Validation and Comparison of Methods to Estimate (Co)Variance Components for MACE

Per Madsen[1], Morten Kargo Sørensen[1], and Thomas Mark[2]
[1] *Danish Institute of Agricultural Sciences, Research Centre Foulum, P.O. Box 50, DK-8830 Tjele, Denmark.*
[2] *Interbull Centre, SLU, Box 7023, 750 07, Sweden*

Introduction

Methods to estimate (co)variance components across countries include the methods described by Sigurdsson *et al*. (1996), Klei & Weigel (1998) and Madsen *et al*. (2000). The first two EM-REML based procedures have earlier been validated on simulated data for traits with a relative high heritability (Sigurdsson *et al*., 1996; Klei & Weigel, 1998). They have also been compared for low heritability traits on field data (Mark *et al*., 2000). Based on these studies it can be concluded that both methods yield unbiased results, when there is sufficient information about the genetic components in the across country inference (well-connected subset of bulls and an intermediate heritability). However, it was also shown that the method of Sigurdsson *et al.* (1996) can underestimate the true value and yield substantially lower estimates compared with the method of Klei & Weigel (1998), when the data are not well connected across countries and heritability is low.

Madsen *et al.* (2000) introduced an AI-REML based method to estimate (co)variances for MACE. This method yielded similar estimates as obtained by the method of Klei & Weigel (1998) on mastitis and milk somatic cell field data. However, it has not previously been validated on simulated data.

The aim of this study was to investigate and compare the performance of these three REML based methods on simulated data for traits of low and moderate heritability with few genetic ties among the simulated populations. Furthermore, we wanted to validate the method of Madsen *et al*. (2000) to produce (co)variances for use in multi-trait-multi-country evaluations, where residual co-variances cannot necessarily be assumed to be zero.

Material and Methods

Simulation

Data were simulated using the stochastic simulation program described by Sørensen *et al*. (1999), and further developed to handle multiple populations by Nielsen *et al.* (2001). This program simulates all individuals in the active breeding populations (ABP = extended nucleus). The remaining parts of the populations are used for progeny test of bulls. For each cow in the ABPs, one lactation was simulated. For test bulls, a daughter yield deviation (DYD) was simulated assuming random mating and based on the assumed progeny group size (Table 1). The mating strategy used in the ABPs was assortative based on predicted breeding values and their economic values given in Table 1. In order to avoid close inbreeding, sire – daughter, full-sib and half-sib matings were not allowed.

For the present study, four populations each evaluating for the same three traits: milk yield (MI), somatic cell (SC), and clinical mastitis (CM), were simulated. Population size, breeding structure and breeding goals are described in Table 1, and these figures roughly reflect the current situation for the four Nordic Red breeds.

Heritabilities and genetic and phenotypic correlations used in the simulation are in Table 2. For cows, only MI was simulated, while both MI, SC and CM were simulated for the tested bulls.

Predictions of breeding values were carried out across populations using a multi-trait Animal Model (AM). For MI, a four trait AM was used,

where MI in each population was treated as a separate trait. For SC and CM, a 8 trait AM was used, were SC and CM in each population was treated as separate traits. In both cases, the true (simulated) (co)-varaince components were used. This corresponds to a perfect international evaluation where breeding values can be used across countries. The breeding scheme was simulated for 20 year.

In order to investigate the effect of selection on the estimated (co)variance components, a dataset without selection was simulated. This simulation was performed by using the pedigree structure from the dataset with selection, and generated new genotypic and phenotypic values for all individuals. Genetic trends based on the simulated breeding values for the cow populations for the two datasets are in Figure 1.

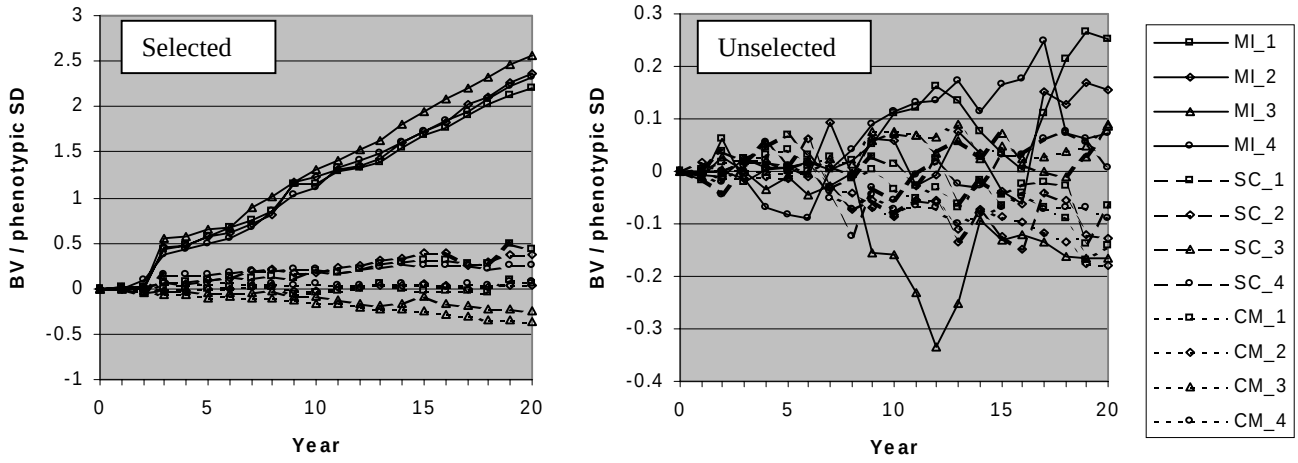
Table 1. Population size and breeding goal for the simulated populations. Economic weights are given as economic weight per phenotypic standard deviation unit.

	Population			
	1	2	3	4
Population size	60.000	200.000	300.000	300.000
Active population (ABP) size	6.000	20.000	30.000	30.000
Bull tested/year	70	158	140	112
Proven Bulls sel./year	8	16	20	20
Daughter group size, Test bulls	80	120	200	250
Economic value MI	1	1	1	1
Economic value SC	0	0	0	0
Economic value CM	1.4	1.4	0	2

Table 2. Heritabilities (on diagonal) genetic correlations (above diagonal) and residual correlations (below diagonal) for the simulations traits.

POP	TRAIT	Population											
		1			2			3			4		
		MI	SC	CM	MI	SC	CM	MI	SC	CM	MI	SC	CM
1	MI	**.25**	-.10	-.30	.93	-.08	-.25	.93	-.08	-.25	.93	-.08	-.25
	SC	.12	**.12**	0.70	-.08	.85	.50	-.08	.85	.50	-.08	.85	.50
	CM	.07	.27	**.04**	-.25	.50	.70	-.25	.50	.70	-.25	.50	.70
2	MI				**.25**	-.10	-.30	0.90	-.08	-25	0.90	-.08	-.25
	SC	0			.12	**.12**	0.70	-.08	.85	.50	-.08	.85	.50
	CM				.07	.27	**.04**	-.25	.50	.70	-.25	.50	.70
3	MI							**.25**	-.10	-.30	0.90	-.08	-.25
	SC	0			0			.12	**.12**	0.70	-.08	.85	.50
	CM							.07	.27	**.04**	-.25	.50	.70
4	MI										**.25**	-.10	-.30
	SC	0			0			0			.12	**.12**	0.70
	CM										.07	.27	**.04**

Figure 1. Genetic trend based on simulated breeding values in the simulated cow populations with and without selection (note the difference in scale).



Links between populations

The genetic links between the four populations expressed as number of common bulls and common ¾ sib groups as well as the mixing of genetic material between the populations are shown in Table 3. Genetic links were slightly stronger compared to the actual Nordic Red populations.

The genetic composition of the four populations in year 20 showed that population 1 only had 20% of its genetic material originating from the base population 1, while 60% originated from base population 2.

Table 3. Genetic links between the simulated populations.

P O	Common bulls and ¾ sib groups[1)]				Origin of germ plasm in percent All animals (year 1-20)[2)]				Year 20[2)]			
	1	2	3	4	1	2	3	4	1	2	3	4
1	1205	40	8	16	57.1	31.7	4.6	6.7	20.8	60.2	4.6	14.4
2	72	2835	21	31	5.8	77.6	7.7	8.8	3.6	68.5	11.9	16.0
3	19	47	2746	15	.0	7.8	81.7	9.6	.7	5.6	77.4	16.3
4	38	61	34	2329	1.3	8.8	8.4	81.5	.6	14.9	13.1	71.4

[1)] Diagonal # proven bulls, below diagonal # common bulls, above diagonal # ¾ sib groups
[2)] Rows correspond to actual population, columns correspond to population of origin.

Prediction of breeding values for parameter estimation

After year 20, new DYDs for all proven bulls were simulated assuming varying progeny group size to reflect the situation in practice where progeny group size varies both for young bulls and for proven bulls selected for further use. For bulls not selected for further use the daughter group size shown in Table 1 was used as the average group size with a standard deviation (SD) of 23, 36, 63 and 80 for the four population respectively. For bulls selected for further use, the new DYDs was simulated for a larger progeny group to imitate that, these bulls would have a second crop of daughters. The new progeny group size used depended on whether the bull was used in the population of birth or not. In the population of birth, the average progeny group size used was 5000 with a SD of 2000. If used in a foreign population, the average progeny group size used was 2500 with a SD of 1000. In all cases, a lower limit of 10 was imposed on the progeny

group size. The new DYDs and the original cow performance records were used for prediction of breeding values in single trait Animal Model within populations.

Predicted breeding values from both datasets were subsequently uni-variately and separate for each population deregressed as described by Jairath *et al.* (1998).

Estimation model and methods

The models for parameter estimation were based on the MACE model of (Schaeffer & Zhang, 1993):

$$\mathbf{y = Cc + ZQg + Zs + e}$$

where **y** is a vector of de-regressed proofs or DYDs, **c** is a vector of fixed country effects, **g** is a vector of phantom group effects, **s** is a vector of random bull effects, and **e** is a vector of random residuals. **C, Z,** and **Q** are design matrices relating de-regressed proofs to countries and bulls, and relating bulls to phantom groups respectively.

Following the same ideas as Klei & Weigel, (1998) where unknown parents were assigned to phantom parent groups on a within country basis, the MACE model becomes:

$$\mathbf{y = ZQf + Zs + e}$$

with the following distributional properties:

$$\begin{pmatrix} \mathbf{y} \\ \mathbf{s} \\ \mathbf{e} \end{pmatrix} \sim \mathbf{MVN}\left\{ \begin{pmatrix} \mathbf{ZQf} \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \mathbf{ZGZ'+R} & \mathbf{GZ'} & \mathbf{R} \\ \text{Symm.} & \mathbf{G} & \mathbf{0} \\ & & \mathbf{R} \end{pmatrix} \right\}$$

where **f** is a vector of phantom group + country effects, $\mathbf{G=G_0 \otimes A}$ is the (co)-variance matrix among elements in **s, A** is the additive relationship matrix**,** and **R** is the residual (co)-variance matrix

(Co)variance components were estimated with the procedures and trait /population combinations as shown in Table 4.

Results and Discussion

The estimated genetic correlations for the dataset with selection are in Table 5. Comparing the first four figures in each cell with the values used in the simulation (Table 2), it can be seen that the estimates in general are biased downwards. For MI (h^2 =.25) the differences between the estimation methods are minor. For SC (h^2 =.12) and CM (h^2 =.04) the downward biases is larger, and again with only minor differences between the methods. The AI-REML procedure also gives asymptotic standard errors of the estimated correlations. By comparing the standard errors in Table 5 with the genetic links in Table 3 it can be seen that the standard error of the estimates describe the strength of the genetic links well.

In order to understand the reason for the downward bias, an analysis with AI-REML based on DYDs was performed (analyses AI-12D). The results from the analyses are also in Table 5 (row five in each cell). These estimates are in close agreement with the values used in the simulation. Analyses on DYDs with the two EM-REML procedures gave estimates (not shown) in close agreement with those obtained by the AI-REML procedure. This could indicate that the deregression procedure did not work as expected for this data. One problem in the deregression could be that it did not entirely account for genetic trend because it only uses SIRE-MGS relationship.

To investigate this, the dataset without selection was analyzed in the same manner as the dataset with selection. The results from these analyses are in Table 6. Again comparing the estimated genetic correlations with the values used in the simulations (Table 2), it can be seen that all methods perform well for MI with estimates close to the simulated ones. For SC and CM, the estimates are again in general biased downward, but less compared with the estimates from the dataset with selection. Analyses AI-12D on DYDs are again in close agreement with the values used for the simulation. This indicates that the deregression procedure also have problems in datasets without selection (no genetic trend) for traits with low heritabilities.

Table 4. Procedures and trait / population combinations used for parameter estimation.

Analyses	Method	Model
EM-S4	EM-REML (Sigurdsson et al., 1996)	Across populations for MI, SC and CM separately $\mathbf{A}^{-1}$ based on SIRE and MGS relationships
EM-SI	EM-REML (Sigurdsson et al., 1996)	Maximum estimate from all possible population combination subsets (the usual Interbull method). $\mathbf{A}^{-1}$ based on SIRE and MGS relationships
EM-K4	EM-REML (Klei & Weigel, 1998)	Across populations for MI, SC and CM separately $\mathbf{A}^{-1}$ based on SIRE and MGS relationships
AI-12	AI-REML (Madsen et al., 2000)	Across populations for MI, SC and CM simultaneously $\mathbf{A}^{-1}$ based on SIRE and DAM relationships traced back to base populations
AI-12D	AI-REML (Madsen et al., 2000)	As D, but on DYD

Conclusions

1. No major difference among compared estimation methods judging from estimated genetic correlations.
2. Estimated genetic correlations are downward biased in data with strong selection. Especially for low heritability traits, and especially when using deregressed proofs in contrast to DYDs.
3. An explanation could be that the deregression procedure did not entirely account for the genetic trend, as it only uses SIRE-MGS relationship in the correction for pedigree informations.
4. Further work is needed to pinpoint the problems with the deregression procedure.
   a) Make replicates of this simulation study.
   b) Make more simulations with varying selection intensity.
   c) Collect proofs and DYDs from a number of countries and estimate correlations from both information sources.
5. MT-Mace is feasible provided appropriate dependent variable is available.

References

Jairath, L., Dekkers, J.M.C., Schaeffer, L.R., Liu, Z., Burnside, E.B. & Kolstad, B. 1998. Genetic evaluation for herd life in Canada. *J. Dairy Sci. 81,* 550-562.

Klei, L. & Weigel, K.A. 1998. A method to estimate estimate correlations among traits in different countries using data on all bulls. *Interbull Bulletin 17,* 8-14.

Madsen, P., Jensen, J. & Mark, T. 2000. Reduced Rank Estimation of (Co)-variance Components For International Evaluation Using AI-REML *Interbull Bulletin 25,* 46-50.

Mark, T., Fikse, W.F., Sigurdsson, A. & Philipsson, J. 2000. Feasibility of International Genetic Evaluations of Dairy Sires for Somatic Cell Count and Clinical Mastitis. *Interbull Bulletin 25,* 154-162.

Nielsen, L.P., Sørensen, M.K., Berg, P. & Noesgaard, J. 2001. Alternative avlsstrategier for rød dansk malkerace. *DJF-rapport nr. 32 - Husdyrbrug,* 99 pp.

Schaeffer, L.R. & Zhang, W. 1993. Multi-trait, across country evaluation of dairy sires. *Interbull Bulletin 8,* 21 pp.

Sigurdsson, A., Banos, G. & Philipsson, J. 1996. Estimation of genetic (co)variance components for international evaluation of dairy bulls. *Acta Agric. Scand. Sect. A, Anim. Sci. 46,* 129-136.

Sørensen, M.K., Berg, P., Jensen, J. & Christensen, L.G. 1999. Stochastic simulation of breeding schemes for total merit in dairy cattle. *Interbull Bulletin 23,* 183-191.

Table 5. Estimated genetic correlations on deregressed proofs from data with selection. Asymptotic standard error of estimates in parentheses[1)].

POP	TRAIT	Population											
		1			2			3			4		
		MI	SC	CM	MI	SC	CM	MI	SC	CM	MI	SC	CM
1	MI				.79 .81 .80 .79(.03) .92(.01)			.79 .91 .77 .64(.07) .94(.02)			.88 .92 .88 .87(.03) .95(.01)		
			.00(.05) .01(.05)	-.08(.05) -.20(.06)		-.10(.07) .05(.04)	-.17(.07) -.14(.05)		-.15(.09) .03(.05)	.06(.09) -.13(.06)		-.04(.06) .09(.04)	-.08(.07) -.12(.05)
	SC					.19 .20 .21			.35 .36 .57			.83 .84 .85	
				.36(.05) .71(.03)	-.08(.06) -.04(.05)	.36(.09) .85(.03)	.16(.10) .56(.05)	.02(.10) -.07(.06)	.38(.12) .78(.05)	-.15(.13) .39(.09)	-.04(.06) -.07(.05)	.83(.04) .89(.03)	.67(.06) .69(.05)
	CM						.08 .14 .15			.03 .06 .24			.60 .61 -.71
					-.22(.06) -.19(.06)	.20(.10) .50(.06)	.45(.09) .72(.05)	-.37(.10) -.26(.07)	.00(.14) .41(.09)	.19(.14) .54(.11)	-.15(.07) -.20(.06)	.01(.08) .53(.06)	.71(.06) .86(.04)
2	MI							.81 .81 .83			.78 .78 .80		
						-.02(.03) -.01(.03)	-.15(.03) -.19(.04)	.74(.04) .91(.02)	-.01(.07) .00(.04)	.18(.07) -.17(.05)	.81(.03) .93(.01)	-.7(.06) .02(.04)	-.16(.07) -.19(.05)
	SC								.17 .17 .23			.23 .31 .26	
							.46(.03) .73(.02)	-.03(.07) -.06(.04)	.59(.08) .86(.03)	.17(.10) .54(.06)	-.10(.06) -.03(.04)	.40(.08) .84(..03)	.36(.10) .57(.05)
	CM									.16 .19 .24			.00 .00 -.02
								-.17(.08) -.25(.05)	.45(.09) .50(.05)	.44(.10) .70(.07)	-.14(.06) -.14(.05)	-.06(.09) .49(.06)	.30(.11) .72(.06)
3	MI										.74 .75 .75		
									-.06(.03) -.07(.03)	-.20(.03) -.28(.03)	.79(.04) .90(.02)	.13(.07) .07(.05)	-.10(.09) -.11(.06)
	SC											.58 .58 .74	
										.49(.02) .69(.02)	-.11(.06) -.11(.04)	.62(.08) .89(.03)	.27(.12) .58(.06)
	CM												-.06 .06 -.13
											-.18(.08) -.23(.06)	-.05(.10) .45(.06)	.06(.15) .64(.08)
4	MI												
												-.4(.03) -.06(.03)	-.14(.03) -.21(.03)
	SC												
													.53(.03) .73(.02)
	CM												

[1)] Analyses: Row 1 = EM-S4, Row 2 = EM-SI, Row 3 =EM-K4, Row 4 = AI-12, and Row 5 = AI-12D.

Table 6. Estimated genetic correlatons on deregressed proofs from data without selection. Asymptotic standard error of estimates in parentheses[1)].

POP	TRAIT	Population											
		1			2			3			4		
		MI	SC	CM	MI	SC	CM	MI	SC	CM	MI	SC	CM
1	MI				.93 .93 .94 -.20(.05) -.14(.04)			.92 .93 .93 .91(.03) .96(.02)			.87 .88 .90 .87(.03) .91(.02)		
					.93(.01) .94(.01)	-.11(.06) -.05(.05)	-.27(.06) -.24(.06)		-.06(.07) -.19(.05)	-.21(.07) -.36(.06)		-.14(.07) -.06(.05)	-.30(.06) -.23(.06)
			-.36(.05) -.34:05)										
	SC					.50 .51 .55			.20 .24 .44			.28 .34 .60	
				.82(.03) .76(.03)	-.18(.05) -.14(.05)	.62(.06) .84(.03)	.49(.08) .50(.06)	-.26(.07) -.04(.05)	.84(.05) .87(.04)	.72(.08) .45(.08)	-.33(.06) -.13.06)	.65(.07) .89(.03)	.83(.05) .53(.06)
	CM						.35 .42 .41			.16 .16 .37			.63 .67 .77
					-.38(.05) -.32(.06)	.42(.08) .52(.06)	.50(.09) .71(.05)	-.34(.07) -.23(.06)	.72(.07) .67(.06)	.69(.08) .78(.07)	-.43(.06) -.30(.07)	.54(.07) .64(.06)	.88(.04) .79(.06)
2	MI							.93 .93 .93			.85 .87 .86		
						-.10(.04) -.10(.04)	-.32(.04) -.32(.04)	.90(.02) .93(.01)	-.09(.07) -.17(.05)	-.13(.07) -.32(.05)	.85(.03) .90(.02)	.07(.06) -.06(.05)	-.20(.06) -.24(.06)
	SC								.65 .70 .71			.41 .41 .48	
							.71(.03) .68(.03)	-.18(.07) -.07(.05)	.50(.10) .80(.04)	.26(.12) .36(.07)	-.19(.07) -.08(.05)	.42(.09) .81(.03)	.48(.09) .41(.06)
	CM									.32 .50 .34			.36 .36 .38
								-.34(.07) -.30(.05)	.36(.11) .50(.06)	.36(.12) .66(.07)	-.29(.07) -.26(.06)	.14(.11) .45(.06)	.45(.09) .65(.07)
3	MI										.87 .87 .88		
									-.12(.03) -.15(.03)	-.30(.03) -.34(.03)	.89(.03) .90(.02)	.03(.07) -.02(.05)	-.25(.06) -.22(.06)
	SC											.54 .56 .62	
										.71(.02) .68(.02)	-.09(.08) -.24(.05)	.55(.09) .91(.03)	.70(.07) .61(.06)
	CM												.73 .73 .80
											-.22(.07) -.38(.06)	.34(.09) .53(.05)	.82(.05) .83(.05)
4	MI												
												-.11(.04) -.10(.03)	-.31(.04) -.31(.04)
	SC												
													.68(.03) .69(.02)
	CM												

[1)] Analyses: Row 1 = EM-S4, Row 2 = EM-SI, Row 3 =EM-K4, Row 4 = AI-12, and Row 5 = AI-12D.