ISSN (online): 2320-9836
ISSN (Print): 2320-9828

Trend and forecasting analysis of area, production and productivity of total pulses in India

Pushpa M. Savadatti

Department of Economic Studies and Planning, Central University of Karnataka, Kalburgi, Karnataka State
pmsavadatti@gmail.com

Abstract

Objectives: To study the trends in growth of area, production and productivity of pulses over a period of time and to forecast the values for these variables.
Methodology: Adopted to study the first objective is to calculate the compound growth rates with the help of exponential growth model. For the second objective univariate time series analysis based on Box-Jenkins (BJ) approach which is popularly known as ARIMA process has been adopted to forecast the values for area, production and productivity of pulses. The present analysis is based on the annual time series data collected from Centre for Monitoring Indian Economy (CMIE) for the period 1966-67 to 2015-16.
Findings: The growth rate analysis indicated that during the study period there is positive and significant growth in area, production and productivity of pulses but relative growth in area is lower compared to that of production and yield. The empirical analysis based on ARIMA indicated that the specifications (i) AR (1), AR (2), MA (1) and MA (2) (ii) AR (1), AR (10), MA (1) and MA (12) (iii) AR (12) and MA (1) were best fitted models for forecasting area, production and productivity of total pulses in India respectively. The model adequacy criteria like correlogram of residuals, root mean square errors, mean absolute percentage error, Theil's inequality coefficient, normality and heteroskedasticity assumptions of residuals validated the results. Based on the selected ARIMA specifications projections for next five years starting from 2016-17 onwards have been made. The forecast results show the increasing trend in production and productivity of pulses but the area under pulse crops shows the growth of near stagnancy during forecast period.
Application: Inferences drawn from the results are firstly these projections help the policy makers and farmers in their decision making. Secondly, concerted efforts are required for effective implementation of government programmes to enhance the area under pulse crops which are influenced by multiplicity of factors and risks, to reduce the widening gap between supply of and demand for pulses in the future.
Keywords: Trend, ARIMA, forecasting, pulses.

1. Introduction

Agriculture is the backbone of the Indian economy as it provides employment to nearly two third of the rural population. Cultivation of pulse crops play a very important role in sustainable agriculture due to their numerous important qualities like improves soil fertility and physical structure, can be grown in mixed / inter-cropping systems, and requires less water as it is a rain fed crop and above all a rich source of vegetable protein for the rural mass. India has the largest share in the world's area and production of pulses. Despite all this, growth rates in area, production and yield of pulse crops in the country have not shown many improvements over the years. The government of India has launched various programmes to increase pulse area, production and productivity. To mention programmes like Pulse Development Scheme (PDS) during IVth plan, National Pulses Development Project (NPSP) during VIIth Plan, Special Food-grains Production Programmer (SFPP) on pulses during 1988-89, Technology Mission on Oilseeds and Pulses (TMOP) during 1990-91, Integrated Schemes of Oilseeds, Pulses, Oil-palm and Maize (ISOPOM) during 2004-05, are few among many implemented by the government. In view of this it would be of interest to analyze the current status of growth in pulse area, production and yield with the help of growth rate analysis and to provide the future trends in these variables with the help of the univariate ARIMA analysis.

ISSN (online): 2320-9836
ISSN (Print): 2320-9828

2. Material and methods

In literature Box-Jenkin's approach popularly known as Auto Regressive Integrated Moving Average (ARIMA) approach has been used very extensively for modeling time series data for agricultural crops. An ARIMA model is used to forecast cultivated area and production of maize in Nigeria and study concluded that projections will help the Nigerian government to make policies with regards to relative price structure, production, consumption pattern, etc. [1]. Another study conducted in Bangladesh, applied ARIMA model for rice production in Bangladesh [2]. An empirical study on agricultural products' price forecasting based on ARIMA model in China showed that ARIMA model can provide high accuracy of short term prediction for cucumber prices in Shandong wholesale market, China [3]. There are number of studies in India also adopted ARIMA process to forecast area, production and yield of different agricultural products in India [4-9]. The present study also used ARIMA modeling to forecast area, production and productivity of total pulses grown in India. The analysis was based on the annual time series data collected for the period 1949-2015 from Centre for Monitoring Indian Economy (CMIE) data sources on area, production and productivity of pulses at all India level.

2.1. Growth rate analysis

To study the trends in growth of pulse area, production and yield an exponential form of growth function as specified below was employed [10].

$Y_t = AB^T U_t$ --------------------------------- (1)

Where,

Y_t = Pulse area, production and yield in the year 't'

$A =$ Intercept indicating Y in the base period

$T =$ Time Variable taking values $1949-50=1, 1950-51=2, \ldots\ldots, 2015-16=67$

$U_t =$ Error term

On taking the natural logarithms equation (1) becomes linear as follows:

$lnY_t = \ln A + T \ln B + lnU_t$ -------------------------- (2)

Equation (2) is written as $Q_t = a + b\,T + V_t$ -------------------- (3)

$Where, \quad Q_t = lnY_t\,; \; a = \ln A; b = \ln B; \; V_t = \ln U_t$

The values of 'a' and 'b' are estimated with the help of Ordinary Least Squares Method (OLS) by applying E-views software.

Then the compound growth rate (CGR) was obtained as follows:

$G = (B-1) * 100$

Where, B is the antilog of 'b'.

2.2. The Box –Jenkins (BJ) methodology

The present study applies BJ forecasting model popularly known as Auto Regressive Integrated Moving Average (ARIMA) models. The BJ methodology consists of four steps [11].

2.2.1. Identification

This step involves the identification of ARIMA (p,d,q) model where, p denotes the number of autoregressive terms, d the number of times the time series data has to be differenced to become stationary and d indicates the number of moving average terms. Thus this step is to find the appropriate values for p,d and q with the help of Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) and correlograms, which are the plots of ACFs and PACFs against the lag length.

2.2.2. Estimation

The BJ methodology is applicable to stationary time-series hence; it is necessary to convert the non-stationary series into stationary series before estimation of the model. Once we identify the appropriate values for p and q, the next step is to estimate the parameters of the ARIMA model that has been selected provisionally with the help of the appropriate estimation methods.

 ISSN (online): 2320-9836
ISSN (Print): 2320-9828

2.2.3. Diagnostic checking

It is necessary to test whether the estimated model fits the data reasonably well or not based on the correlogram of the estimated residuals from the selected model. If the estimated residuals are white noise, then the chosen model may be accepted otherwise another ARIMA model may be selected starting over from the stage of identification. Thus, BJ methodology is an iterative process. These three stages will be repeated until we get a satisfactory ARIMA model.

2.2.4. Forecasting

Once the model is found satisfactory after diagnostic checking the chosen model will be used for forecasting future values. ARIMA models are popular because of their forecasting accuracy.

The mixed autoregressive-moving average process of order (p,q) can be denoted as ARMA (p,q) and can be expressed for the process Y_t as:

$$Y_t = \phi_1 Y_{t-1} + \ldots\ldots\ldots\ldots\ldots\ldots \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots\ldots\ldots\ldots - \theta_q \varepsilon_{t-q} \text{------------------- (4)}$$

$$where,\ \varepsilon_t\ is\ WN\ (0, \sigma^2); WN = white\ noise$$

The annual time series data for the period 1949-50 to 2015-16 have been used for the analysis with the help of E-views software to identify the suitable ARIMA model for forecasting the area, production and productivity of pulse crops in India.

3. Results and discussions

3.1. Growth rate analysis

The results of growth rate analysis indicate that there is significant positive growth rate in area, production and productivity of total pulses in India during the study period (Table 1).

Table 1. Compound growth rates of total pulses in India during 1949-50 to 2015-16

Particulars	Area	Production	Yield
CGR	0.0959**	0.7537*	0.6678*
t-value	2.5069	9.8590	11.5983
P-value	0.0147	0.0000	0.0000
R^2	0.09**	0.60*	0.67*
F -value	6.2858	97.2001	134.5215

,Indicate significant at 1% and 5% level of probability respectively*

The area under total pulses showed positive growth rate during the study period though significant at 5% level the value is not much compared to production and yield which showed positive and significant growth rates and have higher value indicating that the increase in the production has come mainly from the significant increase in yield.

3.2. Forecasting analysis (ARIMA)

3.2.1. Model Identification

The first step in model identification is to check whether the series under consideration are stationary or not. For this purpose, ACF and PACF of the area, production and productivity of total pulses and Augmented Dicky Fuller (ADF) tests were adopted and the results of the ADF tests are presented in the Table 2. Figure 1 depicts the ACF and PACF of area under total pulses which are stationary at levels and ACF cuts off after third lag and PACF after first lag but ACF at lags 10 and 11 seem to be significant. ACF and PACF of production of total pulses which is stationary after first difference are presented in Figure 2.

ACF and PACF both cut off after first lag but it appears that ACF at lag 12 may be significant and PACF at lag 10 may be significant. Figure 3 indicates the ACF and PACF of yield of total pulses which is stationary after first difference. It is clear from this figure that both ACF and PACF cut off after first lag but ACF at lag 12 appears to be significant.

 ISSN (online): 2320-9836
ISSN (Print): 2320-9828

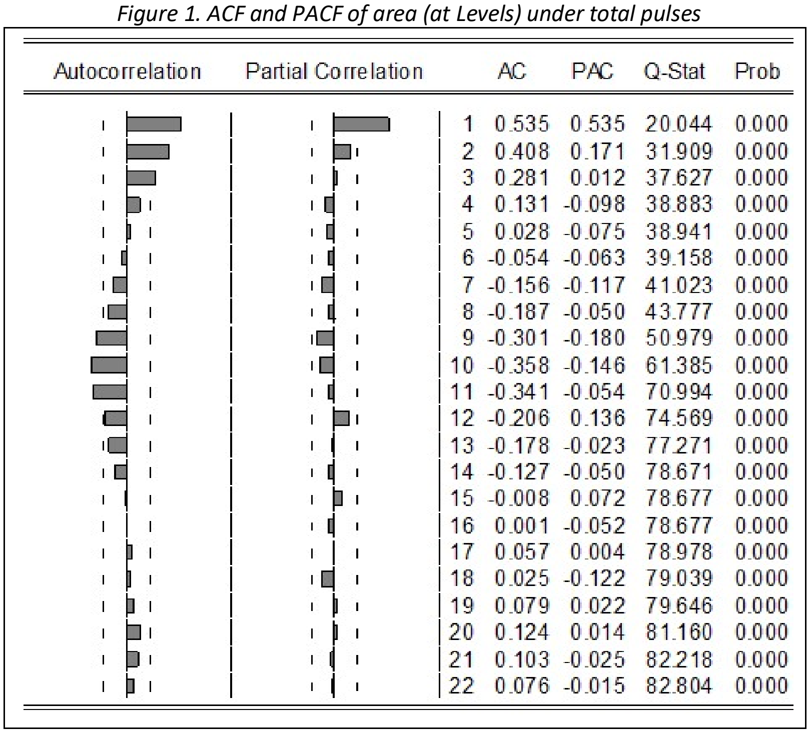
Table 2. ADF Test of stationary for area, production and yield of total pulses in India

Area under Total Pulses (at levels) with constant			
		t-statistics	Probability *
ADF test statistic		-4.338296	0.0009
Test critical values	1% level	-3.533204	
	5% level	-2.906210	
	10% level	-2.590628	
Production of Total Pulses (at levels) with constant			
		t-statistics	Probability *
ADF test statistic		-1.658068	0.4475
Test critical values	1% level	-3.534868	
	5% level	-2.906923	
	10% level	-2.591006	
Production of Total Pulses (first difference) with constant			
		t-statistics	Probability *
ADF test statistic		-14.52556	0.0000
Test critical values	1% level	-2.601024	
	5% level	-1.945903	
	10% level	-1.613543	
Yield of Total Pulses (at levels) with constant			
		t-statistics	Probability *
ADF test statistic		-1.392059	0.5809
Test critical values	1% level	-3.534868	
	5% level	-2.906923	
	10% level	-2.591006	
Yield of Total Pulses (first difference) with constant			
		t-statistics	Probability *
ADF test statistic		-14.25371	0.0000
Test critical values	1% level	-2.601024	
	5% level	-1.945903	
	10% level	-1.613543	

**Mackinnon (1996) one -sided p-values*

Various specifications of the ADF tests have been tried to test for stationarity of the study variables and the final results presented in the Table 2 indicate that Area is stationary at levels and Production and Yield are stationary at first difference. The correlogram of the stationary series are presented in the Figures 1-3.

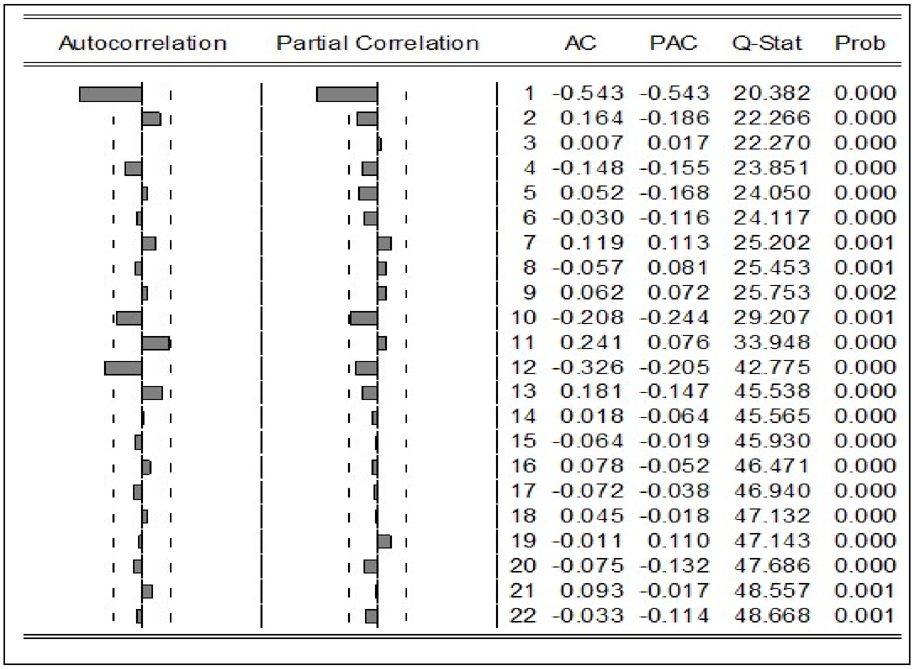
Figure 1. ACF and PACF of area (at Levels) under total pulses



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.535	0.535	20.044	0.000
		2	0.408	0.171	31.909	0.000
		3	0.281	0.012	37.627	0.000
		4	0.131	-0.098	38.883	0.000
		5	0.028	-0.075	38.941	0.000
		6	-0.054	-0.063	39.158	0.000
		7	-0.156	-0.117	41.023	0.000
		8	-0.187	-0.050	43.777	0.000
		9	-0.301	-0.180	50.979	0.000
		10	-0.358	-0.146	61.385	0.000
		11	-0.341	-0.054	70.994	0.000
		12	-0.206	0.136	74.569	0.000
		13	-0.178	-0.023	77.271	0.000
		14	-0.127	-0.050	78.671	0.000
		15	-0.008	0.072	78.677	0.000
		16	0.001	-0.052	78.677	0.000
		17	0.057	0.004	78.978	0.000
		18	0.025	-0.122	79.039	0.000
		19	0.079	0.022	79.646	0.000
		20	0.124	0.014	81.160	0.000
		21	0.103	-0.025	82.218	0.000
		22	0.076	-0.015	82.804	0.000

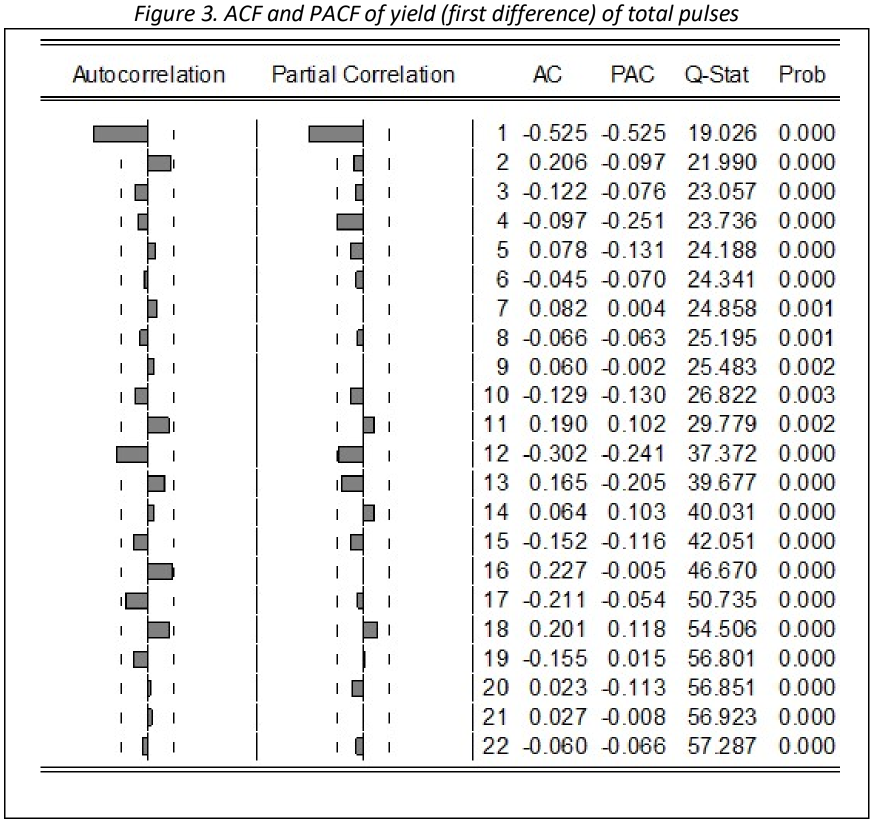
 ISSN (online): 2320-9836
ISSN (Print): 2320-9828

Figure 2. ACF and PACF of production (first difference) total pulses



	AC	PAC	Q-Stat	Prob
1	-0.543	-0.543	20.382	0.000
2	0.164	-0.186	22.266	0.000
3	0.007	0.017	22.270	0.000
4	-0.148	-0.155	23.851	0.000
5	0.052	-0.168	24.050	0.000
6	-0.030	-0.116	24.117	0.000
7	0.119	0.113	25.202	0.001
8	-0.057	0.081	25.453	0.001
9	0.062	0.072	25.753	0.002
10	-0.208	-0.244	29.207	0.001
11	0.241	0.076	33.948	0.000
12	-0.326	-0.205	42.775	0.000
13	0.181	-0.147	45.538	0.000
14	0.018	-0.064	45.565	0.000
15	-0.064	-0.019	45.930	0.000
16	0.078	-0.052	46.471	0.000
17	-0.072	-0.038	46.940	0.000
18	0.045	-0.018	47.132	0.000
19	-0.011	0.110	47.143	0.000
20	-0.075	-0.132	47.686	0.000
21	0.093	-0.017	48.557	0.001
22	-0.033	-0.114	48.668	0.001

Figure 3. ACF and PACF of yield (first difference) of total pulses



	AC	PAC	Q-Stat	Prob
1	-0.525	-0.525	19.026	0.000
2	0.206	-0.097	21.990	0.000
3	-0.122	-0.076	23.057	0.000
4	-0.097	-0.251	23.736	0.000
5	0.078	-0.131	24.188	0.000
6	-0.045	-0.070	24.341	0.000
7	0.082	0.004	24.858	0.001
8	-0.066	-0.063	25.195	0.001
9	0.060	-0.002	25.483	0.002
10	-0.129	-0.130	26.822	0.003
11	0.190	0.102	29.779	0.002
12	-0.302	-0.241	37.372	0.000
13	0.165	-0.205	39.677	0.000
14	0.064	0.103	40.031	0.000
15	-0.152	-0.116	42.051	0.000
16	0.227	-0.005	46.670	0.000
17	-0.211	-0.054	50.735	0.000
18	0.201	0.118	54.506	0.000
19	-0.155	0.015	56.801	0.000
20	0.023	-0.113	56.851	0.000
21	0.027	-0.008	56.923	0.000
22	-0.060	-0.066	57.287	0.000

3.2.2. Model estimation

Based on the various criteria like Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), significance of the parameters, R^2, F-statistics, etc., the optimal ARMA model selected for Area, Production and Yield are presented in the Table 3. Table 3 presents the results of the final ARIMA model selected for the three variables Area, Production and Yield of total pulses with the help of e-views software.

ISSN (online): 2320-9836
ISSN (Print): 2320-9828

Table 3. Optimal ARIMA model selected for area, production and yield of total pulses

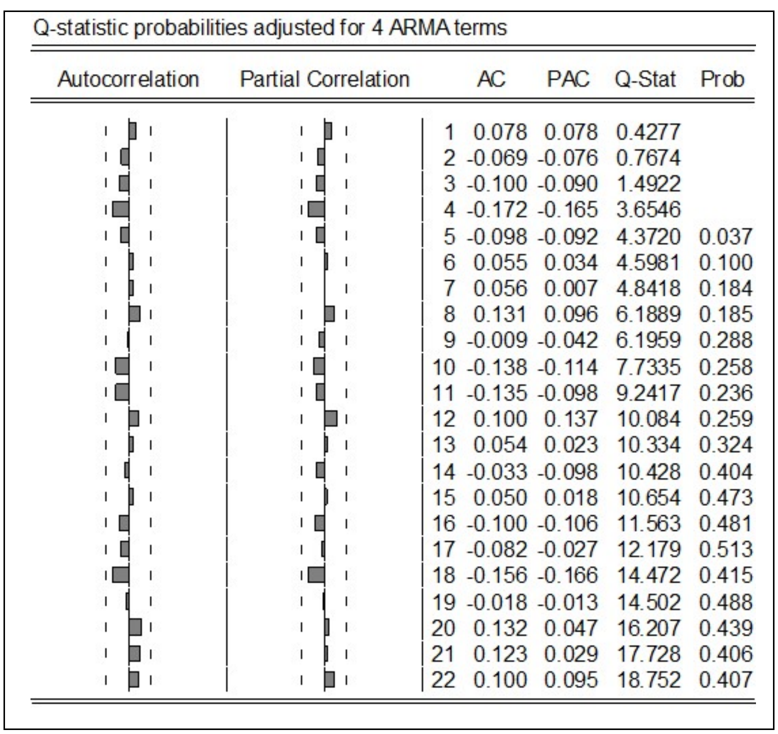
ARMA model for area under total pulses				
Variable	Coefficient	Standard Error	t-Statistics	Probability
C	22768.79	257.8936	88.28751	0.0000
AR(1)	1.867812***	0.028891	64.64938	0.0000
AR(2)	-0.967183***	0.030353	-31.86483	0.0000
MA(1)	-1.727873***	0.073298	-23.57315	0.0000
MA(2)	0.927617***	0.076317	12.15486	0.0000
R^2	0.4809***		AIC	16.90797
Adjusted R^2	0.4383		SIC	17.10540
F-statistics	11.3016			
Probability (F-statistic)	0.0000			
ARMA model for production of total pulses				
Variable	Coefficient	Standard Error	t-Statistics	Probability
C	104.8659	36.91153	2.841007	0.0061
AR(1)	-0.200526	0.230361	-0.870487	0.3875
AR(10)	-0.277461*	0.146348	-1.895898	0.0628
MA(1)	-0.451290*	0.248213	-1.818158	0.0740
MA(12)	-0.347016**	0.131723	-2.634434	0.0107
R^2	0.4497***		AIC	17.29794
Adjusted R^2	0.4038		SIC	17.49700
F-statistics	9.8048			
Probability (F-statistic)	0.000001			
ARMA model for yield of total pulses				
Variable	Coefficient	Standard Error	t-Statistics	Probability
C	4.056306	1.363706	2.974472	0.0042
AR(12)	-0.305126**	0.121134	-2.518905	0.0144
MA(1)	-0.698048***	0.088117	-7.921801	0.0000
R^2	0.3972***		AIC	10.53161
Adjusted R^2	0.3680		SIC	10.66432
F-statistics	13.6184			
Probability (F-statistic)	0.000001			

*******Indicate significance at 1%, 5% and 10% respectively*

3.2.3. Model adequacy check (Diagnostic check)

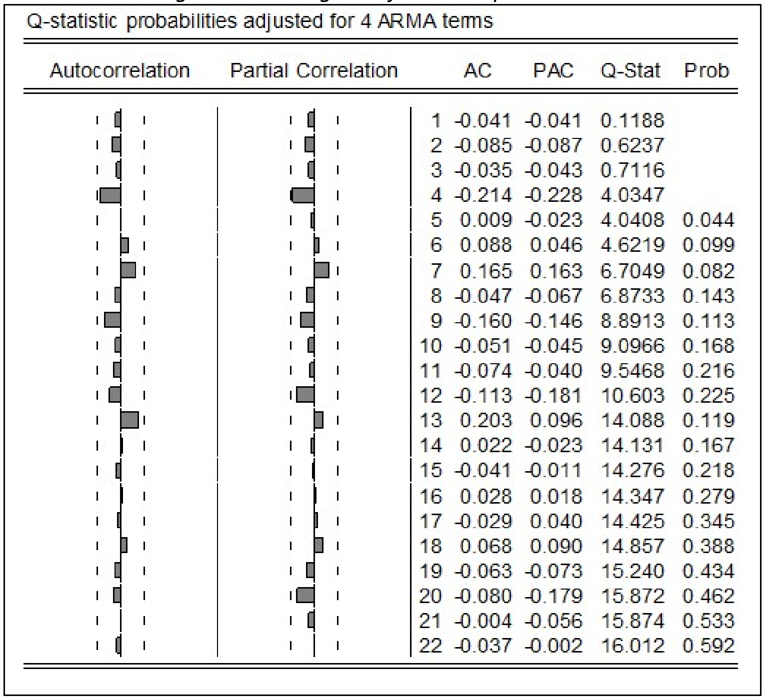
The adequacy of the optimal model selected for estimation is done on the following criteria.

Figure 4. Correlogram of Residuals Area model



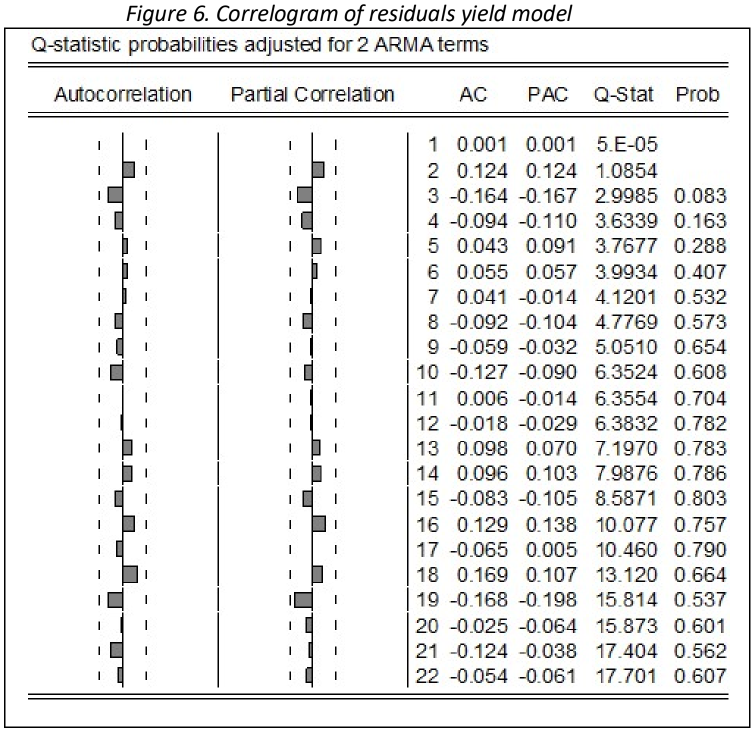
 ISSN (online): 2320-9836
ISSN (Print): 2320-9828

Figure 5. Correlogram of residuals production model



Significance of the estimated parameters: It is clear from the results (Table 3) that in case of Area all the selected AR and MA coefficients are significant at 1 percent level and with respect to production variable all the AR and MA coefficients are significant except AR (1) coefficient. All the coefficients of AR and MA are significant in case of yield variable. R^2 is also significant in all the three cases. Residual analysis: It is necessary to check whether the estimated residuals from the selected ARIMA models are random. The ACF and PACF graphs of the residuals estimated from the final selected ARIMA models for Area, Production and Yield of Total Pulses are presented in Figures 4-6. Figures also indicate Q statistics along with their p-values. The figures of the correlograms of the residuals for the all the selected ARIMA models show that residuals are white noise and p-values associated with the Q–statistics in all the cases indicate that the selected models are adequate and provide good fit to the real data.

Figure 6. Correlogram of residuals yield model



Apart from this other test of the estimated residuals like squared residual tests for testing heteroscedasticity revealed the absence of heteroscedasticity in all the models and tests for normality of the residuals based on Skewness, Kurtosis and Jarque-Bera tests indicate that residuals are fairly normally distributed further substantiating the model adequacy. Multi-co linearity tests show the absence of the problem. Based on the above diagnostic check the models estimated were used for forecasting the area, production and productivity of total pulses in India.

3.2.4. Forecasting

Figures 7-9 represent the actual and fitted values for the Area, Production and Yield of total pulses for the study period. The estimated ARIMA models have been successful in capturing the important movements in the actual values of area, production and yield of total pulses.

Figure 7. Actual and fitted values for area under total pulses

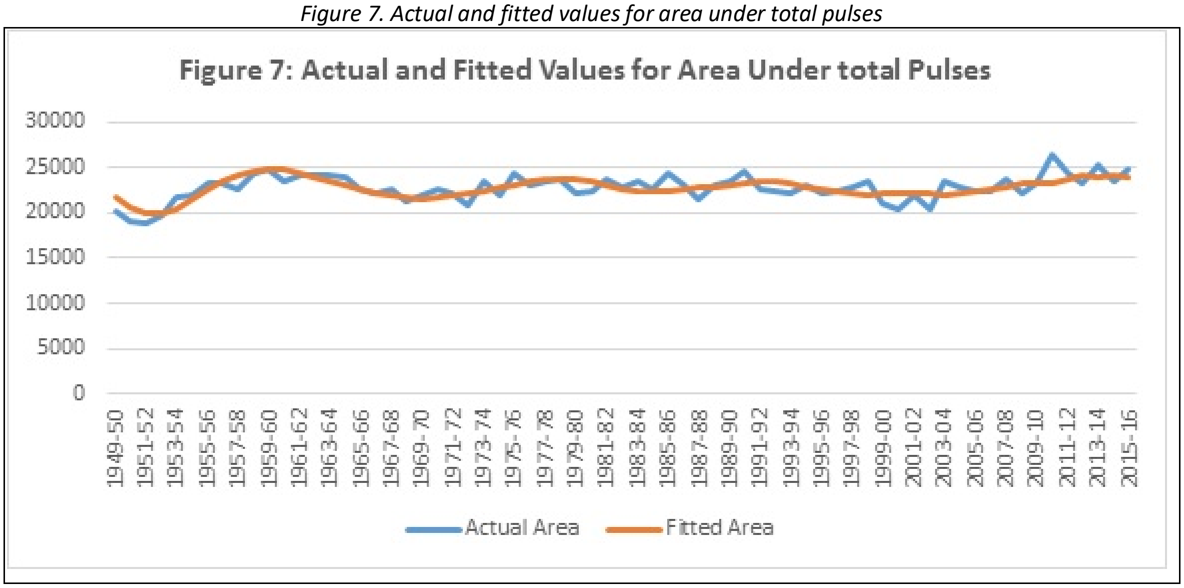
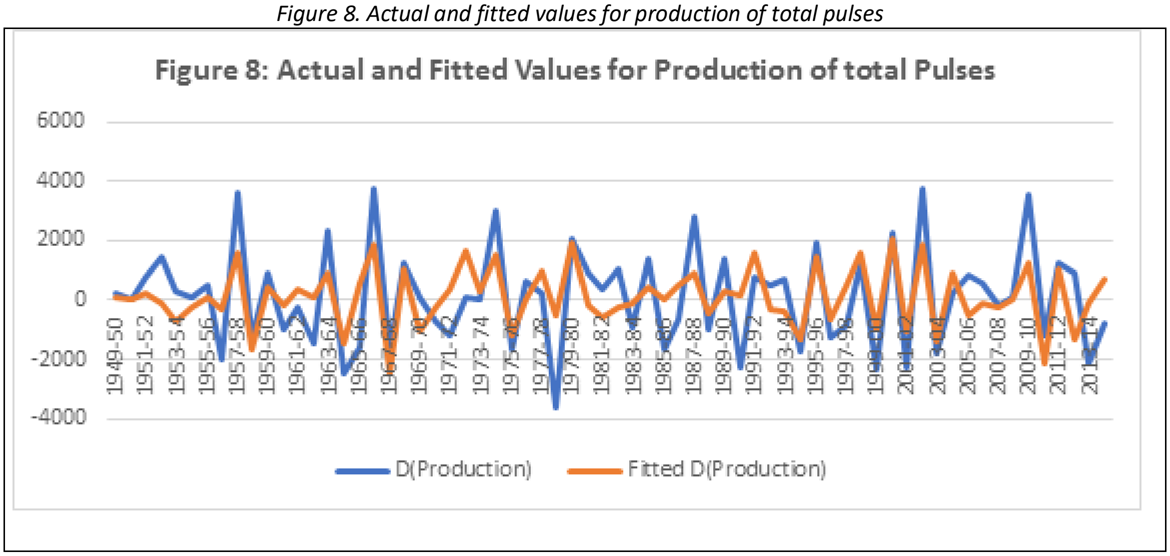


Figure 8. Actual and fitted values for production of total pulses



ARIMA models have been used to produce the average forecasts for the area, production and yield of total pulses in India. The accuracy of the forecasts was tested with the help of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Theil's inequality coefficient. The MAPE for the area, production and yield of total pulses are low indicating forecasting inaccuracy is low. The Theil's inequality coefficient is near to zero in case of all the three models indicating the forecasting accuracy of the models [12]. Table 4 contains the forecast evaluation statistics for the ARIMA models used for forecasting.

 ISSN (online): 2320-9836
ISSN (Print): 2320-9828

Figure 9. Actual and fitted values for yield of total pulses

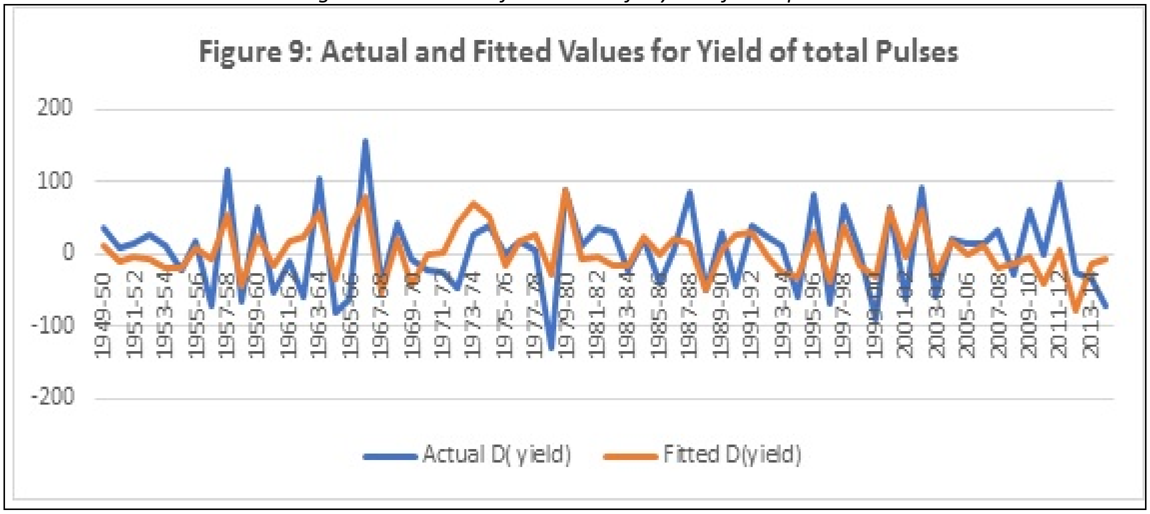


Table 4. Forecast evaluation statistics for the estimated ARIMA models

Variable/Model	RMSE	MAPE	Theil's Inequality Coefficient
Area under Total Pulses AR (1), A (2) , MA (1), MA (2)	1409.17	4.46	0.03
Production of Total Pulses AR (1), AR (10),MA (1), MA (12)	1750.01	12.06	0.06
Yield of Total Pulses AR (12), MA (1)	68.95	11.26	0.06

The forecasted values for area, production and area of total pulses are presented in Table 5.

Table 5. Forecasted values for area, production and yield of total pulses in India

Year	Area ('000' hectares)	Production ('000' tonnes)	Yield (Kgs/hectare)
2016-17	23101.58	16714.78	719.6499
2017-18	23312.96	16818.63	723.9047
2018-19	23463.33	16926.38	727.6623
2019-20	23539.74	17026.85	731.9039
2020-21	23537.04	17136.59	735.8016

The forecast values for the area under total pulses for the period 2016-17 to 2020-21 show a mixed trend early part an increasing trend and then it shows area under pulses may stagnate or decrease. For the production and yield of total pulses forecast values show increasing trend.

4. Conclusion

The present study made an attempt to develop short run forecasting models for Pulse area, production and yield in India. The empirical analysis indicated that the specifications for the (i) Area are AR (1), AR (2), MA (1) and MA (2) (ii) Production AR (1), AR (10), MA (1) and MA (12) (iii) Yield AR (12) and MA (1) were best fitted models for forecasting area, production and productivity of total pulses in India respectively. The annual time series data from 1949-50 to 2015-16 have been used for the analysis. All the essential steps of ARIMA model have been applied for forecasting five period ahead from 2015-16 onwards. The forecast analysis shows that area under pulse crops may increase marginally and stagnate in the future, while production and productivity of pulses show increasing trend in the near future.

The forecast analysis indicates stagnation in area under pulses so there is need for effective implementation of numerous programmes initiated by the government to enhance the area under the crop so that the increasing demand for pulses due to increasing population will be met through production enhancement.

ISSN (online): 2320-9836
ISSN (Print): 2320-9828

5. References

1. M.A. Badmus, O.S. Ariyo. Forecasting cultivated areas and production of maize in Nigerian using ARIMA model. *Asian Journal of Agricultural Sciences.* 2011; 3(3), 171-176.
2. M.A. Awal, M.A.B. Siddique. Rice production in Bangladesh employing by ARIMA model. *Bangladesh Journal of Agricultural Sciences.* 2011; 36(1), 51-62.
3. W. Xin, W. Can W. Empirical study on agricultural products price forecasting based on Internet-based Timely Price Information. *International Journal of Advanced Sciences and Technology.* 2016; 87, 31-36.
4. P.C. Padhan. Application of ARIMA model for forecasting agricultural productivity in India. *Journal of Agriculture and Social Sciences.* 2012*;* 8, 50-56.
5. B. Gurung, S. Panwar, K.N. Singh, R. Banerjee, S.R. Gurung, A.A. Rathore. Wheat yield forecasting using detrended yield over a sub-humid climatic environment in five districts of Uttar Pradesh, India. *Indian Journal of Agricultural Sciences.* 2017; 87(1), 87-91
6. K. Prabhakarn, C. Sivapragasam. Forecasting areas and production of rice. *International Journal of Farm Sciences.* 2014; 4(1), 99-106.
7. S.P. Sahoo, R. Singh. Trend and seasonality in prices and arrivals of Bengal Gram. *Indian Journal of Economics and Development.* 2017; 5(7), 1-5.
8. R. Arivarasi, M. Ganesan. Time series analysis of vegetable production and forecasting using ARIMA Model. *Asian Journal of Science and Technology.* 2015; 6(10), 1844-1848.
9. M. As'sd. Finding the best ARIMA Model to forecast daily peak electricity demand. *Applied Statistics Education and Research Collaboration*. 2012.
10. D.N. Gujarati, Sangeeta (2007). Basic econometrics. *Mc GRAW-Hill International Editions.* 2011; 1-920.
11. R.S. Pindyck, D.L. Rubinfeld. Econometric models and economic forecasts. *Mc GRAW-Hill International Editions.* 2000.

The Publication fee is defrayed by Indian Society for Education and Environment (www.iseeadyar.org)
Cite this article as:
Pushpa M. Savadatti. Trend and forecasting analysis of area, production and productivity of total pulses in India. *Indian Journal of Economics and Development.* Vol 5 (12), December 2017.