Modeling and Forecasting for Agricultural Production – Food grains in India

Dr. P. Sameerabanu[1*], Dr. C. Sekhar[2]

[*1]Department of Mathematics & Statistics, Nalanda College of Agriculture, M.R.Palayam, Tirchy, Tamilnadu, India
[2]Department of Social Science, Nalanda College of Agriculture, M.R.Palayam, Tirchy, Tamilnadu, India
*Corresponding Author

Received:- 14 December 2022/ Revised:- 20 December 2022/ Accepted:- 24 December 2022/ Published: 31-12-2022

Abstract— *Food grains are cultivated in different proportions in different parts of India depending upon its intensity and cropping pattern. It is also useful to know the area, production and yield rates of food grains grown across the country over the years and to predict the future performance of these food grains by providing suitable forecasting methods. The data regarding production of food grains for the period from 1950 to 2021 is to be analyzed by applying forecasting method. Applying the worldwide accepted Box-Jenkins method would be a suitable tool for forecasting the performance of food grains in India. The study will prompt the government to take appropriate and necessary policy measures to sustain and arrest the fluctuating performance of food grains in India and take necessary steps to enhance its overall performance.*

Keywords— Modeling, Forecasting, ARIMA, ACF, PACF, Food Grain.

I. INTRODUCTION

The Green Revolution in India, which began in the late 1960s, helped the country achieve self-sufficiency in food by the late 1980s. The growth of agricultural production exceeded all expectations, reaching a total food grain production of 191.2 million tonnes in 1996–97. While the population grew from 439 million in 1961 to 834 million in 1991, at a rate of 2.2% per year, food grain production nearly kept pace with population growth. However, the struggle to maintain food security for a population that reached one billion in August 1999 continues and given that an estimated 35% of Indians are still below the poverty line (IFPRI, 1999). Other emerging concerns and developments are also beginning to have an impact on the future prospects of agricultural development. The yield of major food crops has either reached its highest level or has started declining. Ground water and surface water sources for irrigation are depleting. The traditional diet, heavy in grains, fails to meet basic nutritional requirements. The adverse effects that modern agriculture has on the environment and human health are a matter of concern for environmentalists, planners, and the public alike. These emerging issues have led many to believe that the original goal of the Green Revolution of 'productivity enhancement' needs to be redefined to include 'sustainability' and 'diversity' in food production. As the limits of growth in traditional food sectors are rapidly being reached, urgent national attention should be given to seeking ways to diversify food production

II. OBJECTIVES

- To analyze the future food grain production in India. Considering ever increasing Population.
- To suggest and use appropriate model for forecasting of food grain production in India.

III. METHODOLOGY

This study has been used on the secondary data obtained from “Ministry of Agriculture and Farmers Welfare, Government of India, 2020-2021”. The study period is from 1950 to 2021. This study attempts to analyze food grains production for the above mentioned years. Auto Regressive Integrative Moving Average (ARIMA) is a model fitted to time series data so that a variable has the ability to represent stationary as well as non-stationary time series. This model is also known as the Box Jenkins model which was introduced in 1960. While implementing this model it is essential to follow the steps involved in building this model. “Gretel” software is used for the analysis of the data.

A pure **Auto Regressive (AR only) model** is one where Yt depends only on its own lags. That is, Yt is a function of the 'lags of Yt'.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \epsilon_1$$

Likewise a pure **Moving Average (MA only) model** is one where Yt depends only on the lagged forecast errors.

$$Y_t = \alpha + \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

where the error terms are the errors of the autoregressive models of the respective lags. The errors Et and E(t-1) are the errors from the following equations:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_0 Y_0 + \epsilon_t$$

$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + .. + \beta_0 Y_0 + \epsilon_{t-1}$$

[Error: The beta coefficients in the second equation above is incorrect.]

That was AR and MA models respectively.

So what does the equation of an ARIMA model look like?

An ARIMA model is one where the time series was differenced at least once to make it stationary and you combine the AR and the MA terms. So the equation becomes:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q}$$

ARIMA model in words:

Predicted Yt = Constant + Lags of Y + Lagged forecast errors

Let's start with finding the 'd'.

Partial autocorrelation of lag (k) of a series is the coefficient of that lag in the autoregression equation of Y.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3}$$

That is, suppose, if Y_t is the current series and Y_t-1 is the lag 1 of Y, then the partial autocorrelation of lag 3 (Y_t-3) is the coefficient of Y_t-3 in the above equation.

Data:

TABLE 1
AGRICULTURAL PRODUCTION – FOODGRAINS (LAKH TONNES)

Year	Total Food Grains (Lakh Tones)
1950	508
1951	520
1952	592
1953	698
1954	680
1955	669
1956	699
1957	643
1958	771
1959	767
1960	820
1961	827
1962	802
1963	806
1964	894
1965	724
1966	742

1967	951
1968	940
1969	995
1970	1084
1971	1052
1972	970
1973	1047
1974	998
1975	1210
1976	1112
1977	1264
1978	1319
1979	1097
1980	1296
1981	1333
1982	1295
1983	1524
1984	1455
1985	1504
1986	1434
1987	1404
1988	1699
1989	1710
1990	1764
1991	1684
1992	1795
1993	1843
1994	1915
1995	1804
1996	1994
1997	1931
1998	2036
1999	2098
2000	1968
2001	2129
2002	1748
2003	2132
2004	1984
2005	2086
2006	2173
2007	2308
2008	2345
2009	2181
2010	2445
2011	2593
2012	2571
2013	2650
2014	2520
2015	2515
2016	2751
2017	2850
2018	2852
2019	2975
2020	3107
2021	3157

IV. ANALYSIS AND INTERPRETATION

4.1 Production under the Food grains:

The first step in a ARIMA model building is to plot the data into a graph. After plotting the graph, it will give an insight of the nature of the data. So in this study the series of data have been plotted in a graph as it is shown below in the figure. It is clear from the graph that data is not stationary and it deviates more from mean and variance. So now it becomes necessary to make the data stationary because it is one of the preconditions of building a model using secondary data to make it stationary. For this purpose, differencing of data is applied starting with first order, if necessary, going for the second order differencing. After going for the first order differencing the data becomes stationary as it is shown in the figure below. Now the data after first order differencing is now suitable for building a model.

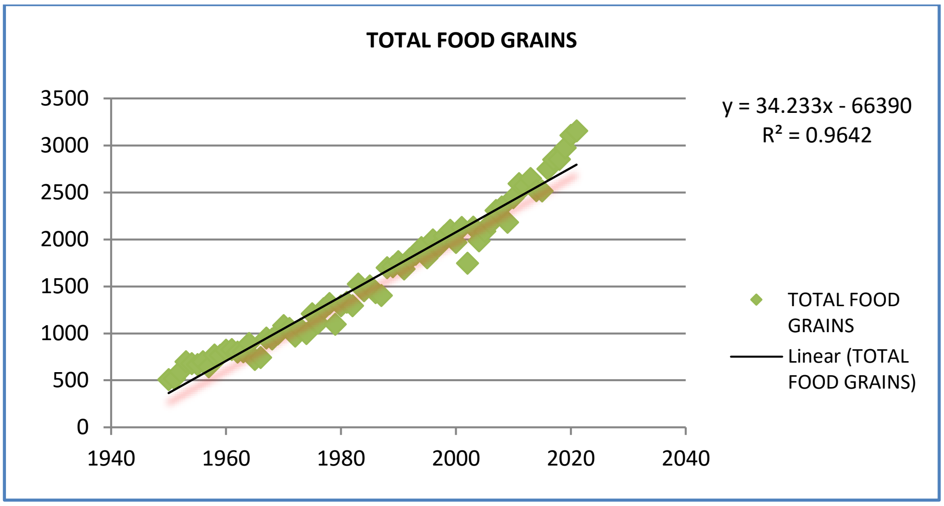

FIGURE 1: Linear Trend for Total Food Grain

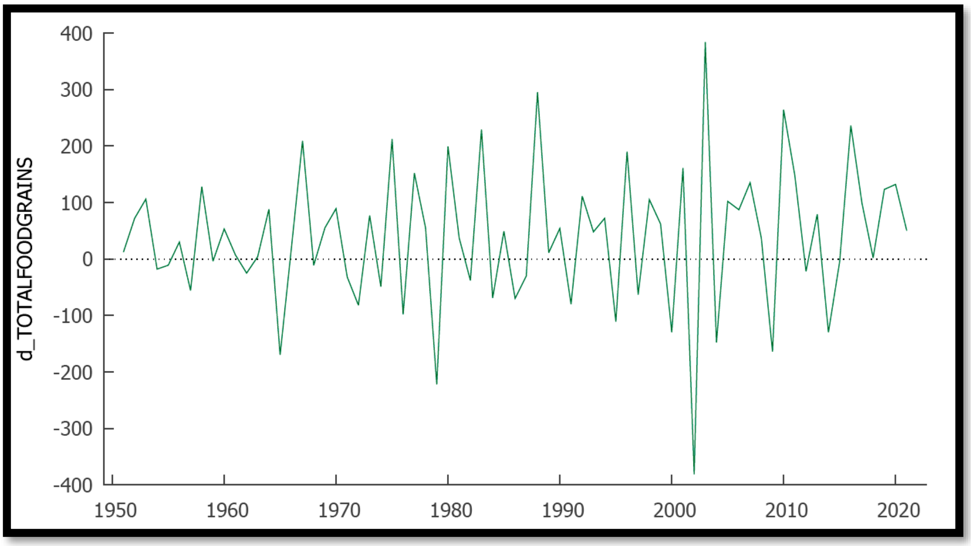

FIGURE 2: Time Series Plot for Stationary

The auto correlation functions and the partial auto correlation functions of the first order difference for the Production under food grains cultivation over the study are given in the above figure in the form of correlogram. The plots in the correlogram show that after the first order difference the data becomes stationary because the spikes are not lying outside the confidence interval which is shown in the dotted lines.

Auto Correlations and Partial Auto Correlations for food grains area and production

***, **, * indicate significance at the 1%, 5%, 10% levels using standard error 1/T^0.5

TABLE 2
AUTOCORRELATION FUNCTION FOR D_TOTAL FOOD GRAINS

LAG	ACF	PACF	Q-stat. [p-value]
1	-0.4992 ***	-0.4992 ***	18.4485 [0.000]
2	0.1085	-0.1873	19.3330 [0.000]
3	-0.0014	-0.0422	19.3331 [0.000]
4	-0.152	-0.2086 *	21.1205 [0.000]
5	0.1968 *	0.0258	24.1623 [0.000]
6	-0.152	-0.0549	26.0058 [0.000]
7	0.1487	0.0814	27.7959 [0.000]
8	-0.0326	0.0855	27.8831 [0.000]
9	-0.0301	0.0483	27.9591 [0.001]
10	0.0184	-0.0063	27.9877 [0.002]
11	0.0613	0.1488	28.3123 [0.003]
12	-0.179	-0.1481	31.1283 [0.002]
13	0.1863	0.0435	34.2287 [0.001]
14	-0.0791	0.0251	34.7977 [0.002]

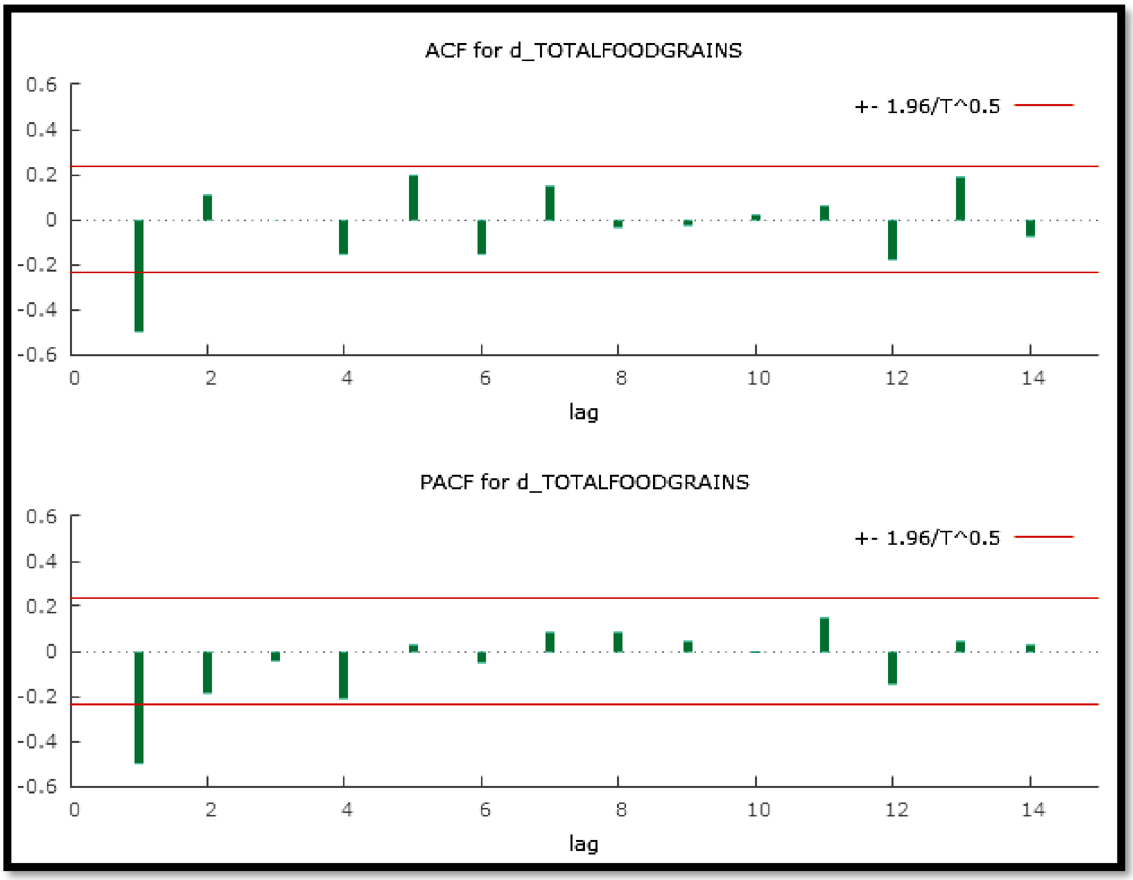

FIGURE 3: Correlogram plot for food grain

The correlogram of production of food grains after first order differentiation, it is clear that the d_total food grain becomes stationary and is of white noise as it shows no significant pattern in the correlogram and all the bars are lying within the confidence level. It is inferred that ARIMA (1,1,1) model will be suitable for the time series. The constructed model is assessed and estimated based on the parameters given, the corresponding diagnostics of the residuals, to select the best model for forecasting the future. Out of the alternate models for the production of food grains ARIMA of (1,1,1) was much suited and appropriate one.

TABLE 3
ESTIMATES OF THE FITTED ARIMA MODEL FOR FOOD GRAINS AREA

Fit Statistic	Mean
Stationary R-squared	0.763
R-squared	0.979
RMSE	106.187
MAPE	5.871
MAE	83.774
Normalized BIC	9.449

St. Stationary R-Square, R-Square, RMSE-Root Mean Square Error, MAPE-Mean Absolute Percentage Error, MAE- Mean Absolute Error.

Model 1: ARIMA, using observations 1952-2021 (T = 70)

Estimated using AS 197 (exact ML)

Dependent variable: (1-L) d_ TOTAL FOODGRAINS

Standard errors based on Hessian

TABLE 4

Parameter	coefficient	std. error	z	p-value
const	0.61161	0.420658	1.454	0.146
phi_1	−0.500054	0.101621	−4.921	8.62e-07 ***
theta_1	−1.00000	0.0387578	−25.80	8.59e-147 ***

	Real	Imaginary	Modulus	Frequency
		AR		
Root 1	-1.9998	0.0000	1.9998	0.5000
		MA		
Root 1	1.0000	0.0000	1.0000	0.0000

The result of the estimated output is presented in table above. It is clear from the table that the model coefficients are significant based on the P value which is below 0.01 percent and all the inverted AR and MA roots satisfy the minimum condition. The constructed model is assessed and estimated based on the parameters taken, the corresponding diagnostics of the residuals, to select the best model for forecasting the future. Out of the alternate models for the productivity of ARIMA of (1,1,1) is suitable and appropriate one.

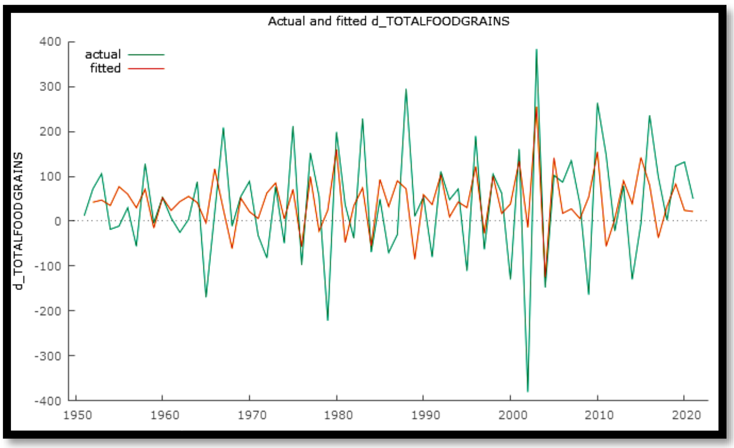

FIGURE 4: Forecasting for Actual and Fitted Value

The above Figure seen to actual and fitted value for first order difference in total food grain data.

TABLE 5
MODEL FIT

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
Total_Foodgrains-Model_1	0	0.763	9.676	16	0.883	0

TABLE 6
EXPONENTIAL SMOOTHING MODEL PARAMETERS

Model			Estimate	SE	t	Sig.
Total_Foodgrains-Model_1	No Transformation	Alpha (Level)	0.345	0.096	3.59	0.001
		Gamma (Trend)	0.087	0.083	1.042	0.301

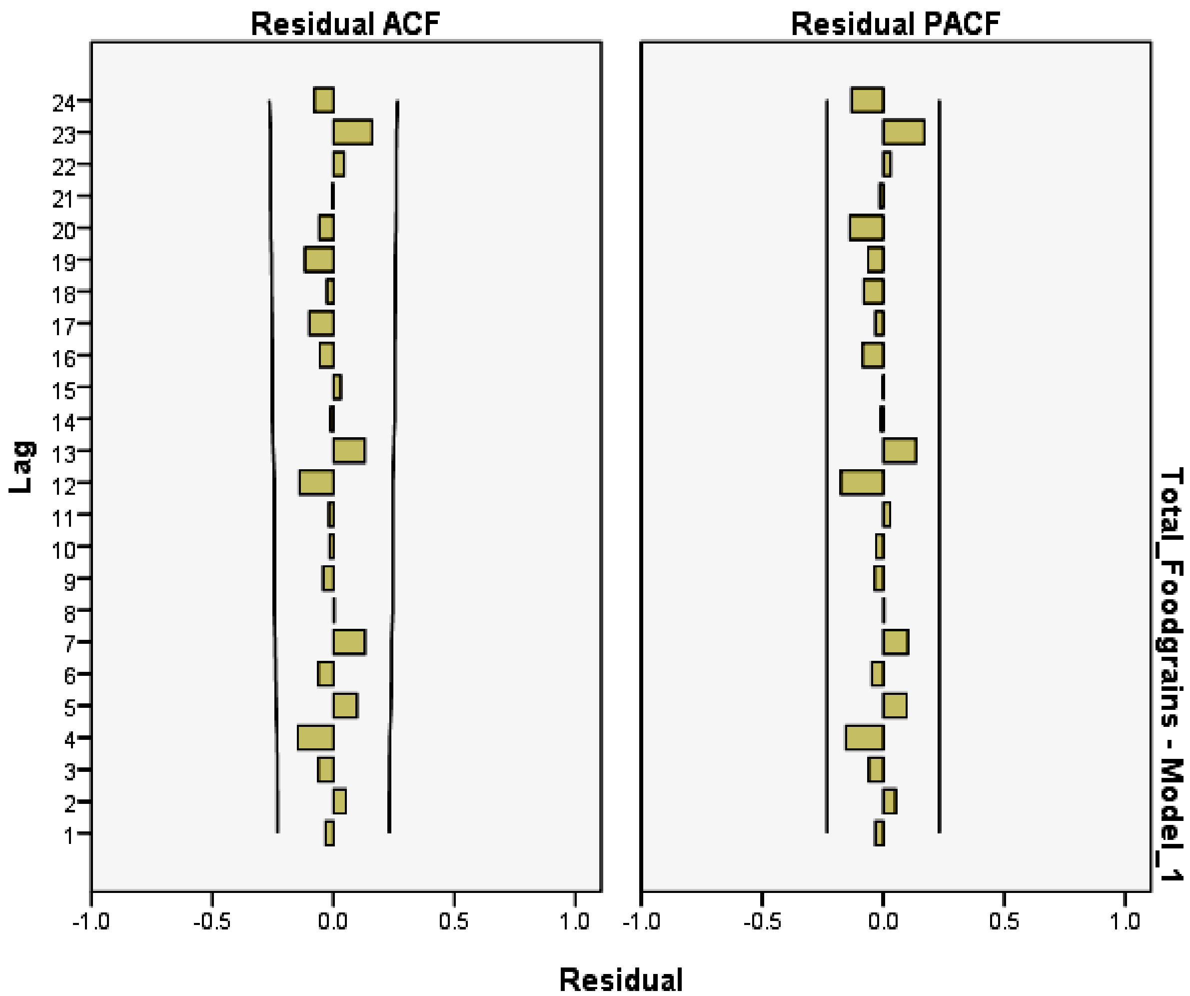

FIGURE 5: ACF and PACF of residuals of fitted ARIMA model for Food grains production

TABLE 7
FORECASTED VALUES OF FOOD GRAINS CULTIVATED AREA AND PRODUCTION WITH 95% CONFIDENCE LEVEL (CL)

Model		2022	2023	2024	2025	2026	2027	2028	2029	2030	2031
Total Food grains	Forecast	**3154**	**3211**	**3269**	**3327**	**3384**	**3442**	**3500**	**3557**	**3615**	**3673**
	UCL	3366	3438	3511	3586	3661	3738	3816	3895	3975	4055
	LCL	2942	2985	3027	3068	3107	3146	3183	3220	3255	3290

LCL – Lower Confidence Level UCL – Upper Confidence Level

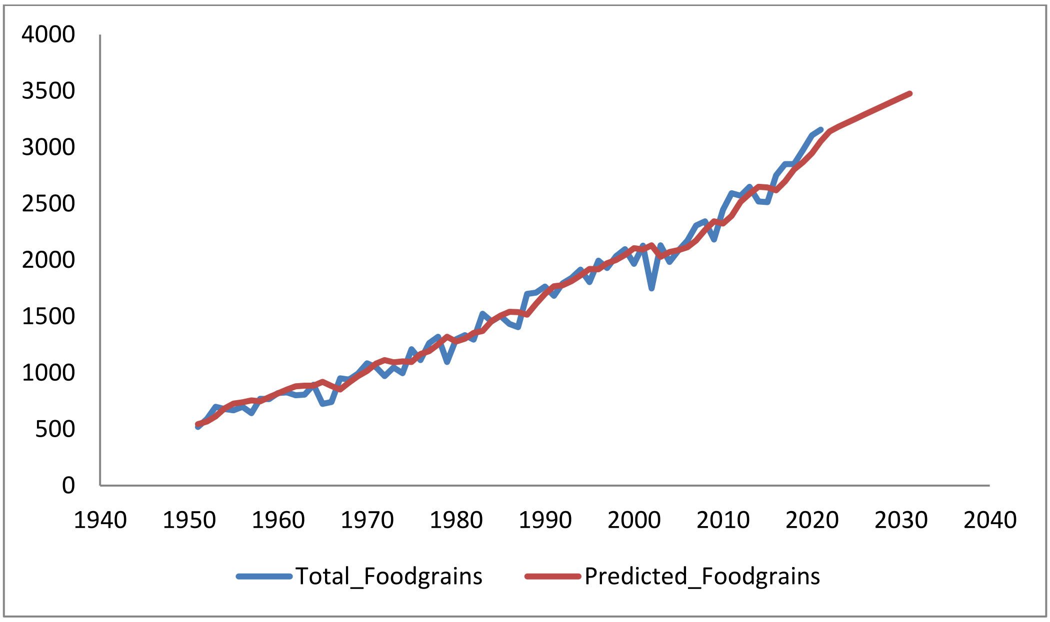

FIGURE 6: Forecasted values of food grains

V. CONCLUSION

In this study the model developed for analysis was ARIMA (1,1,1) for production of food grains ARIMA (1,1,1) was applied. The production and productivity of the different food grain crops have increased during the period under review due to the combined effects on area, production and productivity. In addition to that, productivity can be further increased by adopting appropriate technologies. In light of these findings, following suggestions are recommended for formation of appropriate policies.

REFERENCES

[1] Agricultural Statistic 2015,Government of India, Department of Agricultural, Cooperation & Farmers welfare, Ministry of Agricultural & Farmers welfare. Agricultural Statistics at a glance 2020-21 Government of India, Department of Agricultural, Cooperation & Farmers welfare, Ministry of Agricultural & Farmers welfare.

[2] Biswa ,R. and B. Bhattacharyya.2013."ARIMA modelling to forecast area and production of rice in West Bengal". Journal of Crop and Weed, 9(2):26-31.

[3] Box, G.E.P. and G.M. Jenkin, 1976.Time Series of Analysis, Forecasting and Control, Sam Franscico, Holden Day, California. USA. Directorate of Economics and statistics, Department of Agriculture, Cooperation & Farmers welfare .

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

[5] Gujarati,D.N. 2003." Basic Econometrics".4th Ed. Mc.Graw Hill. New York.USA

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

[7] Norman Badar et al 2015,Major food crops production and yield forecast in Pakistan", Pakistan J. Agric. Res. Vol. 28 No.3, 2015.

[8] Pindhyck, S.R., and Rubinfeld, D.L., Econometric Models and Economic Forecasts, New York: McGraw Hill, Inc., 1991.

[9] Pocket book of Economics & Statistics – 2015. Ministry of Agriculture & Farmers welfare.

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

[11] Directorate of Economics and Statistics, Department of Agriculture and Cooperation, Ministry of Agriculture, Govt. of India. (2015) Agricultural Statistics at a Glance.

[12] Sahu P.K. & Mishra P. (2014) International Journal of Agricultural and Statistical Sciences, 10(2), 425-435.

[13] Sahu P.K., Mishra P., Dhekale B.S., Vishwajith K.P. & Padmanaban K. (2015) American Journal of Applied Mathematics and Statistics, 3(1), 34-48.

[14] Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India (2012) Agriculture at a Glance.

[15] Cuddy J.D.A. & Della V.P.A. (1978) Oxford Bulletin of Economics and Statistics, 40(1), 79-85.

[16] Larson D.W., Jones E., Pannu R.S. & Sheokand R.S. (2004) Food Policy, 29(3), 257-273.

[17] Hasan M.N., Miah M.A.M., Islam M.S., Islam Q.M. & Hossain M.I. (2008) Bangladesh Journal of Agricultural Research, 33(3), 409-417. [10]Box G.E.P. & Jenkins G.M. (1976) Time Series Analysis: Forecasting and Control, Holden-Day, San Fransisco, USA.