International Journal of Farm Sciences **4(1)** :99-106, 2014

Forecasting areas and production of rice in India using ARIMA model

K PRABAKARAN and C SIVAPRAGASAM*

Agricultural College and Research Institute, Madurai, Tamil Nadu
***Department of Applied Research, Gandhigram Rural Institute, Gandhigram**
Email for correspondence: kpraba2020@rediffmail.com

ABSTRACT

Data on rice area and production for the period 1950-51 to 2011-12 were analyzed by time series method. Auto Correlation Function (ACF) and Partial Auto Correlation Function (PACF) were calculated for the data. Appropriate Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) model was fitted. Validity of the model was tested using standard statistical techniques. ARIMA (1, 1, 1) and ARIMA (1, 1, 1) model were used to forecast area and production in India for four leading years. The results also showed area forecast for the year 2015 to be about 44.75 thousand hectare with upper and lower limits 47.53 and 41.97 thousand hectares respectively. The model also showed rice production forecast for the year 2015 to be about 104.37 thousand tonnes with upper and lower limits 115.26 and 93.48 thousand tonnes respectively.

Keywords: Forecasting; area; production; rice; ARIMA; ACF; PACF

INTRODUCTION

Rice (*Oryza sativa* L) is the premier food crop of India. It belongs to the family Gramineae. According to Chatterjee (1948) there are altogether 23 species of genus *Oryza* of which 21 are wild and two viz *Oryza sativa* and *O glaberrima* are cultivated. *O sativa* is grown in all rice growing areas but *O glaberrima* is confined to West Africa only. Thus it indicates that there might have been two centres of origin of cultivated rice, South Eastern Asia (India, Burma and Thailand) and West Africa. It is also identified that Jeypore of Orissa state is the secondary centre of origin.

Among the rice growing countries in the world India has the largest area under rice crop and ranks second in production next to China. It is an important cereal crop of India and is cultivated on about 45.54 m ha area with an annual production of 99.18 million tonnes and a productivity of 1460 kg/ha. Among the rice producing states of India Tamil Nadu ranks sixth in production (5.67 million tonnes) and second in productivity of 3,070 Kg/ha with an area of 1.85 m ha. (Directorates of Economics and Statistics, Department of Agriculture and Corporation, 2010-2011). To meet the demand of increasing population and to maintain self sufficiency there is a need to develop rice varieties with higher yield potential and greater yield stability. Attempts are being made in all parts of rice growing countries to improve the productivity through hybrids and high yielding varieties and to develop suitable varieties for targeted environments.

The country's rice production declined to 89.13 million tonnes in 2009-10 crop year (July-June) from record 99.18 million tonnes in the previous year due to severe drought that affected almost half of the country. India could achieve a record rice production of 100 million tonnes in 2010-11 crop year on the back of better monsoon this year. India's rice production reached to a record high of 104.32 million tonnes in 2011-2012 crop year (July-June).

India is one of the world's largest producers of white rice accounting for 20 per cent of all world rice production. Rice is India's predominant crop and is the staple food of the people of the eastern and southern parts of the country. Rice is the basic food crop and being a tropical plant it flourishes comfortably in hot and humid climate. In 2009-10 total rice production in India amounted to 89.13 million tonnes which was much less than production of previous year 99.18 million tonnes.

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

METHODOLOGY

The annual data on rice cultivated area and production for the period from 1950-51 to 2011-12 were used for forecasting the future values using Auto Regressive Integrated Moving Average (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. The main objective in fitting 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 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 and Newbold 1970).

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. 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 rice cultivated areas and production for the period 1950-51 to 2011-12 were used following the four stages of ARIMA model.

Model identification

For forecasting rice 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 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.

The newly constructed variable X_t can now be examined for stationarity. The graph of X_t was stationary in mean. The next step was to identify the values of p and q. For this the autocorrelation and partial auto correlation coefficients of various orders of X_t were 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. Three tentative ARIMA models were tested and the models which had minimum AIC (Akaike Information Criterion) and SBC (Schwartz Bayesian Criterion) were chosen. The models and corresponding AIC and SBC values are as under:

	ARIMA (p, d, q)	AIC	SBC
Rice area	110	190.534	194.756
	111	189.161	195.494
	112	190.695	199.138
Rice production	110	381.233	385.455
	111	373.137	379.47
	112	375.174	383.618

Therefore most suitable model was ARIMA (1, 1, 1) for rice area and ARIMA (1, 1, 1) for rice production which had lowest AIC and SBC values.

Model estimation and verification

Rice cultivated areas and production model parameters were estimated using SPSS package. Results of estimation are reported in Table 2 and 3.

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 auto correlations of the residuals of various orders. The ACF and PACF of the residual (Fig 3 and Fig 4) also indicate 'good fit' of the model.

Table 1. Auto Correlations and Partial Auto Correlations for rice area and production

Rice area					Rice production				
Lags	ACF	SE	PACF	SE	Lags	ACF	SE	PACF	SE
1	-.410	.125	-.410	.128	1	-.547	.125	-.547	.128
2	.056	.124	-.135	.128	2	.108	.124	-.273	.128
3	-.012	.123	-.051	.128	3	.009	.123	-.100	.128
4	-.269	.122	-.352	.128	4	-.176	.122	-.292	.128
5	.231	.121	-.050	.128	5	.219	.121	-.062	.128
6	-.115	.120	-.086	.128	6	-.171	.120	-.133	.128
7	.110	.119	.014	.128	7	.094	.119	-.082	.128
8	.092	.117	.110	.128	8	.140	.117	.209	.128
9	-.163	.116	-.016	.128	9	-.211	.116	.069	.128
10	.138	.115	.076	.128	10	.116	.115	.026	.128
11	-.056	.114	.124	.128	11	-.009	.114	.151	.128
12	-.114	.113	-.083	.128	12	-.138	.113	-.071	.128
13	.076	.112	-.078	.128	13	.120	.112	-.124	.128
14	-.157	.111	-.173	.128	14	-.075	.111	-.084	.128
15	.255	.109	.094	.128	15	.040	.109	-.155	.128
16	-.098	.108	-.031	.128	16	.013	.108	-.173	.128

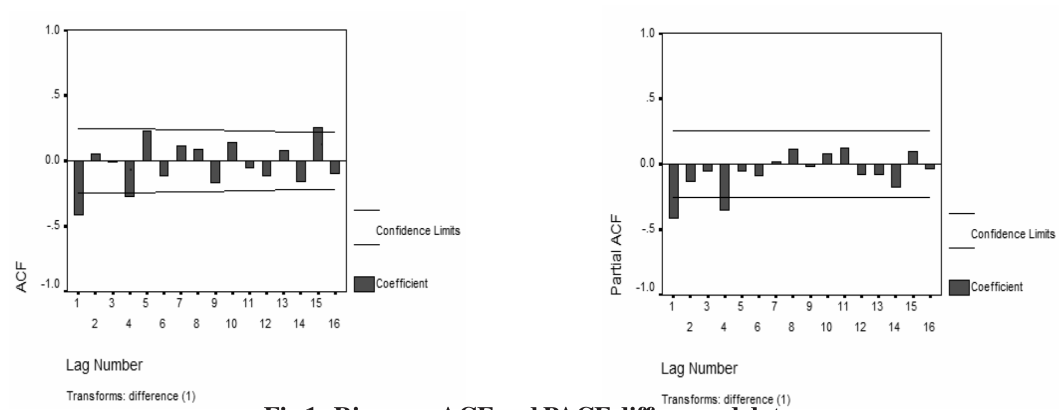

Fig 1. Rice area ACF and PACF differenced data

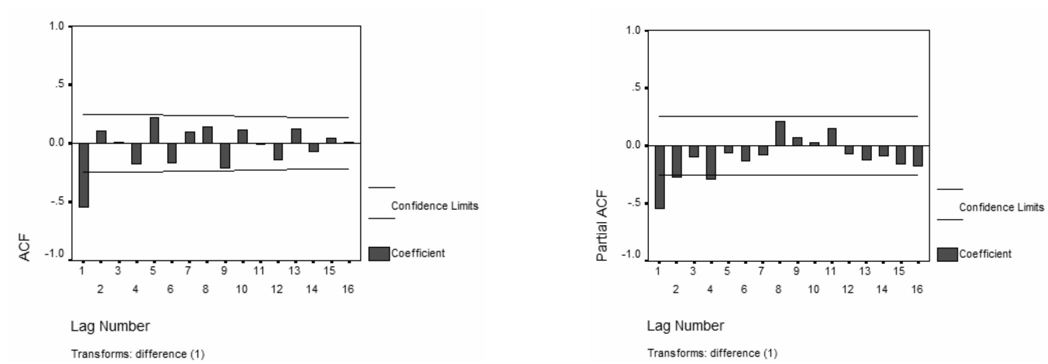

Fig 2. Rice production ACF and PACF differenced data

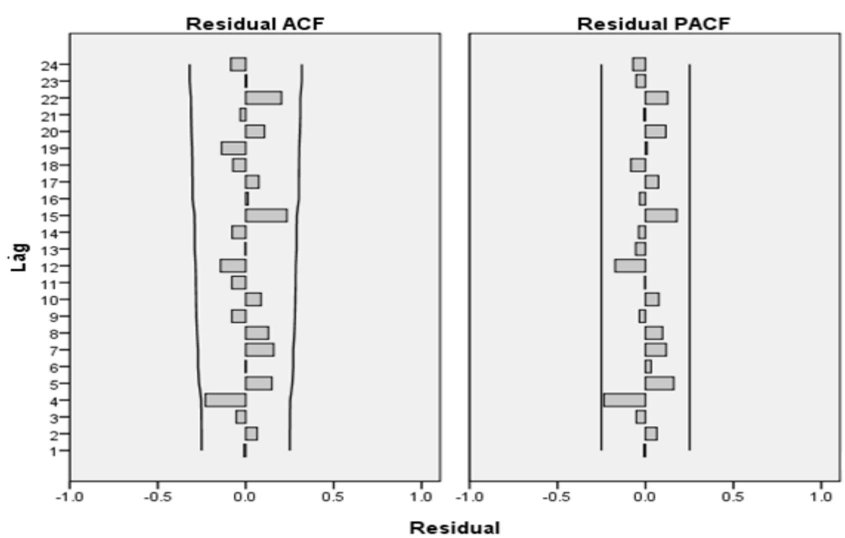

Fig 3. ACF and PACF of residuals of fitted ARIMA model for rice area

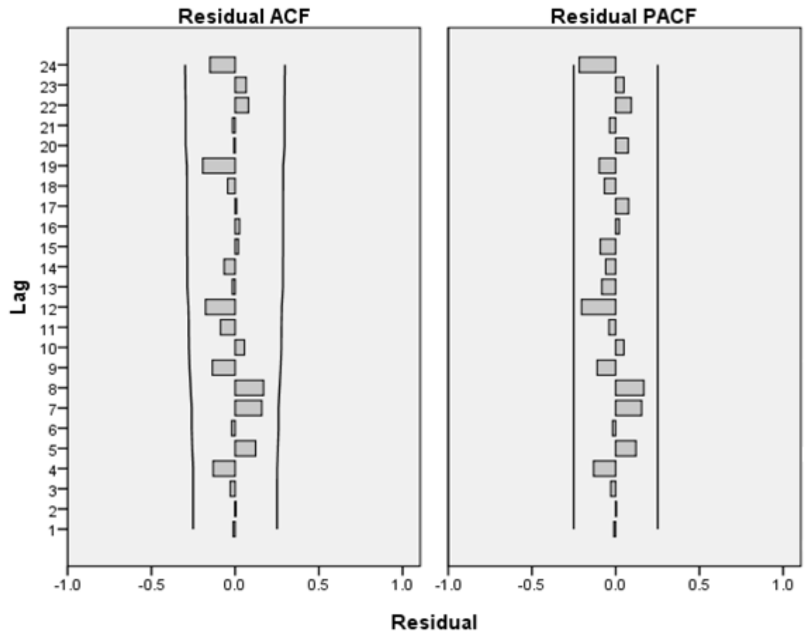

Fig 4. ACF and PACF of residuals of fitted ARIMA model for Rice production

Table 2. Estimates of the fitted ARIMA model for rice area

Model Fit statistics								Box- Ljung Q	
St R-sq	R-sq	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	Sig
0.22	0.94	1.11	2.03	0.79	9.72	4.00	0.42	17.73	0.34

Table 3. Estimates of the fitted ARIMA model for rice production

Model Fit statistics								Box- Ljung Q	
St R-sq	R-sq	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized	Statistics	Sig
0.42	0.96	5.00	6.66	3.58	29.79	18.78	3.42	11.45	0.78

St R-sq: Stationary R-Square, R-Sq: R-Square, RMSE: Root Mean Square Error,

MAPE: Mean Absolute Percentage Error, MAE: Mean Absolute Error, MaxAPE: Maximum Absolute Percentage Error, MaxAE: Maximum Absolute Error

Forecasting with ARIMA model

Mainly 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 measure of the sample period forecasts accuracy was computed. The MAPE for rice cultivated area turned out to be 2.03 and rice production turned out to be 6.66. This measure indicates that the forecasting inaccuracy is low. The forecasts for rice area and production during 2012 to 2015 showing increasing trend are given in Table 4.

Table 4. Forecasted values of rice cultivated areas and production with 95% Confidence Level (CL)

Year	Forecasted value (thousand hectare)	LCL	UCL	Forecasted value (thousand hectare)	LCL	UCL
2012-13	44.11	41.89	46.33	99.72	89.72	109.72
2013-14	44.32	41.85	46.77	101.92	91.79	112.05
2014-15	44.53	41.90	47.15	103.06	92.52	113.61
2015-16	44.75	41.97	47.53	104.37	93.48	115.26

LCL: Lower Confidence Level, UCL: Upper Confidence Level

CONCLUSION

In the study ARIMA (1, 1, 1) and ARIMA (1, 1, 1) were developed models for rice cultivated areas and production of rice respectively. From the forecast available by using the developed model it can be seen that forecasted rice cultivated areas and production were to increase in the next four years. The validity of the forecasted value can be checked when the data for the lead periods become available. The model can be used by researchers for forecasting rice cultivated areas and production in India. However data need to be updated from time to time with incorporation of current values.

REFERENCES

Box GEP and Jenkin 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 Science **18**:185-192.

Granger CWJ and Newbold P 1970. Spectral methods. Econometrica **30**: 424-438.

Makridakis S and Hibbon M 1979. Accuracy of forecasting: an empirical investigation. Journal of Royal Statistical Society A **41(2)**: 97-145.

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: 26.6.2013 ***Accepted: 23.9.2013***