Validation of single-step GBLUP genomic predictions from threshold models using the linear regression method: an application in chicken mortality

Bermann[1], M., Legarra[2], A., Hollifield[1], M.K., Masuda[1], Y., Lourenco[1], D., Misztal[1], I.

[1] Department of Animal and Dairy Science, University of Georgia, Athens, GA, USA.

[2] UMR GenPhySE, INRA, Toulouse, BP 52626, 31326 Castanet Tolosan, France.

INTRODUCTION

- Traditional methods to estimate the accuracy of EBV are not applicable to thresholds models
- The objective of this study was to determine if the LR method (Legarra & Reverter, 2018) could be used to validate genomic threshold models used for analysis of categorical data
- The secondary goal was to estimate the increase in accuracy by adding genomic information

DATA

- The simulated dataset included the survival status and genotypes for 45,000 and 13,500 individuals, respectively
- The real dataset had phenotypes and genotypes for 186,596 and 18,047 individuals, respectively
- Mortality rate was 7% for both datasets

LR METHOD

- Breeding values were estimated with the whole ($\hat{\boldsymbol{u}}_w$) and partial datasets ($\hat{\boldsymbol{u}}_p$)
- The accuracy was estimated by $\sqrt{\frac{cov(\hat{\boldsymbol{u}}_p,\hat{\boldsymbol{u}}_w)}{(1+\bar{F}-2\bar{f})\,\sigma_{u,\infty}^2}}$
- $\bar{F}$ is the average inbreeding coefficient and $\bar{f}$ is the average relationship among focal animals.
- The genetic variance for the focal individuals is denoted by $\sigma_{u,\infty}^2$
- The increase in accuracy was estimated without ratio of variances ($\rho_{A,G}^{-1} - 1$) and with ratio of variances ($\frac{\sigma_A}{\sigma_G}\rho_{A,G}^{-1} - 1$)
- $\rho_{A,G}$ is the correlation between EBV and GEBV
- σ_A and σ_G are the genetic standard deviations estimated with BLUP and ssGBLUP, respectively

RESULTS

Table 1. Accuracies (SD) for each dataset and each scenario

	BLUP		ssGBLUP	
	Estimated value	True value	Estimated value	True value
Simulated dataset	0.45 (0.07)	0.37 (0.08)	0.76 (0.08)	0.65 (0.04)
Real dataset	0.41	-	0.47	-

Table 2. Percentage of increase in accuracy by adding genomic information to the pedigree-based

	Simulated dataset		Real dataset
	Estimated value	True value	Estimated value
Increase without variance ratio	85	75	15
Increase with variance ratio	76	75	22

CONCLUSIONS

- The LR method allows to estimate accuracy when traditional validation methods are not applicable
- Estimates of accuracy are biased due to selection and incorrect estimation of the genetic variance in the focal population (Macedo et al., 2020)
- Ratio of genetic variances should be included when calculating the increase in accuracy due to addition of genomic information
- Small improve in accuracy in the real dataset was due to population structure and low heritability of the trait

REFERENCES

- Legarra, A., & Reverter, A. (2018). https://doi.org/10.1186/s12711-018-0426-6
- Macedo, F., Reverter, A., & Legarra, A. (2020). https://doi.org/10.3168/jds.2019-16603