Golden Research Thoughts
ISSN 2231-5063
Volume-3, Issue-3, Sept-2013

Available online at www.aygrt.isrj.net
DOI : 10.9780/2231-5063/332013/2831

Forecasting Cultivated Areas And Production Of Wheat In India Using ARIMA Model

K. Prabakaran , C. Sivapragasam , C. Jeevapriya And A. Narmatha

Assistant Professor (Agricultural Statistics), Department of Agricultural Economics, Agricultural College &Research Institute, Madurai.
Professor (Statistics), Department of Applied Research, Gandhigram Rural Institute, Gandhigram
Research Scholars, Department of Agricultural Economics, Agricultural College &Research Institute, Madurai.

ABSTRACT. Wheat area and production in India data for the period of 1950-51 to 2011-12 were analyzed by time series methods. 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, 0) model were used to forecast area and production in India for five leading years. The results also shows area forecast for the year 2015 to be about 31.46 thousand hectare with upper and lower limit 34.25 and 31.46 thousand hectares respectively. The model also shows Wheat production forecast for the year 2015 to be about 97.73 thousand tonnes with upper and lower limit 107.55 and 87.92 thousand tonnes respectively.

KEYWORDS: Forecasting, Area, Production, Wheat, ARIMA, ACF, PACF.

INTRODUCTION:

Wheat (*Triticum* spp.) is a cereal grain, originally from the Levant region of the Near East and Ethiopian Highlands, but now cultivated worldwide. In 2010 world production of wheat was 651 million tons, making it the third most-produced cereal after maize (844 million tons) and rice (672 million tons). In 2009, world production of wheat was 682 million tons, making it the second most-produced cereal after maize (817 million tons), and with rice as close third (679 million tons).

This grain is grown on more land area than any other commercial food, World trade in wheat is greater than for all other crops combined. Globally, wheat is the leading source of vegetable protein in human food, having higher protein content than either maize (corn) or rice, the other major cereals. In terms of total production tonnages used for food, it is currently second to rice as the main human food crop and ahead of maize, after allowing for maize's more extensive use in animal feeds.

PRODUCTION AND CONSUMPTION

In 2003, global per capita wheat consumption was 67 kg (150 lb), with the highest per capita consumption of 239 kg (530 lb) found in Kyrgyzstan. In 1997, global wheat consumption was 101 kg (220 lb) per capita, with the highest consumption 623 kg (1,370 lb) per capita in Denmark, but most of this (81%) was for animal feed. Wheat is the primary food staple in North Africa and the Middle East, and is growing in popularity in Asia. Unlike rice, wheat production is more widespread globally though China's share is almost one-sixth of the world.

In the 20th century, global wheat output expanded by about 5-fold, but until about 1955 most of this reflected increases in wheat crop area, with lesser (about 20%) increases in crop yields per unit area. After 1955 however, there was a dramatic ten-fold increase in the rate of wheat yield improvement per year, and this became the major factor allowing global wheat production to increase. Thus technological innovation and scientific crop management with synthetic nitrogen fertilizer, irrigation and wheat breeding were the main drivers of wheat output growth in the second half of the century. There were some significant decreases in wheat crop area, for instance in North America.

Golden Research Thoughts
ISSN 2231-5063
Volume-3, Issue-3, Sept-2013
Available online at www.aygrt.isrj.net
DOI : 10.9780/2231-5063/332013/2831

World trade in wheat is greater than for all other crops combined. Demand of India's wheat in the world shows a rising trend. The country has exported 7,40,746.77 MT of wheat to the world for the worth of Rs.1,023.29 crores during the year of 2011-12. Major Export Destinations (2011-12)**:** Bangladesh, United Arab Emirates, Pakistan, Afghanistan and Thailand

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

METHODOLOGY

The annual data on Wheat cultivated area and production for the period from 1950-51 to 2011-12 were used for forecasting the future values using ARIMA models. 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).

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 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.

1. Identification
2. Estimating the parameters
3. Diagnostic checking and
4. Forecasting
5.

Step 1: Identification: Appropriate values of p, d and q are found first. The tools used for identification are the Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF) and the resulting correlograms and partial correlograms.

The general ARIMA (p, d, q) model is presented in simple form as

$$\varphi (B) \nabla^d X_t = \theta (B) U_t$$

where as B is the backshift operator defined by

$$B^m X_t = X_{t-m} \quad (m = 0,1,2 \ldots p)$$

φ (B) is autoregressive operator of order ' p' defined by

$$\varphi (B) = 1 - \varphi_1 B^1 - \varphi_2 B^2 - \ldots - \varphi_p B^p$$

∇ is the backward difference operator defined by

$$\nabla X_t = X_t - X_{t-1} = (1-B) X_t$$

∇^d means the d^{th} difference of the series values X_t, θ (B) is the moving average operator of order 'q' defined by

$$\theta (B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \ldots - \theta_q B^q$$

U_t is white noise process having a normal probability distribution with mean zero and variance σ^2_u.

An example of ARIMA model is given below to clarify the general representation of the ARIMA (1,1,1) in explaining some features of the general ARIMA (p,d,q) model.

As could be seen in ARIMA(1,1,1) model where p=1, d=1, q=1

$$\varphi (B) = 1 - \varphi_1 B^1$$

$$\nabla^1 = (1-B)^1 = 1 - B \text{ and}$$

$$\theta (B) = 1 - \theta_1 B^1 = 1 - \theta_1 B^1$$

ISSN 2231-5063 Available online at www.aygrt.isrj.net
Volume-3, Issue-3, Sept-2013 DOI : 10.9780/2231-5063/332013/2831

Thus the model becomes

$(1-\varphi_1 B)(1-B) X_t = (1-\theta_1 B^1) U_t$

i.e. $(1-\varphi_1 B^1)(X_t - X_{t-1}) = U_t - \theta_1 U_{t-1}$

i.e. $X_t - \varphi_t X_{t-1} - X_{t-1} + \varphi_t X_{t-2} = U_t - \theta_1 U_{t-1}$

i.e. $(X_t - X_{t-1}) - \varphi (X_t - X_{t-1}) = U_t - \theta_1 U_{t-1}$

i.e. $W_t - \varphi_1 W_{t-1} = U_t - \theta_1 U_{t-1}$

φ_1 and θ_1 are the parameters of the model ARIMA (1, 1, 1).

Similarly φ_i (i = 1,2,….p) and θ_j (j = 1,2…q) are the parameters of the general ARIMA (p, d, q) model. φ_i (i = 1,2…p) are the Autoregressive (AR) parameters and θ_j (j = 1,2…q) are the moving average (MA) parameters

The Autoregressive (AR) operator φ (B) is assumed to be stationary and θ (B), the moving average (MA) operator is assumed to be invertible, $\{a_j\}$ is sequence of independent and identically distributed random variables with mean zero and variance σ^2, and the ARIMA(p, d, q) process becomes ARMA (p, q) process by suitable transformation of the variables and is given by $\varphi(B) w_t = \theta(B) a_t$

i.e., $w_t = \sum_{i=1}^{p} \varphi_i w_{t-i} - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t$

where, w_t consists of (n – d) observations

Step 2: Estimation: Having identified p and q values estimation of parameters of the autoregressive and moving average terms are estimated using simple least squares. The least square criterion for AR(1) is furnished below.

$$(1-\varphi_1 B) \tilde{X}_t = U_t$$

$(1-\varphi_1 B)(X_t - \mu) = U_t$ (Since ($\tilde{X}_t = X_t - \mu$))

$X_t - \varphi_1 X_{t-1} - \mu + \varphi_1 \mu = U_t$

$X_t = \varphi_1 X_{t-1} + \mu - \varphi_1 \mu + U_t$

$X_t = \mu (1 - \varphi_1) + \varphi_1 X_{t-1} + U_t$

Where $\mu (1 - \varphi_1)$ is the constant term

Step 3: Diagnostic checking: Having chosen a particular ARIMA model and having estimated its parameters the fitness of the model is verified. One simple test is to see if the residuals estimated from the model are white noise, if not we must start with other ARIMA model. The residuals were analyzed using Box-Ljung Statistic.

Step 4: Forecasting: One of the reasons for the popularity of the ARIMA modeling is its success in forecasting. In many cases, the forecasts obtained by this method are more reliable than those obtained from the traditional econometric modeling, particularly for short-term forecasts. An Autoregressive Integrated Moving Average Process model is a way of describing how a time series variable is related to its own past value. 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 following tests (Markidakis and Hibbon, 1979) such as Mean square error (MSE) and Mean Absolute percentage error (MAPE).

RESULTS AND DISCUSSION

In this study, we used the data for Wheat cultivated areas and production for the period 1950-51 to 2011-12. 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 Wheat cultivated areas and production as follows.

Model Identification

For forecasting Wheat 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.

The newly constructed variable X_t can now be examined for stationarity. The graph of X_t 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 X_t 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 three tentative ARIMA models and chose that model which has minimum AIC (Akaike Information Criterion) and SBC (Schwartz Bayesian Criterion). The models and corresponding AIC and SBC values are

	ARIMA (p, d, q)	AIC	SBC
Wheat area	110 **111** 112	158.096 **157.151** 159.124	162.318 **163.484** 167.567
Wheat production	**110** 111 112	**314.989** 317.021 318.677	**319.211** 323.354 327.12

So the most suitable model is ARIMA (1, 1, 1) for Wheat area and ARIMA (1, 1, 0) for Wheat production has the lowest AIC and SBC values.

Table:- 1 Auto Correlations and Partial Auto Correlations for Wheat area and production

Wheat area					Wheat production				
Lags	ACF	SE	PACF	SE	Lags	ACF	SE	PACF	SE
1	-.105	.125	-.105	.128	1	-.374	.125	-.374	.128
2	-.097	.124	-.110	.128	2	.153	.124	.015	.128
3	-.068	.123	-.093	.128	3	.041	.123	.119	.128
4	-.108	.122	-.142	.128	4	-.155	.122	-.124	.128
5	.102	.121	.055	.128	5	.215	.121	.119	.128
6	-.055	.120	-.074	.128	6	-.188	.120	-.062	.128
7	-.028	.119	-.049	.128	7	.063	.119	-.042	.128
8	.111	.117	.091	.128	8	.052	.117	.060	.128
9	-.068	.116	-.047	.128	9	-.169	.116	-.104	.128
10	-.175	.115	-.205	.128	10	-.031	.115	-.219	.128
11	.112	.114	.082	.128	11	.083	.114	.087	.128
12	-.078	.113	-.095	.128	12	-.124	.113	-.045	.128
13	.015	.112	-.064	.128	13	.177	.112	.087	.128
14	-.004	.111	-.031	.128	14	-.052	.111	.099	.128
15	.083	.109	.121	.128	15	-071	.109	-.098	.128
16	.110	.108	.053	.128	16	.175	.108	.062	.128

Golden Research Thoughts
ISSN 2231-5063
Volume-3, Issue-3, Sept-2013
Available online at www.aygrt.isrj.net
DOI : 10.9780/2231-5063/332013/2831

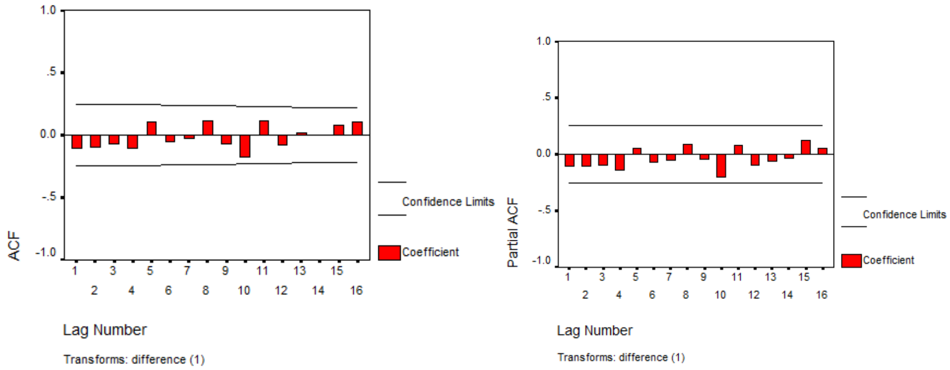

Fig.1 Wheat area ACF & PACF differenced data

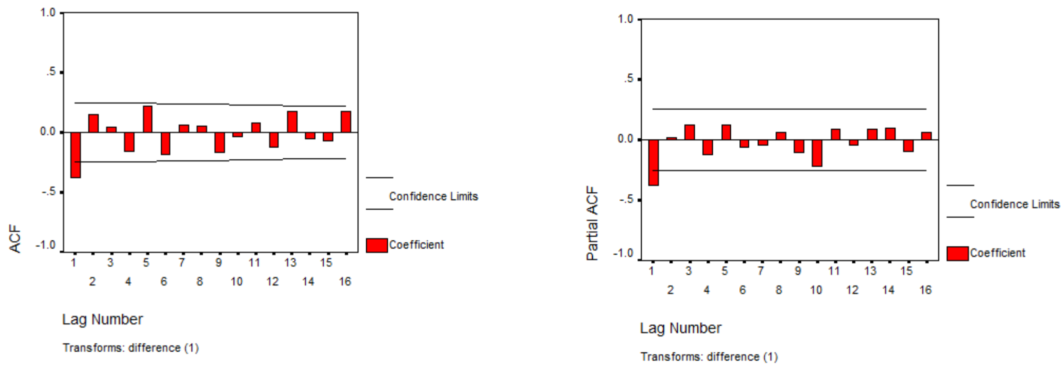

Fig.2 Wheat Production ACF & PACF differenced data

MODEL ESTIMATION AND VERIFICATION

Wheat 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. The ACF and PACF of the residual (Fig 3 and Fig 4) also indicate “good fit” of the model.

Golden Research Thoughts
ISSN 2231-5063
Volume-3, Issue-3, Sept-2013
Available online at www.aygrt.isrj.net
DOI : 10.9780/2231-5063/332013/2831

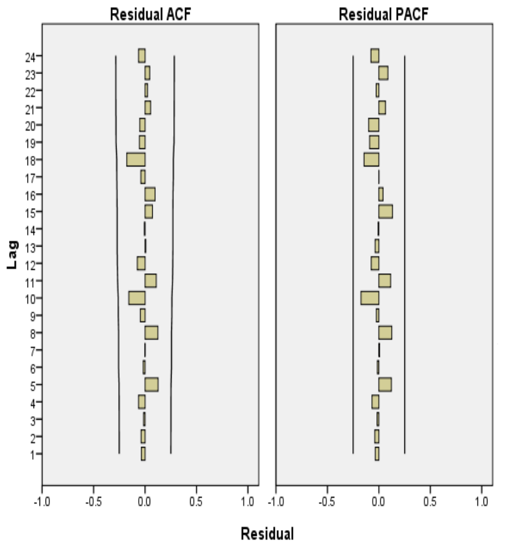

Fig 3 ACF and PACF of residuals of fitted ARIMA model for Wheat area.

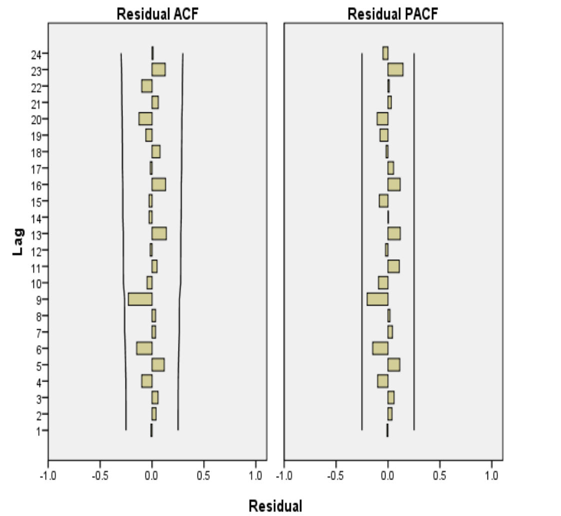

Fig 4 ACF and PACF of residuals of fitted ARIMA model for Wheat production.

Table:-2 Estimates of the fitted ARIMA model for Wheat area

Model Fit statistics								Box- Ljung Q	
St. R-sq.	R-sq.	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	Sig.
0.58	0.98	.86	3.63	0.68	16.56	2.13	.11	10.29	0.85

Table:-3 Estimates of the fitted ARIMA model for Wheat production

Model Fit statistics								Box- Ljung Q	
St. R-sq.	R-sq.	RMSE	MAPE	MAE	MaxAPE	MaxAE	Normalized BIC	Statistics	Sig.
0.15	0.99	3.14	8.98	2.48	39.11	8.03	2.43	11.46	0.83

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

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 Wheat cultivated area turns out to be 3.63 and Wheat production turns out to be 8.98. This measure indicates that the forecasting inaccuracy is low. The forecasts for Wheat area and production during 2012 to 2015 showing increasing trend are given in Table 4.

Table:-4 Forecasted values of Wheat Cultivated Areas and Production with 95%Confidence Level (CL)

Year	Forecasted values (thousand hectares)	LCL	UCL	Forecasted values (thousand tonnes)	LCL	UCL
2012-13	30.25	28.55	31.95	92.73	86.44	99.02
2013-14	30.67	28.43	32.91	95.16	87.80	102.50
2014-15	31.07	28.5	32.91	96.17	87.39	104.94
2015-16	31.46	28.67	34.25	97.73	87.92	107.55

LCL – Lower Confidence Level UCL – Upper Confidence Level

CONCLUSION

In our study the developed model for Wheat cultivated areas and production was found to be ARIMA (1, 1, 1) and ARIMA (1, 1, 0) respectively. From the forecast available by using the developed model, it can be seen that forecasted Wheat cultivated areas and production increases 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 Wheat cultivated areas and production in India. However, it should be updated from time to time with incorporation of current data.

REFERENCES

Box, G.E.P. and G.M. Jenkin, 1976. Time Series of Analysis, Forecasting and Control, Sam Franscico, Holden-Day, California. USA.

Granger, C.W.J. and Newbold, P., 1970, Spectral methods. *Econometrica.* 30: 424-438.

Makridakis, S and Hibbon, M., 1979, Accuracy of forecasting: An empirical investigation. *J.Roy.Statist.Soc.A.* 41(2):97-145.

Rahulamin, M.D., and Razzaque, M.A., “Autoregressive Integrated Moving Average Modeling for Monthly Potato Prices in Bangladesh”, Journal of Financial Management and Analysis, 13(1): 74-80, 2000.

Pindhyck, S.R., and Rubinfeld, D.L., Econometric Models and Economic Forecasts, NewYork: McGraw Hill, Inc., 1991.