Estimation of the Ridge Constant: An Approach Based on the Condition Index[†]

B. A. Carnes, M. E. Ginevan[*], and D. M. Eychner

Division of Biological and Medical Research
Argonne National Laboratory
Argonne, Illinois 60439 USA

[†]Work supported by the U. S. Department of Energy under contract No. W-31-109-ENG-38.

[*]Present address: Mail Stop 1130SS, Nuclear Regulatory Commission, Washington, DC, 20555, USA.

DISCLAIMER

DISCLAIMER

A Monte Carlo simulation was conducted to evaluate the response of ridge regression solutions to increasing collinearity. Specifically, the magnitude of the ridge constant was sensitive to variability in the response vector while appearing insensitive to collinearity. The insensitivity to collinearity was especially apparent when the response vector was oriented toward a minor dimension in the structure of the predictor variables. Limit arguments indicated that, in the minor dimension case, the increasing magnitude of the inner product of the least squares coefficient vector in the denominator of the optimum estimate for the ridge constant was responsible for the insensitivity to collinearity. Since ridge regression was proposed to deal with collinearity, this behavior suggests a defect in the generally used estimate for the optimum ridge constant. To solve this problem, we propose a condition index estimate for the ridge constant based on the eigenvalues of the augmented X'X matrix; this approach ensures sensitivity to collinearity and insensitivity to the behavior of the response vector.

The problem of variance inflation in ordinary least squares (OLS) regression coefficients when linear dependencies (collinearity) exist in the data structure is well known (Silvey, 1969; Marquardt, 1970; McCallum, 1970; Greenberg, 1975; Mason et al., 1975; Gunst and Mason, 1977). Ridge regression (Hoerl et al., 1962, 1964, 1970a, 1970b, 1975) is one of the best known and most often used procedures for dealing with this difficulty. Indeed, simulation results (Lawless and Wang, 1976; Dempster et al., 1977) suggest that ridge-type estimators are superior within the class of biased estimators. It has been suggested that the orientation of β, the magnitude of β relative to σ and the strength of the collinearities (Gunst and Mason, 1976a, 1976b) are important factors affecting the performance of biased estimators.

Simulation

A Monte Carlo simulation was designed to observe the response of ridge solutions to changes in the orientation and variability of the response vector (Y). A simple two predictor variable data structure was employed. Regression algorithms to solve for ridge coefficients

$$\hat{\beta} = (X'X + kI)^{-1}X'Y \qquad (1)$$

followed the format of Gunst and Mason (1977). Estimation of the ridge constant

$$k = p\hat{\sigma}^2/\hat{\beta}'_{LS}\hat{\beta}_{LS} \qquad (2)$$

where p (number of predictor variables) = 2, and $\hat{\sigma}^2$, $\hat{\beta}_{LS}$ are OLS estimators was consistent with previous studies (Hoerl and Kennard, 1975; Hoerl et al., 1975; Gunst and Mason, 1977).

The first step in the simulation was the generation of 30 bivariate random normal data points (X_1, X_2) with means zero, unit variance, and zero correlation. This (30x2) matrix was then obliquely rotated such that the resulting (2x2) correlation matrix of predictor variables could have any desired correlation structure. Eigenvalues and eigenvectors were extracted from the correlation matrix and used to generate two vectors of principal components (PC) scores (major and minor dimensions).

The response vector (Y) was taken as the PC scores on either the major or minor dimension. The PC scores were standardized to unit variance and scaled by the addition of a 30x1 random vector with mean zero and variance C (C=0, 0.1, 0.2, 0.5, 1.0, 2.0). Thus the R^2 of the full regression model ranged from 1.0 to about 0.4. For all simulations $|r_{1Y}| = |r_{2Y}|$; that is, the correlation of the predictors with Y was equal but differed in sign for the case of Y equal to a major dimension.

Eight settings of correlation between predictor variables were selected (r=-.5, -.6, -.7, -.8, -.9, -.95, -.975, -.99). One hundred simulations were generated for each correlation setting and each orientation of the response vector with a minor or major dimension of the data structure. Within each simulation regression, solutions were generated for each of the six variability settings of the response vector. Summary statistics were generated for each set of 100 simulations.

Simulation Results

The criticisms that have been leveled at ridge regression involve the dependence of the ridge constant (k) on the predictor variables (Conniffe and Stone, 1973). However, the most striking behaviors of k emerging from our simulations were (1) the sensitivity of k to increasing variability in the response vector and (2) the nonsensitivity of k to increasing collinearity, especially when the response variable was associated with a minor dimension of the data structure.

As seen from the ridge solution for $\hat{\beta}$ (equation 1), k is added to the main diagonal of the X'X matrix. The eigenvalues, which are diagnostics of collinearity (Silvey, 1969; Greenberg, 1975; Chatterjee and Price, 1977), extracted from the augmented matrix are related to those extracted from the original X'X matrix (Hawkins, 1975; Green and Carroll, 1976).

$$\lambda_R = \lambda_j + k \tag{3}$$

The ridge constant should increase as collinearity becomes more severe (λ_{min} tends toward zero). The condition index, $\lambda_{max}/\lambda_{min}$, from the augmented ridge eigenvalues should become smaller as k becomes larger resulting in an improvement in the condition of X'X and a reduction in the variance inflation of the estimated regression coefficients (Marquardt, 1970). It is not apparent that the response vector should have an impact on the determination of k. However, for any fixed level of collinearity, the variability of the ridge constant was directly proportional to increasing variability in the response variable (Table 1). It was further noted that for a fixed level of variability in the response vector, the magnitude of k was remarkably similar across levels of collinearity, indicating that k responds more to variability in the response variable than to collinearity within the data structure of predictor variables responsible for the variance inflation of regression coefficients (Table 1). Finally, it was noted that the magnitude of k was consistently smaller at all levels of variability in the response variable and levels of collinearity in the minor dimension case when compared with the counterpart in the major dimension (Table 1).

Limit Arguments

The discrepancies in the behavior of $\hat{k}$ in the simulations from the expected behavior led to a closer look at the estimation of k (equation 2).

Attention was focused on the OLS coefficient vectors in the denominator. When unit vectors are employed (as in our simulation) the solution for the coefficient vector can be expressed in terms of correlations.

$$\hat{\beta} = \frac{(r_{1Y} - r_{12}r_{2Y})/(1-r_{12}^2)}{-r_{12}r_{1Y} + r_{2Y}/(1 - r_{12}^2)} \tag{4}$$

Sastry (1970), consistent with other reports (Fox and Cooney, 1954; Klein and Nakamura, 1965), using L'Hopitals rule concluded that

$$\lim \hat{\beta}_1 = \frac{r_{1Y}}{2} = \frac{r_{2Y}}{2} \tag{5}$$

as $r_{12} \rightarrow 1$ or $r_{12} \rightarrow -1$. This limit is consistent with our results when the response vector was associated with a major dimension; however, this limit conflicts with our results when Y was associated with a minor dimension (Table 2).

When the response variable is set equal to the major dimension (c = o in our simulation) $r_{1Y} = -r_{2Y}$. This allows a simplification in the solution for $\hat{\beta}$,

$$\hat{\beta} = \frac{r_{1Y}/(1 - r_{12})}{-r_{1Y}/(1 - r_{12})} \tag{6}$$

At this stage, a limit cannot be determined because the numerator does not remain constant as r_{12} approaches -1. However, all the correlations and, therefore, $\hat{\beta}$ can be expressed as a function of a single angle (Figure 1)

$$r_{1Y} = \cos(\theta_{1Y})$$

$$r_{2Y} = \cos(\theta_{2Y}) = \cos(180 - \theta_{1Y}) \quad (7)$$
$$r_{12} = \cos(180 - 2\theta_{1Y}) \ .$$

Substituting into (6) gives

$$\hat{\beta} = \begin{array}{c} \cos(\theta_{1Y})/(1 - \cos(180 - 2\theta_{1Y})) \\ - \cos(\theta_{1Y})/(1 - \cos(180 - 2\theta_{1Y})) \end{array} \quad (8)$$

As r_{12} approaches -1, θ_{1Y} approaches zero; therefore,

$$\lim \hat{\beta}_1 = 1/2$$
$$\lim \hat{\beta}_2 = -1/2 \ . \quad (9)$$

However, when Y is set equal to a minor dimension, $r_{1Y} = r_{2Y} = r_{iY}$ and

$$\hat{\beta} = r_{iY}/(1 + r_{12}) \ . \quad (10)$$

Again, $\hat{\beta}$ and the correlations can be expressed as a funciton of a single angle (Figure 2)

$$r_{iy} = \cos(\theta)$$
$$r_{12} = \cos(2\theta) \quad (11)$$

which upon substitution yields

$$\hat{\beta} = \cos(\theta)/(1 + \cos 2\theta) \ . \quad (12)$$

As r_{12} approaches -1, θ approaches 90 resulting in no limit for $\hat{\beta}$ when expressed in this form. Applying L'Hopital's rule, we have a limit that does exist as $\theta \rightarrow 90$,

$$\lim \hat{\beta} = \lim \frac{\sin \theta}{2 \sin 2\theta} = \infty. \qquad (13)$$

This limit, while conflicting with Sastry's (1970) results, is consistent with the observed outcome from our simulation (Table 2).

Further, these results have a major impact on the estimation of $\hat{k}$ (equation 2) when the response vector is associated with a minor dimension of the data structure. As X_1 and X_2 become more negatively correlated (r_{12} approaches -1), the denominator of $\hat{k}$, $(\hat{\beta}_{LS}'\hat{\beta}_{LS})$, becomes larger thus forcing $\hat{k}$ to become smaller. This is precisely the opposite behavior one would hope for $\hat{k}$ in a ridge solution when collinearity is increasing.

An Alternative Estimate of k

The aberrant behavior of k demonstrated in our simulation might be circumvented by using the condition index c which is defined as

$$c = \lambda max/\lambda min \qquad (14)$$

(Belsley et al., 1980). When the main diagonal of X'X is augmented, as in equation (1), the resulting condition index c' is given by

$$c' = (\lambda max + \hat{k})/(\lambda min + \hat{k}). \qquad (15)$$

It would seem reasonable to choose $\hat{k}$ such that a relatively well conditioned problem results. [Matrices with condition indices less than 10 (Belsley et al., 1980)]. Once an arbitrary value for c' is chosen one can solve for $\hat{k}$

$$\hat{k} = (\lambda max - c'\lambda min)/(c'-1). \qquad (16)$$

No augmentation would be necessary when the condition index is less than or equal to c. Scaling the ridge constant directly to the eigenstructure ensures sensitivity to collinearity. It would be interesting to compare the ridge solutions thus generated to those from a principal component (PC) regression which is also directly linked to the eigenstructure of the predictor variables. Implementing our proposed decision rule would create a dilemma similar to that found in PC regression. In PC regression, the primary explanatory power of the predictors may be in deleted minor dimensions of the data structure resulting in a much reduced R^2. In this case, the modified ridge solution would likely require a rather large k and would also show a great reduction in R^2. However, when the explanatory power is in minor dimensions where precise estimation is not possible (Silvey, 1969) it seems reasonable that this imprecision should be reflected in poor model performance.

Conclusions

The constant used to augment the main diagonal of the X'X matrix in ridge regression is sensitive to the variability and orientation of the response vector. The smaller $\hat{k}$ in the minor dimension relative to the major dimension reflects the different limits for $\hat{\beta}_{LS}$ in the two cases. A modified ridge estimate based on the condition index should eliminate the aberrant behaviors of $\hat{k}$ observed in this simulation. Association of the response variable with minor dimensions of the data structure indicates deficiencies in the data and should be reflected in the performance of the regression model.

References

Belsley, D. A., E. Kuh, and R. E. Welsch. 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. John Wiley & Sons, New York.

Chatterjee, S and B. Price. 1977. Regression Analysis by Example. John Wiley & Sons, New York.

Conniffe, D. and J. Stone. 1973. A critical view of ridge regression. The Statistician 22:181-187.

Dempster, A. P., M. Schatzoff, and N. Wermuch. 1977. A simulation study of alternatives to ordinary least squares. Journal of the American Statistical Association 72:77-106.

Fox, K. A. and J. F. Cooney. 1954. Effects of intercorrelations upon multiple correlation and regression measures. U. S. Department of Agriculture, Agricultural Marketing Service, Washington, DC.

Green, P. E. and J. D. Carroll. 1976. Mathematical Tools for Multivariate Analysis. Academic Press, New York.

Greenberg, E. 1975. Minimum variance properties of principal component regression. Journal of the American Statistical Association 70:194-197.

Gunst, R. F., J. T. Webster, and R. L. Mason. 1976a. A comparison of least squares and latent root regression estimators. Technometrics 18:75-83.

__________, and R. L. Mason. 1976b. Generalized mean squared error properties of regression estimators. Communication in Statistical Theory and Methodology A5:1501-1508.

__________, and R. L. Mason. 1977. Biased estimation in regression: an evaluation using mean squared error. Journal of the American Statistical Association 72:616-628.

Hawkins, D. M. 1973. On the investigation of alternative regressions by principal components. Applied Statistics 22:275-286.

Hoerl, A. E. 1962. Application of ridge analysis to regression problems. Chemical Engineering Progress 58:54-59.

__________. 1964. Ridge analysis. Chemical Engineering Progress Symposium Series 60:67-78.

__________ and R. W. Kennard. 1970a. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12:55-67.

__________ and R. W. Kennard. 1970b. Ridge regression: applications to nonorthogonal problems. Technometrics 12:69-82.

__________, R. W. Kennard, and K. F. Baldwin. 1975. Ridge regression: some simulations. Communication in Statistics 4:105-123.

Klein, L. R. and M. Nakamura. 1962. Singularity in the equation systmes of econometrics, some aspects of multicollinearity. International Economic Review 3:274-299.

Lawless, J. F. and P. Wang. 1976. A simulation study of ridge and other regression estimators. Communications in Statistics - Theoretical Methods A5(4):307-323.

Marquardt, D. W. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12:591-612.

Mason, R. L., R. F. Gunst, and J. T. Webster. 1975. Regression analysis and problems of multicollinearity. Communications in Statistics 4:277-292.

McCallum, B. T. 1970. Artificial orthogonalization in regression analysis. Review of Economics and Statistics 52:110-13.

Sastry, M. V. R. 1970. Some limits in the theory of multicollinearity. The American Statistician 24:39-40.

Silvey, S. D. 1969. Multicollinearity and imprecise estimation. Journal of the Royal Statistical Society B31:539-552.

TABLE 1

Summary of the simulation results on the behavior of the ridge constant (k), the standard error of the ridge constant ($S_{\hat{k}}$), the condition index (C.I.) for eight correlation settings between predictor variables (r_{12}), six levels of variability in the response vector (Y_1, Y_2, ..., Y_6), and two orientations of the response vector (major and minor dimension).

r_{12}		Major Dimension $\hat{k}$	Major Dimension $S_{\hat{k}}$	Major Dimension C.I.	Minor Dimension $\hat{k}$	Minor Dimension $S_{\hat{k}}$	Minor Dimension C.I.
	Y1	.0000	.0000	3.000	.0000	.0000	3.000
	2	.0201	.0034	2.923	.0069	.0011	2.973
	3	.0505	.0132	2.816	.0178	.0050	2.931
-.50	4	.1016	.0392	2.662	.0385	.0177	2.857
	5	.2068	.1270	2.595	.0966	.1122	2.676
	6	.6497	1.3867	1.870	.4405	1.1238	2.063
	Y1	.0000	.0000	4.000	.0000	.0000	4.000
	2	.0213	.0036	3.848	.0055	.0009	3.959
	3	.0530	.0141	3.649	.0143	.0040	3.896
-.60	4	.1053	.0414	3.375	.0309	.0142	3.785
	5	.2102	.1323	2.966	.0780	.0907	3.510
	6	.6335	1.3287	2.161	.3673	.9482	2.564
	Y1	.0000	.0000	5.666	.0000	.0000	5.666
	2	.0224	.0040	5.342	.0041	.0007	5.604
	3	.0549	.0152	4.945	.0107	.0030	5.506
-.70	4	.1072	.0439	4.438	.0232	.0107	5.332
	5	.2090	.1378	3.750	.0589	.0687	4.901
	6	.5966	1.2149	2.561	.2869	.7479	3.385
	Y1	.0000	.0000	9.000	.0000	.0000	9.000
	2	.0232	.0044	8.168	.0027	.0005	8.893
	3	.0557	.0166	7.130	.0072	.0020	8.722
-.80	4	.1057	.0469	6.234	.0155	.0072	8.424
	5	.1996	.1431	5.004	.0396	.0462	7.678
	6	.5283	1.0257	3.197	.1992	.5231	5.008

(contd.)

r_{12}		Major Dimension			Minor Dimension		
		$\hat{k}$	$S_{\hat{k}}$	C.I.	$\hat{k}$	$S_{\hat{k}}$	C.I.
	Y1	.0000	.0000	19.000	.0000	.0000	19.000
	2	.0232	.0052	15.610	.0014	.0002	18.751
	3	.0530	.0191	12.765	.0036	.0010	18.374
-.90	4	.0959	.0512	10.188	.0078	.0036	17.698
	5	.1720	.1451	7.618	.0199	.0233	16.012
	6	.4039	.7424	4.572	.1038	.2739	9.832
	Y1	.0000	.0000	40.000	.0000	.0000	40.000
	2	.0218	.0062	27.462	.0007	.0001	38.475
	3	.0471	.0215	20.567	.0018	.0005	37.680
-.95	4	.0813	.0537	15.471	.0039	.0018	36.250
	5	.1388	.1387	11.064	.0100	.0117	$32.6\overline{6}$
	6	.2967	.5682	6.480	.0530	.1401	19.447
	Y1	.0000	.0000	79.000	.0000	.0000	79.000
	2	.0195	.0073	44.820	.0003	.0001	78.075
	3	.0396	.0231	31.186	.009	.002	76.290
-.975	4	.0653	.0535	22.595	.0019	.0009	73.491
	5	.1066	.1264	15.818	.0050	.0059	66.000
	6	.2137	.4643	9.169	.0268	.0708	38.645
	Y1	.0000	.0000	199.000	.0000	.0000	199.000
	2	.0156	.0084	78.344	.0001	.0000	197.040
	3	.0292	.0234	51.510	.0004	.0001	191.385
-.99	4	.0457	.0496	36.548	.0008	.0004	$184.3\overline{3}$
	5	.0712	.1069	25.384	.0020	.0023	166.000
	6	.147	.3503	14.683	.0108	.0285	96.192

TABLE 2

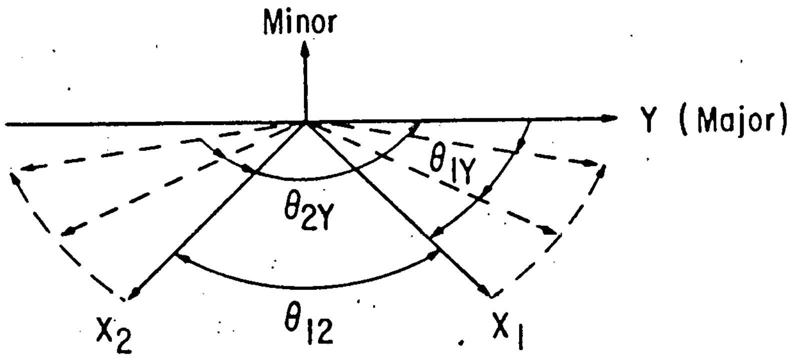
Summary of the correlations (r_{iY}) between the response vector and the predictor variables and the resulting regression coefficients ($\hat{\beta}$) when the response vector (Y) is associated with a minor dimension ($r_{1Y} = r_{2Y} = r_{iY}$; $\hat{\beta}_1 = \hat{\beta}_2 = \hat{\beta}_i$) or major dimension for each of the eight correlation (r_{12}) settings between predictor variables.

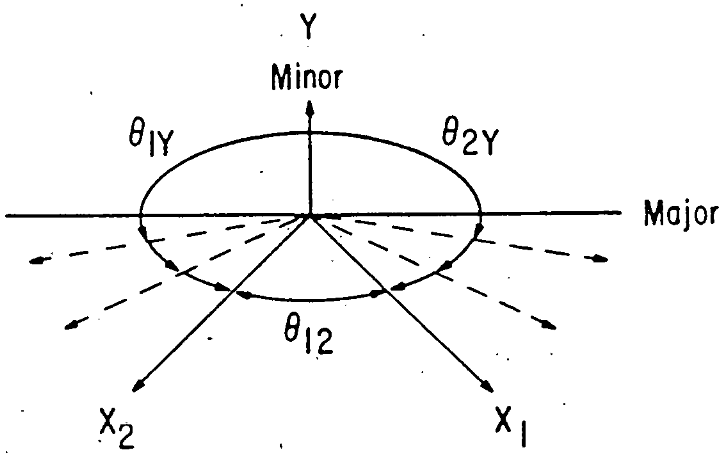
	Minor		Major			
r_{12}	r_{iy}	$\hat{\beta}_i$	r_{1Y}	r_{2Y}	$\hat{\beta}_1$	$\hat{\beta}_2$
-.5	-.5	-1.000	-.866	.866	-.5773	.5773
-.6	-.447	-1.118	-.8944	.8944	-.5590	.5590
-.7	-.387	-1.291	-.9219	.9219	-.5423	.5423
-.8	-.316	-1.581	-.9487	.9487	-.5270	.5270
-.9	-.224	-2.236	-.9747	.9747	-.5130	.5130
-.95	-.158	-3.162	-.9874	.9874	-.5064	.5064
-.975	-.112	-4.472	-.9937	.9937	-.5031	.5031
-.99	-.071	-7.071	-.9975	.9975	-.5012	.5012

Figure Legends

Figure 1. Simulation geometry of the predictor and response vectors when Y is associated with a major dimension.

Figure 2. Simulation geometry of the predictor and response vectors when Y is associated with a minor dimension.


Minor
Y (Major)
θ_{1Y}
θ_{2Y}
θ_{12}
X_2
X_1


Y
Minor
θ_{1Y}
θ_{2Y}
Major
θ_{12}
X_2
X_1