Estimation of Parameters for Three Methods of Heterogeneous Variance Adjustment on 305 Days Lactational Data

***J.Pena*[(1)], *M.A. Ibañez*[(2)]**
[(1)]*Departamento Técnico. CONAFE*
[(2)]*Departamento de Estadística. ETSIA*

Introduction

Meuwissen *et al.* (1996) described a method (method **MEU-1**) for joint estimation of breeding values and phenotypic heterogeneous where fixed and random effects are scaled by a factor exp(γ_i/2) and a linear model is assumed for γ_i. Robert-Granié *et al.* (1999) proposed a linear mixed model assuming heterogeneous residual variances and known constant variance ratios (method **ROBERT**). Both applied an autocorrelation model on the heterogeneity factors γ_i and require knowledge of its parameters, autocorrelation and variance of herd-year.

Pool and Meuwissen (2000) suggested that method of Meuwissen *et al.* (1996) could be applied without scaling fixed effects (method **MEU-2**), similar to Robert-Granié *et al.* (1999), and Gengler and Wiggans (2001) applied this method for correcting heterogeneity of variances in a Test-Day Model.

The objective of this study was to estimate parameters of the model on the heterogeneity factors γ_i with the three methods described before and using simulated data sets with different data structures and a subset of Spanish data for analyzing how accurately this parameters are estimated.

Materials and Methods

Simulated Data

Four different data structures were simulated, as described in Table 1. As in Meuwissen *et al* (1996), heterogeneity factors γ_i were simulated and analysed with an overall mean and a herd-year mean. Records were simulated with a herd-year mean and unrelated individual animal effects. All effects were simulated in a common scale and then scaled to variability level in the different herd-years. Heritability was set to 0.25, overall mean of heterogeneity factors was 13.3, variance of herd-year 0.14 and autocorrelation 0.5. For simplicity, when considering repeated measures within animals, all of them are in the same herd-year.

Ideally, several replicates should have been analyzed for each data structure but in this study only two replicates were obtained. Data sets 1, 2 and 3 are extracted from one big simulated data set of 1000 herds, 10 years and 10 observations per animal. Data set 4 was simulated independently, because was not planned in the initial simulation.

Real data set

Data from a province of northeast Spain were used. It included 150 herds and 14 years of data with an average of 52 lactations per herd-year.

Trait analysed was Kg milk in 305 days, with all lactations projected to 305 days. First five lactations were included. A repeatibility animal model was applied, including fixed effects of month of calving within parity and age within parity and a comparison group defined as Herd-Year-Imported-Parity-Season, build in a flexible manner depending of how many lactations are available. A minimum of 3 lactations in a herd-year was required and average comparison group size was 13. Phantom groups for unknown parents were based on year of birth. Heritability used was 0.28 and repeatibility 0.50. Weights were given based on days in milk.

Heterogeneity factors model include fixed effects of year and parity and random herd-year.

Table1. Simulated data sets with 100 herds, 10 years and 10 animals per herd-year.

	Simulated Data set 1	**Simulated Data Set 2**	**Simulated Data Set 3**	**Simulated Data Set 4**
Number of observations per animal	1	3	10	30
Final number of observations per herd-year	10	30	100	300
Final number of animals	10 000	10 000	10 000	100 000
Final number of observations	10 000	30 000	100 000	300 000

Solving strategy

Effects were solved by IOD and at the end of each IOD iteration scaling factors were estimated. Parameters of the heterogeneity factors model were also estimated in each iteration starting from the first, except for **MEU-1** that needed to use initial values or otherwise variance would converge to zero.

Results and Discussion

Simulated Data Set 1

Methods **MEU-2** and **ROBERT** gave estimates close to true values for autocorrelation and variance of herd-year, but method **MEU-1** show a pattern of autocorrelation getting greater than one and had to be bounded at 0.995 but variance converged to zero.

Simulated Data Set 2

Parameters only were estimated with method **MEU-1**, but again autocorrelation had to be bounded at 0.995 and variance converged to 0.008, far below the true variance.

Simulated Data Set 3

MEU-1 still overestimated autocorrelation but its estimate is much closer to the true values than with data sets with 10 or 30 observations per herd-year. There is a little underestimation of the variance.

Results are shown in Table 2.

Table 2. Estimates of autocorrelation and variance of herd-year with different data sets.

Method	**Simulated Data set 1** (HY size=10)		**Simulated Data Set 2** (HY size=30)		**Simulated Data Set 3** (HY size=100)		**Simulated Data Set 4** (HY size=300)		**REAL DATA SET** (HY size=52)	
	ρ	σ^2_{hy}	ρ	σ^2_{hy}	ρ	σ^2_{hy}	ρ	σ^2_{hy}	ρ	σ^2_{hy}
MEU-1	0.995	0	0.995	0.008	0.656	0.131	0.522	0.136	0.995	0.731
MEU-2	0.525	0.153	-	-	0.543	0.154	-	-	0.903	0.124
ROBERT	0.551	0.141	-	-	0.545	0.153	-	-	0.918	0.122

When analysing other data set from an independent simulation with 1000 herds, 10 years, 10 animals by year and 100 observations by herd-year, estimated autocorrelation was 0.636 and variance 0.116, close to estimates obtained with data set 3. So, the overestimation of autocorrelation was not due to low number of herds. With this data set parameters estimated with **MEU-2** and **ROBERT** were exactly equal and mean and standard deviation of solution of effects were equal to third decimal digit.

Simulated data Set 4

With 300 observations per herd-year estimated autocorrelation and variances with method **MEU-1** were close to true values.

Real Data Set

MEU-1 needed to get autocorrelation fixed at 0.995, but **MEU-2** and **ROBERT** obtained similar values for autocorrelation and variances.

Convergence

In all simulated data sets **MEU-1** needed the biggest number of iterations (between 900 and 1600) and **MEU-2** and **ROBERT** between 14 and 33, but in the real data set, **MEU-2** did converge very much slower than **ROBERT**, that converged in 144 iterations.

Conclusions

It seems **MEU-1** needs a high number of observations by herd-year for obtaining parameters near the true values. When herd-year size is not big enough, there is a trend to overestimate autocorrelation and underestimate variance.

Methods **MEU-2** and **ROBERT** converge faster and gives estimates near to the true values, but with the real data set **MEU-2** did not converge as well as **ROBERT.**

Estimates obtained with **MEU-1** may depend also of solving strategy used, as it was needed to fix initial values for autocorrelation and variance of herd-year during first iterations.

A more detailed simulation with several replicates for each data structure is needed for obtaining more precise conclusions about comparison of methods. Also, and most important, a detailed comparison between solutions of effects obtained with different parameters.

In the real data set it may be of interest to introduce some other fixed effects in the model on the heterogeneity factors, i.e. size of herd-year, similarly to Gengler and Wiggans (2001).

References

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

Robert-Granié, C., Bonaiti, B., Boichard, D. & Barbat, A. 1999. *Livest. Prod. Sci. 60,* 343-357.

Pool, M.H. & Meuwissen, T.H.E. 2000. *Livest. Prod. Sci. 64,* 133-145.

Gengler, N. & Wiggans, G.R. 2001. *Interbull Open Meeting.* Budapest.