Multi-Trait analysis of growth traits: fitting reduced rank models using principal components for Simmental beef cattle

Efetividade da redução da dimensão da matriz de covariância do efeito genético direto na avaliação genética do crescimento em bovinos Simental

Rodrigo Reis Mota¹ Edson Vinicius Costa¹ Paulo Sávio Lopes¹ Mayses Nascimento²³ Luciano Pinheiro da Silva³ Fabyano Fonseca e Silva¹ Luiz Fernando Aarão Marques⁴

ABSTRACT

The aim of this research was to evaluate the dimensional reduction of additive direct genetic covariance matrices in genetic evaluations of growth traits (range 100-730 days) in Simmental cattle using principal components, as well as to estimate (co)variance components and genetic parameters. Principal component analyses were conducted for five different models-one full and four reduced-rank models. Models were compared using Akaike information (AIC) and Bayesian information (BIC) criteria. Variance components and genetic parameters were estimated by restricted maximum likelihood (REML). The AIC and BIC values were similar among models. This indicated that parsimonious models could be used in genetic evaluations in Simmental cattle. The first principal component explained more than 96% of total variance in both models. Heritability estimates were higher for advanced ages and varied from 0.05 (100 days) to 0.30 (730 days). Genetic correlation estimates were similar in both models regardless of magnitude and number of principal components. The first principal component was sufficient to explain almost all genetic variance. Furthermore, genetic parameter similarities and lower computational requirements allowed for parsimonious models in genetic evaluations of growth traits in Simmental cattle.

Key words: computational demand, genetic parameters, heritability.

INTRODUCTION

The success of beef production depends upon several important economic traits,
including breeding to achieve effective genetic variability through selection (MASCIOLI et al., 2000). Typically, a livestock breeder’s decision-making is intuitive; in other words, animals are selected according to specific phenotypes, but trait components and linkages with others are neglected. In this context, principal component analysis may help clarify relationships among variables and improve decision-making (MASCIOLI et al., 2000; NASCIMENTO et al., 2014; OLIVEIRA et al., 2014).

Principal components analysis is a multivariate approach that transforms the original set of variables (the principal components), which are linear combinations of all original variables (MEYER & KIRKPATRICK, 2005), and has elucidated important properties in animal breeding studies. In addition, the principal components approach discounts less discriminant variables (because they are correlated with other variables), and thus, is a useful tool for rank reduction (CRUZ & REGAZZI, 1997).

Different approaches (e.g., principal components and factor analyses) designed to reduce data dimensions in multivariate models have been proposed to facilitate interpretation and computational efficiency. When these approaches are directly applied to the covariance matrices of multivariate mixed models, such models are called “reduced rank models” (BIGNARDI et al., 2012; MEYER, 2005). MEYER (2007a) proposed rank reduction to obtain models more parsimonious (with reduced number of parameters) that are able to generate estimates of variance components and genetic parameters with the same accuracy as models that includes all traits (e.g., the standard full rank model). However, studies using such approach on beef cattle genetic evaluation are still scarce.

Therefore, the aim of this research was to determine the effectiveness of direct genetic effect covariance matrix rank reduction on growth traits (weights at 100, 205, 365, 450, 550, and 730 days of age) in a genetic evaluation of Simmental cattle by using principal components analyses, as well as to estimate genetic parameters of each trait considered in the study.

MATERIALS AND METHODS

The data used in this study refers to Simmental cattle body weight (BW) records from 485 farms and was provided by the Brazilian Association of Simmental and Simbrasil Breeders (ABCRSS) located at Cachoeiro de Itapemirim, Espirito Santo State, Brazil. Records from 60 to 819 days of age from animals born between 1974 and 2006 were used. Dataset was composed of adjusted weights for cattle at 100 (W100), 205 (W205), 365 (W365), 450 (W450), and 730 (W730) days of age, such that weights at 60-149, 150-299, 300-419, 420-499, 500-619, and 620-819 days of age were used to obtain the W100, W205, W365, W450, W550, and W730, respectively.

The BWs were adjusted using the following equation: PPi = Pi + \[\frac{(Pi - Pj)}{(Di)} \times I_i\], where PPi is the standardized weight at age i, Pi is the observed weight closest to standard age Ii, Pj is the previously observed weight at Pi, Ii is the standard age, and Di is the days of age related to the observed weight Pi. Index i varies from i = 1 (at 100 days) to i = 6 (at 730 days).

Records outside the intervals given by the contemporary group averages (+3 standard deviations) were excluded. After filtering, the dataset had 29,510 records for 10,659 animals, sons of 1,289 sires and 7,332 cows. Number of observations (average ± standard deviation) were 6,236 (130.76 ± 40.26), 7,938 (222.56 ± 60.38), 6,070 (338.47 ± 83.77), 3,666 (380.50 ± 90.34), 3,468 (441.10 ± 104.51), and 2,132 (562.72 ± 141.59) for W100, W205, W365, W450, W550, and W730, respectively.

It was used an animal model with contemporary group as a fixed effect, defined as the same sex, year, and birth period (1 = January to March, 2 = April to June, 3 = July to September and 4 = October to December). Classes of cows’ age (days) at calving was a covariate (class 1, from 700 to 1,109; class 2, from 1,110 to 1,519; class 3, from 1,520 to 1,919; and class 4, from 1,920 to 2,340). The following were random effects: direct additive genetic, maternal additive genetic, permanent maternal environment, and residual. In matrix form, the model is described as

\[ y = X\beta + Z_d\delta + Z_m\mu + Z_p + e \]  

where y is the observations vector; \( \beta \) is the fixed effects vector; \( \delta \) is the vector of random direct additive genetic effects; \( \mu \) is the vector of random maternal additive genetic effects; \( p \) is the vector of random permanent maternal environment effects; \( X \) is the incidence matrix of fixed effects; \( Z_d \), \( Z_m \), and \( Z_p \) are the incidence matrices of the following random effects: direct additive genetic, maternal additive genetic, and permanent maternal environment, respectively, and e is the vector of residual effects.
Considering the independence and normality of \( d, m, p, \) and \( e \) random effects, the following assumptions were made:

\[
\begin{pmatrix}
\mathbf{y} \\
\mathbf{d} \\
\mathbf{m} \\
\mathbf{p} \\
\mathbf{e}
\end{pmatrix} =
\begin{pmatrix}
\mathbf{X}
\mathbf{0}
\mathbf{0}
\mathbf{0}
\mathbf{0}
\end{pmatrix}
\begin{pmatrix}
\mathbf{A} & \mathbf{G}_d & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{A} & \mathbf{G}_m & \mathbf{0} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{I}_n \otimes \mathbf{Q} & \mathbf{0} & \mathbf{0} \\
\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_n \otimes \mathbf{R}
\end{pmatrix}
\]

where \( \mathbf{G}_d \) is the (co)variance matrix of random direct additive genetic effects; \( \mathbf{G}_m \) is the (co)variance matrix of random maternal additive genetic effects; \( \mathbf{Q} \) is the (co)variance matrix of random permanent maternal environment effects; \( \mathbf{R} \) is the (co)variance matrix of random residual effects, \( \mathbf{A} \) is the numerator relationship matrix, \( \mathbf{I}_v \) is the identity matrix with order \( v \), where \( v \) is the number of mothers, \( \mathbf{I}_n \) is the identity matrix with order \( n \), where \( n \) is the number of observations, and \( \otimes \) is the Kronecker product operator.

Being a covariance matrix of direct additive genetic effects with order \( k \times k \), \( \mathbf{G}_d \) can be decomposed into \( \mathbf{G}_d = \mathbf{E} \mathbf{A} \mathbf{E}' \) where \( \mathbf{A} \) is the diagonal matrix of \( \mathbf{G}_d \) eigenvalues with \( \lambda_i \) for \( i \) varying from 1 to \( k \) and \( \mathbf{E} = [\mathbf{e}_1 | \mathbf{e}_2 | … | \mathbf{e}_k] \) is the matrix of eigenvectors (MEYER & KIRKPATRICK, 2005). Thus, using \( k = 6 \), the \( \mathbf{G}_d \) matrix has the original size (called the M6 model), but considering \( k = 2, 3, 4, \) and \( 5 \), we have the more parsimonious reduced rank models M2, M3, M4, and M5, respectively. Therefore, for \( k = 6 \), there are 21 parameters (6 variances at the diagonal and 15 covariances outside the diagonal) to estimate \( \mathbf{G}_d \) and for \( k = 2, 3, 4, \) and \( 5 \) the parameters are reduced to 3, 6, 10, and 15, respectively. The model fit was performed via REML using the software WOMBAT (MEYER, 2007b). Models were compared using Akaike information (AIC) and the Schwarz Bayesian information (BIC) criteria.

RESULTS AND DISCUSSION

Although M6 presented lower AIC and BIC values (Table 1), the goodness of fit between models was generally similar, indicating the possibility of a more parsimonious model for genetic evaluations of growth in Simmental cattle. LEGARRA et al. (2004) stated that more parsimonious models are preferred because of the smaller computational demands and lower susceptibility to numerical errors. MEYER (2007a) showed that genetic variance between carcass traits in Angus cattle could be performed using more parsimonious models (i.e., considering a subset of genetic principal components).

Direct additive genetic variance in both models presented equivalent variance component progression along the days of weighting (Figure 1). This direct additive genetic variance behavior may have occurred because animals at advanced ages present genetic values with higher magnitudes, because of their higher phenotypic values (body weight). Therefore, the variance measure of genetic values (direct additive genetic variance) is also directly affected by this magnitude increase, the so-called “scale effect.” This effect can be observed in table 2 with regard to the magnitude of the first principal component coefficient, in which higher values are observed for advanced ages. Similar results were reported by MARQUES et al. (1999) and MOTA et al. (2013) working with Simmental beef cattle. In addition, BOLIGON et al. (2013) reported similar direct additive genetic variance component estimates between all tested models for Nellore cattle. According to these authors, fewer principal components are sufficient to explain almost all genetic variability.

Heritability estimates (Table 2) were higher in advanced ages and remained stable between the tested models (range 0.05-0.30). In general, animals

---

Table 1 - Number of parameters (\( N_p \)); Akaike information criterion (AIC) and Schwarz Bayesian information criterion (BIC) for all tested models.

<table>
<thead>
<tr>
<th>Models</th>
<th>( N_p )</th>
<th>AIC</th>
<th>BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>M6</td>
<td>84</td>
<td>-100,144.332</td>
<td>-100,482.632</td>
</tr>
<tr>
<td>M5</td>
<td>82</td>
<td>-100,142.328</td>
<td>-100,472.475</td>
</tr>
<tr>
<td>M4</td>
<td>78</td>
<td>-100,138.359</td>
<td>-100,452.401</td>
</tr>
<tr>
<td>M3</td>
<td>72</td>
<td>-100,132.694</td>
<td>-100,422.579</td>
</tr>
<tr>
<td>M2</td>
<td>64</td>
<td>-100,126.343</td>
<td>-100,384.019</td>
</tr>
</tbody>
</table>

M6: full rank model; M5, M4, M3, M2: reduced rank models in which the number represents the rank of the (co)variance genetic matrix. Bold font indicates best fit.
at mature ages were less sensitive to environmental effects, such that non-genetic variance (maternal environmental and residual) tended to be less relevant, which reduced the denominator of the heritability expression. MARQUES et al. (2000) reported similar heritability estimates at 100 (0.08), 205 (0.13), 365 (0.19), and 550 (0.24) days of age in Simmental beef cattle. Conversely, BOLIGON et al. (2013) and MEYER (2007a) studying growth traits through reduced rank models in Angus and Nellore cattle, respectively, did not observe significant changes in heritability estimates as reported in this study.

For all tested models, the first principal component explained more than 96% of the entire genetic variance (see Figure 1). The heritability estimates are presented in Table 2, along with the first principal component coefficient (FPCC) and the percentage of the direct additive genetic variance (%var) explained by the principal components for all tested models at 100 (W100), 205 (W205), 365 (W365), 450 (W450), 550 (W550), and 730 (W730) days old.

Table 2 - Heritability estimates ± standard errors, first principal component coefficient (FPCC), and the percentage of the direct additive genetic variance (%var) explained by the principal components for all tested models at 100 (W100), 205 (W205), 365 (W365), 450 (W450), 550 (W550), and 730 (W730) days old.

<table>
<thead>
<tr>
<th>BW</th>
<th>M6</th>
<th>M5</th>
<th>M4</th>
<th>M3</th>
<th>M2</th>
</tr>
</thead>
<tbody>
<tr>
<td>W100</td>
<td>0.05 ± 0.02</td>
<td>0.05 ± 0.02</td>
<td>0.05 ± 0.02</td>
<td>0.05 ± 0.02</td>
<td>0.05 ± 0.02</td>
</tr>
<tr>
<td>W205</td>
<td>0.12 ± 0.03</td>
<td>0.12 ± 0.03</td>
<td>0.12 ± 0.03</td>
<td>0.12 ± 0.03</td>
<td>0.12 ± 0.03</td>
</tr>
<tr>
<td>W365</td>
<td>0.20 ± 0.04</td>
<td>0.20 ± 0.04</td>
<td>0.20 ± 0.04</td>
<td>0.20 ± 0.04</td>
<td>0.20 ± 0.04</td>
</tr>
<tr>
<td>W450</td>
<td>0.24 ± 0.04</td>
<td>0.24 ± 0.05</td>
<td>0.24 ± 0.04</td>
<td>0.23 ± 0.05</td>
<td>0.23 ± 0.04</td>
</tr>
<tr>
<td>W550</td>
<td>0.24 ± 0.05</td>
<td>0.24 ± 0.05</td>
<td>0.24 ± 0.05</td>
<td>0.23 ± 0.05</td>
<td>0.23 ± 0.04</td>
</tr>
<tr>
<td>W730</td>
<td>0.30 ± 0.07</td>
<td>0.30 ± 0.07</td>
<td>0.30 ± 0.07</td>
<td>0.30 ± 0.06</td>
<td>0.30 ± 0.06</td>
</tr>
</tbody>
</table>

FPCC:
- W100: 0.0906
- W205: 0.2092
- W365: 0.3435
- W450: 0.4133
- W550: 0.4743
- W730: 0.6590

%var:
- 96.75
- 96.77
- 96.79
- 98.12
- 99.17

BW: body weights; M6: full rank model; M5, M4, M3, M2: reduced rank models in which the number represents the rank of the (co)variance genetic matrix.

Figure 1 - Genetic variance estimates for models M6, M5, M4, M3, and M2 in which the number represents the rank of the (co)variance genetic matrix for body weights at 100, 205, 365, 450, 550, and 730 days old.
Multi-Trait analysis of growth traits: fitting reduced rank models using principal components for Simmental beef cattle.

Direct additive genetic variance (Table 2). MEYER (2005), using simulated data, reported that the number of principal components included in reduced rank analyses depended on the size of the dataset (i.e., for a small datasets fewer components are needed than for larger datasets, in which a reduction in sample variance to alleviate increased prediction errors is needed).

Direct additive genetic correlation estimates for M6, M5, M4, M3, and M2 models were positive and high (>0.80) as shown in figure 2. In general, these estimates suggest that animal selection for body weight at any age implies an indirect selection response in the same direction as BW, being more efficient in close age groups. These results corroborate those of NEPHAWE (2004), BOLIGON et al. (2009), and BALDI et al. (2010).

CONCLUSION

It was concluded that the first principal component was sufficient to almost entirely explain the direct additive genetic variance. Despite the fact that models with more principal components fit better, because of the similarity in estimated genetic parameters and the lower computational demand, the use of more parsimonious models is indicated for the genetic evaluation of Simmental beef cattle.

Figure 2 - Genetic correlation estimates for M6, M5, M4, M3, and M2 in which the number represents the rank of the (co)variance genetic matrix at 100, 205, 365, 450, 550, and 730 days old.
ACKNOWLEDGEMENTS

The authors thank the Associação Brasileira de Criadores das Raças Simental & Simbrasil for providing the data.

REFERENCES


Ciência Rural, v.46, n.9, set, 2016.