THINK INDIA JOURNAL

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Forecasting of Oilseed Crop Production through Time Series Modelling

Banti Kumar[1], Monika, Rajeev Prakash Bhanot[1]and Manish Kumar Sharma[2]

[1]*Assistant Professor, Department of Mathematics,*

School of Chemical Engineering & Physical Sciences, Lovely Professional University,

Phagwara-144411, Punjab

[2]*Professor and Head, Division of Statistics and CS, SKUAST Jammu-180009, J&K*

Corresponding Author's Email: -bantikumar1573@gmail.com

Abstract

Agriculture is a backbone of India's economy. Oilseeds have been playing a vital role after foodgrains in India's economy in terms of area and production. From 1974-75 to 2018-19, oilseed production have shown an overall increasing trend but due to increasing population we have to import oilseed production from other countries in order tomitigate the demands of our country people. In the present study, an attempt has been made to propose an autoregressive integrated moving average (ARIMA) model to forecast the oilseed production in India which will help to analyse the past and current behaviour of oilseeds production. In this study, an effort has been made to forecast production of oilseeds by using ARIMA model, a most widely used for forecasting time series problems.

Introduction

In India, oilseed is an important crop in terms of both production and consumption. India is one of the leading producer of oilseeds globally and hence oilseed sector occupies an important position in the country's economy. A variety of oilseeds are produced in India such as groundnut, castor seed,sesame, rapeseed and mustard, linseed, soybean and safflower.After food grains, oilseeds provides a great contribution in the Indian agricultural economy in terms of area and production. While oilseeds covered 26087.2 thousand hectares, the area under food grains was 125298.7 thousand hectare during 2015-16 whereas the production of oilseeds was 25250.8 thousand tonnes and that of food grains was 254595.9 thousand tonnes [4].The productivity of oilseeds in India was 1408 kg/ha during 2015-16, whereas it was 3173 kg/ha in USA, 2864 kg/ha in Brazil and 2074 kg/ha in China respectively [7]. The reason behind low and irregular productivity of oilseeds is primarily because of its cultivationon marginal lands where proper irrigation facility may not be available and inputs are applied at low levels.Indian Government is running many development programs, such as Oilseed

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Growers Cooperative Project, National Oilseed and Development Project,Integrated Scheme on Oilseeds, Pulses, Oil Palm and Maize (ISOPOM) and Technology Mission Oilseeds (TMO)etc for improving the this scenario all over the country [5].In this paper, forecastingof oilseeds production has been done for the next five years (2019 to 2023). The model used for forecasting is an Autoregressive Integrated Moving Average (ARIMA) model. Since, this model was first proposed by Box and Jenkins (1976), hence it is also known as Box-Jenkins model [1]. The primary reason behind choosing ARIMA model for forecasting is that it assumes non-zero autocorrelation between the consecutive values of the time series data.The present work has been taken with the objectives

(i) to investigate the past trend in oilseed production in India
(ii) to develop an ARIMA model for oilseeds production in India and
(iii) toforecast the future oil seed production for next 5 years.

Sahu et al., (2010) attempt to analyse the production behaviour along with the total seeds of two major food crops rice and wheat and came up with strategies and programmes for regional cooperation in ensuring food security and reducing hunger and malnutrition in the region. They have forecasted area, production, yield and total seed production which helped to solve the problem of food security and also seed security in SAARC countries in future[8]. Further, they have used the Box – Jenkins ARIMA modelling technique to analyse the time series data from 1961 to 2010. Time series forecasting is an important statistical technique used as a basis for manual and automatic planning in many application domains (Gooijer and Hyndman 2006; Sonawane et al., 2013) [3, 10].

Materials and Methods

The data used for this study is the oilseeds production in India for the last 45 years, i.e., from 1974-75 to 2018-19 which were collected from agricultural statistics by reserve bank of India (RBI), Government of India (RBI statistics).[6]

Autoregressive Integrated Moving Average Model

For better understanding of the ARIMA model, one should know the following terms:Autoregression (AR): is a model obtained by regressing the variable of interest on its own lagged values.Integration (I): It depicts the number of times the differencing has been done to make series stationary.Moving average (MA): represents the output variable depends linearly on the present and various previous values of a stochastic term.

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Correlogram: It is a plot of the autocorrelations and partial correlation versus time lag of a time series. A typical ARIMA models has (p,d,q) order where p denotes the order of Autoregressive process, d denotes the no of times differencing has been done to make the series stationary and q is the order of moving average process. The patterns in a correlogram are used to analyze underlying behavior of a time series. Autocorrelation represents simple correlation between X_t and, say, X_{t+h}, it is a correlation between a series but with a lag. Mathematically, autocorrelation function is given by

$$\rho(p_{or}q) = \frac{\operatorname{cov}(y_t, y_{t-p_{or}q})}{\operatorname{var}(y_t)}$$

Partial autocorrelation is a relationship between two observations after removing the linear relationship of all observations that fall between those two observations.Mathematically,

$$\phi(p) = \frac{\rho_p - \sum_{j=1}^{p-1}\left(\phi_{p-2,j} - \phi_{pp}\phi_{p-2,p-j}\right)\rho_{p-j}}{1 - \sum_{j=1}^{p-1}\left(\phi_{p-2,j} - \phi_{pp}\phi_{p-2,p-j}\right)\rho_j}, p \geq 3$$

Ljung-Box Q statistic: This test is a used to check whether the observation are independent and random over time.The null hypothesis for this test is that the first k autocorrelations are jointly zero, H_0: $\rho_1 = \rho_2 = \ldots = \rho_k = 0$.Alternatively, at least one of $\rho_i$is not zero.Note that it is a test related to autocorrelation.Ljung-Box Q statistic is having asymptotic chi-squaredistribution.

$$Q(p) = T(T+2)\sum_{i=1}^{p}\frac{\rho_i^2}{T-1} \sim \chi_p^2$$

In a time series, time series plot and correlogram can help us in the identification of a suitable model for our time series data.

Stationarity: It means that the statistical properties of a time series generating process do not change over time. More specifically, stationarity means mean, variance, autocorrelation, etc. remains constant over time.A time series $Y_t$is said to be stationary if for all t the following holds true:

$$E(Y_t) = m < \infty, V(Y_t) = E(Y_t - m)^2 = g_0 < \infty \; and \; Cov(Y_t, Y_{t-k}) \\ = E\{(Y_t - m)(Y_{t-k} - m)\} = g_k$$

Unit root test: This test is used to test the stationarity of a time series and check whether the given time series possesses a unit root.Augmented Dickey fuller (1979) test is most widely

ISSN:0971-1260
Vol-22-Issue-37-December -2019

used which presumes that the error term in the Model are correlated [2].ARIMA is a mathematical model designed to forecast data within a time series. Time series is of two types:

- Stationary: A stationary process is one whose statistical properties remains same over time.
- Non- stationary: This process has properties which changes over time. The Box-Jenkins model makesthe non- stationary time series stationary by taking the difference between data points.

Box and Jenkinsproposed a methods for identification, estimation and diagnostic checking of a time series model. This methodology is now referred to as the Box-Jenkins Methodology.

Identification. Model identification is used to determine whether the given time series is stationary or not. Once a stationary model is found, it can be used for forecasting. If the series is not stationary, it is to be made stationary by differencing the series. The original series is replaced by a series of differencing and an ARIMA model is then specified for the differences series.

Estimation and Model Checking: This steps in ARIMA modelling includes diagnostic checking which involves the examination of autocorrelation and partial autocorrelation functions for estimation of residual behaviour. This helps us to identify the areas where the model is inadequate. In this stage, it is tested whether the estimated models has the specifications of univariate time series.

Forecasting: In this step, the estimated model is used to forecast the future values of the given time series.

Results and Discussion:

ISSN:0971-1260
Vol-22-Issue-37-December -2019

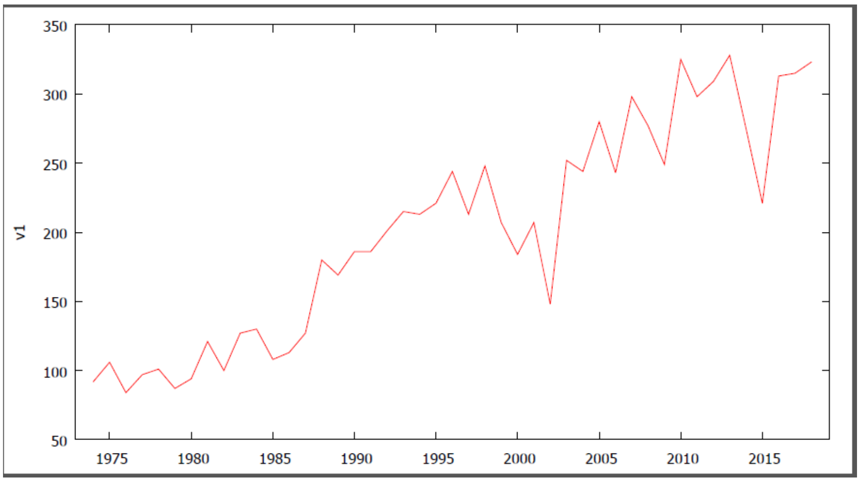


Fig.1: Behaviour of the oilseed production for the period of 1974 to 2018

The figure 1 showed the behaviour of the data of yearly oilseed production. The scenario of the oilseed production have shown an overall increasing trend from 1974-75 to 2018-19.

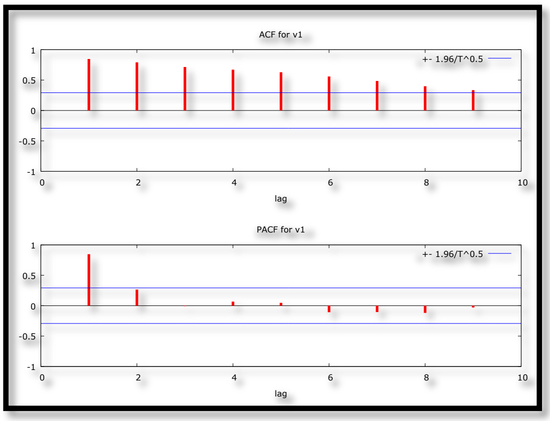
Table 1: Unit Root test for checking the stationary of data

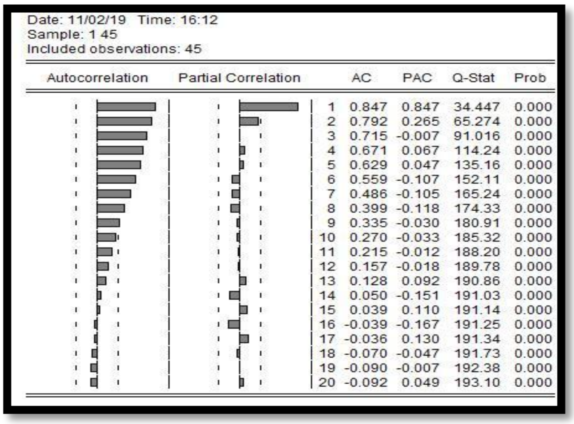
		t-statistic	Sig. value
ADF test statistic		-5.001	0.001
Test critical values	1% level of significance	-4.180	
	5% level of significance	-3.515	
	10% level of significance	-3.188	

ADF test is most widely used test for determining stationary of data. The ADF test has the hypothesis that H_0: the non-stationary of the data against H_1: the data is stationary. The critical value for the rejection of the null hypothesis of unit root test was highly significant with p-value (<.01). The results showed that the data regarding oilseed production was stationary. Therefore, it has been concluded that no differencing of the series was required.

The next step is identification of ARIMA order (p, d, q). The parameters of ARIMA models are estimated and is given in Table 2. Among different models fitted, the adequate model, i.e. ARIMA (1, 0, 0) has been identified based on ACF and PACF plots.

ISSN:0971-1260
Vol-22-Issue-37-December -2019





Date: 11/02/19 Time: 16:12
Sample: 1 45
Included observations: 45

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.847	0.847	34.447	0.000
		2	0.792	0.265	65.274	0.000
		3	0.715	-0.007	91.016	0.000
		4	0.671	0.067	114.24	0.000
		5	0.629	0.047	135.16	0.000
		6	0.559	-0.107	152.11	0.000
		7	0.486	-0.105	165.24	0.000
		8	0.399	-0.118	174.33	0.000
		9	0.335	-0.030	180.91	0.000
		10	0.270	-0.033	185.32	0.000
		11	0.215	-0.012	188.20	0.000
		12	0.157	-0.018	189.78	0.000
		13	0.128	0.092	190.86	0.000
		14	0.050	-0.151	191.03	0.000
		15	0.039	0.110	191.14	0.000
		16	-0.039	-0.167	191.25	0.000
		17	-0.036	0.130	191.34	0.000
		18	-0.070	-0.047	191.73	0.000
		19	-0.090	-0.007	192.38	0.000
		20	-0.092	0.049	193.10	0.000

Fig.2: Correlogram of oilseed production for the period of 1974 to 2018

Further, from Fig.2 it can be seen that the order if AR process is one since the first spike of the partial autocorrelation graph is lying beyond the confidence interval whereas from ACF graph, it can be concluded that the order of MA process is zero since the Autocorrelation spikes are gradually fading away. From the figure 2, it can be seen that the LBQ values increases and significant indicating that there is presence of autocorrelation.

Table 2: Parameter estimationof ARIMA (1, 0, 0) model for oilseed production in India

Variable	coefficient	Std. Error	T-statistic	Probability
C	203.678	49.3235	4.129	0.000
AR(1)	2.444	0.648	3.770	0.001
SIGMA SQ	0.024	0.005	4.652	0.000
Akaike info criterion	455.159			
Schwarz criterion	460.5789			
Hannan-Quinn criterion	457.1795			

An ARIMA model was built using E-views/Gretl software. The estimation of model ARIMA (1,0,0) with parameters have been displayed in the table 2. The estimate of AR (1) (2.444) is found to be statistically significant and positive. Among different models, ARIMA(1,0,0) model was selected on the basis of minimum value of Akiake information criterion(AIC), Schwarz Bayesian Information criterion (SBIC) and Hannan-Quinn criterion(HQ) criteria (Shafaqat 2012) [9].

ISSN:0971-1260
Vol-22-Issue-37-December -2019

Table 3: Forecasted values of Oilseed production using ARIMA (1, 0, 0)

Year	Forecasted Value	Std. Error	Upper confidence Limit	Lower confidence Limit
2019	312.61	34.876	244.25	380.96
2020	303.12	47.223	210.56	395.67
2021	294.46	55.451	185.77	403.14
2022	286.55	61.472	166.07	407.03
2023	279.33	66.072	149.83	408.83

One the basis of fitted ARIMA (1, 0, 0) model, forecasting of oilseed production has been made for next five years i.e., 2019 to 2023 ad have shown that the oilseed production will decrease in the coming years. Hence, govt. has to take necessary steps in order to keep oilseed production on the higher side.

Fig.3: Forecasting of oilseed production for next five years.

From the fig.3, it is clear that the oilseed production has shown a decreasing trend in the future.

Conclusion

In the present study, the ARIMA (1,0,0) was found tobe the best fitted modelas per minimum value of AIC, SBIC and statistical significance of parameters and then used for prediction for next 5 years of oilseed production in India. ARIMA (1,1,0) was used because the reason of its capability to make predictions using the time series data with any kind of patterns and with

ISSN:0971-1260
Vol-22-Issue-37-December -2019

auto correlated consecutive values of the time series. The ARIMA (1, 0, 0) model projected decrease in the oilseed production for coming 5 years from 2019 to 2023.

References:

[1]. G.E.Box and G.M. Jenkins, Time Series Analysis: Forecasting and Control. Holden Day, San Francisco, 1976.

[2]. D.A. Dickey and W.A. Fuller,“Distribution of estimators for Autoregressive Time Series with a Unitroot”. Journal of the American Statistical Association Vol. 74, pp. 427-431, 1979.

[3]. J.G.D. Gooijer and R.J. Hyndman,“25 years of time series forecasting”. International Journal of Forecasting,vol.22, pp. 443–473, 2006.

[4]. http://eands.dacnet.nic.in/latest_20011.htm

[5]. http://salasargroup.com/commodities/oil-seeds/

[6]. https://m.rbi.org.in/Scripts/PublicationsView.aspx?id=19331

[7]. https://www.reportlinker.com/data/series/0PPawmn6Hlc

[8]. P.K. Sahu,“Forecasting production of major food crops in four major SAARC countries”. International Journal of Agricultural and Statistics Science vol.10, pp.71-92, 2010.

[9]. M. Shafaqat, “Forecasting Pakistan exports to SAARC - An application of univariate ARIMA model”. Journal of Contemporary Issues in Business Research, vol.1, pp.41-54. 2012.

[10]. A. Sonawane, M. Hasan, Y. Rajwade, S. Desai, G. Rajurkar, V. Shinde, D. Singh and M. Singh,“Comparison of Neuro-Fuzzy and Regression Models for Prediction of Outflow of on-farm Reservoir”. International Journal of Agriculture, Environment and Biotechnology, vol.6, pp. 311-316. 2013

