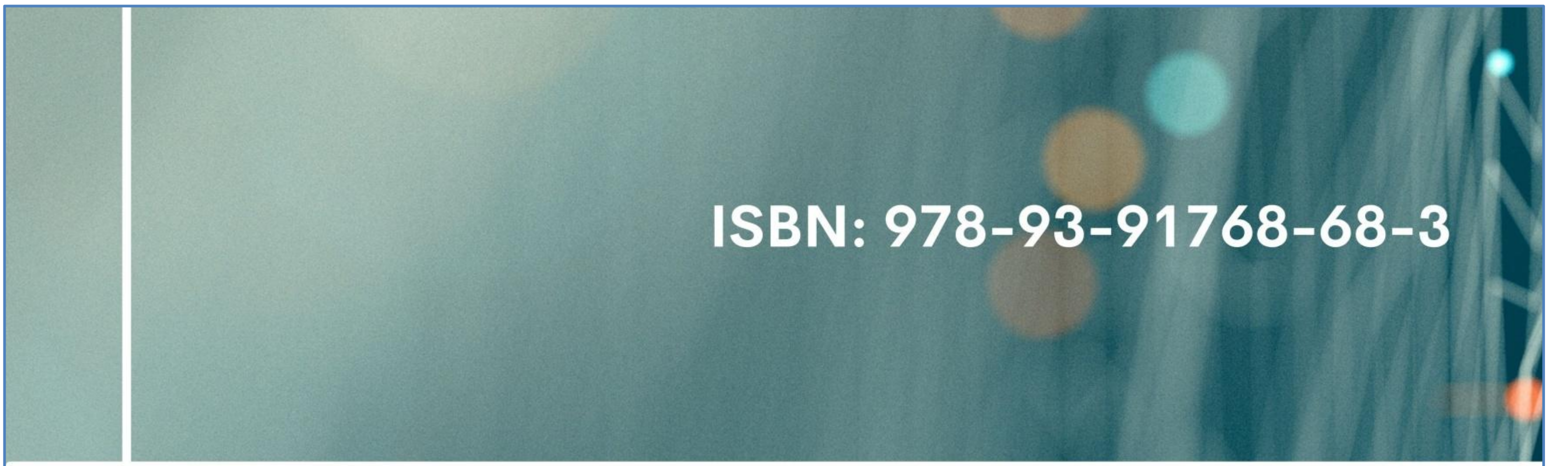


ISBN: 978-93-91768-68-3

Fitting of ARIMA Model For Forecasting Oilseed Crops

Madhu Chhanda Kishan
Dr. Abhiram Dash

Bhumi Publishing

FITTING OF ARIMA MODEL FOR FORECASTING OILSEED CROPS

ISBN: 978-93-91768-68-3

Madhu Chhanda Kishan

Assistant Professor (Agricultural Statistics),

College of Agriculture, Bhawanipatna

Odisha University of Agriculture and Technology

and

Dr. Abhiram Dash

Assistant Professor (Agricultural Statistics),

Department of Agricultural Statistics,

College of Agriculture, Bhubaneswar

Odisha University of Agriculture and Technology

Bhumi Publishing

First Edition: 2021

ISBN: 978-93-91768-68-3

ISBN 978-93-91768-68-3

9 789391 768683

© Copyright reserved by the authors

Publication, Distribution and Promotion Rights reserved by Bhumi Publishing, Nigave Khalasa, Kolhapur

Despite every effort, there may still be chances for some errors and omissions to have crept in inadvertently.

No part of this publication may be reproduced in any form or by any means, electronically, mechanically, by photocopying, recording or otherwise, without the prior permission of the publishers.

The views and results expressed in various articles are those of the authors and not of editors or publisher of the book.

Published by:

Bhumi Publishing,
Nigave Khalasa, Kolhapur 416207, Maharashtra, India
Website: www.bhumipublishing.com
E-mail: bhumipublishing@gmail.com

Book Available online at:

https://www.bhumipublishing.com/books/

PREFACE

This book is a depiction of the research work done by Madhu Chhanda Kishan during her M.Sc. studies under the guidance of Dr. Abhiram Dash. The book lucidly envisages the work on forecasting technique using ARIMA Modelling. The ARIMA model has been developed for important oilseed crops of Odisha to forecast the area, yield and production. The book would be immensely helpful to the researchers in the field of Agricultural Statistics as a reference for their thesis work. This sincere endeavour would cater to the needs of the researchers and render them immense benefit.

ACKNOWLEDGEMENT

Deep sense of gratitude is extended towards the members of advisory committee in the research work of Madhu Chhanda Kishan during M.Sc. studies for their cooperation and valuable advice in carrying out the research work.

The help rendered by staff members of the Department of Agricultural Statistics, College of Agriculture, OUAT, Bhubaneswar during the conduct of research work in the Department is highly acknowledged

We thank M/s Bhumi Publishing, Maharashtra for taking up the publication of the book in neat, flawless and presentable form.

- Madhu Chhanda Kishan

Dr. Abhiram Dash

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

%	Percentage
km	kilometer
kg/ha	kilogram per hectare
ARIMA	Autoregressive Integrated Moving Average
MAPE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
ANN	Artificial Neutral Network
BIC	Bayesian information criterion
AIC	Akaike's Information Criterion
et al.	et alia (and others)
SPSS	Statistical Package For Social Sciences
R2	Coefficient Of Determination
AR	Auto regression
MA	Moving average
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedastic
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
SBC	Schwartz Bayesian Criterion
MSE	Mean squared error

ABSTRACT:

Odisha is one of the agriculture dependent states of India. The oil seeds cover 0.3% of the total cultivated area. This thesis attempted to describe the status of forecasting that is done on the important oilseed crops of Odisha (groundnut, mustard and sesamum) on their area, yield and production for the future years from 2016-17 to 2018-19 by the help of ARIMA(Auto-regressive Integrated Moving average) models. The secondary data regarding the oilseeds are collected for the years from 1975-76 to 2015-16 from various volumes of Odisha Agricultural Statistics. The data for the year from 1975-76 to 2006-07 are used for building of the ARIMA model and for the year from 2007-08 to 2015-16 are kept for cross validation of the selected ARIMA model on the basis of absolute percentage error (APE). The different ARIMA models are judged on the basis of Autocorrelation Function (ACF) and Partial autocorrelation Function (PACF) at various lags. The possible ARIMA models are identified on the basis of significant coefficient of autoregressive and moving average components. The best fitted models for different variables under study are selected on the basis of low value of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). ARIMA (2,2,0) without constant is selected as the best-fitted model for the area for groundnut having absolute percentage error ranges within 10% in most of the cases during cross-validation of the model. For the yield of groundnut (2,2,0) without constant is selected for best-fit ARIMA model having absolute percentage error ranges within 10% in most of the cases during cross-validation of the model. ARIMA (1,2,0) without constant is selected as the best-fitted model for the area for mustard having absolute percentage error less than 10% in most casesduring cross-validation of the model. For the yield of mustard (0,1,1) without constant is selected for best-fit ARIMA model having absolute percentage error ranges from within 10% during cross-validation of the model. ARIMA (1,1,0) with constant is selected as the best-fitted model for the area for sesamum having absolute percentage error ranges from within 10% in most of the cases during cross-validation of the model. For the yield of sesamum (2,1,0) without constant is selected for best-fit ARIMA model having absolute percentage error ranges from within 10% in most of the cases during cross-validation of the model. The forecasted values of area and production of groundnut are found to decrease, while forecasted values yield is found to increase over the future years. The forecasted values of area and production of mustard are found to decrease, while forecasted values yield is found to increase over the future years. The forecasted values of area, yield and production of sesamum are found to decrease over the future years. In groundnut and mustard the production forecast is in accordance with yield forecast and in sesamum the production forecast is in accordance with both yield and area forecasts.

CHAPTER 1

INTRODUCTION

Oilseed crops are grown all over the world. Due to the economical values of these crops, these are considered as the most important crops all over the world. Primarily the oilseeds are grown for edible oil. These are the major sources of lipids for human nutrition as well as industrial purposes too. These are the seeds known for large amount of oils within. Oilseeds constitute a very important group of commercial crops in India. The oil extracted from oilseeds form an important item of our diet and are used as raw materials for manufacturing large number of items like paints, varnishes, hydrogenated oil, soaps, perfumery, lubricants, etc. Oil-cake which is the residue after the oil is extracted from the oilseeds, forms an important cattle-feed and manure. The most commonly grown oilseeds in Odisha are groundnut, mustard, sesame etc.

1.1 Contribution of Agriculture in Economy of Odisha

Odisha is one of the agriculture dependent states of India. Agriculture has a grand effect on the economy of Odisha. It contributes to the state income, 15% to Gross Domestic Product (GDP) of the state. But in the country's scenario the agricultural sector accounts for 18% of the total GDP. Agriculture sector is not limited to provide only food, it's also engaged in employment generation to about 60% of the work force of the State, and collectively 70% of the population is participating in agriculture both directly and indirectly. The total cultivated area is about 61.80 lakh hectares from the total geographical area of 155.71 lakh hectares. About 34% of the geographical area contributes to the net sown area i.e. about 53.31 lakh hectares. There is a noticeable fluctuation and instability in growth rate income of the state due to several hurdles like natural calamities and rainfall fluctuation causes loss in the crop production.

1.2 Concise view of Odisha

Odisha is situated in Eastern part of India. It goes from 17.49^0 N latitude to 22.34^0 N latitude and from 81.27^0 E longitude to 87.29^0 E longitude. It has West Bengal to the north-East and in the East, Jharkhand to the north, Chhattisgarh to the West and north-West and Andhra Pradesh to the South. The total area is 1,55,707 km^2 and extends for 900 kilometres from North to South and 500 kilometres from East to West. Its coastline is 480 kilometres long. The state is divided into 30 districts which are further subdivided into 314 blocks.

1.3 Physiography of Odisha

The physiography of Odisha not only confines the limits of cultivable area but also defines the type of crops which can be grown and decides the source of irrigation and cropping pattern. The state is broadly divided into four physiographic zones.

- Coastal Plains
- Central Tableland
- Northern Plateau
- Eastern Ghats

Coastal Plains

The coastal region is defined by the deltas mainly formed by the rivers Subernarekha, Mahanadi, Bramhani and Baitarani covering 18% of the total area of the state. It takes the charge from North to South having a width in between 24 to 74 kilometres from the sea coast in the districts of Balesore. Cuttack, Puri and some part of Ganjam district too.

Central Tableland

This area runs on the chief basin of river the Mahanadi, the Tira, the Ong and the Tel and a part of catchment of Bramhani in the extreme North East part. Except in the South Western and Eastern part the table land gradually rise from East to West and the elevation is more than 305 meters. It is as large as Northen plateau and consist of undulating and folded topography. This area consist of Bolangir, Southern part of Dhenkanal, northern part of Boudh sub- Division of Phulbani, Athagarh sub-division of Cuttack, Bargarh, Sambalpur and Southern part of Rairakhol sub-division of Sambalpur districts.

Northern Plateau

Thisregion covers 23% of the area of the state consisting the most parts of watersheds of the rivers like the Bramhani, the Baitarani, the Salandi and the Budhabalanga. The districts like Mayurbhanj, Keonjhar, Sundargarh and some part of Dhenkanal comes under this plateau. Some small part of North West of Jajpur sub-division of Cuttack districts and Nilagiri sub division of Balesore and Deogarh, Kuchinda, northern part of Rairakhol sub-divisions of Sambalpur districts also comes under Northern plateau.

Eastern Ghats

This is the largest region that consists of the largest region of the state that is 36% of the total area of the state. The region consist of the hill range belonging to the main line of Eastern ghat along with some Ghats and valleys laying between the hill ranges of elevation between 305 to 475 meters. The districts of Koraput, Kalahandi, Phulbani except northern Boudh and the Western and extreme northern part of Ganjam come under Eastern Ghats. The region has the Tel River and its tributaries in the North and Vasandhara and Nagabali in the South West. South Koraput is a plateau of elevation of more than 610 meters and spread between Nabarangapur and Jaypore sub-division. Small meanderings streams intercept the hills in this area by having flood plain terraced for cultivation. Most of the cultivated areas are sloping upland.

1.4 Oilseed scenario in Odisha

From the net sown area of 54.24 lakh ha., gross cropped area is 90.54 lakh ha. from which oilseeds has the total area of 6.30 lakh ha., 0.3% of the total area of the country according to 2015-16 data. During 2015-16 the production and yield of oilseeds was 5.64 lakh MT and 895 kg/ha in Odisha. Groundnut is the most grown oilseed crop in Odisha and stands first the list in groundnut yield rate of 1750 kg/ha and covers the area of 2.21 lakh ha and production of 3.87 lakh tones in 2015-16 compared the yield rate for the country was 1750 kg/ha. The contribution of mustard with respect to the total area, yield and production is 1.45 lakh ha, 424 kg/ha and 6.16 lakh tonnes respectively. Odisha has the total oilseed yield rate of 928 kg/ha which is less than the total oilseed yield of India i.e. 1153 kg/ha according to the records of 2015-16. So there is enough scope of oilseed production in the state.

In table 1.1 the last 5 years (2011-2016) data of the area and production of important oilseed crops are taken. It shows groundnut, mustard and sesamum has significant contribution in the oilseed statistics of Odisha.Groundnut, mustard and sesamum contributes about 34.8%, 16.75%, 30.79% of the total area and 67.69%, 7.85%, 13.71% of the total production of the oilseed crops of Odisha, respectively. These three crops contribute about 82.34% and 88.98% of total area and production of the total oilseed crops of Odisha, respectively.

1.5 Climate and soil of Odisha

Eight groups of different soils are seen in Odisha: 1.Red loamy soil and red sandy soils, 2. Laterite soils, 3. Red and yellow soils, 4. Coastal alluvial soils including the saline soils, 5. Deltaic alluvial soils, 6. Black soils, 7. Mixed red and black soils, 8. Brown forest soils. The state goes through four meteorological seasons: winter (January to February), pre-monsoon (March to May), South-West monsoon (June to September) and North East monsoon (October–December).

Table 1.1: Area and production of important oilseed crops of Odisha

Important Oilseed crops	Area ('000 ha)	Production ('000 tonnes)
Groundnut	253.35	446.31
Mustard	122.02	51.42
Sesamum	224.18	90.44
Total	728.06	659.28

** The data is based on last 5 years record of 2011-2016

(*Source- Odisha Agricultural statistics)*

1.6 Topography of Odisha

The nature and extent of Agriculture and land use in the state is highly influenced by the physical feature and the micro-topography of the state. The land can be classified into three categories -Low (25.6%), Medium (33.6%) and Upland (40.8%) with various types of soils. The state is divided into 10 agro climatic zone such as North Western Plateau, North Central Plateau, North Eastern coastal plain, East and South Eastern coastal plain, North Eastern Ghats, Eastern Ghats high land, South Eastern Ghats, Western undulating zone, mid central table land, Western central table land.

1.7 Importance of forecasting technique in agriculture

Forecasting in agriculture sector comprises of forecasting of area/yield/production of the crop, also the forewarning of incidence of pest and disease to the crops. For formulation of policies regarding stock, supply and distribution agricultural produce throughout the country the crop yield forecasting is essential. Various methods of forecasting based on the models that use crop biometrical characters, weather parameters, farmers' eye estimates, agro-metrological conditions and remotely sensed crop reflectance observations are used both in separate and integrated approach. Also forecasting based on time series data are of prime importance in policy forming decision regarding various agricultural activities.

1.8 Statistical forecast model in agriculture

The statistical forecast model should be able to provide objective crop forecast with reasonable precisions well in advance before harvest for taking timely decisions. The different statistical forecasting models are

a) Regression Models:

Multiple Linear Regression (MLR) uses plant characters for crop yield forecasting. Weather indices based on MLR model for crop yield and crop pest count forecasting. Logistic regression models for forecasting of qualitative response variable like low or high crop yields, presence absence of crop pests/ disease etc.

b) Probabilistic Models:

Markov chain model for forecasting of crop yields.

c) Time series Models:

Exponential smoothing models for forecasting area and production of crop. Auto-regressive Integrated moving average (ARIMA) Model for forecasting area and production of crops.

1.9 Objective of the study

It is well known that production of a crop production depends on area and yield of crops. So in this study the forecasting of the important oilseeds crops of Odisha such as groundnut, mustard and sesame are obtained for area and yield, from the forecast of the area and yield the forecast of production is obtained.

Thus keeping in view these points, the study is made under following objectives

1) To select and fit possible ARIMA models to data on area and yield of important oilseed crops of Odisha for the year 1975-76 to 2006-07.
2) To select the best fit ARIMA models on basis of significance of coefficients, model diagnosis test and model selection criteria.
3) To cross-validate the selected best fit model by using the available data for the year 2007-08 to 2015-16.
4) To use the best fit model for forecasting the area and yield of important oilseed crops of Odisha for the years 2016-17, 2017-18, 2018-19 after successful cross validation of the selected model.
5) To forecast the production of the important oilseed crops by using the forecast of area and yield.

CHAPTER 2

REVIEW OF LITERATURES

The following literature we have studied which gives an idea of achieving the objectives of study.

Suresh *et al.* (2011) studied on forecasting the sugarcane area, production and productivity of Tamilnadu through fitting of univariate Auto regressive Integrated Moving Average (ARIMA) models. The data on sugarcane area, production and productivity collected from 1950-2007 has been used for present study. ARIMA (1,1,1) model is found suitable for sugarcane area and productivity. ARIMA (2,1,2) is found appropriate for modelling sugarcane production. The performances of models are validated by comparing with actual values. Using the models developed, forecast values for sugarcane area, production and productivity.

Debnath (2013) studied on forecasting of cultivated area and production of cotton in India using Auto regressive Integrated Moving Average (ARIMA) model. Time series data covering the period of 1950-2010 was used for the study. The study revealed that ARIMA (0,1,0), ARIMA (1,1,4) and ARIMA (0,1,1) are the best fitted model for forecasting of cotton area, production and yield in India respectively. The analysis shows that if the present growth rates continue then the cotton area, production and yield in the year 2020 will be 10.92 million hectares, 39.19 million bales of 170 kg of each and 527 kg/ha respectively.

Singh *et al.* (2013) studied paddy area and production in Bastar division of Chhatisgarh data for the period of 1974-75 to2010-2011 were analysed by times 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 (2,1,2) and ARIMA (2,1,0) model were used to forecast paddy area and production in Bastar division of Chhatisgarh for four leading years . The results also shows paddy area forecast for the year 2015 to be about 598.22 thousand hectare with upper and lower limit 665.39 and 531.06 thousand hectares respectively. The model also shows paddy production forecast for the year 2015 to be about 1126.61 thousand tones with upper and lower limit 1430.16 and 823.05 thousand tonnes respectively.

Biswas *et al.* (2014) has studied area, production and productivity data of wheat for Punjab was analysed by time-series method. Yearly wheat yield data for the period of 1950-51 to 2009-10 was used as input to forecast the yield up to the year 2020-21. The Box-Jenkins ARIMA method was put into use to forecast the yield. The validity of the model was tested by standard statistical techniques. The past 60 years data revealed that wheat yield was increased from 0.8t ha-1 in 1950-51 to 4.3t ha-1 in 2009-10. The model projected 15.3 per cent increase in wheat production in years to come by 2020-21 in Punjab. Based on ARIMA output, wheat production of Punjab is likely to increase from 15844.7 thousand tons in 2010-11 to 18271.7 thousand tons in 2020-21.

Sahu *et al.* (2014) have studied under the background of overall food security situation in the SAARC countries, attempts have been made to analyse the production behaviour along with the total seeds of two major food crops rice and wheat. This helped to draw up strategies and programmes for regional cooperation in ensuring food security and reducing hunger and malnutrition in the region also helped for providing seed security in these SAARC countries. In addition to descriptive statistics, ARIMA modelling technique has been used to analyse the information from1961 through 2010. The forecast shows that rice and wheat production for the year 2020 would be about 794 and 777 million tons respectively in the world. In spite of increase in production the study reveals that the yield of rice and wheat in the world would be 4.35 t/ha and 3.4 t/ha in 2020 but the yield of these two crops in SAARC countries, barring one countries in each, will remain far below the world projection. Thus, under the given remote possibility of horizontal expansion, the study emphasizes the need for quantum jump in the per ha yield of these 2 crops for this region. The study also advocates that good quality of seeds in good amount be made available to farmer, otherwise the whole food security of this part if the Globe would be under tremendous risk.

Ilic *et al.* (2016) studied the production of corn in the period from 1947 to 2014 Serbia had an oscillatory trend with significant jumps and falls in production. Temperature fluctuations and changes in the volume of precipitation are the main factors affecting the growth and development of crops and ultimately, the quantity produced. The subject is based upon the forecasting trends in corn production in Serbia. Building on the subject, the purpose of this paper is to create the model for forecasting future corn production and establishing its trends.

Frah (2016) studied on Sudan Production of Sorghum; forecasting 2016-2030 using Auto Regressive Integrated Moving Average ARIMA Model using Box- Jenkins methodology in time series analysis which is the optimal method applied to the pattern. This method consist of four steps namely identification, estimation, diagnostic checking and forecasting by ARIMA models. Future forecasts drawn there show that the sorghum production will be likely to increase in coming year.

Badmus *et al.* (2011) studied on forecasting the cultivated area and production of maize in Nigeria using ARIMA model. Time Series data covering the period of 1970-2005 was used for the study. The data were obtained from the CBN, IFS reports and NBS. The result also shows maize production forecast for the year 2020 to be about 9952.72 tons with upper and lower limits 6479.8 and 13424.64 thousand tons respectively. The model also shows that the maize area would be 9229.74 thousand hectares with lower and upper limit of 7087.67 and 11371.81 thousand hectares respectively by 2020. This projection is important as it helps inform good policies with respect to relative production, price structure as well as consumption of maize in the country. The conclusion from the study is that, total cropped area can be increased in future, if land reclamation and conservation measures are adopted.

Wali *et al.* (2017) studied upon the forecasting of the area and production of cotton in India using the univariate ARIMA model. The time series data on area and production of cotton in India for period of 65 years from 1950-51 to 2015-16 was analysed for the study. The best models were selected by comparing AIC, SBC, normalized BIC, MAPE and maximum values of R^2. The study revealed that ARIMA(0,1,0) and ARIMA (1,1,1) were the best fitted models for forecasting area and production of cotton in India respectively. Selected models were used to forecast area and production of cotton for 4 years from 2017-18 to 2020-21. The analysis showed an increasing trend in area and production of cotton.

Panasa *et al.* (2017) examined the monthly modal prices of maize using Auto Regressive Integrated Moving Average (ARIMA) models so as to determine the most efficient and adequate model for analysing the maize monthly modal prices in Telengana. The result indicates that ARIMA (2,1,1) model is the most adequate and efficient model. Various model selection criterion and diagnostic tests like AIC, BIC and MAPE are used. The forecasted result shows that there are exceptions of increasing maize prices in Bedapalli market next five months (October to February).

Kumar *et al.* (2017) studied on forecasting of sugarcane productivity of Bihar through fitting of well-known Box Jenkins univariate Auto Regressive Integrated Moving Average (ARIMA) model. Time series data on sugarcane productivity in Bihar from

1939-40 to 2014-15 were taken SRI Pusa, Bihar and Indian sugarcane for study. The data on sugarcane productivity in Bihar from the year 1940 to 2010 were utilized to build an ARIMA model and validated through five-year productivity data from 2011-15 Akaike Information Criterion (AIC) was selected for best model selection criteria. ARIMA (0,1,1) model found best suitable model for sugarcane productivity in Bihar based on AIC model selection criteria. The performances of models are validated by comparing with actual values of sugarcane productivity in Bihar data. Using developed ARIMA (0,1,1) model, two years ahead, year 2016 and 2017 sugarcane productivity in Bihar forecasted showing increasing productivity with 4.22% and 5.05% prediction standard error.

Solanki *et al* (2017) used Autoregressive Integrated Moving Average (ARIMA) and Exponential GARCH (EGARCH) model for their estimation procedures for modelling and forecasting of mustard price. For forecasting mustard price ARIMA(0,1,1) model is used which gives reasonable and acceptable forecasts but the study has revealed that the AR (1)-EGARCH (1,1) model outperformed the price forecasting models for mustard prices primarily due to its ability to capture asymmetric volatility pattern.

Poyyamozhi *et al.* (2017) studied on forecasting the cultivated area and production of cotton in India using Autoregressive Integrated Moving Average (ARIMA) model covering the period of 1955-2015 for this study. The study revealed that ARIMA (0,1,0) are the best fitted model for forecasting of cotton production in India respectively. The analysis shows that if present growth rates continue then the cotton production in the year 2025 will be 42.53 million bales of 170 kg of each respectively.

Cenas (2017) studied on analysis of the possibility of improving the accuracy and precision of typical time series model in forecasting future prices of rice crop by combining the techniques of ARIMA and Kalman filter respectively. Using actual rice data collected over a period of five years, the performance of the typical ARIMA model was compared to the combined performance of ARIMA-Kalman filter using Mean Square Error (MSE) and Root Mean Square Error (RMSE) as the bases of comparison. Results of the analysis revealed that a more accurate and precise time series estimates of future price of rice can be achieved when the technique of Kalman filter is combined with typical ARIMA time series model. Further analysis showed that predicted values generated by the combined techniques from out-of-sample forecasts are fairly closer to the actual values. On the basis of the findings of this study, the development of time series software that will work on combining the algorithms of the ARIMA and Kalman filter is recommended.

Hemavathi *et al.* (2017) studies on food grains area and production in India data for the period of 1950-51 to 2014-15 were analysed 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 (1,1,0) and ARIMA (0,1,1) model were used to forecast area and production in India for four leading years. The results also shows area forecast for the year 2019 to be about 124.78million hectare with upper and lower limit 134.27 and 115.29, million hectares respectively. The model also shows food grains production forecast for the year 2019 to be about 271.09 million tonnes with upper and lower limit 294.73 and 247.44 million tonnes respectively.

Darekar *et al.* (2017) presented presents a methodology to forecast prices during harvest period and applied the method to forecast for the kharif 2017-18. The time series data on monthly average prices of paddy from January, 2006to December, 2016 collected from AGMARK was used. ARIMA (Box-Jenkins) model was employed to predict the future prices of paddy. Model parameters were estimated using the R programming software. The performance of fitted model was examined by computing various measures of goodness of fit viz., AIC, BIC and MAPE. The ARIMA model was the most representative model for the price forecast of paddy in overall India. In kharif season the paddy is harvested during September-November. The forecast shows that market prices of paddy, would be ruling the range of Rs. 1600-2200 per quintal in kharif harvesting season 2017-18.

Dash *et al.* (2017) developed appropriate ARIMA models for the time series data on production of food grains in Odisha. The selected model was cross-validated with the available data for the period kept for validation and nit used for model building. The finally selected model after successful cross-validation on basis of MAPE is used for forecasting of production of food grains in Odisha.

Chukwujioke *et al.* (2018) studied the trend analysis of area, yield and production for Cashew in Nigeria. The finding of the study is based on data from the years (1961 to 2016) and was taken from the database of FAO (2018). Three models of trend analysis were applied. The models were Linear Trend Model, Quadratic Trend Model and Cubic Trend Model. The most appropriate Model for trend analysis of the present study was Cubic Trend Model based on the highest R^2 of (95.76%), (95.76%) and (88.12%) for cashew area harvested, production and yield respectively, coupled with the lowest residual sum square and mean square error. Forecasting of the data was done up to 2026. The study represents an insight to national policy makers regarding this essential crop and provides them with a reference range of values in area harvested, yield and production in future so that they may be able to effectively deal with cashew production in Nigeria.

Vijay *et al.* (2018) have studied time series prediction is a vital problem in many applications in nature sciences, agriculture, engineering and economics. The objective of the study is to examine the flexibility of artificial neural network model (ANN) is time series forecasting by comparing with classical time series ARIMA model. The data consist of area and production of pearl millet crop area (''000 ha) and production ("000 MT) from 1955-56 to 2014-15 were collected from "Agricultural Statistics at a Glance 2014-15", Karnataka, India were used in the study to demonstrate the effectiveness of the model. The experiment shows that ANN model outperform the ARIMA models based on root mean (RMSE), MAPE and MSE.

Nath *et al*. (2018) studied on time series modelling to approach forecasting wheat production for India. ARIMA (1,1,0) model was found to the best ARIMA model for the present study. The efforts were made to forecast, the future wheat production for a period up to ten years as accurate as possible, by fitting ARIMA (1,1,0) model to our time series data. The forecast results have shown that the annual wheat production will grow in 2026-27. The wheat production will continuously grow with an average growth rate of approximately 4% year by year.

Sharma (2018) studied on forecasting the future figures and autoregressive integrated moving average (ARIMA) is one of them. Maize is an important cereal of India, keeping in view its importance for rain fed areas of the country and diverse uses. The present study was conducted to forecast maize production for the year 2018 to 2022 based on the estimation of suitable ARIMA model. The analysis of ACF and PACF of differenced series revealed that ARIMA (2,1,0) was the most suitable model for forecasting based on the diagnostics , such as ACF,PACF,AIC,SBC etc. The selected ARIMA model predicted an increase of 13.76 percent increase in maize production in next 5 years 2017to 2022.

CHAPTER 3

MATERIAL AND METHODS

This chapter consists of use of methodology, study period, data sources and the description of the techniques used in the study.

3.1 Period of study

The study period consists of 41 years of data from the year 1975-76 to 2015-16. The data collected for the above time period comprises the area and yield of important oilseed crops such as groundnut, mustard and sesame.

3.2 Sources of data

The database is based on the secondary source data pertaining to the area and yield of important oilseed crops of Odisha like groundnut, mustard and sesamum for the period from 1975-76 to 2015-16. The database is collected from Odisha Agricultural Statistics published by the Directorate Agriculture and food production, Government of Odisha. The area, yield and production are expressed in ’000 ha, kg/ha and ’000 MT respectively (1 ha=10000 m^2, 1 MT = 1000 kg).

3.3 Methodology

ARIMA is a statistical analysis model that uses the time series data to forecast the future trends. It retains a form of regression analysis seeking to predict future movements and the random walks seemingly taken by stocks and the financial market by examining the differences between values in the series instead of using the actual data values. The differenced series have lags referred as “auto-regressive” and forecasted data lags are referred as “moving average”. This model is represented as ARIMA (p,d,q), where p represents order of auto-regression, d shows degree of differencing, q shows the order of moving average.

The Auto Regressive Integrated Moving Average (ARIMA) methodology is also known as Box-Jenkin’s methodology. The Box-Jenkins procedure focuses on fitting a mixed ARIMA model to a given data set. The main objectives of ARIMA model fitting revolves around identifying the stochastic process of the time series and predicting the future values accurately. These methods can also be used in different types of situations that involve the building of models for discrete time series and dynamic systems. This system lags behind in lead times or seasonal series with a large random component. Originally George Box and Gwilym Jenkins have studied ARIMA model extensively during 1968 and their names has been used frequently with general ARIMA process applied for time series analysis, validation, forecasting and control.

However the stochastic model of the series is used to determine the optimal forecast of future values of a time series. A stochastic process can either be stationary or non-stationary. The ARIMA models refer only to a stationary time series but most of the time series are non-stationary. The first stage of Box-Jenkins model is for reducing non-stationary series to a stationary series by taking first or second order differences.

The Box-Jenkins's method is named after the statisticians George Box and Gwilym Jenkins, applies autoregressive moving average ARMA model to stationarised data i.e. ARIMA models to find the best fit of a time-series model to past values of a time series in time series analysis.

3.4 Autoregressive (AR) process

It is similar to the simple linear regression model in which Z_t is the dependent variable and Y_{t-1} is the explanatory variables.

$$\mathbf{Z_t = C + \phi_1 Y_{t-1} + a_t}$$

Where Z_t=time sequenced random variable

C=constant term related to mean (μ) such that $C=\mu(1-\phi_1)$

ϕ_1= relationship of Y_t with Y_{t-1}

a_t = a random shock element at time t

Autoregressive (AR) process is the process that involves past or time-lagged Y terms and the longest time lag associated with Y terms is called the AR order in the process. Thus the above equation is an AR process of order one or AR (1).

3.5 Moving Average (MA) process

MA models can be introduced in several ways. One way is to treat the model as a simple extension of white noise series and the other way is to treat the model as an infinite order AR model with some parameter constraints. Mostly the second approach is adopted. The general equations for moving average model was given as

$$\mathbf{Z_t = C - \theta_1 e_{t-1} + e_t}$$

Where C= constant term related to mean μ

θ = relation of a_t withe_{t-1}

Moving average (MA) is the process with past (time-lagged) random shocks and order of MA is the longest time lag associated with an error term (i.e. a_t). The above equation is an MA process of order one, written as MA (1). For a pure model C=μ. The negative sign attached to$\theta_1$is merely a convention. Whether to use negative or positive sign, it makes no difference.

The formula for calculating autocorrelation is

$$\mathbf{r_t}= \frac{\sum_{t=1}^{n-k}(Z_t-\overline{Z})(Z_{t-k}-\overline{Z})}{\sum_{t=1}^{n}(Z_t-\overline{Z})^2}$$

As long as the data series is stationary the following set of recursive equation gives fairly good estimates of the partial autocorrelation

$\hat{\varphi}_{11}$ = r_1

$$\hat{\varphi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1}\hat{\varphi}_{k-1,j}r_{k-1}}{1-\sum_{j=1}^{k-1}\hat{\varphi}_{k-1,j}r_j}$$

Where k=1,2,3….. and

$\hat{\varphi}_{kj} = \hat{\varphi}_{k-1,j}$-$\hat{\varphi}_{kk}$ $\hat{\varphi}_{k-1,k-j}$ (k=3,4,5……;j=1,2,3…,k-1)

$\hat{\varphi}_{kk}$ = the estimate of the true partial auto correction on coefficient φ_{kk}

r_k= auto correlation coefficient k lag part

$\hat{\varphi}_{kj}$ = estimate of partial autocorrelation coefficient for lag apart when the effect of the intervening lags has been removed.

3.6 fitting of the Box-Jenkins ARIMA model

The autoregressive moving average (ARMA) models was introduced to overcome the difficulty in describing the dynamic structure of the data by fitting Autoregressive (AR) and Moving average (MA) models. Autoregressive integrated moving average (ARIMA) models are the ARMA models that includes the order of differencing (which is done to stationaries the data). The ARIMA model with parameter (p,d,q) is fitted by univariate Box-Jenkin's techniques(Box and Jenkins,1976). This model includes Autoregressive of order *p,* differencing to make stationary series of degree *d* and moving average of order *q*.

3.7 Test for stationarity

If the time series data has constant mean and variance over time then it is stationary. After the original data is plotted, it is verified for stationary, if the data appears to be non-stationary, then the first difference of the data is plotted and checked for stationary. This process is repeated till the data becomes stationary. The maximum order of differencing (d) is usually 2.

3.8 To determine the order of AR (p) and MA (q)

The values of p (order of Auto regression) and q (order of moving average) can be determined by examining the plots of the Autocorrelation and partial autocorrelation of the stationeries value of the variables. The autocorrelation of y at lag k is the correlation between y and itself lagged by k periods, i.e., it is the correlation between y_t and y_{t-k}. The partial autocorrelations of y at lag k is the coefficient of y_LAGk in a regression of y on y_LAG1, y_LAG2, ….., up to y_LAGk. Thus, the partial autocorrelation of y at lag 1 is the same as the autocorrelations of y at lag1.The partial autocorrelation of y at lag 2 is the coefficient of y_LAG2 in a regression of y on y_LAG1 and y_LAG2, and so on. The way to interpret the partial autocorrelation at lag k is that it is the amount of correlation between y and y_LAGk that is not explained by lower-order autocorrelations.

The rules for determining the p and q from the plots of ACF and PACF:

i. If the ACF plots "cuts off sharply" at lag k (i.e., if the autocorrelation is significantly different from zero at lag k and extremely low in significance at the next higher lag and the ones that follow), while there is a more gradual "decay" in the PACF plot (i.e., if the drop off in significance beyond lag k is more gradual), then set q=k and p=0. This is a so-called "MA (q) signature."

ii. On the other hand, if the PACF plot cuts off sharply at lag k while there is a more gradual decay in the ACF plot, then set p = k and q = 0. This is so-called "AR (p) signature."

iii. If there is a single spike at lag 1 in both the ACF and PACF plots, then set p =1 and q=0. If it is positive (this is an AR (1) signature), and set p = 0 and q=1 if it is negative (this is a MA (1) signature).

All the correct form of the model is identified as the highest order AR or MA coefficient should be significant. If the ACF and PACF plot look good but the highest order coefficient is not significant then we should reduce the value of p and q by 1, as the case may be. If there are some significant residual autocorrelations or partial autocorrelations at the first few lag. The rules to be follows are-

i. If there is a “spike” at a low-order lag in the residual ACF plot, then we should increase *q* by 1 and re-fit the model.
ii. Conversely, if there is a spike at a low-order lag in the PACF plot, we should increase *p* by 1 and re-fit the model.

By using Box-Ljung test the adequacy of the selected model is checked. A formal test of the fitness of the model is also done by using Box-Ljung test of the residuals (Ljung and Box (1978)) is done in following manner:

Null hypothesis: H_0: The errors are distributed randomly.

Alternate hypothesis: H_1: The errors are non-random.

The Box-Ljung test statistic, $Q = n(n+2)\sum_{k=1}^{m} \frac{r_k^2}{n-k}$

Where n is the number of observations,

r_k is the estimated autocorrelation of the series at lag k=1,2,….,m

m= number of lags being considered.

Here, the null hypothesis is rejected i.e., the errors are not independent if $\boldsymbol{Q} \geq \chi^2_{1-\alpha,h}$

The null hypothesis is accepted i.e., the errors are independent if $Q < \chi^2_{1-\alpha,h}$

Where, $\chi^2_{1-\alpha,h}$ is the chi-square distribution table value with ‘h’ degrees of freedom and level of significance α such that $P(\chi^2_h > \chi^2_{1-\alpha,h}) = 1-\alpha$

Here, p = number of AR

Q= number of MA

The degree of freedom, h =(m-p-q)

By the help of forecasting tool of Spss 20.0 the Box-Ljung test is done.

The model fit statistics used to select best fit model are:

1. Root Mean Square Error (RMSE)

RMSE = $\frac{\sum_t e_t^2}{n-2}$

2. Mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{\dfrac{\sum_t \left|Y_t - \hat{Y}_t\right|}{yt}}{n} * 100$$

The model which has lowest value of RMSE and MAPE among the model selected ARIMA is considered to be the best-fit model from the given data set. (Dash *et al.,* 2017)

3.9 Cross-validation of the selected model

The cross-validation of the selected model is worked on the 20% of the data that is held up and not used for model building at the end period. For the cross-validation of the model the actual value of the left out period and the forecasted value of the left out period of the selected model are used. Here the data from 1975-76 to 2006-07 are used for model building and data from 2007-08 to 2015-16 are used for cross-validation.
The absolute percentage error is calculated.

$$\text{\% of forecasting error} = \left(\frac{Y - \hat{Y}}{Y}\right) * 100$$

Where, Y= observed value of remaining 8 years

$\hat{Y}$ = The forecasted value of remaining 8 years

The selected best-fit ARIMA model is used for forecasting after the cross-validation. ARIMA techniques are generally used in case of short term forecasting because the prediction for longer periods will have more errors associated with it. So, ARIMA should be used for short term forecasting (Sarika *et al*. 2011).

CHAPTER 4

RESULTS AND DISCUSSION

Data is fitted to the data by the help of ARIMA model on area and yield of groundnut, mustard and sesamum for forecasting. Basically the data used for model building is from the year 1975-76 to 2006-07. The data from 2007-08 to 2015-16 is used for the cross-validation of the selected model. By using the best fit model the forecasting is done for the years 2016-17, 2017-18 and 2018-19.

4.1 Forecasting of area, yield and production of groundnut by fitting appropriate ARIMA model

In figure. 4.1(a), the original plot of data on area under groundnut shows the data is non-stationary that implies it doesn't have constant mean and variance. Thus, the second difference of the data is plotted as the first difference in figure. 4.1(b) was not stationary and shown in figure.4.1 (b) shows that the first difference of data has constant mean and variance, which makes the data stationary.

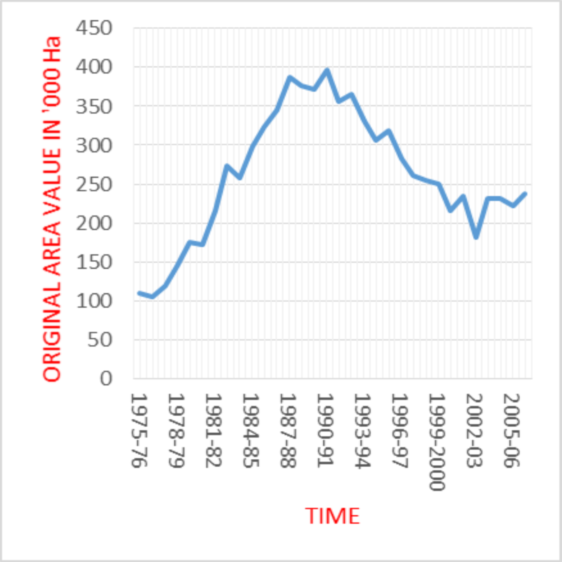
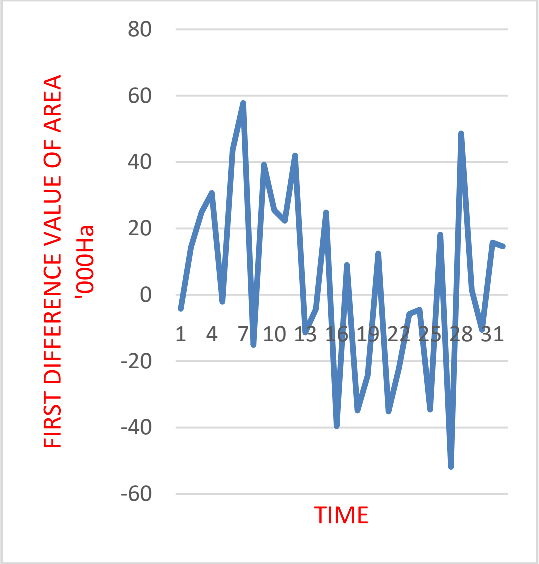


Figure 4.1(a): Plot of original values of area value under groundnut *vs* time



4.1(b): Plot of first difference of area under groundnut *vs* time

The ACF and PACF plot of the second difference values of groundnut area is shown in Figure. 4.3, gives the tentative value of q and p suitable for the area of groundnut is q=0 and p=2. Thus the ARIMA model found to be best fitted for area of groundnut is ARIMA (2,2,0).

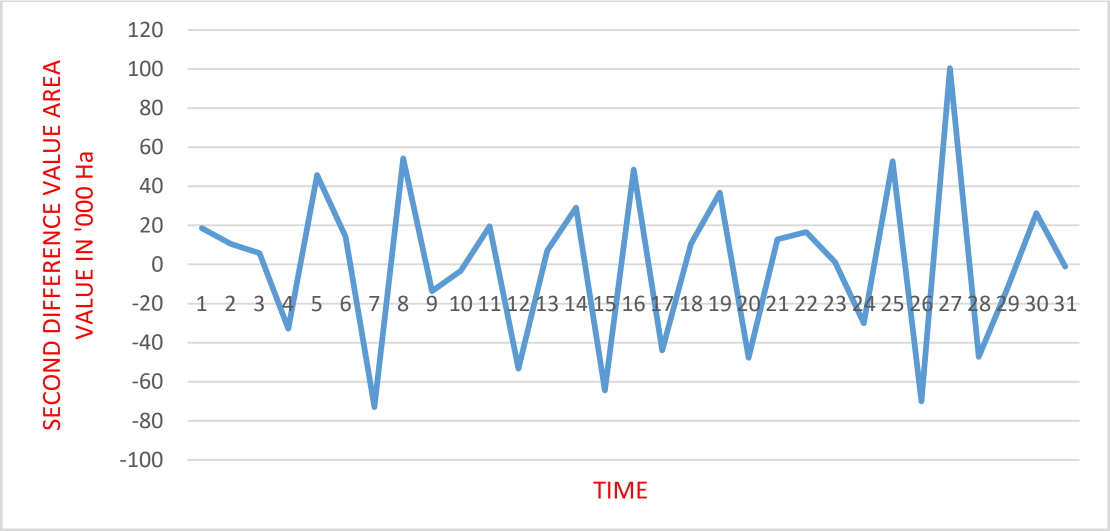


Figure 4.1(c): Plot of second difference values of area under groundnut *vs* time

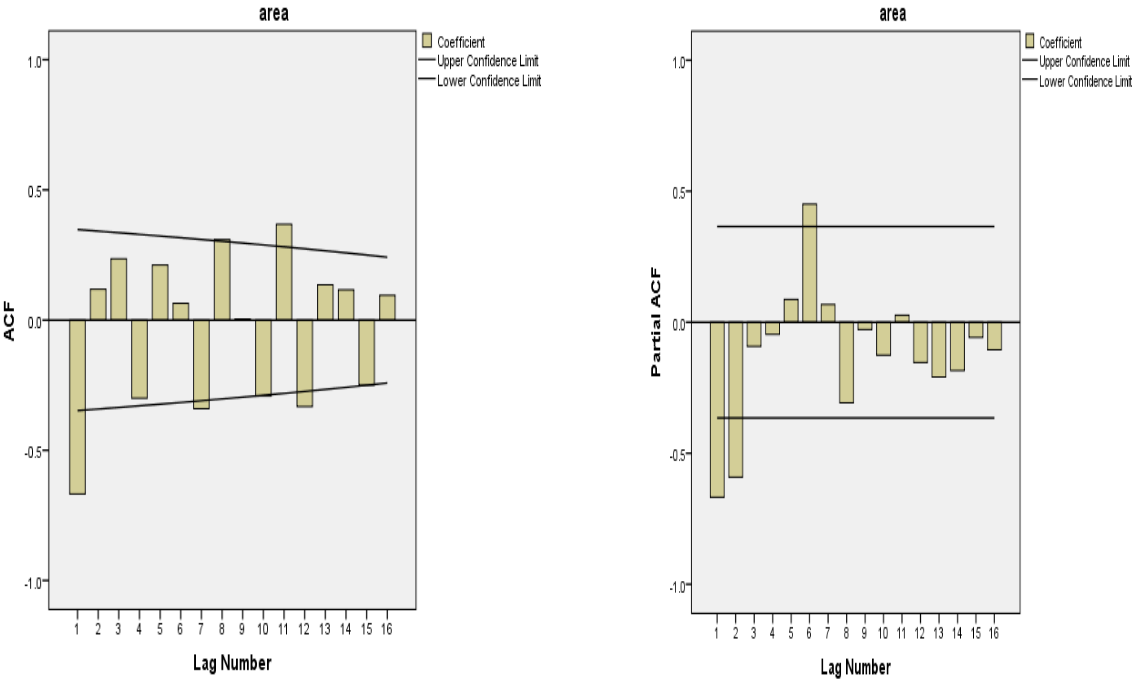


Figure 4.2: ACF and PACF plot of second difference values of area under groundnut

Figure 4.3(a) is the plotted data on original yield under groundnut which is non-stationary that says it doesn't have constant mean and variance. So, the second difference of the data is plotted as the first difference figure. 4.3(b) was not stationary and shown in figure 4.3(c) showing the second difference of data having constant mean and variance making the data stationary.

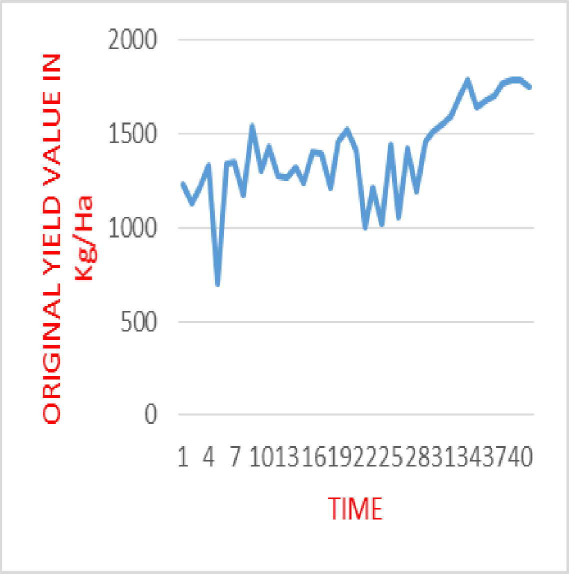
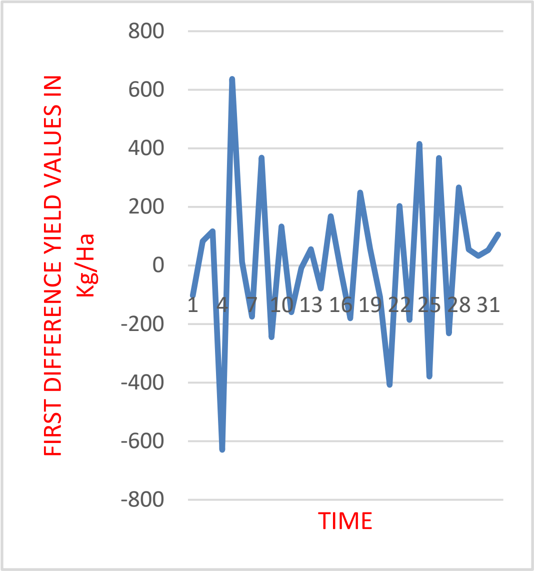


Figure 4.3(a): Plot of original values of yield values of groundnut *vs* time



4.3(b): Plot of first difference yield of groundnut *vs* time

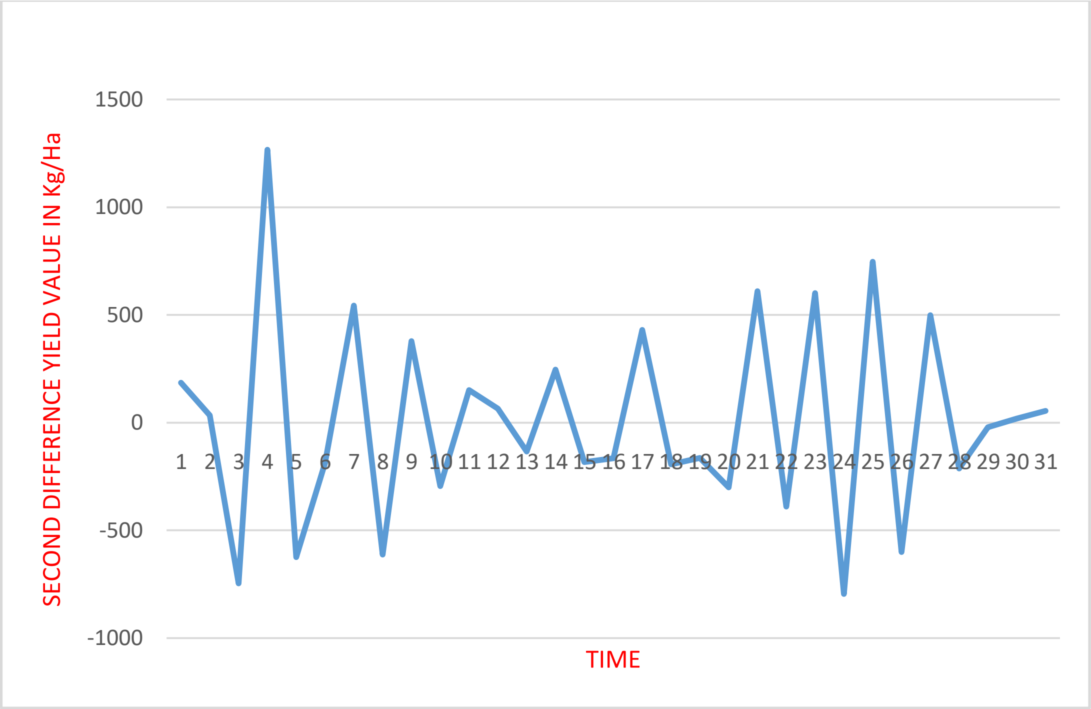


Figure 4.3(c): Plot of second difference values of yield of groundnut *vs* time

The ACF and PACF plot of second difference values of groundnut yield is shown in figure 4.4 suggesting that the tentative value of q and p that would be suitable for yield of groundnut is q= 0 and p=2, so the ARIMA model that is found to be best fitted for yield of groundnut is ARIMA (2,2,0).

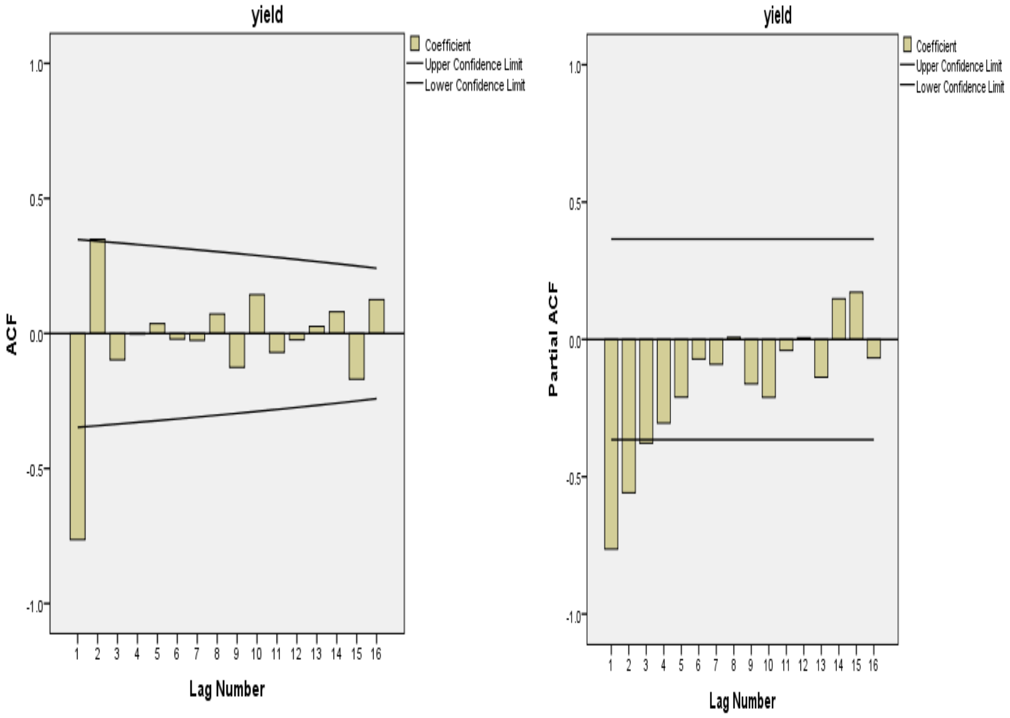


Figure 4.4: ACF and PACF plot of second difference values of yield of groundnut

The study of table 4.1 shows that when ARIMA (2,2,0) is fitted to the data on area under groundnut, the constant is not significant. So, ARIMA (2,2,0) without constant is also fitted. The estimated AR (2) is also found to be significant. Thus, the selected ARIMA model for area under groundnut is ARIMA (2,2,0) without constant. In case of yield when ARIMA (2,2,0) is fitted to the data on yield under groundnut, the constant is not significant. So, ARIMA (2,2,0) without constant is also fitted. The estimated AR (2) is also found to be significant. Thus, the selected ARIMA model for yield under groundnut is ARIMA (2,2,0) without constant.

The study of table 4.2 shows that all the fitted model satisfy the assumption of normality of error as they all have non-significant S-W Statistic and also all the models are found to be adequate due to non-significant Ljung-Box Q statistic. For area under groundnut, the ARIMA (2,2,0) without constant has low value of RMSE and MAPE. So the best-fit model is ARIMA (2,2,0) without constant. In case of yield of groundnut, the ARIMA (2,2,0) without constant has low value of RMSE and MAPE that makes it the best fit model.

Table 4.1: Coefficient of AR and MA components of the fitted ARIMA model considered for forecasting area and yield of groundnut in Odisha

	Best fit ARIMA model	**Constant (μ)**	**Coefficient of autoregressive components**		**Coefficient of moving average components**	
			α_1	α_2	θ_1	θ_2
Area	ARIMA (2,2,0)	0.000 (0.015)	-1.002** (0.146)	-0.546** (0.143)	_	_
	ARIMA (2,2,0) (without constant)	_	-1.002** (0.144)	-0.546** (0.141)	_	_
Yield	ARIMA (2,2,0)	0.000 (0.290)	-1.158** (0.138)	-0.555** (0.138)	_	_
	ARIMA (2,2,0) (without constant)	_	-1.158** (0.136)	-0.555** (0.136)	_	_

(Figures in the parentheses indicate the standard error)

* Significant at 5% level of significance, ** Significant at 1% level of significance.

Table 4.2: Model fit Statistics and Residual Diagnostics of the ARIMA models fitted for area and yield of groundnut in Odisha

	Model	**Model fit Statistics**		**Residual diagnostics**	
		RMSE	**MAPE**	**Ljung-Box Q Statistics**	**Shapiro-Wilk's Statistics**
Area	ARIMA (2,2,0)	25.254	8.064	21.180	0.929
	ARIMA (2,2,0) (without constant)	**24.906**	**8.064**	**21.178**	**0.925**
Yield	ARIMA (2,2,0)	254.095	14.532	26.996	0.912
	ARIMA (2,2,0) (without constant)	**250.265**	**14.516**	**26.992**	**0.917**

(Models highlighted as bold are the best fit models)

The cross validation of the selected best fit ARIMA (2,2,0) without constant model for area under groundnut presented on the table4.3 shows that the absolute percentage error are quite low, thus the selected model is successfully cross validated.

Table 4.3: Cross validation of the selected best fit ARIMA (2,2,0) without constant model for area of groundnut in Odisha

Year	Actual value (in '000 ha) (Y)	Forecasted value (in '000 ha) $(\hat{Y})$	Error $(Y-\hat{Y})$	Absolute Percentage Error $\left(\frac{\lvert Y-\hat{Y}\rvert}{Y}\right)$ X 100
2007-08	251.5	230.52	-20.98	8.34194831
2008-09	256.05	250.46	-5.59	2.18316735
2009-10	243.37	270.4	27.03	11.10654559
2010-11	247.69	250.9	3.21	1.295974807
2011-12	255.14	241.75	-13.39	5.248099083
2012-13	263.41	247.32	-16.09	6.108348202
2013-14	267.68	266.4	-1.28	0.478182905
2014-15	259.23	272.67	13.44	5.184585117
2015-16	221.29	262.4	41.11	18.57743233

The cross validation of the selected best fit ARIMA (2,2,0) without constant model for yield under groundnut presented on the table 4.4 shows that the absolute percentage error are quite low, thus the selected model is successfully cross validated.

Table 4.4: Cross validation of the selected best fit ARIMA (2,2,0) without constant model for yield of groundnut in Odisha

Year	Actual value (in kg/ha) (Y)	Forecasted Value (in kg/ha) $(\hat{Y})$	Error $(Y-\hat{Y})$	Absolute Percentage Error $\left(\frac{\lvert Y-\hat{Y}\rvert}{Y}\right)$ X 100
2007-08	1705.33	1675.82	29.51	1.730456862
2008-09	1790.86	1775.54	15.32	0.855454921
2009-10	1639.15	1916.84	-277.69	16.94109752
2010-11	1679.96	1807.46	-127.5	7.589466416
2011-12	1706.59	1660.58	46.01	2.696019548
2012-13	1774.04	1675.56	98.48	5.551171338
2013-14	1786.95	1837.88	-50.93	2.850107725
2014-15	1787.25	1877.11	-89.86	5.027836061
2015-16	1749.69	1867.71	-118.02	6.745194863

In table 4.5, the forecasted values for area and yield of groundnut are obtained from the respective best fit ARIMA model. It shows that there is a decrease in the forecasted values of area and increase in forecasted value of yield from 2016-17 to 2018-19.

Table 4.5: Forecasted values (with 95% confidence limits) for area and yield of groundnut in Odisha by using the selected ARIMA model

	Year	Forecasted value	Lower confidence limit (95%)	Upper confidence limit (95%)
Area (in ‘000 ha)	2016-17	217.20	175.07	266.60
	2017-18	195.90	144.00	260.85
	2018-19	176.34	112.62	264.18
Yield (kg/ha)	2016-17	1795.02	1152.18	2678.18
	2017-18	1792.09	995.17	2997.32
	2018-19	1858.33	745.57	3918.20

Figure 4.5 and 4.6 that the observed values and the fit values of area and yield of groundnut along with their upper and lower limit as obtained from their last best fit ARIMA model.

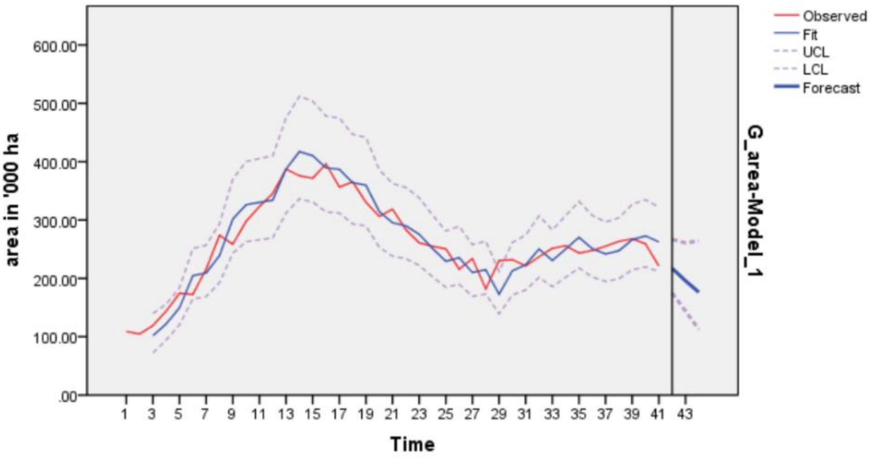


Figure 4.5: Observed and fit values of groundnut area along with upper and lower limit by using best fit ARIMA (2,2,0) without constant model

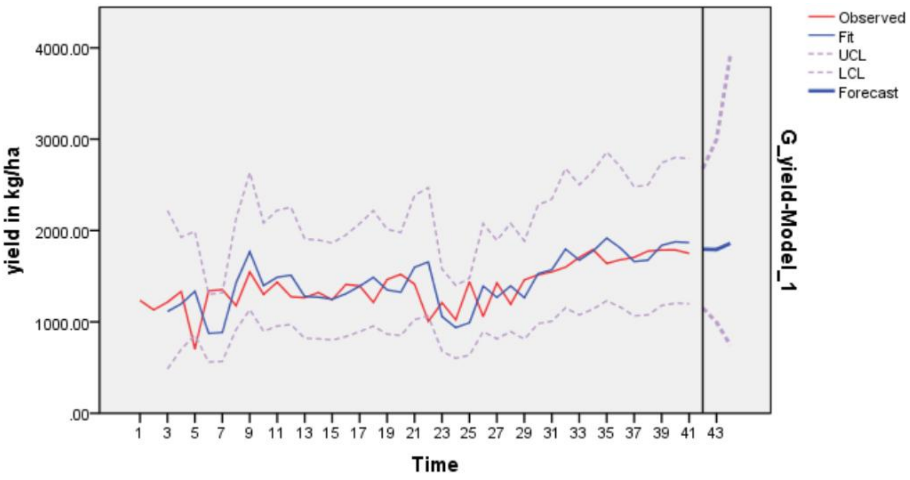


Figure 4.6: Observed and fit values of groundnut yield along with upper and lower limit by using best fit ARIMA (2,2,0) without constant model

In table 4.6, by using the forecasted values for area and yield of groundnut are obtained from the respective best fit ARIMA model the production is calculated. It shows that there is a decrease in the calculated values of production from 20.16-17 to 2018-19.

Table 4.6: Calculated value of production forecast by using the forecasted value of area and yield of groundnut in Odisha

	Year	**Calculated value**
Production in '000 tonnes	2016-17	389.88
	2017-18	351.07
	2018-19	327.69

4.2 Forecasting of area, yield and production of mustard by fitting appropriate ARIMA model

The original plot of data on area under mustard as shown in figure 4.7 (a) explains that the data is non-stationary that says it don't have constant mean and variance. Thus, the second difference of the data is plotted after plotting the first difference figure 4.7(b) of the data which was not stationary and shown in figure 4.7(c) this plot shows the second difference of data is found to be stationary which have constant mean and variance.

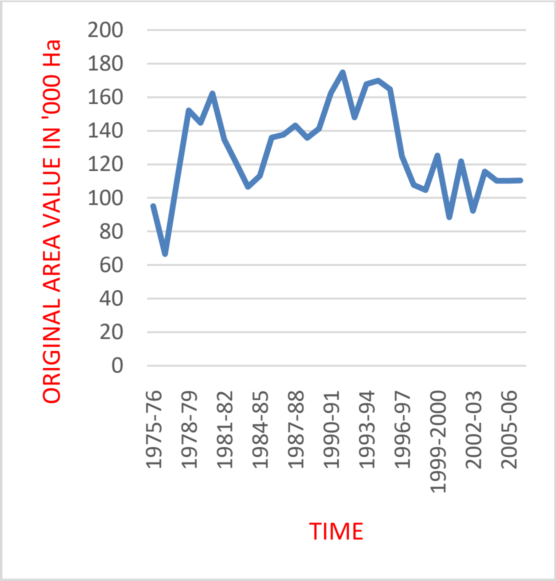


Figure 4.7(a): Plot of original value of area under mustard *vs* time

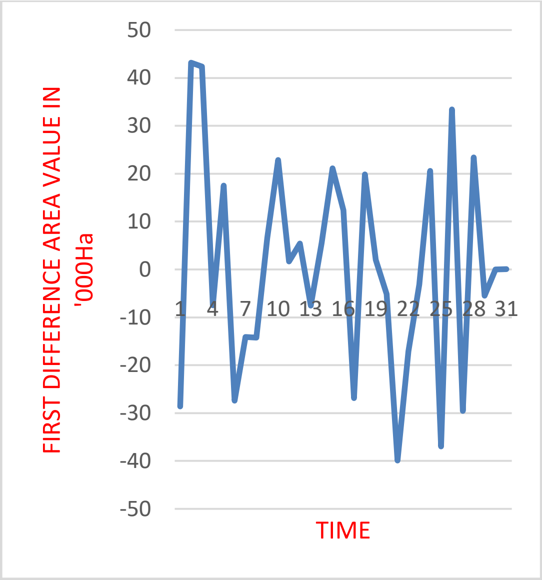


Figure 4.7(b): Plot of first difference value of area under mustard *vs* time

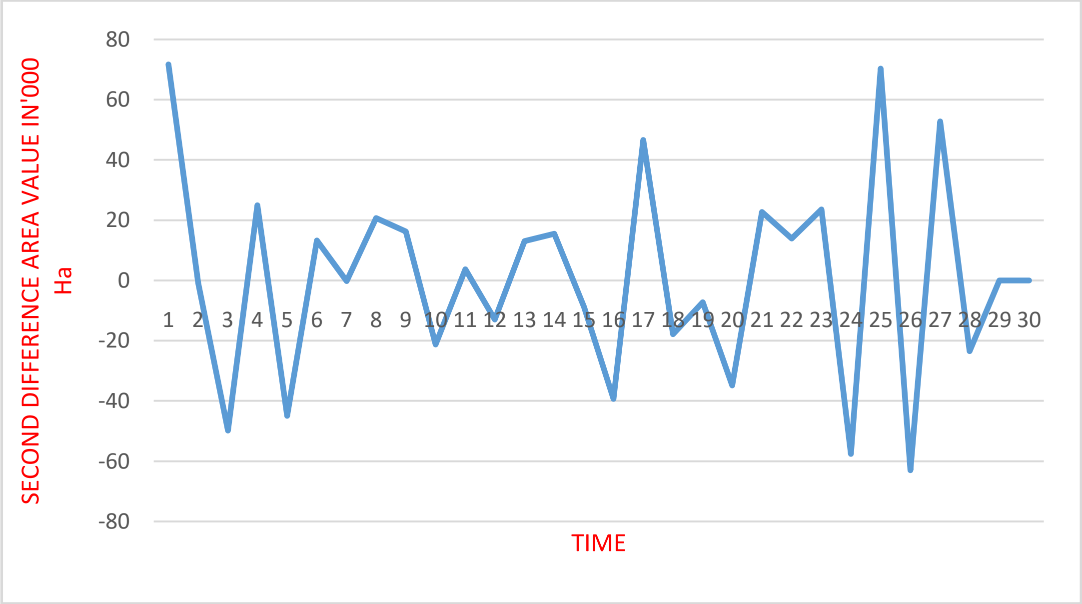


Figure 4.7(c): Plot of second difference value of area under mustard *vs* time

The ACF and PACF plots of the second difference value of mustard area is shown in the figure 4.8 Which shows that the provisional value of q and p that would be satisfactory of mustard are q=0 and p=1. Thus the ARIMA model found fitted for area of mustard is ARIMA (1,2,0).

The original plot of data on yield under mustard as shown in figure 4.9 (a) explains that the data is non-stationary that says it don't have constant mean and variance. Thus, the first difference of the data is plotted and shown in figure 4.9(b), this plot shows the first difference of data is found to be stationary which have constant mean and variance.

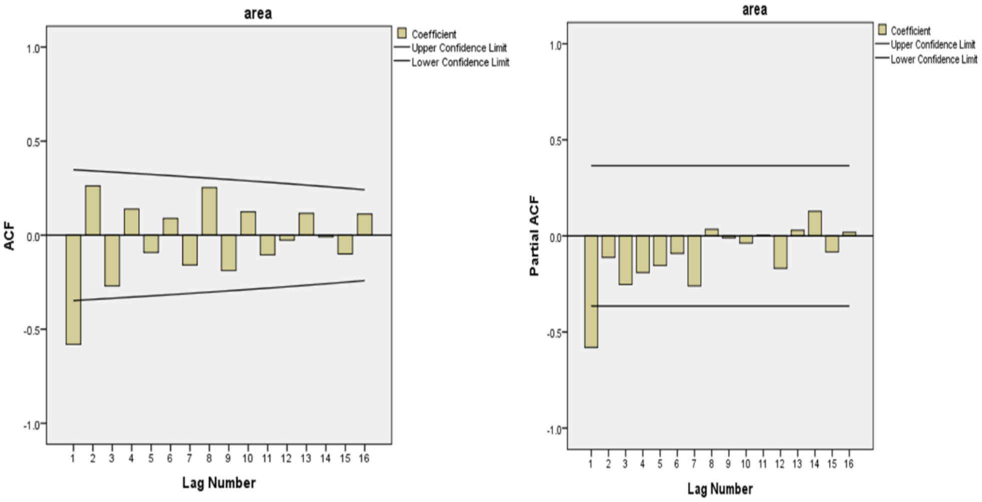


Figure 4.8: ACF and PACF plot of first difference values of area under mustard

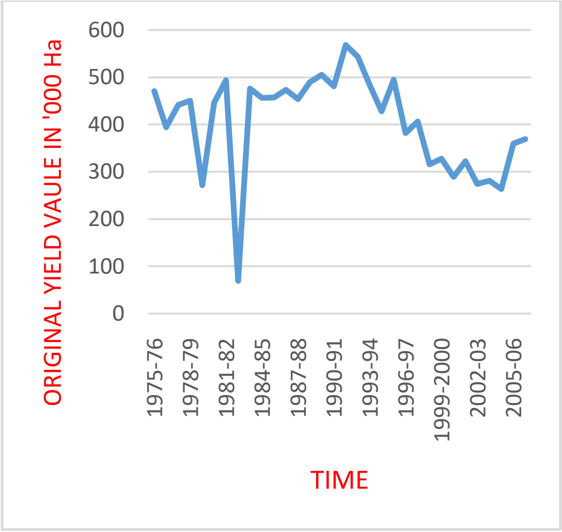
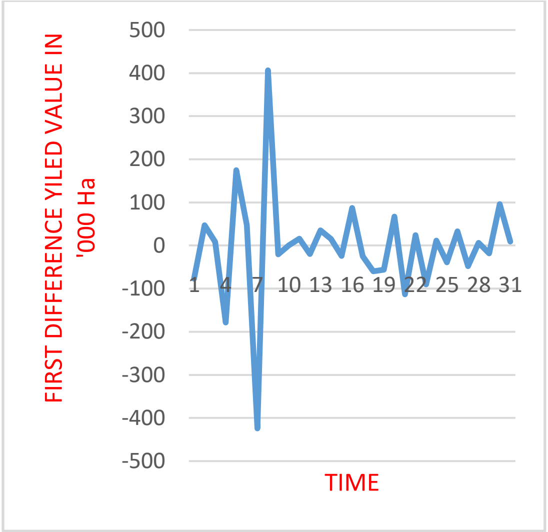


Figure 4.9(a): Plot of original value of yield of mustard *vs* time



4.9(b): Plot of first difference value of yield of mustard *vs* time

The ACF and PACF plots of the first difference value of mustard yield is shown in the figure 4.10 Which shows that the provisional value of q and p that would be satisfactory of mustard are q=1 and p=0. Thus the ARIMA model found fitted for yield of mustard is ARIMA (0,1,1).

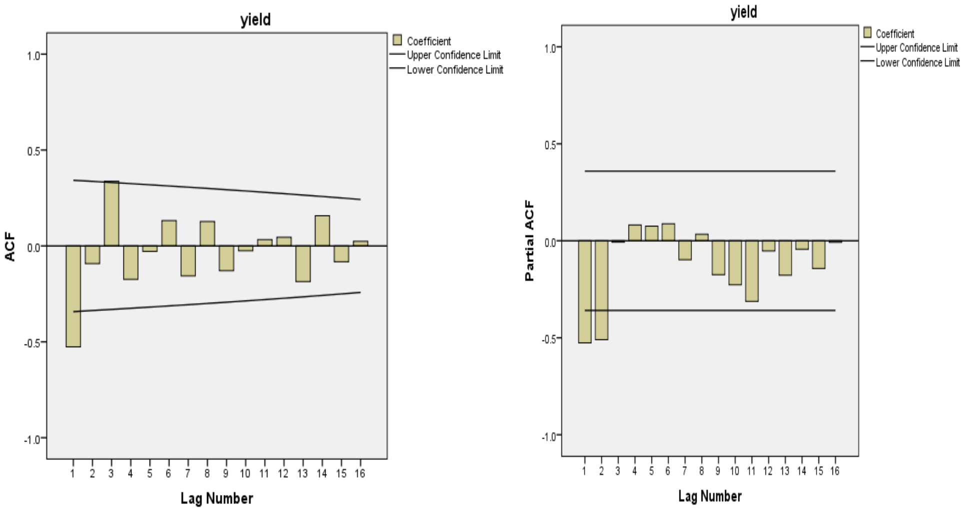


Figure 4.10: ACF and PACF plot of first difference values of yield of mustard

The study of table 4.7 shows that when ARIMA (1,2,0) is fitted to the data on area under mustard, the constant is not significant. So ARIMA (1,2,0) without constant is also fitted. The estimated coefficient of AR (1) is found to be significant. Thus the selected ARIMA model for area under mustard is ARIMA (1,2,0) without constant. In case of yield of mustard ARIMA (0,1,1) is fitted to the data, the constant is not significant. So ARIMA (0,1,1) without constant is also fitted. The estimated coefficient of MA (1) is found to be significant. Thus the selected ARIMA model for mustard yield is ARIMA (0,1,1).

Table 4.7: Coefficient of AR and MA components of the fitted ARIMA model considered for forecasting area and yield of mustard in Odisha

	Best fit ARIMA model	**Constant (μ)**	**Coefficient of autoregressive components**		**Coefficient of moving average components**	
			α_2	**θ_1**	**α_2**	**θ_1**
Area	ARIMA(1,2,0)	0.014 (0.047)	-0.719** (0.113)	–	–	–
	ARIMA(1,2,0) (without constant)	–	-0.722** (0.111)	–	–	–
Yield	ARIMA(0,1,1)	-0.009 (0.033)	–	–	0.801** (0.110)	–
	ARIMA(0,1,1) (without constant)	–	–	–	0.809** (0.108)	–

(Figures in the parentheses indicate the standard error)

* Significant at 5% level of significance, ** Significant at 1% level of significance.

Table 4.8: Model Fit Statistics and Residual Diagnostics of the ARIMA models fitted for area and yield of mustard in Odisha

	Model	Model fit statistics		Residual diagnostics	
		RMSE	MAPE	Ljung – Box Q Statistic	Shapiro-Wilk's Statistic
Area	ARIMA(1,2,0)	31.026	17.190	9.687	0.921
	ARIMA(1,2,0) (without constant)	**30.032**	**17.028**	**9.582**	**0.914**
Yield	ARIMA(0,1,1)	105.224	30.515	12.302	0.923
	ARIMA(0,1,1) (without constant)	**104.254**	**30.855**	**12.818**	**0.917**

(Models highlighted as bold are the best fit models)

Table 4.9: Cross validation of the selected best fit ARIMA(1,2,0) without constant model for area of mustard in Odisha

Year	Actual value (in '000 ha) (Y)	Forecasted Value (in '000 ha) $(\hat{Y})$	Error $(Y-\hat{Y})$	Absolute Percentage Error $\left(\frac{\lvert Y-\hat{Y}\rvert}{Y}\right)$ X 100
2007-08	110.29	110.87	-0.58	0.5258863
2008-09	109.93	110.78	-0.85	0.773219321
2009-10	112.19	110.16	2.03	1.809430431
2010-11	112.45	112.84	-0.39	0.346820809
2011-12	126.67	114.4	12.27	9.686587195
2012-13	116.37	131.32	-14.95	12.84695368
2013-14	145.36	123.94	21.42	14.73582829
2014-15	121.98	145.46	-23.48	19.24905722
2015-16	99.69	136.33	-36.64	36.75393721

The study of table 4.8 Shows that all the fitted model satisfy the assumptions of normality of error as they all have non-significant S-W statistic and also all the models are found to be adequate due to non-significant Ljung-box Q statistic. For area under mustard ARIMA(1,2,0) without constant has low value of RMSE and MAPE ,so the best fit model is ARIMA(1,2,0) without constant. For yield of mustard ARIMA (0,1,1) without constant has low value of RMSE and MAPE, so the bestfit model ARIMA(0,1,1) without constant.

The cross validation of the selected best fit ARIMA(1,2,0) without constant model for area under mustard presented on the table 4.9 shows that the absolute percentage error are quite low, thus the selected model is successfully cross validated.

The cross validation of the selected best fit ARIMA(0,1,1) without constant model for yield under mustard presented on the table 4.10 shows that the absolute percentage error are quite low, thus the selected model is successfully cross validated.

Table 4.10: Cross validation of the selected best fit ARIMA(0,1,1) without constant model for yield of mustard in Odisha

Year	Actual value (in kg/ha) (Y)	Forecasted Value (in kg/ha) $(\hat{Y})$	Error $(Y-\hat{Y})$	Absolute Percentage Error $\left(\frac{\lvert Y-\hat{Y}\rvert}{Y}\right)$ X 100
2007-08	375.1	358.38	16.72	4.457478006
2008-09	382.97	366.79	16.18	4.224874011
2009-10	369.37	375.26	-5.89	1.594607034
2010-11	375.1	379.63	-4.53	1.207677953
2011-12	415.73	384.36	31.37	7.545762875
2012-13	422.02	395.97	26.05	6.172693237
2013-14	423.98	406.81	17.17	4.049719326
2014-15	424	416.19	7.81	1.841981132
2015-16	421.01	423.96	-2.95	0.700695945

In table 4.11, the forecasted values for area and yield of mustard are obtained from the respective best fit ARIMA model. It shows that there is a decrease in the forecasted values of area and yield from 2016-17 to 2018-19.

Figure 4.11 and 4.12 that the observed values and the fit values of area and yield of mustard along with their upper and lower limit as obtained from their last best fit ARIMA model.

Table 4.11: Forecasted values (with 95% confidence limits) for area and yield of mustard in Odisha by using the selected ARIMA model

	Year	Forecasted value	Lower confidence limit (95%)	Upper confidence limit (95%)
Area (in '000 ha)	2016-17	82.94	51.80	126.52
	2017-18	68.59	31.17	132.81
	2018-19	59.87	15.58	164.11
Yield (kg/ha)	2016-17	429.79	195.03	832.88
	2017-18	432.00	193.01	846.26
	2018-19	434.24	191.06	859.72

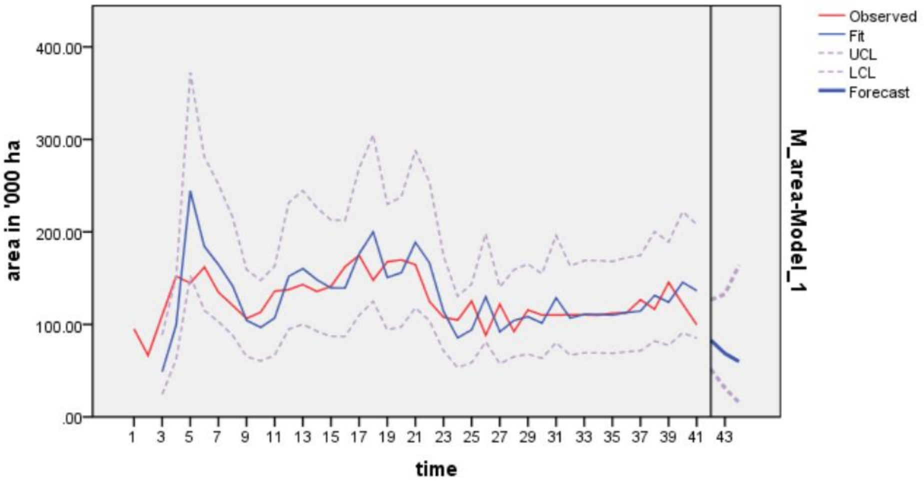


Figure 4.11: Observed and fit values of mustard area along with upper and lower limit by using best fit ARIMA (1,2,0) without constant model

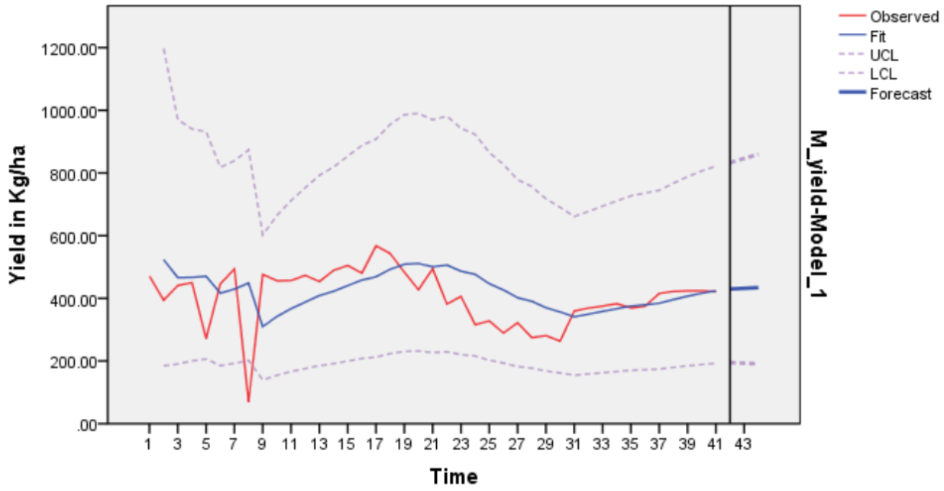


Figure 4.12: Observed and fit value of mustard yield along with upper and lower limit by using best fit ARIMA(0,1,1) without constant model

In table 4.12, by using the forecasted values for area and yield of mustard are obtained from the respective best fit ARIMA model the production is calculated. It shows that there is a decrease in the calculated values of production from 2016-17 to 2018-19.

Table 4.12: Calculated value of production forecast by using the forecasted value of area and yield of mustard in Odisha

	Year	**Calculated value**
Production in ‘000 tonnes	2016-17	34.65
	2017-18	29.63
	2018-19	26.00

4.3 Forecasting of area, yield and production of Sesamum by fitting appropriate ARIMA model

In figure 4.13(a), the original plot of data on area under sesamum shows the data is non-stationary that implies it doesn’t have constant mean and variance. Thus, the first difference of the data is plotted and shown in figure 4.13 (b) shows that the first difference of data has constant mean and variance, which makes the data stationary.

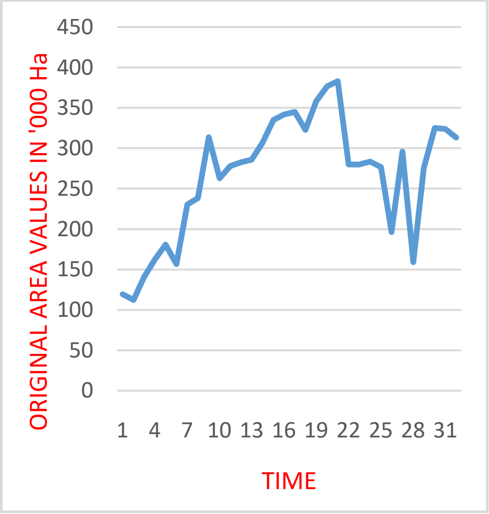
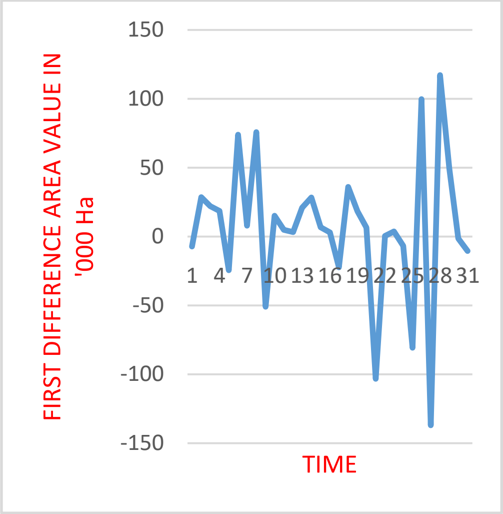


Figure 4.13(a): Plot of original values of area under sesamum *vs* time



4.13(b): Plot of first difference value of area under sesamum *vs* time

The ACF and PACF plot of the second difference values of sesamum area is shown in Figure. 4.14, gives the tentative value of q and p suitable for the area of sesamum is q=0 and p=1. Thus the ARIMA model found to be best fitted for area of groundnut is ARIMA(1,1,0).

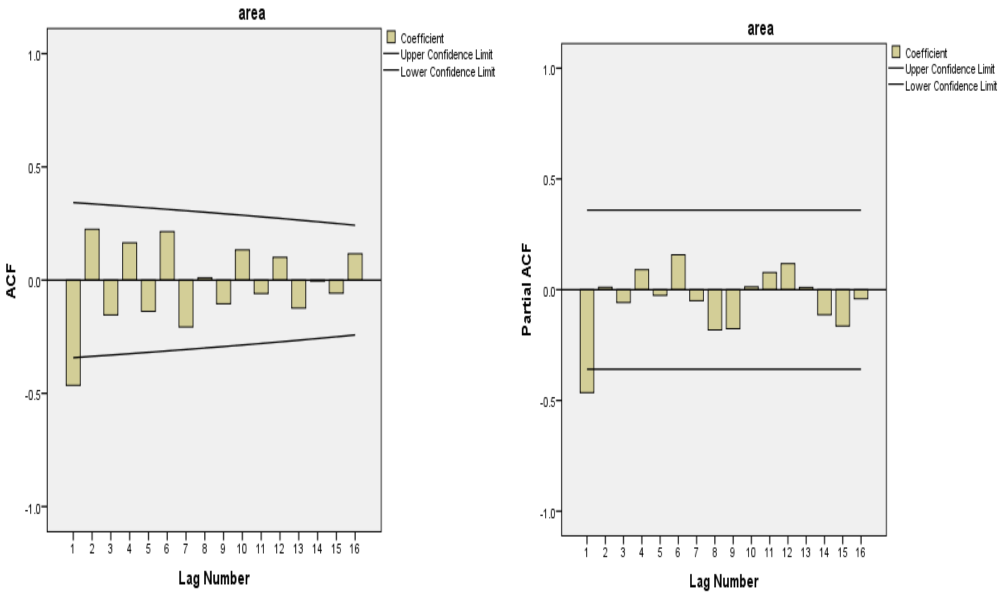


Figure 4.14: ACF and PACF plot of second difference values of area under sesamum

Figure 4.15(a) is the plotted data on original yield under sesamum which is non-stationary that says it doesn't have constant mean and variance. So, the first difference of the data is plotted and shown in figure 4.15(b) showing the first difference of data having constant mean and variance making the data stationary.

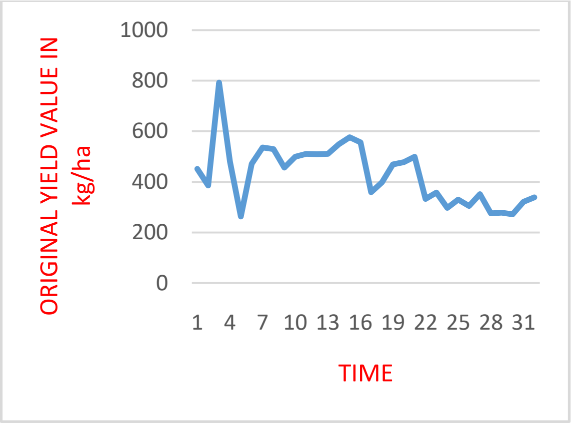


Figure 4.15(a): Plot of original values of yield of sesamum *vs* time

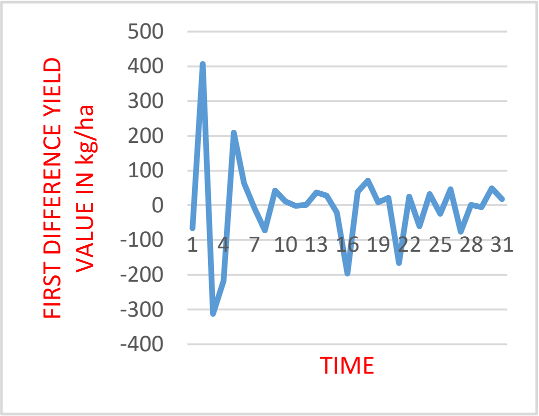


Figure 4.15(b): Plot of first difference value of yield of sesamum *vs* time

The ACF and PACF plot of second difference values of sesamum yield is shown in figure. 4.16 suggesting that the tentative value of q and p that would be suitable for yield of sesamum is q= 0 and p=2, so the ARIMA model that is found to be best fitted for yield of groundnut is ARIMA (2,1,0).

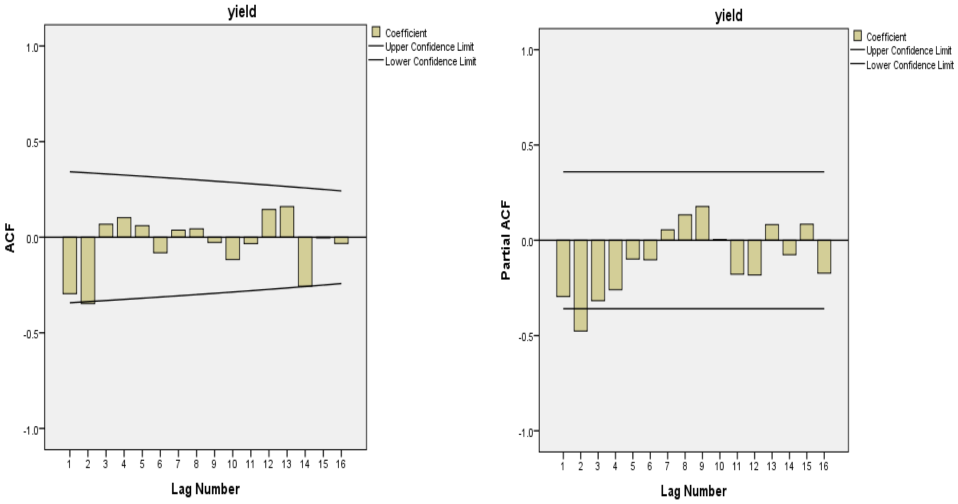


Figure 4.16: ACF and PACF plot of second difference values of yield of sesamum

The study of table 4.13 shows that when ARIMA (1,1,0) is fitted to the data on area under sesamum, the without constant is significant. So,ARIMA(1,1,0) with constant is also fitted. The estimated AR(1) is also found to be significant. Thus, the selected ARIMA model for area under sesamum is ARIMA (1,1,0) with constant. In case of yield when ARIMA (2,1,0) is fitted to the data on yield under sesamum, the constant is not significant. So, ARIMA (2,1,0) without constant is also fitted. The estimated AR (2) is also found to be significant. Thus, the selected ARIMA model for yield under sesamum is ARIMA (2,1,0) without constant.

Table 4.13: coefficients of AR and MA components of the fitted ARIMA model considered for forecasting area and yield of sesamum in Odisha

	Best fit ARIMA model	**Constant (μ)**	**Coefficient of autoregressive components**		**Coefficient of moving average components**	
			α_1	α_2	θ_1	θ_2
Area	ARIMA(1,1,0)	0.097* (0.037)	0.533** (0.138)	–	–	–
	ARIMA(1,1,0) (without constant)	–	-0.450* (0.145)	–	–	–
Yield	ARIMA(2,1,0)	-0.023 (0.038)	-0.383* (0.145)	-0.528** (0.139)	–	–
	ARIMA(2,1,0) (without constant)	–	-0.381* (0.144)	-0.531** (0.139)	–	–

(Figures in the parentheses indicate the standard error)

* Significant at 5% level of significance, ** Significant at 1% level of significance.

The study of table 4.14 shows that all the fitted model satisfy the assumption of normality of error as they all have non-significant S-W Statistic and also all the models are found to be adequate due to non-significant Ljung-Box Q statistic. For area under sesamum, the ARIMA (1,1,0) with constant has low value of RMSE and MAPE. So the best-fit model is ARIMA (1,1,0) with constant. In case of yield of sesamum, the ARIMA (2,1,0) without constant has low value of RMSE and MAPE that makes it the best fit model.

Table 4.14: Model fit Statistics and Residual Diagnostics of the ARIMA models fitted for area and yield of sesamum in Odisha

	Model	**Model fit Statistics**		**Residual diagnostics**	
		RMSE	**MAPE**	**Ljung-Box Q Statistics**	**Shapiro-Wilk's Statistics**
Area	**ARIMA(1,1,0)**	**43.151**	**11.789**	**9.325**	**0.914**
	ARIMA(1,1,0) (without constant)	43.392	12.967	7.048	0.925
Yield	ARIMA(2,1,0)	99.079	14.669	24.828	0.924
	ARIMA(2,1,0) (without constant)	**98.366**	**14.568**	**24.411**	**0.918**

(Models highlighted as bold are the best fit models)

Table 4.15: Cross validation of the selected best fit ARIMA (1,1,0) with constant model for area of sesamum in Odisha

Year	**Actual value (in '000 ha) (Y)**	**Forecasted Value (in '000 ha) $(\hat{Y})$**	**Error $(Y-\hat{Y})$**	**Absolute Percentage Error $\left(\frac{\lvert Y-\hat{Y}\rvert}{Y}\right)$ X 100**
2007-08	304.77	310.85	-6.08	1.994947009
2008-09	298.87	299.69	-0.82	0.274366782
2009-10	284.19	290.85	-6.66	2.343502586
2010-11	260.62	279.53	-18.91	7.255774691
2011-12	235.68	259.83	-24.15	10.24694501
2012-13	230.19	235.37	-5.18	2.250314957
2013-14	212.85	219.37	-6.52	3.06319004
2014-15	233.17	207.65	25.52	10.94480422
2015-16	208.97	206.64	2.33	1.114992583

The cross validation of the selected best fit ARIMA(1,1,0) with constant model for area under sesamum presented on the table 4.15 shows that the absolute percentage error are quite low, thus the selected model is successfully cross validated.

The cross validation of the selected best fit ARIMA(1,2,0) without constant model for yield under groundnut presented on the table4.16 shows that the absolute percentage error are quite low, thus the selected model is successfully cross validated.

Table 4.16: Cross validation of the selected best fit ARIMA (2,1,0) without constant model for yield of sesamum in Odisha

Year	**Actual value (in kg/ha) (Y)**	**Forecasted Value (in kg/ha)** $(\hat{Y})$	**Error** $(Y-\hat{Y})$	**Absolute Percentage Error** $\left(\frac{\lvert Y-\hat{Y}\rvert}{Y}\right)$ **X 100**
2007-08	411.13	309.29	101.84	24.77075378
2008-09	428.58	377.3	51.28	11.96509403
2009-10	383.62	387.68	-4.06	1.058338981
2010-11	385.43	398	-12.57	3.261292582
2011-12	405.85	413.75	-7.9	1.94653197
2012-13	409.05	403.26	5.79	1.415474881
2013-14	403.34	403.38	-0.04	0.009917191
2014-15	405.54	410.12	-4.58	1.129358386
2015-16	392.21	413.89	-21.68	5.527651003

Table 4.17: Forecasted values (with 95% confidence limits) for area and yield of sesamum in Odisha by using the selected ARIMA model

	Year	**Forecasted value**	**Lower confidence limit (95%)**	**Upper confidence limit (95%)**
Area (in '000 ha)	2016-17	204.91	143.73	283.99
	2017-18	189.76	128.09	271.62
	2018-19	179.94	111.37	276.46
Yield (in kg/ha)	2016-17	402.14	259.76	596.86
	2017-18	409.29	243.71	648.43
	2018-19	404.97	238.18	647.84

In table 4.17, the forecasted values for area and yield of sesamum are obtained from the respective best fit ARIMA model. It shows that there is a decrease in the forecasted values of area and yield from 2016-17 to 2018-19.

Figure 4.17 and 4.18 that the observed values and the fit values of area and yield of sesamum along with their upper and lower limit as obtained from their last best fit ARIMA model.

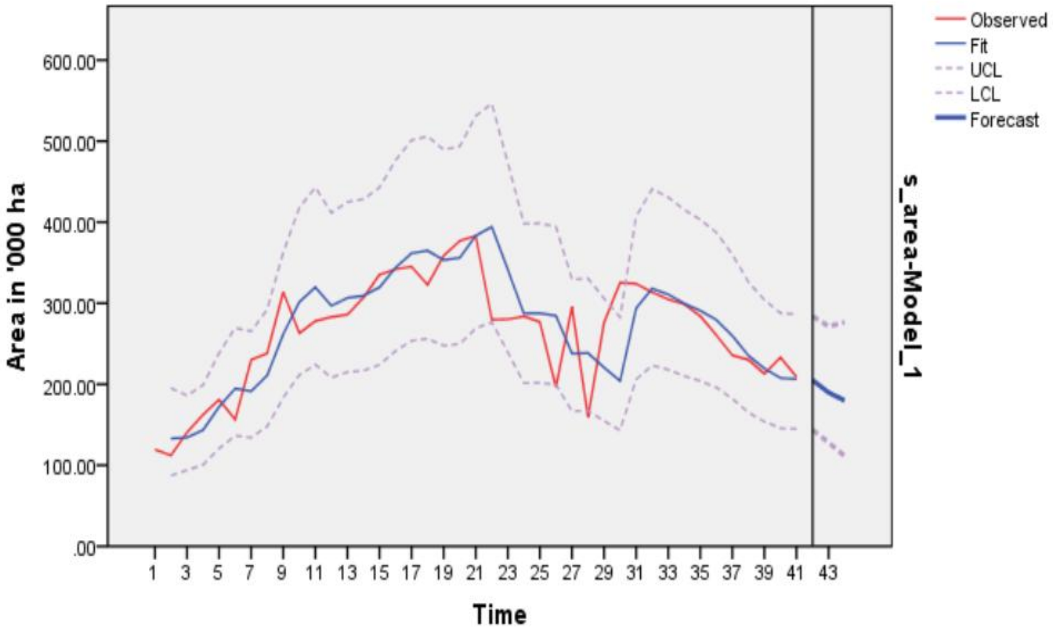


Figure 4.17: Observed and fit values of sesamum area along with upper and lower limit by using best fit ARIMA (1,1,0) with constant model

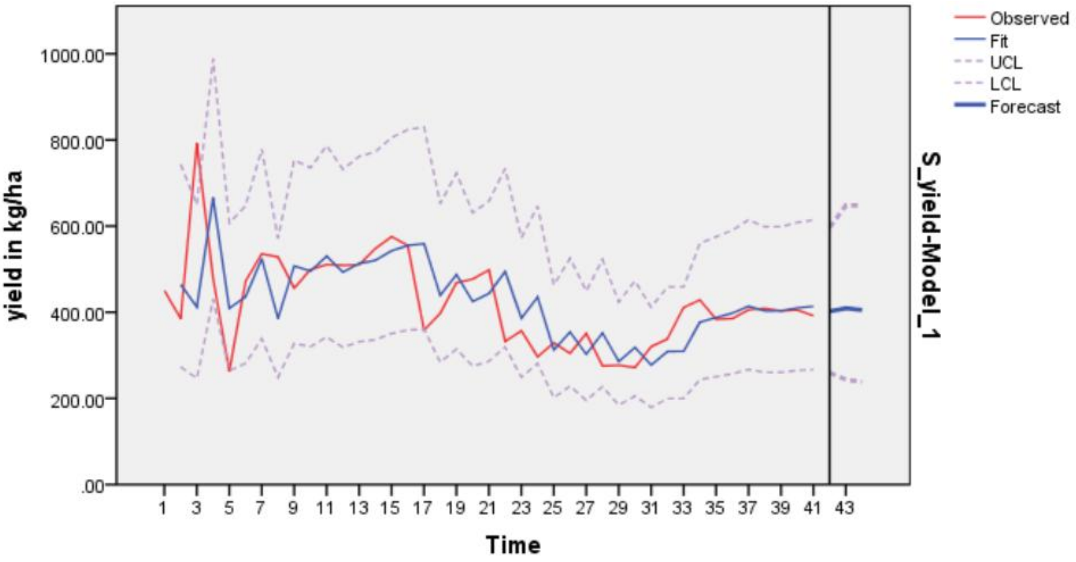


Figure 4.18: Observed and fit values of sesamum yield along with upper and lower limit by using best fit ARIMA (2,1,0) without any constant model

In table 4.18, by using the forecasted values for area and yield of sesamum are obtained from the respective best fit ARIMA model the production is calculated. It shows that there is a decrease in the calculated values of production from 2016-17 to 2018-19.

Table 4.18: Calculated value of production by using the forecasted value of area and yield of sesamum in Odisha

	Year	**Calculated value**
Production in '000 tonnes	2016-17	82.40
	2017-18	77.67
	2018-19	72.87

CHAPTER 5

SUMMARY AND CONCLUSION

ARIMA model is used for the forecast of the area, yield and production of important oilseed crops of Odisha. Generally univariate time series data are used by this model. In this chapter all the steps including the selection of possible ARIMA model, obtaining the best-fit ARIMA model among the selected model forecast the area and yield to calculating the forecasted production value of groundnut, mustard and sesamum are systematically summarised and final drawn conclusion are given below.

ARIMA (2,2,0) and ARIMA (2,2,0) without constant are selected models for forecasting the area under groundnut. ARIMA (2,2,0) is fitted first, the constant is non-significant. ARIMA (2,2,0) without constant is also fitted and that showed the AR (2) is significant. Due to the low RMSE and MAPE ARIMA (2,2,0) without constant is selected as the best-fitted model. In case of yield of groundnut ARIMA (2,2,0) and ARIMA (2,2,0) without constant are selected models. It's found that the constant is non-significant in ARIMA (2,2,0), so ARIMA (2,2,0) without constant is fitted and it is found that AR(2) is significant. Due to low RMSE and MAPE value of ARIMA (2,2,0) is selected as best fitted model for yield of groundnut.

In case of data on area and yield of groundnut, it is found that from the result of cross-validation of selected best fit ARIMA model that the absolute percentage error in almost all cases remains below 10% except a few. This shows that the cross-validation of the respective selected best fit models is successful and these models can be used for future prediction of area and yield. Also the success of cross-validation of area-validation of area and yield shows that the prediction of future values of production obtained from the forecasted values of area and yield is also valid.

The forecasted values of area of groundnut are found to decrease over the future years. The forecasted values yield is found to increase over the future years. The forecasted values of production are found to decrease over the future period.

ARIMA (1,2,0) and ARIMA (1,2,0) without constant are selected model for the forecasting of area under mustard. The constant is not significant in ARIMA (1,2,0) model, so ARIMA (1,2,0) without constant is fitted. It is found that AR (1) is significant. Due to low value of RMSE and MAPE, ARIMA (1,2,0) without constant is selected for the best-fit model. In case of yield of mustard ARIMA (0,1,1) and ARIMA (0,1,1) without constant are selected model for forecasting. As the constant is found to be not significant in ARIMA (0,1,1) , ARIMA (0,1,1) without constant is fitted. It is found that MA(1) is significant. Due to the low value of RMSE and MAPE, ARIMA (0,1,1) without constant is selected for best-fit model.

In case of data on area and yield of mustard, it is found that from the result of cross-validation of selected best fit ARIMA model that the absolute percentage error in almost all cases remains below 10% except a few. This shows that the cross-validation of the respective selected best fit models is successful and these models can be used for future prediction of area and yield. Also the success of cross-validation of area-validation of area and yield shows that the prediction of future values of production obtained from the forecasted values of area and yield is also valid.

The forecasted values of area of mustard are found to decrease over the future years. The forecasted values yield is found to increase over the future years. The forecasted values of production are found to decrease over the future period.

ARIMA (1,1,0) and ARIMA (1,1,0) without constant are selected model for forecasting the area under sesamum. It is found that constant is significant. AR (1) is also significant in both the models ARIMA (1,1,0) and ARIMA (1,1,0) without constant. As the RMSE and MAPE values are low in ARIMA (1,1,0) without constant, it is selected for the best fitted model. In case of yield of sesamum ARIMA (2,1,0) and ARIMA (2,1,0) without constant are selected model for forecasting. It is found that the constant is not-significant in ARIMA (2,1,0), so ARIMA (2,1,0) without constant is fitted and it is found that AR (2) is significant. Due to low value of RMSE and MAPE, ARIMA (2,1,0) without constant is selected as best- fitted model.

In case of data on area and yield of sesamum, it is found that from the result of cross-validation of selected best fit ARIMA model that the absolute percentage error in almost all cases remains below 10% except a few. This shows that the cross-validation of the respective selected best fit models is successful and these models can be used for future prediction of area and yield. Also the success of cross-validation of area-validation of area and yield shows that the prediction of future values of production obtained from the forecasted values of area and yield is also valid.

The forecasted values of area of sesamum are found to decrease over the future years. The forecasted values yield is found to decrease over the future years. The forecasted value of production is found to decrease over the future period.

In groundnut and mustard the decrease in area forecast leads to decrease in the production forecast though the yield forecast increases. This shows that the production forecast is in accordance with yield forecast in case of groundnut and mustard. In case of sesamum decrease in area forecast leads to decrease in production and yield forecasts. This shows that the production forecast is in accordance with both the yield and area forecasts in case of sesamum.

REFERENCES

Badmus MA and Ariyo OS. 2011. Forecasting Cultivated Areas and Production of Maize in Nigerian using ARIMA Model, *Asian Journal of Agricultural Sciences*, 3(3): 171-176

Biswas B, Dhaliwal LK, Singh SP and Sandhu SK. 2014. Forecasting wheat production using ARIMA model in Punjab, *International Journal of Agricultural Sciences*, Vol. 10, Issue 1

Cenas PV. 2017. Forecast of Agricultural Crop Price using Time Series and Kalman Filter Method, *Asia Pacific Journal of Multidisciplinary Research*, Vol. 5,No. 4

Chukwujioke OD and Okiemute AB. 2018. Forecasting of Cashew Area Harvested, Yield and Production Using Trend Analysis, *International Journal of Agricultural Economics*, 3(4): 65-71

Darekar A. and Reddy AA. (2017). Forecasting of Common Paddy Prices in India, *Journal of Rice Research* 2017, Vol 10 No. 1

Dash A, Dhakne DS, and Bhattacharya D. 2017. Forecasting of food grain productionin Odisha by fitting ARIMA model, *Journal of Pharmacognocy and Phytochemistry*, 6(6), 1126-1132

Debnath MK,Bera K and Mishra P. 2013. Forecasting Area, Production and Yield of Cotton in India using ARIMA Model, *STM Journals*, JoPC(2013) 16-20

Frah EAM. 2016. Sudan Production of Sorghum, Forecasting 2016-2030 Using Autoregressive Integrated Moving Average ARIMA Model, *American Journal of Mathematics and Statistics*, 6(4): 175-181

Hemavathi M, Prabakaran K and Meena SP. 2017. Forecasting Food grains Area and Production in India using ARIMA Model, *Statistical Approaches on Multidisciplinary Research,* ISSN: 2349 – 4891

Ilić I, Jovanović S andMili V. 2016. Forecasting corn production in Serbia UsingARIMA model, *Economics of Agriculture*, 4/2016UDC:633.15:658.51:339.564 (497.11)

Kumar M, Raman RK and Kumar S. 2017. Sugarcane Productivity in Bihar- A Forecast through ARIMA Model, *International Journal of Pure and Applied Bioscience*, 5(6): 1042-1051

Nath B, Dhakre DS and Bhattacharya D. 2018. Forecasting wheat production in India: An ARIMA modelling approach, *Journal of Pharmacognosy and Phytochemistry*,8(1): 2158-2165

Panasa V, Kumari RV, Ramakrishna G and Kaviraju S. 2017. Maize Price Forecasting Using Auto Regressive Integrated Moving Average (ARIMA) Model, *International Journal of Current Microbiology and Applied Sciences,* ISSN: 2319-7706 Volume 6 Number 8 (2017) pp. 2887-2895

Poyyamozhi S and Mohideen AK. 2017. Forecasting Of Cotton Production In India using ARIMA Model, *Asia Pacific Journal of Research*, ISSN (Print): 2320-5504, ISSN (Online) : 2347-4793

Sahu K, Mishra P, Dhekale BS, Vishwajith KP and Padmanaban K. 2014. Modelling and Forecasting of Area, Production, Yield and Total Seeds of Rice and Wheat in SAARC Countries and the World towards Food Security, *American Journal of Applied Mathematics and Statistics*, 2015, Vol. 3, No. 1, 34-48

Sharma PK, Dwivedi S, Ali L and Arora RK. 2018. Forecasting Maize Productionin India using ARIMA Model, *Agro Economist - An International Journal Citation*, AE: 5(1): 01-06

Singh DP, Kumar P and Prabakaran K. 2013. Application of ARIMA model for forecasting Paddy production in Bastar division of Chhattisgarh, *American International Journal of Research in Science, Technology, Engineering and Mathematics*, ISSN (Print): 2328-3491, ISSN (Online): 2328-3580, ISSN (CD-ROM): 2328-3629

Solanki P, Sharma R and Sharma M. 2017. Price forecasting of mustard using ARIMA and EGARCH models, *Journal of Oilseed Brassica*, 8 (2) : 181-186.

Suresh KK and Priya SRK. 2011. Forecasting Sugarcane Yield of Tamilnadu Using ARIMA Models, *Sugar Tech*, 13(1):23-26

Vijay N and Mishra GC. 2018. Time Series Forecasting Using ARIMA and ANNModels for Production of Pearl Millet (BAJRA) Crop of Karnataka, India, *International Journal of Current Microbiology and Applied Sciences*, ISSN: 2319-7706 Volume 7 Number 12

Wali VB, Beeraladinni D, and Lokesh H. 2017. Forecasting of Area and Production ofCotton in India: An Application of ARIMA Model, *International Journal ofPure and Applied Bioscience*, 5(5): 341-34

Fitting of ARIMA Model For Forecasting Oilseed Crops

ISBN: 978-93-91768-68-3

About Authors

Madhu Chhanda Kishan did her B.Sc. (Agriculture) Honors in 2013-17 from College of Agriculture, OUAT, Bhubaneswar. She did her M.Sc. (Ag.) in Agricultural Statistics in 2017-19 from Department of Agricultural Statistics, College of Agriculture, OUAT, Bhubaneswar. She conducted her M.Sc. research thesis under the guidance of Dr. Abhiram Dash. She has qualified UGC-NET-JRF in Population studies in 2020 conducted by National Educational Testing Bureau of University Grants Commission. Her area of interest for research is time series analysis. Currently she is working as Assistant Professor (Agricultural Statistics) in College of Agriculture, OUAT, Bhawanipatna.

Dr. Abhiram Dash is working as Agriculture Scientist in the field of Agricultural Statistics as Assistant Professor in the Department of Agricultural Statistics, College of Agriculture, Bhubaneswar, Odisha University of Agriculture and Technology. He has to his credit published more than 45 innovative articles and research papers in reputed national and international journals. He also authored three books related to Agricultural Statistics, R software and Question bank on JRF/SRF in Agricultural Statistics and Computer Applications. He has expertise in forecast modelling of crops, sampling techniques and design of experiments.

Bhumi Publishing

ISBN 978-93-91768-68-3

9 789391 768683