Chapter 6

Prediction Of Breeding Values

The abbreviation `BLUP` stands for **B**est **L**inear **U**nbiased **P**rediction. BLUP-based procedures are the de-facto standard methods to predict breeding values. These procedures use linear mixed effects models to simultaneously estimate fixed effects and predict breeding values. Linear models that contain besides the random error other random effects are called linear mixed effects model. The classification into fixed and random effects is not based on a universally accepted definition. For some effects with distinct and a finite number of classes it is easy to classify them as fixed effects. On the other hand breeding values of animals are always modeled as random effects.

6.1 The Linear Mixed Effects Model

In general the linear mixed effects model has the following structure

$$y = Xb + Zu + e \tag{6.1}$$

where

y	vector of length n with observations
b	vector of length p with fixed effects
u	vector of length q with random breeding values
e	vector of length n with random error terms
X	$n \times p$ incidence matrix linking fixed effects to observations
Z	$n \times q$ incidence matrix linking random breeding values to observations

With a linear mixed effects model, we also have to specify the expected values and the variance-covariance matrices for the random components in the model.

We start with the expected values in vector-notation.

$$E \left[\begin{array}{c} y \\ u \\ e \end{array} \right] = \left[\begin{array}{c} Xb \\ 0 \\ 0 \end{array} \right] \tag{6.2}$$

The variance-covariance matrices can be written as

$$var \left[\begin{array}{c} y \\ u \\ e \end{array} \right] = \left[\begin{array}{rrr} ZGZ^T + R & ZG & 0 \\ GZ^T & G & 0 \\ 0 & 0 & R \end{array} \right] \tag{6.3}$$

6.2 Solutions To The Linear Mixed Effects Model

In the linear mixed effects model (6.1) the vectors b and u are unknown and are of primary interest when using the model in livestock breeding applications. Applying the BLUP properties to the model (6.1) leads to the following solutions for b and u.

$$\hat{u} = GZ^TV^{-1}(y - X\hat{b}) \tag{6.4}$$

where V stands for the variance covariance matrix $var(y)$ of the phenotypic observations. From (6.3), we can see that V can be computed as $V = ZGZ^T + R$. The vector $\hat{b}$ stands for the best linear unbiased estimate of the fixed effects b which are given by

$$\hat{b} = (X^TV^{-1}X)^-X^TV^{-1}y \tag{6.5}$$

where $(X^TV^{-1}X)^-$ stands for a generalized inverse of $(X^TV^{-1}X)$.

Both solutions in (6.4) and (6.5) contain the matrix V^{-1} which has the dimensions $n \times n$ where n stands for the number of observations. In practical breeding programs the number of observations is frequently in the order of 10^6 which makes the computation of V^{-1} practically impossible. CR Henderson and his team found that the solution of the so-called mixed model equations lead to equivalent solutions for $\hat{b}$ and $\hat{u}$. The mixed model equations are given by

$$\left[\begin{array}{lr} X^TR^{-1}X & X^TR^{-1}Z \\ Z^TR^{-1}X & Z^TR^{-1}Z + G^{-1} \end{array} \right] \left[\begin{array}{c} \hat{b} \\ \hat{u} \end{array} \right] = \left[\begin{array}{c} X^TR^{-1}y \\ Z^TR^{-1}y \end{array} \right] \tag{6.6}$$

The application of the linear mixed effects model (6.1) in Livestock Breeding have led to two specially named models depending on what was used as random effects. These models are

1. **sire model**: for each observation the influence of the animals sire was used as random effect.
2. **animal model**: the influence of the genetic component of each animal was used as random effect.

6.3 Sire Model

In the sire model the effect of the sire is taken as a random effect. This leads to the following model

$$y = Xb + Zs + e \tag{6.7}$$

where s is a vector of length q_s with all sire effects. All other components are the same as specified for the model (6.1). The variance-covariance matrix $var(s)$ of all sire effects corresponds to

$$var(s) = A_s * \sigma_s^2 \tag{6.8}$$

where A_s is the numerator relationship matrix only considering relations on the male side of the pedigree. The variance component σ_s^2 is obtained from the variance components estimation using a sire model and it corresponds to $0.25\sigma_a^2$ where σ_a^2 is the genetic additive variance.

The sire model was developed for livestock populations with large half-sib families where a given sire has a large number of offspring. As a result all sires get predicted breeding values. All female animals are not considered in the evaluation and do not get predicted breeding values.

The usage of sire models for predicting breeding values was mainly at the beginning of the introduction of the BLUP methodology. It was motivated by the lower requirements of computing power.

6.4 Animal Model

The so-called animal model uses the breeding value of each animal as a random effect. In matrix-vector notation the animal model has the following structure.

$$y = Xb + Za + e \tag{6.9}$$

where a is a vector of length q_a containing the breeding values of all animals in the pedigree. All other components are the same as specified for the model (6.1). The variance-covariance matrix $var(a)$ of the breeding values is defined as

$$var(a) = A\sigma_a^2 \tag{6.10}$$

where A is the numerator relationship matrix containing all identity by descent relationships in a pedigree. The matrix A is symmetric. The off-diagonal elements $(A)_{ij}$ in row i and column j contain the probability of animals i and j sharing alleles that are identical by descent. Identity by descent of two alleles means that two alleles can be traced back in the pedigree to a common ancestor. As an example the value of $(A)_{ij}$ between two half-sibs i and j is $1/4$ and between two full-sibs it is $1/2$. The diagonal elements $(A)_{ii}$ of matrix A corresponds to $1 + F_i$ where F_i is the inbreeding coefficient of animal i. The inbreeding coefficient corresponds to $1/2$ of the relationship coefficient between parents s and d of i.

6.5 Genomic BLUP

With the rapid increase of animals with genotypic information in the form of SNP-genotypes, this information can be used to predict genomic breeding values. BLUP-based methods can be used to predict genomic breeding values based on genomic information. The genomic BLUP models can have two different forms.

1. marker-effect models where each SNP-locus is taken as a random effect.
2. breeding value based models where genomic breeding values are included as random effects.

As long as the number of genotyped animals is lower than the number of genotypes determined per animal, then model (2) has advantages over model (1) with respect to requirements of computing power. But as the number of genotyped animals increases very rapidly, model (1) might be favorable over model (2).

At this point it is important to note that whether we use genomic information or not, the goal of the prediction of breeding values remains still the same. We want to get a prediction of the genetic potential of each animal. This prediction must be as accurate as possible. The realized values of the predictions are used to rank the animals. Based on these rankings the selection decisions are taken to determine the parents of the future generations of livestock animals.