APPROXIMATE AND 'EXACT' ACCURACIES OF BREEDING VALUE ESTIMATES FOR GROWTH OF BEEF CATTLE FROM RANDOM REGRESSION ANALYSES

Karin Meyer and Bruce Tier

Animal Genetics and Breeding Unit[1], University of New England, Armidale, NSW 2351

SUMMARY

Approximate prediction error covariances among estimates of random regression coefficients for direct genetic effects were obtained for two beef cattle data sets using an extension of the method of Graser and Tier (1997). From these, approximate accuracies of breeding value estimates for birth, 200 day, 400 day and 600 day weight were calculated. Corresponding 'exact' values were determined from the inverse of the coefficient matrix in the mixed model equations (MME), estimating the diagonal blocks of the inverse pertaining to random regression coefficients using Gibbs sampling. Approximate and 'exact' values were contrasted with empirical accuracies, obtained as correlations between true and estimated genetic values in a simulation study. Results showed good agreement between approximate and 'exact' values.
Keywords : Random Regression, genetic evaluation, accuracy, beef cattle

INTRODUCTION

Calculation of accuracies of estimated breeding values (EBV) forms an integral part of genetic evaluation schemes. Whilst 'exact' values could be obtained from the inverse of the coefficient matrix in the mixed model equations, computational requirements for this are generally prohibitive and approximations are used in practice. This paper examines accuracies of EBVs for growth traits of beef cattle from a random regression (RR) model analysis, using a new approximation procedure which provides both sampling variances and covariances among estimated genetic effects. Approximate, 'exact' and empirical accuracies are contrasted for two small data sets.

MATERIAL AND METHODS

Data. Records for weights of beef cattle from birth to 730 days of age were considered. The first data set comprised weights for animals in two herds in the Wokalup selection experiment, which involved monthly weight recording, yielding up to 26 records per animal. The second data set consisted of weights in 16 Hereford herds, in which at least 50% of animals had four or more weights recorded. To investigate the effect of limited numbers of records on the approximation of accuracies, this data set was analysed as

Table 1. Data structure

Number of	Wokalup	Hereford
records	75,829	44,328
animals	7,305	12,839
parents	1,138	9,831
cont.groups	11,417	5,682

is and expanded to twice its size, by adding a ficticious record, assumed to be taken about 100 days later, for each record in the data; see Meyer (2003) for details. Characteristics of the data structure are summarised in Table 1.

Model. The model of analysis fitted a cubic, RR on Legendre polynomials (LP) of age at recording for direct genetic, maternal genetic, direct permanent environmental and maternal permanent environ-

Table 2. Average accuracy of estimates of genetic random regression coefficients[A] from simulation, the inverse of the coefficient matrix and from approximate method

Coeff.	Wokalup				Hereford				Hereford ×2			
	0	1	2	3	0	1	2	3	0	1	2	3
Simulation												
Mean	0.617	0.586	0.486	0.465	0.531	0.495	0.329	0.306	0.555	0.543	0.440	0.409
s.d.	0.018	0.021	0.028	0.028	0.019	0.021	0.037	0.035	0.018	0.019	0.026	0.026
Accuracy from mean prediction error variance												
Inverse	0.621	0.588	0.488	0.467	0.536	0.496	0.325	0.312	0.559	0.544	0.442	0.411
Approx.	0.615	0.582	0.494	0.478	0.550	0.513	0.359	0.341	0.571	0.558	0.466	0.440
α_1^B	0.999	0.982	0.954	0.949	0.948	0.931	0.826	0.851	0.959	0.947	0.902	0.885
R^2 (%)[D]	95.3	95.0	94.6	94.6	94.2	92.0	77.4	81.4	96.2	95.9	93.9	93.4
Accuracy for individuals												
Inverse	0.604	0.570	0.473	0.454	0.530	0.496	0.335	0.317	0.547	0.535	0.431	0.399
Approx.	0.599	0.564	0.478	0.463	0.531	0.497	0.347	0.329	0.548	0.536	0.445	0.418
β_1^C	1.010	0.990	0.942	0.939	0.847	0.814	0.697	0.733	0.881	0.866	0.842	0.842
R^2 (%)	95.3	94.4	93.4	93.9	91.9	89.0	73.7	77.7	94.7	94.0	91.5	91.5

[A] 0: intercept, 1: linear, 2: quadratic, 3: cubic; [B] regression of reliability from inverse on approximation, [C] as [B] but for accuracy, [D] coefficient of determination for β

mental effects. Variances among RR coefficients and heterogeneous measurement error variances were assumed to be given by estimates for Hereford cattle obtained by Meyer (2002). 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. For the Hereford data, CG were divided further by applying an "age slicing" of 45 days up to 300 days, and 60 days for higher ages.

Approximation. A full description of the approximation procedure used is beyond the scope of this paper, and will be published elsewhere. The algorithm is similar to other methods of approximation, in that the informative value of data arising from an individual's own observations is combined with that from close relatives. It differs from methods proposed, for instance, by Graser and Tier (1997) or Jamrozik *et al.* (2000) by not converting these values into "equivalent progeny numbers" on a single trait basis, but retaining them as multiple trait values, so that prediction error covariances can be derived in addition to variances.

Inverse. The diagonal blocks of the inverse of the coefficient matrix in the MME pertaining to sets of RR coefficients for direct genetic effects, were estimated for each data set using a Gibbs sampling algorithm as described by Harville (1999). A total of 400,000 samples were drawn, discarding the first 20,000 samples as burn in.

Simulation. Empirical accuracies for the data sets considered were available from a simulation study, which calculated accuracy as the correlation between true and estimated breeding value across all animals. Results given are means and standard deviations for 800 replicates (Meyer, 2003).

Criteria. Prediction error variances (PEV) of estimated RR coefficients were obtained as diagonal elements of the animal blocks in the inverse coefficient matrix, or their approximations. For inbred animals, PEVs derived from the inverse were scaled by $(1+F_j)^{-1}$ (Tier *et al.*, 1991), with F_j the inbreeding coefficient of animal j. The accuracy for a coefficient was calculated as $\rho_{ij} = \sqrt{1 - PEV_{ij}/Var_i}$, with Var_i the genetic variance for the $i-$th coefficient ($i = 0,1,2,3$ for intercept, linear, quadratic and cubic), PEV_{ij} the estimated PEV for the $i-$th coefficient and $j-$th animal, Under a RR model, EBVs for specific ages can be obtained by evaluating the estimated regression for the ages desired. Whilst this is a polynomial function of age, it is linear in the estimated RR coefficients. Hence, corresponding prediction error variances (PEV_{kj}) of EBVs at ages $k = 0,200,400$ and 600 days were obtained from the estimated covariances among RR coefficients, as the variance of a linear function, and accuracies calculated as $\rho_{kj} = \sqrt{1 - PEV_{kj}/Var_k}$, with Var_k the genetic variance for age k. A second approximate accuracy for EBVs was obtained using the procedure described by Jamrozik *et al.* (2000). For a small number of animals, PEVs estimated from the inverse coefficient matrix were larger than the corresponding genetic variance. Hence, the average of ρ_{ij} and ρ_{kj} was determined omitting those animals. In addition, a corresponding average accuracy was calculated as $\rho_m^* = \sqrt{1 - \overline{PEV_{m.}}/Var_m}$, with $m = i, k$ and $\overline{PEV_{m.}}$ the average PEV for the $m-$th coefficient or EBV across all animals. Regressions of accuracies from the inverse coefficient matrix on their approximations and their coefficients of determination (R^2) were calculated.

RESULTS AND DISCUSSION

Average accuracies of estimated RR coefficients are given in Table 2. Values ρ_i^* derived from $\overline{PEV_{i.}}$ for the direct inverse showed excellent agreement with empirical results. Average ρ_{ij} ($\overline{\rho_{i.}}$) from the inverse for the subset of animals with $PEV_{ij} \leq Var_i$ were consistently slightly less than empirical means, but deviations were well within the range of standard deviations. This implies that, in spite of yielding overestimates of PEV for some animals, the inverse derived by Gibbs sampling provided good estimates of 'exact' accuracies.

In turn, ρ_i^* and $\overline{\rho_{i.}}$ derived from the approximated prediction error covariances agreed closely with their counterparts from the inverse coefficient matrix. R^2s for the regression of 'exact' on approximate accuracies were up to 96%. With regression coefficients close to unity, agreement was best for the Wokalup data set with a high number of records per animal. Conversely, discrepancies were largest and R^2s lowest for quadratic and cubic coefficients for the Hereford data set in which 45% of animals had less than 4 records, the number required to fit a cubic polynomial accurately. Expanding the data set to twice its size reduced this proportion to 7.5% and improved the quality of approximation of accuracies.

Corresponding results for accuracies of EBVs for 0, 200, 400 and 600 days of age, summarised in Table 3, exhibited a very similar pattern. Excellent agreement between values from the inverse coefficient matrix and our approximation implies that not only PEV of RR coefficients but also prediction error covariances among them have been approximated correctly. Approximate values derived using Jamrozik *et al.* (2000)'s procedure tended to be slightly higher than accuracies from our method of approximation, but had similar R^2s.

CONCLUSIONS

Accuracies of genetic evaluation under a RR model can be approximated satisfactorily for the data structure typical for beef cattle. The approximation used is computationally undemanding and appli-

Table 3. Average accuracy of breeding value estimates at selected ages (in days) from simulation, the inverse of the coefficient matrix, from approximate method and Jamrozik *et al.* (2000)'s procedure

Age	Wokalup				Hereford				Hereford $\times 2$			
	0	200	400	600	0	200	400	600	0	200	400	600
Simulation												
Mean	0.639	0.583	0.600	0.613	0.531	0.493	0.522	0.535	0.533	0.506	0.538	0.558
s.d.	0.017	0.020	0.018	0.018	0.020	0.023	0.020	0.019	0.020	0.022	0.019	0.018
Accuracy from mean prediction error variance												
Inverse	0.644	0.590	0.606	0.617	0.531	0.499	0.526	0.539	0.535	0.512	0.542	0.561
Approx.	0.660	0.585	0.600	0.610	0.550	0.516	0.541	0.553	0.553	0.526	0.554	0.573
α_1^A	0.936	1.017	1.004	0.997	0.938	0.940	0.945	0.948	0.941	0.953	0.960	0.960
R^2 (%)^E	96.0	94.6	95.9	95.8	97.0	94.3	94.9	94.7	97.3	95.7	96.2	96.4
Jamrozik	0.673	0.614	0.621	0.630	0.566	0.544	0.560	0.569	0.570	0.553	0.571	0.586
α_2^B	0.900	0.912	0.952	0.962	0.897	0.844	0.884	0.899	0.899	0.860	0.903	0.918
R^2 (%)	96.8	95.9	97.0	96.9	97.1	94.1	95.0	94.7	97.3	95.6	96.3	96.5
Accuracy for individuals												
Inverse	0.632	0.575	0.589	0.600	0.516	0.489	0.518	0.533	0.517	0.496	0.526	0.549
Approx.	0.648	0.571	0.584	0.594	0.531	0.496	0.521	0.534	0.533	0.502	0.528	0.549
β_1^C	0.967	1.023	1.009	1.005	0.935	0.865	0.861	0.849	0.946	0.895	0.893	0.883
R^2 (%)	96.6	95.4	96.1	95.6	96.7	92.8	93.2	92.5	97.2	94.8	95.2	94.9
Jamrozik	0.660	0.597	0.603	0.612	0.547	0.522	0.538	0.549	0.548	0.527	0.544	0.561
β_2^D	0.921	0.935	0.958	0.964	0.906	0.816	0.830	0.824	0.918	0.847	0.863	0.861
R^2 (%)	97.2	96.3	96.8	96.4	96.5	92.8	93.3	92.5	97.0	94.8	95.3	95.0

^{A} regression of reliability from inverse on approximation, ^{B} regression of reliability from inverse on Jamrozik's approximation, ^{C} as ^{A} but for accuracy, ^{D} as ^{B} but for accuracy, ^{E} coefficient of determination

cable to large scale problems. The ability to approximate prediction error covariances is valuable not only in a RR situation, but will facilitate approximation of accuracies of genetic indexes composed of multi-trait EBVs.

REFERENCES

Graser, H.-U. and Tier, B. (1997) *Proc. Ass. Advan. Anim. Breed. Genet.* **12**:547–551.

Harville, D. A. (1999) *Lin. Alg. Appl.* **289**:203–224.

Jamrozik, J., Schaeffer, L. and Jansen, G. B. (2000) *Livest. Prod. Sci.* **66**:85–92.

Meyer, K. (2002) *CD-ROM 7th World Congr. Genet. Appl. Livest. Prod.* Comm. No. 11–01.

Meyer, K. (2003) *Livest. Prod. Sci.* **00**:000–000. (submitted).

Tier, B., Schneeberger, M., Hammond, K. and Fuchs, W. C. (1991) *Proc. Ass. Advan. Anim. Breed. Genet.* **9**:239–241.

ACKNOWLEDGEMENTS