Indian Journal of Agricultural Sciences **84** (7): 856–9, July 2014/Article

Forecasting of growth rates of wheat yield of Uttar Pradesh through non-linear growth models

SANJEEV PANWAR[1], K N SINGH[2], ANIL KUMAR[3], SUSHEEL KUMAR SARKAR[4], RANJEET PAUL[5], ABHISHEK RATHORE[6] and N SIVARAMANE[7]

Indian Agricultural Statistics Research Institute, New Delhi 110 012

Received: 21 January 2013; Revised accepted: 27 March 2014

ABSTRACT

Wheat production in India is about 70 million tonnes per year which counts for approximately 12 per cent of world's production. Being the second largest in population, it is also the second largest in wheat consumption after China, with a huge and growing wheat demand. Major wheat growing states in India are Uttar Pradesh, Punjab, Haryana, Rajasthan, Madhya Pradesh, Gujarat and Bihar. All of north is replenished with wheat cultivation. Uttar Pradesh, the largest wheat growing region of the country, produces around 28 million tonnes of wheat and Bihar produces around 5 million tonnes. The usual parametric approach for growth rate analysis is to assume multiplicative error in the underlying nonlinear geometric model and then fit the linearized model by 'method of least squares'. This paper deals with a critical study of wheat yield of Uttar Pradesh with a non-linear approach. The available data of rice during different years is taken into consideration and different statistical models are fitted for that. The time series data on annual yield of wheat in UP from 1970-2010 were collected from various sources. Growth rates are computed through non-linear models, viz. Logistic, Gompertz and Monomolecular models. Different nonlinear procedures such as Gauss-Newton Method, Steepest-Descent Method, Levenberg-Merquadt Technique and Do Not Use Derivative (DUD) Method were used in this study to estimate the nonlinear growth rates. The results showed that logistic model performed better followed by Gompertz and monomolecular.

Key word: Forecasting, Gauss Newton method, Goodness of fit, Non-linear growth models, Non-linear growth rate, Randomness, Theil statistic

The usual parametric approach for growth rate analysis is to assume multiplicative error in the underlying nonlinear geometric model and then fit the linearized model by 'method of least squares'. This paper deals with a critical study of wheat yield of UP with a non-linear approach. The available data of wheat during different years is taken into consideration and different statistical models are fitted for that. The non-linear models, viz. Logistic, Gompertz and Monomolecular models are applied for wheat of Uttar Pradesh, a comparative study, growth rate, forecasting was conducted for these models.

The time series data on annual production of wheat in UP from 1970-2010 were collected from various sources. Growth rates are computed through non-linear models, viz. Logistic, Gompertz and Monomolecular models.

MATERIALS AND METHODS

Different nonlinear growth models are studied for the purpose of estimating the growth rate and fitting the best model, which will help in better future prediction. The use of RMSE, MAPE and Theil statistics as a measure of goodness of fit and, therefore, as a criterion for choosing the best model,

The compound growth rates should be computed by first identifying the model that describes satisfactorily the path followed by the response variable over time:

The Gompertz model provided better fit for different

Table 1 Fitting of non-linear models for UP

Parameter/ Statistics	Monomolecular	Gompertz	Logistic
k	4110.6005	4200.001	4119.3250
YZ	950.89385	925.89385	935.89283
r	0.02	0.02	0.03
Goodness of Fit			
R^2	0.95	0.95	0.95
RMSE	121.25	121.02	121.31
MAE	93.35	93.05	93.15
MAPE	4.59	4.37	4.28
Theil Statistics	0.33	0.32	0.31

RMSE=Root Mean Square Error, MAE= Mean Absolute Error, MAPE=Mean Absolute Percentage Error

[1] Senior Scientist (e mail: scientist1775@yahoo.co.in), [3] e mail: ajynarcgu72@gmail.com, [6] e mail: a.rathore@ cgiar.org

types of growth pattern observed in banana production growth data. This work will help for better forecasting and in improving the existing models. This also throws light on the likely instability in the forecast. The Gompertz model is based upon a model given by Gompertz (1825) for the hazard in life table, and then used as growth model by Wright (1926). The Gompertz model is very popular and used in various fields such as population studies and animal growth in situations where growth is not symmetrical about the point of inflection. The growth rate (Seber and Wild 1989) is given by

$$\frac{df}{dx} = kf\,(\log \alpha - \log f) \quad (k > 0, \alpha > 0), \qquad (1)$$

and the relative growth rate now declines with log(size). From (1) we obtain

$$f(x) = \alpha \exp\{-e^{-k(x-\gamma)}\} \qquad (2)$$

The time inflection for (2) is at time, $x = \gamma$ with $f=\gamma/e$, at which point the maximum growth rate, $w=k\alpha/e$, occurs, k and γ is the parameters. For Gompertz curve f is

Table 2 Computation of growth rates through non-liner modeling (UP)

Year	Growth Models					
	Monomolecular		Gompertz		Logistic	
	Predicted	Growth rate	Predicted	Growth rate	Predicted	Growth rate
1971	1029.5486	0.0577547	1029.5486	0.0300286	1029.5486	0.0175487
1972	1106.2454	0.0573581	1106.2454	0.0299078	1106.2454	0.0175119
1973	1181.033	0.0807434	1181.033	0.0361942	1181.033	0.0192106
1974	1253.9589	0.0636603	1253.9589	0.031762	1253.9589	0.0180577
1975	1325.0695	0.0511519	1325.0695	0.027938	1325.0695	0.0168864
1976	1394.4099	0.0515478	1394.4099	0.0280684	1394.4099	0.0169293
1977	1462.0242	0.0456678	1462.0242	0.0260593	1462.0242	0.0162425
1978	1527.9554	0.0416438	1527.9554	0.0245859	1527.9554	0.0157028
1979	1592.2453	0.0536507	1592.2453	0.0287497	1592.2453	0.0171501
1980	1654.9349	0.0375922	1654.9349	0.0230098	1654.9349	0.0150896
1981	1716.0639	0.0379751	1716.0639	0.0231631	1716.0639	0.0151509
1982	1775.6713	0.031138	1775.6713	0.0202761	1775.6713	0.0139305
1983	1833.7948	0.0296176	1833.7948	0.0195865	1833.7948	0.0136178
1984	1890.4715	0.0302336	1890.4715	0.0198682	1890.4715	0.0137465
1985	1945.7373	0.0266022	1945.7373	0.0181605	1945.7373	0.0129432
1986	1999.6274	0.0284257	1999.6274	0.0190325	1999.6274	0.0133602
1987	2052.1759	0.0272846	2052.1759	0.0184903	2052.1759	0.0131026
1988	2103.4164	0.0254126	2103.4164	0.017575	2103.4164	0.012655
1989	2153.3814	0.0225213	2153.3814	0.0160924	2153.3814	0.0118945
1990	2202.1026	0.0189986	2202.1026	0.0141597	2202.1026	0.0108336
1991	2249.6109	0.0189986	2249.6109	0.0141597	2249.6109	0.0108336
1992	2295.9367	0.021342	2295.9367	0.0154617	2295.9367	0.0115573
1993	2341.1092	0.0197271	2341.1092	0.0145717	2341.1092	0.0110667
1994	2385.1573	0.0161079	2385.1573	0.0124549	2385.1573	0.0098279
1995	2428.1088	0.0171725	2428.1088	0.0130962	2428.1088	0.0102142
1996	2469.9912	0.0136302	2469.9912	0.010896	2469.9912	0.0088467
1997	2510.831	0.0163232	2510.831	0.0125859	2510.831	0.0099076
1998	2550.6541	0.0159438	2550.6541	0.0123546	2550.6541	0.0097666
1999	2589.486	0.0122812	2589.486	0.0100049	2589.486	0.008258
2000	2627.3512	0.0128317	2627.3512	0.0103723	2627.3512	0.0085033
2001	2664.2738	0.0124037	2664.2738	0.0100871	2664.2738	0.0083132
2002	2700.2773	0.0147073	2700.2773	0.0115856	2700.2773	0.0092882
2003	2735.3846	0.0118787	2735.3846	0.0097327	2735.3846	0.008074
2004	2769.6179	0.016207	2769.6179	0.0125153	2769.6179	0.0098647
2005	2802.9991	0.0142363	2802.9991	0.0112864	2802.9991	0.0090981
2006	2835.5493	0.0128737	2835.5493	0.0104001	2835.5493	0.0085217
2007	2867.2892	0.0115759	2867.2892	0.0095261	2867.2892	0.007933
2008	2898.239	0.009309	2898.239	0.0079227	2898.239	0.0067985
2009	2928.4184	0.0112011	2928.4184	0.0092679	2928.4184	0.0077551
2010	2957.8465	0.0080783	2957.8465	0.0070074	2957.8465	0.0061178

Table 3 Computation of future forecast (UP)

Year	Monomolecular forecast	Gompertz forecast	Logistic forecast
2011	2986.5421	2441.9142	1883.7758
2012	3014.5234	2473.7772	1909.5262
2013	3041.8081	2505.2474	1935.3227
2014	3068.4136	2536.3197	1961.1573
2015	3094.3568	2566.9895	1987.0216
2016	3119.6542	2597.253	2012.9076
2017	3144.3219	2627.1067	2038.8071
2018	3168.3756	2656.5477	2064.7118
2019	3191.8304	2685.5734	2090.6136
2020	3214.7014	2714.182	2116.5041

monomolecular. Also the Gompertz function has the property that any power of f is also Gompertz function.

A model of population growth in which the growth rate is proportional to the size of the population. In the resulting model the population grows exponentially. In reality this model is unrealistic because environments impose limitations to population growth. A more accurate model postulates that the relative growth rate x^1/x decreases when x approaches the carrying capacity K of the environment. The corresponding equation is the so called logistic differential equation:

$$\frac{dx}{dt} = kx\left(1-\frac{x}{k}\right)+e \text{ ; Where,}$$

$$P = \frac{k}{1+A\exp^{kt}} \quad \text{where } A=\frac{k-x_0}{x_0} \qquad (3)$$

It is a symmetric sigmoid (S-shaped) curve.

Monomolecular 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, and can be given by Eq (4):

$$\frac{dx}{dt}=w(K-x) \qquad (4)$$

where, w and K are the intrinsic growth rate and carrying capacity of the system, respectively. Integrating Eq (4), we get Eq (5):

$$x(t) = K - (K - x_0)\exp(-wt) \qquad (5)$$

where, x_0 is the value of x(t) at t =0.

Nonlinear models are more difficult to specify and estimate parameter than linear models. Instead of simply listing exploratory variable, we must write the regression expression, declare parameter names, and guess initial values for them and possibly specific derivatives of the model respects to the parameters. Some models are difficult to fit and there is no guarantee that the procedure will be able to fit the model successfully. By using SAS package, "Proc nlin" fits nonlinear regression models through least square approach. For each nonlinear model to be analyzed, we must specify the following; (i) The names and initial values of the parameters to be estimated. (ii) The model (using a single dependent variable). (iii) Partial derivative of the model with respect to each parameter (except for Doesn't Use Derivative (DUD) method). (iv) The second derivatives of the model with respect to each parameter (Only for the Newton method).

Gauss Newton method, which is widely used, is used for fitting all models in this study. To start the iterative procedure, initial estimates of the parameters of the models were required. Many sets of initial values were tried to ensure global convergence. The iterative procedure was stopped when the reduction between successive residual sums of squares was found to be negligibly small.

Five main methods are available in the literature to obtain estimates of the unknown parameters of a nonlinear regression model. These are (i) Gradient method, (ii) Gauss Newton method, (iii) Steepest-descent method, (iv) Levenberg-Marquardt technique and (v) Donot-Use-Derivatives method. However, in all these methods the following steps are carried out.

Step 1: Starting with a good initial guess of the unknown parameters; a sequence of q's which hopefully converge to q is computed.

Step 2: Error sum of squares expressed as

$$S(\theta) = \sum_{t=1}^{N}\left[Y_T - F_i(\theta)\right]^2 \text{ is minimized with respect}$$

to the current value of θ. The new estimates are obtained.

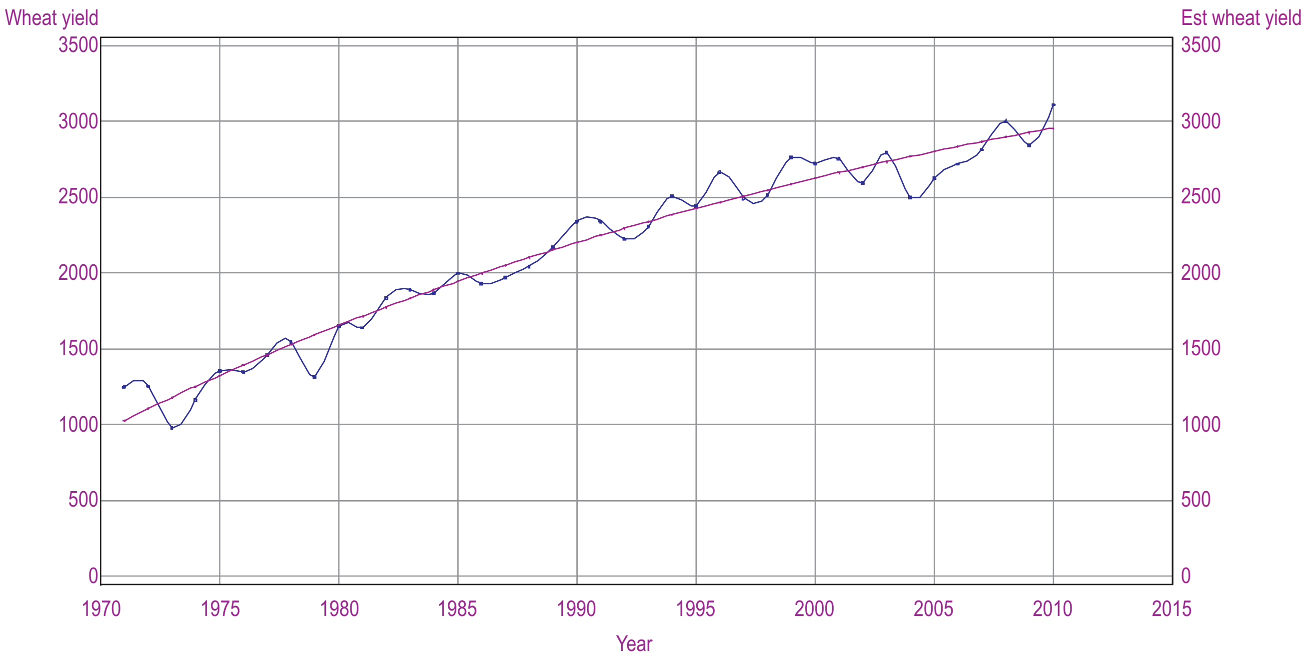
Step 3: By feeding the recently obtained estimates as the initial guess for the next iteration, objective function S(θ) is minimized again to obtain fresh estimates. This procedure is continued till the successive iteration yielded parameter estimate value is close to each other.

The starting value $\theta^{(1)}$, which is initial guess at the minimum θ, can sometimes be suggested by prior information. Sometimes there will be a starting value that tends to work well for a class of problems. Fisher's scoring algorithm for generalized linear models as an iterative re-weighted least square method suggests a uniform starting mechanism for the whole class of models (McCullagh and Nelder 1983). However, it is very difficult to say anything about producing good starting values in general. Draper and Smith (1998) and Ratkowsky (1983) had given a detailed discussion on starting values for nonlinear model.

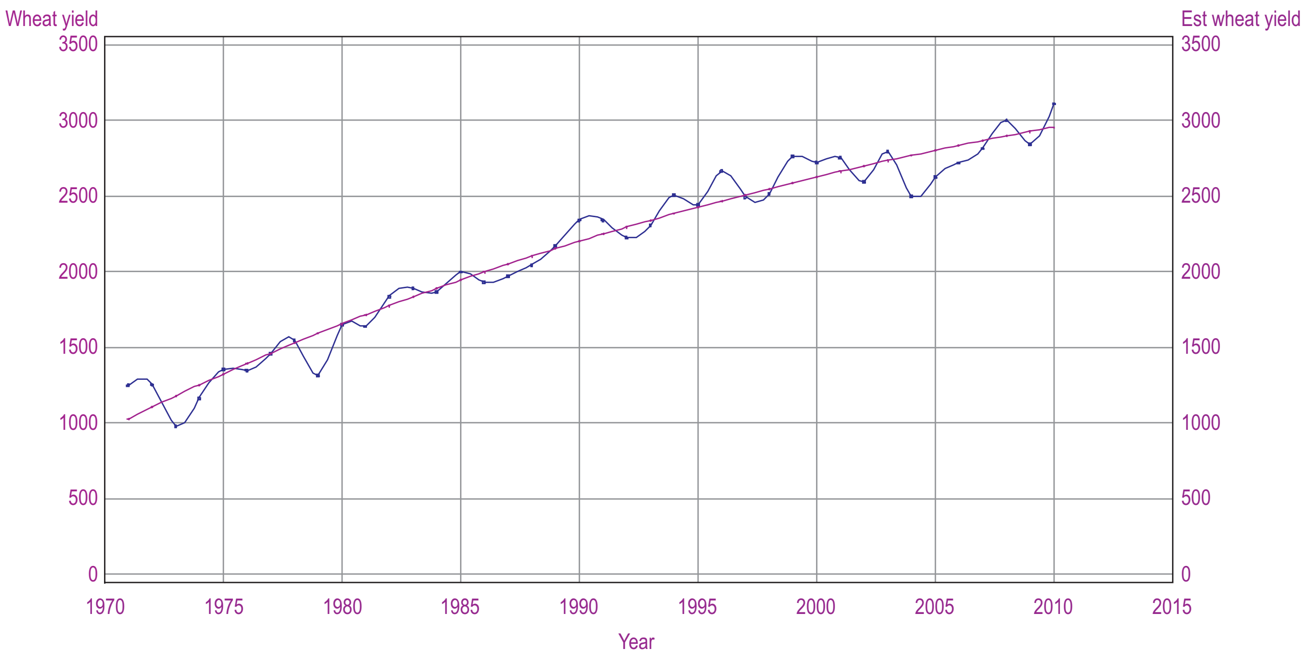
RESULTS AND DISCUSSION

The non-linear models, viz Logistic, Gompertz and Monomolecular models were applied for wheat production in UP. Based on performance of these fits, best non-linear models were chosen for the selected series.

In the first instance, attempts were made to identify the model that best described this data set. It is evident, from the analysis, production of wheat is fluctuating and all three nonlinear models are suitable to fit. The initialization of parameters done using Ratkowsky method and the results are compared using statistics such as, RMSE, MAE and MAPE, Theil statistics, i e The



Logistic Model



Gompertz Model

Gompertz, Logistic and monomolecular models have almost same MAPE (4.37) (Table 1), i e all goodness of fit statistics almost same. So in comparison of existing models, Gompertz or Logistic or Monomolecular models describe wheat production data well. Statistical Analysis System (SAS) software package was employed for data analysis. The annual growth rate for logistic model, Gompertz model and monomolecular model was computed and has been given in the Table 2. Finally, future forecasting also computed indicating on Table 3.

REFERENCES

Chandran K P and Prajneshu. 2004. Computation of growth rates in agriculture: Nonparametric regression approach. *Journal of the Indian Society of Agricultural Statistics* **57**: 382–92.

Dey A K. 1975. Rates of growth of agriculture and industry. *Economic and Political Weekly* **10** (25&26): A26–A30.

Draper N R and Smith H. 1998. *Applied Regression Analysis,* 3rd Edn. John Wiley & Sons, New York, USA.

Prajneshu and Chandran K P. 2005. Computation of compound growth rates in agriculture: Revisited. *Agricultural Economics Research Review* 18 July-December: 317–32.