Random regression models in the evaluation of the growth curve of Simbrasil beef cattle

R.R. Mota[1], L.F.A. Marques[2], P.S. Lopes[1], L.P. da Silva[1], A.M. Hidalgo[3], C.D.S. Leite[1] and R.A. Torres[1]

[1]Departamento de Zootecnia, Universidade Federal de Viçosa, Viçosa, MG, Brasil
[2]Departamento de Zootecnia, Centro de Ciências Agrárias, Universidade Federal do Espírito Santo, Alegre, ES, Brasil
[3]Animal Science Department, Wageningen University, Wageningen, The Netherlands

Corresponding author: R.R. Mota
E-mail: rreismota@hotmail.com

ABSTRACT. Random regression models were used to estimate the types and orders of random effects of (co)variance functions in the description of the growth trajectory of the Simbrasil cattle breed. Records for 7049 animals totaling 18,677 individual weighings were submitted to 15 models from the third to the fifth order including as fixed effects sex, contemporary group, feeding regimen, and type of reproduction and as random effects additive direct genetic effect, animal permanent environment, maternal additive genetic effect, and maternal permanent environment. The best-fit model presented order five to additive direct genetic effect, animal permanent environment, and maternal additive effect, with 6 classes of residual variances, and the maternal permanent environment effect was not significant, likely owing to the low average number of calves per cow. However, the model chosen for the growth curve presents three classes of residual variances, because even not showing the best fit

it is more parsimonious, in addition to promoting a more realistic estimate of heritability.

Key words: (Co)variance functions; Residual variance; Heritability

INTRODUCTION

Selection has been responsible for progressive changes in the Brazilian beef cattle industry (Alencar, 2004). As a consequence, the country occupies a leading position in beef cattle production and in the international market of meat exports as one of the highest producers and exporters (IBGE, 2011).

The Simbrasil breed is among the genetic resources available in Brazil. This breed was introduced in 1945, when crossbreeding between the Simmental and Guzerá breeds began in the municipality of Muqui, Espírito Santo State. It is a synthetic breed generated by alternate breeding from the Simmental and Zebu genotypes. Its performance attributes include robust milk production and carcasses of good classification in addition to strong growth index, ease of management, and adaptation to tropical regions (ABCRSS, 2011).

In Brazil, genetic evaluations for growth traits in beef cattle usually make use of multi-trait analyses for the prediction of breeding values of animals. Generally, weight measurements are taken at various ages, which adjusts the weights for the standard ages and excludes weights that lie outside of the pre-established age intervals. This convention reduces the accuracy of predicted breeding values (Meyer, 2004). Thus, this information is not used effectively in the genetic evaluation of animals (Sarmento et al., 2010).

Recently, Henderson Jr. (1982) and Laird and Ware (1982) have developed random regression models (RRM) as alternatives that describe the trajectory at all points instead of finite points and, unlike the traditional methodologies, consider that productive records of the same individual at different ages to refer to the same trait with different genetic correlations between them (Biassus et al., 2010). The adjustment order for random effects must be carefully analyzed with models more parsimonious because a perfect adjustment could lead to an increase in computational demand and susceptibility to numeric errors (Legarra et al., 2004).

Meyer (1999) has reported strong agreement between analyses when adjusting a phenotypic curve by means of covariance functions both for the direct additive effect and for those that separate the direct additive genetic effect from the permanent environmental effect. Marques et al. (1999), working with Simmental beef cattle, have verified increases in residual variances as animals grew old. This increase may be due to a scale effect in the weight of the animals, which could cause bias in the evaluations if a model of homogeneity of residual variance is adopted (Sarmento et al., 2010), i.e., the use of such model means that variances attributed to random effects are not partitioned correctly (Sousa et al., 2008). The objective of the present study was to estimate the types and orders of random effects of (co)variance functions at the description of the growth trajectory of Simbrasil beef cattle using RRM adjusted by Legendre polynomials.

MATERIAL AND METHODS

The data used in this research were obtained from Simbrasil beef cattle and provided

by the Associação Brasileira de Criadores das Raças Simental e Simbrasil [Brazilian Association of Simbrasil and Simmental Cattle Farmers (ABCRSS)] in the municipality of Cachoeiro de Itapemirim in Espírito Santo State, Brazil. The data were part of a file with 18,677 individual weighings corresponding to 7049 animals born between 1974 and 2006. Weight values were obtained at quarterly intervals by the Controle de Desenvolvimento Ponderal (Weight Development Control) of ABCRSS on 96 beef cattle farms in several states in Brazil. In addition to weight at birth, weight values were collected at approximately 100, 205, 365, 450, 540, and 730 days of age (Figure 1). The averages of the weights at each age assessed are presented in Figure 2. The descriptive statistics for the characteristics studied are presented in Table 1.
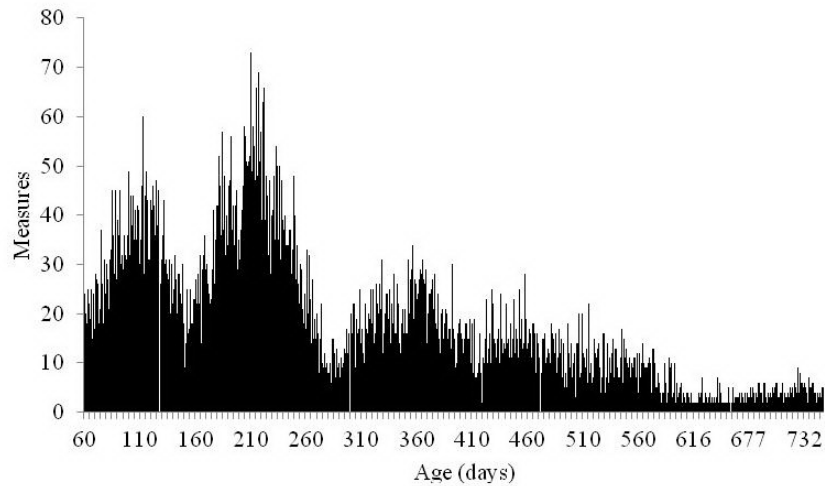

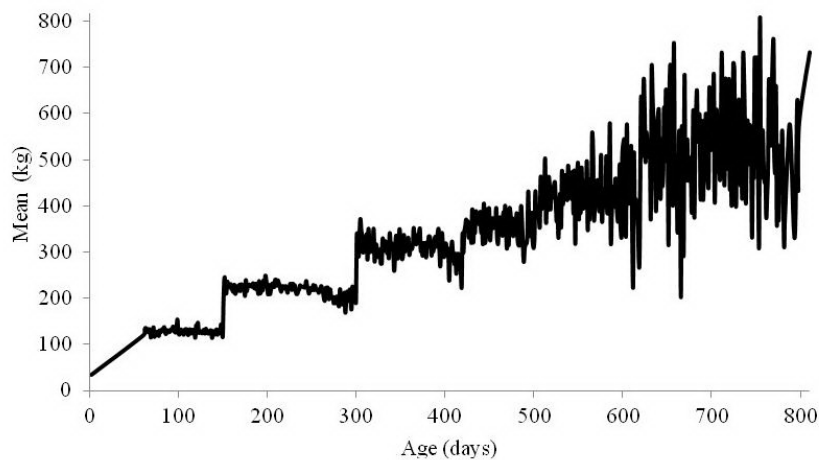**Figure 1.** Number of measures for each age contained in the database.

Figure 2. Mean weights from birth to 810 days old.

Table 1. Descriptive statistics of the weights collected at different ages.

Standard age (inteval of age in days)	Minimum (kg)	Maximum (kg)	Mean (kg)	CV (%)
At birth	25	55	38.599	15.142
100 (60-149)	66.64	208.41	127.772	20.053
205 (150-299)	109.180	335.690	221.444	19.770
365 (300-419)	164.690	532.700	310.868	25.407
450 (420-499)	182.420	647.660	352.531	28.141
540 (500-619)	199.90	769.80	423.048	29.538
730 (620-810)	203.25	950.22	526.36	32.209

CV = coefficient of variation.

All products obtained through embryo transfer were excluded from the weights file, as were as animals with unknown mothers, to make possible the study of maternal effects. Overall, the herds included in the study do not use a mating season, and births are distributed uniformly throughout the year. The structure of residual variances was modeled considering 1 homogeneous class and 6 classes with heterogeneity of variances. The classes used were CL1: for homogeneous; CL2: 1-299 and 300-810 days; CL3: 1-59, 60-299, and 300-810 days; CL4: 1-59, 60-299, 300-499, and 500-810 days; CL5: 1-59, 60-299, 300-419, 420-499, and 500-810 days; CL6: 1-59, 60-149, 150-299, 300-419, 420-499, and 500-810 days, and CL7: 1-59, 60-149, 150-299, 300-419, 420-499, 500-619, and 620-810 days of age, with CLm as the modeling of variance heterogeneity in which m represented the number of classes. The numerator relationship matrix was built with information from all generations available and had 9864 animals.

In matrix notation, the model and its respective assumptions can be described as follows:

$$y = Xb + Z_1a + Z_2m + W_1c + W_2q + e \qquad \text{(Equation 1)}$$

$$E\begin{bmatrix} y \\ a \\ c \\ m \\ q \\ e \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} e\ \text{Var}(a) = K_a \otimes A,\ \text{Var}(c) = K_c \otimes I_{N_d},\ \text{Var}(m) = K_m \otimes A, \text{Var}(q) = K_q \otimes I_{N_m}, \text{Var}(e) = R;$$

where y is a vector of N observations referring to N_d animals; b is a vector that contains the fixed effects; a is a k_a x N_d vector of direct additive genetic random regression coefficients; N_d is the total number of animals in the numerator relationship matrix between the individuals (A); m is the k_m x N_d vector of maternal additive genetic random regression coefficients; c is a k_C x N_d vector of animal permanent environmental random regression coefficients; q is a k_q x N_m vector of maternal permanent environmental random regression coefficients, with N_m equal to the number of females that have progeny with records; e is a vector of random effects, and X, Z_1, Z_2, W_1, and W_1 refer to the incidence matrices of fixed regression coefficients, direct additive genetic, maternal additive genetic, animal permanent environmental, and maternal permanent environmental random regression coefficients. K_a, K_m, K_c, and K_q are, respectively, the (co)variance matrices, between the direct additive genetic, maternal additive genetic, animal permanent environmental, and maternal permanent environmental random regression coefficients. I_{Nd} is an identity matrix of dimension N_d; $\otimes$ is the Kronecker

product, and *R* is a diagonal matrix of residual variances with elements that depend on the structure used. The covariance between the direct additive genetic and maternal effects was assumed to be zero.

The fixed effects used were sex, contemporary group (herd, year, and birth season), type of mating (natural mating or artificial insemination), and feeding regimen (RA1 = animals on pasture with mineralized salt; RA2 = animals on pasture with feed supplementation; RA3 = animals in feedlot). The fixed regressions of order three (quadratic) and the random ones of direct additive, maternal additive, animal permanent environmental, and maternal permanent environmental effects were represented by continuous functions, the covariables of which have been described by Legendre polynomials with orders varying from three to five.

The components of (co)variances and the genetic parameters were estimated using the method of restricted maximum likelihood with the WOMBAT software (Meyer, 2007). The models with various fitting orders for the polynomials were compared using the logarithm of the likelihood function (log L), Akaike information criterion (AIC), and Bayesian information criterion (BIC). AIC and BIC impose penalties according to the number of parameters to be estimated and were defined as follows:

$$AIC = -2\log L + 2\ p, \qquad\qquad\qquad \text{(Equation 2)}$$

$$BIC = -2\log L + p\ \log[N - r(X)] \qquad\qquad \text{(Equation 3)}$$

where *p* is the number of parameters estimated, *N* is the number of observations, *r(X)* is the rank of the incidence matrix of the fixed effect of the model, and *log L* is the logarithm of the likelihood function.

RESULTS AND DISCUSSION

The results of the analyses using various RRM are presented in Table 2. Overall, log L values were lower in less parameterized models (see Table 2); however, some models with a large number of parameters presented lower log L values. Baldi et al. (2010), working with Canchim cattle, have observed that in models in which all random effects (direct additive, maternal additive, animal permanent environmental, and maternal permanent environmental) are included, the amount of adjustment is compromised. With respect to AIC and BIC, we found a tendency for lower values in more parameterized models; however, when the number of parameters was greater, the model did not always provided better fit.

The model with the best fit according to the log L, AIC, and BIC criteria was the one that presented order five for direct additive, animal permanent environmental, and maternal additive effect with 46 parameters (Leg555; see Table 2). Meyer (2003) has recommended model Leg4353_32 with 69 parameters in Angus to estimate covariance functions of weight from birth until 3000 days of age. Boligon et al. (2010), working with weight records of Nellore females for the estimation of covariance functions for weights from birth until adulthood, have observed that according to AIC and BIC criteria, the most suitable models for the description of the trajectory are models Leg4454_5 and Leg4363_5 with 50 and 48 parameters, respectively.

Table 2. Order of polynomial for direct additive genetic effects (k_a) and maternal (k_m) and permanent environmental effects of animal (k_c) and maternal (k_q), number of parameters (N_p), logarithm of the likelihood function (log L) criterion, Akaike information criterion (AIC), and Schwarz Bayesian criterion (BIC).

Model1	k_a	k_c	k_m	k_q	N_p	log L	AIC	BIC
Leg33	3	3	-	-	13	-70,777.024	141,580.048	141,681.222
Leg34	3	4	-	-	17	-70,130.613	140,295.226	140,427.530
Leg45	4	5	-	-	26	-69,049.085	138,150.170	138,352.518
Leg55	5	5	-	-	31	-71,779.071	143,620.142	143,861.404
Leg333	3	3	3	-	19	-63,485.277	127,008.554	127,154.374
Leg3333	3	3	3	3	25	-63,484.272	127,018.544	127,210.414
Leg343	3	4	3	-	23	-63,719.302	127,484.604	127,661.124
Leg443	4	4	3	-	27	-63,748.911	127,551.822	127,759.042
Leg444	4	4	4	-	31	-63,337.531	126,737.062	126,974.982
Leg454	4	5	4	-	36	-66,745.128	133,562.256	133,841.318
Leg455	4	5	5	-	41	-64,029.661	128,141.322	128,455.990
Leg555	5	5	5	-	46	-62,608.523	125,309.046	125,662.088
Leg4444	4	4	4	4	41	-62,976.208	126,034.416	126,349.084
Leg4544	4	5	4	4	46	-62,883.232	125,858.464	126,211.506
Leg5555	5	5	5	5	61	-63,785.973	127,693.946	128,162.110

Leg k_a k_c k_m k_p = order of the covariance function for the direct additive genetic effects (k_a), the permanent environment of animal (k_c), direct additive maternal (k_m), and maternal permanent environmental (k_q).

Such divergence may be due to the nonsignificance of the maternal permanent environment in the present study. This nonsignificance can be attributed to the low number of calves per cow, which was, on average, 1.7. This result is in accordance with those of Pelicioni et al. (2003), who failed to estimate successfully maternal permanent environmental effects when working with a database of Guzerá cattle with an average number of calves per mother of approximately 1.5 calves for weight at birth, 1.3 calves for ages 30 to 210 days, and 1.2 calves from that age onward. Similar results with the same breed have been observed by Bittencourt et al. (2002) and Tanaka et al. (2009), who also worked with an average number of calves per cow below 2.

The adjustment of polynomials for the direct additive, maternal additive, and maternal permanent environmental regression increased as its degree rose. Nevertheless, convergence problems occurred when polynomial orders above five were used, probably associated with eigenvalues close to zero in the random regression coefficients matrix, which according to Nobre et al. (2003) demonstrates that the fitting order is satisfactory.

When the random effects were considered, the model with homogeneity of residual variances (Leg555_1; Table 3) showed the worst fit, with improvement overall of log L, AIC, and BIC criteria, indicating that the behavior of the residual variance is different throughout the growth trajectory. Meyer (1999) and Baldi et al. (2010) have achieved similar results using beef cattle weights that varied from birth to the adulthood. According to the results from the statistical criteria of log L, AIC, and BIC, model Leg3555_6 with 51 parameters generated the best fit (see Table 3). However, this model, with 6 residual classes, resulted in significant increases in the residual and phenotypic variances. Some models, both those with few and those with many classes of residual variances (Figures 3 to 5), presented biased estimates of total and direct additive heritabilities, generally showing values superior to those found in the literature for beef cattle, according to a review by Lira et al. (2008). Therefore, the model with three classes of residual variances was chosen for the residual variance despite lacking the best fitting according to the statistical criteria described above. This model was the most parsimonious and provided the most realistic heritability estimates. These results are in accordance

with those of Baldi et al. (2010), who have published models with a higher number of residual classes that provided better fit; however, these models increased the residual and phenotypic variances, altering the direct heritability.

Table 3. Classes of residual variance, number of parameters (N_p), log-likelihood function (log L), Akaike information criterion (AIC), and Schwarz Bayesian criterion (BIC) for model Leg3555.

Classes	k_a	k_c	k_m	k_q	N_p	log L	AIC	BIC
Leg555_1	5	5	5	-	46	-62,608.523	125,309.046	125,662.088
Leg555_2	5	5	5	-	47	-58,637.578	117,369.156	117,729.872
Leg555_3	5	5	5	-	48	-57,733.999	115,563.998	115,932.390
Leg555_4	5	5	5	-	49	-57,733.502	115,565.004	115,941.107
Leg555_5	5	5	5	-	50	-57,690.182	115,480.364	115,864.106
Leg555_6	5	5	5	-	51	-57,580.401	115,262.802	115,654.218
Leg555_7	5	5	5	-	52	-58,289.810	116,683.620	117,082.712

Leg k_a k_c k_m k_p = order of the covariance function for the direct additive genetic effects (k_a), the permanent environment of animal (k_c), direct additive maternal (k_m), and maternal permanent environmental (k_q).
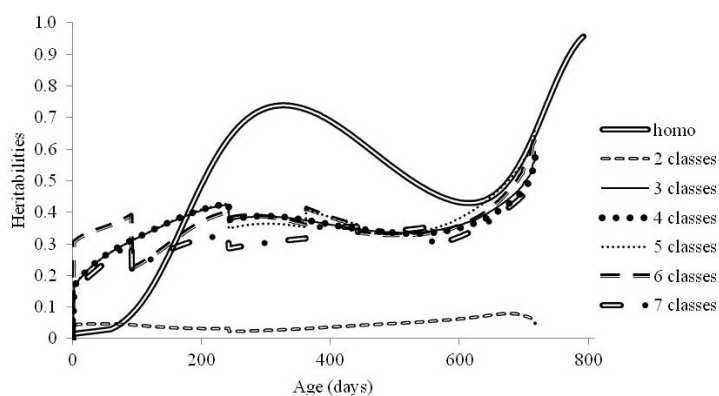

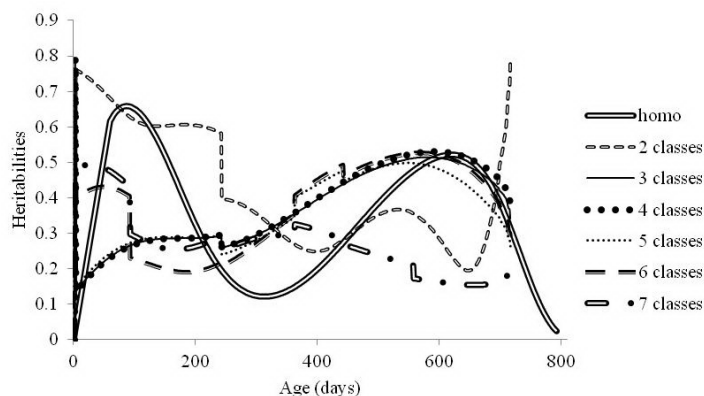**Figure 3.** Total heritabilities in different classes of residual variance.

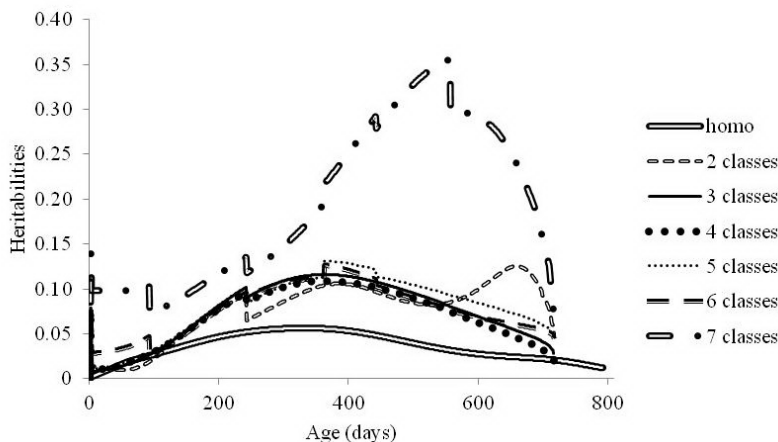**Figure 4.** Direct heritabilities in different classes of residual variance.

Figure 5. Maternal heritabilities in different classes of residual variance.

We recommend model Leg555_3 as the growth trajectory of Simbrasil beef cattle, and it differs from those for other breeds. For instance, Albuquerque and Meyer (2001), when estimating covariance functions in Nellore cattle for weights from birth to approximately 2 years of age, have used model Leg4463, and Baldi et al. (2010), working with Canchim cattle with records of weight from birth to adult age, have recommended model Leg4352.

Heterogeneity of residual variance with three classes must be considered using an RRM of order five for direct additive genetic effect, animal permanent environmental effect, and maternal genetic effect to model the changes according to the age of animals in Simbrasil beef cattle.

ACKNOWLEDGMENTS

REFERENCES

ABCRSS (2011). Associação Brasileira dos Criadores da Raça Simental e Simbrasil. Available at [http://simentalsimbrasil. org.br/?p=raca-simbrasil]. Accessed March 9, 2011.

Albuquerque LG and Meyer K (2001). Estimates of covariance functions for growth from birth to 630 days of age in Nelore cattle. *J. Anim. Sci.* 79: 2776-2789.

Alencar MM (2004). Perspectivas para o Melhoramento Genético de Bovinos de Corte no Brasil. In: Reunião Anual da Sociedade Brasileira de Zootecnia Sociedade Brasileira de Zootecnia, Campo Grande, 358-367.

Baldi F, Albuquerque LG and Alencar MM (2010). Random regression models on Legendre polynomials to estimate genetic parameters for weights from birth to adult age in Canchim cattle. *J. Anim. Breed. Genet.* 127: 289-299.

Biassus IO, Cobuci JA, Costa CN and Rorato PRN (2010). Persistence in milk, fat and protein production of primiparous Holstein cows by random regression models. *Braz. J. Anim. Sci.* 39: 2617-2624.

Bittencourt TCC, Rocha JCMC, Lobo RB and Bezerra LF (2002). Estimação de componentes de (co)variâncias e predição de DEP's para características de crescimento pós-desmama de bovinos da raça Nelore, usando diferentes modelos estatísticos. *Arq. Bras. Med. Vet. Zootec.* 54: 303-308.

Boligon AA, Mercadante ME, Forni S, Lobo RB, et al. (2010). Covariance functions for body weight from birth to maturity in Nellore cows. *J. Anim. Sci.* 88: 849-859.

Henderson CR Jr (1982). Analysis of covariance in the mixed model: higher-level, nonhomogeneous, and random regressions. *Biometrics* 38: 623-640.

IBGE (2011). Instituto Brasileiro de Geografia e Estatística. Available at [http://www.ibge.gov.br/home/estatistica/ indicadores/agropecuaria/producaoagropecuaria/abate-leite-couro-ovos_201004_1.shtm]. Accessed March 30, 2011.

Laird NM and Ware JH (1982). Random-effects models for longitudinal data. *Biometrics* 38: 963-974.

Legarra A, Misztal I and Bertrand JK (2004). Constructing covariance functions for random regression models for growth in Gelbvieh beef cattle. *J. Anim. Sci.* 82: 1564-1571.

Lira T, Rosa EM and Garnero AV (2008). Parâmetros genéticos de características de produtivas e reprodutivas em zebuínos de corte. *Cienc. Anim. Bras.* 9: 1-22.

Marques LFA, Pereira JCC, Oliveira HN, Pereira CS, et al. (1999). Componentes de (co)variância e parâmetros genéticos de características de crescimento da raça Simental no Brasil. *Arq. Bras. Med. Vet. Zootec.* 51: 363-370.

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 (2003). First estimates of covariance functions for lifetime growth of Angus cattle. *Proc. Assoc. Adv. Anim. Breed. Genet.* 15: 395-398.

Meyer K (2004). Scope for a random regression model in genetic evaluation of beef cattle for growth. *Livest. Prod. Sci.* 86: 69-83.

Meyer K (2007). WOMBAT - A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J. Zhejiang Univ. Sci. B* 8: 815-821.

Nobre PR, Misztal I, Tsuruta S, Bertrand JK, et al. (2003). Analyses of growth curves of nellore cattle by multiple-trait and random regression models. *J. Anim. Sci.* 81: 918-926.

Pelicioni LC, Queiroz SA and Albuquerque LG (2003). Estimativas de parâmetros genéticos para pesos ao nascer e mensais até 450 dias em bovinos Guzerá. *Arch. Lat. Prod. Anim.* 11: 34-39.

Sarmento JLR, Torres RA, Lobo RNB, Albuquerque LG, et al. (2010). Modelos de regressão aleatória na avaliação genética do crescimento de ovinos da raça Santa Inês. *Braz. J. Anim. Sci.* 39: 1723-1732.

Sousa JER, Silva MA, Sarmento JLR, Sousa WH, et al. (2008). Homogeneidade e heterogeneidade de variância residual em modelos de regressão aleatória sobre o crescimento de caprinos Anglo-Nubianos. *Pesq. Agropec. Bras.* 43: 1725-1732.

Tanaka ALR, Carvalheiro R, Fries LA and Queiroz SA (2009). Comparação de critérios de seleção para precocidade de crescimento em bovinos da raça Guzerá. *Rev. Bras. Zootec.* 38: 284-291.