

# Appendices to Genetic Evaluation

2006-06-14

**Erling Strandberg and Birgitta Malmfors**

Dept of Animal Breeding and Genetics  
Swedish University of Agricultural Sciences, Uppsala, Sweden

## Contents

Appendix 1. Derivation of selection index equations.....	1
Matrix algebra needed for proof .....	1
Properties of selection index .....	2
Appendix 2. Derivation of variance of index and breeding goal.....	3
Appendix 3. Henderson's mixed model equations to get Best Linear Unbiased Predictions (BLUP) .....	4
Appendix 4. Iterative solutions to the mixed model equations .....	12
The Jacobi and Gauss-Seidel methods.....	12
Example of Jacobi and Gauss-Seidel.....	13
Appendix 5. Setting up the inverse of the relationship matrix directly .....	16
Setting up $A^{-1}$ ignoring inbreeding .....	16
Setting up $A^{-1}$ taking inbreeding into account .....	17
Implication: you get PBVs also for animals without observations .....	17
Appendix 6. Sire models .....	18
Example of sire model .....	19
Sire-maternal grandsire model.....	20
References .....	21

## Appendix 1. Derivation of selection index equations

We will now derive the selection index equations that we have stated to be

$$\mathbf{Pb} = \mathbf{Gv} \quad [1]$$

We start by defining the breeding goal:

$$T = v_1A_1 + v_2A_2 + \dots + v_mA_m = \mathbf{v}'\mathbf{a} \quad [2]$$

where  $\mathbf{v}$  is the vector of economic weights and  $\mathbf{a}$  is the vector of true breeding values with variance-covariance matrix  $\mathbf{C}$ . The index is defined as:

$$\hat{T} = I = b_1X_1 + b_2X_2 + \dots + b_nX_n = \mathbf{b}'\mathbf{x} \quad [3]$$

where  $\mathbf{b}$  is the sought index weights and  $\mathbf{x}$  is the vector of phenotypic measures with variance-covariance matrix  $\mathbf{P}$ . We have also previously defined the matrix  $\mathbf{G}$  which contains the genetic covariances between the measures in the index and the breeding goal traits.

Now, we want to have an index that as closely as possible describes the breeding goal. This can be expressed as that we want to minimize the squared difference between  $I$  and  $T$ ,  $E(T-I)^2$  (similar approach as in least squares). Because both  $T$  and  $I$  have expectations zero, this squared difference is the same as the variance of this difference, which actually is the residual variance from a model  $T = I + e = I + (T-I)$ . So, if we look closer at this squared difference:

$$\begin{aligned} (T-I)^2 &= (\mathbf{v}'\mathbf{a} - \mathbf{b}'\mathbf{x})^2 = \mathbf{v}'\text{var}(\mathbf{a})\mathbf{v} - 2\mathbf{b}'\text{cov}(\mathbf{x},\mathbf{a})\mathbf{v} + \mathbf{b}'\text{var}(\mathbf{x})\mathbf{b} \\ &= \mathbf{v}'\mathbf{Cv} - 2\mathbf{b}'\mathbf{Gv} + \mathbf{b}'\mathbf{Pb} \end{aligned} \quad [4]$$

The way to minimize this expression is to differentiate with respect to the selection index coefficients,  $\mathbf{b}$ , and to set the resulting equation to zero.

$$\frac{\partial (T-I)^2}{\partial \mathbf{b}} = \mathbf{0} - 2\mathbf{Gv} + 2\mathbf{Pb} = \mathbf{0} \quad [5]$$

which after rearrangement leads to the expression we have already described in [1],  $\mathbf{Pb} = \mathbf{Gv}$ .

### Matrix algebra needed for proof

In order to follow the proof just given you need to also know the rules for differentiation of matrices, which actually are very similar to those for scalars. We will just briefly mention the rules used here.

First of all, the derivative of  $\mathbf{v}'\mathbf{Cv}$  with respect to  $\mathbf{b}$  is a vector of zeros, because  $\mathbf{v}'\mathbf{Cv}$  is not a function of  $\mathbf{b}$ . This is the same as when you have scalars.

If we have a multiplication of two vectors  $\mathbf{k}'\mathbf{y}$  and we want to differentiate with respect to  $\mathbf{y}$  then:

$$\frac{\partial(\mathbf{k}'\mathbf{y})}{\partial\mathbf{y}} = \mathbf{k} \quad [6]$$

You recognize this from the differentiation rules for scalars where  $d(ky)/dy = k$ . The only different thing with matrices is that you need to keep in mind the dimensions of the matrices. If you differentiate with respect to a column vector  $\mathbf{y}$  you expect the result to be a vector of the same size and form, i.e.  $\mathbf{k}'$  turns into  $\mathbf{k}$ .

If you have a "quadratic form", e.g.  $\mathbf{y}'\mathbf{A}\mathbf{y}$ , which corresponds to  $Ay^2$  in scalar form, the derivative becomes:

$$\frac{\partial(\mathbf{y}'\mathbf{A}\mathbf{y})}{\partial\mathbf{y}} = 2\mathbf{A}\mathbf{y} \quad [7]$$

which corresponds to the scalar result  $d(Ay^2)/dy = 2Ay$ . Again note that the result  $\mathbf{A}\mathbf{y}$  is a vector of the same form as  $\mathbf{y}$ .

### Properties of selection index

The proof given was set up to give the selection index the property of **minimized expected squared difference** ( $\mathbf{E}(\mathbf{T}-\mathbf{I})^2$ ) between the true and the predicted breeding value, or minimized residual variance from the model  $\mathbf{T} = \mathbf{I} + \mathbf{e}$ . Minimizing the residual variance is the same as maximizing the  $R^2$ -value of that model. The  $R^2$  is identical to the  $(r_{\mathbf{T}\mathbf{I}})^2$ , the squared value of the correlation between the true and the predicted breeding value. So by minimizing we have also **maximized the  $r_{\mathbf{T}\mathbf{I}}$** .

Two other properties that hold for selection index is that it:

- ◆ maximizes the probability of correct ranking between any two individuals
- ◆ maximizes the genetic gain for each round of selection.

## Appendix 2. Derivation of variance of index and breeding goal

You probably already know how to calculate the variance of a variable multiplied by a constant, e.g. the variance of  $kx$ , where  $x$  is a variable with variance  $\sigma_x^2$ , is  $k^2\sigma_x^2$ . Now, if you have vectors to deal with instead this is how it works.

Let's look at the vector multiplication  $\mathbf{b}'\mathbf{x}$ , where  $\mathbf{b}'$  is a row vector of index weights (constants) and  $\mathbf{x}$  is a column vector of phenotypic deviations. We have already defined the variance of  $\mathbf{x}$  to be  $\mathbf{P}$ . Now then, what is the variance of  $\mathbf{b}'\mathbf{x}$ ? From the way it works for scalars (ordinary numbers) one might guess that the answer should be something like  $\mathbf{b}'^2 \text{var}(\mathbf{x})$ , however, that is not correct. The correct answer is:

$$\text{var}(\mathbf{b}'\mathbf{x}) = \mathbf{b}' \text{var}(\mathbf{x})\mathbf{b} = \mathbf{b}'\mathbf{P}\mathbf{b} \quad [8]$$

i.e. instead of squaring the b-values, the b-vectors are on both sides of the variance matrix. Let's take the smallest possible (but yet meaningful) example to see how it works. Assume you have two phenotypic deviations, and thus two index weights:

$$\text{var}(b_1X_1 + b_2X_2)$$

Using the rules for taking variances (see chapter on Statistical Concepts) we get:

$$b_1^2 \text{var}(X_1) + 2b_1b_2 \text{cov}(X_1, X_2) + b_2^2 \text{var}(X_2) \quad [9]$$

Let's then have a look at what  $\mathbf{b}'\mathbf{P}\mathbf{b}$  really means:

$$\begin{bmatrix} b_1 & b_2 \end{bmatrix} \begin{bmatrix} \sigma_{X_1}^2 & \sigma_{X_1, X_2} \\ \sigma_{X_1, X_2} & \sigma_{X_2}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

and if we carry the first multiplication through we get:

$$\begin{bmatrix} b_1\sigma_{X_1}^2 + b_2\sigma_{X_1, X_2} & b_1\sigma_{X_1, X_2} + b_2\sigma_{X_2}^2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

and then finally:

$$b_1^2\sigma_{X_1}^2 + b_1b_2\sigma_{X_1, X_2} + b_1b_2\sigma_{X_1, X_2} + b_2^2\sigma_{X_2}^2$$

Now, this is identical to equation [9] which was what we wanted!

Similarly, we can show that the variance of  $\mathbf{v}'\mathbf{a}$  is  $\mathbf{v}'\mathbf{C}\mathbf{v}$ , where  $\mathbf{C}$  is  $\text{var}(\mathbf{a})$ .

### Appendix 3. Henderson's mixed model equations to get Best Linear Unbiased Predictions (BLUP)

Our aim is to estimate fixed effects and predict breeding values from the model

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

Before we go into the world of Henderson, a few words about **estimation vs. prediction**. The verb "estimate" is used for levels of fixed effects whereas the verb "predict" is used for levels of random effects. Many animal breeders use the term "estimated breeding value", EBV, a term that would probably make a real statistician shudder. Statisticians would argue that a random variable cannot be estimated, because it is not fixed over repeated sampling. What (at least some) statisticians may agree upon is that we make inferences about the realized values of a random variable **given the data**, i.e.  $E(\mathbf{u}|\mathbf{y})$ . To many animal breeders this is quibbling, they are only interested in knowing which animals are best. The advantage of defining animals as random is that we can use prior knowledge about the variance-covariance structure to improve our estimates (sorry, predictions). If we from previous experience know, e.g., the heritability and the relationships among animals, it makes intuitive sense to use this information in order to do a better job in getting the breeding values.

So let's move to the real objective here, what do we want to do?<sup>1</sup> The objective is often expressed as that we want to predict the function:

$$\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u} \tag{10}$$

which is called the *predictand*. This is a way of expressing the problem in such a general way that it becomes almost impossible to understand. What it basically means is that you want to *predict any linear function of the fixed and random effects*. So, before we continue with our quest, we will give some examples of  $\mathbf{K}$  and  $\mathbf{M}$  to clarify this.

First, a very common way of expressing  $\mathbf{K}$  and  $\mathbf{M}$  is to say that we want breeding values for all animals but we do not care about the fixed effects. This would imply that (if there are  $r$  animals in the analysis)  $\mathbf{K}'$  is a  $rxr$  matrix of zeros and  $\mathbf{M}'$  is a  $rxr$  matrix with ones on the diagonal:

$$\begin{bmatrix} 0 & \dots & 0 \\ \cdot & \cdot & \cdot \\ 0 & 0 & 0 \\ \cdot & \cdot & \cdot \\ 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} b_1 \\ \cdot \\ \cdot \\ b_f \end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot \\ \cdot & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & 0 & 1 & 0 \\ 0 & \cdot & \cdot & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_r \end{bmatrix} \tag{11}$$

The end result is a  $rx1$  vector with the breeding values. Another example might be where one has a model where one of the fixed effects contain a genetic level effect, e.g., a breed or country effect. Then one could express the breeding values

for a certain level of that effect. Then  $\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}$  might look something like:

---

<sup>1</sup> The following description is partly based on notes from Larry Schaeffer, e.g. (Schaeffer, 1985)

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_1 \\ \cdot \\ b_f \end{bmatrix} + \begin{bmatrix} 1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot \\ \cdot & 0 & 1 & 0 & \cdot & \cdot \\ \cdot & \cdot & 0 & 1 & 0 & \cdot \\ \cdot & \cdot & \cdot & 0 & 1 & 0 \\ 0 & \cdot & \cdot & \cdot & 0 & 1 \end{bmatrix} \begin{bmatrix} u_1 \\ \cdot \\ \cdot \\ \cdot \\ u_r \end{bmatrix} \quad [12]$$

Now, is this confusing? Don't worry! As we shall see later, the solutions from MME will work regardless of what  $\mathbf{K}$  and  $\mathbf{M}$  we choose, and that is the beauty of it. Think about the opposite situation. If the solutions were dependent on  $\mathbf{K}$  and  $\mathbf{M}$ , we might have to redo the analysis just because we now want to include the fixed effects of breed in our published breeding value. Thankfully, we can do that just by using the estimates of the fixed breed effects directly, as we shall see.

OK, so let's start over again: we want to predict  $\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}$ . To get the best predictor of this we need to know the distribution of the random variable and all moments (the first moment is the mean, the second the variance, but higher order moments exist as well) of the distribution. Then the **best predictor (BP)** given the data is:

**BP: Best Predictor**

$$E(\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u} | \mathbf{y}) \quad [13]$$

which is *best* in the meaning that it has the *smallest mean squared error* (squared difference between predictor and true value) of all predictors.

This overall best predictor may not be possible to get, because we may not know all moments of the distributions. If we restrict the predictor to be a *linear* function of the data, it turns out that we only need to know the mean and the variance of the random variable. Let us assume that the first moment is  $\mathbf{X}\mathbf{b}$  and the second  $\text{Var}(\mathbf{y}) = \mathbf{V}$ , then the **best linear predictor (BLP)** is:

**BLP: Best Linear Predictor**

$$E(\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}) = \mathbf{K}'\mathbf{b} + \mathbf{C}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\mathbf{b}) \quad [14]$$

where

$$\mathbf{C}' = \text{Cov}(\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}, \mathbf{y}) \quad [15]$$

$\mathbf{C}'$  is the covariance between the predictor and the data. The expression  $\mathbf{C}'\mathbf{V}^{-1}$  actually corresponds to a set of ordinary scalar regression coefficients but expressed in matrix terms. The expression describes the regression of the predictor on the data adjusted for fixed effects.

There is a resemblance here to what we did in the selection index theory where the index was calculated as:  $\mathbf{b}'\mathbf{x} = (\mathbf{P}^{-1}\mathbf{G})'\mathbf{x} = \mathbf{G}'\mathbf{P}^{-1}\mathbf{x}$  (if only one trait in goal,  $m=1$ ,  $\mathbf{v}$  disappears,  $\mathbf{G}$  is  $n \times 1$ , and  $\mathbf{x}$  is the vector of phenotypic deviations). (Note by the way that the  $\mathbf{b}$  in the selection index situation (vector of index weights) does not mean the same as the  $\mathbf{b}$  in the mixed model context (vector of fixed effects)). In the table on the next page you can see the similarities between the selection index and the BLP outlined in eq. [14]:

	Selection index	BLP (eq. [14])
Variance-covariance of phenotypic measures	<b>P</b>	<b>V</b>
Covariance between phenotypic measures and what we want to predict (breeding goal)	<b>G</b>	<b>C</b>
Phenotypic measures	<b>x</b>	<b>y-Xb</b>
Predictor	<b>G'P<sup>-1</sup>x</b>	<b>C'V<sup>-1</sup>(y-Xb)</b>

The regression coefficients are then (in both cases) multiplied by the observations adjusted for the mean and the other fixed effects (**y-Xb**). Note that we assume that we know the true values of the fixed effects (**b** in [14]).

**BLUP: Best Linear Unbiased Predictor**

What if we do not know the true values of the fixed effects? Then we have to make sure that the predictor becomes unbiased in some other way. The **best linear unbiased predictor (BLUP)** is:

$$\mathbf{K}\hat{\mathbf{b}} + \mathbf{C}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \quad [16]$$

where:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad [17]$$

The only difference is that we have replaced the true fixed effects by their generalized least-squares estimates, which give an unbiased estimate of the fixed effects.

Now, [16] is not a very convenient expression because **C** seems to be a function of both **K** and **M** (see [15]). Also, the inverse of **V** is needed, and this matrix is of order *N*, the number of observations. This matrix would quickly become impossible to invert even in fairly moderate size animal breeding data. So, we shall try to remove these drawbacks.

We want to find a function of the data, the predictor **L'y**, which predicts the predictand **K'b + M'u** with minimum mean squared error, also making sure that the predictor is unbiased, i.e. that the expectation of the predictor is equal to the expectation of the predictand. This can be expressed in algebra as:

$$E(\mathbf{L}'\mathbf{y}) = \mathbf{L}'\mathbf{X}\mathbf{b} \quad [18]$$

$$E(\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}) = \mathbf{K}'\mathbf{b} \quad [19]$$

because  $E(\mathbf{y})=\mathbf{X}\mathbf{b}$ ,  $E(\mathbf{b})=\mathbf{b}$ , and  $E(\mathbf{u})=\mathbf{0}$ . To get unbiasedness we need:

$$\mathbf{L}'\mathbf{X} = \mathbf{K}' \quad [20]$$

or

$$\mathbf{L}'\mathbf{X} - \mathbf{K}' = \mathbf{0} \quad [21]$$

Note that **L'** is of order  $rxN$  and **K'** is of order  $rxr$ .

Now, the next step is to describe the mean squared error (MSE). Generally,  $MSE = (\text{bias})^2 + (\text{Prediction error variance})$ . However, because we have unbiasedness MSE becomes equal to the prediction error variance.

$$\mathbf{PEV} = \text{Var}(\mathbf{K}'\mathbf{b}+\mathbf{M}'\mathbf{u}-\mathbf{L}'\mathbf{y})$$

$$\begin{aligned}
&= \text{Var}(\mathbf{M}'\mathbf{u}-\mathbf{L}'\mathbf{y}) \\
&= \mathbf{M}'\text{Var}(\mathbf{u})\mathbf{M} + \mathbf{L}'\text{Var}(\mathbf{y})\mathbf{L} - \mathbf{M}'\text{Cov}(\mathbf{u},\mathbf{y})\mathbf{L} - \mathbf{L}'\text{Cov}(\mathbf{y},\mathbf{u})\mathbf{M} \\
&= \mathbf{M}'\mathbf{G}\mathbf{M} + \mathbf{L}'\mathbf{V}\mathbf{L} - \mathbf{M}'\mathbf{G}\mathbf{Z}'\mathbf{L} - \mathbf{L}'\mathbf{Z}\mathbf{G}\mathbf{M}
\end{aligned} \tag{22}$$

Note that **PEV** is a  $r \times r$  matrix, containing prediction error variances on the diagonal for each breeding value, and prediction error covariances on the off-diagonal positions (describing how the prediction error of one breeding value covaries with that of another breeding value).

There are some results hidden in [22] that need be explained before we go on. When we defined the model we should also have defined the expectation and variance of all components of the model. We have defined:

$$\text{Var} \begin{pmatrix} \mathbf{u} \\ \mathbf{e} \end{pmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \tag{23}$$

and we had already defined  $\text{Var}(\mathbf{y}) = \mathbf{V}$ . With the definitions in [23] we can calculate

$$\begin{aligned}
\text{Var}(\mathbf{y}) = \mathbf{V} &= \text{Var}(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}) = \text{Var}(\mathbf{Z}\mathbf{u} + \mathbf{e}) \\
&= \mathbf{Z}\text{Var}(\mathbf{u})\mathbf{Z}' + \text{Var}(\mathbf{e}) + \mathbf{Z}\text{Cov}(\mathbf{u},\mathbf{e}) + \text{Cov}(\mathbf{e},\mathbf{u})\mathbf{Z}' \\
&= \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} + \mathbf{0} + \mathbf{0}
\end{aligned}$$

because the covariance between  $\mathbf{u}$  and  $\mathbf{e}$  is zero (from [23]). Similarly, the covariance between  $\mathbf{y}$  and  $\mathbf{u}$  becomes:

$$\text{Cov}(\mathbf{y},\mathbf{u}) = \text{Cov}(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e},\mathbf{u}) = \mathbf{Z}\text{Var}(\mathbf{u}) = \mathbf{Z}\mathbf{G} \tag{24}$$

$$\text{Cov}(\mathbf{u},\mathbf{y}) = (\mathbf{Z}\mathbf{G})' = \mathbf{G}'\mathbf{Z}' = \mathbf{G}\mathbf{Z}' \tag{25}$$

Now we want to minimize the **PEV** still keeping the unbiasedness. This we can do by adding a relation containing so-called LaGrange multipliers which force  $(\mathbf{L}'\mathbf{X}-\mathbf{K}')$  to be zero:

$$\mathbf{F} = \mathbf{PEV} + (\mathbf{L}'\mathbf{X}-\mathbf{K}')\mathbf{\Phi} \tag{26}$$

Note that because **PEV** is a  $r \times r$  matrix and  $(\mathbf{L}'\mathbf{X}-\mathbf{K}')$  is a  $r \times f$  matrix,  $\mathbf{\Phi}$  must be a  $f \times r$  matrix, otherwise they won't add properly.

So, now we want to find  $\mathbf{L}$  and  $\mathbf{\Phi}$  such that  $\mathbf{F}$  is minimized. This minimization is done by differentiating  $\mathbf{F}$  with respect to both  $\mathbf{L}$  and  $\mathbf{\Phi}$  and setting them to zero:

$$\frac{\partial \mathbf{F}}{\partial \mathbf{L}} = 2\mathbf{V}\mathbf{L} - 2\mathbf{Z}\mathbf{G}\mathbf{M} + \mathbf{X}\mathbf{\Phi} = \mathbf{0} \tag{27}$$

$$\frac{\partial \mathbf{F}}{\partial \mathbf{\Phi}} = \mathbf{X}'\mathbf{L} - \mathbf{K} = \mathbf{0} \tag{28}$$

Because  $\mathbf{L}$  is of order  $N \times r$  the resulting matrix of [27] is also  $N \times r$ , and the order of [28] is  $f \times r$ .

To simplify calculations later we will set  $\mathbf{\theta} = 0.5\mathbf{\Phi}$ , then the first derivative [27] can be written as:



$$\mathbf{VL} = \mathbf{ZGM} - \mathbf{X}\boldsymbol{\theta} \quad [29]$$

We then solve for  $\mathbf{L}$  as

$$\mathbf{V}^{-1}\mathbf{VL} = \mathbf{L} = \mathbf{V}^{-1}\mathbf{ZGM} - \mathbf{V}^{-1}\mathbf{X}\boldsymbol{\theta} \quad [30]$$

Substituting this for  $\mathbf{L}$  into the second derivative [28], then we can solve for  $\boldsymbol{\theta}$  as follows:

$$\begin{aligned} \mathbf{X}'(\mathbf{V}^{-1}\mathbf{ZGM} - \mathbf{V}^{-1}\mathbf{X}\boldsymbol{\theta}) - \mathbf{K} &= \mathbf{0} \\ \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\theta} &= \mathbf{X}'\mathbf{V}^{-1}\mathbf{ZGM} - \mathbf{K} \\ \boldsymbol{\theta} &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZGM} - \mathbf{K}) \end{aligned} \quad [31]$$

Now, we can set this into the equation for  $\mathbf{L}$  [30]:

$$\begin{aligned} \mathbf{L} &= \mathbf{V}^{-1}\mathbf{ZGM} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZGM} - \mathbf{K}) \\ &= \mathbf{V}^{-1}\mathbf{ZGM} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZGM} + \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{K} \end{aligned} \quad [32]$$

If we transpose this to get  $\mathbf{L}'$ :

$$\mathbf{L}' = \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1} + \mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} - \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} \quad [33]$$

and

$$\mathbf{L}'\mathbf{y} = \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}\mathbf{y} + \mathbf{K}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} - \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad [34]$$

Now, remember that the GLS estimator of  $\mathbf{b}$  is (from [17]):

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$$

We can find this expression in [34]:

$$\begin{aligned} \mathbf{L}'\mathbf{y} &= \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}\mathbf{y} + \mathbf{K}'\boxed{(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}} - \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}\mathbf{X}\boxed{(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}} \\ &= \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}\mathbf{y} + \mathbf{K}'\hat{\mathbf{b}} - \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}\mathbf{X}\hat{\mathbf{b}} \end{aligned} \quad [35]$$

We can also see another common part and simplify even more:

$$\begin{aligned} \mathbf{L}'\mathbf{y} &= \boxed{\mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}}\mathbf{y} + \mathbf{K}'\hat{\mathbf{b}} - \boxed{\mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}}\mathbf{X}\hat{\mathbf{b}} \\ &= \mathbf{K}'\hat{\mathbf{b}} + \mathbf{M}'\mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \end{aligned}$$

which then is the BLUP of  $\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}$ . If we let  $\mathbf{K}' = \mathbf{0}$  and  $\mathbf{M}' = \mathbf{I}$ , then the predictor becomes  $\hat{\mathbf{u}}$  and

$$\mathbf{L}'\mathbf{y} = \hat{\mathbf{u}} = \mathbf{GZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\mathbf{b}}) \quad [36]$$

This means that the predictor of  $\mathbf{K}'\mathbf{b} + \mathbf{M}'\mathbf{u}$  is:

$$[\mathbf{K}' \quad \mathbf{M}'] \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} \quad [37]$$

This is good news in that respect that regardless of what  $\mathbf{K}'$  and  $\mathbf{M}'$  are, we can always use the same estimator  $\hat{\mathbf{b}}$  and predictor  $\hat{\mathbf{u}}$ . However, [35] still contains the inverse of  $\mathbf{V}$ , and this is very large and difficult to invert.

To get rid of  $\mathbf{V}^{-1}$  we have to go back to the equations [29] and [28], which contained the derivatives set to zero:

$$\mathbf{V}\mathbf{L} = \mathbf{ZGM} - \mathbf{X}\theta \quad [29]$$

$$\mathbf{X}'\mathbf{L} - \mathbf{K} = \mathbf{0} \quad [28]$$

and write this in matrix form:

$$\begin{bmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{L} \\ \theta \end{bmatrix} = \begin{bmatrix} \mathbf{ZGM} \\ \mathbf{K} \end{bmatrix} \quad [38]$$

and we replace  $\mathbf{V}$  by  $\mathbf{ZGZ}' + \mathbf{R}$  which makes the first equation become:

$$\begin{aligned} (\mathbf{ZGZ}' + \mathbf{R})\mathbf{L} + \mathbf{X}\theta &= \mathbf{ZGM} \\ &= (\mathbf{ZGZ}' + \mathbf{R})\mathbf{L} + \mathbf{X}\theta = \mathbf{ZGM} \end{aligned} \quad [39]$$

$$\begin{aligned} \Rightarrow \mathbf{RL} + \mathbf{ZG}(\mathbf{Z}'\mathbf{L} - \mathbf{M}) + \mathbf{X}\theta &= \mathbf{0} \\ &= \mathbf{RL} + \mathbf{ZS} + \mathbf{X}\theta = \mathbf{0} \end{aligned} \quad [40]$$

In the last step we just set  $\mathbf{S} = \mathbf{G}(\mathbf{Z}'\mathbf{L} - \mathbf{M})$ . Now we solve for  $\mathbf{M}$

$$\mathbf{M} = \mathbf{Z}'\mathbf{L} - \mathbf{G}^{-1}\mathbf{S} \quad [41]$$

and we have a new equation system:

$$\begin{bmatrix} \mathbf{R} & \mathbf{X} & \mathbf{Z} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} \\ \mathbf{Z}' & \mathbf{0} & \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{L} \\ \theta \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{K} \\ \mathbf{M} \end{bmatrix} \quad [42]$$

Now, we solve for  $\mathbf{L}$  in the first equation:

$$\mathbf{RL} + \mathbf{X}\theta + \mathbf{ZS} = \mathbf{0} \quad [43]$$

$$\mathbf{L} = -\mathbf{R}^{-1}\mathbf{X}\theta - \mathbf{R}^{-1}\mathbf{ZS} \quad [44]$$

Note, that we can write  $\mathbf{L}'$  as:

$$\mathbf{L}' = -\theta'\mathbf{X}'\mathbf{R}^{-1} - \mathbf{S}'\mathbf{Z}'\mathbf{R}^{-1} \quad [45]$$

or expressed in matrix algebra:

$$\mathbf{L}' = -\begin{bmatrix} \theta' & \mathbf{S}' \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1} \\ \mathbf{Z}'\mathbf{R}^{-1} \end{bmatrix} \quad [46]$$

and we put that into the two remaining equations:

$$-\mathbf{X}'\mathbf{R}^{-1}\mathbf{X}\theta - \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{S} = \mathbf{K} \quad [47]$$

$$-\mathbf{Z}'\mathbf{R}^{-1}\mathbf{X}\theta - \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}\mathbf{S} - \mathbf{G}^{-1}\mathbf{S} = \mathbf{M} \quad [48]$$

which gives in matrix form again:

$$-\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \theta \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} \mathbf{K} \\ \mathbf{M} \end{bmatrix} \quad [49]$$

Now, we can find a solution to this equation system as:

$$\begin{bmatrix} \theta \\ \hat{\mathbf{S}} \end{bmatrix} = -\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{K} \\ \mathbf{M} \end{bmatrix} \quad [50]$$

or if we take the transpose:

$$\begin{bmatrix} \theta' & \mathbf{S}' \end{bmatrix} = -\begin{bmatrix} \mathbf{K}' & \mathbf{M}' \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \quad [51]$$

Now, we put [56] together with [61]:

$$\begin{aligned} \mathbf{L}'\mathbf{y} &= -\begin{bmatrix} \theta' & \mathbf{S}' \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1} \\ \mathbf{Z}'\mathbf{R}^{-1} \end{bmatrix} \mathbf{y} \\ &= \begin{bmatrix} \mathbf{K}' & \mathbf{M}' \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \end{aligned} \quad [52]$$

We can see that this equation system contains a part  $[\mathbf{K}' \ \mathbf{M}']$  times another part that looks like a solution to a equations system: an inverse times a right-hand side. If we call this solution  $\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix}$ , it would come from an equation system looking like:

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} \quad [53]$$

and it means that the predictor of  $\mathbf{K}'\hat{\mathbf{b}} + \mathbf{M}'\hat{\mathbf{u}}$  is  $\mathbf{K}'\hat{\mathbf{b}} + \mathbf{M}'\hat{\mathbf{u}}$ , where  $\hat{\mathbf{b}}$  and  $\hat{\mathbf{u}}$  come from [53].

Equation [53] is called **Henderson's Mixed Model Equations** or **MME**. Now, we need to take the inverse of the left-hand side coefficient matrix to get the solution. This matrix is of order  $(f+r) \times (f+r)$ , which is smaller than the number of

observations. However, you might argue that  $\mathbf{R}^{-1}$  is still of the same order as number of observations, and now you need that inverse instead of  $\mathbf{V}^{-1}$ . The good news is that we can usually assume that  $\mathbf{R}$  is diagonal, i.e. can be written  $\mathbf{I}\sigma_e^2$ , and is therefore easily inverted. This was actually what we assumed in the main text, and that is why those equations looked even simpler. We will do this simplification here too.

Assume that  $\mathbf{R} = \mathbf{I}\sigma_e^2$ , and thus  $\mathbf{R}^{-1} = \mathbf{I}\frac{1}{\sigma_e^2}$ . Also assume that  $\mathbf{G} = \mathbf{A}\sigma_u^2$  and therefore  $\mathbf{G}^{-1} = \mathbf{A}^{-1}\frac{1}{\sigma_u^2}$ . Then [53] becomes:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X}\frac{1}{\sigma_e^2} & \mathbf{X}'\mathbf{Z}\frac{1}{\sigma_e^2} \\ \mathbf{Z}'\mathbf{X}\frac{1}{\sigma_e^2} & \mathbf{Z}'\mathbf{Z}\frac{1}{\sigma_e^2} + \mathbf{A}^{-1}\frac{1}{\sigma_u^2} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y}\frac{1}{\sigma_e^2} \\ \mathbf{Z}'\mathbf{y}\frac{1}{\sigma_e^2} \end{bmatrix} \quad [54]$$

Now, we multiply both sides with  $\sigma_e^2$  and get the equations:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix} \quad [55]$$

where  $\lambda = \frac{\sigma_e^2}{\sigma_u^2}$ . Equation [55] is the same as equation [37] in the main text.

## Appendix 4. Iterative solutions to the mixed model equations

### The Jacobi and Gauss-Seidel methods<sup>2</sup>

One of the simplest iterative methods used to solve linear equation systems is the Jacobi method and the slight variation called Gauss-Seidel. Let's try to depict the methods in a graphical way. The mixed model equations can be described as below where the coefficient matrix is shaded, **b** is the vector of unknown solutions, and the right-hand side is abbreviated rhs:

$$\begin{array}{|c|} \hline \text{shaded matrix} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \mathbf{b} \\ \hline \end{array} = \begin{array}{|c|} \hline \mathbf{r} \\ \hline \mathbf{h} \\ \hline \mathbf{s} \\ \hline \end{array}$$

Now, we can split the left-hand side coefficient matrix (**C**) up into two parts, the off-diagonal and the diagonal parts:

$$\begin{array}{|c|} \hline \text{shaded matrix with white diagonal} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \mathbf{b} \\ \hline \end{array} + \begin{array}{|c|} \hline \text{shaded diagonal matrix} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \mathbf{b} \\ \hline \end{array} = \begin{array}{|c|} \hline \mathbf{r} \\ \hline \mathbf{h} \\ \hline \mathbf{s} \\ \hline \end{array}$$

where the white space in the coefficient matrix indicates zeroes. Now we can move the left-most part to the right-hand side of the equation:

$$\begin{array}{|c|} \hline \text{shaded diagonal matrix} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \mathbf{b} \\ \hline \end{array} = \begin{array}{|c|} \hline \mathbf{r} \\ \hline \mathbf{h} \\ \hline \mathbf{s} \\ \hline \end{array} - \begin{array}{|c|} \hline \text{shaded matrix with white diagonal} \\ \hline \end{array} \cdot \begin{array}{|c|} \hline \mathbf{b} \\ \hline \end{array}$$

<sup>2</sup> Information in this appendix is based on several sources, e.g. (Misztal, 1999)

In the next step we divide each element of the right-hand side (which is a vector) with the corresponding diagonal element:

$$\mathbf{b} = \left( \begin{matrix} \mathbf{r} \\ \mathbf{h} \\ \mathbf{s} \end{matrix} - \begin{matrix} \text{---} & \text{---} & \text{---} & \text{---} \\ & \text{---} & \text{---} & \text{---} \\ & & \text{---} & \text{---} \\ & & & \text{---} \end{matrix} \mathbf{b} \right) / \begin{matrix} \text{---} & & & \\ & \text{---} & & \\ & & \text{---} & \\ & & & \text{---} \end{matrix}$$

and we have still not done anything illegal. Now comes the trick that makes this into an iterative equation system: we let the  $\mathbf{b}$  on the right-hand side contain the “old” solutions and the  $\mathbf{b}$  on the left-hand side the “new” solutions. For the first iteration we start out with either some “guesses” of solution (e.g. solutions for breeding values from a previous run) or just zeroes. After a certain number of iterations the difference ( $\mathbf{b}_{\text{new}} - \mathbf{b}_{\text{old}}$ ) will be very small and the system is said to have converged.

### *Example of Jacobi and Gauss-Seidel*

Assume we have a model with two fixed effects,  $a$  and  $b$ , each with two levels. The data (10 observations) and the structure has led to the following equation system (in matrix terms it can be summarized as  $\mathbf{X}'\mathbf{X} \mathbf{b} = \mathbf{X}'\mathbf{y}$  or simpler  $\mathbf{C}\mathbf{b} = \mathbf{rhs}$ ):

$$\begin{bmatrix} 6 & 0 & 1 & 5 \\ 0 & 4 & 2 & 2 \\ 1 & 2 & 3 & 0 \\ 5 & 2 & 0 & 7 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 60 \\ 30 \\ 40 \\ 50 \end{bmatrix}$$

with solutions (if we set  $b_2$  to zero):

$$\begin{bmatrix} 8.6364 \\ 3.4091 \\ 8.1818 \\ 0 \end{bmatrix}$$

Now, if we start solving the equation system using *Jacobi* and we start with zeroes for all values in the  $\mathbf{b}$ -vector we get:

$$\begin{aligned} a_1 &= (60 - 0a_2 - 1b_1 - 5b_2) / 6 = 60 / 6 = 10.0 \\ a_2 &= (30 - 0a_1 - 2b_1 - 2b_2) / 4 = 30 / 4 = 7.5 \\ b_1 &= (40 - 1a_1 - 2a_2 - 0b_2) / 3 = 40 / 3 = 13.333 \\ b_2 &= 0 \end{aligned} \quad [56]$$

What exactly are we doing here? Well, the principle is actually quite simple. If we look at the first equation (the row corresponding to  $a_1$ ), we take the right-hand side (which is the sum of 6 observations=60) and adjust it for the other effects in the model (apart from the 6  $a_1$ , it consisted of one  $b_1$  and 5  $b_2$ ) and then we divide by the number of  $a_1$  in that sum. And that principle holds throughout.

Now, the results from the first round of iteration, does not look much like what we know to be the correct answer. On the other hand we started out with really wild guesses. Let's take it another round:

$$a_1 = (60 - 13.333) / 6 = 46.667 / 6 = 7.777$$

$$a_2 = (30 - 2 \times 13.333) / 4 = 3.3334 / 4 = 0.8333$$

$$b_1 = (40 - 1 \times 10 - 2 \times 7.5) / 3 = 15 / 3 = 5.0$$

$$b_2 = 0$$

Note that we are already fairly close for at least  $a_1$ . Note also that we in the calculation of e.g.  $b_1$  use the value from the previous round for  $a_1$  (10.0) although we have just recently calculated a new value (7.777) for that effect (as you will see later, if we had used the new value it would have been a Gauss-Seidel iteration instead). Now, let's look at the result from some more rounds:

$$\begin{bmatrix} 9.166666667 \\ 5 \\ 10.18518519 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.302469136 \\ 2.407407407 \\ 6.944444444 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.842592593 \\ 4.027777778 \\ 8.96090535 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.506515775 \\ 3.019547325 \\ 7.700617284 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 8.716563786 \\ 3.649691358 \\ 8.484796525 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.585867246 \\ 3.257601738 \\ 7.994684499 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.667552583 \\ 3.50265775 \\ 8.299643093 \\ 0 \end{bmatrix}, \begin{bmatrix} 8.616726151 \\ 3.350178453 \\ 8.109043972 \\ 0 \end{bmatrix}$$

As you see it goes a bit up and down, however, after a while it stabilized and after 20 rounds the results are:

$$\begin{bmatrix} 8.636188968 \\ 3.408566905 \\ 8.181170882 \\ 0 \end{bmatrix}$$

i.e. rather close to the exact values.

The *Gauss-Seidel method* is very similar to Jacobi, the only difference is that as soon a new solution is calculated it is used directly (in Jacobi we waited until all estimates of the round were computed). For Gauss-Seidel the first round of computations (given in eq. [56]) would instead become:

$$a_1 = (60 - 0a_2 - 1b_1 - 5b_2) / 6 = 60 / 6 = 10.0$$

$$a_2 = (30 - 0a_1 - 2b_1 - 2b_2) / 4 = 30 / 4 = 7.5$$

$$b_1 = (40 - 10 - 2 \times 7.5 - 0b_2) / 3 = 15 / 3 = 5$$

$$b_2 = 0$$

... and so on.

In some texts you may find a description of the Jacobi algorithm that looks something like (if we describe the equation system as  $\mathbf{Cb}=\mathbf{rhs}$ ):

$$b_i^{(n+1)} = b_i^{(n)} + \frac{rhs_i - \sum_{j=1, m}^m c_{ij} b_j^n}{c_{ii}} \quad [57]$$

In [57], when  $j=i$ ,  $c_{ii} b_i^n / c_{ii}$  is deducted from the  $rhs_i / c_{ii}$  (even though it shouldn't be) and therefore we add this again. Then this algorithm becomes the same as that described in graphic form previously. The algorithm in [57] is easier to program than that shown in the numeric example above. The same algorithm can be used automatically for Gauss-Seidel, if you have only one array for the solutions, and all solutions from 1 up to  $(i-1)$  are from round  $(n+1)$ , and solutions from  $i$  to  $m$  are from round  $n$ .

One can show that the Gauss-Seidel algorithm is actually an iteration on:

$$\mathbf{b}^{(n+1)} = (\mathbf{L} + \mathbf{D})^{-1} (\mathbf{rhs} - \mathbf{U}\mathbf{b}^n) \quad [58]$$

where  $\mathbf{L}$  and  $\mathbf{U}$  are matrices containing the lower and upper off-diagonals of the coefficient matrix  $\mathbf{C}$ , respectively, and  $\mathbf{D}$  is a matrix containing the diagonals of  $\mathbf{C}$ . The Jacobi algorithm, however, is:

$$\mathbf{b}^{(n+1)} = \mathbf{D}^{-1} (\mathbf{rhs} - (\mathbf{L} + \mathbf{U})\mathbf{b}^n) \quad [59]$$

Because  $(\mathbf{L} + \mathbf{D})^{-1}$  is a better approximation of  $\mathbf{C}^{-1}$  (which is the true inverse) than  $\mathbf{D}^{-1}$  is, in general we would expect Gauss-Seidel to perform better than Jacobi.

There is a variety of the Gauss-Seidel called Successive Overrelaxation (SOR) which may give faster convergence:

$$b_i^{(n+1)} = b_i^{(n)} + k \frac{rhs_i - \sum_{j=1, m}^m c_{ij} b_j^n}{c_{ii}} \quad [60]$$

where  $k$  is the relaxation factor (a value between 1 and 2 is usually chosen).



## Appendix 5. Setting up the inverse of the relationship matrix directly

One problem with including the relationship matrix  $\mathbf{A}$  in mixed linear models is that it can become very large, especially in animal models, where each individual is included. Even if it is possible to set up  $\mathbf{A}$  using the tabular method described earlier, the problem comes when you have to invert it. The good news is that there is a method of getting  $\mathbf{A}^{-1}$  directly from a pedigree list of the individuals. For this method to work you need to include all animals in the vector of breeding values ( $\mathbf{a}$ ), also those that have no information but give rise to the relationships, the so-called base population animals. There are ways of setting up  $\mathbf{A}^{-1}$  directly both when ignoring inbreeding and taking inbreeding into account, but the former method is much simpler.

### Setting up $\mathbf{A}^{-1}$ ignoring inbreeding

For the original description see (Henderson, 1975a, 1975b, Henderson, 1976).

The method works as follows. We assume that we have an animal model and that we for each animal have a list with its sire and dam (0 if missing). For simplicity, we will also assume that all individuals have identity number going from 1 up to number of animals,  $n$  (if that is not the case from the beginning they can be re-coded). This means that  $\mathbf{A}$  is of size  $n \times n$ . Note that  $n$  includes the base population animals.

Read the pedigree list one individual at a time. For each individual add values to the positions of  $\mathbf{A}^{-1}$  according to the rules in Table 1. When all animals have been processed the  $\mathbf{A}^{-1}$  is completed.

**Table 1.** Coefficients to be added to  $\mathbf{A}^{-1}$  in an animal model with pedigree information on sire and dam (non-inbred animals)

Identity known for:	Number to be added to position (i=individual)
both sire (s) and dam (d)	2 to (i,i) -1 to (i,s), (s,i), (i,d) and (d,i) 1/2 to (s,s), (d,d), (s,d) and (d,s)
only one parent (p)	4/3 to (i,i) -2/3 to (i,p) and (p,i) 1/3 to (p,p)
none of the parents	1 to (i,i)

The method can also be used for a sire model, where the pedigree information that can be included only comes from the male side (i.e. sire of sire and maternal grand sire of sire) but then the rules in Table 2 must be used instead. Otherwise the procedure is identical.

**Table 2.** Coefficients to be added to  $\mathbf{A}^{-1}$  in a sire model with pedigree information on sire of sire and maternal grandsire of sire (non-inbred animals)

Identity known for:	Number to be added to position (i=individual)
both sire (s) and maternal grandsire (mgs)	16/11 to (i,i) -8/11 to (i,s), (s,i), 4/11 to (s,s) -4/11 to (i,mgs) and (mgs,i) 2/11 to (s,mgs) and (mgs,s) 1/11 to (mgs,mgs)
sire (s) but not mgs	4/3 to (i,i) -2/3 to (i,s) and (s,i) 1/3 to (s,s)
mgs but not sire	16/15 to (i,i) -4/15 to (i,mgs) and (mgs,i) 1/15 to (mgs,mgs)
none of the parents	1 to (i,i)

### Setting up $\mathbf{A}^{-1}$ taking inbreeding into account

There is also a method for setting up  $\mathbf{A}^{-1}$  directly when the animals (parents) are inbred but then animals have to be listed in time order. For more details see (Quaas, 1976).

### Implication: you get PBVs also for animals without observations

The way of setting up  $\mathbf{A}^{-1}$  directly indicates that it is possible to get breeding values also for animals that do not have any phenotypic information of their own. That is exactly what happens for the base populations animals, those animals that are included so that the  $\mathbf{A}^{-1}$  can be set up directly. On the diagonal of the OLS coefficient matrix (left hand side) for such an animal there is just a zero. Similarly, on the right hand side ( $\mathbf{Z}'\mathbf{y}$ ) corresponding to such an animal, there is also a zero. However, when the  $\mathbf{A}^{-1}$  is added to create the MME the information from their relatives is included and their predicted breeding values are based on the relatives' phenotypic information.

## Appendix 6. Sire models

In the simplest sire model, we assume that each sire is mated to a number of females and that each female only has one offspring. A typical example of such a situation is the progeny testing of young dairy bulls for milk production. We can write the model as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zs} + \mathbf{e}$$

and the expectations of the random effects are:

$$\mathbf{s} \sim \mathbf{ND}(\mathbf{0}, \mathbf{A}\sigma_s^2)$$

$$\mathbf{e} \sim \mathbf{ND}(\mathbf{0}, \mathbf{I}\sigma_e^2)$$

where

$$\sigma_s^2 = 0.25h^2\sigma_p^2 = 0.25\sigma_A^2$$

$$\sigma_e^2 = (1 - 0.25h^2)\sigma_p^2 = 0.75\sigma_A^2 + \sigma_E^2$$

The MME expressed in matrix language look basically the same as for the animal model:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{Z} \\ \mathbf{Z}'\mathbf{X} & \mathbf{Z}'\mathbf{Z} + \mathbf{A}^{-1}\lambda \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{s} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{Z}'\mathbf{y} \end{bmatrix}$$

but here  $\lambda = \frac{\sigma_e^2}{\sigma_s^2} = \frac{4-h^2}{h^2}$

Let us look a little closer at the contents of the random part  $\mathbf{Z}'\mathbf{Z}$  (the fixed part is exactly the same as for the animal model):

$$\begin{bmatrix} n_{.1} & 0 & . & . & . \\ 0 & n_{.2} & & & . \\ . & & . & & . \\ . & & & . & 0 \\ 0 & . & . & 0 & n_{.s} \end{bmatrix}$$

The diagonal contains the number of daughters for each sire. Correspondingly the right hand side of the random part contains the sum of observations for each sire.

The relationship matrix in a sire model is based on relationships due to common males in the pedigree, i.e. the fathers and the maternal grandfathers of the sires in the model (see Appendix 5).

### Example of sire model

Let's look at a small example to explain how the sire model works. In the table below we have observations of 305-day milk production from daughters of 4 bulls, spread over two herds. The entries in the table are the sum of the observations with number of observations (daughters) in parentheses, together with the progeny average for each sire, and the corresponding rank.

Sire ( $s_j$ )	Herd ( $h_i$ )			Progeny	
	1	2	Sum	average	Rank
1	89 400 (10)	228 500 (25)	317 900 (35)	9 083	3
2	134 550 (15)	45 850 (5)	180 400 (20)	9 020	4
3	91 000 (10)	111 600 (12)	202 600 (22)	9 209	1
4	44 950 (5)	192 990 (21)	237 940 (26)	9 152	2
Sum	359 900 (40)	578 940 (63)	938 840 (103)		

We can write a model as:

$$y_{ijk} = \mu + h_i + s_j + e_{ijk}$$

We assume that the herd is a fixed effect, and that rams are unrelated so:  $\mathbf{s} \sim \mathbf{ND}(\mathbf{0}, \mathbf{I}\sigma_s^2)$  and  $\mathbf{e} \sim \mathbf{ND}(\mathbf{0}, \mathbf{I}\sigma_e^2)$ , where  $\sigma_s^2 = 0.25\sigma_A^2$  and

$$\sigma_e^2 = 0.75\sigma_A^2 + \sigma_E^2. \text{ With an assumed } h^2 \text{ of } 0.25, \lambda = \frac{4-h^2}{h^2} = 15.$$

After reparameterization where the total mean effect  $\mu$  is set to zero we get MME:

$$\begin{bmatrix} 40 & 0 & 10 & 15 & 10 & 5 \\ 0 & 63 & 25 & 5 & 12 & 21 \\ 10 & 25 & 35+15 & 0 & 0 & 0 \\ 15 & 5 & 0 & 20+15 & 0 & 0 \\ 10 & 12 & 0 & 0 & 22+15 & 0 \\ 5 & 21 & 0 & 0 & 0 & 26+15 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \end{bmatrix} = \begin{bmatrix} 359900 \\ 578940 \\ 317900 \\ 180400 \\ 202600 \\ 237940 \end{bmatrix}$$

The solutions from multiplying the right hand side by the inverse of the coefficient matrix become:

$$\begin{bmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \\ \hat{s}_4 \end{bmatrix} = \begin{bmatrix} 8998.97 \\ 9196.64 \\ -40.11 \\ -16.22 \\ 60.83 \\ -4.49 \end{bmatrix}$$

Again, note that the sum of the four sire effects is zero. To get breeding values, the sire effects are doubled. Note that there is a reranking compared with the ranking based on progeny averages – sire 1 is now ranked as number 4 and has changed places with sire 2. The reason for this is most likely that sire 1 had an unfair advantage when looking at the progeny averages, because he had many more daughters in the good herd (number 2), whereas the opposite was true for sire 2. (Now, of course before estimating the herd effects, we did not really know which herd was the best.) When we adjust for that fact (by applying the model including a herd effect), we get a better description of the true breeding value.

### Sire-maternal grandsire model

The sire model only accounts for the inheritance from the father, not from the mother. The implicit assumption is that all offspring of a sire are from different dams and that each sire is mated to a random sample of dams. For progeny testing of dairy bulls this is commonly a reasonable assumption. However, by including also the maternal grandsire in the model, one can, at least partly, take the inheritance from the maternal side into account. This model (with a herd effect included) could be written as:

$$y_{ijkl} = \mu + h_i + s_j + \frac{1}{2}s_k + e_{ijkl} \quad [61]$$

Note, that there is only one vector of sire effects, i.e. the effect of a certain sire is the same whether it is the father of the offspring or the maternal grandsire (but in the latter case it is halved because it is removed one generation). This type of (additive genetic) sire-maternal grandsire model should not be confused with a sire-maternal grandsire model when one assumes that there are maternal effects. Then there will be one vector of sire effects and one vector of maternal grandsire effects and they will have different genetic interpretation.

In [61] the residual variance has expectation  $\sigma_e^2 = (1 - \frac{1}{4} - \frac{1}{16})\sigma_A^2 + \sigma_E^2$

(instead of  $\sigma_e^2 = (1 - \frac{1}{4})\sigma_A^2 + \sigma_E^2$  in the sire model).

## References

- Henderson, C. R. 1975a. Inverse of a matrix of relationships due to sires and maternal grandsires. *J. Dairy Sci.* 58:1917-1921.
- Henderson, C. R. 1975b. Rapid method for computing the inverse of a relationship matrix. *J. Dairy Sci.* 58:1727-1730.
- Henderson, C. R. 1976. A simple method for computing the inverse a numerator relationship matrix used in prediction of breeding values. *Biometrics.* 32:69-83.
- Misztal, I. 1999. Advanced scientific computing with Fortran 90. Application to animal breeding and beyond. Course material, December 13-17, Gembloux, Belgium.
- Quaas, R. L. 1976. Computing the diagonal elements of a large numerator relationship matrix. *Biometrics.* 32:949-953.
- Schaeffer, L. R. 1985. Course notes on "Estimation of Breeding Values and Population Parameters. April 15-26, 1985. Asker, Norway. Mimeograph.