

ORIGINAL ARTICLE

Estimates of genetic covariance functions for growth of Angus cattle

K. Meyer

Animal Genetics and Breeding Unit (A joint venture with NSW Agriculture), University of New England, Armidale NSW, Australia

Correspondence

Karin Meyer, Animal Genetics and Breeding Unit (A joint venture with NSW Agriculture), University of New England, Armidale NSW 2351, Australia. Tel.: +61 2 6773 3331; Fax: +61 2 6773 3266; E-mail: kmeyer@didgeridoo.une.edu.au

Received: 13 July 2004;
accepted: 8 November 2004

Summary

Estimates of covariance functions and genetic parameters were obtained for growth of Angus cattle from birth to 820 days of age. Data comprised 84 533 records on 20 731 animals in 43 herds, with a high proportion of animals with 4 or more weights recorded. Changes in weights were modelled through random regression on orthogonal polynomials of age at recording. A total of 11 combinations of quadratic, cubic, quartic and quintic polynomials to model direct and maternal genetic effects and permanent environmental effects were considered. Results showed good agreement for all models at ages with many records, but differed at the highest ages and at very early ages with few weights available. Cubic polynomials appeared to be most problematic. The order of polynomial fit for permanent environmental effects of the animal dominated estimates of phenotypic variances and mean squares for residual errors. A model fitting a quartic polynomial for these effects and quadratic polynomials for the other random effects, appeared to be the best compromise between detailedness of the model which could be supported by the data, plausibility of results, and fit, measured as mean square error.

Introduction

Random regression (RR) analyses have become a standard procedure for the genetic analysis of serially recorded traits from livestock improvement schemes. Whilst a large proportion of applications considered test day records of dairy cows, analyses of growth records from meat producing animals are becoming more popular; see Schaeffer (2004) for a recent review. The majority of analyses fitted polynomial functions of time or age at recording as covariables. In particular, Legendre polynomials – as used by Kirkpatrick *et al.* (1990) to estimate genetic covariance functions – have found widespread use. A number of RR analyses of growth records for beef cattle have been reported, mostly fitting a polynomial regression model. Some were limited to the analysis of cow weights (Meyer 1999, 2000; Arango

et al. 2004). Others considered records from birth to weaning or approximately 2 years of age (Meyer 1999, 2001, 2002a; Albuquerque & Meyer 2001a; Nobre *et al.* 2003).

Previous analyses of larger data sets, reported by Nobre *et al.* (2003) and Meyer (2002a), considered a single model of analysis only, fitting cubic regressions on age at recording for all random effects, i.e. direct and maternal, genetic and permanent environmental effects. On the one hand, this choice was a compromise between computational demands and detailedness of the model (Nobre *et al.* 2003). On the other hand, a cubic polynomial is the lowest degree polynomial for which both first and second derivatives, i.e. growth rate and acceleration, vary with age (Meyer 2002a). In addition, the cubic RR model involves almost the same numbers of genetic coefficients as its multivariate counterpart, which treats

birth, weaning, yearling and final weights as four different traits. Lastly, the number of records for typical field data in beef cattle is limited, which restricts the degree of polynomials which can be estimated. Until now, only results from analyses of small data sets have been available for comparison of different orders of polynomial fit. Those analyses suggested use of models with differential orders of fit for individual random effects and a higher degree polynomial for direct permanent environmental effects (Albuquerque & Meyer 2001a; Meyer 2001).

This paper presents estimates of covariance functions for growth of Angus cattle, from analyses of a moderately sized set of field data, selected to have a high portion of animals with four or more records. Analyses examine the effects of different orders of polynomial fit, considering a total of 11 different models.

Materials and methods

Data

Data were records for weights of Australian and New Zealand Angus cattle, extracted from the National Beef Recording Scheme data base in July 2003. Raw data consisted of approximately 1.9 million weight records and pedigree information on close to a million animals. Weights taken from birth to 820 days of age were considered. After basic edits, records were extracted for animals in 43 herds with the highest proportions of animals with multiple weights recorded. Herds selected were those with an average number of records per animal >3.3 , or at least 50% animals with four or more records. Based on inspection of growth rates, further edits eliminated any records clearly 'out-of sequence' for individual animals. For animals flagged as the result of an embryo transfer, identities of recipient or foster dams were determined where possible. If this cow could not be identified, such calves were assigned a dummy 'raising' dam, which was assumed to have only a single offspring. Similarly, a small number of animals with unknown dams were given a dummy, genetic dam.

Recursive deletion of records on animals with less than three observations and records in single record subclasses, left 84 533 records on 20 731 animals. Animals were progeny of 1348 sires and 8167 dams. Figure 1 shows the distribution of records over ages at recording in 10-day intervals, together with corresponding mean weights. Not shown are 17 891 records at age 0, for the 86.3% of animals which had a birth weight record. Current genetic evaluation of beef cattle in Australia considers weights at

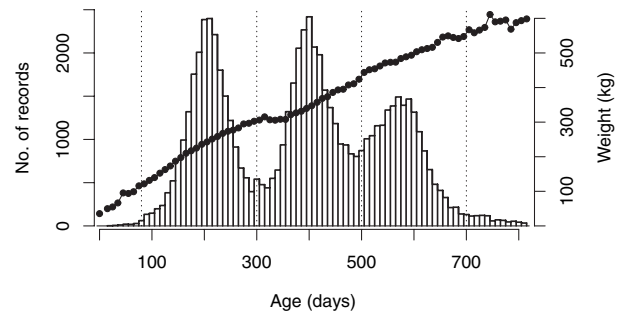


Figure 1 Numbers of records^a (□) and means of observed weights (●) for ages at recording in 10-day intervals. ^aomitting 17 891 birth weight records.

target ages of 200, 400 and 600 days with permissible age ranges of 80–300 days, 301–500 and 501–700 days, respectively. The distribution of records over ages of recording clearly reflects this régime, with very few weights recorded between birth and 80 days of age, and limited numbers of weights available after 700 days.

On average there were 4.08 records per animal. The maximum number of weights recorded per individual was 9. There were 6027, 8963, 4231, 1189 and 321 animals with 3, 4, 5, 6 and 7–9 records, respectively. This resulted in 70.9, 27.7 and 7.3% of animals with at least 4, 5 and 6 weights recorded. At a genetic level, records on parents, progeny and sibs contribute information through the relationship matrix. Counting these as well, corresponding proportions were 98.3, 85.2 and 48.1% of all animals.

Pedigree information for animals with records and their parents was extracted, considering up to five generations backwards. Prior to analyses, pedigrees were 'pruned', i.e. any parents without records and with only a single offspring were treated as unknown. This was performed recursively, resulting in 14 326 parents without records to be taken into account, i.e. direct genetic effects for a total of 35 057 animals to be included in the analysis.

Pedigrees for maternal genetic effects were 'pruned' separately, eliminating any animals without an expression of maternal genetic effects and links to only one other animal. Again, this was performed recursively. This reduced the number of animals in the maternal pedigree and relationship matrix to 20 463. There were 9670 dams with permanent environmental effects to be considered. Of these, 5202 (including 1749 dummy dams) had only one offspring, while 1781, 1012, 680, 376 and 619 cows had 2, 3, 4, 5 and 6 or more, calves with records, respectively.

Analyses

Estimates of (co)variance components were obtained by Bayesian analysis, employing a Gibbs sampling algorithm as implemented in program RRGIBBS (Meyer 2002b). Three Markov chains with 100 000 samples each were drawn for each model, assuming flat priors. Variances, genetic parameters and eigenfunctions for ages in 10-day intervals were calculated for every fourth sample. Estimates were obtained as posterior means across all three chains, disregarding the first 20 000 samples in each chain as 'burn-in' period. Approximate piece-wise, 95% confidence regions for each parameter examined were obtained by ranking samples and determining the cut-off points for the lowest and highest 2.5% of sample values (Chen & Shao 1999). Solutions for all effects in the mixed model were obtained, using the estimates (posterior means) of covariance components for each model. Residuals, i.e. observations adjusted for estimates of the effects affecting each record, and their sums of squares and corresponding mean square errors (MSE) were calculated.

Random effects

The model of analysis fitted a RR on Legendre polynomials of age at recording for four random effects, considering various orders of fit (k). Direct, additive genetic effects (A) were, in turn, modelled through quadratic, cubic and quartic polynomials, involving $k_A = 3, 4$ and 5 RR coefficients, respectively. Direct permanent environmental effects (R) were modelled analogously, considering a quintic polynomial in addition, i.e. $k_R = 3, 4, 5, 6$. Maternal genetic (M) and permanent environmental (C) effects were modelled through either quadratic or cubic polynomials, i.e. $k_M = k_C = 3, 4$. To avoid problems with spurious estimates at the highest ages, their effects were restricted to ages from 0 to 600 days, except for one analysis. This was implemented by setting functions of age at recording, i.e. the covariables in RRs for maternal effects, after 600 days to zero. Direct and maternal genetic effects were assumed to be uncorrelated. This is standard practice in analyses of Australian field data, as cows often remain in their mating groups which can lead to a confounding of sires and contemporary groups (CG), especially if management group recording is poor. This has resulted in implausibly large, negative estimates of direct-maternal genetic correlations, accompanied by inflated estimates of direct and maternal heritabilities, even for simple univariate analyses (Meyer 1997).

Models fitted are denoted in the following by the number of RR coefficients for A, M, R and C , respectively. For example, M4353 describes an analysis fitting quadratic polynomials for both maternal effects, and cubic and quartic polynomials for direct genetic and permanent environmental effects, respectively. A total of 11 combinations of orders of fit, listed in Table 1, were examined. Residuals were considered independently distributed with heterogeneous variances (σ_i^2). Changes in σ_i^2 with age were modelled as a 'step' function with 19 classes (0, 1–30, ..., 271–300, 301–360, ..., 721–780 and 781–820 days) throughout. There were a total of 43 (M3333) to 69 (M5454) covariance components to be estimated for the different combinations.

Fixed effects

Changes in mean with age were modelled through fixed, cubic regressions on Legendre polynomials of age. This involved a separate regression for each level of each of the fixed effects, namely sex, dam age class (in years, treating ages >9 years as one class) and birth type (single versus twin). In addition, CG were fitted as cross-classified fixed effects, with CG defined as herd-sex-management group-year/month subclasses for birth weights, and herd-sex-management group-date of weighing subclasses for other weights. To reduce the range of ages compared directly, CG classes were further subdivided applying an 'age slicing' of 45 days up to 300 days, and 60 days for higher ages. If this subdivision resulted in a subclass with less than five 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 for later weights. In total, there were 8,822 CG levels (after age slicing).

Table 1 Models fitted together with the number of covariance components estimated (p) and residual mean square errors (MSE)

Model	p	MSE
M3333	43	53.6
M5333	52	34.3
M3343	47	36.9
M4343	51	36.4
M4444	59	35.4
M3353	52	30.1
M4353	56	29.6
M5353	61	29.6
M5454	69	28.7
M3363	58	24.7
M5363	67	24.5

Univariate analyses

In addition, eight corresponding univariate analyses were carried out by restricted maximum likelihood, considering ages of 0, 1–200, 101–300, ... and 601–820 days. These considered single records per animal only, fitting direct and maternal genetic effects and maternal permanent environmental effects. Fixed effects considered were CGs, as defined above, birth type and dam age class (heifer versus cow). Dam age was fitted as a fixed, linear and quadratic covariable, and age at recording was accounted for by fitting a linear regression within each sex of animal.

Results

Variance components

Estimates of phenotypic variances (σ_p^2) for the range of ages considered are shown in Figure 2. Variances increased with age and mean. Overall, there was excellent agreement between estimates from different RR models for ages with large numbers of observations, in particular birth weight and the intermediate ages between about 150 and 650 days. As observed in other studies (e.g. Meyer 2001, 2002a; Nobre *et al.* 2003), RR model estimates were somewhat higher than estimates from univariate analyses, increasingly so with increasing age.

Estimates for early ages (after birth) and the highest ages, however, fluctuated considerably. Estimates for the highest ages appeared to be dominated by the order of polynomial fit for permanent environmental effects of the animal, with comparatively little effects associated with the orders of fit for the other random effects. Fitting a cubic regression for this effect ($k_R = 4$), estimates of variance functions increased sharply after about 650 days of age. With

k_R increased to 5 and 6, estimates were reduced to more plausible levels. Estimates of σ_p^2 at the highest ages were smallest for model M3353.

The effects of order of polynomial fit are further illustrated in Figure 3, which presents corresponding results for direct genetic (σ_A^2) and permanent environmental (σ_R^2) variances. Whilst differences in estimates between analyses were again small for ages with many records, effects of the RR model chosen were more prominent for other ages. As above, fitting a cubic polynomial resulted in implausible estimates of variances at the highest ages. For example, at 820 days, estimates of σ_A^2 for $k_A = 4$ rose to 3200 kg² for M4343, and estimates of σ_R^2 for $k_R = 4$ topped 4239 kg² for M3343 (Figure 3 truncated, i.e. not showing these high values).

Increasing k_A to 5 yielded more plausible estimates of σ_A^2 , but a very steep increase at the extreme ages was still evident. Moreover, estimates of σ_A^2 for early ages for a model fitting a quadratic regression for all other effects (M5333) deviated markedly from those obtained from other analyses. Overall, a quadratic polynomial ($k_A = 3$) appeared to be the most sensible choice to model changes in σ_A^2 with age. Except for birth weight and a univariate analysis considering weights after birth to 200 days, estimates of σ_A^2 from RR model analyses were consistently higher than estimates from corresponding univariate analyses.

Similarly, RR estimates of σ_R^2 at the higher ages were smaller for higher orders of fit than for low values of k_R . Variance function estimates for $k_R = 6$ showed oscillations after 600 days, indicating problems with polynomial approximation of growth trajectories. Either $k_R = 5$ or, more conservatively, $k_R = 3$ appeared reasonable choices, although esti-

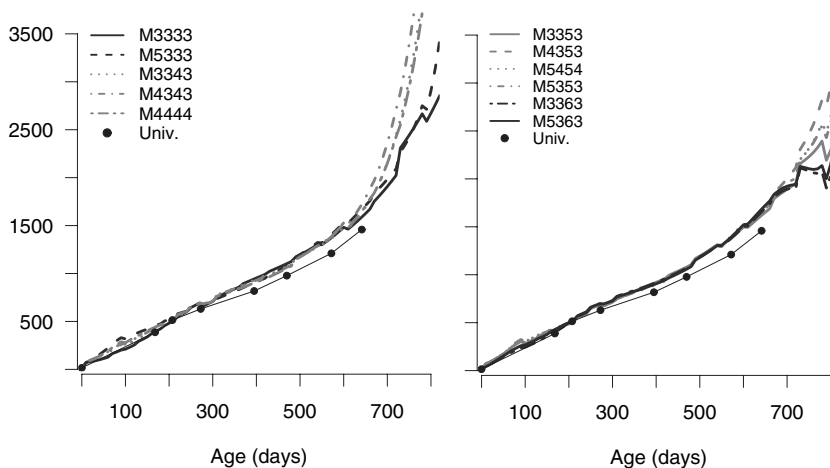


Figure 2 Estimates of phenotypic variances (kg²) from analyses fitting models with different orders of polynomials.

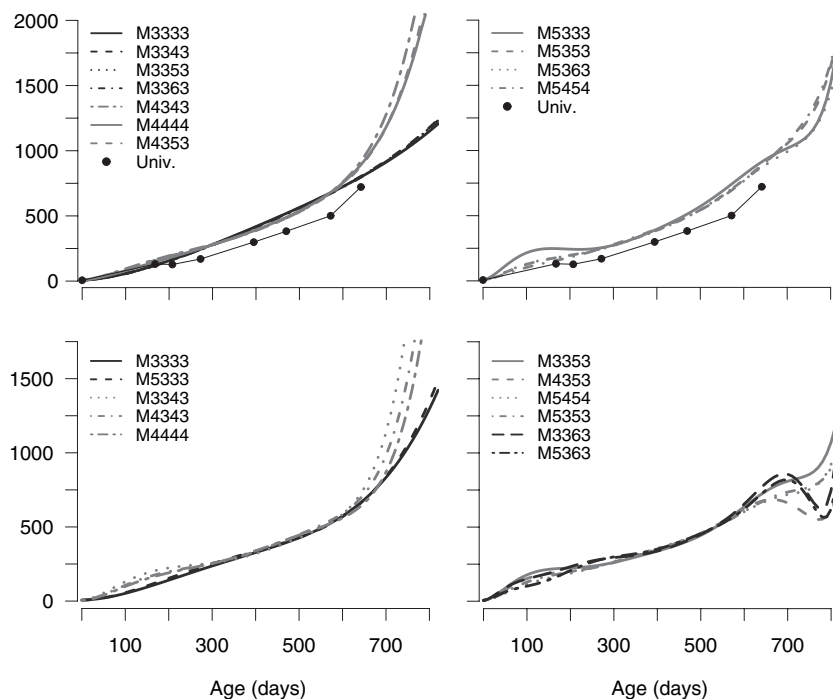


Figure 3 Estimates of direct genetic (top row) and permanent environmental (bottom row) variances (kg^2) from analyses fitting models with different orders of polynomials.

mates of σ_R^2 at early ages (except birth weight) differed markedly for these alternatives. Loosely described, the higher order polynomials had 'more room to move' and thus were more likely to produce small oscillations, where there were few data points to fix estimated growth curves. At 100 days, for example, estimates of σ_R^2 were 178 and 52 kg^2 for analyses M3353 and M3333, respectively. After 100 days, estimates began to converge, but did not coincide fully until about 300 days of age. Furthermore, confidence regions (not shown) for σ_R^2 after 650 days were substantially greater for $k_R = 5$ than for $k_R = 3$, generally encompassing estimates for the latter. For σ_p^2 , however, differences were much less pronounced with respective estimates at 100 days of 279 (M3353) and 210 (M3333) kg^2 , and with estimates being essentially identical by 180 days of age. This suggests differential partitioning in the corresponding maternal and residual variances.

The MSE for all models are summarized in Table 1. As for σ_R^2 and σ_p^2 these depended almost entirely on k_R rather than on the total number of parameters fitted. Only for model M5333, compared with M3333, was there a marked decrease in MSE when k_R was constant but k_A was increased. With the same number of parameters fitted, M3353 had a substantially lower MSE than M5333, indicating that unexplained variation because of permanent environmental effects for $k_R = 3$ might have been

attributed to the higher order genetic polynomials in model M5333.

Estimates of maternal genetic (σ_M^2) and permanent environmental (σ_C^2) variances for selected analyses are shown in Figure 4. Again, there was a distinct effect of order of polynomial fit for the random effect considered, with minor differences due to the model for other effects. For quadratic polynomials, estimated variances were greatest for weights at about 300 days, i.e. mid-way for the range of ages considered. As above, fitting a cubic polynomial caused estimates of variances to increase sharply at the high end of the range of ages, in particular for σ_C^2 . For $k_M = 4$ or $k_C = 4$, highest estimates of maternal variances occurred close to about 200 days. Estimates of σ_C^2 for $k_C = 4$ agreed more closely with univariate results than for $k_C = 3$.

An additional analysis assumed no influence of maternal effects after 500 days for model M4353. Results (not shown) yielded variance functions of similar shape as for M4353 for both σ_M^2 and σ_C^2 , but with the highest estimates occurring around 250 days instead of 300 days. This comparison demonstrates an effect of the range of ages considered, and emphasizes the need to choose this range carefully.

For M3333, estimates of σ_C^2 increased at the highest ages, similar to those for $k_C = 4$. Estimates of σ_C^2 for M3343 (not shown) and M3353 did not, suggesting

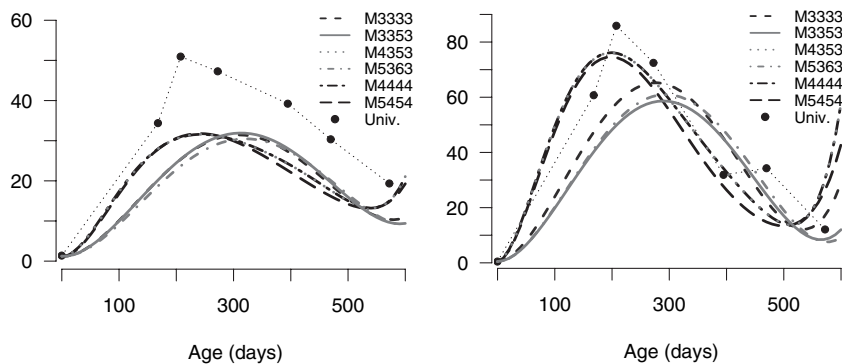


Figure 4 Estimates of maternal genetic (left graph) and permanent environmental (right graph) variances (kg^2) from selected analyses, fitting models with different orders of polynomials.

that this erratic increase might, in part at least, be due to picking up some direct permanent environmental variation which had not been modelled by the low degree polynomial for the corresponding random effect.

Estimates of σ_M^2 were considerably smaller than those from univariate analyses. The reverse was discussed for σ_A^2 . Hence, the latter could, to some extent, reflect differences in the partitioning of direct and maternal variation between the two types of analyses, as different contrasts between relatives were available. With maternal effects and their variances inherently difficult to estimate, this result is perhaps not too surprising. In contrast, Albuquerque & Meyer (2001a) reported consistently slightly higher estimates of σ_M^2 from RR than univariate analyses, whilst Meyer (2001) found good agreement between the two. Nobre *et al.* (2003) obtained markedly higher estimates of both σ_A^2 and σ_M^2 from RR analyses than from multivariate analyses. However, their model allowed for non-zero direct-maternal genetic covariances, and these covariances also differed dramatically between analyses, being negative and large (absolute value) for RR analyses. If estimates of these covariances are biased downwards, they can yield correspondingly inflated estimates of σ_A^2 and σ_M^2 (Meyer 1997). Such a mechanism might have contributed to the discrepancies between analyses encountered by Nobre *et al.* (2003).

Heritabilities and other variance ratios

Estimates of direct heritabilities (h^2) and their approximate confidence regions at selected target ages are contrasted in Figure 5. Differences in estimates directly reflect the pattern in estimates of variances observed above. Whilst all estimates for birth weight were essentially the same, there was a wide range of estimates for h^2 at 100 days of age, attributable to the lack of data at this age and, conse-

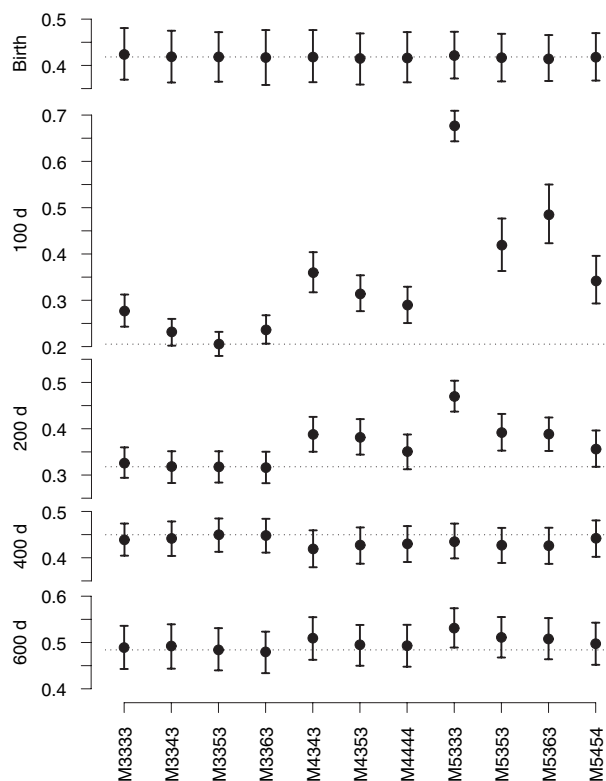


Figure 5 Estimates of heritabilities (●) together with approximate 95% highest posterior density regions (bars) from analyses fitting models with different orders of polynomials, for weights at birth, 100 days, 200 days, 400 days, and 600 days^a. ^aDotted lines give estimates for model M3353.

quently, the dominating effect of the model of analysis. Early growth in beef cattle has received little attention. Albuquerque & Meyer (2001b) reported a univariate estimate of h^2 for weights of Nelore cattle at 100 days of about 0.16, which was lower than any values obtained in this study. Nelore are a later maturing breed than Angus, which might have contributed to this difference. The effects of different polynomial interpolation were still evident at

200 days. With higher estimates of σ_A^2 from RR analyses, corresponding values for h^2 were also consistently greater than the corresponding univariate estimate of 0.24. At 400 and 600 days, estimates from different models again agreed closely, but, similarly, were higher than the respective univariate estimates of 0.37 and 0.41.

Similar relationships held for estimates of maternal heritabilities (m^2) and the proportions of variance because of direct (p^2) and maternal (c^2) permanent environmental effects. Changes in estimates of variance ratios with age for selected analyses are shown in Figure 6, together with approximate 95% credible intervals for model M3353. With a quadratic polynomial fit for direct genetic effects ($k_A = 3$), estimates of h^2 declined sharply after birth and increased slowly subsequently. A fairly similar pattern was observed for data sets with substantial numbers of observations available at early ages, by both Albuquerque & Meyer (2001a) and Meyer (2001). For k_A increased to 5, estimates for this period were substantially higher, but estimates of p^2 were reduced accordingly, indicating a strong sampling correlation between h^2 and p^2 .

With estimates of maternal variances from RR analyses less than from univariate analyses, estimates of m^2 and c^2 for $k_C = 3$ were substantially less as well. Fitting a quadratic polynomial for maternal environmental effects, estimates for c^2 were highest for analysis M3333, which suggests that smaller esti-

mates of σ_C^2 and c^2 for $k_R > 3$ might have contributed to high estimates of σ_R^2 and p^2 .

Correlations

The order of polynomial fit for a random effect also dominated the estimates of its correlation structure, as illustrated in Figure 7 for the direct genetic correlations (r_A) for weight at 200 days with weights at all other ages. For $k_A = 3$, estimates from all four analyses considering this order of fit were essentially identical. The correlation curve was very flat and indicated a broad range of ages for which genetic correlations were very high, i.e. 0.9 or more. With k_A increased to 4 or 5, the width of this interval decreased, and allowed for more fluctuations because of the order of fit for other random effects. Estimates for M5333 were erratic, presumably due to permanent environmental variation not accounted for with $k_R = 3$.

Figure 8 summarizes estimates of direct genetic and phenotypic (r_P) correlations for weights at birth, 200, 400 and 600 days, with weights at all other ages. Estimates and approximate confidence regions given are for model M3353. In addition, estimates of r_P for M3333 and r_A for M5353 are shown. On the whole, estimates agreed with expectations of moderate correlations between birth and later weights and decreasing correlations between other weights with increasing difference in the ages at recording. Whilst

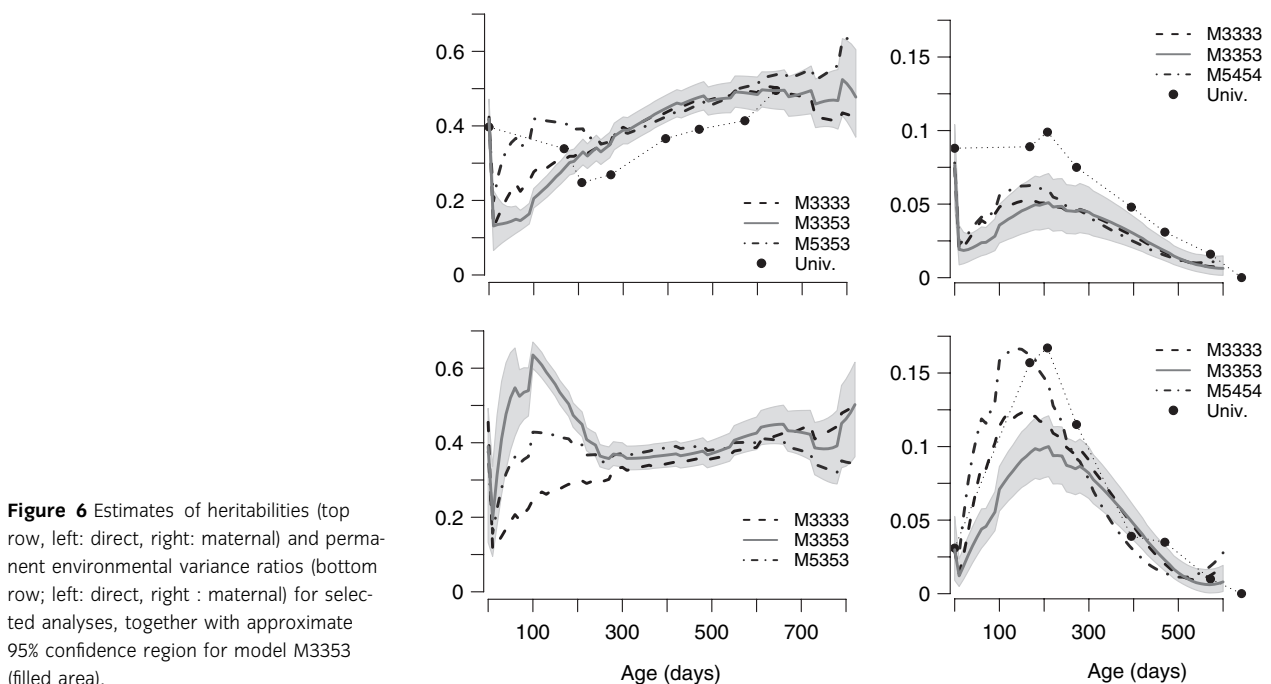


Figure 6 Estimates of heritabilities (top row, left: direct, right: maternal) and permanent environmental variance ratios (bottom row; left: direct, right: maternal) for selected analyses, together with approximate 95% confidence region for model M3353 (filled area).

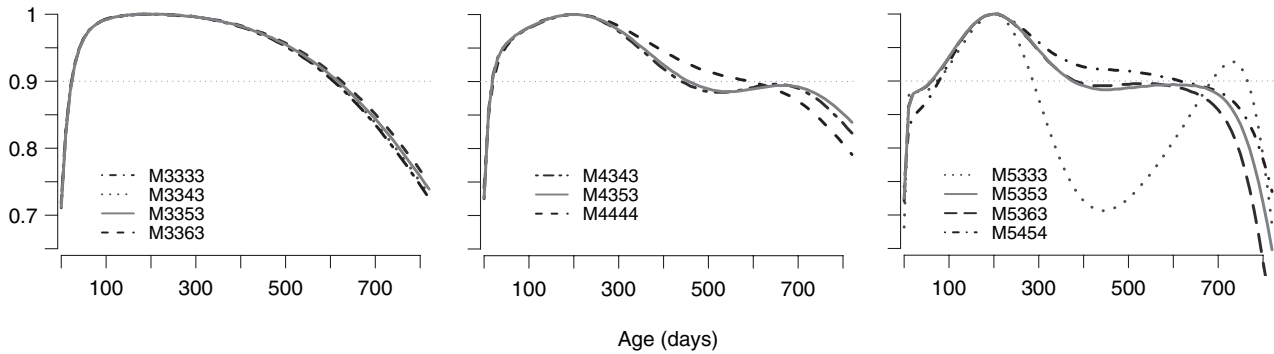


Figure 7 Estimates of direct genetic correlations for weight at 200 days with weights from birth to 820 days, from analyses fitting quadratic ($k_A = 3$, left), cubic ($k_A = 4$, middle) or quartic ($k_A = 5$, right) polynomials for direct genetic effects.

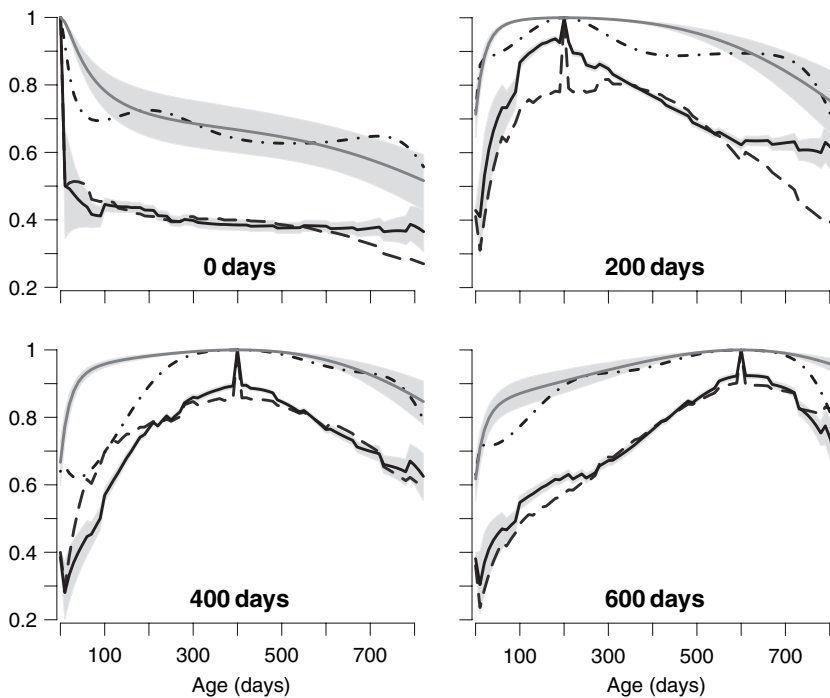


Figure 8 Estimates of direct genetic (top curve) and phenotypic (lower curves) correlations, together with approximate confidence intervals (filled area), for weight at 0 days (top left), 200 days (top right), 400 days (bottom left) and 600 days (bottom right) with weights from birth to 820 days, for analyses fitting model M3353 (solid line), M3333 (dashed line; phenotypic correlation only) and M5353 (dotted line; genetic correlation only).

confidence regions for alternative models, M5353 for r_A and M3333 for r_P (not shown), did not fully overlap with those for M3353, differences in estimates were insubstantial, especially for ages with many records.

Corresponding contour plots of direct genetic and permanent environmental correlations are shown in Figure 9. Clearly, fitting quadratic polynomials enforced a fairly regular decline in correlations with increasing distance in time. Conversely, the additional terms in a quartic polynomial allowed for more changes in estimates of correlations, especially to lower correlations between weights pre- and post-weaning. Fitting quadratic polynomials, estimates of

maternal correlations, both genetic and permanent environmental, were near unity (≥ 0.95) for all ages between 50 and 500 days, but showed steep declines towards birth weight.

Eigenvalues and eigenfunctions

Estimates of the first three eigenvalues of the direct genetic and permanent environmental covariance functions and their approximate confidence regions are summarized in Figure 10. Again, a distinct effect of the order of polynomial fit was evident, with values for a cubic polynomial consistently different. Regardless of the model, the last eigenvalue for each

Figure 9 Contour plots for estimates of direct genetic (r_A ; left) and permanent environmental (r_R ; right) correlations, for analyses fitting a quadratic (top triangle; analyses M3353 for r_A and M3333 for r_R) or quartic (bottom triangle; M5353 for r_A ; and M3353 for r_R) polynomial; contour lines given in steps of 0.05 [solid : 1.0 (diagonal), 0.9, 0.8, ..., dashed: 0.95, 0.85, ...].

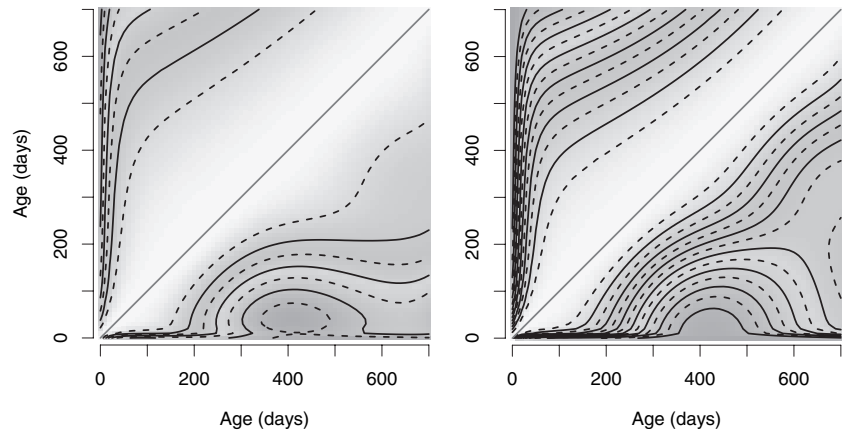
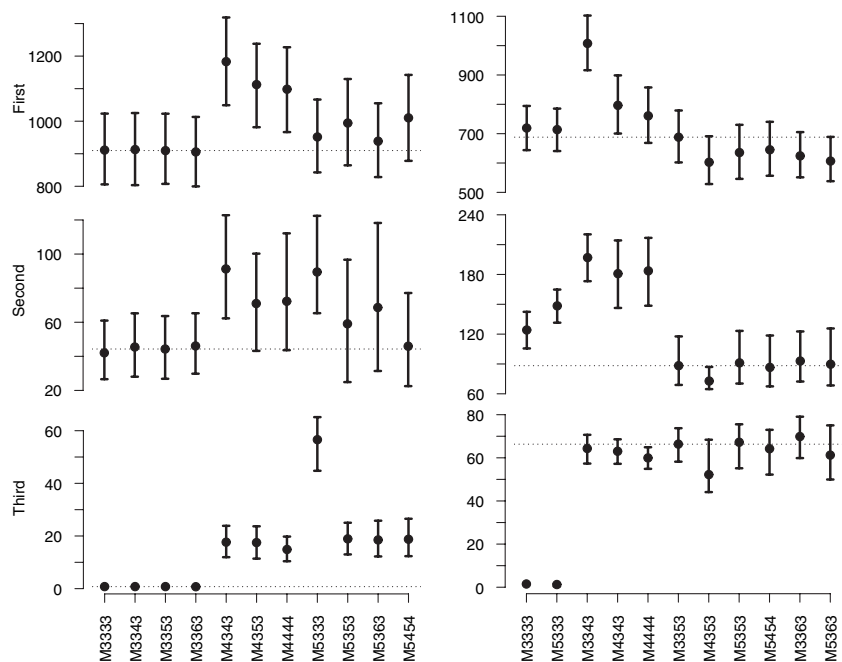


Figure 10 Estimates of the first three eigenvalues (●), together with approximate 95% confidence intervals (bars), for direct genetic (left) and permanent environmental (right) covariance functions^a. ^aDotted lines mark estimates for model M3353.



covariance function estimated was close to zero, which explains the big difference in estimates for the third eigenvalue for quadratic and higher order polynomials. The first genetic eigenvalue explained 92% ($k_A = 5$) to 95% ($k_A = 3$) of the total genetic variation. For the permanent environmental covariance functions, ratios were about 84, 78 and 74% for $k_R = 3, 5$ and 6, respectively.

Figure 11 displays estimates of the first and second direct genetic eigenfunctions for different orders of fit. Estimated confidence regions when fitting a quadratic or cubic polynomial were very narrow, suggesting that the eigenstructure of a covariance function can be estimated more accurately than its coefficient matrix. In addition, the eigenfunctions

obtained from a decomposition of the posterior mean of the covariance matrix among RR coefficients were essentially identical to the posterior mean of the eigenfunctions shown. Again, estimates for $k_A = 3$ and $k_A = 5$ were similar, with the former more or less a smoothed version of the latter.

However, when fitting a quartic polynomial for direct genetic effects, estimated confidence regions were large and irregular. This occurred for all three analyses with $k_A = 5$ and $k_R \geq 5$, i.e. was unlikely to reflect failure of a Markov Chain to fully converge to the posterior distribution. Whilst only 28% of animals had five weights recorded themselves, 85% of animals had five records at the genetic level, when considering records on close relatives. Nevertheless,

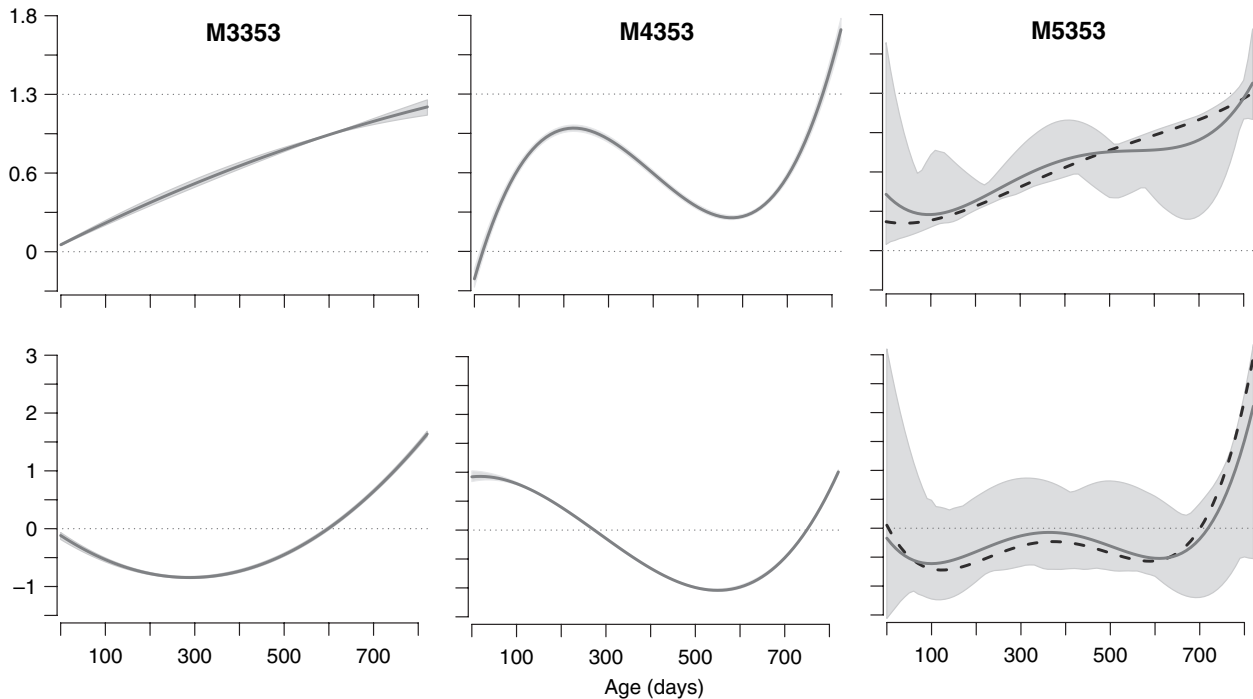


Figure 11 Estimates of the first (top row) and second (bottom row) direct genetic eigenfunctions (solid line), together with approximate 95% confidence region (filled area) and corresponding function obtained from the estimate of the genetic covariance function (dashed line), from analyses fitting a quadratic (left), cubic (middle) or quartic (right) polynomial for direct genetic effects.

this might not have been sufficient to allow quartic polynomials to be fitted at both the genetic and environmental level. Results for analysis M5454 and M5363 were similar to those shown for M5353. Results for M5333, however, were very different, with the estimate of the first eigenfunction similar to that when fitting a cubic RR for direct genetic effects (M4353), and a narrow confidence region. This model was fitted to assess the impact of fitting two quartic polynomials in analysis M5353. Whilst lower confidence regions for this analysis might support the surmise of overparameterization for models M5353, M5454 or M5363, as also noted above, results for model M5333 were highly irregular and thus might not provide a valid comparison.

For both $k_A = 3$ and $k_A = 5$, estimates of the first genetic eigenfunction were positive throughout, i.e. selecting on the basis of the corresponding estimated breeding values is likely to change weights at all ages in the same direction. Moreover, with more than 90% of the total genetic variation attributable to this function, this is the change which is most easily affected. In contrast, the second eigenfunction was close to zero at birth, negative to about 600 days and positive for the highest ages, but explained only about 5% of genetic variation.

Discussion

Results clearly illustrate the effects of the model chosen on estimates from RR analyses. Reassuringly, estimates from different models agreed closely for ages with many observations. Not surprisingly though, for ages with few records the shape of the curve chosen to model growth trajectories dominated estimates. The previously preferred model fitting a RR on cubic polynomials on age at recording (Meyer 2002a) turned out to be an unfortunate choice, which yielded grossly inflated estimates of variances at the highest ages, both for their data and in this study, and implausible estimates of the leading genetic eigenfunctions. Choice of the 'best' model inevitably is a compromise between the features we might want to capture and the amount of information available in the data.

A quadratic polynomial to model variation of individual growth trajectories about the mean would imply that only growth rates or velocity, i.e. first derivatives of the growth curve, varied with age, but not second derivatives, i.e. acceleration. Whilst somewhat restrictive in terms of correlation structure and associated with the highest residual error mean square, a model fitting quadratic polynomials for all

random effects (M3333) gave sensible estimates of variances, heritabilities and eigenfunctions, and would seem to be a conservative choice for the data considered in this study. Fitting a quartic polynomial for some effects would be less restrictive and allow for more fluctuations in correlation estimates. Fitting this for permanent environmental effects of the animal only (M3353) yielded similar estimates to those for an entirely quadratic model, but had considerably lower MSE, and thus would be a judicious choice. Estimates of covariance functions for this model are given in the Appendix. Conversely, fitting a quartic regression for direct genetic effects only (M5333) produced erratic estimates for heritabilities, genetic correlations and eigenfunctions, suggesting that some environmental effects might have affected the genetic covariance function. A model with quartic polynomials for both the genetic and permanent environmental effects of the animal (M5353) exhibited some problems with inflated estimates of variances for the highest ages and large confidence intervals for estimates of genetic eigenfunctions. Furthermore, it did not reduce MSE greatly over the MSE for model M3353, and thus does not appear advantageous.

Models were compared on the basis of estimated variances and genetic parameters only. More formal test criteria exist, but were not calculated. Examples are the deviance information criterion (Spiegelhalter *et al.* 2002), a Bayesian analogue of the information criteria used in a maximum likelihood framework of inference, or Bayes factors (Kaas & Raftery 1995). However, as demonstrated by Jamrozik & Schaeffer (2002) for different RR models for test-day records of dairy cows, different criteria can rank models differently, and it is not clear which statistic is most appropriate for the kind of models considered in this study.

Discrepancies in estimates of genetic variances and heritabilities from RR and from univariate analyses are of concern. Clearly, the data utilized in this study represent a selected sample. In some analyses, higher estimates for later ages from RR analyses have been attributed to removal of selection bias when considering all records simultaneously. However, for our data with many animals recorded until almost 2 years of age, this is not a likely explanation, especially for estimates for ages prior to weaning. Whilst RR analyses use all records available and thus different and possibly many more covariances among relatives than univariate analyses, it is not obvious that they should identify more genetic variation between animals. Analyses fitting RR tend to be

more sensitive to systematic differences not fully modelled than univariate analyses. However, the fixed effect part of the model differed only slightly between analyses, with RR analyses fitting a cubic, orthogonal regression on age at recording while univariate analyses employed a simple linear regression.

Field records of weights of beef cattle have proved inherently difficult in attempts to estimate covariance functions for growth. Generally, records are sparsely distributed over ages at recording. Even for our data set, comprising only weights from a small proportion of herds with comparatively many records, few animals had more than four weights recorded. In addition, records were distributed rather unevenly, with few calves weighed between birth and weaning, resulting in a large cluster of records at the lowest age followed by a period of essentially no recording, and slowly decreasing numbers at the highest ages. This imbalance may well have exacerbated problems of erratic estimates encountered with some models. Nobre *et al.* (2003) found no dramatically inflated variances at the highest ages for a cubic RR model when restricting the analysis to animals without 'missing' records. Similarly, for test day records in dairy cattle, Pool & Meuwissen (2000) reported a reduction in the number of RR terms to be fitted and overall improved estimates, when using records for complete lactations only.

Estimates of covariance functions for data with a better structure, in particular with more weights recorded at early ages, exist. Examples are estimates for experimental data (Meyer 2001) or field data with frequent recording as for the Brazilian Nelore data (Albuquerque & Meyer 2001a; Nobre *et al.* 2003). Whilst variances often differ between populations, estimates of correlations generally vary comparatively little, with differences well within the limits of sampling errors. It might be advantageous to combine information from such analyses with information from field data. In a Bayesian framework of analysis, this could be performed by choosing appropriate priors. However, this might require alternative parameterization, for instance to variance functions and correlation functions rather than covariance functions, and other sampling schemes, and thus might not be straightforward.

Alternatively, we might construct a matrix of covariances for a grid of ages, covering the complete range of ages considered. This would use estimates of variances and correlations from our data for ages with many records, and combine these with literature estimates of correlations for ages with few records. If necessary, 'missing' estimates of variances

could be determined by interpolation (Misztal *et al.* 2000). A modified covariance function could then be estimated from this covariance matrix by least-squares, as suggested by Kirkpatrick *et al.* (1990). An example for the covariance matrix among test-day records of dairy cows is outlined by Misztal *et al.* (2000). Legarra *et al.* (2004) recently presented an application for weights of beef cattle. A more appealing approach might be the likelihood procedure proposed by Mäntysaari (1999). In essence, a grid of covariance estimates among selected ages is treated as if they were mean crossproducts arising from data, and equated to their expectations. The corresponding likelihood function is then readily maximized, using an expectation-maximization (Mäntysaari 1999; Koivula *et al.* 2004) or Method of Scoring (Thompson 1976) type algorithm, constraining the resulting covariance function estimates to be positive definite or have reduced rank, if required.

Spline functions have been advocated as alternative basis functions to orthogonal polynomials for RR analysis. Basically, these are piece-wise, polynomial functions, joined smoothly at selected points, the so-called knots (Dodge 2003). The pieces generally have low degree. In the simplest case, they can be linear. Thus numerical problems because of high orders of fit are reduced. Re-analyses of the data in this study are in progress, fitting a number of RR models with basis functions of B-splines of age at recording as covariables, as suggested by Rice & Wu (2001). Initial results are encouraging. Estimates for ages with many records were essentially the same as those obtained from analyses regressing on Legendre polynomials of age. In addition, the incidence of problems of inflated variance estimates at the highest ages was greatly reduced, and the leading eigenfunctions were estimated more consistently.

Conclusions

Random regression analyses accommodate changes in mean and variance of weights of cattle with age well, and are thus suited to a unified analysis of weights at all ages. However, with the current data structure for weights of beef cattle, the scope for estimating covariance functions for growth from field records is limited to relatively simple models.

Acknowledgements

This work was supported by grant BFGEN.100 of Meat and Livestock Australia Ltd (MLA). We are

indebted to the Australian Angus Society for the use of their data.

References

- Albuquerque L.G., Meyer K. (2001a) Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. *J. Anim. Sci.* **79**, 2776–2789.
- Albuquerque L.G., Meyer K. (2001b) Estimates of direct and maternal genetic effects for weights from birth to 600 days of age in Nelore cattle. *J. Anim. Breed. Genet.* **118**, 83–92.
- Arango J.A., Cundiff L.V., Van Vleck L.D. (2004) Covariance functions and random regression models for cow weight in beef cattle. *J. Anim. Sci.* **82**, 54–67.
- Chen M.-H., Shao Q.-M. (1999) Monte Carlo estimation of Bayesian credible and HPD intervals. *J. Comp. Graph. Stat.* **8**, 69–92.
- Dodge Y., ed. (2003) *The Oxford Dictionary of Statistical Terms*. Oxford University Press, Oxford and New York.
- Jamrozik J., Schaeffer L.R. (2002) Bayesian comparison of random regression models for test-day yields in dairy cattle. *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.* Communication No. 1-03.
- Kaas R.E., Raftery A.E. (1995) Bayes factors. *J. Amer. Stat. Ass.* **90**, 773–795.
- Kirkpatrick M., Lofsvold D., Bulmer M. (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* **124**, 979–993.
- Koivula M., Negussie E., Mäntysaari E.A. (2004) Genetic parameters for test-day somatic cell count at different lactation stages of Finnish dairy cattle. *Livest. Prod. Sci.* **90**, 145–157.
- Legarra A., Misztal I., Bertrand J.K. (2004) Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle. *J. Anim. Sci.* **82**, 1564–1571.
- Mäntysaari E.A. (1999) Derivation of multiple trait reduced random regression (RR) model for the first lactation test day records of milk, protein and fat. In: 50th Annual Meeting. Europ. Ass. Anim. Prod., mimeo, 8 pp.
- Meyer K. (1997) Estimates of genetic parameters for weaning weight of beef cattle accounting for direct-maternal covariances. *Livest. Prod. Sci.* **52**, 187–199.
- Meyer K. (1999) Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. *J. Anim. Breed. Genet.* **116**, 181–205.
- Meyer K. (2000) Random regressions to model phenotypic variation in monthly weights of Australian beef cows. *Livest. Prod. Sci.* **65**, 19–38.
- Meyer K. (2001) Estimates of direct and maternal covariance functions for growth of Australian beef calves from birth to weaning. *Genet. Select. Evol.* **33**, 487–514.

Meyer K. (2002a) Estimates of covariance functions for growth of Australian beef cattle from a large set of field data. *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.* Communication No. 11-01.

Meyer K. (2002b) 'RRGIBBS' – a program for simple random regression analyses via Gibbs sampling. *CD-ROM Seventh World Congr. Genet. Appl. Livest. Prod.* Communication No. 28-27.

Misztal I., Strabel T., Jamrozik J., Mäntysaari E.A., Meuwissen T.H.E. (2000) Strategies for estimating the parameters needed for different test-day models. *J. Dairy Sci.* **83**, 1125–1134.

Nobre P.C.R., Misztal I., Tsuruta S., Bertrand J.K., Silva L.O.C., Lopes P.S. (2003) Analyses of growth curves of Nelore cattle by multiple-trait and random regression models. *J. Anim. Sci.* **81**, 918–926.

Pool M.H., Meuwissen T.H.E. (2000) Reduction of the number of parameters needed for a polynomial random regression model. *Livest. Prod. Sci.* **64**, 133–145.

Rice J.A., Wu C.O. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.

Schaeffer L.R. (2004) Application of random regression models in animal breeding. *Livest. Prod. Sci.* **86**, 35–45.

Spiegelhalter D.J., Best N.G., Carlin B.P., von der Linde A. (2002) Bayesian measures of model complexity and fit. *J. Roy. Stat. Soc. B* **64**, 1–34.

Thompson R. (1976) The estimation of maternal genetic variance. *Biometrics* **32**, 903–917.

Appendix

Estimates of the direct genetic (A), maternal genetic (M), direct permanent environmental (R), and maternal permanent environmental (C) covariance functions for model M3353 were

$$\begin{aligned} \mathcal{A}(a_i^*, a_j^*) &= 744.862\phi_{0i}\phi_{0j} + 342.409(\phi_{0i}\phi_{1j} + \phi_{0j}\phi_{1i}) \\ &\quad - 38.965(\phi_{0i}\phi_{2j} + \phi_{0j}\phi_{2i}) + 190.979\phi_{1i}\phi_{1j} \\ &\quad + 5.080(\phi_{1i}\phi_{2j} + \phi_{1j}\phi_{2i}) + 19.026\phi_{2i}\phi_{2j} \end{aligned}$$

$$\begin{aligned} \mathcal{M}(a_i^*, a_j^*) &= 31.787\phi_{0i}\phi_{0j} + 2.635(\phi_{0i}\phi_{1j} + \phi_{0j}\phi_{1i}) \\ &\quad - 11.402(\phi_{0i}\phi_{2j} + \phi_{0j}\phi_{2i}) + 1.464\phi_{1i}\phi_{1j} \\ &\quad - 0.125(\phi_{1i}\phi_{2j} + \phi_{1j}\phi_{2i}) + 5.075\phi_{2i}\phi_{2j} \end{aligned}$$

$$\begin{aligned} \mathcal{R}(a_i^*, a_j^*) &= 589.278\phi_{0i}\phi_{0j} + 227.361(\phi_{0i}\phi_{1j} + \phi_{0j}\phi_{1i}) \\ &\quad - 11.886(\phi_{0i}\phi_{2j} + \phi_{0j}\phi_{2i}) + 6.999(\phi_{0i}\phi_{3j} \\ &\quad + \phi_{0j}\phi_{3i}) - 40.991(\phi_{0i}\phi_{4j} + \phi_{0j}\phi_{4i}) \\ &\quad + 138.898\phi_{1i}\phi_{1j} + 8.609(\phi_{1i}\phi_{2j} + \phi_{1j}\phi_{2i}) \\ &\quad - 16.332(\phi_{1i}\phi_{3j} + \phi_{1j}\phi_{3i}) - 14.127(\phi_{1i}\phi_{4j} \\ &\quad + \phi_{1j}\phi_{4i}) + 66.337\phi_{2i}\phi_{2j} + 24.170(\phi_{2i}\phi_{3j} \\ &\quad + \phi_{2j}\phi_{3i}) - 17.405(\phi_{2i}\phi_{4j} + \phi_{2j}\phi_{4i}) \\ &\quad + 52.669\phi_{3i}\phi_{3j} + 17.832(\phi_{3i}\phi_{4j} + \phi_{3j}\phi_{4i}) \\ &\quad + 33.551\phi_{4i}\phi_{4j} \end{aligned}$$

$$\begin{aligned} \mathcal{C}(a_i^*, a_j^*) &= 50.023\phi_{0i}\phi_{0j} - 2.150(\phi_{0i}\phi_{1j} + \phi_{0j}\phi_{1i}) \\ &\quad - 23.131(\phi_{0i}\phi_{2j} + \phi_{0j}\phi_{2i}) + 1.771\phi_{1i}\phi_{1j} \\ &\quad + 2.477(\phi_{1i}\phi_{2j} + \phi_{1j}\phi_{2i}) + 12.157\phi_{2i}\phi_{2j} \end{aligned}$$

with ϕ_{ki} denoting the k -th normalized Legendre polynomial of age a_i^* , the i -th age at recording standardized to the interval $[-1:1]$; e.g. $\phi_{0i} = \sqrt{1/2}$, $\phi_{1i} = \sqrt{3/2}a_i^*$ and $\phi_{2i} = \sqrt{45/8}(a_i^*)^2 - \sqrt{5/8}$ (Kirkpatrick *et al.* 1990).

Corresponding estimates of residual variances for the 19 classes were 1.26, 49.78, 65.76, 14.90, 16.78, 20.39, 35.13, 72.62, 111.93, 84.60, 100.45, 97.47, 116.01, 157.81, 113.50, 91.88, 161.55, 337.10 and 52.75 kg².