Simulation Study on Heterogeneous Variance Adjustment for Observations with Different Measurement Error Variance

Pitkänen, T.[1], Mäntysaari, E. A.[1], Nielsen, U. S.[2], Aamand, G. P[3]., Madsen[4], P. and Lidauer, M. H. [1]

[1] *MTT Agrifood Research, Biotechnology and Food Research, Genetic Research, Finland, Jokioinen*
[2] *The Knowledge Centre for Agriculture, Cattle, Denmark*
[3] *Nordic Cattle Genetic Evaluation, Denmark*
[4] *Centre for Quantitative Genetics and Genomics, Aarhus University, Denmark*

Abstract

The Nordic Holstein yield evaluation model describes all available milk, protein and fat test-day yields from Denmark, Finland and Sweden. In its current form all variance components are estimated from observations recorded under conventional milking systems. Also the model for heterogeneity of variance correction is developed for the same observations. As automated milking systems are becoming more popular the current evaluation model needs to be enhanced to account for the different measurement error variances of observations from automated milking systems. In this simulation study different models and different approaches to account for heterogeneous variance when observations have different measurement error variances were investigated. Based on the results we propose to upgrade the currently applied models and to calibrate the heterogeneous variance adjustment method to yield same genetic variance for both milking systems.

Key words: Animal model, heterogeneous variance correction, automated milking system

Introduction

The Nordic Holstein evaluation model uses a multiplicative mixed effect model (Meuwissen 1996, Lidauer *et al.* 2008) to correct for heterogeneous variance (HV). In the past majority of the test-day records were obtained from conventional milking systems (CMS) but during recent years many farms have changed to automated milking systems (AMS). Test-day observations measured from AMS increased from 14% in 2008 to 29% in 2011 for Danish Holstein. The current evaluation model and HV correction does not take into account that observations from AMS have a different measurement error variance (Pitkänen *et al.*, 2012).

The objective of this simulation study was to benchmark different models that takes milking system into account. We also propose a method for calibrating multiplicative mixed effect model to handle different milking systems.

Materials and Methods

Simulation of data

Data from sampled Danish Holstein herds were used as a basis for the simulations. In total 600 herds having a complete data of test-day observations over a twelve years period (2001-2011) were randomly sampled among all herds. For simplicity, only observations for first lactation milk, protein and fat yields were used. For the simulation 40% of the herds were considered as AMS and 60% as CMS herds(Table 1).

Table 1. Summary of data.

	AMS	CMS	Total
Herds	240	360	600
Animals	136 002	109 021	245 031
Records	1 102 550	905 032	2 007 582
N milk	1 094 497	899 015	1 993 512
N protein	1 094 497	899 015	1 993 512
N fat	1 093 469	898 192	1 991 661

The original observations of the data were replaced with simulated observations. The simulation was done in three steps. First, test-day observations without heterogeneity were simulated for stratum *i* based on a multiple trait random regression model:

$$\mathbf{Y^*_i} = \mathbf{hy + ym + lcurve + clys + htd + pe + gen + e},$$

where fixed effects are

hy	herd year,
ym	year month,
lcurve	lactation curve,
clys	calving year season,
	and random effects are
htd	herd test-day,
pe	permanent environment,
gen	genetic animal effect,
e	measurement error.

Variance components used for simulating observations were obtained from the study presented by Pitkänen *et al.* (2012). The compiled heritabilities for 305 day milk, protein and fat yields were 0.39, 0.35, 0.38 for both milking systems even milking system specific measurement error covariance matrices (Pitkänen *et al.*, 2012) were applied.

After simulating test-day observations a heterogeneity factor l_i was simulated for each herd×test month stratum *i* applying the following multiple trait model:

$$\xi_i = \mathbf{b_{i1}} + \mathbf{b_{i2}} + \boldsymbol{\varepsilon_i}$$
$$\mathbf{l_i} = \mathbf{c_m}\ \mathbf{exp(\ -0.5\ \xi_i\ )},$$

where $\mathbf{b_{i1}}$ is a fixed year×month effect, $\mathbf{b_{i2}}$ is a random herd×year effect, $\boldsymbol{\varepsilon_i}$ is a residual effect and $\mathbf{c_m}$ is a scaling factor specific to the milking system. For CMS observations $\mathbf{c_m}$ was 1.0 for each trait and for AMS 0.95 for milk, 1.0 for protein and 1.1 for fat. For the random herd×year effect a first order autoregressive process (AR(1)) was assumed within herd×year classes and considering a correlation between herd×year class effects across traits. The applied correlations between traits were ?? between milk and protein, ?? between milk and fat and ?? between protein and fat.. Residual effects $\boldsymbol{\varepsilon_i}$ for different traits were also correlated within stratum.

Finally, testday observations $\mathbf{Y_i}$ with HV were obtained by:

$$\mathbf{Y_i} = \mathbf{Y^*_i} / \mathbf{l_i}.$$

Ten replicate data sets were simulated.

Solving the multiplicative mixed effect model

The multiplicative mixed effect model can be formulated as

$$\boldsymbol{Y_i}\ \boldsymbol{l_i} = \boldsymbol{X_i b} + \boldsymbol{Z_i u} + \boldsymbol{e_i} \quad [1]$$
$$\boldsymbol{\xi_i} = \boldsymbol{b_{i1}} + \boldsymbol{b_{i2}} + \boldsymbol{\varepsilon_i} \quad [2]$$
$$\boldsymbol{l_i} = \boldsymbol{exp(\ -0.5(\hat{b}_{i1} + \hat{b}_{i2})\ )} \quad [3]$$

where in each cycle mean model [1] and variance model [2] are iterated until convergence. Heterogeneity observations $\boldsymbol{\xi_i}$ for iteration cycle q are calculated after mean model iterations as:

$$\xi_i = \left(v_i^{[q-1]}/u_i^{[q-1]}\right) + \hat{\beta}_1^{[q-1]} + \hat{\beta}_{2i}^{[q-1]},$$
$$v_i^{[q-1]} = \frac{\boldsymbol{y}_i^T \hat{\boldsymbol{e}}_i^{[q-1]} \lambda_i^{[q-1]}}{2\sigma_{em}^2} - \frac{n_i - \tilde{r}_i}{2},$$
$$u_i^{[q-1]} = \frac{\boldsymbol{y}_i^T \boldsymbol{y}_i\ \lambda_i^{2\,[q-1]}}{4\sigma_{em}^2} + \frac{n_i - \tilde{r}_i}{2},$$

where $\widehat{\sigma}_{em}^2$ is a milking system specific base variance for stratum *i*. More detailed description of the solving process can be found from Lidauer et. al (2008).

Applied models

We fitted four different models to the simulated data. Apart from a control model all other models used different variance components than those used in the simulation. It was anticipated that under real situation the true variance components are unknown, which may affect adjustment for HV.

Control model

The control model was exactly the same model that was used for simulating observations **Y***. The control model was used for testing

simulation and it was assumed to yield best results.

HVnoMS

This model applied HV correction but neither the mean model nor the variance model had a milking system interaction. The measurement error co-variance matrix for CMS observations was applied. The model represents the current evaluation model.

HVMS

This was the same model as *HVnoMS* but with a milking system interaction in both mean model and variance model. In the mean model fixed effect **YM** is replaced with **YM×MS** interaction and the measurement error covariance matrix was milking system specific. In the variance model fixed year×month effect is replaced by year×month×milking system effect.

Calibrated HMVS

The model formulation is the same as HVMS. For this model σ^2_{em} needed for the calculation of ξ_i is calibrated so that genetic variances of animals from different milking systems are the same. For the calibration the multiplicative mixed effect model was solved several times. After each run a full model sampling approach (Lidauer *et al.*, 2008) was used to calculate genetic variance of 305 day yields for AMS, s^2_{AMS}, and CMS cows, s^2_{CMS}. The ratio r= s^2_{AMS} / s^2_{CMS} was calculated and σ^2_{em} base variance for CMS stratums was updated by multiplying it with r. The whole process was repeated 5-10 times until ratio r is close to 1.

Comparison of models

The fit of the models was assessed by calculating (1) genetic variance of 305 day breeding values for AMS and CMS cows by the full model sampling approach and compared them to the true simulated variances, (2) correlation of true and estimated 305-day breeding values and (3) proportion of AMS cows in top1000 list. All results are calculated for cows from the latest two birth year class having more than 5 observations. Number of cows was 11420 for AMS 8109 for CMS. All presented results are mean of ten replicates.

Results

Ratios of genetic variance of estimated and true 305 day breeding values are presented in the Table 2. Obtained re-estimated genetic variances from the control model were nearly same as simulated true variances. The model ignoring the milking system (HVnoMS) yielded higher genetic variance in AMS for milk and protein and lower for Fat. Inclusion of milking system specific effects (HVMS) reduced the differences in genetic variances between milking systems.. By definition, the calibrated HMVS model yielded same genetic variances for both milking systems.

There were no big differences in correlations between true and estimated breeding values among HV models (Table 3). HVnoMS resulted lower correlation in FAT compared to all other HV models. The control model yielded the highest correlations as expected

The proportion of AMS cows in the top1000 list differed between models. For the control model the proportion was the same as simulated. HVnoMS resulted in protein and fat top1000 list fewer cows than what has been simulated. The calibrated model yielded for all three traits almostnumber of AMS cows in the top 1000 list as simulated.

Discussion

For the simulated true model we used a lower measurement error variance for milk and protein yield observations and a larger for fat yield observations from AMS compared to observations from CMS. When HV adjustment calculates heterogeneity of observations for milk and protein yields for an AMS stratum it will on average “see” too small heterogeneity because the base variance, to where the heterogeneity is compared, is from CMS which was higher. Therefore HV correction tries to increase variance of AMS milk and protein

observations. For fat it behaves in the opposite because AMS observations had higher measurement error variance than CMS observations. Even the re-estimated genetic variances in HVnoMS were higher for milk and protein yield we did not see in the top1000 cows list an over representation of AMS cows. In fact in the protein list there were less AMS cows than expected. Distribution of cows in fat top list was consistent with the genetic variances as there were too many CMS cows. Even there are differences in genetic variances and top1000 lists between different milking systems we found only very small differences in correlations between true and estimated 305-day breeding values.

When milking system interactions were considered (HVMS) the results looked better. The re-estimated genetic variances for AMS and CMS cows were close to each other but not yet the same. However, there were still a bit too many AMS cows in protein and fat top 1000 cow lists.

For the model with the calibrated HV correction we adjusted the base variance for CMS observations until same genetic variance was obtained from both systems. By doing so, results were closest to the simulated ones. We also saw that genetic variance for AMS cows stayed at same level than in HVMS and genetic variance of CMS cows was changed towards AMS. For the implementation of the approach into the Nordic routine evaluation it is advisable to adjust the base variance for AMS and keeping the base variance for CMS constant. This should yield re-estimated genetic variance closer to the original ones. It is worth to note that calibration of the base variances has to be done only once during model developing work.

Conclusions

Calibration of the heterogeneous variance adjustment method was found to be suitable for correcting heterogeneous variance when different milking systems is present at the data.

Not accounting for different milking systems resulted in too large genetic variance for milk and protein yield breeding values of AMS cows and too low genetic variance for fat yield breeding values.

References

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. *Journal of Animal Breeding and Genetics 125,* 147–159.

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.

Pitkänen, T., Mäntysaari, E.A., Nielsen, U.S., Aamand, G.P., Madsen, P. & Lidauer, M.H. 2012. Measurement error variance of test-day observations from automatic milking systems. *Interbull Bulletin 46,* 103-108.

Table 2. Ratios of genetic variance of estimated and true 305 day breeding values for cows in automated milking system (AMS) and conventional milking system (CMS).

	MILK		PROTEIN		FAT	
MODEL	**AMS**	**CMS**	**AMS**	**CMS**	**AMS**	**CMS**
Control (Y*)	1.01	1.01	1.00	1.02	1.00	1.01
HVnoMS	1.14	0.81	1.02	0.78	0.75	0.89
HVMS	0.70	0.78	0.69	0.76	0.90	0.91
Calibrated HVMS	0.69	0.69	0.69	0.69	0.91	0.90

Table 3. Correlations between true and estimated 305 day breeding values calculated within milking system, (AMS = automated milking system, CMS=conventional) and over all data (ALL).

	MILK			PROTEIN			FAT		
	AMS	**CMS**	**ALL**	**AMS**	**CMS**	**ALL**	**AMS**	**CMS**	**ALL**
Control (Y*)	0.79	0.79	0.79	0.77	0.77	0.77	0.74	0.75	0.75
HVnoMS	0.77	0.77	0.77	0.75	0.75	0.75	0.72	0.74	0.72
HVMS	0.77	0.77	0.77	0.75	0.76	0.75	0.74	0.74	0.74
Calibrated HVMS	0.77	0.77	0.77	0.75	0.76	0.75	0.74	0.74	0.74

Table 4. Percentage of cows in automated milking system in top1000 list.

	MILK	**PROTEIN**	**FAT**
SIMULATED	58	59	58
Control (Y*)	58	59	58
HVnoMS	59	55	40
HVMS	58	61	63
Calibrated HVMS	59	60	59