STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–12.
Published online in International Academic Press (www.IAPress.org)

On the Use of Yeo-Johnson Transformation in the Functional Multivariate Time Series

Sameera Abdulsalam Othman[1], Haithem Taha Mohammed Ali[2,*]

[1] *Department of Mathematics, College of Basic Education, University of Dohuk, Kurdistan Region, Iraq*
[2] *Department of Economics Science, University of Zakho, Kurdistan Region, Iraq*
[3] *Department of Economics, Nawroz University, Kurdistan Region, Iraq*

Abstract Box-Cox and Yeo-Johnson transformation models were utilized in this paper to use density function to improve multivariate time series forecasting. In this article, the transformations are used to improve the forecastability of the nonparametric multivariate time series. The K-Nearest Neighbor function is used in our model, with automatic bandwidth selection using a cross-validation approach and semi-metrics used to measure the proximity of functional data. Then, to decorrelate multivariate response variables, we use principal component analysis. The methodology was applied to two time series data examples with multiple responses. The first example includes three time series datasets of the monthly average of Humidity (H), Rainfall (R) and Temperature (T). The simulation studies are provided in the second example. Mean square errors of predicted values were calculated to show forecast efficiency. The results have proved that applying multivariate nonparametric time series transformed stationary datasets using the Yeo-Johnson model is more efficient than applying the univariate nonparametric analysis to each response independently.

Keywords Box-Cox transformation, Yeo-Johnson transformation, ARIMA models.

DOI: 10.19139/soic-2310-5070-1569

1. Introduction

*T*here are many nonlinear time series structures, so nonparametric regression methods to estimate time series are varied. The Kernel method provides excellent results in the detection of nonlinear dependencies in time series and prediction in smooth regression [1, 2]. , It has been demonstrated that the functional estimation for kernel regression has asymptotic normality under reliance. Simultaneously, Antoniadis [3]. suggested a functional wavelet kernel technique for time series forecasting at the same time. Perez G, Vieu use a semi-functional partial linear model to forecast nonparametric time series [4]. In [5, 6] , the time depending on observations in some datasets is referred to the estimates that based on the principal component technique may be incorrect [7]. As a result, the authors see that this issue may be worsened in certain time series data, particularly those with seasonal variations. However, in actual applications of time series, it has become clear that they are rarely stable, with seasonal fluctuations, trend, and reliance on external influences becoming the norm rather than the exception [8]. These problems increase and become more complicated in multivariate time series. Consequently, Traditional parametric and nonparametric analysis of complex time series now includes data transformation. The authors of present article used the Yeo-Johnson transformations to improve the nonparametric estimation of functional multivariate time series. Our work aims to create a new model that will add Yeo-Johnson transformations in multivariate time series with functional modeling benefits. This article is organized as follows: section two includes the theoretical aspects of

*Correspondence to: samerra abdulsalam(Email: sameera.othman@nuod.ac). Department of Mathematics, College of Basic Education, University of Dohuk, Kurdistan Region, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the Kernel estimation of Multivariate response. section three include some theoretical aspect of the use of Box-Cox transformation BCT and Yeo-Johnson transformation YJT in univariate response. section four includes Asymptotic properties of the functional covariates and multivariate response. while section five include the application part of the article.

2. Kernel estimation of Multivariate response

In Multivariate instances, n random vectors $X_t, , X_2, \ldots, X_n$ with values in R^p are seen. In this case, a multivariate kernel can benefit from kernel local weighting. $K^\star$ is a function from R^p, into R and can be considered to be adequate.The first method is to define $K^\star$as a product of p real kernel functions.$K_t, , K_2, \ldots, K_p \forall u = (u_t, , u_2, \ldots, u_p)^t \in R^p, K^\star(u) = K_t(u_1) \times K_2(u_2) \times \ldots \times K_p(u_p)$. The second way, according to Hardle and Muller [9], is a real kernel function H with a norm (denoted by $\|.\|$ in R^p. as follows: $\forall u \in R^p$ $K^\star(u) = K(\|u\|)$.It is worth noting that if $\|.\|$ is the supermom norm and if $K_1 = K_2, \ldots K_p = 1_[-1_,1]$, both methods coincide taking $1_[0_,1]$. Because the real kernel K is always a positive amount, the $\|u\|)$ real kernel K should also have positive support (i.e.,{v $\in R$ such that $K(u) > 0\} \subset R^+$).When **x** is a fixed vector of R^p the multivariate kernel local weighting consists in translating the n random vectors $x_t, , x_2, \ldots, x_n$ into the n variables $\Delta_t, \Delta_2, \ldots, \Delta_n$:

$$\Delta_i = \frac{1}{h^p} K^\star(\frac{x - x_i}{h}) \tag{1}$$

Where Δ_i are weighted local transformations of the variables x_i, since $\Delta_i = 0$ and x_i is not in some neighborhood of x, the normalization $\frac{1}{h^p}$ is proportional to the volume of the set on which the $x_i^{,}s$ are used into account. A naive functional application of multivariate kernel local weighting ideas would be to convert the n functional random variables $X_t, , X_2, \ldots, X_n$ into the n amounts,

$$\frac{1}{v(h)} K(\frac{d(x, x_i}{h}) \tag{2}$$

Where denotes a real (asymmetrical) kernel and d denotes a semi-metric on E. $v(h)$ V (h) in this expression is the volume of $B(x, h) = \{\acute{x} \in Ed(x, \acute{x})\}$, where (h) is a positive, real bandwidth (depending on n) and K(.) is a symmetrical kernel. The neighbors of the optimal bandwidth, K_{opt},of the KNN estimator are defined by

$$h_{kopt} = \arg\min_h GCV(K)$$

where,

$$GCV = \sum_{i=1}^{n} (y_i - m_{KNN}^{-1}(x_i))^2 \tag{3}$$

With,

$$m_{KNN}^{-i}(\mathbf{x}) = \frac{\sum_{j=1, j\neq i}^{n} y_i K\left(\mathbf{d}_q(\mathbf{x}_j, \mathbf{x})/h_k(x)\right)}{\sum_{j=1, j\neq i}^{n} K\left(\mathbf{d}_q(\mathbf{x}_j, \mathbf{x})/h_k(x)\right)}. \tag{4}$$

The user must resolve the semi-metric (d (.,.)) and the kernel function K(.) [10], selecting the $h_{kopt}(x)$ on x because utilizing the same number of neighbors at every curve, but k_{opt} is the same for any curve. According to [10], semi-metrics best method to compute functional data. The semi-metric built on FPCA is defined as,

$$d_q^{FPCA}(\mathcal{X}_i, \chi_j) = \sqrt{\sum_{k=1}^{q} \left(\int [\mathcal{X}_i(t) - \chi_j(t)] \nu_k(t) dt \right)^2}., \tag{5}$$

3. Box-Cox and Yeo-Johnson Transformations of Univariate Response (BCT and YJT)

To achieve normality of random errors, transformation is used Box and Cox When the series is not stationary [11, 12], taking differences for the data to make the series stationary, and then the series contains positive and negative values. So adopting the Yeo and Johnson method summarized the Box-Cox transformation to incorporate negative and positive values in collections data [13]. They utilized a perfection condition to connect positive and negative data changes, resulting in a one-parameter transformation family [14]. The YJT for $Z \in R$ is provided by, (where Z response of). Consider the univariate time series $\{Z_t \in R\}$. By dividing the time series sample again into $(p-1)$ size $(n = N - s - p + 1)$ statistics samples.

$$\psi(\lambda) = \begin{cases} ((z+1)^{\lambda} - 1) & \lambda \neq 0 \text{ and } z \leq 0 \\ \ln(z+1) & \lambda = 0 \text{ and } z \leq 0 \\ -((-z+1)^{2-\lambda} - 1)/(2-\lambda) & \lambda \neq 2 \text{ and } z < 0 \\ \ln(-z+1) & \lambda = 2 \text{ and } z < 0 \end{cases} \tag{6}$$

Besides, the BCT may not result in an improvement in forecasting execution in some realistic examples of time series [15, 16]. According to CW, Lee JC state [17], it does not reliably create prevalent forecasts. Accordingly, the relationship can depict a standard regression model [18]:

$$Y = m(\mathbf{X}) + \epsilon \tag{7}$$

Where $m(\mathbf{X})$ denotes smooth functional data and ϵ is white noise a sequence consisting of independent identically distributed functions with $E[\varepsilon/\mathbf{X}] = 0$. A kernel estimator is a function that is evaluated at a given function $m(\mathbf{X})$ by,

$$\hat{m}(\mathbf{X}) = \frac{\sum_{i=1}^{n} Y_i K(h^{-1}(\mathbf{d}(\mathbf{X}, \mathbf{X}_i))}{\sum_{i=1}^{n} K(h^{-1}(\mathbf{d}(\mathbf{X}, \mathbf{X}_i))}. \tag{8}$$

Where $\mathbf{X}_i$= $(Z_{(i-p+1)}, \ldots, Z_i)$, $Y_i = Z_{(i+s)}$, $i = p, \ldots, N-s$, $N = n\tau$ for some n $\in$ N and some $\tau > 0$ The application approach incorporates estimating the smooth functional data $m(\mathbf{X})$ in the regression Eq (7). as indicated by the kernel estimator Eq. (8) after changing for each curve in the time series. The Yeo-Johnson Transformations method was used on the stationary time series, whereas the Box-Cox Transformations method was used on the actual time series. As a result, the expression redefined the measurable sample of curves.

$$\psi_{\lambda}(\mathbf{X}_i) = \psi_{\lambda}(Z(t)), (i-1)\tau < t \leq i\tau \tag{9}$$

Also, the response expression by the,

$$\psi_{\lambda}(Y_i) = \psi_{\lambda}(Z(i\tau + s)) \tag{10}$$

where ψ_{λ} denoted a data transformation by the power λ , i=1, ..., n-1. The choice principle acquired for picking the ideal estimate of power λ for each change model is that which corresponds to the most minimum assessments estimates of the mean squares errors of the forecasting the last curve of functional variable as represented by Eq. $MSE(\mathbf{X}_n)$=$(1/s)$ $\sum_{j=1}^{s}$ $(\hat{Z}_j - Z_j)^2$, where, Z_j are the j-th real values and $\hat{Z}_j$denotes the forecast value in the last curve. $\hat{Z}_j$ values are calculated from the inversions of Box-Cox Transformations and Yeo-Johnson Transformations, or from the re-transformation of the modified data metric to the original metric. In univariate, the application algorithm of the Box-Cox Transformations and Yeo-Johnson Transformations models, as well as the nonparametric estimate of the converted functional time series dataset, were the following:

i . Fix τ to define the expressions Eq .(9) and Eq. (10).
ii . To create stationary time series, subtract the seasonality patterns using the differences.
iii . take $\lambda \in \Lambda$, where Λ ={-3,3}.
iv . For each $\lambda \in \Lambda$, The original time series $Z(t)$is transformed using Box-Cox Transformations. Yeo-Johnson Transformations is used to convert the stationary series dataset of k differences Δ^k $Z(t)$ to provide the functional matrices with two explanations $\psi_{\lambda}(\mathbf{X})$ =$[\psi_{\lambda}(Z)]_{n\times\tau}$ and $\psi_{\lambda}(\mathbf{X})$ = $[\psi_{\lambda}(\Delta^k Z)]_{n\times\tau}$, for additional about the grids fille putting together in the R program [19, 20].

Eq.(7) is extended in our regression model below.

$$\psi_\lambda(Y_t)) = m(\psi_\lambda(\mathbf{X}_t)) + \epsilon \tag{11}$$

i=1,..., T, Where $\psi_\lambda(Y_t)$ = $(\psi_\lambda(Y_{t,1}),\ldots,(\psi_\lambda(Y_{t,d}))^T \in \Re^d$,$m(\psi_\lambda(\mathbf{X}_t))$ = $m_1(\psi_\lambda(X_{t,1}),\ldots,m_d(\psi_\lambda(X_{t,d}))^T$ the explanatory variable is functional (that is,$\psi_\lambda(\mathbf{X}_t)$ it accepts values in a potentially infinite-dimensional space). For the kernel approaches suggested in this paper, it is often preferable to ignore the correlation structure entirely the ostensible "working independence estimator", e.g [21]. Adapting a kernel-based strategy described by [18] for estimating $m(X)$ in the time series regression model for multivariate explanatory variables x to a functional situation. Xiao et al. demonstrated that their method is more efficient than the traditional local polynomial method [18]. Because the regression function is nonlinear, the primary idea is to alter the original data. This transformation is dependent on the function $m(\cdot)$. According to the kernel estimator listed below,

$$\hat{m}(\psi_\lambda(\mathbf{X})) = \frac{\sum_{i=1}^n Y_i K(h^{-1}\left(\mathbf{d}(\psi_\lambda(\mathbf{X}), \psi_\lambda(\mathbf{X}_i))\right)}{\sum_{i=1}^n K(h^{-1}\left(\mathbf{d}(\psi_\lambda(\mathbf{X}), \psi_\lambda(\mathbf{X}_i))\right)}. \tag{12}$$

The best value λ^* the power λ is the one that minimizes the $MSE(\mathbf{X}_n)$ of the last functional variable [22, 23]. Our target model below, based on the four-step algorithm mentioned before, is to expand Eq.(12) and convert it into multivariate Case.

$$\hat{m}(\psi^*_\lambda(\mathbf{X})) = \frac{\sum_{i=1}^n Y_i K(h^{-1}\left(\mathbf{d}(\psi^*_\lambda(\mathbf{X}), \psi^*_\lambda(\mathbf{X}_i))\right)}{\sum_{i=1}^n K(h^{-1}\left(\mathbf{d}(\psi^*_\lambda(\mathbf{X}), \psi^*_\lambda(\mathbf{X}_i))\right)}. \tag{13}$$

4. Asymptotic properties of the functional covariates and multivariate response

Suppose that there is a sample $\{(\psi^*_\lambda(X_1), \psi^*_\lambda(Y_1)),\ldots,(\psi^*_\lambda(X_T), \psi^*_\lambda(Y_T))\}$ where $(\psi^*_\lambda(Y_t)$ is a random variable taking its values in a semi-metric space (E, d) of infinite dimension for each t ∈ {1,...,T} and $(\psi^*_\lambda(Y_t) \in \Re^d$, , is the response from the nonparametric regression Eq. (11). Let us now discuss in detail the theoretical framework that allows us to prove the asymptotic results in our research. The error process $\{\epsilon_t\}$ is, as expected independent of the process $\{\psi^*_\lambda(X_t)\}$ and $E[\epsilon_t \mid \chi = \psi^*_\lambda(X_t)] = 0$. Considering that the processes of $(\psi^*_\lambda(X_t), \psi^*_\lambda(Y_t))$ are α - the most generic example of weakly dependent variables. Define F_a^b be as the σ-algebra of events created by the random variables $\{(\psi^*_\lambda(X_t), \psi^*_\lambda(Y_t))\}_{t=a}^b$ and set

$$\sup_{A \in F_{-\infty}^0, B \in F_k^\infty} |\mathrm{pr}\,(A \cap B) - pr\,(A)\,pr(B)| = \alpha(k) \xrightarrow[k\to\infty]{} 0 \tag{14}$$

Let $|\cdot|$ denote the L_1 -norm when applied to a vector; $|y| = \sum_{j=1}^d |y_j|$, $y = (y_1, \cdots, y_d)^T$ when applied to a matrix; and the normal matrix norm when applied to a matrix. Our assumptions are as follows: Let x be a given point in f, and denote by $B(x, h)$ the closed ball of centre x and radius h, namely:

$$B(x, h) = \{x' \in f : d(x, x') \leqslant h\}.$$

The model requires that the probability of $\mathcal{X}$ is such that there exists a non-decreasing function ϕ_x such that:

(H1) $\exists(C_1, C_2), \forall x \in f, \forall\varepsilon > 0,$

$$0 < C_1\phi_x(\varepsilon) \leqslant P(\mathcal{X} \in B(x, \varepsilon)) \leqslant C_2\phi_x(\varepsilon) < \infty.$$

And the joint distribution of $(\mathcal{X}, Y)$ needs to satisfy:

(H2) $\exists C_3, \forall r > 1, E(\|Y\|^r|\mathcal{X}) < C_3 r! < \infty.$

(H3) $\exists C_4, \exists b > 0, \exists\gamma > 0, \forall x, x' \in f,$

$$\|r(x) - r(x')\| \leqslant C_4 d^b(x, x').$$

In addition, we also need the following technical conditions on the kernel function and the bandwidth.

(H4) The kernel function has to be such that:

(i) K is a bounded and Lipschitz continuous function with support $[0, 1)$, and if $K(1) = 0$ it has to fulfill, together with $\phi_x(\cdot)$, the conditions:

(ii) $\exists(C_5, C_6) > 0$, such that $-\infty < C_5 \leqslant K'_{iH} \leqslant C_6 < 0$.

(iii) $\exists C_7 > 0, \exists\gamma_0 > 0, \forall\gamma < \gamma_0$,

$$\int_0^\gamma \phi_x(u)du > C_7\gamma\phi_x(\gamma).$$

(H5) The bandwidth h is a positive sequence such that:

$$\lim_{n\to\infty} h = 0 \quad and \quad \lim_{n\to\infty} \frac{\log n}{n\phi_x(h)} = 0.$$

It is noted that this set of assumptions are different from the case of functional covariate with univariate response and when both explanatory and response variables are functional, because of the different bandwidths for the different components of the response variable.

5. Numerical examples

The methodology was applied to two examples and the data was analyzed using a R software. As clarified, the objective is to develop a methodology for measuring the effectiveness of the two transformation models, Box-Cox Transformations and Yeo-Johnson Transformations, when applied to two time series datasets, for example, multivariate nonparametric investigation of various reactions contrasted with their analysis as a univariate series. Furthermore, mean squared error is used for comparison purposes, $MSE(\mathbf{X}_n)$=$(1/s)$ $\sum_{j=1}^{s}$ $(\hat{Z}_j - Z_j)^2$, where, $\hat{Z}_j$ is the estimated value of the j-th value and Z_j is the real value of the j-th value

5.1. Real data example

In figure (1)., the principal model reflects the monthly average of Temperature (T), Rainfall (R), and Humidity (H) of Ninavah, Iraq, from 1976 to 2000. It discovered that the three-time series is not stationary, as evidenced by the autocorrelation functions values outside of the confidence intervals In figure (2)..

Because the YJT was used on the stationary time series, t was set to equal 8. According to the application methodology and requirements of the R program26, where $\psi^*_\lambda(\mathbf{Z})^{(i)}_{(24*8)}, i = 1, 2, 3.$ represent the (24x8) matrix of (i)th response data, such that the first raw contains the first eight observations of the response i, the second row contains the second eight observations, and so on,

$$\psi^*_\lambda(\mathbf{Z})^{(i)}_{(24*8)} = \begin{pmatrix} \psi^*_\lambda(\mathbf{Z})^{(i)}_{(11)} & \psi^*_\lambda(\mathbf{Z})^{(i)}_{(12)} & \cdots & \psi^*_\lambda(\mathbf{Z})^{(i)}_{(18)} \\ \psi^*_\lambda(\mathbf{Z})^{(i)}_{(21)} & \psi^*_\lambda(\mathbf{Z})^{(i)}_{(22)} & \cdots & \psi^*_\lambda(\mathbf{Z})^{(i)}_{(28)} \\ \vdots & \vdots & \vdots & \vdots \\ \psi^*_\lambda(\mathbf{Z})^{(i)}_{(24)2} & \psi^*_\lambda(\mathbf{Z})^{(i)}_{(24)2} & \cdots & \psi^*_\lambda(\mathbf{Z})^{(i)}_{(24)2} \end{pmatrix} \tag{15}$$

. The data have to put into a new matrix size 24 × 24 in the multivariate case illustrated in Table (1).

Here, we will use the 24th year and predict the 25th based on the 24 prior ones [24]. The horizon of prediction that is denoted by fixed s will be reorganized into two parts. The first is devoted to learning, which contains 23 curves, and the second part represents the testing contains lest curve (24). The functional explanatory sample $\psi^*_\lambda(\mathbf{X}_i)$,$i = 1, \cdots, 23$ will be loaded in the 23× 24 matrix presented in (2):

And a response real sample $\psi^*_\lambda(\mathbf{y}_i)$,$i = 1, \cdots, 23$, that will be loaded into the following 23-dimensional vector

The value $\psi^*_\lambda(\hat{Z})_{576+s}$ is predicted for fixed horizon s.For every estimation of s $\{1, \cdots, 24\}$,the predictions have been fulfilled. It should be noted that in our system, a few parameters must be chosen. This is for the kernel.

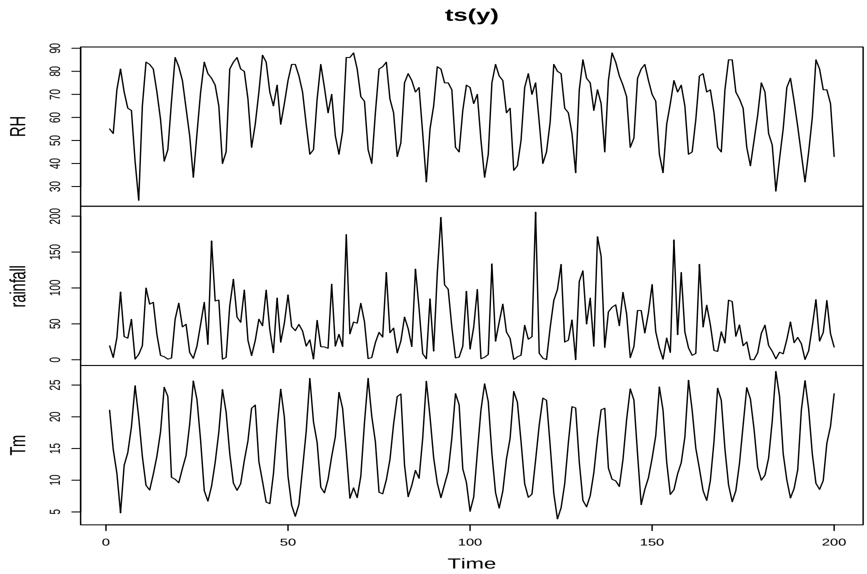


Figure 1. represents the time series of monthly average of H, R, T of Ninavah city in Iraq for the period (1976 – 2000).

Table 1. Table of organization the data matrix in the first example uploaded to the R program

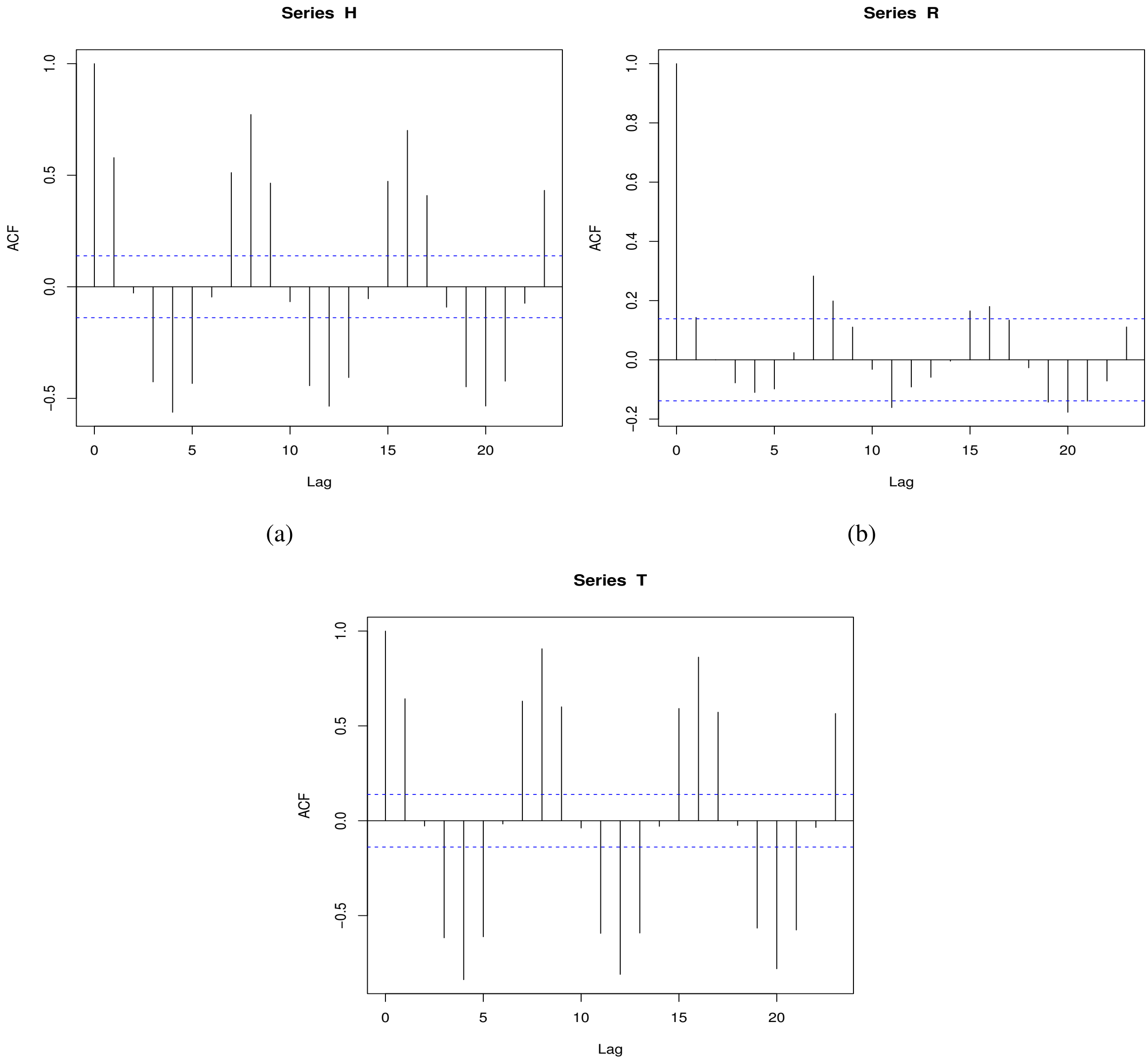
	Col 1	$\cdots$	Col j	$\cdots$	col 24
Row 1	$\psi^*_\lambda(\mathbf{Z})_{(i)}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{(j)}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{(24*8)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Row i	$\psi^*_\lambda(\mathbf{Z})_{1+24(i-1)}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{j+24(i-1)}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{(24i)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Row 24	$\psi^*_\lambda(\mathbf{Z})_{553}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{552+j}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{576}$

Table 2. The horizon of prediction learning, which contains 23 curves

$\psi^*_\lambda(\mathbf{Z})_1$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_j$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{24}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi^*_\lambda(\mathbf{Z})_{1+24(i-1)}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{j+24(i-1)}$	$\cdots$	$\cdots\psi^*_\lambda(\mathbf{Z})_{1+24i}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi^*_\lambda(\mathbf{Z})_{529}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{528+j}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{552}$

$\psi^*_\lambda(\mathbf{Z})_{24+s}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{24i+s}$	$\cdots$	$\psi^*_\lambda(\mathbf{Z})_{552+s}$

A semi-metric is utilized for (local) smoothing parameter selection based on the first functional main components of the data curves, conveyed as far as k-nearest neighbors, see Chapter 7 in[19]. The proposed algorithm was implemented in the R package funopare.knn.gcv, which is accessible at (http://www.sp.ups-tlse.fr/staph/npfda for the univariate -R model). In applying both analyzes, multivariate and univariate nonparametric. The MSE values are shown in Tables (3)to (6). The plots of MSE values of the predicted and original values resulting from the use of multivariate YJT and univariate YJT analysis for three series of real data are shown in figure (3) and (4) respectively. The acquired outputs for the multivariate time series datasets displayed in Table (3) after applying the



(a) (b)

(c)

Figure 2. The ACF plots of the three-time series: (a) H, (b) R and (c) T.

Box-Cox Transformation method to the three-time series data set using the four-step technique described. Table(4) compares MSE between original multivariate data and optimal univariate response parameters of the Box-Cox Transformation.

Table 3. Optimal Box-Cox transformation model parameters and MSE estimations of the functional variable's last curve for multivariate datasets

Time series	H		R		T	
Power parameter	$\lambda = 1$	$\lambda = -1.8$	$\lambda^* = 1$	$\lambda = -0.4$	$\lambda = 1$	$\lambda = -0.6$
$MSE((\mathbf{X}_n)$ of the original and transferred time series	36.8322	11.8353	577.622	483.7307	1.7013	1.5349
$MSE((\mathbf{X}_n)$ after stationaries the original and transferred time series	22.4901	19.8842	457.1106	207.5546	1.4582	1.1711

Table 4. Optimal Box-Cox transformation model parameters and MSE estimations of the functional variable's last curve for univariate datasets.

Time series	H		R		T	
Power parameter	$\lambda = 1$	$\lambda = -1.8$	$\lambda^* = 1$	$\lambda = -0.4$	$\lambda = 1$	$\lambda = -0.6$
$MSE((\mathbf{X}_n)$ of the original and transferred time series	19.7141	14.6442	553.1709	400.638	2.0382	2.4684
$MSE((\mathbf{X}_n)$ after stationaries the original and transferred time series	18.3779	15.0971	559.1246	313.1743	1.7616	1.2831

Table 5. For the three multivariate time series datasets, best power parameters of the two-transformation methods and MSE estimates of variable X_n .

Time series	H		R		T	
Transformation models	BCT	YJT	BCT	YJT	BCT	YJT
Power parameter	$\lambda = -1.8$	$\lambda = 1.6$	$\lambda^* = 0.4$	$\lambda = -2.7$	$\lambda = 0.6$	$\lambda = -0.6$
$MSE((\mathbf{X}_n)$	19.8842	14.5759	207.5545	214.7288	1.1711	0.8847

Table 6. Optimal parameters of univariate YJT-transformation models and $MSE(X_n)$ estimations for three time series datasets.

Time series	H	, R	T
Power parameter	$\lambda = 1.6$	$\lambda = -2.7$	$\lambda = -0.6$
$MSE((\mathbf{X}_n)$ of the multivariate transferred time series	14.5759	214.7288	0.8847
$MSE((\mathbf{X}_n)$ of the univariate transferred time series	22.781	376.654	1.141

As expected, the results in Table (3) show that the mean square error has dropped when using the Box-Cox Transformation compared to its value coming from the original data analysis when $\lambda = 1$. Table (4) shows that MSE for multivariate series in optimal parameter is smaller than MSE for optimal univariate for all series. When the YJT model was utilized in accordance with the identical four-step algorithm, these perplexing results were eliminated. In the attempt to stationarize the actual and converted series using differences, the means square error estimation increased in the actual time series while decreasing in the Yeo Johnson Transformation series. When applied using the identical four-step procedure, the Yeo Johnson Transformation multivariate model generated more accurate predictions with less error than the Box-Cox Transformation multivariate [25], as shown in Table (5). The plots of the actual and predicted values for the most recent curve (25th year) after smoothing the data using the Yeo Johnson Transformation multivariate method for the three-time series are shown in Figure 3. It can be observed that the Yeo Johnson Transformation multivariate functional responses has the smallest mean square error (MSE) than the YJT univariate functional responses model in figure(4) . It indicates that the YJT multivariate has better prediction accuracy. These testing results show that in Table (6),

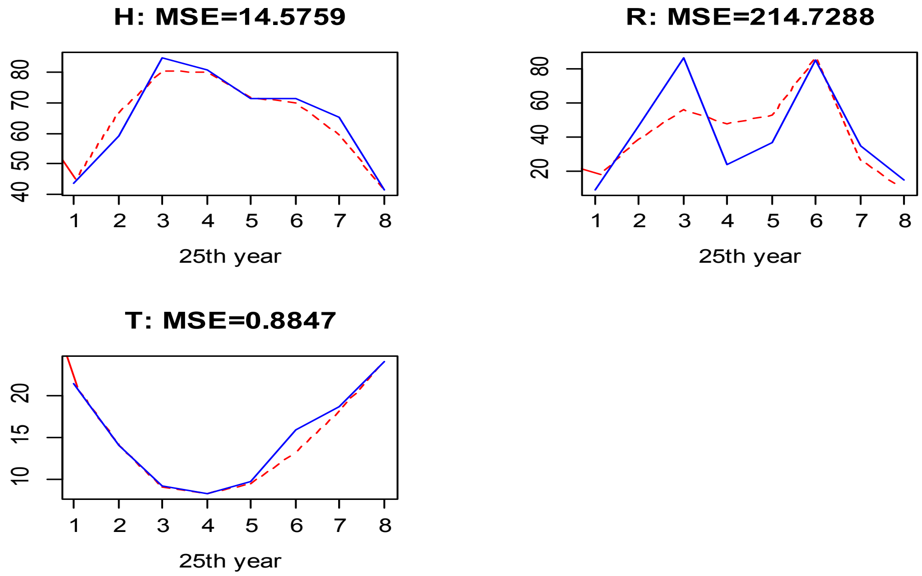


Figure 3. The of the predicted (red) and original (blue) values resulting from the use of YJT univariate analysis for three series of real data.

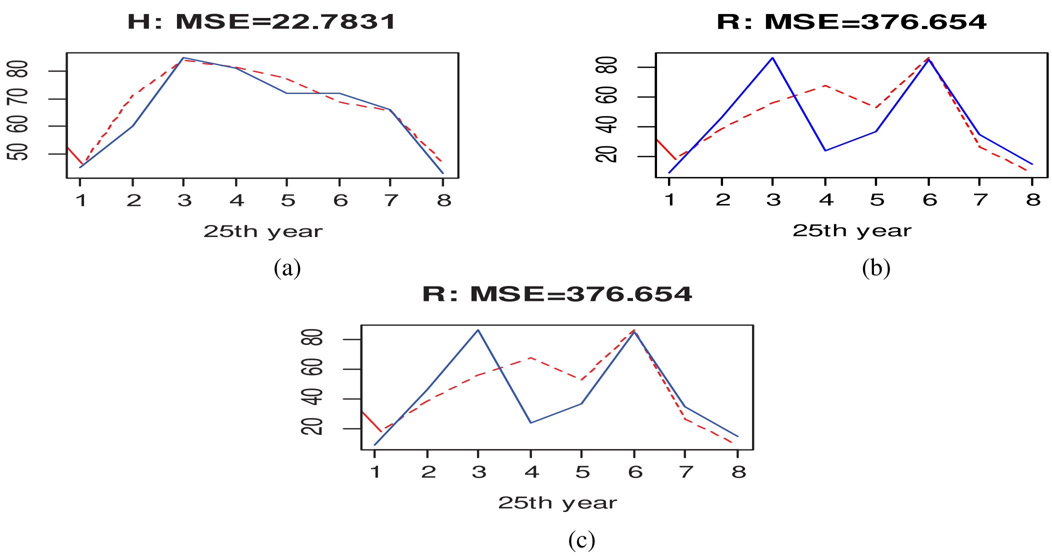


Figure 4. The of the predicted (red) and original (blue) values resulting from the use of YJT univariate analysis for three series of real data.

5.2. Some authorized models

In this section, some hypothetical nonlinear multivariate time series models are selected to apply the proposed power transformation methodologies. The R program was used to generate these models:

$$Y_{t,1} = X_t + X_{t-1}\exp(-X_{t-1}^2) + e_{t_1} \tag{16}$$

$$Y_{t,2} = X_t + \frac{4\exp(X_{t-1})}{1+\exp(X_{t-1}}$$

Under the assumption that

$X_0 = Y_0 = e_0 = 0$
The internal variable was generated as:
$X_t = 0.3X_{(}t-1) - a_t$ where $a_t \sim N(0,1)$
and random errors in the two models as:

$$e_{(t_2} = 0.18e_{(}t-1) + \varepsilon_t + 0.2\varepsilon_{(}t-1)$$
$$e_{(t_2} = 0.58e_{(}t-1)\exp(-e^2_{t-1}) + \varepsilon_t$$

$\varepsilon_t \sim N(0, 0.5)$
Figure (5) shows the generation of the $Y_{t,1}$ and $Y_{t,2}$ series with $\rho = 0.916$ and covariate =1.983. Figure (6) demonstrates the values of the autocorrelation functions (ACF). The obtained results are shown in Table (7), which represents the results of the analysis of the functional nonparametric multivariate time series analysis of the original generated data.As for Table (7), it includes the estimates of MSE and power parameters after data transformation and in figure (7,8), it represents MSE.

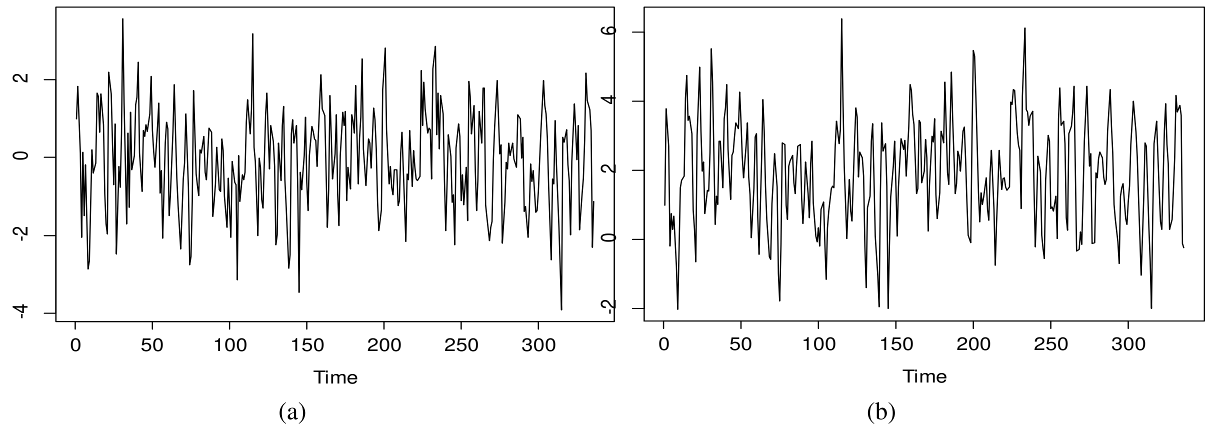


Figure 5. (a) represents the time series of $Y_{(}t,1)$ and (b) represents the time series of $Y_{(}t,2)$.

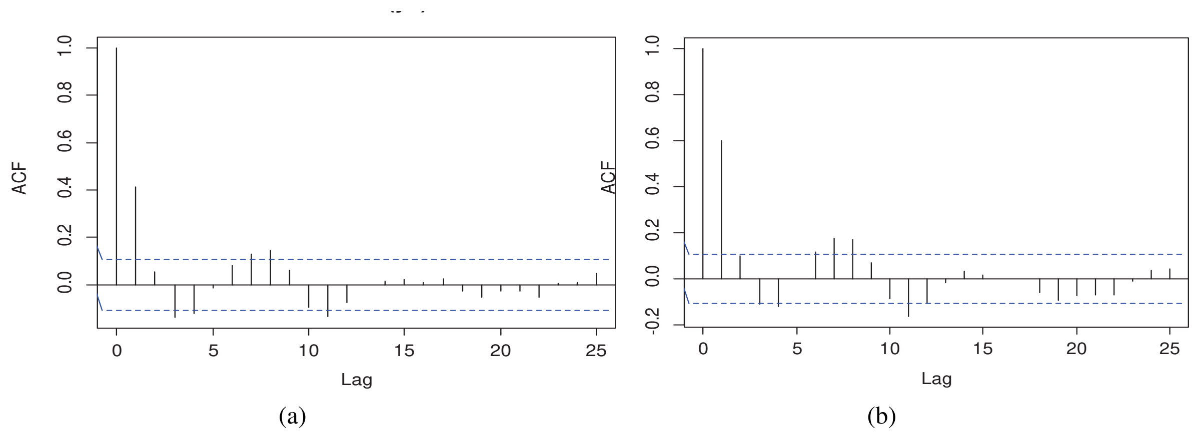


Figure 6. The ACF plots of the two-time series (a) of $Y_{(}t,1)$; (b) of $Y_{(}t,2)$.

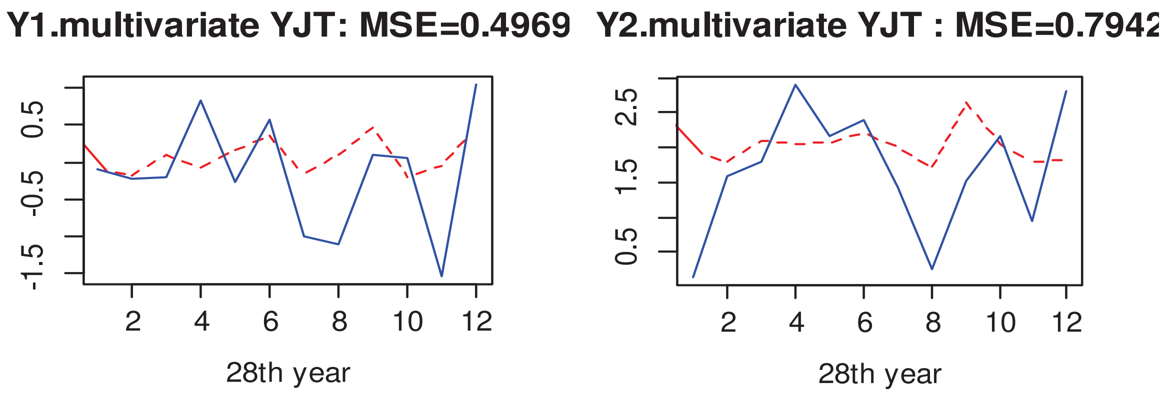
Table 7. MSE estimates of the last curve of functional variable for the multivariate time series and univariate time series for hypothetical models.

Time series	$Y_(t, 1)$	,	$Y_(t, 2)$
$MSE(X_n)$ of the multivariate original time series	0.59277		0.8608
$MSE(X_n)$ of the univariate original time series	0.8099		0.986

Table 8. Optimal parameters of the multivariate and in univariate data YJT -transformation models and MSE estimates of the last curve of functional variable X_n for the two hypothetical models n time series datasets

Time series	$Y_t, 1$	,	$Y_{t,2}$
Power parameter	$\lambda^* = 1.8$		$\lambda^* = 0.4$
$MSE((\mathbf{X}_n)$ of the multivariate transformed time series	0.4969		0.7942
$MSE((\mathbf{X}_n)$ of the univariate transformed time series	0.7526		1.2817

It can be observed that the YJT multivariate functional responses have the smallest mean square error (MSE) than the YJT univariate functional responses model in table (8)



(a) (b)

Figure 7. The plot of the predicted (red) YJT -transformation and actual values (blue) resulting from the use of multivariate analysis for two series of hypothetical models (a) of $Y_(t, 1)$; (b) of $Y_(t, 2)$.

6. Conculusion

This paper contributes to developing new methods for constructing nonparametric estimator of functional regression when the covariate is functional, and the response is multivariate functional. Two methods are considered: one multivariate functional responses is a direct extension of the method for a univariate functional response to multivariate functional responses the other is a new method where the transformation among different functional responses is taken into account. The numerical examples show that the methodology is used to demonstrate the ability of the multivariate YJT nonparametric time series using the K-nearest neighbor method to achieve good fitting compared to the univariate analysis. Reorganizing vectors in data have displaied matrices in a way that improves the ability of multivariate YJT analysis to provide accurate forecasts. It is crucial to notice that the optimum power parameters $\lambda^\star$ of both transformation models are significantly different, despite the fact that YJT represents an enlarged version of the Box-Cox Transformation method. The writers conclude that this

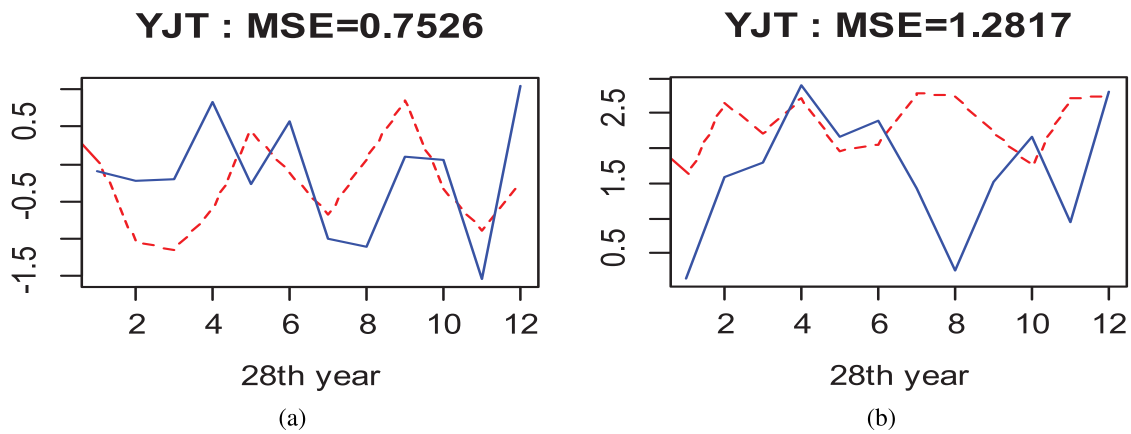


Figure 8. The plot of the predicted (red) YJT -transformation and actual values (blue) resulting from the use of univariate analysis for two series of hypothetical models (a) of $Y_(t, 1)$; (b) of $Y_(t, 2)$.

difference, as well as the amount of displacement in the actual data generated by both methods, was caused by the use of a nonparametric estimation method as an alternative to the parametric method for the hypothesis of normality of transformed response, as well as differences in the level of homogeneity between stationary and non-stationary data. The application methodology presented in this paper illustrates that YJT is a viable alternative to BCT for improving the non-parametric estimation of functional time series dataset. Furthermore, nonparametric estimate of power parameters that is not constrained by the requirements of the probability distribution provides researchers with a variety of choices for ensuring forecast accuracy.

Availability of Data

The datasets that support the paper's results are included in the paper.

Author Contributions

Both authors contributed to and approved the final version of the work.

Funding

There was no outside funding for this study.

Conflicts of Interest

According to the writers, they have no conflicts of interest.

REFERENCES

1. Härdle, W. and Vieu, P., *Kernel regression smoothing of time series*, Journal of Time Series Analysis, 13(3), pp.209-232. 1992.
2. Masry, E, . *Nonparametric regression estimation for dependent functional data*, Stoch Process Their Appl.; 115(1),155–177,2005 Jan.
3. Gannoun A, Saracco J, Yu K .*Nonparametric prediction by conditional median and quantiles*, J Stat Plan Infer; 117 (2), 207-223, 2003 Dec.
4. Aneiros-Perez G, Vieu P .*Nonparametric time series prediction: A semi-functional partial linear modeling.*, J Multivar Anal. 99(5), 834 – 857,2008.
5. Shang HL *Forecasting intraday S&P 500 index returns: A functional time series approach*, J. Forecast;36(7). 741–755,2017 Nov.
6. Shang H, Xu R. *Functional time series forecasting of extreme values, Commun. Stat.: Case Stud. Data Anal*, Appl.; 7(2), 182-199,2021 Jan.
7. Ali, H.T.M. and Othman, S.A., *A New Approach of Power Transformations in Functional Non-Parametric Temperature Time Series*, In Time Series Analysis-New Insights. IntechOpen,2022
8. Chavez-Demoulin, V. & Davison, A. C. . *Modelling Time Series Extremes. REVSTAT–Statistical*, Journal; 10(1): 109–133.2012 March.

9. Häardie W, Müller M . *Multivariate and Semiparametric Kernel Regression. Smoothing and Regression: Approaches, Computation, and Application*, Journal; 24:357-91.2000 Jul.
10. Ferraty F, Vieu P. *Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination.*, Journal; Nonparametr Stat;16(1-2):111-25.2004 Feb 1.
11. Box GE, Cox DR *An analysis of transformations.*, J R STAT SOC B ;26(2):211-43.1964 Jul.
12. He Y, Zheng Y *Short-term power load probability density forecasting based on Yeo-Johnson transformation quantile regression and Gaussian kernel function*, Energy. 1(154):143-56.2018 Jul.
13. Yeo IK, Johnson RA *A new family of power transformations to improve normality or symmetry*, Biometrika 1;87(4):954-959.2000 Dec.
14. Atkinson AC, Riani M, Corbellini A *The Box–Cox Transformation: Review and Extensions*, Stat Sci;36(2):239-55.2021 Apr.
15. Proietti T, Lütkepohl H *Does the Box–Cox transformation help in forecasting macroeconomic time series?, Int*, J. Forecast, 1;29(1):88-99.2013 Jan.
16. He Y, Zheng Y *Short-term power load probability density forecasting based on Yeo-Johnson transformation quantile regression and Gaussian kernel function*, Energy. 1(154):143-56.2018 Jul.
17. Chen CW, Lee JC *On Selecting a Power Transformation in Time-Series Analysis*, J Forecast;16(5):343-54.1997 Sep.
18. Xiao, Z. , Linton, O. , Carroll, R. , Mammen, E *More efficient local polynomial estimation in nonparametric regression with autocorrelated errors*, J. Am. Stat. Assoc. 1;98(464):980-992,2003 Dec.
19. Ferraty F, Vieu P *Nonparametric functional data analysis: theory and practice*, SSBM; 2006 Nov 22.
20. Pek J, Wong O, Wong AC. *Data transformations for inference with linear regression: Clarifications and recommendations*, Pract. Assess. Res. Evaluation;22(1):9,2017.
21. Ruckstuhl AF, Welsh AH, Carroll RJ. *Nonparametric function estimation of the relationship between two repeatedly measured variables*, Statistica Sinica.1:51-71,2000 Jan.
22. Othman, S. A. and Mohammed Ali, H. T. *Improvement of the Nonparametric Estimation of Functional Stationary Time Series Using Yeo-Johnson Transformation with Application to Temperature Curves*, Adv. Math. Phys,30;2021:1-6,2021 Jan.
23. Raymaekers J, Rousseeuw PJ. *Transforming variables to central normality.*, Mach. Learn, 21:1-23,2021 Mar.
24. Youngman BD. *evgam: An R package for generalized additive extreme value models.*, arXiv preprint arXiv:2003.04067.URL: https://arxiv.org/abs/2003.04067, 2020 Mar 9.
25. R Core Team R. *A Language and Environment for Statistical Computing, R Foundation,Vienna,*, Austria; URL: https://www.R-project.org/,2020.