Estimation of Parameters for Heterogeneous Variance Adjustment on Test-Day Data

Sz. Márkus, E.A. Mäntysaari, I. Strandén, M.H. Lidauer
MTT Agrifood Research Finland, Biotechnology and Food Research, Biometrical Genetics, Jokioinen, Finland

1. Introduction

A multiplicative mixed effects model for heterogeneous variance (HV) correction was described by Meuwissen et al. (1996). Fixed and random effects were scaled within herd-year strata (MEU-1). Based on Pool and Meuwissen's (2000) suggestion Gengler and Wiggans (2001) modified MEU-1 to scale only random effects of the model (MEU-2).

Pena and Ibañez (2002) compared these two methods using simulated lactation yield data sets. Their results showed that MEU-1 is more sensitive for the strata size than MEU-2 and MEU-2 converges faster than MEU-1.

The objective of this study was to validate the two HV adjustment methods by a simulation study on test-day data and to estimate heterogeneity model parameters for test-day data of the Nordic yield evaluation.

2. Materials and methods

2.1 Model

The multiplicative mixed effects model of Meuwissen et al. (1996) was tailored to a test-day model as outlined in Lidauer et al. (2008). In our study, data with milk yield test-day observations were simulated based on a single trait test-day model:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Ch} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e}, \qquad [1]$$

where $\mathbf{y}$ has test-day milk yield observations with a HV structure; vector $\mathbf{b}$ has fixed effects (herd year, herd slope, age at calving, days carried calf, test year-month and lactation curve) and random effects vectors $\mathbf{h}$, $\mathbf{a}$, $\mathbf{p}$, and $\mathbf{e}$ are herd-test-day, additive genetic animal, non-genetic animal, and measurement error, respectively; $\mathbf{X}$, $\mathbf{C}$, $\mathbf{Z}$ and $\mathbf{W}$ are the corresponding design matrices. HV was introduced into the simulated data by stratifying the data in $\mathbf{y}$ into herd $i \times$ test-year-month j strata and scaling of the observation of each stratum with a simulated heterogeneity factor.

2.2 Heterogeneous variance adjustment

The first applied HV adjustment method (MEU-1) has been described by Meuwissen et al. (1996) and scales all effects in the model in proportionality. Stratifying the data in $\mathbf{y}$ by herd and test-year-month ij gives:

$$\mathbf{y}_{ij} = \lambda^{-1}_{ij} \cdot (\mathbf{X}_{ij}\mathbf{b} + \mathbf{C}_{ij}\mathbf{h} + \mathbf{Z}_{ij}\mathbf{a} + \mathbf{W}_{ij}\mathbf{p} + \mathbf{e}_{ij}). \qquad [2]$$

The other applied HV adjustment (MEU-2) was based on Meuwissen et al. (1996) but does not scale fixed effects by the heterogeneous adjustment factor:

$$\mathbf{y}_{ij} = \mathbf{X}_{ij}\mathbf{b} + \lambda^{-1}_{ij} \cdot (\mathbf{C}_{ij}\mathbf{h} + \mathbf{Z}_{ij}\mathbf{a} + \mathbf{W}_{ij}\mathbf{p} + \mathbf{e}_{ij}). \qquad [3]$$

For both methods, adjustment factors λ_{ij} were calculated as

$$\lambda_{ij} = \exp(-0.5 \cdot (\beta_{1j} + \beta_{2ip})), \qquad [4]$$

where β values were estimated from the following *variance model*:

$$s_{ijpk} = \mu + \beta_{1j} + \beta_{2ip} + \varepsilon_{ijpk}, \qquad [5]$$

where s_{ijpk} is a heterogeneity observation for herd $\times$ test-year-month stratum ij, μ is an overall mean; β_{1j} is the fixed effect for test-year-month j, β_{2ip} is the random effect of herd $i \times$ test-period p, where test-period p was either test-year m or test-year-month j; and ε_{ijpk} is the random residual.

2.3 Simulation study

2.3.1 Structure of simulated data

To have a realistic data structure, two real data sets with different herd sizes were sampled from Swedish Red Breed herds of the Nordic test-day model yield evaluation. The real data observations were replaced by simulated observations as will be explained below.

The first data set contained 200 large herds having 483 946 first lactation milk yield observations from a time period of 12 years. The average number of observations per herd-test-day was 18. There were 53 741 animals with records and the pedigree had 104 525 animals.

The second data set contained 200 randomly sampled herds having 267 446 first lactation milk yield observations from a time period of 12 years. The average number of observations per herd-test-day was 10. There were 29 320 animals with records and the pedigree had 64 168 animals.

2.3.2 Simulation of milk yield observations

Simulation of milk yields was based on test-day model [1]. Fixed effects were the BLUE solution from a BLUP run on the sampled real data sets applying the same model. Hence, fixed effects were the same for all simulated sets of observations. Random effects were generated from normal distributions using for each random effect the Cholesky decomposition of the corresponding (co)variance matrix. Milk yield observations were generated by summing the corresponding fixed and random effects, and adding a random error term. The applied (co)variance components were from a variance component study for red breeds (Lidauer et al., 2009).

2.3.3 Simulation of the heterogeneity

Heterogeneous adjustment factors were simulated based on the *variance model* [5]. The model includes a fixed effect and a random effect. BLUE solutions from the *variance model*, when applying the HV adjustment methods on the real data were taken as fixed effect $\tilde{\beta}_{1j}$. Random herd×test-period effect $\tilde{\beta}_{2ip}$ were generated from a within herd first order autoregressive process with a given autoregressive correlation parameter (ρ) and a given strata variance (σ^2_{strata}). Hence, homogeneous milk yield observations of a stratum *ij* were divided by the simulated heterogeneous adjustment factor:

$$\tilde{\lambda}_{ij} = \exp(-0.5\cdot(\tilde{\beta}_{1j} + \tilde{\beta}_{2ip})). \qquad [6]$$

2.3.4 Simulated models

Two different *variance models* were simulated in the study. Both models included a fixed year-month effect, whereas the random effect was either herd × test-year (HY-model), or herd × test-year-month (HTM-model). For each *variance model,* two different autocorrelation parameters were used in simulation. The true parameter values in the simulations were 0.6 and 0.9 in case of the HY-model and 0.93 and 0.98 in case of the HTM-model. For each of the described model alternatives, two different strata variances were used: either 0.1 or 0.5. Each model alternatives was applied for both data sets. Five replicates were simulated for each of the 16 model × data alternatives.

2.3.5 Estimation of the variance model parameters

Simulated data was solved with both HV adjustment methods and parameters of the *variance model* were estimated with derivative free REML algorithm.

2.4 Estimation of the variance model parameters from real data

An extension of the large herds data set in point 2.3.1 was used for the multiple trait model analyses. The data set contained 483 946 first and 306 035 second lactation milk, protein and fat yield records. The applied multiple trait test-day model had the same effects as given in model [1] for each trait.

3. Results and discussion

3.1 Simulation study

In case of HY-model both methods estimated the autocorrelation parameters and the herd-year variance with similar precision for the large herds and randomly sampled herds data set (Table 1).

Table 1. Estimates of autocorrelation parameter (ρ) and within herd × test-year variance (σ^2_{hy}) by heterogeneous variance adjustment method (MEU-1 or MEU-2), true within herd × test-year variance (σ^2_{hy}= 0.1 or 0.5), true autocorrelation (ρ= 0.6 or 0.9) and herd size. Averages are from 5 simulated replicates.

True	ρ	0.6		0.9	
values	σ^2_{hy}	0.1	0.5	0.1	0.5
Large herds					
MEU-1	ρ	0.59	0.59	0.88	0.89
	σ^2_{hy}	0.11	0.50	0.11	0.52
MEU-2	ρ	0.61	0.59	0.91	0.90
	σ^2_{hy}	0.10	0.48	0.10	0.49
Random herds					
MEU-1	ρ	0.58	0.60	0.89	0.89
	σ^2_{hy}	0.11	0.51	0.11	0.50
MEU-2	ρ	0.61	0.61	0.92	0.90
	σ^2_{hy}	0.10	0.50	0.10	0.48

In case of HTM-model both methods gave closely the same estimates of the autocorrelation parameter. Method MEU-2 gave closer estimates for the herd-test month variance especially when the simulations were based on the randomly sampled herds data set (Table 2).

In general there were no large differences between the estimates by the two methods. However, MEU-2 showed better convergence and estimated the strata variance better, when the strata size was small.

Table 2. Estimates of autocorrelation parameter (ρ) and within herd × test-year-month variance (σ^2_{htm}) by heterogeneous variance adjustment method (MEU-1 or MEU-2), true within herd × test-year-month variance (σ^2_{htm}= 0.1 or 0.5), true autocorrelation (ρ= 0.93 or 0.98) and herd size. Averages are from 5 simulated replicates.

True	ρ	0.93		0.98	
values	σ^2_{htm}	0.1	0.5	0.1	0.5
Large herds					
MEU-1	ρ	0.93	0.93	0.98	0.98
	σ^2_{htm}	0.11	0.54	0.11	0.52
MEU-2	ρ	0.94	0.93	0.99	0.98
	σ^2_{htm}	0.09	0.49	0.09	0.48
Random herds					
MEU-1	ρ	0.94	0.94	0.99	0.98
	σ^2_{htm}	0.12	0.55	0.13	0.60
MEU-2	ρ	0.95	0.92	0.99	0.98
	σ^2_{htm}	0.09	0.49	0.10	0.47

3.2 Estimation of the variance model parameters from real data

When the *variance model* parameters were estimated using the original large herds data and HY-model, the MEU-2 method gave higher estimates for the autocorrelation parameter. The estimates for the strata variances were similar by both methods. MEU-2 gave higher estimates of the residual variances (Table 3).

In case of HTM-model, MEU-2 gave smaller estimates for the autocorrelation parameters. Same as for the HY-model, MEU-2 estimated similar strata variance as MEU-1 and higher residual variances than MEU-1 (Table 4).

Table 3. Results of the multi-trait estimation of the herd × test-year variance model parameters for Nordic test-day model.

		1. lactation			2. lactation		
		milk	protein	fat	milk	protein	fat
MEU-1	ρ	0.75	0.74	0.72	0.80	0.77	0.78
	σ^2_{hy}	0.07	0.08	0.08	0.07	0.08	0.07
	σ^2_{ε}	0.35	0.36	0.39	0.52	0.54	0.72
	$\sigma^2_{hy}/\sigma^2_{\varepsilon}$	0.22	0.23	0.21	0.13	0.14	0.09
MEU-2	ρ	0.84	0.85	0.83	0.77	0.78	0.82
	σ^2_{hy}	0.07	0.07	0.07	0.07	0.07	0.06
	σ^2_{ε}	1.46	1.55	1.84	1.12	1.24	1.59
	$\sigma^2_{hy}/\sigma^2_{\varepsilon}$	0.05	0.05	0.04	0.07	0.06	0.04

ρ: *autocorrelation parameter;* σ^2_{hy}: *herd × test-year variance;* σ^2_{ε}: *residual variance*

Table 4. Results of the multi-trait estimation of the herd × test-year-month variance model parameters for Nordic test-day model.

		1. lactation			2. lactation		
		milk	protein	fat	milk	protein	fat
MEU-1	ρ	0.98	0.98	0.98	0.98	0.98	0.98
	σ^2_{htm}	0.08	0.09	0.09	0.07	0.08	0.08
	σ^2_{ε}	0.31	0.33	0.38	0.40	0.41	0.57
	$\sigma^2_{htm}/\sigma^2_{\varepsilon}$	0.26	0.27	0.22	0.17	0.19	0.13
MEU-2	ρ	0.93	0.95	0.94	0.95	0.95	0.93
	σ^2_{htm}	0.10	0.09	0.09	0.09	0.09	0.08
	σ^2_{ε}	1.13	1.39	1.66	0.91	1.00	1.58
	$\sigma^2_{htm}/\sigma^2_{\varepsilon}$	0.09	0.06	0.06	0.10	0.09	0.05

ρ: *autocorrelation parameter;* σ^2_{htm}: *herd × test-year-month variance;* σ^2_{ε}: *residual variance*

4. Conclusions

Both tested heterogeneous variance adjustment methods showed the ability to estimate heterogeneity parameters for the Nordic test-day data. However, parameter estimates depended on the applied adjustment method and variance model.

To make a choice between the two heterogeneous variance adjustment methods more investigation will be necessary about the effect of the methods on the estimated breeding values.

Acknowledgement

We thank Nordic Cattle Genetic Evaluation for providing the data and for financial support.

References

Gengler, N., Wiggans, G.R. (2001). Heterogeneity in (Co)Variance Structure of Test-Day Yields. Interbull Bulletin. No.27, 179-183.

Lidauer, M.H., Emmerling, R., Mäntysaari, E.A. (2008). Multiplicative Random Regression Model for Heterogeneous Variance Adjustment in Genetic Evaluation for Milk Yield in Simmental. J. Anim. Breed. Genet. 125:147–159

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). Estimation of Variance Components for Nordic Red Cattle Test-Day Model: Bayesian Gibbs Sampler vs. Monte Carlo EM REML. Interbull meeting. Barcelona, Spain.

Madsen, P., Jensen, J. (2008). An User's Guide to DMU (Version 6, release 4.7). University of Aarhus.

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. (2000). Reduction of the Number of Parameters Needed for a Polynomial Random Regression Test Day Model. Livest. Prod. Sci. 64:133-145

Pena, J., Ibañez, M.A. (2002). Estimation of Parameters for Three Methods of Heterogeneous Variance Adjustment on 305 Days Lactation Data. Interbull Bulletin. No.29, 88-90.