International Journal of Farm Sciences **6(2)** : 286-293, 2016

Forecasting production and area under cultivation for pulses in India using ARIMA model

GAGAN KUMAR

Department of Economics, SLK College (BRA Bihar University)
Sitamarhi 843302 Bihar, India
Email for correspondence: gaganks_slk@yahoo.in

The availability of pulses in India has been inadequate. Every year India has to import about 3.5 million tonnes of pulses for domestic consumption. Pulses are the main source of protein for the common man. Due to erratic production and prices shoot up the pulses go beyond the reach of the poor. Both area and production of pulses have to go up in order to meet the demand. In this article an attempt has been made to forecast the area under cultivation as well as the production level of pulses in near future with the help of ARIMA model using annual data from 1950-51 to 2013-14.

Keywords: Autoregressive; auto-correlation function; partial auto-correlation function; stationarity

INTRODUCTION

Agriculture is the mainstay of Indian economy. Still half of its manpower is dependent on it for livelihood. Though India is one of largest producers of the food grains and pulses, the availability of pulses in particular has been inadequate despite India being the largest producer (18.5 million tonnes) and processor of pulses in the world. It also imports around 3.5 million tonnes annually on an average to meet its ever increasing consumption needs of around 22.0 million tonnes (Patel 2015). Foodgrains output in 2014-15 crop year (July-June) fell by 3.2 per cent to 257.07 million tonnes due to 12 per cent deficit rains and unseasonal rains during February-March (Singh 2015). As a consequence the per capita availability of pulses which was already declining would further decline. Prices have gone up substantially due to such anticipation in recent period. Being the chief source of protein for average Indian a fall in its availability per capita is a cause of concern.

The objective of the study was to develop appropriate ARIMA model for the time series of pulse area and production in India and to make five year forecasts with appropriate prediction interval.

METHODOLOGY

The auto-regressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. These models are fitted to time series data either to better understand the data or to predict future points in the series. The existing study applies Box-Jenkins forecasting model popularly known as ARIMA. The ARIMA is an extrapolation method which requires historical time series data of underlying variable; ARIMA model is commonly used in macro level data analysis. The annual data on pulses cultivated area and production for the period from 1950-51 to 2013-14 were used for forecasting the future values using ARIMA models. The ARIMA methodology is also called as Box-Jenkins methodology (Box and Jenkins 1976). The Box-Jenkins procedure is concerned with fitting a mixed ARIMA model to a given set of data. Gretl software package was used for analysis of data. The main objective in fitting ARIMA model was to identify the stochastic process of the time series and predict the future values accurately. These methods have also been useful in many types of situations which involve the building of models for discrete time series and dynamic systems. However this method is not good for lead times or for seasonal series with a large random component (Granger 1969). Originally ARIMA models were studied extensively by George Box and Gwilym Jenkins during 1968 and their names have frequently been used synonymously with general ARIMA process applied to time series analysis, forecasting and control. However the optimal forecast of future values of a time series are determined by the stochastic model for that series. A stochastic process is either stationary or non-stationary. The first thing to note is that most time series are non-stationary and the ARIMA models refer only to a stationary time series. Since the ARIMA models refer only to a stationary time series, the first stage of Box-Jenkins model is for reducing non-stationary series to a stationary series by taking first order differences.

The main stages in setting up a Box-Jenkins forecasting model are as follows:

1. Identification
2. Estimating the parameters
3. Diagnostic checking
4. Forecasting

RESULTS and DISCUSSION

In the present study the data for cultivated area and production of pulses for the period 1950-51 to 2013-14 were used following the above mentioned stages of ARIMA model. The data have been obtained from the RBI database on Indian economy.

Model identification

For forecasting pulse area and production ARIMA model was estimated only after transforming the variable under forecasting into a stationary series. The stationary series is the one whose values

vary over time only around a constant mean and a constant variance. There are several ways to ascertain this. The most common method is to check stationarity through examining the graph or time plot of the data. Non-stationarity in mean is corrected through appropriate differencing of the data. In this case difference of order 1 was found to be sufficient enough in order to achieve stationarity in mean.

The newly constructed variable Xt can now be examined for stationarity. The graph of Xt was stationary in mean. The next step was to identify the values of p and q. For this the auto-correlation and partial auto-correlation coefficients of various orders of Xt were computed (Table I). The auto-correlation function (ACF) and (PACF) shows that the order of p and q must be 1. Three tentative ARIMA models were tested and the models which had minimum AIC (Akaike Information Criterion), SBC (Schwartz Bayesian Criterion) and Hannan Quinn Criterion were chosen. The models and

Variable	ARIMA (p, d, q)	AIC	SBC	Hannan-Quinn
Pulse area	1, 1, 0	206.56	212.99	209.09
	1, 1, 1	204.77	213.34	208.14
	1, 1, 2	206.65	217.37	210.87
Pulses production	1, 1, 0	228.52	234.94	231..04
	1, 1, 1	227.57	236.15	230.95
	1, 1, 2	229.52	240.24	233.74

their corresponding AIC, SBC and Hannan Quinn Criterion are as under:

Thus the most suitable model was found to be ARIMA (1, 1, 1) for pulse area and ARIMA (1, 1, 1) for pulse production which had lowest AIC, SBC and Hannan Quinn Criterion values.

Model estimation and verification

Pulse production area and production model parameters were estimated using Gretl software. Results of estimation of ACF and PACF are reported in Table 1 (Fig 1) and estimates of fitted ARIMA model in Table 2. The model verification is concerned with checking the residual of the model to see if they contain any systematic pattern which still can be removed to improve on the chosen ARIMA. This was done through examining the auto-correlations and partial correlations of the residuals of various orders. The ACF and PACF of the residual also indicate good fit of the model.

Residual autocorrelation and partial auto-correlation functions- production and

Table 1. Auto-correlations and partial auto-correlations of pulse production and area

LAG	Production				Area			
	ACF	PACF	Q-stat	p-value	ACF	PACF	Q-stat	p-value
1	0.6261***	0.6261***	26.2849	0.000	0.4818***	0.4818***	15.5625	0.000
2	0.5646***	0.2839**	48.0054	0.000	0.3081**	0.0989	22.0293	0.000
3	0.4483***	0.0311	61.9206	0.000	0.2281*	0.0614	25.6317	0.000
4	0.3149**	-0.0967	68.9028	0.000	0.0731	-0.1008	26.0080	0.000
5	0.3497***	0.1747	77.6585	0.000	-0.0478	-0.1085	26.1719	0.000
6	0.3016**	0.0575	84.2847	0.000	-0.0677	-0.0149	26.5055	0.000
7	0.2858**	0.0127	90.3373	0.000	-0.1468	-0.0968	28.1025	0.000
8	0.2182*	-0.0897	93.9287	0.000	-0.2148*	-0.1082	31.5812	0.000
9	0.2013	0.0480	97.0406	0.000	-0.2842**	-0.1564	37.7835	0.000
10	0.1485	-0.0279	98.7654	0.000	-0.3466***	-0.1662	47.1807	0.000
11	0.1632	0.0652	100.8887	0.000	-0.3649***	-0.1350	57.7934	0.000
12	0.1525	-0.0021	102.7773	0.000	-0.1935	0.1095	60.8335	0.000
13	0.2438*	0.2100*	107.7001	0.000	-0.1149	0.0260	61.9274	0.000
14	0.2485**	0.0297	112.9169	0.000	-0.0804	-0.0405	62.4729	0.000
15	0.2497**	0.0115	118.2897	0.000	0.0768	0.0974	62.9813	0.000
16	0.1945	-0.1208	121.6181	0.000	0.0369	-0.1161	63.1011	0.000

area are given in Table 3 and forecast evaluation statistics in Table 4.

Forecasting with ARIMA model

An ARIMA model is used to produce the best weighted average forecasts for a single time series (Rahulamin and Razzaque 2000). The accuracy of forecasts for both ex-ante and ex-post were tested using the tests such as mean square error (MSE) and mean absolute percentage error (MAPE) (Markidakis and Hibbon 1979). ARIMA models are developed basically to forecast the corresponding variable. To judge the forecasting ability of the fitted ARIMA model important measures of the sample period forecasts accuracy was computed. The MAPE for pulse cultivated area turned out to be 4.1 and pulse production turned to be 9.1456. Theil's U statistic is a relative accuracy measure that compares the forecasted results with the results of forecasting with minimal historical data. It also squares the deviations to give more weight to large errors and to exaggerate errors which can help eliminate methods with large errors. Theil's U statistic less than 1 indicates that forecasting technique is better than guessing. This measure indicates that forecasting inaccuracy is low. The forecasts for pulse area and production during 2014 and 2018 showing increasing trend are given in Table 5. Dual residual ACF and PACF for

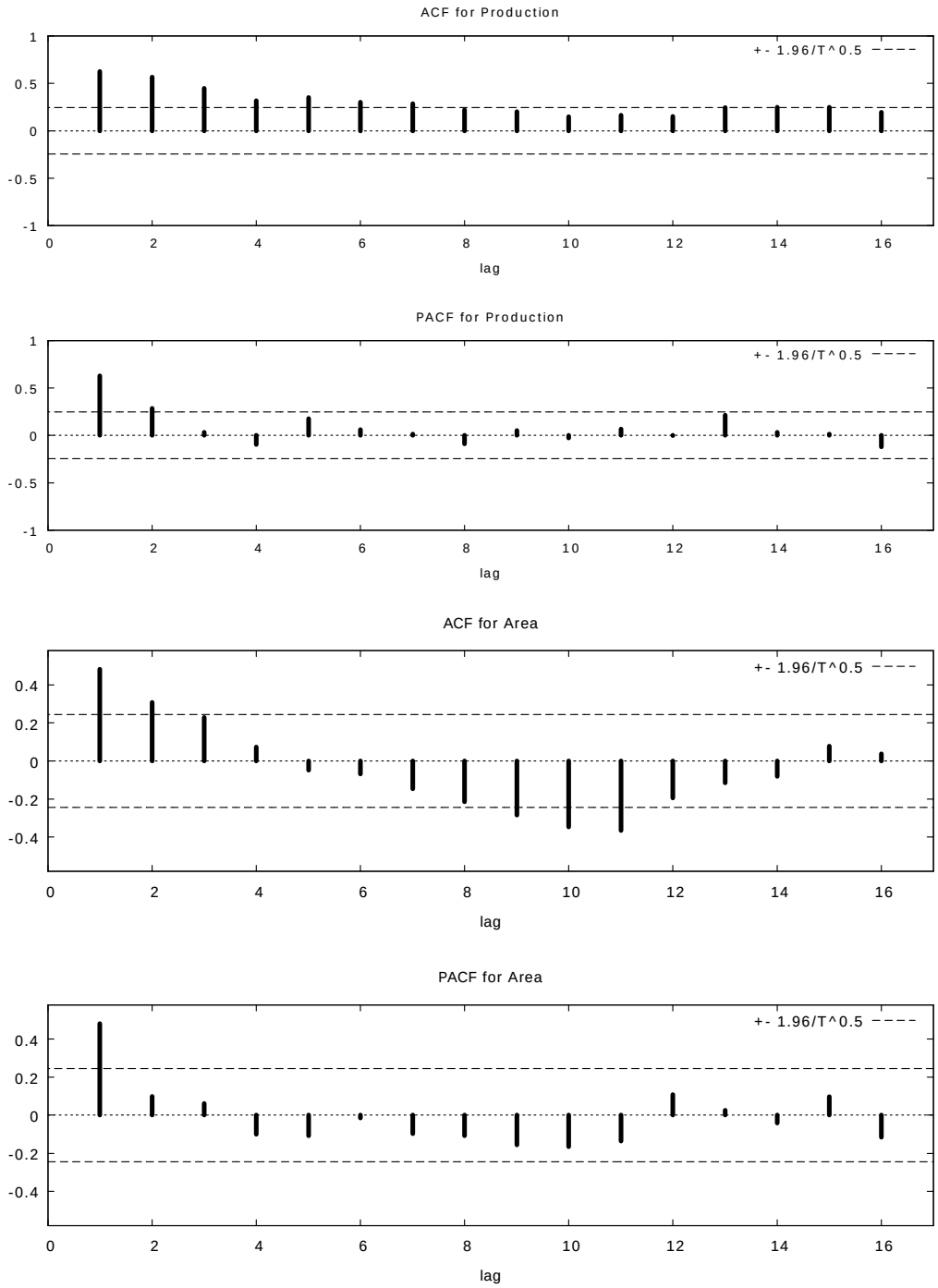

Fig 1. ACF and PACF for area and production of pulses

Table 2. Estimates of fitted ARIMA model for pulse area and production

Model 1: ARIMA, using observations 1951-2013 (T= 63).
Estimated using Kalman filter (exact ML)
Dependent variable: (1-L) Area under pulse production
Standard errors based on Hessian

	Coefficient	SE	z	p-value
Const	0.0856268	0.0774707	1.105	0.2690
phi_1	“0.0251076	0.228461	“0.1099	0.9125
theta_1	“0.461739	0.194617	“2.373	0.0177**

Mean dependent var iable	0.097460
SD dependent var iable	1.292338
Mean of innovations	0.007535
SD of innovations	1.151007
Log-likelihood	“98.38500
Akaike criterion	204.7700
Schwarz criterion	213.3425
Hannan-Quinn	208.1416

	Real	Imaginary	Modulus	Frequency
AR Root 1	-39.8286	0.0000	39.8286	0.5000
MA Root 1	2.1657	0.0000	2.1657	0.0000

Model 1: ARIMA, using observations 1951-2013 (T= 63)

Dependent variable: (1-L) Production of pulses
Standard errors based on Hessian

	Coefficient	SE	z	p-value
Const	0.158381	0.0813619	1.9466	0.05158*
phi_1	“0.244885	0.223935	“1.0936	0.27415
theta_1	“0.429818	0.227631	“1.8882	0.05900*

Mean dependent var iable	0.172381
SD dependent var iable	1.696565
Mean of innovations	0.006745
SD of innovations	1.377198
Log-likelihood	“109.7895
Akaike criterion	227.5789
Schwarz criterion	236.1514
Hannan-Quinn	230.9505

	Real	Imaginary	Modulus	Frequency
AR Root 1	-4.0836	0.0000	4.0836	0.5000
MA Root 1	2.3266	0.0000	2.3266	0.0000

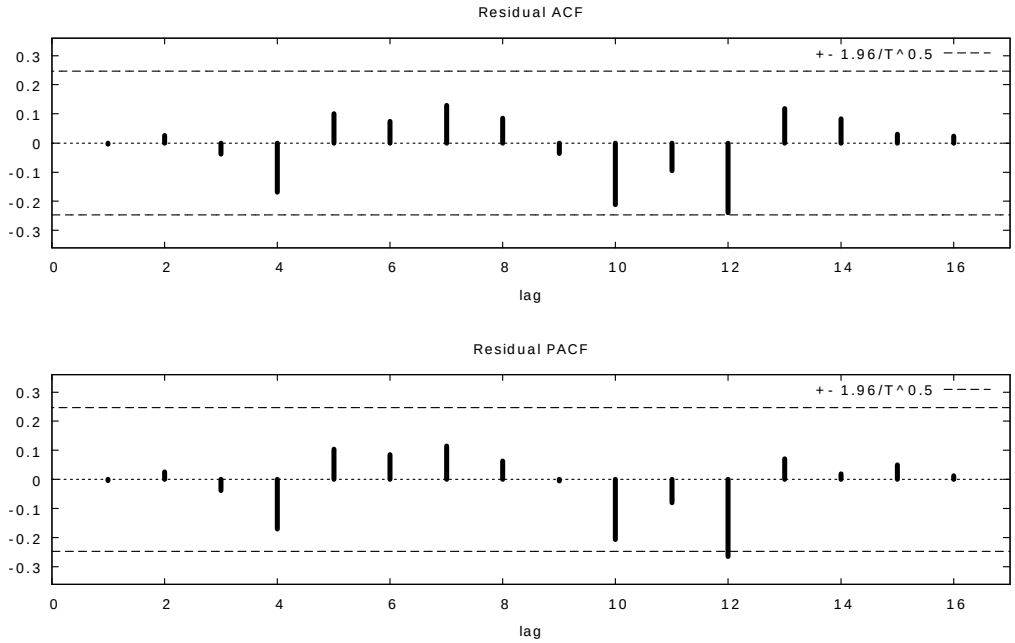

Fig 2. Dual residual ACF and PACF for differenced data on production

Table 3. Residual autocorrelation and partial auto-correlation functions- production and area

LAG	Production/area			
	ACF	PACF	Q-stat	p-value
1.	-0.0071	-0.0071	0.0033	0.954
2.	-0.0585	-0.0586	0.2332	0.890
3.	0.1196	0.1192	1.2093	0.751
4.	0.0697	0.0684	1.5461	0.818
5.	-0.0028	0.0119	1.5466	0.908
6.	0.0455	0.0399	1.6952	0.945
7.	0.0106	-0.0047	1.7034	0.974
8.	0.0648	0.0653	2.0159	0.981
9.	-0.0546	-0.0656	2.2423	0.987
10.	-0.2204*	-0.2276*	5.9955	0.816
11.	-0.2199*	-0.2707**	9.8046	0.548
12.	0.0139	-0.0373	9.8200	0.632
13.	0.0401	0.0833	9.9518	0.698
14.	-0.1174	-0.0239	11.1044	0.678
15.	0.0797	0.1468	11.6464	0.706
16.	-0.0346	-0.0136	11.7506	0.761

Table 4. Forecast evaluation statistics

Item	Mean error	MSE	RMSE	MAE	MPE	MAPE	Theil's U
Production	0.0067447	1.8975	1.3775	1.0743	-1.1941	9.1456	0.78828
Area	0.0075347	1.3264	1.1517	0.93355	-0.12196	4.1	0.89716

MSE= Mean square error, RMSE= Root mean square error, MAE= Mean absolute error, MPE= Maximum percentage error, MAPE= Maximum absolute percentage error

Table 5. Forecasted values of pulse cultivated area and production with 95% confidence level

Year	Prediction (Million ha)	SE	95% interval	Prediction (MT)	SE	95% interval
2014	24.7964	1.15101	22.5404, 27.0523	18.5733	1.37720	15.8740, 21.2725
2015	24.8950	1.29371	22.3594, 27.4306	18.9410	1.44823	16.1026, 21.7795
2016	24.9803	1.42806	22.1814, 27.7793	19.0481	1.59804	15.9160, 22.1803
2017	25.0660	1.55068	22.0267, 28.1052	19.2191	1.71404	15.8596, 22.5785
2018	25.1516	1.66429	21.8896, 28.4135	19.3744	1.82735	15.7929, 22.9559

differenced data on production are given in Fig 2.

CONCLUSION

In this study ARIMA (1, 1, 1) for pulse area and ARIMA (1, 1, 1) for pulse production were developed for prediction. From the forecasts available by using the developed model it can be seen that forecasted pulse cultivated areas and production were to increase in the coming years. The validity of the forecasted value can be checked when the data on the lead periods become available.

REFERENCES

Box GEP and Jenkins GM 1976. Time series of analysis. Forecasting and Control, Sam Franscico, Holden Day, California, USA.

Chatterjee D 1948. A modified key and enumeration of species of *Oryza sativa* L. Indian Journal of Agricultural Sciences **18:** 185-192.

Granger CWJ 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica **30:** 424-438.

Makridakis S and Hibbon M 1979. Accuracy of forecasting: an empirical investigation. Journal of the Royal Statistical Society: Series A **142(2):** 97-145.

Patel A 2015. Pulses productivity and production in India 2015. Agribusiness, January 12, 2015.

Rahulamin MD and Razzaque MA 2000. Autoregressive integrated moving average modeling for monthly potato prices in Bangladesh. Journal of Financial Management and Analysis **13(1):** 74-80.

Received: 11.10.2015 ***Accepted: 27.1.2016***