Experiences with Interbull Test IV: Estimation of Genetic Variance

M. Lidauer, K. Vuori, I. Strandén and E.A. Mäntysaari
MTT Agrifood Research Finland, Biotechnology and Food Science, Biometrical Genetics, Jokioinen, Finland

Abstract

The aim of this work was to access the applicability of Interbull test IV for complex random regression test-day models. Interbull test IV results were compared to results from a method where prediction error variance of Mendelian sampling deviations was obtained by a full model sampling approach. Results from both methods were in good agreement but revealed a bias in estimates of oldest and youngest cow classes.

1. Introduction

The absence of heterogeneous variance adjustment in the national evaluation often yields inflated genetic variance in the most recent birth year groups. Top bulls coming from such groups may receive upwards biased proofs in the international bull comparison.

Fikse *et al.* (2005) proposed a method to test for trends in genetic variance over time. The method is based on an idea where the genetic variance is estimated within strata using breeding values and prediction error variance (PEV) of Mendelian sampling deviations (Van Doormaal *et al.,* 1999). The method approximates PEVs of the Mendelian sampling deviations using the animals' reliability estimates. Typically in dairy cattle evaluations, the reliability estimates are approximated as well.

Fikse *et al.* (2003) found good agreement between genetic variance estimates using either the true PEVs or approximated PEVs based on approximated reliabilities using the method of Misztal and Wiggans (1988). In their study a single trait animal model for first lactation 305-day milk yield was applied. However, the quality of the approximation for complicated multiple trait random regression test-day (TD) model (RRM) evaluations has not been investigated so far. For instance, the Nordic yield evaluation uses a multiple trait RRM that models milk, protein and fat TD yields from all lactations (Mäntysaari *et al.,* 2006; Lidauer *et al.,* 2006). The first three lactations are considered as different traits. Breeding values used for the international bull comparison are obtained by calculating 305-day estimated breeding values (EBV) from random regression animal solutions and weighting EBVs for the 1st, 2nd, and 3rd lactation by 0.5, 0.3 and 0.2, respectively.

The aim of this work was to apply the proposed genetic variance estimation method (Interbull test IV) on a data subset of the Nordic RRM and compare results with a method where PEVs are obtained using a full model sampling (FMS) approach.

2. Material and methods

2.1 Interbull test IV (IB4)

The applied software, provided by Interbull (valMStol Version 1.0.2) estimates the genetic variance within year and gender ($\hat{\sigma}^2_{a_i}$) using the following equation:

$$\hat{\sigma}^2_{a_i} = \frac{\sum_{k=1}^{n_i} d_k \hat{m}_k^2}{n_i - \sum_{k=1}^{n_i} d_k PEV^*(\hat{m}_k)}, \quad [1]$$

where n_i is the number of animals in stratum i; d_k is 2, ¾, or 1 if both parents are known, one parent is missing, or both parents are missing for animal k, respectively; $\hat{m}_k$ is the estimated Mendelian sampling deviation of animal k; and $PEV^*(\hat{m}_k)$ is $PEV(\hat{m}_k)$ expressed in genetic variance units (Fikse *et al.,* 2005).

2.2 Full model sampling approach (FMS)

Different sampling approaches have been proposed when inversion of the coefficient matrix is impossible (e.g. Gracía-Cortéz and Sorensen 2001). In analogy to such ideas, a rather simple method can be implemented if only the average PEV for a group of animals is required. PEVs calculated from a FMS method are close to the true PEVs, if sampling error is negligible small. Equation [1] is equivalent to:

$$\hat{\sigma}^2_{a_i} = \frac{1}{n_i}\sum_{k=1}^{n_i} d_k \left[\hat{m}_k^2 + PEV(\hat{m}_k)\right]. \qquad [2]$$

Given, $PEV(\hat{m}_k) = \left(1 - r^2_{TIm}\right)\frac{\sigma_a^2}{d_k}$, [3]

it can be shown that for sufficiently large strata size n_i or number of replicates (r):

$$\hat{\sigma}^2_{a_i} = \frac{\sum_{k=1}^{n_i} d_k \hat{m}_k^2}{n_i} \frac{1}{r}\sum_{j=1}^{r}\left[\frac{\sum_{k=1}^{n_i} d_k \tilde{m}_{kr}^2}{\sum_{k=1}^{n_i} d_k \hat{\hat{m}}_{kr}^2}\right], \qquad [4]$$

where $\tilde{m}_{kr}$ is the simulated true Mendelian sampling deviation of animal k of replicate r and $\hat{\hat{m}}_{kr}$ is its corresponding estimate from replicate r. Ten full model data replicates where simulated for this analysis.

2.3 Comparison of methods

Comparison of both methods was carried out with data from 5000 randomly sampled Finnish Ayrshire herds from the Nordic yield traits evaluation. The data included 11.2 million TD records from all lactations of 449,160 cows recorded between January 1988 and December 2006. In total, there were 621,775 animals in the pedigree.

The model was the same RRM as used for the official Nordic evaluation (Lidauer *et al.*, 2006). Breeding values were described by reduced rank covariance functions across traits and stage of lactation (Mäntysaari *et al.*, 2006). Heterogeneity of variance was accounted applying the method of Meuwissen *et al.* (1996). Solutions for genetic animal effects were used to derive a 305-day EBV for each biological trait weighted across first three lactations. The required reliability estimates for the animals' EBVs were approximated by the method of Misztal and Wiggans (1988).

Sampling of full model data replicates

A FMS procedure was implemented into the breeding value estimation software MiX99 (Vuori *et al.*, 2006). In the procedure, for each random effect in the model true values are generated from a normal distribution using the Cholesky decomposition of the (co)variance matrix. Phantom parent group effects are set to zero and each animal's true breeding value is generated as the parental average plus the Mendelian sampling deviation. Fixed effects are set to zero. Observations are generated by summing for each record the corresponding effects and adding a random error term.

Ten replicates of full model samples were solved applying the same model to obtain for each animal k a set true ($\tilde{m}_{kr}$) and estimated ($\hat{\hat{m}}_{kr}$) Mendelian sampling deviations.

Estimation of genetic variance

Genetic variances were estimated applying both methods, IB4 and FMS. The estimation was carried out for each birth year and separately for bulls and cows. For a bull it was required to be born after 1980 and to have at least ten daughters with observations beyond days in milk 100. A cow had to have at least one TD observation. Only animals with known parents were included, since animals with missing parents have poor reliability estimates as shown by Fikse *et al.* (2003). The size of genetic variance strata are given in Table 1.

3. Results and discussion

With the FMS approach one replicate was enough to get sufficiently accurate variance estimates for the cow groups, whereas for the bull groups at least ten replicates were needed (Figure 1).

Table 1. Number of animals in birth year strata.

Year	Bulls	Cows	Year	Bulls	Cows
1981	49		1993	92	25422
1982	60		1994	101	26009
1983	124		1995	119	26352
1984	150		1996	116	26430
1985	164	2504	1997	122	26037
1986	176	18177	1998	104	25679
1987	107	19611	1999	113	25663
1988	152	20301	2000	179	23959
1989	158	21741	2001	135	25232
1990	110	23723	2002	7	23459
1991	126	23336	2003		22348
1992	125	24387	2004		12048

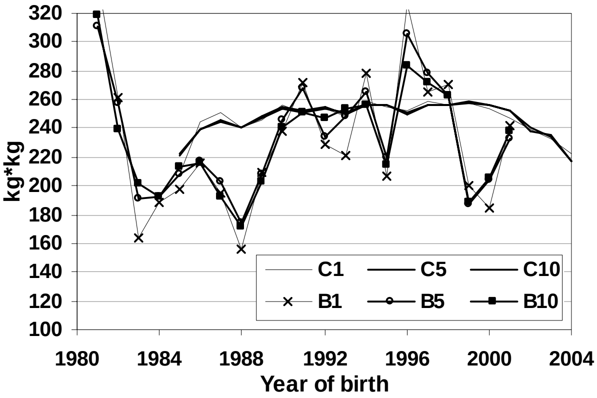

Figure 1. Estimated genetic variances for combined 305-day protein yield given by cow (C) and bull (B) year classes when full model sampling approach was based on one (C1, B1), five (C5, B5), or ten (C10, B10) replicates.

Overall, there was agreement between the estimates from both methods. Estimates for bull groups did not show any systematic difference between the methods, whereas for cow groups systematic differences were noticeable (Figure 2).

Estimates from IB4 were lower than those from FMS for year classes at the beginning and at the end of the time period, which suggests that reliabilities for cows are overestimated for these year groups. However, also estimates from FMS were still lower at the tails than at the middle of the time period.

The lower estimates for cow years at the beginning of the data period may be due to the fact that these year classes (1985, 1986) present a selected part of the cows, namely older calving cows, and therefore might have a smaller genetic variance. Similarly, only earlier calving cows in the most recent year classes (2003, 2004) have observations in the 2^{nd} and 3^{rd} lactation. Again, these cows are genetically more similar and their Mendelian

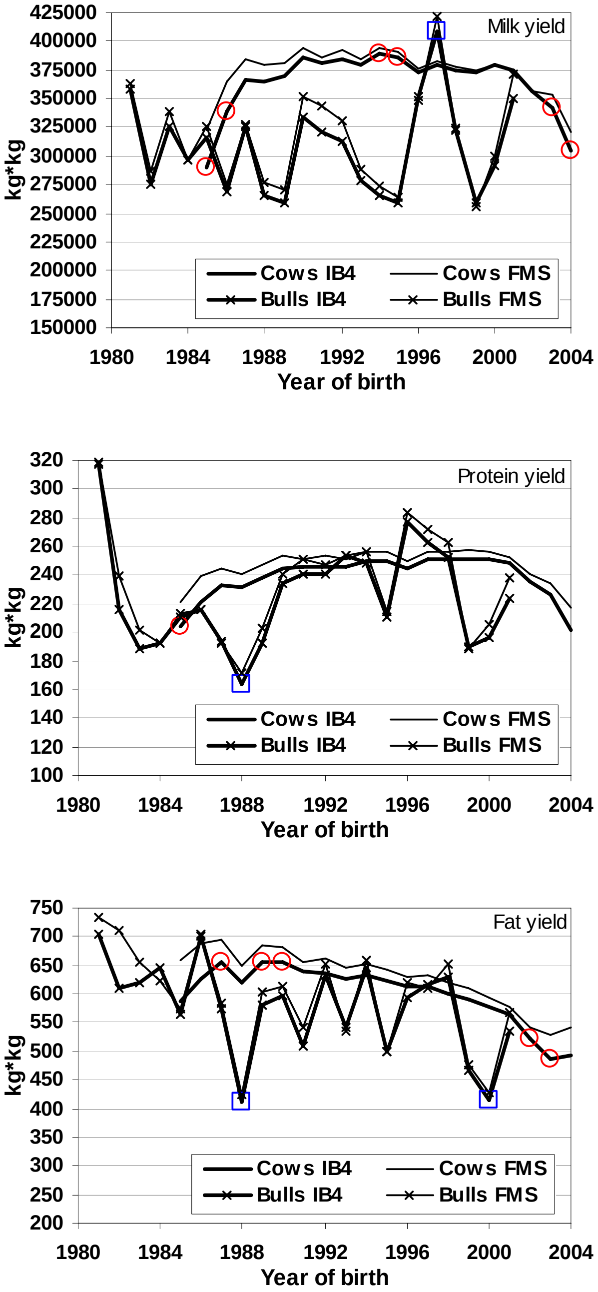

Figure 2. Estimated genetic variance for each biological trait (milk, protein, and fat yield) given for cows and bulls by birth year classes, when estimates are based on either Interbull test IV (IB4) or on a full model sampling (FMS) approach. Estimates marked with a circle (for cow classes) or with a square (for bull classes) were outside the IB4 tolerance interval.

sampling deviations will show less variation than in year classes where cows do not have records in progress.

The IB4 test has not detected a trend in genetic variance estimates for bulls and for protein yield in cows, whereas milk yield and fat yield in cows failed the test. For milk yield six estimates and for fat yield five estimates were outside the tolerance interval (Figure 2).

Considering the found bias in genetic variance estimates at the beginning and end of the time period, IB4 test was repeated using only bulls from the years 1985 to 2000 and cows from the years 1987 to 2002. Then, only fat yield for cows failed the test, which shows a slide downward trend in genetic variance estimates.

Further, it was found that variance estimates based on sires where lower than those based on cows. Since all sires with at least ten daughters were included, it was expected that variance estimates from bulls and cows do not systematically deviate from each other. However, it can be speculated whether selection of test sires could affect the variance.

4. Conclusions

Good agreement was found between genetic variance estimates from the two methods IB4 and FMS. Estimates for cow year classes were biased downwards at the beginning and at the end of the time period. The bias was caused by overestimated cow reliabilities for these classes and by the fact that these classes contained a selected group of animals; either older cows or younger cows. Estimates from cow strata were more consistent across years and therefore better suitable to test for trends in genetic variance. We recommend applying IB4 to all cows with observations from the birth years: starting year of the data minus one year, until most recent year of data minus four years.

Acknowledgements

We thank Nordisk Avelsvärdering NAV (Danish Cattle, Faba Breeding, and Swedish Dairy Association) for financing the study and for providing the data and model.

References

Fikse, W.F., Klei, L., Liu, Z. & Sullivan, P.G. 2003. Procedure for validation of trends in genetic variances. *Interbull Bulletin 31*, 30-36.

Fikse, W.F., Liu, Z. & Sullivan, P.G. 2005. Tolerance values for validation of trends in genetic variances over time. *Interbull Bulletin 33*, 200-203.

García-Cortés, L.A. & Sorensen, D. 2001. Alternative implementations of Monte Carlo EM algorithms for likelihood inferences. *Genet. Sel. Evol. 33*, 443-452.

Lidauer, M., Pedersen, J., Pösö, J., Mäntysaari, E.A., Strandén, I., Madsen, P., Nielsen, U.S., Eriksson, J.-Å., Johansson, K. & Aamand, G.P. 2006. Joint Nordic test day model: evaluation model. *Interbull Bulletin 35*, 103-107.

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

Misztal, I. & Wiggans, G.R. 1988. Approximation of prediction error variances in large-scale animal models. *J. Dairy Sci. 71*, 27-32.

Mäntysaari, E.A., Lidauer, M., Pösö, J., Strandén, I., Madsen, P., Pedersen, J., Nielsen, U.S., Johansson, K., Eriksson, J.-Å. & Aamand, G.P. 2006. Joint Nordic test day model: variance components. *Interbull Bulletin 35*, 97-102.

Van Doormaal, B.J., Kistemaker, G.J. & Sullivan, P.G. 1999. Heterogeneous variances of Canadian bull EBVs over time - appendix. 1999. *Interbull Bulletin 22*, 146-148.

Vuori, K., Strandén, I., Lidauer, M. & Mäntysaari, E.A. 2006. MiX99 – Effective solver for large and complex linear mixed models. *Proc. 8th World Congr. Genet. Appl. Livest. Prod.*, Belo Horizonte, Brazil, CD-ROM *No 27*:33.