Journal of the Nigerian Association of Mathematical Physics
Volume 7(2003)

Effect Of Box-Cox Transformation Parameter On Forecast From ARIMA(p,1,0) models.

N. Ekhosuehi and S. E. Omosigho
Department of Mathematics,
University of Benin, Benin City, Nigeria.

Abstract

In this paper we consider the effect of applying a family of Box-Cox transformations to time series data that exhibit stochastic upward linear trend which admits differencing of order one. We examine the interaction between Box-Cox transformation and forecasting from the ARIMA(p, 1, 0) model. We demonstrate the importance of the Box-Cox transformation parameters $\lambda = 1$, $\lambda = 0.5$ *and* $\lambda = 0$.

Key words: *Box-Cox transformations; Model fitting; Forecasting.*
pp 105-110

1.0 Introduction

The analysis of time series data that exhibit stochastic upward linear trend may involve the use of Box-Cox transformation. In this paper, we show that the choice of the Box-Cox parameter λ can affect the forecasting performance of the resulting ARIMA(p, 1, 0) model. Given the time series observations $\{y_t, t = 1,2,\ldots,n\}$ where $y_t > 0$ the Box-Cox transformation examined in this paper is given by

$$y^{\bullet}_{\lambda,t} = \begin{cases} \ln(y_t), & \lambda = 0 \\ y_t^{\lambda}, & 0 < \lambda \le 1 \end{cases} \tag{1.1}$$

Our motivation is the work of Franses and De Bruin (2002) who observed that in time series modelling it is not clear if the choice of λ influences key parameters in the subsequent model and its out-of-sample forecast. We propose a procedure for obtaining optimal Box-Cox transformation. Our analysis is based on simulated data. We assume we have a time series observations $\{y_t, t = 1,2,\ldots,n\}$ where $y_t > 0$, and that the series display stochastic upward linear trend. The rest of the paper is organised as follows. In section 2, we describe the methodology used in our analysis while the simulation experiment used for generating sets of time series data is presented in section 3. Model development is in section 4 and our results are presented and discussed in section 5. Conclusions are in section 6.

2.0 Methodology

Given time series observations $\{y_t, t = 1,2,\ldots,n\}$ where $y_t > 0$ that exhibit stochastic upward linear trend, obtain

$$\Delta y^{\bullet}_{\lambda,t} = \begin{cases} \ln(y_t) - \ln(y_{t-1}), & \lambda = 0 \\ y_t^{\lambda} - y_{t-1}^{\lambda}, & 0 < \lambda \le 1 \end{cases} \tag{2.1}$$

If the transformed data $\Delta y^{\bullet}_{\lambda,t}$ admits an ARIMA(p, 1, 0) model, then such model can in general be represented as

$$\phi(B)\{(1 - B)\, y_t^{\lambda} - \mu\} = \varepsilon_t, \quad 0 < \lambda \le 1 \tag{2.2}$$

But the traditional ARIMA(p, m, 0), Beran et al.(1998) is known to be of the form:

$$\phi(B)\{(1-B)^m X_t - \mu\} = \varepsilon_t \tag{2.3}$$

where m a positive integer is the number of times X_t has to be differenced to achieve stationarity. For our case, m = 1, $X_t = y_t^{\lambda}$. Now, the model presented in (3) above, can be re-written as

$$X_t = d + \alpha_1(\lambda)X_{t-1} + \ldots + \alpha_p(\lambda)X_{t-p} + \varepsilon_t \tag{2.4}$$

where $\phi(B) = 1 - \sum_{j=1}^{p} \alpha_j(\lambda)B^j$, $d = \mu(1-\alpha_1(\lambda) - \ldots - \alpha_p(\lambda))$, $X_t = \Delta y^{\bullet}_{\lambda,t}$ or $(1-B)y_t^{\lambda}$

and ε_t is normally distributed with mean 0 and variance σ_ε^2. This specification indicates that the parameters of the ARIMA(p, 1, 0) accepted for the transformed data is a function of λ. Our interest is to demonstrate that the choice of λ affects the out-of-sample performance of the resulting model when used to forecast. Our measure of performance is the forecast mean square error. The parameter d in (2.4) is included in the model if it is statistically different from zero. A systematic method for testing if d is statistically different from zero is given by Bowerman and O'Connell(1993, pp.465). Suppose we are able to obtain a specific model for a given transformed data $\Delta y^{\bullet}_{\lambda,t}$ as in (2.4) then we can use the model to obtain a forecast by using

$$\begin{aligned}\hat{X}_{t+k} &= d + \alpha_1(\lambda)X_{t+k-1} + \ldots + \alpha_p(\lambda)X_{t+k-p} \\ &= f(\Delta y^{\bullet}_{\lambda,t})\end{aligned} \tag{2.5}$$

The forecast $\hat{X}_{t+k}$ is for the transformed data. In practice, the interest is on forecast $\hat{y}_{t+k}$ which reflects the observations in levels. Then the forecast $\hat{y}_{t+k}$ is obtained by using the following inverse transformation:

$$\hat{y}_{t+k} = f^{-1}(\Delta y^{\bullet}_{\lambda,t}) = \begin{cases} \exp(\hat{X}_{t+k})y_{t+k-1} & ,\lambda = 0 \\ (y_{t+k-1}^{\lambda} + \hat{X}_{t+K})^{1/\lambda} & ,0 < \lambda \le 1 \end{cases} \tag{2.6}$$

To examine the effect of λ on the performance of the resulting model, a grid search was performed on the set λ = 0, 0.05(0.05),…,1. A summary of the procedure is as follows:

Step 1 Given y_t obtain $X_t = \Delta y^{\bullet}_{\lambda,t}$ for each λ.

Step 2 Derive the ARIMA(p, 1, 0) model for the transformed data X_t

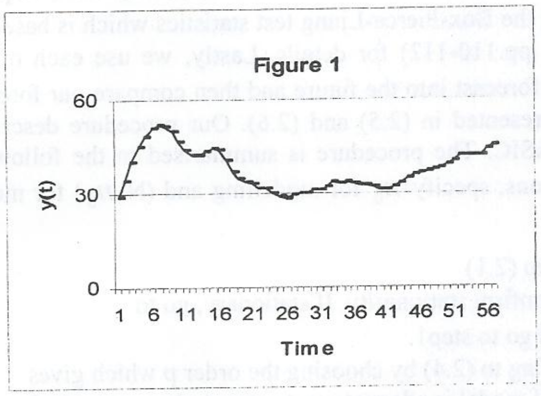
Step 3 Obtain forecast for X_t i.e. $\hat{X}_{t+k}$

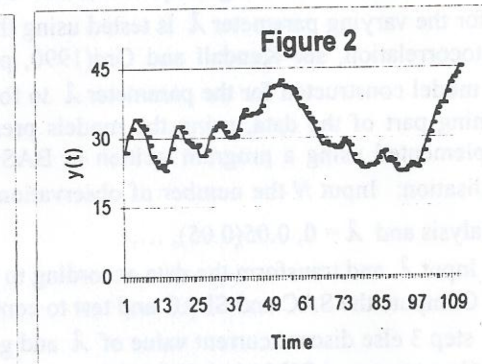
Step 4 Use inverse transform to obtain $\hat{y}_{t+k}$

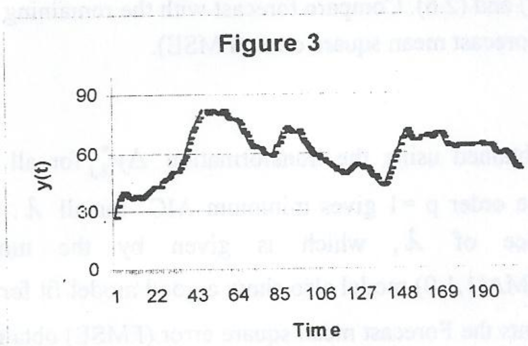
3.0 Simulation Experiment

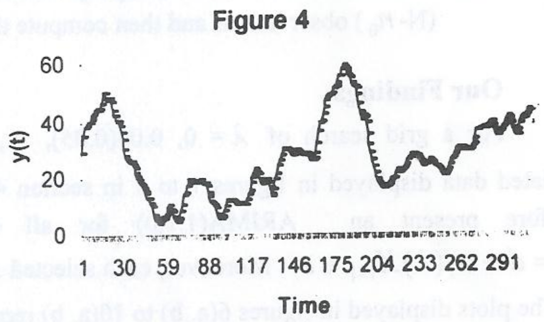
In order to demonstrate the effect of λ on the performance of the resulting forecast, we used simulated time series data having stochastic upward linear trend. The time series observations was generated using the equation $y_t = 1.7y_{t-1} - 0.7y_{t-2} + \varepsilon_t$, (cf. Diggle, 1995; pp.166) where ε_t is a white noise process that is N(0, σ_ε^2). We have chosen $y_0 = 20$ and $y_{-1} = 10$ to ensure that the observations $\{y_t\}$ is non-negative. The sequence ε_t was generated using random number generator which is uniformly distributed in the interval

(0,1) and then using the Box-Muller transformation (Pidd, 1984), to obtain N(μ, σ^2) variates which were then transformed into N(0, 1) variates. Five sets of data having 58,112, 210, 310 and 418 observations using the process described above were generated. We decided to use some part of the data for modelling and the rest observations were used to examine the out-of-sample performance. The five sets of data generated using the simulation method described above are plotted in Figures 1 to 5 below.









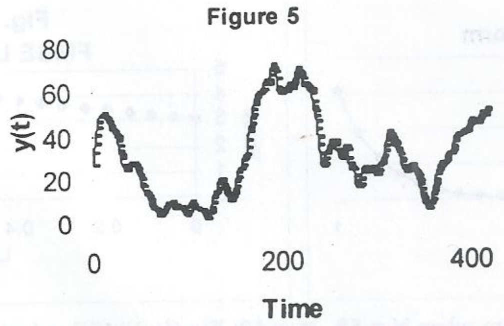


Figure 1 represent the data set having 58 observations, figure 2 represent the data set having 112 observations, figure 3 represent the data set having 210 observavtions, figure 4 represent the data set having 310 observations and figure 5 represent the data set having 418 observations. Each set of data was subjected to the transformation (2.1) above using a grid search of the Box-Cox parameter λ, for $\lambda = 0, 0.05(0.05), \ldots, 1$.

4.0 Model Development

The sample autocorrelation function (SAC) and sample partial autocorrelation function (SPAC) were computed for the transformed data up to a specified length n_0 (keeping the remaining N-n_0 to compare forecast) and tested for stationarity. It is known that SAC or SPAC that refuses to die down quickly is an indication of non-stationarity, (Bowerman and O' Connell, 1993). For each λ in the set of grid search for which stationarity is achieved, we constructed an ARIMA(p,1,0) model by choosing the order p where the AIC function attains its minimum, following the procedure of Shibata(1976) and Pukkila et al.(1990). The adequacy of each model for the varying parameter λ is tested using the Box-Pierce-Ljung test statistics which is based on the residual autocorrelation, see Kendall and Ord(1990, pp.110-112) for details. Lastly, we use each of the ARIMA(p,1,0) model constructed for the parameter λ to forecast into the future and then compare our forecast with the remaining part of the data, using the models presented in (2.5) and (2.6). Our procedure described above, was implemented using a program written in BASIC. The procedure is summarised in the following algorithm.Initialisation: Input N the number of observations, specify n_0 for modelling and (N-n_0) for model performance analysis and $\lambda = 0, 0.05(0.05), \ldots, 1$

Step1: Input λ and transform the data according to (2.1)

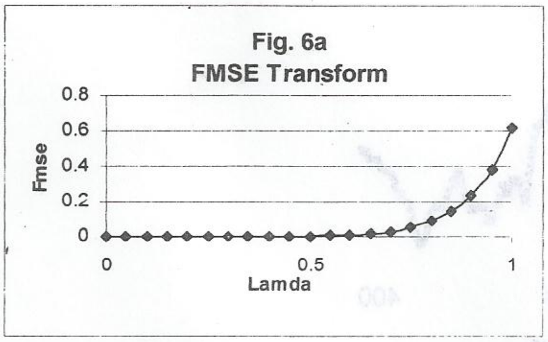
Step2: Compute the SAC and SPAC and test to confirm stationarity. If stationary, go to step 3 else discard current value of λ and go to step1.

Step3: Construct an ARIMA(p,1,0) model according to (2.4) by choosing the order p which gives minimum AIC. Test for model adequacy. If model is adequate go to step 4 else go to step1.

Step4: Forecast into the future using equations (2.5) and (2.6). Compare forecast with the remaining (N-n_0) observations and then compute the Forecast mean square error (FMSE).

5.0 Our Findings

For a grid search of $\lambda = 0, 0.05(0.05), \ldots, 1$ obtained using the transformation $\Delta y^{*}_{\lambda,t}$ for all the simulated data displayed in figures 1 to 5 in section 4, the order p =1 gives minimum AIC for all λ. We therefore present an ARIMA(1,1,0) for all choice of λ, which is given by the model $X_t = d + \alpha_1(\lambda)X_{t-1} + \varepsilon_t$. Moreover, each selected ARIMA(1,1,0) model also show a good model fit for all λ. The plots displayed in figures 6(a, b) to 10(a, b) represents the Forecast mean square error (FMSE) obtained using the remaining part of each of our simulated data set explained in section 4.



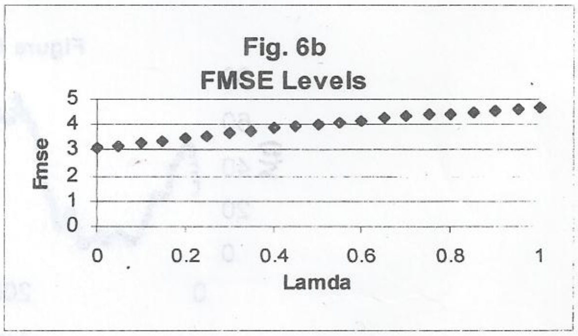
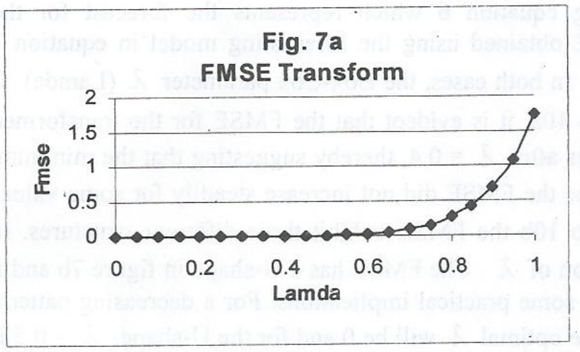


Figure 6a: FMSE transform when N = 58, n_0 = 50; Fig.6b: FMSE at levels when N = 58, n_0 = 50



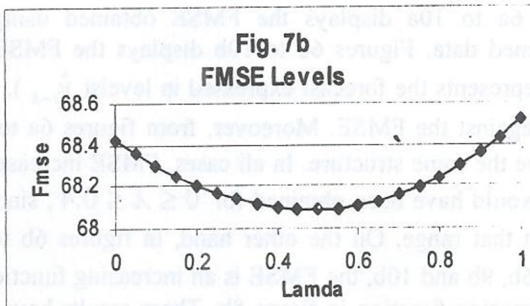
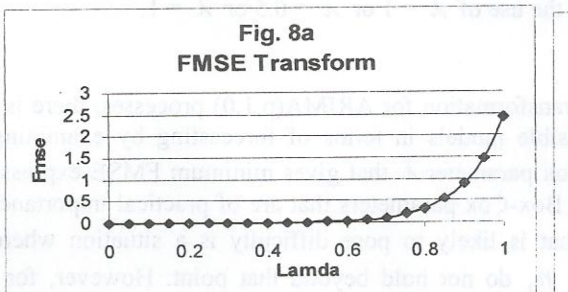


Figure 7a: FMSE transform when N = 112, n_0 =100; **Figure 7b**: FMSE at levels when N = 112, n_0 = 100



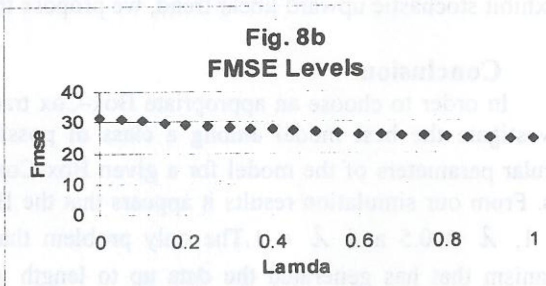
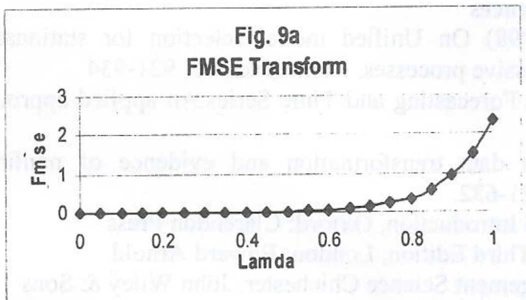


Figure 8a: FMSE transform when N = 210, n_0 = 200; **Figure 8b**: FMSE at levels when N = 210, n_0 = 200



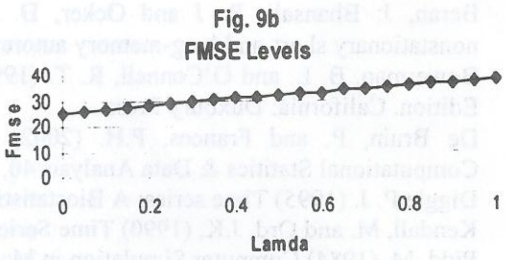
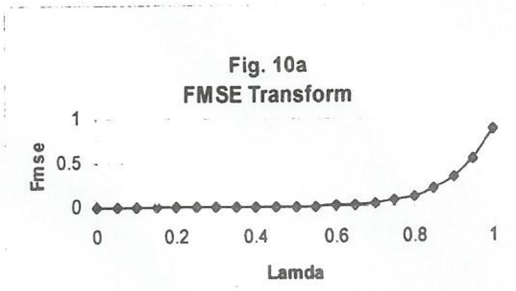


Figure 9a: FMSE transform when N = 310, n_0 =300. **Figure 9b**: FMSE at levels when N = 310, n_0 = 300



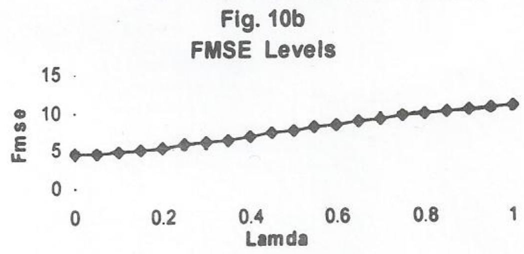


Figure: 10a :FMSE transform when N = 418, n_0 = 400; **Figure 10b**: FMSE at levels when N = 418, n_0 = 400

Figures 6a to 10a displays the FMSE obtained using equation 6 which represents the forecast for the transformed data. Figures 6b to 10b displays the FMSE obtained using the forecasting model in equation 7 which represents the forecast expressed in levels($\hat{y}_{t+k}$). In both cases, the Box-Cox parameter λ (Lamda) is plotted against the FMSE. Moreover, from figures 6a to 10a, it is evident that the FMSE for the transformed data have the same structure. In all cases, FMSE increases after λ = 0.4, thereby suggesting that the minimum FMSE would have been obtained for $0 \le \lambda \le 0.4$, since the FMSE did not increase steadily for some values of λ in that range. On the other hand, in figures 6b to 10b the FMSE exhibit three different structures. In figures 6b, 9b and 10b, the FMSE is an increasing function of λ . The FMSE has a U-shape in figure 7b and it is a decreasing function in figure 8b. These results have some practical implications. For a decreasing pattern, the optimal λ will be 1, and for an increasing pattern, the optimal λ will be 0 and for the U-shape, λ = 0.5 is the optimal value. We conjecture that for practical purposes, these three cases are likely to be the important values for the Box-Cox transformation whenever the interest is on forecasting. Therefore for time series data that exhibit stochastic upward linear trend, we propose the use of λ = 1 or λ = 0.5 or λ = 1.

6.0 Conclusion

In order to choose an appropriate Box-Cox transformation for ARIMA(p,1,0) processes, there is need to investigate the best model among a class of possible models in terms of forecasting by examining the particular parameters of the model for a given Box-Cox parameter λ that gives minimum FMSE expressed in levels. From our simulation results it appears that the Box-Cox parameters that are of practical importance are λ = 1, λ = 0.5 and λ = 1.The only problem that is likely to pose difficulty is a situation where the mechanism that has generated the data up to length n_0 do not hold beyond that point. However, for such processes, forecast based on the Box-Cox transformation may not be optimal. Perhaps regime-switching models will be adequate under such conditions. Such models are currently considered.

References

[1] Beran, J; Bhansali, R. J and Ocker, D. (1998) On Unified model selection for stationary and nonstationary short-and long-memory autoregressive processes. Biometrika, 85, 921-934

[2] Bowerman, B. L. and O'Connell, R. T. (1993) Forecasting and Time Series:An applied approach. 3rd Edition. California: Duxbury Press.

[3] De Bruin, P. and Frances, P.H. (2002) On data transformation and evidence of nonlinearity. Computational Statitics & Data Analysis 40, .621-632.

[4] Diggle P. J. (1995) Time series: A Biostatistical Introduction, Oxford: Clarendon Press

[5] Kendall, M. and Ord, J.K, (1990) Time Series, Third Edition, London: Edward Arnold

[5] Pidd, M. (1984) Computer Simulation in Management Science Chichester: John Wiley & Sons

[5] Pukkila, T.; Koreisha, S.; and Kallinen, A.(1990). The identification of ARMA models, Biometrika, 77, (3), 537-48.

[5] Shibata Ritei (1976) The Selection of the order of an autoregressive model by Akaike's information criterion. Biometrika 63, 117-26.