Research Article

Improvement of the Nonparametric Estimation of Functional Stationary Time Series Using Yeo-Johnson Transformation with Application to Temperature Curves

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

[1]*Department of Mathematics, College of Basic Education, University of Duhok, Kurdistan Region, Iraq*
[2]*Department of Economics, College of Administration and Economics, Nawroz University, Kurdistan Region, Iraq*

Correspondence should be addressed to Sameera Abdulsalam Othman; sameera.othman@uod.ac

In this article, Box-Cox and Yeo-Johnson transformation models are applied to two time series datasets of monthly temperature averages to improve the forecast ability. An application algorithm was proposed to transform the positive original responses using the first model and the stationary responses using the second model to improve the nonparametric estimation of the functional time series. The Box-Cox model contributed to improving the results of the nonparametric estimation of the original data, but the results become somewhat confusing after attempting to make the transformed response variable stationary in the mean, while the functional time series predictions were more accurate using the transformed stationary datasets using the Yeo-Johnson model.

1. Introduction

Forecasting the future is the main function of time series analysis. Proceeding from this idea, researchers have developed several techniques that are concerned with the improvement of accuracy of forecasts by treating the time series as a stochastic process. A functional data analytic approach or so-called a stochastic forecast [1] allows dealing with the observations as a function [2] freely outside of the conditions of parametric and fully nonparametric modeling. This handling of observations in time series data makes it sequential and can be separated into successive time periods [3]. Thus, the dimensions of the time series are reduced with a limited loss of information [4] and represent the data in a linear combination of a few functions (carefully selected) instead of treating the data in its original form as a single vector of values [2], that is, processing and transforming the structure of time series data in line with the structure of regression models. Shang in 2019 showed that with the time dependence of observations in some datasets, the principal

component method may lead to erroneous estimates. Therefore, the two authors believe that this problem may be exacerbated in some time series data, especially those that are characterized by the presence of seasonal changes. However, it has become known in practical applications of time series that they are rarely stationary and that seasonal changes, trend, and dependence on external factors have become the rules, not the exception [5]. For this reason, it can be said that the data transformation has become a part of the traditional parametric and nonparametric analysis of complex time series.

In this article, the two authors have used the Yeo-Johnson transformations to improve the nonparametric estimation of the functional time series. The use of both approaches, transformation, and functional analysis without considering the modeling conditions is an attempt to focus the analyzing goal and the efficiency criterion in the context of forecast ability.

The rest of the article is organized as follows. The Box-Cox and Yeo-Johnson transformations are presented in the next section. The third section contains the formulation of

the problem and the proposed application methodology. The practical examples are included in the fourth section, while the fifth section contained some conclusions.

2. Box-Cox and Yeo-Johnson Transformations

Box and Cox [6] suggested the Box-Cox transformation (BCT) methodology in regression models to reduce anomalies in data, reduce nonlinearity, and achieve normality random errors. The methodology assumes, for any response variable $Z > 0$ and $\lambda \in R$, the transformed variable $\Psi(Z) = (Z^\lambda - 1)/\lambda$ when $\lambda \neq 0$ and $\Psi(Z) = \mathrm{Ln} Z$ when $\lambda = 0$. And when λ is equal to 1, the data is analyzed in its original scale, whereas the case $\lambda = 0$ corresponds to the natural logarithmic transformation of the data. BCT is based on the assumption of the transformed response normality and then defining the probability density function of the original response as a "backward transformed" of change of variables technique.

Yeo and Johnson [7] generalized the BCT to include negative and positive values in datasets [7]. They used a smoothness condition to combine the transformations for positive and negative observations, thus obtaining a one-parameter transformation family [8]. For $Z \in R$, the YJT is given by

$$\psi(Z) = \begin{cases} \dfrac{(Z+1)^\lambda - 1}{\lambda}, & \lambda \neq 0 \text{ and } Z \geq 0, \\[2mm] \mathrm{Ln}(Z+1), & \lambda = 0 \text{ and } Z \geq 0, \\[2mm] \dfrac{-\left((-Z+1)^{2-\lambda} - 1\right)}{2-\lambda}, & \lambda \neq 2 \text{ and } Z < 0, \\[2mm] \mathrm{Ln}(-Z+1), & \lambda = 2 \text{ and } Z < 0. \end{cases} \tag{1}$$

This transformation is appropriate for correcting both left and right skew when $\lambda > 1$ and $\lambda < 1$, respectively, while the linear relationship is achieved when $\lambda = 1$ [9]. Also, Yeo-Johnson transformations (YJT) can hold the properties of the log-mean standardization after the inverse transformation since $\Psi(Z)$ is invertible [10].

In 1970, Box and Jenkins recommended for the first time the use of power transformation in ARIMA models [11]. After that date, many authors took up this topic and made numerous proposals in many mathematical and applied aspects of the time series. Also, some of them indicated some failures in practical cases, for example, the success limitation of the normality assumption of the transformed data, and that it could lead to noticeable improvements in the simplicity of the data models and the accuracy of the estimate [12] especially in the models with skewness for variables [10]. Cook and Olive, [13] and Atkinson [8] also point out that the estimation of the transformation parameter can be particularly sensitive to outliers. And in some practical cases of time series, the BCT may not lead to an improvement in forecasting performance [11], or as Chen and Lee [14] say, it does not consistently produce superior forecasts.

Some problems in practical applications occur for two reasons: the first is the difficulty in obtaining an optimum value of the transformation parameter, so that at the same time, the conditions of the fitting of assumed distribution of the transformed data are met, and the model errors are minimal, while the second is that the transformations lead to a change in the nature of the relationships between the variables of the model, which may lead to a lack of balance between the efficiency of statistical inference and the ability to interpret the sizes of the variables' influence [15].

3. Formulation of the Problem

Let us consider a univariate time series $\{Z_t, t \in R\}$, by redividing the time series sample into $(p - 1)$ statistical samples of size $(n = N - s - p + 1)$. This division allows the time series to be redefined as functional data $\{(\mathbf{X}_i, Y_i)\}_{i=1,..,n}$ in such the variation trends between times of the series are diagnosed through the functional analysis tools [1]. Thus, the relationship can be described as a standard regression model.

$$Y = m(\mathbf{X}) + \varepsilon, \tag{2}$$

where $m(\mathbf{X})$ is the smooth functional data, ε is a sequence of independent identically distributed function white noise sequence in such $E(\varepsilon/\mathbf{X}) = 0$. $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$ are identically distributed as the functional random variable $\mathbf{X}_i = (Z_{i-p+1}, \cdots, Z_i)$ and $Y_i = Z_{i+s}, i = p, \cdots, N - s$ as a scalar response. In order to characterize the relationship, the response Y, given the functional variable $\mathbf{X}$, assumes that $N = n\tau$ for some $n \in N$ and some $\tau > 0$. And then, we get a statistical sample of curves $\mathbf{X}_i = \{Z(t), (i-1)\tau < t \leq i\tau\}$ of size $(n-1)$ and the response $Y_i = Z(i\tau + s), i = 1, \cdots, n - 1$ [16, 17]. The usual nonparametric estimation of the functional relation has several advantages and can be very well adapted to local features of time series data [18] and robustness to functional form misspecification [19]. The kernel regression estimator is evaluated at a given function $m(\mathbf{X})$ by

$$\widehat{m}(\mathbf{X}) = \frac{\sum_{i=1}^n Y_i K\left(h^{-1} d\left(X, \mathbf{X}_i\right)\right)}{\sum_{i=1}^n K\left(h^{-1} d\left(X, \mathbf{X}_i\right)\right)}, \tag{3}$$

where K is a kernel function, h (depending on n) is a positive real bandwidth, and $d\left(X, \mathbf{X}_i\right)$ denotes any semimetric (index of proximity) between the observed curves. The authors suggest several ways to find equation (3) including kernel regression estimator, functional conditional quantiles, and conditional mode.

A number of useful explanatory methods can be used to measure the closeness (proximities) between the curves of the functional variables in a reduced dimensional space. Ferraty and Vieu [16] refer to at least three families of semimetrics to measure $d\left(X, \mathbf{X}_i\right)$, for example, the functional principal component analysis (FPCA) in which the proximity is measured by the square root of the quantity $\int \left(\mathbf{X}_i(t) - \mathbf{X}_j(t)\right)^2 dt$. Also, there is another measure which is based on the second derivative, where the proximity is measured by the square root of the quantity $\int \left(\mathbf{X}_i^{(2)}(t) - \mathbf{X}_j^{(2)}(t)\right)^2 dt$ (Dauxois et al. (1982), Castre et al.

(1986), Ferraty and Vieu [16], and Febrero-Bande and Oviedo de la Fuente (2012)).

Regarding the kernel estimator (3), Wand et al. [20] indicated that it is not working well when the data are asymmetric, as for the standard PCA which may not be the suitable technique to apply when the data distribution is skewed or there are outliers [21]. Therefore, power transformation is considered one of the important alternatives to improve the efficiency of nonparametric estimation of functional data (for more details, see [12, 22, 23]).

Most transformation approaches have a common analytical path, which is the choice of the power transformation model, and propose an algorithm for estimating the power parameters in parallel with the mechanisms of estimating the traditional parameters of the model. Also, there are two common directions of the power parameter's estimation; the first is the parametric direction in which the power parameters are estimated under the statistical modeling assumptions. The most important methodology of this direction is the Box-Cox transformation (BCT) to improve the efficiency of the multiple linear regression model under the normality assumption of transformed response [6]. Also, Wand et al. [20] used the same methodology of Box-Cox to improve the efficiency of density estimation under the assumption of some distributions of the transformed variable [20] (see also [24, 25, 26]).

The second direction is the nonparametric estimation of power parameters without any assumptions about the response and error distributions or what might be called the model-independent approach [27] (see also [28, 29]). In this direction, the power parameters can estimate according to some decision rules such as minimizing or maximizing some indicators of model efficiency.

4. Application Methodology

It is known that the power transformations are important for making the time series stationary in the variance, while the differencing is useful for making the time series stationary in the mean. Generally, none of these approaches can be a substitute for the other. However, sometimes power transformations can make the time series stationary. And because the BCT is used to transform the positive responses, it becomes important to use it to transform the original data as a first stage and then calculate the differences to achieve the stationary of the time series. And as a result, the variance stabilizing obtained from the power transformation will be affected by the differencing process. In this regard, Dittmann and Granger [30] indicate that for every nonstationary process, the polynomial transformations are also nonstationary and have a stochastic trend in mean and invariance. To overcome these problems, the authors believe that the use of YJT will be appropriate to improve forecastability, because it can be used to stabilize the variance in stationary time series. Also, the estimation of the power parameter according to a certain decision rule that we have referred to would be appropriate as long as the issue is related to the nonparametric functional analysis.

The application methodology includes estimating the smooth functional data $m(\mathbf{X})$ in the regression equation (2) according to the kernel estimator equation (3) after transforming the time series dataset. The BCT was applied to the original time series dataset, while the YJT was applied to the stationary time series dataset. So, the statistical sample of curves was redefined by the expression

$$\Psi_\lambda(\mathbf{X}_i) = \{\Psi_\lambda(Z(t)), (i-1)\tau < t \le i\tau\}, \tag{4}$$

and the response by the expression

$$\Psi_\lambda(Y_i) = \Psi_\lambda(Z(i\tau + s)), \tag{5}$$

where $i = 1, \cdots, n - 1$ and Ψ_λ represents a data transformation function by the power parameter λ.

For each transformation model, the decision rule adopted for selecting the optimal estimate of power parameter λ is that which corresponds to the lowest estimates of the mean squares of the forecasting errors of the last curve of functional variable according to the equation $\mathrm{MSE}(\mathbf{X}_n) = (1/s)\sum_{j=1}^{s} (Z_j^\wedge - Z_j)^2$, where $Z_j^\wedge$ and Z_j are the j-th estimated and real values in the last curve. As for $Z_j^\wedge$ values, they are computed from the inversions of BCT and YJT, or what we might call the retransformation from the transformed data metric to the original metric.

So, the application algorithm of BCT and YJT models and nonparametric estimation of the transformed functional time series were as follows:

(1) Fix τ to define expressions (4) and (5)

(2) Remove the seasonality patterns by taking the differences to make the time series stationary

(3) Fix $\lambda \,\epsilon\, \Lambda$, where $\Lambda = \{-3, 3\}$

(4) For each $\lambda \,\epsilon\, \Lambda$, BCT is used to transform the original time series $Z(t)$ and YJT is used to transform the stationary time series of k differences $\Delta^k Z(t)$ to get the two explanatory functional matrices $\Psi_\lambda(\mathbf{X}) = [\Psi_\lambda(Z)]_{n x \tau}$ and $\Psi_\lambda(\mathbf{X}) = [\Psi_\lambda(\Delta^k Z)]_{n x \tau}$, (for more details about the matrices fille organizing in R program, see [16, 31]).

(5) Evaluate the explanatory function estimation of the relationship $\Psi_\lambda(Y_i) = m(\Psi_\lambda(\mathbf{X})) + \varepsilon$ according to the following kernel estimator:

$$\widehat{m}(\Psi_\lambda(\mathbf{X})) = \frac{\sum_{i=1}^n \Psi_\lambda(Y_i) K\left(h^{-1} d\left(\Psi_\lambda(X), \Psi_\lambda(\mathbf{X}_i)\right)\right)}{\sum_{i=1}^n K\left(h^{-1} d\left(\Psi_\lambda(X), \Psi_\lambda(\mathbf{X}_i)\right)\right)}. \tag{6}$$

The optimal value λ^* of the power parameter λ is the one that minimizes the $\mathrm{MSE}(\mathbf{X}_n)$ of the last functional variable.
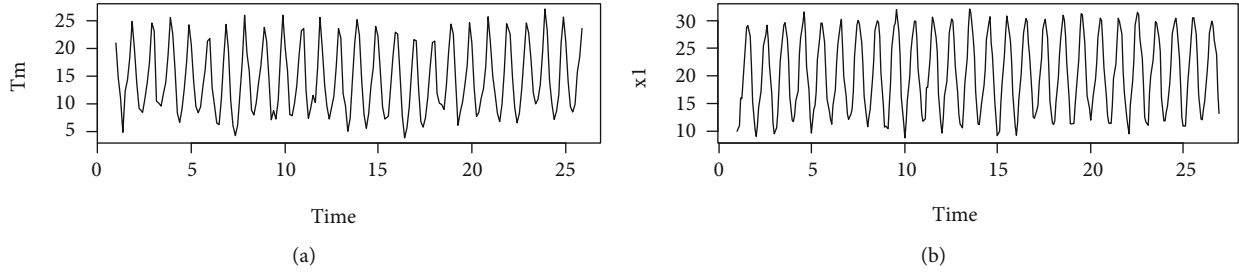
(a)

(b)

FIGURE 1: Plots of the monthly temperature averages series: (a) Nineveh City in Iraq for the period 1976 to 2000; (b) Tunisia (TSN) for the period 1991 to 2015.
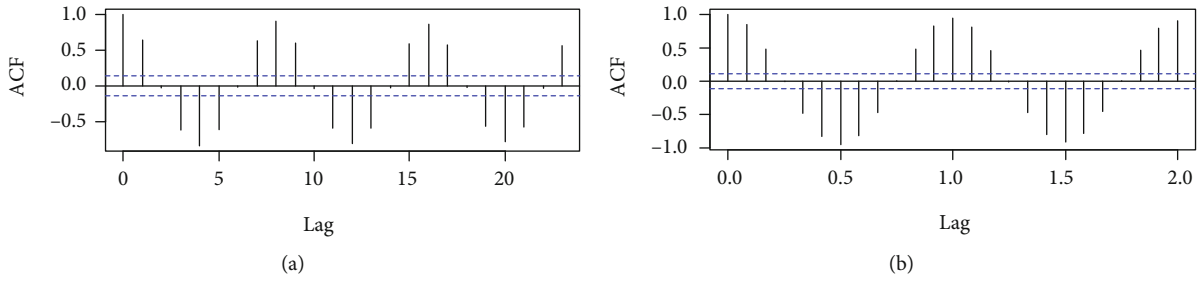

(a)

(b)

FIGURE 2: The ACF plots of the two time series: (a) TSN; (b) TST.

TABLE 1: Optimal parameters of the BCT model and MSE estimates of the last curve of functional variable for the two time series datasets.

Time series	TSN		TST	
Power parameters	$\lambda = 1$	$\lambda^* = 2.1$	$\lambda = 1$	$\lambda^* = 1.7$
MSE($\mathbf{X}_n$) of the original and transferred time series	2.0365	1.5620	0.5214	0.3402
MSE($\mathbf{X}_n$) after making the original and transferred time series stationary	1.7616	1.7494	0.4303	0.5118

TABLE 2: Optimal parameters of the two transformation models and MSE estimates of the last curve of functional variable $\mathbf{X}_n$ for the two time series datasets.

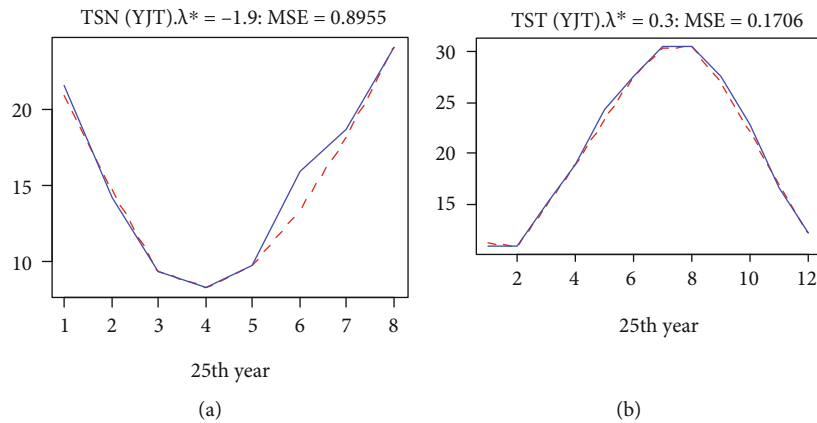| Time series	TSN				TST			
Transformation models	BCT		YJT		BCT		YJT	
	λ	λ^*	λ	λ^*	λ	λ^*	λ	λ^*
Power parameters	1	2.1	1	−1.9	1	1.7	1	0.3
MSE($\mathbf{X}_n$)	1.7616	1.7494	1.7616	0.8955	0.4303	0.5118	0.4303	0.1706

(a)

(b)

FIGURE 3: The plots of the original (blue) and predicted values (red) for the latest curve (25th year) of the inverse of the data transformed by the YJT model for the two time series: (a) TSN; (b) TST.

5. Applications

The two transformation models, BCT and YJT, were applied to two time series examples of monthly temperature averages, and an R program was used to analyze the data. The first time series has a size of 200 observations of Nineveh City in Iraq (TSN) for the eight rainy months in every year. We take the monthly average of the meteorological station of Nineveh for the period 1976 to 2000 (Figure 1(a)). The second has a size of 300 observations of Tunisia (TST) for all months of the period 1991 to 2015. The data can be found at https://climateknowledgeportal.worldbank.org (Figure 1(b)).

It was found that the two time series are not stationary, and this is clearly demonstrated by the values of the autocorrelation functions (ACF) outside the confidence levels in Figure 2.

By applying BCT model to the two time series according to the five-step algorithm suggested in Section 4, we obtained the results shown in Table 1.

As expected, it is evident from the results shown in Table 1 that the estimate of MSE has decreased when using BCT compared to its value resulting from the analysis of the original data when $\lambda = 1$. These confusing results were overcome when the YJT model was used according to the same five-step algorithm.

As for the attempt to make the original and transformed time series stationary by the first-order differences, the MSE estimation increased in the two transformed series and decreased in the original series. By applying the YJT model according to the same five-step algorithm, more accurate predictions of fewer errors were obtained compared to the error estimates obtained by using the BCT model (Table 2). Figure 3 shows the plots of the original and predicted values for the latest curve (25th year) after smoothing the data using the YJT model for the two time series.

6. Conclusions

It is important to note that the optimum power parameters λ^* for both transformation models are significantly different even though YJT represents the extended version of the BCT model. The authors believe that this difference and the amount of displacement in the original data generated by both models were due to the use of a nonparametric estimation method to choose the optimal power parameter as an alternative to the parametric method for the hypothesis of normality of transformed response, in addition to the differences in the level of homogeneity between stationary and nonstationary time series datasets.

The application methodology in this article demonstrates that YGT could be a successful alternative to BCT to improve the nonparametric estimation of the functional time series. Also, the nonparametric estimation of the power parameters not restricted by the conditions of the probability distribution provides the researcher with wide options to ensure the accuracy of the prediction.

Data Availability

The datasets supporting the conclusions of this article are included in the article.

Conflicts of Interest

The authors declare that they have no conflict of interest.

References

[1] H. L. Shang, P. W. F. Smith, J. Bijak, and A. Wisniowski, "A functional data analysis approach for forecasting population: a case study for the United Kingdom," in *Centre for Population Change*, Working paper, number 41, 2013.

[2] L. Kidzinski, "Functional time series," *arXiv: Methodology*, 2015, arXiv:1502.07113v1.

[3] A. Aue, D. D. Norinho, and S. Hörmann, "On the prediction of stationary functional time series," *Journal of the American Statistical Association*, vol. 110, no. 509, pp. 378–392, 2015.

[4] H. L. Shang, "Forecasting intraday S&P 500 index returns: a functional time series approach," *Journal of Forecasting*, vol. 36, no. 7, pp. 741–755, 2017.

[5] V. Chavez-Demoulin and A. C. Davison, "Modelling time series extremes," *Revstat – Statistical Journal*, vol. 10, no. 1, pp. 109–133, 2012.

[6] G. E. P. Box and D. R. Cox, "An analysis of transformations (with discussion)," *Journal of the Royal Statistical Society*, vol. 26, no. 2, pp. 211–252, 1964.

[7] I.-K. Yeo and R. A. Johnson, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, no. 4, pp. 954–959, 2000.

[8] A. C. Atkinson, "The Box-Cox transformation: review and extensions," *Statistical Science.*, pp. ISSN 0883–ISSN 4237, 2020, http://eprints.lse.ac.uk/103537/.

[9] B. Bean, X. Xu, and S. MacEachern, "Transformations and Bayesian density estimation," *Electronic Journal of Statistics*, vol. 10, no. 2, pp. 3355–3373, 2016.

[10] S. Soleymani, "Exact Box-Cox analysis," *A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree in Doctor of Philosophy in Statistics and Actuarial Sciences*, The University of Western Ontario, 2018.

[11] T. Proietti and H. Lutkepohl, "Does the Box-Cox transformation help in forecasting macroeconomic time series?," *International Journal of Forecasting*, vol. 29, no. 1, pp. 88–99, 2013.

[12] H. L. Shang, "Selection of the optimal Box-Cox transformation parameter for modelling and forecasting age-specific fertility," *Journal of Population Research*, vol. 32, no. 1, pp. 69–79, 2015.

[13] R. D. Cook and D. J. Olive, "A note on visualizing response transformations in regression," 2001, https://opensiuc.lib.siu.edu/cgi/viewcontent.cgi?article=1006&context=math_articles.

[14] C. W. S. Chen and J. C. Lee, "On selecting a power transformation in time-series analysis," *Journal of Forecasting*, vol. 16, no. 5, pp. 343–354, 1997.

[15] J. Pek, O. Wong, and A. C. Wong, "Data transformations for inference with linear regression: clarifications and recommendations," *practical assessment, research, and evaluation*, vol. 22, 2017.

[16] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis - Theory and Practice*, Springer Series in Statistics, 2006.

[17] G. Aneiros-perez, R. Cao, and J. M. Vilar-Fernandez, "Functional methods for time series prediction: a nonparametric approach," *Journal of Forecasting*, vol. 30, no. 4, pp. 377–392, 2011.

[18] J. M. Vilar-Fernandez and R. CAO, "Nonparametric forecasting in time series—a comparative study," *Communications in Statistics - Simulation and Computation*, vol. 36, no. 2, pp. 311–334, 2007.

[19] Z. Cai, Q. Li, and J. Y. Park, "Functional-coefficient models for nonstationary time series data," *Journal of Econometrics*, vol. 148, no. 2, pp. 101–113, 2009.

[20] M. P. Wand, J. S. Marron, and D. Ruppert, "Transformations in density estimation," *Journal of the American Statistical Association*, vol. 86, no. 414, pp. 343–353, 1991.

[21] M. Maadooliat, J. Z. Huang, and J. Hu, "Integrating data transformation in principal components analysis," *Journal of Computational and Graphical Statistics*, vol. 24, no. 1, pp. 84–103, 2015.

[22] H. L. Shang, P. W. F. Smith, J. Bijak, and A. Wiśniowski, "A multilevel functional data method for forecasting population, with an application to the United Kingdom," *International Journal of Forecasting*, vol. 32, no. 3, pp. 629–649, 2016.

[23] P. Legendre and D. Borcard, "Box–Cox-chord transformations for community composition data prior to beta diversity analysis," *Ecography*, vol. 41, no. 11, pp. 1820–1824, 2018.

[24] C. Bolance, M. Guillen, and J. P. Nilsen, "Kernel density estimation of actuarial loss functions," *Insurance: Mathematics and Economics*, vol. 32, no. 1, pp. 19–36, 2003.

[25] D. Pitt, M. Guillen, and C. Bolancé, "Estimation of parametric and nonparametric models for univariate claim severity distributions - an approach using R," *XREAP*, 2011, 2011. SSRN: https://ssrn.com/abstract=1856982.

[26] K. M. Sakthivel and C. S. Rajitha, "Kernel density estimation for claim size distributions using shifted power transformation," *International Journal of Science and Research (IJSR)*, vol. 5, no. 4, pp. 2025–2028, 2016.

[27] V. M. Guerrero and R. Perera, "Variance stabilizing power transformation for time series," *Journal of Modern Applied Statistical Methods*, vol. 3, no. 2, pp. 357–369, 2004.

[28] D. Nychka and D. Ruppert, "Nonparametric transformations for both sides of a regression model," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 57, no. 3, pp. 519–532, 1995.

[29] H. T. Alyousif, Department of Economics, C. Administration et al., "Develop a nonlinear model for the conditional expectation of the Bayesian probability distribution (gamma – gamma)," *Journal of Al-Nahrain University Science*, vol. 17, no. 2, pp. 205–212, 2014.

[30] I. Dittmann and C. W. J. Granger, "Properties of nonlinear transformations of fractionally integrated processes," *Journal of Econometrics*, vol. 110, no. 2, pp. 113–133, 2002.

[31] S. A. Othman and H. T. Mohammed Ali, "Studying of nonparametric multivariate time series analysis with applications methodology," *Journal of Xi'an University of Architecture & Technology*, vol. 12, no. 6, pp. 1500–1514, 2020.