

JOURNAL OF ANIMAL SCIENCE

The Premier Journal and Leading Source of New Knowledge and Perspective in Animal Science

Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle

A. Legarra, I. Misztal and J. K. Bertrand

J Anim Sci 2004. 82:1564-1571.

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://jas.fass.org/cgi/content/full/82/6/1564>



American Society of Animal Science

www.asas.org

Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle

A. Legarra, I. Misztal¹, and J. K. Bertrand

Department of Animal and Dairy Science, University of Georgia, Athens 30602-2771

ABSTRACT: Genetic parameters for a random regression model of growth in Gelbvieh beef cattle were constructed using existing estimates. Information for variances along ages was provided by parameters used for routine Gelbvieh multiple-trait evaluation, and information on correlations among different ages was provided by random regression model estimates from literature studies involving Nellore cattle. Both sources of information were combined into multiple-trait estimates; corrected for continuity, smoothness, and general agreement with literature estimates; and extrapolated to 730 d. Covariance functions using standardized

Legendre polynomials were fit for the following effects: additive genetic (direct and maternal), and animal and maternal permanent environment. Residual variances at different ages were fitted using linear splines with three knots. Fit was by least squares. The order of polynomials was varied from third to sixth. Increasing the fit beyond cubic provided small improvements in R^2 and increased the number of small eigenvalues of covariance matrices, especially for the additive effect. Parameters for a random regression model in beef cattle can be constructed with negligible artifacts from literature estimates. Formulas can easily be modified for other types of polynomials and splines.

Key Words: Beef Cattle, Genetic Evaluation, Random Regression

©2004 American Society of Animal Science. All rights reserved.

J. Anim. Sci. 2004. 82:1564–1571

Introduction

Models with covariance functions using polynomials, also known as random coefficient or random regression models (Kirkpatrick et al., 1990), are routinely applied in the genetic evaluation of dairy, but are used less frequently in the beef cattle industry. In the United States, the genetic evaluation of beef cattle is based on a multiple-trait model (birth, 205 d, and 365 d; BIF, 2002). Weights close to given ages are preadjusted, whereas remaining weights are discarded. Random regression models allow for the use of all available records without preadjustment or editing, and would provide estimates of breeding values at any age.

There are few reports of genetic parameter estimates for random regression models in beef cattle (Albuquerque and Meyer, 2001; Meyer, 2002; Nobre et al., 2003a). These estimates seem to contain artifacts due to 1) the tendency of polynomials to provide poor fit at the extremes, 2) an uneven distribution of data points, and 3) the use of small datasets due to high computing costs (Misztal et al., 2000). As a result of those problems,

Nobre et al. (2003b) found that evaluations from random regression models with parameters estimated from the data were worse than from multiple-trait models.

Misztal et al. (2000) described the so-called “constructive approach” to form artifact-free estimates of parameters of random regression models. The basic idea was to assemble functions of variances along the trajectory and of correlations across two trajectories, construct multiple-trait model parameters for a large number of traits, and then fit random regression model parameters, as in Kirkpatrick et al. (1990). In dairy cattle, some parameters of a random regression model were derived from multiple-trait model parameters (Van Der Werf et al., 1998; Kettunen et al., 2000; Emmerling et al., 2002). The purpose of this study was to develop artifact-free parameters for a random regression model in Gelbvieh beef cattle using multiple-trait model parameters estimates for this breed, literature estimates, and heuristics.

Materials and Methods

Model

A multiple-trait model, as in BIF (2002), was assumed. An equivalent random regression model should account for the following sources of variation: animal and maternal additive genetic, animal permanent envi-

¹Correspondence—phone: 706-542-0951; fax: 706-583-0274; e-mail: ignacy@uga.edu.

Received August 29, 2003.

Accepted February 24, 2004.

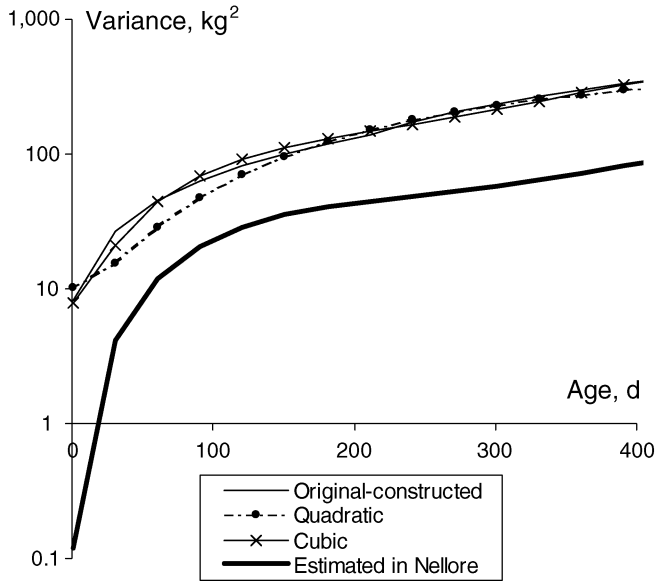


Figure 1. Original and fitted variances for direct additive effect, using polynomials of order 3 (quadratic) and 4 (cubic). The quartic polynomial is overlapped with a cubic polynomial. The estimated Nellore direct variance (Nobre et al., 2003a) is included as a reference.

ronment, maternal permanent environment, and residual. A residual effect in a multiple-trait model is equivalent to permanent environmental and residual effects in a random regression model. Thus, residual (co)variances in a multiple-trait model should be decomposed into permanent environmental and residual variances in the corresponding random regression model.

Algorithm

The algorithm used to estimate the random regression model parameters was as follows:

1. Using literature, for all effects, prepare estimates of variances for weights at several ages and estimates of correlations among different weights;
2. For each random effect except the residual:
 - a) set a vector of ages \mathbf{d} of size n_d
 - b) using variances from 1, obtain a vector of variances \mathbf{v} corresponding to days in \mathbf{d} by linear interpolation or, beyond points obtained from literature, by linear extrapolation.
 - c) similar to b, calculate a matrix of correlations among days in \mathbf{d} : \mathbf{C} .
 - d) Construct a multiple-trait model matrix of (co)variances \mathbf{G}_n , by multiplying each correlation in \mathbf{C} by the correspondent standard deviations of the variances in \mathbf{v} :

$$\mathbf{G}_n = \mathbf{C} \odot (\mathbf{v} \mathbf{v}')^{0.5}$$

where \odot is the Hadamard product.

- e) Let $f_i(x)$ be an i th regression for day x . Matrix Φ is a $n_d \times n_p$ matrix of functions of n_p degree (e.g.,

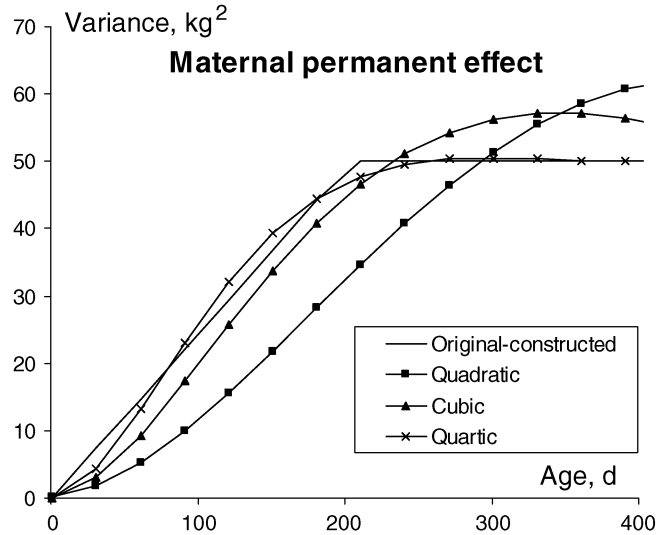
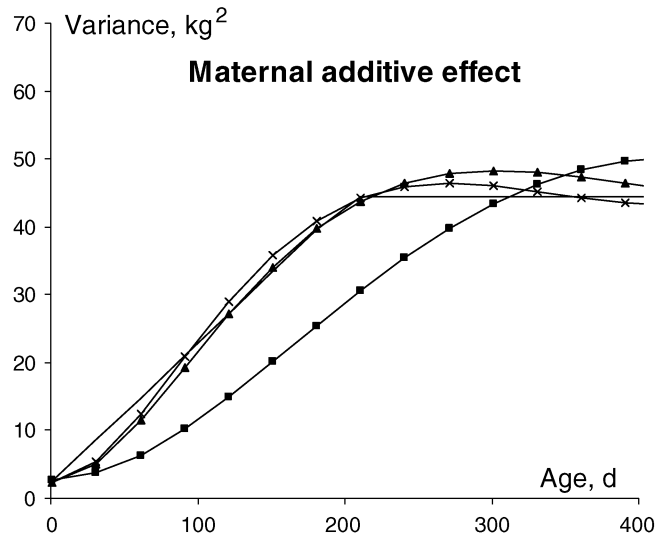


Figure 2. Original and fitted variances for maternal genetic and maternal permanent effects using polynomials of order 3 (quadratic) to 5 (quartic).

polynomials) for all days in \mathbf{d} . Matrices Φ need not be the same for different random effects. Parameters of the random regression model: \mathbf{K}_n of dimension $n_p \times n_p$ can be estimated as in Kirkpatrick et al. (1990) by solving ordinary least squares:

$$\mathbf{G}_n = \Phi \mathbf{K}_n \Phi' + \mathbf{E}$$

where \mathbf{E} is a matrix of residuals. In a linear form:

$$\text{vec}(\mathbf{G}_n) = (\Phi \otimes \Phi) \text{vec}(\mathbf{K}_n) + \mathbf{e}$$

where \otimes is the Kronecker product and vec is an operator that stacks full-stored matrices in vectors (Searle, 1982). Operator vech , which stacks half-stored symmetric matrices, was used by Kirkpatrick et al. (1990); however, in our work, it often

Table 1. Parameters at 1, 205, and 365 d as initially created and as corrected for additive direct and maternal genetic, maternal permanent environment (MPE), and residual effects

Item	Direct			Maternal		
	1 d	205 d	365 d	1 d	205 d	365 d
Genetic parameters						
Direct, 1 d	8.13^a	0.51	0.45	-0.21	0	0
	8.13	0.55	0.46	-0.21	-0.20	-0.20
Direct, 205 d		133.6	0.87	0	-0.21	-0.21
		133.6	0.79	-0.20	-0.20	-0.20
Direct, 365 d			304.5	0	-0.14	-0.14
			304.5	-0.20	-0.20	-0.20
Maternal, 1 d				2.44	0	0
				2.44	0.18	0.12
Maternal, 205 d					44.5	1
					44.5	0.94
Maternal, 365 d						44.5
						44.5
MPE parameters						
1 d	0^a	0	0			
	0.02	0.38	0.28			
205 d		50.0	1			
		50.0	0.97			
365 d			50.0			
			50.0			
Residual parameters						
1 d	9.85^a	0.17	0.12			
	9.85	0.12	0.10			
205 d		344.8	0.69			
		344.8	0.58			
365 d			728.5			
			728.5			

^aParameters as initially created (upper) and as corrected (lower). Variances (in kg²) are on the diagonal in bold type; correlations are on the upper diagonal.

resulted in negative-definite estimates of \mathbf{K}_n . Least squares result in equal weights given to each (co)-variance. A better match for selected points can be obtained by using weighted least squares.

3. Residual effect in the multiple-trait model is modeled in the random regression model as a sum of permanent environment plus random regression model residual. Subsequently the residual (co)variance matrix \mathbf{R}_0 is modeled in the random regression model as a sum of permanent environment (\mathbf{K}_p) and a diagonal error variance (\mathbf{R}_r):

$$\mathbf{R}_0 = \Phi \mathbf{K}_p \Phi' + \mathbf{R}_r$$

where Φ are matrices of functions as defined previously. Let $\text{diag}(\mathbf{R}_r) = \Theta \mathbf{k}_r$, where Θ is a matrix of functions, not necessarily the same as before, and \mathbf{k}_r is a vector of parameters for the residual effect in the random regression model. Solutions of \mathbf{K}_p and \mathbf{k}_r can be obtained by least squares in the following system of equations:

$$\text{vec}(\mathbf{R}_0) = (\Phi \otimes \Phi) \text{vec}(\mathbf{K}_p) + \mathbf{F} \mathbf{k}_r + \varepsilon$$

where \mathbf{F} is a matrix of dimensions $(n_d \times n_d) \times (\text{size of } \mathbf{k}_r)$ such that the i th row of \mathbf{F} is equal to the j th

row of Θ if $i = (j - 1)^2 + 1$, and to a row of zeros otherwise, and where ε is a vector of residuals.

Implementation

The vector of ages was set to $\mathbf{d}' = (1, 31, 61, \dots, 691, 721, 733)$. Literature estimates for variances were those used at the University of Georgia for routine genetic evaluation of Gelbvieh for 1, 205, and 365 d. Correlations among traits were initially obtained from Nobre et al. (2003a) for the missing data set and visually corrected based on common sense (i.e., after visualization of the corresponding patterns between different ages and detection of obvious artifacts such as “holes”); correlations among 1, 205, and 365 d were similar from the two sources. As the variance of maternal permanent environmental effect in the multiple-trait model at birth was considered 0 and the random regression model requires nonzero variances, that variance was arbitrarily set to 0.02 kg². As the estimate of direct-maternal correlations from Nobre et al. (2003a) contained many artifacts, it was arbitrarily set to -0.2. Parameters as prepared are shown on a three-trait scale in Table 1.

Regressions for all effects except the residuals were standardized Legendre polynomials from order 3 (qua-

Table 2. Fit (R^2) and number of small eigenvalues using polynomials of different order, for additive direct and maternal, maternal permanent environment (MPE) and permanent environment (PE) using polynomials of order 3 (quadratic) to 6 (quintic)

Order	Additive		Maternal permanent environment		Permanent environment	
	R^2	Small ^a	R^2	Small ^a	R^2	Small ^a
3	0.976	0	0.942	1	0.962	0
4	0.997	1	0.985	1	0.985	1
5	0.998	4	0.998	2	0.992	1
6	0.999	6	0.998	3	0.995	1

^aAbsolute value of the eigenvalue lower than 10^{-3} times the largest eigenvalue.

dratic) to 6 (quintic). For simplicity, the same degree was fit for all polynomials at the same time. For residuals, regressions were linear splines with knots at 1, 205, and 365 d. Direct and maternal additive genetic parameters were estimated simultaneously.

Therefore the random regression model for the Gelbvieh data to be hypothetically evaluated with the parameters obtained from this work would be:

$$y = (\text{fixed effects}) + \sum_{i=1}^n \phi_i(t)a_i + \sum_{i=1}^n \phi_i(t)m_i + \sum_{i=1}^n \phi_i(t)mpe_i + \sum_{i=1}^n \phi_i(t)pe_i + \varepsilon$$

where y is a weight at a given (standardized) age t ; ϕ_i is the i th order Legendre polynomial as a function of t ; a_i , m_i , mpe_i , and pe_i are the i th random additive direct, additive maternal, maternal permanent effects environment, and animal permanent environment associated with each respective Legendre polynomial, respectively. Variance components for these effects were denoted as \mathbf{K} previously (with different subscripts): n is the order of the Legendre polynomials, and ε is the associated residual whose variance component varies with age and is denoted as \mathbf{R}_r .

Analyses

Evaluation of the estimated parameters was done in several ways: 1) fit, measured as R^2 ; 2) size and sign of the eigenvalues of the parameters; 3) comparison of original and estimated correlations of effects at 1, 205 and 365 d; 4) agreement in plots of original and estimated variances along the trajectory; and 5) agreement in plots of original and estimated correlations among ages. Computer programs for estimation and for visualization were written in Scilab (<http://www-rocq.inria.fr/scilab/>), a matrix-programming environment. These programs are available from the authors upon request.

Results and Discussion

Table 2 shows the R^2 and the number of small (lower than 0.1% of the highest one) eigenvalues for each ef-

fect; the direct and maternal effects are treated as one effect. High-order Legendre polynomials resulted in parameters with very small eigenvalues. Some of these eigenvalues in fifth- and sixth-order polynomials were negative; the methodology does not guarantee positive-definiteness of the result. However, because reducing the rank of the random regression model covariance matrix (i.e., setting these eigenvalues to 0) provided almost identical fit in terms of R^2 , these eigenvalues were unimportant.

The proposed and estimated variances along time for quadratic, cubic, and quartic fits are shown in Figures 1 to 3. The fit appears to be very good even with the quadratic fit for all but the maternal effects. The maternal effects have an assumed abrupt change at 205 d, and this is difficult to model with polynomials. In general, it may be more practical to change the proposed shape of these curves than to increase the dimension of the covariance functions. Figure 3 is arranged to show how

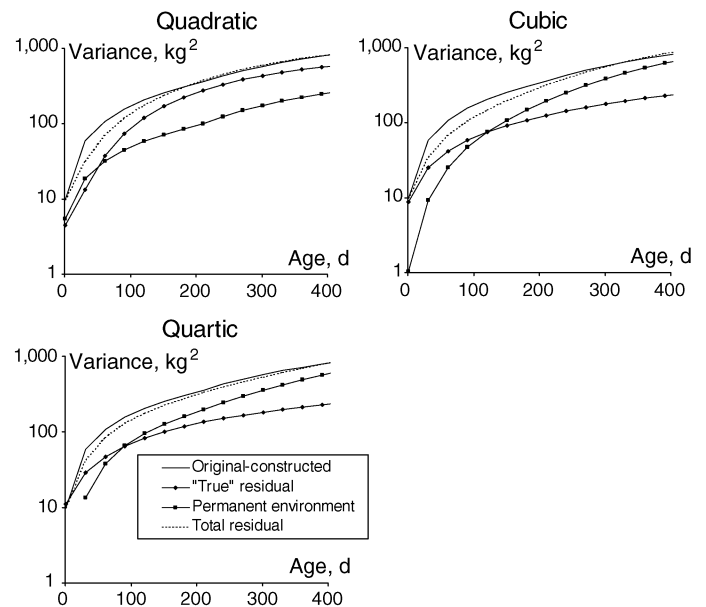


Figure 3. Original and fitted variances for residual effects using polynomials of order 3 (quadratic) to 5 (quartic).

Table 3. Parameters at 1, 205, and 365 d as estimated with the cubic fit for additive direct and maternal genetic, maternal permanent environment (MPE), and residual effects

Item	Direct			Maternal		
	1 d	205 d	365 d	1 d	205 d	365 d
Genetic parameters						
Direct, 1 d	7.85^a	0.55	0.46	-0.22	-0.20	-0.21
Direct, 205 d		144.3	0.80	-0.21	-0.19	-0.20
Direct, 365 d			288.9	-0.21	-0.20	-0.21
Maternal, 1 d				2.34	0.21	0.15
Maternal, 205 d					43.1	0.92
Maternal, 365 d						46.5
MPE parameters						
1 d	0.04^a	0.73	0.64			
205 d		45.6	0.97			
365 d			57.1			
Residual parameters						
1 d	9.86^a	0.13	0.10			
205 d		307.8	0.61			
365 d			769.1			

^aVariances (in kg²) are on diagonal in bold type; correlations are on the upper diagonal.

in a random regression model the sum (the “total” residual) of the permanent environmental effect and the residual is equivalent to the residual in a multiple-trait model.

For (co)variances in this paper, cubic polynomials provide an adequate fit, while avoiding numerical problems. The initial (co)variances were of order 6×6 for the additive effects and 3×3 for the remaining effects, and the cubic fit results were in the order 7×7 and 3×3 , respectively, if dimensions with near-zero eigenvalues are deleted. Slightly higher order of parameters for random regression model than for multiple-trait parameters matrices is usually necessary because Legendre polynomials require higher order to fit less regular functions. One should note that a perfect fit is not required because such a fit would increase computations and susceptibility to numerical errors, and BLUP is robust to slightly incorrect parameters. However, reduction of dimensionality by the elimination of small eigenvalues may result in poor fit at extremes of the trajectory (Foulley and Robert-Granié, 2002). In the case of the cubic fit, such reduction for the direct effect resulted in lower direct variance for birth weight, but negligible differences for (co)variances associated with weaning weight and yearling weight. Reduction for the other effects had much lower influence on the (co)variances. It seems that reduction of dimensionality due to elimination of small eigenvalues may not be prudent in all cases. We believe cubic polynomials to be a good compromise in this case. The values of the parameters (variance components of the random regression model) estimated with the cubic fit are shown in the Appendix.

In general, the agreement between the fit provided by cubic polynomials (Table 3) with those proposed in Table 1 is quite good, with the exceptions of those correlations that were considered as null in the Gelbvieh multiple-trait model but not in the random regression

model. However, their importance in practice is small because of the low correlations (in the maternal additive effect) or variances (in the maternal permanent effect). If necessary, one can achieve a better fit for selected points by using weighted least squares.

Figure 4 shows the pattern of genetic correlations between the additive effects at different times. Holes in such a curve indicate fluctuating correlations, which are unlikely to be biologically realistic. Such a hole can be observed in the pattern of correlations between direct effects at 50 and 150 d of age; however, the additive correlations in this age range are low. The rest of the correlation pattern is reasonably smooth and it is considered as acceptable. “Slices” of this graph at 1 and 205 d of age are shown in Figures 5 and 6, including the original “constructed” multiple-trait model parameters and the fit provided by quadratic to quartic polynomi-

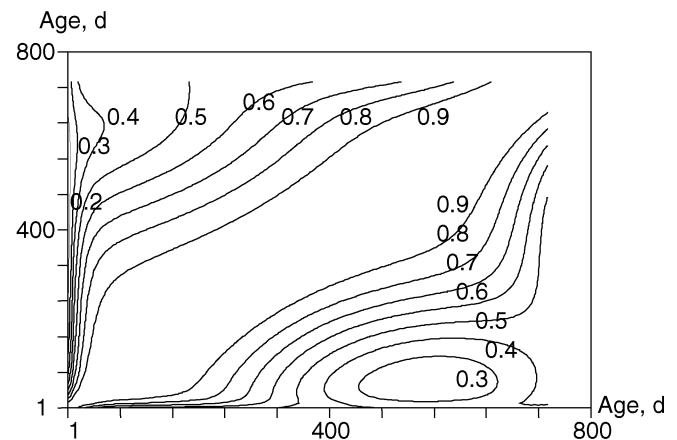


Figure 4. Correlations fitted for Gelbvieh between different ages for effects direct (lower right corner) and maternal (upper left corner), using cubic polynomials.

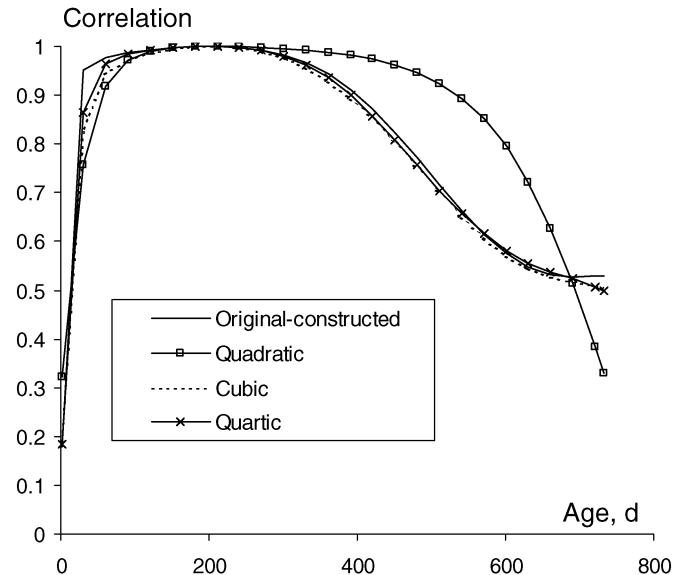
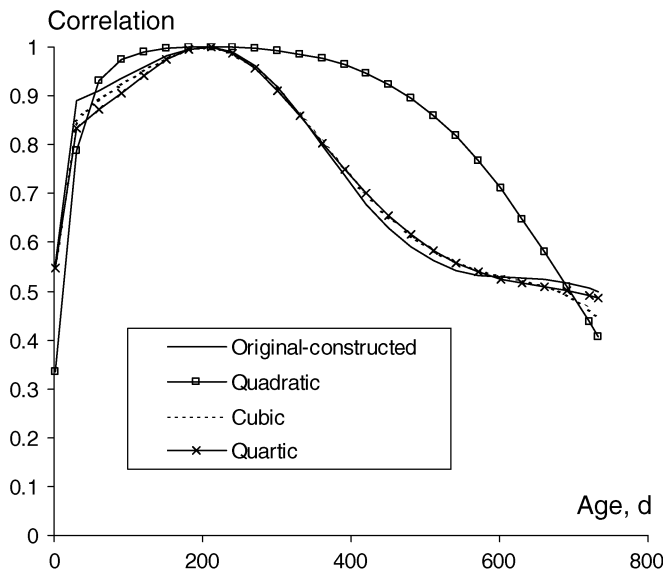
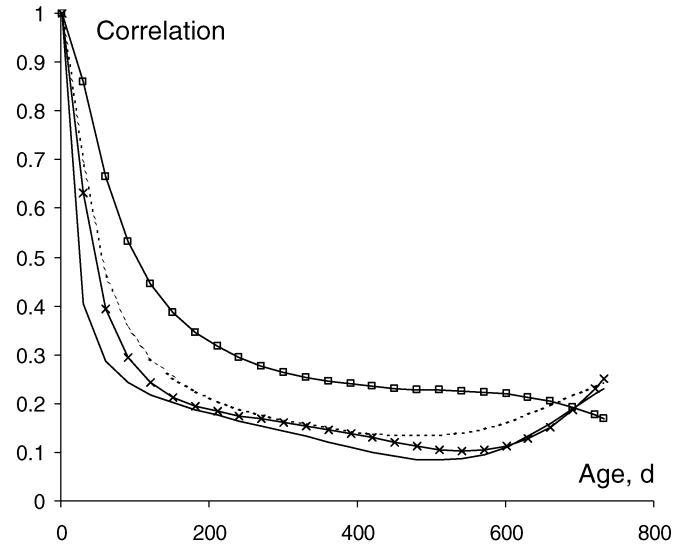
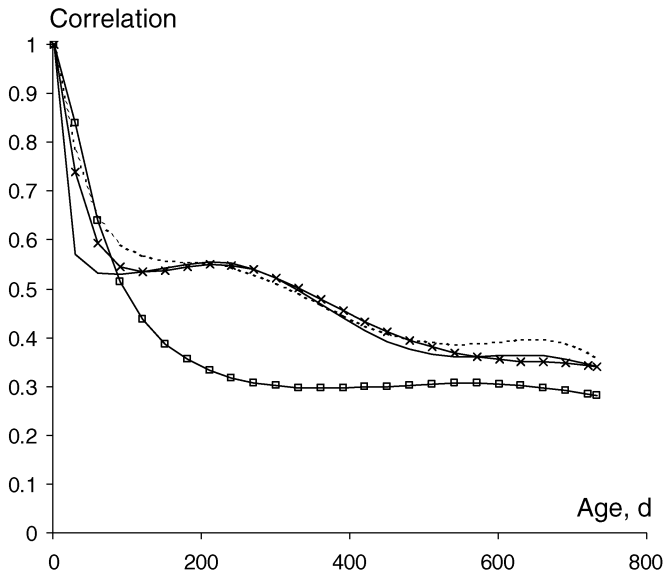


Figure 5. Original and fitted correlations of direct additive effect in d 1 (top) or d 205 (bottom), with effects in other days using polynomials of order 3 (quadratic) to 5 (quartic).

Figure 6. Original and fitted correlations of maternal additive effect in d 1 (top) or d 205 (bottom), with effects in other days using polynomials of order 3 (quadratic) to 5 (quartic).

als. The quadratic polynomials do not provide a good fit, whereas the cubic polynomials provide essentially the same fit as quartic.

Figure 7 shows the direct-maternal genetic correlation at the same age along the trajectory. The pattern estimated in Nellore was extremely erratic and it is shown as a reference. The arbitrary correlation of -0.2 , which was assumed in this study, was well approximated with any polynomial.

Other Polynomials and Splines

Parameters were fit for a random regression model using Legendre polynomials. Other functions, such as

fractional polynomials by Foulley and Robert-Granié (2002) or splines (White et al., 1999; Torres and Quaas, 2001), could have resulted in better fit at a lower dimensionality. In particular, the use of linear splines would allow the use of the multiple-trait model parameters without changes for the genetic and maternal permanent effects, but not for the residual or permanent environment. However, models with splines may have poor numerical properties when points are strongly correlated. Computer programs developed for this study could be easily modified to work with other functions or polynomials.

Use of random regression models with polynomials, splines, or other general functions may lead to highly correlated regressions and subsequently poor numeri-

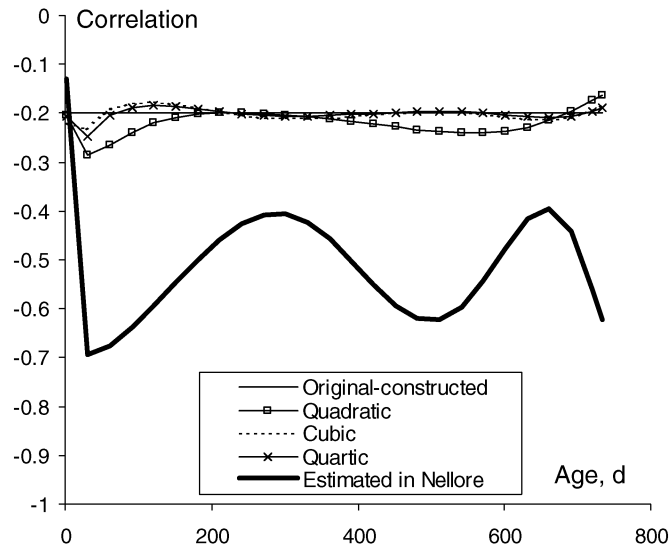


Figure 7. Correlations between direct and maternal additive effects at the same age as estimated in Nellore cattle (Nobre et al., 2003a), as constructed, and as fitted with polynomials of order 3 (quadratic) to 5 (quartic).

cal properties and low convergence rates if solutions are obtained by iteration. Covariates in random regression models can be reparameterized to result in diagonal (co)variances (Van der Werf, 1998) and subsequently much improved numerical properties (Lidauer et al., 2003; Nobre et al., 2003b). For an order of fit high enough, such reparameterized covariates are likely to be very similar regardless of which polynomial/spline/functions were used initially.

Implications

Parameters for random regression models for growth traits in beef cattle can be obtained without estimation from the data by using multiple-trait models' parameter estimates derived from the data or from the literature and common-sense corrections. The calculations allow for testing the necessary degree of fit in negligible amount of time. Cubic Legendre polynomials seem to provide a reasonable fit with minimal artifacts.

Literature Cited

- Albuquerque, L. G., and K. Meyer. 2001. Estimates of covariance functions for growth from birth to 630 days of age in Nellore cattle. *J. Anim. Sci.* 79:2776–2789.
- BIF. 2002. Guidelines for Uniform Beef Improvement Programs. Beef Improv. Fed., Athens, GA.
- Emmerling, R., E. A. Mäntysaari, and M. Lidauer. 2002. Reduced rank covariance functions for a multi-lactation test-day model. CD-ROM communication No. 17-03 in Proc. 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.
- Foulley, J. L., and C. Robert-Granié. 2002. Basic tools for the statistical analysis of longitudinal data via mixed linear models. Course Notes. Montpellier, France.
- Kettunen, A., E. A. Mäntysaari, and J. Pösö. 2000. Estimation of genetic parameters for daily milk yield of primiparous Ayrshire cows by random regression test-day models. *Livest. Prod. Sci.* 66:251–261.
- Kirkpatrick M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124:979–993.
- Lidauer, M., E. A. Mäntysaari, and I. Strandén. 2003. Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livest. Prod. Sci.* 79:73–86.
- Meyer, K. 2002. Estimates of covariance functions for growth of Australian beef cattle from a large set of data. CD-ROM communication No. 11-01 in Proc 7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.
- Misztal, I., T. Strabel, J. Jamrozik, E. A. Mäntysaari, and T. H. E. Meuwissen. 2000. Strategies for estimating the parameters needed for different test-day models. *J. Dairy Sci.* 83:1125–1134.
- Nobre, P. R. C., I. Misztal, S. Tsuruta, J. K. Bertrand, L. O. C. Silva, and P. S. Lopes. 2003a. Analysis of growth curves of Nellore cattle by multiple-trait and random regression models. *J. Anim. Sci.* 81:918–926.
- Nobre, P. R. C., I. Misztal, S. Tsuruta, J. K. Bertrand, L. O. C. Silva, and P. S. Lopes. 2003b. Genetic evaluation of growth in Nellore cattle by multiple-trait and random regression models. *J. Anim. Sci.* 81:927–932.
- Searle, S. R. 1982. *Matrix Algebra Useful for Statistics*. John Wiley and Sons, New York, NY.
- Torres, R. A. A., Jr., and R. L. Quaas. 2001. Determination of covariance functions for lactation traits on dairy cattle using random-coefficient regressions on B-splines. *J. Anim. Sci.*(Suppl. 1)79:112. (Abstr.)
- Van Der Werf, J. H. J., M. E. Goddard, and K. Meyer. 1998. The use of covariance functions and random regressions for genetic evaluation of milk production based on test day records. *J. Dairy Sci.* 81:3300–3308.
- White, I. M. S., R. Thompson, and S. Brotherstone. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.* 82:632–638.

Appendix

Estimates of the parameters (variances and covariances) of the random regression model using cubic Legendre polynomials and values in Table 1 (second row). The term “a” stands for direct additive genetic, “m” for maternal additive genetic, “mpe” for maternal permanent environment, and “pe” for individual animal permanent environment. The subscript of each random effect is the order of the associated term in the polynomial.

Additive genetic (direct and maternal)

Item	a ₁	a ₂	a ₃	a ₄	m ₁	m ₂	m ₃	m ₄
a ₁	413.83	181.33	-48.98	-19.42	-40.24	-5.58	5.29	-3.09
a ₂	181.33	139.83	1.42	-25.42	-15.94	-2.21	2.09	-1.22
a ₃	-48.98	1.42	47.78	21.83	2.01	0.28	-0.26	0.15
a ₄	-19.42	-25.42	21.83	27.17	-0.38	-0.05	0.05	-0.03
m ₁	-40.24	-15.94	2.01	-0.38	58.68	8.03	-12.80	4.69
m ₂	-5.58	-2.21	0.28	-0.05	8.03	9.93	0.84	-2.61
m ₃	5.29	2.09	-0.26	0.05	-12.80	0.84	5.51	-0.93
m ₄	-3.09	-1.22	0.15	-0.03	4.69	-2.61	-0.93	2.56

Maternal permanent environment

Item	mpe ₁	mpe ₂	mpe ₃	mpe ₄
mpe ₁	72.81	11.06	-15.39	6.63
mpe ₂	11.06	4.44	-1.01	0.36
mpe ₃	-15.39	-1.01	5.57	-0.34
mpe ₄	6.63	0.36	-0.34	1.87

Animal permanent environment

Item	pe ₁	pe ₂	pe ₃	pe ₄
pe ₁	730.90	342.30	-125.12	-59.95
pe ₂	342.30	227.60	-6.21	-26.95
pe ₃	-125.12	-6.21	87.53	31.14
pe ₄	-59.95	-26.95	31.14	20.88

Residual variance: Values at days 1, 205, and 365

8.83
121.94
214.40

References

This article cites 10 articles, 7 of which you can access for free at:
<http://jas.fass.org/cgi/content/full/82/6/1564#BIBL>

Citations

This article has been cited by 5 HighWire-hosted articles:
<http://jas.fass.org/cgi/content/full/82/6/1564#otherarticles>