ISSN 1022-4734

JUJSS
Vol. 35, 2019, pp. 65-75

Seasonal Adjustment Based on Neural-Network in Time Series Analysis: A Comparative Study

Atikur Rahman

Assistant Professor, Department of Statistics, Jahangirnagar University, Bangladesh

Zakir Hossain

Department of Statistics, University of Dhaka, Bangladesh

M Lutfor Rahman

Institute of Statistical Research and Training, University of Dhaka, Bangladesh

and Ajit K Majumder

Professor, Department of Statistics, Jahangirnagar University, Bangladesh

Abstract

Time series data analysis without seasonal adjustment provides misleading inferences and hence incorrect results. The linear seasonal autoregressive integrated moving average (*SARIMA*) model is a commonly used method for removing seasonality existing in the data. The non-linear Neural-Network (NN) is an alternative approach for modelling time series data in the presence of seasonality. In this study, we implemented the NN approach via back-propagation algorithm and compared its performance with *SARIMA* (0, 0, 2) (0, 2, 1)$_{12}$ model using the monthly data of paper production in Bangladesh. The NN approach was found to be better because of its smaller mean absolute deviation (MAD), mean absolute percent error (MAPE)and root mean square error (RMSE) as a further improvement of modelling time series data in the presence of seasonal variation.

Keywords: Seasonality, *SARIMA*, Neural-Network, Back-propagation algorithm, MAD, MAPE, RMSE

1. Introduction

The economic and financial time series data usually exhibit both trend and seasonal variations. In 1970, the autoregressive integrated moving average (*ARIMA*) model is introduced by Box and Jenkins for analysing non-stationary time series data with identifying the trend variation (Box and Jenkins, 1976; Stoffer and Dhumway, 2010). In time series analysis, periodicity occurs due to seasonal changes such as monthly, quarterly or weekly (He, 2004). It is practically a significant implication for modelling the seasonal variation exists in the data.

Ignoring the seasonal effect in time series data analysis provides misleading inferences and hence incorrect interpretation of results. Therefore, the seasonal adjustment should be considered for modelling the data prior to further analysis. The seasonal autoregressive integrated moving average (*SARIMA*)is an extension of *ARIMA* model that incorporates the seasonal variation for non-stationary time series data in addition to the trend effect(Box and Jenkins, 1976).Several applications of the *SARIMA* model can be found for modelling monthly time series data in the presence of seasonal variation (Aidoo, 2010; Chang et al., 2012; Saz, 2011).The monthly thunderstorm and rainfall data are analyzed in the context of forecasting by *SARIMA* models in Bangladesh (Islam et al., 2021; Moloy et al., 2018).

The *SARIMA* model is not always suitable for analysing the seasonal time series data due to its linearity (Zhang,2003).Alternatively, the nonlinear neural-network (NN)is a promising method for forecasting both seasonal and trend variations exist in time series data(Wang et al., 2011).The NN method adjust the seasonal variation directly and the deseasonalization is not required prior to the analysis (Zhang and Qi, 2005).The NN approach is computationally less time consuming that provides more accurate and competitive forecasting performance measures (Crone and Dhawan, 2007).Moreover, the NN method provides smaller prediction error than the traditional statistical models for forecasting the seasonal time series data (Hamzaçebi, 2008; Mitrea et al., 2009).

The performance of NN via back-propagation algorithm is found to be better than *SARIMA* for anlaysing data of Korea composite stock price index, monthly air passenger and quarterly energy consumption in the United States (Lee, 2007; Benkachcha et al., 2015; Camara et al., 2016). In Bangladesh, the NN approach is also used for forecasting and modelling the temperature in Sundarban region (Mandal et al., 2019).

The NN and *ARIMA* methods are compared for forecasting the remittances and jute production in Bangladesh (Hossain et al., 2017a, 2017b). The studies report that the NN approach is better than *ARIMA* model for analysing the data. In this study, we implement the NN approach and compare its performance with *SARIMA* for modelling and forecasting the monthly paper production time series data of Bangladesh in the presence of seasonal variation.

2. Methods

Seasonal autoregressive integrated moving average (*SARIMA*) model

The *SARIMA* is awidely used linear model for analyzing time series data when the seasonal fluctuation exhibits in data series. The general form of *SARIMA* model with both non-seasonal and seasonal components can be written as (Angle, 2012)

$$SARIMA\ (p, d, q) \times (P, D, Q)_s,$$

where *p, d* and *q* are orders of non-seasonal while *P, D and Q for seasonal* autoregressive, differencing and moving average, respectively. Here, *s* represents the order of seasonality e.g., s=*12* for monthly data.

3. Neural-Network (NN) approach

The general framework of nonlinear NN approach via back-propagation algorithm with several nodes is shown in Figure 1. The previous lagged observations (y_{t-1},y_{t-2}, … , y_{t-m}) are used as input nodes while the output nodes forecast the future observation (y_t). The intermediate mechanism of algorithm is performed by hidden nodes using different activation functions. The NN training is considered for determining the number of hidden nodes (Cybenko, 1989; Hornik et al., 1990).

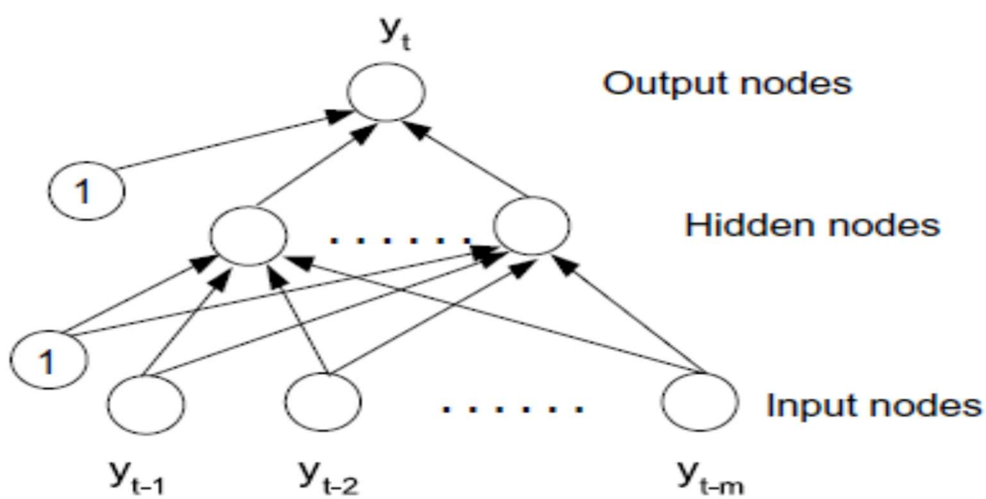

Figure 1. Basic structure of NN approach

The NN training is terminated when the gradient is below 1e-5 and the frequency of validation checks satisfies the default condition of required value 6. In addition, some other criteria are also used as the stopping parameters of NN training e. g., maximum training time, maximum number of training epochs, minimum performance value, minimum gradient magnitude and maximum number of validation checks (Beale, 2018). In this study, we implement the NN approach using the monthly data of industrial commodity (paper) production in Bangladesh. The inputs are considered as the past values of the series without removing the seasonal variation and the output is the future value. Statistical software MATLAB has been used for analyzing data in the current study.

4. Data source and screening

In this study, we use a secondary data set of monthly paper production over the period 1984 to 2013 in Bangladesh extracted from the publication "Monthly Economic Trend - Bangladesh Bank". The data are freely available online at https://www.bb.org.bd/pub/monthly/econtrds/.We first investigate the data used in this study through the time series plot (Figure 2). It is found that there exist periodic and recurrent patterns which indicate clearly non-stationary with some seasonal variations in the data.

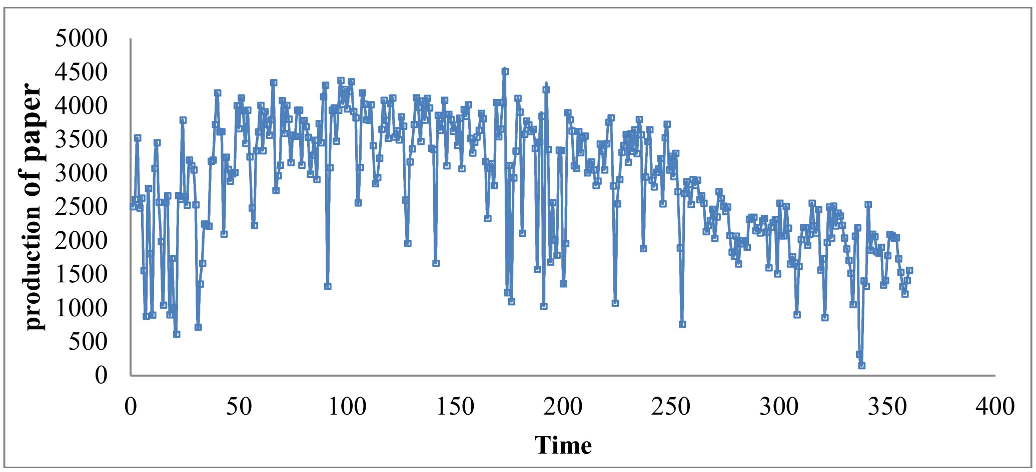

Figure 2. Time series plot of monthly production of paper

The seasonal indices are calculated using the moving average method and summarised in Table 1. It is evident that there are seasonal variations existing in the data as indices are different from 1.

Table1. Seasonal indices of monthly data of paper production in Bangladesh

Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
0.94	0.102	0.107	0.11	0.101	0.91	0.85	0.83	0.101	0.104	0.108	0.97

5. Model fitting and diagnosis

The first and second seasonal differencing are taken in order to make the data series to be stationary. The time series plot, autocorrelation function (ACF) and partial autocorrelation function (PACF) of the first and second seasonally differenced data are presented in Figure 3.

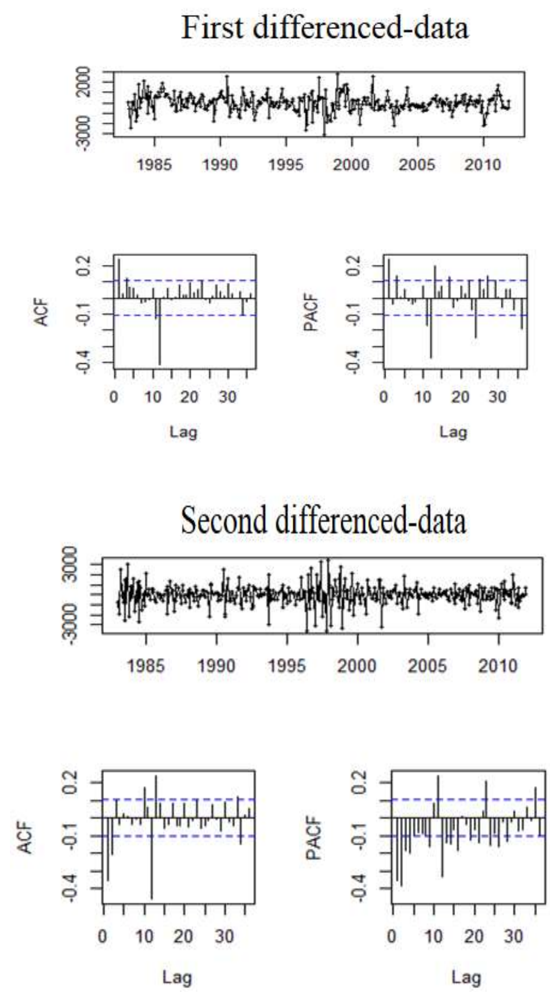

Figure 3. Time series plot, ACF and PACF of the seasonally first and second differenced monthly paper production data

Figure 3 shows that the seasonal second differenced data seems to be stationary. It is observed from the ACF plot (right panel) that there is a significant spike at seasonal lag 12, suggesting a possible seasonal moving average of order 1, MA (1). The PACF plot shows no usual significant seasonal spikes. In the non-seasonal lags, two significant spikes are in the ACF and none is significant in the PACF, suggesting a possible MA(2).Consequently, the initial analysis suggests a possible $SARIMA(0,0,2)(0,2,1)_{12}$ model for analysing data in the presence of seasonal variation used in this study.

We now fit several $SARIMA$ models of different orders including $SARIMA(0,0,2)(0,2,1)_{12}$ model and compute their Akaike information criterion (AIC) values (Akaike, 1973). The $SARIMA(0,0,2)(0,2,1)_{12}$ model is found to be the best optimal because of its minimum AIC value. This model is also investigated by some diagnostics criteria such as standardized residuals, ACF plot of residuals and p-values associated with Ljung-Box statistic (Figure 4) (Ljungand Box, 1978).

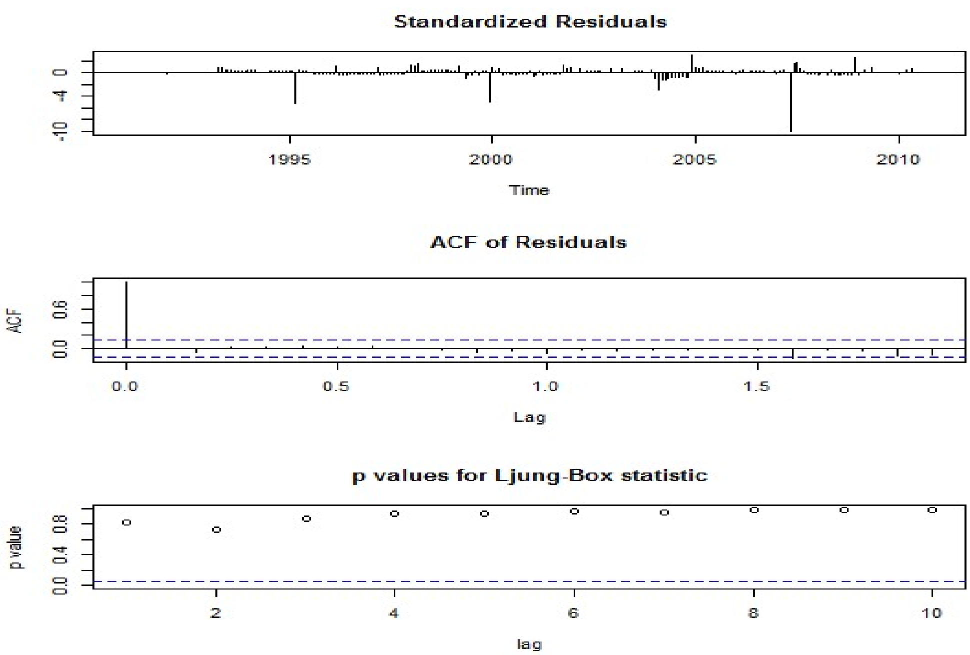

Figure 4. Diagnostic checking of $SARIMA(0,0,2)(0,2,1)_{12}$ model

The standardized residuals plot reveals that the residuals are normally distributed with zero mean and constant variance. The residuals are uncorrelated (ACF of residuals plot) and independent as all p-values of Ljung-Box test are far above the dotted line.

6. Implementing Neural-Network (NN)

We implement the NN approach using the paper production seasonal time series data used in this study. The data series is first divided in to two sets: training and test. We then select the required number of hidden layers for the NN algorithm by calculating both R^2 and adjusted R^2 values (Table 2).

Table 2. R^2 and adjusted R^2 values at different number of hidden layers in NN algorithm

Number of hidden layers	R^2	adjusted R^2
1	0.9751	0.9749
2	0.9773	0.9770
3	0.9846	0.9843
4	0.9956	0.9955
5	**0.9959**	**0.9957**
6	0.9959	0.9958

Table 2 shows that both R^2 and adjusted R^2 values increase with the increasing number of hidden layers. However, these values are found to be the same at hidden layers 5 and 6. Thus, we select the appropriate number of hidden layers 5 for implementing the NN approach to the data. The NN algorithm performs with dividing the total data set of 30 years into training (27 years) and test (3 years) sets. The architectural framework of NN training is outlined in Figure 5.

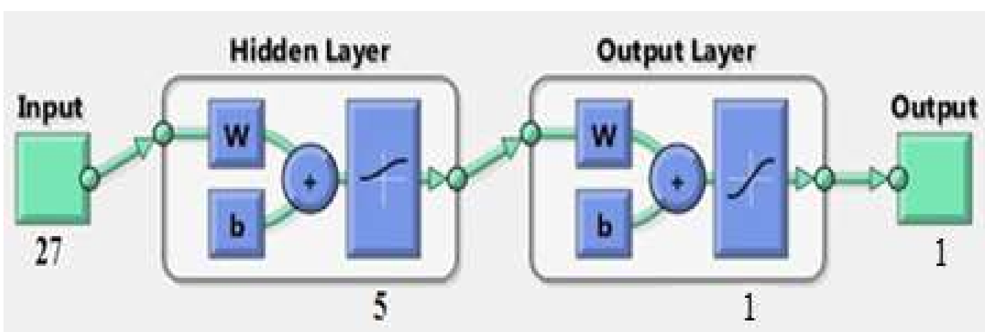

Figure 5. The NN algorithm with a training set of 27 years data and 5 hidden layers

We then present regression plots of training, validation, testing and all data in Figure 6.

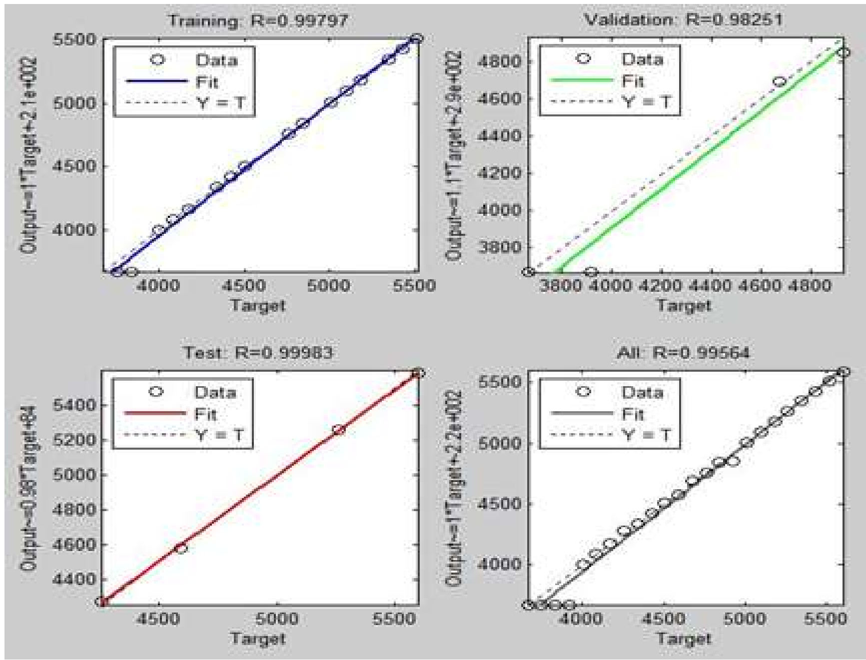

Figure 6. Regression plots of NN approach with training, validation and testing data

It is evident that the dashed line in each plot is close to the solid line which clearly indicates good fit to the data. More precisely, the training plot (upper left panel) indicates a good fit. The plots of validation (upper right panel) and test (lower left panel) also show that values of the correlation coefficient (R) are close to 1those indicate there is a very strong positive linear relationship between outputs and targets.

7. Comparison of NN and SARIMA

We now compute some performance measures: mean absolute deviation (MAD), mean absolute percent error (MAPE) and root mean squared error (RMSE) for both the NN approach and $SARIMA\ (0,0,2)(0,2,1)_{12}$ model.The results are summarized in Table 3.

Table 3. Comparison of performance measures: mean absolute deviation (MAD), mean absolute percent error (MAPE) and root mean squared error (RMSE) for the NN approach and$SARIMA\ (0,0,2)(0,2,1)_{12}$

Methods	Accuracy Measures		
	MAD	MAPE	RMSE
$SARIMA\ (0,0,2)(0,2,1)_{12}$	398.364	22.5186	605.745
NN with 5 hidden layers	**159.563**	**2.3569**	**154.230**

It is observed from Table 3that all three measures: MAD, MAPE and RMSE are found to be substantially smaller for the NN approach than the $SARIMA\ (0,0,2)(0,2,1)_{12}$. It follows that the NN approach performs better than the $SARIMA\ (0,0,2)(0,2,1)_{12}$for modelling and analysis of the seasonal time series paper production data in Bangladesh.

8. Conclusions

In this study, we use the seasonal time series paper production data of Bangladesh to compare performance of the non-linear NN approach and linear *SARIMA* model. The $SARIMA\ (0,0,2)(0,2,1)_{12}$ model is found to be the best choice for analysing the data among various linear *SARIMA* models of different orders because of its minimum AIC value. Moreover, we implement the NN approach to the data with selecting appropriate hidden layers and compare its accuracy measures with $SARIMA\ (0,0,2)(0,2,1)_{12}$ model. The MAD, MAPE and RMSE values are found to be considerably lower for the NN method than $SARIMA\ (0,0,2)(0,2,1)_{12}$.Therefore, the NN approach performs better for modelling and forecasting the monthly paper production time series data in the presence of seasonal variation.

Acknowledgments

The authors would like to acknowledge the central bank of Bangladesh for giving access to use their data.

References

Akaike, H. (1973): Information theory and an extension of the maximum likelihood principle. In Pro-ceedings of the Second International Symposium on Information Theory, Akadémiai Kiadó.

Aidoo, E. (2010): Modelling and Forecasting Inflation Rates in Ghana: An Application of SARIMA Models. Master’s Thesis. Högskolan Dalarna School of Technology and Business Studies, Sweden.

Angel (2012): *STAT510-Applied time Series Analysis*. The Pennsylvania State University, Prienter-Friendly Version.

Beale, M. H., M. T. Hagon and H. B. Demuth (2018): Neural Network Toolbox™ User's Guide, Math works Inc.

Benkachcha, S., J. Benhra and H. Hassani (2015): Seasonal Time Series Forecasting Models based on Artificial Neural Network. International Journal of Computer Applications, 116(20).

Box, G. E. P. and G. M. Jenkins (1976): Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, CA.

Camara A., W. Feixing and L. Xiuqin (2016): Energy Consumption Forecasting Using Seasonal ARIMA with Artificial Neural Networks Models. International Journal of Business and Management, 11(5).

Chang, X., M. Gao, Y. Wang and X. Hou (2012): Seasonal Autoregressive Integrated Moving Average Model for Precipitation Time Series. *Journal of Mathematics and Statistics*, 8(4); 500-505.

Crone, S. F. and R. Dhawan (2007): Forecasting Seasonal Time Series with Neural Networks: A Sensitivity Analysis of Architecture Parameters. International Joint Conference on Neural Networks, Orlando, Florida, USA.

Cybenko, G. (1989): Approximation by superpositions of a sigmoid function. Mathematics of Control Signals and Systems, 2; 303–314.

Hamzaçebi, C. (2008): Improving artificial neural networks' performance in seasonal time series forecasting. *Journal of Information Sciences*, 178;4550–4559.

He, S.Y. (2004): Applied Time Series Analysis. Peking University Press, Beijing.

Hornik, K., M. Stinchcombe and H. White (1990): Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Networks, 3; 551–560.

Hossain, M. M., F. Abdulla and A. K. Majumder (2017a): Comparing the Forecasting Performance of ARIMA and Neural Network Model by using the Remittances of Bangladesh. Jahangirnagar University Journal of Statistical Studies, 34;1-12.

Hossain, M. M., F. Abdulla and Z. Hossain (2017b): Comparison of ARIMA and Neural Network Model to Forecast the Jute Production in Bangladesh. Jahangirnagar University Journal of Science, 40(1); 11-18.

Islam M. M., M. I. Shafique, M. A. Rahman and M. L. Rahman (2021): Forecasting Thunderstorms in Bangladesh: A Time Series Modeling Approach. International Journal of Statistics and Applications, 11(1);1-7. https://doi.org/10.5923/j.statistics.20211101.01

Lee, K. (2007): Neural Network Model vs. SARIMA Model In Forecasting Korean Stock Price Index (KOSPI). Issues in Information Systems, 2. https://doi.org/10.48009/2_iis_2007_372-378

Ljung, G. M. and G. E. P. Box (1978): On a Measure of Lack of Fit in Time Series Models. *Biometrika*, 65(2):297-303.

Mandal, A. K., D. Mallick, R. Sah, A. Ali and M. A. Habib (2019): Artificial Neural Network Models for Temperature Forecasting in Sundarban Region of Bangladesh. Journal of Science and Technology, 17; 81-88.

Mitrea, C. A., C. K. M. Lee and Z. Wu (2009): A Comparison between Neural Networks and Traditional Forecasting Methods: A Case Study. International Journal of Engineering Business Management, 1(2);19- 24.

Moloy, M. D., M. Chowdhury, M. Binyamin and S. K. Mondal (2018): Using SARIMA Approach to Modeling and Forecasting Monthly Rainfall in Bangladesh. International Journal of Sciences & Applied Research, **5(5); 06-14.**

Saz, S. (2011): The Efficacy of SARIMA Models for Forecasting Inflation Rates in Developing Countries: The Case for Turkey. *International Research Journal of Finance and Economics*, 62; 112-142.

Stoffer, D. S. and R. H. Dhumway (2010): Time Series Analysis and its Application. Springer, 10; 596.

Wang, S., L. Yu, L. Tang and S. Wang (2011): A novel seasonal decomposition based least squares support vector regression ensemble learning approach for hydropower consumption forecasting in China. Energy, 36; 6542-6554.

Zhang, G.P. (2003): Time series forecasting using a hybrid ARIMA and neural network model. Neuro computing, 50; 159–175.

Zhang, G. P. and M. Qi (2005): Neural network forecasting for seasonal and trend time series. *European Journal of Operation Research*, 160; 501-514.