ON THE LEAST SQUARES ESTIMATION OF PARAMETERS IN NONLINEAR MODELS

By

ROGER HUGHES MOORE

Bachelor of Science
University of Oregon
Eugene, Oregon
1951

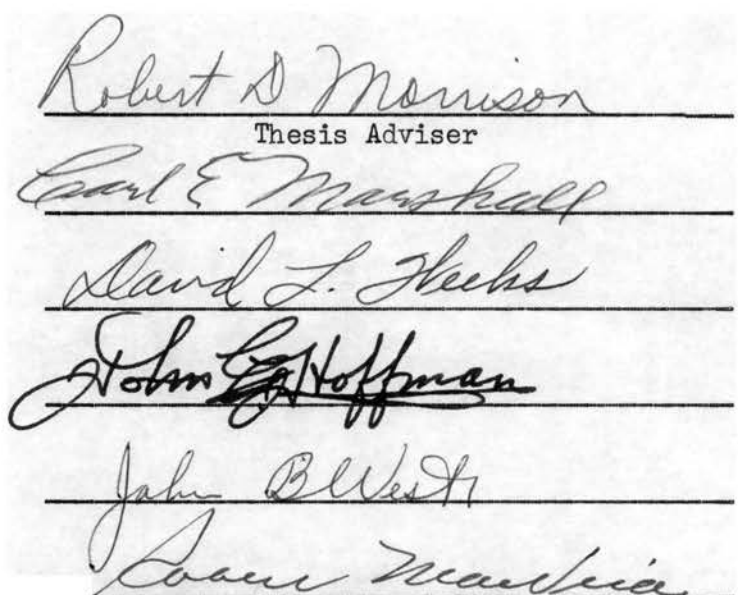
Master of Science
University of Oregon
Eugene, Oregon
1953

Submitted to the Faculty of the Graduate School of
the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
May, 1962

ON THE LEAST SQUARES ESTIMATION OF PARAMETERS

IN NONLINEAR MODELS

Thesis Approved:



Thesis Adviser

Dean of the Graduate School

PREFACE

With regard to the general subject of this thesis, I am perhaps most deeply indebted to Dr. R. Keith Zeigler of the Los Alamos Scientific Laboratory, Los Alamos, New Mexico. It was he who introduced me to the problem of nonlinear estimation and kept encouraging the development of high-speed computer methods for the Gauss iterative technique until they are now in routine use at the LASL and have been initiated at several other research organizations.

When the time came for more detailed study of the problem, I was fortunate to be chosen by the Advisory Committee to participate in the LASL's Advanced Study Program. Financial support from the Laboratory, which is operated by the University of California under contract with the United States Atomic Energy Commission, is gratefully acknowledged by my family and myself. A special note of gratitude is accorded the many members of Laboratory Group T-1, of which Bengt G. Carlson is leader, for their assistance in carrying out the computing reported in this thesis and for their general interest in my progress.

For their courses and for their assistance while acting as my Graduate Advisory Committee, I am indebted to the following Oklahoma State University faculty members: Professors Carl E. Marshall, David L. Weeks, Robert A. Hultquist, John L. Folks, John E. Hoffman, and John B. West.

Dr. Robert D. Morrison was my thesis adviser. His guidance is gratefully acknowledged.

Even with all the help given by those listed above and by many others who remain unmentioned but not forgotten, these would have been long and dreary years without my family. To my parents, Frank and Nana Moore, who managed to see me through my undergraduate years and have since assisted with astoundingly keen insight: I trust that this accomplishment is as satisfying to you as it is to me. To my kids, Mike and Dan and Susie, who never really got used to my not being at home every night: I hope I can make it up and that in twenty years you will understand. To my wife Eileen, who kept me going with cajolery and affection and who typed the final draft of this thesis: I can only restate my love.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

The estimation of parameters that appear in various functional relationships is of general statistical and scientific interest. For many problems, the method of least squares is a desirable and satisfactory technique. Historically, it seems that Karl Friedrich Gauss was the driving force behind the development and enlargement of the method. It is of interest to note that a translation of Gauss' least squares work has been prepared by Trotter (1957).

As one peruses this material, it appears that Gauss was on the verge of stating the principle of maximum likelihood. Of course, the clear enunciation of this principle had to wait for Fisher (1922, 1925) to state it in some generality. Indeed, it is now clear that Gauss' least squares criterion for the estimation of parameters in many relationships is equivalent to maximum likelihood when certain of the variables in the relationships are assumed to be normally distributed.

Out of this background has grown an area of statistical interest called Linear Hypotheses, which includes such seemingly diverse concepts as multiple regression and analysis of variance. An important tool in Linear Hypotheses is the concept of model building, in which a mathematical expression is prepared in an effort to describe observational data. This expression is the "model" for the data and is used

extensively in preparing the theoretical bases for tests of hypotheses, confidence intervals, estimation procedures, and similar statistical matters.

If y is a random variable, X is a mathematical variable, and α is a parameter, it will be said that the equation $H(y, X, \alpha) = 0$ is a model. If, in addition, we can assume that y is normally distributed with mean μ and variance σ^2, we will speak of the following: "The model $H(y, X, \alpha) = 0$ where $y \sim N(\mu, \sigma^2)$." Moreover, each of the arguments may be multi-dimensional. For example, y may be a 2-element vector with a particular bivariate distribution, and α may be a vector of parameters. Based on these considerations, we may define a model as follows:

> Definition I.1. A model is a mathematical equation involving random variables, mathematical variables, and parameters. Information about the distributions of the random variables is considered part of the model.

A long-accepted practice has been the division of models into two classes: those that are "linear" and those that are not. The definition given by James and James (1959, p. 235) says, "An equation or expression is said to be linear in a certain variable [their emphasis] if it is of the first degree in that variable." Using this concept, we can state the next definition:

> Definition I.2. A linear model is a model that is linear in the random variables and in the parameters.

We note particularly that a model may be linear even if it contains nonlinear functions of the mathematical variable X. For example, suppose we have $H(y, X, \alpha) = y - (\alpha_1 + \alpha_2 X) = 0$. This model is linear because the random variable y and both parameters, α_1 and α_2, appear in the function H in a linear fashion. But, by our definition, the model $G(y, X, \alpha) = y - (\alpha_1 + \alpha_2 X^2) = 0$ is also linear. Both of these models are forms of a general polynomial model. On the other hand, suppose $F(y, X, \alpha) = y - \sin(\alpha X) = 0$ is prescribed. This function is linear in the random variable y, but it is nonlinear in the parameter α. This particular function, then, qualifies as a nonlinear model.

Just as Graybill (1961, pp. 103-5) classifies linear models into at least five useful types, so does it serve our present purpose to consider the classification of nonlinear models. One step is the subdivision of nonlinear models into three types: (1) nonlinear only in the parameters, (2) nonlinear only in the random variables, and (3) nonlinear in both. Further subdivision seems to be a prodigious task because of the infinite variety of statistical models.

We have chosen to limit our considerations in this work to models of the type which are nonlinear only in the parameters. With this in mind, we are able to propose a generalization of model 1 as presented by Graybill (1961, p. 103). We shall call this generalization "models of Type I."

Definition I.3. A model of type I is any function $y = f(X; \alpha) + e$ in which y is an observable random variable, X is a J-element vector of known mathematical variables, α is a K-element vector of unknown parameters, and e is an unobservable random variable whose mean is assumed to be zero.

The semicolon separating the arguments X and α in the function f is introduced only to emphasize the difference between the mathematical variables and the parameters. We make particular note of the fact that this type of model preserves the linearity of the random variables as required by our self-imposed limitation, while no restrictions are placed on the parameters.

Model 1, as defined by Graybill, fits this definition, and any results obtained for models of Type I apply immediately to his model. To see this, recall that the linear model 1 is given as

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i X_i + e.$$

If we define X as a vector whose transpose is $X' = (1, X_1, X_2, \ldots, X_k)$ and β a vector whose transpose is $\beta' = (\beta_0, \beta_1, \beta_2, \ldots, \beta_k)$ and let $J = K = (k + 1)$, we see how Graybill's model 1 is a special case of models of Type I.

There is another model which qualifies as meeting Graybill's specifications and, thereby, ours. This is the model in which the vector X is itself a function of another vector of mathematical variables, say X* whose transpose is $(X^*)' = (X_1^*, X_2^*, \ldots, X_K^*)$. Setting $X_1 = f_1(X^*)$, $X_2 = f_2(X^*)$, ..., and $X_k = f_k(X^*)$, we can obtain the model given by $y = \beta_0 + \beta_1 f_1(X^*) + \beta_2 f_2(X^*) + \ldots + \beta_k f_k(X^*) + e$ $= f(X^*; \beta) + e$.

The main reason for bringing this last model into our discussion is to point out that a model should be formulated in terms of its basic arguments. There is no advantage, for example, to be gained in talking about the model "$y = \alpha X + e$ where $X = \sin X^*$" when we can

just as well (and without losing sight of our objectives) refer to the model "$y = \alpha \sin X^* + e$." Of course, we must realize that α is the parameter and that X^* is the mathematical variable--but that realization must go with any model we form.

Many functions belong to the Type I class. Whenever a function also belongs to Graybill's model 1, it means that the parameters appear linearly in the function. The estimation of parameters in this function, whether it be a hyperplane or an analysis of variance model, usually requires little more than application of results already contained in a highly developed statistical domain. Such problems therefore have, for the most part, solutions that are both esthetically pleasing and practically satisfying.

The same cannot be said for a model in which even one of the parameters appears nonlinearly. The basic difficulty, as Turner (1959, pp. 3-8) points out, is simply that sufficient estimators of nonlinear parameters do not exist in general. In addition, even if we were willing to accept a standard technique of estimation, such as least squares, we would be faced with having to solve systems of nonlinear equations. This is hardly a pleasing prospect.

Trying to circumvent the problems associated with solving nonlinear equations, many authors have proposed other estimating procedures. These invariably require certain restrictions on the values allotted to the vector X. Sometimes it is necessary also to perform a transformation on the model, so that linearity of the random variables is violated.

By way of illustration, consider the "sum of exponentials" model

$$y = \sum_{k=1}^{K/2} a_k e^{\lambda_k X} + e$$

where K is an even integer. With reference to our Type I, we note that X is a scalar, the K-element vector of parameters has elements $a_1, \lambda_1, a_2, \lambda_2, \ldots, a_{K/2}, \lambda_{K/2}$, and the random variable e has mean zero. This model has a great many uses, particularly in the physical and biological sciences. Attesting to its usefulness is the great amount of literature available in which various authors discuss the problem of estimating the parameters. From Prony's method, discussed by Whittaker and Robinson (1944, pp. 369-70), to a fairly recent proposal by Cornell (1956), a recurrent theme appears: the data must be equally spaced with respect to values of X, the number of points at which observations are made often should be some multiple of K/2 (the number of "periods"), and, having taken all these precautions, the experimenter must beware of the all-too-many cases in which estimates of the parameters are complex numbers. Hildebrand (1956, pp. 380-1) gives a particularly discouraging example in which Prony's method gives estimates which are either real or complex, depending upon a slight degree of rounding in the observations y. This obstacle never faces the practitioner who remains loyal to linear models.

CHAPTER II

THE GENERAL LEAST SQUARES PROBLEM AND THE GAUSS ITERATIVE METHOD

In a manner similar to Chapter 6 of Graybill (1961), consider the frequency function $g(y; X_1, X_2, \ldots, X_J; \alpha_1, \alpha_2, \ldots, \alpha_K)$ of a random variable y, which depends on J known quantities $X_1, X_2, \ldots, X_J$ and on K unknown parameters $\alpha_1, \alpha_2, \ldots, \alpha_K$. We shall assume that the expected value of y is a function of the X_j and the α_k; i.e., $E(y) = f(X; \alpha)$ where X and α are vectors whose elements are, respectively, the X_j and the α_k. Furthermore, we assume nothing in particular about the variance of y, denoting it temporarily only by σ^2_y.

If the transformation $e = y - f(X; \alpha)$ is performed, it follows that e is a random variable such that $E(e) = 0$ and the variance of e is σ^2_y. The transformation could also be written $y = f(X; \alpha) + e$. We note especially that this is the form required for a model to be of Type I.

In order to estimate the parameter vector α, we will consider having taken a random sample of size I from the distribution of y. We note that a sample consists of the selection of a J-element vector X and a corresponding random observation y. Denoting the general element of the sample by i, we obtain a system of relationships among

the observations:

$$y_i = f(X_i; \alpha) + e_i, \qquad i = 1, 2, \ldots, I,$$

where X_i is a vector of J elements and α is a vector of K elements.

This can be written in matrix notation as

$$Y = F(X; \alpha) + E$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_I \end{bmatrix}; \qquad F(X; \alpha) = \begin{bmatrix} f(X_1; \alpha) \\ f(X_2; \alpha) \\ \cdots \\ f(X_I; \alpha) \end{bmatrix}; \qquad E = \begin{bmatrix} e_1 \\ e_2 \\ \cdots \\ e_I \end{bmatrix};$$

$$X_i = \begin{bmatrix} X_{i1} \\ X_{i2} \\ \cdots \\ X_{iJ} \end{bmatrix}, \quad i = i, 2, \ldots, I; \qquad \text{and} \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_K \end{bmatrix}.$$

To obtain the least squares estimate of the parameter vector α, we could use an adaptation of Aitkin's proposal (1934-5). That is, suppose the error vector E has a known covariance matrix, say V. Then, the least squares estimate of α would be the vector, say a, such that the generalized sum of squares $[Y - F(X; \alpha)]' V^{-1} [Y - F(X; \alpha)]$ is minimized. This vector a is also the maximum likelihood estimate of α when E is assumed to be distributed as an I-variate normal whose mean vector is null and whose covariance matrix is V.

For our work, we shall assume that the elements of E have zero for their expected values, have the same variance σ^2, and are uncorrelated. Thus, V is simply σ^2 times the identity matrix and the generalized sum of squares is $\sigma^{-2} [Y - F(X; \alpha)]' [Y - F(X; \alpha)]$. However, to avoid continued reference to the factor σ^{-2}, we will be concerned with minimizing the sum of squares of the errors, E'E. Because the vector Y and the vectors X_i are fixed for a given sample, we will use the notation $Q(\alpha) = E'E = [Y - F(X; \alpha)]' [Y - F(X; \alpha)]$, indicating that the sum of squares is a function of the parameter vector. When, in addition, we impose the condition that Y is an I-variate normal, the maximum likelihood estimate of σ^2 is Q(a)/I, where Q(a) denotes the minimum of the sum of squares of the errors. Furthermore, Cramér (1951, pp. 500-4) states that under certain general conditions the estimating vector a is normally distributed with mean α and a specific covariance matrix.

In order to make possible the statement of a formal minimization procedure, certain assumptions and notation will be given as the discussion developes. The first of these assumptions follows.

> Assumption II.1. The function $f(X_i; \alpha)$, $i = 1, 2, \ldots, I$, has continuous first and second derivatives with respect to each element of the vector α.

Therefore, for a given set of observations (y_i, X_i) where $i = 1, 2, \ldots, I$, the sum of squares of the errors is a continuous and twice-differentiable function of the parameter vector α.

P will denote an IxK matrix of the first order partial derivatives of the vector $F(X; \alpha)$ with respect to the vector α. In general, the

matrix P will be a function of α. Thus, $P(\alpha) = [p_{ik}(\alpha)]$ where the element in the $\underline{i}$th row and $\underline{k}$th column is $p_{ik}(\alpha) = \partial f(X_i; \alpha)/\partial \alpha_k$.

With Assumption II.1, we see that a necessary condition that $Q(\alpha)$ has a minimum at $\alpha = a$ is that $\partial Q/\partial \alpha_k = 0$ for $k = 1, 2, \ldots, K$ when the derivatives are evaluated at the point $\alpha = a$. This system of K equations in K unknowns can be summarized in matrix notation, after dropping the common factor 2, as

$$P'(a)[Y - F(X; a)] = \varphi \qquad \text{(II.1)}$$

where $P'(a)$ is the transpose of the matrix $P(\alpha)$ evaluated at $\alpha = a$, $F(X; a)$ is the vector $F(X; \alpha)$ evaluated at $\alpha = a$, and φ is a null vector of K elements. These equations are commonly called the "normal equations."

Formally, then, all we need to do to find the least squares estimate of α is to solve this system of equations for the elements of the vector a. In general, however, Equations (II.1) are nonlinear in the vector α. Although a great deal of mathematical effort and ingenuity has been devoted to the problem of solving simultaneous nonlinear equations, each set of such equations poses its own difficulties and no really satisfying general approach has been forthcoming.

Gauss' technique of expanding the model $y = f(X; \alpha) + e$ in a first-order Taylor's series in α and iteratively seeking a value of α that minimizes $Q(\alpha)$ is a well-known and widely used procedure. Nevertheless, for the sake of completeness, it is worthy of some discussion here. We begin by examining a form of Graybill's model 1 and noting how the least squares estimates of its parameters are obtained.

The model we will consider is of the matrix form

$$Y = X\beta + E \qquad \text{(II.2)}$$

where Y is an I-element vector of observed random variables, X is an IxK matrix of observed mathematical variables, β is a K-element vector of unknown parameters to be determined, and E is an I-element vector of unobservable random variables which have a constant variance and are uncorrelated. For this linear model Equations (II.1) simplify and take the form $(X'X)b = X'Y$. If, in addition, we assume that the matrix X has rank K, we know that X'X also has rank K and therefore the least squares estimate of β is $b = (X'X)^{-1}X'Y$. Moreover, because X'X is positive definite and independent of β, we know that the vector b so determined is the unique value at which the sum of squares, $Q(\beta)$, is a minimum.

Let us return now to our models of Type I and examine Gauss' method. We begin by choosing in some manner a vector a_0 which we hope in some sense is "close" to the minimizing vector a. We rewrite the model $y = f(X; \alpha) + e$ in the form of a first-order Taylor's series about the point a_0 to obtain a new model

$$y = f(X; a_0) + \sum_{k=1}^{K} \frac{\partial f(X; a_0)}{\partial \alpha_k} \delta_{k,1} + \underline{e}.$$

Whenever a_0 appears in an expression, such as $f(X; a_0)$, it means that the expression is to be evaluated at a_0. The symbol $\delta_{k,1} = (\alpha_k - a_{k,0})$ where $a_{k,0}$ is the $\underline{k}$th element of the vector a_0. The symbol $\underline{e}$ contains the error e and all additional error introduced in the expansion of the function. Thus, we have $\underline{e} = e +$ (all terms in the series for

$f(X; \alpha)$ of order higher than the first).

Suppose now that this series expansion has been performed for each of the I observations (y_i, X_i). We define F_0 to be the vector $F(X; \alpha)$ evaluated at $\alpha = a_0$. We also let P_0 be the matrix $P(\alpha)$ at $\alpha = a_0$. We let δ_1 be a K-element vector whose kth element is $\delta_{k,1}$. Finally, we let $\underline{E}_0$ be an I-element vector whose ith element is $\underline{e}_i$. Then, in matrix notation, the model can be written $Y = F_0 + P_0\delta_1 + \underline{E}_0$. To finally obtain the form (II.2), we let $Z_0 = Y - F_0$ and write our new model as

$$Z_0 = P_0\delta_1 + \underline{E}_0. \qquad \text{(II.3)}$$

If we can assume that P_0 is a matrix of rank K, we can apply the theory applicable to Graybill's model 1 and obtain a least squares estimate of δ_1 by minimizing $q(\delta_1) = \underline{E}_0'\underline{E}_0 = (Z_0 - P_0\delta_1)'(Z_0 - P_0\delta_1)$. This estimate, say d_1, is given by $d_1 = (P_0'P_0)^{-1}P_0'Z_0$. Now, because we have written the vector $\delta = \alpha - a_0$, we will take $a_1 = a_0 + d_1$ as an "improved" estimate of the vector a which we have assumed to be the value of α which minimizes $Q(\alpha)$.

The procedure is now repeated. That is, we replace the initial estimating vector a_0 by the improved vector a_1. The vector F and the matrix P are now evaluated at a_1 and denoted by F_1 and P_1, respectively. The vector of corrections is now denoted by δ_2. Thus, we are now in the position of finding the estimate, say d_2, which minimizes $q(\delta_2) = \underline{E}_1'\underline{E}_1 = (Z_1 - P_1\delta_2)'(Z_1 - P_1\delta_2)$. The new estimate, say a_2, of the minimizing vector a is defined by $a_2 = a_1 + d_2$.

Continued repetition of the procedure leads to the mth iteration in which the mth vector of corrections is computed by minimizing

$q(\delta_m) = (\underline{E}_{m-1})'(\underline{E}_{m-1}) = (Z_{m-1} - P_{m-1}\delta_m)'(Z_{m-1} - P_{m-1}\delta_m)$. From this, we obtain the equation of the estimate of δ_m as

$$d_m = [(P_{m-1})'(P_{m-1})]^{-1}(P_{m-1})'(Z_{m-1}).$$

When, after say the $\underline{M}$th iteration, the estimating vector correction d_M is deemed "sufficiently small," we will say that the procedure has converged and that $a_M = a_{M-1} + d_M$ is our estimate of the true minimizing vector a.

In order that the foregoing may be a useful procedure, we must make Assumption II.2 as follows:

Assumption II.2. Each of the matrices $P_0, P_1, \ldots, P_m, \ldots$ is of rank K.

If this assumption is not made, there is no guarantee that there is a unique solution for each system of equations and the procedure is vitiated.

This minimization procedure may be characterized as follows: We have a function, say $Q(\alpha)$, of the vector α of K variables. We assume the existence of a vector, say a, at which the function attains a minimum. Starting with a trial vector, say a_0, we generate a sequence of vectors by the method given above such that the vector a is a limit point of the sequence $a_0, a_1, a_2, \ldots, a_m, \ldots$. In addition, we are assuming that the first time $[a_m - a_{m-1}] = [d_m]$ is "small" implies that $[a_{m'} - a]$ is also "small" for $m' = m + 1, m + 2, \ldots$.

The procedure is open to mathematical criticism; but, at the same time, the method has been demonstrated by application to be useful and profitable when it is properly handled. Deming (1943) has devoted most

of his book to the routine use of the method, even when the vector X is itself a random variable. Rao's (1952, pp. 165-72) method of "scoring," given as a general technique for solving maximum likelihood equations, is the same sort of procedure. Moore and Zeigler (1959, pp. 28-31) give several examples of functions that have yielded to careful administration of the method.

Well-known, but nevertheless worth noting in terms of our present notation, is the fact that the iterative method gives an exact result in a single iteration when applied to the linear model $Y = X\alpha + e$. The demonstration is straightforward. Any estimate of a, say a_0, is chosen. The matrix P_0 is independent of a_0 and we write $P_0 = X$. We have the vector $Z_0 = Y - Xa_0$. Then, the vector of corrections is $d_1 = (X'X)^{-1}X'Z_0 = (X'X)^{-1}X'(Y - Xa_0)$. The new estimate of the least squares estimate of α is $a_1 = a_0 + d_1 = a_0 + (X'X)^{-1}X'Y - (X'X)^{-1}X'Xa_0$ $= (X'X)^{-1}X'Y$. But it has already been noted that this is the estimate of the vector α obtained by direct application of least squares methods. Hence, $a_1 = a$, and the correspondence between the techniques for this model has been established.

CHAPTER III

SUFFICIENT CONDITIONS FOR MULTIPLE MINIMA OF THE SUM OF SQUARES SURFACE

It was pointed out in Chapter II that linear models, perhaps after some reparameterization, lead to least squares solutions that are unique. Hence, there is no concern about multiple minima and the possibility of choosing a local maximum or saddle point as a least squares estimate of the parameter vector α.

However, this is not the case for nonlinear models. Each proposed model presents a new problem for the statistician. Even after applying the Gauss iterative techniques, he must satisfy himself and his client that his solution does meet the requirements of being a global minimum. Showing uniqueness of the result is also a difficult problem. In general, a detailed examination of the sum of squares surface is required. Sometimes, it is possible to call on certain theorems about convex functions; but these are difficult to work with because the sum of squares is often only locally convex and defining the neighborhoods of convexity is a complicated procedure.

It is the purpose of this chapter to state and prove a theorem about a class of functions that will lead to several points in the parameter space at which the sum of squares is minimized. Following the theorem, some space will be devoted to an example and discussion

of a difficulty implied by the theorem.

Because the values of the mathematical variable X are of no direct concern in this theorem, the notation will be simplified by writing F_α and P_α for the vector $F(X; \alpha)$ and the matrix $P(\alpha)$, respectively, defined in Chapter II. The notation for the transpose of a vector or a matrix will be used where convenient; e.g., we will speak interchangeably of α and its transpose α'.

Theorem. Let the following conditions be satisfied:

1. The parameter vector α is divisible into $(k + 1)$ subvectors $\beta_1, \beta_2, \ldots, \beta_k, \beta_{k+1}$.
2. The length of the subvector $\beta_{i'}$ is the same as the length of the subvector $\beta_{i''}$; $i' \neq i''$; $i', i'' = 1, 2, \ldots, k$. The subvector β_{k+1} may contain no elements.
3. With this subdivision of α, the function to be fitted may be written

$$f(X; \alpha) = \sum_{i=1}^{k} g(X; \beta_i) + h(X; \beta_{k+1}).$$

4. The sum of squares is minimized at the point $a' = (b_1' \; b_2' \; \ldots \; b_k' \; b_{k+1}')$, where the length of b_i is the same as that of β_i.

Then there are at least k! points at which the sum of squares is minimized.

Proof. Consider the vector, say $\underline{a}$, formed by permuting the first k subvectors of the vector a. The new vector can be written

$$\underline{a}' = (b_{(1)}' \; b_{(2)}' \; \cdots \; b_{(k)}' \; b_{k+1}'),$$

where $(i) \neq i$ for at least one i in the set $\{1, 2, \ldots, k\}$. It will be shown that $\underline{a}$ is also a minimizing vector for the sum of squares.

By conditions 1, 2, and 3 and the convention adopted just prior to stating the theorem,

$$F_\alpha = \sum_{i=1}^{k} G_{\beta_i} + H_{\beta_{k+1}} .$$

Hence, $F_a = F_{\underline{a}}$.

Let R be the matrix of partial derivatives of the vector F_α with respect only to the elements of $\beta_{i'}$, and let S be similarly defined for the elements of $\beta_{i''}$. It follows that $R_{\beta_{i'}} = S_{\beta_{i'}}$. Therefore, if T is the matrix of partial derivatives of F_α with respect only to the elements of β_{k+1}, the matrix P can be partitioned so that

$$P = (R \quad R \quad \ldots \quad R \quad T)$$

where the matrix R is repeated k times.

Since condition 4 implies that the normal equations are satisfied at the point given by the vector a, it follows that $P_a'(Y - F_a) = \varphi$, where φ is a null vector. However, because of the partitioning of P, this single equation is equivalent to the following (k + 1) matrix equations:

$$R_{b_i}'(Y - F_a) = \varphi, \; i = 1, 2, \ldots, k, \qquad \text{(III.1)}$$

$$\text{and} \quad T_{b_{k+1}}'(Y - F_a) = \varphi.$$

Now consider the matrix products

$$R_{b_{(i)}}'(Y - F_a), \; (i) = 1, 2, \ldots, k, \qquad \text{(III.2)}$$

where the subvectors $b_{(i)}$ are selected from the vector $\underline{a}$. These matrix

products are, by definition, permutations of the set of products in Equations (III.1). Hence, it follows that the products (III.2) are null vectors. The vector $\underline{a}$ is therefore a solution to the normal equations. Since there are k! possible permutations of the minimizing vector a, there are at least k! points at which the sum of squares is minimized.

Motivation and inspiration for the theorem were derived from a paper by Cornfield, Steinfeld, and Greenhouse (1960). These men were concerned with the fitting of experimental data by the sum of exponentials model given in Chapter I. In particular, they were interested in estimating the parameters α_1, λ_1, α_2, and λ_2 in the model

$$y = \alpha_1 e^{-\lambda_1 t} + \alpha_2 e^{-\lambda_2 t} + e.$$

This is an example of the kind of models with which the theorem above is concerned. Thus, if $(\hat{\alpha}_1, \hat{\lambda}_1, \hat{\alpha}_2, \hat{\lambda}_2)$ is a point at which the sum of squares is a minimum, then the point $(\hat{\alpha}_2, \hat{\lambda}_2, \hat{\alpha}_1, \hat{\lambda}_1)$ is also a minimum.

The main concern of Cornfield and his fellow workers with respect to this result is the implication that there is a third stationary point at some intermediate position which does not give a minimum sum of squares. They show, for the model considered, how to find such a third stationary point. It seems to be their contention that the iterative least squares method may sometimes converge to this third stationary point and thus lead to estimates of the parameters which do not minimize the sum of squares.

To examine this conjecture, a set of data was prepared by the

author of this thesis for a double exponential model. The values of y were exact to at least seven significant figures. Following the instructions of Cornfield and his co-authors, the third stationary point was found. These values were then used as the starting values for the parameters in a routine calculation using the program described by Moore and Zeigler (1959). Contrary to the fears of Cornfield, et al, the iterative procedure (as carried out by the Moore-Zeigler program) experienced no difficulty in "sliding away" from the false point and converging to the true values of the parameters.

CHAPTER IV

A FORTRAN II PROGRAM FOR PERFORMING SAMPLING INVESTIGATIONS OF THE CHARACTERISTICS OF ESTIMATORS OBTAINED FROM GENERAL MODELS

Researchers in the field of Linear Hypotheses often have one great advantage over those who are concerned with nonlinear models: the estimators of the parameters in a linear model usually are themselves linear functions of the random variables. This being the case, it is possible to state precise theoretical results concerning such matters as the distributions of the parameters and tests of hypotheses about the model in question. Graybill (1961) summarizes a general theory that covers most models and problems of this sort that the statistician is willing to admit to his domain of interest.

However, this is not the case when one is dealing with nonlinear models. Nearly every model carries with it certain idiosyncracies and difficulties that may not occur in any other model. As a result, a considerable amount of applied statistical literature has been devoted to articles which present special techniques for particular models.

Examination of the properties of estimators obtained from iterative methods is hampered by the basic intractability of the form of the estimators. This is, of course, a consequence of the fact that

the estimators are the results of a sequence of approximations. Owing to the lack of suitable forms, the usual elegant techniques for linear estimators are not applicable.

To overcome these difficulties, one may resort to using "Monte Carlo" or distribution sampling methods. These methods are among the oldest of all statistical devices and were brought into prominence early in the 20th century by "Student's" work with the t distribution. It is unfortunately true that these techniques are generally cumbersome and tedious. However, the introduction of extremely high-speed computers was accompanied by their being heralded as devices which would remove the drudgery from large scale sampling investigations. Teichroew (1953) gives a discussion of this subject, along with a bibliography containing most of the pertinent literature available at the time of publication.

In this chapter we will discuss a computer program written expressly for the purpose of examining the properties of estimators of parameters in nonlinear models. The overall code name for the program is TEE. It was written by the author of this thesis under the auspices of the Los Alamos Scientific Laboratory, Los Alamos, New Mexico. Written in the FORTRAN II coding system, the program is designed to allow investigation of various characteristics of the parameters in nearly any desired model with almost any proposed distribution of the random variables in the model.

In principle, the program is completely general. This generality is achieved by dividing the program into several basic parts. By so doing, each part becomes a subroutine that can be modified easily as various conditions arise. For example, it sometimes may be convenient

to have the computer prepare the values of the mathematical variables X, while at other times the investigator may be interested in values of X that are most conveniently prepared externally. To allow for such situations, the array of mathematical variables is selected by a subroutine. Within the subroutine, the proper X-array is prepared or loaded according to a key number which is one of the subroutine's arguments.

Even the most basic part of the program, that in which the parameters are estimated by a particular method, is a subroutine. Because it was desired to write a program that would allow the comparison of estimating procedures, such a subroutine was an absolute necessity because of the impossibility of even beginning to code into the program all possible estimating procedures. In the present work, we have been interested in least squares estimators. Hence, the subroutine for their estimation is an adaptation of a basic least squares program, long in use at Los Alamos and elsewhere, which is described in detail by Moore and Zeigler (1959). Should another estimating procedure be proposed or modifications required on this one, it will be a simple matter to prepare a different subroutine or to make the modifications.

Because it is impossible to anticipate the desires of an investigator, the output of the program is also contained in a subroutine. Actually, this subroutine must be entered several times during the running of a problem. First, it must be entered to prepare for the storage of results that will be forthcoming. Next, it must be entered at the end of the estimation procedure in order to store these results. Finally, it must be entered to prepare and record for examination the properties of the estimators in which there is interest.

Details of the preparation and operation of this program are not included in this thesis. Still in its developmental stage, the program requires considerable reorganization and refinement before being made available for general usage. It is expected to be issued as a Los Alamos Scientific Laboratory Report when it is in its final form and operating writeups can be prepared.

CHAPTER V

CONCERNING THE SINGLE EXPONENTIAL MODEL AND THE SELECTION OF A SET OF PARAMETERS FOR EXAMINATION

The sum of exponentials model

$$y = \sum_{k=1}^{K/2} \alpha_k e^{\lambda_k x} + e \,, \qquad (V.1)$$

given in the earlier chapters as an example of a nonlinear model, is an extremely useful one. By varying the value of K and the signs of the α's and the λ's, it is possible to represent a great many types of curves. In various areas of investigation, the parameters have special and useful interpretation. For example, the λ's are the "decay constants" of the several isotopes in the material being measured in radioactive decay studies. Many functions of the parameters, such as the "mean life" and the "half-life" of an isotope or its initial abundance, are of interest to the experimenter. The model obtained from (V.1) by setting $K = 4$, $\lambda_1 = \lambda$, and $\lambda_2 = 0$ is sometimes called the Mitcherlisch Law. This model, which is written $y = \alpha_1 e^{\lambda x} + \alpha_2 + e$, is often used to fit data from fertilizer experiments. The parameter α_2 is the asymptotic yield for large amounts of fertilizer, x, while λ measures the rate of change in effect and α_1 is a "positioning" parameter.

In the remainder of this thesis, we shall be concerned with the simplest form of the exponential model; i.e., we take K = 2, drop the subscript k, and obtain the model

$$y = \alpha e^{\lambda x} + e, \qquad (V.2)$$

where e is independently distributed with mean zero and variance σ^2. If we assume that the initial estimates of α and λ are a_0 and c_0, respectively, and that the number of observations is I, we obtain the matrices Z_0, P_0, and δ_1 which are required by Equation (II.3). Thus,

$$Z_0 = \begin{bmatrix} y_1 - a_0 e^{c_0 x_1} \\ y_2 - a_0 e^{c_0 x_2} \\ \cdots \\ y_I - a_0 e^{c_0 x_I} \end{bmatrix}, \qquad P_0 = \begin{bmatrix} e^{c_0 x_1} & a_0 x_1 e^{c_0 x_1} \\ e^{c_0 x_2} & a_0 x_2 e^{c_0 x_2} \\ \cdots & \cdots \\ e^{c_0 x_I} & a_0 x_I e^{c_0 x_I} \end{bmatrix},$$

$$\text{and} \quad \delta_1 = \begin{bmatrix} \delta a_1 \\ \delta c_1 \end{bmatrix}.$$

We note that P_0 is of rank 2 if $a_0 \neq 0$ and $x_{i'} \neq x_{i''}$ for at least one pair i' and i". Thus, Assumption I.2 is satisfied. The estimates of δa_1 and δc_1, say Δa_1 and Δc_1, are obtained by solving the normal equations. This solution can be written in matrix form with the elements expressed as sums:

$$\begin{bmatrix} \Delta a_1 \\ \Delta c_1 \end{bmatrix} = \begin{bmatrix} \Sigma(e^{2c_0x_i}) & a_0\Sigma(x_i e^{2c_0x_i}) \\ a_0\Sigma(x_i e^{2c_0x_i}) & a_0^2\Sigma(x_i^2 e^{2c_0x_i}) \end{bmatrix}^{-1}$$

$$\times \begin{bmatrix} \Sigma(e^{c_0x_i})(y_i - a_0 e^{c_0x_i}) \\ a_0\Sigma(x_i e^{c_0x_i})(y_i - a_0 e^{c_0x_i}) \end{bmatrix} \qquad (V.3)$$

where the index of summation runs over the values i = 1, 2, ..., I.

Estimates of the parameters α and λ are obtained from these results as $a_1 = a_0 + \Delta a_1$ and $c_1 = c_0 + \Delta c_1$. These estimates take the place of a_0 and c_0 in Equation (V.3), and the solutions obtained in the second iteration are denoted Δa_2 and Δc_2. The process is continued until such a time, say the Mth iteration, that the solutions at that stage, say Δa_M and Δc_M, are "sufficiently small." The values a_M and c_M at that stage are taken to be the least squares estimates of α and λ and are denoted by a and c.

The variances of these estimates can be examined in a manner similar to that used in linear models. Deming (1943) and Rao (1952) have previously used this idea. The asymptotic covariance matrix for a and c is

$$\sigma^2 \left[(P_{\alpha,\lambda})'(P_{\alpha,\lambda})\right]^{-1} \qquad (V.4)$$

where $(P_{\alpha,\lambda})$ is the matrix P_0 evaluated at the true values of the

parameters. The variance of the errors, σ^2, is estimated by summing the squares of the deviations of the observations from the fitted function and dividing by (I - 2). Thus, the estimate of the variance is

$$s^2 = \frac{\Sigma(y_i - ae^{cx_i})^2}{I - 2} . \qquad (V.5)$$

The factor (I - 2) is used in an attempt to obtain an unbiased estimate of σ^2. Finally, then, to obtain estimates of the variances of the parameter estimates, we evaluate the matrix $(P_{a,\lambda})$ using a and c to give $(P_{a,c})$, substitute this result and that of Equation (V.5) into Equation (V.4), and find that an estimate of the covariance matrix is

$$s^2 \left[(P_{a,c})'(P_{a,c})\right]^{-1} . \qquad (V.6)$$

An experimenter interested in the single exponential model should usually have some idea of the values of the parameters α and λ and also of the domain of values that the independent variable x is likely to assume. Thus, for the most part, it will be known whether the decay constant λ will be positive or negative and that each x_i will lie in some interval, say (X_0, X_1). Moreover, knowing approximately the value of λ, it is possible to have some idea of the change that will occur in the response y as x varies between X_0 and X_1. Furthermore, it is reasonable to be concerned only with positive values of α.

In order to discuss several sets of parameters on a common basis, we note that the points x belonging to the interval (X_0, X_1) map into the points t belonging to the interval (0, 1) by means of the

transformation $t = (x - X_0)/(X_1 - X_0)$. We use this relationship to change the model $y = \alpha e^{\lambda x} + e$ for x in the interval (X_0, X_1) to the model

$$y = \beta e^{\delta t} + e \qquad (V.7)$$

for t belonging to the interval (0, 1). It can be shown that the relationships among the parameters can be expressed as

$$\lambda = \delta/(X_1 - X_0)$$

$$\text{and} \quad \alpha = \beta e^{-\lambda X_0} \quad .$$

Thus, from estimates of β and δ, which we henceforth call b and d, we can obtain estimates of α and λ. In particular, we note that the estimates of λ and δ are linear functions of each other. We may state, therefore, such expressions as $E(d) = (X_1 - X_0)E(c)$, where the operator $E(\ )$ is the usual expectation. When $X_0 = 0$, α and β are identical.

Suppose now we consider the ratio, say R, of the value of the model at X_1 to its value at X_0; i.e.,

$$R = \frac{\alpha e^{\lambda X_1}}{\alpha e^{\lambda X_0}} = e^{\lambda(X_1 - X_0)} \quad .$$

Solving this expression for λ, we obtain

$$\lambda = \frac{\log R}{(X_1 - X_0)} \quad .$$

Thus, if we wish to simulate a model in which the ratio of the function at the endpoints of the domain of x is known, we may find the value of

λ which corresponds to that ratio. Because we know the parameter δ in terms of λ, we can simulate the same effect in the model (V.7) by taking

$$\delta = \log R.$$

Based on these considerations and arbitrarily choosing five values of β, the following 25 combinations of β and δ were selected for detailed examination: $\beta = 10, 50, 100, 500, 1000$; and $\delta = \log (1/4), \log (1/2), \log 1 = 0, \log 2, \log 4$. Using these values of δ, we examined situations in which the model evaluated at $t = 1$ ranges from one-fourth to four times its value at $t = 0$.

With respect to the choice of the values of δ, it is somewhat unrealistic to use an exponential model which affords the chance that some of the observations y will be negative. Because we wished to assume that the errors in the observations are normally distributed with mean zero and unit variance, it was necessary to select models so that only a few negative values would occur. The lowest value attained in the set of 25 models is 2.5. Thus, the probability of generating a negative observation for that value is about 0.0062, and the chance of a disturbance due to negative numbers is small.

It was decided to limit the study of these populations to the case in which 10 values of t are uniformly distributed on the interval $(0, 1)$. We therefore chose $t_i = (i - 1)/9$ for $i = 1, 2, \ldots, 10$. Twenty-five exact functions were generated, one for each pair of the chosen β and δ. A total of 500 samples for each function were then generated by adding a random normal deviate from $N(0, 1)$ to each point of each (β, δ) pair.

The random normal deviates were generated by the method given by Box and Muller (1958):

> Let U_1 and U_2 be independent random variables from the same rectangular density function on the interval (0, 1). Consider the random variables:
>
> $$W_1 = (-2 \log_e U_1)^{1/2} \cos 2\pi U_2$$
>
> $$W_2 = (-2 \log_e U_1)^{1/2} \sin 2\pi U_2 .$$
>
> Then W_1 and W_2 will be a pair of independent random variables from the same normal distribution with zero mean and unit variance.

Box and Muller indicate that this method gives higher accuracy than other standard methods and compares favorably with the other methods in the matter of computing speed. The generation of the uniform random variables required for this study was accomplished by the use of the "residue-class" method which seems to have first been published by Lehmer (1949). A subroutine incorporating these two methods was used in the TEE computer program discussed in Chapter IV.

The entire program was operated on an IBM 7090 with magnetic tape input and output and under a supervising monitor system. Less than 110 seconds were required to generate the 500 samples and to record the sampling characteristics of the least squares estimates of each pair of β and δ. Discussion of the results of the experiment is divided into three parts: (1) some general observations, (2) characteristics of the estimate of β, and (3) characteristics of the estimate of δ.

CHAPTER VI

SOME GENERAL OBSERVATIONS ABOUT DISTRIBUTION SAMPLING FOR THE MODEL $y = \beta e^{\delta t} + e$

The problem of determining at what stage convergence of the iterative least squares procedure can be said to have occured was purposely left vague in Chapter II. Any number of possible criteria are available. Returning to the notation of Chapter II in which α is a vector of parameters and a_M is the estimate of α obtained at the Mth iteration, we might use the criterion that we will be satisfied when $|Q(a_{M-1}) - Q(a_M)| < u$ where $Q(a_{M-1})$ and $Q(a_M)$ are the sums of squares for the (M - 1)th and the Mth iterations and u is an arbitrarily "small" scalar. However, this is difficult to provide for in a computer because of the problem of trying to anticipate a reasonable choice for the constant u and because the difference between two consecutive sums of squares may be obscured by the size of the observations.

Experience at the Los Alamos Scientific Laboratory has led to the following criterion for convergence: Let $|a_{M-1}|$ and $|a_{M-1} - a_M|$ denote the vectors of the absolute values of the elements of the vector a_{M-1} and the difference vector $(a_{M-1} - a_M)$. Let v be an arbitrarily small scalar that is chosen in advance. The condition $|a_{M-1} - a_M| < v|a_{M-1}|$ means that every element of $|a_{M-1} - a_M|$ is strictly less than its

corresponding element in $v|a_{M-1}|$. Then, if this condition is satisfied, we take a_M as the least squares estimate of the vector α. Normally, $v = 10^{-6}$ has proved satisfactory for this test for a great many functions and a great many sets of observations. However, in the course of the present study, it was found that this convergence criterion could not be satisfied for some samples when $\delta = 0$. A detailed examination showed that the reason for failure in these cases was the oscillatory nature of the sequence of estimates of the parameter δ. To remedy this condition, the criterion was reduced to $v = 10^{-3}$ for the samples in which $\delta = 0$. Because the entire program was coded in single-precision, it is felt that increased accuracy in calculation--by either increasing word size or using multiple word precision--would result in satisfactory convergence. In Chapters VII and VIII, we shall see that this relaxed criterion seems to have had little or no affect on the bias and the variances of the estimates.

As a practical point, it is well to establish a limit on the number of iterations that will be allowed for any one set of observations. Again by experience, it has been fairly well determined that, unless convergence has been attained within 20 or 25 iterations, some additional analysis is in order. Sometimes this consists merely of checking the input data for transcription errors. Of more serious consequence is a basic programming error or the failure of the data to even remotely resemble the function being fitted.

It has been noted by many writers, among them Levenberg (1944) and Hartley (1961), that the occurence of large oscillations has been detrimental to the increased usage of the Gauss iterative scheme. Both of these authors have proposed remedies. But, as Hartley points out in

his paper, Levenberg's idea does not lend itself well to computer methods. Hartley's own modification, justifiable mathematically under some fairly general conditions, is subject somewhat to the same objection. However, this can be overcome in a variety of ways, and the Hartley proposal was considered by this author as a possibly desirable feature. A pilot study, using 100 samples from the single exponential model with $\beta = 10$ and $\delta = \log (1/4)$, was conducted to contrast Hartley's procedure with the basic Gauss method. In terms of the programs developed at Los Alamos, the running time for Hartley's modification was about twice that of the unmodified method. There was no change in the results, and very little reduction in the number of iterations required for convergence was exhibited. The same convergence criterion was used in both trials.

Other methods of oscillation control are mentioned by Hartley (1961). Still another, adopted originally as an expedient and since used routinely, has been suggested by Moore and Zeigler (1959, p. 43). This is an arbitrary technique for forcing each element of the estimating vectors to have the same sign as that assigned to the corresponding element of the vector of initial estimates a_0. Although successful from a practical standpoint, the method still lacks the statement and proof of the conditions under which it may be expected to be mathematically precise.

In the light of these observations, it was decided to carry out the sampling experiment with no attempt to control the oscillations. As it turned out for these models, every one of the 12,500 samples in this study converged to the required criterion. Since we were not concerned with methods of obtaining the initial estimates of the

parameters, the values used to prime the procedure were the true values of the parameters.

The number of iterations required for convergence for each sample was recorded, and a histogram of these values was formed for each of the 25 models. Based on these histograms, estimates of the average and of the range of the number of iterations required were computed. These results are shown in Table I. It is to be remembered that convergence is not possible until at least two iterations have occured and that a maximum of 20 iterations were permitted. In addition, the values in the table probably should be interpreted as lower bounds for "real data" since the initial values of the estimates were taken as the exact values of the true parameters.

Table I shows that: (1) for a given β, the average number of iterations essentially decreases as δ increases; (2) the average number of iterations is a decreasing function of β for fixed values of δ; and (3) the variability of the number of iterations decreases as β and/or δ increase. Thus, we see that satisfactory and rapid convergence of the iterative least squares method depends to a large extent both on the true function and upon the convergence criterion. No information was obtained for situations in which the initializing estimate of the parameter vector was not the value used in the model formation. However, graphical or linearization methods have proved quite satisfactory in practical problems because they are generally fairly efficient when dealing with the single exponential model. In an automatic computer it seldom matters if several iterations are needed for convergence.

The results of estimating the error variance, σ^2, for each group

TABLE I

THE NUMBER OF ITERATIONS REQUIRED FOR CONVERGENCE FOR VARIOUS MODELS OF THE FORM $y = \beta e^{\delta t} + e$

True Parameters		Number of Iterations Required		
β	δ	Mean	Minimum	Maximum
10	log (1/4)	6.09	4	10
	log (1/2)	5.80	4	8
	log 1 = 0*	4.20*	2	5
	log 2	5.10	3	7
	log 4	4.65	3	6
50	log (1/4)	4.51	3	6
	log (1/2)	4.47	3	5
	log 1 = 0*	3.78*	2	4
	log 2	4.08	3	5
	log 4	3.94	3	4
100	log (1/4)	4.10	3	5
	log (1/2)	4.03	3	5
	log 1 = 0*	3.61*	2	4
	log 2	3.93	3	5
	log 4	3.83	3	4
500	log (1/4)	3.60	3	4
	log (1/2)	3.66	3	4
	log 1 = 0*	2.55*	2	4
	log 2	3.30	3	4
	log 4	3.05	3	4
1000	log (1/4)	3.23	3	4
	log (1/2)	3.24	3	4
	log 1 = 0*	2.09*	2	4
	log 2	3.03	3	4
	log 4	3.00	3	3

*The convergence criterion was $v = 10^{-3}$. For unmarked models, the criterion was $v = 10^{-6}$.

of 500 samples were recorded. The estimate, s^2, was obtained from Equation (V.5) with I = 10. Had the model been linear in two parameters, we would have been able to state that s^2 is independent of the estimators b and d and that the quantity $z = (I - 2)s^2/\sigma^2$ is distributed as a chi-square variable with $(I - 2) = 8$ degrees of freedom. An idea of whether this statement is approximately correct for this nonlinear model can be obtained by comparing the sampling distribution of a variable $w = s^2/\sigma^2$ with a "chi-square-over-degrees-of-freedom" distribution with 8 degrees of freedom. Because the tails of the sampling distribution are generally of interest, the endpoints of the intervals into which the values of w were to be sorted were chosen so that the smaller probabilities in the tails could be examined. It was convenient to use 16 such intervals, with the one at either end chosen so that five out of 500 would be the expected number of occurences. If s^2 were an unbiased estimator of σ^2, we would expect the average value of w to be unity. A chi-square goodness-of-fit test was run for each of the 25 sampling distributions. A summary of these results is given in Table II.

The most important fact to be gleaned from Table II is that the average value, $\bar{w}$, of $w = s^2/\sigma^2$ is near unity. This indicates a small bias when the sample size is 10 and the independent variable is equally spaced over the interval (0, 1). The next point is that the distribution of w is very close to that of 1/8 times a chi-square with 8 degrees of freedom. This is indicated by noting that none of the goodness-of-fit chi-squares exceeds the upper 5 per cent critical value and that their average is 12.36 which is not far from the expected value of 15 which comes from the use of fifteen intervals.

TABLE II

SUMMARY OF THE DISTRIBUTION OF $w = s^2/\sigma^2$ WHERE $\sigma^2 = 1$

True Parameters			Goodness-of-Fit	Number Below Lower	Number Below Upper
β	δ	$\bar{w}$	Chi-Square*	1% Point	1% Point
10	log (1/4)	1.0021	14.25	8	4
	log (1/2)	1.0164	12.43	3	5
	log 1 = 0	0.9685	8.01	3	4
	log 2	1.0547	22.29	7	11
	log 4	0.9732	19.31	8	1
50	log (1/4)	0.9662	8.93	8	5
	log (1/2)	1.0088	11.73	5	4
	log 1 = 0	0.9774	21.35	12	7
	log 2	1.0265	9.31	5	4
	log 4	0.9907	8.29	7	2
100	log (1/4)	0.9884	7.65	2	5
	log (1/2)	0.9919	10.47	7	5
	log 1 = 0	1.0154	6.27	5	8
	log 2	1.0390	14.45	5	4
	log 4	1.0060	8.07	6	8
500	log (1/4)	0.9776	16.95	7	3
	log (1/2)	0.9722	10.93	5	2
	log 1 = 0	1.0024	6.93	6	5
	log 2	1.0309	16.93	4	5
	log 4	0.9689	22.59	3	7
1000	log (1/4)	1.0352	17.75	5	6
	log (1/2)	1.0043	6.65	6	3
	log 1 = 0	1.0363	10.53	3	4
	log 2	0.9923	5.69	6	5
	log 4	0.9915	11.23	5	8
Averages Over Entire Sampling Experiment		1.0014	12.36	5.6	5

*Based on 16 intervals of unequal probability. Upper 5 per cent critical value for 15 degrees of freedom is 25.0.

Finally, it is seen that, on the average, the tails of the sample distributions agree well with the hypothetical chi-square-over-degrees-of-freedom distribution.

CHAPTER VII

CONCERNING CHARACTERISTICS OF THE ESTIMATOR b OF THE PARAMETER β IN THE MODEL $y = \beta e^{\delta t} + e$

The asymptotic variance of the estimator of β is defined as the value obtained by selecting the first diagonal element of the matrix (V.4), where β, δ, and t are used in place of α, λ, and x. Thus, it is the quantity in the first row and the first column of the matrix

$$\sigma^2 \begin{bmatrix} \Sigma(e^{2\delta t_i}) & \beta\Sigma(t_i e^{2\delta t_i}) \\ \beta\Sigma(t_i e^{2\delta t_i}) & \beta^2\Sigma(t_i^2 e^{2\delta t_i}) \end{bmatrix}^{-1} \qquad \text{(VII.1)}$$

where the index of summation runs over the values i = 1, 2, ..., I. We shall denote the value so obtained by σ_b^2.

Two other estimates of the variance of b were obtained from the sampling experiment. The first of these is computed from each of the 500 estimates of β. In general, if N estimates of β are available for a given pair of parameters, we compute

$$S_b^2 = \frac{\sum_{n=1}^{N} (b_n - \bar{b})^2}{N}$$

where b_n is the nth estimate of β and $\bar{b} = (\sum_{n=1}^{N} b_n)/N$. The other estimate of the variance of b is obtained by putting the iterated estimates b and d into the matrix (VII.1) upon the completion of each fit and averaging the results to get

$$\overline{s_b^2} = \sum_{n=1}^{N} (s_{b_n}^2)/N$$

where $s_{b_n}^2$ is the nth estimate yielded by the procedure.

In order to examine the distribution of b, it was assumed that the estimates would be normally distributed with mean β and variance σ_b^2, where the latter term has been defined as the asymptotic variance of b. Although this is true for infinite sample sizes, one of the objects of the investigation was to examine the question of whether a sample of size 10 is large enough for equally spaced values of the independent variable to allow the use of asymptotic results. Following the suggestions of Cochran (1952), the values of b were sorted into 20 equally probable cells whose endpoints were obtained under this assumption of normality. Thus, (1/20) x 500 = 25 of the b_n were expected in each cell. The quantity $\bar{b}$ was taken as the estimate of the expected value of b, and the bias was taken as the difference $(\bar{b} - \beta)$. To examine the magnitude of the bias, its absolute value was compared with $\sigma_b/\sqrt{500}$ which is approximately the standard deviation of $\bar{b}$. The pertinent results are given in Table III. Examination of the matrix (VII.1) shows that σ_b^2 is not a function of β but is a function of δ. This can be seen in Table III under the column headed by $\sigma_b/\sqrt{500}$.

We note from Table III that there is no apparent bias or trend of

TABLE III

THE ESTIMATOR b OF THE PARAMETER β

True Parameters β	δ	$\bar{b}$	Bias (b − β)	$\sigma_b/\sqrt{500}$	Goodness-of-Fit Chi-Square*
10	log (1/4)	9.98	-0.02	0.033	19.84
	log (1/2)	10.04	0.04	0.030	21.84
	log 1 = 0	10.02	0.02	0.026	28.48
	log 2	9.96	-0.04	0.022	18.08
	log 4	10.01	0.01	0.018	14.40
50	log (1/4)	50.03	0.03	0.033	18.48
	log (1/2)	50.00-	-0.00	0.030	19.20
	log 1 = 0	49.99	-0.01	0.026	10.72
	log 2	50.01	0.01	0.022	14.16
	log 4	50.01	0.01	0.018	22.00
100	log (1/4)	99.95	-0.05	0.033	28.56
	log (1/2)	100.03	0.03	0.030	25.52
	log 1 = 0	99.99	-0.01	0.026	19.28
	log 2	100.02	0.02	0.022	12.48
	log 4	99.98	-0.02	0.018	10.72
500	log (1/4)	500.06	0.06	0.033	18.40
	log (1/2)	499.99	-0.01	0.030	17.12
	log 1 = 0	500.00+	0.00	0.026	37.68**
	log 2	500.00-	-0.00	0.022	20.48
	log 4	500.02	0.02	0.018	19.52
1000	log (1/4)	1000.02	0.02	0.033	21.12
	log (1/2)	1000.02	0.02	0.030	14.80
	log 1 = 0	1000.03	0.03	0.026	20.72
	log 2	1000.01	0.01	0.022	12.96
	log 4	1000.00-	-0.00	0.018	17.36

*Based on 20 equally probable intervals. Upper 5 per cent critical value for 19 degrees of freedom is 30.1.

**Significant at 5 per cent level.

bias for the various values of β and δ. In every case, the absolute value of the bias is less than twice the quantity $\sigma_b/\sqrt{500}$, which implies that $\bar{b}$ is not far from β. Of the 25 results, 10 show negative bias and 15 show positive. This, too, is reasonable under the assumption that $\bar{b}$ is an unbiased estimate of β. As for the chi-square goodness-of-fit results, only one is significantly larger than the 5 per cent critical value. In general, these sampling results indicate that bias, if any, is small and that the distribution of the estimator is nearly normal.

The square roots of σ_b^2, S_b^2, and $\overline{s_b^2}$ are given for each parameter set in Table IV. If the asymptotic standard deviation, σ_b, is used as a standard, we see from Table IV that just about the same numbers of S_b and $(\overline{s_b^2})^{1/2}$ are above that standard as are below it. This indicates that all three quantities are essentially the same. For practical purposes, then, it seems that the procedure of estimating the variance of b by substituting the estimates of the parameters into the formula for the asymptotic variance is satisfactory for the models considered herein.

If we handle the question of testing a hypothesis about the parameter β, say H_0: $\beta = \beta_0$, in the same manner as when we have a linear regression situation, we arrive at a test statistic

$$t' = (b - \beta_0)/s_b$$

where b is the iterated estimate of β and s_b is obtained by substituting the estimates b and d into the asymptotic variance formula. The 500 values of t' were tabulated for each parameter set (using the parametric value of β for β_0) as though t' were distributed as

TABLE IV

SQUARE ROOTS OF THE ESTIMATES OF THE VARIANCE OF

THE ESTIMATOR b OF THE PARAMETER β

True Parameters β	δ	σ_b	S_b	$\sqrt{\bar{s}_b^2}$
10	log (1/4)	0.740	0.755	0.741
	log (1/2)	0.669	0.674	0.674
	log 1 = 0	0.588	0.580	0.579
	log 2	0.498	0.499	0.511
	log 4	0.402	0.382	0.397
50	log (1/4)	0.740	0.752	0.727
	log (1/2)	0.669	0.706	0.698
	log 1 = 0	0.588	0.587	0.581
	log 2	0.498	0.490	0.505
	log 4	0.402	0.416	0.400
100	log (1/4)	0.740	0.711	0.735
	log (1/2)	0.669	0.635	0.666
	log 1 = 0	0.588	0.584	0.592
	log 2	0.498	0.514	0.507
	log 4	0.402	0.401	0.403
500	log (1/4)	0.740	0.706	0.732
	log (1/2)	0.669	0.639	0.659
	log 1 = 0	0.588	0.632	0.588
	log 2	0.498	0.503	0.505
	log 4	0.402	0.416	0.396
1000	log (1/4)	0.740	0.717	0.752
	log (1/2)	0.669	0.638	0.670
	log 1 = 0	0.588	0.592	0.598
	log 2	0.498	0.474	0.496
	log 4	0.402	0.407	0.401

Student's t with 8 degrees of freedom. Sixteen cells of varying probabilities were used. A chi-square goodness-of-fit value was obtained for each sampling distribution to give some indication of how t' differed from t. The cells at each end of the distribution were expected to contain five values (one per cent of the number of samples). The results relating to the parameter β are given in Table V. We see that there is very little, if any, evidence for rejecting the conjecture that t' has nearly a Student's t distribution with 8 degrees of freedom. The average numbers above and below the one per cent points are nearly five in each case. Thus, for forming tests of significance and confidence intervals, it appears that the tabulated values of Student's t can be used.

TABLE V

DISTRIBUTION OF t' FOR THE ESTIMATOR b OF THE PARAMETER β

True Parameters β	δ	Goodness-of-Fit Chi-Square*	Number Below Lower 1% Point	Number Above Upper 1% Point
10	log (1/4)	11.07	8	2
	log (1/2)	15.61	3	5
	log 1 = 0	19.26	5	6
	log 2	18.49	10	2
	log 4	11.97	8	4
50	log (1/4)	14.15	3	6
	log (1/2)	17.45	6	4
	log 1 = 0	8.81	5	8
	log 2	6.33	5	5
	log 4	15.45	3	8
100	log (1/4)	17.89	6	3
	log (1/2)	20.03	4	2
	log 1 = 0	23.91	5	3
	log 2	14.45	5	4
	log 4	11.44	5	4
500	log (1/4)	11.73	3	2
	log (1/2)	10.87	6	5
	log 1 = 0	20.27	3	7
	log 2	14.69	3	9
	log 4	17.35	3	4
1000	log (1/4)	14.06	3	5
	log (1/2)	6.71	4	5
	log 1 = 0	11.29	7	4
	log 2	17.21	3	2
	log 4	9.33	6	9
Averages Over Entire Sampling Experiment		14.39	4.9	4.7

*Based on 16 intervals of unequal probability. Upper 5 per cent critical value for 15 degrees of freedom is 25.0.

CHAPTER VIII

CONCERNING CHARACTERISTICS OF THE ESTIMATOR d OF THE PARAMETER δ IN THE MODEL $y = \beta e^{\delta t} + e$

The decimal representations of the natural logarithms for δ are presented in Table VI. In the computer generation of the models, all eight significant digits were used in floating point form.

TABLE VI

NATURAL LOGARITHMIC AND DECIMAL EQUIVALENTS FOR VALUES OF δ

Natural Logarithmic Representation	Decimal Representation
log (1/4)	-1.3862944
log (1/2)	-0.69314717
log 1	0
log 2	+0.69314717
log 4	+1.3862944

The asymptotic variance of the estimator of δ, denoted by σ_d^2, is that value obtained by selecting the second diagonal element of the matrix (VII.1). Two other estimates of the variance of d are defined analogously to the two given for the estimator b. The first, which is

the sampling variance of d, is denoted by s_d^2; the second, which is the mean of several estimates of the variance of d, is denoted by $\overline{s_d^2}$.

The individual estimates of δ were sorted into 20 cells of equal probability, the endpoints of which were selected by assuming that the estimator d is normally distributed with mean δ and variance σ_d^2. The quantity $\bar{d}$, the average of the estimates of δ, was taken as an estimate of the expected value of d, and the bias was estimated by the difference $(\bar{d} - \delta)$. Pertinent results are given in Table VII. Examination of the matrix (VII.1) shows that σ_d^2 is a function of both β and δ; for a fixed value of δ, σ_d^2 is inversely proportional to the square of β. This behavior may be seen in Table VII under the column headed by $\sigma_d/\sqrt{500}$.

Perhaps the most striking result shown in Table VII is the propensity toward d being negatively biased as an estimator of δ. Eighteen of the 25 results show this negative bias. The absolute values of the bias, nevertheless, do not exceed twice $\sigma_d/\sqrt{500}$, and we are thus led to feel that the biases are probably small. Only one of the goodness-of-fit chi-square values is significant at the 5 per cent level. Hence, we conclude that there is little evidence against the estimator d being nearly normally distributed with mean δ and variance σ_d^2.

Table VIII contains the square roots of the asymptotic variance, σ_d^2, and the other two estimates, s_d^2 and $\overline{s_d^2}$. The table shows that any of the three seems to serve equally well, thus justifying the use of the asymptotic variance formula for samples of size 10 and equally spaced values of the independent variable.

As in the case of the estimator of β, we define a psuedo-t

TABLE VII

THE ESTIMATOR d OF THE PARAMETER δ

True Parameters			Bias		Goodness-of-Fit
β	δ	$\bar{d}$	$(\bar{d} - \delta)$	$\sigma_d/\sqrt{500}$	Chi-Square*
10	-1.38629	-1.38641	-0.00012	0.00912	19.44
	-0.69315	-0.69415	-0.00100	0.00633	11.44
	0	-0.00399	-0.00399	0.00443	12.24
	0.69315	0.69859	0.00544	0.00316	24.16
	1.38629	1.38555	-0.00074	0.00228	20.88
50	-1.38629	-1.38689	-0.00060	0.00182	21.68
	-0.69315	-0.69345	-0.00030	0.00127	9.68
	0	0.00061	0.00061	0.00089	15.52
	0.69315	0.69252	-0.00063	0.00063	22.00
	1.38629	1.38604	-0.00025	0.00046	15.68
100	-1.38629	-1.38645	-0.00016	0.00091	20.64
	-0.69315	-0.69279	0.00036	0.00063	27.68
	0	0.00019	0.00019	0.00044	11.60
	0.69315	0.69279	-0.00036	0.00032	11.28
	1.38629	1.38651	-0.00022	0.00023	21.28
500	-1.38629	-1.38660	-0.00031	0.00018	12.96
	-0.69315	-0.69314	0.00001	0.00013	16.96
	0	-0.00001	-0.00001	0.00009	25.28
	0.69315	0.69320	0.00005	0.00006	14.72
	1.38629	1.38622	-0.00007	0.00005	30.80**
1000	-1.38629	-1.38645	-0.00016	0.00009	15.20
	-1.69315	-0.69318	-0.00003	0.00006	19.04
	0	-0.00007	-0.00007	0.00004	23.76
	0.69315	0.69316	0.00001	0.00003	13.52
	1.38629	1.38629-	-0.00000	0.00002	22.72

*Based on 20 equally probable intervals. Upper 5 per cent critical value for 19 degrees of freedom is 30.1.

**Significant at 5 per cent level.

TABLE VIII

SQUARE ROOTS OF THE ESTIMATES OF THE VARIANCE OF

THE ESTIMATOR d OF THE PARAMETER δ

True Parameters

β	δ	σ_d	S_d	$\sqrt{s_d^2}$
10	-1.38629	0.2040	0.2217	0.2071
	-0.69315	0.1415	0.1443	0.1439
	0	0.0991	0.0981	0.0978
	0.69315	0.0708	0.0695	0.0728
	1.38629	0.0510	0.0480	0.0503
50	-1.38629	0.0408	0.0429	0.0401
	-0.69315	0.0283	0.0299	0.0284
	0	0.0198	0.0196	0.0196
	0.69315	0.0142	0.0146	0.0143
	1.38629	0.0102	0.0104	0.0101
100	-1.38629	0.0204	0.0200	0.0203
	-0.69315	0.0142	0.0125	0.0141
	0	0.0099	0.0096	0.0101
	0.69315	0.0071	0.0073	0.0072
	1.38629	0.0051	0.0051	0.0051
500	-1.38629	0.0041	0.0039	0.0040
	-0.69315	0.0028	0.0028	0.0028
	0	0.0020	0.0021	0.0020
	0.69315	0.0014	0.0014	0.0014
	1.38629	0.0010	0.0010	0.0010
1000	-1.38629	0.0020	0.0019	0.0021
	-0.69315	0.0014	0.0014	0.0014
	0	0.0010	0.0010	0.0010
	0.69315	0.0007	0.0007	0.0007
	1.38629	0.0005	0.0005	0.0005

statistic for testing the hypothesis that the exponential parameter is equal to some specified value; i.e., H_0: $\delta = \delta_0$. Thus, for each trial of the 25 models under investigation, we computed the quantity

$$t' = (d - \delta)/s_d$$

and tabulated the results into 16 cells of unequal probability as though t' were distributed as Student's t with 8 degrees of freedom. Table IX contains the pertinent results of the tabulation. The appearance of three goodness-of-fit chi-squares that are significant at the 5 per cent level is perhaps disconcerting, but it seems that t' has nearly the distribution of a t variable. However, probability statements involving such an assumption should be guarded. The tendency, at least in the tails of the sampling distribution, seems to be in the direction of too many occurences both above and below a specified percentage point. Thus, if an experimenter were to form a symmetric 98 per cent confidence interval under the assumption that t' has a t distribution, he would actually be using a slightly smaller confidence level.

TABLE IX

DISTRIBUTION OF t' FOR THE ESTIMATOR d OF THE PARAMETER δ

True Parameters β	δ	Goodness-of-Fit Chi-Square*	Number Below Lower 1% Point	Number Above Upper 1% Point
10	-1.38629	29.39**	5	12
	-0.69315	11.11	4	3
	0	12.43	6	6
	0.69315	15.49	4	6
	1.38629	19.11	7	5
50	-1.38629	17.51	10	6
	-0.69315	9.31	9	5
	0	12.62	7	4
	0.69315	21.37	8	6
	1.38629	9.71	7	7
100	-1.38629	11.17	5	6
	-0.69315	25.37**	3	2
	0	17.65	1	3
	0.69315	12.49	4	7
	1.38629	9.96	5	4
500	-1.38629	12.27	4	1
	-0.69315	7.45	4	5
	0	37.83**	12	8
	0.69315	17.79	4	6
	1.38629	19.32	5	3
1000	-1.38629	21.67	6	4
	-0.69315	9.36	6	2
	0	20.81	2	8
	0.69315	13.29	1	5
	1.38629	17.05	9	3
Averages Over Entire Sampling Experiment		16.46	5.5	5.1

*Based on 16 intervals of unequal probability. Upper 5 per cent critical value for 15 degrees of freedom is 25.0.

**Significant at 5 per cent level.

CHAPTER IX

SUMMARY AND EXTENSIONS

In this thesis we have examined the general problem of estimating parameters in nonlinear models by the method of least squares. Because of the many different types of models, it seems impossible to give general rules of procedure.

The well-known Gauss iterative method is advocated because its value has been demonstrated in a variety of situations and because of the existence of several worthwhile modifications to the basic technique. However, general approval of the method does not lift from the statistician's shoulders the responsibility for assuring himself and his client that the results are satisfactory. In particular, a class of functions is exhibited which guarantees the existence of several points in the parameter space at which the minimum sum of squares is attained.

In order to examine the small-sample characteristics of any estimation procedure for nonlinear functions, a large-scale FORTRAN II computer program has been prepared to base such an examination on Monte Carlo methods. The single exponential model, $y = \beta e^{\delta t} + e$ where t lies in the interval $(0, 1)$ and e is independently distributed as a normal distribution with zero mean and unit variance, is taken as an example. Over a wide range of β and δ, the model is examined and the

results discussed. In general, it appears that the least squares estimators of β and δ are very nearly normally distributed and that biases for each are small. The usual t statistic for forming confidence intervals and tests of significance seems to be satisfactory for both parameters.

A great deal of work remains to be done with respect to estimation in nonlinear models. Each model presents its own special problems. The Monte Carlo attack is feasible with super-speed computers, and the program discussed in this thesis seems to have a great deal of application when used on computers which are at least as fast as the IBM 7090.

Assurance of convergence in any iterative scheme is important. Hartley (1961) has done some work in this direction, but it seems certain that some models might lead to less restrictive conditions than he has used for general functions. Effort might be profitably expended in the search for efficient methods of obtaining initial estimates of the parameters.

Until exact distributions are available for the estimators and for tests of hypotheses about the parameters, we will have to be satisfied with what might be termed "good enough" methods. That is, we will need some useful limits on sample sizes, numbers of iterations, approximate distributions, and similar matters. It is likely that solutions to these problems will be required for each type of nonlinear model.

A SELECTED BIBLIOGRAPHY

Aitkin, A. E. (1934-5). On least squares and linear combinations of observations. Proceedings of the Royal Society of Edinburgh LV, 42-8.

Box, G. E. P. and Muller, M. E. (1958). A note on the generation of random normal deviates. Annals of Mathematical Statistics 29, 610-1.

Cochran, W. G. (1952). The χ^2 Test of Goodness of Fit. Annals of Mathematical Statistics 23, 315-45.

Cornell, R. G. (1956). A New Estimation Procedure for Linear Combinations of Exponentials. ORNL-2120, Office of Technical Services, U. S. Department of Commerce, Washington 25, D. C.

Cornfield, J., Steinfeld, J., and Greenhouse, S. W. (1960). Models for the interpretation of experiments using tracer compounds. Biometrics 16, 212-34.

Cramér, H. (1951). Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

Deming, W. E. (1943). Statistical Adjustment of Data. John Wiley and Sons, Inc., New York.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. Philosophical Transactions of the Royal Society A222, 309-68.

Fisher, R. A. (1925). Theory of statistical estimation. Proceedings of the Cambridge Philosophical Society 22, 700-25.

Graybill, F. A. (1961). An Introduction to Linear Statistical Models, Volume I. McGraw-Hill Book Company, Inc., New York.

Hartley, H. O. (1961). The modified Gauss-Newton method for the fitting of non-linear regression functions by least squares. Technometrics 3, 269-80.

Hildebrand, F. B. (1956). Introduction to Numerical Analysis. McGraw-Hill Book Company, Inc., New York.

James, G. and James, R. C., Eds. (1959). Mathematics Dictionary. D. Van Nostrand Company, Inc., Princeton, New Jersey.

Lehmer, D. H. (1949). Mathematical methods in large-scale computing units. Proceedings of a Second Symposium on Large-Scale Digital Calculating Machinery. Harvard University Press, Cambridge, Massachusetts.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. Quarterly of Applied Mathematics 2, 164-8.

Moore, R. H. and Zeigler, R. K. (1959). The Solution of the General Least Squares Problem with Special Reference to High-Speed Computers. LA-2367, Office of Technical Services, U. S. Department of Commerce, Washington 25, D. C.

Rao, C. R. (1952). Advanced Statistical Methods in Biometric Research. John Wiley and Sons, Inc., New York.

Teichroew, D. (1953). Distribution Sampling with High Speed Computers. North Carolina State College Ph.D. Thesis.

Trotter, H. F. (1957). Gauss's Work (1803-1826) on the Theory of Least Squares. AEC-tr-3049, United States Atomic Energy Commission Technical Information Service Extension, Oak Ridge, Tennessee.

Turner, Jr., M. E. (1959). The Single Process Law: A Study in Nonlinear Regression. North Carolina State College Ph.D. Thesis, University Microfilms, Inc., Ann Arbor, Michigan.

Whittaker, E. and Robinson, G. (1944). The Calculus of Observations. Blackie and Son Limited, London.

VITA

Roger Hughes Moore

Candidate for the Degree of

Doctor of Philosophy

Thesis: ON THE LEAST SQUARES ESTIMATION OF PARAMETERS IN NONLINEAR MODELS

Major Field: Statistics

Biographical:

Personal Data: Born in Portland, Oregon, April 15, 1929, the son of Frank and Nana Moore.

Education: Attended grade school in Waldport and Ashland, Oregon; attended high school in North Bend and Gold Beach, Oregon, graduating from the latter in 1947; received the Bachelor of Science degree from the University of Oregon, with a major in Mathematics, in June, 1951; received the Master of Science degree from the University of Oregon, with a major in Mathematics and Statistics, in June, 1953; attended the Graduate Center of the University of New Mexico in Los Alamos, New Mexico; completed requirements for the Doctor of Philosophy degree in May, 1962.

Professional Experience: Joined the Los Alamos Scientific Laboratory, Los Alamos, New Mexico, as a Research Assistant in July, 1953; was promoted to Staff Member in January, 1955; served as a statistician for a variety of scientific disciplines and administrative functions, with constantly increasing emphasis on computer applications; was chosen in 1960 to complete the requirements for the Doctor of Philosophy degree with the support of the Los Alamos Scientific Laboratory's Advanced Study Program.