

Inverse Numerator Relationship Matrix Approximation¹

J. F. TAYLOR and M. A. TOMASZEWSKI
Department of Animal Science
Texas A&M University
College Station 77843

ABSTRACT

Genetic evaluation by BLUP requires knowledge of the inverse numerator relationship matrix among breeding values. When pedigree information is complete, the true relationship matrix can be determined and an exact inverse found. When pedigree information is incomplete, an exact inverse of an approximate relationship matrix is obtainable. The usual algorithm for the inverse numerator relationship matrix may be derived assuming each unidentified parent to be a different base population animal.

INTRODUCTION

Despite a considerable amount of literature pertaining to the joint use of genetic grouping and the inverse numerator relationship matrix (NRM) in genetic evaluation models (2, 7, 10, 11, 12), it remains unclear as to just how selection is accounted for in these models. Henderson (3) suggested that use of the inverse NRM provides a reduction in prediction error variance over assuming animals unrelated but omitted discussion of bias in predicted breeding values (BV). It can be readily demonstrated that predicted BV, \hat{u} , are unbiased regardless of the NRM used, in the sense that $E[\hat{u}] = \mathbf{0}$ provided $E[\mathbf{u}] = \mathbf{0}$, although this requires repeated sampling of animals rather than of observations on existing animals. Because ignoring relationship information causes predicted BV to be regressed toward the base population mean (zero), it is unclear how predicted BV are

“unbiased” in this case. Intuitively, this seems to require the number of observations on each animal to tend to infinity in which case conditioning on realized BV, $E[\hat{u}|\mathbf{u}]$, tends to \mathbf{u} , and \hat{u} is conditionally consistent for \mathbf{u} . This was probably recognized by Henderson (3), who included genetic groups in sire evaluation when relationships were ignored. Henderson (3) stated that use of relationships reduced the number of groups necessary to account for genetic trend. That this is the case is not obvious; however, we present an example in which use of complete relationship information causes the midparental BV to be the base for regression of progeny of selected parents rather than the base generation mean.

An additional problem arises in the computation of the inverse NRM when pedigree information is incomplete. Henderson's algorithm (5) provides the correct inverse NRM when pedigree information is complete, but an exact inverse of an approximate NRM is obtained for incomplete pedigree information. Further, the approximation is based upon the assumption that each unidentified parent is a different, base population individual, which is an undesirable assumption when populations have been selected directionally for a number of generations. An example, albeit contrived, provides evidence for this concern.

Although genetic grouping is an intuitively attractive solution to this problem, it is not clear how genetic groups should be interpreted when some animals from selected generations have complete pedigree information and others have missing information. Further, Henderson's inverse (5) of an approximation of the NRM will assure an inconsistency exists between generations formed by the NRM and those formed by grouping. Henderson's approximation (5) assigns all animals with both parents unidentified as members of the base generation, whereas grouping strategies attempt to allocate these and other individuals with incomplete

Received June 13, 1988.

Accepted October 28, 1988.

¹Technical Article 23829 of the Texas Agricultural Experiment Station, College Station 77843. Project 2491, a contribution to Southern Regional Project S49, Genetic Methods of Improving Dairy Cattle for the South.

pedigree information to different generation groups according to the likely number of ancestral generations in the pedigree. The consequences of this for sire evaluation, where directional selection for production has been practiced for a number of generations, are unknown but warrant further examination.

THE INVERSE NUMERATOR RELATIONSHIP MATRIX

Consider the partitioning $u' = [u'_0 \ u'_1 \ \dots \ u'_q]$ where u_k represents BV of animals assigned to generation k by pedigree. Base generation animals in u_0 have both parents unidentified. Each animal in u_k must have parents in preceding generations, with at least one parent in u_{k-1} .

Assume no inbreeding and that all animals in generations $k > 0$ have both parents identified. The latter assumption will be relaxed later. Quaas (8, 9) suggested the recursive relationship $u = 1/2 Pu + \phi$, where for $q = 2$ generations:

$$\begin{bmatrix} u_0 \\ u_1 \\ u_2 \end{bmatrix} = 1/2 \begin{bmatrix} 0 & 0 & 0 \\ P_{10} & 0 & 0 \\ P_{20} & P_{21} & 0 \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} u_0 \\ \phi_1 \\ \phi_2 \end{bmatrix} \quad [1]$$

In general, P_{ij} ; $i = 1, \dots, q$ and $j = 0, \dots, i-1$ are known matrices with rows containing 1 in positions defining parents in generation j of animals in generation i ; and 0 otherwise. For l_i a unity vector the same length as u_i , complete pedigree information assures:

$$l_i = 1/2 \sum_{j=0}^{i-1} P_{ij} l_j$$

for $i > 0$. The vectors ϕ_i represent Mendelian sampling of parental gametes as deviations from the midparent defined by:

$$1/2 \sum_{j=0}^{i-1} P_{ij} u_j$$

Define $Var(u) = A\sigma_a^2$ for A , the numerator relationship matrix defined by Henderson (5),

and σ_a^2 , the additive genetic variance. For $u = 1/2 Pu + \phi$, we have $(I - 1/2 P)u = \phi$ and, hence, $u = (I - 1/2 P)^{-1} \phi$ since $(I - 1/2 P)^{-1}$ exists and may be formed recursively (8). Elements of ϕ are uncorrelated by definition and $Var(\phi) = D\sigma_a^2$ where for $q = 2$ and complete pedigree information:

$$D = \begin{bmatrix} I & 0 & 0 \\ 0 & 1/2I & 0 \\ 0 & 0 & 1/2I \end{bmatrix}$$

Hence, from $u = (I - 1/2 P)^{-1} \phi$ we have:

$$Var(u) = (I - 1/2 P)^{-1} Var(\phi)(I - 1/2 P')^{-1}$$

or:

$$A = (I - 1/2 P)^{-1} D (I - 1/2 P')^{-1}$$

hence:

$$\begin{aligned} A^{-1} &= (I - 1/2 P') D^{-1} (I - 1/2 P) \\ &= D^{-1} - P' - P + 1/2 P'P \end{aligned} \quad [2]$$

Henderson's algorithm (5) for rapid computation of A^{-1} with complete pedigree information follows from [2] as shown by Quaas (8).

INVERSE NUMERATOR RELATIONSHIP MATRIX APPROXIMATION

When pedigree information is incomplete, P is misspecified; u is incorrectly ordered by generation and may also be incomplete. In this case, exact inverses of approximate NRM may be derived. Suppose that u is augmented by missing parents with BV in u_m such that:

$$\begin{bmatrix} u_m \\ u \end{bmatrix} = 1/2 \begin{bmatrix} 0 & 0 \\ T & P \end{bmatrix} \begin{bmatrix} u_m \\ u \end{bmatrix} + \begin{bmatrix} u_m \\ \phi \end{bmatrix} \quad [3]$$

Known relationships among animals in u are defined by P , whereas T and u_m may be defined variously to specify missing parents.

Null rows of T corresponding to u_0 in u defines u_m as base generation animals. Alternative representations to [3] could allow a generation ordering of animals in u_m and, hence, of animals in u , as in assigning animals to genetic groups (7, 11, 12). ϕ is as before.

For animals in u_m unrelated:

$$\text{Var} \begin{bmatrix} u_m \\ u \end{bmatrix} = \begin{bmatrix} I & A_{mu} \\ A'_{mu} & A_{uu} \end{bmatrix} \sigma_a^2 \quad \text{and} \quad \text{Var} \begin{bmatrix} u_m \\ \phi \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & D \end{bmatrix} \sigma_a^2 \quad \text{for } D \text{ as before.}$$

From the approach used in the derivation of [2] use of the recursive relationship [3] yields:

$$\begin{bmatrix} I & A_{mu} \\ A'_{mu} & A_{uu} \end{bmatrix}^{-1} = \begin{bmatrix} I + 1/4T'D^{-1}T & -1/2T'D^{-1}(I - 1/2P) \\ -1/2(I - 1/2P')D^{-1}T & (I - 1/2P')D^{-1}(I - 1/2P) \end{bmatrix} \quad [4]$$

Absorbing the relationships involving u_m [5, equation (20)] provides:

$$\begin{aligned} A_{uu}^{-1} &= (I - 1/2P') [D^{-1} - 1/4D^{-1}T (I + 1/4T'D^{-1}T)^{-1} T'D^{-1}] (I - 1/2P) \\ &= (I - 1/2P')(D + 1/4TT')^{-1} (I - 1/2P) \end{aligned} \quad [5]$$

With no missing pedigree information, definition of $T = 0$ assures equivalency of [2] and [5].

Henderson's Approximation

Suppose that u is ordered by pedigree as defined previously and that all animals with both parents unidentified are included in u_0 . Each animal with a single parent unidentified has that parent assigned as a unique animal in u_m . Then T is null except for a diagonal band for animals in u_i ; $i > 0$ with elements 0 for animals with both parents identified or 1 for animals with a single unidentified parent. Consequently, $(D + 1/4TT')^{-1}$ is diagonal with diagonals 1 if both parents are unidentified, 4/3 if one parent is identified, or 2 if both parents are identified. Henderson's algorithm (5) follows from [5].

PREDICTION OF BREEDING VALUES

The usual form for the mixed linear model for BV is:

$$y = X\beta + Zu + e$$

$$E \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} u \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_a^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix}$$

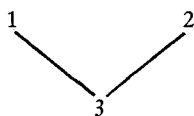
for which solutions to the mixed model equations (MME):

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + A^{-1}\lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad [6]$$

with $\lambda = \sigma_e^2/\sigma_a^2$ provide $X\hat{\beta}$ as BLUE of $X\beta$ and \hat{u} as BLUP of realized values of u .

Henderson (4) has generalized [6] to include various classes of selection models in which moments of the joint normal distribution of y , u , and e are altered by the selection practiced. Quaas and Pollak (10) demonstrated that genetic group models arise as a particular case of Henderson's (4) $L'u$ selection. In practice, we are concerned with the assumption that $E[u] = 0$ when selection has occurred. Because $u = (I - 1/2P)^{-1} \phi$ with P known, we have $E[u] = (I - 1/2P)^{-1} E[\phi]$. Hence, $E[\phi] = 0$ is sufficient for $E[u] = 0$. That is, regardless of parental selection, [6] applies provided base generation breeding values and within generation Mendelian sampling effects are drawn at random from populations with null means. If, within generation, animals with the better combinations of parental gametes are selected to produce records, $E[\phi] \neq 0$, and conditional models of the form (4) are necessary.

Henderson (3) suggested that use of A^{-1} in [6] reduced the need for genetic group models. Intuitively, this seems to be the case when pedigree information is complete and P is completely specified, under the assumption that $E[\phi] = 0$. Consider the pedigree:



for which:

$$A^{-1} = \begin{bmatrix} 3/2 & 1/2 & -1 \\ 1/2 & 3/2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

From [6] we have $(Z'Z + A^{-1}\lambda)\hat{u} = Z'(y - X\hat{\beta})$ and under the simple animal model $y_{ij} = \mu + u_i + e_{ij}$, the equation for animal 3 with n observations is:

$$(n + 2\lambda)\hat{u}_3 - \lambda\hat{u}_1 - \lambda\hat{u}_2 = y_3 - n\hat{\mu} \quad [7]$$

where y_3 is the total of observations for animal 3, and u_1 , u_2 , and u_3 are BV of animals 1, 2, and 3, respectively. Rearrangement of [7] yields:

$$\begin{aligned} \hat{u}_3 &= 1/2(\hat{u}_1 + \hat{u}_2) \\ &+ (y_3 - n\hat{\mu} - n/2\hat{u}_1 - n/2\hat{u}_2)/ \\ &(n + 2\lambda) \end{aligned} \quad [8]$$

Although BV of animals 1 and 2 are regressed toward the null base population mean, inclusion of A^{-1} causes the Mendelian effect for animal 3:

$$\hat{\phi}_3 = (y_3 - n\hat{\mu} - n/2\hat{u}_1 - n/2\hat{u}_2)/(n + 2\lambda)$$

to be regressed toward the midparent. Although $\hat{u}_1 + \hat{u}_2 = 0$ in this example, this is not generally the case with more than two animals comprising the base population. Kennedy et al. (6) provide a similar example.

In sire or animal models, but not sire-maternal grandsire models with missing pedi-

gree information, the row sum of Z is a unity vector $1_r'Z' = 1_N'$ for $Z N \times r$. Any model including μ or a fixed factor cross classified with the random animal effect, has 1_N as a column of X or as a row sum across levels of the factor. For these models [6] yields $1_r'A^{-1}\hat{u} = 0$. Because $1_r'A^{-1} = (A^{-1}1_r)'$ with complete pedigree information we have from [2]:

$$A^{-1}1_r = (I - 1/2P')D^{-1}(I - 1/2P)1_r$$

Because by generation:

$$1_i = 1/2 \sum_{j=0}^{i-1} P_{ij}1_j$$

for $i > 0$ we have $(I - 1/2P)1_r = w = [1_0' 0_1' \dots 0_q']'$. Because $(I - 1/2P')D^{-1}$ is upper triangular, we have $A^{-1}1_r = w$ and $1_r'A^{-1}\hat{u} = 0$ requires the sum of predicted BV of base generation animals to be null. This is consistent with the model parameterization with $E[u] = 0$, which requires $E[\phi] = 0$ and, hence, $E[u_0] = 0$. However, $E[u] = 0$ does not seem to imply that BV of animals in selected generations u_i $i > 0$ are drawn at random from populations with null means. Equation [8] suggests that as the number of observations on each animal increases, predicted BV converge to true BV, even when directional parental selection has occurred.

When pedigree information is incomplete, Henderson's approximation (5) for A^{-1} using result [4] leads to $1_r'A^{-1}\hat{u} = 0$ requiring the sum of BV of base generation animals and of missing parents in u_m treated as base generation animals to sum to zero. Absorbing relationships for u_m as in [5] leads to prediction of BV of missing parents from existing parent-progeny BV and w will contain nonzero elements in appropriate parent-progeny positions for animals (progeny) with a missing parent. Because incomplete pedigree information is the norm rather than exception, genetic grouping is necessary to avoid treating missing parents as base generation animals.

EXAMPLE

Table 1 contains true BV, pedigrees, generation number, and observations on 11 animals with two generations of parental selection. Additive genetic variance among base generation animals (1 to 6) is $\sigma_a^2 = 140$ and true BV

TABLE 1. True breeding values, pedigrees, generation, and observation.

Animal	Breeding value	Sire ¹	Dam ¹	Generation ²	Observation
1	-15	0	0	0	78
2	-10	0	0	0	103
3	-5	0	0	0	114
4	5	0	0	0	93
5	10	0	0	0	87
6	15	0	0	0	125
7	-7	3	2	1	95
8	-4	3	4	1	123
9	9	5	4	1	91
10	22	5	6	1	109
11	11	10	8	2	113

¹ Unidentified parents coded as 0.

² Base, first, and second generations.

of first (7 to 10) and second (11) generation animals generated from [1] with Mendelian effects normally and independently distributed (NID) (0, $\sigma_a^2/2$). Errors were NID (0, σ_e^2) with $\sigma_e^2 = 280$ and a single observation per animal generated additively according to $y_{ij} = \mu + u_i + e_{ij}$ for $\mu = 100$ and u_i true BV.

The inverse NRM from [2] for animals ordered 1 through 11 is:

$$A^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3/2 & 1/2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 2 & 1/2 & 0 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 2 & 1/2 & 0 & 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 2 & 1/2 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 3/2 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 0 & 5/2 & 0 & 1/2 & -1 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 1/2 & 0 & 5/2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 2 \end{bmatrix} \quad [10]$$

$1'A^{-1} = [1'_6 \ 0'_4 \ 0'_1]$ identifying base generation animals and with 0_j a vector of zeros of length j .

The MME in [6] for the one-way random model are constructed with $X = 1_{11}$, $Z = I_{11}$, $\hat{\beta} = \hat{\mu}$, $\hat{u}' = [\hat{u}_1 \dots \hat{u}_{11}]'$, $\lambda = 2.0$, and A^{-1} as in [10]. The unique solution to this system is $\hat{\mu} = 101.66$ and $\hat{u}' = [-7.89 \ -0.77 \ 5.76 \ -1.05 \ -4.87 \ 8.83 \ .66 \ 6.65 \ -4.50 \ 3.55 \ 6.35]$. The simple mean of predicted BV within generations provides evidence of genetic trend as 0, 1.59, and 6.35, respectively.

Suppose now that animals 3 and 8 are missing from the pedigrees of animals 8 and 11, respectively. Approximation [5] provides:

$$A_{uu}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3/2 & 1/2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 3/2 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 11/6 & 1/2 & 0 & 0 & -2/3 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 2 & 1/2 & 0 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1/2 & 3/2 & 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2/3 & 0 & 0 & 0 & 4/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 0 & 0 & 0 & 7/3 & -2/3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -2/3 & 4/3 \end{bmatrix}$$

Now $I'A_{uu}^{-1} = [1 \quad 1 \quad 1 \quad 2/3 \quad 1 \quad 1 \quad 0 \quad 2/3 \quad 0 \quad -1/3 \quad 2/3]$
 and $I'A_{uu}^{-1}\hat{u} = \hat{u}_1 + \hat{u}_2 + \hat{u}_3 + \hat{u}_4 + \hat{u}_5 + \hat{u}_6 + 2/3(\hat{u}_8 - 1/2\hat{u}_4) + 2/3(\hat{u}_{11} - 1/2\hat{u}_{10}) = 0.$

Suppose u_{m1} represents the BV of the missing parent of animal 8. We have $u_8 = 1/2u_4 + 1/2u_{m1} + \phi_8$ and, hence, $Var(u_8 - 1/2u_4) = 3/4\sigma_a^2$ and $Cov(u_{m1}, u_8 - 1/2u_4) = 1/2\sigma_a^2$. Because u_{m1} has no observations, the regression of u_{m1} on $u_8 - 1/2u_4$ is $2/3$ providing $\hat{u}_{m1} = 2/3(\hat{u}_8 - 1/2\hat{u}_4)$ as the pedigree evaluation of the missing parent.

Replacing A^{-1} by A_{uu}^{-1} in the MME provides $\hat{\mu} = 102.79$ and $\hat{u}' = [-8.26 \quad -1.07 \quad 2.60 \quad -1.80 \quad -5.32 \quad 8.44 \quad -0.95 \quad 4.86 \quad -5.21 \quad 3.12 \quad 3.92]$. Predicted BV of missing parents are $2/3(\hat{u}_8 - 1/2\hat{u}_4) = 3.84$ and $2/3(\hat{u}_{11} - 1/2\hat{u}_{10}) = 1.57$ and the sum $I'A_{uu}^{-1}\hat{u} = 0$ as required. The simple means of predicted BV within generations are now -0.90 , $.46$, and 3.92 , respectively. Although genetic trend remains evident, the BV of animal 11 is undervalued by treating the missing parent as from the base population. The negative mean of predicted BV of true base generation animals reveals that missing parents were on average superior to the base generation null mean. That is, these animals were selected.

Although the artificiality of this example is apparent, it serves to illustrate the potentially undesirable property of utilizing an approximate A^{-1} , which assigns missing parents to the base population.

DISCUSSION

It seems that use of the inverse numerator relationship matrix avoids the need for genetic grouping when pedigree information is complete and the better progeny are not selected to produce records. However, missing pedigree information in genetic evaluation is the norm rather than exception. Consequently, formulation of genetic groups as in other work (7, 11, 12) is essential to circumvent potentially undesirable consequences associated with inverse NRM approximation. However, since solutions for group effects are dependent on the structure of A^{-1} (1, 7), it is evident that a theoretical examination of grouping models is required in order that animal breeders understand what group solutions estimate. This is particularly

true for approximations of A^{-1} where it is known a priori that certain missing parents are not base generation members but are forced to be so. In this case, an obvious inconsistency exists between generations defined by the inverse NRM and by genetic groups.

REFERENCES

- 1 Everett, R. W., and J. F. Keown. 1985. Mixed model sire evaluation with dairy cattle—experience and genetic gain. *J. Anim. Sci.* 59:529.
- 2 Famula, T. R., E. J. Pollak, and L. D. Van Vleck. 1983. Genetic groups in dairy sire evaluation under a selection model. *J. Dairy Sci.* 66:927.
- 3 Henderson, C. R. 1975. Use of relationships among sires to increase accuracy of sire evaluation. *J. Dairy Sci.* 58:1731.
- 4 Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423.
- 5 Henderson, C. R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69.
- 6 Kennedy, B. W., L. R. Schaeffer, and D. A. Sorenson. 1988. Genetic properties of animal models. *J. Dairy Sci.* 71(Suppl. 2):17.
- 7 Pollak, E. J., and R. L. Quaas. 1983. Definition of group effects in sire evaluation models. *J. Dairy Sci.* 66:1503.
- 8 Quaas, R. L. 1984. Linear prediction *in* Use of mixed models for prediction and for estimation of (co)variance components. BLUP school handbook. R. L. Quaas, R. D. Anderson, and A. R. Gilmour, ed. Anim. Genet. Breeding Unit, Univ. New England, Armidale, NSW, 2351, Australia.
- 9 Quaas, R. L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71:1338.
- 10 Quaas, R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. *J. Dairy Sci.* 64:1868.
- 11 Robinson, G. K. 1986. Group effects and computing strategies for models for estimating breeding values. *J. Dairy Sci.* 69:3106.
- 12 Thompson, R. 1979. Sire evaluation. *Biometrics* 35:339.