Electronic Journal of Applied Statistical Analysis
EJASA (2012), Electron. J. App. Stat. Anal., Vol. 5, Issue 2, 137 – 150
e-ISSN 2070-5948, DOI 10.1285/i20705948v5n2p137

USING ITERATIVE LINEAR REGRESSION MODEL TO TIME SERIES MODELS

Zouaoui Chikr-el-Mezouar[*(1)], **Mohamed Attouch**[(2)]

[(1)]*Department of Exact Science, Bechar University, Algeria*
[(2)]*Department of Mathematics, Djillali Liabès University, Algeria*

Received 16 September 2010; Accepted 07 September 2011
Available online 26 April 2012

Abstract*: This paper presents an iterated liner regression model and compares its forecasting performance with the traditional liner regression (LR) and Box-Jenkins ARIMA models using two well-known time series datasets: airline data and sunspot data. The difference between iterated LR and traditional LR is the former considers the error and uses it as dependent variables again to reduce the error rate until error rate is very small. The results show that the performance of iterated LR is slightly better than Box-J model and much better than traditional LR models.*

Keywords*: Linear regression, iterated linear regression, time series and bilinear models.*

1. Introduction

Linear Regression (LR) has been used to model time series in various fields of applications including identification and classification and dynamical systems. The complexity observed and encountered in time series suggests the use of LR which have been proven to be capable of modeling linear relationship without a priori assumption of the nature of the relationship. Faraway and Chatfield (1998) fits a variety of neural network models to the well-known airline data and compared the resulting forecasts with those obtained from the Box-Jenkins, holt-Winters methods and linear regression.
Many commercial packages are available for fitting LR models. Here we have used the MATLAB package and the MINITAB package (release 14).
From our experience, we found that the final results of the fitting and forecasting depend on the

[*] E-mail: chikrtime@yahoo.fr

choice of these explanatory variables.
In this work the explanatory variables is determined by determining the best subset autoregressive (AR) model to the data using the Box-Jenkins (BJ) technique and the AIC and BIC criterions for the model selection. Assuming that the resulting model is:

$$X_t = \phi_0 + \phi_1 X_{t-k_1} + \phi_2 X_{t-k_2} + \ldots + + \phi_p X_{t-k_p} + a_t .$$

Then we choose explanatory variables as $X_{t-k_1}, X_{t-k_2}, \ldots, X_{t-k_p}$. Section 2 defined Iterated Linear Regression method. In the following section we discuses a case studies. Finally conclusions are drawn.

2. Iterated Linear Regression Method (ILR)

Suppose that the time series being generated through a MA(q), or ARMA(p,q) or bilinear BL(p,o,m,k). Now, we propos a new method for fitting ARMA or bilinear models to observed data $\{X_t\ ,\ t=1,2,\ldots,N\}$ by entering the error terms as the explanatory variables and call this method as “iterated linear regression” or “repeated residual linear regression”. The method for fitting an ARMA(p,q) model is described in the following steps:

1. Fit an LR model using the AR terms $X_{t-k_1}, X_{t-k_2}, \ldots, X_{t-k_p}$.as the explanatory variables, then obtain the resulting residuals (errors), the error sum of squares and denote them by $\hat{a}_1^{(1)} = \left(X_t - \hat{X}_t\right)$ and $SSE^{(1)}$ respectively. We propose that these errors are initial explanatory errors in ARMA.
2. Fit an LR model using $X_{t-k_1}, X_{t-k_2}, \ldots, X_{t-k_p}$., $\hat{a}_{t-1}^{(1)}, \hat{a}_{t-2}^{(1)}, \ldots, \hat{a}_{t-p}^{(1)}$ as explanatory variables, then obtain the resulting the new residuals (errors) and the new error sum of squares and denote them by $\hat{a}_t^{(2)}$ and $SSE^{(2)}$ respectively.
3. Repeat the fitting step [2] until the mean square errors converge (after i iterations) if: $\left(\left|SSE^{(i)} - SSE^{(i-1)}\right| / \left|SSE^{(i)}\right|\right) < 0.000001$.

For example, an ARMA(2,1) model: $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \theta_0 + a_t - \theta_1 a_{t-1}$, is fitted, given a realization $\{X_t\ ,\ t=1,2,\ldots,N\}$ as follows:

1. Fit an LR AR(2) model choosing X_t as dependent variable and X_{t-1}, X_{t-2} as the explanatory variables. Calculate the corresponding residuals (errors) $a_{t-1}^{(1)}$, the error sum of squares $SSE^{(1)}$.
2. Fit an ARMA (2,1) model by choosing X_t as the dependent variable and X_{t-1}, X_{t-2} and $a_{t-1}^{(1)}$ as the explanatory variables. Calculate the corresponding residuals (errors) $\hat{a}_t^{(2)}$, the error sum of squares $SSE^{(2)}$.
3. Repeat step (2) until the convergence occurs. We say That the procedure converges at

step *i* (after *i* iterations) if: $\left(\left|SSE^{(i)} - SSE^{(i-1)}\right| / \left|SSE^{(i)}\right|\right) < 0.000001$.

The same procedure is applied for the bilinear time series models. For example to fit the bilinear model BL(2,0,1,1): $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \theta_0 + a_t + \beta_{11} X_{t-1} a_{t-1}$, given a realization $\{X_t$, $t=1,2,\ldots,N\}$, we proceed as follows:

1. Fit an AR (2) model by choosing X_t as the dependent variables, X_{t-1}, X_{t-2} as the explanatory variables. Calculate the corresponding residuals (errors) $a_{t-1}^{(1)}$ and the corresponding residual sum of squares $SSE^{(1)}$.
2. Fit an BL(2,0,1,1) model by choosing X_t as the dependent variables, X_{t-1}, X_{t-2} ,and $X_{t-1} a_{t-1}^{(1)}$ as the explanatory variables. Calculate the corresponding residuals (errors) $\hat{a}_t^{(2)}$, the error sum of squares $SSE^{(2))}$.
3. Repeat step (2) until the convergence occurs. We say that the procedure converges at step *i* (after *i* iterations) if: $\left(\left|SSE^{(i)} - SSE^{(i-1)}\right| / \left|SSE^{(i)}\right|\right) < 0.000001$.

3. Case Studies

Recently, we have been witnessing almost exponential growth in the applications of LR fitting to real time series data. Naturally, some of these applications are more successful than others. In all cases, the experiences reported are very valuable. Some of these applications have attracted the most attention among linear time series analysts.
Choosing the best model will involve obtaining some important values in fitting and forecasting phases like SSE (Sum Squared Errors), AIC (Akaike Information Criterion) which will be compared to their corresponding values when applying the new LR method.
For our illustration and comparisons, we consider two sets of real data, airline data and sunspot numbers. For each set of data, the total number of observations is denoted by *T*, the first *N* are used for fitting the models and the remaining observations *T-N (=M)* are used for predictions. The effective number of observations is denoted by *n (n=N-maximum lag)*. For each model fitted, we compute the following statistics [2]:

a) *SSE*, the sum of squared residuals: $SSE = \sum_{t=1}^{N} \left(X_t - \hat{X}_t\right)^2$,

   where, X_t *and* $\hat{X}_t$ are the true and predicted output.

b) $\hat{\sigma} = \sqrt{SSE/(n - \nu)}$, where ν is the number of parameters used.

c) the Akaike information criterion (AIC): $AIC = n \ln\left(\frac{SSE}{n}\right) + 2\nu$.

d) The Bayesian information criterion (BIC): $BIC = n\ln\left(\frac{SSE}{n}\right) + \nu + \nu\ln(n)$.

e) ARE_t, the average relative error (used in fitting phase): $ARE_t = \frac{1}{n}\sum_{t=\gamma+1}^{N}\left(\frac{\left|X_t - \hat{X}_t\right|}{\left|X_t\right|}\right)*100$,

where γ, is the maximum lag.

f) ARE_f, the average relative error (used in forecasting phase): $ARE_f = \frac{1}{M}\sum_{t=1}^{M}\left(\frac{\left|X_t - \hat{X}_t\right|}{\left|X_t\right|}\right)*100$.

g) MSE_t, the mean sum of squared residuals (used in fitting phase): $MSE_t = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \hat{X}_i\right)$.

h) MSE_f , the mean sum of squared residuals (used in forecasting phase): $MSE_f = \frac{1}{M}\sum_{i=1}^{M}\left(X_i - \hat{X}_i\right)$.

3.1 The Airline Data

The airline data comprises monthly totals of international airline passengers from January 1949 to December 1960 (see [1] and [2]). Figure 1(a) shows that the data have an upward trend together with seasonal variation whose size is roughly proportional to the local mean level (called multiplicative seasonality). The presence of seasonality was one reason for choosing ([2]) this data set. A common approach to dealing with this type of seasonality is to choose a transformation, usually logarithms, to make the seasonality additive (figure 1(b)). We denote the original data by {$\mathbf{x_t}$} and the transformed data by *{y_t} ($y_t=ln(x_t)$)*.

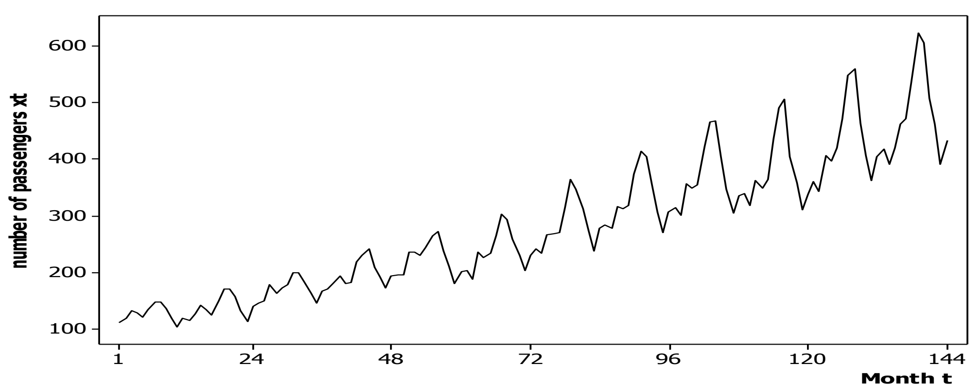

Fig 1 (a). Airline data monthly totals (in thousand) of international airline passengers from January 1949 to December 1960: natural logarithms.

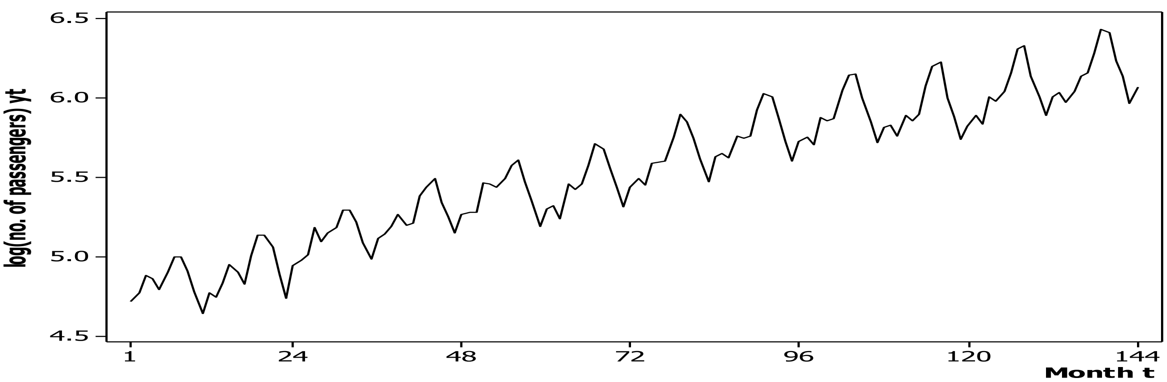

Fig 1 (a). Airline data monthly totals (in thousand) of international airline passengers from January 1949 to December 1960: natural logarithms.

i. Box-Jenkins Model:

The standard Box-Jenkins analysis ([1] and [3]) involves taking natural logarithms of the data following by seasonal and non-seasonal differencing to make the series stationary. A special type of seasonal autoregressive integrated moving average (SARIMA) model, of order $(0,1,1)\times(0,1,1)_{12}$ in the usual notation ([1]), is then fitted which has the form:

$$(1-B)(1-B^{12})y_t = (1-0.614B)(1+0.370B^{12})a_t,$$

in which we have $\hat{\theta}_1 = 0.396, \hat{\Theta}_1 = 0.614$. This model is often called the “airline model” and is used as the yardstick for future comparisons, thought other SARIMA models could be found with a similar fit and forecast accuracy. For the airline model fitted to the airline data with N=132 and M=12, the MINITAB package (release 14) gave the following values (after back-transforming all forecasts from the model for the logged data into the original units):

$$SSE = 10789, \quad \hat{\sigma} = 9.522, \quad ARE_t = 2.90278, \quad AIC = 540.35, \quad BIC = 547.91,$$
$$SS_{IS} = 4328, \quad ARE_f = 2.6516$$

ii. Classical LR Model

Using linear regression to fit a model with $y_{t-1}, y_{t-12}, y_{t-13}$ as the explanatory variables and y_t as dependent variable. The linear regression equation (for the data after scaling by dividing by 100 which changes the constant but not the other coefficients) is:

$$y_t = 0.0322 + 0.7824y_{y_{t-1}} + 1.0720y_{t-12} - 0.8394y_{t-13}$$

iii. Iterated LR Model

Using linear regression to fit a model with $y_{t-1}, y_{t-12}, y_{t-13}, e_{t-1}, e_{t-9}$ as the explanatory variables and y_t as dependent variable. The linear regression equation (for the data after scaling by dividing by 100 which changes the constant but not the other coefficients) is:

$$y_t = 0.0187 + 0.8384 y_{t-1} + 1.0804 y_{t-12} - 0.949 y_{t-13} - 0.1983 \hat{e}_{t-1} + 0.1826 \hat{e}_{t-9}$$

Table 1. Contains the results of the classical LR and iterated LR models for the airline data *N*=132, *n*=119, *M*=12 months.

	n-p		Measures of fit				Forecast accuracy	
		SSE	$\hat{\sigma}$	ARE	AIC	BIC	ARE	SS_{IS}
$y_{t-1}, y_{t-12}, y_{t-13}$	4	1.1807	0.1013	3.0765	-540.94	-525.83	3.3828	0.5073
$y_{t-1}, y_{t-12}, y_{t-13}$ e_{t-1}, e_{t-9}	6	1.1150	0.0993	3.0244	-543.76	-521.09	3.0194	0.4110

n-p: number of parameters.

From the table above we can see that ILR model is better than the LR in training and forecasting using to airline data.

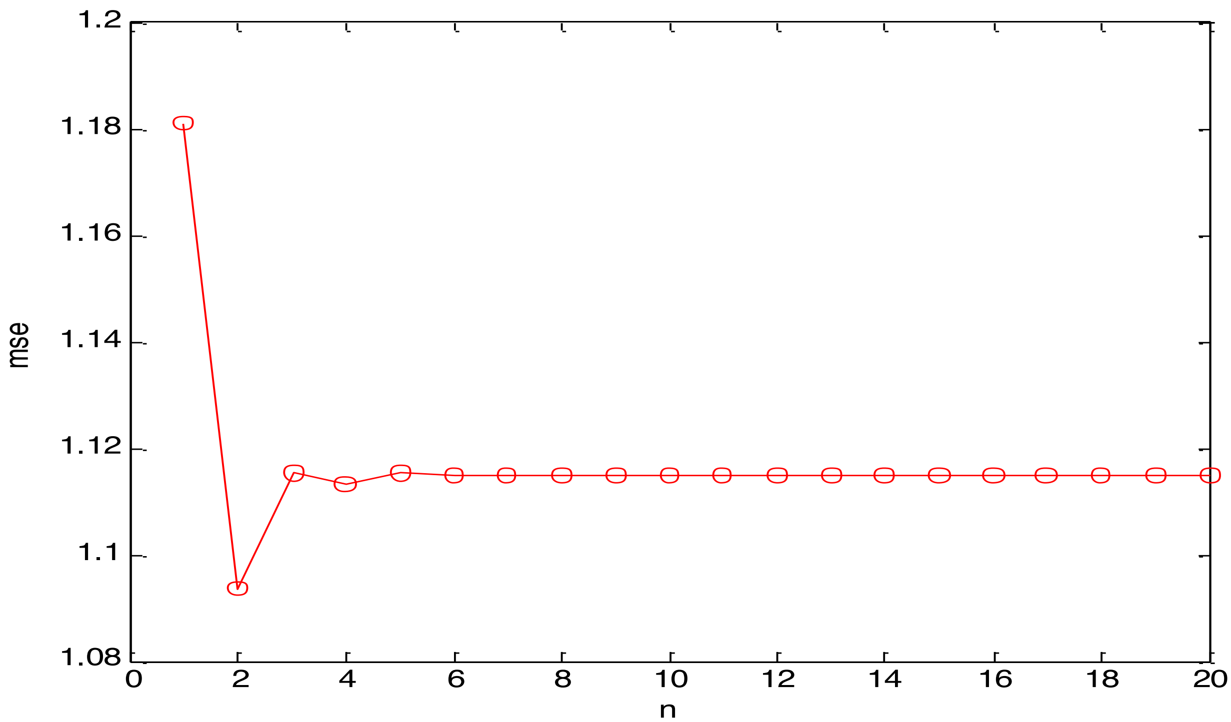

Fig 2. Convergence of MSE.

From the figure above we can see that the convergence of MSE accurate after the fourth iteration.

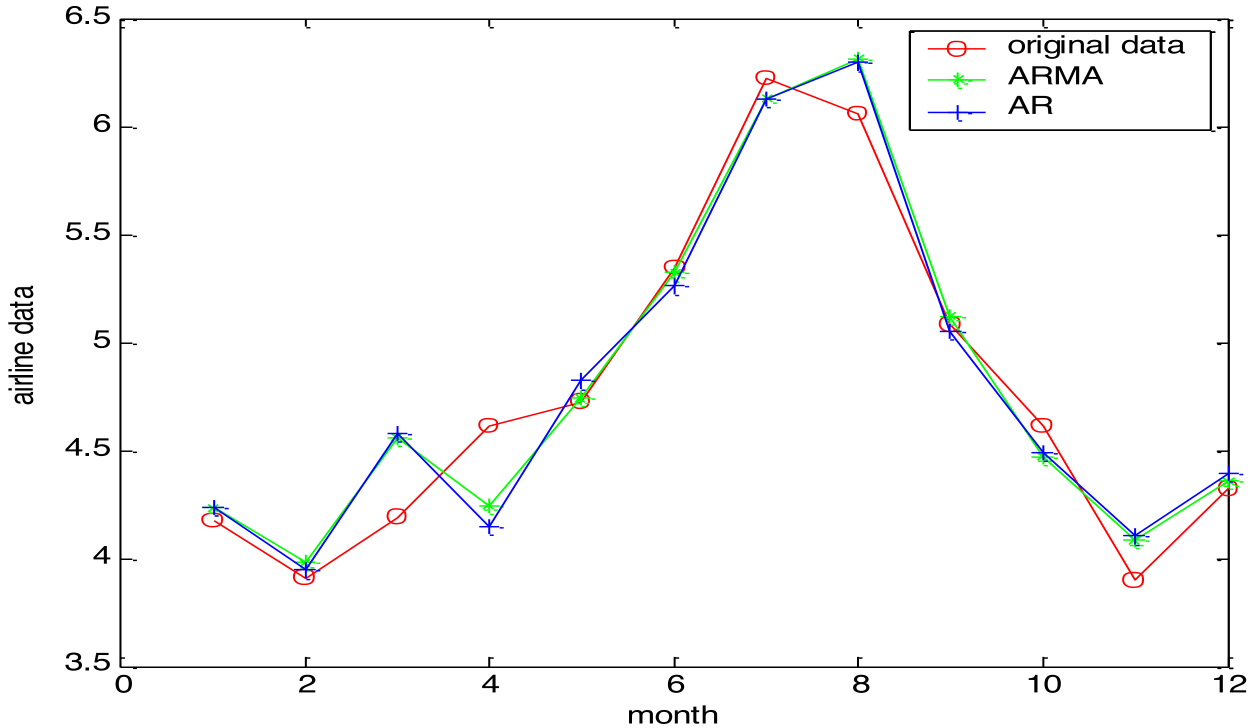

Fig 3. Forecasting of airline data.

From the figure above we can see that the ILR model is better than the LR model in forecasting.

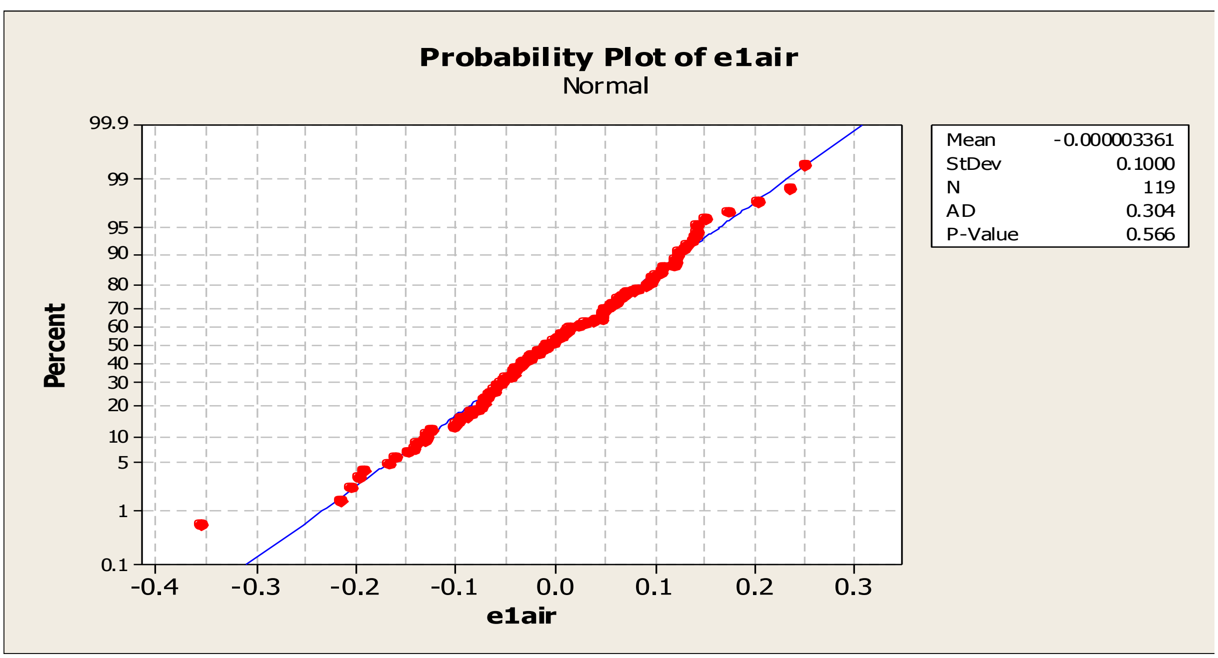

Fig 4 Normality test of the errors obtained from LR model.

From the figure above we can see that the errors obtained from LR model are normally distributed.

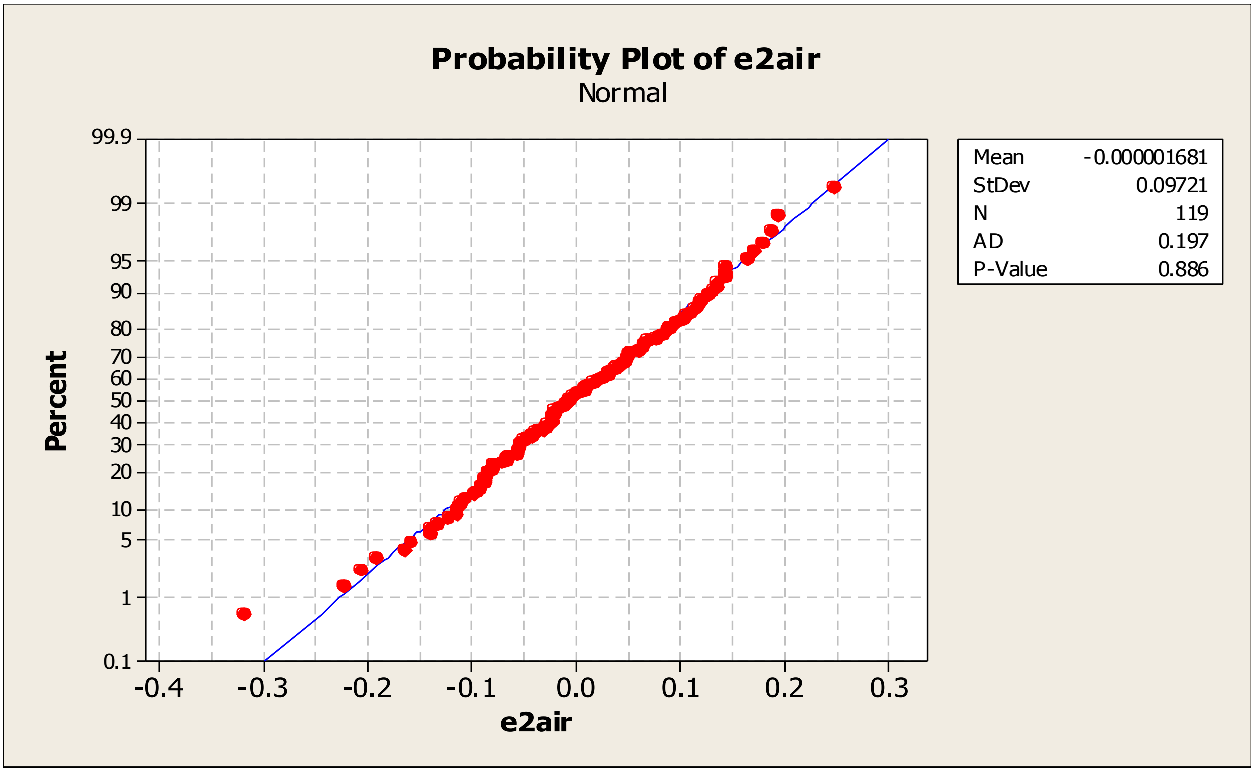

Fig 5. Normality test of the errors obtained from Iterated LR model.

From the figure above we can see that the errors obtained from ILR model are normally distributed.
From the above results, we see that the iterative LR approach is better than both the classical LR models and the Box-Jenkins model in both fitting and forecasting.

3.2 Sunspot data

The data we consider is the classic series of the Wolf yearly sunspot numbers for the years 1700-1988. This series has a certain historic interest for statisticians, see, [4] and the references in it and [5]. Scientists believe that the sunspot numbers affect the weather of the earth and hence human activities such as agriculture, telecommunications, and others. It is believed by many scientists that this series has an eleven year cycle. The plot of the series, Fig.6, and the autocorrelation function indicate that the series is stationary in the mean. However, a square root transformation is suggested to be applied for the series to be stationary also in the variance. The square root of the data is shown in Fig.7.
Subba Rao and Gabr (1984) fitted a subset AR and a subset BL models to the original annual sunspot numbers of 1700-1920, N=221, using the databank available at the University of Manchester Institute of Science & Technology, UK, at the time which contained the data only for the period 1700-1955. They used the next 35 observations (1921-1955) for predictions. The fitted subset AR model is : $X_t - 1.2496X_{t-1} + 0.551X_{t-2} - 0.145X_{t-9} = a_t$ with the mean squared error of fitting, *MSE=203.21* and mean squared error of prediction *214.1*.

The fitted subset BL model is

$$X_t - 1.5012X_{t-1} + 0.767X_{t-2} - 0.1152X_{t-9} = 6.886 - 0.1458X_{t-2}a_{t-1} + 0.006312X_{t-8}a_{t-1} - 0.007152X_{t-1}a_{t-3} - 0.006047X_{t-4}a_{t-3} + 0.003619X_{t-1}a_{t-6} + 0.004334X_{t-2}a_{t-4} + 0.001782X_{t-3}a_{t-2} + a_t$$

with the mean squared error of fitting, MSE=124.33 and mean squared error of prediction 123.77.

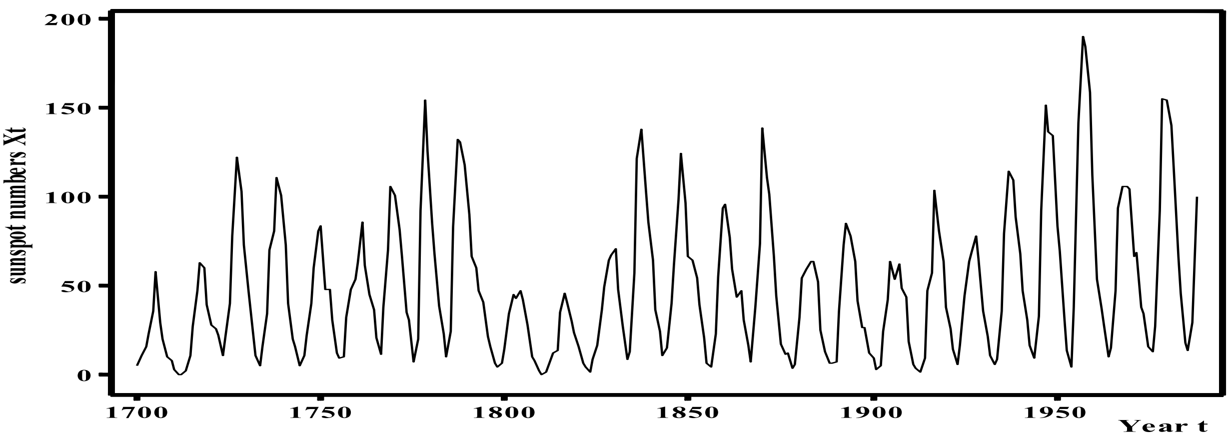

Fig 6. Walfer sunspot numbers for the years 1700-1988.

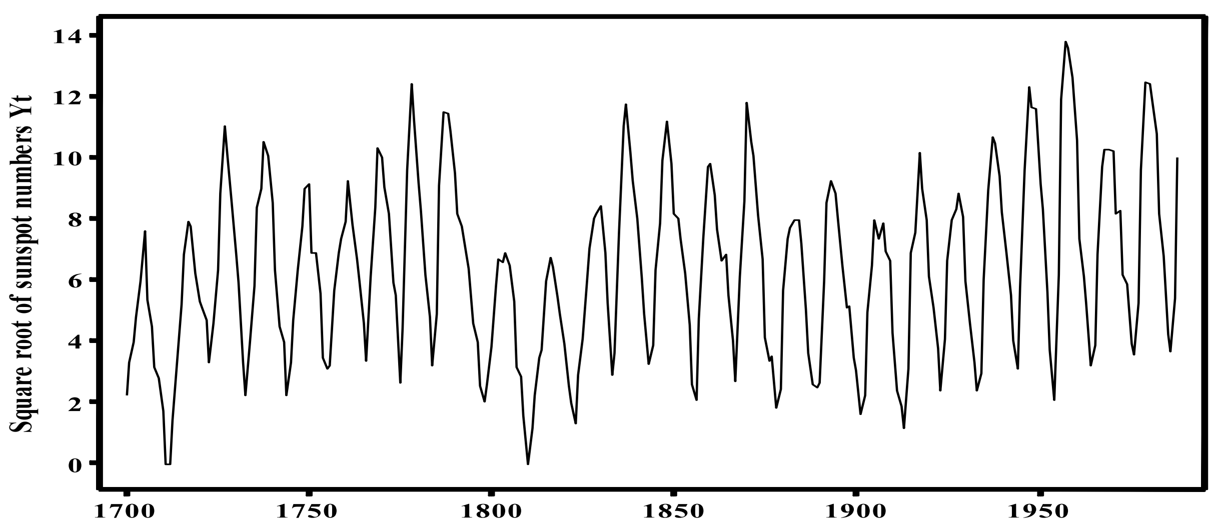

Fig 7. Square root of the sunspot numbers for the years 1700-1988.

Wei (1990) fitted the following Box-Jenkins model to the square root of annual sunspot numbers of 1700-1984 (N=285);

$$\left(1-1.17B+0.46B^2+0.21B^9\right)\left(Y_t-6.3\right)=a_t$$

with the mean squared error of fitting, *MSE=1.362***.**

Now we use the annual sunspot numbers with same fitting period, 1700-1979, the same prediction period 1980-1987 and the same square root transformation.

i. Box-Jenkins Model

Using the BJ methodology on the square root of sunspot data for the period 1700-1979, *N=280* for fitting, we found that the best ARMA model according to both, the AIC and BIC criterion's, is given by:

$$Y_t-1.233Y_{t-1}+0.4609X_{t-2}-0.2283X_{t-9}=a_t-0.1726a_{t-1}$$

with mean squared error, MSE= var(a_t) =1.14 and mean squared error of prediction (after back-transforming all forecasts accuracy from the model for the square root data into the original units) is *137.36*.

ii. Classical LR Model

The linear Regression model is fitted to the square root of sunspot data for the period 1700-1979, *N=280* and then predictors are calculated for the next 8 (M) observations. The equation obtained is:

$$Y_t=0.6733+1.2752Y_{t-1}-0.5433Y_{t-2}+0.1590Y_{t-9}$$

iii. Iterated LR Model

The equation obtained is:

$$Y_t=0.7267+1.2724Y_{t-1}-0.5464Y_{t-2}+0.1575Y_{t-9}+0.011Y_{t1}e_{t-4}$$
$$-0.0069Y_{t-4}e_{t-5}-0.1080e_{t-6}$$

Table 2 contains the results of the classical LR and iterated LR models sunspot data for the period 1700-1979, N=280 and then predictors are calculated for the next 8 (M) observations.

	n-p*	Measures of fit				Forecast accuracy	
		MSE	$\hat{\sigma}$	AIC	BIC	SS_{IS}	MSS_{IS}
$y_{t-1}, y_{t-2}, y_{t-9}$	4	1.1245	1.0706	31.87	50.49	1093.6	136.6992
$y_{t-1}, y_{t-2}, y_{t-9}, y_{t-1}e_{t-4}, y_{t-4}e_{t-5}, e_{t-6}$	7	1.1017	1.0674	34.54	65.04	787.0	98.37

*n-p: number of parameters.

From the table above we can see that ILR model is better than the LR in training and forecasting using to sunspot data.

Table 3 Contains one step ahead predictions of sunspot numbers (1980-7) with 1979 as the base year using the different models.

		Subset AR		BL	
Year	Observation	Prediction	Errors	Prediction	Errors
1980	154.7	159.8028	-5.1028	155.2547	-0.5547
1981	140.5	122.7683	17.7317	128.7530	11.7470
1982	115.9	100.2049	15.6951	106.6250	9.2750
1983	66.6	79.1174	-12.5174	75.2996	-8.6996
1984	45.9	34.2955	11.6045	30.0708	15.8292
1985	17.9	29.6227	-11.7227	29.5248	-11.6248
1986	13.4	10.3749	3.0251	10.4166	2.9834
1987	29.2	20.9005	8.2995	19.5876	9.6124
MSE			136.6992		98.3758

The table above shows that the MSE obtained by using ILR model is smule than the obtained by using LR.

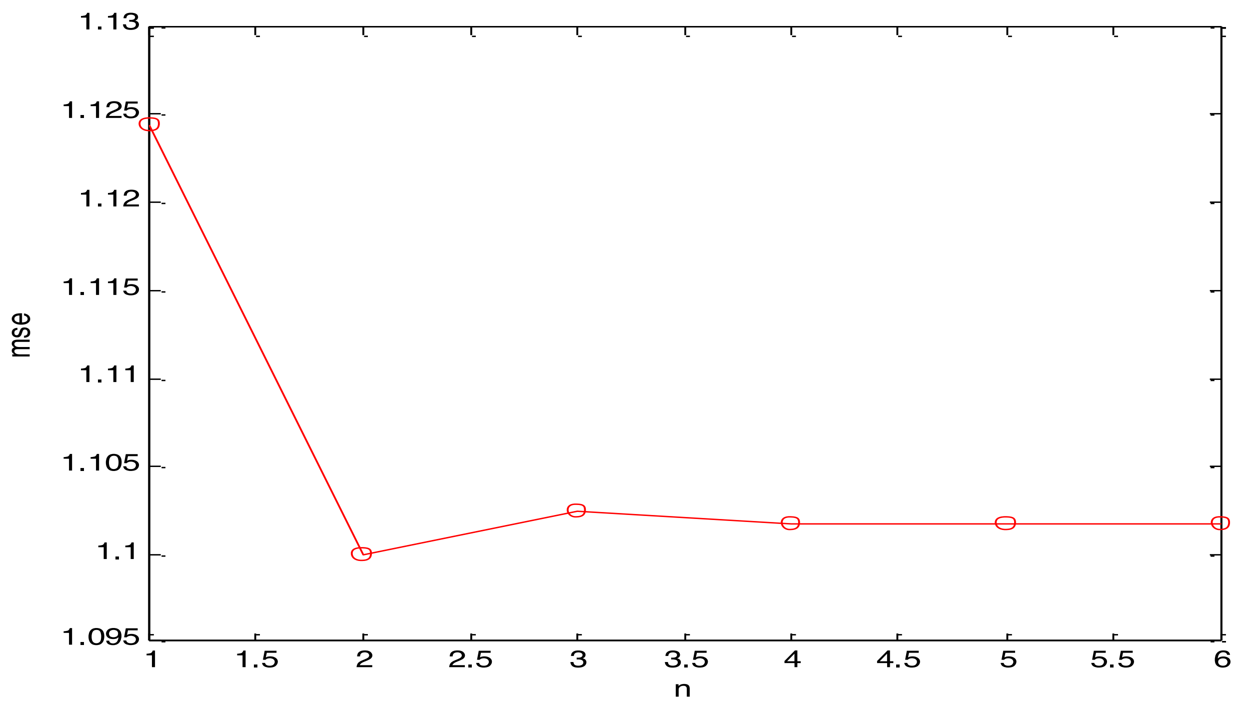

Fig 8. Convergence of MSE.

From the figure above we can see that the convergence of MSE accurate after the fourth iteration.

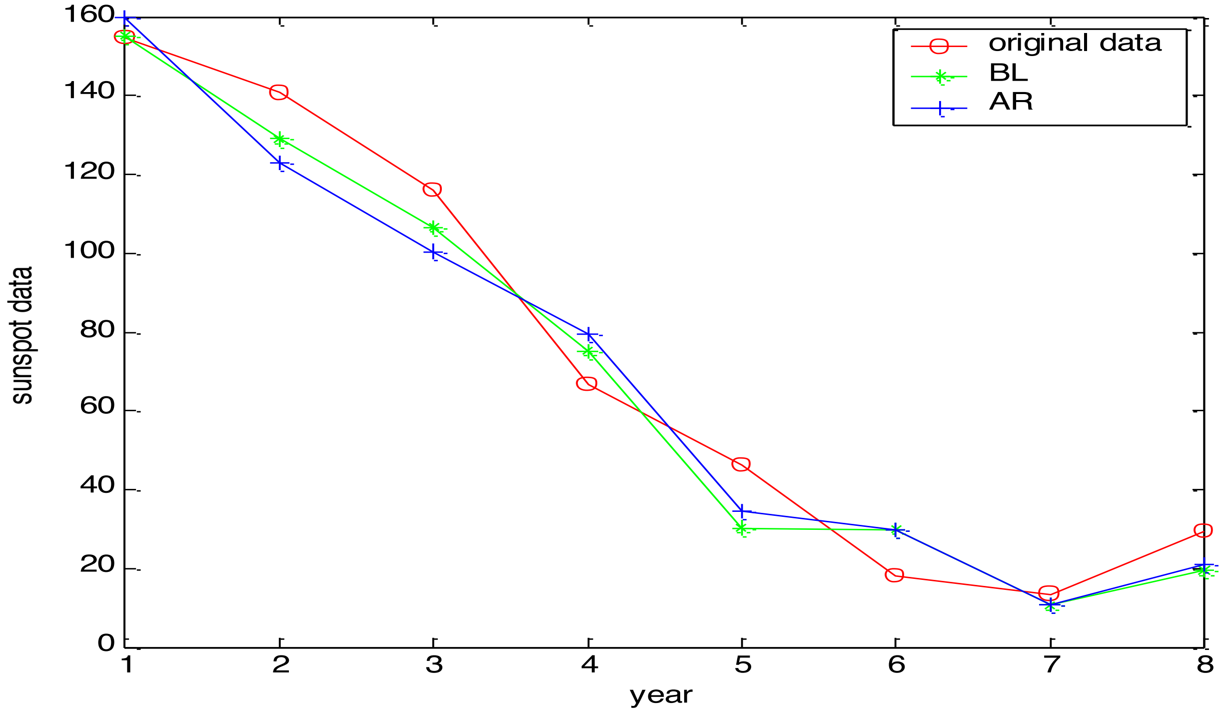

Fig 9. Forecasting sunspot data.

From the figure above we can see that the ILR model is better than the LR model in forecasting.

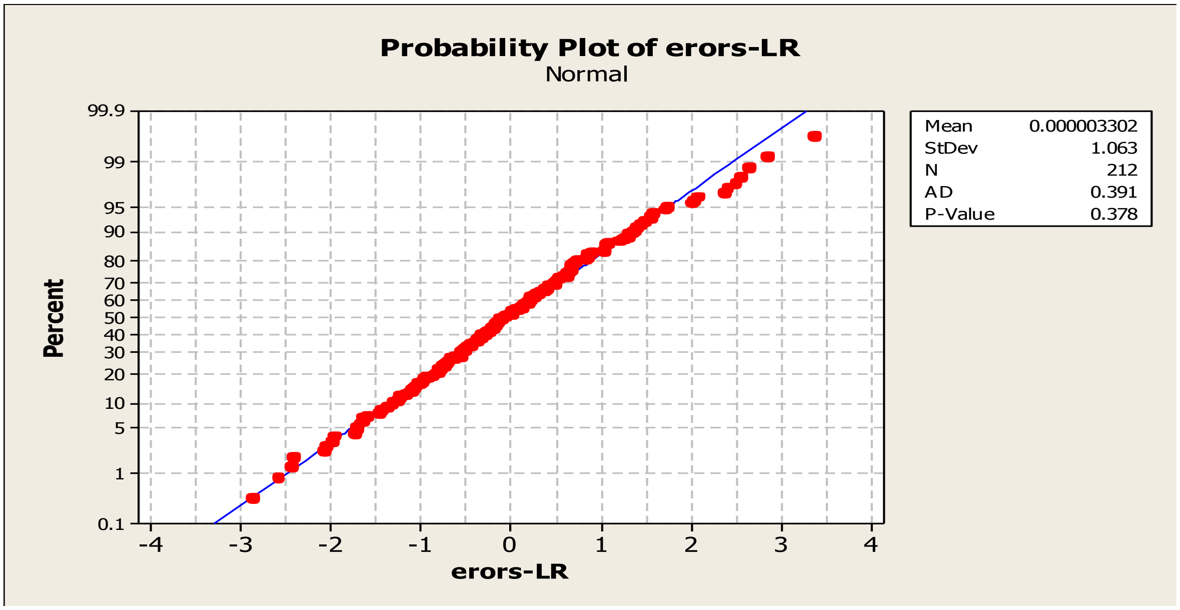

Fig 10. Normality test of the errors obtained from LR model.

From the figure above we can see that the errors obtained from LR model are normally distributed.

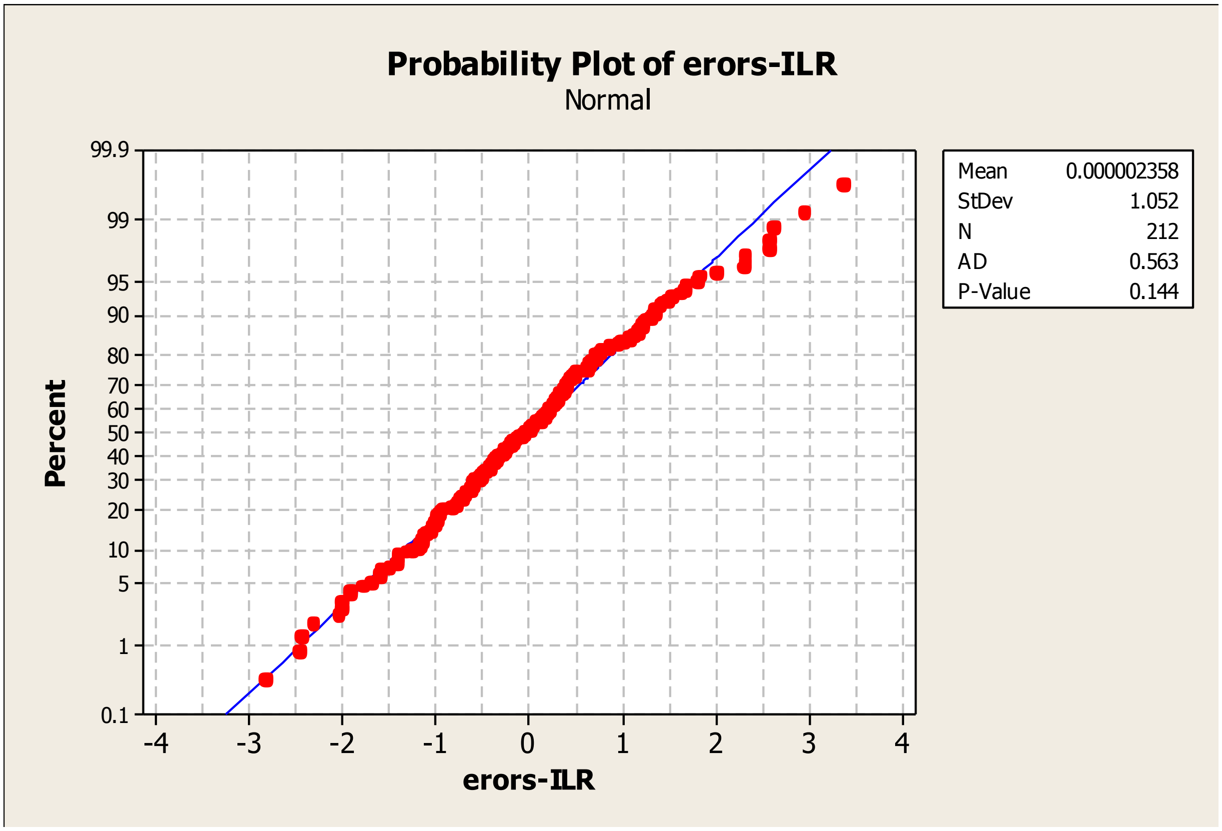

Fig 11. Normality test of the errors obtained from Iterated LR model.

From the figure above we can see that the errors obtained from ILR model are normally distributed.
From the above results, we see that the iterative LR approach is better than both the classical LR models and the Box-Jenkins model in both fitting and forecasting.

4. Conclusions

1. Fitting and forecasting using iterated LR method were better compared to the classical LR method.
2. The results obtained by the application of the iterated LR method to airline data was found to be better than that obtained by Faraway and Chatfiled (1998)
3. The results obtained by the application of the iterated LR method to the sunspot numbers were found to be better compared to that obtained by the subset AR and subset BL (Subba Rao and Gabr (1984)).
4. The errors obtained from LR and ILR are normally distributed (see Fig10 and Fig11).

References

[1]. Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis: Forecasting and Control 3rd Ed.*, Englewood Cliffs, New Jersey: Prentice Hall.
[2]. Faraway, J. and Chatfield, C. (1998). Time series Forecasting with Neural Networks: A Comparative Study Using the Airline Data. *Appl. Statist.*, 47 (2), 231-250.
[3]. Harvey, A.C. (1993). *Time Series Models 2nd Ed*. Hemel Hempstead: Harvester Wheatsheaf.
[4]. Subba Rao, T., Gabr, M.M. (1984). An Introduction to Bispectral Analysis and Bilinear Time Series Models. *Lecture Notes in Statistics*, 24, Berlin: Springer-Verlag.
[5]. Wei, W.W.S. (1990). *Time series analysis: Univariate and multivariate methods*. Addison-Wesley Publishing Company, Inc