Safflower Yield Forecasting in India – An application of Auto Regressive Integrated Moving Average (ARIMA) Model

K.Alivelu, P.Padmavathi, C.Sarada, P.Lakshmamma and M. Santha Lakshmi

ICAR-IIOR, Rajendranagar, Hyderabad, Telangana, India

**Abstract**

The present study was carried out for forecasting the safflower productivity of India by fitting univariate Auto Regressive Integrated Moving Average (ARIMA) models. The data on safflower yield collected from 1965-66 to 2013-14 has been used for present study. The order of an ARIMA model is usually denoted by the notation ARIMA (*p,d,q*), where pis the order of the autoregressive part d is the order of the differencing qis the order of the moving-average process. For different values of p and q, various ARIMA models were fitted and appropriate model was chosen corresponding to minimum value of Akaike information criteria (AIC), Schwarz-Bayesian information criteria (SBC). ARIMA (1, 1, 2) model was found suitable for all India safflower yield with minimum MAPE (5.4).

Key words: Safflower, ARIMA model, Yield.

**Introduction**

Safflower (*Carthamus tinctorius* L.), commonly known as kardi is one of the important *rabi* oilseed crops of the country. India is in first place in terms of area and production in the world with an area of 1.5 lakh ha and production of 1.13 lakh tonnes with a yield of 638 kg/ha. (2013-14). Safflower is mainly grown in Maharashtra, Karnataka to some extent in Gujarat and parts of Andhra Pradesh, Madhya Pradesh, Orissa, Bihar etc.Forecasting of crop yield based on time series data is an important task and facilitates efficient planning of cropping systems. A time series is a collection of observations of well-defined data items obtained through repeated measurements over time. The main objectives of time series analysis are to develop a model and estimate the parameters and forecast the future values of time series.

ARIMA models have been used for forecasting rice productivity and production of Odessa ([Rahul Tripathi](http://www.hindawi.com/13410876/), 2014), Sugarcane yield of Tamilnadu (Suresh, 2011), Forecasting of irrigated crops like potato, mustard and wheat were forecasted using ARIMA models (Sahu 2006). Apart from agricultural crops milk production in India was forecasted using time-series modelling techniques (Pal et al. 2007). The objective of our present study was using ARIMA models developed by Box and Jenkins (1976) to forecast safflower yield of India.

**Materials and Methods**

Time series data covering the period of 1965-2014 was used for the Study. The data were collected from Indiaagristat. One of time series models which is popular and mostly used is ARIMA model. ARIMA (p, d, q) model is a mixture of Autoregressive (AR) and Moving average (MA) model . The general form of ARIMA (p, d, q) described by Judge et al (1988). ARIMA methodology is categorized in to identification, estimation, diagnostic checking and forecasting stages.

**Identification**

The first step in applying ARIMA model is to check for stationarity i.e. series remains constant level over time. The method to check for stationarity is to plot the data and check the autocorrelation function. If the graph of ACF (Auto Correlation Function) cuts of or dies down quickly, the series is stationary. The non-stationary series made stationary by differencing the data series.

This is done by subtracting the observation in the current period from the previous one. If this transformation is done only once to a series it is said to be first differencing i.e d=1. This process essentially eliminates the trend if the series is growing at a fairly constant rate. If it is growing at an increasing rate, the series has to be differenced again. Another important procedure in identification stage is to decide the values of p and q. They can be estimated by observing the graphs of ACF and PACF (Pindyk and Rubinfeld, 1991).

**Estimation, diagnostic checking and forecasting**

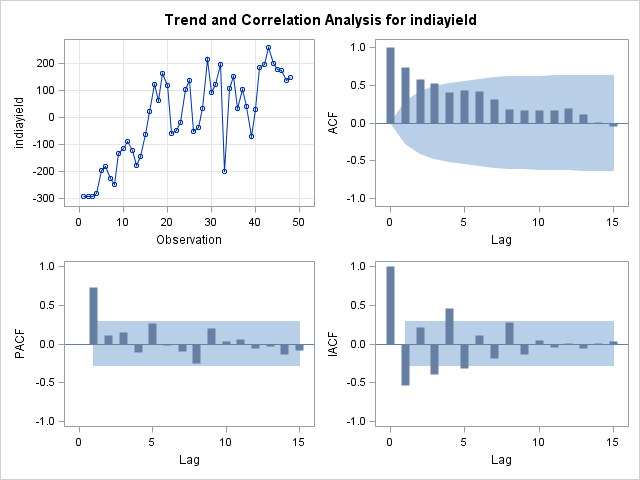
The model was estimated using SAS 9.3 software with PROC ARIMA procedure. The first check was by plotting the ACF of residuals of the fitted model. When the graph shows no trend with rectangular scatter around a zero horizontal level then the model was best fitted model. Second check was straight line graph when normal scores were plotted against residuals. Another check was residuals were plotted against fitted values the graph should be having no pattern. Lowest Akaike Information Criteria (AIC) also used to select the best model. Mean Absolute Percentage Error (MAPE) was used as measure of accuracy of the models. Using the best model, forecasts were made.

**Results and Discussion**

The fig1. Shows the trend, ACF and PACF of safflower yield in India over 49 years. From the figure it is evident that the ACF declines very slowly and after the first lag the PACF drops and all lags after that are statistically in significant. It indicates the non stationarity in the mean yield The table1. results concluded that the highly significant chi square value indicating the absence of white noise (uncorrelated) of the series.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table1. Autocorrelation- Check for White Noise** | | | | | | | | | |
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 88.55 | 6 | <.0001 | 0.729 | 0.583 | 0.527 | 0.398 | 0.432 | 0.413 |
| 12 | 103.74 | 12 | <.0001 | 0.309 | 0.179 | 0.166 | 0.167 | 0.169 | 0.189 |

**Fig.1 Trend, ACF and PACF of safflower yield in India**



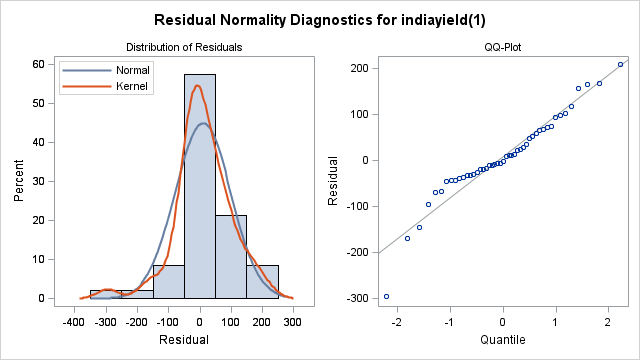
So the non stationary series made to stationary by differencing technique. Differencing changes the variable under consideration (Chatfield, 1975 and Cressie(1988). PAC function of the first differenced series was used to determine p to be 1. Several tentative models were fitted with different q values and the best fitted model is selected based on the minimum AIC, SBC and MAPE. ARIMA (1, 1, and 2) model was found suitable for all India safflower yield with minimum MAPE (5.4). After fitting the best model, residuals were tested for autocorrelation analysis . By observing the insignificant values, it can be concluded that the fitted model is a good fit.

On the basis of fitted model the mathematical model obtained as

yt = 0.562 yt-1 + εt + 0.105 εt-1 – 0.755 εt-2

Normality test was done by plotting the histogram of residuals indicating the best fit of the model. ARIMA(1, 1, 2) was taken for 30 years ahead forecasts for safflower India yield which are given in table3 with standard error and upper lower confidence limits. Forecasts of safflower yield showed an increasing trend from 690 kg/ha in 2015 to 831 kg/ha. The area under safflower is coming down from 2.79 lakh ha during 11th plan to 1.78 lakh ha during 2013-14. Safflower cultivation in non-traditional areas like rice fallows may add additional area under safflower and following good production practices and implementation of technology can achieve increasing trend in yield.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table2. Autocorrelation Check of Residuals | | | | | | | | | |
| To Lag | Chi-Square | DF | Pr > ChiSq | Autocorrelations | | | | | |
| 6 | 3.73 | 3 | 0.29 | 0.03 | 0.06 | -0.07 | -0.17 | 0.03 | 0.17 |
| 12 | 7.57 | 9 | 0.58 | 0.05 | -0.08 | -0.06 | -0.02 | 0.00 | 0.22 |
| 18 | 11.58 | 15 | 0.71 | 0.09 | 0.01 | -0.03 | -0.19 | -0.09 | 0.01 |
| 24 | 15.83 | 21 | 0.78 | -0.02 | -0.08 | -0.13 | -0.16 | -0.01 | 0.01 |

**Fig2. Normality test for residuals** 

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 3. Forecast of safflower yield** | | | | | | |
| Year | Forecast | Std Error | 95% Confidence Limits | | | |
| 2012 | 662.13 | 119.992 | 426.95 | 897.311 | ±235.18 |  |
| 2013 | 671.535 | 120.52 | 435.321 | 907.749 | ±236.21 |  |
| 2014 | 680.939 | 121.045 | 443.696 | 918.182 | ±237.24 |  |
| 2015 | 690.343 | 121.568 | 452.075 | 928.611 | ±238.26 |  |
| 2016 | 699.747 | 122.088 | 460.459 | 939.036 | ±239.28 |  |
| 2017 | 709.152 | 122.607 | 468.847 | 949.456 | ±240.30 |  |
| 2018 | 718.556 | 123.123 | 477.239 | 959.872 | ±241.31 |  |
| 2019 | 727.96 | 123.637 | 485.636 | 970.284 | ±242.32 |  |
| 2020 | 737.364 | 124.149 | 494.037 | 980.692 | ±243.32 |  |
| 2021 | 746.769 | 124.659 | 502.442 | 991.096 | ±244.32 |  |
| 2022 | 756.173 | 125.167 | 510.851 | 1001.5 | ±245.32 |  |
| 2023 | 765.577 | 125.672 | 519.264 | 1011.89 | ±246.3133 |  |
| 2024 | 774.981 | 126.176 | 527.681 | 1022.28 | ±247.3006 |  |
| 2025 | 784.386 | 126.678 | 536.102 | 1032.67 | ±248.2839 |  |
| 2026 | 793.79 | 127.178 | 544.527 | 1043.05 | ±249.2634 |  |
| 2027 | 803.194 | 127.675 | 552.955 | 1053.43 | ±250.2389 |  |
| 2028 | 812.598 | 128.171 | 561.388 | 1063.81 | ±251.2108 |  |
| 2029 | 822.003 | 128.665 | 569.824 | 1074.18 | ±252.1789 |  |
| 2030 | 831.407 | 129.157 | 578.264 | 1084.55 | ±253.1432 |  |

**References**

Box G E P and Jenkins G M 1976. Time series analysis. Forecasting and control. San Francisco, Holden-Day, California, USA.

Chatfield C 1975. The analysis of time series: Theory and Practice. Chapman and Hall, London.

Cressie N 1988. A graphical procedure for determining non-stationary in time series. JAS.83:1018-15.

Judge G G R Carter Hill, Willium E G and I Helmut I 1988. Introduction to the theory and practice of econometrics 2 nd ed., John Wiley and Son, INC, New York, Toronto, Singapore.

Pal Satya, Ramasubramanian, V and Mehta S C 2007. Statistical models for forecasting milk production in India. Journal of Indian Society of Agricultural Statistics. 61(2): 80-83.

Pindyk, R S Daniel and Rubinfed L 1991. Economic models and economic forecasts. 3 rd ed. McGraw Hill International editions (Economic Survey) New York, USA.

Rahul Tripathi Nayak A K Raja R 2014 “Forecasting Rice Productivity and Production of Odisha, India, Using Autoregressive Integrated Moving Average Models,” Advances in Agriculture, vol. 2014, Article ID 621313, 9 pages,

Suresh K K and Priya S R K 2011. Forecasting sugarcane yield of Tamilnadu using ARIMA models, Sugar Tech, vol. 13, no. 1, pp. 23–26.