Forecasting Rice Production of Bangladesh Using ARIMA and Artificial Neural Network Models

Abira Sultana and Murshida Khanam

Department of Statistics, University of Dhaka, Dhaka-1000, Bangladesh

Abstract

Forecasting behavior of Econometric and Machine Learning models has recently attracted much attention in the research sector. In this study an attempt has been made to compare the forecasting behavior of Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) using univariate time series data of annual rice production (1972 to 2013) of Bangladesh. Here, suitable ARIMA has been chosen from several selected ARIMA models with the help of AIC and BIC values. A simple ANN model using backpropagation algorithm with appropriate number of nodes or neurons in a single hidden layer, adjustable threshold value and learning rate, has been constructed. Based on the RMSE, MAE and MAPE values, the results showed that the estimated error of ANN is much higher than the estimated error of chosen ARIMA. So, according to this study, it can be said that the ARIMA model is better than ANN model for forecasting the rice production in Bangladesh.

I. Introduction

Artificial neural networks modeling, a new technique for inference and forecasting attracted much attention in recent times particularly in economics and finance. Such models can simply get an explanation for complicated problems which are free from the supposition of linearity that is sometimes achieved to make the conventional methods amenable[1].

Economic data and the techniques of statistical inference are the basis of construction of the Econometric models. Also the Economic theories assist to fabricate the Econometric models[2]. Economists use Econometric models to forecast the future developments in the Economy[3].

That is why, in this paper we tried to make a comparison between the performance of forecasting behavior of a constructed neural network model and ARIMA model. For this purpose we used rice production data of Bangladesh as in an agro based country like Bangladesh, rice production is a crucial part of the national economy. The economy of Bangladesh is increasing and a number of factors influence the growth of economy. Agriculture is an important component of GDP. Around 13.75 percent of the total GDP in FY 2017-18 is being contributed by agriculture, which is also the central part of the economy[4]. Half of the agricultural GDP as well as one-sixth of the national income have been supplied by the Rice sector in Bangladesh. Almost 48% of rural employment is provided by the rice production. Moreover, it meets one-third of the total calorie demand of an average person. Rice plantation covers 75% of the cropped area and over 80% of the total irrigated area. Thus, the livelihood of the people of Bangladesh is essentially dependent on rice[5].

Bangladesh exported 50000 tons of rice to Sri Lanka in 2014 and stopped rice import until 2015[6]. About 10.59 million tons of rice was produced in Bangladesh in the year 1971 and in that time the total population of the country was 70.88 millions. Even so, the country is currently producing on 25.0 million tons rice to give food to 135 million people[5]. However Bangladesh is not steady in rice production remaining to repetition of natural calamities. Because in 2017, Vietnam exported rice to Bangladesh as the production of rice was not a sufficient amount[6]. For this reason a forecasting or future prediction of rice production is immensely needed to meet up the requirement of the people of our country.

A series of related works have been done by several researchers till date. The accuracy of forecasting Jute production (1972 to 2013) in Bangladesh of ARIMA and ANN models has been compared[7]. The result showed that ANN model performs better than ARIMA model to forecast the jute production in Bangladesh. Another study has been undertaken to fit the best ARIMA model that could be used to forecast the Rice production of Bangladesh such as in Aus, Boro, Aman season covering the entire country[8]. The study compares the original series and the forecasted series. The analysis showed that ARIMA model gives good forecasting for short term analysis. The performance of ANN with exponential smoothing and ARIMA models in forecasting rice exports from Thailand has been compared in other study[9]. This study employed the various accuracy measures RMSE, MSE, MAE and MAPE to determine that the models can produce reasonable results to predict future.

This work differs the other mentioned researches in many ways. In this paper, the forecasting behavior of selected ARIMA and ANN models is compared using univariate time series data of annual rice production (1972 to 2013) and the best of these two models considering the values of RMSE, MAPE and MAE is selected for further forecasting of Rice production. In section 2 we discussed about the data and variables that have been used in this work. Section 3 covers a brief discussion regarding the models and the theories that have been used. After that section 4 contains all the relevant tables and figures that contains the results of this analysis. Finally section 4 discusses the overall findings of the study.

---

*Author for correspondence. e-mail: *murshida@du.ac.bd*

II. Models and Theories

The two models that have been considered in the present study are ARIMA and ANN Models.

ARIMA Model

ARIMA model popularly known as the Box-Jenkins (BJ) methodology which is the most general class of models for forecasting a time series data. The data can be made stationary by differencing or taking logarithm. A series that needs to be differenced to make stationary is called "integrated"[10].

A series is called ARIMA (p,d,q) if we need to difference the series d times to make it stationary before applying ARMA (p,q) model. Here, p denotes the number of autoregressive terms and q denotes the number of moving average terms[11].

ANN Model

An ANN or connectionist system is computing system vaguely inspired by the biological neural networks that constitute animal brains. Such system learns to perform tasks by considering examples, generally without being programmed with any task specific rules[12].

An ANN is reinforcement on a group of linked units or nodes called artificial neurons like the neurons in a biological brain. Every association is like the synapse in a biological brain. This transmits a sign from one artificial neuron to another[12].

Components of ANN

Several components are involved to construct an ANN. Some are: Neurons, Connections and Weights, Propagation function, Learning rule and Learning algorithm[13, 14].

Stationarity Test for the Data Set

Before starting the analysis of the study, we have tested the stationarity of the data set. The overall stationarity test is conducted using two tests, Graphical Analysis and Augmented Dickey-Fuller (ADF) test[10].

Evaluation of Forecasting Method

There exist many statistics to evaluate the forecast error of any Time Series or Econometric model. Let us suppose that, Y_t be the actual observation and f_t is the fitted value. Then the following statistical measures can be considered to evaluate the forecasting .

Root Mean Square Error (RMSE) of n predicted values can be defined as,

$$\text{RMSE} = \sqrt{E\big((f_t - Y_t)^2\big)} \qquad (1)$$

Mean Absolute Error (MAE) of predicted values are given by,

$$\text{MAE} = \frac{1}{n}\sum_{t=1}^{n}|f_t - Y_t| = \frac{1}{n}\sum_{t=1}^{n}|e_t| \qquad (2)$$

Mean Absolute Percentage Error (MAPE) which measures the goodness of fit of a model to a data set is given by,

$$\text{MAPE} = \frac{\frac{1}{n}\sum_{t=1}^{n}|f_t - Y_t|}{Y_t} * 100 \qquad (3)$$

III. Data and Variable

In this study time series annual data of Rice production of Bangladesh ('000 tons) from 1972 to 2013 was obtained from Food and Agricultural Organization of United Nations (FAO), Bangladesh website[15] with proper permission. The collected data have been cross checked with the data from Bangladesh Statistical Year Book[16] published by Bangladesh Bureau of Statistics (BBS).

According to the requirements of the study and availability of the data, one single variable has been considered for this research work and that is the annual Rice Production of Bangladesh.

IV. Results

To check the stationarity of the data set, we conducted two relevant tests called Graphical analysis and Augmented Dickey- Fuller (ADF) test. Both of the test results have showed that the data set for the production of rice of Bangladesh has become stationary after taking the first difference.

In order to identify the tentative ARIMA model for the data set, after identifying the order of difference we need to plot the Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) to get the value of p and q. The ACF and PACF plots have come up with five possible models MA(1), MA(2), ARMA(1,1), ARMA(1,2), ARMA(2,2). Among which ARIMA (1, 1, 1) has been selected as the best model for further comparison considering the lower AIC and BIC values. Table 1 contains the AIC and BIC values of the five selected models.

Table 1. AIC and BIC values of selected ARIMA Models.

Criteria	ARIMA (0,1,1)	ARIMA (0,1,2)	ARIMA (1,1,1)	ARIMA (1,1,2)	ARIMA (2,1,2)
AIC	579.91	581.23	**577.81**	578.87	580.85
BIC	582.96	583.81	**582.39**	584.97	588.48

Here minimum values of AIC, BIC indicate that ARIMA (1,1,1) as the best selected model to forecast the rice production of Bangladesh.

Table 2. Estimated values of ARIMA (1,1,1).

	AR1	MA1	σ^2 (estimated)	log-likelihood	AIC
ARIMA(1,1,1)	0.9997	-0.9896	1121147	-285.91	577.81
S.E.	0.0021	0.0376			

Table 2 shows the estimated values of ARIMA (1,1,1). Here we are going to observe the forecasting behavior of our fitted model ARIMA (1,1,1)

Table 2 along with the graph (fig1) shows the forecasted value obtained by the ARIMA (1,1,1) model.

Table 3. The forecasted and actual value of Rice Production of Bangladesh using ARIMA (1,1,1) model.

Year	Actual Value	Forecasted Value	Error of ARIMA (1,1,1) Model	MAPE
2007	28801.73	27634.92	1166.81	
2008	31176.91	28074.13	3102.78	
2009	32112.05	28513.22	3598.83	
2010	33390.82	28952.18	4438.64	10.68
2011	33768.21	29391.02	4377.19	
2012	33681.50	29829.74	3851.76	
2013	34350.50	30268.33	4082.17	

The MAPE value bigger than 10 (10.68) indicates good forecasting[17].
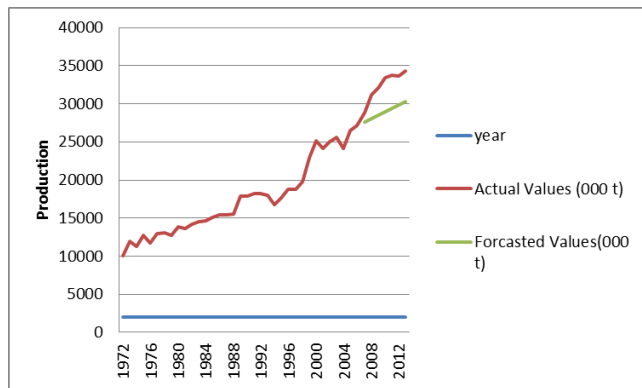

Fig. 1. Actual value along with the Forecasted value using ARIMA (1,1,1).

Table 2 and figure1 represents the actual values from 1972 to 2006 along with the predicted values from 2007 to 2013 and the error of measuring these values.

The data set has been divided into two sets, training set and test set, to identify the suitable ANN model. The training data set has been used to fit the model and the test data set has been used to evaluate the final model that was fitted on the training dataset[18]. To construct the proper ANN model the backpropagation algorithm has been used along with one hidden layer and three nodes or neurons. We have considered the learning rate 0.10 and the threshold point 0.01 respectively to get the desire ANN.

The forecasted values along with the actual values and the errors of measuring these values are given in table 3 and graphically represented in figure 2

Table 4. Forecasted value for the rice production of Bangladesh using ANN Model.

Year	Actual Value	Forecasted value	Error of ANN	MAPE
2007	28801.73	17692.25	11109.48	
2008	31176.91	17692.25	13484.66	
2009	32112.05	17692.25	14419.80	
2010	33390.82	17692.25	15698.57	83.52
2011	33768.21	17692.25	16075.96	
2012	33681.50	17692.25	15989.25	
2013	34350.50	17692.25	16658.25	

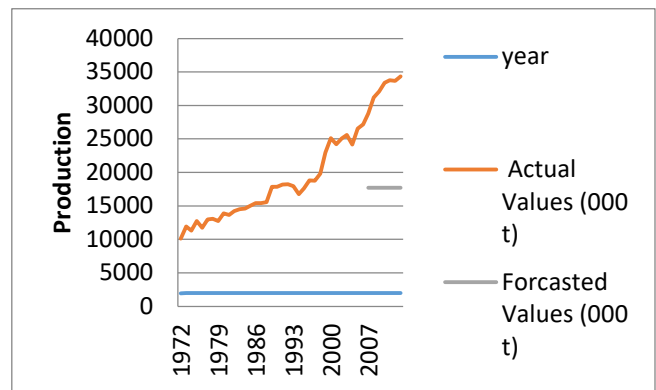The MAPE value bigger than 50 (83.52) indicates inaccurate forecasting[17].

Fig. 2. Actual value along with the forecasted value using ANN.

Table 2 and figure 1 represents the actual values from 1972 to 2006 along with the forecasted values from 2007 to 2013 and the error of measuring these values using ANN Model. Here, due to the selection process of ANN Model, which includes, number of nodes, learning rate and threshold value, the forecasted values are all same. One reason for this type of forecasted values is the way of accepting the data by the ANN Model we established here.

Comparison of the Approaches: We have compared the both approaches, ARIMA and ANN regarding their forecasting behavior using three accuracy measures such as RMSE, MAE and MAPE.

Table 5. Forecasting Accuracy indices for Selected ARIMA and ANN Model.

Accuracy Measure	ARIMA	ANN
RMSE	3670.38	14886.34
MAE	3516.88	14776.57
MAPE	10.68	83.52

Considering table 4 and table 5 we can draw the conclusion that ARIMA model performed better than ANN Model.

Forecasting: Since ARIMA model performed better than ANN using the type of given data set, we reconsidered the ARIMA Model to perform further forecasting of recent years and upcoming year.

Table 6. Final Forecasting of Recent Years and Future Years Using ARIMA(1,1,1) Model.

Year	Forecasted value	Lower 80	Higher 80	Lower 95	Higher 95
2014	30706.81	26496.56	34917.06	24267.78	37145.83
2015	31145.16	26629.79	35660.52	24239.50	38050.81
2016	31583.39	26771.83	36394.95	24224.74	38942.03
2017	32021.49	26921.07	37121.91	24221.07	39821.91
2018	32459.48	27076.32	37842.64	24226.64	40692.31
2019	32897.34	27236.62	38558.06	24240.01	41554.66
2020	33335.08	27401.22	39268.93	24260.03	42410.13
2021	33772.69	27569.52	39975.87	24285.75	43259.63
2022	34210.19	27741.00	40679.37	24316.42	44103.95
2023	34647.56	27915.27	41379.85	24351.40	44943.72

The above table contains the forecasted values along with the higher and lower 80 and 95 level values for the year 2014 to 2023.
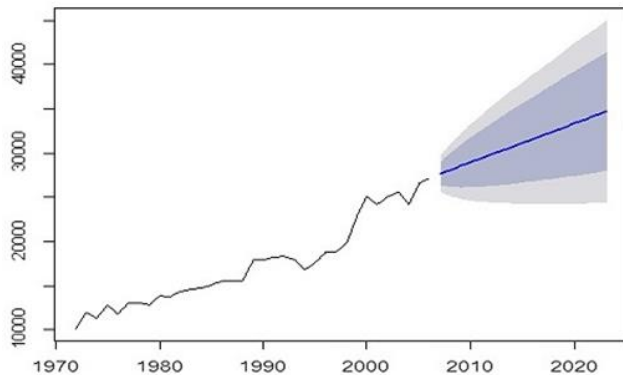

Fig. 3. Forecasts for ARIMA (1,1,1).

The above plot contains the forecasted values and patterns of the future rice production of Bangladesh.

V. Discussion

The basic motives of this study were to compare the forecasting behavior of an econometric model and a model of machine learning and apply the best selected model using the given data set to conduct a forecasting for the future rice production of Bangladesh.
According to the estimation process we employed two tests to check the stationarity of the only one variable, rice production. Since the data are non-stationary, we have taken first differences to make the data stationary.

In search for the suitable ARIMA model to modeling and forecasting the rice production, ACF and PACF plots recommended five possible ARIMA models, among them ARIMA (1,1,1) has been selected according to the AIC and BIC values.

To identify the suitable ANN model we have divided the data set into training data set and test data set. Where training set has been used to train up the model for modeling purpose and the test set has met the forecasting purpose. Backpropagation algorithm has been used to construct the simple ANN model with one hidden layer and three nodes or neurons. Moreover, 0.10 and 0.01 have been adapted respectively as the learning rate and threshold point to get the desire form of ANN model.

After that, to determine the best forecasting model we have obtained three forecasting evaluation techniques namely RMSE, MAE and MAPE. The results based on these three measures of error have showed that the performance of ARIMA (1,1,1) model is better than ANN model with single hidden layer and three neurons using this time series data set of rice production (table 5).

Finally, with the help of preferred ARIMA (1,1,1) model, we have estimated the future production of rice for some recent years and future years and presented the results in table 6 and represented graphically in figure 3.

In conclusion, it can be said that the ARIMA (1, 1, 1) model is better than ANN model regarding the comparison between them to forecast the rice production in Bangladesh.

The other finding from the forecasting analysis is that the rice production will be increased in the upcoming years.

VI. Conclusion

According to this study, ARIMA model has been recommended to forecast the time series annual data. Because in this study, ARIMA model gives better result comparing with the ANN model for forecasting the Rice production of Bangladesh.

In this paper, univariate data has been considered to compare ARIMA model and ANN model. It can be suggested to do further research considering bivariate data and make a comparison between ARIMA model and ANN model. It can be recommended to make comparison with other model such as VAR model, VECM model, Holt's smoothing techniques etc. Moreover, this work is a collaboration of artificial intelligence and econometric models. Works like this need to be executed more in the field of several researches.

References

1. Moshiri, S. and Cameron, N., 2000. Neural Network Versus Econometric Models in Forecasting Inflation. *Journal of Forecasting*, *19*, 201-217.

2. Buying Energy and Nonfuel Minerals. https://www. sciencedirect.com/topics/economics-econometrics-and-finance/econometric-model

3. Forecasting and Econometric Models. https://www. econlib.org/library/Enc/ForecastingandEconometricMo dels.html

4. National Accounts Statistics, Provisional Estimates of GDP, 2017-18 and Final Estimates of GDP, 2016-17.

5. Bangladesh Rice Knowledge Bank, http://www. knowledgebank-brri.org/riceinban.php

6. Structural changes in Asian rice trade: Lessons for Bangladesh, https://www.thedailystar.net/business/ structural-changes-asian-rice-trade-lessons-bangladesh-1538269.

7. Hossain, M.M., A. Faruq, and Hossain, Z., 2017. Comparison of ARIMA and Neural Networks Model to Forecast the Jute Production in Bangladesh. *Jahangirnagar University Journal of Science, 40(1), 11-18*.

8. Hamjah, M.A., 2014. Rice Production in Bangladesh: An Application of Box-Jenkins ARIMA Model. Mathematical theory and Modeling, *ISSN, 4(4), 2224-5804*.

9. Henry, C. Co. and Boosarawongse, R., 2007. Forecasting Thailand's Rice Export: Statistical Techniques Vs. Artificial Neural Networks. *Computers and Industrial Engineering, 53, 610-627*.

10. Gujarati, D. N. and Porter, D. C. 2009. Basic Econometrics. 5th edi. New York: McGraw Hill Book Co.

11. James D. Hamilton., 1994. Time Series Analysis.

12. A Gentle Introduction to Neural Networks series- Part 1. https://towardsdatascience.com/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc.

13. Activation Functions in Neural Networks towards Data Science. https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6.

14. Types of Activation Functions in Neural Networks and Rationale behind it. https://www.analyticsindiamag. com/most-common-activation-functions-in-neural-networks-and-rationale-behind-it/.

15. Ricepedia; The online authority on rice. http://ricepedia.org./index.php/bangladesh

16. Bangladesh Bureau of Statistics (BBS) 1978 to 2013.

17. MAPE, Interpretation, https://www.researchgate.net /figure/nterpretation-of-typical-MAPE-values_tbl1_257812432

18. About Train, Validation and Test Sets in Machine Learning. https://towardsdatascience.com/train-validation-and-test-sets-72cb40cba9e7.