Available online at www.isas.org.in
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 63(2) 2009 159-164

Modelling for Growth Pattern in Crossbred Cattle along with Autocorrelation

Surendra Singh[1], L.M. Bhar[1], A.K. Paul[1*], Satish Chand Sharma[2] and N. Sivaramane[1]
[1]*Indian Agricultural Statistics Research Institute, New Delhi*
[2]*J.V. College, Baraut, Bagpat*

(Received: October 2007, Revised: April 2009, Accepted: July 2009)

SUMMARY

Research on cattle growth is one of the important studies in the animal sciences. In the present study data on body weight were taken from birth to an age of 36 months for double cross Friesian × Sahiwal (F×S) and tripple cross Friesian × Sahiwal × Hariana (F×S×H) cattle. The data is found to contain heteroscedasticity of error variance and autocorrelation. The data in the study were unequal spaced which made impossible to use convention autocorrelation technique and hence a technique has been developed for finding out autocorrelation for unequal spaced data. Two nonlinear models, Logistic and Gompertz models are fitted to estimate growth rate and other parameters. Models are modified incorporating heterogeneity of error variance along with autocorrelation. Both the models were fitted under homoscedastic error structure and heteroscedastic error structure along with autocorrelation for comparison. It is found that parameters estimated under heteroscedastic error structure along with autocorrelation are better than the models fitted under homoscedastic error structure. Growth rate was found to be better for F×S×H breed than F×S breed. Maturity weight was found to be more for F×S breed than F×S×H breed. The results shows that Gompertz model outperformed Logistic model and correcting the models under homoscedastic error structure to heteroscedastic error structure has greatly improved the estimates.

Keywords: Homoscedastic, Heteroscedastic, Autocorrelation, Crossbred, Non-linear model, cattle growth, Logistic model, Gompertz model.

1. INTRODUCTION

Growth is a complex biological process that must be evaluated carefully if a profitable combination of growth and efficiency is to be realized. Knowledge relating to birth weight, mature weight, maturing rate and the point of inflection of the growth curve in various breeds and crosses is useful for cattle growth in various breeds so that producers can select breed combinations that will produce the most efficient growth pattern for their operations.

The growth pattern of cattle has been mainly studied under homoscedastic error structure, reported in the literature. But data of cattle growth generally violate assumption of homoscedasticity, i.e., error have common variance. Therefore, the purpose of this study is to compare the growth pattern under homoscedastic and heteroscedastic error structure along with autocorrelation.

Growth models are used to predict rates and change in the shape of the organism. They can be applied in determining the food requirements so as to get a desired growth. The estimated parameters of growth function can evaluate various growth characteristics of animal. Comparison of nonlinear models for weight age data in cattle has been done

*Corresponding author : A.K. Paul
E-mail address : amrit_66@rediffmail.com

under homoscedastic error structure ((Brown *et al.* (1972), Brown *et al.* (1976) and Kolluru *et al.* (2003)). A number of such nonlinear models are available, but comparisons of models are needed to find most appropriate one. Such comparisons were made among weight–age models for animals. Kolluru (2000) studied only Logistic model under heteroscedastic error condition for cattle growth. Therefore, there is a need to study other models also. Two models are taken for the present study; these are Logistic and Gompertz models. When heterogeneity of variance is evident, ordinary least square estimate of parameters may be inefficient as well as when weights are collected over time for each cattle, serial correlation is often present. In the present study Logistic and Gompertz models are fitted incorporating heteroscedasticity of variance along with autocorrelation.

2. MATERIAL AND METHODS

2.1 Data Description

Data used in the study were collected from history sheets of cattle from birth to 36 months of age from military dairy farms at Dehradun for Friesian × Sahiwal breed for 40 cattle. For Friesian × Sahiwal × Hariana breed data for 35 cattle were taken for comparing the growth pattern.

Let us consider the following model

$$y_j = f(\boldsymbol{X}_j, \beta) + e_j \qquad (2.1)$$

where y_j is the j^{th} observation, X_j is covariate vector, β is parameter vector, e_j is the error term and f is a non-linear function. In the present study two non-linear models have been considered viz. Logistic and Gompertz Models. The functional forms of these models are as follows

(i) Logistic model: $f(t, \beta) = \dfrac{\beta_1}{1+\beta_2 exp(-\beta_3\, t)}$

(ii) Gompertz model: $f(t, \beta) = \beta_1 \exp(-\beta_2 e^{-\beta_3 t})$

Here in $f(\boldsymbol{X}_j, \beta)$, the covariate vector $\boldsymbol{X}_j$ is replaced by t, which is the only covariate in the model. Usually, it is assumed that (i) the errors e_j have zero means, (ii) the errors e_j are uncorrelated, (iii) the errors e_j have common variance and (iv) the errors e_j are normally distributed. In case of the animal growth data, many a times, the above assumption are violated; errors are generally correlated and do not have common variance. In the present study we, therefore, fitted these models considering non-constant and correlated error variances. However, for the sake of comparison, models are also fitted under homoscedastic error variance.

2.2 Models with Heteroscedastic Error Structure along with Autocorrelation

2.2.1 Heteroscedastic error structure

In the present study data revealed heteroscedasticity of error variance. Heteroscedasticity of variance is tested by Rank correlation test

$$R_e = 1 - \{6\Sigma d_i^2 / (n\,(n^2 - 1))\}$$

where, d_i = difference between the ranks of corresponding value of y_i and e_i. A high rank correlation suggests the presence of heteroscedasticity.

Let us consider the model as given in (2.1). As mentioned earlier assumption of constant intra-individual response variance is violated frequently for growth data. Generally, growth data often exhibit constant coefficient of variation rather than constant variance (Davidian and Giltinan 1995); that is, variance is proportional to the squares of the mean response. In this case, a more appropriate assumption is

$$E(y_j) = f(\boldsymbol{X}_j, \beta),\ V(y_j) = \sigma^2\,[f(\boldsymbol{X}_j, \beta)]^2 \qquad (2.2)$$

where, σ a scale parameter, is the coefficient of variation. Under such situation, where variance is nonconstant across the response range, it is assumed that the variances of y_j are known up to a constant of proportionality, (Davidian and Giltinan 1995), that is,

$$V(y_j) = \sigma^2/w_j \qquad (2.3)$$

for some constants w_j, $j = 1, \ldots, n$. This type of setting might arise in the case where the responses y_j are themselves averages of w_j uncorrelated replicate measurements, with all such measurements having common variance σ^2. Under this model, except for the multiplicative constant σ^2, variance is known up to the value of the regression parameter β, 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 β by a suitable estimate, and to apply the weighted least squares idea.

The OLS estimator $\hat{\beta}_{OLS}$ is a natural choice to use for construction of estimated weights. An estimator for β that takes into account the assumed mean-variance relationship may be obtained by forming estimated weights (Davidian and Giltinan 1995)

$$\hat{w}_j = \frac{1}{f^2(\boldsymbol{X}_j, \hat{\beta}_{OLS})} \tag{2.4}$$

2.2.2 Auto-correlated structure

In the present study, the presence of autocorrelation in the data was checked by Durbin–Watson test. The Durbin–Watson statistic is given by

$$d = \sum_{u=2}^{n} (e_u - e_{u-1})^2 / \sum_{u=1}^{n} e_u{}^2$$

which ranges between 0 and 4. Value of d near 2 indicates no autocorrelation, a value towards 0 indicates positive autocorrelation, while the value towards 4 indicates negative autocorrelation. When there is evidence for the presence of autocorrelation, we find the value of auto correlation and model is fitted accordingly.

Correlation among observations on a given individual (cattle) is more likely to be present in this context (weight). In many cases, a systematic pattern of correlation is evident, which may be characterized accurately by a relatively simple model. To accommodate intra-individual correlation, a description of the assumed correlation pattern among the elements of e (error vector) is made. This assumption will be simple if the observations are taken at equal intervals. But the situation is quite complex when the unequally spaced observations are considered in the present study. Suppose that

$$\text{Corr}\ (e) = \Gamma(\alpha) \tag{2.5}$$

where the correlation matrix $\Gamma(\alpha)$ is a function of a vector of correlation parameters α. The choice of a suitable correlation matrix depends on the nature of the repeated measurement factor. As a special case where the repeated observations are taken over time, standard models for serial correlation patterns are available, i.e, the autoregressive (AR) model of order one. For definiteness, the observations are assumed to be indexed in the order in which they were collected. In the simplest case where the n repeated measurements are equally spaced in time, if the correlation between two adjacent observation is α, then the correlation between any two measurements j_1 and j_2 is given by

$$\text{Corr}\left(e_{j_1}, e_{j_2}\right) = \alpha^{|j_1 - j_2|} \tag{2.6}$$

The AR(1) correlation pattern may be generalized to accommodate situations where the observations are not equally spaced (see, Liang and Zeger (1986) and Chi and Reinsel (1989)) . If j_1 and j_2 are measurements taken at times t_{j_1} and t_{j_2} respectively where $j_1 \neq j_2$ then

$$\text{Corr}\ (e_{j_1}, e_{j_2}) = \alpha^{|t_{j_1} - t_{j_2}|} \tag{2.7}$$

This may be expressed by the correlation matrix as

$$\Gamma(\alpha) = \begin{bmatrix} 1 & \alpha^{(t_2-t_1)} & \alpha^{(t_3-t_1)} & . & . & . & \alpha^{(t_n-t_1)} \\ & 1 & \alpha^{(t_3-t_2)} & . & . & . & \alpha^{(t_n-t_2)} \\ & & . & . & & & . \\ & & & . & . & & . \\ & & & & . & . & . \\ & & & & & . & \alpha^{(t_n-t_{n-1})} \\ & & & & & & 1 \end{bmatrix} \tag{2.8}$$

The method of estimation of α is described in the next sub-section.

2.2.3 Auto-correlation for unequally spaced observations

When observations are taken on subjects at arbitrary time points, there must be an underlying continuous time process (Jones 1981, Diggle 1988, Chi and Reinsel 1989). For equally spaced observations, there may or may not be an underlying continuous time process. Unequally spaced observations differ from equally spaced observations with some missing observations in that there is no basic sampling interval. The mathematical model for a continuous time AR(1) process, denoted as CAR(1), (Jones and Boadi-Boateng 1991), is given by

$$\frac{\mathrm{d}}{\mathrm{dt}}\varepsilon(t) + \alpha\varepsilon(t) = G\eta(t) \tag{2.9}$$

where G is a constant, G^2 is variance per unit time, $\eta(t)$ is continuous time 'white noise', $\Sigma(t)$ is error at t^{th} time and α is the correlation between two adjacent observations. A model for $\eta(t)$ is a differential equation given by

$$\eta(t) = \frac{d}{dt}\omega(t) \qquad (2.10)$$

where $\omega(t)$ is Brownian motion or a Wiener process.

Combining equations (2.9) and (2.10), we get

$$d\varepsilon(t) + \alpha\varepsilon(t)dt = Gd\omega(t) \qquad (2.11)$$

This equation is solved by integration. A solution of (2.11) with the random input removed is given by (see, Jones 1993)

$$\frac{d}{dt}\varepsilon(t) + \alpha\varepsilon(t) = 0 \qquad (2.12)$$

If (2.12) is integrated from time t_1 to time t_2, the solution is a prediction

$$\varepsilon(t_2) = \exp\{-\alpha(t_2 - t_1)\}\varepsilon(t_1) \qquad (2.13)$$

The solution, (2.13) is now in the form of a discrete time AR(1) process with an auto regression coefficient α. The equation (2.13) can be generalized as

$$\varepsilon(t_n) = \exp\{-\alpha(t_n - t_{n-1})\}\varepsilon(t_{n-1}) \qquad (2.14)$$

Fitting equation (2.14) by nonlinear modeling techniques, one can estimate α.

2.2.4 Computational aspects

Once the weights are calculated by (2.4) by using OLS estimator $\hat{\beta}_{OLS}$, they form the weight matrix for all observations. Let us denote this matrix as W, a diagonal matrix whose diagonal elements are the weights estimated through (2.4). Then the variance-covariance for y under heteroscedastic model is

$$\text{Cov}(y) = \sigma^2 \boldsymbol{W} \qquad (2.15)$$

For the model with heteroscedastic error variance along with auto-correlation, the variance-covariance matrix of y becomes

$$\text{Cov}(y) = \sigma^2 \boldsymbol{W}^{1/2}\, \Gamma(a)\, \boldsymbol{W}^{1/2} = \Sigma \text{ (say)} \qquad (2.16)$$

where $\Gamma(\alpha)$ is as given in (2.8). After estimating α by the model (2.14), it is incorporated in (2.16).

Now applying Generalized Least Squares principle, the one-step ahead estimates of parameters can be obtained by minimizing

$$(y - f(\boldsymbol{X}, \beta_{OLS}))' \Sigma^{-1} (y - f(\boldsymbol{X}, \beta_{OLS})) \qquad (2.17)$$

After getting new estimates of β, weights are again estimated and another set of parameters are estimated through (2.17). The process is continued till the values of β converges. Final value of estimates of β is represented as β_{GLS}.

2.3 Measure of Model Adequacy

The empirical comparison of models can be made using with goodness of fit statistics such as RMSE and RMAPE. Lower the values of RMSE and RMAPE better are the models. It is concluded that the model which has minimum RMSE and RMAPE gives better parameters of the fitted model. For calculating the RMSE and RMAPE following formulas are used.

Root mean squared error (RMSE)

$$= \left[\sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n-p}\right]^{1/2} \qquad (2.18)$$

Relative mean absolute percentage error (RMAPE)

$$= \frac{1}{n}\sum_{i=1}^{n} \frac{|Y_i - Y|}{Y_i} \times 100 \qquad (2.19)$$

where Y_i is original value, $\hat{Y}_i$ is predicted values or estimated value and n is the total number of observations, p is the number of parameters.

3. RESULTS AND DISCUSSION

Models are first fitted under homoscedastic error structure. For this purpose SAS package Version 9.1 has been used. In the present study the data revealed heteroscedasticity of error variance as rank correlation is found to be 0.6920 for F×S×H breed and 0.6210 for F×S breed. In the present study data also have auto-correlation, when checked using Durbin–Watson test. The Durbin–Watson statistic (d) is found to be 0.8120

Table 1. Auto-correlation of different breed by different model

Name of Breed	Model	Auto-correlation
F×S	Logistic	0.1254
F×S	Gompertz	0.1330
F×S×H	Logistic	0.4448
F×S×H	Gompertz	0.4319

Table 2. Parameter estimates of models under homoscedastic error variance and heteroscedastic error variance along with auto-correlation of Friesian × Sahiwal breed at Dehradun station

Parameters	Under homoscedastic error variance		Under heteroscedastic error variance along with autocorrelation	
	Logistic model	Gompertz model	Logistic model	Gompertz model
β_1	354.4000 (15.6193)	382.5000 (15.7104)	299.9152 (3.3298)	328.5822 (1.7363)
β_2	6.6555 (0.8399)	2.3190 (0.0966)	9.4134 (0.1504)	2.5333 (0.0064)
β_3	0.1534 (0.0142)	0.0927 (0.0072)	0.2643 (0.0035)	0.1377 (0.0009)
Goodness of fit statistics				
RMSE	16.7320	11.3109	0.1210	0.0908
RMAPE	21.4587	13.4961	12.9276	8.1148
Autocorrelation	-	-	0.1254	0.1330

Note: Figures in the brackets indicate standard errors.

for F×S×H breed and 0.8394 for F×S breed. α was estimated by fitting equation (2.14) by nonlinear modeling technique. The value of α (auto-correlation) is obtained through NLIN option of SAS procedure. The estimated values of autocorrelation for different breeds are given in Table 1.

For fitting models under heteroscedastic error structure along with autocorrelation, program is written in SAS/IML language.

It can be observed from Table 2 that for F×S breed at Dehradun farm RMSE (11.3109) is less for Gompertz model than RMSE (16.7320) by Logistic model and similarly RMAPE(13.4961) is Less for Gompertz model than RMAPE(21.4587) by Logistic model, this shows that results of Gompertz model are better than logistic model under homoscedastic error condition. The data of the breed are having heteroscedasticity of variance and autocorrelation which was tested by Durbin Watson test. Auto correlation for this breed is found to be 0.1254 and 0.1330 of Logistic and Gompertz models respectively.When the results under homoscedastic error structure and heteroscedastic error structure along with autocorrelation are compared it is found that RMSE and RMAPE under heteroscedastic error structure with auto correlation are found less than homoscedastic error structure for both models, this shows that results for heteroscedastic error structure are better than homoscedastic error structure.

From Table 3 it is observed that for F×S×H breed RMSE (23.9571) by Gompertz model is less than RMSE (24.3640) by logistic model under homoscedastic error structure and RMAPE (7.1689) by Gompertz model is less than RMAPE (13.3725) by Logistic model, so results of Gompertz model are better than Logistic model under homoscedastic error structure. RMSE (0.2942) and RMAPE (7.2071) of Logistic model and RMSE (0.2975) and RMAPE (4.3462) of Gompertz model under heteroscedastic error structure along with auto correlation are less than RMSE and RMAPE under homoscedastic error structure. This shows that results under heteroscedastic error structure with autocorrelations are better than homoscedastic error structure. Mature weight is found to be more for F × S breed than mature weight of F × S × H breed under homoscedastic error structure where as mature weight found more for F × S × H breed than F × S breed under heteroscedastic error structure along with autocorrelation. Growth rate was found to be better for F × S × H than F × S breed.

Table 3. Parameter estimates of different models under homoscedastic error variance and heteroscedastic error variance along with auto-correlation of Friesian × Sahiwal × Hariana breed at Dehradun station

Parameters	Under homoscedastic error variance		Under heteroscedastic error variance along with autocorrelation	
	Logistic model	Gompertz model	Logistic model	Gompertz model
β_1	327.8000 (17.2990)	349.2000 (23.9764)	362.7186 (8.6022)	475.8925 (5.2318)
β_2	8.5121 (1.9712)	2.5636 (0.2814)	13.3927 (1.4666)	2.9731 (0.1243)
β_3	0.1949 (0.0262)	0.1156 (0.0176)	0.2429 (0.0179)	0.0941 (0.0100)
Goodness of fit statistics				
RMSE	24.3640	23.9571	0.2942	0.2975
RMAPE	13.3725	7.1689	7.2071	4.3462
Autocorrelation	-	-	0.4448	0.4319

Note: Figures in the brackets indicate standard errors.

REFERENCES

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. *J. Anim. Sci*., **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. *J. Anim. Sci.*, **43**, 810-818.

Chi, E.M. and Reinsel, G.C. (1989). Models for longitudinal data with random effects and AR(1) errors. *J. Amer. Statist. Assoc.,* **84**, 452-459.

Davidian, D. and Giltinan, M. (1995). *Non Linear Models for Repeated Measurement Data*. Chapman Hall, London.

Diggle , P.J. (1988). An approach to the analysis of repeated measurement data. *Biometrics,* **45**, 1255-1258.

Jones, R.H. and Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics,* **47**, 161-175.

Jones, R.H. (1993). *Longitudinal Data with Serial Correlation: A State-Space Approach.* Chapman and Hall, London.

Kolluru, R. (2000). On some aspect of growth patterns of crossbred cattle. Unpublished M.Sc. Thesis, Indian Agricultural Research Institute, New Delhi.

Kolluru, R., Rana, P.S. and Paul, A.K. (2003). Modelling for growth pattern in crossbred cattle. *J. Anim. Sci.,* **73(10),** 1174-1179.

Liang, K.Y. and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika,* **73,** 13-22.

SAS (1990). SAS Users' Guide Version 6. SAS Institute Incorporation, U.S.A.