Chapter 13
Effects of Selection

C. R. Henderson

1984 - Guelph

1 Introduction

The models and the estimation and prediction methods of the preceding chapters have not addressed the problem of data arising from a selection program. Note that the assumption has been that the expected value of every element of $\mathbf{u}$ is 0. What if $\mathbf{u}$ represents breeding values of animals that have been produced by a long-time, effective, selection program? In that case we would expect the breeding values in later generations to be higher than in the earlier ones. Consequently the expected value of $\mathbf{u}$ is not really $\mathbf{0}$ as assumed in the methods presented earlier. Also it should be noted that, in an additive genetic model, $\mathbf{A}\sigma_a^2$ is a correct statement of the covariance matrix of breeding values if no selection has taken place and $\sigma_a^2 =$ additive genetic variance in an unrelated, non-inbred, unselected population. Following selection this no longer is true. Generally variances are reduced and the covariances are altered. In fact, there can be non-zero covariances for pairs of unrelated animals. Further, we often assume for one trait that $Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$. Following selection this is no longer true. Variances are reduced and non-zero covariances are generated. Another potentially serious consequence of selection is that previously uncorrelated elements of $\mathbf{u}$ and $\mathbf{e}$ become correlated with selection. If we know the new first and second moments of $(\mathbf{y}, \mathbf{u})$ we can then derive BLUE and BLUP for that model. This is exceedingly difficult for two reasons. First, because selection intensity varies from one herd to another, a different set of parameters would be needed for each herd, but usually with too few records for good estimates to be obtained. Second, correlation of $\mathbf{u}$ with $\mathbf{e}$ complicates the computations. Fortunately, as we shall see later in this chapter, computations that ignore selection and then use the parameters existing prior to selection sometimes result in BLUE and BLUP under the selection model. Unfortunately, comparable results have not been obtained for variance and covariance estimation, although there does seem to be some evidence that MIVQUE with good priors, REML, and ML may have considerable ability to control bias due to selection, Rothschild et al. (1979).

2 An Example of Selection

We illustrate some effects of selection and the properties of BLUE, BLUP, and OLS by a progeny test example. The progeny numbers were distributed as follows

	Treatments	
Sires	1	2
1	10	500
2	10	100
3	10	0
4	10	0

We assume that the sires were ranked from highest to lowest on their progeny averages in Period 1. If that were true in repeated sampling and if we assume normal distributions, one can write the expected first and second moments. Assume unrelated sires, $\sigma_e^2 = 15$, $\sigma_s^2 = 1$ under a model,

$$y_{ijk} \;=\; s_i \;+\; p_j \;+\; e_{ijk}.$$

With no selection

$$E\begin{pmatrix} \bar{y}_{11} \\ \bar{y}_{21} \\ \bar{y}_{31} \\ \bar{y}_{41} \\ \bar{y}_{12} \\ \bar{y}_{22} \end{pmatrix} = \begin{pmatrix} p_1 \\ p_1 \\ p_1 \\ p_1 \\ p_2 \\ p_2 \end{pmatrix}, \; Var = \begin{pmatrix} 2.5 & 0 & 0 & 0 & 1 & 0 \\ & 2.5 & 0 & 0 & 0 & 1 \\ & & 2.5 & 0 & 0 & 0 \\ & & & 2.5 & 0 & 0 \\ & & & & 1.03 & 0 \\ & & & & & 1.15 \end{pmatrix}.$$

With ordering of sires according to first records the corresponding moments are

$$\begin{pmatrix} 1.628 & + & p_1 \\ .460 & + & p_1 \\ -.460 & + & p_1 \\ -1.628 & + & p_1 \\ .651 & + & p_2 \\ .184 & + & p_2 \end{pmatrix} \text{ and } \begin{pmatrix} 1.229 & .614 & .395 & .262 & .492 & .246 \\ & .901 & .590 & .395 & .246 & .360 \\ & & .901 & .614 & .158 & .236 \\ & & & 1.229 & .105 & .158 \\ & & & & .827 & .098 \\ & & & & & .894 \end{pmatrix}.$$

Further, with no ordering $E(\mathbf{s}) \;=\; \mathbf{0}, Var(\mathbf{s}) \;=\; \mathbf{I}$. With ordering these become

$$\begin{pmatrix} .651 \\ .184 \\ -.184 \\ -.651 \end{pmatrix} \text{ and } \begin{pmatrix} .797 & .098 & .063 & .042 \\ & .744 & .094 & .063 \\ & & .744 & .098 \\ & & & .797 \end{pmatrix}.$$

These results are derived from Teicheroew (1956), Sarhan and Greenberg (1956), and Pearson (1903).

Suppose $p_1 = 10$, $p_2 = 12$. Then in repeated sampling the expected values of the 6 subclass means would be

$$\begin{pmatrix} 11.628 & 12.651 \\ 10.460 & 12.184 \\ 9.540 & -- \\ 8.372 & -- \end{pmatrix}$$

Applying BLUE and BLUP, ignoring selection, to these expected data the mixed model equations are

$$\begin{pmatrix} 40 & 0 & 10 & 10 & 10 & 10 \\ & 600 & 500 & 100 & 0 & 0 \\ & & 525 & 0 & 0 & 0 \\ & & & 125 & 0 & 0 \\ & & & & 25 & 0 \\ & & & & & 25 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{pmatrix} = \begin{pmatrix} 400.00 \\ 7543.90 \\ 6441.78 \\ 1323.00 \\ 95.40 \\ 83.72 \end{pmatrix}$$

The solution is $[10.000, \quad 12.000, \quad .651, \quad .184, \quad -.184, \quad -.651]$, thereby demonstrating unbiasedness of $\hat{\mathbf{p}}$ and $\hat{\mathbf{s}}$. The reason for this is discussed in Section 13.5.1.

In contrast the OLS solution gives biased estimators and predictors. Forcing $\sum \hat{s}_i = 0$ as in the BLUP solution we obtain as the solution

$$[10.000, \quad 11.361, \quad 1.297, \quad .790, \quad -.460, -1.628].$$

Except for $\hat{p}_1$ these are biased. If OLS is applied to only the data of period 2, $s_1^o - s_2^o$ is an unbiased predictor of $s_1 - s_2$. The equations in this case are

$$\begin{pmatrix} 600 & 500 & 100 \\ 500 & 500 & 0 \\ 100 & 0 & 100 \end{pmatrix} \begin{pmatrix} p_2^o \\ s_1^o \\ s_2^o \end{pmatrix} = \begin{pmatrix} 7543.90 \\ 6325.50 \\ 1218.40 \end{pmatrix}.$$

A solution is $[0, \quad 12.651, \quad 12.184]$. Then $s_1^o - s_2^o = .467 = E(s_1 - s_2)$ under the selection model. This result is equivalent to a situation in which the observations on the first period are not observable and we define selection at that stage as selection on $\mathbf{u}$, in which case treating $\mathbf{u}$ as fixed in the computations leads to unbiased estimators and predictors. Note, however, that we obtain invariant solutions only for functions that are estimable under a fixed $\mathbf{u}$ model. Consequently p_2 is not estimable and we can predict only the difference between s_1 and s_2.

3 Conditional Means And Variances

Pearson (1903) derived results for the multivariate normal distribution that are extremely useful for studying the selection problem. These are the results that were used in the example in Section 13.2. We shall employ the notation of Henderson (1975a), similar to that of Lawley (1943), rather than Pearson's, which was not in matrix notation. With no selection $[\mathbf{v}_1' \ \ \mathbf{v}_2']$ have a multivariate normal distribution with means,

$$\begin{pmatrix} \mu_1' & \mu_2' \end{pmatrix}, \quad \text{and} \ Var \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}' & \mathbf{C}_{22} \end{pmatrix}. \tag{1}$$

Suppose now in conceptual repeated sampling $\mathbf{v}_2$ is selected in such a way that it has mean $= \mu_2 + \mathbf{k}$ and variance $= \mathbf{C}_s$. Then Pearson's result is

$$E_s \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} \mu_1 + \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{k} \\ \mu_2 + \mathbf{k} \end{pmatrix}. \tag{2}$$

$$Var_s \begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{11} - \mathbf{C}_{12}\mathbf{C}_0\mathbf{C}_{12}' & \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_s \\ \mathbf{C}_s\mathbf{C}_{22}^{-1}\mathbf{C}_{12}' & \mathbf{C}_s \end{pmatrix}, \tag{3}$$

where $\mathbf{C}_0 = \mathbf{C}_{22}^{-1}(\mathbf{C}_{22} - \mathbf{C}_s)\mathbf{C}_{22}^{-1}$. Henderson (1975) used this result to derive BLUP and BLUE under a selection model with a multivariate normal distribution of $(\mathbf{y}, \mathbf{u}, \mathbf{e})$ assumed. Let $\mathbf{w}$ be some vector correlated with $(\mathbf{y}, \mathbf{u})$. With no selection

$$E \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{d} \end{pmatrix}, \tag{4}$$

$$Var \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{V} & \mathbf{ZG} & \mathbf{R} & \mathbf{B} \\ \mathbf{GZ}' & \mathbf{G} & \mathbf{0} & \mathbf{B}_u \\ \mathbf{R} & \mathbf{0} & \mathbf{R} & \mathbf{B}_e \\ \mathbf{B}' & \mathbf{B}_u' & \mathbf{B}_e' & \mathbf{H} \end{pmatrix}, \tag{5}$$

and

$$\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}, \ \mathbf{B} = \mathbf{ZB}_u + \mathbf{B}_e.$$

Now suppose that in repeated sampling $\mathbf{w}$ is selected such that $E(\mathbf{w}) = \mathbf{s} \neq \mathbf{d}$, and $Var(\mathbf{w}) = \mathbf{H}_s$. Then the conditional moments are as follows.

$$E \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\boldsymbol{\beta} + \mathbf{Bt} \\ \mathbf{B}_u\mathbf{t} \\ \mathbf{s} \end{pmatrix}, \tag{6}$$

where $\mathbf{t} = \mathbf{H}^{-1}(\mathbf{s} - \mathbf{d})$.

$$Var \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{w} \end{pmatrix} = \begin{pmatrix} \mathbf{V} - \mathbf{BH}_0\mathbf{B}' & \mathbf{ZG} - \mathbf{BH}_0\mathbf{B}' & \mathbf{BH}^{-1}\mathbf{H}_s \\ \mathbf{GZ}' - \mathbf{BH}_0\mathbf{B}' & \mathbf{G} - \mathbf{B}_u\mathbf{H}_0\mathbf{B}_u' & \mathbf{B}_u\mathbf{H}^{-1}\mathbf{H}_s \\ \mathbf{H}_s\mathbf{H}^{-1}\mathbf{B}' & \mathbf{H}_s\mathbf{H}^{-1}\mathbf{B}_u' & \mathbf{H}_s \end{pmatrix}, \tag{7}$$

where $\mathbf{H}_0 = \mathbf{H}^{-1}(\mathbf{H} - \mathbf{H}_s)\mathbf{H}^{-1}$.

4 BLUE And BLUP Under Selection Model

To find BLUE of $\mathbf{K}'\boldsymbol{\beta}$ and BLUP of $\mathbf{u}$ under this conditional model, find linear functions that minimize diagonals of $Var(\mathbf{K}'\boldsymbol{\beta})$ and variance of diagonals of $(\hat{\mathbf{u}} - \mathbf{u})$ subject to

$$E(\mathbf{K}'\boldsymbol{\beta}^o) \ = \ \mathbf{K}'\boldsymbol{\beta} \text{ and } E(\hat{\mathbf{u}}) \ = \ \mathbf{B}_u\mathbf{t}.$$

This is accomplished by modifying GLS and mixed model equations as follows.

$$\begin{pmatrix} \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{V}^{-1}\mathbf{B} \\ \mathbf{B}'\mathbf{V}^{-1}\mathbf{X} & \mathbf{B}'\mathbf{V}^{-1}\mathbf{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \mathbf{t}^o \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\ \mathbf{B}'\mathbf{V}^{-1}\mathbf{y} \end{pmatrix} \tag{8}$$

BLUP of $\mathbf{k}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}$ is

$$\mathbf{k}'\boldsymbol{\beta}^o + \mathbf{m}'\boldsymbol{\beta}_u\mathbf{t}^o + \mathbf{m}'\mathbf{G}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^o - \mathbf{B}\mathbf{t}^o).$$

Modified mixed model equations are

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1`}\mathbf{Z} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{B}_e \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{B}_e - \mathbf{G}^{-1}\mathbf{B}_u \\ \mathbf{B}_e'\mathbf{R}^{-1}\mathbf{X} & \mathbf{B}_e'\mathbf{R}^{-1}\mathbf{Z}^{-}\mathbf{B}_u'\mathbf{G}^{-1} & \mathbf{B}_e'\mathbf{R}^{-1}\mathbf{B}_e + \mathbf{B}_u'\mathbf{G}^{-1}\mathbf{B}_u \end{pmatrix}$$

$$\begin{pmatrix} \boldsymbol{\beta}^o \\ \mathbf{u}^o \\ \mathbf{t}^o \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{Z}'\mathbf{R}^{-1`}\mathbf{y} & \mathbf{B}_e'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}'. \tag{9}$$

BLUP of $\mathbf{k}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}$ is $\mathbf{k}'\boldsymbol{\beta}^o + \mathbf{m}'\mathbf{u}^o$. In equations (8) and (9) we use $\mathbf{u}^o$ rather than $\hat{\mathbf{u}}$ because the solution is not always invariant. It is necessary therefore to examine whether the function is predictable. The sampling and prediction error variances come from a g-inverse of (8) or (9). Let a g-inverse of the matrix of (8) be

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

then

$$Var(\mathbf{K}'\boldsymbol{\beta}^o) \ = \ \mathbf{K}'\mathbf{C}_{11}\mathbf{K}. \tag{10}$$

Let a g-inverse of the matrix of (9) be

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{C}_{13} \\ \mathbf{C}_{12}' & \mathbf{C}_{22} & \mathbf{C}_{23} \\ \mathbf{C}_{13}' & \mathbf{C}_{23}' & \mathbf{C}_{33} \end{pmatrix},$$

Then

$$Var(\mathbf{K}'\boldsymbol{\beta}^o) \ = \ \mathbf{K}'\mathbf{C}_{11}\mathbf{K}. \tag{11}$$

$$Cov(\mathbf{K}'\boldsymbol{\beta}^o, \ \hat{\mathbf{u}} - \mathbf{u}) \ = \ \mathbf{K}'\mathbf{C}_{12}. \tag{12}$$

$$Var(\hat{\mathbf{u}} - \mathbf{u}) \ = \ \mathbf{C}_{22}. \tag{13}$$

$$Cov(\mathbf{K}'\boldsymbol{\beta}^o, \ \hat{\mathbf{u}}') \ = \ \mathbf{K}'\mathbf{C}_{13}\mathbf{B}_u'. \tag{14}$$

$$Var(\hat{\mathbf{u}}) \ = \ \mathbf{G} - \mathbf{C}_{22} + \mathbf{C}_{23}\mathbf{B}_u' + \mathbf{B}_u\mathbf{C}_{23}' - \mathbf{B}_u\mathbf{H}_0\mathbf{B}_u'. \tag{15}$$

Note that (10), ..., (13) are analogous to the results for the no selection model, but (14) and (15) are more complicated. The problems with the methods of this section are that $\mathbf{w}$ may be difficult to define and the values of $\mathbf{B}_u$ and $\mathbf{B}_e$ may not be known. Special cases exist that simplify the problem. This is true particularly if selection is on a subvector of $\mathbf{y}$, and if estimators and predictors can be found that are invariant to the value of $\boldsymbol{\beta}$ associated with the selection functions.

5 Selection On Linear Functions Of y

Suppose that whatever selection has occurred has been a consequence of use of the record vector or some subvector of $\mathbf{y}$. Let the type of selection be described in terms of a set of linear functions, say $\mathbf{L}'\mathbf{y}$, such that

$$E(\mathbf{L}'\mathbf{y}) \ = \ \mathbf{L}'\mathbf{X}\boldsymbol{\beta} + \mathbf{t},$$

where $\mathbf{t} \neq \mathbf{0}$. $\mathbf{t}$ would be $\mathbf{0}$ if there were no selection.

$$Var(\mathbf{L}'\mathbf{y}) \ = \ \mathbf{H}_s.$$

Let us see how this relates to (9).

$$\mathbf{B}_u \ = \ \mathbf{G}\mathbf{Z}'\mathbf{L}, \ \mathbf{B}_e \ = \ \mathbf{R}\mathbf{L}, \ \mathbf{H} \ = \ \mathbf{L}'\mathbf{V}\mathbf{L}.$$

Substituting these values in (9) we obtain

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{L} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{0} \\ \mathbf{L}'\mathbf{X} & \mathbf{0} & \mathbf{L}'\mathbf{V}\mathbf{L} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \hat{\mathbf{u}} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{L}'\mathbf{y} \end{pmatrix}. \quad (16)$$

5.1 Selection with $\mathbf{L}'\mathbf{X} \ = \ \mathbf{0}$

An important property of (16) is that if $\mathbf{L}'\mathbf{X} \ = \ \mathbf{0}$, then $\hat{\mathbf{u}}$ is a solution to the mixed model equations assuming no selection. Thus we have the extremely important result that whenever $\mathbf{L}'\mathbf{X} \ = \ \mathbf{0}$, BLUE and BLUP in the selection model can be computed by using the mixed model equations ignoring selection. Our example in Section 2 can be formulated as a problem with $\mathbf{L}'\mathbf{X} \ = \ \mathbf{0}$. Order the observations by sires within periods. Let

$$\mathbf{y}' \ = \ [\bar{y}_{11.}', \ \bar{y}_{21.}', \ \bar{y}_{31.}', \ \bar{y}_{41.}', \ \bar{y}_{12.}', \ \bar{y}_{22.}'].$$

According to our assumptions of the method of selection

$$\bar{y}_{11.} \ > \bar{y}_{21.} \ > \bar{y}_{31.} \ > \bar{y}_{41.}.$$

Based on this we can write

$$\mathbf{L}' = \begin{pmatrix} \mathbf{1}_{10}' & -\mathbf{1}_{10}' & \mathbf{0}_{620}' & \\ \mathbf{0}_{10}' & \mathbf{1}_{10}' & -\mathbf{1}_{10}' & \mathbf{0}_{610}' \\ \mathbf{0}_{20}' & \mathbf{1}_{10}' & -\mathbf{1}_{10}' & \mathbf{0}_{600}' \end{pmatrix}$$

where

$$\begin{aligned} &\mathbf{1}_{10}' \text{ denotes a row vector of 10 one's.} \\ &\mathbf{0}_{620}' \text{ denotes a null row vector with 620 elements, etc.} \end{aligned}$$

It is easy to see that $\mathbf{L}'\mathbf{X} = \mathbf{0}$, and that explains why we obtain unbiased estimators and predictors from the solution to the mixed model equations.

Let us consider a much more general selection method that insures that $\mathbf{L}'\mathbf{X} = \mathbf{0}$. Suppose in the first cycle of selection that data to be used in selection comprise a subvector of $\mathbf{y}$, say $\mathbf{y}_s$. We know that $\mathbf{X}_s\boldsymbol{\beta}$, consisting of such fixed effects as age, sex and season, causes confusion in making selection decisions, so we adjust the data for some estimate of $\mathbf{X}_s\boldsymbol{\beta}$, say $\mathbf{X}_s\boldsymbol{\beta}^o$ so the data for selection become $\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta}^o$. Suppose that we then evaluate the i^{th} candidate for selection by the function $\mathbf{a}_i'(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta}^o)$. There are c candidates for selection and s of them are to be selected. Let us order the highest s of the selection functions with labels 1 for the highest, 2 for the next highest, etc. Leave the lowest $c - s$ unordered. Then the animals labelled 1, ..., s are selected, and there may, in addition, be differential usage of them subsequently depending upon their rank. Now express these selection criteria as a set of differences, of $\mathbf{a}'(\mathbf{y}_s - \mathbf{X}_s\boldsymbol{\beta}^o)$,

$$1 - 2,\ 2 - 3, (s-1) - s,\ s - (s+1), ...,\ s - c.$$

Because $\mathbf{X}_s\boldsymbol{\beta}^o$ is presumably a linear function of $\mathbf{y}$ these differences are a set of linear functions of $\mathbf{y}$, say $\mathbf{L}'\mathbf{y}$. Now suppose $\boldsymbol{\beta}^o$ is computed in such a way that $E(\mathbf{X}_s\boldsymbol{\beta}^o) = \mathbf{X}_s\boldsymbol{\beta}$ in a no selection model. (It need not be an unbiased estimator under a selection model, but if it is, that creates no problem). Then $\mathbf{L}'\mathbf{X}$ will be null, and the mixed model equations ignoring selection yield BLUE and BLUP for the selection model. This result is correct if we know $\mathbf{G}$ and $\mathbf{R}$ to proportionality. Errors in $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{R}}$ will result in biases under a selection model, the magnitude of bias depending upon how seriously $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{R}}$ depart from $\mathbf{G}$ and $\mathbf{R}$ and upon the intensity of selection. The result also depends upon normality. The consequences of departure from this distribution are not known in general, but depend upon the form of the conditional means.

We can extend this description of selection for succeeding cycles of selection and still have $\mathbf{L}'\mathbf{X} = \mathbf{0}$. The results above depended upon the validity of the Pearson result and normality. Now with continued selection we no longer have the multivariate normal distribution, and consequently the Pearson result may not apply exactly. Nevertheless with traits of relatively low heritability and with a new set of normally distributed errors for each new set of records, the conditional distribution of Pearson may well be a suitable approximation.

6 With Non-Observable Random Factors

The previous section deals with strict truncation selection on a linear function of records. This is not entirely realistic as there certainly are other factors that influence the selection decisions, for example, death, infertility, undesirable traits not recorded as a part of the data vector, $\mathbf{y}$. It even may be the case that the breeder did have available additional records and used them, but these were not available to the person or organization attempting to estimate or predict. For these reasons, let us now consider a different selection model, the functions used for making selection decision now being

$$\mathbf{a}_i^{'}\ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^o)\ +\ \alpha_i$$

where α_i is a random variable not observable by the person performing estimation and prediction, but may be known or partially known by the breeder. This leads to a definition of $\mathbf{w}$ as follows

$$\mathbf{w}\ =\ \mathbf{L}'\mathbf{y} + \boldsymbol{\theta}.$$

$$Cov(\mathbf{y}, \mathbf{w}') \ = \ \mathbf{B} = \mathbf{VL} + \mathbf{C},\ \text{where } \mathbf{C} = Cov(\mathbf{y}, \boldsymbol{\theta}'). \tag{17}$$

$$Cov(\mathbf{u}, \mathbf{w}') \ = \ \mathbf{B}_u = \mathbf{GZ}'\mathbf{L} + \mathbf{C}_u,\ \text{where } \mathbf{C}_u = Cov(\mathbf{u}, \boldsymbol{\theta}') \tag{18}$$

$$Cov(\mathbf{e}, \mathbf{w}') \ = \ \mathbf{B}_e = \mathbf{RL} + \mathbf{C}_e,\ \text{where } \mathbf{C}_e = Cov(\mathbf{e}, \boldsymbol{\theta}'). \tag{19}$$

$$Var(\mathbf{w}) \ = \ \mathbf{L}'\mathbf{VL} + \mathbf{L}'\mathbf{C} + \mathbf{C}'\mathbf{L} + \mathbf{C}_\theta,\ \text{where } \mathbf{C}_\theta = Var(\boldsymbol{\theta}). \tag{20}$$

Applying these results to (9) we obtain the modified mixed model equations below)

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}'\mathbf{L} + \mathbf{X}'\mathbf{R}^{-1}\mathbf{C}_e \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{Z}\mathbf{R}^{-1}\mathbf{C}_e - \mathbf{G}^{-1}\mathbf{C}_u \\ \mathbf{L}'\mathbf{X} + \mathbf{C}_e^{'}\mathbf{R}^{-1}\mathbf{X} & \mathbf{C}_e^{'}\mathbf{R}^{-1}\mathbf{Z}' + \mathbf{C}_u^{'}\mathbf{G}^{-1} & \boldsymbol{\psi} \end{pmatrix}$$

$$\begin{pmatrix} \boldsymbol{\beta}^o \\ \hat{\mathbf{u}} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{L}'\mathbf{y} + \mathbf{C}_e^{'}\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}, \tag{21}$$

where $\boldsymbol{\psi}\ =\ \mathbf{L}'\mathbf{VL} + \mathbf{C}_e^{'}\mathbf{R}^{-1}\mathbf{C}_e + \mathbf{C}_u^{'}\mathbf{G}^{-1}\mathbf{C}_u + \mathbf{L}'\mathbf{C} + \mathbf{C}'\mathbf{L}$.

Now if $\mathbf{L}'\mathbf{X} = \mathbf{0}$ and if $\boldsymbol{\theta}$ is uncorrelated with $\mathbf{u}$ and $\mathbf{e}$, these equations reduce to the regular mixed model equations that ignore selection. Thus the non-observable variable used in selection causes no difficulty when it is uncorrelated with $\mathbf{u}$ and $\mathbf{e}$. If the correlations are non-zero, one needs the magnitudes of $\mathbf{C}_e$, $\mathbf{C}_u$ to obtain BLUE and BLUP. This could be most difficult to determine. The selection models of Sections 5 and 6 are described in Henderson (1982).

7 Selection On A Subvector Of $\mathbf{y}$

Many situations exist in which selection has occurred on $\mathbf{y}_1$, but $\mathbf{y}_2$ is unselected, where the model is

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1\boldsymbol{\beta} \\ \mathbf{X}_2\boldsymbol{\beta} \end{pmatrix} + \begin{pmatrix} \mathbf{Z}_1\mathbf{u} \\ \mathbf{Z}_2\mathbf{u} \end{pmatrix} + \begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix},$$

$$Var\begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{12} & \mathbf{R}_{22} \end{pmatrix}.$$

Presumably $\mathbf{y}_1$ are data from earlier generations. Suppose that selection which has occurred can be described as

$$\mathbf{L}'\mathbf{y} = (\mathbf{M}' \;\; \mathbf{0})\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{pmatrix}.$$

Then the equations of (16) become

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{X}_1'\mathbf{M} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1} & \mathbf{0} \\ \mathbf{M}'\mathbf{X}_1 & \mathbf{0} & \mathbf{M}'\mathbf{V}_{11}\mathbf{M} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \hat{\mathbf{u}} \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{M}'\mathbf{y}_1 \end{pmatrix} \quad (22)$$

Then if $\mathbf{M}'\mathbf{X}_1 = \mathbf{0}$, unmodified mixed model equations yield unbiased estimators and predictors. Also if selection is on $\mathbf{M}'\mathbf{y}_1$ plus a non-observable variable uncorrelated with $\mathbf{u}$ and $\mathbf{e}$ and $\mathbf{M}'\mathbf{X}_1 = \mathbf{0}$, the unmodified equations are appropriate.

Sometimes $\mathbf{y}_1$ is not available to the person predicting functions of $\boldsymbol{\beta}$ and $\mathbf{u}$. Now if we assume that $\mathbf{R}_{12} = \mathbf{0}$,

$$\begin{aligned} E(\mathbf{y}_2 \mid \mathbf{M}'\mathbf{y}_1) &= \mathbf{Z}_2\mathbf{G}\mathbf{Z}_1'\mathbf{M}\mathbf{k}. \\ E(\mathbf{u} \mid \mathbf{M}'\mathbf{y}_1) &= \mathbf{G}\mathbf{Z}_1'\mathbf{M}\mathbf{k}, \end{aligned}$$

where

$$\mathbf{k} = (\mathbf{M}'\mathbf{V}_{11}\mathbf{M})^{-1}\mathbf{t},$$

$\mathbf{t}$ being the deviation of mean of $\mathbf{M}'\mathbf{y}_1$ from $\mathbf{X}_1'\boldsymbol{\beta}$. If we solve for $\boldsymbol{\beta}^o$ and $\mathbf{u}^o$ in the equations (23) that regard $\mathbf{u}$ as fixed for purposes of computation, then

$$E[\mathbf{K}'\boldsymbol{\beta}^o + \mathbf{T}'\mathbf{u}^o] = \mathbf{K}'\boldsymbol{\beta} + E[\mathbf{T}'\mathbf{u} \mid \mathbf{M}'\mathbf{y}_1]$$

provided that $\mathbf{K}'\boldsymbol{\beta} + \mathbf{T}'\mathbf{u}$ is estimable under a fixed $\mathbf{u}$ model.

$$\begin{pmatrix} \mathbf{X}_2'\mathbf{R}_{22}^{-1}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}_{22}^{-1}\mathbf{Z}_2 \\ \mathbf{Z}_2'\mathbf{R}_{22}^{-1}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}_{22}^{-1}\mathbf{Z}_2 \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \mathbf{u}^o \end{pmatrix} = \begin{pmatrix} \mathbf{X}_2'\mathbf{R}_{22}^{-1}\mathbf{y}_2 \\ \mathbf{Z}_2'\mathbf{R}_{22}^{-1}\mathbf{y}_2 \end{pmatrix}. \quad (23)$$

This of course does not prove that $\mathbf{K}'\boldsymbol{\beta}^o + \mathbf{T}'\mathbf{u}^o$ is BLUP of this function under $\mathbf{M}'\mathbf{y}_1$ selection and utilizing only $\mathbf{y}_2$. Let us examine modified mixed model equations regarding $\mathbf{y}_2$ as the data vector and $\mathbf{M}'\mathbf{y}_1 = \mathbf{w}$. We set up equations like (21).

$$\begin{aligned}\mathbf{B}_e &= Cov[\mathbf{e}_2, \mathbf{y}_1'\mathbf{M}] = \mathbf{0} \text{ if we assume } \mathbf{R}_{12} = \mathbf{0}. \\ \mathbf{B}_u &= Cov(\mathbf{u}, \mathbf{y}_1'\mathbf{M}) = \mathbf{G}\mathbf{Z}_1'\mathbf{M}.\end{aligned}$$

Then the modified mixed model equations become

$$\begin{pmatrix} \mathbf{X}_2'\mathbf{R}_{22}^{-1}\mathbf{X}_2 & \mathbf{X}_2'\mathbf{R}_{22}^{-1}\mathbf{Z}_2 & \mathbf{0} \\ \mathbf{Z}_2'\mathbf{R}_{22}^{-1}\mathbf{X}_2 & \mathbf{Z}_2'\mathbf{R}_{22}^{-1}\mathbf{Z}_2 + \mathbf{G}^{-1} & -\mathbf{Z}_1'\mathbf{M} \\ \mathbf{0} & -\mathbf{M}'\mathbf{Z}_1 & \mathbf{M}'\mathbf{Z}_1\mathbf{G}\mathbf{Z}_1'\mathbf{M} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \mathbf{u}^o \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_2'\mathbf{R}_{22}^{-1}\mathbf{y}_2 \\ \mathbf{Z}_2'\mathbf{R}_{22}^{-1}\mathbf{y}_2 \\ \mathbf{0} \end{pmatrix}. \tag{24}$$

A sufficient set of conditions for the solution to $\boldsymbol{\beta}^o$ and $\mathbf{u}^o$ in these equations being equal to those of (23) is that $\mathbf{M}' = \mathbf{I}$ and $\mathbf{Z}_1$ be non-singular. In that case if we "absorb" $\boldsymbol{\theta}$ we obtain the equations of (23).

Now it seems implausible that $\mathbf{Z}_1$ be non-singular. In fact, it would usually have more rows than columns. A more realistic situation is the following. Let $\bar{\mathbf{y}}_1$ be the mean of smallest subclasses in the $\bar{\mathbf{y}}_1$ vector. Then the model for $\bar{\mathbf{y}}_1$ is

$$\bar{\mathbf{y}}_1 = \bar{\mathbf{X}}_1\boldsymbol{\beta} + \bar{\mathbf{Z}}_1\mathbf{u} + \mathbf{e}_1.$$

See Section 1.6 for a description of such models. Now suppose selection can be described as $\mathbf{I}\bar{\mathbf{y}}_1$. Then

$$\begin{aligned}\mathbf{B}_e &= \mathbf{0} \text{ if } \mathbf{R}_{12} = \mathbf{0}, \text{ and} \\ \mathbf{B}_u &= \bar{\mathbf{Z}}_1.\end{aligned}$$

Then a sufficient condition for GLS using $\mathbf{y}_2$ only and computing as though $\mathbf{u}$ is fixed to be BLUP under the selection model and regarding $\mathbf{y}_2$ as that data vector is that $\bar{\mathbf{Z}}_1$ be non-singular. This might well be the case in some practical situations. This is the selection model in our sire example.

8 Selection On u

Cases exist in animal breeding in which the data represent observations associated with $\mathbf{u}$ that have been subject to prior selection, but with the data that were used for such selection not available. Henderson (1975a) described this as $\mathbf{L}'\mathbf{u}$ selection. If no selection on the observable $\mathbf{y}$ vector has been effected, BLUE and BLUP come from solution to equations (25).

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & -\mathbf{L} \\ \mathbf{0} & -\mathbf{L}' & \mathbf{L}'\mathbf{G}\mathbf{L} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \mathbf{u}^o \\ \boldsymbol{\theta} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{0} \end{pmatrix} \tag{25}$$

These reduce to (26) by "absorbing" $\boldsymbol{\theta}$.

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} - \mathbf{L}(\mathbf{L}'\mathbf{G}\mathbf{L})^{-1}\mathbf{L}' \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \mathbf{u}^o \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix} \tag{26}$$

The notation $\mathbf{u}^o$ is used rather than $\hat{\mathbf{u}}$ since the solution may not be unique, in which case we need to consider functions of $\mathbf{u}^o$ that are invariant to the solution. It is simple to prove that $\mathbf{K}'\boldsymbol{\beta}^o + \mathbf{M}'\mathbf{u}^o$ is an unbiased predictor of $\mathbf{K}'\boldsymbol{\beta} + \mathbf{M}'\mathbf{u}$, where $\boldsymbol{\beta}^o$ and $\mathbf{u}^o$ are some solution to (27) and this is an estimable function under a fixed $\mathbf{u}$ model

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \mathbf{u}^o \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}. \tag{27}$$

A sufficient condition for this to be BLUP is that $\mathbf{L} = \mathbf{I}$. The proof comes by substituting $\mathbf{I}$ for $\mathbf{L}$ in (26). In sire evaluation $\mathbf{L}'\mathbf{u}$ selection can be accounted for by proper grouping. Henderson (1973) gave an example of this for unrelated sires. Quaas and Pollak (1981) extended this result for related sires. Let $\mathbf{G} = \mathbf{A}\sigma_s^2$. Write the model for progeny as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{Z}\mathbf{Q}\mathbf{g} + \mathbf{Z}\mathbf{S} + \mathbf{e},$$

where $\mathbf{h}$ refers to fixed herd-year-season and $\mathbf{g}$ to fixed group effects. Then it was shown that such grouping is equivalent to no grouping, defining $\mathbf{L} = \mathbf{G}^{-1}\mathbf{Q}$, and then using (25). We illustrate this method with the following data.

group	sire	n_i	y_i
1	1	2	10
	2	3	12
	3	1	7
2	4	2	6
	5	3	8
3	6	1	3
	7	2	5
	8	1	2
	9	2	8

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & .5 & .5 & 0 & .25 & .25 & 0 & .125 \\ & 1 & 0 & 0 & .5 & 0 & 0 & .25 & 0 \\ & & 1 & .25 & 0 & .5 & .125 & 0 & .25 \\ & & & 1 & 0 & .125 & .5 & 0 & .0625 \\ & & & & 1 & 0 & 0 & .5 & 0 \\ & & & & & 1 & .0625 & 0 & .5 \\ & & & & & & 1 & 0 & .03125 \\ & & & & & & & 1 & 0 \\ & & & & & & & & 1 \end{pmatrix}.$$

Assume a model $y_{ijk} = \mu + g_i + s_{ij} + e_{ijk}$. Let $\sigma_e^2 = 1$, $\sigma_s^2 = 12^{-1}$, then $\mathbf{G} = 12^{-1}\mathbf{A}$. The solution to the mixed model equations with μ dropped is

$$\hat{\mathbf{g}} = (4.8664,\ 2.8674,\ 2.9467),$$

$$\hat{\mathbf{s}} = (.0946,\ -.1937,\ .1930,\ .0339,\ -.1350,\ .1452,\ -.0346,\ -.1192,\ .1816).$$

The sire evaluations are $\hat{g}_i + \hat{s}_{ij}$ and these are (4.961, 4.673, 5.059, 2.901, 2.732, 3.092, 2.912, 2.827, 3.128).

$$\mathbf{Q}' = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

This gives

$$\mathbf{L}' = \begin{pmatrix} 12 & 16 & 12 & -8 & -8 & -8 & 0 & 0 & 0 \\ -8 & -8 & 0 & 20 & 20 & 0 & -8 & -8 & 0 \\ 0 & 0 & -8 & -8 & -8 & 12 & 16 & 16 & 8 \end{pmatrix} = \mathbf{G}^{-1}\mathbf{Q},$$

and

$$\mathbf{L}'\mathbf{G}\mathbf{L} = \begin{pmatrix} 40 & -16 & -8 \\ & 40 & -16 \\ & & 52 \end{pmatrix}.$$

Then the equations like (25) give a solution

$$\mu^o = 2.9014,$$

$$\mathbf{s}^o = (2.0597,\ 1.7714,\ 2.1581,\ 0, -.1689,\ .1905,\ .0108,\ -.0739,\ .2269),$$

$$\boldsymbol{\theta} = (1.9651,\ -.0339,\ .0453).$$

The sire evaluation is $\mu^o + s_i^o$ and this is the same as when groups were included.

9 Inverse Of Conditional A Matrix

In some applications the base population animals are not a random sample from some population, but rather have been selected. Consequently the additive genetic variance-covariance matrix for these animals is not $\sigma_a^2\mathbf{I}$, where σ_a^2 is the additive genetic variance in the population from which these animals were were taken. Rather it is $\mathbf{A}_s\sigma_{a*}^2$, where $\sigma_{a*}^2 \neq \sigma_a^2$ in general. If the base population had been a random sample from some population, the entire $\mathbf{A}$ matrix would be

$$\begin{pmatrix} \mathbf{I} & \mathbf{A}_{12} \\ \mathbf{A}_{12}' & \mathbf{A}_{22} \end{pmatrix}. \tag{28}$$

The inverse of this can be found easily by the method described by Henderson (1976). Denote this by

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}' & \mathbf{C}_{22} \end{pmatrix}. \tag{29}$$

If the Pearson result holds, the **A** matrix for this conditional population is

$$\begin{pmatrix} \mathbf{A}_s & \mathbf{A}_s\mathbf{A}_{12} \\ \mathbf{A}_{12}'\mathbf{A}_s & \mathbf{A}_{22} - \mathbf{A}_{12}'(\mathbf{I} - \mathbf{A}_s)\mathbf{A}_{12} \end{pmatrix} \tag{30}$$

The inverse of this matrix is

$$\begin{pmatrix} \mathbf{C}_s & \mathbf{C}_{12} \\ \mathbf{C}_{12}' & \mathbf{C}_{22} \end{pmatrix}, \tag{31}$$

$$\text{where } \mathbf{C}_s = \mathbf{A}_s^{-1} - \mathbf{C}_{12}\mathbf{A}_{12}', \tag{32}$$

and $\mathbf{C}_{12}$, $\mathbf{C}_{22}$ are the same as in (29)

Note that most of the elements of the inverse of the conditional matrix (31) are the same as the elements of the inverse of the unconditional matrix (29). Thus the easy method for $\mathbf{A}^{-1}$ can be used, and the only elements of the unconditional **A** needed are those of $\mathbf{A}_{12}$. Of course this method is not appropriate for the situation in which $\mathbf{A}_s$ is singular. We illustrate with

$$\text{unconditional } \mathbf{A} = \begin{pmatrix} 1.0 & 0 & .2 & .3 & .1 \\ & 1.0 & .1 & .2 & .2 \\ & & 1.1 & .3 & .5 \\ & & & 1.2 & .2 \\ & & & & 1.3 \end{pmatrix}.$$

The first 2 animals are selected so that

$$\mathbf{A}_s = \begin{pmatrix} .7 & -.4 \\ -.4 & .8 \end{pmatrix}.$$

Then by (30) the conditional **A** is

$$\begin{pmatrix} .7 & -.4 & .1 & .13 & -.01 \\ & .8 & 0 & .04 & .12 \\ & & 1.07 & .25 & .47 \\ & & & 1.117 & .151 \\ & & & & 1.273 \end{pmatrix}.$$

The inverse of the unconditional **A** is

$$\begin{pmatrix} 1.103168 & .064741 & -.136053 & -.251947 & -.003730 \\ & 1.062405 & .003286 & -.170155 & -.143513 \\ & & 1.172793 & -.191114 & -.411712 \\ & & & .977683 & -.031349 \\ & & & & .945770 \end{pmatrix}.$$

The inverse of the conditional $\mathbf{A}$ is

$$\begin{pmatrix} 2.103168 & 1.064741 & -.136053 & -.251947 & -.003730 \\ & 1.812405 & .003286 & -.170155 & -.143513 \\ & & 1.172793 & -.191114 & -.411712 \\ & & & .977683 & -.031349 \\ & & & & .954770 \end{pmatrix}.$$

$$\mathbf{A}_s^{-1} = \begin{pmatrix} 2.0 & 1.0 \\ 1.0 & 1.75 \end{pmatrix}, \ \mathbf{C}_{12} = \begin{pmatrix} -.136053 & -.251947 & -.003730 \\ .003286 & -.170155 & -.143513 \end{pmatrix},$$

$$\mathbf{A}_{12}' = \begin{pmatrix} .2 & .1 \\ .3 & .2 \\ .1 & .2 \end{pmatrix},$$

and

$$\mathbf{A}_s^{-1} - \mathbf{C}_{12}\mathbf{A}_{12}' = \begin{pmatrix} 2.103168 & 1.064741 \\ & 1.812405 \end{pmatrix},$$

which checks with the upper 2×2 submatrix of the inverse of conditional $\mathbf{A}$.

10 Minimum Variance Linear Unbiased Predictors

In all previous discussions of prediction in both the no selection and the selection model we have used as our criteria linear and unbiased with minimum variance of the prediction error. That is, we use $\mathbf{a}'\mathbf{y}$ as the predictor of $\mathbf{k}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}$ and find $\mathbf{a}$ that minimizes $E(\mathbf{a}'\mathbf{y} - \mathbf{k}'\boldsymbol{\beta} - \mathbf{m}'\mathbf{u})^2$ subject to the restriction that $E(\mathbf{a}'\mathbf{y}) = \mathbf{k}'\boldsymbol{\beta} + E(\mathbf{m}'\mathbf{u})$. This is a logical criterion for making selection decisions. For other purposes such as estimating genetic trend one might wish to minimize the variance of the predictor rather than the variance of the prediction error. Consequently in this section we shall derive a predictor of $\mathbf{k}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{u}$, say $\mathbf{a}'\mathbf{y}$, such that $E(\mathbf{a}'\mathbf{y}) = \mathbf{k}'\boldsymbol{\beta} + E(\mathbf{m}'\mathbf{u})$ and has minimum variance. For this purpose we use the $\mathbf{L}'\mathbf{y}$ type of selection described in Section 5. Let

$$\begin{aligned} E(\mathbf{L}'\mathbf{y}) &= \mathbf{L}'\mathbf{X}\boldsymbol{\beta} + \mathbf{t}, \ \mathbf{t} \neq \mathbf{0}. \\ Var(\mathbf{L}'\mathbf{y}) &= \mathbf{H}_s \neq \mathbf{L}'\mathbf{V}\mathbf{L}. \end{aligned}$$

Then

$$\begin{aligned} E(\mathbf{y} \mid \mathbf{L}'\mathbf{y}) &= \mathbf{X}\boldsymbol{\beta} + \mathbf{V}\mathbf{L}(\mathbf{L}'\mathbf{V}\mathbf{L})^{-1}\mathbf{t} \equiv \mathbf{X}\boldsymbol{\beta} + \mathbf{V}\mathbf{L}\mathbf{d}. \\ E(\mathbf{u} \mid \mathbf{L}'\mathbf{y}) &= \mathbf{G}\mathbf{Z}'\mathbf{L}(\mathbf{L}'\mathbf{V}\mathbf{L})^{-1}\mathbf{t} \equiv \mathbf{G}\mathbf{Z}'\mathbf{L}\mathbf{d}. \\ Var(\mathbf{y} \mid \mathbf{L}'\mathbf{y}) &= \mathbf{V} - \mathbf{V}\mathbf{L}(\mathbf{L}'\mathbf{V}\mathbf{L})^{-1}(\mathbf{L}'\mathbf{V} - \mathbf{H}_s)(\mathbf{L}'\mathbf{V}\mathbf{L})^{-1}\mathbf{L}'\mathbf{V} \equiv \mathbf{V}_s. \end{aligned}$$

Then we minimize $Var(\mathbf{a}'\mathbf{y})$ subject to $E(\mathbf{a}'\mathbf{y}) = \mathbf{k}'\boldsymbol{\beta} + \mathbf{m}'\mathbf{G}\mathbf{Z}'\mathbf{L}\mathbf{d}$. For this expectation to be true it is required that

$$\mathbf{X}'\mathbf{a} = \mathbf{k} \text{ and } \mathbf{L}'\mathbf{V}\mathbf{a} = \mathbf{L}'\mathbf{Z}\mathbf{G}\mathbf{m}.$$

Therefore we solve equations (33) for $\mathbf{a}$.

$$\begin{pmatrix} \mathbf{V}_s & \mathbf{X} & \mathbf{VL} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} \\ \mathbf{L}'\mathbf{V} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \boldsymbol{\theta} \\ \boldsymbol{\phi} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{k} \\ \mathbf{L}'\mathbf{ZGm} \end{pmatrix} \tag{33}$$

Let a g-inverse of the matrix of (33) be

$$\begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} & \mathbf{C}_{13} \\ \mathbf{C}_{12}' & \mathbf{C}_{22} & \mathbf{C}_{23} \\ \mathbf{C}_{13}' & \mathbf{C}_{23}' & \mathbf{C}_{33} \end{pmatrix}. \tag{34}$$

Then

$$\mathbf{a}' = \mathbf{k}'\mathbf{C}_{12}' + \mathbf{m}'\mathbf{GZ}'\mathbf{LC}_{13}'.$$

But it can be shown that a g-inverse of the matrix of (35) gives the same values of $\mathbf{C}_{11}, \mathbf{C}_{12}, \mathbf{C}_{13}$. These are subject to $\mathbf{L}'\mathbf{X} = \mathbf{0}$,

$$\begin{aligned} \mathbf{C}_{11} &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} - \mathbf{L}(\mathbf{L}'\mathbf{V1})^{-1}\mathbf{L}'. \\ \mathbf{C}_{12} &= \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \ \mathbf{C}_{13} = \mathbf{L}(\mathbf{L}'\mathbf{VL})^{-1}\mathbf{L}'. \end{aligned}$$

Consequently we can solve for $\mathbf{a}$ in (35), a simpler set of equations than (33).

$$\begin{pmatrix} \mathbf{V} & \mathbf{X} & \mathbf{VL} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} \\ \mathbf{L}'\mathbf{V} & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \boldsymbol{\theta} \\ \boldsymbol{\phi} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{k} \\ \mathbf{L}'\mathbf{ZGm} \end{pmatrix}. \tag{35}$$

By techniques described in Henderson (1975) it can be shown that

$$\mathbf{a}'\mathbf{y} = \mathbf{k}'\boldsymbol{\beta}^o + \mathbf{m}'\mathbf{GZ}'\mathbf{Lt}^o$$

where $\boldsymbol{\beta}^o$, $\mathbf{t}^o$ are a solution to (36).

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} & \mathbf{0} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{L}'\mathbf{VL} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}^o \\ \hat{\mathbf{u}} \\ \mathbf{t}^o \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{L}'\mathbf{y} \end{pmatrix}. \tag{36}$$

Thus $\boldsymbol{\beta}^o$ is a GLS solution ignoring selection, and $\mathbf{t}^o = (\mathbf{L}'\mathbf{VL})^{-1}\mathbf{L}'\mathbf{y}$. It was proved in Henderson (1975a) that

$$\begin{aligned} Var(\mathbf{K}'\boldsymbol{\beta}^o) &= \mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{K} = \mathbf{K}'\mathbf{C}_{11}\mathbf{K}, \\ Cov(\mathbf{K}'\boldsymbol{\beta}^o, \ \mathbf{t}') &= \mathbf{0}, \text{ and} \\ Var(\mathbf{t}) &= (\mathbf{L}'\mathbf{VL})^{-1}\mathbf{H}_s(\mathbf{L}'\mathbf{VL})^{-1}. \end{aligned}$$

Thus the variance of the predictor, $\mathbf{K}'\boldsymbol{\beta}^o + \mathbf{m}'\hat{\mathbf{u}}$, is

$$\mathbf{K}'\mathbf{C}_{11}\mathbf{K} + \mathbf{M}'\mathbf{GZ}'\mathbf{L}(\mathbf{L}'\mathbf{VL})^{-1}\mathbf{H}_s(\mathbf{L}'\mathbf{VL})^{-1}\mathbf{L}'\mathbf{ZGM}. \tag{37}$$

In contrast to BLUP under the $\mathbf{L}'\mathbf{y}$ ($\mathbf{L}'\mathbf{X} = \mathbf{0}$) selection model, minimization of prediction variance is more difficult than minimization of variance of prediction error because the former requires writing a specific $\mathbf{L}$ matrix, and if the variance of the predictor is wanted, an estimate of $Var(\mathbf{L}'\mathbf{y})$ after selection is needed.

We illustrate with the following example with phenotypic observations in two generations under an additively genetic model.

Time	
1	2
y_{11}	y_{24}
y_{12}	y_{25}
y_{13}	y_{26}

The model is

$$\begin{aligned} y_{ij} &= t_i + a_{ij} + e_{ij}. \\ Var(\mathbf{a}) &= \begin{pmatrix} 1 & 0 & 0 & .5 & .5 & 0 \\ & 1 & 0 & 0 & 0 & .5 \\ & & 1 & 0 & 0 & 0 \\ & & & 1 & .25 & 0 \\ & & & & 1 & 0 \\ & & & & & 1 \end{pmatrix}. \end{aligned}$$

This implies that animal 1 is a parent of animals 4 and 5, and animal 2 is a parent of animal 6. Let $Var(\mathbf{e}) = 2\mathbf{I}_6$. Thus $h^2 = 1/3$. We assume that animal 1 was chosen to have 2 progeny because $y_{11} > y_{12}$. Animal 2 was chosen to have 1 progeny and animal 3 none because $y_{12} > y_{13}$. An $\mathbf{L}$ matrix describing this type of selection and resulting in $\mathbf{L}'\mathbf{X} = \mathbf{0}$ is

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

Suppose we want to predict

$$3^{-1} \; (-1 \; -1 \; -1 \; 1 \; 1 \; 1) \; \mathbf{u}.$$

This would be an estimate of the genetic trend in one generation. The mixed model coefficient matrix modified for $\mathbf{L}'\mathbf{y}$ is

$$\begin{pmatrix} 1.5 & 0 & .5 & .5 & .5 & 0 & 0 & 0 & 0 & 0 \\ & 1.5 & 0 & 0 & 0 & .5 & .5 & .5 & 0 & 0 \\ & & 2.1667 & 0 & 0 & -.6667 & -.6667 & 0 & 0 & 0 \\ & & & 1.8333 & 0 & 0 & 0 & -.6667 & 0 & 0 \\ & & & & 1.5 & 0 & 0 & 0 & 0 & 0 \\ & & & & & 1.8333 & 0 & 0 & 0 & 0 \\ & & & & & & 1.8333 & 0 & 0 & 0 \\ & & & & & & & 1.8333 & 0 & 0 \\ & & & & & & & & 6 & -3 \\ & & & & & & & & & 6 \end{pmatrix}.$$

The right hand sides are $\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} & \mathbf{L}'\mathbf{y} \end{pmatrix}'$. Then solving for functions of $\mathbf{y}$ it is found that BLUP of $\mathbf{m}'\mathbf{u}$ is

$$(.05348\ .00208\ -.05556\ .00623\ .00623\ -.01246)\mathbf{y}.$$

In contrast the predictor with minimum variance is

$$[\ .05556\ \ 0\ \ -.05556\ \ 0\ \ 0\ \ 0\ ]\mathbf{y}.$$

This is a strange result in that only 2 of the 6 records are used. The variances of these two predictors are .01921 and .01852 respectively. The difference between these depends upon $\mathbf{H}_s$ relative to $\mathbf{L}'\mathbf{VL}$. When $\mathbf{H}_s = \mathbf{L}'\mathbf{VL}$, the variance is .01852.

As a matter of interest suppose that $\mathbf{t}$ is known and we predict using $\mathbf{y} - \mathbf{X}\boldsymbol{\beta}$. Then the BLUP predictor is

$$(-.02703 - .06667 - .11111 .08108 .08108 .08108)(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

with variance $= .09980$. Note that the variance is larger than when $\mathbf{X}\beta$ is unknown. This is a consequence of the result that in both BLUP and in selection index the more information available the smaller is prediction error variance and the larger is the variance of the predictor. In fact, with perfect information the variance of $(\mathbf{m}'\hat{\mathbf{u}})$ is equal to $Var(\mathbf{m}'\mathbf{u})$ and the variance of $(\mathbf{m}'\mathbf{u} - \mathbf{m}'\hat{\mathbf{u}})$ is 0. The minimum variance predictor is the same when $\mathbf{t}$ is known as when it is unknown. Now we verify that the predictors are unbiased in the selection model described. By the Pearson result for multivariate normality,

$$E(\mathbf{y}_s) = \frac{1}{18}\begin{pmatrix} 12 & 6 \\ -6 & 6 \\ -6 & -12 \\ 2 & 1 \\ 2 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} \mathrm{d}_1 \\ \mathrm{d}_2 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \mathrm{t}_1 \\ \mathrm{t}_2 \end{pmatrix},$$

and

$$E(\mathbf{u}_s) = \frac{1}{18} \begin{pmatrix} 4 & 2 \\ -2 & 2 \\ -2 & -4 \\ 2 & 1 \\ 2 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \mathrm{d}_1 \\ \mathrm{d}_2 \end{pmatrix}.$$

It is easy to verify that all of the predictors described have this same expectation. If $\mathbf{t} = \boldsymbol{\beta}$ were known, a particularly simple unbiased predictor is

$$3^{-1}\ (-1 - 1 - 1\ 1\ 1\ 1)\ (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

But the variance of this predictor is very much larger than the others. The variance is 1.7222 when $\mathbf{H}_s = \mathbf{L}'\mathbf{V}\mathbf{L}$.