

Animal Genetics and Breeding Unit¹, University of New England, Armidale, NSW 2351, Australia

Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models

BY B. TIER and K. MEYER

Summary

A method for approximating prediction error variances and covariances among estimates of individual animals' genetic effects for multiple-trait and random regression models is described. These approximations are used to calculate the prediction error variances of linear functions of the terms in the model. In the multiple-trait case these are indexes of estimated breeding values, and for random regression models these are estimated breeding values at individual points on the longitudinal scale. Approximate reliabilities for terms in the model and linear functions thereof are compared with corresponding reliabilities obtained from the inverse of the coefficient matrix in the mixed model equations. Results show good agreement between approximate and 'true' values.

Zusammenfassung

Approximative Schätzung der Standardfehler der Kovarianzen zwischen additiv genetischen Tiereffekten in Mehrmerkmals- und zufälligen Regressionsmodellen

Beschrieben wird eine Methode zur approximativen Schätzung der Standardfehler der Varianzen und Kovarianzen zwischen Schätzwerten für tierindividuelle genetische Effekte für Mehrmerkmals- und zufällige Regressionsmodelle. Diese approximative Methode wird benutzt, um die Restvarianz von linearen Funktionen der Effekte im Modell zu schätzen. Im Falle von Mehrmerkmalsmodellen sind dies Indices der geschätzten Zuchtwerte, und für zufällige Regressionsmodelle sind dies geschätzte Zuchtwerte an individuellen Punkten der Funktion. Die approximativen Sicherheiten der Effekte im Modell und linearen Funktionen dieser werden mit den Sicherheiten aus der Inversen der Koeffizientenmatrix im Gleichungssystem des gemischten Modells verglichen. Die Ergebnisse zeigen eine gute Übereinstimmung zwischen den approximativen und den 'wahren' Werten.

Introduction

Most genetic evaluation schemes involve models which fit several additive genetic effects for each animal. These are generally multiple trait models, which combine information on correlated traits, to predict the genetic merit of animals for many traits simultaneously. For example, BREEDPLAN, the Australian genetic evaluation scheme for beef cattle, provides the estimated breeding values (EBVs) for many traits simultaneously (JOHNSTON et al. 1999), currently 25.

Recently, there has been an increased interest in the so-called random regression (RR) models for traits which are recorded repeatedly per animal, such as longitudinal data. RR models are similar to multiple-trait models in that a number of correlated additive genetic effects, namely regression coefficients, are estimated for each individual. Estimates of genetic RR coefficients provide a complete trajectory of genetic merit and EBVs for any

¹AGBU is a joint venture of NSW Agriculture and University of New England.

point on the longitudinal scale can be obtained by evaluating the regression equations at that point.

When comparing EBVs, we are interested not only in their values but also in their credibility. Accuracies or reliabilities describe the value of the information used when calculating EBVs, and are determined by the prediction error variances (PEV) of the EBVs. If the inverse of the coefficient matrix in the mixed model equations (MME) is known, PEVs can be calculated directly from the diagonal elements of the inverse. Generally, however, direct inversion is feasible only for small populations, even if sparse matrix techniques (MISZTAL and PEREZ-ENCISCO 1993; THOMPSON et al. 1994) are used. HARVILLE (1999) described a method for approximating selected elements of the inverse of a matrix with a Gibbs sampler, which is suitable for much larger problems, but this is computationally demanding. For instance, MEYER (2004) used this procedure to obtain accuracies of genetic evaluation for an RR model involving more than 2.2 million equations. For routine applications with millions of equations, however, inverting the coefficient matrix of the MME matrix is infeasible with either method. Consequently, a variety of methods have been developed for approximating PEVs and accuracies.

For animal model genetic evaluation, approximation procedures progressed from simple approaches which approximated the diagonal of the inverse coefficient matrix for single-trait analyses (MISZTAL and WIGGANS 1988; MEYER 1989) and multiple-trait models (TIER et al. 1991; GENGLER and MISZTAL 1996) to methods based on summarizing all information available for an individual into a single number, and using this to derive an accuracy. Sources of information for an animal are its own records, records on progeny and other descendants, parents and more distant ancestors, and other collateral relatives. These can be summarized into equivalent numbers of progeny (ENP), by treating each animal as if it were a sire in a sire model equivalent to the animal model considered. The ENP concept has been applied to single-trait (KOOTs et al. 1997), multiple-trait (GRASER and TIER 1997; HARRIS and JOHNSON 1998; LIU et al. 2001), and RR models (JAMROZIK et al. 2000); see JAMROZIK et al. 2000 for a recent overview and further references. Recently, FOUILLOUX and LALOË (2001) suggested a method to determine prediction error variances of breeding values or of contrasts between breeding values empirically by sampling. This method requires solving the MME numerous times which is currently impractical for very large populations.

To approximate PEVs of EBVs derived from RR models, JAMROZIK et al. (2000) extended the idea of ENP to EBVs calculated as a function of the estimated regression coefficients. This involves two steps. First, the diagonal block of the coefficient matrix in the MME corresponding to each individual is calculated. Secondly, the accuracy at a particular function is calculated from this, and translated into an ENP. Thereafter, the method is similar to the others cited above, involving two passes through the pedigree combining the values for all sources of data. This is performed separately for each function of interest, but gives no approximation for accuracies of the RR coefficients.

Whilst PEVs of RR can be derived, in essence, analogous to those in a multi-trait model, corresponding prediction error covariances (PEC) are required to obtain PEVs of EBVs at individual points as the variance of a linear function of correlated RR coefficients. Another scenario where we might be interested in PECs as well as PEVs is the approximation of PEVs of selection indexes, which summarize information from numerous EBVs obtained in a multi-trait genetic evaluation. For instance, BREEDOBJECT (BARWICK and HENZELL 1998) provides customized selection indexes involving the EBVs generated by BREEDPLAN. Availability of PECs would allow the accuracies to be calculated for any customized index. However, the problem of approximating PECs has received little attention.

This paper describes a method to approximate both PEVs and PECs simultaneously, extending the method of ENPs to replace the single number by a matrix of contributions for each individual. Approximate accuracies for EBVs and linear functions of EBVs are

obtained from this matrix, for both RR and multiple-trait models, and compared with results from other methods for several data sets.

Approximating prediction error (co)variances

Let \mathbf{C} denote the coefficient matrix of the MME. As emphasized above, prediction error (co)variances between effects estimated are given by the corresponding elements of its inverse, \mathbf{C}^{-1} , diagonal elements in the case of PEVs. Approximation methods suggested so far have thus attempted to adjust diagonal elements of \mathbf{C} for ‘links’ with other effects in the model, so that reciprocals of the adjusted diagonals closely resemble the diagonal elements of \mathbf{C}^{-1} .

For a single-trait, single-record model, for instance, MEYER (1989) selected the coefficients of \mathbf{C} corresponding to an animal, its parents and the fixed subclass in which it had a record. Inverting the resulting 4×4 matrix (assuming both parents to be known) then yielded an approximation of the corresponding elements of \mathbf{C}^{-1} . Clearly, the quality of approximation depended on the amount of non-zero off-diagonal elements of \mathbf{C} , linking the animal and parents considered with other effects in the model, which were ignored. For animals with numerous progeny, these numbers could be substantial. Hence a simple, sequential procedure was suggested, which first adjusted diagonals of animals with records for limited subclass sizes, then accumulated adjustments to parents’ diagonals for limited information on their progeny, and finally adjusted diagonals of progeny for the adjusted diagonals of their parents, taking care not to double count the animal considered.

The principles involved in the more recently suggested procedures (see JAMROZIK et al. 2000) and our method are no different in that we need to account for information on each animal provided by its own records, its parents and other known ancestors, its progeny and further descendants, and any collateral relatives. This generally involves at least two passes through the pedigree, from youngest to oldest (‘upwards’) and from oldest to youngest (‘downwards’). In contrast to most previous methods, with the exception of LIU et al. (2001) however, we are trying to approximate the $k \times k$ diagonal block for each animal corresponding to all additive genetic effects fitted.

Models of analysis

Multi-trait model

Consider k traits with a single record per trait. In the usual notation, let

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (1)$$

denote the multi-trait animal model, with \mathbf{y} the vector of observations, \mathbf{b} the vector of fixed effects, \mathbf{a} the vector of additive genetic values, \mathbf{e} the vector of residuals, and \mathbf{X} and \mathbf{Z} the incidence matrices relating observations to effects. Assume that all vectors are ordered according to traits within animal, and let

$$\begin{aligned} \text{var}(\mathbf{a}) &= \mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0 \\ \text{var}(\mathbf{e}) &= \mathbf{R} = \Sigma^+ \mathbf{R}_i \end{aligned}$$

where \mathbf{G}_0 is the $k \times k$ matrix of genetic covariances among traits, and \mathbf{R}_i is the residual covariance matrix among the traits recorded for the i th animal. \mathbf{R}_i is equal to \mathbf{R}_0 , the $k \times k$ residual covariance among all traits, for animals with all k traits recorded, or the submatrix of \mathbf{R}_0 obtained by deleting rows and columns pertaining to traits not recorded on the i th animal. ‘ \otimes ’ denotes the Kronecker product and ‘ Σ^+ ’ stands for the direct matrix sum. \mathbf{A} is the numerator relationship matrix between animals. Assume further that animals are in

order of age or generation from oldest to youngest, i.e. values for parents in \mathbf{a} always precede those of their progeny.

Random regression model

With repeated records per animal and trait in an RR model, we need to expand (1) to include permanent environmental (PE) effects of the animal

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{a} + \mathbf{W}\mathbf{p} + \mathbf{e} \quad (2)$$

with \mathbf{y} , \mathbf{b} , \mathbf{e} and \mathbf{X} as above, and \mathbf{p} and \mathbf{a} the vectors of RR coefficients for animals' PE and additive genetic effects, respectively. Assume there are k covariables used to model the animals' genetic effects; \mathbf{W} and \mathbf{Z} are incidence matrices containing covariables relating regression coefficients to the functions of the continuous scale (time) along which observations have been recorded.

Residuals e_j in (2) represent temporary environmental effects are commonly assumed to be independently distributed with variances σ_m^2 . For a single-trait RR model

$$\begin{aligned} \text{var}(\mathbf{p}) &= \mathbf{P} = \mathbf{I} \otimes \mathbf{P}_0 \\ \text{var}(\mathbf{a}) &= \mathbf{G} = \mathbf{A} \otimes \mathbf{G}_0 \\ \text{var}(\mathbf{e}) &= \mathbf{R} = \text{diag}\{\sigma_m^2\} \end{aligned}$$

where \mathbf{P}_0 is the $k \times k$ of permanent environmental covariances among RR coefficients. This model can be readily extended to multiple traits.

Value of observations for an animal

The first step of the approximation involves determining the value of animals' own records. Let \mathbf{D}_i , of size $k \times k$, denote the block representing the contribution of records for animal i to information on its own EBVs. This is derived from the data part of the MME, and is model dependent.

Multi-trait model

With PE due to the animal included in the residual in (1), \mathbf{D}_i is the submatrix of \mathbf{C} corresponding to the animal's genetic effects

$$\mathbf{D}_i = \mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i \quad (3)$$

with \mathbf{Z}_i the submatrix of \mathbf{Z} and for the i th animal and \mathbf{R}_i as described above. In essence this is \mathbf{R}_i expanded to dimension $k \times k$ by adding the rows and columns of zeroes for missing traits.

Genetic evaluation models generally include some 'contemporary group' effects among the fixed effects fitted, e.g. herd-test-day effects for dairy cattle data or herd-paddock-sex of calf-date of weighing subclasses for weight records of beef cattle. Definition (3) does not account for limited subclass sizes. When individual i has few contemporaries, (3) should be modified to be

$$\mathbf{D}_i = \mathbf{Z}'_i (\mathbf{R}_i^{-1} - \mathbf{R}_i^{-1} (\mathbf{S}_i^{-1}) \mathbf{R}_i^{-1}) \mathbf{Z}_i \quad (4)$$

where \mathbf{S}_i is the block of \mathbf{C} pertaining to the contemporary groups of which animal i is a member. This discounts the value of the observations to accommodate the limited number

of contrasts between this animal and others, and is the multi-trait equivalent to replacing ‘1’ by $(n - 1)/n$ in a univariate, single-record scenario (with n the subclass size).

Random regression model

To obtain the equivalent in the RR model, we need to ‘absorb’ the animals’ PE effects into the corresponding genetic effects

$$\mathbf{D}_i = \mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i - \mathbf{Z}'_i \mathbf{R}_i^{-1} \mathbf{W}_i (\mathbf{W}'_i \mathbf{R}_i^{-1} \mathbf{W}_i + \mathbf{P}_0^{-1})^{-1} \mathbf{W}'_i \mathbf{R}_i^{-1} \mathbf{Z}_i \quad (5)$$

with \mathbf{W}_i the submatrix of \mathbf{W} for animal i . As above, contributions from observations in limited subclass size can be adjusted by using a weight of $w_m = (n_m - q)/n_m$ for the m th record, with n_m the size of subclass to which the record belongs and q the number of ‘repeated’ records it has in that subclass. If \mathbf{R}_i is diagonal, this is achieved simply by replacing \mathbf{R}_i with $\mathbf{R}_i^* = \text{diag}\{w_m \sigma_j^2\}$ in (5).

Value of observations on descendants

In the second step, we accumulate the value of records on progeny and other descendants for each animal, processing the pedigree from youngest to oldest. Conceptually, this is obtained by assuming that each progeny has only one parent known and that this parent has no further information, building the MME for the animal and the parent, and then ‘absorbing’ the animal equations into those of the parent. Let \mathbf{E}_i denote the $k \times k$ block of contributions for animal i and p_i the number of progeny it may have. \mathbf{E}_i can be calculated as

$$\mathbf{E}_i = \frac{1}{3} \mathbf{G}_0^{-1} - \frac{4}{9} \mathbf{G}_0^{-1} \left(\mathbf{D}_i + \sum_{l=1}^{p_i} \mathbf{E}_l + \frac{4}{3} \mathbf{G}_0^{-1} \right)^{-1} \mathbf{G}_0^{-1} \quad (6)$$

This block is accumulated separately for both sire and dam of animal i . As the pedigree is processed ‘upwards’ any block \mathbf{E}_l required for the progeny of animal i has already been fully determined. Note that an animal contributes the same quantity to each of its parents.

Equation (6) is adequate if animal i has been directly contrasted with a relatively few half-sibs. If the animal’s records were in contemporary groups which included many of its half-sibs, \mathbf{D}_i in (6) would give an overestimate of the individual’s contribution to its parents. As above, we can discount the informative value of records by weighting contributions with a factor determined by the proportion of sibs in a subclass. Let \mathbf{Q}_i denote a diagonal matrix of weights $w_m \leq 1$. For the multi-trait model, $w_m = \sqrt{(n_m - q_m)/n_m}$ with n_m the total number of records in the subclass for trait m and q_m the number of sibs of animal i in the subclass. Replacing \mathbf{D}_i in (6) by $\mathbf{D}_i^* = \mathbf{Q}_i \mathbf{D}_i \mathbf{Q}_i$ then reduces the value of \mathbf{E}_i appropriately.

Value of observations on ancestors

In the third and final step, we accumulate the values of parents, ancestors and collateral relatives for each animal by processing the pedigree from oldest to youngest.

In the previous step, the value of descendants for all animals was accumulated. Hence the block \mathbf{E}_j for parent j of animal i includes the contribution for animal i . This has to be removed prior to accumulating the contribution of the parent to progeny i , to avoid double counting. The adjusted block is

$$\mathbf{E}_j^* = \frac{1}{3}\mathbf{G}_0^{-1} - \frac{4}{9}\mathbf{G}_0^{-1} \left(-\mathbf{E}_i + \mathbf{F}_j + \frac{4}{3}\mathbf{G}_0^{-1} \right)^{-1} \mathbf{G}_0^{-1} \quad (7)$$

where \mathbf{F}_j is the $k \times k$ block for parent j in which contributions from all sources of information has been accumulated. As the pedigree is processed ‘downwards’, blocks \mathbf{F}_j are already finalized when the contribution of parent j to animal i is to be calculated.

The ‘final’ block \mathbf{F}_i for an animal is then the sum of blocks for its parents, ‘adjusted’ for the animal, the block for the contribution from its own records, and the blocks for its progeny

$$\mathbf{F}_i = \sum_{j=1}^{t_i} \mathbf{E}_j^* + \mathbf{D}_i + \sum_{l=1}^{p_i} \mathbf{E}_l \quad (8)$$

where $t_i = 0, 1$, or 2 denotes the number of known parents for animal i .

Prediction error covariances

Matrices \mathbf{T}_i of approximate PEV and PEC for the k genetic values estimated for animal i are then obtained as

$$\mathbf{T}_i = (\mathbf{F}_i + \mathbf{G}_0^{-1})^{-1} \quad (9)$$

Let pev_{im} denote the m th diagonal element of \mathbf{T}_i and var_m the corresponding genetic variance, i.e. the m th diagonal element of \mathbf{G}_0 . Reliabilities of estimated genetic RR coefficients and multi-trait EBVs are then calculated as

$$\rho_{im}^2 = 1 - \frac{\text{pev}_{im}}{\text{var}_m} \quad (10)$$

This ignores any inbreeding, as the approximation procedure did not account for inbreeding. More generally, the denominator in (10) is the variance of the animal’s true genetic value. This depends on the animal’s degree of inbreeding. Hence, if inbreeding has been accounted for as is the case for pev_{im} derived from the inverse of \mathbf{C} , (10) is replaced by

$$\rho_{im}^2 = 1 - \frac{\text{pev}_{im}}{(1 + \mathcal{F}_i)\text{var}_m} \quad (11)$$

for EBV m for animal i with inbreeding coefficient \mathcal{F}_i (TIER et al. 1991).

Indexes of EBVs from multiple-trait models are obtained by multiplying a vector of weights by the vector of EBVs. Similarly, EBVs for specific points on the longitudinal scale can be obtained by evaluating the regression equation given by the estimated RR coefficients for animal i . These are linear functions of the EBVs. Hence, the genetic and prediction error variances can be calculated as

$$\text{pev}_{i\mathcal{J}} = \boldsymbol{\phi}' \mathbf{T}_i \boldsymbol{\phi} \quad (12)$$

and

$$\text{var}_{\mathcal{J}} = \boldsymbol{\phi}' \mathbf{G}_0 \boldsymbol{\phi} \quad (13)$$

where $\boldsymbol{\phi}$ denotes the vector of weights for index \mathcal{J} (multiple-trait model) or the vector of covariables evaluated at the point of interest \mathcal{J} (RR model).

Application

To examine the usefulness of this method, approximated reliabilities for EBVs, RR coefficients and indexes were compared with 'true' values and the approximation of JAMROZIK et al. (2000). True values were calculated from the inverse of the coefficient matrix of the MME.

Data sets

Four data sets were used to examine the qualities of the method of approximating the prediction error covariances. Three data sets were used to test the method for RR models and one for multiple traits and their indexes.

For the RR models, weights of beef cattle from birth to 730 days of age were considered. The first data set (Hereford) comprised a single Hereford herd with 4013 records on 1095 animals, with the majority of animals in the data having four records. The second data set (Wokalup) comprised the weights for animals in two herds in the Wokalup selection experiment, which involved monthly weight recording, yielding up to 26 records per animal (see MEYER et al. 1993 for details). The third data set (Murray Grey) consisted of weights for all Murray Grey herds registered with Australia's National Beef Recording Scheme. Less than 30% of these animals had more than two weights recorded.

The final data set (Multiple Trait) consisted of the same Murray Grey herds but in this instance five traits were analysed. They were 600-day weight, and two 'carcass traits' measured by live ultrasound scanning between 300 and 700 days of age – eye muscle area and fat depth on the rump. As implemented in BREEDPLAN, carcass traits for bulls were treated as different, correlated traits to those of heifers or steers, yielding five traits in total. Two arbitrary vectors of index weights, $\phi'_1 = \{1, 1, 1, 1, 1\}$ and $\phi'_2 = \{1, 2, 2, 0.5, 0.5\}$, were used to generate indexes from the EBVs. Characteristics of the data structure are summarized in Table 1.

Table 1. Data structure

	Hereford	Wokalup	Murray Grey	Multiple trait	
Recorded animals	1095	7305	117 977	28 768	
Ancestors	750	1138	58 396	21 659	
Contemporary groups	230	11 417	54 263	7407	
Total records	4013	75 829	227 219	47 655	
Animals with					
1 record	7	800	58 396	17 838	
2 records	17	545	26 840	2973	
3 records	312	158	19 755	7957	
4 records	759	271	10 776	0	
>4 records	0	5531	2170	0	
<i>Multiple trait analysis by trait</i>					
		Eye muscle area		Rump fat	
	600-day weight	Heifers ^a	Bulls	Heifers ^a	Bulls
Number of records	25 809	5700	5214	5744	5188
Contemporary groups	4856	584	690	588	689

^aIncluding steers.

Analyses

Model

The model of analysis for the first three data sets fitted a cubic RR on Legendre polynomials (LP) of age at recording for direct genetic, maternal genetic, direct permanent environmental and maternal permanent environmental effects. For the three RR analyses, the assumed variances among RR coefficients and heterogeneous measurement error variance estimates for Hereford cattle were those estimated by MEYER (2002), and are presented in Table 2.

Fixed effects fitted were contemporary groups (CG) and a quartic regression on LP of age, with CG defined as herd–sex–management group–year–month of weighing subclasses for birth weights, and herd–sex–management group–date of weighing subclasses otherwise. CG were divided further by applying an ‘age slicing’ of 45 days up to 300 days, and 60 days for higher ages.

For the multiple-trait analysis a simple model with a direct effect of the animal, a contemporary group defined analogously, and a linear regression on age at recording and a quadratic regression on dam age, fitted as a fixed effects, was used for all traits. Genetic and residual covariances among the traits are shown in Table 3.

Prediction error variances and covariances

Diagonal blocks of the inverse of the coefficient matrix in the MME, pertaining to sets of RR coefficients for direct genetic effects and for the multiple-trait analysis, were obtained for each data set in up to three different ways.

- *Inverse.* Direct inversion of the coefficient matrix using sparse matrix inversion was feasible only for the first data set.

Table 2. Covariances among random regression coefficients for genetic (upper triangle) and permanent environmental (lower triangle) effects^a

	Direct effects				Maternal effects					
	Intercept	Linear	Quadratic	Cubic	Intercept	Linear	Quadratic	Cubic		
Intercept	776.11	607.96	353.57	51.47	20.66	274.52	121.01	34.36	-2.83	18.49
Linear	338.55	281.31	244.85	71.37	22.53	36.85	9.94	16.87	2.97	3.48
Quadratic	-34.00	79.78	138.09	69.12	28.63	-16.85	2.25	8.13	5.79	1.64
Cubic	31.28	8.48	51.52	49.11	16.38	61.15	8.48	-0.81	16.19	5.97

^aFrom Meyer (2002).

Table 3. Genetic (upper triangle) and residual (lower triangle) covariances among traits in the multiple trait model^a

	600-day weight	Eye muscle area		P8 fat depth		
		Heifers	Bulls	Heifers	Bulls	
600-day weight	924	395.9	24.3	28	0.92	-0.049
Eye muscle area						
Heifers	77	21.3	9.4	9.02	0.256	-0.024
Bulls	71	0	33.2	12	0.336	-0.209
P8 fat depth						
Heifers	19.9	2.59	0	2.59	1.66	0.921
Bulls	9.33	0	1.2	0	0	0.888

^aFrom Meyer et al. (2003).

- *Harville*. A Gibbs sampling algorithm suggested by HARVILLE (1999) was used to approximate the diagonal blocks of interest for all data sets. For this, a total of 400 000 samples were drawn, discarding the first 20 000 samples as 'burn in'. For a small number of animals, PEVs estimated using this method were larger than the corresponding genetic variance.
- *Approximate*. The approximation method described above but including only the direct genetic and permanent environmental effects.

For the 'approximate' analysis of the RR data, covariances due to maternal direct and permanent environmental effects were added to the covariances due to the direct permanent environmental effects.

Prediction error variances of EBVs for 0-, 200-, 400- and 600-day weights from the RR analyses and for two indexes of EBVs from the multiple-trait analysis were obtained using (12).

Reliabilities

Reliabilities for PEVs were calculated using (11) for the 'inverse' and 'Harville' methods and (10) for the approximate method. Reliabilities of EBVs from the RR analyses were also approximated using the method of JAMROZIK et al. (2000), modified to accommodate small numbers of contemporaries (*Jamrozik method*).

Criteria

Methods were compared by regressing one on another. Harville's method was compared with the inverse method. The 'approximate' and Jamrozik methods were compared with the inverse, or with Harville's method when direct inversion was impossible.

Results

Results from the Hereford data set presented in Table 4 show that the Gibbs sampling procedure (Harville) provides a good estimate of the the 'true' reliabilities of the EBVs at individual ages. The average reliabilities over all animals are similar, and the regression of the reliabilities obtained from the direct inverse (inverse) on those given by Harville's procedure show a high degree of correspondence between the two methods. The slopes are 1, the intercepts are essentially 0 and almost all ($\geq 99\%$) of the variation is explained by the regression. Results for reliabilities of the first two RR coefficients are similar. Agreement between inverse methods for the two higher order coefficients is slightly lower, but this does not affect the agreement between the reliabilities of the EBVs as these linear functions are dominated by the intercept and linear coefficients.

Reliabilities are overestimated by the approximate method. This is shown by the mean values, which are higher for the approximate method than either of those obtained from the inverse or Harville methods, and by the regression of approximate on inverse reliabilities. The intercepts are 0 and the linear coefficients are < 1 , indicating that the higher the reliability the more it is overestimated. The regression models account for most ($\geq 97.5\%$) of the variation. The relationship between the inverse and approximate methods is illustrated in Fig. 1 which contains the scatter plot of reliabilities for EBVs for 200-day weight. With lower overall mean values and higher slopes, the approximate method performed marginally better than the method of Jamrozik at approximating the reliabilities of EBVs, although R^2 values were essentially the same. The approximate method overestimates the reliabilities of the higher order coefficients more than the lower order RR coefficients, but once again this has little effect on the reliabilities of the EBVs.

Table 5 shows close correspondence between the reliabilities produced by Harville's and approximate methods for both RR coefficients and EBVs in the Wokalup data. The

Table 4. Comparisons of reliabilities of estimated breeding values (EBVs) and random regression (RR) coefficients in the single Hereford herd

	EBVs (age in days)				RR coefficients (degree)			
	0	200	400	600	0	1	2	3
	<i>Mean reliabilities</i>							
Inverse	0.321	0.275	0.310	0.331	0.324	0.289	0.137	0.120
Harville	0.320	0.270	0.306	0.325	0.318	0.282	0.132	0.120
Approximate	0.360	0.312	0.348	0.368	0.363	0.329	0.179	0.158
Jamrozik	0.377	0.346	0.370	0.385				
	<i>Regression of inverse on Harville</i>							
$\beta_{I,H}$	1.0015	1.0078	1.0083	1.0086	1.0082	1.0053	0.9685	0.9451
Intercept	0.0010	0.0022	0.0023	0.0031	0.0035	0.0060	0.0093	0.0062
R^2 (%)	99.9	99.4	99.4	99.3	99.2	98.7	94.8	95.5
	<i>Regression of inverse on approximate</i>							
$\beta_{I,A}$	0.9042	0.8855	0.8945	0.8957	0.8940	0.8718	0.7644	0.7579
Intercept	-0.0041	-0.0029	-0.0013	0.0002	-0.0001	0.0024	0.0007	-0.0004
R^2 (%)	98.0	97.5	98.5	98.9	98.7	98.9	97.5	96.8
	<i>Regression of inverse on Jamrozik</i>				<i>Regression of Harville on approximate</i>			
$\beta_{I,J}$	0.8670	0.8031	0.8434	0.8576	0.8799	0.8564	0.7420	0.7616
Intercept	-0.0054	-0.0032	-0.0015	-0.0001	-0.0011	-0.0001	-0.0004	-0.0006
R^2 (%)	98.0	97.4	98.6	99.1	98.0	97.7	90.9	91.4

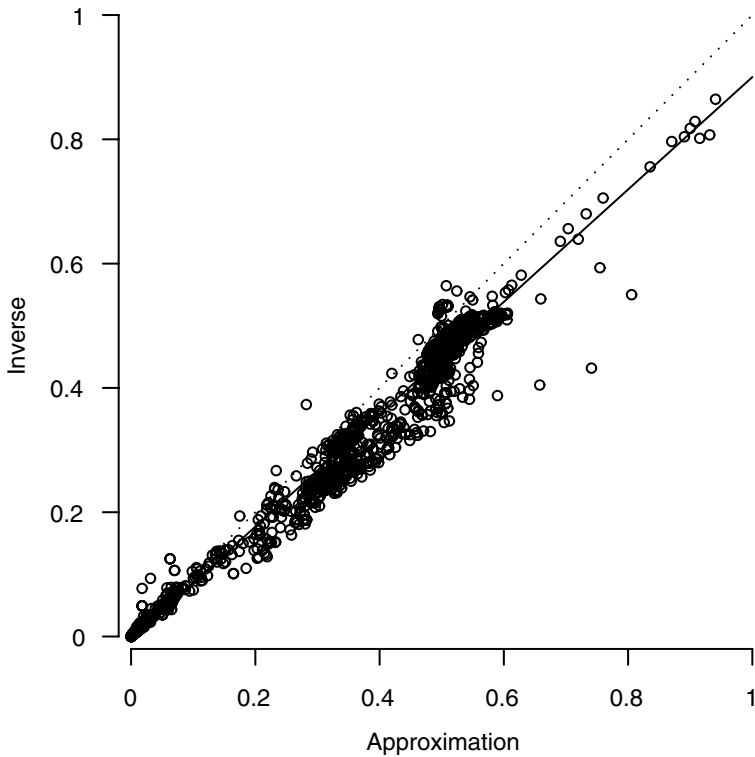


Fig. 1. Reliabilities for estimated breeding values for 200 days for a single Hereford herd from approximation and inverse of the coefficient matrix, together with linear regression of inverse on approximation

Table 5. Comparisons of reliabilities of random regressions (RR) coefficients and estimated breeding values for Wokalup and Murray Grey data sets. Reliabilities of EBVs and PEVs derived from Harville, Approximate and Jamrozik (EBV-only) methods

Order	Wokalup				Murray Grey			
	0	1	2	3	0	1	2	3
Random regression coefficients								
<i>Mean reliabilities</i>								
Harville	0.386	0.346	0.238	0.218	0.280	0.230	0.110	0.105
Approximate	0.378	0.339	0.244	0.228	0.278	0.222	0.097	0.095
<i>Regression of Harville on Approximate</i>								
$\beta_{H,A}$	0.999	0.982	0.954	0.949	0.956	0.950	0.939	0.906
Intercept	0.008	0.013	0.005	0.001	0.014	0.020	0.019	0.019
R^2 (%)	95.3	95.0	94.6	94.6	88.6	85.1	69.4	71.5
Estimated Breeding Values								
Age	0	200	400	600	0	200	400	600
<i>Mean reliabilities</i>								
Harville	0.415	0.348	0.367	0.381	0.304	0.262	0.275	0.279
Jamrozik	0.453	0.377	0.386	0.397	0.349	0.310	0.311	0.315
Approximate	0.436	0.342	0.360	0.372	0.316	0.265	0.274	0.275
<i>Regression of Harville on Jamrozik</i>								
$\beta_{H,J}$	0.900	0.912	0.952	0.962	0.894	0.885	0.925	0.923
intercept	0.007	0.005	0.001	-0.001	-0.009	-0.012	-0.013	-0.012
R^2 (%)	96.8	95.9	97.0	96.9	88.8	86.9	88.4	87.8
<i>Regression of Harville on approximate</i>								
$\beta_{H,A}$	0.936	1.017	1.004	0.997	0.911	0.955	0.954	0.954
intercept	0.007	0.000	0.006	0.009	0.016	0.009	0.014	0.016
R^2 (%)	96.0	94.6	95.9	95.8	90.2	88.5	89.5	89.0

Table 6. Average reliability of breeding value and index estimates from Harville and Approximate methods for multiple traits

Trait	1	2	3	4	5	Index 1	Index 2
<i>Mean reliabilities</i>							
Harville	0.315	0.187	0.171	0.191	0.179	0.302	0.303
Approximate	0.282	0.164	0.162	0.155	0.149	0.272	0.260
<i>Regression of Harville on approximate</i>							
$\beta_{H,A}$	0.957	0.951	0.951	0.960	0.978	0.970	0.980
Intercept	0.044	0.031	0.016	0.042	0.034	0.038	0.032
R^2	91.4	84.6	84.0	90.3	88.7	90.2	89.0

reliabilities overestimated for birth weight, but are well estimated for EBVs at later weights. Most of the variation is explained by the regression which has a slope of almost 1 for the later EBVs. For the Wokalup data, the approximate method provides a better estimate than that of Jamrozik. For the Murray Grey data set, the accuracy of the quadratic and cubic RR coefficients is low, because there are few animals with more than two records. As a consequence, the quality of the approximation is not as good for the higher order RR coefficients as shown by the lower R^2 values.

Results of the multiple-trait analysis are shown in Table 6. Considerably more records were available on the first trait than on other traits. The regressions of reliabilities obtained

from Harville's method on those from the approximate method show that when data are available higher reliabilities are overestimated and more variation is explained than when reliabilities are calculated primarily from correlated information. The approximate reliabilities of the indexes – functions of the PEVs and PECs – corresponded well to those obtained using Harville's method.

Discussion

A simple method for approximating blocks of the inverse of the coefficient matrix of the MME with an animal model has been developed. Approximate PEVs and PECs can be used to approximate the reliabilities of RR coefficients and linear functions thereof. Approximations correspond well to the 'true' values derived from the inverse of the coefficient matrix, despite amalgamating covariances due to both maternal effects with the direct permanent environmental effect in the RR models. The method is also well suited to approximate the reliabilities of indexes of EBVs in multiple-trait models.

When there are no covariances between direct and maternal effects, approximate reliabilities for these effects can be computed independently. To use the procedure described to approximate PECs of maternal genetic effects, the (co)variances for both genetic and permanent direct environmental effects would need to be added to the permanent maternal environmental (co)variances.

The method slightly overestimates the higher reliabilities (>0.5). This is not surprising because only first-order relationships amongst contemporaries are taken into account when valuing the data with respect to the individual and its parents, and animals must have many descendants to achieve a high accuracy. The most deviant points in Fig. 1 belong to animals with three generations of descendants which are all included in the same set of contemporaries. These great grand-progeny contribute nothing to this ancestor's EBVs, but in the approximation of reliability this is ignored. In reality, these animals are no longer available for selection and are consequently of limited interest.

The reliabilities of RR coefficients and EBVs and the quality of their approximation depend upon the quantity and quality of the available data. Poor estimation of the higher order coefficients and their reliabilities in the Murray Grey data is not surprising as few animals have more than two observations, whereas with many observations per animal in the Wokalup data set, the higher order coefficients are better estimated. Poorly estimated higher order RR coefficients have little effect on the reliabilities of the EBVs in the Hereford and Murray Grey data sets. This is due to the relatively small contribution of these coefficients to the EBVs, compared with the constant and linear terms.

By approximating the diagonal blocks of the inverse of the MME in this way, it is simple to compute the reliability of EBVs at any point on the longitudinal axis, unlike the method of JAMROZIK *et al.* (2000) which computes the value for each point independent of the others. It may be possible to infer these diagonal blocks for RR models using JAMROZIK *et al.* (2000) from the reliabilities of a number of different EBVs.

For large-scale analyses with a considerable number of traits (or RR coefficients), the $k \times (k + 1)/2$ PECs for each animal translate to large storage requirements. For highly correlated genetic effects, however, a substantial reduction in the number of variables which need to be considered can be achieved by reparameterizing to the most important principal components only, i.e. the linear functions of the original variables given by the eigenvectors of \mathbf{G}_0 pertaining to the largest eigenvalues. For the 22 direct genetic effects in a multivariate genetic evaluation of Australian beef cattle considered by MEYER *et al.* (2004), for instance, the first nine genetic principal components accounted for 99.1% of the genetic variation, i.e. applying this reparameterization, the number of PECs to be stored for each animal could be reduced from 253 (for $k = 22$) to 45 (for $k = 9$), with very little loss in genetic information.

Conclusions

Accuracies of genetic evaluation under RR and multiple-trait models can be approximated satisfactorily for the data structure typical for beef cattle. The approximation used is computationally undemanding and applicable to large-scale problems. The ability to approximate PECs is valuable not only in a RR situation, but will facilitate approximation of accuracies of genetic indexes composed of multi-trait EBVs.

Acknowledgements

This work was supported by Meat and Livestock Australia under grant BFGEN.100.

References

- BARWICK, S.; HENZELL, A., 1998: Breedobject: Breeding objective and indexing software for beef breeding. In: Proc. Sixth World Congr. Genet. Appl. Livest. Prod., vol. 27, Armidale, Australia, pp. 445–446.
- FOUILLOUX, M. N.; LALOË, D., 2001: A sampling method for estimating the accuracy of predicted breeding values in genetic evaluation. *Genet. Select. Evol.* **33**: 473–486.
- GENGLER, N.; MISZTAL, I., 1996: Approximation of reliability for multiple-trait animal models with missing data by canonical transformation. *J. Dairy Sci.* **79**: 317–328.
- GRASER, H.-U.; TIER, B., 1997: Applying the concept of number of effective progeny to approximate accuracies of prediction derived from multiple trait analyses. *Proc. Ass. Advan. Anim. Breed. Genet.* **12**: 547–551.
- HARRIS, B.; JOHNSON, D., 1998: Approximate reliability of genetic evaluations under an animal model. *J. Dairy Sci.* **81**: 2723–2728.
- HARVILLE, D. A., 1999: Use of the Gibbs sampler to invert large, possibly sparse, positive definite matrices. *Lin. Alg. Appl.* **289**: 203–224.
- JAMROZIK, J.; SCHAEFFER, L. R.; JANSEN, G. B., 2000: Approximate accuracies of prediction from random regression models. *Livest. Prod. Sci.* **66**: 85–92.
- JOHNSTON, D. J.; TIER, B.; GRASER, H.-U.; GIRARD, C., 1999: Presenting BREEDPLAN version 4.1. *Proc. Ass. Advan. Anim. Breed. Genet.* **13**: 193–196.
- KOOTS, K. R.; SCHAEFFER, L. R.; JANSEN, G. B., 1997: Approximate accuracy of genetic evaluation under an animal model. *J. Dairy Sci.* **80** (Suppl. 1): 226 (abstract).
- LIU, Z.; REINHARDT, F.; REENTS, R., 2001: The effective daughter contribution concept applied to multiple trait models for approximating reliability of estimated breeding values. *Interbull Bull.* **27**: 41–46.
- MEYER, K., 1989: Approximate accuracy of genetic evaluation under an individual animal model. *Livest. Prod. Sci.* **21**: 87–100.
- MEYER, K., 2002: Estimates of covariance functions for growth of Australian beef cattle from a large set of field data. CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod. Communication No. 11-01. Montpellier, France.
- MEYER, K., 2004: Scope for a random regression model in genetic evaluation of beef cattle for growth. *Livest. Prod. Sci.* **86**: 68–83.
- MEYER, K.; CARRICK, M. J.; DONNELLY, B. J. P., 1993: Genetic parameters for growth traits of Australian beef cattle from a multi-breed selection experiment. *J. Anim. Sci.* **71**: 2614–2622.
- MEYER, K.; JOHNSTON, D. J.; GRASER, H.-U., 2004: Estimates of the complete genetic covariance matrix for traits in multi-trait genetic evaluation of Australian Hereford cattle. *Aust. J. Agric. Res.* (in press).
- MISZTAL, I.; PEREZ-ENCISCO, M., 1993: Sparse matrix inversion for restricted maximum likelihood estimation of variance components by expectation–maximization. *J. Dairy Sci.* **76**: 1479–1483.
- MISZTAL, I.; WIGGANS, G. R., 1988: Approximation of prediction error variance in large-scale animal models. *J. Dairy Sci.* **71** (Suppl. 2): 27–32.
- THOMPSON, R.; WRAY, N. R.; CRUMP, R. E., 1994: Calculation of prediction error variances using sparse matrix methods. *J. Anim. Breed. Genet.* **111**: 102–109.
- TIER, B.; SCHNEEBERGER, M.; HAMMOND, K.; FUCHS, W. C., 1991: Determining the accuracy of estimated breeding values in multiple trait animal models. *Proc. Ass. Advan. Anim. Breed. Genet.* **9**: 239–241.

Author's address: BRUCE TIER (for correspondence), Animal Genetics and Breeding Unit, University of New England, Armidale, NSW 2351, Australia. E-mail: btier@une.edu.au