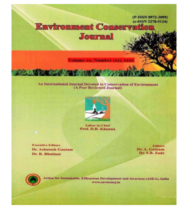
Environment Conservation Journal 24 (1):225-231, 2023

Journal homepage: https://www.environcj.in/

Environment Conservation Journal

ISSN 0972-3099 (Print) 2278-5124 (Online)



Historical pattern of rice productivity in India

Bhola Nath ✉

Department of Agricultural Statistics, Institute of Agriculture, Visva-Bharati, West Bengal, India

D. Bhattacharya

Department of Agricultural Statistics, Institute of Agriculture, Visva-Bharati, West Bengal, India

ARTICLE INFO

Received : 19 April 2022
Revised : 31 July 2022
Accepted : 28 August 2022

Available online: 15 January 2023

Key Words:
ARIMA
Forecast
Mann Kendal's trend
MAPE
Productivity

ABSTRACT

Forecast of productivity (yield) has an importance over production and area separately because it depends on both. Trend of the same reveals the necessity of the resources to be managed, for increasing yield in future. The forecast values of the series are obtained using autoregressive integrated moving average (ARIMA) model and the trend is determined by the means of Mann Kendall's trend test. In the present work we have found that the productivity of rice for overall country shows an increasing trend. Mann Kendal's trend analysis reported that the productivity has a steadily increasing trend which was also evident from the Sen's slope coefficient (*Q*). ARIMA (1,1,1) model with constant was found to be appropriate model for forecasting the productivity of rice. The forecast values were obtained for the subsequent four years starting from 2018 to 2021. Forecast error was also calculated and it was found to be less than 2 per cent i.e., 1.36 per cent.

Introduction

Rice is one among the most essential crops for human consumption which is grown and consumed by almost each and every part of India. India itself has an area of 43.79 million hectare with a rate of production of 168.50 million tonnes and average productivity of 3.85 tonnes per hectare (FAOSTAT, 2017). Rice has a potential to grow in different diversified climates so it is considered to be an important crop which also provide the food security to the country. A proper trend analysis and forecast for such a crucial crop has potential significance on many accounts as food securities and the management of storage and transportation facilities (Tripathi *et al.*, 2014). To know the maximum possible potential of yield as well an area and production for the optimum harvest of a crop, there should be a proper knowledge of ecology, appropriate advancements in that region.

In the present investigation we have taken productivity of rice as a variable under study. Why "productivity" rather than "production"? The answer is that the productivity is the result of relationship between area and production. At present almost all the possible resources (i.e., area, technologies) are being used in their optimum levels, then there is no further scope of increment in them. The only option is to fulfil the food requirements of the increasing population is increase the yield per unit of land. Keeping the above views in mind, the present study deals with the following objectives: (*i*) to determine the trend (*ii*) to forecast and validate the rice productivity.

An upward rise in price can be seen with the decrease in production which reduces marketable surplus. The adverse effect of income on farmers can also be seen with the increase in production i.e., as production increases the price decreases or *vice-versa*. To determine the inflation rate, salaries, wages and various policies related decisions, price plays an important role. The managements like surplus and deficit, which would stabilize the price in long term, can be achieved roughly by proper forecast of the production and also ensures the profit to the farmer. The latest trend in yield as well as production and area should be well studied for getting an idea about the future requirements of storability and transportation. Sometimes the actual series does not hold the assumptions imposed on

Corresponding author E-mail: *bnagstat@gmail.com*

Doi:https://doi.org/10.36953/ECJ.12292333

error term. In that case, a transformation of the data set or the different techniques from the class of non-parametric approaches can be used (Nath *et al.*, 2020). Mann Kendal's trend test is a non-parametric procedure which is considered here for estimation of possible trend (Tripathi *et al.*, 2014). Trend analysis of productivity as well as area and production of garlic in Dindigul district of Tamil Nadu, India has been attempted by Manoharan and Ramalakshmi (2015). They have used Mann Kendal's trend test, Sen's slope coefficient, simultaneously for determination of magnitude of trend in productivity of rice. In the present work trend has been obtained by using Mann Kendal's trend test.

Now a day's modelling like remote sensing and simulation are being used widely for forecasting of the crop acreage and production. Sometimes, forecast for the same is needed much before the planting or harvesting of the crop. In that situation, ARIMA model which is based on the historical data (time series) can be used for forecasting the acreage, production and yield of crop. Contreras *et al.* (2003) used ARIMA methodology and obtained the forecast values for next-day electricity prices in the market of California and central Spain for daily markets namely, spot market and long-term market. Forecast values of productivity, production and yield of rice for the state of Odisha were attempted by Tripathi *et al.* (2014) by using ARIMA models for time series data starting from 1950-51 to 2008-09. Box-Jenkins' ARIMA model for forecasting the production of wheat for India was used by Nath *et al.* (2019). They selected the best appropriate ARIMA model by using minimum value of the AIC and MAPE. By using the selected model, forecast values for subsequent years can be obtained. Many studies available in the literature to justify that the vigilant and detailed selection of ARIMA model can produce a precise forecast to univariate time series. So, for forecasting the productivity of rice can be done by using ARIMA (p, d, q) model.

Data Collection

The data on productivity of rice for India were collected from the secondary source i.e., Data Net India Pvt. Ltd. Dataset comprise of time series data on annual productivity of rice (tonnes per hectare) starting from 1963 to 2017.The whole data set is divided in two parts viz., training set and validation set. The training data set has the observations starting from 1963 to 2014 and the remaining observations were retained for model verification under validation data set.

Trend Analysis

Mann Kendall trend test

Initially, this test was developed by Mann and Kendall separately and subsequently they derived the distribution of the test statistic also. The possible trend in a particular time series can be tested by using Mann-Kendall trend test which is a rank-based non-parametric method (Kolliesuah *et al.*, 2020).

The test statistic can be defined as,

$$s=\sum_{i=1}^{n-1}\sum_{j=(i+1)}^{n}\operatorname{sgn}(x_j - x_i) \quad (1)$$

where, x_i and x_j are the sequential data values, n is the total number of observations and

$$\operatorname{sgn}(\theta)=\begin{cases}+1, & \text{if } \theta>0\\ 0, & \text{if } \theta=1\\ -1, & \text{if } \theta<1\end{cases} \quad (2)$$

Significance level and the estimate of slope magnitude are the two parameters of this test. Significance level shows the strength of trend whereas the estimate of slope magnitude reflects the magnitude as well as the magnitude of the trend. For a random variable having *i.i.d.* $N(0,1)$ without any tie in data values, $E(S) = 0$ and

$$\sigma_S^2=\frac{n(n-1)(2n+5)}{18} \quad (3)$$

If some data values are tied, then the correction to $\sqrt{\sigma_S^2}$ can be done as,

$$\sigma_S^2 = \frac{n(n-1)(2n+5)-\sum_{i=1}^{n} t_i(i)(i-1)(2i+5)}{18} \quad (4)$$

Where, t_i denotes the number of ties of extent i. For n larger than 10, the test statistic can be modified as,

$$Z_s=\begin{cases}\frac{S-1}{\sqrt{\sigma_S^2}}, & \text{for } S>0\\ 0, & \text{for } S=0\\ \frac{S+1}{\sqrt{\sigma_S^2}}, & \text{for } S<0\end{cases} \quad (5)$$

where, Z_s is the standard normal variate.

Sen's slope coefficient

The magnitude of slope of trend can also be obtained by applying the Sen's slope coefficient (Ghimire *et al.*, 2018). Sen's estimate for slope is associated with the Mann-Kendall test as,

$$\beta=\text{median}\left(\frac{x_j-x_i}{j-i}\right), \quad \text{for all } j>i, \quad (6)$$

The median of these N values of β_i is represented as Sen's estimator of slope which is defined as,

$$Q_i = \begin{cases} \beta_{(N+1)/2}, & \text{if } N \text{ is odd} \\ \frac{1}{2}\left(\beta_{\frac{N}{2}} + \beta_{\frac{N+2}{2}}\right), & \text{if } N \text{ is even} \end{cases} \quad (7)$$

The positive and negative values of Q indicate an upward and downward trend, respectively.

ARIMA Modelling

To obtain the forecast values of equally distant time series, the well-known ARIMA modelling approach can be used. An ARIMA model forecasts the value of a dependent series by considering the linear combination of its own past values. Suppose the model is ARIMA(1,1,1), then it can be written mathematically as,

$$Y_t = \mu + \phi_1 Y_{t-1} + \theta_1 \varepsilon_{t-1}, \quad (8)$$

where, μ= the mean term, Y_t= the response variable observed at time t, ϕ_1= coefficient of the AR component, θ_1= coefficient of the MA component and ϵ_t= error term. The model given in (8) can be extended for, p number of AR components, q number of MA components and d, difference taken to make the series stationary, respectively and can be expressed as the model given in (9).

$$Y_t = \mu + \phi_i Y_{t-i} + \theta_i \varepsilon_{t-i}, \quad (9)$$

(***i***) ***Identification stage:*** The ARIMA model needs a stationary time series which can be performed using Augmented Dickey-Fuller (ADF) test. Then, the tentative values of the number of parameters "*p*"and "*q*" are to be decided. These values can be decided by looking at the significant spikes in autocorrelation function (ACF) and partial autocorrelation function (PACF). At this stage, one or more tentative models can be chosen for the available data.

(***ii***) ***Estimation stage:*** The tentative orders of the ARIMA models are needed to be estimated. These estimates can be obtained by using different packages like, SPSS, R etc. Presently, R is considered in this study.

(***iii***) ***Diagnostic checking:*** The best fit model can be obtained by considering the following information criteria:

(a) Significance of the parameters: Significance test forall the estimates of the parameters in the model should be obtained.

(b) Akaike Information Criteria (AIC): AIC can be estimated as,AIC = $(-2 \log L + 2m)$, where $m = p + q$ and L is the likelihood function.

(c)Model evaluation: Candidate models can be further evaluated by using mean absolute percent error(MAPE) which is defined in (10).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{Y}-Y|}{Y} \times 100, \quad (10)$$

where, $\hat{Y}$ is the forecast value, Y is actual value of response variable and n is the total number of observations.

(d) Residual diagnosis: The residuals are to be diagnosed for autocorrelation and normal distribution. These can be done by ACF, PACF and normal Q-Q plot, respectively.

Empirical Findings

Mann Kendall's trend analysis

As has already been discussed that the it is a nonparametric test, then it becomes an important task to test the normality of the residuals of the concerned series. The normality of the error distribution of the series has been tested by using Shapiro-Wilk's test and Normal Q-Q plot (Nath *et al.*, 2020). Shapiro Wilk's test and normal Q-Q plot suggested that the series does not follow the normality of the error distribution (Table-1 and Figure-1). Under this test, the null hypothesis about normal distribution is rejected because *p*-value is less than 0.05 also the observations are deviating much from the normal line in normal Q-Q plot.

It is evident from Table-2 that the time series data under study possess an upward trend as Q is found to be positive and further it can be checked by forecasting the actual series using ARIMA model.

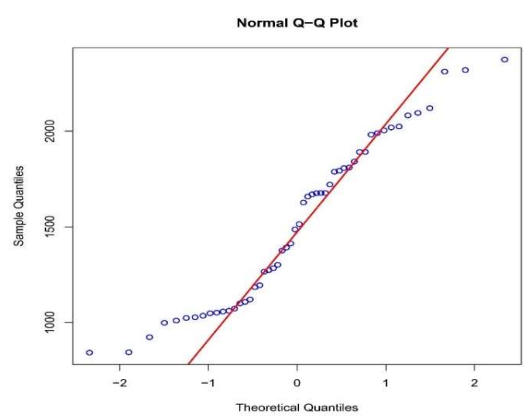


Figure 1: Normal Q-Q plot for the actual time series data

Table1: Descriptive measures for the actual time series

Variable	Mean	Median	Skewness	Kurtosis
Productivity	1516.71	1500.50	0.1963	-1.2467

Shapiro-Wilk's test	
Test statistic (*W*) value	Significance (*p*-value)
0.9405	0.011*

*significant at 5% level of significance.

Table 2: Mann Kendall's trend and Sen's slope coefficient summary

Test Statistic (*Z*) value	Observations (*n*)	*p*-value
8.538	52	0.001
S	σ_S^2	τ
1083	16058.33	0.817
Sen's slope coefficient (*Q*)		
Q	Lower limit	Upper limit
27.528	25.667	29.500

Table 3: Augmented Dickey Fuller test summary

	Statistic value	Lag order	*p*-value
Actual series	-2.8314	3	0.24
Series with $d = 1$	-5.9535	3	0.01**

**Significant at 1% level of significance.

Table 4: List of candidate ARIMA models.

Model	Coefficient	Estimate (se)	AIC
ARIMA (1,1,0)	*AR(1)*	-0.5633 (0.1131)	682.63
ARIMA (1,1,0) with Constant	*AR(1)*	-0.5855 (0.1107)	681.88
	Constant	27.3948 (16.1793)	
ARIMA (1,1,1)	*AR(1)*	-0.3275 (0.2163)	682.97
	MA(1)	-0.3535 (0.2176)	
ARIMA (1,1,1) with Constant	***AR(1)***	**-0.1397 (0.2395)**	**677.56**
	MA(1)	**-0.7132 (0.2319)**	
	Constant	**26.9253 (6.3578)**	

Table 5: MAPE for cross validation of fitted model

Year	Forecast	Actual	APE
2015	2305.28	2295.00	0.4459
2016	2337.88	2305.00	1.4066
2017	2364.02	2417.00	2.2413
MAPE (%)		**1.3646**	

Table 6: Box-Ljung test summary.

Test	χ^2 statistic value	*df*	*p*-value
Box-Ljung	1.3522	19	1

Table 7: Forecast summary obtained from ARIMA (1,1,1) with constant model

Year	Forecast	90% (C.I.)		95% (C.I.)		99% (C.I.)	
		Low	High	Low	High	Low	High
2018	2391.05	2081.14	2700.97	2021.76	2760.34	1905.73	2876.38
2019	2417.96	2099.65	2736.28	2038.66	2797.26	1919.48	2916.44
2020	2444.89	2118.41	2771.37	2055.86	2833.91	1933.62	2956.15
2021	2471.81	2137.37	2806.26	2073.30	2870.33	1948.07	2995.56

Note- C. I. stand for Confidence Intervals.

Fitting of ARIMA model

Model identification

ADF Test: Under ADF test the construction of the hypotheses is done as,

H_0: the series is not stationary against H_1: the series is stationary

Then this hypothesis was tested for the actual time series and it was found that the actual time series was not a stationary time series. So, the first order difference (i.e., $d = 1$) of the actual time series was taken and ADF test was applied to that series and it was found that the differenced time series was stationary. First order difference ($d = 1$) means we have generated a differenced time series of current year (Y_t) and immediate previous year values [i.e., $Y = Y_t - Y_{t-1}$]. The test result, is given in Table-3.

ACF and PACF

Approximate ARIMA models can be decided by the deciding the values of (p, d, q) As we have discussed earlier that choosing the order of ARIMA(p, d, q) is the way to get approximate ARIMA models. The actual series was not stationary, then first order differenced time series was tried for stationarity check and it produced significant result. That produces the value of "d" as "1" for the present time series. Now, the AR (p) and MA(q) orders are needed to be defined and it would be possible only by looking at the significant spikes of the ACF and PACF plots for the differenced time series.

Figure-2 produces the MA(q) order of 1 to 13, as all the 13 lags have significant spikes and AR(q) order of 1 and 2 as lags 1 and 2 are having significant spikes. It becomes difficult to identify an appropriated ARIMA (p, d, q) model from the ACF and PACF plots of actual time series because the actual time series was non-stationary. Therefore, the ACF and PACF plots of first differenced time series (Figure-3) was considered for determination of the orders and MA(q) order i.e., q = 0, 1 and AR(p) order i.e., p = 1. These values were found to be the approximate values of the parameters to be considered for building the model for concerned time series.

Model identification

The four approximate ARIMA (p, d, q) models were found to be the appropriate models based on the looking at the ACF and PACF plots. The detailed list of the expected models with their respective AIC values are given in the Table-4. Table-4 reveals that ARIMA (1,1,1) with constant was found to have the minimum AIC (677.56) value among four candidate ARIMA models. So, ARIMA (1,1,1) with constant model was found to be the appropriate ARIMA model for the present time series data. Then this said model has been checked for error diagnostics and diagnostic check summary is given in the Table-5.

Further, forecast values have been obtained by using ARIMA (1,1,1) model with constant after the cross validation of the model based on MAPE (Table-5). The MAPE is found to be less than 2 per cent (i.e., 1.36%).

Test for the autocorrelation in Residuals

Now in next step we would check presence of autocorrelation among the residuals obtained from the fitted ARIMA(1,1,1) model with constant by using "Box-Ljung" test (Table-6). The large p-values (more than the α=0.05)of the test suggests that H_0could not be rejected and it may lead to the conclusion that the autocorrelation functions were found to be non-significant among lags 1 to 20. Here the degrees of freedom are 19 because 20 lags have been used under ACF and PACF plot. Thus, it can be concluded the assumption of presence of serial correlation is rejected in the fitted ARIMA(1,1,1) model with constant.Residuals were further checked for the normal distribution by plotting histogram (Figure-4) and normal Q-Q plots (Figure-5).From figures 4 and 5, it can be clearly said that residuals are normally distributed. On the basis of normal Q-Q plot of standard residuals (Figure-5) in the fitted ARIMA (1,1,1) model, it can be concluded that standard errors are roughly constant with respect to mean and variance overtime.To check whether there is any auto correlation present in forecast errors, the plots of ACF and PACF are obtained. It can easily be seen that there is no any lag with significant spike in both the plots i.e., ACF and PACF plots (Figure-6).

Forecast using fitted ARIMA (1,1,1) model with constant

The forecast has been obtained from the selected ARIMA (1, 1, 1) with constant model, which was found to be the appropriate ARIMA model for the present time series data. Forecast values (Table-7) with the lowest and highest values for the confidence intervals of 90, 95 and 99 percent are

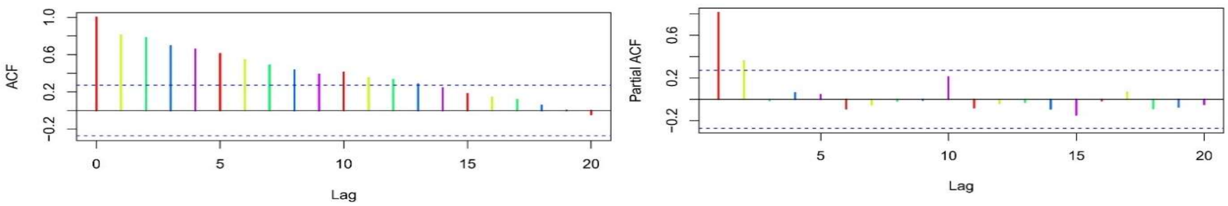


Figure 2: ACF and PACF plot for the actual series

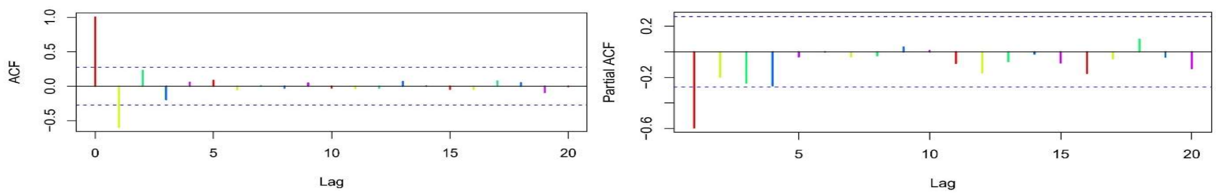


Figure 3: ACF and PACF plots of the differenced time series

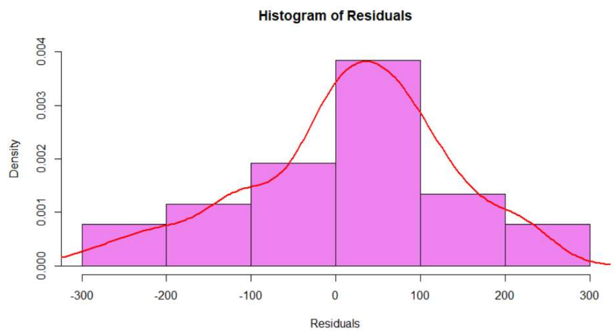


Figure 4: Histogram of residuals obtained in fitted ARIMA (1,1,1) model with constant

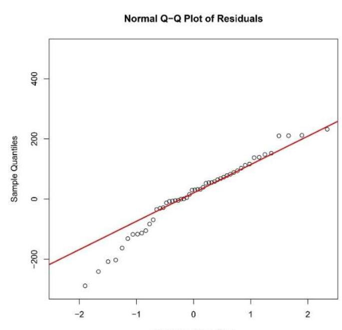


Figure 5: Normal Q-Q plot of residuals obtained in ARIMA (1, 1, 1) model with constant

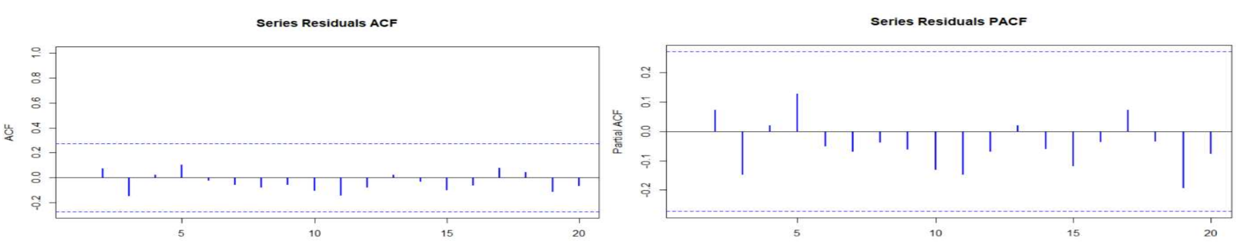


Figure 6: ACF and PACF plots of residuals obtained in ARIMA (1,1,1)

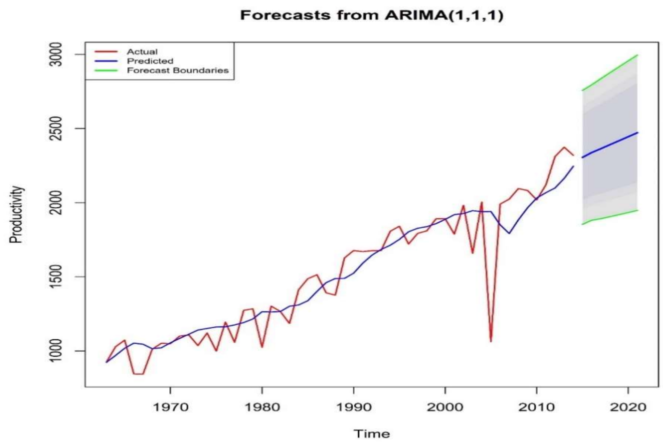


Figure 7: Plot of forecast with ARIMA (1,1,1) model with constant

obtained. The plot of actual and forecast values are also obtained which shows steadily increasing trend during the period of forecast (Figure-7).

Conclusion

Mann Kendal's trend test shows a significant presence of trend in the data on productivity of rice which is also evident from Sen's *Q* statistic. The positive value of *Q* indicates the presence of increasing trend. Upward trend can also be seen in the plot of forecast provided in the Figure-7 which has been obtained by using ARIMA (1,1,1) model with constant. The forecast values are showing the average growth rate of approximately 1.11 percent indicating the estimate of productivity of 2471.81 quintals per hectare for the year of 2021. To increase the potential productivity of rice suitable varieties according to the different ecological conditions can be introduced in farmer's field along with the nutrient requirement and agronomic management practices. Based on the forecast and validation results, it may be concluded that ARIMA (1,1,1) model with constant captures the real behaviour of the data.

Acknowledgement

Authors are highly grateful to the editors and reviewers for their invaluable comments.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Contreras, J., Espinola, R.F., Nogales, J. & Conejo, A.J. (2003). ARIMA models to predict next-day electricity prices. *IEEE Transactions on Power Systems*, 18(3), 1014-20.

Ghimire, D., Lamsal, G., Paudel, B., Khatri, S. & Bhusal, B. (2018). Analysis of trend in area, production and yield of major vegetables of Nepal. *Trends in Horticulture*, 1(2), 1-11.

Kolliesuah, N.P., Saysay, J.L., Zinnah, M.M., Freeman, A.T. & Chinenye, D. (2020). Trend analysis of production, consumption and export of cashew crop in West Africa. *African Crop Science Journal*, 28(s1), 187-202.

Manoharan, M.& Ramalakshmi, T. (2015). Cost, returns and resource use efficiency in garlic cultivation in dindigul district. *International Journal of Business and Administration Research Review*, 3(10), 16-25.

Nath, B., Dhakre, D. S. & Bhattacharya, D. (2019). Forecasting wheat production in India: An ARIMA modelling approach. *Journal of Pharmacognosy and Phytochemistry*, 8(1), 2158-65.

Nath, B., Dhakre, D. S. & Bhattacharya, D. (2020). The challenge of selecting the best forecasting model for a time series data. *Journal of Food, Agriculture and Environment*, 18(2), 97-102.

Tripathi, R., Nayak, A.K., Raja, R., Shahid, M., Kumar, A., Mohanty, S., Panda, B.B., Lal, B. & Gautam, P. (2014). Forecasting Rice Productivity and Production of Odisha, India, Using Autoregressive Integrated Moving Average Models. *Advances in Agriculture*, (1), 1-9.