Journal of Animal and Veterinary Advances 3 (12): 856-863, 2004

Fixing Collinearity Instability Using Principal Component and Ridge Regression Analyses in the Relationship Between Body Measurements and Body Weight in Japanese Black Cattle

A. E. O. Malau-Aduli, [1]M. A. Aziz, [1]T. Kojima, [1]T. Niibayashi, [1]K. Oshima and [1,2]M. Komatsu
School of Agricultural Science, University of Tasmania, Private Bag 54 Hobart, Tasmania 7001, Australia; [1]Laboratory of Animal Breeding and Reproduction, Department of Livestock and Grassland Science, National Agricultural Research Centre for Western Region, 60 Yoshinaga, Kawai, Oda, Shimane 694-0013 Japan; [2]Department of Animal Breeding and Reproduction, National Institute of Livestock and Grassland Science,2 Ikenodai, Tsukuba, Ibaraki 305-0901 Japan

Abstract: Monthly measurements of withers height (WHT), hip height (HIPHT), body length (BL), chest width (CHWD), shoulder width (SHWD), chest depth (CHDP), hip width (HIPWD), lumbar vertebrae width (LUVWD), thurl width (THWD), pin bone width (PINWD), rump length (RUMPLN), cannon circumference (CANNCIR) and chest circumference (CHCIR) from birth to yearling age, were utilised in principal component and ridge regression analyses to study their relationship with body weight in Japanese Black cattle with an objective of fixing the problem of collinearity instability. The data comprised of a total of 10,543 records on calves born between 1937 and 2002 within the same herd under the same management. Simple pair wise correlation coefficients between the body measurements revealed positive, highly significant ($P<0.001$) values of 0.98 between WHT and HIPHT, HIPWD and LUVWD, while the lowest correlation of 0.50 was between CHDP and SHWD. Severe collinearity problems as portrayed by variance inflation factors (VIF) above 10 were evident in all body measurements ranging from 11.25 in PINWD to 46.94 in LUVWD except for SHWD (1.80), CHDP (3.70), CHWD (7.11) and CANNCIR (7.33). Principal component and ridge regression analyses allowed the derivation of new and more stable regression coefficients that overcame the problem of collinearity. Of all the body measurements studied, hip height was shown to be the least important for predicting the body weight of Japanese Black cattle, while SHWD and CHWD were the most important.

Key words: Principal component, ridge regression, body weight, body measurements, Japanese Black Cattle

Introduction

The use of simple and multiple linear least squares regression analyses for the prediction of body weight from body measurements in cattle is important for taking management decisions related to selection for growth. In dairy cows, the estimation of body weight from body size measurements and body condition score has been reported by Heinrichs *et al.* (1992), Enevoldsen and Kristensen (1997), Kertz *et al.* (1997) and Koenen and Groen (1998). In beef cattle, similar research has been conducted and reported by Gilbert *et al.* (1993), Wilson *et al.* (1997) and Vargas *et al.* (2000). Similar information in Japanese Black cattle is scanty, and where available, it is mostly limited to carcass traits (Mukai *et al.* 1995, Mukai *et al.* 2000, Karnuah *et al.* 2001, Smith *et al.* 2001 and Sosa *et al.* 2002).

In any regression analysis, the partial regression coefficients and partial sums of squares for any independent variable are dependent on which other independent variables are in the model. Thus, the changes in regression coefficients and sums of squares as other variables are added to, or removed from, the model could be large and this is mainly because the independent variables are not mutually orthogonal. Lack of orthogonality could be extreme such that two or more independent variables become very nearly linearly dependent thereby creating the problem of collinearity. Collinearity causes instability in the regression coefficients because the estimates can change markedly as a result of small changes in the estimation data. The instability is reflected in very large standard errors for the partial regression coefficients. Frequently, none of the individual partial regression coefficients will be significantly different from zero, even though their combined effect is highly significant. Such estimates lead to poor prediction and can be difficult to interpret in terms of the underlying biological process (Rook *et al.*, 1990).

Principal component and ridge regression are two commonly used regression methods to computationally attack the problem of collinearity. Ridge regression does this by reducing the apparent magnitude of the correlations (Hoerl and Kennard 1970a, 1970b, Hoerl *et al.* 1975, Marquardt and Snee 1975, Smith and Campbell, 1980). Principal component regression on the other hand, approaches the collinearity problem from the point of view of eliminating from consideration, those dimensions of the X-space that are causing the collinearity problem. This is similar in concept to dropping an independent variable from the model when there is insufficient dispersion in that variable to contribute meaningful information on Y. However, in principal component regression, the dimension dropped from consideration is defined by a linear combination of the variables rather than by a single independent variable (Rawlings *et al.* 1998). Our aim in this paper was to study of the relationship between body weight and body measurements in Japanese Black cattle so as to ascertain the existence or otherwise of collinearity instability. If detected, the second objective was to fix the problem using principal component and ridge regression analyses.

Materials and Methods

Animals, location and management: Japanese Black cattle kept at the Department of Livestock and Grassland Science, National Agricultural Research Centre for Western Region, Oda, Shimane Prefecture, Japan, were utilised for this study. The management practices in this herd had been described previously (Shimada *et al.* 1992).

Data: Records of monthly body weight and body measurements from birth to yearling age of Japanese Black cattle born between 1937 and 2002 were analysed. The body measurements were: withers height (WHT), hip height (HIPHT), body length (BL), chest width (CHWD), shoulder width (SHWD), chest depth (CHDP), hip width (HIPWD), lumbar vertebrae width (LUVWD), thurl width (THWD), pin bone width (PINWD), rump length (RUMPLN), cannon circumference (CANNCIR) and chest circumference (CHCIR). We adopted similar body measurement methods used by Magnabosco *et al.* (2002). A total of 10,543 observations on these 13 variables was utilised in the final analysis after editing the data.

Statistical Analysis: In all analyses, body weight was treated as the dependent variable (Y) while the body measurements were the independent variables (X) and within animal variation was used since the analysis was of within animal means (Table 1). Prior to subjection to multiple regression analyses, the following mixed model was used in general linear models (PROC GLM) procedures (SAS 2002) to statistically adjust for the effects of age, sex, season, year, sire, dam and season-year of birth since the data were collected from male and female animals of different ages across seasons in different years:

$$Y_{ijklmno} = \mu + A_i + SEX_j + S_k + Y_l + SIRE_m + DAM_n + (SY)_{kl} + e_{ijklmno}$$

where $Y_{ijklmno}$ = body weight of the o^{th} calf of the i^{th} age group of the j^{th} sex born within the k^{th} season of the l^{th} year belonging to the m^{th} sire and n^{th} dam,
μ = the overall mean,
A_i = fixed effect of the i^{th} age group (i=1, 3),
SEX_j = fixed effect of the j^{th} sex (j=1, 2),
S_k = fixed effect of the k^{th} season of birth (k=1, 4),
Y_l = fixed effect of the l^{th} year of birth group (l=1, 10),
$SIRE_m$ = random effect of sire,
DAM_n = random effect of dam,
$(SY)_{kl}$ = season-year effect,
$e_{ijklmno}$ = random error associated with each record with a mean of 0 and variance 6^2_e.

Age was coded as 0 month (birth), 1-5 months (pre-weaning) and 6 months and above (post-weaning). Sex was coded as 1 and 2 to represent male and female respectively. Season of birth was coded as winter (December – February), Spring (March – May), Summer (June – August) and Autumn (September – November). Year of birth was coded into ten groups of 10 year-intervals between 1937 and 2002.

As a first indication of the severity of collinearity, simple pairwise correlation coefficients between all the 13 independent body measurement variables were obtained along with their variance inflation factors (VIF) using SAS (2002) as depicted in Tables 2 and 3 respectively. When a correlation matrix is inversed, the diagonal elements (also known as the VIFs), are computed as follows:

VIF = $(1/1-R^2)$ where R^2 = coefficient of determination. Rook *et al.* (1990) stated that VIF in excess of 10 indicates severe collinearity which leads to unstable estimation of the associated least squares regression coefficients. To be able to identify the major sources of variation among the independent variables and eliminate collinearity problems, principal component analysis was utilised. The description of the method below is based on those of Rawlings *et al.* (1998) and Rook *et al.* (1990):

Principal component analysis can simply be put as: W = XZ,

where W = $n \times p$ matrix of principal components, X = $n \times p$ matrix of original variables and Z = $p \times p$ matrix satisfying Z ' (X ' X) Z = Ë and Z ' Z = I ; Ë = diagonal matrix of the ordered eigenvalues/ latent roots ($ë_1$ = $ë_2$ = ...$ë_p$) of X ' X ; n = number of animals or samples.

The first column of Z is the first singular vector of X or the first eigenvector of X ' X. Thus the coefficients in the first eigenvector define the particular linear function of the columns of X (the original variables) that generates the

Table 1: Means and standard deviations (cm) of the independent body measurement variables in Japanese Black cattle

Body measurement	Mean	Standard deviation
WHT	84.21	11.33
HIPHT	88.04	11.31
BL	84.09	16.33
CHWD	22.09	5.16
SHWD	26.69	7.05
CHDP	34.48	9.16
HIPWD	22.14	5.11
LUVWD	17.21	3.99
THWD	25.91	5.19
PINWD	15.10	3.72
RUMPLN	28.77	5.41
CANNCIR	11.82	1.67
CHCIR	100.24	19.62

*WHT = Withers height HIPHT = Hip height BL = Body length
CHWD = Chest width SHWD = Shoulder width CHDP = Chest depth
HIPWD = Hip width LUVWD = Lumbar vertebrae width THWD = Thurl width
PINWD = Pin bone width RUMPLN = Rump length CANNCIR = Cannon circumference
CHCIR = Chest circumference

Table 2: Simple pairwise correlation coefficients of body measurements in Japanese Black cattle

	WHT	HIPHT	BL	CHWD	SHWD	CHDP	HIPWD	LUVWD	THWD	PINWD	RUMP	CANN	CHCIR
WHT	0.98	0.97	0.90	0.64	0.76	0.94	0.96	0.93	0.91	0.94	0.84	0.94	
HIPHT	0.98		0.97	0.90	0.64	0.76	0.93	0.95	0.93	0.90	0.93	0.84	0.93
BL	0.97	0.97		0.90	0.63	0.77	0.94	0.96	0.93	0.91	0.94	0.85	0.94
CHWD	0.90	0.90	0.90		0.65	0.74	0.90	0.92	0.89	0.87	0.89	0.81	0.90
SHWD	0.64	0.64	0.63	0.65		0.50	0.63	0.64	0.62	0.61	0.63	0.55	0.63
CHDP	0.76	0.76	0.77	0.74	0.51		0.80	0.80	0.82	0.77	0.82	0.82	0.83
HIPWD	0.94	0.93	0.94	0.90	0.63	0.80		0.98	0.97	0.93	0.97	0.88	0.97
LUVWD	0.96	0.95	0.96	0.92	0.64	0.80	0.98		0.97	0.94	0.97	0.92	0.97
THWD	0.93	0.93	0.93	0.90	0.62	0.82	0.97	0.97		0.95	0.97	0.88	0.97
PINWD	0.91	0.90	0.91	0.87	0.61	0.77	0.93	0.94	0.95		0.93	0.90	0.94
RUMP	0.94	0.93	0.94	0.90	0.63	0.82	0.97	0.97	0.97	0.93		0.90	0.98
CANN	0.84	0.84	0.85	0.81	0.55	0.82	0.88	0.88	0.92	0.88	0.90		0.91
CHCIR	0.94	0.93	0.94	0.90	0.63	0.83	0.97	0.97	0.97	0.94	0.98	0.91	

* All correlations were highly significant ($P<0.01$)

first column of W. The second column of W is obtained using the second eigenvector of X ' X, and so on. Note that W ' W = L. Thus W has the property that all its columns are orthogonal (L is a diagonal matrix so that all off-diagonal elements, the sum of products between columns of W, are zero). The sum of squares of the ith column of W is $ë_i$, the ith diagonal element of L. Thus if X is an *n x p* matrix of observations on *p* variables, each column of W is a new variable defined as a linear transformation of the original variables. The ith new variable has sums of squares $ë_i$ and all are pairwise orthogonal.

The linear transformation in principal component analysis is to a set of orthogonal variables such that the first principal component accounts for the largest possible amount of the total dispersion (proportion of variation), measured by $ë_1$, the second principal component accounts for the largest possible amount of the remaining dispersion $ë_2$, and so forth (Table 4). The total dispersion is given by the sum of all the eigenvalues, which is equal to the sum of squares of the original variables; tr (X ' X) = tr (W ' W) = $Óë_i$.

Recall that the original variables were transformed to the principal components using the latent vectors W = XZ. The original regression model Y = bo + b_1X_1 + ...b_pX_p can be restated in terms of the standard variables as Y = xβ + ì, where ì = n x 1 vector of residuals. The model can be restated in terms of principal components as: Y = Wá + ì, where W = XZ and á = Z ' β. It is therefore possible to calculate regression coefficients in terms of the principal components (á) and transform them back to the β using the relationship β = Zá. If a principal component with a very small latent root is present, there is collinearity in the data. By setting the coefficient of this component to zero when this component is excluded from the regression, collinearity is removed.

Ridge regression combats the collinearity problem by artificially reducing the mean square errors to values lower than those of the ordinary least squares estimates thereby resulting in smaller standard errors and higher accuracy.

Table 3: Parameter estimates and variance inflation factors (VIF) of body measurements for the prediction of body weight in Japanese Black cattle*

Variable	Estimate	s.e.	Sig.	Tolerance	R2	VIF	Remarks
Intercept	-157.33	1.61	0.0001	.	.	0.00	-
WHT	0.89	0.05	0.0001	0.03	0.97	38.62	Severe collinearity
HIPHT	0.06	0.05	0.1970	0.03	0.97	34.39	Severe collinearity
BL	0.87	0.03	0.0001	0.05	0.95	21.89	Severe collinearity
CHWD	1.28	0.05	0.0001	0.14	0.86	7.11	Non-collinearity
SHWD	-0.14	0.02	0.0001	0.56	0.44	1.80	Non-collinearity
CHDP	0.16	0.02	0.0001	0.27	0.73	3.70	Non-collinearity
HIPWD	1.77	0.10	0.0001	0.03	0.97	30.72	Severe collinearity
LUVWD	4.80	0.16	0.0001	0.02	0.98	46.94	Severe collinearity
THWD	-1.11	0.10	0.0001	0.03	0.97	32.45	Severe collinearity
PINWD	-0.45	0.09	0.0001	0.09	0.91	11.25	Severe collinearity
RUMPLN	-0.46	0.10	0.0001	0.03	0.97	31.89	Severe collinearity
CANNCIR	2.46	0.15	0.0001	0.14	0.86	7.33	Non-collinearity
CHCIR	-0.37	0.03	0.0001	0.03	0.97	38.22	Severe collinearity

* Variable acronyms as spelt out in Table 1

VIF in excess of 10 indicates severe collinearity (Rook *et al.* 1990)

Table 4: Variance proportions and eigenvalues of the partial regression coefficients in the prediction of body weight in Japanese Black cattle

No..	EValues	Intercept	WHT	HIPHT	BL	CHWD	SHWD	CHDP	HIPWD	LUVWD	THWD	PINWD	RUMP	CANN	CHCIR
1.	13.869	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2.	0.051	0.039	0.004	0.001	0.000	0.005	0.003	0.028	0.002	0.001	0.000	0.007	0.000	0.002	0.000
3.	0.036	0.002	0.000	0.000	0.000	0.001	0.873	0.025	0.000	0.000	0.000	0.000	0.000	0.002	0.000
4.	0.020	0.001	0.000	0.000	0.002	0.018	0.076	0.659	0.002	0.002	0.000	0.011	0.000	0.002	0.000
5.	0.008	0.000	0.001	0.001	0.003	0.593	0.037	0.046	0.003	0.000	0.006	0.142	0.001	0.013	0.001
6.	0.005	0.012	0.004	0.004	0.049	0.313	0.001	0.004	0.029	0.012	0.001	0.293	0.002	0.071	0.001
7.	0.004	0.000	0.008	0.008	0.066	0.019	0.002	0.130	0.047	0.001	0.017	0.392	0.023	0.093	0.015
8.	0.002	0.256	0.001	0.000	0.139	0.014	0.004	0.069	0.098	0.000	0.002	0.072	0.003	0.521	0.005
9.	0.001	0.008	0.000	0.000	0.005	0.001	0.002	0.005	0.315	0.053	0.238	0.012	0.244	0.221	0.003
10.	0.001	0.000	0.001	0.001	0.017	0.010	0.001	0.006	0.250	0.146	0.434	0.012	0.134	0.001	0.191
11.	0.0009	0.262	0.082	0.094	0.683	0.016	0.000	0.000	0.029	0.122	0.024	0.001	0.000	0.005	0.063
12.	0.0008	0.368	0.042	0.042	0.028	0.008	0.001	0.016	0.039	0.544	0.034	0.003	0.010	0.007	0.339
13.	0.0007	0.044	0.004	0.001	0.002	0.002	0.000	0.012	0.181	0.120	0.243	0.047	0.583	0.059	0.379
14.	0.0003	0.007	0.857	0.849	0.003	0.000	0.000	0.000	0.006	0.000	0.003	0.000	0.000	0.005	0.001

Thus the ridge estimates tend to be more stable as they are not affected by slight variations in the estimation data. The regression model can be defined in terms of X and Y as the matrices of the standardized independent and dependent variables respectively, and can be represented by: Y = Xβ + ì

where Y = *n x 1* vector of observations on a response variable

X = *n x p* matrix of observations on *p* explanatory variables

β = *p x 1* vector of regression coefficients

ì = *n x 1* vector of residuals

If X and Y are scaled such that X ' X and X ' Y are matrices of the correlation coefficients, then the least squares estimator of β = $(X ? X)^{-1}$ X ' Y. It is known that: p

E [(ˆβ - β) ? (ˆβ - β)] = 6^2 Ó $ë_j^{-1}$ J=1

where $ë_1$ = $ë_2$ = $ë_p$ are the eigenvalues of **X ' X**. This is a measure of the squared distance of the estimated regression coefficients from their true values, in other words, the total mean squared error. Thus, small eigenvalues are evidence of collinearity because when one or more eigenvalues are small, the total mean squared error is large indicating imprecision in the least squares estimates.

Ridge regression leads to an alternative estimator with a lower mean square error. The ridge regression estimator is indexed by a parameter *k* > 0 such that:

ˆβ (*k*) = $(X ' X + kI)^{-1}$ X ' Y

= $(X ' X + kI)^{-1}$ X ' Xβ where *k*I is a diagonal matrix with all elements consisting of an arbitrary small constant *k*. The total mean squared error is: E [(ˆβ(*k*) - β) ? (ˆβ(*k*) - β)]

= 6^2 trace [$(X ' X + kI)^{-1}$ X ' X$(X ' X + kI)^{-1}$] + k^2 β? $(X ' X + kI)^{-2}$ βp

= 6^2 Ó $ë_i (ë_i + k)^{-2}$ + k^2 β? $(X ' X + kI)^{-2}$ β i=1

The first term in this equation is the total variance of ˆβ(*k*) while the second term is the square of the bias. The bias increases with *k*, while the total variance falls. In other words, larger values of *k* reduce multicollinearity, but increase

Table 5: Principal components, eigenvalues (altent roots), eigenvectors (latent vectors) and proportions of variation of body measurements (redefined regression after adjusting intercept of the original variables) in the prediction of body weight in Japanese Black cattle

	Principal Components												
	1	2	3	4	5	6	7	8	9	10	11	12	13
EValue	11.398	0.586	0.364	0.184	0.129	0.110	0.079	0.038	0.033	0.026	0.020	0.017	0.016
Var Prop	0.877	0.045	0.028	0.014	0.010	0.008	0.006	0.003	0.002	0.002	0.002	0.001	0.001
EVector													
WHT	0.286	0.017	-0.251	0.276	-0.216	0.289	0.069	-0.288	0.105	0.017	0.058	0.315	-0.667
HIPHT	0.285	0.018	-0.261	0.288	-0.234	0.291	0.069	-0.414	0.030	-0.004	-0.081	-0.235	0.619
BL	0.287	0.003	-0.227	0.215	-0.183	0.177	0.101	0.809	-0.257	0.027	-0.082	0.117	0.099
CHWD	0.276	0.092	-0.181	0.224	0.907	-0.003	0.049	-0.029	-0.029	0.017	-0.036	0.055	0.023
SHWD	0.201	0.928	0.294	-0.071	-0.078	0.024	0.051	0.007	0.006	0.006	0.006	-0.001	0.003
CHDP	0.248	-0.252	0.753	0.516	-0.034	-0.128	-0.009	0.009	0.026	0.023	0.047	0.010	0.010
HIPWD	0.289	-0.049	-0.091	-0.092	-0.075	-0.331	0.142	0.061	0.567	0.331	0.092	0.413	0.239
LUVWD	0.292	-0.029	-0.123	-0.005	-0.021	-0.137	-0.332	0.160	0.266	-0.072	0.400	-0.737	-0.240
THWD	0.290	-0.088	0.030	-0.218	-0.042	-0.143	-0.130	-0.160	-0.470	0.687	-0.234	-0.196	-0.136
PINWD	0.282	-0.055	-0.071	-0.364	-0.052	-0.293	-0.086	-0.009	0.059	-0.112	0.055	0.091	0.029
RUMP	0.291	-0.067	-0.008	-0.126	-0.048	-0.197	0.814	-0.174	-0.524	-0.407	0.460	0.242	0.108
CANN	0.271	-0.207	0.316	-0.507	0.140	0.686	-0.318	0.057	0.144	-0.024	0.060	0.033	0.044
CHCIR	0.291	-0.071	0.021	-0.105	-0.048	-0.197	-0.074	-0.025	0.090	-0.483	-0.734	-0.091	-0.116

Table 6: Ridge estimator (k), residual mean square error (RMSE) and ridge coefficients of body measurements in the prediction of body weight in Japanese Black cattle

Ridge Coefficients															
k	RMSE	Intercept	WHT	HIPHT	BL	CHWD	SHWD	CHDP	HIPWD	LUVW	THWD	PINWD	RUMP	CANN	CHCIR
0.0	9.8239	-157.325	0.886	0.064	0.873	1.285	-0.141	0.163	1.771	4.802	-1.107	-0.453	-0.464	2.460	-0.368
0.2	10.846	0.764	0.572	0.498	0.448	1.113	0.019	0.133	0.885	1.484	0.306	0.448	0.425	1.175	0.113
0.4	11.334	-158.575	0.495	0.456	0.368	0.977	0.103	0.172	0.816	1.259	0.468	0.649	0.534	1.330	0.145
0.6	11.698	-153.316	0.451	0.426	0.329	0.901	0.151	0.199	0.783	1.160	0.530	0.730	0.572	1.443	0.156
0.8	12.017	-148.618	0.422	0.403	0.305	0.849	0.180	0.216	0.760	1.098	0.559	0.768	0.587	1.514	0.161
1.0	12.320	-144.279	0.401	0.386	0.287	0.810	0.200	0.228	0.741	1.053	0.574	0.788	0.593	1.557	0.163
1.2	12.619	-140.200	0.384	0.371	0.274	0.780	0.210	0.236	0.725	1.018	0.582	0.797	0.594	1.582	0.163

the bias, whereas a *k* of zero produces the least squares estimates. Since the aim of ridge regression is to pick a value of *k* for which the reduction in total variance is not exceeded by the increase in bias, the question becomes of determining what value of *k* should be used. The most commonly used procedure is to calculate the ridge regression coefficients for a set of values of *k* (Table 6) and plot the resulting regression coefficients against *k*. These plots, called ridge plots (Figure 1), often show large changes in the estimated coefficients for smaller values of *k*, which, as *k* increases, ultimately stabilize or "settle down" to a steady progression towards zero. An optimum value for *k* is said to occur when these estimates appear to "settle down" (Freund and Wilson 1998). An alternative quantitative measure of the stability of the ridge trace known as the Index of Stability of Relative Magnitudes (ISRM) was used by Rook *et al.* (1990). It was calculated as follows: $ISRM = Ó_i [(p (ë_i + k_i))^2 / œë_i) - 1]^2$ where $œ = Ó_i ë_i (ë_i + k_i)^2$.

Principal component and ridge regression analyses following the methods exhaustively described above, were carried out using PROC REG and PROC PRINCOMP procedures (SAS 2002) to analyse our data.

Results

Overall, the average body weight of the Japanese Black cattle utilised for this study was 91.79 kg. The highest body measurement of 100.24 cm was recorded for chest circumference while the lowest was 11.82 cm for cannon circumference (Table 1). In order to appraise the relationship between these body measurements, simple pairwise correlation coefficients were computed (Table 2). The correlations were all positive and highly significant ranging from 0.50 between CHDP and SHWD to the highest value of 0.98 between WHT and HIPHT, RUMPLN and CHCIR and HIPWD and LUVWD. It was obvious from Table 3 that all the partial regression coefficient estimates were highly significant ($P>0.0001$) for predicting body weight, except hip height ($P>0.197$). By statistical implication, hip height might as well be chucked out of the predictors without any significant consequence on body weight. The coefficients of determination (R^2) that generally indicate the relative precision or accuracy of the prediction were all high (up to 0.98) except for SHWD (0.44). However, the variance inflation factors (VIFs) gave the first indication of the existence of severe collinearity in 9 out of the 13 independent body measurement variables investigated (Table 3). As a further confirmation of the multicollinearity problem, principal component analysis was carried out. The variance proportions and eigenvalues

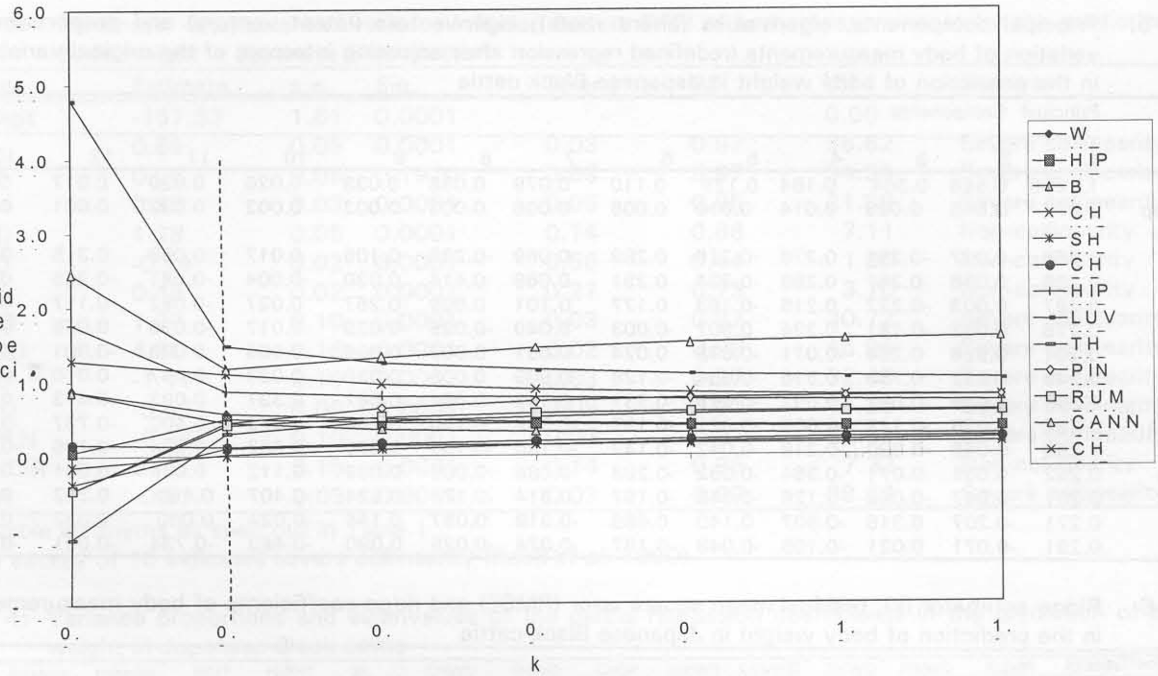

Fig. 1: Ridge trace plots of k value chosen as stable (..........)

of the partial regression coefficients are shown in Table 4. A close examination of this table reveals that there were 4 relatively small eigenvalues of 0.0009, 0.0008, 0.0007 and 0.0003 for components 11, 12, 13 and 14 respectively. These components with small eigenvalues had large variance proportions of 0.683 for BL, 0.849 for HIPHT and 0.857 for WHT indicating their involvement in multicollinearity. In the case of WHT and HIPHT, the large variance proportions of 0.857 and 0.849 in component 14, indicated a very strong, in-built correlation pattern among these two variables. Furthermore, the variables LUVWD in component 12 and RUMPLN in component 13 showed somewhat large variances of 0.544 and 0.583 respectively, also indicating that these variables were also involved in multicollinearity.

To remedy the severe multicollinearity problem using principal component analysis, principal components, eigenvalues, eigenvectors and variance proportions of the redefined regression after adjusting the intercept of the original variables, are presented in Table 5. It was obvious that the first principal component accounted for most of the total variance (0.877) with an eigenvalue of 11.398. It was also evident that the multicollinearity problems portrayed in Table 4 were adequately overcome as demonstrated by the uniform spread of the variance proportions from principal component 2 (0.045) right through to principal component 13 (0.001). The eigenvectors are simply coefficients for the transformation that show how the principal component variables relate to the original variables (Freund and Wilson 1998). It shows that the highest positive relationship 0.928 was between SHWD in principal component 2 followed by CHWD in principal component 5 (0.907). The highest negative relationships of -0.737 and -0.734 were observed in LUVWD in principal component 12 and CHCIR in principal component 11, respectively. The negative sign indicated antagonism between the principal component variables and the original variables. The implication was that since the highest positive relationship was obtainable for CHWD and SHWD, these two body measurements were probably the most important predictors for body weight in this study. However, a definite statement could only be made after examining the ridge estimator (k) and the new ridge coefficients (Table 6) as demonstrated by the ridge plot in Fig. 1. It shows that at k = 0.2, stability was achieved in all the body measurement coefficients having eliminated collinearity. Figure 1 clearly indicated that of all the body measurements utilised for the prediction of body weight, CHWD and SHWD were the most stable and HIPHT the least.

Discussion

The body measurements of cattle are known to be affected by many factors including, but not limited to, sex, breed, age and seasonal variations. For instance, Cestnik (2001) found that in Istrian cattle, withers height in bulls and cows averaged 145 and 138 cm respectively, indicating that the males had higher withers than females. In Charolais breed, Tozser *et al.* (2001) reported that withers height, rump width, chest girth and body length averaged 132.2, 52.1, 194.5 and 177.2 cm respectively in 6.8 years old cows. Similarly, in non-descript local bullocks, Varade and Ali (2001) reported that body measurements of chest girth, abdominal girth, body length and withers height averaged 161.25, 66.28, 148.29 and 130.46 cm respectively and were significantly affected by age. Rodriguez *et al.* (2001) found significant differences among ages for thoracic height, depth and circumference in which Creoli cattle with two teeth had lower body measurements than those with all dentition. Average withers height, body length and rump length for this Creoli breed were 119.17, 137.93 and 31.84 cm respectively, while in Jersey crossbreds, Roy *et al.* (2001) found that body length, heart girth and withers height averaged 146-150, 161-170 and 121-125 cm respectively. Therefore, it was necessary in our study, to adjust for age, sex, season and year effects. The present study showed that at an average body weight of 91.79 kg from birth to yearling age, the Japanese Black cattle had average withers height, hip height, body length and chest circumference of 84.21, 88.04, 84.09 and 100.24 cm respectively. These values fall within the range reported by Mukai *et al.* (1995) in a study of the genetic relationship between withers height, chest girth, chest depth, thurl width, body weight, daily gain and carcass traits in Japanese Black calves.

Our observation of significantly positive correlations between the body measurements was in agreement with Varade *et al.* (2001) and Enevoldsen and Kristensen (1997) who also reported that body measurements in cattle were significantly and positively correlated with each other. The implication is that in selection programs for growth, these relationships are very useful in indicating whether there are antagonisms between two traits incorporated in a selection index or not. In the present study where very high positive correlations of 0.98 were observed, it means selecting for one of the two correlated traits would automatically lead to an indirect selection for the other. A closely related and useful research tool is the utilisation of these body measurements in multiple regressions to predict body weight as exemplified in Holstein veal calves by Wilson *et al.* (1997) who used body length, heart girth, withers height and hip width at different ages to predict body weight.

The existence of severe collinearity as indicated by the VIFs implied that there would be high instability of the regression coefficients as a result of small changes in the estimation data. Therefore, such estimates would lead to poor prediction and possibly difficult interpretation of the underlying biological process. Freund and Wilson (1998) stated that principal components with large variances (eigenvalues) help in the interpretation of results of a regression where collinearity exists. However, when trying to diagnose the reasons for collinearity, the focus is on the principal components with very small eigenvalues because variables in multicollinearity are identifiable by their relatively large variance proportions with small eigenvalues. The variance proportions indicate the relative contribution from each principal component to the variance of each regression coefficient. Consequently, the existence of a relatively large contribution to the variances of several coefficients by a component with small eigenvalue may indicate which variables contribute to the overall multicollinearity. For easier comparison among coefficients, the variance proportions are standardized to sum up to 1. The highest positive relationship was obtainable for CHWD and SHWD, therefore, these two body measurements are probably the most important predictors for body weight in this study. However, a definite statement can only be made after examining the ridge estimator (k) and the new ridge coefficients (Table 6) as demonstrated by the ridge plot in Fig. 1. It shows that at $k = 0.2$, stability was achieved in all the body measurement coefficients having eliminated collinearity. Figure 1 clearly indicated that of all the body measurements utilized for the prediction of body weight, CHWD and SHWD were the most stable and HIPHT the least.

In conclusion, this study has demonstrated that the problem of multicollinearity in the relationship between body weight and body measurements in Japanese Black cattle can be solved by principal component and ridge regression analyses. Furthermore, of all the body measurements, CHWD and SHWD were the most stable and therefore, the most important in the prediction of body weight while HIPHT was the most unstable, hence the least important.

Acknowledgement

This study was carried out when the senior author was at the National Agricultural Research Center for Western Region (WeNARC) Oda, Shimane Prefecture, Japan. We are grateful to the Director, WeNARC Oda. The authors are grateful to the Japanese Society for the Promotion of Science (JSPS) for funding the research fellowships to Drs A.E.O. Malau-Aduli and M. Abdel-Aziz. We also thank Mrs Yuko Arima for partly assisting with data entry of monthly body measurements and body weights into the computer.

References

Cestnik, V., 2001. An interesting possibility of saving Istrian cattle from extinction. Veterinarska Stanica, 32: 163-166.

Enevoldsen, C. and T. Kristensen, 1997. Estimation of body weight from body size measurements and body condition scores in dairy cows. J. Dairy Sci., 80: 1988-1995.

Freund, R. J. and W. J. Wilson, 1998. Regression analysis: Statistical modelling of a response variable. Academic Press, San Diego, California, USA, pp: 444.

Gilbert, R. P., D. R. Bailey and N. H. Shannon, 1993. Linear body measurements of cattle before and after twenty years of selection for postweaning gain when fed two different diets. J. Anim. Sci., 71: 1712-1720.

Heinrichs, A. J., G. W. Rogers and J. B. Cooper, 1992. Predicting body weight and wither height in Holstein heifers using body measurements. J. Dairy Sci., 75: 3576-3581.

Hoerl, A. E. and R. W. Kennard, 1970a. Ridge regression: Biased estimation for non-orthogonal problems. Technometrics, 12: 55-67.

Hoerl, A. E. and R. W. Kennard, 1970b. Ridge regression: Applications to non-orthogonal problems. Technometrics 12: 69-82.

Hoerl, A. E., R. W. Kennard and K. F. Baldwin, 1975. Ridge regression: Some simulations. Communications in Statistics, 4: 105-124.

Karnuah, A. B., K. Moriya, N. Nakanishi, T. Nade, T. Mitsuhashi and Y. Sasaki, 2001. Computer image analysis for prediction of carcass composition from cross-sections of Japanese Black steers. J. Anim. Sci., 79: 2851-2856.

Kertz, A. F., L. F. Reutzel, B. A. Barton and R. L. Ely, 1997. Body weight, body condition score and wither height of prepartum Holstein cows and birth weight and sex of calves by parity: A database and summary. J. Dairy Sci., 80: 525-529.

Koenen E. P. and A. F. Groen, 1998. Genetic evaluation of body weight of lactating Holstein heifers using body measurements and conformation traits. J. Dairy Sci., 81: 1709-1713.

Magnabosco, D. U., M. Ojala, A. De los Reyes, R. D. Sainz, A. Fernandes and T. R. Famula, 2002. Estimates of environmental effects and genetic parameters for body measurements and weight in Brahman cattle raised in Mexico. J. Anim. Breed. Genet., 119: 221-228.

Marquardt D. W. and R. D. Snee, 1995. Ridge regression in practice. The American Statistician, 29: 3-19.

Mukai, F., K. Oyama and S. Kohno, 1995. Genetic relationships between performance test traits and field carcass traits in Japanese Black cattle. Livestock Prod. Sci., 44: 199-205.

Mukai F., A. Yamagami and K. Anada, 2000. Estimation of genetic parameters and trends for additive, direct maternal genetic effects for birth and calf market weights in Japanese Black beef cattle. Anim. Sci. J., 71: 462-469.

Rawlings, J. O., S. G. Pantula and D. A. Dickey, 1998. Applied regression analysis: A research tool. 2nd Edition, Springer-Verlag, New York, 657pp.

Rodriguez, M., G. Fernandez, C. Silveira and J. V. Delgado, 2001. Ethnic study of Uruguayan Creoli bovine. 1. Biometric analysis. Archivos de Zootecnia, 50: 113-118.

Rook, A. J., M. S. Dhanoa and M. Gill, 1990. Prediction of the voluntary intake of grass silages by beef cattle. 2. Principal component and ridge regression analyses. Anim. Prod., 50: 439-454.

Roy, P. K., R. C. Saha and R. B. Singh, 2001. Influence of body weight at calving and body measurements on various economic traits of Jersey crossbreds maintained under hot humid conditions of West Bengal. Indian J. Dairy Sci., 54: 53-56.

SAS, 2002. Statistical Analysis System, SAS Institute Inc., Cary, North Carolina, USA.

Shimada, K., Y. Izaike, O. Suzuki, M. Kosugiyama, N. Takenouchi, K. Ohshima and M. Takahashi, 1992. Effect of milk yield on growth of multiple calves in Japanese Black cattle (Wagyu). Asian-Austral. J. Anim. Sci., 5: 717-722.

Smith, G. and F. Campbell, 1980. A critique of some ridge regression methods. J. Amer. Stat. Assoc., 75: 74-81.

Smith, S. B., M. Zembayashi, D. K. Lunt, J. O. Sanders and C. D. Gilbert, 2001. Carcass traits and microsatellite distributions in offspring of sires from three geographical regions of Japan. J. Anim. Sci., 79: 3041-3051.

Sosa, J. M., P. L. Senger and J. J. Reeves, 2002. Evaluation of American Wagyu sires for scrotal circumference by age and body weight. J. Anim. Sci., 80: 19-22.

Tozser, J., Z. Domokos, L. Alfoldi, G. Hollo and J. Rusznak, 2001. Body measurements of cows in a Charolais herd with different genotypes. Allattenyesztes es Takarmanyozas, 50: 15-22.

Varade, P. K. and S. Z. Ali, 2001. Body measurements on non-descript bullocks of Buldhana District in Maharashtra. J. Appl. Zool. Res., 12: 71-72.

Vargas, C. A., M. A. Elzo, C. C. Chase and T. A. Olson, 2000. Genetic parameters and relationships between hip height and weight in Brahman cattle. J. Anim. Sci., 78: 3045-3052.

Wilson, L. L., C. L. Egan and T. L. Terosky, 1997. Body measurements and body weights of special-fed Holstein veal calves. J. Dairy Sci., 80: 3077-3082.