Chapter 11
Best Linear Unbiased Prediction
- Animal Models-

Julius Van der Werf

Key Terms and Concepts

Fixed and random effects
Mixed model equations
Numerator relationship matrix

Introduction

Before going to the full mixed model we take one intermediate step and consider a random model first, with observations only affected by animals' breeding values. This will show how random effects are estimated in linear models.

In contrast to fixed effects, which are estimated as (differences between) corrected means, random effects are somewhat regressed towards the mean, the same principle as when we estimate breeding values and assign only a part of a phenotypic difference toward the breeding value.
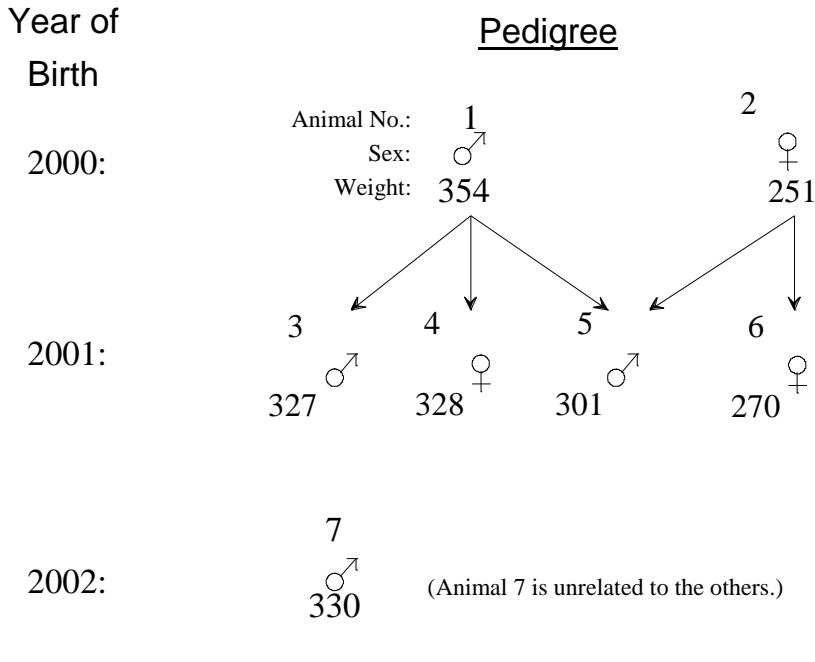
The other aspect about random effects is that they can be correlated to each other, e.g. two breeding values are correlated if the animals have a additive genetic relationship. This correlation can be taken into account using a matrix with additive genetic relationships between animals. We will show that using such a matrix results in using information from relatives in the estimation procedure.

In the second part of this lecture, we present a mixed model, where random and fixed effects are jointly fitted and estimated. The solutions of the mixed model are BLUP EBV for individual animals, and an example will demonstrate that these solutions make sense based on what we have learned so far.

Example of a Random Model

We will fit a random model with observations only affected by animals' breeding values. A model with only random effects is strictly not possible as observations are always affected by at least one fixed effect (the mean), but we will consider observations as deviations from the mean (the mean is assumed known).

We will use the same example as in the previous topic, but additionally, we also consider a pedigree structure among the animals, as some of them are genetically related.

Year of
Birth

<u>Pedigree</u>

2000:

Animal No.: 1   2
Sex: ♂   ♀
Weight: 354   251

2001:

3   4   5   6
♂   ♀   ♂   ♀
327   328   301   270

2002:

7
♂
330   (Animal 7 is unrelated to the others.)

Using the same example data set as in the previous topic:

$$y \quad = \quad Z \quad u \quad + \quad e$$

$$\begin{pmatrix} 354 \\ 251 \\ 327 \\ 328 \\ 301 \\ 270 \\ 330 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{pmatrix}$$

7 x 1   =   7 x 7   7 x 1   +   7 x 1

The Z matrix contains elements which relate the 7 observations (in the rows) to the 7 breeding values (in the columns). The Z matrix is a design matrix, like the X matrix in the fixed model. In this example, each animal in u has exactly one observation (no repeat measures or missing data), and Z is simply an identity matrix. If an animal would have no observations, we would have a column with zero's only (the animal would still be included in

the model, because it may be genetically related to other animals, see later). Animals with more observations would have more "1" values in their column.

The Need to Shrink

Now, how do we find solutions for breeding values in this random model?
If we treated this as a fixed model and used

$$\hat{u} = (Z'Z)^{-1}Z'y \qquad \text{then we would have } \hat{u} = Y$$

... as $(Z'Z)^{-1}Z'$ equals the identity matrix. **This is obviously wrong.** As the model does not fit the mean we should express data as deviations from the average:

$$\hat{u} = (Z'Z)^{-1}Z'(y-\overline{y}) \qquad \text{then we would have } \hat{u} = y - \overline{y}$$

Which is equivalent to $\hat{A} = P$ (P as a deviation) **which we know is still wrong!** P contains effects due to genes and to environment. We penalise for likely 'luck' in the environment by regressing by h², to give $\hat{A} = h^2 P$.

So: We cannot use $\hat{u} = (Z'Z)^{-1}Z'(y - \overline{y})$ because it does not regress or *shrink* the observations to account for luck. Here is a solution:

The 'wrong' single-animal version is: $\hat{A} = \dfrac{V_A}{V_A} P$

(which we can see again is wrong as the correct EBV should be $\hat{A} = h^2 P$

Correct this to regress properly: $\hat{A} = \dfrac{V_A}{V_A + V_E} P \;=\; h^2 P$

Divide top and bottom by V_A: $\hat{A} = \dfrac{1}{1+\frac{V_E}{V_A}} P$

And in a linear model this is: $\hat{u} = (Z'Z + \frac{V_E}{V_A} . I)^{-1} Z'(y - \overline{y})$

Or: $\hat{u} = (Z'Z + \lambda.I)^{-1} Z'(y - \overline{y})$

where I is an identity matrix. Therefore, if we add the variance ratio $\lambda = \frac{V_E}{V_A}$ to the diagonals of Z'Z we achieve that each animal effect is estimated by regressing its deviation towards the mean. Note that λ is smaller for higher values of heritability, therefore the regression will be stronger for lower values of heritability.

To account for relationships a matrix with additive genetic relationships among animals is used. We add the inverse of this matrix, multiplied by λ, to Z'Z. An informal derivation is give below, and was first presented by Dr. C.R. Henderson in the early 1960's (Henderson, 1973). The random model equations are then

Or: $\hat{u} = (Z'Z + \lambda A^{-1})^{-1} Z'(y - \overline{y})$

Where A^{-1} is the inverse numerator relationships matrix (also often called NRM). The term nominator refers to the definition of this relationship. The additive genetic relationship between animals (incl. with themselves) is between 0 and 1 with no inbreeding, and can be between 0 and 2 if inbreeding is accounted for.

Note that BLP is the same as the classical selection index, except that there is a custom set

of index weights for each candidate animal whose breeding value is to be estimated, depending on its particular set of information sources available.

Informal derivation of BLUP equations, and proof of equivalence between selection index and BLUP

$$\hat{u} \;=\; cov(u,y). \; var(y)^{-1} \; y \qquad\qquad selection\ index$$

var(u) = G
var(e) = R
var(y) = var(Zu+e) = ZGZ' + R
cov(u,y)= ZG

$$\hat{u} = \; G \;\; Z' (\; Z G Z' \;+\; R \;)^{-1} \; y$$

dimensions: *a.1 = a.a a.o (o.a a.a a.o   o.o) o1*
                    (**o**bservations and **a**nimals)

Note that $R = I.V_E$.

"Divide" by *G* and rearranging gives:
(note that this is not a formal derivation, there are some missing steps, but it gives the general idea)

$$\hat{u} = (\; Z'Z \;+\; G^{-1} \; I.V_E)^{-1} \; Z' \; y$$

Note that $G = A.V_A$, and $G^{-1} = A^{-1}. 1/V_A$

$$\hat{u} = (\; Z'Z \;+\; A^{-1} \, (\tfrac{V_E}{V_A}) \,)^{-1} \; Z' y \qquad\qquad BLP$$

(BLP = Best Linear Prediction, and stands for BLUP without fixed effects)

The Additive Genetic Relationships Matrix

We want to estimate breeding values as $\hat{u} \;=\; (Z'Z + \lambda A^{-1})^{-1} Z'(y - \overline{y})$, hence we need the inverse of the Numerator Relationships Matrix (A).

First we can calculate A, the Numerator Relationship Matrix for the example. Therefore we can calculate the coefficient of relationship for each pair of animals, then put the results in A. For example, the coefficient of relationship between parent and offspring (e.g. Animals 1 and 3) is 1/2. So $a_{1,3}$ equals 1/2, and so forth to give:

$$A = \begin{array}{ccccccc}
1 & 0 & \tfrac{1}{2} & \tfrac{1}{2} & \tfrac{1}{2} & 0 & 0 \\
0 & 1 & 0 & 0 & \tfrac{1}{2} & \tfrac{1}{2} & 0 \\
\tfrac{1}{2} & 0 & 1 & \tfrac{1}{4} & \tfrac{1}{4} & 0 & 0 \\
\tfrac{1}{2} & 0 & \tfrac{1}{4} & 1 & \tfrac{1}{4} & 0 & 0 \\
\tfrac{1}{2} & \tfrac{1}{2} & \tfrac{1}{4} & \tfrac{1}{4} & 1 & \tfrac{1}{4} & 0 \\
0 & \tfrac{1}{2} & 0 & 0 & \tfrac{1}{4} & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}$$

Note that A is symmetrical: $a_{3,1}$ also equals 1/2.   This way of building A is correct as long as no animal is inbred, but a simple correction is available to correct for inbreeding.

For large data sets, inverting A can be a real problem.

[This problem does not exist for inverting $(Z'Z + \lambda.A^{-1})$, as some computing tricks have been

developed to solve for $\hat{u}$ iteratively, without making the inversion - see later]

Fortunately, Henderson has shown how to 'build' A^{-1} directly. This saves a lot of computing time for doing the inversion (see next page for more detail).

$$A^{-1} = \begin{pmatrix} 13/6 & 1/2 & -2/3 & -2/3 & -1 & 0 & 0 \\ 1/2 & 11/6 & 0 & 0 & -1 & -2/3 & 0 \\ -2/3 & 0 & 4/3 & 0 & 0 & 0 & 0 \\ -2/3 & 0 & 0 & 4/3 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 2 & 0 & 0 \\ 0 & -2/3 & 0 & 0 & 0 & 4/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

The last matrix is the final A^{-1}. Inverting it (on a computer) yields A as originally given!

Summary of basic rules for creating the inverse of the relationships matrix These rules assume no inbreeding, but they can be modified to account for it

	For each animal which is to have an estimate of u, add to A^{-1}		
	Both parents known	One parent known	Neither parent known
Own diagonal	2	4/3	1
Parent x animal	-1	-2/3	
Parents' diagonals	1/2	1/3	
Parent x parent	1/2		

Example of building A^{-1}

Animals 1 and 2 have no known parents so our first 2 steps in building A are:

      Step 1: add a "1" to the (1,1) element of A-1
      Step 2: add a "1" to the (2,2) element of A-1

Animals 3 and 4 have animal 1 as a sire, but dams are unknown:

      Step 3a: add a "4/3" to the (3,3) element of A-1
      Step 3b: add a "1/3" to the (1,1) element of A-1
      Step 3c: add a "-2/3" to the (1,3) and (3,1) elements of A-1
      Step 4a: add a "4/3" to the (4,4) element of A-1
      Step 4b: add a "1/3" to the (1,1) element of A-1
      Step 4c: add a "-2/3" to the (1,4) and (4,1) elements of A-1

$$\begin{pmatrix} 5/3 & 0 & -2/3 & -2/3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -2/3 & 0 & 4/3 & 0 & 0 & 0 & 0 \\ -2/3 & 0 & 0 & 4/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 13/6 & 1/2 & -2/3 & -2/3 & -1 & 0 & 0 \\ 1/2 & 11/6 & 0 & 0 & -1 & -2/3 & 0 \\ -2/3 & 0 & 4/3 & 0 & 0 & 0 & 0 \\ -2/3 & 0 & 0 & 4/3 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 2 & 0 & 0 \\ 0 & -2/3 & 0 & 0 & 0 & 4/3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

      A^{-1} after step 4 A^{-1} *after step 7*

Animal 5 has both parents known: 1 and 2

Step 5a:	add a "2" to the (5,5) element of A-1	
Step 5b:	add a "1/2" to the (1,1) element of A-1	
	add a "1/2" to the (2,2) element of A-1	
	add a "1/2" to the (2,1) and (1,2) elements of A-1	
Step 5c:	add a "-1" to the (1,5) and (5,1) elements of A-1	
	add a "-1" to the (2,5) and (5,2) elements of A-1	

Animal 6 has one dam

Step 6a:	add a "4/3" to the (6,6) element of A-1
Step 6b:	add a "1/3" to the (2,2) element of A-1
Step 6c:	add a "-2/3" to the (2,6) and (6,2) elements of A-1

Animal 7 is unrelated

Step 7	add a "1" to position (7,1)

Solutions to the Random Model

<u>What we want is</u> $\hat{u} = (Z'Z + \lambda A^{-1})^{-1} Z'(y - \bar{y})$

Note that as $Z = I$ in our example (since all animals have 1 record only),
$Z'Z = I$ and $Z'Y = Y$.

Also assuming $h^2 = 0.5$ then $\lambda = \dfrac{V_E}{V_A} = \dfrac{(1-h^2)V_P}{h^2 V_P} = \dfrac{(1-h^2)}{h^2} = 1$

$\hat{u} = (\qquad Z'Z \qquad + \qquad \lambda A^{-1} \qquad)^{-1} \qquad Z'(y - \bar{y})$

and since in our case Z is the identity matrix I

$\hat{u} = (\qquad I \qquad + \qquad \lambda A^{-1})^{-1} \qquad (y - \bar{y})$

Note: $\bar{y} = 308.7$

$$\begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \end{pmatrix} = \left(\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} {}^{13}\!/\!_6 & {}^1\!/\!_2 & -{}^2\!/\!_3 & -{}^2\!/\!_3 & -1 & 0 & 0 \\ {}^1\!/\!_2 & {}^{11}\!/\!_6 & 0 & 0 & -1 & -{}^2\!/\!_3 & 0 \\ -{}^2\!/\!_3 & 0 & {}^4\!/\!_3 & 0 & 0 & 0 & 0 \\ -{}^2\!/\!_3 & 0 & 0 & {}^4\!/\!_3 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 2 & 0 & 0 \\ 0 & -{}^2\!/\!_3 & 0 & 0 & 0 & {}^4\!/\!_3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 45.3 \\ -57.7 \\ 18.3 \\ 19.3 \\ -7.7 \\ -38.7 \\ 21.3 \end{pmatrix}$$

$$\begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \end{pmatrix} = \begin{pmatrix} .410 & 0.030 & .117 & .117 & .127 & -.008 & 0 \\ -.030 & .435 & -.008 & -.008 & .135 & .124 & 0 \\ .117 & -.008 & .462 & .033 & .036 & -.002 & 0 \\ .117 & -.008 & .033 & .462 & .036 & -.002 & 0 \\ .127 & .135 & .036 & .036 & .421 & .039 & 0 \\ -.008 & .124 & -.002 & -.002 & .039 & .464 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & .5 \end{pmatrix} \begin{pmatrix} 45.3 \\ -57.7 \\ 18.3 \\ 19.3 \\ -7.7 \\ -38.7 \\ 21.3 \end{pmatrix}$$

The solution is BLP, as we have ignored fixed effects and just taken deviations from a general mean. BLP is the same as the classical selection index, except that there is a custom set of index weights for each candidate animal whose breeding value is to be estimated.

The result is:
$$
\begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \end{pmatrix} = \begin{pmatrix} 24.0 \\ -32.6 \\ 14.7 \\ 15.3 \\ -5.4 \\ -25.9 \\ 10.6 \end{pmatrix}
$$

Some Comments to the solutions

Refer to the coefficient matrix on the last page, we could call these the index weights. The weights for $\hat{u}_i$ are in the ith row.

1.   Look at animal No. 7 in the example data set, and ignore any effects of year. With no relatives to help him, his EBV is quite simply:

$EBV_7 = h^2(y - \overline{y}) = 0.5(330 - 308.72) = 10.64$.

There are no off-diagonals in the NRM for animal 7. You can see that the 'custom index weights' for animal 7 involve no use of information from other animals, as expected.

2.   Note also that the weights for all other animals make no use of information from animal 7, again as expected.

3.   The diagonals of the index weight matrix have high values - as h² is high (0.5) animals gain most from their own phenotypes. Note that there are diminishing returns (lower diagonal values) from own phenotype as more information from relatives is available.

4.   Animal 1 leans on its three offspring.

5.   For animal 1, there is a negative weight on animal 2's phenotype. This makes sense: E.g. If animal 2 is very good indeed, then any superiority in animal 5 is likely to be due to animal 2 rather than animal 1. So, for a given phenotype of animal 5, the better animal 2 is, the lower animal's breeding value is likely to be.

The Mixed Model
Overview

Fixed Model:	$y = Xb + e$	$\hat{b} = (X'X)^{-1} \quad X'y$
Random Model:	$y = Zu + e$	$\hat{u} = (Z'Z + \lambda A^{-1})^{-1} Z'y$
	- where A = Relationships Matrix and $\lambda = \frac{V_E}{V_A}$	
	We say that $\hat{u}$ is BLP of u. u is the vector of TBV's and $\hat{u}$ is the vector of EBV's	
Mixed Model:	$y \quad = Xb + Zu + e$	
	$= $ fixed effects + breeding values + residual	

Mixed Model Equations:

$$\begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

We say that $\hat{u}$ is BLUP of u.

In this section we combine the two models in the two previous sections: the fixed model and the random model. In mixed models, the breeding values (the random effects) are estimated using the principle of regression, using information from all possible relatives, and correcting them for one, or possibly more fixed effects. We will use again the same example as before, estimating only one fixed effect (the year effect).

The mixed model is a mixture of a fixed and a random model. Both fixed effects (eg. the mean effect and the year effects) and random effects (usually animals' breeding values) **are fitted in the same model and estimated simultaneously in the same analysis.**

Example of a mixed model - following the same data as in the previous sections (with only year effect in the fixed model), just 'add' the two procedures:

$$
\begin{array}{ccccccccc}
y & = & X & b & + & & Z & u & + & e
\end{array}
$$

$$
\begin{pmatrix} 354 \\ 251 \\ 327 \\ 328 \\ 301 \\ 270 \\ 330 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} b_{mean} \\ b_{Y2000} \\ b_{Y2001} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{pmatrix}
$$

$$
\begin{array}{ccccccccc}
7 \times 1 & = & 7 \times 3 & 3 \times 1 & + & & 7 \times 7 & 7 \times 1 & + & 7 \times 1
\end{array}
$$

What does this mean?

For example, observation 1: The model says that 354Kg is made up of 1 dose of bmean, 1 dose of b_{2000}, no dose of b_{2001}, one dose of u_1, no dose of any of u_2 ... u_7, plus whatever is left over undescribed by these effects (e1). The task is to get estimates of bmean, b2000, b2001, and u_1 ... u_7.

Henderson showed that the mixed model equations can be solved like this:

$$\begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

Notice that this is simply related to the equivalent estimates in fixed and random models:

Fixed Model: $\hat{b} = (X'X)^{-1}X'Y$

Random Model: $\hat{u} = (Z'Z + A^{-1}\lambda)^{-1}Z'Y$

The X'Z and Z'X blocks in the coefficient matrix (the matrix to be inverted) provide a connection between the fixed and random effects. If they were full of zeros, the results would be the same as if two separate models had been fitted (one fixed and one random, as in the previous sections).

The values in these blocks let the analysis account for the fact that, for example, progeny

which are very good because of being born in a good year do not overly increase their parents' EBV's. To solve for $\hat{b}$ and $\hat{u}$ we need X'Z and its transpose, Z'X.   In the current example, Z is equal to the identity matrix, and so X'Z = X' and Z'X = X.

$$
\begin{pmatrix} \hat{b} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ Z'y \end{pmatrix}
$$

$$
\begin{pmatrix} \hat{b}_{mean} \\ \hat{b}_{2000} \\ \hat{b}_{2001} \\ \\ \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \end{pmatrix} = \begin{bmatrix} 7 & 1 & 3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 \\ 3 & 1 & 5 & 0 & 0 & 1 & 1 & 1 & 1 & -1 \\ \\ 1 & 1 & 0 & 1\tfrac{9}{6} & \tfrac{1}{2} & -\tfrac{2}{3} & -\tfrac{2}{3} & -1 & 0 & 0 \\ 1 & 1 & 0 & \tfrac{1}{2} & 17/6 & 0 & 0 & -1 & -\tfrac{2}{3} & 0 \\ 1 & 0 & 1 & -\tfrac{2}{3} & 0 & 7/3 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & -\tfrac{2}{3} & 0 & 0 & 7/3 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & -1 & 0 & 0 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 & -\tfrac{2}{3} & 0 & 0 & 0 & \tfrac{7}{3} & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix}^{-1} \begin{pmatrix} 2161 \\ 275 \\ 896 \\ \\ 354 \\ 251 \\ 327 \\ 328 \\ 301 \\ 270 \\ 330 \end{pmatrix} = \begin{pmatrix} 311.94 \\ -9.15 \\ -8.90 \\ \\ 28.26 \\ -28.85 \\ 18.34 \\ 18.77 \\ -0.87 \\ -22.40 \\ 0 \end{pmatrix}
$$

Notice that:

1. The X'X and X'y   blocks are as for the fixed effects analysis.

2. Z'Z + A^{-1} is taken from the random model, with I = 1 as h² = 0.5, but presented already summed.

3. X'Z = X', Z'X = X   and   Z'Y = Y, as Z (and Z') is the identity matrix.

4. Because the mean is fitted, raw data can be used (in Z'y = y) rather than deviations from the overall mean, as used for the random model in the previous topic.

5. Once inverted, the coefficient matrix constitutes a set of custom index weights:

    The elements of first three rows are multiplied by the elements of the (X'y plus Z'Y) vector to give the estimates of the three fixed effects.
    The elements of the last seven rows are the index weights for estimating breeding values. Note that there are index weights in the "Z'X block" which are used to account for fixed effects when calculating EBV's.

    Looking at the result vector, notice that:

    1.   b$_{2002}$ = -(-9.15) - (-8.90) = 18.06 kg

2. Whereas the fixed model estimated 2001 to be 4 kg greater as an effect than 2000, this mixed model puts 2001 only about 0.25 kg ahead - a serious discrepancy of about 3.75 kg. This is because the average EBV of 2001 animals (+3.46 kg) is about 3.75 kg ahead of the average EBV of 2000 animals (-0.295 kg).

    BLUP has determined that the difference observed in the means of 2000 and 2001 is largely due to genetic effects rather than environmental effects. This determination makes some sense when inspecting the pedigree diagram on the first page of the fixed model topic. Animal 1 looks to be quite superior to animal 2 (given the effects fitted in the BLUP model) and has a notably higher EBV. It leaves three offspring in 2001, to animal 2's two offspring - such that 2001 has a better representation of 'good genes' rather than 'bad genes'. The BLUP analysis takes this all into account, with the resulting discrepancy in year effects

between the fixed and mixed models.

3   Animal 7 has a zero EBV. This is because it cannot be fairly compared to any other animal, as they are all born in different years. If animal 7 had had a relative born in a different year, it would have got a non-zero EBV, based on that relative's EBV.

Summary

Best Linear Unbiased Prediction (BLUP) is the name of a method that is used worldwide to give estimated breeding values (EBV's) for commercially important traits.

BLUP uses all available information to estimate an animal's EBV, i.e. information from all genetically related animals, and possibly from correlated traits (if multi-trait BLUP is used). Furthermore, BLUP corrects for fixed effects such as herd, year or season of production etc., it accounts for unequal use of the best sires in different herds, for selection and non-random mating.

BLUP relies on correct knowledge of genetic parameters (heritability, correlations), and on a good data structure.

The principle of BLUP is based on a combination of two techniques:

1)   Selection index, where the phenotypic information about an animal is used to estimate a breeding value by regression.
2)   Linear models as used in statistical analysis to estimate effects such as herd, year, season, age etc. in order to correct data when estimating breeding values.

References

Henderson, C.R. (1973). Sire Evaluation and Genetic Trends. In *Proceedings of the Animal Breeding and Genetics Symposium in Honor of J. L. Lush*. American Society for Animal Science, Blackburgh, Champaign, Illinois, pp10-41.

Simm, G. (2000). Genetic Improvement of Cattle and Sheep. Farming Press, Miller Freeman, UK..