Rastislav Potocký; Van Ban To

Confidence regions in nonlinear regression models

Applications of Mathematics, Vol. 37 (1992), No. 1, 29–39

Persistent URL: `http://dml.cz/dmlcz/104489`

CONFIDENCE REGIONS IN NONLINEAR REGRESSION MODELS

RASTISLAV POTOCKÝ, TO VAN BAN

(Received June 8, 1990)

Summary. New curvature measures for nonlinear regression models are developed and methods of their computing are given. Using these measures, more accurate confidence regions for parameters than those based on linear or quadratic approximations are obtained.

Keywords: Nonlinear regression, confidence regions, weak intrinsic and weak parameter-effects curvatures.

1. INTRODUCTION

Consider the usual nonlinear model

$$y_t = f(x_t, \Theta) + \varepsilon_t \qquad t = 1, \ldots, n \tag{1}$$

where $y = (y_1, \ldots, y_n)'$ is a vector of observations, $x_t = (x_{t1}, \ldots, x_{tk})$ are some control variables, $\Theta = (\Theta_1, \ldots, \Theta_p)'$ is a vector of unknown parameters and ε_t are independent normally distributed random variables, $\varepsilon_t \sim N(0, \sigma^2)$. The function f is supposed to have a known form, nonlinear in Θ.

For the model (1) the least squares estimates $\hat{\Theta}$ are the values of the parameters which minimize the sum of squares $S(\Theta) = \sum_{t=1}^{n} \big(y_t - f(x_t, \Theta)\big)^2$, or in vector notation $S(\Theta) = \|y - \eta(\Theta)\|^2$, where $\eta(\Theta) = \big(\eta_1(\Theta), \ldots, \eta_n(\Theta)\big)'$ and $\eta_t(\Theta) = f(x_t, \Theta)$.

Let $e = e(\Theta) = y - \eta(\Theta)$ be the error vector. As shown by Gallant [7], if Θ is the true value, then the $100(1-\alpha)\%$ exact confidence region for Θ includes all values satisfying $e(\Theta)'P(\Theta)e(\Theta) \leqslant \sigma^2\chi^2(p, \alpha)$ if σ^2 is known or $e(\Theta)'P(\Theta)e(\Theta) \leqslant ps^2F(p, \nu, \alpha)$ if σ^2 is estimated independently by s^2. (Here, P is the projection matrix for the column space of the $n \times p$ matrix of partial derivatives of η with respect to Θ.) Since it is difficult to display exact regions in the case of more than two parameters, methods of obtaining approximate confidence regions are of considerable value.

The approximate likelihood region for Θ (i.e. the confidence region based on likelihood ratio) is defined as the set of values Θ for which

$$S(\Theta) - S(\hat{\Theta}) \leqslant \delta^2 \tag{2}$$

where δ is a number not depending on Θ. Typically, one would put $\delta^2 = c^2\varphi^2$ where $\varphi = \sigma$ and $c = \sqrt{\chi^2(p,\alpha)}$ if σ^2 is known or $\varphi = \sqrt{p}\hat{\sigma}$ and $c = \sqrt{F(p,\nu,\alpha)}$ if σ^2 is estimated by $\hat{\sigma}^2$ based on ν *df*. (Here, $\chi^2(p,\alpha)$ and $F(p,\nu,\alpha)$ denote the critical values of χ^2—and F-distributions with p, and p and ν *df* and tail area probability α, respectively; φ is called the standard radius.) As shown by Beale and confirmed by Bates and Watts in [1], the right hand side of (2) should contain the so-called intrinsic curvature N_Φ which, however, can be neglected in practice.

Of course, the simplest approximation to (2) is the L-region

$$(\Theta - \hat{\Theta})' V_{.}' V_{.} (\Theta - \hat{\Theta}) \leqslant c^2\varphi^2 \tag{3}$$

where $V_{.} = \left\| \frac{\partial \eta}{\partial \Theta} | \hat{\Theta} \right\|$.

Box and Coutie [3] recommended another simple region, namely

$$(\Theta - \hat{\Theta})' \left\| \frac{1}{2} \frac{\partial^2 S}{\partial \Theta^2} | \hat{\Theta} \right\| (\Theta - \hat{\Theta}) \leqslant c^2\varphi^2, \tag{4}$$

as an approximate likelihood region for Θ (the so-called CL-region).

Having used a quadratic approximation to the solution locus instead of the usual linear approximation and curvature measures developed by Bates and Watts [1], Hamilton et al. [8] gave confidence regions corresponding to ellipsoids on the tangent plane at the least squares point which are more accurate than the regions (3) and (4). However, Cook and Witmer [6] showed examples of models for which the Bates-Watts methodology did not work.

The aim of this article is to develop criteria which enable us to find more accurate confidence regions than those developed by Bates and Watts [2] and Hamilton et al. [8]. For this reason the weak intrinsic, weak parameter-effects and total curvatures are introduced and methods of their computing are given. In the last part of the article we discuss advantages of our approach.

Let $\mathbf{A}^{\cdot}_{\cdot\cdot} = (A^i_{jk})$, $i = 1, \ldots, n$, $j = 1, \ldots, p$, $k = 1, \ldots, q$ be an $n \times p \times q$ array. Its i-th face $A^i_{\cdot\cdot} = (A^i_{jk})$ is a $p \times q$ matrix and its jk-th column $A^{\cdot}_{jk} = (A^1_{jk}, \ldots, A^n_{jk})'$ is a n-vector. The array $\mathbf{A}^{\cdot}_{\cdot\cdot}$ can be written as

$$\left\| \begin{matrix} A^{\cdot}_{11} & \ldots & A^{\cdot}_{1q} \\ & \ldots & \\ A^{\cdot}_{p1} & \ldots & A^{\cdot}_{pq} \end{matrix} \right\| \quad \text{or, equivalently, as} \quad \left\| \begin{matrix} A^1_{\cdot\cdot} \\ \vdots \\ A^n_{\cdot\cdot} \end{matrix} \right\| .$$

In what follows we consider the square bracket multiplication introduced by Bates and Watts [1]. If E is an $m \times n$ matrix and T is an $n \times p \times q$ array, then the elements of the i-th face M_i, $i = 1, \ldots, m$ of the $m \times p \times q$ array $\mathrm{M} = [E][\mathrm{T}]$ are $E_i T_{jk}$, $j = 1, \ldots, p$, $k = 1, \ldots, q$ where E_i is the i-th row of E and T_{jk} is the jk-th column of T.

Let E, F, G, H, J, U be $k \times n$, $l \times p, q \times k$, $n \times n$, $r \times k$ and $s \times n$ matrices, respectively. The following properties will be used throughout the paper:

a) $[JE][\mathbf{A}^{\cdot}_{\cdot\cdot}] = [J][[E][\mathbf{A}^{\cdot}_{\cdot\cdot}]]$
b) $[U][F\mathbf{A}^{\cdot}_{\cdot\cdot}G] = F[U][\mathbf{A}^{\cdot}_{\cdot\cdot}]G$ (Cook and Goldberg [5])
c) $f'(h'\mathbf{A}^{\cdot}_{\cdot\cdot}k) = h'[f'][\mathbf{A}^{\cdot}_{\cdot\cdot}]k$
d) $f'H(h'\mathbf{A}^{\cdot}_{\cdot\cdot}k) = f'(h'[H][\mathbf{A}^{\cdot}_{\cdot\cdot}]k)$
e) $[E][a\mathbf{A}^{\cdot}_{\cdot\cdot}] = a[E][\mathbf{A}^{\cdot}_{\cdot\cdot}]$
f) $F(a\mathbf{A}^{\cdot}_{\cdot\cdot})G = aF\mathbf{A}^{\cdot}_{\cdot\cdot}G$

for all $f \in \mathbf{R}^n$, $h \in \mathbf{R}^p$, $k \in \mathbf{R}^q$, $a \in \mathbf{R}^1$.

Let $V_{.}$ and $\mathbf{W}^{\cdot}_{\cdot\cdot}$ denote the $n \times p$ matrix and the $n \times p \times p$ array of the first and second derivatives of the model function $\eta(\Theta)$ with elements

$$V_{ij} = \frac{\partial \eta_i}{\partial \Theta_j}|\hat{\Theta} \text{ and } W^i_{jk} = \frac{\partial \eta_i}{\partial \Theta_j \Theta_k}|\hat{\Theta},$$

respectively.

We assume that the rank of $V_{.}$ is p and $V_{.} = UR$ is the unique orthogonal-triangular decomposition of $V_{.}$ ($Q\mathbf{R}$-decomposition), where the columns of the $n \times p$ matrix U form an orthogonal basis for $V_{.}$ and R is an upper triangular matrix with $\mathbf{R}_{ii} > 0$, $i = 1, \ldots, p$. Then the parameter-effects array for the parameter Θ at $\hat{\Theta}$ is defined as follows (see [1]):

$$\mathbf{A}^{T\cdot}_{\cdot\cdot} = [U'][L'\mathbf{W}^{\cdot}_{\cdot\cdot}L] \tag{5}$$

where $L = R^{-1}$.

Definition 1. By an $n \times p \times q \times r$ array we understand

$$\mathbf{A}^{\cdot}_{\cdot\cdot\cdot} = \left\| \begin{matrix} \mathbf{A}^1_{\cdot\cdot\cdot} \\ \vdots \\ \mathbf{A}^n_{\cdot\cdot\cdot} \end{matrix} \right\| .$$

where $\mathbf{A}^i_{\cdot\cdot\cdot}$, $i = 1, \ldots, n$ are $p \times q \times r$ arrays.

The premultiplication $Q\mathbf{A}^{\bullet}_{\dots}$, the postmultiplication $\mathbf{A}^{\bullet}_{\dots}Q$, the $*$-multiplication $Q*\mathbf{A}^{\bullet}_{\dots}$ and the square-bracket multiplication $[Q][\mathbf{A}^{\bullet}_{\dots}]$, where Q is a matrix, mean the summation over the second, third, first subscripts and the superscript, respectively.

Definition 2. The $p \times p \times p \times p$ array

$$\mathbf{A}^{T\cdot}_{\dots} = [U'][L' * (L'\mathbf{H}^{\bullet}_{\dots}L)] \tag{6}$$

where U, L are as in (5) and $H^{\bullet}_{\dots}$ is the $n \times p \times p \times p$ array of the third derivatives of the model function η with the elements

$$\eta^{i}_{jkl} = \frac{\partial \eta_i}{\partial \Theta_j \partial \Theta_k \partial \Theta_l}|\hat{\Theta}$$

is called the four dimensional parameter-effects array.

3. LIKELIHOOD REGIONS

Let $\mathsf{V}^{(4)}$ denote the array with the elements

$$\eta^{i}_{jklm} = \frac{\partial \eta_i}{\partial \Theta_j \partial \Theta_k \partial \Theta_l \partial \Theta_m}|\hat{\Theta}.$$

Using the Taylor-series expansion up to the term od degree 4 we obtain

$$e = y - \eta(\Theta) = \hat{e} - \left(V.\varphi + \frac{1}{2}\varphi'\mathbf{W}^{\bullet}_{..}\varphi + \frac{1}{6}\varphi' * (\varphi'\mathbf{H}^{\bullet}_{\dots}\varphi) + \frac{1}{24}\mathsf{V}^{(4)}\varphi^{(4)} + \dots \right) \tag{7}$$

where $\hat{e} = e(\hat{\Theta})$, $\varphi = \Theta - \hat{\Theta}$. Using $\hat{e}'V. = 0$ we get

$$\begin{aligned} e'e = \hat{e}'\hat{e} + \varphi'(V'_{.}V. - [\hat{e}'][\mathbf{W}^{\bullet}_{..}])\varphi + \left[(V.\varphi)'(\varphi'\mathbf{W}^{\bullet}_{..}\varphi) - \frac{1}{3}\hat{e}'(\varphi' * (\varphi'\mathbf{H}^{\bullet}_{\dots}\varphi))\right] \\ + \left[\frac{1}{3}\varphi'V'_{.}(\varphi' * (\varphi'\mathbf{H}^{\bullet}_{\dots}\varphi)) + \frac{1}{4}\|\varphi'\mathbf{W}^{\bullet}_{..}\varphi\|^2 - \frac{1}{12}\hat{e}'\mathsf{V}^{(4)}\varphi^{(4)}\right] + \dots \end{aligned} \tag{8}$$

In the rest of the article it is assumed that

1) $\eta(\Theta)$ is a continuous function in Θ with finite derivatives up to and including degree 4 over the whole parameter space Ω.
2) The vector $\partial S(\Theta)/\partial\Theta$ vanishes only at the one point $\hat{\Theta} \in \Omega$.
3) $\hat{e}'(\varphi * (\varphi'\mathbf{H}^{\bullet}_{\dots}\varphi))$, $\hat{e}'\mathsf{V}^{(4)}\varphi^{(4)}$ and the Beale's nonlinearity N_{Φ} (see [1]) can be neglected.

From (2) and (8) we obtain

Lemma 1. *Under the above assumptions the approximate likelihood region is*

$$\begin{aligned}&\varphi'(V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}])\varphi + \varphi' V_{\cdot}'(\varphi'\mathbf{W}^{\cdot}_{\cdot\cdot}\varphi)\\ &+ \left[\frac{1}{3}\varphi' V_{\cdot}'(\varphi' * (\varphi'\mathbf{H}^{\cdot}_{\cdot\cdot\cdot}\varphi)) + \frac{1}{4}\|\varphi'\mathbf{W}^{T\cdot}_{\cdot\cdot}\varphi\|^2\right] + \ldots \leqslant c^2\rho^2\end{aligned} \tag{9}$$

where the neclected terms are of degree 5 and higher in φ *and* $\mathbf{W}^{T\cdot}_{\cdot\cdot} = [P^T][\mathbf{W}^{\cdot}_{\cdot\cdot}] = [V_{\cdot}(V_{\cdot}'V_{\cdot})^{-1}V_{\cdot}'][\mathbf{W}^{\cdot}_{\cdot\cdot}]$.

Proof. $\|\varphi,\mathbf{W}^{\cdot}_{\cdot\cdot}\varphi\|^2 = \|\varphi'\mathbf{W}^{T\cdot}_{\cdot\cdot}\varphi\|^2 + \left\|\varphi'[I - P^T][\mathbf{W}^{\cdot}_{\cdot\cdot}]\varphi\right\|^2 = \|\varphi'\mathbf{W}^{T\cdot}_{\cdot\cdot}\varphi\|^2$, because $N_{\clubsuit} = 0$ implies $\|\varphi'\mathbf{A}^{N\cdot}_{\cdot\cdot}\varphi\| = 0$, where $\mathbf{A}^{N\cdot}_{\cdot\cdot}$ is the intrinsic curvature array (see [1], (2.30)) and consequently $\left\|\varphi'[I - P^T][\mathbf{W}^{\cdot}_{\cdot\cdot}]\varphi\right\| = 0$. □

Put $\varphi^* = \lambda h$, $\varphi_1 = \lambda_1 h$, $\varphi_2 = \lambda_2 h$ where h is a unit vector in $\mathbf{R}^p$ and $\lambda, \lambda_1, \lambda_2 \geqslant 0$ are such that φ^*, φ_1 and φ_2 are boundary points of the regions (9), (3) and (4), respectively. Evidently, $\lambda_1 = (\lambda_1)_h = c\rho\|V_{\cdot}h\|^{-1}$ and $\lambda_2 = (\lambda_2)_h = c\rho\big(h'(V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}])h\big)^{-1/2}$, because $\left\|\frac{1}{2}\frac{\partial^2 S}{\partial\Theta^2}|\hat{\Theta}\right\| = V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}]$. Inserting $\varphi^*\lambda h$ into (9) we have

$$\begin{aligned}&\lambda^2 h'(V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}])h + \lambda^3 h' V_{\cdot}'(h'\mathbf{W}^{\cdot}_{\cdot\cdot}h)\\ &+ \lambda^4\left[\frac{1}{3}h' V_{\cdot}'(h' * (h'\mathbf{H}^{\cdot}_{\cdot\cdot\cdot}h)) + \frac{1}{4}\|h\mathbf{W}^{T\cdot}_{\cdot\cdot}h\|^2\right] + \ldots = c^2\rho^2.\end{aligned} \tag{10}$$

If we now "invert" (10), expressing λ as a power series in $c\rho$ we get

$$\begin{aligned}\lambda = \lambda_h = (\lambda_1)_h \Bigg\{ &\left[\frac{h'(V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}])h}{h'V_{\cdot}'V_{\cdot}h}\right]^{-\frac{1}{2}} - \frac{c\rho}{2}\frac{\|V_{\cdot}h\| h' V_{\cdot}'(h'\mathbf{W}^{\cdot}_{\cdot\cdot}h)}{\left[h'(V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}])h\right]^2}\\ &+ \frac{c^2\rho^2\|V_{\cdot}h\|}{8\left[h'(V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}])h\right]^{7/2}}\Big[5\big(h'V_{\cdot}'(h'\mathbf{W}^{\cdot}_{\cdot\cdot}h)\big)^2 - 4h'(V_{\cdot}'V_{\cdot} - [\hat{e}'][\mathbf{W}^{\cdot}_{\cdot\cdot}])\\ &\times h\left(\frac{1}{3}h'V_{\cdot}'(h' * (h'\mathbf{H}^{\cdot}_{\cdot\cdot\cdot}h)) + \frac{1}{4}\|h'\mathbf{W}^{T\cdot}_{\cdot\cdot}h\|^2\right)\Big] + \ldots\Bigg\}.\end{aligned} \tag{11}$$

The second factor on the right-hand side of (11), being $\lambda_h/(\lambda_1)_h = \left(\frac{\lambda}{\lambda_1}\right)_{h'}$ is the "radius" ratio that does not depend on the length of h.

Theorem 1. *Let* d *be a unit vector in* $\mathbf{R}^p$. *Then*

$$\lambda_{Ld} = (\lambda_1)_{Ld}\left\{1 + \alpha(d) - \frac{c\rho}{2}\Gamma(d) + \frac{c^2\rho^2}{2}\beta(d) + \ldots\right\} \tag{12}$$

where

$$\alpha(d) = (1 - d'Bd)^{-1/2} - 1,$$
$$B = L'[\hat{e}'][\mathbf{W}^{\boldsymbol{\cdot}}_{\cdot\cdot}]L,$$
$$\Gamma(d) = d'(d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d)/(1 - d'Bd)^2,$$
$$\beta(d) = \frac{1}{4}(1 - d'Bd)^{-7/2}\Big[5\big(d'(d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d)\big)^2 - 4(1 - d'Bd)$$
$$(13) \qquad \times\left(\frac{1}{3}d'(d' * (d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d')) + \frac{1}{4}\|d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d\|^2\right)\Big], \quad (\lambda_1)_{Ld} = c\rho\|Ld\|.$$

Proof. Using the fact that $V.Ld = Ud$ and $\|V.Ld\| = 1$, we obtain

$$(14) \qquad [(V.Ld)']\,[(Ld)' * ((Ld)'\mathbf{H}^{\boldsymbol{\cdot}}_{\cdot\cdot\cdot}Ld)] = d'\big(d' * (d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d)\big).$$

Then inserting $h = Ld$ into (11) we get (12). □

Definition 3. $\Delta^N = \max_{\|d\|=1}|\alpha(d)|$, $\boldsymbol{\Delta}^T = \max_{\|d\|=1}\frac{c\rho}{2}|\Gamma(d)|$, $\Delta = \max_{\|d\|=1}\left|\alpha(d) - \frac{c\varphi}{2}\boldsymbol{\Gamma}(d)\right|$ are called the weak intrinsic, weak parameter-effects and total curvatures, respectively.

There is a close relationship between the Bates-Watts curvatures and those proposed by us. Definition 3 says that $\boldsymbol{\Delta} = \max_{\|d\|=1}|\alpha(d)|$, $\alpha(d) = (1 - d'Bd)^{-1/2}$ where $B = L'[\hat{e}'][\mathbf{W}^{\boldsymbol{\cdot}}_{\cdot\cdot}]L$. Using the properties of the square bracket multiplication we obtain $B = [\hat{e}'][L'\mathbf{W}^{\boldsymbol{\cdot}}_{\cdot\cdot}L]$. For the residual vector $\hat{e}$ we have

$$\hat{e} = P_N\hat{e} = \big(N(N'N)^{-1}N'\big)\hat{e} = NN'\hat{e}$$

where the columns of the $n \times (n-p)$ matrix N form an orthogonal basis for the space orthogonal to the tangent plane. Hence

$$B = [\hat{e}'][L'\mathbf{W}^{\boldsymbol{\cdot}}_{\cdot\cdot}L] = [(NN'\hat{e})'][L'\mathbf{W}^{\boldsymbol{\cdot}}_{\cdot\cdot}L]$$
$$= [\hat{e}'N]\big[L'[N'][\mathbf{W}^{\boldsymbol{\cdot}}_{\cdot\cdot}]L\big] = [\hat{e}'N][\mathbf{A}^{N\cdot}_{\cdot\cdot}],$$

i.e. B is the $p \times p$ matrix obtained from the square bracket multiplication of the rotated residual vector $N'\hat{e}$ and the intrinsic curvature array $\mathbf{A}^{N\cdot}_{\cdot\cdot}$. It follows that $\boldsymbol{\Delta}^N$ is a function of the intrinsic curvature array proposed by Bates and Watts. The same argument can be used in the case of $\boldsymbol{\Delta}^T$ and $\boldsymbol{\Delta}$.

Moreover, an inequality holds between $\boldsymbol{\Delta}^T$ and the Bates-Wats maximum intrinsic curvature Γ^T (analogously for $\boldsymbol{\Delta}^N$ and and $\boldsymbol{\Gamma}^N$). We have $\boldsymbol{\Gamma}^T = \max_{\|d\|=1}\sqrt{p}\hat{\sigma}\|d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d\|$ (see e.g. [1], [5]). On the other hand,

$$\boldsymbol{\Delta}^T = \max_{\|d\|=1}\frac{c\rho}{1}|\Gamma(d)| = \frac{c\sqrt{p}\,\hat{\sigma}}{2}\max_{\|d\|=1}\frac{|(d'(d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d))|}{(1 - d'Bd))^2}$$
$$\leqslant \frac{c\sqrt{p}\,\hat{\sigma}}{2}\max_{\|d\|=1}\|d'\mathbf{A}^{T\cdot}_{\cdot\cdot}d\| = \frac{\sqrt{F(p, n-p, \alpha)}}{2}\boldsymbol{\Gamma}^T,$$

since $1 - d'Bd \doteq 1$. Consequently, $\boldsymbol{\Delta}^T \leqslant \frac{1}{2}\sqrt{F(p, n-p, \alpha)}\boldsymbol{\Gamma}^T$.

If we may neglect the term $\frac{c^2\rho^2}{2}\beta(d)$ in (11) then

$$\left|\left(\frac{\lambda}{\lambda_1}\right)_h - 1\right| \leqslant \boldsymbol{\Delta} \leqslant \boldsymbol{\Delta}^N + \boldsymbol{\Delta}^T \tag{15}$$

and

$$\left|\left(\frac{\lambda}{\lambda_2}\right)_h - 1\right| \leqslant \boldsymbol{\Delta}^T \qquad \forall h \in \mathbf{R}^p. \tag{16}$$

Inserting the different values od d into (12) we obtain the bounds of the approximate likelihood region.

4. Computing Nonlinearities

Evidently $\boldsymbol{\Delta}^N = \max_i \left|1 - (1-\lambda_i)^{-1/2}\right| \approx \max \frac{1}{2}|\lambda_i|$, where λ_i are the eigenvalues of the symmetric matrix $B = L'[\hat{e}'][\mathbf{W}^{\cdot}_{..}]L = L'[\hat{e}_1\mathbf{W}^1_{..} + \ldots + \hat{e}_n\mathbf{W}^n_{..}]L$. For finding likelihood regions we suggest that $\boldsymbol{\Delta}^N$ can be neglected if it is not greater then 0.03 (cf. [1]).

For calculating $\boldsymbol{\Delta}^T$ we note that

$$\boldsymbol{\Delta}^{T\cdot} \max_{\|h\|=1} \frac{c\rho}{1} \frac{h'(h'\mathbf{A}^{T\cdot}_{..}h)}{(1-h'Bh)^2} \approx \frac{c\rho}{2}\frac{\Lambda}{(1-\hat{h}'B\hat{h})^2},$$

where $\hat{h}$ maximizes $|\Lambda(h)|$, $\Lambda(h) = h'(h'\mathbf{A}^{T\cdot}_{..}h)$, $\Lambda = |\Lambda(\hat{h})|$. Since $\|h\| = 1$, $\Lambda(h)$ gives a local extremum at a point h^* if h^* has the same direction as the gradient $\nabla\Lambda(h)$. We have

$$\nabla\Lambda(h) = \left(\frac{\partial\Lambda(h)}{\partial_h}\right)' = h'\mathbf{A}^{T\cdot}_{..}h + 2h'\begin{Vmatrix} h' & (\mathbf{A}^{T\cdot}_{..})^1 \\ & \vdots \\ h' & (\mathbf{A}^{T\cdot}_{..})^p \end{Vmatrix} = h'\mathbf{C}^{\cdot}_{..}h$$

where $\mathbf{C}^{\cdot}_{..} = (c^k_{ij})$, with $c^k_{ij} = (A^T_{..})^k_{ij} + 2(A^T_{..})^i_{jk}$, $i, j, k = 1, \ldots, p$. Note that $\Lambda(h)$ is an odd function of h. Hence the algorithm for finding h^* can be described as follows:

(1) choose an initial direction h_i, $\|h_i\| = 1$;

(2) calculate $g_i = \nabla\Lambda(h_i)$ and $\tilde{g}_i = g_i/\|g_i\|$;

(3) If $\tilde{g}_i'h_i \leqslant 0.9999$ then set $h_{i+1} = (3\tilde{g}_i + h_i)/\|3\tilde{g}_i + h_i\|$ and repeat (2), otherwise $\Lambda = |\tilde{g}_i'(\tilde{g}_i'\mathbf{A}^{T\cdot}_{..}\tilde{g}_i)|$ (cf. [1]).

In practice we choose $h_i = (1, 0, \ldots, 0)'$, …, $(0, \ldots, 0, 1)'$ $\left(\frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}, 0, \ldots, 0\right)$, …, $\left(\frac{1}{\sqrt{2}}, \ldots, \pm\frac{1}{\sqrt{2}}\right)'$, …, $\left(0, \ldots, \frac{1}{\sqrt{2}}, \pm\frac{1}{\sqrt{2}}\right)$ and compare the results obtained. We claim that we have found the global maximum. Then $\boldsymbol{\Delta}^T = \frac{c\rho}{2}\frac{\Lambda}{(1-\tilde{g}_i'B\tilde{g}_i)^2}$.

For calculating Δ we note that $\Delta \approx \Delta^T$ if $\Delta^N \leqslant 0.03$. When $\Delta^N > 0.03$, then

$$\Delta \approx \max_{\|h\|=1} \left\| \frac{1}{2} h' B h - \frac{cp}{2} \frac{h'(h' \mathbf{A}^{T}_{\cdot\cdot} \cdot h)}{(1 - h'Bh)^2} \right\|$$
$$\approx \left\| \frac{1}{2} h^{*\prime} B h^{*} - \frac{cp}{2} \frac{h^{*\prime}(h^{*} \mathbf{A}^{T}_{\cdot\cdot} \cdot h^{*})}{(1 - h^{*\prime}Bh^{*})^2} \right\|,$$

where h^* maximizes $\left|\frac{1}{2}h'Bh - \frac{cp}{2}h'(h'\mathbf{A}^{T}_{\cdot\cdot}\cdot h)\right|$. We note that $h'Bh$ is an even function in h and $\Sigma(h) = \frac{1}{2}h'Bh - \frac{cp}{2}h'(h'\mathbf{A}^{T}_{\cdot\cdot}\cdot h)$ reaches its maximum (minimum) at h^* for which $h^{*\prime}(h^{*\prime}\mathbf{A}^{T}_{\cdot\cdot}\cdot h^*) \leqslant 0$ ($\geqslant 0$). The gradient of $\Sigma(h)$ is $\nabla\Sigma(h) = \left(\frac{\partial \Sigma(h)}{\partial h}\right)' = (h'B)' - \frac{cp}{2}h'\mathbf{C}^{\cdot}_{\cdot\cdot}h$.

The algorithm for getting $\Sigma_1 = \max_{\|h\|=1} \Sigma(h)$ can be described as follows:

(1) Choose an initial direction $h^*{}_i$, $\|h^*{}_i\| = 1$;
(2) calculate $h_i = -h^*{}_i \operatorname{sgn}\left((h^*{}_i)'(h^{*\prime}_i\mathbf{A}^{T}_{\cdot\cdot}\cdot h^*{}_i)\right)$;
(3) calculate $g_i - \nabla\Sigma(h_i)$ and $\tilde{g}_i = g_i/\|g_i\|$
(4) if $\tilde{g}_i h_i \leqslant 0.9999$, then set $h^*{}_{i+1} = (3\tilde{g}_i + h_i)/\|3\tilde{g}_i + h_i\|$ and repeat (2), otherwise

$$\Sigma_1 = \frac{1}{2}h'_i B h_i - \frac{cp}{2}\frac{h'_i(h'_i\mathbf{A}^{T}_{\cdot\cdot}\cdot h_i)}{(1 - h_i B h_i)^2}.$$

The initial directions are chosen as above and the results obtained are compared. The algorithm for calculating $\Sigma_2 = \min \Sigma(h)$ is similar but in the second step we put $h_i = h^*{}_i \operatorname{sgn}\left(h^*{}_i(h^{*\prime}_i\mathbf{A}^{T}_{\cdot\cdot}\cdot h^*{}_i)\right)$. Then $\Sigma = \max(|\Sigma_1|, |\Sigma_2|)$.

5. Assessing the significance of curvatures

We suggest the following rule for the choice of the appropriate confidence region (cf. [4]).

1) If $\Delta \leqslant 0.15$, curvature effects may be ignored and the L-region (3) accepted.
2) If $0.15 < \Delta \leqslant 1/3$, the L-region (3) is accepted for rough analysis only. If further $\Delta^T \leqslant 0.15$, the CL-region (4) should be used.
3) If $0.15 < \Delta \leqslant 1/3$, $0.15 < \Delta^T$, the approximate region (12) is recommended.
4) If $1/3 < \Delta$, a suitable reparametrization should be used. Our experience has shown that we ought to calculate $\Delta^* = \left|\alpha(\hat{d}) - \frac{cp}{2}\Gamma(\hat{d}) + \frac{c^2p^2}{2}\beta(\hat{d})\right|$, where $\hat{d}$ maximizes $\left|\alpha(\hat{d}) - \frac{cp}{2}\Gamma(\hat{d})\right|$, in order to obtain more information about the model.

Example. The Fieller-Creasy problem. Let

$$\eta_1(\Theta) = \ldots = \eta_n(\Theta) = \Theta_1$$
$$\eta_{n+1}(\Theta) = \ldots = \eta_{2n}(\Theta) = \Theta_1\Theta_2$$

and suppose that σ^2 is known, $\sigma^2/n = 1$, $c^2 = \chi^2\ (2; 0.05) = 6$. It can be shown that $B = 0$, $\Delta^N = 0$, $\mathbf{H}^{\bullet}_{\dots} = 0$ and $(\lambda_1)_{Ld} = \frac{\sqrt{6}}{\hat{\Theta}_1\sqrt{1+\hat{\Theta}_2^2}}\left(\hat{\Theta}_1^2(d_1 - \hat{\Theta}_2 d_2)^2 + (1+\hat{\Theta}_2^2)^2 d_2^2\right)^{1/2}$, $\alpha(d) = 0$,

$$-\frac{c\rho}{2}\Gamma(d) = \sqrt{6}\ d_2 \frac{\hat{\Theta}_2 d_1^2 + (1-\hat{\Theta}_2^2)d_1 d_2 - \hat{\Theta}_2 d_2^2}{\hat{\Theta}_1\sqrt{1+\hat{\Theta}_2^2}}$$

$$\frac{c^2\rho^2}{2}\beta(d) = \frac{15}{4}\left(\rho\Gamma(d)\right)^2 - 3\left(\frac{d_2(d_1 - \hat{\Theta}_2 d_2)}{\hat{\Theta}_1}\right)^2 .$$

The results for different values of $(\hat{\Theta}_1, \hat{\Theta}_2)$ are listed in Table 1 (cf. [6]).

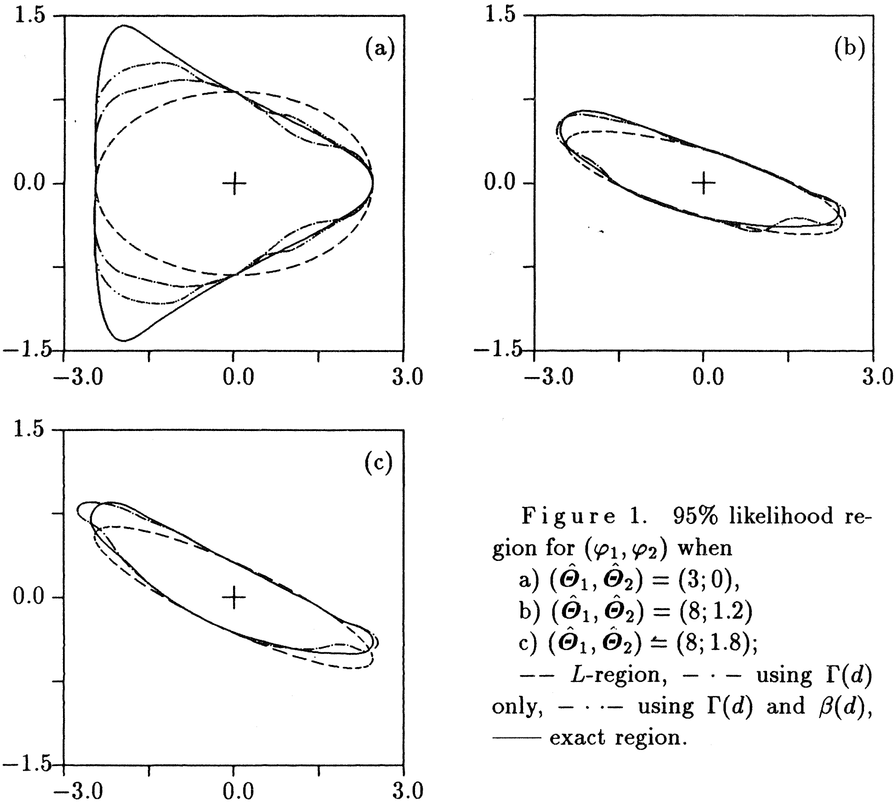
TABLE 1
Curvatures in the Fieller-Creasy model

$\hat{\Theta}_1$	$\hat{\Theta}_2$	Γ^T	c^{-1}	Δ^T	Δ^*
3	0	0.33	0.41	0.314	0.485
8	1.2	0.35	0.41	0.238	0.324
8	1.8	0.48	0.41	0.305	0.431
8	2.4	0.63	0.41	0.374	0.559
8	3.0	0.77	0.41	0.440	0.712

In the first case the value Γ^T (the Bates-Watts parameter-effect curvature) is smaller than the cutoff c^{-1} but Δ^T is large (> 0.25). By the rule, the L-region is not accepted which agrees with the findings in [6]. Since $\Delta^T < 1/3$, we construct the region (12). Δ^* is smaller than 0.5 so we conclude that (12) is acceptable.

In the second case Δ^T is large but not seriously: $0.15 < \Delta^T < 0.25$, Δ^* is relatively small (< 0.35). We hope that there is no big mistake in using the L-region. The L-region in the third case differs from the corresponding exact region. The region (12), however, agrees with it.

In the last two cases we have $\Delta^T > 1/3$ and $\Delta^* > 0.5$. Neither the L-regions nor the regions (12) should be used. This, again, agrees with the findings in [6].



F i g u r e 1. 95% likelihood region for (φ_1, φ_2) when
a) $(\hat{\Theta}_1, \hat{\Theta}_2) = (3; 0)$,
b) $(\hat{\Theta}_1, \hat{\Theta}_2) = (8; 1.2)$
c) $(\hat{\Theta}_1, \hat{\Theta}_2) \doteq (8; 1.8)$;
—— L-region, — · — using $\Gamma(d)$ only, — · ·— using $\Gamma(d)$ and $\beta(d)$, —— exact region.

6. Discussion

Hamilton et al. [8] have constructed simple confidence regions, but not for the original parameter Θ. Instead, their results concern a parameter τ which is a nonlinear function of Θ.

Clarke [4] presented methods of constructing regions with higher precision. However, his investigations deal with a single parameter Θ_i, not with the whole vector Θ.

The advantage of our approach is that it leads to confidence regions for the original parameter Θ. Moreover, our results agree with those obtained by Clarke [4] for $p = 1$.

It should also be pointed out that the total curvature introduced in Definition 3 has to be taken into account in order to obtain more accurate confidence regions for parameters in nonlinear regression models.

References

[1] *D. M. Bates, D. G. Watts*: Relative curvature measures of nonlinearity, Journal of the Royal Statistical Society, Ser. B *42* (1980), 1–25.

[2] *D. M. Bates, D. G. Watts*: Parameter transformations for improved approximate confidence regions in nonlinear least squares, The Annals of Statistics *9* (1981), 1152–1167.

[3] *G. E. P. Box, G. A. Coutie*: Application of digital computers in the exploration of functional relationships, Proc. I.E.E. *103* no. 1, part B, suppl. (1956), 100–107.

[4] *G. P. I. Clarke*: Marginal curvatures and their usefulness in the analysis of nonlinear regression models, Journal of the American Statistical Association *82* (1987), 844–850.

[5] *R. D. Cook, M. L. Goldberg*: Curvatures for parameter subsets in nonlinear regression, The Annals of Statistics *14* (1986), 1399–1418.

[6] *R. D. Cook, J. A. Witmer*: A note on parameter–effects curvature, Journal of the American Statistical Association *80* (1985), 872–878.

[7] *A. R. Gallant*: Nonlinear regression, The American Statistician *29* (1975), 73–81.

[8] *D. C. Hamilton, D. G. Watts, D. M. Bates*: Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions, The Annals of Statistics *10* (1982), 386–393.

Súhrn

OBLASTI SPOĽAHLIVOSTI PRE NELINEÁRNE REGRESNÉ MODELY

RASTISLAV POTOCKÝ, TO VAN BAN

V článku sú uvedené algoritmy na výpočet novonavrhnutých mier zakrivenia pre nelineárne modely. Pomocou týchto mier možno získať presnejšie oblasti spoľahlivosti pre parametre modelu v porovnaní s tými, ktoré sú založené na lineárnej alebo kvadratickej aproximácii modelovej funkcie.

Author's address: Doc. RNDr. Rastislav Potocký CSc., RNDr. To Van Ban, MFF UK, Mlynská dolina, 842 15 Bratislava.