Proc. Assoc. Advmt. Anim. Breed. Genet. 22:557-560

WEIGHTING OF GENOMIC AND PEDIGREE RELATIONSHIPS IN SINGLE STEP EVALUATION OF CARCASS TRAITS IN AUSTRALIAN SHEEP

A.J. McMillan[1] and A.A. Swan[1]

[1]Animal Genetics Breeding Unit*, University of New England, New South Wales, Australia

SUMMARY

An issue for implementation of single step genomic evaluations is how to weight genomic and pedigree relationships in modelling genetic co-variance. A weighting parameter lambda ranging between 0 and 1 can be used in the statistical model, with higher values corresponding to greater weighting of genomic information. We investigated appropriate values of lambda for a range of carcass traits in terminal sire sheep breeds, using the accuracy and bias of genomic prediction of breeding values as criteria. The accuracy generally increased with lambda, although the "optimal" value of lambda at the maximum accuracy varied widely, covering almost the entire range of possible values across traits. Accuracy typically approached an asymptote towards the optimal lambda, so a wide range of values could be used with minimal loss of prediction accuracy. The bias in Estimated Breeding Values (EBVs) increased with lambda, such that EBVs over-predicted phenotypic performance at high values of lambda.

INTRODUCTION

Evaluations utilising genomic information in the form of blended EBVs have been available to Australian sheep breeders since 2011 (Swan *et al.* 2012). However utilising all available information on animals including phenotypes, genotypes and pedigree information in routine Australian sheep analysis were desired but had not previously been accomplished. In 2016 large scale multi-trait single step analyses (Legarra *et al.* 2014) were implemented for carcass and live weight traits in the three major breed evaluations, Terminal sires, Maternal breeds, and Merinos. These analyses include 17 traits, with pedigrees in excess of 2 million animals, and SNP genotypes for up to 15 thousand animals.

One of the issues for the implementation of single step in routine evaluations is how to optimally combine genomic and pedigree information for genotyped animals, since it is often argued that SNP genotypes do not explain all of the genetic variation (Goddard *et al.* 2011). To accommodate this, the variance of breeding values for genotyped animals can be modelled as $(\lambda G + (1-\lambda)A_{22})\sigma_u^2$, where G is the genomic relationship matrix calculated from SNP genotypes, A_{22} is the pedigree relationship matrix between genotyped animals, σ_u^2 is the genetic variance, and λ (lambda) is a weighting factor between 0 and 1. This variance matrix can be used in single step analyses, and often a high value of lambda, between 0.95 and 0.99, is used for the pragmatic reason that the resulting modified genomic relationship matrix can be reliably inverted. However, the broader questions remain, what is an appropriate value for lambda, and does lambda vary between traits? In this paper we use cross-validation to investigate the accuracy of genomic predictions across a range of lambda values for a range of carcass traits important for the terminal sire single step evaluation.

* AGBU is a joint venture of NSW Department of Primary Industries and University of New England

MATERIALS AND METHODS

A subset of animals from the Australian Terminal sire sheep evaluation were chosen, derived from the genomic reference population first established by the Sheep CRC (Van der Werf 2010). These animals have both genotypes and phenotypes for the traits studied. Key traits from the single step carcass analysis for terminal sires were investigated, including post-weaning weight (pwt), post-weaning eye muscle depth from live animal scanning (pemd), post-weaning fat depth from live animal scanning (pfat), hot carcass weight (hcwt), carcass eye Muscle Depth (cemd) carcass C-site fat depth (ccfat), lean meat yield (lmy), intra-muscular fat (imf), and shear force at day 5 (sf5). A summary of the animals recorded per trait and total animals in the pedigree is shown in Table 1.

Table 1: Data summary for terminal sire cross-validation analyses, with size of pedigree, number of animals recorded (and genotyped), number of Poll Dorset/White Suffolk animals with records (PD/WS rec), number of sires (PD/WS sires), and number of cross-validation sets (PD/WS ncv).

Trait	Pedigree	Records	PD rec	PD sires	PD ncv	WS rec	WS sires	WS ncv
pwt	28826	7714	3764	247	12	2567	169	8
pemd	28825	7713	3764	247	12	2566	169	8
pfat	28820	7712	3763	247	12	2567	169	8
hcwt	31774	8976	4298	248	14	2981	170	9
cemd	31345	8720	4172	248	13	2896	170	9
ccfat	31191	8630	4132	248	13	2868	170	9
lmy	22752	5254	2416	85	8	1658	56	5
imf	29952	8088	3905	215	13	2770	154	9
sf5	30764	8374	4017	248	13	2814	170	9

For each of these traits, the procedure involved estimating SS-GBLUP (Single Step Genomic BLUP) REML variance components using the Wombat software package (Meyer 2007) for values of lambda ranging between 0 and 1 in increments of 0.1. Animals with phenotypes were then allocated to cross-validation groups of approximately 300, stratified within two breeds, Poll Dorset (PD) and White Suffolk (WS). Animals were allocated to breeds based on the breed content of their sires. In addition, progeny from the same sire family were always allocated to the same cross-validation group, such that no animal in a cross-validation set would have half-sibs in the training data. Within these strata animals were allocated to groups at random, and the same groupings were used for all values of lambda. Summaries of the cross-validation schemes are shown in Table 1.

SS-GBLUP analyses were carried out for each cross-validation set across the range of lambda values specified above, using the 's1step' option in Wombat. Phenotypes for animals in the cross-validation set were omitted from the training data, but their pedigree and genotype data were included in the analysis in order to obtain their EBVs. Prediction accuracy was then calculated as the correlation between these EBVs and their phenotypes (adjusted for fixed effects). To approximate the correlation between True Breeding Value and EBV, these correlations were then scaled by the square root of the heritability of the trait, which was assumed to be the heritability estimated in the absence of genomic information. EBV bias was also calculated for each cross-validation set as the slope of the regression of phenotype on EBV (the expected value of the slope is 1, and if the estimate is less than 1 then EBVs over-predict phenotypic performance). Prediction accuracies and bias was then averaged across the cross-validation sets.

RESULTS AND DISCUSSION

Table 2 shows for each trait the estimated heritability (lambda = 0), the maximum cross-validation accuracy for breeds (r_{max}), the value of lambda where the maximum cross-validation

accuracy was observed, and the range in lambda values where the accuracy varied by ±0.01. Figure 1 displays accuracy across the range of lambda values by sire breed.

Table 2: Terminal sire lambda cross-validation summary, with estimated heritability (h^2), maximum cross-validation accuracy (r_{max}), $\lambda_{max} = \lambda$ at r_{max}, and range in lambda where accuracy varied by ±0.01 around r_{max} (λ_{low} to λ_{high}).

Trait	h^2	$r_{max(PD)}$	$r_{max(WS)}$	$\lambda_{max(PD)}$	$\lambda_{max(WS)}$	$\lambda_{low(PD)}$	$\lambda_{low(WS)}$	$\lambda_{high(PD)}$	$\lambda_{high(WS)}$
pwt	0.29	0.32	0.27	0.50	0.20	0.20	0.10	0.95	0.60
pemd	0.35	0.37	0.28	0.80	0.95	0.40	0.70	0.95	0.95
pfat	0.24	0.28	0.29	0.95	0.95	0.70	0.70	0.95	0.95
hcwt	0.14	0.41	0.27	0.20	0.95	0.20	0.80	0.50	0.95
cemd	0.21	0.34	0.21	0.60	0.90	0.40	0.50	0.95	0.95
ccfat	0.28	0.21	0.22	0.60	0.60	0.40	0.40	0.95	0.95
lmy	0.49	0.26	0.39	0.80	0.60	0.60	0.40	0.95	0.80
imf	0.60	0.35	0.28	0.80	0.60	0.50	0.40	0.95	0.95
sf5	0.37	0.28	0.24	0.60	0.50	0.30	0.30	0.95	0.95

Lambda values at maximum accuracy were 0.5 or greater, except pwt (WS) and hcwt (PD). As the maximum accuracy was approached, the response surface was generally asymptotic (see Figure 1), such that the range encompassing r_{max} ± 0.01 was large. Therefore, accuracy was relatively insensitive over a large range of lambda values especially beyond 0.5.

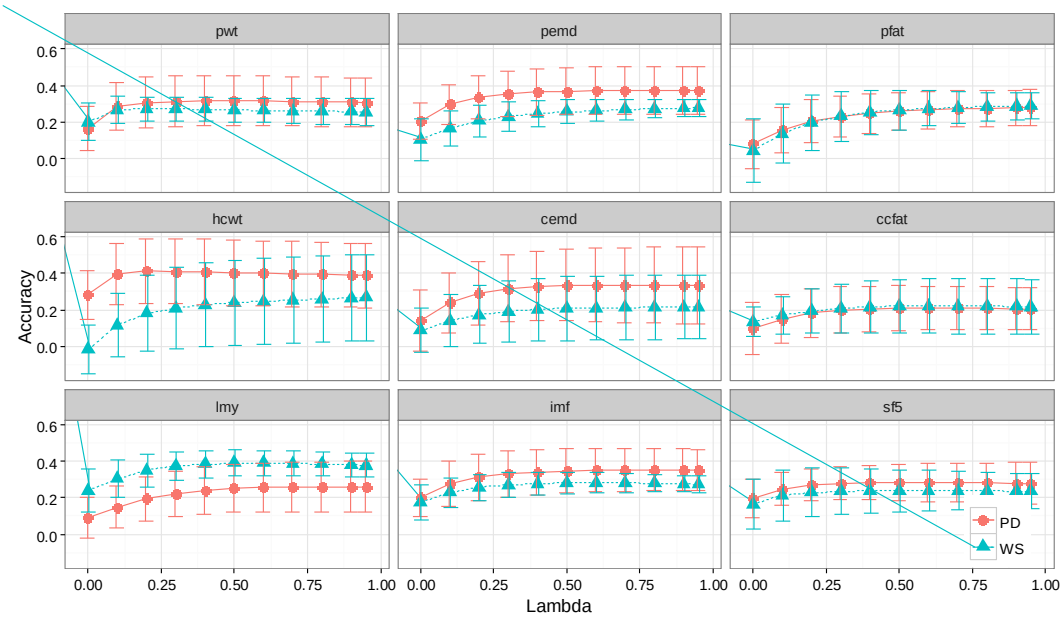

Figure 1: Accuracy versus lambda by sire breed in terminal sires (PD=Poll Dorset, WS = White Suffolk). Error bars show ± 1 standard deviation.

The slope of the regression of phenotype on EBV was used to assess the bias of EBVs across the range of lambda values and is shown in Figure 2. Results show some variation between traits and sire breeds within traits, but there is a clear general trend that the bias increases with lambda. That is, higher values of lambda lead to EBVs which over-predict phenotypic performance. In selection

cohorts with a mix of genotyped and un-genotyped contemporaries, this may lead to genotyped animals being incorrectly favoured. It is uncertain to us why the bias increases with lambda, but it may be due to an increasing influence of small genomic relationships in the G matrix which are due to identity by state rather than identity by descent genome sharing.

Correlations between EBVs for different lambda values were also calculated for different classes of animals, including progeny tested sires, and animals with and without phenotypes. For EBVs calculated with lambda of 0.5 and 0.95, these correlations ranged between 0.96 and 0.99, demonstrating that a wide range of lambda values between 0.5 and approaching 1 can be used with minimal impact on the ranking of animals.

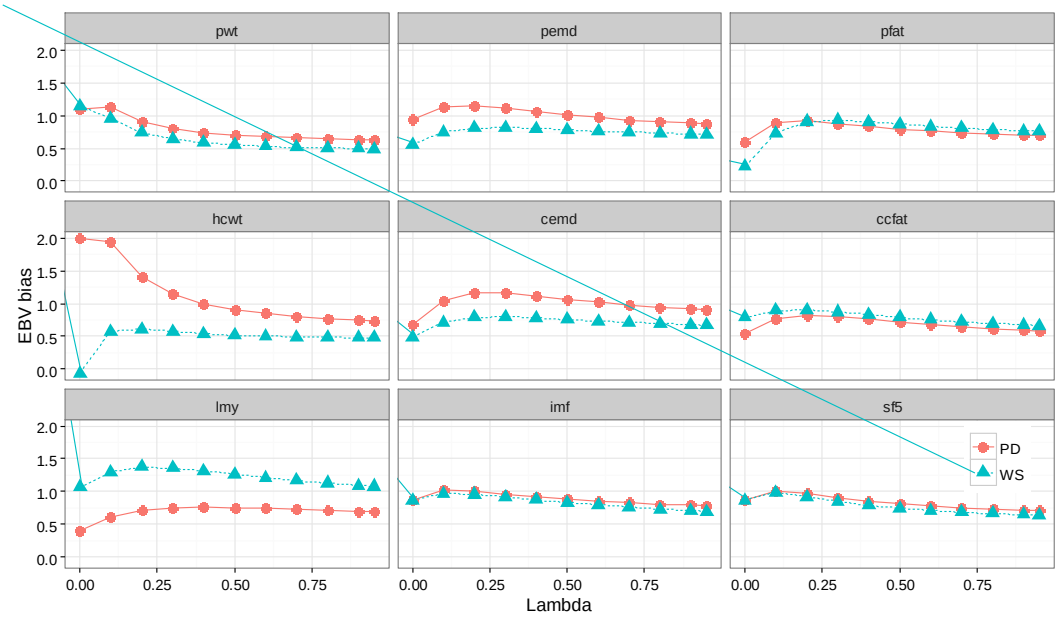

Figure 2: EBV bias versus lambda by sire breed in terminal sires (PD = Poll Dorset, WS = White Suffolk).

Given the relatively large window for insensitivity of prediction accuracy, high correlation of EBVs between lambda 0.5 and 0.95 and the levels of bias in EBVs when lambda is high we have initially used a value of 0.5 for lambda in routine industry evaluations. More research on this issue is warranted, including the impact of lambda in multi-trait models.

ACKNOWLEDGEMENTS

This research was funded by Meat and Livestock Australia. The authors acknowledge the contributions of the Sheep CRC Information Nucleus and industry-funded research flocks.

REFERENCES

Meyer K. (2007) *Journal of Zhejiang University. Science B* **8**: 815.
Swan, A.A., Johnston, D.J., Brown, D.J., Tier, B. and Graser, H-U. (2012) *Ani. Prod. Sci.* **52**: 126.
Legarra A., Christensen O.F., Aguilar I., and Misztal I. (2014) *Lives. Sci.* **166**: 54.
Goddard M.E., Hayes B.J. and Meuwissen T.H.E. (2011) *Jour. Ani. Breed. Genet.* **128**: 409.
Van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010). *Ani. Prod. Sci.* **50**: 998.