

Calculation of Accuracies of EBV using Sampling in a Large Multiple Breed Data Set

J.M. Hickey^{1,2,3}, H.A. Mulder³, D.I. Wilhelmus³, I. Strandén⁴ and R.F. Veerkamp³

¹Grange Beef Research Centre, Teagasc, Dunsany, Co. Meath, Ireland

²School of Agriculture, Food and Veterinary Medicine, University College Dublin, Ireland

³Animal Sciences Group, PO Box 65, 8200 AB, Lelystad, the Netherlands

⁴AgriFood Research Finland MTT, Animal Production Research, Jokioinen, Finland

Abstract

A tool was developed to calculate true across breed accuracies of breeding values using sampling. Both the accuracy of breed effects and the accuracy of the within breed breeding values were important in determining across breed accuracies. Within breed accuracies estimated using sampling were compared to those approximated using the Tier and Meyer (2004) method as implemented in MiX99 (Strandén *et al.*, 2001). The Tier and Meyer method overestimated high and low accuracies and underestimated those of intermediate value, however these errors were small.

Introduction

In multiple breed genetic evaluations across breed accuracy (r_{ab}) of EBVs primarily depend on the accuracy with which breed effects (r_b) and individual genetic deviations within a breed (r_{wb}) are estimated, with sampling covariances among these effects being of minor importance (Van Vleck *et al.*, 1992).

Calculation of true r_{ab} and r_{wb} using the exact method is infeasible due to the size of most national genetic evaluation data sets. Methods to approximate r_{wb} generally perform well but may give biased estimates for animals with certain data structures (e.g. Tier and Meyer, 2004). Unbiased approximations of r_{wb} can be calculated using sampling (Garcia-Cortes *et al.*, 1995; Fouilloux and Laloe, 2001).

Currently no method exists to approximate r_{ab} . Firstly this work aimed to extend the sampling method of Fouilloux and Laloe (2001) to a multiple breed, multiple trait scenario, and secondly to compare the r_{wb} calculated using sampling with those approximated by the Tier and Meyer (2004) method as implemented in the MiX99 (Strandén *et al.*, 2001).

Materials and methods

The sampling method of Fouilloux and Laloe (2001) to calculate true r_{wb} was extended to a multiple breed, multiple trait scenario.

Simulation of true breed group values

All animals in the pedigree trace to a founder breed group. For each breed group g represented in the pedigree, a vector $\mathbf{b}_g$ of breed group values, one for each trait, is simulated from normal distributions which are defined relevant to the population.

Simulation of true breeding values

For each animal i in the pedigree a vector $\mathbf{t}_i$ of true breeding values for each of n traits is simulated. The vector $\mathbf{t}_i$ is the sum of $\mathbf{b}_i$, a vector of true breed group values, and $\mathbf{u}_i$, a vector of true additive genetic values, both of which depend upon the status of i 's parents j and k .

If both j and k are unknown, each element in $\mathbf{b}_g$ of animal i is given the average of its

founder breed groups and each element of $\mathbf{u}_i$ is simulated as $\mathbf{L}_G \mathbf{z}$. $\mathbf{L}_G$ is obtained by Cholesky decomposition of $\mathbf{V}_G$, the genetic covariance matrix of the traits and $\mathbf{z}$ is a multivariate random-sampled vector with a mean of zero and a covariance matrix $\mathbf{I}$.

If one parent, say j is known, then $\mathbf{b}_i$ is given the average of the breed group value of the known parent, $\mathbf{b}_j$, and the founder breed group value for the missing parent, $\mathbf{b}_g$, while $\mathbf{u}_i$ equals $0.5(\mathbf{u}_j) + \sqrt{0.75}(\mathbf{L}_G \mathbf{z})$. If both parents j and k are known, then $\mathbf{b}_i$ is taken to be the average $\mathbf{b}_j$ and $\mathbf{b}_k$ while $\mathbf{u}_i$ equals $0.5(\mathbf{u}_j) + 0.5(\mathbf{u}_k) + \sqrt{0.5}(\mathbf{L}_G \mathbf{z})$. This results in a matrix of true breeding values with a distribution $N(\mathbf{Qm}, \mathbf{A} \otimes \mathbf{V}_G)$, where $\mathbf{Q}$ is an incidence matrix relating animals to $\mathbf{m}$, the means of the founder breed groups of which they are comprised, and $\mathbf{A}$ is the relationship matrix between all animals in the pedigree.

Simulation of true phenotypic values

A vector $\mathbf{y}_i$, of phenotypic values for each trait, is generated for each animal i as $\mathbf{y}_i = \mathbf{t}_i + \mathbf{e}_i$ where $\mathbf{e}_i = \mathbf{L}_E \mathbf{z}$ is a vector of random residual values for each trait, where $\mathbf{L}_E$ is obtained by Cholesky decomposition of $\mathbf{V}_E$, the residual covariance matrix for the traits. Values of fixed effects do not affect the distribution of random variables (Garcia-Cortes *et al.*, 1995) and are simulated with values of zero.

Simulation of estimated breed group values and estimated breeding values

By solving mixed model equations, set up using breed groups in the relationship matrix matching those simulated and fixed effects defined the same as those used to define the data structure, $\hat{\mathbf{B}}$ and $\hat{\mathbf{U}}$ matrices of estimated breed group values of estimated breeding values can then be obtained.

Sampling process and calculation of true accuracies

The whole process is repeated several times and r_b is calculated as the correlation between the true and estimated breed effects, r_{wb} is calculated as the correlation between true and estimated breeding value within breed, and r_{ab} is calculated as the correlation between true and estimated breeding value across breed, across all the replicates. As the number of replications increases estimates of accuracy converge to their true values. Calculations of standard errors of correlations determined 350 replicates to be sufficient (results not shown). Within breed breeding values are calculated by subtracting the breed effects from the across breed breeding values.

Application to data

This method was applied to the Irish multiple breed beef cattle data set used for the January 2007 routine genetic evaluation. Fifteen traits related to beef production were evaluated using data on purebred and crossbred animals of thirty five breeds, of which eight dominated. Most of the 493,092 animals with records on at least one trait only had information on subsets of traits (e.g. carcass conformation (**conf**) (304,589 records), weaning weight (**weight**) (52,161 records), and feed intake (**intake**) (2,491 records)). Different breeds also tended to have records on particular subsets. Most of the information came from crossbreds for some breeds and traits, for others it came from purebreds. Fifteen breed groups were defined, one for each of the fourteen most numerous breeds and one for the remaining breeds. Breed group values were simulated with a mean of zero and variance equal to the variance of the breed group solutions from the January 2007 routine evaluation. The $\mathbf{V}_G$ and $\mathbf{V}_E$ from the same evaluation were also used, therefore assuming that they were homogeneous across all breeds. The phenotypes were given the same data structure as that of the January 2007 evaluation.

Mixed model equations were solved using PEST (Groeneveld, 1990). The r_{wb} and r_{ab} were calculated for the AI sires in the Irish Cattle Breeding Federation database of the seven most numerous breeds in the data set only.

Table 1. Breed variance and genetic variance used to generate true breed group values and true breeding values. Accuracy of breed group effect (r_b), average within breed (r_{wb}), and average across breed (r_{ab}) accuracy for AI sires for three traits and three breeds

	Carcass conformation			Weaning weight			Feed Intake		
Breed Variance	1.30 units ²			59.47 kg ²			0.07 units ²		
Genetic variance	0.97 units ²			1040.00 kg ²			0.33 units ²		
Breed	r_b	r_{wb}	r_{ab}	r_b	r_{wb}	r_{ab}	r_b	r_{wb}	r_{ab}
Aberdeen Angus	0.99	0.69	0.88	0.50	0.60	0.57	0.16	0.57	0.35
Holstein	0.99	0.83	0.93	0.26	0.62	0.57	0.02	0.65	0.46
Limousin	0.99	0.77	0.90	0.57	0.79	0.75	0.46	0.71	0.61

Comparison to Tier and Meyer

The r_{wb} for AI sires calculated using sampling (r_{wbSA}) were compared to within breed accuracies approximated by the Tier and Meyer (2004) method as implemented in MiX99 (r_{wbTM}) by regressing the r_{wbSA} on the r_{wbTM} . MiX99 cannot be used to approximate r_{ab} .

Results

Sampling method

For illustrative purposes results (Table 1) are presented for only three breed groups, Aberdeen Angus, Holstein, and Limousin, and for three traits representative of the different data structures, conf, weight, and feed intake.

Large differences in r_b were observed for different breeds and traits. Breed group effects were well estimated for some traits (e.g. conf in all breeds (0.99)) moderately estimated for others (e.g. weight in Aberdeen Angus (0.50)) and poorly estimated for some traits (e.g. intake in Holstein (0.02)). Average r_{wb} tended to be higher when the corresponding r_b was well estimated (e.g. r_{wb} for conf in Holstein (0.83)). However, when the breed effects were poorly estimated it was still possible to have relatively high average r_{wb} . For example intake had an r_{wb} accuracy of 0.65 in Holsteins which is comparable to the value of 0.71 for Limousin despite the large difference in their respective r_b (0.02 versus 0.46).

Average r_{ab} depended on both the r_b effect and the average r_{wb} . Where r_b was greater than

that of the average r_{wb} the average r_{ab} was greater than r_{wb} . The effect was the opposite where r_b was lower than r_{wb} . For example r_{ab} for Holstein is higher than r_{wb} for conf (0.93 versus 0.83) but lower for intake (0.46 versus 0.65). Ratios of breed variance and genetic variance and their sampling covariances affected r_{ab} .

Comparison to Tier and Meyer

For all fifteen traits significant quadratic (e.g. Table 2) and in some cases higher order terms existed for the regressions of r_{wbSA} on r_{wbTM} . Where r_{wbSA} was high (>0.90) or low (circa <0.30) r_{wbTM} were overestimated. Intermediate values were slightly underestimated. The values for the regression coefficients agreed with the average errors within certain ranges of r_{wb} (Table 3). However the magnitude of the errors was small, even for traits with few records. The quality of the approximation was not as good for traits with lower numbers of records, as shown by the lower values of R^2 .

Table 2. Coefficients for quadratic regression (Int = intercept, β = slope, R^2 = R-squared) of within breed accuracies for carcass conformation (CC), weaning weight (WW) and feed intake (FI) for all AI sires in the data set calculated using sampling on within breed accuracies approximated using Tier and Meyer.

	CC	WW	FI
Int	-0.03 ¹	-0.02 ¹	-0.04 ¹
B	1.13 ¹	1.14 ¹	1.22 ¹
β^2	-0.12 ¹	-0.15 ¹	-0.24 ¹
R^2	0.98	0.96	0.94

¹p < 0.001

Table 3. Average error (Tier and Meyer – True) in accuracy approximated by the Tier and Meyer method for different ranges of true accuracy for carcass conformation (CC), weaning weight (WW) and feed intake (FI).

Accuracy range	CC	WW	FI
0.99 - 0.90	0.01	0.01	0.02
0.79 - 0.70	-0.01	0.00	0.00
0.59 - 0.50	0.00	-0.02	-0.01
0.39 - 0.30	0.02	0.00	0.00
0.19 - 0.10	0.03	0.02	0.04

Discussion

Sampling method

Accuracies for multiple breed EBVs were estimated using sampling. The r_{ab} appeared to be weighted averages of r_b and r_{wb} with the weighting depending on the ratio of their variances and the sampling covariance between them (Van Vleck *et al.*, 1992). Where there is poor partitioning of a total breeding value into its breed and individual genetic components these covariances may be important. The extent to this affected our results needs to be quantified.

Information on correlated traits was important in determining levels of r_{wb} . While the breeds and traits which had most phenotypic records had the highest average r_{wb} , the breeds and traits with vastly less records did not have vastly lower average r_{wb} . However as breed effects are modeled by fixed breed groups r_b does not benefit from information on correlated traits. Modeling breed groups as random effects could be considered.

In a multiple breed breeding program r_b , r_{wb} and r_{ab} influence the response to selection. While r_{wb} animals may be acceptable (e.g. feed intake in Holstein) and genetic gain can be made within a breed, the r_b and consequently r_{ab} may be low and efficiency of across breed selection would be reduced.

Comparison to Tier and Meyer

The Tier and Meyer method as implemented in MiX99 accounts only for one fixed effect,

in this case contemporary group, yet it provides good approximations of r_{wb} with only minor bias being observed. Lower R^2 for traits with lower numbers of records may be partially due to increasing standard errors of r_{wbSA} with reducing r_{wb} .

Conclusions

Within and across breed accuracy of EBVs can be calculated using sampling. Further work is required to quantify the effect that the ratio of breed group variance to genetic variance on the relevance of r_b and r_{wb} to r_{ab} . The Tier and Meyer method as implemented in MiX99 provides good approximations of within breed accuracy.

Acknowledgements

The Irish Cattle Breeding Federation are acknowledged for providing the data.

References

- Fouilloux, M.N. & Laloe, D. 2001. A sampling method for estimating the accuracy of predicted breeding values in genetic evaluation. *Gen. Sel. Evol.* 33, 473-486.
- Garcia-Cortes, L.A., Moreno, C., Varona, L. & Altarriba, J. 1995. Estimation of prediction-error variances by resampling. *J. Anim. Breed. Genet.* 112, 176-182.
- Groeneveld, E. 1990. *PEST User's Manual*. Urbana, Illinois, University of Illinois.
- Strandén, I., Lidauer, M., Mäntysaari, E.A. & Pösö, J. 2001. Calculation of Interbull weighting factors for the Finnish test day model. *Interbull Bulletin* 26, 78-79.
- Tier, B. & Meyer, K. 2004. Approximating prediction error covariances among additive genetic effects within animals in multiple-trait and random regression models. *J. Anim. Breed. Genet.* 121, 77-89.
- Van Vleck, L.D., Hakim, A.F., Cundiff, L.V., Koch, R.M., Crouse, J.D. & Boldman, K.G. 1992. Estimated breeding values for meat characteristics of crossbred cattle with an animal model. *J. Anim. Sci.* 70, 363-371.

