Random Regression Test-Day Model in The Netherlands

A.P.W. de Roos, A.G.F. Harbers and G. de Jong
NRS, P.O. Box 454, 6800 AL Arnhem, The Netherlands
Roos.S@cr-delta.nl

Introduction

NRS will introduce a random regression test-day model for the official genetic evaluation of dairy production traits in The Netherlands in May 2002. This paper deals with the estimation of the genetic parameters and the developments with respect to the national genetic evaluation.

Parameter estimation

The genetic parameters have been estimated in two steps. In the first step, the parameters for milk, fat and protein yield have been estimated separately, but multi-trait for lactations 1, 2 and 3. In the second step, the rank of the genetic and permanent environmental covariance matrices was reduced and the parameters were re-estimated in one multi-trait analysis.

STEP 1: single trait runs for milk, fat and protein

The data set comprised 500,000 test-day records from 23,700 cows on 490 herds (Pool & Meuwissen, 2001). The model included four fixed effects: year x season of calving, parity x age at calving, weekly classes of days in milk (DIM) and herd x testdate (HTD). Both the genetic and permanent environmental effect were modelled with 4th order Legendre polynomials for every lactation. Using three lactations, this resulted in 15 genetic and 15 permanent environmental solutions per animal. Residuals were divided into 10 classes within lactation, to account for heterogeneous residual variances across lactation stages. Gibbs sampling techniques were used to estimate all parameters.

Reduction of rank

To reduce computer requirements the rank of the genetic and permanent environmental covariance matrices was reduced by setting the smallest eigenvalues equal to zero. It is important that the goodness-of-fit of the reduced rank model is not much worse than that of the full rank model. Therefore, different reduced rank models were compared with the full rank model, with respect to how well they can predict 305-day breeding values when testday records after DIM 90 are omitted. The mean of squared differences between 'true' and estimated breeding values was allowed to increase by maximum of 1%. The number of regression coefficients after rank reduction is given in Table 1.

Table 1. Number of parameters before and after rank reduction

| trait	Full rank model		Reduced rank model	
	genetic	permanent environment	genetic	permanent environment
Milk	15	15	7	10
Fat	15	15	4	7
Protein	15	15	4	7
total	45	45	15	24

STEP 2: one multi trait run for milk, fat and protein

In step 2 a new data set was used comprising 850,000 test-day records from 44,000 cows from 544 herds. Only test-day records without missing observations between DIM 5 and 335 were included. Lactations were required to have at least 6 test-day records including one before DIM 45 and one after DIM 300. Cows were at least 50% Holstein and had 2 known parents and at least 9 paternal half-sibs. HTD's had at least 8 records. The pedigree contained 97,000 animals of which 474 were bulls with lactating progeny. Unknown parents were assigned to phantom groups, based on selection path, year of birth, breed and country of origin.

Results

Heritabilities for daily yields are given in Figure 1. Generally, genetic and permanent environmental variances for daily yield were constant in mid-lactation and somewhat higher at the ends of lactation 2 and 3. Genetic correlations among yields in mid-lactation (DIM 80 to 250) were above .80 in lactation 1, .65 in lactation 2 and .55 in lactation 3. Genetic correlations between DIM 5 and 335 were around .40 in lactation 1 and .10 in lactation 2 and 3. Genetic correlations between milk and protein yields were constant during the lactation, around .85. Genetic correlations between milk and fat yields were around .6 in beginning and end of the lactation and around .4 in mid-lactation.

Figure 1. Heritability of 24h milk, fat and protein yield
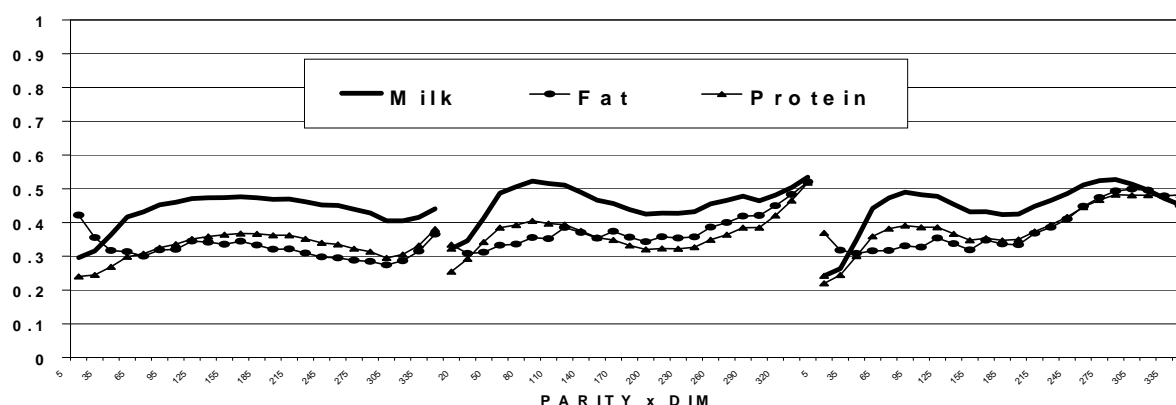
Heritabilities for 305-day yield are shown in Table 2. Note that 305-day yield is not measured on cows, so the interpretation of these heritabilities is not straightforward. Because correlations between different parities and DIM are highest for the genetic effect, lower for the permanent environmental effect and zero for the residual effect, heritabilities of traits that are a combination of many test-days (e.g. 305-day yield) are higher than the heritabilities of the underlying test-days. Genetic correlations between the lactations were similar for all traits, i.e. around .90 (lactation 1 & 2), 0.74 (1 & 3) and 0.95 (2 & 3). Lactations 1, 2 and 3 were combined into one breeding value with weights: 0.410, 0.329 and 0.261, respectively. These weights were calculated taking the frequency of expression, the moment of expression and the relation with lactations 4 and higher into account. Genetic correlations between 'overall' 305-day yield were 0.41 (milk & fat), 0.84 (milk & protein) and 0.63 (fat & protein).

Table 2. Heritabilities of 305-day yield

Lactation	milk	fat	protein
1	.55	.48	.46
2	.55	.55	.46
3	.54	.54	.49

A breeding value for persistency has been defined as: $\sum_{i=61}^{305} EBV_i - (305 - 60)EBV_{60}$. Persistency in lactation 1 has a genetic correlation around 0.63 with persistency in lactation 2 and 3, whereas the genetic correlation between persistency in lactation 2 and 3 is around .98.

National genetic evaluation

Model

In the national genetic evaluation a random regression test-day model will be used, in which 24h milk, fat and protein yield in lactation 1-3 are analysed in one multi-trait analysis. The model contains 15 genetic and 24 permanent environmental regression coefficients (Table 1). The obtained parameters, as described above, are used in the genetic evaluation.

In a first test run the following fixed effect were included with number of levels:

1. Parity x DIM — 993
2. Parity x age x year x season of calving — 2,244
3. Parity x age x year x season of calving x class of DIM — 2,772
4. Parity x stage of pregnancy — 27
5. Parity x heterosis — 30
6. Parity x recombination — 27
7. HTD — 3,164,612

The effect of parity x DIM contains a class for every day within parity and describes the average lactation curve. Fixed effect 2 accounts for differences in level of production between groups of cows that have the same lactation shape. Fixed effect 3 accounts for differences in shape of lactation. In fixed effect 2, many different classes for age at calving will be distinguished and in fixed effect 3 there will be many different classes for year-season of calving. All fixed effects are included as cross-classified effects, not as fixed regressions.

Breeding value estimation

The breeding value estimation program includes a preconditioned conjugate gradient (PCG) algorithm. Strandén & Lidauer (1999) have shown the superiority of PCG over other solving algorithms in convergence rate and computing time per iteration, especially when models are relatively complex. A disadvantage is the higher memory requirements.

A first test run has been performed in February 2001. The size of the data and model is described in Table 3. The evaluation was run on a Compaq AlphaServer type ES40 with 12.8 Gb RAM. Each iteration took around 4.5 minutes of CPU. Convergence was reached after approximately 200 iterations. The PCG vectors required almost 6 Gb of RAM.

Table 3. Size of data set and model in test run

# test-day records[*]	74,281,904
# lactations	8,391,648
# cows with test-day records	4,093,340
# herds	34,060
# herd-testdates	3,164,612
# animals in pedigree	5,915,100
# traits	3
# genetic regression coefficients	15
# permanent environmental regression coefficients	24
# mixed model equations	196,491,060

[*] Test-day records were available since 1990, all breeds are included.

Heterogeneous variances across herd-testdates

A adjustment for heterogeneous non-genetic variances across herd-testdates will be performed simultaneously with the genetic evaluation, similar as in the current lactation model in The Netherlands (Meuwissen *et al.*, 1996). Although the method is still under study, it will approximately operate as follows:

1. Compute r = yield – fixed effects – genetic effect.
   Note that r = permanent environment effect + residual.
2. Standardise r, by dividing r by the standard deviation of the permanent environment effect plus residual of observations from cows in the genetic base populations (Black-and-White cows born in 1995 with at least 87.5% HF) on the same parity and DIM. After standardisation, the standard deviation of r should be 1.
3. Compute the standard deviation of r for all HTD's.
4. Compute an adjustment factor (F) for every HTD. This factor is computed from the standard deviations of a group of HTD's on the same herd.
5. Compute an adjusted yield as:
   fixed effects + genetic effect + (permanent environment effect + residual)*F.
6. Continue solving algorithm with adjusted yield.

This process is repeated after every 10 iterations. Eventually, the obtained estimated breeding values will be standardised to the variance of the data from cows in the base population. Besides HTD, there may also be a adjustment for heterogeneous variances across other effects, e.g. age at calving.

Reliabilities

The estimation of reliabilities has been based on the method described by Jamrozik *et al.*, (2000) and has been extended with the absorption of the HTD effect, to account for contemporary group size. The latter has been described by Strandén *et al.* (2001). Computation of reliabilities in the test run (Table 3) took 9.5 hours.

Conclusions

Genetic parameters for a multi-trait random regression test-day model for milk, fat and protein yield in lactation 1-3 have been estimated. A reduction of the rank has been performed to reduce the computer requirements, but a satisfying goodness of fit was assured.

The national genetic evaluation program is based on a fast PCG algorithm. An adjustment for heterogeneous non-genetic variances across HTD's will be performed simultaneously with the genetic evaluation.

References

Jamrozik, J., Schaeffer, L.R. & Jansen, G.B. 2000. Approximate accuracies of prediction from random regression models. *Livest. Prod. Sci. 66,* 85-92.

Meuwissen, T.H.E., De Jong, G. & Engel, B. 1996. Joint estimation of breeding values and heterogeneous variances of large data files. *J. Dairy Sci. 79,* 310-316.

Pool, M.H. & Meuwissen, T.H.E. 2001. Stepwise reducing the number of fitted parameters in multiple-trait lactation random regression test-day models. *J. Dairy Sci.* (submitted).

Strandén, I. & Lidauer, M. 1999. Solving large mixed linear models using preconditioned conjugate gradient iteration. *J. Dairy Sci. 82,* 2779-2787.

Strandén, I., Lidauer, M., Mäntysaari, E.A. & Pösö, J. 2001. Calculation of Interbull weighting factors for the Finnish test day model. *Interbull Bulletin 26,* 78-79.