Optimization of Test Day Models for Genetic Evaluation with Small Herd Sizes

***R. Emmerling* [1] *, K.-U. Götz* [1] *, G. Thaller* [2] *, L. Dempfle* [3]**

[1)]Bavarian Institute of Animal Production, Prof.-Dürrwaechter-Platz 1, 85586 Poing-Grub
emmerl@blt.bayern.de.
[2)]Institute of Animal Breeding TU München, 85350 Freising-Weihenstephan.
[3)]Institute of Animal Science TU München, 85350 Freising-Weihenstephan.

Introduction

The use of test day models including single test day yields has become more and more important for breeding value estimation for milk production traits in the recent years. With the transition from aggregated lactation yields to test day yields the amount of data is increasing drastically. However, test day model approaches offer better opportunities to correct for environmental influences. These new opportunities lead to a wide range of effects that have to be examined when validating a model. Depending on the data structure in some cases problems may be caused due to small fixed effect subclasses, especially concerning the contemporary group (Meyer et al., 1989; Ptak and Schaeffer, 1993; Reents et al. 1995; Swalve, 1995). Therefore, the question whether to include interactions between fixed effects in the model is especially important for cattle populations with an unfavourable herd structure.

The goal of genetic evaluation should be to maximize genetic progress per time unit. One of the factors that directly influences selection response is the correlation between true and estimated breeding values. Aim of the present study was to analyse how this correlation is affected if interaction effects are included into the evaluation model. The method is demonstrated with the interaction between lactation and herd test day (HTD) as an example. The analysis is based on test day milk yields of the Braunvieh breed in Southern Germany.

Material

Data for the analysis included about 6.4 million test day yields of 397000 cows from 1990 to 1997. Two data sets were created by randomly sampling 40 complete herds from two geographical regions with different herd structures. One region contained herds near and within the mountain region and the other contained herds under more intensive conditions. The data sets included test day yields from first to third lactation. The model for the yield of one can contain the same or different herd test day effects for all lactations. Characteristics of the data are given in Table 1. Data sets included test day yield information from 2477 and 1782 cows, respectively. Parents without records were added to the pedigree file, so that the total number of animals was 5645 and 4145, respectively. Average daily milk yield of the Braunvieh cows was 18.0 kg (s = 5.4 kg), 0.74 kg fat (s = 0.23 kg) and 0.63 kg protein (s = 0.18 kg).

Table 1. Structure of herd test days (HTD) and number of yield records per HTD for having only one HTD effect and for having one HTD effect for each lactation for the two data sets (DAT I und DAT II)

	one HTD for lactations 1 to 3		HTD only lactation 1		HTD only lactation 2		HTD only lactation 3	
	DAT I	DAT II	DAT I	DAT II	DAT I	DAT II	DAT I	DAT II
number of records	39.853	27.971	16.826	11.221	13.118	9.197	9.909	7.553
number of HTD	3.349	3.360	3.200	3.138	3.122	3.055	2.995	2.850
average records/HTD	11,9	8,3	5,26	3,58	4,20	3,01	3,31	2,65
percentage of HTD with								
1 record	2,1	3,2	10,2	15,2	13,4	20,0	21,6	26,4
≤ 2 records	4,7	6,5	22,5	37,2	30,0	46,4	42,4	54,8
≤ 5 records	17,7	26,0	62,0	83,0	74,5	91,2	86,3	93,9
≤ 10 records	49,7	71,7	90,9	99,3	97,2	99,7	99,3	99,9

Methods and model

The statistical model for analysing the test day data was a multiple trait test day model with repeated observations within each lactation. The first, second and third lactation were regarded as different traits (Reents, 1995). The influence of stage of lactation was corrected with a function of four regression coefficients (Ali and Schaeffer, 1987), which were fitted within subclasses of cows. These subclasses contained animals with same parity, similar calving age (three classes), calving season (two), and calving interval (five).

The statistical model for the analysis of test day observations was:

$$\mathbf{y_{ijklmnop} = L_i + HTD_{jk} + LHTD_{ijk} + (LCSA)_{ilmn} + b_{1ilmn}\,(D/c) + b_{2ilmn}\,(D/c)^2 + b_{3ilmn}\,\ln(c/D) + b_{4ilmn}\,[\ln(c/D)]^2 + A_{io} + P_{io} + e_{ijklmnop}}$$

where $y_{ijklmnop}$ is the pth milk yield of cow o in lactation i; L_i is the fixed effect of lactation i; HTD_{jk} is the herd effect in herd j on test day k over all lactations; $LHTD_{ijk}$ is the interaction effect between lactation i and HTD_{jk}; $(LCSA)_{ilmn}$ is the fixed 'lactation' x 'calving interval' x 'season of calving‘ x 'age of calving' interaction effect, with a total of 90 subclasses; b_{1ilmn} and b_{2ilmn} are fixed regression coefficients on the linear and quadratic effects of (D/c), where D is the lactation stage in days and c is a constant (c=381); b_{3ilmn} and b_{4ilmn} are fixed regression coefficients on the linear and quadratic effects of ln(c/D); A_{io} is the random additive genetic effect of cow o in lactation i; P_{io} is the random permanent environmental effect of cow o within lactation i, and $e_{ijklmnop}$ is the random residual effect.

Since $LCSA_{ilmn}$ already contains the effect of the lactation, the main effect of lactation need not be included in the analysis. The interaction LHTD was included as a random effect with different operational values in separate analyses. The following three situations are especially interesting:

i. The variance of the interaction effect was set to a very small value, which is converging to a model without interaction.

ii. The variance of the interaction effect was set to a very large value, which is converging to a model with fix interaction.

iii. The variance of the interaction effect was set to the true variance.

Only in the third case the approach has BLUP properties, otherwise the approach can be looked at as empirical BLUP. The method of investigation is based on the calculation of the correlation between true and estimated breeding value:

$$r_{HI} = \frac{\mathrm{Cov}\,(u,\hat{u}')}{\sqrt{\mathrm{Var}(u)\,\mathrm{Var}(\hat{u})}}$$

where Var(u) is the variance of true breeding values. The covariance between true and estimated breeding values ($Cov(u,\hat{u}')$) and the variance of the estimated breeding values ($Var(\hat{u})$) were calculated from blocks from the inverted left hand side (LHS) of the mixed model equations (MME) (see also Appendix):

$$Cov(u,\hat{u}') = G - \tilde{T}_{22} \qquad [1]$$

$$Var(\hat{u}) = G - \tilde{T}_{22} + \tilde{T}_{21}\tilde{F}^{-1}\Delta F\tilde{F}^{-1}\tilde{T}_{12} \qquad [2]$$

where

- G: matrix of additive genetic (co-)variances;
- $\tilde{T}_{22}$: block of the inverse of the LHS of the MME, that corresponds to the ($Z_1R^{-1}Z_1 + G^{-1}$)-block of the animal effects in the LHS, where Z_1 is the design matrix of the animal effects;
- $\tilde{T}_{21}$: block of the inverse of the LHS, that corresponds to the off-diagonal block between the interaction LHTD and the ($Z_1R^{-1}Z_1 + G^{-1}$)-block of the animal effects in the LHS, $\tilde{T}_{12}$ is the transpose of $\tilde{T}_{21}$;
- $\tilde{F}^{-1}$: inverse of the assumed covariance matrix of LHTD, where $\tilde{F} = I\tilde{\sigma}^2_F$;
- ΔF: difference between the true variance of LHTD (F) and the operational value ($\tilde{F}$).

$\tilde{T}$ is the inverse of the coefficient matrix of the MME where the operational values $\tilde{G}$, $\tilde{R}$, and $\tilde{F}$ are applied. The correlation r_{HI} for single animals was calculated by aggregating the variances and covariances from equations [1] and [2] with equal weights for lactations. The variance of the estimated breeding values depends on the operational value and on the true variance of the interaction LHTD. Therefore, it is possible to calculate for a given operational value the correlations r_{HI} for several possible true variance components. The blocks of the inverse LHS were calculated with the software FSPAK (PEREZ-ENCISO et al., 1994) and the FORTRAN 90 interface FSPAK90 (MISZTAL und PEREZ-ENCISO, 1998).

Results and discussion

The statistical model was applied to the two data sets in separate runs assuming three different operational values for interaction LHTD ($\tilde{\sigma}^2_F = 10^{-6}$, 1.0 und 10^{9} kg^2). Correlations were calculated for the 'true' values from zero to 100 kg^2. The results are summarized in Table 2 for all cows with test day observations. If $\tilde{\sigma}^2_F$ was assumed very high the results converged to results obtained from a model with fix LHTD effect. In the opposite case when $\tilde{\sigma}^2_F$ was set to a very low value the results converged to results from a model without the interaction effect.

Table 2. Average correlation r_{HI} for animals with test day information in the two data sets DAT I (n=2477) and DAT II (n=1782) for three different operational values in the analysis over different true variances of the interaction LHTD

operational values of σ^2_F (kg^2)		true variance of interaction LHTD (kg^2) $\sigma^2_F = 0.0$	$\sigma^2_F = 1.0$	$\sigma^2_F = 5.0$	$\sigma^2_F = 10$	$\sigma^2_F = 50$	$\sigma^2_F = 100$
$\tilde{\sigma}^2_F = 10^{-6}$	DAT I	0.6827	0.6816	0.6774	0.6722	0.6361	0.5997
	DAT II	0.6719	0.6707	0.6661	0.6605	0.6223	0.5846
$\tilde{\sigma}^2_F = 1.0$	DAT I	0.6825	0.6818	0.6788	0.6751	0.6483	0.6200
	DAT II	0.6718	0.6709	0.6673	0.6630	0.6323	0.6008
$\tilde{\sigma}^2_F = 10^{9}$	DAT I	0.6669	0.6669	0.6669	0.6669	0.6669	0.6669
	DAT II	0.6482	0.6482	0.6482	0.6482	0.6482	0.6482

The comparison of the results with quasi fixed and quasi ignored LHTD effect shows the consequences of ignoring LHTD. The analysis with an operational value of 10^9

caused up to 0.016 (DAT I) and 0.024 (DAT II) smaller correlations for cows with yield information for true variances in the range from zero to 10 kg^2. The observed differences were higher in the second data set with the more unfavourable herd structure. When the true variance exceeded the 10 kg^2, the average correlations from the model with the quasi fixed effect were higher than those from analysis with the quasi ignored effect. The analysis where a value of 1.0 kg^2 was assumed for the operational value gave similar results as compared to the analysis with the operational value 10^{-6} kg^2.

To obtain an idea of a realistic range for the true variance of the investigated interaction, larger data sets were analysed with a fixed model. The estimates for variance of interaction were between 5.4 and 5.8 kg^2 in different data sets. These values can be regarded as upper bounds for the true variance of the interaction in the population, because the expected value of the variance of the estimates is greater than the variance of the true effect.

Including LHTD as a fixed effect in the model results in a comparison of animals within a certain lactation and HTD. The consequence of this restriction is a large percentage of subclasses with only one or two observations (see Table 1). Correlations between true and estimated breeding values depending on the herd size are presented in Table 3. The increase of correlation r_{HI} assuming a true state of nature of σ_F^2 = 5.0 kg^2 was between 0.010 and 0.037. When excluding the interaction LHTD, correlation increased especially in smaller herds, as expected.

Table 3. Differences of correlation r_{HI} between the analysis with quasi ignored and quasi fixed interaction effect in different herdsizes in the second dataset

herds with x calvings in DAT II	(% of cows)	true variance of interaction LHTD (kg^2) σ_F^2 = 0.0	σ_F^2 = 1.0	σ_F^2 = 5.0	σ_F^2 = 10	σ_F^2 = 50	σ_F^2 = 100
<= 20	(2.4)	0.079	0.077	0.072	0.066	0.022	-0.020
21-30	(11.5)	0.043	0.042	0.037	0.031	-0.007	-0.045
31-40	(11.4)	0.030	0.029	0.025	0.019	-0.020	-0.058
41-50	(17.3)	0.022	0.021	0.016	0.011	-0.025	-0.062
51-60	(19.2)	0.019	0.018	0.013	0.008	-0.031	-0.070
61-80	(38.0)	0.016	0.015	0.010	0.004	-0.034	-0.071

Conclusions

Based on the results of this analysis the correlation between true and estimated breeding values could be increased by ignoring the interaction LHTD when the true variance of this effect is between zero and 10 kg^2 milk. The amount of increase depends on the amount of available information within herds. To evaluate a model it is necessary to take into account the correlation r_{HI} as well as the bias of estimated breeding values. Treating LHTD as a random interaction effect the risk of getting biased breeding values

exists, but this is assumed to be very unlikely. Another aspect is the increase of the number of equations in the MME when the interaction effects are included in the model. In the investigated data sets the increase is about 25 percent of the dimension of the MME. Considering the small risk of bias and the advantages of higher correlations r_{HI} and a computationally less demanding procedure, the model without the interaction effect LHTD seems to be well suited for the unfavourable herd structure in Southern Germany.

References

Ali, T.E. & Schaeffer, L.R. 1987. Accounting for covariances among test day milk yield in dairy cows. Can. J. Anim. Sci. 67, 637-

644.
Meyer, K., Graser, H.-U. & Hammond, K. 1989. Estimates of genetic parameters for first lactation test day production of Australian Black and White cows. Livest. Prod. Sci. 21, 177-199.
Misztal, I. & Perez-Enciso, M. 1998. FSPAK90 - a Fortran 90 interface to sparse-matrix package FSPAK with dynamic memory allocation and sparse matrix structure. Proc. 6th World Cong. Gen. Appl. Livest. Prod. 27, 467-468.
Perez-Enciso, M., Misztal, I. & Elzo, M.A. 1994. FSPAK- an interface for public domain sparse matrix subroutines. Proc. 5th World Cong. Gen. Appl. Livest. Prod. 22, 77-78. Ptak, E. & Schaeffer, L.R. 1993. Use of test day yields for genetic evaluation of dairy sires and cows. Livest. Prod. Sci. 34, 23-34.
Reents, R., Jamrozik, J., Schaeffer, L.R. & Dekkers, J.C.M. 1995. Estimation of genetic parameters for test day cell score. J. Dairy Sci. 78, 2847-2857.
Swalve, H.H. 1995. The effect of test day models on the estimation of genetic parameters and breeding values for dairy yield traits. J. Dairy Sci. 78, 929-938.

Acknowledgment

Financial support from 'Bayerisches Staatsministerium für Ernährung, Landwirtschaft und Forsten' is gratefully acknowledged.

Appendix

The developement of the method is shown on a simple example with one random effect (u) and one effect, that should be investigated (α).

True state of nature

$y = 1\mu + X\alpha + Zu + e$

$$Var(e) = R = I\sigma_e^2 \; ; \qquad 0 \le \sigma_e^2 \le \infty \; ;$$

$$Var(u) = G = I\sigma_u^2 \; ; \qquad 0 \le \sigma_u^2 \le \infty \; ;$$

$$Var(\alpha) = F = I\sigma_\alpha^2 \; ; \qquad 0 \le \sigma_\alpha^2 \le \infty \; ;$$

$Var(y) = XFX' + ZGZ' + R$

$Cov(y,u') = ZG$

Estimation procedure

Setting up the MME :

$$\begin{bmatrix} 1'\tilde{R}^{-1}1 & 1'\tilde{R}^{-1}X & 1'\tilde{R}^{-1}Z \\ X'\tilde{R}^{-1}1 & X'\tilde{R}^{-1}X + \tilde{F}^{-1} & X'\tilde{R}^{-1}Z \\ Z'\tilde{R}^{-1}1 & Z'\tilde{R}^{-1}X & Z'\tilde{R}^{-1}Z + \tilde{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{\alpha} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 1'\tilde{R}^{-1}y \\ X'\tilde{R}^{-1}y \\ Z'\tilde{R}^{-1}y \end{bmatrix}$$

Symbolize the designmatrices as :

$$[1 \quad X \quad Z] = W$$

Denote the inverse of the coefficient matrix as :

$$\begin{bmatrix} 1'\tilde{R}^{-1}1 & 1'\tilde{R}^{-1}X & 1'\tilde{R}^{-1}Z \\ X'\tilde{R}^{-1}1 & X'\tilde{R}^{-1}X + \tilde{F}^{-1} & X'\tilde{R}^{-1}Z \\ Z'\tilde{R}^{-1}1 & Z'\tilde{R}^{-1}X & Z'\tilde{R}^{-1}Z + \tilde{G}^{-1} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} \tilde{T}_{00} & \tilde{T}_{01} & \tilde{T}_{02} \\ \tilde{T}_{10} & \tilde{T}_{11} & \tilde{T}_{12} \\ \tilde{T}_{20} & \tilde{T}_{21} & \tilde{T}_{22} \end{bmatrix} = \begin{bmatrix} \tilde{T}_{0\_} \\ \tilde{T}_{1\_} \\ \tilde{T}_{2\_} \end{bmatrix} = \begin{bmatrix} \tilde{T}_{\_0} & \tilde{T}_{\_1} & \tilde{T}_{\_2} \end{bmatrix} = \tilde{T}$$

Define the matrix $\tilde{P}^{-1}$ and the vector $\hat{t}$:

$$\tilde{P}^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \tilde{F}^{-1} & 0 \\ 0 & 0 & \tilde{G}^{-1} \end{bmatrix} \qquad \hat{t} = \begin{bmatrix} \hat{\mu} \\ \hat{\alpha} \\ \hat{u} \end{bmatrix}$$

Then the MME and solutions can be written as :

$$\left[W'\tilde{R}^{-1}W + \tilde{P}^{-1}\right]\hat{t} = W'\tilde{R}^{-1}y$$

$$\hat{t} = \tilde{T}\,W'\tilde{R}^{-1}y \quad .$$

By definition :

$$\tilde{T}\left[W'\tilde{R}^{-1}W + \tilde{P}^{-1}\right] = I$$

and it follows :

$$\tilde{T}\left[W'\tilde{R}^{-1}W\right] = \tilde{T}\left[\left(W'\tilde{R}^{-1}W + \tilde{P}^{-1}\right) - \tilde{P}^{-1}\right] = I - \tilde{T}\tilde{P}^{-1}$$

That can be written as:

$$\tilde{T}\left[W'\tilde{R}^{-1}W\right] = \begin{bmatrix} I & -\tilde{T}_{01}\tilde{F}^{-1} & -\tilde{T}_{02}\tilde{G}^{-1} \\ 0 & I - \tilde{T}_{11}\tilde{F}^{-1} & -\tilde{T}_{12}\tilde{G}^{-1} \\ 0 & -\tilde{T}_{21}\tilde{F}^{-1} & I - \tilde{T}_{22}\tilde{G}^{-1} \end{bmatrix}$$

and:

$$\tilde{T}_{2\_}\,W'\tilde{R}^{-1}X = -\tilde{T}_{21}\tilde{F}^{-1} \qquad (1)$$

$$\tilde{T}_{2\_}\,W'\tilde{R}^{-1}Z = I - \tilde{T}_{22}\tilde{G}^{-1} \qquad (2)$$

Calculation of the covariance between estimated and true u :

Since $\hat{u} = \tilde{T}_{2\_}\,W'\tilde{R}^{-1}y$

we have

$$Cov(\hat{u}, u') = \tilde{T}_{2\_}\,W'\tilde{R}^{-1}ZG = (I - \tilde{T}_{22}\tilde{G}^{-1})G$$

under the condition $G = \tilde{G}$ we get:

$$Cov(\hat{u}, u') = G - \tilde{T}_{22}$$

Calculation of the variance of the estimated u :

$$Var(\hat{u}) = \tilde{T}_{2\_}\,W'\tilde{R}^{-1}\,[\,XFX' + ZGZ' + R\,]\tilde{R}^{-1}W\,\tilde{T}_{\_2}$$

$$\begin{aligned} Var(\hat{u}) = {} & \tilde{T}_{2\_}\,W'\tilde{R}^{-1}XFX'\tilde{R}^{-1}W\tilde{T}_{\_2} \\ & + \tilde{T}_{2\_}\,W'\tilde{R}^{-1}ZGZ'\tilde{R}^{-1}W\tilde{T}_{\_2} \\ & + \tilde{T}_{2\_}\,W'\tilde{R}^{-1}R\tilde{R}^{-1}W\tilde{T}_{\_2} \end{aligned}$$

Under the condition

$G = \tilde{G}$ and $R = \tilde{R}$ and $F = \tilde{F} + \Delta F$ and using of equation (1) and (2) :

$$\begin{aligned} Var(\hat{u}) = {} & \tilde{T}_{21}\tilde{F}^{-1}\tilde{T}_{12} + \tilde{T}_{21}\tilde{F}^{-1}\Delta F\tilde{F}^{-1}\tilde{T}_{12} \\ & + G - \tilde{T}_{22} - \tilde{T}_{22} + \tilde{T}_{22}\tilde{G}^{-1}\tilde{T}_{22} \\ & - \tilde{T}_{21}\tilde{F}^{-1}\tilde{T}_{12} + \tilde{T}_{22} - \tilde{T}_{21}\tilde{G}^{-1}\tilde{T}_{12} \end{aligned}$$

$$Var(\hat{u}) = G - \tilde{T}_{22} + \tilde{T}_{21}\tilde{F}^{-1}\Delta F\tilde{F}^{-1}\tilde{T}_{12}$$

Thus under the condition

$G = \tilde{G}$ and $R = \tilde{R}$ and $F = \tilde{F} + \Delta F$ we get:

$$Cov(\hat{u}, u') = G - \tilde{T}_{22}$$

$$Var(\hat{u}) = G - \tilde{T}_{22} + \tilde{T}_{21}\tilde{F}^{-1}\Delta F\tilde{F}^{-1}\tilde{T}_{12}$$