AN ABSTRACT OF THE THESIS OF

HUNG-FU CHIANG (Name) for the M. S. (Degree) in Statistics (Major)

Date thesis is presented June 10, 1965

Title ESTIMATION OF INTEGER-VALUED PARAMETERS

Abstract approved Redacted for Privacy
(Major professor)

This thesis is a study of the problem of calculating the least-squares estimates of the parameters in a linear regression model when the parameter space is restricted to integer points within a given polyhedron. General observations are made about estimation of integer-valued parameters, and an integer-quadratic programming algorithm is given for calculating the estimates. Consistency of this method of estimation is proved.

Two numerical examples are given to illustrate the method in the case of simple linear regression.

ESTIMATION OF INTEGER-VALUED PARAMETERS

by

HUNG-FU CHIANG

A THESIS

submitted to

OREGON STATE UNIVERSITY

in partial fulfillment of
the requirements for the
degree of

MASTER OF SCIENCE

June 1966

APPROVED:

Redacted for Privacy

Associate Professor of Statistics

In Charge of Major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented June 10, 1965

Typed by Carol Baker

ACKNOWLEDGMENTS

The author wishes to express his sincerest appreciation to Dr. Donald Guthrie, Jr. and to Dr. Donald R. Jensen for their assistance and guidance.

TABLE OF CONTENTS

LIST OF FIGURES

ESTIMATION OF INTEGER-VALUED PARAMETERS

I. INTRODUCTION

Background

This thesis presents a discussion of least-squares estimation in a linear regression model of parameters which are known *a-priori* to lie within a set of lattice points bounded by a convex polyhedral region formed by a system of linear inequalities. For purposes of discussion, the lattice is taken to be the points with integer-valued coordinates, but most of the methods and discussion will be valid for more general structures.

Hammersley (4) has discussed estimation of the integer-valued mean of a normal distribution with reference to estimating the molecular weight of insulin. There are several examples of problems arising in the physical sciences which might require the extension of his work to the full linear regression model.

By way of illustrating systems which involve integer-valued parameters, consider the following examples from the physical sciences.

Studies in crystallography utilize x-ray diffraction techniques and the relationship

$$n\lambda = 2d \sin \theta$$

where λ is the length of electromagnetic waves which strike the face of a crystal at an angle θ, and d is interplanar spacing in the crystal. Because of the phenomenon of interference in diffracted light, the parameter n is required to be an integer.

A related problem in wave mechanics is that of a vibrating body. Resonance is vibration of the body at one of its natural frequencies. In the case of a vibrating string, the natural frequencies are known to satisfy the relationship

$$f_n = n\sqrt{T/\mu}/2L$$

where f_n is a harmonic frequency of order n, n is an integer, T is tension of the system, μ is mass per unit length, and L is length of the string. In engineering applications of this subject to rigging, resonance of a particular system under a given load may prompt interest in determining order (n) of the harmonic at which resonance has occurred.

The phenomenon of absorption and emission of light of characteristic wavelengths in atomic species is based on a discrete number of orbits which may be occupied by electrons. In the case of hydrogen, parameters of the system are related by the equation

$$mvr = nh/2\pi$$

where m and v are mass and velocity of the electron, respectively, r is radius of the orbit, h is Planck's constant, and n is an integer called the quantum number.

Two principal works on the theory of estimation in such cases are especially pertinent to the present discussion. Hammersley (4) has considered the problem of estimating integer-valued parameters in a single-parameter family of distributions. Hocking (5) has discussed the least-squares estimation of regression parameters constrained to lie within a known convex region, but not restricted to lattice points.

In this thesis we explore only one principle of estimation -- the least-squares principle -- without claiming any optimal properties for the estimates so produced. In Hammersley's paper and the published discussion following its presentation several points are made concerning the single-parameter case which may tend to support the validity of least-squares estimation in the case of linear regression models involving several parameters.

Hammersley proposed to estimate the integer-valued mean of a normal distribution by the "rounded-off" value of the sample mean, that is, the integer nearest to the sample mean. In the single-parameter case this can easily be shown to be the least-squares estimate (by symmetry of the parabola defining the quadratic to be minimized), as well as the maximum likelihood estimate. In the single-parameter case, therefore, least squares estimates will possess properties in common with Hammersley's nearest-integer estimate.

Taking the usual first step in evaluating an estimation procedure, Hammersley determined an asymptotic expression for the variance of the nearest-integer estimate. In discussing the paper C. Stein observed that this estimate possesses a property which is almost that of minimizing the variance. The property is that the estimation procedure minimizes the maximum (with respect to the unknown mean) of the variance. Stein illustrates this as follows: Let T be any estimator, and let

$$\text{Var}\,(T) = f_T(\mu).$$

Then define

$$g(T) = \underset{-\infty<\mu<\infty}{\text{maximum}}\; f_T(\mu)\,.$$

For Hammersley's estimate the function $f_T(\mu)$ is constant, and $g(T)$ assumes a minimum value, thus making the estimate a "minimax variance" estimate. Stein also pointed out that there are estimators with smaller variance for certain values of μ, but none of which is uniformly better than Hammersley's estimate.

Another interesting feature of Hammersely's estimate was pointed out in the discussion by D. V. Lindley. If a "loss" function is defined to be zero when the correct integer parameter is chosen and unity when any other value is chosen, then the nearest-integer (least-squares) estimator is the estimator which minimizes the maximum expected loss.

One peculiar feature of such estimation problems is the fact that one may choose as his estimate exactly the true parameter point. In classical estimation in continuous parameter spaces this event happens with probability zero, and asymptotic results are stated in terms of ϵ-intervals about the true parameter. For the case in which the parameter is constrained to lattice points, however, the property of consistency, for example, is measured in terms of the convergence of the probability that the estimate exactly coincides with the parameter. Hammersley showed that this probability is of extremely small order in n, much smaller than in the case of estimation in the continuous parameter space.

Exploring the possibility of finding a sufficient estimate, Hammersely showed that no integer-valued sufficient estimate exists. In the discussion period, however, it was observed by C. A. B. Smith that while an integer-valued sufficient statistic may not exist, this does not invalidate the sufficiency of the sample mean and variance.

The above arguments then make a good case for the use of least-squares in the single-parameter problem. The purpose of this thesis is to investigate computational and other aspects of the least-squares estimation procedure for several parameters constrained to lattice points. In so doing, the author acknowledges existence of competing criteria which may be applied to estimation problems of this kind; however, a critical evaluation of alternative procedures would be the subject of another study.

Classical Least Squares Estimation

In the usual linear regression problem, the estimation of a p-coordinate vector-valued[1] parameter

$$\underline{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

is based on n scalar observations $Y_1, \cdots, Y_n$ which have means[2]

$$E(Y_i) = \sum_{j=1}^{p} \beta_j X_{ij}, \quad i = 1, 2, \cdots, n, \tag{1.1}$$

1 Throughout this thesis, whenever a symbol is to indicate a vector, that symbol will be underlined. All vectors are assumed to be column vectors. Matrices are represented by capital letters with dimensions indicated in the text.

2 The letter E is reserved exclusively to denote the mathematical expectation operator.

and variances σ^2. The constants X_{ij} are assumed to be fixed and known. In the following discussion they will form an $n \times p$ matrix X of rank $p < n$. The assumption that X is of "full" rank may be made without loss of generality; cases not of "full" rank may be reduced to the above case as discussed by Graybill (2). The value of σ^2 is assumed not to depend on the values of the constants X_{ij}, which frequently are called the design parameters.

The design parameters X_{ij} play the role of independent variables, whereas the Y_i are dependent variables. It is further assumed that the Y_i, which are subject to stochastic variations, may be written

$$Y_i = \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i, \quad i = 1, 2, \cdots, n \tag{1.2}$$

where the "errors" ϵ_i have mean zero, variances σ^2, and are uncorrelated. In order to conveniently test statistical hypotheses and form confidence regions it is frequently assumed that the ϵ_i are normally distributed.

The least squares method of estimation of the parameter $\underline{\beta}$ is based on the minimizing of the quadratic form

$$Q(\underline{\beta}) = (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}) \tag{1.3}$$

with respect to $\underline{\beta}$. Elementary calculus gives as the minimizing value of $\underline{\beta}$ the least squares estimate

$$\hat{\underline{\beta}} = (X'X)^{-1}X'\underline{Y} = S^{-1}X'\underline{Y} \tag{1.4}$$

where

$$S = X'X .$$

In the following S will be known as the design matrix. S is symmetric, $p \times p$, non-singular, and positive definite.

If the ϵ_i are normally distributed, then the least squares estimate $\hat{\underline{\beta}}$ has a multivariate normal distribution with mean $\underline{\beta}$ and covariance matrix

$$\Sigma_{\hat{\beta}} = \sigma^2 S^{-1} .$$

For the details of the above discussion, the reader is referred to Graybill (2).

Restricted Parameter Spaces

In his Ph. D. dissertation (5), Hocking considered the least-squares estimation of regression parameters which are known to be within a convex subset of p-dimensional Euclidean space. He considered the evaluation of the estimates, tests of statistical hypotheses about them, and the determination of confidence sets for them.

In this thesis, the parameter space is considered to be restricted to a convex polyhedral subset of p-space obtained by imposing the m linear inequalities

$$A\underline{\beta} \leq \underline{C} \tag{1.5}$$

where A is an $m \times p$ matrix of rank $m < p$. The introduction of these m linear constraints complicates the determination of the least squares estimates considerably. Should the unrestricted minimizing value of $\underline{\beta}$ of $Q(\underline{\beta})$ satisfy the constraints, it would also be the restricted minimizing value, but if $\underline{\beta}$ does not satisfy (1.5), then procedures involved more than differential calculus are necessary. One known method of finding the restricted minimum of $Q(\underline{\beta})$ is the quadratic programming algorithm of Wolfe (10) discussed in detail also by Hadley (3).

As an additional restriction on the parameter space, suppose that the only reasonable values for $\underline{\beta}$ are those where all coordinates are integers. This further complicates the location of the minimum of $Q(\underline{\beta})$, since the only values considered must satisfy both the linear inequalities and have integer-valued coordinates.

Two numerical examples are presented for $p = 2$ in Part IV. In these examples, the restricted parameter set is shown graphically.

The basic problem of this thesis is, therefore, the determination of least squares estimates of regression parameters known to be integers and to satisfy certain linear inequalities. (By setting $m = 0$ one can study the problem of estimating otherwise unconstrained integer-valued parameters).

II. DETERMINATION OF LEAST-SQUARES ESTIMATES

In the following we discuss the location of the minimum of the quadratic form $Q(\underline{\beta})$, defined above, over the restricted parameter set where the coordinates of $\underline{\beta}$ must be integers and must satisfy the linear inequalities $A\underline{\beta} \leq \underline{C}$.

One naive method would be to find the unrestricted least-squares estimate $\hat{\underline{\beta}}$ and round each coordinate $\hat{\beta}_j$ to its nearest integer. This method has two drawbacks: (1) while $\hat{\underline{\beta}}$ may satisfy the linear constraints, the estimate so constructed may not; and, more seriously, (2) the estimates so constructed may not actually be the least-squares integer, as evidenced by the numerical examples in Part IV.

As an alternative we propose the integer-quadratic programming technique of Kunzi and Oettli (8). This method combines the usual quadratic programming algorithms with a "cutting plane" approach similar to that of Gomory (1) for linear programs. For purposes of the discussion of this method, we rewrite

$$Q(\underline{\beta}) = \underline{Y}'\underline{Y} - 2\underline{Y}'X\underline{\beta} + \underline{\beta}'S\underline{\beta} \qquad (2.1)$$

since $\underline{Y}'\underline{Y}$ does not involve $\underline{\beta}$, by letting $\underline{T} = 2X'\underline{Y}$ we may equivalently consider the minimization of

$$Q^*(\underline{\beta}) = \underline{\beta}'S\underline{\beta} - T'\underline{\beta} . \qquad (2.2)$$

Before explicitly giving the algorithm, let us consider a geometrical description for the case $p = 2$ as illustrated in Figure 1. Suppose the unrestricted minimum of $Q^*(\underline{\beta})$ is outside the admissible domain. Since S is positive definite, $Q^*(\underline{\beta})$ is strictly convex and its horizontal cross-sections are ellipses. Let $\hat{\underline{\beta}}$ be the unconstrained minimum (the least squares solution). All the ellipses have their centers at $\hat{\underline{\beta}}$. Expand one of these ellipses, say $Q^*(\underline{\beta}^{(1)})$, from the center $\hat{\underline{\beta}}$ to the first point with integral coordinates which also satisfies the linear constraints.

The iterative procedure consists of expanding, at each step, the ellipse and its circumscribed polyhedron p^k (Figure 1) until this polyhedron meets a feasible parameter point with integral coordinates; this point is identified by the symbol $\underline{\beta}^{(k+1)}$. Let λ be the parameter of expansion; the polyhedron $p^{(k)}(\lambda)$ circumscribes the ellipsoid $Q^*(\lambda)$. Extreme points of p^k may be removed from consideration by cutting the circumscribing polygon p^k with a hyperplane through the point $\underline{\beta}^{k+1}$ and tangent to $Q^*(\lambda)$ at some point $\tilde{\underline{\beta}}^{(k+1)}$. The result is a new polyhedron with $(k+1)$ faces. If, for $\lambda = \lambda^{(k+1)}$, the point $\underline{\beta}^{(k+1)}$ with integral coordinates is exactly at the point $p^k(\lambda)$, then $\underline{\beta}^{(k+1)}$ is outside of $p^{(k+1)}(\lambda^{(k+1)})$ and the expansion begins again. The optimal point must belong to the polyhedron and to the ellipsoid. The initial polyhedron is merely a half-space, having a hyperplane tangent to the ellipsoid as its only

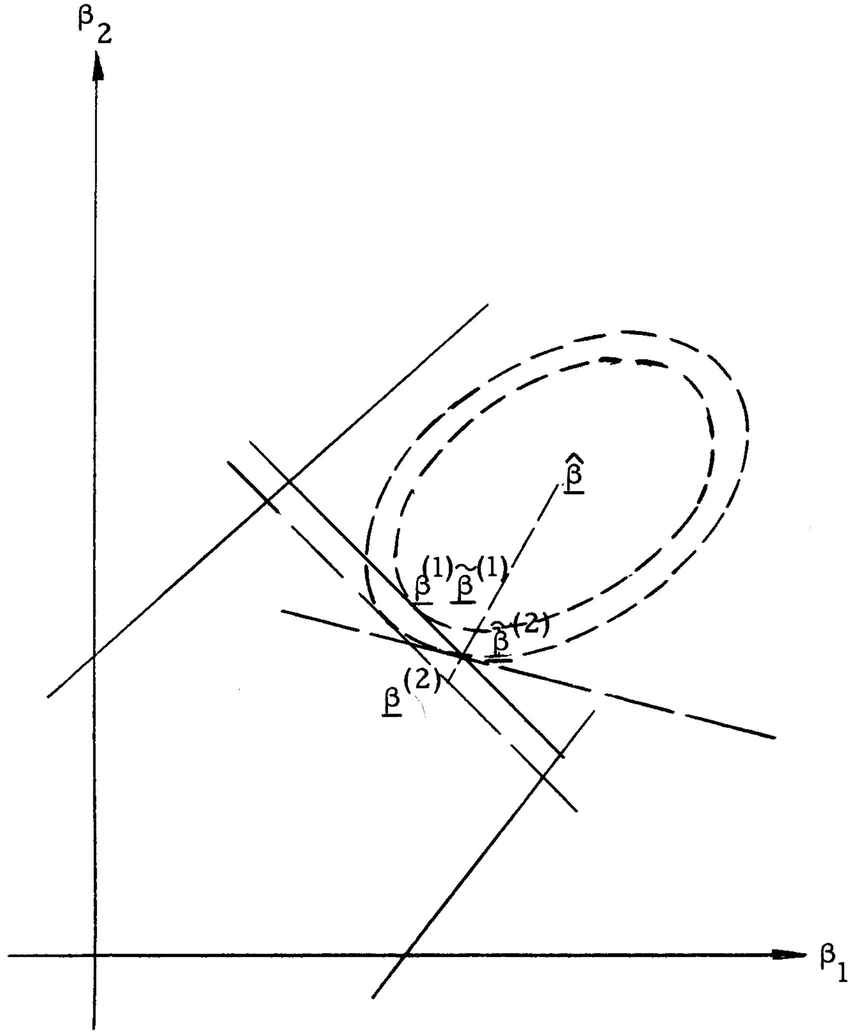

Figure 1. Illustration for the case $p = 2$.

surface.

The complete algorithm involves the following steps:

Step 1. Determine the unrestricted minimum $\underline{\hat{\beta}}$ by the method of least-squares. If $\underline{\hat{\beta}}$ is an integer, the solution is complete. If not, go to:

Step 2. Solve for $\underline{\beta}^{(1)}$, the solution of the quadratic program not restricted to be integers. If $\underline{\beta}^{(1)}$ is an integer, the problem is finished and $\underline{\beta}^{(1)}$ is the optimum solution. If not, proceed to:

Step 3. A set of integral values $\underline{\beta}^{(k)}$ for $k \geq 2$ can be determined by iteration, according to the recurrence rule:

At the beginning, set $\underline{\tilde{\beta}}^{(1)} = \underline{\beta}^{(1)}$; then with $\underline{\beta}^{(k-1)}$ and $\underline{\tilde{\beta}}^{(k-1)}$ (note that $\underline{\beta}^{(0)} = \underline{\hat{\beta}}$), we have to solve the mixed linear programming problem with integer coefficients.

Minimize λ (a scalar unrestricted in sign), (2. 3)

subject to the conditions:

$$A\underline{\beta} \leq \underline{C} \qquad (2.4)$$

$$t^{j'}\underline{\beta} - a_j\lambda \leq b_j \qquad (j = 1, 2, \cdots, k-1) \qquad (2.5)$$

where

$$t^j = -T + 2S\underline{\tilde{\beta}}^{(j)} \qquad \text{(p x 1 vector)}$$

$$a_j = t^{j'}(\underline{\tilde{\beta}}^{(j)} - \underline{\hat{\beta}}) \qquad \text{(scalar)}$$

$$b_j = t^{j'}\underline{\hat{\beta}} \qquad \text{(scalar)}$$

This problem may be transformed into the successive solution of linear programs purely in integers. Let $(\underline{\beta}^{(k)}, \lambda^{(k)})$ be the solution of this program (a system of p+1 values). Then set

$$\underline{\tilde{\beta}}^{(k)} = \underline{\hat{\beta}} + \mu_k(\underline{\beta}^{(k)} - \underline{\hat{\beta}}) \qquad (2.6)$$

where

$$\mu_k = \{[Q^*(\underline{\beta}^{(1)}) - Q^*(\underline{\hat{\beta}})]/[Q^*(\underline{\beta}^{(k)}) - Q^*(\underline{\hat{\beta}})]\}^{1/2}$$

and begin a new cycle.

<u>Step 4.</u> If $\underline{\beta}^{(m)} = \underline{\beta}^{(k)}$, for $k \leq m-1$, the iterations stop, and $\underline{\beta}^{(m)}$ is the optimal solution.

Since Step 3 is an all-integer programming problem, Gomory's (1) requirement that all the coefficients appearing in the linear program be rational must be satisfied. If the quantity μ_k is irrational, a rational approximation would have to be formed. In the case where the unrestricted solution satisfies the linear constraints, the algorithm will not start with a half-space but with a polyhedron determined by suitably chosen surfaces.

III. DISTRIBUTION OF THE LEAST SQUARES ESTIMATE

In any estimation problem the criteria for choosing any particular estimator are based on the sampling distribution of the various possible estimates. It is well known (see reference (2)) that the usual least squares estimates in the unrestricted parameter space have the desirable property of being minimum variance unbiased estimates. Finding the exact sampling distributions of estimates restricted to lattice points, however, is so involved that there appears little hope of finding explicit closed expressions for such distributions. Hammersley observed, however, that consistency of integer-valued estimates, at least in the one-dimensional case is very easy to verify; he further noted the consistency to be of special type -- instead of considering the probability of being within some small neighborhood of the true parameter point, one may consider the probability of achieving the parameter point exactly. Hammersley was unable in the one-dimensional case to find small-sample distributions of his estimates; the task appears to be insurmountable in higher dimensions.

Large sample consistency of the least-squares estimate is, however, quite easy to prove. Let $\hat{\hat{\underline{\beta}}}$ be the restricted least squares estimate obtained by the method of the previous section, and consider the following theorem.

Theorem: Let the sample size n be increased in such a way that for some $p \times p$ positive definite matrix S_o, the equality

$$\frac{S}{n} = S_o , \tag{3.1}$$

holds for all n, where S is the design matrix. Then

$$P(\hat{\hat{\underline{\beta}}} = \underline{\beta}) \rightarrow 1 \quad \text{as} \quad n \rightarrow \infty. \tag{3.2}$$

Before proving the theorem let us make the observation that the condition imposed on S is merely one to preserve its full rank. While it would be possible to have the matrix S be of full rank for every finite n and to have the limiting matrix value of (S/n) be of less than full rank, a more reasonable requirement which realizes the full potential of the observations at each successive stage is that given. Furthermore, as observed during the course of the proof, all that is necessary is that $|S/n|$ be bounded away from zero. In the case of a simple linear regression where $p = 2$, $X_{i1} = 1$, and $X_{i2} = X_i$,

$$\frac{S}{n} = \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{\Sigma X_i^2}{n} \end{bmatrix}$$

The condition on S merely has the effect of bounding $\Sigma X_i^2/n$

away from zero and infinity.

Proof: Let $\underline{\beta}^*$ be any point in the restricted parameter space with $\underline{\beta}^* \neq \underline{\beta}$. Consider the random variable

$$\begin{aligned} W_n &= \frac{1}{n}[Q(\underline{\beta}^*) - Q(\underline{\beta})] \\ &= \frac{2}{n}\underline{Y}'X(\underline{\beta} - \underline{\beta}^*) + \underline{\beta}^{*\prime} S_o \underline{\beta}^* - \underline{\beta}' S_o \underline{\beta} \end{aligned} \qquad (3.3)$$

Note that W is a linear function of $\underline{Y}$ and since

$$E(\underline{Y}) = X\underline{\beta}, \qquad (3.4)$$

then

$$\begin{aligned} E(W_n) &= \frac{2}{n}\underline{\beta}' X' X(\underline{\beta} - \underline{\beta}^*) + \underline{\beta}^* S_o \underline{\beta}^* - \underline{\beta}' S_o \underline{\beta} \\ &= 2\underline{\beta}' S_o(\underline{\beta} - \underline{\beta}^*) + \underline{\beta}^{*\prime} S_o \underline{\beta}^* - \underline{\beta}' S_o \underline{\beta} \\ &= (\underline{\beta} - \underline{\beta}^*)' S_o (\underline{\beta} - \underline{\beta}^*) > \underline{0} \end{aligned} \qquad (3.5)$$

since S_o is positive definite. Furthermore, since the covariance matrix of $\underline{Y}$ is $\sigma^2 I$,

$$\begin{aligned} Var(W_n) &= (\frac{2}{n})^2 (\underline{\beta} - \underline{\beta}^*)' \ X' I X(\underline{\beta} - \underline{\beta}^*) \\ &= \frac{4}{n}[(\underline{\beta} - \underline{\beta}^*)' S_o (\underline{\beta} - \underline{\beta}^*)] \end{aligned} \qquad (3.6)$$

and

$$\mathrm{Var}(W_n) \to 0 \quad \text{as} \quad n \to \infty.$$

Using Tchebysheff's inequality, we see that $Q(\underline{\beta}^*) - Q(\underline{\beta})$ tends in probability to a positive number, therefore

$$P(Q(\underline{\beta}^*) > Q(\underline{\beta})) \to 1 \tag{3.7}$$

for any $\underline{\beta}^* \neq \underline{\beta}$. Since $\hat{\hat{\underline{\beta}}}$ is the point in the restricted parameter space which minimizes $Q(\underline{\beta})$,

$$P(\hat{\hat{\underline{\beta}}} = \underline{\beta}^*) \to 0 \tag{3.8}$$

for all $\underline{\beta}^* \neq \underline{\beta}$ and hence

$$P(\hat{\hat{\underline{\beta}}} = \underline{\beta}) \to 1. \tag{3.9}$$

The theorem is proved, and consistency of the restricted least squares estimate is established.

In order to fully investigate other large and small sample properties of the least-squares and other estimation procedures, further analysis is required beyond the scope of this thesis. In the small sample case, one appropriate method would seem to be the Monte Carlo simulation of the process, comparing the accuracy (by some standard) of this procedure to others. Such an analysis would, of course, require formulating other criteria and approaches, a large task in itself. The complexity of the mathematical analysis required is clearly of high degree, but this author hopes to pursue the problem more fully at some later date.

IV. NUMERICAL EXAMPLES

Example 1. In order to illustrate the estimating method presented above, let us consider the following artificial example for the case $p = 2$. The objective is to find integer-valued estimates of $\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ subject to the constraints

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \underline{\beta} \leq \begin{bmatrix} 5 \\ 3 \end{bmatrix}$$

that is,

$$\beta_1 + \beta_2 \leq 5$$
$$\beta_2 \leq 3$$

Figure 2 shows these constraints and the admissible parameter points.

Suppose that

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad Y = \begin{bmatrix} 2 \\ 3 \\ 8 \\ 10 \\ 15 \\ 5 \\ 11 \\ 8 \\ 2 \\ 5 \end{bmatrix}$$

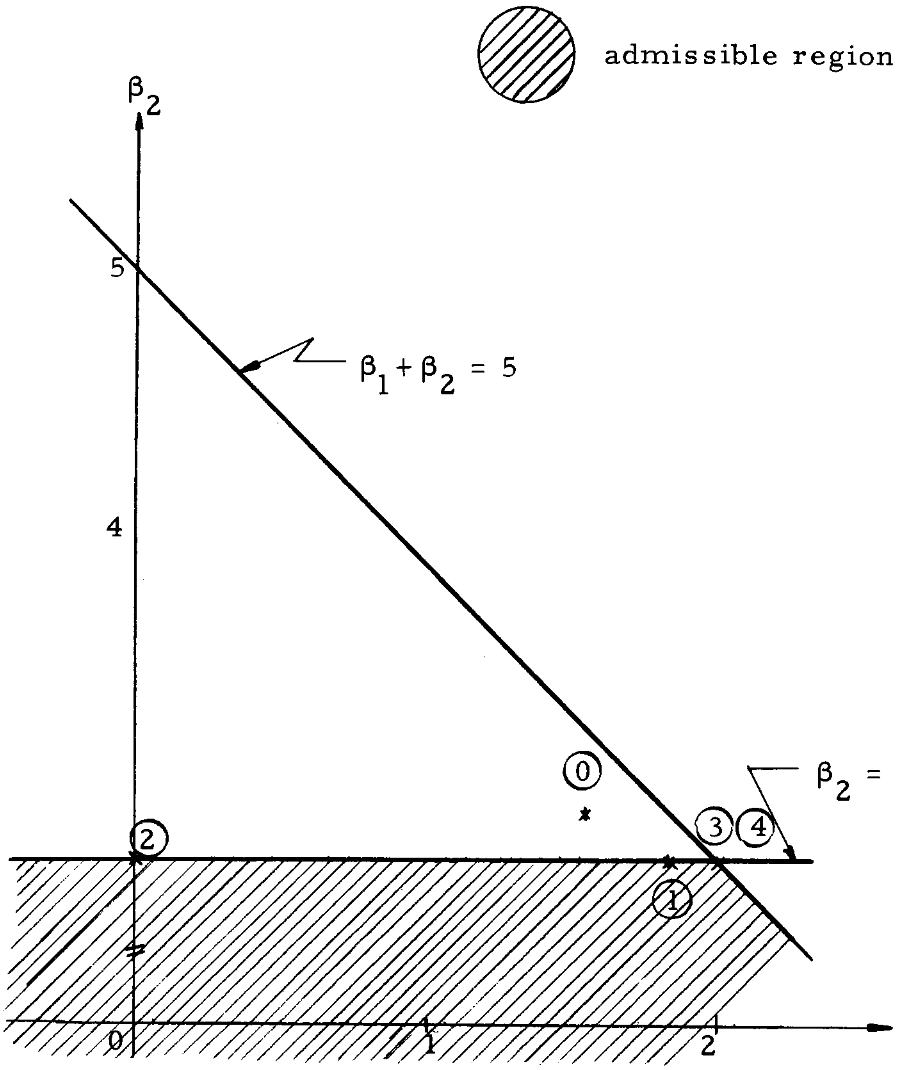

Figure 2. Admissible points of Example 1.

whereupon

$$S = \begin{bmatrix} 10 & 17 \\ 17 & 45 \end{bmatrix}, \quad S^{-1} = \frac{1}{161} \begin{bmatrix} 45 & -17 \\ -17 & 10 \end{bmatrix},$$

$$X'\underline{Y} = \begin{bmatrix} 69 \\ 168 \end{bmatrix}$$

We therefore must minimize

$$Q(\underline{\beta}) = 641 - [138 - 336] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + [\beta_1 \quad \beta_2] \begin{bmatrix} 10 & 17 \\ 17 & 45 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

Solving for the unrestricted least-squares estimator, we find

$$\hat{\underline{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1.547 \\ 3.149 \end{bmatrix}$$

and $Q(\underline{\beta}) = 4.927$

This is the point ⓪ in Figure 2.

After three iterations using Wolfe's method for quadratic programming [Hadley (3, p. 221-230)], we obtain as the coordinates which minimize the function, subject to the linear but not the integral constraints, the point

$$\underline{\beta}^{(1)} = \begin{bmatrix} \beta_1^{(1)} \\ \beta_2^{(1)} \end{bmatrix} = \begin{bmatrix} 1.807 \\ 3.000 \end{bmatrix}$$

for which $Q(\underline{\beta}^{(1)}) = 5.6$. This is the point ① in Figure 2.

By solving a sequence of integer programs, we now find the restricted minimum of $Q(\underline{\beta})$. To begin, we set the equalities $\underline{\tilde{\beta}}^{(1)} = \underline{\beta}^{(1)}$ and $\underline{\beta}^{0} = \underline{\beta}$, whence

$$t^{(1)} = \begin{pmatrix} 0.140 \\ -4.562 \end{pmatrix}$$

$$a_1 = 0.716$$

$$b_1 = -14.149$$

and our first integer program becomes;

$$\text{Minimize} \quad \lambda^{(1)} = \lambda_1^{(1)} - \lambda_2^{(1)} \qquad \lambda_1^{(1)}, \lambda_2^{(1)} \geq 0$$

$$\text{Subject to} \quad \beta_1 + \beta_2 \leq 5$$

$$\beta_2 \leq 3$$

$$0.196\beta_1 - 6.372\beta_2 - (\lambda_1^{(1)} - \lambda_2^{(1)}) \leq -19.761 \quad \text{and}$$

β_1, β_2 are integers.

After two iterations, the above procedure yields

$$\lambda^{(1)} = 0.645$$

$$\underline{\beta}^{(2)} = \begin{pmatrix} \beta_1^{(2)} \\ \beta_2^{(2)} \end{pmatrix} = \begin{pmatrix} 0 \\ 3 \end{pmatrix}$$

and $Q(\underline{\beta}^{(2)}) = 36$. This is the point ② in Figure 2.

The multiplier μ_3 is then found to be 0.147 and

$$\underline{\tilde{\beta}}^{(3)} = \begin{bmatrix} 1.320 \\ 3.127 \end{bmatrix}$$

which leads us to the following

$$\text{Minimize} \quad \lambda^{(2)} = \lambda_1^{(2)} - \lambda_2^{(2)}$$

$$\text{Subject to} \quad \beta_1 + \beta_2 \leq 5$$

$$\beta_2 \leq 3$$

$$-3.741\,\beta_1 - 6.863\,\beta_2 - (\lambda_1^{(2)} - \lambda_2^{(2)}) \leq 27.398 \quad \text{and}$$

β_1, β_2 are integers.

Solution of this minimization problem gives the results

$$\lambda^{(2)} = -0.673$$

$$\underline{\beta}^{(3)} = \begin{bmatrix} \beta_1^{(3)} \\ \beta_2^{(3)} \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

and $Q(\underline{\beta}^{(3)}) = 6$. The solution $\underline{\beta}^{(3)}$ is the point ③ in Figure 2.

Again, by calculation we have $\mu_4 = 0.792$ and

$$\underline{\tilde{\beta}}^{(4)} = \begin{bmatrix} 1.906 \\ 3.044 \end{bmatrix}$$

The new integer program is:

$$\text{Minimize} \quad \lambda^{(3)} = \lambda_1^{(3)} - \lambda_2^{(3)}$$

$$\text{Subject to} \quad \beta_1 + \beta_2 \leq 5$$

$$\beta_2 \leq 3$$

$$3.796\beta_1 + 8.764\beta_2 - 0.341(\lambda_1^{(3)} - \lambda^{(2)}) \leq 33.006 \quad \text{and}$$

$\beta_1, \ \beta_2$ are integers.

The solution of the new integer program is

$$\underline{\beta}^{(4)} = \begin{pmatrix} \beta_1^{(4)} \\ \beta_2^{(4)} \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

which is seen to be the same as $\underline{\beta}^{(3)}$. Thus the linearly constrained integer-valued estimate of $\underline{\beta}^{(4)}$; that is

$$\hat{\hat{\underline{\beta}}} = \underline{\beta}^{(4)} = \begin{pmatrix} \beta_1^{(4)} \\ \beta_2^{(4)} \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

and $Q(\underline{\beta})$ is equal to 6.

Example 2. In Example 1, we have made a restriction $\beta_3 \leq 3$. In that example, it was not necessary to use integer programming. In the example we use the data presented in Example 1, but modify the constraints to be

$$\beta_1 + \beta_2 \leq 5$$

$$\beta_2 \leq 2.5$$

$$\beta_1 \geq 0$$

$$\beta_2 \geq 0$$

and β_1, β_2 are integers.

Note that $\underline{\beta}$ in a regression model is not necessarily non-negative. Because any number may be written as the difference of two non-negative numbers, the restriction $\underline{\beta} \geq \underline{0}$ may always be imposed, provided the numbers of parameters is increased accordingly. Wolfe's method gives the solution of this quadratic program, after four iterations, as

$$\underline{\beta}_1^{(1)} = \begin{pmatrix} \beta_1^{(1)} \\ \beta_2^{(2)} \end{pmatrix} = \begin{pmatrix} 2.5 \\ 2.5 \end{pmatrix}$$

and $Q(\underline{\beta}^{(1)}) = 12.25$ from which we find

$$t^{(1)} = \begin{pmatrix} -3 \\ -26 \end{pmatrix}$$

$$a_1 = 14.015$$

$$b_1 = -86.515$$

Therefore, the next problem is to solve the following:

Minimize $\lambda = \lambda_1 - \lambda_2$

Subject to $\beta_1 + \beta_2 \leq 5$

$$\beta_2 \leq 2.5$$

$$-3\beta_1 - 26\beta_2 - 14.015(\lambda_1 - \lambda_2) \leq -86.515$$

$$\beta_1, \beta_2, \lambda_1, \lambda_2 \geq 0$$

and β_1, β_2 must be integers.

Using the Simplex method for linear programming we obtain

$$\beta_1 = 2.5; \quad \beta_2 = 2.5 \quad \text{and} \quad \lambda_1 = 1 .$$

Since β_1 and β_2 both appeared in the basic set, they must be integers, but the slack variables β_3, β_4 (which are introduced to make the inequality constraints as equalities ones) are not necessarily integers. Therefore this problem belongs to the mixed integer-continuous variable class. According to Gomory's algorithm [Hadley (3, p. 282-285)] we introduce

$$\beta_3 + \beta_4 \geq 0.5$$

as the first cut which gives

$$\beta_1 = 2; \quad \beta_2 = 2.5 \quad \text{and} \quad \beta_3 = 0.5.$$

Since β_2 is not integral, another cut should be introduced which is found to be

$$\beta_4 \geq 0.5.$$

Then we find

$$\underline{\beta}^{(2)} = \begin{bmatrix} \beta_1^{(2)} \\ \beta_2^{(2)} \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

which gives $Q(\underline{\beta}^{(2)}) = 29.$

Again, as in Example 1, we calculate

$$\mu_2 = 0.552$$

$$\tilde{\underline{\beta}}^{(2)} = \begin{bmatrix} 2.349 \\ 2.515 \end{bmatrix}$$

and

$$t^{(2)} = \begin{bmatrix} -5.510 \\ -29.784 \end{bmatrix}$$

$$a_2 = 14.464$$

$$b_2 = -102.324$$

By the same procedure we obtain

$$\underline{\beta}^{(3)} = \begin{bmatrix} \beta_1^{(3)} \\ \beta_2^{(3)} \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

Hence $\underline{\beta}^{(3)} = \underline{\beta}^{(2)}$ so that

$$\hat{\hat{\underline{\beta}}} = \underline{\beta}^{(3)} = \begin{bmatrix} \beta_1^{(3)} \\ \beta_2^{(3)} \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$$

will be the restricted least-squares estimate.

Figure 3 provides a sketch of Example 2. Point ⓪ is, as before, the unrestricted least-squares estimate. Point ① is the least-squares estimate where $\underline{\beta}$ is restricted by the linear inequalities. Point ③ is the integer least-squares estimate. The ellipse sketched in Figure 3 is the cross section of

$$Q(\underline{\beta}) = Q(\hat{\hat{\underline{\beta}}})$$

to show the "shape" of the quadratic form $Q(\underline{\beta})$.

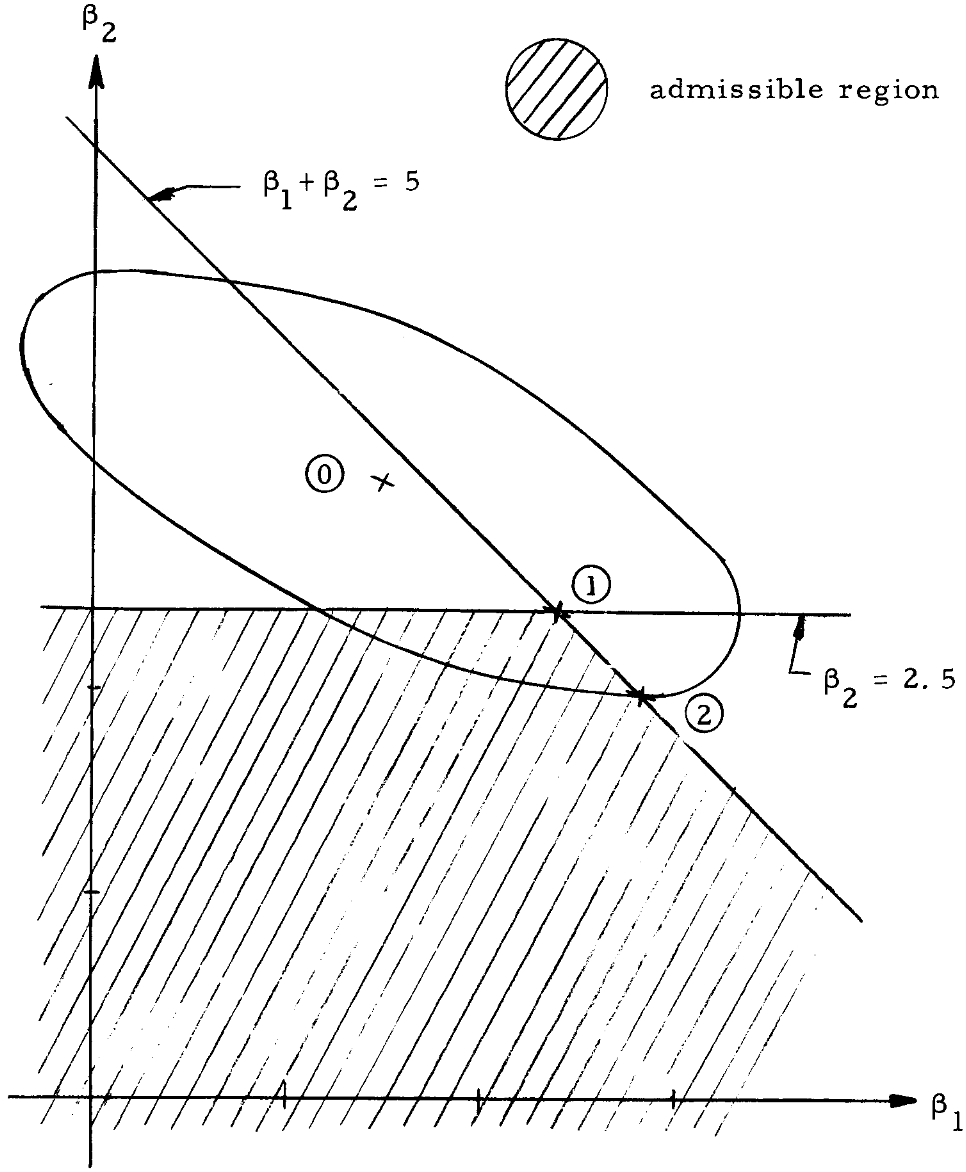

Figure 3. Admissible points of Example 2.

BIBLIOGRAPHY

1. Gomory, R. E. Outline of an algorithm for integer solution to linear programs. Bulletin of the American Mathematical Society 64:275-278. 1958.

2. Graybill, F. A. An introduction to linear statistical models. Vol. I. New York, McGraw-Hill, 1961. 463 p.

3. Hadley, G. Nonlinear and dynamic programming. Reading, Addison-Wesley, 1964. 484 p.

4. Hammersley, J. M. On estimating restricted parameters. Journal of the Royal Statistical Society, Ser. B, 12:192-240. 1950.

5. Hocking, R. R. Mathematical programming in statistical estimation theory. Ph. D. thesis. Ames; Iowa State University, 1962. 125 p.

6. Jensen, D. R. Linear hypothesis with discrete parameters: simple linear regression, bounds on least-squares estimators. Corvallis, Oregon State University, Department of Statistics, 1964. (Mimeographed)

7. Kuhn, H. W. and A. W. Tucker. Nonlinear programming. In: Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950. Berkeley, University of California Press, 1950. P. 481-492.

8. Kunzi, H. P. and W. Oettli. Une methode de resolution de programmes quadratiques en nombres entiers sous des liasions lineaires et pour des fonctions objectives strictement convexes. Academie des Sciences, Paris. Comptes Rendus 252:1415-1417. 1961.

9. Mood, A. M. and F. A. Graybill. Introduction to the theory of statistics. 2d ed. New York, McGraw-Hill, 1963. 443 p.

10. Wolfe, P. Recent developments in nonlinear programming. Santa Monica, Rand Corp., 1962. 40 p. (Rand Report, R-401-PR.)