

MULTIBREED GENETIC EVALUATION OF GROWTH IN BEEF CATTLE USING RANDOM REGRESSION LINEAR SPLINE MODELS

J. P. Sánchez, I. Aguilar, J. K. Bertrand, I. Misztal

ABSTRACT

The objective of this study was to examine the feasibility of using random regression, spline (**RR-spline**) models for fitting growth traits in a multibreed beef cattle population. To meet the objective, the evaluation results from the RR-spline model were compared to the multi-trait (**MT**) model when both were fit to a data set (1.8 million records and 1.1 million animals) provided by the American Gelbvieh Association. The effect of prior information on the predicted breeding values of sires was also investigated. When both MT and RR-spline models were applied to a data set containing records for weaning weight (WWT) and yearling weight (YWT) within specified age ranges, the rankings of bulls direct breeding values (as measured via Pearson correlations) provided by both models were comparable, with slightly greater differences in the re-ranking of bulls observed for YWT evaluations (≥ 0.99 for BWT and WWT and ≥ 0.98 for YWT). For maternal effects, the estimated correlations were slightly smaller. The inclusion of prior information had negligible effect in the overall ranking for bulls with greater than 20 progeny records; however, the effect of prior information for breeds or groups poorly represented in the data was very important. Pearson correlations for direct and maternal WWT and YWT ranged from 0.95 to 0.98 when comparing evaluations of data sets where out of range age records were removed or retained. Using the random regression approach allows for the retention of records that are normally eliminated in MT models because they are outside of normal age ranges in multibreed populations.

INTRODUCTION

Multibreed (MB) genetic evaluation for beef growth traits in the U.S. uses an approach presented by Klei et al. (1996) that employs Bayesian methodology, which combines literature estimates of breed and heterosis effects with data to provide across-breed predictions of breeding value. Currently, the MB evaluation uses multi-trait (**MT**) models that treat weights recorded at different ages as different traits. The use of MT models creates the need for the establishment of age ranges and the elimination of measurements recorded outside them; this could lead to large data losses as more animals are measured outside this age range. Recently linear spline, random regression (**RR-spline**) models have been proposed as an alternative to MT models for the genetic evaluation of growth data of varying ages in single breed populations (Bohmanova et al., 2005; Robbins et al., 2005; Iwaisaki et al., 2005), and Misztal (2006) has shown properties of RR-spline models. The general aim of this study was to determine the suitability of RR-spline models for evaluating growth in MB populations.

MATERIAL AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the data were obtained from an existing data base. The data used in this study was provided by the American Gelbvieh Association. Two data sets were formed. D1 only used weaning weights (**WWT**) and yearling weights (**YWT**) normally used in MT models, ie, weights taken between the age ranges of 160-250 days and 305-410 days, respectively. D2 contained WWT and YWT that ranged from 50-285 days and 286-600 days, respectively. The number of records in D1 for BWT, WWT and YWT were 841,848, 641,243, and 164,399, respectively. There was a 17% and 35% increase in WWT and YWT records, respectively, from D1 to D2. However, nearly 71% and 76% of the records measured outside D1 age ranges were within 20 days of the age range bounds.

Specific information for the MB model employed in the present study is provided by Legarra et al. (2007); therefore a detailed description of the MB model will not be presented. However, the general idea of the evaluation procedure is to use a Bayesian approach, first proposed by Klei et al. (1996), that combines literature estimates of breed and heterosis effects with data to provide across-breed predicted breeding values. In the evaluation procedure the prior mean of the heterosis and breed effects were an average of literature values, and the distribution was taken to be normal with a variance that allowed more or less weight to be given to the literature values. The weight given to the literature relative to the data is established in a somewhat arbitrary manner in most MB genetic evaluation systems.

Under the MT model the record of trait t , belonging to the i^{th} animal, offspring of the d^{th} dam, and produced in the j^{th} contemporary group was explained by this linear model.

$$y_{ijid} = \text{fixed}_{ijid} + a_{ii} + mg_{id} + me_{id} + e_{ijid} ,$$

where

$$\text{fixed}_{ijid} = CG_{ij} + \sum_{k=1}^3 (age_{ii} - MES_t)^k * \beta_{k,MES} + \sum_{k=1}^6 Dh_{ki} * D\beta_{tk} + \sum_{k=1}^6 Mh_{kd} * M\beta_{tk} ,$$

In the random part of the model a_{ii} refers to the additive genetic effect of animal i for trait t , mg_{id} is the maternal additive genetic effect of dam d for trait t , me_{id} is the maternal environmental effect of dam d for trait t and e_{ijid} is the random residual effect for the trait t of the animal i . In the fixed part of the model for both MT and RR-Spline, CG_{ij} is the contemporary group (**CG**) effect j for the trait t . The definitions for contemporary group for each trait were the standard ones employed in the beef industry. $\sum_{k=1}^3 (age_{ii} - MES_t)^k * \beta_{k,MES}$ is a cubic polynomial function of the deviation of the actual age of the animal at time of weighing to the expected measurement ages, MES_t , (205 d for WWT and 365 d for YWT). This function is not considered for BWT. Dh_{ki} is the coefficient of the expected fraction of F_1 heterosis in individual i for the group combination k . These coefficients were computed based on the assumption that heterosis effects are primarily due to genetic dominance interactions, as indicated by Klei et al. (1996). The proportion of F_1 heterosis in a dominance

model is determined by the probable proportion of loci where the alleles received from the parents are from different breeds.

Under the RR-spline model, the t^{th} measurement on animal i , offspring of the dam d , and produced in the j^{th} contemporary group was explained by this linear model.

$$y_{ijd} = \text{fixed}_{ijd} + \sum_{k=1}^3 \phi_k(t_{ii}) \cdot a_{ik} + \sum_{k=1}^3 \phi_k(t_{ii}) \cdot mg_{dk} + \sum_{k=1}^3 \phi_k(t_{ii}) \cdot p_{ik} + \sum_{k=1}^3 \phi_k(t_{ii}) \cdot me_{dk} + e_{ijd},$$

where

$$\text{fixed}_{ijd} = CG_j + \sum_{k=1}^3 \phi_k(t_{ii}) \cdot A\beta_k + \sum_{h=1}^6 \sum_{k=1}^3 \phi_k(t_{ii}) \cdot Dh_{hi} \cdot D\beta_h + \sum_{h=1}^6 \sum_{k=1}^3 \phi_k(t_{ii}) \cdot Mh_{hi} \cdot M\beta_h,$$

The random effect part of the model was fitted using linear spline functions, with knots at 1, 205, and 365 days of age. Thus, a_{ik} , mg_{dk} , p_{ik} , and me_{dk} represent the coefficients for the knot k in the linear splines describing the additive genetic value for animal i , the additive genetic maternal value of dam d , the permanent environmental effect of animal i and the maternal environmental effect of dam d , respectively. $\phi_k(t_{ii})$ represents the value of the k^{th} covariate at time t in individual i . These covariates were computed as described by Misztal (2006). When t_{ii} was beyond the last knot, the covariate for this last knot was computed assuming an increasing function of time ($(t_{ii} - 365)/(365 - 205)$) for direct effects and constant (1) for the maternal effects

The measurement age of the animal was fitted using a three knot linear spline, with covariates as in the direct genetic random part of the model, and coefficients $A\beta_k$. The heterosis effects were considered to be a function of time. This function was again a linear spline with knots at 1, 205 and 365 days for the direct heterosis, similar to the random direct part of the model. Beyond the last knot, the covariate $\phi_3(> 365)$ was computed assuming the same increasing function of time. For the maternal heterosis this function was considered, like for the maternal random part, to be constant. The variance-covariance matrix used in the analysis is reported by Legarra et al. (2004), and because the knots in the RR-spline model were placed at 1, 205 and 365 days, the variances-covariances for the direct and maternal genetic effects and the maternal permanent environmental effects were equal to those used in the MT analysis. The residual variance in the MT model was split into two terms in the RR-spline model. One of the terms accounted for the variance of the permanent environmental effects plus the residual covariances among the traits in the MT model. The other term was the residual variance in the RR-spline model which was assumed to be heterogeneous and was described as a function of age.

D1 was analyzed using both a MT and RR-spline model, and D2 was analyzed using only the RR-spline model. In both analyses, BWT and WWT were pre-corrected for the age of the dam according to the equations for the Gelbvieh breed (Beef Improvement Federation, 1996). For both models, different prior parameters were assessed, but the same value was employed for both BOF and heterosis effects. Values were selected to place weak, medium and strong weight on prior information. All the analyses were carried out using a modified

version of BLUP90IOD (Tsuruta et al., 2001) with modifications consisting of the incorporation of prior information; details on these modifications were described by Legarra et al. (2007). Comparisons between models were done via Pearson correlations between predicted breeding values of bulls.

RESULTS AND DISCUSSION

Pearson correlations between the EBV of sires with more than 20 offspring with BWT records obtained after MT or RR-spline analysis of D1 for each of three scenarios of prior information are shown in Table 1. Fairly good agreement was observed between models with correlations ranging from 0.92 to 0.99. But as expected, the lowest values were observed for WWT and YWT due to the age variation within these two traits. For BWT there was no variation in age, thus the models were almost the same. When the top 100 ranked bulls were compared across methods (Table 1), it was observed that for all the scenarios of prior information for direct WWT and YWT effects, between 12 and 25 bulls were dropped from the top 100 ranked bulls after using MT when compared to RR-spline model. For maternal effects, the number of bulls that differed in the top 100 between the two models was slightly larger, between 16 and 39. However, most of the bulls that differed in the top 100 for direct and maternal breeding value were bulls that were close to the threshold and needed only to change a small number of places to drop out of, or enter, the top 100.

Table 2 shows Pearson correlations between breeding values of bulls with more than 20 offspring with BWT records estimated using RR-spline analysis of D1 and D2 under the three scenarios of prior information studied. In this case the differences between correlations are due to the different data sets used, which reflected differences in the amount of information. The correlations for direct and maternal BWT were high (1.0) across all scenarios. The correlations for direct and maternal WWT and YWT ranged from .95 to .98 across traits with very little differences observed within a trait for the different prior information scenarios. The number of animals dropped from the top 100 bull ranks after using D2 is also shown in Table 2. The results from this table indicate that changes in the breeding values for some bulls occur due to the additional data provided when animals that are usually eliminated for out of age ranges at weaning and yearling are no longer eliminated.

RR-spline models appear to be a valuable tool for fitting growth in beef cattle; previous studies in single breed populations have shown their desirable attributes in terms of convergence properties and relative freedom from estimation artifacts (Bohmanova et al., 2005; Iwaisaki et al., 2005, Misztal et al., 2006; Robbins et al., 2005). Also their results are directly comparable to those obtained after MT analysis when the knots are placed at the usual age measurement points. Our results suggest that random regression, linear spline models can also be useful in MB populations. In general, this study indicates that the assumed prior variance does not have a large impact in the ranking of bulls with moderate to large offspring groups with records for direct effects, but for maternal effects, the weight given to the priors appears to be more important. Although not shown, the results also indicate that the weight given to the priors could have a large impact on animals within breeds that are poorly represented in the data.

CONCLUSIONS

The application of random regression using linear splines has been extended to the genetic evaluation of a MB population. The application of these models could be useful in genetic evaluation by allowing the inclusion of out of range age records that are eliminated in current multi-trait analyses, which should lead to increases in the accuracy of evaluation and genetic progress. The use of prior information for heterosis and additive genetic effects of breeds is mandatory, and although the overall rankings are relatively robust to this information, the evaluations of particular breeds may be greatly affected.

LITERATURE CITED

- Beef Improvement Federation. 1996. Guidelines for uniform beef improvement programs. (7th Ed.). Beef Improvement Federation, Colby, KS, USA.
- Bohmanova, J., I. Misztal, J. K. Bertrand. 2005. Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. *J. Anim. Sci.* 83:62-67.
- Iwaisaki, H., S. Tsuruta, I. Misztal, J. K. Bertrand. 2005. Genetic parameter estimated with multitrait and linear spline-random regression using Gelbvieh early growth data. 83:757-763.
- Klei L., R. L. Quaas, E. J. Pollak, and B. E. Cunningham. 1996. Multiple breed evaluation. Pages 106-113 in Proc. Research Symposium and Annual Meeting, Beef Improvement Federation. Birmingham, AL, USA.
- Legarra, A., I. Misztal, and J. K. Bertrand. 2004. Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle. *J. Anim. Sci.* 82:1564-1571.
- Legarra, A., J. K. Bertrand, T. Strabel, R. L. Sapp, J. P. Sánchez, and I. Misztal. 2007. Multi-breed genetic evaluation in a Gelbvieh population. *J. Anim. Breed. Genet.* 124:286-295.
- Misztal, I. 2006. Properties of random regression models using linear splines. *J. Anim. Breed. Genet.* 123:74-80.
- Robbins, K. R., I. Misztal, J. K. Bertrand. 2005. A practical longitudinal model for evaluating growth in Gelbvieh cattle. *J. Anim. Sci.* 83:29-33.
- Tsuruta, S., I. Misztal, and I. Strandén. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. *J. Anim. Sci.* 79:1166-1172.

Table 1. Pearson Correlations between predicted breeding values of bulls with more than 20 offspring with records at birth obtained using MT and RR-spline models with data set D1¹ and under three scenarios of prior information².

Trait ³	Prior Information		
	Weak	Medium	Strong
D-BWT	1.00 2	1.00 2	1.00 1
M-BWT	1.00 4	1.00 5	1.00 2
D-WWT	0.99 15	0.99 13	0.99 12
M-WWT	0.99 20	0.98 16	0.98 18
D-YWT	0.98 18	0.99 13	0.98 25
M-YWT	0.96 27	0.92 35	0.93 39

¹ D1 = The ranges of ages for considering a record as valid was 160-250 and 305-410 days for WWT and YWT.

² In bold are shown the number of bulls present in the top 100 after RR-spline analysis which are not in the top 100 after MT analysis.

³ D-BWT = Direct Birth Weight, M-BWT = Maternal Birth Weight, D-WWT = Direct Weaning Weight, M-WWT = Maternal Weaning Weight, D-YWT = Direct Yearling Weight, M-YWT = Maternal Yearling Weight.

Table 2. Pearson Correlations between EBVs of bulls with more than 20 offspring with record at birth predicted using data sets D1 and D2¹, using RR-spline model and under three scenarios of prior information².

Trait ³	Prior Information		
	Weak	Medium	Strong
D-BWT	1.00 3	1.00 4	1.00 4
M-BWT	1.00 7	1.00 8	1.00 6
D-WWT	0.96 24	0.96 23	0.96 22
M-WWT	0.96 27	0.96 30	0.96 30
D-YWT	0.96 23	0.96 24	0.97 25
M-YWT	0.96 29	0.95 31	0.98 24

¹ D1 = The ranges of ages for considering a record as valid was 160-250 and 305-410 days for WWT and YWT, D2 = The ranges of ages for considering a record as valid was 50-280 and 281-600 days for WWT and YWT.

² In bold are shown the number of bulls present in the top 100 after the analysis using D2 which are not in the top 100 after the analysis of D1.

³ D-BWT = Direct Birth Weight, M-BWT = Maternal Birth Weight, D-WWT = Direct Weaning Weight, M-WWT = Maternal Weaning Weight, D-YWT = Direct Yearling Weight, M-YWT = Maternal Yearling Weight.