Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/040111

Linear Models for S-Shaped Growth Curves

Alok Kumar Shukla[1], *Subhash Kumar Yadav*[2,*] *and Vishwas Tiwari*[1]

[1] Department of Statistics, D.A-V. College, Kanpur, U.P., India
[2] Department of Mathematics and Statistics (A Centre of Excellence)Dr. RML Avadh University, Faizabad-224001, U.P., India

Received: 29 Oct. 2014, Revised: 30 Dec. 2014, Accepted: 2 Jan. 2015
Published online: 1 Mar. 2015

Abstract: The S-shaped or sigmoidal models have two important members as Logistic model and Gompertz model. These models are very useful in studies of medical science, actuarial science and biological sciences. The proposed linear models are the alternative of Logistic model and Gompertz model respectively. Linear models satisfy the property of best linear unbiased estimator (BLUE) and removed the problem of nonlinear least squares estimation mentioned by Ratkowsky (1983, 1989) and Bates & Watts (1980). The goodness of fit for the proposed linear models have been verified with the help of several published data sets.

Keywords: Linear Model, Gompertz Model, Logistic Model, Sigmoidal Models

1 Introduction

One of the important categories of non-linear models is the Sigmoidal or S-shaped models in which the Logistic and Gompertz models are very important members. The Logistic model is very important among research studies relating to the fields of medical sciences, population studies etc. The Gompertz model is useful in actuarial studies, biological studies and also in economic phenomena. These models possess neither a minima nor maxima but have a point of inflection. They start from a point and increase to attain maximum growth rate at its inflection point and then growth rate decreases and model approaches an asymptotic value. A sigmoidal model may be symmetrical or asymmetrical about its inflection point. Ratkowsky (1983) has described deterministic components of Logistic and Gompertz models in the following form.
Logistic Model:

$$Y = \frac{1}{\alpha + \beta\rho^X} \tag{1}$$

Gompertz Model:

$$Y = exp\left(\alpha + \beta\rho^X\right) \tag{2}$$

We can also rewrite (1) and (2) as,
Logistic Model:

$$\frac{1}{Y} = \alpha + \beta\rho^X \tag{3}$$

Gompertz Model:

$$log\,Y = \alpha + \beta\rho^X \tag{4}$$

Equation(3) is reciprocal transformation of Y in equation (1) defines logistic model and equation (4) is *log* transformation of Y in equation (2) describes Gompertz model. Logistic model is symmetrical about its point of inflection whereas Gompertz model is asymmetrical about its point of inflection. Vierra and Hoffman (1977) have remarked that a model having a point of inflection at a fixed point is not desirable. They advised to use models which allow degree of asymmetry in it.

Draper and Smith (1998) have classified Logistic and Gompertz models as intrinsically non-linear models. Ratkowsky (1983, 1989) and Draper and Smith (1998) have listed several models which are reparameterization of Gompertz and Logistic models. These reparametrized models are similar in their behavioural properties.

* Corresponding author e-mail: drskystats@gmail.com

Depending on the properties of least square estimators of parameters of nonlinear models, Ratkowsky (1983) has classified nonlinear models as far-from-linear or close-to-linear. If least square estimates of parameters of nonlinear models are close to being unbiased having distribution very close to normal distribution and having variance very close to minimum variance bound, the nonlinear models are classified as close-to-linear. On the contrary, if least square estimators of nonlinear models are badly biased with distribution far-from-normal, having variance in large excess of minimum variance bound, the nonlinear models are classified as far-from-linear. Ratkowsky (1983) has remarked that in close-to-linear model, there is always a sure convergence in whatever methods used for parameter estimation. He further remarked that the convergence is not only faster but also to a minima which is a global minima. These models show close-to-linear behaviour. The reparameterizations containing expected value parameters are more cumbersome in appearance than original expressions and expected value parameters may appear more than once, even though the models with expected value parameters are still preferred because of close-to-linear behaviour.

Bates and Watts (1980) developed two measures of accessing nonlinearity, one is intrinsic nonlinearity and another is parameter effects. They have shown that suitable reparameterization of nonlinear model can minimize parameter effects whereas it has no effect on intrinsic nonlinearity. Ratkowsky (1983) remarked that most of the nonlinear models of interest possess low intrinsic nonlinearity. The main component responsible for nonlinearity in a nonlinear model is parameter effects and suitable reparameterization of nonlinear model can substantially reduce it and a nonlinear model can be brought into a form which will be behaving as close-to-linear model. A nonlinear model which is far-from-linear model may have more than one minima in its residual sum of square surface and if in such a nonlinear model convergence occurs at all, it occurs to a local minima and not to a global minima which make inferences from parameter estimates doubtful.

It is thus absolutely clear that the main thrust is on choosing nonlinear models which are close-to-linear in their behaviour. These facts have provided a great deal of motivation to use linear models.

A linear model is one whose parameters appear linearly in it and in nonlinear model, at least one parameter appears nonlinearly. A model is said to be linear in its parameters if the first order partial differentials of the model with respect to its parameters are independent of parameters. In nonlinear model at least one first order partial differential of the model contains parameter(s).

2 Proposed Linear Models

Motivated by above statement, we have used two models linear in their parameters as the alternative to Logistic and Gompertz models. Shukla *et al.* (2011) have proposed the linear models as an alternative of asymptotic regression model.

$$Y = \alpha + \beta\rho^{X},\ 0 < \rho < 1 \tag{5}$$

$$Y = a + \frac{b}{X} + \frac{c}{X^2},\ X > 0 \tag{6}$$

Equation(5) is asymptotic regression model and equation (6) is alternative linear model for asymptotic regression model proposed by Shukla *et al.* (2011). We have written only deterministic part of the model ignoring subscripts representing number of observation on Y and X with parameters a , b and c. All these models admit an additive error term, which is assumed to be independently and identically distributed random variable with mean zero and fixed variance having no autocorrelation, heteroscedasticity and multicolliearity. If error term is assumed to be normally distributed, the maximum likelihood estimates of parameters can also be obtained.

Right Hand Side (RHS) of equation(5) is replacing by $a+\frac{b}{X}+\frac{c}{X^2}$ then we find equation (6).

Such as this expression $a+\frac{b}{X}+\frac{c}{X^2}$ is replace in equation (3), and then we have

$$\frac{1}{Y} = a + \frac{b}{X} + \frac{c}{X^2},\ X > 0 \tag{7}$$

This is linear model for Logistic model.

Again RHS of equation(4) replacing by expression $a+\frac{b}{X}+\frac{c}{X^2}$,then we have

$$log\, Y = a + \frac{b}{X} + \frac{c}{X^2},\ X > 0 \tag{8}$$

This is linear model for Gompertz model.

Thus models (7) and (8) are linear models as alternative for nonlinear Logistic and Gompertz models respectively.

We can postulate statistical models for (7) and (8) as:

$$\frac{1}{Y} = a + \frac{b}{X} + \frac{c}{X^2} + U_i,\ X > 0 \& i = 1, 2, \ldots, n \tag{9}$$

$$log\ Y = a + \frac{b}{X} + \frac{c}{X^2} + U_i,\ X > 0 \& i = 1, 2, \ldots, n \tag{10}$$

The random variable U_i's are assumed to be independently and identically normally distributed random variable with mean zero and fixed variance (σ^2) with no heteroscedasticity, autocorrelation and multicollinearity.The constants a , b and c are the unknown parameters of the models (9) and (10).

2.1 Goodness of fit of proposed linear models

Coefficient of Determination -R^2

A way of assessing the regression model was to see how much of the total sum of squares had fallen in to the sum of squares due to the regression. The Coefficient of Determination-R^2 is defined as,

$$R^2 = \frac{SS_R}{SS_T} = 1 - \frac{SS_{Res}}{SS_T}$$

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - \overline{Y}\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}$$

Where SS_T measures the total variability in the observations Y, SS_R as the regression or model sum of squares and SS_{Res} as the residual or error sum of squares .The value of R^2 very close to 1 implies that most of the variability in Y has been explained by the fitted model. Thus observance of a high R^2 value indicates a good fit. Oftenly it is multiplies by 100 so as to be expressed it in percentage form.

Residual Mean Square -s^2

The best criterion to choose a model is the residual mean square(s^2). The residual mean square(s^2) is an unbiased estimator of σ^2. Its expression for our models (9) and (10) will be,

$$s^2 = \frac{SS_{Res}}{n-3} = \frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n-3} = \frac{\sum_{i=1}^{n} e_i^2}{n-3}$$

Where n have pairs of data (Y_i, X_i) $(i = 1, 2, \ldots, n)$. The value of s^2 indicate that error due to regression or model. A smaller value of s^2 reflects the appropriateness of the fitted model.

Mean Absolute Error (M.A.E.)

The mean absolute error which is average of absolute error is defined as,

$$M.A.E. = \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{n} = \frac{\sum_{i=1}^{n}\left|e_i\right|}{n}$$

A smaller value of $M.A.E.$ preferred in any data sets to which a model has been fitted.

Using statistical software SPSS 17.0 for fitting proposed linear statistical models (9) and (10) by multiple regression analysis.

3 Empirical Study

The models (9) and (10) are fitted for examine goodness over nonlinear Logistic and Gompertz model, data were obtain various published papers in different journals. Secondary data is more reliable for adequacy and appropriateness of proposed model.

Table 1: Model (9)

$S.No.$	R^2	s^2	$M.A.E.$
1	99.394	$6.2563E-7$	0.000643
2	98.441	$4.64422E-7$	0.000529
3	99.605	$7.32468E-8$	0.000173
4	98.096	$3.7636E-7$	0.000416
5	97.728	$4.59694E-7$	0.000459
6	99.979	0.0000246	0.003010

Data Source: S.No.1 from H. Singh *et al.* (1991), S.No.2 from M. Grossman *et al.* (1985), S.No.3 from S. Vieira and R. Hoffman (1977), S.No.4 from T. Sengul and S. Kiraz (2005) for male, S.No. 5 from T. Sengul and S. Kiraz (2005) for female, S.No. 6 from M.G. Kundu and A.K. Paul (2010)

Table 2: Model (10)

$S.No.$	R^2	s^2	$M.A.E.$
1^{*0}	99.883	0.00108	0.01587
2^{*}	98.759	0.00839	0.04398
3^{*}	97.518	0.02332	0.09749
4^{*}	98.759	0.00839	0.04398
5^{**}	99.373	0.00527	0.04774
6^{***}	99.576	0.00734	0.05520

Data Source: $*0$ Kutner *et al.*(2004) p 627, $*$Kutner *et al.*(2004) p 569 , $**$ Ratkowsky (1983) p 88 data , $***$ Draper and Smith (1998) p 562 of sample 8

The table-1 shows goodness of fit for model (9) to Logistic data sets. The table-2 shows goodness of fit for model (10) to Gompertz data sets. The tables gives the values of coefficient of determination (R^2), the values of residual mean square (s^2) and mean absolute error ($M.A.E.$). The residuals follow the assumptions of zero mean, normal distribution and fixed variance and there is no autocorrelation. Also there is no multicollinearity in the data.

4 Conclusion

In the present manuscript two linear regression models have been proposed as the alternatives of Logistic and Gompertz nonlinear regression models. The proposed linear models are proved theoretically as well as numerically as good alternative linear models to mentioned nonlinear models. Applications of linear models to data sets are very convenient as famous least squares method of estimation is directly applicable and parameter estimates possess good statistical properties. The predictions, constructions of confidence intervals and test of significance procedures etc can be carried out very conveniently, using proposed linear regression models. Thus the proposed linear regression models should be used the alternatives of Logistic and Gompertz nonlinear regression models in practice.

Acknowledgements

The authors are very much thankful to the editor-in-chief and the unknown learned referees for critically examining the manuscript and giving valuable suggestions to improve the manuscript.

References

[1] D.M. Bates and D.G. Watts, Relative Curvature Measures of Non-Linearity, Journal of Royal Statistical Society, **B-42, 1-16, Discussion**, 16-25 (1980).
[2] N.R. Draper and H. Smith , Applied Regression Analysis, 3rd **Ed.**, John Wiley (1998).
[3] M. Grossman, B.B. Bohren and V.L. Anderson , Logistic Growth Curve of Chickens: a Comparison of Techniques to Estimate Parameters, J. Heredity,**76**, 397-399 (1985).
[4] M.G. Kundu and A.K. Paul , Phenotypic and Genetic Variability of Estimated Growth Curve Parameters in Pigs, J. Indian Soc. of Agr. Stat., **64(3)**, 349-358 (2010).
[5] M. H. Kutner, C. J. Nachtsheim and J. Netar, Applied Linear Regression Models, Mc Graw Hill,Singapore (2004).
[6] D. A. Ratkowsky , Non-Linear Regression Modeling, Marcel Dekker (1983).
[7] D. A. Ratkowsky , Hand Boook of Non-Linear Regression Models, Marcel Dekker (1989).
[8] T. Sengul and S. Kiraz , Non-Linear Models for Growth Curves in Large White Turkey, Turk. J.Vet. Anim. Sci.,**29**, 331-337 (2005).
[9] A.K. Shukal, S.K. Yadav and G.C. Misra , A Linear Model for Asymptotic Growth Curve, Elixir Statistics,**41**, 5991-5993 (2011).
[10] H. Singh, A.K. Nivsarkar and C.L. Suman , Growth Curve Analysis and Genetic Parameters of Body Weight in Indigenous Guinea Fowl, Indian J. Poult. Sci.,**26**, 20-25 (1991).
[11] W.L. Stevens , Asymptotic Regression, Biometrics,**7**, 247-267 (1951).
[12] S. Vieira and R.Hoffman , Comparison of the Logistic and The Gompertz Growth Functions Considering Additive and Multiplicative Error Terms, Appl. Stat.,**26**, 143-148 (1977).