The Annals of Statistics
1982, Vol. 10, No. 2, 386-393

ACCOUNTING FOR INTRINSIC NONLINEARITY IN NONLINEAR REGRESSION PARAMETER INFERENCE REGIONS

BY DAVID C. HAMILTON,[1] DONALD G. WATTS[2] AND DOUGLAS M. BATES[3]

University of Wisconsin, Dalhousie University and Queen's University

Joint confidence and likelihood regions for the parameters in nonlinear regression models can be defined using the geometric concepts of sample space and solution locus. Using a quadratic approximation to the solution locus, instead of the usual linear approximation, it is shown that these inference regions correspond to ellipsoids on the tangent plane at the least squares point. Accurate approximate inference regions can be obtained by mapping these ellipsoids into the parameter space, and measures of the effect of intrinsic nonlinearity on inference can be based on the difference between the tangent plane ellipsoids and the sphere which would be obtained using a linear approximation.

1. Introduction.

1.1 *Likelihood and confidence regions.* In a nonlinear regression model, the expected response in the tth experiment depends on the settings of the k independent variables $\mathbf{x}_t = (x_{t1}, x_{t2} \cdots, x_{tk})'$, and on a p-dimensional vector of unknown parameters, $\boldsymbol{\theta}$, through a known response function $f(\boldsymbol{\theta}, \mathbf{x}_t)$. We assume that observed responses deviate from the expected responses due to independent normally distributed random errors with zero mean and constant variance for all $\mathbf{x}_t$. The maximum likelihood, or least squares, estimates $\hat{\boldsymbol{\theta}}$ minimize the sum of squares function

$$S(\boldsymbol{\theta}) = \textstyle\sum_{t=1}^{n} \{y_t - f(\boldsymbol{\theta}, \mathbf{x}_t)\}^2. \tag{1.1}$$

The response for n experiments is denoted by the vector $\mathbf{y} = (y_1, \cdots, y_n)'$ which lies in the n-dimensional sample space, and the corresponding expected responses may be similarly collected into the vector $\boldsymbol{\eta}(\boldsymbol{\theta}) = (f(\boldsymbol{\theta}, \mathbf{x}_1), \cdots, f(\boldsymbol{\theta}, \mathbf{x}_n))'$. The set of all possible expected response vectors, as $\boldsymbol{\theta}$ varies, comprise a p-dimensional surface called the solution locus (Box and Lucas, 1959). Therefore the estimates $\hat{\boldsymbol{\theta}}$ can be simply characterized geometrically as the parameter vector whose image on the solution locus, $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$, is closest to $\mathbf{y}$. In spite of its simple geometric definition, $\hat{\boldsymbol{\theta}}$ must usually be obtained by minimizing (1.1) using an iterative computer program.

Joint inference (i.e. confidence or likelihood) regions for the parameters also have simple geometric definitions as the image, in parameter space, of inference regions on the solution locus. The likelihood region, consisting of all values of θ for which

$$S(\boldsymbol{\theta}) - S(\hat{\boldsymbol{\theta}}) \le \delta^2, \tag{1.2}$$

contains all values of $\boldsymbol{\theta}$ which correspond to points on the solution locus which are within the distance $\sqrt{\{S(\hat{\boldsymbol{\theta}}) + \delta^2\}}$ of $\mathbf{y}$. Beale (1960) recommended such a region as an approximate $100(1-\alpha)\%$ confidence region with δ^2 equal to $\sigma^2\chi^2(p; \alpha)$ if σ^2 is known, and $ps^2F(p, \nu; \alpha)$ if σ^2 is estimated by s^2 with ν degrees of freedom.

Received December 1980; revised July 1981, November 1981.

[1] Research supported by NSERC and the Dalhousie University Graduate Studies Research Development Fund.

[2] Research supported by NSERC and the Queen's Advisory Research Council.

[3] Research supported by US Army Contract DAAG 29-80-C-0041 and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

Key words and phrases. Intrinsic curvature, nonlinear regression, approximate inference regions.
AMS 1970 *subject classifications.* Primary 62J02, secondary 62F25.

Exact confidence regions based on the χ^2 and F distributions are obtained by comparing squared lengths of projections of the error vector $\mathbf{e}(\boldsymbol{\theta}) = \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})$ onto orthogonal subspaces of sample space. If $\boldsymbol{\theta}$ is the true value, $\mathbf{e}(\boldsymbol{\theta})'P(\boldsymbol{\theta})\mathbf{e}(\boldsymbol{\theta})$ and $\mathbf{e}(\boldsymbol{\theta})'\{I - P(\boldsymbol{\theta})\}\mathbf{e}(\boldsymbol{\theta})$ have independent $\sigma^2\chi^2$ distributions with p and $n - p$ degrees of freedom, where $P(\boldsymbol{\theta})$ is the matrix of projection onto the tangent plane at $\boldsymbol{\eta}(\boldsymbol{\theta})$. Exact confidence regions therefore include all values $\boldsymbol{\theta}$, or points $\boldsymbol{\eta}(\boldsymbol{\theta})$, satisfying

$$\mathbf{e}(\boldsymbol{\theta})'P(\boldsymbol{\theta})\mathbf{e}(\boldsymbol{\theta}) \le \sigma^2\chi^2(p; \alpha)$$

if σ^2 is known, or

$$\mathbf{e}(\boldsymbol{\theta})'P(\boldsymbol{\theta})\mathbf{e}(\boldsymbol{\theta}) \le ps^2F(p, \nu; \alpha) \tag{1.3}$$

if σ^2 is estimated independently by s^2. For the case of no replications, one may use

$$\mathbf{e}(\boldsymbol{\theta})'P(\boldsymbol{\theta})\mathbf{e}(\boldsymbol{\theta}) \le f\mathbf{e}(\boldsymbol{\theta})'\mathbf{e}(\boldsymbol{\theta})/(1 + f) \tag{1.4}$$

where $f = pF(p, n - p; \alpha)/n - p$. These regions have been discussed by Beale (1960), Gallant (1975, 1976) and Sundararaj (1978). Similar exact regions are described by Williams (1962), Halperin (1963) and Hartley (1964).

Determination of likelihood and exact confidence regions usually requires substantial computation. In addition, it is difficult to display exact regions when there are more than two parameters, so there is considerable value in obtaining adequate approximate inference regions which can be easily presented and summarized.

1.2 *The linear approximation.* A linear Taylor's approximation of the model function is often used to derive approximate inference regions. For convenience, we assume that the data $\mathbf{y}$, the model, and all the derivatives have been divided by $\rho = s\sqrt{p}$, so that the approximate inference region is

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'V.'V.(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \le F, \tag{1.5}$$

where $V.$ is the $n \times p$ matrix of partial derivatives with ijth entry

$$v_{ij} = \frac{\partial f}{\partial \theta_j}(\boldsymbol{\theta}, x_i)|_{\hat{\boldsymbol{\theta}}}, \quad i = 1, \cdots, n, j = 1, \cdots p,$$

and F is $F(p, n - p; \alpha)$ for likelihood or no-replications confidence regions and is $F(p, \nu; \alpha)$ for a confidence region based on ν degrees of freedom from replications. The approximation (1.5) assumes that, over the region of interest, the mapping of $\boldsymbol{\theta}$ into $\boldsymbol{\eta}(\boldsymbol{\theta})$ is

$$\hat{\boldsymbol{\eta}} + V.(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}). \tag{1.6}$$

Thus the tangent plane projection of the solution locus inference region is a sphere and the mapping from the tangent plane to the parameter space is linear. Using geometric reasoning, Beale (1960) and Bates and Watts (1980) note that this approximation will be acceptable only if the solution locus is sufficiently flat to be replaced by the tangent plane and if straight, parallel equispaced lines in the parameter space map into straight, parallel equispaced lines on the solution locus. They used second derivatives of the model function to derive curvature measures of intrinsic and parameter effects nonlinearity to assess the validity of these assumptions.

1.3 *Geometrical approach to approximating inference regions.* Geometrical reasoning reveals that there can be two stages in obtaining approximate inference regions. The first involves finding the solution locus inference region and obtaining its projection onto the tangent plane, and the second involves mapping that tangent plane region into the parameter plane. The latter problem is addressed in a separate paper (Bates and Watts, 1981). In this paper we focus on the effects of intrinsic nonlinearity on the tangent plane projection of solution locus inference regions, and derive improved approximate regions.

2. Improved approximate tangent plane projections of solution locus inference regions. In this section we derive improved approximate projections of solution locus inference regions onto the tangent plane, using a quadratic approximation. These improved approximate projection regions are found to be ellipsoids, in contrast with the spherical projections obtained with the linear approximation.

Consider a point $\boldsymbol{\theta}$ in the parameter space. This point maps through the nonlinear transformation, $\boldsymbol{\eta}(\boldsymbol{\theta})$, to the point on the tangent plane with coordinates

$$\tau = U'\{\boldsymbol{\eta}(\boldsymbol{\theta}) - \hat{\boldsymbol{\eta}}\}. \tag{2.1}$$

In (2.1) the columns of U form an orthonormal basis for the tangent plane (1.6) and U is related to $V_.$ by $U = V_.L$, where L is a $p \times p$ matrix and is the Jacobian of the mapping from $\boldsymbol{\theta}$ to τ at $\hat{\boldsymbol{\theta}}$.

The coordinates, τ, provide a natural reference system for the solution locus and approximations to it. An approximation to the solution locus by a quadratic surface centered at $\hat{\boldsymbol{\eta}}$ is

$$\tilde{\boldsymbol{\eta}}(\tau) = \hat{\boldsymbol{\eta}} + U\tau + [N][\tau' A \tau]/2, \tag{2.2}$$

where the columns of the $n \times (n - p)$ matrix N are an orthonormal basis for the space orthogonal to the tangent plane, A is the intrinsic curvature array, and the square bracket multiplication involves summation over the numerator index (Bates and Watts, 1980). This simple expression results because, as shown in the appendix, $d\boldsymbol{\eta}/d\tau|_{\boldsymbol{0}} = U$ and $d^2\boldsymbol{\eta}/d\tau^2|_{\boldsymbol{0}} = [N][A]$.

Exact solution locus inference regions are defined by equations (1.2) through (1.4). Approximate solution locus inference regions and their projections onto the tangent plane can be defined by substituting approximate error vectors $\tilde{\mathbf{e}} = \mathbf{y} - \tilde{\boldsymbol{\eta}}(\tau)$ for $\mathbf{e}(\boldsymbol{\theta})$ in those equations. The approximate error vector at the point $\boldsymbol{\eta}(\tau)$ is

$$\tilde{\mathbf{e}} = \hat{\mathbf{e}} - U\tau - [N][\tau' A \tau]/2. \tag{2.3}$$

2.1 *Likelihood regions.* The squared total length of $\tilde{\mathbf{e}}$ is

$$\tilde{\mathbf{e}}'\tilde{\mathbf{e}} = \hat{\mathbf{e}}'\hat{\mathbf{e}} + \tau'(I - B)\tau + (\tau' A \tau)'(\tau' A \tau)/4 \simeq \hat{\mathbf{e}}'\hat{\mathbf{e}} + \tau'(I - B)\tau, \tag{2.4}$$

where $B = [\hat{\mathbf{e}}'N][A]$ is the $p \times p$ matrix obtained from the inner product of the rotated residual vector $N'\hat{\mathbf{e}}$ and the intrinsic curvature array. The matrix B is referred to as the "effective residual curvature matrix" because it gives the effective normal curvatures relative to the residual vector $\hat{\mathbf{e}}$.

Equation (2.4) may be explained geometrically as in Figure 1a, which shows, relative to the origin $\hat{\boldsymbol{\eta}}$, the three dimensions spanned by the vectors $U\tau$, $\mathbf{q} = [N][\tau' A \tau]/2$ and $\hat{\mathbf{e}}$. The point $\tilde{\boldsymbol{\eta}}$ projects onto the plane defined by the vectors $\mathbf{q}$ and $\hat{\mathbf{e}}$ at the terminus of $\mathbf{q}$, so that

$$\tilde{\mathbf{e}}'\tilde{\mathbf{e}} = \tau'\tau + (\hat{\mathbf{e}} - \mathbf{q})'(\hat{\mathbf{e}} - \tau) = \tau'\tau + \hat{\mathbf{e}}'\hat{\mathbf{e}} + \mathbf{q}'\mathbf{q} - 2\hat{\mathbf{e}}'\mathbf{q}. \tag{2.5}$$

The value of $\mathbf{q}'\mathbf{q}$ is $\|\tau' A \tau\|)^2/4$ and $2\hat{\mathbf{e}}'\mathbf{q}$ equals $\tau' B \tau$, which gives the equality in (2.4).

Now consider the vectors $\hat{\mathbf{e}}$, $\mathbf{q}$ and $\hat{\mathbf{e}} - \mathbf{q}$ in the plane defined by $\hat{\mathbf{e}}$ and $\mathbf{q}$, as shown in Figure 1b, where we have exaggerated the size of $\mathbf{q}$ relative to $\hat{\mathbf{e}}$ for illustrative purposes. A linear approximation to the solution locus would yield an approximate error vector with squared length

$$\|\tau\|^2 + \|\hat{\mathbf{e}}\|^2 = \|\tau\|^2 + OX^2,$$

the quadratic approximation would yield an approximate error vector with squared length

$$\|\tau\|^2 + \|\hat{\mathbf{e}} - \mathbf{q}\|^2 = \|\tau\|^2 + OY^2,$$

and the approximation used in (2.4) would yield a squared length

$$\|\tau\|^2 + OZ^2.$$

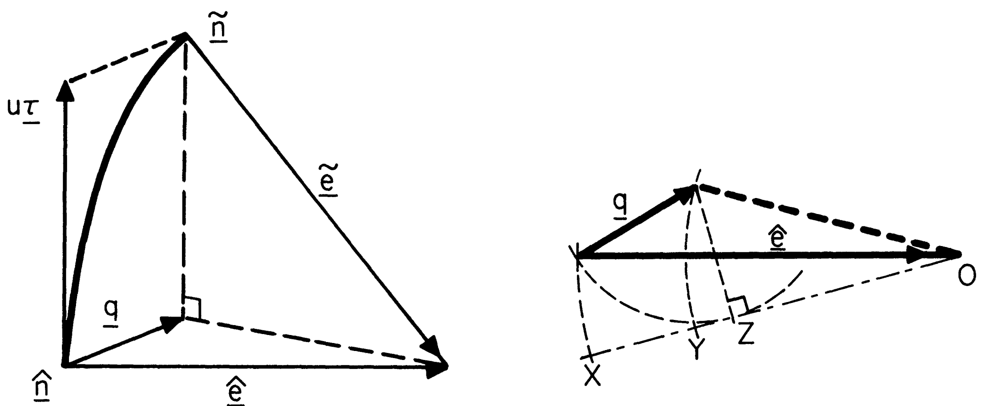

FIG. 1a FIG. 1b

FIG. 1. *Approximating* $\|\tilde{\mathbf{e}}\|$.

Thus, neglecting the term in (2.4) which involves fourth powers of the length of τ corresponds to adjusting the linear approximation by the segment XZ instead of XY.

Inserting (2.4) into the scaled version of (1.2) gives the approximate solution locus region defined by

$$\tau'(I - B)\tau = F(p, n - p; \alpha). \tag{2.5}$$

Since $I - B$ is positive definite, as shown below, the approximate tangent plane projection of the solution locus likelihood region is an ellipsoid. Beale (1960, equation (2.3), page 46) also noted that the sum of squares contours were approximately ellipsoidal in the τ coordinates.

Box and Coutie (1956) recommended using

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'W(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq F$$

as an approximate confidence region for $\boldsymbol{\theta}$, where

$$W = \tfrac{1}{2}\, \partial^2 S/\partial\boldsymbol{\theta}^2 \,|_{\hat{\theta}}.$$

In the notation of this paper,

$$W = V.'V. - [\mathbf{e}'][V..] = (L^{-1})'(I - B)L^{-1}.$$

Thus the confidence region based on the second derivatives of the sum of squares function may be considered the image in parameter space of the ellipsoidal tangent plane likelihood region (2.5), assuming there is no parameter effects nonlinearity so that there is a linear mapping from the tangent plane to the parameter plane. Since $S(\hat{\boldsymbol{\theta}})$ is a local minimum, it also follows that W, and hence $(I - B)$, is positive definite.

2.2 *Confidence regions with replications.* Similar approximations can be obtained for the exact confidence regions (1.3) and (1.4) by examining the components of the error vector relative to the quadratic solution locus approximation. Differentiating (2.2) with respect to τ reveals that the tangent plane to the quadratic surface at $\boldsymbol{\eta}(\tau)$ is defined by the columns of a matrix $(U + NM)$, where the $(n - p) \times p$ matrix $M = A\tau$ has ijth entry $\sum_{r=1}^{p} a_{ijr}\tau_r$. The projection matrix onto this plane is

$$(U + NM)(I + M'M)^{-1}(U + NM)',$$

since $(U + NM)'(U + NM) = (I + M'M)$, and the squared length of the tangential component of the error vector $\mathbf{e}$ is

$$\tau'(I - B + M'M/2)'(I + M'M)^{-1}(I - B + M'M/2)\tau. \tag{2.6}$$

Assuming that terms involving $M'M$ and its products may be neglected, since they produce fourth order and higher products of τ, the approximate tangent plane region for the exact confidence region (1.3) is bounded by values of τ satisfying

$$\tau'(I-B)^2\tau = \chi^2(p;\alpha)$$

or

$$\tau'(I-B)^2\tau = F(p,\nu;\alpha), \tag{2.7}$$

which is the equation of an ellipsoid.

2.4 *Confidence regions with no replications.* When the variance is unknown and cannot be estimated independently, the length of the orthogonal component of the error vector must also be considered. Using equations (1.4), (2.4) and (2.6), the tangent plane coordinates of the boundary of the approximate solution locus confidence region satisfy

$$\begin{aligned}\tau'(I-B+M'M/2)(I+M'M)^{-1}(I-B+M'M/2)\tau \\ = \{\hat{e}'\hat{e}+\tau'(I-B)\tau+\tau'M'M\tau/4\}f/(1+f),\end{aligned} \tag{2.8}$$

and, assuming as before that terms in $M'M$ may be ignored, (2.8) simplifies to

$$\tau'(I-B)\{I-(1+f)B\}\tau = F(p,n-p;\alpha), \tag{2.9}$$

which is also the equation of an ellipsoid, provided $I-(1+f)B$ is positive definite.

The tangent plane ellipsoids derived above depend on the residual vector-solution locus configuration through the matrix B, whose entries give the projections of the acceleration vectors on the residual vector. If the solution locus curves towards (away from) the residual vector all the projections are positive (negative) and all the eigenvalues of B are positive (negative) and the tangent plane ellipsoid is larger (smaller) than the linear approximation sphere. This is intuitively correct because then the solution locus is curving towards (away from) the residual vector and so more points on the solution locus are nearer (father from) the observation point than if the solution locus was a plane.

3. Assessing the effect of intrinsic nonlinearity on inference regions. The extent to which the tangent plane inference ellipsoid differs from the linear approximation sphere gives a direct indication of the effects of intrinsic nonlinearity on inference. Each of the matrices $(I-B)$, $(I-B)^2$ and $(I-B)\{I-(1+f)B\}$ has the same eigenvectors as B, and has eigenvalues which are simple functions of the eigenvalues, $\lambda_1 < \cdots < \lambda_p$, of B. Therefore the axes of the different inference ellipsoids point in the same directions and the length of the ith axis, as a proportion of the radius of the linear approximation sphere, is

$(1-\lambda_i)^{-1/2}$	for likelihood regions,
$(1-\lambda_i)^{-1}$	for confidence regions based on replications,
$[(1-\lambda_i)\{1-(1+f)\lambda_i\}]^{-1/2}$	for no-replication confidence regions.

The extreme axis length ratios, obtained using λ_1 and λ_p, can therefore be used to assess directly the effect of intrinsic nonlinearity on likelihood and confidence regions.

Previous attempts have been made to assess the influence of intrinsic nonlinearity on likelihood and confidence regions. Beale (1960) proposed that intrinsic nonlinearity be accounted for by inflating the right hand side of (1.2) by the squared factor

$$m^2 = 1 + \frac{n(p+2)}{p(n-p)}N_\phi \tag{3.1}$$

to ensure that the coverage probability of the likelihood region is at least as large as the nominal confidence level. N_ϕ was therefore proposed as a measure of intrinsic nonlinearity, and this role can be extended to the factor m.

Bates and Watts (1980) showed that N_ϕ is one quarter of the mean squared intrinsic curvature, and hence the factor m is easily calculated from the array A. They also proposed that intrinsic nonlinearity be measured by the maximum intrinsic curvature, Γ^N, which is also easily calculated from A.

The axis ratios for likelihood and no-replications confidence regions and the intrinsic nonlinearity measures discussed above have been calculated for 18 data sets analyzed in Bates and Watts (1981) and are shown below in Table 1, which reveals several interesting features. First, Beale's tangent plane inflation factor m is always extremely close to unity, suggesting that the confidence level associated with the likelihood region (1.2) is very close to the nominal level. Second, the maximum relative curvature $\Gamma^N\sqrt{F}$ exceeds .5 for only two data sets, suggesting that the solution locus may be adequately approximated by a plane in 16 of the 18 cases at the 95% confidence level. Note that data set 21, for which the ellipsoidal approximation breaks down at the 95% level as shown below, has an extreme value for m and $\Gamma^N\sqrt{F}$. Turning next to the likelihood axis ratios in columns 4 and 5, we see that, except for the low value of .86 with data set 21 and the high values of 1.08 and 1.09, all the axis ratios are within .05 of unity. Thus the effect of intrinsic nonlinearity on likelihood regions is small. For the no-replications confidence region, however, 11 of the axis ratios in columns 6 and 7 deviate more than .1 from unity so that intrinsic nonlinearity has greater influence on confidence regions, as expected from the expressions.

Comparing the various columns, we see a positive correlation between Beale's factor m and the scaled maximum curvature $\Gamma^N\sqrt{F}$. This correlation was noted in Bates and Watts (1980), who suggested the dependence $N_\phi\,(\gamma_{rms}^N)^2/4 \simeq (\Gamma^N)^2/(1+p)^2$. Beale's inflation factor and the maximum curvature of Bates and Watts both are negatively correlated with the smallest axis ratios in columns 3 and 5. Finally, there is little relationship between m and $\Gamma^N\sqrt{F}$ and the largest axis ratios in columns 4 and 6.

In deriving the approximate tangent plane inference ellipsoids, terms involving fourth powers of the length of τ were ignored. A crude upper bound for the proportion of the ignored term relative to the quadratic term can be calculated from the expression in the previous paragraph as

$$\frac{(\Gamma^N)^2F}{4(1-\lambda_p)^2} \simeq F(\gamma_{rms}^N)^2\frac{(1+p)^2}{16(1-\lambda_p)^2}.$$

TABLE 1

Axis Ratios and Nonlinearity Measures (95% Confidence Level)

data set	m	$\Gamma^N\sqrt{F}$	axis ratios: likelihood smallest	likelihood largest	confidence smallest	confidence largest
1	1.00	.08	1.00	1.00	1.00	1.00
2	1.00	.13	1.00	1.00	1.00	1.00
3	1.00	.20	1.00	1.08	1.00	1.28
4	1.00	.15	1.00	1.05	1.00	1.16
5	1.01	.47	.95	1.00	.87	1.00
9	1.00	.41	.99	1.03	.98	1.10
13	1.00	.03	1.00	1.00	.99	1.00
14	1.00	.29	.96	1.00	.89	1.01
15	1.00	.16	1.00	1.02	.99	2.81
16	1.00	.08	.99	1.00	.97	1.00
17	1.00	.43	.92	1.00	.82	1.00
18	1.00	.00	1.00	1.00	1.00	1.00
19	1.00	.03	1.00	1.01	.99	1.02
20	1.00	.32	1.00	1.03	1.00	1.06
21	1.02	1.43	.86	1.03	.72	1.07
22	1.00	.14	1.00	1.09	1.00	1.27
23	1.00	.15	.95	1.01	.88	1.03
24	1.00	.54	.95	1.02	.88	1.05

This proportion was calculated for the data sets in Table 1, and was found to be less than 6% for every data set except data set 21, which had a value of 49%. Thus the ellipsoidal approximation is acceptable for the 95% regions except for that one data set.

In conclusion, in this paper we have shown that intrinsic nonlinearity can be accounted for by using ellipsoids rather than spheres on the tangent plane and that the shape of the ellipsoid is determined by the components of the acceleration vectors in the direction of the residual vector. The ellipsoids are therefore directly related to the actual design and responses from an experiment, and hence are specific to the particular inference situation. We have also shown that summary intrinsic curvature measures, such as the maximum or mean square curvature, are not closely related to the more direct indicators presented here.

APPENDIX

The first and second partial derivatives of $\boldsymbol{\theta}$ with respect to τ are used in Section 2 to obtain a second-order Taylor's series approximation to the mapping from τ to $\boldsymbol{\theta}$. Differentiating the identity $\tau(\boldsymbol{\theta}(\tau)) = \tau$ twice using the chain rule gives

$$(d\tau/d\boldsymbol{\theta})(d\boldsymbol{\theta}/d\tau) = I, \quad (d\boldsymbol{\theta}/d\tau)'(d^2\tau/d\boldsymbol{\theta}^2)(d\boldsymbol{\theta}/d\tau) + [d\tau/d\boldsymbol{\theta}][d^2\boldsymbol{\theta}/d\tau^2] = \mathbf{0},$$

where the square brackets denote numerator multiplications (Bates and Watts, 1980). Rearranging and substituting known quantities yields

$$d\boldsymbol{\theta}/d\tau\,|_0 = (d\tau/d\boldsymbol{\theta}\,|_{\hat{\theta}})^{-1} = (U'V.)^{-1} = L$$

and

$$\begin{aligned} d^2\boldsymbol{\theta}/d\tau^2\,|_0 &= -[d\boldsymbol{\theta}/d\tau]_0[(d\boldsymbol{\theta}/d\tau)'(d^2\tau/d\boldsymbol{\theta}^2)(d\boldsymbol{\theta}/d\tau)]_0 \\ &= -[L][L'[U'][V..]L] = -[L][A^T_{..}], \end{aligned}$$

where $V..$ is the array of second derivatives and $A^T_{..}$ is the parameter-effects array derived in Bates and Watts (1980).

The first and second partial derivatives of $\boldsymbol{\eta}(\boldsymbol{\theta}(\tau))$ with respect to τ may also be obtained. Differentiating using the chain rule gives

$$d\boldsymbol{\eta}/d\tau\,|_0 = (d\boldsymbol{\eta}/d\boldsymbol{\theta})\,|_{\hat{\theta}}(d\boldsymbol{\theta}/d\tau)\,|_0 = V.L = U$$

and

$$\begin{aligned} d^2\boldsymbol{\eta}/d\tau^2\,|_0 &= (d\boldsymbol{\theta}/d\tau\,|_0)(d^2\boldsymbol{\eta}/d\boldsymbol{\theta}^2\,|_{\hat{\theta}})(d\boldsymbol{\theta}/d\tau\,|_{\hat{\theta}}) + [d\boldsymbol{\eta}/d\boldsymbol{\theta}]\,|_{\hat{\theta}}[d^2\boldsymbol{\theta}/d\tau^2\,|_0] \\ &= L'V..L - [V.][L][A^T_{..}] \\ &= [I - UU'][L'V..L] = [NN'][L'V..L] = [N][A^N_{..}], \end{aligned}$$

where the columns of N form an orthonormal basis for the space orthogonal to the tangent plane and $A^N_{..}$ is the intrinsic curvature array (Bates and Watts, 1980).

REFERENCES

BATES, D. M. and WATTS, D. G. (1980). Relative curvature measures of nonlinearity, (with discussion). *J. Roy. Statist. Soc. Ser. B* **40** 1–25.

BATES, D. M. and WATTS, D. G. (1981). Parameter transformations for improved approximate confidence regions in nonlinear least squares. *Ann. Statist.* **9** 1152–1167.

BEALE, E. M. L. (1960). Confidence regions in nonlinear estimation (with discussion). *J. Roy. Statist. Soc. B* **22** 41–88.

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 No.* 1, 100–107.

BOX, G. E. P. and LUCAS, H. L. (1959). Design of experiments in nonlinear situations. *Biometrika* **46** 77–90.

GALLANT, A. R. (1975). Nonlinear regression. *Amer. Statist.* **29** 2, 73–81.

GALLANT, A. R. (1976). Confidence regions for the parameters of a nonlinear regression model. Inst. of Stats. Mimeo. Series No. 1077, North Carolina State University, Raleigh.
HALPERIN, M. (1963). Confidence interval estimation in nonlinear regression. *J. Roy. Statist. Soc. Ser. B* **25** 330–333.
HARTLEY, H. O. (1964). Exact confidence regions for the parameters in nonlinear regression laws. *Biometrika* **51** 347–353.
SUNDARARAJ, N. (1978). A method for confidence regions for nonlinear models. *Austral. J. Statist.* **20** 270–274.
WILLIAMS, E. J. (1962). Exact fiducial limits in nonlinear estimation. *J. Roy. Statist. Soc. Ser. B* **24** 125–139.

DAVID C. HAMILTON
DEPARTMENT OF MATHEMATICS, STATISTICS AND COMPUTING SCIENCE
DALHOUSIE UNIVERSITY
HALIFAX, NOVA SCOTIA

DONALD G. WATTS
DEPARTMENT OF MATHEMATICS AND STATISTICS
QUEEN'S UNIVERSITY
KINGSTON, ONTARIO

DOUGLAS M. BATES
DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN