

ESTIMATES OF COVARIANCE FUNCTIONS FOR GROWTH OF ANGUS CATTLE FROM RANDOM REGRESSION ANALYSES FITTING B-SPLINE FUNCTIONS

Karin Meyer

Animal Genetics and Breeding Unit¹, University of New England, Armidale, NSW 2351

SUMMARY

Estimates of variance components and genetic parameters for growth of Angus cattle from birth to 820 days of age were obtained from a random regression analysis. Trajectories were modelled through quadratic B-spline functions with 7 random regression coefficients for direct genetic and permanent environmental effects, and 5 coefficients for the corresponding maternal effects. Data comprised a large set of field records, selecting only animals with at least 4 weights recorded. Results showed smooth estimates of variances, increasing with weight and age, with good interpolation for early ages with few records and little end-of-range problems at the highest ages. On the whole, estimates of genetic parameters were consistent with literature results. However, estimates of heritabilities for weaning weight were higher for direct and lower for maternal effects than usually found, indicating a different partitioning between direct and maternal variances in random regression analyses than in standard multivariate analyses.

Keywords : Genetic parameters, beef cattle, growth, random regression, B-splines

INTRODUCTION

Random regression (RR) analyses have become a standard procedure for analyses of longitudinal data from livestock recording schemes, such as weights of animals recorded at different ages. However, applications to field data for beef cattle so far have suffered from sparsity of records, and numerical problems inherent in the use of orthogonal polynomials to model trajectories. This paper presents a RR analysis for a large set of fields records of Angus cattle, severely edited to include only animals with at least four weights recorded, i.e., as good a structure as might be obtained. As suggested by Rice and Wu (2001) regression on B-spline basis functions of age at recording is used to model trajectories. These are expected to be more robust to problems often encountered in RR analyses of sparse data than high order polynomials; see Meyer (2005a) for a brief review.

MATERIAL AND METHODS

Data. Records consisted of weights of Angus cattle from birth to 820 days of age, extracted from the National Beef Recording Scheme data base in July 2004. Raw data comprised close to 2 million weight and 1 million pedigree records. Initial edits eliminated implausible records, records more than 3 standard deviations from the respective weekly mean, and records clearly 'out of sequence'. Subsequently, all records in contemporary groups (CG) with less than 3 observations and for all animals with less than four weights were disregarded recursively. This left 284,330 records on 64,307 animals, i.e., an average of 4.43 records per animal, with a mean weight of 293.4 kg and mean age of 302.2 days. Birth weight was available for 58,161 animals. Figure 1 shows the distribution of weights for other ages at recording, together with corresponding means.

¹ AGBU is a joint venture of NSW Department of Primary Industries and the University of New England

There were 42,052 (65.4%), 18,328 (28.5%) 3,208 (5.0%) and 719 (1.1%) animals with 4, 5, 6 and 7 – 9 records, respectively. Animals with records were progeny of 3,539 sires. Pedigree information for parents not in the data was obtained for up to 4 generations backwards, resulting, after pruning, in 104,905 animals in the analysis. Whilst only 34.6% of animals in the data had 5 or more records themselves, 88.4% had 5 or more records ‘at a genetic level’, i.e., counting information on close relatives. Corresponding proportions for 6 or more and 7 or more records were 48.7% and 11.4%.

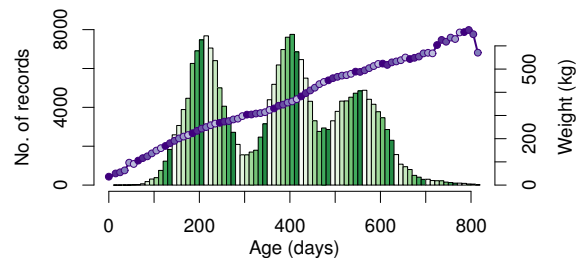


Figure 1. Distribution of records and means (●) for ages in 10-day intervals.

Analyses. Estimates of (co)variance components were obtained by Bayesian analysis, employing a Gibbs sampling algorithm, as implemented in RRGIBBS (Meyer 2002b). Three Markov chains with a total of 2.4 million samples were generated, assuming flat bounded priors. Variances and genetic parameters for ages in 10 day intervals were calculated thinning to every 20th sample. Estimates were obtained as posterior means, disregarding the first 30,000 samples in each chain as “burn-in” period, and point-wise approximate 95% highest posterior density (HPD) regions were determined.

Random effects. The model of analysis fitted a RR on quadratic B-spline basis functions of age. Direct genetic (A) and permanent environmental (PE) trajectories were modelled with 7 regression coefficients each, dividing the range of ages into 5 equal intervals of 164 days. Maternal genetic (M) and permanent environmental (C) effects were considered to affect records up to 500 days only. Placing knots at equal intervals of 166.7 days yielded 5 RR to be fitted for each. Direct and maternal effects were assumed to be uncorrelated. This allowed pedigrees for maternal effects to be pruned separately, resulting in effects for 54,773 genetic dams to be fitted. Of these, 31,519 had progeny in the data, and 10,460 had records themselves as well. Assigning dummy identities to a small proportion of animals in the data with unknown dams (or foster dams of a different breed) yielded 34,960 permanent environmental dams to be considered. Residual effects were assumed independently distributed with heterogeneous temporary environmental variances (σ_i^2). Changes in σ_i^2 with age were modelled as a step function with 19 classes (0, 1–60, 61–90, ..., 271–300, 301–360, ..., 721–780 and 781–820 days). This gave a total of 105 covariance components to be estimated.

Fixed effects. Changes in mean with age were modelled through fixed regressions on quadratic B-spline functions of age with knots as for direct, random effects. These were nested within sex, dam age class (in years, treating ages ≥ 7 years as one class) and birth type (single vs. twin), resulting in a total of 77 regression coefficients. In addition, CGs were fitted as cross-classified fixed effects with 11,098 levels. CG were defined as herd-sex-management group-year/month of weighing subclasses for birth weights, and herd-sex-management group-date of weighing subclasses for other weights. CG classes were further subdivided, applying an “age slicing” of 45 days up to 300 days, and 60 days for higher ages. If this resulted in a subclass with less than 5 records for the highest ages in the CG, this last subclass was merged with the previous age subclass, provided the range of ages did not exceed 54 days for weights up to 300 days, and 72 days otherwise.

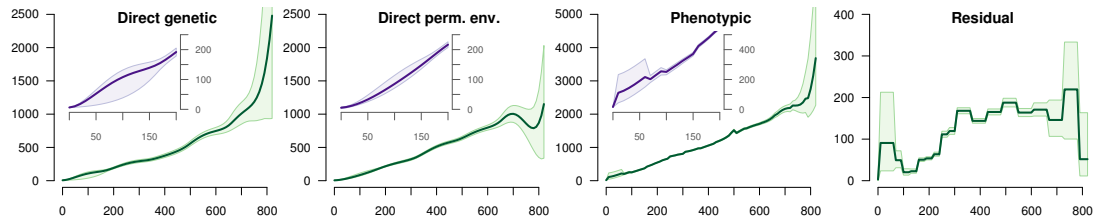


Figure 2. Estimates of variance components (kg^2) together with 95% HPD regions (shaded area).

RESULTS AND DISCUSSION

Estimates of phenotypic, direct additive genetic and permanent environmental, and residual variances (σ^2) are shown in Figure 2, together with their approximate 95% HPD regions. The magnitude of confidence regions clearly reflects the paucity of records after birth to about 120 days of age (see insets) and after 700 days. Overall, estimates of variances increased steadily with age and mean. Variance functions for the early ages with few records were smooth and plausible i.e., regression on B-spline functions yielded a good interpolation for the interval with little information available. Preliminary analyses (not shown) were carried out for a subset of the data (considering ages up to 820 days), in- and excluding records from birth to 80 days. Estimates of variances for direct effects from the two analyses were virtually identical, indicating that lack of records in the first interval did not affect estimates for subsequent ages adversely. Estimates of the direct genetic variances after about 700 days were highly variable, with some indication of a steep increase for the last 30 days. A corresponding drop in estimates of σ^2 suggested problems in partitioning variation for these ages with few observations. Disregarding this, estimates of temporary environmental variances after 300 days were essentially constant. Small 'knot effects', i.e., changes in variance functions at knots at 164, 328, 492 and 656 days, were evident, especially for direct genetic variances.

Corresponding estimates of variance ratios are summarised in Figure 3. Estimates of direct heritabilities were 39.7%, 35.1%, 35.6% and 43.8% for weights at birth, 200, 400 and 600 days, respectively, with narrow confidence regions throughout. Whilst estimates for birth weight and weights after 400 days agreed well with values commonly reported for yearling and final weights for Angus cattle (e.g. Koots et al. 1994), estimates for weaning weight (200 days) were higher than most literature values. The reason for this is unknown. Similar estimates were found in previous RR analyses of a smaller, less severely edited set of Angus data, both for analyses fitting a RR on Legendre polynomial (Meyer 2005b) and B-spline basis functions (Meyer 2005c). Moreover, estimates were consistently higher than those

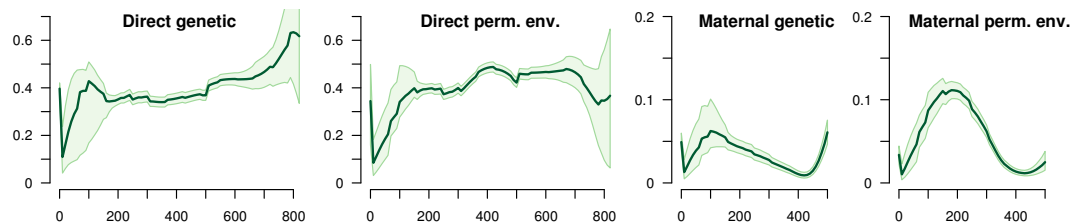


Figure 3. Estimates of variance ratios together with 95% HPD regions (shaded area).

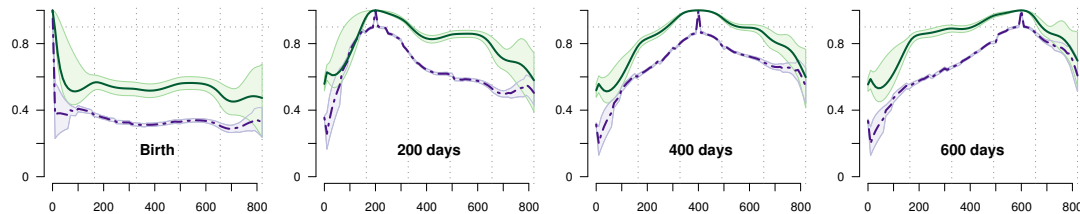


Figure 4. Estimates of genetic (solid line) and phenotypic (dashed line) correlations.

obtained for weaning weight from univariate analyses of the same data. In contrast, Meyer (2002a) reported close agreement of estimates for heritabilities at 200 days between RR and univariate analyses of Hereford data.

Conversely, estimates of maternal heritabilities obtained previously (Meyer 2005b) were lower than expected, and the same is observed here with an estimate of only 4.3% (95% HPD : 3.5 – 5.1%) for 200 day weight. This indicates a different partitioning between direct and maternal genetic effects in RR analyses than in standard, uni- or multivariate analyses. A sharp rise in maternal heritability estimates after 450 days of age is inexplicable and implausible, and shows that B-spline functions are not a panacea for all ‘end-of-range’ problems in RR analyses. In contrast, estimates of the proportion of variance due to permanent environmental effects showed little evidence of such problems, and the estimate of 11.1% (95% HPD : 10.1 – 12.0%) for 200 days is in good agreement with literature values.

Figure 4 displays estimates of direct genetic and phenotypic correlations between weights at birth, 200, 400 and 600 days and all other weights. Vertical lines show the position of knots. Again slight ‘knot effects’ can be observed for genetic correlations. Overall, estimates were similar to previous results for other breeds (Albuquerque and Meyer 2001; Nobre et al. 2003; Meyer 2002a), showing a genetic correlation close to 0.5 between birth and later weights, and genetic correlations above 0.9 (dotted horizontal line), among records at ages taken up to 190 days apart.

CONCLUSIONS

Results show that B-spline functions are well suited to RR analyses of sparse growth data, yielding, overall, plausible and consistent estimates. Further work is required to examine the scope for ‘clamping’ maternal trajectories at zero at the upper limit of the ages for which they are considered valid, and to investigate the mechanisms affecting partitioning of direct and maternal variation.

REFERENCES

- Albuquerque, L. G. and Meyer, K. (2001) *J. Anim. Sci.* **79**:2776.
- Koots, K. R., Gibson, J. P., Smith, C. and Wilton, J. W. (1994) *Anim. Breed. Abstr.* **62**:309.
- Meyer, K. (2002a) *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.* Comm. No. 11–01.
- Meyer, K. (2002b) *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.* Comm. No. 28–27.
- Meyer, K. (2005a) *Austr. J. Exp. Agric.* **45**:000 (in press).
- Meyer, K. (2005b) *J. Anim. Breed. Genet.* **122**:73.
- Meyer, K. (2005c) *Genet. Select. Evol.* **37**:000 (in press).
- Nobre, P. C. R., Misztal, I., Tsuruta, S., Bertrand, J. K., Silva, L. O. C. and Lopes, P. S. (2003) *J. Anim. Sci.* **81**:918.
- Rice, J. A. and Wu, C. O. (2001) *Biometrics* **57**:253.