International Journal of Chemical Studies

P-ISSN: 2349–8528
E-ISSN: 2321–4902
IJCS 2019; 7(6): 1819-1825

Received: 01-09-2019
Accepted: 05-10-2019

K Arulpandiyan
PG Scholar in Agricultural Statistics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

K Prabakaran
Assistant Professor (Agricultural Statistics), Department of Agricultural Economics, Agricultural College & Research Institute, Madurai, Tamil Nadu, India

Corresponding Author:
K Arulpandiyan
PG Scholar in Agricultural Statistics, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India

Forecasting area and production of major rainfed pulse crops in Madurai district

K Arulpandiyan and K Prabakaran

Abstract
The study was carried out to forecast the major rainfed pulse crops like green gram and red gram cultivated area and production in Madurai district. The data collected for the period of 2003-04 to 2017-18 in Madurai district selected pulse crop were used. Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) model was fitted to analyze the data. Validity of the model was tested using standard statistical techniques. ARIMA (0, 1, 1) and ARIMA (0, 1, 1) model were used to forecast green gram cultivated area and production and ARIMA (2, 1, 0) and ARIMA (2, 1, 0) model were used to forecast red gram cultivated area and production in Madurai district for next five years. The result also shows green gram cultivated area forecast for the year 2022 to be about 4357.15 thousand hectare with upper and lower limit 7018.80 and 1659.51 thousand hectares and red gram cultivated area forecast for the year 2022 to be about 1746.03 thousand hectare with upper and lower limit 2808.87 and 683.192 thousand hectares respectively. The model also shows green gram production forecast for the year 2022 to be about 4578.70 thousand tonnes with upper and lower limit 7560.39 and 1597.01 thousand tonnes and red gram production forecast for the year 2022 to be about 1385.10 thousand tonnes with upper and lower limit 2958.76 and 188.548 thousand tonnes respectively.

Keywords: Forecasting, area, production, green gram, red gram, ARIMA

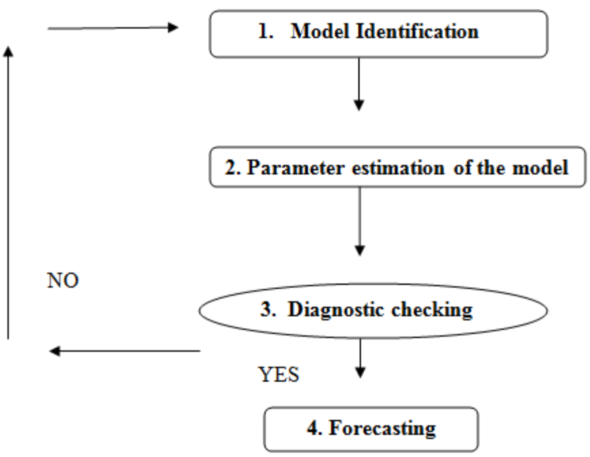
Introduction
In India 60 percent of the total cultivated area comes under rainfed condition. Rainfed crops account for 48 percent area under food crops and 68 percent under non-food crops in India. Rainfed agriculture supports an estimated 40 percent of population (484 million) and has a large share of cropped area under rice (42 percent), pulses (77 percent), oilseeds (66 percent) and coarse cereals (85 percent). In Tamil Nadu, The agriculture sector continues to be the backbone of the State economy, providing 44 percent of the total employment in Tamil Nadu. Of the total gross cropped area of 57.53 lakh hectares, the gross area irrigated was 32.38 lakh hectares (56 percent) and the rest (44 percent) was under rainfed cultivation.
Madurai is the second most important district in the state of Tamil Nadu, where significant progress has been made in the development of agriculture. Madurai district is basically agrarian and agriculture is the main occupation and also, majorly cultivated on rainfed pulse crops like Green gram and Red gram in Madurai district.
Forecasting has an important role in the management of agriculture. Univariate time series modeling have been useful in developing the forecasting for production of the crops. During the last few decades many sophisticated statistical forecasting models have been developed due to the availability of advanced computers. One of such models includes the Autoregressive Integrated Moving Average (ARIMA) models (Box-Jenkins, 1976) [1]. These ARIMA models have been widely used in attempts to forecast the area and production of major rainfed crops in Madurai district. Autocorrelation function (ACF), partial autocorrelation function (PACF) and spectral density function indicate non-stationary of the series.
In this study attempts have been made to examine the class of ARIMA models that may best fit the area and production of major rainfed pulse crops in Madurai district and the forecast skill of the best fitted model have been also studied.

Materials and Methods
The annual data on major rainfed pulse crops like red gram and green gram cultivated area and production for the period from 2003-04 to 2017-18 and data collected from Regional statistical office of economics and statistics, Madurai district were used for forecasting the future values

using ARIMA models. Originally ARIMA models have been 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. The ARIMA methodology is also called as Box-Jenkins methodology. The Box-Jenkins procedure is concerned with fitting a mixed Auto Regressive Integrated Moving Average (ARIMA) model to a given set of data. The main objective in fitting this ARIMA model is 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 situation which involve the building of models for discrete time series and dynamic systems. But, this method was not good for lead times or for seasonal series with a large random component (Granger and Newbold, 1970) [2].

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 model 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 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.



Step 1: Identification of the model

Identification of the model for ARIMA (p,d,q) is based on the concepts of time-domain and frequency-domain analysis i.e. autocorrelation function (ACF), partial autocorrelation function (PACF) and spectral density function. Once the order of differencing has been diagnosed and the differenced univariate time series can be analysed by the method of both time-domain and frequencydomain approach (Cressie, 1988) [3].

Step 2: Estimation of parameters

The appropriate p, d and q values of the model and their statistical significance can be judged by t-distribution. A model with minimum values of RMSE, MAPE, AIC, BIC, Q-statistics and with high R-square, may be considered as an appropriate model for forecasting. The model selection criteria includes Akaike Information criterion (AIC) and Schwarz's Bayesian Information criterion (SBC), Mean squared error (MSE), Root Mean squared error (RMSE), Mean absolute error (MAE) and Minimum Absolute Percentage Error (MAPE).

Step 3: Diagnostic checking

Considerable skill is required to choose the actual ARIMA (p,d,q) model so that the residuals estimated from this model are white noise. So the autocorrelations of the residuals are to be estimated for the diagnostic checking of the model. These may also be judged by Ljung-Box statistic under null hypothesis that autocorrelation co-efficient is equal to zero.

Step 4: Forecasting

ARIMA models are developed basically to forecast the corresponding variable. The entire data is segregated in two parts, one for sample period forecasts and the other for post-sample period forecasts. The former are used to develop confidence in the model and the latter to generate genuine forecasts for use in planning and other purposes.

Result and Discussions

In this study we used the data for major rainfed pulse crops like red gram and green gram cultivated area and production during the period 2003-04 to 2017-18 in Madurai district. As we have earlier stated that development of ARIMA model for any variable involves four steps: Identification, Estimation, Verification and Forecasting. Each of these four steps is now explained for rainfed pulse crops like red gram and green gram cultivated area and production as follows.

Model Identification

For forecasting rainfed pulse crops like red gram and green gram cultivated area and production, ARIMA model 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 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 is non-stationary. Non stationarity in mean is connected through appropriate differencing of the data. In this case difference of order 1 was sufficient to achieve stationarity in mean.

Crop		ARIMA	AIC	SBC
Green gram	Area	(0, 1, 1)*	225.4371	227.3542
		(2, 1, 0)	226.5077	229.0640
	Production	(0, 1, 1)*	232.8510	234.7681
		(2, 1, 0)	232.9540	235.5102
Red gram	Area	(0, 1, 1)	211.2287	213.1459
		(2, 1, 0)*	209.7466	212.3028
	Production	(0, 1, 1)	228.4856	231.5849
		(2, 1, 0)*	223.6598	226.2160

*indicated that selected model based on lowest AIC and SBC values

The newly constructed variable Xt can now be examined for stationarity. The graph of Xt was stationary in mean. The next step is to identify the values of p and q. For this, the autocorrelation and partial auto correlation coefficients of various orders of Xt are computed (Table 1). The Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) (Fig 1 and Fig 2) show that the order of p and q can at most be 1. We entertained two tentative ARIMA models and choose that model which has minimum AIC (Akaike Information Criterion) and SBC (Schwartz Bayesian Criterion). The models and corresponding AIC and SBC values are given in the above table.

The most suitable model is ARIMA (0, 1, 1) and ARIMA (0, 1, 1) for green gram area and production and ARIMA (2, 1, 0) and ARIMA (2, 1, 0) for red gram area and production has the lowest AIC and SBC values.

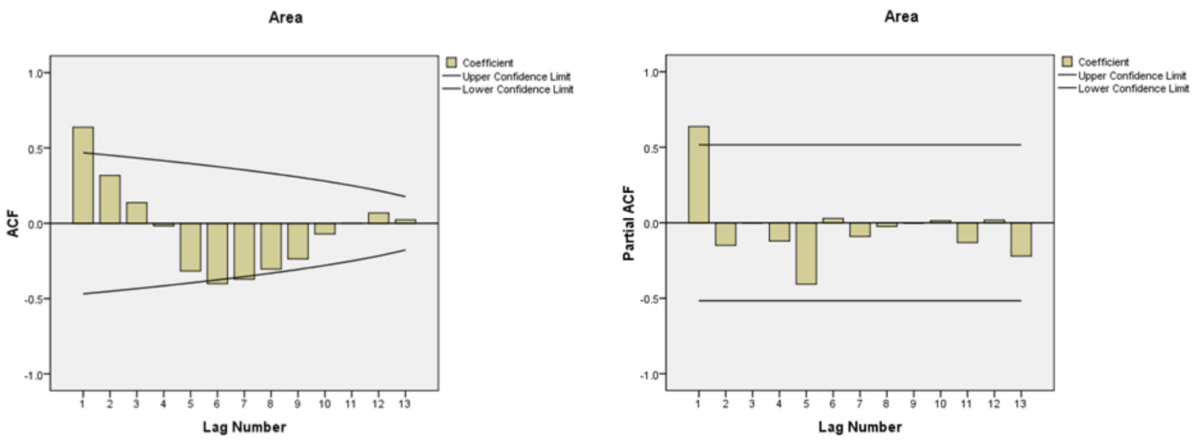
Table 1: Auto Correlations and Partial auto correlations of red gram and green gram cultivated area and production

Greengram area					Redgram area				
Lag	**ACF**	**Std. Error (a)**	**PACF**	**Std. Error**	**Lag**	**ACF**	**Std. Error (a)**	**PACF**	**Std. Error**
1	.638	.234	.638	.258	1	.601	.234	.601	.258
2	.318	.226	-.150	.258	2	.409	.226	.075	.258
3	.137	.217	000	.258	3	.366	.217	.147	.258
4	-.017	.208	-.121	.258	4	.058	.208	-.365	.258
5	-.317	.198	-.406	.258	5	-.125	.198	-.146	.258
6	-.401	.188	.029	.258	6	-.102	.188	.074	.258
7	-.371	.177	-.091	.258	7	-.092	.177	.169	.258
8	-.304	.166	-.025	.258	8	-.275	.166	.336	.258
9	-.237	.153	.000	.258	9	-.337	.153	-.291	.258
10	-.071	.140	.013	.258	10	-.237	.140	087	.258
11	.002	125	.-131	.258	11	-.349	.125	.008	.258
12	.069	108	.017	.258	12	-.240	.108	.216	.258
13	.023	.089	-.221	.258	13	-.098	.089	-.211	.258

Green gram production				
Lags	**ACF**	**SE**	**PACF**	**SE**
1	-.412	.241	-.412	.267
2	-.253	.231	-.509	.267
3	.276	.222	-.147	.267
4	-.216	.211	-.393	.267
5	.090	.200	-.217	.267
6	.078	.189	-.229	.267
7	-.004	.177	.034	.267
8	-.071	.164	-.050	.267
9	.000	.149	.042	.267
10	.015	.134	-.025	.267
11	-.002	.116	039	.267
12	.003	.094	-.057	.267

Redgram production				
Lags	**ACF**	**SE**	**PACF**	**SE**
1	.710	.234	.710	.258
2	.534	.226	.060	.258
3	.221	.217	-.361	.258
4	.131	.208	.174	.258
5	-.035	.198	-.132	.258
6	-.046	.188	.007	.258
7	-.199	.177	-.220	.258
8	-.282	.166	-.190	.258
9	-.363	.153	.055	.258
10	-.381	.140	-.168	.258
11	-.356	.125	.006	.258
12	-.251	.108	.086	.258
13	-.121	.089	.054	.258

Green gram ACF and PACF differenced data



Red gram area ACF and PACF differenced data

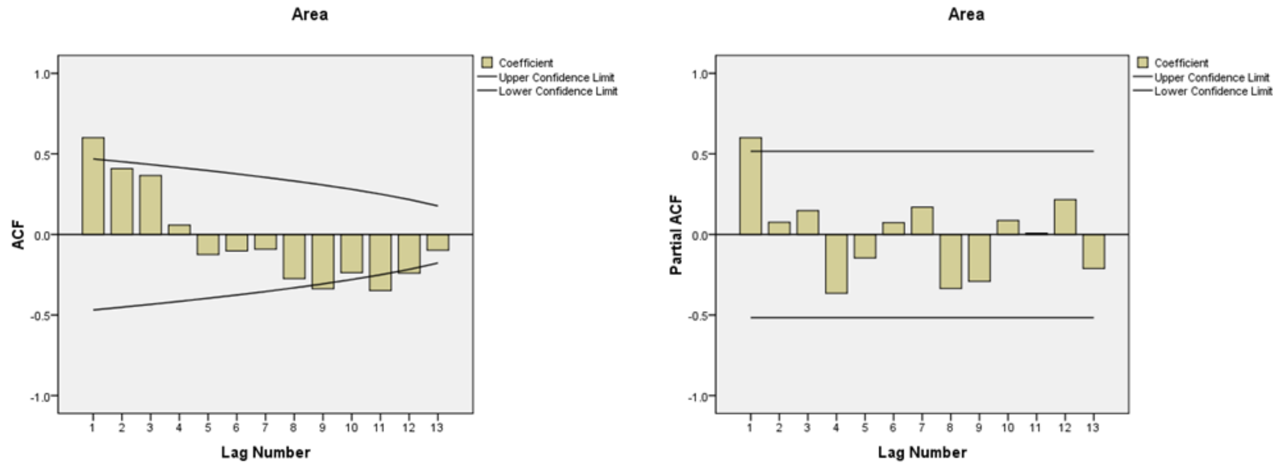
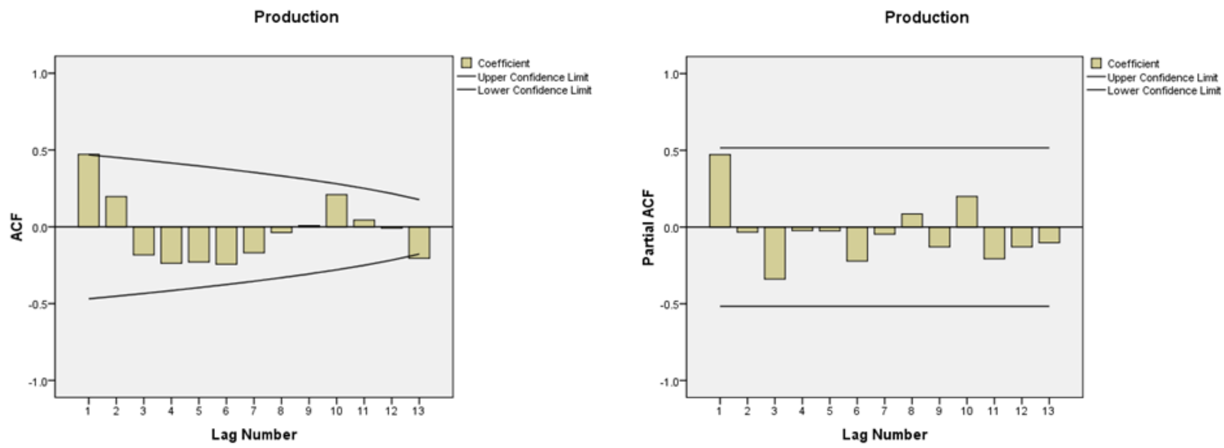


Fig 1: Green gram and red gram cultivated area ACF and PACF differenced data

Green gram crop production ACF and PACF differenced data



Red gram Crop production ACF and PACF differenced data

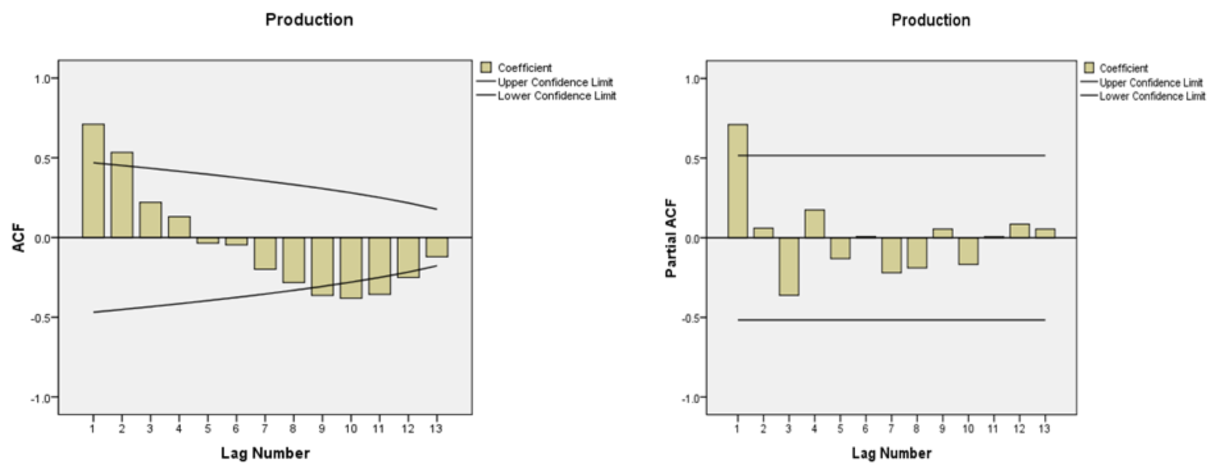


Fig 2: Green gram and red gram production ACF and PACF differenced data

Model estimation and verification

Rainfed pulse crops like green gram and red gram crops cultivated areas and production model parameters were estimated using SPSS package. Results of estimation are reported in Table 2 and Table 3. The model verification is concerned with checking the residual of the model to see if they contain any systematic pattern which still can he removed to improve on the chosen ARIMA. This is done through examining the auto correlations and partial auto correlations of the residuals of various orders. For this purpose, the various correlations up to 13 lags were computed and the same along with their significance which is tested by

Box-Ljung test provided in Table 4 and Table 5. As the results indicated majority of correlations are not significantly different from zero at a reasonable level for rainfed pulse crops like red gram and green gram area and none of these correlations significantly different from zero at a reasonable level for rainfed pulse crops like red gram and green gram production also. This proves that the selected ARIMA model is an appropriate model. The ACF and PACF of the residual (Fig 3) also indicate "good fit" of the model.

Table 2: Estimates of the fitted ARIMA model for green gram and red gram area

Fit statistics	Green gram	Red gram
St. R-sq	.442	.684
R-sq	.442	.836
RMSE	625.977	457.993
MAPE	15.802	17.605
MaxAPE	61.682	49.701
MAE	453.177	303.955
MaxAE	10.22	720.207
Normalized BIC	13.420	12.976

Table 3: Estimates of the fitted ARIMA model for green gram and red gram production

Fit statistics	Green gram	Red gram
St. R-sq	.362	.590
R-sq	.362	.590
RMSE	642.991	662.743
MAPE	72.144	32.949
MaxAPE	72.144	70.134
MAE	518.256	460.134
MaxAE	11.95	920.610
Normalized BIC	13.293	13.895

St. R-Sq	Stationary R-Square
R-Sq	R-Square
RMSE	Root Mean Square
MAPE	Mean Absolute Percentage Error
MaxAPE	Maximum Absolute Percentage Error
MAE	Mean Absolute Percentage Error
MaxAE	Maximum Absolute Error

Table 4: ACF and PACF of residuals for green gram and red gram cultivated area

Green gram								
Lag	ACF	Std. Error(a)	Box-Ljung Statisics			Lag	PACF	Std. Error
			Value	df	Sig.(b)			
1	.638	.234	7.409	1	.006	1	.638	.258
2	.318	.226	9.390	2	.009	2	-.150	.258
3	.137	.217	9.791	3	.020	3	000	.258
4	-.017	.208	9.797	4	.044	4	-.121	.258
5	-.317	.198	12.355	5	.030	5	-.406	.258
6	-.401	.188	16.910	6	.010	6	.029	.258
7	-.371	.177	21.305	7	.003	7	-.091	.258
8	-.304	.166	24.668	8	.002	8	-.025	.258
9	-.237	.153	27.048	9	.001	9	.000	.258
10	-.071	.140	27.307	10	.002	10	.013	.258
11	.002	125	27.307	11	.004	11	.-131	.258
12	.069	108	27.709	12	.006	12	.017	.258
13	.023	.089	27.779	13	.010	13	-.221	.258

Red gram								
Lag	ACF	Std. Error(a)	Box-Ljung Statisics			Lag	PACF	Std. Error
			Value	df	Sig.(b)			
1	.601	.234	6.570	1	.010	1	.601	.258
2	.409	.226	6.849	2	.007	2	.075	.258
3	.366	.217	12.697	3	.005	3	.147	.258
4	.058	.208	12.775	4	.012	4	-.365	.258
5	-.125	.198	13.173	5	.022	5	-.146	.258
6	-.102	.188	13.467	6	.036	6	.074	.258
7	-.092	.177	13.736	7	.056	7	.169	.258
8	-.275	.166	16.481	8	.036	8	.336	.258
9	-.337	.153	21.315	9	.011	9	-.291	.258
10	-.237	.140	24.179	10	.007	10	087	.258
11	-.349	.125	31.930	11	.001	11	.008	.258

12	-.240	.108	36.830	12	.000	12	.216	.258
13	-.098	.089	38.057	13	.000	13	-.211	.258

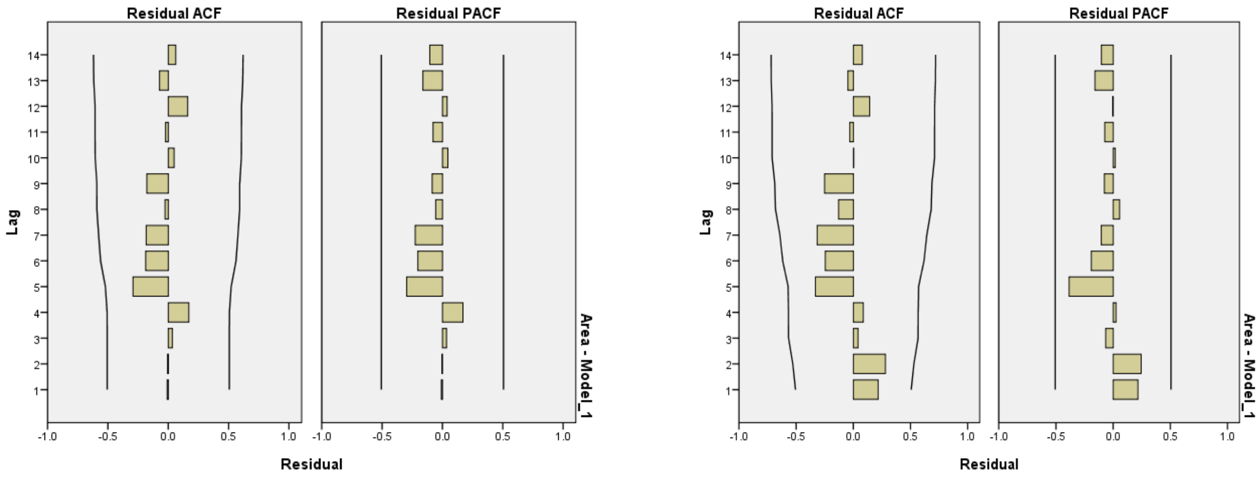
Table 5: ACF and PACF of residuals for green gram and red gram production

Green gram								
Lag	**ACF**	**Std. Error(a)**	**Box-Ljung Statisics**			**Lag**	**PACF**	**Std. Error**
			Value	**df**	**Sig.(b)**			
1	-.412	.241	2.925	1	.087	1	-.412	.267
2	-.253	.231	4.117	2	.128	2	-.509	.267
3	.276	.222	5.670	3	.129	3	-.147	.267
4	-.216	.211	6.720	4	.151	4	-.393	.267
5	.090	.200	6.922	5	.227	5	-.217	.267
6	.078	.189	7.093	6	.312	6	-.229	.267
7	-.004	.177	7.093	7	.419	7	.034	.267
8	-.071	.164	7.282	8	.507	8	-.050	.267
9	.000	.149	7.282	9	.608	9	.042	.267
10	.015	.134	7.294	10	.697	10	-.025	.267
11	-.002	.116	7.294	11	.775	11	039	.267
12	.003	.094	7.296	12	.837	12	-.057	.267

Red gram								
Lag	**ACF**	**Std. Error(a)**	**Box-Ljung Statisics**			**Lag**	**PACF**	**Std. Error**
			Value	**df**	**Sig.(b)**			
1	.710	.234	9.182	1	.002	1	.710	.258
2	.534	.226	14.776	2	.001	2	.060	.258
3	.221	.217	15.809	3	.001	3	-.361	.258
4	.131	.208	16.208	4	.003	4	.174	.258
5	-.035	.198	16.238	5	.006	5	-.132	.258
6	-.046	.188	16.299	6	.012	6	.007	.258
7	-.199	.177	17.564	7	.014	7	-.220	.258
8	-.282	.166	20.470	8	.009	8	-.190	.258
9	-.363	.153	20.072	9	.002	9	.055	.258
10	-.381	.140	33.469	10	.000	10	-.168	.258
11	-.356	.125	41.565	11	.000	11	.006	.258
12	-.251	.108	46.926	12	.000	12	.086	.258
13	-.121	.089	48.784	13	.000	13	.054	.258

ACF and PACF of residuals of fitted ARIMA model for green gram crop cultivated area and production

Area **Production**



ACF and PACF of residuals of fitted ARIMA model for red gram crop cultivated area and production

Area **Production**

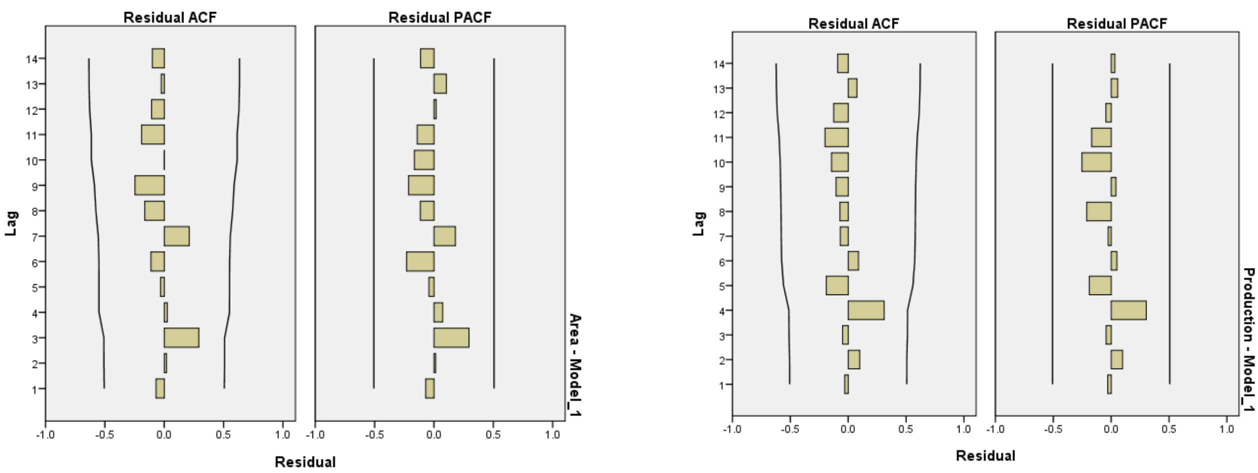


Fig 3: ACF and PACF of residuals of fitted ARIMA model for green gram and red gram cultivated area and production

Forecasting with ARIMA model

ARIMA models are developed basically to forecast the corresponding variable. To judges the forecasting ability of the fitted ARIMA model important measure of the sample period forecasts accuracy was computed. The Mean Absolute Percentage Error (MAPE) for green gram cultivated area turns out to be 15.802 and production turns out to be 72.144 and red gram cultivated area turns out to be 17.605 and production turns out to be 32.949. This measure indicates that the forecasting inaccuracy is low. The forecasts for red gram and green gram cultivated area and production during 2018-19 to 2022-23 showing increasing trend are given in table 6.

Table 6: Forecasted values of green gram and red gram cultivated area and production with 95% Confidence Level (CL)

Green gram						
Year	**Area (000'ha)**	**LCL**	**UCL**	**Production (000'tonnes)**	**LCL**	**UCL**
2018-19	4198.58	2997.45	5399.72	3775.23	2211.97	5338.48
2019-20	4238.22	2549.10	5927.35	3976.09	1962.28	5989.91
2020-21	4277.87	2213.03	6342.70	4176.96	1796.39	6557.53
2021-22	4317.51	1935.50	6699.52	4377.83	1679.90	7075.75
2022-23	4357.15	1695.51	7018.80	4578.70	1597.01	7560.39

Red gram						
Year	**Area (000'ha)**	**LCL**	**UCL**	**Production (000'tonnes)**	**LCL**	**UCL**
2018-19	1560.59	790.975	2330.20	1080.97	24.6671	2137.28
2019-20	1620.02	722.991	2517.05	1181.15	114.368	2476.66
2020-21	1669.72	693.418	2646.02	1263.12	170.443	2696.68
2021-22	1711.28	682.169	2739.39	1330.21	188.811	2849.22
2022-23	1746.03	683.192	2808.87	1385.10	188.548	2958.76

LCL – Lower Confidence Level
UCL – Upper Confidence Level

Conclusions

In our study the developed model for green gram cultivated area and production was found to be ARIMA (0, 1, 1) and ARIMA (0, 1, 1) and red gram cultivated area and production was found to be ARIMA (2, 1, 0) and ARIMA (2, 1, 0) respectively. The model can be used by researchers for forecast green gram and red gram cultivated areas and production in Madurai district of Tamil Nadu. From the forecast available by using the developed model, it can be seen that forecasted green gram and red gram crop cultivated area and production are increasing for the next five years.

Acknowledgement

I wish to express my indebtedness and deep sense of gratitude to my beloved and esteemed Chairman of the advisory committee, Dr. K. Prabakaran, Assistant Professor (Agricultural Statistics), of the Department of Agricultural Economics, Agricultural College & Research Institute, Madurai, Tamil Nadu, who has been a constant source of care, encouragement and enthusiasm. With lots of love and affection, I thank my parents Mr. M. Krishnan and Mrs. K. Thangaveni for providing me an opportunity in my educational life.

References

1. Box GE, Jenkins GM. Time Series Analysis Forecasting and Control. Holden Day. San Fran., USA, 1976.
2. Granger CWJ, Newbold P. Spectral methods. Econometrica. 1970; 30:424-438.
3. Cressie N. A Graphical Procedure for Determining Non-stationary in Time Series. JASA. 1988; 83:1108-15.
4. Biswas R, Bhattacharyya B. ARIMA modeling to forecast area and production of rice in West Bengal. Journal of Crop and Weed. 2013; 9(2):26-31.