Original Research Article

Forecasting of Area, Yield and Production of Sunflower in Odisha

Abhiram Dash1*, R.K. Patnaik2, S.K. Dwibedi3, B.S. Bishoyi4 and P. Shatpathy5

1Department of Agril. Statistics, College of Agriculture, Chiplima, India
2College of Agriculture, Chiplima, India
3Department of Agronomy, College of Agriculture, Bhubaneswar, India
4Department of Agronomy, College of Agriculture, Chiplima, India
5Department of Plant Physiology, College of Agriculture, Chiplima, India

*Corresponding author

A B S T R A C T

Forecasting of sunflower production is of utmost importance in state of Odisha. Forecasted values of area and yield of sunflower are used to forecast production of sunflower. Forecasting done by the help of best fit regression model is quite reliable. Selecting the appropriate models that would fit the data is done through the help of scatter plot of the data. The model to be considered for best fit has to satisfy the error assumptions and must have all the estimated coefficients significant. Among the models satisfying the error assumptions and having all the estimated coefficients significant, the one having highest value of R², highest value of adjusted R² and lowest value of RMSE is considered to be the best fit model. This best fit model should be used for forecasting. The compound model and quadratic model are found to be the best fit models for forecasting area and yield of sunflower in Odisha.

Keywords
Forecasting of area, Yield, Production

Accepted: 07 October 2018
Available Online: 10 November 2018

Introduction

The central government through Rastriya Krishi Vikash Yojana has been promoting the cultivation of oil seeds on a public. The impressive fatty acid content in sunflower oil helps in maintaining a balance in our body. Furthermore, it does not contain any saturated fats which help in controlling the cholesterol levels in your body. Sunflower oil is rich in Vitamin A and Vitamin E which help in promoting skin health. Forecasting of sunflower production is of utmost importance to meet the need of healthy edible oil in odisha. Forecasting of production of sunflower seeds is done by using the forecasted values of area and yield of sunflower. Forecasting through best fit regression model are quite reliable.

Materials and Methods

The analysis is based on the secondary data on area and yield of sunflower crop of Odisha for the period from 1975-76 to 2015-16. The data are obtained from various volumes of Odisha Agricultural Statistics published by the Directorate of Economics and Statistics, Government of Odisha. The data for the period from 1975-76 to 2007-08 are used to
build model for forecasting future values regarding area and yield of sunflower crop in Odisha. The models found appropriate to fit the data on area and yield of sunflower crop in Odisha are selected from the scatter plot of the data. In all the selected models used for the present study time is considered to be the independent variable.

The parametric growth models, both linear and non-linear models are employed (Montgomery et al., (2001); Ratkowsky (1990); Draper and Smith (1998)). The following models are used to fit the data on area and yield of sunflower crop in Odisha: linear model (ii) power model (iii) compound model (iv) logarithmic model (v) quadratic model (vi) cubic model (polynomial model of degree two) and (vii) cubic model (polynomial model of degree three).

In all the models $Y_t$ is the value of the variable in time $t$, $\beta_0$ and $\beta_1$ are the parameters of the models used in the study and $\varepsilon_t$ is the random error component. Brief descriptions of different models are given below:

**Linear model**

$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$

**Power model**

$Y_t = \beta_0 t^{\beta_1} \exp(\varepsilon_t)$.
The form of power model after logarithmic transformation is $\ln(Y_t) = \ln(\beta_0) + \beta_1 \ln(t) + \varepsilon_t$

**Compound model**

$Y_t = \beta_0. \beta_1^t. \exp(\varepsilon_t)$.
The form of the compound model after logarithm transformation is $\ln(Y_t) = \ln(\beta_0) + \ln(\beta_1). t + \varepsilon_t$

**Logarithmic model**

$Y_t = \beta_0 + \beta_1. \ln(t) + \varepsilon_t$

**Quadratic model**

$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$, where $\beta_2$ is the parameter of the model.

**Cubic model**

$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \varepsilon_t$, where $\beta_2$ and $\beta_3$ are the parameters of the model.

Using Ordinary Least Squares technique, the estimated values of the coefficients $\beta_0, \beta_1, \beta_2$ and $\beta_3$ are obtained for the fitted models. The estimated values of $\beta_0, \beta_1, \beta_2$ and $\beta_3$ are written as $b_0, b_1, b_2$ and $b_3$ respectively.

The significance of the coefficients of the model and the significance of the model as a whole have been tested by the following procedure:

To test the significance of the coefficients of the model using their respective estimate $t$-tests with appropriate degrees of freedom have been performed. The appropriate test statistic is $t = \frac{b_j}{SE(b_j)}$ which follows a ‘t’ distribution with $(n - p)$ degrees of freedom, where ‘n’ is the number of observations and ’p’ is the number of parameters involved in the model, $SE(b_j)$ is the standard error of $b_j$, $j = 0, 1, 2$.

If absolute value of Calculated value of $t \geq$ tabulated $t_\alpha$ at $n-2$ d.f. then $H_0$ is rejected, otherwise accepted.

In order to carry out the above tests we have to assume that errors follow normal distribution with constant variance and are independently distributed.

Here we have considered the following statistical tests for testing the assumptions regarding errors in the model:
Durbin-Watson test for testing independence of residuals.

Park’s test for testing homoscedasticity of residuals.

Shapiro-Wilk’s test for testing normality of residuals.

The model to be considered for selection must satisfy the assumptions regarding the errors, have overall significance, and the parameters are to be significant (Dash et al., 2017).

Next the model fit statistics, viz., $R^2$ and RMSE are computed for the purpose of model selection.

Among the models fitted for the dependent variable, the model which has highest $R^2$, highest adjusted $R^2$ and lowest RMSE is considered to be the best fit model for that variable.

$$\frac{SSM}{SSE}$$

Note that, $R^2 = \frac{SSM}{SSE}$, where, SSM is the sum of square due to model; SSE is the sum of square due to error.

$$SSM = \sum_{i=1}^{n}(\hat{y}_t - \bar{y})^2, \quad SSE = \sum_{i=1}^{n}(y_t - \hat{y}_t)^2,$$

Where $y_t$ and $\hat{y}_t$ are respectively the actual and estimated values of the response variable at time $t$, $\bar{y}$ is the mean of $y_t$.

Root Mean Square Error is defined as RMSE

$$\sqrt{\frac{\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}{(n-p)}},$$

The data for the period from 2008-09 to 2015-16 are held up for validation of the selected model and thus not used for model building. After successful cross validation of the selected best fit model, the forecast values of area and yield of sunflower crop in Odisha are obtained for the years from 2016-17 to 2020-21. Using the forecast values of area and yield, the forecast values of production of sunflower for the future years i.e. 2016-17 to 2020-21 are obtained in the following manner: Forecast value of production (in ‘000 tonnes) = Forecast value of area (in’000 ha) x Forecast value of yield (in kg/ha).

**Results and Discussion**

From study of the data using scatter diagram shown in figures 1 and 2, some models (such as, linear, power, compound, logarithmic, quadratic and cubic models) are tried to find the appropriate model which can capture the behaviour of the data very closely for the purpose of forecasting.

From the study of Table 1, it is found that in case of the models fitted for area under sunflower crops in Odisha, among all the fitted models the compound model is found to satisfy the independency of errors, homoscedasticity of errors, normality of errors. Also the estimated coefficients of the compound model are significant. Also the value of $R^2$ for compound model is quite high and RMSE of compound model is quite low.

Thus the compound model is considered to be the best fit model for data on area under sunflower crop in Odisha. Though the value of $R^2$ for cubic model fitted to data on area under sunflower crop in Odisha is higher than that of compound model and RMSE of cubic model is lower than that of compound model, the compound model is selected to be the best model because the cubic model has insignificant estimated coefficients.
Table 1 Estimated coefficients, model fit statistics and model diagnostics criteria of the models fitted to area under sunflower crop in Odisha

| Model (Area) | Coefficient estimates | Model fit statistics | Model diagnostics criteria |
|--------------|------------------------|----------------------|----------------------------|
|              | $b_0$ | $b_1$ | $b_2$ | $b_3$ | $R^2$ | RMSE | MAPE | Shapiro-Wilk’s Statistic | Durbin-Watson Statistic | Coefficient of $\ln(t)$ in Park’s test of heteroscedasticity |
| Linear       | -1.195 (0.237) | 0.345 (0.001) | - | - | 0.597 | 2.738 | 70.132 | 0.912 | 0.865* | 0.222 |
| Compound     | 0.812 (0.001) | 1.083 (0.001) | - | - | 0.703 | 2.457 | 46.368 | 0.902 | 1.826 | 1.751 |
| Quadratic    | 2.09 (0.136) | -0.218 (0.247) | 0.017 (0.004) | - | 0.697 | 2.413 | 48.586 | 0.911 | 1.097* | 2.046 |
| Cubic        | -0.201 (0.915) | 0.535 (0.261) | -0.038 (0.241) | 0.01 (0.09) | 0.712 | 2.333 | 59.847 | 0.912 | 1.173* | 1.496 |
| Power        | 0.444 (0.011) | 0.759 (0.001) | - | - | 0.604 | 3.3.85 | 59.951 | 0.921 | 0.577** | 3.314* |
| Logarithmic  | -3.297 (0.098) | 3.09 (0.001) | - | - | 0.378 | 3.403 | 101.68 8 | 0.923 | 0.601** | 0.045 |

Table 2 Estimated coefficients, model fit statistics and model diagnostics criteria of the models fitted to yield of sunflower crop in Odisha

| Models (Yield) | Coefficient estimates | Model fit statistics | Model diagnostics criteria |
|----------------|------------------------|----------------------|----------------------------|
|                | $b_0$ | $b_1$ | $b_2$ | $b_3$ | $R^2$ | RMSE | MAPE | Shapiro-Wilk’s Statistic | Durbin-Watson Statistic | Coefficient of $\ln(t)$ in Park’s test of heteroscedasticity |
| Linear         | 367.19 (0.001) | 13.17 (0.001) | - | - | 0.631 | 97.328 | 15.736 | 0.889* | 1.078 | -0.327 |
| Compound       | 397.90 (0.001) | 1.022 (0.001) | - | - | 0.625 | 126.48 | 20.549 | 0.879* | 0.683 | 0.079 |
| Quadratic      | 503.13 (0.001) | -10.134 (0.001) | 0.685 (0.001) | - | 0.755 | 80.614 | 11.719 | 0.901 | 1.578 | 0.717 |
| Cubic          | 465.89 (0.001) | 2.106 (0.897) | -0.201 (0.856) | 0.017 (0.42) | 0.670 | 81.017 | 11.96d | 0.912 | 1.612 | 0.976 |
| Power          | 352.31 (0.001) | 0.188 (0.001) | - | - | 0.383 | 92.136 | 14.592 | 0.887* | 1.195 | -0.105 |
| Logarithmic    | 295.47 (0.001) | 114.697 (0.001) | - | - | 0.377 | 122.77 | 19.704 | 0.872* | 0.705 | 0.446 |
Table 3: Cross validation of the selected model for forecasting area and yield of sunflower

| Year  | Area (Compound model) | Yield (Quadratic model) |
|-------|-----------------------|-------------------------|
|       | Actual Value (in '000 ha) | Predicted Value (in '000 ha) | Absolute Percentage Error | Actual Value (in kg/ha) | Predicted Value (in kg/ha) | Absolute Percentage Error |
| 2008-09 | 14.85 | 13.17 | 11.32 | 925.25 | 950.93 | 2.78 |
| 2009-10 | 17.73 | 15.18 | 14.44 | 901.86 | 988.09 | 9.56 |
| 2010-11 | 20.63 | 18.27 | 11.44 | 1047.99 | 1026.62 | 2.04 |
| 2011-12 | 23.69 | 21.45 | 9.45 | 1180.67 | 1066.52 | 9.67 |
| 2012-13 | 26.32 | 26.73 | 1.57 | 1189.97 | 1107.79 | 6.91 |
| 2013-14 | 24.88 | 26.12 | 4.98 | 1193.33 | 1150.44 | 3.59 |
| 2014-15 | 21.69 | 19.62 | 9.54 | 1195.02 | 1194.45 | 0.05 |
| 2015-16 | 16.03 | 18.25 | 13.83 | 1189.02 | 1239.83 | 4.27 |

MAPE 9.565 MAPE 4.858

Table 4: Forecast values of area and yield of sunflower in Odisha obtained from the selected regression model and forecast values of production of sunflower obtained from the forecast values of area and yield

| Year  | Area (in '000 ha) | Yield (in kg/ha) | Production (in '000 tonnes) |
|-------|-------------------|------------------|-------------------------------|
| 2016-17 | 23.01 | 1226.59 | 28.22 |
| 2017-18 | 24.91 | 1234.72 | 30.76 |
| 2018-19 | 25.98 | 1284.21 | 33.36 |
| 2019-20 | 26.21 | 1335.08 | 34.99 |
| 2020-21 | 27.63 | 1387.32 | 38.33 |

Figure 1: Scatter plot of area under sunflower in Odisha
Figure 2: Scatter plot of yield of sunflower in Odisha
Figure 3: Actual and forecasted values of yield of sunflower in Odisha
Figure 4: Actual and forecasted values of area under sunflower in India
The study of Table 2 reveals that in case of the models fitted for yield of sunflower crops in Odisha, among all the fitted models the quadratic model is found to satisfy the independency of errors, homoscedasticity of errors, normality of errors. Also the estimated coefficients of the quadratic model are significant.

Also the adjusted $R^2$ for quadratic model is the highest and RMSE of quadratic model is the lowest among all the models fitted to yield of sunflower crop in Odisha. Thus the quadratic model is considered to be the best fit model for data on area under sunflower crop in Odisha.

The result of cross-validation test of the selected best fit model is given in the Table 3. The study of the table reveals that the estimated values of area for the period of cross-validation (i.e. 2008-09 to 2015-16) are also very close to the actual values with APE less than 10% and MAPE just less than 10%.

The estimated values of yield for the period of cross-validation are very close to the actual values with APE less than 15% and MAPE just less than 5%. This shows that the selected best fit models for area and yield of sunflower in Odisha gives good estimate values of the respective variables.

Using the best fit models for area and yield of sunflower crop in Odisha, the forecast values of area and yield of sunflower crop in Odisha are obtained for the future years i.e. 2016-17 to 2020-21.

Using the forecast values of area and yield, the forecast values of production of sunflower crop is obtained for the future years. The forecast values of area, yield and production of sunflower in odisha for the period from 2016-17 to 2020-21 are presented in Table 4. Figures 3 and 4 shows the actual and forecasted values of area and yield of sunflower in Odisha, respectively.

The models found appropriate to fit the data on area and yield of sunflower in Odisha are selected from the scatter plot of the data. Various model selection criteria like $R^2$, adjusted $R^2$ and RMSE and various model diagnostic criteria like Shapiro-Wilk’s test, Park’s test and Durbin-Watson test are used to select the best fit model.

After successful cross-validation of the selected best fit model, the forecast values of area and yield of sunflower in Odisha for the future years are obtained by using the best fit model. The forecast values of production of sunflower in odisha are obtained from the forecast values of area and yield.

The forecast values of area, yield and production of sunflower in odisha for future years shows an increasing trend.

References

Dash, A., D.S. Dhakre and Bhattacharya, D. 2017. Analysis of Growth and Instability in Rice Production of Odisha by Fitting Appropriate Statistical Models. Int.J.Curr.Microbiol.App.Sci. 6(10): 3313-3322.

Draper, N.R. and Smith, H. (1998): Applied Regression Analysis, 3rd Edition, John Wiley and Sons, New York, USA.

Montgomery, D. C., Peck, E. A. and Vining, G. G. (2001): Introduction to Linear Regression Analysis, 3rd Edition, New York, John Wiley & Sons, USA.

Ratkowsky, D.A. (1990): Handbook of Non-linear Regression Models, Marcel Dekker, New York.

How to cite this article:

Abhiram Dash, R.K. Patnaik, S.K. Dwibedi, B.S. Bishoyi and Shatpathy, P. 2018. Forecasting of Area, Yield and Production of Sunflower in Odisha. Int.J.Curr.Microbiol.App.Sci. 7(11): 618-623. doi: https://doi.org/10.20546/ijcmas.2018.711.075