CONFIDENCE REGIONS IN SEMILINEAR REGRESSION

by

Mark Knowles
Alza Corporation
and
David Siegmund
Stanford University

TECHNICAL REPORT NO. 7
JANUARY 1990

Prepared Under Contract 89-0355
For The Air Force Office of Scientific Research

Reproduction in Whole or in Part is Permitted
for any Purpose of the United States Government

Approved for public release; distribution unlimited

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
CONFIDENCE REGIONS IN SEMILINEAR REGRESSION

by

Mark Knowles
Alza Corporation
and
David Siegmund
Stanford University

TECHNICAL REPORT NO. 7
JANUARY 1990

PREPARED UNDER CONTRACT 89-0355
FOR THE AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

Also prepared under National Security Agency Grant MDA 904-89-H-2040 and issued as Technical Report No. 6, Department of Statistics, Stanford University; and under National Science Foundation Grant DMS 89-05874 and issued as Technical Report No. 340, Department of Statistics, Stanford University

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
1. Introduction.

For the general nonlinear regression model \( y_i = \eta_i(\lambda) + \varepsilon_i \) (\( i = 1, \ldots, m \)), where \( \varepsilon_1, \ldots, \varepsilon_m \) are independent \( N(0, \sigma^2) \), there exists a variety of methods for obtaining confidence regions for the parameter vector \( \lambda \) or components thereof. Some of the methods give exact regions, at least in special cases (e.g. Halperin, 1963), whereas others are based on asymptotic approximations (Beale, 1960; Jennrich, 1969). A common characteristic is that one understands them in terms of local, linear or quadratic, approximations to the surface \( \eta(\lambda) \), which can be misleading when the curvature of the surface is large. Various diagnostic measures of curvature (Beale, 1960; Bates and Watts, 1980) assist one in deciding whether these local approximations are adequate.

In this paper we consider the very special case that all parameters but one enter into \( \eta \) linearly, say \( \lambda = (\theta, \beta, \beta_1, \ldots, \beta_p) \) and

\[
\eta_i(\lambda) = \sum_j \beta_j x_{ij} + \beta f_i(\theta). \tag{1}
\]

A number of common models are of this special form. Examples are (i) the Michaelis-Menten model, in which \( \lambda = (\theta, \beta) \) and

\[
\eta_i(\lambda) = \beta x_i / (\theta + x_i); \tag{2}
\]

(ii) the asymptotic growth model, in which \( \lambda = (\theta, \beta) \) and

\[
\eta_i(\lambda) = \beta \{1 - \exp(-\theta x_i)\}; \tag{3}
\]

and (iii) the broken line regression model, in which \( \lambda = (\theta, \beta, \beta_1, \beta_2) \) and

\[
\eta_i(\lambda) = \beta_1 + \beta_2 x_i + \beta (x_i - \theta)^+. \tag{4}
\]

In fact, for the most part we consider the still simpler case that \( p = 0 \), so \( \lambda = (\theta, \beta) \) and

\[
\eta_i(\lambda) = \beta f_i(\theta), \tag{5}
\]

which includes (2) and (3); and we leave a detailed discussion of (4) to a future paper. For the model (5) we provide accurate approximations to likelihood ratio confidence regions
for \( \theta \) and jointly for \((\theta, \beta)\), which do not require additional assumptions about the design or sample size and are valid even when there is considerable nonlinearity. Our method uses a modification of the upcrossings argument developed by Johnstone and Siegmund (1989) in their discussion of hypothesis tests for the existence of a nonlinear parameter. In general it gives an upper bound for the non-coverage probability and hence a conservative confidence region, although in many particular cases the approximation seems very good and the region essentially exact.

Although our method requires somewhat more numerical computation than others, the amount is not a serious hindrance to its use in the class of simple models considered here. However, the computational requirements become much more onerous when more than one nonlinear parameter is involved, and hence one goal of this paper is to investigate whether the virtues of our method make its further development worthwhile. A second goal is to use our likelihood ratio regions to get a clearer picture of the strengths and weaknesses of other methods when curvature is large and hence local approximations are suspect.

The paper is organized as follows. Section 2 contains our basic theoretical results. In Section 3 we consider a family of artificial examples for which we can easily perform a large number of computations. For these examples we compare confidence regions obtained by our method with Halperin’s (1963) exact regions and with approximate regions obtained by assuming an \( F \)-like version of the likelihood ratio statistic has an \( F \) distribution. In Section 4 we apply our method to some standard “real” examples. Section 5 tries to draw some general conclusions, and details of some mathematical derivations are given in two appendices.

2. Mathematical Results.

In vector notation our model is \( y = \beta f(\theta) + \varepsilon \), where \( \varepsilon \) is \( N(0, \sigma^2 I_m) \). The scalar parameters \( \theta, \beta \), and \( \sigma^2 \) are all unknown. It will be convenient to put \( \gamma(\theta) = f(\theta)/\|f(\theta)\| \), so as \( \theta \) varies in its set of possible values \( \gamma \) traces out a curve in \( S^{m-1} \), the unit sphere in \( \mathbb{R}^m \). Given a value of \( \theta \), the maximum likelihood estimates of \( \beta \) and \( \sigma^2 \) are \( \hat{\beta}_\theta = \)
\[ \langle f(\theta), y \rangle / \| f(\theta) \|^2 \]

and

\[ \hat{\sigma}^2 = m^{-1} \| y - \hat{\beta}_f f(\theta) \|^2 = m^{-1} \| y \|^2 (1 - \langle \gamma(\theta), U \rangle^2), \]

where \( U = y / \| y \| \in S^{m-1} \). The likelihood ratio statistic for testing \( H_0 : \theta = \theta_0 \) is

\[ \Lambda(\theta_0) = -\frac{1}{2} m \log \left( \min_{\theta} \frac{\hat{\sigma}^2}{\hat{\sigma}^2_{\theta_0}} \right), \]

which by (6) equals

\[ -\frac{1}{2} m \log \left\{ (1 - \max_{\theta} \langle \gamma(\theta), U \rangle^2) / (1 - \langle \gamma(\theta_0), U \rangle^2) \right\}. \]

A 100(1 - \alpha)\% confidence region for \( \theta \) is the set of all values \( \theta_0 \) not rejected by an \( \alpha \) level likelihood ratio test. In order that the probability under \( H_0 \) of the event \( \Lambda(\theta_0) > b \) not depend on the unknown nuisance parameters \((\beta, \sigma)\) we evaluate it conditionally given the sufficient statistics under \( H_0, \| y \| \) and \( \langle f(\theta_0), y \rangle \). Hence our \((1 - \alpha)100\%\) confidence region for \( \theta \) is the set of all \( \theta_0 \) such that

\[ \Pr \{ \Lambda(\theta_0) > \{ \Lambda(\theta_0) \}_{\text{obs}} \| y \|, \langle f(\theta_0), y \rangle \} > \alpha. \]

Since this probability does not depend on the values of \( \beta \) and \( \sigma^2 \), we assume that \( \beta = 0 \) and \( \sigma^2 = 1 \), so \( y \) is standard normal. Then \( U = y / \| y \| \) is uniformly distributed on \( S^{m-1} \). Hence by (8) and the slight generalization of Basu’s Theorem given in the Appendix, we see that our confidence region is the set of all \( \theta_0 \) such that

\[ \Pr \left\{ \max_{\theta} \left| \langle \gamma(\theta), U \rangle \right| > w \left| \langle \gamma(\theta_0), U \rangle \right| \right\} > \alpha, \]

where \( w \) denotes the observed value of \( \max_{\theta} \left| \langle \gamma(\theta), U \rangle \right| \) and \( U \) is uniform on \( S^{m-1} \).

This definition has a natural geometric interpretation: given that \( U = y / \| y \| \) is a certain geodesic distance in \( S^{m-1} \) from the hypothesized direction \( \gamma(\theta_0) \), we reject \( \theta_0 \) if in fact \( U \) lies sufficiently close to \( \gamma(\theta) \) or \(-\gamma(\theta)\) for some alternative value of \( \theta \). The reason both \( \gamma \) and \(-\gamma\) enter our considerations is that the unknown nuisance parameter \( \beta \) can be positive or negative. The probability in (9) has a related geometric interpretation. Given \( \langle \gamma(\theta_0), U \rangle = z \in (-w, w) \) this probability is the relative volume of that part of the \( m - 2 \).
dimensional sphere in $S^{m-1}$ of geodesic radius $\cos^{-1}(z)$ about $\gamma(\theta_0)$ which is contained inside the tubes of geodesic radius $\cos^{-1}(w)$ about the curves $\gamma$ and $-\gamma$. See Figure 1.

By an easy adaptation of the standard upcrossing argument (e.g. Leadbetter, Lindgren, and Rootzén, 1983; Johnstone and Siegmund, 1989) we have

$$\Pr\left\{ \max_{\theta_0 \leq \theta \leq \theta_1} \langle \gamma(\theta), U \rangle > w \mid \langle \gamma(\theta_0), U \rangle = z \right\} \leq \int_{\theta_0}^{\theta_1} E\left\{ \langle \dot{\gamma}(\theta), U \rangle^+ \mid \langle \gamma(\theta), U \rangle = w, \langle \gamma(\theta_0), U \rangle = z \right\} \Pr\left\{ \langle \gamma(\theta), U \rangle = w \mid \langle \gamma(\theta_0), U \rangle = z \right\} d\theta,$$

(10)

where $\dot{h}(\theta) = dh/d\theta$ and $\Pr(X = x)$ denotes the probability density function of the random variable $X$ evaluated at $x$.

In order to express the integral on the right hand side of (10) in a form suited for numerical evaluation it is helpful to introduce the following notation. Let $\rho(\theta) = \langle \gamma(\theta_0), \gamma(\theta) \rangle$,

$$\mu = \dot{\rho}(z - \rho w)/(1 - \rho^2),$$

(11)

and

$$\tau^2 = \{ \| \dot{\gamma} \|^2 - \dot{\rho}^2/(1 - \rho^2) \} \{ 1 - (w^2 + z^2 - 2wz)/(1 - \rho^2) \},$$

(12)

Also let

$$f_{k-1}(x) = \frac{\Gamma(k/2)}{\pi^{1/2}\Gamma((k-1)/2)} \left\{ (1 - x^2)^+ \right\}^{(k-3)/2}$$

(13)

denote the probability density function of $U_1^{(k)}$, where $U^{(k)} = (U_1^{(k)}, \ldots, U_k^{(k)})^t$ is uniformly distributed on $S^{k-1}$; and let

$$F_{k-1}(x) = \int_{-\infty}^{x} f_{k-1}(t) dt.$$  

(14)

Note that $F_k(x) = \Pr\{ t_k \leq \sqrt{kx}/\sqrt{(1 - x^2)} \} (|x| < 1)$, where $t_k$ has Student’s distribution with $k$ degrees of freedom.

Theorem 1. Let $U$ be uniform on $S^{m-1}$. For arbitrary $\theta_0 < \theta_1$ and $-1 < z < w < 1$ the inequality (10) holds with
\[ \text{pr}\{ (\gamma(\theta), U) = w \mid (\gamma(\theta_0), U) = z \} = f_{m-2} \left( \frac{w - \rho(\theta)z}{\sqrt{[(1 - z^2)(1 - \rho^2(\theta))]}} \right) / \sqrt{[(1 - z^2)(1 - \rho^2(\theta))]} \]

and

\[ E\{ (\gamma(\theta), U)^+ \mid (\gamma(\theta_0), U) = z, (\gamma(\theta), U) = w \} = \mu F_{m-3}(\mu/\tau) + (m - 2)^{-1}\tau f_{m-1}(\mu/\tau). \]

Here \( \mu, \tau, f_k \) and \( F_k \) are defined by (11) – (14).

**Remarks.** (i) The derivation of Theorem 1 is discussed in Appendix 1.

(ii) As discussed by Johnstone and Siegmund (1989) in the context of testing \( \beta = 0 \), there is often equality in (10). In particular there is equality whenever the tube radius, \( \cos^{-1}(w) \), is so small that the tube has no self-overlap. Moreover, the numerical examples of Knowles and Siegmund (1989) in the testing context as well as those in the following section indicate the inequality (10) provides a very good approximation in many problems where exact equality does not hold.

(iii) In order to approximate the probability in (9), the inequality (10) must be used four times: for \( \theta_1 > \theta_0 \) and \( \theta_1 < \theta_0 \), for \( \gamma \) and for \( -\gamma \). The calculation for \( -\gamma \) is most easily effected by changing \( z \) to \(-z\) and leaving \( \gamma \) unchanged.

We turn now to the problem of finding a confidence region for the pair \((\theta, \beta)\). The log likelihood ratio statistic for testing \( H_0 : \theta = \theta_0, \beta = \beta_0 \) is

\[ \Lambda(\theta_0, \beta_0) = -\frac{1}{2} m \log \left( \frac{\hat{\sigma}^2}{\sigma^2_0} \right), \]

where \( \hat{\sigma}^2 \) is defined in (6) and

\[ \hat{\sigma}^2 = m^{-1} \| y - \beta_0 f(\theta_0) \|^2. \]
The sufficient statistic under $H_0$ is $\|y - \beta_0 f(\theta_0)\|$, so our confidence region is the set of all $(\theta_0, \beta_0)$ such that

$$\text{pr}_{\theta_0, \beta_0} \left[ \Lambda(\theta_0, \beta_0) > \{\Lambda(\theta_0, \beta_0)\}_{\text{obs}} \|y - \beta_0 f(\theta_0)\| \right] > \alpha. \tag{17}$$

The subscripts $\theta_0, \beta_0$ are a reminder that the probability may depend on the hypothesized value of $(\theta, \beta)$, but by sufficiency it does not depend on the unknown nuisance parameter $\sigma^2$.

To evaluate the probability on the left hand side of (17), observe that

$$\Lambda(\theta_0, \beta_0) = -\frac{1}{2} m \log \left( \min_{\theta} \hat{\sigma}_0^2 / \hat{\sigma}_0^2 \right) - \frac{1}{2} m \log \left( \hat{\sigma}_0^2 / \hat{\sigma}_0^2 \right)$$

$$= \Lambda(\theta_0) + \Lambda'(\theta_0, \beta_0),$$

say, where $\Lambda(\theta_0)$ is defined in (7). Hence for any $b > 0$

$$\text{pr}_{\theta_0, \beta_0} \left\{ \Lambda(\theta_0, \beta_0) > b \|y - \beta_0 f(\theta_0)\| \right\} = \text{pr}_{\theta_0, \beta_0} \left\{ \Lambda'(\theta_0, \beta_0) > b \|y - \beta_0 f(\theta_0)\| \right\}$$

$$+ E_{\theta_0, \beta_0} \left[ \text{pr}_{\theta_0, \beta_0} \left\{ \Lambda(\theta_0) > b - \Lambda'(\theta_0, \beta_0) \|y - \beta_0 f(\theta_0)\|, \langle f(\theta_0), y \rangle \right\} \right.$$}

$$\times \left( \Lambda'(\theta_0, \beta_0) < b \right) \|y - \beta_0 f(\theta_0)\|]. \tag{18}$$

The first term on the right hand side of (18) is just the significance level of the likelihood ratio test of $\theta = \theta_0, \beta = \beta_0$ against $\theta = \hat{\theta}_0, \beta$ arbitrary, and is easily evaluated. Since the inner conditioning in the second term involves a statistic equivalent to $\|y\|, \langle f(\theta_0), y \rangle$, we can bound the conditional probability by means of Theorem 1.

To carry out these evaluations put $c^2 = 1 - \exp(-2b/m)$. Since

$$\|y - \hat{\beta}_0 f(\theta_0)\|^2 = \|y - \beta_0 f(\theta_0)\|^2 - \langle \gamma(\theta_0), y - \beta_0 f(\theta_0) \rangle^2,$$

we find by simple algebra that

$$\text{pr}_{\theta_0, \beta_0} \left\{ \Lambda'(\theta_0, \beta_0) > b \|y - \beta_0 f(\theta_0)\| \right\}$$

$$= \text{pr}_{\theta_0, \beta_0} \left\{ (\gamma(\theta_0), y - \beta_0 f(\theta_0))^2 / \|y - \beta_0 f(\theta_0)\|^2 > c^2 \|y - \beta_0 f(\theta_0)\| \right\}$$

$$= \text{pr} \left\{ |U_1^{(m)}| > c \right\}, \tag{19}$$

where, as above, $U_1^{(m)}$ denotes the first coordinate of a point $U^{(m)}$ which is uniformly distributed on $S^{m-1}$.  

6
Now consider the final term in (18). For the conditional probability we can condition on $\|y - \beta_0 f(\theta_0)\|$ and $\langle \gamma(\theta_0), y - \beta_0 f(\theta_0) \rangle / \|y - \beta_0 f(\theta_0)\|$. It will be convenient to consider specific values, say $\|y - \beta_0 f(\theta_0)\| = \eta$ and $\langle \gamma(\theta_0), y - \beta f(\theta_0) \rangle / \|y - \beta_0 f(\theta_0)\| = \xi$. The reasoning leading to (19) shows that taking the outer expectation means integrating with respect to $\Pr\{U_1^{(m)} \in d\xi\} = f_{m-1}(\xi)d\xi$, cf. (13), over the range $|\xi| \leq c$.

Simple algebra shows that conditionally

$$\Lambda(\theta_0) > b - \Lambda'(\theta_0, \beta_0)$$

if and only if

$$\max_\theta \langle \gamma(\theta), y / \|y\| \rangle^2 > \frac{c^2 + 2\delta_0 \xi \eta^{-1} + \delta_0^2 \eta^{-2}}{1 + 2\delta_0 \xi \eta^{-1} + \delta_0^2 \eta^{-2}},$$

where $\delta_0 = \beta_0 \|f(\theta_0)\|$. Hence by sufficiency the second term on the right hand side of (18) equals

$$\int_{|\xi| < c} \Pr\left\{ \max_\theta \langle \gamma(\theta), U \rangle^2 > \frac{c^2 + 2\delta_0 \xi \eta^{-1} + \delta_0^2 \eta^{-2}}{1 + 2\delta_0 \xi \eta^{-1} + \delta_0^2 \eta^{-2}} \langle \gamma(\theta_0), U \rangle \right\} \frac{\xi + \delta_0 \eta^{-1}}{(1 + 2\delta_0 \xi \eta^{-1} + \delta_0^2 \eta^{-2})^{1/2}} f_{m-1}(\xi)d\xi. \quad (20)$$

This calculation is summarized as

**Theorem 2.** Let $\Lambda(\theta_0, \beta_0)$ be defined by (16). Let $b > 0$ and $c^2 = 1 - \exp(-2b/m)$. For each $\eta > 0$

$$\Pr_{\theta_0, \beta_0}\left\{ \Lambda(\theta_0, \beta_0) > b \|y - \beta_0 f(\theta_0)\| = \eta \right\} \quad (21)$$

equals the sum of (19) and (20), where $\delta_0 = \beta_0 \|f(\theta_0)\|$. With the help of Theorem 1 one can obtain an upper bound for the conditional probability in (20) and hence for (21).

3. A Class of Artificial Examples.

In this section we discuss a class of examples devised for its geometric comprehensibility and computational simplicity. Cases of both large and small curvature are included.
We compare the confidence regions described in the preceding section with two competing suggestions. The first of these is inversion of the likelihood ratio test, with the significance level evaluated by assuming that an $F$-like version of the likelihood ratio statistic has an $F$ distribution. For example, to obtain a confidence region for $\theta$ with $\beta$ and $\sigma^2$ as nuisance parameters, consider

$$\frac{\| y - \eta(\theta_0, \hat{\beta}) \|^2 - \| y - \eta(\hat{\theta}, \hat{\beta}) \|^2}{\| y - \eta(\hat{\theta}, \hat{\beta}) \|^2}. \quad (22)$$

This is a version of the likelihood ratio statistic (7) for testing $\theta = \theta_0$, and if $\eta$ were linear in $\theta$ and $\beta$ its null distribution would be $(m - 2)^{-1} F_{1, m-2}$. Hence the set of all $\theta_0$ for which (22) is less than the $(1 - \alpha)^{th}$ quantile of the distribution of $(m - 2)^{-1} F_{1, m-2}$ is approximately a $(1 - \alpha)100\%$ confidence region provided the curvature of $\eta$ is not too great.

The second competing confidence region is that suggested by Halperin (1963), which is exact for the models considered in this paper. To find a confidence region for $\theta$ in the model (5), let $g = \hat{f} - \langle \hat{f}, f \rangle f / \| f \|^2$ and observe that

$$\frac{\langle g(\theta_0), y \rangle^2 / \| g(\theta_0) \|^2}{\| y - \langle g(\theta_0), y \rangle g(\theta_0) / \| g(\theta_0) \|^2 - \langle f(\theta_0), y \rangle f(\theta_0) / \| f(\theta_0) \|^2 \|^2} \quad (23)$$

is distributed as $(m - 2)^{-1} F_{1, m-2}$ when $\theta = \theta_0$. Hence an exact $(1 - \alpha)100\%$ confidence region is the set of all $\theta_0$ for which (23) is less than the $(1 - \alpha)^{th}$ quantile of $(m-2)^{-1} F_{1, m-2}$.

The geometric interpretation of (23) for testing $\theta = \theta_0$ is very simple: $\theta_0$ is rejected if the squared length of the projection of the residual $y - \hat{\beta}_{\theta_0} f(\theta_0)$ along the component of the tangent to $f$ at $\theta_0$ which is orthogonal to $f(\theta_0)$ is long compared to the sum of squared lengths of its projections in directions orthogonal to both $f(\theta_0)$ and $\hat{f}(\theta_0)$. Such a procedure is reasonable if $f$ is well-approximated by its tangent line in a sufficiently large neighborhood of $\theta_0$.

For the rest of this section we assume

$$\gamma(\theta) = f(\theta) = \begin{pmatrix} \sqrt{(1 - h^2)} \cos \{ \theta / \sqrt{(1 - h^2)} \}, \sqrt{(1 - h^2)} \sin \{ \theta / \sqrt{(1 - h^2)} \}, h, 0, \ldots, 0 \end{pmatrix}^t. \quad (24)$$

where $0 \leq h < 1$ is known and $|\theta| < \pi \sqrt{(1 - h^2)}$. Obviously $\gamma$ is a circle of circumference $2\pi \sqrt{(1 - h^2)}$ defined by $\{ x : x \in S^{m-1}, x_1^2 + x_2^2 = 1 - h^2, x_3 = h \}$. If $h = 0$ the model (24)
is trivially linear in the parameters $\mu_1 = \beta \cos \theta$ and $\mu_2 = \beta \sin \theta$. In this case the three methods described above all give the same confidence region. Cases of large curvature occur when $h$ is close to one, so $\gamma$ itself has large curvature, or when $h > 0$ and $\beta$ is close to zero, so $\theta$ is almost not identifiable. For future reference it will be helpful to record the following elementary geometric observations.

Obviously $\gamma$ is a unit speed curve, i.e., $\|\dot{\gamma}\| = 1$. The Frenet-Serret equation for $\dot{\gamma}$, when $\gamma$ is considered a curve in $S^{m-1}$, is (cf. Do Carmo, p. 261)

$$\ddot{\gamma} = k_g N_1 - \gamma,$$  \hspace{1cm} (25)

where $N_1$ is a unit normal to $\dot{\gamma}$ which lies in the tangent space of $S^{m-1}$ and $k_g$ is the geodesic curvature of $\gamma$. Simple calculation shows that $k_g = h/\sqrt{(1 - h^2)}$ and

$$N_1 = (-h \cos \{\theta/\sqrt{(1 - h^2)}\}, -h \sin \{\theta/\sqrt{(1 - h^2)}\}, \sqrt{(1 - h^2)}, 0, \ldots, 0)^t.$$  \hspace{1cm} (26)

To evaluate one of the confidence regions defined above, for example the region given by (9), we begin with the observed value of $\max_\theta |\langle \gamma(\theta), U \rangle| = w$, where $U = y/\|y\|$ as before. Then in principle we choose a trial value of $\theta_0$, which gives the observed value of $\langle \gamma(\theta_0), U \rangle = z$ and evaluate the probability (9) to see if $\theta_0$ belongs to the confidence region. For the calculations given below for the model (24) it is often technically more convenient first to choose a trial value of $z$, evaluate (9), and then solve for $\theta_0$. The key relation involving $U$, $z$, and $\theta_0$ is as follows. Assume $\langle \gamma(\hat{\theta}), U \rangle = w > 0$. Since $U \in S^{m-1}$ and $\langle \gamma(\hat{\theta}), U \rangle = 0$ we can write

$$U = w\gamma(\hat{\theta}) + \sqrt{(1 - w^2)} \sum_{i=1}^{m-2} \xi_i N_i,$$  \hspace{1cm} (27)

where the $N_i$ are mutually orthogonal unit vectors in the tangent space of $S^{m-1}$ at $\gamma(\hat{\theta})$ and are orthogonal to $\dot{\gamma}(\hat{\theta})$. Since $\gamma(\theta_0)$ is in the subspace spanned by $\gamma(\hat{\theta})$, $\dot{\gamma}(\hat{\theta})$, and $N_1$ given in (26), its inner product with $N_k$ vanishes for all $2 \leq k \leq m - 2$. Hence by (26) and (27)

$$z = \langle \gamma(\theta_0), U \rangle = w \langle \gamma(\theta_0), \gamma(\hat{\theta}) \rangle + \xi_1 \sqrt{(1 - w^2)} \langle \gamma(\theta_0), N_1 \rangle$$

$$= w \left[ h^2 + (1 - h^2) \cos \left\{ (\hat{\theta} - \theta_0)/\sqrt{(1 - h^2)} \right\} \right]$$

$$+ \xi_1 h \sqrt{\left\{ (1 - w^2)(1 - h^2) \right\} \left[ 1 - \cos \left\{ (\hat{\theta} - \theta_0)/\sqrt{(1 - h^2)} \right\} \right]},$$
\[
\hat{\theta} - \theta_0 = \sqrt{(1 - h^2)} \cos^{-1} \left[ \frac{z - wh^2 - \xi_1 h \sqrt{\{(1 - h^2)(1 - w^2)\}}}{w(1 - h^2) - \xi_1 h \sqrt{\{(1 - h^2)(1 - w^2)\}}} \right].
\]

(28)

From (28) we see that the confidence region depends not only on the minimal geodesic distance from $\gamma$ or $-\gamma$ to $U$, but also on the direction. A large positive value of $\xi_1$ indicates that $U$ is in the direction of the north pole $x_3 = 1, x_1 = x_2 = x_4 = \cdots = x_m = 0$ from $\gamma$. In this case one rightfully expects the confidence region to be larger than if $U$ were an equal distance from $\gamma$ in the direction of the equator.

Table 1 contains some numerical examples of confidence regions for $\theta$ in the model (24). The columns headed (9), (22), and (23) give the sizes of 95% confidence regions as a percentage of the size of the parameter space, $2\pi \sqrt{(1 - h^2)}$, for the three methods defined above in the correspondingly numbered displays. In most cases $\xi_1$ is taken to have its mean value of zero, but in some cases it is $\pm 0.2$. Since $\sum \xi_k^2 = 1$, these cases correspond to roughly a one standard deviation departure from the mean. The first row, where $h = 0.01$, is included to illustrate agreement of the three methods when $h = 0$. Since evaluation of the right hand side of (10) by the method given in Appendix 1 leads to indeterminate expressions when $h = 0$, the value $h = 0.01$ was used instead. The column headed “Prob” gives our approximation for the probability of rejecting $\theta_0$, for $\theta_0$ at a boundary of the confidence region. The column headed MC contains an asterisk when it can be shown that our approximation of (9) via (10) is exact, and otherwise contains a 99999 repetition Monte Carlo estimate in order to demonstrate that the approximation is usually quite good even when it is not exact. In the last row, where the confidence set (9) is the entire parameter space, the parenthetical entry in the column headed Prob gives the minimum value attained by our approximation to (9), and the entry under MC shows that the approximation is not unnecessarily conservative.

| Table 1 About Here |

The procedure (23) is quite volatile and includes the entire parameter set in its confidence region in some cases which the other methods do not find problematic. Bates and Watts (1988, pp. 223-9) have made similar observations and recommend that this method
not be used in practice. The method (22) appears to do surprisingly well. Following the formulation of Beale (1960), Hamilton and Wiens (1987) provide a justification for the assumed distribution of (22) by showing that if $\sigma$ is small (22) is essentially the same as (23). In a linear model these two are in fact identical. However, the differences in behavior they exhibit here makes us feel uncomfortable with this justification. A quite different comparison of (10) and (22) is given in Appendix 2.

Table 2 and Figures 2, 3, and 4 are concerned with joint confidence regions for $(\theta, \beta)$. In addition to the region defined in (17), with the required probability calculations carried out as indicated in Theorems 1 and 2, we also consider regions analogous to (22) and (23). The analogue of (22) is obtained by assuming that

$$\frac{||y - \eta(\theta_0, \beta_0)||^2 - ||y - \eta(\hat{\theta}, \hat{\beta})||^2}{||y - \eta(\hat{\theta}, \hat{\beta})||^2}$$  \hspace{1cm} (29)

has approximately the distribution of $2(m - 2)^{-1}F_{2,m-2}$ when $\theta = \theta_0$, $\beta = \beta_0$. The analogue of (23) is based on the observation that

$$\frac{\langle f(\theta_0), y - \beta_0 f(\theta_0) \rangle^2/\|f(\theta_0)\|^2 + \langle g(\theta_0), y \rangle^2/\|g(\theta_0)\|^2}{\|y\|^2 - \langle g(\theta_0), y \rangle^2/\|g(\theta_0)\|^2 - \langle f(\theta_0), y \rangle^2/\|f(\theta_0)\|^2}$$  \hspace{1cm} (30)

has exactly the distribution of $2(m - 2)^{-1}F_{2,m-2}$ when $\theta = \theta_0$, $\beta = \beta_0$.

Figure 2 involves relatively small curvature. The regions are approximately elliptical as one expects by analogy with the linear case, and the three regions all agree. Figure 3 involves larger curvature. The regions defined by (17) and (29) show very good, although no longer perfect agreement, but the butterfly shaped region defined by (30) has responded poorly to the stress of moderate curvature. Figure 4 involves still larger curvature, and we have displayed only the regions of (17) and (29). Although the approximate $F$ region defined by (29) is somewhat smaller than the region defined by (17), the agreement is surprisingly good.

We have performed a Monte Carlo experiment which verified that the confidence regions defined by (17) have the correct coverage probability for the examples of Table 2. The details are omitted.
Remark. For the convenience of not having to consider end effects, we have allowed $\theta$ in (24) to have the full range of values, $|\theta| \leq \pi \sqrt{1 - h^2}$, so the curve $\gamma$ is a circle. In some cases this assumption introduces peculiarities which probably do not occur often and which we therefore have not discussed in detail. For $h$ close to 0 the curves $\gamma$ and $-\gamma$ are close to each other. Indeed, in the limiting case $h = 0$ they coincide, and to have a unique parametrization of our model we must either assume that $\beta$ is positive or restrict the range of $\theta$, so that $\gamma(\theta)$ is a semicircle. When $h$ is small, because of the proximity of $-\gamma(\theta)$ to $\gamma\{\theta + \pi \sqrt{1 - h^2}\}$, the confidence region defined by (9) may have two components, one centered at $\hat{\theta}$ and the other at $\hat{\theta} + \pi \sqrt{1 - h^2}$. This is also true of the regions defined by (22) and (23). In fact the entry under (9) in the next to last row of Table 1 and that under (22) in the last row involve confidence regions consisting of the union of two intervals. The exact joint confidence region defined by (30) exhibits similar behavior even when $h$ is not close to 0. For the generally well-behaved example of Figure 2, the region determined by (30) actually has a second, not displayed component containing values of $\beta$ close to zero and $\theta$ close to the extremities of the parameter space at $\pm 2.25$. The regions determined by (17) and (29) do not seem to exhibit this kind of behavior, at least not in cases of small or moderate curvature. In spite of its artificiality, our model (24) with $h = 0$ and $\theta$ restricted to make $\gamma$ a semicircle differs only slightly from that of the Fieller-Creasy problem, viewed as a problem of nonlinear regression (cf. Cook and Goldberg, 1986). There too the curve $\gamma$ is a great semicircle, so the three methods (9), (22), and (23) all give the same confidence region for the ratio of normal means with common unknown variance.

It is interesting to see how the diagnostics suggested by Bates and Watts (1980) apply to these examples. Since our procedures are all invariant with respect to re-parameterization, we consider only Bates’ and Watts’ “intrinsic curvature” which is defined at each $(\theta_0, \beta_0)$ to be the maximum normal curvature of the surface $\eta(\theta, \beta) = \beta \gamma(\theta)$. Suppose $\alpha(t) = \beta(t) \gamma\{\theta(t)\}$ is a curve in the surface with $\alpha(0) = \beta_0 \gamma(\theta_0)$. Denoting differentiation with respect to $t$ by a prime, we have $\alpha' = \beta' \gamma(\theta) + \beta \theta' \dot{\gamma}(\theta)$ and $\alpha'' = \beta'' \gamma(\theta) + (2 \beta' \theta' + \beta \theta'') \dot{\gamma}(\theta) + \beta \theta' \ddot{\gamma}(\theta)$. By (25) the component of $\alpha''$ normal to the tangent space spanned by $\partial \eta / \partial \theta = \beta \dot{\gamma}(\theta)$ and $\partial \eta / \partial \beta = \gamma(\theta)$ is $\beta \theta' \ddot{\gamma}(\theta)$, with norm
\[ |\beta'(\theta)^2 k_g| \] This is maximized subject to \( 1 = \| \alpha' \|^2 = (\beta')^2 + (\beta\theta')^2 \) by setting \( \beta' = 0 \), so \( \theta' = 1/\beta \). Hence the maximizing value at \((\theta_0, \beta_0)\) is \( |k_g(\theta_0)/\beta_0| \). For a \((1 - \alpha)100\%\) confidence region the Bates-Watts diagnostic is

\[
\{ |k_g(\hat{\theta})|/\hat{\beta} \} \hat{\sigma} \sqrt{2F_{2, m-2}(1 - \alpha)} \tag{31}
\]

where \( \hat{\sigma} \) is a suitable estimator of \( \sigma \) and \( F_{2, m-2}(p) \) denotes the \( p \)th percentile of the appropriate \( F \) distribution. For the model (24) and \( \hat{\sigma}^2 = (m - 2)^{-1} \| y - \eta(\hat{\theta}, \hat{\beta}) \|^2 \) (31) reduces to

\[
\frac{h \sqrt{(1 - \max(\gamma(\theta), U)}^2 \sqrt{(1 - h^2) \max(\| \gamma(\theta), U)\}}}{\sqrt{2F_{2, m-2}(1 - \alpha)/(m - 2)}} \tag{32}
\]

This quantity should be small compared to unity. Bates and Watts (1980) suggest that confidence regions based on (29) should be adequate if (32) is less than 1/2. In fact, for the first row in Table 2 and \( \alpha = .05 \) it is .36, while it is about .97 for the second row. For the third row it is about 1.2. In these examples the recommended cutoff of Bates and Watts seems more conservative than necessary.

Unfortunately excellent agreement between the approximate region defined by (29) and the exact region defined by (30) cannot be assumed to follow automatically from a satisfactory Bates-Watts diagnostic. Although the diagnostic does not depend on the value of \( \xi_1 \), as in the one parameter case the exact region defined by (30) is sensitive to changes in \( \xi_1 \), while the regions determined by (17) and (29) are much less so. For example, for the data in the first row of Table 2, for \( |\xi_1| \leq 0.5 \), which represents variation of about two standard deviations, and for \( \beta \) about equal to \( \hat{\beta} \), the confidence regions of (17) and (29) change by about 10 – 15\% from the region in Figure 1 and are always essentially identical. Changes in the region defined by (30) are in the range of 25 – 50\%, so that this region is sometimes smaller and sometimes much larger than the other two.

A similar diagnostic for the problem of finding a confidence region for \( \theta \) has been suggested by Cook and Goldberg (1986). However, their definition does not seem completely satisfactory since it assigns a positive, possibly large curvature to the model (24) with \( h = 0 \) and to the model of the Fieller-Creasy problem, for which the Bates-Watts curvature is zero and where the regions defined by (9), (22), and (23) are all identical.
4. Other Examples.

In this section we consider some examples from the literature of nonlinear regression and compare confidence regions obtained by our method with those obtained by others. For simplicity we consider only the one dimensional problem of confidence regions for \( \theta \), so our method is defined by (9) and (10).

Our data and models are borrowed from Bates and Watts (1988). They are the Treated and Untreated Puromycin data of Appendix 1 Table A1.3, fit by the Michaelis-Menten model (2), and the BOD data of Appendix 1 Table A1.4, fit by the asymptotic growth model (3). Bates and Watts discuss the Treated Puromycin data and the BOD data extensively, the first as an example involving small curvature, where linear approximations work well, and the second as an example of large curvature, where they do not. Hamilton (1986) discusses the Untreated Puromycin data.

Table 3 compares the confidence regions obtained from (9) and (10) with the approximate \( F \) regions defined by (22), which were calculated by the authors cited. The column headed \( \alpha \) gives for \( \theta_0 \) equal to the lower and upper endpoints, respectively, of the approximate \( F \) regions the conditional \( p \)-values of the likelihood ratio test evaluated by the prescription of Theorem 1.
These examples are consistent with those in Section 3. In each case the approximate region defined by (22) is smaller than our region. The discrepancy is small or large according as the curvature is small or large. In Section 3 it was a simple matter to demonstrate by analysis or simulation that our regions were essentially exact, and we believe that to be the case here as well.

In the first two rows of Table 3, which involve small curvature, the conditional likelihood ratio region defined by (9) and (10) is about 11 - 13% longer than the approximate F region defined by (22). For the second row Hamilton (1986) also gives the exact confidence region defined by (23). It turns out to be (0.0293, 0.0735), which is much closer to our interval than that defined by (22).

Although our 95% region for the BOD data is unbounded to the right, it would be bounded for all confidence levels no larger than 94%. Our 90% region is (0.1872, 1.519), which is contained within Bates and Watts's approximate 95% region.

5. Discussion.

In this paper we have obtained approximate conditional likelihood ratio confidence regions for $\theta$ and for $(\theta, \beta)$ in the simple model (5). In cases of small to moderately large curvature the regions appear to be essentially exact; in general they are conservative. In cases of extremely large curvature they may in principle be quite conservative, although this possibility was not realized in any of those cases we checked by simulation.

In simple examples we have compared our region with two others which have been frequently studied. The exact region of Halperin (1963) is quite volatile, even in cases of moderate curvature. It probably should not be used, except perhaps as a diagnostic to indicate those cases where one can feel confident that the coverage probability of some approximate region actually is what it is supposed to be. In our examples the approximate F region is usually slightly anti-conservative and becomes more so when curvature is large. Generally speaking its performance appears to be better than previous studies explain theoretically, since standard asymptotic analyses (e.g., Johansen, 1984, Chapter 5) in effect show that for small $\sigma$ this procedure and the poorly behaved Halperin region
are closely related. An alternative asymptotic analysis of our regions, which shows their relation to and explains the usually slight anti-conservatism of the approximate \( F \) regions, is given in Appendix 2.

Rather different asymptotic calculations of Cox and McCullagh (1986) suggest that when there are at least three parameters there are cases when the approximate \( F \) regions are conservative. See also Hamilton and Wiens (1987). It would be interesting to make a systematic comparison of these different asymptotic analyses.

Possible future projects are more systematic development of the method of this paper for important special cases of the model (1) and extension of the method to deal with higher dimensional problems. The broken line regression model (4) is one important special case in which the methods of this paper are clearly applicable. Hinkley (1971) has discussed standard large sample theory. Extension of our method to higher dimensional cases requires substantially more numerical integration. In the long run there may be some advantage to developing approximations to the required probabilities. See Appendix 2.

In a more speculative vein, it may also be possible to use the procedure of this paper as a diagnostic, say by setting all nonlinear parameters except one equal to their maximum likelihood values and then comparing a confidence region for the remaining nonlinear parameter with that obtained by the approximate \( F \) method. Alternatively, if all but one of the nonlinear parameters in a complex model seem to vary more or less linearly, one might approximate the given model by one which is linear in all parameters but the problematic one and apply the methods of this paper to the approximating model.
Appendix 1
Mathematical Derivations

Here we give details for two of the arguments used in Section 2. The first is a slight generalization of Basu’s Theorem (Lehmann, 1986, p. 191). The second is the explicit evaluation of the right hand side of (10) given in Theorem 1.

**Proposition.** If $S$ is sufficient and boundedly complete for the family of distributions $P_{\theta}$, $\theta \in \Theta$, and if $(T_1, T_2)$ is ancillary, then for any $\theta_0 \in \Theta$

$$P_{\theta_0}(T_1 \in A|T_2, S) = P_{\theta_0}(T_1 \in A|T_2) \quad \text{a.e.} \ P_{\theta}, \theta \in \Theta.$$

**Proof.** It suffices to show

$$E_{\theta}[E_{\theta_0}\{f_1(T_1)|T_2, S\}f_2(T_2)f_3(S)] = E_{\theta}[E_{\theta_0}\{f_1(T_1)|T_2\}f_2(T_2)f_3(S)] \quad (A.1)$$

for all $\theta$ and all bounded functions $f_1$, $f_2$, $f_3$. By two applications of Basu’s Theorem, we see that the left hand side of (A.1) equals

$$E_{\theta}[E_{\theta_0}\{f_1(T_1)f_2(T_2)|S\}f_3(S)] = E_{\theta_0}\{f_1(T_1)f_2(T_2)\}E_{\theta}\{f_3(S)\}
= E_{\theta_0}[E_{\theta_0}\{f_1(T_1)|T_2\}f_2(T_2)]E_{\theta}\{f_3(S)\}
= E_{\theta}[E_{\theta_0}\{f_1(T_1)|T_2\}f_2(T_2)f_3(S)],$$

which completes the proof.

In order to evaluate the integral on the right hand side of (10) we use repeatedly the facts that if $U^{(k)} = (U_1^{(k)}, \ldots, U_k^{(k)})^t$ is distributed uniformly on $S^{k-1}$, then (i) $U_1^{(k)}$ has the probability density function $f_{k-1}$ given in (13), and (ii) conditional on $U_1^{(k)}, \ldots, U_j^{(k)}$ the random variable $(U_{j+1}^{(k)}, \ldots, U_k^{(k)})^t$ is uniformly distributed on a $k-j-1$ dimensional sphere of radius $\{1 - \sum_j^k U_i^{(k)2}\}^{1/2}$.

Let $\rho(\theta) = \langle \gamma(\theta), \gamma(\theta_0) \rangle$, $e_1 = \gamma(\theta_0)$, and $e_2 = \{\gamma(\theta) - \rho(\theta)\gamma(\theta_0)\}/\{1 - \rho^2(\theta)\}^{1/2}$, so $e_1$ and $e_2$ are orthonormal and

$$\gamma(\theta) = \rho(\theta)e_1 + \sqrt{1 - \rho^2(\theta)}e_2.$$
Hence

$$\langle \gamma(\theta), U \rangle = \rho(\theta)U_1 + \sqrt{1 - \rho^2(\theta)} U_2,$$

where $U_i = \langle e_i, U \rangle$ ($i = 1, 2$). Given $U_1 = z$ the conditional density of $U_2$ is

$$f_{m-2}[x/\sqrt{(1 - z^2)}/\sqrt{(1 - z^2)}]$$

and hence

$$\text{pr}\{\langle \gamma(\theta), U \rangle = w | \langle \gamma(\theta_0), U \rangle = z\} = \text{pr}[\rho(\theta)z + \sqrt{1 - \rho^2(\theta)} U_2 = w | U_1 = z]$$

is given by (15).

It is easy to verify that

$$\dot{\gamma}(\theta) = \dot{\rho}e_1 - \left\{\rho\dot{\rho}/\sqrt{(1 - \rho^2)}\right\}e_2 + \sqrt{\|\dot{\gamma}\|^2 - \dot{\rho}^2/(1 - \rho^2)}e_3,$$

where $e_3 = \left\{\dot{\gamma} - \dot{\rho}e_1 + \rho\dot{\rho}e_2/\sqrt{(1 - \rho^2)}\right\}/\sqrt{\|\dot{\gamma}\|^2 - \dot{\rho}^2/(1 - \rho^2)}$ is a unit vector orthogonal to both $e_1$ and $e_2$. Hence given $\langle \gamma(\theta_0), U \rangle = z$, $\langle \gamma(\theta), U \rangle = w$ we have

$$\langle \dot{\gamma}(\theta), U \rangle = \dot{\rho}U_1 - \left\{\rho\dot{\rho}/\sqrt{(1 - \rho^2)}\right\}U_2 + \sqrt{\|\dot{\gamma}\|^2 - \dot{\rho}^2/(1 - \rho^2)}U_3$$

$$= \dot{\rho}(z - \rho w)/\sqrt{(1 - \rho^2)} + \sqrt{\|\dot{\gamma}\|^2 - \dot{\rho}^2/(1 - \rho^2)}U_3,$$

where $U_3 = \langle e_3, U \rangle$ is the first coordinate of a point uniformly distributed on an $m-3$ dimensional sphere of radius

$$\sqrt{1 - (w^2 + z^2 - 2\rho zw)/(1 - \rho^2)}.$$

Hence the conditional density of $\langle \dot{\gamma}(\theta), U \rangle$ is given by

$$\tau^{-1}f_{m-3}\{(x - \mu)/\tau\},$$

where $\mu$ and $\tau$ are given in (11) and (12). The evaluation of $E\{\langle \dot{\gamma}(\theta), U \rangle^+ | \langle \gamma(\theta_0), U \rangle = z, \langle \gamma(\theta), U \rangle = w\}$ indicated in Theorem 1 now follows by integration.
Appendix 2
Approximations to (10) and (20)

This appendix contains heuristic approximations to (10) and (20). Our goal is not to give alternative computational formulae, although good, simply evaluated approximations would be useful, especially if one were to extend the methods of this paper to higher dimensional problems. Rather we want to show the relation of (10) and (20) to the analogous $F$ distributions of linear regression theory and to explain the usually small but consistent anti-conservatism we have observed in the confidence regions defined by (22) and (29). Although we do not attempt a rigorous mathematical formulation of our results, it should be noted that they are tail approximations and in principle are quite unlike the approximations of Beale (1960) and others, which depend only on the local behavior of the regression function in a neighborhood of the true value of the unknown parameter and which to first order do not depend on the curvature of the regression surface.

We begin with the observation that the argument $u(\theta) = \{w - \rho(\theta)z\}/\sqrt{[(1 - z^2)(1 - \rho^2(\theta))]}$ of the second factor in the integrand of (10), as evaluated in Theorem 1, is positive and for large $w$ makes its principal contribution to the integral in a small neighborhood of the point $\theta^*$ where $u(\theta)$ is minimized.

It is easy to see by differentiation that $\theta^*$ satisfies

$$\rho(\theta^*) = z/w. \quad (A.2)$$

Typically there are two values of $\theta^*$, one less than $\theta_0$ and the other greater than $\theta_0$. For convenience we consider the value $\theta^* > \theta_0$. A two term Taylor series expansion and some algebra show that in a neighborhood $|\theta - \theta^*| < \delta$, where $\delta = \delta(m)$ converges to zero at a rate slightly slower than $m^{-1/2}$,

$$f_{m-2} \left( \frac{w - \rho(\theta)z}{\sqrt{[(1 - z^2)(1 - \rho^2(\theta))]} \right)$$

$$\sim C_{m-2} \left( \frac{1 - w^2}{1 - z^2} \right)^{(m-4)/2} \exp \left[ - (\theta - \theta^*)^2 \{m\rho^2(\theta^*)w^4/(1 - w^2)(w^2 - z^2)\}/2 \right], \quad (A.3)$$

where $C_{k-1} = \Gamma(k/2)/[\sqrt{\pi}\Gamma((k - 1)/2)]$ denotes the normalizing constant in (13).
For $\mu$ defined by (11), by using (A.2) we can easily see that $\mu(\theta^*) = 0$ and

$$\mu(\theta) \sim -(\theta - \theta^*) \hat{\rho}^2(\theta^*) w^3 / (w^2 - z^2)$$  \hspace{1cm} (A.4)

as $\theta \to \theta^*$.

It is notationally convenient to put $\hat{\rho}(\theta^*) = \hat{\rho}^*$ and $\tau(\theta^*) = \tau^*$, where $\tau(\theta)$ is defined in (12).

We assume that $\tau^* > 0$, although the possibility that $\tau^* = 0$ is also of interest. If the curve $\gamma$ were a geodesic, $\tau$ would be identically zero, and the right hand side of (10) would have to be interpreted as a limit. In that case the left hand side would exactly equal

$$\text{pr} \left[ U_1^{(m-1)} > \sqrt{\{(w^2 - z^2)/(1 - z^2)\}} \right] \sim \frac{C_{m-2}(1 - w^2)^{(m-2)/2}}{m(1 - z^2)^{(m-3)/2}\sqrt{(w^2 - z^2)}}$$  \hspace{1cm} (A.5)

as $m \to \infty$. For $\tau$ close to zero the right hand side of (15) is approximately $\max(\mu, 0)$; and using this approximation, (A.3), and (A.4) one can show that the right hand side of (10) asymptotically equals the right hand side of (A.5).

We now suppose $\tau^* > 0$ and observe that the functions defined in (13) and (14) satisfy

$$f_{m-1}(t/\sqrt{m}) \sim \sqrt{m} \varphi(t) \quad \text{and} \quad F_{m-3}(t/\sqrt{m}) \to \Phi(t)$$  \hspace{1cm} (A.6)

as $m \to \infty$, where $\varphi$ and $\Phi$ are the standard normal density and distribution functions respectively.

By means of (A.4) and (A.6) we can approximate the right hand side of (15) for $\theta$ close to $\theta^*$. Since the integral in (10) need only be evaluated in a neighborhood of $\theta^*$ of size slightly larger than $m^{-1/2}$, with the help of (A.3) we see after a change of variable that this integral

$$\sim \frac{C_{m-2}\tau^*(1 - w^2)^{(m-3)/2}}{m\hat{\rho}^* w(1 - z^2)^{(m-3)/2}} \int_{-\infty}^{\infty} \{(y/\Delta)\Phi(y/\Delta) + \varphi(y/\Delta)\} \exp(-y^2/2)dy,$$

as $m \to \infty$, where

$$\Delta = \{\tau^*/(w\hat{\rho}^*)\} \sqrt{\{(w^2 - z^2)/(1 - z^2)\}}.$$  \hspace{1cm} (A.7)

20
Evaluating this last integral, we obtain heuristically for any $\theta_1 > \theta^*$

$$\Pr\left\{ \max_{\theta_0 < \theta < \theta_1} \langle \gamma(\theta), U \rangle > w \mid \langle \gamma(\theta_0), U \rangle = z \right\} \sim \frac{C_{m-2}(1 - w^2)^{(m-2)/2}}{m(1 - z^2)^{(m-3)/2} \sqrt{(w^2 - z^2)}} \sqrt{(1 + \Delta^2)}. \quad (A.8)$$

Note that (A.8) exceeds the right hand side of (A.5) by precisely the factor $\sqrt{(1 + \Delta^2)}$.

Some additional manipulation of (A.7) shows that $\Delta$ in close to zero if $w$ is close to one, $z$ is close to $w$, or both. In these cases one should expect the essentially exact region defined by (9) and (10) the approximate region defined by (22) to agree.

It is now relatively easy to substitute (A.8) into (20) and integrate to obtain an analogous tail approximation to (18). We must initially exclude from the range of integration values of $|\xi|$ close to $c$, where in the obvious notation $z = z(\xi)$ is close to $w = w(\xi)$ and the approximation (A.8) is inappropriate because the probability is not small. However, the excluded region can be made negligibly small, and the effect is the same as if we merely substitute (A.8) into (20) and integrate. Simple algebra shows that

$$(w^2 - z^2)/(1 - z^2) = (c^2 - \xi^2)/(1 - \xi^2).$$

From the indicated substitution and integration we see that as $m \to \infty$ the expression (20)

$$\sim (1 - c^2)^{(m-2)/2} \int_{|\xi| < c} \sqrt{\left\{ (1 + \Delta^2)/(c^2 - \xi^2) \right\}} d\xi. \quad (A.9)$$

In this case, if we put $\Delta = 0$ the integral in (A.9) equals $\pi$ and we obtain not only asymptotically but exactly the relevant $F$ probability in the form

$$\Pr\{U_1^{(m-1)^2} + U_2^{(m-1)^2} > c^2\} = (1 - c^2)^{(m-2)/2}. $$
References


Table 1

Confidence Regions for $\theta$ in the Model (24)

<table>
<thead>
<tr>
<th>h</th>
<th>m</th>
<th>w</th>
<th>$\xi_1$</th>
<th>(9)</th>
<th>(22)</th>
<th>(23)</th>
<th>Prob.</th>
<th>MC</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>10</td>
<td>0.80</td>
<td>0.0</td>
<td>0.15</td>
<td>0.15</td>
<td>0.15</td>
<td>0.050</td>
<td>*</td>
</tr>
<tr>
<td>0.9</td>
<td>10</td>
<td>0.95</td>
<td>0.0</td>
<td>0.20</td>
<td>0.20</td>
<td>0.22</td>
<td>0.050</td>
<td>*</td>
</tr>
<tr>
<td>0.9</td>
<td>10</td>
<td>0.90</td>
<td>0.0</td>
<td>0.32</td>
<td>0.31</td>
<td>1.00</td>
<td>0.050</td>
<td>*</td>
</tr>
<tr>
<td>0.9</td>
<td>10</td>
<td>0.80</td>
<td>0.0</td>
<td>0.61</td>
<td>0.53</td>
<td>1.00</td>
<td>0.050</td>
<td>0.048</td>
</tr>
<tr>
<td>0.7</td>
<td>10</td>
<td>0.80</td>
<td>0.0</td>
<td>0.31</td>
<td>0.30</td>
<td>0.38</td>
<td>0.050</td>
<td>*</td>
</tr>
<tr>
<td>0.6</td>
<td>10</td>
<td>0.75</td>
<td>0.0</td>
<td>0.39</td>
<td>0.33</td>
<td>1.00</td>
<td>0.051</td>
<td>0.050</td>
</tr>
<tr>
<td>0.8</td>
<td>15</td>
<td>0.80</td>
<td>0.0</td>
<td>0.26</td>
<td>0.25</td>
<td>0.28</td>
<td>0.050</td>
<td>*</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>0.2</td>
<td>0.29</td>
<td>0.28</td>
<td>1.00</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>-0.2</td>
<td>0.23</td>
<td>0.23</td>
<td>0.21</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.7</td>
<td>0.0</td>
<td>0.28</td>
<td>0.24</td>
<td>0.25</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>0.2</td>
<td>0.30</td>
<td>0.26</td>
<td>0.32</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>&quot;</td>
<td>&quot;</td>
<td>&quot;</td>
<td>-0.2</td>
<td>0.26</td>
<td>0.23</td>
<td>0.21</td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.6</td>
<td>0.0</td>
<td>0.85</td>
<td>0.35</td>
<td>0.39</td>
<td>0.050</td>
<td>0.049</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.55</td>
<td>0.0</td>
<td>1.00</td>
<td>0.58</td>
<td>1.00</td>
<td>(0.088)</td>
<td>(0.091)</td>
</tr>
</tbody>
</table>
Table 2
Data Related to Figures 2, 3, and 4

<table>
<thead>
<tr>
<th>$h$</th>
<th>$2\pi \sqrt{1 - h^2}$</th>
<th>$m$</th>
<th>$|y|$</th>
<th>$\hat{\theta}$</th>
<th>$\langle \gamma(\hat{\theta}, U) \rangle$</th>
<th>$\hat{\beta}_{\hat{\theta}}$</th>
<th>$\xi_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7</td>
<td>4.49</td>
<td>15</td>
<td>5.0</td>
<td>0.0</td>
<td>0.90</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>0.8</td>
<td>3.77</td>
<td>11</td>
<td>5.0</td>
<td>0.0</td>
<td>0.80</td>
<td>4.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.8</td>
<td>3.77</td>
<td>10</td>
<td>5.0</td>
<td>0.0</td>
<td>0.76</td>
<td>3.8</td>
<td>0.0</td>
</tr>
</tbody>
</table>
Table 3
Selected Confidence Regions for $\theta$

<table>
<thead>
<tr>
<th>Data</th>
<th>Confidence</th>
<th>(9)</th>
<th>(22)</th>
<th>$\alpha$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puromycin $(T)$</td>
<td>99%</td>
<td>(0.0385, 0.1013)</td>
<td>(0.0408, 0.0972)</td>
<td>0.018, 0.016</td>
</tr>
<tr>
<td>Puromycin $(U)$</td>
<td>95%</td>
<td>(0.0295, 0.0733)</td>
<td>(0.0314, 0.0701)</td>
<td>0.079, 0.075</td>
</tr>
<tr>
<td>BOD</td>
<td>95%</td>
<td>(0.0989, $+\infty$)</td>
<td>(0.132, 1.77)</td>
<td>0.065, 0.084</td>
</tr>
</tbody>
</table>
Figure 1. Geometric Interpretation of (9)
Figure 2. Joint Confidence Region: Small Curvature
Figure 3. Joint Confidence Region: Moderate Curvature
Figure 4. Joint Confidence Region: Large Curvature
**CONFIDENCE REGIONS IN SEMILINEAR REGRESSION**

**AUTHOR(s)**
Mark Knowles and David Siegmund

**PERFORMING ORGANIZATION NAME AND ADDRESS**
Department of Statistics - Sequoia Hall
Stanford University
Stanford, California 94305-4065

**CONTROLLING OFFICE NAME AND ADDRESS**
Air Force Office of Scientific Research
Mathematical & Information Science Division
Bolling Air Force Base, Washington, DC 20332

**REPORT DATE**
January 1990

**NUMBER OF PAGES**
30pp.

**DISTRIBUTION STATEMENT (of this Report)**
Approved for public release; distribution unlimited.

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

**SUPPLEMENTARY NOTES**

**KEY WORDS (Continue on reverse side if necessary and identify by block number)**
Confidence region, nonlinear regression, upcrossings

**ABSTRACT (Continue on reverse side if necessary and identify by block number)**
For the simple semilinear regression model $y = \beta_f(\theta) + \epsilon$, we obtain conservative approximations to likelihood ratio confidence regions for $\theta$ and joint regions for $(\theta, \beta)$ which in many cases are essentially exact. To approximate the required probability we adapt the upcrossings approach of Johnstone and Siegmund (1989). In simple examples we compare our regions with two others which have been widely discussed in the literature.