ON LINEAR MODELS WITH RESTRICTIONS ON PARAMETERS

Shayle R. Searle
Departments of Biometrics and Statistics
Cornell University
Ithaca, N. Y., 14853 U.S.A.

BU-1450-M

September 1999

Key Words: estimation, full model, hypotheses, reduced model

Abstract

Two procedures are discussed (and shown equivalent) for obtaining the best linear unbiased estimator of $\boldsymbol{\beta}$ in the linear model $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta},\ \sigma^2\mathbf{I})$ in the presence of linear restrictions on $\boldsymbol{\beta}$.

1. The Model, with Linear Restrictions

1.1 The model

We deal with the familiar linear model $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ with $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta},\ \sigma^2\mathbf{I})$ for $\mathbf{y}$ being a vector of N data values and $\boldsymbol{\beta}$ a vector of p fixed effects parameters. $\mathbf{X}$ of order $N \times p$ is taken as having rank $p - m$ for $m > 0$. Linear restrictions are taken as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{c} \tag{1}$$

for $\mathbf{H}$ and $\mathbf{c}$ known, with $\mathbf{H}_{m\times p}$ being of maximal row rank with its rows linearly independent of those of $\mathbf{X}$, meaning that there are no non-null vectors $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$ for which $\boldsymbol{\nu}'\mathbf{H} = \boldsymbol{\lambda}'\mathbf{X}$. The oft-used purpose of such restrictions is to reduce a model which is not of full rank to being one that is. An example for $\mathrm{E}(y_i) = \mu + \alpha_i$ is to use the restriction $\sum \alpha_i = 0$.

1.2 Restrictions contrasted with hypotheses

It is well known when testing the hypothesis

$$H{:}\, \mathbf{K}'\boldsymbol{\beta} = \mathbf{t} \tag{2}$$

that the estimator of $\boldsymbol{\beta}$ under the hypothesis is

$$\boldsymbol{\beta} = \boldsymbol{\beta}^0 - (\mathbf{X}'\mathbf{X})^-\mathbf{K}[\mathbf{K}'(\mathbf{X}'\mathbf{X})^-\mathbf{K}]^{-1}(\mathbf{K}'\boldsymbol{\beta}^0 - \mathbf{t}) \tag{3}$$

where $(\mathbf{X}'\mathbf{X})^-$ is a generalized inverse of $(\mathbf{X}'\mathbf{X})^-$ and $\boldsymbol{\beta}^0 = (\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$. In developing (3), the fact that $\mathbf{K}\boldsymbol{\beta} = \mathbf{t}$ of (2) is a hypothesis means that $\mathbf{K}'\boldsymbol{\beta}$ is an estimable function so that $\mathbf{K}'$ has the form $\mathbf{T}'\mathbf{X}$ for some $\mathbf{T}'$; i.e., $\mathbf{K}' = \mathbf{T}'\mathbf{X}$.

Thus the $\mathbf{H}$ of restrictions (1) contrasts sharply with $\mathbf{K}'$ of the hypothesis (2): rows of $\mathbf{H}$ are linearly independent (as elucidated in Section 1.1) of those of $\mathbf{X}$, whereas rows of $\mathbf{K}'$ are, in fact, linear combinations of (i.e., linearly dependent on) rows of $\mathbf{X}$. Further, although $\mathbf{H}$ and $\mathbf{K}$ both have full row rank, $\mathbf{H}$ has maximal row rank whereas $\mathbf{K}'$ does not necessarily have that property. It is these two distinctions of $\mathbf{H}$ from $\mathbf{K}'$ which lead to an estimator of $\boldsymbol{\beta}$ under restrictions $\mathbf{H}\boldsymbol{\beta} = \mathbf{c}$ being different from that under the hypothesis $H{:}\,\mathbf{K}'\boldsymbol{\beta} = \mathbf{t}$ (even when $\mathbf{t} \equiv \mathbf{c}$).

2. Estimation

Estimation of $\boldsymbol{\beta}$ under restrictions $\mathbf{H}\boldsymbol{\beta} = \mathbf{t}$ can be done either by what we call the full model procedure or by the reduced model procedure. These procedures are presented and shown equivalent. In each case the basis of estimation is best linear unbiased estimation (BLUE) or, equivalently, maximum likelihood (ML) based on assuming that $\mathbf{y}$ has a multivariate normal distribution.

2.1 The full model procedure

In this procedure one has to minimize $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})$ subject to $\mathbf{H}\boldsymbol{\beta} = \mathbf{c}$. This can be achieved by minimizing $(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\theta}'(\mathbf{H}\boldsymbol{\beta}-\mathbf{c})$ with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ where $2\boldsymbol{\theta}$ is a vector of Lagrange multipliers (the 2 for pure convenience). This yields equations

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{c} \end{bmatrix} \tag{4}$$

where $\hat{\boldsymbol{\beta}}$ is the desired estimator. On defining

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{H}' \\ \mathbf{H} & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}, \tag{5}$$

Searle (1971, pp. 21-3) gives considerable detail (from a number of references) for deriving the $\mathbf{B}$-submatrices in (5). In particular, on using

$$\mathbf{Q}^{-1} = (\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})^{-1}, \tag{6}$$

which exists because $[\mathbf{X}' \quad \mathbf{H}']$ has full row rank, those details lead to

$$\mathbf{B}_{11} = \mathbf{Q}^{-1} - \mathbf{D}(\mathbf{D}'\mathbf{H}'\mathbf{H}\mathbf{D}^{-1}\mathbf{D}' \quad \text{and} \quad \mathbf{B}_{12} = \mathbf{D}(\mathbf{H}\mathbf{D}]^{-1}. \tag{7}$$

For $\mathbf{D}$ in these results we can use

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{H}' \quad \text{and } \mathbf{H}\mathbf{D} = \mathbf{I} \tag{8}$$

as noted by Pringle and Rayner (1967), and Chipman (1964). Then (8) gives

$$\mathbf{H}\mathbf{Q}^{-1}\mathbf{H}' = \mathbf{I}. \tag{9}$$

Using this in

$$\mathbf{H}\mathbf{Q}^{-1}(\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H}) = \mathbf{H}$$

based on (6), gives $\mathbf{H}\mathbf{Q}^{-1}\mathbf{X}'\mathbf{X} = \mathbf{0}$ and hence

$$\mathbf{H}\mathbf{Q}^{-1}\mathbf{X}' = \mathbf{0}. \tag{10}$$

Then substituting (8) into (7) and then into (5) and (4) gives

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \mathbf{B}_{11}\mathbf{X}'\mathbf{y} + \mathbf{B}_{12}\mathbf{c} \\ &= (\mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{H}'\mathbf{H}\mathbf{Q}^{-1})\mathbf{X}'\mathbf{y} + \mathbf{Q}^{-1}\mathbf{H}'\mathbf{c} \\ &= \mathbf{Q}^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{H}'\mathbf{c}). \end{aligned} \tag{11}$$

2.2 The reduced model procedure

The reduced model procedure reduces the model $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ by replacing in $\boldsymbol{\beta}$ a subset of elements of $\boldsymbol{\beta}$ which can, from the restrictions $\mathbf{H}\boldsymbol{\beta} = \mathbf{c}$, be expressed in terms of the other elements of $\boldsymbol{\beta}$. We then apply BLUE or ML directly to the model for $\mathbf{y}$ which results from this replacement. To do this necessitates partitioning $\mathbf{H}$, $\mathbf{X}$ and $\boldsymbol{\beta}$.

2.2a Reducing the model

With $\mathbf{H}_{m\times p}$ having full row rank, write it as

$$\mathbf{H} = [\mathbf{H}_1 \quad \mathbf{H}_2] = \mathbf{H}_1[\mathbf{I} \quad \mathbf{H}_1^{-1}\mathbf{H}_2] \tag{12}$$

for $\mathbf{H}_1$ of order $m \times m$ and nonsingular (ignoring the possible need for permuting columns of $\mathbf{H}$ so that $\mathbf{H}_1$ is nonsingular). Similarly write $\mathbf{X} = [\mathbf{X}_1 \quad \mathbf{X}_2]$ and $\boldsymbol{\beta} = [\boldsymbol{\beta}_1' \quad \boldsymbol{\beta}_2']'$ for $\mathbf{X}_1$ being $N \times m$ and $\boldsymbol{\beta}_1$ being $m \times 1$. Then from the restrictions $\mathbf{H}\boldsymbol{\beta} = \mathbf{H}_1\boldsymbol{\beta}_1 + \mathbf{H}_2\boldsymbol{\beta}_2 = \mathbf{c}$,

$$\boldsymbol{\beta}_1 \;=\; \mathbf{H}_1^{-1}\mathbf{c} - \mathbf{H}_1^{-1}\mathbf{H}_2\boldsymbol{\beta}_2. \tag{13}$$

This gives

$$\mathbf{y} \;=\; \mathbf{X}\boldsymbol{\beta} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 = (\mathbf{X}_2 - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{H}_2)\boldsymbol{\beta}_2 + \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c}. \tag{14}$$

Define

$$\mathbf{S} = \mathbf{X}_2 - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{H}_2 = \mathbf{X}\begin{bmatrix} -\mathbf{H}_1^{-1}\mathbf{H}_2 \\ \mathbf{I} \end{bmatrix}. \tag{15}$$

Then the model equation (14) can be written as

$$\mathrm{E}(\mathbf{y} - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c}) = \mathbf{S}\boldsymbol{\beta}_2. \tag{16}$$

We call this the reduced model.

2.2b Estimation

Estimating $\boldsymbol{\beta}_2$ from (16) relies on the following lemma.

Lemma 1: $\mathbf{S}$ has full column rank.

Proof (due to Bittner, 1974) With the rank p matrix

$$\begin{bmatrix} \mathbf{H} \\ \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 \\ \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} \mathbf{H}_1^{-1} & \mathbf{0} \\ -\mathbf{X}_1\mathbf{H}_1^{-1} & \mathbf{I} \end{bmatrix} \text{ being nonsingular,}$$

we have from the identity

$$\begin{bmatrix} \mathbf{H}_1^{-1} & \mathbf{0} \\ -\mathbf{X}_1\mathbf{H}_1^{-1} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{H} \\ \mathbf{X} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & \mathbf{H}_1^{-1}\mathbf{H}_2 \\ \mathbf{0} & \mathbf{S} \end{bmatrix}$$

that the left-hand side has rank p and the right-hand side has rank m + rank(**S**). Therefore $p = m + \text{rank}(\mathbf{S})$, so that **S** has rank $p - m$, which is its number of columns. Q.E.D.

Applying least squares to (16) and using the lemma in doing so gives the estimator

$$\tilde{\beta}_2 = (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{y} - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c}) \tag{17}$$

and then from (13) we get

$$\tilde{\beta}_1 = \mathbf{H}_1^{-1}\mathbf{c} - \mathbf{H}_1^{-1}\mathbf{H}_2\tilde{\beta}_2$$

so that

$$\tilde{\beta} = \begin{bmatrix} \tilde{\beta}_1 \\ \tilde{\beta}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{H}_1^{-1}\mathbf{c} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} -\mathbf{H}_1^{-1}\mathbf{H}_2 \\ \mathbf{I} \end{bmatrix} (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{y} - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c}). \tag{18}$$

3. Equality of the two procedures

The full model procedure yields $\hat{\beta} = \mathbf{Q}^{-1}(\mathbf{X}'\mathbf{y} + \mathbf{H}'\mathbf{c})$ of (11) as the estimator. This looks surprisingly different from $\tilde{\beta}$ of (18) yielded by the reduced model procedure. Yet, as one would hope and expect, the two estimators are the same — as is now shown, using the following lemma.

Lemma 2: $\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{X} = \mathbf{X}$.

Proof Because **S** is rectangular with full column rank, linear relationships exist among its rows which (on ignoring the possible need for permuting rows) can be represented as

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{R}\mathbf{S}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{R} \end{bmatrix} \mathbf{S}_1 \tag{19}$$

for $\mathbf{S}_1$ nonsingular, and some $\mathbf{R}$. But (15) shows $\mathbf{S}$ as a product with $\mathbf{X}$ being the pre-multiplier. Therefore from (19) it is possible to write $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{I} \\ \mathbf{R} \end{bmatrix} \mathbf{X}_0 \tag{20}$$

for some $\mathbf{X}_0$. Straightforward algebra now reduces $\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'\mathbf{X}$ to $\mathbf{X}$, based on substitutions from (19) and (20), and on recalling that with $\mathbf{R}$ being real (as it is), $\mathbf{I} + \mathbf{R}'\mathbf{R}$ is nonsingular. Q.E.D.

It is now easy to show that $\mathbf{Q}\tilde{\boldsymbol{\beta}} = \mathbf{Q}\hat{\boldsymbol{\beta}}$:

$$\begin{aligned}
\mathbf{Q}\tilde{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H})\tilde{\boldsymbol{\beta}} \\
&= \{\mathbf{X}'[\mathbf{X}_1 \quad \mathbf{X}_2] + \mathbf{H}'[\mathbf{H}_1 \quad \mathbf{H}_2]\} \begin{bmatrix} \mathbf{H}_1^{-1}\mathbf{c} \\ \mathbf{0} \end{bmatrix} \\
&= \quad + [\mathbf{X}'\mathbf{X} + \mathbf{H}'\mathbf{H}] \begin{bmatrix} -\mathbf{H}_1^{-1}\mathbf{H}_2 \\ \mathbf{I} \end{bmatrix} (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{y} - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c}), \text{ from (18)} \\
&= \mathbf{X}'\mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c} + \mathbf{0} + \mathbf{H}'\mathbf{c} + \mathbf{0} + \mathbf{X}'\mathbf{X} \begin{bmatrix} -\mathbf{H}_1^{-1}\mathbf{H}_2 \\ \mathbf{I} \end{bmatrix} (\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{y} - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c}) \\
&= \mathbf{X}'\mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c} + \mathbf{H}'\mathbf{c} + \mathbf{X}'\mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}'(\mathbf{y} - \mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c}), \text{ from (15)} \\
&= \mathbf{X}'\mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c} + \mathbf{H}'\mathbf{c} + \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}_1\mathbf{H}_1^{-1}\mathbf{c} \text{ from Lemma 2} \\
&= \mathbf{H}'\mathbf{c} + \mathbf{X}'\mathbf{y} \\
&= \mathbf{Q}\hat{\boldsymbol{\beta}}, \text{ from (11).}
\end{aligned}$$

Thus $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$, and so the full model and the reduced model procedures are equivalent.

References

Bittner, A.C. (1974) Exact linear restrictions on parameters in a linear regression model. *The American Statistician* **28**, 35-36.

Chipman, J.S. (1964) On least squares with insufficient observations. *Journal of the American Statistical Association* **59**, 1078-1111.

Rayner, A.A. and Pringle, R.M. (1967) A note on generalized inverses in the linear hypothesis not of full rank. *Annals of Mathematical Statistics* **38**, 271-273.

Searle, S.R. (1971, 1998) *Linear Models*, Wiley & Sons, New York.