Non-linear statistical modeling for area, production and productivity of rice crop in Chhattisgarh


Introduction

Rice (oryza sativa L.) is the world’s most important food crop after wheat. It is extensively grown in humid tropical and subtropical region of the world. India has the largest area under rice. In India, the total food grain production for the year 2010-11 was 241 million tones, out of which, rice contributed about 95.32 million tones. Therefore, it needs to be increased by improved cultivars and agro-techniques. In India, where a lot of agro-climatic variations exists, forecast of long term production possibilities at the national level are of little values as a guide for detailed planning. To help macro level planning, in this research, it is proposed to develop nonlinear regression models for area, production and productivity of rice for three agro-climatic zones of Chhattisgarh viz., Chhattisgarh plains, Northern hills and Bastar plateau.
Resource and Research Methods

Linear and nonlinear models:

A linear model is one which is directly proportional to input. In such a model, all the parameters appear linearly. In contrast to this, nonlinear model is one in which at least one of the parameters appears nonlinearly. For example, following equations:

\[ Y(t) = A + Bt \]  \hspace{1cm} \ldots (i)
\[ Y(t) = A + Bt + Ct^2 \]  \hspace{1cm} \ldots (ii)

represents a linear model, where as equation:

\[ Y(t) = Y(0) \exp(-At) \]  \hspace{1cm} \ldots (iii)

represents a nonlinear model. A consequence of nonlinearity is that it has curved solution locus, whereas linear model has straight line solution locus. We consider two types of nonlinear models given by following equation, respectively:

\[ Y(t) = \exp(A + Bt^2) \]  \hspace{1cm} \ldots (iv)
\[ Y(t) = \exp(-Bt) - \exp(-At) \]  \hspace{1cm} \ldots (v)

Although, both these models are nonlinear as the parameters A and B appear nonlinearly but they are of essentially different character. Eq. (iv) can be transformed, by taking natural algorithm, into the form:

\[ \ln Y(t) = A + Bt^2 \]

which is linear in parameters. The model given in Eq. (iv) is called intrinsically linear since it can be transformed into linear form. Eq. (iii) however, cannot be converted into a form which is linear in parameters. Such a model is said to be intrinsically nonlinear (Draper and Smith, 1981).

Some important nonlinear growth models:

Logistic model:

This model postulates that, in the beginning, growth take an exponential rate, as in the well known Malthusian model. Subsequently, a “deterrent force” comes into play. The graph of this model is elongated S-shaped and has a point of inflexion at half the carrying capacity.

Let \( y(t) \) denote the variable under study like area, production and productivity of rice at time \( t \). Also, let \( \alpha > 0 \) denote the intrinsic growth rate and \( \lambda \) the carrying capacity of the environment. Then this model is represented by:

\[ y(t) = \frac{\lambda}{1 + \beta \exp(-\alpha t)} \]

where,

\[ \beta = \frac{[\lambda - y(0)]/y(0)}{and \ y(0)} \text{is the value of y} \ (\text{at} \ t = 0). \]

Gompertz model:

This model has sigmoid type of behavior and is found quite useful full in the biological work. However, unlike logistic model, this is not symmetric about its point of inflexion. This model is given by:

\[ y(t) = \frac{\lambda}{1 + \beta \exp(-\alpha t)} \]

where symbols have their usual meaning.

Monomolecular model:

This model describes the progress of a growth situation in which it is believed that the rate of growth at any time is proportional to the resources yet to be achieved by the equation. This model is represented by the equation:

\[ y(t) = \lambda - (\lambda - \beta) \exp(-\alpha t) \]

where symbols have their usual meaning.

Richard’s model:

It is a four parameter growth model. It is proposed by Richards (1959) and is represented by:

\[ y(t) = \frac{\lambda}{[1 + \beta \exp(-\alpha t)]^{\delta}} \]

The upper sign within the brackets is applicable when \( \delta \) is positive and lower sign when \( \delta \) lies in the range \(-1 < \delta < 0\). Richards model is a generalization of logistic (when \( \delta = 0 \)), Gompertz (when \( \delta = 0 \)) and monomolecular (when \( \delta = -1 \)).

MMF Model:

Another four parameter model is MMF (Morgan Mercer Florin) model. This model is represented by:

\[ y(t) = (\alpha \beta \lambda y^b)/(\beta + y^b) \]

Examination of residuals:

After filling the non-linear statistical model, the next important step is to see whether the assumptions made regarding the error term are valid or not. This is being done by examining residuals. It is evident from the definition that residuals are the differences between what is actually observed, and what is predicted by the model i.e. the amount which is unexplained by the model. If our fitted model is correct, the residuals should exhibit tendencies that tend to confirm the assumptions we have made, or at least, should not exhibit denial of the assumptions.

Various goodness of fit of nonlinear statistical models:

Co-efficient of determination \((R^2)\):

The goodness of fit is examined by using the co-efficient of determination \((R^2)\). Kvalseth (1985) examined the different forms of \(R^2\) available in the literature. Eight different forms of \(R^2\) have been mentioned in his paper. One of the main conclusion of the paper is that

\[ R^2 = 1 - \text{residual SS/ correctedSS} \]

\ldots (1)

is the most appropriate for nonlinear statistical model.
This $R^2$ would be used as the coefficient of determination for goodness of fit. The potential range of values of this $R^2$ is well defined with end points corresponding to perfect fit and complete lack of fit, such as $0 < R^2 < 1$, where $R^2 = 1$ corresponds to perfect fit and $R^2 = 0$ for any reasonable model specification. For nonlinear models its value can be negative, if the selected model fits worse than the mean.

Root mean square error (RMSE):
The RMSE is defined as
\[
RMSE = \sqrt{\frac{\sum(Y - \bar{Y})^2}{n}} .......(2)
\]
The smaller the value of RMSE, better the model

Mean absolute error (MAE):
The MAE is defined as
\[
MAE = \frac{\sum|Y - \bar{Y}|}{n} .......(3)
\]
The smaller the value of MAE, better is the model.

Mean square error (MSE):
The MSE is defined as
\[
MSE = \frac{\sum(Y - \bar{Y})^2}{(n-p)}
\]

Correlation co-efficient between observed and predicted values [$r(Y, \bar{Y})$]:
It is the simple product moment correlation coefficient between observed and predicted values. The greater the value, the better is the model.

One step ahead forecasting (OSAF):
In OSAF method the last observation is not considered and the model is fitted to the data set. The last value is predicted from the model and it is compared with the actual value. The percent forecast error is defined as:
\[
PCFE = \frac{|Y(t) - \bar{Y}(t)|}{Y(t)} \times 100
\]
where, $Y(t)$ is the observed value and is the predicted value. The smaller the value of PCFE, better is the model.

Collection of data:
State level time series data pertaining to area, production and productivity of rice for the period of 25 years i.e. 1975-76 to 2000-01 were collected from various secondary sources. The required data for the study were collected from the Agricultural statistics, published by the Directorate of Agriculture, Government of M.P., Bhopal and the Basic Agricultural statistics published by the commissioner, land records and settlement Gwalior, Madhya Pradesh. From Chhattisgarh at a glance, published by the Directorate of Economics and Statistics, Government of C.G., Raipur and Krishi Sankhyaiki Sarni, published by the Commissioner, land records and settlement.

Research Findings and Discussion

There three- parameter models viz., monomolecular, logistic and Gompertz and two four-parameter viz., Morgan mercer florin (MMF) and Richards model were considered for studying rice production in Chhattisgarh State and also for developing an appropriate nonlinear statistical growth model which would be able to describe the path of rice production and also to forecast future production. Initial estimates of the parameters were obtained by different methods. Many sets of initial parameter values were tried to ensure convergence as well as stability of the parameter estimates.

Goodness of fit statistics viz., root mean square error (RMSE), root absolute error (MAE), mean square error (MSE) and correlation $r(Y, \bar{Y})$ were computed using MS-Excel. In case of one step ahead forecasting (OSAF), the set of initial parameter values for which the iterative procedure converged, was used to fit the model. The results of fitting various models in three agro-climatic zones viz., Chhattisgarh plain, Bastar plateau and Northern hills region and whole Chhattisgarh State. Normal Q-Q plot of residuals of best fitted models for all agro climatic zones and over all state.

Nonlinear area under rice:
All the five models fitted well to the data. Examination of residuals using run test and Shapiro-Wilk test showed that residuals were random and the normality assumptions not violated in all the models. As not much variation was observed among the models in respect to $R^2$, RMSE, MSE, MAE and $r(Y, \bar{Y})$. Though MMF model had highest $R^2$ value, but the per cent forecast efficiency was not good. As far as Richards’s model is concerned, there was no gain in goodness of fit statistics. Hence, monomolecular, logistic and Gompertz model may be considered for explain the area of Chhattisgarh plain.

In this case all five models fitted well to the data. Examination of residuals reveals that there was no evidence against the assumptions of randomness and normality of residuals for all the five models. Comparing three two-parameter models viz., monomolecular, logistic and Gompertz to two four-parameter models viz., Richards and MMF model, we found that there was no gain in using fourth parameter. $R^2$ values were same for all the models and MSE was more in case of four parameter models, which established the fact, that three-parameter models may be appropriate. Therefore, we can consider either monomolecular or logistic or Gompertz models for studying the area data of rice for Bastar plateau.

Both four parameters models viz., Richards and MMF model and three parameter models; monomolecular, logistic and Gompertz fitted very well to these data sets. Examination of residuals reveals that there was no evidence against the assumptions of randomness and normality of residuals. The goodness of fit statistics reveals that there was virtually no difference among the models as far as $R^2$, RMSE, MAE $r(Y, \bar{Y})$
and rare concerned. However, MSE was marginally high in case of Gompertz and MMF model compared to monomolecular, logistic and Richards’s model. It implies that the fourth parameter in the MMF model and third parameter in the Gompertz model do not help to improve the fit. Also gain was not substantial in case of four parameter model i.e. Richards model compared to three parameter models. Hence, either monomolecular or logistic model can be used for describing the rice area data of Northern hills.

Except, logistic model the four remaining models viz., monomolecular, Gompertz, Richards and MMF fitted well to the data. There was no evidence to reject the assumptions of randomness and normality of residuals for these models. As, there was no substantial difference in other goodness of fit statistics for the four models. Comparing two three parameter models viz., monomolecular and Gompertz models to two four parameter models i.e. Richards and MMF models, we found that there was no gain in using the fourth parameter. Therefore, we considered either Gompertz or monomolecular model for studying the rice area data of Chhattisgarh State as a whole.

Nonlinear statistical models for studying production of rice:

All the models fitted reasonably well to the production data of rice for Chhattisgarh plains. Run test indicated randomness of residuals and Shapiro-wilk test reveals that the normality assumption not violated for all models. R² value for MMF was low as compared to rest of the models. As far as Richards’ model is concerned the fourth parameter do not help to improve the fit as evidenced by R² and MSE. Hence, we abandoned Richards’s model and further studied only the three models each of which has three parameter comparison of various goodness of fit statistics viz., R², RMSE, MAE and PCFE showed that there was only negligible difference among these models. As a result, at this stage it was not possible to single out an appropriate model for these data sets. We can consider any one out of these three models viz., logistic, Gompertz and monomolecular for describe the data sets regarding production data of Chhattisgarh plains.

All the five models viz., monomolecular, logistic, gompertz, Richards and MMF fitted well to the data. However, RMSE and MAE were marginally high in case of Richards and MMF model. We find that there was no significant difference in the values of R² and other goodness of fit statistics, for the models monomolecular, logistic, gompertz, Richards and MMF. However, the disadvantage with richards and MMF model is that it involved an extra parameter and there was no gain due to this parameter as evidence by MSE. Therefore, either monomolecular, logistic or gompertz models describe the production data of rice for Bastar plateau.

All five models viz., monomolecular, logistic, gompertz, Richards and MMF fitted well to the data. Examination of residuals reveals that there was no evidence against the assumptions of randomness and normality of residuals. The goodness of fit statistics reveals that there was marginally difference among the models as far as R², RMSE, MAE and MSE are concerned. We found that there was no gain in using fourth parameter. Therefore, can considered either monomolecular, logistic or gompertz for studying the rice production data of northern hills.

Except MMF model the remaining four models fitted well to the data. There was no evidence to reject the assumptions of randomness and normality of residuals for these models. As there was no substantial difference in other goodness of fit statistics and R² values for all four models. It implies that the fourth-parameter in the model do not help to improve the fit. Hence, either monomolecular or logistic or Gompertz model describes the rice production data for whole Chhattisgarh State.

Nonlinear statistical models for studying productivity of rice:

Monomolecular and Richards models fitted well as compared other models to the data. A comparative study reveals that monomolecular model was significantly better than Richards model on the basis of goodness of fit statistics. It implies that the fourth parameter, in Richard’s model, does not help to improve the fit. Hence, monomolecular model was best for describing data sets of Chhattisgarh plain.

All the five models fitted well to the data sets. There was no evidence of violation of the assumption regarding residuals that these are random and normally distributed for all the five models. However, the disadvantage with both four-parameter models viz., Richards and MMF models was that it involved an extra parameter and there was no gain due to this parameter as evidence by MSE. Therefore, either monomolecular or logistic or Gompertz models describes the production data of rice for Bastar plateau.

In the case all the four models except MMF model fitted well to the data. R² values and goodness of fit statistics revealed marginally difference among the models. However, there was no gain in using four parameter models. It implies that the Richards model do not help to improve the fit. Hence, either monomolecular or logistic or gompertz model describe the rice yield data for northern hills.

In this case only logistic and richards’s model fitted well to the data. Examination of residuals revealed that RMSE and MAE values were same for the both models. However, MSE and r (Y, Ŷ) values were better found for logistic model. Hence, logistic is the most appropriate model for the data set of whole Chhattisgarh yield.

Conclusion:

Mathematical models play an important role in agricultural research. In this investigation, various nonlinear growth models viz., monomolecular, logistic, Gompertz, Richards and MMF were studied. One drawback of all these models was that they were deterministic in nature. These models can be replaced by adding an errors term on the right hand side of
the underlying equations. Though adding an error term on the right hand side of the model makes it more realistic, it also results in considerable difficulties in the estimation of parameters. The assumptions made for the error are that these are independent and normally distributed.

**Literature Cited**


