Contributed paper

DETERMINATION OF OPTIMUM WEIGHTING FACTORS FOR SINGLE-STEP GENETIC EVALUATION VIA GENETIC VARIANCE PARTITIONING

J.A. Torres-Vázquez, A.M. Samaraweera, M G. Jeyaruban, D.J. Johnston, and V. Boerner

Animal Genetics Breeding Unit*, University of New England, Armidale, NSW, 2351 Australia

SUMMARY

It is important in single-step genetic evaluations to use appropriate lambdas (λ) for calculating weighted average of NRM (numerator relationship matrix) and GRM (genomic relationship matrix) in joint relationship matrix. λ is usually estimated using a single-trait cross-validation procedure. However, it can be shown that a univariate single-step model applying a scalar λ is simply a condensed form of an extended model containing two genetic factors, factor $H \sim N(0, H)$ and factor $A \sim N(0, A)$, where the partitioning of the total genetic variance reflects λ. For multivariate single-step genetic evaluation, this model condensation implies that all involved genetic variances may yield the same λ, which is highly unlikely. Hence, it is required to estimate λ by accounting for its heterogeneity using the extended model for variance component estimation. This study used an extended single-step model to estimate variances and λs for calving difficulty (CD), gestation length (GL), and birth weight (BW) using Australian Angus data. A total of 129,851 animals with 45,575 genotypes were analysed. Initial variances obtained from a pedigree-only model were then used as starting values for the extended single-step model assigning 90% of the genetic variance to factor A and 10% to factor H. Since CD is a categorical trait with three categories, a threshold model-Gibbs sampling method was used to estimate variances. Heritability estimates for the extended single-step model were very similar to those from the pedigree only model implying that the single-step model was not explaining more variation in the data than the pedigree only model. For CD, GL, and BW, the total heritability estimates were 0.39 ± 0.04, 0.68 ± 0.02, and 0.44 ± 0.01, respectively. For the same traits, the total maternal heritability estimates were 0.17 ± 0.02, 0.11 ± 0.01, and 0.09 ± 0.01, respectively. In contrast, to the Gibbs sampling starting values, the genetic variance was partitioned between A and H such that direct genetic λ estimates for CD, GL, and BW were 0.36 ± 0.05, 0.62 ± 0.03, 0.75 ± 0.03, respectively. Maternal genetic λ estimates ranged from 0.01 ± 0.01 (for BW) to 0.05 ± 0.01 (for CD). The results imply that λ values are heterogeneous in multivariate single-step genomic evaluation. Further studies are needed to investigate the consequences of using heterogenous λ values for direct genetic and maternal genetic components in multivariate single-step evaluation in terms of model dimensions, solver convergence rate, and model forward predictive ability.

INTRODUCTION

Genomic selection has been implemented in Australia's BREEDPLAN genetic evaluation using single-step genomic methods (Johnston *et al*. 2018). An important step in genomics-assisted genetic evaluation involves the use of unbiased genetic parameters, including weighing factors or lambdas (λ), which affect the accuracies of genomic estimated breeding values (GEBVs). Optimum λ estimates are usually obtained using single-trait cross-validation procedures, although other methods have been suggested (Zhang *et al*. 2018).

It can be shown that a single-step model involving a scalar λ is simply a condensed form of an extended model containing two genetic factors, $u_A \sim N(0, A \otimes \Sigma_A)$ and $u_H \sim N(0, H \otimes \Sigma_H)$ where A is the

*A joint venture of NSW Department of Primary Industries and the University of New England

pedigree derived numerator relationship matrix, H is the joint relationship matrix, and Σ_A and Σ_H are covariance matrices. The total genetic variance (Σ_T) is equal to $\Sigma_A + \Sigma_H$, where it is assumed that $\Sigma_H \oslash \Sigma_T \equiv ii'\lambda$, where i is an identity vector of respective dimension. However, in multivariate models or univariate models which contain several genetic factors (e.g. direct, maternal), $\Sigma_H \oslash \Sigma_T \equiv ii'\lambda$ is very unlikely. Heterogeneous λs can be estimated and accounted for by using the extended single-step model in variance component estimation and best linear unbiased prediction. To our knowledge, no studies have published λs estimated for categorical traits such as calving difficulty using extended multivariate linear-threshold single-step models. The objective of this study was to estimate variances and optimum λs with an extended single-step model for calving difficulty (CD) together with gestation length (GL), and birth weight (BW) using Australian Angus data.

MATERIALS AND METHODS

Data. Phenotypes, genotypes, and pedigree for this study were obtained from the data extract submitted by Angus Australia for BREEDPLAN evaluation. Phenotypic data included CD, GL, and BW, with GL and BW were pre-adjusted for sex and age of the dam (Graser *et al*. 2005). GL and BW were measured in days and kilograms, respectively. CD was scored using three categories (Jeyaruban *et al*. 2016). Unassisted birth was represented by score of 1, while easy and hard pull were represented by scores of 2 and 3, respectively.

Contemporary groups (CGs) were formed according to the BREEDPLAN format (Graser *et al*. 2005). For each trait, CGs with less than 5 animals were discarded from the analysis. Further for CD, CGs with single CD score were eliminated. If the proportion of any score was less than 5% within a CG, that CG was discarded. The average number of observations per CG, for CD, GL, and BW were 84, 37, and 40, respectively.

The data consisted of 129,851 animals with phenotypes. Frequencies for CD scores 1, 2, and 3 were 78,653 (89.2%), 6,565 (7.4%), and 2,929 (3.3%), respectively. The number of dams with phenotypes for CD, GL, and BW were 7,536, 2,038, and 8,448, respectively. The pedigree included ancestors over 4 generations and consisted of 327,395 animals with 27,145 sires and 186,339 dams. A total of 45,575 animals were genotyped for 56,009 SNP markers. The GRM was constructed according to VanRaden (2008) and adding 0.0001 to the diagonal to guarantee positive definiteness.

Analyses. The data were analysed with multivariate model $y = Xb + Zu + Wp + e$ (model p) and multivariate extended single-step model $y = Xb + Zu_A + Zu_H + Wp + e$ (model h), where y is a vector of phenotypic observations of GL, BW, and CD, X is a block-diagonal design matrix linking fixed effects to their respective observations; Z is a block-diagonal design matrix linking direct and maternal genetic effects in $u \sim N(0, A \otimes \Sigma_G)$, $u_A \sim N(0, A \otimes \Sigma_A)$, and $u_H \sim N(0, H \otimes \Sigma_H)$ to their respective observations where Σ_G, Σ_A, and Σ_H are co-variance matrices, A is the pedigree derived numerator relationship matrix, and H is the joint relationship matrix using the genomic relationship matrix (G) obtained as $G + I.001$, and I is an identity matrix; W is a block-diagonal design matrix linking maternal permanent environmental effects in $p \sim N(0, I \otimes \Sigma_p)$ to their respective observations, and $e \sim N(0, I \otimes \Sigma_E)$ is a vector of residuals. λ was estimated as $\Sigma_H \oslash \Sigma_T$, where Σ_T is the total genetic variance obtained from model h.

Given the categorical nature of CD, a multivariate threshold model Gibbs sampling approach was used to estimate Σs in models p and h. Starting values for Σs for model p were obtained from heuristic partitioning of the observed phenotypic covariance matrix. Starting values for Σs for model h were derived from the results of model p, where $\Sigma_H = \Sigma_G \times 0.1$ and $\Sigma_A = \Sigma_G \times 0.9$. This implies that $\Sigma_T = \Sigma_H + \Sigma_A$. In all case, the prior degrees of freedom were zero. Results for models were confirmed by a second analysis with starting values $\Sigma_H = \Sigma_G \times 0.9$ and $\Sigma_A = \Sigma_G \times 0.1$. Thresholds for CD were fixed to be 0 and 1, and the residual variance was unconstrained. For all models, single chains of 200,000 iterations were sampled, and the first 50,000 samples were discarded as burn-in. To avoid autocorrelation, every 20th sample was stored, and a total of 7,500 samples were kept for

computing posterior means and standard deviations.

RESULTS AND DISCUSSION

Descriptive statistics for the studied traits are presented in Table 1. Table 2 presents the parameters derived from model p and h. The direct genetic and maternal genetic heritability estimates for CD, GL, and BW obtained from model p were similar to those obtained from model h when using Σ_T (Table 2), implying that model h does not explain additional variation in the data compared to model p. The highest direct genetic heritability was obtained for GL (0.68), followed by BW (0.44) and CD (0.39), and the maternal heritability estimates ranged from 0.09 (for BW) to 0.17 (for CD).

Table 1. Descriptive statistics for calving difficulty (CD, score), gestation length (GL, days), and birth weight (BW, kg)

Trait	Animals	% genotyped	Mean	Min*	Max*	SD*
CD	88,147	4	1.1	1	3	0.4
GL	43,140	44	280.2	259.8	296.8	4.6
BW	102,864	42	37.5	16.0	65.9	5.0

*Min, minimum; Max, maximum; SD, standard deviation; ‡with phenotypes

Table 2. Variances (σ^2), covariance (σ), heritabilities (h^2), and correlation (r) derived from pedigree only (p) and extended single-step (h) models for calving difficulty (CD), gestation length (GL), and birth weight (BW)

Parameter*	Model p			Model h		
	CD	GL	BW	CD	GL	BW
σ_d^2	2.56 ± 0.28	12.69± 0.74	7.26 ± 0.30	2.59 ± 0.28	13.23± 0.59	7.83 ± 0.28
σ_m^2	1.15 ± 0.16	1.89 ± 0.32	1.56 ± 0.13	1.15 ± 0.16	2.11±0.22	1.61 ± 0.13
$\sigma_{d,m}$	-0.68 ± 0.17	-1.87 ± 0.38	-0.45 ± 0.15	-0.71 ± 0.17	-2.26±0.30	-0.64 ± 0.13
σ_c^2	0.30 ± 0.13	0.66 ± 0.24	0.37 ± 0.10	0.31 ± 0.14	0.53±0.20	0.45 ± 0.11
σ_e^2	3.32 ± 0.23	6.01 ± 0.39	9.14 ± 0.18	3.26 ± 0.21	5.97±0.34	8.67 ± 0.16
σ_p^2	6.66 ± 0.23	19.38± 0.23	17.90± 0.11	6.61 ± 0.21	19.5 ± 0.22	17.92± 0.12
$h_d^{2\ddagger}$	0.38 ± 0.04	0.65 ± 0.03	0.41 ± 0.02	0.39 ± 0.04	0.68 ± 0.02	0.44 ± 0.01
$h_m^{2\ddagger}$	0.17 ± 0.02	0.10 ± 0.02	0.09 ± 0.01	0.17 ± 0.02	0.11 ± 0.01	0.09 ± 0.01
$r_{d,m}$	-0.39 ± 0.07	-0.38 ± 0.05	-0.13 ± 0.04	-0.41 ± 0.07	-0.43 ± 0.03	-0.18 ± 0.03

*Direct genetic (d), maternal genetic (m), permanent environment effect of dam (c), residual (e) and total phenotypic (p) effects; ‡Total heritability estimates for model h.

A higher proportion of direct genetic variance is explained by the genomic factor than the polygenic factor in model h for BW and GL yielding λ values 0.75 and 0.62, respectively (Table 3 and Table 4). In contrast, the polygenic factor in model h explained the highest proportion of additive genetic variance for CD (Table 3) yielding a lower λ (0.36) than for BW and GL (Table 4). The observed pattern of variance partitioning between polygenic and genomic factors in model h for CD versus GL and BW suggests that λ values are highly sensitive to availability of genotypic information for each trait (Table 1), and therefore, heterogeneous in multivariate single-step genomic evaluation.

The maternal genetic λ estimates were near zero, and ranged from 0.01 for BW to 0.05 for CD (Table 4) due to greater maternal genetic variance partitioning for polygenic factor in model h (Table 3). This could be a result of low number of genotyped dams per trait in the current dataset, and therefore, maternal genetic λ values are expected to rise in the future if the number of genotyped

dams increases.

Table 3. Genetic variances (σ^2), covariance (σ), correlation (r), and heritabilities (h^2) for direct genetic (d) and maternal genetic (m) components accounted by the polygenic and genomic factors using the extended single-step model (model h) for calving difficulty (CD), gestation length (GL), and birth weight (BW)

	Polygenic factor			Genomic factor		
	CD	GL	BW	CD	GL	BW
σ_d^2	1.66 ± 0.30	4.99 ± 0.56	1.97 ± 0.26	0.93 ± 0.04	8.23 ± 0.27	5.86 ± 0.15
σ_m^2	1.10 ± 0.17	2.03 ± 0.22	1.60 ± 0.12	0.06 ± 0.01	0.08 ± 0.01	0.02 ± 0.01
$\sigma_{d,m}$	-0.54 ± 0.18	-1.84 ± 0.30	-0.61 ± 0.13	-0.17 ± 0.02	-0.42 ± 0.04	-0.02 ± 0.03
h_d^2	0.25 ± 0.04	0.25 ± 0.03	0.11 ± 0.01	0.14 ± 0.01	0.42 ± 0.01	0.33 ± 0.01
h_m^2	0.17 ± 0.02	0.10 ± 0.01	0.09 ± 0.01	0.01 ± 0.03	0.00 ± 0.01	0.00 ± 0.01
$r_{d,m}$	-0.40 ± 0.09	-0.58 ± 0.05	-0.35 ± 0.06	-0.76 ± 0.03	-0.53 ± 0.04	-0.08 ± 0.11

Table 4. Estimates of lambdas (λ) for direct and maternal genetic components for selected traits using the multivariate linear-threshold single-step model (model h)

Λ	Calving difficulty	Gestation length	Birth weight
Direct genetic	0.36 ± 0.05	0.62 ± 0.03	0.75 ± 0.03
Maternal genetic	0.05 ± 0.01	0.04 ± 0.01	0.01 ± 0.01

Our results suggest considering different λs for each trait rather than one global value across traits. An extended multivariate single-step model allows estimation of heterogeneous λs in variance component estimation. However, further studies are needed to investigate the consequences of using heterogenous λ estimates for multivariate evaluations in terms of model dimensions, solver convergence rate, and model forward predictability.

CONCLUSIONS

By using an extended multivariate linear-threshold single-step model, heterogeneous direct genetic λs were obtained for GL, BW, and CD, which ranged from 0.36 (CD) to 0.75 (BW). Maternal genetic λ estimates ranged from 0.01 for BW to 0.05 for CD. Results suggest employing an extended single-step model with variance partitioning between genomic and polygenic factors accounting for heterogeneous λs in future BREEDPLAN genomic evaluation for the studied traits.

ACKNOWLEDGEMENTS

The authors would like to thank the Meat and Livestock Australia (MLA) for their financial support through L.GEN.1704, and Angus Australia for providing data for this study.

REFERENCES

Graser H.-U., Tier B., Johnston D.J., and Barwick S.A. (2005). *Aust. J. Exp. Agric.* **45**: 913.

Jeyaruban M.G., Johnston D., Tier B., and Graser H.-U. (2016) *Anim. Prod. Sci.* **56**: 927.

Johnston D.J, Ferdosi M., Connors N., Boerner V., Cook J., Girard C., Swan A., and Tier B. (2018). In *Proc. 11th World Cong. Genet. Appl. Livest. Prod.* **11:** 269.

VanRaden P.M. (2008). *J. of Dairy Sci.* **91**: 4414.

Zhang Y., Swan A., and Banks R. (2018). In *Proc. 11th World Cong. Genet. Appl. Livest. Prod.* **11:** 348.