J. Anim. Breed. Genet. ISSN 0931-2668

ORIGINAL ARTICLE

Properties of random regression models using linear splines

I. Misztal

Animal and Dairy Science, University of Georgia, Athens, GA, USA

Correspondence
I. Misztal, Animal and Dairy Science, University of Georgia, Athens, GA 30602, USA.
Tel: +1 706 542 0951; Fax: +1 706 583 0274;
E-mail: ignacy@uga.edu

Received: 23 August 2005;
accepted: 10 January 2006

Summary

Properties of random regression models using linear splines (RRMS) were evaluated with respect to scale of parameters, numerical properties, changes in variances and strategies to select the number and positions of knots. Parameters in RRMS are similar to those in multiple trait models with traits corresponding to points at knots. RRMS have good numerical properties because of generally superior numerical properties of splines compared with polynomials and sparser system of equations. These models also contain artefacts in terms of depression of variances and predictions in the middle of intervals between the knots, and inflation of predictions close to knots; the artefacts become smaller as correlations corresponding to adjacent knots increase. The artefacts can be greatly reduced by a simple modification to covariables. With the modification, the accuracy of RRMS increases only marginally if the correlations between the adjacent knots are ≥0.6. In practical analyses the knots for each effect in RRMS can be selected so that: (i) they cover the entire trajectory; (ii) changes in variances in intervals between the knots are approximately linear; and (iii) the correlations between the adjacent knots are at least 0.6. RRMS allow for simple and numerically stable implementations of genetic evaluations with artefacts present but transparent and easily controlled.

Introduction

Many analyses by random regression models (RRM) use Legendre polynomials (Kirkpatrick *et al.* 1990; Van der Werf 1997) as basis functions. These polynomials are able to model a variety of curves for variances and covariances but they also have undesirable properties. Fit at the extremes of the trajectory may be poor. Curves at points of the trajectory with few records are likely to contain artefacts (Misztal *et al.* 2000; Nobre *et al.* 2003a; Meyer 2005a). Parameters on a scale of polynomials require conversions to the original scale to determine whether these are realistic or not. If they contain many artefacts, they can be converted to the original scale, corrected, and then converted back. However, this process can be complicated, can introduce additional artefacts, and the resulting matrices of parameters may be singular (Legarra *et al.* 2004). Finally, there is a problem with numerical stability in genetic evaluation, as convergence may be problematic with large data sets (Robbins *et al.* 2005). Stability may be improved by reparameterization to diagonal variances but at the expense of some additional complexity (Lidauer *et al.* 2003; Nobre *et al.* 2003b). In maternal models where the genetic correlation between direct and maternal effects is assumed present, only a partial diagonalization is possible (Bohmanova *et al.* 2005). Rank reduction (i.e. elimination of dimensions with very small eigenvalues), which is usually performed with the diagonalization may result in large changes for some points on the trajectory, especially if there are large changes in variances along the trajectory (Bohmanova *et al.* 2005). Successful estimation of variances with Legendre polynomials is possible but requires large data sets, even distribution of data points on the trajectory, and careful modelling of other effects (Druet *et al.* 2003).

Problems with the Legendre polynomials are due to many factors: rapid changes of high-order terms at the extremes, poor modelling capability of asymmetrical functions, lack of information to estimate a very large number of parameters, and sensitivity of (co)variance curves to each of the many parameters. Therefore, several alternatives were proposed. Robert-Granié *et al*. (2002) advocated the use of fractional polynomials where basis functions also include roots and logs. Subsequently, changes at the extremes can be more moderate. White *et al.* (1999) used smoothing polynomials and Torres & Quaas (2001) used B-splines with 10 knots. Each coefficient of a spline function affects only a fraction of the trajectory, resulting in possibly better numerical properties and fewer estimation artefacts. Also, knots in splines can be chosen corresponding to the pattern of changes along the trajectory resulting in smaller dimensionality; knots can be denser in regions of fast changes, and sparser in regions of slow changes.

Meyer (2005b) looked at the theory of B-splines. Meyer also estimated parameters for a range of models using linear to cubic splines as well as Legendre polynomials. Meyer found the models fitting splines to be superior to those with polynomials, with quadratic splines being the best compromise. Models using cubic splines resulted in the lowest mean squared error but the largest artefacts. The numerical advantages of splines are smaller with increasing order of splines. For example, numerical properties with cubic splines and cubic polynomials are likely to be similar.

Recently, several studies at the University of Georgia looked at the use of linear splines (Bohmanova *et al.* 2005; Iwaisaki *et al.* 2005; Robbins *et al.* 2005) for growth in beef. Compared with a model using cubic Legendre polynomials, the model using linear splines (RRMS) with three knots was much easier to set up and had superior numerical properties. Part of the simplicity was due to parameters being on the same scale as in multiple traits.

In growth of beef, the knots were placed at points of maximum concentration of records, i.e. birth, weaning and yearling weights. If RRMS is to be applied for traits with more balanced distribution of records, an important question is how to select the number and location of knots. Too many knots will increase complexity, while too few would reduce accuracy. More knots can be placed in regions of fast (co)variance changes and fewer in regions with slower changes, e.g. based on changes on the phenotypic level. However, changes at the mean and variance scales are not necessarily the same. The aim of this paper was to investigate properties of RRMS with focus on methods to determine the number and placement of knots.

Materials and methods

Model

Assume n knots at points T_i, $i = 1, \ldots, n$. Let $y_{..jk.t}$ be the kth observation for subject j at time t. RRMS, showing two random effects only, can be presented as:

$$y_{..jk.t} = \cdots + \sum_{i=1}^{n} \varphi_i(t) a_{ij} + \sum_{i=1}^{n} \varphi_i(t) p_{ij} + e_{jk},$$

$$\mathrm{var}(a_j) = G_a, \quad \mathrm{var}(p_j) = G_p, \quad \mathrm{var}(e_{jk}) = \sigma^2_{e,k},$$

where $\varphi_i(t)$ is the ith covariate at time t, a_{ij} (p_{ij}) is the ith coefficient for the first (second) effect of subject j, a_j (p_j) is the vector of all coefficients for the first (second) effect of subject j, and e_{jk} is the residual for the kth record of subject j. The first effect could be additive genetic and the second permanent environmental. The covariates are:

$$\text{if} \qquad t = T_i : \varphi_i(t) = 1, \quad \varphi_j = 0, \qquad j \neq i,$$

$$\text{if} \quad T_i \leq t < T_{i+1} : \varphi_i(t) = \alpha, \quad \varphi_{i+1}(t) = 1 - \alpha,$$

$$\alpha = \frac{T_{i+1} - t}{T_{i+1} - t_i}, \qquad \varphi_j = 0,\ j < i \quad \text{and} \quad j > \mathrm{i} + 1$$

The formulas above assume that $T_1 \leq t < T_n$. If $t < T_1$ or $t > T_n$, one can use truncation or linear extrapolation.

Equivalence of variances with multiple trait model

The model above can be written in a matrix notation as:

$$y_{..jk.t} = \cdots + \Phi(t) a_j + \Phi(t) p_j + e_{jk},$$

where $\Phi(t)$ is a vector of covariables at point t. Assume that data points occur only at knots and that observation k occurs at knot k. Because

$$\varphi_i(T_k) = \begin{cases} 1 & \text{if } i = \mathrm{k} \\ 0 & \text{if } i \neq \mathrm{k} \end{cases}$$

the equation for an observation at knot k ($t = T_k$), dropping now redundant index t is:

$$y_{..jk} = \cdots + a_{jk} + p_{jk} + e_{jk}.$$

Assume that animal j has one observation at each knot. The model for all observations of that animal is:

$$\begin{bmatrix} y_{\cdot j1} \\ y_{\cdot j2} \\ \vdots \\ y_{\cdot jn} \end{bmatrix} = \cdots + \begin{bmatrix} a_{j1} \\ a_{j2} \\ \vdots \\ a_{jn} \end{bmatrix} + \begin{bmatrix} p_{j1} \\ p_{j2} \\ \vdots \\ p_{jn} \end{bmatrix} + \begin{bmatrix} e_{j1} \\ e_{j2} \\ \vdots \\ e_{jn} \end{bmatrix} = \cdots + a_j + p_j + e_j.$$

Because of no repeated records, the residual and permanent environment effects are confounded and can be merged. In a vector notation, this results in:

$$\begin{bmatrix} y_{\cdot j1} \\ y_{\cdot j2} \\ \vdots \\ y_{\cdot jn} \end{bmatrix} = \cdots + a_j + e_j^M,$$

$$\mathrm{var}(e_j^M) = R_0 = G_p + \mathrm{diag}([\sigma^2_{e,1}, \sigma^2_{e,2}, \ldots, \sigma^2_{e,n}]),$$

where e_j^M is a new vector of residual effects.

The above equations are those of a multiple trait model (MTM). Thus, under some conditions, the MTM and RRMS models are similar. When data points occur only at the knots traits, and traits in a MTM are defined at knot points, the animal effects in both models are identical. Furthermore, the residual effect in the MTM is equal to the sum of the permanent environment and residual effects in the RRMS. Following, parameters of both models are the same for the animal effect, with residual covariance matrix in the MTM being a sum of permanent environment and residual covariance matrices in the RRMS. These derivations can be generalized to more random effects and missing observations at some knots. Based on the study by Iwaisaki *et al.* (2005), variances from the RRMS are similar to those in the MTM even if data points occur outside but close to knot points.

With good literature estimates for MTM, it is possible to create satisfactory parameters for RRMS without re-estimation or conversion. In such a case, one needs to split the residual variance in MTM into the permanent and residual variances in RRM. Any split that results in a positive-definite covariance matrix for the permanent environment effect will result in the same solutions to animal effects if data points occur at knots only. If data points are scattered, both variances need to be estimated or derived from estimates for data points other than at knots.

Computing properties

With linear splines, convergence when solving mixing model equations was superior than in models with Legendre polynomials, and the cost of one round of iteration was lower (Bohmanova *et al.* 2005; Robbins *et al.* 2005). This was due to several factors. First, models with splines generally have better numerical properties, because each coefficient in splines affects only a fraction of the trajectory. Secondly, while in polynomials of order n all covariables are generally non-zero, in linear splines only covariables associated with two adjacent knots are non-zero. This increases the sparsity of the left-hand side of the mixed model equations (LHS). For example, assume a model involving m sets of covariables of size n, plus q additional effects. The number of contributions to the LHS from one observation will be $(mn + q)^2$ when polynomials are used and no more than $(2m + q)^2$ when linear splines are used. Assuming $m = 3$, $n = 4$ and $q = 5$, the number of contributions to LHS is 289 and 121 respectively.

Covariance functions with linear splines

Let us examine changes in (co)variances and correlations along the trajectory. Because, at most, only two adjacent knots have non-zero coefficients for any observation, it is sufficient to consider changes only within two knots. Let t be a point on a trajectory, $T_1 = 0 \le t \le 1 = T_2$. The value of one level of one effect (e.g. one animal), omitting indices, is

$$a(t) = \Phi(t)a = \begin{bmatrix} 1-t & t \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = (1-t)a_1 + ta_2.$$

Let

$$\mathrm{var}(a) = G = (g_{ij}).$$

Then

$$\mathrm{var}[a(t)] = (1-t)^2 g_{11} + 2t(1-t)g_{12} + t^2 \times g_{22}$$

$$\mathrm{cov}[a(0), a(t)] = (1-t)g_{11} + tg_{12}.$$

To further simplify the formulas, assume that observations on the trajectory have been standardized so that $g_{11} = g_{22} = 1$, $g_{12} = \rho$, where $\rho = \mathrm{corr}(a_1, a_2)$. Then

$$\mathrm{var}[a(t)] = (1-t)^2 + 2t(1-t)\rho + t^2$$

$$\mathrm{var}[a(0)] = 1; \quad \mathrm{var}[a(1)] = 1; \quad \mathrm{var}[a(0.5)] = 0.5 + 0.5\rho$$

$$\mathrm{cov}[a(0), a(1)] = \rho$$

$$\mathrm{cov}[a(0), a(t)] = 1 - t + \rho t.$$

Figure 1 shows changes of variances across the trajectory for different values of ρ. As the correlation decreases, the shape is more concave. For a given ρ, the depression in the middle ($t = 0.5$) is

$$\frac{\mathrm{var}[a(0.5)]}{\mathrm{var}[a(0)]} = \frac{(1+\rho)}{2}.$$

The changes in the covariance between the beginning of the trajectory and a given point are linear. However, the corresponding changes in the correlation are convex (see Figure 2). This convexity increases as ρ decreases.

Shapes of variance and covariance functions are desired to be smooth. With linear splines, these shapes will be smoother if correlations among the adjacent knots are high, potentially implying a large number of knots. However, having too many knots would lead to more computations and potential numerical problems. In a MTM, Schaeffer (1984) found that large inaccuracies in genetic variance and correlations caused negligible differences in predicted breeding values. In a simulation study, Strabel & Jamrozik (2002) simulated the data using one set of parameters and predicted breeding values using many sets of incorrect parameters. The loss of accuracy of breeding values was <0.02. Assuming that dips in the genetic variances up to 20% can be tolerated, the correlations between points should not be <0.6. Subsequently, knots can be chosen in such a manner so that: (i) extremes are covered; and (ii) the correlations between the adjacent knots are high but not too high, e.g. between 0.6 and 0.8. The above rules are derived indirectly assuming that the 'true' variance functions are approximately linear between adjacent knots. If changes in 'true' variances along the trajectory are fast and nonlinear, additional knots may be required.

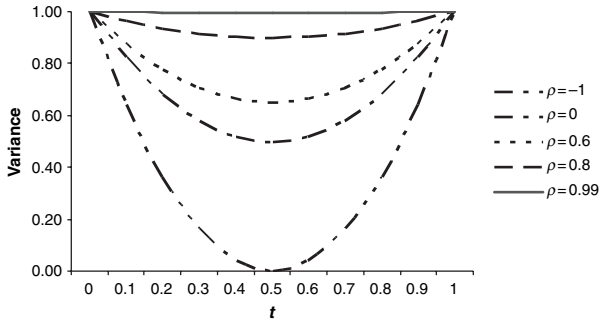

Figure 1 Changes in variances across the trajectory of a two-knot random regression model using linear splines for different correlations between the extreme knots.

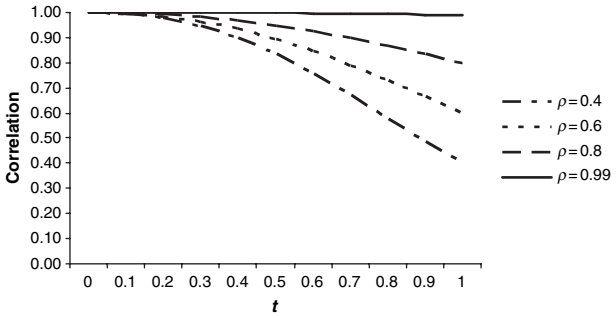

Figure 2 Corr [$a(0)$,$a(t)$] of a two-knot random regression model using linear splines for different correlations between the extreme knots.

Changes in variances between the knots can be made nearly linear if covariables are modified. One possible modification is changing the vector of covariables from $[(1-t) \quad t]$ to $[(1-t)^q \quad t^q]$, where q is a constant that makes the variance more linear. Subsequently,

$$a(t) = (1-t)^q a_1 + t^q a_2, \quad 0 \le q \le 1.$$

Then

$$\mathrm{var}[(a(t)] = (1-t)^{2q} + 2\rho t^q (1-t)^q + t^{2q}$$

$$\mathrm{cov}[a(0), a(t)] = (1-t)^q + \rho t^q.$$

In particular, q can be chosen so that $\mathrm{var}[a(0.5)] = 1$. Then:

$$q = \frac{\log[2(1+\rho)]}{[2\log(2)]}.$$

Figures 3 and 4 presents changes of variances (correlations) for a few values of q for $\rho = 0.5$. For $q = 0.79$, the variance is almost flat.

In case of many knots, the q parameter will be different for each pair of knots. Assume n knots and let

$$\rho_i = \mathrm{corr}(a_i, a_{i+1}) \quad \text{and} \quad q_i = \frac{\log[2(1+\rho_i)]}{[2 \times \log(2)]}.$$

For observation at point $t{:}t_i \le t \le 1 = t_{i+1}$, the unmodified vector if covariables will be:

$$[0 \ldots \quad 0 \quad \alpha \quad 1-\alpha \quad 0 \ldots \quad 0],$$

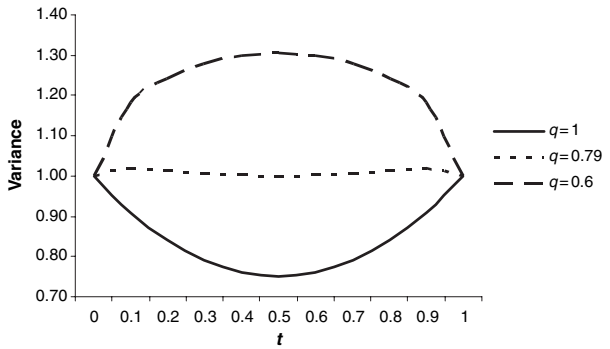

Figure 3 Variance of a two-knot random regression model using linear splines for different values of the modification parameter q with $\rho = 0.5$.

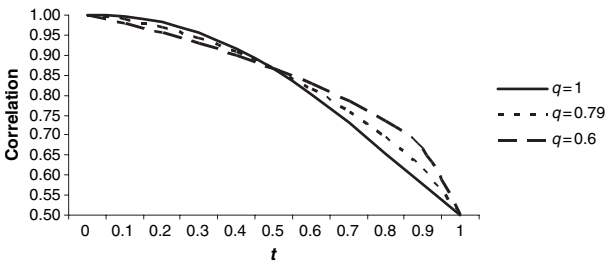

Figure 4 Corr $[a(0),a(t)]$ of a two-knot random regression model using linear splines for different values of the modification parameter q with $\rho = 0.5$.

where $\alpha = (t_{i+1} - t)/t_{i+1} - t_i)$, while the modified vector will be:

$$\left[0 \ldots \quad 0 \quad \alpha^{qi} \quad (1-\alpha)^{qi} \quad 0 \ldots \quad 0\right].$$

In a study by Bohmanova *et al.* (2005), knots were at 1, 205 and 365 days, $\text{corr}(a_1,a_2) = 0.55$, $\text{corr}(a_2,a_3) = 0.79$. Then, $q_1 = 0.82$, $q_2 = 0.92$, and the vectors of modified covariables at days 100 and 300 are:

$$\begin{aligned}\Phi(100) &= [0.512^{0.82} \quad 0.488^{0.82} \quad 0],\\ \Phi(300) &= [0 \quad 0.406^{0.92} \quad 0.594^{0.92}].\end{aligned}$$

Numerical comparisons

The purpose of numerical comparisons was to evaluate the effect of correlations among adjacent knots on variances and accuracy of predictions by RRMS. This was performed by simulating the data with a near-continuous covariance function (approximated by five knots) and analysing with the model used in the simulation and with two-knot models. Data were simulated for a model with one random effect and five equally spaced knots $t_i = i$, i=1,5. Variances were computed according to the formula:

$$g_{ij} = \frac{1-(1-\rho)|i-j|}{4} \text{ with}$$
$$\rho = \{0.99, 0.9, 0.8, 0.6, 0.4, 0.2, 0.0\},$$

where the correlation between the extreme points was ρ and the residual variance was set to 10. Observations were generated for 1000 unrelated sires with 10 observations per sire. Each observation included records at each of five knots. Solutions to sire values were calculated for models with: (i) five knots as used to generate the data (K5); and (ii) two knots ($t_1 = 1$, $t_2 = 5$). For the model with two knots, either regular (K2) or modified covariables (K2M) were used; the modification included the values of q so that $\text{var}[a(3)] = 1$. The five-knot model is the model for comparisons as it was the model used to generate the data. The two-knot models allow for analyses of loss of accuracy as a function of the correlation between the extreme knots. Statistics computed were accuracies of sire predictions for $t = 1$ and $t = 3$ and the variance of predictions at the same points. The first point ($t = 1$) is estimated directly by the two-knot model while the second point is only approximated. Results are averages over 25 replicates.

Results and discussion

Figure 5 presents the variance of the prediction at one knot $(\text{var}[\hat{a}(1)])$.The variance is rising with increasing ρ as adjacent points contribute more information. Compared with K5, the variance with K2 is inflated although this inflation is decreasing with increasing ρ. The use of K2M reduces the inflation drastically. Figure 6 presents the variance of the prediction in the middle of the trajectory $(\text{var}[\hat{a}(3)])$.This variance is rising with ρ as before; however, the variance with K2 is deflated although this deflation is decreasing with increasing ρ. Again, the use of K2M reduces the deflation drastically. The inflation or deflation with K2 can result in inflated/deflated proofs in between the knots. This may be a problem if differences between specific points are used in selection. The effect of inflation and deflation on breeding values will partially cancel out if selection is on an average of many points.

Figures 7 and 8 present accuracies computed as $\text{corr}[(\hat{u}(1), u(1)]^2$and $\text{corr}[(\hat{u}(3), u(3)]^2$respectively. As expected, the accuracies of K2 and K2M are lower

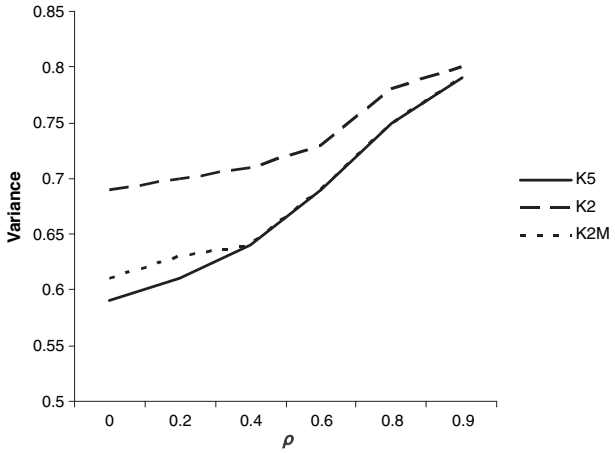

Figure 5 Variance of the prediction of $a(1)$ for models with five knots (K5), two knots and regular (K2) or modified covariables (K2M).

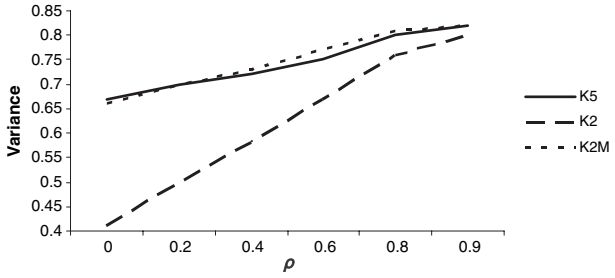

Figure 6 Variance of the prediction of $a(3)$ for models with five knots (K5), two knots and regular (K2) or modified covariables (K2M).

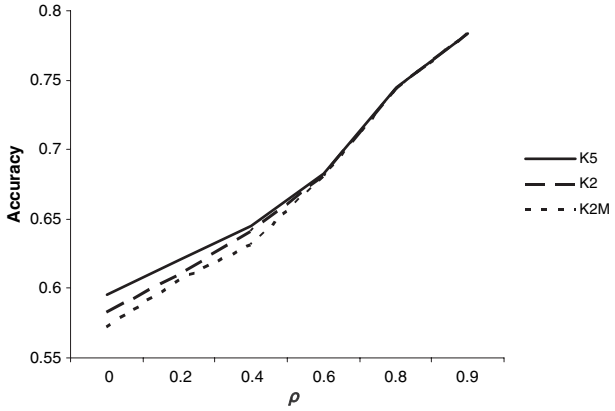

Figure 7 Accuracy of the prediction of the extreme point for models with five knots (K5), two knots and regular (K2) or modified covariables (K2M).

than those of K5; however, differences are small. K2 is marginally more accurate than K2M for $t = 1$ and the reverse is true for $t = 3$; differences between K5 and K2M are <0.005 for $\rho \geq 0.6$.

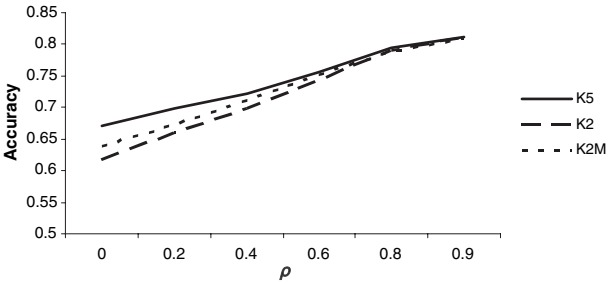

Figure 8 Accuracy of the prediction of the middle point for models with five knots (K5), two knots and regular (K2) or modified covariables (K2M).

The results from the simulation suggest a way to select knots in RRMS. Assume that accurate covariance functions are known. The extreme knots would bracket all points on the trajectory occurring in the data. Additional knots would be added so that correlations between adjacent knots would be in the range of 0.6–0.8. If parameters are to be estimated, two strategies are available. The first one is to initially use a large number of knots spaced equally, and then eliminate highly correlated knots. Such an approach may be expensive. Another possibility is to start with a small number of knots, estimate variances, and then add/deduct/shift knots while applying appropriate q values so that in the end the correlations between the adjacent points are in the range of 0.6–0.8. An extra heuristics could be adding extra points if the curves for variances/covariances appear too constrained or removing some points if curves appear erratic.

The above procedures assume that data points on the trajectory are equally distributed. If the distribution has distinct peaks, e.g. as in weights of beef cattle, the only knots to use may be those that correspond to those peaks. In this case, the use of the q parameter will not be important.

The procedures are derived assuming a certain structures of covariance functions, and especially equal variances at the knots. Indirectly, this assumes nearly linear changes in variances between any two adjacent knots. With rapid changes in variances, additional knots may be needed.

Parameter estimation using linear splines may be time consuming as many steps could be required to determine the optimal placement of knots and their parameters. In particular, the optimal knots for each effect may be different although practical differences from using same knots for every effect may be negligible.

Random regression models using linear splines may be especially useful for a national genetic

evaluation if covariance functions are already known. Then, parameters for RRMS can be easily constructed, and the evaluation can be run using the original model. A disadvantage of RRMS would be artefacts as described; however, these artefacts are more obvious than with polynomials and then easily controlled by changing the number and positions of knots.

Conclusions

Random regression models with linear splines have advantages and disadvantages. Advantages include easy creation of parameters and relatively robust computing. Disadvantages are lack of smoothness of breeding values and the need to adjust the number and positions of knots carefully while controlling the artefacts. RRMS may be useful as the first model in an implementation of a new genetic evaluation using the RRM. It is likely to be more accurate than repeatability or MTMs while being only marginally more complicated. If its limitations are found to be too restrictive, incrementally more complex models can be tried, e.g. using quadratic splines as in Meyer (2005b).

Acknowledgements

Helpful discussions with Bernt Guldbrandtsen and discussions and corrections by Karin Meyer are gratefully appreciated. Discussions with and assistance by Jesus Arango, Jarmila Bohmanova and Shogo Tsuruta are gratefully acknowledged.

References

Bohmanova J., Misztal I., Bertrand J.K. (2005) Studies on multiple trait and random regression models for genetic evaluation of beef cattle for growth. *J. Anim. Sci.*, **83**, 62–67.

Druet T., Jaffrézic F., Boichard D., Ducrocq V. (2003) Modeling lactation curves and estimation of genetic parameters for first lactation test-day records of French Holstein cows. *J. Dairy Sci.*, **86**, 2480–2490.

Iwaisaki H., Tsuruta S., Misztal I., Bertrand J.K. (2005) Genetic parameters estimated with multi-trait and linear spline-random regression models using Gelbvieh early growth data. *J. Anim. Sci.*, **83**, 499–506.

Kirkpatrick M., Lofsvold D., Bulmer M. (1990) Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics*, **124**, 979–993.

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.

Lidauer M., Mantysaari E.A., Stranden I. (2003) Comparison of test-day models for genetic evaluation of production traits in dairy cattle. *Livest. Prod. Sci.*, **79**, 73–86.

Meyer K. (2005a) Estimates of genetic covariance functions for growth of Angus cattle. *J. Anim. Breed. Genet.*, **122**, 73–85.

Meyer K. (2005b) Random regression analyses using B-splines to model growth of Australian Angus cattle. *Genet. Sel. Evol.*, **37**, 473–500.

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

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

Nobre P.R.C., Misztal I., Tsuruta S., Bertrand J.K., Silva L.O.C., Lopes P.S. (2003b) Genetic evaluation of growth in Nellore cattle by multiple trait and random regression models. *J. Anim. Sci.*, **81**, 927–932.

Robbins K.R., Misztal I., Bertrand J.K. (2005) A practical longitudinal model for evaluating growth in Gelbvieh cattle. *J. Anim. Sci.*, **83**, 29–33.

Robert-Granié C., Maza E., Rupp R., Foulley J.L. (2002) Use of fractional polynomial for modelling somatic cell scores in dairy cattle. In: Proceedings of the 7th World Congress of Genetics Applied to Livestock Production, Montpellier (France), 19–23 August 2002, CD-ROM, comm. no 16–05, Montpellier, France.

Schaeffer L.R. (1984) Sire and cow evaluation under multiple trait models. *J. Dairy Sci.*, **67**, 1567–1580.

Strabel T., Jamrozik J. (2002) The effect of incorrect estimated variance-covariance components on genetic evaluation of dairy cattle with random regression models. In: Proceedings of the 7th World Congress of Genetics Applied to Livestock Production, Montpellier, France, 19–23 August 2002, CD-ROM com. no 01–09, Montpellier, France.

Torres R.A., Quaas R.L. (2001) Determination of covariance functions for lactation traits on dairy cattle using random-coefficient regressions on B-splines. *J. Anim. Sci.*, **79**, Suppl. 1, 112, abstract.

Van der Werf M. (1997) Random Regressions in Animal Breeding. Course notes available at: http://www-personal.une.edu.au/~jvanderw/ranregrcourse.htm. Accessed 28 May 2005.

White I.M., Thompson R., Brotherstone S. (1999) Genetic and environmental smoothing of lactation curves with cubic splines. *J. Dairy Sci.*, **82**, 632–638.