NONLINEAR ESTIMATION

by

Aaron Hicks Booker

A Dissertation Submitted to the
Graduate Faculty in Partial Fulfilment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1963

TABLE OF CONTENTS

I. INTRODUCTION

The subject of this thesis is the estimation of parameters through independent observations of the sum of a known function of the parameters and an error term which is assumed normal with mean zero and constant variance. The historical method of treating such a problem in estimation is that of least squares, or what is equivalent for this problem, maximum likelihood.

The concept of maximum likelihood came into prominence as a statistical method of estimation as a result of the publications of R. A. Fisher. Fisher's paper (5) of 1922 suggests that in many circumstances the method of maximum likelihood may produce sufficient estimates or estimates which contain all the information in the sample. To facilitate the transition to maximum likelihood from the method of moments, Fisher published in 1925 (6) a comparison of maximum likelihood with the method of moments in terms of efficiency. The paper of 1934 (7) is concerned primarily with conditions for existence of sufficient statistics and their distribution. In none of these papers is a maximum likelihood estimate more uniquely defined than as a solution of the likelihood equation $\frac{\partial}{\partial\theta} \ln \pi g(y_i, \theta) = 0$ where y_i is the observed random variable with the distribution $g(y_i, \theta)$ and θ is the unknown parameter.

The existing mathematical theory of gradient and relaxation methods was sufficient to solve systems of equations such as the likelihood equations. However, the computational complexity was such that the advent of high speed computers

was necessary to make the solutions practical. The knowledge that the solution of the likelihood equation is now computationally practical gives increased importance to the investigation of the statistical properties of the solution. Chapter II attempts to partially answer this question by applying recent developments in the theory of stochastic convergence to derive certain asymptotic properties for each step in an iterative process of solving for the maximum likelihood estimates. Thus a method of estimation is developed in Chapter II which avoids the search for the absolute maximum of the likelihood equation and yet yields statistics which have the same properties as those known for the maximum likelihood estimator.

Methods of constructing exact confidence regions for the parameter are developed and criteria for evaluating the relative merits of the various methods are discussed in Chapter III. There are several results stated in Chapters II and III which require only slight modifications of proofs given in the literature. These proofs are contained in the appendix.

II. NONLINEAR LEAST SQUARES POINT ESTIMATION

1. Introduction - Summary

We are given a set of N responses y_t which have arisen from a nonlinear regression model

$$y_t = f(x_t, \theta) + e_t; \quad t = 1, 2, \ldots, N \tag{1}$$

Here x_t denotes the t^{th} 'fixed' input vector of k elements giving rise to y_t, whilst θ is an m-element unknown parameter vector with elements θ_i and the e_t are a set of N independent error residuals from $N(0, \sigma^2)$ with σ^2 unknown. The expectations of the y_t, are therefore the k + m variable functions $f(x_t, \theta)$ which will be assumed to satisfy certain regularity conditions. The problem is to estimate θ notably by 'Least Squares'.

In this chapter we shall develop an iterative method of solution of the Least-Squares equations which has the following properties

(a) the computational procedure is convergent for finite N

(b) the resulting estimators are asymptotically '100% efficient' as $N \to \infty$.

In sections 2-4 we give a survey of our results leaving the mathematical proofs to sections 5-7 whilst in section 8 we illustrate our method with an example.

Although our theoretical development is oriented towards our specific goals certain results are proved in a somewhat more general form. Some of our theory will be seen to correspond to well known theorems on stochastic limits which have

to be reproved because of certain modifications which we require.

2. The formulation of the large sample theory of the least square point estimator

The estimation of θ by nonlinear least squares (here identical with maximum likelihood estimation) gives rise to the minimization of

$$Q(\theta) \equiv \sum_t [y_t - f(x_t, \theta)]^2 = \text{Min.} \qquad (2)$$

with associated 'least squares equations'

$$\frac{\partial Q(\theta)}{\partial \theta_i} = Q_i(\theta) = 0 \qquad (3)$$

$$i = 1, 2, \ldots, m.$$

Whilst there are iterative methods of solving the nonlinear least squares equations 3 (see e.g. Hartley (8)) it will, in general, not be known whether the solution θ of 3 so obtained is a local minimum of 2 or the absolute minimum, and it is only for this absolute minimum of $Q(\theta)$ that asymptotic optimality properties have been established. The exhaustive scanning of the parameter space is usually computationally impractical, particularly when the number of parameters is > 3 and the conditions on Q ensuring uniqueness of the solutions of equations 3 are usually not satisfied. A method of estimation is therefore developed which avoids the search for the absolute minimum of 2 and yet yields two estimators, $\tilde{\theta}$ and $\mathring{\theta}$, which are asymptotically 100% efficient under fairly general assumptions. The method consists of splitting the N observations into m groups of (say) n observations each so that $N = mn$ and the responses $y_{h\tau}$ arise from k-dimensional inputs $x_{h\tau}$ ($h = 1, 2, \ldots, m$; $\tau =$

1, 2, ..., n). The convex closures $C_h(n)$ of the $x_{h\tau}$ must be disjoint. (See more specific formulation in section 6). The method then consists of two steps:-

Step (i) Construct a consistent estimator θ^* of θ by solving the system of m nonlinear equations .

$$\bar{y}_h = \bar{f}(h, \theta^*) \tag{4}$$

where

$$\bar{y}_h = n^{-1} \sum_{\tau} y_{h\tau}; \quad \bar{f}(h, \theta) = n^{-1} \sum_{\tau} f(x_{h\tau}, \theta)$$

Step (ii) Using θ^* as a 'starting' value carry out <u>one</u> iteration step of the standard Gauss Newton iteration applied to 2 to obtain the 100% efficient estimator $\tilde{\theta}$. As an alternative starting with θ^* the modified Gauss Newton iteration by Hartley (8) may be carried to convergence yielding a local (or absolute) minimum of $Q(\theta)$ at $\theta = \mathring{\theta}$ which is likewise asymptotically 100% efficient. Under certain additional assumptions Huzurbazar (9) showed that $\mathring{\theta}$ will be asymptotically the unique consistent solution of the likelihood equations 3 and hence yield, asymptotically the absolute minimum of $Q(\theta)$.

The main result lies in establishing the consistence of θ^* under very general conditions on f. Computationally the solution of 4 is achieved by driving $\bar{Q}(\theta) = \sum_h [\bar{y}_h - \bar{f}(h, \theta)]^2$ to its minimum value of zero with the help of the 'modified Gauss Newton' iteration. Certain computational shortcuts are introduced. It will be noted that when $f(x, \theta)$ is

linear in θ the estimators $\tilde{\theta}$ and $\mathring{\theta}$ agree with the standard BLUE least squares estimator irrespective of what starting value is used. Briefly we make the following assumptions. (For a more specific statement of our assumptions see section 6.)

The first partial derivatives

$$f_i \equiv \frac{\partial f}{\partial \theta i}(x, \theta) \tag{5}$$

are continuous functions of x and θ where θ is confined to a certain closed convex region, S, of the m-space and the x_t are confined to certain convex closures $C_h(n)$ of the k space. We also assume that the N x m matrix of first partial derivatives f_i has full rank, viz

$$F = [f_i(x_t, \theta)] \quad \text{has rank m} \tag{6}$$

for all θ in S and any set of k-vectors x_t for t = 1, 2, ..., N of which at least m are distinct. Certain minor additional assumptions concerning the function $f(x, \theta)$ will be described in section 3 below.

We shall be concerned with the asymptotic behavior of the above estimators of θ as the sample size $N \to \infty$. More specifically we shall assume for convenience that N is a multiple of m, i.e. that

$$N = n\,m \quad \text{and} \quad n \to \infty \tag{7}$$

Moreover, we shall assume that it will be possible to split the set of x_t vectors into m groups of n vectors $x_{h\tau}$ ($h = 1, 2, \ldots, m$; $\tau = 1, 2, \ldots, n$) in such a way that the convex closures $C_h(n)$ containing the $x_{h\tau}$ are disjoint,

uniformly bounded in n and that the minimum distance of any two points lying in different $C_h(n)$ is bounded away from zero as $n \to \infty$. These restrictions are of a very mild character and can usually be satisfied in a great variety of ways.

A method of finding a solution θ^* of 4 will be given in section 3 and the consistency of θ^* will be proved in section 6.

3. The consistent estimator θ^*

For the computation of the consistent estimator θ^* we require the following lemma, the proof of which is given in section 6.

Lemma:- If we denote the first partial derivatives of the group average (see 4) by

$$\bar{f}_i(h, \theta) = \frac{\partial \bar{f}}{\partial \theta i} \quad (h, \theta) \tag{8}$$

it follows from 6 that the m x n Jacobian of the $\bar{f}_i$ has rank m, i.e. that

$$\text{Rank} \; | \bar{f}_i(h, \theta) | = m \tag{9}$$

for i = 1, 2, ..., m; h = 1, 2, ..., m, and for all θ in S and all $x_{h\tau}$ sets with properties specified in 1.

The estimator θ^* has been defined as a solution of the m nonlinear equations 4 and will be obtained as the absolute minimum of the Least Squares form

$$\bar{Q}(\theta) = \sum_{h=1}^{m} [\bar{y}_h - \bar{f}(h, \theta)]^2 = \min \tag{10}$$

It is clear that any stationary point of $\bar{Q}(\theta)$ is a solution of 4. For a stationary point θ must satisfy the equations

$$0 = \bar{Q}_i(\theta) = \frac{\bar{Q}(\theta)}{\theta_i} = -2 \sum_{h=1}^{m} [\bar{y}_h - \bar{f}(h, \theta)]\bar{f}_i(h, \theta) \qquad (11)$$

for i = 1, 2, ..., m.

Now since the Jacobian $\bar{f}_i(h, \theta)$ has rank m (see 9) any root of the system 11 must satisfy $\bar{y}_h - \bar{f}(h, \theta) = 0$ that is equations 4. Various iterative methods are now available for computing a stationary point of the Least Squares form $\bar{Q}(\theta)$. For example the 'Modified Gauss Newton Iteration' (Hartley (8)) will converge to a stationary point if, in addition to 9, the following assumptions are made about $\bar{Q}(\theta)$.

(i) It is possible to find a 'starting value' ${}_o\theta$ in S such that

$$\bar{Q}({}_o\theta) < \lim \inf \bar{Q}(\theta) \text{ for } \theta \text{ in } \bar{S} \qquad (12)$$

where $\bar{S}$ is the complement of S.

(ii) No two stationary points of $\bar{Q}(\theta)$ yield identical values of $\bar{Q}(\theta)$, which means that 4 has a unique solution. (13)

The above two assumptions 12 and 13, in conjunction with 9, are sufficient to ensure the convergence of the modified Gauss Newton Iteration[1] to a solution of 11 and hence of 4. For a description of these see Hartley (8).

4. The asymptotically efficient estimators $\tilde{\theta}$ and $\overset{o}{\theta}$.

The estimator $\tilde{\theta}$ which is the result of a single Gauss Newton iteration with θ^* as a starting point is obtained as a correction vector $D = \tilde{\theta} - \overset{o}{\theta}$ to θ^* from the m linear equations.

$$\sum_{j=1}^{m} \sum_{h\tau} f_i(x_{h\tau}, \theta^*) f_j(x_{h\tau}, \theta^*) D_j = \sum_{h\tau} [y_{h\tau} - f(x_{h\tau}, \theta^*)] f_i(x_{h\tau}, \theta^*) \qquad (14)$$

[1] If assumption (ii) is not satisfied, the Modified Gauss Newton Method still converges for a subset of the iterative solutions.

The rank of equations 14 is m by the assumption 6. The estimator $\overset{o}{\theta}$ is the limit of the modified Gauss Newton iteration (see Hartley (8)) with θ^* as starting point. Both $\tilde{\theta}$ and $\overset{o}{\theta}$ can be shown to be asymptotically 100% efficient, for $\tilde{\theta}$ this is done in section 7. No such properties can be assured for a stationary point of $Q(\theta)$ (i.e. solution of 3) or indeed for a local minimum of $Q(\theta)$. The symptotic and approximate variances and covariances of both $\tilde{\theta}$ and $\overset{o}{\theta}$ are given by

$$\text{Cov}\overset{o}{\theta}_i\overset{o}{\theta}_j \doteq \text{Cov}\,\tilde{\theta}_i\tilde{\theta}_j \doteq \sigma^2\left[\sum_t f_i(x_t, \theta)f_j(x_t, \theta)\right]^{-1} \qquad (15)$$

and may be estimated by substituting $\tilde{\theta}$ or $\overset{o}{\theta}$ for θ and $Q(\overset{o}{\theta})/(N-m)$ or $Q(\tilde{\theta})/(N-m)$ for σ^2.

5. Some theorems on stochastic limits

The following theorem is a slight modification of theorem 1 given in Mann and Wald[1] (11).

Theorem 1: A sequence of scalar (vector) functions $f_n(x_n)$ of a random vector x_n is such that[2]

$$f_n(x_n) = \text{op}\,[r(n)]\,(\text{Op}) \qquad (16)$$

if and only if for every $\epsilon > 0$ there is a sequence of regions $R_n(\epsilon)$ such that

$$\begin{aligned} &\text{(i)} \quad f_n(a_n) = o[r(n)]\ (0) \quad \text{when } a_n \subset R_n \\ &\text{(ii)} \quad P[x_n \subset R_n\ (\epsilon)] \geq 1-\epsilon \text{ for } n > N(\epsilon) \end{aligned} \qquad (17)$$

---

[1] We understand that the present modification is fully proved in Lecture notes by H. Chernoff.

[2] The Op inside ( ) and subsequent symbols inside ( ) represent alternative forms of the theorems.

Corollary 1.1: Let $x_n^{(j)} = Op[f_j(n)]$ for $j = 1, 2, \ldots, r$ and $R_n(\epsilon)$ be a sequence of subsets of the $k(n)$-dimensional space where $y_n = (y_n^{(1)}, y_n^{(2)}, \ldots, y_n^{(k(n))})$ is such that $P[y_n \subset R_n(\epsilon)] \geq 1 - \epsilon$ for sufficiently large n. Let $g_n(x^{(1)}, \ldots, x^{(r)}, y_n)$ be a sequence of functions such that for every $\epsilon > 0$, $g_n(a_n, b_n) = O[f(n)](o)$ if $a_n^{(j)} = O[f_j(n)]$ and $b_n \subset R_n(\epsilon)$. Then

$$g_n(x_n, y_n) = Op[f(n)] \text{ op } [f(n)] \tag{18}$$

Proof: Let $f_n(x_n)$ of Theorem 1 be $f_n(x_n) = x_n$. Then by (ii) there exist regions for x_n which can be combined with the given regions for y_n to satisfy (ii) for (x_n, y_n). Condition (i) of Theorem 1 is given by the hypothesis of corollary 1.1 and 18 follows directly from Theorem 1.

Corollary 1.2: Let y_n, x_n, z_n be sequences of stochastic vectors with dimensions $k(n)$, r, r respectively. Let $R_n(\epsilon)$ be a sequence of regions such that $P[y_n \subset R_n(\epsilon)] \geq 1 - \epsilon$ for sufficiently large n and $x_n = Op(1)$, $z_n - x_n = Op[f(n)]$ where $\lim_{n \to \infty} f(n) = 0$. Let $G_n(x_n, y_n)$ be a sequence of functions and define $H_n(y_n, x_n, z_n)$ by

$$H_n(y_n, x_n, z_n) = G_n(y_n, x_n) - G_n(y_n, z_n) - T_{sn}(y_n, x_n, z_n) \tag{19}$$

where T_{sn} is the s^{th} order multiple Taylor expansion of G_n about (y_n, x_n) and evaluated at (y_n, z_n). If $G_n(y_n, x_n)$ has continuous and uniformly bounded $(s + 1)^{th}$ order partial derivatives with respect to x provided y_n is in $R_n(\epsilon)$, then $H_n(y_n, x_n, z_n) = op[f^s(n)]$[1].

[1] For $s = 0$ replace the condition on $z_n - x_n$ by $(z_n - x_n) = op(1)$, $f(n)$ bounded.

Proof: Make the following identification of the quantities in corollary 1.1 and 1.2.

Corollary 1.1	Corollary 1.2
x_n	x_n, $x_n - z_n$
y_n	y_n
$f_j(n)$ for x_n	1
$f_j(n)$ for $x_n - z_n$	$f(n)$
$f(n)$	$f^s(n)$
$g_n(x_n, y_n)$	$H_n(y_n, x_n, z_n)$

Thus it is only necessary to show that for every $\varepsilon > 0$ and for any sequency a_n, b_n, c_n such that $a_n \subset R_n(\varepsilon)$, $c_n - b_n = O[f(n)]$, it follows that $H_n(a_n, b_n, c_n) = o[f^s(n)]$. That is, we must verify the property of our function H_n which is stipulated by the o property of the g_n function in corollary 1.1 to which it corresponds.

Since H_n is the remainder term in Taylor's formula for functions of several variables and the mixed $(s + 1)^{th}$ order partial derivatives are bounded by, say B, the sequence H_n can be written

$$\left|H_n(a_n, b_n, c_n)\right| \leq B/(s + 1)! \left(\sum_{i=1}^{r} c_n^{(i)} - b_n^{(i)}\right)^{s+1}. \quad (20)$$

Using the inequality

$$\left(\sum_{i=1}^{r} u_i\right)^N \leq \left(\sum_{i=1}^{r} u_i^2\right)^{N/2} r^{N/2}$$

it follows that

$$\left|H_n(a_n, b_n, c_n)\right| \leq Br^{(s+1)/2}/(s+1)! \left[\sum_{i=1}^{r} (c_n^{(i)} - b_n^{(i)})^2\right]^{(s+1)/2} \quad (21)$$

and consequently

$$\left|H_n(a_n, b_n, c_n)\right| \leq O(\left| c_n - b_n \right|^{s+1}) = o[f^s(n)] \quad (22)$$

where $\left|c_n - b_n\right|$ denotes the modulus of the vector $c_n - b_n$.

6. The consistency of the estimator θ^*

We now return to the model of section 1, that is we consider the nonlinear regression law

$$y_{h\tau} = f(x_{h\tau}, \theta) + e_{n\tau} \quad (23)$$

under the following assumptions:

(i) The convex closures $C_h(n)$ of the $x_{h\tau}$ in the k-dimensional space are contained for all n in convex bounded spaces C_h which (for different h) are disjoint.

(ii) The functions $f_i(x, \theta)$, $f_{ij}(x, \theta)$, and $f_{ijk}(x, \theta)$ are continuous, bounded functions of x and θ for all x $C_h(n)$ and θ S.

(iii) The N by m matrix with elements $f_i(x_\tau, \theta)$ has rank m for all θ S and any set x for $\tau = 1, 2, \ldots, N$ where at least m of the x vectors are distinct.

Note that the lemma of section 3 obtains

(iii′) The m by m matrix $F_n = [\bar{f}_i(h, \theta)]$ has rank m for θ S and $x_{h\tau}$ satisfying (i).

Proof of lemma: Suppose that the $\bar{f}_i(h, \theta)$ had a rank < m for some point θ in S and for some set of $x_{h\tau}$. Then we would have a set of u_i with $\sum_{i=1}^{m} u_i^2 > 0$ and

$$\sum_{i=1}^{m} u_i \bar{f}_i(h, \theta) = 0 \quad (24)$$

for all h = 1, 2, ..., m.

Consider the function

$$G(x) = \sum_{i=1}^{m} u_i f_i(x, \theta) \tag{25}$$

Now from 24 we have for every h = 1, 2, ..., m that

$$n^{-1} \sum_{\tau} G(x_{h\tau}) = n^{-1} \sum_{i} u_i \sum_{\tau} f_i(x_{h\tau}, \theta) = \sum_{i} u_i \bar{f}_i(h, \theta) = 0 \tag{26}$$

But 26 implies that the m group means of the n values of $G(x_{h\tau})$ are zero for every h. It follows that

$$\min_{\tau} G(x_{h\tau}) \leq 0 \leq \max_{\tau} G(x_{h\tau}) \tag{27}$$

Since G(x) is continuous it must take on the value zero at some point $\bar{x}_h$ in the closure $C_h(n)$. That is, we must have

$$0 = G(\bar{x}_h) = \sum_{i=1}^{m} u_i f_i(\bar{x}_h, \theta) \tag{28}$$

for h = 1, 2, ..., m.

Now since the closures $C_h(n)$ are disjoint, equations 28 would contradict assumption (iii). This proves the Lemma.

We now prove

Theorem 2: For any θ S, $|F_1| = |f_i(x, \theta)|$ has the same sign for any two m-vectors ${}_1x$ and ${}_2x$ whose h^{th} elements ${}_1x_h$, ${}_2x_h$ are in C_h.

Proof: Suppose ${}_1x_h$, ${}_2x_h$ C_h for h = 1, 2, ..., m and $|F_1({}_1x_h, \theta)| > 0$, $|F_1({}_2x_h, \theta)| < 0$. Then consider the function of q

$$G(q) = |F_1[({}_1x_h(1-q) + {}_2x_h q), \theta]| . \tag{29}$$

We have $G(0) > 0$ and $G(1) < 0$ and hence, because of the convexity of each C_h, there is a q^* such that $G(q^*) = 0$. Thus

$$\left|F_1[{}_1x_h(1-q^*) + {}_2x_h q^*, \theta]\right| = 0 \tag{30}$$

which contradicts (iii′) for $n = 1$ since ${}_1x_h(1-q) + {}_2x_h q$ is in C_h.

Next we prove

Theorem 3: There is no subsequence F_k of the sequence F_n such that $\lim_{k\to\infty} \left|F_k\right| = 0$.

Proof: Suppose

$$F_k = n_k^{-1} \sum_{\tau=1}^{n_k} f_i(x_{h\tau}, \theta) \tag{31}$$

is such that $\lim_{k\to\infty} \left|F_k\right| = 0$. The determinant $\left|F_k\right|$ may be expressed as the sum of $(n_k)^m$ determinants, say F_p, where $p = 1, 2, \ldots, (n_k)^m$ corresponding to the $(n_k)^m$ ways of choosing t from each column. Thus $\left|F_k\right|$ is the mean of $(n_k)^m$ determinants each in the form of an F_1. By Theorem 2, all these F_1 values have the same sign and hence

$$\text{mod } \left|F_k\right| \geq \min (\text{mod } \left|F_1\right| , \ldots, \text{mod } \left|F_{n_k}\right|) =$$

$$\text{mod } \left|F_{\ell(k)}\right| \tag{32}$$

Thus $\lim_{k\to\infty} \left|F_{\ell(k)}\right| = 0$ which contradicts Theorem 2 and the compactness of the C_h. It follows that $\left|F_n\right|$ is bounded away from zero.

Next we prove

Theorem 4: Let θ^* be any consistent estimate of θ and define

$$I_n(\theta) = -n^{-1}\sigma^{-2} \sum_{h\tau} f_i(x_{h\tau}, \theta)\, f_j(x_{h\tau}, \theta). \tag{33}$$

then we have

$$\begin{aligned} n^{-1}[L''(\theta) - I_n(\theta)] &= o_p(1) \\ n^{-1}[L''(\theta^*) - L''(\theta)] &= o_p(1) \\ n^{-1}[L''(\theta^*)] &= O_p(1) \end{aligned} \tag{34}$$

where $L''(\theta)$ is the m by m matrix of second partial derivatives of the likelihood function

$$L(\theta) = \log \prod_{h\tau} (2\pi)^{-1/2}\sigma^{-1}\exp\left\{-1/2\sigma^2[y_h - f(x_{h\tau}, \theta)^2]\right\}. \tag{35}$$

Proof: Define

$$z_\tau = \sum_{h=1}^{m} \left\{ [y_{h\tau} - f(x_{h\tau}, \theta)]f_{ij}(x_{h\tau}, \theta) - f_i f_j \right\}/\sigma^2 \tag{36}$$

so that $n^{-1}L''(\theta) = n^{-1} \sum_\tau z_\tau$.

Since

$$\begin{aligned} E\, z_\tau &= -\sum_h f_i f_j/\sigma^2 \\ \mathrm{Var}\, z_\tau &= \sum_h f_{ij}^2/\sigma^2 \end{aligned} \tag{37}$$

it follows from assumption (ii) that Var (z_τ) is bounded and since the z_τ are independent, Loeve (10, p.234) showed that

$$\bar{z} - E(\bar{z}) = n^{-1}L''(\theta) - I_n(\theta) = o_p(1) \tag{38}$$

Also from assumption (ii) it follows that $I_n(\theta)$ is bounded so that

$$n^{-1}L''(\theta) = O_p(1) \tag{39}$$

Denote the element ij of $n^{-1}L''(\theta)$ by $u_n^{(ij)}(y, x, \theta)$. Identify the functions G_n of corollary 1.2 with the sequence $u_n^{(ij)}(y, x, \theta)$ and also

Corollary 1.2	Theorem 4	
y_n	$(y_{h\tau}, x_{h\tau})$	
x_n	θ	
z_n	θ^*	for n, m.

It can be shown that regions $R_n(\epsilon)$ exist such that the conditions of corollary 1.2 are satisfied for s = 0, i.e., by assumption (ii) the elements of $n^{-1}L''(\theta)$ are continuous, bounded functions of θ for any $y \subset R_n(\epsilon)$, $\theta \subset S$ and $(z_n - x_n) = op(1)$. Thus

$$u_n^{(ij)}(y, x, \theta^*) - u_n^{(ij)}(y, x, \theta) = op(1) \tag{40}$$

or equivalently,

$$n^{-1}[L''(\theta^*) - L''(\theta)] = op(1) \tag{41}$$

and the combination of 39 and 41 by the rules of algebra concerning OP and op given by Chernoff (1) obtains

$$n^{-1}L''(\theta^*) = Op(1) \tag{42}$$

Next we prove

Theorem 5: The elements of $I_n^{-1}(\theta)$ are bounded.[1]

Proof: It is required to show that the sequence $|I_n(\theta)|$ is bounded away from zero. Suppose there is a subsequence ν of n with $\lim |I\nu(\theta)| = 0$. As $\nu \to \infty$ all m^2 elements of $I\nu(\theta)$

[1]The elements of $I_n(\theta)$ are defined by 33.

are bounded. Hence for a subsequence μ of ν all elements of $I_\mu(\theta)$ converge. Write $I_\mu(\theta) = U_{ij}(\mu, \theta)$, then $\lim_{\mu\to\infty} U_{ij}(\mu, \theta) = U_{ij}(\infty, \theta)$. Since $\left|U(\infty, \theta)\right| = 0$, there exists a vector u such that

$$\sum_{i=1}^{m} u_i^2 > 0 \text{ and}$$

$$\sum_{i=1}^{m} u_i U_{ij}(\infty, \theta) = 0 \text{ for } j = 1, 2, \ldots, m. \tag{43}$$

The μ input vectors $x_{h\tau}$ depend on μ. Since all $x_{h\tau}$ lie in the bounded regions C_h for all τ and μ there exists a subsequence ω of μ (and hence of ν) such that we are able to select from each of the m groups one $x_{h\tau}$ say $x_{h\tau(\omega)}$ for which we have

$$\lim_{\omega\to\infty} x_{h\tau(\omega)} = x_h \text{ for each } h = 1, 2, \ldots, m, \tag{44}$$

and where x_h lies in the closed C_h.

Consider now the function

$$G(x) = \sum_{i=1}^{m} u_i f_i(x, \theta). \tag{45}$$

Using G we obtain for the above subsequence

$$\omega^{-1} \sum_{h\tau} G^2(x_{h\tau}) = \sum_{ij=1}^{m} [\omega^{-1} \sum_{h\tau} f_i(x_{h\tau}, \theta) f_j(x_{h\tau}, \theta)] u_i u_j \tag{46}$$

which can be written

$$\omega^{-1} \sum_{h\tau} G^2(x_{h\tau}) = \sum_{ij=1}^{m} U_{ij}(\omega, \theta) u_i u_j. \tag{47}$$

It follows that for the above sequence

$$\lim_{\omega\to\infty} \omega^{-1} \sum_{h\tau} G^2(x_{h\tau}) = \lim_{\omega\to\infty} \sum_{ij=1}^{m} U_{ij}(\omega, \theta)u_i u_j = 0 \qquad (48)$$

and since $\lim_{\omega\to\infty} x_{h\tau} = x_h$ for h = 1, 2, ..., m

we infer

$$G(x_h) = 0 \quad \text{for } h = 1, 2, \ldots, m. \qquad (49)$$

But 49 implies that

$$\sum_{i=1}^{m} u_i f_i(x_h, \theta) = 0 \quad \text{for} \quad h = 1, 2, \ldots, m \qquad (50)$$

which would contradict assumption (iii) since the x_h, which lie in C_h, are all distinct.

Next we prove

Theorem[1] 6: The statistic θ^* defined as a solution of

$$\bar{y}_h = \bar{f}(h, \theta^*) \qquad (51)$$

is such that

$$\theta^* - \theta = o_p(n^{-1/2}). \qquad (52)$$

Proof: Define $\delta = \theta^* - \theta$. Then

$$\bar{e}_h = \bar{y}_h - \bar{f}(h, \theta)$$

$$= \bar{f}(h, \theta^*) - \bar{f}(h, \theta)$$

$$\bar{e}_h = \sum_{i=1}^{m} \bar{f}_i(h, \theta)\delta_i + 1/2 \sum_{ij} \bar{f}_{ij}[h, \bar{\theta}(h)]\delta_i\delta_j \qquad (53)$$

where $\bar{\theta}(h)$ is on the line segment joining θ and θ^*.

Let

$$A(h) = -1/2\bar{f}_{ij}[h, \bar{\theta}(h)] = (a_{ij}) \qquad (54)$$

Now since F_n is nonsingular by assumption (iii),

---

[1]This theorem clearly establishes the consistency of θ^*.

$$\delta = F_n^{-1}(\bar{e} + \delta' A(h)\delta). \tag{55}$$

For every $\epsilon > 0$, define

$$R_n(\epsilon) = S(\bar{e} : \max |\bar{e}_{h.}| < C(\epsilon)n^{-1/2}) \tag{56}$$

where $C(\epsilon)$ is such that

$$\sigma^2/c^2 < 1 - (1 - \epsilon)^{1/m} \tag{57}$$

Then

$$\Pr[\bar{e} \subset R_n(\epsilon)] = (\int_{-c}^{c} \exp(-\frac{u^2}{2\sigma^2}) \frac{du}{\sqrt{2\pi}\sigma})^m \tag{58}$$

and by Tchebysheff's inequality

$$\Pr[\bar{e} \subset R_n(\epsilon)] > (1 - \sigma^2/c^2)^m > 1 - \epsilon \tag{59}$$

We now define $\delta(\bar{e})$ as the function of the m-vector $\bar{e}$ given by 53 and it remains to show $\delta(b_n) = 0(n^{-1/2})$ for $b_n \subset R_n(\epsilon)$. Let

$$a = \max_{ijhn} a_{ij} \tag{60}$$

and from theorem 3 it is possible to define

$$r = \max_{ijhn} \text{element of } F_n^{-1} \tag{61}$$

We now define

$$q = mrCn^{-1/2} \tag{62}$$

$$t = 1/(2m^3ar) \tag{63}$$

and

$$b = t - t(1 - 2q/t)^{1/2} \tag{64}$$

a solution of

$$q + x^2/2t = x. \tag{65}$$

Thus

$$n^{1/2}b/t = n^{1/2} - (n-\alpha\sqrt{n})^{1/2} \tag{66}$$

where $\alpha = 2mrC/t$.

We now use L'Hospital's rule and the identity $(a^{1/2} - b^{1/2})(a^{1/2} + b^{1/2}) = a - b$ to obtain

$$\lim_{n\to\infty} (n)^{1/2}b/t = \lim_{n\to\infty} \frac{\alpha n^{1/2}}{n^{1/2} + [n-\alpha(n)^{1/2}]^{1/2}} \tag{67}$$

$$= \lim_{n\to\infty} \frac{\alpha}{1 + \frac{n^{1/2} - \alpha/2}{n^{1/2} - \alpha n^{1/2}}} = \alpha/2$$

and $b = 0(n^{-1/2})$.

Let M be sufficiently large that

$$M^{-1/2} < 1/4m^4r^2Ca \tag{68}$$

and for $n > M$

$$m^3rab \leq 1. \tag{69}$$

Then for $b_n \subset R_n(\epsilon)$ and $n > M$ define the sequence $({}_s\delta)$ as follows

$$o^{\delta} = F_n^{-1}b_n$$

$$s+1^{\delta} = F_n^{-1}(b_n + {}_s\delta' A(h) {}_s\delta). \tag{70}$$

thus

$$\max_i \left|s+1^{\delta}{}_i\right| \leq mr \max_t \left|b_{nt}\right| + m^3ar \max_i \left|{}_s\delta_i\right|^2 \tag{71}$$

Or

$$ {}_{s+1}q' \leq q + m^3 a r ({}_s q')^2 \tag{72} $$

where

$$ {}_s q' = \max_i \left| {}_s\delta_i \right| . \tag{73} $$

Now

$$ {}_o q' \leq q \leq b \text{ and if } {}_s q' \leq b, \text{ then} $$
$$ {}_{s+1}q' \leq q + m^3 a r ({}_s q')^2 \leq q + m^3 a r b^2 = b. \tag{74} $$

Write

$$ {}_{s+1}\delta - {}_s\delta = 1/2 F_n^{-1} \left[{}_s\delta' A ({}_s\delta - {}_{s-1}\delta) + ({}_s\delta' - {}_{s-1}\delta') A {}_{s-1}\delta \right) \tag{75} $$

so that

$$ \max_i \left| {}_{s+1}\delta_i - {}_s\delta_i \right| \leq 1/2 m r (m^2 a {}_s q' + m^2 a {}_{s+1} q') \max_i \left| {}_s\delta_i - {}_{s-1}\delta_i \right| \tag{76} $$
$$ \leq m^3 r a b \max_i \left| {}_s\delta_i - {}_{s-1}\delta_i \right| . $$

Thus it is clear that $({}_s\delta)$ converges to ${}_\infty\delta$ a solution of 55. Since $b = O(n^{-1/2})$, it follows from 74 and 73 that for any sequence $b_n \subset R_n(\epsilon)$,

$$ {}_\infty\delta(b_n) = O(n^{-1/2}). \tag{77} $$

Hence

$$ {}_\infty\delta(\bar{e}) = Op(n^{-1/2}) \tag{78} $$

and $\theta^* - \theta = Op(n^{-1/2})$ by theorem 1.

7. <u>Asymptotic 100% efficiency of the estimator $\tilde{\theta}$.</u>

Consider the statistic $\tilde{\theta} = \theta^* + D$ where the quantity D

is defined by the equation

$$-L''(\theta^*)D = L'(\theta^*) \tag{79}$$

Let $\hat{\theta}$ be the 'asymptotically efficient statistic' which satisfies $L'(\hat{\theta}) = 0$.[1] Then it is sufficient for asymptotic efficiency of $\tilde{\theta}$ to show that

$$n^{+1/2}(\theta^* + D - \hat{\theta}) = o_p(1) \tag{80}$$

since

$$x_n = n^{+1/2}(\hat{\theta} - \theta) \tag{81}$$

$$y_n = n^{+1/2}(\theta^* + D - \hat{\theta}) \tag{82}$$

then one obtains the following equation upon using a result given by Doob (3)

$$d_\infty n^{+1/2}(\hat{\theta} - \theta) = d_\infty(x_n + y_n) = d_\infty n^{+1/2}(\tilde{\theta} - \theta). \tag{83}$$

Identify the vectors y, x, z of corollary 1.2 with (y, x), $\hat{\theta}$, θ^*, and let $G_n = n^{-1}L'(\theta)$. The sequence of functions G_n has bounded and continuous second order partial derivatives under assumption (ii). Also, $(\theta^* - \hat{\theta}) = O_p(n^{-1/2})$ so that $f(n) = n^{-1/2}$ and $\lim f(n) = 0$. For $s = 1$, corollary 1.2 obtains

$$H_n(y, \hat{\theta}, \theta^*, x) = n^{-1}L'(\hat{\theta}) - n^{-1}L'(\theta^*) - n^{-1}L''(\theta^*)(\hat{\theta} - \theta^*)$$
$$= o_p(n^{-1/2}) \tag{84}$$

Since $L'(\hat{\theta}) = 0$, 84 can be written using 79

$$n^{1/2}(\theta^* + D - \hat{\theta}) = n[L''(\theta^*)]^{-1} o_p(1) \tag{85}$$

---

[1] For the asymptotic distribution of the maximum likelihood estimator we refer to the literature. We are here merely concerned with proving that the asymptotic variance-covariance of $\tilde{\theta}$ agrees with that of $\hat{\theta}$ to order $O(n^{-1})$.

From Theorem 4

$$n^{-1}L'' (\theta^*) - I_n(\theta) = op(1) \tag{86}$$

Since the elements of $n[L'' (\theta^*)]^{-1}$ are rational functions of the elements of $n^{-1}L'' (\theta^*)$ and $I_n^{-1}(\theta)$ was shown to be bounded in Theorem 5, it follows by Slutsky's theorem as given by Cramer (2)[1] that

$$n[L''(\theta^*)]^{-1} - I_n^{-1}(\theta) = op(1) \tag{87}$$

and consequently

$$n[L''(\theta^*)]^{-1} = OP(1) \tag{88}$$

Now 85 can be written

$$n^{1/2}(\theta^* + D - \hat{\theta}) = op(1) \tag{89}$$

which establishes the asymptotic efficiency of θ.

The modified Gauss Newton method employes the corrective vector D* defined by

$$[\sum_{h\tau} f_i(x_{h\tau}, \theta^*) f_j(x_{h\tau}, \theta^*)]D^* = \sum_{h\tau} [y_h - f(x_{h\tau}, \theta^*)] f_i(x_{h\tau}, \theta^*). \tag{90}$$

By subtracting 79 from 90 one obtains

$$n^{1/2}(D^* - D) = [\sigma^{-2} I_n^{-1}(\theta^*)][n^{-1} \sum_{n\tau} (y - f) f_{ij}][n^{1/2} D] \tag{91}$$

Let G_n of corollary 1.2 be $I_n(\theta^*)$ where we identify x with θ, z with θ^*, and y with $x_{h\tau}$. Since $(\theta^* - \theta) = Op(n^{-1/2})$,

---

[1] A modification of the result is necessary to cover the case where the constants vary with n but satisfy the regularity assumptions (i) and (ii) and alteration of the proof is given in the appendix.

$f(n) = n^{-1/2}$ and G_n has continuous bounded first order partial derivatives with respect to θ, it follows that for s = 0,

$$I_n(\theta^*) - I_n(\theta) = op(1) \tag{92}$$

Again using Slutzky's theorem,

$$I_n^{-1}(\theta^*) - I_n^{-1}(\theta) = op(1) \tag{93}$$

but $I_n^{-1}(\theta) = Op\ (1)$ from Theorem 5 so that

$$I_n^{-1}(\theta^*) = Op(1) \tag{94}$$

Now we apply corrollary 1.2 to the expression

$$G_n(y, x, \theta^*) = n^{-1} \sum_{n\tau} [y_{h\tau} - f(x_h, \theta^*)\ f_{ij}(x_{h\tau}, \theta^*) \tag{95}$$

where we identify y with $(y, x_{h\tau})$, x with θ, and z with θ^* it follows that

$$G_n(y, x, \theta^*) - G_n(y, x, \theta) = op(1). \tag{96}$$

Since $G_n(y, x, \theta)$ is the mean of independent normal random variables

$$\sum_h [y_{h\tau} - f(x_{h\tau}, \theta)]f_{ij}(x_{h\tau}, \theta) \tag{97}$$

each having mean zero and bounded variance, it follows that

$$n^{-1} \sum_{h\tau} [y_{h\tau} - f(x_{h\tau}, \theta)]f_{ij}(x_{h\tau}, \theta) = op(1). \tag{98}$$

and consequently

$$G_n(y, x, \theta^*) = op(1). \tag{99}$$

The right hand side of 91 can be written by using 94 and 99 as follows

$$n^{1/2}(D^* - D) = Op(1)op(1) = op(1) \tag{100}$$

When one identifies $n^{1/2}(D^* - D)$ and $n^{1/2}(\theta^* + D - \hat{\theta})$ with x_n and y_n then equation 83 yields

$$d \circ n^{1/2}(\theta^* + D - \hat{\theta}) = d \circ n^{1/2}(\theta^* + D^* - \hat{\theta}). \qquad (101)$$

Thus the correction vector D* could be used and retain the asymptotic properties of $\hat{\theta}$.

Since it has been shown that $(\tilde{\theta} - \theta) = Op(n^{-1/2})$, a correction of $\tilde{\theta}$ by D defined by

$$-L''(\tilde{\theta})D = L'(\tilde{\theta}) \qquad (102)$$

will produce an asymptotic 100% efficient estimate of θ. Thus each step in the Gauss Newton iterative produce is consistent and asymptotically 100% efficient, provided we fix an upper bound for the maximum number of steps as $n \to \infty$. For all applications of the Gauss Newton iteration it is completely satisfactory to assume that the number of steps is held below a finite, although possibly large, upper bound.

8. A numerical example

In order to illustrate the alogarithms described in sections 2 to 3 we use the exponential law with zero symptotes.

$$y_{h\tau} = \theta_1 \exp(\theta_2 x_{h\tau}) + e_{h\tau} \qquad (103)$$

for $\theta_1 = 1$ and $\theta_2 = -1$. Using a table of random normal deviates $N(0, \sigma^2)$ for the $e_{h\tau}$ and the equidistant series of x-values $x_{h\tau} = (.04) + (\tau - 1)(.05) + .5(h - 1)$ for $h = 1, 2$ and $\tau = 1, \ldots, 10$ we obtain the data shown in Table 1.

The linear equations in δ of the modified Gauss Newton method for θ^* are

Table 1

Data

t	h	τ	x_t	e_t	$\exp(-x_t)$	y_t
1	1	1	.04	-0.84	.96 079	.12 079
2	1	2	.09	1.65	.91 393	2.56 393
3	1	3	.14	-0.38	.86 936	.48 936
4	1	4	.19	-0.38	.82 696	.44 696
5	1	5	.24	-0.74	.78 663	.04 663
6	1	6	.29	0.20	.74 826	.94 826
7	1	7	.34	-1.13	.71 177	- .41 823
8	1	8	.39	0.31	.67 706	.98 706
9	1	9	.44	-0.33	.64 404	.31 404
10	1	10	.49	0.18	.61 263	.79 263
11	2	1	.54	-0.99	.58 275	- .40 725
12	2	2	.59	-0.64	.55 433	- .08 567
13	2	3	.64	-0.26	.52 729	.26 729
14	2	4	.69	0.00	.50 158	.50 158
15	2	5	.74	1.75	.47 711	2.32 711
16	2	6	.79	-1.89	.45 384	-1.43 616
17	2	7	.84	-0.88	.43 171	- .44 829
18	2	8	.89	-0.64	.41 066	- .22 934
19	2	9	.94	-0.74	.39 063	- .34 937
20	2	10	.99	1.08	.37 158	1.45 158

$$\sum_h [\bar{y}_h - \bar{f}(h, {}_o\theta)]\bar{f}_1(h, {}_o\theta) = \delta_1 \sum_h \bar{f}_1^2 + \delta_2 \sum_h \bar{f}_1\bar{f}_2$$

$$\sum_h [\bar{y}_h - \bar{f}(h, {}_o\theta)]\, \bar{f}_1(h, {}_o\theta) = \delta_1 \sum_h \bar{f}_1\bar{f}_2 + \delta_2 \sum_h \bar{f}_2^2 \qquad (104)$$

where

$$\bar{y}_n = 1/10 \sum_{\tau=1}^{10} y_{h\tau}$$

$$\bar{f}(h, {}_o\theta) = 1/10 \sum_{\tau=1}^{10} {}_0\theta_1 \exp({}_0\theta_2 x_{h\tau}) \tag{105}$$

$$\bar{f}_1(h, {}_o\theta) = 1/10 \sum_{\tau=1}^{10} \exp({}_0\theta_2 x_{h\tau})$$

$$\bar{f}_2(h, {}_o\theta) = 1/10 \sum_{\tau=1}^{10} {}_0\theta_1 x_{h\tau} \exp({}_o\theta_2 e_{h\tau})$$

The form $\bar{Q}(\theta)$ is evaluated at ${}_o\theta + \delta$, ${}_o\theta + .9\delta$, ..., ${}_o\theta + .1\delta$, ${}_o\theta + .09\delta$, ... accepting ${}_1\theta$ as the first value such that $\bar{Q}(\theta)$ is reduced. The values in the iterative computation of θ^* are shown in Table 2.

Table 2

Computation of θ^*

Iteration Number	${}_i\theta_1$	${}_i\theta_2$	${}_i\delta_1$	${}_i\delta_2$	$\bar{Q}({}_i\theta)$
1	1.05 34	-1.98 88	.05 377	-.98 88	.85 110
2	1.18 86	-2.68 68	.13 524	-.69 79	.02 164
3	1.23 76	-2.87 04	.04 902	-.18 35	4.13E-5
4	1.24 04	-2.87 88	.00 281	-.00 847	1.8E-10
5	1.24 04	-2.87 88	5.7E-6	-1.4E-5	1.7E-20
6	1.24 04	-2.87 88	5.2E-11	-1.2E-10	0

The computed θ^* is then taken as the starting point in the solution of

$$\sum_t [y_t - f(x_t, \theta^*)]f_1(x_t, \theta^*) = \delta_1 \sum_t f_1^2 + \delta_2 \sum_t f_1 f_2$$

$$\sum_t [y_t - f(x_t, \theta^*)]f_2(x_t, \theta^*) = \delta_1 \sum_t f_1 f_2 + \delta_2 \sum_t f_2^2. \quad (106)$$

The form $Q(\theta)$ is evaluated at $\theta = \theta^* + 2^{-k}\delta$ for $k = 0, 1, \ldots,$ until a value is obtained for which $Q(\theta)$ is reduced. This value then becomes the starting point for the next iteration in the solution of $\overset{o}{\theta}$. The calculation of θ is shown in Table 3.

Table 3

Computation of $\tilde{\theta}$ and $\overset{o}{\theta}$

Iteration Number	$i\theta_1$	$i\theta_2$	$i\delta_1$	$i\delta_2$	$\bar{Q}(i\theta)$
1	1.09 69	-2.59 22	-.14 353	.28 660	15.512
2	1.09 51	-2.56 29	-.11 185	.02 933	15.512
3	1.09 45	-2.56 06	-.00 055	.00 235	15.512
4	1.09 45	-2.56 04	-.00 005	.00 021	15.512

III. EXACT NONLINEAR CONFIDENCE REGIONS

1. Some small sample results for linear least squares

It is well known that if k = m and f is linear in θ then 1 of Chapter II may be written

$$y = X\theta + e \tag{107}$$

where y, θ and e are respectively N x 1, m x 1 and N x 1 vectors and the inputs x_{ti} form the elements of the N x m input matrix X here assumed of rank m. In this case the unique solution of 3 are the least squares estimators

$$\hat{\theta} = (X'X)^{-1}X'y \tag{108}$$

which are BLUE (best, linear, unbiased) and represent a set of m statistics jointly sufficient for the construction of exact joint confidence regions for some or all of the θ_i as well as for the plotting of likelihood contours. We confine our discussion here to the most frequently used ellipsoidal type of confidence regions for the complete θ-vector. Using the decomposition of the sum of squares e'e of the e_t into 'regression' and 'residual'components we write

$$e'e = \text{Reg}(e) + \text{Res}(e) \tag{109}$$

where the first component

$$\text{Reg}(e) = (X'e)'(X'X)^{-1}(X'e) \tag{110}$$

is of rank m and is distributed as $\sigma^2\chi^2$ for m degrees of freedom, while the second component

$$\text{Res}(e) = e'e - \text{Reg}(e) \tag{111}$$

has rank N - m and is independently distributed as $\sigma^2\chi^2$ for N - m degrees of freedom. Accordingly an exact 100α% confidence region for θ is given by

$$\text{Reg}(y - X\theta)/\text{Res}(y - X\theta) \leq mF(\alpha;\ m,\ N - m)/(N - m) \tag{112}$$

where $F(\alpha; m, N - m)$ is the upper $100\alpha\%$ point of F for m, N-m degrees of freedom. It shoud be noted that because of the sufficiency of $\hat{\theta}$ the residual $Res(y - X\theta)$ does not involve θ but only X and y. Confidence regions such as 112 are here taken in the classical sense and the restrictive assumptions under which they are meaningful should be borne in mind.

2. Exact confidence regions in nonlinear estimation

We now return to the nonlinear model 1 and first note that an exact $100\alpha\%$ confidence region of the type 112 can immediately be constructed. Take any decomposition of the sum of squares $e'e$ into two quadratic forms Reg(e) which we assume to be a rank m and $Res(e) = e's - Reg(e)$ of rank N - m, then these two forms will, by Cochran's theorem, be independently distributed as $\sigma^2\chi^2$ for m and N - m degrees of freedom and hence the statement

$$Pr.\left\{Reg.[y - f(x, \theta)]/Res[y - f(x, \theta)] \leq \frac{m}{N - m} F(\alpha; m, N-m)\right\} \tag{113}$$

represents an exact $100\alpha\%$ confidence region for θ. The analogy of 113 to Fiellers (4) quadratic confidence intervals for a ratio of normal means is apparent. The question now arises as to what decomposition of $e'e$ should be chosen, for although the statement 113 is exact for any such decomposition it may be useless as a confidence region. The particular definition of Reg(e) given by 110, which is available in linear estimation theory is statistically meritorious because it is based on a jointly sufficient set of statistics $\hat{\theta}$. In the nonlinear case no such set of sufficient statistics will in general be available. In fact it has been shown that under certain regularily conditions for the $f(x_t, \theta)$ the multivariate normal distribution of the y_t will admit a set

of m statistics jointly sufficient for θ, if and only if $f(x_t, \theta)$ is 'essentially linear'. By this we mean that

$$f(x_t, \theta) = \sum_{i=1}^{m} w_i(\theta) u_{ti} \tag{114}$$

where the $w_i(\theta)$ are continuous functions of the θ_i and the N x m matrix U of the u_{ti} has rank m and does not depend on the θ. While in general $f(x_t, \theta)$ will not be representable in the form 114 it will usually be possible to at least approximately represent the $f(x_t, \theta)$ as m term linear forms of parameter functions $w_i(\theta)$. Particularly desirable are moreover such reparametrisations $w_i = w_i(\theta)$ which represent a one to one mapping of the m-dimensional θ-space to the m-dimensional w-space. In general such approximations can be obtained by truncating expansions into multiinfinite series of complete functions, while special choices of $w_i(\theta)$ suited to the particular $f(x_t, \theta)$ at hand will often give better approximations. As a simple example consider the exponential regression

$$f(x_t, \theta) = \theta_3 + \theta_1 e^{\theta_2 x_t} \tag{115}$$

with $x_t = -\frac{N-1}{2}, \ldots, 0, \ldots, \frac{N-1}{2}$ and N odd.

By expanding $e^{\theta_2 x_t}$ we obtain

$$f(x_t, \theta) \doteq w_1 u_{1t} + w_2 u_{2t} + w_3 u_{3t}$$

where $w_1 = \theta_1\theta_2$, $w_2 = \theta_1\theta_2^2$, $w_3 = \theta_3 + \theta_1$ (116)

$$u_{1t} = x_t,\ u_{2t} = 1/2x_t^2,\ u_{3t} = 1$$

Proceeding with the general case and having constructed an approximation of type 114 to $f(x_t, \theta)$, we now choose as the quadratic form

$$\text{Reg}(e) = (U'e)'\ (U'U)^{-1}(U'e) \tag{117}$$

and

$$\text{Res}(e) = e'e - \text{Reg}(e) \tag{118}$$

for construction of an exact 100α% confidence region 113. It should be noted that the precision of the approximation 114 does in no way affect the exactness of the probability statement 113. However, the denominator quadratic form Res $y - f(x_t, \theta)$ will in general be dependent on θ although with a 'good approximation' 114 this dependence will be 'slight'. If $f(x_t, \theta)$ is linear in θ to start with this method yields, of course, the standard confidence region based on Reg(e) which is invariant with regard to linear transformations. In the nonlinear case there will of course not be a unique choice of U or of Reg(e) given by 117.

In our example of an exponential regression 115 using the approximation 116 to determine U let

$$z_{1t} = x_t,\ z_{3t} = 1,\ z_{2t} = \xi_t = x_t^2 - (N^2 - 1)/12$$

and

$$B_{33} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & -\frac{N^2-1}{12} & 1 \end{pmatrix} \qquad \text{so that } UB = Z$$

where

$$Z = \begin{pmatrix} x_1 & \xi_1 & 1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ x_N & \xi_N & 1 \end{pmatrix} .$$

Thus

$$\begin{aligned} \text{Reg}(e) &= e'UBB^{-1}(U'U)^{-1}B^{-1}B'U'e \\ &= (Z'e)'\,(Z'Z)^{-1}(Z'e). \end{aligned} \tag{119}$$

and since ξ_t are the second degree orthogonal polynomials,

$$Z'Z = \begin{pmatrix} \Sigma x^2 & 0 & 0 \\ 0 & \Sigma \xi^2 & 0 \\ 0 & 0 & N \end{pmatrix} \tag{120}$$

$$Z'e = \begin{pmatrix} \Sigma x_t e_t \\ \Sigma \xi_t e_t \\ \Sigma e_t \end{pmatrix}$$

so that

$$\text{Reg}(e) = ne^{-2} + (\Sigma x_t^2)^{-1}(\Sigma e_t x_t)^2 + (\Sigma \xi_t^2)^{-1}(\Sigma e_t \xi_t)^2. \tag{121}$$

Using this definition of Reg(e) in 117, 118, and 113 and spelling out the matrix notation we obtain a confidence region for θ as follows;

$$0 \leq c\Sigma_t (y_t - \theta_3 - \theta_1 e^{\theta_2 x_t})^2 - (1 + c)\left\{N(\bar{y} - \theta_3 - N^{-1}\theta_1 \Sigma_t e^{\theta_2 x_t})^2\right. \tag{122}$$

$$+ S^{-1}\left(\sum_t y_t x_t - \theta_1 \sum_t x_t e^{\theta_2 x_t}\right)^2 + S_2^{-1}\left(\sum_t y_t \xi_t - \theta_1 \sum_t \xi_t e^{\theta_2 x_t}\right)^2 \Big\} \qquad (122)$$

where

$$c = \frac{3}{N-3} F(\alpha;\ 3,\ N-3)$$

$$S_1 = \sum_t x_t^2; \qquad S_2 = \sum_t \xi_t^2 \qquad (123)$$

The complete tabulation of confidence regions such as 122 in the θ_1, θ_2, θ_3 space is, of course, a computational tedium. However, it is comparatively easy to test for any particular θ whether or not it lies in the confidence region.

For the numerical example of Chapter I, a 95% confidence region is defined by

$$\text{Reg}(e)/\text{Res}(e) \leq \frac{2}{18}(3.55) \qquad (124)$$

where the U matrix defining Reg(e) by 117 is

$$\begin{pmatrix} .04 & 1 \\ .09 & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ .94 & 1 \\ .99 & 1 \end{pmatrix}$$

Of course the approximation of e^{-x} by $1-x$ is not too good for $0 \leq x \leq 1$. The method described in section 1 was illustrated by a 3-parameter exponential and involved a three term quadratic as an approximation to $\theta_3 - \theta_1 e^{-\theta_2 x}$. To

test if $\theta_2 = -1$ and $\theta_1 = 1$ is in the 95% region,

$$\text{Reg}(e) = \begin{pmatrix} \Sigma\, x_t e_t \\ \Sigma\, e_t \end{pmatrix}' \begin{pmatrix} 6.967 & 10.3 \\ 10.3 & 20. \end{pmatrix}^{-1} \begin{pmatrix} \Sigma\, x_t e_t \\ \Sigma\, e_t \end{pmatrix} \tag{125}$$

$$= \begin{pmatrix} -2.588 \\ -4.57 \end{pmatrix} \begin{pmatrix} .6015 & -.30977 \\ -.30977 & .20953 \end{pmatrix} \begin{pmatrix} -2.588 \\ -4.57 \end{pmatrix}$$

$$= 1.075$$

$\text{Res}(e) = \Sigma\, e_t^2 - \text{Reg}(e) = 17.169 - 1.075 = 16.093$ and $\text{Reg}(e)/\text{Res}(e) \leq .394$ so $\theta_2 = -1$, $\theta_1 = 1$ lies inside the exact 95% confidence region.

A better approximation of $\theta_1 e^{\theta_2 x_t}$ is obtained by defining U and W as in 116 with the exception $w_3 = \theta_1$ since θ_3 in the model of Chapter I is known to be zero. An exact 95% confidence region is now determined by

$$U'U = \begin{pmatrix} \Sigma\, x_t^2 & 1/2\Sigma\, x_t^3 & \Sigma\, x_t \\ 1/2\Sigma\, x_t^3 & 1/4\Sigma\, x_t^4 & 1/2\Sigma\, x_t^2 \\ \Sigma\, x_t & 1/2\Sigma\, x_t^2 & 20 \end{pmatrix} = \begin{pmatrix} 6.967 & 2.650 & 10.3 \\ 2.650 & 1.075 & 3.48 \\ 10.3 & 3.4835 & 20. \end{pmatrix} \tag{126}$$

and again for $\theta_2 = -1$, $\theta_1 = 1$.

$$\text{Reg}(e) = \begin{pmatrix} \Sigma\, x_t e_t \\ 1/2\Sigma\, x_t^2 e_t \\ \Sigma\, e_t \end{pmatrix}' \quad (U\ U)^{-1} \quad \begin{pmatrix} \Sigma\, x_t e_t \\ 1/2\Sigma\, x_t^2 e_t \\ \Sigma\, e_t \end{pmatrix}$$

$$= \begin{pmatrix} -2.588 \\ -.82338 \\ -4.570 \end{pmatrix}' \begin{pmatrix} 10.27 & -18.774 & -2.0192 \\ -18.774 & 36.455 & 3.319 \\ -2.0192 & 3.319 & .51175 \end{pmatrix} \begin{pmatrix} -2.588 \\ -.82338 \\ -4.570 \end{pmatrix}$$

$$= 1.1944$$

Res(e) = 17.169 - 1.1944 = 15.975

so that Reg(e)/Res(e)$\leq$ .565 and again the parameter point $\theta_2 = -1$, $\theta_1 = 1$ is inside the confidence region.

An asymptotic 95% confidence region for θ_1 and θ_2 is given by

$$\begin{pmatrix} \theta_1 - \tilde{\theta}_1 \\ \theta_2 - \tilde{\theta}_2 \end{pmatrix} \begin{pmatrix} \sum_t e^{2x_t\theta_2} & \sum_t \theta_1 x_t e^{2\theta_2 x_t} \\ \sum_t \theta_1 x_t e^{2\theta_2 x_t} & \sum_t \theta_1^2 x_t^2 e_t^{2\theta_2 x_t} \end{pmatrix} \begin{pmatrix} \theta_1 - \tilde{\theta} \\ \theta_2 - \tilde{\theta}_2 \end{pmatrix} \qquad (128)$$

$$- \frac{2}{18} Q(\tilde{\theta}) F(.95;\ 2,\ 18) \leq 0.$$

To test if $\theta_1 = 1$, $\theta_2 = -1$ is inside 128 for the data of the example in Chapter II,

$$\begin{pmatrix} 1. & - & 1.0945 \\ -2.5604 & + & 1. \end{pmatrix}' \begin{pmatrix} 3.5852 & .80574 \\ .80574 & .31801 \end{pmatrix} \begin{pmatrix} - & .0945 \\ -1.5604 \end{pmatrix} - (15.512)(.394) \qquad (129)$$

$$< 0$$

so the true parameter point is contained inside the asymptotic confidence region.

Tables 4 and 5 indicate the values of θ_1, θ_2 which satisfy 124 and 128 for two grid sizes.

Tabular value	θ Satisfies
0	Neither
1	124
2	128
3	124 and 128

In looking at these tables it must be remembered that it is not to be expected that the confidence regions should coincide, not even approximately. They are constructed from different forms of Reg(e) and only the exact one has a confidence coefficient of 0.95 exactly. Even if they were both exact they may well differ with regard to the regions in the θ-space they cover. Table 6 indicates the values of θ_1, θ_2 which satisfy

$$\text{Reg(e)/Res(e)} \leq \frac{3}{17}F(95\%,\ 3,\ 17) \qquad (130)$$

for U defined by 116 in the following way

$$(1) \implies 130$$
$$(0) \implies \theta_1,\ \theta_2 \not< 130.$$

Since the U matrix given by 116 refers to a three parameter problem, the confidence region will be a three dimensional region in the θ_1, θ_2, θ_3 space. The two dimensional region in the θ_1, θ_2 plane defined by 130 represents the cross section of the three dimensional region in the plane $\theta_3 = 0$.

3. Approximate linearization of the regression law by Lagrangean interpolation

In this section we develop a method of constructing the 'essentially linear' approximation of the type 114 to the nonlinear regression law $f(x_t, \theta)$ in the special case of a

single input variable x_t i.e., for k = 1. We shall make the following regularity and continuity assumptions about the behavior of $f(x, \theta)$

(a) $f(x, \theta)$ has continuous derivatives up to order m with regard to its argument x over the range $x_t(\min) \leq x \leq x_t(\max)$ and for all θ.

(b) $f(x, \theta)$ has continuous 1st order derivatives

$$\frac{\partial f(x, \theta)}{\partial \theta_j} = f_j(x, \theta) \tag{131}$$

with regard to all m elements θ_j of the parameter vector θ,

(c) For any set of distinct input values x_i the matrix

$$[f_j(x_i, \theta)] \tag{132}$$

has rank m for all θ. Moreover, the following assumption is made about the input values x_t.

(d) At least m of the N input variables x_t are distinct.

We now choose a set of m 'representative values' x_i covering the x-range. These may or may not coincide with some of the x_t and must have the property that

$$\begin{aligned} & x_t(\min) \leq x_i \leq x_t(\max) \\ \text{and} \quad & x_i \neq x_{i'}, \\ & \text{for } i = 1, 2, \ldots, m \text{ and } i \neq i' \end{aligned} \tag{133}$$

Apart from 133 the choice of the x_i is governed by considerations of making the remainder term 141 in the Lagrangean approximation (now to be developed) as small as possible. A specially convenient choice is to make the

x_i equidistant and of the form

$$x_i = x_t(\min) + (i - 1/2)m^{-1}[x_t(\max) - x_t(\min)]. \qquad (134)$$

The N values $f(x_t, \theta)$ of our regression function will now be approximated by their Lagrangean interpolates between the $f(x_i, \theta)$. Writing the standard Lagrangean interpolation coefficients in the form

$$u(x, i) = \prod_{i' \neq i}^{m} (x - x_{i'})/ \prod_{i' \neq i}^{m} (x_i - x_{i'}) \qquad (135)$$

we obtain the Lagrangean interpolate between the $f(x_i, \theta)$ in the form

$$f_{lag}(x_t, \theta) = \sum_{i=1}^{m} u(x_t, i) f(x_i, \theta). \qquad (136)$$

Identifying $u(x_t, i)$ in 136 and 135 with u_{ti} in 114 and $f(x_i, \theta)$ with $w_i(\theta)$ we recognize 136 as an 'essentially linear' approximation to $f(x_t, \theta)$. However, we have to prove that the rank of (u_{ti}) is m and examine the conditions under which the reparametrization

$$w_i(\theta) = f(x_i, \theta) \qquad (137)$$

represents a unique mapping of the θ space into the w-space. Finally some discussion of the remainder term in the approximation 136 is appropriate.

Consider then the elements of the N x m matrix

$$u_{ti} \equiv u(x_t, i) = \prod_{i' \neq i}^{m} (x_t - x_{i'})/ \prod_{i' \neq i}^{m} (x_i - x_{i'}). \qquad (138)$$

If the rank of the u_{ti} were smaller than m there would exist a set of c_i with $\sum c_i^2 = 1$ such that

$$\sum_{i=1}^{m} c_i u_{ti} = 0 \quad t = 1, 2, \ldots, N \tag{139}$$

But 138 and 139 would imply that the $(m-1)^{th}$ degree polynomial $\sum_{i=1}^{m} c_i u(x, i)$ would have all x_t as roots. This would only be possible if fewer than m of the x_t were distinct which would contradict assumption (d).

It should be noted for any i' for which $x_{i'}$ coincides with an x_t, we automatically have $u_{t',i'} = 1$ and $u_{t',i} = 0$ for $i \neq i'$ and in such a case

$$f_{lag}(x_t, \theta) = f(x_{t'}, \theta).$$

The mapping 137 of the θ space into the w-space will be 'locally' one to one because of condition(c). A much stronger sufficient condition for uniqueness 'in the large' would be given by the definiteness of the quadratic form

$$\pm \sum_{ij} f_j(x_i, \theta) V_i V_j > 0 \tag{140}$$

for all θ and V_i with $\sum V_i^2 = 1$.

The remainder term of the Lagrangean interpolation formula 136 will provide a gauge for the difference between $f_{lag}(x_t, \theta)$ and $f(x_t, \theta)$. From standard finite difference calculus we obtain

$$f(x_t, \theta) - f_{lag}(x_t, \theta) = \prod_{i=1}^{m} (x_t - x_i) f_m(\xi, \theta)/m! \tag{141}$$

where f_m is the m^{th} derivative of f with regard to x and $x_t(\min) < \xi < x_t(\max)$.

It should be stressed again that the accuracy of the approximation given by 141 does in no way affect the exactness of the confidence region 113 if the quadratic forms Reg(e) and Res(e) are defined by 117 and 118 and the matrix $U = [u(x_t, i)]$ by 135. However, any failure in the accuracy of the approximation 136 will result in larger departures of the boundary of the confidence region from contours of constant likelihood:- For the contours of constant likelihood are given by

$$e'e = [y - f(x, \theta)]'[y - f(x, \theta)] = \text{const} \qquad (142)$$

and these agree with the boundary of the confidence region 113 if $\text{Res}[y - f(x, \theta)]$ does not involve θ. But this latter condition is certainly satisfied if $f(x, \theta) = f_{lag}(x, \theta)$. We do not enter here into the question as to why we consider it a desirable principle that the boundaries of confidence regions should approximately coincide with contours of constant likelihood.

4. Selectivity of exact nonlinear confidence regions

A confidence region is said to be most selective or 'shortest' in the Neyman sense provided that the region is exact and the probability of covering false values of the parameter is less than the probability for any other region.

In the nonlinear model $y = f(x, \theta) + e$ the transformation

$$f(x, \theta) = UW + z \qquad (143)$$

obtains

$$y = UW + z + e$$

where $U(N \times m)$ is dependent only on x, $W(m \times 1)$ dependent on θ and z is the error of approximating $f(x, \theta)$ by UW, in

general dependent on both x and θ. The exact confidence region

$$Q_1/Q_2 \leq \frac{m}{N - m} F(\alpha;\ m,\ N - m) \tag{144}$$

was constructed using the equation

$$\begin{aligned} e'e &= (U'e)'(U'U)^{-1}U'e + e'[I - U(U'U)^{-1}U']e \\ &= e'Ae + e'Be \\ &= Q_1(e) + Q_2(e) \end{aligned} \tag{145}$$

Let θ be the true parameter, θ_o be an arbitrary value in the parameter range and define $e_o = y - f(x,\ \theta_o)$, then

$$e_o'e_o = Q_1(e_o) + Q_2(e_o).$$

Thus Q_1, Q_2 are independently distributed $X'^2_{\lambda_1,m}$, $X'^2_{\lambda_2,N-m}$ respectively where

$$\begin{aligned} \lambda_1 &= [f(x,\ \theta) - f(x,\ \theta_o)]'A[f(x,\ \theta) - f(x,\ \theta_o)] \\ \lambda_2 &= [f(x,\ \theta) - f(x,\ \theta_o)]'B[f(x,\ \theta) - f(x,\ \theta_o)] \end{aligned} \tag{146}$$

Rewriting λ_1 and λ_2 using 131

$$\begin{aligned} \lambda_1 &= [W - W(\theta_o)]'(U'U)[W - W(\theta_o)] + [z - z(\theta_o)]'A[z - z(\theta_o)] \\ &\quad + 2[W - W(\theta_o)]'U'[z - z(\theta_o)] \\ \lambda_2 &= [z - z(\theta_o)]'B[z - z(\theta_o)] \end{aligned} \tag{147}$$

If the function $f(x,\ \theta)$ is linear in θ or essentially linear then λ_2 and the last two terms of λ_1 vanish so that the non-centrality is concentrated in Q_1. Since $[Q_1(\theta_o)/m]/[Q_2(\theta_o)/(N-m)]$ has doubly non-central F distribution

$$\frac{N-m}{m} e^{\frac{-(\lambda_1+\lambda_2)}{2}} \sum_{r=0}^{\infty}\sum_{s=0}^{\infty} \frac{\lambda_2^s \lambda_1^r}{2^{r+s} r!s!} J_{m+2r,N+2s-m}^{(N-m)x/m} \tag{148}$$

where $J_{m1,m2}(x)$ is the distribution function of x_{m1}^2/x_{m2}^2, the probability of θ_o satisfying equation 144 is the integral of 148 from 0 to $F(\alpha;\ m,\ N-m)$.

In order to study the dependence of this probability on the non-centralities λ_1, λ_2 given by 146 it is convenient to employ certain analytic approximations. Patnaik (12) showed that a non-central X'^2 statistic based on n degrees of freedom is approximately distributed as δX_v^2 where X_v^2 is a central X^2 statistic based on v degrees of freedom with v given by

$$v = (n + \lambda)^2/(n + 2\lambda) \tag{149}$$

and the scale factor δ is given by

$$\delta = (n + 2\lambda)/(n + \lambda). \tag{150}$$

It follows that

$$\Pr\left\{Q_1/Q_2 \le \frac{m}{N-m} F(\alpha;\ m,\ N-m)\right\} \tag{151}$$

$$\doteq \Pr\left\{F(v_1, v_2) \le \frac{1 + \lambda_2/(N-m)}{1 + \lambda_1/m} F(\alpha;\ m,\ N-m)\right\}$$

where $F(v_1, v_2)$ is an F-ratio statistic based on

$$v_1 = (m + \lambda_1)^2/(m + 2\lambda_1)$$

$$v_2 = (N - m + \lambda_2)^2/(N - m + 2\lambda_2). \tag{152}$$

It is clear from 151 and 152 that the probability 151 will <u>decrease</u> if λ_2 is kept small while λ_1 is <u>increased</u>. The

major effect is the scale factor $[1 + \lambda_2/(N - m)]/[1 + \lambda_1/m]$ which would decrease with increasing λ_1 and with λ_2 kept moderately small. The second effect is the formula for v which can be written

$$v = n[1 + \lambda^2/n(n + 2\lambda)]$$

so that both v_1 and v_2 will exceed m and N - m respectively but v_2 only slightly so since N - m is large.

The effect of increasing both degrees of freedom on the probability 151 can be assessed from an inspection of tables of % points of F and will be found almost negligible compared with the effect of the scale factor.

The merits or demerits of confidence regions can therefore be essentially assessed by an evaluation of the scale factor

$$\frac{1 + \lambda_2/(N - m)}{1 + \lambda_1/(m)} \tag{153}$$

in conjunction with the formulas 146 for λ_1 and λ_2 which indicate their dependence on the matrices A and B and hence through 145 on the choice of Reg(e).

This assessment of the merits of confidence regions is particularly gratifying. For if our definition of Reg(e) yields a larger scale factor than an alternative definition of Reg(e), the former will be superior to the latter for <u>all</u> levels of confidence coefficients. However, the comparative values of the scale factor may depend on the 'false' value of θ. Finally our approximate formulas show qualitatively

that a definition of Reg(e) coinciding with likelihood contours is advantageous.

A somewhat less satisfactory indicator of the merits of confidence regions is given by the mean of the non-central F distribution 148 which, however, can be calculated exactly. If $F = m_2 x_{m1}^2 / m_1 x_{m2}^2$, then

$$E(m_1 F/m_2) = m_1/(m_2 - 2). \tag{154}$$

The mean of 148 is

$$\frac{N - m}{m} e^{\frac{-(\lambda_1 + \lambda_2)}{2}} \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} \frac{\lambda_1^r \lambda_2^s}{2^{r+s} s! r!} [(m + 2r)/(N + 2s - m - 2)] \tag{155}$$

which can be written

$$\frac{N - m}{m} (m + \lambda_1)[e^{-\lambda_2/2} \sum_{s=0}^{\infty} \frac{\lambda_2^s}{2^s s!} \frac{1}{N + 2s - m - 2}]. \tag{156}$$

Let $\lambda_2/2 = u^2$, then

$$\sum_{s=0}^{\infty} \frac{\lambda_2^s}{2^s s!} \frac{1}{N + 2s - m - 2} = \sum_{s=0}^{\infty} \frac{u^{2s}}{s!} \frac{1}{N - m - 2 + 2s} \tag{157}$$

Integrating the identity

$$u^{N - m - 3} e^{u^2} = \sum_{s=0}^{\infty} \frac{u^{N - m - 3 + 2s}}{s!} \tag{158}$$

term by term obtains

$$\int_0^u x^{N - m - 3} e^{x^2} dx = \sum_{s=0}^{\infty} \frac{u^{N - m - 2 + 2s}}{s!} \frac{1}{N - m - 2 + 2s} \tag{159}$$

so that 156 can be written

$$\frac{N-m}{m}(m+\lambda_1)\left[e^{-\lambda_2/2}\ \frac{\lambda_2}{2}\ \frac{-(N-m-2)}{2}\ \int_0^{(\lambda_2/2)^{1/2}} x^{N-m-3}e^{x^2}dx\right] \tag{160}$$

and in series form when $N - m$ is even and ≥ 4

$$\frac{N-m}{m}(m+\lambda_1)\left\{1/2[\lambda_2/2)^{-1} - \ell(\lambda_2/2)^{-2} + \ldots + (-1)^{\ell}\,\ell!(\lambda_2/2)^{-(\ell+1)} - (-1)^{\ell}\,\ell!(\lambda_2/2)^{-(\lambda+1)}e^{-\lambda_2/2}]\right\} \tag{161}$$

where $\ell = (N - m - 4)/2$.

5. Some remarks about alternative approaches to the problem

We should mention that a recent paper by E.J. Williams (14) implies that in certain situations Reg(e) may be constructed in analogy to 110 by

$$\text{Reg}(e) = (F'e)'(F'F)^{-1}(F'e) \tag{162}$$

where $F = [f_i(x_t, \theta)]$ is the N x m matrix of differentials. This definition of Reg(e) is a quadratic from in the e_t whose coefficients depend on θ. Nevertheless Res(e) = e'e - Reg(e) can be shown to only slightly depend on θ provided the second differentials f_{ij} are small. Williams only deals with the case of a single parameter θ, assumes that an independent 'error mean square' is available for the estimation of σ^2, does not use Res(e) at all and regards Reg(e) as the 'natural' way of constructing confidence intervals. It should be noted that the use of the F-matrix 162 in place of the U matrix in 145 introduces a θ-dependence into the matrix elements of A and B.

It is clear that the dependence on θ of the coefficients of Reg(e) or Res(e) does in no way affect the exactness of the confidence region 144, since for a given and fixed θ the quadratic forms would still follow the x^2 distribution. Reference should also be made to the paper by Turner, M.E., Monroe, R.J., and Lucas, H.L. (13) where confidence intervals are based on an expansion in powers of f.

Table 4

Exact and asymptotic 95% confidence regions

θ_2 ↓ θ_1 →	1.0			1.2			1.4			1.6			1.8			2.0			2.2			2.4			2.6		
- .50	3	3	3	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .55	3	3	3	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .60	3	3	3	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .65	3	3	3	3	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .70	3	3	3	3	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .75	3	3	3	3	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .80	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .85	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .90	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- .95	3	3	3	3	3	3	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.00	3	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.05	3	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.10	3	3	3	3	3	3	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.15	3	3	3	3	3	3	3	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.20	3	3	3	3	3	3	3	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.25	3	3	3	3	3	3	3	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.30	3	3	3	3	3	3	3	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.40	3	3	3	3	3	3	3	3	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.45	3	3	3	3	3	3	3	3	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.50	3	3	3	3	3	3	3	3	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.55	3	3	3	3	3	3	3	3	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.60	3	3	3	3	3	3	3	3	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.65	3	3	3	3	3	3	3	3	3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.70	3	3	3	3	3	3	3	3	3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.75	3	3	3	3	3	3	3	3	3	3	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.80	3	3	3	3	3	3	3	3	3	3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.85	3	3	3	3	3	3	3	3	3	3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.90	3	3	3	3	3	3	3	3	3	3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-1.95	3	3	3	3	3	3	3	3	3	3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.00	3	3	3	3	3	3	3	3	3	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.05	3	3	3	3	3	3	3	3	3	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.10	3	3	3	3	3	3	3	3	3	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.15	3	3	3	3	3	3	3	3	3	3	3	3	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.20	3	3	3	3	3	3	3	3	3	3	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.25	3	3	3	3	3	3	3	3	3	3	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.30	3	3	3	3	3	3	3	3	3	3	3	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.35	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.40	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.45	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.50	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0
-2.55	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0
-2.60	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0
-2.65	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0
-2.70	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0

Table 5

Exact and asymptotic 95% confidence regions

$\theta_2\downarrow$ $\theta_1\rightarrow$	-4		-2		0		2		4		6		8		10		12		14		16		18		20		22		24
10	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 1	0	0	0	0	3	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 2	0	0	0	0	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 3	0	0	0	0	3	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 4	0	0	0	0	1	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 5	0	0	0	0	1	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 6	0	0	0	0	1	3	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 7	0	0	0	0	1	1	3	3	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 8	0	0	0	0	1	1	3	3	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
- 9	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-10	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-11	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-12	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
-13	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
-14	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
-15	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0
-16	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
-17	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
-18	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-19	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-20	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-21	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-22	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-23	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-24	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-25	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-26	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-27	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-28	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-29	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-30	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-31	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
-34	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table 6

Exact 95% region using U with rank 3

θ_2 \ θ_1	-6	-4	-2	0	2	4	6	7	10	12	14	16	18	20	22
10	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
9	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
8	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
7	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
6	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
5	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
4	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
3	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
2	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
1	0 0 0	0 0 0	1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
0	0 0 0	0 0 0	1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 1	0 0 0	0 0 0	1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 2	0 0 0	0 0 0	1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 3	0 0 0	0 0 0	1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 4	0 0 0	0 0 0	1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 5	0 0 0	0 0 0	1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 6	0 0 0	0 0 0	1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 7	0 0 0	0 0 0	1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 8	0 0 0	0 0 0	1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
- 9	0 0 0	0 1	1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-10	0 0 0	0 1	1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-11	0 0 0	0 1	1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-12	0 0 0	0 1	1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-13	0 0 0	0 1	1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-14	0 0 0	0 1	1 1	1 1 1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-15	0 0 0	0 1	1 1	1 1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-16	0 0 0	0 1	1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-17	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-18	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-19	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-20	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-21	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-22	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-23	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-24	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-25	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-26	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-27	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-28	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-29	0 0 0	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-30	0 0 1	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-31	0 0 1	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-32	0 0 1	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-33	0 0 1	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0
-34	0 0 1	1	1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	1 1 1	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0	0 0 0

IV. APPENDIX

The notation and following definitions were used by H.B. Mann and A. Wald (11).

Definition 1: $a_n = o[f(n)]$ if $\lim_{n \to \infty} |a_n|/f(n) = 0$

where f(n) is a positive real function defined on the positive integers.

2: $a_n = O[f(n)]$ if $|a_n| \leq Mf(n)$ for all n and a fixed value $M>0$.

3: $x_n = o_p[f(n)]$ if for every $\epsilon > 0$

$$\lim_{n \to \infty} P[|x_n|/f(n) < \epsilon] = 1$$

4: $x_n = O_p[f(n)]$ if for every $\epsilon > 0$ there exists an A_ϵ such that

$P[|x_n| \leq A_\epsilon f(n)] \geq 1 - \epsilon$ for every n.

5: $d^\infty(y_n) = d(y)$ reads the asymptotic distribution of y_n is that of y.

Next we prove

Theorem 1: A sequence of functions $f_n(x_n)$ of a random vector x_n is such that

$$f_n(x_n) = o_p[r(n)] \quad (Op) \tag{163}$$

if and only if for every $\epsilon > 0$ there is a sequence of regions $[R_n(\epsilon)]$ such that

(i) $f_n(a_n) = o[r(n)]$ when $a_n \subset R_n(\epsilon)$ (0)

(ii) $P[x_n \subset R_n(\epsilon)] \geq 1 - \epsilon$ for $n >$ some $N(\epsilon)$. (164)

Proof: First it will be proved that 164 $\Rightarrow$ 163.

For $\epsilon > 0$, there exists an $N(\epsilon)$ such that

$$\sup\left\{\left|f_n(a_n)\right| \ : \ a_n \subset R_n(\epsilon)\right\} = s_n(\epsilon)$$

is finite for $n \geq N(\epsilon)$. Suppose otherwise, then for every $N(\epsilon)$ there is an $n > N(\epsilon)$ such the $s_n(\epsilon)$ is not defined. Thus for an infinite number of values of n there exist a_n's in $R_n(\epsilon)$ such that $\left|f_n(a_n)\right| > nr(n)$ which contradicts (i). For $n \geq N(\epsilon)$ let $a_n(\epsilon) \subset R_n(\epsilon)$ such that $\left|f_n\,[a_n(\epsilon)]\right| \geq s_n(\epsilon)/2$. By (i), for $\delta > 0$, there exists an $N(\epsilon, \delta)$ such that for $n \geq N(\epsilon, \delta)$, $\left|f_n[a_n(\epsilon)]\right| /r(n) < \delta/2$. Then for all $a_n \subset R_n(\epsilon)$,

$$\left|f_n(a_n)\right| \leq s_n(\epsilon) \leq 2\left|f_n[a_n(\epsilon)]\right| < \delta r(n). \tag{165}$$

Thus $R_n(\epsilon) \subset S\left\{x_n: \left|f(x_n)\right| /r(n) \leq \delta\right\}$ and it follows that for $n > \max\left\{N(\epsilon), N(\epsilon, \delta)\right\}$

$$P\left\{\left|f_n(x_n)\right| /r(n) \leq \delta\right\} \geq P\left\{x_n \subset R_n(\epsilon)\right\} \geq 1 - \epsilon \tag{166}$$

Since this is true for every $\epsilon > 0$,

$$\lim_{n\to\infty} P\left\{\left|f_n(x_n)\right| /r(n) \leq \delta\right\} = 1 \text{ or } f_n(x_n) = o_p[r(n)].$$

To prove that 163 $\Rightarrow$ 164, define:

$$c_n(\epsilon) = \min\left\{c: P[\left|f_n(x_n)\right| \leq c] \geq 1 - \epsilon\right\}. \tag{167}$$

By 163, for $\epsilon > 0$ and $\delta > 0$ there is an $N(\epsilon, \delta)$ such that for $n > N(\epsilon, \delta)$

$$P\ f_n(x_n)\ /r(n) < \delta\ \geq 1 - \epsilon \tag{168}$$

and it follows that

$$c_n(\epsilon) \leq r(n)\delta \tag{169}$$

therefore $$c_n(\epsilon) = o[r(n)]. \tag{170}$$

Let $R_n(\epsilon) = S\left\{a_n: |f_n(a_n)| \leq c_n(\epsilon)\right\}$. It is clear that the $R_n(\epsilon)$ satisfy the conditions of 164.

Note that the dimensionality of $R_n(\epsilon)$ is the number of components in the vector x_n which may be a function of n.

The following lemma and theorem establish the modified form of Slutsky's theorem as used in Section 7 of Chapter I.

Lemma: Let $B_n = (b_n^{(1)}, \ldots, b_n^{(p)})$ and $A_n = (a_n^{(1)}, \ldots, a_n^{(p)})$ be vector sequences contained in a compact set S such that

$$\lim_{n\to\infty} \left\{B_n - A_n\right\} = 0. \tag{171}$$

Then for any rational function G of p variables defined over S it follows that

$$\lim_{n\to\infty} [G(B_n) - G(A_n)] = 0. \tag{172}$$

Proof: Since G is a rational function defined over the compact set S, G is uniformly continuous over S. Thus for any $\epsilon > 0$, there is a $\delta(\epsilon)$ such that for every x and x' for which $|x - x'| < \delta(\epsilon)$ it is true that $|G(x) - G(x')| < \epsilon$. From 171 there is an N_o such that for $N > N_o$, $|A_n - B_n| < \delta(\epsilon)$, therefore for $N > N_o$, $|G(A_n) - G(B_n)| < \epsilon$.

Next we prove

Theorem 7: Let $x_n = (x_n^{(1)}, \ldots, x_n^{(p)})$ be a sequence of stochastic vectors and $A_n = (a_n^{(1)}, \ldots, a_n^{(p)})$ be a sequence of constants contained in a compact set S satisfying $(x_n - A_n) = op(1)$. Any rational function G of the components of x_n

such that G(x) is defined for $x \subset S$ satisfies,

$$G(x_n) - G(A_n) = op(1). \tag{173}$$

Proof: First apply Theorem 1 where

$$f_n(x_n) = x_n - A_n = op(1). \tag{174}$$

Thus, for every $\epsilon > 0$, there is a sequence of regions $R_n(\epsilon)$ such that

$$\begin{aligned} &(i) \quad B_n - A_n = o(1) \quad \text{when} \quad B_n \subset R_n(\epsilon) \subset S \\ &(ii) \quad P[x_n \subset R_n(\epsilon)] \geq 1 - \epsilon. \end{aligned} \tag{175}$$

Now (i) and the hypothesis that G(x) is defined satisfy the hypothesis of the preceeding lemma to obtain

$$(i') \quad G(B_n) - G(A_n) = o(1).$$

Since (i') and (ii) satisfy the hypothesis for Theorem 1 when $f_n'(x_n) = G(x_n) - G(A_n)$, it follows that

$$f_n'(x_n) = op(1)$$

or

$$G_n(x_n) - G(A_n) = op(1).$$

It was shown in Chapter II that under certain regularity conditions the asymptotic distribution of $\tilde{\theta}$ and $\overset{\circ}{\theta}$ is that of $\hat{\theta}$, the absolute maximum of the likelihood equation. The distribution of $\hat{\theta}$ is now considered by

Theorem 8: Let $\hat{\theta}$ be the solution of $L'(\theta) = 0$ which yields the absolute minimum of $Q(\theta)$. If $(\hat{\theta} - \theta) = Op(n^{-1/2})$, then $n^{1/2}(\hat{\theta} - \theta)$ has an asymptotic multivariate normal distribution with mean zero and variance-covariance matrix

$$-I_n^{-1} = n\sigma^2 [\sum_{h\tau} f_i(x_{h\tau}, \theta) f_j(x_{h\tau}, \theta)]^{-1}. \tag{176}$$

Proof: Using the following table of identification,

Corollary 1.2	Theorem 8
y_n	y_n, x_n
x_n	θ
z_n	$\hat{\theta}$
$f(n)$	$n^{-1/2}$
$G_n(y_n, x_n)$	$n^{-1}L'(\theta)$

(177)

the conditions of corollary 1.2 are satisfied for s = 1 which obtains

$$n^{-1}L'(\hat{\theta}) - n^{-1}L'(\theta) - n^{-1}L''(\theta)(\hat{\theta} - \theta) = op(n^{-1/2}) \quad (178)$$

which can be written

$$n^{1/2}(\hat{\theta} - \theta) = [-n^{-1}L''(\theta)]^{-1}[op(1) + n^{-1/2}L'(\theta)]. \quad (179)$$

From 39 of section 7 of Chapter II

$$n^{-1}L''(\theta) = Op(1) \quad (180)$$

so that the asymptotic distribution of $n^{1/2}(\hat{\theta} - \theta)$ is that of

$$[-n^{-1}L''(\theta)]^{-1}[n^{-1/2}L'(\theta)].$$

From 87 of Section 7 of Chapter II

$$[n^{-1}L''(\theta)]^{-1} - I_n^{-1}(\theta) = op(1) \quad (181)$$

and since $n^{-1/2}L'(\theta)$ is multivariate normal with mean zero and variance-covariance matrix $-I_n(\theta)$, it follows that $n^{1/2}(\hat{\theta} - \theta)$ is asymptotically multivariate normal with mean zero and variance-covariance matrix $-I_n^{-1}(\theta)$.

V. BIBLIOGRAPHY

1. Chernoff, H. Large sample theory: parametric case. Annals of Mathematical Statistics. 27: 1-22. 1956.

2. Cramer, H. Mathematical methods of statistics. Princeton, New Jersey. Princeton University Press. 1945.

3. Doob, J.L. Limiting distributions of certain statistics. Annals of Mathematical Statistics. 6: 160-169. 1935.

4. Fieller, E.C. The biological standardization of insulin. Supplement of the Journal of the Royal Statistical Society. 7, Supplement: 1-23. 1940.

5. Fisher, R.A. The goodness of fit of regression formulae and the distribution of regression coefficients. Journal of the Royal Statistical Society. 85: 597-612. 1922.

6. Fisher, R.A. Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society. 22: 700-725. 1925.

7. Fisher, R.A. Two new properties of mathematical likelihood. Proceedings of the Royal Society. Series A, 144: 285-307. 1934.

8. Hartley, H.O. Modified Gauss Newton method for the fitting of nonlinear regression functions. Technometrics. 3: 269-280. 1961.

9. Hurzurbazar, V.S. The likelihood equation consistency and the maxima of the likelihood function. Annals of Eugenics. 14: 185-198. 1948.

10. Loeve, M. Probability theory. Princeton, New Jersey. D. Van Nostrand Company. 1955.

11. Mann, H.B. and Wald, A. On stochastic limit and order relationships. Annals of Mathematical Statistics. 13: 217-226. 1943.

V. BIBLIOGRAPHY (CONTINUED)

12. Patnaik, P.B. The non-central X^2 and F distributions and their applications. Biometrika. 36: 202-232. 1949.

13. Turner, M.E., Monroe, R.J., and Lucus, H.L. Generalized asymptotic regression and nonlinear estimation. Biometrics. 17: 120-129. 1961.

14. Williams, E.J. Exact fiducial limits in nonlinear estimation. Journal of the Royal Statistical Society. Series B, 24: 125-139. 1962.