

The influence of multicollinearity on crossbreeding parameter estimates for weaning weight in beef cattle

S.J. Schoeman^{1#}, M.A. Aziz² and G.F. Jordaan¹

¹Department of Animal Sciences, University of Stellenbosch, Private Bag X1, Matieland 7602, South Africa

²Department of Animal Production, Faculty of Agriculture, Alexandria University, Egypt

Abstract

Data on 17258 weaning weight records of calves from a crossbreeding project were utilized to investigate the problem of collinearity and its effect on the estimation of direct, maternal and the non-additive genetic effects. Several criteria were used to detect the near-dependency among the independent variables. The results indicated that there was a near-dependency among both the direct and the maternal genetic effects causing unstable estimates. It was attempted to solve the collinearity problem using ridge regression. An improved model fit was evident at a ridge value of $k = 0.8$ with large reductions in standard errors and estimates with more meaningful biological interpretation. Ridge regression is recommended for the estimation of crossbreeding effects where inevitable collinearity amongst the independent variables is evident.

Keywords: Beef cattle, crossbreeding, genetic effects, ridge regression

[#]Corresponding author. E-mail: sjsc@sun.ac.za

Introduction

Crossbreeding parameter estimates are important in setting up effective crossbreeding systems. It helps the producer to fully utilize genetic differences between breeds. Genetic effects in terms of breed additive, breed maternal, individual heterosis and maternal heterosis have widely been estimated and used for breed characterization and for predicting the performance of crosses that have not actually been tested (Alenda *et al.*, 1980; Dillard *et al.*, 1980; Robison *et al.*, 1981; Schoeman *et al.*, 1993; Skrypzeck *et al.*, 2000a).

In most studies of this nature, multiple regression procedures were used, especially where pedigree records are not available. The proportion of genes contributed by each breed was used as the independent variables. However, the possibility of linear dependencies among the independent variables, or collinearity, and its effect and means to handle this have not been addressed. The impact of collinearity on least squares can be serious. In such cases the estimates of the regression coefficients may differ greatly from the parameters they are estimating, even to the point of having incorrect sign (Rawlings, 1988). It impairs the usefulness of the regression analysis with respect to the interpretation of the regression coefficients. If there is a collinearity problem, standard procedures used to estimate regression coefficients therefore fail.

The problem of multicollinearity is often solved by selection or omission of highly correlated variables. For the animal breeder, however, the most accurate estimates of all crossbreeding parameters and their relative magnitudes are of interest. Variable selection is therefore not an appropriate option.

The objective of this study was to assess the impact of a possible collinearity problem amongst the independent or x -variables on crossbreeding parameter estimates and ways to overcome the problem.

Material and Methods

Weaning weight records of calves were obtained from the initial crossbreeding phase of the composite breed development programme of the Johannesburg Metropolitan Council. The environment and management of the herd, as well as replacement and selection procedures, were described in detail by Paterson *et al.* (1980), while the various crossbreeding combinations and breeds involved were described by Schoeman *et al.* (2000) and Skrypzeck *et al.* (2000a; b).

The data consisted of 17258 weaning weight records which were recorded between 1968 and 1992. Five breeds, namely the Afrikaner (AF), Hereford, Angus, Simmentaler (ST) and Charolais (CH) were used in the crossbreeding programme to produce 129 different breed groups of calves. No distinction was made between

the Hereford and the Angus, as earlier studies (Fredeen *et al.*, 1982; Schoeman, 1996; Tosh *et al.*, 1999) did not detect important differences between them. Therefore, they were pooled and considered as one breed (HA).

The data were classified into sex (male and female), age of dam (ranged from 2 to 9 years), herd-year-season (HYS) and breed groups as discreet variables. Some of the HYS subclasses contained only a small number of observations. Thus, they were pooled with the next HYS subclass having the same season in the following year within the same farm. Season of birth was recorded as winter born or summer born. Age at weaning was also recorded.

Least squares analysis of variance was conducted using Proc. GLM of SAS (2000). The initial model fitted, included the fixed effects of HYS, sex, age of dam as well as weaning age as a covariate and all one-way interactions. All interactions with no effect ($P > 0.05$) were subsequently removed from the model. The model was:

$$Y_{ijkl} = \mu + \text{HYS}_j + S_k + D_l + A_{ge} + e_{ijkl}$$

where:

Y_{ijkl} = the record,

μ = the overall mean,

HYS_j = the effect of the j^{th} HYS contemporary group,

S_k = the effect of the k^{th} sex of calf,

D_l = the effect of l^{th} age of dam,

A_{ge} = the age at weaning, and

e_{ijkl} = the random error which is assumed to be randomly and independently distributed with mean zero and variance σ_e^2 .

Subsequently, after eliminating a few outliers on the predicted observations, a multiple regression analysis was conducted on them, using the Proc. REG procedure of SAS (2000). The model fitted to the adjusted data included the genetic effects, which were defined in terms of breed additive, breed maternal, average individual and average maternal heterosis. The coefficients used for the genetic effects were the proportions of the genes contributed by each breed to either the calf or the dam. These proportions were considered as continuous variables and were calculated as follows: $g_i = \frac{1}{2} (\alpha_i^s + \alpha_i^d)$, $h_{ij} = \alpha_i^s \alpha_j^d + \alpha_j^s \alpha_i^d$, where g_i and h_{ij} denote the coefficients for the direct and heterotic effects, respectively, and α_i^s and α_i^d the gene proportions of breed i in the sire and dam of the calf, respectively (Wolf *et al.*, 1995). Since the mating plan did not allow for the estimation of individual and maternal heterosis, due to a lack of observations in specific crosses, only average individual and maternal heterosis could be estimated. The following regression model was used:

$$Y = \beta_0 + \beta_1 \text{DAF} + \beta_2 \text{DCH} + \beta_3 \text{DST} + \beta_4 \text{DHA} + \beta_5 \text{MAF} + \beta_6 \text{MCH} + \beta_7 \text{MST} + \beta_8 \text{MHA} + \beta_9 \text{HI} + \beta_{10} \text{HM} + e$$

where:

Y = the adjusted or predicted record,

β_0 = the constant (the intercept),

$\beta_1, \beta_2, \beta_3, \beta_4$ = the regression coefficients of breed additive effects,

DAF, DCH, DST, DHA = the percentages of genes contributed by AF, CH, ST and HA, respectively,

$\beta_5, \beta_6, \beta_7, \beta_8$ = the regression coefficients of breed maternal effects,

MAF, MCH, MST, MHA = the percentages of genes contributed by dams of AF, CH, ST and HA, respectively,

β_9 = the regression coefficient of average individual heterosis, due to the interaction of two alleles at the same locus, with alleles being from different breeds,

HI = the percentage of genes contributed to the average individual heterosis,

β_{10} = the regression coefficient of average maternal heterosis,

HM = the percentage of loci in the dam with one gene from one breed and the other from a different breed,

e = the error term.

Several diagnostic procedures were subsequently used for detecting collinearity (Rawlings, 1988; Bowerman & O'Connell, 1990). These procedures included simple correlation coefficients between the independent variables, the variance inflation factor (VIF) which is an indicator of the severity of multicollinearity and the eigenanalysis of X, or principle component analysis. These analyses were conducted using the Proc. PRINCOMP of SAS (2000). Ridge regression was fitted as a remedial procedure for the estimation of the ridge point estimates. This procedure allows an approach to detect collinearity and simultaneously estimating new coefficients through the artificial reduction of the correlations between the related X-variables to overcome collinearity by adding a constant (k) to the diagonal of the correlation matrix among predictors before inverting it for least squares estimation (Bowerman & O'Connell, 1990). The k-values were incremented by 0.2 from zero to 1.2 so as to minimize the residual sum of squares while a quantitative measure of stability, called the "Index of Stability of Relative Magnitude (ISRM)", proposed by Vinod (1976) was used to locate the position (k-value) where the best estimates were obtained.

Results and Discussion

Regression coefficients and their standard errors for the direct effects, maternal effects and average individual and maternal heterotic effects are presented in Table 1. The estimates of the direct effects corresponded to what was expected according to the results of earlier studies (Alenda *et al.*, 1980; Dillard *et al.*, 1980; Schoeman *et al.*, 1993; Skrypzeck *et al.*, 2000a). The Charolais, as expected, has the largest direct effect while the Simmentaler has the largest maternal effect. Standard errors are fairly large, especially for the maternal effects. The large negative direct effects vs. the large positive maternal effects for all breeds, furthermore, complicated the biological interpretation thereof. The same applied to the average individual and average maternal heterotic effects, where only the individual heterotic component was significant ($P < 0.05$).

Table 1 Regression coefficients (\pm s.e.), variance inflation factors (VIF) and tolerance values for direct effects, maternal effects and heterotic effects for weaning weight in crossbred calves

Genetic effects	Regression coefficient (\pm s.e.)	VIF	Tolerance value
Intercept	144.0 (93.6)	-	-
DAF ^a	-277.7 (48.7)**	1386	0.0007
DCH	-228.3 (48.7)**	4132	0.0002
DHA	-266.7 (48.9)**	5512	0.0002
DST	-240.0 (48.8)**	6609	0.0002
MAF	301.3 (73.4)**	11240	0.0001
MCH	297.2 (73.3)**	4069	0.0002
MHA	301.7 (73.6)**	19402	0.0001
MST	312.3 (73.3)**	13396	0.0001
HI	1.61 (0.70)*	1.78	0.56
HM	0.50 (0.53)	1.90	0.53

^a Abbreviations: D - direct effects; M – maternal genetic effects of AF – Afrikaner, CH – Charolais; HA – Hereford-Angus and ST – Simmentaler. HI – individual heterotic effect; MH – maternal heterotic effect

* $P < 0.01$, ** $P < 0.001$

According to Rawlings (1988), large standard errors of the partial regression coefficients may indicate near linear dependencies, or collinearity, among the X-variables, in which case they contribute overlapping information for describing the dependent variable (Y). In such cases, the regression coefficients become extremely unstable and are very sensitive to small random errors in Y and may fluctuate widely as individual independent variables are either added or removed from the model. It thus gives a misleading impression of the importance of individual X-variables and might even appear less important than they really are.

Pairwise correlation coefficients between the coefficients of the individual different effects and pairwise variance inflation factors ($1 / (1 - r^2)$) are presented in Table 2. Correlation coefficients of the direct effects between different breeds were all negative and varied between -0.03 and -0.58 . Those between the maternal

effects were also negative, with the exception of the correlation between MAF and MHA which was 0.50. Those negative correlations are obvious, since they involve mutually exclusive proportions and are thus obviously not independent. The within-breed correlations between the direct and maternal effects were positive and ranged from 0.48 to 0.92, while those between breeds were all negative and varied from -0.01 to -0.45. A few highly correlated variables were thus observed, which clearly suggested the existence of a possible collinearity problem. These included those between DCH and DST, DCH and MCH, MHA and HM, MST and MHA, and especially between DAF and MAF, with VIF's which varied from 1.43 to 6.67. However, when several variables are involved in linear dependencies, individual pairwise correlations do not need to be large to indicate collinearity (Rawlings, 1988). Although the correlation matrix is a useful starting point for understanding the structure of the X-space, it has certain limitations. It is in the first place incapable of diagnosing the coexistence of near-dependencies among the various individual X-variables. In addition, it cannot support evidence of a collinearity problem in the absence of a high correlation. Further diagnostic procedures are therefore needed.

Table 2 Correlation coefficients above the diagonal and variance inflation factors (VIF) below the diagonal between the proportions of the direct genetic, maternal genetic and heterotic effects for weaning weight in crossbred calves between Afrikaner, Charolais, Hereford-Angus and Simmentaler cattle

	Genetic effects									
	DAF ^a	DCH	DHA	DST	MAF	MCH	MHA	MST	HI	HM
DAF	-	-0.03	-0.40	-0.07	0.92	-0.11	-0.45	-0.24	0.34	0.32
DCH	1.00	-	-0.22	-0.58	-0.01	0.55	-0.07	-0.21	0.20	0.13
DHA	1.19	1.05	-	-0.56	0.42	-0.12	0.57	-0.23	-0.36	-0.43
DST	1.01	1.51	1.46	-	-0.03	-0.28	-0.26	0.48	-0.01	0.15
MAF	6.67	1.00	1.21	1.00	-	-0.11	0.50	-0.24	0.32	0.37
MCH	1.01	1.43	1.01	1.08	1.01	-	-0.25	-0.14	-0.03	0.23
MHA	1.25	1.01	1.48	1.07	1.33	1.07	-	-0.59	0.11	-0.67
MST	1.06	1.05	1.06	1.30	1.06	1.02	1.53	-	-0.40	0.33
HI	1.13	1.04	1.14	1.00	1.11	1.00	1.01	1.19	-	-0.04
HM	1.11	1.02	1.23	1.02	1.16	1.06	1.81	1.12	1.00	-

^a Abbreviations: See Table 1

Variance inflation factors of b_i (VIF) of the model and tolerance values are simple diagnostic procedures for detecting overall collinearity problems, with

$$VIF_j = 1 / (1 - R_j^2),$$

where R_j^2 is the coefficient of determination from the regression of one (X_j) of the dependent variables on the others. These values are also presented in Table 1. In all cases, except for HI and HM, the VIFs were considerably larger than the cut-off value of 10, proposed by Gill (1986), while the corresponding tolerance values were much less than the proposed value of 0.10. These results thus indicate a serious collinearity problem among these variables, i.e. some variables have a trivial impact on the estimates of the parameters, while the contribution of others is redundant, however, without identifying specific near-dependencies among the explanatory variables. The VIFs should, however, also be evaluated relative to the fit of the model. Models with lower R^2 -values ($R^2 = 20.9\%$ in this study), also have considerably lower VIFs than the proposed 10, thus causing poor parameter estimates (Freund & Wilson, 1998).

For the analysis of multiple regression, where collinearity is evident, principle component (PC) analysis is suggested (Gill, 1986; Rawlings, 1988; Khattree & Naik, 2000). It is a procedure that creates a set of new uncorrelated variables which are linearly related to the original variables without altering the total variability thereof. Principle component analysis derives a number of linear combinations of a set of variables that retains as much of the information in the original variables, i.e. it seeks collections of related variables. They are orthogonal rotations of the original variables and are used to uncover approximate linear dependencies among the independent variables (X_i). However, it does not reflect the "true" underlying sources of the variability among the X_i 's. Principle components are derived from the eigenvalues and eigenvectors of the correlation

matrix. Eigenvalues, along with the cumulative eigenvalue proportions and condition indices of the first eight PCs are presented in Table 3, while the elements of the eigenvectors of the (co)variance matrix between the PCs and the original variables are presented in Table 4.

Table 3 Eigenvalues of the correlation matrix, cumulative contributions and condition indices of the genetic effects for weaning weight in crossbred calves between Afrikaner, Charolais, Hereford-Angus and Simmentaler cattle.

Principle component	Eigenvalue	Cumulative proportion	Condition index
PC ₁	3.22	0.32	1.00
PC ₂	2.36	0.56	1.17
PC ₃	1.86	0.74	1.31
PC ₄	1.15	0.86	1.67
PC ₅	0.54	0.91	2.44
PC ₆	0.44	0.96	2.72
PC ₇	0.33	0.99	3.13
PC ₈	0.09	1.00	5.82

Table 4 Elements of the eigenvectors of the (co)variance matrix of the correlations between the components and the original variables of the genetic effects for weaning weight in crossbred calves

Genetic effects	Principle components							
	PC ₁	PC ₂	PC ₃	PC ₄	PC ₅	PC ₆	PC ₇	PC ₈
DAF ^a	0.40	0.28	-0.23	-0.31	0.08	0.02	-0.17	0.03
DCH	0.04	0.34	0.51	0.21	0.02	-0.47	-0.12	-0.30
DHA	-0.44	0.04	-0.02	-0.48	-0.11	0.03	0.24	0.41
DST	0.19	-0.44	-0.28	0.41	0.06	0.33	-0.04	-0.15
MAF	0.42	0.27	-0.22	-0.30	0.06	0.05	-0.12	-0.07
MCH	0.05	0.16	0.58	0.14	-0.06	0.64	-0.15	0.29
MHA	-0.47	0.18	-0.22	0.17	0.09	0.05	0.24	-0.37
MST	0.15	-0.54	0.14	-0.01	-0.12	-0.46	-0.09	0.34
HI	0.14	0.36	-0.21	0.51	0.24	-0.17	0.34	0.58
HM	0.40	-0.10	0.24	-0.13	-0.18	0.07	0.82	-0.21

^a Abbreviations: See Table 1

Eigenvalues, which provide measures of the amount of variation in the dimensions of the datapoints dispersed in the multidimensional X-space, were large for the first seven PCs, accounting for 99% of the standardised variance (Table 3). It thus appears that eight sample PCs effectively summarise the total variance. Large condition indices, which are simply the eigenvalue of a PC relative to the sum of all the eigenvalues, also indicate a collinearity problem (Besley *et al.*, 1980), so does the multicollinearity index ($mci = 1.15$) suggested by Thisted (1980). The first four PCs accounted for 86% of the total variation, while the next four accounted for only 14% thereof. Small eigenvalues (or large condition indices) indicate possible problems with multicollinearity.

The first PC, which accounted for 32% of the total variation (Table 3), seems to represent a contrast between mainly the AF and HA, while in the second PC, differences were primarily between the AF, CH and ST (Table 4). The eigenvector corresponding to the first eigenvalue (PC₁) defines the dimensions causing the collinearity problem as:

$$\begin{aligned}
 &0.40 \text{ DAF} - 0.44 \text{ DHA} \\
 \text{and} &0.42 \text{ MAF} - 0.47 \text{ MHA} \\
 &\text{or} \\
 &0.40 \text{ DAF} - 0.47 \text{ MHA} \\
 \text{and} &0.42 \text{ MAF} - 0.44 \text{ DHA}
 \end{aligned}$$

These variables are primarily responsible for the near-singularities as shown by the fact that they are very similar in magnitude, but opposite in sign, or the sum of the coefficients which is in each case close to zero. Likewise, the third PC, which accounted for an additional 12% of the variation, seems to highlight the differences between AF and ST with CH. PC₄, PC₇ and PC₈ were greatly affected by the heterotic effects, although their contributions to the overall variance were very small. From PC₈, having the smallest eigenvalue, the linear dependencies amongst the direct and maternal effects within each breed are evident (Table 3).

Although most of the elements of the PCs have little obvious or meaningful interpretation, it is clear that there were near-dependencies among the direct effects on the one hand and the maternal effects on the other hand. Likewise, the direct effects, as well as the maternal effects, were all collinear among themselves, all contributing redundantly to the regression coefficients leading to the instability in the regression results and the very high standard errors. Such near-dependencies may arise from an inbuilt mathematical constraint on variables that forces them to add to the constant, inadequate sampling or bad experimental design such as a lack of information on some the pure breeds and an inadequate number of crossbred combinations (Rawlings, 1988; Sölkner & James, 1990). All these could have contributed to the problem in this study.

Ridge coefficients of the direct, maternal and non-additive genetic effects fitted on increasing k-values are shown in Figure 1. A large degree of instability was observed in these coefficients at low values of k (k = 0 – 0.4), but stabilizes thereafter. Ridge regression coefficients converged towards zero as k increases.

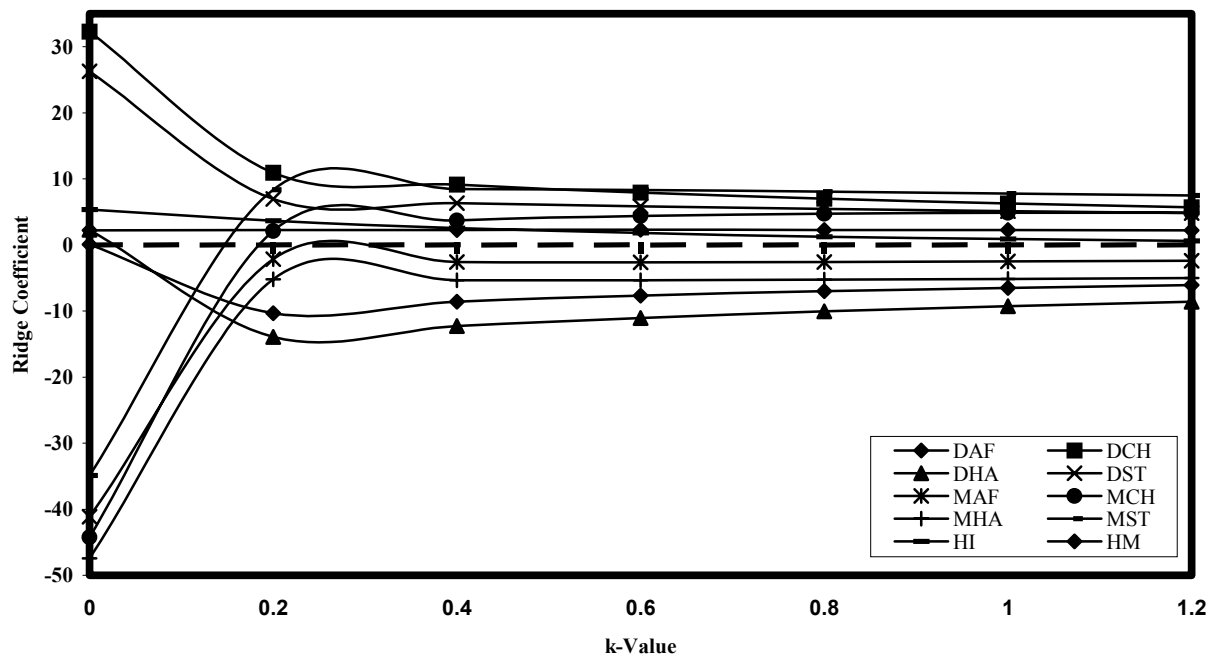


Figure 1 Ridge regression coefficients of direct, maternal, average individual heterosis and average maternal heterosis for weaning weight with increasing k-values (Abbreviations: See Table 1).

Means squares error (MSE), the Index of Stability of Relative Magnitude (ISRM) and variance inflation factors of the different genetic effects are presented in Table 5. Means squares error slightly increased with increasing k-values, while the ISRM values decreased and reached a stable point at k = 0.8, which could be regarded as the point of maximum model improvement. Likewise, the VIFs decreased sharply with increasing k-values from k = 0 (which are the VIFs from the ordinary least squares estimates) to considerably lower estimates at higher k-values. Differences among the VIFs also decreased between the different genetic effects with increasing k-values. This indicated the successful elimination of the collinearity problem.

Table 5 Means squares error (MSE), Index of Stability of Relative Magnitude (ISRM) and variance inflation factors (VIF) of direct, maternal, average individual and average maternal heterotic effects from ridge regression with k-values between 0 and 1.2 for weaning weight in crossbred calves

k-value	MSE	ISRM	VIFs of the genetic effects									
			DAF ^a	DCH	DHA	DST	MAF	MCH	MHA	MST	HI	HM
0	25.9	64.01	1386	4132	5512	6609	11240	4069	19402	13396	1.78	1.90
0.20	26.2	5.00	0.73	0.54	0.49	0.37	0.56	0.71	0.41	0.51	0.83	0.89
0.40	26.6	3.61	0.34	0.36	0.31	0.26	0.28	0.46	0.25	0.33	0.52	0.54
0.60	27.0	3.23	0.22	0.26	0.23	0.20	0.19	0.33	0.17	0.23	0.36	0.37
0.80	27.4	3.16	0.16	0.21	0.17	0.17	0.14	0.25	0.14	0.18	0.27	0.27
1.00	27.8	3.23	0.13	0.17	0.14	0.14	0.11	0.20	0.11	0.15	0.21	0.21
1.20	28.2	3.38	0.11	0.14	0.12	0.12	0.10	0.17	0.09	0.12	0.17	0.17

^a Abbreviations: See Table 1

Ridge regression coefficients of the direct, maternal and heterotic effects at $k = 0.8$ are presented in Table 6. These estimates indicated a substantial improvement in model fit with large reductions in standard errors as compared to the estimates from the ordinary least squares analysis (Table 1). They also provide a more meaningful biological interpretation with regard to the magnitude and signs of these estimates. Contrary to the large negative estimates of the direct effects and the large positive maternal effects from the ordinary least squares analysis, ridge regression estimates were considerably lower and more in line with what could be expected (Cunningham & Magee, 1988; Schoeman *et al.*, 1993). Both direct and maternal effects of both the Afrikaner and HA were negative, while those of both the Charolais and Simmentaler were positive, with the Charolais having the largest direct effect and the Simmentaler the largest maternal effect. Estimates of the heterotic effects were also significant ($P \leq 0.01$), with the estimate of the maternal heterotic effect larger than that of the individual heterotic effect. This is also in accordance to theoretical expectations for weaning weight (Gregory & Cundiff, 1980).

Table 6 Estimates of ridge regression coefficients (\pm s.e.) at $k = 0.8$ of direct, maternal, average individual and average maternal heterotic effects for weaning weight

Intercept	Genetic effects										
	DAF ^a	DCH	DHA	DST	MAF	MCH	MHA	MST	HI	HM	
121.7**	-7.2**	8.1**	-10.0**	7.2**	-1.9**	6.9**	-5.7**	8.5**	1.3**	2.4**	
(0.93)	(0.65)	(0.41)	(0.32)	(0.31)	(0.33)	(0.69)	(0.21)	(0.29)	(0.33)	(0.24)	

^a Abbreviations: See Table 1

** $P \leq 0.01$

Conclusions

In the analysis of crossbreeding data using regression procedures, the breed contributing fractions or explanatory variables are inevitably not independent, the severity thereof also influenced by the data structure. Collinearity or near-singularity among these x -variables may cause a problem in analysing crossbreeding data, as is proved in this investigation. Such collinearity could allow “important” variables to be replaced in the model with “incidental” variables that are involved in near-singularity. In such case, the regression analysis provides little indication of the relative importance of the independent variables and interpretation of estimates in such cases should be done with caution. Ridge regression eliminated the collinearity problem and yielded more stable and accurate estimates also having a more meaningful biological interpretation.

References

- Alenda, R., Martin, M.R., Lasley, J.F. & Eilersieck, M.R., 1980. Estimation of genetic and maternal effects in crossbred cattle of Angus, Charolais and Hereford parentage. I. Birth and weaning weight. *J. Anim. Sci.* 50, 226-234.

- Besley, D.A., Kuh, E. & Welsh, R.E., 1980. Regression Diagnostics: Identifying influential data and sources of collinearity. Wiley & Sons, Inc., New York, 474 pp.
- Bowerman, B.L. & O'Connell, R.T., 1990. Linear Statistical Models – an applied approach (2nd ed.). PWS-Kent Publishing Co., Boston, Massachusetts. 1024 pp.
- Cunningham, B.E. & Magee, W.T., 1988. Breed-direct, breed-maternal and non-additive genetic effects for preweaning traits in crossbred calves. *Can. J. Anim. Sci.* 68, 83-92.
- Dillard, E.U., Rodrigues, O. & Robison, O.W., 1980. Estimation of additive and non-additive direct and maternal genetic effects from crossbreeding beef cattle. *J. Anim. Sci.* 50, 653-663.
- Fredeen, H.T., Weiss, G.M., Rahnfeld, G.W., Lawson, J.E. & Newman, J.A., 1982. Environmental and genetic effects on preweaning performance of calves from first-cross cows. II. Growth traits. *Can. J. Anim. Sci.* 62, 51-67.
- Freund, R.J. & Wilson, W.J., 1998. Regression Analysis – Statistical modelling of a response variable. Academic Press, London. 444 pp.
- Gill, J.L., 1986. Outliers, residuals and influence in multiple regression. *J. Anim. Breed. Genet.* 103, 161-175.
- Gregory, K.E. & Cundiff, L.V., 1980. Crossbreeding in beef cattle: Evaluation of systems. *J. Anim. Sci.* 51, 1224-1242.
- Khattree, R. & Naik, D.N., 2000. Multivariate data reduction and discrimination with SAS Software, Cary, N.C.: SAS Institute Inc. (1st ed.). 558 pp.
- Paterson, A.G., Venter, H.A.W. & Harwin, G.O., 1980. Pre-weaning growth of British, *Bos indicus*, Charolais and dual-purpose type cattle under intensive pasture conditions. *S. Afr. J. Anim. Sci.* 10, 135-142.
- Rawlings, J.O., 1988. Applied regression analysis. A research tool (1st ed.). Wadsworth, Inc., Belmont, California. 553 pp.
- Robison, O.W., McDaniel, B.T. & Rincon, E.J., 1981. Estimation of direct and maternal additive and heterotic effects from crossbreeding experiments in animals. *J. Anim. Sci.* 52, 44-50.
- SAS, 2000. Statistical Analysis System user's Guide (Version 8.1). SAS Institute Inc., Cary, North Carolina.
- Schoeman, S.J., 1996. Characterization of beef cattle breeds by virtue of their performances in the National Beef Cattle Performance and Progeny Testing Scheme. *S. Afr. J. Anim. Sci.* 26, 15-19.
- Schoeman, S.J., Van Zyl, J.G.E. & De Wet, R., 1993. Direct and maternal additive and heterotic effects in crossbreeding Hereford, Simmentaler and Afrikaner cattle. *S. Afr. J. Anim. Sci.* 23, 61-66.
- Schoeman, S.J., Jordaan, G.F. & Skrypzeck, H., 2000. The influence of proportion of Simmentaler breeding in a multibreed synthetic beef cattle population on preweaning growth traits. *S. Afr. J. Anim. Sci.* 30, 98-109.
- Skrypzeck, H., Schoeman, S.J., Jordaan, G.F. & Naser, F.W.C., 2000a. Estimates of crossbreeding parameters in a multibreed beef cattle crossbreeding project. *S. Afr. J. Anim. Sci.* 30, 193-203.
- Skrypzeck, H., Schoeman, S.J., Jordaan, G.F. & Naser, F.W.C., 2000b. Pre-weaning growth traits of the Hereford breed in a multibreed composite beef cattle population. *S. Afr. J. Anim. Sci.* 30, 220-229.
- Sölkner, J. & James, J.W., 1990. Optimum design of crossbreeding experiments. II. Efficiency of designs without all purebreds. *J. Anim. Breed. Genet.* 107, 421-430.
- Thisted, R.A., 1980. Comment: A critique of some ridge regression methods. *J. Am. Stat. Assoc.* 75, 81-86.
- Tosh, J.J., Kemp, R.A. & Ward, D.R., 1999. Estimates of direct and maternal genetic parameters for weight traits and backfat thickness in a multibreed population of beef cattle. *Can. J. Anim. Sci.* 79, 433-439.
- Vinod, H.D., 1976. Application of new ridge regression methods to a study of bell system scale economics. *J. Am. Stat. Assoc.* 71, 835-841.
- Wolf, J., Distl, O., Hyamek, J., Grosshans, T. & Seeland, G., 1995. Crossbreeding in farm animals. V. Analysis of crossbreeding plans with secondary crossbred generations. *J. Anim. Breed. Genet.* 112, 81-94.