Partitioning Estimated Breeding Values in Animal Models

G. J. Kistemaker
Canadian Dairy Network, Guelph, Ontario, Canada
kistemaker@cdn.ca

Introduction

The Canadian test day model (CTDM) has been used since February 1999. This model is a 12 trait model with milk, fat, protein and SCS in each of the first three lactations. Breeding values within each trait are described using the 3-parameter Wilmink curve. As a result, this model has 36 genetic parameters for each animal. The 3 parameters for each lactation are used to calculate lactation breeding values and combined breeding values for milk, fat, protein, SCS and lactation persistency. Observations made in first lactation will affect breeding values in later lactations but the weight which is put on an animals own records relative to the parent average and/or progeny contributions have not been calculated. The purpose of this paper is to partition an animal's breeding value with respect to the sources of information (observations, parent average and progeny) which contribute to it and then calculate a relative weight on the data relative to the pedigree contributions. The CTDM is used as an example but the derivation applies to all animal models.

Methods

The MME for the CTDM which is used at CDN can be written as:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z} & \mathbf{X'R^{-1}W} \\ \mathbf{Z'R^{-1}X} & \mathbf{Z'R^{-1}Z + A^{-1} \otimes G^{-1}} & \mathbf{Z'R^{-1}W} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}Z} & \mathbf{W'R^{-1}W + I^{-1} \otimes P^{-1}} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z'R^{-1}y} \\ \mathbf{W'R^{-1}y} \end{bmatrix}$$

In order to partition the breeding values, animals are sorted by birth date. To partition the breeding values of animal i, animals are divided into three groups:

1. All animals older than animal i
2. Animal i
3. All animals younger than animal i

Partition **Z, a** and $\mathbf{A^{-1}}$ using these three groups:

$$\mathbf{Z} = [\,\mathbf{Z_1}\ \ \mathbf{Z_2}\ \ \mathbf{Z_3}\,]$$
$$\mathbf{a} = [\,\mathbf{a_1}\ \ \mathbf{a_2}\ \ \mathbf{a_3}\,]$$
$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{A}^{11} & \mathbf{A}^{12} & \mathbf{A}^{13} \\ \mathbf{A}^{21} & \mathbf{A}^{22} & \mathbf{A}^{23} \\ \mathbf{A}^{31} & \mathbf{A}^{32} & \mathbf{A}^{33} \end{bmatrix}$$

The MME can then be written as:

$$\begin{bmatrix} \mathbf{X'R^{-1}X} & \mathbf{X'R^{-1}Z_1} & \mathbf{X'R^{-1}Z_2} & \mathbf{X'R^{-1}Z_3} & \mathbf{X'R^{-1}W} \\ \mathbf{Z_1'R^{-1}X} & \mathbf{Z_1'R^{-1}Z_1 + A^{11} \otimes G^{-1}} & \mathbf{Z_1'R^{-1}Z_2 + A^{12} \otimes G^{-1}} & \mathbf{Z_1'R^{-1}Z_3 + A^{13} \otimes G^{-1}} & \mathbf{Z_1'R^{-1}W} \\ \mathbf{Z_2'R^{-1}X} & \mathbf{Z_2'R^{-1}Z_1 + A^{21} \otimes G^{-1}} & \mathbf{Z_2'R^{-1}Z_2 + A^{22} \otimes G^{-1}} & \mathbf{Z_2'R^{-1}Z_3 + A^{23} \otimes G^{-1}} & \mathbf{Z_2'R^{-1}W} \\ \mathbf{Z_3'R^{-1}X} & \mathbf{Z_3'R^{-1}Z_1 + A^{31} \otimes G^{-1}} & \mathbf{Z_3'R^{-1}Z_2 + A^{32} \otimes G^{-1}} & \mathbf{Z_3'R^{-1}Z_3 + A^{33} \otimes G^{-1}} & \mathbf{Z_3'R^{-1}W} \\ \mathbf{W'R^{-1}X} & \mathbf{W'R^{-1}Z_1} & \mathbf{W'R^{-1}Z_2} & \mathbf{W'R^{-1}Z_3} & \mathbf{W'R^{-1}W + I^{-1} \otimes P^{-1}} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a_1} \\ \mathbf{a_2} \\ \mathbf{a_3} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{X'R^{-1}y} \\ \mathbf{Z_1'R^{-1}y} \\ \mathbf{Z_2'R^{-1}y} \\ \mathbf{Z_3'R^{-1}y} \\ \mathbf{W'R^{-1}y} \end{bmatrix}$$

Solutions for the animal of interest are in $\mathbf{a_2}$ and therefore we only need:

$$\begin{bmatrix} \mathbf{Z_2'R^{-1}X} & \mathbf{Z_2'R^{-1}Z_1 + A^{21} \otimes G^{-1}} & \mathbf{Z_2'R^{-1}Z_2 + A^{22} \otimes G^{-1}} & \mathbf{Z_2'R^{-1}Z_3 + A^{23} \otimes G^{-1}} & \mathbf{Z_2'R^{-1}W} \end{bmatrix} \begin{bmatrix} \mathbf{b} \\ \mathbf{a_1} \\ \mathbf{a_2} \\ \mathbf{a_3} \\ \mathbf{p} \end{bmatrix} = \begin{bmatrix} \mathbf{Z_2'R^{-1}y} \end{bmatrix}$$

The Z matrices relate each observation to one animal, therefore $\mathbf{Z_2'R^{-1}Z_1}$ and $\mathbf{Z_2'R^{-1}Z_3}$ are null matrices and this equation can be rewritten as:

$$\left[\mathbf{Z_2'R^{-1}Z_2} + \mathbf{A^{22}} \otimes \mathbf{G^{-1}}\right]\mathbf{a_2} = \mathbf{Z_2'R^{-1}}[\mathbf{y} - \mathbf{Xb} - \mathbf{Wp}] - \left[\mathbf{A^{21}} \otimes \mathbf{G^{-1}}\right]\mathbf{a_1} - \left[\mathbf{A^{23}} \otimes \mathbf{G^{-1}}\right]\mathbf{a_3}$$

Then $\mathbf{a_2}$ can be calculated using:

$$\mathbf{a_2} = \mathbf{D^{-1}Z_2'R^{-1}}[\mathbf{y} - \mathbf{Xb} - \mathbf{Wp}] - \mathbf{D^{-1}}\left[\mathbf{A^{21}} \otimes \mathbf{G^{-1}}\right]\mathbf{a_1} - \mathbf{D^{-1}}\left[\mathbf{A^{23}} \otimes \mathbf{G^{-1}}\right]\mathbf{a_3}$$

where $\mathbf{D} = \mathbf{Z_2'R^{-1}Z_2} + A^{22}\mathbf{G^{-1}}$

In this partitioning:

$\mathbf{D^{-1}Z_2'R^{-1}}$ are weights on test day records
$-\mathbf{D^{-1}}\left[\mathbf{A^{21}} \otimes \mathbf{G^{-1}}\right]$ are weights on parameters in $\mathbf{a_1}$
$-\mathbf{D^{-1}}\left[\mathbf{A^{23}} \otimes \mathbf{G^{-1}}\right]$ are weights on parameters in $\mathbf{a_3}$

The three matrices with weights are large but most values are 0. There are non-zero values for test day records on animal i only. Vectors $\mathbf{A^{21}}$ and $\mathbf{A^{23}}$ have non-zero values for the parents, progeny and mates of animal i only. If both parents and all mates are known then the sum of the elements in $-\mathbf{A^{21}}$ and $-\mathbf{A^{23}}$ is equal to A^{22}. The total weight on pedigree contributions can be written as $A^{22}\mathbf{D^{-1}G^{-1}}$ when parents and mates are known. If some parents and/or mates are unknown then this total weight would also include the weights on missing animals which would have estimated breeding values of zero.

Each of the 36 genetic parameters for animal i in the CTDM, and the published EBVs which are calculated from these 36 parameters, may have contributions from all 36 estimated genetic parameters which make up the total pedigree contribution. In addition they have contributions from all observations on animal i for each trait which was observed.

Relative Weight on an Animal's Own Records

These weights by itself do not have any meaning since they are combining parameters which have different variances and are correlated. The main question to answer is the relative weight which is put on the pedigree contribution relative to the test day records. The relative weight on the pedigree contributions for each EBV can be calculated as the covariance between the animals EBV and the pedigree EBV divided by the variance of the pedigree EBV:

W_{ped}= Cov(EBVi,EBVped)/ Var(EBVped)

where: $Var(EBV_{ped}) = \mathbf{W'GW}$
$Cov(EBVi,EBVped) = A^{22}\mathbf{W'D^{-1}W}$
$\mathbf{W}$ = a vector with weights used to combine the estimated parameters into the EBV for the trait of interest (e.g. milk yield in first lactation).

Relative weight on an animal's own data contributions can be calculated as:

$$W_{data} = 1 - W_{ped}$$

The contribution from the parent average can be calculated as:

$$W_{pa} = W_{ped} * -(A^{2sire} + A^{2dam})/A^{22}$$

where: A^{2sire} and A^{2dam} are the elements in $\mathbf{A^{21}}$ corresponding to the sire and dam of animal i.

Relative Weight on Progeny Test Day Records

For bull's the main interest is the weight which is put on the progeny test day records relative to all other sources which contribute to the bulls proof. This weight is affected by the weight on the progeny breeding values relative to the parent average and by the weight which the test day records received when the progeny breeding values were calculated.

The weight on progeny test day records is:

$$W_{progdata} = \frac{W_{prog}\overline{W_{data}}}{1 - W_{prog}\overline{W_{pa}}}$$

where: W_{prog} is the weight on the progeny EBVs when calculating the bull's EBV.

$\overline{W_{data}}$ is the weighted average (using elements in $\mathbf{A}^{23}$ as weights) of W_{data} over all progeny.

$\overline{W_{pa}}$ is the weighted average (using elements in $\mathbf{A}^{23}$ as weights) of W_{pa} over all progeny.

Results

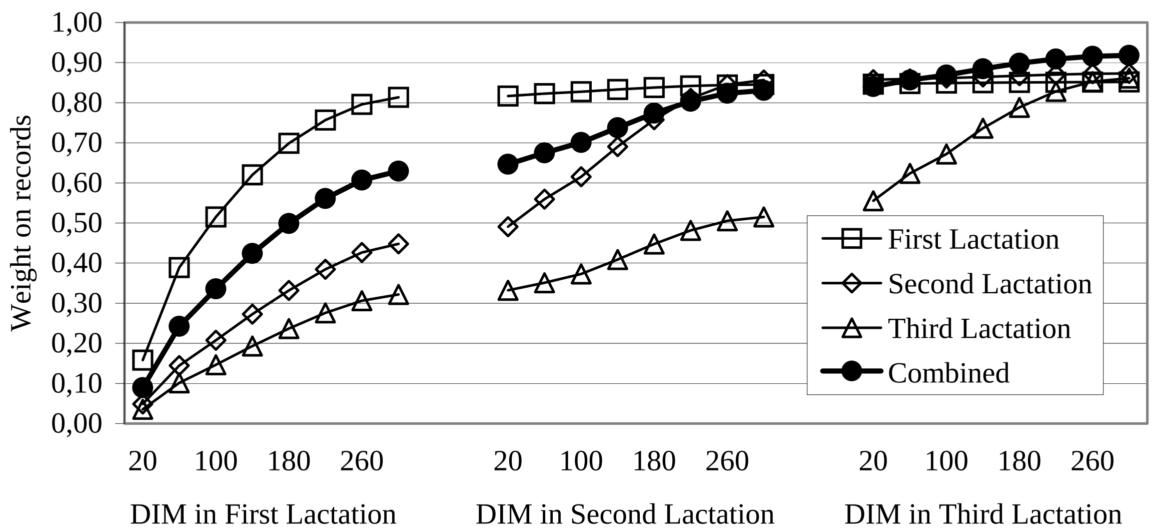
All results presented are for a situation where the parents of all animals are known and are not inbred. It is assumed that milk, fat, protein and SCS are measured at every test. Variances and covariances used are those currently used in the CTDM for the Holstein population.

Relative Weight on an Animal's Own Records

Weights were calculated for a cow which had her first test day record at 20 DIM and had a test every 40 DIM until 300 DIM. This cow was also tested on the same DIM in the second and third lactation. The relative weights on the cows test day records were calculated after each test was added.

Figure 1 shows the weight on test day records for 1st, 2nd and 3rd lactation protein EBVs and for the combined protein EBV.

Figure 1: Weight on Own Records when Calculating Protein EBVs



- When this cow had her first test day record at 20 DIM this test day record received a weight of .15 when estimating 1st lactation protein EBV, a weight of .04 when estimating 3rd lactation protein EBV and a weight of .09 when estimating combined protein EBV.
- After this cow added a second test day record at 60 DIM there was a total weight of .39 on these two test day records when estimating 1st lactation protein EBV, a weight of .10 when estimating 3rd lactation EBV and a weight of .24 when estimating combined EBV.

When this cow completed her first lactation there was a total weight of .81 on the 8 test day records in first lactation when calculating 1st lactation protein EBV and a weight of .31 on her first lactation test day records when calculating the 3rd lactation protein EBV. The data contribution when calculating the combined EBV is .63. The weight on the data when calculating the combined protein EBV increases to .92 when this cow had 8 test day records in all three lactations.

Figure 2 shows the weights on the data when calculating EBVs for lactation persistency (calculated using the EBV for milk yield on 280 DIM relative to EBV for milk yield on 60 DIM). Weights on the data for lactation persistency were lower than weights on the data for protein (Figure 1). Weights on the data increased when cows added test day records around 60 or around 280 DIM but there was not much change when test days records between 100 and 200 DIM were added.

Figure 2: Weight on a Cows Own Records when Calculating Persistency EBVs

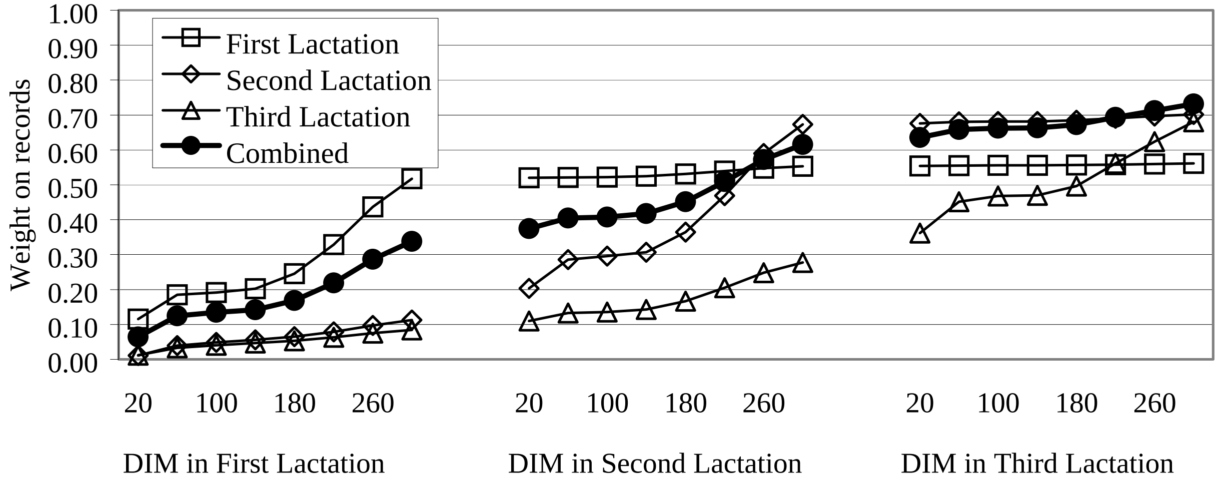
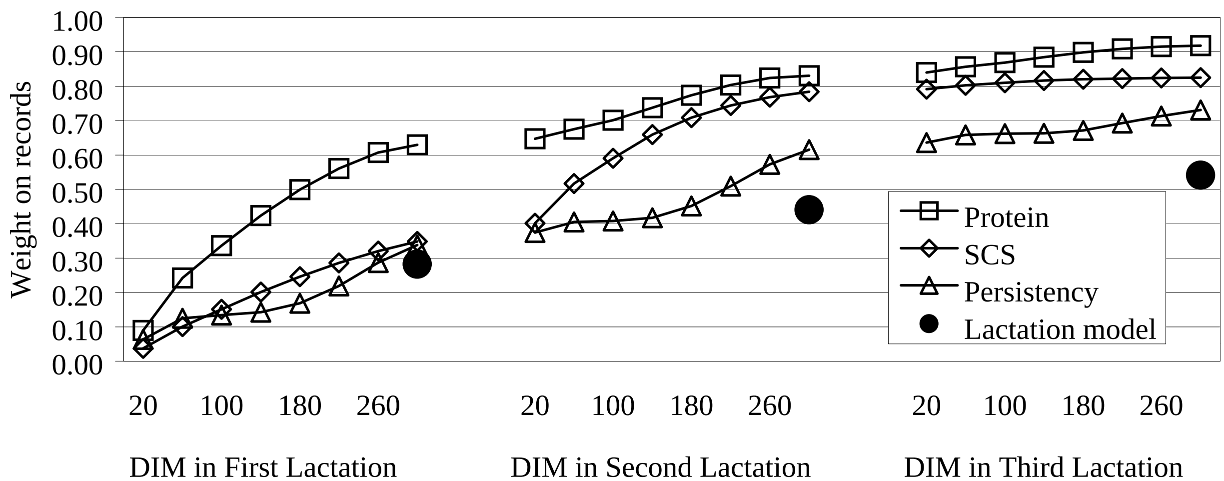


Figure 3: Weight on Own Records when Calculating Combined EBVs



Weights on the data for combined protein, lactation persistency and SCS are shown in figure 3 together with the weight on lactation records in the lactation model. This figure clearly shows that in the CTDM there is a higher weight on the data then in the lactation model which was previously used in Canada. This difference is a combination of the difference in models (test day records vs. lactation records) and a larger ratio of genetic to residual variance in the CTDM compared to the lactation model.

Relative Weight on Progeny Test Day Records

Weights on daughter data records when calculating combined protein EBVs were calculated for bulls with 20, 50, 100 or 1000 daughters (Figure 4). Daughters were assumed to have their first test day record at 20 DIM and added one test day record every 40 DIM. Weights on daughter data records were calculated after each test day record was added. Figure 4 also shows weights on lactation records in the lactation model when daughters have one, two or three lactation records (solid symbols). There was a rapid increase in the weight on test day records when daughters added their first 4 records. As soon as a bull has 1000 daughters with one test day record the weight on these test day records was > .95, when these daughters have 4 test day records the weight was > .99.

Summary

This paper presented a method to separate EBVs in contributions from observations and pedigree. Relative weights on data contributions for animals with their own records and for animals with progeny records were calculated. The method was presented using the Canadian test day model as an example but can be applied to any animal model.

Figure 4: Weight on Progeny Records in Combined Protein Proofs