The Annals of Statistics
1981, Vol. 9, No. 6, 1152–1167

PARAMETER TRANSFORMATIONS FOR IMPROVED APPROXIMATE CONFIDENCE REGIONS IN NONLINEAR LEAST SQUARES

BY DOUGLAS M. BATES AND DONALD G. WATTS

University of Wisconsin and Queen's University

In a previous paper, it was shown that parameter-effects nonlinearities of a nonlinear regression model–experimental design–parameterization combination can be quantified by means of a parameter-effects curvature array A based on second derivatives of the model functon. In this paper, the individual terms of A are interpreted and local compensation methods are suggested. A method of computing the parameter-effects array corresponding to a transformed set of parameters is given and we discuss how this result could be used to determine reparameterizations which have zero local parameter-effects nonlinearity.

1. Introduction. In an earlier paper (Bates and Watts, 1980), we developed curvature measures for intrinsic and parameter-effects nonlinearities of a model-design-parameterization combination. The curvature measures were developed for the usual nonlinear model in which the relationship between the values of a response y_t, $(t = 1, 2, \cdots, n)$ and some control variables $\mathbf{x}_t = (x_{t1}, x_{t2}, \cdots, x_{tr})'$ can be written

$$y_t = f(\mathbf{x}_t, \boldsymbol{\theta}) + \varepsilon_t, \qquad t = 1, 2, \cdots, n. \tag{1.1}$$

In (1.1), $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_p)'$ is a set of unknown parameters and ε_t are random normal errors assumed to have zero mean, to have constant variance, and to be independent of one another.

Because the model is nonlinear, the solution locus (Box and Lucas, 1959) described by

$$\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \eta_2(\boldsymbol{\theta}), \cdots, \eta_n(\boldsymbol{\theta}))', \tag{1.2}$$

where $\eta_t(\boldsymbol{\theta}) = f(\mathbf{x}_t, \boldsymbol{\theta})$, is not a plane as in the linear case, but is a curved surface dictated by the model function and design combination. The non-planarity of the solution locus, or intrinsic curvature, has been discussed by Beale (1960), Bates and Watts (1980) and Hamilton, Watts and Bates (1981). The last reference presents methods for accounting for the effects of intrinsic nonlinearity in constructing approximate confidence or likelihood regions. In this paper, we consider parameter-effects nonlinearities and show how their effects on linear approximation confidence regions may be modified by reparameterization so as to provide improved linear approximation confidence regions. In Section 2 we give interpretations of the individual terms in the array A and show how local compensation of parameter effects may be effected. Next, in Section 3 we show that the array $\tilde{A}$, corresponding to a nonlinear reparameterization $\boldsymbol{\beta} = G(\boldsymbol{\theta})$, may be computed efficiently from the original array, A, corresponding to the parameters $\boldsymbol{\theta}$, minus a correction term which depends on the first and second derivative terms $G.$ and $G..$ In Section 4 we discuss how the above results could be used to determine reparameterizations which have zero local parameter-effects nonlinearities and hence would yield good linear-approximation confidence regions. One promising approach appears to be that of using a particular class of transformation, such as Ross's expected-value transformation, to give zero or small parameter-effects nonlinearity. The more general problem of measuring the curvature of a

Received May 1980; revised March 1981.

AMS 1970 *subject classifications.* Primary 62J02; secondary 62F25.

Key words and phrases. Parameter-effects curvatures, reparameterization, expected-value transformations.

statistical problem to indicate the nonlinearity in the sense that locally optimum methods fail to be globally optimal, is considered in Efron (1975, 1978) and the methods derived there are applied in Efron and Hinkley (1978). N. Reid and D. Hinkley, in their discussion of Bates and Watts (1980), both demonstrate that the intrinsic curvature of the nonlinear solution locus is a special case of Efron's statistical curvature. The implications of this and other aspects of parameter transformations are discussed in Section 5.

2. The components of parameter-effects nonlinearity.

2.1. *The parameter-effects curvature array.* For a nonlinear model of the form (1.1), the least squares estimates $\hat{\boldsymbol{\theta}}$ are the values of the parameters which minimize the sum of squares

$$S(\boldsymbol{\theta}) = \sum_{t=1}^{n} \{y_t - f(\mathbf{x}_t, \boldsymbol{\theta})\}^2. \tag{2.1}$$

In vector notation, equation (2.1) can be written

$$S(\boldsymbol{\theta}) = \| \mathbf{y} - \boldsymbol{\eta}(\boldsymbol{\theta})\|^2, \tag{2.2}$$

where the double vertical bars denote the length of a vector, so that the least squares estimates are those values such that $\boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$ is the point on the solution locus which is closest to $\mathbf{y}$. If the solution locus is relatively flat near $\boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$, so that it can be reasonably approximated by the tangent plane, a $1 - \alpha$ confidence region consists of those values of $\boldsymbol{\theta}$ for which

$$\| \boldsymbol{\eta}(\boldsymbol{\theta}) - \boldsymbol{\eta}(\hat{\boldsymbol{\theta}})\|^2 \leq ps^2F, \tag{2.3}$$

where $F = F(p, \nu; \alpha)$ is the upper α probability point of Fisher's F distribution with p and ν degrees of freedom, and s^2 is an estimate of σ^2 based on ν degrees of freedom. Since the form of equation (2.3) is that of a sphere, we regard the square root of the term on the right as a radius. To avoid a dependence on the confidence level, we call ρ the standard radius, where

$$\rho^2 = ps^2. \tag{2.4}$$

In the following development, we assume that the estimates have been obtained and that the data have been scaled so that $\rho = 1$.

To determine the curvature array, we evaluate the first and second derivatives of the model function, $V_.$ and $V_{..}$, with entries

$$(V_.)_{ti} = \partial\eta_t/\partial\theta_i |_{\hat{\theta}} \tag{2.5}$$

and

$$(V_{..})_{tij} = \partial\eta_t/\partial\theta_i\partial\theta_j |_{\hat{\theta}} \tag{2.6}$$

and form an orthogonal-triangular decomposition of $V_.$ (see Kennedy and Gentle, 1980),

$$V_. = QR \tag{2.7}$$

where Q is n by n orthogonal and R is p by n with zeros below the main diagonal. The upper p by p submatrix of R is denoted R_1 and its inverse by $L = R_1^{-1}$.

Now we wish to perform some linear algebra manipulations with the second derivative array $V_{..}$, but normal matrix multiplication notation is ambiguous in this case. We therefore distinguish one type of summation using square brackets, so that $[Q'][V_{..}]$ indicates a summation over the index t in terms like $(V_{..})_{tij}$ (the only summation which is consistent with the dimensions). We also call this a numerator multiplication since the summation is over the term occurring in the numerator of the partial derivative. Multiplications written without square brackets indicate summation over the second index (for premultiplication) or the third index (for postmultiplication). Note that the square bracket and regular multiplications commute.

Using this notation, the curvature array $A_{..}$ can be written

$$A_{..} = [Q'][L'V_{..}L], \tag{2.8}$$

which can be regarded as the second derivative array for the parameters

$$\boldsymbol{\phi} = L'(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{2.9}$$

in a rotated set of sample space coordinates.

The first p faces of $A_{..}$ constitute the parameter-effects curvature array $A^T_{..}$. We also denote the first p columns of Q by the symbol $U_.$, so $A^T_{..}$ can be written

$$A^T_{..} = [U'_.][L'V_{..}L]. \tag{2.10}$$

The matrix $U_.$ is the derivative of $\boldsymbol{\eta}$ with respect to $\boldsymbol{\phi}$ evaluated at $\boldsymbol{\phi} = \mathbf{0}$ and its columns ($\mathbf{u}_i$, $i = 1, 2, \cdots, p$) provide an orthogonal basis for the tangent plane to the solution locus at $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$. The corresponding second derivative array is

$$U_{..} = L'V_{..}L, \tag{2.11}$$

composed of vectors ($\mathbf{u}_{ij}$, $i, j = 1, \cdots, p$).

The advantage of the $\boldsymbol{\phi}$ parameters is that if $\mathbf{d} = (d_1, d_2, \cdots, d_p)'$ is a unit vector, then the relative parameter effects curvature corresponding to the direction $\mathbf{d}$ is simply $\gamma^T_{\mathbf{d}} = \| \mathbf{d}'A_{..}\mathbf{d} \|$. The maximum curvature over all directions $\mathbf{d}$ is defined to be the parameter effects curvature, Γ^T. Since we will only be dealing with parameter effects, we delete the superscript T and the subscripts $\cdot\cdot$ on the parameter effects array in this paper.

2.2 *Tangent plane coordinates.* The usual approach to obtaining an approximate $1 - \alpha$ confidence region is to assume that both parameter-effects and intrinsic nonlinearities are negligible, and to define a region based on a linear approximation as

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'V'_.V_.(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \le F, \tag{2.12}$$

where $V_.$ is calculated for the scaled responses. In Hamilton *et al* (1981), it is shown that confidence regions can be approximated as elliptical regions on the tangent plane to the solution locus at $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}(\hat{\boldsymbol{\theta}})$ and this approximation does not require any assumptions on parameter-effects. It is further shown that this procedure compensates for intrinsic nonlinearity. If the region is based on an estimate of σ from replications, this region is contained in

$$\| \tilde{\boldsymbol{\eta}}(\boldsymbol{\theta}) - \hat{\boldsymbol{\eta}} \| \le m\sqrt{F}, \tag{2.13}$$

where $\tilde{\boldsymbol{\eta}}(\boldsymbol{\theta})$ is the projection of the point $\boldsymbol{\eta}(\boldsymbol{\theta})$ onto the tangent plane at $\hat{\boldsymbol{\eta}}$, m is the inflation factor expressed as

$$m = (1 - \lambda_p)^{-1}, \tag{2.14}$$

and λ_p is the largest eigenvalue of $B = [\mathbf{y} - \hat{\boldsymbol{\eta}}]'[U_{..}]$. We call this a conservative confidence region.

In the transformed sample space coordinates, the projection of $\boldsymbol{\eta}(\boldsymbol{\theta}) - \hat{\boldsymbol{\eta}}$ onto the tangent plane is simply $Q'\{\boldsymbol{\eta}(\boldsymbol{\theta}) - \hat{\boldsymbol{\eta}}\}$ with the last $n - p$ coordinates set to zero. Letting $\boldsymbol{\tau}$ be the tangent plane coordinate vector we have

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

so that $\| \boldsymbol{\tau} \| = \| \tilde{\boldsymbol{\eta}}(\boldsymbol{\theta}) - \hat{\boldsymbol{\eta}} \|$. Then the conservative confidence region is simply the disk on the tangent plane given by

$$\| \boldsymbol{\tau} \| \le m\sqrt{F} = R_\alpha, \tag{2.15}$$

where R_α is the radius of the $1 - \alpha$ conservative confidence region disk. We may also express the coordinates $\boldsymbol{\tau}$ in terms of the $\boldsymbol{\phi}$ parameters as

$$\boldsymbol{\tau} = U'_.\{\boldsymbol{\eta}(\hat{\boldsymbol{\theta}} + L\boldsymbol{\phi})) - \hat{\boldsymbol{\eta}}\} = H(\boldsymbol{\phi}). \tag{2.16}$$

Note from (2.16) that $H(\mathbf{0}) = \mathbf{0}$ and

$$\partial H/\partial \phi|_0 = U'_{.}\partial \eta/\partial \phi|_0 = U'_{.}U_{.} = I. \tag{2.17}$$

The second derivative of H is a $p \times p \times p$ array with i, j, k entry

$$\begin{aligned} \partial^2 \tau_i/\partial \phi_j \partial \phi_k|_0 &= \partial^2\{u_i'(\eta - \hat{\eta})\}/\partial \phi_j \partial \phi_k|_0 \\ &= u_i' u_{jk} = (A)_{ijk} = a_{ijk}, \end{aligned} \tag{2.18}$$

that is

$$\partial^2 H/\partial \phi^2|_0 = [U'_{.}][U_{..}] = A. \tag{2.19}$$

Thus the second-order Taylor series approximation to H is

$$H(\phi) = \phi + (\phi' A \phi)/2. \tag{2.20}$$

Equation (2.20) may be used to give an alternate interpretation for the curvature measure Γ. From (2.13) the conservative confidence region in the ϕ coordinates will consist of all values of ϕ for which

$$\|H(\phi)\| \leq R_\alpha,$$

and so using a first order approximation, the conservative confidence region consists of all ϕ such that $\|\phi\| \leq R_\alpha$. To test the validity of this approximation we can compare the length of the quadratic to that of the linear term. For the quadratic expansion, the length of the quadratic term is

$$\begin{aligned} \|\phi' A \phi\|/2 &= \|\phi\|^2 \|\mathbf{d}' A \mathbf{d}\|/2 \\ &= \|\phi\|^2 \gamma_d/2, \end{aligned}$$

where the unit vector $\mathbf{d} = \phi/\|\phi\|$, and so the ratio of the lengths of the quadratic to the linear components has a maximum of $R_\alpha \Gamma/2$. Thus $R_\alpha \Gamma/2$ indicates the usefulness of the linear approximation for a $1 - \alpha$ confidence region, a value exceeding ½ or ¼ indicating unacceptable nonlinearity at level α because the quadratic term is comparable to the linear term rather than being a small correction to the linear term.

To assess the effect of Γ or of particular terms of A on a confidence region, we compare them to $1/R_\alpha$. Terms which exceed $1/R_\alpha$ in magnitude reveal serious departures of the linear approximation from the true confidence region. This approach has the advantage that once A and Γ have been calculated, the adequacy of a linear approximation confidence region for different confidence levels α_1, α_2 etc. can be determined simply by comparing the curvatures a_{ijk} and Γ to the curvatures $1/R_{\alpha_1}$, $1/R_{\alpha_2}$ and so on.

2.3 *Interpreting the array A*. Just as we can trace parameter curves on the solution locus by varying one coordinate of ϕ while holding the others fixed, so may we trace parameter curves on the tangent plane. If the linear approximation to H were exact, these parameter curves would be straight, equispaced, parallel lines and all entries in A would be zero: non-zero entries indicate failure of the parameter curves to be straight or equispaced or parallel. In fact, precise interpretation of each of the terms in the array A is possible. For convenience, we denote the terms $(A)_{ijk}$ simply by a_{ijk}.

To interpret the terms a_{ijk} of the array A, consider a two dimensional example with tangent vectors $\mathbf{u}_1$ and $\mathbf{u}_2$ and second derivative vectors $\mathbf{u}_{11} = a_{111}\mathbf{u}_1 + a_{211}\mathbf{u}_2$, $\mathbf{u}_{12} = a_{112}\mathbf{u}_1 + a_{212}\mathbf{u}_2$ and $\mathbf{u}_{22} = a_{122}\mathbf{u}_1 + a_{222}\mathbf{u}_2$. Now the point $\phi = \mathbf{0}$ coincides with the point $\tau = \mathbf{0}$, and at that point $\partial\tau/\partial\phi_1 = \mathbf{u}_1$. At another point $\phi = \boldsymbol{\delta}$ the new tangent vectors $\partial\tau/\partial\phi_1$ and $\partial\tau/\partial\phi_2$ will be approximately

$$\partial\tau/\partial\phi_1 \sim \partial\tau/\partial\phi_1|_0 + \delta_1(\partial^2\tau/\partial\phi_1^2) + \delta_2(\partial^2\tau/\partial\phi_1\partial\phi_2) = \mathbf{u}_1 + \delta_1\mathbf{u}_{11} + \delta_2\mathbf{u}_{12}$$

and

$$\partial\tau/\partial\phi_2 \sim \mathbf{u}_2 + \delta_1\mathbf{u}_{12} + \delta_2\mathbf{u}_{22}.$$

Expanding gives

$$\partial\tau/\partial\phi_1 \sim \mathbf{u}_1(1 + a_{111}\delta_1 + a_{112}\delta_2) + \mathbf{u}_2(a_{211}\delta_1 + a_{212}\delta_2),$$

$$\partial\tau/\partial\phi_2 \sim \mathbf{u}_1(a_{112}\delta_1 + a_{122}\delta_2) + \mathbf{u}_2(1 + a_{212}\delta_1 + a_{222}\delta_2).$$

Thus, a_{111} gives the change in the $\mathbf{u}_1$ direction of the $\partial\tau/\partial\phi_1$ vector due to a unit change in $\partial\tau/\phi_1$; that is, terms of the form a_{iii} cause changes in length only. For this reason we refer to a_{iii} as *compansion* terms, since they cause compression or expansion of scale along a ϕ_i parameter line. The term a_{211} gives the change in the $\mathbf{u}_2$ direction of the $\partial\tau/\partial\phi_1$ vector due to a unit change in ϕ_1; that is, terms of the form a_{jii} ($j \neq i$) cause changes in the $\mathbf{u}_j$ direction of the ϕ_i parameter lines as we move along them. We refer to these as *arcing* terms. The term a_{212} gives the change in the $\mathbf{u}_2$ direction of the $\partial\tau/\partial\phi_1$ vector due to a unit change in ϕ_2; that is, terms of the form a_{jij} cause changes in the $\mathbf{u}_j$ direction of the ϕ_i parameter curves as we move across the ϕ_j parameter curves. We call these *fanning* terms since the ϕ_i parameter lines will appear to fan out from a common point on the τ_i axis. Since $a_{jji} = a_{jij}$, terms of the form a_{jji} also cause ϕ_i fanning.

With two parameters, only compansion, arcing and fanning can occur. With more than two parameters, only one more type of parameter effect can occur – when all the subscripts are different. A term such as a_{ijk} causes a change in the $\mathbf{u}_i$ direction of the $\partial\tau/\partial\phi_j$ tangent vector due to a unit change in ϕ_k. We refer to these as *torsion* terms since they cause a twisting of the (ϕ_j, ϕ_k) parameter surface, where a parameter surface – analogous to a parameter curve – is the set of points generated by holding all ϕ_i's except two constant and varying those two.

2.4 *An example.* The data given in Section 3.1 of Bates and Watts (1980) for the Michaelis-Menten model

$$f(\mathbf{x}, \boldsymbol{\theta}) = \theta_1 x/(\theta_2 + x)$$

yields a parameter effects array

$$A = \begin{pmatrix} 0 & -.292 \\ & -.163 \end{pmatrix}$$
$$\quad\;\; \begin{pmatrix} 0 & -.081 \\ & -.716 \end{pmatrix}$$

with $\Gamma = 0.771$. Since each face of the array A is symmetric, we only display the upper triangular part; so, for example, in the above array, $a_{212} = a_{221} = -.081$.

Examination of the terms in A reveals the following: the ϕ_2 parameter curves will be perfectly uniformly spaced since the ϕ_1 compansion term (a_{111}) is zero. On the other hand, since the ϕ_2 compansion term (a_{222}) is $-.716$, the ϕ_1 parameter curves will be markedly compressed together as ϕ_2 increases. The small ϕ_2 arcing term (a_{122}) of $-.163$ will cause little curving of the ϕ_2 parameter curves and the ϕ_1 parameter curves will be straight because the ϕ_1 arcing term (a_{211}) is zero. The small ϕ_1 fanning term ($a_{221} = a_{212}$) of $-.081$ causes little convergence of the ϕ_1 parameter curves, but the larger value for the ϕ_2 fanning term ($a_{121} = a_{112}$) of $-.292$ causes considerable convergence of the ϕ_2 parameter curves. The parameter curves drawn on the tangent plane, shown in Figure 1, behave as predicted. (To avoid confusion in interpreting this diagram, recall that a parameter curve is associated with the parameter which is varying; therefore the lines labelled $\phi_1 = 0$, $\phi_1 = 1$, and so on are actually ϕ_2 parameter curves.)

For the 95% confidence level, a conservative confidence region corresponds to a circle of radius $R_{0.05} = 2.27$ on the solution locus, as shown by the circle in Figure 1. This region is much too large for the uniform coordinate assumption to be acceptable, and so we expect considerable discrepancy between the linear approximation and exact 95% confidence

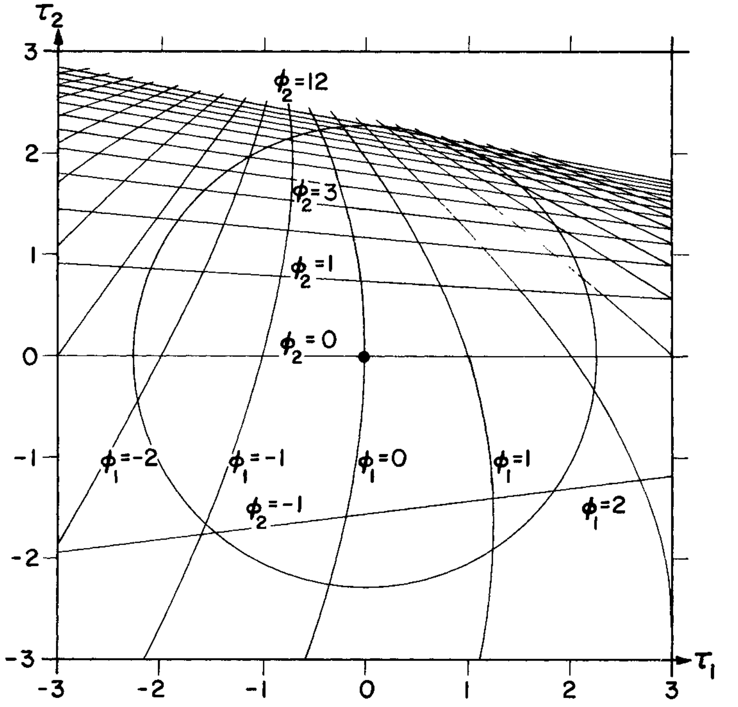

FIG. 1. ϕ parameter curves on the tangent plane.

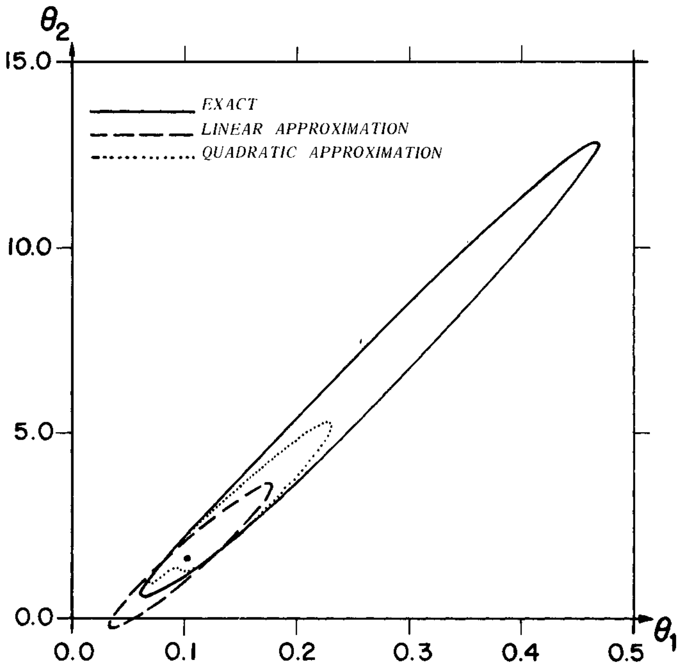

FIG. 2. Exact and approximate 95% confidence regions for Michaelis-Menten example.

regions. This inadequacy of the linear approximation at the 95% level is consistent with the value $\Gamma = 0.771$ greatly exceeding $1/R_{0.05} = 0.441$. In addition, we may compare individual terms in A to the curvature $1/R_{0.05}$ to see if there are any severe parameter effects at level 0.05. In this case, the 95% radius of curvature (0.441) is exceeded by the ϕ_2 compansion term (−.716), so we again conclude that there would be serious effects due to nonlinearity. Figure 2 further illustrates the serious inadequacy of the linear approximation parameter confidence region (shown by the dashed line) when compared to the exact region (solid line).

2.5 *Compensating for particular parameter effect nonlinearities.* As shown in (2.15), the conservative confidence region is easily expressed in terms of the τ parameters. However, for the purposes of the experimenter and statistician, expression of the confidence region in the $\boldsymbol{\theta}$ parameters is much more valuable. If the inverse of the transformation H were available, we would be able to map the region in the τ coordinates into a region in the ϕ coordinates and thence into the $\boldsymbol{\theta}$ coordinates. Unfortunately, calculation of this inverse would be an extremely difficult, if not impossible, task in most practical situations.

However, our knowledge of the types of parameter effects measured by the array A and their relationship to the transformation H can be used to determine approximations to H^{-1} which allow us to compensate for particular parameter effects. For instance, in the example above the most damaging parameter effect is compansion so we may be led to approximate H^{-1} by

$$\tilde{H}_i^{-1}(\tau) = \{-1 + \exp(a_{iii}\tau_i)\}/a_{iii}, \qquad i = 1, \ldots, p,$$

which compensates for compansion but does not alter any of the other parameter effects, as shown in Bates (1978). The confidence region in the $\boldsymbol{\theta}$ parameters obtained this way, bounded by $\{\hat{\boldsymbol{\theta}} + L\tilde{H}^{-1}(\tau){:}\, \|\tau\| = R_\alpha\}$, will have been compensated locally for compansion, but there will be no compensation for other parameter effects and no guarantee that the compansion compensation will apply over the entire region of interest $\{\tau{:}\ \|\tau\| = R_\alpha\}$.

A slightly more general approach which compensates locally for all of the parameter effects is to use a second order Taylor series approximation

$$\tilde{H}^{-1}(\tau) = \tau - (\tau' A \tau)/2.$$

This is an easily expressed transformation but unfortunately it can lead to curious behavior of the confidence region: we have seen instances where regions calculated using this approximate inverse had loops on the boundary and other bizarre behavior.

Both of these approaches and other possible types of local parameter effects compensation suffer from the defect that they are local methods and are symptomatic in nature. That is, they are attempts to combat the symptoms of the parameter effects rather than attack the root causes. As such, it is not clear over what region these compensation methods will be effective and therefore how large a confidence region can be approximated using these methods. For example, Figure 2 shows the 95% confidence region obtained by the quadratic approximation to H^{-1} and, 'while it is an improvement on the linear approximation, it is still clearly inadequate. Local compensation methods also suffer from the defect that the parameters created in this way, if we regard the approximate inverse $\tilde{H}^{-1}$ as effectively creating a new set of parameters, are artificial and hence would not usually have a direct meaning for the experimenter or statistician. In general we have found that reparameterization is a better approach to the problem of dealing with parameter effects.

3. Determining a new parameter-effects array. The most effective method of dealing with parameter effects is to reparameterze the model in an advantageous way, where advantageous implies meaningful parameters with small parameter effects. In some cases, there are recommended transformations (Draper ad Smith, 1966; Guttman and

TABLE 1
Effect of transformations for exponential models

	Parameter-effects nonlinearities					
Data set[1]	*Original*		*Transformation 1*[2]		*Transformation 2*[3]	
	γ_{rms}	Γ	γ_{rms}	Γ	γ_{rms}	Γ
5	0.302	0.398	—[4]	—	0.179	0.207
9	2.095	3.405	—	—	0.916	1.492
13	0.402	0.659	0.088	0.146	0.317	0.521
14	1.375	2.241	0.284	0.449	0.329	0.531
16	2.650	5.935	1.024	2.296	1.544	3.467
18	16.533	36.970	—	—	12.516	27.989
19	1.086	2.487	0.606	1.366	0.885	2.045
22	1.487	4.196	4.369	12.424	3.600	10.211

[1] Refers to data sets of Table 3.1, Bates and Watts (1980).
[2] Subtracting the average x from the exponent: e.g. $\theta_1\exp(\theta_2 x)$ becomes $\beta_1\exp(\beta_2(x-\bar{x}))$, so $\beta_2=\theta_2$, $\beta_1=\theta_1\exp(\theta_2\bar{x})$.
[3] Taking logarithms of the exponent parameter: e.g. $\theta_1\exp(\theta_2 x)$ becomes $\beta_1\exp\{(\exp\beta_2)x\}$, so $\beta_1=\theta_1$, $\beta_2=\ln(\theta_2)$.
[4] Indicates the transformation does not apply.

Meeter, 1965): for example, Table 1 shows the effect of applying two such transformations to models with exponential terms. In most cases the parameter effects are decreased, but in some cases the parameter effects are actually increased.

Because there is little guidance available as to the choice of a transformation and its effects in a particular situation, it may be necessary to experiment with many transformations. While it is usually possible to reexpress the model in terms of the new parameters and then recalculate all the derivatives and the A array, this would be inefficient, especially when evaluating the effects of a number of transformations. To facilitate evaluation of the effects of a reparameterization, therefore, we derive formulas which express a new parameter-effects array in terms of the original.

Suppose we wish to determine the parameter-effects array A corresponding to a reparameterization in which the new parameters $\boldsymbol{\beta}$ are nonlinear transformations of the θ's,

$$\boldsymbol{\beta}=G(\boldsymbol{\theta}) \tag{3.1}$$

or

$$\beta_i=G_i(\boldsymbol{\theta}),\qquad i=1,2,\cdots,p. \tag{3.2}$$

We assume the inverse transformation is

$$\boldsymbol{\theta}=S(\boldsymbol{\beta}), \tag{3.3}$$

or

$$\theta_i=S_i(\boldsymbol{\beta}),\qquad i=1,2,\cdots,p, \tag{3.4}$$

and write the $p\times p$ Jacobian matrices as $S.$ and $G.$ with elements $\partial S_i/\partial\beta_j$ and $\partial G_i/\partial\theta_j$ respectively. The $p\times p\times p$ second derivative arrays are written as $S..$ and $G..$ with elements $\partial^2 S_i/\partial\beta_j\partial\beta_k$ and $\partial^2 G_i/\partial\theta_j\partial\theta_k$ respectively: a term with subscripts i, j, k resides in the ith face, jth row and kth column.

Using the chain rule for differentiation, the new tangent vectors at the least squares estimates $\hat{\boldsymbol{\beta}}=G(\hat{\boldsymbol{\theta}})$ are

$$\mathbf{b}_i=\partial\boldsymbol{\eta}/\partial\beta_i\,|_{\hat{\beta}}=\textstyle\sum_{j=1}^{p}(\partial\boldsymbol{\eta}/\partial\theta_j\,|_{\hat{\theta}})(S.)_{ji}=\sum_{j=1}^{p}\mathbf{v}_j(S.)_{ji}, \tag{3.5}$$

where $\mathbf{v}_j$ is the jth column of $V_.$. Equivalently, we have

$$B_. = (\mathbf{b}_1, \mathbf{b}_2, \cdots, \mathbf{b}_p) = V_. S_{..} \tag{3.6}$$

Similarly the new second derivative vectors are

$$\begin{aligned}\mathbf{b}_{ij} &= \partial^2\eta/\partial\beta_i\partial\beta_j \,|\, \hat{\beta} \\ &= (\partial/\partial\beta_j) \textstyle\sum_{r=1}^{p} \mathbf{v}_r (S_.)_{ri} \\ &= \textstyle\sum_{r=1}^{p} \mathbf{v}_r (S_{..})_{rij} + \sum_{r=1}^{p} \sum_{s=1}^{p} (S_.)_{ri} \mathbf{v}_{rs} (S_.)_{sj},\end{aligned}$$

where $\mathbf{v}_{rs}$ is the rsth vector in the array $V_{..}$. Thus we may write

$$B_{..} = [V_.][S_{..}] + S_.' V_{..} S_{..}$$

Now $A = [U_.'][U_{..}]$, where $U_. = V_. L$, $U_{..} = L' V_{..} L$ and L is chosen so that $U_.' U_. = I$. We may likewise orthonormalize $B_.$ by M, say, to give $W_. = B_. M$ subject to $W_.' W_. = I$. One suitable choice is $M = S_.^{-1} L$ so that

$$A = [W_.'][W_{..}] = [(B_. M)'][M' B_{..} M]$$

and from (3.6) and (3.7),

$$\begin{aligned}\tilde{A} &= [L'(S_.^{-1})' S_.' V_.'][L'(S_.^{-1})'([V_.][S_{..}] + S_.' V_. S_.) S_.^{-1} L] \\ &= [U_.'][L'(S_.^{-1})'[V_.][S_{..}] S_.^{-1} L] + [U_.'][L' V_{..} L] \\ &= [U_.'][L'(S_.^{-1})'[V_.][S_{..}] S_.^{-1} L] + A. \qquad (3.8)\end{aligned}$$

But the square bracket and regular multiplications commute, so the first term in (3.8) can be written

$$[U_.' V_.][L'(S_.^{-1})' S_{..} S_.^{-1} L] = -[L^{-1}][L' T L] \tag{3.9}$$

where we define

$$T = -(S_.^{-1})' S_{..} S_.^{-1}. \tag{3.10}$$

Since this is in the form of a curvature, that is acceleration/(velocity)^2, we refer to T as a transformation curvature array.

Thus, the new array equals the original array minus an adjustment,

$$\tilde{A} = A - [L^{-1}][L' T L]. \tag{3.11}$$

This result was derived independently by Clarke (1980). But the term T is given in terms of derivatives with respect to $\boldsymbol{\beta}$, whereas it would be more convenient to have it in terms of the original parameters $\boldsymbol{\theta}$. By writing

$$\boldsymbol{\beta} = G(S(\boldsymbol{\beta}))$$

and differentiating with respect to $\boldsymbol{\beta}$ twice, it is easy to show that

$$S_.' G_{..} S_. = -[G_.][S_{..}], \tag{3.12}$$

so, from (3.10),

$$T = [G_.^{-1}][G_{..}]. \tag{3.13}$$

Thus

$$\tilde{A} = A - [L^{-1}][L'[G_.^{-1}][G_{..}] L] \tag{3.14}$$

which is the final result.

Two important points should be noted. First, to determine a new array A it is only necessary to have the original array A and the matrix L from previous calculations, a total of only $p^2(p + 3)/2$ values. Equation (3.14) is therefore an extremely efficient form for computing the effects of different transformations. Second, the term T is expressed in terms of the original parameters $\boldsymbol{\theta}$, which is very helpful when evaluating a transformation because it avoids having to determine the inverse transformation S.

EXAMPLE 3.1. Data set 15 from Bates and Watts (1980), taken from Meyer and Roth (1972), has a model of the form

$$f(\mathbf{x}, \boldsymbol{\theta}) = \theta_1\theta_3 x_1/(1 + \theta_1 x_1 + \theta_2 x_2).$$

At $\hat{\boldsymbol{\theta}} = (3.1320, 15.160, 0.77998)'$, the matrix L is

$$L = \begin{bmatrix} 0.000 & 0.000 & -1.400 \\ 0.000 & 1.047 & -0.314 \\ -0.018 & 0.028 & 0.261 \end{bmatrix}$$

and the array A is

$$A = \begin{matrix} \begin{bmatrix} 0.00 & -0.04 & -0.33 \\ & -0.05 & 0.69 \\ & & 12.77 \end{bmatrix} \\ \begin{bmatrix} 0.00 & -0.02 & 0.00 \\ & -0.03 & 0.10 \\ & & -0.03 \end{bmatrix} \\ \begin{bmatrix} 0.00 & 0.00 & -0.02 \\ & -0.02 & -0.07 \\ & & 0.22 \end{bmatrix} \end{matrix}$$

so that $\Gamma = 12.8$. This is clearly unacceptable since $1/R_{0.05} = 0.23$, so the radius of the 95% confidence disk is more than 50 times the minimum radius of curvature of a parameter curve.

One obvious transformation is to express the product $\theta_1\theta_3$ as a single parameter so $\beta_1 = \theta_1$, $\beta_2 = \theta_2$, $\beta_3 = \theta_1\theta_3$. Then

$$G_{.} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \theta_3 & 0 & \theta_1 \end{bmatrix}$$

and $G_{..}$ is zero except for $(G_{..})_{313} = (G_{..})_{331} = 1.0$. Then

$$\tilde{A} = \begin{matrix} \begin{bmatrix} 0.00 & -0.04 & 0.11 \\ & -0.05 & 0.00 \\ & & -0.05 \end{bmatrix} \\ \begin{bmatrix} 0.00 & -0.02 & 0.00 \\ & -0.03 & 0.10 \\ & & -0.03 \end{bmatrix} \\ \begin{bmatrix} 0.00 & 0.00 & -0.02 \\ & -0.02 & -0.07 \\ & & 0.21 \end{bmatrix} \end{matrix}$$

with $\Gamma = 0.264$. This is a considerable improvement over the original parameterization and reveals that the projection of the joint confidence region onto the (θ_1, θ_3) plane is essentially bounded by a pair of hyperbolae of the form $\theta_1\theta_3 =$ constant. That is, the confidence region is like an elliptical pancake of roughly constant thickness which has been lifted at one end. Since this transformation reduces all the dominant parameter effects and markedly reduces the parameter-effects curvature measure, the confidence region given by a linear approximation in the $\boldsymbol{\beta}$ parameters would be much closer to the actual region than that given by a linear approximation in the original parameters. Furthermore, mapping the good approximate confidence region in $\boldsymbol{\beta}$ through the exact nonlinear inverse transformation should give a superior approximate confidence region in $\boldsymbol{\theta}$.

4. Procedures for obtaining zero parameter-effects.

4.1. *Introduction.* In some cases the form of the model function suggests a transformation to try, such as in the above example, whereas in other cases there may be recommended transformations, as discussed previously. Other approaches can be proposed, based on the development of Section 3: that is, we try to find a reparameterization such that the new array A is zero.

Setting $\tilde{A}$ equal to zero in (3.11) and solving for the resulting "target" transformation curvature array T^* gives

$$T^* = [L][(L')^{-1}AL^{-1}] \tag{4.1}$$

which may be written in terms of the original parameters and derivative matrices as

$$T^* = [(V_.'V_.)^{-1}V_.'][V_{..}], \tag{4.2}$$

again using the relations $U_. = V_.L$, $U_{..} = L'V_{..}L$, and $U_.'U_. = I$. Note that the right-hand side of (4.2) consists of the least squares coefficients of the acceleration vectors regressed on the first derivative vectors. From (3.13), then, the target transformation $\boldsymbol{\beta}^* = G^*(\boldsymbol{\theta})$ should satisfy

$$[G_.^{*-1}][G_{..}^*] = T^* \tag{4.3}$$

or, by premultiplying by $G_.^*$,

$$G_{..}^* = [G_.^*][T^*] \tag{4.4}$$

which expresses the second derivatives of the target transformation in terms of its first derivatives and the acceleration regression coefficients.

The above equations suggest procedures for obtaining transformations which produce zero curvature locally. The first method is to attempt to solve these equations completely generally and so to derive a global solution for G^* using the algebraic form of (4.2) and (4.4). A second approach is to obtain a particular solution to (4.4) in which (4.2) is evaluated at $\hat{\boldsymbol{\theta}}$, that is to solve the system of coupled second order equations

$$G_{..}^* = [G_.^*][T^*(\tilde{\theta})]. \tag{4.5}$$

A third procedure is to solve equation (4.4) with the transformations restricted to a particular form or chosen from a special class of functions, and then for a specific data set, get the particular solution evaluated at $\hat{\boldsymbol{\theta}}$. We consider each of these procedures below and discuss their relative merits in Section 5. For convenience we delete the asterisk on the target transformation and its derivatives, and on the target transformation array.

4.2. *A general procedure.* The general procedure and its basic impracticality can be demonstrated using a one-parameter example. Suppose that the model function is $f(\mathbf{x}, \boldsymbol{\theta}) = \exp(\theta_1 x)$ and n observations are obtained at $x_1, x_2 \dots, x_n$. Then the target transformation array can be written

$$T(\boldsymbol{\theta}) = \textstyle\sum_{t=1}^{n} \{x_t^3 \exp(2\theta_1 x_t)\} / \sum_{t=1}^{n} \{x_t^2 \exp(2\theta_1 x_t)\},$$

and so the general solution is

$$d^2g/d\theta_1^2 = (dg/d\theta_1) \textstyle\sum_{t=1}^{n} \{x^3 \exp(2\theta_1 x_t)\} / \sum_{t=1}^{n} \{x^2 \exp(2\theta_1 x_t)\}.$$

This would appear to be a difficult equation to solve in general, and even more difficulty could be expected with more parameters, since the answer requires solving a set of coupled nonlinear second-order differential equations. In fact, we suspect that there may not be a general solution since such a solution would imply that there is a global transformation of parameters which would reduce the parameter-effects curvature to zero everywhere on the solution locus. In addition, it seems highly likely that the parameters resulting from such a procedure would not have meaningful interpretations, nor would the transformations be invertible; that is, it would not be possible to express the θ's as closed-form functions of the β's. Because of the apparent difficulties inherent in this approach, we do not pursue it further here; nevertheless, exploration in this area may yield interesting and valuable insights into the reparameterization process.

4.3. *Local transformations using* $T(\hat{\boldsymbol{\theta}})$. Transformations which will yield zero parameter-effects curvature near $\hat{\boldsymbol{\theta}}$ can be derived by solving (4.5) in which case a simpler set of linear coupled partial differential equations results. These equations may not have solutions, however, and so they will be of limited usefulness.

To illustrate the approach and the difficulties associated with it, we consider a two parameter model. In this case, each of the transformations β_1 and β_2 must satisfy the equations

$$\partial^2\beta/\partial\theta_1^2 = t_{111}\partial\beta/\partial\theta_1 + t_{211}\partial\beta/\partial\theta_2 \tag{4.6}$$

$$\partial^2\beta/\partial\theta_1\partial\theta_2 = t_{122}\partial\beta/\partial\theta_1 + t_{212}\partial\beta/\partial\theta_2 \tag{4.7}$$

$$\partial^2\beta/\partial\theta_2^2 = t_{122}\partial\beta/\partial\theta_1 + t_{222}\partial\beta/\partial\theta_2. \tag{4.8}$$

Solutions to these equations may be determined by letting

$$g_1 = \partial\beta/\partial\theta_1, \qquad g_2 = \partial\beta/\partial\theta_2, \quad \text{and} \quad \mathbf{g} = (g_1, g_2)'.$$

Then (4.6), (4.7) and (4.8) can be written

$$\partial\mathbf{g}/\partial\theta_1 = T_1\mathbf{g}, \tag{4.9}$$

$$\partial\mathbf{g}/\partial\theta_2 = T_2\mathbf{g}, \tag{4.10}$$

where T_i has elements t_{ijk}, $i, j, k = 1, 2$. Differentiating (4.9) with respect to θ_2 and substituting (4.10) while differentiating (4.10) with respect to θ_1 and substituting (4.9), and exploiting the fact that the order of differentiation is interchangeable, implies that the solution must satisfy

$$(T_1T_2 - T_2T_1)\mathbf{g} = \tilde{T}\mathbf{g} = 0. \tag{4.11}$$

Now (4.11) will have no solutions if the rank of $\tilde{T}$ is 2, will have a solution (but one which cannot satisfy boundary conditions) if the rank is 1, and will have proper solutions if the rank is 0. In this last case, $T_1T_2 = T_2T_1$ and the solution is of the form

$$\mathbf{g} = \begin{pmatrix} b_1\{\exp(m_1\theta_1 + m_2\theta_2) - 1\} \\ b_2\{\exp(k_1\theta_1 + k_2\theta_2) - 1\} \end{pmatrix},$$

where m_1, m_2 and k_1, k_2 are the eigenvalues of T_1 and T_2 respectively, and b_1 and b_2 satisfy

$$\begin{pmatrix} m_1 & m_2 \\ k_1 & k_2 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The condition $T_1T_2 = T_2T_1$ is a severe one, however, and one which we suspect will be rarely satisfied.

The method may be extended to higher order models, but the solution will have to satisfy multiple conditions $T_iT_j = T_jT_i$, $i, j = 1, 2, \ldots, p$, $i \neq j$, and hence there is little likelihood of any solutions existing. In addition, where solutions exist, the new parameters will be exponential combinations of $\boldsymbol{\theta}$ and will almost surely have no meaning to the experimenter or the statistician. The transformations will generally not be invertible either. For these reasons, we are reluctant to pursue this approach further at this time; nevertheless, they do suggest that there is a restricted class of transformations required so that perhaps subsets of them may be used effectively. It can also be seen, by referring to Section 2 and to Bates (1978), that these transformations compensate specifically for compansion and fanning parameter-effects nonlinearities.

Possibly the most important observation to be made, though, is that analytic solutions may exist which provide zero parameter curvature near $\hat{\boldsymbol{\theta}}$ and hence that there is a parameter metric in which the linear approximation should be accurate under the assumption that higher-order derivatives are smaller than the second-order ones. In other words, while Beale (1960) pointed out that one can envisage a parameterization for which there are no second-order parameter effects curvatures, the above result provides an explicit solution.

4.4. *Restricted transformations.* The third procedure we consider is that of selecting a transformation of restricted form and "tuning" it to produce small curvatures for the particular data set at hand. As an example, we consider the expected-value transformations suggested by Ross (1970): in this approach, the expected values of the model function at particular design points are selected as parameters for the model. However, the effectiveness of such a reparameterization will depend substantially on the choice of design points for which the expected value is used. In our procedure, it is these design points which are tuned so as to produce small curvatures.

This procedure is best described by means of an example, so we consider the Michaelis-Menten model of Section 2.4, with

$$f(x, \boldsymbol{\theta}) = \theta_1 x/(\theta_2 + x),$$

in which the parameter θ_1 occurs conditionally linearly (that is, $\partial f/\partial\theta_1$ does not depend on θ_1). The transformation to expected value parameters is obtained by choosing values r and s and setting

$$\beta_1 = \theta_1 r/(\theta_2 + r), \qquad (4.12)$$

$$\beta_2 = \theta_1 s/(\theta_2 + s).$$

For this transformation the array $[G_{\cdot}^{-1}][G_{\cdot\cdot}]$ is

$$T = \begin{matrix} \begin{pmatrix} 0 & 0 \\ & -2\theta_1/\{(\theta_2 + r)(\theta_2 + s)\} \end{pmatrix} \\ \begin{pmatrix} 0 & 1/\theta_1 \\ & -2(2\theta_2 + r + s)/\{(\theta_2 + r)(\theta_2 + s)\} \end{pmatrix} \end{matrix} \qquad (4.13)$$

while the transformation target array $T^*(\hat{\boldsymbol{\theta}}) = [(V_{\cdot}'V_{\cdot})^{-1}V_{\cdot}'][V_{\cdot\cdot}]$

$$T^*(\hat{\boldsymbol{\theta}}) = \begin{matrix} \begin{pmatrix} 0 & 0 \\ & a \end{pmatrix} \\ \begin{pmatrix} 0 & 1/\hat{\theta}_1 \\ & b \end{pmatrix} \end{matrix} \qquad (4.14)$$

where a and b are constants determined by the particular experimental design and parameter estimates. They are in fact the coefficients of $\mathbf{v}_{22}$ regressed on $\mathbf{v}_1$ and $\mathbf{v}_2$. Equating T and $T^*(\hat{\boldsymbol{\theta}})$ and solving for r and s gives

$$r = -\hat{\theta}_2 + [b - \{b^2 + 8(a/\hat{\theta}_1)\}^{1/2}]/2(a/\hat{\theta}_1), \qquad (4.15)$$

$$s = -\hat{\theta}_2 + [b + \{b^2 + 8(a/\hat{\theta}_1)\}^{1/2}]/2(a/\hat{\theta}_1).$$

Under the transformation (4.12) subject to (4.15), $\tilde{A} = 0$ and so the linear approximation confidence region in $\boldsymbol{\beta}$ will be good: mapping the linear approximation region into the $\boldsymbol{\theta}$ plane should therefore give an accurate approximate region in $\boldsymbol{\theta}$.

Using the data from Section 3.1 of Bates and Watts (1980) again, we find that $r = 0.396$ and $s = 1.945$ (values that are very close to the design points $x = 0.4$ and $x = 2.0$). The exact 95% confidence region (solid line) and the conservative region derived from the linear approximation in the $\boldsymbol{\beta}$ parameters using this r and s (short dashes) are shown in Figure

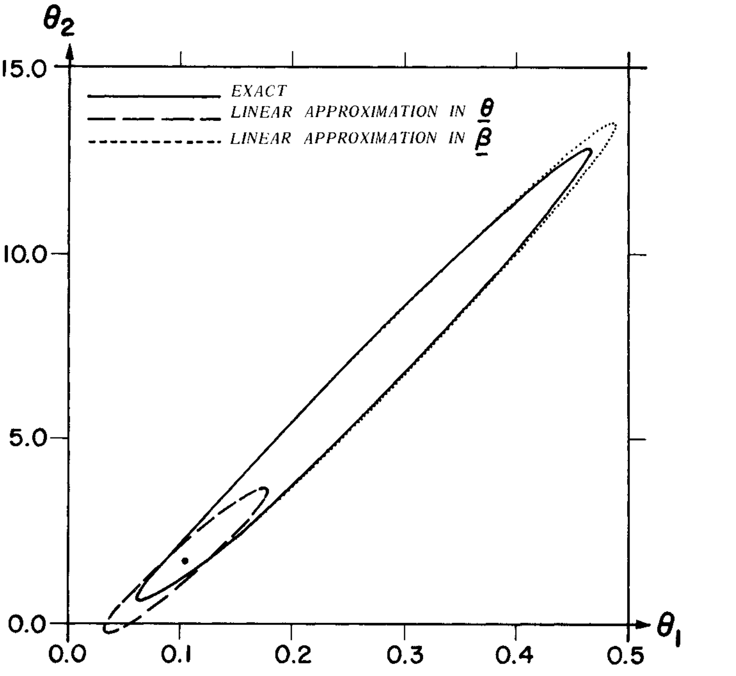

FIG. 3. Exact and approximate 95% confidence regions for the Michaelis-Menten example.

3. The elliptical region from the linear approximation in the $\boldsymbol{\theta}$ parameters (long dashes) is seen to be very poor compared to the transformed linear approximation in the $\boldsymbol{\beta}$ parameters. In addition, the linear approximation confidence region is much simpler and less expensive to produce than the exact region. We note, for interest, that in this case $T_1 T_2 - T_2 T_1$ is of full rank and hence no solution to the differential equation (4.5) exists.

A similar approach can be used for any two-parameter model in which one of the parameters is conditionally linear because $T^*(\hat{\boldsymbol{\theta}})$ and $[G_{\cdot}^{-1}][G_{\cdot\cdot}]$ will always be of the form (4.13) and (4.14), and so an expected-value transformation can always be found which will render the parameter-effects curvature zero.

With higher order models it becomes more difficult to use expected-value parameters since transformation back to the original parameters may not be explicit. In some cases it may be possible to restrict the expected-value transformations to make them invertible. For example, in the three parameter model $f(x, \boldsymbol{\theta}) = \theta_1 + \theta_2 \exp(\theta_3 x)$, the expected-value transformation is not invertible (that is, the inverse does not have a closed form) but it can be made so if the three values of x are chosen to be equally spaced as suggested by Ross (1978). The parameters $\boldsymbol{\beta}$ would then be

$$\beta_1 = \theta_1 + \theta_2 \exp\{\theta_3(s - t)\},$$

$$\beta_2 = \theta_1 + \theta_2 \exp(\theta_3 s),$$

$$\beta_3 = \theta_1 + \theta_2 \exp\{\theta_3(s + t)\}.$$

While this transformation will not eliminate parameter-effects completely, it can substantially reduce them.

5. Discussion. In this paper we have discussed the problems of reparameterizing nonlinear models so as to obtain accurate linear approximation confidence regions for the parameters. A result is derived which expresses the parameter-effects $\tilde{A}$ under a reparameterization $\boldsymbol{\beta} = G(\boldsymbol{\theta})$ in terms of the parameter-effects array A for the original parameters

$\boldsymbol{\theta}$ and the first and second derivatives $G.$ and $G..$ of the transformation. This result provides an extremely efficient computational procedure for determining the effects of different transformations and, as shown in Section 4, can be used to determine reparameterizations for which the new parameters will have zero parameter-effects curvature. This implies that in the new parameters the linear approximation confidence region should be accurate, under the assumption that higher-order derivatives are small. In particular, we show that the expected-value transformations of Ross (1970, 1978) can be used advantageously to produce zero or small parameter-effects curvatures.

We also present other methods for obtaining parameterizations which provide accurate linear approximation confidence regions. The first involves using the A array and compensating for particular nonlinear parameter effects, the second uses an impractical global transformation procedure, and the third requires the solution to a set of coupled partial differential equations. These methods are apparently of little practical use, mainly because they will involve transformations which do not permit ready interpretation of the parameters. The two methods based on partial differential equations also suffer from the disadvantage that in the one case they will be difficult, if not impossible, to solve, and in the second case the solutions may in fact not exist. When solutions do exist in the second case, these solutions will also suffer from basic non-interpretability of the parameters. Nevertheless, they could be used to obtain accurate approximate marginal confidence regions for the individual parameters or joint parameter regions by using a mapping procedure.

Except for the papers of Box (1960), Ross (1970, 1978) and Gillis and Ratkowsky (1978), we are unaware of much work on the use of parameter transformations in multi-parameter, nonlinear regression. The recent paper by Clarke (1980) should be mentioned. This is an important area for research but, of course, the ideal situation would be one in which the experimental design is selected so that the original parameters will themselves have small parameter-effects curvatures. Designing nonlinear experiments for small curvatures is a difficult and challenging problem, and one which clearly is due for attention.

The wider problem of the choice of parameters for a general statistical model can be approached through the concept of statistical curvature introduced by Efron (1975). The intrinsic curvature of the solution locus is just a special case of Efron's statistical curvature and can be related to the extent to which locally most powerful tests are globally powerful. For example, a locally most powerful unbiased level α test of $H:\boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $A:\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ is based upon the length of the component of the residual vector at $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$ parallel to the tangent plane to the solution locus at $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$. A most powerful level α test of $H:\boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $A:\boldsymbol{\theta} = \boldsymbol{\theta}_1$ will be based upon the component of the residual vector at $\boldsymbol{\eta}(\boldsymbol{\theta}_0)$ in the direction of $\boldsymbol{\eta}(\boldsymbol{\theta}_1) - \boldsymbol{\eta}(\boldsymbol{\theta}_0)$ so the extent to which a locally most powerful test retains its power globally is related to the extent to which the tangent plane approximates the solution locus.

Efron (1975) deals with the one-parameter case (although extensions to multi-parameter cases are mentioned in J. Reeds' discussion of that paper) and examination of the parameter effects in a one-parameter situation is relatively easy since the only type of effect possible is compansion. From (4.2) or (5.4) of Efron (1975) it can be seen that the compansion term is

$$a = \nu_{11}(\theta)/(i_\theta)^{3/2}$$

in Efron's notation. This quantity will be related to the local symmetry of the likelihood function and a transformation which reduces this quantity will create a parameter that approaches its asymptotic behavior more rapidly. Under a transformation of the parameter from θ to β this quantity would transform as

$$\tilde{a} = a - (\partial^2\beta/\partial\theta^2)(\partial\beta/\partial\theta)^{-1}(i_\theta)^{-1/2}$$

which is analogous to (3.14) with $\partial\beta/\partial\theta$ for $G.$, $\partial^2\beta/\partial^2\theta$ for $G..$, and $(i_\theta)^{-1/2}$ for L. In fact, the whole concept of measuring both intrinsic and parameter-effects curvatures by an array $A_{..}$ in the multi-parameter situation can be extended to Efron's statistical curvature, but that is the subject of another paper.

Acknowledgements. The authors are grateful to Professor Cedric Schubert, Department of Mathematics and Statistics at Queen's University, for indicating the method of solution of the partial differential equations given in Section 4.3. The work reported here has been supported by the University of Alberta General Research Fund, the Queen's Advisory Research Council, and the Natural Sciences and Engineering Research Council of Canada through research grants.

REFERENCES

BATES, D. M. (1978). Curvature measures of nonlinearity, Ph.D. Thesis, Queen's University, Kingston, Canada.

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

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

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

BOX, G. E. P. (1960). Fitting empirical data. *Ann. N.Y. Acad. Sci.* **86** 792.

CLARKE, G. P. Y. (1980). Moments of the least squares estimators in a non-linear regression model. *J. Roy. Statist. Soc. Ser. B* **42** 227–237.

DRAPER, N. R. and SMITH, H. (1966). *Applied Regression Analysis.* Wiley, London.

EFRON, B. (1975). Defining the curvature of a statistical problem. *Ann. Statist.* **3** 1189–1242.

EFRON, B. (1978). The geometry of exponential families. *Ann. Statist.* **6** 362–375.

EFRON, B. and HINKLEY, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* **65** 457–482.

GILLIS, P. R. and RATKOWSKY, D. A. (1978). The behavior of estimators of the parameters of various yield-density relationships. *Biometrics* **34** 191–198.

GUTTMAN, I. and MEETER, D. A. (1965). On Beale's measures of nonlinearity. *Technometrics* **7** 623–637.

HAMILTON, DAVID C., WATTS, DONALD G. and BATES, DOUGLAS M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions. To appear in *Ann. Statist.*

KENNEDY, WILLIAM J. and GENTLE, JAMES, E. (1980). *Statistical Computing.* Dekker, New York.

MEYER, R. R. and ROTH, P. M. (1972). Modified damped least squares: an algorithm for non-linear estimation. *J. Inst. Math. Appl.* **9** 218–233.

ROSS, G. J. S. (1970). The efficient use of function minimization in non-linear maximum-likelihood estimation. *Appl. Statist.* **19** 205–221.

ROSS, G. J. S. (1978). Exact and approximate confidence regions for functions of parameters in non-linear models. In COMPSTAT 78, Third Symposium on Computation. (Corstein and J. Hermans, eds.) Physica-Verlag, Vienna.

DEPARTMENT OF STATISTICS
COLLEGE OF LETTERS AND SCIENCE
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706

DEPARTMENT OF MATHEMATICS AND STATISTICS
QUEENS UNIVERSITY
KINGSTON, ONTARIO K7L 3N6