ESTIMATING VARIANCE COMPONENTS AND BREEDING VALUES IN A DAIRY POPULATION WITH SMALL HERD SIZE

PREISINGER, R., CLAUS, J. and KALM, E.
Institute of Animal Breeding and Animal Husbandry
Christian - Albrechts - University, Kiel, Germany

SUMMARY

Using the principle of the direct sire comparison method (BLUP), sire evaluations and variance components were calculated assuming herd x year x season and herdclass x year x season models for population with small herd size. Herd x year x season will be replaced by a classification containing herd level as an equivalent for single herds to reduce loss of records.
From models with herd classes, one receives biased predictors since certain fixed effects between herds within a given are ignored. The bias of random effects is calculated by using the herd class x year x season model as a reduced herd x year x season model. Estimates for the additive genetic variance will be reduced, since adjustment for the genetic merit of different herds is made inaccessible when using herd class x year x season.

INTRODUCTION

The BLUP procedure has replaced older procedures for dairy sire evaluation, which can be used very differently under various circumstances. BLUP is a principle of estimation, such that the statistical model is in no way completely specified. Models differ in the number of traits and the way they take account of fixed and random effects.
Especially in populations with small herd size, such as in central Europe, the models differ in taking account for herd x year - and herd x year x season effects (HYS).
The objective of this paper is to present the consquences of estimating variance components and ranking sires using alternative models that include HYS - effects in different ways.

ALTERNATIVES FOR USING HERD EFFECTS IN THE MODEL

Including the three - way interaction of herd x year x season in the model proves most useful in obtaining the best estimators. Use of fixed HYS in the model in place of herd x year effects improve the efficiency in removing the bias caused by season effects on sire proofs.
In populations with small herd size, we often will have a record with no contemporary record in a HYS. If herds are treated as fixed, no contribution for estimating breeding values will be obtained from these animals. The same is true if all records in one HYS are made by paternal half - sisters. The effective number of daugthers in response to the actual number of daughters decreases and the variance of prediction error increas.

Another possibility would be to treat HYS as random. But if the choice of sire to be used in different herds based on some prior knowledge or selection occurs, there might be some correlation between the effects of herds and sires. Thus, a model which treats herds as fixed effects to obtain unbiased predictors (HENDERSON, 1975)is necessary.
Investigations by DEMPFLE (1982) showed that it would be useful to replace the herd x year x season classification by a classification containing herd level as an equivalent for single herds. With this herd class x year x season (HCYS) classification, the residual error variance was only slightly higher than when using a herd x year x season classification, whereas the effective daughter number was much higher.
Setting up herd classes would be to treat herds with similar herd x year x season effect as one herd. This suggestion assumes that estimates of herd effects are available. Since they are not available, we have to use the average herd yield to classify the herds. In a model with herd class x year x season instead of HYS, certain fixed effects will be ignored. It is generally known that ignoring certain fixed effects that are actually present in the model leads to biased estimators (HENDERSON, 1975).

CALCULATION OF THE BIAS

A model which includes HYS - effects will definitely be the best one. A model which includes only herd class x year x season (HCYS) represent a reduced one.
A full herd class model without bias would be:

$$\underline{y} = \underline{X}_1\underline{b}_1 + \underline{X}_2\underline{b}_2 + \underline{Zu} + \underline{e} \qquad \text{model 1}$$

$\underline{y}$ = vector of observations

$\underline{b}_1$ = vector of fixed herd class x year x season effects (HCYS)

$\underline{b}_2$ = vector of fixed herd effects within herd class x year x season

$\underline{u}$ = vector of random sire effects

$\underline{e}$ = vector of random residuals

$\underline{X}_1$, $\underline{X}_2$, $\underline{Z}$, represent the corresponding design matrices

Despite the inclusion of herd classes, for every HYS in model 1 estimators will be obtained.

Using an ordinary herd class model, for example

$$\underline{y} = \underline{X}_1\underline{b} + \underline{Zu} + \underline{e} \qquad \text{model 2}$$

The variation between herds and HYS will not be taken into account axactly.

Estimators for model 1 can be obtained from Mixed Model Equations:

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 & X_1'Z \\ X_2'X_1 & X_2'X_2 & X_2'Z \\ Z'X_1 & Z'X_2 & Z'Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X_1'y \\ X_2'y \\ Z'y \end{bmatrix}$$

$$G^{-1} = I * \frac{\sigma_e^2}{\sigma_s^2} \qquad \text{Cov}(u,e) = 0 \qquad \text{Cov}(s_i,s_j) = 0$$

Solutions for the MME can be derived from a g - inverse of the coefficient matrix. The simplest way to find a g - inverse of a singular matrix is to delete the minimum number of rows and columns required to obtain a full rank submatrix, invert this submatrix, and substitute "zeros" for the deleted rows and columns (SEARLE, 1966).
After deleting the herd classes from the coefficient matrix, the result would be an ordinary HYS - model. Deleting herds within herd class x year x season, we obtain model 2.

The equations to be solved for the herdclass model are:

$$\begin{bmatrix} X_1'X & X_1'Z \\ Z'X_1 & Z'Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b}_1^* \\ \hat{u}^* \end{bmatrix} = \begin{bmatrix} X_1'y \\ Z'y \end{bmatrix}$$

Let a g - inverse of this coefficient matrix be:

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{12}' & C_{22} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}$$

The variance of the error in prediction is equal to or less than that resulting from the solution of a HYS - model. The estimators and predictors are biased.
Their expectations under the true model are:

$$E(k_1'b_1^*) = k_1'b_1 + k_1'L'b_2$$

$$L' = C_1T \qquad T = \frac{X_1'X_2}{Z'X_2}$$

$K_1'\hat{b}_1^*$ have expectations k_1b_1 plus some linear function of b_2

(HENDERSON, 1984). For this example:

$$E(k_1\hat{b}_1^*) = k_1'b_1 + k'C_1Tb_2$$
$$= k_1'b_1 + k'(C_{11}X_1'X_2 + C_{12}Z'X_2)\ b_2$$
$$E(\hat{u}^*) = 0 + C_2Tb_2$$
$$= 0 + (C_{12}'X_1'X_2 + C_{22}Z'X_2)\ b_2$$

Of course the true values of b_2 are never known. Consequently predictors or estimates by alternative models cannot be compared with these true values to see which model is most accurate. For calculating the bias we will use the solutions obtained from the HYS - model as an approximation for the unknown true values. It would also be possible to use simulated values for b_2.

MATERIALS AND METHODS

In this analysis, 15,620 first lactation records from a period of three years were used to calculate deviation in ranking and bias of the estimators using different herd class x year x season models. Lactations were precorrected for age at calving and extended to 305 days. 176 sires with at least 20 daughters were included in the investigation. In the HYS - model 5265 HYS groups were set up (5 seasons for each year). Every HYS contains almost two records from two different sires.
For the same set of data, HYS were replaced by 254 (HCYS1) or 645 (HCYS2) milk yield groups (300 kg steps or 100 kg respectively per region and season). For fat yield, 265 (HCYS1) or 558 (HCYS2) fat yield groups (10 kg steps or 5 kg respectively per region and season) were set up.
The necessary variance components were estimated by Restricted Maximum Likelihood (REML) (SEARLE, 1979).

RESULTS AND DISCUSSION

Table 1:

Spearman correlation coefficients between models including herd x year x season (HYS) or herd class x year x season (HCYS) as fixed effects

trait	HCYS1	HCYS2
milk yield kg (HYS)	.920	.930
fat yield kg (HYS)	.938	.955

Spearman correlation coefficients were calculated to compare ranking sires with a HYS - model and the two different herd class x year x season models (HCYS1 with 254 classes for milk and 265 for fat yield; HCYS2 with 645 classes for milk and 558 for fat yield). Using smaller herd classes correia-tion coefficients are hardly higher than with fewer classes (Table 1).

In sire evaluation, we calculate the differences between the daugthers of bull j and the contemporaries within each herd or herd class. We although pay attention to the genetic merit of the contemporaries. Using herdclasses we have to assume that the mean genetic value of all herds in one herdclass is the same. But this is not true.

Estimates of variance components for milk yield and fat yield with different models show that adjustment for the genetic merit of different herds is made inaccessible when using herd class x year x season models. Consquently, we get higher estimates for sire component of variance using HYS, as shown in table 2.

Table 2:

Estimates of heritabilties and variance components for milk and fat yield with different models

Model		milk yield variance	h^2	fat yield variance	h^2
HYS	σ_s^2	50 462	.349	82	.353
	σ_e^2	527 126		851	
HCYS1	σ_s^2	35 452	.255	68	.288
	σ_e^2	520 576		912	
HCYS2	σ_s^2	36 652	.265	71	.292
	σ_e^2	515 825		900	

In agreement with DEMPFLE (1982), using more and smaller herd classes error component of variance is reduced. Different ranking results from the high variation of HYS estimators within herd class x year x season. The high variation of the bias of the sire effects indicates large differences in sire effects between HYS and HCYS models. Nevertheless it could hardly be reduced by using smaller herd classes (Table 3).

Table 3:

Standard deviation of the bias and HYS - effects nested in one herd class for the two different herd class x year x season models				
model	milk yield		fat yield	
	s_{bias}	s_{HYS}	s_{bias}	s_{HYS}
HCYS1	63	330	2.9	17.6
HCYS2	60	290	2.4	16.9

CONCLUSIONS

Substituting a herd class x year x season classification for a herd x year x season classification leads to biased estimators since certain fixed effects are ignored. Estimates for the additive genetic variance component are reduced. A higher difference in ranking can be calculated for the cow effects. The predictors result from only one herd x year x season. There is no possibility to compensate for a bias in prediction by using records from different herd classes. It is only useful to set up herdclasses for HYS groups with less than 2 or 3 animals from just one sire so as to reduce the total number of records lost. All the other HYS groups will themselves be included in the model, since an increase in the number of records in a HYS class above 4 or 5 records improved the accuracy of sire proofs only slightly.

REFERENCES

DEMPFLE, L. (1982): Zuchtwertschätzung beim Rind mit einer ausführlichen Darstellung der BLUP-Methode. Beiheft zur Zeitschrift für Tierzüchtung und Züchtungsbiologie, Heft 3.

HENDERSON, C.R. (1975): Comparison of Alternative Sire Evaluation Methods. Journal of Animal Science, Vol. 41, 760 - 770

HENDERSON, C.R. (1984): Applications of Linear Models in Animal Breeding, University of Guelph.

SEARLE, S.R. (1966): Matrix Algebra for the Biological Sciences, Verlag: John Wiley and Sons, Inc., New York.

SEARLE, S.R. (1979): Notes on Variance Component Estimation. Biometrics Unit, Cornell Univ., Ithaca, New York.