Journal of Experimental Biology and Agricultural Sciences

<http://www.jebas.org>

ISSN No. 2320 – 8694

STATISTICAL MODELING AND FORECASTING OF ARIMA AND ARIMAX MODELS FOR FOOD GRAINS PRODUCTION AND NET AVAILABILITY OF INDIA

Soumik Ray^{1*}, Banjul Bhattacharyya²

¹Department of Agricultural Economics and Statistics, Centurion University of Technology and Management, M.S Swaminathan School of Agriculture, Paralakhemundi, Odisha-761211

²Department of Agricultural Statistics, Bidhan Chandra Krishi Viswavidyalaya, Faculty of Agriculture, Krishi Viswavidyalaya, Mohanpur, Nadia, West Bengal, India

Received – April 11, 2020; Revision – June 07, 2020; Accepted – June 23, 2020

Available Online – June 25, 2020

DOI: [http://dx.doi.org/10.18006/2020.8\(3\).296.309](http://dx.doi.org/10.18006/2020.8(3).296.309)

KEYWORDS

Parametric regression model

ARIMA

ARIMAX

Normality test

Forecasting

ABSTRACT

Agricultural development policies in India have aimed at reducing hunger, food insecurity, malnourishment and poverty at a rapid rate. The present work is designed with specific objectives to study the trend analysis of rice, wheat and total food grain in India for the period starting from 1950-2019. For stochastic trend model estimation, time series parametric regression models *i.e.* Linear model, Quadratic model, Exponential model, Logarithmic model, Auto Regressive Integrated Moving Average (ARIMA) and Auto Regressive Integrated Moving Average with explanatory variables (ARIMAX) were analyzed for estimating an appropriate econometric model to capture the trend of major food grain viz. rice, wheat, total food grain production and net availability of the country. Several goodness of fit criteria viz. Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE), and maximum R-squared values was worked out for finding best fitted models. Kolmogorov-Smirnov (K-S) test and Run-test were used to estimate the 'Normality' and 'Independence' of residuals of all data series respectively. By using the best fitted models, it was observed that the availability of rice (70.05 kg/year), wheat (70.73 kg/year) and total food grain (182.96 kg/year) will decrease in 2021 as comparatively to this year.

* Corresponding author

E-mail: raysoumik4@gmail.com (Soumik Ray)

Peer review under responsibility of Journal of Experimental Biology and Agricultural Sciences.

Production and Hosting by Horizon Publisher India [HPI]
(<http://www.horizonpublisherindia.in/>).
All rights reserved.

All the articles published by [Journal of Experimental Biology and Agricultural Sciences](#) are licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](#) Based on a work at www.jebas.org.

1 Introduction

Agriculture plays an important role in Indian economy, 65% of Indian population depend on agriculture and allied sectors and currently it contributes 16-17% of the GDP (Wagh & Dongre, 2016). During the last forty years, the share of Indian agriculture in gross domestic product has decreased, but extensive use of HYV seeds, modern irrigation, technology and fertilizer have contributed in increasing the agricultural productivity and achieving self-sufficiency in meeting food demand (Ray & Bhattacharyya, 2018). It is needless to mention that agriculture plays a key role in providing food availability globally (and nationally and locally in some agriculture-based countries); food production is the base for food security. It is an important source of income to purchase food; and also, may provide foods with high nutritional status. India's population will be reached 1.5 billion by 2030, the challenge facing the country is to produce more and more from diminishing irrigation water resources and per capita arable land and expanding abiotic and biotic stresses (Swaminathan & Bhavani, 2013). India currently produces about 283.37 million tonne (MT) of food grains to meet the needs of a population of 1.366 billion. Rice is one of the chief grains of India and this country has the largest area under rice cultivation, as it is one of the principal food crops. It is in fact the dominant crop of the country. India is becoming one of the most world's largest producers of rice, contributing for 20% of all world rice production. For the year 2019, it was calculated that India has produced 116.40 million tonne (MT) of rice with an increase of 3% over the previous year. Next to rice, wheat is the most important food grain of India and is staple food of millions of Indians, especially in the northern and north-western parts of the country. It is rich in protein, vitamin, and carbohydrate and provides balanced food. India is the 4th largest producer of wheat in the world after Russia, USA and China and accounts for 8.7% world's total production of wheat (Ramadas et al., 2019). Wheat is grown in India with a production of 102.20 Million tonne (MT) with an increase of 2.3% over the previous year in 2019. Uttar Pradesh, Punjab, Haryana, Madhya Pradesh, Rajasthan, Bihar, Maharashtra, Gujarat, Karnataka, West Bengal, Uttarakhand, Himachal Pradesh and Jammu & Kashmir contribute about 99.5% of total wheat production in the country. Rest 0.5% of total wheat production comes from Remaining States, namely, Jharkhand, Assam, Chhattisgarh, Delhi and other North Eastern States

Per capita net availability of food grains refers to the physical availability of food stocks in desired quantities. Using food grains as a proxy for food, availability of food grain is given by domestic production net of feed, seed, wastage, net imports as well as draw-down of stocks. Physical availability of food grains in any location within a nation mostly depends on storage, transport infrastructure

and market integration within the national territory. Per capita net availability of food grains relates to the supply of food through production, distribution, exchange and also relates to the productivity and population.

A number of studies have been conducted by several authors in different form to analyze the production, productivity behavior of major crop with some specific objective with respect to India as well as other countries. Among these some most common are (Sharma, 2010; Biswas & Bhattacharyya, 2013; Yogarajah et al., 2013; Vishwajith et al., 2014; Prabakaran & Sivapragasm, 2014; Sajid et al., 2015).

Most of the studies have emphasized analyzing the trends of major crops with ARIMA and ARIMAX model to check their future behavior. This research study made an attempt to examine the future behavior of production and per capita net availability of rice, wheat and total food grain by using ARIMA and ARIMAX models. Different parametric trend model i.e. linear, quadratic, cubic, exponential, logarithmic are also applied on the basis of past information. With the application of Box-Jenkins methodology (1976), it has taken as valuable process to utilize the forecast behavior of rice, wheat and total food grain data series. The study has also been made an attempt to compare the above methods with the help of actual data series.

2 Materials and Methods

Data with respect to production, productivity of rice, wheat and total food grain in India for the period of 1950-51 to 2018-19 and per capita net availability data series for the period 1950-51 to 2016-17 has been collected from the main website www.IndiaStat.com, Directorate of Economic and Statistics, Department of Agriculture and Corporation, Government of India. Population data series for period of 1950-51 has been collected from census of India publication also consider here in this study. As this study is mainly dealing with time series analysis, present data series have been verified initially for existence of outlier and randomness. Descriptive statistics are used to examine the basic feature of the data in any research work. The selected descriptive statistics i.e. mean, standard error, standard deviation, skewness, kurtosis, minimum, maximum along with average simple growth rate have been used to explain behavior of each data set in this study. Average simple growth rate (SGAR%) has been formulated as Dhekale et al. (2014)

$$SGAR(\%) = \frac{X_t - X_0}{X_0 \times n} \times 100$$

Where X_t is the value of the series for the last period, X_0 is the value of the series for the first period and n is the number of the periods.

Grubbs test has been applied for detecting outlier in this time series data as the test is particularly useful in case of large sample and easy to follow. SPSS software has been used for testing the outlier.

2.1 Regression model

Some of the parametric and non-parametric trend models are also used to this study for checking the trend behavior of the data series. The models along with their equations are given below (Ray & Bhattacharyya, 2016):

2.1.1 Linear model

A linear model is one in which all the parameters appear linearly and it is formulated as $X_t = a + bt + e_t$.

2.1.2 Quadratic model

The quadratic model can be used to model a series which “takes off” or a series which “dampens”. It expressed as $X_t = a + bt + ct^2 + e_t$.

2.1.3 Exponential model

The equation of exponential model is $X_t = a[Exp(bt)] + e_t$.

2.1.4 Logarithmic model

The equation of logarithmic model is given by $X_t = a + b \ln(t) + e_t$.

In order to apply these models, e_t is expressed as error term which is independently and identically normally distributed. In all the trend models model significant was tested by F test and individual regression coefficient is testing using t test. The best fitted model is selected on the basis of maximum value of R^2 and minimum values of RMSE, MAPE, MAE, AIC and SBC.

2.2 Time series analysis

Generally, time series, as a stochastic process, is an ordered sequence of observation taken at successive equally spaced points of time. The time series data can be univariate in case of crop production under consideration. The ARIMA (Auto Regressive Integrated Moving Average) class model is only applied to a univariate time series data. This method of time series modelling is often referred to as the Box-Jenkins approach (Box & Jenkins, 1976). At least 50 observations are required for estimating a good ARIMA model and a reasonably large sample size is required for a seasonal time series (Pankratz, 1983). With the ARIMA models forecast are made using the past of the process and are particularly suitable for short term forecasting and also forecasting seasonally enriched series. Box-Jenkins models are only reasonable for stationary time series with equi-spaced discrete time intervals.

2.2.1 Box-Jenkins Auto Regressive Integrated Moving Average (ARIMA) Models

Box-Jenkins (BJ) methodology (Box and Jenkins of Time Series Analysis: Forecasting and Control) is used here for time series

analysis which is technically known as the ARIMA methodology. The ARIMA Model Includes: The Autoregressive (AR) model, The Moving Average (MA) Model, The ARMA Model.

2.2.1.1 The Autoregressive (AR) Model

The Simplest form of the ARIMA model is called the autoregressive model. Here z_t stand for the value of a stationary time series at time t , that is, a time series that has no trend, but fluctuates about a constant value referred to as the *level* of the series. By autoregressive, we assume that current z_t values depend on *past* values from the same series. In symbols, at any t ,

$$z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t$$

$$Z_t = C + \varepsilon_t + \sum_{i=1}^p \phi_i z_{t-i}$$

Where C is the constant level, z_{t-1} , z_{t-2} , ..., z_{t-p} are past series values (lags), the ϕ 's are coefficients (similar to regression coefficients) to be estimated, and ε_t is a random variable with mean zero and constant variance. The ε_t 's are assumed to be independent and represent random error. Some of the ϕ 's may be zero. If z_{t-p} is the furthest lag with a nonzero coefficient, the AR model is said to be of order p , denoted $AR(p)$.

2.2.1.2 The Moving Average (MA) Model

z_t can also be modeled as a linear combination of white noise stochastic error terms. This type of model is known as a moving average (MA) model. If z_t is considered as a weighted average of the uncorrelated ε_t 's, $MA(q)$ moving average component of order q , which relates each z_t value to the q residuals of the q previous z estimates may be expressed as

$$z_t = e_t - q_1 e_{t-1} - q_2 e_{t-2} - \dots - q_q e_{t-q}$$

2.2.1.3 The ARMA Model

The AR and MA models for stationary series to account for both past values and past shocks may be combined. Such a model is called an $ARMA(p, q)$ model with p order AR terms and q order MA terms. Thus, an $ARMA(p, q)$ model is written as

$$z_t = C + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t - q_1 e_{t-1} - q_2 e_{t-2} - \dots - q_q e_{t-q}$$

2.2.2 Augmented Dickey Fuller (ADF) test (Stationarity test)

Dickey & Fuller (1979) established the Augmented Dickey Fuller test and it can be presented as

$$\Delta Y_t = \alpha + \beta_t + \delta Y_{t-1} + \sum \lambda_i \Delta Y_{t-i} + e_t$$

Where ΔY_t is the first difference of Y and α allows for a non-zero intercept or drift component i.e., constant, t is included to allow for deterministic trend as the data may be trend stationary. The null hypothesis here is Y_t has a unit root ($H_0: \delta=0$) against δ is negative. Thus, the test consists of testing the negativity of δ in above equation. The test statistics is given by

$$DF_t = \frac{\hat{\delta}}{SE(\hat{\delta})}$$

It can be compared to the relevant critical value for the Dickey-Fuller Test. If the test statistics is less than the critical value, then the null hypothesis of $\delta=0$ is rejected and the data is stationary.

2.2.3 ARIMA models

Most real time series show a trend, an average increase or decrease over time which means that they are Φ non-stationary i.e., they are integrated. Series also show cyclic behavior. Trends and cycles can be removed from a series through differencing. By differencing several times and/or at different lags, most series can be converted to a stationary series and then ARMA model for z_t is applied. Thus, the combined model for the original univariate time series, which involves autoregression, moving average, and integration, is termed as *ARIMA* (p, d, q) model (model of orders p, d , and q) with p AR terms, d differences, and q MA terms. The ARIMA model is often a parsimonious description of the behavior of a series. Box-Jenkins method consists of the following steps,

2.2.3.1 Identification

To identify the model of ARIMA (p, d, q) concepts of time-domain and frequency-domain analysis i.e. autocorrelation function (ACF), partial autocorrelation function (PACF) and spectral density function.

The autocorrelation function (ACF) and partial ACF (PACF) are very important for the definition of the internal structure of the analyzed series. The ACF $\rho(k)$ at lag k of the z_t series is the linear correlation coefficient between z_t and z_{t-k} , calculated for $k=0, 1, 2, \dots$

$$\rho_k = \frac{\text{Cov}(z_t, z_{t-k})}{\sqrt{\text{Var}(z_t) \text{Var}(z_{t-k})}}$$

The PACF is defined as the linear correlation between z_t and z_{t-k} , controlling for possible effects of linear relationships among values at intermediate lags.

Once the order of differencing has been diagnosed and the differenced univariate time series can be analyzed by the method of both time-domain and frequency-domain approach.

2.2.3.2 Estimation

Having identified the appropriate p and q value the next stage is to estimate the parameter of the autoregressive and moving average terms included in the model. The appropriate p, d and q values of the model and their statistical significance can be judged by t -distribution.

2.2.3.3 Diagnostic checking

Diagnostic checking consists of evaluating the adequacy of the estimated model. 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 are also judged by Ljung-Box statistic under null hypothesis that autocorrelation co-efficient is equal to zero. The Ljung-Box statistic, in case of large samples which follows a chi-square distribution with m degrees of freedom, is given by

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2 m$$

2.2.4 Estimation of parameter of ARIMAX model

ARIMAX model is a generalization of ARIMA model and is capable of incorporating an extraneous input variable. Nonlinear least square method employed to estimate the parameters of ARIMAX model. Following Bierens (1987), above equation can be written as follows;

$$y_t = \mu / \left(1 + \sum_{s=1}^r \gamma_s \right) + \left[\left(\sum_{s=1}^p \alpha_s L^s + \sum_{s=1}^q \gamma_s L^s \right) / \left(1 + \sum_{s=1}^r \gamma_s L^s \right) \right] y_t + \left[\left(1 + \sum_{s=1}^q \beta_s L^s \right) / \left(1 + \sum_{s=1}^r \gamma_s L^s \right) \right] x_t + e_t$$

$$\theta_0 = (\mu, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q, \gamma_1, \dots, \gamma_r)$$

$$z_t = (y_t, x_t), \phi(\theta_0) = \mu / \left(1 + \sum_{s=1}^r \gamma_s \right)$$

The above equation can be written as an ARX model:

$$y_t = \phi(\theta) + \sum_{s=1}^{\infty} \eta_s(\theta) z_{t-s} + e_t$$

Assuming that only $z_1, \dots, z_n$ are observed, nonlinear least square estimator of θ_0 , e_t is error term, and p , q and r are natural numbers specified in advance. The first step in building an ARIMAX model consists of identifying a suitable ARIMA model for the endogenous variable. The ARIMAX model concept requires testing for stationarity of endogenous variable before modeling.

2.2.4.1 Diagnostic checking

Diagnostic checking of ARIMAX model is same as ARIMA model by using the visualization of ACF and PACF graphs of residuals. One can accept the particular estimate fit, if the residuals estimated from this model white noise.

2.3 Model selection criteria using goodness of fit statistics

Among the competitive Box-Jenkins ARIMA model, ARIMAX model best model is selected on the basis of maximum R^2 , minimum root means square error (RMSE), minimum mean absolute percentage error (MAPE) and mean absolute error (MAE). Any model which has fulfilled most of the above criteria is selected. This section provides definitions of the goodness-of-fit measures used in time series modeling.

2.3.1 R-squared

An estimate of the proportion of the total variation in the series that is explained by the model. This measure is most useful when the series is stationary. Positive values mean that the model under consideration is better than the baseline model.

$$R^2 = \frac{\sum_{i=1}^n (\hat{X}_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

2.3.2 Root Mean Square Error (RMSE)

The square root of mean square error. A measure of how much dependent series varies from its model-predicted level, expressed in the same units as the dependent series.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{n}}$$

2.3.3 Mean Absolute Percentage Error (MAPE)

A measure of how much dependent series varies from its model-predicted level. It is independent of the units used and can therefore be used to compare series with different units.

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right|}{n} \times 100$$

2.3.4 Mean absolute error (MAE)

Measures how much the series varies from its model-predicted level. MAE is reported in the original series units.

$$MAE = \frac{\sum_{i=1}^n |X_i - \hat{X}_i|}{n}$$

2.4 Prediction and Forecasting

In statistics, prediction is a one of the important part of statistical inference known as predictive inference. Prediction can be considered within any of the several applications to statistical inference. Indeed, description of statistics may provide a means of transferring knowledge about a sample of a population to the whole population with other related populations, which is not the same as prediction over time. When the information is transferred to specific points in time, the process is superiorly known as forecasting. Fully formed statistical models for stochastic simulation purposes, so as to generate alternative versions of the time series, representing what might happen over non-specific time-periods in the future. Simple or fully formed statistical models to describe the likely outcome of the time series in the immediate future, given knowledge of the most recent outcomes (forecasting).

3 Results and Discussion

Univariate production and per capita net availability of rice, wheat and total food grain were analyzed using SAS (Version 9.3), SPSS and XLSTAT and the following results and discussion are obtained.

3.1 Per se performance

Descriptive statistics such as mean, standard deviation, skewness, kurtosis, simple growth rate percentage etc. are useful methods to check the basic behavior of data (Das et al., 2019). Since 1950 the production under rice has increased from 20.58 MT to 116.40 MT, registering a simple growth rate of almost 6.65% per annum. It also observed that the total food grains have increased from 50.83 MT to 283.37 MT with an average 153.09 MT and average simple growth rate of almost 6.53% per year. In case of wheat production, simple growth rate is almost 21.17% per annum, which is comparatively high than rice production. In this study, the production dataset of rice, wheat and total food grains indicate positive skewness and negative kurtosis, which means that there has been increasing order during early half of the study period and it, remain steady for a long time. This theory employed by Vishwajith et al. (2019). On the other hand, net availability of total food grains of India varied from 140.70 kg/year to 189.10 kg/year with an annual simple growth rate 0.47 percent. On an average of the series remains 165.69 kg/year during the study period. In case of rice, the data series has increased from 56.20 kg/year to 84.80 kg/year, registering a simple growth rate of almost 0.30% per

annum. In case of wheat net availability, simple growth rate is almost 3.17% per annum, which is comparatively high than other data series. The main reason of high growth rate of net availability is due higher production and productivity of wheat. The Negative kurtosis value of rice, wheat and total food grain indicates platykurtic in nature and negative value of skewness which reveals that has been marginal and consistent of availability during the later phase of investigation (Vishwajith et al., 2019). All the data series was tested for outlier by Grubbs method (Das et al., 2019). It was observed that the number of extreme observations (*i.e.* outlier) in the present data sets were zero, which is depicted in Table 1.

3.2 Parametric, Non-Parametric, ARIMA and ARIMAX Models

Before analyzing by ARIMA and ARIMAX, four parametric regression models were also fitted on the data series and the values of precision coefficient are given in Table 2 & 3. From the tables all the four models were almost equally precise, however

Table 1 Descriptive Statistics of Agricultural Production and per capita net availability of food grains in India

Parameters	<i>Rice</i>		<i>Wheat</i>		<i>Total food grain</i>	
	Prod.	Availability	Prod.	Availability	Prod.	Availability
Mean	63.26	70.00	46.60	46.94	153.09	165.69
Standard Error	3.39	0.76	3.64	1.78	8.14	1.27
Standard Deviation	28.34	6.25	30.47	14.56	68.10	10.44
Kurtosis	-1.31	-0.02	-1.29	-1.05	-1.13	-0.06
Skewness	0.21	-0.11	0.22	-0.27	0.29	-0.12
Minimum	20.58	56.20	6.18	21.10	50.83	140.70
Maximum	116.40	84.80	102.20	74.90	283.37	189.10
Outlier(Grubbs test)	No	No	No	No	No	No
SGAR%	6.65	0.30	21.17	3.17	6.53	0.47

Note: Prod.= Production(MT) Availability= Per capita net availability(Kg/Year)

Table 2 Fitting of linear and Non linear model of Agriculture production (million tonnes)

Rice							
Model	Parameter Estimation			Goodness of Fit			
	a	b1	b2	RMSE	MAPE	MAE	R ²
Linear	15.59	0.98		5.30	8.86	4.23	0.96
Quadratic	20.67	0.84	0.01	4.75	6.63	3.51	0.97
Exponential	23.85	0.98		5.79	8.17	4.66	0.96
Logarithmic	-18.18	24.25		13.47	25.16	11.45	0.72
Wheat							
Model	Parameter Estimation			Goodness of Fit			
	a	b1	b2	RMSE	MAPE	MAE	R ²
Linear	-5.44	0.98		4.80	21.84	3.86	0.97
Quadratic	1.33	0.57	0.43	3.85	13.40	3.12	0.98
Exponential	7.40	0.97		10.51	17.54	7.65	0.85
Logarithmic	-40.00	25.57		14.92	67.13	12.41	0.70
Food grain							
Model	Parameter Estimation			Goodness of Fit			
	a	b1	b2	RMSE	MAPE	MAE	R ²
Linear	38.47	0.98		11.27	7.66	9.06	0.96
Quadratic	52.45	1.88	0.01	9.56	5.69	7.32	0.97
Exponential	58.89	0.98		11.70	6.67	9.14	0.96
Logarithmic	-40.42	56.97		31.47	23.75	26.12	0.73

Significant at 5% level.

Table 3 Fitting of linear and Non linear model of per capita net availability (kg/year)

Rice							
Model	Parameter Estimation			Goodness of Fit			
	a	b1	b2	RMSE	MAPE	MAE	R ²
Linear	65.931	0.123		5.73	7	4.78	0.36
Quadratic	61.827	0.508	-0.006	5.46	6.31	4.29	0.54
Exponential	65.664	0.002		5.75	7.05	4.83	0.36
Logarithmic	60.336	2.983		5.53	6.51	4.45	0.44
Wheat							
Model	Parameter Estimation			Goodness of Fit			
	a	b1	b2	RMSE	MAPE	MAE	R ²
Linear	23.108	0.701		5.04	8.95	3.96	0.86
Quadratic	15.818	1.385	-0.011	3.97	6.75	2.9	0.91
Exponential	24.826	0.017		6.53	11.62	5.27	0.77
Logarithmic	2.295	13.512		6.04	12.56	4.71	0.8
Food grain							
Model	Parameter Estimation			Goodness of Fit			
	a	b1	b2	RMSE	MAPE	MAE	R ²
Linear	112.857	1.133		9.23	4.76	7.74	0.29
Quadratic	59.465	5.71	-0.066	8.58	4.29	6.97	0.59
Exponential	82.034	1.015		9.24	4.79	7.81	0.27
Logarithmic	150.055	4.553		8.77	4.48	7.3	0.61

Significant at 5% level

Quadratic model was superior for all data series of production and net availability based on goodness of fit criteria of models *i.e.* maximum value of R², minimum value of RMSE, MAPE, MAE.

After consideration of these four parametric regression models, ARIMA model for production data series and ARIMAX model for per capita net availability data series was employed in addition. Firstly, for applying the most time series models, the dataset must be checked of its stationarity (Prabakaran & Sivapragasm, 2014). Augmented Dickey Fuller (ADF) test is applied for testing the stationarity for all the data sets (Dickey & Fuller 1979). It is observed that all the dataset is presented as non-stationary at the level point which is insignificant at 5% level of significance. After differencing one the test is going to highly significant at 5% level of significance for production data series and at 1% level of significance for net availability data series respectively, which is depicted in Table 4. Thus, it is confirmed that first differencing for all series are perfect for modeling and forecasting.

After fixing the value of d as 1, values of p and q were determined as per autocorrelation and partial autocorrelation consideration. It

was observed that for rice production and net availability data series, only one significant spike for ACF at lag1 and two significant spikes PACF at lag 1 and lag2 was present. On the other hand, for net availability for wheat, two significant spikes for both ACF and PACF at lag1 and lag2 was formed. For total food grain for production and net availability data series, only one significant spike was present for both ACF and PACF at lag1 respectively, which are depicted in Figure1 & 2.

After determining the autocorrelation and partial autocorrelation, the best fitted ARIMA (p,d,q) was selected on the basis of maximum value of R², minimum value of RMAE, MAPE, MAE (Das et al., 2019). It was found that ARIMA (1,1,2) fitted well for rice and ARIMA (1,1,1) fitted well for both wheat and total food grain production data series respectively. The per capita net availability of food grains is strongly related with population of India and production and productivity of all food grains. As per capita net availability of different crops may be treated as the function of population, production and productivity, so ARIMAX model may be considered for modeling and forecasting of these data series. ARIMAX(1,1,2) model was fitted well for rice data

Figure 1 Correlogram of ACF and PACF for first differenced of production dataset

Figure 2 Correlogram of ACF and PACF for first differenced of per capita net availability dataset

Table 4 ADF test for food grains production and per capita net availability in India

Production				
Stationarity Test	Data type	ADF statistics	P value	Decision
Rice	Level	-2.4656	0.3337	Data Non- Stationary
	1st Difference	-4.1461	0.0083	Data Stationary
Wheat	Level	-2.5622	0.2891	Data Non- Stationary
	1st Difference	-4.3979	0.0035	Data Stationary
Total Food grain	Level	-2.1881	0.4791	Data Non- Stationary
	1st Difference	-6.0142	< 0.0001	Data Stationary
Significant at 5% level				
Per capita net availability				
Stationarity Test	Data type	ADF statistics	P value	Decision
Rice	Level	-2.477	0.3289	Non- Stationary
	1st Difference	-6.209	< 0.0001	Stationary
Wheat	Level	-1.7843	0.6918	Non- Stationary
	1st Difference	-8.6402	< 0.0001	Stationary
Total Food grain	Level	-2.9507	0.147	Non- Stationary
	1st Difference	-6.4668	< 0.0001	Stationary
Significant at 1% level				

Table 5 Best fitted ARIMA and ARIMAX models for Production and Net Availability datasets

Production									
Crops	ARIMA Model	R ²	RMSE	MAPE	MAE	Ljung- Box(Q) test			
						Statistics	DF	Critical Value	Sig
Rice	1,1,2	0.965	4.871	6.145	3.44	13.896	15	25.00	0.533
Wheat	1,1,1	0.980	3.391	8.124	2.536	15.484	16	26.30	0.489
Total food grain	1,1,1	0.971	10.49	5.898	8.182	13.672	17	27.59	0.69
Per Capita Net Availability									
Crops	ARIMAX Model	R ²	RMSE	MAPE	MAE	Ljung- Box(Q) test			
						Statistics	DF	Critical Value	Sig
Rice	1,1,2	0.793	3.41	3.422	2.394	14.062	15	25.00	0.521
Wheat	2,1,2	0.886	4.596	6.775	3.28	12.339	14	23.69	0.579
Total food grain	1,1,1	0.650	6.824	3.028	5.024	13.724	17	27.59	0.619

at 5% level of significant

series, whereas ARIMAX(2,1,2) model was considered as best fitted model for wheat and ARIMAX(1,1,1) was selected as best fitted model for net availability of total food grain data series which is presented in Table 5.

Ljung-Box Q statistic was also applied for estimating the diagnostic check of this best fitted model. It was observed that autocorrelation coefficient is independent for all best fitted models which also

depicted in Table 5. From the residual ACF and PACF plots of ARIMA and ARIMAX, it was clear that all autocorrelations and partial autocorrelations lie between 95% control limits as shown in Figure3. This also confirmed the 'good fit' of this selected model. For checking normality of residuals, K-S test was applied. It was observed that the calculated value of the test statistic was D_n (Cal.) < D_n (Tab.) for all ARIMA and ARIMAX models, says that the residuals are normally distributed given in Table 6. For checking the

Figure 3 Residual ACF and PACF of best fitted ARIMA and ARIMAX models

Table 6 Tests of normality and randomness of residuals

Test for Normality. (Kolmogorov-Smirnov test)						Run test		
Statistics	p-value	Critical Value			Z-Statistics	No of Runs	P-Value	
		5%	2%	1%				
Production (Best fitted ARIMA)								
Rice	0.148	0.088	0.163	0.183	0.195	0.733	32	0.463
Wheat	0.071	0.854				1.221	30	0.221
Total Foodgrain	0.081	0.723				0.488	33	0.625
Per capita net availability (Best fitted ARIMAX)								
Rice	0.154	0.133	0.177	0.198	0.212	0.549	26	0.582
Wheat	0.074	0.876				0.264	29	0.791
Total Foodgrain	0.081	0.820				1.328	35	0.185

Table 7 Model validation and forecasting for production and per capitaneet availability

Year	Production (MT)					
	Rice		Wheat		Total Food grain	
	A	P	A	P	A	P
2016	104.40	108.48	92.30	91.41	251.60	264.61
2017	109.70	108.79	98.50	92.25	275.10	263.79
2018	112.80	110.01	99.90	98.32	281.70	270.72
2019	116.40	112.11	102.20	101.29	283.37	278.59
2020		114.55		103.29		283.96
2021		116.22		104.78		287.28
2022		117.47		106.14		290.60
2023		118.85		107.54		293.93
Year	Per Capita Net Availability (Kg/year)					
	Rice		Wheat		Total Food grain	
	A	P	A	P	A	P
2014	72.3	76.32	66.8	62.51	178.60	176.88
2015	67.9	71.81	61.3	64.27	169.80	174.15
2016	67.2	73.30	72.9	66.94	177.70	171.50
2017	69.5	72.62	74.9	69.44	189.10	179.29
2018		70.42		69.74		183.53
2019		70.73		68.64		180.65
2020		78.19		71.62		188.77
2021		70.05		70.73		182.96

Note- A=Actual. P= Predicted

randomness of residuals, Run test was performed and it was observed that the probability value was greater than the 5% level of significance (*i.e.* >0.05) indicating residuals were distributed independently also. Thus, it can be concluded that the residuals are independent and follow normal distribution.

3.3 Forecasting

Finally, model validation and forecasting were done for both production and net availability of India from 2016-17 to 2023 and

2014-15 to 2021 respectively by using best fitted ARIMA and ARIMAX models. First four years dataset was used for validation of the model can be regarded as in sample forecast and last four years dataset was used for prediction purpose, which is probably known as out sample forecast (Ray et al., 2016; Dhekale et al., 2017), which is depicted in Table 7. By using best fitted ARIMA and ARIMAX models, it was observed that the actual and predicted graphs were closely related and predicted graph were lies within the 95% confidence intervals as captured in Figure 4. From the forecasted values, it can be concluded that for a few coming

Figure 4 Forecasting of production and per capita net availability by using ARIMA and ARIMAX models

years production of rice, wheat and total food grain will follow an increasing trend and it have been estimated as 118.85(MT), 107.54(MT) and 293.93(MT) respectively for the year 2023. But the availability of rice (70.05 kg/year), wheat (70.73 kg/year) and total food grain (182.96 kg/year) will decrease in next year as comparatively to this year. Thus, it can be said that India has produce higher production of rice, wheat, total food grain but the availability of this major food grains is changed steadily.

Conclusion

Time series analysis for forecasting may be considered as a useful practice of a model for prediction of future values based on previously observed values. The analysis of best fitted ARIMA and ARIMAX model and predicted forecasting pattern can play vital role to deal with future food security scenario and planning for policy makers in India. Finally, better management practice, technological

improvement, high government policies i.e. price support programmes, agricultural funding, etc. and enhancing relationship between research workers and farmers maybe important factors in sustaining this trend of production for long term.

Conflict of Interest

Authors would hereby like to declare that there is no conflict of interests that could possibly arise.

References

- Bierens HJ (1987) ARIMAX model specification testing with an application to unemployment in the Netherland. *Journal Econometrics* 35:161-90.
- Biswas R, Bhattacharyya B (2013) ARIMA modeling to forecast area and production of rice in West Bengal. *Journal of Crop and Weed* 9(2): 26-31.
- Box GEP, Jenkins GM (1976) *Time Series Analysis, Forecasting and Control*. 2nd Edition, Holden-Day, San Francisco, USA.
- Das SS, Ray S, Sen A, Siva GS and Das S (2019) Statistical study on modelling and forecasting of jute production in West Bengal, India. *International Journal of Current Microbiology and Applied Science* 8(7): 1719-1730.
- Dhekale BS, Sahu PK, Viswajith KP, Mishra P (2017) Analysis of growth, instability, modelling and forecasting of cotton production scenario in India. *Indian Journal of Economic and Development* 13 (2a): 211-216.
- Dhekale BS, Sahu PK, Viswajith KP, Mishra P, Noman MD (2014) Modeling and forecasting for tea production in West Bengal. *Journal of Crop and Weed* 10 (2) : 94-103.
- Dickey DA, Fuller WA (1979) Distribution of estimators for Autoregressive Time Series with a Unit root. *Journal of the American Statistical Association* 74: 427-431.
- Pankratz A (1983) *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*. Wiley series in Probability and Mathematical Statistics, USA.
- Prabakaran K, Sivapragasam C (2014) Forecasting areas and production of rice in India using ARIMA model. *International Journal of Farm Sciences* 4(1): 99-106.
- Ramadas S, Kumar TMK, Singh GP (2019) Wheat production in India: Trends and prospects. Shah F(ed). *Recent Advances in Grain Crops Research*, Intech Open, London, UK.
- Ray S, Bhattacharyya B (2016) A statistical investigation on analysis of food consumption pattern in India. *Journal of Crop and Weed* 12(3): 47-54.
- Ray S, Bhattacharyya B (2018) Statistical investigation of food grains demand and supply in India. *International Journal of Agriculture Science* 10 (11): 6200-6205.
- Ray S, Bhattacharyya B, Pal S (2016) Statistical modeling and forecasting of food grain in effects on public distribution system: An application of ARIMA model. *Indian Journal of Economic and Development* 12(4): 739-744.
- Sajid A, Badar N, Fatima H (2015) Forecasting production and yield of sugarcane and cotton crops of Pakistan for 2013-2030. *Sarhad Journal of Agriculture* 31(1):1-9.
- Sharma M (2010) ARIMA model for food grains production of India. *Southern Economist* 48(18):17-20.
- Swaminathan MS, Bhavani RV (2013) Food production and availability- Essential prerequisites for sustainable food security. *Indian Journal of Medical Research* 138(3): 383-391.
- Vishwajith KP, Dhekale BS, Sahu PK, Mishra P, Noman Md (2014) Time series modeling and forecasting of pulses production in India. *Journal of Crop and Weed* 10 (2) : 147-154.
- Vishwajith KP, Sahu PK, Mishra P, Monika D, Dubey A, Singh RB, Dhekale BS, Fatih C, Suman (2019) Modelling and forecasting of mung production in India. *Current Journal of Applied Science and Technology* 34(1): 1-19.
- Wagh R, Dongre A (2016) Agricultural sector: Status challenges and it's role in Indian economy. *Journal of Commerce & Management Thought* 7(2): 209-218.
- Yogarajah B, Elankumaran C, Vigneswaran R (2013) Application of ARIMAX Model for Forecasting Paddy Production in Trincomalee District in Sri Lanka. *Proceedings of the Third International Symposium, SEUSL: 6-7 July 2013, Oluvil, Sri Lanka*.