*J. Ind. Soc. Agril. Statist.*
*59(3), 2005 : 237–242*

# On Appropriate Reparameterization of a Nonlinear Statistical Model

C. Sarada and Prajneshu[1]
*Central Institute of Brackishwater Aquaculture, Chennai*
(Received : November, 2005)

**SUMMARY**

Procedure is thoroughly discussed for arriving at the appropriate reparameterization in respect of a nonlinear statistical model with a view to achieving *close-to-linear* (Ratkowsky (1990)) behaviour. Details are given to compute the two root-mean-square measures of curvature for assessment of nonlinearity for all parameters combined as well as marginal curvatures for individual parameters. The method for identifying appropriate reparameterization is then discussed through the use of "Expected-value parameters" and "Simulation" techniques. As an illustration, a nonlinear statistical model for aphid population growth (Prajneshu (1998)) is considered.

*Key words :* Expected-value parameters, Marginal curvature, Measures of nonlinearity, Profile-t plots, Reparameterization, Simulation studies.

## 1. INTRODUCTION

It is well recognized that any type of statistical inquiry in which principles from some body of knowledge enter seriously into the analysis is likely to lead to a 'Nonlinear model'. Accordingly, in recent years, a number of such models have been developed in various disciplines. These models are generally "mechanistic" in nature as the underlying parameters have specific biological interpretations. Quite often these arise as the solutions of differential, or integro-differential, or difference equations. Usually the functional forms obtained as such are fitted to the data resulting in highly nonlinear behaviours. Unfortunately, hardly any attention is paid to the various reparameterizations and consequently, the parameter estimates generally do not satisfy, even remotely, any optimum properties, like that of unbiasedness, minimum variance, distribution being normal (Ratkowsky (1990)). Therefore, determination of the optimum reparameterization for a nonlinear model is of great relevance.

Accordingly, purpose of the present paper is to make a systematic study of above aspect. Next section discusses briefly two root mean square measures of curvature as well as marginal curvatures for individual parameters. Procedure for providing guidelines for replacing nonlinear parameters by ones with better statistical properties is also described. In Section 3, as an illustration, a nonlinear statistical model, applied to aphid count data on potato leaves, is considered. S-Plus software package and a FORTRAN program, written by Kang and Rawlings (1998), are used for data analysis. Optimum parameterizations of original parameters of the nonlinear statistical model for describing dynamics of aphid population growth are obtained. Residual analysis for the model with independently and identically distributed (i.i.d.) error terms revealed presence of first order autoregressive (AR(1)) errors. Finally, the parameterized model with AR(1) errors not only provided an extremely good fit but also ensured desirable properties for the parameter estimates.

## 2. PROCEDURE FOR REPARAMETERIZATION

### 2.1 Measures of Nonlinearity

In order to examine the extent of nonlinearity of a model, the two root mean square (rms) measures of curvature, viz. intrinsic curvature (IN) and parameter-

[1] *Indian Agricultural Statistics Research Institute, New Delhi - 110 012*

effects curvature (PE) need to be computed. The details are given in Bates and Watts (1988). IN is inherent to the model and can not be changed by any reparameterization, while PE is dependent on a particular parameterization. A low IN value ensures nearly unbiased predicted values of the response variate. Furthermore, low values of IN as well as PE ensure joint confidence regions for the parameters to be close to being ellipsoidal and confidence-limits for individual parameters to be close to being symmetrical.

PE needs to be computed only when IN is within permissible limits. A low value of PE indicates that the model exhibits *close-to-linear* (Ratkowsky (1990)) behaviour; otherwise an appropriate reparameterization should be identified so as to reduce PE as much as possible. It is desirable that both IN and PE measures multiplied by $\sqrt{F_{p,n-p}(\alpha)}$ should be small, say less than 0.3. Here p is the number of parameters of the model, n is the number of observations and $F_{p,n-p}(\alpha)$ denotes the upper $\alpha\%$ point of $F_{p,n-p}$-distribution.

One disadvantage of the above curvature measures is that they cannot identify those parameters which are primarily responsible for nonlinear behaviour of the model in a multiparameter situation. Accordingly, as a next step, marginal curvatures for individual parameters may be computed (Clarke (1987)). As a rule of thumb, if marginal curvature multiplied by $t_{n-p}(.05)$, where $t_{n-p}(.05)$, the upper 0.05 quantile of $t_{n-p}$-distribution, is less than 0.1, curvature effects may be ignored and linear approximation would suffice. Kang and Rawlings (1998) showed that, for a given parameterization, marginal curvature for transformed parameter can be computed without determining inverse transformation.

As pointed out by Bates and Watts (1988), extent of nonlinearity of individual parameter estimates can be assessed graphically by employing profile-t plots. These plots graphically reveal the extent of nonlinearity in individual parameter estimates. For a nonlinear model, profile-t function $\tau(\theta_p)$ can be computed as follows: Suppose the vector of parameters is $\theta = (\theta_1, \theta_2)^T$ and we wish to test $H_0: \theta_1 = \theta_{10}$. Let $\hat{\theta}$ be the overall least squares estimate, $\hat{\hat{\theta}}_{2|1}$, be the conditional least-squares estimate of $\theta_2$ with $\theta_1 = \theta_{10}$ and put $\hat{\hat{\theta}}(\theta_{10}) = (\theta_{10}, \hat{\hat{\theta}}_{2|1})^T$. Then sum of square function is given as

$$R(\theta_{10}) = RSS(\hat{\hat{\theta}}(\theta_{10})) - RSS(\hat{\theta}) \qquad (2.1)$$

where RSS denotes the residual sum of squares. The profile - t function defined as signed square root

$$\tau(\theta_{10}) = \text{Sign}(\theta_{10} - \hat{\theta})\sqrt{R(\theta_{10})}/S \qquad (2.2)$$

has an approximate t-distribution under $H_0$. Confidence-interval for $\theta_i$ is the set $\{\theta | -t < \tau(\theta_i) < t\}$, where t is approximate percentage point of $t_{n-p}$ distribution.

Plots of this profile-t function provides exact likelihood intervals for individual parameters and, in addition, it reveals how nonlinear situation is. For coordinate directions along which the approximate linear methods are accurate, plots of nonlinear t-statistic, $\tau(\theta_i)$ against $\theta_i$ over several standard deviations on either side of maximum likelihood should be straight. Any deviations from straightness serve as a warning that linear approximation may be misleading in that direction.

### 2.2 Choice of Suitable Reparameterization

After obtaining information about those parameters which deviate from linear behaviour, the next step is to choose suitable parameterization. Unfortunately, there are very few guidelines available for replacing such parameters by ones with better properties. Ratkowsky (1990) suggested that one way of achieving this is to find Expected-value parameters while another way is to perform Simulation studies.

#### *2.2.1 Expected-value Parameters*

These can be obtained by choosing p-values of explanatory variable X, p being the number of parameters. The new parameters are the expected- value parameters $y_1, y_2, \ldots, y_p$ after eliminating original parameters. They seem to be best kind of parameters as they invariably lead to parameters with good estimation properties. However, it has got a limitation that their defining equations often cannot be solved for the 'old' parameters in terms of the expected-value parameters. Therefore, one cannot always obtain an algebraic expression for the reparameterized model. This prevents universal application of this procedure.

#### *2.2.2 Simulation Studies*

By taking parameter estimates and estimate of residual variance about regression line of a model as true values, a large number, say 1000, of simulated data sets may be generated by allowing error term to change randomly (which is having normal distribution and required variance) at each value of explanatory variable for each data set. The estimates of parameters are then obtained for each generated data set and histograms are plotted for individual parameters. These histograms would provide an idea about parameters which are behaving nonnormally and also indicate probable reparameterization. A long right-hand tail suggests replacement of the parameter in the model function by an exponential of the parameter, while one with a long left-hand tail suggests replacement of parameter by logarithm of the parameter. Furthermore, distributional properties of each parameter can be studied by computing skewness coefficient $(\upsilon_1)$, kurtosis coefficient $(\upsilon_2)$ and bias.

For 1000 data sets, $\upsilon_1$ is approximately distributed as $N(0, (6/1000)^{1/2})$ and $\upsilon_2$ is approximately distributed as $N(0,(24/1000)^{1/2})$. Value of $\upsilon_1 > \pm 0.152$ (at 5% level) or $\pm$ 0.200 (at 1% level) and value of $\upsilon_2 > \pm 0.304$ (at 5% level) or $\pm$ 0.399 (at 1% level) indicates significance of skewness and kurtosis respectively. Further, absolute value of %bias greater than 1% is an indicator of nonlinear behaviour of the parameter.

## 3. RESULTS AND DISCUSSION

As an illustration, mean data of *Aphis gossypii* glover per 100 potato leaves recorded at Central Research Station, Modipuram, India at weekly intervals during 1983 to 1987 is considered (Verma and Parihar (1991)). Relevant data is reproduced in Table 1 for ready reference. The following nonlinear statistical model, hereinafter referred to as 'Model I' (Prajneshu (1998)) is fitted to this data:

$$N(t) = ae^{bt}(1+de^{bt})^{-2} + \varepsilon \quad (3.1)$$

where the error term $\varepsilon$ is assumed to be independently and identically normally distributed. Standard software packages, like SAS, SPSS can be employed to achieve the task. However, none of the packages, except perhaps S-plus, contain ready-made programs for computing rms curvature measures of nonlinearity and profile-t plots. Therefore, this package is used for statistical analysis of present data. The details are given in Venables and Ripley (1999).

The estimates of parameters a, b and d of Model I, along with their asymptotic standard errors and 95% confidence-intervals are presented in the second column of Table 2. The rms curvature measures are also reported in the same column. It may be noted that intrinsic nonlinearity (0.21) is within acceptable limits. However, parameter-effects nonlinearity (12.23) is extremely high, indicating thereby the need for reparameterization of parameters. In order to identify the parameters which are causing such a high value, marginal curvatures of individual parameters are computed using FORTRAN program NLIN-CURV.ED of Kang and Rawlings (1998). A subroutine MODEL has to be supplied to the program, which requires first and second derivatives of Model I, given below:

$$\partial N(t)/\partial a = e^{bt}/(1+de^{bt})^2 \quad (3.2)$$

$$\partial N(t)/\partial b = tae^{bt}(1-de^{bt})/(1+de^{bt})^3 \quad (3.3)$$

**Table 1.** *Aphis gossypii* glover per 100 potato leaves at weekly intervals during 1983-87

| Time (weeks) | Number of aphids per 100 leaves |
|---|---|
| 0 | 0.0 |
| 1 | 6.5 |
| 2 | 7.6 |
| 3 | 6.6 |
| 4 | 13.0 |
| 5 | 5.3 |
| 6 | 6.7 |
| 7 | 16.3 |
| 8 | 16.0 |
| 9 | 16.0 |
| 10 | 20.6 |
| 11 | 35.6 |
| 12 | 53.0 |
| 13 | 86.8 |
| 14 | 75.3 |
| 15 | 33.8 |
| 16 | 17.3 |
| 17 | 8.0 |
| 18 | 4.5 |
| 19 | 0.0 |
| 20 | 0.5 |
| 21 | 0.3 |
| 22 | 1.5 |

**Table 2.** Parameter estimation, rms curvatures measures, marginal curvature, measures of nonlinearity and simulation studies

| | | Model I | Model II |
|---|---|---|---|
| (i) | Parameter estimation: | | |
| | a | 0.0018 | − 6.34 |
| | | (0.0017) | (0.99) |
| | | [0.001, 0.0070] | [−8.17, − 4.76] |
| | b | 0.92 | 0.92 |
| | | (0.08) | (0.08) |
| | | [0.800, 1.0700] | [0.80, 1.07] |
| | d | $0.0^555$ | − 12.13 |
| | | $(0.0^559)$ | (1.67) |
| | | [0.000, 0.0002] | [− 14.03, − 10.04] |
| (ii) | rms Curvature effects: | | |
| | (IN) $\sqrt{F_{3,20}(.05)}$ | 0.21 | 0.21 |
| | (PE) $\sqrt{F_{3,20}(.05)}$ | 12.23 | 0.24 |
| (iii) | (Marginal curvatures) $t_{20}(.05)$: | | |
| | a | 0.78 | 0.08 |
| | b | 0.07 | 0.07 |
| | d | 0.82 | 0.07 |
| (iv) | Simulation Studies: | | |
| | Skewness: | | |
| | a | 0.61** | $-0.18^{NS}$ |
| | d | 0.61** | $-0.10^{NS}$ |
| | Kurtosis: | | |
| | a | − 1.06** | $0.29^{NS}$ |
| | d | 1.05** | $0.27^{NS}$ |
| | %Bias: | | |
| | a | − 23.30** | $0.02^{NS}$ |
| | d | − 24.02** | $0.01^{NS}$ |

Note: Figures in parentheses are asymptotic standard errors. Figures in square brackets are 95% lower and upper confidence limits respectively.

** indicates significant at 1% level and NS indicates non-significant.

$$\partial N(t)/\partial d = -2a(e^{bt})^2/(1+de^{bt})^3 \quad (3.4)$$

$$\partial^2 N(t)/\partial a^2 = 0 \quad (3.5)$$

$$\partial^2 N(t)/\partial a \partial b = tae^{bt}(1-de^{bt})/\ (1+de^{bt})^3 \quad (3.6)$$

$$\partial^2 N(t)/\partial b^2 = t^2 ae^{bt}[(1-de^{bt})^2 - 2de^{bt}]/(1+de^{bt})^4 \quad (3.7)$$

$$\partial^2 N(t)/\partial a \partial d = -2(e^{bt})^2/(1+de^{bt})^3 \quad (3.8)$$

$$\partial^2 N(t)/\partial b \partial d = -2ta(e^{bt})^2(2-de^{bt})/(1+de^{bt})^4 \quad (3.9)$$

$$\partial^2 N(t)/\partial d^2 = 6a(e^{bt})^3/(1+de^{bt})^4 \quad (3.10)$$

Detail of MODEL subroutine supplied to the main program NLIN_CURV.ED is given in Annexure-I of the website www.ciba.nic.in/cs/annex.htm (Sarada (2005)).

The results of computation in respect of Model I are presented in column 2 of Table 2. Marginal curvatures for parameter estimates of a (0.78) and d (0.82) are significantly higher than 0.1 while that for the parameter b (0.07) is within acceptable limits. Profile-t plots (Fig. 1) also reveal that parameters a and d are deviating from planar assumption, suggesting thereby the need for a suitable reparameterization for these parameters. Unfortunately, Expected-value parameters cannot be obtained for (3.1) as parameters b and d cannot be eliminated. So, with a view to having an idea about reparameterization, simulation studies are carried out. One thousand simulated samples are generated and coefficient of skewness, kurtosis and %bias are calculated for parameters a and d and the results are presented in column 2 of Table 2. The histograms plotted for these parameters are presented in Fig. 2. The two histograms are having long right-hand tail, which implies that parameters should be replaced by their exponentials.

Accordingly, the model in (3.1) is reparameterized as

$$N(t) = e^{a+bt}(1+e^{d+bt})^{-2} + \varepsilon \quad (3.11)$$

which is referred to as Model II in subsequent discussion. The marginal curvatures for the transformed model are computed by NLIN-CUR.ED program by supplying to it TRANSF subroutine, which is given in Annexure-II of the website www.ciba.nic.in/cs/annex.htm (Sarada (2005)). All the above statistics are now computed for Model II and the results are presented in column 3 of Table 2. The intrinsic nonlinearity (0.21) remains the same, as expected.

Further, the marginal curvatures of all the three parameter estimates are now within acceptable limits, and the parameter-effects nonlinearity reduces to 0.24, which being less than 0.3, indicates that now it is not significant. Further, the % bias for parameter estimates of a and d has also reduced drastically. The profile-t plots

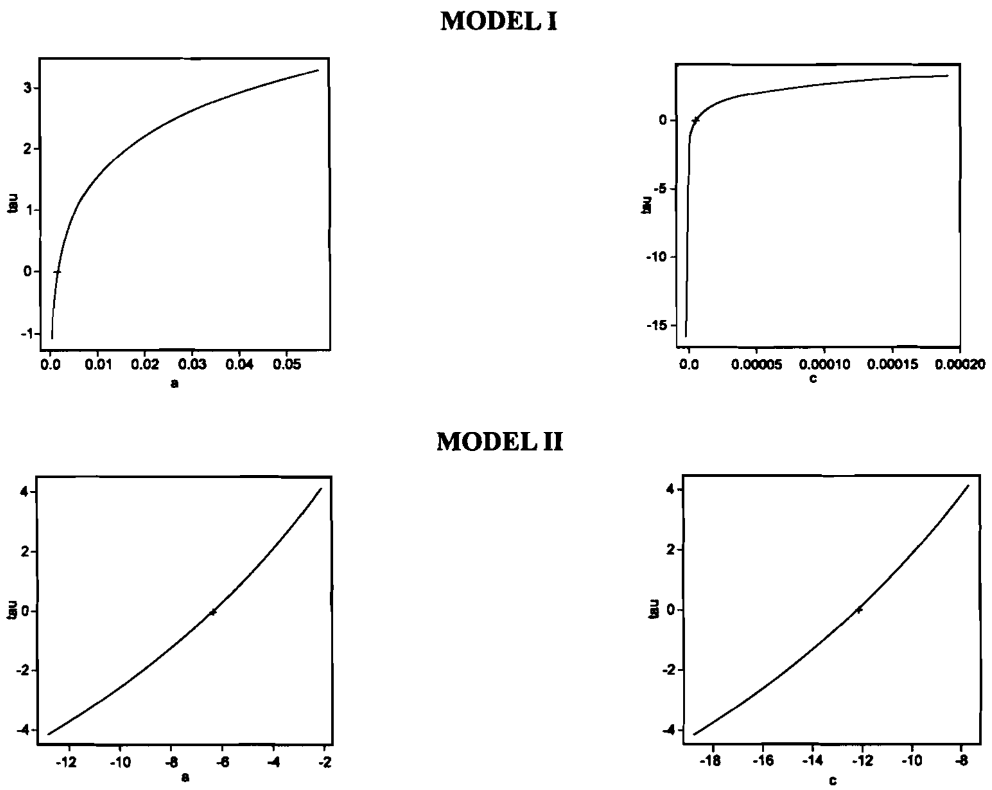


**Fig. 1**. Profile-t plots for Model I and Model II

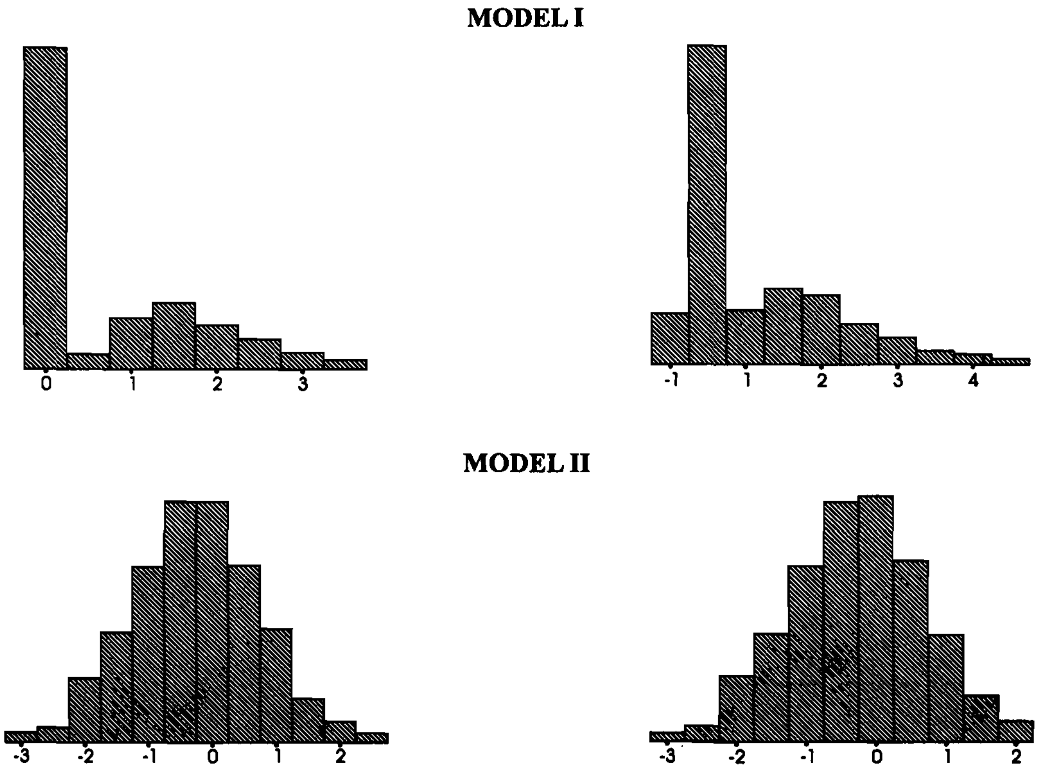


**Fig. 2**. Histograms of 1000 simulated standardized parameter estimates of Model I and Model II

are also now nearly straight (Fig. 1) and histograms for the two parameter estimates are also symmetric (Fig. 2). Thus, Model II is a suitable reparameterized model for the given situation. A perusal of Table 2 shows that reparameterization has resulted in a substantial decrease in percentage asymptotic standard errors for parameter estimates. In fact, for parameter estimate of a, it has decreased from 98.9% to 15.6% while for the parameter estimate of d, it has come down from 107.4% to 8.5%. Similar results hold in respect of width of confidence-intervals and these have also become more symmetric.

Further, results of residual analysis for Model II are presented in column 2 of Table 3. The value of Run test statistic (-2.98) lies in the critical region at 5% level and so errors are not independent. Durbin-Watson test statistic (0.77) near to zero indicates possibility of AR(1) error structure. Therefore, Model II with AR(1) error structure is fitted to the data and the results are reported in column 3 of Table 3. A perusal indicates that the assumption of independence of errors is now not rejected at 5% level. Hence Model II with AR(1) error structure is appropriate for describing the given data.

**Table 3**. Summary statistics for fitting Model II to *Aphis gossypii* glover data

| | Statistics/Error | i.i.d. | AR (1) |
|---|---|---|---|
| (i) | Parameter estimates: | | |
| | a | – 6.36 | – 11.88 |
| | | (1.02) | (1.55) |
| | b | 0.93 | 1.32 |
| | | (0.08) | (0.11) |
| | d | – 12.15 | – 17.62 |
| | | (1.06) | (1.51) |
| | AR (1) | | 0.89 |
| | | | (0.10) |
| (ii) | Goodness of fit statistics: | | |
| | RMSE | – | 4.70 |
| (iii) | Residual analysis: | | |
| | Run test (Z) | – 2.98 | 0.00 |
| | Durbin-Watson test | 0.72 | 2.21 |

**ACKNOWLEDGEMENT**

Authors are grateful to the referee for valuable comments.

**REFERENCES**

Bates, D.M. and Watts, D.G. (1988). *Nonlinear Regression Analysis and its Applications*. John Wiley, New York.

Clarke, G.P.Y. (1987). Marginal curvatures and their usefulness in the analysis of nonlinear regression models. *J. Amer. Statist. Assoc.*, **82**, 844-850.

Kang, G. and Rawlings, J. (1998). Marginal curvatures for function of parameters in nonlinear regression. *Statis. Sin.*, **8**, 467-476.

Prajneshu (1998). A nonlinear statistical model for aphid population growth. *J. Ind. Soc. Agril. Statist.*, **51**, 73-80.

Ratkowsky, D.A. (1990). *Handbook of Nonlinear Regression Models*. Marcel Dekker, New York.

Sarada, C. (2005). www.ciba.nic.in/cs/annex.htm

Venables, W.N. and Ripley, B.D. (1999). *Modern Applied Statistics with S-plus: Data Analysis (Statistics and Computing)*. Third edition. Springer–Verlag, New York.

Verma, K.D. and Parihar, S.B.S. (1991). Build up of the vector *Aphis gossypii* glover on potato. *J. Aphidol.*, **5**, 16-18.