Asymptotic Normality of the Robust Equivariant Estimator for Functional Nonparametric Models

Ibrahim M. Almanjahie, Omar Fetitah, Mohammed Kadi Attouch, and Hayat Louhab

1Department of Mathematics, College of Science, King Khalid University, Abha 62529, Saudi Arabia
2Statistical Research and Studies Support Unit, King Khalid University, Abha 62529, Saudi Arabia
3Laboratory of Statistics and Stochastic Processes, University of Djillali Liabes, BP 89, Sidi Bel Abbes 22000, Algeria

Correspondence should be addressed to Ibrahim M. Almanjahie; imalmanjahi@kku.edu.sa

Received 6 May 2021; Revised 3 November 2021; Accepted 14 March 2022; Published 10 May 2022

1. Introduction

Nonparametric kernel regression is a familiar tool to explore the underlying relationship between response variables and covariates. In the functional data studies, these estimators are given by Ferraty and Vieu [1] and Ferraty and Vieu [2]. As in parametric regression estimation, the kernel estimator may be affected by outliers; hence, considering robustness is essential.

In this context, let \((X_i, Y_i)_{i=1, \ldots, n}\) be a sequence of strictly stationary dependent random variables identically distributed as \((X, Y)\), a random pair valued in \(F \times \mathbb{R}\), where \(F\) is a semi-metric space. Let \(d\) denote the semi-metric. The purpose of this work is to study the nonparametric estimation of the robust regression \(\theta(x)\) when the scale parameter is unknown. In fact, for any \(x \in F\), \(\theta(x)\) is defined as zero with respect to the parameter \(a\) of the following equation:

\[
\Psi(x, a, \sigma) = \left[ \psi_a \left( \frac{Y - a}{\sigma} \right) \right]_{X = x} = 0,
\]

where \(\psi_a\) is a real-valued function that satisfies some regularity conditions, to be stated later, and \(\sigma\) is a robust measure of conditional scale. In what follows, we assume, for all \(x \in F\), that the robust regression \(\theta(x)\) exists and is unique (see, for instance, Boente and Fraiman [3]).

We point out that the robustification method is an old subject in statistics. Among many papers dealing with the robust nonparametric estimates of regression function in finite dimension case, one can refer, for example, to key works of Robinson (1984), and Collomb and Härdle [4]. We refer to Laïb and Ould Saïd [5] for some results on multivariate time series (mixing and ergodicity conditions), and Roozbeh [6] and Roozbeh et al. [6–8] for multicollinearity in the least trimmed squares regression and outliers effects.

The robust regression is widely studied in nonparametric functional statistics. Indeed, it was firstly introduced by Azzedine et al. [9]; they proved the almost complete convergence of this model in the independent and identically distributed (i.i.d.) case. The asymptotic normality of their model has been established later by Attouch et al. [10–12] in
both dependent and independent cases and unknown scale parameters. However, all these results are obtained in the complete data. In the incomplete data cases, we can refer to Derrar et al. (2020) and the references therein.

The main objective of this paper is to generalize the results of Boente and Vahnovanb [13]. Precisely, we prove the asymptotic normality of the constructed estimator by combining the ideas of robustness with those of unknown scale parameters. This result is obtained under standard conditions allowing us to explore the different structural axes of the subject, such as the robustness of the regression function and the correlation of the observation. We emphasize that, contrary to the usual case where the scale parameter is fixed, it must be estimated here, which makes it more difficult to establish the asymptotic properties of the estimator. However, although this difference is more important in the context of this work, we have been able to overcome it.

The paper is organized as follows. Section 2 is dedicated to the presentation of the robust estimator with unknown scale parameters. The needed assumptions and notations are given in Section 3. We state our main results in Section 4, and the proofs of the main results are relegated to the appendix. Sections 5 and 6 are devoted to the simulation and real data applications of our proposed methods. The conclusion is stated in Section 7.

2. The Robust Equivariant Estimators and Their Related Functional

Let us consider a functional stationary ergodic process \( Z_i = (X_i, Y_i)_{i=1,...,n} \) (see Laïb and Louani [14] for some definitions and examples). When the scale parameter is unknown, the robust estimator may be constructed based on two steps. Firstly, we estimate the scale parameter \( \sigma \) by the conditional median of the absolute deviation from the conditional median, that is,

\[
\sigma = s(x) = \text{MED}(|Y - m(x)||X = x) = \text{MAD}(F(\cdot|X = x)),
\]

where \( F(\cdot|X = x) = \mathbb{P}(Y \leq \cdot|X = x) \) is the conditional distribution of \( Y \) given that \( X = x \) and \( m(x) = \text{MED}(Y|X = x) = \inf\{y \in \mathbb{R}: F(y|X = x) \geq 1/2\} \) is the median of the conditional distribution. Then, for \( x \in \mathcal{F} \), the kernel estimator \( \mathbb{F}(x) = \text{MAD}(\mathbb{F}(\cdot|X = x)) \), where \( \mathbb{F}(y|X = x) \) is given by

\[
\mathbb{F}(y|X = x) = \frac{\sum_{i=1}^{n} K(h^{-1}d(x, X_i)) I_{(-\infty, y]}(Y_i)}{\sum_{i=1}^{n} K(h^{-1}d(x, X_i))},
\]

where \( K \) is a kernel function and \( h = h_n \) is a sequence of positive numbers which goes to zero as \( n \) goes to infinity. Next, the kernel estimator \( \theta(x) \), of the robust regression \( \theta(x) \), is the zero, with respect to \( a \), of the equation

\[
\Psi(x, a, \tilde{s}(x)) = 0,
\]

where

\[
\Psi(x, a, \tilde{s}(x)) = \sum_{i=1}^{n} K(h^{-1}d(x, X_i)) \psi_s(Y_i - a/\tilde{s}(x)) \sum_{i=1}^{n} K(h^{-1}d(x, X_i))
\]

(5)

3. Notation, Conditions, and Comments

Throughout the paper, when no confusion is possible, we will denote by \( C_i \) for \( i \in \{1, 2, 3, 4\} \) some strictly positive generic constants. \( x \) stands for a fixed point in \( \mathcal{F}, N_x \) denotes a fixed neighborhood of \( x \), and we set \( \lambda_p(u, a, \sigma) = \mathbb{E}[\psi_x^p(Y - a/\sigma)|X = u] \) and \( \Gamma_p(u, a, \sigma) = \mathbb{E}[\psi_x^p|Y = a/\sigma|X = u] \), for \( p \in [1, 2] \). We state the following conditions:

A1. The function \( \psi_x \) is a continuously differentiable function, being strictly monotone and bounded w.r.t. the second component, and its derivative \( \partial \psi_x(y - a/\sigma(x))/\partial a \) is bounded and continuous at \( \theta(x) \) uniformly in \( y \).

A2. There exists a nonnegative differentiable function \( \phi \) and a nonnegative function \( g \) such that \( \mathbb{P}(X \in B(x, r)) = \phi(r)g(x) + o(\phi(r)) \) where \( B(x, r) = \{x' \in \mathcal{F}|d(x, x') < r\} \).

A3. The function \( \lambda_p(\ldots, \cdot) \) satisfies a Lipschitz condition of order one; that is, there exists a strictly positive constant \( b_p \) such that

\[
\forall a \in \mathbb{R}, |\lambda_p(u_1, a, \sigma) - \lambda_p(u_2, a, \sigma)| \leq C_3d(u_1, u_2)^{b_p}.
\]

(6)

A4. The function \( \Gamma_p(\ldots, \cdot) \) satisfies a Lipschitz condition of order one; that is, there exists a strictly positive constant \( d_p \) such that \( \forall (u_1, u_2) \in N_x \times N_x, \forall a \in \mathbb{R}, |\Gamma_p(u_1, a, \sigma) - \Gamma_p(u_2, a, \sigma)| \leq C_3d(u_1, u_2)^{d_p} \).

A5. The functions \( f_p(x) \) and \( p(x) \) are bounded functions on \( C \) such that \( A_{p} = \inf_{x \in C} p(x) > 0 \) and \( A_{p} = \inf_{x \in C} f_p(x) > 0 \). Moreover, \( p(x) \) is a continuous function in a neighborhood of \( C \).

A6. (i) \( F(y|X = x) \) is a continuous function of \( x \) in a neighborhood of \( C \). Furthermore, it satisfies the following equicontinuity condition:

\[
\forall \varepsilon > 0, \exists \delta > 0: |u - v| < \delta \quad \Rightarrow \quad \sup_{x \in C} (|F(u|X = x) - F(v|X = x)|) < \varepsilon.
\]

(ii) \( F(y|X = x) \) is symmetric around \( \theta(x) \) and a continuous function of \( y \) for each fixed \( x \).

A7. The kernel \( K \) is a differentiable function supported on \( [0, 1] \). Its derivative \( K' \) exists and is such that there exist two constants \( C_1 \) and \( C_2 \) with \( -\infty < C_1 < K'(t) < C_2 < 0 \) for \( 0 \leq t \leq 1 \).
4. Main Results

The following result in Proposition 1 ensures uniform consistency of the regression function \( \theta(x) \), when the smoothing is based on either local medians or local M-smoothers, while Theorem 1 deals with the asymptotic normality of the proposed estimator. The proofs of these asymptotic results are postponed to the appendix.

**Proposition 1.** Assume that A5 to A7 hold. Moreover, assume that A9 holds for kernel weights and that A8 and A10 hold for nearest neighbor with kernel weights. Then, for any compact set C,

(a) under A1 and A6(ii), we have that 
\[
\sup_{x \in C} |\hat{\theta}_{\text{MED}}(x) - \theta(x)| \xrightarrow{a.s.} 0.
\]

(b) if, in addition, \( F(y|X = x) \) have a unique median at \( \theta(x) \), we have that
\[
\sup_{x \in C} |\hat{\theta}(x) - \theta(x)| \xrightarrow{a.s.} 0.
\]

We will now set \( \psi_x(t) = \sigma_x(t/\sigma) \).

**Lemma 1.** Under A1 and A6(ii), if there exists a real constant \( C > 0 \) and an increasing function \( \phi(h) \) such that 
\[ (n\phi(h))^{-1} \sum_{i=1}^{n} W_{i}^{2}(x, X_{1}, ..., X_{n}) \leq C, \]
with \( W_{i}(x) = K_{i}/\sum_{i=1}^{n} K_{i} \), then, we have that 
\[
(n\phi(h))^{1/2} \sum_{i=1}^{n} W_{i}(x) \left[ \psi_{x}(Y_{i} - \theta(x)) - \psi_{x}(Y_{i} - \theta(x)) \right] \xrightarrow{p} 0,
\]
for any sequence \( \sigma_{n} = \sigma_{n}(x) \) such that \( \sigma_{n}(x) \rightarrow \sigma = \sigma > 0 \) in probability.

**Theorem 1.** Under A1 to A4, A7, and A9, if in addition \( \hat{\theta}(x) \xrightarrow{p} \theta(x) \) and \( \tilde{z}(x) \xrightarrow{p} \sigma(x) \), for any \( x \in \mathcal{A} \), then, we have 
\[
\left( \frac{n\phi(h)}{\sigma^{2}(x, \theta(x))} \right)^{1/2} \left\{ \hat{\theta}(x) - \theta(x) - B_{n}(x) \right\} \xrightarrow{d} N(0, 1) \text{ as } n \rightarrow \infty,
\]
where
\[
B_{n}(x) = \frac{h}{\phi(h)\alpha_{1}\Gamma_{1}(x, \theta(x), \sigma)} \cdot \int_{0}^{1} K(t) \varphi_{e,x}(th) \phi(t) dt + o(1),
\]
\[
\sigma^{2}(x, \theta(x)) = \frac{\alpha_{2}\lambda_{2}(x, \theta(x), \sigma)}{\alpha_{1} \sigma(g(x))(\Gamma_{1}(x, \theta(x), \sigma))^{2}},
\]
with \( \varphi_{e,x}(t) = \mathbb{E}[\psi_{x}^{\prime}(Y - \theta(x)/\sigma)|d[X, x] = t] \), \( \alpha_{j} = -\int_{0}^{1} (K^{-1})^{j}(z) \varphi(z) dz = 1/2 \), \( \mathcal{A} = \{ z \in \mathcal{F}, g(z) \lambda_{2}(z, \theta(x), \sigma) \Gamma_{1}(z, \theta(x), \sigma) \neq 0 \} \), and \( \xrightarrow{d} \) means the convergence in distribution.

**4.1. Application to Conditional Confidence Interval.** An important application of the asymptotic normality result is the building of confidence intervals for the true value of \( \theta(x) \) given that curve \( X = x \). A plug-in estimate for the asymptotic standard deviation \( \sigma(x, \theta(x)) \) can be obtained using the estimators \( \hat{\lambda}_{1}(x, \hat{\theta}(x), \hat{\sigma}(x)) \) and \( \hat{\sigma}(x, \hat{\theta}(x), \hat{\sigma}(x)) \) of \( \lambda_{1}(x, \theta(x), \sigma(x)) \) and \( \sigma(x, \theta(x), \sigma(x)) \), respectively. We get
\[
\hat{\sigma}_{n}(x, \hat{\theta}(x)) = \left\{ \hat{\beta}_{2}\hat{\lambda}_{2}(x, \hat{\theta}(x), \hat{\sigma}(x))/(\hat{\beta}_{1}) \hat{g}(x)\hat{\Gamma}_{1}(x, \hat{\theta}(x), \hat{\sigma}(x)) \right\}^{1/2}.
\]
Then, \( \hat{\sigma}_{n}(x, \hat{\theta}(x)) \) can be used to get the following approximate \( (1 - \zeta) \) confidence interval for \( \theta(x) \):
This score function is known. We take \( \tilde{\theta} (x) \pm t_{1-\xi/2} \times \left( \frac{\bar{a}_n^2 (x, \tilde{\theta} (x))}{mn\Psi (x, h)} \right)^{1/2} \),  

\begin{equation}
\tilde{\theta} (x) \pm t_{1-\xi/2} \times \left( \frac{\bar{a}_n^2 (x, \tilde{\theta} (x))}{mn\Psi (x, h)} \right)^{1/2},
\end{equation}

where \( t_{1-\xi/2} \) denotes the \( 1 - \xi/2 \) quantile of the standard normal distribution. Here, we point out that the estimators and \((x, \tilde{\theta} (x), \tilde{\psi} (x))\) can be estimated by

\[ \tilde{\lambda}_2 (x, \tilde{\theta} (x), \tilde{\psi} (x)) = \frac{\sum_{i=1}^n K \left( \frac{h^{-1} d(x, X_i)}{d(x, X_i)} \right)}{\sum_{i=1}^n K \left( \frac{h^{-1} d(x, X_i)}{d(x, X_i)} \right)} \]

\[ \tilde{f}_1 (x, \tilde{\theta} (x), \tilde{\psi} (x)) = \frac{\sum_{i=1}^n K \left( \frac{h^{-1} d(x, X_i)}{d(x, X_i)} \right) \partial \frac{d\theta}{d\psi} (Y_i - \tilde{\theta} (x) \tilde{\psi} (x))}{\sum_{i=1}^n K \left( \frac{h^{-1} d(x, X_i)}{d(x, X_i)} \right)} \]

We estimate \( \beta_1 \) and \( \beta_2 \) by \( \tilde{\beta}_1 = 1/n\phi (x, h) \) and \( \tilde{\beta}_2 = 1/n\phi (x, h) g (x) \sum_{i=1}^n K \left( \frac{h^{-1} d(x, X_i)}{d(x, X_i)} \right) \).

This last estimation is justified by the fact that, under A2, A7, and A9, we have (see Ferraty and Vieu [2002, p. 44])

\[ \frac{1}{\phi (h) g (x)} E \left| \frac{K (x)}{\theta_j (x)} \right| \xrightarrow{\mathcal{N} (0, 1)} x_j, \quad j = 1, 2. \]

Then, by the asymptotic normality result in Section 4, we have

\[ \Lambda_n (x) = \left( \frac{n\phi (x, h)/\sigma^2 (x, \tilde{\theta} (x))}{1/2} \right) \left( \tilde{\theta} (x) - \theta (x) \right) \xrightarrow{p} \mathcal{N} (0, 1) \]  

as \( n \xrightarrow{} \infty \).

5. Simulation Study

The purpose of this section is to apply our method to an example of simulated data. More precisely, our main aim is to compare the sensitivity to the outliers of the classical kernel smoothing techniques. In this simulation, we take 200 samples of the curve \( X \) which is represented in Figure 1.

The scalar response variable is defined by \( Y = r (X) + e \), where \( r (X) \) is the non-linear regression model, and \( r (X) = 2.5 X (1 + \cos (\pi W)) \), where \( e \) is \( \mathcal{N} (0, 0.5) \). The selection of the bandwidth \( h \): \( h_k \) is an important and basic problem in all kernel smoothing techniques. In this simulation, for both methods, the optimal bandwidths were chosen by the cross-validation method on the \( k \) nearest neighbors, where \( h_{k_{opt}} \) is the bandwidth corresponding to the optimal number of neighbors obtained by a cross-validation procedure:

\[ h_k = \min \{ h \in \mathbb{R}^+ \mid \sum_{i=1}^n I_{B (x, h)} (X_i) = k \}, \]

with

\[ k_{opt} = \arg \min_k CV (k), \quad CV (k) = \sum_{i=1}^n \left( Y_i - \tilde{Y} (X_i) \right)^2, \]

where \( \tilde{Y} (X_i) \) is the leave-one-out density estimator of the CKE, KRE, and ERE (we refer the reader to Ferraty and Vieu [2] for more details). We choose the quadratic kernel:

\[ K (x) = \frac{3}{2} \left( 1 - x^2 \right) \text{I}_{|1, 0|}. \]

Another essential point for ensuring a good behavior of the method is using a semi-metric that is well adapted to the kind of data that we have to deal with. Here, we used the semi-metric defined by the \( L_1 \) distance between the first derivatives of the curves \( d_{\text{der}} (\cdot, \cdot) \) (for further discussion, see Ferraty and Vieu [2]).

\[ d_{\text{der}} (x_j, x_j) = \sqrt{\int_0^1 (x'_j (t) - x'_j (t))^2 dt, \forall x_j, x_j \in \mathcal{F}}. \]

This choice is motivated by the regularity of the curves \( X \). To compare these two methods, we split the data randomly into two subsets: training sample \( \{ X_i, Y_i \}_{i=1}^n \) and test sample \( \{ X_j, Y_j \}_{j=1}^n \) for measuring the forecasting accuracy of the responses.

To verify the theoretical results, it is possible to visualize the data histogram and then compare its shape to the normal density for a fixed \( X_{x} \in f \) (we carry out 10 independent replications).

The histogram of \( \Lambda_n (X_{x}) \) is almost symmetric around zero point and well-shaped like the standard normal density. Thus, the simulation results indicate that \( \Lambda_n (X_{x}) \) obeys the standard normal law when \( n \) is large (see Figure 2).

In order to construct confidence bands, we proceed using the following algorithm:
Step 1. For each $X_j$ in the test sample, we calculate the estimator $\hat{\theta}(X_j)$ using the training sample.

Step 2. For all $j \in J$, we define the confidence bands by

$$\left[ \hat{\theta}(X_j) - z_{0.975} \times \left( \frac{\hat{\sigma}^2(X_j, \hat{\theta}(X_j))}{n \phi(X_j, h)} \right)^{1/2}, \hat{\theta}(X_j) + z_{0.975} \times \left( \frac{\hat{\sigma}^2(X_j, \hat{\theta}(X_j))}{n \phi(X_j, h)} \right)^{1/2} \right],$$

(22)

where $z_{0.975}$ is the 2.5% quantile of a standard normal distribution.

Step 3. We present our results by plotting the extremities of the predicted values versus the true values and the confidence bands.

The results are presented in Figures 3 and 4, where the solid black curve connects the true values and the dashed blue curve connects the lower and upper predicted confidence interval (IC). Clearly, Figure 3 shows the excellent behavior of our functional forecasting procedure for the ERE method in the absence of outliers.

The performance of the estimators is described by the mean squared error (MSE):

$$\text{MSE} = \frac{1}{\#(J)} \sum_{j \notin J} \left( Y_j - \hat{Y}(X_j) \right)^2,$$

(23)

where $\#(J)$ is the length of testing sample $J$, and $\hat{Y}(X_j)$ means the regression estimators of CKE, KRE, and ERE calculated at the point $X_j$. The main feature of our approach is illustrated in the second case when we perturb the data by introducing some outliers as indicated below. In this part, following the same approach in Sinova et al. [18], let $\{\epsilon_i\}_{i=1}^n$ be independent sequences of random variables as in Bernoulli (with parameter $\epsilon_i \in [5\% : 10\%]$) and contamination size constant $M$ be either 5 or 25.

The data generating model is given by

$$Y_i^* = Y_i + M \epsilon_i, 1 \leq i \leq n.$$

(24)

In all these cases, we arrived at the same conclusion. Usually in the presence of outliers, the ERE regression shows better behavior than that of the CKE and KRE methods. Even if the MSE of the three methods increases substantially with the $\epsilon_i$ and with the value of contamination size constant $M$, it remains very low for the ERE method. The results are given in Table 1 and Figure 4 where we only present the case of $\epsilon_i = 10\%$ and $M = 25$, because the results are very similar to the other cases.

6. Real Data Analysis

We focus now on the comparison of our estimator ERE ($\hat{\theta}(x)$) with CKE ($\hat{r}(x)$) and KRE ($\tilde{\theta}(x)$) estimators for some spectroscopic datasets. For instance, let us consider a sample of 108 soil samples measured by near-infrared reflectance (NIR). Each sample is illuminated by a light beam at 1050 equally spaced wavelengths ($t_1, \ldots, t_{1050}$) in the range of 400–2500 nanometres (nm) with a 2 nm resolution. The dataset is available at http://www.models.kvl.dk/NIRsoil. For each wavelength $t$ and each soil sample $i$, the absorption $X_i(t)$ of radiation is measured. The $i$th discretized spectrometric curve is given by $X_i(t_1), \ldots, X_i(t_{1050})$; Figure 5 displays the 108 NIR reflectance spectra curves, where the red ones correspond to the outlier curves.

Moreover, to determine the chemical and microbiological properties of soil, ergosterol concentration was determined through High-Performance Liquid
Chromatography (HPLC), which is taken in the following as a response variable $Y$. The latter variable is affected by the presence of some outliers as shown in Figure 6.

In this application, we use the MAD-median rule (cf. Wilcox [19] for detecting outliers). Recall that this method will refer to declaring $Y_i$ an outlier if

$$\frac{|Y_i - M|}{MAD \times 0.6745} > C,$$

where $M$ is the sample median and MAD is the median absolute deviation given by

\begin{tabular}{|c|c|c|c|c|}
\hline
Contamination size constant $M$ & Parameter $\epsilon_0$ & MSE (ERE) & MSE (KRE) & MSE (CKE) \\
\hline
5 & 0.05 & 1.3892 & 0.9501 & 0.8108 \\
5 & 0.10 & 2.1959 & 1.7996 & 0.8685 \\
25 & 0.05 & 3.8701 & 1.8444 & 0.8912 \\
25 & 0.10 & 9.4202 & 2.3472 & 0.9269 \\
\hline
\end{tabular}
Moreover, $C$ is taken to be $\frac{1}{\sqrt{2.975}}$: the square root of the quantile of a chi-squared distribution with one degree of freedom. We applied this method to our sample data and the result shows that the MAD-median method identifies 21 outliers. The goal now is to predict the ergosterol concentration from the observation of a new spectrometric curve. To this end, we observe $n = 108$ pairs $(X_i, Y_i)_{i=1,...,n}$, where $X_i$ (resp., $Y_i$) is the $i$th spectrometric curve (resp., response).

The prediction problem is very simple and can be formulated via the model $Y_i = r(X_i) + \epsilon_i$ for $i = 1, \ldots, n$. Then, we split the observations $(X_i, Y_i)_{i=1,...,n}$ into two parts: one is the training sample containing $(X_i, Y_i)_{i=1,...,86}$ for modeling, and the other is $(X_j, Y_j)_{j=87,...,108}$ which is a testing sample for measuring the forecasting accuracy of the responses. In the modeling procedures, we choose the asymmetrical quadratic kernel function $K(u) = \frac{3}{4}(12/11 - u^2)^{\frac{1}{2}}(u)$ and the semi-metric $\rho_{d^{\text{deriv}}}(...) \rho_{0,1}(u)$ of the first derivative. As in all smoothing methods, the choice of the smoothing parameter has a crucial role in the computational issues. In this illustration, we use the cross-validation procedure described in Rachdi and Vieu (2007) for which the bandwidth $h$ is chosen via the following rule:

$$h_{\text{opt}} = \arg \min_h CV(h) = \arg \min_h \sum_{i=1}^{n} (Y_i - \hat{Y}^{(-i)}(X_i))^2,$$

where $\hat{Y}^{(-i)}$ is the leave-one-out-curve estimator of the CKE, KRE, and ERE. Finally, we predict the $(Y_j)_{j=87,...,108}$ by the three methods above and check the performance of both models by computing the mean squared prediction error (MSE), and the relative mean squared prediction error (RMSE):

$$\text{RMSE} = \frac{1}{22} \sum_{j \text{the testing sample}} \left( \frac{Y_j - \hat{Y}(X_j)}{Y_j} \right)^2,$$

where $\hat{Y}(X_j)$ means the regression estimators of CKE, KRE, and ERE calculated at the point $X_j$. Then, we display in Table 2 and Figure 7 these errors where it appears clearly that the ERE is quite more performant than the KRE and CKE. That is, the classical kernel method is susceptible to the presence of outliers.
Table 2: MSE and RMSE of the CKE, KRE, and ERE.

<table>
<thead>
<tr>
<th></th>
<th>CKE</th>
<th>KRE</th>
<th>ERE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean squared error (MSE)</td>
<td>6.898781</td>
<td>6.216957</td>
<td>4.807382</td>
</tr>
<tr>
<td>Relative mean squared error (RMSE)</td>
<td>0.000908</td>
<td>0.000819</td>
<td>0.000640</td>
</tr>
</tbody>
</table>

Next, the obtained prediction results are shown in Figure 8, while we give in Figure 9 the 95% predictive intervals of the concentrations for the ergosterol concentration in the sample test. Figure 9 also shows clearly a good behavior of the ERE estimator compared to the KRE and KRE ones. It appears in the band of confidence intervals which is very large for the CKE (0 to 200) and KRE (60 to 120) as compared to the ERE (80 to 100). This conclusion shows the good performance of our asymptotic normality.

Figure 7: Box plots of the MSE (a) and RMSE (b) of the CKE, KRE, and ERE.

Figure 8: Prediction of the last 15 testing samples. The solid black curve connects the true values.
7. Conclusion

This work generalizes the results of Boente et al. [13] in real response random variables to the functional response random variables. Our results are applied to derive asymptotic normality of the predictor estimate and build conditional confidence intervals. The results were obtained under sufficient standard conditions that allow one to explore different structural axes of the subject, such as the functional naturalness of the model and the data as well as the robustness of the regression function.

Appendix

Proof of Proposition 1. To prove Proposition 1, the authors begin by fixing some notations. If for any measurable $A \subseteq \mathbb{R}$, note that $\hat{\phi}_A(x) = \tilde{r}_A(x)/\tilde{p}(x)$ where

$$\tilde{r}_A(x) = \sum_{i=1}^n W_{i,n}(x) I_A(y_i),$$

$$\tilde{p}(x) = \sum_{i=1}^n W_{i,n},$$

$$W_{i,n}(x) = \frac{K(d(x_i,x)/h_n)}{\sum_{j=1}^n K(d(x_j,x)/h_n)}.$$  \hspace{1cm} (A.1)

(a) Following Theorem 3.3 in Boente and Fraiman [20], we will only need to show that

$$\sup_{x \in C} \sup_{y \in \mathbb{R}} \left| \int \hat{r}(y|x) - r(y,x) \right| \overset{a.s.}{\longrightarrow} 0, \hspace{1cm} (A.5)$$

$$\sup_{x \in C} \left| \tilde{p}(x) - p(x) \right| \overset{a.s.}{\longrightarrow} 0, \hspace{1cm} (A.6)$$

where $r(y,x) = \phi_{(-\infty,p)}(x) = p(x)F(y|X = x)$, $	ilde{r}(y,x) = \hat{\phi}_{(-\infty,p)}(x)$, with $\hat{\phi}_{(-\infty,p)}(x)$ and $\tilde{p}(x)$ being defined in (A1) and (A2), respectively. The weights $W_{i,n}$ are the kernel weights given by (A.3) or the nearest neighbor with kernel weights

$$W_{i,n}(x) = \frac{K(d(x_i,x)/H_n(x))}{\sum_{j=1}^n K(d(x_j,x)/H_n(x))}.$$ \hspace{1cm} (A.7)

Note that (A.6) can be derived for kernel weights using Proposition 2 in Collomb [21]. Now, (6) follows from A5 and the inequality

$$\sup_{x \in C} \sup_{y \in \mathbb{R}} \left| \int \hat{F}(y|X = x) - F(y|X = x) \right|$$

$$\leq \sup_{x \in C} \sup_{y \in \mathbb{R}} \left| \int \hat{r}(y,x) - r(y,x) \right| + \sup_{x \in C} \left| \tilde{p}(x) - p(x) \right|$$

$$\leq A_p A_{\tilde{p}},$$ \hspace{1cm} (A.8)

where $A_p = \inf_{x \in C} p(x)$ and $A_{\tilde{p}} = \inf_{x \in C} \tilde{p}(x)$.

(b) The equicontinuity condition required in A6 and the uniqueness of the conditional median imply that $\theta(x)$ is a continuous function of $x$, and thus for any fixed $a \in \mathbb{R}$, the function $h_n(x) = F(a + \theta(x)|X = x)$ will also be continuous as a function of $x$. Given $\epsilon > 0$, let $0 < \delta < \epsilon$ be such that

$$|u - v| < \delta \Rightarrow \sup_{x \in C} \left| \int F(u|X = x) - F(v|X = x) \right| < \frac{\epsilon}{2},$$ \hspace{1cm} (A.9)
Then, from the uniqueness of the conditional median and (A.9), we get that
\[
\frac{1}{2} < F(\theta(x) + \delta | X = x) < \frac{1}{2} + \epsilon, \quad (A.10)
\]
\[
\frac{1}{2} - \epsilon < F(\theta(x) - \delta | X = x) < \frac{1}{2}, \quad (A.11)
\]
which implies that
\[
\frac{1}{2} < \hat{F}(\theta(x) + \delta | X = x) < \frac{1}{2} + \text{cand} \frac{1}{2} - \epsilon \quad (A.14)
\]
\[
< \hat{F}(\theta(x) - \delta | X = x) < \frac{1}{2}
\]
Therefore, \( \sup_{x \in C} [\hat{\theta}_{ME} (x) - \theta(x)] \leq \delta < \epsilon \) which concludes the proof. \( \square \)

**Proof of Lemma 1.** For each fixed \( x \), the authors define
\[
H_t(u) = \psi_{\sigma_e}(u) - \psi_{\sigma}(u), I_t(u)
\]
\[
J_n^+(t) = (n \phi(h))^{-1/2} \sum_{i=1}^n W_m(x, X_1, \ldots, X_n)H_t(Y_i - \theta(x)),
\]
\[
\hat{J}_n^+(t) = J_n^+(t) - \mathbb{E}(J_n^+(t)),
\]
\[
J_n^-(t) = (n \phi(h))^{-1/2} \sum_{i=1}^n W_m(x, X_1, \ldots, X_n)I_t(Y_i - \theta(x)),
\]
\[
\hat{J}_n^-(t) = J_n^-(t) - \mathbb{E}(J_n^-(t)).
\]
(A.15)
The proof will be complete if we show
\[
\lim_{d \to 0} \lim_{n \to \infty} \mathbb{P}\left( \sup_{0 \leq t \leq d} | \hat{J}_n^+(t) | > \epsilon \right) = 0, \quad (A.16)
\]
\[
\lim_{d \to 0} \lim_{n \to \infty} \sup_{0 \leq t \leq d} \mathbb{E} | J_n^+(t) | = 0, \quad (A.17)
\]
and the analogous result with \( J_n^+(t) \) replaced by \( J_n^-(t) \). Equation (A.17) follows easily from the dominated convergence theorem. In order to show (13), it is enough to prove that the sequence \( I_n^+(t) \) of random variables on the space \( C[0,1] \) of continuous function on \( [0,1] \) is tight. According to Theorem 12.3 of Billingsley [22], it suffices to verify the following:

(i) The sequence \( \{I_n^+(0)\} \) is tight.

(ii) There exist constants \( \gamma > 0 \) and \( \alpha > 1 \) and a nondecreasing continuous function \( F \) on \( [0,1] \) such that
\[
\mathbb{E}\left| \hat{J}_n^+(t_2) - \hat{J}_n^+(t_1) \right|^a \leq (F(t_2) - F(t_1))^a
\]
holds for all \( 0 \leq t_1 < t_2 \) and \( n \) large enough. Since \( \hat{J}_n^+(0) = 0 \), (i) follows:
\[
\mathbb{E}\left( \left( \hat{J}_n^+(t_2) - \hat{J}_n^+(t_1) \right)^2 \right) = \mathbb{E}\left( \left( H_{t_2}(u) - H_{t_1}(u) \right)^2 \right).
\]
(A.19)

As in lemma A of Fraiman [23], the authors have that for \( t_2 > t_1 \)

\[
|H_{t_2}(u) - H_{t_1}(u)| \leq d^\ast (t_2 - t_1) + \int_{-a(\sigma+t_1)}^b \sup_{s \in [\sigma+t_1]} |\psi''(s)| \frac{s}{\sigma+t} - \frac{s}{\sigma+t_1} ds,
\]
\[
\leq d^\ast (t_2 - t_1) + \alpha \left( \frac{\alpha}{\sigma+t_1} \right) \left( \frac{-1}{\sigma+t} \right) \left( \frac{b^2}{2} \right) \text{ for } b > a(\sigma+t_1),
\]
\[
= d^\ast t_2 - \frac{db^2}{2} (\sigma+t) - \left( d^\ast t_1 - \frac{db^2}{2} (\sigma+t_1) \right),
\]
(A.20)

\[
\text{In addition,}
\]
\[
|H_{t_2}(u) - H_{t_1}(u)| \leq d^\ast t - \frac{db^2}{2} (\sigma+t).
\]
\[ d^* = (aC + |\Psi_N'(-a)|) \] is a positive constant. (A.21)

Then,

\[ |H_{t_2}(u) - H_{t_1}(u)| \leq F(t_2) - F(t_1), \] (A.22)

which implies that \( \mathbb{E}(J_n(t_2) - J_n(t_1))^2 \leq C(F(t_2) - F(t_1))^2 \).

Finally, a similar argument shows that the same result holds for \( J_n(t) \).

**Proof of Theorem 1.** For \( i = 1, \ldots, n \), the authors consider the quantity \( K_i = K(h^{-1}d(x, X_i)) \), and using Taylor’s expansion of order one, the authors get the following:

\[ \tilde{\theta}(x) - \theta(x) = \frac{\tilde{\Psi}_N(x, \theta(x), \tilde{s}(x))}{\Psi_N'(x, \tilde{\xi}_n, \tilde{s}(x))}, \] (A.23)

with

\[ (n\phi(h))^{1/2}
\begin{pmatrix}
\tilde{\Psi}_N(x, \theta(x), \tilde{s}(x)) - \Psi_N(x, \theta(x), \sigma)
\end{pmatrix}
\xrightarrow{p} 0, \quad (n\phi(h))^{1/2}
\begin{pmatrix}
\tilde{\Psi}_N(x, \theta(x), \tilde{s}(x)) - \tilde{\Psi}_N(x, \theta(x), \tilde{s}(x)\big)
\end{pmatrix}
\xrightarrow{p} a_1, \]

where

\[ \Psi_N(x, \theta(x), \sigma) = \frac{1}{n\phi(h)} \sum_{i=1}^{n} K_i \psi_i \left( \frac{Y_i - \theta(x)}{\sigma} \right), \]

\[ a_1 = \frac{1}{n\phi(h)} 2\tilde{s}(x)^{-1} h g(x) \int_0^1 K(z) \varphi_{x,s}(hz) \phi'(hz) dz + o(\phi(h)), \]

To prove (a), the authors consider the following decomposition:

\[ |\tilde{\Psi}_N'(x, \tilde{\xi}_n, \tilde{s}(x)) - \Gamma_1(x, \theta(x), \tilde{s}(x))| \leq |\tilde{\Psi}_N'(x, \xi_n, \tilde{s}(x)) - \tilde{\Psi}_N'(x, \theta(x), \tilde{s}(x))|, \]

\[ + |\tilde{\Psi}_N'(x, \xi_n, \tilde{s}(x)) - \Gamma_1(x, \theta(x), \tilde{s}(x))|, \]

Concerning the first term, observe that \( |\tilde{\Psi}_N'(x, \xi_n, \tilde{s}(x)) - \tilde{\Psi}_N'(x, \theta(x), \tilde{s}(x))| \leq \sup_{y \in R} |\partial \psi_i(y, \xi_n, \tilde{s}(x))| \frac{dy}{dt} - \partial \psi_i(y, \theta(x), \tilde{s}(x))| \frac{dy}{dt} \) and because \( \partial \psi_i(y - a/\tilde{s}(x)) \) \( \partial t \) is continuous at \( \theta(x) \) uniformly in \( y \), the use of the convergence in probability of \( \tilde{\Psi}_N'(x) \) to 1 shows that the first term of (15) converges in probability to 0.
However, the limit of the second term is obtained by evaluating the bias term of \( \bar{\Psi}_N(x, \theta(x), \bar{s}(x)) \). By stationarity, we have

\[
\mathbb{E}[\bar{\Psi}_N(x, \theta(x), \bar{s}(x)) - \Gamma_1(x, \theta(x), \bar{s}(x))]
\leq \mathbb{E} \left[ \frac{1}{\phi(h)} K_1 \mathbb{E}_{B(x,h)} (X_i) \left( \Gamma_1 (X_i, \theta(x), r\bar{s}(x)) - t\Gamma_1 n(x, \theta(x), \bar{s}(x)) \right) \right].
\]

(A.29)

Under A4 and the boundness of \( K \), the authors get

\[
\mathbb{E} \left[ \frac{1}{\phi(h)} K_1 \mathbb{E}_{B(x,h)} (X_i) \left( \Gamma_1 (X_i, \theta(x), r\bar{s}(x)) - t\Gamma_1 n(x, \theta(x), \bar{s}(x)) \right) \right] \leq C h^{d_i}.
\]

(A.30)

From this, it follows that the proof of (a) (b) can be deduced from Lemma 1, (c) follows from using Taylor's expansion of order two. Effectively, denote \( Z_i = \psi_x' (\epsilon_i) [\theta(x_i) - \theta(x)] \), where \( \epsilon_i = (y_i - \theta(x_i))/\sigma(x_i) \).

\[
\begin{aligned}
& + \frac{1}{2} \frac{1}{\bar{s}(x)^2} \frac{1}{\sqrt{n} \phi(h)} \sum_{i=1}^{n} K \left( \frac{d(x_i, x)}{h} \right) \psi''_x \left( \frac{\xi_i}{\bar{s}(x)} \right) \left[ \theta(x_i) - \theta(x) \right]^2 \\
& = \frac{1}{\bar{s}(x) \sqrt{n} \phi(h)} \sum_{i=1}^{n} K \left( \frac{d(x_i, x)}{h} \right) Z_i \\
& + \frac{1}{2} \frac{1}{\bar{s}(x)^2} \frac{1}{\sqrt{n} \phi(h)} \sum_{i=1}^{n} K \left( \frac{d(x_i, x)}{h} \right) \left[ \sigma(x_i) - \sigma(x) \right] \psi''_x \left( \frac{\xi_i}{\bar{s}(x)} \right) \left[ \theta(x_i) - \theta(x) \right] \\
& + \frac{1}{2} \frac{1}{\bar{s}(x)^2} \left[ \sigma(x) - \bar{s}(x) \right] \frac{1}{\sqrt{n} \phi(h)} \sum_{i=1}^{n} K \left( \frac{d(x_i, x)}{h} \right) \psi''_x \left( \frac{\xi_i}{\bar{s}(x)} \right) \left[ \theta(x_i) - \theta(x) \right] \\
& + \frac{1}{2} \frac{1}{\bar{s}(x)^2} \frac{1}{\sqrt{n} \phi(h)} \sum_{i=1}^{n} K \left( \frac{d(x_i, x)}{h} \right) \psi''_x \left( \frac{\xi_i}{\bar{s}(x)} \right) \left[ \theta(x_i) - \theta(x) \right]^2 \\
= S_{2n} + S_{3n} + S_{4n} + S_{4n},
\end{aligned}
\]

with \( \xi_i \) and \( \xi_i \) being intermediate points. Using the boundness of \( \psi''_x \), the Lipschitz continuity of \( \theta \), and the continuity of \( \sigma \) together with the consistency of the scale estimator, the authors get that \( S_{3n} \rightarrow 0 \) for \( 2 \leq j \leq 4 \).

On the other hand,

\[
\mathbb{E}[\bar{s}(x)S_{4n}] = n (n\phi(h))^{-1/2} \mathbb{E} [K_1 Z_i].
\]

(A.32)

By conditioning with respect to real variable \( d(x, X_i) \), the authors get

\[
\mathbb{E}[\bar{s}(x)S_{4n}] = n (n\phi(h))^{-1/2} \mathbb{E} [K_1 \mathbb{E}[Z_i|d(X_i, x)]].
\]

(A.33)

The integration w.r.t. the distribution of the real variable \( d(x, X) \) shows that

\[
\mathbb{E}[\bar{s}(x)S_{4n}] = n (n\phi(h))^{-1/2} h g(x) \int_0^1 K(h^{-1} t) \phi_{\sigma, X}(t) \phi'(t) dt + o(\phi(h)).
\]

(A.34)

It follows from the usual change of variables \( h^{-1} t = s \) that
\[ E[\tilde{S}(x)S_{in}] = n(1/n\phi(h))^{1/2}h\varphi(x) \int_0^1 K(s)\varphi_{s,x}(hs)\phi'(hs)ds + o(\phi(h)). \]  

(A.35)

Then, \( S_{in} \overset{p}{\longrightarrow} a_1. \)

In order to prove (d), let \( Z = \psi_x(y_t - \theta(x))/\sigma. \)

Therefore, the main point is to calculate the following limit:

\[ \text{Var}[\bar{\Psi}_N(x, t, \sigma)] = \frac{1}{n\phi^2(h)} \sum_{i=1}^n \text{Var}[K_i\psi_i(t, \sigma)] \]

\[ = \frac{1}{n\phi^2(h)} \left[ K_1^2\psi_x^2(Y - t/\sigma) \right] - \frac{1}{n\phi^2(h)} \left( \frac{1}{\phi^2(h)} \left[ K_1^2\psi_x(Y - t/\sigma) \right]^2 \right) \]

(A.37)

Accordingly,

\[ (n\phi(h))^{1/2} \text{Var}[\bar{\Psi}_N(x, t, \sigma)] = \frac{E\left[ K_1^2 \psi_x(Y - t/\sigma) \right]}{n^{1/2} \phi^{3/2}(h)} \left( \frac{E\left[ K_1 \psi_x(Y - t/\sigma) \right]^2}{E\left[ K_1^2 \right]} \right) \]

\[ - \frac{(n\phi(h))^{1/2}}{n} \left( \frac{E\left[ K_1 \psi_x(Y - t/\sigma) \right]}{\phi(h)} \right)^2. \]

(A.38)

It suffices now to show that

\[ \forall t \in \mathbb{R}, \frac{K_1\psi_x(Y - t/\sigma)}{\phi(h)} \rightarrow \lambda_1(x, t, \sigma). \]  

Indeed, a simple calculation gives us

\[ \left| \frac{E\left[ K_1 \psi_x(Y - t/\sigma) \right]}{\phi(h)} - \lambda_1(x, t, \sigma) \right| = \frac{1}{\phi(h)} \left| E\left[ K_1 \right]_B(x, h) \right| (\lambda_1(X_1, t, \sigma) - \lambda_1(x, t, \sigma)). \]  

(A.40)

Moreover, by A3 (with \( y = 1 \)) the authors have

\[ \left| I_1(X_1) \right| \lambda_1(X_1, t, \sigma) - \lambda_1(x, t, \sigma) \right| \leq c h^3. \]  

(A.41)

Then,

\[ \left| \frac{E\left[ K_1 \psi_x(Y - t/\sigma) \right]}{\phi(h)} - \lambda_1(x, t, \sigma) \right| \leq c h^3 \rightarrow 0. \]  

(A.42)

Consequently, (20) holds.

On the other hand, for the first term of the right hand side of (19), by using analogous arguments to those considered to derive (20), we show that

\[ \frac{E\left[ K_1^2 \psi_x^2(Y - t/\sigma) \right]}{E\left[ K_1^2 \right]} \rightarrow \lambda_2(x, t, \sigma). \]  

(A.43)

Indeed, by (A.1), we can easily get
Finally, again by (A.3) (with $\gamma = 2$), we get

$$\frac{1}{\mathbb{E}[K_1]} \mathbb{E}[K_1^{1/2}(X_1) \lambda_2(X_1, t, \sigma) - \lambda_2(x, t, \sigma)] \leq C h^b.$$  \hfill (A.45)

Next, by using the distribution of the real variable $Z = d(x, X)$, the authors can show that, under (A.2) and (A.4) (see Ferraty and Vieu (2006, p. 44)),

$$\forall j > 0 \mathbb{E}[K_1] = -g(x) \int_0^1 (K_1) (t \phi(th)dt + o(\phi(h))).$$  \hfill (A.46)

It follows that

$$\mathbb{E}\left[\frac{K_1}{n^{1/2} \phi^{1/2}(h)} \right] \longrightarrow g(x) \frac{\alpha}{n^{1/2} \phi^{1/2}(h)},$$

Then, using (20)–(22), we get

$$n \phi(h) \frac{1/2 \text{Var}(\Psi_n(x, t, \sigma))}{\mathbb{E}\left[\frac{K_1}{n^{1/2} \phi^{1/2}(h)} \right]} \longrightarrow g(x) \frac{\alpha_2 \lambda_2(x, t, \sigma)}{n^{1/2} \phi^{1/2}(h)}.$$  \hfill (A.48) \hfill \square

**Data Availability**

The real data that support the findings of this study are available at http://www.models.kvl.dk/NIRsoil.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

The authors thank and extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through the Research Groups Program under grant no. R.G.P. 1/177/43.

**References**


