International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2024; SP-9(3): 108-114

www.mathsjournal.com
Received: 03-03-2024
Accepted: 09-04-2024

Ajay Kumar
College of Horticulture, Maharana Pratap Horticultural University, Karnal, Haryana, India

Satish Manda
SCPK RRC, Maharana Pratap Horticultural University, Karnal, Haryana, India

Raj Kumar
College of Horticulture, Maharana Pratap Horticultural University, Karnal, Haryana, India

Rajender Kumar
Central Library, Maharana Pratap Horticultural University, Karnal, Haryana, India

Nitin Tanwar
Department of Statistics, Lady Shri Ram College for Women, University of Delhi, Delhi, India

Corresponding Author:
Satish Manda
SCPK RRC, Maharana Pratap Horticultural University, Karnal, Haryana, India

Forecasting sugarcane productivity of India using ARIMA models

Ajay Kumar, Satish Manda, Raj Kumar, Rajender Kumar and Nitin Tanwar

Abstract
India has a well-established system of collecting agricultural statistics. Sugarcane is one of the important commercial crops in India. Crop productivity forecasts and crop production estimates are necessary for national food security including early determination of the import/export plan and price and to provide timely information for optimum management of growing crops. This paper attempts forecasting the sugarcane productivity of India using the Univariate Auto Regressive Integrated Moving Average (ARIMA) models and the time series data taken from 1980-81 to 2022-23.The results of the study revealed that the ARIMA model (3, 2, 0) have been selected as best models among all the models for prediction of the sugarcane productivity on the basis of Schwarz's Bayesian Information Criterion, Mean absolute prediction error (MAPE), MAE and R^2 for sugarcane productivity of India. The performances of models are validated by comparing with actual values.

Keywords: Sugarcane, ARIMA, crop productivity, forecast, R square

Introduction

Sugarcane (*Saccharum officinarum*) is widely grown crop in India. It provides employment to over a million people directly or indirectly besides contributing significantly to the national exchequer. India has the largest area under sugarcane cultivation in the world and the world's second largest producer of sugarcane next only to Brazil. In India, Sugarcane is grown as a Kharif Crop. Broadly there are two distinct agro-climatic regions of sugarcane cultivation in India, *viz*., tropical and subtropical. The tropical sugarcane region consists of sugarcane agro climatic zone 4 (Peninsular zone) and 5 (Coastal zone) which includes the states of Maharashtra, Andhra Pradesh, Tamil Nadu, Karnataka, Gujarat, Madhya Pradesh, Goa, Pondicherry and Kerala. In the case of sub-tropical sugarcane region: Around 55 per cent of total cane area in the country is in the sub-tropics. U.P, Bihar, Haryana and Punjab comes under this region.

Sugarcane belongs to bamboo family of plants and is indigenous to India. It is the main source of sugar, gur and khandsari. About two-thirds of the total sugarcane produced in India is consumed for making gur and khandsari and only one third of it goes to sugar factories. It also provides raw material for manufacturing alcohol. It is a long duration crop and requires 10 to 15 and even 18 months to mature, depending upon the geographical conditions. It requires hot and humid climate with average temperature of 21-27 °C and 75-150 cm rainfall.

Review of literature

Suresh and Priya (2011) [14] forecasted the sugarcane area, production and productivity of Tamil Nadu through fitting of univariate Auto Regressive Integrated Moving Average (ARIMA) models using data from 1950–2007 and found that ARIMA (1, 1, 1) model is suitable for sugarcane area and productivity as well as ARIMA(2, 1, 2) is found appropriate for modeling sugarcane production. Hossain and Abdulla (2016) [4] applied Box-Jenkins ARIMA model on potato production in Bangladesh over the period 1971 to 2013 and conclude that the best selected ARIMA model for forecasting the potato production for whole Bangladesh is ARIMA (0, 2, 1).

Kumar *et al.* (2017) [7] forecasted the sugarcane (*Saccharum officinarum*) productivity of Bihar through fitting Box Jenkins univariate Auto Regressive Integrated Moving Average (ARIMA) model on time series data on sugarcane productivity in Bihar from 1939-40 to 2014-15 and on the basis of Akaike information criterion (AIC), best model selected is ARIMA (0, 1, 1). Khan *et al.* (2015) [5] applied ARIMA methodology to forecast the rice production in Pakistan based on time series data 1993 to 2015 and showed that ARIMA (2, 1, 1) model is more suitable for next 15 years forecasting. Mehmood and Ahmad (2013) developed a univariate ARIMA model by using Box-Jenkins methodology to forecast the area of mangoes in Pakistan for a period 1961 to 2009. The analysis found ARIMA (0, 1, 0) as an appropriate model to forecast the area in Pakistan.

Materials and Methods

The Data on sugarcane productivity (In tonne/ha) of India for a period from 1980-81 to 2022-23 has been collected from Indiastat.com to develop ARIMA models and find out best among them. Then with the help of selected models, next year productivity of India also predicted.

Box-Jenkins' Autoregressive Integrated Moving Average methodology

Box-Jenkins ARIMA forecasts are based only on past value of the variable being forecast. The Box-Jenkins procedure for fitting a good forecasting model consist of three stages i.e. i) Identification stage ii) Estimation stage and iii) Diagnostic checking

The general functional form of ARIMA model may be expressed as

$$\phi_p(B)\Delta^d Y_t = C + \theta_q(B)a_t$$

Where,

Y_t= Variable under forecasting

B = Lag operator

a = Error term ($Y_t - \hat{Y}_t$, where $\hat{Y}t$ is the estimated value of Y_t)

t = the time subscript

$\phi_p(B)$ = Non-seasonal AR

$(1-B)^d$ = Non-seasonal difference

$\theta_q(B)$ = Non-seasonal MA

Different statistical measures have been used for comparison and validation of the fitted models as described below:

Bayesian Information Criterion (BIC)

It is a criterion for model selection among a finite set of models, based on maximum likelihood function L of the estimated model, and is defined by formula

$$BIC = -2.\ln L + k \ln(n)$$

Where k is the number of parameters to be estimated and n is the number of years for which forecasting has been done.

Root Mean Square Error (RMSE)

It is used as a measure of comparing two models and its formula is given as

$$RMSE = \left[\left\{\frac{1}{n}\sum_{i=1}^{n}(O_i - E_i)^2\right\}\right]^{\frac{1}{2}}$$

Mean Absolute Percentage Error (MAPE)

This measures the accuracy of a fitted model and is given as

$$MAPE = \frac{100}{n} \times \sum_{i=1}^{n}\left|\frac{O_i - E_i}{O_i}\right|$$

Where O_i and E_i are the observed and forecasted values and n is the number of years for which forecasting has been done.

Results and Discussion

Identification of order for Autoregressive and Moving average polynomial

Identification involves the determination of the appropriate order of AR and MA polynomials i.e. values for p and q. The orders were determined from the autocorrelation functions and partial autocorrelation functions of the stationary series. The sugarcane productivity (s) data was found to be non-stationary from the Table 1. Almost all the autocorrelations upto (n/4) th lags significantly different from zero confirmed non-stationarity. The plotting of the acfs (Figure 1) also indicates that the acfs decline gradually implying non-stationarity. Thus the crop productivity series considered here were found to be non-stationary. The non-stationary data series was transformed into stationary series by the second differencing of the original data series. Differencing of order one i.e. d=2 was enough for getting an approximate stationary series.

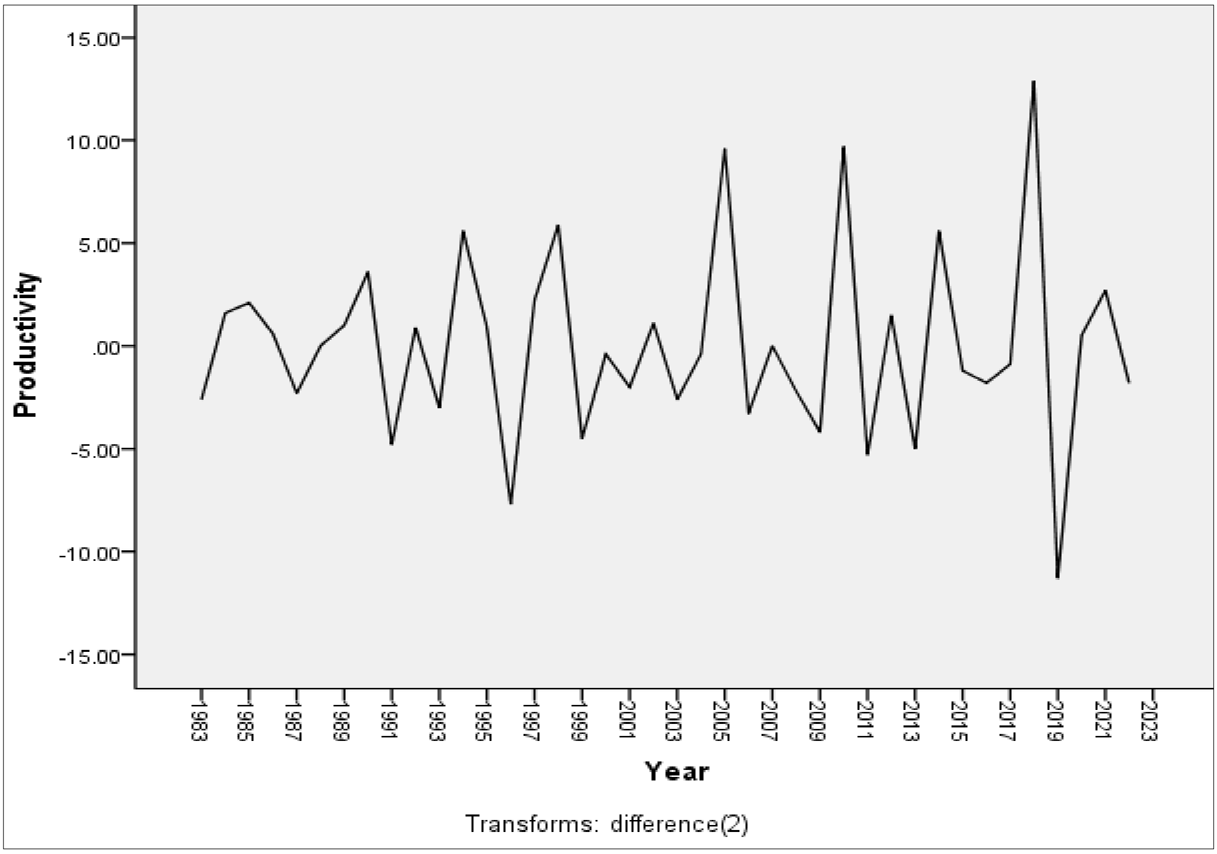
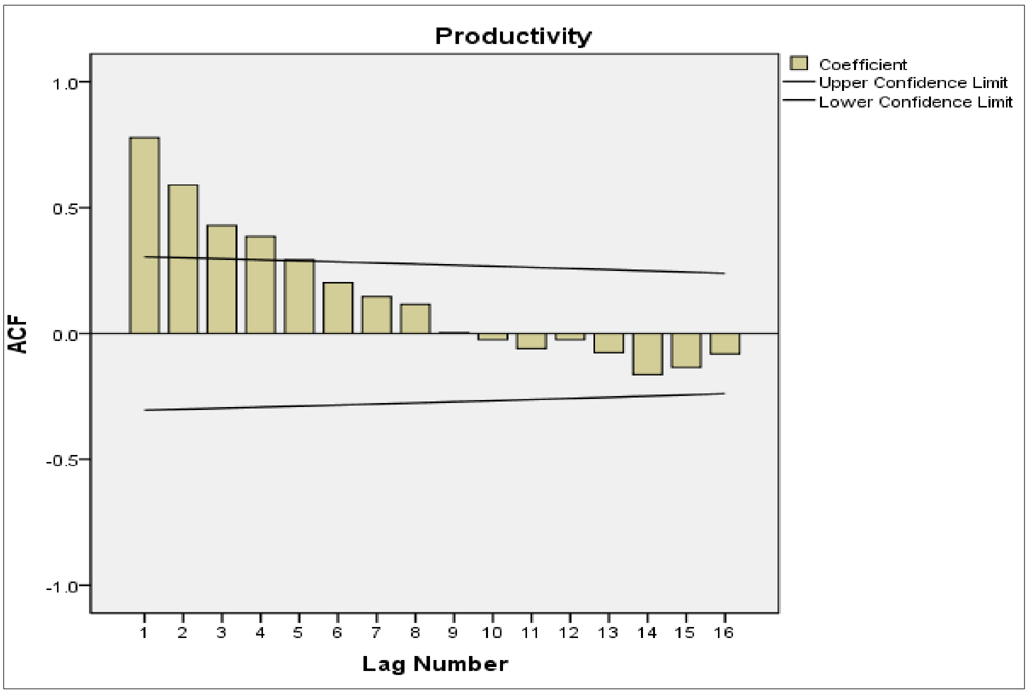


Table 1: Autocorrelation for Sugarcane productivity

Lag	Autocorrelation	Std. Error	Box-Ljung Statistic		
			Value	DF	Sig
1	.778	.152	26.042	1	<0.01
2	.590	.150	41.443	2	<0.01
3	.429	.148	49.787	3	<0.01
4	.385	.146	56.703	4	<0.01
5	.293	.144	60.822	5	<0.01
6	.202	.142	62.845	6	<0.01
7	.146	.140	63.934	7	<0.01
8	.116	.138	64.640	8	<0.01
9	.003	.136	64.641	9	<0.01
10	-.025	.134	64.675	10	<0.01



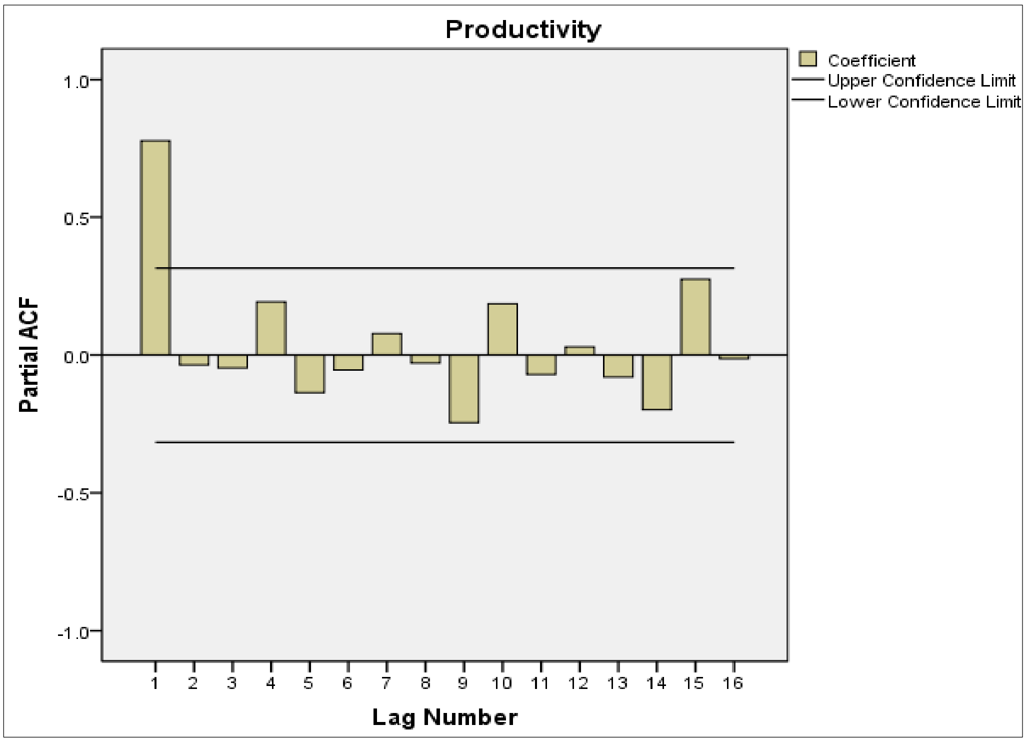
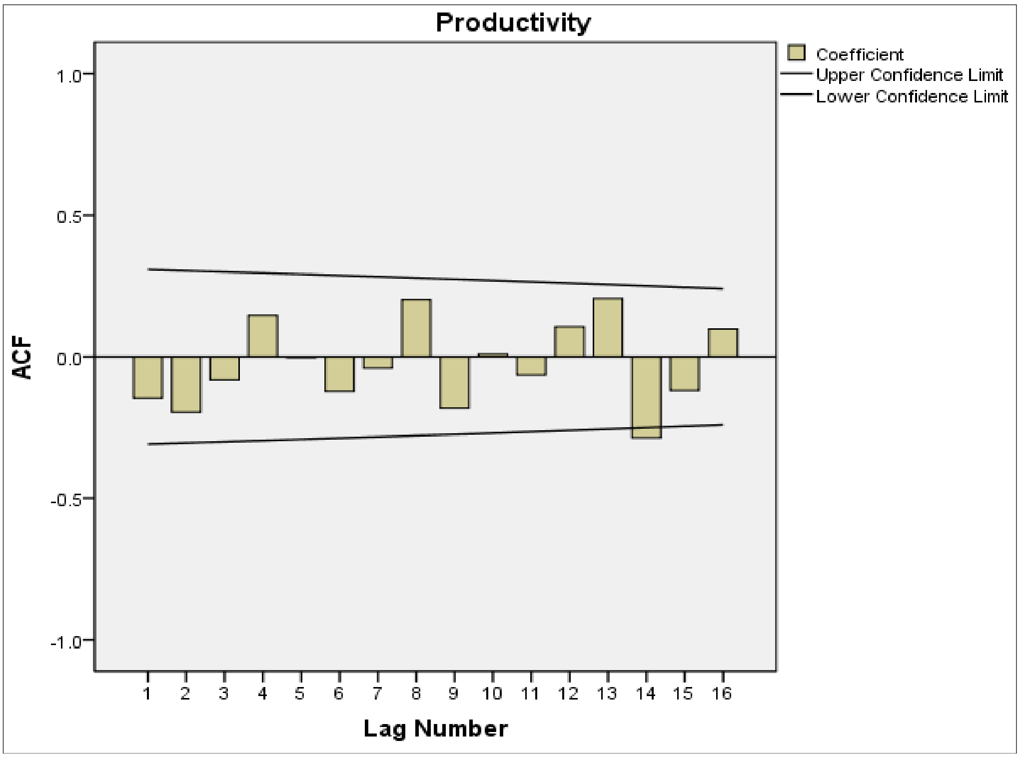


Fig 1: Autocorrelation and Partial autocorrelation for Sugarcane productivity



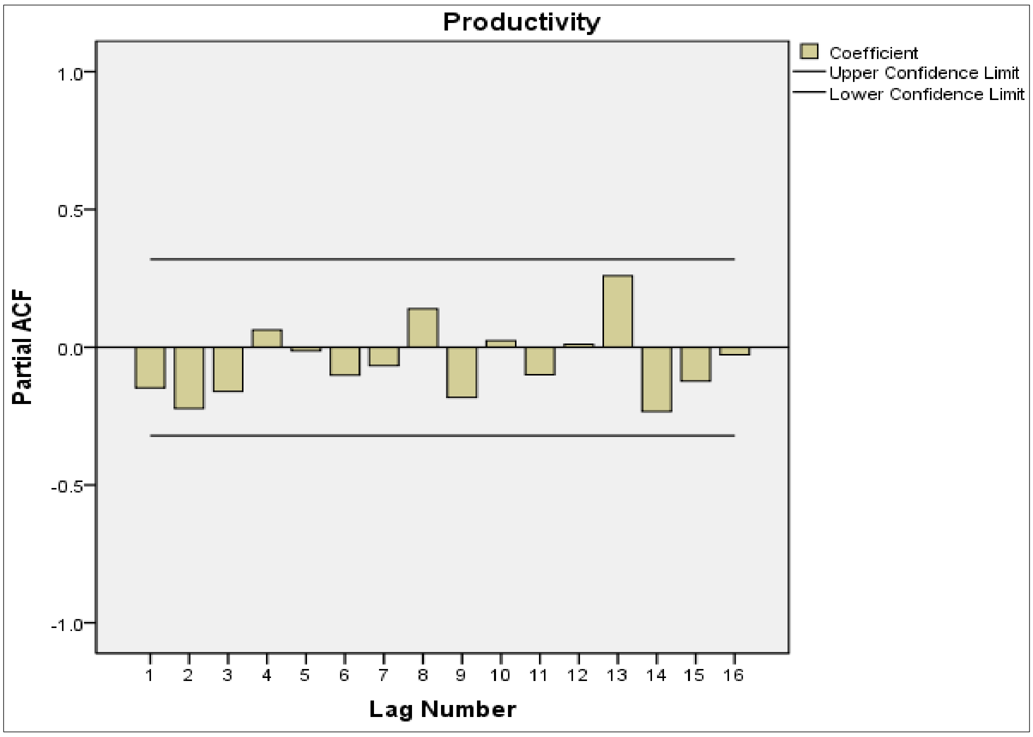


Fig 2: Autocorrelation and Partial autocorrelation for Sugarcane productivity after differencing

Parameter Estimates

After experimenting with different lags of the moving average and auto regressive processes; different ARIMA model were fitted for estimating the sugarcane productivity. The models ARIMA (3, 3, 0), ARIMA (4, 2, 0) and ARIMA (3, 2, 0) were considered in the identification stage and ARIMA estimation was carried out using a non-linear least squares (NLS) approach. The relatively popular method for estimation of parameters under ARIMA structure is due to Marquardt who has designed a powerful algorithm for estimation through iterative improvement. Parameter estimates of the selected models shown in Table 2 are less than one (Needed for convergence) and also satisfy the stationarity and invertibility conditions under ARIMA structure.

Table 2: Tentative ARIMA models for Sugarcane productivity

Model		Estimate	Std. Error	Approx Prob.
ARIMA (3, 3, 0)	AR (1)	-0.85	0.15	<0.01
	AR (2)	-0.65	0.17	<0.01
	AR (3)	-0.40	0.15	<0.01
ARIMA (4, 2, 0)	AR (1)	-0.98	0.17	<0.01
	AR (2)	-0.90	0.21	<0.01
	AR (3)	-0.65	0.22	<0.01
	AR (4)	-0.29	0.19	0.13
ARIMA (3, 2, 0)	AR (1)	-1.20	0.15	<0.01
	AR (2)	-1.01	0.19	<0.01
	AR (3)	-.54	0.16	<0.01

Table 3: Selection Criteria vales for choosing ARIMA models

Models	RMSE	R-squared	MAE	MAPE	BIC
ARIMA(3,3,0)	4.89	0.53	3.37	4.92	3.65
ARIMA(4,2,0)	3.55	0.76	2.50	3.67	2.99
ARIMA(3,2,0)	3.46	0.78	2.42	3.56	3.04

After experimenting with different lags of the moving average and the autoregressive processes, ARIMA (3, 2, 0) was found to be the best fit for wheat productivity estimation.

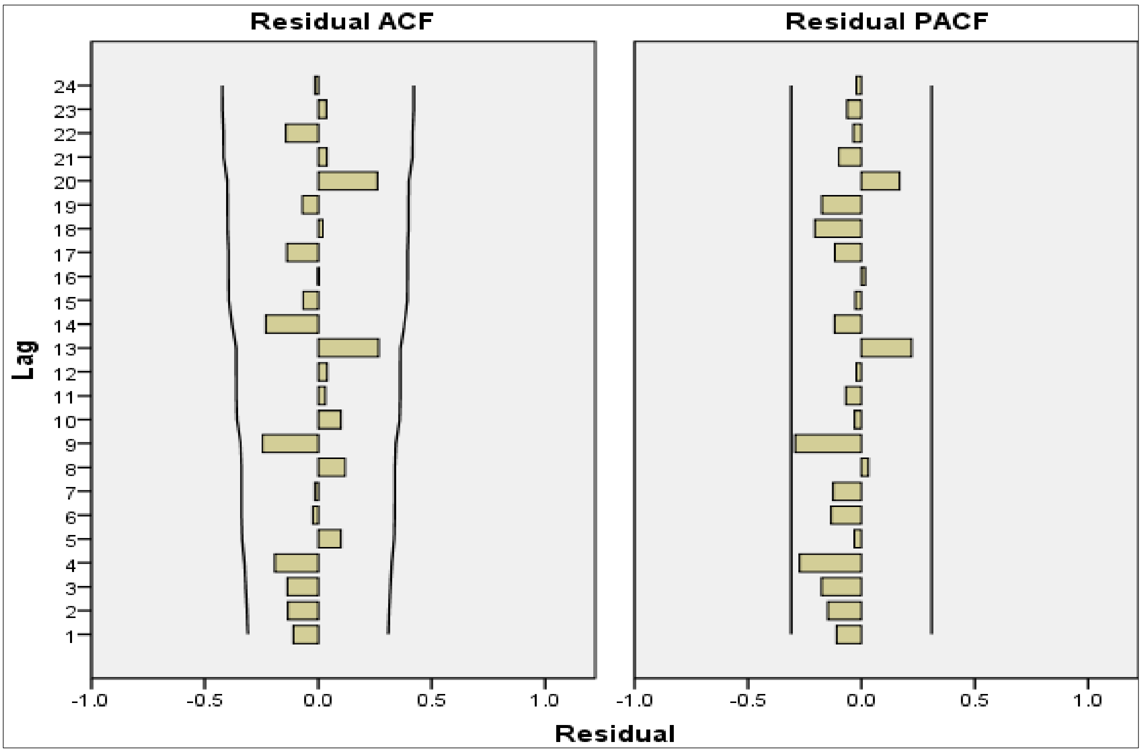


Fig 3: Residual ACF and PACF plot based on fitted ARIMA model

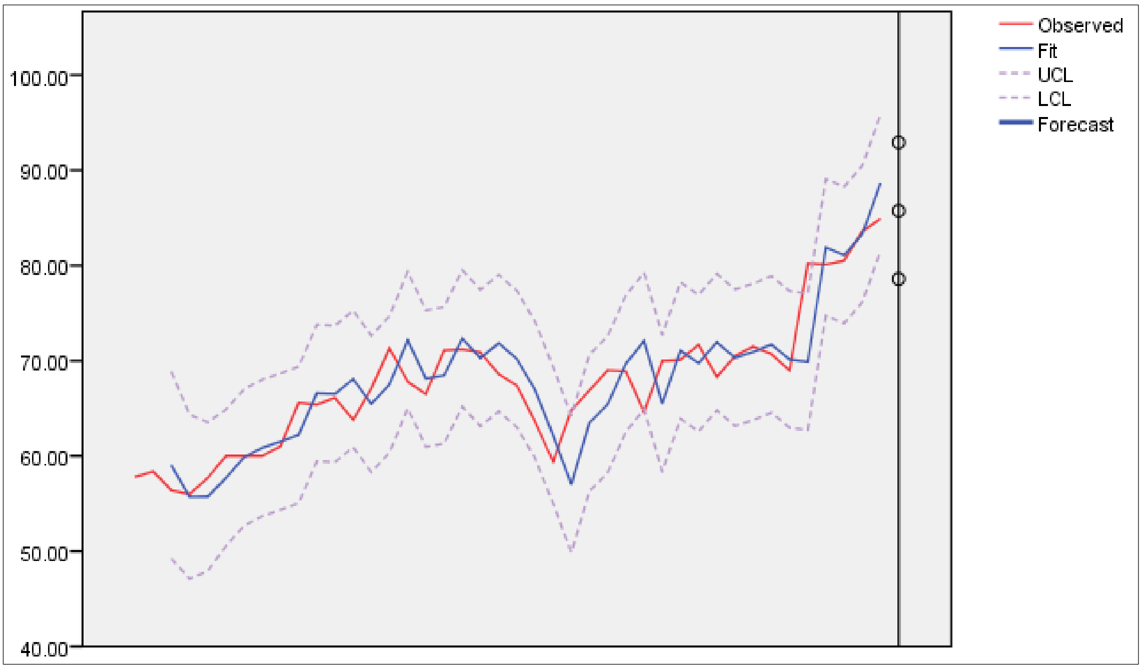


Fig 4: Observed and Estimated productivity graph based on fitted ARIMA model

Table 4: Estimated Sugarcane productivity based on ARIMA (3, 2, 0) model and their associated percentage deviation (RD %) = 100× (Obs. Productivity- Est. Productivity) / Obs. Productivity)

Forecast Year	Observed Productivity (tonne/ha)	Estimated Productivity (tonne/ha)	Percent relative deviation
2022-23	83.30	85.75	-2.94

Conclusion

In this study, Sugarcane productivity for India is forecasted using ARIMA techniques. From the above results and summary, ARIMA (3, 2, 0) provides satisfactory results for the targeted year. So, it is conclude that ARIMA (3, 2, 0) model be the best model to forecast sugarcane productivity in India. For the next year ie.2023-24, the productivity of sugarcane will be 87.25 tonne per hectare.

References

1. Box GEP, Pierce AD. Distribution of residual autocorrelations in autoregressive integrated moving average time series models. J Am. Stat. Assoc. 1970;65(332):1509-1526.
2. Box GEP, Jenkins GM. Time series analysis: Forecasting and Control. San Francisco: Holden Day; c1976.

3. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: Forecasting and control. San Francisco: Holden Day; c1994.
4. Hossain MM, Abdulla F. Forecasting Potato Production in Bangladesh by ARIMA Model. J Adv. Stat., 2016, 4(1).
5. Khan K, Khan G, Shaikh SA, Lodhi AS, Jilani G. ARIMA modelling for forecasting of rice production: A case study of Pakistan. J Sci. Technol. 2015;4(8):117-120.
6. Kumar A, Deepankar PK, M J, Kumar A. Wheat productivity forecasting in Haryana: A time series approach. Bull Environ Pharmacol. Life Sci. 2019;8(3):63-69.
7. Kumar M, Raman RK, Kumar S. Sugarcane Productivity in Bihar- A Forecast through ARIMA Model. Int. J Pure Appl. Biosci. 2017;5(6):1042-1051.
8. Kumarasinghe HPASS, Peiris BL. Forecasting annual tea production in Sri Lanka. Trop Agric. Res. 2018;29(2):184-193.
9. Mehmood S, Ahmad Z. Time series model to forecast area of mangoes from Pakistan: An application of univariate ARIMA model. Acad. Contemp. Res. J. 2013;2(1):01-07.
10. Padhan PC. Application of ARIMA Model for forecasting agricultural productivity in India. J Agric. Soc. Sci. 2012;8:50-56.
11. Panse VG. Trends in areas and productivity of principal crops in India. Agric. Situat. India. 1952;7:144-48.
12. Panse VG. Recent trends in the productivity of rice and wheat in India. Indian J Agric. Econ. 1959;14:11-38.
13. Panse VG. Productivity trends of rice and wheat in first two five year plans in India. J Indian Soc. Agric. Stat. 1964;16(1):41-50.
14. Suresh KK, Priya SRK. Forecasting Sugarcane Productivity of Tamil Nadu Using ARIMA Models. Sugar Tech. 2011;13(1):23-26.
15. Verma U, Dabas DS, Grewal MS, Singh JP, Hooda RS, Yadav M, *et al*. Crop productivity forecasting in Haryana: 1986 to 2010. Summary Report, Department of Soil Science, CCS HAU, Hisar (Haryana); c2011. p. 141-148.
16. Yule GU. On the method of investigating periodicities in disturbed series, with special reference to wolfers sunspot numbers. Philos. Trans R Soc. London, Ser. A. 1927;226:267-298.