1 Solutions with Linear Restrictions

If the model has less than full rank, the estimate of β can be biased in general, and only certain linear combinations of β are estimable. Those linear combinations correspond exactly to linear combinations of $\mathrm{E}(x_i^T\beta)$, where x_i is a row of X, or, in the coordinate-free approach, to linear combinations of the elements of the mean vector μ. This suggests several sensible approaches to the problem of estimation with rank deficient models.

One possibility is to choose any $\hat{\beta}$ and proceed with the characterizations and use of estimable functions. This is potentially complex, especially in unbalanced models with many factors. The book by Searle (1971), for example, exemplifies this approach.

As an alternative, one can consider redefining the problem as follows. Given a fixed linear model $Y = X\beta + \varepsilon$ with X less than full rank, find an appropriate basis for $\mathcal{C}(X)$. If that basis is given by $\{z_1, \ldots, z_r\}$, and the matrix whose columns are the z_i is Z, then fit the full rank model $Y = Z\gamma + \varepsilon$. All estimable functions in the original formulation are of course still estimable. This corresponds exactly to the coordinate-free approach that is at the heart of these notes.

In the one-way anova example, we can simply delete the column of 1s to produce a full rank model. Both R and Arc set $\alpha_1 = 0$ to get a full rank model, but Splus uses a different method, at least by default.

Occasionally, the βs may have some real meaning and we don't wish to remove columns from X. In this case we might produce a unique full rank solution by placing restrictions on β of the form

$$d_i^T\beta = 0$$

The restricted normal equations are then

$$X^TX\tilde{\beta} = X^Ty$$

$$d_i^T\tilde{\beta} = 0, i = 1, 2, \ldots, t \geq p - r$$

To choose the d_i, it makes sense to require that the estimable functions in the original problem be the same as those in the constrained problem. We know that $c^T\beta$ is estimable if and only if $c \in \mathcal{C}(X^T)$, so this is equivalent to $d_i \notin \mathcal{C}(X^T)$. Otherwise, we would be restricting estimable functions.

For a general statement, let $\Delta^T = (d_1, \ldots, d_t), t \geq p - r$ be the matrix specifying the restrictions. Then:

Theorem 1 *The system:*

$$\begin{pmatrix} X \\ \Delta \end{pmatrix}\beta = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \textit{ or } \begin{pmatrix} X^TX \\ \Delta \end{pmatrix}\beta = \begin{pmatrix} X^Ty \\ 0 \end{pmatrix}$$

has a unique solution $\hat{\beta}$ for all μ and y if and only if:

1. $\rho\left[\begin{pmatrix} X \\ \Delta \end{pmatrix}\right] = p$
2. $\mathcal{C}(X^T) \cap \mathcal{C}(\Delta^T) = 0$. *This says that all functions of the form $a^T\Delta\beta$ are not estimable.*

Proof. Only an informal justification is given. Part 1 guarantees the uniqueness of a solution. The set of solutions to the unrestricted normal equations is given by $\hat{\beta}_0 + \mathrm{N}(X^T X)$ for some β_0. If we can, we ensure that the solution to the restricted normal equations, which is now unique, is an element of this flat, we are done. As long as the rows of Δ lie in the space $\mathrm{N}(X^T X)$, then a restriction is placed on $N(X^T X)$ but not on $\mathcal{C}(X^T)$. Thus, Part (1) ensures uniqueness, and Part (2) ensures that the resulting estimate is an element of the original flat. QED

Here are some additional results concerning linear estimation with restrictions, stated without proof.

Corollary 1 *Consider a linear model parameterized as $Y = X\beta + \varepsilon$, and an additional set of restrictions $\Delta\beta = 0$. Let $\tilde{\beta}$ be a vector of parameters that simultaneously satisfies:*

$$X\beta = X\tilde{\beta} \text{ and } \Delta\tilde{\beta} = 0$$

The first condition says that β and $\tilde{\beta}$ give the same μ-vector and the second specifies the conditions. These are conditions on parameters, not on estimates. Then there is a unique $\hat{\beta}$ for every β provided that the conditions of Theorem 1 are satisfied. That is, there is a unique OLS *estimate that satisfies the side conditions.*

Corollary 2 *If the conditions of Theorem 1 are satisfied, so that the $\tilde{\beta}_j$ are functions of the β_j determined uniquely by the conditions of Corollary 1, then the $\tilde{\beta}_j$ are estimable functions.*

Proof. We first obtain a formula for the $\tilde{\beta}_j$ in terms of the β_j. For any β, let $\tilde{\beta}$ be the unique solution to

$$X\beta = X\tilde{\beta} \text{ and } \Delta\tilde{\beta} = 0$$

or equivalently,

$$\begin{aligned}
\begin{pmatrix} X \\ \Delta \end{pmatrix} \tilde{\beta} &= \begin{pmatrix} X\beta \\ 0 \end{pmatrix} \\
\begin{pmatrix} X^T & \Delta^T \end{pmatrix} \begin{pmatrix} X \\ \Delta \end{pmatrix} \tilde{\beta} &= \begin{pmatrix} X^T & \Delta^T \end{pmatrix} \begin{pmatrix} X\beta \\ 0 \end{pmatrix} \\
(X^T X + \Delta^T \Delta)\tilde{\beta} &= \begin{pmatrix} X^T & \Delta^T \end{pmatrix} \begin{pmatrix} \mu \\ 0 \end{pmatrix}
\end{aligned}$$

By Theorem 1, $G = \begin{pmatrix} X \\ \Delta \end{pmatrix}$ has full column rank, and hence we can write

$$\begin{aligned}
\tilde{\beta} = (G^T G)^{-1} G^T \begin{pmatrix} \mu \\ 0 \end{pmatrix} &= [X^T X + \Delta^T \Delta]^{-1} X^T \mu \\
&= [X^T X + \Delta^T \Delta]^{-1} X^T X \beta \qquad (1)
\end{aligned}$$

and thus the restricted parameter vector is an explicit linear combination on the unrestricted parameter vector. An unbiased estimate of $\tilde{\beta}$ is thus $[X^T X + \Delta^T \Delta]^{-1} X^T \hat{\mu}$ and all linear combinations of $\tilde{\beta}$ are estimable, including elements of $\tilde{\beta}$.

To see that the estimable functions are the same as those in the original problem, write

$$\tilde{c}^T \tilde{\beta} = (\tilde{c}^T [X^T X + \Delta^T \Delta]^{-1}) X^T \mu = (\tilde{c}^T [X^T X + \Delta^T \Delta]^{-1})^T X \beta = c^T \beta$$

for $c = (X^T X)[X^T X + \Delta^T \Delta]^{-1} \tilde{c}$. QED

1.1 More on one way anova

For the one way anova model, the "usual constraint" is of the form $\sum a_i \alpha_i = 0$. Most typically, one takes all the $a_i = 1$, which comes from writing:

$$\begin{aligned} y_{ij} &= \mu_i + \varepsilon_{ij} \\ &= \bar{\mu} + (\mu_i - \bar{\mu}) + \varepsilon_{ij} \\ &= \mu + \alpha_i + \varepsilon_{ij} \end{aligned}$$

Now if $X = (J, X_1, \ldots, X_p)$; then

$$X^T y = \begin{pmatrix} \sum y_i \\ y_{1+} \\ \vdots \\ y_{p+} \end{pmatrix}$$

$$X^T X = \begin{pmatrix} n & n_1 & \cdots & n_p \\ n_1 & n_1 & \cdots & 0 \\ \vdots & & \ddots & \\ n_p & 0 & \cdots & n_p \end{pmatrix}$$

$$\Delta^T \Delta = \begin{pmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} (0, 1, \cdots, 1)$$

Suppose that $\hat{\beta}_0$ is any unrestricted least squares solution and $\hat{\mu} = X\hat{\beta}_0$ is the estimated mean vector. $\hat{\mu}$ is of course the same for any choice of $\hat{\beta}$. Then the restricted least squares solution can be found by substituting $\hat{\mu}_0$ for $X\hat{\beta}$ in (1). A bit of algebra will verify that:

$$\hat{\hat{\beta}}^T = (\bar{y}_{++}, \bar{y}_{1+} - \bar{y}_{++}, \ldots, \bar{y}_{p+} - \bar{y}_{++})$$