RIDGE REGRESSION ESTIMATION PROCEDURES APPLIED TO CANONICAL CORRELATION ANALYSIS

BU-509-M

May, 1974

Donald A. Anderson* and Edward J. Carney**
Biometrics Unit
Cornell University
Ithaca, New York 14850

* On leave from the University of Wyoming.
** On leave from the University of Rhode Island.

RIDGE REGRESSION ESTIMATION PROCEDURES APPLIED TO CANONICAL CORRELATION ANALYSIS

The principles of ridge regression analysis are extended to canonical correlation analysis in order to counteract the inherent instability of the usual procedures. The mathematical formulation of the ridge procedure is presented, and the effect on the instability of estimates is evaluated via a Monte Carlo simulation from a specified population. The simulation demonstrates the instability of the usual analysis, and indicates that the ridge analysis produces dramatic improvements in the stability and interpretability of the resultant estimates.

The study of relationships between two sets of random variables was begun in a brief paper by Hotelling (1935). His definitive study of the problem appeared the next year in a fifty-seven page paper, Hotelling (1936), which still stands as a key reference in multivariate literature.

The most familiar situation of this type is multiple regression analysis wherein one of the sets contains only one variable, the 'dependent' variable, and the second set contains several 'independent' variables. Regression procedures are designed to determine the relationship between independent and dependent variables. A common application is estimation of the mean, or prediction, of the dependent variable given values of the independent variables. On occasion, one would like to interpret directly the functional relationship or even individual regression coefficients.

Suppose there are several dependent variables which are to be predicted from the same set of independent variables. One approach is simply to perform the usual regression analysis with each of the dependent variables separately on the common set of independent variables. This procedure, while widely used, neglects the inner relationships between dependent variables. A more satisfactory procedure is a multivariate regression analysis which simultaneously provides estimates of all regression coefficients and an estimate of the variance covariance matrix of the vector of dependent variables. A third and perhaps more natural procedure is to determine that linear combination of the dependent variables which is "most predictable" in terms of the independent variables. That is, to obtain linear combinations of both dependent and independent variables such that these two combinations have maximum correlation. The process is repeated to obtain second, third, $\cdots$, pairs of linear combinations which at each step have maximum correlation. This third procedure is essentially the method of canonical correlation analysis proposed by Hotelling (1935, 1936).

Canonical correlation analysis has obvious mathematical appeal. Its acceptance as a useful statistical tool; however, is substantially less than might be expected. One reason for this is that estimates of canonical variates are, for small or moderate samples, highly unstable and individually uninterpretable. A simple illustration of this phenomenon in the multiple linear regression in the presence of multicolinearity.

The instability of canonical analysis was examined by Thorndike and Weiss (1973) by partitioning a relatively large sample into two parts, applying the analysis to each part, and cross validating the resultant canonical correlations and canonical variates. From their study they conclude that the instability is present and that more careful attention to canonical analysis method is required. Suggested methods include obtaining samples of sufficient size to permit at least a holdout group for cross validation.

For multiple regression in the presence of multicolinearity, ridge regression estimation as introduced by Hoerl and Kennard (1970a, b) has been shown to circumvent many of the difficulties of instability and non-interpretability. These estimates produce a substantially smaller expected mean square error of the estimates while allowing a slight bias in the estimates. The purpose of this paper is to apply the principles of ridge estimation to canonical correlation analysis. A brief discussion of canonical correlation is given before introducing the ridge estimates of the canonical variates and their correlations. The ridge estimates are assessed via a Monte Carlo simulation from a given correlation structure with two sets of five variables each.

The results of the simulation demonstrate very dramatically the instability of the usual estimates of canonical variates. In fact, the length of the difference between the estimated and the true canonical variate can easily be longer than the true vector itself, and often two to three times as long. Thus the estimated vector can lie in any quadrant, and for interpretation is completely useless. Further, it is shown that there is an upward bias in the estimates of the canonical correlations tending to indicate stronger relationships than are actually present.

For the case under simulation, the ridge regression estimates are shown to be much more stable (by factors of several hundred per cent in some cases). The mean length of the difference between observed and true canonical variate is reduced from over one hundred per cent to the order of thirty to fifty per cent. The probability of obtaining estimates of canonical variates close to the true variates with ridge estimates is therefore much greater. For purposes of interpretation, the ridge estimates are more reliable yielding coefficients with correct sign and approximate correct relative magnitudes. For small to moderate, $k_1, k_2 \leq .3$, values of the ridge constants the estimates of the canonical correlation are more nearly unbiased. For larger values of the ridge constants, the estimates are biased downward.

We conclude from the extensive mathematical and empirical results on ridge regression analysis and from the limited extent of the simulation results of this paper that for moderate (or small) sample sizes the usual canonical correlation analysis is unsatisfactory. The estimates are generally uninterpretable and there is a high probability of over estimates of the canonical correlations. The application of ridge analysis can provide more interpretable results and does in fact provide a more complete analysis.

Canonical Correlation Analysis

Let $\underline{y}(p_1 \times 1)$ and $\underline{x}(p_2 \times 1)$ be two sets of random variables, $p_1 \leq p_2$, whose joint covariance matrix

$$(2.1) \qquad \mathrm{Cov}\begin{pmatrix} \underline{y} \\ \underline{x} \end{pmatrix} = \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

exists and is positive definite. Then there exists a matrix

$$(2.2) \qquad D = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

such that

$$(2.3) \qquad D\,\Sigma\,D' = \Phi = \begin{bmatrix} I & \Phi_{12} \\ \Phi'_{12} & I \end{bmatrix},$$

where

$$(2.4) \qquad \Phi_{12} = \begin{bmatrix} \rho_1^2 & & & & \cdot & \\ & \rho_2^2 & & & \cdot & \\ & & \cdot & & \cdot & 0 \\ & & & \cdot & & \\ & & & \cdot\,\rho_{p_1}^2 & \cdot & \end{bmatrix}.$$

In (2.4) the values $\rho_1^2, \rho_2^2, \cdots, \rho_{p_1}^2$ are the characteristic roots of

$$(2.5) \qquad \Sigma_{11}^{-1/2}\,\Sigma_{12}\,\Sigma_{22}^{-1}\,\Sigma_{21}\,\Sigma_{11}^{-1/2} = T ,$$

and $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_{p_1}$ are the canonical correlations.

Denote by $A^*(p_1 \times p_1)$ the matrix whose rows are characteristic vectors of T corresponding to $\rho_1, \rho_2, \cdots, \rho_{p_1}$. Then

$$(2.6) \qquad A = A^* \Sigma_{11}^{-1/2} = \begin{bmatrix} \underset{\sim}{\alpha}'_1 \\ \underset{\sim}{\alpha}'_2 \\ \vdots \\ \underset{\sim}{\alpha}'_{p_1} \end{bmatrix}$$

is the matrix whose rows are the canonical variates for the first set. Further, let B^* be a matrix satisfying

$$(2.7) \qquad A^* \Sigma_{11}^{-1/2} \Sigma_{12} \Sigma_{22}^{-1/2} = \Phi_{12} B^* ,$$

the first p_1 rows being determined and the remaining p_2-p_1 being any orthogonal completion. Then

$$(2.8) \qquad B = B_2^* \Sigma_{22}^{-1/2} = \begin{bmatrix} \underset{\sim}{\beta}' \\ \underset{\sim}{\beta}'_2 \\ \vdots \\ \underset{\sim}{\beta}'_{p_1} \\ \underset{\sim}{\delta}_1 \\ \vdots \\ \underset{\sim}{\delta}_{p_2-p_1} \end{bmatrix}$$

has as its first p_1 rows the coefficients for the canonical variates of the second set.

Thus for the population we have p_1 pairs of canonical variates $(\underline{\alpha}_i' \underline{y}, \underline{\beta}_i' \underline{x})$ with canonical correlation ρ_i, $i = 1, 2, \cdots, p_1$. The canonical variates have unit variance, are uncorrelated with other variates with the same set, and corr $(\underline{\alpha}_i' \underline{y}, \underline{\beta}_j' \underline{x}) = 0$ if $i \neq j$. The collection of canonical variates accounts for all existing linear relations among the two sets of variates. The structure of these variates can yield considerable insight to the structure of the population.

In practice it is necessary to estimate the canonical variates and the canonical correlations from sample data. Thus if $\hat{\Sigma}$, $\hat{\Sigma}_{11}$, etc. denote estimates we have the sample covariance matrix

$$(2.9) \qquad \hat{\Sigma} = \begin{bmatrix} \hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22} \end{bmatrix}$$

and

$$(2.10) \qquad \hat{\Sigma}_{11}^{-1/2} \hat{\Sigma}_{12} \hat{\Sigma}_{22}^{-1} \hat{\Sigma}_{21} \hat{\Sigma}_{11}^{-1/2} ,$$

the estimates obtained from these as with population values (2.1) and (2.5).

It would appear from the form of (2.5) that if there is a high degree of multicolinearity evidenced in Σ, the estimates obtained from (2.9) and (2.10) will be highly unstable with different samples from the same population yielding vastly different estimates of the canonical variates. Individual coefficients may be excessively large, too small, or even of incorrect sign. The expected mean square errors

$$(2.11)\quad E\{(\hat{\underline{\alpha}}_i-\underline{\alpha}_i)'(\hat{\underline{\alpha}}_i-\underline{\alpha}_i)\} \quad \text{and} \quad E\{(\hat{\underline{\beta}}_i-\underline{\beta}_i)'(\hat{\underline{\beta}}_i-\beta)\} \qquad i=1,2,\cdots,p_1$$

will be large. Thus the experimenter attempting to discover the structural relations among the random variables will be frustrated or even led to erroneous conclusions.

For most applications, stable estimates with small expected mean square error and small bias are desired. In the special case $p_1 = 1$ which corresponds to multiple linear regression, ridge regression estimates have been shown by Hoerl and Kennard (1970a, b), to produce estimates with these properties. We consider next the extension of ridge estimation procedures to the canonical correlation problem.

Ridge Analysis for Canonical Correlation

Suppose first that the random variables in the population have been standardized so that Σ is the correlation matrix. Further, let $\hat{\Sigma}$ denote the sample correlation matrix, an estimate of Σ. The ridge estimates of the canonical variates and correlations are based on

$$(3.1)\qquad \begin{bmatrix} \hat{\Sigma}_{11}+k_1 I & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22}+k_2 I \end{bmatrix}$$

rather than $\hat{\Sigma}$, where k_1 and k_2 are constants. Thus the squares of the ridge estimates of canonical correlations, say $\tilde{\rho}_i^2(k_1,k_2)$ $i = 1, 2, \cdots, p_1$, are characteristic roots of

$$(3.2)\qquad \left(\tilde{\Sigma}_{11}+k_1 I\right)^{-1/2} \tilde{\Sigma}_{12}\left(\tilde{\Sigma}_{22}+k_2 I\right)^{-1} \tilde{\Sigma}_{21}\left(\tilde{\Sigma}_{12}+k_1 I\right)^{-1/2} .$$

The corresponding ridge estimates of coefficients in the canonical variates $\tilde{\underline{\alpha}}_i(k_1,k_2)$ and $\tilde{\underline{\beta}}_i(k_1,k_2)$ are obtained computationally from (3.2) as from (2.5) in section 2. Thus the ridge estimates are well defined as functions of k_1 and k_2.

To investigate the stability, mean square error, and bias of the ridge estimates a simulation experiment is conducted. Each set is taken to have five variables and the population correlation matrix is given by

$$(3.3)\quad \Sigma = \begin{bmatrix} 1.0 & -.107 & .089 & .309 & .530 & .613 & .085 & .506 & .013 & .003 \\ & 1.0 & .089 & .309 & .530 & .103 & .595 & -.005 & .523 & .003 \\ & & 1.0 & .346 & .587 & .248 & .228 & .426 & .446 & .003 \\ & & & 1.0 & .735 & .384 & .459 & .459 & .384 & .004 \\ & & & & 1.0 & .553 & .553 & .553 & .553 & -.022 \\ & & & & & 1.0 & .184 & .633 & .143 & .000 \\ & & & & & & 1.0 & .143 & .633 & .000 \\ & & & & & & & 1.0 & .184 & .000 \\ & & & & & & & & 1.0 & .000 \\ & & & & & & & & & 1.000 \end{bmatrix}$$

The characteristic roots of Σ_{11}, and Σ_{22} are, respectively,

characteristic Roots Σ_{11}: 2, 1.5, 1, .4, .1, and

characteristic Roots Σ_{22}: 2.4, 1.6, .5, .4, .1 .

The population canonical correlations are

$$.8, .6, .4, .2, .11 ,$$

and the population canonical variates have coefficients given by

$$A = \begin{bmatrix} \underset{\sim}{\alpha}_1' \\ \underset{\sim}{\alpha}_2' \\ \underset{\sim}{\alpha}_3' \\ \underset{\sim}{\alpha}_4' \\ \underset{\sim}{\alpha}_5' \end{bmatrix} = \begin{bmatrix} .274 & .274 & .245 & .202 & .407 \\ .672 & -.672 & .000 & .000 & .000 \\ .492 & .492 & -.846 & .017 & -.039 \\ .477 & .477 & .429 & -1.181 & .115 \\ -1.846 & -1.846 & -1.668 & -1.318 & -4.052 \end{bmatrix}$$

$$B = \begin{bmatrix} \underset{\sim}{\beta}_1' \\ \underset{\sim}{\beta}_2' \\ \underset{\sim}{\beta}_3' \\ \underset{\sim}{\beta}_4' \\ \underset{\sim}{\beta}_5' \end{bmatrix} = \begin{bmatrix} .357 & .359 & .357 & .356 & -.007 \\ .437 & -.437 & .437 & -.437 & .000 \\ .781 & .784 & -.781 & -.784 & .003 \\ .877 & -.873 & -.876 & .874 & -.016 \\ -.014 & .014 & .009 & -.019 & -1.000 \end{bmatrix} .$$

The Simulation Results

Eighty samples of size n = 50 were generated from a multivariate normal $N(\underset{\sim}{0},\Sigma)$. For each sample the ridge estimates of canonical correlations and variates were calculated for $k_1,k_2 = 0, .1, .2, .3, .4, .5, .75$, and 1.0.

The square error of the estimate $\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2)$ from $\underset{\sim}{\alpha}_i$ $i = 1, 2, \cdots, 5$

$$(4.1) \quad \Delta^2_{yi}(k_1,k_2) = (\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2)-\underset{\sim}{\alpha}_i)'(\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2)-\underset{\sim}{\alpha}_i) = \sum(\tilde{\alpha}_{ij}(k_1,k_2)-\alpha_{ij})^2$$

was obtained for each sample*. Similarly, we define $\Delta^2_{xi}(k_1,k_2)$ for the

*Since canonical variates are determined only up to sign, the square error was calculated also for $-\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2)$ and the minimum value taken.

$\underset{\sim}{\beta}_i$. In some applications only the structure of the canonical variates in terms of relative magnitudes of the coefficients within a variate is of interest. In such cases we might consider the estimates as determined only up to a constant multiple. An appropriate square error for such applications is

$$(4.2) \qquad (c\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2) - \underset{\sim}{\alpha}_i)'(c\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2) - \underset{\sim}{\alpha}_i)$$

where c is chosen to minimize the value. It is easily seen that

$$(4.3) \qquad c = \tilde{\underset{\sim}{\alpha}}'(k_1,k_2)\underset{\sim}{\alpha}_i \ / \ \tilde{\underset{\sim}{\alpha}}_i'(k_1,k_2)\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2)$$

$$= \tilde{\underset{\sim}{\alpha}}'(k_1,k_2)\underset{\sim}{\alpha}_i \ / \ \|\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2)\|^2 \ .$$

Then (4.2) reduces to

$$(4.4) \qquad \underset{\sim}{\alpha}_i'\underset{\sim}{\alpha}_i[1 - (\tilde{\underset{\sim}{\alpha}}_i'(k_1,k_2)\underset{\sim}{\alpha}_i)^2 \ / \ \|\tilde{\underset{\sim}{\alpha}}_i(k_1,k_2)\|^2\|\underset{\sim}{\alpha}_i\|^2]$$

$$= \|\underset{\sim}{\alpha}_i\|^2[1 - R^2_{yi}(k_1,k_2)] \ ,$$

hence this minimum square error is expressed as a fraction of the square of the length of the true population canonical variate. Similarly, $\|\underset{\sim}{\beta}_i\|^2[1 - R_{xi}(k_1,k_2)]$ for the second set of variates.

In order that square error terms be expressed in a similar scale for all variates we shall express them in terms of a multiple of the square of the length of the true canonical variate, that is

$$\Delta^2_{yi}(k_1,k_2) = \|\underset{\sim}{\alpha}_i\|^2[\Delta^2_{yi}(k_1,k_2) \ / \ \|\underset{\sim}{\alpha}_i\|^2] \ .$$

In Table A.1 we present the average values of $[\Delta^2_{yi}(k_1,k_2) \ / \ \|\underset{\sim}{\alpha}_i\|^2]$ and $[1 - R_{yi}(k_1,k_2)]$, averaged over the eighty samples. The table is arranged in five sections corresponding to the five pairs of canonical variates. The table values for a given pair of canonical variates $(\underline{\alpha}_i, \underline{\beta}_i)$ and (k_1, k_2) are arranged as

$$\begin{array}{ll} \Delta^2_{yi}(k_1,k_2) \;/\; \|\underline{\alpha}_i\|^2 & \qquad \Delta^2_{xi}(k_1,k_2) \;/\; \|\underline{\beta}_i\|^2 \\ \\ 1 - R^2_{yi}(k_1,k_2) & \qquad 1 - R^2_{xi}(k_1,k_2) \end{array} .$$

All table values have three decimal places, the decimals being omitted for space consideration. To summarize the effect of the ridge estimation we extract from Table A.1 only the $k_1 = k_2 = k$ values of $\Delta^2_{yi}(k,k) \;/\; \|\underline{\alpha}_i\|^2$ and $\Delta^2_{xi}(k,k) \;/\; \|\underline{\beta}_i\|^2$. This is not to suggest that $k_1 = k_2$ is a good choice, rather that the smaller table permits a quicker comparison of values.

TABLE 4.1

Average Square Error of Canonical Variates, n = 50

		0.0	.10	.20	.30	.40	.50	.75	1.0
	Variate								
	First	1.372	.282	.203	.169	.147	.127	.111	.111
	Second	2.191	.356	.242	.198	.175	.160	.156	.165
Set 1	Third	3.420	.644	.462	.385	.350	.330	.314	.315
	Fourth	6.070	1.270	.742	.564	.494	.461	.437	.437
	Fifth	.296	.397	.479	.536	.579	.611	.668	.705
	First	.449	.262	.195	.159	.135	.116	.102	.102
	Second	.799	.476	.345	.275	.233	.205	.182	.180
Set 2	Third	.799	.381	.370	.366	.373	.385	.416	.445
	Fourth	.516	.505	.531	.546	.560	.573	.604	.630
	Fifth	1.881	1.571	1.363	1.233	1.113	1.061	.948	.886

The effect of the ridge estimation procedure on mean square error is dramatically evidenced in Tables A.1 and 4.1. The usual estimate $(k_1,k_2) = (0,0)$ gives rise to an average square error of $\|\underline{\alpha}_1\|^2$ (1.372).

Thus we are led to conclude that the mean square length of the difference between the estimated and the true first canonical variate is greater in magnitude than the length of the vector itself. The instability of the estimate is clearly established by this value and interpretations from such an estimate could be totally misleading. Even small values of k_1 produce substantial reductions in the average square error, for example

$$\Delta_1^2(.1,.1) = \|\underline{\alpha}_1\|^2(.282) \quad \text{and} \quad \Delta_1^2(.3,.3) = \|\underline{\alpha}_1\|^2(.169) .$$

It is also observed that for the first canonical variate, the average square error continues to decrease as k_1 and k_2 increase, the smallest value attained being $\|\underline{\alpha}_1\|^2(.111)$.

The effect on second, third, and fourth canonical variates from the first set is even more pronounced with (0,0) values being respectively, 2.191, 3.420, and 6.074. For these variates we observe decreases and then increases with (k_1,k_2). Only the last variate with a small canonical correlation starts low and increases.

The initial values of square error for variates in the second set are not nearly so large. Substantial decreases in average square error are noted, however; and the smallest values attained are remarkably close to the corresponding values for variates in the first set.

The very nature of the ridge estimation procedure forces smaller values for estimates of the canonical correlations since the usual procedure is to obtain variates with "maximum" correlation with respect to the sample correlation matrix. In Table 4.2 we present a summary of this effect by giving the average estimated canonical correlation over eighty samples for the $k_1 = k_2$ values. Recall that the true population canonical correlations are .8, .6, .4, .2, and .11, respectively.

TABLE 4.2

Averages of Estimated Canonical Correlations

	0.0	.10	.20	.30	.40	.50	.75	1.0
First	.843	.796	.758	.725	.696	.669	.610	.561
Second	.682	.611	.562	.522	.488	.458	.398	.353
Third	.493	.404	.351	.312	.281	.256	.209	.178
Fourth	.285	.196	.160	.137	.120	.108	.085	.071
Fifth	.095	.056	.041	.033	.028	.024	.018	.015

When $(k_1,k_2) = (0,0)$ we observe a rather definite bias upward in the estimates of canonical correlations. This is to be expected since the method of estimation is a maximization and tends to pick up relationships peculiar to the sample. The estimates were most nearly unbiased at $k_1 = k_2 = .10$, and for larger values the estimates show a definite bias downward. It should be pointed out, however, that real interest for the experimenter is the correlation of his estimated canonical variates with respect to the population correlation matrix and not the sample correlation matrix. In practice this cannot be computed. In this simulation, however, we observed that the true correlations of the mean estimated canonical variates for all k_1,k_2 pairs over the eighty samples were essentially the same (within 0.02). On individual samples, ridge estimates had higher correlations with the population correlation matrix than did the usual estimate. In Table A.2 we present these true correlations for n = 50 and values $k_1 = k_2 = 0.0, .1, .2, .3, .4, .5, .75, 1.0$ for one random sample generated in addition to those in the simulation.

The reader will note canonical correlations moving toward population values and toward zero correlation between canonical variates not in the same pair.

The simulation was repeated for sample sizes of $n = 25$ and $n = 100$. At $n = 25$ the effect of the ridge analysis was more pronounced in the relative sense, but even the ridge estimates produced considerable instability. We would conclude that while ridge analysis improves the stability the resultant estimates are still likely to be unsatisfactory. In Table 4.3 we present the mean square error values for $k_1 = k_2$, as in $n = 100$, Table 4.1. We note that the initial values are smaller than with samples of size $n = 50$. The effect of the ridge estimates is still toward a smaller mean square error with minimum values smaller than with $n = 50$, and these minimum values were attained for smaller values of k_1 and k_2. Initial biases on the estimates of canonical correlations were slightly smaller at $k_1 = 0$, $k_2 = 0$ the means over 80 samples being

$$.82, .64, .46, .25, .09 \quad .$$

A downward bias in the estimate of canonical correlation was observed even when $k_1 = .1$ and $k_2 = .1$. The indication is that with increased sample size the usual estimates are not as unstable, but the effect of the ridge estimates is toward greater stability and in fact give greater stability than for smaller sample sizes.

TABLE 4.3

Average Square Error of Canonical Variates, n = 100

		0.0	.10	.20	.30	.40	.50	.75	1.0
	Variate								
	First	.659	.124	.090	.078	.074	.072	.074	.081
	Second	.707	.175	.130	.113	.107	.106	.115	.132
Set 1	Third	2.173	.348	.259	.229	.217	.213	.218	.232
	Fourth	4.686	.870	.459	.361	.324	.310	.312	.328
	Fifth	.196	.335	.431	.500	.548	.584	.645	.685
	First	.159	.102	.078	.066	.060	.057	.059	.059
	Second	.429	.244	.175	.142	.125	.117	.113	.112
Set 2	Third	.261	.198	.198	.209	.226	.243	.287	.285
	Fourth	.450	.434	.434	.434	.474	.493	.534	.530
	Fifth	1.575	1.251	1.092	.944	.858	.813	.730	.731

References

Hoerl, A. E. and Kennard, R. W. Ridge regression: biased estimation for nonorthogonal problems. Technometrics, 1970a, 12, 55-66.

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

Hotelling, Harold. The most predictable criterion. Journal of Educational Psychology, 1935, 26, 139-142.

Hotelling, Harold. Relations between two sets of variates. Biometrika, 1936, 28, 321-377.

Thorndike, R. M. and Weiss, D. J. A study of the stability of canonical correlations and canonical components. Educational and Psychological Measurement, 1973, 33, 123-134.

TABLE A.1

Average Square Error of Ridge Estimates, n = 50

k_1 \ k_2	0.0		.10		.20		.30		.40		.50		.75		1.0	
	First Canonical Variates $(\underline{\alpha}_1, \underline{\beta}_1)$															
0.0	1372	449	1281	291	1245	222	1228	186	1222	166	1220	154	1221	141	1224	139
	491	291	472	224	462	186	455	163	453	148	453	138	453	122	453	114
.10	313	393	282	262	267	206	258	175	252	157	248	145	242	133	240	133
	238	266	220	206	210	173	203	153	199	140	196	130	191	115	189	107
.20	234	363	214	247	203	195	197	166	192	149	189	139	185	129	183	129
	194	252	179	196	171	166	166	147	162	134	159	125	155	111	152	103
.30	198	343	182	235	174	186	169	159	166	143	164	133	161	124	160	126
	171	243	159	189	153	160	148	142	145	130	142	121	139	107	136	100
.40	176	329	162	225	155	177	150	150	147	135	145	126	143	118	146	122
	157	236	146	183	140	154	136	136	133	124	131	116	128	103	127	097
.50	159	317	147	215	139	168	134	142	130	126	127	116	120	106	116	107
	146	229	136	177	130	148	125	130	122	117	119	108	113	093	109	084
.75	137	295	128	201	123	157	119	133	116	118	115	110	111	102	111	104
	126	218	118	167	113	140	109	122	107	111	105	102	102	089	100	082
1.0	131	284	124	194	120	152	118	129	116	115	114	107	112	100	111	102
	115	212	109	163	105	136	102	119	100	108	099	100	096	087	095	080

TABLE A.1 (cont.)

k_1 \ k_2	Second Canonical Variates $(\underline{\alpha}_2, \underline{\beta}_2)$															
	0.0		.10		.20		.30		.40		.50		.75		1.0	
0.0	2191	799	2102	515	2047	383	2011	310	1989	270	1973	248	1949	224	1936	220
	587	455	561	362	545	306	534	269	526	245	522	228	515	203	510	189
.10	411	754	356	476	330	356	313	289	303	251	296	229	288	208	284	205
	329	423	289	332	271	283	258	250	249	226	244	210	237	185	233	171
.20	315	726	265	461	242	345	227	281	217	244	211	223	204	203	200	201
	275	409	237	323	219	275	207	243	199	220	194	204	187	180	183	167
.30	278	708	232	449	212	337	198	275	189	239	184	218	177	199	175	198
	251	399	214	316	197	269	186	238	178	216	173	200	166	176	162	164
.40	258	694	215	438	196	329	183	268	175	233	170	212	164	193	163	195
	235	392	199	309	182	263	171	232	164	211	159	195	153	172	151	161
.50	247	683	206	429	187	321	175	261	166	226	160	205	151	184	147	183
	223	385	187	302	171	256	160	225	151	204	145	188	136	163	132	149
.75	239	667	200	415	185	311	175	255	168	221	163	201	156	182	152	181
	205	374	170	292	155	249	145	220	138	199	133	184	125	160	121	147
1.0	246	660	208	408	195	308	186	252	180	219	175	200	169	180	165	180
	196	368	161	288	148	246	139	217	132	197	128	182	119	159	115	146

TABLE A.1 (cont.)

k_1 \ k_2	0.0	.10	.20	.30	.40	.50	.75	1.0
	Third Canonical Variates ($\underline{\alpha}_3,\underline{\beta}_3$)							
0.0	3420 799	3565 450	3673 444	3768 448	3838 454	3874 469	4078 504	4174 539
	648 455	647 389	645 390	640 393	636 397	636 406	636 423	637 443
.10	681 436	644 381	647 378	640 380	638 389	639 402	646 436	654 467
	434 360	405 343	399 344	392 344	387 347	385 351	384 362	385 371
.20	497 425	462 372	462 370	448 370	443 380	441 393	441 427	443 458
	381 352	351 335	347 337	338 336	333 338	330 341	328 351	329 359
.30	431 418	398 366	395 361	385 366	380 376	377 389	375 423	376 454
	356 347	326 330	321 331	313 331	308 334	305 337	303 346	302 354
.40	396 413	365 361	361 356	355 362	350 373	347 387	344 421	344 452
	341 343	311 327	306 327	298 328	294 331	291 334	288 343	287 351
.50	376 408	345 357	342 353	337 360	333 371	330 385	327 419	327 450
	330 340	300 323	296 324	289 326	285 329	281 332	278 341	277 349
.75	356 401	326 349	325 348	321 356	318 367	316 381	314 416	313 447
	314 333	283 316	280 319	274 322	270 325	267 328	263 337	261 345
1.0	356 398	326 344	325 344	322 352	320 365	318 379	316 414	315 445
	303 329	273 312	270 316	265 319	261 322	258 326	254 335	252 342

TABLE A.1 (cont.)

k_1 \ k_2	0.0		.10		.20		.30		.40		.50		.75		1.0	
	Fourth Canonical Variates (α_4, β_4)															
0.0	6074	516	6042	545	6028	566	6011	577	5988	593	5998	608	5963	643	5973	818
	817	436	817	479	820	507	818	523	816	538	816	551	814	580	677	609
.10	1283	482	1270	505	1278	534	1268	547	1265	563	1264	579	1261	611	1259	637
	546	419	541	453	547	478	544	492	544	505	545	516	547	536	550	550
.20	725	485	729	505	742	531	740	551	738	571	738	588	738	621	739	645
	447	417	451	450	458	472	457	489	457	503	457	515	458	536	460	550
.30	552	479	549	498	559	523	564	546	569	567	571	585	573	619	575	643
	399	414	398	445	405	467	406	486	407	500	408	512	411	534	413	548
.40	477	477	474	494	482	517	488	540	494	560	499	578	501	613	503	639
	369	412	366	442	374	464	377	482	379	497	380	509	383	531	385	545
.50	441	474	437	489	445	513	450	535	456	556	461	573	466	609	467	636
	349	411	346	440	353	462	357	480	360	494	362	506	365	529	368	543
.75	409	465	408	482	414	508	418	531	421	550	425	567	437	604	437	632
	318	406	315	435	324	459	327	476	330	490	334	500	339	524	342	540
1.0	404	451	408	472	414	501	417	524	420	545	423	562	432	598	437	630
	298	397	298	430	307	456	310	474	313	487	316	498	324	518	327	538

TABLE A.1 (cont.)

k_1 \ k_2	Fifth Canonical Variates $(\underline{\alpha}_5, \underline{\beta}_5)$															
	0.0		.10		.20		.30		.40		.50		.75		1.0	
0.0	296	1881	299	1614	301	1440	304	1322	304	1234	304	1166	307	1060	313	1000
	246	713	248	723	250	731	251	737	252	742	251	746	252	757	260	764
.10	399	1850	397	1571	397	1391	398	1263	398	1175	398	1103	399	993	398	926
	191	692	189	698	189	702	189	704	189	706	188	708	188	711	188	713
.20	477	1802	478	1556	479	1363	480	1241	480	1153	481	1086	482	974	482	910
	162	683	164	689	165	694	167	697	168	699	168	701	169	703	170	705
.30	534	1784	534	1521	535	1349	536	1233	537	1141	538	1077	540	966	540	902
	152	679	152	684	155	689	158	692	160	695	161	697	163	699	163	701
.40	575	1792	575	1513	576	1339	577	1218	579	1133	580	1068	582	961	581	898
	146	677	147	682	149	685	152	688	155	691	157	694	159	697	160	698
.50	607	1787	607	1510	608	1331	609	1213	610	1125	611	1061	613	958	613	895
	143	676	144	680	145	683	148	685	151	688	153	691	157	695	158	697
.75	662	1778	662	1500	663	1322	663	1200	664	1114	665	1049	668	948	669	889
	136	674	138	677	140	679	141	681	143	682	146	684	153	691	154	694
1.0	697	1765	699	1492	699	1329	700	1194	700	1107	701	1043	703	939	705	886
	129	670	133	675	135	677	137	679	138	680	140	680	147	685	151	692

TABLE A.2

True Correlations of Estimated Canonical Variates

$k_1 = 0.0 \quad k_2 = 0.0$

.679	-.248	-.107	-.047	-.100
-.177	.643	-.096	.050	.143
-.046	-.141	.403	.187	-.110
.032	.151	-.065	.233	-.062
.042	.151	-.021	.112	.093

$k_1 = .10 \quad k_2 = .10$

.743	-.191	-.086	-.023	-.128
-.141	.603	-.094	.024	.119
-.018	-.152	.388	.205	-.114
.010	.162	-.037	.241	-.073
.005	.162	-.000	.091	.092

$k_1 = .20 \quad k_2 = .20$

.765	-.149	-.078	-.020	-.135
-.109	.591	-.101	.013	.106
-.002	-.153	.381	.208	-.117
-.009	.172	-.023	.242	-.078
-.013	.163	.008	.086	.092

$k_1 = .30 \quad k_2 = .30$

.775	-.123	-.072	-.019	-.138
-.088	.588	-.104	.006	.098
.008	-.151	.377	.209	-.119
-.022	.177	-.013	.242	-.081
-.024	.164	.014	.084	.092

$k_1 = .40 \quad k_2 = .40$

.781	-.105	-.069	-.018	-.140
-.073	.586	-.106	.000	.092
.016	-.149	.374	.210	-.120
-.032	.180	-.006	.242	-.083
-.031	.163	.019	.083	.091

$k_1 = .50 \quad k_2 = .50$

.785	-.092	-.066	-.018	-.140
-.062	.585	-.107	-.004	.088
.021	-.146	.372	.210	-.122
-.040	.182	.000	.242	-.085
-.037	.163	.022	.082	.091

$k_1 = .75 \quad k_2 = .75$

.790	-.070	-.061	-.017	-.141
-.042	.584	-.108	-.011	.081
.030	-.141	.369	.211	-.125
-.053	.184	.010	.241	-.089
-.046	.161	.027	.081	.090

$k_1 = 1.0 \quad k_2 = 1.0$

.792	-.055	-.058	-.016	-.141
-.030	.583	-.109	-.015	.077
.036	-.137	.367	.212	-.127
-.062	.184	.017	.240	-.091
-.052	.160	.031	.080	.090