MULTIPLE LINEAR REGRESSION VIEWPOINTS
VOLUME 11, NUMBER 3 SUMMER 1982

ON USING THE AVERAGE INTERCORRELATION AMONG PREDICTOR VARIABLES AND EIGENVALUE ORIENTATION TO CHOOSE A REGRESSION SOLUTION

Beverly Mugrage
The University of Akron

Jesse Marquette
The University of Akron

Isadore Newman
The University of Akron

Background

Since its introduction in 1970 by Arthur Hoerl, the efficacy of ridge regression has been vigorously debated by statisticians. Notable are the debates in the Journal of the American Statistical Association, JASA, in 1980 (Smith and Campbell) and in Technometrics in 1979 (Draper and Van Nostrand). Much research among proponents of ridge regression concentrated on comparisons of various ridge regression solutions. Dempster, Schatzoff, and Wermuth (1977) compared 57 varietios of ridge regression; Galarneau-Gibbons (1981) compared ten of the most promising ridge algorithma. Both were simulation studies.

Presented at AERA 1982, MLR Special Interest Group
Not refereed by editorial staff

Since the introduction of the Monte Carlo method in 1949 by von Neumann and Ulam, simulation studies have been frequently used in statistics to solve problems otherwise difficult or expensive to solve. Monte Carlo simulation can be adapted to any situation for which a model representing reality can be designed and for which a mechanism to simulate this model can be effected.

Analysis of the recent literature of ridge regression reveals essential agreement that ridge regression is an appropriate alternative to least squares regression when predictor variables are highly intercorrelated. Another theme is common. Many researchers from Newhouse and Oman in 1971 to Galarneau-Gibbons in 1981 also suggest that the orientation of the beta vector with respect to the eigenvectors corresponding to the largest and the smallest eigenvalue of the $X'X$ matrix determines the relative performance of ordinary least squares estimators and ridge estimators.

Purpose of this Study

The question of the predictive values of the orientation of beta and/or the average absolute intercorrelation among independent variables in guiding an investigator's choice of regression method is interesting and important. The availability of a computer simulation capable of producing data with given R^2 and average absolute intercorrelation made study of this question possible. The simulation was designed for the 1979 comparison of shrinkage formuli by Newman, McNeil, Garver, and Seymour.

Methods

Twelve populations of 1,000 cases were generated representing four different values of intercorrelation among predictor variables (0.80, 0.50, 0.30, 0.15) and three different values of R^2 (0.50, 0.30, 0.05). From each population 220 samples were drawn with replacement. There were 50 cases per sample.

For each sample generated, Marquette and Du Fala's statistical package ADEPT (1979) was used to calculate the ordinary least squares solution, the principal components solution and three ridge solutions. The ridge solutions chosen were the Lawless-Wang solution, the McDonald-Galarneau solution, and a Hoerl-Kennard-Baldwin solution. The Hoerl-Kennard-Baldwin solution is important historically and because of its good performance in previous studies. The Lawless-Wang solution is a Bayesian solution derived from the assumptions $Y \sim N(X\beta, \sigma^2 I)$ and $\beta \sim N(0, \sigma_\beta^2 I)$ with the ridge parameter $k = \sigma^2/\sigma_\beta^2$ estimated by $k = ps^2/\sum \lambda_i \gamma_i^2$. The McDonald-Galarneau solution is an iterative solution which estimates the true length of the beta vector by $Q = \hat{\beta}'\hat{\beta} - s^2 \sum \lambda_i^{-1}$ and then picks k to minimize $|\hat{\beta}'(k)\hat{\beta}(k) - Q|$. This procedure defaults to ordinary least squares if Q is negative. These three methods of determining k were different enough in derivation to be interesting to compare.

The study was a 3 x 4 x 5 factorial design. There were three values for R^2, four for average absolute intercorrelation and five regression methods.

The various regression solutions were ranked on four criteria:

1. Average variance of regression coefficients.
2. Error in regression coefficients as measured by $(\beta-\hat{\beta})'(\beta-\hat{\beta})$.

3. Mean square error.

4. Shrinkage of R^2 upon cross-validation.

For each sample, solutions were ranked from one to five with smaller rank indicating more desireable solution. Ranks were then summed for each solution on all criteria to give an overall measure of quality of solution.

The orientation of the coefficient vector, beta, with respect to the eigenvector associated with the largest eigenvalue of the $X'X$ matrix was calculated for each sample. For some populations the range of values for the orientation was small enough to cause computational difficulty in the computer packages used in this study. For this reason, the orientation of beta was categorized and interaction between regression method and the orientation of beta was determined using two-way analysis of variance. The decision to categorize the orientation of beta is discussed further in the results section.

Results

Since this study was exploratory, a significance level of $\alpha = .05$ was used. When multiple comparisons were made, the correction suggested by Newman and Fry, $\alpha = .05/n$, was applied (Newman and Fry, 1972). All tests were two-tailed.

Error in Beta

For all populations with high average absolute intercorrelation, $|\bar{r}| = .80$, the error in beta as measured by $(\beta-\hat{\beta})'(\beta-\hat{\beta})$ was significantly different for ordinary least squares regression and each of the ridge solutions tested. For high multicollinearity, the error in beta for each ridge solution was significantly different from that of every other ridge solution with only one exception: Lawless-Wang error in coefficients was not significantly different from that of Hoerl-Kennard-Baldwin for the population with R^2=.50 and $\overline{|r|}$ =.80. For each of the populations with high multicollinearity, Lawless-Wang regression produced the smallest error in coefficients while ordinary least squares and principal components regression accounting for 100 percent of the trace produced the largest error in coefficients.

For moderate multicollinearity (0.50 and 0 .30), there was always a significant difference between the error in beta for ordinary least squares and each ridge solution's error in beta. The error for the complete principal components solution also was significantly different from that of each of the ridge solutions. Error in beta did not differ significantly for OLS and complete principal components solutions.

For low multicollinearity ($\overline{|r|}$ =.15), ordinary least squares regression and complete principal components regression produced significantly different error of beta from each other as well as from each ridge solution.

For graphic representation of these results, see Figure 1.

Variance of Betas

For each population, for any given method, the coefficients of each independent variable formed a distribution. Thus if beta 1 is the coefficient of the first independent variable, a distribution for the ordinary least squares beta would exist, as well as one for the Lawless-Wang beta 1, the Hoerl-Kennard-Baldwin beta 1, and the McDonald-

FIGURE 1

Error in Regression Coefficients as a Function of Solution Type, R^2, and $\overline{|r|}$

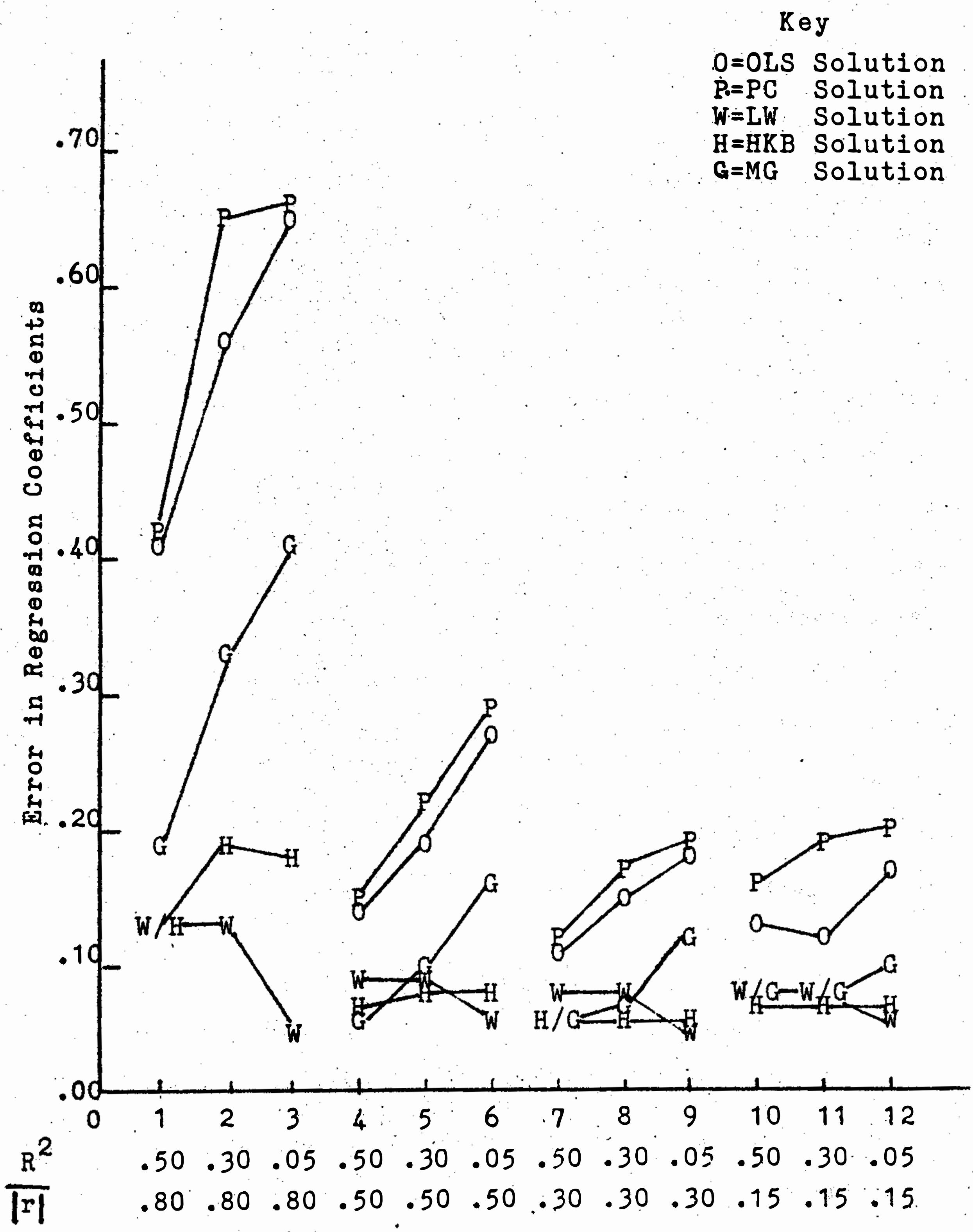


TABLE 1

Summary of Results of Cochran's Test for Variance of Betas

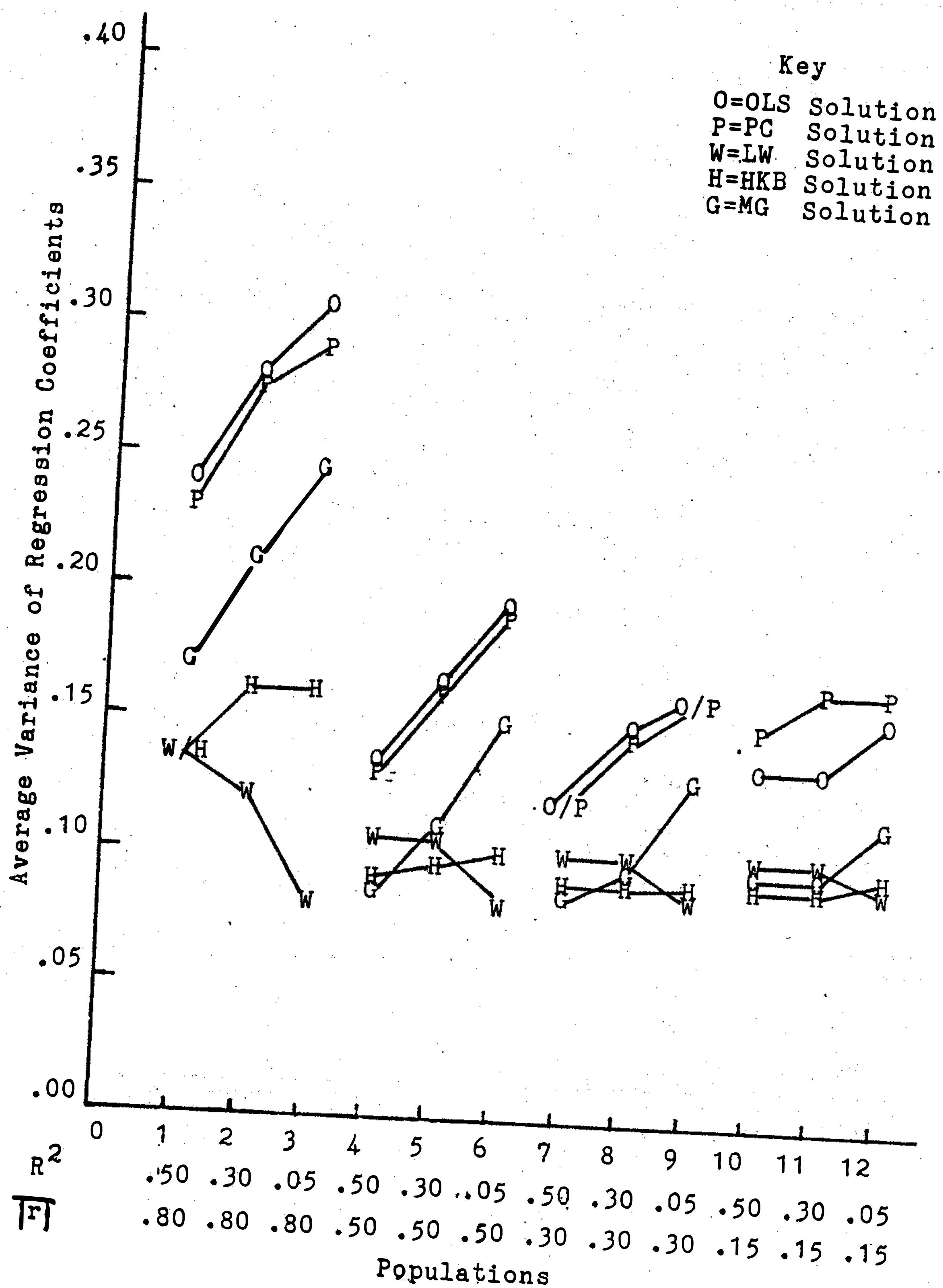
Population Parameters $R^2/\|r\|$	Cochran's G for						
	β_1	β_2	β_3	β_4	β_5	β_6	β_7
.50/.80	.3480	.2661	.2979	.2785	.2827	.2825	.2695
.30/.80	.3537	.2881	.2509	.3113	.2950	.3081	.2671
.05/.80	.3300	.2850	.3134	.3038	.2982	.2842	.3021
.50/.50	.3052	.2512	.2686	.2969	.2581	.2491	.2741
.30/.50	.3221	.2587	.2689	.2981	.2767	.2749	.2726
.05/.50	.3322	.2759	.3006	.3049	.3196	.2753	.2841
.50/.30	.2785	.2485	.2718	.2638	.2512	.2415	.2621
.30/.30	.2870	.2534	.2805	.2710	.2535	.2706	.2697
.05/.30	.2983	.3099	.2903	.2862	.2644	.2666	.2656
.50/.15	.2506	.2831	.2873	.2738	.2571	.2728	.2473
.30/.15	.3308	.2753	.2904	.2639	.2771	.2921	.2377
.05/.15	.2748	.2806	.2746	.2619	.2639	.2801	.2845

All tests significant

Critical Region: $G > G_{.05} = .2360$

FIGURE 2

Average Variance of Regression Coefficients as a Function of Solution Type, R^2, and $\overline{|r|}$



Galarneau beta 1. Variances of these distributions were compared using Cochran's test, normality having been verified with a chi square test and sample size being equal. The results appear in Table 1. Cochran's test for each estimated beta for every population showed that the four variances compared were not all equal. To examine the relationship among the variances more closely, multiple comparisons $\alpha = .05/n$ was used for .05 significance. This is the correction suggested by Newman and Fry (1972).

For high multicollinearity (0.80) the variance of the ordinary least squares beta was significantly different from that of Lawless-Wang or Hoerl-Kennard-Baldwin beta for each independent variable. The ordinary least squares beta variance was higher than that of any ridge beta variance for each of the betas for the seven independent variables.

For all population (R^2 = 0.50, 0.30, 0.05) with high multicollinearity the Lawless-Wang estimator was always significantly different from that of the McDonald-Galarneau estimator and for R^2 = 0.05, it was significantly different from both of the other two ridge estimators. See Figure 2 for graphic representation of this information.

Shrinkage Upon Cross-Validation

The shrinkage in R^2 upon cross-validation was not significantly different among the various regression solutions for eight of the twelve populations including the population with R^2 = 0.50 and high average absolute intercorrelation (0.80). For the other two populations (R^2 = 0.30, and R^2 = 0.05) with high multicollinearity there was a significant difference in shrinkage of R^2 upon cross-validation

FIGURE 3

Shrinkage in R^2 upon Cross-Validation as a Function of Solution Type, R^2, and $\overline{|r|}$

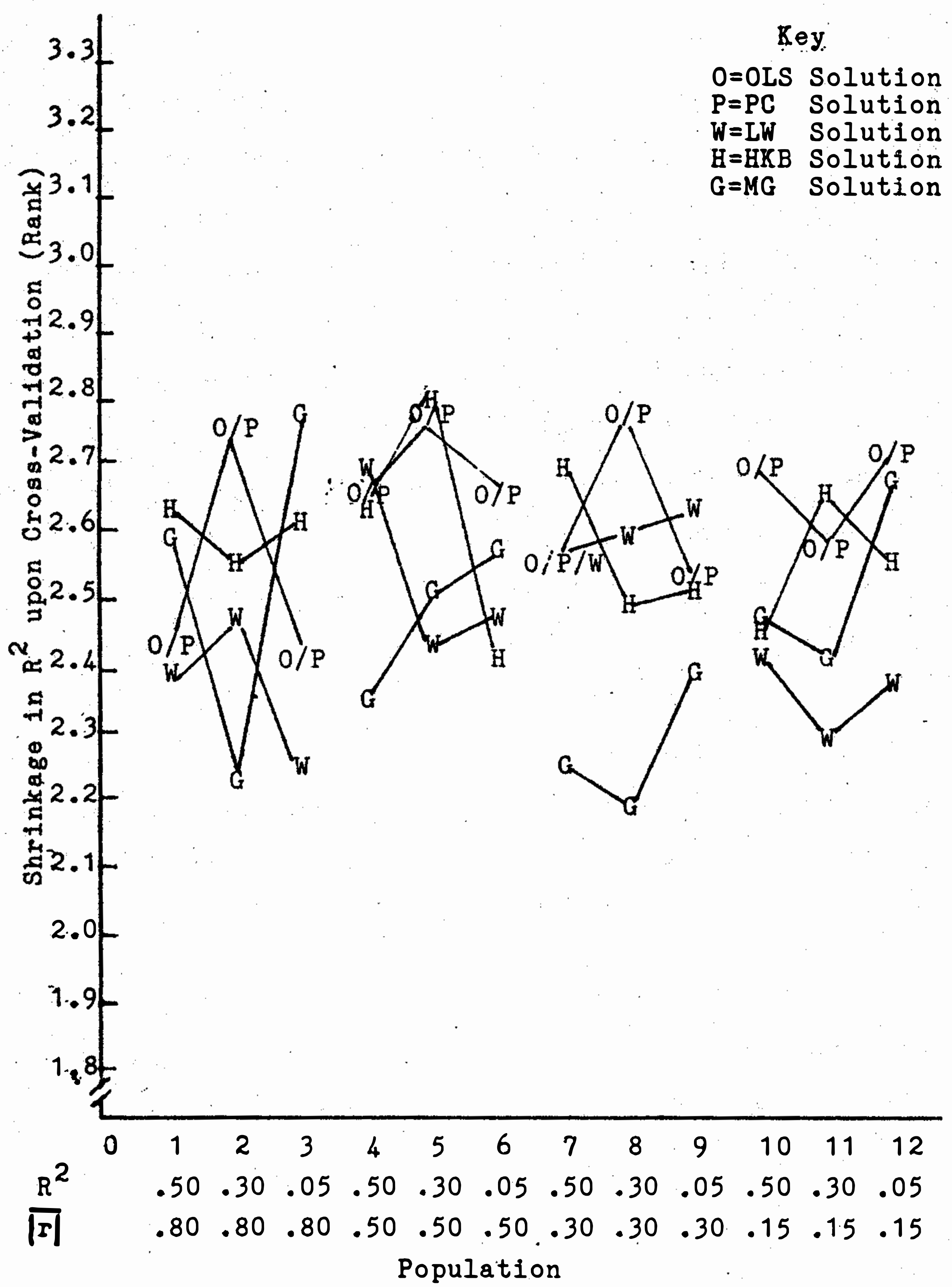
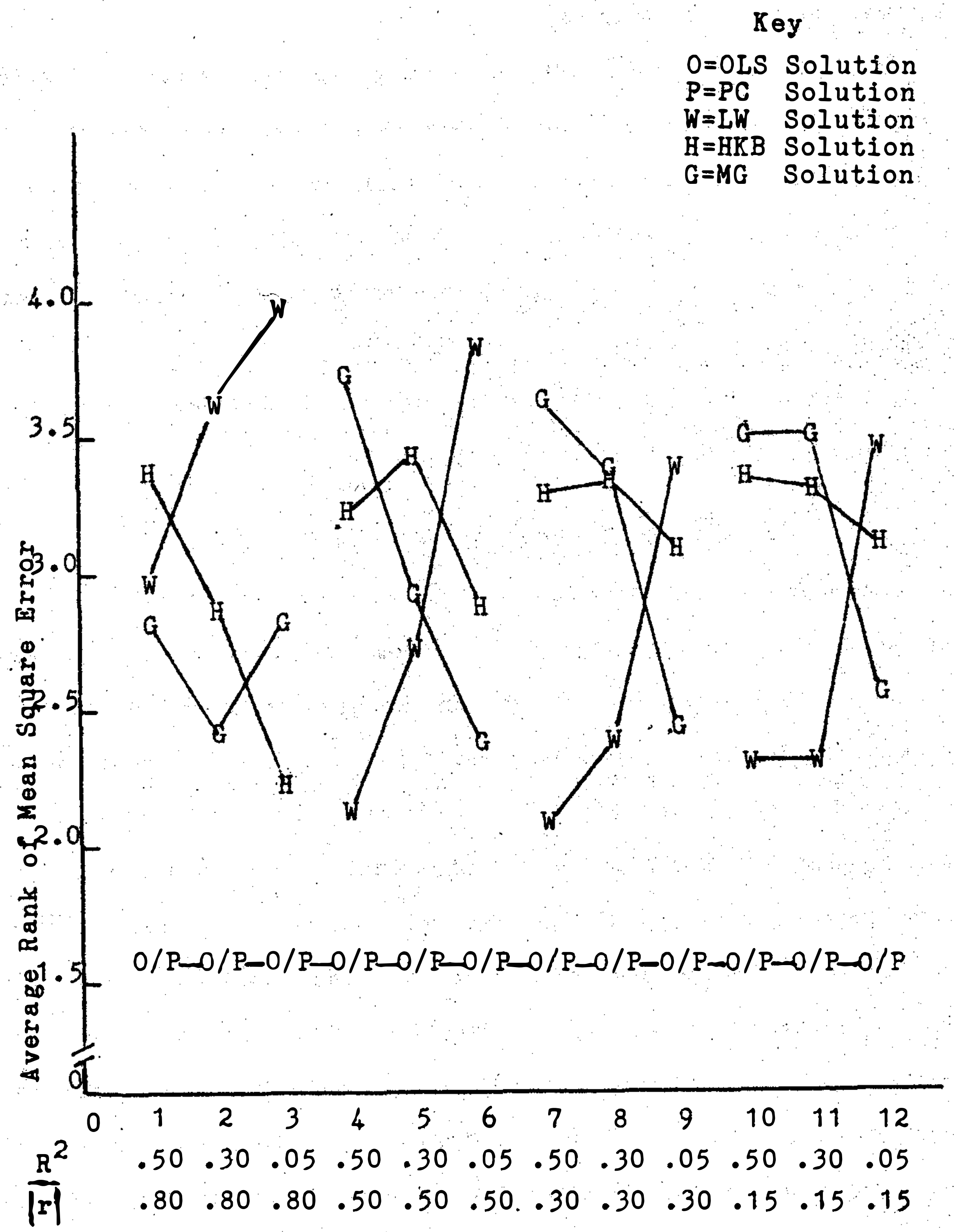


FIGURE 4

Average Rank of Mean Square Error as a Function of Solution Type, R^2, and $\overline{|r|}$



between ordinary least squares and at least some of the ridge solutions. For $R^2 = 0.30$, the ordinary least squares R^2 shrunk more than the ridge solutions and for $R^2 = 0.05$, the ordinary least squares R^2 shrunk less than the other estimators.

There is no evidence in the results of this study indicating that the ridge regression R^2 shrunk less than the ordinary least squares R^2 for populations with high multicollinearity, a situation in which ridge regression is commonly used. The actual value of the shrunken R^2 may be more useful than the value of the shrinkage of R^2 upon cross-validation.

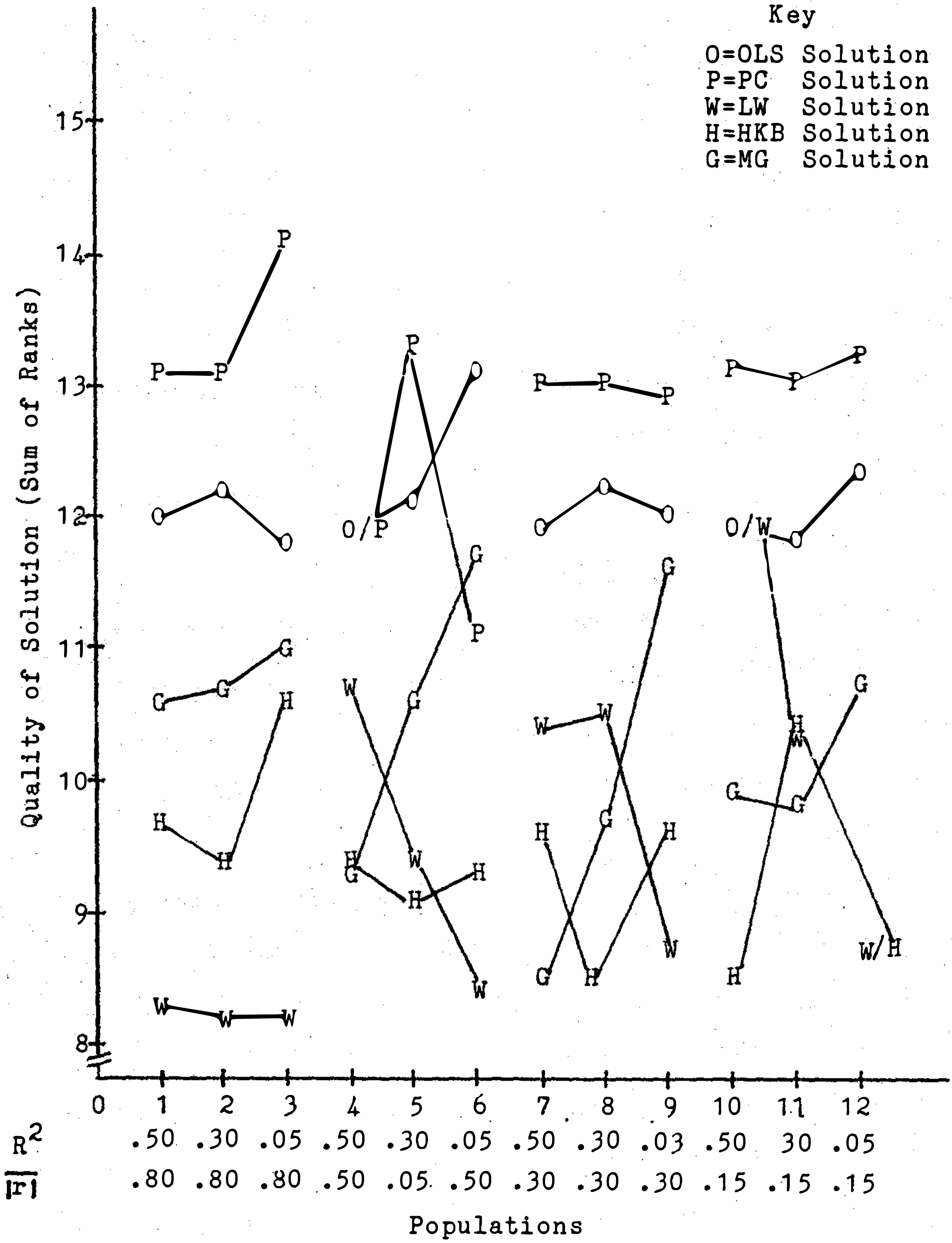
See Figure 3 for graphic representation of shrinkage for varying R^2 and average absolute intercorrelation.

Means of R^2 Before and After Cross-Validation

Knowledge of the shrinkage in R^2 upon cross-validation may be less valuable than knowledge of the final value of R^2 upon cross-validation. The value of the shrunken R^2 gives a lower bound on R^2. Shrinkage in R^2 is of less interest. For this reason, means of R^2 before and after cross-validation were calculated for the ordinary least squares solution and the ridge solutions for each population. Before cross-validation, R^2 for the ordinary least squares solution was greatest. The values of R^2 for the ridge solutions were only slightly smaller. After cross-validation values for R^2 among the solutions were again close in value. McDonald-Galarneau ridge regression produced the largest R^2 after cross-validation for six of the twelve populations. Ordinary least squares regression and Lawless-

FIGURE 5

Quality of Solution as a Function of Solution Type, R^2, and $\overline{|r|}$



Wang regression produced the highest R^2 for three populations each.

MSE

As expected, the mean square error for ordinary least squares regression was significantly different from that of the ridge solutions for all populations. Generally, the MSE for the ridge solutions were significantly different from each other. For only three populations (R^2=0.50, $\overline{|r|}$ =0.80; R^2=0.30, $\overline{|r|}$ =0.50; R^2=0.30, $\overline{|r|}$ =0.15), was there no significant difference among ridge MSE. Of the ridge solution, the Lawless-Wang solutions had the lowest MSE for six of the twelve populations, McDonald-Galarneau for five, and the Hoerl-Kennard-Baldwin solution for only one of the twelve populations. Graphic representation of MSE for various values of R^2 and r is seen in Figure 4.

Overall Solution Quality

If overall quality is measured by the sum of ranks, analyses of variance indicated a significant F-ratio with a probability of 0.00000 for all populations. Representation of overall quality of solution as a function of R^2 and average absolute intercorrelation occurs in Figure 5.

A good solution was operationally defined as one whose sum of ranks was less than the mean sum of ranks. The number of good solutions for each method for each population are given in Table 2. For average absolute intercorrelation of 0.80 and R^2 = 0.50, 214 of 200 Lawless-Wang solutions were considered good compared with 157 of 220 Hoerl-Kennard-Baldwin solutions and 108 McDonald-Galarneau solutions. For all other highly multicollinear populations, results were similar.

TABLE 2

NUMBER OF GOOD SOLUTIONS

Population 1: R^2=.50, $|r|$ =.80

Number of Samples: 214

Type of Solution	Number of Good Solutions
Ordinary Least Squares (OLS)	27
Principal Components (PC)	13
Lawless and Wang Ridge (LW)	214
Hoerl, Kennard and Baldwin Ridge (HKB)	157
McDonald and Galarneau (MG)	108

Population 2: R^2:.30, $|r|$ =.80

Number of Samples: 212

Type of Solution	Number of Good Solutions
OLS	23
PC	9
LW	211
HKB	163
MG	91

Population 3: R^2=.05, $|r|$ =.80

Number of Samples: 216

Type of Solution	Number of Good Solutions
OLS	22
PC	10
LW	216
HKB	152
MG	135

Population 4: R^2=.50, $|r|$ =.50

Number of Samples: 215

Type of Solution	Number of Good Solutions
OLS	33
PC	4
LW	81
HKB	177
MG	184

NUMBER OF GOOD SOLUTIONS

Population 5: R^2=.30, $|r|$ =.50

Number of Samples: 215

Type of Solution	Number of Good Solutions
OLS	33
PC	3
LW	165
HKB	214
MG	100

Population 6: R^2= .05, $|r|$ =.50

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	0
PC	69
LW	217
HKB	169
MG	43

Population 7: R^2=.50, $|r|$ =.30

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	38
PC	7
LW	109
HKB	163
MG	204

Population 8: R^2=.30, $|r|$ =.30

Number of Samples: 217

Type of Solution	Number of Good Solutions
OLS	27
PC	5
LW	113
HKB	208
MG	157

TABLE 2

NUMBER OF GOOD SOLUTIONS

Population 9: R^2=.05, $|r|$ =.30

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	39
PC	8
LW	217
HKB	166
MG	99

Population 10: R^2=.50, $|r|$ =.15

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	40
PC	4
LW	115
HKB	216
MG	153

Population 11: R^2=.30, $|r|$ =.15

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	45
PC	7
LW	121
HKB	208
MG	65

Population 12: R^2=.05, $|r|$ =.15

Number of Samples: 219

Type of Solution	Number of Good Solutions
OLS	27
PC	8
LW	174
HKB	217
MG	101

with Lawless-Wang regression producing the largest number of good solutions. For these same populations, in every case, principal components accounting for 100 percent of the trace produced the fewest good solutions followed by ordinary least squares regression.

One must be cautious in interpreting overall quality of solution done as a sum of ranks. In summing ranks, equal weighting is imposed on the criteria for good solution: variance of beta error in beta, shrinkage upon cross-validation, and MSE. This stacks the deck against the OLS solution and the principal components solution accounting for 100 percent of the trace. Theory tells us that ridge should outperform OLS on two of the four criteria used.

Orientation of the Beta Vector

To test for interaction of the orientation of the beta vector and method of regression solution, the orientation of beta was categorized and two-way analyses of variance were run. Categorization of the orientation became necessary because the small range the orientation exhibited in some populations presented serious computational difficulties using the ADEPT model comparison and DPLINEAR. For highly multicollinear data the interaction between the orientation of beta and method was nonsignificant. Significant interaction occurred for $R^2=0.05$, $\overline{|r|}=0.30$, and $R^2=0.50$, and $\overline{|r|}=0.15$ only. For these levels of intercorrelation, ridge regression would rarely be considered the method of choice. Orientation of the beta vector appears of little usefulness in choosing among ridge regression methods for highly multicollinear data.

Conclusions

The results of this study indicate that for high degrees of multicollinearity, when stability and interpretability of coefficients is important, ridge regression is an attractive alternative to least squares regression. Low error and small variance of coefficients make ridge regression a useful device for anyone wanting to interpret beta weights for any reason, a device that should prove useful to social science investigators attempting to look at "causation" through correlation as in path analysis. Lawless-Wang ridge regression performed especially well on criteria for stability of coefficients in this study.

The major advantage to ridge regression is <u>not</u> in prediction nor in hypothesis testing but in applications for which the sign or interpretability of coefficients is important.

Principal components using all components was equivalent to the OLS solution in production of R^2, $\hat{Y}$, and MSE. It was not equivalent in variance or error of regression coefficients. For the principal component solution variance of coefficients increased rapidly as components associated with lower eigenvalues were added. Evidence from this experiment supports the use of a cut-off in using principal components regression (Rummel,1970). More work needs to be done concerning appropriate placement of such a cut-off.

Values for R^2 before cross-validation and values for R^2 after cross-validation were close for ordinary least squares and the ridge solutions tested in this study. The value of R^2 after cross-validation seems a more appropriate way of comparing solutions than shrinkage in R^2 upon cross-validation.

The orientation of the eigenvector associated with the largest eigenvalue of the $X'X$ matrix with respect to the population beta vector does not appear to be useful in choosing among ordinary least squares regression, principal components regression accounting for 100 percent of the trace, Lawless-Wang ridge regression, Hoerl-Kennard-Baldwin ridge regression, or McDonald-Galarneau ridge regression.

It is clear from this study that the quality of a solution as determined by error in coefficients, variance of coefficients, MSE or R^2 after cross-validation depends upon the characteristics of the population. There is a strong dependence upon the degree of multicollinearity. Within a given multicollinearity, there is a dependence upon the R^2 of the population.

Ridge regression has a distinct advantage over OLS when stability and interpretibility of coefficients is important but not for purposes of prediction or hypothesis testing.

REFERENCES

Draper, Norman and Van Nostrand, R. Craig Ridge regression and James Stein estimaton: review and comments. Technometrics, 1979, 21, 451-465.

Gibbons, Diane Galarneau A simulation study of some ridge estimators. Journal of the American Statistical Association, 1981, 76, 131-9.

Marquette, Jesse and Du Fala, Margaret ADEPT: Advanced Date Enquiry Package Time-Shared, 1979.

Newhouse, Joseph and Oman, Samuel An Evaluation of Ridge Estimators, Santa Monica, CA: Rand, 1971, (R-716 PR).

Newman, Isadore, and Fry J. A response to "A Note on Multiple Comparisons" and a comment on shrinkage. Multiple Linear Regression Viewpoints, 1972, 2, 36-38.

Newman, Isadore, McNeil, Keith A., Garver, Thomas & Seymour, Gayle A. A Monte Carlo evaluation of estimated parameters of five shrinkage formuli. American Educational Research Association, 1979, 1-18.

Rummel, Rudolph J. Applied Factor Analysis, Evanston, Illinois: Northwestern University Press, 1970.

Smith, Gary and Campbell, Frank A critique of some ridge regression methods. Journal of the American Statistical Association, 1980, 75, 74-81.