

## Using ARIMA Model to Forecast the Area, Yield and Production of Arhar in Odisha

Balaga Divya<sup>1</sup> and Abhiram Dash<sup>2\*</sup>

<sup>1</sup>P.G. Scholar, Department of Agricultural Statistics, College of Agriculture, OUAT, Bhubaneswar (Odisha), India.

<sup>2</sup>Assistant Professor, Department of Agricultural Statistics, College of Agriculture, OUAT, Bhubaneswar (Odisha), India.

(Corresponding author: Abhiram Dash<sup>\*</sup>)

(Received 09 July 2022, Accepted 11 August, 2022)

(Published by Research Trend, Website: [www.researchtrend.net](http://www.researchtrend.net))

**ABSTRACT:** Arhar is one of the important pulse crop of Odisha. It has very high nutritive value and thus contribute towards the nutritional security of the state. Forecasting of arhar production is very much necessary to enable the agriculture planners to formulate appropriate policies regarding the cultivation of the crop. The present research is carried out on forecasting area, yield and production of arhar in Odisha by using ARIMA model.

ARIMA, the most widely used model for forecasting is used in the study. The data on area, yield and production of arhar are collected from 1970-71 to 2019-20 are used to fit the models found suitable from ACF and PACF plots. The ACF and PACF plots are obtained from stationarized data. The best fit model was selected on basis of significance of estimated coefficients, model diagnostic tests and model fit statistics. The selected best fit model was cross validated by refitting the model by leaving last 5 years, 4 years, upto last 1 year data and obtaining one step ahead forecast for the years 2015-16 to 2019-20. After successful cross validation the selected best fit model is used for forecasting the area, yield and production of arhar in Odisha for the future years 2020-21, 2021-22, 2022-23.

The ARIMA model found to be best fit for area, yield and production of arhar are ARIMA(1,1,2), ARIMA(1,1,0), ARIMA(1,1,0) respectively. All these selected models are fitted without constant as the constant term is insignificant for all these cases. The forecast values shows that area, yield and production of arhar in Odisha remain stagnant in future years with variation in lower and upper class interval of the forecast values.

**Keywords:** ARIMA, cross validation, forecast, model diagnostics, model fit statistics.

### INTRODUCTION

The production of pulses plays a pivotal role in nutritional security as well as agrarian economy of the state of Odisha. The major pulses grown in Odisha are green gram, arhar, horse gram, etc. In Odisha the area under arhar is 130 thousand hectares and its production is 140 thousand MT contributing 7.02% and 10.76% of overall area and production of pulses in Odisha respectively [agricultural statistics at a glance, 2020]. In Odisha the cultivation of arharis concentrated in Cuttack, Puri, Kalahandi, Koraput, Dhenkanal, Balangir, Rayagada, Naupada and Sambalpur districts. Among them Ganjam district stands first with respect to area and production of arhar in Odisha.

As arhar occupies an important position among the pulse crops in Odisha, a timely and accurate forecast of

area and production of such important pulse crops is valuable in terms of agricultural policy decisions and food and nutritional security of the people. Various researchers have been contributing in this area of research. Mishra *et al.* (2021) studied the trend in the production of total pulses in major growing states in India using ARIMA. Vishwajit *et al.* (2018) studied about the modelling and forecasting of arhar in major arhar growing states in India using ARIMA and other models. Shah *et al.* (2017) conducts a study to forecast the substantial food crop production in Khyber Pakhtunkhwa, Pakistan; the secondary data were utilized by applying ARIMA forecasting strategy. They found that the outcome of the ARIMA model was sufficient. The present study focuses on the forecasting the area, yield and production of arhar crop in Odisha

using ARIMA models. Kumari *et al.* (2022) deals with comparison of different statistical models to predict area, production and yield of citrus in Gujarat state. The study revealed that ARIMA model was superior to explain the area, production and yield of citrus.

## MATERIALS AND METHODS

The secondary data on area, yield and production of arhar are collected for the state of Odisha (kharif and rabi seasons combined) for the period 1970-71 to 2019-20 from *Five Decades of Odisha Agriculture Statistics* published by Directorate of Agriculture and Food Production, Odisha.

Autoregressive Integrated moving Average is a statistical model which is used to predict the future trends. The ARMA models, which includes the order of differencing (which is to stationarize the data) is known as Autoregressive integrated moving average (ARIMA) models. A non-seasonal ARIMA model is classified as an "ARIMA (p,d,q)" model, where, the parameters p,d,q are the non-negative integers where p is the number of autoregressive terms, d is the number of nonseasonal differences necessary for stationarizing the data, and q is the number of moving average terms. Thus, the ARIMA (p,d,q) model can be represented by the following general forecasting equation:

$$Y_t = \mu + \sum_{i=1}^p \alpha_i Y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Where  $\mu$  is a mean,  $\alpha_1, \alpha_2, \dots, \alpha_p$  and  $\theta_1, \theta_2, \dots, \theta_q$  are the parameters of the model, p is the order of autoregressive term and q is the order of moving average term and  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-j}$  are noise error terms.

### Model identification:

The ARIMA model is estimated only if the data under study is stationary, it can be tested by using Augmented Dickey-Fuller test. If it is not stationary then it should be converted into stationary series by differencing the data at suitable lag. Usually the data is stationarized

after 1 or 2 differencing. After stationarizing the data, the Auto Correlation Factor(ACF) and Partial Auto Correlation Factor(PACF) plots are used to identify tentative Auto Regression (AR) and Moving Average (MA) orders. Various tentative models based on identified AR and MA orders are fitted and parameters are estimated. After fitting the tentative models normality and independency of the residuals of the fitted models is tested by using Shapiro-Wilk's test and Box-pierce test respectively. Then the tentative models satisfying the normality and independency of residuals are compared by using the model fit statistics such as mean absolute percentage error(MAPE), root mean square error(RMSE) and Akaike's Information criteria corrected (AICc) which are mathematically as follows:

$$\text{Root mean square error (RMSE): } \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y_t)^2}{n}}$$

$$\text{Mean absolute percentage error (MAPE): } \frac{100}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \text{ (Mishra, et al., 2020)}$$

Where  $\hat{y}_t$  = forecasted value,  $y_t$  = actual value and n = number of times the summation iteration happens

$$\text{Akaike's information criteria corrected (AICc): } \text{AIC} + \frac{2K^2 + 2K}{n - k - 1}$$

Where k denotes the number of parameters and n denotes the sample size.

The model with lowest RMSE, MAPE and AICc values is selected as the best fit ARIMA model among selected tentative models and it is taken for forecasting.

## RESULTS AND DISCUSSION

The data on area, yield and production of arhar crop was tested for the presence of stationarity by using Augmented Dickey Fuller test and the results are presented in Table 1. The test results confirmed that the data was not stationary and made stationary by differencing at lag 2.

**Table 1: Test of stationarity of data on area, yield and production of arhar in Odisha.**

Variable	Original series		First order differenced series	
	ADF test statistic	P value	ADF test statistic	P value
Area	-1.474	0.782	-3.359	0.045
Yield	-2.083	0.540	-3.799	0.028
Production	-1.603	0.733	-4.864	0.010

The next step was to identify the order of AR and MA terms such as p and q using the ACF and PACF plots. Different tentative models were identified using the

orders of AR and MA terms. Figs. 1, 2 and 3 shows the ACF and PACF plots of first order difference of area, yield and production of arhar crop in Odisha.

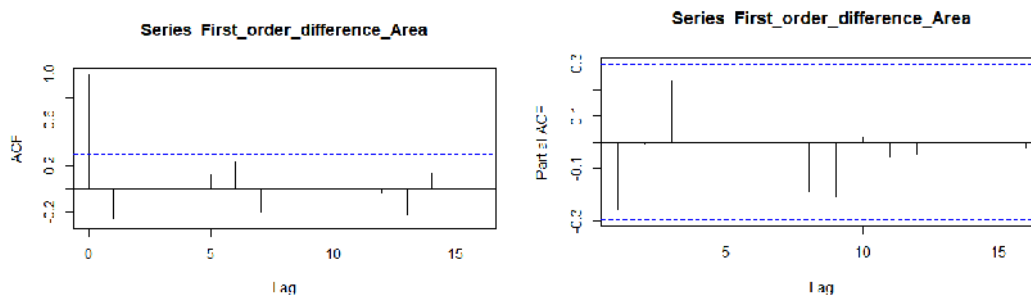


Fig. 1. ACF and PACF plot of first order difference of area under arhar.

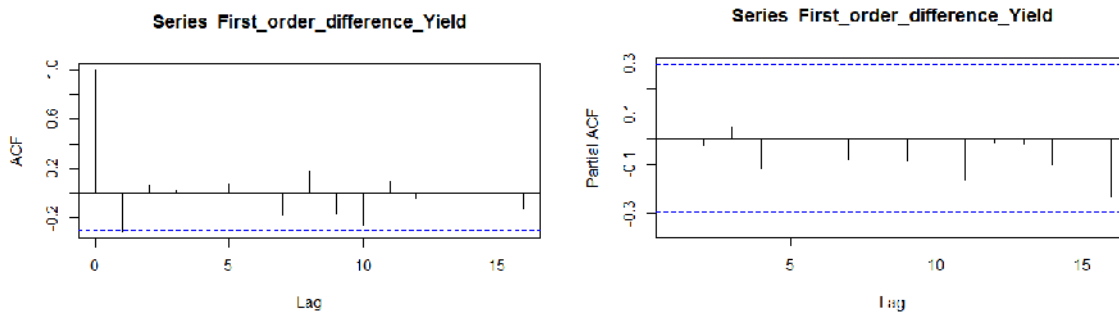


Fig. 2. ACF and PACF plot of first order difference of yield of arhar.

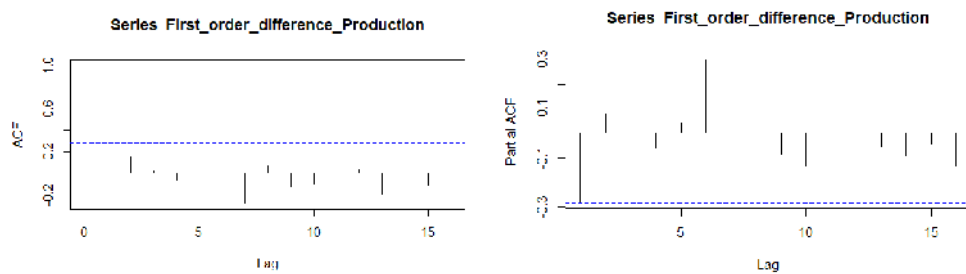


Fig. 3. ACF and PACF plot of first order difference of production of arhar.

The tentative models of area and their estimated coefficients along with error measures are shown in the Table 2. The study of the table reveals that

ARIMA(1,1,2) without constant model has all the estimated coefficients significant.

Table 2: Parameter estimates of the selected ARIMA (p,d,q) model fitted to area under arhar for Odisha.

ARIMA(p,d,q)	Constant	$\Phi_1$	$\Phi_2$	$\Phi_3$	$\theta_1$	$\theta_2$	$\theta_3$
ARIMA (1,1,2)	---	0.825** (0.113)	---	---	1.252** (0.165)	0.548** (0.141)	---
ARIMA (2,1,2)	---	1.137** (0.440)	-0.290 (0.358)	---	-1.494** (0.371)	0.752** (0.270)	---
ARIMA (1,1,3)	---	0.783** (0.157)	---	---	-1.147** (0.214)	0.353 (0.236)	0.154 (0.159)
ARIMA (1,1,0)	---	-0.230 (0.139)	---	---	---	---	---
ARIMA (0,1,1)	---	---	---	---	-0.184 (0.116)	---	---

Figures inside the parentheses represents the standard error of the parametric estimators.

\*\*'- at 5% significance level, \*\*\*'- at 1% significance level

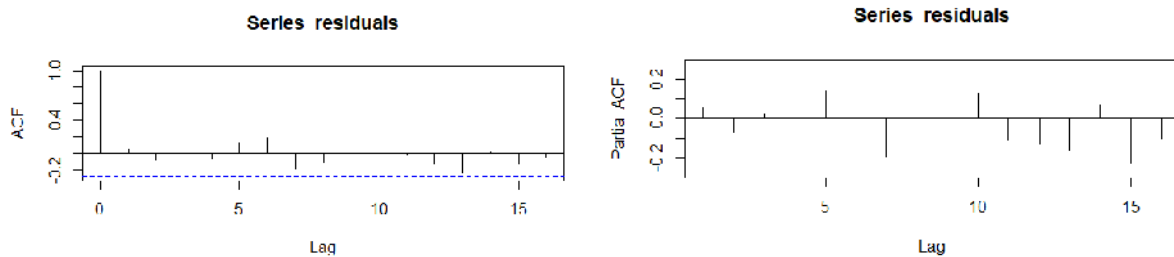
Table 3 shows the model diagnostics test and model fit statistics for the fitted ARIMA models. ARIMA (1,1,2) model satisfies both the test of normality and

independency of residuals. Thus this model is selected to be the best fit model for area under arhar crop.

**Table 3: Model fit statistics of the ARIMA (p,d,q) model fitted to area under arhar.**

ARIMA(p,d,q)	Shapiro-wilk test		Box – pierce test		AICc	RMSE	MAPE
	W	p-value	$\chi^2$	p-value			
ARIMA (1,1,2)	0.891	0.673	0.131	0.717	379.99	10.28	6.23
ARIMA (2,1,2)	0.876	0.079	0.047	0.828	381.40	10.34	6.47
ARIMA (1,1,3)	0.888	0.002	0.009	0.922	381.55	10.37	6.61
ARIMA (1,1,0)	0.896	0.043	0.061	0.804	382.20	11.32	7.07
ARIMA (0,1,1)	0.903	0.819	0.263	0.607	382.60	11.38	7.13

W - Shapiro-wilk test statistic  $\chi^2$  - Box – Pierce test statistic



**Fig. 4.** ACF and PACF of residuals from selected ARIMA (1,1,2) model for area under arhar.

The tentative ARIMA models of yield and their estimated coefficients along with error measures are shown in the Table 4. The study of the table reveals that

ARIMA (1,1,0) and ARIMA (0,1,1) without constant model has all the estimated coefficients significant.

**Table 4: Parameter estimates of the ARIMA (p,d,q) model fitted to yield of arhar for Odisha.**

ARIMA(p,d,q)	Constant	$\Phi_1$	$\Phi_2$	$\Phi_3$	$\theta_1$	$\theta_2$	$\theta_3$
ARIMA (1,1,0)	---	-0.408** (0.129)	---	---	---	---	---
ARIMA(0,1,1)	---	---	---	---	-0.283** (0.105)	---	---
ARIMA (0,1,0)	---	---	---	---	---	---	---
ARIMA (1,1,0)	10.372 (8.445)	-0.269 (0.143)	---	---	---	---	---
ARIMA (0,1,1)	10.235 (7.910)	---	---	---	-0.265 (0.144)	---	---

Figures inside the parentheses represents the standard error of the parametric estimators.

\*\* - at 5% significance level, \*\*\* - at 1% significance level

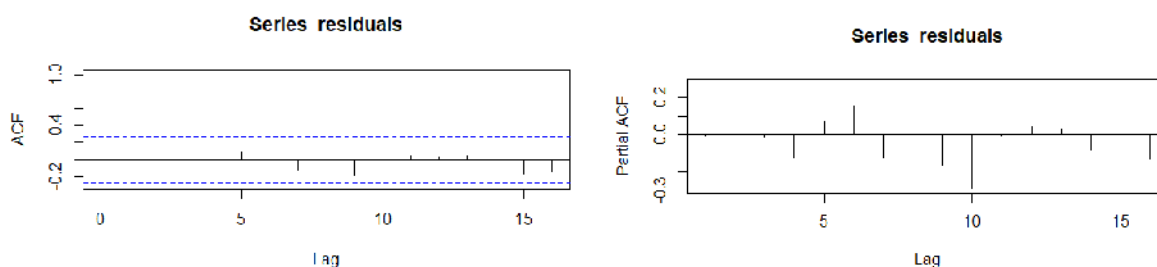
Table 5 shows the model diagnostics test and model fit statistics for the fitted ARIMA models for yield of arhar. ARIMA (1,1,0) without constant model satisfies both the test of normality and independency of residuals. The RMSE, MAPE and AICc are less for ARIMA (1,1,0) without constant model. Thus, this

model is selected to be the best fit model for production of arhar crop. Fig. 5 also shows that none of the autocorrelations and partial autocorrelations of residuals are significant. This further confirms the selection of the respective best fit models.

**Table 5: Model fit statistics of the ARIMA (p,d,q) model fitted to yield of arhar.**

ARIMA(p,d,q)	Shapiro-wilk test		Box – pierce test		AICc	RMSE	MAPE
	W	p-value	<sup>2</sup>	p-value			
ARIMA (1,1,0)	0.948	0.088	0.002	0.966	567.59	30.914	7.532
ARIMA (0,1,1)	0.954	0.052	0.040	0.841	567.84	31.877	8.018
ARIMA (0,1,0)	0.966	0.165	2.928	0.087	568.17	77.275	8.855
ARIMA (1,1,0)	0.948	0.028	0.007	0.932	568.41	73.970	8.219
ARIMA (0,1,1)	0.957	0.071	0.0003	0.986	568.55	74.077	8.314

W - Shapiro-wilk test statistic <sup>2</sup> - Box – Pierce test statistic



**Fig. 5.** ACF and PACF of residuals from selected ARIMA (1,1,0) model for yield of arhar.

The tentative models of production and their estimated coefficients along with error measures are shown in the Table 6. The study of the table reveals that ARIMA

(0,1,1) and ARIMA (1,1,0) constant model has the estimated coefficients significant.

**Table 6: Parameter estimates of the ARIMA (p,d,q) model fitted to production of arhar for Odisha.**

ARIMA(p,d,q)	Constant	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\theta_1$	$\theta_2$	$\theta_3$
ARIMA (1,1,0)	2.363 (1.512)	- 0.286* (0.136)	----	----	----	----	----
ARIMA (1,1,0)		-0.249 (0.137)	----	----	----	----	----
ARIMA (0,1,1)	----	----	----	----	-0.191 (0.117)	----	----
ARIMA (0,1,1)	2.342 (1.512)	----	----	----	-0.231* (0.119)	----	----
ARIMA (0,1,0)	----	----	----	----	----	----	----

Figures inside the parentheses represents the standard error of the parametric estimators.

\*- at 5% significance level, \*\*\*- at 1% significance level

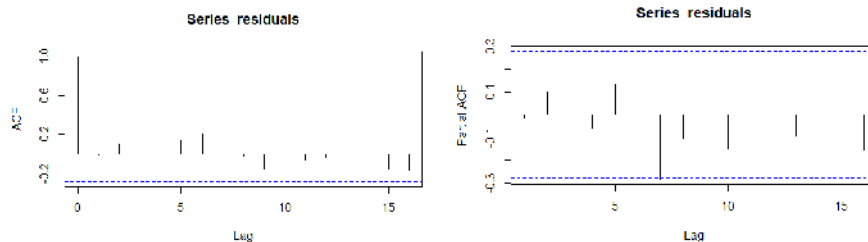
Table 7 shows the model fit statistics and model diagnostics test for the fitted ARIMA models for production of arhar. ARIMA (1,1,0) with constant

model satisfies both the test of normality and independency of residuals. Thus this model is selected to be the best fit model for production of arhar crop.

**Table 7: Model fit statistics of the ARIMA (p,d,q) model fitted to production of arhar.**

ARIMA(p,d,q)	Shapiro-wilk test		Box – pierce test		AICc	RMSE	MAPE
	W	p-value	<sup>2</sup>	p-value			
ARIMA (1,1,0)	0.856	0.213	0.016	0.896	401.20	13.426	12.053
ARIMA (1,1,0)	0.861	0.076	0.017	0.896	401.23	13.751	12.834
ARIMA (0,1,1)	0.872	0.032	0.349	0.554	401.99	13.862	13.135
ARIMA (0,1,1)	0.867	0.612	0.083	0.772	402.04	13.547	13.306
ARIMA (0,1,0)	0.901	0.012	4.101	0.042	402.26	14.216	14.241

W - Shapiro-wilktest statistic <sup>2</sup> - Box – Pierce test statistic



**Fig. 6.** ACF and PACF of residuals from selected ARIMA (1,1,0) model for production of arhar.

**Cross Validation.** In the following table, we have done stepwise cross validation by considering the best fitted ARIMA model of respective variables of arhar crop of Odisha. The APE(absolute percentage error) of area under arhar is found to be in the range between 0 to 11 and the MAPE(mean APE) is found to be 3.8544 for

area under arhar crop. Similarly for yield the APE range is found between 0.5 to 13 and MAPE is 5.688 and that for production, APE range is between 1 to 17 and MAPE is 5.0659. These results shows that the selected ARIMA models are successfully cross validated.

**Table 8: Cross validation of area, yield and production of arhar.**

Year	Area			Yield			Production		
	Actual	Predicted	APE	Actual	Predicted	APE	Actual	Predicted	APE
2015-16	138.29	137.87	0.31	886	897.37	1.28	122.52	123.98	1.19
2016-17	135.54	138.29	2.02	884	889.77	0.65	119.82	122.86	2.54
2017-18	137.89	135.54	1.71	897	884.63	1.38	123.69	120.54	2.54
2018-19	144.06	137.89	4.28	1022	892.91	12.63	147.23	122.65	16.69
2019-20	128.63	142.718	10.95	1124	983.56	12.49	144.58	141.17	2.35
MAPE	1.854			5.688			5.065		

The appropriate ARIMA models which are represented in the previous tables were used to forecast the area,

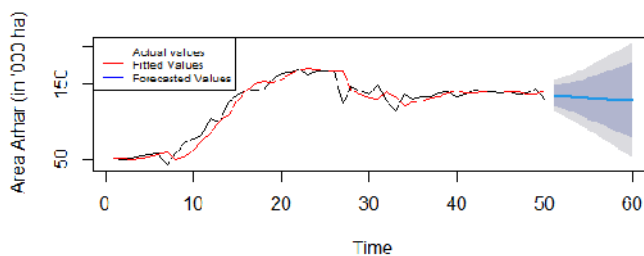
yield and production of arhar crop in Odisha for the years 2020-21, 2021-22 and 2022-23.

**Table 9: Forecast values of area, yield and production of arhar in Odisha for the year 2020-21 to 2022-23.**

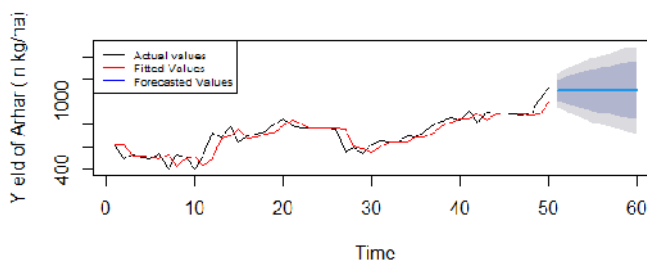
Year	Area ('000ha)			Yield (kg/ha)			Production ('000tonnes)		
	Forecasted	95 % confidence interval		Forecasted	95 % confidence interval		Forecasted	95 % confidence interval	
		Low CI	High CI		Low CI	High CI		Low CI	High CI
2020-21	128.63	105.58	151.67	1124	971.01	1276.99	141.16	113.39	168.94
2021-22	128.63	96.04	161.21	1124	907.63	1340.36	142.72	108.12	177.33
2022-23	128.63	88.72	168.54	1124	859.01	1388.99	142.32	101.07	183.58

CI denotes the class interval

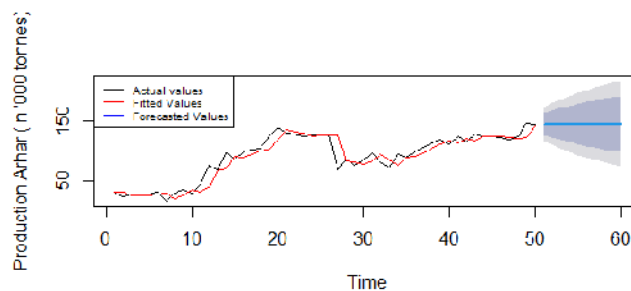
Figures 7,8, and 9 shows the actual, fitted and forecast values of area, yield and production of arhar in Odisha



**Fig. 7.** Actual with fitted and forecasted values of area under arhar from ARIMA (1,1,2) without constant model.



**Fig. 8.** Actual with fitted and forecasted values of yield of arhar from ARIMA (1,1,0) without constant model.



**Fig. 9.** Actual with fitted and forecasted values of production of arhar from ARIMA (1,1,0) without constant model.

## CONCLUSION

ARIMA (1,1,2) without model, ARIMA (1,1,0) without constant model and ARIMA (1,1,0) without constant model are found to be the best fit model for area, yield and production of arhar in Odisha. These selected models are used for forecasting of area, yield and production of arhar in Odisha. The forecast values shows that area, yield and thus production of arhar in Odisha remain stagnant in future years with variation in lower and upper class interval of the forecast values.

## FUTURE SCOPE

The study regarding use of ARIMA model in forecasting can be further compared with Artificial Neural Network models in case of data having both linear and non-linear pattern.

**Acknowledgement.** The authors are highly thankful to the staffs of Department of Agril. Statistics, C.A., Bhubaneswar for their moral support in conducting the research study.

**Conflict of Interest.** None.

## REFERENCES

- Mishra, P., Khatib, A, Sardar, I., Mohammad, J., Karakaya, K., Dash, A., Ray, M., Narshimhaiah, L. and Dubay, A. (2020). Modeling and Forecasting of Sugarcane Production in India. *Sugar Tech*, 23(6): 1317-1324
- Mishra, P., Yonar, A., Yonar, H., Kumari, B., Abotale, M., Das, S. S. and Patil, S. G. (2021). State of the art in total pulse production in major states of India using ARIMA techniques. *Current research in food science*, (4): 800-806
- Kumari, Prity, Parmar, D. J., Kumar, Sathish, M., Mahera, A. B. and Lad, Y. A. (2022). Comparison of Statistical Models for Prediction Area, Production and Yield of Citrus in Gujarat. *Biological Forum – An International Journal*, 14(2): 690-695.
- Shah, S. A. A. Zeb, N. and Alamgir, K. (2017). Forecasting major food crops production in Khyber Pakhtunkhwa, Pakistan. *Journal of Applied and Advanced Research*, 2 (1): 21–30.
- Vishwajith, K. P., Sahu, P. K., Mishra, P., Dhekale, B. S. and Singh, R. B. (2018). Modelling and forecasting of arhar production in India. *International journal of agricultural statistical sciences*, (14):73-86.

**How to cite this article:** Balaga Divya and Abhiram Dash (2022). Using ARIMA Model to Forecast the Area, Yield and Production of Arhar in Odisha. *Biological Forum – An International Journal*, 14(3): 1179-1185.