



## Heterogeneous genetic (co)variances in simulated closed herds under selection

Daniela Andressa Lino-Lourenço<sup>1\*</sup>, Carlos Antonio Lopes de Oliveira<sup>2</sup>, Elias Nunes Martins<sup>2</sup>, Meiby Carneiro de Paula Leite<sup>3</sup>, Fabiana Martins Costa Maia<sup>4</sup> and Alexandra Inês dos Santos<sup>2</sup>

<sup>1</sup>Programa de Pós-graduação em Zootecnia, Universidade Estadual de Maringá, Av. Colombo, 5790, 87020-900, Maringá, Paraná, Brazil. <sup>2</sup>Departamento de Zootecnia, Centro de Ciências Agrárias, Universidade Estadual de Maringá, Maringá, Paraná, Brazil. <sup>3</sup>Centro de Ciências Agrárias, Ambientais e Biológicas, Universidade Federal do Recôncavo da Bahia, Cruz das Almas, Bahia, Brazil. <sup>4</sup>Departamento de Zootecnia, Universidade Tecnológica Federal do Paraná, Dois Vizinhos, Paraná, Brazil. \*Author for correspondence. E-mail: dandilino@gmail.com

**ABSTRACT.** Assuming that selection in closed herds can promote reduction in additive genetic variance, multiple regression models were used to estimate this change in additive genetic (co)variance component, over the years when the selection was done. Weights at 550 days (W550) were studied using simulated data of herds submitted to 20 years of selection. (Co)variance components were estimated assuming that the weight at 550 days was a new trait every five years, by multiple-trait analyses involving four traits in the animal model. Three multiple regression equations were fitted—RMI, RMM, RMF—estimating thus the additive genetic (co)variance components for the 20 years of selection and eight years prior to the selection process. The initial years of each generation of selection were used as a covariate in the RMI. In the RMM, intermediate years were used, and the final years were considered in the RMF. The equations showed high coefficients of determination. However, there was no difference in the adjustment between the models. It was observed that the multiple regression models can be used in the estimation of genetic (co)variance components, when heteroscedasticity is assumed over time due to the selection process.

**Keywords:** (co)variance components, heteroscedasticity, Bayesian inference, multiple regression.

### (Co)variâncias genéticas heterogêneas em rebanhos simulados fechados sob seleção

**RESUMO.** Assumindo que a seleção em rebanhos fechados pode promover a redução da variância genética aditiva, foi estudada a possibilidade do uso de um modelo de regressão múltipla para estimar os componentes de (co)variância genética aditiva, ao longo dos anos em que a seleção foi praticada. Foram utilizados dados simulados de peso aos 550 dias em dez rebanhos de bovinos de corte submetidos à seleção por 20 anos. Assumindo que a cada cinco anos o peso aos 550 dias era uma nova característica, por meio de análises multivariadas envolvendo quatro características, em um modelo animal, foram estimados componentes de (co)variância. Foram ajustadas três equações de regressão múltipla, RMI, RMM, RMF, estimando componentes de (co)variância genética aditiva para 20 anos de seleção e para oito anos anteriores à seleção. Na RMI, foram utilizados os anos iniciais de cada geração de seleção, para a RMM os anos intermediários e na RMF os anos finais como covariável. As equações apresentaram altos coeficientes de determinação, no entanto, não houve diferença de ajuste entre os três modelos. Observou-se que os modelos de regressão múltipla podem ser usados na estimação dos componentes de (co)variância genética quando se admite heterocedasticidade ao longo do tempo, causada pela seleção.

**Palavras-chave:** componentes de (co)variância, heterocedasticidade, inferência Bayesiana, regressão múltipla.

### Introduction

In animal improvement programs, particularly in genetic evaluations, it is frequently assumed that the variances remain constant over generations of selection, however in closed herds, it is expected that selection changes not only the mean of traits but also their additive genetic variance.

Considering that the mean and variance describe statistically the basic characteristics of a population, it is expected changes in these parameters when the

population structure is modified. The difficulty to assume heteroscedasticity over the generations lies in obtaining precise estimates of the (co)variance components, since the number of observations at each heteroscedasticity class is reduced with increasing the number of classes. With a very large number, the computational effort to estimate the (co)variance components becomes very high, and if the classes are less numerous the genetic connections are weaker, leading to inaccurate estimates of the components.

According to Carneiro Júnior et al. (2007), if the different types of heteroscedasticity are not taken into account, the accuracy can be reduced when estimating the variance components, leading to errors in the evaluation of the animals, resulting in lower genetic progress. In this way, this research aimed at studying the possibility of using the regression analysis to estimate the components of additive genetic (co)variance for the weight at 550 days, considering the heteroscedasticity over the year of selection in simulated and closed herds with overlapping of generations.

## Material and methods

The data set consisted of simulated herds of beef cattle, using Fortran language in F90 compiler, subjected to 20 years of selection. The breeding herd was formed by 1,500 dams and 38 bulls, keeping a bull-dam ratio 1:40. From the second year, was simulated the use of artificial insemination in 50% of the cows, with ten bulls used for artificial insemination and 19 bulls for natural breeding. Birth rate established was 90% and survival was 95% by the start of reproduction. The culling rate was variable and determined by the number of empty cows at the end of breeding season. Only the primiparous cows could remain in the breeding herd if they were empty. The young animals could be included into this herd, from 22 months-old.

It was simulated the weaning weight (WW) and the weight at 550 days (W550), as well as the breeding values for the direct effect of weaning weight and weight at 550 days, and the maternal effect of weaning weight. Nevertheless, only the trait weight at 550 days was the focus of the present study.

The following simulated identifiable environmental effects were considered: sex (male or female), birth season (early, middle or end of birth season), year of birth and age of dam at calving (aod), in months. To assign the levels of the identifiable effects of environment, was used the uniform distribution for the effects of sex (2 levels) and birth season (3 levels). The levels of the effects of year of birth and age of dam at calving were selected by sorting from a uniform distribution only for the animals belonging to the base population. The magnitude of the effects of age of dam at calving (aod) was determined by the regression equations  $WW = 0.1(aod) - 0.0004(aod)^2$  and  $W550 = 0.06(aod) - 0.00024(aod)^2$ .

The breeding values were simulated from a multivariate normal distribution with dimension equal to 3. The residuals were generated through a

multivariate normal distribution of dimension 2, whereas the maternal permanent environmental effects were simulated from a normal distribution. In order to generate the initial breeding values, it was used the structure of genetic (co)variance described below,

$$G_0 = \begin{bmatrix} \sigma_{aD}^2 & \sigma_{aDm} & \sigma_{aDS} \\ \sigma_{aDm} & \sigma_m^2 & \sigma_{aSm} \\ \sigma_{aDS} & \sigma_{aSm} & \sigma_{aS}^2 \end{bmatrix} = \begin{bmatrix} 165 & -55 & 135 \\ -55 & 95 & 0 \\ 135 & 0 & 215 \end{bmatrix}$$

where:

$\sigma_{a_i}^2$  is the additive genetic variance for the trait  $i$ ,  $i=WW(D)$  and  $W550(S)$ ;

$\sigma_m^2$  is the maternal additive genetic variance of the trait WW;

$\sigma_{aDS}$  is the genetic covariance between the direct effects of WW and W550;

$\sigma_{a_m}$  is the genetic covariance between the direct effects of traits  $i$  with the maternal effects of WW, with  $i=WW(D)$ ,  $W550(S)$ .

The residual variances adopted were 365 and 450  $kg^2$  for WW and W550, respectively, and the value of residual covariance was 150  $kg^2$ .

Each year the genetic evaluation was undertaken to guide the selection, using the mixed model equations (HENDERSON, 1984) for prediction of breeding values using a multi-trait analysis in the MTDFREML software (Multiple Trait Derivative Free Restricted Maximum Likelihood) (BOLDMAN et al., 1995).

For the classification of the candidates to selection, it was used an empirical selection index with weights of 0.3, 0.3, and 0.4, respectively, for the direct and maternal breeding values for WW and direct for W550. The mating of animals selected was randomly performed, with restriction to control the increased levels of inbreeding, by preventing mating between parents and offspring and between full and half-siblings.

After establishing the conditions of simulation and selection, ten replications were performed, totaling ten herds with average of 31,198 animals. In order to obtain the genetic (co)variances between the years of birth, for the weight at 550 days, first it was performed the estimation of variance components of W550 of each generation of selection (5 years) and covariance of W550 between generations, using the software MTGSAM (Multiple Trait Gibbs Sampling in Animal Model), (VAN TASSEL; VAN VLECK, 1995), in a multi-trait analysis, where the weight at 550 days was treated as four distinct traits according to the generation in which the animal was born, with

W55-g1 being the weight at 550 days for the animals born from the year 1 to year 5; W550-g2 for those born between years 6 and 10; W550-g3, between 11 and 15; and W550-g4, between 16 and 20. It was considered that one animal only had an observation for the trait referring to the class it belonged, with missing data in the traits corresponding to the other classes. Sex, birth season, and year of birth were considered identifiable environmental effects, and aod was considered as covariate.

For the identifiable environmental effects, it was assumed non-informative priors, with uniform initial distribution, i.e., all the values have the same probability of occurrence. The distribution of genetic effects and residuals was considered normal multivariate. In the case of breeding values, it was considered the known covariance structure given by the relationship matrix. The definitions for the elements of the animal model and of the joint distribution of  $Y$  (observations),  $a$  (genetic values) and  $e$  (residuals) were, respectively:

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} X_1 & 0 & 0 & 0 \\ 0 & X_2 & 0 & 0 \\ 0 & 0 & X_3 & 0 \\ 0 & 0 & 0 & X_4 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 & 0 & 0 \\ 0 & Z_2 & 0 & 0 \\ 0 & 0 & Z_3 & 0 \\ 0 & 0 & 0 & Z_4 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

where:

$y_1, y_2, y_3, y_4$  are the vectors of observations, for the weight at 550 days, referring to generations 1, 2, 3 and 4, respectively;

$X_1, X_2, X_3, X_4$  are the incidence matrices of the identifiable environmental effects, for the weight at 550 days, referring to generations 1, 2, 3 and 4, respectively;

$\beta_1, \beta_2, \beta_3, \beta_4$  are the vectors of identifiable environmental effects, for the weight at 550 days, referring to generations 1, 2, 3 and 4, respectively;

$Z_1, Z_2, Z_3, Z_4$  are the incidence matrices of the random effects, for the weight at 550 days, referring to generations 1, 2, 3 and 4, respectively;

$a_1, a_2, a_3, a_4$  are the vectors of direct genetic effects, for the weight at 550 days, referring to generations 1, 2, 3 and 4, respectively;

$e_1, e_2, e_3, e_4$  are the vectors of random effects, for the weight at 550 days, referring to generations 1, 2, 3 and 4, respectively;

with joint distribution:

$$\begin{bmatrix} y \\ a \\ e \end{bmatrix} \sim NMV \left\{ \begin{bmatrix} X\beta \\ 0 \\ 0 \end{bmatrix}; \begin{bmatrix} ZGZ'+R & ZG & R \\ GZ' & G & 0 \\ R & 0 & R \end{bmatrix} \right\}$$

where:

$$G = G_0 \otimes A$$

$\otimes$  is the Kronecker product;

$G_0$  is the matrix of genetic (co)variance of  $i$  generations, for the weight at 550 days old, given as follows:

$$G_0 = \begin{bmatrix} \sigma_{a_1}^2 & \sigma_{a_1 a_2} & \sigma_{a_1 a_3} & \sigma_{a_1 a_4} \\ \sigma_{a_2 a_1} & \sigma_{a_2}^2 & \sigma_{a_2 a_3} & \sigma_{a_2 a_4} \\ \sigma_{a_3 a_1} & \sigma_{a_3 a_2} & \sigma_{a_3}^2 & \sigma_{a_3 a_4} \\ \sigma_{a_4 a_1} & \sigma_{a_4 a_2} & \sigma_{a_4 a_3} & \sigma_{a_4}^2 \end{bmatrix}$$

where:

$\sigma_{a_i}^2$  is the additive genetic variance of the generation  $i$ ;  $i = 1, 2, 3$  and  $4$ ;

$\sigma_{a_i a_j}$  is the genetic covariance between the generations  $i$  and  $j$ ;  $i = 1, 2, 3$  and  $4$ , and  $j = 1, 2, 3$  and  $4$ .

$$R = R_0 \otimes I$$

where:

$I$  is the identity matrix of order equal to the number of animals;

$R_0$  is the residual variance matrix of the  $i$  generations, for the weight at 550 days old, given as follows:

$$R_0 = \begin{bmatrix} \sigma_{e_1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{e_2}^2 & \dots & 0 \\ \vdots & \vdots & \sigma_{e_3}^2 & \vdots \\ 0 & 0 & \dots & \sigma_{e_4}^2 \end{bmatrix}$$

where:

$\sigma_{e_i}^2$  is the residual variance of the generation  $i$ ;  $i = 1, 2, 3$  and  $4$ .

For the genetic (co)variance components it was assumed that  $G$  has inverted Wishart distribution ( $IW$ ).

For each of the replications (herds) it was generated a Gibbs chain of 1,000,000 cycles with samples being stored every 100 cycles, after the elimination of 50,000 initial cycles, generating chains of 9,500 samples of the (co)variance components. The convergence of Gibbs sampling chains was verified using the method from Heidelberger and Welch (1983), which, in the first instance, compares the Gibbs chain with a hypothetical chain of stationary distribution, then verifies whether the means of the samplings are

within a threshold of the credibility interval established. The test is available in CODA library (Convergence Diagnosis and Output Analysis, version 0.4), developed by Cowles et al. (1995), and implemented in the R software (R DEVELOPMENT CORE TEAM, 2009).

In order to obtain the genetic (co)variances for W550 for the years when the selection was performed, it was employed a multiple regression model in which:

$$Y = b_0 + b_1x_1 + b_2x_2 + b_3x_1^2 + b_4x_2^2 + b_5x_1x_2,$$

where:

- Y is the (co)variance component to be estimated;
- $b_0$  is the general constant;
- $b_1, b_2, b_3, b_4$  and  $b_5$  are the regression coefficients;
- $x_1$  and  $x_2$  are the years of birth of the animals.

Through programs developed in Fortran language, F90 compiler, and using the sampling of genetic (co)variance components for generation of selection provided by the software MTGSAM, it was possible to estimate the multiple regression coefficients under three conditions: RMI (Initial Multiple Regression) in which the years representing the generations were the initial ones; RMM (Intermediate Multiple Regression) in which the intermediate years was used; and the RMF (Final Multiple Regression) in which the final years were the representative in the X matrix.

The estimation was carried out by the method of generalized least squares (GLS), as:

$$\beta = (X'V^{-1}X)^{-1} X'V^{-1}y,$$

where:

- $\beta$  is the vector of the regression coefficients;

X is the incidence matrix of the fixed effects of the year representing each generation of selection. For the RMI, the values were 1, 6, 10, and 16; for the RMM were 3, 8, 13, and 18; for the RMF the years considered were 5, 10, 15, and 20.

V is the matrix of (co)variance between the genetic (co)variance components between the four generations, estimated by the software MTGSAM;

y is the vector of posterior means, estimated by the software MTGSAM;

The multiple regression models produce a response surface that allow, from the existing points, to make inferences about any point within the studied range.

For each herd, three equations were fitted, which enabled obtaining genetic (co)variance matrices of W550 between the years of birth of the animals, under the situations RMI, RMM, RMF, considering that the (co)variance components of W550 for the animals born in the eight years prior to the first year

of selection, and from the first to the fifth year of selection, have received values of variance and covariance of the first year of selection. For the year of animal's birth, it was considered total amplitude of 28 years.

In order to observe the dispersion of the (co)variance components and to know whether the values estimated by the regression were consistent with the reality and whether they were within the interval set at 90% of the total probability, it was calculated the credibility interval at 90% using the software R (R DEVELOPMENT CORE TEAM, 2009). A comparison was performed between the credibility intervals of genetic variances between generations to check if in fact the variances could be considered heterogeneous.

The inference power on the genetic variance components for the years of birth and genetic (co)variance between the years of animal's birth, obtained through multiple regression, was tested by the coefficient of determination, represented by the ratio between the sum of the squares of the (co)variance values estimated by the regression for the four generations, and the sum of the squares of the (co)variance values provided by the MTGSAM, for the same four generations.

The difference between the inference power of the equations RMI, RMM and RMF was tested via Bayesian methodology and implemented in the software WinBugs (SPIEGELHALTER et al., 2003).

## Results and discussion

Additive genetic variance components estimated per generation for W550, using the software MTGSAM, are listed in Table 1. From the generation 1 to 4, the mean genetic variances were 159.04, 127.13, 81.25, and 67.49, respectively. It was observed a reduction in genetic variances for W550 with the advancing generations of selection, and reduction of genetic covariance between the generations, proportional to the distance between them.

For Gomez-Raya and Burnside (1990), the higher the accuracy of selection, the greater is the reduction of variances, considering the accuracy of selection defined as the genetic correlation between the true breeding value and the predicted one. According to Quinton and Smith (1995), the sharp decrease in genetic variances is related with the use of information from relatives to predict breeding values, by the use of the relationship matrix that can generate co-selection, increasing the probability of selecting related animals, so that the losses in genetic variability are associated to high levels of inbreeding

in populations under selection. In the present case, the losses were less sharp because the inbreeding was kept at low levels by preventing the mating between parents and offspring and between siblings.

**Table 1.** Additive genetic variance components estimated per generation for W550, using the software MTGSAM.

Components*	Herds				
	1	2	3	4	5
$\sigma_{a_1}^2$	148.70	182.61	139.62	182.28	159.45
$\sigma_{a_2}^2$	132.76	119.56	127.19	116.69	137.02
$\sigma_{a_3}^2$	66.60	117.10	63.08	81.20	59.47
$\sigma_{a_4}^2$	66.01	91.68	56.20	59.73	48.66
$\sigma_{a_1a_2}$	77.13	78.92	74.77	77.25	77.96
$\sigma_{a_1a_3}$	28.68	66.50	19.08	48.32	27.03
$\sigma_{a_1a_4}$	19.63	26.73	13.91	21.47	15.14
$\sigma_{a_2a_3}$	26.42	50.73	19.03	42.45	24.08
$\sigma_{a_2a_4}$	24.42	25.16	17.50	19.86	12.30
$\sigma_{a_3a_4}$	12.02	29.90	17.67	18.37	8.16
Components*	6	7	8	9	10
$\sigma_{a_1}^2$	144.40	168.93	151.08	144.95	168.40
$\sigma_{a_2}^2$	112.12	128.12	116.14	142.1081	139.61
$\sigma_{a_3}^2$	78.28	88.38	71.95	91.71	94.70
$\sigma_{a_4}^2$	61.87	84.82	55.49	79.80	70.68
$\sigma_{a_1a_2}$	65.12	91.86	63.88	98.82	97.04
$\sigma_{a_1a_3}$	36.36	53.09	22.70	42.66	53.18
$\sigma_{a_1a_4}$	20.21	31.40	18.52	27.08	23.59
$\sigma_{a_2a_3}$	25.16	49.53	16.54	47.07	50.94
$\sigma_{a_2a_4}$	26.73	32.61	10.99	24.92	24.91
$\sigma_{a_3a_4}$	17.56	22.78	10.21	18.52	13.60

\* $\sigma_{a_i}^2$  is the additive genetic variance for the generation i and  $\sigma_{a_{ij}}$  the additive genetic covariance between generations i and j.

Ferraz Filho et al. (2002) found value of 225.06 kg<sup>2</sup> for the additive genetic variance for the weight at 550 days, in Tabapuá animals. In Nelore animals, Van Melis et al. (2003) registered for the genetic variance of weight at 550 days, value of 205.60 kg<sup>2</sup>. These values are higher than those found herein, but the cited authors considered variance homogeneity over the years.

**Table 2.** Credibility intervals (IC) for genetic variances of generation of selection.

Herd	Components*			
	$\sigma_{a_1}^2$	$\sigma_{a_2}^2$	$\sigma_{a_3}^2$	$\sigma_{a_4}^2$
1	121.24 - 178.91	106.48 - 160.29	49.05 - 86.32	47.50 - 87.06
2	154.52 - 211.47	97.40 - 143.98	92.38 - 143.36	67.71 - 118.71
3	113.84 - 166.80	102.60 - 153.88	46.44 - 83.33	40.00 - 75.00
4	152.33 - 214.14	94.53 - 141.36	61.31 - 102.65	42.07 - 80.71
5	133.45 - 187.71	109.36 - 166.07	42.77 - 78.21	33.09 - 66.41
6	119.98 - 170.87	88.57 - 137.89	59.11 - 99.33	44.47 - 81.34
7	142.51 - 196.55	104.73 - 153.70	68.99 - 109.96	63.30 - 109.41
8	125.14 - 178.32	92.74 - 141.92	52.21 - 95.10	38.47 - 74.39
9	122.27 - 169.38	119.38 - 166.67	70.68 - 114.83	58.69 - 104.57
10	140.94 - 196.98	114.93 - 166.39	72.19 - 120.87	51.76 - 92.09
Mean	132.62 - 187.11	103.07 - 153.22	61.51 - 103.40	48.71 - 88.97

\* $\sigma_{a_i}^2$  is the additive genetic variance for the generation i.

The credibility intervals (IC) for the estimates of additive genetic variance for the generation of selection can be seen in Table 2.

It was observed that for the herds 1 and 7, the IC were close to the variances of generations 3 and 4; in the herd 2, the variances for the generation 2 and 3 were very close; and in the herd 9 the variances for the first and second generations were closer, indicating homoscedasticity in these cases. In the rest of the herds there is heterogeneity of variances, since the IC does not have very close values.

According to Winkelman and Schaeffer (1988), the non-consideration of variance heterogeneity in different herds, of different regions, different levels of management and production, and with varied genetic composition, can lead to a biased process of genetic evaluation and selection, which could affect the choice of animals that will produce lower genetic gain when used in genetic improvement programs.

Likewise, if the population is under continuous process of selection, each generation has changes in genetic variances and covariances modifying hence the genetic response. The non-consideration of this factor can also be a source of bias in genetic evaluations. Thus, from the (co)variance components for generation of selection, the three multiple regression equations were fitted and allowed the estimation of the heterogeneous components of genetic (co)variance for each herd, per year of animal's birth.

When tested the degree of fitness and the inference power of the multiple regression used to estimate the components for each year based on the components for generation, it was obtained high coefficients of determination ( $r^2$ ), for all the herds, with mean of 0.946, 0.958, and 0.958 and standard deviation of 0.032, 0.025, and 0.025 for RMI, RMM and RMF, respectively. The coefficients of determination are listed in Table 3.

**Table 3.** Values of  $r^2$  for each herd, mean  $r^2$  and standard deviation for RMI, RMM, RMF.

Herd	$r^2$		
	RMI	RMM	RMF
1	0.914	0.935	0.935
2	0.950	0.942	0.942
3	0.905	0.935	0.935
4	0.985	0.981	0.981
5	0.934	0.957	0.957
6	0.945	0.957	0.957
7	0.970	0.973	0.973
8	0.961	0.978	0.978
9	0.927	0.949	0.949
10	0.967	0.977	0.977
Mean	0.946	0.958	0.958
Standard deviation	0.032	0.025	0.025

It was possible to estimate 28 variance components, from the ten (co)variance components for the four generations of selection. It was also estimated 378 covariance components, totaling 406 components for each one of the three models, in each herd. The estimation of a great number of components by using programs developed for genetic evaluation is computationally costly, but it can be avoided using multiple regression models to access all or just components of interest. Thus, in the process of genetic evaluation, using the BLUP (best linear unbiased predictor), the relationship matrix can be properly weighted according to the structure of genetic (co)variance assumed.

The genetic variances estimated for each year of birth, per herd, did not present the same trend of the genetic variances per generation of selection, which were reduced as the generations have passed, for all the herds. From the year 1 to year 8, i.e., in the eight years prior to starting the process of selection, the variances were considered constant and equal to the first year of selection. It was verified a reduction in the variances, per year, from the 13<sup>th</sup> to the 28<sup>th</sup> year, for the herds 1, 3, 5, 9 and, 10. For the herd 2, it decreased from the year 13 to 18, increasing from 19 to 28. In the herd 4, the reduction was from the Year 13 to 23, increasing from 24 to 28. In the herds 6 and 8, it was reduced from the 13<sup>th</sup> to the 25<sup>th</sup> year and increased from the 26<sup>th</sup> to the 28<sup>th</sup>. For the herd 7, the decrease was from 13<sup>th</sup> to 21<sup>st</sup> year, increasing from the 22<sup>nd</sup> to the 28<sup>th</sup>. Nevertheless, when the average trend of the variances between all the herds was verified, it was observed the same pattern obtained for generation of selection, being reduced over the years (Table 4). This reduction can be explained by the moderate increase of inbreeding and reduction in the effective population size, which were altered from 0 to 6.4% and from 148.25 to 24.48, respectively, along the years of selection.

**Table 4.** Mean additive genetic variance between the ten herds for RMI, RMM, RMF.

Component*	Mean		
	RMI	RMM	RMF
$\sigma^2_{a_9}$	201.82	180.04	161.21
$\sigma^2_{a_{10}}$	190.72	169.78	150.20
$\sigma^2_{a_{11}}$	180.04	159.95	139.86
$\sigma^2_{a_{12}}$	169.78	150.55	130.18
$\sigma^2_{a_{13}}$	159.95	141.57	121.16
$\sigma^2_{a_{14}}$	150.55	133.02	112.80
$\sigma^2_{a_{15}}$	141.57	124.89	105.11
$\sigma^2_{a_{16}}$	133.02	117.19	98.08
$\sigma^2_{a_{17}}$	124.89	109.92	91.71
$\sigma^2_{a_{18}}$	117.19	103.07	86.00
$\sigma^2_{a_{19}}$	109.92	96.64	80.96
$\sigma^2_{a_{20}}$	103.07	90.65	76.57
$\sigma^2_{a_{21}}$	96.64	85.07	72.85
$\sigma^2_{a_{22}}$	90.64	79.92	69.80
$\sigma^2_{a_{23}}$	85.07	75.20	67.40
$\sigma^2_{a_{24}}$	79.92	70.91	65.67
$\sigma^2_{a_{25}}$	75.20	67.03	64.60
$\sigma^2_{a_{26}}$	70.91	63.59	64.19
$\sigma^2_{a_{27}}$	67.04	60.57	64.45
$\sigma^2_{a_{28}}$	63.59	57.98	65.37

\* $\sigma^2_{a_i}$  is the additive genetic variance for the year  $i$ , where  $i=9$  is the first year of selection.

In general, it was found a reduction in the genetic covariances between the years within the same generation, as the years became distant, which can be explained by the decrease in the connection between the animals born in years apart. Once it was observed a reduction in the variances, the covariances reduction is also an indicative that, according to Gianola et al. (1992), when there is variance heterogeneity, surely there must be covariance heterogeneity.

Comparing the trend of genetic variances over the generations with the variances over the years, it could be verified that the RMI regression was the one that best estimated the variance from the year 1 to the year 9, being the ninth year in which the selection began to be practiced. However, it underestimated the variances of the remaining years. The RMM and RMF regressions in general overestimated the genetic variance for W550 prior to the beginning of selection process, and have estimated better the variances during the process of selection.

When it was verified whether the credibility intervals for genetic variation in generation of selection had included the estimates of genetic variances for year of animal's birth, it was observed that for the RMI, on average, only 47.5% of the estimates were within the respective ICs; for the RMM, this number rose to 84.5%; for the RMF, 100% of the estimates were within the respective

ICs. It points out that some estimates obtained for genetic variance for year, through RMI and RMM may not represent their real values.

The main goal of a regression model is to predict values of  $y$  for a given value of  $x$ , within the data range, even when the values of  $x$  do not have assigned observed value. In the same way, this principle can be applied when the modeling is for (co)variances. Therefore, instead of estimating variances for all the generations, using a more complex modelling when heteroscedasticity is assumed, in analyses that require a long processing time, it may be possible only the estimation of extreme values and the use of multiple regression for interpolation, thus reducing the computational time used.

Some authors have sought alternative methods to obtain more adequate estimates of variance components, since this process becomes inaccurate when involving great data sets or a great number of components in a same model. In this way, Rekaya et al. (2003) presented the methodological development of a structural model (SM) to estimate the genetic (co)variances for milk production in Holstein cattle in five regions of United States. In this study, the SM has proven to be advantageous in situations where there was a limited number of information to estimate the genetic (co)variances mainly due to the weak genetic links between the regions. Another advantage the authors found for this modeling was the possibility of incorporate external information to the data, such as climatic conditions, management and genetic similarities between the regions, in order to explore patterns in genetic (co)variances. The estimates of genetic (co)variances and heritability obtained by the SM were much more precise than those estimated by the multi-trait model, since the first model had retained lower amount of parameters.

High values of  $r^2$ , obtained for the RMF and the occurrence of 100% of estimates of genetic variance of W550 for year within the credibility interval for the variance in generation of selection, show that the regression has been effectively fitted, indicating that the variance components in each year and the genetic covariance among the years can be estimated accurately using the multiple regression. Thus, from the moment when the selection started to be practiced, the multiple regression is a tool that can be used to estimate the components under situations where the (co)variance heterogeneity exists and should be considered. For years prior to selection, the most indicated in this case is to use the estimates obtained through the software MTGSAM.

## Conclusion

The results indicated that the multiple regression model, considering the (co)variance components as a function of the generations under selection, using the final years of each generation as covariates, here represented by RMF, can be implemented successfully in Bayesian procedures to estimate (co)variance components when the heteroscedasticity due to the effect of selection is assumed over time.

## References

- BOLDMAN, K. G.; KRIESE, L. A.; VAN VLECK, L. D.; VAN TASSELL, C. P.; KACHMAN, S. D. **A manual for use of MTDFREML**. A set of programs to obtain estimates of variances and covariances [Draft]. Lincoln: USDA, Agricultural Research Service, 1995.
- CARNEIRO JÚNIOR, J. M.; ASSIS, G. M. L.; EUCLYDES, R. F.; TORRES, R. A.; LOPES, P. S. Estimação de componentes de variância utilizando-se inferência Bayesiana e freqüentista em dados simulados sob heterogeneidade de variâncias. **Revista Brasileira de Zootecnia**, v. 36, n. 5, p. 1539-1548, 2007.
- COWLES, K.; BEST, N.; VINES, K. **Convergence diagnosis and output analysis**. Cambridge: MRC Biostatistics Unit, 1995. (Version 0.40).
- FERRAZ FILHO, P. B.; RAMOS, A. A.; SILVA, L. O. C.; SOUZA, J. C.; ALENCAR, M. M.; MALHADO, C. H. M. Tendência genética dos efeitos direto e materno sobre os pesos à desmama e pós-desmama de bovinos da raça Tabapuã no Brasil. **Revista Brasileira de Zootecnia**, v. 31, n. 2, p. 635-640, 2002.
- GIANOLA, D.; FOULLEY, J. L.; FERNANDO, R. L.; HENDERSON, C. R.; WEIGEL, K. A. Estimation of heterogeneous variances using empirical bayes methods: theoretical considerations. **Journal of Dairy Science**, v. 75, n. 10, p. 2805-2823, 1992.
- GOMEZ-RAYA, L.; BURNSIDE, E. B. The effect of repeated cycles of selection on genetic variance, heritability, and response. **Theoretical Applied Genetics**, v. 79, n. 4, p. 568-574, 1990.
- HEIDELBERGER, P.; WELCH, P. D. Simulation run length control in the presence of an initial transient. **Operations Research**, v. 31, n. 6, p. 1109-1144, 1983.
- HENDERSON, C. R. **Applications of linear models in animal breeding**. Ontario: University of Guelph, 1984.
- QUINTON, M.; SMITH, C. Comparison of evaluation-selection systems for maximizing genetic response at same level of inbreeding. **Journal of Animal Science**, v. 73, n. 8, p. 2208-2212, 1995.
- R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, 2009. Available from: <<http://www.R-project.org>>. Accessed on: Mar. 22, 2010.
- REKAYA, R.; WEIGEL, K. A.; GIANOLA, D. Bayesian estimation of parameters of a structural model for genetic covariances between milk yield in five regions of the

United States. **Journal of Dairy Science**, v. 86, n. 5, p. 1837-1844, 2003.

SPIEGELHALTER, D.; THOMAS, A.; BEST, N.; LUNN, D. **WinBUGS User Manual**: Version 1.4. Cambridge: MRC Biostatistics Unit, 2003.

VAN MELIS, M. H.; ELER, J. P.; SILVA, J. A. V.; FERRAZ, J. B. S. Estimaco de parmetros genticos em bovinos de corte utilizando os mtodos de mxima verossimilhana restrita e R. **Revista Brasileira de Zootecnia**, v. 32, n. 6, p. 1624-1632, 2003.

VAN TASSEL, C. P.; VAN VLECK, L. D. **A manual for use of MTGSAM**. A set of FORTRAN programs to apply Gibbs sampling to animal models for variance

component estimation [Draft]. Lincoln: USDA, Agricultural Research Service, 1995.

WINKELMAN, A.; SCHAEFFER, L. R. Effect of heterogeneity of variance on dairy sire evaluation. **Journal of Dairy Science**, v. 71, n. 11, p. 3033-3039, 1988.

*Received on August 19, 2010.*

*Accepted on April 20, 2011.*

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.