

ON THE GOODNESS OF ESTIMATED PARAMETERS
IN THE RATIO MODEL

By

PAUL EUGENE PULLEY, JR.

Bachelor of Science
Oklahoma State University
Stillwater, Oklahoma
1957

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1959

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, 1965

MAY 31 1966

ON THE GOODNESS OF ESTIMATED PARAMETERS
IN THE RATIO MODEL

Thesis Approved:

Robert D Morrison

Thesis Adviser

David J. Weeks

Paul R. Schmidt

R. B. Deal

F. Wayne Johnson

James H. Lant

J. W. Boyce

Dean of the Graduate School

581466

PREFACE

A survey of the literature concerning the subject of nonlinear estimation, in many instances seems to be a collection of specific cases where a solution (not necessarily the "best") has been arrived at through some prior knowledge of the ranges of the parameters being sought. It is also noted that even if the errors associated with the observations are quite small, the researcher may get an answer which at best is suspect.

The author's interest in the area of nonlinear estimation was aroused when he was confronted with a problem of estimation wherein the parameters sought were formation constants of complex ions in a transition metal complex. None of the literature was of any assistance in solving this problem. Consequently, the search for a solution was an expedition into the land of relative minima. Even though a reasonable answer was ultimately found, there was no assurance that it was the best or even that it was unbiased. Prior knowledge of an analytical form made the results acceptable to the researcher.

In the example just cited, the author made no representations regarding the "goodness" of the solution. Indeed, the method of solution was quite shakey, and, as a result, no inferences could be drawn. The "goodness" of a statistical procedure used to arrive at a solution should be evaluated. Bias and variance are relative under the general heading of "goodness" of an estimator. An estimate can be unbiased

and still be of little value if the variance of the estimate is large. Alternatively, one may have an estimate with little or no variance, but if it has an unknown bias, it can be equally without value. Desirable estimates or statistics are those with small variance and no bias.

In the material which follows, the author intends to avoid, where possible, the use of seemingly endless proofs of non-existence and instead substitute an example to show non-existence. Lastly, unless one agrees on some basic concepts, the process of argument is endless. With this idea in mind, the author will spend little time in restating those concepts with which practicing statisticians are already familiar.

ACKNOWLEDGMENTS

Because I lost most of my eyesight when I was not quite fourteen years old, I am indebted far more than most to the many people who have helped immeasurably during my entire college career and in the work involved in this dissertation. My mother and later my wife have, among many other labors in my behalf, read aloud all my texts, and I doubly appreciate their efforts knowing that the material they read was usually boring to them and often read when they were weary from other tasks. In spite of the advice of many that mathematics was too exacting a discipline for most students and virtually impossible for the blind, Dr. James Zant of the Oklahoma State University mathematics department allowed me to enroll to begin the study which led to my career in mathematics. His encouragement, advice and understanding have been invaluable. Words fall short in attempting to express my gratitude to Dr. L. Wayne Johnson, head of the mathematics department at Oklahoma State University, for all his help but especially for placing me in the first university research computing facility in Oklahoma. Here my strong interest in electronics and my training in mathematics found expression. I am indebted indeed to Dr. Robert D. Morrison, professor of statistics and chairman of my committee, for his endless hours of advice and counsel, both in the preparation of this dissertation and in other matters as well; to Dr. Robert Mac Vicar, University Vice-President and Dean of the Graduate

School, and to Professor George H. White, former Director of the Office of General Education, for their advice, encouragement and confidence; to Dr. Roy B. Deal, professor of mathematics, Dr. David L. Weeks, professor of statistics, and Professor Arlo Schmidt for the privilege of studying under them, having them serve on my committee and knowing them; to Edgar Butler for his assistance in the computing work and in the preparation of graphs; to Beverly Richardson for the painstaking work in typing this paper; to Dr. Louis P. Varga, professor of chemistry, for presenting a problem to me which stirred my interest in the subject of this dissertation, and to Oklahoma State University Computer Center for the use of the computer.

TABLE OF CONTENTS

Chapter	Page
I. THE LEAST SQUARES METHOD FOR LINEAR MODELS	1
II. LEAST SQUARES FOR THE NON-LINEAR MODELS.	4
III. FINDING A SOLUTION FOR THE RATIO MODEL.	16
Some Preliminary Investigations	17
Generating the Errors	19
IV. SOME PROPERTIES OF ESTIMATED PARAMETERS BY A SYSTEMATIC SEARCH PROCEDURE	22
V. AN ALTERNATIVE APPROACH TO NON-LINEAR ESTIMATION.	24
VI. SUMMARY AND CONCLUSIONS	31
BIBLIOGRAPHY	34
APPENDIX	35

LIST OF FIGURES

Figure		Page
1.	Distribution of Estimated Parameters for Generating Model $y = \frac{1}{1+x} + e$ and Estimating Model $y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + e$	36
2.	Distribution of Estimated Parameters for Generating Model $y = \frac{x}{1+2x} + e$ and Estimating Model $y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + e$	37
3.	Distribution of Estimated Parameters for Generating Model $y = 0.3 + e$ and Estimating Model $y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + e$	38
4.	Graphic Representation of $\frac{\partial \sum e^2}{\partial \alpha_1}$ as a Function of α_1	39
5.	Enlargement of Discontinuity Region of Figure 4	40

CHAPTER I

THE LEAST SQUARES METHOD FOR LINEAR MODELS

In the method of least squares, observations $y_i (i=1, \dots, n)$ are drawn from a population where each y_i is assumed to have the form

$$y_i = f_i(\theta_1, \theta_2, \dots, \theta_k, \dots, \theta_p) + \varepsilon_i,$$

where θ_i is a function of an unknown parameter β_i and a known function of x_i . The distribution of the random variable ε_i may or may not be known. The sum of squares

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - f_i(\theta_1, \theta_2, \dots, \theta_p))^2$$

is formed and the values of the β_k that minimize the $\sum_{i=1}^n \varepsilon_i^2$ are the least squares estimates of the β_k . Graybill (1961) shows that in many important cases when the distribution of the ε_i is unknown, the least squares method gives rise to estimates of the β 's which are unbiased, consistent, and, under certain conditions have minimum variance. It can be shown, in some instances, that the least squares estimate of the β 's and the maximum likelihood estimate of the β 's are the same.

In this and all succeeding sections, use will be made of matrix algebra. Summation notation will be used only when it is deemed necessary for clarity.

Let us assume that the random vector $\begin{matrix} e \\ n \times 1 \end{matrix}$ has mean \emptyset (null vector)

and $\sigma^2 I$ as its variance-covariance matrix. The distribution of the vector e will be unknown or unspecified. This removes the possibility of obtaining maximum likelihood estimators for the $(p \times 1)$ vector of the β_k 's, say β . In ordinary linear least squares procedures, one finds estimates of unknown parameters under the restriction that $e'e$ is minimum and each θ_k is some linear function of the form

$$\theta_k = \beta_k h_k(x_i)$$

where β_k is unknown and $h_k(x_i)$ is some known or assumed function of a known value x_i where x_i is a value of a single variable x . Let $\begin{matrix} Y \\ n \times 1 \end{matrix}$ be a vector of observations and let X be a matrix of known values where the i -th row of X is a $(1 \times p)$ row vector whose j -th element is some function of the x_i value that is associated with the i -th element in the Y vector. Let this function of x_i be represented by $h_j(x_i)$. The model could then be written as

$$Y = X\beta + e$$

in which $\begin{matrix} \beta \\ p \times 1 \end{matrix}$ is a vector of unknown parameters to be estimated. The i -th element in the Y vector could be represented by

$$y_i = \sum_{j=1}^p \beta_j h_j(x_i) + \varepsilon_i$$

Returning to the criterion that $e'e$ is to be minimized, we have

$$e'e = (Y - X\beta)'(Y - X\beta) = Y'Y - 2\beta'X'Y + \beta'X'X\beta.$$

The β that minimizes $e'e$ is found by obtaining the solution to

$$\frac{\partial e'e}{\partial \beta} = \phi.$$

From this we get

$$\hat{\beta} = (X'X)^{-1} X' Y$$

and also obtain an estimate of σ^2 by

$$\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n - p}$$

where n = the number of observations of y , and p is the number of parameters used to minimize $e'e$. The estimate for σ^2 can also be written

$$\hat{\sigma}^2 = \frac{Y'(I - X(X'X)^{-1}X')Y}{n - p}$$

Recalling that the distribution of e was not specified, it is not possible to make any assumptions about all functions of the random variable e . Graybill (1961), in discussing the Gauss-Markoff theorem, shows that under quite general conditions the least squares estimate of β in the model

$$Y = X\beta + e$$

is the best (minimum variance) linear (linear function of the vector Y) unbiased estimate of β if the following two conditions on the distribution of e are satisfied:

- a) The expected value of $e = \emptyset$, and
- b) The expected value of $ee' = \sigma^2 I$.

CHAPTER II

LEAST SQUARES FOR THE NON-LINEAR MODELS

Consider now the case where the unknown parameters are to be estimated using the general model

$$y = \frac{p(\beta, x)}{q(\alpha, x)} + \epsilon$$

where $p(\beta, x)$ and $q(\alpha, x)$ are polynomials in known values of x . The β 's and α 's are unknown coefficients (vectors), and the y values are observed. This model in general describes a large class of asymptotic functions. One of the simpler forms of the ratio of two linear functions is the bilinear model

$$y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + \epsilon.$$

This simple model may be used to describe many different types of data sets. Once again no assumption will be made about the distribution of ϵ other than that the expected value of $\epsilon = 0$, and the expected value of $\epsilon^2 = \text{constant}$.

We can write, using matrix notation,

$$\begin{array}{ccccccc} Y & = & \Gamma & X & \beta & + & e \\ n \times 1 & & n \times n & n \times p & p \times 1 & & n \times 1 \end{array}$$

or

$$Y - \Gamma X \beta = e$$

where Γ is a diagonal matrix whose elements are

$$\gamma_{ii} = [q_i(\alpha, x_i)]^{-1} = \left[1 + \sum_{j=1}^q \alpha_j h_j(x_i) \right]^{-1}.$$

In the least squares procedure we wish to minimize

$$e'e = Y'Y - 2\beta'X'\Gamma Y + \beta'X'\Gamma^2 X \beta.$$

Taking the partials of $e'e$ with respect to the β_k and α_j and

equating to zero, we have:

$$\frac{\partial e'e}{\partial \beta} = 0 = -2X'\Gamma Y + 2X'\Gamma^2 X \hat{\beta}.$$

$$\frac{\partial e'e}{\partial \alpha_0} = 0 \quad \text{since } \alpha_0 = \text{constant} = 1.$$

$$\frac{\partial e'e}{\partial \alpha_j} = 2\beta'X'[x]_j \hat{\Gamma}^2 Y - 2\beta'X'[x]_j \hat{\Gamma}^3 X \beta = 0.$$

where $[x]_j$ is a diagonal matrix whose elements, x_i^j , are the coefficients of the α_j . Now

$$\frac{\partial e'e}{\partial \beta} = 0$$

can be solved for $\hat{\beta}$, i.e.,

$$\hat{\beta} = (X'\Gamma^2 X)^{-1} X'\Gamma Y \quad \text{or} \quad \hat{\beta}' = Y'\Gamma X(X'\Gamma^2 X)^{-1}.$$

This solution is valid for any Γ so long as the γ_{ii} are bounded and $(X'\Gamma^2 X)^{-1}$ exists. Indeed Γ is nothing more than a transformation on the matrix X , and $\hat{\beta}$ is the least squares solution for a given Γ . This means that for a given Γ all the theorems concerning the general linear hypothesis models apply.

The procedure for estimating the α 's in the Γ matrix for a given model must be considered. Ahlfors (1963) states that "the sum and the product of two analytic functions are again analytic. The same is true of the quotient $\frac{f(z)}{g(z)}$ of two analytic functions, provided that

$g(z)$ does not vanish. ... Since the sum and product of analytic functions are again analytic, it follows that every polynomial is an analytic function. As long as Γ is bounded and the model is analytic, then

$$\begin{aligned} \frac{\partial^2 e'e}{\partial \alpha \partial \beta} &= \frac{\partial^2 e'e}{\partial \beta \partial \alpha} \\ &= 2 X' [x]_j \Gamma^2 Y - 4 X' [x]_j \Gamma^3 X \beta. \end{aligned}$$

One can readily see that if the denominator of the model has a zero within the range of the x_i , the model is not a continuous function. For example, consider the model

$$y = \frac{1}{1 + \alpha x} + \varepsilon.$$

This model is not defined when

$$\alpha = \frac{-1}{x_i}.$$

It seems only logical then to restrict oneself to parameter values of α_j that yield continuous non-zero values for $q(\alpha, x)$. It was noted earlier that the least squares estimate for the vector β , given Γ , was minimum variance and unbiased. The residual sum of squares of deviation for a given matrix Γ is

$$R = Y'Y - Y' \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma Y.$$

The problem arises when one attempts to find the vector α that will minimize R . Since $Y'Y$ is known, minimizing R is equivalent to maximizing $Y' \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma Y$. Before embarking on the quest for the estimate of the vector α , some preliminary observations should be made:

1. If the estimation model is to be continuous, every diagonal element of Γ must be non-zero and have the same sign. The proof is by inspection. Consider the model

$$y = \frac{p(\beta, x)}{q(\alpha, x)} + \varepsilon,$$

where

$$[q_i(\alpha, x_i)]^{-1} = \left(\sum_{j=0}^q \alpha_j x_i^j \right)^{-1} = \gamma_{ii}.$$

If the model is to be continuous, γ_{ii}^{-1} can never equal zero over the range of x . In addition, all γ_{ii} must have the same sign. If the sign of $q(\alpha, x)$ changes over the range of x , then $q(\alpha, x)$ contains at least one zero in the interval of x . This comes directly from the intermediate value theorem as treated by Johnson and Kiokemeister (1963).

2. It can be shown that the diagonal matrix Γ has at most $q < n$ elements that are equal. Every element of Γ is an inverse polynomial, and Ahlfors states that "a rational function $R(z)$ of order p has p zeros and p poles, and every equation $R(z) = a$ has exactly p roots." Since we are interested only in real values of a , we find that the number of times that the γ 's can take on the value a is less than or equal to p .

Let us now attempt to maximize the expression

$$R^*_{1 \times 1} = Y' \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma Y.$$

For a given set of Y and X and suitable restrictions such that the γ_{ii} are continuous over the range of x in $q(\alpha, x)$, we have R^* as a continuous function of (α, x) . Taking the partial of R^* with respect

to the α_j , we have

$$\begin{aligned} \frac{\partial R^*}{\partial \alpha_j} &= \left[\frac{\partial Y' \Gamma X}{\partial \alpha_j} \right] \left[(X' \Gamma^2 X)^{-1} X' \Gamma Y \right] + \\ & \left[Y' \Gamma X \right] \left[\frac{\partial (X' \Gamma^2 X)^{-1}}{\partial \alpha_j} \right] \left[X' \Gamma Y \right] + \left[Y' \Gamma X (X' \Gamma^2 X)^{-1} \right] \left[\left(\frac{\partial X' \Gamma Y}{\partial \alpha_j} \right) \right]. \end{aligned}$$

Noting that Y and X are constants and $\Gamma = \gamma_{ii}$ is diagonal, we can look at the three factors of $\frac{\partial R^*}{\partial \alpha_j}$ separately. The first and third are readily differentiable, but the second is not.

Consider a non-singular matrix $A_{p \times p}$ which contains the parameter α_j . We have

$$\begin{aligned} AA^{-1} &= I, \\ \left[\frac{\partial A}{\partial \alpha_j} \right] A^{-1} + A \frac{\partial A^{-1}}{\partial \alpha_j} &= 0, \\ \frac{\partial A^{-1}}{\partial \alpha_j} &= \left[-A^{-1} \right] \left[\frac{\partial A}{\partial \alpha_j} \right] \left[A^{-1} \right]. \end{aligned}$$

Let $A^{-1} = (X' \Gamma^2 X)^{-1}$. We have $A = X' \Gamma^2 X$ and

$$\frac{\partial A}{\partial \alpha_j} = -2 X' [x]_j \Gamma^3 X$$

where $[x]_j$ is a diagonal ($n \times n$) matrix. Thus

$$\frac{\partial (X' \Gamma^2 X)^{-1}}{\partial \alpha_j} = 2(X' \Gamma^2 X)^{-1} X' [x]_j \Gamma^3 X (X' \Gamma^2 X)^{-1}.$$

We also have

$$\frac{\partial Y' \Gamma X}{\partial \alpha_j} = -Y' [x]_j \Gamma^2 X$$

and

$$\frac{\partial X' \Gamma Y}{\partial \alpha_j} = -X' [x]_j \Gamma^2 Y.$$

Combining these results, we have

$$\begin{aligned} \frac{\partial R^*}{\partial \alpha_j} &= Y' [x]_j \Gamma^2 X (X' \Gamma^2 X)^{-1} X' \Gamma Y \\ &\quad + 2Y' \Gamma X (X' \Gamma^2 X)^{-1} X' [x]_j \Gamma^3 X (X' \Gamma^2 X)^{-1} X' \Gamma Y \\ &\quad - Y' \Gamma X (X' \Gamma^2 X)^{-1} X' [x]_j \Gamma^2 Y. \end{aligned}$$

Let $Q = \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma$ and note that

$$QQ = \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma = \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma = Q$$

Thus, $Q_{n \times n}$ is idempotent and less than full rank. Rewriting $\frac{\partial R^*}{\partial \alpha_j}$

in terms of Q and equating to zero, we have

$$\frac{\partial R^*}{\partial \alpha_j} = -Y' [x]_j \Gamma Q Y + 2Y' Q [x]_j \Gamma Q Y - Y' Q [x]_j \Gamma Y = 0$$

or letting

$$[x]_j \Gamma = \Gamma [x]_j = D$$

we have

$$2 Y' (Q - I) D Q Y = 0.$$

The matrix of the quadratic form is the matrix $A_{n \times n}$ in the function $s = Y' A Y$ where s is a scalar. In ordinary least squares procedures the matrix A contains no parameters to be estimated. It has been shown that this is not the case when the estimation model has the ratio form. Consider the quadratic form

$$s = Y' (2Q - I) D Q Y - Y' Q D Y - 2Y' [(Q - I) D Q] Y.$$

In this form

$$A_{n \times n} = [(Q - I) D Q]$$

which is not necessarily symmetric since D is a row operator on Q in the product DQ . It would be nice if an acceptable vector α existed

which could, in association with the known x_i , make the matrix

$$A = [(Q - I) DQ] = \phi.$$

Obviously, if $D = kI$, we would automatically have a solution since

$$[(Q - I) kIQ] = [(Q - I) IQk] = [Q^2 - Q] k = \phi.$$

D was defined to be $[x]_j \Gamma$ where

$$d_{ii} = x_i^j \gamma_{ii} = x_i^j \left[1 + \sum_{j=1}^q \alpha_j x_i^j \right]^{-1}.$$

If $D = kI$ is to be a solution, we have n such equations or polynomials that must be satisfied. But this is impossible since for any set of α_j we have a polynomial of degree $q < n$ which must attain the same value at n distinct points such that $D = kI$. With $D = kI$ eliminated as a possible solution, attention should be given to the entire matrix of the quadratic form. If a general solution exists wholly within the matrix of the quadratic form, then a solution should exist for the specific case having the estimation model

$$y = \frac{\beta}{1 + \alpha x} + \varepsilon.$$

In matrix notation the model can be written

$$Y = \Gamma J \beta + e$$

where J is a column vector of ones and

$$\gamma_{ii} = (1 + \alpha_1 x_i)^{-1}.$$

Also

$$Q = \Gamma J (J' \Gamma^2 J)^{-1} J' \Gamma$$

and

$$D = [x]_1 \Gamma.$$

Here

$$q_{ij} = (\sum \gamma_{ii}^2)^{-1} \gamma_{ii} \gamma_{jj},$$

and

$$a_{ii} = x_i \gamma_{ii}.$$

Thus, in the matrix equation $(Q - I)DQ = \phi$, we must have each element of the product equal to zero. Let us first examine the diagonal elements of this product assuming at least one of the $x_i = 0$ and all of the $x_i \geq 0$. The i -th diagonal element of the product is

$$\phi_{ii} = (\sum \gamma_{ii}^2)^{-1} \gamma_{ii}^2 \sum_{j=1}^n d_{jj} \gamma_{jj}^2 - d_{ii} (\sum \gamma_{ii}^2)^{-1} \gamma_{ii}^2.$$

Now since all of the γ_{ii} have the same sign and none of the $\gamma_{ii} = 0$, we can substitute for d_{ii} and γ_{ii} . We have

$$\sum_{j=1}^n d_{jj} \gamma_{jj}^2 = \sum_{j=1}^n x_j (1 + a_{1j} x_j)^{-3}.$$

Thus ϕ_{ii} cannot equal zero for any $x_i = 0$. The off diagonal elements follow the same pattern, and hence, the possibility of a general solution wholly within the matrix $[(Q-I)DQ]$ can be eliminated. Also to be remembered is the fact that for more complex estimation models (i.e., more parameters in $q(\alpha, x)$), two or more equations with the general matrix form $[(Q-I)DQ]$ must simultaneously be satisfied. These equations differ only in the factor D which has been eliminated as a source for a solution.

Consider now the problem of solving the simultaneous equations

$$\frac{\partial e'e}{\partial \beta} = \phi, \quad \frac{\partial e'e}{\partial \alpha_j} = 0.$$

We have already indicated that

$$\hat{\beta} = (X' \Gamma^2 X)^{-1} X' \Gamma Y$$

is minimum variance unbiased for any given Γ . Now examine

$$\frac{\partial e'e}{\partial \alpha_j} = 2 \beta' X' [x]_j \Gamma^2 Y - 2 \beta' X' [x]_j \Gamma^3 X \beta$$

which retains the same form across all j except for the $[x]_j$. One

can let $\alpha_0 = 1$ for all models. If we substitute our best estimate for β , which is the best for any Γ , we have

$$\begin{aligned} \frac{\partial e'e}{\partial \alpha_j} &= 2 Y' \Gamma X (X' \Gamma^2 X)^{-1} X' [x]_j \Gamma^2 Y \\ &\quad - 2 Y' \Gamma X (X' \Gamma^2 X)^{-1} X' [x]_j \Gamma^3 X (X' \Gamma^2 X)^{-1} X' \Gamma Y \end{aligned}$$

or, after substituting Q and D as before we have

$$\frac{\partial e'e}{\partial \alpha_j} = 2 Y' Q D Y - 2 Y' Q D Q Y.$$

Equating to zero we have precisely the same situation as before, namely a whole set of equations of the form $2 Y' (Q - I) D Q Y = 0$ for each j which must be satisfied.

Consider the case where $Y^* = Y - K$ where K is unspecified. We still want to maximize $Y' \Gamma X (X' \Gamma^2 X)^{-1} X' \Gamma Y$ or find the vector α such that $2 Y' (Q - I) D Q Y = 0$. Let

$$\begin{aligned} Y^{*'} [(Q - I) D Q] Y^* &= (Y - K)' [(Q - I) D Q] (Y - K) \\ &= Y' [(Q - I) D Q] Y - 2 K' [(Q - I) D Q] Y + K' [(Q - I) D Q] K = 0. \end{aligned}$$

If

$$Y' [(Q - I) D Q] Y = 0,$$

then

$$K' [(Q - I) D Q] (Y - \frac{1}{2} K) = 0$$

for some K . One might then choose an α and solve for K , and if $K = kJ$, we would have a solution for the model

$$y - k = \frac{p(\beta, x)}{q(\alpha, x)} + \epsilon$$

where the vector K is based on the chosen α . It should be possible to set up some convergent algorithm based on K as a function of α that would correct the vector K to \emptyset . Unfortunately, it appears that such a convergent procedure would be applicable only in the case where the estimation model contained only one α_j . In the case where more than one α_j is present, there is no guarantee that all K_j would be the same for a given vector α .

The number of possible solutions in the ratio model must be considered. It can be shown that a polynomial of degree n has n roots, some of which may be duplicates or may be complex. If we consider the model

$$y = \frac{\beta_0}{1 + \alpha x} + \epsilon$$

in matrix form, we have

$$Y = \Gamma J \beta_0 + e$$

where $J' = (1, 1, \dots)$. It follows that

$$\frac{\partial e'e}{\partial \alpha} = 2\beta_0 J' [x]_j \hat{\Gamma}^2 Y - 2\beta_0 J' [x]_j \hat{\Gamma}^3 J \beta_0 = 0$$

Converting this to summation notation, we have

$$\beta_0 \sum_{i=1}^n x_i \hat{\gamma}_{ii}^2 y_i = \beta_0^2 \sum x_i \hat{\gamma}_{ii}^3$$

If $\beta_0 \neq 0$, we have the result

$$\beta_0 \sum x_i \hat{\gamma}_{ii}^3 - \sum x_i \hat{\gamma}_{ii}^2 y_i = 0.$$

This may be written in a form which is the difference of two polynomials over the common denominator $\prod_{i=1}^n (1 + \hat{\alpha} x_i)$ where n is the number of distinct values of x . Since the restriction is made that each γ_{ii} has the same sign and is not equal to zero, the coefficient of β_0 in the first term becomes

$$\sum_{i=1}^n x_i \left[\prod_{\substack{j=1 \\ j \neq i}}^n (1 + \hat{\alpha} x_j) \right]^3$$

which is a polynomial of degree $(3n - 3)$ in α . The second term becomes

$$\frac{\sum_{i=1}^n x_i \gamma_{ii}^2 y_i}{\prod_{i=1}^n (1 + \hat{\alpha} x_i)} = \frac{\sum_{i=1}^n x_i y_i \prod_{\substack{j=1 \\ j \neq i}}^n (1 + \hat{\alpha} x_j)^2}{\prod_{i=1}^n (1 + \hat{\alpha} x_i)}$$

has degree $(3n - 2)$ in α . Now consider the equation in which

$$\frac{\partial e^t e}{\partial \beta_0} = \hat{\beta} (J' \Gamma^2 J) - J' \Gamma Y = 0.$$

Written in summation notation,

$$\hat{\beta}_0 = (\sum \gamma_{ii}^2)^{-1} \sum \gamma_{ii} y_i.$$

By substituting for γ_{ii}^2 and finding the common denominator one finds

$$\beta_0 = \frac{\left[\prod_{i=1}^n (1 + \alpha x_i) \right] \sum_{i=1}^n y_i \left[\prod_{\substack{j=1 \\ j \neq i}}^n (1 + \alpha x_j) \right]}{\sum_{i=1}^n \prod_{\substack{j=1 \\ j \neq i}}^n (1 + \alpha x_j)^2}$$

The above expression is of the first degree in α . The expression

$$\hat{\beta}_0 \sum x_i \gamma_{ii}^3 - \sum x_i \gamma_{ii}^2 y_i = 0$$

is a polynomial in α of degree $(3n - 2)$. We now have to find all the

zeros of a polynomial of degree $(3n - 2)$ in α in order to determine if there exists an α and β_0 that will minimize $e'e$. There is no guarantee that only one absolute minimum exists. There is no guarantee that any solution lies within an acceptable space for α .

The generalization can now be made to the case where the model contains two or more α_j . We have, in the case for $(\beta_0, \alpha_1, \alpha_2)$, two simultaneous polynomials. They are

$$\frac{\partial e'e}{\partial \alpha_1} = 0 = \sum_{i=0}^k c_i \hat{\alpha}_1^{k-i} \hat{\alpha}_2^i$$

and

$$\frac{\partial e'e}{\partial \alpha_2} = 0 = \sum_{i=0}^k c_i^* \hat{\alpha}_1^{k-i} \hat{\alpha}_2^i$$

where $k = (3n - 2)$ and c_i, c_i^* are functions of x_i and y_i which are derived from the data and estimation model. It can be heuristically argued that each of the polynomials can have either no solution or an infinity of solutions in an acceptable space for α_1 and α_2 . There is also the condition that among the infinity of roots, should they exist, the only acceptable ones are those for which both polynomials are simultaneously satisfied.

CHAPTER III

FINDING A SOLUTION FOR THE RATIO MODEL

It has been demonstrated that even in the simplest case of estimating β_0 and α_1 for the ratio model

$$y = \frac{\beta_0}{1 + \alpha_1 x} + \varepsilon,$$

there are $(3n - 2)$ possible solutions in an acceptable region. More elaborate models contain much higher degree polynomials, and for more than one α_j , we also have the problem of satisfying the partial equations simultaneously. Therefore, it was decided to restrict this study to the bilinear model

$$y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + \varepsilon.$$

It was desired to find, if possible, some estimate of the "goodness" of the estimated parameters and what, if any, inferences could be made about them. To do this, the Gauss-Seidel relaxation procedures for various models with no error were first examined. These iterative methods are often used with good results in many scientific applications. They can be found in elementary texts on numerical analysis. These methods do have their drawbacks, and in the cases of the bilinear models tried, they were found to be unsatisfactory. In these attempts to find solutions for bilinear models, it was found that under

all possible permutations (six in the case for $\beta_0, \beta_1, \alpha_1$), the rates of convergence were extremely slow even with good starting values. In many instances there was no convergence at all. It should be noted that better results were obtained with these methods when the dependent parameter in a given equation was the variable with respect to which the partial e'e was taken. There seems to be a logical reason for this since, in the case of ordinary linear regression models, these procedures have a tendency to diverge if the wrong dependent parameter is considered in a given set of normal least squares equations.

Some Preliminary Investigations

To determine, in some degree, the difficulties in obtaining a solution for the ratio model, it was decided that an investigation should be made in order to find as many solutions as possible for different data sets from different models with and without added error. In each case, the estimating model was bilinear, but the data generating function or model was not necessarily bilinear. One of the criteria for ordinary linear regression is that the expected value of e is \emptyset and that the estimate $\hat{\beta}$ forces the sum of the residuals to zero when β_0 is present. It seems reasonable to require the same result for the sum of residuals in the non-linear case. Accordingly, we can attempt (with no guarantee of success) to force this result if we always use the estimate for β , say $\hat{\beta}$, as a function of Γ . Thus

$$\hat{\beta} = (X' \Gamma^2 X)^{-1} X' \Gamma Y$$

can be substituted in the expression

$$\frac{\partial e'e}{\partial \alpha_1} = 2 \hat{\beta}' X' [x]_1 \Gamma^2 Y - 2 \hat{\beta}' X' [x]_1 \Gamma^3 X \hat{\beta} =$$

$$2Y' \Gamma X (X' \Gamma^2 X)^{-1} X' [x]_1 \Gamma^2 Y - 2Y' X (X' \Gamma^2 X)^{-1} X' [x]_1 \Gamma^3 X (X' \Gamma^2 X)^{-1} X' \Gamma Y$$

which has only one parameter α_1 . Any zero of this expression will satisfy the least squares regression requirements except for the fact that each zero is either a relative minimum, relative maximum or point of inflection for the residual sum of squares function. Since the number of possible solutions increases with n , it was decided to take a direct approach and systematically search out as many roots or solutions as could be found for a given set of eleven data points. In each search for roots, resolution of 0.001 was used for α_1 on the interval $-100.0 \leq \alpha_1 \leq 100.0$ and the values of $\hat{\beta}_0$, $\hat{\beta}_1$, α_1 , $\Delta_i = y_i - \hat{y}_i$, $\Sigma \Delta$, and $\frac{\Sigma \Delta^2}{8}$ were tabulated. This procedure yields real valued roots for the parameters. Any other root α_1^* that was within 0.001 of a previously discovered root for α would be bypassed. It should be pointed out that the interval for α_1 included all the points of discontinuity. The eleven x values always used were (0, 1, 2, . . . , 10). Thus the discontinuous values for α_1 were (-1, -1/2, . . . , -1/10). Actually, any values for α_1 such that $\alpha_1 \leq -0.1$ causes the bilinear estimation model to have points of discontinuity over the range of x .

Some of the generating models used in this phase of the study were:

$$y = \frac{1}{1+x}$$

$$y = \frac{x}{1+x}$$

$$y = \frac{1+x}{1+2x}$$

$$y = \frac{6+7x}{1+5x}$$

$$y = \frac{x^2}{3+x}$$

$$y = \frac{1}{1+x} + \epsilon$$

$$y = 6 + \epsilon.$$

When no error was added to a bilinear function, it was found that although several roots existed, there was only one where $\frac{\sum \Delta^2}{8} = 0$. When errors were added, it was discovered that, among the roots found, there existed more than one of equal merit from a data description standpoint. It was also found that in some cases, the number of solutions obtained increased as the added error decreased. This is not saying that bad data can be fitted with more ease. It is cited only to point up the fact that good data and a poor selection of the estimation model may cause additional problems in estimation. One test run on a set of data generated by the model $y = 6 + \epsilon$, where the $\sigma^2 = 0.01$, yielded 141 roots of which four were of equal merit. Any one of the 141 solutions was reasonable.

Generating the Errors

Reference has been made to "added errors." It is not practical to use random numbers from a published table when dealing with a problem of some magnitude. There are mathematical procedures which furnish a satisfactory substitute. These methods, though not random in the sense of being unrepeatable, are quickly computable

and have an extremely large cycle. The procedures described below produce numbers u_n that are approximately uniformly distributed on the unit interval $(0, 1)$. It is relatively easy to generate random numbers having other than rectangular distributions from a source that is rectangular or uniform. Let u be uniformly distributed on $(0, 1)$. To convert u to a normally distributed variable with mean μ and variance σ^2 , we note that the variable u has mean $= 1/2$ and variance $= 1/12$. By use of the central limit theorem it is seen that $\epsilon = (\bar{u} - \frac{1}{2}) \sqrt{12n}$ approaches a normal distribution with mean zero and variance one as n increases without bound. If $n = 12$, we have $\epsilon = \sum_{i=1}^{12} u_i - 6$. Preliminary tests for $n = 12$ proved to be a large enough sample from $f(u)$ to give a satisfactory normal distribution for ϵ .

To obtain the uniform random numbers u , one has only to call upon the theory of power residuals. Any computing subroutine that has a finite input will eventually repeat. This makes it imperative for any systematic procedure to have a very long cycle. Congruent numbers (modular numbers) are defined as elements of a set such that for every x, y in the set, the quantity $(x-y)$ is exactly divisible by some fixed value m , which is the modulus of the set. A full discussion of the theory is rather long. IBM Corporation gives a thorough analysis and description of the method and theory. The conclusion was made that among all methods which were statistically acceptable, the following procedure has the most merit. Since the random errors sought were to be computed on a base 10 computer, it was determined that a multiplier $m = 10011$ and a starter $u_0 = 1234567373$ would be

Satisfactory. This choice has a cycle of 500,000,000. Any number m , of the form $m = 200 T \frac{1}{2} R$ may be used. T is any positive integer, and some of the acceptable R values are 3, 11, 13, 19, 21, 27. The starting value u_0 can be any 10-digit number not ending in 0 and not divisible by 2 or 5. Then, computationally, u_{i+1} consists of the ten low-order positions of the product mu_i . The decimal point for each value of u is considered to be at the left so that $0 < u < 1$.

CHAPTER IV

SOME PROPERTIES OF ESTIMATED PARAMETERS

BY A SYSTEMATIC SEARCH PROCEDURE

The computer used in this study has a relatively fast (4.5 μ s) access time, and in spite of this speed, considerable computing time was used for this investigation. All preliminary computing was done with programs written in Fortran (4) compiling language. This method of programming, though costly in machine time, allowed for quick (one shot) testing of different methods over several data sets. The determination of bias and covariance of estimated parameters presents an entirely different problem. In Monte Carlo procedures, one has to repeat an algorithm enough times to establish, within reason, some trend or probability. Because of the repetitive nature of this part of the study, it was decided to do the balance of the programming in symbolic machine language, which is generally six to ten times as efficient as Fortran. It was also decided to restrict the study to three generating models:

$$y = \frac{1}{1+x} + \epsilon$$

$$y = 0.3 + \epsilon$$

$$y = \frac{x}{1+2x} + \epsilon$$

Each data set was taken over the same x values, namely

$x_i = (0, 1, 2, \dots, 10)$. The added errors ε_i were drawn (generated) from a quasi-normal distribution having a mean = 0 and standard deviation = 0.1. This sets the c.v. (coefficient of variation) at approximately 30 per cent. The estimation model for each test was

$$y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + \varepsilon$$

with the acceptable region for α_1 restricted to the interval $(-0.08 < \alpha_1 < 4.0)$. All the estimates were tabulated for each data set, and the best estimate from each set was selected for the study of bias and covariance.

All three generating models were reasonably well behaved. The prior knowledge of the driving forces (parameters) for each data set made the selection of the acceptability space quite easy. Such prior knowledge is unusual in most cases of non-linear estimation. There were 1088 data sets tried of which 905 had solutions in the acceptable space. In each of the selected sets of estimates $(\beta_0, \beta_1, \alpha_1)$, there was a sufficiently good set such that $\Sigma \Delta \leq 10^{-6}$ and $\frac{\Sigma \Delta^2}{8} \leq 0.01$ which was essentially the mean and variance of the quasi-normal added errors. In other words, the data sets were adequately described by the parameters that were estimated. The fitted line in each instance passed inside the data points as indicated by the randomness of the distribution of the signs for the Δ_i over the range of x .

CHAPTER V
AN ALTERNATIVE APPROACH TO
NON-LINEAR ESTIMATION

Gauss method (sometimes called Gauss-Seidel, Seidel, Gauss-Newton, or Newton-Raphson) is primarily based on linearizing the estimation model with respect to the desired parameters by means of a truncated Taylor series. Starting values or initial estimates are substituted for the parameters in the series or normal equations, and a set of correction factors is obtained. These correctors are then weighted and added to the previous estimates and the process repeated. Hopefully, this sequence of estimates will converge to or stabilize on an acceptable set of parameter estimates. Two of the more recent papers on the Gauss method are quite good. Moore and Ziggler (1959) indicate that the "least squares" solutions derived by this method are usually quite satisfactory and that the inverse of the matrix can be used for estimating variances and covariances if the y_i are assumed to be normally distributed. They state, "From these estimates we can compute single and joint confidence intervals for the parameters as indicated for the linear problem." Hartley (1959) states that "when the regression function is non-linear in the parameters, both the theory and the practice of the estimation procedure is considerably more difficult." In the presentation of his modification

to the Gauss method and in the accompanying proofs, he assumes certain conditions which are present in the ratio estimation model. Some of the conditions are:

1. The estimation model is continuous, and both first and second partial derivatives exist.
2. The residual sum of squares of deviation function is always greater than 0 for all observed vectors and estimating parameters in a bounded convex set S of the parameter space.
3. The assumption is made that it is possible to find a suitable starting vector that, along with the properties of convex sets, will guarantee convergence.

With some additional assumptions, Hartley proves that an absolute minimum can be achieved.

By means of his assumptions and convex theory, Hartley proves that all of the first partial derivatives in the sequence approach zero as a limit. This satisfies part of the requirements for a least squares solution. He indicates that it is highly unlikely that two or more sets of estimation parameters yield the same residual sum of squares. He states, "It is also clear that an absolute minimum of Q is inside S ." Here Q is the residual sum of squares function, and S is the convex set described above. The statements regarding the uniqueness of both the estimates and the absolute minimum are based on the fact that the matrix of the normal equations is positive definite." . . .

If the region S . . . can be chosen to be the region containing the absolute minimum, . . . then our iterative process will converge to the vector θ^+ yielding the absolute minimum of Q ." Here θ^+ is the solution vector.

The preceding, from Hartley (1959), is a logical and effective presentation of the theory. The difficulty arises when one must necessarily find the starting vector θ_0 that resides in the convex set S . Local minima may be relatively near to one another so that the "wide grid" suggested by Hartley may have to be changed to a fine one. Even then there is no assurance that the local minimum found is not adjacent to one where the function Q is nearer to zero. Hartley adds, "If there is a problem in which the absolute minimum is not unique for all large samples, the least squares principle ceases to be an appropriate method for estimation since it will be incapable of distinguishing between the two solutions. . . . It is only the unique vector which yields the absolute minimum of Q that is of interest in statistical estimation theory."

Consider now the bilinear estimation model

$$y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + \varepsilon$$

from the standpoint of the Gauss procedure. The factor $(1 + \alpha_1 x)^{-1}$ in this model can be expanded into or written as a Taylor series. We have:

$$(1 + \alpha_1 x)^{-1} = \sum_{i=0}^{\infty} (-\alpha_1 x)^i$$

which is analytic everywhere except for the point $(\alpha_1 x) = -1$. Unfortunately, the Gauss method calls for a truncated series, and for $(\alpha_1 x) > 1$ we have term for term divergence. Since it is not feasible to restrict $(\alpha_1 x)$ to the interval $(0 \leq \alpha_1 x < 1)$, we must then use the factor $(1 + \alpha_1 x)^{-1}$ itself for the "linearizing" and evaluation parts in the Gauss procedure. Taking the cue from the ordinary linear

least squares procedure, we have the function

$$Q = \sum \epsilon^2 = \sum \left[y - \frac{p(\beta, x)}{q(\alpha, x)} \right]^2 = \sum \left[y - \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} \right]^2.$$

If we have sufficiently good starting values, the ratio model could be treated as if it were linear. Remembering the rule for compound (two factor) differentiation, we have for the bilinear estimation model the vector

$$\frac{\partial \epsilon}{\partial (\beta, \alpha)} = \left(\frac{-1}{1 + \alpha_1^* x}, \frac{-x}{1 + \alpha_1^* x}, \frac{\beta_0^* + \beta_1^* x}{(1 + \alpha_1^* x)^2} \right).$$

Here the superscript * is used to indicate a parameter held constant for purposes of differentiation. Except for the factor $(1 + \alpha_1 x)^2$ we have a "linearized" vector if starting values are substituted for $(\beta_0^*, \beta_1^*, \alpha_1^* = \alpha_1)$. Thus the "linearized" bilinear estimation model becomes a "linear" model with parameter vector $(\beta_0, \beta_1, \alpha_1)'$. This new model has starting values $(\beta_0^*, \beta_1^*, \alpha_1^*)$ within it, and it would be sheer folly to solve for $(\hat{\beta}_0, \hat{\beta}_1, \hat{\alpha}_1)$. We should then be interested in some method of obtaining a correction for the starting values. From elementary numerical analysis (Salvadori and Baron (1955)), we can obtain such a corrector. Consider the set of simultaneous equations

$$\begin{array}{ccc} A & X & = & C \\ n \times n & n \times 1 & & n \times 1 \end{array}$$

where A is non-singular. A starting value X^* is chosen and the product $A X^* = C^*$ is evaluated. Then the system

$$A X_{\Delta} = C - C^*$$

is formed and a solution is obtained for the vector X_{Δ} . The correction

X_{Δ} is then weighted and added to the vector X^* to obtain a better estimate for the vector X . The weight is usually ≤ 1 , and the sequence is repeated until $C - C^* \doteq \emptyset$. If $C - C^* = \emptyset$, then $X_{\Delta} = \emptyset$ and the solution X has been obtained.

We can now write Q in the "linearized" form

$$Q = \sum \left[y - \frac{\beta_0 + \beta_1 x}{(1 + \alpha_1 x)} + \frac{\alpha_1^* x (\beta_0^* + \beta_1^* x)}{(1 + \alpha_1^* x)^2} \right]^2.$$

In matrix notation the new Q takes the familiar form

$$Q = e'e = (Y - Z\underline{\rho})'(Y - Z\underline{\rho})$$

where the rows of Z have the

$$Z_i = \left(\frac{1}{(1 + \alpha_1^* x_i)}, \frac{x_i}{(1 + \alpha_1^* x_i)}, \frac{-x_i(\beta_0^* + \beta_1^* x_i)}{(1 + \alpha_1^* x_i)^2} \right)$$

and the vector

$$\underline{\rho}' = (\beta_0, \beta_1, \alpha_1).$$

But, because of the starting values in the matrix Z , we want a correction factor $\underline{\rho}^*$ such that $w\underline{\rho}^* + (\beta_0^*, \beta_1^*, \alpha_1^*)$ will converge to the parameter $(\hat{\beta}_0, \hat{\beta}_1, \hat{\alpha}_1)$ that will minimize Q . Accordingly, we can define

$$y_i^* = \frac{\beta_0^* + \beta_1^* x_i}{1 + \alpha_1^* x_i}$$

and let

$$\epsilon_i^* = y_i - y_i^*.$$

Thus $\underline{\rho}^*$ is the correction vector that will satisfy the ordinary linear regression model:

$$e^* = Z_p^* + e.$$

The weight, usually positive and ≤ 1 can also be determined by finding the minimum point of a quadratic curve passing through three successive values of Q . This minimum of the quadratic function used to determine w is suggested by Hartley (1959) and is used in his proof of convergence to the absolute minimum for Q .

The modified Gauss procedure as outlined above has its faults. The matrix Z can become singular due to poor starting values, a bad combination of data and estimation model, or unfortunate oscillation in the correction vector. It also drops the requirement that $\Sigma \Delta = 0$. It does have the distinct advantage, in the case of good starting values, that it simultaneously satisfies $\frac{\partial e'e}{\partial (\beta, \alpha)} = 0$ when, the estimation models become more complex when for example, there are two or more α_j . Since the matrix $Z'Z$ is positive definite, the procedure will obtain the parameter estimates that coincide with local minima. Theoretically, the method is sound even though it fails when the estimation model is exact and there is no added error. This is due to the fact that the matrix $Z'Z$ becomes singular as the estimates for the parameters approach the true value. The theory requires that Q is greater than 0 for all estimation models and corresponding data sets.

The systematic search procedure which obtains all possible real valued estimates yielded some 905 solutions or parameter sets across three data generating models. In each search there was the requirement that $\Sigma \Delta = 0$ and $\Sigma \Delta^2$ be minimum. All of the parameter sets were quite good from a data description standpoint. The normal

equations were always satisfied for the data set that had estimation parameters in an acceptable region. It logically follows that these estimates should be good starting values for testing the modified Gauss procedure. With the relaxation of the restriction $\sum \Delta = 0$, one should expect a shift in the parameter values between the two procedures. It would appear that even better estimates would be obtained under the relaxation. The exact opposite occurred. The least squares criterion requires that $\sum \Delta^2$ be minimum. Out of the 905 trials of the modified Gauss procedure, there were only three sets where the sums of squares of deviation were essentially equal to the search procedure and only one where there was any improvement. The sum of deviations was never zero.

CHAPTER VI

SUMMARY AND CONCLUSIONS

The results of this study should point up the fact that procedures in non-linear estimation are far from what is statistically desirable. It is not practicable to embark on a systematic search for a solution in the absence of large-scale computing facilities. When the estimation model has the bilinear form, one has to search out all possible parameter sets before one can be assured that the best sets have been found. This has its analogy in the theory of stepwise regression wherein one attempts to seek out the best model to describe a particular data set and bases a decision on the sums of squares removed by different sets of parameters. Unless all possible models have been tried, one cannot be sure that the best model or set of best models has been found. There is also the problem of propagated errors that accompanies any numerical procedure that is programmed into a computer. Any well-designed computer has built-in electronic logic checking circuits that will set error indicators if the computer makes an internal mistake. The computer will do exactly what it is programmed to do. Propagated errors come from poor programming techniques or insufficient numerical analysis of the problem from a computation standpoint. Some of the most innocuous looking analytic functions or computing algorithms that are mathematically sound do

not yield readily to good computer programming. To this Pandora's computing box add the flatness of the polynomials derived from the data used and bilinear estimation model, and you have a situation in which the parameters estimated can be suspect. The error propagation problems are the same for the modified Gauss procedure but to a higher degree. This is due to the greater amount of computing that is necessary in building the $Z'Z \mid Z'e^*$ matrix and finding $\underline{\rho}^*$ for each iteration. The theory and methods of controlling propagated errors in high-speed computation is a discipline in itself. Consequently, much thought went into the preparation of the symbolic machine language programs used in this study.

The bias in the estimates cannot be due to the programming or the computer. The fault, from a statistical standpoint, lies either with the estimation procedure or with the least squares criterion. The added errors had a near normal distribution, but the distribution of the estimated parameters was evidently not normal under repeated sampling. In the tabulations and graphs exhibited in the Appendix, one will see little similarity between the estimated parameters and the driving forces (parameters) in the generating models. The relatively small acceptable space for the parameters served only to narrow the parameter distribution and hence make a bad situation look a little better. There was adequate evidence that numerous estimates were outside the acceptable space and that these estimates could be as good or better than the ones selected. The space restriction also caused approximately one-third of the data sets to be rejected on the basis of not having an acceptable solution. Both of the estimation procedures

were apparently successful in describing the data even though the modified Gauss procedure sometimes diverged and, almost without exception, had a sizeably larger sum of squares of deviation. In other words, the modified Gauss procedure was not as good as the systematic search procedure from the least squares standpoint.

There is still much work to be done and many questions to be answered in the theory of non-linear estimation. Throughout the entire study, as set forth in this paper, one of the answers became increasingly evident. It is that no statistical inferences can be associated with the estimated parameters in the ratio model. All one can do is to attempt to describe the data with the knowledge that other estimates can exist that are as good or better and that a different procedure can sometimes ferret out still better estimates.

BIBLIOGRAPHY

Ahlfors, Lars V. Complex Analysis. New York: McGraw-Hill Book Company, Inc., 1953, pp. 38 - 44.

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

Hartley, H. O. The Modified Gauss-Newton Method for the Fitting of Non-Linear Regression Functions by Least Squares. Ames, Iowa: Iowa State University Statistical Laboratory, 1959. (Technical Report No. 2.3).

Hille, Eninar. Analytic Function Theory, Vol. 1. Dallas: Ginn and Company, 1959, p. 207.

IBM Reference Manual: Random Number Generation and Testing and Programs Related Thereto.

Johnson, Richard and Fred L. Kiokemeister. Calculus with Analytic Geometry. Boston: Allyn and Bacon Inc., 1960, p. 217.

Moore, R. H. and R. K. Zeigler. The Solution of the General Least Squares Problem with Special Reference to High-Speed Computers. Los Alamos, New Mexico: Los Alamos Scientific Laboratory, 1959, Chapter 2.

APPENDIX

SUMMARY OF DISTRIBUTION DATA OF THE ESTIMATED
PARAMETERS IN THE BILINEAR ESTIMATION MODEL

Generating model $y = \frac{1}{1+x} + e$ Number of trials 506
 Number having solutions 505
 Average number of solutions 1.00

Estimating model $y = \frac{\beta_0 + \beta_1 x}{1 + a_1 x} + e$

Acceptable Region $-0.08 \leq a_1 \leq 4.0$

	β_0	β_1	a_1
Means of estimated parameters	.997	.236	1.111
Covariance of estimated parameters	.0095	.0016	.0216
		.0124	.0585
Mean of $\hat{\sigma}^2 =$.01006		.3537
Variance of $\hat{\sigma}^2 =$.0000230		

Generating model $y = \frac{x}{1+2x} + e$ Number of trials 265
 Number having solutions 200
 Average number of solutions 1.01

Estimating model $y = \frac{\beta_0 + \beta_1 x}{1 + a_1 x} + e$

Acceptable Region $-0.08 \leq a_1 \leq 4.0$

	$\hat{\beta}_0$	$\hat{\beta}_1$	\hat{a}_1
Means of estimated parameters	.014	.766	1.496
Covariance of estimated parameters	.0138	-.0222	-.0147
		.1944	.4097
Mean of $\hat{\sigma}^2 =$.00987		.8816
Variance of $\hat{\sigma}^2 =$.0000245		

Generating model $y = 0.3 + e$ Number of trials 317
 Number having solutions 200
 Average number of solutions 1.26

Estimating model $y = \frac{\beta_0 + \beta_1 x}{1 + a_1 x}$

Acceptable Region $-0.08 \leq a_1 \leq 4.0$

	$\hat{\beta}_0$	$\hat{\beta}_1$	\hat{a}_1
Means of estimated parameters	.290	.171	.558
Covariance of estimated parameters	.0073	-.0034	-.0078
		.0479	.1509
Mean of $\hat{\sigma}^2 =$.01088		.4882
Variance of $\hat{\sigma}^2 =$.0000313		

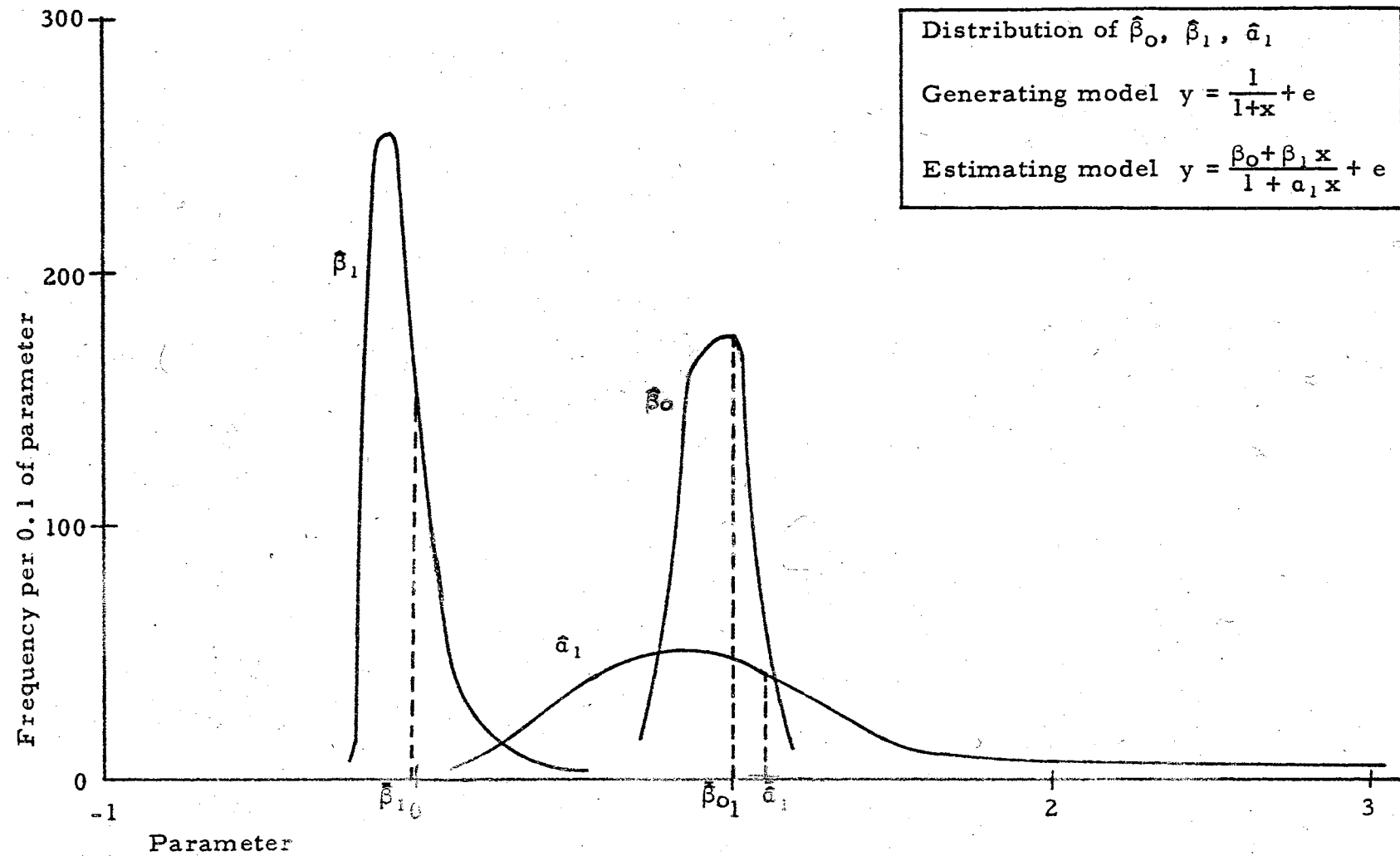


Fig. 1 Distribution of Estimated Parameters for Generating Model $y = \frac{1}{1+x} + e$ and Estimating Model $y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + e$.

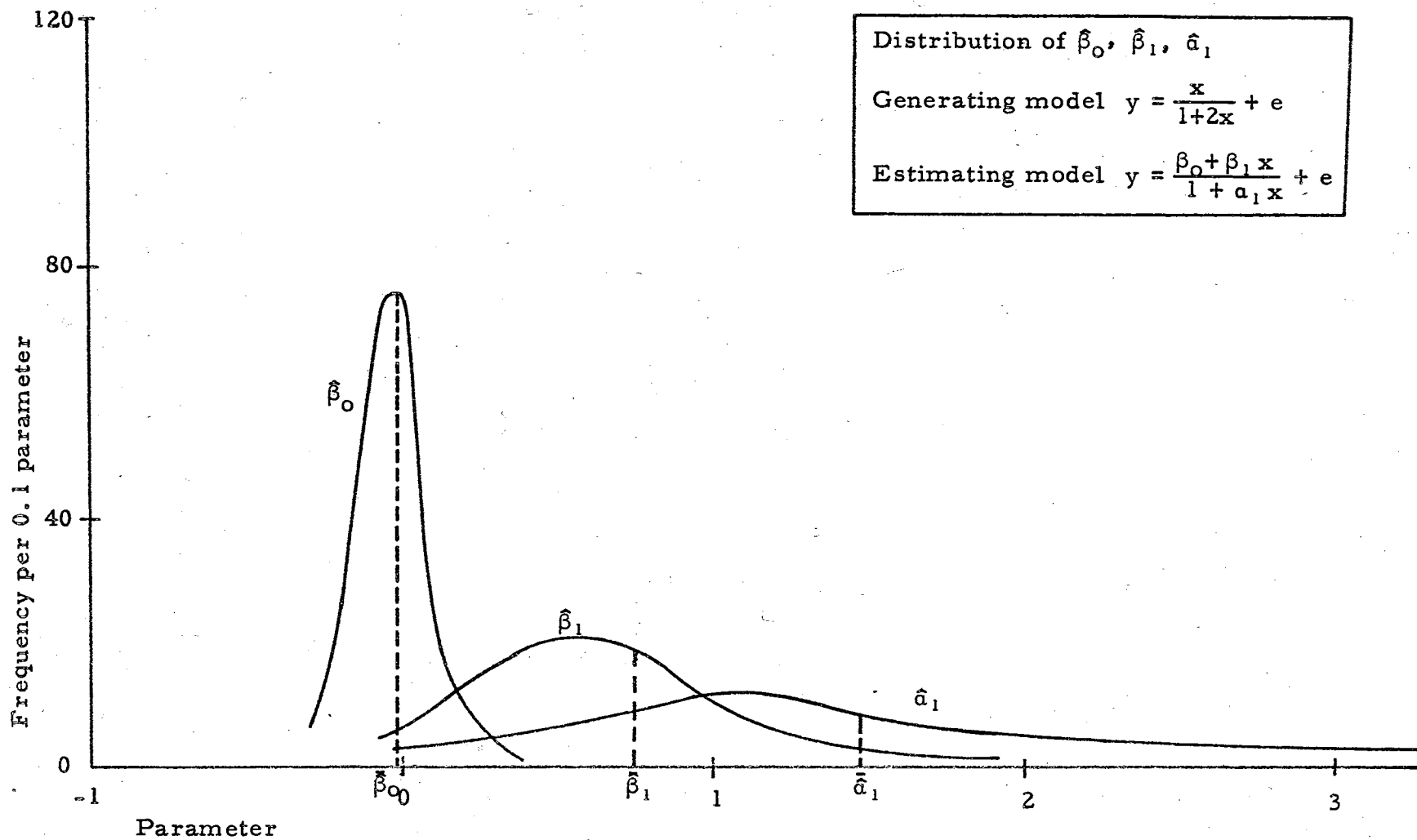


Fig. 2 Distribution of Estimated Parameters for Generating Model $y = \frac{x}{1+2x} + e$ and Estimating Model $y = \frac{\beta_0 + \beta_1 x}{1 + a_1 x} + e$.

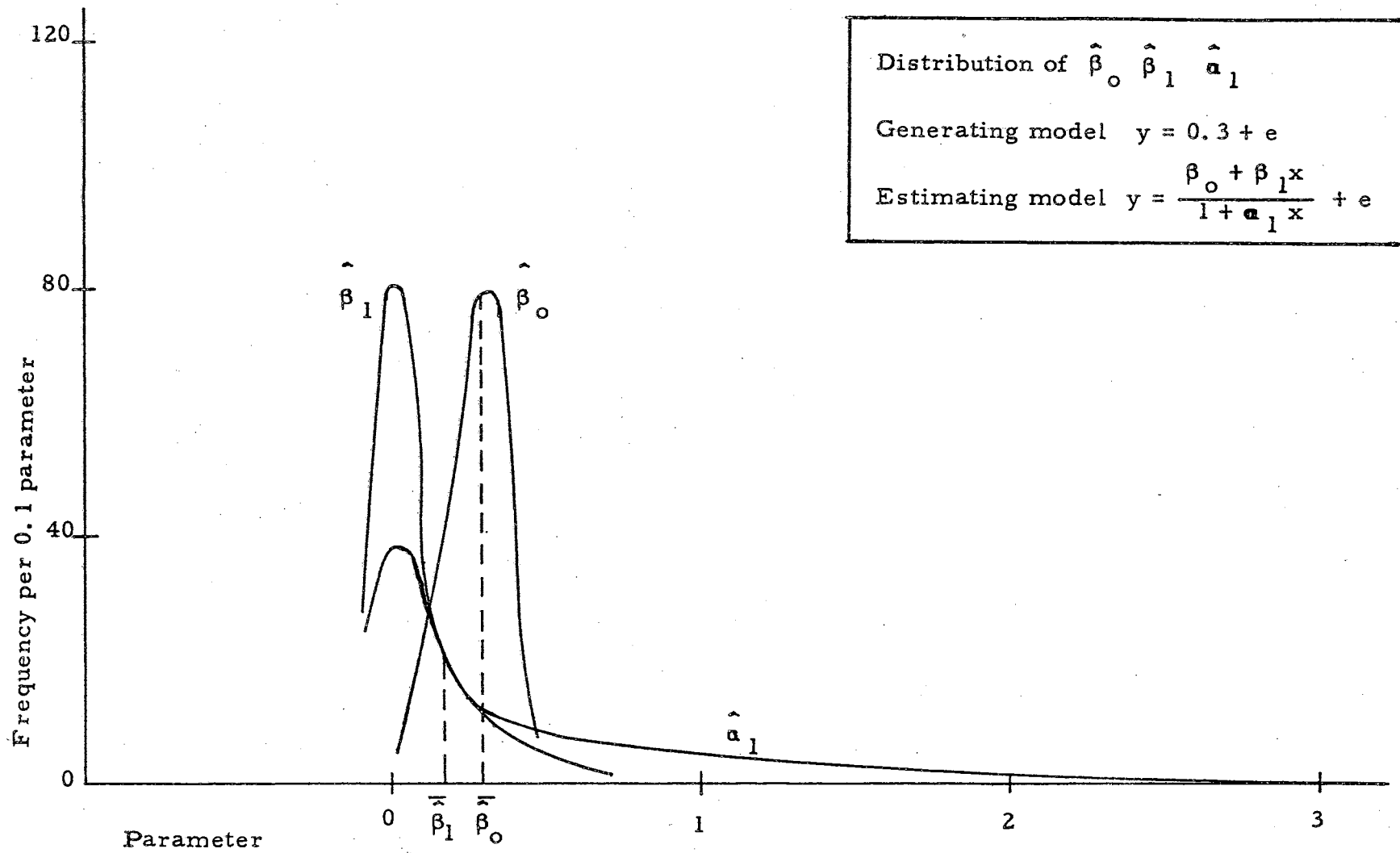


Fig. 3 Distribution of Estimated Parameters for Generating Model $y = 0.3 + e$ and Estimating Model $y = \frac{\beta_0 + \beta_1 x}{1 + \alpha_1 x} + e$.

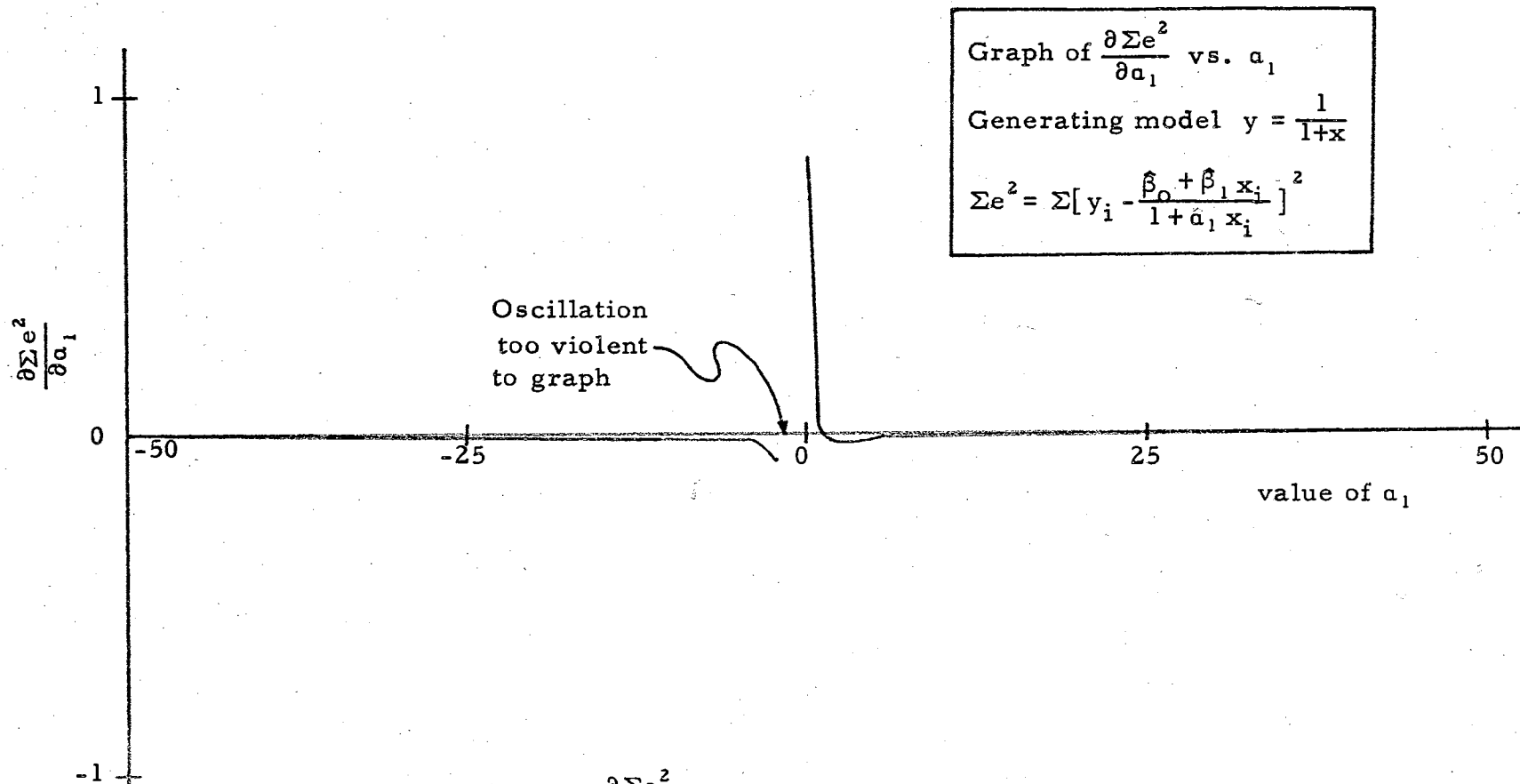


Fig. 4 Graphic Representation of $\frac{\partial \Sigma e^2}{\partial a_1}$ as a Function of a_1 . (The least squares estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are obtained as functions of a_1 and are substituted in a manner such that $\frac{\partial \Sigma e^2}{\partial a_1}$ is a function of a_1 alone. Every zero of $\frac{\partial \Sigma e^2}{\partial a_1}$ satisfies the least squares normal equations in a manner such that the lack of fit $\Sigma(y_i - y_i^*) = 0$ at each zero of $\frac{\partial \Sigma e^2}{\partial a_1}$. The generating model is $y = \frac{1}{1+x}$ for $x=0, 1, \dots, 10$. The estimating model is $y_i^* = \frac{\hat{\beta}_0 + \hat{\beta}_1 x_i}{1 + a_1 x_i} + e$. Computing resolution is $\Delta a_1 = 0.001$.)

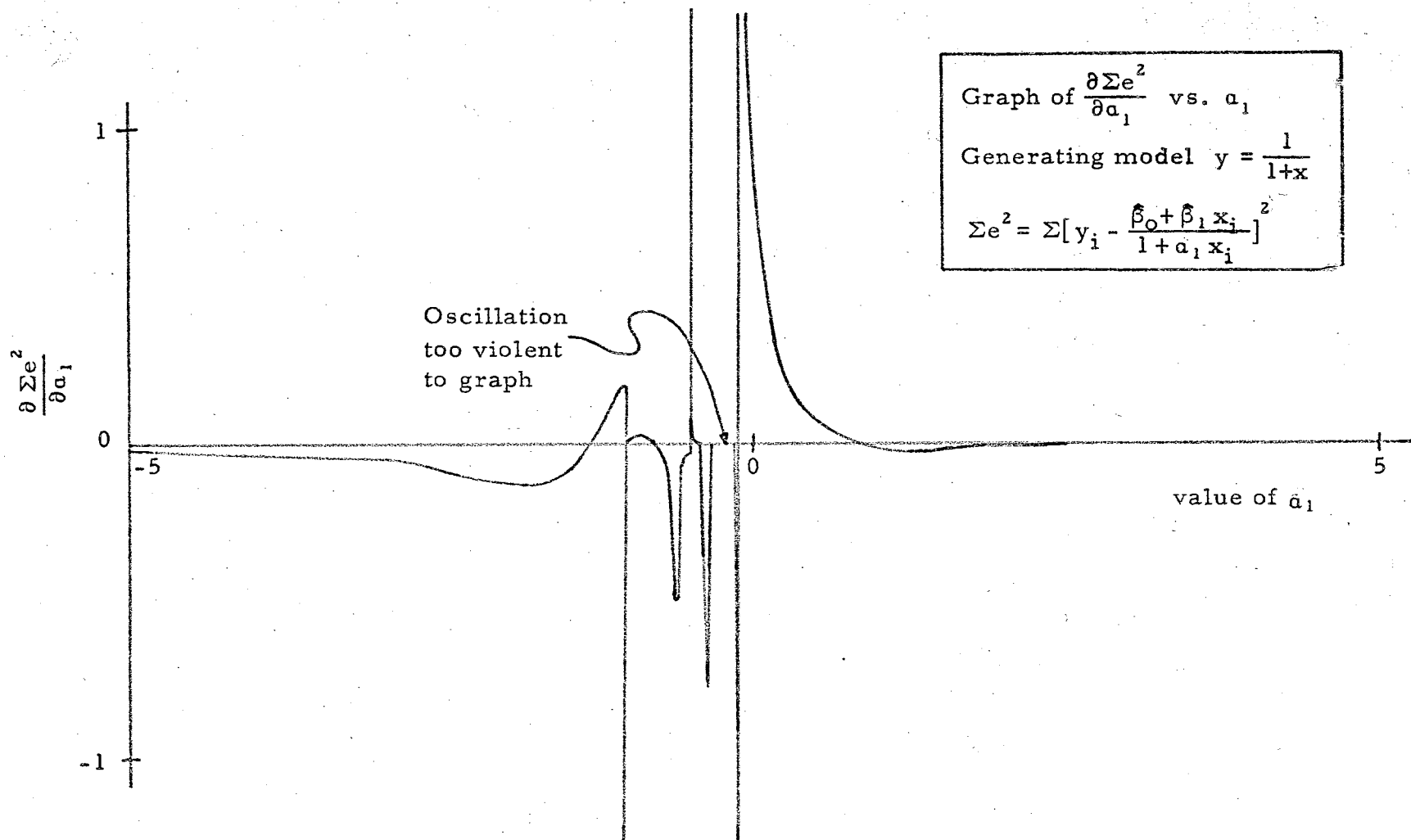


Fig. 5 Enlargement of Discontinuity Region of Fig. 4. (Note that magnification by 10 still leaves area of violent oscillations ungraphable.)

VITA

Paul Eugene Pulley, Jr.

Candidate for the Degree of

Doctor of Philosophy

Thesis: ON THE GOODNESS OF ESTIMATED PARAMETERS IN THE
RATIO MODEL

Major Field: Mathematics and Statistics

Biographical:

Personal Data: Born in Oklahoma City, Oklahoma, November 14, 1924, the son of Paul Eugene and Nelle R. Pulley.

Education: Attended grade school and junior high school in Oklahoma City; graduated from Classen High School in Oklahoma City in 1942; received the Bachelor of Science degree from the Oklahoma State University, with a major in mathematics, in May 1957; received the Master of Science degree from the Oklahoma State University, with a major in mathematics, in May, 1959; completed requirements for the Doctor of Philosophy degree in May, 1965.

Professional experience: Assisted in the initial installation of the Computing Center at the Oklahoma State University in 1956; worked as a graduate assistant in the Computing Center at the Oklahoma State University; did programming research for International Business Machines Corporation in New York City during the summer of 1957; helped set up an automation system for the Dallas Regional Office of the United States Department of Agriculture during the summer of 1958; served as Assistant Director of the Computation Center at the University of Kansas from June, 1959 to September, 1960; worked as staff assistant and then Assistant Director in the Computing Center at the Oklahoma State University from September, 1960 to September, 1964; has accepted an appointment as Assistant Director of the Computing Center and instructor in biostatistics at the University of Mississippi School of Medicine; is an associate member of Sigma Xi; member of Pi Mu Epsilon, Mathematical Association of America, and Society of Industrial and Applied Mathematics.