Effect of Heterogeneous Variance Adjustment in an Across Country Evaluation of Brown Swiss

M. Lidauer[1], E.A. Mäntysaari[1], R. Emmerling[2], and I. Strandén[1]
[1] *Agrifood Research Finland MTT, Animal Production Research, 31600 Jokioinen, Finland*
[2] *Bavarian Institute of Animal Production, Postfach 1180, 85580 Poing-Grub, Germany*

1. Introduction

An across country genetic evaluation for production traits, based on a random regression (RR) test-day (TD) model (TDM), was developed for Brown Swiss cattle of Austria, Baden-Württemberg and Bavaria. The countries comprise very heterogeneous herd environments due to large differences in climate and topography as well as in herd management. Consequently, variance of TD yields is highly heterogeneous across countries, regions and herds.

Accounting for heterogeneous variance (HV) should improve the reliability of ranking of animals and the acceptability of the new TDM by breeding organisations. A simultaneous estimation of breeding values and of HV by a multiplicative mixed model (MMM) (Meuwissen *et al.*, 1996) was found preferable. Recently, it was shown that this approach is feasible for large RR TDM (De Roos *et al.*, 2001; Lidauer *et al.*, 2001).

The objective of this work was to elaborate the importance of accounting for HV in the across country evaluation. Therefore, results from two breeding value estimations, with and without HV adjustment, were compared.

2. Material and methods

2.1 Data

The joint TD data were all lactation records from the year 1990 onwards, produced by 1.02 million Brown Swiss cows. In total there were about 23 million milk yield observations. The joint pedigree comprised of 1.6 million animals.

Table 1. Number (n, million) of milk yield (kg) observations, average within-herd mean (Mean), and within-herd standard deviation (SD) given by country.

	n	Mean	SD
Austria	5.2	16.5	4.8
Baden-Württemberg	3.6	17.7	5.5
Bavaria	13.7	18.0	5.4

The milk yields from Austria were on average 8% lower than those from Germany (Table 1). However, across regions differences were even more pronounced with up to 5kg difference between the best Bavarian region and the high alpine Austrian region.

2.2 Models

Breeding value models

Estimated breeding values (EBV) for milk yield were obtained from two evaluations. First, the developed multiple-trait reduced rank RR TDM (Emmerling *et al.*, 2002) ignoring HV was used. Second, HV was accounted in the same model by applying a MMM. Both models considered TD yields of the first (F), second (S) and all later (L) lactations as different traits and had the same effects in the model. However, in the MMM, observations were scaled by multiplicative adjustment factors. TD observations, which belonged to the same production year × month × region × parity class and to the same herd × TD (HTD) class were assumed to be homogeneous and represented a stratum i. Heterogeneity was assumed across strata. This gave the following MMM:

$$\begin{bmatrix}\mathbf{y}_{F_i}\lambda_{F_i}\\ \mathbf{y}_{S_i}\lambda_{S_i}\\ \mathbf{y}_{L_i}\lambda_{L_i}\end{bmatrix}=\begin{bmatrix}\mathbf{X}_{F_i}\mathbf{b}_F\\ \mathbf{X}_{S_i}\mathbf{b}_S\\ \mathbf{X}_{L_i}\mathbf{b}_L\end{bmatrix}+\begin{bmatrix}\mathbf{T}_{F_i}\\ \mathbf{T}_{S_i}\\ \mathbf{T}_{L_i}\end{bmatrix}[\mathbf{h}]+\begin{bmatrix}\mathbf{Z}_{F_i}\\ \mathbf{Z}_{S_i}\\ \mathbf{Z}_{L_i}\end{bmatrix}[\mathbf{a}]+\begin{bmatrix}\mathbf{U}_{F_i}\\ \mathbf{U}_{S_i}\\ \mathbf{U}_{L_i}\end{bmatrix}[\mathbf{p}]+\begin{bmatrix} \\ \\ \mathbf{V}_{L_i}\mathbf{w}_L\end{bmatrix}+\begin{bmatrix}\mathbf{e}_{F_i}\\ \mathbf{e}_{S_i}\\ \mathbf{e}_{L_i}\end{bmatrix}, \qquad [1]$$

where λ_{Fi}, λ_{Si} and λ_{Li} are multiplicative adjustment factors for observations in stratum *i*; $\mathbf{b}_F$, $\mathbf{b}_S$ and $\mathbf{b}_L$ contained the fixed effects of production year × month × region × parity (YMRP), second order polynomial on calving age within region × parity, third order polynomial on days carried calf within region × parity, and regression function on days in milk within calving year × calving season × region × parity. There were 5 parity classes, where the fifth class included all parities from the fifth onwards. Effect **h** included the HTD effect, where HTD classes were defined across parities. Random effects **a**, **p**, and $\mathbf{w}_L$ included RR coefficients for the breeding values, the animal environmental effects within and across first, second and later lactations, and the animal environmental effects within each later lactation, respectively. The residuals were $\mathbf{e}_{Fi}$, $\mathbf{e}_{Si}$ and $\mathbf{e}_{Li}$. The reduced rank RR variance components were the same as given in Emmerling *et al.* (2002).

Heterogeneity model

The heterogeneity of variance in the TD data was modelled by the following linear model:

$$s_{ijk} = \beta_{1ij} + \beta_{2ik} + \varepsilon_{ijk}, \qquad [2]$$

where s_{ijk} is a heterogeneity for stratum *i* that was calculated as given in Meuwissen *et al.* (1996), β_{1ij} is a fixed YMRP classification, β_{2ik} is a random HTD classification and ε_{ijk} is the residual. YMRP and HTD were defined same as in [1]. For the random HTD a first order autoregressive process was assumed. The used variance ratio $\hat{\sigma}^2_\varepsilon / \hat{\sigma}^2_{\beta_2}$ for the random effect β_{2ik} was 5.62 and the autoregressive correlation parameter ρ_{HTD} was set to 0.80.

The multiplicative adjustment factor for observations on trait $T \in \{F, S, L\}$ in stratum *i* was calculated as $\lambda_{Ti} = \exp(-0.5\mathbf{S}_{Ti}\boldsymbol{\beta})$, where $\boldsymbol{\beta}$ contained estimates for the heterogeneity and $\mathbf{S}_{Ti}$ was a vector summing the estimates associated with stratum *i*.

2.3 Solving the MMM

The largest advantage in accounting for HV by MMM is that the method preserves HV in the data, which is explained by the TDM. This requires solving both, the breeding values and the adjustment factors simultaneously. A detailed description of the implemented solving strategy will be presented at the 7th WCGALP (Lidauer *et al.*, 2002). It follows the strategy given by Lidauer *et al.* (2001) with some modifications.

2.4 Comparison of results

The MMM method scales the observations with respect to the error variance, i.e., it homogenises the within-stratum variance of residuals across strata. For the validation, the changes in standard deviation (SD) of the residuals were monitored. Further, 305-d EBVs were derived from the solutions for the additive genetic animal effects. Changes in the EBVs were investigated for the combined EBVs, which were formed by weighting the EBVs for first, second and later lactations equally.

3. Results and Discussion

3.1 Standard deviation of residuals

The MMM accounted HV among time, later lactations, herds and regions. Within-herd SD of residuals were more similar when accounting for HV (Table 2) but they remained higher for regions with small herd sizes. The larger within-herd variation of herds with bull dams was adjusted to the same level as for all other herds. Consequently, this will also reduce a possible bias in EBVs of young sires as was found in related studies.

Table 2. Average herd size, within-herd mean (Mean) and average within-herd standard deviation (SD) of observations, residuals and scaling factors pertaining to first lactation milk yields from the years 1996-1998, grouped by herds without and with bull dams and by regions.

		Herd	1st lac. Milk yield		SD of residuals		Scaling
Country	Region	Size	Mean	SD	HV ignored	HV accounted	factor
Herds without bull dams							
Bavaria	1	12.2	16.8	3.9	1.07	1.11 (1.11)[1]	1.11
	2	16.3	16.6	3.8	1.13	1.15 (1.17)	1.09
	3	15.4	16.6	3.9	1.15	1.14 (1.16)	1.06
Baden-Württemberg	10	14.3	16.3	3.9	1.16	1.14 (1.15)	1.05
Austria	12	4.9	16.5	3.7	1.03	1.10 (1.02)	1.16
	13	7.6	16.2	3.8	1.19	1.18 (1.13)	1.09
	14	8.6	15.9	3.9	1.22	1.20 (1.14)	1.07
	15	4.9	12.0	3.0	1.16	1.33 (1.18)	1.27
Herds with bull dams							
	1	17.5	19.2	4.4	1.22	1.13 (1.17)	0.98
	3	27.0	19.0	4.3	1.28	1.15 (1.21)	0.96
	10	26.3	18.8	4.5	1.32	1.14 (1.20)	0.91
	13	11.9	18.6	4.2	1.25	1.17 (1.17)	1.01

[1] Values in parenthesis are from a test run with modifications in the algorithm to account for lost degrees of freedom due to the estimation of HTD means in the TDM.

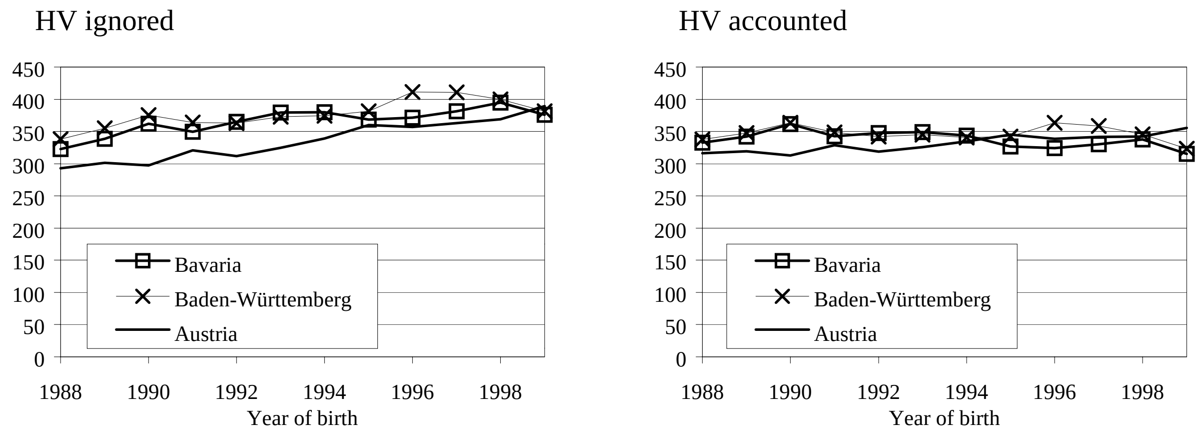

Figure 1. Standard deviation (in kg) of estimated breeding values for cows when heterogeneous variance was ignored or accounted in the breeding value estimation. Cows are grouped by country and year of birth.

3.2 Changes in EBVs

Applying MMM removed the increasing trend in the SD of EBVs over time; i.e., it corrected the scaling effect due to increasing milk yields over time (Figure 1). Hence, also genetic trends were 6 to 8% lower. These changes were similar for all three countries.

Over all, correlations between the EBVs from the two evaluations were close to unity. Correlations were above 0.99 for EBVs of bulls, above 0.98 for EBVs of cows from Germany and above 0.97 for EBVs of cows from Austria.

Changes in single EBVs were largest for the EBVs from the extremes of the distribution. For few animals the differences were two SD. However, for most of the animals (87%) changes were blow 0.2 SD. The share of Austrian cows in the group of 10 000 best cows increased by 2%, because the records with lower mean and lower SD from Austria were scaled upwards.

3.3 Within strata error variances

The applied MMM was the same as outlined by Meuwissen *et al.* (1996). The method estimates the error variances within strata using a Maximum Likelihood approach. Consequently, the error variance estimates are not corrected with respect to the rank of the coefficient matrix of the TDM. This is sufficient if herds are of reasonable sizes. However, it leads to an underestimation of the error variances in case of very small herds. Thus, the scaling factors will become overestimated, which explains the inflated SD of residuals for the regions with small herd size (Table 2).

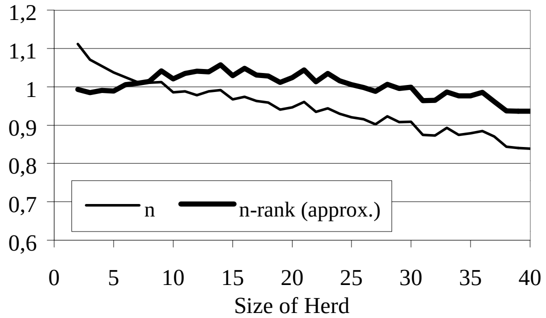

Figure 2. Mean of scaling factors without (n) and with (n-rank(approx.)) considering the rank for the estimation of error variances.

To overcome this problem a modification in the algorithm was tested, which approximates the loss in degrees of freedom due to the estimation of fixed effects in the TDM. In the modified algorithm, $n_{ijk}/n_{i.k}$ was added to each stratum's heterogeneity observation s_{ijk}; where n_{ijk} is the number of observations in stratum *ijk* and $n_{i.k}$ is the number of observations in the HTD class. The modification made scaling factors independent from the herd size (Figure 2) and the SD of residuals were no longer inflated (Table 2, values in parenthesis). Consequently, the SD of EBV of Austrian cows as well as the share of Austrian cows on the list of top cows decreased.

4. Conclusions

Accounting for heterogeneous variance by a MMM was found useful for the across country evaluation. Accounting for the rank of the coefficient matrix of the TDM when estimating within strata error variances was found important and should be considered for the future evaluations.

Acknowledgement

We thank Theo H. E. Meuwissen for his useful comments.

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.

Emmerling, R., Lidauer, M. & Mäntysaari, E.A. 2002. Multiple lactation random regression test-day model for Simmental and Brown Swiss cattle in Germany and Austria. *Interbull Meeting, Interlaken*, May 26-27,Switzerland.

Lidauer, M. & Mäntysaari, E.A. 2001. A multiplicative random regression model for test-day data with heterogeneous variance. *Interbull Bulletin 27*, 167-171.

Lidauer, M., Emmerling, R. & Mäntysaari, E.A. 2002. Accounting for heterogeneous variance in a test-day model for joint genetic evaluation of Austrian and German Simmental cattle. *Proc. 7th WCGALP*.

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.