Change-point Problems
IMS Lecture Notes – Monograph Series (Volume 23, 1994)

CONFIDENCE REGIONS IN BROKEN LINE REGRESSION

BY DAVID O. SIEGMUND AND HEPING ZHANG
Stanford University and Yale University

The broken line regression model is viewed as a special case of nonlinear regression. Following the methodology of Knowles, Siegmund, and Zhang (1991), we discuss procedures for constructing confidence regions. Our method involves inversion of the likelihood ratio test. A slightly conservative bound is obtained for the level of the test given the values of statistics which are sufficient for the nuisance parameters when the parameters of interest are fixed. We use a number of published data sets and simulations to compare our method with the approximate F method, which is based on the assumption that a formal analogue of the F statistic has approximately an F distribution, and with the Bayesian method of Smith and Cook (1980).

1. Introduction. The broken line regression model

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 (x_i - \theta)^+ + \varepsilon_i, \tag{1}$$

where $a^+ = \max(a, 0)$ and ε_i $(i = 1, \cdots, m)$ are independent $N(0, \sigma^2)$, has been discussed by a number of authors. Some of these consider the model (1) as a special case of nonlinear regression (e.g., Ratkowsky (1983) p. 122 ff., and Seber and Wild (1989) p. 447 ff.), while others have addressed it directly (e.g., Hinkley (1971), Feder (1975), Smith and Cook (1980)). An interesting special case considered by Hinkley (1971) is

$$y_i = \alpha - \beta (x_i - \theta)^- + \varepsilon_i \quad (i = 1, \cdots, m), \tag{2}$$

where $a^- = -(a - a^+)$. In (2) the parameters θ and α have natural interpretations: α is the maximum mean response produced by an input x, and θ is the minimum input required to produce this expected response.

Direct asymptotic analysis of (1) leads to technical difficulties, due to the lack of smoothness of the likelihood function (Hinkley (1971), Feder (1975)); and as we show by an example below, Hinkley's asymptotic version of the

AMS 1991 Subject Classification: Primary 62J02; Secondary 53A07

Key words and phrases: Approximate F method, broken line regression, conditional likelihood ratio, upcrossing.

likelihood ratio statistic can behave poorly when used in small samples. Smith and Cook (1980) gave a non-asymptotic, Bayesian treatment which relies on numerical integration to obtain exact posterior credible sets for the change-point, θ. Their method does not appear to have been used for problems of inference involving θ in conjunction with other parameters, although there is in principle no impediment, except the computational one, to such an adaptation.

In the literature which views the model (1) as a special case of nonlinear regression, one tries to avoid the lack of smoothness of the likelihood function by smoothing the corner of the regression function through use of an interpolating function governed by a smoothing parameter. See Seber and Wild (1989, p. 447 ff.) for a discussion of several versions of this technique. For a number of reasons this does not seem to be a completely satisfactory solution of the problem. Although one can argue on *a priori* grounds that a smooth regression function is more realistic, there are usually inadequate data near the change-point to make a definitive choice of the smoothing technique, and the smoothing parameter which the method introduces frequently does not have a useful physical interpretation. Also, this method only avoids the lack of smoothness in a certain formal sense. Since the new likelihood function may have very large curvature near design points, where the smoothing takes place, the use of standard inferential tools of nonlinear regression, linear approximations supplemented by curvature diagnostics, can be misleading, as an example given below shows. Also the method does not seem to be easily adapted to multiple regression problems, e.g., broken plane regression (cf. Siegmund and Zhang (1993)).

The purpose of this paper is to give small sample, conservative confidence regions for θ and conservative joint regions for θ and other parameters of the model (1), which for all practical purposes appear to be exact. We do this by applying the method of Knowles, Siegmund, and Zhang (1991), who considered the special class of nonlinear regression models

$$y_i = \beta f_i(\theta) + \varepsilon_i \quad (i = 1, \cdots, m)$$

with θ a scalar parameter. We also review two alternative methods with which we compare our procedure.

The paper is organized as follows. Section 2 contains our basic theoretical results. In Section 3 we give a brief description of two other methods for the broken line regression model (1). One, the approximate F method, is based on the assumption that an obvious analogue of the usual F statistic of linear regression analysis has approximately an F distribution. The other is the Bayesian analysis of Smith and Cook (1980). In Section 4, some "real" data are examined. To get more insights into our results, Section 5 gives the results of some simulations. Section 6 contains our concluding comments. Detailed proofs of the results in Section 2 are given in an Appendix.

2. Confidence Regions. It is obvious that (1) is a special case of the nonlinear regression model

$$y = Xa + \beta f(x, \theta) + \varepsilon, \tag{3}$$

where $y = (y_1, \cdots, y_m)'$, $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_m)'$, X is an $m \times r$ matrix of full rank with entries depending only on the design variables $x = (x_1, \cdots, x_m)'$, a is a vector of linear parameters and $f(x, \theta)$ is a vector of nonlinear functions of θ and x. For example, in (1), we have

$$X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_m \end{pmatrix},$$
$$a = (\beta_0, \beta_1)',$$
$$f(x, \theta) = (0, \cdots, 0, x_{\tau+1} - \theta, \cdots, x_m - \theta)',$$

if, say, $x_1 \leq \cdots \leq x_\tau < \theta \leq x_{\tau+1} \leq \cdots \leq x_m$. If the covariance matrix of the errors is of the form of a product of the unknown σ^2 and a known matrix, the standard technique of linear regression analysis allows us to transform the problem to one in which the known matrix is the identity. Our methods have nothing to say about heteroscedascity which depends in a more complicated way on unknown parameters.

To construct a confidence region for θ or the pair (β, θ), it is convenient to reduce the model (3) to the case considered by Knowles, Siegmund, and Zhang (1991), where $a = 0$. One possibility is to use a minor variant of the standard technique for reducing a linear model to canonical form (cf. Lehmann (1986), p. 366). Let Q be an $m \times (m - r)$ matrix having orthogonal columns which span the orthogonal complement of the column space of X. Then $Q'X = 0$ and $Q'\varepsilon$ is $N(0, \sigma^2 I)$. Consider the reduced model

$$Q'y = \beta Q'f + Q'\varepsilon,$$

where there are now $m - r$ observations and we have eliminated r nuisance parameters.

It is also possible to obtain joint confidence regions for θ and arbitrary combinations of the other parameters. Motivated by (2), where θ and α have natural physical interpretations, we shall consider joint confidence regions for θ and a single component of a. By a similar transformation we can without loss of generality assume that

$$y = \alpha g + \beta f(\theta) + \varepsilon, \tag{4}$$

where g is a vector not depending on θ and $f(\theta) = f(x, \theta)$.

It is worth noting that in specific cases it is not necessary to find the matrix Q, and it may in fact be convenient to carry out the reduction somewhat differently. For example, to eliminate α in (4), we consider $\tilde{y} = y - \langle g, y\rangle g/\|g\|^2$, which satisfies

$$\tilde{y} = \beta \tilde{f}(\theta) + \tilde{\varepsilon},$$

where $\tilde{f}(\theta) = f(\theta) - \langle g, f(\theta)\rangle g/\|g\|^2$ and $\tilde{\varepsilon} = \varepsilon - \langle g, \varepsilon\rangle g/\|g\|^2$. Here and in the following the angular brackets denote the usual scalar product in Euclidean space and the double vertical bars denote the Euclidean norm. With this reduction we retain m observations, but the distribution of $\tilde{\varepsilon}$ is singular, since it is concentrated on the $m-1$ dimensional orthogonal complement of g. However, since the distribution of $\tilde{\varepsilon}$ is spherically symmetric, the theory developed below applies without change.

Suppose now that we want to find a confidence region for the pair (θ, α) in the reduced model (4). Since this problem was not discussed by Knowles, Siegmund and Zhang (1991), we indicate the relevant calculations here.

Let

$$y^* = y - \alpha_0 g,\ \gamma(\theta) = f(\theta)/\|f(\theta)\|,\ \gamma_g = g/\|g\|, \tag{5}$$

and

$$\tilde{\gamma}(\theta) = \frac{\gamma(\theta) - \langle g, \gamma(\theta)\rangle g/\|g\|^2}{(1 - \langle g, \gamma(\theta)\rangle^2/\|g\|^2)^{\frac{1}{2}}}. \tag{6}$$

The likelihood ratio statistic for testing $H_0 : \theta = \theta_0, \alpha = \alpha_0$ is

$$\Lambda(\theta_0, \alpha_0) = -\frac{1}{2} m \log(\min_\theta \hat{\sigma}_\theta^2/\hat{\sigma}_0^2) \tag{7}$$

where

$$\hat{\sigma}_\theta^2 = m^{-1}(\|y\|^2 - \langle y, \gamma_g\rangle^2 - \langle y, \tilde{\gamma}(\theta)\rangle^2), \tag{8}$$

and

$$\hat{\sigma}_0^2 = m^{-1}(\|y^*\|^2 - \langle y^*, \gamma(\theta_0)\rangle^2). \tag{9}$$

The sufficient statistics under H_0 are $\|y^*\|$ and $\langle y^*, f(\theta_0)\rangle$. Our conditional likelihood ratio confidence region is the set of all (θ_0, α_0) such that

$$P_{\theta_0,\alpha_0}[\Lambda(\theta_0, \alpha_0) > b \,\big|\, \|y^*\|, \langle y^*, f(\theta_0)\rangle] > 1 - \delta, \tag{10}$$

where $b = \{\Lambda(\theta_0, \alpha_0)\}_{obs}$. By sufficiency this conditional probability does not depend on the unknown nuisance parameters β and σ^2. Observe that

$$\begin{aligned}\Lambda(\theta_0, \alpha_0) &= -\frac{1}{2} m \log \frac{\min_\theta \hat{\sigma}_\theta^2}{\hat{\sigma}_{\theta_0}^2} - \frac{1}{2} m \log \frac{\hat{\sigma}_{\theta_0}^2}{\hat{\sigma}_0^2} \\ &\triangleq \Lambda(\theta_0) + \tilde{\Lambda}(\theta_0, \alpha_0),\end{aligned}$$

where

$$\Lambda(\theta_0) = -\frac{1}{2} m \log \frac{\min_\theta \hat{\sigma}_\theta^2}{\hat{\sigma}_{\theta_0}^2}$$

and

$$\tilde{\Lambda}(\theta_0, \alpha_0) = -\frac{1}{2} m \log \frac{\hat{\sigma}_{\theta_0}^2}{\hat{\sigma}_0^2}.$$

Then the left hand side of (10) is equal to

$$\begin{aligned}&P_{\theta_0,\alpha_0}[\tilde{\Lambda}(\theta_0, \alpha_0) > b \big| \, \|y^*\|, \langle y^*, f(\theta_0)\rangle] \\ &+ E_{\theta_0,\alpha_0}[P_{\theta_0,\alpha_0}\{\Lambda(\theta_0) > b - \tilde{\Lambda}(\theta_0, \alpha_0) \big| \, \|y\|, \langle y, f(\theta_0)\rangle, \langle y, g\rangle\} \\ &\quad \times \mathbf{1}\{\tilde{\Lambda}(\theta_0, \alpha_0) \le b\} \big| \, \|y^*\|, \langle y^*, f(\theta_0)\rangle].\end{aligned} \tag{11}$$

The term $\tilde{\Lambda}(\theta_0, \alpha_0)$ is just a version of the likelihood ratio statistic for testing $\alpha = \alpha_0$ in the model (4) with $\theta = \theta_0$, which is a linear hypothesis. Hence $\tilde{\Lambda}(\theta_0, \alpha_0)$ is a monotonic function of $|t|$ with $m - 2$ degrees of freedom, both conditionally and unconditionally. The distribution of $\Lambda(\theta_0)$ is much more complicated. However, by virtue of conditioning on $\|y^*\|$, $\langle y^*, f(\theta_0)\rangle$, $\langle y^*, g\rangle$ or equivalently on $\|y\|$, $\langle y, f(\theta_0)\rangle$, $\langle y, g\rangle$, which are sufficient for σ, β, α when $\theta = \theta_0$ is known, we see that $\tilde{\Lambda}(\theta_0, \alpha_0)$ is a constant and the conditional distribution of $\Lambda(\theta_0)$ does not depend on unknown parameters. A geometric interpretation of the conditional probability in (11) is given below, and a sharp upper bound is given in an Appendix. Here we summarize the result.

Let $U^{(n)} = (u_1^{(n)}, \cdots, u_n^{(n)})'$ denote a generic random variable uniformly distributed on the unit sphere S^{n-1}. Let f_{n-1} and F_{n-1} denote the density and distribution functions of $u_1^{(n)}$, respectively.

THEOREM 1. *Let* $c^2 = 1 - \exp\{-2b/m\}$.

(i) *The first term in (11) is*

$$P[|u_1^{(m-1)}| > c]. \tag{12}$$

(ii) *Let* $\|y^*\| = \eta$, $\langle \gamma(\theta_0), y^* \rangle = \phi$,

$$\frac{g'(I-\gamma(\theta_0)\gamma'(\theta_0))y^*}{(\|g\|^2 - \langle g, \gamma(\theta_0)\rangle^2)^{\frac{1}{2}} \|(I-\gamma(\theta_0)\gamma'(\theta_0))y^*\|} = \xi,$$

and

$$\zeta = \xi(\|g\|^2 - \langle g, \gamma(\theta_0)\rangle^2)^{\frac{1}{2}}(\eta^2 - \phi^2)^{\frac{1}{2}} + \phi\langle g, \gamma(\theta_0)\rangle + \alpha_0\|g\|^2.$$

Then the second term in (11) is

$$\int_{|\xi|\le c} P[\max_\theta \langle \tilde\gamma(\theta), U^{(m)}\rangle^2 > w^2 \Big| \langle \tilde\gamma(\theta_0), U^{(m)}\rangle = z, \qquad \langle \gamma_g, U^{(m)}\rangle = v] f_{m-2}(\xi) d\xi, \tag{13}$$

where

$$w^2 = 1 - \frac{\zeta^2/\|g\|^2 + (1-c^2)(\eta^2-\phi^2)}{\eta^2 + (2\zeta - \alpha_0\|g\|^2)\alpha_0}, \tag{14}$$

$$z = \frac{\phi + \langle g, \gamma(\theta_0)\rangle(\alpha_0 - \zeta/\|g\|^2)}{(1 - \langle g, \gamma(\theta_0)\rangle^2/\|g\|^2)^{\frac{1}{2}}(\eta^2 + (2\zeta - \alpha_0\|g\|^2)\alpha_0)^{\frac{1}{2}}}, \tag{15}$$

and

$$v = \frac{\zeta}{\|g\|(\eta^2 + (2\zeta - \alpha_0\|g\|^2)\alpha_0)^{\frac{1}{2}}}. \tag{16}$$

(iii) *The left hand side of (10) is the sum of (12) and (13).*

REMARK. The conditional probability in (13) has an interesting geometric interpretation. Conditional on $\langle \gamma_g, U^{(m)}\rangle = v$, $U^{(m)}$ has the representation $U^{(m)} = v\gamma_g + (1-v^2)^{\frac{1}{2}}\tilde U^{(m-1)}$, where $\tilde U^{(m-1)}$ is uniformly distributed on the $m-2$ dimensional unit sphere in the $m-1$ dimensional space orthogonal to γ_g. Since $\langle \gamma_g, \tilde\gamma(\theta)\rangle = 0$ for all θ, the conditional probability in (13) equals

$$P[\max_\theta \langle \tilde\gamma(\theta), \tilde U^{(m-1)}\rangle^2 > w^2/(1-v^2) \Big| \langle \tilde\gamma(\theta_0), \tilde U^{(m-1)}\rangle = z/(1-v^2)^{\frac{1}{2}}]. \tag{17}$$

The condition in (17) specifies that $\tilde U^{(m-1)}$ is on a sphere of geodesic radius $\cos^{-1}[z/(1-v^2)^{\frac{1}{2}}]$ about $\tilde\gamma(\theta_0)$; and $\langle \tilde\gamma(\theta), \tilde U^{(m-1)}\rangle^2 > w^2/(1-v^2)$ for some θ if and only if $\tilde U^{(m-1)}$ is within a tube of radius $\cos^{-1}[w/(1-v^2)^{\frac{1}{2}}]$ about either $\tilde\gamma$ or $-\tilde\gamma$. Hence the conditional probability (17) is the proportion of the sphere intersected by the tubes (cf. Figure 1 of Knowles, Siegmund, and Zhang (1991)). In the trivial case that $\tilde\gamma$ is a great circle this intersection is the union of two spherical caps, and (17) equals

$$2\{1 - F_{m-2}[(w^2 - z^2)^{\frac{1}{2}}/(1 - v^2 - z^2)^{\frac{1}{2}}]\}. \tag{18}$$

Since there does not seem to be a direct geometric method for computing the probability (17) in general, the following theorem uses Rice's formula for the expected number of upcrossings of a level to give an upper bound. Its proof is sketched in an Appendix. Knowles, Siegmund, and Zhang (1991) give a more detailed discussion, evidence that the method usually gives very close to exact results, and an approximation in the form of a perturbation of (18) reflecting the extent to which $\tilde{\gamma}$ fails to be a geodesic.

THEOREM 2. *Let*

$$\rho = \langle \tilde{\gamma}(\theta), \tilde{\gamma}(\theta_0) \rangle, \tag{19}$$

$$\mu = \dot{\rho}(z - w\rho)/(1 - \rho^2), \tag{20}$$

and

$$\tau^2 = (\|\dot{\tilde{\gamma}}\|^2 - \dot{\rho}^2/(1 - \rho^2))(1 - z^2 - v^2 - (w - \rho z)/(1 - \rho^2)). \tag{21}$$

Then

$$\begin{aligned} &P[\max_\theta \langle \tilde{\gamma}(\theta), U^{(m)} \rangle > w \Big| \langle \tilde{\gamma}(\theta_0), U^{(m)} \rangle = z, \langle \gamma_g, U^{(m)} \rangle = v] \\ &\leq (1 - z^2 - v^2)^{-\frac{1}{2}} \int_\theta [\mu F_{m-4}(\frac{\mu}{\tau}) + (m-3)^{-1} \tau f_{m-3}(\frac{\mu}{\tau})] \\ &\times f_{m-3}(\frac{w - \rho z}{\sqrt{(1 - \rho^2)(1 - z^2 - v^2)}})(1 - \rho^2)^{-\frac{1}{2}}\, d\theta. \end{aligned} \tag{22}$$

REMARK. In order to approximate (17), (22) must be used twice: once for $\tilde{\gamma}$ and once for $-\tilde{\gamma}$. In practice the curve $-\tilde{\gamma}$ usually contributes a negligible amount to the overall probability, but in some cases it can be important.

3. Review of Two Other Methods.

3.1. Approximate F Method. Consider the general nonlinear regression model $y = h(\lambda) + \varepsilon$ and suppose that the vector λ is partially specified by the hypothesis $H_0 : \lambda = \lambda_0$. Let $\hat{\lambda}$ and $\hat{\lambda}_0$ denote maximum likelihood estimators without restriction and restricted by the hypothesis, respectively. In the case of a linear function h the ratio $[\|y - h(\hat{\lambda}_0)\|^2 - \|y - h(\hat{\lambda})\|^2]/\|y - h(\hat{\lambda})\|^2$ is, up to normalization, the familiar F statistic for testing the hypothesis H_0. The approximate F method is based on the assumption that this same distribution is approximately valid even if h is nonlinear.

For example, for the model (4) with the notation introduced in (8) and (9), let

$$F = \frac{m-3}{2} \cdot \frac{\hat{\sigma}_0^2 - \hat{\sigma}_{\hat{\theta}}^2}{\hat{\sigma}_{\hat{\theta}}^2}. \tag{23}$$

In a linear model this statistic F would be distributed as $F_{2,m-3}$ when $\theta = \theta_0, \alpha = \alpha_0$. For the approximate F method we pretend it has this same distribution under the nonlinear model. To the extent that it does, the set of all (θ_0, α_0) for which $F \leq F_{2,m-3}(1-\delta)$ is approximately a $1-\delta$ confidence region for (θ, α).

Like the conditional likelihood ratio region of Section 2 the approximate F region is based on the likelihood ratio statistic. However, the boundaries of the region are determined by the hypothesized unconditional distribution of the statistic, whose true distribution cannot be used because it depends on unknown nuisance parameters. In the case of a linear model (this is essentially the trivial case in (18) where the curve $\tilde{\gamma}$ is a great circle) the conditional and unconditional distributions coincide, and the two methods yield the same confidence region. It seems reasonable that the two regions should be close to each other in cases of "small" nonlinearity. What is somewhat surprising is that with one qualification the regions turn out to be close to one another even if there is substantial nonlinearity.

Although the approximate F method is highly regarded in the nonlinear regression literature, we believe that the justification traditionally offered for attaching a confidence level to the regions thus obtained is not as strong as one would hope. There is a sequence of papers (Beale (1960), Johansen (1984), Hamilton and Wiens (1987)) showing that as σ tends to 0 the approximate F statistic has in the limit the appropriate F distribution. This is the case of small nonlinearity. The regression surface can be well approximated by its tangent plane at the true parameter value, and in the limit the nonlinearity vanishes. A consequence is that in the limit the approximate F region agrees with other regions, notably the Halperin (1963) region, which are known to behave badly when the regression function is moderately nonlinear (Bates and Watts (1988, pp. 223-9), Knowles, Siegmund, Zhang (1991), Section 5 below). The asymptotic theory also contains local curvature corrections which allow one to adjust for nonlinearity; but the broken line model is locally flat so for our problem these local curvature corrections are zero, except at design points where they do not exist. This asymptotic analysis leaves us in the uncomfortable state that the claimed coverage probability of the approximate F region is justified when the region is close to other regions which experience tells us can behave quite poorly.

In addition to these asymptotic analyses, the Monte Carlo study of Donaldson and Schnabel (1987) reports that the true coverage probability of the approximate F method is in all the cases they considered close to the nominal level. However, they consider only one case of large (intrinsic) curvature and restricted their study to confidence regions for the entire parameter vector.

For the restricted class of models they considered, Knowles, Siegmund, and Zhang (1991) found good agreement between the conditional likelihood ratio and the approximate F method, even in the presence of large curvature, except when the boundary of the parameter space plays a role. The essential geometric insight provided by analysing the conditional likelihood ratio method is this. The approximate F method implicitly assumes that the curve $\tilde{\gamma}$ is a great circle. In this case the probability (17) is given by (18). When the curve is not a great circle, the cap on the sphere of geodesic radius $\cos^{-1}(z/(1-v^2)^{1/2})$ about $\tilde{\gamma}(\theta_0)$ cut off by the tube around $\tilde{\gamma}$ is itself not spherical, so (17) and (18) are unequal, and in most cases (17) is slightly larger. However, in some cases the curve $\tilde{\gamma}$ ends before exiting from the sphere. Then the cap cut off by the tube can be very small, or not exist at all, depending on the minimum distance to the sphere from the curve, and contrary to what the curvature corrections of the small σ theory suggest (17) can be much smaller than (18). It appears to be only this second case which leads to substantial discrepancies between the conditional likelihood ratio and approximate F methods, and then the approximate F method is overly conservative (cf. Section 5).

Since the approximate F method is more easily evaluated than our method, it presumably would be preferred if the two methods behave similarly. Hence a numerical comparison of the two methods provides empirical evidence for the case of broken line regression regarding the validity of the assumed coverage probability of the approximate F method. We also learn when to expect discrepancies between the actual and nominal levels of the approximate F method. In these cases one may prefer our method to obtain a more precise confidence region. Alternatively one might use it as a diagnostic at a few parameter values to see if the approximate F method is performing well or to determine what kind of adjustments to the approximate F region are required to make its nominal and actual levels more consistent.

For the broken line regression models it is easy to obtain the approximate F region numerically. For the model (4), we can in fact proceed analytically, as follows. For fixed $\theta = \theta_0$ and $\alpha = \alpha_0$, the residual sum of squares is

$$\begin{aligned} m\hat{\sigma}_0^2 &= \|y^*\|^2 - \langle \gamma(\theta_0), y^* \rangle^2 \\ &= (\|g\|^2 - \langle g, \gamma(\theta_0)\rangle^2)\alpha_0^2 \\ &\quad - 2\langle g, (I - \gamma(\theta_0)\gamma'(\theta_0))y\rangle \alpha_0 + \|y\|^2 - \langle y, \gamma(\theta_0)\rangle^2. \end{aligned} \tag{24}$$

where $\gamma(\theta)$ and y^* are defined in (5).

By (23) the boundary of the approximate F region satisfies

$$\hat{\sigma}_0^2 = \left(\frac{2}{m-3}F_{2,m-3}(1-\delta)+1\right)\hat{\sigma}_{\hat{\theta}}^2. \tag{25}$$

Since $\hat{\sigma}_0^2$ is a quadratic polynomial in α_0, the boundary points for given θ_0 are the real solutions of

$$a(\theta_0)\alpha_0^2 - 2b(\theta_0)\alpha_0 + c(\theta_0) = 0, \tag{26}$$

where

$$a(\theta_0) = \|g\|^2 - \langle g, \gamma(\theta_0)\rangle^2, \tag{27}$$

$$b(\theta_0) = \langle g, (I - \gamma(\theta_0)\gamma'(\theta_0))y\rangle, \tag{28}$$

and

$$c(\theta_0) = \|y\|^2 - \langle y, \gamma(\theta_0)\rangle^2 - m(\frac{2}{m-3}F_{2,m-3}(1-\delta)+1)\hat{\sigma}_{\hat{\theta}}^2. \tag{29}$$

Hinkley (1971) considered the model (2), for which he pointed out that the approximate F region for (θ, α) can be obtained by solving a quadratic equation like (26). However, some other aspects of his procedure, which is based on asymptotic considerations, seem to have undesirable consequences for small sample sizes. See the discussion in Section 4 and Figure 1.

3.2. Bayesian Method. Smith and Cook (1980) gave a Bayesian analysis of the broken line regression model (1). One crucial assumption concerns the prior distribution for σ, namely, $\nu\lambda/\sigma^2 \sim \chi^2(\nu)$, for some choices of ν, the degrees of freedom, and λ, a constant. They used $\lambda = 0$, which can be understood as $\lambda \to 0$, and $\nu = 0$ or -4. For the smaller the value of ν the confidence interval is more conservative. We shall also use these choices in our examples.

4. Examples. In this section, three data sets are considered. Model (1) seems reasonable for all of them. The conditional likelihood ratio (CLR) method, the approximate F (AF) method and the Bayesian method are used to find approximate confidence intervals for the change-point θ. The results are compared for different confidence levels. For Data 1, model (2) is also considered. Confidence intervals for the change-point θ and joint confidence regions for the change-point θ and the maximum response α are derived by the CLR and AF methods, and compared with Hinkley's (1971) results. Smith and Cook's (1980) analysis concerns only the change-point θ, and we do not attempt to adapt their method to obtain joint regions for θ and α. For Data 2, we only make inferences on the change-point, which is the parameter of interest in that case.

EXAMPLE 1. In this example, we study Hinkley's (1971) data from Pool and Borchgreving (1964). First, we consider model (1). Table 1 gives the 0.95 and 0.99 confidence intervals for θ obtained from different methods. The numbers in AF are given by Hinkley (1971).

Table 1: Confidence Intervals for Hinkley's Data: Model 1

method	ν	0.95	0.99
CLR		(4.044, 5.238)	(3.689, 5.502)
AF		(4.068, 5.202)	(3.641, 5.441)
Bayesian	0	(4.179, 5.134)	(4.061, 5.299)
Bayesian	-4	(4.117, 5.204)	(4.007, 5.390)

Model (2) seems adequate for these data (cf. Hinkley (1971)), under which we also computed confidence intervals for θ, see Table 2.

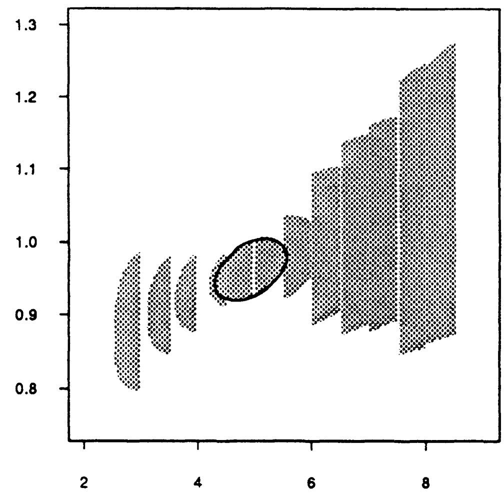
Table 2: Confidence Intervals for Hinkley's Data: Model 2

method	0.95	0.99
CLR	(4.441, 5.418)	(4.245, 5.700)
AF	(4.450, 5.389)	(4.253, 5.655)

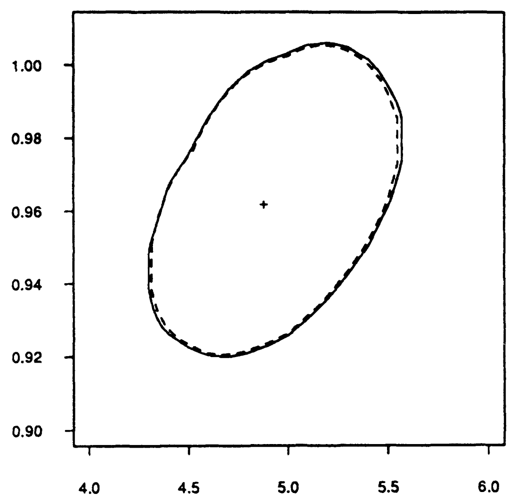
As shown in the tables, the results obtained by the CLR and AF methods are fairly close although the conditional likelihood ratio method gives slightly more conservative intervals than the approximate F does. However, even if we use the conservative prior of $\nu = -4$, the Bayesian intervals appear to be slightly narrower than those of the CLR and AF methods.

Now we turn to the joint confidence region for (θ, α). Figure 1 shows the 0.95 confidence regions under model (2). The shaded area is obtained from Hinkley's (1971) original argument (after correcting what we believe are minor typographical errors in the expressions for $S_t'(y, w)$, where $\bar{x}_T$ should be w, for $H_t'(y)$, where $\bar{x}_t$ should have an asterisk, and for $K_t'(y)$, where $\tilde{\beta}_{0t}$ should not be squared). The dotted line is from the approximate F method described in Section 3.1; and the solid line is from the conditional likelihood ratio method given in Section 2. As one can see from the figure, the region from Hinkley's (1971) method looks peculiar. But the other two regions are more or less elliptic, which is what we should expect since these data are well behaved.

EXAMPLE 2. Since the original data are not available, the numerical values in Table 3 are inferred from Figure 1 of Smith and Cook (1980). In this



(a)



(b)

Figure 1: 0.95 confidence region for Hinkley's (1971) data. In (a), the shaded area is from Hinkley's (1971) method, the coincident solid and dashed lines are the conditional likelihood ratio region and the approximate F rgion. The latter are plotted in (b) again. The plus sign is the location of MLE of (θ, α).

application, model (1) seems reasonable. The change-point θ corresponds to the time at which a rejection occurs after a patient has received a kidney transplant. See Smith and Cook (1980) for more details.

Table 3: Renal Transplant Data
(Smith and Cook, 1980)

	Patient A	Patient B
x	y	
1	47.5	36.0
2	57.0	45.5
3	61.0	50.0
4	71.0	60.0
5	67.2	73.3
6	54.4	71.0
7	48.3	66.7
8	43.2	60.0
9		30.5
10		18.3

Again we computed conditional likelihood ratio, approximate F, and two Bayesian intervals for θ. The results are given in Table 4. The less conservative Bayesian intervals are somewhat shorter than other three, which are reasonably consistent with one another.

Table 4: 95% Interval for Renal Transplant Data

method	ν	Patient A	Patient B
CLR		(3.649, 4.655)	(5.550, 6.984)
AF		(3.609, 4.691)	(5.547, 7.319)
Bayesian	0	(3.748, 4.531)	(6.069, 6.892)
Bayesian	-4	(3.502, 4.731)	(5.747, 7.156)

EXAMPLE 3. We now consider Data Set 3 in Table 6.18 of Ratkowsky (1983). In contrast to the two previous examples which have been modeled in the literature as broken line regressions, in analyzing these and similar data, Ratkowsky suggests models which allow for a smooth transition between linear regimes. The specific model he discusses is the "bent hyperbola" model,

$$y = \beta_0 + \beta_1 x + \beta_2[(x-\theta)^2 + \delta]^{\frac{1}{2}} + \varepsilon, \tag{30}$$

proposed by Griffiths and Miller (1973). Other possibilities are discussed by Seber and Wild (1989, pp. 465-480). The broken line regression model (1) is the special case of (30) with $\delta = 0$. Although the data are well behaved and seem to be well described by this model, the evidence that δ is different from zero is weak (Ratkowsky, 1983, p. 125). Moreover, the notion of a smooth transition between linear regimes, while intuitively appealing, does not seem to help us understand these data, where the transition occurs over a very short part of the range of the independent variable. We have used the model (1). Table 5 contains the four interval estimates we have discussed above for the change-point, θ. Figure 2 contains conditional likelihood ratio and approximate F confidence regions for (θ, β_2). All methods are in close agreement.

Table 5: Confidence Interval for Ratkowsky's Data

method	ν	0.90	0.99
CLR		(17.582, 18.631)	(17.268, 18.901)
AF		(17.602, 18.615)	(17.283, 18.897)
Bayesian	0	(17.670, 18.575)	(17.384, 18.818)
Bayesian	-4	(17.630, 18.609)	(17.318, 18.862)

We were originally attracted to these data by Ratkowsky's (1983, p. 125) report of a large value for the Bates-Watts (1980) intrinsic curvature diagnostic. Since the conventional interpretation of a large intrinsic curvature is that the coverage probability of the approximate F method may not be consistent with the nominal value, these data seemed like an interesting test case for our methods. The satisfactory performance of the approximate F method seen in Table 5 and Figure 2 seems to provide support for the hypothesis of Knowles, Siegmund, and Zhang (1991) that its coverage probability is close to the nominal value even if there is large curvature, unless the confidence region contains values of θ close to the ends of the curve $\tilde{\gamma}(\theta)$ defined in (6) (cf. Section 5).

However, a closer look makes one dubious of the usefulness of the usual intrinsic curvature diagnostic for the model (30) when δ is close to 0. For the limiting case of broken line regression, where $\delta = 0$, the regression function is linear between design points, but has a discontinuous derivative at design points. Hence the intrinsic curvature for that model equals zero, except at the design points, where it is undefined. It seems plausible that the intrinsic curvature diagnostic may be quite misleading for the model (30) when it is close to the broken line model (1).

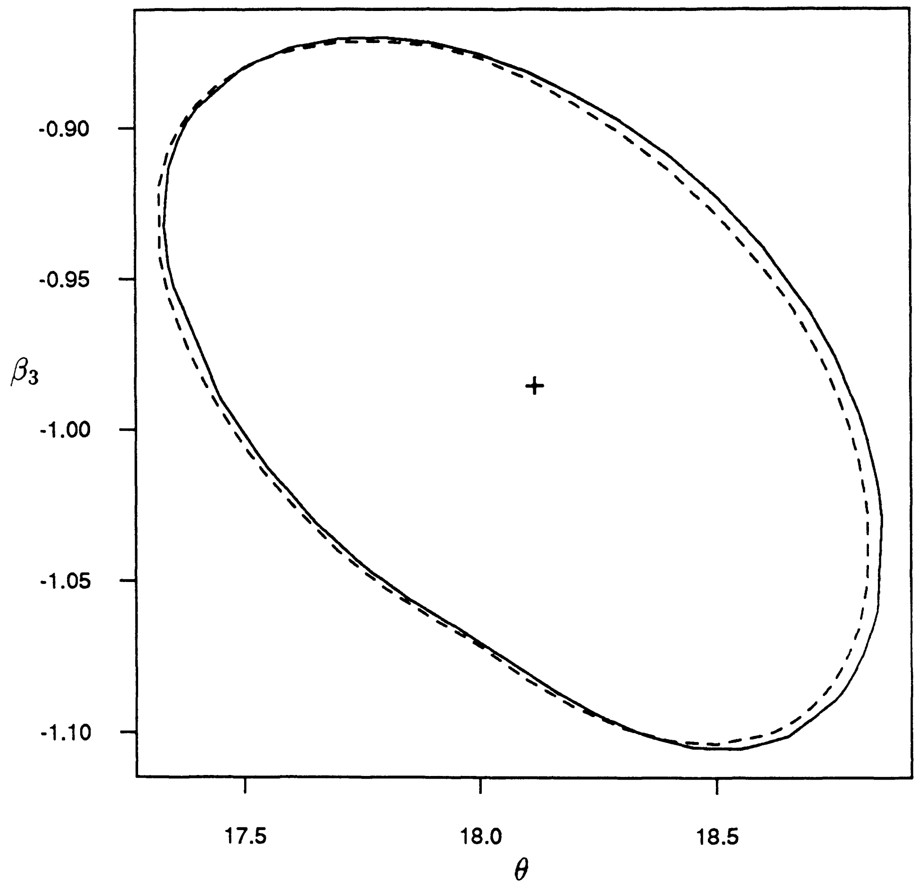


Figure 2: 0.95 joint confidence region for Ratkowsky's data. The solid line is the conditional likelihood ratio region, and the dashed line is the approximate F region.

Table 6: Smooth Model for Ratkowsky's Data

$\delta(\hat{\theta})$	method	0.90	0.99
0.1	AF	(17.602, 18.615)	(17.283, 18.897)
(18.144)	CLR	(17.586, 18.627)	(17.272, 18.900)
	curvature	0	0
0.5	AF	(17.608, 18.618)	(17.270, 18.990)
(18.124)	CLR	(17.604, 18.622)	(17.263, 18.993)
	curvature	1.251	1.865
3.4362	AF	(17.903, 19.167)	(17.499, 19.581)
(18.522)	CLR	(17.901, 19.168)	(17.497, 19.583)
	curvature	0.424	0.631

To test this last hypothesis we considered a related model suggested by Tishler and Zang (1981), which has some advantages for our purposes. The response y is given by

$$y = \beta_0 + \beta_1 x + \beta_2 q_\delta(x - \theta) + \varepsilon. \tag{31}$$

where

$$q_\delta(z) = \begin{cases} 0 & \text{if } z \le -\delta, \\ (z+\delta)^2/4\delta & \text{if } -\delta < z \le \delta, \\ z & \text{if } z > \delta. \end{cases} \tag{32}$$

This model is similar to (30) in the sense of providing a smooth transition between linear regimes and being equal to (1) in the limit when $\delta = 0$. Moreover, the smoothing parameter δ has a clear interpretation: outside an interval of length 2δ centered at θ the regression function is exactly linear and hence is easily compared with (1). In Table 6 we consider three different values of δ. Since Tishler and Zang (1981) regard δ as a convenient technical device to facilitate numerical computation of the maximum likelihood estimators of the other parameters and suggest arbitrary small values which can be perturbed without substantially changing those estimators, we use $\delta = 0.1$ and $\delta = 0.5$. The third value is $\delta = 3.4362$, which is the maximum likelihood estimator and would presumably be used by someone who finds the smooth model (31) preferable to (1). As measured by an F ratio, the maximum likelihood value gives a marginally better fit ($p = 0.038$) than $\delta = 0$. We have regarded these values of δ as fixed in computing the entries in Table 6. Since the independent variable ranges from $x = 1$ to $x = 27$, the differences in the estimated values of θ seem insignificant. However the variability in the Bates-Watts (1980) curvature diagnostic, computed from an appropriate modification of display (28) of Knowles, Siegmund and Zhang (1991), is quite substantial. The coverage probability of the approximate F method appears to be quite close to the nominal value in all cases.

5. Simulations. In this section we compare the performance of the conditional likelihood ratio and approximate F methods on simulated data. For simplicity we consider the model (1) with $\beta_0 = \beta_1 = 0$ and put $\beta_2 = \beta$. In the preceding examples the data were well behaved, and there was close agreement between the two methods. Here we shall be particularly interested in problematic data. It is clear that a small value of β makes it difficult to estimate θ accurately. Difficulties also arise if the change-point θ is near to the end of the range of the independent variable. According to Knowles, Siegmund, and Zhang (1991) the conditional likelihood ratio and approximate F methods should give essentially the same intervals in the first case, which is more or less a case of large curvature, although as indicated above the concept of curvature is itself problematic for the model (1). However, the two methods may disagree in the second case.

Table 7: A Random Sample from $N(0, 1)$

-0.56340998	1.32382202	-0.87364787	-1.70070076	-1.42179060
0.79639786	-0.24871901	-0.82794911	0.74958265	1.09769726
-2.23353744	0.06592534	-0.31559977	-1.24223769	-0.96793693

For our experiments we put $x_i = i$ for $i = 1, 2, \cdots, 15$ and generated a single sample of size fifteen from the standard normal distribution to serve as a common set of residuals for different values of θ and β. See Table 7 for the simulated sample. Table 8 contains some basic results. The column headed w gives the observed value of (14). Table 9 contains 0.90 and 0.95 confidence intervals for θ obtained by the conditional likelihood ratio and approximate F methods. To indicate the disagreement between these methods, we also report the ratios $(a_1 - a)/(b - a)$ and $(b - b_1)/(b - a)$, where (a, b) and (a_1, b_1) are the conditional likelihood ratio and approximate F confidence intervals, respectively. In cases marked by an asterisk, at least one of the confidence intervals included the endpoint of the parameter space, so the ratio does not have a simple interpretation.

Table 8: Key Parameters in Simulated Models

θ	β	$\hat{\theta}$	$\hat{\beta}$	w
7.5	1.0	7.318	0.847	0.9444
7.5	0.5	7.056	0.347	0.7768
3.5	0.5	4.000	0.464	0.9359
10.5	1.5	11.00	1.478	0.9064
12.5	2.0	12.711	1.674	0.7404

Table 9: Confidence Intervals for Simulated Data

θ	β	method	0.90	0.95
7.5	1.0	CLR	(5.5496, 8.6864)	(5.0568, 8.9476)
		AF	(5.3212, 8.6279)	(4.7922, 8.8742)
		difference	-7.28%, 1.86%	-6.801%, 1.89%
7.5	0.5	CLR	(3.0040, 10.9989)	(1.7623, 11.5662)
		AF	(2.6959, 10.4185)	(1.3335, 11.4637)
		difference	-3.85%, 7.26%	-4.37%, 1.05%
3.5	0.5	CLR	(2.5207, 5.6039)	(1.7715, 5.9547)
		AF	(2.0473, 5.6644)	(1.1234, 6.0301)
		difference	-15.35%, -1.96%	-21.02%, -2.45%
10.5	1.5	CLR	(9.7707, 11.7838)	(7.6834, 12.6443)
		AF	(10.0243, 11.7294)	(7.5714, 12.5662)
		difference	14.87%, 3.19%	2.24%, 1.56%
12.5	2.0	CLR	(10.5596, 13.9837)	(10.1110, 15.0000)
		AF	(10.6332, 15.0000)	(10.1234, 15.0000)
			2.15%, -29.68%*	0.25%, 0.00%*

As expected, the two methods are in fairly good agreement even when β is small and hence the confidence interval is large, except when the interval includes points near the end of the parameter set. In that case the disagreement can be substantial, and the usually slightly longer conditional likelihood ratio interval can be substantially shorter.

For the data of the third row of Table 9 we have also computed the intervals suggested by Halperin (1963). Although, as indicated above, these intervals have been severely criticized, the facts that they have exactly the claimed coverage probability and in the linear case agree with intervals obtained from the F statistic give them an intuitive appeal. The 0.90 and 0.95

intervals are (1.0,6.09) and (1.0,7.1), respectively. They are much longer than both the AF and CLR intervals.

It is also interesting to evaluate the conditional coverage probability of the approximate F intervals. For the data in the first row of Table 9, at the end points of the 0.95 AF intervals, our conditional evaluation (upper bound) of the nominally 0.05 probabilities are 0.035 and 0.061 at the left and right endpoints respectively. Thus although there is a relatively minor discrepancy between the AF and CLR intervals, the discrepancy as measured by the differences in tail probabilities seems comparatively large. The positive discrepancy at the right hand end point is a curvature effect, whereas the negative one at the left hand end point must be a boundary effect (cf. the geometric comparison of the AF and CLR probability evaluations in Sections 2 and 3). The size of the boundary effect is surprising, since the left end point does not seem especially close to the boundary of the parameter set.

Other sample sizes and parameter values gave similar results.

6. Discussion. We have shown that the conditional likelihood ratio method of Knowles, Siegmund, and Zhang (1991) can be implemented to provide confidence regions in an important class of nonlinear regression models having a substantial literature of their own – the broken line models. Since this method is more complicated numerically than the approximate F method, we have also attempted to provide empirical support for the conjecture of Knowles, Siegmund, and Zhang (1991) that the simpler method yields close to the same confidence regions, even in cases of large curvature, provided the confidence region does not involve values of θ near the ends of the curve $\tilde{\gamma}(\theta)$. We have also reviewed the Bayesian method of Smith and Cook (1980), which does not seem to offer any particular advantages unless one wants a Bayesian method. Since one presumably would prefer to use the simpler approximate F method unless the accuracy of its coverage probability is in doubt, a possible use of our method is as a diagnostic, which would be substituted for the approximate F method only in those cases where there appears to be a substantial discrepancy between the nominal and actual coverage probabilities of that method.

It is tempting to extrapolate these findings to more general nonlinear regression models and conjecture that the approximate F regions will essentially always have about the nominal coverage probability, except when the boundary of the parameter space is involved. However, it should be kept in mind that our method and the accompanying geometric intuition apply only to cases where all nonlinear parameters are included in the confidence region so by appropriate conditioning one can in principle obtain an exact region.

Acknowledgment. The research of the first author was supported in part by NSA Grant MDA 904-89-H-2040 and NSF Grant DMS 91-04432.

Appendix: Proof of the Theorems.

Proof of Theorem 1.

(i) Note that $\tilde{\Lambda}(\theta_0, \alpha_0)$ is the log likelihood ratio statistic under the hypothesis $H_0' : \alpha = \alpha_0$ for the model

$$y = \alpha g + \beta f(\theta_0) + \varepsilon.$$

Let $\Gamma_0 = I - \gamma(\theta_0)\gamma'(\theta_0)$ be projection to the orthogonal complement of $\gamma(\theta_0)$. It follows from linear regression theory that

$$m(\hat{\sigma}^2_{\theta_0} - \hat{\sigma}^2_0) = -\frac{\langle \Gamma_0 g, y^* \rangle^2}{\|\Gamma_0 g\|^2},$$

and hence

$$\begin{aligned} &P_{\theta_0,\alpha_0}[\tilde{\Lambda}(\theta_0, \alpha_0) > b \Big| \, \|y^*\|, \langle \gamma(\theta_0), y^* \rangle] \\ &= P_{\theta_0,\alpha_0}[\frac{\langle \Gamma_0 g, y^* \rangle^2}{\|\Gamma_0 g\|^2 \|\Gamma_0 y^*\|^2} > c^2 \Big| \, \|\Gamma_0 y^*\|] \\ &= P[|u_1^{(m-1)}| > c]. \end{aligned}$$

(ii) We want to evaluate

$$P_{\theta_0,\alpha_0}[\Lambda(\theta_0) > b - \tilde{\Lambda}(\theta_0, \alpha_0) \Big| \, \|y^*\|, \langle \gamma(\theta_0), y^* \rangle, \tilde{\Lambda}(\theta_0, \alpha_0)]. \tag{33}$$

For convenience, consider the specific values

$$\|y^*\| = \eta, \; \langle \gamma(\theta_0), y^* \rangle = \phi, \; \frac{\langle \Gamma_0 g, y^* \rangle}{\|\Gamma_0 g\| \; \|\Gamma_0 y^*\|} = \xi. \tag{34}$$

Under the conditions in (34), we have

$$\begin{aligned} \hat{\sigma}^2_0 &= m^{-1}(\eta^2 - \phi^2), && (35) \\ \langle g, y \rangle &= \xi(\|g\|^2 - \langle g, \gamma(\theta_0) \rangle^2)^{\frac{1}{2}}(\eta^2 - \phi^2)^{\frac{1}{2}} + \phi \langle g, \gamma(\theta_0) \rangle + \alpha_0 \|g\|^2 && (36) \\ &\triangleq \zeta, && (37) \end{aligned}$$

and

$$\|y\| = (\eta^2 + (2\zeta - \alpha_0 \|g\|^2)\alpha_0)^{\frac{1}{2}}. \tag{38}$$

Some manipulation shows that

$$\Lambda(\theta_0) > b - \tilde{\Lambda}(\theta_0, \alpha_0)$$

if and only if

$$\max_\theta \langle \tilde{\gamma}(\theta), y \rangle^2 \; > \; \|y\|^2 - \langle y, g \rangle^2 / \|g\|^2 - (1 - c^2) m \hat{\sigma}^2_0. \tag{39}$$

For $\theta = \theta_0$, $\|y^*\|$, $\langle\gamma(\theta_0), y^*\rangle$ and $\tilde{\Lambda}(\theta_0, \alpha_0)$ are the sufficient statistics for α, β and σ. Hence to find (33), we can assume $y \sim N(0, I)$, and it follows from (35)-(39), that

$$(30) = P[\max_\theta \langle\tilde{\gamma}(\theta), U^{(m)}\rangle^2 > w \Big| \langle\tilde{\gamma}(\theta_0)\rangle, U^{(m)} = z, \langle\gamma_g, U^{(m)}\rangle = v], \qquad (40)$$

where w, z and v defined in (14), (15), and (16).

PROOF OF THEOREM 2. Our proof is an application of Rice's formula in the form of the inequality

$$P[\max_{t_0 \le t \le t_1} Z_t > w] = \int_{t_0}^{t_1} E(\dot{Z}_t^+ \Big| Z_t = w) g_t(w)\, dt, \qquad (41)$$

where g_t is the probability density function of Z_t, $\dot{h} = dh/dt$, and Z_{t_0} is restricted to be less than w. See Knowles, Siegmund, and Zhang (1991) for a more complete discussion. We apply (41) to $Z_\theta = \langle\tilde{\gamma}(\theta), U^{(m)}\rangle$ conditional on $\langle\tilde{\gamma}(\theta_0), U^{(m)}\rangle = z$, $\langle\gamma_g, U^{(m)}\rangle = v$, and we consider only values $\theta > \theta_0$. Theorem 2 also involves $\theta < \theta_0$ which can be treated similarly.

Recall $\rho(\theta)$ defined in (19) and define orthonormal vectors

$$e_1 = \tilde{\gamma}(\theta_0), \qquad (42)$$
$$e_2 = (\tilde{\gamma}(\theta) - \rho(\theta)\tilde{\gamma}(\theta_0)/\sqrt{(1-\rho^2)}, \qquad (43)$$
$$e_3 = (\dot{\tilde{\gamma}} - \dot{\rho}e_1 + \rho\dot{\rho}e_2/\sqrt{(1-\rho^2)})/\sqrt{\{\|\dot{\tilde{\gamma}}\|^2 - \dot{\rho}^2/\sqrt{(1-\rho^2)}\}}. \qquad (44)$$

Since γ_g is orthogonal to $\tilde{\gamma}(\theta)$ for all θ and hence orthogonal to $\dot{\tilde{\gamma}}(\theta)$, let $e_4 = \gamma_g$. Let $u_i^{(m)} = \langle e_i, U^{(m)}\rangle$ $(i = 1, \cdots, 4)$. Then by (42)-(44)

$$\langle\tilde{\gamma}(\theta), U^{(m)}\rangle = \rho u_1^{(m)} + \sqrt{(1-\rho^2)} u_2^{(m)}, \qquad (45)$$
$$\langle\dot{\tilde{\gamma}}(\theta), U^{(m)}\rangle = \dot{\rho}(z - \rho w)/\sqrt{(1-\rho^2)} + \sqrt{\{\|\dot{\tilde{\gamma}}\|^2 - \dot{\rho}^2/\sqrt{(1-\rho^2)}\}} u_3^{(m)}. \qquad (46)$$

Conditional on $\langle\tilde{\gamma}(\theta_0), U^{(m)}\rangle = z$ and $\langle\gamma_g, U^{(m)}\rangle = v$, namely, $u_1^{(m)} = z$, $u_4^{(m)} = v$, the density of $u_2^{(m)}$ is

$$f_{m-3}(x/\sqrt{(1-z^2-v^2)})/\sqrt{(1-z^2-v^2)} \qquad (47)$$

and hence by (45)

$$P[\ \langle\tilde{\gamma}(\theta), U^{(m)}\rangle = w \mid \langle\tilde{\gamma}(\theta_0), U^{(m)}\rangle = z,\ \langle\gamma_g, U^{(m)}\rangle = v]$$
$$= \frac{1}{\sqrt{\{(1-\rho^2)(1-z^2-v^2)\}}} f_{m-3}\left(\frac{w - \rho z}{\sqrt{\{(1-\rho^2)(1-z^2-v^2)\}}}\right). \qquad (48)$$

Now, given $\langle\tilde{\gamma}(\theta_0), U^{(m)}\rangle = z$, $\langle\gamma_g, U^{(m)}\rangle = v$ and $\langle\tilde{\gamma}(\theta), U^{(m)}\rangle = w$, $u_3^{(m)}$ is the first coordinate of a point uniformly distributed on an $m-4$ dimensional sphere of radius

$$\surd\{1 - z^2 - v^2 - \frac{(w-\rho z)^2}{1-\rho^2}\}. \tag{49}$$

Hence by (46) and (49) the conditional density of $\langle\dot{\tilde{\gamma}}(\theta), U^{(m)}\rangle$ is given by

$$\tau^{-1} f_{m-4}(\frac{x-\mu}{\tau}),$$

where μ and τ are given in (20) and (21). By calculation we get

$$\begin{aligned} &E[\,\langle\dot{\tilde{\gamma}}(\theta), U^{(m)}\rangle^+ \big|\, \langle\tilde{\gamma}(\theta_0), U^{(m)}\rangle = z, \langle\gamma_g, U^{(m)}\rangle = v, \langle\tilde{\gamma}(\theta), U^{(m)}\rangle = w] \\ &= \mu F_{m-4}(\frac{\mu}{\tau}) + (m-3)^{-1}\tau f_{m-2}(\frac{\mu}{\tau}). \end{aligned} \tag{50}$$

Theorem 2 now follows by substitution of (48) and (50) into (41).

REFERENCES

BATES, D. M. and WATTS, D. G. (1980). Relative curvature measures of nonlinearity (with discussion) *J. Roy. Statist. Soc. B* **42**, 1–25.

BATES, D. M. and WATTS, D. G. (1988). *Nonlinear Regression Analysis and its Applications.* New York: Wiley.

DONALDSON, J. R. and SCHNABEL, R. B. (1987). Computational experience with confidence regions and confidence intervals for nonlinear least squares, *Technometrics* **29**, 67–82.

FEDER, P. I. (1975). On asymptotic distribution theory in segmented regression problems, *Ann. Statist.* **3**, 49–83.

GRIFFITHS, D. A., and MILLER, A. J. (1973). Hyperbolic regression – a model based on two-phase piecewise linear regression with smooth transition between regimes, *Commun. Statist.* **2**, 561–569.

HALPERIN, M. (1963). Confidence interval estimation in nonlinear regression, *J. Roy. Statist. Soc. B* **25**, 330–333.

HAMILTON, D. (1986). Confidence regions for parameter subset in nonlinear regression, *Biometrika* **74**, 423–425.

HINKLEY, D. (1971). Inference in two-phase regression, *J. Amer. Statist. Assoc.* **66**, 736–743.

JOHNSTONE, I., and SIEGMUND, D. (1989). On Hotelling's formula for the volume of tubes and Naiman's inequality, *Ann. Statist.* **17**, 184–194.

KNOWLES, M., SIEGMUND, D., and ZHANG, H. P. (1991). Confidence regions in semilinear regression, *Biometrika* **78**, 15–31.

LEHMANN, E. L. (1986). *Testing Statistical Hypotheses*, 2nd ed., John Wiley, New York.

POOL, J. and BORCHGREVINK, C. F. (1964) Comparison of rat liver response to coumarin administered in vivo versus in vitro, *Amer. J. Phys.* **206(1)** 229–238.

RATKOWSKY, D. A. (1983), *Nonlinear regression models.* New York: Marcel Dekker, Inc.

SEBER, G. A. F., and WILD, C. J. (1989), *Nonlinear regression.* New York: Wiley.

SIEGMUND, D., and ZHANG, H. P. (1993). The expected number of local maxima of a random field and the volume of tubes, to appear in *Ann. Statist.*

SMITH, A. F. M., and COOK, D. G. (1980). Straight lines with a change point: A Bayesian analysis of some renal transplant data, *Appl. Statist.* **29**, 180–189.

TISHLER, A., and ZANG, I. (1981). A new maximum likelihood algorithm for piecewise regression, *J. Amer. Statist. Assoc.* **76**, 980–987.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CA 94305

DEPARTMENT OF EPIDEMIOLOGY AND PUBLIC HEALTH
YALE UNIVERSITY SCHOOL OF MEDICINE
60 COLLEGE STREET
NEW HAVEN, CT 06510