$

Lemma 4: Under the same conditions of the previous theorem, the estimator $\hat{\theta}_x$ exists and unique a.s. for all large n and there exists $\zeta_1 > 0$, such that

$$\sum_{n\ge 1}\mathbb{P}\left\{\inf_{k_{1,n}\le k\le k_{2,n}}\Gamma_x''(\xi_n) < \zeta_1\right\} < \infty.$$

For the sake of shortness, the proof of the intermediate results is given in brevity. The proofs of Lemmas 1, 3 and 4 are omitted. It can be obtained by combing the ideas of Lemma 2 to those of [Attouch, Kaid and Louhab (2019)]. So, the proof of Lemma 2 is sufficient to prove the claimed result.

Proof of Lemma 2:

By the compactness of $[\theta_x - \delta,\quad \theta_x + \delta]$, we have

$$[\theta_x - \delta,\quad \theta_x + \delta] \subset \bigcup_{j=1}^{d_n}(y_j - l_n, y_j + l_n)$$

with $l_n = n^{-1/2}$ and $d_n = O(n^{1/2})$. Then the monotony of $\mathbb{E}[\hat{\Psi}_N(x,\cdot)]$ and $\hat{\Psi}_N(x,\cdot)$ gives, for $1 \le j \le d_n$, that

$$\mathbb{E}\left[\hat{\Psi}_N(x, y_j - l_n)\right] \leq \sup_{t\in(y_j-l_n, y_j+l_n)} \mathbb{E}\left[\hat{\Psi}_N(x,t)\right] \leq \mathbb{E}\left[\hat{\Psi}_N(x, y_j + l_n)\right]$$

$$\hat{\Psi}_N(x, y_j - l_n) \leq \sup_{t\in(y_j-l_n, y_j+l_n)} \hat{\Psi}_N(x,t) \leq \hat{\Psi}_N(x, y_j + l_n).$$

Now, from the condition (A2), we obtain, for any $y_1, \quad y_2 \in [\theta_x - \delta, \theta_x + \delta]$, that $\left|\mathbb{E}\left[\hat{\Psi}_N(x, y_1)\right] - \mathbb{E}\left[\hat{\Psi}_N(x, y_2)\right]\right| \leq C|y_1 - y_2|$.

It follows that

$$\sup_{t\in[\theta_x-\delta,\ \theta_x+\delta]} \left|\hat{\Psi}_N(x,t) - \mathbb{E}\left[\hat{\Psi}_N(x,t)\right]\right|$$
$$\leq \max_{1\leq j\leq d_n} \max_{z\in\{y_j-l_n, y_j+l_n\}} \left|\hat{\Psi}_N(x,z) - \mathbb{E}\left[\hat{\Psi}_N(x,z)\right]\right| + 2Cl_n.$$

Evidently,

$$l_n = o\left(\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right).$$

So, all it remains to prove is that

$$\sup_{a_n\leq h\leq b_n} \max_{1\leq j\leq d_n} \max_{y\in\{t_j-l_n, t_j+l_n\}} \left|\hat{\Psi}_N(x,y) - \mathbb{E}\left[\hat{\Psi}_N(x,y)\right]\right| = O\left(\left(\frac{\log(n)}{n\ \ \phi_x(a_n)}\right)^{1/2}\right), a.co.$$

$$\mathbb{P}\left(\sup_{a_n\leq h\leq b_n} \max_{1\leq j\leq d_n} \max_{z\in\{t_j-l_n, t_j+l_n\}} \left|\hat{\Psi}_N(x,z) - \mathbb{E}\left[\hat{\Psi}_N(x,z)\right]\right| > \eta\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right)$$

$$\leq\ 2d_n \max_{1\leq j\leq d_n} \max_{z\in\{t_j-l_n, t_j+l_n\}} \mathbb{P}\left(\sup_{a_n\leq h\leq b_n} \left|\hat{\Psi}_N(x,z) - \mathbb{E}[\hat{\Psi}_N(x,z)]\right| > \eta\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right).$$

Now, we look at the quantity

$$\mathbb{P}\left(\sup_{a_n\leq h\leq b_n} \left|\hat{\Psi}_N(x,z) - \mathbb{E}[\hat{\Psi}_N(x,z)]\right| > \eta\sqrt{\frac{\log n}{n\phi_x(a_n)}}\right),$$

for all $z = t_j \mp l_n,\ 1 \leq j \leq d_n$. The proof of the above inequality is based on the Bernstein's inequality for empirical processes, i.e.,

$$\alpha_n(K) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(K_i\psi(Y_i,t) - \mathbb{E}[K_i\psi(Y_i,t)]),$$

where $K_i = K(h^{-1}d(x,X_i))$. Then, we get, for all $z = t_j \mp l_n, \quad 1 \leq j \leq d_n$, that

$$\mathbb{P}\left\{\sup_{a_n\leq h\leq b_0}\sqrt{\frac{n\phi_x(h)}{\log n}}\left|\hat{\Psi}_N(x,z) - \mathbb{E}[\hat{\Psi}_N(x,z)]\right| \geq {\eta_0}'\right\} \leq \log(n)n^{-C'\eta_0^2}. \tag{5}$$

Consequently, a good choice of η_0 allows deducing the inequality in Eq. (5).

2.1 Simulation part

The main propose of this part is to evaluate the behavior of our *k*NN estimator given in Eq. (2) for α-mixing data sample against the NW kernel estimator defined in Ferraty et al. [Ferraty and Vieu (2006)]. Let use the following nonparametric functional relation:

$Y_i = m(X_i) + \varepsilon_i, i = 1, \ldots, 200$

where $m(X_i) = \left(\int_0^{\pi/3} X_i'(t)dt\right)^2$, $X_i'(t)$ is the first derivative of $X_i(t)$, $(\varepsilon_i)_{i=1}^{200}$ are Gaussian random variables $\mathcal{N}(0, 0.05)$. The functional samples $X_i(t)$ (displayed in Fig. 1) are calculated by

$$X_i(t) = a_i t^2 + \sin\left(b_i\left(t - \frac{\pi}{3}\right)\right), \quad i = 1,2, \ldots, 200, t \in \left[0, \frac{\pi}{3}\right],$$

where $(a_i)_{i=1}^{200}$ are independent identically distributed random variables and drown by $\mathcal{U}(0, \pi/9)$, $b_i = 1/4b_{i-1} + \xi_i$, $(\xi_i)_{i=1}^{200}$ are i.i.d. according to $\mathcal{N}(0,1)$ and independent to $(a_i)_{i=1}^{200}$ and $(b_i)_{i=1}^{200}$ respectively, where b_0 is centered and reduced Gaussian r.v.

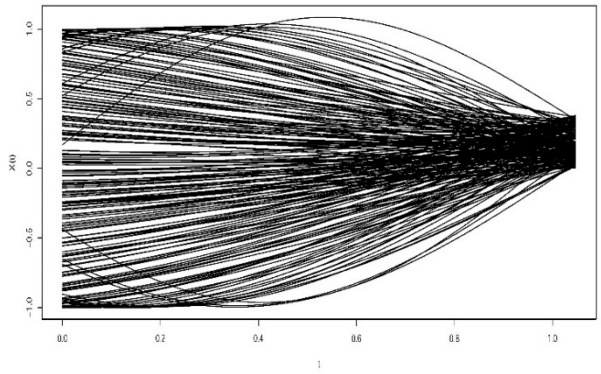

Figure 1: The functional samples $X_i(t), t \in [0, \pi/3]$ for $i = 1, \ldots, 200$

We need to choose an appropriate semi-metric $d(.,.)$, a kernel $K(.)$ and the smoothing parameter k_{opt} for the *k*NN functional estimator, and h_{opt} for the NW kernel estimator. For that, we opted for the asymmetric quadratic kernel defined as $K(u) = \frac{3}{4}\left(\frac{12}{11} - u^2\right)\mathbb{1}_{[0,1]}(u)$. Thus, because of the smoothness of curves $X_i(t)$, we opt for the semi-metric based on the first derivative:

$$d^{deriv}(X_i, X_j) = \sqrt{\int_0^{\pi/3} \left(X_i'(t) - X_j'(t)\right)^2 \mathrm{d}t}, \quad \forall X_i, X_j \in \mathcal{F}$$

The prediction procedure is as follows. Our sample is divided into two parts, the first $(X_i, Y_i)_{i=1}^{150}$ which is used for modeling; the rest of the sample will be used to the verification of the quality of prediction. Thus, from the generated functional samples, we can calculate the optimal parameter k_{opt} for the classical *k*NN kernel, the robust *k*NN and the optimal

parameter h_{opt} for the NW robust estimator by cross-validation procedures. To proceed, the k_{opt} for the *k*NN kernel is defined by

$k_{opt} = \text{argmin}_k CV_1(k)$, where $CV_1(k) = \sum_{i=1}^{n}\left(Y_i - \hat{m}_{(-i)}^{kNN}(\chi)\right)^2$ and

$$\hat{m}_{(-i)}^{kNN}(X) = \frac{\sum_{j=1,j\neq i}^{n} Y_j\, K\left(\frac{d^{deriv}(X_j, X)}{H_{n,k}(X)}\right)}{\sum_{j=1,j\neq i}^{n} K\left(\frac{d^{deriv}(X_j; X)}{H_{n,k}(X)}\right)}.$$

For the robust *k*NN, we define $k_{opt} = \text{argmin}_k CV_2(k)$, where $CV_2(k) = \sum_{i=1}^{n}\left(Y_i - \hat{r}_{(-i)}^{kNN}(X)\right)^2$ and

$$\hat{r}_{(-i)}^{kNN}(X) = \underset{t}{\text{argmin}} \frac{\sum_{j=1,j\neq i}^{n} \rho(Y_j, t) K\left(\frac{d^{deriv}(X_j, X)}{H_{n,k}(X)}\right)}{\sum_{j=1,j\neq i}^{n} K\left(\frac{d^{deriv}(X_j, X)}{H_{n,k}(X)}\right)}.$$

For the NW robust method, we can define $h_{opt} = \text{argmin}_h CV_3(h)$ where

$$CV_3(h) = \sum_{i=1}^{n}\left(Y_i - \hat{r}_{(-i)}^{kernel}(\chi)\right)^2$$

and

$$\hat{r}_{(-i)}^{kernel}(X) = \underset{t}{\text{argmin}} \frac{\sum_{j=1,j\neq i}^{n} \rho(Y_j, t) K\left(\frac{d^{deriv}(X_j, X)}{h}\right)}{\sum_{j=1,j\neq i}^{n} K\left(\frac{d^{deriv}(X_j, \chi)}{h}\right)}.$$

The theoretical results of these methods for selecting the smoothing parameter are obtained by Härdle et al. [Härdle and Vieu (1992)] when the data are dependent and by Rachdi et al. [Rachdi and Vieu (2007)] when the data are functional.

Using the prediction samples, we calculate the prediction values of the response variables noted as $(Y_i)_{i=151}^{200}$. Thus, the mean square error (MSE) of the predicted responses for each method is illustrated in Fig. 2, where we can notice that the prediction of the *k*NN estimator is more precise than that of the NW kernel estimator under α-mixing dependency.

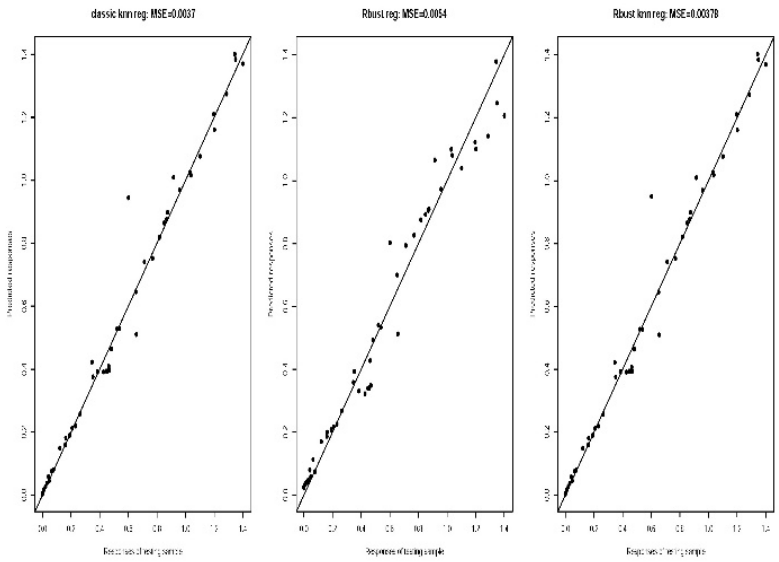

Figure 2: The prediction of the *k*NN classical method, NW robust method, and *k*NN robust method

The similar results are also showed when the sample sizes are n=300 and n=500. To further explore the performances of the two approaches, we carry out M=100 independent replications of the experiment for kNN estimator and NW kernel estimator when the sample sizes are $n = 200$, $n = 300$ and $n = 500$ respectively. In each case, let the testing sample sizes be 50. Figs. 3 and 4 show the bean-plots of the MSE of the prediction values by the three methods for the different sample sizes with and without outliers. One can see that, for each method, the MSE of the prediction values decreases as the sample sizes n increase. Furthermore, let evaluate the average of MSE of *k*NN estimator and kernel estimator. The obtained results are given in Tab. 1. We observe in Tab. 1 that in the presence of typical values, the *k*NN robust regression gives better results than the *k*NN regression and robust regression. Indeed, the MSE values of both methods increase significantly compared to the number of disturbing points. However, the MSE remains very low for the *k*NN robust method.

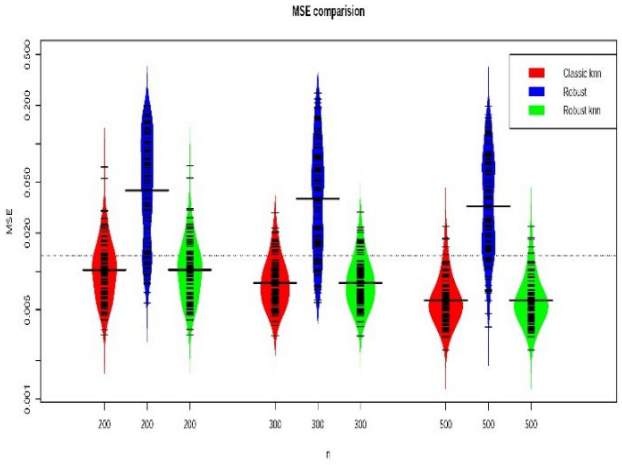

Figure 3: The bean-plots of the MSE of the prediction values by the three methods for the different sample sizes without outliers

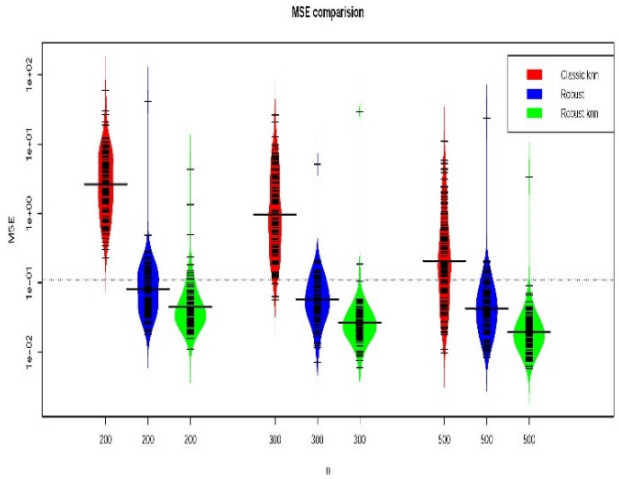

Figure 4: The bean-plots of the MSE of the prediction values with different sample sizes in the presence of outliers for the three methods estimation

Table 1: The comparison of different methods when artificial outliers taint data

Number of perturbations	0 value	5 values	15 values
MSE *k*NN reg.	0.004335144	4.399523	14.09368
MSE Robust reg.	0.01468679	0.0173451	0.1109583
MSE *k*NN Robust reg.	0.004128343	0.00998825	0.05887536

2.2 Real data application

This short section presents an example of using the *k*NN robust regression to predict the octane in gasoline spectrometric curves. To clarify that, let consider a sample of 60 gasoline samples; the used data is available in the R-package PLS. For any wavelength t ($(t_1, \dots, t_{401})$ obtained by a light beam for the near-infrared range $900 - 1700\text{nm}$), and for each gasoline sample i, denote by $X_i(t)$ the measured absorption of radiation. Fig. 5 displays the 60 spectrometric curves obtained by the discretized data curves $X_i(t_1), \dots, X_i(t_{401})$.

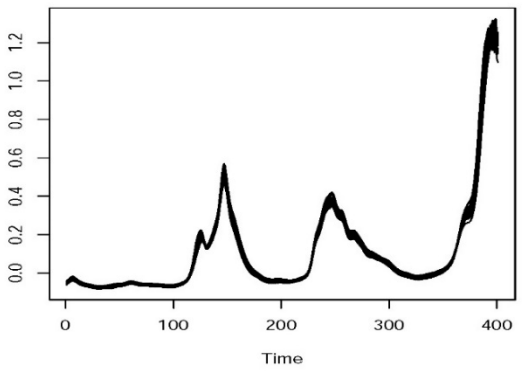

Figure 5: 60 spectra curves at 401 equally spaced wavelengths

We suppose that for each gasoline sample, the quantity of octane is known. Our goal is to predict the octane number given a new spectrometric curve. For that, we suppose that we have n pairs $(X_i, Y_i)_{i=1,\dots,n}$ where X_i (resp. Y_i) is the i-th spectrometric curve (resp. response). Thus, we can formulate this prediction problem by the following model

$Y_i = r(X_i) + \epsilon_i$ for $i = 1, \dots, n$.

Then, we split our observations $(X_i, Y_i)_{i=1,\dots,60}$ into two samples, the training sample contains $(X_i, Y_i)_{i=1,\dots,30}$ for modeling, and the rest of the sample $\left(X_j, Y_j\right)_{j=31,\dots,60}$ is the testing sample used to determine the quality of the prediction. To calculate the estimator for both methods, we take the quadratic kernel function $K(u) = \frac{3}{4}\left(\frac{12}{11} - u^2\right)\mathbb{1}_{[0,1]}(u)$, and the semi-metric $d^{deriv}(.,.)$. Similar to the CV-procedures in the simulation study we select the parameters k_{opt} (for the *k*NN) and h_{opt} (for the kernel). Finally, we predict the $\left(Y_j\right)_{j=31,\dots,60}$ for the above two methods and calculate the MSE of the predicted values; the results are given in Figs. 6 and 7. Note that we can see in these figures that the calculated MSE with the robust *k*NN method is smaller than the one obtained with the robust kernel, and the predicted plot by *k*NN method estimation is very close to the real case than that given by the kernel method estimation. Hence, we can deduce that the robust *k*NN method gives better results than the kernel method because the first one takes into account the proximity of the curves.

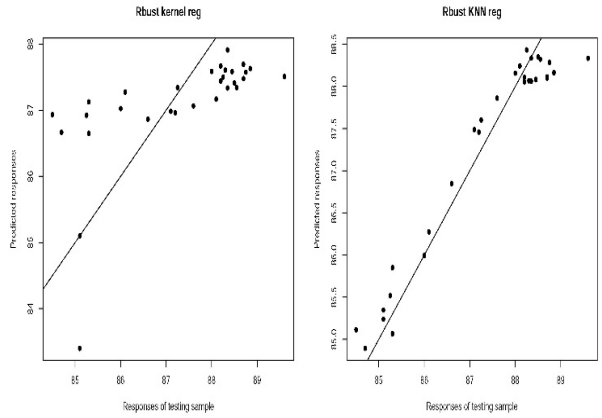

Figure 6: Prediction for both methods

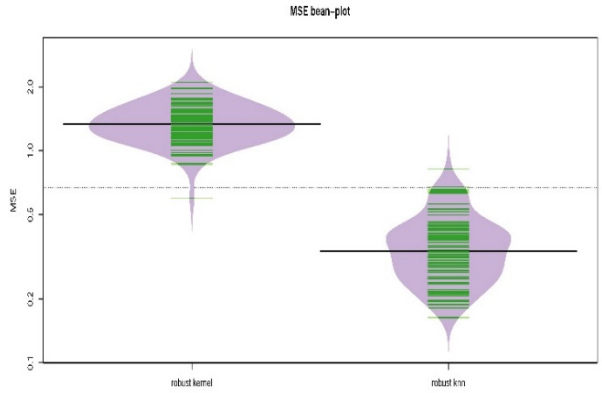

Figure 7: The bean-plots of the MSE for both methods

3 Conclusion

The uniform *k*NN consistency method is an alternative smoothing approach that permits to obtain an adaptive estimator for many statistical problems, such as the bandwidth selection. However, unlike to the multivariate case, the uniform consistency is not a simple extension of the pointwise method because this latter requires some additional tools and techniques. Notice that the difficulty of this issue lies in the fact that the bandwidth parameter in the *k*NN method is a random variable.

The main novelty of this contribution is to estimate the regression function by combining two important statistical tools: the robust method and the *k*NN procedures. This combination permits to build a new estimator which inherits the advantages for both approaches. Moreover, in the application part, we have shown that *k*NN robust estimator is easy to implement and more robust and performant than it's classical one such as the *k*NN regression estimator or the robust regression estimator. More precisely, we have carried

out two numerical studies. The first one examines the behavior of the *k*NN robust regression estimator in the presence of outliers in data. While the other deals with real data application and illustrates the importance of this combined method estimation to the prediction problem.

As a smoothing method, the preference of the local linear approach is closely linked to the bandwidths parameters choice; linked, here, by the number of neighbor *k*. The adaptation of all these ideas to the local linear estimate of the robust regression is an important prospect to investigate in the future.

Another open question, we can consider the more general case when the scale parameter is unknown, and data come from a functional time series (dependency, ergodicity). 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.

In conclusion, we can say that the constructed estimator's behavior is not affected by the number of outliers in the data set. The combination of the *k*NN algorithm and the robust method permits the reduction of the effect of outliers in data compared to the classical kernel approach.

Acknowledgment: The authors grateful to the referees for their time and effort in providing very help and valuable comments and suggestion which leads to improve the quality of the paper.

Funding Statement: The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through the Research Groups Program under grant number R.G.P. 2/67/41. I. M. A. and M. K. A. are the authors who received the grant.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Almanjahie, M. I.; Chikr Elmezouar, Z.; Laksaci, A. (2020): Functional causality between oil prices and GDP based on big data. *Computers Materials & Continua*, vol. 63, no. 2, pp 593-604.

Almanjahie, M. I.; Chikr Elmezouar, Z.; Laksaci, A.; Rachdi, M. (2018): *k*NN local linear estimation of the conditional cumulative distribution function: dependent functional data case. *Comptes Rendus Mathematique Academic des Sciences Paris*, vol. 356 , no. 10, pp. 1036-1039.

Attouch, M.; Laksaci, A.; Ould Saïd, E. (2010): Asymptotic normality of a robust estimator of the regression function for functional time series data. *Journal of The Korean Statistical Society*, vol. 39, no. 4, pp. 489-500.

Attouch, M.; Laksaci, A.; Ould Saïd, E. (2012): Robust regression for functional time series data. J*ournal of The Japan Statistical Society*, vol. 42, no. 3, pp. 125-143.

Attouch, M. K.; Kaid, Z.; Louhab, H. (2019): Asymptotic normality of a robust kernel estimator of the regression function for functional ergodic data: case of an unknown scale parameter. *Comptes Rendus Mathematique Academic des Sciences Paris*, vol. 357, no. 5, pp. 478-481.

Azzedine, N.; M. Laksaci, A.; Ould Saïd, E. (2008): On the robust nonparametric regression estimation for functional regressor. *Statistics & Probability Letters*, vol. 78, no. 18, pp. 3216-3221.

Belarbi, F.; Chemikh, S.; Laksaci, A. (2018): Local linear estimate of the nonparametric robust regression in functional data. *Statistics & Probability Letters*, vol. 134, no. 3, pp. 128-133.

Bosq, D. (2000): Linear processes in function spaces theory and applications. *Lecture Notes in Statistics*, vol. 149, Springer.

Burba, F.; Ferraty, F.; Vieu, P. (2009): k-nearest neighbor method in functional nonparametric regression. *Journal of Nonparametric Statistics*, vol 21, no. 4, pp. 453-469.

Chen, J.; Zhang, L. (2009): Asymptotic properties of nonparametric M-estimation for mixing functional data. *Journal of Statistical Planning and Inference*, vol. 139, no. 2, 533-546.

Chen, Y. T.; Wang, J.; Xia, R. L.; Zhang, Q.; Cao, Z. H. et al. (2019): The visual object tracking algorithm research based on adaptive combination kernel. *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 12, pp. 4855-4867.

Derrar, S.; Laksaci, A.; Ould Saíd, E. (2020): M-estimation of the regression function under random left truncation and functional time series model. *Statistical Papers*, vol. 61, no. 3, 1181-1202.

Ferraty, F.; Vieu, P. (2006): Nonparametric Functional Data Analysis: Theory and Practice. *Series in Statistics*. Springer.

Gheriballah, A.; Laksaci, A.; Rouane, R. (2010): Robust nonparametric estimation for spatial regression. *Journal of Statistical Planning and Inference*, vol. 140, no. 7, pp. 1656-1670.

Gheriballah, A.; Laksaci, A.; Sekkal, S. (2013): Nonparametric M-regression for functional ergodic data. *Statistics & Probability Letters*, vol. 83, no. 3, pp. 902-908.

Kara-Zaïtri, L.; Laksaci, A.; Rachdi, M.; Vieu, P. (2017): Data-driven kNN estimation for various problems involving functional data. *Journal of Multivariate Analysis*, vol. 153, no. 1, pp. 176-188.

Li, W. J.; Cao, Y. X.; Chen, J.; Wang, J. X. (2017): Deeper local search for parameterized and approximation algorithms for maximum internal spanning tree. *Information and Computation*, vol. 252, pp. 187-200.

Li, W. J.; Feng, Q. L.; Chen, J.; Hu, S. (2017): Improved kernel results for some FPT problems based on simple observations. *Theoretical Computer Science*, vol. 657, pp. 20-27.

Li, W. J.; Liu, H. Y.; Wang, J. X.; Xiang, L. Y.; Yang, Y. J. (2019): An improved linear kernel for complementary maximal strip recovery: Simpler and smaller. *Theoretical Computer Science*, vol. 786, pp. 55-66.

Li, W. J.; Zhu, B. H. (2018): A 2k-vertex kernel for vertex cover based on crown decomposition. *Theoretical Computer Science*, vol. 739, pp. 80-85.

Maronna, R.; Douglas Martin, R.; Yohai, V. J.; Salibián-Barrera, M. (2019): *Robust Statistics: Theory and Methods (with R),* 2nd Edition. Wiley.

Rachdi, M.; Laksaci, A.; Almanjahie, I. M.; Chikr-Elmezouar, Z. (2020): FDA: theoretical and practical efficiency of the local linear estimation based on the *k*NN smoothing of the conditional distribution when there are missing data. *Journal of Statistical Computation and Simulation*, vol. 90, no. 8, 1479-1495.

Zhu, T. F.; Lin, Y. P.; Liu, Y. H.; Zhang, W.; Zhang, J. M. (2019): Minority oversampling for imbalanced ordinal regression. *Knowledge-Based Systems*, vol. 166, pp. 140-155.