RESEARCH ARTICLE Agricultural Science Digest, Volume 40 Issue 4: 404-407 (December 2020)

ARIMA Modelling for Forecasting of Rice Production: A Case Study of India

Sunali Mahajan, Manish Sharma, Amit Gupta

10.18805/ag.D-5029

ABSTRACT

Background: India has been the top exporter in global rice trade, accounting for more than 20% of the export in the last four years. With the increasing demand for rice globally, exports are increasing from India. This factor is helping to enhance the production capacities of rice in India. The rice production has increased by 3.5 times in the last 60 years. So, forecasting of any agricultural produce play a major role in optimal decision formulae for government and agricultural sector in India. This study aimed to increase the production of rice for next four years, by using an appropriate model.

Methods: To develop an appropriate model for rice production, time series data from 1950-51 to 2016-17 have been used, through the Box Jenkins auto regressive integrated moving average (ARIMA) methodology.

Result: The diagnostic tests showed that ARIMA (0 2 2) model is appropriate for forecasting on the basis of the significance of the model, parameters, Akaike Information Criteria (AIC), Schwartz Bayesian Information Criterion (SBIC) and R^2. The forecasted results suggested that there are expectations of increasing the rice production in India. The proposed forecasting model will help us to estimate the rice production for future and found that rice production of India would become 110.64 MT in 2020-21. This study clearly indicates that forecasting is very useful for policy makers to anticipate future needs of rice and its export. Moreover, this would also prove helpful in shaping the national policy of economic growth and self-sufficiency in food grain.

Key words: Akaike Information Criteria, ARIMA, Forecast, Rice production, Schwartz Bayesian Information Criterion.

INTRODUCTION

After independence, agricultural production has been brought up by bringing additional area under cultivation, extension of irrigation facilities, use of better seeds and modern techniques, water management and plant protection but there is still a large disparity between the per capita income in the farm sector and the non-farm sector. Therefore, it is essential to deal with those issues which impact the food grains production mainly comprised of rice, wheat and maize. As we talk about rice, rice is one of the main staple foods of the large portion of the world's population. It plays an important role to meet the desperate needs of food demand both in urban and rural areas. A lot of efforts have been made by the government of India to increase the productivity of rice to meet food demand and boost export. In 2009-10, total rice production in India amounted to 89.13MT which was much less than the production104.32 MT in 2011-2012 crop years (July-June). In India, rice is grown in 43.86 Mha, the production level is 104.80 MT and the productivity is about 2390 kg/ha (Agricultural Statistics at a Glance-2015). The area under rice crop was 30.81 Mha in 1950-51 which has increased to 43.86 Mha during 2014-15 which is nearly 142 per cent higher. The rice production has registered an appreciable increase from 20.58 MT in 1950-51 to 104.86 MT during 2014-15, which is nearly 5 times. The scenario of first five highest producing countries of rice in the world is given below:

Division of Statistics and Computer Science, Sher-e-Kashmir University of Agricultural Sciences and Technology, Jammu-180 009, Jammu and Kashmir, India.

Corresponding Author: Sunali Mahajan, Division of Statistics and Computer Science, Sher-e-Kashmir University of Agricultural Sciences and Technology, Jammu-180 009, Jammu and Kashmir, India. Email: sunali12mahajan@gmail.com

How to cite this article: Mahajan, S., Sharma, M. and Gupta, A. (2020). ARIMA Modelling for Forecasting of Rice Production: A Case Study of India. Agricultural Science Digest. 40(4): 404-407.

Submitted: 29-08-2019 **Accepted:** 23-05-2020 **Published:** 29-09-2020

(Area: million ha and Production: MT)

Country	Area	Production
China	30.58	205.21
India	43.86	104.80
Indonesia	13.84	71.29
Bangladesh	11.78	51.50
Vietnam	7.90	44.04
Total for world	107.96	476.84

Source: Agricultural statistics at a Glance 2015.

There are improved technologies and various interventions which could be adapted to increase the productivity in the country. Cultivation of hybrid rice has

potential to increase the productivity and needs to be promoted. Even in J&K state, the foodgrains production has more than trebled, since the year 1950-51. Also, according to National Food Security Act (NFSA) there is a need of 9.50 lakh MT of food grains instead of 7.51 lakh MT to fulfill the requirement of J&K population (Daily Excelsior Report, July 8, 2015). Rising population and per capita income are pushing up the demand, which needs to be met through enhanced productivity per unit area, input and time, the annual increase in demand in India. On the basis of these issues, the study has been taken for the forecasting of rice production in India with the objectives, to develop appropriate ARIMA model for the time series of rice production in India and to make forecasts with appropriate prediction interval. Further, it allows the policy makers to predict the future shortage of food grain in the country and could take necessary measures in future to avoid food shortage.

MATERIALS AND METHODS

This study used the annual time series data of rice production from 1950-51 to 2016-17. An appropriate modelling for forecasting of rice production has been used. Among all the techniques, Box and Jenkins methodology or ARIMA modelling (originally studied by George Box and Gwilym Jenkins during 1968) is commonly used for forecasting purpose and has been introduced by Box and Jenkins (1976). In 1973, Granger and Newbold, told that the ARIMA modelling is not good for lead times or for seasonal series. This technique combines the two specifications into one equation *i.e*. Autoregressive Process (AR) and Moving Average Process (MA). The general structure of ARIMA model is; ARIMA (p, d, q), where 'p' and 'q' are the order of the autoregressive and moving average process respectively while 'd' is the order of differencing. The mathematical form of ARIMA (p, d, q) model is:

$$Y_t = C + (F_1 Y_{t-1} + \cdots + F_p Y_{t-p}) - (\theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}) + e_t.$$

Where,

C is the constant, Y_{t-1} ,...., Y_{t-p} are lags, θ_1 are the parameters of the equation and e_t is the random error.

However, the AR model of order p is

$$Y_t = C + (F_1 Y_{t-1} + \cdots + F_p Y_{t-p}) + e_t.$$

Similarly, the MA structure of order q is

$$Y_t = e_t - (\theta_1 e_{t-1} + \cdots + \theta_q e_{t-q})$$

The forecasting of ARIMA model involved different steps such as (i) identification of the model or specification of the model, (ii) estimation and (iii) diagnostic. The foremost step in the process of modeling is to check for the stationarity of the series by using appropriate tests like Augmented Dickey-Fuller (Dickey and Fuller, 1981 and Fuller, 1996), as the estimation procedures are available only for stationary series. If the original series is non stationary then first of all it should be made stationary through appropriate differencing because an ARIMA model is applicable only to a stationary time series.

In the second step after identification of the parameters (p, d, q) the series is subjected to fitting of the appropriate ARIMA (p, d, q) model. The procedure for fitting the model involves transforming the series through appropriate differencing, if non-stationary and then subjecting the differenced series to fitting. Choice of parameters is on the basis of significant ACFs and PACFs. In the estimation stage, various parameters of the model has been evaluated.

At last, diagnostic checking of the data has been done through (i) Akaike Information Criteria (Akaike, 1979): AIC can be written as

$$AIC = -2 \log L + 2n,$$

Where,

L is the likelihood function and n is the number of hyper parameters estimated from the model. (ii) Schwarz-Bayesian Information Criteria (Schwarz, 1978): As an alternative to AIC, sometimes SBIC is also used which is given by

$$SBIC = -2 \log L + n \log T,$$

Where,

T is total number of observations. Lower the value of these statistics better is the fitted model and (iii) R^2.

Once the appropriate ARIMA model has been fitted, one can examine the goodness of fit by means of plotting the ACF and PACF of residuals of the fitted model and come up with forecasting.

RESULTS AND DISCUSSION

The annual time series data of rice production for the last sixty five years were used in the prediction model. For this purpose, both graphics and empirical methods were used. Fig 1 showed the line chart of rice production in India which reveals that there are many ups and downs in the data. Long term increasing pattern of data indicates that it is non stationary with mean (59.36), standard deviation (25.46) and trend with ADF value (-6.05). Further, ADF test have showed being the non-significant accept the null hypotheses of non-stationarity *i.e*., data are non-stationary and the p- values of Ljung-Box Q values are significant which indicates that the residuals are dependent whereas, Fig 2 represents the correlogram

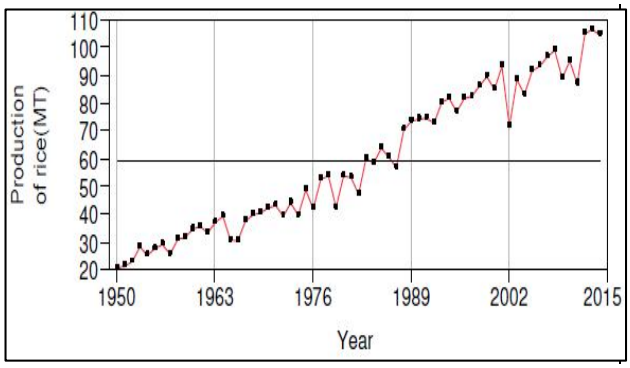


Fig 1: Behaviour of the annual production of rice in India.

for yearly rice production data through ACF and PACF plots in which spikes are coming outside from the insignificant zone and fails to follow the assumption of randomness of the data. So, forecasting rice production through ARIMA model has been estimated after transforming the variable. The logarithmic and differenced techniques are utilized to make variable stationary. Line chart, Correlogram and ADF test have been utilized again after taking 2nd order difference of variable.

Fig 3 shows line chart of the yearly rice production data after taking logarithm and 2nd order differencing. The mean of rice production has no change so it has constant variation throughout the time period. Thus, stationarity has been achieved through differencing, the mean and standard deviation were constant whereas, the trend ADF (-17.41). Also, Fig 4 showed that the order of p and q can at most be 1 and 2 through ACF and PACF plots as in ACF plot only one spike is outside the range whereas in PACF two spikes showed very small values (*i.e*, largly deviated from the cutoff range).

Several ARIMA models were developed on the basis of Box-Jenkins methodology as shown in Table 1. Among them the best five model have been proposed on the basis of minimum AIC (Akaike Information Criterion), SBIC (Schwartz Bayesian Information Criterion) and R^2. The model ARIMA (0 2 2) being significant have lowest AIC (-216.40) and SBIC(-209.97) values with the value of R^2 (0.956).

The estimates of the parameter are shown in Table 2 having MA (1) as 1.7833, MA (2) as -0.7834, which were positively and negatively significant respectively. Further, ARIMA (0 2 2) model has MAPE (1.8514) and – 2 log likelihood (-222.3997) which indicate the best fit of the model. The next step is checking the model verification with checking the residual of the model which was done through examining the auto correlations and partial auto correlations of the residuals.

In Fig 5, the ACF and PACF plots of the residual indicate 'good fit' of the model. Moreover, p-values of Ljung-Box Q test are greater than the significance level (0.05), the residuals are now independent and the model fulfills the assumption of randomness.

Forecasting with ARIMA model

Fig 6 showed graphical representation of forecasting of ARIMA (0 2 2) model, which depicts the upward trend and

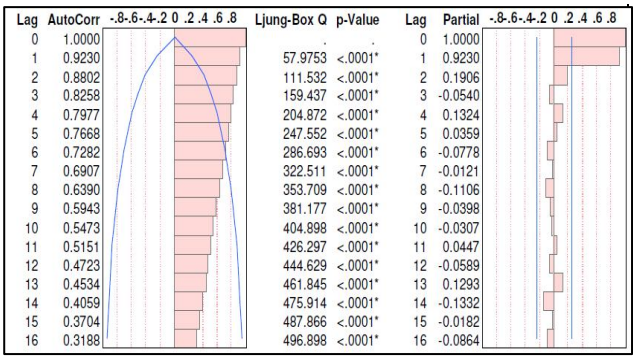


Fig 2: Correlogram of annual production of rice in India.

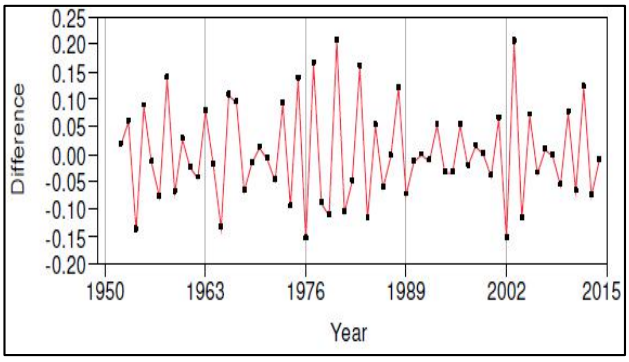


Fig 3: Behaviour of data after taking logarithm and differencing of order 2.

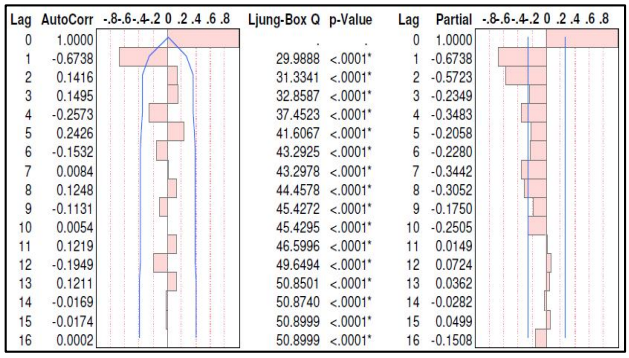


Fig 4: Correlogram of data after taking logarithm and differencing of order 2.

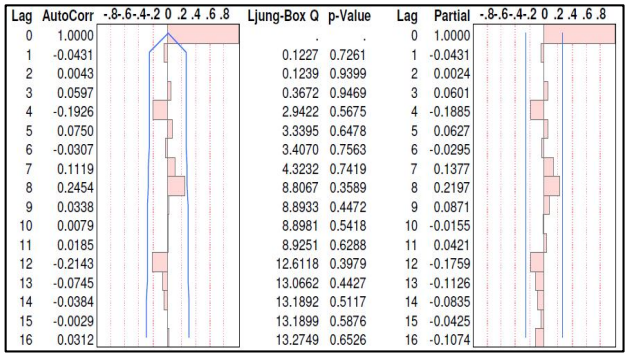


Fig 5: ACF and PACF plots of ARIMA (0 2 2) for annual production of rice crop.

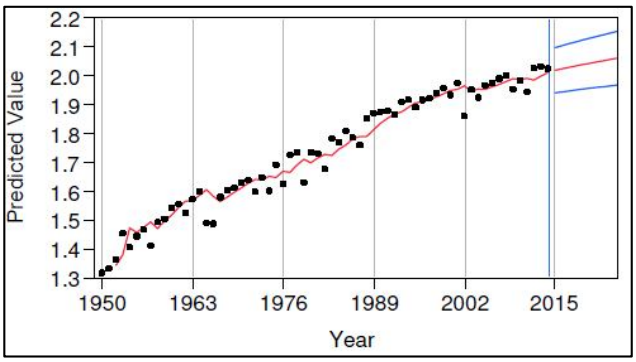


Fig 6: Forecasting graph of ARIMA (0 2 2) for annual production of rice crop.

Table 1: Different models for annual production of rice in India.

Model	Intercept	Significance of parameters/model	AIC	SBIC	R^2
ARIMA(0 2 2)	Yes	Significant	-216.40	-209.97	0.956
ARIMA(1 2 2)	Yes	Non-significant	-214.44	-205.87	0.953
ARIMA(2 2 2)	Yes	Non-significant	-212.45	-201.73	0.951
ARIMA(2 2 1)	No	Significant	-210.82	-204.39	0.942
ARIMA(2 0 0)	Yes	Significant	-205.21	-198.69	0.891

Table 2: Parameter estimates of ARIMA (0 2 2) for annual production of rice in India.

Term	Lag	Estimate	Std error	t ratio	p-value	MAPE	-2logliklehood	Constant estimate
MA1	1	1.7833	0.1112	16.04	<.0001			
MA2	2	-0.7834	0.1053	-7.44	<.0001	1.85	-222.39	-0.0002
Intercept	0	-0.0002	0.0001	-2.27	0.0268			

Table 3: Prediction table of ARIMA (0 2 2) for the production of rice in India.

Year	Actual (MT)	Forecasted values(MT)	L-95	U-95
2010-11	95.32	97.18	83.25	120.15
2011-12	87.10	98.09	84.17	121.48
2012-13	105.24	96.68	82.73	119.40
2013-14	106.54	99.87	86.14	124.33
2014-15	104.80	102.68	88.99	128.44
2015-16	104.52	104.47	90.69	130.88
2016-17	104.32	105.75	92.28	135.07
2017-18	109.70*	107.01	93.92	139.35
2018-19		108.24	95.62	143.73
2019-20		109.45	97.37	148.23
2020-21		110.64	99.16	152.83
2021-22		111.80	101.01	157.55
2022-23		112.92	102.91	162.40
2023-24		114.02	104.86	167.36
2024-25		115.09	106.86	172.46
2030-31		120.79	119.97	205.97

the forecasted value for the year 2020-21 may be 110.64 MT. The forecasts for rice production during 2010-11 to 2024-25 showing an increasing trend in Table 3.

After forecasting, the validity of the forecasted value has been checked for the year 2017-18. The estimated value for 2017-18 was 107.01 MT but the actual figure is 109.70 MT. This little difference in point estimation is because of various policy initiatives taken by government and normal monsoon in 2017-18 (www.economictimes.indiatimes.com). Rainfall deficit during the current monsoon season is now at 10 per cent, according to the Indian Meteorological Department (www.thehindu.com). According to them, some areas got extra rainfall, some areas were deficient, but in spite of the patchy rains, 2018-19 targets for rice is 113 MT which are marginally higher than last year's harvest (Minister of State for Agriculture Parshottam Rupala at the Ministry's Annual Conference, on February 28, 2019). The lower and upper bound values have also been evaluated for the forecasted values.

CONCLUSION

The estimated model ARIMA (0 2 2) for production of rice in India is as

$$\hat{Y}_t = -0.0002 + 2Y_{t-1} - Y_{t-2} - 1.7833\, e_{t-1} + 0.7834\, e_{t-2}$$

As the population increases over time gradually, therefore it is necessary to plan to meet the requirements of nation. For this purpose, forecasting is the key tool to alarm about the need of nation in advance. On the basis of above model, we have found that the rice production of India would become 115.09 MT in 2024-25 but it may increase more when the weather will be in the favour of farmers and even when the government take more policy initiatives which will boost and educate the farmers regarding the crop. In 2017-18 the actual production is 109.70MT and the forecasted production was 107.01 MT. The difference in the production of rice in the year 2017-18 may be due to irregular monsoon as well as due to policy initiatives taken by the government. This model will be used for forecasting of rice production time to time with the incorporation of current data.

REFERENCES

Akaike, H. (1979). A bayesian extension of the minimum AIC procedure of autoregressive model fitting. Biometrica. 66: 237-242.

Box, G.E.P., Jenkin, G.M. (1968). Discrete models for forecasting and control. In: Encyclopedia of Linguistics, Information and Control, 1-6, Oxford, Pergamon Press.

Box, G.E.P., Jenkin, G.M. (1976). Time Series of Analysis, Forecasting and Control, San Franscico, Holden-Day, California, USA.

Dickey, D., Fuller, W. (1981). Likelihood Ratio Statistics for Auto--regressive Time Series with a Unit Root. Econometrica. 49: 1057-1072.

Fuller, W. (1996). Introduction to Statistical Time Series, Second Edition. John Wiley, New York.

Granger, C. W. J., Newbold, P. (1973). Some comments on the evaluation of forecasts. Applied Economics. 5: 35-47.

http://eands.dacnet.nic.in

https://www.jmp.com

https://www.thehindu.com

Schwarz, G. (1978). Estimating the dimension of a model. Annals of Statistics. 6: 461-464.