STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
*Stat., Optim. Inf. Comput.*, Vol. **11**, June 2023, pp 570–579.
Published online in International Academic Press (www.IAPress.org)

# On the Use of the Power Transformation Models to Improve the Temperature 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] *College of Administration Economics, University of Zakho, Kurdistan Region, Iraq.Scientific Research and Development Center, Nawroz University, Kurdistan Region, Iraq*

**Abstract** The aim of this paper is to select an appropriate ARIMA model for the time series after transforming the original responses. Box-Cox and Yeo-Johnson power transformation models were used on the response variables of two time series datasets of average temperatures and then diagnosed and built the appropriate ARIMA models for each time-series. The authors treat the results of the model fitting as a package in an attempt to decide and choose the best model by diagnosing the effect of the data transformation on the response normality, significant of estimated model parameters, forecastability and the behavior of the residuals. The authors conclude that the Yeo-Johnson model was more flexible in smoothing the data and contributedto accessing a simple model with good forecastability.

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

**DOI:** 10.19139/soic-2310-5070-1333

## 1. Introduction

*M*ostly, there is no data set that can be ready for statistical modeling due to some violations of the conditions of statistical analyses. Therefore, the authors resort to one of three options to overcome the problems resulting from the lack of conditions, which is the use of robust procedures or data transformation in the context of parametric analyses or the use of non-parametric procedures in the context of free distribution analyses. In creating statistical models, the literature on transforming data from its original scale to a new scale aims to estimate a simple model with normal errors and constant error variance [1, 2]. This modification of the values and shape of a distribution function [3]. it also aims to improve model output when the violation of the conditions is not significant but effective. In 1964, Box and Cox proposed the Box-Cox transformation (BCT) and the approach of the normality assumption of the transformed response so as to boost multiple linear regression and removnon-linearity, heteroscedasticity and skewness [4] The subject of power transformations has been studied extensively with emphasis on approaches to employing estimation methods to choose decision rules that help in determining the estimators of the parameters of statistical models and the optimal power parameters[for more important aspects and recent approaches to the power transformation approaches, see Atkinson et al] [5]. This article is concerned with the use of power transformations to improve the forecastability of the time series models in light of the problem of the multiple and overlapping approaches estimating the power parameters. The two parametric power transformation models BCT and Yeo-Johnson Transformation (YJT) were used to transform the response variable in two real datasets of temperature time series. In the second section, we provide some theoretical reviews on the use of power transformations in time

*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)

series and some mathematical aspects of BCT and YJT models. As for the third section, it included the proposed computational algorithm. The fourth section included the practical application, while the fifth section included the conclusions.

## 2. Power Transformation in Time Series Models

Since 1970, Box and Jenkins recommended the use of power transformations in ARIMA models so that the modelling procedure involved the data transformation as a preliminary analysis and an iterative three-stage process; identification, estimation and checking the model [6]. Hipel and McLeod discussed the use of power transformation in seasonal and non-seasonal ARIMA model [7]. Chen and Lee discussed the selection of appropriate power transformation in ARMA models by using the Bayesian treatment to estimate the model and the power parameter [8]. There are two approaches to using power transformation to improve the forecastability of time series models. The first is to perform the transformation before fitting the model [6]. While the second is a complicated procedure because the process of estimating the power parameter is conducted in conjunction with the estimation of the model parameters and the other fitting processes of the model [9, 10, 11]. This complexity in the second approach results from the fact that the autocorrelation structure of the transformed time series and its variations, is not independent of the selection of the power parameter [12]. The simple power transformation model for positive random variable $Z$ is, $Z^{\langle\lambda\rangle} = Z^{\lambda}$ if $\lambda \neq 0$ and $Z^{\langle\lambda\rangle} = \log\langle Z\rangle$ if $\lambda = 0$ [2]. This transformation is monotone and it preserves the order of data for $\lambda > 0$. Thus, when $\lambda = 1$ the data is analyzed in its original scale, and $\lambda = 0$ corresponds to a logarithmic transformation. While the BCT model was more complex but it is more suitable for mathematical treatment. For any Y $> 0$ and $\lambda\in \mathbb{R}$ , the BCT is given by,

$$Z^{(\lambda)}(\lambda, Z) = \begin{cases} Z^{\lambda} - 1\backslash\lambda & \text{if } \lambda \neq 0 \\ \log(Z) & \text{if } \lambda = 0 \end{cases} \tag{1}$$

where $\lambda \in \mathbb{R}$ and usually estimated according to a hypothetical statistical model by maximum likelihood under the assumption of normality transformed response. The initial assumption of BCT restricted to positive data. Additionally, BCT restricts the sample space of the transformed variable by the two inequalities $Z^{\lambda} > -1\backslash\lambda$ if $\lambda > 0$ and $Z^{\lambda} > -1\backslash\lambda$ if $\lambda < 0$ so that it is not consistent with the normality assumption of the transformed variable [13]. Yeo and Johnson [14] generalized the BCT within the case of the negative random variable to achieve the advantage of working without having to fret about the domain of Z. The YJT for a fixed $Z^(\lambda) : R \rightarrow R$ is defined by,

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

Where $\lambda$ is the power parameter likewise the BCT. This transformation can hold the properties of the log-mean standardization after the inverse-transformation since $Z^{\langle\lambda\rangle}\ \langle\lambda, Z\rangle$ is invertible. The basic assumption in all the power transformation methodologies assumes, as Box and Cox first proposed in 1964, that the transformed data variable distributed according to the normal distribution with mean $\mu$ and variance $\sigma^2$. As a result, the original data will have an unknow probability density function of form $f_Z(Z^{\langle\lambda\rangle}; \lambda, \mu, \sigma^2)$, $J(\lambda, Z)$. As for the methods of estimating the power parameter $\lambda$ , one of the criteria is to choose the estimator that maximizes the following log-likelihood function $(LLF)$ of the PDF of the original variable $Z$,

$$L(Z_i) = \frac{-n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(Z^{\lambda} - \mu)^2 + \log J(\lambda, Z) \tag{3}$$

Where $\hat{\sigma^2}(\lambda)$ is the variance estimator of $Z^{(\lambda)}$. As for the use of YJT transformation to improve the forecastability of the time series the authors have used it in a few limited cases for some time series approaches. Matias et al,[15] used YJT to address the hyper normality in generalized autoregressive conditional heteroskedasticity process. Proietti and Riani [16] address the problem of seasonal adjustment of a nonlinear transformation of the original time series in structural time series models for which the linear Gaussian model holds, see also Riani, [10]. Othman and Mohammed Ali,[11] proposed an application algorithm to transform the positive original responses and the stationary responses to improve the nonparametric estimation of the functional time series using BCT and YJT.

## 3. The Computational Algorithm

On the subject of the estimating power parameter, Maximum Likelihood Estimation (MLE) and Bayesian estimation methods are still popular. There are three other methodologies for selecting the optimum power parameter, depending on some statistics and mathematical properties as a decision rules according to the type of the data: The first, is to use various test statistics such as the significant or highest p-value of the goodness of fit test statistic of transformed data normality [17, 18] and Bartlett's test of the homogeneity of variance [19]. The second methodology is used some of the efficiency indicators, such as the minimization of the Mean Square Error (Daga and Ilk, and Othman and Mohammed Ali[19, 11],, maximization of the coefficient of determination (Alyousif. and Abduahad, [20] and variance stabilizing Victor and Perera, [21]. While the third methodology is to use some mathematical properties of the probability distributions such as the minimization of the Kullback-Leibler divergence in non-Gaussian multivariate data Joachimi and Taylor [21] and some tests of concavity conditions of cost functions in the economics model Koebel, et al, [22]. But the expected difficulties are the attempt to estimate the power parameter according to two or more criteria (Othman and Mohammed Ali, [11]) because there is a significant possibility that there is no feasible region that achieves this optimization. The problem can be exacerbated when one of the two criteria is to achieve data normality, while the second is to improve one of the criteria of statistical model efficiency.

According to the hypothesis of normality of transformed response variable $Z_t^{(\lambda)}$ in ARMA process, the MLE of the scale and location parameters of response PDF are given by, $E(Z_t^{(\lambda)}) = \mu(\hat{\lambda};\Theta)$, and $Var(Z_t^{(\lambda)}/Z_{t-i}^{(\lambda)}), \varepsilon_{t-j}) = \sigma_\varepsilon^2$ where $\Theta = (\hat{\phi}_0, \hat{\phi}_1, \cdots, \hat{\phi}_p, \hat{\theta}_0, \hat{\theta}_1, \cdots, \hat{\theta}_q)', \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$, and $\hat{\lambda}$ is the Maximum Likelihood estimator of the power parameter $\lambda$ according to the LLF (3), in such,

$$\mu(\hat{\lambda};\Theta) = \hat{\phi}_0 + \sum_{i=1}^{p} \hat{\phi}_i Z_{t-i}^{(\lambda)} + \sum_{i=1}^{p} \hat{\theta}_j \hat{\varepsilon}_{t-j} \tag{4}$$

The MLE for the relationship model of the original response $Z_t$ is the back transform of the model equation (4),

$$\mu^{-1}(\hat{\lambda};\Theta) = \hat{\phi}_0 + \sum_{i=1}^{p} \hat{\phi}_i Z_{t-i} + \sum_{i=1}^{p} \hat{\theta}_j \hat{\varepsilon}_{t-j} \tag{5}$$

Where the original response $Z_(t-i)$ can be specified in the case of the BCT as follows,

$$Z_t = \begin{cases} (\lambda Z_t^{(\lambda)} + 1)^{\frac{1}{\lambda}} & \text{if } \lambda \neq 0 \\ exp(Z_t^{(\lambda)}) & \text{if } \lambda = 0 \end{cases} \tag{6}$$

While in the case of the YJT, it is specified as follows,

$$Z_t = \begin{cases} (\lambda Z_t^{(\lambda)} + 1)^{\frac{1}{\lambda}} - 1 & \lambda \neq 0 \text{ and } Z_t^{(\lambda)} \leq 0 \\ exp(Z_t^{(\lambda)}) - 1 & \lambda = 0 \text{ and } Z_t^{(\lambda)} z \leq 0 \\ 1 - (-Z_t^{(\lambda)}(2-\lambda) + 1)^{\frac{1}{2-\lambda}} & \lambda \neq 2 \text{ and } Z_t^{(\lambda)} < 0 \\ 1 - exp(-Z_t^{(\lambda)}) & \lambda = 2 \text{ and } Z_t^{(\lambda)} < 0 \end{cases} \tag{7}$$

whereas the error variance $\varepsilon^2$ that is not explained by the estimated model equation (5) of the original response is,

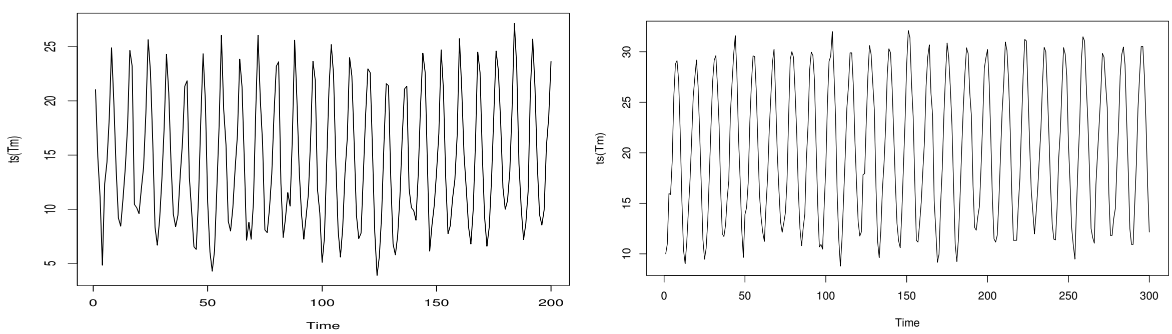
$$\varepsilon_t^2(\hat{\lambda};\Theta) = E\{Z_t - \mu^{-1}(\hat{\lambda};\Theta)\}^2 \tag{8}$$

The following proposed computational algorithm was developed in two directions: the first is to use the BCT on the original positive data, and then to deal with the stationary situation of the transformed data. As for the second direction, it is the use of an unconstrained YJT with a positive condition on the stationary data directly,

i Choosing possible values of the power parameter $\lambda$ such as the values of the interval (-2,2).
ii Apply BCT's model (1) to transform the original positive response $Z_t$ to $Z_t^{(\lambda)}$ and YJT's model (2) to transform the stationary response $\Delta^d Z_t$ to $(\Delta^d Z_t)^{(\lambda)}$.
iii Select the value of $\lambda$ , denoted by $\hat{\lambda}$ , that corresponds to the largest value of the LLF (3).
iv Stationaries $Z_t^{(\lambda)}$ by BCT to get the stationary transformed response $\Delta^d Z_t^{(\lambda)}$.
v Calculate Shapiro–Wilk test statistics for the new responses $\Delta^d Z_t^{(\lambda)}$ and $(\Delta^d Z_t)^{(\lambda)}$.
vi Fit an ARIMA model to the new response shown in step (v) and obtain $\sigma_\varepsilon^2$ which is the MSE.
vii Derivation the back-transformed time series from the fitting model within the original data domain to obtain the fitting model of $Z_t$.
viii Check the possibility of achieving the both constraints; the normality of the errors and the least MSE of the fitting model of $Z_t$.

## 4. The Applications

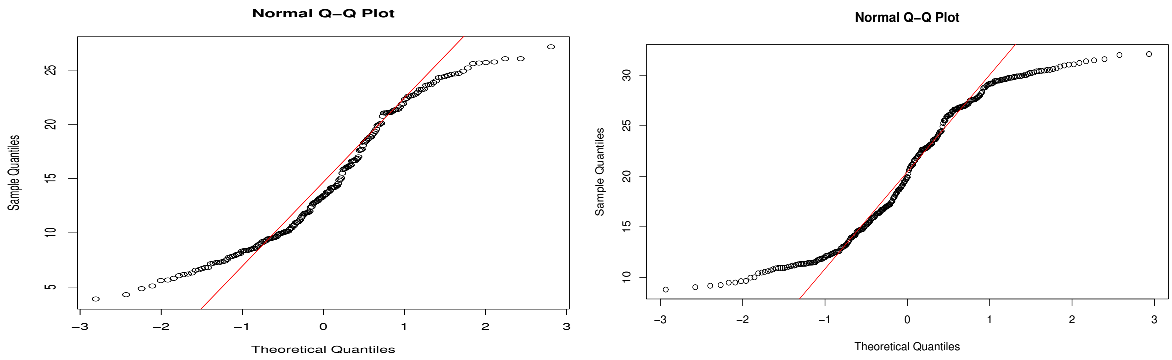
The two transformation models, BCT and YJT were applied to two time series datasets of monthly temperature averages. The first time series includes 200 observations of Ninavah city (TSN) in Iraq for the period 1976 to 2000 (Figure (1)). While the second timeseries includes 300 observations of Tunisia (TST) for the period 1991 to 2015 (Figure (1b)) [Othman and Mohammed, 2021][11]. R package was used to implement the proposed algorithm (https://climateknowledgeportal.worldbank.org).




(a) (b)

Figure 1. The plots of the two monthly temperature average series: (a) TSN (b) TST [11]

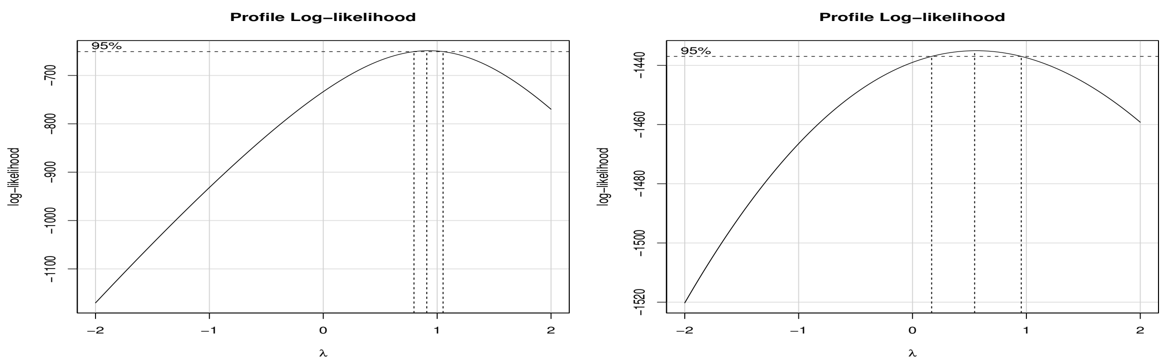
The S-shaped in Q-Q plots in figure (2) indicates that the two dataset examples are characterized by a Light-tailed and symmetric distributions. We can note that the datasets do not match the straight line that represents the normal state from the side of the two tails at the top and bottom. Seasonal changes in weather time series generally lead to the dispersion data relative to a normal distribution. According to the third step of the computational algorithm, the optimal power parameters of the two models of BCT and YJT were estimated for the two datasets by using the




(a) (b)

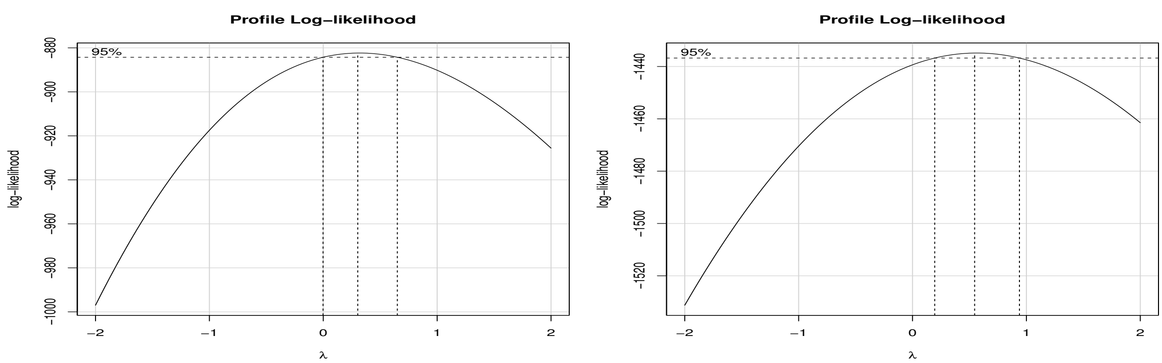
Figure 2. Q-Q plots of the two-time series: (a) TSN; (b) TST.

log-likelihood function (3). Figures (4) and (3) displays the log-likelihood convex curves for the two datasets as a function of the selected values interval (-2,2) of the power parameter $\lambda$.




(a) (b)

Figure 3. The log-likelihood curve relative to the selected $\lambda$ values of YJT for the two datasets (a) TSN (b) TST




(a) (b)

Figure 4. The log-likelihood curve relative to the selected $\lambda$ values of BCT for the two datasets (a) TSN (b) TST

Table (1)shows Kolmogorov-Smirnov and Shapiro-Wilk normality tests for the original and transformed TSN and TST datasets using the BCT and YJT transformation models. This table shows that $\lambda = 0.16$ and 0.72 maximizes the LLFs of the two datasets by using the BCT model. Whereas in the case of using the YJT model, the LLFs is maximized when $\lambda = 0.93$ and 0.56. Although the $\lambda$ values were estimated as maximizing the LLF, they did

Table 1. The normality tests of the original and transformed datasets.

| Datasets | Responces | Optimal | K-Smirnov | | Shapiro-Wilk | |
|---|---|---|---|---|---|---|
| | | | Stat. | p-value | State | p-value |
| TSN | $z_t$ | 1.00 | 0.0953 | 0.0002 | 0.9441 | 8.5E-7 |
| | $\Delta^d Z_t^{(\lambda)} \| BCT$ | 0.16 | 0.0840 | 0.0020 | 0.9650 | 0.0001 |
| | $\left(\Delta^d Z_t\right)^{(\lambda)} \| YJT$ | 0.93 | 0.9790 | 0.0000 | 0.9881 | 0.1250 |
| TST | $z_t$ | 1.00 | 0.1060 | 2.0E-8 | 0.9262 | 5.0E-11 |
| | $\Delta^d Z_t^{(\lambda)} \| BCT$ | 0.72 | 0.1102 | 4.8E-9 | 0.9280 | 8.1E-11 |
| | $\left(\Delta^d Z_t\right)^{(\lambda)} \| YJT$ | 0.56 | 0.8730 | 0.0000 | 0.8092 | 2.2E-16 |

not shift the two datasets sufficiently to the normal shape. But based on the slightly increased p-values for the two tests of the transformed TSN dataset by the BCT model, it can be said that the transformation has accommodated the data in a shape that is less skewed than the original data. This conclusion could also include the use of YJT according to the Kolmogorov-Smirnov test, while the large p-value of the Shapiro-Wilk test indicates that the data are completely normal. As for the TST dataset, the shifting achieved by the two transformation models was not sufficient to match the data or harmonized it to the normal shape or even to approximate normality. This potential intersection between the principle of maximizing the LLF and the fulfilment of the data normality condition may result from the sensitivity of the maximum likelihood estimator to outliers [23](Raymaekers and Rousseeuw) or when the data are extremely skewed. According to the sixth step of the computational algorithm and using the R software package, we are based on the standard analysis of the time series to determine the tentative models of the two datasets according to the AIC criterion. The Maximum Likelihood estimation method was used to estimate the appropriate models for the three cases: the original responses, the transformed responses by the BCT model and the transformed stationary responses by the YJT model. For the two transformation models, in addition to the data normality goal, the subsequent decision rules after the traditional time series building are the indicators of the significant of the estimated as well as reliability and efficiency. Ultimately, the authors considered the errors of normality and the mean squared errors of the last year prediction to be the critical factors in adopting the appropriate model for the datasets under study. Tables (2) and (3) lists all the cases considered in this article. Figures (5) and (6) shown autocorrelation (ACF) function plots and residual histograms for all fitted models.

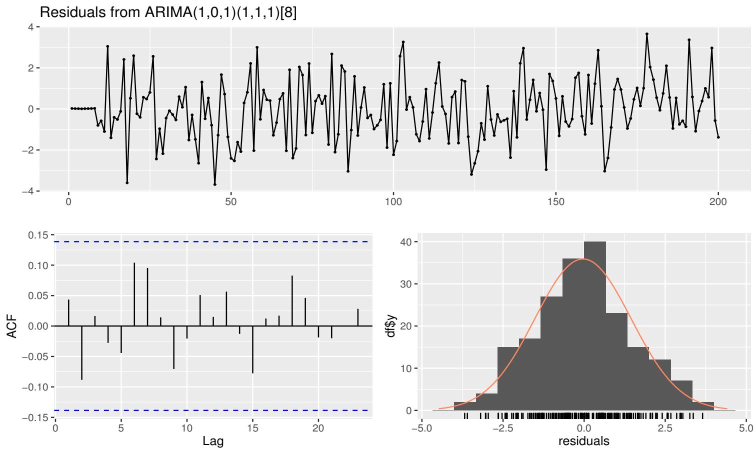
Table 2. The appropriate models, parameter estimates and some efficiency indicators resulting from the use of the two transformation models of the TSN Dataset.

| Responces | Estimated Model | Parameters Estimate | | SE | z-value | p-value | MSE |
|---|---|---|---|---|---|---|---|
| $z_t$ | ARIMA $(1,0,0)(1,1,1)_8$ | AR | 0.3026 | 0.0709 | 4.2705 | 0.0000 | 2.2405 |
| | | AR Sea. | 0.0897 | 0.0796 | 1.1270 | 0.2598 | |
| | | MA Sea. | -0.1000 | 0.0851 | -11.7475 | 0.0000 | |
| $\Delta^d Z_t^{(\lambda)} \vert BCT$ | ARIMA $(1,0,0)(1,1,1)_8$ | AR | 0.3565 | 0.0687 | 5.1922 | 0.0000 | 1.9940 |
| | | AR Sea. | 0.1437 | 0.0801 | 1.7937 | 0.0729 | |
| | | MA Sea. | -0.1000 | 0.0696 | -14.3666 | 0.2200 | |
| $(\Delta^d Z_t)^{(\lambda)} \vert YJT$ | AR(1) | AR | 0.2561 | 0.0711 | 3.5995 | 0.0003 | 1.9305 |

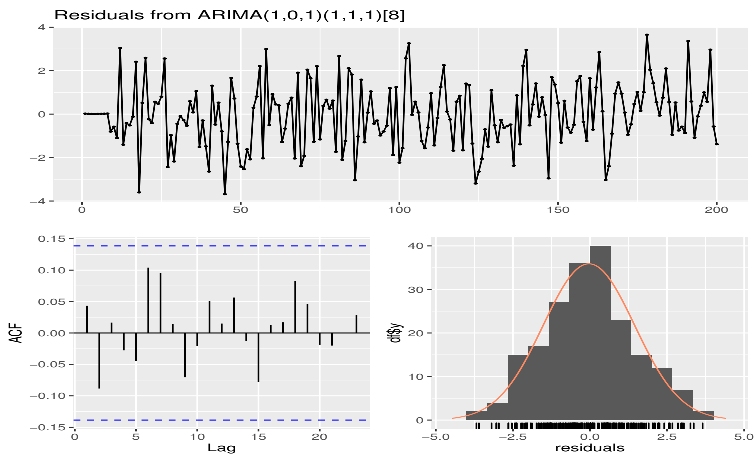
Table 3. The appropriate models, parameter estimates and some efficiency indicators resulting from the use of the two transformation models of the TST Dataset

| Responces | Estimated Model | Parameters Estimate | | SE | z-value | p-value | MSE |
|---|---|---|---|---|---|---|---|
| $z_t$ | ARIMA $(1,0,1)(1,1,1)_{12}$ | AR | 0.2637 | 0.1559 | 1.6916 | 0.0907 | 0.3244 |
| | | MA | 0.0853 | 0.1588 | 0.5370 | 0.5913 | |
| | | AR Sea. | -0.0157 | 0.0779 | -0.2015 | 0.8403 | |
| | | MA Sea. | -0.8509 | 0.0571 | -14.9044 | 0.0000 | |
| $\Delta^d Z_t^{(\lambda)} \vert BCT$ | ARIMA $(1,0,1)(1,1,1)_{12}$ | AR | 0.2442 | 0.1524 | 1.6024 | 0.1091 | 0.3002 |
| | | Ma | 0.1086 | 0.1538 | 0.7062 | 0.4801 | |
| | | AR Sea. | -0.0184 | 0.0764 | -0.2404 | 0.8100 | |
| | | MA Sea. | -0.8617 | 0.0546 | -15.7808 | 0.2000 | |
| $(\Delta^d Z_t)^{(\lambda)} \vert YJT$ | AR(1) | AR | 0.0711 | 4.2112 | 4.2112 | 0.0000 | 0.2992 |

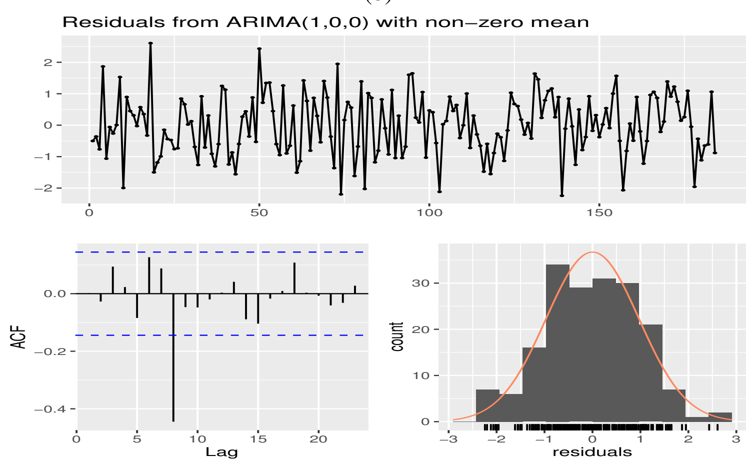
In this article, the authors treat the results of time series modeling as a package in an attempt to decide and choose the best model by diagnosing the effect of data transformation on the components of model estimation and prediction results. This treatment tries to balance and mix the normalization goal of the data (Table (1)) with the indicators in tables(2) and (3) and the behavior of the residuals in figures (5) and (6). According to increasing p-values of the statistics of the Shapiro-Wilk and Kolmogorov-Smirnov tests in table (1), the BCT model succeeded in transforming the TSN dataset to the normal or semi-normal shape. While the YJT did not succeed in transforming the same dataset to a normal shape. On the other hand, for the same dataset, the p-values of the z-statistics in table (2) indicate the high significance of the estimated parameters of the fitted model of transformed data by the YJT model. In addition to the simplicity of the model and its forecastability, which is evident by the MSE values for the last year's predictions, which is the smallest among its peers. Also, the normality of the error in Figure (5c) supports this conclusion. This analysis fully applies to the TST dataset transformed by the YJT model, whose fitted model was also shown to be of the same simplicity, significant and high forecastability, although the transformation did not produce an optimal power parameter capable of transforming it to the normal shape, see table (3) and figure (6c).




(a)

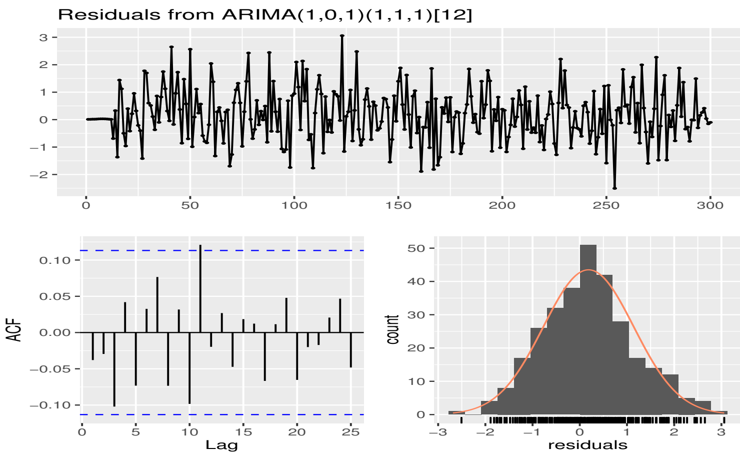



(b)

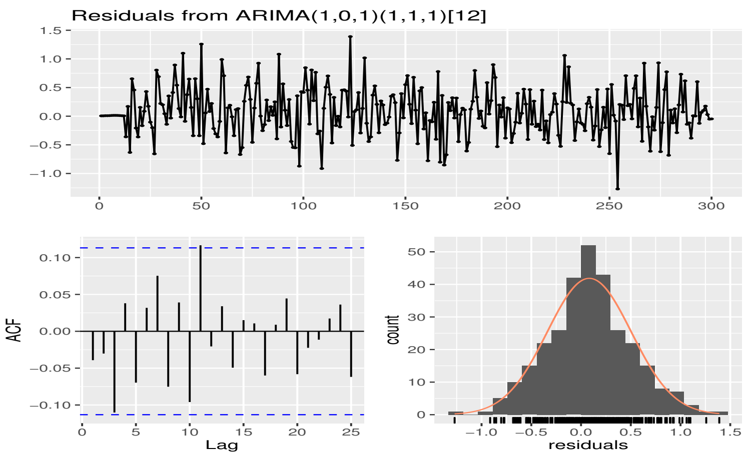



(c)

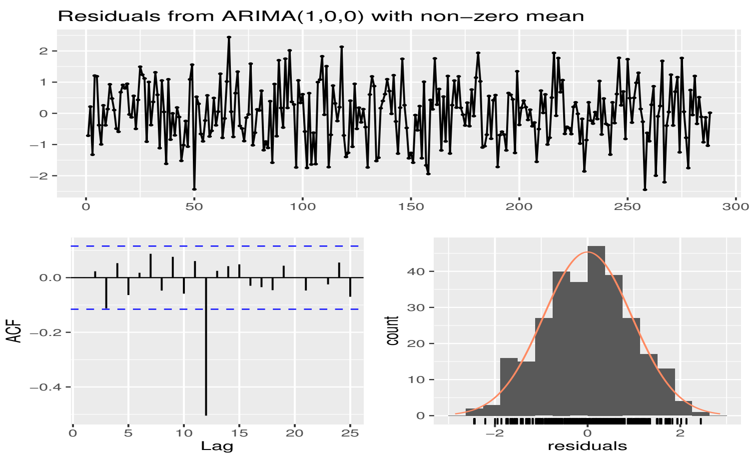
Figure 5. ACF plots and residuals histograms of the fitted ARIMA models of the TSN datasets. (a) Original response (b) Transformed response by BCT. (c) Transformed response by YJT.




(a)




(b)




(c)

Figure 6. ACF plots and residuals histograms of the fitted ARIMA models of the TST datasets. (a) Original response (b) Transformed response by BCT. (c) Transformed response by YJT.

## 5. Conclusion

It is known that the power parameter estimation resulting from maximizing the LLF sometimes does not shift the data enough to make it normal. In cases where the normality of the data is achieved, it is clear that the quality of other modeling efficiency criteria may not be achieved and vice versa. Therefore, it can be said that there is no feasible solution area for all fitting quality criteria and transformation goals. The authors believe that these difficulties stem from the use of power transformation to transform the response variable before the model fitting. While the use of traditional efficiency criteria after modeling the transformed datasets contributed to the success in choosing a simple model. The analysis proved that the YJT model was more flexible in smoothing the data and contributed to accessing a simple model with good forecastability.

REFERENCES

1. Bartlett, Maurice S, *The use of transformations*, journal Biometrics, vol. 3, no. 1, pp. 39–52, 1947.
2. Tukey, John W, *On the comparative anatomy of transformations*, The Annals of Mathematical Statistics, pp. 602–632, 1957.
3. Dike, AO and Otuonye, EL and Chikezie, DC, *The nth Power Transformation of the Error Component of the Multiplicative Time Series Model*, Journal of Advances in Mathematics and Computer Science pp. 1–15, 2016.
4. Box, George EP and Cox, David R, *An analysis of transformations*, Journal of the Royal Statistical Society: Series B (Methodological), vol. 26, no. 2, pp.2–243, 1964.
5. Atkinson, Anthony C and Riani, Marco and Corbellini, Aldo, *The Box–Cox Transformation: Review and Extensions*, Statistical Science.vol.36,no.2, pp.239–255,2021.
6. Box, GEP and Jenkins, GM, *Time series Analysis. Forecasting and Control*, Halden-Day,San Francisco,1970.
7. Hipel, Keith W and McLeod, A Ian, *Time series modelling of water resources and environmental systems*, 1994.
8. Chen, Cathy WS and Lee, Jack C, *On Selecting a Power Transformation in Time-Series Analysis frequency information*, Journal of Forecasting,vol.16,no. 5, pp.343–354,1997.
9. Guerrero, Victor M and Perera, Rafael, *Variance stabilizing power transformation for time series*, Journal of Modern Applied Statistical Methods,vol.3,no.2, pp.9, 2004.
10. Riani, Marco, *Robust transformations in univariate and multivariate time series*, Econometric Reviews, vol.28,no.1-3 pp.262–278, 2008.
11. Othman, Sameera Abdulsalam and Ali, Haithem Taha Mohammed, *Improvement of the Nonparametric Estimation of Functional Stationary Time Series Using Yeo-Johnson Transformation with Application to Temperature Curves* Advances in Mathematical Physics, vol.2021,2021.
12. Hopwood, William S and McKeown, James C and Newbold, Paul, *Time series forecasting models involving power transformations*, Journal of Forecasting,vol. 3,no.1 pp.57–61, 1984.
13. Terasaka, Takahiro and Hosoya, Yuzo, *A modified Box-Cox transformation in the multivariate ARMA model*, Journal of the Japan Statistical Society, vol. 37,no.1 pp.1–28, 2007.
14. Yeo, In-Kwon and Johnson, Richard A, *A new family of power transformations to improve normality or symmetry*, Biometrika, vol. 87, no. 4, pp. 954–959, 2000.
15. J. M. Matias, W. Gonzallez-Manteiga, J. Tadonez, C. Ordonez, *Managing distribution changes in time series prediction*, Journal of computational and applied mathematics, vol.191,no.2, pp. 206–215, 2006.
16. Proietti, Tommaso and Riani, Marco, *Transformations and seasonal adjustment*, Journal of Time Series Analysis, vol.30,no.1,pp.47–69,2009 1992.
17. O. Asar, O. Ilk, O. Dag, *Estimating Box-Cox power transformation parameter via goodness-of-fit tests* Communications in Statistics-Simulation and Computation, vol. 46,no.1, pp. 91–105, 2017.
18. VA, Jorge I and Correa, Juan C and Marmolejo-Ramos, Fernando and others, *A new approach to the Box-Cox transformation*, Frontiers in Applied Mathematics and Statistics, 2015.
19. Dag, Osman and Ilk, Ozlem, *An algorithm for estimating Box–Cox transformation parameter in ANOVA*, Communications in Statistics-Simulation and Computation, vol. 46,no.8,pp.6424–6435,2017.
20. Alyousif, Haithem Taha and Abduahad, Fedaa Noeel, *Develop a Nonlinear Model for the Conditional Expectation of the Bayesian Probability Distribution (Gamma–Gamma)*, Al-Nahrain Journal of Science,vol. 17,no.2,pp.205–212,2017, 2014.
21. Joachimi, Benjamin and Taylor, AN, *Forecasts of non-Gaussian parameter spaces using Box–Cox transformations*, Monthly Notices of the Royal Astronomical Society, vol. 416, no. 2, pp. 1010–1022, 2011.
22. B. Koebel, M. Falk and F, *Imposing and testing curvature conditions on a Box–Cox cost function*, Journal of Business and Economic Statistics,vol. 21, no. 2, pp. 319–335, 2003.
23. Raymaekers, Jakob and Rousseeuw, Peter J, *Transforming variables to central normality*, Machine Learning,pp. 1–23, 2021.