Using Factor Analysis Modeling Multiple Traits in Genetic Improvement of Nelore Beef Cattle‡

M. J. Yokoo,[1] G. de los Campos,[2] G. J. M. Rosa,[3] F. F. Cardoso,[1]
B. P. Sollero,[1] L. L. Cardoso,[1] R. B. Lobo,[4] and L. G. Albuquerque[5]
[1]Embrapa South Animal Husbandry and Sheep Research Unit (CPPSul), Bagé, RS, Brazil,
[2]University of Alabama at Birmingham, Birmingham, AL, USA, [3]University of Wisconsin, Madison, WI, USA,
[4]University of São Paulo (USP), Ribeirão Preto, SP, Brazil, [5]São Paulo State University (UNESP), Jaboticabal, SP, Brazil.
‡ Paper with financial support by National Counsel of Technological and Scientific Development (CNPq-Brasil), Brasília.

ABSTRACT: Genetic parameters for ultrasound carcass and growth traits were estimated by factor analyses used as a special case of structural equation models in a Bayesian framework. Data were analyzed using the standard multi-trait mixed models with sire model (Model 1; SMTMs). The factor analyses (FA) were done by four alternative FA models. The results indicate that FA models could estimate breeding values of the bulls practically equal relative to the SMTMs. The FA models may reduce the ranking model and give a parsimonious estimation of genetic covariance matrices. Although the FA models may reduce covariance matrices ranks and give a parsimonious estimation of dispersion parameters, these models have to be tested in order to implement the benefits, as an alternative of SMTMs.
Keywords: animal breeding; carcass; genetic parameters; structural equation models; ultrasound

INTRODUCTION

The continuous increasing of records and traits in genetic evaluation schemes for beef cattle generally involve multi-trait mixed models analyses, which requires genetic links among these traits. It leads to statistical and computational difficulties in estimating the genetic (co)variance matrix needed to have accurate breeding values. Structural equation models (SEM, e.g., Wright 1921) are multivariate models adapted to obtain more parsimonious quantitative genetic mixed-effects models (Gianola and Sorensen (2004)). SEM can be understood like a term that does not denote a particular statistical technique, but a number of techniques and procedures used together aiming to model some covariance structure.

In animal breeding, specifically in this paper, SEM was used to model the estimated genetic and residual (co)variance matrix. These models can be viewed as an extension of the standard multi-trait mixed models (SMTM, e.g., Henderson and Quaas (1976)) that are capable of expressing functional networks among traits. Gianola and Sorensen (2004) discussed the use of recursive and simultaneous equation models (special case of SEM) acting on phenotypes. Alternatively, the Factor Analysis (FA) may be used as another special case of SEM to represent the genetic covariance matrix (Jöreskog (1970)). FA can be used to model genetic effects in the context of a multivariate linear mixed model for reducing the dimension of the estimated genetic (co)variance matrix, obtaining a more parsimonious model without reducing dimension of the original records (e.g., de los Campos and Gianola (2007)).

The objective of this study was to consider FA (a special case of SEM) acting on genetic and residual effects separately to estimate genetic and residual (co)variance parameters modeling traits and estimating genetic parameters in genetic improvement of Brazilian Nelore beef cattle.

MATERIALS AND METHODS

Data. Data of 2,700 animals were provided by the Nelore Breeding Program - Nelore Brazil (PMGRN) and collected from 2002 to 2004 on ten farms located in six Brazilian states. Animals were born from 2000 to 2002. The following real-time ultrasound carcass measures were collected: longissimus muscle area (LMA) and backfat thickness (BF), both obtained from a cross-sectional image on the longissimus dorsi muscle, measured between the 12th and 13th ribs; and rump fat thickness (RF), measured at the intersection between the gluteus medium and biceps femoris muscles located between the hooks and pin bones. Backfat thickness was evaluated at the 3/4 position from the chine bone end of the longissimus muscle, using the cross-sectional ribeye image. Other traits recorded included: body weight (BW), hip height (HH), both measured at the date of ultrasound scanning, and 450-days of age standardized scrotal circumference (SC). With the exception of SC, traits were measured in animals ranging from 480 to 629 days of age. A description of traits in Brazilian Nelore cattle are presented in Table 1.

Factor analysis. A vector of random variables ($\mathbf{u}_i$), in a standard FA model, can be described as a linear combination of fewer unobservable random variables called common factors ($\mathbf{f}_i$) with the unobservable incidence matrix ($\mathbf{\Lambda}$) of factor loadings plus a vector of trait-specific factors ($\boldsymbol{\delta}_i$) peculiar to each i. In compact notation, $\mathbf{u}_i = \Lambda\mathbf{f}_i + \boldsymbol{\delta}_i$ (Equation 1). Thus, to work with the entire data these equation can be written as, $\mathbf{u} = (\mathbf{I}_n \otimes \mathbf{\Lambda})\ \mathbf{f} + \boldsymbol{\delta}$ (Equation 2), where $\mathbf{u} = (\mathbf{u}'_1, \ldots, \mathbf{u}'_n)$, $\mathbf{f} = (\mathbf{f}'_1, \ldots, \mathbf{f}'_n)$, and $\boldsymbol{\delta} = (\boldsymbol{\delta}'_1, \ldots, \boldsymbol{\delta}'_n)$. This gives covariance matrix of $\mathbf{u}$ under the FA model like, $Cov(\mathbf{u}_i) = \mathbf{\Sigma}_u = \mathbf{\Lambda\Lambda'} + \mathbf{\Psi}$, with $\mathbf{\Psi}$ = Diag $\{\ \mathbf{\Psi}_i\ \}$ = diagonal matrix of specific variances. The marginal distribution of $\mathbf{u}_i$ is, $\mathbf{u}_i \overset{iid}{\sim} N[\mathbf{0}, \mathbf{\Lambda\Lambda}'+\mathbf{\Psi}]$.

Consider now a SMTM for p traits measured on each of n subjects, the equation for the entire data set is, $\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \boldsymbol{\varepsilon}$ (Equation 3), where $\mathbf{y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_n)'$, $\mathbf{X} = (\mathbf{X}'_1, \ldots, \mathbf{X}'_n)'$, $\mathbf{Z}$ = Diag $\{\mathbf{Z}_i\}$, $\mathbf{u} = (\mathbf{u}'_1, \ldots, \mathbf{u}'_n)'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \ldots, \boldsymbol{\varepsilon}'_n)'$. A standard probability assumption in quantitative genetics or the joint distribution of random

Table 1. Descriptive statistics of ultrasound carcass and growth traits in Nelore cattle.

Traits[§]	No. of records	Mean ± SD[¥]	No. of sires	No. of dams	No. CG[£]
LMA, cm^2	2,770	48.05 ± 8.36	231	2,552	243
BF, mm	2,577	1.87 ± 1.07	226	2,397	253
RF, mm	2,566	2.95 ± 1.94	226	2,384	252
SC, mm	1,340	245.87 ± 30.22	106	1,009	88
HH, cm	2,349	136.06 ± 5.04	226	2,308	250
BW, kg	2,942	339.69 ± 65.98	236	2,683	302

[§] LMA = longissimus muscle area; BF = backfat thickness; RF = rump fat thickness; SC = standardized scrotal circumferences at 450 days of age; BW and HH = weight and hip height obtained at the time of scanning, respectively.
[¥] SD = standard deviation.
[£] No.CG = number of contemporary groups.

effects is $p(\boldsymbol{\varepsilon},\mathbf{u}) = N[\boldsymbol{\varepsilon}|\mathbf{0},\mathbf{I}\otimes\mathbf{R}_0]N[\mathbf{u}|\mathbf{0},\mathbf{A}\otimes\mathbf{G}_0]$, where $\mathbf{R}_0$ and $\mathbf{G}_0$ are within-subject and within-sire residual and additive (co)variance matrices, respectively, $\mathbf{A}$ is the relationship matrix and $\otimes$ is the direct product operator.

To implement the FA as a special case of SEM to model the genetic and the residual (co)variance matrix, it was assumed that equation 2 holds the vector of random additive genetic effects ($\mathbf{u}$) in equation 3, likewise for the vector of random residual effects ($\boldsymbol{\varepsilon}$) so that, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}[(\mathbf{I}\otimes\boldsymbol{\Lambda})\mathbf{f}] + \mathbf{Z}\boldsymbol{\delta} + [(\mathbf{I}_\varepsilon\otimes\boldsymbol{\Lambda}_\varepsilon)\mathbf{f}_\varepsilon] + \boldsymbol{\delta}_\varepsilon$, where $\boldsymbol{\Lambda}$, $\mathbf{X}$ and $\mathbf{Z}$ are as before, $\boldsymbol{\beta}$ is a systematic effect and $\mathbf{f}$ and $\boldsymbol{\delta}$ are interpreted as vectors of common and specific additive genetic effects, respectively, similarly for the random residual effects ($\boldsymbol{\varepsilon}$). Combining the assumptions of the FA model described above with those of SMTM leads to the random effects joint distribution of, $p(\boldsymbol{\varepsilon},\mathbf{u}) = N[\boldsymbol{\varepsilon}|\mathbf{0},\mathrm{I}_\varepsilon\otimes(\Lambda_\varepsilon\Lambda'_\varepsilon+\Psi_\varepsilon)]N[\mathbf{u}|\mathbf{0},\mathrm{A}\otimes(\Lambda\Lambda'+\Psi)]$, where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}_\varepsilon$ are the matrix of additive genetic and residual factor loads, respectively, $\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}_\varepsilon$ are the diagonal matrix of specific additive genetic and residual variances, respectively.

Statistical analyses. Data were analyzed using a SMTM sire model (Model 1: SMTMs). Systematic effects included: contemporary groups (defined as animals of the same sex, except for SC, born in the same herd, year and season, and reared within the same management group), age of animal at scanning (linear effect, except for SC), and age of dam (linear and quadratic effects for BF, RF, HH, and BW).

The FA were done in Model 2 (FA2F, with two factors for the matrix of additive genetic and residual loading factors) and in Model 3 (FA3F, with three factors for the matrix of additive genetic and residual loading factors). Based on the results of these Models 2 and 3 more models were generated (Models 4 and 5). Model 4 (FA2G) had two factors only for the matrix of additive genetic loading factors and in the residual matrix it was considered as in SMTMs. The Model 5 (FA2R) had two factors only for the matrix of residual loading factors, and the additive genetic matrix was considered as in SMTMs.

All models were implemented in a Bayesian framework. Inferences were based on 160,000 samples from the posterior distribution obtained after discarding 40,000 samples as burn in, and thinned every 10th iteration. Convergence was checked by inspection of trace plots of dispersion parameters. After estimating all the genetic parameters the *Spearman* and *Pearson* correlation was calculated between the sire breeding values for all traits to compare all these models.

Table 2. Estimated parameters to compare different Bayesian models.

Models[†] Parameters[§]	SMTMs	FA2F	FA3F	FA2G	FA2R
DIC	**37,874.7**	38,036.8	38,918.0	**37,898.3**	38,043.4
pD	456.4	**455.2**	566.0	**431.2**	499.7
Mean(L)	**-18,709.2**	-18,790.7	-19,176.0	**-18,733.5**	-18,771.8

[†]SMTMs=Standard multi-trait mixed models with sire model; FA2F=Model with two factors for the matrix of additive genetic and residual loading factors; FA3F=Model with with three factors for the matrix of additive genetic and residual loading factors; FA2G=Model with two factors only for the matrix of additive genetic loading factors and in the residual matrix it was considered as in SMTMs; FA2R=Model with two factors only for the matrix of residual loading factors, and the additive genetic matrix was considered as in SMTMs;
[§] DIC=Deviance Information Criterion; pD= The estimated numbers of effective parameters; Mean(L)= The posterior mean of the log-likelihood;

RESULTS AND DISCUSSION

Factor analysis. The estimated parameters to compare different Bayesian models are shown in Table 2. The Deviance Information Criterion (DIC, Spiegelhalter et al. (2002)) favored SMTMs and FA2G over the other three models. In the same sense, the posterior mean of the log-likelihood (Mean(L)) fitted considerably better the data from these both models (SMTMs and FA2G) than FA2R, FA2F and FA3F. Indicating that FA2G might be an alternative to the SMTMs in genetic evaluation schemes for beef cattle involving multi-trait estimates.

As expected, the estimated numbers of effective parameters (pD, Spiegelhalter et al. (2002)) was higher in FA3F. However, the FA2G had fewer effective numbers of parameters (Table 2), indicating a more parsimonious model than the other. The number of parameters (p) of $\boldsymbol{\Sigma}_u$, in the FA are, $p = q + mq - m(m-1)/2$ parameters, where $q \times m$ are the size of $\boldsymbol{\Lambda}$ matrix of factor loads. FA3F has 21 p in each matrix (additive genetic and residual), while FA2F has 17 p in each matrix (additive genetic and residual). In SMTMs, the p of $\mathbf{G}_0$ are, $p = q(q+1)/2$ parameters, where $q \times q$ is the size of $\mathbf{G}_0$ matrix and in SMTMs has 21 p in each matrix (additive genetic and residual). Depending on the data set, i.e. the number of records and number of traits the FA models may be used as special case of SEM to reduce covariance matrices rank in the model and give a parsimonious estimation of genetic parameters compared to SMTM. In this paper FA3F had poorer fit to the data compared to the other models; however, FA2G was effective compared to SMTMs. Generally, when more factors are included in the model, it might better explain the relationships between traits; however, when it has a smaller number of traits, fewer factors can be precisely estimated. Because we have only six traits and the number of records of SC used is limited (1,340), this may, partially, explain the worst DIC, pD and Mean(L) of FA3F.

Table 3. Estimates of heritability (diagonal), genetic (above diagonal) and residual (below diagonal) correlations, with standard deviation obtained from the factor analysis (Model 4: FA2G). FA2G had two factors only for the matrix of additive genetic loading factors and in the residual matrix it was considered as in SMTMs (standard multi-trait mixed models with sire model).

Traits†	LMA	BF	RF	BW	HH	SC
LMA	0.25±0.05	0.10±0.12	0.07±0.10	0.03±0.08	-0.08±0.11	0.03±0.08
BF	0.14±0.02	0.31±0.08	0.31±0.13	-0.01±0.14	-0.33±0.12	0.08±0.15
RF	0.11±0.02	0.58±0.01	0.23±0.06	0.00±0.11	-0.25±0.12	0.05±0.12
BW	0.49±0.02	0.22±0.02	0.15±0.02	0.22±0.05	0.03±0.12	0.01±0.07
HH	0.11±0.02	-0.01±0.02	-0.03±0.02	0.42±0.02	0.30±0.08	-0.07±0.13
SC	0.21±0.04	-0.02±0.07	0.01±0.07	0.39±0.04	0.20±0.05	0.40±0.11

†See Table 1 for abbreviations.

As discussed by Smith et al. (2001) and Kirkpatrick and Meyer (2004) the FA model usually reduces the rank of covariance matrices, but the **Ψ** matrix has to be close to zero, when specific effects are assumed absent. This provides a mixed model formulation with less than full rank covariance matrices.

Genetic Parameters. In Tables 3 and 4 are shown estimates of heritabilities (diagonal), genetic (above diagonal) and residual correlations (below diagonal), obtained from the FA2G and the SMTMs, respectively. The heritabilities and genetic breeding values estimated with FA2F, FA3F, FA2G and FA2R (not shown) were similar to those reported in Table 3. The exceptions are most heritabilities estimated by SMTMs (Table 4), which yield slightly larger estimates (except for SC), but with the same pattern and within one standard deviation of those obtained by FA models. The *Spearman* and *Pearson* correlation of the breeding values of all traits and all models ranged between 0.94 and 1.00 showing that to select based on these breeding values, any of those FA models would be a good strategy to estimate such parameters.

As expected, the residual correlations between the traits in Model 4 (FA2G; Table 3) were practically equal as the SMTMs (Table 4). Nevertheless, most of the genetic correlations between the traits in Model 4 (FA2G; Table 3) were somewhat lower compared with the SMTMs (Table 4). On the other hand, these genetic correlations obtained by Model 4 (FA2G; Table 3) have the same pattern compared with the SMTMs (Table 4) but with a lower magnitude. Only when genetic correlations were of lower magnitude the FA models could estimate the corresponding covariance reasonably well, suggesting that FA models would not be a good alternative to estimate high correlations as discussed by Kirkpatrick and Meyer (2004).

Table 4. Estimates of heritability (diagonal), genetic (above diagonal) and residual (below diagonal) correlations, with standard deviation obtained from the standard multi-trait mixed models with sire model (Model 1: SMTMs).

Traits†	LMA	BF	RF	BW	HH	SC
LMA	0.29±0.06	0.13±0.15	0.05±0.16	0.35±0.13	-0.14±0.15	0.19±0.17
BF	0.14±0.02	0.47±0.10	0.62±0.10	0.03±0.17	-0.52±0.12	0.09±0.22
RF	0.11±0.02	0.57±0.01	0.35±0.08	0.07±0.17	-0.42±0.14	0.05±0.19
BW	0.49±0.02	0.22±0.02	0.15±0.02	0.28±0.06	0.24±0.15	0.09±0.17
HH	0.11±0.02	-0.01±0.02	-0.03±0.02	0.42±0.02	0.38±0.08	-0.17±0.21
SC	0.21±0.04	-0.02±0.07	0.02±0.07	0.39±0.04	0.20±0.05	0.40±0.10

†See Table 1 for abbreviations.

CONCLUSION

Results suggest that factor analyses used as a special case of structural equation models could estimate breeding values of the bulls practically equal to standard multi-trait mixed models using sire model. These models have to be tested in order to implement the benefits, as an alternative to standard multi-trait mixed models using sire model. Depending on the data set, namely the number of records and traits, the factor analyses can reduce covariance matrices ranks and give a parsimonious estimation of genetic dispersion parameters compared to standard multi-trait mixed models, especially if the covariances between the traits are low.

LITERATURE CITED

de los Campos, G. and Gianola, D. (2007). Genet. Sel. Evol., 39:481-494.

Gianola, D. and Sorensen, D. (2004). Genetics, 167:1407-1424.

Henderson, C. R. and Quaas, R. L. (1976). J. Anim. Sci., 43:1188-1197.

Jöreskog, K. G. (1970). Biometrika, 57(2):239-251.

Kirkpatrick, M. and Meyer, K. (2004). Genetics, 168:2295-2306.

Meyer, K. (2009). Genet. Sel. Evol., 41:1-21.

Smith, A. B., Cullis, B. R., Thompson, R. (2001). Biometrics, 57:1138-1147.

Spiegelhalter, D.J., Best, N.G., Carlin, B. P., and van der Linde, A. (2002). J. of the Roy. Stat. Soc. B, 64:583-639.

Wright, S. (1921). J. Agric. Res. 201: 557–585.