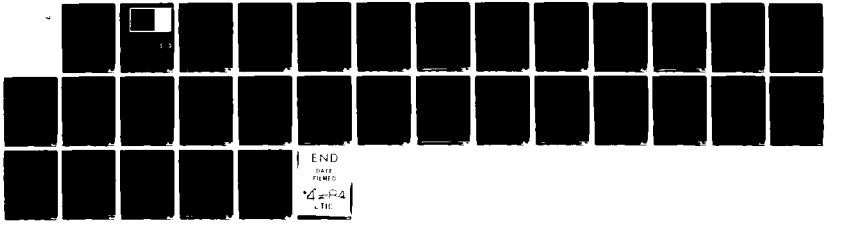
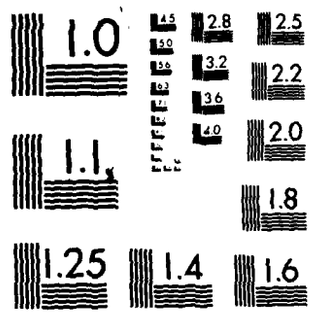
AD-A139 239 BIAS IN NONLINEAR REGRESSION(U) WISCONSIN UNIV-MADISON 1/1
MATHEMATICS RESEARCH CENTER R D COOK ET AL. FEB 84
MRC-TSR-2645 DAAG29-80-C-0041

UNCLASSIFIED F/G 12/1 NL



END
DATE FILMED
4-84
DTIC



MRC Technical Summary Report #2645

Bias in Nonlinear Regression

R.D. Cook, C.L. Tsai and B.C. Wei

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

February, 1984

(Received November 28, 1983)

Sponsored by

U.S. Army Research Office
P.O. Box 12211
Research Triangle Park
North Carolina 27709

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

BIAS IN NONLINEAR REGRESSION

R.D. Cook, C.L. Tsai* and B.C. Wei†

Technical Summary Report #2645

February 1984

ABSTRACT

We investigate the biases of the residuals and the maximum likelihood parameter estimates from standard, normal-theory nonlinear regression models. Emphasis is placed on determining the influence of individual cases on the biases and on understanding how the residual biases can affect the usefulness of standard diagnostic methods. It is shown that the various bias expressions in the literature are equivalent, that the biases in nonlinear regression can be studied usefully in the context of linear regression, and that diagnostic plots using residuals can be misleading because of substantial residual biases. For a class of partially nonlinear models, it is shown that the maximum intrinsic curvature (Bates and Watts 1980) is closely related to the residual expectations. Finally, the model associated with power transformations of single explanatory variables in linear regression is investigated in further detail and several numerical illustrations are presented.

AMS (MOS) Subject Classification: 62J02

Key Words: Diagnostics, Influence, Intrinsic Curvature, Influence, Residuals, Transformations

Work Unit Number 4 (Statistics and Probability)

*Department of Statistics and Operations Research, New York University, New York, NY 10006.

†Department of Mathematics, Nanjing Institute of Technology, Nanjing, Jiangsu, The People's Republic of China.

Sponsored in part by the United States Army under Contract No. DAAG29-80-C-0041.

SIGNIFICANCE AND EXPLANATION

The statistical analysis of a collection of data is usually based on a specified model, a mathematical formula describing the behavior of the data up to a few unknown parameters which are to be estimated from the data. Much is known about the statistical behavior of parameter estimates and other important statistics that arise from analyses based on models that are linear functions of the parameters, while relatively little is known when the underlying model is nonlinear in the parameters.

The purpose of this paper is to study the bias of the parameter estimates and other important statistics that stem from analyses based on nonlinear models with normal errors.

BIAS IN NONLINEAR REGRESSION

R.D. Cook, C.L. Tsai* and B.C. Wei†

1. INTRODUCTION

It is well known that the maximum likelihood estimators of the parameters in standard nonlinear regression models are generally biased estimators of the true parameter values. Other quantities that are useful in analyses of nonlinear regression models can be characterized in the same way. The ordinary residuals, for example, generally have nonzero expectations and in this sense are also biased.

The past studies of the various biases in nonlinear regression are certainly useful, but they do not fully exploit recognizable aspects of the structure of the problem in a way that might facilitate understanding; nor do they emphasize points of view that provide for an appreciation of the potential importance of the biases in practice. The effects of residual biases on standard diagnostic plots, for example, are apparently unknown. In this paper we describe a relatively simple structure for investigating the nature of the parameter and residual biases in nonlinear regression.

In Section 2, we briefly review the past results on the parameter bias, discuss their interpretation and use in practice, and show that individual cases can have a substantial influence on the bias. In Section 3 we investigate the expectation of the residual vector and related quantities such as the change in the residual expectations that result from case deletion. In particular, we show that bias in nonlinear regression can be usefully studied in the context of linear regression.

The transition from linear to fully nonlinear models often seems quite abrupt. This transition can be smoothed by investigating intermediate, partially nonlinear models. In Section 4, we study a special class of partially nonlinear models that occurs frequently in practice and covers many of the illustrative examples in the statistical literature. For this class of models, interesting relationships occur between the residuals and other diagnostic statistics. The expectations of the residuals, for example, are shown to be closely related to the Bates-Watts measure of intrinsic curvature. In the remainder of this section, we establish notation and briefly review relevant background material.

*Department of Statistics and Operations Research, New York University, New York, NY 10006.
†Department of Mathematics, Nanjing Institute of Technology, Nanjing, Jiangsu, The People's Republic of China.

Sponsored in part by the United States Army under Contract No. DAAG29-80-C-0041.

The standard nonlinear regression model can be represented as

$$y_i = f(\underset{\sim}{x}_i, \underset{\sim}{\theta}) + \varepsilon_i, \quad i = 1,2,\ldots,n \tag{1}$$

where $\underset{\sim}{x}_i$ represents a vector of known explanatory variables associated with the i-th observable response y_i, $\underset{\sim}{\theta}$ is a p-vector of unknown parameters, the response function f is assumed to be known, continuous and twice differentiable in $\underset{\sim}{\theta}$, and the errors are assumed to be independent, identically distributed normal random variables with mean 0 and variance σ^2. For this model, the maximum likelihood estimator $\hat{\underset{\sim}{\theta}}$ of $\underset{\sim}{\theta}$ can be found by minimizing the objective function

$$J(\underset{\sim}{\theta}) = \sum_{i=1}^{n} \left(y_i - f(\underset{\sim}{x}_i, \underset{\sim}{\theta})\right)^2 . \tag{2}$$

Kennedy and Gentle (1980) discuss computational methods for obtaining $\hat{\underset{\sim}{\theta}}$; for our purposes we assume that $\hat{\underset{\sim}{\theta}}$ is available. The asymptotic behavior of $\hat{\underset{\sim}{\theta}}$ is investigated by Wu (1981) who provides additional references. The usual estimator of σ^2 is $s^2 = J(\hat{\underset{\sim}{\theta}})/(n - p)$.

For notational convenience, let $f_i = f(\underset{\sim}{x}_i, \underset{\sim}{\theta})$, $i = 1,2,\ldots,n$, and let $\underset{\sim}{V}$ denote the $n \times p$ matrix with elements $f_i^r = \partial f_i/\partial\theta_r$, $i = 1,2,\ldots,n$, $r = 1,2,\ldots,p$. Unless indicated otherwise, all derivatives are evaluated at the true parameter values. Various quadratic expansions used in the following sections involve the $p \times p$ matrices $\underset{\sim}{W}_i$, $i = 1,2,\ldots,n$; the elements of $\underset{\sim}{W}_i$ are $f_i^{rs} = \partial^2 f_i/\partial\theta_r\partial\theta_s$, $r,s, = 1,2,\ldots,p$.

2. BIAS OF $\hat{\underset{\sim}{\theta}}$.

Numerous approximations for the bias vector $E\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}$ are available in the literature. Cox and Snell (1968) derive an order n^{-1} approximation for the bias in an expanded class of models that includes (1) as a special case. Their derivation is based on a quadratic expansion of the likelihood equations and thus requires third partial derivatives of the log likelihood. Box (1971) used a quadratic expansion of the residuals to obtain an approximation for the bias in a multivariate version of (1). For the single

response case, Bates and Watts (1980) provide a connection between Box's bias approximation and their parameter-effects curvature array. More recently, further approximations for the bias have been offered by Clarke (1980), Hougaard (1981) and Amari (1982). Clarke and Hougaard deal specifically with model (1) while Amari considers curved exponential families that include (1) as a special case. Clarke's derivation of the bias approximation is similar in spirit to that of Cox and Snell (1968), while Hougaard's result is apparently a straightforward application of Skovgaard (1981).

The derivations of the various bias approximations are based on different notations, approaches and degrees of generality. Clarke (1980) mentions that his result agrees with that of Box (1971), but otherwise the relationships between these approximations are apparently unknown. As shown in the Appendix, these bias approximations are, in fact, identical for model (1). Further, it is shown that, apart from differences in notation, the Bates-Watts (1980) form for the bias is the same as that given by Clarke (1980).

A useful numerical study of the bias in various yield-density models is given by Gillis and Ratkowsky (1978; see also Ratkowsky, 1983) who conclude that the actual bias (as obtained through simulation) is accurately reflected by the now common bias approximation.

Since the bias forms discussed above are all equivalent, any of them can be used as an aid in nonlinear regression analyses. For the purposes of interpretation, however, it is helpful to exploit the relationship with linear regression: From the Appendix it follows that the bias approximation $\underset{\sim}{b} = E\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}$ can be expressed as

$$\underset{\sim}{b} = (\underset{\sim}{V}^T\underset{\sim}{V})^{-1}\underset{\sim}{V}^T\underset{\sim}{d} \tag{3}$$

where $\underset{\sim}{d}$ is an n-vector with elements $-\sigma^2 tr[(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}\underset{\sim}{W}_i]/2$, $i = 1,2,\ldots,n$. Thus, the bias $\underset{\sim}{b}$ is simply the coefficients from the ordinary least squares regression of $\underset{\sim}{d}$ on the columns of $\underset{\sim}{V}$. Further, $\underset{\sim}{d}$ is essentially the expected difference between linear and quadratic approximations of the estimated response function. To see this, let $\underset{\sim}{f}(\underset{\sim}{\theta})$ denote the n-vector with elements f_i, $i = 1,2,\ldots,n$, and expand $\underset{\sim}{f}(\hat{\underset{\sim}{\theta}})$ about $\underset{\sim}{\theta}$,

$$\underset{\sim}{f}(\hat{\underset{\sim}{\theta}}) = \underset{\sim}{f}(\underset{\sim}{\theta}) + \underset{\sim}{V}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) + \frac{1}{2}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})^T\underset{\sim}{W}(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) \tag{4}$$

where $\underset{\sim}{W}$ is an $n \times p \times p$ array with i-th face $\underset{\sim}{W}_i$, $i = 1,2,\ldots,n$. Multiplication involving three dimensional arrays is defined as in Bates and Watts (1980) so that the third term of (4) is an n-vector with elements $(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})^T \underset{\sim}{W}_i (\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})/2$, $i = 1,2,\ldots,n$. The expected difference between the linear and quadratic approximations of $\underset{\sim}{f}(\hat{\underset{\sim}{\theta}})$ is thus

$$-\frac{1}{2} E(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})^T \underset{\sim}{W} (\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}) \approx \underset{\sim}{d} . \tag{5}$$

These results indicate that the bias will be small if the elements of $\underset{\sim}{d}$ are sufficiently close to zero, so that the model is essentially linear, or if $\underset{\sim}{d}$ is orthogonal to the tangent plane, i.e. the column space of $\underset{\sim}{V}$.

Because $\underset{\sim}{b}$ can be simply interpreted as the coefficient vector from an ordinary least squares regression, we can now employ a variety of the diagnostic methods available in linear regression to a study of the bias $\underset{\sim}{b}$ in nonlinear regression. In particular, added variable plots (Cook and Weisberg, 1982) for the components of $\underset{\sim}{b}$, or $\hat{\underset{\sim}{b}}$ obtained by substituting estimates for parameters, may prove to be particularly useful. Further, the form of $\underset{\sim}{b}$ allows for the rather straightforward development of methods for investigating the effects of individual cases on the determination of $\underset{\sim}{b}$ or $\hat{\underset{\sim}{b}}$.

To investigate the influence of individual cases on $\underset{\sim}{b}$, some additional notation is required. The subscript (i) means "with the i-th case deleted" so that, for example, $\underset{\sim}{V}_{(i)}$ is the $(n-1) \times p$ matrix formed by deleting the i-th row of $\underset{\sim}{V}$ and

$$\underset{\sim}{b}_{(i)} = (\underset{\sim}{V}_{(i)}^T \underset{\sim}{V}_{(i)})^{-1} \underset{\sim}{V}_{(i)}^T \underset{\sim}{d}_{(i)} . \tag{6}$$

To display the effects of deleting the i-th case, we express $\underset{\sim}{b}_{(i)}$ as a function of the full data, as is usual in this kind of investigation. The general methods for doing this are the same as those used in linear regression, but an added complication arises here since all the elements of $\underset{\sim}{d}$ change when a case is deleted.

Define the n-vector

$$\underset{\sim}{\delta}_i = -\frac{\sigma^2}{2} \frac{\underset{\sim}{v}_i^T (\underset{\sim}{V}^T \underset{\sim}{V})^{-1} \underset{\sim}{W} (\underset{\sim}{V}^T \underset{\sim}{V})^{-1} \underset{\sim}{v}_i}{1 - h_i} \tag{7}$$

where $\underset{\sim}{v}_i^T$ is the i-th row of $\underset{\sim}{V}$ and h_i is the i-th diagonal element of the tangent plane hat matrix $\underset{\sim}{P}_v = \underset{\sim}{V}(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}\underset{\sim}{V}^T$. Further, let $\underset{\sim}{\delta}^i$ and $\underset{\sim}{d}^i$ denote $(n-1) \times 1$ vectors formed by deleting the i-th components from $\underset{\sim}{\delta}_i$ and $\underset{\sim}{d}$, respectively. Then $\underset{\sim}{d}_{(i)} = \underset{\sim}{d}^i + \underset{\sim}{\delta}^i$ and it follows that

$$\underset{\sim}{b}_{(i)} - \underset{\sim}{b} = \underset{\sim}{b}_\delta - \frac{(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}\underset{\sim}{v}_i\alpha_i}{1 - h_i} \tag{8}$$

where $\underset{\sim}{b}_\delta$ is the coefficient vector from the ordinary least squares regression of $\underset{\sim}{\delta}_i$ on $\underset{\sim}{V}$, and α_i is the i-th residual from the regression of $\underset{\sim}{d} + \underset{\sim}{\delta}_i$ on $\underset{\sim}{V}$.

Equation (8) has several general features in common with well known, analogous expressions from linear regression. For example, the tangent plane leverage components h_i can be interpreted in much the same way as the leverage components from linear regression so that remote points on the tangent plane can have a substantial influence on the bias. The vector $\underset{\sim}{\delta}_i$ reflects the change in the expected difference between linear and quadratic approximations of $\underset{\sim}{f}(\hat{\underset{\sim}{\theta}})$ when the i-th case is deleted. If h_i is large then individual components of $\underset{\sim}{\delta}_i$ may be large and removal of the i-th case may result in a substantial change in the agreement between the linear and quadratic approximations.

3. RESIDUAL BIAS

Although the errors in model (1) have expectation zero, the expectations of the residuals $e_i = y_i - f(\underset{\sim}{x}_i, \hat{\underset{\sim}{\theta}})$, $i = 1,2,\ldots,n$, are generally nonzero and in this sense the residuals are biased.

Taking the expectation on both sides of (4), using $\underset{\sim}{b}$ to approximate $E(\hat{\underset{\sim}{\theta}}) - \underset{\sim}{\theta}$, and using $E(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})(\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta})^T \approx \sigma^2(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}$ yields the following approximation $\underset{\sim}{E}$ for the expectation of the residual vector $\underset{\sim}{e} = (e_i)$,

$$\underset{\sim}{E} = (\underset{\sim}{I} - \underset{\sim}{P}_v)\underset{\sim}{d}. \tag{9}$$

This approximation agrees with the results of Cox and Snell (1968).

A useful pattern appears from (3) and (9): From (3) the parameter bias $\underset{\sim}{b}$ is simply the coefficient vector from the ordinary least squares regression of $\underset{\sim}{d}$ on $\underset{\sim}{V}$, while from (9) we see that $E(\underset{\sim}{e})$ is just the vector of residuals from this same regression. Other useful analogies with linear regression can be developed as well: Suppose that $f(\underset{\sim}{x}, \hat{\underset{\sim}{\theta}})$ is to be used as an estimate of $f_x = f(\underset{\sim}{x}, \underset{\sim}{\theta})$ at a point $\underset{\sim}{x}$ that does not occur in the data. Let $\underset{\sim}{v}_x = \partial f_x/\partial\underset{\sim}{\theta}$ and $\underset{\sim}{W}_x = \partial^2 f_x/\partial\underset{\sim}{\theta}^2$. Then $f(\underset{\sim}{x}, \underset{\sim}{\theta}) - Ef(\underset{\sim}{x}, \hat{\underset{\sim}{\theta}}) \doteq d_x - \underset{\sim}{v}_x^T\underset{\sim}{b}$ where $d_x = -\sigma^2 tr[(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}\underset{\sim}{W}_x]/2$ is the "response" at $\underset{\sim}{x}$ and $\underset{\sim}{v}_x^T\underset{\sim}{b}$ is the "estimate" at $\underset{\sim}{x}$ from the regression of $\underset{\sim}{d}$ on $\underset{\sim}{V}$. In such problems, it may be useful to plot $\underset{\sim}{d}_x - \underset{\sim}{v}_x^T\underset{\sim}{b}$ as a function of $\underset{\sim}{x}$ to find if there are regions of substantial bias.

Bates and Watts (1980) express $\underset{\sim}{b}$ in terms of the diagonal elements of the faces of the parameter-effects curvature array $\underset{\sim}{A}^T$. Similarily, $\underset{\sim}{E}$ can be expressed in terms of an orthonormal basis for the null space of $\underset{\sim}{V}$ and the diagonal elements the intrinsic curvature array, $\underset{\sim}{A}^N$. Let $\underset{\sim}{V} = \underset{\sim}{Q}\underset{\sim}{R}$ denote the QR decomposition of $\underset{\sim}{V}$ and partition $\underset{\sim}{Q} = [\underset{\sim}{U}, \underset{\sim}{N}]$ where the columns of the $n \times (n - p)$ matrix $\underset{\sim}{N}$ form an orthonormal basis for the null space of $\underset{\sim}{V}$. Then

$$\underset{\sim}{E} \doteq -\frac{\sigma^2}{2}\underset{\sim}{N}\sum_{i=1}^{p}\underset{\sim}{A}_{ii}^N \tag{10}$$

where $\underset{\sim}{A}_{ii}^N$ is an $(n - p) \times 1$ vector with elements a_{jii}, $j = 1,2,\ldots,n-p$, and a_{jii} is the i-th diagonal element of the j-th face of the intrinsic curvature array $\underset{\sim}{A}^N$. (In forming (10) we have used the original rather than scaled data as in Bates and Watts, 1980). Since $\underset{\sim}{E}$ depends only on $\underset{\sim}{N}$ and the elements of the intrinsic curvature array, it is invariant under transformations of the parameters, as expected.

The influence of the i-th case on the determination of $\underset{\sim}{E}$ can be found by using the results on $\underset{\sim}{b}_{(i)}$ from Section 2. Let $e_{j(i)}$ denote the j-th residual based on the data without the i-th case, $j = 1,2,\ldots,n$, and let d_j and δ_{ij} denote the j-th elements of $\underset{\sim}{d}$ and $\underset{\sim}{\delta}_i$, respectively. Then $E(e_{j(i)}) = d_j + \delta_{ij} - \underset{\sim}{v}_j^T\underset{\sim}{b}_{(i)}$ and using (8) it follows that the change in the residual expectation is

$$E(e_{j(i)} - e_j) = \delta_{ij} - \underset{\sim}{v}_j^T \underset{\sim}{b}_\delta - h_{ij}\, a_i/(1 - h_i) \tag{11}$$

where h_{ij} is the ij-th element of $\underset{\sim}{H}$. When $i = j$ equation (11) gives the change in bias of $f(\underset{\sim}{x}_i, \hat{\underset{\sim}{\theta}})$ when the i-th case is deleted,

$$E(e_{i(i)} - e_i) = \frac{\delta_{ii} - \underset{\sim}{v}_i^T \underset{\sim}{b}_\delta}{1 - h_i} - \frac{h_i}{1 - h_i} E(e_i) \ . \tag{12}$$

Clearly, remote points on the tangent plane can have a substantial influence on the average behavior of the residuals. Notice also that $\delta_{ii} - \underset{\sim}{v}_i^T \underset{\sim}{b}_\delta$ is simply the i-th residual from the regression of $\underset{\sim}{\delta}_i$ on $\underset{\sim}{V}$.

4. SPECIAL MODELS

4.1 Partially Nonlinear Models.

Further insight into the behavior of the bias can be gained by considering the partially nonlinear response function

$$\underset{\sim}{f}(\underset{\sim}{\theta}) = \underset{\sim}{Z}\underset{\sim}{\alpha} + \beta \underset{\sim}{g}(\gamma) \tag{13}$$

where $\underset{\sim}{Z}$ is a known, full rank $n \times (p - 2)$ matrix, $\underset{\sim}{\theta}^T = (\underset{\sim}{\alpha}^T, \beta, \gamma)$ and, as indicated, β and γ are scalars. This class of response functions occurs often in practice and in the statistical literature; see, for example, Bates and Watts (1980) Gallant (1975), Vølund (1978) and Stone (1980). In particular, (13) allows for transformation of a single explanatory variable in linear regression. Let $\underset{\sim}{g}$, $\underset{\sim}{g}'$ and $\underset{\sim}{g}''$ denote n-vectors with elements g_i, $\partial g_i/\partial\gamma$ and $\partial^2 g_i/\partial\gamma^2$, respectively. Application of the results of Section 2 yields

$$\underset{\sim}{V} = [\underset{\sim}{Z}, \underset{\sim}{g}(\gamma), \beta \underset{\sim}{g}'(\gamma)] \tag{14}$$

$$\underset{\sim}{d} = -\operatorname{cov}(\hat{\beta}, \hat{\gamma})\underset{\sim}{g}'(\gamma) - \tfrac{1}{2}\beta \operatorname{var}(\hat{\gamma})\underset{\sim}{g}''(\gamma) \tag{15}$$

and

$$\underset{\sim}{b} = -\beta^{-1}\operatorname{cov}(\hat{\beta}, \hat{\gamma})\, \underset{\sim}{\ell}_p - \tfrac{1}{2}\beta \operatorname{var}(\hat{\gamma})(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}\underset{\sim}{V}^T\underset{\sim}{g}''(\gamma) \tag{16}$$

where $\underset{\sim}{\ell}_p$ is the p-th standard basis vector, and $\operatorname{var}(\hat{\gamma})$ and $\operatorname{cov}(\hat{\beta},\hat{\gamma})$ are the indicated large sample variance and covariance, i.e. the appropriate elements of $\sigma^2(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}$.

Notice that $(\underset{\sim}{V}^T\underset{\sim}{V})^{-1}\underset{\sim}{V}^T\underset{\sim}{g}''$ is simply the coefficient vector from the regression of $\underset{\sim}{g}''$ on $\underset{\sim}{V}$ and that $\text{cov}(\hat{\beta},\hat{\gamma})$ contributes only to the bias of $\hat{\gamma}$.

The behavior of $\underset{\sim}{b}$ as a function of γ is complicated and difficult to predict. For the remaining parameters, however, several useful conclusions can be obtained from (16). First, $\underset{\sim}{b}$ does not depend on $\underset{\sim}{\alpha}$, as expected. Second, as a function of β and σ^2, the bias of each of the components of $\hat{\underset{\sim}{\alpha}}$ and the bias of $\hat{\beta}$ are proportional to σ^2/β since $\text{var}(\hat{\gamma})$ is proportional to σ^2/β^2. Third, as a function of β and σ^2, the bias of $\hat{\gamma}$ is proportional to σ^2/β^2 since $\text{cov}(\hat{\beta},\hat{\gamma})$ is proportional to σ^2/β. Clearly, values of β close to zero can result in substantially biased estimators while for sufficiently large values of β the bias will be negligible. This reinforces the usual empirical notion that a transformation γ an explanatory variable in linear regression will be well determined only when the r· ession coefficient β is large relative to σ.

The second-order expectation of $\underset{\sim}{e}$ for model (13) follows immediately from (,

$$\underset{\sim}{E} = -\frac{1}{2}\beta\,\text{var}(\hat{\gamma})(\underset{\sim}{I} - \underset{\sim}{P}_{\underset{\sim}{V}})\underset{\sim}{g}''(\gamma) \quad . \tag{17}$$

As a function of β and σ^2, $\underset{\sim}{E}$ is proportional to σ^2/β.

There is an interesting relationship between the maximum intrinsic curvature Γ^N (Bates and Watts, 1980) and the residual expectation (17). The maximum intrinsic curvature can be written as

$$\Gamma^N = \max \frac{\|(\underset{\sim}{I} - \hat{\underset{\sim}{P}}_{\underset{\sim}{V}})\ddot{\underset{\sim}{f}}_{\underset{\sim}{k}}\|}{\|\dot{\underset{\sim}{f}}_{\underset{\sim}{k}}\|^2}\, s\sqrt{p} \tag{18}$$

where $\ddot{\underset{\sim}{f}}_{\underset{\sim}{k}}$ and $\dot{\underset{\sim}{f}}_{\underset{\sim}{k}}$ are the acceleration and velocity vectors, respectively, and the maximum is taken over all directions $\underset{\sim}{k} = (k_i) \in R^p$. We have added the "hat" to $\underset{\sim}{P}_{\underset{\sim}{V}}$ as a reminder that all derivatives involved in (18) are evaluated at the maximum likelihood estimates. Evaluating Γ^N for model (13) we find that

$$\Gamma^N = \max \frac{k^2|\hat{\beta}|\,\|(\underset{\sim}{I} - \hat{\underset{\sim}{P}}_{\underset{\sim}{V}})\underset{\sim}{g}''(\hat{\gamma})\|}{\|\underset{\sim}{k}^T\hat{\underset{\sim}{V}}^T\hat{\underset{\sim}{V}}\underset{\sim}{k}\|^2}\, s\sqrt{p} = |\hat{\beta}|\,\hat{\text{var}}(\hat{\gamma})\|(\underset{\sim}{I} - \hat{\underset{\sim}{P}}_{\underset{\sim}{V}})\underset{\sim}{g}''(\hat{\gamma})\|\sqrt{p}/s \quad . \tag{19}$$

From (17) and (19) it follows that

$$\|\hat{\underset{\sim}{E}}\|/s = I^N/2\sqrt{p} \tag{20}$$

where $\hat{\underset{\sim}{E}}$ is $\underset{\sim}{E}$ evaluated at the maximum likelihood estimates. Thus, for model (13) the maximum intrinsic curvature also describes the length of the vector of estimated residual biases. This simple relation does not seem to hold for more complicated models.

4.2 Power Transformations of Explanatory Variables.

In this section we investigate the special case of model (13) that corresponds to selecting a power transformation for a single explanatory variable x in linear regression. Emphasis is placed on obtaining an understanding of the role that the values of x have in determining the parameter and residual biases.

The problem of determining a power transformation for a single explanatory variable can be approached in the framework of model (13) with

$$g_i(\gamma) = g(x_i,\gamma) = (x_i^\gamma - 1)/\gamma \ . \tag{21}$$

We investigate the role of the x_i's by modifying (21) to allow for a systematic alteration of these values. Specifically, for $\gamma \neq 0$ we use $g(cx_i,\gamma)$, $c > 0$, while for $\gamma = 0$ we find it more convenient to use $g(x_i^c,\gamma)$. The introduction of the known constant c is intended to allow for a comparison of a limited number of designs and should not be confused with rescaling of a common design.

When $\gamma = 0$ and $g_i = g(x_i^c,\gamma)$ we find $\underset{\sim}{V} = [\underset{\sim}{Z}, c\log x_i, \beta c^2(\log x_i)^2/2]$ and $g''(x_i^c,0) = c^3(\log x_i)^3/3$. Next, let $\underset{\sim}{b}^1 = (b_i^1)$ and $\underset{\sim}{E}^1$ denote the parameter and residual biases, respectively, when $\sigma = \beta = c = 1$. Then a little algebra will verify that

$$\begin{aligned} b_i &= \sigma^2 b_i^1/\beta c \ , \quad i = 1,2,\ldots,p-2 \ , \\ b_{p-1} &= \sigma^2 b_{p-1}^1/\beta c^2 \end{aligned} \tag{22}$$

and

$$b_p = \sigma^2 b_p^1/\beta^2 c^3$$

where the biases b_{p-1} and b_p correspond to β and γ respectively. Further,

$$\underset{\sim}{E} = \sigma^2 \underset{\sim}{E}^1/\beta c \ . \tag{23}$$

The results in (22) and (23) suggest that when $\gamma = 0$ increasing the ratio $\max x_i/\min x_i$ will reduce all biases, with $\hat{\gamma}$ benefitting from the greatest reduction. For example, comparing the designs $\{x_i\}$ and $\{x_i^2\}$, we see that the bias for $\hat{\gamma}$ from $\{x_i\}$ will be 8 times larger than the bias from $\{x_i^2\}$.

When $\gamma = 0$ and $g_i = g_i(cx_i,\gamma)$, $\underset{\sim}{E}$ does not depend on c. For example the designs $\{x_i\}$ and $\{2x_i\}$ will yield the same residual expections when $\gamma = 0$.

When $\gamma \neq 0$ we have been unable to find an informative expression for $\underset{\sim}{b}$ when using $g(cx_i,\gamma)$. However, when $\underset{\sim}{Z}$ contains a column of 1's it is relatively straightforward to deal with $\underset{\sim}{E}$ since the column space of $\underset{\sim}{V}$ is the same as the column space of $\underset{\sim}{V}^* = [\underset{\sim}{Z}, x_i^\gamma, x_i^\gamma \log x_i]$, and except for the term $c^\gamma x_i^\gamma (\log x_i)^2/\gamma$ all addends comprising $\underset{\sim}{g}''(\gamma) = [g''(cx_i,\gamma)]$ are contained in the column space of $\underset{\sim}{V}^*$. Thus, when constructing the projection in (17) we may use $\underset{\sim}{V}^*$ and $g''(cx_i,\gamma) = c^\gamma x_i^\gamma (\log x_i)^2/\gamma$. From this it follows that when $g_i = g_i(cx_i,\gamma)$

$$\underset{\sim}{E} = \sigma^2 \underset{\sim}{E}^1/\beta c^\gamma . \tag{24}$$

Clearly, the effect of replacing $\{x_i\}$ with $\{cx_i\}$ depends strongly on γ and c. For example, when $\gamma > 0$ the intrinsic curvature for the design $\{2x_i\}$ will be smaller than that for $\{x_i\}$, while when $\gamma < 0$ the reverse is true.

5. ILLUSTRATIONS

In this section, we present several examples to illustrate selected results of the previous sections.

To illustrate the nature of the residual bias (9), we use model (13) with $\underset{\sim}{g}(\gamma)$ given by (21), $\underset{\sim}{Z} = \underset{\sim}{1}$, $n = 20$ and $x_i = 2.2(1)21.2$. Plots of $\underset{\sim}{E}^1$ versus x for various values of γ are shown in Figure 1. For reference, we also give the maximum intrinsic curvature Γ^N for $\sigma = \beta = 1$ and each value of γ. The values of Γ^N and the scales of the y-axes can be converted for other values of σ and β, and certain other values for the explanatory variable by using (20), (23) and (24). The plots of $\underset{\sim}{E}$ are clearly patterned, but the display changes in both shape and magnitude as γ varies. At

$\gamma = -1$, Γ^N and the magnitudes of the residual expectations are relatively large, while at $\gamma = 2$ the residual expectations seem unimportant.

Box and Hill (1974) describe a weighted analysis based on a linearized version of the model

$$f_i = \frac{\theta_0\theta_2(x_{i2} - x_{i3}/1.632)}{1 + \theta_1 x_{i1} + \theta_2 x_{i2} + \theta_3 x_{i3}} \qquad (25)$$

for $i = 1,2,\ldots,24$. We use this model and the data provided by Box and Hill (1974) to illustrate the effects of individual cases on the parameter bias. The ordinary least squares estimate of $\underset{\sim}{\theta}^T = (\theta_0,\ldots,\theta_3)$ is $\underset{\sim}{\hat{\theta}}^T = (35.9, .071, .038, .167)$, and the large sample estimates of the corresponding standard errors are 8.21, .179, .100 and .416, respectively. The estimated bias vector, obtained by substituting estimates for parameters in (3), is $\underset{\sim}{\hat{b}}^T = (1.997, .438, .245, 1.02)$. Several of these biases are large relative to the parameter estimates, even when the standard errors are taken into account.

An index plot of the θ_3 component of $\underset{\sim}{\hat{b}}_{(i)} - \underset{\sim}{\hat{b}}$ is given as Figure 2. Clearly, cases 20 and 22 play a substantial role in reducing the bias of $\hat{\theta}_3$. Comparing Figure 2 with Figure 3, an index plot of the $\hat{h}_i$, we see that the three cases (20, 22 and 24) with the largest effects on the bias of $\hat{\theta}_3$ also have the three largest values of $\hat{h}_i$ (see equation 8). Removal of any of these three cases, particularly case 20, would cause a substantial change in the agreement between the linear and quadratic approximations of (25). This can be seen, for example, by inspecting the elements of $\underset{\sim}{\hat{\delta}}_{20}$, which are substantial, or by comparing the curvatures with and without case 20: $\Gamma^N = .088$, $\Gamma^T = 140.11$, $\Gamma^N_{(20)} = .12$ and $\Gamma^T_{(20)} = 352.77$.

To illustrate the use of added variable plots associated with the regression in (3), we use the Michaelis-Menton model $f_i = \theta_1 x/(\theta_2 + x)$ in combination with the data from Bates and Watts (1980). However, for emphasis the first case ($x = 2$, $y = .0615$) is deleted in this example so that there are $n = 11$ cases. The estimates based on these data are $\hat{\theta}_1 = .08898$, $\hat{\theta}_2 = 1.3668$ and $s = .00425$. The large sample standard errors for $\hat{\theta}_1$ and $\hat{\theta}_2$ are .0165 and .4079, respectively. Figure 4 gives the added variable plot for θ_1. The first five points in this plot have two replicates, while the final

point in the upper, right corner is replicated only once. The clear indication from this figure is that the outlying point is having a substantial influence on the bias for $\hat{\theta}_1$. Further calculation confirms this indication: For the full (11 cases) data $\hat{\underline{b}}^T = (.0026, .074)$ while the bias without the outlying case (x = 2, y = .0527) and evaluated at the estimates from the full data is $\hat{b}^T = (.0432, .898)$. Also, without the outlying case $\hat{\theta}_1 = .0990$ and $\hat{\theta}_2 = 1.5693$, and the respective large sample standard errors are .0808 and 1.656. The substantial increase in the standard error of $\hat{\theta}_2$ partially accounts for the change in the bias of $\hat{\theta}_2$ when the outlying case is removed.

DISCUSSION

The development of diagnostic methods for linear regression is dependent on a thorough study and characterization of the exact small sample behavior of a few fundamental building blocks such as the ordinary residuals and related statistics. In nonlinear regression, the small sample behavior of the corresponding building blocks is generally intractable so that some degree of approximation is necessary. We have found the various quadratic approximations in this paper to be useful aids for understanding selected aspects of nonlinear regression problems. In principle, these approximations can be extended to a higher accuracy, although the practical usefulness of such extensions is unclear.

Recall that the residual expectations are strongly dependent on the intrinsic array A^N. The maximum intrinsic curvatures calculated by Bates and Watts (1980, Table 2), and Ratkowsky (1983) indicate that the residual expectations will generally be negligible. This may be a reflection of the quality of published studies rather than a reflection of the intrinsic linearity of statistical models since the illustrations in Section 5 show that there is no intrinsic reason why the intrinsic curvature cannot be large.

Finally, the results of this paper are presented from a diagnostic view, but may also be useful in other contexts. The updating analog of (8), for example, may be useful for

searching the factor space to find a few additional runs that would substantially reduce the bias.

APPENDIX

Equivalence of Bias Approximations

To show the equivalence of the five bias approximations discussed in Section 2, we develop Cox and Snell's (1968) result for model (1). The approximations of Box (1971), Clarke (1980), Hougaard (1981) and Amari (1982) will be obtained at intermediate steps in this development.

Let m_{rs}, $r,s = 1,\ldots,p$, denote the elements of $\underset{\sim}{M} = (\underset{\sim}{V}^T\underset{\sim}{V})^{-1}$ and let ℓ_i denote the log likelihood corresponding to the i-th observation from model (1) so that the total log likelihood is $\ell = \sum_i \ell_i$. Here and in the expressions that follow, summations involving the index i are over the integers 1 to n, while summations involving any other index are over the integers 1 to p. We assume that σ^2 is known when constructing ℓ.

Cox and Snell's (1968) bias approximation for $\hat{\theta}_s$, $s = 1,\ldots,p$, can now be written as

$$b_s = \frac{\sigma^4}{2} \sum_r \sum_t \sum_u m_{rs} m_{tu} (K_{rtu} + 2J_{t,ru}) \qquad (A.1)$$

where $K_{rtu} = E\partial^3\ell/\partial\theta_r\partial\theta_t\partial\theta_u$ and $J_{t,ru} = \sum_i E(\partial\ell_i/\partial\theta_t)(\partial^2\ell_i/\partial\theta_r\partial\theta_u)$.

The approximation error for (A.1) is $o(n^{-1})$. Evaluating K_{rtu} and $J_{t,ru}$ for model (1), we find that

$$K_{rtu} = -\frac{1}{\sigma^2} \sum_i (f_i^r f_i^{tu} + f_i^t f_i^{ru} + f_i^u f_i^{rt})$$

and

$$J_{t,ru} = \frac{1}{\sigma^2} \sum_i f_i^t f_i^{ru} .$$

Substituting these forms in (A.1) and simplifying yields

$$b_s = -\frac{\sigma^2}{2} \sum_i \sum_{r,t,u} m_{rs} m_{tu} f_i^r f_i^{tu} . \qquad (A.2)$$

Apart from trivial changes in notation, this is Hougaard's (1981) bias approximation.

Next, since

$$\sum_{t,u} m_{tu} f_i^{tu} = \mathrm{tr}(\underset{\sim}{M}\, \underset{\sim}{W}_i)$$

equation (A.2) can be rewritten as

$$b_s = -\frac{\sigma^2}{2} \sum_i \sum_r m_{rs} f_i^r \mathrm{tr}(\underset{\sim}{M}\, \underset{\sim}{W}_i)$$

or equivalently

$$\underset{\sim}{b} = -\frac{\sigma^2}{2} \underset{\sim}{M} \sum_i \underset{\sim}{v}_i \mathrm{tr}(\underset{\sim}{M}\, \underset{\sim}{W}_i) \qquad (A.3)$$

where $\underset{\sim}{v}_i^T$ is the i-th row of $\underset{\sim}{V}$. Apart from notation, equation (A.3) is the bias approximation derived by Box (1971).

To obtain Clarke's (1980) expression for the bias, we first express (A.3) in full matrix form

$$\underset{\sim}{b} = \underset{\sim}{M}\, \underset{\sim}{V}^T \underset{\sim}{d} \qquad (A.4)$$

where $\underset{\sim}{d}$ is an n-vector with elements $-\frac{\sigma^2}{2} \mathrm{tr}(\underset{\sim}{M}\, \underset{\sim}{W}_i)$. Next we express (A.4) in terms of the QR-decomposition $\underset{\sim}{V} = \underset{\sim}{U}\underset{\sim}{R}$, where $\underset{\sim}{R}$ is a $p \times p$ nonsingular, upper triangular matrix and $\underset{\sim}{U}$ is $n \times p$ with orthogonal columns, $\underset{\sim}{U}^T\underset{\sim}{U} = \underset{\sim}{I}$:

$$\underset{\sim}{b} = \underset{\sim}{L}\, \underset{\sim}{U}^T \underset{\sim}{d} \qquad (A.5)$$

where $\underset{\sim}{L} = \underset{\sim}{R}^{-1}$. The i-th element of $\underset{\sim}{d}$ can be represented as

$$d_i = -\frac{\sigma^2}{2} \mathrm{tr}(\underset{\sim}{L}^T\, \underset{\sim}{W}_i\, \underset{\sim}{L}) = \sum_j d_{ij}$$

where d_{ij} is the j-th diagonal element of $-\frac{\sigma^2}{2} \underset{\sim}{L}^T\, \underset{\sim}{W}_i\, \underset{\sim}{L}$. Finally, defining $\underset{\sim}{d}_j$ to be an n-vector with elements d_{ij}, i = 1,2,...n, we obtain

$$\underset{\sim}{b} = \underset{\sim}{L} \sum_j \underset{\sim}{U}^T \underset{\sim}{d}_j \qquad (A.6)$$

which is Clarke's (1980) form. In Clarke's notation, $\underset{\sim}{K} = \underset{\sim}{L}$ and $B_{ijj} = \underset{\sim}{u}_i^T \underset{\sim}{d}_j$ where $\underset{\sim}{u}_i$ is the i-th column of $\underset{\sim}{U}$. Further, in the notation of Bates and Watts (1980), $\underset{\sim}{a}_{jj} / s^2 p = \underset{\sim}{U}^T \underset{\sim}{d}_j$ so that, apart from differences in notation, Clarke's form is the same

as that given by Bates and Watts.

Amari (1982, Theorem 9) gives the bias of an efficient estimator from a curved exponential family. For the maximum likelihood estimator $\hat{\underset{\sim}{\theta}}$, Amari's form immediately reduces to

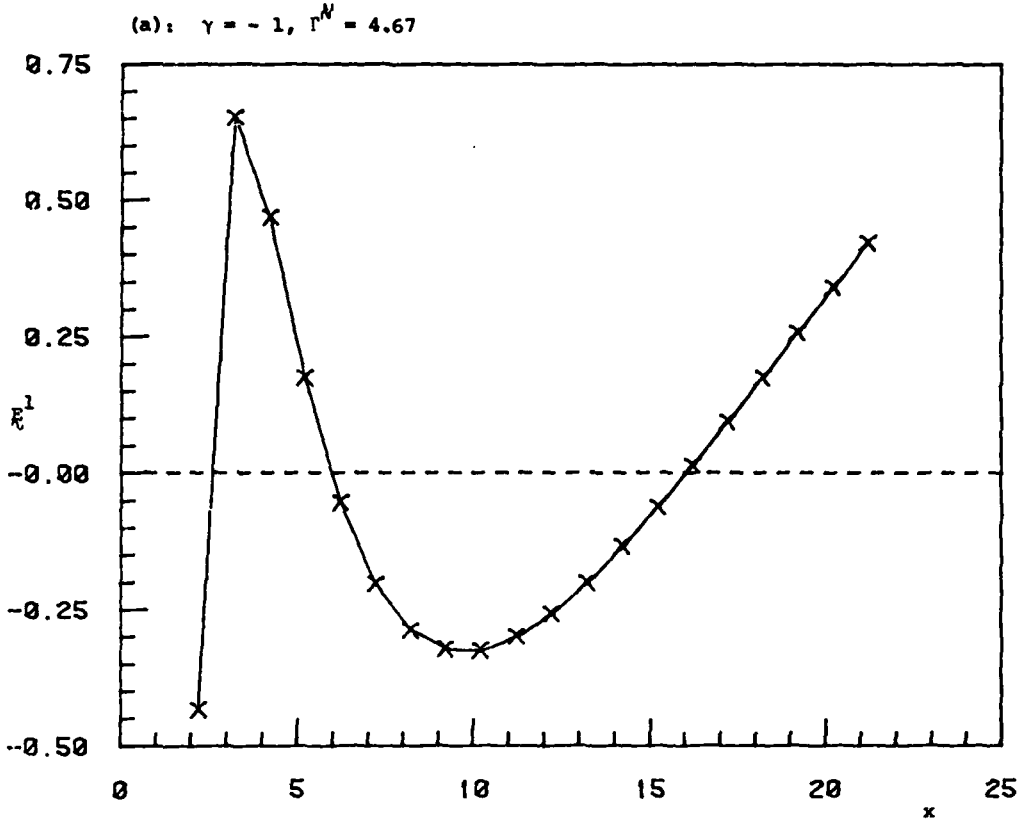
$$b_s = -\frac{\sigma^4}{2} \sum_{r,t,u} m_{rs} m_{tu} \overset{m}{\Gamma}_{tur} \tag{A.7}$$

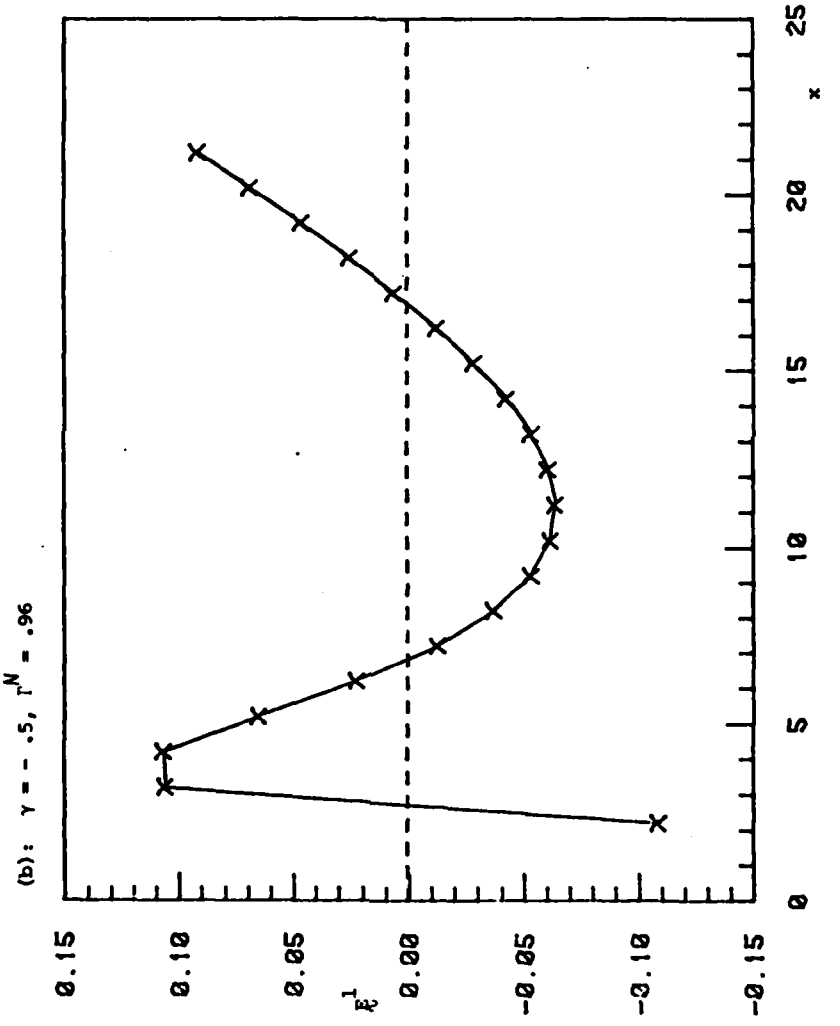
where $\overset{m}{\Gamma}_{tur}$ is the mixture connection. From the definition of the mixture connection, we have

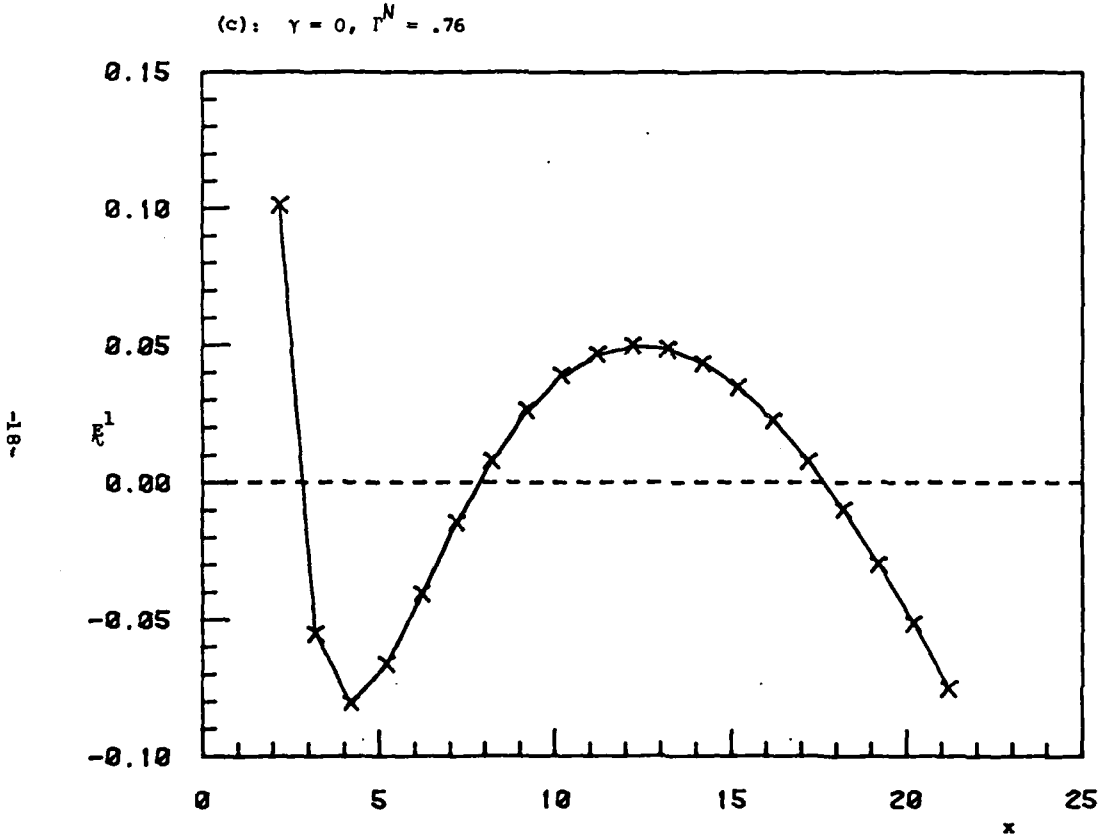
$$\overset{m}{\Gamma}_{tur} = E\{\frac{\partial^2 \ell}{\partial\theta_t \partial\theta_u} \cdot \frac{\partial \ell}{\partial\theta_r}\} + E\{\frac{\partial \ell}{\partial\theta_t} \frac{\partial \ell}{\partial\theta_u} \frac{\partial \ell}{\partial\theta_r}\} .$$

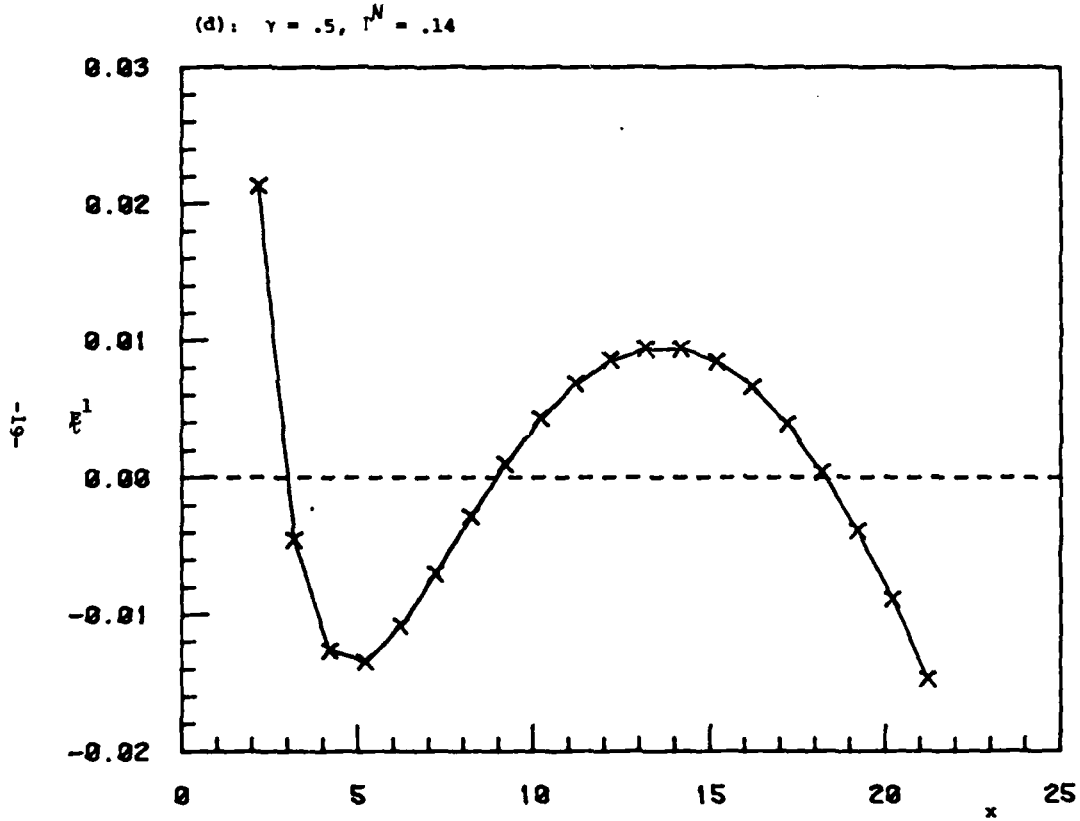
For model (1) the second term of the mixture connection equals zero and it is not difficult to verify that the first term is $\sum_i f_i^{tu} f_i^r/\sigma^2$. Substituting this into (A.7) yields (A.2) so that for model (1) Amari's bias approximation is equivalent to the four forms discussed above.

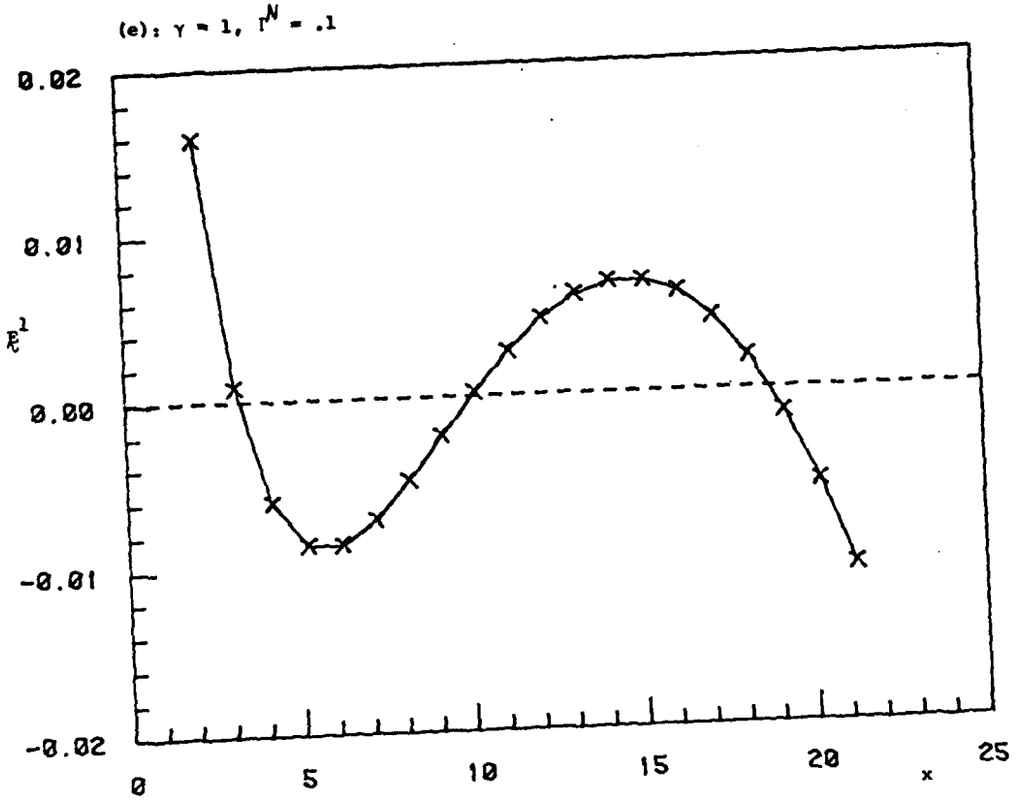
Figure 1(a) - 1(f): Scatter plots of $\hat{\epsilon}_i^1$ versus x for the power transformation model (21) and six values of γ.




(b): γ = - .5, Γ^N = .96
0.15
0.10
0.05
0.00
-0.05
-0.10
-0.15
E^1
0
5
10
15
20
25
x


(c): γ = 0, Γ^N = .76
0.15
0.10
0.05
0.00
-0.05
-0.10
ξ_c^1
0
5
10
15
20
25
x


(d): γ = .5, Γ^N = .14
0.03
0.02
0.01
0.00
-0.01
-0.02
$\frac{1}{\varphi}$ E^1
0
5
10
15
20
25
x


(e): γ = 1, Γ^N = .1
0.02
0.01
0.00
-0.01
-0.02
E_ℓ^1
0
5
10
15
20
25
x

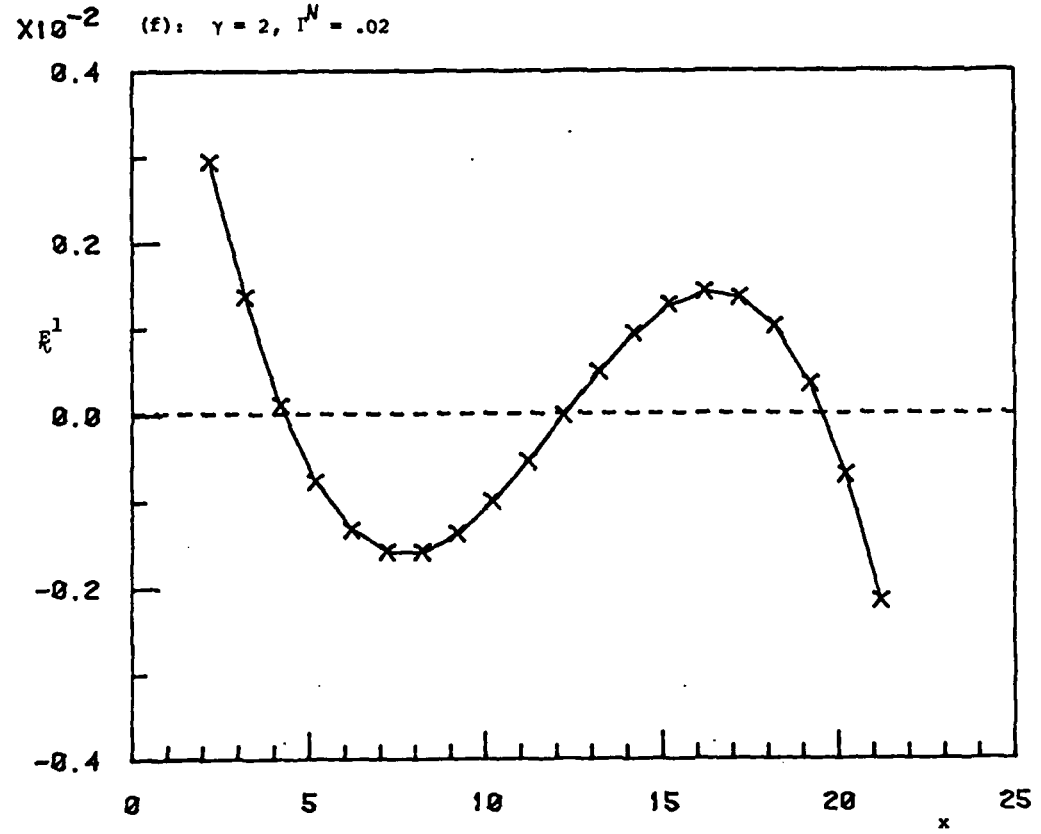
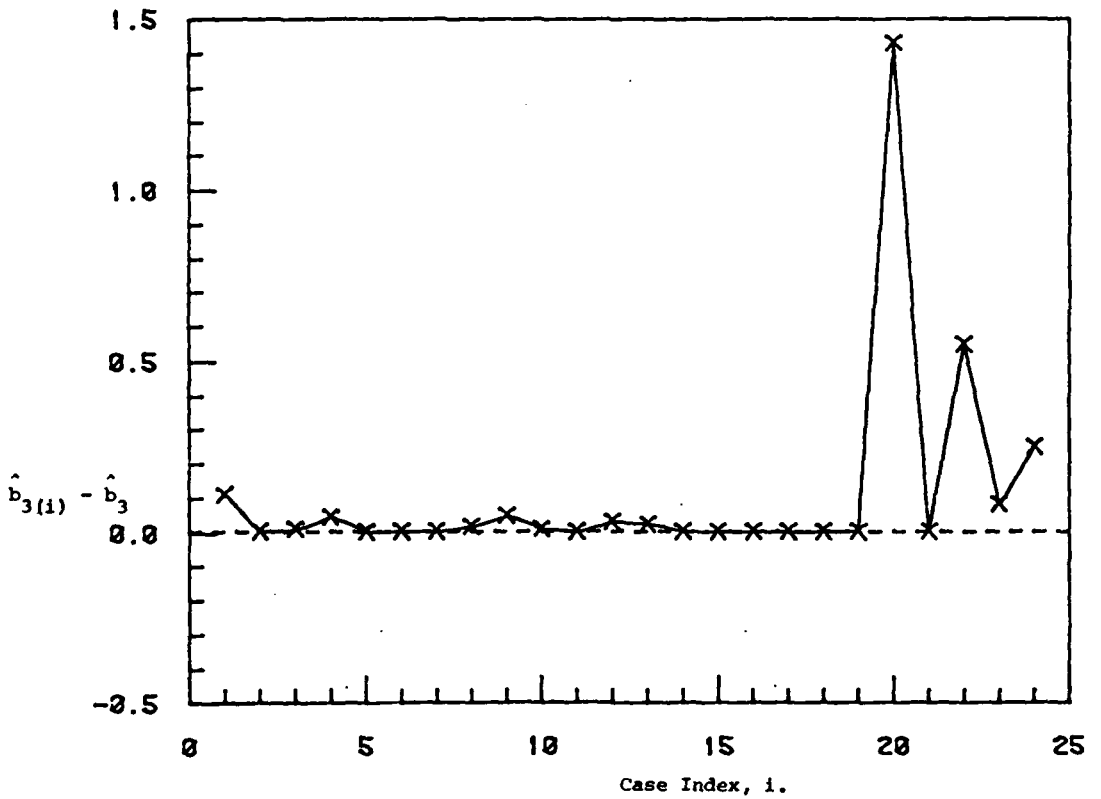

X10^-2
(ε): γ = 2, Γ^N = .02
0.4
0.2
0.0
-0.2
-0.4
ε^1
0
5
10
15
20
25
x

Figure 2: Index plot of $(\hat{b}_{3(i)} - \hat{b}_3)$ for model (25).



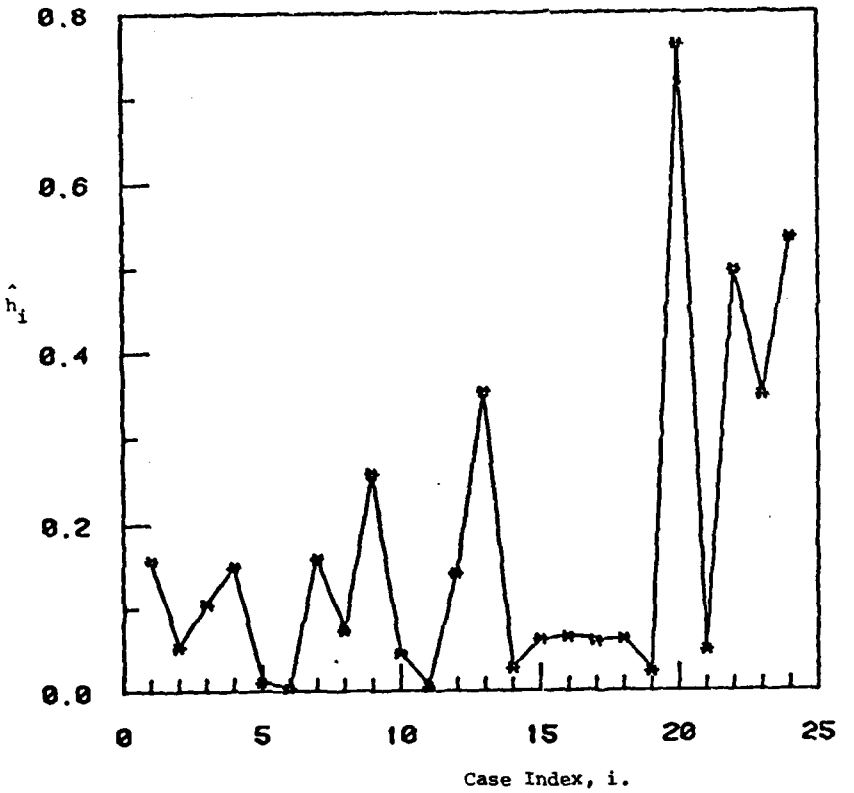
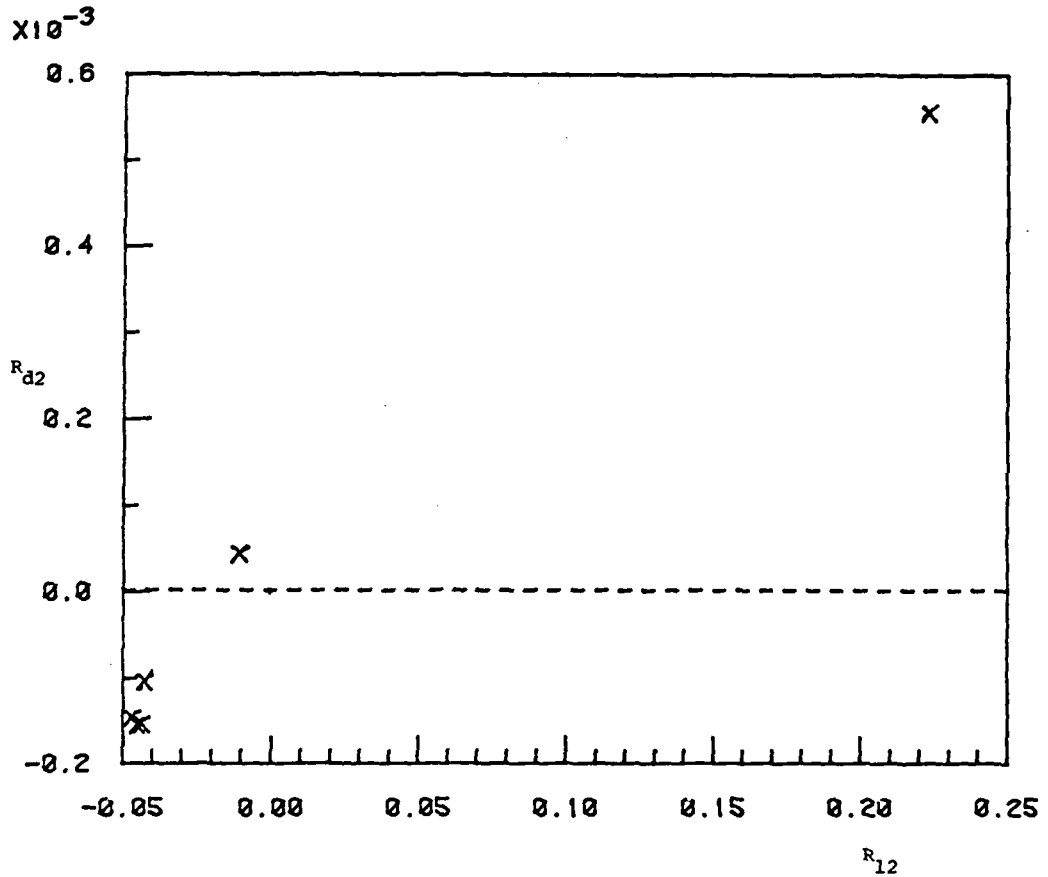


Figure 3: Index plot of $\hat{h}_i$ for model (25).

Figure 4: Added variable plot for the bias of θ_1 in the Michaelis-Menton model; R_{12} = residuals from (f_i^1) on (f_i^2), R_{d2} = residuals from $\underset{\sim}{d}$ on (f_i^2).



REFERENCES

Amari, S. (1982), "Differential Geometry of Curved Exponential Families-Curvature and Information Loss," The Annals of Statistics, 10, 375-385.

Bates, D.M. and Watts, D.G. (1980), " Relative Curvature Measures of Nonlinearity," Journal of the Royal Statistical Society B, 42, 1-25.

Box, G.E.P. and Hill, J.H. (1974), "Correcting Inhomogeneity of Variance with Power Transformation Weighting," Technometrics, 16, 385-389.

Box, M.J. (1971), "Bias in Nonlinear Estimation," Journal of the Royal Statistical Society B, 32, 171-201.

Clarke, G.P.Y. (1980), "Moments of the Least Squares Estimators in a Non-linear Regression Model," Journal of the Royal Statistical Society B, 42, 227-237.

Cook, R.D. and Weisberg, S. (1982), Influence and Residuals in Regression, Chapman and Hall: New York and London.

Cox, D.R. and Snell, E.J. (1968), "A General Definition of Residuals," Journal of the Royal Statistical Society B, 30, 248-275.

Gallant, A.R. (1975), "Nonlinear Regression," The American Statistician, 29, 73-81.

Gillis, P.R., and Ratkowsky, D.A. (1978), "The Behaviour of Estimators of the Parameters of Various Yield-Density Relationships," Biometrics, 34, 191-198.

Hougaard, P. (1981), "The Appropriateness of the Asymptotic Distribution in a Non-linear Regression Model in Relation to Curvature," Research Report 81/9, Statistical Research Unit, Universitetsparken 5, DK-2100 Copenhagen, Denmark.

Kennedy, W. and Gentle, J. (1980), Statistical Computing, Marcel Dekker, Inc.: New York and Basel.

Ratkowsky, D.A. (1983), Nonlinear Regression Modeling, Marcel Dekker, Inc.: New York and Basel.

Skovgaard, I.M. (1981), "Edgworth Expansions of the Distributions of Maximum Likelihood Estimators in the General (non i.i.d.) Case," Scandenavian Journal of Statistics, 8, 227-236.

Stone, M. (1980), "Discussion of the Paper by Dr. Bates and Professor Watts," Journal of the Royal Statistical Society B, 42, 17-19.

Vølund, A. (1978), "Application of the Four-parameter Logistic Model to Bioassay: Comparison with Slope Ratio and Parallel Line Models," Biometrics, 34, 357-365.

Wu, C.F. (1981), "Asymptotic Theory of Nonlinear Least Squares Estimation," Annals of Statistics, 9, 501-513.

RDC/CLT/BCW/mj

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 2645	2. GOVT ACCESSION NO. AD-A139239	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) BIAS IN NONLINEAR REGRESSION		5. TYPE OF REPORT & PERIOD COVERED Summary Report - no specific reporting period
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R.D. Cook, C.L. Tsai and B.C. Wei		8. CONTRACT OR GRANT NUMBER(s) DAAG29-80-C-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS Mathematics Research Center, University of Wisconsin 610 Walnut Street Madison, Wisconsin 53706		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS Work Unit Number 4 - Statistics and Probability
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P.O. Box 12211 Research Triangle Park, North Carolina 27709		12. REPORT DATE February 1984
		13. NUMBER OF PAGES 26
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Diagnostics, Influence, Intrinsic Curvature, Influence, Residuals, Transformations

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We investigate the biases of the residuals and the maximum likelihood parameter estimates from standard, normal-theory nonlinear regression models. Emphasis is placed on determining the influence of individual cases on the biases and on understanding how the residual biases can affect the usefulness of standard diagnostic methods. It is shown that the various bias expressions in the literature are equivalent, that the biases in nonlinear regression can be studied usefully in the context of linear regression, and that diagnostic plots using residuals can be misleading because of substantital residual biases. For a class of partially nonlinear models, it is shown that the maximum intrinsic curvature (Bates and Watts 1980) is closely related to the residual expectations. Finally, the model associated with power transformations of single explanatory

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE (cont.)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT (cont.)

variables in linear regression is investigated in further detail and several numerical illustrations are presented.