Chapter 9
Biased Estimation and Prediction

C. R. Henderson

1984 - Guelph

All methods for estimation and prediction in previous chapters have been unbiased. In this chapter we relax the requirement of unbiasedness and attempt to minimize the mean squared error of estimation and prediction. Mean squared error refers to the sum of prediction error variance plus squared bias. In general, biased predictors and estimators exist that have smaller mean squared errors than BLUE and BLUP. Unfortunately, we never know what are truly minimum mean squared error estimators and predictors because we do not know some of the parameters required for deriving them. But even for BLUE and BLUP we must know **G** and **R** at least to proportionality. Additionally for minimum mean squared error we need to know squares and products of $\boldsymbol{\beta}$ at least proportionally to **G** and **R**.

1 Derivation Of BLBE And BLBP

Suppose we want to predict $\mathbf{k}_1'\boldsymbol{\beta}_1 + \mathbf{k}_2'\boldsymbol{\beta}_2 + \mathbf{m}'\mathbf{u}$ by a linear function of $\mathbf{y}$, say $\mathbf{a}'\mathbf{y}$, such that the predictor has expectation $\mathbf{k}_1'\boldsymbol{\beta}_1$ plus some function of $\boldsymbol{\beta}_2$, and in the class of such predictors, has minimum mean squared error of prediction, which we shall call BLBP (best linear biased predictor).

The mean squared error (MSE) is

$$\mathbf{a}'\mathbf{R}\mathbf{a} + (\mathbf{a}'\mathbf{X}_2 - \mathbf{k}_2')\boldsymbol{\beta}_2\boldsymbol{\beta}_2'(\mathbf{X}_2'\mathbf{a} - \mathbf{k}_2) + (\mathbf{a}'\mathbf{Z} - \mathbf{m}')\mathbf{G}(\mathbf{Z}'\mathbf{a} - \mathbf{m}). \quad (1)$$

In order that $E(\mathbf{a}'\mathbf{y})$ contains $\mathbf{k}_1'\boldsymbol{\beta}_1$ it is necessary that $\mathbf{a}'\mathbf{X}_1\boldsymbol{\beta}_1 = \mathbf{k}_1'\boldsymbol{\beta}_1$, and this will be true for any $\boldsymbol{\beta}_1$ if $\mathbf{a}'\mathbf{X}_1 = \mathbf{k}_1'$. Consequently we minimize (1) subject to this condition. Differentiating (1) with respect to **a** and to an appropriate Lagrange Multiplier, we have equations (2) to solve.

$$\begin{pmatrix} \mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2' & \mathbf{X}_1 \\ \mathbf{X}_1' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{ZGm} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{k}_2 \\ \mathbf{k}_1 \end{pmatrix}. \quad (2)$$

a has a unique solution if and only if $\mathbf{k}_1'\boldsymbol{\beta}_1$ is estimable under a model in which $E(\mathbf{y})$ contains $\mathbf{X}_1\boldsymbol{\beta}_1$. The analogy to GLS of $\boldsymbol{\beta}_1$ is a solution to (3).

$$\mathbf{X}_1'(\mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2')^{-1}\mathbf{X}_1\boldsymbol{\beta}_1^* = \mathbf{X}_1'(\mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2')^{-1}\mathbf{y}. \quad (3)$$

Then if $\mathbf{K}_1'\boldsymbol{\beta}_1$ is estimable under a model, $E(\mathbf{y})$ containing $\mathbf{X}_1\boldsymbol{\beta}_1$, $\mathbf{K}_1'\boldsymbol{\beta}_1^*$ is unique and is the minimum MSE estimator of $\mathbf{K}_1'\boldsymbol{\beta}_1$. The BLBE of $\boldsymbol{\beta}_2$ is

$$\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2'(\mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2')^{-1}(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1^*) \tag{4}$$

$\equiv \boldsymbol{\beta}_2^*$, and this is unique provided $\mathbf{K}_1\boldsymbol{\beta}_1$ is estimable when $E(\mathbf{y})$ contains $\mathbf{X}_1\boldsymbol{\beta}_1$. The BLBP of $\mathbf{u}$ is

$$\mathbf{u}^* = \mathbf{G}\mathbf{Z}'(\mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2')^{-1}(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1^*), \tag{5}$$

and this is unique. Furthermore BLBP of

$$\mathbf{K}_1'\boldsymbol{\beta}_1 + \mathbf{K}_2'\boldsymbol{\beta}_2 + \mathbf{M}'\mathbf{u} \text{ is } \mathbf{K}_1'\boldsymbol{\beta}_1^* + \mathbf{K}_2'\boldsymbol{\beta}_2^* + \mathbf{M}'\mathbf{u}^*. \tag{6}$$

We know that BLUE and BLUP can be computed from mixed model equations. Similarly $\boldsymbol{\beta}_1^*$, $\boldsymbol{\beta}_2^*$, and $\mathbf{u}^*$ can be obtained from modified mixed model equations (7), (8), or (9). Let $\boldsymbol{\beta}_2\boldsymbol{\beta}_2' = \mathbf{P}$. Then with $\mathbf{P}$ singular we can solve (7).

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_2 + \mathbf{I} & \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\beta}_2^* \\ \mathbf{u}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{7}$$

The rank of this coefficient matrix is rank $(\mathbf{X}_1)+p_2+q$, where $p_2 =$ the number of elements in $\boldsymbol{\beta}_2$. The solution to β_2^* and u^* is unique but β_1^* is not unless $\mathbf{X}_1$ has full column rank. Note that the coefficient matrix is non-symmetric. If we prefer a symmetric matrix, we can use equations (8).

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_2\mathbf{P} & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_2\mathbf{P} + \mathbf{P} & \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_2\mathbf{P} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\alpha}_2^* \\ \mathbf{u}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{P}\mathbf{X}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{8}$$

Then $\boldsymbol{\beta}_2^* = \mathbf{P}\boldsymbol{\alpha}_2^*$. The rank of this coefficient matrix is rank $(\mathbf{X}_1)$ + rank $(\mathbf{P}) + \mathbf{q}$. $\mathbf{K}_1'\boldsymbol{\beta}_1^*$, $\boldsymbol{\beta}_2^*$, and $\mathbf{u}^*$ are identical to the solution from (7). If $\mathbf{P}$ were non-singular we could use equations (9).

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_2 + \mathbf{P}^{-1} & \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\beta}_2^* \\ \mathbf{u}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{9}$$

The rank of this coefficient matrix is rank $(\mathbf{X}_1) + p_2 + q$.

Usually $\mathbf{R}$, $\mathbf{G}$, and $\mathbf{P}$ are unknown, so we need to use guesses or estimates of them, say $\tilde{\mathbf{R}}$, $\tilde{\mathbf{G}}$, and $\tilde{\mathbf{P}}$. These would be used in place of the parameter values in (2) through (9).

In all of these except (9) the solution to $\boldsymbol{\beta}_2^*$ has a peculiar and seemingly undesirable property, namely $\boldsymbol{\beta}_2^* = k\tilde{\boldsymbol{\beta}}_2$, where k is some constant. That is, the elements of $\boldsymbol{\beta}_2^*$ are proportional to the elements of $\tilde{\boldsymbol{\beta}}_2$. Also it should be noted that if, as should always be the case, $\mathbf{P}$ is positive definite or positive semi-definite, the elements of $\boldsymbol{\beta}_2^*$ are "shrunken" (are nearer to 0) compared to the elements of the GLS solution to $\boldsymbol{\beta}_2$ when $\mathbf{X}_2$ is full column rank. This is comparable to the fact that BLUP of elements of $\mathbf{u}$ are smaller in absolute value than are the corresponding GLS computed as though $\mathbf{u}$ were fixed. This last property of course creates bias due to $\boldsymbol{\beta}_2$ but may reduce mean squared errors.

2 Use Of An External Estimate Of β

We next consider methods for utilizing an external estimate of $\boldsymbol{\beta}$ in order to obtain a better unbiased estimator from a new data set. For this purpose it will be simplest to assume that in both the previous experiments and the present one the rank of $\mathbf{X}$ is $r \leq p$ and that the same linear dependencies among columns of $\mathbf{X}$ existed in both cases. With possible re-ordering the full rank subset is denoted by $\mathbf{X}_1$ and the corresponding $\boldsymbol{\beta}$ by $\boldsymbol{\beta}_1$. Suppose we have a previous solution to $\boldsymbol{\beta}_1$ denoted by $\boldsymbol{\beta}_1^*$ and $E(\boldsymbol{\beta}_1^*) = \boldsymbol{\beta}_1 + \mathbf{L}\boldsymbol{\beta}_2$ where $\mathbf{X} = (\mathbf{X}_1 \;\; \mathbf{X}_2)$ and $\mathbf{X}_2 = (\mathbf{X}_1\mathbf{L})$. Further $Var(\boldsymbol{\beta}_1^*) = \mathbf{V}_1$. Assuming logically that the prior estimator is uncorrelated with the present data vector, $\mathbf{y}$, the GLS equations are

$$(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1 + \mathbf{V}_1^{-1})\hat{\boldsymbol{\beta}}_1 = \mathbf{X}_1'\mathbf{V}^{-1}\mathbf{y} + \mathbf{V}_1^{-1}\boldsymbol{\beta}_1^*. \tag{10}$$

Then BLUE of $\mathbf{K}'\boldsymbol{\beta}$, where $\mathbf{K}'$ has the form $(\mathbf{K}_1' \;\; \mathbf{K}_1'\mathbf{L})$ is $\mathbf{K}_1'\hat{\boldsymbol{\beta}}_1$, and its variance is

$$\mathbf{K}_1'(\mathbf{X}_1'\mathbf{V}^{-1}\mathbf{X}_1 + \mathbf{V}_1^{-1})^{-1}\mathbf{K}_1. \tag{11}$$

The mixed model equations corresponding to (10) are

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_1 + \mathbf{V}_1^{-1} & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{y} + \mathbf{V}_1^{-1}\boldsymbol{\beta}_1^* \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}. \tag{12}$$

3 Assumed Pattern Of Values Of β

The previous methods of this chapter requiring prior values of every element of $\boldsymbol{\beta}$ and resulting estimates with the same proportionality as the prior is rather distasteful. A possible alternative solution is to assume a pattern of values of $\boldsymbol{\beta}$ with less than p

parameters. For example, with two way, fixed, cross-classified factors with interaction we might assume in some situations that there is no logical pattern of values for interactions. Defining for convenience that the interactions sum to 0 across each row and each column, and then considering all possible permutations of the labelling of rows and columns, the following is true for the average squares and products of these interactions. Define the interaction for the ij^{th} cell as α_{ij} and define the number of rows as r and the number of columns as c. The average values are as follows.

$$\alpha_{ij}^2 = \gamma, \tag{13}$$

$$\alpha_{ij}\alpha_{ij'} = -\gamma/(c-1), \tag{14}$$

$$\alpha_{ij}\alpha_{i'j} = -\gamma/(r-1), \tag{15}$$

$$\alpha_{ij}\alpha_{i'j'} = \gamma/(c-1)(r-1). \tag{16}$$

Then if we have some prior value of γ we can proceed to obtain locally minimum mean squared error estimators and predictors as follows. Let $\mathbf{P}$ = estimated average value of $\boldsymbol{\beta}_2\ \boldsymbol{\beta}_2'$. Then solve equations (7), (8) or (9).

4 Evaluation Of Bias

If we are to consider biased estimation and prediction, we should know how to evaluate the bias. We do this by looking at expectations. A method applied to (7) is as follows. Remember that $\mathbf{K}_1\boldsymbol{\beta}_1^*$ is required to have expectation, $\mathbf{K}_1'\boldsymbol{\beta}_1+$ some linear function of $\boldsymbol{\beta}_2$. For this to be true $\mathbf{K}_1'\boldsymbol{\beta}_1$ must be estimable under a model with $\mathbf{X}_2\boldsymbol{\beta}_2$ not existing. $\boldsymbol{\beta}_2^*$ and $\mathbf{u}^*$ are required to have expectation that is some linear function of $\boldsymbol{\beta}_2$.

Let some g-inverse of the matrix of (7) be

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{C}_{13} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{C}_{23} \\ \mathbf{C}_{31} & \mathbf{C}_{32} & \mathbf{C}_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \end{pmatrix}. \tag{17}$$

Then

$$E(\mathbf{K}_1'\boldsymbol{\beta}_1^*) = \mathbf{K}_1'\boldsymbol{\beta}_1 + \mathbf{K}_1'\mathbf{C}_1\mathbf{T}\boldsymbol{\beta}_2, \tag{18}$$

where

$$\mathbf{T} = \begin{pmatrix} \mathbf{X}_1'\tilde{\mathbf{R}}^{-1}\mathbf{X}_2 \\ \mathbf{P}\mathbf{X}_2'\tilde{\mathbf{R}}^{-1}\mathbf{X}_2 \\ \mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{X}_2 \end{pmatrix}.$$

$$E(\boldsymbol{\beta}_2^*) = \mathbf{C}_2\mathbf{T}\boldsymbol{\beta}_2. \tag{19}$$

$$E(\mathbf{u}^*) = \mathbf{C}_3\mathbf{T}\boldsymbol{\beta}_2. \tag{20}$$

Then the biases are as follows.

$$\begin{aligned}
\text{For } \mathbf{K}_1\boldsymbol{\beta}_1^*, \text{ bias} &= \mathbf{K}_1'\mathbf{C}_1\mathbf{T}\boldsymbol{\beta}_2. &(21)\\
\text{For } \boldsymbol{\beta}_2^*, \text{ bias} &= (\mathbf{C}_2\mathbf{T}-\mathbf{I})\boldsymbol{\beta}_2. &(22)\\
\text{For } \mathbf{u}^*, \text{ bias} &= \mathbf{C}_3\mathbf{T}\boldsymbol{\beta}_2. &(23)
\end{aligned}$$

If the equations (8) are used, the biases are the same as in (21), (22), and (23) except that (22) is premultiplied by $\mathbf{P}$, and $\mathbf{C}$ refers to a g-inverse of the matrix of (8). If the equations of (9) are used, the second term of $\mathbf{T}$ is $\mathbf{X}_2'\tilde{\mathbf{R}}^{-1}\mathbf{X}_2$, and $\mathbf{C}$ refers to the inverse of the matrix of (9).

5 Evaluation Of Mean Squared Errors

If we are to use biased estimation and prediction, we should know how to estimate mean squared errors of estimation and prediction. For the method of (7) proceed as follows. Let

$$\mathbf{T} = \begin{pmatrix} \mathbf{X}_1'\tilde{\mathbf{R}}^{-1}\mathbf{X}_2 \\ \mathbf{P}\mathbf{X}_2'\tilde{\mathbf{R}}^{-1}\mathbf{X}_2 \\ \mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{X}_2 \end{pmatrix}. \quad (24)$$

Note the similarity to the second "column" of the matrix of (7). Let

$$\mathbf{S} = \begin{pmatrix} \mathbf{X}_1'\tilde{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{P}\mathbf{X}_2'\tilde{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\tilde{\mathbf{R}}^{-1}\mathbf{Z} \end{pmatrix}. \quad (25)$$

Note the similarity to the third "column" of the matrix of (7). Let

$$\mathbf{H} = \begin{pmatrix} \mathbf{X}_1'\tilde{\mathbf{R}}^{-1} \\ \mathbf{P}\mathbf{X}_2'\tilde{\mathbf{R}}^{-1} \\ \mathbf{Z}'\tilde{\mathbf{R}}^{-1} \end{pmatrix}. \quad (26)$$

Note the similarity to the right hand side of (7). Then compute

$$\begin{pmatrix} \mathbf{C}_1\mathbf{T} \\ \mathbf{C}_2\mathbf{T}-\mathbf{I} \\ \mathbf{C}_3\mathbf{T} \end{pmatrix} \boldsymbol{\beta}_2\boldsymbol{\beta}_2'(\mathbf{T}'\mathbf{C}_1' \quad \mathbf{T}'\mathbf{C}_2'-\mathbf{I} \quad \mathbf{T}'\mathbf{C}_3')$$
$$+ \begin{pmatrix} \mathbf{C}_1\mathbf{S} \\ \mathbf{C}_2\mathbf{S} \\ \mathbf{C}_3\mathbf{S}-\mathbf{I} \end{pmatrix} \mathbf{G} \quad (\mathbf{S}'\mathbf{C}_1' \quad \mathbf{S}'\mathbf{C}_2' \quad \mathbf{S}'\mathbf{C}_3'-\mathbf{I})$$
$$+ \begin{pmatrix} \mathbf{C}_1\mathbf{H} \\ \mathbf{C}_2\mathbf{H} \\ \mathbf{C}_3\mathbf{H} \end{pmatrix} \mathbf{R} \quad (\mathbf{H}'\mathbf{C}_1' \quad \mathbf{H}'\mathbf{C}_2' \quad \mathbf{H}'\mathbf{C}_3')$$

$$= \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} & \mathbf{B}_{13} \\ \mathbf{B}_{21} & \mathbf{B}_{22} & \mathbf{B}_{23} \\ \mathbf{B}_{31} & \mathbf{B}_{32} & \mathbf{B}_{33} \end{pmatrix} = \mathbf{B}. \tag{27}$$

Then mean squared error of

$$(\mathbf{M}_1' \;\; \mathbf{M}_2' \;\; \mathbf{M}_3') \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\beta}_2^* \\ \mathbf{u}^* - \mathbf{u} \end{pmatrix}$$

$$= (\mathbf{M}_1' \;\; \mathbf{M}_2' \;\; \mathbf{M}_3') \;\; \mathbf{B} \begin{pmatrix} \mathbf{M}_1 \\ \mathbf{M}_2 \\ \mathbf{M}_3 \end{pmatrix}. \tag{28}$$

Of course this cannot be evaluated numerically except for assumed values of $\boldsymbol{\beta}$, $\mathbf{G}$, $\mathbf{R}$. The result simplifies remarkably if we evaluate at the same values used in (7), namely $\boldsymbol{\beta}_2\boldsymbol{\beta}_2' = \tilde{\mathbf{P}}$, $\mathbf{G} = \tilde{\mathbf{G}}$, $\mathbf{R} = \tilde{\mathbf{R}}$. Then $\mathbf{B}$ is simply

$$\mathbf{C} \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12}\mathbf{P} & \mathbf{C}_{13} \\ \mathbf{C}_{21} & \mathbf{C}_{22}\mathbf{P} & \mathbf{C}_{23} \\ \mathbf{C}_{31} & \mathbf{C}_{32}\mathbf{P} & \mathbf{C}_{33} \end{pmatrix}. \tag{29}$$

$\mathbf{C}$ and $\mathbf{C}_{ij}$ are defined in (9.17).

When the method of (8) is used, modify the result for (7) as follows. Let a g-inverse of the matrix of (8) be

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{C}_{13} \\ \mathbf{C}_{21} & \mathbf{C}_{22} & \mathbf{C}_{23} \\ \mathbf{C}_{31} & \mathbf{C}_{32} & \mathbf{C}_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \\ \mathbf{C}_3 \end{pmatrix} = \mathbf{C}. \tag{30}$$

Substitute $\tilde{\mathbf{P}}\mathbf{C}_2\mathbf{T} - \mathbf{I}$ for $\mathbf{C}_2\mathbf{T} - \mathbf{I}$, $\tilde{\mathbf{P}}\mathbf{C}_2\mathbf{S}$ for $\mathbf{C}_2\mathbf{S}$, and $\tilde{\mathbf{P}}\mathbf{C}_2\mathbf{H}$ for $\mathbf{C}_2\mathbf{H}$ and proceed as in (28) using the $\mathbf{C}_i$ from (29). If $\mathbf{P} = \tilde{\mathbf{P}}$, $\mathbf{G} = \tilde{\mathbf{G}}$, $\mathbf{R} = \tilde{\mathbf{R}}$, $\mathbf{B}$ simplifies to

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{C} \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}. \tag{31}$$

If the method of (9) is used, delete $\mathbf{P}$ from $\mathbf{T}$, $\mathbf{S}$, and $\mathbf{H}$ in (24), (25), and (26), let $\mathbf{C}$ be a g-inverse of the matrix of (9), and then proceed as for method (7). When $\mathbf{P} = \tilde{\mathbf{P}}$, $\mathbf{G} = \tilde{\mathbf{G}}$, and $\mathbf{R} = \tilde{\mathbf{R}}$, the simple result, $\mathbf{B} = \mathbf{C}$ can be used.

6 Estimability In Biased Estimation

The traditional understanding of estimability in the linear model is that $\mathbf{K}'\boldsymbol{\beta}$ is defined as estimable if some linear function of $\mathbf{y}$ exists that has expectation $\mathbf{K}'\boldsymbol{\beta}$, and thus this linear function is an unbiased estimator. But if we relax the requirement of unbiasedness, is the above an appropriate definition of estimability? Is any function of $\boldsymbol{\beta}$ now estimable? It seems reasonable to me to restrict estimation to functions that could be estimated if we had no missing subclasses. Otherwise we could estimate elements of $\boldsymbol{\beta}$ that have no relevance to the experiment in question. For example, treatments involve levels of protein in the ration. Just because we invoke biased estimation of treatments would hardly seem to warrant estimation of some treatment that has nothing to do with level of protein. Consequently we state these rules for functions that can be estimated biasedly.

1. We want to estimate $\mathbf{K}_1'\boldsymbol{\beta}_1 + \mathbf{K}_2'\boldsymbol{\beta}_2$, where a prior on $\boldsymbol{\beta}_2$ is used.

2. If $\mathbf{K}_1'\boldsymbol{\beta}_1 + \mathbf{K}_2'\boldsymbol{\beta}_2$ were estimable with no missing subclasses, this function is a candidate for estimation.

3. $\mathbf{K}_1'\boldsymbol{\beta}_1$ must be estimable under a model in which $E(\mathbf{y}) = \mathbf{X}_1\boldsymbol{\beta}_1$.

4. $\mathbf{K}_1'\boldsymbol{\beta}_1 + \mathbf{K}_2'\boldsymbol{\beta}_2$ does not need to be estimable in the sample, but must be estimable in the filled subclass case.

Then $\mathbf{K}_1'\boldsymbol{\beta}_1^o + \mathbf{K}_2'\boldsymbol{\beta}_2^o$ is invariant to the solution to (7),(8), or (9). Let us illustrate with a model

$$y_{ij} = \mu + t_i + e_{ij} \quad , \quad i = 1, 2, 3.$$

Suppose that the numbers of observations per treatment are (5, 3, 0). However, we are willing to assume prior values for squares and products of $t_1,\ t_2,\ t_3$ even though we have no data on t_3. The following functions would be estimable if $n_3 > 0$,

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ t_1 \\ t_2 \\ t_3 \end{pmatrix}.$$

Further with $\boldsymbol{\beta}_1$ being just μ, and $\mathbf{K}_1'$ being 1, and $\mathbf{X}_1' = (1\ 1\ 1)$, $\mathbf{K}_1'\boldsymbol{\beta}_1$ is estimable under a model $E(y_{ij}) = \mu$.

Suppose in contrast that we want to impose a prior on just t_3 . Then $\boldsymbol{\beta}_1' = (\mu\ t_1\ t_2)$ and $\boldsymbol{\beta}_2 = t_3$. Now

$$\mathbf{K}_1'\boldsymbol{\beta}_1' = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \mu \\ t_1 \\ t_2 \end{pmatrix}.$$

But the third row represents a non-estimable function. That is, μ is not estimable under the model with $\beta_1' = (\mu\ t_1\ t_2)$. Consequently $\mu + t_3$ should not be estimated in this way.

As another example suppose we have a 2×3 fixed model with $n_{23} = 0$ and all other $n_{ij} > 0$. We want to estimate all six $\mu_{ij} = \mu + a_i + b_j + \gamma_{ij}$. With no missing subclasses these are estimable, so they are candidates for estimation. Suppose we use priors on $\boldsymbol{\gamma}$. Then

$$(\mathbf{K}_1' \;\; \mathbf{K}_2') \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ a_1 \\ a_2 \\ b_1 \\ b_2 \\ b_3 \\ \boldsymbol{\gamma} \end{pmatrix}.$$

Now $\mathbf{K}_1'\boldsymbol{\beta}_1$ is estimable under a model, $E(y_{ijk}) = \mu + a_i + b_j$. Consequently we can by our rules estimate all six μ_{ij} . These will have expectations as follows.

$$E(\hat{\mu}_{ij}) = \mu + a_i + b_j + \text{ some function of } \gamma \;\neq\; \mu + a_i + b_j + \gamma_{ij}.$$

Now suppose we wish to estimate by using a prior only on γ_{23}. Then the last row of $\mathbf{K}_1'\boldsymbol{\beta}$ is $\mu + a_2 + b_3$ but this is not estimable under a model

$$E \begin{pmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{21} \\ y_{22} \\ y_{23} \end{pmatrix} = \begin{pmatrix} \mu + a_1 + b_1 + \gamma_{11} \\ \mu + a_1 + b_2 + \gamma_{12} \\ \mu + a_1 + b_3 + \gamma_{13} \\ \mu + a_2 + b_1 + \gamma_{21} \\ \mu + a_2 + b_2 + \gamma_{22} \\ \mu + a_2 + b_3 \end{pmatrix}.$$

Consequently we should not use a prior on just γ_{23}.

7 Tests Of Hypotheses

Exact tests of hypotheses do not exist when biased estimation is used, but one might wish to use the following approximate tests that are based on using mean squared error of $\mathbf{K}'\boldsymbol{\beta}^o$ rather than $Var(\mathbf{K}'\boldsymbol{\beta}^o)$.

7.1 $Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$

When $Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$ write (7) as (32) or (8) as (33). Using the notation of Chapter 6, $\mathbf{G} = \mathbf{G}_*\sigma_e^2$ and $\mathbf{P} = \mathbf{P}_*\sigma_e^2$.

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 & \mathbf{X}_1'\mathbf{Z} \\ \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{X}_1 & \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{X}_2 + \mathbf{I} & \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X}_1 & \mathbf{Z}'\mathbf{X}_2 & \mathbf{Z}'\mathbf{Z} + \mathbf{G}_*^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\beta}_2^* \\ \mathbf{u}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix}. \quad (32)$$

The corresponding equations with symmetric coefficient matrix are in (33).

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2\tilde{\mathbf{P}}_* & \mathbf{X}_1'\mathbf{Z} \\ \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{X}_1 & \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{X}_2\tilde{\mathbf{P}}_* + \tilde{\mathbf{P}}_* & \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X}_1 & \mathbf{Z}'\mathbf{X}_2\tilde{\mathbf{P}}_* & \mathbf{Z}'\mathbf{Z} + \mathbf{G}_*^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1^* \\ \boldsymbol{\alpha}^* \\ \mathbf{u}^* \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \tilde{\mathbf{P}}_*\mathbf{X}_2'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{pmatrix} \tag{33}$$

Then $\boldsymbol{\beta}_2^* = \tilde{\mathbf{P}}_*\boldsymbol{\alpha}_2^*$.

Let a g-inverse of the matrix of (32) post-multiplied by

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \equiv \mathbf{Q}$$

or a g-inverse of the matrix (33) pre-multipled and post-multiplied by $\mathbf{Q}$ be denoted by

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix},$$

where $\mathbf{C}_{11}$ has order $p \times p$ and $\mathbf{C}_{22}$ has order $q \times q$. Then if $\tilde{\mathbf{P}}_* = \mathbf{P}_*$, mean squared error of $\mathbf{K}'\boldsymbol{\beta}^*$ is $\mathbf{K}'\mathbf{C}_{11}\mathbf{K}\sigma_e^2$. Then

$$(\mathbf{K}'\boldsymbol{\beta}^* - \mathbf{c})'[\mathbf{K}'\mathbf{C}_{11}\mathbf{K}]^{-1}(\mathbf{K}'\hat{\boldsymbol{\beta}}^* - \mathbf{c})/s\ \hat{\sigma}_e^2$$

is distributed under the null hypothesis approximately as F with $s,\ t$ degrees of freedom, where s = number of rows (linearly independent) in $\mathbf{K}'$, and $\hat{\sigma}_e^2$ is estimated unbiasedly with t degrees of freedom.

7.2 $Var(\mathbf{e}) = \mathbf{R}$

Let g-inverse of (7) post-multiplied by

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \equiv \mathbf{Q}$$

or a g-inverse of (8) pre-multiplied and post-multiplied by $\mathbf{Q}$ be denoted by

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}.$$

Then if $\tilde{\mathbf{R}} = \mathbf{R}, \tilde{\mathbf{G}} = \mathbf{G}$, and $\tilde{\mathbf{P}} = \mathbf{P}$, $\mathbf{K}'\mathbf{C}_{11}\mathbf{K}$ is the mean squared error of $\mathbf{K}'\boldsymbol{\beta}^*$, and $(\mathbf{K}'\boldsymbol{\beta}^* - \mathbf{c})'(\mathbf{K}'\mathbf{C}_{11}\mathbf{K})^{-1}(\mathbf{K}'\boldsymbol{\beta}^* - \mathbf{c})$ is distributed approximately as χ^2 with s degrees of freedom under the null hypothesis, $\mathbf{K}'\boldsymbol{\beta} = \mathbf{c}$.

8 Estimation of P

If one is to use biased estimation and prediction, one would usually have to estimate $\mathbf{P}$, ordinarily a singular matrix. If the elements of $\boldsymbol{\beta}_2$ are thought to have no particular pattern, permutation theory might be used to derive average values of squares and products of elements of $\boldsymbol{\beta}_2$, that is the value of $\mathbf{P}$. We might then formulate this as estimation of a variance covariance matrix, usually with fewer parameters than $t(t+1)/2$, where t is the order of $\mathbf{P}$. I think I would estimate these parameters by the MIVQUE method for singular $\mathbf{G}$ described in Section 9 of Chapter 11 or by REML of Chapter 12.

9 Illustration

We illustrate biased estimation by a 3-way mixed model. The model is

$$y_{hijk} = r_h + c_i + \gamma_{hi} + u_j + e_{ijk},$$
$$r, c, \gamma \text{ are fixed, } Var(\mathbf{u}) = \mathbf{I}/10, \; Var(\mathbf{e}) = 2\mathbf{I}.$$

The data are as follows:

	Levels of j			
hi subclasses	1	2	3	$y_{hi..}$
11	2	1	0	18
12	0	1	1	13
13	1	0	0	7
21	1	2	1	26
22	0	0	1	9
$y_{..j.}$	25	27	21	

We want to estimate using prior values of the squares and products of γ_{hi}. Suppose this is as follows, ordering i within h, and including γ_{23}.

$$\begin{pmatrix} .1 & -.05 & -.05 & -.1 & .05 & .05 \\ & .1 & -.05 & .05 & -.1 & .05 \\ & & .1 & .05 & .05 & -.1 \\ & & & .1 & -.05 & -.05 \\ & & & & .1 & -.05 \\ & & & & & .1 \end{pmatrix}.$$

The equations of the form

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

are presented in (34).

$$\frac{1}{2}\begin{pmatrix} 6 & 0 & 3 & 2 & 1 & 3 & 2 & 1 & 0 & 0 & 0 & 3 & 2 & 1 \\ & 5 & 4 & 1 & 0 & 0 & 0 & 0 & 4 & 1 & 0 & 1 & 2 & 2 \\ & & 7 & 0 & 0 & 3 & 0 & 0 & 4 & 0 & 0 & 3 & 3 & 1 \\ & & & 3 & 0 & 0 & 2 & 0 & 0 & 1 & 0 & 0 & 1 & 2 \\ & & & & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ & & & & & 3 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & 0 \\ & & & & & & 2 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ & & & & & & & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ & & & & & & & & 4 & 0 & 0 & 1 & 2 & 1 \\ & & & & & & & & & 1 & 0 & 0 & 0 & 1 \\ & & & & & & & & & & 0 & 0 & 0 & 0 \\ & & & & & & & & & & & 4 & 0 & 0 \\ & & & & & & & & & & & & 4 & 0 \\ & & & & & & & & & & & & & 3 \end{pmatrix} \begin{pmatrix} \mathbf{r} \\ \mathbf{c} \\ \boldsymbol{\gamma} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} 38 \\ 35 \\ 44 \\ 22 \\ 7 \\ 18 \\ 13 \\ 7 \\ 26 \\ 9 \\ 0 \\ 25 \\ 27 \\ 21 \end{pmatrix} \frac{1}{2} \qquad (34)$$

Note that γ_{23} is included even though no observation on it exists.

Pre-multiplying these equations by

$$\begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \equiv \mathbf{T}$$

and adding $\mathbf{I}$ to the diagonals of equations (6)-(11) and $10\mathbf{I}$ to the diagonals of equations (12)-(14) we obtain the coefficient matrix to solve for the biased estimators and predictors. The right hand side vector is

$$(19, 17.5, 22, 11, 3.5, -.675, .225, .45, .675, -.225, -.45, 12.5, 13.5, 10.5)'.$$

This gives a solution of

$$\begin{aligned} \mathbf{r}^* &= (3.6899, 4.8607), \\ \mathbf{c}^* &= (1.9328, 3.3010, 3.3168), \\ \boldsymbol{\gamma}^* &= (.11406, -.11406, 0, -.11406, .11406, 0), \\ \mathbf{u}^* &= (-.00664, .04282, -.03618). \end{aligned}$$

Note that

$$\begin{aligned} \textstyle\sum_i \gamma^*_{ij} &= 0 \text{ for i } = 1,\ 2,\ \text{and} \\ \textstyle\sum_j \gamma^*_{ij} &= 0 \text{ for j } = 1,\ 2,\ 3. \end{aligned}$$

These are the same relationships that were defined for γ.

Post-multiplying the g-inverse of the coefficient matrix by $\mathbf{T}$ we get (35) ... (38) and the matrix for computing mean squared errors for $\mathbf{M}'(\mathbf{r}^*,\ \mathbf{c}^*,\ \boldsymbol{\gamma}^*,\ \mathbf{u}^*)$. The lower 9×9 submatrix is symmetric and invariant reflecting the fact that $\boldsymbol{\gamma}^*$, and $\mathbf{u}^*$ are invariant to the g-inverse taken.

Upper left 7×7

$$\begin{pmatrix}
.26181 & -.10042 & .02331 & 0 & .15599 & -.02368 & .00368 \\
-.05313 & .58747 & -.22911 & 0 & .54493 & .07756 & .00244 \\
-.05783 & -.26296 & .41930 & 0 & -.35232 & -.02259 & -.00741 \\
.56640 & .61368 & -.64753 & 0 & -1.02228 & .00080 & -.03080 \\
-.29989 & .13633 & .02243 & 0 & 2.07553 & .07567 & .04433 \\
-.02288 & .07836 & -.02339 & 0 & .07488 & .08341 & -.03341 \\
-.02712 & -.02836 & .02339 & 0 & .07512 & -.03341 & .08341
\end{pmatrix} \tag{35}$$

Upper right 7×7

$$\begin{pmatrix}
.02 & .02368 & -.00368 & -.02 & -.03780 & -.01750 & -.00469 \\
-.08 & -.07756 & -.00244 & .08 & -.01180 & -.01276 & -.03544 \\
.03 & .02259 & .00741 & -.03 & -.01986 & -.02631 & .00617 \\
.03 & -.00080 & .03080 & -.03 & .02588 & -.01608 & -.04980 \\
-.12 & -.07567 & -.04433 & .12 & -.05563 & .01213 & .00350 \\
-.05 & -.08341 & .03341 & .05 & -.00199 & .00317 & -.00118 \\
-.05 & .03341 & -.08341 & .05 & .00199 & -.00317 & .00118
\end{pmatrix} \tag{36}$$

Lower left 7×7

$$\begin{pmatrix}
.05 & -.05 & 0 & 0 & -.15 & -.05 & -.05 \\
.02288 & -.07836 & .02339 & 0 & -.07488 & -.08341 & .03341 \\
.02712 & .02836 & -.02339 & 0 & -.07512 & .03341 & -.08341 \\
-.05 & .05 & 0 & 0 & .15 & .05 & .05 \\
-.01192 & .01408 & -.04574 & 0 & -.08l51 & -.00199 & .00199 \\
-.03359 & -.02884 & -.01023 & 0 & .02821 & .00317 & -.00317 \\
-.05450 & -.08524 & .05597 & 0 & .05330 & -.00118 & .00118
\end{pmatrix} \tag{37}$$

Lower right 7×7

$$\begin{pmatrix}
.10 & .05 & .05 & -.10 & 0 & 0 & 0 \\
 & .08341 & -.03341 & -.05 & .00199 & -.00317 & .00118 \\
 & & .08341 & -.05 & -.00199 & .00317 & -.00118 \\
 & & & .10 & 0 & 0 & 0 \\
 & & & & .09343 & .00537 & .00120 \\
 & & & & & .09008 & .00455 \\
 & & & & & & .09425
\end{pmatrix} \tag{38}$$

A g-inverse of the coefficient matrix of equations like (8) is in (39) ... (41).

This gives a solution (−1.17081, 0, 6.79345, 8.16174, 8.17745, 0, 0, .76038, −1.52076, 0, 0, −.00664, .04282, −.03618). Premultiplying this solution by $\mathbf{T}$ we obtain for $\boldsymbol{\beta}_1^*$, (-1.17081, 0, 6.79345, 8.16174, 8.17745), and the same solution as before for $\boldsymbol{\beta}_2^*$ and $\mathbf{u}^*$; $\boldsymbol{\beta}_1$ is not estimable so $\boldsymbol{\beta}_1^*$ is not invariant and differs from the previous solution. But estimable functions of $\boldsymbol{\beta}_1$ are the same.

Pre and post-multiplying (39) ... (41) by $\mathbf{T}$ gives the matrix (42) ... (43). The lower 9×9 submatrix is the same as that of (38) associated with the fact that $\boldsymbol{\beta}_2^*$ and $\mathbf{u}^*$ are unique to whatever g-inverse is obtained.

Upper left 7×7

$$\begin{pmatrix} 1.00283 & 0 & -.43546 & -.68788 & -1.07683 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 & 0 \\ & & .51469 & .32450 & .51712 & 0 & 0 \\ & & & 1.20115 & .72380 & 0 & 0 \\ & & & & 3.34426 & 0 & 0 \\ & & & & & 0 & 0 \\ & & & & & & 0 \end{pmatrix} \tag{39}$$

Upper right 7×7 and (lower left 7×7)'

$$\begin{pmatrix} .65838 & .68324 & 0 & 0 & -.026 & -.00474 & .03075 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.30020 & -.39960 & 0 & 0 & -.03166 & -.03907 & -.02927 \\ -.14427 & -.71147 & 0 & 0 & .01408 & -.02884 & -.08524 \\ -1.64509 & -.70981 & 0 & 0 & -.06743 & -.00063 & -.03l94 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{40}$$

Lower right 7×7

$$\begin{pmatrix} 12.59603 & -5.19206 & 0 & 0 & -.01329 & .02112 & -.00784 \\ & 10.38413 & 0 & 0 & .02657 & -.04224 & .01567 \\ & & 0 & 0 & 0 & 0 & 0 \\ & & & 0 & 0 & 0 & 0 \\ & & & & .09343 & .00537 & .00120 \\ & & & & & .09008 & .00455 \\ & & & & & & .09425 \end{pmatrix} \tag{41}$$

Upper left 7×7

$$\begin{pmatrix} 1.00283 & 0 & -.43546 & -.68788 & -1.07683 & -.10124 & .00124 \\ & 0 & 0 & 0 & 0 & 0 & 0 \\ & & .51469 & .32450 & .51712 & .05497 & -.00497 \\ & & & 1.20115 & .72380 & .07836 & -.02836 \\ & & & & 3.34426 & .15324 & .04676 \\ & & & & & .08341 & -.03341 \\ & & & & & & .08341 \end{pmatrix} \tag{42}$$

Upper right 7×7 and (lower left 7×7)'

$$\begin{pmatrix} .1 & .10124 & -.00124 & -.1 & -.026 & -.00474 & .03075 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -.05 & -.05497 & .00497 & .05 & -.03166 & -.03907 & -.02927 \\ -.05 & -.07836 & .02836 & .05 & .01408 & -.02884 & -.08524 \\ -.2 & -.15324 & -.04676 & .2 & -.06743 & -.00063 & -.03194 \\ -.05 & -.08341 & .03341 & .05 & -.00199 & .00317 & -.00118 \\ -.05 & .03341 & -.08341 & .05 & .00199 & -.00317 & .00118 \end{pmatrix} \tag{43}$$

Lower right 7×7 is the same as in (38).

Suppose we wish to estimate $\mathbf{K}'(\boldsymbol{\beta}_1' \;\; \boldsymbol{\beta}_2')'$, which is estimable when the $r \times c$ subclasses are all filled, and

$$\mathbf{K}' = \begin{pmatrix} 6 & 0 & 2 & 2 & 2 & 2 & 2 & 2 & 0 & 0 & 0 \\ 0 & 6 & 2 & 2 & 2 & 0 & 0 & 0 & 2 & 2 & 2 \\ 3 & 3 & 6 & 0 & 0 & 3 & 0 & 0 & 3 & 0 & 0 \\ 3 & 3 & 0 & 6 & 0 & 0 & 3 & 0 & 0 & 3 & 0 \\ 3 & 3 & 0 & 0 & 6 & 0 & 0 & 3 & 0 & 0 & 3 \end{pmatrix} /6.$$

Pre-multiplying the upper 11x11 submatrix of either (35) to (38) or (42) to (43) by $\mathbf{K}'$ gives identical results shown in (44).

$$\begin{pmatrix} .44615 & .17671 & .15136 & .19665 & .58628 \\ & .91010 & .08541 & .38312 & 1.16170 \\ & & .32993 & .01354 & .01168 \\ & & & .76397 & .09215 \\ & & & & 2.51814 \end{pmatrix} \tag{44}$$

This represents the estimated mean squared error matrix of these 5 functions of $\boldsymbol{\beta}$.

Next we illustrate with another set of data the relationships of (3), (4), and (5) to (7). We have a design with 3 treatments and 2 random sires. The subclass numbers are

	Sires	
Treatments	1	2
1	2	1
2	1	2
3	2	0

The model is

$$y_{ijk} = \mu + t_i + s_j + x_{ijk}\beta + e_{ijk}.$$

where β is a regression and x_{ijk} the associated covariate.

$$\begin{aligned} \mathbf{y}' &= (5\ 3\ 6\ 4\ 7\ 5\ 4\ 8), \\ \text{Covariates} &= (1\ 2\ 1\ 3\ 2\ 4\ 2\ 3). \end{aligned}$$

The data are ordered sires in treatments. We shall use a prior on treatments of

$$\begin{pmatrix} 2 & -1 & -1 \\ & 2 & -1 \\ & & 2 \end{pmatrix}.$$

$$Var(\mathbf{e}) = 5\mathbf{I}, \text{ and } Var(\mathbf{s}) = \mathbf{I}.$$

We first illustrate the equations of (8),

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_1 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}_2 & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{pmatrix} =$$

$$\begin{pmatrix} 1.6 & 3.6 & .6 & .6 & .4 & 1.0 & .6 \\ & 9.6 & .8 & 1.8 & 1.0 & 2.2 & 1.4 \\ & & .6 & 0 & 0 & .4 & .2 \\ & & & .6 & 0 & .2 & .4 \\ & & & & .4 & .4 & 0 \\ & & & & & 1.0 & 0 \\ & & & & & & .6 \end{pmatrix}. \quad (45)$$

and

$$\begin{pmatrix} \mathbf{X}_1'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{X}_2'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} = (8.4\ \ 19.0\ \ 2.8\ \ 3.2\ \ 2.4\ \ 4.8\ \ 3.6)'. \quad (46)$$

These are ordered, μ, $\boldsymbol{\beta}$, $\mathbf{t}$, $\mathbf{s}$. Premultiplying (45) and (46) by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 & 0 \\ & & 2 & -1 & -1 & 0 & 0 \\ & & & 2 & -1 & 0 & 0 \\ & & & & 2 & 0 & 0 \\ & & & & & 1 & 0 \\ & & & & & & 1 \end{pmatrix}$$

we get

$$\begin{pmatrix} 1.6 & 3.6 & .6 & .6 & .4 & 1.0 & .6 \\ 3.6 & 9.6 & .8 & 1.8 & 1.0 & 2.2 & 1.4 \\ .2 & -1.2 & 1.2 & -.6 & -.4 & .2 & 0 \\ .2 & 1.8 & -.6 & 1.2 & -.4 & -.4 & .6 \\ -.4 & -.6 & -.6 & -.6 & .8 & .2 & -.6 \\ 1.0 & 2.2 & .4 & .2 & .4 & 1.0 & 0 \\ .6 & 1.4 & .2 & .4 & 0 & 0 & .6 \end{pmatrix}, \tag{47}$$

and

$$(8.4 \;\; 19.0 \;\; 0 \;\; 1.2 \;\; -1.2 \;\; 4.8 \;\; 3.6)'. \tag{48}$$

The vector (48) is the right hand side of equations like (8). Then the coefficient matrix is matrix (47) + dg(0 0 1 1 1 1 1). The solution is

$$\begin{aligned} \mu^* &= 5.75832, \\ \beta^* &= -.16357, \\ (\mathbf{t}^*)' &= (-.49697 \;\; -.02234 \;\; .51931), \\ (\mathbf{s}^*)' &= (-.30146 \;\; .30146). \end{aligned}$$

Now we set up equations (3).

$$\mathbf{V} = (\mathbf{ZGZ}' + \mathbf{R}) = \begin{pmatrix} 6 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ & 6 & 0 & 1 & 0 & 0 & 1 & 1 \\ & & 6 & 0 & 1 & 1 & 0 & 0 \\ & & & 6 & 0 & 0 & 1 & 1 \\ & & & & 6 & 1 & 0 & 0 \\ & & & & & 6 & 0 & 0 \\ & & & & & & 6 & 1 \\ & & & & & & & 6 \end{pmatrix}. \tag{49}$$

$$\mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2' = \begin{pmatrix} 2 & 2 & 2 & -1 & -1 & -1 & -1 & -1 \\ & 2 & 2 & -1 & -1 & -1 & -1 & -1 \\ & & 2 & -1 & -1 & -1 & -1 & -1 \\ & & & 2 & 2 & 2 & -1 & -1 \\ & & & & 2 & 2 & -1 & -1 \\ & & & & & 2 & -1 & -1 \\ & & & & & & 2 & 2 \\ & & & & & & & 2 \end{pmatrix}. \tag{50}$$

$$(\mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2')^{-1} =$$

$$\begin{pmatrix} .1525 & -.0475 & -.0275 & -.0090 & .0111 & .0111 & -.0005 & -.0005 \\ & .1525 & -.0275 & -.0090 & .0111 & .0111 & -.0005 & -.0005 \\ & & .1444 & .0214 & -.0067 & -.0067 & .0119 & .0119 \\ & & & .1417 & -.0280 & -.0280 & -.0031 & -.0031 \\ & & & & .1542 & -.0458 & .0093 & .0093 \\ & & & & & .1542 & .0093 & .0093 \\ & & & & & & .1482 & -.0518 \\ & & & & & & & .1482 \end{pmatrix}. \quad (51)$$

The equations like (4) are

$$\begin{pmatrix} .857878 & 1.939435 \\ 1.939435 & 5.328690 \end{pmatrix} \begin{pmatrix} \boldsymbol{\mu}^* \\ \boldsymbol{\beta}^* \end{pmatrix} = \begin{pmatrix} 4.622698 \\ 10.296260 \end{pmatrix}. \quad (52)$$

The solution is (5.75832 - .163572) as in the mixed model equations.

$$(\mathbf{y} - \mathbf{X}_1\boldsymbol{\beta}_1^*) = \mathbf{y} - \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \\ 1 & 3 \\ 1 & 2 \\ 1 & 4 \\ 1 & 2 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} 5.75832 \\ -.163572 \end{pmatrix} = \begin{pmatrix} -.59474 \\ -2.43117 \\ .40526 \\ -1.26760 \\ 1.56883 \\ -.10403 \\ -1.43117 \\ 2.73240 \end{pmatrix}.$$

$$\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2'(\mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2')^{-1} =$$

$$\begin{pmatrix} .1426 & .1426 & .1471 & -.0725 & -.0681 & -.0681 & -.0899 & -.0899 \\ -.0501 & -.0501 & -.0973 & .1742 & .1270 & .1270 & -.0766 & -.0766 \\ -.0925 & -.0925 & -.0497 & -.1017 & -.0589 & -.0589 & .1665 & .1665 \end{pmatrix}.$$

Then $\mathbf{t}^*$ = (-.49697 -.02234 .51931)′ as before.

$$\mathbf{GZ}'(\mathbf{V} + \mathbf{X}_2\boldsymbol{\beta}_2\boldsymbol{\beta}_2'\mathbf{X}_2')^{-1} =$$

$$\begin{pmatrix} .0949 & .0949 & -.0097 & .1174 & .0127 & .0127 & .0923 & .0923 \\ -.0053 & -.0053 & .1309 & -.0345 & .1017 & .1017 & .0304 & .0304 \end{pmatrix}.$$

Then $\mathbf{u}^*$ = (-.30146 .30146)′ as before.

Sections 9 and 10 of Chapter 15 give details concerning use of a diagonal matrix in place of $\mathbf{P}$.

10 Relationships Among Methods

BLUP, Bayesian estimation, and minimum mean squared error estimation are quite similar, and in fact are identical under certain assumptions.

10.1 Bayesian estimation

Let $(\mathbf{X} \;\; \mathbf{Z}) = \mathbf{W}$ and $(\boldsymbol{\beta}' \;\; \mathbf{u}') = \boldsymbol{\gamma}'$. Then the linear model is

$$\mathbf{y} = \mathbf{W}\boldsymbol{\gamma} + \mathbf{e}.$$

Let $\mathbf{e}$ have multivariate normal distribution with null means and $Var(\mathbf{e}) = \mathbf{R}$. Let the prior distribution of $\boldsymbol{\gamma}$ be multivariate normal with $E(\boldsymbol{\gamma}) = \boldsymbol{\mu}$, $Var(\boldsymbol{\gamma}) = \mathbf{C}$, and $Cov(\boldsymbol{\gamma}, \mathbf{e}') = \mathbf{0}$. Then for any of the common loss functions, that is, squared loss function, absolute loss function, or uniform loss function the Bayesian estimator of γ is the solution to (53).

$$(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{C}^{-1})\; \hat{\gamma} = \mathbf{W}'\mathbf{R}^{-1}\mathbf{y} + \mathbf{C}^{-1}\boldsymbol{\mu}. \tag{53}$$

Note that $\hat{\boldsymbol{\gamma}}$ is an unbiased estimator of $\boldsymbol{\gamma}$ if estimable and $E(\boldsymbol{\gamma}) = \boldsymbol{\mu}$. See Lindley and Smith (1972) for a discussion of Bayesian estimation for linear models. Equation (53) can be derived by maximizing $f(\mathbf{y}, \boldsymbol{\gamma})$ for variations in $\boldsymbol{\gamma}$. This might be called a MAP (maximum a posteriori) estimator, Melsa and Cohn (1978).

Now suppose that

$$\mathbf{C}^{-1} \rightarrow \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}$$

and prior on $\boldsymbol{\mu} = \mathbf{0}$. Then (53) becomes the mixed model equations for BLUE and BLUP.

10.2 Minimum mean squared error estimation

Using the same notation as in Section 10.1, the minimum mean squared error estimator is

$$(\mathbf{W}'\mathbf{R}^{-1}\mathbf{W} + \mathbf{Q}^{-1})\gamma^o = \mathbf{W}'\mathbf{R}^{-1}\mathbf{y}, \tag{54}$$

where $\mathbf{Q} = \mathbf{C} + \boldsymbol{\mu}\boldsymbol{\mu}'$. Note that if $\boldsymbol{\mu} = \mathbf{0}$ this and the Bayesian estimator are identical. The essential difference is that the Bayesian estimator uses prior $E(\boldsymbol{\beta})$, whereas minimum MSE uses only squares and products of $\boldsymbol{\beta}$.

To convert (54) to the situation with prior on $\boldsymbol{\beta}_2$ but not on $\boldsymbol{\beta}_1$, let

$$\mathbf{Q}^- = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}.$$

The upper left partition is square with order equal to the number of elements in $\boldsymbol{\beta}_1$.

To convert (54) to the BLUP, mixed model equations let

$$\mathbf{Q}^- = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix},$$

where the upper left submatix is square with order p, the number of elements in $\boldsymbol{\beta}$. In the above results $\mathbf{P}$ may be singular. In that case use the technique described in previous sections for singular $\mathbf{G}$ and $\mathbf{P}$.

10.3 Invariance property of Bayesian estimator

Under normality and with absolute deviation as the loss function, the Bayesian estimator of f($\boldsymbol{\beta}$, $\mathbf{u}$) is $f(\boldsymbol{\beta}^o, \hat{\mathbf{u}})$, where $(\boldsymbol{\beta}^o, \hat{\mathbf{u}})$ is the Bayesian solution (also the BLUP solution when the priors are on $\mathbf{u}$ only), and f is any function. This was noted by Gianola (1982) who made use of a result reported by DeGroot (1981). Thus under normality any function of the BLUP solution is the Bayesian estimator of that function when the loss function is absolute deviation.

10.4 Maximum likelihood estimation

If the prior distribution on the parameters to be estimated is the uniform distribution and the mode of the posterior distribution is to be maximized, the resulting estimator is ML. When $\mathbf{Zu} + \mathbf{e} = \boldsymbol{\epsilon}$ has the multivariate normal distribution the MLE of $\boldsymbol{\beta}$, assumed estimable, is the maximizing value of $k \exp[-.5\ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]$. The maximizing value of this is the solution to

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y},$$

the GLS equations. Now we know that the conditional mean of $\mathbf{u}$ given $\mathbf{y}$ is

$$\mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Under fairly general conditions the ML estimator of a function of parameters is that same function of the ML estimators of those same parameters. Thus ML of the conditional mean of $\mathbf{u}$ under normality is

$$\mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^o),$$

which we recognize as BLUP of $\mathbf{u}$ for any distribution.

11 Pattern Of Values Of P

When $\mathbf{P}$ has the structure described above and consequently is singular, a simpler method can be used. A diagonal, non-singular $\mathbf{P}$ can be written, which when used in mixed model equations results in the same estimates and predictions of estimable and predictable functions. See Chapter 15.