Growth pattern of crossbred cattle under homoscedastic and heteroscedastic error variance condition

Article *in* The Indian journal of animal sciences · May 2008

Lalmohan Bhar
Indian Agricultural Statistics Research Institute

Amrit Paul
Indian Agricultural Statistics Research Institute

Indian Journal of Animal Sciences **78** (5): 560–564, May 2008

Growth pattern of crossbred cattle under homoscedastic and heteroscedastic error variance condition

SURENDRA SINGH[1], L M BHAR[2] and A K PAUL[3]

Indian Agriculture statistical Research Institute, New Delhi 110 012 India

Received: 20 February 2007; Accepted: 5 January 2008

Key words:Double cross, Growth models, Heteroscedasticity, Homoscedasticity

Growth studies are very important for the livestock production because growth is the foundation on which production of milk, meat and wool rests. These studies serve as an aid in assessing the maximum production potential of livestock and play a significant role in animal production and welfare.

Growth models are used to predict rates and change in the shape of the organism. Comparison of nonlinear models for weight-age data in cattle has been studied under homoscedsticity (Brown *et al.* 1972, Brown *et al.* 1976, Alessandra *et al.* 2002, Kolluru *et al.* 2003). A number of such nonlinear models are available, but comparison of models are needed to find most appropriate model. Such comparisons were made among weight-age models for animals. Kolluru (2000) studied only Logistic model under heteroscedastic error condition. There is a need to study other model also, hence logistic and Gompertz models were taken for the present study.

Data used in the study were collected from Agra station for Friesian×Sahiwal breed. Data for 40 cattle were collected from birth to 36 months of age for comparing the growth pattern. Growth pattern of Friesian×Sahiwal breed has been studied by the following models:

Logistic model

$$X_t = \frac{\beta_1}{1+\beta_2 \exp(-\beta_3 t)}$$

β_1- asymptotic weight; β_2–scaling parameter; β_3-rate of maturity; X_t is dependent variable (weight); t-time (age).

Gompertz model

$$X_t = \beta_1 \exp(-\beta_2 e^{-\beta_3 t})$$

Fitting of nonlinear models under homoscedastic error structure, based on some assumptions as explained below:

Present address: [1-3]

Let us consider the following model,

$$Y_j = f(X_j, \beta) + \varepsilon_j$$

is covariate vector and is ß is parameter vector and f is a non-linear function. Usually, it is assumed that (i) errors ε_j have zero means, (ii) errors ε_j are uncorrelated, (iii) the errors ε_j has common variance, (iv) the errors ε_j are normally distributed.

We explore the various nonlinear interative techniques for estimation of parameters for the models used in the study. There are 4 main methods for nonlinear estimation (i) linearization method, (ii) gradient method, (iii) Levenberg Marquardt method (1963), (iv) Don't use derivative (DUD) method. Theoretical details of the methods are given in SAS User's guide (1990). Procedure NLIN is available in SAS Package to fit a nonlinear model by any one of these procedures, based on relative merits and demerits.

Idea behind the present work is to give best fitted model for the cattle growth and to have an idea about how the cattle weight behaves over time, by plotting the data. The scatter plot of data showed an 'S' shaped curve, so the nonlinear sigmoidal models discussed are fitted to the growth data. Values of the parameter estimates with their asymptotic standard error are presented in Table 1.

The empirical comparison of models can be done with goodness of fit statistics R^2, MSE and RMSE. Higher the value of R^2 and lower the values of other measures of statistics better are the models. It is concluded that the model which has minimum RMSE and maximum R^2, will be the best fitted model.

R^2 =1–(Residual sum of squares/total sum of squares),

$$\text{residual sum of squares} = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$\text{total sum of squares} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2,$$

Y_i dependent variables (weights), $\hat{Y}_i$ is predicted values or estimated value, $\overline{Y}$ is mean of dependent variables, n = no.

Table 1. Parameter estimates of logistic and Gompertz models under homoscedastic and heteroscedastic error structure

Parameter	Logistic	Gompertz	Heterscedasticity of error variance (logistic model)	Heteroscedasticity of error variance (Gompetz model)
α	410.5000	446.1000	343.9066	315.6594
	(13.6002)	(12.8978)	(4.0048)	(1.4062)
ß	8.0829	2.5350	10.4364	2.5214
	(0.8657)	(0.0807)	(0.1656)	(0.0051)
γ	0.1623	0.0951	0.2531	0.1354
	(0.0114)	(0.0051)	(0.0034)	(0.0008)
Goodness of fit statistics				
R^2	0.9887	.9956	0.9999	0.9999
MSE	226.0819	85.8952	0.0235	0.0072
RMSE	15.0360	9.3217	0.1534	0.0851

of observation, p= no. of parameter to be estimated.

$$\text{Root mean squared error (RMSE)} = \left[\sum_{i=1}^{n}\frac{(Y_i-\hat{Y}_i)^2}{n-p}\right]^{1/2}$$

As in case of the animal growth data, many a times, the above assumption of nonlinear models under homoscedasticity is violated. In the present study data revealed heteroscedasticity. This was tested by using Rank correlation test. It can be observed from Table 2, that the rank correlation is high, which reveals the heteroscedasticity of error variance is present in the growth data.

As the data were heteroscedastic, OLS is not to be used any more. So to estimate the parameters we go for using the procedure explained in Model with heteroscedastic error

Table 2. Rank correlation test for Frisian×Sahiwal breed at Agra Station

t	y	pred	res	Abs	r2	r1	d=r1-r2	d*d
0	24.36	21.443	2.9165	2.9165	9	1	-8	64
0.25	27.71	25.729	1.9808	1.9808	8	2	-6	36
0.5	30.79	29.994	0.7963	0.7963	1	3	2	4
1	37.64	38.459	-0.819	0.8188	2	4	2	4
2	54.07	55.134	-1.064	1.064	4	5	1	1
3	70.14	71.469	-1.329	1.3289	5	6	1	1
5	106.43	103.12	3.3118	3.3118	10	7	-3	9
7	138.07	133.41	4.6633	4.6633	11	8	-3	9
9	161.36	162.33	-0.975	0.9745	3	9	6	36
11	178.79	189.9	-11.11	11.111	16	10	-6	36
13	205.21	216.11	-10.9	10.898	15	11	-4	16
15	242.93	240.95	1.9771	1.9771	7	12	5	25
18	273.93	275.67	-1.74	1.7395	6	13	7	49
21	312.92	307.32	5.5957	5.5957	13	14	1	1
24	340.77	335.92	4.8527	4.8527	12	15	3	9
30	395.43	383.92	11.512	11.512	17	16	-1	1
36	410	419.67	-9.672	9.6716	14	17	3	9
								310
			R=	0.6201				

r_1 denots Rank of time (independent variable); r_2 denotes rank of residuals, abs denotes absolute value.

structure. For the procedure discussed a computer program has been developed in IML (Interactive Matrix Language)-a module of SAS and the parameters under heteroscedastic error structure are estimated. Parameters are estimated and weights are predicted, using homoscedastic error structures. Results are given in Table 1.

Test for heteroscedasticity of variance

Various tests are available in literature (Koutsoyannis 1993) for testing heteroscedasticity in data. We will consider here rank correlation test which is easy to use both computationally and conceptually. It is the simplest test and is used for both small and large samples. Technique involved following steps;

Consider the model

$$Y = f(X, \beta)+e \quad (1)$$

(i) Regress Y on X and obtain residuals.

(ii) Order the residuals (ignoring their sign) and X values in ascending or descending order and compute the rank correlation coefficient by following formula

$$R_{e,x} = 1-\{6\sum d_i^2/(n\ (n^2-1))\} \quad (2)$$

d_i, difference between the ranks of corresponding pairs of X and e_j;

n, Number of observations in the sample.

A high rank correlation suggests the presence of heteroscedasticity.

Models with hetroscedastic error structure

Logistic and Gomperatz model are used for developing appropriate models under heteroscedastic error structure for growth of cattle, as it is well known that in most of the cases growth data follows logistic and Gompertz model. However, as far as cattle growth is concerned, it is assumed that the errors are independently distributed with constant variance. However, this assumption rarely meets in the reality, particularly for growth models.

Let us consider the following model,

$$Y_j = f(X_j, \beta)+e_j \quad (3)$$

X_j is covariate vector, ß is parameter vector and f is a non-linear function. Usually, it is assumed that the errors have zero means and uncorrelated, the errors ε_j have common variance and are normally distributed.

Though the first assumption ensures that the model f for mean response is correctly specified. This assumption is rarely called into question, as it is usually the case that the form of the covariate-response relationship is fairly well understood, especially for nonlinear relationships, where the model may result directly from theoretical considerations.

The remaining 3 assumptions are fairly restrictive and may not hold in some applications. For the present data, which are observed over time, the specification of uncorrelated errors may be unrealistic. Assumption of constant intra-individual response variance is violated frequently in practice. For example, growth data, often exhibit constant coefficient of variation rather than constant variance (Davidian and Giltinan 1995); that is variance proportional to the squares of the mean response. In this case, a more appropriate assumption would be:

$$E(Y_j)=f(X_j, ß),\ V(Y_j)=(CV)^2[f(X_j,ß)]^2 \quad (4)$$

Where, CV< the scale parameter, is the coefficient of variation.

Since we are considering that heterogeneity of variance is evident in growth data of cattle, this can be verified by applying the testing procedure as given in Test for heteroscedasticity of variance, we found that heterogeneity is present in data. Hence if we apply ordinary least square method, the parameter estimates would be inefficient relative to method that considers heteroscedasticity.

Motivating our discussion of this issue and of how the classical least squares are nonconstant across the response range such that the variances of Y are known up to a constant of proportionality as:

$$E(Y_j)=f(X_j, ß),\ V(Y_j)=\sigma^2/w_j \quad (5)$$

for some constants W_j, j=1, ..., n

Under this setting with the assumption of independent normal errors, it is straightforward to show that the maximum likelihood estimator ß of parameter is the value $\hat{ß}_{WLS}$ that minimized normal equation

$$\sum_{j=1}^{n} W_j\{Y_j - f(X_j, ß)\}^2 \quad (6)$$

estimating equivalently $\hat{ß}_{wls}$ by the equation

$$\sum_{j=1}^{n} W_j(Y_j - (f(X_j, ß))f_ß(X_j, ß)=0$$

$\hat{ß}_{wls}$ is weighted least square estimate

As the ordinary least squares, the maximum likelihood estimation for σ^2 is usually replaced by

$$\hat{\sigma}_{WLS}^2 = \frac{1}{n-p}\sum_{j=1}^{n} W_j\{Y_j - f(X_j, \hat{\beta}_{WLS})\}^2 \quad (7)$$

where p are the number of parameters to be estimated.

For definiteness, consider the constant coefficient of variation in equation (4), except for the multiplicative constant σ^2 variance is known up to the value of the regression parameter ß, which appears through the mean response. An obvious approach is thus to take advantage of the functional form for a variance to construct estimated weights, replacing ß by a suitable estimate, and to apply the weighed least squares idea. The OLS estimator $ß_{OLS}$ is a natural choice to use for construction of estimated weights. Formally, an estimator for ß that takes into account the assumed mean-variance relationship may be obtained by forming estimated weights.

$$\hat{W}_j = \frac{1}{f^2(X_j, ß_{OLS})} \quad (8)$$

And then solving (6), using the $\hat{W}_j$ in place of W_j intuition suggests that the resulting estimator for ß will be preferred to that obtained by ordinary least squares.

This example is a special case of a very general class of methods for estimation of ß known in the context of nonlinear regression as generalized least squares (GLS). GLS method can be characterized by the following scheme.

Step 1. Estimate ß by a preliminary estimator $\hat{\beta}_p$ e.g. the OLS estimator $\hat{\beta}_{OLS}$

Step 2 Form estimated weights $\hat{W}_j = \frac{1}{f^2(X_j, \hat{ß}_{OLS})}$

Step 3. Using the weights from step 2, re-estimate ß by the weighted least squares. Treating the resulting estimator as a new preliminary estimator, return to step 2. Denote the final estimate by $\hat{\beta}_{GLS}$ and calculate variance by formula given above in (7).

Computational aspect

For implementation of above explained procedure we will consider functional form of mean response as

$$f(X_j ß) = \frac{ß_1}{1+ß_2\exp(-ß_3 X_j)} \quad (9)$$

$$ß=\begin{pmatrix} ß_1 \\ ß_2 \\ ß_3 \end{pmatrix}$$

This is Logistic function.

$$f(X_j, ß) = ß_1\exp(-ß_2 e^{-ß_3 X_j}) \quad (10)$$

$$ß=\begin{pmatrix} ß_1 \\ ß_2 \\ ß_3 \end{pmatrix}$$

This is Gompertz function.

Logistic and Gompertz model are fitted for all sets of data estimated parameters denoted by $ß_{OLS}$ by using SAS software.

A program written in SAS/IML is used to obtain parameter estimate under heteroscedastic error variance models.

Steps are given as follows:

Step 1

The preliminary estimates of ß calculated by Gauss-Newton procedure. This procedure has been discussed with full detail SAS USER'S guide. After calculating preliminary estimates of ß we will go for step 2.

Step 2

In step 2, the motive is to form weights to use them for weighted least square procedure. Since functional form of mean response in known, we will estimate the value of mean response for each ordinary least squares estimate. After calculating estimated values, weights are formed by using equation (8) for functions of equation (9) and (10). Once weights have been calculated we go for step 3. $\hat{w}$ are formed by using equation (8) for functions of equation (9) and (10). Once weights have been calculated we go for step 3.

Step 3

Third step consists of applying (weighted least squares) W.L.S. to obtain estimates by using weights obtained from previous step. Now our objective function is

$$O(ß)= [Y-f(X,ß)]' W^{-1}[Y-f(X,ß) \quad (11)$$

where O (ß) is objective function to be minimized, W is a diagonal matrix, whose diagonal elements are the weights, Y is vector of observations and f is mean response.

To get estimate of ß we will minimize objective function i.e, equation (11). Because closed from solution of the generalized least squares estimating equations are rarely available, computation of nonlinear least squares estimates required the use of iterative numerical Newton-Rephson technique for quadratic Taylor Series expansion, which is given by

$$ß=ß^*-J^{-1} (ß^*)S(ß^*) \quad (12)$$

where J_{pxp} is known as Hessian matrix, whose elements are second order partial derivatives of the objective function with respect to parameters, such that

$$J_{k1,k2} = \left(\frac{\partial^2}{\partial\beta_{k1}\partial_{k2}} \right) O(\beta) \quad (13)$$

where O (ß) is objective function to be minimized, J (k_1, k_2) is (k_1, k_2)th element in Hessian Matrix, S($ß^*$) is partial derivative matrix of objective function at approximate value of ß.$ß^*$ is approximate value of ß and S_{px1} is a vector of partial derivative of objective function with respect to parameter vector such that

$$S_i = \left(\frac{\partial}{\partial\beta_i} \right) O(\beta) \quad (14)$$

Computation of the Hessian matrix may be quite burdensome. Actually J (ß) is replaced by its expectation. In general, the expectation matrix will be easier to calculate than J (ß). The expectation of J(ß) can be written as

$$2Z'W^{-1}Z(ß) \quad (15)$$

where Z_{nxp} is matrix of first order partial derivatives of mean response with respect to parameters such that

$$Z_{ij} = (\frac{\partial}{\partial\beta_j}) f(X_j,\beta),\ i = 1, 2, \dots n, j=1, 2, \dots p \quad (16)$$

where, n, number of observations, and p, number of parameters to be estimated.

Vectors can be computed by the formula given below,

$$S(ß) = 2Z'W^{-1} [Y-f(ß)] \quad (17)$$

Since it is an interative process so first we will estimate value of S and J by preliminary estimates of ß =$ß_{PRL}$ i.e, $ß_{OLS}$ and will use them in equation (12). After getting new estimates of ß (sayß*) and use again in equation (12), till the values of ß converges. Final value of estimates of ß will be represented as $ß_{GLS}$.

It is evident from Table 1 that for Friesian × Sahiwal breed at Agra station RMSE (9.3217) is less for Gompertz model than RMSE (15.0360) of Logistic model, which shows that results of Gompertz model are better than Logistic model under homoscedastic error structure condition, hence growth rate is batter for Logistic model. As data having heteroscedasticity of variance, so models are modified, incorporating heteroscedasticity of variance. When results are compared under homoscedastic error condition and heteroscedastic error condition, RMSE is found less for heteroscedastic error condition. This shows that when model fitted under heteroscedastic error condition, it gives better results than homoscedastic error condition.

SUMMARY

Different growth models are fitted in growth data for Friesian×Sahiwal breed at Agra station. Gompertz model gave better fit than Logistic model. The GLS estimates are found to be more precise than OLS estimates for both Logistic as well as Gompertz model under heteroscedastic error condition.

REFERENCES

Alessandra F, Bergamasco, Luiz Henrique de Aquino, Joel Augusto Muniz, Febyano Fonseca e Silva. 2002. *Growth Curve of Holstein Heifers Females*. American society of Agricultural and Biological Engineers, St. Joseph, Michigan

Brown J E, Brown C J and Butts W T. 1972. A discussion of the aspects of weight, mature weight and rate of maturing in Hereford and Angus cattle. *Journal of Animal Science* **34**: 525.

Brown J E, Fitzhugh H A and Cartwright T C. 1976. A comparison of Nonlinear models for describing weight-age relationships in cattle. *Journal of Animal science* **43**: 810–18.

Davidian David and Giltinan M. 1995. *Non linear models for Repeated Measurement Data*. London: Chapman Hall, pp. 11–

210.

Draper N R and Smith H. 1966. *Applied Regression Analysis*. New York: Wiley

Kolluru R. 2000. 'On some aspect of growth patterns of crossbred cattle.' Thesis submitted to Indian Agricultural Statistics Research Institute, New Delhi 110 012.

Ramesh Kolluru, Rana P S and Paul A K. 2003. Modelling for growth pattern in crossbred cattle. *Journal of animal Science* **73**(10): 1174–79.

Leaflet A S, Hassen A, Wilson D E, Rouse G H and Trait R G. 2004. Use of linear and Nonlinear growth curves to describe body weight changes of young Angus bulls and heifers. Iowa state University Animal Industry report.

Lambe N R, Navajas E A, Simm G and Bunger L. 2006. A genetic investigation of various growth models to describe growth of lambs of two contrasting breeds. *Journal of Animal Science* **84**: 2642–54.

Marquardt D L. 1963. An algorithm for least squares estimation of non-linear parameters *Journal of Society for Indian Applied Mathematics* **2**: 431.

SAS. 1990. Users' Goude Varsion 6. Rdn. 4. SAS Institute Incorporation, USA.