

RESTRICTED MAXIMUM LIKELIHOOD ESTIMATION OF VARIANCES AND COVARIANCES FROM SELECTED DATA

H. SIMIANER, WEST GERMANY

Institute of Animal Breeding and Genetics, Bismarckstr. 16,
D-6300 Giessen (Federal Republic of Germany)

SUMMARY

For application in a multi-trait BLUP procedure, variances and covariances of first, second and third lactation records were estimated. This was done by a restricted maximum likelihood (REML) procedure with Cholesky-transformed data, including all observations. Initial values were taken from an analysis with Henderson's method 3, only using data of cows, that have been recorded for these three lactations.

REML-estimates of genetic and residual variances and covariances were 10 to 25% higher than the initial values. Estimated heritabilities decreased about 10%, while genetic and phenotypic correlations were changed unsystematically within a narrow range. The results show, that it is worthwhile including all data available in parameter estimation to remove bias due to selection.

INTRODUCTION

Estimation of variances and covariances from selected data yields certain implications. On the one hand it is no trivial task to run a good estimation procedure, especially for covariance components. On the other hand reliable estimates of population parameters are needed, when mixed model techniques are applied to selected data. Along with certain technical problems, there are also some theoretical inconsistencies, for instance the antagonism of unbiasedness and estimation within parameter space. These implications have been recently discussed by Henderson (1984a,b). A special type of selection exists in dairy records of subsequent lactations. Between first and second lactation, there is certainly selection based on both dairy performance of first lactation and other criteria. So, if estimation of variances and covariances only uses data from cows with first and second lactation records, estimates are expected to be biased downwards (Meyer and Thompson, 1984). The effect on estimates of parameters such as heritabilities and correlations cannot be predicted that easy. It is one objective of this study to get a rough idea of the size of these biases, when real-life data are analyzed with different estimation procedures.

MATERIAL AND METHODS

The data available for this study consisted of 153,885 records of 80,679 German Friesian cows (33,280 with only first, 21,592 with first and second and 25,807 with first, second and third lactation). Milk-, fat- and protein yield were recorded for each lactation.

Using only the data of those cows with three lactations recorded, a method 3 estimation of variances and covariances (Henderson, 1955) was carried out. Least-squares-effects were used to correct the data for those fixed effects, that were not applicable to all observations.

Furthermore, method 3 estimates of variances and covariances were used as initial values for a restricted maximum likelihood (REML) procedure, including also data from cows, which have been culled after their first or second lactation. As suggested by Schaeffer (1985), REML was applied after Cholesky-transformation of the observation vector.

The multi-trait model, allowing for missing observations, is (Arnason, 1985):

$$y = Xb + Wu + Ze \quad (1)$$

with

y = observation vector
 b = vector of fixed effects
 u = vector of random effects
 e = vector of random error effects

X, W, Z = design matrices corresponding to b, u and e

Variances and covariances are:

$$\text{var} \begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} W(V^*A)W' + Z(I^*E)Z' & W(V^*A) & Z(I^*E) \\ & V^*A & 0 \\ & & Z(I^*E)Z' \end{bmatrix}$$

with

V = numerator relationship matrix
 A = matrix of additive-genetic variances and covariances of the multiple traits
 E = matrix of residual variances and covariances of the multiple traits
 * = Kronecker-product (Searle, 1966)

As E has to be positive definite, it can be subjected to Cholesky-decomposition:

$$E = TT'$$

where T is a lower triangular matrix.

Cholesky-transformation of the observation vector y leads to y_t :

$$y_t = Z(I^*T^{-1})y$$

and so variances and covariances become:

$$A_t = T^{-1}AT^{-1}$$

$$E_t = T^{-1}ET^{-1} = T^{-1}TT'T^{-1} = I$$

For requirements of this transformation see van der Werf (1983) and Arnason (1985).

The i-th round of REML-iteration starts with A_i and E_i , and so Cholesky-decomposition leads to:

$$y_{ti} = Z(I * T_i^{-1})y$$

$$E_{ti} = I$$

$$A_{ti} = T_i^{-1} A_i T_i^{-1}$$

Then mixed model equations are:

$$\begin{bmatrix} X'X & X'W \\ W'X & W'W + (V * A_{ti})^{-1} \end{bmatrix} \begin{bmatrix} \hat{b}_{ti} \\ \hat{u}_{ti} \end{bmatrix} = \begin{bmatrix} X'y_{ti} \\ W'y_{ti} \end{bmatrix} \quad (2)$$

After solving these equations and inverting their left hand side, genetic variances and covariances are estimated:

$$\hat{A}_{ti} = (U'_{ti} U_{ti} + \sum_{m=1}^q C_{mmi}) / q$$

with

$$U_{ti} = \begin{bmatrix} \hat{u}'_{ti1} \\ \hat{u}'_{ti2} \\ \vdots \\ \hat{u}'_{tiq} \end{bmatrix} \quad \text{Matrix of the transposed vectors of estimated breeding values}$$

C_{mmi} = diagonal block of the m-th animal in the inverted left hand side of (2)

q = number of individuals in u

Residual covariances between traits are assumed to be zero (Schaeffer, Wilton and Thompson, 1978). This is only acceptable with Cholesky-transformed data. Therefore only diagonal elements of \hat{E}_{ti} have to be estimated:

$$\hat{e}_{tij} = (y'_{tij} y_{tij} - \hat{b}'_{tij} X'_j y_{tij} - \sum_{m=1}^q \hat{u}'_{tij} W'_j y_{tij}) / (N_j - r(X_j))$$

where index j refers to matrix- and vector partitions defined for the j -th trait and

$$\begin{aligned} \hat{e}_{tij} &= j\text{-th diagonal element of } \hat{E}_{ti} \\ N_j &= \text{number of observations in trait } j \\ r(X_j) &= \text{rank of } X \text{ defined for trait } j \end{aligned}$$

Finally

$$A_{i+1} = T_i \hat{A}_{ti} T_i'$$

$$E_{i+1} = T_i \hat{E}_{ti} T_i'$$

As in every round of iteration mixed model equations (2) only differ in A_{ti} and y_{ti} , most of the computations in setting up (2) have only once to be done. This is, in fact, a considerable reduction of computational labour.

RESULTS AND DISCUSSION

The model applicated to the described data was

$$Y_{ijklmnopq} = \mu + m_i + c_j + a_k + p_l + f_m + h_n + g_o + s_{op} + e_{ijklmnopq}$$

with

- y = phenotypical observation
- μ = overall mean
- m_i = month of calving
- c_j = year of calving
- a_k = age at first calving
- p_l = length of the preceding calving interval
- f_m = length of the following calving interval
- h_n = herd class
- g_o = genetic group of sire
- s_{op} = sire within genetic group; random
- e = residual error; random

The complete model was run for the method 3 analysis of variance with the program of Harvey (1977). For effects printed in boldface, data were corrected before running the REML-procedure with initial values of variances and covariances taken from this analysis. Iteration was stopped, when the change of residual variances from one round to another fell below 0.1%. Convergence happened to occur quite rapidly. By far the biggest change of estimates was to observe in the first round of iteration.

Results of both method 3- and REML estimation of variances and covariances and parameters for the trait milk yield are given in table 1. In each pair of values the bigger one is marked by bold printing.

As expected variances and covariances increased when using all observations for estimation. Typically the increase in residual variances and covariances was fairly larger (up to 25%) as the one in genetic variances and covariances (up to 11%). Consequently heritabilities calculated from method 3 variances were overestimated about 10%. For milk yield, estimated genetic and phenotypical correlations happened to increase - with one exception - when all observations were included. Considering results for fat- and protein yield in addition, increase and decrease of estimated correlations seemed to appear quite unsystematically, the change always being within a narrow range of $\pm 2\%$ maximum.

REML-estimates of genetic and phenotypical parameters are lying within the range that Maijala and Hanna (1974) have found to be probably the realistic one. A comparison of these results to those of theoretical investigations and simulation studies (Meyer and Thompson, 1984; Rothschild, Henderson and Quaas, 1979) is difficult to make, as there is no knowledge about selection criteria, selection intensities etc. in the real-life data case.

In tendency, bias due to selection of observations within animals seems to be removed or at least reduced by using a REML-procedure.

Table 1: Estimates of variances, covariances and parameters for milk yield in first, second and third lactation from Henderson's method 3 and REML analysis

Estimates	Method 3			REML					
	Lact.	1	2	3	1	2	3		
genetic variances and covariances (kg ²)	1	35842	30492	30391	39966	33694	33423		
	2		40344	38359				43348	41315
	3			44019					
residual variances and covariances (kg ²)	1	429775	238049	228654	516681	286185	274891		
	2		619209	334016				767223	411260
	3			704625					
heritabilities		.308	.245	.235	.287	.214	.220		
additive-genetic correlations	1		.802	.765		.809	.772		
	2			.911				.917	
phenotypical correlations	1		.485	.439		.476	.448		
	2			.530				.545	

REFERENCES

- ARNASON, T. 1985. Practical aspects in the estimation of breeding values for multi-trait selection. 36th Annual Meeting of the EAAP, Kallithea, Greece, September 30 - October 3, 1985
- HARVEY, W.R. 1977. User's guide for LSML76. Ohio State University, polycopy
- HENDERSON, C.R. 1955. Estimation of variance and covariance components. *Biometrics* 9: 226
- HENDERSON, C.R. 1984a. Estimation of variances and covariances under multiple trait models. *J. Dairy Sci.* 67: 1581
- HENDERSON, C.R. 1984b. Application of linear models in animal breeding. Guelph, Ontario, University of Guelph
- MAIJALA, K. and HANNA, M. 1974. Reliable phenotypic and genetic parameters in dairy cattle. Proc. First World Congress Genet. Appl. Livest. Prod., Madrid, Vol. I: 541
- MEYER, KARIN and THOMPSON, R. 1984. Bias in variance and covariance component estimators due to selection on a correlated trait. *Z. Tierzüchtg. Züchtgsbiol.* 101: 33

- ROTHSCHILD, M.F., HENDERSON, C.R. and QUAAS, R.L. 1979. Effects of selection on variances and covariances of simulated first and second lactations. *J. Dairy Sci.* **62**: 996
- SCHAEFFER, L.R. 1985. Notes on the course 'Advances in estimating breeding values and population parameters'. Berlin, March 18 - 29, 1985
- SCHAEFFER, L.R., WILTON, J.W. and THOMPSON, R. 1978. Simultaneous estimation of variance and covariance components from multitrait mixed model equations. *Biometrics* **34**: 199
- SEARLE, S.R. 1966. Matrix algebra for the biological sciences. New York, London, Sydney: Wiley and sons
- VAN DER WERF, J. 1983. Single and multiple trait models for sire evaluation in a population under selection. Wageningen, Agricultural University, Department of Animal Science, thesis