Livestock Science 142 (2011) 188–194

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Multicollinearity in genetic effects for weaning weight in a beef cattle composite population

Raphael Antonio Prado Dias [a], Juliana Petrini [a], José Bento Sterman Ferraz [b], Joanir Pereira Eler [b], Rachel Santos Bueno [b], Ana Luísa Lopes da Costa [a], Gerson Barreto Mourão [a,*]

[a] University of São Paulo, Piracicaba, SP, PO Box 9 Campus ESALQ, CEP: 13.418-900, Brazil
[b] University of São Paulo, Pirassununga, SP, Campus FZEA, CEP: 13.635-900, Brazil

ARTICLE INFO

Article history:
Received 26 November 2010
Received in revised form 13 July 2011
Accepted 23 July 2011

Keywords:
Genetic evaluation
Ridge regression
Least squares
Crossbred

ABSTRACT

In the genetic evaluation of composite cattle or multiracial populations, the additive and non-additive genetic effects need to be estimated given their importance in developing strategies for crossing. However, multicollinearity, defined as the presence of strong linear correlations between the explanatory variables, is an obstacle to obtaining these estimates, since it produces unstable regression coefficients with large standard errors when the least square method is used, leading to erroneous inferences. Thus, the objective of this study was to detect possible collinearity involving the covariates of genetic effects, to assess them by the ridge regression (RR) method, and to compare results with estimates obtained by the least squares (LS) method. Weaning weight data of composite Montana Tropical bovine born between 1994 and 2008 were used. Some covariates from the model were involved in two strong and three weak collinearities. The RR method was used as an alternative to the LS method and showed better results. After using RR, the average variance inflation factors reduced from 16 to 5.3 and yielded more accurate estimates, with smaller standard errors than those obtained by the LS.

1. Introduction

In an effort to assist the improvement of livestock, many studies are based on genetic analysis to obtain estimates of variance components and genetic parameters of production and functional traits, as well as to evaluate the effects of heterosis and recombination loss in the case of crossbred animals. Thus, in the genetic evaluation of multibreed populations, both additive and non-additive genetic effects need to be estimated.

The most commonly used statistical method to derive prediction equations is the least squares (LS) method. However, when strong linear relationships between the covariates exist, a problem known as multicollinearity, the estimates of regression coefficients for LS tend to be unstable, often with large standard errors and can lead to mistaken inferences (Bergmann and Hohenboken, 1995).

An alternative method of estimation used in the informative analysis when there is multicollinearity (Draper and Smith, 1998) is the ridge regression (RR, Hoerl and Kennard, 1970), which consists of the addition of non-negative coefficients to the principal diagonal of the correlation matrix, reducing or eliminating the linear dependencies. Ridge estimators are biased, but estimates obtained by this method are more accurate, i.e., they present lower standard errors and are more stable than those obtained by LS in the presence of multicollinearity.

This study aims to identify possible dependencies among covariates included in the model for weaning weight (WW) of composite Montana Tropical beef cattle, to obtain estimates of direct, maternal and heterosis effects for this trait using LS and RR methods, and to compare the results from both approaches.

2. Material and methods

Data of weaning weight (WW, kg) were obtained from a database of Montana Tropical composite cattle reared in Brazil

* Corresponding author. Tel.: +55 19 34294009; fax: +55 19 34294215.
E-mail address: gbmourao@usp.br (G.B. Mourão).

1871-1413 © 2011 Elsevier B.V. Open access under the Elsevier OA license.
doi:10.1016/j.livsci.2011.07.016

and Uruguay, totalizing 191,036 records of animals born between 1994 and 2008. From the original database, only animals with valid measurements were used to estimate additive and non-additive genetic effects. Individuals without parentage information were deleted from the database. Animals were identified and weighed at weaning around 7 months of age. The databank controlled information on performance, breed composition and pedigree of all animals. Animals were grouped into contemporary groups (CG) that consider the year of birth, farm, management group within farm, and sex.

The animals were kept in tropical pastures, the majority in acid soils with *Brachiaria* spp. grass. A salt and mineral supplementation was given during all years. During the dry season, some farms also supplemented with a mineral salt enriched with a protein source.

Due to the large number of breeds involved in the base population of the Montana Tropical composite, the individuals from different breed compositions were grouped according to the NABC system (Ferraz et al., 1999; Mourão et al., 2007), where breeds are classified into biological types as follows: *Type N*: *Bos indicus* or Zebu breeds such as Gyr, Guzerat, Indubrazil, Nellore, Tabapuan, and other Zebu breeds from Africa, like Boran. These breeds have a high hardiness, parasite resistance, good carcass and are mainly represented by the Nellore cows. *Type A*: *Bos taurus* cattle, originally adapted to the tropics by natural or artificial selection and descended from animals introduced by colonists. Animals from these breeds have a high fertility and adaptability to the tropical climate and have some good meat quality traits. *Type B*: *B. taurus* breeds with British origin such as Angus, Devon and Hereford. These breeds contribute with sexual precocity and finishing, carcass conformation traits, carcass, and meat quality and growth. *Type C*: *B. taurus* breeds from continental Europe, including Charolais, Limousin, Simmental and other breeds. These breeds have a high potential for growth, yield, and carcass quality (Table 1).

In order to predict the additive genetic effects associated with biological types, their percentages were used, given as the proportion of each biological type in the genetic composition of the calf and the dam. Similarly, the effects of heterozygosity were obtained by a linear relationship to the coefficients of direct heterozygosity (*HD*) and maternal total (*HM*), which were obtained by the following equations:

$$H_D = 1 - \sum_{i=1}^{4} S_i \times D_i; H_M = 1 - \sum_{i=1}^{4} MGS_i \times MGD_i,$$

where 4 is the number of biological types (N, A, B, C); S_i, D_i, MGS_i and MGD_i are fractions of the i^{th} biological type of sire, dam, grandsire and granddam respectively.

The genetic analysis was based on the statistical model:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Fv} + \mathbf{Za} + \mathbf{Wm} + \mathbf{Sc} + \mathbf{e} \quad (1)$$

where *y* is the vector of observations; **X**, **F**, **Z**, **W** and **S** are the incidence matrices; **β** is the vector associated with a constant, the additive direct effects associated with individual (A, B or C) and maternal (AM, BM or CM) biological type compositions, the direct heterozygosity (NxA, NxB, NxC, AxB, AxC and BxC), the maternal total heterozygosity (HM) and the age at weaning (AW); **v** is the vector associated with the cubic effect of the Julian day (JD), the age of the dam when the calf was born (AOD, Table 2) and the contemporary groups (CG); **a** is the vector of random additive direct effects; **m** is the vector of random additive maternal effects; **c** is the vector of the uncorrelated random effects of maternal permanent environment and **e** is the vector of residual effects, NID (0, $\hat{\sigma}_e^2$). The relationship matrix was formed by a total of 398,301 animals, 2509 sires, and 210,565 dams in 7.42 generations, considering the first generation as one. The pedigree file was constituted by purebred animals, which are the founders of this composite population, crossbred and composite individuals. The number of generations was calculated based in the methodology described by Brinks et al. (1961).

The covariate associated with the direct and maternal additive genetic effects of the biological type N was excluded from the statistical model because the sum of the biological type fractions is equal to one. Therefore, the effects of the

Table 1
Number of observations (N) in each genetic group based on the NABC system.

Genetic group	Necessary condition	N
3/4	60%>N<90%	2987
	60%>A<90%	111
	60%>B<90%	2848
	60%>C<90%	182
Montana Tropical®	18.75<N<31.25% and 18.75<A<31.25% and 18.75<B<31.25% and 18.75<C<31.25%	12,991
	18.75<N<31.25% and 18.75<A<31.25% and 43.75<B<56.25% and C<6.25%	3111
	18.75<N<31.25% and 43.75<A<56.25% and B<6.25% and 18.75<C<31.25%	17,383
	N<37.25% and 12.50<A<87.50% and B≤75% and C≤75% and (N+A)≥25% and (B+C)≤75%	84,960
Purebred	N≥90%	5237
	A≥90%	378
	B≥90%	2209
	C≥90%	2
F_1	40≤N≤60% and 40≤A≤60%	2519
	40≤N≤60% and 40≤B≤60%	41,500
	40≤N≤60% and 40≤C≤60%	5946
Other	Every breed composition that does not comply with the conditions above	8672

Table 2
Minimum and maximun age (months) and number of observations for weaning weight (WW) for each dam age at calving (AOD, months).

Class of AOD	Age of dam at calving		Absolute frequency
	Minimum	Maximum	WW
1	–	≤27	17,833
2	>27	≤41	30,351
3	>41	≤59	25,657
4	>59	≤119	52,090
5	>119	≤143	6574
6	>143	≤167	2862
7	>167	–	1890
Total			137,257

biological types A, B and C were estimated as deviations of the additive effects of N.

The estimates of the variance components for animal ($\hat{\sigma}_a^2$), maternal ($\hat{\sigma}_m^2$), maternal permanent environmental ($\hat{\sigma}_{ep}^2$) and residual effects ($\hat{\sigma}_e^2$), as well as the covariance between animal and maternal genetics effects ($\hat{\sigma}_{am}$) were 103.28 kg^2; 54.28 kg^2; 63.52 kg^2; 341.73 kg^2, and −11.80 kg^2, respectively, by the LS method which estimates the genetic covariates, and 104.11 kg^2; 50.96 kg^2; 66.64 kg^2; 341.65 kg^2, and −11.91 kg^2, respectively, by the RR method.

The solution to the Model (1) was obtained using the procedure described below. The solutions for **v**, **a**, **m** and **c** vectors were obtained as:

$$\mathbf{y}_{(t)} = \mathbf{F}\mathbf{v}_{(t)} + \mathbf{Z}\mathbf{a}_{(t)} + \mathbf{W}\mathbf{m}_{(t)} + \mathbf{S}\mathbf{c}_{(t)} + \mathbf{e}_{1(t)} \quad (2)$$

where t denotes the t^{th} iteration and $\mathbf{y}_{(t)} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{(t-1)}$. In the first iteration, the values obtained by the LS method were set to $\hat{\boldsymbol{\beta}}$. Using either method LS or RR, the solutions for β were obtained from the following model:

$$\mathbf{y}_{(t)} = \mathbf{X}\boldsymbol{\beta}_{(t)} + \mathbf{e}_{2(t)} \quad (3)$$

where

$$\mathbf{y}_{(t)} = \mathbf{y} - \mathbf{F}\hat{\mathbf{v}}_{(t)} - \mathbf{Z}\hat{\mathbf{a}}_{(t)} - \mathbf{W}\hat{\mathbf{m}}_{(t)} - \mathbf{S}\hat{\mathbf{c}}_{(t)}.$$

The solutions $\hat{\mathbf{a}}, \hat{\mathbf{m}}$ and $\hat{\mathbf{c}}$ were obtained in the first step of the t^{th} iteration.

The previous steps were carried out ten times and the absolute differences between the last two solutions were analyzed for β. In ridge regression, coefficients *k* are added to the main diagonal of correlation matrix, aiming to decrease linear dependences between covariates. Because the vector **K**, formed by these coefficients *k*, is obtained empirically, this iterative process was developed in order to obtain a vector **K** that minimizes the variance of parameters and the bias of these. Thus, if we consider the vector **K** equal to the null vector, the process converges in the first iteration.

In the process to detect multicollinearity, to know which covariates were involved in the possible relation of *quasi*-dependence, the value of the variance inflation factor (VIF) was used, which is obtained for the predicted variable X_i as follows: $VIF_i = 1/(1-R_i^2)$, where R_i^2 is the determination coefficient of the X_i linear regression with respect to the other variables. The VIF for the ordinary least squares method are the diagonal elements of the inverse simple correlation matrix, which represents an increase of variance due to the presence of multicollinearity. Usually values greater than 10 indicate the presence of multicollinearity and, therefore, problems in estimation (Chatterjee et al., 2000).

To detect the number of collinearities, the condition index (CI) was calculated for each eigenvalue:

$$CI_i = (\lambda_{max}/\lambda_i)^{1/2},$$

where λ_{max} is the largest eigenvalue and λ_i is the i^{th} eigenvalue of the correlation matrix. High CI values indicate dependence among the variables because if λ_i is close to zero, and since the determinant of the correlation matrix is the product of the eigenvalues, the determinant of this matrix will also be near zero indicating multicollinearity. According to Belsley (1991) CI values between 10 and 30 indicate a weak multicollinearity and values greater than 30 indicate a strong multicollinearity.

The covariates involved in each collinearity were detected from the proportion decomposition of the variance associated to the eigenvalues, based on how much the variance of the estimated parameter is associated with each eigenvalue. According to Belsley (1991), $Var\left(\hat{\boldsymbol{\beta}}\right) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$ where σ^2 is the estimated residual variance, **V** are the eigenvectors matrix, ∧ are the eigenvalues diagonal matrix. Writing $\mathbf{V} = v_{ij}$ the variance of the i^{th} element of β, the variance of each parameter estimate can be written as the sum of the *p* components with each number associated with an eigenvalue as follows: $Var\left(\hat{\beta}_i\right) = \sigma^2 \sum_{j=1}^{p} v_{ij}^2/\lambda_j$, where *p* is the number of predictor variables. When λ_j is small, the components of variance associated with dependence will be relatively large compared with the other components of variance. A high proportion of two or more coefficients associated with some eigenvalues provides evidence that the corresponding dependencies are causing problems.

Letting $t_{ij} = v_{ij}^2/\lambda_j$ and $t_i = \sum_{j=1}^{p} t_{ij}$, the proportion of variance of the i^{th} regression coefficient associated with the j^{th} component of this decomposition will be obtained by $\pi_{ij} = t_{ij}/t_i$, always with i = 1, 2, …, p. One recommendation is to identify the eigenvalues with CI greater than 30 (Belsley, 1991). Variables with the variance decomposition proportion (π_{ij}) greater than 0.5 for each of these eigenvalues are candidates for linear dependence. Measures of multicollinearity are obtained after the predictor variables standardization, as recommended by Freund and Littell (2000).

In the general Model (3) the β value can be estimated as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ by the LS method, so the estimates and their variances may be uncertain in the presence of multicollinearity. Alternatively, the RR consists in the addition of coefficients to the main diagonal of the correlation matrix ($\mathbf{X}'\mathbf{X}$), causing a decrease in the variance estimates. The estimate of β for RR was obtained as follows: $\hat{\boldsymbol{\beta}}^* = (\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}\mathbf{X}'\mathbf{y}$, where $\mathbf{K} = \mathrm{diag}(k_1, k_2, \ldots, k_p)$, $k_i \geq 0$. When $k_i = 0$, for all i, $\hat{\boldsymbol{\beta}}^*$ is reduced to the least square estimator.

The matrix of variance and covariance of $\hat{\boldsymbol{\beta}}^*$ was estimated as follows:

$$Var\left(\hat{\boldsymbol{\beta}}^*\right) = (\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}\hat{\sigma}^2,$$

where $\hat{\sigma}^2$ is the estimate of σ^2 by the least square estimator and VIF were the diagonal elements of the $(\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}$ matrix.

The mean square error (MSE) is the difference between the values of $\hat{\boldsymbol{\beta}}^*$ and $\hat{\boldsymbol{\beta}}$ squared as follows: $\mathrm{MSE} = E\left[\left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right) \times \left(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}\right)\right] = \mathrm{trace}\left[Var\left(\hat{\boldsymbol{\beta}}^*\right)\right] + \boldsymbol{\beta}'(\mathbf{Z} - \mathbf{I})'(\mathbf{Z} - \mathbf{I})\boldsymbol{\beta}$. MSE = estimation variance + (bias)^2, where $\mathbf{Z} = (\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}\mathbf{X}'\mathbf{X}$.

The RR is defended when the introduction of biases in the estimates is balanced by a decrease in the estimation of the error variance, resulting in a lower MSE compared to the LS method (Hoerl and Kennard, 1970).

The ridge regression is performed using the transformed data, so they have a mean equal to 0 and a variance equal to 1. Thus, after the analysis, the estimates were transformed back to the original scale.

After verifying the existence of strong multicollinearity and the covariables of each, the steps referred to the Models (2) and (3) were executed 10 times with the objective to obtain the stabilization of the estimates. The estimates of the regression coefficients of the β vector obtained by the LS and RR methods and their standard errors (Fig. 3), the k_i elements of the diagonal **K** matrix, as well as the absolute difference between the estimates of regression coefficients obtained by RR in the last two iterations were presented (Table 3).

The determination of an ideal value for the ridge parameter **K** which results in a smaller MSE than the one obtained by the LS, depends on the value of the β parameter vector and the error variance σ^2, which are both unknown (Hoerl and Kennard, 1970). Consequently, **K** must be determined empirically. Many methods have been proposed to obtain the appropriate values but there is no consensus on which method is most appropriate. In this work the ridge parameter **K** was estimated by the method adopted by Roso et al. (2005), where the k_i elements of the **K** diagonal matrix were estimated by $k_i = \theta \times VIF_i / max(VIF)$, where VIF_i is the variance inflation factor of the i^{th} explanatory variable; $max(VIF)$ is the largest VIF and θ is the minimum value between zero and one, therefore, the maximum VIF is less than 10.

Two alternative strategies of analysis using the Model (1) are: 1) an additive-dominant model that included the additive effects, the heterozygosity and the age at weaning that were estimated by LS; 2) an additive-dominant model using the RR model with the same effects included previously but estimated by the RR.

If the ridge parameter increases, the variance decreases and the bias increases. This is a known relationship between the ridge parameter, the variance and the bias of the RR estimate. Since $E\left(\hat{\beta}\right) = \beta$ and $E\left(\hat{\beta}^*\right) = (\mathbf{X}'\mathbf{X} + \mathbf{K})^{-1}\mathbf{X}'\mathbf{X}\beta = \mathbf{H}\beta$, a measure of the bias of regression vector $\hat{\beta}^*$ was calculated as $1 - (||\mathbf{H}|| / ||\mathbf{I}|| \times 100)$, where denotes Euclidean norm. A measure of bias close to zero for a particular RR method indicates less bias in the estimates.

Table 3
K element values of the diagonal matrix and the absolute difference between the estimates of regression coefficients obtained by RR in the last two iterations $\left|\hat{\beta}^*_{(10)} - \hat{\beta}^*_{(9)}\right|$.

Covariates[a]	K	$\left\|\hat{\beta}^*_{(10)} - \hat{\beta}^*_{(9)}\right\|$	Convergence
A	0.01618	0.00223	10^{-2}
B	0.01102	0.00006	10^{-4}
C	0.03346	0.00018	10^{-3}
AM	0.00533	0.00363	10^{-2}
BM	0.01555	0.00307	10^{-3}
CM	0.01221	0.00337	10^{-2}
NxA	0.00666	0.00038	10^{-3}
NxB	0.01367	0.00318	10^{-2}
NxC	0.01926	0.00252	10^{-2}
AxB	0.00532	0.00225	10^{-2}
AxC	0.01347	0.00003	10^{-4}
BxC	0.00501	0.00127	10^{-2}
HM	0.00218	0.00178	10^{-2}

[a] A, B and C are the additive effects associated with the individual biological composition types for A, B and C, respectively; AM, BM and CM are the maternal additive effects associated with the maternal biological composition types for AM, BM and CM, respectively; NxA, NxB, NxC, AxB, AxC and BxC are the direct heterozygosity and HM is the maternal total heterozygosity. The additive genetic effects were estimated as deviations of biological type N to replace 100% of the alternative biological type.

3. Results and discussion

The response variable WW had 191,036 observations with an average of 198.4 kg and a standard deviation of 36.7 kg. Cattle were weighed when they were between 140 and 290 days old. The maximum and minimum weights were 97 kg and 300 kg, respectively.

According to the VIF, the additive direct effect A (22.7), B (15.5), and C (46.9); additive maternal effect BM (21.8), CM (17.1) and the heterozygosis NxB (19.2), NxC (27.0) and AxC (18.9) could be involved in collinearity because their variances were being inflated more than tenfold compared with the estimates obtained if the vector of covariates were orthogonal (Fig. 1). Rodriguez-Almeida et al. (1997) working with populations of bulls and cows from Red Poll, Hereford, Angus, Limousin, Braunvieh, Pinzgauer, Gelbvieh, Simmental, and Charolais breeds also noted the correlation between additive genetic and maternal effects. The multicollinearity involving the effects of the biological types can be the result of the Montana Tropical breeding program structure, since an animal is only included in this breed if it has certain proportions of each biological type in its genetic composition. With the necessity of directing the crossings, it is possible that the whole inheritance of certain biological type is derived exclusively from a dam or a sire, in order to maintain the breed required percentage. Thus, the connection between these variables is natural. Besides, each proportion of one biological type depends on the other to complete the genetic composition of the animal (Roso et al., 2005).

Through the condition indices, obtained from the eigenvalues, six linear dependences between the explanatory variables were observed: two strong, with CI values above 30, and four weak, with CI values between 10 and 30 (Fig. 2).

From the variance-decomposition proportions associated with the largest condition index (CI = 34.5) only the covariate BM showed a higher value (0.57) than the threshold (0.5). However, the explanatory variables AM (0.47), NxB (0.48), and NxC (0.46) may also be involved in a strong linear relationship, since the threshold of 0.5 is an empirical value. Furthermore, a VIF greater than 10 was determined for these covariates. Considering the second largest condition index (CI = 32.0), only the intercept (0.77) and the covariate AW (0.71) are involved in a linear dependence. However the VIF associated with the explanatory variable AW was the lowest (1.04). This was due to the fact that VIF has been calculated from the matrix (**X′X**) centered and standardized without the intercept column, i.e., although the covariate AW was close to the orthogonality with the genetic covariates, in relation to the intercept it was not.

Some of the explanatory variables that have VIF values greater than 10 and which were not involved in a strong collinearity may be in a weak collinearity. The same behavior was observed by Roso et al. (2005) for the pre-weaning weight gain from purebred animals and crossbred with Angus, Blonde d'Aquitaine, Charolais, Gelbvieh, Hereford, Limousin, Maine-Anjou, Salers, Shorthorn and Simmental. Another possibility is that the diagnostic method of the variance inflation factor overestimates the presence of multicollinearity, since there is no defined limit to distinguish between the VIF values that can be considered high and those that can be considered low, which results in a limitation to the use of this method, due to its inability to distinguish between the various coexisting *quasi-*

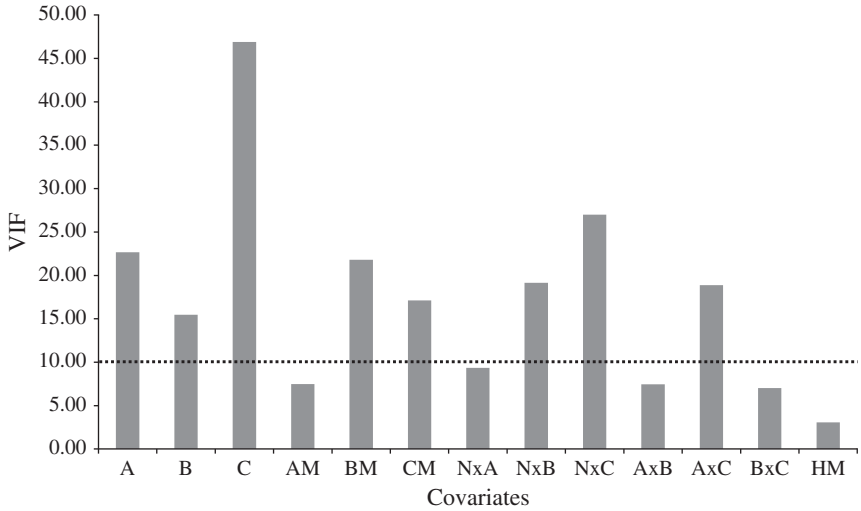

Fig. 1. Variance inflation factor (VIF) for the explanatory covariates considered in the design matrix X of Model (1). The dotted line (VIF>10) is an indicative of involvement in collinearity (A, B and C are the additive effects associated with the individual biological composition types for A, B and C, respectively; AM, BM and CM are the maternal additive effects associated with the maternal biological composition types for AM, BM and CM, respectively; NxA, NxB, NxC, AxB, AxC and BxC are the direct heterozygosity and HM is the maternal total heterozygosity).

dependence. In this way, despite that the variance inflation factor is the most commonly used method to diagnose multicollinearity because of the easier interpretation of the results, it is necessary to compare these results to those obtained by other methods.

The effects of the explanatory variables did not differ in direction for the RR and LS methods. Thus, it is possible to infer that the presence of multicollinearity in the model does not change the management of the crossings aiming at genetic progress of weaning weight. However, there are differences in the magnitude of these estimates and their standard errors. With this, the genetic gain with the increase of certain biological types on the genetic composition of the animal is overestimated when the LS method is employed. Moreover, the results are less reliable when compared to those obtained by RR.

The effects of the explanatory variables A, B and C were higher than those of N. Similarly, in Franke et al. (2001) and Williams et al. (2010), the additive genetic effects for Angus, Hereford, Charolais, Simmental, Gelbvieh were larger than those for Brahman. This result was expected due to the larger size of European and adapted breeds when compared to the Zebu breeds, which is related to a higher potential for preweaning growth of the former breeds.

The explanatory variables AM and BM had negative effects while the CM had a positive effect on the weaning weight when compared to the maternal effect of biological type N, possibly because of the low magnitude of maternal effects for the breeds

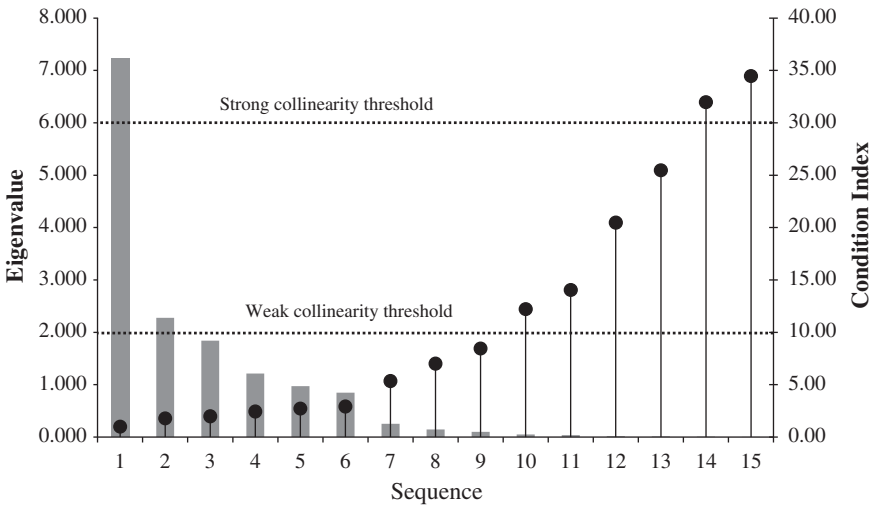

Fig. 2. Eigenvalues (■) and condition index (CI, ●) of the X matrix of Model (1). The lines are indicative of the presence of collinearity. The numbers in the abscissa correspond to the sequence of the eigenvalues and their respective condition indexes, where one represents the first eigenvalue obtained from the correlation matrix, with the higher value, up to the number 15, which represents the last eigenvalue obtained, with the lower value.

within these groups (Franke et al., 2001; Legarra et al., 2007; Williams et al., 2010). Thus, the use of dams with higher fractions of N and C is favorable for obtaining heavier calves at weaning.

All heterosis effects contributed positively to the trait, with increases between one and 10 kg. The BxC heterozygosity had the lowest heterosis effect, probably a result from the genetic similarity among these biological types (Brandt et al., 2009). Equally, Williams et al. (2010) found lower values of heterosis for crosses between British and Continental breeds compared to the heterosis estimates for Zebu×British cross and Continental×Zebu cross.

A positive effect was also estimated for maternal heterosis, suggesting that the use of composite dams is beneficial to the weaning weight (Vergara et al., 2009).

The magnitude of the error for the estimated $\hat{\beta}^*$ can be observed (Table 3) according to the absolute difference between the estimated regression coefficients for the explanatory variables obtained by RR in the last two iterations. To estimate the regression coefficient of the explanatory variable, the difference between the solution iterations was 0.00006 and the error between the estimates was less than 10^{-4}. All **K** values were obtained with smaller convergence than 10^{-2}.

There was evidence that the method used to estimate the covariates by RR was better than the one by LS, because it was observed that the greatest VIF (9.99) obtained by RR was below the threshold of 10 and the VIF average (5.25) obtained by RR was closer to one than the VIF average (16.03) obtained by LS. Consequently, the regression coefficients obtained by RR had lower standard errors when compared to those obtained by LS (Fig. 3), indicating higher efficiency of the RR method to reduce the regression coefficient variability.

The sum of squared deviations (SQD) obtained by RR (44,018,682) was lower than that obtained by LS (44,088,623). This small difference was expected since the multicollinearity does not affect the sum of squared deviations but the parameter estimates.

Table 4
Estimates of the covariates effects age at weaning (AW, days), Julian day (JD, days) and the variable class of age of the dam when the calf was born (AOD) on weaning weight (in kg) using the least squares (LS) and ridge regression (RR) methods.

Covariate	LS	RR
AW (linear coef.)	0.477	0.503
JD (linear coef.)	1.657	1.513
JD^2 (quadratic coef.)	−0.012	−0.010
JD^3 (cubic coef.)	−0.00014	−0.00015
AOD_1	−25.81	−25.63
AOD_2	−14.98	−14.94
AOD_3	−4.99	−4.96
$AOD_{4\ (reference)}$	0.00	0.00
AOD_5	−1.75	−1.85
AOD_6	−5.58	−5.75
AOD_7	−12.13	−12.33

The bias of 9.8%, considered small, was similar to the one obtained by Carvalheiro et al. (2006) (11.1%) who studied the weight gain in pre-weaning calves from Nellore×Hereford crossings.

A linear effect was observed ($p<0.01$) for age at weaning (AW), a cubic ($p<0.05$) for Julian day (JD) and both effects for the categorical variables AOD and CG. The estimates of AOD are presented as deviations from AOD_4 which represents the highest level of efficiency when the dam reaches maturity (Table 4).

The effects of AOD were important with an adjustment magnitude of approximately 26 kg, which was detected between AOD_1 and AOD_4. Zampar et al. (2006) obtained similar values for this classification variable, although they did not consider the effect of Julian day in their model. Szabó et al. (2006) also determined a significant effect of the age of the dam at calving on weaning weight, noting an increase in weight up to 5 years old followed by a decrease on the trait. There was little expected difference between the estimates considering the LS and the RR, because these variables apparently were not considered to be involved in collinearity.

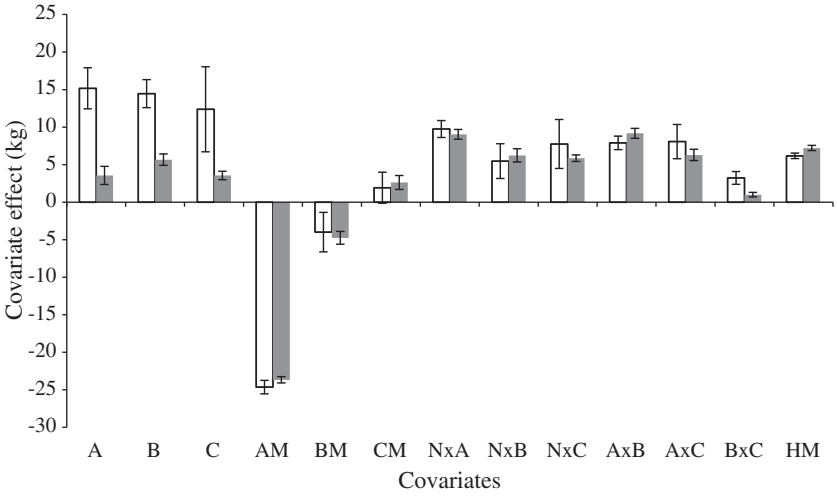

Fig. 3. Effects of covariates (kg) using the genetic Model (1), with their standard errors estimated by the least squares methods (LS, □) and ridge regression (RR, ■) for the response variable to weaning weight (A, B and C are the additive effects associated with the individual biological composition types for A, B and C, respectively; AM, BM and CM are the maternal additive effects associated with the maternal biological composition types for AM, BM and CM, respectively; NxA, NxB, NxC, AxB, AxC and BxC are the direct heterozygosity and HM is the maternal total heterozygosity. The additive genetic effects were estimated as deviations of biological type N to replace 100% of the alternative biological type).

4. Conclusions

The estimates of regression coefficients obtained by the ridge regression were more reliable because of the smaller standard errors than those obtained by the least squares means, thus the inclusion of bias was offset by increased accuracy.

The choice of the **K** diagonal matrix elements in the ridge regression depends on the data and on the explanatory variables of the model and should be weighed for the explanatory variables according to the collinearity to avoid the inclusion of an unnecessary bias.

Acknowledgments

This research was supported by FAPESP, CNPq, CAPES, and CFM-Leachman Pecuária Ltda.

The authors acknowledge the "Grupo de Melhoramento Genético Animal e Biotecnologia" (Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga-SP, Brazil) for its support and providing the database.

References

Belsley, D.A., 1991. Conditioning Diagnostics, Collinearity and Weak Data in Regressionfirst ed. John Wiley and Sons, Inc., New York.

Bergmann, J.A.G., Hohenboken, W.D., 1995. Alternatives to least squares in multiple linear-regression to predict production traits. J. Anim. Breed. Genet. 112, 1–16.

Brandt, H., Müllenhoff, A., Lambertz, C., Erhardt, G., Gauly, M., 2009. Estimation of genetic and crossbreeding parameters for preweaning traits in German Angus and Simmental beef cattle and the reciprocal crosses. J. Anim. Sci. 88, 80–86.

Brinks, J.S., Clark, R.T., Rice, F.J., 1961. Estimation of genetic trends in beef cattle. J. Anim. Sci. 20, 903.

Carvalheiro, R., Pimentel, E.C.G., Cardoso, V., Queiroz, S.A., Fries, L.A., 2006. Genetic effects on preweaning weight gain of Nelore-Hereford calves according to different models and estimation methods. J. Anim. Sci. 84, 2925–2933.

Chatterjee, S., Hadi, A.S., Price, B., 2000. Regression Analysis by Examplethird ed. John Wiley and Sons, Inc., New York.

Draper, N.R., Smith, H., 1998. Applied Regression Analysisthird ed. John Wiley and Sons, Inc., New York.

Ferraz, J.B.S., Eler, J.P., Golden, B.L., 1999. Análise genética do composto Montana Tropical. Rev. Bras. Reprod. Anim. 23, 111–113.

Franke, D.E., Habet, O., Tawah, A.L., Williams, A.R., Derouen, S.M., 2001. Direct and maternal genetic effects on birth and weaning traits in multibreed cattle data and predicted performance of breed crosses. J. Anim. Sci. 79, 1713–1722.

Freund, R., Littell, R.C., 2000. SAS System for Regressionthird ed. SAS Inst., Inc. Cary, NC.

Hoerl, A.E., Kennard, R.W., 1970. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12, 55–67.

Legarra, A., Bertrand, J.K., Strabel, T., Sapp, R.L., Sanchez, J.P., Misztal, I., 2007. Multi-breed genetic evaluation in a Gelbvieh population. J. Anim. Breed. Genet. 124, 286–295.

Mourão, G.B., Ferraz, J.B., Eler, J.P., Balieiro, J.C.C., Bueno, R.S., Mattos, E.C., Figueiredo, L.G.G., 2007. Genetic parameters for growth traits of a Brazilian *Bos taurus*×*Bos indicus* beef composite. Genet. Mol. Res. 6, 1190–1200.

Rodriguez-Almeida, F.A., Van Vleck, L.D., Gregory, K.E., 1997. Estimation of direct and maternal breed effects for prediction of expected progeny differences for birth and weaning weights in three multibreed populations. J. Anim. Sci. 75, 1203–1212.

Roso, V.M., Schenkel, F.S., Miller, S.P., Schaeffer, L.R., 2005. Estimation of genetic effects in the presence of multicollinearity in multibreed beef cattle evaluation. J. Anim. Sci. 83, 1788–1800.

Szabó, F., Nagy, L., Dákay, I., Márton, M., Tötök, M., Bene, Sz, 2006. Effects of breed, age of dam, birth season and sex on weaning weight of beef calves. Livest. Sci. 103, 181–185.

Vergara, O.D., Elzo, M.A., Ceron-Muñoz, M.F., Arboleda, E.M., 2009. Weaning weight and post-weaning gain genetic parameters and genetic trends in a Blanco Orejinegro–Romosinuano–Angus–Zebu multibreed cattle population in Colombia. Livest. Sci. 124, 156–162.

Williams, J.L., Aguilar, I., Rekaya, R., Bertrand, J.K., 2010. Estimation of breed and heterosis effects for growth and carcass traits in cattle using published crossbreeding studies. J. Anim. Sci. 88, 460–466.

Zampar, A., Mourão, G.B., Ferraz, J.B.S., Eler, J.P., Balieiro, J.C.C., Bueno, R.S., Pedrosa, V.B., Mattos, E.C., Figueiredo, L.G.G., 2006. Effects of classes of dam at calving on growth traits in a *Bos taurus*×*Bos indicus* composite beef cattle population. Proceedings of the 8th World Congress on Genetics Applied to Livestock Production, Belo Horizonte, Brazil. CD-ROM.