

CONSTRAINED LINEAR MODELS: A SAS IMPLEMENTATION

by

Thomas M. Gerig and A. Ronald Gallant

Institute of Statistics
Mimeograph Series No. 899
Raleigh - November 1973

BIOMATHEMATICS TRAINING PROGRAM

Constrained Linear Models: A SAS Implementation

by

Thomas M. Gerig and A. Ronald Gallant

ABSTRACT

This technical report contains the documentation of a Statistical Analysis System (SAS) implementation of the Constrained Linear Models estimation method reported in Mimeo Series Number 814. A monograph briefly describing the computational method and relevant theory is included to make this report self contained.

THE CLM PROCEDURE

The CLM procedure is designed to analyze linear models whose parameters are subject to linear equality constraints. No rank conditions are imposed on the X matrix or on the restriction matrix. The methods applied are discussed in the attached monograph.

For a given model, sets of linear parametric functions can be estimated and linear hypotheses can be tested. The procedure automatically checks to see if linear parametric functions are estimable or specified, and checks all constraints and hypothesis equations for consistency.

OUTPUT

For each model entered, the CLM procedure gives model dimensions, error variance information, restriction specification, and parameter estimates including standard errors and an indication of whether the estimates are unbiased or specified by the restrictions. The variance-covariance or correlation matrix of the estimated parameters may be requested.

Each request for the estimation of linear parametric functions results in the printing of the combining matrix and the estimates with standard errors and an indication of whether the estimates of the linear functions are unbiased or specified by the restrictions. The variance-covariance matrix or correlation matrix of the estimated linear functions may be requested.

Each request to test a linear hypothesis results in the estimation of the associated linear functions (with output as described in the previous paragraph) in addition to the printing of the hypothesized values of the linear functions and the appropriate F value, degrees of freedom, and p -value. The variance-covariance or correlation matrix of the estimated linear functions may be requested.

THE PROCEDURE CLM STATEMENT

The PROCEDURE CLM statement is of the form

```
PROC CLM  <DATA=data_set_name> ;
```

PROCEDURE INFORMATION STATEMENTS

VARIABLES Statement

The VARIABLES statement tells which SAS variables will be used in the analysis to follow and only these variables will be included in the internal computation of the sums of squares and cross products matrix.

DROP Statement

The DROP statement serves the same purpose as the VARIABLES statement except to specify which SAS variables not to include.

BY Statement

The BY statement works as explained in the SAS User's Guide.

CLASSES and ID Statements

The CLASSES and ID statements may not be used with this procedure.

PARMCARDS Statement

There are five types of parameter cards: MODEL, CONSTRAINTS (or RESTRICTIONS), ESTIMATE, TEST, and OUTPUT.

Each MODEL statement must be immediately followed by a CONSTRAINTS statement (if there are any constraints).

The ESTIMATE statement causes linear parametric functions to be estimated (under the current model) and the TEST statement causes linear hypotheses to be tested (in the framework of the current model). There is no limit to the number of ESTIMATE or TEST statements that can be used. The OUTPUT statement causes predicted values and residuals to be saved in an output file.

PARAMETER CARD SPECIFICATION

MODEL, CONSTRAINT, ESTIMATE, and TEST Statements

MODEL, CONSTRAINT, ESTIMATE, and TEST statements are of the form

```
MODEL numeric_variable = numeric_variable <additional numeric variables>
      NOINT
      </NOMEAN> <VAR> <CORR> #
```

CONSTRAINTS

```
<linear combination of independent numeric variables = constant>
      :
      :
      <linear combination of independent numeric variables = constant> #
```

ESTIMATE

```
<linear combination of independent numeric variables = ? >
      :
      :
      <linear combination of independent numeric variables = ? >
      </VAR> <CORR> #
```

TEST

```
<linear combination of independent numeric variables = constant> #
      :
      :
      <linear combination of independent numeric variables = constant>
      </VAR> <CORR> #
```

Options and Parameters

NOMEAN	When this option is absent, the procedure will
NOINT	include an intercept term in the model. When
	NOMEAN or NOINT is specified, no intercept will
	be included.
VAR	This option causes the variance-covariance matrix
	of the appropriate set of estimates to be printed.
CORR	This option causes the correlation matrix of the
	appropriate set of estimates to be printed.

OUTPUT Statements

OUTPUT statements are of the form

```

OUTPUT OUT=data_set_2
  <PREDICTED predicted_variable_name>
  <RESIDUALS residual_variable_name> #

```

where predicted_variable_name and residual_variable_name are valid SAS names.

An OUTPUT statement refers to the current MODEL statement. The OUTPUT statement tells SAS to build a new data set to be named data_set_2 containing all the variables in the data set to which CIM is applied in addition to other new variables named in the OUTPUT statement. If PREDICTED and a SAS variable name are given, the predicted values for the current model will be included in the new data set under the given SAS variable name. Similarly, if RESIDUALS and a SAS variable name are given, the residuals for the current model will be included in the new data set under the given SAS variable name.

JOB CONTROL STATEMENT REQUIREMENTS AT TUCC

```

//job_name      JOB  xxx.yyy.zzz, programmer_name
//STEP1         EXEC SAS
//SAS.STEPLIB    DD DSN=NCS.ES.B4139.GALLANT.SASLIB, DISP=SHR
//              DD DSN=NCS.SAS, DISP=SHR
//SAS.SYSIN      DD *
:
:
SAS  statements and data cards
:
:
/*

```

EXAMPLE

The following example is discussed in the attached monograph.

```
//CIMEG      JOB xxx.yy.zzzzz,programmer_name
//STEP1      EXEC SAS
//SAS.STEPLIB DD DSN=NCS.ES.B4139.GALLANT.SASLIB,DISP=SHR
//           DD DSN=NCS.SAS,DISP=SHR
//SAS.SYSIN   DD *
INPUT Y 5-10 X 15-20;
BETA_1=0; BETA_2=0; BETA_3=0; BETA_4=0; BETA_5=0;
IF X<12 THEN GO TO A; IF X>=12 THEN GO TO B;
A: BETA_1=1; BETA_2=X; BETA_3=X**2; RETURN;
B: BETA_4=1; BETA_5=X; RETURN;
CARDS;
      0.46      0.50
      0.47      1.50
      0.56      2.50
      :         :
      0.99      70.50
      1.04      71.50

PROC CLM;
PARMCARDS;
MODEL Y = BETA_1 BETA_2 BETA_3 BETA_4 BETA_5 / NOINT #
CONSTRAINTS
      BETA_1 + 12*BETA_2 + 144*BETA_3 - BETA_4 - 12*BETA_5 = 0
      BETA_2 + 24*BETA_3 - BETA_5 = 0 #
ESTIMATE
      BETA_1 + 12*BETA_2 + 144*BETA_3 = ?
      BETA_2 + 24*BETA_3 = ? #
TEST
      BETA_5 = 0 #
OUTPUT OUT=DATA_2 PREDICTED YHAT RESIDUAL EHAT #
PROC PRINT DATA=DATA_2;
/*
```

```
*****  
*  
* MODEL DIMENSIONS  
*  
*****
```

DEPENDENT VARIABLE = Y

NUMBER OF OBSERVATIONS = 72

NUMBER OF PARAMETERS IN THE MODEL = 5

NUMBER OF RESTRICTIONS = 2

NUMBER OF INDEPENDENT RESTRICTIONS = 2

MAXIMUM NUMBER OF LINEARLY INDEPENDENT ESTIMABLE LINEAR FUNCTIONS = 5

MAXIMUM NUMBER OF UNSPECIFIED LINEARLY INDEPENDENT ESTIMABLE LINEAR FUNCTIONS = 3

```
*****  
*  
* ERROR VARIANCE  
*  
*****
```

ERROR SUM OF SQUARES = 3.79124350D-02

ERROR VARIANCE ESTIMATE = 5.49455580D-04

DEGREES OF FREEDOM ASSOCIATED WITH THE ESTIMATE OF ERROR VARIANCE = 69

* * *

LEFT HAND SIDE, RR, OF THE RESTRICTIONS RR*B=R

* * *

```
*****  
*                                     *  
*      RIGHT HAND SIDE, R, OF THE RESTRICTIONS RR#B=R      *  
*                                     *  
*****
```

RCW	1	0.0
RCW	2	0.0

```
*****  
*  
*  
*  
*  
*  
*****  
PARAMETER ESTIMATES  
*****  
*****
```

VARIABLE	COEFFICIENT	STANDARD ERROR	COMMENTS
BETA_1	0.4235225	0.01565432	UNBIASED
BETA_2	0.05500032	0.003006579	UNBIASED
BETA_3	-0.002126766	0.0001284148	UNBIASED
BETA_4	0.7297768	0.006731528	UNBIASED
BETA_5	0.003957931	0.0001542794	UNBIASED

```
*****
*
*      E S T I M A T E  G * B
*
*****
```

```
*****
*
*      G
*
*****
```

```

BETA_1  BETA_2  BETA_3  BETA_4  BETA_5
ROW 1      1.0000  12.0000  144.0000  0.0      0.0
ROW 2      0.0      1.0000  24.0000  0.0      0.0

```

```
*****
*
*      E S T I M A T E  O F  G * B
*
*****
```

ROW	ESTIMATE	STANDARD ERROR	COMMENTS
1	0.772720	0.005123311	UNBIASED
2	0.0039579	0.0001542794	UNBIASED

```
*****  
*  
*          ESTIMATE G*B  
*  
*****
```

```
*****  
*  
*          G  
*  
*****
```

BETA_1 BETA_2 BETA_3 BETA_4 BETA_5

ROW 1 0.0 0.0 0.0 0.0 1.0000

```
*****  
*  
*          ESTIMATE OF G*B  
*  
*****
```

ROW	ESTIMATE	STANDARD ERROR	COMMENTS
1	0.003957931	0.0001542794	UNBIASED

STATISTICAL ANALYSIS SYSTEM

DATA DATA0001

10:27 TUESDAY, NOVEMBER 6, 1973

```
*****
*
*      T E S T   H : G * B = G * B 0
*
* *****
* *****
```

```
*****
*
*      G*B0
*
* *****
* *****
```

ROW 1 0.0

F STATISTIC FOR TESTING $H:G*B=G*B0 = 658.1434$
WITH 1 AND 69 DEGREES OF FREEDOM

PROBABILITY OF OBSERVING A LARGER F = 0.0

OBS	Y	X	BETA_1	BETA_2	BETA_3	BETA_4	BETA_5	YHAT	EHAT
1	0.46	0.5	1	0.5	0.25	0	0.0	.450491	.0095090
2	0.47	1.5	1	1.5	2.25	0	0.0	.501238	-.0312378
3	0.56	2.5	1	2.5	6.25	0	0.0	.547731	.0122690
4	0.61	3.5	1	3.5	12.25	0	0.0	.589971	.0200293
5	0.61	4.5	1	4.5	20.25	0	0.0	.627957	-.0179569
6	0.67	5.5	1	5.5	30.25	0	0.0	.661690	.0083104
7	0.68	6.5	1	6.5	42.25	0	0.0	.691169	-.0111687
8	0.78	7.5	1	7.5	56.25	0	0.0	.716394	.0636057
9	0.69	8.5	1	8.5	72.25	0	0.0	.737366	-.0473664
10	0.74	9.5	1	9.5	90.25	0	0.0	.754085	-.0140849
11	0.77	10.5	1	10.5	110.25	0	0.0	.766550	.0034501
12	0.78	11.5	1	11.5	132.25	0	0.0	.774761	.0052386
13	0.75	12.5	0	0.0	0.00	1	12.5	.779251	-.0292510
14	0.80	13.5	0	0.0	0.00	1	13.5	.783209	.0167911
15	0.78	14.5	0	0.0	0.00	1	14.5	.787167	-.0071668
16	0.82	15.5	0	0.0	0.00	1	15.5	.791125	.0280752
17	0.77	16.5	0	0.0	0.00	1	16.5	.795083	-.0250827
18	0.80	17.5	0	0.0	0.00	1	17.5	.799041	.0009594
19	0.81	18.5	0	0.0	0.00	1	18.5	.802999	.0070014
20	0.78	19.5	0	0.0	0.00	1	19.5	.806957	-.0269565
21	0.87	20.5	0	0.0	0.00	1	20.5	.810914	.0590856
22	0.80	21.5	0	0.0	0.00	1	21.5	.814872	-.0148724
23	0.83	22.5	0	0.0	0.00	1	22.5	.818830	.0111697
24	0.81	23.5	0	0.0	0.00	1	23.5	.822788	-.0127892
25	0.88	24.5	0	0.0	0.00	1	24.5	.826746	.0532538
26	0.81	25.5	0	0.0	0.00	1	25.5	.830704	-.0207041
27	0.83	26.5	0	0.0	0.00	1	26.5	.834662	-.0046620
28	0.82	27.5	0	0.0	0.00	1	27.5	.838620	-.0186200
29	0.82	28.5	0	0.0	0.00	1	28.5	.842578	-.0225779
30	0.86	29.5	0	0.0	0.00	1	29.5	.846536	.0134642
31	0.82	30.5	0	0.0	0.00	1	30.5	.850494	-.0304937
32	0.85	31.5	0	0.0	0.00	1	31.5	.854452	-.0044517
33	0.88	32.5	0	0.0	0.00	1	32.5	.858410	.0215904
34	0.86	33.5	0	0.0	0.00	1	33.5	.862368	-.0023675
35	0.91	34.5	0	0.0	0.00	1	34.5	.866325	.0436745
36	0.87	35.5	0	0.0	0.00	1	35.5	.870283	-.0002834
37	0.87	36.5	0	0.0	0.00	1	36.5	.874241	-.0042413
38	0.87	37.5	0	0.0	0.00	1	37.5	.878199	-.0081993
39	0.85	38.5	0	0.0	0.00	1	38.5	.882157	-.0321572
40	0.90	39.5	0	0.0	0.00	1	39.5	.886115	.0138849
41	0.87	40.5	0	0.0	0.00	1	40.5	.890073	-.0200731
42	0.91	41.5	0	0.0	0.00	1	41.5	.894031	.0159690
43	0.90	42.5	0	0.0	0.00	1	42.5	.897989	.0020111
44	0.93	43.5	0	0.0	0.00	1	43.5	.901947	.0280531
45	0.89	44.5	0	0.0	0.00	1	44.5	.905905	-.0159048
46	0.89	45.5	0	0.0	0.00	1	45.5	.909863	-.0198627
47	0.92	46.5	0	0.0	0.00	1	46.5	.913821	.0061794
48	0.89	47.5	0	0.0	0.00	1	47.5	.917779	-.0277786
49	0.92	48.5	0	0.0	0.00	1	48.5	.921737	-.0017365
50	0.96	49.5	0	0.0	0.00	1	49.5	.925694	.0343056
51	0.92	50.5	0	0.0	0.00	1	50.5	.929652	-.0096524
52	0.91	51.5	0	0.0	0.00	1	51.5	.933610	-.0236103
53	0.95	52.5	0	0.0	0.00	1	52.5	.937568	.0124318
54	0.93	53.5	0	0.0	0.00	1	53.5	.941526	-.0115262
55	0.93	54.5	0	0.0	0.00	1	54.5	.945484	-.0154841

OBS	Y	X	BETA_1	BETA_2	BETA_3	BETA_4	BETA_5	YHAT	EHAT
56	0.98	55.5	0	0	0	1	55.5	0.94944	-.0305580
57	0.95	56.5	0	0	0	1	56.5	0.95340	-.0034000
58	0.97	57.5	0	0	0	1	57.5	0.95736	.0126421
59	0.97	58.5	0	0	0	1	58.5	0.96132	.0086842
60	0.96	59.5	0	0	0	1	59.5	0.96527	-.0052738
61	0.97	60.5	0	0	0	1	60.5	0.96923	.0007683
62	0.94	61.5	0	0	0	1	61.5	0.97319	-.0331896
63	0.96	62.5	0	0	0	1	62.5	0.97715	-.0171476
64	1.03	63.5	0	0	0	1	63.5	0.98111	.0488945
65	0.99	64.5	0	0	0	1	64.5	0.98506	.0049366
66	1.01	65.5	0	0	0	1	65.5	0.98902	.0209787
67	0.99	66.5	0	0	0	1	66.5	0.99298	-.0029793
68	0.99	67.5	0	0	0	1	67.5	0.99694	-.0069372
69	0.97	68.5	0	0	0	1	68.5	1.00090	-.0308951
70	1.01	69.5	0	0	0	1	69.5	1.00485	.0051469
71	0.99	70.5	0	0	0	1	70.5	1.00881	-.0188110
72	1.04	71.5	0	0	0	1	71.5	1.01277	.0272311

COMPUTING METHODS FOR LINEAR MODELS
SUBJECT TO LINEAR PARAMETRIC CONSTRAINTS

by

THOMAS M. GERIG and A. RONALD GALLANT

¹/ Assistant Professor of Statistics and Assistant Professor of Economics and Statistics, respectively. North Carolina State University, Raleigh, N.C. 27607

ABSTRACT

An efficient and accurate computational form for β which minimizes $SSE(\beta) = (y - X\beta)'(y - X\beta)$ subject to $R\beta = r$ using the Moore-Penrose g -inverse is given. No rank conditions are imposed on R or X . The results are applied (i) to estimate the parameters in a linear model which are subject to linear equality constraints and (ii) to obtain the generalized inverse of $X'X$ which yields a solution of the normal equations subject to non-estimable constraints on the parameters.

1. INTRODUCTION

We shall consider the linear model

$$y = X\beta + e$$

subject to the linear equality constraints

$$R\beta = r .$$

In some instances, the restrictions to be imposed are known to be true a priori. It is common in this case to reparametrize using the restrictions to an unconstrained model or to use the method of Lagrange multipliers; see Pringle and Rayner (1971). Our approach is more direct and, computationally, rapid and accurate. In addition, no rank conditions are imposed on X or R . An interesting by-product of our method arises when X is not of full rank and the constraints are of the form $R\beta = 0$ and are non-estimable. In this case the method produces a g -inverse $(X'X)^-$ such that $\tilde{\beta} = (X'X)^-X'y$ solves the normal equations and satisfies the restrictions. This can be used to find the g -inverse which gives estimates which satisfy "the usual constraints".

In the remainder of this section we will give our notation and set forth some properties of matrices which will be used later on.

The descriptions of the matrices appearing in the model above are as follows: y is an $(n \times 1)$ vector of observations, X is an $(n \times p)$ matrix of known constants, β is a $(p \times 1)$ vector of unknown constants, e is an $(n \times 1)$ vector of uncorrelated random variables each having mean zero and variance σ^2 , R is a $(q \times p)$ matrix of known constants, and r is a $(q \times 1)$ vector of known constants. The equations $R\beta = r$ are assumed to have at least one solution in β .

DEFINITION. (Searle, 1971, p. 16). Let A be an $(m \times n)$ matrix. Then there exists a matrix A^+ of order $(n \times m)$ which satisfies $AA^+A = A$, $A^+AA^+ = A^+$, $(AA^+)' = AA^+$, and $(A^+A)' = A^+A$. The matrix A^+ is unique and is called the Moore-Penrose g -inverse of A .

The following notation will be used throughout the paper. If A is an arbitrary $(m \times n)$ matrix, let

$$P_A = A^+A,$$

$$Q_A = I - P_A,$$

$$P_A^* = AA^+,$$

and

$$Q_A^* = I - P_A^*$$

whose dimensions are $(n \times n)$, $(n \times n)$, $(m \times m)$, and $(m \times m)$ respectively. The ranks of these matrices are, respectively, $\text{rank}(A)$, $n - \text{rank}(A)$, $\text{rank}(A)$, and $m - \text{rank}(A)$. If a is an arbitrary $(n \times 1)$ vector, let

$$\|a\| = (a'a)^{\frac{1}{2}}.$$

Notation which is specific to the model $y = X\beta + e$ subject to $R\beta = r$ is as follows:

$$\tilde{\beta} = (XQ_R)^+(y - XR^+r) + R^+r,$$

$$\text{SSE}(\beta) = (y - X\beta)'(y - X\beta) = \|y - X\beta\|^2,$$

and

$$\hat{\sigma}^2 = \text{SSE}(\hat{\beta}) / (n - \text{rank}(XQ_R))$$

whose dimensions are, respectively, $(p \times 1)$, (1×1) , and (1×1) .

The following matrix relations may be verified using the four defining properties of the Moore-Penrose g-inverse. Much of this verification may be found in Pringle and Rayner (1971) and Searle (1971).

P_A , Q_A , P_A^* , Q_A^* are symmetric and idempotent,

$$(A')^+ = (A^+)',$$

$$A^+(A^+)' = (A'A)^+,$$

$$(A'A)^+A' = A'(AA')^+ = A^+,$$

$$RR^+r = r \text{ provided } R\beta = r \text{ are consistent,}$$

$$R(\beta - R^+r) = 0 \text{ provided } R\beta = r,$$

$$Q_R(\beta - R^+r) = (\beta - R^+r) \text{ provided } R\beta = r,$$

$$P_R R^+r = R^+r,$$

$$Q_R(XQ_R)^+ = (XQ_R)^+,$$

$$Q_R(Q_R X' X Q_R)^+ Q_R = (Q_R X' X Q_R)^+,$$

$$P_R R^+ = R^+.$$

and

$$R\tilde{\beta} = r.$$

In the remaining sections we will assume that the above relations are known and will use them repeatedly without reference to this section.

2. MINIMIZATION OF ERROR SUM OF SQUARES SUBJECT TO A LINEAR CONSTRAINT

In this section, we demonstrate that the vector $\tilde{\beta}$, defined in the previous section, minimizes $(y - X\beta)'(y - X\beta)$ subject to the consistent set of linear equations $R\beta = r$. Next, we discuss a method of computation based on a FORTRAN subroutine by Businger and Golub (1969).

THEOREM 1. $\tilde{\beta}$ minimizes $SSE(\beta)$ subject to the consistent constraints $R\beta = r$.

PROOF. We will first verify that $R\tilde{\beta} = r$. There is a $\bar{\beta}$ such that $R\bar{\beta} = r$ since we assumed a consistent set of constraint equations. Then

$$\begin{aligned} R\tilde{\beta} &= R(XQ_R)^+(y - XR^+r) + RR^+r \\ &= RQ_R(XQ_R)^+(y - XR^+r) + RR^+r \\ &= 0 + RR^+R\bar{\beta} = R\bar{\beta} = r. \end{aligned}$$

We now verify that $SSE(\tilde{\beta}) \leq SSE(\bar{\beta})$ provided $\bar{\beta}$ satisfies $R\bar{\beta} = r$.

$$\begin{aligned}
\text{SSE}(\tilde{\beta}) &= \|y - X P_R \tilde{\beta} - X Q_R \tilde{\beta}\|^2 \\
&= \|y - X R^+ r - X Q_R \tilde{\beta} + X Q_R (\tilde{\beta} - \bar{\beta})\|^2 \\
&= \|y - X R^+ r - X Q_R \tilde{\beta}\|^2 + \|X Q_R (\tilde{\beta} - \bar{\beta})\|^2 \\
&\quad + 2(y - X R^+ r - X Q_R \tilde{\beta})' (X Q_R) (\tilde{\beta} - \bar{\beta}) \\
&= \|y - X P_R \tilde{\beta} - X Q_R \tilde{\beta}\|^2 + \|X Q_R (\tilde{\beta} - \bar{\beta})\|^2 \\
&\quad + 2(y - X R^+ r)' [I - (X Q_R) (X Q_R)^+] (X Q_R) (\tilde{\beta} - \bar{\beta}) \\
&= \|y - X \tilde{\beta}\|^2 + \|X Q_R (\tilde{\beta} - \bar{\beta})\|^2 + 0 \\
&\geq \text{SSE}(\tilde{\beta}) . \quad \square
\end{aligned}$$

The computation of the minimizing vector $\tilde{\beta}$ requires only elementary matrix computations provided one has a means for computing the Moore-Penrose g-inverse. This may be obtained using the singular value decomposition of a matrix.

For a given matrix A of order $(m \times n)$ with $m \geq n$ we may decompose A as

$$A = U S V' ,$$

where U is $(m \times n)$, S is an $(n \times n)$ diagonal matrix with non-negative elements, V' is $(n \times n)$ and

$$I_n = U'U = V'V = VV'.$$

This is called the singular value decomposition of A . Let s_i denote the diagonal elements of S . Set $s_i^+ = 1/s_i$ if $s_i > 0$, set $s_i^+ = 0$ if $s_i = 0$ and form the diagonal matrix S^+ from the elements s_i^+ . Then the Moore-Penrose g-inverse of A is given by

$$A^+ = VS^+U'$$

and the rank of A^+ is the same as the number of non-zero s_i . (If $m < n$ compute $B = (A')^+$ using this method and set $A^+ = B'$.)

A listing of a FORTRAN subroutine to obtain the singular value decomposition of A may be found in Businger and Golub (1969). The subroutine as listed is for a COMPLEX matrix A , but we had no difficulty in converting it to REAL*8 from the COMPLEX version. We have had good results using an IBM 370/165 setting the parameters $ETA = 1.D - 14$ and $TOL = 1.D - 60$; we set $s_i = 0$ if $s_i < s_1 \times (1 \times 10^{-13})$.

If y and X are too large for storage in fast core but the sums of squares and cross-products $y'y$, $X'y$, and $X'X$ can be stored, then the computational formula

$$\hat{\beta} = (Q_R'X'XQ_R)^+(X'y - X'XR^+r) + R^+r$$

may be used. If the formula

$$\tilde{\beta} = (XQ_R)^+(y - XR^+r) + R^+r$$

is feasible, its use should improve the accuracy of the computations by avoiding unnecessary matrix multiplications.

3. LINEAR MODELS SUBJECT TO LINEAR PARAMETRIC CONSTRAINTS

The statistical properties of $\tilde{\beta}$ when used as an estimate of β in the linear model

$$y = X\beta + e$$

subject to the a priori consistent constraints

$$R\beta = r$$

are spelt out in the following four theorems. These results are fairly well known, see for example Pringle and Rayner (1971, p. 98). We have included proofs which depend heavily on the properties of the Moore-Penrose g-inverse in the Appendix; we believe that these proofs are new.

THEOREM 2. There is a $\bar{\beta}$ of the form

$$\bar{\beta} = Ay + c$$

such that $\mathcal{E}(\lambda'\bar{\beta}) = \lambda'\beta$ for every β satisfying the consistent equations $R\beta = r$ if and only if there are vectors δ and ρ such that

$$\lambda' = \delta'X + \rho'R.$$

THEOREM 3. Let $\bar{\beta}$ be any estimator of the form $\bar{\beta} = Ay + c$ and λ be of the form $\lambda' = \delta'X + \rho'R$. If $\mathcal{E}(\lambda'\bar{\beta}) = \lambda'\beta$ for all β satisfying the consistent equations $R\beta = r$ then

$$\text{Var}(\lambda'\tilde{\beta}) \leq \text{Var}(\lambda'\bar{\beta}).$$

THEOREM 4. $\mathcal{E}(\text{SSE}(\tilde{\beta})) = [n - \text{rank}(XQ_R)]\sigma^2$ provided $R\beta = r$.

THEOREM 5. Let e be distributed as a multivariate normal with mean 0 and variance-covariance matrix $\sigma^2 I$. If Λ is a matrix of the form

$$\Lambda' = \Delta'X + P'R$$

then $\Lambda'\tilde{\beta}$ is distributed as a (possibly singular) multivariate normal with mean $\Lambda'\beta$ and variance-covariance matrix $\sigma^2 \Lambda'CA$ where

$$C = (XQ_R)^+(XQ_R)^+.$$

Moreover, $\text{SSE}(\tilde{\beta})/\sigma^2$ is independently distributed having a chi-squared distribution with $n - \text{rank}(XQ_R)$ degrees freedom.

COROLLARY. Let e be distributed as a multivariate normal with mean 0 and variance-covariance matrix $\sigma^2 I$. If Λ is a matrix of the form $\Lambda' = \Delta'X + P'R$ then under the hypothesis

$$H: \Lambda'\beta = \Lambda'\beta^0,$$

where $R\beta^0 = r$, the statistic

$$S = \frac{(\Lambda'\tilde{\beta} - \Lambda'\beta^0)'(\Lambda'CA)^+(\Lambda'\tilde{\beta} - \Lambda'\beta^0)/f_1}{\tilde{\sigma}^2}$$

where $f_1 = \text{rank}(\Lambda'CA)$ has the F-distribution with f_1 numerator degrees freedom and $f_2 = n - \text{rank}(XQ_R)$ denominator degrees freedom provided $f_1 > 0$.

In a typical application of the preceding theory, one might be interested in estimating the linear function $\lambda'\beta$ for the model $y = X\beta + e$ subject to

$R\beta = r$ assuming that the errors are normally distributed. More likely than not, the $(n + q \times p)$ matrix $\begin{pmatrix} X \\ R \end{pmatrix}$ will have rank p so that λ will, of necessity, be of the form $\lambda' = \delta'X + \rho'R$. (If the matrix $\begin{pmatrix} X \\ R \end{pmatrix}$ does not have rank p the condition must be checked directly.) By Theorem 3, the best linear unbiased estimate of $\lambda'\beta$ is $\lambda'\tilde{\beta}$. The corresponding estimate of variance is $\tilde{\sigma}^2 \lambda' C \lambda$ with $n - \text{rank}(XQ_R)$ degrees freedom by Theorem 4. Note, however, that if λ' is of the form $\lambda' = \rho'R$ then $\lambda' C \lambda = 0$ and $\lambda'\tilde{\beta} = \lambda'R^+ r$. That is, if $\lambda' = \rho'R$ then $\lambda'\beta$ is specified by the restrictions to be the constant $\lambda'R^+ r$; the data contribute nothing to the estimate. One may test a hypothesis of the form

$$H: \Lambda'\beta = \Lambda'\beta^0$$

where $R\beta^0 = r$, and

$$\Lambda' = \Delta'X + P'R$$

by applying the Corollary in a straightforward fashion.

As can be seen from the above, the application of Theorem 1 to the estimation of an estimable function $\lambda'\beta$ requires the computation of

$$\tilde{\beta} = (XQ_R)^+(y - XR^+r) + R^+r,$$

$$C = (XQ_R)^+(XQ_R)^+,$$

and

$$\tilde{\sigma}^2 = (y - X\tilde{\beta})'(y - X\tilde{\beta}) / (n - \text{rank}(XQ_R)).$$

In those applications for which the matrices y and X are too large for storage in fast core, one may compute the sum of squares and cross-product matrices $y'y$, $X'y$, and $X'X$ and substitute the formulas:

$$\tilde{\beta} = (Q_R' X' X Q_R)^+ (X'y - X'X R^+ r) + R^+ r,$$

$$C = (Q_R' X' X Q_R)^+,$$

and

$$\tilde{\sigma}^2 = (y'y - \tilde{\beta}' X' y + \tilde{\beta}' X' X \tilde{\beta}) / (n - \text{rank}(C)).$$

To illustrate the computations we will consider fitting a grafted polynomial to data collected in a nutrition study (Eppright et al., 1972). The data shown in Figure 1 are preschool boys' weight/height ratios ($= y$) plotted against age ($= x$); the numeric values are given in Table I. We will assume a model which is linear for ages above twelve months and quadratic for ages below. Further, we will require that the response function be continuous and have continuous first derivative in x . More formally, the model is

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_t^2 \quad 0 \leq x < 12$$

$$= \beta_4 + \beta_5 x_t \quad 12 \leq x \leq 72$$

subject to a continuity constraint at the point of join

$$\beta_1 + \beta_2 \cdot 12 + \beta_3 \cdot 144 - \beta_4 - \beta_5 \cdot 12 = 0$$

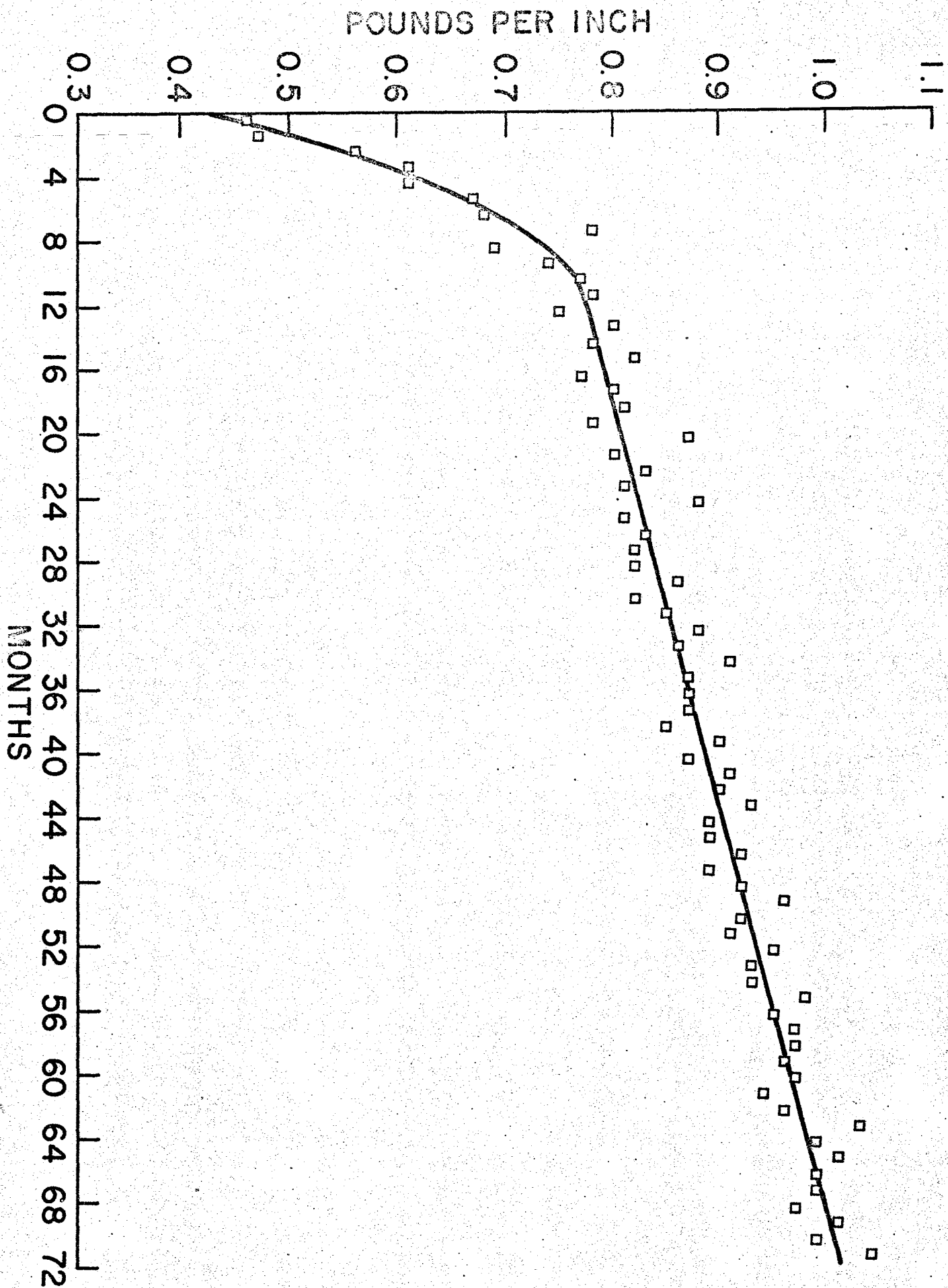


TABLE I

Boys' Weight/Height vs. Age

W/H	AGE	W/H	AGE	W/H	AGE
0.46	0.5	0.88	24.5	0.92	48.5
0.47	1.5	0.81	25.5	0.96	49.5
0.56	2.5	0.83	26.5	0.92	50.5
0.61	3.5	0.82	27.5	0.91	51.5
0.61	4.5	0.82	28.5	0.95	52.5
0.67	5.5	0.86	29.5	0.93	53.5
0.68	6.5	0.82	30.5	0.93	54.5
0.78	7.5	0.85	31.5	0.98	55.5
0.69	8.5	0.88	32.5	0.95	56.5
0.74	9.5	0.86	33.5	0.97	57.5
0.77	10.5	0.91	34.5	0.97	58.5
0.78	11.5	0.87	35.5	0.96	59.5
0.75	12.5	0.87	36.5	0.97	60.5
0.80	13.5	0.87	37.5	0.94	61.5
0.78	14.5	0.85	38.5	0.96	62.5
0.82	15.5	0.90	39.5	1.03	63.5
0.77	16.5	0.87	40.5	0.99	64.5
0.80	17.5	0.91	41.5	1.01	65.5
0.81	18.5	0.90	42.5	0.99	66.5
0.78	19.5	0.93	43.5	0.99	67.5
0.87	20.5	0.89	44.5	0.97	68.5
0.80	21.5	0.89	45.5	1.01	69.5
0.83	22.5	0.92	46.5	0.99	70.5
0.81	23.5	0.89	47.5	1.04	71.5

and a continuity constraint on the first derivative in x at the join point

$$\beta_2 + \beta_3 \cdot 24 - \beta_5 = 0.$$

The values of $\tilde{\beta}$, C , and $\tilde{\sigma}^2$ computed from these data are:

$$\tilde{\beta}' = (.424, .055, -.00213, .730, .00396),$$

$$C = \begin{bmatrix} .446 & -.0812 & .00342 & -.0469 & .000930 \\ -.0812 & .0165 & -.000702 & .0199 & -.000396 \\ .00342 & -.000702 & .0000300 & -.000900 & .0000183 \\ -.0469 & .0199 & -.000900 & .0825 & -.00171 \\ .000930 & -.000396 & .0000183 & -.00171 & .0000433 \end{bmatrix},$$

and

$$\tilde{\sigma}^2 = .000549$$

with $72 - 3 = 69$ degrees freedom. The fitted equation is plotted against the data in Figure 1. If, say, one wished to estimate the slope of the equation for ages past twelve months one would consider the estimation of $\lambda'\beta$ with $\lambda' = (0, 0, 0, 0, 1)$. Since $\begin{pmatrix} X \\ R \end{pmatrix}$ has rank 5, λ is of the form $\lambda' = \delta'X + \rho'R$. The estimate of slope is $\lambda'\tilde{\beta} = \tilde{\beta}_5 = .00396$ with standard error

$$(\lambda' C \lambda' \tilde{\sigma}^2)^{\frac{1}{2}} = (c_{55} \tilde{\sigma}^2)^{\frac{1}{2}} = .000154.$$

4. SOLVING NORMAL EQUATIONS SUBJECT TO NON-ESTIMABLE CONSTRAINTS

Consider the problem of solving the normal equations

$$X'X\beta = X'y$$

subject to the non-estimable parametric constraints

$$R\beta = 0.$$

By non-estimable, we mean that there do not exist non-zero vectors δ and ρ such that

$$\delta'X = \rho'R.$$

Our solution consists of showing that

$$C = (Q_R'X'XQ_R)$$

is a generalized inverse of $X'X$ provided that the restrictions $R\beta = 0$ are non-estimable. It then follows (Searle, 1971, p. 8) that

$$\tilde{\beta} = CX'y$$

is a solution of the normal equations. We have seen earlier that $R\tilde{\beta} = 0$ thus satisfying the non-estimable parametric constraints.

THEOREM 6. If there do not exist non-zero vectors δ and ρ such that $\delta'X = \rho'R$ then

$$C = (Q_R' X' X Q_R)^+$$

is a generalized inverse of $X'X$.

PROOF. Let

$$T = \begin{bmatrix} I & -XR^+ \\ 0 & I \end{bmatrix}$$

$$Z = \begin{bmatrix} XQ_R \\ R \end{bmatrix} = T \begin{bmatrix} X \\ R \end{bmatrix}.$$

Since T is non-singular, the ranks of Z and $\begin{pmatrix} X \\ R \end{pmatrix}$ are the same. Since the linear spaces generated by the rows of X and R are disjoint by hypothesis we have that $\text{rank}\begin{pmatrix} X \\ R \end{pmatrix} = \text{rank}(X) + \text{rank}(R)$. Since the linear spaces generated by the rows of XQ_R and R are orthogonal $\text{rank}(Z) = \text{rank}(XQ_R) + \text{rank}(R)$. Thus $\text{rank}(XQ_R) = \text{rank}(X)$. Since the columns of XQ_R are linear combinations of the columns of X and the ranks of the two matrices are equal, there is a non-singular matrix S such that $XQ_R = XS$. Thus,

$$X'X(Q_R'X'XQ_R)^+X'X = S'^{-1}S'X'XQ_R(Q_R'X'XQ_R)^+Q_R'X'XS^{-1}$$

$$= S'^{-1}Q_R'X'XQ_R(Q_R'X'XQ_R)^+Q_R'X'XQ_RS^{-1}$$

$$= S'^{-1}Q_R'X'XQ_RS^{-1} = S'^{-1}S'X'XS^{-1}$$

$$= X'X. \quad \square$$

As an example of the use of this theorem in applications consider a complete randomized block design with two treatments and two blocks:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ b_1 \\ b_2 \end{bmatrix} + e$$

subject to the usual non-estimable parametric constraints.

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The Moore-Penrose g-inverse of R is

$$R^+ = \begin{bmatrix} 0 & 0 \\ \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$$

so that

$$Q_R' X' X Q_R = \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

Thus,

$$C = \begin{bmatrix} \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & -\frac{1}{4} & 0 & 0 \\ 0 & -\frac{1}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & -\frac{1}{4} \\ 0 & 0 & 0 & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}.$$

It is easy to verify that

$$\begin{bmatrix} \tilde{\mu} \\ \tilde{\sigma}_1 \\ \tilde{\sigma}_2 \\ \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix} = CX'y = \frac{1}{4} \begin{bmatrix} y_{11} + y_{12} + y_{21} + y_{22} \\ y_{11} + y_{12} - y_{21} - y_{22} \\ -y_{11} - y_{12} + y_{21} + y_{22} \\ y_{11} - y_{12} + y_{21} - y_{22} \\ -y_{11} + y_{12} - y_{21} + y_{22} \end{bmatrix},$$

and that these are the usual parameter estimates.

REFERENCES

- Businger, P. A. and Golub, G. H. (1969). "Singular value decomposition of a complex matrix". Communications of the ACM, 12, 564-65.

- Eppright, E. S., Fox, H. M., Fryer, B. A., Lamkin, G. H., Vivian, V. M., and Fuller, E. S. (1972). "Nutrition of infants and preschool children in the north central region of the United States of America". World Review of Nutrition and Dietetics, 14, 269-332.
- Pringle, R. M. and Raynor, A. A. (1971). Generalized Inverse Matrices with Applications to Statistics. Hafner Publishing Company, New York.
- Searle, S. R. (1971). Linear Models. John Wiley and Sons, Inc., New York.

APPENDIX

PROOF OF THEOREM 2. (If) Let $\lambda' = \delta'X + \rho'R$

$$e(\lambda'\beta) = \lambda'(XQ_R)^+(X\beta - XR^+r) + \lambda'R^+r$$

$$= (\delta'X + \rho'R)Q_R(XQ_R)^+X(\beta - R^+r) + (\delta'X + \rho'R)R^+r$$

$$= \delta'XQ_R(XQ_R)^+X(\beta - R^+r) + \delta'XR^+r + \rho'RR^+r$$

$$= \delta'XQ_R(XQ_R)^+XQ_R(\beta - R^+r) + \delta'XR^+r + \rho'R\beta$$

$$= \delta'XQ_R(\beta - R^+r) + \delta'XR^+r + \rho'R\beta$$

$$= \delta'X(\beta - R^+r) + \delta'XR^+r + \rho'R\beta$$

$$= (\delta'X + \rho'R)\beta$$

$$= \lambda'\beta$$

(Only if) If β is of the form

$$\beta = R^+ r + Q_R \gamma$$

then $R\beta = r$ for all choices of γ . We will take β to be of this form and examine the consequences of various choices of γ under the assumption that there is a

$$\bar{\beta} = Ay + c$$

such that $\varepsilon(\lambda'\bar{\beta}) = \lambda'\beta$ for all γ . Under this assumption, for all γ

$$\lambda'AXR^+r + \lambda'AXQ_R\gamma + \lambda'c = \lambda'R^+r + \lambda'Q_R\gamma.$$

First set $\gamma = 0$, hence

$$\lambda'AXR^+r + \lambda'c = \lambda'R^+r,$$

so that

$$\lambda'AXQ_R\gamma = \lambda'Q_R\gamma$$

for all choices of γ . By successive choice of the elementary vectors for γ we obtain

$$\lambda'AXQ_R = \lambda'Q_R$$

whence

$$\begin{aligned}
\lambda' &= \lambda'AXQ_R + \lambda'P_R \\
&= \lambda'AX + \lambda'AXP_R + \lambda'P_R \\
&= [\lambda'A]X + [\lambda'AXR^+ + \lambda'R^+]R \\
&= \delta'X + \rho'R . \quad \square
\end{aligned}$$

PROOF OF THEOREM 3. From the proof of the previous theorem we have

$\lambda'AXQ_R = \lambda'Q_R$. The variance of $\lambda'\tilde{\beta}$ is

$$\begin{aligned}
\text{Var}(\lambda'\tilde{\beta}) &= \lambda'(XQ_R)^+(XQ_R)^+\lambda\sigma^2 \\
&= \lambda'Q_R(XQ_R)^+(XQ_R)^+Q_R\lambda\sigma^2 \\
&= \lambda'Q_R(Q_RX'XQ_R)^+Q_R\lambda\sigma^2 .
\end{aligned}$$

Let $W = XQ_R$, the variance of $\lambda'\tilde{\beta}$ is

$$\begin{aligned}
\text{Var}(\lambda'\tilde{\beta}) &= \lambda'AA'\lambda\sigma^2 \\
&= \lambda'A[P_W^* + Q_W^*]A'\lambda\sigma^2 \\
&= \lambda'A(XQ_R)(XQ_R)^+A'\lambda\sigma^2 + \lambda'AQ_W^*A'\lambda\sigma^2 \\
&= \lambda'A(XQ_R)(Q_RX'XQ_R)^+(XQ_R)^+A'\lambda\sigma^2 + \lambda'AQ_W^*A'\lambda\sigma^2 \\
&= \lambda'Q_R(Q_RX'XQ_R)^+Q_R\lambda\sigma^2 + \lambda'AQ_W^*A'\lambda\sigma^2 \\
&= \text{Var}(\lambda'\tilde{\beta}) . \quad \square
\end{aligned}$$

PROOF OF THEOREM 4. Let $W = XQ_R$. Then

$$\begin{aligned}
 \text{SSE}(\tilde{\beta}) &= \|y - X\tilde{\beta}\|^2 \\
 &= \|y - X(XQ_R)^+(y - XR^+r) - XR^+r\|^2 \\
 &= \|y - XQ_R(XQ_R)^+(y - XR^+r) - XR^+r\|^2 \\
 &= \|Q_W^*e + Q_W^*X(\beta - R^+r)\|^2.
 \end{aligned}$$

Now

$$Q_W^*X(\beta - R^+r) = Q_W^*W(\beta - R^+r) + Q_W^*XP_R(\beta - R^+r) = 0$$

since $Q_W^*W = 0$ and $P_R(\beta - R^+r) = 0$ provided $R\beta = r$. We now have that

$$\text{SSE}(\tilde{\beta}) = e'Q_W^*e,$$

where Q_W^* is symmetric and idempotent with rank $Q_W^* = n - \text{rank}(W)$. Thus

$$E(e'Q_W^*e) = [n - \text{rank}(W)]\sigma^2. \quad \square$$

PROOF OF THEOREM 5. We may rewrite β as follows.

$$\begin{aligned}
\Lambda' \tilde{\beta} &= \Lambda' (XQ_R)^+ (X\beta + e - XR^+r) + \Lambda'R^+r \\
&= \Lambda' (XQ_R)^+ e + \Lambda' (XQ_R)^+ X(\beta - R^+r) + \Lambda'R^+r \\
&= \Lambda' (XQ_R)^+ e + (\Lambda'X + P'R)Q_R (XQ_R)^+ X(\beta - R^+r) + (\Lambda'X + P'R)R^+r \\
&= \Lambda' (XQ_R)^+ e + \Lambda'XQ_R (XQ_R)^+ X(\beta - R^+r) + \Lambda'XR^+r + P'RR^+r \\
&= \Lambda' (XQ_R)^+ e + \Lambda' (XQ_R) (XQ_R)^+ XQ_R (\beta - R^+r) + \Lambda'XR^+r + P'RB \\
&= \Lambda' (XQ_R)^+ e + \Lambda'XQ_R (\beta - R^+r) + \Lambda'XR^+r + P'RB \\
&= \Lambda' (XQ_R)^+ e + \Lambda'X(\beta - R^+r) + \Lambda'XR^+r + P'RB \\
&= \Lambda' (XQ_R)^+ e + (\Lambda'X + P'R)\beta \\
&= \Lambda' (XQ_R)^+ e + \Lambda'\beta .
\end{aligned}$$

Thus, $\Lambda' \tilde{\beta}$ has the (possibly singular) multivariate normal distribution with mean $\Lambda'\beta$ and variance-covariance $\sigma^2 \Lambda' C \Lambda$.

Let $W = XQ_R$. From the proof of Theorem 4 we have that $SSE(\tilde{\beta})/\sigma^2 = e'Q_W^*e/\sigma^2$ where Q_W^* is symmetric and idempotent with rank $n - \text{rank}(XQ_R)$. Thus, $SSE(\tilde{\beta})/\sigma^2$ has the chi-squared distribution with $n - \text{rank}(XQ_R)$ degrees freedom.

Consider the random variables $(XQ_R)^+ e + \beta$ and $Q_W^* e$. Since their covariance matrix $\sigma^2 (XQ_R)^+ Q_W^* = 0$ they are independent. As a consequence, the random variables $\Lambda' \tilde{\beta} = \Lambda' (XQ_R)^+ e + \Lambda'\beta$ and $SSE(\tilde{\beta}) = e'Q_W^*Q_W^*e$ are independent. $\square$