# Heterogeneous Variance Adjustment in Across-Country Genetic Evaluation with Country-Specific Heritabilities

*M.H. Lidauer*[*], E.A. Mäntysaari[*], J. Pösö[†], J.-Å. Eriksson[‡], U.S. Nielsen[§] and G.P. Aamand[**]

## Introduction

Correct ranking of animals in across-country genetic evaluation requires careful design of both, the model for breeding value estimation and the adjustment for heterogeneous variance. The Nordic across-country genetic evaluation for yield traits of Danish, Finnish and Swedish red dairy cattle is based on a multiple-trait random regression test-day model, which considers observations from different countries as different traits. This was supported by analysis of variances, which yielded for the same biological traits differences in heritabilities across countries (Lidauer et al., 2009). However, for the breeding value estimation it was decided to impose a genetic correlation of unity and to have the same genetic variances across countries. Both conventions may depart somewhat from the true genetic nature of the breeding population but enhance genetic comparison of animals across the three countries. The genetic correlation of unity is modelled in the breeding value estimation model, whereas the same genetic variances can be secured through adjustment for heterogeneous variance.

Heterogeneous variance is accounted by a multiplicative mixed model approach (Meuwissen et al., 1996). The method solves the model for breeding values and the model for estimating the heterogeneity of variance simultaneously with the aim to scale phenotypic observations until re-estimated residual variances are the same for all the strata. Because variance ratios remain unaffected by the method, it is possible to calibrate the heterogeneous variance adjustment method to yield the same genetic variances across countries, given the genetic variances can be estimated from the estimated breeding values (EBV).

Currently, genetic variances are standardized across the countries based on standard deviations (SD) of the bull breeding values. In this study we explored the usefulness of estimating the genetic variances from cows' EBVs applying a full model sampling method and using these estimates for calibrating the heterogeneous variance adjustment method to obtain homogeneous genetic variances across the countries.

## Material and methods

**Data.** Data from the spring 2009 yield evaluation of Nordic red cattle were used for this study. The data consisted of 67.9 million test-day records on milk, protein and fat yield from

[*] MTT Agrifood Research, Jokioinen, Finland
[†] Faba Service, Vantaa, Finland
[‡] Swedish Dairy Association, Stockholm, Sweden
[§] The Danish Agricultural Advisory, Aarhus, Denmark
[**] NAV Nordic Cattle Genetic Evaluation, Aarhus, Denmark

Finnish Ayrshire (FAY), Red Danish Cattle (RDM), Swedish Red Breed (SRB), and Finnish Holstein and Finncattle. Latter two breeds were included to increase Finnish herd contemporary groups. The pedigree consisted of 4.3 million animals and genetic differences between breeds were modelled by phantom parent groups accounting for breed of origin, selection path and time.

**Multiplicative random regression model.** Test-day (TD) observations from milk, protein and fat yields, from the first three lactations and from the countries Denmark, Finland and Sweden were considered as 27 different traits. TD observations were stratified by herd × test-month classes within traits. Hence, the applied multiplicative model was

$$\mathbf{y}_{ti}\lambda_{ti} = \mathbf{X}_{ti}\mathbf{b}_t + \mathbf{T}_{ti}\mathbf{h}_t + \mathbf{Z}_{ti}\mathbf{a} + \mathbf{U}_{ti}\mathbf{p} + \mathbf{V}_{ti}\mathbf{w} + \mathbf{e}_{ti}\,, \qquad [1]$$

where $\mathbf{y}_{ti}$ contains observations of trait $t$ in stratum $i$ and $\lambda_{ti}$ is the corresponding multiplicative adjustment factor. The vector $\mathbf{b}_t$ contains fixed effects, $\mathbf{h}_t$ contains random herd×TD effects, vectors **a**, **p** and **w** contain the reduced rank random regression coefficients for the additive genetic animal effects, the non-genetic animal effects across lactations and within later lactations, respectively, and $\mathbf{e}_{ti}$ contains the random residuals. Applied variance components (Lidauer et al. 2009) differed across countries. Largest difference was found for first lactation protein yield, for which the compiled heritability for a 305-d yield was 0.33 for FAY but 0.43 for SRB. The heterogeneity of variance in the data was modelled by a fixed test-month effect and a random herd×test-year effect with a 1st order autoregressive process between years within herds. Solving of the multiplicative random regression model (RRM) was carried out in the same manner as describe by Lidauer et al. (2008).

**Adjustment of heterogeneous genetic variance.** The multiplicative model converges to a set of solutions that fulfils the equality: $n_{ti}\hat{\sigma}^2_{e_t} = \mathbf{y}^{\mathrm{T}}_{ti}\lambda_{ti}\hat{\mathbf{e}}_{ti}$, where $n_{ti}$ is the number of observations for trait $t$ in stratum $i$ and $\hat{\sigma}^2_{e_t}$ is the residual variance for trait $t$ (Meuwissen et al., 1996). Because the method scales all effects in proportionality to the residual variance, it is possible to set up an equivalent condition of the form $n_{ti}\hat{\sigma}^2_{e_t}\alpha_t = \mathbf{y}^{\mathrm{T}}_{ti}\lambda_{ti}\hat{\mathbf{e}}_{ti}$, where $\alpha_t$ is a calibration factor that will expand or reduce in proportionality the variances of all random effects included in the model for trait $t$. This allows to implement an iterative process with the aim to find a set of $\alpha_t$ 's, which yields homogeneous genetic variances across countries. The applied iterative procedure was of the form:

1) round $q$=1; $\alpha_t^{[q]} = 1.0$
2) solve the multiplicative RRM [1].
3) re-estimate the genetic variance $\hat{\sigma}^{2[q]}_{a_t}$ using the Mendelian sampling deviations of cows.
4) update $\alpha^{[q+1]}_{t.c} = \alpha^{[q]}_{t.c}\,\hat{\sigma}^{2[q]}_{a_{t.BASE}} \Big/ \hat{\sigma}^{2[q]}_{a_{t.c}}$, where $t.c$ is one of the nine traits of country $c$, and where the nine Sweden traits are considered as the base country traits.
5) if differences in re-estimated genetic SDs are larger than 1% continue with step 2).

**Re-estimation of genetic variance.** The genetic variance ($\hat{\sigma}^{2[q]}_{a_t}$) was re-estimated from Mendelian sampling deviations of cows born in 2002 to 2004. Groups of size 50067, 180573 and 159961 were extracted for RDM, FAY, and SRB, respectively. A full model sampling method was implemented to obtain for each cow group the average prediction error variance of the Mendelian sampling deviations (Lidauer et al., 2008). Only one replicate was carried out for sampling the average prediction error variance, because sampling error can be neglected due to the large group sizes. Thus, re-estimation simplified to the form:

$$\hat{\sigma}^{2[q]}_{a_t} = \frac{1}{n_t}\sum_{k=1}^{n_t} d_k \hat{m}^{2[q]}_{kt}\left[\sum_{k=1}^{n_t} d_k \tilde{m}^{2}_{kt} \Big/ \sum_{k=1}^{n_t} d_k \hat{\tilde{m}}^{2}_{kt}\right],$$

where $\hat{\sigma}^{2[q]}_{a_t}$ is the re-estimated genetic variance for trait $t$ in round $q$; $n_t$ is the cow group size for trait $t$; $d_k$ is the inverse of the variance proportion that has not been explained by known parents of animal $k$ and is 2, 4/3, or 1 if both parent are known, one parent is missing, or both parents are missing, respectively; $\hat{m}_{kl}$ is the Mendelian sampling deviation for animal $k$ and trait $t$, calculated from 305-d EBVs, where an animal's 305-d EBV was calculated as a sum over daily breeding values from days in milk 8 to 312; and the term within the brackets accounts for the average prediction error variance, where $\tilde{m}_{kt}$ is a generated true Mendelian sampling deviation for animal $k$ and $\hat{\tilde{m}}_{kt}$ is its corresponding estimate from analysis of the sampled data applying model [1]. A full model sampling procedure, explained by García-Cortéz and Sorensen (2001), was tailored to model [1]. Random effects in model [1] were generated from a normal distribution and applying for correlated random factors the Cholesky decomposition of the corresponding (co)variance matrix. Each animal's true additive genetic effect was generated as the parental average plus a Mendelian sampling deviation. Fixed effects and phantom parent group effects were set to zero. Observations were generated by summing the corresponding effects and adding a random error term. All $\lambda_{ti}$ values were set to 1.0.

## Results and discussion

Genetic variances were acceptable homogeneous across birth years but heterogeneous across countries after the multiplicative RRM was solved for the first time (figure 1). Genetic SD for RDM (Danish traits) and FAY (Finnish traits) deviated from the corresponding genetic SD for SRB (Swedish traits) between -3% and +13% (table 1). Six adjustment rounds were necessary to obtain calibration factors ($\alpha_t$), which yielded sufficiently homogeneous genetic variances. As show in table 1, genetic SD of traits for RDM and FAY differed not more than ±1% from those of the corresponding traits for SRB.

Correlations between the cows' EBVs obtained before and after the adjustment for across-country heterogeneous genetic variance, including EBVs of all cows from the five most recent birth years, were as high as ≥0.9973. Correlations were highest for SRB cows (≥0.9998) and RDM cows (0.9997) and lower for FAY cows (≥0.9966). Across-country adjustment affected most on traits for FAY, for which heritabilites were lower.

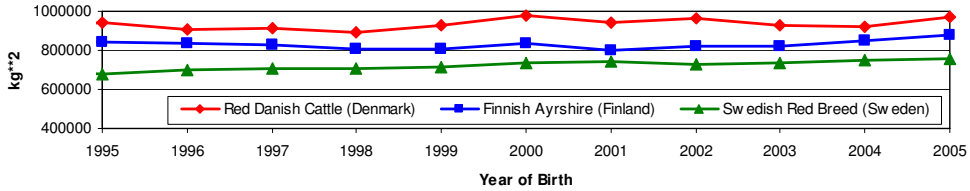


**Figure 1: Re-estimated genetic variance for 2nd lactation milk yield (on 305-d basis) before adjustment for across-county heterogeneity of genetic variance**

**Table 1: Re-estimated genetic standard deviations for milk yield traits before and after adjustment of heterogeneous genetic variances across the countries(= across breeds)[α]**

| | M 1 | M 2 | M 3 | P 1 | P 2 | P 3 | F 1 | F 2 | F 3 |
|---|---|---|---|---|---|---|---|---|---|
| ***Before adjustment*** | | | | | | | | | |
| Red Danish Cattle | 808 | 969 | 1009 | 21.8 | 29.5 | 31.0 | 28.6 | 37.9 | 41.9 |
| Finnish Ayrshire | 769 | 910 | 950 | 20.4 | 27.5 | 29.2 | 29.1 | 37.9 | 41.9 |
| Swedish Red Breed | 713 | 858 | 899 | 21.0 | 27.9 | 29.6 | 27.6 | 35.8 | 39.6 |
| ***After adjustment*** | | | | | | | | | |
| Red Danish Cattle | 739 | 894 | 933 | 21.2 | 28.6 | 30.1 | 27.6 | 36.4 | 40.1 |
| Finnish Ayrshire | 742 | 895 | 936 | 21.0 | 28.6 | 30.5 | 27.4 | 36.2 | 40.2 |
| Swedish Red Breed | 740 | 889 | 931 | 21.3 | 28.3 | 30.0 | 27.7 | 36.0 | 39.8 |

[α]Genetic standard deviation in kg for 305-d milk (M), protein (P) and fat (F) yield for first three lactations (1, 2, 3).

## Conclusion

The applied method, where genetic variances, re-estimated from Mendelian sampling deviations of cows, were used to calibrate the heterogeneous variance adjustment factors was found useful and yielded homogeneous genetic variances across the countries. The method can be applied for any heterogeneous variance adjustment method, which targets on the scaling of the phenotypic observations.

## References

García-Cortés, L.A., Sorensen, D. (2001). *Genet. Sel. Evol.,* **33**, 443-452.

Lidauer, M.H., Emmerling, R., Mäntysaari, E.A. (2008). *J. Anim. Breed. Genet.*, **125**, 3: 147-159.

Lidauer, M.H., Madsen, P., Matilainen, K., Mäntysaari, E.A, Strandén, I., Thompson, R., Pösö, J., Pedersen, J., Nielsen, U.S., Eriksson, J.-Å., Johansson, K., Aamand, G.P. (2009). *Interbull Bulletin*, **40**, 37-41.

Meuwissen, T. H. E., De Jong, G., Engel, B. (1996). *J. Dairy Sci.,* **79**, 310-316.