Computational Experience With Confidence Regions and Confidence Intervals for Nonlinear Least Squares

Janet R. Donaldson and Robert B. Schnabel

Center for Applied Mathematics
National Bureau of Standards
Boulder, CO 80303-3328

Department of Computer Science
University of Colorado
Boulder, CO 80309

We present the results of a Monte Carlo study of the leading methods for constructing approximate confidence regions and confidence intervals for parameters estimated by nonlinear least squares. We examine three variants of the linearization method, the likelihood method, and the lack-of-fit method. The linearization method is computationally inexpensive, produces easily understandable results, and is widely used in practice. The likelihood and lack-of-fit methods are much more expensive and more difficult to report. In our tests, both the likelihood and lack-of-fit methods perform very reliably. All three variants of the linearization method, however, often grossly underestimate confidence regions and sometimes significantly underestimate confidence intervals. The linearization method variant based solely on the Jacobian matrix appears preferable to the two variants that use the full Hessian matrix because it is less expensive, more numerically stable, and at least as accurate. The Bates and Watts curvature measures are consistent with our results.

1. INTRODUCTION

This article presents the results of an empirical study comparing several methods for constructing confidence regions and confidence intervals about parameters estimated by nonlinear least squares. The methods compared are the lack-of-fit method, the likelihood method, and three variants of the linearization method.

The need for confidence regions and intervals commonly arises in data-fitting applications, where a response variable $y_i$ observed with unknown error $e_i$ is fit to $m$ fixed predictor variables $x_i$, using a function $f(x_i; \theta)$ that can be either linear or nonlinear in the $p$ parameters $\theta$. The function $f(x_i; \theta)$ is linear in $\theta$ if it can be written as

$$f(x_i; \theta) = x_i^T \theta = \sum_{j=1}^{p} x_{ij} \theta_j, \quad i = 1, \ldots, n.$$  

Otherwise it is nonlinear. The methods analyzed in this study are identical when $f(x_i; \theta)$ is linear in $\theta$; otherwise they are not.

When the error $e_i$ is additive, the response variable can be modeled by

$$Y = F(\hat{\theta}) + \hat{e},$$

where $Y$ denotes a column vector with $i$th component $y_i$, $F(\hat{\theta})$ denotes a column vector with $i$th component $f(x_i; \hat{\theta})$, $\hat{e}$ denotes a column vector with $i$th component $\hat{e}_i$, and $\hat{\theta}$ denotes the true but unknown value of the parameters. The least squares estimator of $\theta$ is the parameter value, denoted by $\hat{\theta}$, that minimizes the sum of the squares of the residuals, where the residuals $r_i(\theta)$, are given by

$$r_i(\theta) = y_i - f(x_i; \theta).$$

Thus

$$\hat{\theta} = \arg \min \ S(\theta),$$

where $S(\theta)$ is the residual sum of squares:

$$S(\theta) = \sum_{i=1}^{n} r_i(\theta)^2 = R(\theta)^T R(\theta),$$

with $R(\theta) = Y - F(\theta)$ denoting a column vector with $i$th component $r_i(\theta)$, and $R(\theta)^T$ denoting the transpose of $R(\theta)$.

In our study, we assume that the model is correct.
and that the errors are normal, iid random variables with mean 0 and variance $\sigma^2$—that is, distributed as $N(0, \sigma^2I)$. Then the least squares estimator $\hat{\theta}$ is the maximum likelihood estimator of the parameters $\theta$ of the $p$-variate normal density function.

$$L(Y) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2}\frac{1}{\sigma^2}||Y - F(\theta)||^2\right).$$

In practice, the estimated values of the parameters $\hat{\theta}$ will not equal the true values $\theta$ because of the random errors, $\epsilon_i$, in the data. Since $\theta$ is a random variable, however, it may be possible to indicate with some specific confidence level $1 - \alpha$ in what region about $\hat{\theta}$ we might reasonably expect $\theta$ to be. Such regions are known as $100(1 - \alpha)%$ confidence regions. A joint confidence region for all of the parameters is defined using a function

$$CR_{\alpha}: Y \rightarrow \text{a region in } R^p$$

that satisfies

$$\Pr[\theta \in CR_{\alpha}(Y)] = 1 - \alpha.$$ 

Similarly, a confidence interval for an individual parameter $\theta_j$ is defined using a function

$$CI_{\alpha,j}: Y \rightarrow \text{an interval in } R$$

that satisfies

$$\Pr[\theta_j \in CI_{\alpha,j}(Y)] = 1 - \alpha.$$ 

It is also desirable to construct confidence intervals and confidence regions with the smallest possible size such that the preceding conditions are satisfied. "Good" confidence intervals and confidence regions are often based on minimal sufficient statistics.

Various methods have been proposed for calculating confidence regions and intervals for parameters estimated by nonlinear least squares. These include several variants of the linearization method, the likelihood method, and the lack-of-fit method, as well as the jackknife and bootstrap methods (see, e.g., Bard 1974; Draper and Smith 1981; Efron 1982; Gallant 1976). This article compares three variants of the linearization method, the likelihood method, and the lack-of-fit method. We review these methods briefly in Section 2. The jackknife and bootstrap methods are computationally intensive resampling methods that entail the solution of a large number of related parameter estimation problems. We have not included these two methods in this article. Readers are referred to Efron and Gong (1983) for a comparison of the jackknife and bootstrap methods, and to Duncan (1978) for a comparison of the jackknife and likelihood methods.

Methods that, for all functions $f(x_i; \theta)$ and confidence levels $1 - \alpha$, are statistically guaranteed to contain the true value $100(1 - \alpha)%$ of the time are called exact; all other methods are called approximate. All of the methods analyzed in this article are equivalent and exact for linear models. For nonlinear models, only the lack-of-fit method for computing confidence regions is exact; the other methods for computing confidence regions and all of the methods for computing confidence intervals are approximate. The linearization regions and intervals appear to be the most approximate for nonlinear models, but they are also far less expensive to compute than the likelihood or lack-of-fit regions and intervals, and they are the predominant methods implemented in production software. Some nonlinear least squares packages, including NL2SOL (Dennis, Gay, and Welsch 1981), include three variants of the linearization method, which differ only in that the variance-covariance matrix of the estimated parameters is approximated in three different ways, namely

$$\hat{\Sigma}_{\theta} = s^2(J(\hat{\theta})^TJ(\hat{\theta}))^{-1},$$

$$\hat{\Sigma}_{\theta} = s^2H(\hat{\theta})^{-1},$$

or

$$\hat{\Sigma}_{\theta} = s^2H(\hat{\theta})^{-1}(J(\hat{\theta})^TJ(\hat{\theta}))H(\hat{\theta})^{-1},$$

where $s^2 = S(\hat{\theta})/(n - p)$ is the estimated residual variance, $J(\hat{\theta})$ is the Jacobian matrix of $F(\theta)$ at $\hat{\theta}$ (i.e., the $n \times p$ matrix with $(i,j)$th element $J_{i,j}(\theta) = \partial f(x_i; \theta)/\partial \theta_j$ evaluated at $\hat{\theta}$), and $H(\hat{\theta})$ is the Hessian matrix of $S(\theta)$ at $\hat{\theta}$ (i.e., the $p \times p$ matrix with $(j,k)$th element $H_{j,k}(\theta) = \partial^2 S(\theta)/\partial \theta_j \partial \theta_k$ evaluated at $\hat{\theta}$).

Sections 3–6 of this article describe and analyze a Monte Carlo study that compares all of these methods for computing confidence regions and intervals on 20 nonlinear models. The study is used to estimate how often the true parameter values are contained in the confidence regions and confidence intervals constructed using a given method. The actual percentage of the nominally $100(1 - \alpha)%$ confidence regions and intervals that are found to contain the true value is known as the observed coverage. The observed coverage will generally depend on the method used to construct the confidence regions and confidence intervals; on the nominal confidence level, $1 - \alpha$; on the degree of nonlinearity of the function, $f(x_i; \theta)$; and, to a small extent, on the number of replications in the simulation. If the experiment used to generate the data is repeated a large number of times under the same conditions, and if $CR_{\alpha}$ and $CI_{\alpha,j}$ are exact and the model is correct, then the observed coverage will approach the nominal coverage. When $CR_{\alpha}$ and $CI_{\alpha,j}$ are only approximate, the observed coverage will not necessarily approach the nominal coverage, although one would hope that the difference between the observed and nominal coverage for a reasonable approximate method would be small for most functions.
No similar study of this magnitude appears to have been reported previously. The properties of confidence regions and confidence intervals computed using the linearization, likelihood or lack-of-fit methods have been analyzed by several authors, including Bates and Watts (1980), Beale (1960), Cook and Witmer (1985), Duncan (1978), Gallant (1976), Guttman and Meeter (1965), and Jennrich (1969). The literature includes numerous warnings regarding the possible inaccuracy of the approximate methods, but it contains little empirical data to illustrate the size of the discrepancies between observed and nominal coverage that might be expected. In those studies that do contain empirical data on confidence regions and intervals, the largest reported differences between the observed and nominal coverage is only 9% for a 95% confidence region computed using the linearization method; it is even smaller for the likelihood method (Gallant 1976). In many practical applications, potential differences of 9% might not be cause for concern. Evidence of much larger differences, however, would indicate the need for improved methods. Our results provide such evidence.

Our Monte Carlo study has several purposes. First, we wish to determine whether the observed coverage of the linearization method is significantly affected by how the variance–covariance matrix is computed. Second, we wish to determine whether the approximate confidence regions and confidence intervals constructed using the linearization and likelihood methods and the approximate confidence intervals constructed using the lack-of-fit method have observed coverage significantly different from nominal. In particular, we want to know whether the frequently used linearization method is significantly better or worse than the more expensive likelihood and lack-of-fit methods. Section 3 describes how we designed our study to answer these questions. The results are presented and discussed in Section 4. We have also investigated how effective the diagnostics of Bates and Watts (1980) are in predicting when the confidence regions produced by the linearization and likelihood methods should be reliable; this part of the study is the subject of Section 5.

Our study is oriented toward nonlinear least squares software developers who need assurance that the methods they implement are reasonable for a wide variety of problems. We make only the customary assumptions that the model is correct and that the errors are normally distributed. We do not assume that we can change the representation of the functions analyzed. Readers interested in using reparameterization to improve their results are referred to Ratkowsky (1983) and Kass (1984).

The conclusions that we draw from this study are presented in Section 6. The first conclusion is that among the variants of the linearization method, the one using $\bar{V}_a$ is the best choice because it is the cheapest, and for each data set tested it gives results that are never considerably worse and are sometimes considerably better than the other variants. The second conclusion is that even the best linearization method can be very poor; confidence regions with observed coverage as low as 12.4% for a nominal 95% region and confidence intervals with observed coverage as low as 75.0% for a nominal 95% interval are reported. In contrast, for each of the data sets tested, the confidence regions and confidence intervals constructed using the likelihood method and lack-of-fit methods are close to nominal. Finally, our study indicates that the diagnostics of Bates and Watts (1980) appear quite successful at predicting when linearization confidence regions will be poor. Our recommendations as to how nonlinear least squares software should calculate confidence regions and intervals, in light of these conclusions, are also given in Section 6.

2. BACKGROUND

This section briefly discusses methods for constructing confidence regions and confidence intervals. First, we give a very quick survey of confidence regions and confidence intervals for linear least squares. Next, we describe the two different ways function nonlinearity can affect the solution locus. Then we review the linearization, likelihood, and lack-of-fit methods for constructing confidence regions and confidence intervals when the model is nonlinear. For a more complete discussion, see Bard (1974), Donaldson (1985), Draper and Smith (1981), and Gallant (1976).

Linear Least Squares

When $f(x_i; \theta)$ is linear in the parameters $\theta$, then $f(x_i; \theta) = x_i^{T}\theta$. Consequently, the Jacobian matrix of $F(\theta)$ is $X$, an $n \times p$ matrix with ith row $x_i^{T}$. If we assume that $X$ is of full rank, then $X^{T}X$ is nonsingular and the linear least squares estimators can be expressed in closed form by

$$\hat{\theta} = (X^{T}X)^{-1}X^{T}y.$$ 

When $\tilde{e} \sim N(0, \sigma^2I)$, a 100(1 – $\alpha$)% confidence region for $\theta$ contains those values $\theta$ for which

$$S(\theta) - S(\hat{\theta}) \leq s^2pF_{p, n-p, 1-\alpha}.$$  

\[(2.1)\]

TECHNOMETRICS, FEBRUARY 1987, VOL. 29, NO. 1
Equation (2.1) is equivalent to
\[(\theta - \hat{\theta})^T X^T X (\theta - \hat{\theta}) \leq s^2 p F_{p,n-p,1-\alpha} \] (2.2)
for all linear models, showing that the shape of the confidence regions for \( \hat{\theta} \) is ellipsoidal for all linear models.

A 100(1 - \alpha)% confidence interval for \( \theta_j \) contains those values \( \theta_j \) for which
\[|\theta_j - \hat{\theta}_j| \leq s \sqrt{\left( X^T X \right)_{jj}^{-1} B_{n-p,1-\alpha/2}},\] (2.3)
where \((X^T X)_{jj}^{-1}\) is the \((j,j)\)th element of the inverse of \(X^T X\). The limits of this confidence interval can be shown to be those values \( \theta_j \) that maximize \((\theta_j - \hat{\theta}_j)^2\) subject to
\[S(\theta) - S(\hat{\theta}) = s^2 (t_{n-p,1-\alpha/2})^2 = s^2 F_{n-p,1-\alpha/2}.\] (2.4)

### Nonlinearity and the Solution Locus

The solution locus, or estimation space, of \( f(x_i; \theta) \), \( i = 1, \ldots, n \), consists of all points with coordinates expressible as
\[f(x_1; \theta), f(x_2; \theta), \ldots, f(x_n; \theta),\]
where the \( x_i \) (\( i = 1, \ldots, n \)) are the fixed values of the predictor variables and \( \theta \) is allowed to vary over all possible values of the \( p \) unknown parameters. The solution locus is planar if there exists a reparameterization of \( f(x_i; \theta) \) that makes the function linear in the \( p \) parameters. Otherwise, the solution locus is curved.

A coordinate grid on the solution locus can be formed by tracing the paths obtained when each parameter is individually allowed to vary while all other parameters are held fixed. The coordinate grid is curvilinear whenever the function \( f(x_i; \theta) \) is nonlinear in one or more of its parameters. It is linear only when the function itself is linear.

Curvature of the solution locus is called “intrinsic” curvature (Bates and Watts 1980; Beale 1960). Curvature of the coordinate grid is called “parameter-effects” or simply “parameter” curvature (Bates and Watts 1980). Intrinsic curvature is not affected by reparameterization; parameter-effects curvature is. Linear functions have zero parameter-effects curvature and zero intrinsic curvature. Nonlinear functions always have nonzero parameter-effects curvature and can have either zero or nonzero intrinsic curvature—that is, a planar or curved solution locus, respectively.

### Nonlinear Least Squares

When the function is nonlinear, the least squares estimators of the parameters cannot in general be expressed in closed form and must instead be computed by iterative techniques. Construction of exact confidence regions and confidence intervals also is much more difficult, so approximate methods are frequently used. The following gives a brief description of the leading methods: linearization, likelihood, and lack-of-fit.

#### Linearization Methods

Linearization methods for constructing confidence regions and confidence intervals assume that the nonlinear function can be adequately approximated by an affine, or linear, approximation to the function at the solution. That is, this method assumes that the solution locus is planar and that the coordinate grid is linear throughout the area to be covered by the confidence regions and confidence intervals. Under this assumption, linear least squares theory tells us that the confidence region for \( \theta \) consists of those values \( \theta \) for which
\[(\theta - \hat{\theta})^T \hat{V}^{-1} (\theta - \hat{\theta}) \leq s^2 p F_{p,n-p,1-\alpha/2},\]
whereas a confidence interval for \( \theta_j \) (\( j = 1, \ldots, p \)) consists of those values \( \theta_j \) for which
\[|\theta_j - \hat{\theta}_j| \leq \hat{\nu}_{jj}^{1/2} t_{n-p,1-\alpha/2},\]
where \( \hat{\nu} \) is the estimated variance-covariance matrix of the parameters and \( \hat{\nu}_{jj} \) is the \((j,j)\)th element of \( \hat{\nu} \).

Three approximations to \( \hat{\nu} \) are frequently used. These are
\[\hat{\nu}_a = s^2 (J(\hat{\theta})^T J(\hat{\theta}))^{-1},\] (2.5)
\[\hat{\nu}_b = s^2 H(\hat{\theta})^{-1},\] (2.6)
and
\[\hat{\nu}_c = s^2 H(\hat{\theta})^{-1} (J(\hat{\theta})^T J(\hat{\theta})) H(\hat{\theta})^{-1},\] (2.7)
where \( J(\theta) \) is the Jacobian matrix of \( F(\theta) \) at \( \hat{\theta} \); \( H(\theta) \) is the Hessian matrix of \( S(\theta) \) at \( \hat{\theta} \); and \( s^2 \) is the residual variance, \( s^2 = S(\hat{\theta})/(n - p) \).

Approximation (2.5) is the most common approximation to \( \hat{\nu} \). It is computed by approximating \( F(\theta) \) by the affine approximation around \( \hat{\theta} \),
\[F(\theta) \approx F(\hat{\theta}) + J(\hat{\theta})(\theta - \hat{\theta}),\]
and then directly applying the linear least squares theory.

Approximation (2.6) can be obtained using maximum likelihood theory. For large samples, maximum likelihood estimators are asymptotically distributed as the \( p \)-variate normal with variances and covariances given by \( \hat{\nu} \), where
\[\hat{\nu}^{-1} = -E \left( \frac{\partial^2 \log L(Y)}{\partial \theta_j \partial \theta_k} \right) \]
evaluated at \( \hat{\theta} \). It is straightforward to show that \( \hat{\nu}_b^{-1} \) approaches \( \hat{\nu}^{-1} \) as \( n \to \infty \).

Approximation (2.7) can be obtained from sensitivity analysis. If the observations \( Y \) are changed to \( Y + \epsilon \), then, to within terms linear in \( \epsilon \), \( \hat{\theta} \) will be
changed to
\[ \hat{\theta}(\hat{e}) = \hat{\theta} - H(\hat{\theta})^{-1}J(\hat{\theta})^T \hat{e} \]
Solving
\[ \hat{V} = \text{cov} (\hat{\theta}(\hat{e})) = E((\hat{\theta}(\hat{e}) - \hat{\theta})(\hat{\theta}(\hat{e}) - \hat{\theta})^T) \]
yields \( \hat{V} = \hat{V}_0 \).

When certain regularity conditions are met (Jennrich 1969), each of these approximations to \( \hat{V} \) asymptotically will approach the true variance-covariance matrix of the model. Note also that these approximations differ only when
\[ \sum_{i,j} r_{ij}(0) \left( \frac{\partial^2 f(x_i; \theta)}{\partial \theta_j \partial \theta_k} \right) \]
evaluated at \( \hat{\theta} \), is nonzero. In particular, for linear functions, each of these representations of \( \hat{V} \) is equal to
\[ s^2(J(\hat{\theta})^TJ(\hat{\theta}))^{-1} = s^2(X^TX)^{-1}. \]

For nonlinear functions, \( \hat{V}_0 \) is said to use observed information and \( \hat{V}_n \) is said to use expected information. (This difference can also be viewed as the difference between using observed and expected Fisher information.)

Linearization methods have the advantage that their resulting confidence regions and intervals are simple and inexpensive to construct and that they produce bounded, convex confidence regions. In addition, the information needed to construct confidence regions and intervals using this method can be parsimoniously summarized by the \( p \times p \) matrix \( \hat{V} \), and it is well understood by users familiar with linear least squares. Because the linearization methods assume that both the intrinsic curvature and the parameter-effects curvature of \( f(x_i; \theta) \) are 0, however, we expect that the linearization methods could sometimes produce observed coverages very far from the expected nominal coverage. The results of our Monte Carlo study show this to be true.

Likelihood Method. The likelihood method is another approximate method for producing confidence regions and confidence intervals. The likelihood method confidence region for \( \theta \) consists of those values \( \theta \) for which
\[ S(\theta) - S(\hat{\theta}) \leq s^2 F_{p \times p, 1 - \alpha}, \]
This is analogous to Equation (2.1) for confidence regions for the parameters of a linear function, although when \( f(x_i; \theta) \) is nonlinear in the parameters the resulting confidence region is no longer ellipsoidal. The likelihood method confidence interval for \( \theta_j \) is the interval bounded by the points that maxi-

mize \( \theta_j \) subject to
\[ S(\theta) - S(\hat{\theta}) \leq s^2 F_{p \times p, 1 - \alpha}. \]
This confidence interval is the projection onto the appropriate parameter axis of the preceding region and is analogous to Equation (2.4) for confidence intervals in the case of linear least squares.

When the solution locus is planar, the confidence regions (but not necessarily the confidence intervals) constructed using the likelihood method are equivalent to the lack-of-fit confidence regions and, therefore, are exact. In addition, likelihood method confidence regions and intervals have the desirable property that they are constructed from contours of constant likelihood and that the regions and intervals are not affected by reparameterization of the function \( f(x_i; \theta) \). Thus we might expect the likelihood method to produce confidence regions and confidence intervals with observed coverage closer to nominal than those produced using the linearization methods. The likelihood method has several computational disadvantages, however. Both the confidence regions and confidence intervals produced using the likelihood method can be disjoint and unbounded, because the contours of a nonlinear function can be disjoint and unbounded. The method also is computationally expensive, requiring the evaluation of \( F(\theta) \) at a sufficient number of points to produce a contour. Finally, when the data arrays are large, it can be awkward to publish the information necessary to reconstruct the confidence region because this information is not succinctly summarized as it is in the case of the linearization method.

Lack-of-Fit Method. The lack-of-fit method can be used to produce exact joint confidence regions for all \( p \) of the parameters and to produce approximate confidence intervals and confidence regions for subsets of the parameters. Let
\[ P(\theta) = J(\hat{\theta})(J(\hat{\theta})^TJ(\hat{\theta}))^{-1}J(\hat{\theta})^T. \]
The lack-of-fit method is based on the fact that the quadratic forms
\[ Q_1(\hat{\theta}) = P(\hat{\theta})^T P(\hat{\theta}) \frac{R(\hat{\theta})}{\sigma^2} \]
and
\[ Q_2(\hat{\theta}) = P(\hat{\theta})^T (1 - P(\hat{\theta})) R(\hat{\theta}) \frac{1}{\sigma^2} \]
are independent chi-squared random variables with \( p \) and \( n - p \) degrees of freedom, respectively. Therefore,
\[ \frac{Q_1(\hat{\theta})}{Q_2(\hat{\theta})/(n - p)} \]
is distributed as \( F_{p, n - p, 1 - \alpha} \), so an exact 100(1 - \alpha)%
confidenceregionconsistsofallvalues\(\theta\)suchthat

\[
\frac{R(\theta)^T P(\theta) R(\theta)}{R(\theta)^T (1 - P(\theta)) R(\theta)} < \frac{p}{n-p} F_{p,n-p, \gamma, \alpha}
\]

Note that the lack-of-fit method does not require that the least squares solution be found prior to constructing the confidence region.

Similarly, a lack-of-fit method confidence interval for the \(j\)th parameter consists of those values \(\theta_j\) for which there exist values of \(\theta_k\) (\(k = 1, \ldots, j-1, j+1, \ldots, p\)) such that for these \(p\) parameter values, \(\theta\),

\[
\frac{S^2(\hat{\theta}_{j+1}) - S^2(\hat{\theta}_{j-1})}{S^2(\hat{\theta}_{j-1})/(n-p)} \leq F_{1,n-p,1-\gamma,\alpha}
\]

where \(S^2(\hat{\theta}_{j+1})\) is the residual sum of squares obtained when \(R(\theta)\) is linearly fit to all of the columns of \(J(\theta)\) excluding the \(j\)th, and \(S^2(\hat{\theta}_{j-1})\) is the residual sum of squares obtained when \(R(\theta)\) is linearly fit to \(J(\theta)\). This interval is exact if \(f(x_i; \theta)\) is linear in \(\theta_k\) (\(k = 1, \ldots, j-1, j+1, \ldots, p\)); otherwise it is approximate (see Halperin 1963).

The lack-of-fit method is even more computationally expensive than the likelihood method, requiring the evaluation of both \(F(\theta)\) and \(J(\theta)\) at a sufficient number of points to produce a contour. As is the case for the likelihood method, the information needed to construct the confidence regions cannot be succinctly summarized for publication. Moreover, the confidence regions and confidence intervals constructed using the lack-of-fit method are guaranteed to contain every minimum, maximum, and/or saddle point of the likelihood surface (see Gallant 1976). This makes the lack-of-fit method structurally undesirable.

3. THE MONTE CARLO STUDY

This section briefly describes how our Monte Carlo study was conducted. Full details were provided by Donaldson (1985).

The Monte Carlo method uses the computer to simulate the results of repeating an experiment many times to obtain a large sample from which the statistical properties of a system can be examined. The errors, \(e\), for our study were produced using the Marsaglia and Tsang (1984) pseudonormal random number algorithm as implemented by James Blue and David Kahanar of the National Bureau of Standards Scientific Computing Division. The response variable, \(Y\), was constructed with \(ith\) component

\[
y_i = f(x_i; \theta) + \epsilon_i.
\]

The least squares estimate, \(\hat{\theta}\), was calculated using NL2SOL, an unconstrained quasi-Newton code for nonlinear least squares (Dennis et al. 1981). Starting values for NL2SOL were set to the true values of the parameters, \(\theta\), and the stopping criteria for the convergence tests based on the relative change in the parameters and in the sum of squares both were set to \(10^{-5}\).

For each confidence region or interval method and each derivative configuration being analyzed, we recorded whether the true values of the parameters were contained within the confidence regions and confidence intervals for each realization of the data. Determining whether the true parameter values lay within the confidence regions and confidence intervals about the least squares estimates fortunately did not require that we construct the full confidence regions and confidence intervals for each confidence level and method. Instead, we simply calculated the smallest confidence level, \(1 - \omega\), such that a 100(1 - \(\omega\))% confidence region or confidence interval constructed using the method being analyzed would contain the true parameter values. When \(\omega > \alpha\), the true value did not lie in the 100(1 - \(\omega\))% confidence region or confidence interval; when \(\omega \leq \alpha\), it did. The values \(1 - \omega\) were obtained using the hypothesis tests corresponding to the formulas for confidence regions and intervals given in Section 2, and the appropriate cumulative distribution functions; the procedures were described in detail by Donaldson (1985). The cumulative distribution functions were obtained from the STARPAC subprogram library (Donaldson and Tryon 1983).

The observed coverage, \(\gamma_s\), for the particular nominal confidence level, method, and system under analysis is the percentage of the total number of realizations of the data, \(N\), for which \(\omega \leq \alpha\). When \(N\) is large, the standard deviation of \(\gamma_s\) can be approximated using the normal approximation to the binomial distribution. In this study we used \(N = 500\), so the maximum standard deviation of the observed coverage at any coverage level is approximately 2.2%.

Note that substituting a new realization of the data for one that could not be completely analyzed because either (a) the nonlinear least squares algorithm did not converge or (b) the test statistics could not be computed for every method being analyzed, is a form of censoring that will bias the observed coverages obtained. In our analysis, we adjusted the value of \(\delta\) for each data set so that every realization could be completely analyzed: therefore, the results reported in this article are not derived from censored data. But this procedure may have had the effect of reducing the intrinsic curvature of the data sets studied. We will try to remove this limitation in future work.

We computed the observed coverage for four nominal confidence levels, .50, .75, .95, and .99. In this
CONFIDENCE REGIONS FOR NONLINEAR LEAST SQUARES

article we only include results for level .95, although we comment briefly in Section 4 on our results at the other levels. Data for the full study were given by Donaldson (1985).

The references for the data sets used in our Monte Carlo study are given in the Appendix and were described in detail by Donaldson (1985). With only two exceptions, the functions and data that compose our data sets have been taken from Duncan (1978), Guttman and Meeter (1965), Himmelblau (1970), and Ratkowsky (1983). The standard deviation of the errors of some of the data sets has been adjusted to allow us to analyze successfully each realization of the data for each data set. The two data sets not taken from the published literature are identified as 8ACA and 9AAG. Data set 8ACA was created especially for this study by generalizing function 3 to a larger number of parameters. Data set 9AAG involves a microwave absorption line function taken from a consulting session at the National Bureau of Standards in Boulder, Colorado.

The number of parameters in the 20 data sets analyzed ranges from 2 to 8 and the ratio of the number of parameters to the number of observations ranges from 2/42 to 3/5. Although these data sets are often troublesome, they are mostly real world problems that have not been made artificially difficult.

Each data set was analyzed twice to allow us to examine the effect of increasing the standard deviation of the errors. In the first analysis, \( \hat{\theta} \sim N(0, \sigma^2 I) \); in the second analysis, \( \hat{\theta} \sim N(0, (\eta \sigma)^2 I) \), where \( \eta \) is approximately the largest number <10 for which every realization of the data could be successfully analyzed. The methods analyzed in the second analysis were the same as in the first except that variants (2.6) and (2.7) of the linearization method were excluded from the second analysis because, when \( \eta > 1.0 \), we were frequently unable to compute the required test statistics using these two variants.

Computation of the linearization method and the lack-of-fit method requires that certain derivatives be available. The Jacobian matrix of \( F(\theta) \) is used by both the linearization and lack-of-fit methods. Variants (2.6) and (2.7) of the linearization method use the Hessian matrix of \( S(\theta) \) as well. In practice, analytic derivatives often are not available. Therefore, in our study each method was implemented and analyzed using three different derivative configurations. These configurations are (a) the Jacobian matrix and the Hessian matrix, both approximated by finite differences; (b) the Jacobian matrix computed analytically and the Hessian matrix computed by finite differences; and (c) both the Jacobian matrix and the Hessian matrix computed analytically. For derivative configurations (a) and (b), the variance-covariance matrix needed by the linearization method was re-computed directly from NL2SOL (Dennis et al. 1981). For configuration (c), it was constructed outside of NL2SOL. For details on the formulas used to compute the finite-difference derivative approximations, see Donaldson (1985).

We ran our Monte Carlo study in single precision on a 60-bit word-length computer. All subroutines extracted from other sources were used without modification except for NL2SOL, which was changed for this study in two important ways. First, we disabled the two tests within NL2SOL used to detect near singularity. Second, we used the STAR-PAC (Donaldson and Tryon 1983) user interface to NL2SOL. With this interface, the finite difference approximation to the Jacobian matrix is computed with the optimal derivative step sizes selected using the algorithm developed by Schnabel (1982), thus maximizing the number of correct digits in each element of the finite difference Jacobian matrix.

4. RESULTS AND OBSERVATIONS

This section presents the results of our Monte Carlo study of the lack-of-fit method, the likelihood method, and the three variants of the linearization method. The section is divided into a discussion of confidence regions and confidence intervals. For each, we also make a number of observations about the results. The conclusions we draw from our analysis are discussed in the next section.

Confidence Regions

Results. The results for nominally 95% confidence regions constructed using each of the methods analyzed in this study with \( \hat{\theta} \sim N(0, \sigma^2 I) \) are graphically displayed in Figure 1. For each data set, the observed coverage is plotted against the method and derivative configuration used to obtain it.

The three derivative configurations are labeled DC1, DC2, and DC3 in this and the following figures. DC1 denotes use of finite difference approximations for both the Jacobian and the Hessian matrices; DC2 denotes use of the analytic Jacobian matrix and the finite-difference Hessian matrix; and DC3 denotes use of the analytic Jacobian matrix and Hessian matrix. Since the computations required to calculate the lack-of-fit method results and the likelihood method results using derivative configurations DC2 and DC3 are the same, these results are displayed together.

Figure 2 shows the analogous results for \( \hat{\theta} \sim N(0, (\eta \sigma)^2 I) \). As noted in Section 3, variants (2.6) and (2.7) of the linearization method are excluded from the analysis displayed in Figure 2 because computational difficulties were encountered for these variants when the variance of the errors was increased.

A conservative 95% critical region about the nomi-
Figure 1. Observed Coverage for Nominally 95% Confidence Regions With $\hat{\theta} \sim N(0, \sigma^2 I)$ Versus Method, by Derivative Configuration.

The nominal confidence level is indicated on each plot by a pair of horizontal lines that represent the values $100(1 - x) \pm 4.4$, where 4.4 is twice the maximum standard deviation of the observed coverage at any coverage level. This critical region provides a quick means of determining whether any of the observed coverages for each method are significantly different from the nominal confidence level at the 5% level. When the method used to construct the confidence regions and confidence intervals is exact, we expect that the observed coverage for 95% of all possible data sets will lie within this critical region.

Observations. Figures 1 and 2 show that the lack-of-fit and likelihood method confidence regions are quite reliable and that the results are not affected by use of finite difference derivatives. In all of our tests, they produced observed coverages that seldom vary from nominal by an amount that is significant at the 5% level. In fact, for these data sets, there is only one instance [data set 3AAA, $\hat{\theta} \sim N(0, \sigma^2 I)$] when the difference between the nominal and observed coverages produced using these two methods is greater than 5%, and in this instance, the observed coverage is greater than nominal, not less.

The three variants of the linearization method, on the other hand, frequently produced far less reliable confidence regions, although, as discussed subsequently the results still do not appear to be affected by the use of finite-difference derivatives. The difference between the nominal and observed coverages obtained using the linearization methods often are considerably more than 20%, which is a difference that many, if not most, users would find unacceptable. The largest discrepancies were produced by data sets 11AAB, 14ACG, and 15AAA.

By comparing Figure 1 with Figure 2, it is apparent that increasing the variance of the errors does, in fact, increase the differences between observed and nominal coverage for all methods. Our tests at confidence levels .50, .75, and .99, which are not reported in detail here, also showed that the spread between the observed and nominal coverage obtained using the linearization method increases as the nominal confidence level is increased.

The large differences for some data sets between the observed coverage of confidence regions constructed using the likelihood method and those obtained using the linearization method may be explained by the difference in the shape of the two regions. The likelihood method confidence region corresponds to the boundary and interior of a contour of the sum of squares surface—that is, a contour of constant likelihood—whereas the linearization method confidence regions are always ellipsoidal. We plotted these pairs of contours for various data sets, and sometimes the differences were very large. Contours for data sets 3AAA and 14ACG were given by Donaldson (1985). Cook and Witmer (1985) also graphically compared linearization method contours and exact contours for other examples and found similar results.
Figure 2. Observed Coverage for Nominally 95\% Confidence Regions With $\theta \sim N(0, (\theta_i^*)^T)$ Versus Method, by Derivative Configuration.

Figure 1 also indicates that the observed coverage obtained using variants (2.5), (2.6), and (2.7) of the linearization method are nearly identical. The results of two-sided paired-sample $t$ tests indicate that there are no statistically significant differences at the 5\% level between the observed coverages obtained using any of the variants of the linearization method with any of the derivative configurations. The same results were obtained for our tests at the .50, .75, and .99 confidence levels.

Confidence Intervals

Results. Figures 3 and 4 provide information for confidence intervals that is analogous to that shown in Figures 1 and 2 for confidence regions. The observed coverages plotted are the smallest of the $p$
confident interval coverages obtained for each data set. Figure 3 displays the observed confidence interval results for nominally 95% confidence levels when \( \hat{\epsilon} \sim N(0, \sigma^2 I) \); Figure 4 shows the results when \( \hat{\epsilon} \sim N(0, (\eta \sigma)^2 I) \), excluding linearization method variants (2.6) and (2.7) as was done for the linearization method confidence regions.

Observations. Figure 3 shows that for confidence intervals the best results are obtained using the lack-of-fit and likelihood methods and the worst results are obtained using the linearization method, as was the case for confidence regions. The lack-of-fit and likelihood methods produce confidence intervals that seldom vary from nominal by an amount that is significant at the 5% level and never are less than nominal by more than 5.0%. Again, use of a finite difference Jacobian matrix does not appear to affect the results for these two methods.

The three variants of the linearization method, on the other hand, frequently produce far less reliable confidence intervals than the lack-of-fit and likelihood methods. Disturbing differences between observed and nominal coverages occur when each of the variants of the linearization method is used to construct confidence intervals. The observed coverage for a nominally 95% confidence interval is as low as 75.0% (data set 15AAA), 44.0% (data set 9AAG), and 10.8% (data set 9AAG) for variants (2.5), (2.6), and (2.7), respectively. For most of the data sets tested in our study, however, the span between observed and nominal coverage produced by the three variants of the linearization method is considerably less for confidence intervals than for linearization method confidence regions constructed for the parameters of the same data set. This is especially true when derivative configurations DC2 and DC3 are used.

One reason linearization method confidence intervals have better coverage than linearization method confidence regions is that, when the parameter estimates are correlated with each other, a number of points may be included in the linearization method confidence intervals but not in the confidence regions. Note, however, that if a confidence interval were computed for the linear combination of the parameters given by the eigenvector corresponding to the minor axis of the linearization method confidence region ellipsoid, then the linearization method confidence interval observed coverage should approximately equal that of the linearization method confidence region. In our Monte Carlo study, we actually computed the linearization method confidence interval observed coverage for this linear combination of the parameters. In every case, the observed coverage that we obtained for the confidence interval for this linear combination was approximately equal to that of the linearization method confidence region observed coverage.

The use of finite differences to approximate both the Jacobian matrix and the Hessian matrix appears to significantly degrade the confidence interval re-
results for linearization variants (2.6) and (2.7). Figure 3 shows that, although there is no striking difference in the results obtained using the three variants of the linearization method with derivative configurations DC2 and DC3, variants (2.6) and (2.7) degrade significantly more than variant (2.5) when using DC1—that is, finite difference Jacobian and Hessian matrices. A two-sided paired-sample t test was used to determine whether, for a given derivative configuration, the observed coverages obtained using the different linearization method variants are statistically different at the 5% significance level. The results indicate that when derivative configurations DC2 and DC3 are used, the differences in the results obtained using variants (2.5), (2.6), and (2.7) are not statistically significant at the 5% level. When both the Jacobian matrix and the Hessian matrix are approximated using finite differences (derivative configuration DC1), however, the differences between the results using methods (2.5) and (2.6), (2.5) and (2.7) and (2.6) and (2.7) are significant at the 5.6%, 1.4%, and .4% levels, respectively.

Comparing Figures 3 and 4 shows that as the variance of the errors is increased, the differences between observed and nominal coverage also are increased, as was the case for the confidence region results. This increase is not as pronounced for confidence intervals as for confidence regions, however. The results at confidence levels .50, .75, .95, and .99 also showed that as the nominal confidence level approaches 100%, the spread between observed and nominal coverages obtained using the linearization method is increased.

5. DIAGNOSTIC TOOLS

The preceding section demonstrates a pressing need for diagnostics to warn users when the commonly used linearization method confidence region will not have adequate coverage. In addition, it would be useful to have a warning to indicate when the approximate likelihood method may be inadequate. Bates and Watts (1980) proposed measures of nonlinearity that provide such diagnostics.

According to Bates and Watts, when their relative measure of parameter effects curvature is small compared with the critical value \( (F_{p,n-p,0.05})^{-1/2} \), then the linear coordinate grid assumption is valid over the region of interest and, therefore, the linearization method confidence region should be adequate. Similarly, when their relative measure of intrinsic curvature is small compared with the same critical value, then the assumption that the solution locus is planar is valid over the region of interest and, therefore, the likelihood method confidence region should be adequate.

In Figure 5 we plot the 20 confidence region observed coverages obtained using linearization method variant (2.5) with analytic derivatives (derivative configuration DC3) and \( \hat{e} \sim N(0, (\eta\hat{e})^2I) \) against the Bates and Watts relative measure of parameter ef-
effects curvature. Likewise, in Figure 6 we plot the corresponding 20 likelihood method confidence region observed coverages against the Bates and Watts relative measure of intrinsic curvature. The relative curvature measures were computed at the true parameter values using the true variance of the errors. We expect that we would have obtained reasonably similar results had we evaluated the relative curvature measures using the estimated parameters and variance from the least squares solutions, although we have not made these calculations. In these plots, we have scaled the measures of parameter effects curvature and intrinsic curvature by dividing the measure by the appropriate critical value. Thus, in both of these plots, a scaled curvature measure of less than 1 indicates that the relative measure was less than the critical value; a value greater than 1 indicates the curvature exceeded the critical value.

It is clear from Figure 5 that the Bates and Watts parameter effects curvature measure is strongly correlated with the observed coverage obtained using the linearization method. Furthermore, in all three data sets in which the parameter effects curvature is less than the critical value, the observed confidence region is within 3.8% of nominal; in all cases in which the parameter effects curvature is greater than 10 times the critical value, the observed coverage is unsatisfactorily low—that is, between 82.6% and 12.4%. Data sets with parameter effects curvature between 1 and 10 times the critical value had observed confidence region coverage between 83.2% and 91.6%. From these results, it appears that the Bates and Watts parameter effects curvature is a reliable, if perhaps stringent, indicator of when the linearization method will produce reliable confidence regions. Readers are referred to Cook and Witmer (1985) for a more detailed examination of the Bates and Watts parameter effects curvature measure.

Figure 6 shows that all but 1 of the 20 data sets tested in this study have intrinsic curvature that is less than the critical value, which means that each of these data sets is nearly planar. For nearly planar data sets, we expected good observed coverage from the likelihood method, and, as Figure 6 shows, that is what we got. Since none of our data sets has high intrinsic curvature, however, we do not know how the likelihood method will perform when the solution locus is not nearly planar. We cannot assume that the accurate results obtained in our study using the likelihood method will necessarily carry over to data sets with large intrinsic curvature.

Cook, Tsai, and Wei (1984) provided an example that has scaled parameter effects curvature of 934.5 and scaled intrinsic curvature of 8.4. Both the parameter effects curvature and intrinsic curvature of this data set exceed any curvature measure that we observed in the 20 data sets in our study. For this data set, we computed observed confidence region coverages of 19.0% and 95.0% using the linearization method and likelihood methods, respectively. Whereas the linearization method confidence region observed coverage is very far from nominal, as we
would expect based on the parameter effects curvature of this model, the likelihood method confidence region observed coverage is not. We cannot conclude anything from this one observation. It is clear, however, that additional analysis of data sets with high intrinsic curvature would be useful to further assess the effect of a nonplanar solution locus on the likelihood method.

6. CONCLUSIONS

Based on our computational study, we can draw conclusions about (a) the comparison among the three variants of the linearization method, (b) the reliability of linearization methods for calculating confidence regions and confidence intervals, and (c) the reliability of the likelihood and lack-of-fit methods for calculating confidence regions and confidence intervals.

When using the linearization method to construct confidence regions and intervals, our Monte Carlo study has shown no clear-cut difference in the observed coverage of one variant as compared with another. In our tests, the only statistically significant difference among the results produced by the three linearization variants was in constructing confidence intervals with finite difference Jacobian and Hessian matrices; here variant (2.5) was superior to variants (2.6) and (2.7). We found no empirical evidence that one should prefer variants (2.6) or (2.7) even though they may be appealing from a theoretical point of view. Therefore, we conclude that variant (2.5) of the linearization method, which is computed using

\[ \hat{\Sigma}_a = s^2((J(\hat{\theta}))^T J(\hat{\theta}))^{-1}, \] (6.1)

is the best variant to use for constructing both confidence regions and confidence intervals, because it is simpler, less expensive, and more numerically stable to compute than variants (2.6) or (2.7), which use

\[ \hat{\Sigma}_b = s^2 H(\hat{\theta})^{-1}, \] (6.2)

and

\[ \hat{\Sigma}_c = s^2 H(\hat{\theta})^{-1} (J(\hat{\theta}))^T J(\hat{\theta}) H(\hat{\theta})^{-1}, \] (6.3)

respectively. Variant (2.5) is simpler and less expensive because it only requires the Jacobian matrix of the model function at the solution and not the additional second-order terms that are also required to form the Hessian matrix. It is more stable because it can be formed by inverting the upper triangular factor \( R \) of the QR factorization of the Jacobian matrix rather than by calculating the inverse of the Hessian matrix; the former calculation can be expected to lose roughly half as many digits as the latter in finite precision arithmetic (see Stewart 1973).

The linearization method is not always an adequate method for approximating confidence regions and confidence intervals for the parameters of a nonlinear model, however. The results presented in the preceding section show just how poor the linearization method can be in some cases. Although there are many examples in which the linearization method's observed coverage differs from nominal by only a very small amount, there are also many cases in which the observed coverage is far lower than the nominal. In our tests, the best linearization method variant, (2.5), produced observed coverages as low as 12.4% for nominal 95% confidence regions and 75.0% for nominal 95% confidence intervals. These differences appear to be because of the elliptical shape of the linearization method confidence region being a poor approximation to the shape of the true contour.

Users will continue to use the linearization method, however, because it is readily available in software packages and provides a concise representation of the information needed to construct confidence regions and intervals. The erratic results obtained in our study when using the linearization method lead us to conclude that users of nonlinear least squares software must be helped to assess cautiously the results that they obtain using the linearization method. The results of the preceding section show that the diagnostic tools proposed by Bates and Watts (1980) are very likely to be successful in indicating cases in which the linearization method confidence regions are likely to be unreliable. In these cases, more reliable methods, such as the likelihood or lack-of-fit methods, are required to produce accurate confidence regions or intervals.

Our study shows that the lack-of-fit and likelihood methods produce observed coverages acceptably close to nominal in every test case. Although the difficulties and expense associated with using these two methods to compute confidence intervals make it unlikely that they will ever routinely replace the commonly used linearization method for this purpose, they appear to be reliable alternatives that should be considered when diagnostics show that linearization confidence regions are unreliable. It is not as difficult and expensive to construct confidence intervals using the lack-of-fit or likelihood methods, and we believe that producers of nonlinear least squares software should consider this possibility. (Constructing these intervals requires the solution of a series of nonlinearly constrained optimization problems; it may be necessary to construct special purpose software to solve these problems as efficiently as possible.) Performing hypothesis tests using the likelihood or lack-of-fit methods is computationally simple, so we recommend that one of these
two methods be employed for hypothesis tests whenever possible.

Users may prefer the likelihood method to the lack-of-fit method even though it is approximate and the lack-of-fit method is exact, because the likelihood method has more desirable structural characteristics than the lack-of-fit method. Our study provides no empirical evidence that the results produced by the likelihood method are inferior to those produced by the lack-of-fit method. This does not guarantee that similar results will be obtained on other data sets, however. In particular, the results of the diagnostic test proposed by Bates and Watts (1980) showed that all of our data sets have low intrinsic curvature, which is precisely the situation when likelihood methods are expected to be very reliable. The additional data set that we analyzed with high intrinsic curvature also produced likelihood method confidence region observed coverage close to nominal. Additional analysis is required to determine whether the likelihood method is reliable for data sets with high intrinsic curvature and to determine whether the Bates and Watts measure of intrinsic curvature is a useful tool for indicating when the likelihood method confidence regions are likely to be unreliable.

In addition to diagnostics, it appears that there is a need for new methods for estimating confidence regions that are both reliable and easy to report. We are especially interested in investigating two methods that would result in conservative elliptical confidence regions. The first method is to find the minimal magnification of the (95%) linearization confidence region that encloses the (95%) likelihood or lack-of-fit confidence region. This would require the solution of a constrained optimization problem with one nonlinear equality constraint. The second method is to find the smallest volume ellipse that encloses the desired likelihood or lack-of-fit confidence region. This would require the solution of a semi-infinite programming problem—that is, an optimization problem with an infinite set of constraints.

7. SUMMARY

We have presented the results of a Monte Carlo study comparing the linearization, likelihood, and lack-of-fit methods for constructing confidence regions and confidence intervals. Our results indicate that the linearization method should be constructed using the linearization method can be essentially meaningless.

Our study shows that the likelihood and lack-of-fit methods, on the other hand, produced consistently good results for the data sets tested. Because the likelihood method is approximate, however, it is not clear that the good results we obtained with it will necessarily be characteristic of all data sets. Moreover, because of the undesirable structural characteristics of the lack-of-fit method, it is unlikely to be used routinely. In cases in which accuracy is of extreme importance, however, it may be a useful tool to have.

Because of the uncertainty associated with the linearization and likelihood methods, we also have briefly examined how the Bates and Watts (1980) curvature measures relate to the confidence region observed coverages that we obtained in this study. Our results show that the Bates and Watts parameter effects curvature appears to provide an excellent indication of when the linearization method may produce less than satisfactory results. Our results are not as conclusive, however, about the relation between intrinsic curvature and likelihood method coverage, since the solution loci for all of our data sets were nearly planar.

ACKNOWLEDGMENT

This research was supported by Army Research Office Contract DAAG 29-84-K-0140.

APPENDIX

Data set 1 (Identifier 2AAA); p = 2, n = 12.
Reference: Guttman and Meeter (1965), model \( \eta_2 \), page 628.
Function:
\[
y_i = \theta_1 \exp(\theta_2 x_i) - \exp(-\theta_2 x_i) + \hat{e}_i.
\]

Data set 2 (Identifier 3AAA); p = 2, n = 12.
Reference: Guttman and Meeter (1965), model \( \eta_3 \), page 628.
Function:
\[
y_i = 1 - \theta_1 \exp(\theta_2 x_i) - \theta_2 \exp(-\theta_2 x_i) + \hat{e}_i.
\]

Data set 3 (Identifier 4AAB); p = 2, n = 24.
Function:
\[
y_i = 1,000 \theta_2 x_i + \hat{e}_i.
\]

Data set 4 (Identifier 5AAF); p = 4, n = 18.
Function:

\[ y_i = \frac{12\theta_3^2(\beta x_i)^\beta \exp[\theta_3^2(1 - \beta x_i)]}{x_i(2\pi)^{1/2}(12\theta_1 + 1)} + \frac{12\theta_3^2(1 - \beta x_i)^\beta \exp[\theta_3^2(1 - \beta x_i)]}{x_i(2\pi)^{1/2}(12\theta_2 + 1)} + \epsilon_i, \]

where \( \beta = \theta_3 + (1 - \theta_3)\theta_4 \).

Data set 5 (Identifier 6AAA); \( p = 3, n = 13 \).

Function:

\[ y_i = \theta_1 x_{i,1} + \theta_2 \exp(\theta_3 x_{i,2}) + \epsilon_i. \]

Data set 6 (Identifier 8ACA); \( p = 4, n = 24 \).
Reference: None.

Function:

\[ y_i = \sum_{k=2}^{4} \sum_{j=1}^{k-1} \left( \frac{\theta_k \exp(-\theta_k x_i) - \theta_k \exp(-\theta_k x_j)}{\theta_k - \theta_k} \right) + \epsilon_i. \]

Data set 7 (Identifier 9AAG); \( p = 8, n = 25 \).

Function:

\[ y_i = \theta_1 - \frac{(\theta_3 - \theta_3 x_i + (2\theta_2 - \theta_2)\theta_3 x_i^2)}{(1 + (2\theta_2 - \theta_2)\theta_3 x_i^2)(1 + (2\theta_2 - \theta_2)\theta_3 x_i^2)} \]
\[ + \frac{2\theta_2 \theta_3 x_i (2\theta_2 - \theta_2)\theta_3 x_i^2}{(1 + (2\theta_2 - \theta_2)\theta_3 x_i^2)(1 + (2\theta_2 - \theta_2)\theta_3 x_i^2)} + \epsilon_i. \]

Data set 8 (Identifier 11AAB); \( p = 4, n = 9 \).

Function:

\[ y_i = \frac{\theta_2 \exp(\theta_3) + \theta_2 x_i^\theta_3}{\theta_3 + x_i^\theta_3} + \epsilon_i. \]

Data set 9 (Identifier 12AAB); \( p = 4, n = 9 \).

Function:

\[ y_i = \frac{\theta_2 \exp(\theta_3) + \theta_2 x_i^\theta_3}{\exp(\theta_3) + x_i^\theta_3} + \epsilon_i. \]

Data set 10 (Identifier 14ACG); \( p = 3, n = 10 \).

Function:

\[ y_i = \frac{-\exp(\theta_2 - \theta_3 x_i)}{\theta_2 + \theta_3 x_i} + \epsilon_i. \]

Data set 11 (Identifier 14ABG), \( p = 3, n = 21 \).

Function:

\[ y_i = \frac{-\ln(\theta_1 + \theta_3 x_i)}{\theta_1} + \epsilon_i. \]

Data set 12 (Identifier 14AAG); \( p = 3, n = 42 \).

Function:

\[ y_i = \frac{-\ln(\theta_1 + \theta_3 x_i)}{\theta_3} + \epsilon_i. \]

Data set 13 (Identifier 15AAA); \( p = 3, n = 16 \).
Reference: Ratkowsky (1983), model 6.11, pages 120 and 58.

Function:

\[ y_i = \frac{-\theta_1 + \theta_2 x_i}{\theta_1 + \theta_2 x_i} + \epsilon_i. \]

Data set 14 (Identifier 16AAF); \( p = 5, n = 27 \).

Function:

\[ y_i = \theta_1 + \theta_2 x_i + \theta_3 (x_i - \theta_4)^2 + \theta_3^{1/2} + \epsilon_i. \]

Data set 15 (Identifier 17AAA); \( p = 2, n = 42 \).

Function:

\[ y_i = -\ln(\theta_1 + \theta_2 x_i) + \epsilon_i. \]

Data set 16 (Identifier 18AAA); \( p = 3, n = 9 \).

Function:

\[ y_i = \theta_1 \exp[-\exp(\theta_2 - \theta_3 x_i)] + \epsilon_i. \]

Data set 17 (Identifier 19AAA); \( p = 3, n = 9 \).

Function:

\[ y_i = \frac{-\theta_1}{1 + \exp(\theta_2 - \theta_3 x_i)} + \epsilon_i. \]

Data set 18 (Identifier 20AAG); \( p = 4, n = 9 \).

Function:

\[ y_i = \theta_1 [1 + \exp(\theta_2 - \theta_3 x_i)]^{-\theta_3} + \epsilon_i. \]

Data set 19 (Identifier 21AAA); \( p = 4, n = 9 \).
Function:

\[ y_i = \theta_1 - \theta_2 \exp(-\theta_3 x_i^2) + \epsilon_i. \]

Data set 20 (Identifier 22AAB); \( p = 3, n = 5 \).


Function: \( y_i = \theta_1 - \theta_2 x_i^2 + \epsilon_i. \)

[Received June 1985. Revised July 1986.]

REFERENCES


