1963

Linear estimation in convex parameter spaces

William Thomas Lewish

Iowa State University
This dissertation has been 64-3961 microfilmed exactly as received

LEWISH, William Thomas, 1926—
LINEAR ESTIMATION IN CONVEX PARAMETER SPACES.

Iowa State University of Science and Technology
Ph.D., 1963
Mathematics

University Microfilms, Inc., Ann Arbor, Michigan
LINEAR ESTIMATION IN CONVEX PARAMETER SPACES

by

William Thomas Lewish

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

Major Subject: Statistics

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1963
# TABLE OF CONTENTS

## I. INTRODUCTION
- The Problem 1
- Historical Background 4
- Review of Some Linear and Quadratic Programming Procedures 7

## II. PROPERTIES OF THE QUADRATIC PROGRAMMING ESTIMATOR 15
- The "Minimum Distance" Property of $\beta$ 15
- A "Comparison" of $\beta$ and $\hat{\beta}$ 19
- Spherical Confidence Region 22
- Asymptotic Properties 22
- Confidence Regions for $\beta_0$ 25
- Generalized Least Squares 26

## III. ASYMPTOTIC BIAS AND MEAN SQUARE ERROR 27
- Two Parallel Restricting Planes 27
- Two Non-Parallel Restricting Planes 35

## IV. LINEAR COMBINATIONS OF $\beta$ 59
- An Alternative Estimator 59
- MSE of a Simple Linear Combination 60

## V. A NUMERICAL EXAMPLE 65
- Transformation and Projection 66
- Quadratic Programming Solution 73
- The Asymptotic Bias and the Variance 80
- Estimation of the Variance 81
VI. FITTING POLYNOMIALS WITH NON-NEGATIVE DERIVATIVES 85
   A. Special Case of the Cubic 86
   B. Approximate Solution for any Polynomial 88
   C. The Infinite Range Case of Odd Order Polynomials 103
   D. Application of Restrained Least Squares to Frequency Graduation 104

VII. ACKNOWLEDGMENT 116

VIII. LITERATURE CITED 117
I. INTRODUCTION

A. The Problem

In this thesis we are concerned with the estimation of parameters in a linear model "under constraints". More specifically we consider the classical linear model

\[ y = X\beta + \epsilon \]

whose components are specified in B below. The classical Least Squares estimation of the elements \( \beta_i \) of the parameter vector \( \beta \) in the above model covers both, the case when the \( \beta_i \) are "free" or "not constrained" and the case when the \( \beta_i \) are known to satisfy linear equations. In contrast we consider in this thesis the problem of estimating the parameter point \( \beta \) (with coordinates \( \beta_i \)) when \( \beta \) is known to lie in a convex region of the \( \beta \)-space. We begin by illustrating the estimation problem by describing an example of its occurrence.

A chemical mixture is known to be comprised of a number of \( p \) "ingredients" \( j = 1, 2, \ldots, p \) which are mixed in unknown proportions. A chemical analysis for the mixture reveals a number, \( n \), of chemical "characteristics" \( y_i \) \( (i = 1, 2, \ldots, n) \) for the mixture. It is required to estimate the unknown weight proportions \( \beta_j \) \( (j = 1, 2, \ldots, p) \) with which the \( p \) ingredients occur in the mixture. Denote by \( x_{ij} \) the value of the \( i \)-th characteristic if a unit weight of the \( j \)-th ingredient is subjected to the analysis. Then,
assuming that the ingredients contribution to the mixture are additive, the observed value of the i-th characteristic in the mixture will be given by the linear model

\[ y_i = \sum_{j=1}^{P} \beta_j x_{ij} + e_i \]  

(1)

It should be noted that the error term \( e_i \) is here considered the only error. Actually the model should more likely be

\[ y_i = \sum_{j=1}^{P} \beta_j X_{ij} + e_i \]

(2)

where \( X_{ij} \) is the true amount of the i-th characteristic in the j-th ingredient. Thus \( x_{ij} = X_{ij} + E_{ij} \), that is both dependent and independent variables are subject to error. See [15] for a summary of this situation. All the results presented in this thesis are based on the assumption that \( E_{ij} = 0 \).

We further assume that the scales in which the \( y_i \) are measured are standardized and that the observed values of \( y_i \) differ from their expected values by independent residuals \( e_i \) of equal variance. Under such circumstances we have a standard regression model and one could estimate the \( \beta_j \) by Least Squares i.e. by minimizing

\[ Q(\beta) = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{P} \beta_j x_{ij})^2 = (y - X\beta)'(y - X\beta) \]

(3)
However, the estimated regression coefficients ($\hat{\beta}_j$ say) computed by the classical Least Squares estimation i.e. from the "normal" regression equations obtained by differentiation of (3), would take no account of our a priori knowledge that the proportions are not negative i.e. that

$$\beta_j \geq 0 \quad j = 1, 2, \ldots, p$$

(4)

Further, since the $\beta_j$ are weight proportions we have for the sum of proportions

$$\sum_{j=1}^{p} \beta_j \leq 1$$

(5)

The inequality sign taking account of the fact that the mixture may contain an unknown proportion of "inert" ingredient not contributing to any of the characteristics.

If we take account of the inequalities (4) and (5) the Least Squares principle is, in many situations, still an appropriate method for the computation of the estimates of the $\beta_j$. This leads to the problem of finding the minimum of the quadratic form $Q(\beta)$ given by (3) but restricting the "parameter point" to the tetrahedral section of the $\beta_j$ space defined by (4) and (5). This is a problem in "quadratic programming" in which the positive definite objective function is the Least Squares quadratic form which is to be minimized and in which the "restrictions" (4) and (5) define a convex
space bounded by planes. This is precisely the special situation for which numerous methods of "quadratic programming" have been developed. Some of these will be described in Section C below.

The determination of the composition of a mixture emitting gamma rays is another application of our problem. In [17] several methods of estimating the unknown proportions of the components in the mixture, including, of course, the method of Least Squares are presented and evaluated. In the framework of this application it was stated that estimated negative proportions implies an incomplete model or inaccurate measurements. By this argument, the problem of estimation in a constrained parameter space is bypassed.

B. Historical Background

Consider the general linear hypothesis [12, 13]

\[ y = X\beta + e \]  

(6)

where \( y \) is an nxl vector of observations, \( X \) is an nxp matrix of known constants, \( \beta \) is an pxl vector of unknown constants or parameters which are to be estimated and \( e \) is an nxl vector of errors. If \( \varepsilon(e) = 0 \) and if \( \varepsilon(e'e) = \sigma^2 I \), the best linear unbiased estimate of the unknown vector \( \beta \) is obtained by minimizing the residual sum of squares where

\[ Q = e'e = (y - X\beta)'(y - X\beta) \]  

(7)
The vector of estimates will be given by

$$\hat{\beta} = (X'X)^{-1} X'y$$  \hspace{1cm} (8)

and the variance-covariance matrix of \( \hat{\beta} \) can be written as

$$\epsilon[\hat{\beta}_1 - \epsilon(\hat{\beta}_1)]' \epsilon[\hat{\beta}_1 - \epsilon(\hat{\beta}_1)]' = (X'X)^{-1} \sigma^2.$$  \hspace{1cm} (9)

An unbiased estimate of the unknown variance is \( \hat{\sigma}^2 \) which can be computed from

$$\hat{\sigma}^2 = \frac{Q_{\min}}{n-p}.$$  \hspace{1cm} (10)

If \( \epsilon \) is NID \((0, \sigma^2)\), the above described Least Squares procedure for estimating \( \hat{\beta} \) is equivalent to the method of Maximum Likelihood.

It is apparent that the Least Squares procedure selects estimates from the unrestricted \( p \)-dimensional parameter space. In some instances, however, it is known a priori that the parameters must lie in a subspace of the unrestricted parameter space. If the restrictions imposed on the parameters constitute a set of \( q \) linear equations, the set of restrictions can be written \( A\hat{\beta} = k \) where \( A \) is an \( q \times p \) nonsingular matrix of known coefficients and \( k \) is an \( q \times 1 \) vector of known constants [13]. The Least Squares solution [16] can be obtained by transforming the problem in terms of new observations, \( u \), and a new matrix of constants \( W \) assuming that \( q \leq p \). Let the matrix \( A \) and the vector \( \beta \) be partitioned such that
\[ A\beta = \begin{bmatrix} A_1 & A_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = A_1\beta_1 + A_2\beta_2 = k \] (11)

where the inverse of \( A_1 \) exists.

Since

\[ \beta_1 = A_1^{-1}(k - A_2\beta_2) \] (12)

the model becomes

\[ y = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \begin{bmatrix} A_1^{-1}(k - A_2\beta_2) \\ \beta_2 \end{bmatrix} + e \] (13)

Upon simplification, we obtain

\[ (y - X_1A_1^{-1}k) = (X_2 - X_1^{-1}A_2) \beta_2 + e \]

or

\[ u = W\beta_2 + e \] (14)

Of course

\[ u = y - X_1A_1^{-1}k \]

and

\[ W = X_2 - X_1A_1^{-1}A_2 \]

Through these manipulations the model (14) reduces to (6) and we now proceed with the previously described procedure to obtain estimates of \( \beta_2 \). Knowing \( \hat{\beta}_2 \), \( \hat{\beta}_1 \) can be obtained from (12).
Rather than consider linear equalities, as was done above let us now examine the Least Squares procedure in which the parameters are restrained by inequalities, or more generally, restrained to a convex subspace. We find that very little has appeared in the statistical literature on this subject. Fortunately within the last two decades, techniques have been developed by mathematicians for maximizing or minimizing functions of parameters subject to the condition that the parameters are elements of a convex subspace. If the function to be minimized (or maximized) is linear, a situation that would be encountered if the sum of absolute deviations were minimized, the formulation and solution is called linear programming. When the function to be minimized (or maximized) is quadratic, as we have in this thesis, the generic term quadratic programming is applied.

C. Review of Some Linear and Quadratic Programming Procedures

Linear programming has been extremely successful in numerous problems associated with economic analysis. Many of the quadratic programming techniques reduce the quadratic function to be minimized to a linear form in order to effect a solution. Consequently a brief statement of the linear programming problem is most appropriate. In matrix notation the formulation is to find a vector \( x \) which
minimizes \( z = a'x \) \hspace{2cm} (15a)

subject to the constraints \( x_i > 0 \quad i = 1, \ldots, n \) \hspace{2cm} (15b)

and \( Ax \leq b \) \hspace{2cm} (15c)

The coefficient of the matrix \( A \) as well as the element of the vector \( a \) and \( b \) are assumed to be known constants. A detailed description of the techniques of linear programming, in particular the extremely popular Simplex method can be found in many sources, for example [3, 8].

With regard to regression analysis, Wagner [19] summarized and clarified the well known fact that the minimization of the sum of absolute deviations of a \( p \)-variable regression model can be recast into the linear programming model with \( n \) restrictions where \( n \) is the number of observations. He showed that the Simplex procedure was directly applicable but that \( n + 2p \) restrictions were required. A reduced linear programming problem with \( n \) restrictions was obtained if a bounded variable algorithm was employed. In this work no statistical properties of the estimates were obtained.

In its most general form the quadratic programming problem is to find a vector \( x \) which minimizes (or maximizes)

\[
F = \lambda p'x + \frac{1}{2} xCx
\]

where

\[
x_i \geq 0 \quad i = 1, \ldots, n
\] \hspace{2cm} (16b)

and

\[
a'x \leq b.
\] \hspace{2cm} (16c)
A, c, b, and p are known matrices and vectors. The parameter λ, which in our application will be 1, is used to introduce flexibility into the general problem.

Since

\[ Q = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta \] (17)

and \( y'y \) is a constant in the minimization of \( Q \) as a function of \( \beta \), our Least Squares model, in quadratic programming terminology, would be

\[
F = \frac{1}{2} (Q - y'y) \\
\delta = -y'X, \quad \lambda = 1 \\
C = X'X
\] (18) (19) (20)

We see that the elements of \( C \) are the coefficients of the normal equation and \( \delta \) is the vector of coefficients of the right-hand side of the normal equations. The criteria \( F \) must consequently be minimized in order to minimize \( Q \).

The quadratic programming problem, its algorithms and areas of application have been extensively studied. Almost all this work has been restricted to solutions in convex regions of the \( \beta \)-space or where the matrix of the quadratic form, \( C \), is positive or positive semi-definite. As is well known, if (17) is minimized, the associated \( \beta \) point will be either a local minimum or a saddle point. If \( C \) is positive definite or positive semi-definite we are assured that no saddle points are possible and in addition, the local minimum is also the global minimum.
Wolfe's [21] procedure for minimizing a quadratic objective function is to replace the quadratic function by an equivalent linear function obtained by combining the primal and the dual problem. The linearity of the gradient of the quadratic function is also used. His Theorem 2 shows that a solution of

\[ v'x = 0 \]  
\[ Cx - v + A'U + \lambda p' = 0 \]

is also a solution of (16a), (16b), and (16c). The vectors \( u \) and \( v \) are such that \( v(n \times 1) \geq 0 \) but \( u(m \times 1) \) can have positive or negative elements.

The main advantage of this approach is that the Simplex method — modified only slightly — can be used. However, for solving an \( m \) by \( n \) quadratic problem (\( n \) variables with \( m \) constraints) an \( (m+n) \) by \( (m+3n) \) table of coefficients is required. Fortunately it contains many zero entries.

Beale [1] initiated the work on minimizing a quadratic function by utilizing the fact that the gradient vector of a quadratic form is linear. Thus he was able to apply immediately the Simplex method. Essentially a "feasible solution" i.e. a solution satisfying the constraints is obtained at each iteration. Then at each iteration a quadratic function whose first and second derivatives at the feasible solution are the same as those of the given function
is minimized. The procedure converges in a finite number of iterations.

Charne's and Lemke's [4] main interest in quadratic programming was to solve the Least Squares problem but they generalized the algorithm to include objective function that were "separable" and convex, that is, \[ F = \sum_{i=1}^{n} f_i(x_i) \]
where \( f_i(x_i) \) is a convex function. Each \( f_i \) was replaced by a series of connected linear lines. However, due to the large number of zeros contained in the matrix of coefficients, the Simplex method was modified to utilize this advantage.

Frank and Wolfe [7] provide an algorithm that replaces the original quadratic objective function by a linear system augmented by Lagrangian multipliers (linearity of the gradient vector). The maximum in the linear system is known to be zero. The existence of solutions to the original problem is due to the boundedness of its objective function, and hence is equivalent to the feasibility of the linear constraints in the linear problem. The linear system is solved by the Simplex method. These authors also generalized their algorithm which is called "the gradient - and - interpolation method" to consider any convex objective function.

Hartley [9] proposed an algorithm that extended the results of Charnes and Lemke. Hartley's procedure not only includes separable convex objective function but also
separable convex restraints. His procedure is to approximate to any desired degree of accuracy all functions by connected straight lines. Although the matrix of coefficients become quite large an appreciable reduction in the problem is achieved by rephrasing the problem in terms of the dual. The simplex method is used. Zoutendijk [22] also treats the same case using the method of feasible direction (see below).

Zoutendijk's [22] method which he calls the method of feasible directions makes use of steep ascent. A "best" direction for each iteration is obtained by solving a small linear or quadratic programming problem. The length of the step to be taken in the "best" direction is obtained by requiring that the function, F, increase as much as possible from a given feasible solution. This is a one-dimensional maximization problem. An extension to problems involving nonlinear convex constraints is also considered as well as other special cases. In contrast to most other methods which converge in a finite number of iterations, the feasible direction method is nonterminating and thus requires a criterion for stopping if the approximation to the solution is reasonably close.

Houthakker [11] presents a finite iterative method in which the conventional calculus approach, equating the partial derivatives of the function, F, to zero and solving, is repeatedly applied. Assurance that the constraints are
are not violated, for example, all solution elements must be positive, is absorbed in a variable parameter which is called "the capacity". A capacity is calculated for each partial derivative. The smallest capacity which is less than a critical capacity indicates that the associated variable becomes effective, i.e. becomes nonzero or must be altered. Thus the method consists of successive determination of critical points and the associated set of effective variables. Once the latter are known, maximization proceeds by conventional calculus methods.

An interesting feature of this method is the ease in which the calculations can be initiated. Using a Taylor approximation, the quadratic terms are neglected and only the linear terms of the objective function enter into the first step.

Just recently Hartley and Hocking [10] obtained an algorithm which minimizes a quadratic form subject to strictly convex restraints, that is convex functions that are not lines or planes.

For nonconvex and nonlinear objective functions see Dantzig [6]. It is apparent that there are many quadratic programming algorithms currently available. Although some authors [4, 19] indicate the application to Least Squares theory, the primary interest is in the development of the algorithms. Apparently no work has been done on ascertaining
the statistical properties of the estimates that are calculated.
II. PROPERTIES OF THE QUADRATIC PROGRAMMING ESTIMATOR

Some small sample and asymptotic properties of $\tilde{\beta}$, the quadratic programming estimator, will be developed in which the only assumption is that the restricted parameter space, $S$, is convex. More specific results and more specialized properties for $\tilde{\beta}$ can undoubtedly be obtained if $S$ is a particular region, say a hypercube, a hypersphere or a half space. Some such specialized regions will be considered later.

It will be noted that the Least Squares estimator $\hat{\beta}$ is extensively used and explicitly associated with $\tilde{\beta}$. This is considered realistic and desirable since an extensive theory has been developed for the Least Squares estimator which will assist in the development of additional properties of $\tilde{\beta}$.

A. The "Minimum Distance" Property of $\tilde{\beta}$

Theorem 1: The minimization of the residual sum of squares, $Q$, in the convex region $S$ is identical with determining that point $\tilde{\beta}$ in $S$ which is "nearest" to the Least Squares estimator $\hat{\beta}$. The concept of "nearest" refers to the metric in which the elements of $\hat{\beta}$ are independently distributed with equal variance.

Proof: Consider the model

$$ y = X \beta + e \quad (1) $$
Let $T$ (pxp) and $U$ (nxp) be transformations such that

$$T'X'T = I_p = U'U$$

(2)

The model (1) can be rewritten as

$$y = XTT^{-1} \beta + e = U\gamma + e$$

(3)

where

$$\gamma = T^{-1} \beta$$

(4a)

$$U = XT$$

(4b)

Making the usual assumption that $E(e) = 0$ and $E(ee') = \sigma^2I$, the Gauss-Markoff Theorem [12, 13] can be applied to obtain the Least Squares solution of the transformed model (3).

Thus

$$\hat{\gamma} = (U'U)^{-1} U'y = U'y.$$  

(5)

In the transformed parameter space, the $\gamma$-space, the quadratic form that is minimized is

$$Q_{L.S.} = (y - U\hat{\gamma})'(y - U\hat{\gamma})$$

(6)

Denote by $Q_p$ the quadratic form minimized by the quadratic programming algorithm. $\hat{\beta}$ is the vector of estimates determined by quadratic programming. The algebraic decomposition identity shown below follows easily.
\[ Q_p = (y - \bar{\gamma})' (y - \bar{\gamma}) \]  
\[ = [(y - \bar{\gamma}) + U(\gamma - \bar{\gamma})]' [(y - \bar{\gamma}) + U(\gamma - \bar{\gamma})] \]
\[ = (y - \bar{\gamma})' (y - \bar{\gamma}) + 2(y - \bar{\gamma})' U(\gamma - \bar{\gamma}) \]
\[ + (\bar{\gamma} - \gamma)' U U(\gamma - \bar{\gamma}) \]
\[ Q_p = Q_{L.S.} + (\hat{\gamma} - \bar{\gamma})' (\hat{\gamma} - \bar{\gamma}) \]

From (4a) we see that \( \bar{\gamma} \) lies in the plane of \( \hat{\gamma} \). In addition \( \gamma \) lies in the plane of \( \bar{\gamma} \) as shown in the development of Theorem 2. Since \( Q_{L.S.} \) is a constant, it follows from (7b) that \( Q_p \) is a function of only \( \bar{\gamma} \) over the convex region \( S_\gamma \) and \( \bar{\gamma} \) appears only in the distance squared term

\[ (\hat{\gamma} - \bar{\gamma})' (\hat{\gamma} - \bar{\gamma}) = \sum_{i=1}^{p} (\hat{\gamma}_i - \bar{\gamma}_i)^2 \]

Thus minimizing \( Q_p \) is equivalent to minimizing the distance term. In other words, the quadratic programming technique selects from the acceptable region \( S \), that solution vector (or point) \( \bar{\gamma} \) that is nearest to the unrestrained Least Squares solution vector (or point) \( \hat{\gamma} \). Of course this property of minimum distance applies to a specific metric which has been shown to be that metric in which \( U'U \) is the identity matrix \( I \). In this metric the elements of \( \hat{\gamma} \) have equal variances, \( \sigma^2 \), and all their covariances will be zero.

Whether the minimum distance property of \( \bar{\gamma} \) can be used as a computational procedure in the general case is quite doubtful. It will usually be an effort to obtain the matrix \( T \) and much more difficult to minimize the distance in the
\( \gamma \)-space. It is apparent the \( T \) can be the orthonormal matrix of \( X'X \). However, \( T \) can also be the upper triangular matrix defined such that

\[
X'X = T'T
\]

If only a few inequality restrictions are operative, the latter decomposition of \( X'X \) can be used to reduce dimensionality of the \( \gamma \)-space considerably. In the resulting 1 or 2-dimensional parameter space, one can easily make use of the minimum distance property. Chapter V contains an example illustrating this procedure. When many restrictions must be satisfied, the convex region \( S \) is then the intersection of many restricting surfaces. Either of the two approaches using \( T \) explicitly will then become awkward.

In what follows we assume without loss of generality that for the original matrix \( X \) we already have \( X'X = I \) (rather than \( U'U = I \)) so that we can use the original \( \beta \) metric.

In the usual unrestricted Least Squares procedure, the addition or deletion of parameters or observations can be made rather easily using the previously computed results [13]. With \( \beta \) however, the entire quadratic programming algorithm must be used. Any change in the number of observations or parameters means that a new problem has been proposed. A new metric will be obtained under these circumstances. Thus if two estimates \( \beta_1 \) and \( \beta_2 \) are obtained
γ-space. It is apparent the \( T \) can be the orthonormal matrix of \( X'X \). However, \( T \) can also be the upper triangular matrix defined such that

\[ X'X = T'T \]  

(8)

If only a few inequality restrictions are operative, the latter decomposition of \( X'X \) can be used to reduce dimensionality of the \( \gamma \)-space considerably. In the resulting 1 or 2-dimensional parameter space, one can easily make use of the minimum distance property. Chapter V contains an example illustrating this procedure. When many restrictions must be satisfied, the convex region \( S \) is then the intersection of many restricting surfaces. Either of the two approaches using \( T \) explicitly will then become awkward.

In what follows we assume without loss of generality that for the original matrix \( X \) we already have \( X'X = I \) (rather than \( U'U = I \)) so that we can use the original \( \beta \) metric.

In the usual unrestricted Least Squares procedure, the addition or deletion of parameters or observations can be made rather easily using the previously computed results [13]. With \( \beta \) however, the entire quadratic programming algorithm must be used. Any change in the number of observations or parameters means that a new problem has been proposed. A new metric will be obtained under these circumstances. Thus if two estimates \( \beta_1 \) and \( \beta_2 \) are obtained
for two separate sets of observations, two transformations matrix $T_1$ and $T_2$ are also implied. If the data are combined and a new solution $\hat{\beta}_3$ is obtained by quadratic programming, then a new transformation matrix $T_3$ is involved which cannot be explicitly expressed in terms of $T_1$ and $T_2$.

B. A "Comparison" of $\beta$ and $\hat{\beta}$

Theorem 2: If $S$ is convex and the true parameter $\beta$ is in $S$, then $|\beta - \beta|^2 < |\hat{\beta} - \beta|^2$.

Proof: Let $\beta$ denote the true parameter vector which lies in $S$, $\hat{\beta}$, the unrestricted Least Squares vector of estimates, and $\bar{\beta}$, the vector of estimates given by the quadratic programming procedure. The lines through the points $(\hat{\beta}_i, \beta_i)$ and $(\bar{\beta}_i, \beta_i)$ are respectively

$$\beta_i^{(1)} = \beta_i + p (\hat{\beta}_i - \beta_i)$$

$$\beta_i^{(2)} = \beta_i + q (\bar{\beta}_i - \beta_i).$$

Since both $\beta_i$ and $\bar{\beta}_i$ are in $S$ and since $S$ is convex, the line $\beta_i^{(2)}$ also lies in $S$. The 2-dimensional plane determined by these two lines will be

$$\beta_i^0 = \beta_i + q (\bar{\beta}_i - \beta_i) + p (\hat{\beta}_i - \beta_i).$$

Consequently, the points $\bar{\beta}_i$, $\hat{\beta}_i$, and $\beta_i$ lie in the above plane.
Consider the triangle formed by the 3 points as shown by Figure 1a below.

![Figure 1a](image)

Let \( \alpha \) be the angle \( \beta \hat{S} \hat{P} \), it will now be shown that

\[
\frac{\pi}{2} \leq \alpha \leq \pi.
\]

Assume the contrary, that is \( \alpha < \frac{\pi}{2} \) as shown by Figure 1(b). There would be at least one point, say the foot of the perpendicular dropped from \( \hat{P} \), onto the line \( \hat{P}(2) \) which would be nearer to \( \hat{P} \) than to \( \hat{S} \). Furthermore, this point would be in \( S \). However this cannot be, since by Theorem 1, \( \hat{S} \) is the point in \( S \) that is nearest \( \hat{P} \). Hence \( \frac{\pi}{2} \leq \alpha < \pi \). From elementary trigonometry, we know that in a triangle with angles \( A, B \) and \( C \) and sides opposite \( a, b, \) and \( c \) respectively that

\[
a^2 = b^2 + c^2 - 2bc \cos A \]

where \( A \) is our \( \alpha \). Since \( \cos \alpha \) is negative, it follows immediately that
$a^2 > b^2$ or $|\hat{\beta} - \beta| \geq |\beta - \beta|$ since the inequality must hold for all $i$. We have of course assumed that the vectors $\hat{\beta}$, $\beta$ and $\beta$ are all distinct. If either $\hat{\beta} = \beta$ or $\beta = \beta$, the theorem is trivially true.

The theorem is obviously not true if $S$ is not convex as can be seen from Figure 2 below.

![Figure 2. S is not Convex](image)

Theorem 2 can be rewritten as

$$\sum_{i=1}^{p} (\beta_i - \beta_i)^2 \leq \sum_{i=1}^{p} (\hat{\beta}_i - \beta_i)^2$$  \hspace{1cm} (9)

since

$$|\hat{\beta} - \beta|^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta) = \sum (\beta_i - \beta_i)^2$$

and

$$|\beta - \beta|^2 = (\beta - \beta)'(\beta - \beta) = \sum (\beta_i - \beta_i)^2$$

From (9) it follows that

$$\sum_{i=1}^{p} \text{MSE} (\beta_i) = E|\beta - \beta|^2 \leq E|\hat{\beta} - \beta|^2 = \sum_{i=1}^{p} \text{Var} \hat{\beta}_i$$  \hspace{1cm} (10)
Attention is called to the fact that (10) does not imply the stronger statement that
\[
\text{MSE (} \hat{\beta}_i \text{)} \leq \text{Var (} \hat{\beta}_i \text{)}
\]
unless \( p = 1 \) or if \( \beta_i \ (i = 2, \ldots, p) \) are unrestricted.

C. Spherical Confidence Region

Theorem 3: Any spherical confidence region based on \( \hat{\beta} \) is not larger than that based on \( \hat{\beta} \).

Proof: To each \( \hat{\beta} \) there is attached one and only one corresponding value of \( \beta \) which from Theorem 2 has the property that \( \| \beta - \beta \| \leq \| \hat{\beta} - \beta \|^2 \). It follows that for a given spherical confidence region with radius \( R \),

\[
P_r \left( \| \beta - \beta \|^2 \leq R^2 \right) > P_r \left( \| \hat{\beta} - \beta \|^2 \leq R^2 \right) \tag{11}
\]
or if \( \alpha \) is the confidence coefficient

\[
\alpha = P_r \left( \| \beta - \beta \|^2 \leq D_1 \right) = P_r \left( \| \hat{\beta} - \beta \|^2 \leq D_2 \right) \tag{12}
\]

where \( D_1 \leq D_2 \).

Consequently the confidence coefficient associated with a particular spherical confidence region \( R \) computed from the estimator \( \hat{\beta} \) cannot be smaller than that computed from the Least Squares estimator \( \hat{\beta} \).

D. Asymptotic Properties

In this section we consider that \( n \) replicates of \( y \) are available. The transformed Least Squares model of immediate
concern will then be

\[ \bar{y} = X\beta + \bar{\epsilon} \]

where

\[
\begin{bmatrix}
\bar{y}_1 \\
\bar{y}_2 \\
\vdots \\
\bar{y}_N
\end{bmatrix} \quad \text{and} \quad \bar{\epsilon} = 
\begin{bmatrix}
\bar{\epsilon}_1 \\
\bar{\epsilon}_2 \\
\vdots \\
\bar{\epsilon}_N
\end{bmatrix}
\]

Thus \( \bar{y}_i \) is the average response for the \( i \)-th set of observations and \( \bar{\epsilon}_i \) is the average error associated with \( \bar{y}_i \). We make the usual assumption that \( \bar{\epsilon}_i \) is NID \( (0, \sigma^2) \) for all \( i \). There is no loss in generality by assuming a common variance.

1. **Consistency**

   Theorem 4: The quadratic programming estimators, \( \hat{\beta}_i \), are consistent. Since \( \hat{\beta} \) is known to be consistent [5] and since \( |\hat{\beta} - \beta_0| \leq |\hat{\beta} - \beta_0| \) it follows that \( \hat{\beta} \) is consistent.

2. **Asymptotic normal behavior**

   Three possible general cases arise since the population mean \( \beta_0 \) can lie within the convex region \( S \), on a simple boundary of \( S \) or at the intersection of two or more boundaries. Only the first case will be considered here.
Theorem 5. When $\beta_0$ is an interior point of $S$, then with probability approaching unity as $n$ increases, the quadratic programming estimator $\hat{\beta}$ is distributed $N(\beta_0, \frac{\sigma^2}{n})$.

Proof: Let $R$ be the radius of the largest sphere about $\beta_0$ which is contained in $S$. Then it follows that

$$P_r \{\hat{\beta} \text{ lies in } S\} \geq P_r \{||\hat{\beta} - \beta_0||^2 \leq R^2\} \tag{13}$$

If $S$ is a spherical region the above becomes an equality, otherwise the inequality holds. The right hand term can be written as

$$P_r \left\{x_\beta^2 \leq \frac{nR^2}{\sigma^2}\right\} = P_r \left\{x_\beta^2 \leq \frac{nR^2}{\sigma^2}\right\} \tag{14}$$

So we have

$$P_r \{\hat{\beta} \text{ lies inside } S\} \geq P_r \left\{x_\beta^2 \leq \frac{nR^2}{\sigma^2}\right\}$$

as $n \to \infty$

$$P_r \left\{x_\beta^2 \leq \frac{nR^2}{\sigma^2}\right\} \to 1$$

and as a consequence

$$P_r \{\hat{\beta} \text{ lies outside } S\} \to 0$$

Therefore, the distributions of $\hat{\beta}$ and $\beta$ tend to be equal in probability. Hence it follows that since $\hat{\beta}$ is $N(\beta_0, \frac{\sigma^2}{n})$, then

$$P_r \{\beta \text{ is } N(\beta_0, \frac{\sigma^2}{n})\} \to 1 \text{ as } n \to \infty \tag{15}$$
E. Confidence Regions for $\beta_0$

It has been assumed that the population mean of $\beta$ lies in the convex region $S$. The question arises whether we can utilize this information for the construction of confidence regions without employing the point estimator $\hat{\beta}$. These new confidence regions which will be developed below will not be the "spherical" regions of Section C above, but will usually have smaller volumes in our metric. They can be derived in the following manner.

Consider two "events" $A$ and $B$ where:

Event A: that $\beta$ lies in $S$

Event B: $|\beta - \hat{\beta}|^2 \leq p \hat{\sigma}^2 \frac{F_p, n-p}{p}$ ($\alpha$)

In $\beta$ the estimator $\hat{\sigma}^2$ is the familiar residual mean square estimator of $\sigma^2$ based on $n-p$ d.f. and $F_{p, n-p}$ ($\alpha$) is the $100\alpha\%$ point of the $F$-distribution with $p$ and $n-p$ d.f.

Clearly

$$P_r(A) = 1 \text{ and } P_r(B) = 1 - \alpha$$

Moreover for the probability that $A$ or $B$ occurs is

$$P_r(A + B) = 1.$$ 

Hence for the probability that $A$ and $B$ to occur, we have

$$P_r(A \cdot B) = P_r(A) + P_r(B) - P_r(A+B)$$

$$= 1 + (1 - \alpha) - 1$$

$$= 1 - \alpha$$
For the event A and B to occur $\beta$ must lie in the intersection of S and the sphere described by B and so this intersection is an exact confidence region with a $1 - \alpha$ confidence coefficient. If $\hat{\beta}$ is so far away from S that the intersection of S with the sphere is empty, the only meaningful conclusion is that the assumption that $\beta$ is in S must be refuted, that is, the reference set S is to be rejected.

F. Generalized Least Squares

From Least Squares theory it is known [13] that if $\varepsilon(\varepsilon \varepsilon') = K \sigma^2$, where K is a known nonsingular matrix (nxn), there exists a matrix T(nx n) such that $TKT' = I$. Hence transforming the original model we obtain $Ty = TX \beta + Te$. $Ty$ is a new vector of observations, $TX$ is the transformed matrix of known coefficients, and $Te$ is the transformed vector of errors. The transformed model and the transformed set of restrictions $TA \beta \leq Tk$ can now be subject to all the manipulations described in the previous sections.
III. ASYMPTOTIC BIAS AND MEAN SQUARE ERROR

In this chapter we return to the general case in which the Least Squares estimators, $\hat{\beta}_1$, have a multivariate density $(N(\beta, \sigma^2\Sigma)$ where $\beta$ is a $k$-dimensional vector of means and $\sigma^2\Sigma$ is the $p \times p$ variance-covariance matrix.

A. Two Parallel Restricting Planes

We now assume that the convex space $S$ in which $\beta$ lies is bounded by two parallel planes, that is

$$K_1 \leq a'\beta \leq K_2$$

(1)

where $a'$ is a unitary vector. Now the transformation matrix $T$ given by (II, 2) may, of course, be multiplied by an orthogonal matrix without changing its properties. Let it therefore be so chosen that the transformation of the vector $a'$ coincides with the $\gamma_1$ axis. This implies, since

$$a'\beta = a'T\gamma = (Ta')'\gamma$$

that $(T'a)'$ should become a row vector of the form

$$(T'a)' = (e, 0, 0, \ldots, 0).$$

(2)

We can then write (1) in the form

$$k_1 = \frac{K_1}{e} \leq \gamma_1 \leq \frac{K_2}{e} = k_2$$

(3)

In the following we shall, for convenience, let $x_1 = \gamma_1$ and $x_1$ denote the population mean in the $\gamma$-space.
Since the estimators $x_i$ are independently distributed by virtue of (2) it follows that the following expectations are immediately obtained

\begin{align*}
\mathbb{E}(x_i - x_i) &= 0 \quad i = 2, \ldots, p \quad (4a) \\
\mathbb{E}(x_i - x_i)^2 &= 1 \quad i = 2, \ldots, p \quad (4b) \\
\mathbb{E}((x_i - x_i)(x_j - x_j)) &= 0 \quad i, j = 1, \ldots, p \quad (4c) \\
\text{and } i \neq j
\end{align*}

There remains the evaluation of $\mathbb{E}(x_1)$ and $\text{Var}(x_1)$ from a modified univariate normal density. Consider the weighted density function

\begin{align*}
g(x_1) &= g(x_1 \mid x_1 < k_1) P(x_1 < k_1) + g(x_1 \mid k_1 x_1 < k_2): \\
P(k_1 \leq x_1 \leq k_2) + g(x_1 \mid x_1 < k_2) P(x_1 > k_2) \quad (5a)
\end{align*}

where

\begin{align*}
g(x_1 \mid x_1 < k_1) &= k_1 \\
g(x_1 \mid x_1 > k_2) &= k_2 \\
P(x_1 < k_1) &= N(k_1; X_1, 1) = N_1 \\
P(x_1 > k_2) &= 1 - N(k_2; X_1, 1) = (1 - N_2) = Q_2
\end{align*}

and

\begin{align*}
g(x_1 \mid k_1 \leq x_1 \leq k_2) &= n(x_1; X_1, 1).
\end{align*}

The density of $x_1$ can be described as follows. For $k_1 \leq x_1 \leq k_2$, $x_1$ is distributed normally, $n(x_1; X_1, 1)$ with points of truncation at $k_1$ and $k_2$. If $x_1 < k_1$, $x_1$ is taken
as $k_1$ with a cumulative normal weight of $N_1$. Similarly if $x_1 > k_2$, the estimate of $x_1$, is taken to be $k_2$ with a cumulative normal weight of $Q_2$. Thus $x_1$ is distributed normally with spikes of height $N_1$ and $Q_2$ at $k_1$ and $k_2$ respectively. It is assumed that the population mean $X_1$ lies between $k_1$ and $k_2$. The density can be written in a more useful form as

$$f(x_1; X_1) = \begin{cases} k_1 \text{ with probability } N_1 & \text{if } x_1 < k_1 \\ (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (x_1 - X_1)^2\right\} & \text{if } k_1 \leq x_1 \leq k_2 \\ k_2 \text{ with probability } Q_2 & \text{if } x_1 > k_2 \end{cases}$$

(5b)

1. **The first moment of $x_1$**

From (5b) it follows that

$$\varepsilon(x_1) = k_1 N_1 + \int_{k_1}^{k_2} x_1 (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} (x_1 - X_1)^2\right\} \, dx_1 + k_2 Q_2$$

After integration and rearrangement, we have

$$\varepsilon(x_1) = n_1 - n_2 - (X_1 - k_1) N_1 + X_1 N_2 + k_2 Q_2$$

(6)

where

$$n_1 = n \left(k_1; X_1, 1\right)$$

and

$$n_2 = n \left(k_2; X_1, 1\right).$$
2. The second moment, variance, and MSE of $x_1$

Again using the density function (5b), we obtain

$$
\varepsilon(x_1^2) = k_1^2 N_1 + \int k_2 \int (2\pi)^{-\frac{1}{2}} x_1^2 \exp\{-\frac{1}{2}(x_1 - x_1)^2\} \, dx_1 + k_2^2 Q_2
$$

$$
= k_1^2 N_1 + k_2^2 Q_2 + (k_1 + x_1) n_1 - (k_2 + x_1) n_2
$$

$$
+ (1 + x_1^2) (N_2 - N_1)
$$

(7)

Since

$$
\text{Var}(x_1) = \varepsilon(x_1^2) - \varepsilon^2(x_1)
$$

the variance of $x_1$ is obtained by using (6) and (7). After substitution and rearrangement we have that

$$
\text{Var}(x_1) = (k_1 + x_1 - 2k_2) n_1 + (k_2 - x_1) n_2 + (1 + x_1^2 - 2k_1 k_2 + k_2^2) N_2
$$

$$
- (1 + x_1^2 - k_1^2 - 2k_2 x_1 + 2k_2 k_1) N_1 -
$$

$$
\{n_1 - n_2 - (X_1 - k_1) N_1 - (k_2 - X_1) N_2\}^2
$$

(8)

The mean square error (MSE) which is defined as

$$
\text{MSE}(x_1) = \varepsilon(x_1^2) - 2X_1 \varepsilon(x_1) + X_1^2
$$

can also be obtained from (6) and (7). Upon substitution and rearrangement we find that

$$
\text{MSE}(x_1) = (k_1 - X_1) n_1 - (k_2 - X_1) n_2 + N_2 - N_1 +
$$

$$
(X_1 - k_2)^2 Q_2 + (X_1 - k_1)^2 N_1
$$

(9)
Thus (6), (7), and (9) which give the $\varepsilon(x_1)$, $\text{Var}(x_1)$ and $\text{MSE}(x_1)$ represent the basic equations which will be used to investigate some special cases of the derived equations.

3. **Monotonic behavior of MSE** ($x_1$)

It will be shown that the derivative of the MSE($x_1$) with respect to $k_1$ and $k_2$ are monotonic functions. This result along with sections 4, 5, and 6 which follow supports the obvious conclusion that as the boundaries of restriction are moved away from the mean the MSE approaches the unrestricted variance.

From the differentiation of (9) we obtain

$$\frac{d \text{MSE}(x_1)}{dk_1} = -n_1 - (k_1 - x_1)^2 n_1 - n_1 - 2 (x_1 - k_1) N_1 + (x_1 - k_1)^2 n_1$$

$$= -2n_1 - 2(x_1 - k_1) N_1 = -2 [n_1 + (x_1 - k_1) N_1]$$

Since $k_1 \leq x_1$, all the terms in the bracket are positive. Thus it is shown that the derivative of the MSE with respect to $k_1$ must always be negative.

Consider now

$$\frac{d \text{MSE}(x_1)}{dk_2} = -(x_1 - k_2)^2 n_2 - 2 Q_2 (x_1 - k_2) + (k_2 - x_1)^2 n_2$$

$$= 2 (k_2 - x_1) Q_2$$

Since $k_2 > x_1$, the derivative will always be positive.
4. Behavior of \( \varepsilon(x_1) \), \( \text{Var}(x_1) \) and \( \text{MSE}(x_1) \) as \( k_1 \rightarrow -\infty \)

As \( k_1 \rightarrow -\infty \) (6), (8) and (9) simplify to

\[
\varepsilon(x_1) = -n_2 + x_1 N_2 + k_2 Q_2
\]

\[
\text{Var}(x_1) = (k_2 - x_1) n_2 + (x_1^2 - 2 x_1 k_2 + k_2^2) N_2 - 
\left[ n_2 + (k_2 - x_1) N_2 \right]^2
\]

\[
\text{MSE}(x_1) = (x_1 - k_2)^2 (Q_2) - (k_2 - x_1) n_2 + N_2
\]

If \( x_1 = k_2 \) then \( N_2 = \frac{1}{2} \) and \( n_2 = (2\pi)^{-\frac{1}{2}} \). Equations (10), (11), and (11) will reduce to

\[
\varepsilon(x_1) = k_2 - (2\pi)^{-\frac{1}{2}} = x_1 - (2\pi)^{-\frac{1}{2}}
\]

\[
\text{Var}(x_1) = \frac{1}{2} \left( 1 - \frac{1}{n} \right) = 0.34084
\]

\[
\text{MSE}(x_1) = \frac{1}{2}
\]

In this special case i.e., when the restricting plane passes through the mean \( x_1 \), a negative constant bias of \( (2\pi)^{-\frac{1}{2}} \) is introduced in the expectation of \( x_1 \). More important the MSE is a half of the variance of the unrestricted Least Squares estimator of \( x_1 \). We see here an important gain in precision of the restricted estimator.

5. Behavior of \( \varepsilon(x_1) \), \( \text{Var}(x_1) \) and \( \text{MSE}(x_1) \) as \( k_1 \rightarrow -\infty \) and \( k_2 \rightarrow +\infty \)

Equations (10), (11), and (12) will simplify upon re-arrangement to
\[ e(x_1) = x_1 \]
\[ \text{Var}(x_1) = 1 = \text{MSE}(x_1) \]

These equations can be anticipated since as \( k_1 \to -\infty \) and \( k_2 \to +\infty \), the distribution of \( x_1 \) will be the complete univariate normal with mean \( x_1 \) and variance equal to 1.

6. Behavior of \( e(x_1) \), \( \text{Var}(x_1) \) and \( \text{MSE}(x_1) \) as \( k_2 \to +\infty \)

As \( k_2 \to +\infty \) equations (6), (8) and (9) upon rearrangement reduce to

\[ e(x_1) = x_1 + n_1 - (X_1 - k_1) N_1 \]  \hspace{1cm} (14a)
\[ \text{Var}(x_1) = (k_1 - X_1) n_1 + Q_1 + (X_1 - k_1)^2 N_1 - \left[ n_1 - (X_1 - k_1) N_1 \right]^2 \]  \hspace{1cm} (14b)
\[ \text{MSE}(x_1) = (k_1 - X_1) n_1 + Q_1 + (X_1 - k_1)^2 N_1 \]  \hspace{1cm} (14c)

7. Bias and variance of the original variables

Since from (2) we have

\[ \beta = T \gamma \]

so it follows directly that

\[ \text{Bias} (\beta) = T \left[ \begin{array}{c} \text{bias} (\gamma_1) \\ 0 \\ \vdots \\ 0 \end{array} \right] = \text{bias} (\gamma_1) \cdot T \cdot 1 \]  \hspace{1cm} (15)
where $T_1$ is the first column of the matrix $T$. Thus the bias in one estimate will upon retransformation usually enter into all the original estimates. It may fortuitously happen that some of the elements of $T_1$ will be zero. Then the biases associated with these zero elements will also be zero.

With regard to the variances and covariances we have

$$[\tilde{\beta} - e(\tilde{\beta})] [\tilde{\beta} - e(\tilde{\beta})]' = T e [\tilde{\gamma} - e(\tilde{\gamma})] [\tilde{\gamma} - e(\tilde{\gamma})]' T'$$

$$= T \begin{bmatrix} V(\tilde{\gamma}_1) & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} T'$$

$$= T \left[ I + \begin{bmatrix} V(\tilde{\gamma}_1) - 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right] T'$$

$$= T I T' + [V(\tilde{\gamma}_1) - 1][T_1, 0, \cdots, 0][T_1, T_2, \cdots, T_p]'$$

$$= \Sigma \sigma^2 + \Sigma^2 [V(\tilde{\gamma}_1) - 1] T_1 T_1'$$

Since $T_1 T_1'$ which is (pxp) will have $t_{i1}^2$ on the diagonal of the $i$-th row and since $[V(\tilde{\gamma}_1) - 1]$ will be negative, it is apparent that all the variances of $t$ will be decreased. Covariances may increase or decrease depending on the combination
of signs associated with elements $t_{11}$ and $t_{j1}$.

B. Two Non-Parallel Restricting Planes

Once again let $\beta_1$ have a multivariate density $N(\beta, \sigma^2 \Sigma)$ where $\beta$ is a $p$ dimensional vector of means and $\sigma^2 \Sigma$ is the $p \times p$ variance-covariance matrix. The two linear restrictions will be denoted by

$$a_1' \beta \leq \ell_1$$
$$a_2' \beta \leq \ell_2$$

where $a_1' \neq k a_2'$. Thus the restricting planes are not parallel. Consider the transformation

$$\beta = T_1 y$$

such that $y$ has a multivariate density $N(0, I)$. The restrictions become

$$C_1' y \leq m_1$$
$$C_2' y \leq m_2.$$
\[ C_0 \mathbf{q} = \begin{bmatrix} C_1' \\ C_2' \end{bmatrix} \mathbf{q} = 0 \]

where \( \mathbf{q} \) is a column vector of unknowns, and \( C_1' \) and \( C_2' \) have been normalized. The matrix \( C_0 \) is of rank 2 so there are \( p-3 \) variables say \( q_3, ..., q_p \) can be arbitrarily selected. Let \( C_3' \) be one of the normalized possible solutions to the above equations. Then it follows that

\[
\begin{align*}
C_1' C_3 &= 0 \\
C_2' C_3 &= 0 \\
C_3' C_3 &= 1
\end{align*}
\]

Repeating the same process but including the new solution, we solve

\[ C_1 \mathbf{q} = \begin{bmatrix} C_1' \\ C_2' \\ C_3' \end{bmatrix} \mathbf{q} = 0 \]

Let \( C_4' \) be one of the possible normalized solutions. It follows that

\[
\begin{align*}
C_1' C_4 &= 0 \\
C_2' C_4 &= 0 \\
C_3' C_4 &= 0 \\
C_4' C_4 &= 1
\end{align*}
\]

Carrying this construction to termination, a transformation, \( C, \)
is obtained such that

\[ C_i' C_j = 0 \]
\[ C_2' C_j = 0 \quad j = 3, \ldots, p \]
\[ C_i' C_1 = 1 \quad i = 1, \ldots, p \]
\[ C_1' C_2 = r \]

Let \( C = T_2 \) and let \( y = T_2 w \). The variance-covariance matrix of \( w \) becomes

\[
\begin{bmatrix}
1 & r & 0 & \cdots & 0 \\
r & 1 & 0 & \cdots & 0 \\
0 & 0 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & I_{p-2} \\
0 & 0 & & & 0
\end{bmatrix}
\]

To diagonalize the variance-covariance matrix let

\[ w = T_3 v \]

where

\[
T_3 =
\begin{bmatrix}
1 & 0 & 0 & \cdots & 0 \\
-r & 1 & 0 & \cdots & 0 \\
0 & 0 & \ddots & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & I_{p-2} \\
0 & 0 & & & 0
\end{bmatrix}
\]

The variance-covariance matrix of \( v \) becomes
Letting \( v = T_4 y \)

the variance-covariance matrix becomes \( I \) and the quadratic form reduces to \( y' y \).

It can be shown that the above series of transformations

\[ \beta = T_1 T_2 T_3 T_4 y \]

reduces the restraints to

\[ y_1 \leq k_1 \]

\[ a y_1 + b y_2 \leq k_2. \]

As before, for convenience, let \( x = y \).

The region which satisfies the above inequalities is the acceptable parameter space which will be called \( S_1 \). We now partition the unacceptable space into 3 parts in order to facilitate the work which is to follow. In the \( x_1 x_2 \)-plane, the point of intersection of the two restricting lines will be

\[ [k_1, \frac{1}{b} (k_2 - a k_1)] = (k_1, p) \]
A line perpendicular to \( ax_1 + bx_2 = k_2 \) and through the point of intersection is given by

\[-bx_1 + ax_2.\]

The unacceptable space contained between \( ax_1 + bx_2 = k_2 \) and the above perpendicular will be denoted by \( S_2 \). \( S_3 \) is that parameter space between \( x_1 = k_1 \) and \( x_2 = p \) and \( S_4 \) is the remaining region and lies between \( x_2 = p \) and the above described perpendicular line. Hence the \( x_1x_2 \)-plane will have been partitioned into four regions which is shown in Figure 3.

![Figure 3. The Four Partitioned Regions](image)

The joint density function of \( x_1 \) and \( x_2 \) will be

\[
f(x_1, x_2) = f(x_1, x_2 \mid (x_1, x_2) \in S_1)P[(x_1, x_2) \in S_1] + \]
\[
f(x_1, x_2 \mid (x_1, x_2) \in S_2)P[(x_1, x_2) \in S_2] + \]
\[
f(x_1, x_2 \mid (x_1, x_2) \in S_3)P[(x_1, x_2) \in S_3] + \]
\[
f(x_1, x_2 \mid (x_1, x_2) \in S_4)P[(x_1, x_2) \in S_4].
\]
Again using the normal distribution the four terms on the right-hand side are evaluated in the following way:

In $S_1$ we have

$$f[x_1, x_2 \mid (x_1, x_2) \in S_1] \cdot P[(x_1, x_2) \in S_1]$$

$$= \frac{1}{2\pi e} \frac{1}{\frac{1}{2\pi e} (x_1^2 + x_2^2)} \cdot P[(x_1, x_2) \in S_1]$$

$$= \frac{1}{2\pi e} \cdot n(x_1) \cdot n(x_2)$$

In $S_4$ we have

$$f[x_1, x_2 \mid (x_1, x_2) \in S_4] \cdot P[(x_1, x_2) \in S_4] = P[(x_1, x_2) \in S_4] = p_4.$$ Since

$$f[x_1, x_2 \mid (x_1, x_2) \in S_4] = 1$$

because if the estimates fall in $S_4$, the point of intersection of the two non-parallel planes will be accepted as the quadratic programming estimates. This follows because of the minimum distance property. In $S_3$ we have that

$$f[x_1, x_2 \mid (x_1, x_2) \in S_3] \cdot P[(x_1, x_2) \in S_3] =$$

$$\int_{k_1}^{\infty} n(x_1) \, dx_1 \cdot n(x_2)$$

$$\frac{1}{P[(x_1, x_2) \in S_3]} \cdot P[(x_1, x_2) \in S_3] = Z(k_1) \cdot n(x_2).$$
Thus if the estimates fall in region $S_3$, $x_1$, will be assigned the values $k_1$ and $x_2 = x_2$.

Consider now the density in $S_2$. For convenience, the axis is rotated so that the new axis are perpendicular and parallel to the boundary of $S_2$. Let the new axis be $X_1$ and $X_2$, then

$$
\begin{bmatrix}
X_1 \\
X_2
\end{bmatrix} = \begin{bmatrix}
a & b \\
-b & a
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}
$$

The transformation matrix is orthogonal so the Jacobian will be unity. Thus

$$
f[x_1, x_2 | (x_1, x_2) \in S_2] P [(X_1, X_2) \in S_2] =
$$

$$
\int_{k_2}^{\infty} \left( \frac{1}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{1}{2}X_2^2} \ dx_2 \cdot \left( \frac{1}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{1}{2}X_1^2} \ dx_1
$$

$$
P[(X_1, X_2) \in S_2] \cdot P[(X_1, X_2) \in S_2] =
$$

$$
\left[ 1 - F_2 (k_2) \right] \left( \frac{1}{2\pi} \right)^{\frac{1}{2}} e^{-\frac{1}{2}X_1^2}
$$

where

$$
-\infty \leq X_1 < \frac{a}{b} k_2 - \frac{k_1}{b} = q.
$$

The notation $F_2$ denotes the cumulative normal density from $-\infty$ to $k_2$ in terms of the coordinate $X_2$. In $S_2$, the estimates are modified so that $X_2 = k_2$ and $X_1$ is distributed as shown above.
The density can now be written as

\[
f(x_1, x_2) = \begin{cases} 
  n(x_1)n(x_2) & \text{if } (x_1, x_2) \in S_1 \\
  (1-F_2)n(x_1) & \text{if } (x_1, x_2) \in S_2 \\
  X_2 = k_2 \text{ with probability } (1-F_2)F(q) \\
  Z(k_1)n(x_2) & \text{if } (x_1, x_2) \in S_3 \\
  x_1 = k_1 \text{ with probability } Z(k_1)n(p) \\
  x_1 = k_1, \text{ with probability } P_4 & \text{if } (x_1, x_2) \in S_4 \\
  x_2 = p \text{ with probability } P_4 
\end{cases}
\]

(15)

Because of the transformation \( x_3, \ldots, x_p \) are all independent and distributed as \( N(0, 1) \).

In the following section, the moments will be derived for two of the four possible situations, namely when the angle \( \alpha \) is obtuse and when it is acute, but the acceptable parameter space lies below the plane \( ax_1 + bx_2 = k_2 \). Of course, the other two cases are similar but the acceptable space is above the restricting planes.

1. \( \alpha \) is obtuse
   a. The first moment of \( x_1 \). In \( S_1 \), we find using the density (15) that

\[
\varepsilon_1(x_1) = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} x_1 n(x_1)n(x_2) \, dx_1 \, dx_2.
\]
Upon integrating by parts and completing the square, we obtain

\[ \varepsilon_1(x_1) = -n(k_1) F(p) - an(k_2) F(k_2; a, k_2, b^2) \]  \hspace{1cm} (16)

In \( S_2 \), the first moments in the transformed variables are

\[ \varepsilon(X_1) = (1-F_2) \int_{-\infty}^{q} X_1 n(X_1) dX_1 = -n(q) (1-F_2) \]  \hspace{1cm} (17)

\[ \varepsilon(X_2) = (1-F_2) \int_{-\infty}^{q} k_2 n(X_1) dX_1 = k_2 (1-F_2) F_1(q) \]  \hspace{1cm} (18)

Denote \( F_1(q) \) by \( F \) where the subscript indicates the transformed variable \( X_1 \). Since

\[ \varepsilon_2(x_1) = a \varepsilon(X_1) - b \varepsilon(X_2) \]

we have, upon substitution of (17) and (18) into the above, that

\[ \varepsilon_2(x_1) = -(1-F_2) [a n(q) + b k_2 F] \]  \hspace{1cm} (19)

From (15), the expectation of \( x_1 \) in \( S_3 \) is

\[ \varepsilon_3(x_1) = k_1 Z(k_1) F(p) \]  \hspace{1cm} (20)

From (15) we obtain directly that \( \varepsilon_4(x_1) = k_1 F_4 \). The regional expectations are summed to obtain the over-all expectation of \( x_1 \) since the regions are disjoint. Thus

\[ \varepsilon(x_1) = \sum_{i=1}^{4} \varepsilon_i(x_1) \]
Upon substitution, we obtain

\[
\varepsilon(x_1) = -n(k_1) F(p) - an(k_2) F(k_1; ak_2, b^2) + k_1 Z(k_1) F(p) - (1-F_2) (an(q) + bF_1) + k_1 P_4.
\]

(21)

b. The first moment of \(S_2\). In region \(S_1\), we have that

\[
\epsilon_1(x_2) = \int \int x_2 n(x_1) n(x_2) \, dx_1 \, dx_2.
\]

Integrating by parts, completing the square and simplifying, we obtain

\[
\epsilon_1(x_2) = -bn(k_2) F(k_1; ak_2, b^2).
\]

(22)

Using (17) and (18) and the fact that

\[
\varepsilon_2(x_2) = a \varepsilon(X_1) + b \varepsilon(X_2)
\]

we obtain in \(S_2\) the following

\[
\varepsilon_2(x_2) = (1-F_2) [a k_2 F_1 - b n(q)]
\]

(23)

In \(S_3\) it follows that

\[
\varepsilon_3(x_2) = Z(k_1) \int x_2 n(x_2) \, dx_2 = Z(k_1) n(p)
\]

(24)

Region \(S_4\) yields directly from (15) that

\[
\varepsilon_4(x_2) = p P_4.
\]
Hence combining the four expectations produces

\[ \varepsilon(x_2) = \text{sn}(k_2) F(k_1; ak_2, b^2) - Z(k_1) n(p) + p F_4 + (1-F_2) [ak_2 F_1 - bn(q)] \]

(25)

c. The second moment and MSE of \( x_1 \). In \( S_1 \) we have

\[ \varepsilon_1(x_1^2) = \int \int x_1^2 n(x_1) n(x_2) \, dx_1 \, dx_2. \]

Integrating by parts and completing the square, we obtain

\[ \varepsilon_1(x_1^2) = \int n(z) F \left( \frac{k_2}{b} - \frac{a}{b} z \right) \, dz - k_1 n(k_1) F(p) + a n(k_2) [b^2 n(k_1; ak_2, b^2) - a k_2 F(k_1; ak_2, b^2)]. \]

With regard to \( S_2 \), we find that

\[ \varepsilon(x_1^2) = (1-F_2) \int X_1^2 n(X_1) \, dx_1 \]

\[ = (1-F_2) (F_1-q n(q)). \]

In like manner

\[ \varepsilon(x_2^2) = (1-F_2) \int k_2^2 n(X_1) \, dx_1 \]

\[ = (1-F_2) k_2^2 F_1. \]
The mixed moment in $S_2$ is

$$\varepsilon(X_1X_2) = (1-F_2) \int_{-\infty}^{\infty} k_2 X_1 n(X_1) \, dX_1 = k_2 \varepsilon(X_1).$$

Substituting for $\varepsilon(X_1)$ the mixed moment becomes

$$\varepsilon(X_1X_2) = (1-F_2) k_2 n(q).$$

Since

$$x_1 = ax_1 - bx_2$$

it follows that

$$\varepsilon(x_1^2) = a^2 \varepsilon(x_1^2) + b^2 \varepsilon(x_2^2) - 2ab \varepsilon(x_1x_2).$$

Substituting into this expression, we obtain

$$\varepsilon_2(x_1^2) = (1-F_2) (2abk_2 - b^2q) n(q) +$$

$$(1-F_2)(a^2 + b^2k_2^2) F_1. \tag{26}$$

In region $S_3$ since $x_1 = k_1$, we obtain directly from (15) that

$$\varepsilon_3(x_1^2) = Z(k_1) k_1^2 N(p). \tag{27}$$

Similarly in region $S_4$, (15) gives directly

$$\varepsilon_4(x_1^2) = k_1^2 F_4. \tag{28}$$

As before, the expectations for each region are combined so that
\( \varepsilon(x_1^2) = \text{MSE}(x_1) = \int_{-\infty}^{k_1} n(z) F\left( \frac{k_2}{b} - \frac{a}{b} z \right) \, dz - k_1 n(k_1) F(p) + a n(k_2) [b^2 n(k_1; a k_2, b^2) - a k_2 F(k_1; a k_2, b^2)] + (1-F_2)(2 ab k_2 - b^2 q) n(q) + (1-F_2)(a^2 + b^2 k_2) F_1 + k_1^2 Z(k_1) F(p) + k_1^2 P_4 \)  

(29)

d. The second moment and MSE of \( x_2 \).

In a manner similar to that used in the preceding section, we obtain

\( \varepsilon_1(x_2^2) = \int_{-\infty}^{k_1} n(z) F\left( \frac{k_2}{b} - \frac{a}{b} z \right) \, dz - b^2 k_2 n(k_2) F(k_1; a k_2, b^2) - ab^2 n(k_1) n(k_1; a k_2, b^2) \)

\( \varepsilon_2(x_2^2) = (1-F_2)(b^2 + a^2 k^2) F_1 - (1-F_2)(q b^2 + 2 ab k_2) n(q) \)

(30)

\( \varepsilon_3(x_2^2) = z(k_1) [F(p) - p n(p)] \)  

(31)

\( \varepsilon_4(x_2^2) = p^2 P_4 \)  

(32)

Combining in the usual manner will yield

\( \varepsilon(x_2^2) = \text{MSE}(x_2) = \int_{-\infty}^{k_1} n(z) F\left( \frac{k_2}{b} - \frac{a}{b} z \right) \, dz - b^2 k_2 n(k_2) F(k_1; a k_2, b^2) - ab^2 n(k_1) n(k_1; a k_2, b^2) + (1-F_2)(b^2 + a^2 k^2) F_1 - (1-F_2)(q b^2 + 2 ab k_2) n(q) + Z(k_1)[F(p) - p n(p)] + p^2 P_4 \)  

(33)
e. The first mixed moment. We find that

\[ \varepsilon_1(x_1x_2) = bn(k_2) \left[ b^2n(k_1; ak_2, b^2) - ak_2 F(k_1; ak_2, b^2) \right] \]

\[ \varepsilon_2(x_1x_2) = ab(1-k_2^2)(1-F_2) F_1 + (1-F_2)(b^2k_2 - a^2k_2- abq)n(q) \]

\[ \varepsilon_3(x_1x_2) = -k_1 Z(k_1) n(p) \]

\[ \varepsilon_4(x_1x_2) = k_1 p \ P_4 \]

Combining in the usual manner we obtain the regional expectation of the first mixed moment, namely

\[ \varepsilon(x_1x_2) = bn(k_2) \left[ b^2n(k_1; ak, b^2) - ak_2 F(k_1; ak_2, b^2) \right] + \]

\[ ab(1-k_2^2)(1-F_2) F_1 - k_1 Z(k_1) n(p) + k_1 p \ P_4 + \]

\[ (1-F_2)(b^2k_2 - a^2k_2 - abq)n(q) \]

f. Variance and Covariance of \( x_1 \) and \( x_2 \). By making use of previously derived equations, the variances and covariance of \( x_1 \) and \( x_2 \) can be easily obtained. Since

\[ \text{Var}(x_1) = \text{MSE}(x_1) - \varepsilon^2(x_1) \]

we have, after combining (21) and (29), that
\[
\begin{align*}
\text{Var}(x_1) &= \int_{-\infty}^{k_1} n(z) F\left[\frac{k_2}{b} - \frac{a}{b} z\right] \, dz - k_1 n(k_1) F(p) + \\
&\quad \text{an}(k_2) \left[ b^2 n(k_1; ak_2, b^2) - ak_2 F(k_1; ak_1, b^2) \right] + \\
&\quad (1-F_2) (2ab k_2 - b^2 q) n(q) + a^2 b^2 k_2^2 (1-F_2) F_1 + \\
&\quad k_1^2 Z(k_1) F(p) + k_1^2 P_4 - [n(k_1) F(p) - \text{an}(k_2) F(k_1; ak_2, b^2)] - \\
&\quad (1-F_2) (an(q) + bF_1) + k_1 P_4 + k_1 Z(k_1) F(p)]^2
\end{align*}
\]

By combining equations (25) and (33) in a similar manner, we obtain

\[
\begin{align*}
\text{Var}(x_2) &= \int_{-\infty}^{k_1} n(z) F\left[\frac{k_2}{b} - \frac{a}{b} z\right] \, dz - b^2 k_2 n(k_2) F(k_1; ak_2, b^2) - \\
&\quad ab^2 n(k_1) n(k_1; ak_2, b^2) + a^2 b^2 k_2^2 (1-F_2) F_1 + p^2 P_4 - \\
&\quad (qb^2 + 2abk_2) (1-F_2) n(q) + Z(k_1) F(p) - pZ(k_1) n(p) - \\
&\quad [bn(k_2) F(k_1; ak_2, b^2) - k_1 n(p) Z(k_1) + p P_4 + \\
&\quad (1-F_2) (ak_2 F_1 - bn(q))]^2
\end{align*}
\]

Now combining (37), (21) and (25), the covariance of \(x_1\) and \(x_2\) is obtained from (40).

\[
\text{Cov}(x_1x_2) = \varepsilon(x_1x_2) - \varepsilon(x_1) \varepsilon(x_2).
\]
Limiting expectation when \( k_2 \to \infty \). This limiting situation corresponds to the case treated in section A in which there is only one restricting plane located to the right of the population mean of 0. The following simplifications can now be made:

\[
F(k_1; ak_2, b^2) = 0 \\
n(k_1; ak_2, b^2) = 0 \\
P_4 = 0 \\
1-F_2 = 0 \\
F_1 = 1 \\
n(q) = 0
\]

Imposing these conditions on (21) and (25), we obtain

\[
\varepsilon(x_1) = -n(k_1) + k_1 [1-F(k_1)] \tag{41}
\]

and

\[
\varepsilon(x_2) = 0. \tag{42}
\]

If in (9), the mean is set to zero \( (x_1 = 0) \) and also recalling that in (9) to set \( k_2 = k_1 \) to maintain a consistent notation, it will be obvious that (9) and (41) are identical. Equation (42) is an obvious result and provides an algebraic check since if the second restricting plane is removed, the variable \( x_2 \) will become unrestricted and be univariately distributed \( N(0, 1) \).
2. $\alpha$ is acute

Many of the results derived for the previous section can be used directly. The expectations evaluated for regions $S_2$ and $S_3$ are applicable with no change. Expectations in $S_4$ also remain unchanged except that $P_4$ is different and has the value

$$P_4 = Z(p) - \int \int n(x_2) \ P\left[ \frac{\alpha}{b} x_2 - \frac{a}{b} \right] dx$$

instead of

$$P_4 = \int \int n(x_1) n(x_2) \ dx_1 \ dx_2$$

It remains to evaluate the expectations in $S_1$ and this will now be done. It follows directly that

$$\varepsilon_1(x_1) = \int \int x_1 n(x_1) n(x_2) \ dx_1 \ dx_2$$

$$\varepsilon_1(x_2) = \int \int x_2 n(x_1) n(x_2) \ dx_1 \ dx_2$$

$$\varepsilon_1(x_1^2) = \int \int x_1^2 n(x_1) n(x_2) \ dx_1 \ dx_2$$
\[ e_1(x_2^2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(x_1) n(x_2) \, dx_1 \, dx_2 \]

\[ e_1(x_1 x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(x_1) n(x_2) \, dx_1 \, dx_2 \]

where

\[ K = k_2 - \frac{b}{a} x_2. \]

After integrating, the above expectations will be

\[ e_1(x_1) = -n(k_1) F(p) + an(k_2) F(p; bk_2, a^2) \quad (43) \]

\[ e_1(x_2) = -n(p) F(k_1) + n(p) \left( F\left( \frac{k_2}{a} - \frac{b}{a} \right) - bn(k_2) F(p; bk_2, a^2) \right) \quad (44) \]

\[ e_1(x_1^2) = F(k_1) F(p) - k_1 n(k_1) F(p) - \int_{-\infty}^{p} n(z) F(k_2 \frac{b z}{a}) \, dz + \]

\[ a^2 k_2 n(k_2) F(p; bk_2, a^2) + a^2 bn(k_2) n(p; bk_2, a^2) \quad (45) \]

\[ e_1(x_2^2) = F(k_1) F(p) - pn(p) F(k_1) + pn(p) \left( F\left( \frac{k_2}{a} - \frac{b}{a} p \right) - \right) \]

\[ \int_{-\infty}^{p} n(z) F\left( \frac{k_2}{a} - \frac{b z}{a} \right) \, dz + b^2 k_2 n(k_2) F(p; bk_2, a^2) - a^2 n(k_2) n(p; bk_2, a^2) \quad (46) \]

\[ e_1(x_1 x_2) = n(k_2) n(p) + ab k_2 n(k_2) F(p; bk_2, a^2) - a^3 n(k_2) n(p; bk_2, a^2) \quad (47) \]
a. The first moment of $x_1$. Combining (19), (20), and (43) and since

$$
\varepsilon_1(x_1) = k_1 P_4
$$

we obtain

$$
\varepsilon(x_1) = -n(k_1) F(p) + an(k_2) F(p; bk_2, a^2) + k_1 P_4 + k_1 Z(k_1) F(p) - (1-F_2)(an(q) + bk_2 F_1)
$$

(48)

b. The first moment of $x_2$. Combining (23), (24), and (44) and since

$$
\varepsilon_1(x_2) = p P_4
$$

we find that

$$
\varepsilon(x_2) = -n(p) F(k_1) + n(p) F\left(k_2\frac{a}{a} + \frac{b}{p}\right) - bn(k_2) F(p; bk_2, a^2) + (1-F_2)(ak_2 F_1 - bn(q)) + k_1 n(p) Z(k_1) + p P_4
$$

(49)

c. The second moment and MSE of $x_1$. Combining (26), (27) and (45), we obtain

$$
\varepsilon(x_1^2) = \text{MSE}(x_1) = F(k_1) F(p) - k_1 n(k_1) N(p) - \int n(z) F\left(k_2\frac{a}{a} - \frac{b}{z}\right) dz + a^2 k_2 n(k_2) F(p; bk_2, a^2) + a^2 bn(k_2) n(p; bk_2, a^2) + (1-F_2)(2 ab k_2 - a^2 p) n(p) + (1-F_2)(a^2 + b^2 k_2^2) F_1 + k_1^2 F(p) Z(k_1) + k_1^2 P_4
$$

(50)
d. The second moment and MSE of $x_2$. Combining (30), (31), (32), and (46) we obtain

$$
\varepsilon(x_2^2) = \text{MSE}(x_2) = F(x_1)F(p) - pn(p) F(k_1) + pn(p) F\left(\frac{k_2 - b}{a}\right)
$$

$$
- \int_{-\infty}^{k} n(z) F\left(\frac{z - b}{a}\right) dz + b^2 k n(k_2) F(p; b, a^2) -
$$

$$
a^2 n(k_2) n(p; b, a^2) + (1-F_2)(b^2 + a_2 k_2) F_1 -
$$

$$
(1-F_2)(b_2^2 + 2ab k_2) n(q) + Z(k_1)\{N(p) - p n(p)\} + p^2 F_4
$$

(51)

e. The first fixed moment. Combining (34), (35), (36) (47) we find that

$$
\varepsilon(x_1 x_2) = n(k_2) n(p) + an(k_2)\{b k_2 F(p; b, a^2) - a^2 k(p; b, a^2)\}
$$

$$
ab(1-F_2)(1-k_2^2) F_1 + (1-F_2)(b_2^2 k_2 - abq - a^2 k_2) n(q) +
$$

$$
p k_1 F_4 - k_1 n(p) Z(k_1)
$$

(52)

f. Variance and Covariance of $x_1$ and $x_2$. To avoid repetition, the variances and covariances will not be explicitly stated. However, they can be obtained directly using (48), (49), (50), (51), and (52) and the fact that

$$
\text{Var}(x) = \varepsilon(x^2) - \varepsilon^2(x).
$$

The covariance can be obtained from the same equations and
the identity

\[ \text{Cov}(x_1, x_2) = \varepsilon(x_1 x_2) - \varepsilon(x_1) \varepsilon(x_2). \]

g. \textbf{Limiting expectations as } b \rightarrow 0. \text{ When } b \text{ approaches } 0, \text{ we have } a \text{ approaching } -1 \text{ and the convex region is contained between two parallel planes. Consequently the expectations should reduce to those obtained in section A. The following limiting conditions will be obtained}

\[
\begin{align*}
F(p) &= 1 & & F_4 = 0 \\
F(p; bk_2, a^2) &= 1 & & n(q) = 0 \\
1-F_2 &= F(k_2) & & F_1 = 1 \\
n(p) &= 0 & & p \rightarrow \infty
\end{align*}
\]

It must be kept in mind that because of a difference in notation between sections A and B certain changes must be carried out. Thus

\[
(k_1)_A = (k_2)_B \\
(k_2)_A = (k_1)_B
\]

We also find that the \( x_2 \) axis of section A becomes the \( x_1 \) axis of section B. And the \( x_1 \) axis of section A becomes the \( x_2 \) axis of section B. Consequently an interchange in expectations will follow.

If the above conditions are applied, we find that (48) reduces to (5) and (50) reduces to (8). In (5) and (8) the
mean must, of course, be zero. That is, in the limit when \( b = 0 \) the first two moments of \( x_1 \) for the case when \( \alpha \) is acute reduce to the first two moments obtained when there are 2 parallel restricting planes. The first two moments for \( x_2 \) become 0 and 1 respectively, and the \( \text{Cov}(x_1 \ x_2) \) becomes zero. These results for \( x_2 \) follow since \( x_1 \) and \( x_2 \) will be independently distributed in the limiting case and \( x_2 \) will be distributed \( N(0, 1) \).

h. Retransformation to original variables. Recall that the series of transformations from \( \beta \)-space to \( \gamma \)-space is given by

\[ \bar{\beta} = T_1 \ T_2 \ T_3 \ T_4 \ \gamma = T\gamma. \]

We then have that

\[ \varepsilon(\bar{\beta}-\beta) = T \begin{bmatrix} \text{bias in } \bar{\gamma}_1 \\ \text{bias in } \bar{\gamma}_2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = (\text{bias in } \bar{\gamma}_1) \ T_1 + (\text{bias in } \bar{\gamma}_2) \ T_2 \]

As before \( T_1 \) and \( T_2 \) are column vectors of the first and second columns of \( T \) respectively. It is apparent that the above equation allows immediate generalization to

\[ \varepsilon(\bar{\beta}-\beta) = \sum_{i=1}^{m} (\text{bias in } \bar{\gamma}_1) \ T_{i1} \]
where \( m \) is the number of inequality restrictions that the parameters are required to satisfy. In a similar manner, the variance-covariance matrix of the original variables can be written as:

\[
\begin{bmatrix}
\text{Var}(\tilde{\gamma}_1) & \text{Cov}(\tilde{\gamma}_1 \tilde{\gamma}_2) & 0 & \ldots & 0 \\
\text{Cov}(\tilde{\gamma}_1 \tilde{\gamma}_2) & \text{Var}(\tilde{\gamma}_2) & 0 & \ldots & 0 \\
0 & 0 & \ddots & \ddots & \vdots \\
0 & \ddots & \ddots & \ddots & \ddots \\
0 & \ddots & \ddots & \ddots & \ddots \\
\end{bmatrix}
\]

where

\[
h_{11} = \text{Var}(\tilde{\gamma}_1)-1 \\
h_{12} = \text{Cov}(\tilde{\gamma}_1 \tilde{\gamma}_2) \\
h_{22} = \text{Var}(\tilde{\gamma}_2)-1
\]

Finally we obtain the variance-covariance matrix as

\[
\sigma^2 \left\{ \Sigma + h_{11} T_1 T'_1 + h_{12} (T_1 T'_2 + T_2 T'_1) + h_{22} T_2 T'_2 \right\}
\]
As in the case of the bias, the generalization is immediate, namely with \( m \) restrictions, the variance-covariance matrix would be

\[
\sigma^2 \Sigma + \sigma^2 \sum_{i=1}^{m} \sum_{j=1}^{m} h_{ij} T_{i} T_{j}'
\]

where

\[
h_{ii} = \text{Var}(\tilde{Y}_i) - 1
\]

\[
h_{ij} = \text{Cov}(\tilde{Y}_i \tilde{Y}_j) \quad i \neq j
\]
IV. LINEAR COMBINATIONS OF $\beta$

With regard to linear combinations of quadratic programming estimators there are not many results. For example, it is well known [13] that the Least Squares estimate of $q'\beta$ will be $q'\hat{\beta}$, that is

$$\hat{q'\beta} = q'\hat{\beta}$$  \hspace{1cm} (1)

An analogous situation is not apparent for $\beta$ since, in general,

$$\hat{q'\beta} \neq q'\hat{\beta}$$  \hspace{1cm} (2)

In fact minimizing the residual sum of square with $q'\beta = k$ as a restriction will in general produce different estimates of $\beta$ than if the restriction were omitted from the minimization procedure.

A. An Alternative Estimator

Of general interest, although essentially trivial, is the fact that it is always possible to find a

$$\hat{\gamma} = \hat{q'\beta}$$

such that

$$\text{MSE}(\hat{\gamma}) \leq \text{Var}(\hat{\gamma})$$  \hspace{1cm} (3)

where

$$\hat{\gamma} = q'\hat{\beta}.$$
The procedure is as follows. Two linear combinations are obtained such that

\[ L_1 = \text{Min}(q'g) \quad \text{in } S \]  
\[ L_2 = \text{Max}(q'g) \quad \text{in } S \]  

The estimate of the linear combination, \( \tilde{y} \), will be

\[ \tilde{y} = \begin{cases} 
L_1 & \text{if } \hat{y} \leq L_1 \\
\hat{y} & \text{if } L_1 \leq \hat{y} \leq L_2 \\
L_2 & \text{if } \hat{y} > L_2 
\end{cases} \]  

Thus if the linear combination \( q'\hat{x} \) passes through the convex region then it would be selected as the estimator of \( y \) otherwise \( L_1 \) or \( L_2 \) would be selected. If the region \( S \) is simple to visualize then \( L_1 \) and \( L_2 \) can be obtained by inspection. If so obtained then \( S \) need not be convex. If \( S \) is a region in \( n \)-dimensions, where \( n \) is quite large, then \( L_1 \) and \( L_2 \) would be obtained using a linear programming technique. In this situation \( S \) must be convex. The result (5) for this estimator follows directly from (II, A and B).

**B. MSE of a Simple Linear Combination**

Consider the linear form

\[ y = q'\hat{x}. \]
Under certain circumstances it will be shown that

$$\text{MSE}(\hat{\gamma}) \leq \text{Var}(\hat{\gamma})$$

Suppose $\hat{\beta}$ lies partially in $S$ such that $\hat{\beta}_1(i = 3, \ldots, p)$ is acceptable but $\hat{\beta}_1$ and $\hat{\beta}_2$ are not acceptable. We have immediately that

$$\tilde{\beta}_1 = \beta_1 \text{ for } i = 3, \ldots, p$$

and only $\hat{\beta}_1$ and $\hat{\beta}_2$ are of direct concern. Consider the 2-dimensional plane containing $\hat{\beta}_1$ and $\hat{\beta}_2$. This plane can be drawn as shown in Figure 4 in which the intersection of the restricting plane with the $\hat{\beta}_1 \hat{\beta}_2$-plane is a line. The unacceptable region will be taken as the region lying to the right of the boundary and will be denoted by $\overline{S}$.

---

Figure 4. The Restrained Space
The density can always be transformed so that $\hat{\beta}_1$ is distributed $N(\beta_1; 1)$ and $\beta_2$ is distributed $N(\beta_2; 1)$. We also assume that the mean lies at the intersection of the linear form and the boundary. We have that

$$\hat{\gamma} = a_1 \hat{\beta}_1 + a_2 \hat{\beta}$$

and

$$\tilde{\gamma} = a_1 \tilde{\beta}_1 + a_2 \tilde{\beta}_2$$

The perpendicular distance from the above points to $\gamma$ is $(\hat{\gamma} - \gamma)$ and $(\tilde{\gamma} - \gamma)$. We shall show that

$$\varepsilon(\tilde{\gamma} - \gamma)^2 \leq \varepsilon(\hat{\gamma} - \gamma)^2.$$ 

We confine our interest to the boundary of $S$ (since $\beta = \hat{\beta}$ in $S$). Consider 2 planes cutting the density surface at a distance $\pm d$ from the mean. Due to symmetry

$$f[x_1, x_2 = \frac{1}{a_2} (k_2 - a_1 k_1) + d; k_1, \frac{1}{a_2} (k_2 - a_1 k_1)] =$$

$$f[x_1, x_2 = \frac{1}{a_2} (k_2 - a_1 k_1) - d; k_1, \frac{1}{a_2} (k_2 - a_1 k_1)] = g(x_1, k_1).$$

We obtain

$$(\hat{\gamma} - \gamma)^2 = F^2 \cos^2 \theta = (d-x)^2 \cos^2 \theta$$

when $\hat{\gamma}$ lies below $\gamma$. When $\tilde{\gamma}$ lies above $\gamma$ we obtain

$$(\tilde{\gamma} - \gamma)^2 = (d + x)^2 \cos^2 \theta.$$
We then have
\[ \varepsilon(\hat{\gamma} - \gamma)^2 = \cos^2 \phi \int (d-x_1)^2 g(x_1, k_1) dx_1 + \int (d+x)^2 g(x_1, k_1) dx_1. \]

Upon integration, we find that
\[ \varepsilon(\hat{\gamma} - \gamma)^2 = (1 + k_1^2 + d^2) \cos^2 \phi. \]

Considering \( \tilde{\gamma} \), we will obtain
\[ \varepsilon(\tilde{\gamma} - \gamma)^2 = \cos^2 \phi \int d^2 g(x_1, k_1) dx_1 + \int d^2 g(x_1, k_1) dx_1 \]
\[ = d^2 \cos^2 \phi. \]

Hence we find that
\[ \varepsilon(\hat{\gamma} - \gamma)^2 - \varepsilon(\gamma - \gamma)^2 = (1 + k_1^2) \cos^2 \phi \]
and it is apparent that the right hand side is always positive. Since this must hold for all values of \( d \) it follows that
\[ \text{MSE}(\tilde{\gamma}) \leq \text{Var}(\gamma). \]

If the mean is on the boundary but not at the intersection, the above results and conclusions again follow. This is so since the densities
\[ f(x_1, x_2 = \frac{1}{a_2} (k_2 - a_1 k_1) \pm d; k_1, \beta_2) = g(x_1, k_1) \]

and this case reduces to the one discussed above.

If the population mean is inside \( S, \beta_1 < k_1 \), we find in an analogous manner that

\[ \varepsilon(\bar{y} - \gamma)^2 - \varepsilon(\hat{\gamma} - \gamma)^2 = \cos^2 \theta \int_{\frac{x_1}{k_1}} \frac{-\frac{1}{2}(x-\beta_1)^2}{\sqrt{2\pi}} \, dx. \]

Since the right hand side is always positive it follows that \( \text{MSE}(\hat{\gamma}) \leq \text{Var}(\gamma) \). In summary then we have established the large sample property that if the population mean lies inside or on the boundary of \( S \), the mean square error of \( \hat{\gamma} \) will on the average be less than the variance of \( \hat{\gamma} \).
V. A NUMERICAL EXAMPLE

To illustrate the algorithms and theorems that have been presented, an artificially constructed example will be presented. Let

\[
\begin{bmatrix}
0.4 \\
0.2 \\
1.5 \\
0.3 \\
0.8 \\
-0.2 \\
0.2 \\
-0.1 \\
0.6 \\
-0.2 \\
\end{bmatrix}
\quad
\begin{bmatrix}
0 & 1 & 0 \\
0 & 1 & 0 \\
2 & 1 & 0 \\
0 & 1 & 1 \\
1 & 1 & 1 \\
0 & 0 & 2 \\
0 & 1 & 0 \\
0 & 0 & 1 \\
0 & 2 & 0 \\
0 & 0 & 1 \\
\end{bmatrix}
\]

The normal equations are

\[
\begin{bmatrix}
5 & 3 & 1 \\
3 & 10 & 2 \\
1 & 2 & 8 \\
\end{bmatrix}
\begin{bmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\hat{\beta}_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
3.8 \\
4.6 \\
0.4 \\
\end{bmatrix}
\]

(1)

The estimates are restricted to satisfy the inequalities

\[
0 \leq \hat{\beta}_1 \leq 1 \\
0 \leq \hat{\beta}_2 \leq 1 \\
0 \leq \hat{\beta}_3 \leq 1 \\
0 \leq \hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3 \leq 1.
\]

(2a)  (2b)  (2c)  (2d)
It follows directly that the Least Squares solution is

\[
\begin{bmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2 \\
\hat{\beta}_3
\end{bmatrix}
= \begin{bmatrix}
5 & 3 & 1 \\
3 & 10 & 2 \\
1 & 2 & 8
\end{bmatrix}^{-1}
\begin{bmatrix}
3.8 \\
4.6 \\
0.4
\end{bmatrix}
\]

\[
= \frac{1}{310}
\begin{bmatrix}
76 & -22 & -4 \\
-22 & 39 & -7 \\
-4 & -7 & 41
\end{bmatrix}
\begin{bmatrix}
3.8 \\
4.6 \\
0.4
\end{bmatrix}
= \begin{bmatrix}
0.6 \\
0.3 \\
-0.1
\end{bmatrix}
\]  

(3)

All inequalities excepting one, 2c, are satisfied. Several numerical procedures will now be used to obtain the \( \hat{\beta} \) estimators.

A. Transformation and Projection

When only one or two restrictions are not satisfied it may be more convenient to by-pass quadratic programming algorithms and linearly transform the regression coefficients. Let

\[
\eta = C\beta
\]  

(4)

where \( C \) is an upper triangular matrix and is defined as

\[
X'X = C'C.
\]

This decomposition of \( X'X \) is a well-known application of the Choleski or square root transformation of a real symmetric square matrix. It readily follows that
In the transformed space we have new regression coefficients

\[ \eta_1 = \sqrt{5} \beta_1 + \frac{3\beta_2}{\sqrt{5}} + \frac{\beta_3}{\sqrt{5}} \]  
\[ (5a) \]

\[ \eta_2 = \frac{\sqrt{41} \beta_2}{\sqrt{5}} + \frac{7\beta_3}{\sqrt{205}} \]  
\[ (5b) \]

\[ \eta_3 = \frac{\sqrt{310} \beta_3}{\sqrt{41}} \]  
\[ (5c) \]

subject to the restrictions

\[ 0 \leq \frac{\eta_1}{\sqrt{5}} - \frac{3\eta_2}{\sqrt{205}} - \frac{4\eta_3}{\sqrt{12710}} \leq 1 \]  
\[ (6a) \]

\[ 0 \leq \frac{\sqrt{5}\eta_2}{\sqrt{41}} - \frac{3\eta_3}{\sqrt{12710}} \leq 1 \]  
\[ (6b) \]

\[ 0 \leq \frac{\sqrt{41}\eta_3}{\sqrt{310}} \leq 1 \]  
\[ (6c) \]

\[ 0 \leq \frac{\eta_1}{\sqrt{5}} + \frac{2\eta_2}{\sqrt{205}} + \frac{20\eta_3}{\sqrt{12710}} \leq 1 \]  
\[ (6d) \]
From (5a), (5b), (5c), the Least Squares estimates of the transformed regression coefficients can be obtained. We find that

$$\hat{\eta}_1 = \frac{3.8}{\sqrt{5}}; \hat{\eta}_2 = \frac{11.6}{\sqrt{205}}; \hat{\eta}_3 = -0.1 \frac{310}{41}$$

It can be verified that all transformed restraints are satisfied except (6c), that is, $\hat{\eta}_3$ must be positive. The minimum distance from the point $\hat{\eta}$ in the $S_\eta$ space is obtained by perpendicular projection. This will be accomplished by inspection to give

$$\tilde{\eta}_1 = \hat{\eta}_1 = \frac{3.8}{\sqrt{5}}; \tilde{\eta}_2 = \hat{\eta}_2 = \frac{11.6}{\sqrt{205}}; \tilde{\eta}_3 = 0.$$ 

Retransforming to the original variables, we obtain the restricted solution given below.

$$\tilde{\eta} = \mathbf{c}^{-1} \tilde{\eta}$$

$$\begin{bmatrix}
\tilde{\beta}_1 \\
\tilde{\beta}_2 \\
\tilde{\beta}_3
\end{bmatrix} =
\begin{bmatrix}
\frac{1}{\sqrt{5}} & \frac{-3}{\sqrt{205}} & \frac{-4}{\sqrt{12710}} \\
0 & \frac{1}{\sqrt{41}} & \frac{-7}{\sqrt{12710}} \\
0 & 0 & \frac{1}{\sqrt{310}}
\end{bmatrix}
\begin{bmatrix}
3.8 \\
\frac{11.6}{\sqrt{205}} \\
0
\end{bmatrix} =
\begin{bmatrix}
\frac{121}{205} \\
\frac{11.6}{41} \\
0
\end{bmatrix} =
\begin{bmatrix}
0.590,244 \\
0.282,927 \\
0
\end{bmatrix} (7)$$
It is apparent above that the $\tilde{\beta}$ estimators have all been modified. One would be tempted to use the estimates

$$\bar{\beta}_1 = \hat{\beta}_1 \quad (8a)$$
$$\bar{\beta}_2 = \hat{\beta}_2 \quad (8b)$$
$$\bar{\beta}_3 = 0 \quad (8c)$$

Although this procedure provides estimates that are in $S$, the criteria of minimum residual sum of squares is not satisfied. See the example illustrating this point that is given below.

The increase in the residual sum of squares can be easily found since

$$Q_{Q.P.} - Q_{L.S.} = (\hat{\eta} - \tilde{\eta})' (\hat{\eta} - \tilde{\eta}) = \sum_{i=1}^{3} (\hat{\eta}_i - \tilde{\eta}_i)^2$$

$$= 0 + 0 + \left(0.1 \sqrt{\frac{310}{41}}\right)^2 = \frac{310}{41} = 0.0756,098$$

In terms of the original regression coefficients, the increase would be the same but would be computed in the following way:

$$Q_{Q.P.} - Q_{L.S.} = (\hat{\beta} - \tilde{\beta})' S (\hat{\beta} - \tilde{\beta}) = .0756,098 \quad (9)$$

since
\[
\begin{bmatrix}
.009,756 \\
.017,073 \\
-.1
\end{bmatrix}
\begin{bmatrix}
5 & 3 & 1 \\
3 & 10 & 2 \\
1 & 2 & 8
\end{bmatrix}
\begin{bmatrix}
.009,756 \\
.017,073 \\
-.1
\end{bmatrix}
\]

\[
\begin{bmatrix}
-.000,001 \\
-.000,002 \\
-.756,098
\end{bmatrix}
\begin{bmatrix}
.009,756 \\
.017,073 \\
-.1
\end{bmatrix}
\approx 0 + 0 + 0.0756,098
\]

When the negative estimates are set to zero (8a), (8c) the residual sum of squares will be denoted by \( \bar{Q} \) and we have

\[
\bar{Q} - Q_{L.S.} = [0, 0, .1] \begin{bmatrix} 5 & 3 & 1 \\ 3 & 10 & 2 \\ 1 & 2 & 8 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ .1 \end{bmatrix} = 0.08 \quad (10)
\]

Comparing (9) and (10) we see that

\[
Q_{Q.P.} \leq \bar{Q}
\]

It should be noted that \( Q_{Q.P.} = \bar{Q} \) when orthogonal polynomials are being used to fit the data. That is, all negative regression coefficients are set to zero and all positive least squares regression coefficients retained.

To numerically illustrate that \( \tilde{\beta} \) does not depend upon which linear transformation is used, let us consider a second linear transformation of the \( \beta \)'s. It will be shown that this transformation will provide estimates when transformed back
to the $\beta$-space which will be exactly $\tilde{\beta}_1$, $\tilde{\beta}_2$, $\tilde{\beta}_3$ as given above.

Let $b = T\tilde{\beta}$

where

$$T = \begin{bmatrix} 1.290,994 & 2.427,876 & 2.128,020 \\ 1.581,138 & -1.076,162 & -0.029,478 \\ 0.912,871 & 1.716,768 & -1.862,972 \end{bmatrix}$$

and

$$T^{-1} = \begin{bmatrix} 0.116,742 & 0.464,387 & 0.126,004 \\ 0.165,771 & -0.246,932 & 0.193,263 \\ 0.209,967 & 0 & -0.296,938 \end{bmatrix} \quad (11)$$

As in the previous transformation, we again require that $X'X = T'T$.

The transformed restrictions become

$$0 \leq 0.116,743b_1 + 0.464,387b_2 + 0.126,004b_3 \leq 1 \quad (12a)$$
$$0 \leq 0.165,771b_1 - 0.246,932b_2 + 0.193,263b_3 \leq 1 \quad (12b)$$
$$0 \leq 0.209,967b_1 - 0.296,938b_3 \leq 1 \quad (12c)$$
$$0 \leq 0.492,481b_1 + 0.217,455b_2 + 0.022,329b_3 \leq 1 \quad (12d)$$

and the Least Squares estimates, obtained from (11) are

$$\hat{b}_1 = 1.290,157$$
$$\hat{b}_2 = 0.628,782$$
$$\hat{b}_3 = 1.249,050$$

Again we find that the third restriction, equation (12c) is not satisfied since
The Figure 5 below will aid in the discussion.

The line

\[ b_1 = 2.173,369 - \frac{0.209,967}{0.296,938} b_3 \]

passes through \((\hat{b}_1, \hat{b}_2)\) and is perpendicular to the restricting equation

\[ 0.209,967b_1 - 0.296,938b_3 = 0 \]  (13)

At the point of intersection of these two lines we have

\[ \frac{0.296,938}{0.209967} b_3 = 2.173,369 - \frac{0.209,967}{0.296,938} b_3. \]

Therefore \(\hat{b}_3 = 1.024,536\).
With $\beta_3$ known, $\tilde{\beta}_1$ is obtained from (3). Consequently we have the following modified Least Squares estimators.

$$\begin{align*}
\tilde{\beta}_1 &= 1.448,912 \\
\tilde{\beta}_2 &= \tilde{\beta}_2 \\
\tilde{\beta}_3 &= 1.024,536
\end{align*}$$

Retransforming back, we have that

$$\bar{\beta} = T^{-1}\tilde{\beta}$$

$$\begin{bmatrix}
\tilde{\beta}_1 \\
\tilde{\beta}_2 \\
\tilde{\beta}_3
\end{bmatrix} =
\begin{bmatrix}
0.116,743 & 0.464,387 & 0.126,004 \\
0.165,771 & -0.246,932 & 0.193,263 \\
0.209,967 & 0 & -0.296,938
\end{bmatrix}
\begin{bmatrix}
1.448,912 \\
.628,782 \\
1.024,536
\end{bmatrix}$$

From which we obtain the same estimates that were calculated previously, see (7), namely

$$\begin{bmatrix}
\tilde{\beta}_1 \\
\tilde{\beta}_2 \\
\tilde{\beta}_3
\end{bmatrix} =
\begin{bmatrix}
0.590,244 \\
0.282,926 \\
0
\end{bmatrix}$$

B. Quadratic Programming Solution

The quadratic programming solution is obtained as a solution of three successive linear programming problems in which different side conditions are imposed. An excellent exposition of the algorithm used here is given [21] which the reader may wish to consult. We shall proceed directly to the formulation of the problem and its solution.
Restrictions (3a, b, c, d) can be rearranged in the below form.

\begin{align*}
0 & \leq \beta_1 \quad (14a) \\
0 & \leq \beta_2 \quad (14b) \\
0 & \leq \beta_3 \quad (14c) \\
0 & \leq \beta_4 \quad (14d)
\end{align*}

A slack variable \( \beta_4 \) has been introduced to convert inequality (3d) to an equation (4e). Because of the slack variable, the quadratic form to be minimized is modified slightly. Thus the solution vector is sought which minimizes

\[
\begin{bmatrix}
-0.38, -4.6, -0.4, 0
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{bmatrix}
+ \begin{bmatrix}
5 & 3 & 1 & 0 \\
3 & 10 & 3 & 0 \\
1 & 2 & 8 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4
\end{bmatrix}
\]

Note that the algorithm automatically assures that the first four conditions, the requirement of positive estimates, are satisfied. Only condition (4e) is explicitly used.

Consider the first LP problem, Table 1, in which the objective function to be minimized to zero is

\[
\sum_{i=1}^{1} w_i = w
\]

since \( i = 1 \) (one restriction, 14d).

Initial basis is formed from the coefficients of \( z^1 \), \( z^2 \) and \( w \). In addition, \( v, u, \) and \( \mu \) are not allowed to enter the
basis. When w is driven to zero, the columns containing unused components of $z^1$ and $z^2$ are discarded. The w column is also discarded and Table 2 is obtained.

Commencing with Table 2, the second LP problem is to minimize to zero

$$\Sigma z = z^1_1 + z^1_2$$

(16)

under the restrictions that $v'\beta = 0$ and $\mu$ must not enter the basis.

If Table 2 is examined, it can be seen that all the above conditions are satisfied. Thus in obtaining Table 2 as the solution to problem 1 we have also obtained the solution to problem 2. We now proceed to the third LP problem.

Minimize the objective function

$$-\mu$$

(17)

in Table 2 subject to $v'\beta = 0$. Also $z^1_1$ and $z^1_2$ are not allowed to enter the basis once they are driven out. This condition is equivalent to discarding the columns for $z^1_1$ and $z^1_2$ which shall be done. Only three basis changes are required and the Simplex computation has been completed since $\mu > 1$. Table 3 summarizes the results.

Since the solution is desired for $\lambda = 1$ linear interpolation between iteration 5 and 6 will be required. The linear interpolation formula of Wolfe, his equation 21, is
for
\[ \mu^k \leq \lambda \leq \mu^{k+1} \]

\[ x = \frac{\mu^{k+1} - \lambda}{\mu^{k+1} - \mu} x^k + \frac{\lambda - \mu^k}{\mu^{k+1} - \mu} x^{k+1} \]  

(18)

Note that the exponent denotes the iteration count. For the example we have

\[ k = 5 \]
\[ \lambda = 1 \]
\[ \mu^k = 0 \]
\[ \mu^{k+1} = 41/35.8 \]

and the interpolation formula (19) simplifies to

\[ x = \frac{35.8}{41} x^{k+1} \]  

(20)

Using (20) and the pertinent values from Table 3 we obtain

\[ \beta_1 = \frac{35.8}{41} \cdot \frac{12.1}{17.9} = \frac{24.2}{41} = 0.590244 \]

\[ \beta_2 = \frac{35.8}{41} \cdot \frac{5.8}{17.9} = \frac{11.6}{41} = 0.282927 \]

Since \( v_3 \) is in the basis at a non-zero value \( \beta_3 \) must be 0. These are the same estimates of \( \beta_1, \beta_2, \beta_3 \) that were obtained in the previous section.
Table 1. The Initial Simplex Tableau

| Basis | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\beta_4$ | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $u$ | $\mu$ | $z_1^1$ | $z_2^1$ | $z_3^1$ | $z_4^1$ | $z_1^2$ | $z_2^2$ | $z_3^2$ | $z_4^2$ | $w$ |
|-------|-----------|-----------|-----------|-----------|-----------|-------|-------|-------|-------|-----|-------|--------|--------|--------|--------|--------|--------|--------|--------|
| $w$   | 1         | 1         | 1         | 1         | 1         | 1     | 0     | 0     | 0     | 0   | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 1      |
| $z_1^1$ | 0         | 5         | 3         | 1         | 0         | -1    | 0     | 0     | 0     | 1   | -3.8  | 1      | 0      | 0      | 0      | -1     | 0      | 0      | 0      |
| $z_2^1$ | 0         | 3         | 10        | 2         | 0         | -1    | 0     | 0     | 0     | 1   | -4.6  | 0      | 1      | 0      | 0      | 0      | -1     | 0      | 0      |
| $z_3^1$ | 0         | 1         | 2         | 8         | 0         | 0     | 0     | 0     | -1    | 0   | -0.4  | 0      | 0      | 1      | 0      | 0      | 0      | -1     | 0      |
| $z_4^1$ | 0         | 0         | 0         | 0         | 0         | 0     | 0     | 0     | 0     | -1  | 0     | 0      | 0      | 0      | 1      | 0      | 0      | 0      | -1     | 0      |
Table 2. Simplex Tableau After the First LP Solution

<table>
<thead>
<tr>
<th>Basis</th>
<th>$P_0$</th>
<th>$\tilde{P}_1$</th>
<th>$\tilde{P}_2$</th>
<th>$\tilde{P}_3$</th>
<th>$\tilde{P}_4$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
<th>$v_4$</th>
<th>$u$</th>
<th>$u$</th>
<th>$z^1_1$</th>
<th>$z^1_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{P}_4$</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>-7</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$z^1_1$</td>
<td>0</td>
<td>0</td>
<td>-7</td>
<td>-39</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>5</td>
<td>-4</td>
<td>0</td>
<td>-1.8</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$z^1_2$</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>-22</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>3</td>
<td>-2</td>
<td>0</td>
<td>-3.4</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>$\tilde{P}_1$</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>-0.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$u$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Table 3. A Summary of the Solution

<table>
<thead>
<tr>
<th>Iteration</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$v_1$</th>
<th>$v_2$</th>
<th>$v_3$</th>
<th>$v_4$</th>
<th>$u$</th>
<th>$\mu$</th>
<th>$z_1$</th>
<th>$z_2$</th>
<th>$z_3$</th>
<th>$z_4$</th>
<th>$w$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>12.1</td>
<td>5.8</td>
<td>15.5</td>
<td>0</td>
<td>$\frac{41}{35.8}$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ \frac{17.9}{17.9} \]
C. The Asymptotic Bias and the Variance

From (III, 15 and 16) we have shown that

Bias is $\bar{\beta} = T$ (Bias in transformed orthogonal space)

Recall that for our example we had made the transformation

$$\beta = C^{-1}\eta.$$  \hfill (21)

For calculation of the bias, however, we require an orthogonal transformation from the original parameter space to an orthogonal space which allows for the projecting of the probability mass.

So let

$$\eta = 0_\gamma$$  \hfill (22)

be the required orthogonal transformation. Combining (21) and (22) we have

$$\beta = C^{-1}0_\gamma T = T_\gamma.$$  \hfill (23)

From (III, 16), the asymptotic covariance matrix was given as

$$\frac{\sigma^2}{N}\left\{\Sigma + [\text{Var}(\gamma) - 1] T A T'\right\} = \frac{\sigma^2}{N} (\Sigma + [\text{Var}(\gamma) - 1] T_3 T'_3$$

$$= \frac{\sigma^2}{N}\left\{\Sigma - \frac{1}{2} (1 + \frac{1}{\pi}) T_3 T'_3\right\}$$  \hfill (24)

since from (III, Section 4), the variance was shown to be

$$\text{Var}(\gamma) = \frac{\sigma^2}{N} \left\{\frac{1}{2} (1 - \frac{1}{\pi})\right\}$$

if the population mean is on the boundary.
Similarly, the asymptotic bias is obtained from (III, 15) as
\[
(bias \text{ is } \beta) = -\frac{1}{\sqrt{2\pi}} T.3
\] (25)

Hence from (24), the asymptotic variance of the i-th regression coefficient is
\[
\text{Var}(\beta_i) = \frac{s^2}{N} \left\{ \text{Var}(\beta_i) - (1 + \frac{1}{n}) t^2_{i3} \right\}.
\] (26)

In this form the reduction in variance is explicitly shown.

D. Estimation of the Variance

Since the estimation of variances and covariances is an a posterior operation, it is not appropriate to disregard information about the true parameter value \( \gamma_0 \) which is contained in \( \hat{\gamma} \) or \( \hat{\gamma}^* \). Inference about \( \gamma_0 \) may be based on whether or not a regular boundary point is in the confidence sphere of \( \gamma_0 \) around \( \hat{\gamma} \). Hence we have two cases:

Case I. If \( \hat{\gamma} \) permits the inference that \( \gamma_0 \) is an interior point, estimate
\[
\frac{s^2}{N} \text{ by }
\]
\[
\frac{s^2}{N} = (y'y - \hat{y}'\hat{y})/(n-p)
\] (27)
and use
\[
e \left\{ (\gamma - \gamma_0) (\gamma - \gamma_0)' \right\} = \frac{s^2}{N} I
\]
or for \( \beta \)
\[
e \left\{ (\beta - \beta_0) (\beta - \beta_0)' \right\} = \frac{s^2}{N} S^{-1}
\] (28)
That is $\hat{\beta}$ is the quadratic programming estimator but the estimate of the variances and covariance and $s^2$ are the usual Least Squares estimates.

Case II. If $\hat{Y}$ permits the inference that $Y_0$ is a regular boundary point, estimate $\frac{\sigma^2}{N}$ from (27) and use for the mean square and product matrix

$$\varepsilon \{(\hat{Y} - Y_0) (\hat{Y} - Y_0)\}' = \frac{\sigma^2}{N} (I_p - \frac{1}{2} vv')$$

(29)

In the original parameter space (29) becomes

$$\varepsilon \{(\hat{\beta} - \beta_0) (\hat{\beta} - \beta_0)\}' = \frac{\sigma^2}{N} \left\{s^{-1} - \frac{1}{2} (C^{-1}v)(C^{-1}v)' \right\}$$

(30)

where the vector $v$ is determined as follows. We have

$$TAT' = (C^{-1}0') A (C^{-1}0')'$$

(31)

The matrix $A$ has all zero elements except $a_{33} = 1$ so that

$$A = A'$$

and also

$$A = A^2 = A' A = A A'.$$

We then have

$$TAT' = C^{-1}0' A' A 0 C^{-1} = C^{-1}(0' A') (A0) (C^{-1})'$$

$$= C^{-1} vv' (C^{-1})' = (C^{-1}v) (C^{-1}v)' .$$

So $v$ is a vector of unit length normal to the tangential plane through a point on the boundary which is $\hat{Y}$ and nearest $\hat{Y}$. If $\hat{Y}$ is outside the parameter space then
For our example we have

\[ \tilde{\gamma}_1 = \hat{\gamma}_1 \]

\[ \tilde{\gamma}_2 = \hat{\gamma}_2 \]

\[ \tilde{\gamma}_3 \neq \hat{\gamma}_3 \]

so that (32) becomes

\[
v = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}
\]

and

\[
C^{-1}v = \begin{bmatrix}
\frac{4}{\sqrt{12710}} \\
\frac{7}{\sqrt{12710}} \\
-\frac{41}{\sqrt{310}}
\end{bmatrix}
\]

The adjustment in the variance-covariance matrix becomes

\[
\frac{1}{2}(C^{-1}v)(C^{-1}v)' = \frac{1}{2}
\begin{bmatrix}
\frac{16}{12710} & \frac{28}{12710} & -\frac{4}{310} \\
\frac{28}{12710} & \frac{49}{12710} & -\frac{7}{310} \\
-\frac{4}{310} & -\frac{7}{310} & \frac{41}{310}
\end{bmatrix}
\]
Since

\[ S^{-1} = C^{-1}(C^{-1})' = \begin{bmatrix}
\frac{3116}{12710} & \frac{-902}{12710} & \frac{-4}{310} \\
\frac{-902}{12710} & \frac{1599}{12710} & \frac{-7}{310} \\
\frac{-4}{310} & \frac{-7}{310} & \frac{41}{310}
\end{bmatrix} \]

the variance-covariance matrix of the regression coefficient (30) for our example becomes

\[
\frac{s^2}{N} \left( S^{-1} - \frac{1}{2}(C^{-1}v)(C^{-1}v)' \right) = \begin{bmatrix}
\frac{3108}{12710} & \frac{-930}{12710} & \frac{-2}{310} \\
\frac{-930}{12710} & \frac{1574.5}{12710} & \frac{-3.5}{310} \\
\frac{-2}{310} & \frac{-3.5}{310} & \frac{20.5}{310}
\end{bmatrix} \frac{s^2}{N}
\]
VI. FITTING POLYNOMIALS WITH NON-NEGATIVE DERIVATIVES

This application of constrained Least Squares is of general interest in several areas. We shall develop in great detail the Least Square estimation of the coefficient of a cubic with a non-negative first derivative. For the fitting of higher order polynomials, several interesting aspects will be presented. The fitting of polynomials, as an approximation to a more awkward or unknown function, simplifies the Least Square estimation. However, at the same time the approximation may be acceptable only if the fitted polynomial has a positive derivative. The application we have in mind is Edgeworth's "method of translation" as described by Buehler [2]. Specifically, a cubic variate transformation is desired which must be monotonic. Consequently, the cubic must have a non-negative first derivative over the whole variate range. It is mentioned by Buehler that the non-monotonicity of the cubic variate transformation (change in sign of the derivative) represents a shortcoming of this approach. In fact, the example illustrating the method actually had a negative derivative just beyond the range of interest of the variate so the polynomial must have reflected this. In what follows, Buehler's example will be reworked as an illustration of constrained estimation so that a non-negative derivative is assured over a greater range of the variate.
A. Special Case of the Cubic

Consider the cubic and its derivative

\[ y(x) = a + bx + cx^2 + dx^3 \]  \hspace{1cm} (1)

\[ y'(x) = b + 2cx + 3dx^2 \]  \hspace{1cm} (2)

For \( y'(x) \) to be positive for all values of \( x \), we immediately know that

\[ b \geq 0 \]  \hspace{1cm} (3a)

\[ d \geq 0 \]  \hspace{1cm} (3b)

For the derivative to be positive, the discriminant of the quadratic polynomial describing the derivative must be imaginary. The discriminant condition apart from a constant can be written as

\[ c^2 - 3bd < 0 \]  \hspace{1cm} (4)

Note that if either \( b \) or \( d \) are identically zero, it is impossible to satisfy (4). We have now achieved a necessary and sufficient condition for (2) to be positive, that is,

\[ b > 0 \]  \hspace{1cm} (5a)

\[ d > 0 \]  \hspace{1cm} (5b)

\[ c^2 - 3bd < 0 \]  \hspace{1cm} (5c)

The constraints must be convex if the quadratic programming algorithm of Hartley and Hocking [10] is to be used. Obviously, (5a) and (5b) are convex. It remains to be shown that (5c) is also convex.
Theorem: If \( b > 0 \) and \( d > 0 \) the imaginary discriminant is convex. Given that \( b > 0 \), \( d > 0 \) and that
\[
g_1 = c_1^2 - 3b_1d_1 < 0 \quad (6a)
\]
\[
g_2 = c_2^2 - 3b_2d_2 < 0 \quad (6b)
\]
it must be shown that
\[
(\alpha c_1 + 3c_2)^2 - 3(\alpha b_1 + 3b_2)(\alpha d_1 + 3d_2) < 0 \quad (7)
\]
where
\[
0 < \alpha < 1, \quad 0 < \beta < 1, \text{ and } \alpha + \beta = 1.
\]
Expanding (7) and rearranging we obtain
\[
\alpha^2g_1 + \beta^2g_2 + \alpha \beta (2c_1c_2 - 3b_1d_2 - 3d_1b_2).
\]
The first two terms are always negative by hypothesis. Now consider the third term. Multiplying (6a) by (6b) and simplifying gives
\[
c_1^2c_2^2 \leq 9b_1b_2d_1d_2
\]
or
\[
|c_1c_2| \leq 3 \left( (b_1d_2) (b_2d_1) \right)^{1/2} \leq \frac{3}{2} (b_1d_2 + b_2d_1)
\]
since the arithmetic mean exceeds the geometric means. Hence the third term is never positive and the theorem is established.

The present approach of using the discriminant is, of course, confined to the cubic. In the next section we shall develop a general method applicable to any polynomial. Buehler's example, although it is a cubic, is used as an illustration there.
B. Approximate Solution for any Polynomial

1. Some convergence proofs

We shall confine our attention to a useful property of an approximate solution which is characterized by confining \( x \) to a finite range.

Suppose we have the argument \( x \) confined to a finite range, say

\[-X \leq x \leq X \] (8)

and in the interval we desire a non-negative derivative, i.e.

\[ y'(x) \geq 0 \] (9)

We shall impose the inequality (9) at a finite set of grid points covering the range from \(-X\) to \(X\) and will let the grid interval tend to zero.

Let a partition, called the Nth partition, of the interval (8) such that

\[ x_t = tx^{2^{-N}} \]

where;

\[ t = -2^N, -2^{N+1}, -2^N + 2, \ldots, 0, \ldots, 2^N - 2, 2^{N-1}, 2^N. \]

This yields \( 1 + 2^N \) linear restraints upon the coefficients of the polynomial. We wish to minimize

\[ Q_N(\beta) = \sum_{j=1}^{n} \left( y_j - \sum_{i=0}^{k} \beta_i x_j^i \right)^2 \] (10)

where the \( j \)-tuples \((y_j, x_j)\) are the observed "responses" and
"input factors" subject to (9) which is written as

\[ y'(x_t) = \sum_{i=1}^{k} \beta_i x_t^{i-1} \geq 0 \]  

(11)

Suppose that another partition, call it the (N+1)-st partition, is made. Specifically, let

\[ x_{t'} = t' x_2^{-1}(N+1) \]

where:

\[ t' = -2^{N+1}, -2^{N+1}-1, -2^{N+1}-2, \ldots, 2^{N+1}-2, 2^{N+1}-1, 2^{N+1}. \]

The (N+1)-st partition produces \(1 + 2^N\) values of \(x_t\), that are identical with the \(x_t\) values given in the N-th partition. The additional \(2^N\) points are located half-way between the pair of points of the N-th partition as shown in Figure 6.

\[ \begin{align*}
\cdots & \cdots \\
 x_2 & x_4 & x_6 & x_8 & \cdots \\
\cdots & \cdots 
\end{align*} \]

Figure 6. Partition of the Interval

Thus, the (N+1)-st partition of the range of \(x\) contains the N-th partition. We now have the \(1 + 2^N\) linear restraint upon the derivatives that were obtained from the N-th partition plus the \(2^N\) linear restraints obtained from the "half way" points. Because of the (N+1)-st minimization problem contains all the restraints of the N-th problem, the polynomial yielding the minimum \(Q_{N+1}\) is a competitor admitted to the N-th
problem, i.e.

\[ Q_N \leq Q_{N+1} \]

Suppose we truncate the sequence of minimizations at some value. We have produced the sequence of solutions

\[ Q_1 \leq Q_2 \leq \cdots \leq Q_N \leq Q_{N+1} \leq \cdots \leq Q_k \]  

(12)

where \( Q_k \) represents the last term of the finite sequence, in which every solution satisfies the requirements of a non-negative derivative.

We wish to show that

\[ \lim_{k \to \infty} Q_k = Q^* \]  

(13)

That is the sum of squares corresponding to the truncated solution can be made as close to the true solution \( Q^* \) as one desires by selecting a sufficiently fine partition of the finite interval.

First we show that the sequence (12) is bounded, that is

\[ Q_N \leq L \text{ for all } N \]  

(14)

Proof: Select any polynomial, \( P^+(x) \), which has a positive derivative over the range \(-X \leq x \leq X\). For example, the polynomial

\[ P^+(x) = 2x \]

will suffice. Now \( P^+(x) \) is a possible competitor for the
minimization problem for any \( N \), that is for all partitions of the interval. Hence, it immediately follows that

\[
Q_N \leq Q_P^+(x) = L \tag{15}
\]

The inequality (15) shows that the non-decreasing sequence of residual sums of squares is bounded. Hence (12) becomes

\[
Q_1 \leq Q_2 \leq \ldots \leq Q_N \leq \ldots \leq Q_k \leq L \tag{16}
\]

and (13) follows immediately.

Now consider the behavior of the sequence of estimators corresponding to the sequence of partitions. We wish to show that

\[
\lim_{k \to \infty} \beta_1(k) = \beta^*_1 \tag{17}
\]

First we show that the \( \beta_1(N) \) is bounded. Rewriting the quadratic form (10) as

\[
(y - x\hat{\beta})'(y - x\hat{\beta}) = (y - x\hat{\beta})'(y - x\hat{\beta}) + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)
\]

we see that provided \( X'X \) is non-degenerate, that the quadratic form is positive definite in the \( \hat{\beta}_1 \), it follows that the statement

\[
Q_N [\beta_1, \beta_2, \ldots] \leq L \text{ for all } N
\]

implies the existence of a constant, \( M \), such that

\[
|\beta_1(N)| \leq M \text{ for all } i \text{ and } N.
\]

Employing the Bolzano-Weirestrass Theorem, we know that there
is at least one limit point in the bounded closed interval $[-M, M]$. Any such limit point constitutes a solution of the problem and (17) follows immediately. It has not been shown that a unique limiting solution exists and this may be quite difficult to verify since the restraints do not enter into the solution explicitly.

2. **Buehler's example reworked**

Consider the polynomial

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + e$$  \hspace{1cm} (18)

Because of the correlated structure of the error $e$, Buehler used the weight Least Square procedure of Aitken. Thus, the normal equations are written as:

$$(X'WX) \beta = X'Wy$$  \hspace{1cm} (19)

where $W$ is the matrix of weight. The normal equations given by Buehler are:

$$
\begin{bmatrix}
.95406 & .83558 & 2.5221 & 4.3259 \\
.83558 & 4.1751 & 5.7359 & 29.630 \\
2.5221 & 5.7359 & 33.243 & 54.764 \\
4.3259 & 29.630 & 54.764 & 424.69 \\
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\beta_2 \\
\beta_3 \\
\beta_4 \\
\end{bmatrix} =
\begin{bmatrix}
- .017926 \\
2.4759 \\
3.3333 \\
18.844 \\
\end{bmatrix}
$$

and the unrestricted Least Squares solution is:
We immediately see that condition (3b), i.e., the requirement that the coefficient of $x^3$ be positive, is not satisfied. Also because $\hat{\beta}_4$ is negative the discriminant of the derivative is positive and has the value

\[ 4(.037488)^2 - 12(.69962)(-.0019207) = .021746 \]

Since the unrestricted estimate of $\beta_1$, is greatly negative, the restricted solution will undoubtedly be negative also.

The quadratic programming algorithm requires a positive vector of solutions so that we must transform the vector of parameters slightly. Instead of solving

\[ y = X\beta + e \]

directly we will let

\[ \beta = m + \delta \]

where $m$ is an arbitrarily selected vector of constants which will insure that the elements of the solution vector $\delta$ will all be non-negative.

The model (18) becomes

\[ y = Xm + X\delta + e \]
or

\[ z = y - X_m = X_\delta + e. \]

The restraints upon the derivative

\[ A_\delta \geq 0 \]

become

\[ A_\delta \geq -A_m. \]

The right hand sides of the normal equations for transformed model becomes

\[ X'Wz = X'Wy - X'WXm. \]

Let

\[ m' = (-1, 0, 0, -1) \]

then

\[ -X'WXm = \begin{bmatrix} 5.2766 \\ 30.4861 \\ 57.1976 \\ 429.518 \end{bmatrix} \]

and we immediately have

\[ X'Wz = \begin{bmatrix} 5.2585 \\ 32.963 \\ 60.529 \\ 448.38 \end{bmatrix} \]

The coefficients of the normal equations are not altered by
the transformation except that the solution vector will be in terms of $\delta$.

We will now describe the procedure of selecting the restrictions. The roots of the derivative of the unrestrained solution are -6.29 and 19.30. Between these two roots the derivative is positive but for $x$ values less than -6.29 and greater than 19.30 the derivative will always be negative. Therefore, a set of four $x$ values were selected in the range where the derivative was negative. So we

$$z' = \delta_2 + 2\delta_3 x + 3\delta_4 x^2 \geq 0$$

where $x$ takes on the values of $\pm 8.0$ and $\pm 9.0$ for Case I. The matrix of restrictions for Case I then becomes

$$
\begin{bmatrix}
0 & 1 & -16 & 192 \\
0 & 1 & -18 & 243 \\
0 & 1 & 16 & 192 \\
0 & 1 & 18 & 243
\end{bmatrix}
\begin{bmatrix}
\delta_1 \\
\delta_2 \\
\delta_3 \\
\delta_4
\end{bmatrix}
= 
\begin{bmatrix}
192 \\
243 \\
192 \\
243
\end{bmatrix}
$$

when the vector on the right hand side is $-A\mathbf{m}$.

For case II, the $x_2$ takes the values of $\pm 8, \pm 9, \pm 10, \pm 11$ and the matrix of restriction becomes
From Wolfe's algorithm [21], we obtain the unbounded solution vectors shown in Table 4.

Table 4. Quadratic Programming Results

<table>
<thead>
<tr>
<th></th>
<th>Case I</th>
<th>Case II</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\delta^g)</td>
<td>(\delta^{g+1})</td>
<td>(\delta^g)</td>
</tr>
<tr>
<td>(L)</td>
<td>.99921108</td>
<td>.99973740</td>
</tr>
<tr>
<td>(\delta_1)</td>
<td>.28313529</td>
<td>.00116517</td>
</tr>
<tr>
<td>(\delta_2)</td>
<td>.69638347</td>
<td>.00286578</td>
</tr>
<tr>
<td>(\delta_3)</td>
<td>.03125255</td>
<td>.00012861</td>
</tr>
<tr>
<td>(\delta_4)</td>
<td>.99944923</td>
<td>.0041296</td>
</tr>
</tbody>
</table>

When the solution is unbounded, Wolfe shows that the solution for any \(\lambda > L\) is given by

\[
\delta_i = \delta_i^g + (\lambda - L) \delta_i^{g+1}.
\]
Since we wish the solutions for \( \lambda = 1 \) we would have

\[
\delta_i = \delta_i^g + (1-L) \delta_i^{g+1}.
\]

Since the term

\[
(1-L) \delta_i^{g+1}
\]

is quite small, it was ignored and the solution vector becomes

\[
\tilde{\delta} \approx \delta^g = \begin{bmatrix} 1 - \beta_1 \\ \beta_2 \\ \beta_3 \\ 1 - \beta_4 \end{bmatrix}
\]

In terms of the \( \beta \)'s, the solutions are shown in Table 5. It is apparent that the restriction \( \beta_4 \geq 0 \) is the main source of the difficulty. We will now allow the possibility of \( \beta_4 = 0 \) and determine the coefficients. This will be Case III.

The triangular decomposition of

\[
X'WX = C'C
\]

is

\[
C = \begin{bmatrix}
0.976760 & 0.855461 & 2.582108 & 4.428826 \\
0 & 1.855609 & 1.900729 & 13.926054 \\
0 & 0 & 4.791968 & 3.518109 \\
0 & 0 & 0 & 14.098348
\end{bmatrix}
\]

and the inverse is
So in the transformed parameter space of $a$ where 

$$a = c^T \beta$$

the Least Squares estimates are 

$$\hat{a} = c^T \beta$$

or specifically 

$$\hat{a}_1 = -0.018353$$
$$\hat{a}_2 = 1.342727$$
$$\hat{a}_3 = 0.172884$$
$$\hat{a}_4 = -0.027078$$

Again we let 

$$\tilde{a}_1 = \hat{a}_1$$
$$\tilde{a}_2 = \hat{a}_2$$
$$\tilde{a}_3 = \hat{a}_3$$
$$\tilde{a}_4 = 0$$

Since 

$$\tilde{\beta} = c^{-1} \tilde{a}$$

we obtain the minimum distance estimators
\[ \beta_1 = -0.71540 \\
\beta_2 = 0.686649 \\
\beta_3 = 0.036078 \\
\beta_4 = 0.0 \]

That is, the cubic polynomial is not applicable and the fitted quadratic becomes

\[ y = -0.71540 + 0.686649x + 0.036078x^2 \]

the derivative is

\[ y' = 0.686649 + 0.072156x \]

and is non-negative for all values of \( x \) greater than -9.5. From a practical point of view, Case III would be the preferred solution since it can be obtained with ease relative to Cases I and II. The Case III solution is only slightly more difficult to obtain than the ordinary Least Squares solution. We see that the restraint of a non-negative derivative excludes the cubic polynomial as a model for this set of data.

From Table 5 we see that the restrained solutions are very similar to the unrestrained solutions. The troublesome negative \( -\beta_4 \) still remains for Case I and Case II but its value has been made progressively smaller. Cases I and II show that \( \beta_4 \) is approaching zero thereby indicating that under this method of restraining the solution, a quadratic polynomial
is more appropriate. The Case III solution confirms this.

Table 5. Estimates of Coefficients

<table>
<thead>
<tr>
<th></th>
<th>Unrestrained</th>
<th>Case I</th>
<th>Case II</th>
<th>Case III</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>-.72192</td>
<td>-.716864</td>
<td>-.716716</td>
<td>-.715540</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>.69962</td>
<td>.696383</td>
<td>.696750</td>
<td>.686649</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>.037488</td>
<td>.031252</td>
<td>.031269</td>
<td>.036078</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>-.0019207</td>
<td>-.000548</td>
<td>-.000024</td>
<td>-.0.0</td>
</tr>
</tbody>
</table>

It remains to be seen whether the algorithm of Hartley and Hocking utilizing the restriction of a convex discriminant gives a more meaningful solution. Tables 6 and 7 show selected values of the fitted polynomial and the associated derivatives. Within the region of Buehler's interest, the four polynomials practically coincide. It is only at the left end, values of $x$ near and beyond $-5$, do we find any appreciable departure. The derivatives, on the other hand, show much greater sensitivity and the effect of the negative constants can be seen at the large negative values of $x$. As expected, the restraints prevent the derivatives of the restricted polynomials from dropping below zero within the range of the restraints.

With regard to the variance-covariance matrix of the restricted estimates, the same procedure can be used as was used in (V, 6). In this case the $v$ vector is
Table 6. Selected Values of the Polynomials

<table>
<thead>
<tr>
<th>Unrestrained</th>
<th>Case I</th>
<th>Case II</th>
<th>Case III</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-11</td>
<td>-1.325</td>
<td>-3.866</td>
<td>-4.565</td>
</tr>
<tr>
<td>-10</td>
<td>-2.049</td>
<td>-4.001</td>
<td>-4.533</td>
</tr>
<tr>
<td>-9</td>
<td>-2.582</td>
<td>-4.053</td>
<td>-4.437</td>
</tr>
<tr>
<td>-8</td>
<td>-2.936</td>
<td>-4.007</td>
<td>-4.277</td>
</tr>
<tr>
<td>-7</td>
<td>-3.124</td>
<td>-3.872</td>
<td>-4.053</td>
</tr>
<tr>
<td>-6</td>
<td>-3.155</td>
<td>-3.652</td>
<td>-3.766</td>
</tr>
<tr>
<td>-4</td>
<td>-2.798</td>
<td>-2.967</td>
<td>-3.002</td>
</tr>
<tr>
<td>-3</td>
<td>-2.432</td>
<td>-2.510</td>
<td>-2.529</td>
</tr>
<tr>
<td>-2</td>
<td>-1.956</td>
<td>-1.980</td>
<td>-1.985</td>
</tr>
<tr>
<td>-1</td>
<td>-1.382</td>
<td>-1.381</td>
<td>-1.382</td>
</tr>
<tr>
<td>0</td>
<td>-0.722</td>
<td>-0.717</td>
<td>-0.717</td>
</tr>
<tr>
<td>1</td>
<td>0.013</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td>2</td>
<td>0.812</td>
<td>0.797</td>
<td>0.802</td>
</tr>
<tr>
<td>3</td>
<td>1.662</td>
<td>1.639</td>
<td>1.654</td>
</tr>
<tr>
<td>4</td>
<td>2.553</td>
<td>2.534</td>
<td>2.569</td>
</tr>
<tr>
<td>5</td>
<td>3.474</td>
<td>3.478</td>
<td>3.546</td>
</tr>
<tr>
<td>6</td>
<td>4.410</td>
<td>4.468</td>
<td>4.584</td>
</tr>
<tr>
<td>7</td>
<td>5.354</td>
<td>5.501</td>
<td>5.684</td>
</tr>
<tr>
<td>8</td>
<td>6.291</td>
<td>6.574</td>
<td>6.846</td>
</tr>
<tr>
<td>9</td>
<td>7.211</td>
<td>7.683</td>
<td>8.069</td>
</tr>
<tr>
<td>10</td>
<td>8.102</td>
<td>8.824</td>
<td>9.353</td>
</tr>
</tbody>
</table>
Table 7. Selected Values of the Derivatives

<table>
<thead>
<tr>
<th>x</th>
<th>Unrestrained</th>
<th>Restrained</th>
<th>Restrained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Case I</td>
<td>Case II</td>
<td>Case III</td>
</tr>
<tr>
<td>-11</td>
<td>-.822</td>
<td>-.190</td>
<td>.000</td>
</tr>
<tr>
<td>-10</td>
<td>-.626</td>
<td>-.093</td>
<td>.064</td>
</tr>
<tr>
<td>-9</td>
<td>-.442</td>
<td>.001</td>
<td>.128</td>
</tr>
<tr>
<td>-8</td>
<td>-.269</td>
<td>.091</td>
<td>.192</td>
</tr>
<tr>
<td>-7</td>
<td>-.108</td>
<td>.178</td>
<td>.255</td>
</tr>
<tr>
<td>-6</td>
<td>.042</td>
<td>.262</td>
<td>.319</td>
</tr>
<tr>
<td>-5</td>
<td>.181</td>
<td>.343</td>
<td>.382</td>
</tr>
<tr>
<td>-4</td>
<td>.308</td>
<td>.420</td>
<td>.445</td>
</tr>
<tr>
<td>-3</td>
<td>.423</td>
<td>.494</td>
<td>.508</td>
</tr>
<tr>
<td>-2</td>
<td>.527</td>
<td>.565</td>
<td>.571</td>
</tr>
<tr>
<td>-1</td>
<td>.619</td>
<td>.632</td>
<td>.634</td>
</tr>
<tr>
<td>0</td>
<td>.760</td>
<td>.696</td>
<td>.697</td>
</tr>
<tr>
<td>1</td>
<td>.769</td>
<td>.757</td>
<td>.759</td>
</tr>
<tr>
<td>2</td>
<td>.827</td>
<td>.815</td>
<td>.822</td>
</tr>
<tr>
<td>3</td>
<td>.876</td>
<td>.869</td>
<td>.884</td>
</tr>
<tr>
<td>4</td>
<td>.907</td>
<td>.920</td>
<td>.946</td>
</tr>
<tr>
<td>5</td>
<td>.930</td>
<td>.968</td>
<td>1.008</td>
</tr>
<tr>
<td>6</td>
<td>.942</td>
<td>1.012</td>
<td>1.069</td>
</tr>
<tr>
<td>7</td>
<td>.942</td>
<td>1.053</td>
<td>1.131</td>
</tr>
<tr>
<td>8</td>
<td>.931</td>
<td>1.091</td>
<td>1.192</td>
</tr>
<tr>
<td>9</td>
<td>.908</td>
<td>1.126</td>
<td>1.254</td>
</tr>
<tr>
<td>10</td>
<td>.873</td>
<td>1.157</td>
<td>1.315</td>
</tr>
<tr>
<td>11</td>
<td>.827</td>
<td>1.185</td>
<td>1.376</td>
</tr>
</tbody>
</table>
and the inverse of C has been previously given. In general, one must orthogonize $X'WX$ to determine the vector $v$ but with only one simple restriction, the above procedure is more convenient.

C. The Infinite Range Case of Odd Order Polynomials

Consider the fitting of polynomials of odd degrees, say of degrees $n+1$. The derivative can be expressed as

$$Y'(x) = \frac{n}{2} \prod_{i=1}^{n/2} (a_i x^2 + b_i x + c_i)$$

For the imaginary discriminant to exist for all quadratic terms, the roots of (21) must pair up as complex conjugates. However, it is very difficult to express this requirement explicitly in the fitting of the constant of the polynomial since the factorization required in (a) cannot be accomplished until after the constants are known, and (b) the relationship between the fitted constants and the coefficients $a_i$, $b_i$, and $c_i$ involve the symmetric functions of the roots and is exceedingly awkward to manipulate. It may be possible, however, to develop an iterative technique.
D. Application of Restrained Least Squares to Frequency Gradiation

We now consider the problem of graduating data with the Gram-Charlies series. We restrict ourselves to the Type A series in which the frequency function of interest is expressed as an infinite series of derivatives of the normal distribution. In practice, the series will be truncated after four or five terms. An excellent description of the Gram-Charlies series can be found in Kendall and Stuart [14] whose notation and example we shall employ.

1. The standard solution as a least squares problem

Let the density function of interest be represented by

\[ f(x) = \sum_{j=0}^{k} C_j H_j(x) \alpha(x) \]  

(22)

where

\[ \alpha(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \]

\( H_j(x) \) is the \( j \)-th order Hermite polynomial.

\( C_j \) are coefficients to be determined.

The \( C_j \)'s are to be determined so as to minimize the weighted mean square integral. Thus the quadratic form to be minimized becomes

\[ Q = \int_{-\infty}^{\infty} \left[ f(x) - \sum_{j=0}^{k} C_j \alpha(x) H_j(x) \right]^2 \frac{1}{\alpha(x)} \, dx \]  

(23)
where $\frac{1}{\alpha(x)}$ is the appropriate weight. Differentiating (23) with respect to $C_i$ yields the $i$-th normal equation, i.e.,

$$\frac{\partial Q}{\partial C_i} = 2 \int_{-\infty}^{\infty} \left[ f(x) - \sum_{j=0}^{h} \alpha(x) H_j(x) \right] dx = 0$$

which can be written as

$$\int_{-\infty}^{\infty} \sum_{j=0}^{h} C_j \alpha(x) H_j(x) H_l(x) dx = \int_{-\infty}^{\infty} f(x) H_l(x) dx. \quad (24)$$

Interchanging the order of integration and summation in (24) and using the well known orthogonality property

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) 2(x) dx = \begin{cases} 0 & m \neq n \\ n! & m = n \end{cases}$$

the $i$-th normal equation (24) reduces to

$$C_i = \frac{1}{i!} \int_{-\infty}^{\infty} f(x) H_i(x) dx \quad (25)$$

Since the Hermite polynomials are polynomials in $x$, the $C_i$ are linear functions of the moments of the unknown density function. Unfortunately, there is no assurance that the $C_i$ calculated from (25) will insure that the calculated frequencies, will be non-negative. Kendall and Stuart provide an interesting example in which 3, 4, and 5 terms of (22) are fitted to the distribution of lengths of beans. With three
terms all calculated frequencies were positive but both the
4 and 5 term graduations gave one negative general frequency
at the high tail. Thus they show that employing additional
terms of (22) does not guarantee the elimination of negative
frequencies. In the following section, the restrained method
of fitting the constants which will soon be described will be
illustrated using the same data.

2. The restrained solution

In the restrained solution of (22) the quadratic form is
identical with (23) but now its minimization is restrained by
linear inequalities. Specifically, we wish to minimize

\[
Q = \int_{-\infty}^{\infty} \left[ f(x) - \sum_{j=0}^{k} C_j \alpha(x) H_j(x) \right]^2 \frac{1}{\alpha(x)} \, dx
\]  

subject to

\[
f(x_t) = \sum_{j=0}^{k} C_j \alpha(x_t) H_j(x_t) \geq 0
\]  

for all \( x_t \).

The \( x_t \) represent an arbitrarily selected but meaningful
grid of points in the finite interval

\[-X \leq x \leq X\]

The inequalities (27) insure that the calculated frequencies
will be positive or zero at the specified points.
To determine the explicit structure of (26) to be minimized with respect to the $C_j$, we expand the integral to obtain

$$Q = \int_{-\infty}^{\infty} \left[ f^2(x) - 2f(x) \sum_{j=0}^{k} C_j \alpha(x) H_j(x) \right] \frac{1}{\alpha(x)} \, dx \quad (28)$$

From (28), the constant term, i.e., the term not containing the unknown coefficient can be transposed to the left hand side. Then we have

$$F = Q - \int_{-\infty}^{\infty} f^2(x) \frac{1}{\alpha(x)} \, dx \quad (29)$$

The linear coefficients of $C_j$ are obtained from the middle expression and are given by

$$-2 \int_{-\infty}^{\infty} f(x) H_j(x) \, dx \quad (30)$$

Here the coefficients of $C_j$ can be expressed as the moments of the distribution since $H_j(x)$ is a polynomial in $x$ as shown in Table 8.
Table 8. Coefficients of $C_j$

<table>
<thead>
<tr>
<th>Subscript</th>
<th>$H_j(x)$</th>
<th>Coefficient</th>
<th>Standardized Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>-2</td>
<td>-2</td>
</tr>
<tr>
<td>1</td>
<td>$x$</td>
<td>$-2\mu_1$</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>$x^2-1$</td>
<td>$-2(\mu_2-1)$</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>$x^3-3x$</td>
<td>$-2(\mu_3-\mu_1)$</td>
<td>$-2\mu_3$</td>
</tr>
<tr>
<td>4</td>
<td>$x^4-6x^2+3$</td>
<td>$-2(\mu_4-6\mu_2+3)$</td>
<td>$-2(\mu_4-3)$</td>
</tr>
<tr>
<td>5</td>
<td>$x^5-10x^3+15x$</td>
<td>$-2(\mu_5-10\mu_3+15\mu_1)$</td>
<td>$-2(\mu_5-10\mu_3)$</td>
</tr>
</tbody>
</table>

The standardized coefficients follow by setting $\mu_1 = 0$ and $\mu_2 = 1$. The usual practice, of course, is to estimate these population moments by the sample moments.

The second order coefficients are given in the product of the two summations, i.e.

$$\int \sum_{j=0}^{\infty} C_j \alpha(x) H_j(x) \cdot \sum_{i=0}^{k} C_1 \alpha(x) H_1(x) \frac{1}{\alpha(x)} \, dx$$

Utilizing the orthogonality conditions previously described, the coefficient of $C_j$ is given as $j!$. Hence, for $k=5$, the matrix of the squared terms becomes

$$\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 2! & 0 & 0 & 0 \\
0 & 0 & 0 & 3! & 0 & 0 \\
0 & 0 & 0 & 0 & 4! & 0 \\
0 & 0 & 0 & 0 & 0 & 5!
\end{bmatrix}$$
The quadratic form to be minimized (26) can now be written for $k=5$, as

$$\frac{1}{2}F = \begin{bmatrix} -1 & 0 & 0 & -\mu_3 & -(\mu_4-3) & -(\mu_5-10\mu_3) \\ 0 & C_1 & 0 & C_3 & C_4 & C_5 \\ 0 & C_1 & C_2 & C_3 & C_4 & C_5 \\ -\mu_3 & C_3 & C_2 & C_3 & C_4 & C_5 \\ -(\mu_4-3) & C_4 & C_3 & C_3 & C_4 & C_5 \\ -(\mu_5-10\mu_3) & C_5 & C_4 & C_5 & C_5 & C_5 \end{bmatrix} \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{bmatrix} = \begin{bmatrix} 1 \\ C_1 \\ C_2 \\ 2 \\ 6 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ C_1 \\ C_2 \\ 2 \end{bmatrix} \begin{bmatrix} C_0 \\ C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \end{bmatrix}$$

(33)

The quadratic form for $k\neq 5$ can be obtained from (33) by inspection.

The data to be graduated is given in Table 9 and the unrestrained graduating function given by Kendall and Stuart is

$$\hat{f}(x) = 9440 \frac{\alpha(x)}{\sqrt{\mu_2}} \sum_{j=0}^{5} \hat{c}_j H_j(x)$$

(34)

The first two moments of $z = \text{twice the length}$ are

$$\mu_1 = 28.809216$$
$$\mu_2 = 3.238425$$

and the standardized higher moments are

$$\mu_3 = -.910569$$
$$\mu_4 = 4.862944$$
$$\mu_5 = -12.574125$$
Table 9. Graduation Using Unrestricted and Restricted Gram-Charlier Series

<table>
<thead>
<tr>
<th>Length</th>
<th>Observed Frequency</th>
<th>Refined Unrestricted Fit</th>
<th>Unrestricted Fit</th>
<th>Restricted Fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>mm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;17.75</td>
<td>-</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-4.1</td>
</tr>
<tr>
<td>17.5</td>
<td>1</td>
<td>-15.2</td>
<td>-9.4</td>
<td>1.1</td>
</tr>
<tr>
<td>17.0</td>
<td>6</td>
<td>13.7</td>
<td>10.3</td>
<td>14.1</td>
</tr>
<tr>
<td>16.5</td>
<td>55</td>
<td>116.6</td>
<td>110.5</td>
<td>101.3</td>
</tr>
<tr>
<td>16.0</td>
<td>275</td>
<td>370.4</td>
<td>362.0</td>
<td>409.3</td>
</tr>
<tr>
<td>15.5</td>
<td>1129</td>
<td>926.2</td>
<td>907.6</td>
<td>1068.5</td>
</tr>
<tr>
<td>15.0</td>
<td>2082</td>
<td>1833.0</td>
<td>1839.8</td>
<td>1867.4</td>
</tr>
<tr>
<td>14.5</td>
<td>2294</td>
<td>2506.4</td>
<td>2559.9</td>
<td>2209.5</td>
</tr>
<tr>
<td>14.0</td>
<td>1787</td>
<td>2082.6</td>
<td>2118.8</td>
<td>1793.3</td>
</tr>
<tr>
<td>13.5</td>
<td>929</td>
<td>921.3</td>
<td>896.1</td>
<td>1049.2</td>
</tr>
<tr>
<td>13.0</td>
<td>437</td>
<td>199.0</td>
<td>167.5</td>
<td>510.1</td>
</tr>
<tr>
<td>12.5</td>
<td>199</td>
<td>132.1</td>
<td>130.2</td>
<td>247.1</td>
</tr>
<tr>
<td>12.0</td>
<td>115</td>
<td>178.1</td>
<td>184.1</td>
<td>114.8</td>
</tr>
<tr>
<td>11.5</td>
<td>70</td>
<td>117.0</td>
<td>117.1</td>
<td>42.4</td>
</tr>
<tr>
<td>11.0</td>
<td>36</td>
<td>43.5</td>
<td>41.3</td>
<td>11.2</td>
</tr>
<tr>
<td>10.5</td>
<td>18</td>
<td>10.0</td>
<td>9.1</td>
<td>2.0</td>
</tr>
<tr>
<td>10.0</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9.5</td>
<td>1</td>
<td>1.7</td>
<td>1.4</td>
<td>.3</td>
</tr>
<tr>
<td>9.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>9440</td>
<td>9440</td>
<td>9441.4</td>
<td>9440.1</td>
</tr>
</tbody>
</table>
Substituting these sample moments into (25) gives the values of $c_j$ contained in Table 10.

Table 10. Estimates of the Coefficients in a Four Term Gram-Charlier Series

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Restrained</th>
<th>Unrestrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_0$</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>$c_1$</td>
<td>0.010015</td>
<td>0.0</td>
</tr>
<tr>
<td>$c_2$</td>
<td>0.009105</td>
<td>0.0</td>
</tr>
<tr>
<td>$c_3$</td>
<td>-0.068508</td>
<td>-0.151762</td>
</tr>
<tr>
<td>$c_4$</td>
<td>0.086362</td>
<td>0.077622</td>
</tr>
<tr>
<td>$c_5$</td>
<td>-0.025840</td>
<td>-0.028903</td>
</tr>
</tbody>
</table>

The restrained solution will be determined by locating the point in the restricted space nearest to the Least Squares solution. From Table 9, we see that the first negative graduation occurs at a length of 17.5 mm. Now the restriction at $X = 17.5$ is

$$5.634321 + 19.382929c_1 + 61.046205c_2 + 171.24281c_3 + 405.96416c_4 + 711.61068c_5 > 0.$$ (35)

The restricting planes can all be determined from Table 11. For example at $X = 17.5$ we find that the coefficient of $c_5$ is

$$\frac{9440}{\sqrt{\mu_2}} \kappa(x) H_5(x) = \frac{(9440)(0.00107407)(126.29934)}{\sqrt{3.238425}} = 711.61068$$
### Table 11. Normal Ordinate and Hermitian Coefficients

<table>
<thead>
<tr>
<th>Length $X$</th>
<th>$\alpha(x)$</th>
<th>$H_1(x)$</th>
<th>$H_2(x)$</th>
<th>$H_3(x)$</th>
<th>$H_4(x)$</th>
<th>$H_5(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.5</td>
<td>0.0010740</td>
<td>3.4401607</td>
<td>10.8347062</td>
<td>30.3928099</td>
<td>72.0520342</td>
<td>126.2993428</td>
</tr>
<tr>
<td>17.0</td>
<td>0.0062259</td>
<td>2.8844701</td>
<td>7.3201679</td>
<td>15.3458656</td>
<td>22.3041871</td>
<td>2.9522993</td>
</tr>
<tr>
<td>16.5</td>
<td>0.0265017</td>
<td>2.3287794</td>
<td>4.4232138</td>
<td>5.6431308</td>
<td>-1.1280344</td>
<td>-22.8706871</td>
</tr>
<tr>
<td>16.0</td>
<td>0.0828386</td>
<td>1.7730888</td>
<td>2.1438440</td>
<td>0.2550482</td>
<td>-5.9793088</td>
<td>-11.6220387</td>
</tr>
<tr>
<td>15.5</td>
<td>0.1901451</td>
<td>1.2173981</td>
<td>-1.8479394</td>
<td>-3.6958531</td>
<td>2.8924327</td>
<td></td>
</tr>
<tr>
<td>15.0</td>
<td>0.3205019</td>
<td>0.6617075</td>
<td>-0.5621431</td>
<td>-1.6953894</td>
<td>0.5645774</td>
<td>7.1551428</td>
</tr>
<tr>
<td>14.5</td>
<td>0.3967065</td>
<td>0.106168</td>
<td>-0.9887604</td>
<td>-0.3168590</td>
<td>1.5783508</td>
<td></td>
</tr>
<tr>
<td>14.0</td>
<td>0.3607980</td>
<td>-0.4496737</td>
<td>-0.7997935</td>
<td>1.2580943</td>
<td>1.8276484</td>
<td>-5.8542229</td>
</tr>
<tr>
<td>13.5</td>
<td>0.2460727</td>
<td>-1.0053644</td>
<td>0.0107576</td>
<td>1.9999135</td>
<td>-2.0429147</td>
<td>-5.9457803</td>
</tr>
<tr>
<td>13.0</td>
<td>0.1179629</td>
<td>-1.5610550</td>
<td>1.4368929</td>
<td>0.8790411</td>
<td>-5.6829104</td>
<td>5.3551714</td>
</tr>
<tr>
<td>12.5</td>
<td>0.0424577</td>
<td>-2.1167457</td>
<td>3.4806124</td>
<td>-3.1340800</td>
<td>2.0977868</td>
<td>20.5964365</td>
</tr>
<tr>
<td>12.0</td>
<td>0.0112218</td>
<td>-2.6724363</td>
<td>6.1419161</td>
<td>-11.0690072</td>
<td>11.1554692</td>
<td>14.4637475</td>
</tr>
<tr>
<td>11.5</td>
<td>0.0021780</td>
<td>-3.2281270</td>
<td>9.4208040</td>
<td>-23.9552979</td>
<td>49.0683324</td>
<td>-62.5776177</td>
</tr>
<tr>
<td>11.0</td>
<td>0.0003104</td>
<td>-3.7838176</td>
<td>13.317276</td>
<td>-42.8225093</td>
<td>122.0807389</td>
<td>-290.6412194</td>
</tr>
<tr>
<td>10.5</td>
<td>0.0000324</td>
<td>-4.3395083</td>
<td>17.8313324</td>
<td>-68.7001987</td>
<td>244.631063</td>
<td>-786.7778387</td>
</tr>
<tr>
<td>10.0</td>
<td>0.0000024</td>
<td>-4.8951989</td>
<td>22.9629729</td>
<td>-102.6179233</td>
<td>433.4462335</td>
<td>-1711.3338610</td>
</tr>
<tr>
<td>9.5</td>
<td>0.0000001</td>
<td>-5.4508896</td>
<td>28.7121976</td>
<td>-145.6052405</td>
<td>707.5415010</td>
<td>-3274.309590</td>
</tr>
<tr>
<td>9.0</td>
<td>0.0000000</td>
<td>-6.0065802</td>
<td>35.0790064</td>
<td>-198.6917076</td>
<td>1088.2206708</td>
<td>-5741.7179761</td>
</tr>
</tbody>
</table>
By definition, we observe that $C_0$ equals unity.

Consider the transformation

$$y_j = \sqrt{j!} C_j$$

that transforms the diagonal matrix of the quadratic form to unity. The restricting plane (35) is transformed to its normal form and becomes

$$0.04174885 + 0.14362276y_1 + 0.3198501y_2 + 0.5180288y_3 +$$

$$0.6140236y_4 + 0.48134321y_5 = 0$$

(37)

Now the coefficients of the $y_j$'s are the direction cosines, $\alpha_j$. Let

$$D = \ell + \sum_{j=1}^{5} \alpha_j \hat{y}_j = 0.0417489 - 1.11480 = -0.069731$$

Since

$$\cos \theta_j = \alpha_j = \frac{\hat{y}_j - \tilde{y}_j}{D}$$

we have immediately that

$$\tilde{y}_j = \hat{y}_j - D \alpha_j .$$

Upon retransformation, we obtain

$$\tilde{C}_j = \frac{\tilde{y}_j}{\sqrt{j!}} .$$

Table 12 summarizes the computations described above.
Table 12. Summary of the Direct Projection Method

<table>
<thead>
<tr>
<th>j</th>
<th>$\hat{y}_j$</th>
<th>$-D^\alpha_j$</th>
<th>$\tilde{y}_j$</th>
<th>$\gamma_j$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>.010015</td>
<td>.010015</td>
<td>.010015</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>.022303</td>
<td>.022303</td>
<td>.009105</td>
</tr>
<tr>
<td>3</td>
<td>-.371740</td>
<td>.036123</td>
<td>-.335617</td>
<td>-.068508</td>
</tr>
<tr>
<td>4</td>
<td>.380272</td>
<td>.042816</td>
<td>.423088</td>
<td>.086362</td>
</tr>
<tr>
<td>5</td>
<td>-.316623</td>
<td>.033564</td>
<td>-.283059</td>
<td>-.025840</td>
</tr>
</tbody>
</table>

As a check upon the calculation, we note that

$$\tilde{D} = l + \sum_{j=1}^{5} \alpha_j \hat{y}_j = 0$$

That is the distance from the restricted point to the restricting plane should be zero. This is obvious since the restricted point actually lies in the restricting plane.

Carrying out the above computation, we find that

$$\tilde{D} = .041749 + .143623(.010015) + .319850(.022303) + .518029(-.335617) + .614024(.423088) + .481343(-.283059) -$$

$$\tilde{D} = -.0000006$$

which agrees within the computational accuracy to the theoretical value of zero.

From Table 9 we see that the restricted solution has adjusted the graduation to produce a slightly positive value of 0.1 at a length of 17.5 instead of the -9.4 obtained using
the unrestricted estimators. There appears to be no consistent pattern in the restricted graduation. That is one might suppose the other upper tail graduations would be increased also but we find that the 16.5 mm. length, a decrease in fit was obtained. However, this is an exception since in general the upper tail graduations have been increased.

It is interesting to note the fit of -0.4 at lengths greater than 17.5. The above computation was repeated using the restricting plane for \( X = 18.0 \) and the restricted solution of

\[
\begin{align*}
\bar{c}_0 &= 1.0 \\
\bar{c}_1 &= 0.008521 \\
\bar{c}_2 &= 0.015959 \\
\bar{c}_3 &= -0.133346 \\
\bar{c}_4 &= 0.092030 \\
\bar{c}_5 &= -0.021373
\end{align*}
\]

was obtained. The graduations corresponding to lengths of 18.0, 17.5 and 17.0 were -2.4, -9.8, and -16.1. It is apparent from this that the restriction for \( X = 18.0 \) produced a non-convex region. Note also that this restriction is beyond the observed data and corresponded to extrapolation. Hence reasonable care must be exercised in selecting the restrictions.
VII. ACKNOWLEDGMENT

I acknowledge help and encouragement from Professor H. O. Hartley at all stages of this work which was generously supported by the National Science Foundation Research Grant No. 14236.
VIII. LITERATURE CITED


9. Hartley, H. O. Nonlinear programming for separable objective functions and constraints. (Dittoed.) Paper presented at a Symposium on Mathematical Programming (Linear Programming and Recent Extensions) sponsored by the Rand Corporation, Santa Monica, California, March 16-20, 1959. Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology.

10. ______ and R. Hocking. Convex programming by tangential approximation. [To be published in Management Science ca. 1963].


13. The theory of least squares and the general linear hypothesis. (Dittoed.) Ames, Iowa, Statistical Laboratory, Iowa State University of Science and Technology. ca. 1947.


