Tatra Mt. Math. Publ. **26** (2003), 227–236

TATRA MOUNTAINS
Mathematical Publications

APPROXIMATE CONFIDENCE REGIONS FOR PARAMETER SUBSETS IN A NON-LINEAR MODEL

Rastislav Potocký

ABSTRACT. Using a cubic approximation to the solution locus instead of the usual linear or quadratic approximations, improved approximate confidence regions for parameter subsets in nonlinear regression models are derived. Examples are given which show that these regions have true coverage closer to the nominal level than regions currently used.

1. Introduction

Consider the usual nonlinear regression model

$$y_a = f(x_a, \theta) + e_a\,, \qquad a = 1, \dots, n\,, \tag{1}$$

$Y = (y_1, \dots y_n)^T$ denotes the vector of observations, the function $f(\theta) = \big(f(x_1, \theta), \dots, f(x_n, \theta)\big)$ has a known form dependent on p unknown parameters, $\theta = (\theta_1, \dots \theta_p)^T$, x_a are known vector-valued variables, e_a are independent and normally distributed random errors with zero mean and variance σ^2.

The distribution of the least squares estimator of θ has been obtained by Pázman in [5]. The problem of finding acceptable confidence regions for θ has been discussed by many authors (see, e.g., [1], [6]) but is not yet satisfactorily solved. An interesting approach can be found in Kubáček [4].

Suppose that θ can be partitioned as $\theta = \big(\theta_{(1)}, \theta_{(2)}\big)$ where $\theta_{(1)}$ consists of the first p_1 elements and $\theta_{(2)}$ of the last p_2 elements of θ; $p_1 + p_2 = p$ where only $\theta_{(2)}$ is of primary interest. The aim of the paper is to give an approximate confidence region for $\theta_{(2)}$ (or a vector function of $\theta_{(2)}$). Our method, using a cubic approximation to the solution locus is based on the four-dimensional curvature array which provides (along with the Bates-Watts array) all information necessary to construct regions with higher precision. As confirmed by examples

2000 Mathematics Subject Classification: Primary 62J02; Secondary 62F25.
Keywords: likelihood regions parameter-effect array.

This work was supported by the grant VEGA no. 1/7295/20.

given in the second part of the paper, these regions are more accurate than those commonly used. Moreover a criterion is developed to distinguish cases when the linear approximation region may be used for both those in which an improved confidence region is preferable and those where a parameter transformation is necessary. It is shown that this criterion is more reliable than that using the Bates-Watts cutoff.

2. Confidence regions

Let $S(\theta)$ be the usual sum of squares corresponding to the model (1) and $\theta^\star = (\theta^\star_{(1)}, \theta^\star_{(2)})$ be the least squares value of θ. Let $g(\theta_{(2)}) = (g_1(\theta_{(2)}), \dots g_{p_1}(\theta_{(2)}))$ be the value of $\theta_{(1)}$ which minimizes $S(\theta)$ for given $\theta_{(2)}$. Put $\alpha(\theta_{(2)}) = (g(\theta_{(2)}), \theta_{(2)})$ and $h(\theta_{(2)}) = f(\alpha(\theta_{(2)}))$. As discussed in [3], the approximate confidence region based on the likelihood ratio (the likelihood region for short) is such a set of values $\theta_{(2)}$ for which

$$\big(Y - h(\theta_{(2)})\big)^T \big(Y - h(\theta_{(2)})\big) - S(\theta^\star) \le c^2 \rho^2, \tag{2}$$

where $\rho = \sigma$ and $c^2 = \chi^2(p_2, \delta)$ if σ is known or $\rho^2 = p_2 s^2$ and $c^2 = F(p_2, n - p, \delta)$ if σ^2 is estimated by $s^2 = S(\theta^\star)/(n-p)$. Here $\chi^2(p_2, \delta)$ and $F(p_2, n-p, \delta)$ are the upper δ-probability points of χ^2 and F-distribution, respectively, ρ is the standard radius. Let $V, V.., V...$ be the first, second and third derivatives of $f(\theta)$ with elements V_i^a, V_{ij}^a, V_{ijk}^a. Suppose that the rank of V is p, and $V = UR$ is the orthogonal triangular decomposition of V. Then the three-dimensional parameter-effects array (the Bates-Watts curvature array) is defined as $A.. = [U^T][L^T V.. L]$ and the four-dimensional parameter-effects array is defined as $A... = [U^T]\big[L^T \bullet (L^T V...\ L)\big]$, where $L = R^{-1}$ and the square bracket and $\bullet$ multiplications of an array by a matrix are defined in [6]. Let the derivative $\frac{d\alpha}{d\theta_{(2)}}\big|\theta^\star_{(2)}$ be identified with the matrix Δ. The corresponding arrays of the second and third derivatives of α with respect to $\theta_{(2)}$ evaluated in $\theta^\star_{(2)}$ will be denoted by $\Delta..$ and $\Delta...$, respectively. Then the cubic approximation of $h(\theta_{(2)})$ about $\theta^\star_{(2)}$ can be written as

$$\begin{aligned} h(\theta_{(2)}) = {} & f(\theta^\star) + V\Delta\Phi_2 + 1/2\{\Phi_2^T\Delta^T V..\,\Delta\Phi_2 + V(\Phi_2^T\Delta..\Phi_2)\} \\ & + 1/6\{\Phi_2^T\Delta^T \bullet (\Phi_2^T\Delta^T V...\,\Delta\Phi_2) \\ & + 3\Phi_2^T\Delta^T V..(\Phi_2^T\Delta..\Phi_2) + V(\Phi_2^T \bullet (\Phi_2^T\Delta...\Phi_2))\} + \dots, \end{aligned} \tag{3}$$

where $\Phi_2 = \theta_{(2)} - \theta^\star_{(2)}$.

Partition V as $V = (V_1, V_2)$, $U = (U_1, U_2)$, where V_i and U_i are $n \times p_i$ matrices, $i = 1, 2$. Put $V^\circ.. = L^T V..\,L$ and $V^\circ... = L^T \bullet (L^T V...\ L)$. Let

$\vartheta_2 = R_{22}\Phi_2$, R_{22} being the lower part of R. Finally, partition $V^{\circ}_{..}$ and $V^{\circ}_{...}$ are as follows

$$V^{\circ}_{..} = \begin{pmatrix} V^{\circ}_{11} & V^{\circ}_{12} \\ V^{\circ}_{21} & V^{\circ}_{22} \end{pmatrix}, \quad V^{\circ}_{...} = \begin{pmatrix} V^{\circ}_{111} & V^{\circ}_{112} \\ V^{\circ}_{121} & V^{\circ}_{122} \\ V^{\circ}_{211} & V^{\circ}_{212} \\ V^{\circ}_{221} & V^{\circ}_{222} \end{pmatrix}.$$

Using this, the expansion of $h(\theta_{(2)})$ can be rewritten as

$$\begin{aligned} h(\theta_{(2)}) = f(\theta^{\star}) + U_2\vartheta_2 &+ 1/2\{\vartheta_2^T[P_{U_2}]\,[V^{\circ}_{22}]\vartheta_2 - U_1[U_2\vartheta_2]^T[V^{\circ}_{12}]\vartheta_2\} \\ &+1/6\Big\{(\vartheta_2^T \bullet (\vartheta_2^T[P_{U_2}]\,[V^{\circ}_{222}]\vartheta_2) - 3\vartheta_2^T[P_{U_2}]\,[V^{\circ}_{21}] \\ &\qquad \times (U_1^T(\vartheta_2^T V^{\circ}_{22}\vartheta_2) + 2[U_2\vartheta_2]^T[V^{\circ}_{12}]\vartheta_2)) \\ &\qquad - 3(N[(\Phi_2^T\Delta^T V_{..}\Phi_2 + V(\Phi_2^T\Delta_{..}\Phi_2))^T][V_{..}]\Delta\Phi_2 \,(4) \\ &\qquad + N \bullet ((U_2\vartheta_2)^T[\Phi_2^T\Delta^T V_{..}]\,[V_{...}]\Delta\Phi_2) \\ &\qquad + N[\Phi_2^T\Delta^T V^T][V_{..}](\Phi_2^T\Delta_{..}\Phi_2))\Big\} + \dots, \end{aligned}$$

where the neglected terms are of degree 4 and higher in Φ_2, P_{U_2} is the orthogonal projection. The exact form of N is omitted as it is not needed in what follows. Using the first and second terms of h only we obtain the simplest approximate likelihood region, namely

$$\Phi_2^T R_{22} R_{22}\Phi_2 \le c^2\rho^2$$

(the so-called L-region). Using (4) once more we obtain

$$\begin{aligned} &(h(\theta_{(2)}) - h(\theta^{\star}_{(2)}))^T(h(\theta_{(2)}) - h(\theta^{\star}_{(2)})) \\ &\quad = (U_2\vartheta_2)^T(U_2\vartheta_2) \\ &\qquad + (U_2\vartheta_2)^T(\vartheta_2 T[P_{U_2}]\,[V^{\circ}_{22}]\vartheta_2) \\ &\qquad + 1/3(U_2\vartheta_2)^T\{\vartheta_2^T \bullet (\vartheta_2^T[P_{U_2}]\,[V^{\circ}_{222}]\vartheta_2) - 3\vartheta_2^T[P_{U_2}]\,[V^{\circ}_{21}](U_1^T(\vartheta_2^T V^{\circ}_{22}\vartheta_2) \\ &\qquad\qquad + 2[U_2\vartheta_2]^T[V^{\circ}_{12}]\vartheta_2)\} \\ &\qquad + 1/4\|\vartheta_2^T[U_2^T]\,[V^{\circ}_{22}]\vartheta_2\|^2 + \|[U_2\vartheta_2]^T[V^{\circ}_{12}]\vartheta_2\|^2 + \dots \\ &\quad = \vartheta_2^T\vartheta_2 + \vartheta_2^T(\vartheta_2^T A^2_{22}\vartheta_2) + 1/3\,\vartheta_2^T(\vartheta_2^T \bullet (\vartheta_2^T A^2_{222}\vartheta_2)) \\ &\qquad - \vartheta_2^T[\vartheta_2^T]\,[A^2_{21}](\vartheta_2^T A^1_{22}\vartheta_2) + 1/4\|\vartheta_2^T A^2_{22}\vartheta_2\|^2 - \|[\vartheta_2^T]\,[A^2_{12}]\vartheta_2\|^2 + \dots \end{aligned}$$

As $(Y - h(\theta_{(2)}))^T(Y - h(\theta_{(2)})) - S(\theta^{\star})$ approximately equals $(h(\theta_{(2)}) - h(\theta^{\star}_{(2)}))^T(h(\theta_{(2)}) - h(\theta^{\star}_{(2)}))$ (2) can be rewritten as

$$\begin{aligned} &\vartheta_2^T\vartheta_2 + \vartheta_2^T(\vartheta_2^T A^2_{22}\vartheta_2) + 1/3\vartheta_2^T(\vartheta_2^T \bullet (\vartheta_2^T A^2_{222}\vartheta_2)) - \vartheta_2^T[\vartheta_2]^T[A^2_{21}](\vartheta_2^T A^1_{22}\vartheta_2) \\ &\qquad + 1/4\|\vartheta_2^T A^2_{22}\vartheta_2\|^2 - \|[\vartheta_2]^T[A^2_{12}]\vartheta_2\|^2 + \dots \le c^2\rho^2, \end{aligned} \tag{5}$$

where A^2_{22}, A^2_{21}, A^1_{22}, A^2_{12} and A^2_{222} are subarrays of $A..$ and $A...$, respectively.

Put $\Psi_2 = \lambda b$, $\Psi_2^\star = \lambda^\star b$ where b is a unit vector of $\mathbb{R}^{p_2}$, $\lambda, \lambda^\star \geq 0$, and Ψ_2 and $\Psi_2^\star$ are bound points of the region, (5) and the L-region, respectively. λ and $\lambda^\star$, being called the radii of these regions,depend only on the direction of b. We can therefore write $\lambda(b)$ and $\lambda(b)$. Put $b^\circ = R_{22}b$. Evidently $\lambda^\star(b) = c\rho/||b||$. Inserting $\Psi_2 = \lambda b$ into (5) we get

$$\lambda^2 b^{\circ T} b^\circ + \lambda^3 b^{\circ T}\big(b^{\circ T} A^2_{22} b^\circ\big) + \lambda^4\big\{1/4\big\|\big(b^{\circ T} A^2_{22} b^\circ\big)\big\|^2 - \big\|[b^{\circ T}][A^2_{12}]b^\circ\big\|^2 \\ - b^{\circ T}[b^{\circ T}][A^2_{21}]\big(b^{\circ T} A^1_{22} b^\circ\big) + 1/3 b^{\circ T}\big(b^{\circ T} \bullet \big(b^{\circ T} A^2_{222} b^\circ\big)\big)\big\} + \cdots = c^2\rho^2. \quad (6)$$

Now we invert (6) expressing λ as a power series in $c\rho$

$$\lambda(b) = \lambda^\star(b)\big\{1 - (c\rho/2) b^{\circ T}(b^{\circ T} A^2_{22} b^\circ)/||b^\circ||^3 \\ + \big(c^2\rho^2/(8||b^\circ||^6)\big)5\big(b^{\circ T}(b^{\circ T} A^2_{22} b^\circ)\big)^2 \\ - 4||b^\circ||^2\big(1/4\big\|(b^{\circ T} A^2_{22} b^\circ)\big\|^2 - \big\|\big(\big[b^{\circ T}\big]\big[A^2_{12}\big]b^\circ\big)\big\|^2 \\ - b^{\circ T}\big[b^{\circ T}\big]\big[A^2_{21}\big](b^{\circ T} A^1_{22} b^\circ) + 1/3 b^{\circ T} \bullet \big(b^{\circ T}\big[b^{\circ T}\big]\big[A^2_{222}\big]b^\circ\big)\big)\big\} + \ldots$$

Rewrite this equation putting $d = b^\circ/||b^\circ||$. We obtain

THEOREM 1. *Let d be a unit vector in $\mathbb{R}^{p_2}$. Then*

$$\lambda(d) = \lambda^\star(d)\big\{1 - (c\rho/2)\Gamma(d) + (c^2\rho^2/2)\beta(d) + \ldots\big\}, \quad (7)$$

where the neglected terms are of degree 4 and higher in $c\rho$, and

$$\Gamma(d) = d^T\big[d^T\big]\big[A^2_{22}\big]d,$$

$$\beta(d) = 5/4\big(d^T\big[d^T\big]\big[A^2_{22}\big]d\big)^2 - 1/4\big\|(d^T A^2_{22} d)\big\|^2 + \big\|\big[d^T\big]\big[A^2_{12}\big]d\big\|^2 \\ + d^T\big[d^T\big]\big[A^2_{21}\big]\big(d^T A^1_{22} d\big) - 1/3\, d^T \bullet \big(d^T\big[d^T\big]\big[A^2_{222}\big]d\big),$$

$$\lambda^\star(d) = c\rho||R_{22}^{-1}d|| \quad (8)$$

(7) provides a method of the construction of the likelihood region: giving the vector d different values we obtain the bounds of this region.

Our experience shows that the contribution of $\beta(d)$ to (7) can be neglected in many cases met in practice.

The second factor in the right-hand side of (7) is the ratio of the radii of the likelihood and L-regions. Therefore

$$M = \max\big\{\big|-(c\rho/2)\Gamma(d) + (c^2\rho^2/2)\beta(d)\big|\colon ||d|| = 1\big\} \quad (9)$$

indicates the total nonlinearity of the model.

We can use any of the following numbers to approximate M

$$M_1 = c/2M_s^T, \text{ where } M_s^T = \rho \max\{|\Gamma(d)|\colon \|d\| = 1\},$$

$$M_2 = \max\{c/2M_s^T, \big|(-c\rho/2)\Gamma(d^\circ) + (c^2\rho^2/2)\beta(d^\circ) + 1/3 d^{\circ T} \bullet \big(d^{\circ T}[d^{\circ T}][A_{222}^2]d^\circ\big)\big|\},$$

$$M_3 = \max\{c/2M_s^T, \big|(-c\rho/2)\Gamma(d^\circ) + (c^2\rho^2/2)\beta(d^\circ)\big|\},$$

where d° maximizes $|\Gamma(d)|$. It is clear that

$$M_s^T \le \max\{\big\|(d^T A_{22}^2 d)\big\|\colon \|d\| = 1\} = \Gamma_s^T, \tag{10}$$

where Γ_s^T is the maximum parameter-effects curvature for $\theta_{(2)}$ as defined in [2]. Basing on this we call M_s^T the weak parameter-effects curvature. Following [1] we suggest a very rough rule for the choice of the confidence region for $\theta_{(2)}$:

1. If $M \le 0.15$, curvature effects may be ignored and the L-region accepted.
2. If $0.15 < M \le 1/3$, the region based on (7) is recommended.
3. If $1/3 < M$, a reparametrization is needed.

In [1] Clarke gave a likelihood interval for the scalar parameter θ_p with the bound points defined by the equation

$$\theta_p - \theta_p^\star = (g^{pp}\rho^2)^{1/2} c^\sim\big(1 - (c^\sim\rho)/2\big)\Lambda + (c^\sim\rho)^2/2\tau + \dots \tag{11}$$

where $c^\sim = \pm c.$, $\Lambda = A_{pp}^p$, $\tau = -(A_{pp}^p) + \|A_{.p}^p\|^2 + A_{p.}^p A_{pp}^{\cdot} - 1/3 A_{ppp}^p$, g^{pp} is the lower element of $(V^T V)^{-1}$. It is worth mentioning that his result follows from the Theorem 1 for $p_2 = 1$ and $\theta_{(2)} = \theta_p$.

Consider a non-linear transformation depending on $\theta_{(2)}$ only, namely

$$\beta_{(1)} = \theta_{(1)}; \quad \beta_{(2)} = s(\theta_{(2)}) = \big(s(\theta_{(2)}), \dots, s_{p_2}(\theta_{(2)})\big).$$

Denote by $H.$, $H..$ and $H...$ the first derivatives matrix, the second and third derivatives arrays of $s(\theta_2)$, respectively. Assuming that the QR-decomposition of $R_{22}H^{-1}$ is $X_2 K_{22}$ we obtain the following result.

Theorem 2. *Three-dimensional parameter-effects array for the new parameter β is $B..$ with the elements*

$$\begin{bmatrix} A_{11}^1 & A_{12}^1 X_2 \\ X_2^T A_{21}^1 & X_2^T\big(A_{22}^1 - L_{22}^T[R_{12}H^{-1}][H..]L_{22}\big)X_2 \\ [X_2^T][A_{11}^2] & [X_2^T][A_{12}^2]X_2 \\ X_2^T[X_2^T][A_{21}^2] & [X_2^T][X_2^T A_{22}^{2\star} X_2] \end{bmatrix},$$

where

$$A_{22}^{2*} = A_{22}^2 - L_{22}^T[L_{22}^{-1}H^{-1}][H..]L_{22}; \quad L_{22} = R_{22}^{-1}.$$

Following the same argument as in Theorem 1 we find that confidence region for $s(\theta_{(2)})$ is

$$\Phi^T\Phi + \Phi^T(\Phi^T B^2_{22}\Phi) + 1/3\,\Phi^T(\Phi^T \bullet (\Phi^T B^2_{222})) - \Phi^T[\Phi^T][B^2_{21}](\Phi^T B^1_{22}\Phi) \\ - \big\|[\Phi^T][B^2_{12}]\Phi\big\|^2 + 1/4\|\Phi^T B^2_{22}\Phi\|^2 + \dots \le c^2\rho^2\,,$$

where $\Phi = K_{22}\big(s(\theta_{(2)}) - s(\theta^\star_{(2)})\big)$, B^2_{22}, B^2_{21}, B^2_{12}, B^1_{22} are the elements of the new three-dimensional parameter-effects array given in Theorem 2, B^2_{222} is the element of the new four-dimensional parameter-effects array, c and ρ (the standard radius) are defined in the usual way. Thus we need only one element of $B\dots$, namely B^2_{222}, to construct the confidence region. For this the following lemma is needed.

Lemma 1.

$$B^2_{222} = A^2_{222} - [L^{-1}_{22}H.^{-1}]\big[L^T_{22} \bullet (L^T_{22}H\dots L_{22})\big] \\ - \big([(A^2_{21}L^{-1}_{22} + A^{2\star}_{22}L^{-1}_{22})H.^{-1}][L^T_{22}H..L_{22}]\big)^\sim,$$

where $L_{12} = -R^{-1}_{11}R_{12}R^{-1}_{22}$ and and where the following convention has been used: given an array M, $M^\sim$ is defined to be the array with the elements $(M^\sim)^i_{jk} = M^i_{jk} + M^j_{ki} + M^k_{ij}$.

Put now $\Phi = \Lambda(h)h$; $\|h\| = 1$. By using Theorem 1, Theorem 2, and the above lemma we have the final result.

Theorem 3. *Let h be a unit vector in R^{p_2}. Then*

$$\Lambda(h) = \Lambda^\star(h)(1 - (c\rho/2)\Gamma(h) + (c^2\rho^2/2)\beta(h) + \dots\,,$$

where the neglected terms are of degree 4 and higher in $c\rho$, and

$$\Gamma(h) = h^T[h^T][A^{2\star}_{22}]h\,,$$

$$\beta(h) = 5/4\Gamma^2(h) - 1/4\big\|h^T[A^{2\star}_{22}]h\big\|^2 + \big\|[h^T][A^2_{12}]h\big\|^2 \\ + h^T[h^T][A^2_{21}](h^T A^{1\star}_{22}h) - 1/3h^T\big(h^T \bullet \big(h^T[h^T][B^2_{222}]h\big)\big)\,,$$

$$A^{1\star}_{22} = A^1_{22} - L^T_{22}[R_{12}H.^{-1}][H..]L_{22}\,,$$

$$\Lambda^\star(h) = c\rho\|H.L_{22}h\|\,.$$

The theorem is used in the same way as Theorem 1 to give the confidence region for $s(\theta_{(2)})$. In addition it gives a deeper insight into the problem of finding such a transformation of the original parameter that the confidence region is computed easily. We see, for example, that the requirement $\Gamma(h) = 0$ is equivalent to $R^T_{22}[L_{22}][A^2_{22}]R_{22} = [H.^{-1}][H..]$. For $\beta(h)$ a similar argument can be used.

3. Illustrations

Example 1. The Fieller-Creasy problem. The model function is $f(x_a, \theta) = \theta_1 x_a + \theta_1 \theta_2 (1 - x_a)$ where x_a is an indicator variable that takes the values 1 and 0. We assume the equal sample sizes n for both populations and the known common variance σ^2 such that $\sigma^2/n = 1$. We have $c^2 = \chi^2(2, 0.05) = 6$. By using Theorem 1 ($p_1 = 0, p_2 = p$) we obtain the values of M for different $(\theta_1^\star, \theta_2^\star)$. They are listed in Table 1.

θ_1^*	θ_2^*	Γ^T	c^{-1}	M
3	0	0.34	0.41	0.314
8	1.2	0.35	0.41	0.238
8	1.8	0.48	0.41	0.305
8	2.4	0.63	0.41	0.374
8	3.0	0.77	0.41	0.440

In the first case the value of the Bates-Watts parameter-effects curvature is smaller than the cutoff c^{-1}. According to the Bates-Watts criterion, the L-region should agree with the likelihood region,whose level is exact in this case. As Fig. 1a shows, this is not the case. On the other hand, our criterion leads to the conclusion that the region mentioned in the Theorem 1 should be used, since $0.15 < M < 1/3$. Fig. 1a confirms the good agreement between this region and the likelihood region. In the second case M lies between 0.15 and 0.25. We hope that there is not a big mistake in using L-region. Again Fig. 1b supports this conclusion. The L-region in the third case differs from the corresponding likelihood region. Since $0.15 < M < 1/3$ we recommend the region based on (7) which agrees well with the likelihood region.

In the last two cases M is large and hence, both the L-region and the region based on (7) should not be used. Instead, a reparametrization is needed. We recall that all our conclusions agree with findings in [2]. Let us return to the first case and concentrate now on θ_2. As Fig. 1a shows, the likelihood interval for θ_2 will be about 67 % longer than the corresponding linear approximation interval. Using (11) of this paper leads to an improved interval which is almost identical with the likelihood one.

Example 2. Consider the three-parameter model $f(x_a, \theta) = \theta_1 + \theta_2 \exp(\theta_3 x_a)$ in combination with 27 observations reported in [7] and concentrate on the subset (θ_1, θ_3). Fig. 2 gives 86 % likelihood and linear confidence regions as well as the region based on Theorem 1. (The "unusual" 86 % level is chosen to ensure that the corresponding univariate marginal regions have a nominal 95 % coverage rate; see [2]). In this case $\Gamma_s^T = 1.07$ is above the Bates-Watts cutoff $0,542$. In spite of this the linear region seems to approximate the likelihood

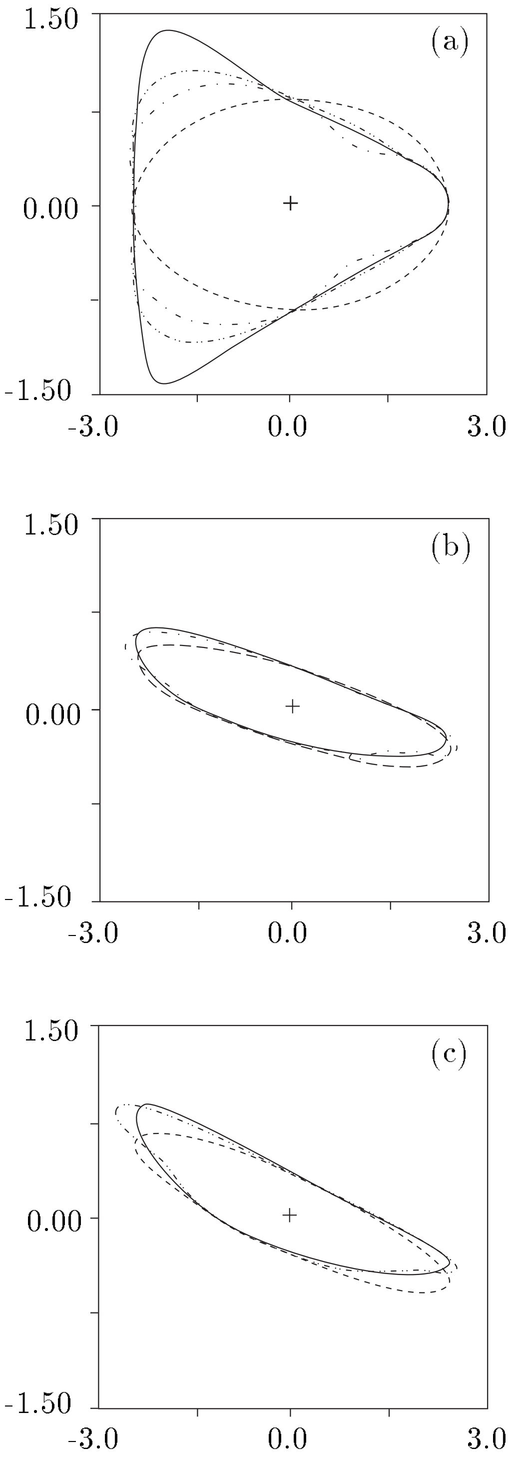

FIGURE 1. 95 % likelihood region for (ϕ_1, ϕ_2); $\phi_i = \Theta_1 - \Theta_1^*$, $i = 1, 2$, when a) $(\Theta_1^*, \Theta_2^*) = (3; 0)$ b) $(\Theta_1^*, \Theta_2^*) = (8; 1, 2)$ c) $(\Theta_1^*, \Theta_2^*) = (8; 1, 8)$. $---$ L-region, $- \cdot - \cdot -\cdot$ (7) using $\Gamma(d)$ only, $- \cdot \cdot - \cdot \cdot - \cdot \cdot$ using $\Gamma(d)$, —— exact region.

region relatively well. Nevertheless, according to the rule presented in this paper the region based on Theorem 1 should be used. As Fig. 2 shows this choice ensures even closer agreement than the linear region.

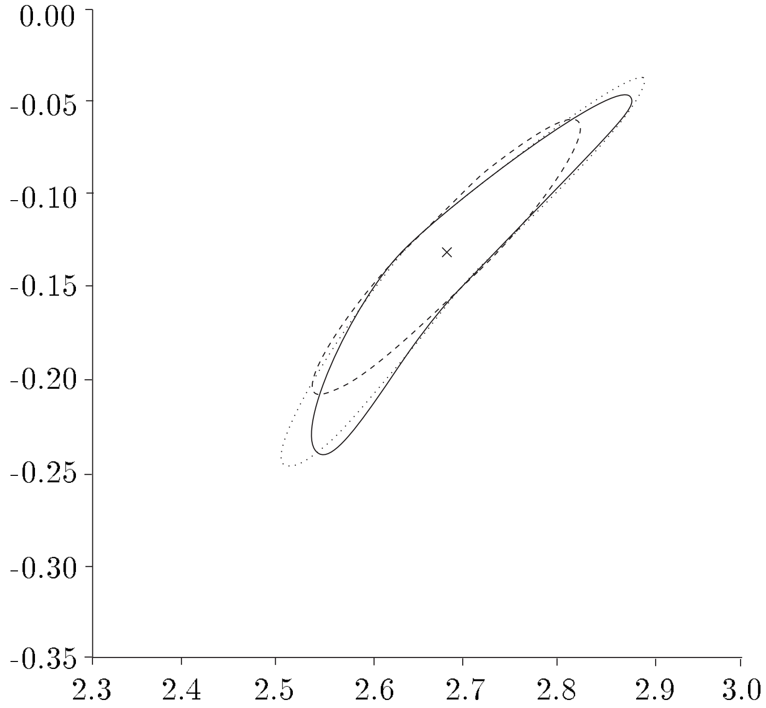

FIGURE 2. 86 % confidence regions for (Θ_1, Θ_3) from Example 2. — — — L-region, $\cdots\cdot$ based on Theorem 1, —— likelihood.

4. Discussion

Hamilton [3] has constructed simple likelihood regions, but not for the original parameter. Instead, his results deal with a parameter which is a nonlinear function of the original one. A back transformation is therefore needed which, however, can be difficult to find. Clarke [1] presented a method of constructing regions with higher precision. His investigations, however, deal with a scalar parameter only.

The advantage of Theorem 1 is that it leads to approximate likelihood regions for the original vector parameter. Moreover we need only some elements of $A_{..}$ and $A_{...}$, namely A_{12}^2, A_{22}^2, A_{22}^1 and A_{222}^2, to construct the likelihood region (7).

REFERENCES

[1] CLARKE, G. P. I.: *Marginal curvatures and their usefulness in the analysis of nonlinear regression models*, J. Amer. Statist. Asoc. **82** (1987), 844–850.

[2] COOK, R. D.—GOLDBERG, M. L.: *Curvatures for parameter subsets in nonlinear regression*, Ann. Statist. **14** (1986), 1399–1418.

[3] HAMILTON, D.: *Confidence regions for parameter subsets in nonlinear regression*, Biometrika **73** (1986), 57–64.

[4] KUBÁČEK, L.: *Models with a low nonlinearity*, Probastat 94, Tatra Mt. Math. Publ. **7** (1996), 149–155.

[5] PÁZMAN, A.: *Nonlinear Statistical Models*, Kluwer Acad. Publ., Dordrecht, 1993; Ister Science Press, Bratislava, (1993).

[6] POTOCKÝ, R.—VAN BAN, T.: *Confidence regions in nonlinear regression models*, Appl. Math. **37** (1992), 29–39.

[7] RATKOWSKY, D. A.: *Nonlinear Regression Modeling*, Marcel Dekker, Inc., New York, 1983.

Received May 2, 2002

Department of Probability and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University
Mlynská dolina
SK–842 48 Bratislava
SLOVAKIA

E-mail: potocky@fmph.uniba.sk