

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Papers and Publications in Animal
Science

Animal Science Department

January 1980

Estimating a Genetic Covariance from Least Squares Solutions

G. R. Wiggans
Cornell University

R. L. Quaas
Cornell University

L. Dale Van Vleck
University of Nebraska-Lincoln, dvan-vleck1@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/animalscifacpub>



Part of the [Animal Sciences Commons](#)

Wiggans, G. R.; Quaas, R. L.; and Van Vleck, L. Dale, "Estimating a Genetic Covariance from Least Squares Solutions" (1980). *Faculty Papers and Publications in Animal Science*. 374.

<https://digitalcommons.unl.edu/animalscifacpub/374>

This Article is brought to you for free and open access by the Animal Science Department at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Papers and Publications in Animal Science by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Estimating a Genetic Covariance from Least Squares Solutions

G. R. WIGGANS¹, R. L. QUAAS,
and L. D. VAN VLECK
Department of Animal Science
Cornell University
Ithaca, NY 14853

ABSTRACT

An unbiased and invariant procedure is described for estimating the covariance between effects of two traits or of random variables such as sires in two environments when the residual effects for one trait are not correlated with those of the other trait. When the covariance is desired between two classes in the model, such as sire of fetus and sire of cow, the data can be divided randomly into two sets to allow use of this procedure to estimate the covariance since the residuals in the two data sets would be uncorrelated. The method is to obtain solutions by ordinary least squares for each trait separately as if the random effects of interest were actually fixed and then to compute a sample covariance between the two sets of solutions.

INTRODUCTION

A problem of recurring interest to animal breeders is the estimation of the genetic correlation between the expression of genotypes in two environments (2). One approach is to consider the measurements separate traits. Typical examples are when dairy sires have daughters in temperate and in tropical regions (6), when the daughters are fed all roughage diets or high concentrate rations (13), or when they are in different management systems (5, 9, 12). The same daughters do not have records in both environments so some standard methods of estimating covariance components do not apply.

A similar situation exists for a pair of traits in the usual sense when the observations are all

from different animals; i.e., no animal has both traits measured (1). The similarity of these situations lies in the fact that the "errors" associated with records are uncorrelated. Therefore, any covariance between observations on different traits is genetic in nature.

Schaeffer et al. (11) have discussed the use of restricted maximum likelihood (8) for such situations. They give an iterative computing algorithm based on "mixed model equations." The purpose of this note is to describe an invariant unbiased estimator based on a bilinear form that is computed readily from solutions to least squares equations.

METHOD

Consider a model

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & O \\ O & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & O \\ O & Z_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

where $[\beta'_1 \beta'_2]$ are fixed effects and $[u'_1 u'_2]$ are random vectors representing sire values with null means, variance,

$$\begin{bmatrix} I\sigma_{u_1}^2 & I\sigma_{u_1 u_2} \\ I\sigma_{u_1 u_2} & I\sigma_{u_2}^2 \end{bmatrix},$$

and elements uncorrelated to those of $[e'_1 e'_2]$. $[e'_1 e'_2]$ is also a random vector with null means and variance

$$\begin{bmatrix} I\sigma_{e_1}^2 & O \\ O & I\sigma_{e_2}^2 \end{bmatrix}.$$

Also, $\text{rank}(X_i Z_i) = \text{rank}(X_i) + \text{rank}(Z_i) - 1$.

Received May 7, 1979.

¹Animal Improvement Programs Lab, USDA, BARC East, Beltsville, MD 20705.

This rank condition implies that all sires are "connected" (10) and, thus, rules out models in which sires are nested within a fixed effect. The reason for this requirement will be noted later.

This model would be appropriate for an animal breeding situation where y_1 and y_2 represent records on different groups of sires' daughters in different environments or records on different traits, and u_1 and u_2 are the sires' values in the two environments. An estimate of $\sigma_{u_1 u_2}$ is desired as a measure of the similarity between sire effects in the two environments. The β 's represent fixed effects such as those associated with year and season of freshening and herd.

The proposed method is to solve the ordinary least squares equations regarding the u_i as fixed effects and then to compute the sample covariance between the sire solutions. That is, to solve

$$\begin{bmatrix} X_i' X_i & X_i' Z_i \\ Z_i' X_i & Z_i' Z_i \end{bmatrix} \begin{bmatrix} \tilde{\beta}_i \\ \tilde{u}_i \end{bmatrix} = \begin{bmatrix} X_i' y_i \\ Z_i' y_i \end{bmatrix}$$

for $i = 1, 2$ and then compute

$$\begin{aligned} \hat{\sigma}_{u_1 u_2} &= \tilde{u}_1' (I - 1/c J_c) \tilde{u}_2 / (c - 1) \\ &= [\sum \tilde{u}_{1i} \tilde{u}_{2i} - (1/c)(\sum \tilde{u}_{1i})(\sum \tilde{u}_{2i})] / (c - 1), \end{aligned}$$

where c is the number of sires and J_c is a c^2 matrix of ones. This estimator is unbiased and invariant to the solutions which are not unique, as will be illustrated in an example. The proof of these statements follows that of Henderson et al. (4) for the same properties of Henderson's Method 2 (3) estimator of a component of variance. Instead of repeating that proof, which is tedious, the similarity to the Method 2 estimator will be shown.

Any solutions to the least squares equations must satisfy

$$Z_i' X_i \tilde{\beta}_i + Z_i' Z_i \tilde{u}_i = Z_i' y_i.$$

Hence,

$$\tilde{u}_i = (Z_i' Z_i)^{-1} Z_i' (y_i - X_i \tilde{\beta}_i),$$

which is the vector of totals of progeny records

that have been "adjusted for the fixed effects." The vector of adjusted records, $y_i - X_i \tilde{\beta}_i$, is that which is used in Method 2. The quadratic form used (along with the residual sum of squares) in Method 2 is

$$\begin{aligned} & (y_i - X_i \tilde{\beta}_i)' Z_i' (Z_i' Z_i)^{-1} Z_i' \\ & (I - 1/N J_N) Z_i' (Z_i' Z_i)^{-1} Z_i' (y_i - X_i \tilde{\beta}_i) \\ &= \tilde{u}_i' Z_i' (I - 1/N J_N) Z_i \tilde{u}_i \\ &= \sum_j n_j \tilde{u}_{ij}^2 - (\sum_j n_j \tilde{u}_{ij})^2 / N, \end{aligned}$$

where n_j is the number of progeny of the j th sire, and N is the total number of observations. Because of the similarity in the way that records are adjusted for fixed effects, the proposed estimator of the covariance is invariant to the constraints needed to obtain least squares solutions for any model under which Method 2 is a valid estimation procedure for variance components. To insure invariance, the rank conditions on X_i and Z_i of the model are necessary.

The major distinction between the bilinear form of $\hat{\sigma}_{u_1 u_2}$ and the quadratic form of Method 2 is that the latter is a weighted (by number of progeny) sum of squares of sires' solutions as opposed to an unweighted sum of cross products. A covariance estimator based on a weighted sum of cross products can be constructed, but it is not obvious what the weights should be, although logically they should be some function of progeny numbers. An expression for the variance of a weighted estimator was derived but did not prove useful in determining optimal weights. There is, perhaps, some justification for the unweighted estimator in that it is analogous to the "D₃-best" estimator of a variance component (7) which uses an unweighted sum of squares of the \tilde{u} 's.

The procedure can be generalized for the case where some elements of u_1 (u_2) are not correlated to an element of u_2 (u_1). In the example, this would happen if all sires do not have progeny in both environments.

For this case, define u_1^* and u_2^* to be the $c^* \times 1$ vectors of solutions for the c^* sires with progeny in the first and second environments. Then an unbiased estimator of $\sigma_{u_1 u_2}$ that is

invariant to the solutions to the least squares equations is

$$\hat{\sigma}_{u_1 u_2} = [\Sigma u_{1i}^* u_{2i}^* - (1/c^*) \\ (\Sigma u_{1i}^*)(\Sigma u_{2i}^*)] / (c^* - 1).$$

The proof follows the same lines as that given for the case when all sires have progeny in both environments.

Application

The technique was applied to a sire evaluation problem where the different environments were defined as herds where differing proportions of net energy in the ration were derived from concentrate feeds (13). The estimated covariances were used in calculating correlations as a measure of how similar sire effects were in the different feeding environments.

Example

A simple example may illustrate the method more clearly. Suppose the model for measurements on the two traits is

$$y_{1ij} = \mu_1 + s_{1i} + w_{1ij}$$

$$y_{2ik} = \mu_2 + s_{2i} + w_{2ik}$$

where μ_1 and μ_2 are constants for traits 1 and 2, s_{1i} and s_{2i} are the random effects associated with sire i for traits 1 and 2, and w_{1ij} and w_{2ik} are the residual effects for the j th and k th progeny of sire i for traits 1 and 2. No progeny have records for both traits. Traits 1 and 2 may be the same trait in two different environments. The s 's and w 's have expected values of zero and variances $\sigma_{s_1}^2$, $\sigma_{s_2}^2$, $\sigma_{w_1}^2$, and $\sigma_{w_2}^2$. The covariance between s_{1i} and s_{2i} is $\sigma_{s_1 s_2}$ for all i . All other covariances are zero.

Suppose the ordinary least squares equations for the two traits are:

$$\begin{bmatrix} 20 & 5 & 3 & 12 \\ 5 & 5 & 0 & 0 \\ 3 & 0 & 3 & 0 \\ 12 & 0 & 0 & 12 \end{bmatrix} \begin{bmatrix} \tilde{\mu}_1 & \tilde{\mu}_2 \\ \tilde{s}_{11} & \tilde{s}_{21} \\ \tilde{s}_{12} & \tilde{s}_{22} \\ \tilde{s}_{13} & \tilde{s}_{23} \end{bmatrix} = \begin{bmatrix} 2000 & 2200 \\ 500 & 600 \\ 360 & 400 \\ 1140 & 1200 \end{bmatrix}$$

With the constraints $\tilde{\mu}_1 = 0$ and $\tilde{\mu}_2 = 0$, the solutions for the s 's are

$$\begin{bmatrix} 100 \\ 120 \\ 95 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 120 \\ 133 \\ 100 \end{bmatrix}$$

The solutions in terms of the models, where $M(\)$ denotes the model, are

$$M(\tilde{s}_{11}) = \mu_1 + s_{11} + \bar{w}_{11}.$$

$$M(\tilde{s}_{12}) = \mu_1 + s_{12} + \bar{w}_{12}.$$

$$M(\tilde{s}_{13}) = \mu_1 + s_{13} + \bar{w}_{13}.$$

$$M(\tilde{s}_{21}) = \mu_2 + s_{21} + \bar{w}_{21}.$$

$$M(\tilde{s}_{22}) = \mu_2 + s_{22} + \bar{w}_{22}.$$

$$M(\tilde{s}_{23}) = \mu_2 + s_{23} + \bar{w}_{23}.$$

The models for the solutions for the random effects for trait 1 each contain the same function of the fixed effects as also will be true for more complicated models and similarly the solutions for trait 2 all contain the same functions of fixed effects.

The estimate of the covariance is

$$\frac{\sum_{i=1}^3 \tilde{s}_{1i} \tilde{s}_{2i} - (\Sigma \tilde{s}_{1i})(\Sigma \tilde{s}_{2i})}{3 - 1} = 197.5.$$

$$E[\Sigma \tilde{s}_{1i} \tilde{s}_{2i}] = 3\mu_1 \mu_2 + 3\sigma_{s_1 s_2}$$

and

$$E\left[\frac{(\Sigma \tilde{s}_{1i})(\Sigma \tilde{s}_{2i})}{3}\right] = 3\mu_1 \mu_2 + \sigma_{s_1 s_2}$$

since the w_{1ij} and w_{2ik} are all uncorrelated.

A different set of constraints will change the solutions and the models for the solutions, but the estimate of the covariance will be the same with expected value $\sigma_{s_1 s_2}$.

The constraints, $\hat{s}_{13} = 0$ and $\hat{s}_{23} = 0$, result in the solutions for the s 's as 5, 25, 0, and 20, 33, 0 for traits 1 and 2 with models for the solutions excluding the w terms which are

defined by the model as uncorrelated between traits 1 and 2:

$$\begin{aligned} M(\hat{s}_{11}) &= s_{11} - s_{13} & M(\hat{s}_{21}) &= s_{21} - s_{23} \\ M(\hat{s}_{12}) &= s_{12} - s_{13} & M(\hat{s}_{22}) &= s_{22} - s_{23} \\ M(\hat{s}_{13}) &= 0 & M(\hat{s}_{23}) &= 0 \end{aligned}$$

The estimate of the covariance is as before 197.5. The expectations of the sum of products and correction term are:

$$E[\sum \hat{s}_{1i} \hat{s}_{2i}] = 4\sigma_{s_1 s_2}$$

and

$$E\left[\frac{(\sum \hat{s}_{1i})(\sum \hat{s}_{2i})}{3}\right] = 2\sigma_{s_1 s_2},$$

so that

$$E[\hat{\sigma}_{s_1 s_2}] = \sigma_{s_1 s_2}.$$

This example simply illustrates the proof of unbiasedness and invariance of this method of estimating a genetic covariance.

ACKNOWLEDGMENT

This research was supported, in part, by a grant from Eastern Artificial Insemination Cooperative, Inc., Ithaca, NY.

REFERENCES

1 Calo, L. L., L. D. Van Vleck, R. E. McDowell, and

- P. D. Miller. 1973. Simultaneous selection for milk and beef production among Holstein-Friesians. *J. Dairy Sci.* 56:1080.
- 2 Falconer, D. S. 1952. The problem of environment and selection. *Am. Nature* 86:293.
- 3 Henderson, C. R. 1953. Estimation of variance and covariance components. *Biometrics* 9:226.
- 4 Henderson, C. R., S. R. Searle, and L. R. Schaeffer. 1974. The invariance and calculation of Method 2 for estimating variance components. *Biometrics* 30:583.
- 5 Mason, I. L., and A. Robertson. 1956. The progeny testing of dairy bulls at different levels of production. *J. Agric. Sci.* 47:367.
- 6 McDowell, R. E., G. R. Wiggans, J. K. Camoens, L. D. Van Vleck, and D. G. St. Louis. 1976. Sire comparison for Holsteins in Mexico versus the United States and Canada. *J. Dairy Sci.* 59:298.
- 7 Olsen, A., J. Seely, and D. Birkes. 1976. Invariant quadratic unbiased estimation for two variance components. *Ann. Stat.* 4:878.
- 8 Patterson, H. D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545.
- 9 Robertson, A., L. K. O'Connor, and J. Edwards. 1960. Progeny testing dairy bulls at different management levels. *Anim. Prod.* 2:141.
- 10 Schaffer, L. R. 1975. Disconnectedness and variance component estimation. *Biometrics* 31:969.
- 11 Schaeffer, L. R., J. W. Wilton, and R. Thompson. 1978. Simultaneous estimation of variance and covariance components from multi-trait mixed model equations. *Biometrics* 34:199.
- 12 Van Vleck, L. D. 1963. Genotype and environment in sire evaluation. *J. Dairy Sci.* 46:983.
- 13 Wiggans, G. R., and L. D. Van Vleck. 1978. Evaluation of sires in herds feeding differing proportions of concentrates and roughages. *J. Dairy Sci.* 61:246.