Model Validation and Rank Reduction of Covariance Matrices in the Random Regression Test-Day Model in The Netherlands

A.P.W. de Roos, E.P.C. Koenen, A.G.F. Harbers and G. de Jong
NRS, P.O. Box 454, 6800 AL Arnhem, The Netherlands

1. Introduction

As of May 2002, the national genetic evaluation for milk, fat and protein yield in The Netherlands is performed with a random regression test-day model. Several authors described test-day models, including many different fixed effects (Emmerling *et al.*, 2000; Lidauer *et al.*, 2000; Schaeffer *et al.*, 2000). De Roos and Pool (2001) described the rank reduction of the genetic and permanent environmental covariance matrices. The aim of this study was to find the fixed effects that should be included in the Dutch test-day model and to analyse the effects of the rank reduction on breeding values of bulls.

2. Material and methods

2.1 Model validation

A genetic evaluation for milk yield was performed with 74.3 million test-day records from 4.1 million cows on 34,060 herds, recorded between May 1990 and June 2000 (De Roos *et al.*, 2001). Based on literature and common sense an initial model was constructed, including 7 fixed effects:

1. parity x days in milk (DIM)
2. parity x age x year x season of calving
3. parity x age x year x season of calving x class of DIM
4. parity x stage of pregnancy
5. parity x heterosis
6. parity x recombination
7. herd-testdate (HTD)

The genetic and permanent environmental effects were modelled with random regressions (De Roos *et al.*, 2001).

All residuals corresponding to the test-day records were used for testing the significance of other fixed effects. It was assumed that fixed effects not included in the model would explain a significant part of the variance of the residuals. Significances of fixed effects were tested in univariate analyses using F-statistics.

2.2 Rank reduction

A multi-trait genetic evaluation of milk, fat and protein yield in lactations 1, 2 and 3 was performed, using the same fixed effects as described above. The genetic parameters for this analysis were estimated in two steps:

Step 1. Single trait, multi-lactation analyses, with fourth-order Legendre polynomials for every lactation;

Step 2. Rank reduction within trait, and re-estimation of covariance matrices in one multi-trait, multi-lactation analysis (De Roos & Pool, 2001).

In step 1, the model included 5 regressions per lactation and 3 lactations, i.e. 15 regressions per trait. In the canonical decomposition step of the rank reduction, 15 'new' regressions are created of which only those with the largest eigenvalues are used in step 2.

3. Results

3.1 Model validation

Table 1 shows the fixed effects that were tested for explaining variance of the residuals. Fixed effects HTD x parity and herd x parity did not explain a significant part of the variance in the residuals. This means that the herd effect is not different for cows in different parities and can therefore be modelled with HTD. The use of HTD instead of HTD x parity is preferred because it results in larger contemporary groups.

Interactions of herd, herd-size and herd-region with DIM or class of DIM were all significant. This indicates that management not only affects the level of production but also the shape of the lactation curve.

Table 1. Fixed effects that were analysed.

Fixed effect	significance
HTD x parity	ns
herd x parity	ns
herd x class of DIM	***
herd-size x DIM	***
herd-region x DIM	***
calving interval in the previous lactation	***
calving interval in the previous lactation x class of DIM	***
days dry [a]	***
days dry x class of DIM	***
heterosis x breed combination [b]	ns
recombination x breed combination	ns

[a] interval between last test-day and calving date
[b] breed combination is Holstein x Dutch Frisian, Holstein x MRY and other
*** P<0.01, ns P>0.10

The effects of calving interval in the previous lactation and days dry were highly significant, especially with an interaction with class of DIM. The solutions for these effects were negative for calving intervals shorter than 300 days and for less than 50 days dry. The solutions for calving intervals longer than 300 days and for more than 50 days dry were almost zero. Cows with calving intervals of less than 300 days or cows that were less than 50 days dry may often be cows that have aborted. It is very likely that this abortion has a negative effect on milk production in the subsequent lactation, especially in the beginning of the lactation.

The interactions of heterosis and recombination with breed combination were not significant. This indicates a multi-breed evaluation is valid with one heterosis and recombination effect applied to all breed combinations.

3.2 Rank reduction

The three regression functions with the largest eigenvalues for protein yield are shown in Figure 1. The corresponding eigenvalues were 30,918, 11,678 and 1,821 g^2, respectively. After the rank reduction the regression functions are defined over all three lactations, whereas before the rank reduction every lactation had 5 separate functions, i.e. the fourth-order Legendre polynomials.

An animal with an estimated breeding value (EBV) of +1 g for regression function 1 will have EBVs for daily protein yield between 0.18 and 0.35 g in lactation 1, between 0.29 and 0.51 g in lactation 2 and between 0.32 and 0.59 g in lactation 3.

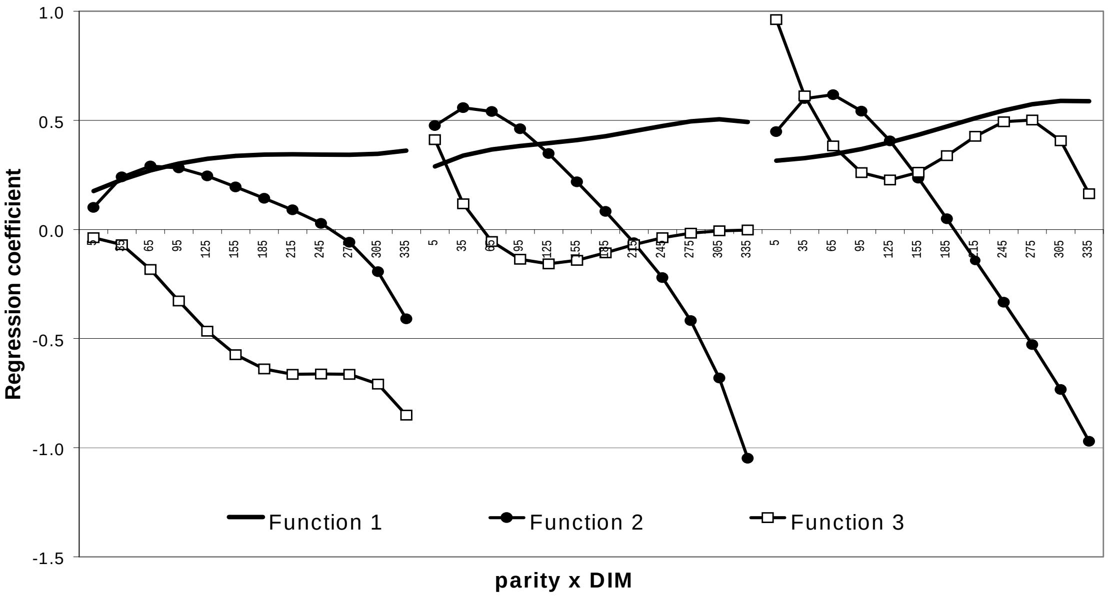

Figure 1. Three regression functions with the largest eigenvalues for protein yield in lactations 1, 2 and 3. Regression functions 1, 2 and 3 model the overall mean, persistency within lactation and progress in production over lactations, respectively.

This indicates that regression function 1 models the overall mean, because the regression coefficients are positive and relatively constant over all parities and DIM. Similarly, regression function 2 models persistency, because the regression coefficients in the beginning of the lactation are opposite to the coefficients at the end of the lactation. Regression function 3 models the progress in production over lactations, because the coefficients in lactation 1 are negative and the coefficients in lactation 3 are positive.

4. Discussion

4.1 Model validation

The significance of herd x class of DIM indicates differences in persistency across herds. This may be caused by differences in feeding systems, for example herds with flat feeding systems may have a ration that is nutritionally poor for fresh cows but rich for cows at the end of their lactation.

Gengler and Wiggans (2001) and De Roos *et al.* (2002) showed that adding a random regression effect for herd x 2-year in the parameter estimation has a large effect on the genetic covariance structure. The random regression effect for herd has a relatively large variance at the beginning and end of the lactation, whereas the genetic and permanent environmental variances are constant during the lactation. When the random regression effect for herd was not included in the model, the genetic and permanent environmental variances were high at the beginning and end of the lactation, which is not consistent with results from multi-trait analysis (Van der Werf *et al.*, 1998). Based on these results, De Roos *et al.* (2002) proposed to include a random regression effect of herd x year in the national genetic evaluation of production traits in The Netherlands. The interaction between herd, herd-size or herd-region with DIM will therefore not be included as fixed effects in the model.

To account for systematic effects on milk yield due to abortion, the effect of days dry x class of DIM is included in the Dutch test-day model. The effect of days dry is preferred above calving interval because an abortion often results in a short dry period, but it may not always result in a short calving interval. The solutions of days dry x class of DIM were obtained from the national genetic evaluation of May 2002. The estimated difference in milk yield between a dry period of less than 21 days compared to a dry period of more than 60 days was 9 kg at the beginning of the lactation and 4 kg at the end.

4.2 Rank reduction

After a genetic evaluation with the reduced rank multi-trait model, yield deviations (YDs) for every test-day record were computed as YD = test-day yield – fixed effects - ½ genetic effect of dam. For bulls, the YDs of daughters were averaged for all test-day records on daughters within 4 lactation stages (5-50, 51-120, 121-200 and 201-335 DIM) and 3 lactations. Subsequently, daughter yield deviations (DYDs) were approximated for every lactation by multiplying the lactation stage averages with the length of the intervals (i.e. 46, 70, 80 and 135 days).

Some bulls had EBVs for 305-day yield in lactations 2 and 3 that could not really be explained from their parent average and the approximated DYDs of their daughters. These bulls were often bulls with only heifer daughters and parents with low reliabilities or bulls with a very large number of second crop heifer daughters. An example of such a bull is shown in Table 2. This bull had 10,371 producing daughters of which 90% were heifers, 9% were second parity cows and only 1% were third parity cows. Almost all second parity cows had short lactations.

Table 2. Example of a bull that got strange EBVs in the reduced rank test-day model.

lactation	approx. DYD [a]	EBV [b]
1	31.4	30
2	27.1	23
3	21.2	9

[a] approximated DYD for 305-day protein yield (kg)
[b] EBV for 305-day protein yield (kg)

The EBV of this bull was 30 kg for 305-day protein yield in lactation 1, which is very close to the approximated DYD (31.4 kg). However, this is not the case for the EBVs in lactation 2 and 3, which are 4.1 and 12.2 kg lower than the approximated DYDs, respectively. The EBV can be lower than the approximated DYD because there is culling among the daughters, but it is not expected that this has such dramatic effects as showed in this example.

A possible reason for these strange EBVs may be the reduced rank regression functions that are defined over all lactations. Before the rank reduction, the EBV for 305-day yield in lactation 3 is mainly based on the first crop daughters that have observations in lactation 3. In the reduced rank model, however, the second crop heifer daughters have a direct effect on the regressions that also describe lactations 2 and 3. This means that the EBV for 305-day protein yield in lactation 3 is largely based on the second crop daughters, because they outnumber the first crop daughters. The reduced rank model uses the linear combination of the regression functions that best describes the lactation curves of the daughters. This may not be optimal to describe the genetic curve of lactation 3, but there is only a small amount of data in lactation 3 compared to the amount of data in lactation 1. The reduced rank model can therefore give unexpected results for bulls with a very unbalanced distribution daughters over lactations.

5. Conclusions

The random regression test-day model was improved by adding the fixed effect of days dry x class of DIM to model the effect of abortion on production in the subsequent lactation. The effect of HTD x parity was not significant after HTD was already in the model. Also the interactions of heterosis and recombination with breed combination were not significant.

It is recommended to include a herd x year x lactation stage effect to model differences in persistency across herds. In the Dutch test-day model a random regression effect of herd x year will be added, as proposed by Gengler and Wiggans (2001).

In the reduced rank random regression model as described by De Roos *et al.* (2001), regression functions are defined over all lactations. This may result in EBVs that cannot be explained from the yield deviations of the daughters when the amount of data in lactation 1 is much larger than in lactation 3. Therefore, the random regression test-day model that is used in The Netherlands for the national genetic evaluation is a multi-lactation, full rank model and single trait for milk, fat and protein.

References

De Roos, A.P.W., Harbers, A.G.F. & De Jong, G. 2001. Random regression test-day model in The Netherlands. *Interbull Bulletin 27,* 155-158.

De Roos, A.P.W. & Pool, M.H. 2001. Test-day model parameters for dairy cattle in The Netherlands. *52nd EAAP annual meeting,* Budapest, Hungary.

De Roos, A.P.W., Harbers, A.G.F. & De Jong, G. 2002. Herd specific random regression curves in a test-day model for protein yield in dairy cattle. *7th WCGALP*, Montpellier, France (accepted).

Emmerling, R., Dempfle, L. & Götz, K.-U. 2000. Validation of fixed effects in a testday model. *51st EAAP annual meeting,* The Hague, The Netherlands.

Gengler, N. & Wiggans, G.R. 2001. Variance of effects of lactation stage within herd by herd yield. *J. Dairy Sci. 84 (Suppl. 1),* 216.

Lidauer, M., Mäntysaari, E.A., Strandén, I. & Pösö, J. 2000. Multiple-trait random regression test-day model for all lactations. *Interbull Bulletin 25,* 81-86.

Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J. & Van Doormaal, B.J. 2000. Experience with a test-day model. *J. Dairy Sci. 83,* 1135-1144.

Van der Werf, J.H.J., Goddard, M.E. & Meyer, K. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci. 81,* 3300-3308.