

9-2005

ANALYSIS OF NONLINEAR REGRESSION MODELS: A CAUTIONARY NOTE

Shyamal D Peddada

National Institutes of Health, Research Triangle Park, NC

Joseph K Haseman

National Institutes of Health, Research Triangle Park, NC

Follow this and additional works at: https://scholarworks.umass.edu/dose_response

Recommended Citation

Peddada, Shyamal D and Haseman, Joseph K (2005) "ANALYSIS OF NONLINEAR REGRESSION MODELS: A CAUTIONARY NOTE," *Dose-Response: An International Journal*: Vol. 3 : Iss. 4 , Article 7.

Available at: https://scholarworks.umass.edu/dose_response/vol3/iss4/7

This Article is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Dose-Response: An International Journal by an authorized editor of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

ANALYSIS OF NONLINEAR REGRESSION MODELS: A CAUTIONARY NOTE

Shyamal D. Peddada and Joseph K. Haseman □ Biostatistics Branch, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC

□ Regression models are routinely used in many applied sciences for describing the relationship between a response variable and an independent variable. Statistical inferences on the regression parameters are often performed using the maximum likelihood estimators (MLE). In the case of nonlinear models the standard errors of MLE are often obtained by linearizing the nonlinear function around the true parameter and by appealing to large sample theory. In this article we demonstrate, through computer simulations, that the resulting asymptotic Wald confidence intervals cannot be trusted to achieve the desired confidence levels. Sometimes they could underestimate the true nominal level and are thus liberal. Hence one needs to be cautious in using the usual linearized standard errors of MLE and the associated confidence intervals.

Keywords: confidence interval, coverage probability, variance estimation

1. INTRODUCTION

Linear and nonlinear statistical models are widely used in many applications to describe the relationship between a response variable Y and an explanatory variable X . A statistical model is said to be linear if the mean response is a linear function of the unknown parameters, otherwise it is said to be a nonlinear model. For example, in the context of fertilizer trials the mean yield of corn is sometimes modeled as a function of dosage X by the quadratic function

$$a + bX - \frac{c}{2}X^2$$

The above model is linear in the unknown parameters a , b and c . In animal carcinogenicity studies and risk assessment, often researchers model the mean response to different doses X of a chemical by the following Hill model (Kim *et al.*, 2002, Portier *et al.*, 1996, Walker *et al.*, 1999):

$$a + b \frac{X^d}{c^d + X^d}, \quad (1)$$

where a represents the baseline response, $a + b$ denotes the maximum response, c denotes the ED_{50} (i.e. effective dose corresponding to 50% of

Address correspondence to Shyamal D. Peddada, Biostatistics Branch, National Institute of Environmental Health Sciences (NIH), Research Triangle Park, NC 27709. E-mail: peddada@niehs.nih.gov.

Analysis of Nonlinear Regression Models

the maximum response from the baseline response) and d is the slope parameter. Since some of the parameters enter the above model nonlinearly this is a nonlinear model.

One of the purposes of fitting regression models is to draw inferences on unknown parameters, or their functions, which have some physical interpretation. For example, in the case of fertilizer trials a researcher is often interested in estimating the “optimum dose” which maximizes the corn yield. From the above quadratic function, this parameter is given by b/c , a nonlinear function of the regression parameters b and c . In the case of animal carcinogenicity studies, in addition to estimating a , b , c and d , researchers are often interested in estimating the effective dose corresponding to e % of the maximum response from the baseline response. This parameter is denoted by ED_e . Typical parameters of interest are ED_{01} and ED_{10} (Portier *et al.*, 1996, Walker *et al.*, 1999), which are nonlinear functions of a , b , c and d .

A key step in the statistical inference on unknown parameters of a model is to compute the standard errors of various estimates. If the statistical model is either nonlinear or the parameter of interest in a linear model is a nonlinear function of the regression parameters, then the *approximate* standard errors are usually derived by using the first order term in a suitable Taylor’s series expansion. Once the approximate standard errors are obtained the Wald type confidence intervals such as those given in (3), (4) (see the Appendix) are derived. Such confidence intervals are used very extensively in applications.

Suppose θ is an unknown parameter of interest and suppose $\hat{\theta}$ is its MLE with standard error $S.E.(\hat{\theta})$. The *coverage probability* of a $(1 - \alpha) \times 100\%$ confidence interval for a parameter θ , where z_α is the suitable critical value, is described as follows. Suppose for each random realization of data one was to construct the above confidence interval, then the *coverage probability* of the confidence interval is the proportion of all such intervals that contain the true parameter θ . A confidence interval is said to be *accurate* if $(1 - \alpha) \times 100\%$ of all such intervals contain θ . A confidence interval formula is said to be *liberal* if its coverage probability is less than $(1 - \alpha)$ and is said to be *conservative* if its coverage probability exceeds $(1 - \alpha)$.

Under some conditions on the linear model the Wald confidence intervals (3) and (4) are accurate when the parameter of interest is a linear function of the regression parameters. However, if the parameter is either a non-linear function of the regression parameters or if the model is a nonlinear model, then they are not necessarily accurate, unless the sample sizes are “very large”. Basically the large sample theory confidence intervals are derived by “linearizing” the nonlinear function. This is accomplished by approximating the function by the first order derivative term in the Taylor series expansion of the nonlinear function. Hence for (3) and (4) to be accurate it is important that the second and higher

S. D. Peddada and J. K. Haseman

order terms in the Taylor series expansion are “negligible” in comparison to the first order term. The effect of the second order term is known as the “curvature effect.”

The purpose of this article is to demonstrate through computer simulations that, in some instances, the standard error of the MLE based on the above linearization process can be a severe underestimate of the true standard error of the MLE. Consequently, (3) and (4) can be liberal and are not trustworthy. As an alternative, we consider confidence intervals based on the sandwich estimator of the covariance matrix of MLE introduced in Zhang (1997) and Zhang *et al.* (2000a). We notice that, to some extent, the intervals based on Zhang *et al.* (2000a) methodology correct this problem.

A second problem that is often associated with nonlinear regression analysis is the numerical computation of maximum likelihood estimates. Usually the computation of MLE is based on an iterative process that requires carefully chosen initial starting points to avoid convergence to local optima. Depending upon the nonlinear function this can be a challenging problem. For a good description regarding this issue one may refer to Ratkowsky (1990). Usually it is highly recommended to apply the iterative process by choosing a large number of starting points and choose the best solution among all such solutions. It is important to note that a poor approximation to the true MLE may result in a poor estimate of the standard error, thus compounding the previously mentioned concern regarding the estimation of standard errors.

All notations and formulas are provided in the Appendix of the paper.

2. ESTIMATION OF STANDARD ERRORS AND CONFIDENCE INTERVALS

2.1 Curvature effects and the coverage probability problem

Several authors have noted that the “usual formulas” underestimate (or overestimate) the true standard errors of the estimates of parameters in a nonlinear model. This results in very high (or very low) false positive rates when performing test of hypothesis and liberal (or too conservative) confidence intervals. That is, the confidence intervals could be too narrow or too wide. For example, Simonoff and Tsai (1986) performed extensive simulation studies using three different nonlinear models to demonstrate that the true coverage probability of the confidence regions (3) and (4) can be much below the desired nominal levels. In some cases, when there are no outliers present, the coverage probability can be as low as 0.75 for a 95% nominal level and the coverage probability can drop to about 0.149 when outliers are present. A similar phenomenon was also observed in Zhang (1997) and Zhang *et al.* (2000a) for several growth and hormone models for boys during puberty. A consequence of these liberal confidence intervals is a very high false positive rate in the context of testing hypotheses. Thus the P values based on such standard errors can

Analysis of Nonlinear Regression Models

be smaller than the true P values and hence a researcher may declare significance even though there is no significant effect.

Several authors such as Bates and Watts (1988) and Ratkowsky (1990) have discussed the effect of curvature on the accuracy of (3) and (4). A very detailed investigation of these effects is provided in these books. The curvature effects can be decomposed into two components, the *intrinsic effect* γ_{\max}^N that is due to the shape of the response function, and the *parametric effect* γ_{\max}^P that is due to the functional form of the parameters. Smaller the values of γ_{\max}^N and γ_{\max}^P , lower the curvature effects. Estimators for these two parameters along with a simple F test to detect the severity of the two curvature effects can be found in Ratkowsky (1990) and in Seber and Wild (1989). Let

$$c^* = \frac{1}{2 \sqrt{F_{\alpha, p, n-p}}},$$

where $F_{\alpha, p, n-p}$ is the $(1 - \alpha)$ th percentile of central F distribution with $(p, n - p)$ degrees of freedom. If the estimated value of γ_{\max}^N (γ_{\max}^P) $< c^*$ then it suggests that the intrinsic (parametric) curvature effect is not significant.

If the intrinsic curvature is severe then one may want to consider an alternative nonlinear model to describe the dependence of Y on X . On the other hand, if the parametric curvature is severe then one may re-parameterize so that the resulting parametric form is subject to less curvature effect. A potential drawback with this solution is that although the parametric curvature may be reduced due to re-parameterization, the experimenter may find the new parameters difficult to interpret. Thus it is often a challenge to analyze data using nonlinear statistical models. Either the parameters have good physical interpretation but hard to perform inferences on or the parameters resulting from re-parameterization are difficult to interpret but easy to perform inferences on!

Example 2.1 (Seber and Wild, 1989): Consider the following two-parameter Hill equation:

$$a \frac{X}{X + b}$$

Based on an experimental data discussed in Seber and Wild (1989) the authors concluded that the intrinsic curvature in the above function is not very severe but the parametric effect curvature is very severe. For this reason they re-parameterized the model as:

$$\frac{X}{cX + d}, \text{ where } c = 1/a \text{ and } d = b/a.$$

S. D. Peddada and J. K. Haseman

Under this re-parameterization it is reasonable to perform statistical inferences on c and d using the standard methods.

Simonoff and Tsai (1986) performed very extensive simulations studying eight different jackknife based methods some of which “adjust” for the curvature. They found that the jackknife confidence interval centered at the average of the pseudo-values \mathbf{P}_{1i} (see (5) in the Appendix) with covariance matrix based on the pseudo-values \mathbf{P}_{2i} (see (6) in the Appendix), which account for the curvature effects, performed best. Simonoff and Tsai (1986) called this procedure the RLQM procedure. Zhang (1997) and Zhang *et al.* (2000a) considered an alternative sandwich estimator $\hat{\mathbf{V}}_{Z_1}$ (see (9) in the Appendix) for the covariance matrix of MLE. Based on simulation studies involving nonlinear models for growth curve, growth hormone and testosterone for boys during puberty, Zhang (1997) found that the confidence intervals based on (6) perform better than those based on RLQM in terms of the coverage probability.

2.2 A simulation study

Since the Hill model (2) is widely used in the context of animal carcinogenicity studies and in the risk assessment of various chemicals we base our simulation studies on this model. In this subsection we demonstrate that the Wald confidence intervals (4), denoted by C_{MLE} , can sometime be liberal for some individual parameters. We only provide simulation results for (4) because in view Simonoff and Tsai (1986) and Zhang (1997) the results are expected to be even worse for (3). We compared the coverage probabilities of C_{MLE} with the coverage probabilities of the confidence intervals introduced in Zhang (1997) and Zhang *et al.* (2000a), which are denoted as C_Z . For simplicity we take the baseline response to be zero and hence consider the following 3-parameter Hill model:

$$Y_{ij} = b \frac{x_{ij}^d}{x_{ij}^d + c^d} + \varepsilon_{ij},$$

where $i = 1, 2, 3, 4, 5, j = 1, 2, \dots, m$, and the random errors ε_{ij} are identically and independently distributed as standard normal variables. To understand the effect of sample size on the coverage probabilities, we considered several different patterns of m , the number of animals at each dose group. In this article we report results corresponding to $m = 3$ (small sample size), $m = 10$ (moderately large sample size), and $m = 20$ (a large sample size). As in Walker *et al.* (1999) the five dose groups considered in this paper are (0, 3.5, 10.7, 35.7, 125).

We considered 5 different patterns of parameters for the Hill model (see Figure 1) with different amounts of curvature effects. In addition to summarizing the coverage probabilities for the two methods of confidence

Analysis of Nonlinear Regression Models

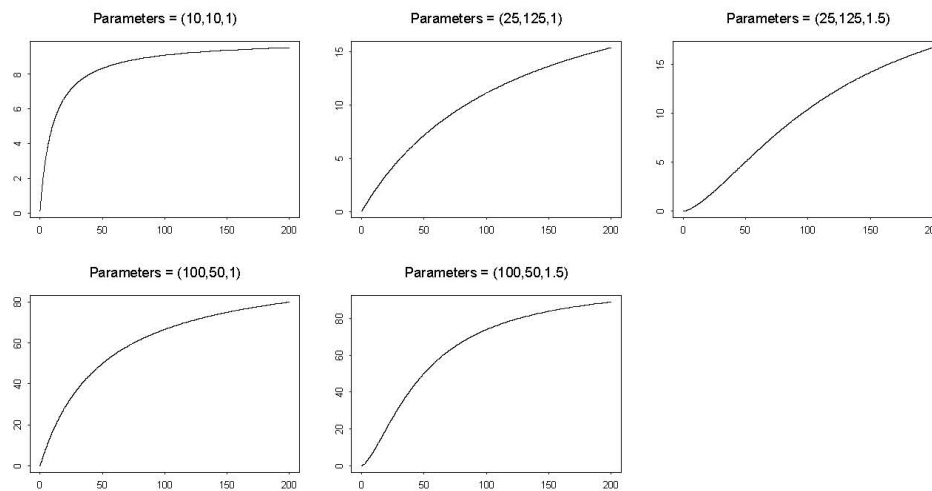


FIGURE 1 Hill model for different patterns of parameters.

intervals for each of the parameters, in Table 1 we also provide the median of the estimated curvature effects for each pattern based on 1000 simulation runs and the value of c^* for each m . We estimated the two curvature effects using the FORTRAN code provided in Ratkowsky (1990). All minimizations were performed using the subroutine AMOEBA provided in Press *et al.* (1989) with several initial starting values for performing the minimization.

We note from Table 1 that, apart from the case of $(b, c, d) = (10, 10, 1)$, there are no serious intrinsic curvature effects but there can be severe parametric curvature effects. In the case $(b, c, d) = (10, 10, 1)$ the median of the estimated value of $\gamma_{\max}^N = 0.35$ which exceeds $c^* = 0.27$. In this situation we notice that ED_{01} cannot be estimated with accurate standard error. The coverage probability of MLE is only 0.86, which is much smaller than the nominal level of 0.95. The remaining parameters appear to be estimated reasonably well by the MLE. The method of Zhang *et al.* (2000a) seems to improve the coverage probability.

Both, MLE as well as Zhang *et al.* (2000a), procedures are affected by the severity of the parametric curvature effects. Of the two methods, the methodology of Zhang *et al.* (2000a) performs better. In the worst case when $m = 3$ and $(b, c, d) = (25, 125, 1)$ both procedures perform very poorly for estimating the parameters b, c and ED_{10} , although the procedure of Zhang *et al.* (2000a) is better. As the sample size per dose group increases from $m = 3$ to $m = 20$ the parametric curvature effects decrease and hence the coverage probabilities tend to improve. When there is very little parametric curvature effect the methods tend to attain the nominal

S. D. Peddada and J. K. Haseman

TABLE 1 Comparison of the coverage probabilities of C_{MLE} and C_Z .

(b, c, d)	Curvature ($\gamma_{\max}^N, \gamma_{\max}^P$)	Method	Coverage probability				
			b	c	d	ED_{01}	ED_{10}
$c^* = 0.27$		$m = 3$					
(25,125,1)	(.14,55.09)	C_{MLE}	.75 (17.73)	.71 (142.11)	.98 (.68)	.97 (1.63)	.85 (8.10)
		C_Z	.80 (22.97)	.75 (184.94)	1.00 (.86)	.99 (2.09)	.92 (10.95)
(25,125,1.5)	(.17,33.20)	C_{MLE}	.66 (18.77)	.87 (352.35)	1.00 (2.51)	.98 (15.03)	.89 (47.88)
		C_Z	.75 (25.00)	.92 (470.19)	1.00 (3.36)	.98 (20.44)	.94 (63.51)
(10,10,1)	(.35,3.95)	C_{MLE}	.92 (2.22)	.90 (6.28)	.96 (.60)	.86 (.24)	.93 (1.02)
		C_Z	.95 (2.93)	.94 (8.14)	.99 (.79)	.90 (.32)	.98 (1.34)
(100,50,1)	(.01,1.14)	C_{MLE}	.92 (8.33)	.91 (10.00)	.93 (.08)	.94 (.11)	.93 (.45)
		C_Z	.97 (10.94)	.97 (13.18)	.97 (.11)	.97 (.13)	.98 (.59)
(100,50,1.5)	(.02,.42)	C_{MLE}	.92 (5.98)	.92 (5.47)	.94 (.13)	.94 (.40)	.92 (.64)
		C_Z	.98 (7.89)	.97 (7.27)	.98 (.17)	.98 (.52)	.98 (.85)
$c^* = 0.3$		$m = 10$					
(25,125,1)	(.08,63.24)	C_{MLE}	.85 (16.76)	.82 (158.04)	.99 (.36)	.98 (.76)	.85 (8.94)
		C_Z	.86 (17.96)	.83 (169.20)	.99 (.38)	.98 (.81)	.86 (9.63)
(100,50,1)	(.006,.65)	C_{MLE}	.94 (4.71)	.94 (5.69)	.95 (.05)	.95 (.06)	.95 (.25)
		C_Z	.96 (5.07)	.96 (6.12)	.96 (.05)	.97 (.06)	.97 (.27)
$c^* = 0.30$		$m = 20$					
(25,125,1)	(.06,48.63)	C_{MLE}	.89 (12.48)	.86 (120.65)	.98 (.25)	.99 (.47)	.87 (6.89)
		C_Z	.89 (12.91)	.87 (124.81)	.99 (.26)	.99 (.49)	.88 (7.14)
(100,50,1)	(.004,.46)	C_{MLE}	.95 (3.35)	.94 (4.04)	.94 (.03)	.95 (.04)	.96 (.18)
		C_Z	.96 (3.47)	.95 (4.19)	.96 (.03)	.96 (.04)	.96 (.18)

Median widths of the confidence intervals are provided within parentheses. Nominal level = 0.95.

level of 0.95. However, as in the case of $(b, c, d) = (25, 125, 1)$, when the parametric curvature effect is large the convergence to the nominal level is very slow. Even with a sample of size 20 per group we do not seem to attain the nominal level of 0.95. Among the five parameters, the slope parameter d is often estimated conservatively, the coverage probability usually exceeding the nominal level of 0.95. On the other hand the rest of the parameters are often estimated liberally. The worst affected parameters are the maximum of the Hill model, i.e. b , the ED_{50} , i.e. c , and ED_{10} .

3. CONCLUSIONS

Statistical analysis of nonlinear regression models are routinely performed in applied sciences using the standard asymptotic methods which are based on linearization of the nonlinear model around the unknown parameter. Often data analysts and researchers do not pay

Analysis of Nonlinear Regression Models

attention to the some of the subtle assumptions underlying such analysis. As evidenced in the simulation studies reported in this paper and in Simnoff and Tsai (1986), this may result in underestimation of the standard errors and extremely high false positive rates and liberal or narrow confidence intervals. Consequently, one cannot trust the results obtained from such analyses.

The purpose of this article is to caution researchers and data analysts against the potential problems with nonlinear models. Although at the moment there is no satisfactory methodology for estimating standard errors of MLE, the methodology proposed in Zhang (1997) and in Zhang *et al.* (2000a) is perhaps an improvement over the existing procedure.

The problem is further complicated if there is heteroscedasticity in the data (i.e. variance of Y is not constant over all observations). In such situations EPA's BMD software (USEPA, 2001) uses the method of maximum likelihood estimation by modeling the variance of Y as a power function of the mean of Y . Although this is a common practice, it can potentially introduce bias due to model mis-specification. As an alternative to this procedure, Zhang (1997) and Zhang *et al.* (2000a) introduced the sandwich estimator (10) for the covariance matrix of MLE, which is asymptotically consistent. This estimator is based on a procedure developed in Peddada and Smith (1997) for the covariance matrix of the MLE in a linear model with heteroscedastic errors. Some optimality and asymptotic properties of this procedure are discussed in Peddada (1993) and Peddada and Smith (1997). Thus as an alternative to the procedure used in EPA's BMD software [15] one may derive the standard errors of MLE using (10) and obtain the corresponding asymptotic confidence intervals.

Given the extensive usage of nonlinear models in practice we believe there is a need for further methodological work in this field. Perhaps one may want to explore Bayesian methods that do not rely on linearization of the nonlinear model.

ACKNOWLEDGEMENTS

The authors thank Drs. Beth Gladen (NIEHS), David Umbach (NIEHS) and Joanne Zhang (FDA) for carefully reading this manuscript and making several comments that substantially improved the presentation of the manuscript.

REFERENCES

- Bates, D., and Watts, D. G. (1988). *Nonlinear Regression Analysis and its Applications*, Wiley, New York, NY.
 Box, G. E. P. and Coutie, G. A. (1956). Application of digital computers in the exploration of functional relationships. *Proc. I. E. E.*, 103, Part B, *Suppl.* 1, 100 - 107.
 Fox, T., Hinkley, D., and Larntz, K. (1980). Jackknifing in Nonlinear Regression. *Technometrics*, 22, 29-33.

S. D. Peddada and J. K. Haseman

- Kim, A., Kohn, M., Portier, C., and Walker, N. (2002). Impact of Physiologically Based Pharmacokinetic Modeling on Bench Mark Dose Calculations for TCDD-Induced Biochemical Responses. *Regulatory Toxicology and Pharmacology*, **36**, 287-296.
- Peddada, S. D. (1993). Jackknife variance estimation and bias reduction. *Handbook of Statistics*, **9**, 723—744; C. R. Rao, ed., Elsevier Science Publishers.
- Peddada, S. D., and Patwardhan, G. (1992). Jackknife variance estimators in linear models. *Biometrika*, **79**, 654-657.
- Peddada, S. D., and Smith, T. (1997). Consistency of a class of variance estimators in linear models under heteroscedasticity. *Sankhya*, **59**, 1-10.
- Portier, C., Sherman, C., Kohn, M., Elder, L., Kopp-Schneider, A., Maronpot, R., and Lucier, G. (1996). Modeling the Number and Size of Hepatic Focal Lesions Following Exposure to 2,3,7,8-TCDD. *Toxicology and Applied Pharmacology*, **138**, 20-30.
- Press, W., Flannery, B., Teukolsky, S., Vetterling, B. (1989). *Numerical Recipes: The art of scientific computing (FORTRAN version)*.
- Ratkowsky, D. A. (1990). *Handbook of Nonlinear Regression Models, STATISTICS: textbooks and monographs*, Marcel Decker Inc., New York, NY.
- Seber, G. A. F., and Wild, C. J. (1989). *Nonlinear Regression*, Wiley, New York, NY.
- Shao, J (1990). Asymptotic theory in heteroscedastic nonlinear models. *Statistics and Probability Letters*, **10**, 77-85.
- Shao, J (1992). Consistency of least squares estimator and its jackknife variance estimator in nonlinear models. *The Canadian Journal of Statistics*, **20**, 415-428.
- Simonoff, J. S., and Tsai. C. (1986). Jackknife-based Estimators and Confidence Regions in Nonlinear Regression. *Technometrics*, **28**, 103-112.
- USEPA (2000). Exposure and Human Health Reassessment of 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD) and Related Compounds. (September 2000 Draft). Part II: Health Assessment of 2, 3, 7, 8-Tetrachlorodibenzo-p-dioxin (TCDD) and Related Compounds. EPA/600/P-00/001 Be. *National Center for Environmental Assessment, Office of Research and Development, U.S. Environmental Protection Agency*, Washington D.C.
- USEPA (2001). Help Manual for Benchmark Dose Software Version 1.3. EPA 600/R-00/014F, *Office of Research and Development*, Washington D.C. 20460.
- Walker, N., Portier, C., Lax, S., Crofts, F., Li, Y., Lucier, G., Sutter, T. (1999). Characterization of the Dose-Response of CYP1B1, CYP1A1 and CYP1A2 in the Liver of Female Sprague-Dawley Rats Following Chronic Exposure to 2,3,7,8-Tetrachlorodibenzo-p-dioxin. *Toxicology and Applied Pharmacology*, **154**, 279-286.
- Zhang, J. (1997). Analysis of Nonlinear Fixed and Random Effects Models with Applications to Statural Growth and Hormonal Changes in Boys at Puberty. *Ph.D. dissertation, University of Virginia*, Charlottesville, Virginia.
- Zhang, J., Peddada, S. D., and Rogol, A. (2000a). Estimation of Parameters in Nonlinear Regression Models. *Statistics For the 21st Century*, Eds. C. R. Rao and G. Szekely, 459-483.
- Zhang, J., Peddada, S. D., Malina, R., and Rogol, A. (2000b). A Longitudinal Assessment of Hormonal and Physical Alterations During Normal Puberty in Boys VI. Modeling of Growth Velocity, Mean Growth Hormone (GH Mean) and Serum Testosterone (T). *American Journal of Human Biology*, **12**, 814-824.

APPENDIX

In a nonlinear regression model $\mathbf{Y} = f(\mathbf{X}, \theta) + \varepsilon$, where θ is a $p \times 1$ vector of unknown parameters, \mathbf{Y} is an $n \times 1$ response vector, \mathbf{X} is the matrix of explanatory variables and ε is a normal random vector with mean $\mathbf{0}$ and covariance matrix $\sigma^2 \mathbf{I}$. Suppose $\hat{\theta}$ is the MLE of θ and $\hat{\theta}_{(i)}$ is the MLE of θ when the i th observation is deleted from the calculation of MLE.

Let $\hat{\mathbf{V}}$ denote the standard asymptotic covariance matrix of $\hat{\theta}$, $\hat{\mathbf{V}}'_i$ denote the i th row vector of $\hat{\mathbf{V}}$ and let $\hat{\mathbf{V}}_{..}$ be an $n \times p \times p$ second derivative array where $v_{ist} = \partial^2 f(\mathbf{X}_i, \theta) / \partial \theta_s \partial \theta_t$ evaluated at $\hat{\theta}$. For a $m \times n$ matrix \mathbf{A} and an $n \times p \times p$ array $\mathbf{B} = \{(\mathbf{b}_{rs})\}$, we define $[\mathbf{A}][\mathbf{B}] = \{(\mathbf{A}\mathbf{b}_{rs})\}$.

Analysis of Nonlinear Regression Models

The standard MLE based asymptotic $(1 - \alpha) \times 100\%$ confidence region for the parameter $\hat{\theta}$ is given by (see Seber and Wild, 1989)

$$\frac{(\hat{\beta} - \beta)'(\hat{\mathbf{V}}'\hat{\mathbf{V}})(\hat{\beta} - \beta)}{p\hat{\sigma}^2} \leq F_{\alpha, p, n-p}, \quad (3)$$

where $\hat{\sigma}^2$ is proportional to the maximum likelihood estimator of σ^2 , $F_{\alpha, p, n-p}$ is the $(1 - \alpha)$ th percentile of central F distribution with $(p, n - p)$ degrees of freedom. The corresponding confidence intervals for individual components θ_i are obtained by

$$\hat{\theta}_i \pm t_{\alpha/2, n-p} \hat{\sigma} \hat{\mathbf{V}}_{ii}, \quad (4)$$

where $t_{\alpha/2, n-p}$ is the $(1 - \alpha/2)$ th percentile of central t distribution based on $n - p$ degrees of freedom and $\hat{\mathbf{V}}_{ii}$ is the i th diagonal element of $\hat{\mathbf{V}}$.

A variety of jackknife procedures have been considered in the literature (Fox *et al.*, 1980, Simnoff and Tsai, 1986). In particular, Simnoff and Tsai (1986) considered eight different jackknife procedures to construct confidence intervals. Their RLQM procedure, which uses the Box and Coutie (1956) modification of the covariance matrix of MLE for dealing with curvature effects, is based on the following pseudo-values:

$$\mathbf{P}_{1i} = \hat{\theta} + n(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1} \frac{\hat{\mathbf{V}}_i \hat{\varepsilon}_i}{(1 - \hat{h}_{ii})}, \quad (5)$$

where $\hat{h}_{ii} = \hat{\mathbf{V}}_i'(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}_i$.

$$\mathbf{P}_{2i} = \hat{\theta} + \frac{n\mathbf{T}_i^{-1}\hat{\mathbf{V}}_i\hat{\varepsilon}_i}{1 - \hat{h}_{ii}^*}. \quad (6)$$

Here $\mathbf{T}_i = \hat{\mathbf{V}}'\hat{\mathbf{V}} - [\hat{\varepsilon}'_{(i)}][\hat{\mathbf{V}}_{(i)}]$, and $\hat{h}_{ii}^* = \hat{\mathbf{V}}_i'\mathbf{T}_i^{-1}\hat{\mathbf{V}}_i$. For each $j = 1, 2$, the jackknife point estimator $\hat{\theta}_{jj}$ is given by

$$\hat{\theta}_{jj} = \frac{1}{n} \sum_{i=1}^n \mathbf{P}_{ji} \quad (7)$$

and a jackknife point estimator of the covariance matrix is given by

$$\mathbf{V}_{jj} = \frac{1}{n(n-p)} \sum_{i=1}^n (\mathbf{P}_{ji} - \hat{\theta}_{jj})(\mathbf{P}_{ji} - \hat{\theta}_{jj})'. \quad (8)$$

S. D. Peddada and J. K. Haseman

As an alternative to the above methodology, Zhang (1997) and Zhang *et al.* (2000a) proposed a class of sandwich estimators that are based on the estimators introduced in Peddada (1993), Peddada and Patwardhan (1992), and in Peddada and Smith (1997) for linear models.

For homoscedastic errors, i.e. $Var(Y_i) = \sigma^2$, Zhang (1997) and Zhang *et al.* (2000a, 2000b) considered the following sandwich estimator for estimating the covariance matrix of $\hat{\theta}$

$$\hat{V}_{Z_1} = \frac{n\hat{\sigma}^2}{n-p} (\hat{V}'\hat{V})^{-1} \sum_{i=1}^n \frac{1}{1-\hat{h}_{ii}} \hat{V}_i \hat{V}_i' (\hat{V}'\hat{V})^{-1}, \quad (9)$$

where $\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n \hat{\varepsilon}_i^2$.

Remark 4.1. Suppose $\mathbf{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})'$ with

$$Y_{ij} = f(\mathbf{X}_{ij}, \theta) + \varepsilon_{ij}, j = 1, 2, \dots, n_i, i = 1, 2, \dots, k,$$

where ε_{ij} are *i.i.d.* with $E(\varepsilon_{ij}) = 0$ and $Var(\varepsilon_{ij}) = \sigma_i^2$. Then Zhang *et al.* (2000a) proposed the following class of variance estimators:

$$\hat{V}_{Z_2} = (\hat{V}'\hat{V})^{-1} \sum_{i=1}^k \frac{\hat{\varepsilon}_i' \hat{\varepsilon}_i}{n_i - \delta_{ii}} \hat{V}_i' \hat{V}_i (\hat{V}'\hat{V})^{-1}, \quad (10)$$

where δ_{ii} is some function of $T\hat{V}_i'(\hat{V}'\hat{V})^{-1}\hat{V}_i$ such that $\delta_{ii} \rightarrow 0$ as $\sum_{i=1}^k n_i \rightarrow \infty$, \hat{V}_i is a $n_i \times p$ matrix of partial derivatives of $f(\mathbf{X}_{ij}, \theta)$ with respect to θ , and $\hat{\varepsilon}_i = (\hat{\varepsilon}_{i1}, \hat{\varepsilon}_{i2}, \dots, \hat{\varepsilon}_{in_i})'$ with $\hat{\varepsilon}_i = \mathbf{Y}_{ij} - f(\mathbf{X}_{ij}, \hat{\theta})$.

Zhang *et al.* (2000a) deduced the asymptotic properties of the above estimators along the lines of Shao (1990, 1992).