

Estimates of Relationship Between Milk Yield and Age at First Calving in Cattle Using Multivariate REML

V.P.S. Chauhan and Prem Narain

Project Directorate on Cattle (ICAR), Meerut, U.P. 250 005.

Summary

Relationship between milk yield and age at first calving in cattle were calculated from (co) variances estimated by multivariate restricted maximum likelihood with equal design matrix fitting a unbalanced mixed model with herd, year and calendar month of calving as fixed effects, and sire and residual errors as independent random effects. Univariate analyses were also carried out to compare the variances estimated by multivariate analysis. The data consisted of first lactation records on 733 purebred progeny of 69 Haryana sires in 3 herds from 1969 to 1981. The genetic correlation between yield and age was estimated to be 0.82 ± 0.07 , but the phenotypic correlation was only 0.37 ± 0.05 . Large differences were observed in the estimates of variances for milk yield from univariate and multivariate analysis, but for age at first calving the variances were quite similar.

Key Words : Canonical transformation, genetic and phenotypic correlation, heritability, multivariate, univariate, restricted maximum likelihood (REML).

Introduction

In dairy cattle, milk yield is influenced by age at calving. Cows with higher the age at first calving tend to yield more milk, however, higher age at calving is not economical in dairy industry. This association is a biological phenomenon because the higher age at first calving gives a cow an opportunity for increase in body capacity and udder development.

This study was undertaken to estimate the (co)variance between milk yield in first lactation and age at first calving. Analysis were also conducted to compare the estimates of variances from univariate and multivariate restricted maximum likelihood (REML) techniques. The studies on relationship between the above two traits are mostly from methods 2 and 3 of Henderson [4]. The heritabilities and genetic and phenotypic correlations were also estimated.

• Principal Scientist, IARI, Pusa, New Delhi - 110 012.

2. Materials and Methods

The data consisted of first lactation records of 733 purebred progeny of 69 Haryana bulls from 1969 to 1981, collected from Madhurikund, Mathura and Babugarh herds in Uttar Pradesh.

The univariate and multivariate analysis with equal design matrices for both traits were carried out using an unbalanced mixed model with herd, year and calendar month of calving as fixed effects, sires and residual error as independent random effects. The sire is a genetic effect and accounts for one-fourth of the additive genetic (co)variance.

Between and within sire components of (co)variances were estimated using the REML procedure of Patterson and Thompson [8], together with expectation-maximization (EM) algorithm (Dempster *et al.* [2]). This algorithm ensures the estimates of (co)variances to be positive definite. Because of equal design matrices for both traits it was possible to transform the observations, sire, and residual (co)variances to a canonical scale (Thompson [12]; Hayes and Hill [3]). After absorbing the fixed effects the coefficient matrix for sires was tridiagonalised to speed up the convergence. Procedures in detail for tridiagonalisation and estimating (co)variance components with REML are outlined by Thompson [13], Lin [5], Meyer [6], Smith and Graser [11]. Approximate standard errors of the estimates were the lower bounds as they were estimated using the inverse of the sires coefficient matrix (Meyer [6]). Analysis was carried out using a computer program of Meyer [7].

The following model with equal design matrices for all traits was used.

$$y = (I_q * X) b + (I_q * Z) u + e$$

where :

y is vector of observations for yield and age;

I_q is identity matrix of order equal to number of traits (q);

X is design matrix for the fixed effects;

b is vector of unknown fixed effects of herd, year, and month of calving;

Z is design matrix for sires;

u is vector of unknown random effects for sires;

e is vector of random residual effects;

* denotes the direct product operation (Searle [9]).

$$\text{Where, } E \begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} (I^* X)b \\ 0 \\ 0 \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} V & G^*Z & R^*I_N \\ Z'^*G & G^*I_S & 0 \\ R^*I_N & 0 & R^*I_N \end{bmatrix}$$

where G , R and V are the sire, residual, and phenotypic (co)variance matrices, respectively, and $V = G^*ZZ' + R^*I_N$. The I_S and I_N are the identity matrices of order equal to number of sires and total number of observations, respectively.

Canonical transformation enables the multivariate analysis with equal design matrices and with one random factor in addition to random residuals to be carried out as univariate analysis. If the G is positive Semi-definite and R is positive definite, there exists a non-singular matrix, Q , such that

$$QQ' = G^c \quad \text{and} \quad QRQ' = R^c$$

where G^c is a diagonal matrix and R^c is an identity matrix. The elements of the diagonal matrix, G^c , are the eigenvalues of $R^{-1}G$ and corresponding rows of the Q matrix are the normalized eigenvectors (Anderson [1]). The diagonal elements of G^c are the new genetic variances (Hayes and Hill [3]), inverse of which was added to the diagonal elements of the sires' part of the transformed mixed model equations. On a transformed scale the data vector and the parameters to be estimated for the above model are :

$$y^c = Qy, \quad b^c = Qb, \quad u^c = Qu, \quad e^c = Qe,$$

$$\text{Var}(u^c) = \text{Var}(Qu) = QRQ',$$

$$\text{Var}(e^c) = \text{Var}(Qe) = QRQ', \quad \text{and}$$

$$\text{Cov}(u^c, e^c) = 0.$$

Estimates of the fixed and random effects were back-transformed to the original scale by premultiplying the estimates on the canonical scale with Q^{-1} . Similarly the estimates of sire and residual (co)variances were back-transformed as

$$G = Q^{-1}G^cQ^{-1} \quad \text{and} \quad R = Q^{-1}R^cQ^{-1}$$

The matrices G , R , Y , u , b and e with a superscript c indicate the values on a canonical scale. A procedure for calculating the Q matrix

utilizing symmetric matrices is given by Schaeffer [10]).

Starting values of sire and residual (co)variances were taken from preliminary analysis of these data. Criterion of convergence was chosen as .000001 percent average minimum change in the estimates of between and within sire (co)variances.

3. Results and Discussion

The estimates of between and within sire (co)variances from univariate and multivariate analysis at convergence are in Table 1. For age at calving the between sire variance was larger than the within sire variance from univariate as well as from multivariate analysis. The variances for age at calving from both analysis were similar, suggesting that the milk yield did not effect the variability in age at calving, which is quite logical. But the between sire variances for milk yield were different from univariate and multivariate analysis, although the within sire variances were not much different. It showed that the age at calving was responsible for variability in milk yield. The estimate of covariance component between yield and age at calving was large (1700.29).

Table 1. Estimates of between and within sire (co)variance components for milk yield and age at first calving from multivariate and univariate analysis.

| Traits | Multivariate | | Univariate | |
|--------------|--------------|-----------|------------|-----------|
| | Between | Within | Between | Within |
| Yield, Yield | 19,754.85 | 86,130.87 | 7,206.07 | 86,805.54 |
| Age, Age | 218.45 | 19.92 | 218.56 | 19.91 |
| Yield, Age | 1,700.29 | 140.62 | - | - |

Table 2 contains the estimates of heritability and genetic and phenotypic correlations. Genetic correlation was high and positive (0.82 ± 0.07), but the phenotypic correlation was small (0.37 ± 0.05). From multivariate analysis the heritability of milk yield was high, and for age at first calving it was much higher than the parameter space (0 to 1). This might be because of sampling. However, with the univariate analysis the heritability of milk yield was only

0.31±0.12, but for the age at first calving it was the same as from ultivariate analysis since the variances were similar.

Table 2. Genetic and phenotypic correlations and heritabilities of milk yield and age at first calving in Haryana cattle from multivariate analysis¹.

| Trait | Milk yield | Age at calving |
|----------------|------------|----------------|
| Milk yield | 0.75±0.16 | 0.37±0.05 |
| Age at calving | 0.82±0.07 | 0.67±0.06 |

¹Heritabilities on the diagonal, genetic correlation on lower and phenotypic correlation on upper off diagonal. Heritabilities of yield and age from univariate analysis were 0.31±0.12 and 3.67±0.06, respectively.

The Q matrix for age and yield at convergence was as follows.

$$Q = \begin{bmatrix} 0.223797 & 0.000038 \\ -0.026691 & -0.003427 \end{bmatrix}$$

Corresponding eigenvalues on the canonical scale were 10.97004 and 0.007658 giving heritabilities of 3.67 and 0.003 for age and yield, respectively. This implies that a compound variable (C) with following linear combination of age at first calving (A) and milk yield (M) would explain most genetic variation for these two traits.

$$C = 0.223792*A + 0.000038*M$$

This linear combination has the higher heritability and is the one in which maximum genetic progress could be achieved through selection. However, for achieving maximum genetic progress it is important that relative economic weights for age and yield may be accounted for. Then the compound variable may be modified as

$$C = 0.223792*Aw_1 + 0.000038*Mw_2$$

where w_1 and w_2 are relative economic weights for age at first calving and milk yield, respectively. However, we are not aware of the reports on selection for such canonical compound variables and the responses thereof.

REFERENCES

- [1] Anderson, T.W., 1958. An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York.
- [2] Dempster, A.P., N.M. Laird, and D.B. Rubin., 1977. Maximum likelihood from incomplete data via the EM algorithm. *J.R. Stat. Soc., Ser. B.*, **38**, 1.
- [3] Hayes, J.F., and W.G. Hill., 1980. A reparameterization of a genetic selection index to locate its sampling properties. *Biometrics* **36**, 237.
- [4] Henderson, C.R., 1953. Estimation of variance and covariance components. *Biometrics* **9**, 226.
- [5] Lin, C.Y., 1987. Application of singular value decomposition to restricted maximum likelihood estimation of variance components. *J. Dairy Sci.* **70**, 2680.
- [6] Meyer, K., 1985. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, **41**, 153.
- [7] Meyer, K., 1986. Restricted maximum likelihood to estimate genetic parameters - in practice. Pages 454-459 in *Proc. 3rd World Congr. Genet. Appl. Livest. Prod.*, **12**. Lincoln, Nebraska.
- [8] Patterson, H.D. and R. Thompson., 1971. Recovery of interblock information when block sizes are unequal. *Biometrika*, **54**, 545.
- [9] Searle, S.R., 1966. Matrix Algebra for the Biological Sciences, John Wiley and Sons, New York.
- [10] Schaeffer, L.R., 1986. Estimation of variances and covariances within the allowable parameter space. *J. Dairy Sci.*, **69**, 187.
- [11] Smith, S.P. and H.U. Graser., 1986. Estimating variance components in a class of mixed models by restricted maximum likelihood. *J. Dairy Sci.*, **69**, 1156.
- [12] Thompson, R. 1977. Estimation of quantitative genetic parameters. Pages 639-658 in *Proc. Int. Conf. Quant. Genet.*, Iowa State Univ. Press, Ames.
- [13] Thompson, R., 1982. Methods of estimation of genetic parameters. Page 95-103 in *Proc. 2nd World Congr. Genet. Appl. Livest. Prod.*, **5**. Madrid, Spain.