THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

CONFIDENCE REGIONS FOR THE PARAMETERS OF A NONLINEAR REGRESSION MODEL

by

A. Ronald Gallant

Institute of Statistics Mimeograph Series No. 1077

July 1976

CONFIDENCE REGIONS FOR THE PARAMETERS OF A NONLINEAR REGRESSION MODEL

by

A. RONALD GALLANT*

*A. Ronald Gallant is Associate Professor of Statistics and Economics, Department of Statistics, North Carolina State University, Raleigh, North Carolina. Much of this work was done while on leave at the Graduate School of Business, University of Chicago. I am indebted to Professors R. R. Bahadur, Albert Madansky, B. Peter Pashigian, David L. Wallace, and Arnold Zellner for helpful discussions of the topic. All errors are my own.

ABSTRACT

Three methods in common use for finding confidence regions for the parameters of a nonlinear regression model are studied. These are the lack of fit method, the linearization method, and the maximum likelihood method. The objective is to determine which of the three methods is the preferred choice in applications. The conclusion is that the linearization method has the best structural characteristics but its nominal confidence level is apt to be inaccurate. A choice of the likelihood ratio method is a compromise between structural characteristics and accuracy. The lack of fit method would only be chosen if accuracy was of extreme concern.

1. INTRODUCTION

Three methods in common use for finding confidence regions for the parameters of a nonlinear regression model

$$y_t = f(x_t, \theta^*) + e_t \qquad (t = 1, 2, \ldots, n)$$

are studied. They are the lack of fit method, the linearization method, and the maximum likelihood method. The objective of the study is to determine which of the three methods is the preferred choice in applications. For this purpose, it is assumed that the functional form of the response function $f(x, \theta)$ is known and that the errors e_t $(t = 1, 2, \ldots, n)$ are independently and normally distributed with mean zero and unknown variance σ^2. The unknown parameter θ^* is a p-dimensional vector known to be contained in the convex, compact parameter space Θ. The asterisk is used to indicate that the true, but unknown, value of the parameter is meant; its omission means that θ is to be regarded as a variable for the purpose of; *e.g.*, differentiation. The inputs x_t $(t = 1, 2, \ldots, n)$ are k-dimensional vectors whose values are known.

There is a considerable difference in the burden of the notation between the more general case of finding confidence regions for a subset of the parameters of the model and the restricted case of a joint confidence region for all the parameters; see, *e.g.*, [5, Secs. 4 and 5]. Attention will be confined to the latter case for the most part to avoid obscurity due to a complicated notation. Little of substance is affected by this restriction, excepting those cases where the lack of fit method cannot be applied.

Let y denote the observed values of y_t $(t = 1, 2, \ldots, n)$ arranged as an n-dimensional column vector;

$$y = (y_1, y_2, \ldots, y_n)' \quad (n \times 1) .$$

Consider a family of (Borel-measurable) test functions $\varphi(y,\theta)$ defined by

$$\varphi(y,\theta) = \begin{cases} 1 & \text{reject } H: \theta^* = \theta \\ 0 & \text{accept } H: \theta^* = \theta . \end{cases}$$

As is well-known, a confidence procedure is in a one-to-one correspondence with such a family of tests. This correspondence is: for observed y put in the confidence region R_y those θ in Θ for which $H: \theta^* = \theta$ is accepted; viz.,

$$R_y = \{\theta \in \Theta: \varphi(y,\theta) = 0\} .$$

The probability that R_y includes θ is given by

$$P(\theta \in R_y) = P[\varphi(y,\theta) = 0] .$$

The nonlinear regression models considered here are presumed to satisfy the regularily conditions of [4, 6]. Now, were θ^* to , conceptually, range over Θ some values of θ^* will not permit the satisfaction of these regularily conditions. For example, those θ^* on the boundary of Θ will not; typically, those θ^* which do not will be a set of Lebesque measure zero. For those conceptual choices of θ^* which do permit the satisfaction of the regularity conditions, the three families of tests considered will have the property that

$$\lim_{n\to\infty} P[\varphi(y,\theta^*) = 0|\theta^*,\sigma^2] = 1-\alpha$$

where α is a preassigned number between zero and one. The notation $P(\cdot|\theta^*,\sigma^2)$ indicates probability computed according to the n-dimensional multivariate normal distribution with mean vector

$$f(\theta) = [f(x_1,\theta), f(x_2,\theta), \ldots, f(x_n,\theta)]' \qquad (n \times 1)$$

and variance-covariance matrix $\sigma^2 I$.

The procedure for finding confidence intervals associated with such a family of tests will achieve the nominal $(1-\alpha) \times 100\%$ confidence level in the limit if the true value permits satisfaction of the regularily conditions.

An approximate $(1-\alpha) \times 100\%$ confidence procedure is defined to be a procedure with this property for all θ^* in Θ excepting a subset of Θ with Lebesque measure zero. On the other hand, an exact $(1-\alpha) \times 100\%$ confidence procedure is a procedure for which the equality

$$P[\varphi(y,\theta^*) = 0|\theta^*,\sigma^2] = 1-\alpha$$

holds for all θ^* in Θ , all $\sigma^2 > 0$, and all n large enough for the test to be defined, typically for n larger than p . Lack of fit methods, defined later, usually have this property. Obviously, an exact confidence procedure is included within the definition of an approximate procedure.

Approximate confidence procedures, adjusting the definition suitably for context, are in common use in a wide variety of situations. Nonetheless, some have argued that the use of such procedures is invalid; see the discussion and references in [1, p. 305]. The thrust argument is that, since θ^* and σ^2 are unknown, a confidence statement should be simultaneously correct for all θ^* in Θ and all $\sigma^2 > 0$; the asserted probability should bound $P[\varphi(y,\theta^*) = 0|\theta^*,\sigma^2]$ from below. If an approximate procedure is to meet this requirement it must be a uniform approximate procedure. That is, it must be true for the family of tests defining it that

$$\lim_{n\to\infty} \inf_{\theta^*,\sigma^2} P[\varphi(y,\theta^*) = 0|\theta^*,\sigma^2] = 1-\alpha \ .$$

The objection to approximate confidence procedures can, perhaps, be made more forcefully by noting that if the equation above is not true then there does not exist any sample size such that

$$P(\theta^* \in R_y|\theta^*,\sigma^2) > 1-\alpha-\varepsilon$$

simultaneously for all $\theta^* \in \Theta$ and $\sigma^2 > 0$.

The response to this objection is the argument that θ^* and σ^2 are fixed in any given application. It is not necessary to be able to make confidence statements which are simultaneously correct for all values of the parameters whatever they might be; it is only necessary that they be correct for the parameters which obtain. When an approximate confidence procedure is used in an application there will, in fact, be a sample size such that

$$P(\theta^* \in R_y|\theta^*,\sigma^2) > 1-\alpha-\varepsilon \ .$$

This is all that is required for validity.

One may not accept the counter argument. In the present context this would require that lack of fit confidence procedures be employed by default. A uniform asymptotic theory is not available for the linearization and maximum likelihood families of tests. Moreover, it does not appear likely that a uniform asymptotic theory can be developed which has sufficient generality to be useful in applications. The difficulty is that many nonlinear regression models obtained from substantive considerations have singularities over regions of the parameter space Θ natural to the problem; see, e.g., [3, 8, 9].

Expected length, area, or volume, according as to dimension of θ, is commonly used to compare confidence procedures and is a tractable criterion here. The criterion will be termed expected volume regardless of dimension and is defined by

$$\text{Expected volume} = \int_{\mathbb{R}^n} \int_{R_y} dm(\theta)\, dP(y|\theta^*, \sigma^2)$$

$$= \int_{\mathbb{R}^n} \int_{\Theta} [1-\varphi(y,\theta)]\, dm(\theta)\, dP(y|\theta^*, \sigma^2)$$

where $m(\theta)$ denotes Lebesque measure on the p-dimensional real numbers. Recall that Θ is compact so existence of the integral is assured. Given two approximate confidence procedures with the same nominal $(1-\alpha) \times 100\%$ confidence level, the one with the smaller expected volume is judged the better of the two. As Pratt [19] shows, by interchanging the order of integration,

$$\text{Expected volume} = \int_{\Theta} \int_{\mathbb{R}^n} [1-\varphi(y,\theta)]\, dP(y|\theta^*, \sigma^2)\, dm(\theta)$$

$$= \int_{\Theta} P[\varphi(y,\theta) = 0|\theta^*, \sigma^2]\, dm(\theta) .$$

The integrand is the probability of covering θ,

$$c_{\varphi}(\theta) = P[\varphi(y,\theta) = 0|\theta^*,\sigma^2] ,$$

and is analgous to the operating characteristic curve of a test. The essential difference between the coverage probability function $c_{\varphi}(\theta)$ and the operating characteristic function lies in the treatment of the hypothesized value θ and the true value of the parameter θ^*. For the coverage function, θ^* is held fixed and θ varies; the converse is true for the operating characteristic function.

A comparison of two approximate confidence procedures corresponding to the two families of tests $\varphi(y,\theta)$ and $\Psi(y,\theta)$ can be effected by comparing their coverage functions. If it can be shown that $c_{\varphi}(\theta)$ is smaller than $c_{\Psi}(\theta)$ for all θ then the confidence procedure corresponding to the family $\varphi(y,\theta)$ is the better of the two. Failing to show this, the integrals of the coverage functions over Θ must be compared.

Some notation which is used throughout the discussion is set forth here for convenient reference.

Notation: Given the regression model

$$y_t = f(x_t,\theta^*) + e_t \qquad (t = 1, 2, \ldots, n) ,$$

the observations

$$(y_t,x_t) \qquad (t = 1, 2, \ldots, n) ,$$

and the hypothesis

$$H: \theta^* = \theta$$

define:

$y = (y_1, y_2, \ldots, y_n)'$ $(n \times 1)$,

$f(\theta) = [f(x_1,\theta), f(x_2,\theta), \ldots, f(x_n,\theta)]'$ $(n \times 1)$,

$e = (e_1, e_2, \ldots, e_n)'$ $(n \times 1)$,

$SSE(\theta) = [y-f(\theta)]'[y-f(\theta)] = \Sigma_{t=1}^{n} [y_t - f(x_t,\theta)]^2$,

$F(\theta)$ = the $n \times p$ matrix with typical element $(\partial/\partial\theta_j)f(x_t,\theta)$ where t is the row index and j the column index,

$P(\theta) = F(\theta)[F'(\theta)F(\theta)]^{-1}F'(\theta)$ $(n \times n)$,

$Q(\theta) = I - P(\theta)$ $(n \times n)$,

$C(\theta) = [F'(\theta)F(\theta)]^{-1}$ $(p \times p)$,

$\delta(\theta) = f(\theta^*) - f(\theta)$ $(n \times 1)$,

$\hat{\theta}$ = the $p \times 1$ vector minimizing $SSE(\theta)$ over Θ ,

$s^2 = SSE(\hat{\theta})/(n-p)$,

$\hat{C} = C(\hat{\theta})$,

$F = F(\theta^*)$,

$$P = P(\theta^*) ,$$

$$Q = Q(\theta^*) .$$

<u>Distributions</u>. Let $p(i,\lambda)$ denote the Poisson density function with mean λ. Let $g(t;\nu,\lambda)$ denote the non-central chi-squared density function with ν degrees freedom and noncentrality parameter λ as defined in [10, p. 74]. Let $G(t;\nu,\lambda)$ be the corresponding distribution function. Let $F'(t;\nu_1,\nu_2,\lambda)$ denote the noncentral F-distribution function with ν_1 numerator degrees-freedom, ν_2 denominator degrees-freedom and noncentrality parameter λ as defined in [10, p. 77]. The central F-distribution function with corresponding degrees freedom is denoted by $F(t;\nu_1,\nu_2)$. The upper α percentage of the central F-distribution is denoted by $F_\alpha(\nu_1,\nu_2)$; F_α will denote $F_\alpha(p,n-p)$ throughout. The critical point of the likelihood ratio test is denoted by

$$c^* = 1 + pF_\alpha/(n-p) .$$

2. THREE COMMON METHODS FOR FINDING CONFIDENCE REGIONS

The three methods considered are obtained from families of tests for the hypothesis of location

$$H: \theta^* = \theta \quad \text{against} \quad A: \theta^* \neq \theta$$

regarding σ^2 as a nuisance parameter. The confidence regions associated with these families would be identical if the response function $f(x,\theta)$ were linear in the parameters. Thus, the problem of deciding which to employ in an application arises solely from the nonlinearity of the regression model.

If interest centers in some subvector τ of the parameter θ partioned according to

$$\theta = (\rho', \tau')'$$

then the requisite families of tests would be generated from the hypothesis

$$H: \tau^* = \tau \quad \text{against} \quad A: \tau^* \neq \tau$$

regarding ρ and σ^2 as nuisance parameters. Of the three methods, the lack of fit procedure is affected the most substantially in the shift to the general case. The details are given in [5, Secs. 4 and 5].

2.1 Lack of Fit Method

If the hypothesis $H: \theta^* = \theta$ is true then the components of the n-vector

$$\hat{e} = y - f(\theta)$$

constitute a random sample from the normal distribution with mean zero and unknown variance σ^2. In principle, any test of the null hypothesis

$$H': \hat{e}_t \quad (t = 1, 2, \ldots, n) \quad \text{a r.s. from a} \quad n(0, \sigma^2)$$

is a test of $H: \theta^* = \theta$ against $A: \theta^* \neq \theta$. In practice, one is interested in a test which is sensitive to departures from H' due to changes in location of the form $\mathcal{E}(\hat{e}) = \delta(\theta)$ rather than a test which is extremely sensitive to, e.g., non-normality. A test which is sensitive to what are, in the present context, irrelevant departures from the null hypothesis H' can have the undesirable property of generating empty confidence regions with positive probability, even were the assumptions true. It would seem desirable to avoid confidence procedures with this property if possible.

These are the essential ideas in the articles by Williams [23] and Halperin [12]. Turner et al. [22] and Hartley [15] exploit the same idea but do not necessarily guarantee by the method of construction that their confidence regions cannot be empty with positive probability. These citations are not exhaustive; applications of similar ideas can be traced back further through the references in these articles.

The family of tests recommended by Halperin for finding an exact $(1-\alpha) \times 100\%$ confidence region for θ^* is based on the F-statistic

$$R(\theta) = \{[y-f(\theta)]'P(\theta)[y-f(\theta)]/p\}/\{[y-f(\theta)]'Q(\theta)[y-f(\theta)]/(n-p)\} .$$

It may be computed using a linear regression program by regressing $y-f(\theta)$ on $F(\theta)$ with zero intercept; the numerator is the regression mean square with p degrees freedom and the denominator is the error mean square with $n - p$ degrees freedom. The test rejects when $R(\theta) > F_\alpha(p,n-p)$ so that the family of tests defining the confidence procedure is

$$\varphi_R(y,\theta) = \begin{cases} 1 & R(\theta) > F_\alpha \\ 0 & R(\theta) \le F_\alpha . \end{cases}$$

The probability that the procedure covers θ is

$$c_R(\theta) = 1 - \int_0^\infty G[t/(c^*-1);\ n-p, \lambda_{R2}]g(t;p,\lambda_{R1})dt$$

where

$$\lambda_{R1} = \delta'(\theta)P(\theta)\delta(\theta)/(2\sigma^2) ,$$

$$\lambda_{R2} = \delta'(\theta)Q(\theta)\delta(\theta)/(2\sigma^2) .$$

This expression is obtained by recognizing that $R(\theta)$ has the doubly noncentral F-distribution with noncentrality parameters λ_{R1} and λ_{R2} .

Assuming that the least squares estimate $\hat{\theta}$ is not a boundary point of the parameter space Θ 1/ , the test statistic $R(\theta)$ takes on the value zero at $\hat{\theta}$ - note that $\nabla\, SSE(\theta) = -2F'(\theta)[y-f(\theta)]$. Consequently, the lack of fit confidence region will contain $\hat{\theta}$ and be nonempty.

The statistic $R(\theta)$ also takes on the value zero at local minima and maxima of the sum of squares function. Thus, in an application it is possible that a lack of fit confidence region consists of a union of disjoint subsets of Θ . Williams [23] illustrates this situation with an example. This feature may or may not be disconcerting to statisticians but it is, in my experience, disconcerting to clients.

In many instances, an attribute of the response function $f(x,\theta)$ is that there is a region A of the p-dimensional real numbers with infinite volume $(m(A) = \infty)$ such that $\delta'(\theta)\delta(\theta) \le a$ for all $\theta \in A$. Consequently, $c_R(\theta)$ will be bounded away from zero on A and

$\int_A c_R(\theta)dm(\theta) = \infty$ 2/. In these cases, expected volume is more a characteristic of the volume of the parameter space Θ than an intrinsic property of the confidence procedure.

If the restrictions embodied in the convex compact set Θ are natural to the situation, this is not a problem. It is then useful to know that one confidence procedure has smaller expected volume than another. On the other hand, if the bound on $\|\theta\|$ implied in the assumption that Θ is compact is merely an artifact of the regularity conditions rather than intrinsic to the problem, expected volume may lose meaning as a basis for comparison.

2.2 Linearization Method

The least squares estimator is asymptotically normally distributed with mean vector θ^* and a variance-covariance matrix estimated consistently by $s^2\hat{C}$ 3/. This suggests the use of the statistic

$$S(\theta) = (\hat{\theta}-\theta)'\hat{C}^{-1}(\hat{\theta}-\theta)/(ps^2)$$

for testing $H: \theta^* = \theta$ against $A: \theta^* \neq \theta$. The test rejects when $S(\theta) > F_\alpha(p,n-p)$ so that the family of tests defining the associated confidence procedure is

$$\varphi_S(y,\theta) = \begin{cases} 1 & S(\theta) > F_\alpha \\ 0 & S(\theta) \leq F_\alpha . \end{cases}$$

The term linearization test stems from the following observation, which is useful in computations. The test may be computed as the standard F-test for the hypothesis $H: \beta = \theta$ against $A: \beta \neq \theta$ in the linear model $z = X\beta + e$ obtained by expanding $f(x,\theta)$ in a first order Taylor's series about the least squares estimate $\hat{\theta}$, $z = y - f(\hat{\theta}) + F(\hat{\theta})\hat{\theta}$, and $X = F(\hat{\theta})$. The least squares estimate, itself, and the statistic s^2 may be computed using either Hartley's [14] or Marquardt's [18] algorithm.

The probability that the procedure covers θ may be approximated by [6]

$$c_S(\theta) = 1 - \int_0^\infty G[t/(c^*-1);n-p,0]g(t;p,\lambda_S)dt$$

where

$$\lambda_S = (\theta^*-\theta)'C^{-1}(\theta^*-\theta)/(2\sigma^2) .$$

The converage function may be computed using charts of the non-central F-distribution such as in [20, p. 438-455].

Confidence regions obtained from the linearization family of tests are nonempty and contain the least squares estimate. Neglecting cases where $\{\theta: \varphi_S(y,\theta) = 0\}$ intersects the boundary of Θ, a linearization confidence region is an ellipsoid with center at $\hat{\theta}$ and eigen vectors of $\hat{C}$ as axes. In short, a linearization confidence region has the familiar features of confidence regions customarily employed in linear regression analysis.

The expected volume of the linearization confidence procedure, computed from the approximation, may be bounded independently of the volume of the parameter space Θ as follows:

$$\int_{\Theta} c_S(\theta)dm(\theta) \le \int c_S(\theta)dm(\theta)$$

$$= 2\{\Gamma(n/2)/[\Gamma(p/2)\Gamma((n-p)/2)]\}$$

$$\cdot [\det(C)]^{1/2}[2\pi\sigma^2 F_\alpha/(n-p)]^{p/2} .$$

Thus, the linearization procedure can not have an expected volume depending primarily on the volume of Θ as is possible with the lack of fit procedure, to the extent the approximation of the coverage probability by $c_S(\theta)$ is valid.

2.3 Likelihood Ratio Method

The likelihood ratio test for the hypothesis $H: \theta^* = \theta$ against $A: \theta^* \neq \theta$ employs the statistic

$$T(\theta) = SSE(\theta)/SSE(\hat{\theta})$$

and rejects when $T(\theta) > c^*$. The family of tests defining the associated confidence region is

$$\varphi_T(y,\theta) = \begin{cases} 1 & T(\theta) > c^* \\ 0 & T(\theta) \le c^* . \end{cases}$$

The denominator $SSE(\hat{\theta})$ of $T(\theta)$ may be computed using Hartley's [14] or Marquardt's [18] algorithm, as mentioned previously. The numerator $SSE(\theta)$ may be computed by using θ as a starting value with either of these algorithms and limiting the number of iterations to one.

The probability that the likelihood ratio procedure covers θ may be approximated by [4]

$$c_T(\theta) = 1 - \int_0^\infty G[t/(c^*-1)+2c^*\lambda_{T2}/(c^*-1)^2;n-p,\lambda_{T2}/(c^*-1)^2]$$

$$\times\ g(t;p,\lambda_{T1})dt$$

where

$$\lambda_{T1} = \delta'(\theta)P\delta(\theta)/(2\sigma^2)$$

$$\lambda_{T2} = \delta'(\theta)Q\delta(\theta)/(2\sigma^2)\ .$$

This function is partially tabulated in [4].

Confidence regions obtained using the likelihood ratio family of tests are nonempty and contain the least squares estimate. The sum of squares surface $SSE(\theta)$ may have local minima. When these minima are below $c^* SSE(\hat{\theta})$ it is possible that the likelihood ratio confidence region will consist of a union of disjoint regions. Note, however, that the likelihood ratio confidence region _may_ include these local minima; the lack of fit confidence region _must_ include them as well as include local maxima and saddle points. One would expect that confidence regions consisting of disjoint unions will occur with less frequency in applications using the likelihood procedure than using the lack of fit procedure.

A feature which the likelihood ratio procedure shares with the lack of fit procedure is that, for the same reasons, the integral $\int_A c_T(\theta)dm(\theta)$ may be infinite for some response functions. As noted before, expected volume then becomes more a characteristic of the parameter space Θ than an intrinsic property of the confidence procedure, to the extent the approximation of the coverage probability by $c_T(\theta)$ is valid[4/].

3. EXAMPLES

The difficulty in attempting a general comparison of the three confidence procedure is that several factors influence the outcome: the form of the response function $f(x,\theta)$, the configuration of the design points $\{x_t: t = 1, 2, \ldots, n\}$, the sample size n, the magnitude of the error variance σ^2, and the actual location of the true parameter θ^*. At one extreme, these factors may interact such that the regression model is so nearly like a model which is linear in the parameters that there are no essential differences between the three methods: confidence contours will be ellipsoidal and nearly coincident, coverage probabilities computed according to the asymptotic approximations will be accurate and nearly equal. At the other extreme are regression models which are so nonlinear that confidence contours can be expected to differ markedly among the three methods and asymptotic approximations will be so inaccurate as to be useless.

Beale [2] has derived measures of nonlinearity which are designed to reflect the influence of these factors. Examples have been examined by Guttman and Meeter [11] to assess the effectiveness of Beale's measures as a measure of the coincidence of linearization and likelihood ratio contours. They find that the measures achieve a fair degree of success[5/]. More so, if attention is restricted to the measures computed from second order partial derivatives in θ of the response function $f(x,\theta)$. Interestingly, Beale obtained the measure N_φ ($\tilde{N}_\varphi$

in Guttman and Meeter) in Section 2 of his article by studying the coincidence of the lack of fit and likelihood ratio contours in an effort to approximate $P[\varphi_T(y,\theta^*) = 0|\theta^*,\sigma^2]$.

The concern here is with the sampling characteristics of the three confidence procedures, so an interest in Beale's measures centers more in their ability to predict the accuracy of the asymptotic approximations of the previous section rather than as a measure of coincidence of contours. This use is more in line with Beale's intention. However, coincidence of linearization and likelihood ratio contours and accuracy of asymptotic approximations in finite samples are not orthogonal dimensions of statistical behavior. In the Addendum to their article, Sprott and Kalbfleisch [21] discuss the relationship between coincidence and accuracy and present an example where a transformation designed to improve coincidence also improves accuracy. Thus, it is expected that Guttman and Meeter's examples, which differ markedly in coincidence of linearization and likelihood ratio contours will display differences both in contours of the coverage functions and in accuracy of approximation.

The two models chosen for reexamination are defined in Table 1; the designations $\mathfrak{M}_2$ and $\mathfrak{M}_3$ are as in Guttman and Meeter. These models differ with respect to design points and response function; variance and sample size are the same in both cases. Model $\mathfrak{M}_3$ is the more nonlinear; Beale's nonlinearity measures evaluated at θ^* using analytic derivatives are: $N_\theta = .00171$ and $N_\varphi = .00053$ for $\mathfrak{M}_2$ and $N_\theta = .87954$ and $N_\varphi = .00333$ for $\mathfrak{M}_3$.

The natural parameter space for these models is $\theta_1,\theta_2 > 0$. To achieve compactness, θ_1 and θ_2 must be bounded from above and all boundary points included. In view of the symmetry of Model η_3 it is reasonable to impose the inequality $\theta_1 \geq \theta_2$ as well[6/].

The coverage functions for these models are evaluated over the region ± 5 standard deviations of the least squares estimator from θ^* and shown as Tables 2 and 3. Contours, obtained by a quadratic interpolation from these tables, are plotted as Figures A and B. An impression of the relative size and location of the tabled regions with respect to the original scaling of the models may be had by reference to Figure C. The shapes of the confidence regions themselves, corresponding to a realization of the errors, are plotted as Figure 2.6 of Guttman and Meeter [11]; they look much the same as the 75% coverage contours of Figures A and B after allowing for the differences in scaling.

There are three interesting aspects of these computations. The first comes as no surprise, for the nearly linear Model η_2 the contours of the coverage functions are nearly coincident while for the relatively more nonlinear Model η_3 the departures are fairly substantial. The second is the near coincidence of the coverage functions for the lack of fit and likelihood ratio procedures for both models; they are not distinguishable within the tolerances used in the construction of Figures A and B. The third is the inequality

$$c_T(\theta) \leq c_R(\theta)$$

which holds over the tabled region for both models. This indicates that the likelihood ratio procedure is preferable to the lack of fit procedure as it covers false values of θ with equal or smaller probability in all instances.

These examples describe relative yields of two substances in a chemical process at time x and, as such, satisfy the inequality

$$0 < f(x,\theta) < 1 .$$

It follows that

$$\delta'(\theta)\delta(\theta) < 12$$

over the set $A = \{\theta: 0 \le \theta_1, 0 \le \theta_2, \theta_1 \ge \theta_2\}$; hence, the integrals $\int_A c_R(\theta)dm(\theta)$ and $\int_A c_T(\theta)dm(\theta)$ are both infinite[7,8]. As mentioned previously, the comparison of the expected volume of the linearization procedure, $\int_\Theta c_S(\theta)dm(\theta)$, with the expected volume of the lack of fit or maximum likelihood procedure is of dubious value when this situation obtains; the linearization procedure may, in principle, be shown to be the best by taking the volume of Θ suitable large.

Despite this qualification, the expected volume $\int_\Theta c(\theta)dm(\theta)$ over the region

$$\Theta = \{\theta: |\theta_1 - 1.4| \le 5\sigma_1, |\theta_2 - .4| \le 5\sigma_2\}$$

has intuitive appeal as a summary measure of performance over a relevant subset of the parameter space. This measure is the unweighted average of the probability of covering false values of θ over the rectangle Θ centered at θ^* . Its interpretation is analagous to that of average power of a test with respect to a uniform, informative prior, and it may be used as a numeric aid to the interpretation of Figures A and B. Expected volumes for the two examples and three procedures are given in Table 4. They were computed by numerical integration from the entries of Tables 2, 3[9] with respect to the original scaling, not standard deviation scaling.

As seen from the table, there is no difference of any importance in expected volume among the three procedures for Model η_2, confirming the visual impression conveyed by Figure A. The situation changes with Model η_3; the ordering from best to worst is linearization, likelihood ratio, and lack of fit with little difference between the likelihood ratio and lack of fit procedures. This ordering is difficult to discern visually from Figure B without the aid of the expected volume computation.

The last question of interest is the effect of the variation in degree of nonlinearity between Model η_2 and Model η_3 on the accuracy of the asymptotic approximations used to study the statistical behavior of the confidence procedures. Table 5 relates to this question; the right hand sides of

$$c_S(\theta) \doteq P[\varphi_S(y,\theta) = 0 | \theta^*, \sigma^2]$$

$$c_T(\theta) \doteq P[\varphi_T(y,\theta) = 0 | \theta^*, \sigma^2] ,$$

computed by Monte-Carlo integration using the control variate method of variance reduction [13, p. 59-60] is compared with $c_S(\theta)$ and $c_T(\theta)$ in the table.

The parameter values were chosen for inclusion in Table 5 by random selection from those parameter combinations in Tables 2 and 3 having non-blank entries, subject to the constraint that there be at least one point satsifying: $0 < c_S(\theta), c_T(\theta) \leq .25$; $.25 < c_S(\theta), c_T(\theta) \leq .5$; $.5 < c_S(\theta), c_T(\theta) \leq .75$; $.75 < c_S(\theta), c_T(\theta) < .95$; and $c_S(\theta), c_T(\theta) = .95$.

The coverage function $c_T(\theta)$ of the likelihood ratio procedure is a remarkably accurate approximation of the actual coverage probability; its accuracy is negligably affected by the difference in nonlinearity between the Models η_2 and η_3. For any practical purpose, the nominal confidence level of 95% is achieved in both cases; overall, the approximation

$$c_T(\theta) \doteq P[\varphi_T(y,\theta) = 0|\theta^*,\sigma^2]$$

is entirely satisfactory.

The converse is true of the linearization procedure. For the nearly linear Model η_2, the approximation

$$c_S(\theta) \doteq P[\varphi_S(y,\theta) = 0|\theta^*,\sigma^2]$$

is only marginally acceptable overall. The agreement of the nominal 95% confidence level with the confidence level actually achieved is close enough for practical purposes, but the hypothesis

$$H: \ P[\varphi_S(y,\theta^*) = 0|\theta^*,\sigma^2] = .95$$

may be rejected at the 5% level[10/]. For the more nonlinear Model η_3, the approximation is completely unsatisfactory[11/].

These findings are in agreement with the results reported in [6]. There also, the accuracy of the likelihood ratio approximation was found to be better than that of the linearization approximation.

4. COMPARISON

The argument that approximate confidence procedures are invalid shall be set aside for the moment. Were it accepted, the lack of fit procedure must be employed and there is no basis for discussion.

The linearization procedure has the most desirable structural characteristics provided there is agreement that a convex confidence region is preferable to a (possibly) disconnected region and that a bounded region is preferable to a (possibly) unbounded region. Moreover, the linearization region is the simplest to construct in applications, especially when the region is an interval.

Arrayed against these desirable structural characteristics is the serious fault revealed by Monte-Carlo simulation: there is apt to be an uncomfortably large discrepancy between the actual and nominal confidence level of the linearization procedure. These same simulations indicate that the likelihood ratio procedure suffers little, if any, from this deficiency and, of course, the lack of fit procedure is completely free of it, being an exact procedure. One might, then, choose either the likelihood ratio or lack of fit procedure in preference to the linearization procedure.

The structural characteristics of the likelihood ratio procedure are certainly more desirable than those of the lack of fit procedure. While both may be unbounded and disconnected, the likelihood ratio confidence region has the merit of having a boundary which coincides with a likelihood contour. In contrast, a lack of fit region contains neighborhoods of points corresponding to every ripple of the likelihood surface - every local minimum, every local maximum, and every saddle point. It is hard to find any merit in this. The only justification for a choice of the lack of fit procedure is that it is exact.

In summary, the linearization method has the best structural characteristics but its nominal confidence level is apt to be inaccurate. A choice of the likelihood ratio method is a compromise between structural characteristics and accuracy. The lack of fit method would only be chosen if accuracy were of extreme concern.

FOOTNOTES

[1/]This is true in applications more often than it might appear at first glance. Boundaries often correspond to degenerate forms of the response function so that there is always an interior point with a smaller sum of squares than any boundary point. Further, $\hat{\theta}$ is interior to Θ with probability tending to one as sample size increases under the regularity conditions.

[2/]A simple example is $f(x,\theta) = \theta_1 e^{-\theta_2 x}$ where all $x_t > 0$ and $\theta_1, \theta_2 \geq 0$. Take as A the set $\{\theta: .5 \leq \theta_1 \leq 1, 0 \leq \theta_2\}$.

[3/]More formally, $\hat{\theta}$ converges almost surely to θ^*, s^2 converges almost surely to σ^2, $[(1/n)F'(\hat{\theta})F(\hat{\theta})]^{-1}$ converges almost surely to a matrix Σ^{-1}, and $\sqrt{n}\,(\hat{\theta} - \theta^*)$ converges in distribution to a p-variate normal with mean vector zero and variance-covariance matrix $\sigma^2\Sigma^{-1}$ [7].

[4/]In some instances, such as the examples of the next section, it is possible to verify directly that $\inf_{\theta \in A} P[\varphi_T(y,\theta) = 0 | \theta^*, \sigma^2] > 0$; reliance on the approximation is then avoided.

[5/]In Table 2.3 of their article, read $\mathfrak{N}_\varphi = .0011$ for $\mathfrak{N}_\varphi = .011$.

6/ Table 3 and Figure B do not reflect the imposition of the bound $\theta_1 \geq \theta_2$. To impose it in Table 3, delete all entries in Table 3 with column headings -5, -4.5, -4, and -3.5 . To impose it in Figure B, draw a line through the points (-4, -3.3) and (-3.1, 5).

7/ Smaller bounds can be established on subsets. For Model η_3 the bound $a = .034$ holds on the set $A = \{(\theta_1,\theta_2): 25 \leq \theta_1, .3 \leq \theta_2 \leq .305\}$.

8/ In this case, it is possible to verify that $\int_A \int_{\mathbb{R}^n} [1 - \varphi_T(y,\theta)]dP(y|\theta^*,\sigma^2)dm(\theta)$ is infinite without relying on the approximation $c_T(\theta)$. It may be shown that there is an open set E such that for all e in E the $\sup_{\theta\epsilon A} \|e + \delta(\theta)\|^2 \leq c^* \inf_{\theta\epsilon A} \|e + \delta(\theta)\|^2$. This implies that $\varphi_T(y,\theta) = 0$ for $y = f(\theta^*) + e$, hence, the inner integral is bounded away from zero on A and the double integral is infinite.

9/ Before rounding; the inequality $\theta_1 \geq \theta_2$ was imposed on Model η_3 .

10/ The z-statistic is $(.9475 - .95)/.001227 = -2.04$.

11/ Two attempts to remedy this deficiency were employed. The first was to use $C(\theta)$ instead of $\hat{C}$ in the formula for $S(\theta)$. The second was to use the correction suggested by Beale [2, Eq. 2.21]. Both attempts were failures. In fairness to Beale it must be appended that his correction is for the likelihood ratio method not the linearization method.

1. Examples of Guttman and Meeter

a. Model η_2

$$f(x, \theta) = \begin{cases} \theta_1 (e^{-x\theta_2} - e^{-x\theta_1})/(\theta_1 - \theta_2) & \theta_1 \neq \theta_2 \\ \theta_1 x e^{-x\theta_1} & \theta_1 = \theta_2 \end{cases}$$

$\theta^* = (1.4, .4)$

$\{x_t\} = \{.25, .5, 1, 1.5, 2, 4, .25, .5, 1, 1.5, 2, 4\}$

$n = 12$

$\sigma^2 = (.025)^2$

b. Model η_3

$$f(x, \theta) = \begin{cases} 1 - (\theta_1 e^{-x\theta_2} - \theta_2 e^{-x\theta_1})/(\theta_1 - \theta_2) & \theta_1 \neq \theta_2 \\ 1 - e^{-x\theta_1} - x\, \theta_1\, e^{-x\theta_1} & \theta_1 = \theta_2 \end{cases}$$

$\theta^* = (1.4, .4)$

$\{x_t\} = \{1, 2, 3, 4, 5, 6, 1, 2, 3, 4, 5, 6\}$

$n = 12$

$\sigma^2 = (.025)^2$

2. Coverage Probabilities for the Model η_2.

$(\theta_2 - .4)/\sigma_2$	$(\theta_1 - 1.4)/\sigma_1$ -5.0	-4.5	-4.0	-3.5	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
A. LACK OF FIT																					
5.0						.003	.007	.013	.021	.030	.039	.045	.047	.045	.039	.032	.024	.017	.011	.007	.004
4.5					.003	.008	.017	.030	.047	.065	.080	.090	.092	.087	.076	.062	.046	.032	.021	.013	.007
4.0				.003	.008	.018	.036	.062	.093	.125	.150	.165	.166	.155	.135	.109	.081	.056	.037	.022	.012
3.5			.002	.006	.017	.037	.071	.117	.168	.217	.253	.272	.271	.251	.217	.175	.131	.091	.059	.035	.020
3.0			.004	.013	.033	.070	.127	.198	.274	.341	.387	.408	.401	.370	.319	.257	.193	.135	.087	.052	.029
2.5		.002	.008	.024	.059	.119	.205	.306	.405	.486	.538	.556	.542	.498	.430	.348	.262	.183	.118	.071	.039
2.0		.004	.015	.042	.096	.185	.302	.430	.545	.633	.683	.696	.674	.618	.536	.435	.329	.230	.149	.089	.049
1.5	.002	.007	.024	.066	.144	.262	.408	.554	.676	.760	.804	.809	.780	.716	.622	.506	.383	.268	.172	.102	.056
1.0	.003	.012	.037	.095	.197	.342	.508	.662	.780	.854	.888	.886	.851	.762	.679	.553	.417	.289	.165	.106	.058
0.5	.004	.017	.051	.125	.249	.413	.589	.741	.849	.911	.935	.927	.887	.812	.702	.567	.423	.290	.182	.105	.056
0.0	.006	.022	.065	.152	.290	.464	.641	.786	.883	.934	.950	.935	.888	.806	.689	.548	.401	.270	.166	.093	.048
-0.5	.008	.027	.076	.170	.313	.488	.659	.794	.882	.926	.935	.912	.855	.762	.637	.494	.352	.230	.137	.075	.037
-1.0	.009	.030	.080	.175	.315	.480	.639	.765	.845	.883	.884	.850	.781	.678	.550	.412	.283	.178	.102	.054	.026
-1.5	.009	.030	.078	.165	.292	.440	.583	.695	.767	.797	.789	.744	.664	.557	.434	.311	.205	.123	.068	.034	.016
-2.0	.009	.027	.068	.142	.248	.373	.492	.587	.646	.666	.648	.594	.513	.412	.306	.210	.131	.075	.040	.019	.009
-2.5	.007	.022	.054	.111	.192	.287	.378	.450	.493	.501	.476	.423	.350	.268	.190	.123	.073	.040	.020	.009	.004
-3.0	.005	.016	.038	.077	.133	.197	.259	.306	.331	.330	.304	.260	.206	.150	.101	.062	.035	.018	.009	.004	.002
-3.5	.004	.010	.024	.048	.081	.119	.154	.180	.191	.186	.166	.136	.102	.071	.045	.026	.014	.007	.003		
-4.0	.002	.006	.013	.025	.042	.061	.079	.090	.093	.088	.075	.059	.042	.028	.017	.009	.005	.002			
-4.5		.003	.006	.012	.019	.027	.034	.037	.038	.034	.028	.021	.014	.009	.005	.003					
-5.0			.002	.004	.007	.010	.012	.013	.012	.011	.008	.006	.004	.002							
B. LINEARIZATION																					
5.0						.002	.003	.006	.010	.015	.020	.024	.025	.025	.022	.018	.013	.008	.005	.002	
4.5					.002	.005	.010	.017	.027	.039	.050	.058	.061	.059	.051	.041	.029	.019	.011	.005	.002
4.0				.003	.006	.013	.026	.043	.065	.089	.110	.124	.127	.120	.104	.081	.058	.037	.021	.011	.005
3.5			.003	.007	.016	.032	.058	.093	.134	.175	.209	.228	.230	.214	.183	.143	.102	.065	.037	.019	.008
3.0		.002	.006	.016	.035	.067	.115	.175	.240	.301	.347	.370	.366	.337	.287	.224	.159	.102	.058	.029	.013
2.5	.002	.005	.013	.032	.067	.123	.199	.288	.378	.455	.508	.529	.518	.474	.403	.316	.224	.143	.081	.041	.018
2.0	.003	.010	.026	.058	.115	.199	.306	.422	.530	.614	.666	.682	.660	.603	.514	.403	.287	.183	.104	.051	.022
1.5	.006	.017	.043	.093	.175	.288	.422	.556	.670	.752	.796	.803	.773	.706	.603	.474	.337	.214	.120	.059	.025
1.0	.010	.027	.065	.134	.240	.378	.530	.670	.781	.852	.886	.884	.848	.773	.660	.518	.366	.230	.127	.061	.025
0.5	.015	.039	.089	.175	.301	.455	.614	.752	.852	.912	.935	.926	.884	.803	.682	.529	.370	.228	.124	.058	.024
0.0	.020	.050	.110	.209	.347	.508	.666	.796	.886	.935	.950	.935	.886	.796	.666	.508	.347	.209	.110	.050	.020
-0.5	.024	.058	.124	.228	.370	.529	.682	.803	.884	.926	.935	.912	.852	.752	.614	.455	.301	.175	.069	.039	.015
-1.0	.025	.061	.127	.230	.366	.518	.660	.773	.848	.884	.886	.852	.781	.670	.530	.378	.240	.134	.065	.027	.010
-1.5	.025	.059	.120	.214	.337	.474	.603	.706	.773	.803	.796	.752	.670	.556	.422	.288	.175	.093	.043	.017	.006
-2.0	.022	.051	.104	.183	.287	.403	.514	.603	.660	.682	.666	.614	.530	.422	.306	.199	.115	.058	.026	.010	.003
-2.5	.018	.041	.081	.143	.224	.316	.403	.474	.518	.529	.508	.455	.378	.288	.199	.123	.067	.032	.013	.005	.002
-3.0	.013	.029	.058	.102	.159	.224	.287	.337	.366	.370	.347	.301	.240	.175	.115	.067	.035	.016	.006	.002	
-3.5	.008	.019	.037	.065	.102	.143	.183	.214	.230	.228	.209	.175	.134	.093	.058	.032	.016	.007	.003		
-4.0	.005	.011	.021	.037	.058	.081	.104	.120	.127	.124	.110	.089	.065	.043	.026	.013	.006	.003			
-4.5	.002	.005	.011	.019	.029	.041	.051	.059	.061	.058	.050	.039	.027	.017	.010	.005	.002				
-5.0		.002	.005	.008	.013	.018	.022	.025	.025	.024	.020	.015	.010	.006	.003	.002					
C. LIKELIHOOD RATIO																					
5.0						.003	.007	.013	.021	.030	.038	.044	.046	.044	.038	.031	.023	.016	.010	.006	.003
4.5					.003	.008	.017	.030	.047	.064	.080	.089	.091	.086	.075	.060	.044	.030	.019	.011	.006
4.0				.003	.008	.018	.036	.062	.093	.124	.150	.164	.165	.154	.133	.106	.079	.054	.034	.020	.011
3.5			.002	.006	.016	.037	.071	.117	.168	.216	.253	.272	.270	.250	.215	.172	.128	.088	.056	.033	.018
3.0			.004	.013	.033	.070	.127	.198	.274	.340	.387	.408	.401	.369	.317	.255	.190	.131	.084	.049	.027
2.5		.002	.008	.024	.058	.119	.205	.306	.405	.486	.537	.556	.542	.497	.429	.345	.259	.179	.115	.068	.037
2.0		.004	.014	.041	.096	.185	.302	.430	.545	.632	.683	.696	.673	.618	.535	.433	.326	.226	.145	.085	.046
1.5	.002	.007	.024	.065	.143	.262	.407	.554	.676	.760	.804	.809	.779	.716	.621	.505	.381	.265	.169	.099	.053
1.0	.003	.011	.036	.093	.196	.342	.508	.662	.780	.854	.888	.886	.851	.781	.679	.551	.415	.267	.181	.105	.056
0.5	.004	.016	.050	.123	.247	.412	.568	.741	.849	.911	.935	.927	.887	.812	.702	.566	.422	.288	.180	.102	.053
0.0	.006	.021	.063	.150	.288	.463	.640	.786	.883	.934	.950	.935	.888	.806	.689	.547	.400	.268	.163	.091	.046
-0.5	.007	.026	.073	.167	.311	.487	.658	.794	.882	.926	.935	.912	.855	.762	.637	.494	.351	.228	.135	.073	.036
-1.0	.008	.028	.078	.172	.312	.479	.639	.764	.845	.883	.884	.850	.781	.678	.549	.411	.282	.177	.101	.053	.025
-1.5	.008	.028	.075	.162	.289	.438	.582	.695	.767	.797	.789	.743	.664	.557	.433	.311	.204	.123	.067	.034	.015
-2.0	.008	.025	.066	.139	.245	.370	.491	.586	.646	.666	.647	.594	.512	.412	.306	.209	.131	.075	.039	.019	.008
-2.5	.006	.020	.052	.108	.189	.284	.377	.449	.492	.500	.476	.423	.350	.268	.189	.123	.073	.040	.020	.009	.004
-3.0	.005	.014	.036	.074	.130	.194	.257	.305	.330	.329	.304	.260	.206	.150	.100	.062	.035	.018	.008	.004	
-3.5	.003	.009	.022	.045	.078	.116	.153	.179	.191	.185	.165	.136	.102	.071	.045	.026	.014	.007	.003		
-4.0	.002	.005	.012	.024	.041	.060	.077	.089	.093	.087	.075	.059	.042	.027	.016	.009	.005	.002			
-4.5		.002	.005	.011	.018	.026	.033	.037	.037	.034	.028	.021	.014	.009	.005	.003					
-5.0			.002	.004	.007	.009	.011	.012	.012	.011	.008	.006	.004	.002							

NOTE: Blanks indicate values less than .0015. To convert from standard deviation scaling use $\sigma_1 = .052957$ and $\sigma_2 = .014005$.

3. Coverage Probabilities for the Model η_3.

	$(\theta_1 - 1.4)/\sigma_1$																				
$(\theta_2 - .4)/\sigma_2$	-5.0	-4.5	-4.0	-3.5	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
A. LACK OF FIT																					
5.0						.255	.316	.005													
4.5						.128	.488	.030													
4.0						.044	.594	.124	.002												
3.5						.010	.585	.330	.015												
3.0							.450	.584	.094	.002											
2.5							.242	.744	.323	.028											
2.0							.078	.737	.648	.168	.014										
1.5							.012	.543	.853	.507	.115	.012									
1.0								.242	.848	.836	.442	.115	.017	.002							
0.5								.050	.607	.934	.821	.459	.155	.036	.006						
0.0								.004	.227	.785	.950	.838	.541	.248	.086	.025	.007	.002			
-0.5									.030	.367	.808	.933	.861	.652	.396	.199	.086	.034	.013	.005	.002
-1.0										.058	.359	.712	.858	.845	.729	.551	.368	.222	.124	.060	.034
-1.5										.002	.046	.230	.488	.663	.721	.692	.608	.494	.376	.272	.188
-2.0												.017	.082	.206	.339	.434	.479	.479	.449	.399	.342
-2.5													.002	.012	.038	.079	.127	.170	.202	.221	.227
-3.0																.002	.005	.011	.019	.028	.037
-3.5																					
-4.0																					
-4.5																					
-5.0																					
B. LINEARIZATION																					
5.0	.022	.024	.012	.002																	
4.5	.024	.055	.057	.027	.005																
4.0	.012	.057	.117	.116	.055	.011															
3.5	.002	.027	.116	.218	.209	.100	.020	.002													
3.0		.005	.055	.209	.357	.330	.160	.033	.002												
2.5			.011	.100	.330	.517	.467	.230	.049	.004											
2.0				.020	.160	.467	.673	.599	.301	.064	.005										
1.5				.002	.033	.230	.599	.800	.706	.361	.077	.005									
1.0					.002	.049	.301	.706	.887	.779	.402	.084	.006								
0.5						.004	.064	.361	.779	.935	.815	.416	.084	.005							
0.0							.005	.077	.402	.815	.950	.815	.402	.077	.005						
-0.5								.005	.084	.416	.815	.935	.779	.361	.064	.004					
-1.0									.006	.084	.402	.779	.887	.706	.301	.049	.002				
-1.5										.005	.077	.361	.706	.800	.599	.230	.033	.002			
-2.0											.005	.064	.301	.599	.673	.467	.160	.020			
-2.5												.004	.049	.230	.467	.517	.330	.100	.011		
-3.0													.002	.033	.160	.330	.357	.209	.055	.005	
-3.5														.002	.020	.100	.209	.218	.116	.027	.002
-4.0																.011	.055	.116	.117	.057	.012
-4.5																	.005	.027	.057	.055	.024
-5.0																		.002	.012	.024	.022
C. LIKELIHOOD RATIO																					
5.0						.245	.314	.005													
4.5						.122	.485	.030													
4.0						.041	.589	.124													
3.5						.009	.579	.329	.015												
3.0							.444	.583	.094	.002											
2.5							.237	.743	.323	.028											
2.0							.075	.736	.648	.168	.014										
1.5							.012	.540	.852	.507	.115	.012									
1.0								.240	.848	.836	.442	.115	.017	.002							
0.5								.049	.607	.934	.621	.459	.155	.036	.006						
0.0								.004	.226	.785	.950	.838	.541	.247	.085	.024	.006	.002			
-0.5									.030	.367	.808	.933	.861	.651	.394	.195	.084	.032	.012	.004	.002
-1.0										.058	.359	.712	.858	.844	.726	.546	.360	.213	.116	.060	.030
-1.5										.002	.046	.230	.488	.662	.719	.689	.600	.482	.360	.253	.170
-2.0												.017	.082	.206	.338	.432	.473	.470	.434	.380	.318
-2.5													.002	.012	.038	.078	.125	.167	.196	.211	.213
-3.0																.002	.005	.011	.018	.026	.034
-3.5																					
-4.0																					
-4.5																					
-5.0																					

NOTE: Blanks indicate values less than .0015. To convert from standard deviation scaling use $\sigma_1 = .27395$ and $\sigma_2 = .029216$.

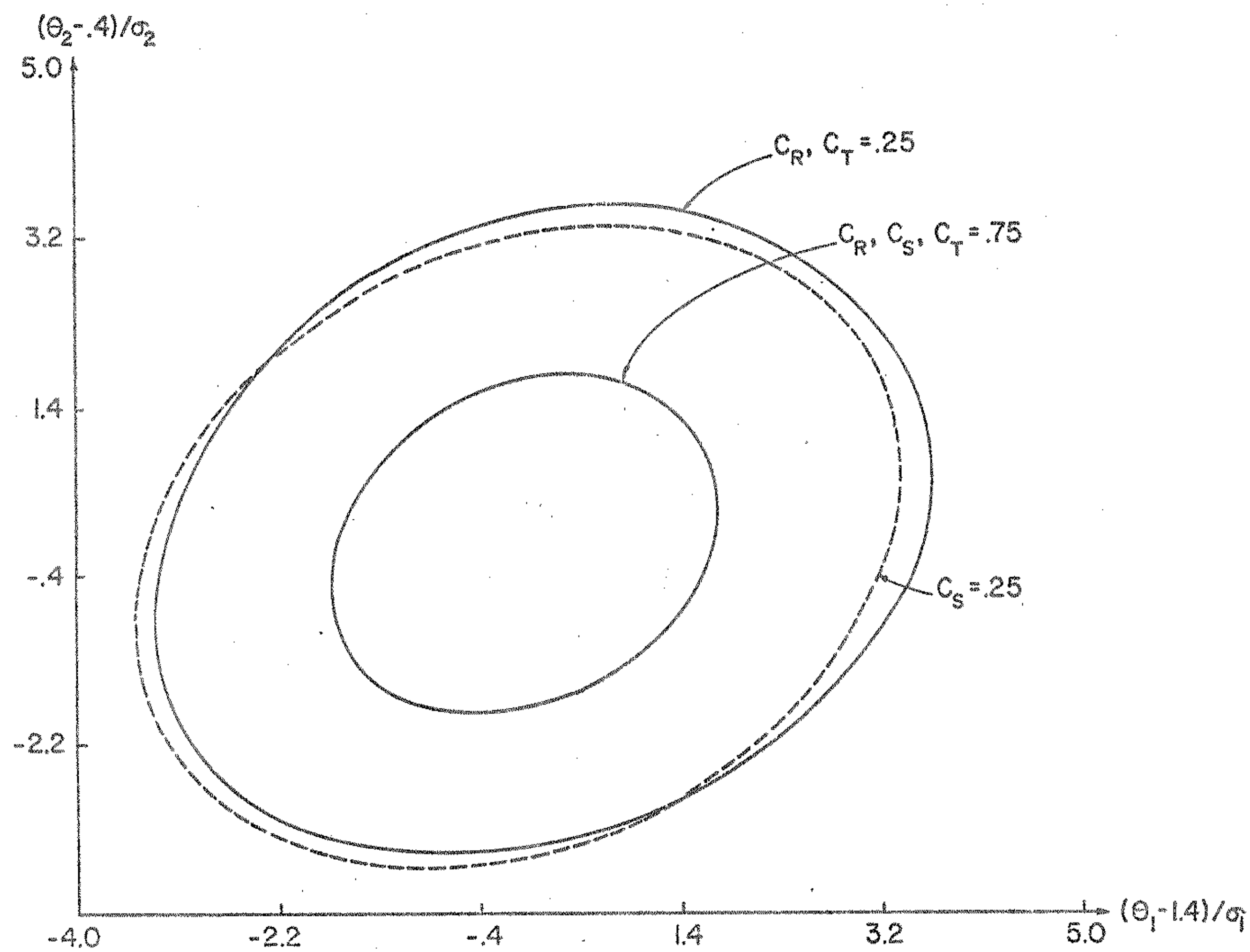
4. Expected Volume of the Lack of Fit, Linearization, and Likelihood Ratio Confidence Procedures

Expected Volume	Model	
	η_2	η_3
Lack of Fit Procedure	.01857	.07187
Linearization Procedure	.01847	.06838
Likelihood Ratio Procedure	.01851	.07132
Parameter Space	.07417	.69235

5. Monte-Carlo Estimates of Coverage Probabilities

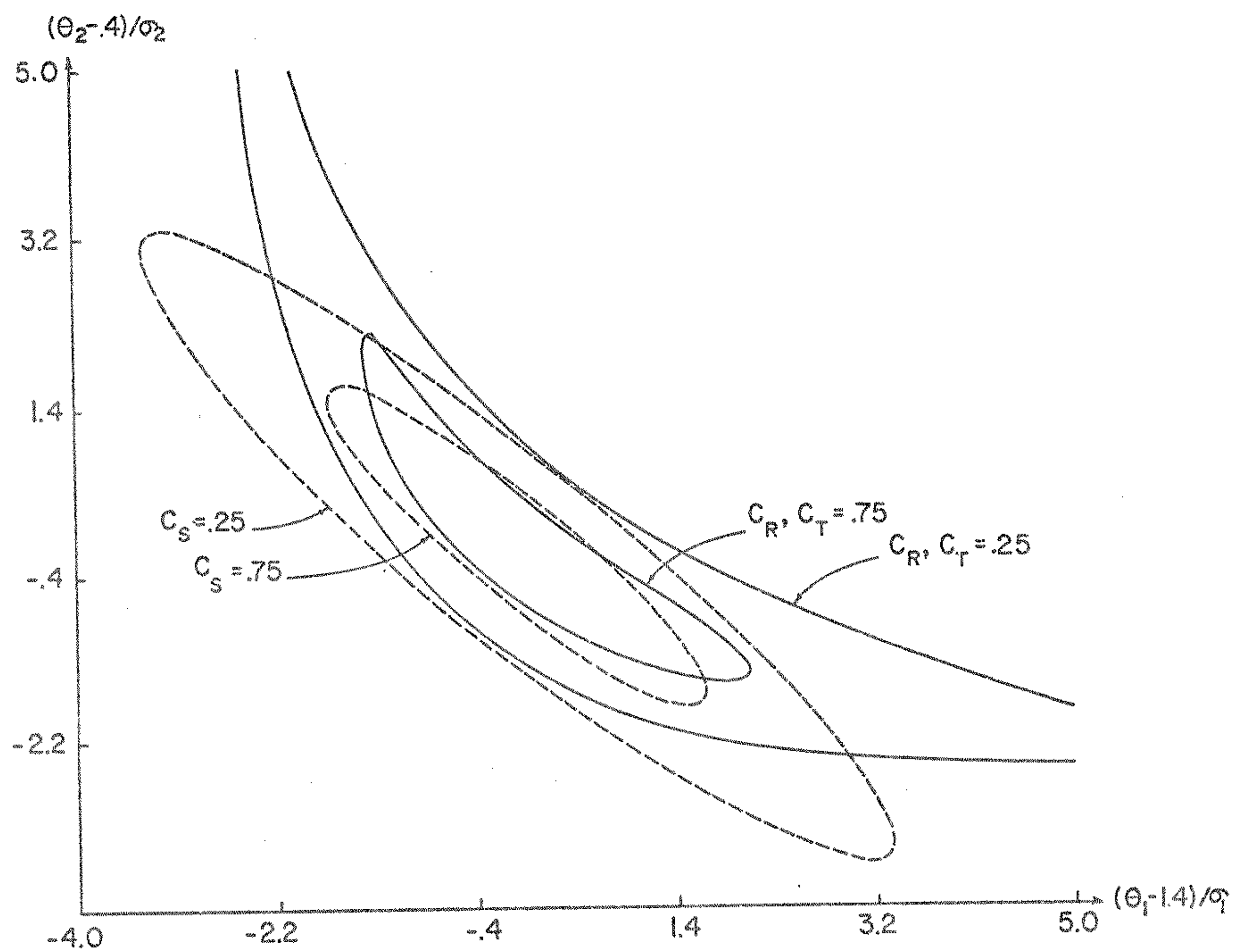
		Linearization			Likelihood Ratio		
			Monte-Carlo Estimate			Monte-Carlo Estimate	
$(\theta_1 - 1.4)/\sigma_1$	$(\theta_2 - .4)/\sigma_2$	$c_S(\theta)$	$P[\phi_S(y, \theta) = 0]$	Std. Err.	$c_T(\theta)$	$P[\phi_T(y, \theta) = 0]$	Std. Err.
			a. Model η_2				
-4.5	1.0	.0275	.0165	.0017	.0111	.0107	.0020
-3.0	0.5	.3009	.2842	.0027	.2472	.2477	.0035
-1.5	-1.5	.7057	.7262	.0023	.6949	.6952	.0017
1.5	-0.5	.7521	.7461	.0018	.7621	.7621	.0016
3.0	-4.0	.0062	.0052	.0008	.0045	.0052	.0006
2.0	3.0	.2873	.2878	.0017	.3171	.3200	.0028
-1.5	1.0	.6705	.6777	.0022	.6619	.6632	.0015
0.5	-0.5	.9115	.9110	.0016	.9115	.9108	.0009
0.0	0.0	.9500	.9475	.0012	.9500	.9493	.0008
			b. Model η_3				
-2.5	0.5	.0036	.0460	.0009	.0000	.0000	.0000
-1.0	0.0	.4016	.5478	.0074	.2262	.2263	.0060
2.0	-1.5	.5987	.5417	.0062	.7193	.7218	.0071
0.5	-1.0	.7790	.7953	.0056	.7123	.7108	.0041
4.5	-3.0	.0055	.1050	.0012	.0264	.0248	.0025
0.0	1.0	.4016	.2873	.0054	.4415	.4436	.0032
-2.0	3.5	.0205	.2355	.0022	.5793	.5808	.0078
-0.5	1.0	.7790	.6290	.0055	.8359	.8440	.0040
0.0	0.0	.9500	.8655	.0034	.9500	.9498	.0012

A. Coverage Probabilities for the Model η_2.



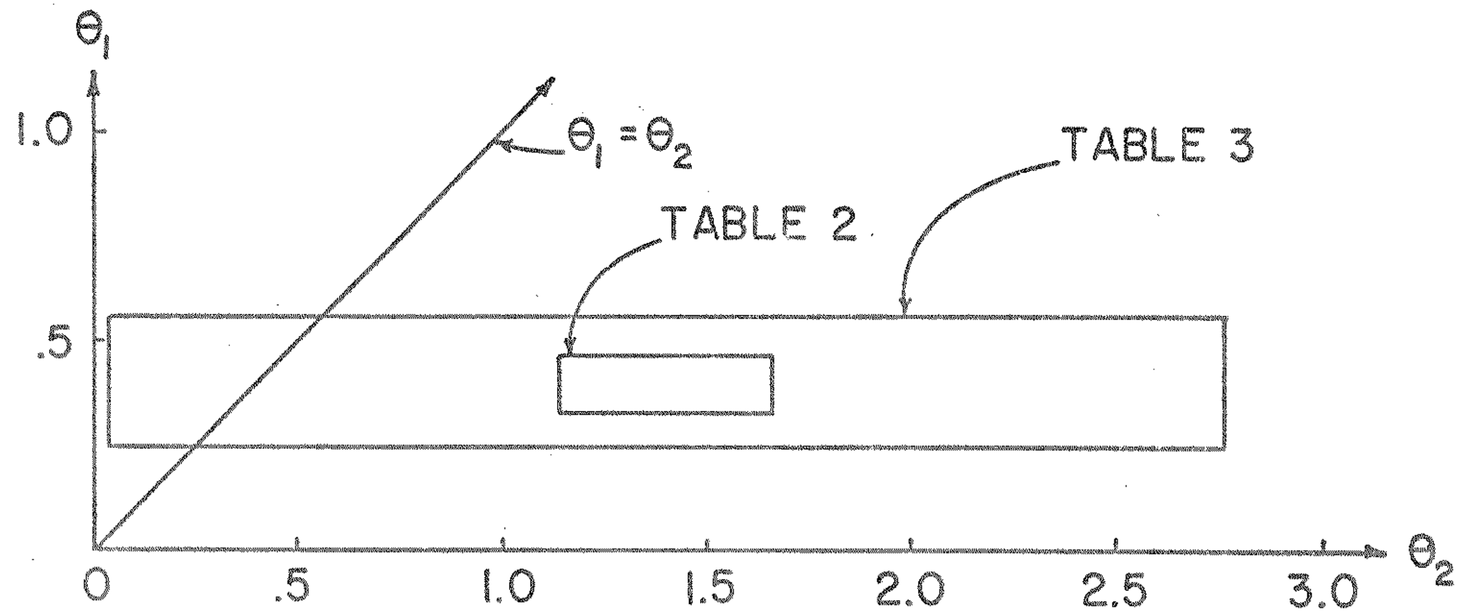
NOTE: To convert from standard deviation scaling use $\sigma_1 = .052957$ and $\sigma_2 = .014005$.

B. Coverage Probabilities for the Model η_3.



NOTE: To convert from standard deviation scaling use σ_1 = .27395 and σ_2 = .029216 .

C. Relative Size and Location of the Tabled Regions.



REFERENCES

[1] Bahadur, R. R., "Rates of Convergence of Estimates and Test Statistics", The Annals of Mathematical Statistics, 38 (April 1967), 303-23.

[2] Beale, E. M. L., "Confidence Regions in Non-Linear Estimation", Journal of the Royal Statistical Society Series B, 22, No. 1 (1960), 41-76.

[3] Feder, Paul I., "The Log Likelihood Ratio in Segmented Regression", The Annals of Statistics, 3 (January 1975), 84-97.

[4] Gallant, A. Ronald, "The Power of the Likelihood Ratio Test of Location in Nonlinear Regression Models", Journal of the American Statistical Association, 70 (March 1975), 198-203.

[5] Gallant, A. Ronald, "Nonlinear Regression", The American Statistician, 29 (May 1975), 73-81.

[6] Gallant, A. Ronald, "Testing a Subset of the Parameters of a Nonlinear Regression Model", Journal of the American Statistical Association, 70 (December 1975), 927-32.

[7] Gallant, A. Ronald, "Inference for Nonlinear Models" Mimeograph Series No. 875 (revised), Institute of Statistics, North Carolina State University, 1975.

[8] Gallant, A. Ronald, "Testing a Nonlinear Regression Specification: A Nonregular Case," Mimeograph Series No. 1017, Institute of Statistics, North Carolina State University, 1975.

[9] Gallant, A. Ronald and Fuller, Wayne A., "Fitting Segmented Polynomial Regression Models Whose Join Points Have to be Estimated", Journal of The American Statistical Association, 68 (March 1973), 144-47.

[10] Graybill, Franklin A., An Introduction to Linear Statistical Models, Vol. 1, New York: McGraw-Hill Book Company, 1961.

[11] Guttman, Irwin and Meeter, Duane A., "On Beale's Measures of Non-linearity", Technometrics, 7, No. 4 (November 1965), 623-37.

[12] Halperin, Max, "Confidence Interval Estimation in Non-Linear Regressinn", Journal of the Royal Statistical Society Series B, 25, No. 2 (1963), 330-3.

[13] Hammersly, J. M. and Handscomb, D. C., Monte Carlo Methods, New York: John Wiley and Sons, Inc., 1964.

[14] Hartley, H. O., "The Modified Gauss-Newton Method for the Fitting of Nonlinear Regression Functions by Least Squares", Technometrics, 3 (May 1961), 269-80.

[15] Hartley, H. O., "Exact Confidence Regions for the Parameters in Non-Linear Regression Laws", Biometrika, 51, Nos. 3 and 4 (1964), 347-53.

[16] I.B.M. Corporation, System 360 Scientific Subroutine Package, Version III, White Plains, N.Y.: Technical Publications Department, 1968.

[17] International Mathematical and Statistical Libraries, Inc. IMSL Library 1, Fifth Ed., Houston: International Mathematical and Statistical Libraries, Inc., 1975.

[18] Marquardt, Donald W., "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", Journal of the Society for Industrial and Applied Mathematics, 11 (June 1963), 431-41.

[19] Pratt, John W., "Length of Confidence Intervals", Journal of the American Statistical Association, 56 (September 1961), 549-67.

[20] Scheffé, Henry, The Analysis of Variance, New York: John Wiley and Sons, Inc., 1959.

[21] Sprott, D. A. and Kalbfleisch, John D., "Examples of Likelihoods and Comparison with Point Estimates and Large Sample Approximations," Journal of the American Statistical Association, 64 (June 1969), 468-84.

[22] Turner, Malcolm E., Monroe, Robert J., and Lucas, Henry L., "Generalized Asymptotic Regression and Non-Linear Path Analysis", Biometrics, 17 (March 1961), 120-49.

[23] Williams, E. J., "Exact Fiducial Limits in Non-Linear Estimation", Journal of the Royal Statistical Society Series B, 24, No. 1 (1962), 125-39.

APPENDIX

NUMERICAL METHODS EMPLOYED

The computations were performed on an IBM 370/168, University of Chicago, and on IBM 370/165, Triangle Universities Computation Center, using double precision throughout; the exceptions were the use of two single precision routines, CDTR [16] and GGNOF [17].

The computations for Tables 2 and 3 were performed using the series expansion [10, p. 76]

$$c_T(\theta) = 1 - \Sigma_{i=0}^{\infty} \Sigma_{j=0}^{\infty} p(i;\lambda_{T2}/[c^* - 1]^2)p(j;\lambda_{T1})$$

$$\times \int_0^\infty G(t/[c^* - 1] + 2c^*\lambda_{T2}/[c^* - 1]^2; n-p+2i, 0)g(t;p+2j, 0)dt .$$

The Poisson weights were computed using DLGAM [16] and the integral was evaluated using CDTR [16] and DQL16 [16]. Similar methods were used to compute $c_R(\theta)$ and $c_S(\theta)$.

Simpson's rule, applied to the entries of Tables 2 and 3 (before rounding), was used to obtain the integrals shown in Table 4. The necessary adjustment to accomodate the boundary $\theta_1 \geq \theta_2$ was by means of a trapazoid rule.

The probabilities $P[\varphi_S(y,\theta) = 0|\theta^*,\sigma^2]$ and $P[\varphi_T(y,\theta) = 0|\theta^*,\sigma^2]$ of Table 5 were computed by 4000 Monte Carlo trials using the control variate method of variance reduction [13, p. 59-60]. The control variate was

$$V(\theta) = [(F'e + \theta^* - \theta)'C(F'e + \theta^* - \theta)/p]/[e'Qe/(n-p)]$$

with corresponding test function

$$\varphi_V(e,\theta) = \begin{cases} 1 & V(\theta) > F_\alpha \\ 0 & V(\theta) \leq F_\alpha \end{cases} .$$

Let i index the 4000 trials and let

$$\hat{p}_S = (1/4000)\ \Sigma_{i=1}^{4000}\ \varphi_S(y_i,\theta)\ ,$$

$$\hat{p}_T = (1/4000)\ \Sigma_{i=1}^{4000}\ \varphi_T(y_i,\theta)\ ,$$

$$\hat{p}_V = (1/4000)\ \Sigma_{i=1}^{4000}\ \varphi_V(e_i,\theta)\ ,$$

$$\hat{\sigma}_S^2 = (1/3999)\ \Sigma_{i=1}^{4000}\ [\varphi_S(y_i,\theta) - \hat{p}_S]^2\ ,$$

$$\hat{\sigma}_T^2 = (1/3999)\ \Sigma_{i=1}^{4000}\ [\varphi_T(y_i,\theta) - \hat{p}_T]^2\ ,$$

$$\hat{\sigma}_V^2 = (1/3999)\ \Sigma_{i=1}^{4000}\ [\varphi_V(e_i,\theta) - \hat{p}_V]^2\ ,$$

$$\hat{\sigma}_{VS} = (1/3999)\ \Sigma_{i=1}^{4000}\ [\varphi_V(e_i,\theta) - \hat{p}_V][\varphi_S(y_i,\theta) - \hat{p}_S]\ ,$$

$$\hat{\sigma}_{VT} = (1/3999)\ \Sigma_{i=1}^{4000}\ [\varphi_V(e_i,\theta) - \hat{p}_V][\varphi_T(y_i,\theta) - \hat{p}_T]\ ,$$

$$p_V = P[\varphi_V(e,\theta) = 0 | \theta^*,\sigma^2]\ (= c_S(\theta))\ .$$

The Monte-Carlo estimates were

$$P[\varphi_S(y,\theta) = 0 | \theta^*,\sigma^2] = \begin{cases} \hat{p}_S + p_V - \hat{p}_V & .5 < \hat{\sigma}_{VS}/(\hat{\sigma}_V\hat{\sigma}_S) \\ \hat{p}_S & \text{otherwise} \end{cases}$$

with estimated ~~standard error~~ variance

$$\widehat{SE}^2 = \begin{cases} \{p_V(1-p_V) + c_S(\theta)[1-c_S(\theta)] - 2\hat{\sigma}_{VS}\}/4000 & .5 < \hat{\sigma}_{VS}/(\hat{\sigma}_V\hat{\sigma}_S) \\ c_S(\theta)[1-c_S(\theta)]/4000 & \text{otherwise} \end{cases}$$

and

$$P[\varphi_T(y,\theta) = 0|\theta^*,\sigma^2] = \begin{cases} \hat{p}_T + p_V - \hat{p}_V & .5 < \hat{\sigma}_{VT}/(\hat{\sigma}_V\hat{\sigma}_T) \\ \hat{p}_T & \text{otherwise} \end{cases}$$

with estimated ~~standard error~~ variance

$$\widehat{SE}^2 = \begin{cases} \{\hat{p}_V(1-\hat{p}_V) + c_T(\theta)[1-c_T(\theta)] - 2\hat{\sigma}_{VT}\}/4000 & .5 < \hat{\sigma}_{VT}/(\hat{\sigma}_V\hat{\sigma}_T) \\ c_T(\theta)[1-c_T(\theta)]/4000 & \text{otherwise} \end{cases}$$

The pseudo-random number generator employed was GGNOF [17]. The modified Gauss-Newton algorithm [14] from a start of θ^* was used to compute $\hat{\theta}$ for each Monte-Carlo trial; excepting 14 instances when the algorithm failed to converge for the computations related to Model η_3 . For these 14 cases, $\hat{\theta}$ was computed by grid search over the (θ_1,θ_2) pairs of Table 3.