

Estimation of Across Country Genetic Parameters for MACE Based on DYD's or Deregressed Proofs

Per Madsen¹, and Thomas Mark²

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

A method to estimate (co)variances for MACE based on AI-REML was introduced by Madsen *et al.* (2000). This method yielded similar estimates as obtained by the method of Klei & Weigel (1998) on mastitis and milk somatic cell field data. A simulation study to validate the AI-REML and two EM-REML methods was conducted by Madsen *et al.* (2001). The simulation study included 4 populations with different degree of exchange of germ plasm and 3 traits with different heritability and selection intensity. Deregressed proofs (**DER**) and daughter yield deviations (**DYD**) calculated after 20 years of selection was used to validate the estimation procedure. The main conclusions from the previous study were:

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 DER in contrast to DYD.
3. An explanation could be that the deregression procedure did not entirely account for the genetic trend, as it only uses SIRE-MGSrelationship in the correction for pedigree informations.

4. Further work is needed to pinpoint the problems with the deregression procedure.
5. MT-Mace is feasible provided appropriate dependent variable is available.

The aim of current study was to investigate if the observed bias in the estimated parameters also could be found on real data.

Material and Methods

Data

Data in form of national genetic evaluation results and DYD for milk yield for Holstein bulls were obtained from the Untied Kingdom (**GBR**), Hungary (**HUN**), Italy (**ITA**), the Netherlands (**NLD**), and the United States (**USA**). The total number of bulls with data in at least one country was 47850. The pedigree file included 74941 animals and a total 252 phantom parent groups (**PHG**) were defined based on selection path, birth year of bull and country of first registration. The large number of PHG lead to many groups with a limited genetic contribution to animals with data. The contributions of the PHG to bulls with data expressed as gene proportions are summarised in Table1.

Table 1. Summary of PHG contribution to bulls with data expressed as percent of genes.

	Contribution in percent of genes						
	>3	2-3	1-2	.5-1	.1-.5	.05-.1	<0.05
Number of PHG	2	4	29	24	62	71	60

The amount of genetic ties and number of bulls with data in each of the countries are shown in

Table 2. Numerous genetic ties connected the five countries.

Table 2. Number of common bulls (below diagonal), common $\frac{3}{4}$ sib families (above diagonal) and total number of bulls with data (on diagonal) included in parameter estimation and prediction of breeding values.

	GRB	HUN	ITA	NLD	USA
GBR	6758	249	581	623	776
HUN	207	1979	290	297	419
ITA	410	236	4604	708	1314
NLD	529	213	519	7689	1120
USA	605	363	895	817	29993

National evaluation results were deregressed within trait and country (Jairath *et al.*, 2001). The deregression procedure was tested by performing three MACE evaluations, a single country “MACE” (I_1) using data from USA only, a joint MACE on all available data, but with genetic correlations set to zero (I_2), and MACE using estimated genetic correlations (I_3). I_1 should show that the deregression followed by a MACE is a reversible process when the same pedigree informations are used in the two steps. I_2 should show the impact of different genetic grouping in within country analysis compared with MACE and I_3 should show the influence of correlated informations from other countries. These three runs were conducted on both DER and DYD.

Estimation model and methods

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

$$\mathbf{y} = \mathbf{C}\mathbf{c} + \mathbf{Z}\mathbf{Q}\mathbf{g} + \mathbf{Z}\mathbf{s} + \mathbf{e}$$

where \mathbf{y} is a vector of DER or DYD, \mathbf{c} is a vector of fixed country effects, \mathbf{g} is a vector of phantom group effects, \mathbf{s} is a vector of random bull effects, and \mathbf{e} is a vector of random residuals. \mathbf{C} , \mathbf{Z} , and \mathbf{Q} are design matrices relating DER or DYD to countries and bulls, and relating bulls to phantom groups respectively.

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

$$\mathbf{y} = \mathbf{Z}\mathbf{Q}\mathbf{f} + \mathbf{Z}\mathbf{s} + \mathbf{e}$$

with the following distributional properties:

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

where \mathbf{f} is a vector of phantom group + country effects, $\mathbf{G} = \mathbf{G}_0 \otimes \mathbf{A}$ is the (co)-variance matrix among elements in \mathbf{s} , \mathbf{A} is the additive relationship matrix, and \mathbf{R} is the residual (co)-variance matrix.

(Co)variance components were estimated using AI-REML (Madsen *et al.*, 2002) and were based either DER or DYD.

Results and Discussion

Results from the test of the deregression procedure showed that deregression and MACE are reversible processes when using the same pedigree information (and PHG). The main reason for differences between nat'l and int'l proofs are due to correlated information from other countries. I_1 and I_2 are almost identical with the nat'l proofs (X), the largest absolute difference are .24 and 96.6, respectively (Table 3). For I_3 the largest absolute difference is 588.39 or .78 SD units. The corresponding analysis based on DYD yielded differences between nat'l proofs and the MACE proofs of up to 776.5, 776.7 and 798.5 for I_1 , I_2 and I_3 , respectively, but the correlation between nat'l proofs and the MACE proof were all $>.99$.

Table 3. Comparison of national genetic evaluation results (**X**) with results from MACE applied to only milk production in the USA (**I₁**), with MACE results using r_G 's set to zero (**I₂**) and with MACE results using estimated r_G 's (**I₃**). 17,269 bulls with milk production variable in the USA were considered. MACE runs use deregressed national evaluation results as dependent variable.

	Difference (I-X)					
	Mean	SD	Mean	Max absolute value	SD	Correlation (X,I)
I ₁	120.3	755.2	0.003	0.24	0.031	1.0000
I ₂	120.3	755.3	-0.002	96.90	3.552	1.0000
I ₃	116.0	755.7	4.277	588.39	50.303	0.9978

The estimated genetic correlations based on DER and DYD for all bulls tested are in Table 4. Corresponding estimated obtained on a “well-connected” subset were very similar (not shown). The estimates based on DER were also in close agreement with results obtained on a “well-

connected” subset using the EM-REML methods of Klei & Weigel (1998) and of Sigurdsson *et al.* (1996) (results not shown). The small standard error of the estimated genetic correlations shows that there were many genetic links between the 5 Holstein populations.

Table 4. Estimated genetic correlations for milk yield for Holstein in five countries from model with phantom parent groups. Results based on de-regressed proofs (above diagonal) and on DYD (below diagonal). Asymptotic standard error of estimates in parentheses.

	GBR	HUN	ITA	NLD	USA
GBR		.871 (.026)	.921 (.009)	.941 (.008)	.896 (.009)
HUN	.784 (.028)		.930 (.020)	.839 (.028)	.883 (.020)
ITA	.902 (.010)	.844 (.023)		.938 (.008)	.967 (.004)
NLD	.897 (.011)	.683 (.035)	.873 (.012)		.924(.008)
USA	.856 (.010)	.803 (.022)	.946 (.004)	.821 (.012)	

The correlations based on DER were in all cases larger than the corresponding estimate based on DYD (Table 4). This is in contrast to the results from the simulation study of Madsen *et al.* (2001) where estimates based on DYD were numerically larger and in all cases in closer agreement with the parameters used in the simulation than the estimates based on DER.

One reason for this discrepancy could be that in the simulation study all animals could be traced back to the four base populations, so only four PHG were defined. In the present study, only few of the

estimated PHG effects were significantly different from zero. To investigate the effect of the large number of PHG, genetic correlations were also estimated ignoring PHG.

In general, estimated r_G (Table 5) were also larger when based on DER compared to estimates based on DYD. By comparing the results in Table 4 and 5 it can be seen, that PHG have a substantial effect on some of the estimated correlations. With PHG in the model, $r_G(\text{GBR},\text{ITA})$ based on DER were .921 while in the model without PHG it was .784.

Table 5. Estimated genetic correlation's for milk yield for Holstein in five countries from model without phantom parent groups. Results based on de-regressed proofs (above diagonal) and on DYD (below diagonal). Asymptotic standard error of estimates in parentheses.

	GBR	HUN	ITA	NLD	USA
GBR		.837 (.021)	.784 (.014)	.899 (.009)	.822 (.011)
HUN	.771 (.024)		.919 (.016)	.886 (.020)	.889 (.017)
ITA	.801 (.014)	.877 (.018)		.902 (.009)	.966 (.003)
NLD	.815 (.015)	.764 (.027)	.884 (.011)		.928 (.007)
USA	.794 (.012)	.824 (.018)	.948 (.004)	.847 (.010)	

The analyses on real data did not support the findings of Madsen *et al.* (2001). A closer look on the simulation study revealed possible reasons for the apparent superiority of DYD over DER for estimation of genetic correlations:

1. In the simulation program of Sørensen *et al.* (1999), progeny groups of bulls were not simulated. Instead an observation on the bull itself was simulated as the breeding value of the bull plus twice a progeny group average. This means that the residual variance of an observation on a bull was proportional to $n/4$, where n is the number of progenies. This was also the weight used in prediction of breeding values in the simulation program and assumed for DYD, but in the deregression and estimation of genetic correlation the observations was weighted by n .
2. The deregression procedure was based on a SIRE-MGS relationship, while the estimation was based one a SIRE-DAM relationship.

Results based on DER are therefore wrong in Madsen *et al.* (2001), but results based on DYD should be correct.

Conclusions

1. The deregression and MACE procedures applied by Interbull are reversible processes when using the same pedigree information in the two steps
2. Definition of PHG can have a large impact on the estimated genetic correlations
3. The downward bias in estimated genetic correlations when based on DER found by Madsen *et al.* (2001) could not be verified on

real data. This is manly due to a mismatch between the simulation model, deregression and estimation model.

4. DER is recommended as dependent variables in MACE

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 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.
- Madsen, P., Sørensen, M.K. & Mark, T. 2001. Validation and Comparison of Methods to Estimate (Co)Variance Componants for Mace. *Interbull Bulletin* 27, 73-79.
- Schaeffer, L.R. & Zhang, W. 1993. Multi-trait, across country evaluation of dairy sires. *Interbull Bulletin* 8, 21pp.
- 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.
- Sørensen, M.K., Berg, P., Jensen, J. & Christensen, L.G. 1999a. Stochastic simulation of breeding schemes for total merit in dairy cattle. GIFT-workshop, Wageningen, The Netherlands. *Interbull Bulletin* 23, 183-191.