8010135

DREACHSLIN, JANICE L.

THE EFFECT OF SAMPLE SIZE ON RIDGE REGRESSION AS A DATA ANALYTIC TECHNIQUE FOR USE WITH ILL CONDITIONED INPUT MATRICES

Wayne State University PH.D. 1979

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106 18 Bedford Row, London WC1R 4EJ, England

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✓.

1. Glossy photographs _____
2. Colored illustrations _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Print shows through as there is text on both sides of page _____
6. Indistinct, broken or small print on several pages _____ throughout _____
7. Tightly bound copy with print lost in spine _____
8. Computer printout pages with indistinct print ✓
9. Page(s) _____ lacking when material received, and not available from school or author _____
10. Page(s) _____ seem to be missing in numbering only as text follows _____
11. Poor carbon copy _____
12. Not original copy, several pages with blurred type _____
13. Appendix pages are poor copy _____
14. Original copy with light type _____
15. Curling and wrinkled pages _____
16. Other _____

THE EFFECT OF SAMPLE SIZE ON RIDGE REGRESSION AS A DATA ANALYTIC TECHNIQUE FOR USE WITH ILL CONDITIONED INPUT MATRICES

by

Janice L. Dreachslin

A DISSERTATION

Submitted to the Office for Graduate Studies,

Graduate Division of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

1979

MAJOR: EDUCATION (EDUCATIONAL EVALUATION AND RESEARCH)

APPROVED BY:

Donald L. Marcotte Oct. 22, 1979

Adviser Date

Alfred Rogers

Joseph Z. Park Jr.

Thomas J. Duggan

ACKNOWLEDGEMENTS

Members of my doctoral committee, especially my major adviser, Donald Marcotte, deserve thanks for their advice and assistance. John Kolvereid of General Motors Corporation, who introduced me to ridge regression, reviewed and assessed my progress, and provided guidance, made the completion of this dissertation possible. Diane Gibbons of General Motors Corporation and others who assisted in the development of computer software are also appreciated.

Family and friends, through providing moral support, helped to create an atmosphere conducive to progress. In particular, my father's appreciation of the value of education encouraged me to pursue my graduate studies. Special thanks are reserved for a dear friend who coaxed and cajoled me into the successful completion of this dissertation.

ABSTRACT

When prediction is the objective of educational research, a goal of data analysis is to obtain a regression equation which retains its predictive power upon application to a second sample from the same population. When predictor variables exhibit high positive intercorrelations, obtained regression coefficients may fluctuate widely from sample to sample (Kerlinger & Pedhazur, 1973). In addition, variance accounted for (R^2) is overestimated, the variability of the p sample beta weights is inflated (Claudy, 1972), and the estimated squared length of the coefficient vector ($\beta'\beta$) exceeds the population value of $\beta'\beta$ (Hoerl & Kennard, 1970a). Least squares analysis produces regression equations with the above described problems when data are multicollinear (Hoerl & Kennard, 1970a; Claudy, 1972; Kerlinger & Pedhazur, 1973). Claudy (1972) contends that the problems which result from the application of ordinary least squares analysis to multicollinear data are increased as sample size decreases.

Hoerl (1962) contends that ridge regression should be used in lieu of ordinary least squares analysis to reduce and/or eliminate the above described problems when data are multicollinear. Marquardt and Snee (1975) state that the sizes of the maximum variance inflation factor and

the minimum eigenvalue are the best indicators of multicollinear predictors. The higher the maximum variance inflation factor, the more multicollinear the data. The magnitude of the maximum variance inflation factor and the minimum eigenvalue are inversely related.

Ridge regression requires the addition of a constant $0<k\leq 1$ to the diagonal of the input correlation matrix of predictor variables ($X'X$) prior to inversion. Regression equations which are calculated through the use of $(X'X+k)^{-1}$ as opposed to $X'X^{-1}$ behave more like equations which are calculated given orthogonal predictors, reduce the maximum variance inflation factor and increase the minimum eigenvalue (Hoerl & Kennard, 1970a, 1970b).

Ordinary least squares analysis and ridge regression are compared across samples from the same population with respect to stability of estimated beta coefficients, magnitude of the maximum variance inflation factor, value of the minimum eigenvalue, magnitude of R^2 before and after cross-validation, standard deviation of each set of estimated beta coefficients, the estimated squared length of the coefficient vector, and the subjective evaluation of how reasonable the signs and absolute values of the estimated beta coefficients are. The descriptive comparisons outlined are conducted with samples from the same population which have the following number of subjects per predictor variable: 5, 10, 20, 30 and 40. The following ridge k selection procedures are employed: Inspection of the ridge trace

which graphically displays ridge estimated coefficients mapped against k, and the point estimation procedures proposed by Hoerl (1962), McDonald and Galarneau (1975) and Hoerl and Kennard (1976).

The data set consists of raw scores on the following subtests of the Comprehensive Test of Basic Skills (CTBS), Level 1, Form S for 391 fourth grade students: Mathematics Concepts, Reading Vocabulary, Reading Comprehension, Language Spelling, Language Mechanics, and Language Expression. The Mathematics Concepts subtest score is the criterion, while the five reading and language subtest scores are the predictors. Data are multicollinear.

Results and conclusions include the following:

1. Ridge regression is effective in increasing the minimum eigenvalue and reducing the maximum variance inflation factor from the levels obtained with k=.000, i.e., ordinary least squares analysis.

2. The ridge results behave more like results obtained given orthogonal predictors than the ordinary least squares analysis results in that the standard deviation of the five sample beta weights and the estimated squared length of the coefficient vector are both reduced.

3. Inflation of the standard deviation of the five sample beta weights and estimated $\beta'\beta$ is greater at N=25 and N=50 than for the larger sample sizes so that larger k must be employed to obtain comparable results.

4. Sample R^2 at k=.000 is inflated with respect to population R^2 and the inflation increases with decreasing sample size.

5. In general, shrinkage upon cross-validation is less for the ridge selected k's than for least squares analysis.

6. Coefficients change sign from a theoretically invalid negative to positive as k>0 for N=25 and N=50.

7. Neither least squares analysis nor ridge regression produce consistently more stable regression coefficients.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Statement of the Problem

Two types of research problems in education commonly utilize the multiple regression model as a data analytic technique. The first type of research problem has prediction as the objective of research. Ordinary least squares analysis is applied to sample data in order to statistically combine scores on two or more predictor variables and predict a subject's score on a criterion variable. The second type of research problem has the identification of an optimal subset of predictors from a larger set of predictors as the objective of research (Halinski & Feldt, 1970).

When prediction is the objective of educational research, a goal of data analysis is to obtain beta coefficients for the prediction equation which are stable upon application to a second sample from the same population. Kerlinger and Pedhazur (1973, p. 77), state that, "when a variable is added to a regression equation, all the regression coefficients change. In addition, regression coefficients may change from sample to sample as a result of sampling fluctuation, especially when the independent variables are highly correlated. All this means, of course, that substantive interpretation of regression coefficients is difficult and dangerous, and it

becomes more difficult and dangerous as predictors are more highly correlated with each other." Pedhazur (1975) states that predictor variables almost always exhibit positive correlations in studies of educational effects.

Correlated predictor variables are common in undesigned experiments in which the researcher cannot manipulate the independent variables or control extraneous variables directly or by randomization. Undesigned experiments are a predominant mode of research in education (Pedhazur, 1970).

Claudy (1972) observes that applications of ordinary least squares analysis in undesigned experiments violate a key assumption of the statistical technique. In classic ordinary least squares analysis theory the predictor variables are treated as fixed or controlled by the experimenter through an experimental design (Halinski & Feldt, 1970). When levels of the predictor variables are not fixed, intercorrelations among predictors result. Applications of ordinary least squares analysis to data sets in which levels of the predictor variables are not fixed or controlled by the experimenter, i.e., when predictors are correlated, result in two types of errors:

1. The population R^2 is over-estimated by the sample R^2.

2. The variability of the sample beta weights is inflated as compared with the variability of the population beta weights. (Claudy, 1972).

Claudy, (1972, p. 314) states that much research has been done in an attempt to overcome the problem of over-estimation of R^2 but, "little attention has been paid to the inflation of the variability of the n sample beta weights as compared with the actual variability of the n population beta weights, and in fact this is a seldom mentioned finding." Pedhazur

(1975, p. 261) concludes, "when multicollinearity is relatively large, slight changes in the pattern of the intercorrelations, which may be due to measurement or specification or sampling errors, may result in substantial changes in the magnitudes of the regression coefficients." Claudy (1972) identifies a need for research directed toward the development of regression techniques which reduce the variability of sample beta coefficients when predictor variables are correlated.

Hoerl (1962) contends that ridge regression, a biased estimation technique, should be used in lieu of ordinary least squares analysis to estimate population beta coefficients for use in the regression equation when the predictor variables are correlated. Given an ill-conditioned input correlation matrix (predictor variables (X'X), the total variance of estimated beta obtained through ordinary least squares analysis may be too large to be of practical utility in prediction with subsequent samples. Ridge regression results in estimated beta coefficients with smaller mean square error than the estimated beta coefficients obtained through ordinary least squares analysis (Mayer & Willke, 1973).

Population beta (β) coefficients are usually estimated through Gauss-Markov-linear functions of the criterion variable (Y) which are unbiased and have minimum variance (Hoerl & Kennard, 1970a). The estimated beta ($\hat{\beta}$) coefficients derived from the application of ordinary least squares analysis are obtained from the equation which minimizes the residual sum of squares, i.e., the sum of the squared differences between observed and estimated scores on the criterion (Simon, 1975).

Ordinary least squares which estimates population beta (β) coefficients through Gauss-Markov-linear functions of the criterion variable is a good procedure if the sum of squares and crossproducts matrix to be factored is nearly a unit matrix when in the form of a correlation matrix. If the input matrix which is to be factored is not nearly a unit matrix, "the least squares estimates often do not make sense when put into the context of the physics, chemistry, and engineering of the process which is generating the data" (Hoerl & Kennard, 1970a, p. 55).

When predictor variables are uncorrelated, the input correlation matrix (X'X) is a unit or identity matrix. An input correlation matrix with non-zero off diagonal elements, i.e., with correlated predictor variables, "is said to be ill conditioned and the original experimental design is said to be non-orthogonal" (Simon, 1975, p. 22).

Undesigned experiments in which some experimental variables are not controlled by the experimenter result in input correlation matrices which are ill conditioned. The predictor variables in an undesigned experiment are correlated mathematically, complicating interpretation of the estimated beta ($\hat{\beta}$) coefficients derived from an application of ordinary least squares analysis to the input correlation matrix (Simon, 1975).

Marquette and Dufala (1978) state that multicollinearity of the predictor variables results in ill conditioning of the input correlation matrix. Application of ordinary least squares analysis to an ill conditioned input matrix results in estimated beta ($\hat{\beta}$) coefficients with large standard errors which reduce the utility of the estimated beta ($\hat{\beta}$) coefficients

for prediction purposes. Marquardt and Snee (1975) conclude that ordinary least squares achieves good fit to the estimation data when the input matrix is ill conditioned, but often destroys good prediction of new data.

Marquardt and Snee (1975) identify two indicators of an ill conditioned input correlation matrix. The two indicators of an ill conditioned input correlation matrix are the sizes of the maximum variance inflation factor (VIF) and the minimum eigenvalue (λ_{min}) of the X'X matrix. Marquardt and Snee (1975, p. 4) state that, "The maximum variance inflation factor is the best single measure of the conditioning of the data. For any predictor orthogonal to all other predictors, the inflation factor is 1.0". The variance inflation factor for each predictor is an indicator of the influence of the zero order correlations of the predictor with each of the other predictors in the X'X matrix on the variance of the coefficient of the predictor. The variance inflation factors are the elements on the diagonal of the inverse $(X'X)^{-1}$ of the input correlation matrix.

Marquardt and Snee (1975, p. 5) report that, "a seriously non-orthogonal (or 'ill conditioned') problem is characterized by the fact that the smallest eigenvalue (λ_{min}), is very much smaller than unity." The variance inflation factor is an indicator of how close the smallest eigenvalue of the X'X matrix is to zero. The magnitude of the VIF and the value of λ_{min} reflect the magnitude of the correlations among predictors in the input correlation matrix. The magnitude of the VIF and the value of λ_{min} are inversely related.

Hoerl and Kennard (1975a) report that lower bounds for the average squared distance from the estimated beta coefficients ($\hat{\beta}$) to the population beta (β) coefficients is σ^2/λ_{min}. With one or more small eigenvalues, the distance from $\hat{\beta}$ to β tends to be too large for the $\hat{\beta}$ calculated through ordinary least squares to be of practical utility for prediction purposes. Simon (1975, pp. 46-47) concludes that, "When the predictor variables are uncorrelated, the eigenvalues, λ_i, are each equal to one. In that case, the average squared distance between estimated and true beta coefficients will be equal to the error variance of the data multiplied by the number of variables, p, involved. However, when the predictors are correlated, as in the case of the undesigned experiment, some of the eigenvalues become very small and their reciprocals very large. This increases the average squared distance between the estimated and true beta coefficients." Hoerl and Kennard (1970a) provide the following formula for the expected squared distance between $\hat{\beta}$ and β:

$$E(L^2) = (\hat{\beta}-\beta)' \ (\hat{\beta}-\beta) = \sigma^2 \sum_{i=1}^{p} (1/\lambda_i)$$

where:

$E(L^2)$ = average squared distance between estimated and true beta coefficients
$\hat{\beta}$ = vector of estimated beta coefficients
β = vector of true beta coefficients
σ^2 = error variance of the input data
ρ = number of predictor variables
λ_i = eigenvalue of the ith predictor variable

Marquardt and Snee (1975) observe that, with an ill conditioned input matrix, the ordinary least squares solution yields $\hat{\beta}$ coefficients which are too large and with signs which may reverse with small changes in the data. The probability of $\hat{\beta}$ coefficients which are too large in absolute value and

which may have the wrong sign increases as the deviation of the predictor vectors from orthogonality increases. "In multiple linear regression, the effect of nonorthogonality of the prediction vectors is to pull the least squares estimates of the regression coefficients away from the true coefficients that one is trying to estimate" (Hoerl & Kennard, 1970b, p. 82).

Marquardt and Snee (1975) identify two classic methodologies for dealing with ill conditioned data. The first classic methodology is least squares which fits estimation data well but may result in unstable coefficients and concomitant shrinkage when applied to a new sample. The second classic methodology is variable selection, which, "implies a simplistic two-valued classification logic wherein any predictor variable may be either important or unimportant. Large prediction biases can result from elimination of 'non-significant' predictors." (Marquardt & Snee, 1975, p. 4).

Hoerl and Kennard (1970a) identify the following procedures traditionally used to untangle ill conditioned input matrices: stepwise regression, principal components, and computation of all 2^p regressions, where p is the number of predictor variables. Hoerl and Kennard (1970a, p. 58) conclude that, "with the occasional exception of principal components, these methods don't really give an insight into the structure of the factor space and the sensitivity of the results to the particular set of data at hand."

Hoerl (1962) suggests ridge regression as an alternative to ordinary least squares analysis in order to diminish the

inflation and instability associated with ordinary least squares estimates of population beta (β) coefficients when the input matrix is ill conditioned. The estimated ridge coefficient ($\hat{\beta}^*$) is a biased estimate of population beta since the expected value of $\hat{\beta}^*$ is not population beta (β). The formula for ridge estimation of population beta coefficients from an input correlation matrix is as follows (Hoerl & Kennard 1970a):

$$\beta^* = [X'X + kI]^{-1} X'Y; \; 0 \leq k \leq 1$$

where:

β^*	=	pxl vector of ridge estimates of the population beta coefficients when p is the number of predictor variables
$(X'X+kI)^{-1}$	=	the inverse of the pxp input correlation matrix of predictor variables with the constant k; $0 \leq k \leq 1$ added to the diagonal
$X'Y$	=	the pxl vector of correlations between predictor variables and the criterion variable

Ridge regression adds the constant k to each diagonal element of the input correlation matrix prior to computing the ridge estimate of population beta in order to reduce the reciprocals of the eigenvalues of X'X and produce more stable estimates of population betas as compared with the ordinary least squares estimates of population betas. Mayer and Willke (1973) state that given $k>0$, there is at least one k which results in estimates of population beta with smaller mean square error than the total variance of the least squares estimator of population beta.

Ridge regression uses the achievement of small mean square error of estimated beta as the criterion for selec-

tion of $\hat{\beta}^*$. Marquardt and Snee (1975) conclude that utilizing a biased estimator to achieve the criterion of small mean square error is acceptable if the introduction of bias in estimated beta results in a major reduction of the variance of estimated beta as compared with the unbiased ordinary least squares estimate of population beta. The sum of squares of residuals is an increasing function of k, whereas the length of $\hat{\beta}^*$ is a decreasing function of k (Marquardt & Snee, 1975).

Simon (1975, p. 49) explains that, "adding a constant k to the correlation matrix diagonal has the effect of adding k to the eigenvalues of the variance component. For the very small eigenvalues, the addition of even a small k can do much to decrease the size of the reciprocals of the eigenvalues and to decrease the squared distance between estimated and true beta coefficients."

Hoerl and Kennard (1970a, 1970b) contend that, given an ill conditioned input matrix, ridge regression will result in a better prediction equation than ordinary least squares analysis as evidenced by the following:

1. The estimated coefficients are closer to the true coefficients on the average

2. The signs of the estimated coefficients are more meaningful

3. Point estimates can be made with smaller mean square error

4. The ridge coefficients are more stable upon application to new data

Hoerl and Kennard (1970a) provide the following formula for the expected value of the squared distance between the ridge estimate of beta and population beta:

$$E\left[L^2_i\ (k)\right] = \sigma^2 \sum_i^p \lambda_i \left|(\lambda_i + k)^2 + k^2\ \beta'\ (X'X + kI)^{-2}\beta\right.$$

where:

$E\left[L^2(k)\right]$	=	average squared distance between ridge estimates and true beta coefficients
σ^2	=	error variance of the input data
λ_i	=	eigenvalue of the ith predictor variable
k	=	constant
$X'X$	=	input correlation matrix of predictor variables
β	=	vector of true (population) beta coefficients
I	=	the identity matrix

The first component of the equation represents the variance, whereas the second accounts for the bias due to the addition of the constant k to the diagonal of the input correlation matrix. When k=0, the second component is also equal to zero and the resulting $E\left[L^2_i\ (0)\right]$ is the expected squared distance between the ordinary least squares estimate of beta and population beta.

Hoerl and Kennard (1970a) state that the more ill conditioned the input correlation matrix, the further the ridge estimate of population beta can move from the unbiased estimate of population beta without an appreciable increase in the residual sum of squares. Thus, the application of ridge analysis to a severely ill conditioned input matrix will not result in an appreciable decrease in the amount of criterion variance accounted for from the amount of criterion variance accounted for through ordinary least squares analysis. Hoerl and Kennard (1970a) recommend altering the least squares estimate of population beta through the selection of a value

of the constant k which will shorten the length of the regression vector while minimizing the increase of the residual sum of squares. As k increases, the variance error decreases more rapidly than the bias increases, therefore a value of k can be selected for which the mean square error of the ridge estimated betas will be smaller than for estimated betas obtained through ordinary least squares analysis.

Marquardt and Snee (1975) describe the boundedness assumption implicit in the use of ridge estimation. Lease squares implies the assumption of an unbounded uniform prior distribution on the coefficient vector. Since, when choosing a biased estimator through ridge estimation, the X'X matrix is in correlation form, the population value of any regression coefficient is seldom greater than three. The regression coefficient vector is, thus, finite. Ridge estimation places boundedness requirements on the coefficient vector. The existence of a $k>0$ such that the mean square error of the ridge coefficient is less than the mean square error of the ordinary least squares coefficient is dependent upon the boundedness assumption (Marquardt, 1970).

Hoerl and Kennard (1970a) recommend utilization of the ridge trace which is a two dimensional plot of the estimated values of population beta ($\hat{\beta}^*_{(k)}$) and $\phi^*_{(k)}$, the residual (error) sum of squares, for selected values of the constant k in the interval $[0,1]$ to aid in the selection of an optimum value of k for the development of the regression equation to be used in prediction. The ridge trace follows a path through the sum of squares surface so that for a fixed value of the

residual sum of squares (ϕ), a value for estimated beta is chosen which has minimum length. In practice, the researcher chooses values of $k \geq 0$ and then computes the residual sum of squares for each selected value of the constant k. The selection of a specific k value from observation of the ridge trace is a subjective decision.

Hoerl and Kennard (1970a) suggest the following criteria for selection of the constant k from the values of k for which $\hat{\beta}^*$ and $\phi^*_{(k)}$ are displayed on the ridge trace:

1. At the selected value of k absolute values of the estimated beta coefficients are reasonable given knowledge of the factors for which the coefficients represent rates of change

2. At the selected value of k the system stabilizes and the coefficient vectors change little beyond the selected value of k

3. At the selected value of k estimated beta coefficients with signs that appeared unreasonable at k=0, given knowledge of the factors, change to the expected sign

4. At the selected value of k the residual sum of squares is not large relative to the minimum residual sum of squares or relative to a reasonable residual sum of squaes, given knowledge of the system under study

Methods for point estimation of k are also available (Lindley & Smith, 1972; Mallows, 1973; Farebrother, 1975; Kennard & Baldwin, 1975; McDonald & Galarneau, 1975; Obenchain, 1975; Hoerl & Kennard, 1976; Lawless & Wang,

1976; Golub, Heath & Wahba, 1977; Dempster, Schatzoff, & Wermuth, 1977). Consensus is not achieved as to the best method for the selection of k (Hoerl, 1979).

Marquardt and Snee (1975, p. 6) state that, "the ridge estimate gives the smallest regression coefficients consistent with a given degree of increase in the residual sum of squares." The ridge trace graphically displays the $\hat{\beta}^*$ so that the researcher can observe which coefficients are sensitive to the data. The ridge trace has one curve or trace per estimated beta coefficient. The value of k where the estimated beta coefficients stabilize results in a set of coefficients which are not sensitive to small changes in the estimation data (Marquardt & Snee, 1975). Marquardt and Snee (1975) contend that the increase in the residual sum of squares with increasing values of k is not relevant if the object of the regression procedure is to develop stable coefficients for use in future prediction rather than to obtain the closest fit possible to the estimation data. Given orthogonal predictors, coefficients stabilize at k=0 so that the ordinary least squares estimates of population betas may be used.

Hoerl and Kennard (1970b) use a ten predictor example to illustrate the application of ridge analysis to a non-orthogonal problem previously analyzed by Gorman and Toman in illustrating a procedure for selecting a subset of the predictor variables to use in a prediction equation. Fifteen regressions are computed varying k in the interval $|0,1|$.

The eigenvalues of X'X reflect significant interfactor correlations with the ridge trace showing instability and overestimation of population beta with k=0. Since the estimated regression coefficients stabilize between k=.20 and k=.30, k=.25 is selected. Hoerl and Kennard (1970b) report that utilization of Gorman and Toman's method of selecting the best subset of predictor variables does not eliminate the tendency to overestimate population betas and to produce unstable coefficients when using an ordinary least squares procedure. Through examination of the ridge trace, Hoerl and Kennard (1970b) select different factors to eliminate than are selected through the Gorman and Toman variable subset selection procedure. The factors selected for elimination through examination of the ridge trace are driven toward zero as k>0.

Utilizing corn yield data previously analyzed by Laird and Cady, Marquardt and Snee (1975) demonstrate that ridge regression results in a stable prediction model and better illustrates the role of all the variables in the model than the full, stepwise, and PRESS procedures applied to the same data. Maximum variance inflation factors of 180, 122, and 12 are found for the full, stepwise, and PRESS models respectively. Ridge analysis results in a model which keeps all predictor variables in the equation and reduces the correlations among predictor variables. Ridge results suggest a nonlinear model which is consistent with the physical background of the corn yield problem and illustrates the roles of all the predictor variables in accounting for criterion variance.

Barcikowski and Stevens (1976) recommend the use of ridge regression to produce more interpretable canonical variates. Ordinary least squares analysis, ridge regression, and directed ridge regression, a procedure which involves adding the constant k only to those elements in X'X with relatively small eigenvalues, are applied to investment equations previously analyzed by Christ (Guilkey & Murphy, 1975). Results show little change in R^2 among the three procedures but smaller standard errors of the estimated betas are obtained through the ridge methods.

Walton, Newman, and Fraas (1978) apply ridge regression and least squares analysis in the prediction of counselor practicum ratings, utilizing a sample of 93 counselor education students. Predictor variables include undergraduate grade point average, final graduate grade point average, Miller Analogies test score, and type of undergraduate institution. Simple, squared, and interaction predictor variables are used. Subjects are randomly assigned to prediction and estimation groups. Results produce a shrunken ordinary least squares R^2 which is more similar to the true population estimate of R^2 than the shrunken or non-shrunken ridge R^2. Ridge regression coefficients are less sensitive to the particular variable scores in the estimation data which results in a smaller R^2 but produces greater stability upon application of the regression equation to a new sample.

Walton, Newman, and Fraas (1978, p. 12) conclude that, "If the research project requires stable coefficients as would be the case in making a point prediction over different

samples, ridge regression may be the appropriate analytical tool. However, if the purpose of the research project is to test a hypothesis, the use of multiple linear regression would be more appropriate." Observing the stability of ordinary least squares estimates of population beta after cross-validation or correction for shrinkage is recommended. If relatively stable results are obtained, ridge analysis might be unnecessary (Walton, Newman, & Fraas, 1978).

Marquardt and Snee (1975) identify two types of data sets to which ridge regression is applicable. The first type of data set has correlated predictor variables due to the collection of historical data without an experimental design. The second type of data set has correlated predictor variables due to physical and mathematical constraints such as missing values, correlated errors, split plotting, and nonconstant variance which occurred within an experimental design.

Simon (1975, p. 15) states that, "In the behavioral sciences, unlike the physical sciences, performance cannot be examined or evaluated independently of the context in which it occurs and can only be described or predicted as a function of this context. The more generalizable data therefore will be derived from experiments in which critical context factors are varied rather than held constant." With the exception of Walton, Newman, and Fraas' (1978) application of ridge regression in the prediction of counselor practicum ratings, no applications of ridge regression to prediction problems within education were located.

Claudy (1972, p. 314) reports that, "Preliminary empirical studies indicated that the variability of sample beta weights" obtained through ordinary least squares analysis "is inflated at all sample sizes, but that the degree of inflation decreases with increasing sample size." A review of the ridge regression literature in which ridge analysis is employed with real world data and sample N's are reported reveal N's which range from 29 to 228 with a mean subject to predictor variable ratio of 6.46 to 1.00 (Hoerl & Kennard, 1970b; Marquardt & Snee, 1975; Price, 1977). The largest subjects to predictor variable ratio reported in the ridge regression literature is 13.84 to 1.00 (Hoerl & Kennard, 1970b). No applications of ridge regression to samples of varying N from the same population were located.

Overview of the Study

Ordinary least squares and ridge regression data analytic procedures are applied to the problem of prediction of Comprehensive Test of Basic Skills (CTBS) Mathematics Concepts subtest score from the CTBS Reading Vocabulary, Reading Comprehension, Language Spelling, Language Mechanics, and Language Expression subtest scores. The total number of subjects is 391. Data are provided by the Educational Testing Service on a sample of fourth grade students to whom the CTBS was administered in Spring, 1978. There are no missing data.

Least squares and ridge regression data analytic procedures are repeated with the following number of subjects randomly assigned to estimation and prediction groups: 25, 50, 100, 150, 120. Differences in maximum variance inflation factors(VIF), minimum eigenvalues and the standard deviation of the set of five estimated beta coefficients are observed and described with k=0 and at the selected ridge values of k. Sample estimates of the squared length of the coefficient vector ($\beta'\beta$) are calculated for each of the twenty regression equations obtained for each estimation and prediction sample and compared with k=0 and at the selected ridge values of k. Shrinkage in R^2 is described and compared between the two data analytic procedures before and after cross-validation within estimation and prediction groups of the same N.

A descriptive comparison and analysis of obtained results between the two data analytic procedures and across the varying number of subjects per estimation and prediction group is planned. Criteria for comparing the results of the two proce-

dures are reported in the ridge regression literature (Hoerl, 1962; Hoerl & Kennard, 1970a, 1970b; Marquardt & Snee, 1975; Walton, Newman, & Fraas, 1978) and include the following: stability of estimated beta coefficients across samples from the same population, value of the obtained VIF and λ_{min} among procedures, amount of reduction in R^2 from the least squares analysis estimate of R^2 obtained when ridge regression is applied both before and after cross-validation and subjective evaluation of how reasonable the signs and absolute values of the estimated beta coefficients are.

Definition of Terms

Double Cross-Validation -- Given R^2 and the regression equation obtained on two samples from the same population, the regression equation obtained from each sample is applied to predictor scores from the other sample. A cross-validated R^2 is calculated for each of the two samples. Through comparison of the two original R^2 and the two cross-validated R^2, the researcher may evaluate the utility of the original regression equation for use in the prediction of criterion scores, given a set of predictor scores, for other subjects from the same population. The researcher is concerned with the amount of shrinkage in R^2 upon cross-validation.

Ill Conditioned Input Matrix -- An ill conditioned input matrix is a matrix of correlations between predictors in which the off-diagonal correlations are greater than zero, i.e., in which predictor variables are not orthogonal or independent.

Multicollinearity -- Multicollinearity is the condition which exists when predictor variables are not orthogonal or independent.

Ordinary Least Squares Analysis -- Given scores for N subjects from the same population on a criterion variable and two or more predictor variables, ordinary least squares analysis reduces the data to a regression equation. Each predictor variable is mathematically assigned a regression coefficient or weight. The squared correlation between the subjects' obtained and predicted criterion scores is the percentage of criterion variance accounted for (R^2). The

regression equation may be used to mathematically combine predictor scores for any subject from the population in order to estimate her or his criterion score. Ordinary least squares analysis is an unbiased procedure since the expected value of each sample regression coefficient is the corresponding population regression coefficient. Ordinary least squares analysis minimizes the error sum of squares.

Ridge Regression -- Given a matrix of correlations among two or more predictor variables (X'X) and a vector of correlations between each predictor and the criterion (X'Y), ridge regression reduces the data to a regression equation. X'X and X'Y are calculated based upon scores for N subjects from the same population on a criterion variable and two or more predictor variables. Each predictor variable is mathematically assigned a regression coefficient or weight. The squared correlation between the subjects' obtained and predicted criterion scores is the percentage of criterion variance accounted for (R^2). The regression equation may be used to mathematically combine predictor scores for any subject from the population in order to estimate her or his criterion score Ridge regression is a biased procedure since the expected value of each sample regression coefficient is not the corresponding population regression coefficient. Ridge regression biases the estimates of population coefficients through adding a constant $0 \leq k \leq 1$ to each diagonal element of X'X prior to reducing the data to a regression equation.

CHAPTER II

Review Of The Literature

Instability of beta coefficients obtained through ordinary least squares analysis upon application to a second sample from the same population is a problem when prediction is the research objective. Instability of beta weights is particularly noted when predictor variables are highly intercorrelated, i.e., when the data set exhibits multicollinearity (Kerlinger & Pedhazur, 1973). When sample size is small, beta coefficients obtained through ordinary least squares analysis are less stable than beta coefficients obtained using a larger estimation group (Claudy, 1972). Ridge regression is proposed as a data analytic technique which produces more stable beta coefficients and, consequently, better prediction upon application of the equation to a new sample from the same population than ordinary least squares analysis (Hoerl & Kennard, 1970a). A review of the ridge regression literature in which sample N's are reported reveals a mean subjects to predictor variable ratio of 6.46 to 1.00 (Hoerl & Kennard, 1970b; Marquardt & Snee, 1975; Price, 1977) with a maximum subjects to predictor variable ratio of 13.84 to 1.00 (Hoerl & Kennard, 1970b).

Literature which deals with the problem of multicollinearity is reviewed. The ridge regression data analytic tech-

nique is described and the effectiveness of ridge regression as compared to ordinary least squares analysis in the production of stable coefficients for use in prediction is presented as described in the literature. Finally, the literature dealing with the relationship between sample size and the predictive power of obtained regression equations is reviewed.

Multicollinearity

Farrar and Glauber (1967) define multicollinearity as departure from orthogonality of the independent variables. When predictor variables are orthogonal, the zero order correlations (rx_ix_j) between predictors are 0.00 and the input correlation matrix of predictor variables (X'X) is in the form of a unit or identity matrix. When X'X has non-zero off-diagonal elements, i.e., when predictor variables exhibit intercorrelation, the input matrix "is said to be ill conditioned and the original experimental design is said to be non-orthogonal" (Simon, 1975, p. 22).

In experimental designs independent variables are manipulated by the experimenter so that the predictors are operating independently for purposes of the study and random assignment of subjects to treatment is used to insure that uncontrolled variables are not systematically affecting obtained scores on the dependent variable (Blalock, 1963). The assumption that predictor variables are independent is basic to least squares analysis. Claudy (1972) observes that applications of ordinary least squares analysis in undesigned experiments violate a key assumption of the statis-

tical technique. In classic least squares analysis theory the predictor variables are treated as fixed or controlled by the experimenter so that intercorrelation among predictors does not occur (Claudy, 1972). Although the least squares analysis technique is robust with respect to departures from other assumptions, the model is sensitive to departures from orthogonality of the independent variables (Farrar & Glauber, 1967).

The independent or predictor variables in non-experimental research are usually intercorrelated (Pedhazur, 1975). Price (1977) states that independent variables are non-orthogonal in most applications of ordinary least squares analysis. Undesigned experiments in which the researcher cannot manipulate the predictor variables or control extraneous variables directly or through randomization are a predominant mode of educational research (Pedhazur, 1970).

Kerlinger and Pedhazur (1973) note that missing data which produce unequal within cell N's in an experimental design result in non-orthogonal independent variables. The following three sources of multicollinearity in non-experimental designs are noted by Mason, Gunst, and Webster (1975):

1. An over-defined model
2. Sampling techniques
3. Physical constraints on the model or in the population

McNeil and Spaner (1971) state that inclusion of predictor variables representative of higher order, i.e., non-linear, relationships between the dependent variable and predictor variables results in a non-orthogonal design.

Mason, Gunst, and Webster (1975) state that an over-defined model is a model in which the number of subjects is less than the number of predictor variables. Multicollinearity is caused by faulty sampling techniques when the data collected by the experimenter at a given point in time is incomplete, i.e., the model is not fully or correctly defined, or the observed relationship between predictors is specific to the time at which the data were obtained (Mason, Gunst, & Webster, 1975). Physical constraints on the regression model occur when a constant but unknown relationship exists between predictor variables, e.g., when the sum of the predictors must be constant although the value of individual predictors may vary (Mason, Gunst, & Webster, 1975). McNeil and Spaner (1971) state that higher order predictor vectors are necessarily correlated with the linear vector and approve of the inclusion of higher order predictor vectors in the regression equation whenever theoretical or empirical justification exists.

Marquardt and Snee (1975) identify two classifications of data which result in ill conditioned input matrices:

1. Data sets with correlated predictor variables due to the collection of historical data without a design

2. Data sets from an experimental design which have correlated predictor variables due to physical and mathematical constraints such as missing values, correlated errors, split plotting, and non-constant variance

"Multicollinearity constitutes a threat--and often a very serious threat--both to the proper specification and the effective estimation of the type of structural relationship commonly sought through the use of regression techniques"

(Farrar & Glauber, 1967, p. 93). Pedhazur (1975) describes how measurement and specification errors affect multicollinearity. Only in the bivariate case does measurement error result in underestimation of the regression parameters. In multiple regression, measurement error affects the intercorrelations among the predictors. Specification error also affects intercorrelations among the predictors. Specification errors occur when an inappropriate model is postulated, and when the error term is, in fact, correlated with the predictors.

"While it is possible to control specification errors in experimental research by the process of randomization, no such controls are possible in non-experimental research. As is well known, addition or deletion of variables that are correlated with variables originally in the equation will alter the magnitudes, and possible the signs, of the regression coefficients . . . . Interpretation of regression coefficients as indices of effects when specification errors are relatively large may grossly distort the reality of the situation" (Pedhazur, 1975, pp. 263-64).

Mason, Gunst, and Webster (1975) warn researchers who use ordinary least squares analysis with multicollinear data of the following:

1. Prediction outside the region of multicollinearity can result in inaccurate prediction

2. Substantive interpretation of individual regression coefficients is not recommended since the $\hat{\beta}$ are unreliable

Farrar and Glauber (1967) describe multicollinearity mathematically. As interdependence between predictors in the input correlation matrix (X'X) increases, X'X approaches a singular matrix. Since the inverse $(X'X)^{-1}$ of a singular matrix cannot be obtained, a unique solution to the regression equation cannot be found. Diagonal elements of $(X'X)^{-1}$ become infinitely large and the variances of $\hat{\beta}$ for the multicollinear variables also become infinite. "The mathematics, in its brute and tactless way, tells us that explained variance can be allocated completely arbitrarily between linearly dependent members of a completely singular set of variables, and almost arbitrarily between members of an almost singular set." (Farrar & Glauber, 1967, p. 93).

Pedhazur (1975) concludes that small changes in the pattern or magnitude of intercorrelations among predictors can result in substantial differences in obtained regression coefficients. Therefore, although $\hat{\beta}$ is defined as the estimated expected change in Y for each unit change in X in standard score form, when predictors are intercorrelated such straightforward interpretation of $\hat{\beta}$ is not possible (Pedhazur, 1975). As stated by Farrar and Glauber (1967, pp. 93-94), "Attempts to apply regression techniques to highly multicollinear independent variables generally result in parameter estimates that are markedly sensitive to changes in model specification and sample coverage. . . . The increase in sample standard errors for multicollinear regression coefficients virtually assures a tendency for relevant variables to be discarded from regression equations."

Mason, Gunst, and Webster(1975) identify six indicators of an ill conditioned input matrix.

1. The magnitude of the zero order correlations between predictors (rx_ix_j). The larger the rx_ix_j, the greater the multicollinearity.

2. The magnitude of the squared multiple correlation coefficient obtained using all of the predictors except x_j ($R^2_{(j)}$). A high degree of multicollinearity is indicated when the difference between R^2 and the largest $R^2_{(j)}$ is small.

3. Comparison of the overall F statistic with the individual t statistics obtained when predicting the criterion from an individual predictor. Multicollinearity exists if the individual t's are insignificant while the overall F is significant.

4. Calculation of the determinant ($|X'X|$) of X'X. The closer $|X'X|$ is to 0, the greater the degree of multicollinearity. The value of $|X'X|$ ranges between 0 and N.

5. Calculation of an R^2 through regressing each predictor on the remaining p-1 predictors. If any of the obtained R^2 are close to unity, multicollinearity exists.

6. Observation of the magnitude of the diagonal elements of $(X'X)^{-1}$. The larger the magnitude, the greater the multicollinearity.

Marquardt and Snee (1975) state that the two best indicators of an ill conditioned input matrix are the sizes of the maximum variance inflation factor and the minimum eigenvalue (λ_{min}). When two variables are orthogonal, each VIF is 1.00. When variables are correlated, the VIF can range

to infinity. Variance inflation factors are the diagonal elements of $(X'X)^{-1}$. Mason, Gunst, and Webster (1975) state that the VIF is the best of the six measures which they identify for detecting multicollinearity. The maximum variance inflation factor and the minimum eigenvalue are inversely related (Marquardt & Snee, 1975).

Ill conditioned input matrices are indicated by an λ_{min} which is close to zero. Every eigenvalue for an orthogonal matrix equals 1.00. Given a non-orthogonal matrix, the eigenvalues may be larger or smaller than one. The more ill conditioned the matrix, the greater the range of eigenvalues. The sum of the eigenvalues at k=0 is always approximately equal to p, i.e., the number of predictor variables (Simon, 1975).

Five solutions to the problem of multicollinearity are proposed.

1. Redefine the model by eliminating some predictor variables (Mason, Gunst, & Webster, 1975)

2. Compute all 2^p regressions, where p is the number of predictor variables (Hoerl & Kennard, 1970a)

3. Utilize a variable selection technique, e.g., forward stepwise regression (Marquardt & Snee, 1975)

4. Employ principal components analysis (Hoerl & Kennard, 1970a; Mason, Gunst, & Webster, 1975)

5. Gather additional data (Mason, Gunst, & Webster, 1975)

Redefining the model involves subjective judgments by the experimenter which may result in the elimination of significant predictor variables (Mason, Gunst, & Webster, 1975).

Computation of all 2^p regressions inflates experiment-wise error (Winer, 1971). Results obtained through utilization of variable selection techniques are affected by small changes in the sample data (Marquardt & Snee, 1975). Additional data is often unavailable due to lack of funds, changes in the population and other uncontrollable factors. If the additional data collected rarely is present in the population under study, the regression equation may be poorly specified through allowing the rarely present data to inordinately influence the resultant equation (Mason, Gunst, & Webster, 1975). Hoerl and Kennard (1970a), p. 58) conclude that, "with the occasional exception of principal components, these methods don't really give an insight into the structure of the factor space and the sensitivity of the results to the particular set of data at hand."

Due to the greater generalizability of obtained results, Simon (1975) recommends the use of multicollinear predictors despite the problems of specification and interpretation which multicollinear data sets present. Hoerl (1962) suggests ridge regression as an alternative to ordinary least squares analysis in order to diminish the inflation and instability associated with ordinary least squares estimates of population beta coefficients when the input matrix is ill conditioned.

Ridge Regression

Hoerl (1962) first proposed ridge regression as an alternative to ordinary least squares analysis for use in prediction studies when the input correlation matrix of predictor variables (X'X) is ill conditioned. Hoerl and Kennard (1970a) contend

that ridge regression, when used with an ill conditioned X'X, will obtain a prediction equation with the following advantages when compared to the prediction equation obtained through the application of ordinary least squares analysis:

1. Estimated coefficients are closer to the true coefficients on the average

2. The signs of the coefficients are more meaningful

3. Point estimates can be made with smaller mean square error

4. The ridge coefficients are more stable, i.e., more similar to new sets of coefficients obtained through analysis of other data sets from the same population

The standard ordinary least squares analysis model is as follows (Hoerl & Kennard, 1970a):

$$Y = X\beta + e$$

where:

Y = nxl vector of criterion scores when n is the number of observations

X = nxp matrix of predictor scores when p is the number of predictor variables

β = pxl vector of beta coefficients

e = nxl vector of residuals for which $E(e)=0$ and $E(e'e) = \sigma^2 I$ when σ^2 is the error variance in the population and I is the identity matrix

Given the ordinary least squares analysis model, the coefficients are estimated according to the following formula (Conniffe & Stone, 1974):

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

where:

$\hat{\beta}$ = pxl vector of ordinary least squares estimates of the population beta coefficients when p is the number of predictor variables

$(X'X)^{-1}$ = the inverse of the pxp input correlation matrix of predictor variables

$X'Y$ = the pxl vector of correlations between predictor variables and the criterion variable

Hoerl and Kennard (1970a) propose the following alternative ridge estimates of the population β coefficients:

$$\hat{\beta}^* = (X'X + kI)^{-1} X'Y; \quad 0 \leq k \leq 1$$

where:

$\hat{\beta}^*$ = pxl vector of ridge estimates of the population beta coefficients when p is the number of predictor variables

$(X'X+kI)^{-1}$ = the inverse of the pxp input correlation matrix of predictor variables with the constant k; $0 \leq k \leq 1$ added to the diagonal

$X'Y$ = the pxl vector of correlations between predictor variables and the criterion variable

Hoerl and Kennard (1970a, p. 56) state that, "if the shape of the factor space is such that reasonable data collection results in an X'X with one or more small eigenvalues, the distance from $\hat{\beta}$ to β will tend to be large." Hoerl and Kennard (1970a) provide the following formula for the average squared distance between $\hat{\beta}$ and β:

$$E(L^2) = \sigma^2 \sum_{i=1}^{p} (1|\lambda_i)$$

where:

$E(L^2)$ = average squared distance between the estimated and true beta coefficient

σ^2 = error variance of the input data

p = number of predictor variables

$1|\lambda_i$ = reciprocal of the eigenvalue of the ith predictor variable

Following is the formula for the average squared distance between $\hat{\beta}^*$ and β (Hoerl & Kennard, 1970a):

$$E\left[L^2(k)\right] = \sigma^2 \sum_{i}^{p} \lambda_i \left| (\lambda i+k)^2 + k^2 \beta' \ (X'X+kI)^{-2} \beta \right.$$

where:

$E[L^2(k)]$	= average squared distance between ridge estimated and true beta coefficients
σ^2	= error variance of the input data
λi	= eigenvalue of the ith predictor variable
k	= constant
X'X	= input correlation matrix of predictor variables
β	= vector of true (population) beta coefficients
I	= the identity matrix

The first component of the equation represents the error variance, whereas the second accounts for the bias due to the addition of the constant k to the diagonal of the input correlation matrix of predictor variables. When k=0, the second component is also equal to zero and the resulting $E[L^2(k)]$ is the expected squared distance between the ordinary least squares estimate of beta and population beta.

As k increases, the error variance decreases more rapidly than the bias increases. Therefore, at some value of $0<k\geq 1$ the mean square error of $\hat{\beta}^*$ will be less than the mean square error of $\hat{\beta}$ (Simon, 1975).

Through evaluating the equation for the mean square error of $\hat{\beta}$ and the equation for the mean square error of $\hat{\beta}^*$, the effect of adding k to the diagonal of $X'X^{-1}$ may be seen (Simon, 1975). As the eigenvalues become smaller their reciprocals become larger, thus increasing the mean square error of $\hat{\beta}$. Adding the constant k to the diagonal of $X'X^{-1}$ decreases the value of the reciprocals of the eigenvalues and concomitantly decreases the mean square error of $\hat{\beta}^*$. The sum of the recipro-

cals of the eigenvalues divided by p shows how many times greater the mean square error of estimated beta is for a non-orthogonal as compared to an orthogonal design. Through reduction of the sum of the reciprocals of the eigenvalues, ridge regression produces $\hat{\beta}^*$ with smaller mean square error than $\hat{\beta}$. Therefore, the ridge equation behaves more like an orthogonal system and $\hat{\beta}^*$ are more stable than $\hat{\beta}$ (Simon, 1975).

Hoerl and Kennard (1970a, p. 58) conclude that, "the worse the conditioning of X'X, the more $\hat{\beta}$ can be expected to be too long. On the other hand, the worse the conditioning, the further one can move from $\hat{\beta}$ without an appreciable increase in the residual sum of squares. . . . It seems reasonable that if one moves away from the minimum sum of squares point, the movement should be in a direction which will shorten the length of the regression vector."

Figure 1 illustrates the relationship between the error variance of $\hat{\beta}$, the error variance of $\hat{\beta}^*$, the squared bias, and the constant k (Hoerl & Kennard, 1970a). Whereas total variance decreases with increasing values of k, the squared bias increases. Figure 1 shows that values of $0<k\leq 1$ exist for which the mean square error of $\hat{\beta}^*$ is less than the mean square error of $\hat{\beta}$.

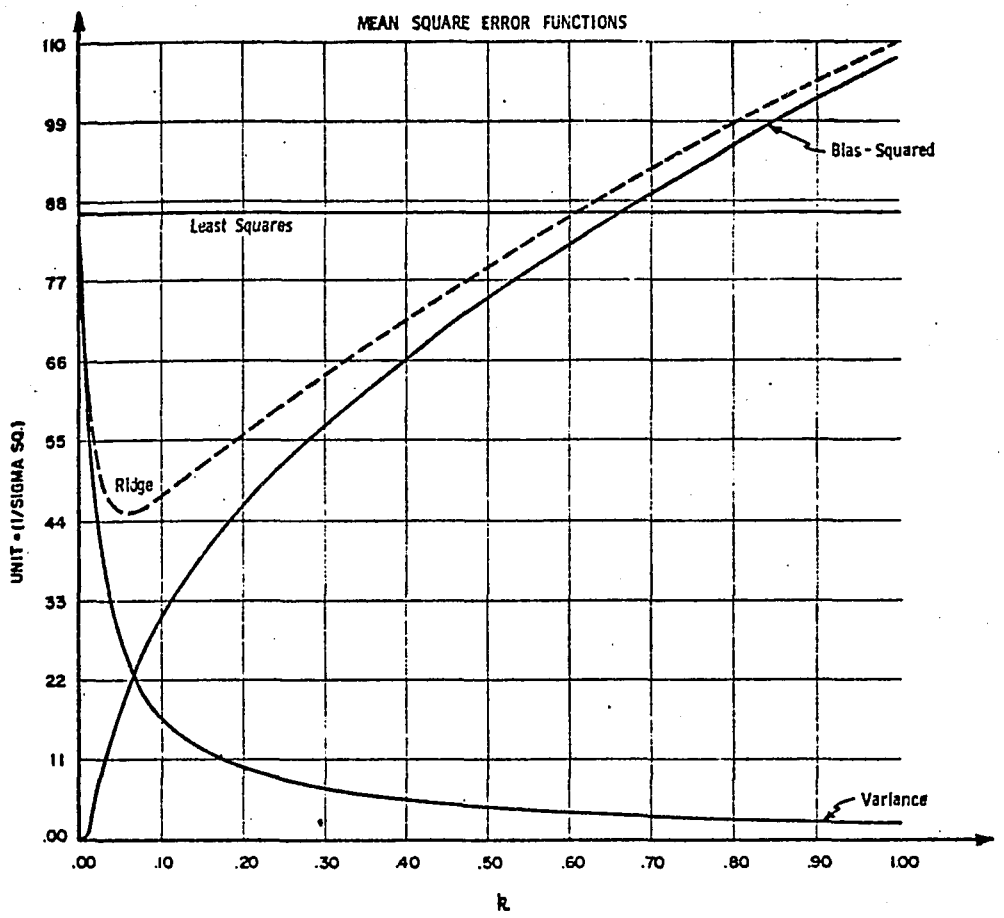

Figure 1

Note. From "Ridge Regression: Biased Estimation for Nonorthogonal Problems" by Arthur E. Hoerl and Robert W. Kennard, Technometrics, February 1970, Vol. 12, No. 1, 55-67. Reprinted by permission.

Hoerl and Kennard (1970a) recognize that the existence of a $0<k\leq 1$ which will result in $\hat{\beta}^*$ estimates with smaller mean square error than $\hat{\beta}$ estimates, is dependent upon the assumption of boundedness on $\beta'\beta$, i.e., the squared length of the population regression vector. A specific upper bound cannot be assigned to $\beta'\beta$, however, "$\beta'\beta$ does not become infinite in practice, and one should be able to find a value or values for k that will put $\hat{\beta}^*$ closer to β than is $\hat{\beta}$. In

other words, unboundedness in the strict mathematical sense, and practical unboundedness are two different things (Hoerl & Kennard, 1970a, p. 63).

Marquardt and Snee (1975) state that when X'X is in correlation form, the population value of any regression coefficient is seldom greater than three. Therefore, the regression coefficient vector is, in practice, finite.

Mason, Gunst and Webster (1975) note that not only is the mean square error of $\hat{\beta}$ too large when X'X is ill conditioned, but the least squares estimates ($\hat{\beta}$) of β also tend to be too large in absolute value. "The ridge estimate," state Marquardt and Snee (1975, p. 6), "gives the smallest regression coefficient consistent with a given degree of increase in the residual sum of squares."

Ordinary least squares estimates of β are often large due to the multicollinearity of X'X. Thus, with an ill conditioned X'X, obtained least squares coefficients do not reflect the importance of the associated independent variables in prediction of the criterion variable (Mason, Gunst, and Webster, 1975). Simon (1975) further states that least squares analysis, when used with a nonorthogonal X'X, produces some coefficient with an incorrect opposite, i.e., not theoretically valid opposite, sign in order to compensate for other coefficients which are too large in absolute value.

Hoerl and Kennard (1970a) recommend that researchers utilize the ridge trace as an aid in the selection of k. The ridge trace is a two dimensional plot of the $\hat{\beta}^*$ and the residual sum of squares ϕ^* for selected values of $0 \leq k \leq 1$

(Hoerl & Kennard, 1970b). The ridge trace follows a path through the sum of squares surface so that for a fixed value of ϕ^*, a value for $\hat{\beta}^*$ is chosen which has minimum length. In practice the researcher chooses values of $0 \leq k \leq 1$ and then computes the residual sum of squares for each selected value of the constant k. Marquardt and Snee (1975, p. 11) state that "sensitivity analysis is an aim of ridge regression" and the ridge trace displays $\hat{\beta}^*$ graphically so that the researcher can see which coefficients are sensitive to the data. No more than ten coefficients should be plotted on a given ridge trace (Marquardt & Snee, 1975).

Figure 2 is an example of the ridge trace on which data analyzed by Hoerl and Kennard (1970b) are plotted. Coefficients stabilize as k increases. If k were increased to infinity, computed coefficients would, at some point, equal zero (Simon, 1975).

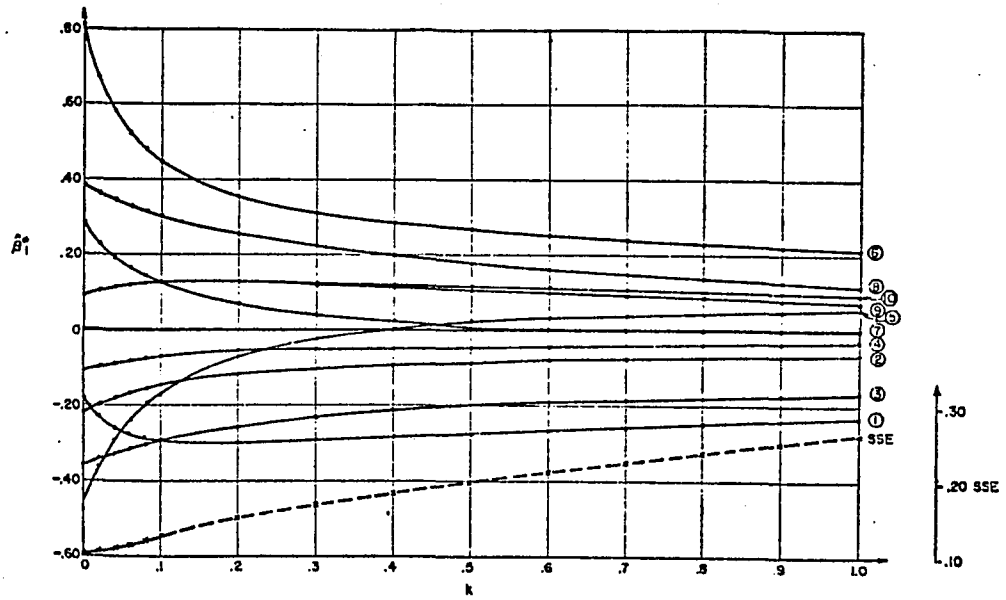

Figure 2

Note. From "Ridge Regression: Applications to Non-orthogonal Problems" by Arthur E. Hoerl and Robert W. Kennard, Technometrics, February 1970, Vol. 12, No. 1, 69-82. Reprinted by permission.

The selection of a value of k according to the observational procedure recommended by Hoerl and Kennard (1970a) is a subjective process. Given the ridge trace, Hoerl and Kennard recommend use of the following criteria for the selection of k:

1. At the selected value of k absolute values of the estimated beta coefficients are reasonable given knowledge of the factors for which the coefficients represent rates of change

2. At the selected value of k the system stabilizes and the coefficient vectors change little beyond the selected value of k

3. At the selected value of k estimated beta coefficients with signs that appear unreasonable at k=0, given knowledge of the factors, changed to the expected sign

4. At the selected value of k the residual sum of squares is not large relative to the minimum residual sum of squares or relative to a reasonable residual sum of squares, given knowledge of the system under study

The ridge trace has one curve or trace per estimated beta coefficient. The value of k where the estimated beta coefficients stabilize results in a set of coefficients which are not sensitive to small changes in the estimation data (Marquardt & Snee, 1975). Marquardt and Snee (1975) contend that the increase in the residual sum of squares with increasing values of k is not relevant if the objective of the regression procedure is to develop stable coefficients for use in future prediction rather than to obtain the closest fit possible to the estimation data. Given orthogonal predictors, coefficients stabilize at k=0 so that the ordinary least squares estimates of β may be used.

Methods for point estimation of k are also available (Lindley & Smith, 1972; Mallows, 1973; Farebrother, 1975; McDonald & Galarneau, 1975; Hoerl, Kennard, & Baldwin, 1975; Obenchain, 1975; Hoerl & Kennard, 1976; Lawless & Wang, 1976; Golub, Heath & Wahba, 1977; Dempster, Schatzoff & Wermuth, 1977). Consensus is not achieved as to the best method for the selection of k (Hoerl, 1979).

Gibbons (1978) applies ten methods for the point estimation of k to a data set previously analyzed by McDonald and Schwing. The range of k values selected vary between 0.004 and 1.0. Gibbons (1978, pp. 21-22) states, "This disparity between the specified k values highlights the need for the systematic study of these estimators."

Within the range of selected k values in Gibbon's (1978) application of selection procedures to the McDonald and Schwing data, the selection technique attributed to McDonald and Galarneau (1975) selects a low k, the Hoerl and Kennard technique (1976) selects a low medium value of k, and an operationalization of the technique suggested by Hoerl (1962) which purports to reduce the length of the coefficient vector without an appreciable increase in the error sum of squares selects a high level of k.

The selection technique developed by McDonald and Galarneau (1975) chooses the value of k which minimizes $|\hat{\beta}^{*\prime}\hat{\beta}^{*}-Q|$ and defaults to the least squares estimator if Q is negative. The expression $|\hat{\beta}^{*\prime}\hat{\beta}^{*}-Q|$ is evaluated for selected values of k. The equation for Q is as follows:

$$Q = \hat{\beta}'\hat{\beta}-(Y-X\hat{\beta})'(Y-X\hat{\beta})/(n-p-1)\Sigma\lambda_i^{-1}$$

where:

$\hat{\beta}'\hat{\beta}$	= the estimated squared length of the coefficient vector at k=0
$(Y-X\hat{\beta})'(Y-X\hat{\beta})$	= the error (residual) sum of squares at k=0
n	= the number of observations or subjects
p	= the number of predictor variables
$\Sigma\lambda_i^{-1}$	= the sum of the reciprocals of the eigenvalues at k=0

The Hoerl and Kennard (1976) technique is an iterative procedure. A point estimate of k is selected in the Gibbons (1978) application through the use of a 10^{-4} convergence criterion to successive k values, defaulting to k=0 if convergence is not obtained in thirty iterations. The algorithim (Hoerl & Kennard, 1976) is as follows:

$$k_o = p((Y-X\hat{\beta})'(Y-X\hat{\beta})|n-p-1)|\hat{\beta}'\hat{\beta}$$

$$k_1 = p((Y-X\hat{\beta})'(Y-X\hat{\beta})|n-p-1)|\hat{\beta}^*(k_o)'\hat{\beta}^*(k_o)$$

$$\vdots$$

$$k_t = p((Y-X\hat{\beta})'(Y-X\hat{\beta})|n-p-1)|\hat{\beta}^*(k_{t-1})\hat{\beta}(k_{t-1})$$

where:

p	= the number of predictor variables
$(Y-X\hat{\beta})'(Y-X\hat{\beta})$	= the error (residual) sum of squares at k=0
n	= the number of observations or subjects
p	= the number of predictor variables
$\hat{\beta}'\hat{\beta}$	= the squared length of the coefficient vector at k=0
$\hat{\beta}^{*\prime}(k_o)\hat{\beta}^*(k_o)$	= the squared length of the coefficient vector at $k=k_o$
$\hat{\beta}^{*\prime}(k_{t-1})\hat{\beta}^*(k_{t-1})$	= the squared length of the coefficient vector at $k=k_{t-1}$
t	= the maximum number of iterations

The selection technique proposed by Hoerl (1962) is operationalized by Gibbons (1978) for applications purposes as follows. The level of k selected is that level which maximizes the second derivative of the square root of the residual sum of squares with respect to the length of the coefficient vector. The following equation is evaluated for selected values of k:

$$d^2\left((Y-X\hat{\beta}^*)'(Y-X\hat{\beta}^*)\right)^{1/2}\left|d(\hat{\beta}^{*\prime}\hat{\beta}^*)\right| =$$

$$\left((Y-X\hat{\beta}^*)'(Y-X\hat{\beta}^*)\right)^{-1/2}\{-k^2(\hat{\beta}^{*\prime}\hat{\beta}^*)\Big|$$

$$(Y-X\hat{\beta}^*)'(Y-X\hat{\beta}^*)+\hat{\beta}^{*\prime}\hat{\beta}^*\,\Big|\,\hat{\beta}^{*\prime}(X'X+kI)^{-1}\hat{\beta}^*\Big)-k\}$$

where:

$(Y-X\hat{\beta}^*)'(Y-X\hat{\beta}^*)$	=	the residual sum of squares at a selected value of k
$\hat{\beta}^{*\prime}\hat{\beta}^*$	=	the estimated squared length of the coefficient vector at a selected level of k
$(X'X+kI)$	=	the input correlation matrix of predictors with a selected value of k added to each diagonal element

Newhouse and Oman (1971) caution against the use of ridge regression until a mathematical procedure for the selection of k is rigorously established. Newhouse and Oman (1971, p. 14) conclude, ". . . it is probable that a rule for determining k could be defined, such that β_k $(\hat{\beta}^*)$ would be better than OLS$(\hat{\beta})$. Until the properties of such a ridge estimator are rigorously derived, however, we would caution the reader against using ridge analysis to estimate regression coefficients." Since the value of k selected is dependent upon the sample at hand, Newhouse and Oman (1971) conclude that the assumption that k is independent of the random error vector is violated.

Conniffe and Stone (1974) contend that ridge regression ignores the warning implicit in an ill conditioned X'X and state that, since variables are subject to measurement error, a small eigenvalue may, in a practical sense, indicate singularity of X'X. Conniffe and Stone (1974, p. 182) state that, "Hoerl and Kennard's proof that the mean square error of $\hat{\beta}^*$ is less than that of $\hat{\beta}$ for certain values of k is certainly correct. However, the proof is valid only if the appropriate value of k is assumed known. The ridge procedure involves the estimation of k so Hoerl and Kennard have not proved that

the mean square error of $\hat{\beta}^*$ is in fact smaller than that of $\hat{\beta}$." The authors (Conniffe and Stone, 1974) recommend that researchers avoid ridge regression and instead employ traditional methods for dealing with an ill conditioned X'X such as respecification of the model, additional data collection, and factor analysis.

Obenchain (1974, p. 441) states that, "The calculation, examination, and interpretation of a ridge trace for an ill conditioned regression problem is an invaluable data analytic tool. The insights one may gain by performing ridge analysis are neither different from nor more profound than those which can be developed by other means; one is merely forcing oneself to obtain additional information about the fitted coefficients so that the results of the regression analysis will not be naively misinterpreted."

Banerjee and Carr (1971) recommend that the mean square error of $\hat{\beta}^*$ should be compared with a modified mean square error. The following formula is proposed:

$$E(L^2) \text{ modified} = \sigma^2 \sum_{i}^{p} [1/(\lambda i+k)]$$

where:

$E(L^2)$ modified	=	modified average squared distance between estimated and true beta coefficients
σ^2	=	error variance of the input data
p	=	number of predictor variables
$1/(\lambda i+k)$	=	reciprocal of the eigenvalue of the ith predictor variable plus the constant k

Even given the proposed modification of $E(L^2)$, the $\hat{\beta}^*$ have smaller mean square error than the $\hat{\beta}$, although the magnitude of the difference is decreased (Banerjee and Carr, 1971).

Simon (1975, p. 43) concludes that, "it appears that an optimal choice of k (or interval of k values) is an open question at this time unless one has prior knowledge about the length and/or direction of the unknown coefficient vector.

Although the problem of how and where to select the best value of k has not yet been resolved, the overall superiority of ridge regression over least squares regression in the analysis of nonorthogonal data has not been seriously questioned." Hoerl (1979) confirms that consensus is not achieved as to the best method for the selection of k.

Hoerl and Kennard (1970b) apply ridge analysis to two regression problems with nonorthogonal predictors which were previously analyzed by other researchers. The first application is to a ten predictor example previously analyzed by Gorman and Toman who utilized a variable selection procedure. Hoerl and Kennard (1970b) illustrate that the variable selection procedure employed by Gorman and Toman fails to eliminate the tendency to overestimate β and also fails to produce a stable $\hat{\beta}$.

Hoerl and Kennard (1970b) plot the ridge trace for the ten predictor example. The ridge trace shows overestimated and unstable $\hat{\beta}$ at $k=0$ with predictors 5 and 6 exhibiting rapid decreases in absolute value as $k>0$. The correlation between predictors 5 and 6 is 0.84. Hoerl and Kennard (1970b) therefore conclude that the negative sign which predictor 5 has at $k=0$ is erroneous and the fact that predictor 5 is driven toward zero as $k>0$ probably reflects the fact that predictors 5 and 6 are measures of the same factor. The authors (Hoerl & Kennard, 1970b) also note that predictor 1 becomes the strongest negatively signed predictor as $k>0$. The system stabilizes with $.2<k<.3$.

The variable selection procedure employed by Gorman and Toman resulted in the elimination of predictors 9, 10, 4, and 1, but the ridge trace for the reduced system shows that overestimation and instability of $\hat{\beta}$ are still present. And, in fact, the estimated squared length of the coefficient vector ($\beta'\beta$) is now 1.91 as compared with 1.35 before the four predictors were eliminated. Based upon the original ridge trace, Hoerl and Kennard (1970b) eliminate predictors 5 and 7 which helps control the overestimation and instability of $\hat{\beta}$.

Hoerl and Kennard (1970b), however, do not recommend the use of variable selection procedures." In this example, the system stabilizes in the region k = 0.2 to k = 0.3; therefore, choose k = 0.25 as a stable point solution . . . . Then view the system as one of ten controlled factors with the coefficients as the 'best' estimates.Factors with small effects have small coefficients.To 'discard' a factor, set it at its average value for all predictions, which is the equivalent of setting the coefficient equal to zero (Hoerl & Kennard, 1970b, pp. 74-75)."

Using a thirteen predictor example, Hoerl and Kennard (1970b) compare a variable selection procedure guided by the ridge trace with principal components analysis used by Jeffers in a former analysis of the same data. The variance accounted for by the reduced ordinary least squares solution suggested by the ridge trace is 69.55% as compared with the 63.97% accounted for by Jeffers' principal component solution.

Zaitzeff (1971) compares the results of stepwise regression and a ridge regression procedure through which predictors with $\hat{\beta}^*$ of less than .05 at the selected value of k are eliminated. Zaitzeff (1971) subsequently applies ridge analysis to the reduced set of predictors which remain after variable elimination based upon the first ridge analysis. Zaitzeff (1971) notes that the resultant ridge equation is more stable and the signs of coefficients are more meaningful than in the stepwise regression model.

Guilkey and Murphy (1975) compare the results of a directed ridge estimator (DRE) procedure which adds k only to those diagonal elements of $X'X^{-1}$ which correspond to small eigenvalues with ridge regression and ordinary least squares analysis. Data are investment equations developed in a text by Christ. Twenty equations with different numbers of predictors, degrees of multicollinearity and successful fit using an ordinary least squares procedure are presented by Christ. Guilkey and Murphy (1975) analyze the equations through the DRE and ridge regression procedures as well as with ordinary least square analysis.

No difference in sign or significance of any coefficients are noted among the three methods when applied to the data set with the least multicollinearity although the standard errors of estimated β are slightly reduced with the ridge methods. When applied to the most multicollinear equation, the ridge methods result in changes in the signs of coefficients and reduced standard errors of estimated β when compared with the ordinary least squares equation (Guilkey & Murphy, 1975).

Marquardt and Snee (1975) apply ordinary least squares analysis, ridge regression, and a generalized inverse procedure to a nine predictor example in which the percent conversion of n-heptane to acetylene is the criterion. The coefficients stabilize by k=.05. Overestimated least squares coefficients are reduced in the ridge and generalized inverse models. Marquardt and Snee (1975) conclude that the nine-term ridge model and the generalized inverse model predict percent conversion of n-heptane to acetylene more accurately upon extrapolation within the ranges covered by the predictor variables individually than either the nine or reduced five term ordinary least squares model.

Marquardt and Snee (1975) demonstrate the application of ridge regression to two data sets with 33 and 15 predictor variables respectively. The criterion in the first example is corn yield. Ridge results are compared with the full, stepwise, and PRESS models previously applied by Cady and Allen. Marquardt and Snee (1975) conclude that the ridge model, which keeps all thirty-three predictors in the equation, predicts well and demonstrates the role of all variables in the model. Marquardt and Snee (1975) base their conclusions upon results obtained when the ridge equation obtained with an estimation sample is applied to a prediction sample.

Similar results are obtained when Marquardt and Snee (1975) apply ridge regression to the fifteen predictor problem. Ridge coefficients are more stable and their signs more meaningful. The ridge model works in practice as a prediction model whereas the least squares and stepwise regression models do not (Marquardt and Snee, 1975).

Price (1977) compares the results of ordinary least squares analysis and ridge regression in the prediction of employee satisfaction from five predictor variables. The data are multicollinear. Based on the criteria for selection of k through inspection of the ridge trace (Hoerl & Kennard, 1970a), Price (1977) selects k=.05. Price (1977, p. 765) concludes, ". . . ridge solutions can be dramatically different from regular least squares solutions when the predictors are not orthogonal. The different solutions each suggest an associated different interpretation of the data. Certainly, the least squares solution that shows incorrect signs for potentially important variables is unsatisfactory. For example, the role of X_5, applies company policy fairly, in the analysis given above has a meaningful interpretation only in the ridge solution . . . . ridge regression produces solutions that are reasonable and that suggest directions for further investigation that may not be apparent from the regular least square solution."

Walton, Newman, and Fraas (1978) apply ridge regression and least squares analysis in the prediction of counselor practicum ratings, utilizing a sample of 93 counselor education students. Predictor variables include undergraduate grade point average, final graduate grade point average, Miller Analogies test score, and type of undergraduate institution. Simple, squared, and interaction predictor variables are used. Subjects are randomly assigned to prediction and estimation groups. Results produce a shrunken ordinary least squares R^2 which is more similar to the true population esti-

mate of R^2 than the shrunken or non-shrunken ridge R^2. Ridge regression coefficients are less sensitive to the particular variable scores in the estimation data which results in a smaller R^2 but produces greater stability upon application of the regression equation to a new sample.

Walton, Newman, and Fraas (1978, p. 12) conclude that, "If the research project requires stable coefficients as would be the case in making a point prediction over different samples, ridge regression may be the appropriate analytical tool. However, if the purpose of the research project is to test a hypothesis, the use of multiple linear regression would be more appropriate." Observing the stability of ordinary least squares estimates of population beta after cross-validation or correction for shrinkage is recommended. If relatively stable results are obtained, ridge analysis might be unnecessary (Walton, Newman, & Fraas, 1978).

Sample Size

Claudy (1972, p. 314) concludes that, "Preliminary empirical studies indicated that the variability of sample beta weights is inflated at all sample sizes, but that the degree of inflation decreases with increasing sample size." Overfitting of the regression surface to the sample data results when the assumption that predictor variables are independent is violated (Farrar & Glauber, 1967).

Claudy (1972) states that overfitting of the regression surface results in the following two types of errors:

1. Population R^2 is overestimated by sample R^2.

2. The variability of $\hat{\beta}$ is inflated relative to the variability of β

Claudy (1972, p. 314) further states that, "little attention has been paid to the inflation of the variability of the n sample beta weights as compared with the actual variability of the n population beta weights,and in fact this is a seldom mentioned finding."

Cureton (1962) proposes a least deviant procedure to reduce the variability of sample beta weights. The least deviant procedure necessitates two samples from the same population and involves the following steps:

1. Estimate β for each subsample
2. Arrange the $\hat{\beta}_i$ from both samples in rank order
3. Calculate the median
4. Use from each pair of $\hat{\beta}_i$, the $\hat{\beta}_i$ which is closest to the median.

Claudy (1972) conducts an empirical study within which Cureton's (1962) least deviant procedure, ordinary least squares, the average variance-reduction procedure, criterion correlations, and equal raw score weights are compared. Both the least deviant and average variance reduction procedures are inferior to the other procedures in predictive effectiveness at smaller sample sizes. Claudy (1972) concludes that, given an X'X with high positive correlations, equal raw score weights should be used for samples of less than fifty. Ordinary least squares $\hat{\beta}$ should be used with larger samples.

Fisher (1924) proposes the following formula for the estimation of the expected value of sample multiple R^2:

$$R^2 = 1 - \frac{N-n}{N-1} \quad (1-\rho^2)$$

where:
R^2 = expected value of the sample multiple correlation coefficient squared
N = number of subjects
n = number of predictor variable
ρ^2 = population multiple correlation coefficient squared

The formula shows that the degree of overestimation of population R^2 is inversely related to sample size.

A review of the ridge regression literature in which the technique is applied to real world data and sample N's are reported reveal N's which range from 29 to 228 with a mean subject to predictor variable ratio of 6.46 to 1.00 (Hoerl & Kennard, 1970b; Marquardt & Snee, 1975; Price, 1977). The largest subjects to predictor variable ratio reported in the ridge regression literature is 13.84 to 1.00 (Hoerl & Kennard, 1970b). Table 1 reports sample size, number of predictors, and the number of subjects per predictor variable for ridge regression studies in which the technique is applied to real world data and sample N's are reported.

Table 1

Sample Size, Number of Predictors, and Number of Subjects Per Predictor as Reported in Marquardt and Snee (1975), Hoerl and Kennard (1970b) and Price (1977)

Study	Sample Size	No. of Predictors	No. of Subjects Per Predictor
Marquardt and Snee (1975)	228	33	6.91
	29	15	1.93
Hoerl and Kennard (1970b)	36	10	3.6
	180	13	13.84
Price (1977)	30	5	6.00

With the exception of Walton, Newman, and Fraas' (1978) application of ridge regression in the prediction of counselor practicum ratings, no applications of ridge regression to prediction problems with education were located. No applications of ridge regression to samples of varying N from the same population were located.

CHAPTER III

Method And Procedures

Ordinary least squares analysis and ridge regression are compared across samples from the same population with respect to stability of estimated beta coefficients, magnitude of the maximum variance inflation factor, value of the minimum eigenvalue, magnitude of R^2 before and after cross-validation, standard deviation of each set of estimated beta coefficients, the estimated squared length of the coefficient vector, and the subjective evaluation of how reasonable the signs and absolute values of the estimated beta coefficients are. The descriptive comparisons outlined are conducted with samples from the same population which have the following number of subjects per predictor variable: 5, 10, 20, 30, and 40. First, the data set is described. Secondly, the methodology, including reference to the computer programs utilized which are contained in appendices A through H, is outlined. Thirdly, research questions are stated.

The Data Set

Data consist of raw scores on the following subtests of the Comprehensive Test of Basic Skills (CTBS), Level 1, Form S for 391 fourth grade students: Mathematics Concepts, Reading Vocabulary, Reading Comprehension, Language Spelling,

Language Mechanics, and Language Expression. The Mathematics Concepts subtest score is the criterion, while the five reading and language subtest scores are predictors. Data are provided by Educational Testing Services, Princeton, New Jersey. Students were administered the CTBS in Spring, 1978. There are no missing data.

Table 2 reports test-retest reliability coefficients for five subtests of the CTBS/S, Level 1. The test was administered at an interval of four weeks. Data are reported in the Comprehensive Tests of Basic Skills, Forms S and T All Levels, Technical Bulletin No. 2 (1977).

Table 2

Test-Retest Reliability Coefficients for Five Subtests of the CTBS/S, Level 1

Subtest	Reliability Coefficient
Mathematics Concepts	.89
Reading Vocabulary	.91
Reading Comprehension	.90
Language Spelling	.83
Language Mechanics	.79
Language Expression	.85

Table 3 reports internal consistency reliability coefficients calculated with the Kuder-Richardson formula 20 (KR20) for five subtests of the CTBS/S Level 1 for grades 3.7 and 4.7. Data are reported in the Comprehensive Tests of Basic Skills Form S All Levels, Technical Bulletin No. 1 (1974).

Table 3

Internal Consistency Reliability Coefficients for Five Subtests of the CTBS/S, Level 1 for Grades 3.7 and 4.7

Subtest	Reliability Coefficient	
	Grade 3.7	Grade 4.7
Mathematics Concepts	.89	.89
Reading Vocabulary	.92	.93
Reading Comprehension	.94	.94
Language Spelling	.89	.91
Language Mechanics	.85	.86
Language Expression	.90	.91

Content validation procedures are described in the Comprehensive Tests of Basic Skills Form S All Levels, Technical Bulletin No. 1 (1974). Form S was developed in order to update CTBS/Q and CTBS/R. Therefore, content validation began with a critical review of CTBS/Q. Items were selected for inclusion in CTBS/S, Level 1 based upon the review of CTBS/Q and the following four criteria:

1. The item's discrimination in difficulty among grades 2.7, 3.7, and 4.7

2. The item's discrimination in difficulty among five achievement groups

3. The point-biserial correlation coefficient between the item and total test score

4. The item's effect on total test reliability

The CTBS/S, Level 1 was reviewed by black and Spanish-speaking teachers, content and curriculum experts, and specialists in minority education to eliminate items which were considered to be biased against black or Spanish-speaking students.

Methodology

The PL/I source program which is listed in Appendix A is used to randomly select the following number of subjects per sample: 25, 50, 100, 150, 200. Two samples at each sample size are selected. Selection is without replacement within each sample and with replacement across samples. Consequently, each student's observed score is treated as her/his true score. Thus, the effects of measurement unreliability are not considered.

The FORTRAN source program listed in Appendix B inputs the sample raw scores from each of the ten samples as well as for N=391 and outputs the correlation matrix of predictor variables (X'X), the vector of correlations between each predictor variable and the criterion (X'Y) and the estimated mean and standard deviation for raw scores on each variable.

A FORTRAN source program which is listed in Appendix C is used to calculate twenty regression equations, varying k in the range [0,1].

The first equation is the equation obtained through ordinary least squares analysis. Nine equations use values of k from k=0.01 through k=0.09, incrementing k by 0.01 at each iteration. Ten additional equations are calculated for values of k from k=0.1 through k=1.0, incrementing k by 0.1

at each iteration. The twenty regression equations are calculated for each of the ten samples as well as for N=391. Input is the matrix of intercorrelations among predictors and between predictors and the criterion, i.e., X'X and X'Y.

In addition, the standard deviation and averages of each raw score variable, the raw scores for the sample, and the formats used on input and output are read in. The output of this program consists of the value of k, the residual sum of squares (RSS), R^2, and $\hat{\beta}^*$ for each of the twenty levels of k.

The FORTRAN source program for the calcomp plotter which is contained in Appendix D is used to output the ridge trace for the twenty regression equations calculated for each of the ten samples as well as for N=391.

The ridge trace plots the error sum of squares and the estimated beta coefficients at each of the twenty values of k for which a regression equation was calculated.

The source program inputs the file containing the output from the program in Appendix C, i.e., k, RSS, R^2, and $\hat{\beta}^*$ for each of the twenty levels of k. This program also inputs a file that contains the constants used to set up the parameters for the plotter. The output of this program is a computer drawn, graphic representation of the ridge trace. The error sum of squares and the estimated beta coefficients are mapped against k on separate graphs.

An observational value of k is selected based upon the following four criteria suggested by Hoerl and Kennard (1970a):

1. At the selected value of k absolute values of the estimated beta coefficients are reasonable given knowledge

of the factors for which the coefficients represent rates of change

2. At the selected value of k the system stabilizes and the coefficient vectors change little beyond the selected value of k

3. At the selected value of k estimated beta coefficients with signs that appeared unreasonable at k=0, given knowledge of the factors, change to the expected sign

4. At the selected value of k the residual sum of squares is not large relative to the minimum residual sum of squares or relative to a reasonable sum of squares, given knowledge of the system under study

The FORTRAN source program listed in Appendix E selects a point estimate of k according to the selection procedures developed by Hoerl (1962), McDonald and Galarneau (1975), and Hoerl and Kennard (1976) which are described in Chapter II for each of the ten samples as well as for N=391.

This program inputs the file containing K, RSS, R^2, and $\hat{\beta}^*$ for each level of k, the file containing X'X and X'Y, the file of raw scores, the file with the standard deviations and averages for each raw score variable, and a file that contains the value of k=0 and the value of k picked by observation as a result of the graphic mapping previously described.

The Hoerl (1962) and McDonald and Galarneau (1975) procedures are applied to each of the twenty values of k previously input. Both of these procedures thus choose a single existent k value that best fits their selection criteria.

The Hoerl and Kennard (1976) procedure performs at most thirty iterations of the equation and produces K, RSS, R^2,

and $\hat{\beta}^*$ for each iteration. The final k_i is selected whenever the difference between k_i and k_{i+1} is less than 10^{-4}.

The first output file contains the value of k=0, the value of k selected by observation and those selected by the three previously described procedures. These k values are output in ascending order. Each k is paired with the name of the procedure that selected it. The k selected by the Hoerl and Kennard (1976) procedure is also linked with the number of iterations performed to select that k. The second output file is used to order the twenty k's which were input and the k selected by Hoerl and Kennard (1976) with its associated RSS, R^2, and $\hat{\beta}^*$. Thus, this file contains twenty-one values of k and associated RSS, R^2, and $\hat{\beta}^*$ for each.

The FORTRAN source program listed in Appendix F is used to calculate and output the inverse of X'X and the eigenvalues for each original input correlation matrix of predictors. The same FORTRAN source program is used to output $(X'X + kI)^{-1}$ and the associated eigenvalues for each selected value of k>0. The eigenvalues and $X'X^{-1}$ are calculated for each of the ten samples as well as for N=391.

The FORTRAN source program listed in Appendix G is used to cross-validate the regression equations for each selected k including k=0 and to output $\hat{\beta}^{*\prime}\hat{\beta}^*$ and the average and standard deviation of the five sample beta weights for all twenty-one levels of k.

Dorans, Drasgow, and Tucker (1978) identify the following two acceptable cross-validation procedures:

1. The application of estimation sample raw score coefficients to prediction sample raw scores

2. The application of estimation sample standard score coefficients to prediction sample scores which are standardized based upon the mean and standard deviation of the estimation sample.

The FORTRAN source program listed in Appendix G employs the first of these procedures.

Each of the ten original samples serves in turn as both the estimation and prediction group since double cross-validation as described by Kerlinger and Pedhazur (1973, p. 284) is employed. The standard regression weights obtained through ordinary least squares analysis, i.e., with k=0, for each sample are converted to raw score weights. The raw score weights are applied to raw scores for the other sample of the sample size. The cross-validated R^2 is calculated. Similarly, the standardized ridge regression weights at the selected values of k for each sample are converted to raw score weights and the resultant regression equation is applied to raw scores for the other sample of the same size. The cross-validated R^2 is calculated.

The FORTRAN source program listed in Appendix G converts the estimated beta coefficients to estimated raw score coefficients, produces the cross-validated R^2, and calculates the average and standard deviation of the five beta weights obtained with k=0 and at each selected level of k for all ten samples. The FORTRAN source program listed in Appendix G also estimates the squared length of the coefficient vector ($\beta'\beta$) for each

level of k for which the source programs in Appendices C & D calculate a regression equation. All the described statistics are calculated for each of the ten samples and for N=391.

The FORTRAN source program for the calcomp plotter which is contained in Appendix H is used to plot estimated $\beta'\beta$ at all levels of k for each of the ten samples and for N=391. The input to this program is the file containing estimated $\beta'\beta$ for each level of k. The output is a computer drawn graphic representation of estimated $\beta'\beta$ mapped against k.

All computer software contained in Appendices A-H has been validated through the analysis of test data with known results.

A descriptive comparison and analysis of obtained results between the two data analytic procedures and across the varying N's is planned.

Criteria for comparing the results of the two procedures are reported in the ridge regression literature (Hoerl, 1962; Hoerl and Kennard, 1970a, 1970b; Marquardt and Snee, 1975; Walton, Newman, and Fraas, 1978) and include the following: stability of estimated beta coefficients across samples from the same population, value of the obtained VIF and λ_{min} among procedures, amount of reduction in R^2 from the least squares analysis estimate of R^2 obtained when ridge regression is applied both before and after cross-validation and subjective evaluation of how reasonable the signs and absolute values of the estimated beta coefficients are.

Research Questions

Question 1: Is the maximum variance inflation factor obtained through inversion of the input correlation matrix with each of the selected values of k added to the diagonal smaller than the maximum variance inflation factor obtained through inversion of the original input correlation matrix at the following sample sizes: 25, 50, 100, 150, 200, 391?

Question 2: Is the minimum eigenvalue obtained through analysis of the input correlation matrix with each of the selected values of k added to the diagonal larger than the minimum eigenvalue obtained through analysis of the original input correlation matrix at the following sample sizes: 25, 50, 100, 150, 200, 391?

Question 3: Is the standard deviation of the five sample beta weights obtained through ridge regression at the selected values of k less than the standard deviation of the five sample beta weights obtained through ordinary least squares analysis at the following sample sizes: 25, 50, 100, 150, 200, 391?

Question 4: Is the estimated squared length of the coefficient vector which is calculated based upon ridge regression coefficients at the selected values of k less than the estimated squared length of the coefficient vector calculated based upon ordinary least squares regression coefficients at the following sample sizes: 25, 50, 100, 150, 200, 391?

Question 5: Is the amount of shrinkage in the R^2 obtained through ridge regression upon cross-validation with a prediction group at the selected values of k less than the amount of

shrinkage in the R^2 obtained through ordinary least squares analysis upon cross-validation with a prediction group at the following sample sizes: 25, 50, 100, 150, 200?

Question 6: Are the estimated beta coefficients obtained through ridge regression at each selected level of k more similar to a second set of estimated beta coefficients at the level of k selected by the same selection procedure in the other sample of the same size than estimated coefficients obtained through ordinary least squares analysis at the following sample sizes: 25, 50, 100, 150, 200?

CHAPTER IV

Results

Introduction

When prediction is the objective of educational research, a goal of data analysis is to obtain a regression equation which retains its predictive power upon application to a second sample from the same population. When predictor variables exhibit high positive intercorrelations, obtained regression coefficients may fluctuate widely from sample to sample (Kerlinger & Pedhazur, 1973). In addition, variance accounted for (R^2) is overestimated, the variability of the p sample beta weights is inflated (Claudy, 1972), and the estimated squared length of the coefficient vector ($\beta'\beta$) exceeds the population value of $\beta'\beta$ (Hoerl & Kennard, 1970a). Least squares analysis produces regression equations with the above described problems when data are multicollinear (Hoerl & Kennard, 1970a; Claudy, 1972; Kerlinger & Pedhazur, 1973).

Hoerl (1962) contends that ridge regression should be used in lieu of ordinary least squares analysis to reduce and/or eliminate the above described problems when data are multicollinear. Marquardt and Snee (1975) state that the sizes of the maximum variance inflation factor and the minimum eigenvalue are the best indicators of multicollinear predictors. The higher the maximum variance inflation factor,

the more multicollinear the data. The magnitudes of the maximum variance inflation factor and the minimum eigenvalue are inversely related. The addition of a constant $0<k\leq 1$ to the diagonal of $X'X$ reduces the maximum variance inflation factor and increases the minimum eigenvalue so that regression equations which are calculated through the use of $(X'X+k)^{-1}$ as opposed to $X'X^{-1}$ behave more like equations which are calculated given orthogonal predictors (Hoerl & Kennard, 1970a, 1970b). The following ridge k selection procedures are described in Chapter 2: Observation, Hoerl (1962), McDonald and Galarneau (1975), and Hoerl and Kennard (1976).

Data are eleven samples which consist of students' raw scores on the following subtests of the Comprehensive Test of Basic Skills (CTBS), Level 1, Form S: Mathematics Concepts, Reading Vocabulary, Reading Comprehension, Language Spelling, Language Mechanics, and Language Expression. Mathematics Concepts score serves as the criterion. Two samples at each of the following sample sizes are selected from the total available sample which consists of 391 subjects' scores: 25, 50, 100, 150, 200. Research questions are stated in Chapter 3.

A descriptive comparison and analysis of obtained results between least squares analysis and ridge regression and across the varying number of subjects per estimation and prediction group is planned. All results are reported by sample for research questions 1-5. Results for research question 6 are reported across the two samples of the same size. First, the mean and standard deviation of each raw score variable and the matrix of correlations among predic-

tors and between each predictor and the criterion are presented. Second, the variance accounted for (R^2), maximum variance inflation factor (VIF), minimum eigenvalue (λmin), standard deviation of the p sample estimated beta coefficients ($\hat{\beta}$), and the estimated squared length of the coefficient vector (est.$\beta'\beta$) are reported and discussed. Third, plots of the ridge trace and estimated squared length of the coefficient vector by k are presented and discussed. For all samples except N=391, the cross-validated R^2 and shrinkage are presented and discussed. Finally, estimated beta coefficients for the k selected by each selection procedure are reported and discussed for each pair of samples at every sample size except N=391. Results are then summarized and discussed by research question between procedures and across sample sizes.

Results

Table 4 reports the mean and standard deviation for each raw score variable while Table 5 contains the matrix of correlations among predictors and between each predictor and the criterion for the data set which consists of 391 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 4

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=391

Variable	Mean	Standard Deviation
Mathematics Concepts*	15.225	6.007
Reading Vocabulary	21.115	9.791
Reading Comprehension	24.180	11.268
Language Spelling	28.790	8.894
Language Mechanics	10.625	5.201
Language Expression	16.275	7.562

* Criterion

Table 5

Matrix of Correlations Among Predictors and Between Each Predictor and the Criterion for N=391

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts*	1.000	.749	.766	.692	.730	.790
Reading Vocabulary	.749	1.000	.821	.676	.676	.796
Reading Comprehension	.766	.821	1.000	.693	.707	.821
Language Spelling	.692	.676	.693	1.000	.626	.703
Language Mechanics	.730	.676	.707	.626	1.000	.782
Language Expression	.790	.796	.821	.703	.782	1.000

* Criterion

Table 6 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures range between .021 and .300. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 5 depicts the decrease in estimated $\beta'\beta$ with increasing k.

Table 6

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=391

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.7094	4.649	.1620	.029	.184
Hoerl & Kennard (1976)	.021	.7095	4.141	.1828	.027	.181
Observation	.050	.7095	3.592	.2120	.024	.178
McDonald & Galarneau (1975)	.100	.7094	2.931	.2620	.021	.173
Hoerl (1962)	.300	.7092	1.703	.4620	.015	.156

Table 7 reports variable identification symbols by predictor variable name to aid in interpretation of the ridge traces which are presented in Figures 4, 7, 10, 13, 16, 19, 22, 25, 28, 31 and 34. The ridge trace for N=391, i.e., Figure 4,shows little variability among the $\hat{\beta}^*_i$ at any level

of k, however, the variability of the p sample beta weights does decrease with increasing levels of k as indicated by the obtained values of $\hat{\beta}$ reported in Table 6 and the ridge trace depicted in Figure 4. Figure 3 shows little variability in the error sum of squares mapped against k. Reduction in R^2 with increasing k as reported in Table 6 is consequently small.

Table 7

Ridge Trace Variable Identification Symbols Reported by Variable Name

Variable Name	Symbol
Reading Vocabulary	Diamond
Reading Comprehension	Triangle
Language Spelling	Circle
Language Mechanics	Box
Language Expression	Plus Sign

Table 8 reports the mean and standard deviation for each raw score variable while Table 9 contains the matrix of correlations among predictors and between each predictor and the criterion for data set A which consists of 25 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

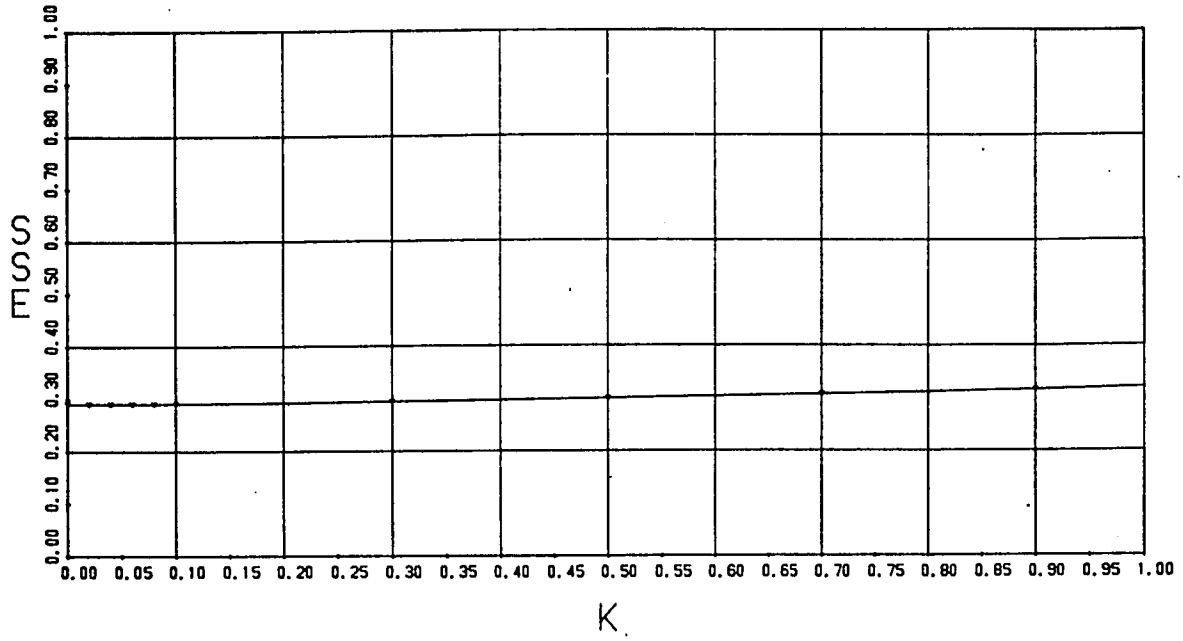

FIGURE 3
ERROR SUM OF SQUARES

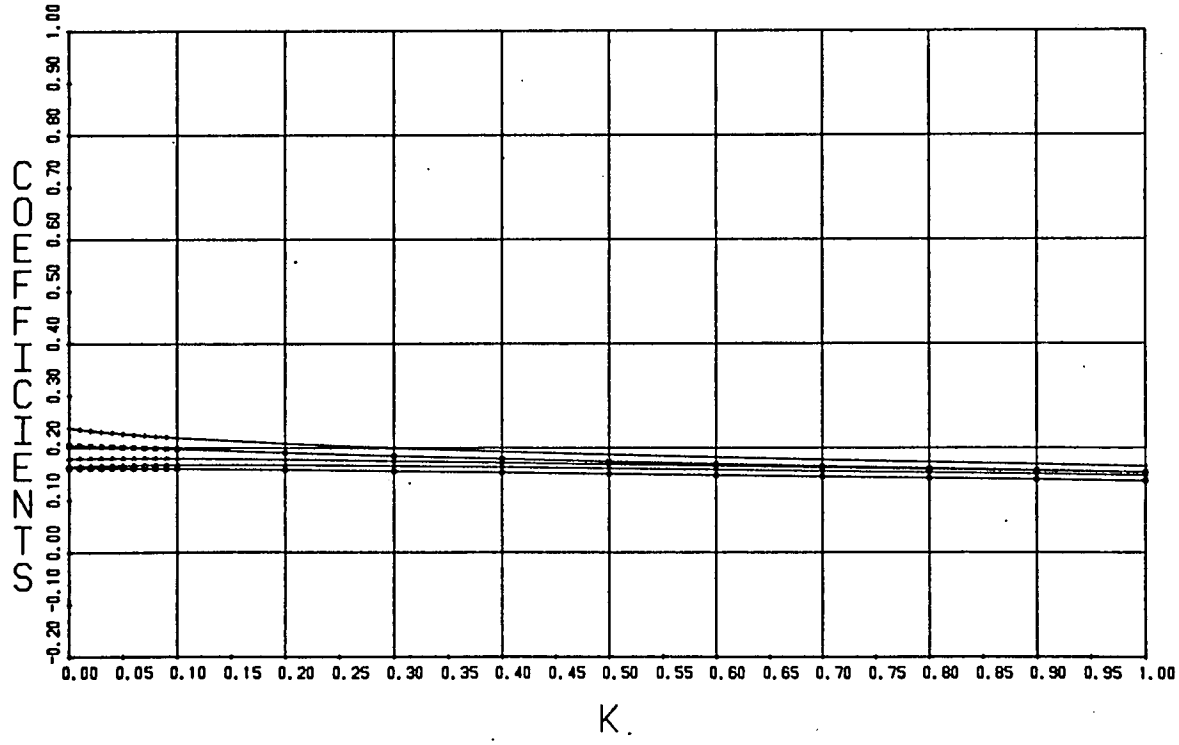

FIGURE 4
SAMPLE SIZE=391 SET A

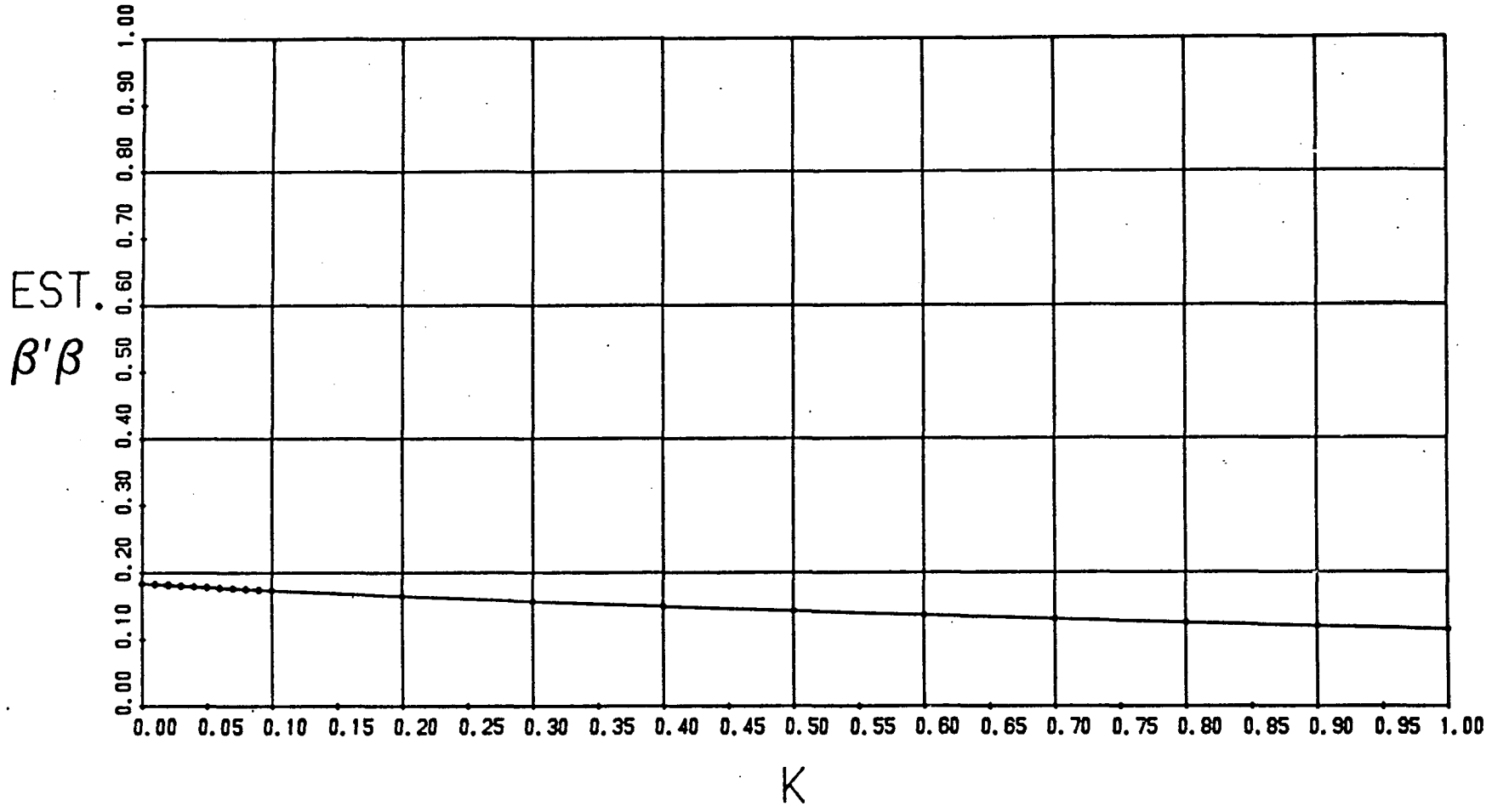

FIGURE 5
SAMPLE SIZE=391 SET A

Table 8

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=25, Set A

Variable	Mean	Standard Deviation
Mathematics Concepts*	14.040	6.043
Reading Vocabulary	18.000	9.839
Reading Comprehension	20.880	10.974
Language Spelling	27.200	9.579
Language Mechanics	8.400	5.052
Language Expression	14.160	7.226

* Criterion

Table 9

Matrix of Correlations Among Predictors and Between Each Predictor and the Criterion for N=25, Set A

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts*	1.000	.792	.779	.696	.820	.724
Reading Vocabulary	.792	1.000	.716	.682	.721	.770
Reading Comprehension	.779	.716	1.000	.520	.760	.757
Language Spelling	.696	.682	.520	1.000	.617	.582
Language Mechanics	.820	.721	.760	.617	1.000	.817
Language Expression	.724	.770	.757	.582	.817	1.000

* Criterion

Table 10 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .200 and .900. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 8 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=.000 is more than twice estimated $\beta'\beta$ at k=.000 for N=391. Estimated $\beta'\beta$ for N=25, Set A approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.300<k<.400$.

Figure 6 shows greater variability in the error sum of squares mapped against k for N=25, Set A than Figure 3 shows for N=391. Reduction in R^2 with increasing k as reported in Table 10 results in an obtained decrease of .021 in the variance accounted for at k=.900 as compared with k=.000. Observation of the ridge trace depicted in Figure 7 shows a negative coefficient for the predictor variable Language Expression which changes sign with $.100<k<.200$. All four ridge k selection procedures selected a $k>.100$. No coefficients have negative signs for N=391.

Table 10

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=25 Set A

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.8010	4.003	.1621	.187	.381
McDonald & Galarneau (1975)	.200	.7932	1.975	.3621	.088	.224
Hoerl & Kennard (1976)	.248	.7915	1.767	.4102	.079	.211
Observation	.700	.7827	0.904	.8621	.038	.152
Hoerl (1962)	.900	.7800	0.747	1.0621	.031	.138

Table 11 reports the results of cross-validation of the regression equation obtained for N=25 Set A. The regression equation was cross-validated with N=25 Set B. Shrinkage decreases with increasing levels of k. The obtained difference between shrinkage at K=.000 and shrinkage at k=.900 is .0244.

Table 11

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=25 Set A

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.8010	.6931	.1079
McDonald & Galarneau (1975)	.200	.7932	.7003	.0929
Hoerl & Kennard (1976)	.248	.7915	.7002	.0913
Observation	.700	.7827	.6974	.0853
Hoerl (1962)	.900	.7800	.6965	.0835

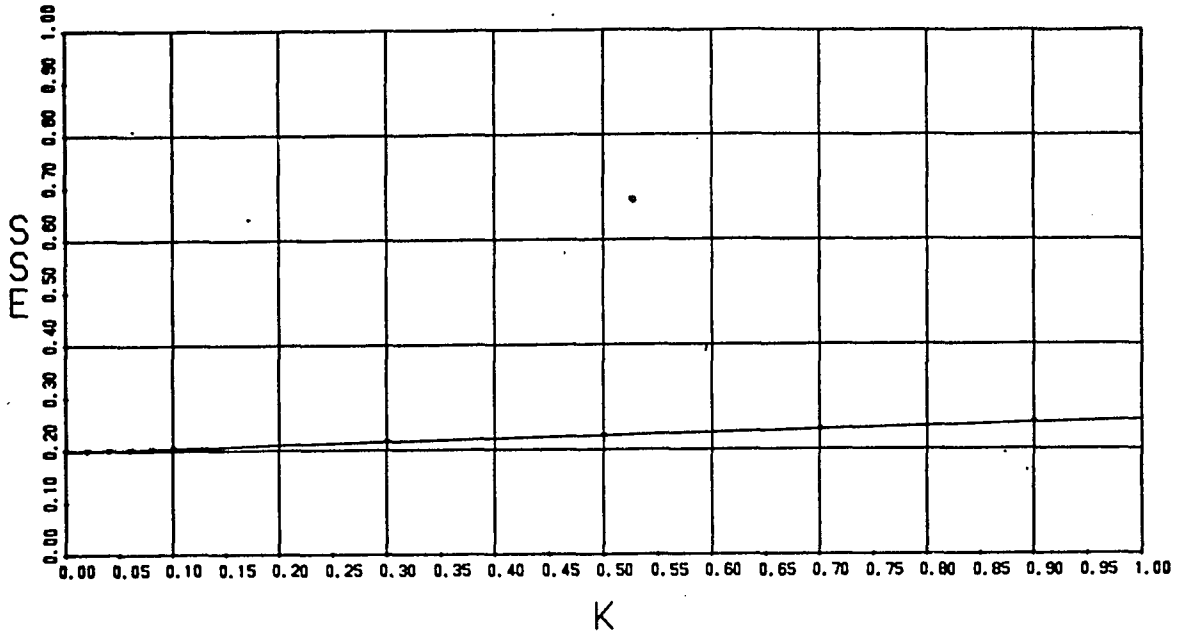

FIGURE 6
ERROR SUM OF SQUARES

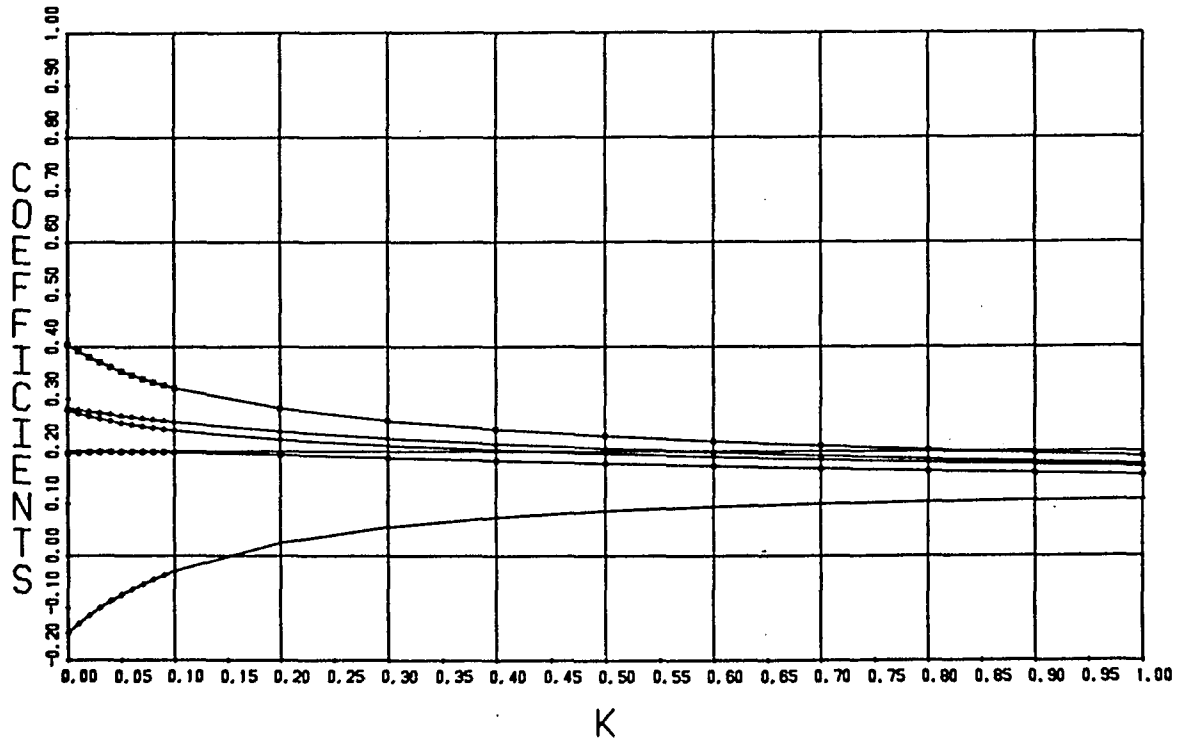

FIGURE 7
SAMPLE SIZE= 25 SET A

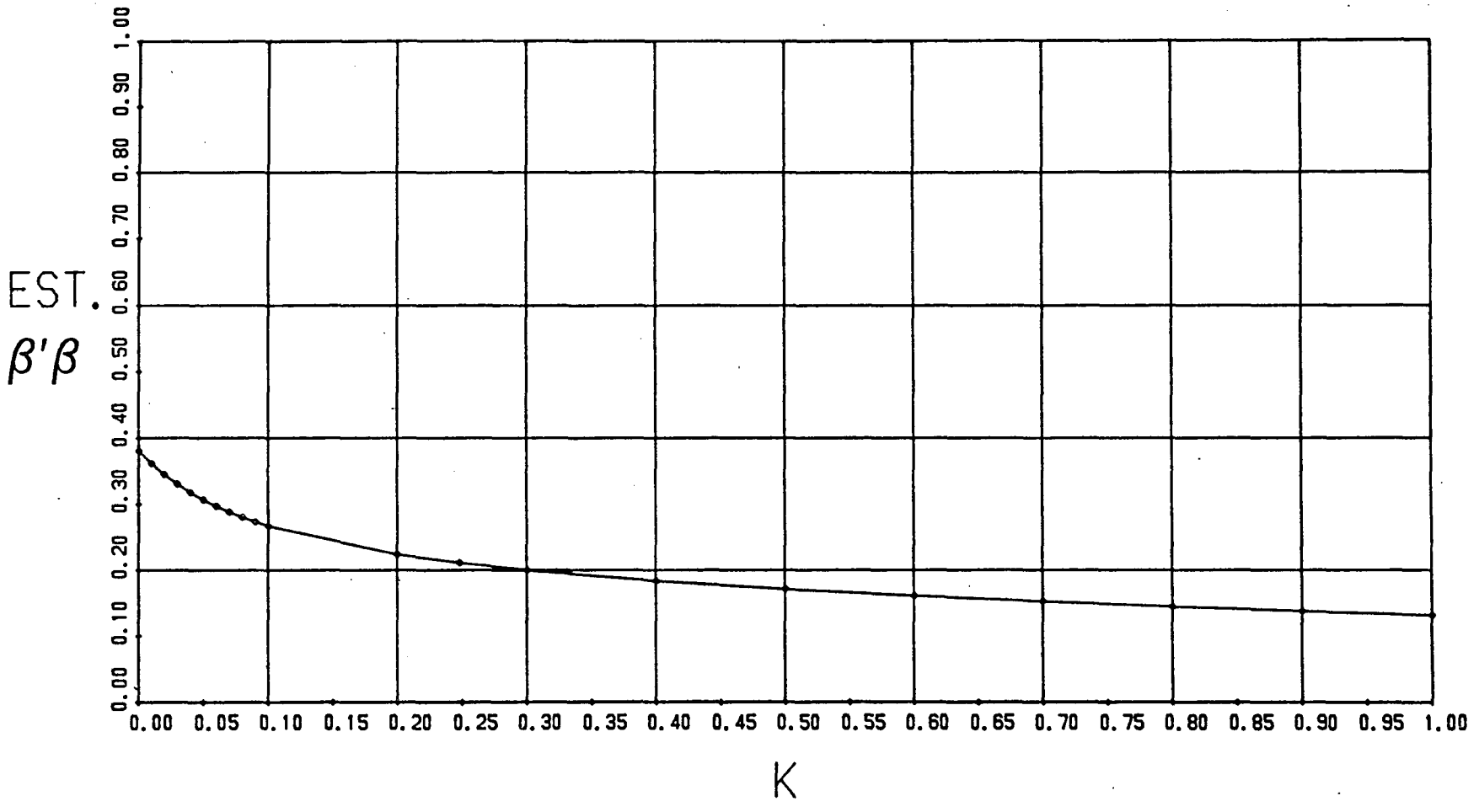

FIGURE 8
SAMPLE SIZE= 25 SET A

Table 12 reports the mean and standard deviation for each raw score variable while Table 13 contains the matrix of correlations among predictors and between each predictor and the criterion for data set B which consists of 25 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 12

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=25 Set B

Variable	Mean	Standard Deviation
Mathematics Concepts*	13.560	5.441
Reading Vocabulary	17.880	8.576
Reading Comprehension	20.480	11.140
Language Spelling	26.760	9.517
Language Mechanics	9.440	5.367
Language Expression	14.480	7.885

* Criterion

Table 13

Matrix of Correlations Among Predictors and Between Each Predictor and the Criterion for N=25, Set A

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts*	1.000	.792	.779	.696	.820	.724
Reading Vocabulary	.792	1.000	.716	.682	.721	.770
Reading Comprehension	.779	.716	1.000	.520	.760	.757
Language Spelling	.696	.682	.520	1.000	.617	.582
Language Mechanics	.820	.721	.760	.617	1.000	.817
Language Expression	.724	.770	.757	.582	.817	1.000

* Criterion

Table 14 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .070 and 1.000. The obtained range of k for N=391 is .279 as compared with an obtained range of .700 for N=25, Set A and .930 for N=25 Set B. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decreases as k increases. Figure 11 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=0 is more than three times estimated $\beta'\beta$ at k=0 for N=391. Estimated $\beta'\beta$ for N=25 Set B approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.400<k<.500$.

Figure 9 shows greater variability in the error sum of squares mapped against k for N=25 Set B than Figure 3 shows for N=391. As reported in Table 14, estimated $\beta'\beta$ is longer and $\hat{S}$ larger at k=0 for N=25, Set B than for N=25 Set A (Table 10). Reduction in R^2 with increasing k as reported in Table 14 shows an obtained decrease of .0476 in the variance accounted for at k=1.000 as compared with k=.000. Observation of the ridge trace depicted in Figure 10 shows negative coefficients for the predictor variables Reading Vocabulary and Language Expression which change signs as $k>.000$.

The variable Reading Vocabulary retains a negative coefficient at the k selected by the McDonald and Galarneau (1975) procedure while changing sign at or before the k selected by the three other ridge k selection procedures. No coefficients have negative signs for N=391.

Table 14

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=25 Set B

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.7625	4.632	.1482	.284	.603
McDonald & Galarneau (1975)	.070	.7572	3.284	.2182	.199	.385
Hoerl & Kennard (1976)	.287	.7378	1.696	.4358	.105	.218
Observation	.700	.7208	0.937	.8482	.055	.147
Hoerl (1962)	1.000	.7149	0.707	1.1482	.040	.123

Table 15 reports the results of cross-validation of the regression equation obtained for N=25 Set B. The regression equation was cross-validated with N=25 Set A. Shrinkage decreases with increasing levels of k. The regression equation calculated on the estimation group accounts for more variance in the prediction group, i.e., N=25 Set A, than in the estimation group for k=.287, .700, and 1.000.

Table 15

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=25 Set B

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.7625	.7161	.0464
McDonald & Galarneau (1975)	.070	.7572	.7472	.0100
Hoerl & Kennard (1976)	.287	.7378	.7681	-.0303
Observation	.700	.7208	.7725	-.0517
Hoerl (1962)	1.000	.7149	.7729	-.0580

Table 16 reports the estimated beta coefficients for each predictor variable by selection procedure together with the zero order correlation between sets of estimated beta coefficients by selection procedure. The correlation between Set A and Set B for least squares analysis is .5436. Correlations between sets of coefficients for the four ridge selection procedures are less than .5436. The obtained correlation between sets of coefficients decreases as k increases. When coefficients are ordered by magnitude, three changes in order occur across Sets A and B given k=.000, whereas four changes in order occur across Sets A and B, given the k selected by each of the ridge k selection procedures.

Table 16

Estimated Beta Coefficients ($\hat{\beta}^*_i$) for each Predictor Variable, k and the Correlation Between Set A and Set B Coefficients (r) Reported by Selection Procedure For N=25 Sets A and B

Selection Procedure	Selected k	Set	$\hat{\beta}^*_1$	$\hat{\beta}^*_2$	$\hat{\beta}^*_3$	$\hat{\beta}^*_4$	$\hat{\beta}^*_5$	r
Least Squares Analysis	.000	A	.278	.282	.196	.404	-.148	
	.000	B	-.105	.103	.405	.644	-.049	.5436
McDonald & Galarneau (1975)	.200	A	.222	.238	.193	.282	.025	
	.070	B	-.022	.100	.350	.500	.038	.4889
Hoerl & Kennard (1976)	.248	A	.216	.231	.191	.270	.041	
	.287	B	.068	.109	.273	.337	.114	.4030
Observation	.700	A	.184	.191	.167	.210	.100	
	.700	B	.106	.116	.212	.244	.134	.3000
Hoerl (1962)	.900	A	.175	.180	.158	.196	.106	
	1.000	B	.111	.114	.187	.212	.133	.2772

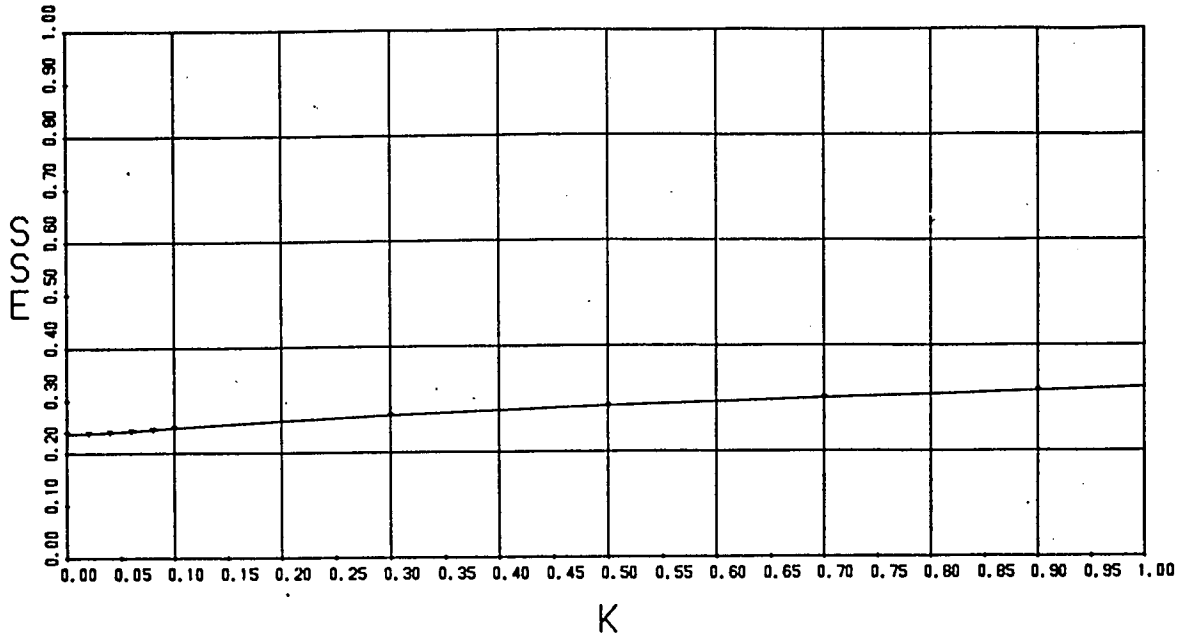

FIGURE 9
ERROR SUM OF SQUARES

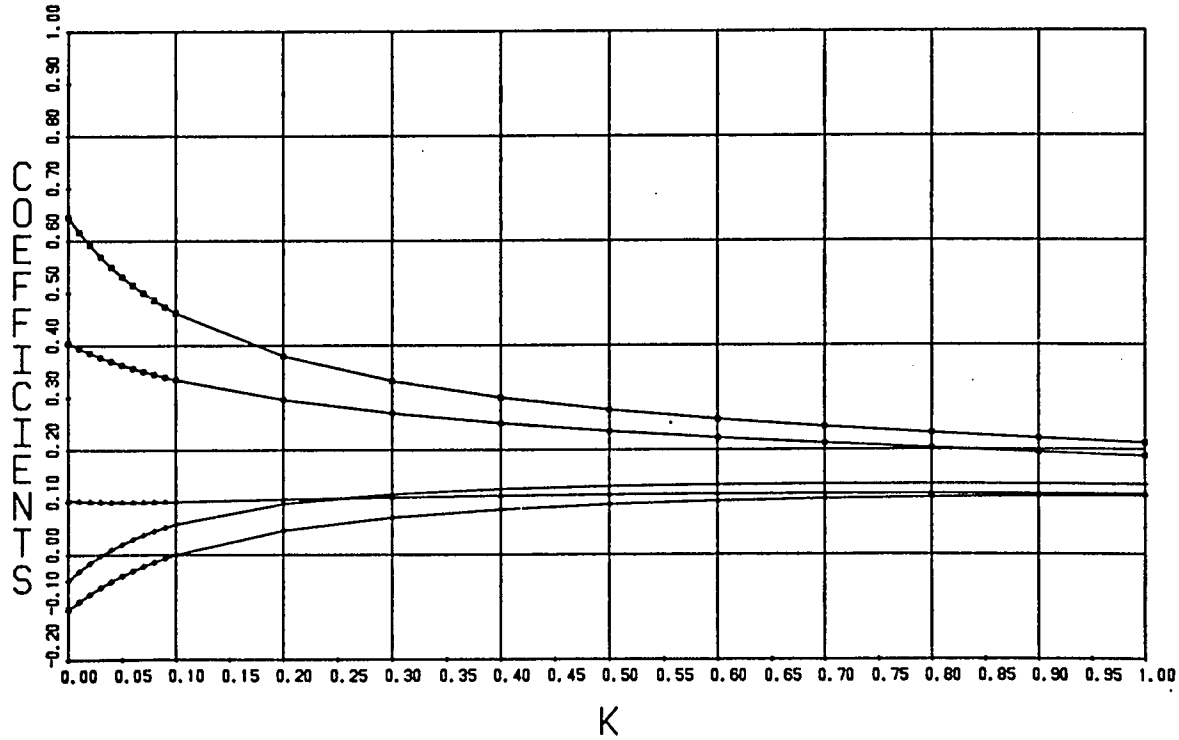

FIGURE 10
SAMPLE SIZE= 25 SET B

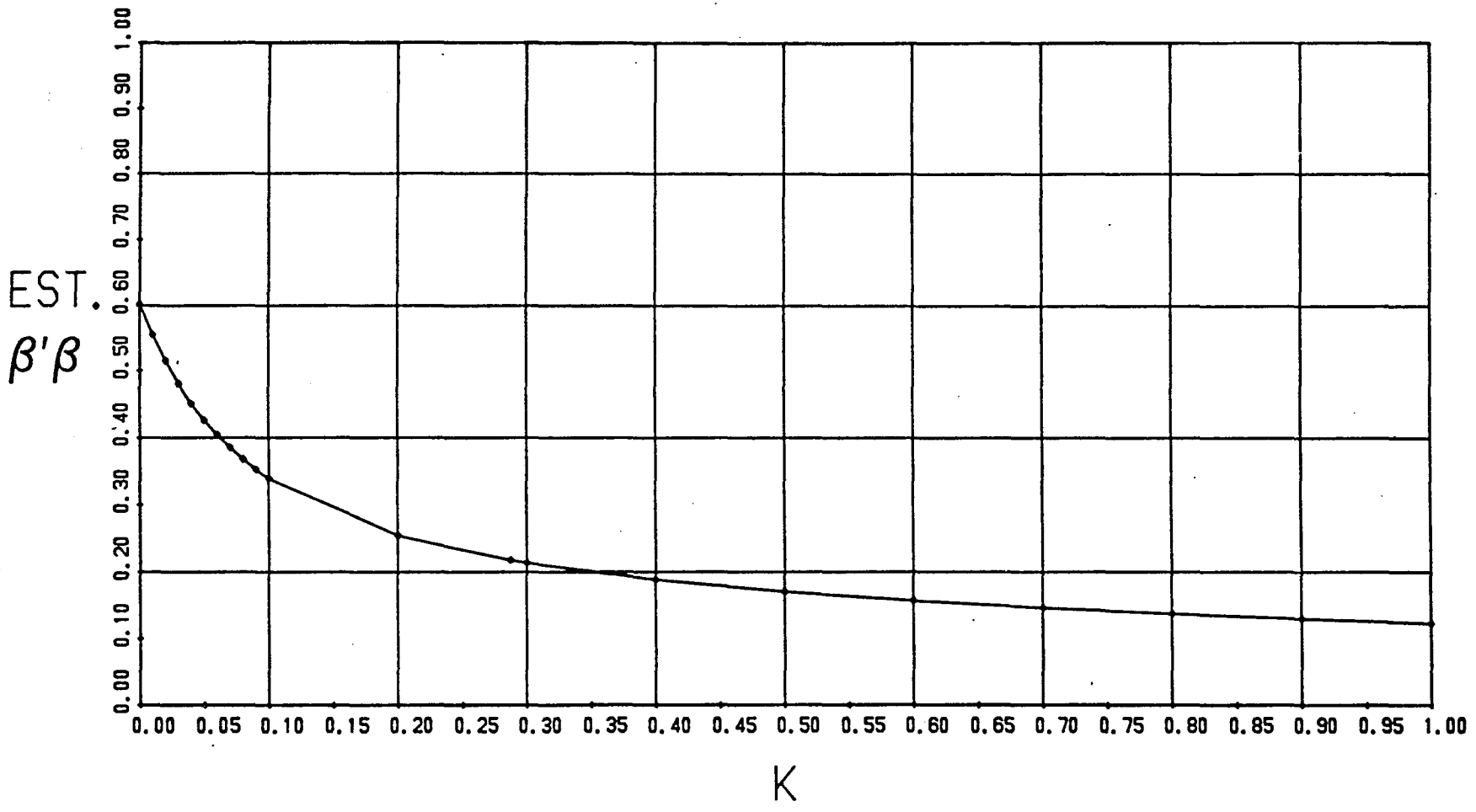

FIGURE 11
SAMPLE SIZE= 25 SET B

Table 17 reports the mean and standard deviation for each raw score variable while Table 18 contains the matrix of correlations among predictors and between each predictor and the criterion for data set A which consists of 50 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 17

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=50 Set A

Variable	Mean	Standard Deviation
Mathematics Concepts*	15.480	5.957
Reading Vocabulary	20.540	9.888
Reading Comprehension	24.660	10.935
Language Spelling	29.340	9.715
Language Mechanics	11.580	5.793
Language Expression	16.200	7.416

* Criterion

Table 18

Matrix of Correlations Among Predictors and Between Each Predictor and the Criterion for N=50 Set A

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts *	1.000	.794	.720	.791	.854	.837
Reading Vocabulary	.794	1.000	.718	.728	.729	.800
Reading Comprehension	.720	.718	1.000	.688	.767	.846
Language Spelling	.791	.728	.688	1.000	.701	.736
Language Mechanics	.854	.729	.767	.701	1.000	.854
Language Expression	.837	.800	.846	.736	.854	1.000

* Criterion

Table 19 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .067 and .900. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 14 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=0 is approximately twice estimated $\beta'\beta$ at k=0 for N=391. Estimated $\beta'\beta$ for N=50, Set A approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.400<k<.500$.

Figure 6 shows little variability in the error sum of squares mapped against k for N=50 Set A. Reduction in R^2 with increasing k as reported in Table 19 results in an obtained decrease of .021 in the variance accounted for at k=.900 as compared with k=.000. Observation of the ridge trace depicted in Figure 13 shows a negative coefficient for the predictor variable Reading Comprehension which changes sign with $.150<k<.200$. The ridge k selection procedure proposed by Hoerl (1962) and the observation procedure result in the selection of a $k>.150$ whereas the Hoerl and Kennard (1976) and McDonald and Galarneau (1975) procedures selected $k<.150$. No coefficients have negative signs for N=391.

Table 19

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=50 Set A

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.8307	6.474	.1140	.177	.354
Hoerl & Kennard (1976)	.067	.8291	4.133	.1812	.138	.286
McDonald & Galarneau (1975)	.080	.8286	3.870	.1940	.133	.277
Observation	.600	.8138	1.114	.7140	.051	.162
Hoerl (1962)	.900	.8100	.800	1.0140	.037	.139

Table 20 reports the results of cross-validation of the regression equation obtained for N=50 Set A. The regression equation was cross-validated with N=50 Set B. Shrinkage decreases with increasing k. The obtained difference between shrinkage at k=.000 and shrinkage at k=.900 is .0156.

Table 20

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=50 Set A

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.8307	.7490	.0817
Hoerl & Kennard (1976)	.067	.8291	.7516	.0775
McDonald & Galarneau (1975)	.080	.8286	.7517	.0769
Observation	.600	.8138	.7463	.0675
Hoerl (1962)	.900	.8100	.7441	.0659

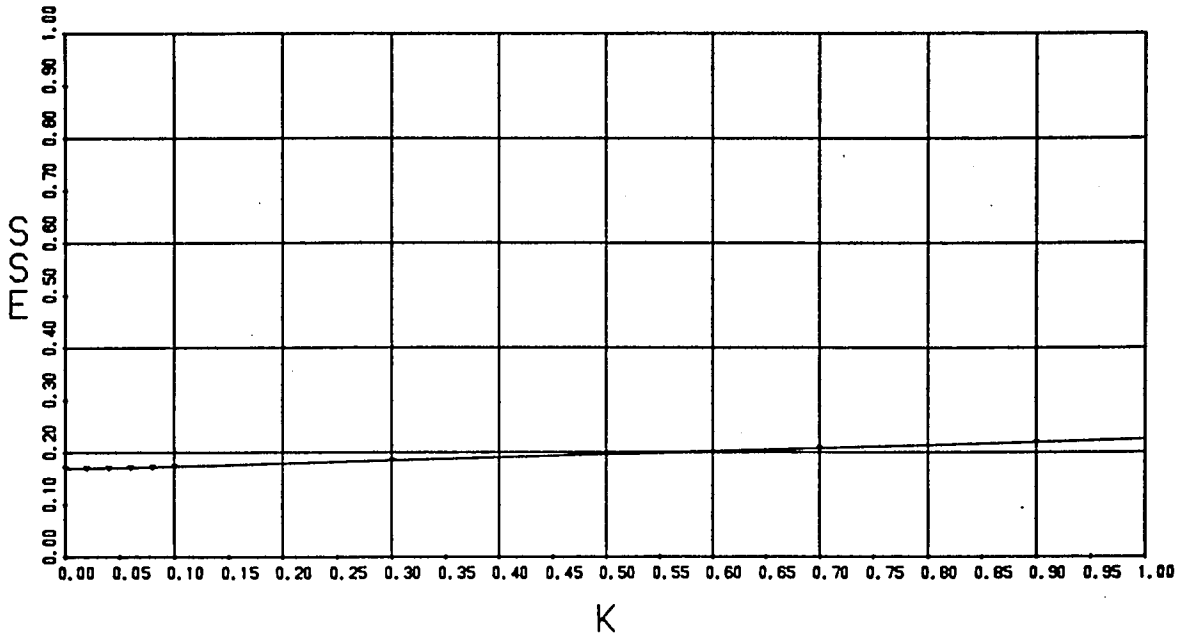

FIGURE 12
ERROR SUM OF SQUARES

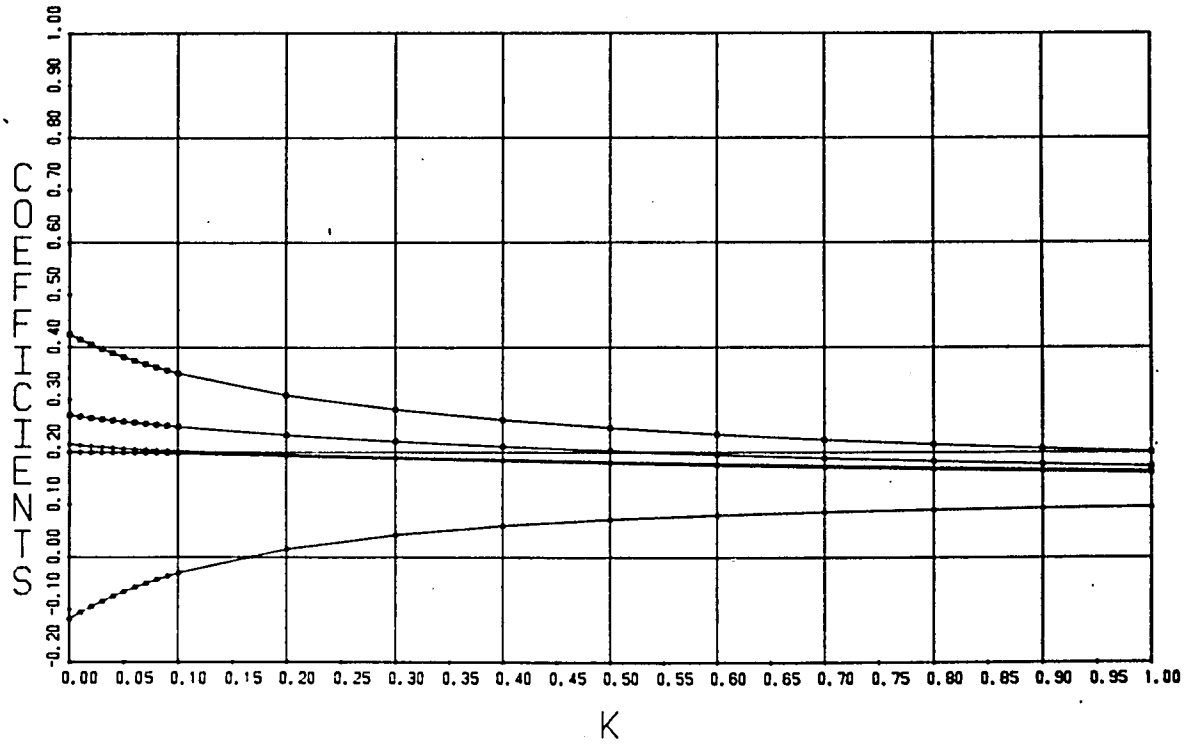

FIGURE 13
SAMPLE SIZE= 50 SET A

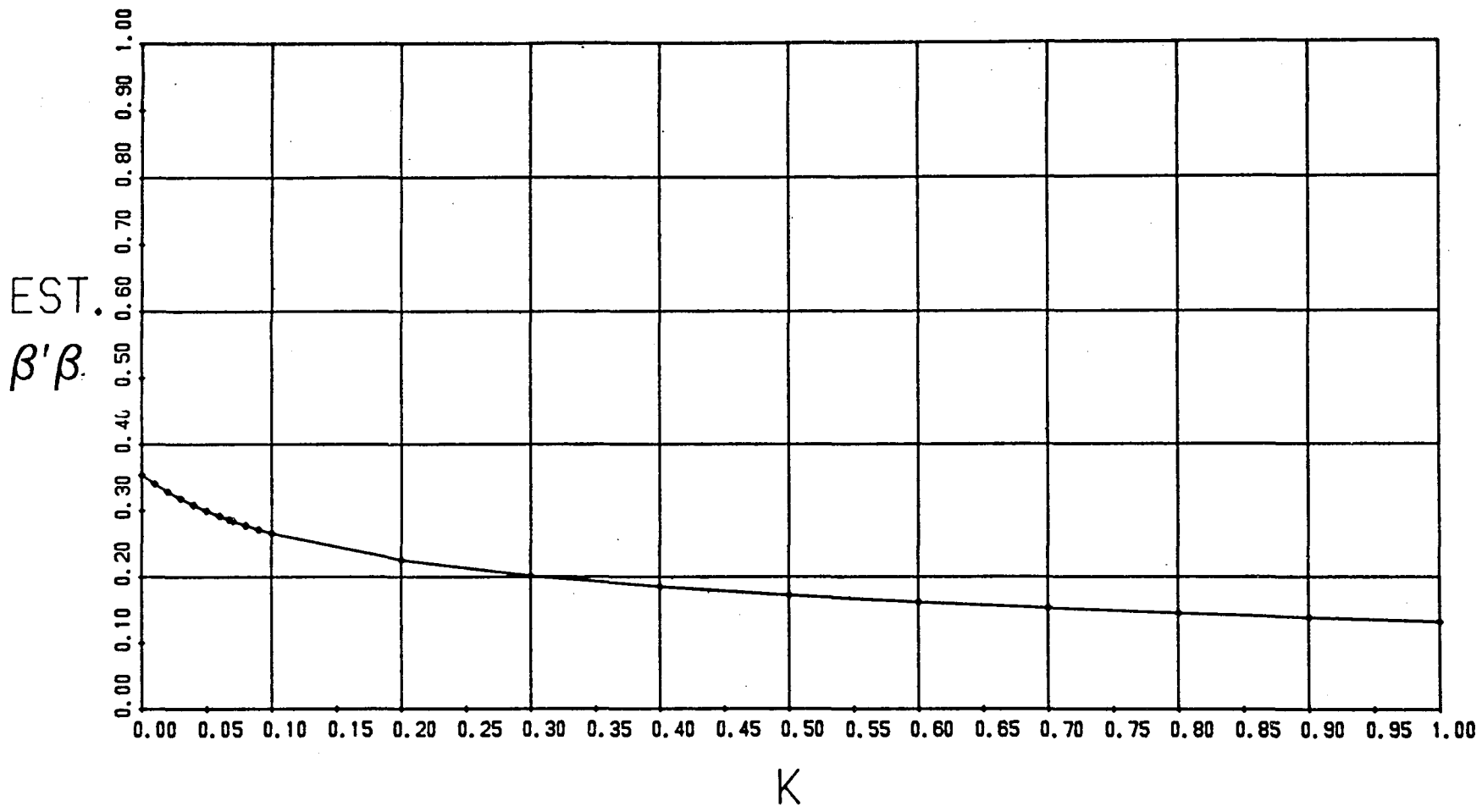

FIGURE 14
SAMPLE SIZE= 50 SET A

Table 21 reports the mean and standard deviation for each raw score variable while Table 22 contains the matrix of correlations among predictors and between each predictor and the criterion for data set B which consists of 50 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 21

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=50 Set B

Variable	Mean	Standard Deviation
Mathematics Concepts *	15.920	5.314
Reading Vocabulary	21.380	9.376
Reading Comprehension	24.700	11.416
Language Spelling	30.300	9.345
Language Mechanics	10.820	4.877
Language Expression	16.760	6.834

* Criterion

Table 22

Matrix of Correlations Among Predictors and Between Each Predictor and the Criterion for N=50 Set B

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts *	1.000	.704	.753	.791	.770	.832
Reading Vocabulary	.704	1.000	.807	.704	.685	.777
Reading Comprehension	.753	.807	1.000	.769	.756	.843
Language Spelling	.791	.704	.769	1.000	.685	.725
Language Mechanics	.770	.685	.756	.685	1.000	.816
Language Expression	.832	.777	.843	.725	.816	1.000

* Criterion

Table 23 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the five ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .050 and 1.000. The obtained range of k for N=391 is .279 as compared with an obtained range of .833 for N=50 Set A and .950 for N=50 Set B. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 17 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=.000 is more than twice estimated $\beta'\beta$ at k=.000 for N=391. Estimated $\beta'\beta$ for N=50, Set B approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.300<k<.400$.

Figure 15 shows a monotonic increase in the error sum of squares as $k>.000$. Reduction in R^2 with increasing k as reported in Table 23 shows an obtained decrease of .0254 in the variance accounted for at k=1.000 as compared with k=.000. Observation of the ridge trace depicted in Figure 16 shows a negative coefficient for the predictor variable Reading Comprehension which changes sign with $.080<k<.100$. The ridge k selection procedures proposed by Hoerl and Kennard (1976) and McDonald and Galarneau (1975) result in the selection of a $k<.080$, whereas the observation and Hoerl (1962) procedures selected $k>.080$. No coefficients have negative signs for N=391.

Table 23

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=50 Set B

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.7772	4.975	.1391	.208	.401
McDonald & Galarneau (1975)	.050	.7755	3.761	.1891	.161	.310
Hoerl & Kennard (1976)	.095	.7730	3.104	.2339	.135	.267
Observation	.600	.7566	1.072	.7391	.050	.151
Hoerl (1962)	1.000	.7518	.716	1.1391	.034	.123

Table 24 reports the results of cross-validation of the regression equation obtained for N=50 Set B. The regression equation was cross-validated with N=50 Set A. At every level of k including k=.000, the cross-validated R^2 is larger than the R^2 obtained on the estimation group, i.e., N=50 Set B. The increase in R^2 upon cross-validation is greater for the ridge k selection procedures than for least squares analysis and rises monotonically as k increases. The increase in R^2 is 2.73% at k=.000 and 5.22% at k=1.000.

Table 24

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=50 Set B

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.7772	.8045	-.0273
McDonald & Galarneau (1975)	.050	.7755	.8082	-.0327
Hoerl & Kennard (1976)	.095	.7730	.8091	-.0361
Observation	.600	.7566	.8059	-.0493
Hoerl (1962)	1.000	.7518	.8040	-.0522

Table 25 reports the estimated beta coefficients for each predictor variable by selection procedure together with the zero order correlation between sets of estimated beta coefficients by selection procedure. The correlation between Set A and Set B for least squares analysis is .5495. The correlations between Set A and Set B at the Hoerl and Kennard (1976) and McDonald and Galarneau (1975) selected k levels are slightly larger with increases of .0064 and .0057 respectively when compared with the least squares r whereas the obtained correlations between Sets A and B at the observation and Hoerl (1962) selected k's are slightly smaller than .5495, i.e., .5331 and .5256 respectively. When coefficients are ordered by magnitude two changes in order occur across Sets A and B for k=0 and at the k selected by the Hoerl and Kennard (1976), McDonald and Galarneau (1975), and observation ridge k selection procedures. Four changes in order occur across Sets A and B at the k selected by the Hoerl (1962) procedure.

Table 25

Estimated Beta Coefficients ($\hat{\beta}^*_i$) for each Predictor Variable, k and the Correlation Between Set A and Set B Coefficients (r) Reported by Selection Procedure For N=50 Sets A and B

Selection Procedure	Selected k	Set	$\hat{\beta}^*_1$	$\hat{\beta}^*_2$	$\hat{\beta}^*_3$	$\hat{\beta}^*_4$	$\hat{\beta}^*_5$	r
Least Squares Analysis	.000	A	.200	-.118	.270	.425	.215	
	.000	B	.018	-.081	.382	.178	.465	.5495
Hoerl & Kennard (1976)	.067	A	.199	-.052	.255	.370	.205	
	.095	B	.051	.020	.321	.195	.350	.5559
McDonald & Galarneau (1975)	.080	A	.199	-.043	.252	.362	.204	
	.050	B	.037	-.015	.345	.191	.391	.5552
Observation	.600	A	.176	.081	.195	.234	.178	
	.600	B	.104	.114	.218	.177	.221	.5331
Hoerl (1962)	.900	A	.164	.094	.178	.207	.168	
	1.000	B	.109	.120	.187	.161	.190	.5256

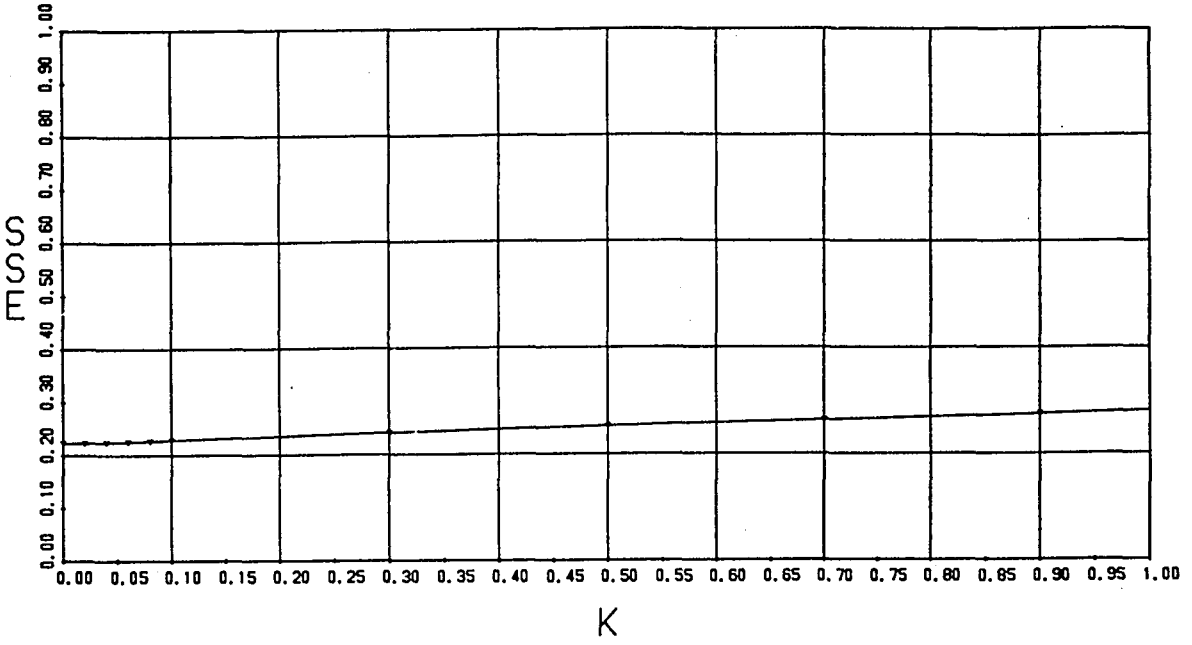

FIGURE 15
ERROR SUM OF SQUARES

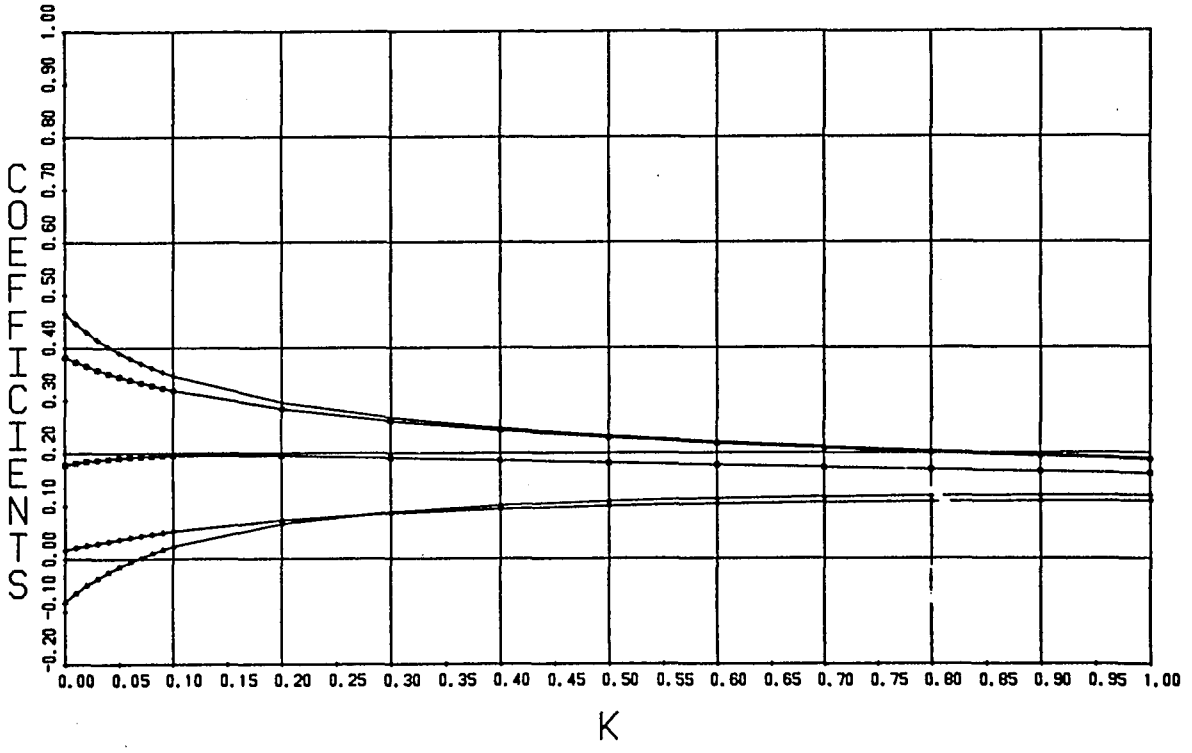

FIGURE 16
SAMPLE SIZE= 50 SET B

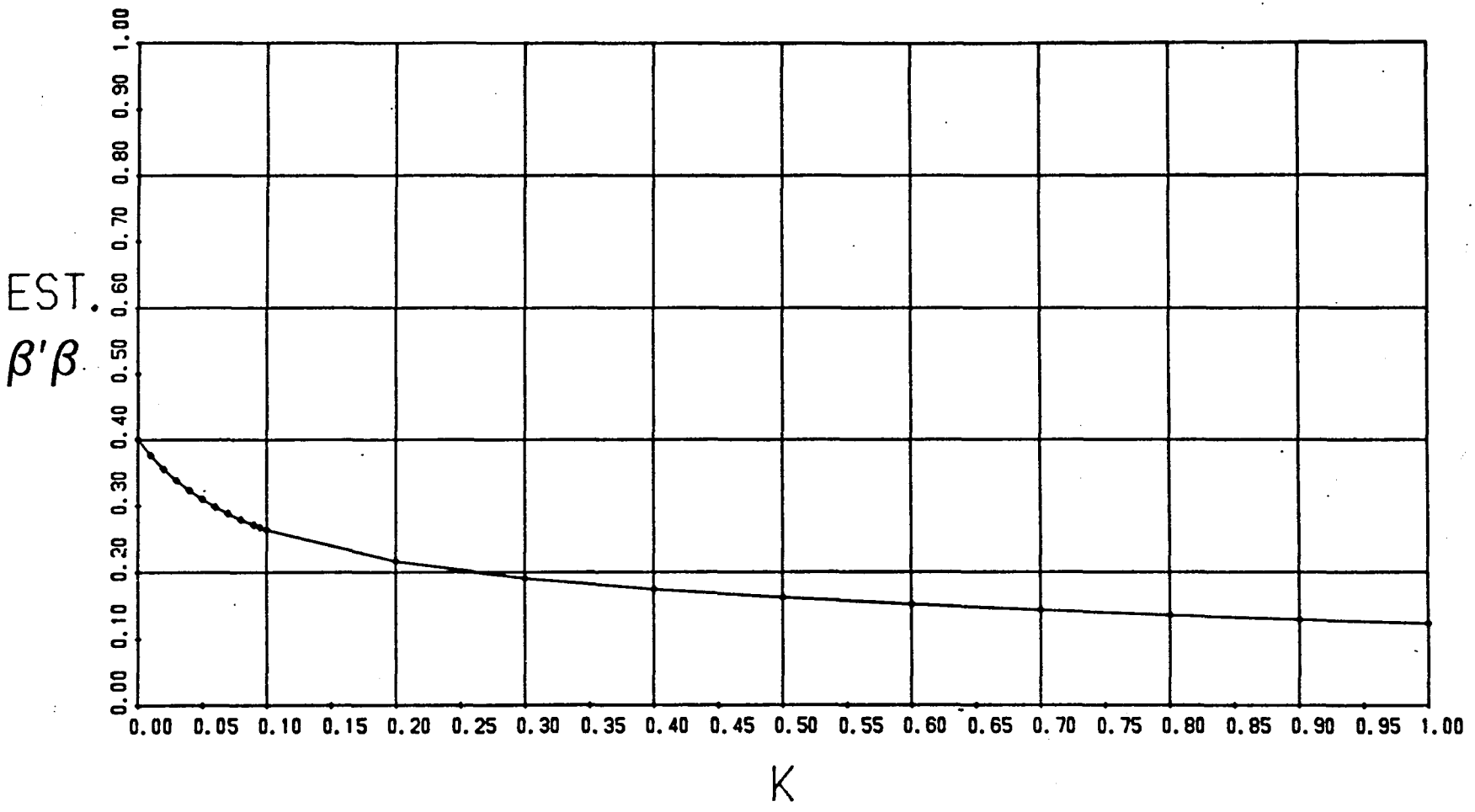

FIGURE 17
SAMPLE SIZE= 50 SET B

Table 26 reports the mean and standard deviation for each raw score variable while Table 27 contains the matrix of correlations among predictors and between each predictor and the criterion for data Set A which consists of 100 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 26

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=100 Set A

Variable	Mean	Standard Deviation
Mathematics Concepts *	16.200	5.483
Reading Vocabulary	22.010	9.613
Reading Comprehension	24.390	11.092
Language Spelling	29.940	9.066
Language Mechanics	11.060	5.073
Language Expression	17.180	7.415

* Criterion

Table 27

Matrix of Correlations Among Predictors and
Between Each Predictor and the Criterion for N=100 Set A

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts *	1.000	.748	.761	.716	.710	.798
Reading Vocabulary	.748	1.000	.829	.688	.672	.744
Reading Comprehension	.761	.829	1.000	.715	.660	.773
Language Spelling	.716	.688	.715	1.000	.668	.701
Language Mechanics	.710	.672	.660	.668	1.000	.789
Language Expression	.798	.744	.773	.701	.789	1.000

* Criterion

Table 28 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .077 and .600. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 20 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=.000 is .216 as compared with .184 for N=391. Estimated $\beta'\beta$ for N=100 Set A approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.100<k<.200$.

Figure 18 shows little variability in the error sum of squares mapped against k for N=100 Set A. Reduction in R^2 with increasing k as reported in Table 28 results in an obtained decrease of .0031 in the variance accounted for at k=.600 as compared with k=.000. No coefficients change sign as $k>0$ and no predictor variable has a negative coefficient as can be seen in Figure 19. Negative signs which change as $k>0$ are noted at N=25 and N=50. No coefficients have negative signs for N=391.

Table 28

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=100 Set A

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.7221	4.107	.1545	.083	.216
Hoerl & Kennard (1976)	.077	.7217	2.886	.2312	.060	.193
McDonald & Galarneau (1975)	.300	.7201	1.592	.4545	.034	.163
Hoerl (1962)	.600	.7190	1.015	.7545	.022	.140
Observation	.600	.7190	1.015	.7545	.022	.140

Table 29 reports the results of cross-validation of the regression equation obtained for N=100, Set A. The regression equation was cross-validated with N=100 Set B. Shrinkage decreases with increasing k. The obtained difference between shrinkage at k=.000 and shrinkage at k=.600 is .010.

Table 29

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=100 Set A

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.7221	.6831	.039
Hoerl & Kennard (1976)	.077	.7217	.6863	.035
McDonald & Galarneau (1975)	.300	.7201	.6889	.031
Hoerl (1962)	.600	.7190	.6898	.029
Observation	.600	.7190	.6898	.029

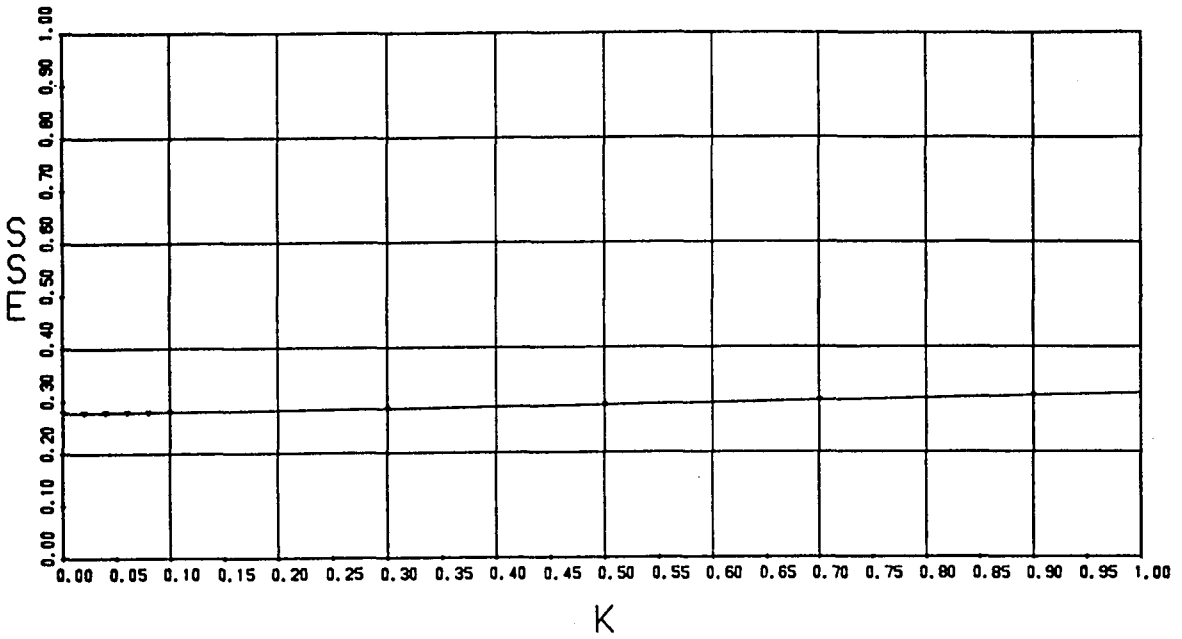

FIGURE 18
ERROR SUM OF SQUARES

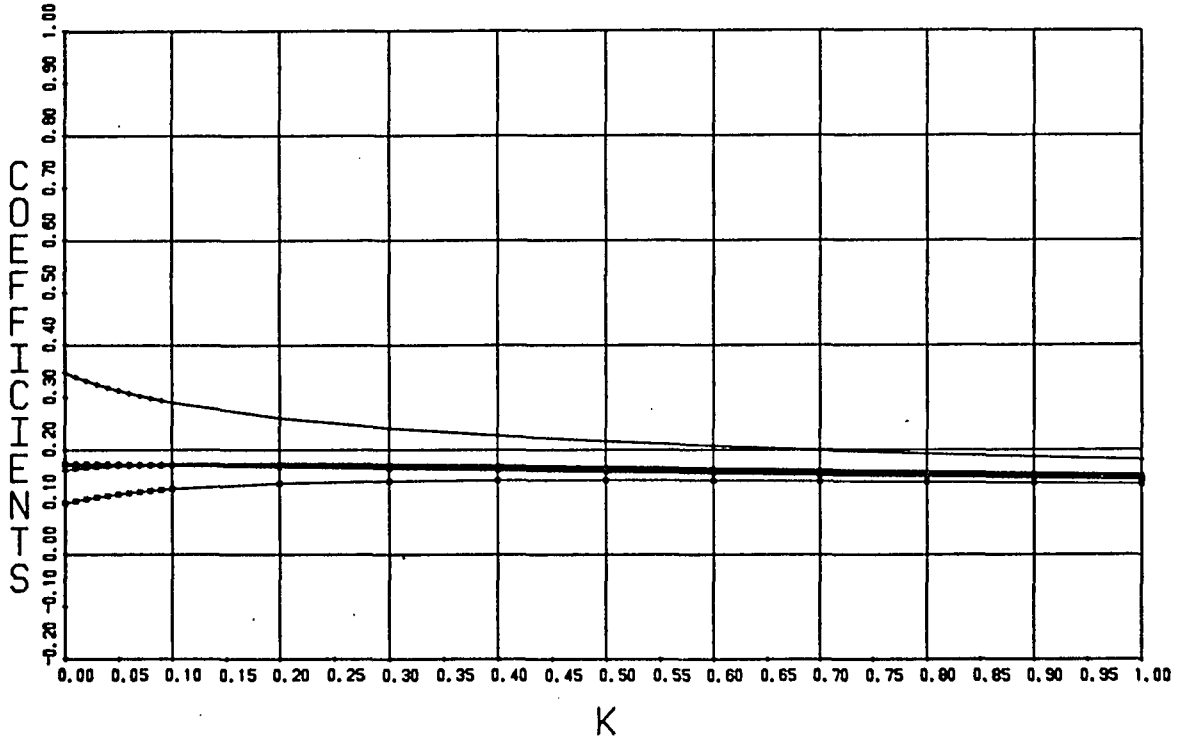

FIGURE 19
SAMPLE SIZE=100 SET A

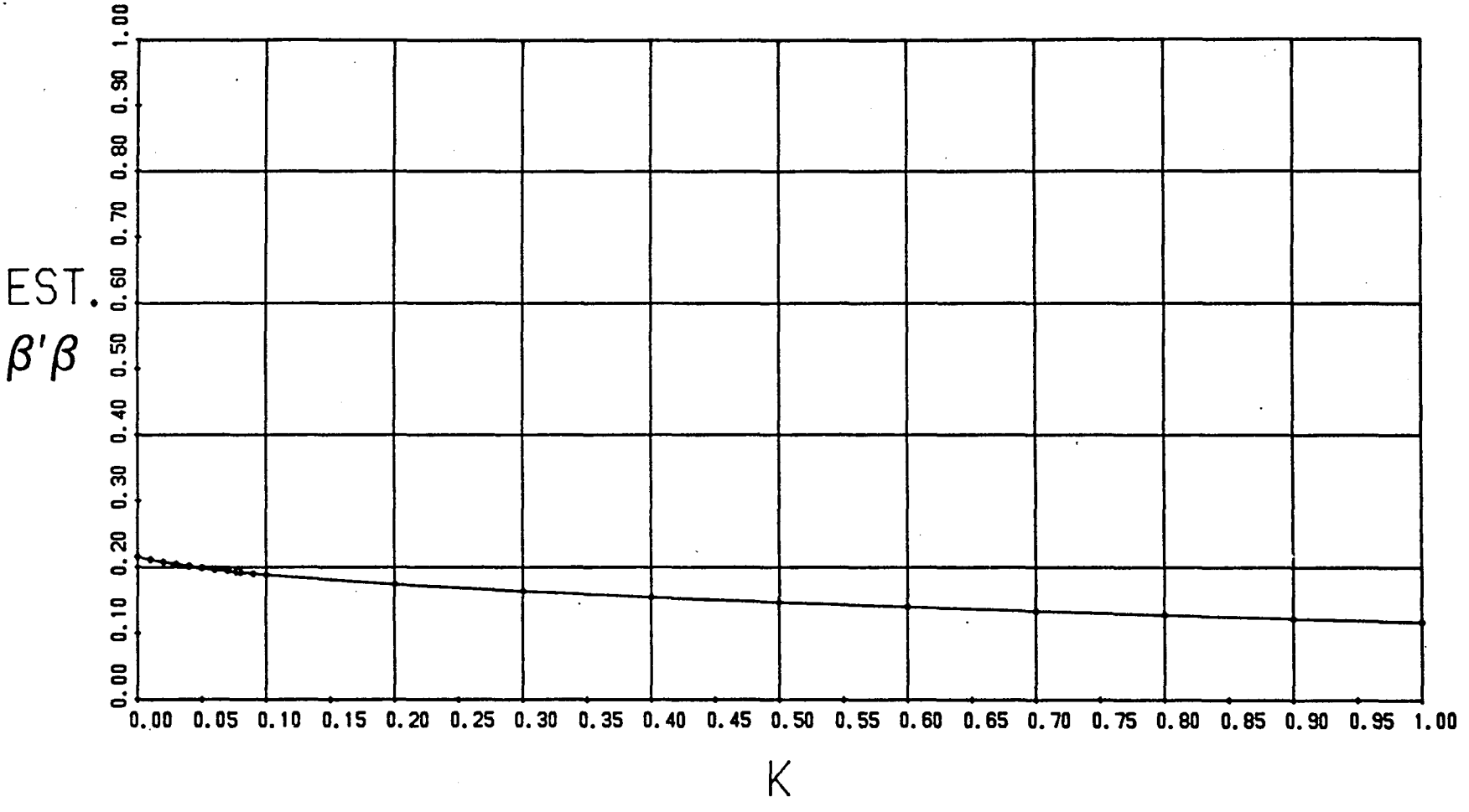

FIGURE 20
SAMPLE SIZE=100 SET A

Table 30 reports the mean and standard deviation for each raw score variable while Table 31 contains the matrix of correlations among predictors and between each predictor and the criterion for data set B which consists of 100 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 30

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=100 Set B

Variable	Mean	Standard Deviation
Mathematics Concepts *	16.180	5.428
Reading Vocabulary	22.100	9.869
Reading Comprehension	25.040	10.855
Language Spelling	30.420	8.123
Language Mechanics	10.910	5.010
Language Expression	17.150	7.192

* Criterion

Table 31

Matrix of Correlations Among Predictors and
Between Each Predictor and the Criterion for N=100 Set B

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts *	1.000	.763	.741	.681	.696	.752
Reading Vocabulary	.763	1.000	.819	.694	.679	.782
Reading Comprehension	.741	.819	1.000	.643	.681	.791
Language Spelling	.681	.694	.643	1.000	.539	.650
Language Mechanics	.696	.679	.681	.539	1.000	.765
Language Expression	.752	.782	.791	.650	.765	1.000

* Criterion

Table 32 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .050 and .600. The obtained range of k for N=391 is .279 as compared with an obtained range of .523 for N=100 Set A and .550 for N=100 Set B. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 23 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=0 is .186. For N=391, the estimated squared length of the coefficient vector at k=0 is .184. Estimated $\beta'\beta$ for N=100 Set B approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.010<k<.020$.

Figure 21 shows a slow monotonic increase in the error sum of squares as $k>.000$. Reduction in R^2 with increasing k as reported in Table 32 shows an obtained decrease of .0006 in the variance accounted for at k=.600 as compared with k=.000. Observation of the ridge trace depicted in Figure 22 shows little variation among estimated coefficients. Coefficients converge as $k>.000$. No predictor variable has a negative sign. Negative signs which change as $k>0$ are noted at N=25 and N=50. No coefficients have negative signs for N=391.

Table 32

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=100 Set B

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.6926	4.022	.1737	.028	.186
Observation	.050	.6926	3.206	.2237	.022	.180
Hoerl & Kennard (1976)	.093	.6925	2.729	.2672	.019	.175
Hoerl (1962)	.400	.6922	1.350	.5737	.009	.149
McDonald & Galarneau (1975)	.600	.6920	1.022	.7737	.007	.135

Table 33 reports the results of cross-validation of the regression equation obtained for N=100 Set B. The regression equation was cross-validated with N=100 Set A. At every level of k including k=.000, the cross-validated R^2 was larger than the R^2 obtained on the estimation group, i.e., N=100, Set B. The increase in R^2 upon cross-validation is greater for the ridge k selection procedures than for least squares analysis and rises monotonically as k increases. The increase in R^2 is 1.96% at k=.000 and 2.34% at k=.600.

Table 33

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=100 Set B

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.6926	.7122	-.0196
Observation	.050	.6926	.7131	-.0205
Hoerl & Kennard (1976)	.093	.6925	.7136	-.0211
Hoerl (1962)	.400	.6922	.7150	-.0228
McDonald & Galarneau (1975)	.600	.6920	.7154	-.0234

Table 34 reports the estimated beta coefficients for each predictor variable together with the zero order correlation between sets of estimated beta coefficients by selection procedure. All obtained correlations between sets of estimated β are negative. Due to the convergence of estimated coefficients as $k>0$, slight changes in estimated β produce significant differences in the order by magnitude of estimated β when Sets A and B are compared.

Table 34

Estimated Beta Coefficients ($\hat{\beta}^*_i$) for each Predictor Variable, k and the Correlation Between Set A and Set B Coefficients (r) Reported by Selection Procedure For N=100 Sets A and B

Selection Procedure	Selected k	Set	$\hat{\beta}^*_1$	$\hat{\beta}^*_2$	$\hat{\beta}^*_3$	$\hat{\beta}^*_4$	$\hat{\beta}^*_5$	r
Least Squares Analysis	.000	A	.170	.162	.174	.098	.347	
	.000	B	.229	.155	.211	.195	.164	-.4038
Hoerl & Kennard (1976)	.077	A	.170	.171	.171	.121	.300	
	.093	B	.213	.162	.200	.185	.169	-.2700
McDonald & Galarneau (1975)	.300	A	.168	.173	.165	.140	.242	
	.600	B	.178	.158	.166	.159	.162	-.0588
Hoerl (1962)	.600	A	.160	.165	.155	.141	.207	
	.400	B	.188	.162	.177	.167	.167	-.2141
Observation	.600	A	.160	.165	.155	.141	.207	
	.050	B	.220	.160	.205	.189	.168	-.4819

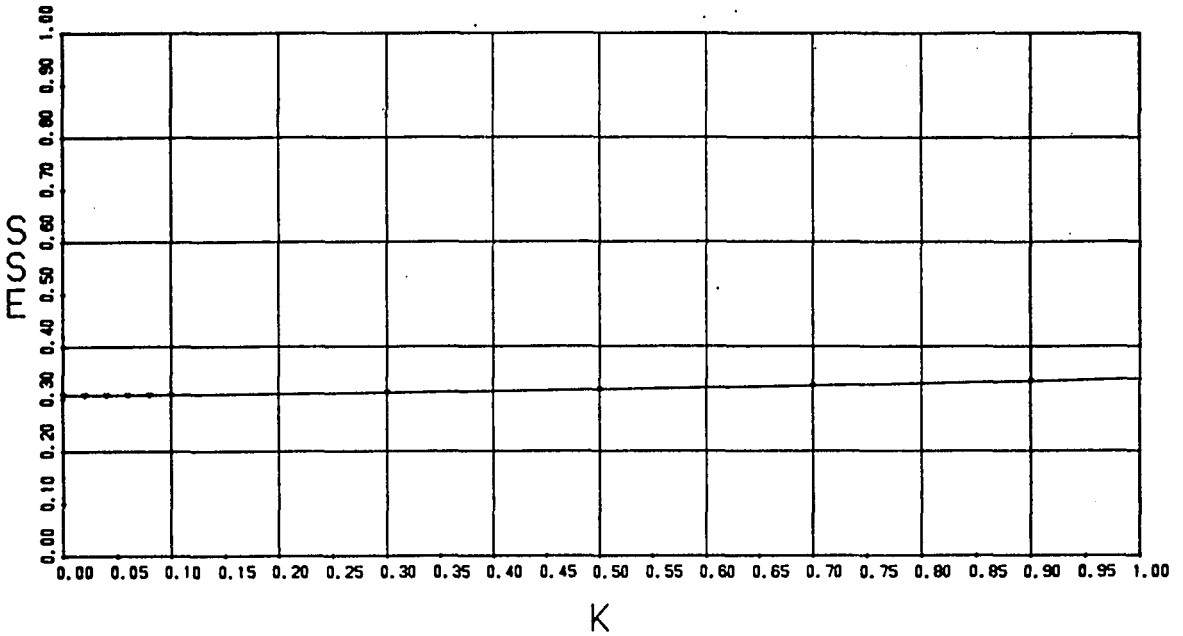

FIGURE 21
ERROR SUM OF SQUARES

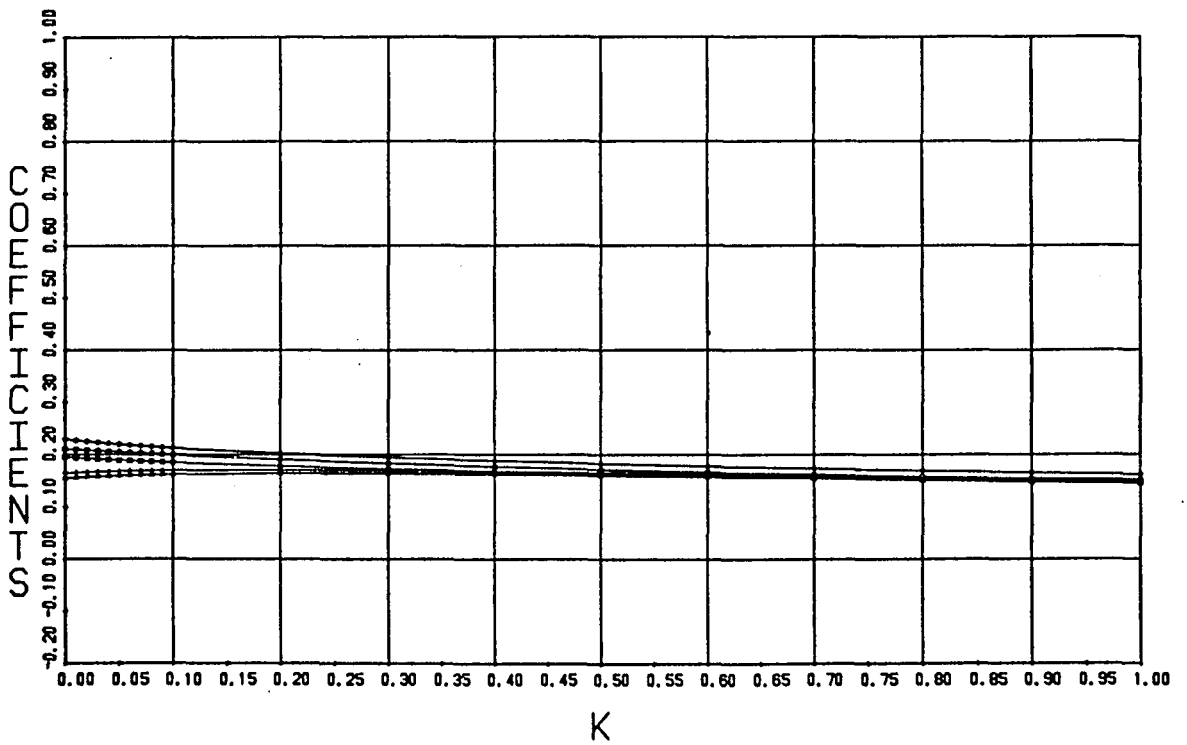

FIGURE 22
SAMPLE SIZE=100 SET B

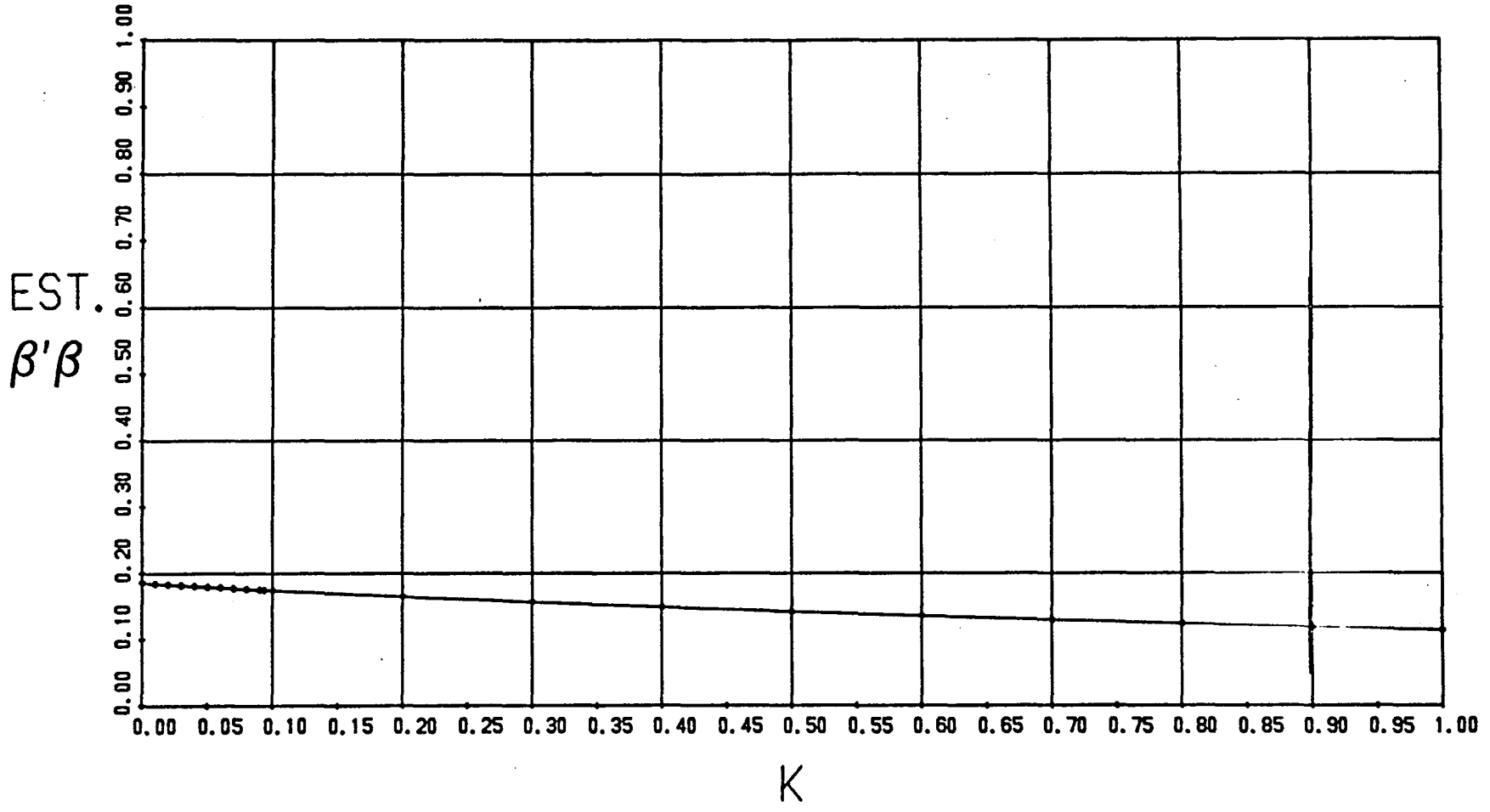

FIGURE 23
SAMPLE SIZE=100 SET B

Table 35 reports the mean and standard deviation for each raw score variable while Table 36 contains the matrix of correlations among predictors and between each predictor and the criterion for data Set A which consists of 150 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 35

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=150 Set A

Variable	Mean	Standard Deviation
Mathematics Concepts *	15.973	5.833
Reading Vocabulary	21.960	9.762
Reading Comprehension	25.033	11.497
Language Spelling	29.693	8.764
Language Mechanics	11.067	5.157
Language Expression	17.013	7.439

* Criterion

Table 36

Matrix of Correlations Among Predictors and
Between Each Predictor and the Criterion for N=150 Set A

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts *	1.000	.745	.803	.679	.746	.783
Reading Vocabulary	.745	1.000	.816	.689	.714	.795
Reading Comprehension	.803	.816	1.000	.721	.721	.822
Language Spelling	.679	.689	.721	1.000	.638	.738
Language Mechanics	.746	.714	.721	.638	1.000	.753
Language Expression	.783	.795	.822	.738	.753	1.000

* Criterion

Table 37 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .077 and .600. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 26 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=0 is .231 as compared with .184 for N=391. Estimated $\beta'\beta$ for N=150 Set A approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.100<k<.200$.

Figure 24 illustrates the monotonic increase in the error sum of squares as $k>0$. Reduction in R^2 with increasing k as reported in Table 28 results in an obtained decrease of .0056 in the variance accounted for at k=.700 as compared with k=.000. No coefficients change sign as $k>0$ and no predictor variable has a negative coefficient as can be seen in Figure 25. Negative signs which change as $k>0$ are noted at N=25 and N=50. No coefficients have negative signs for N=391.

Table 37

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λ min), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=150 Set A

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.7245	4.323	.1703	.104	.231
Hoerl & Kennard (1976)	.046	.7242	3.471	.2158	.086	.210
McDonald & Galarneau (1975)	.090	.7236	2.914	.2603	.074	.197
Observation	.600	.7193	1.047	.7703	.030	.139
Hoerl (1962)	.700	.7189	.9329	.8703	.027	.133

Table 38 reports the results of cross-validation of the regression equation obtained for N=150 Set A. The regression equation was cross-validated with N=150 Set B. Shrinkage decreases with increasing k. The obtained difference between shrinkage at k=.000 and shrinkage at k=.700 is .015.

Table 38

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=150 Set A

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.7245	.6811	.0434
Hoerl & Kennard (1976)	.046	.7242	.6844	.0398
McDonald & Galarneau (1975)	.090	.7236	.6863	.9373
Observation	.600	.7193	.6905	.0288
Hoerl (1962)	.700	.7189	.6906	.0283

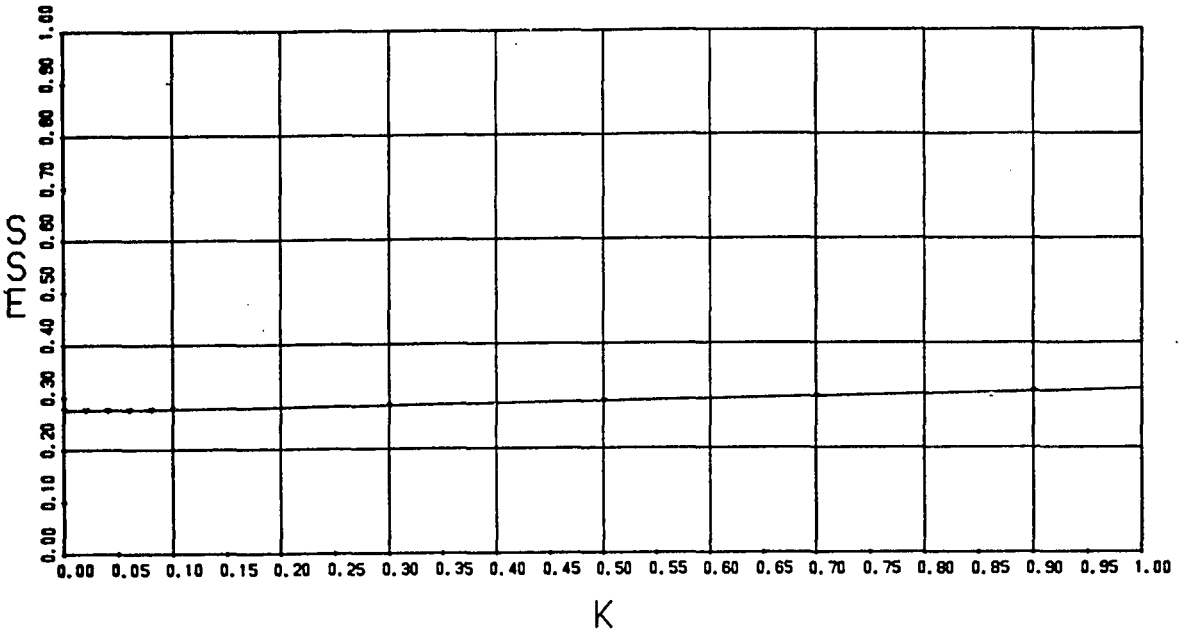

FIGURE 24
ERROR SUM OF SQUARES

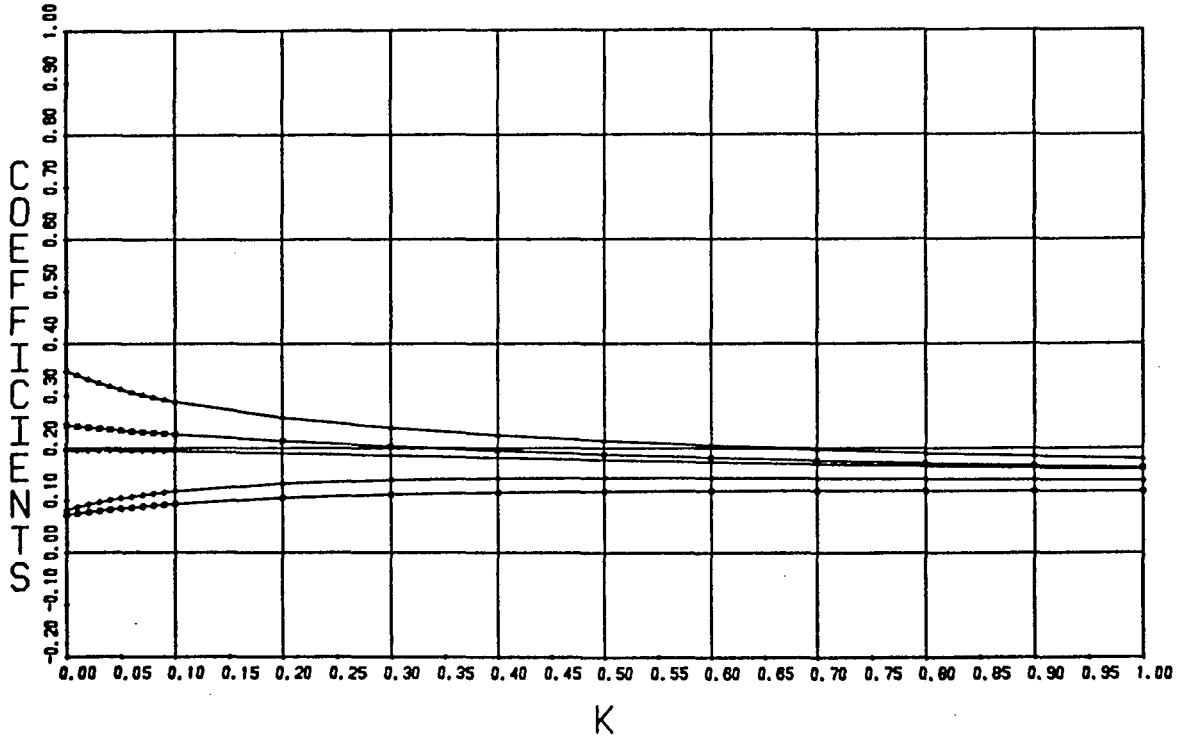

FIGURE 25
SAMPLE SIZE=150 SET A

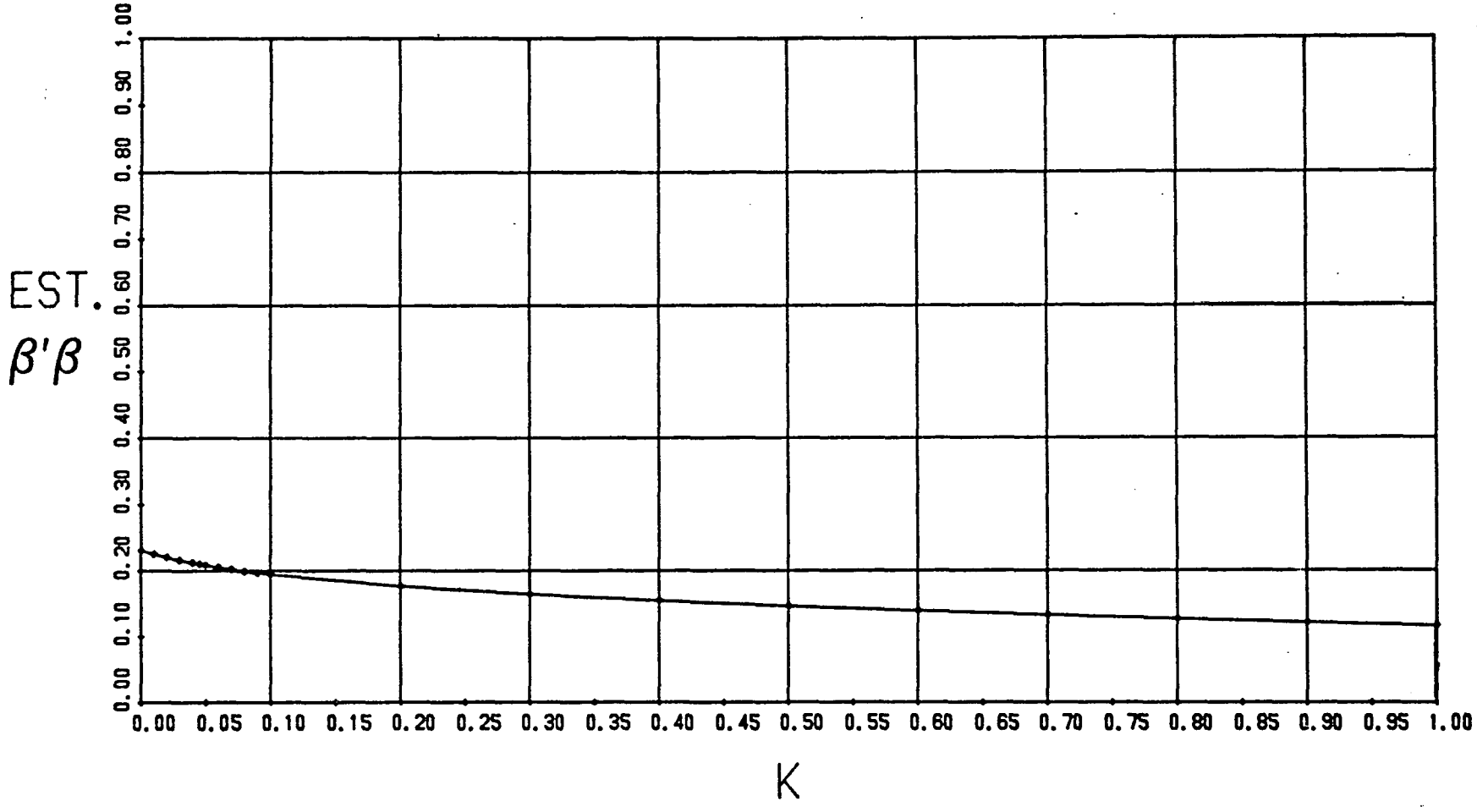

FIGURE 26
SAMPLE SIZE=150 SET A

Table 39 reports the mean and standard deviation for each raw score variable while Table 40 contains the matrix of correlations among predictors and between each predictor and the criterion for data set B which consists of 150 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 39

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=150 Set B

Variable	Mean	Standard Deviation
Mathematics Concepts *	15.207	5.672
Reading Vocabulary	19.880	9.190
Reading Comprehension	23.780	11.147
Language Spelling	29.473	8.491
Language Mechanics	10.360	5.239
Language Expression	16.213	6.967

* Criterion

Table 40

Matrix of Correlations Among Predictors and
Between Each Predictor and the Criterion for N=150 Set B

	Mathe-matics Concepts	Reading Vocabulary	Reading Compre-hension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts *	1.000	.714	.757	.672	.730	.769
Reading Vocabulary	.714	1.000	.789	.642	.654	.742
Reading Comprehension	.757	.789	1.000	.664	.740	.808
Language Spelling	.672	.642	.664	1.000	.613	.687
Language Mechanics	.730	.654	.740	.613	1.000	.763
Language Expression	.769	.742	.808	.687	.763	1.000

* Criterion

Table 41 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .050 and .400. The obtained range of k for N=391 is .279 as compared with an obtained range of .654 for N=150 Set A and .350 for N=150 Set B. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 29 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector is .184. For N=391, the estimated squared length of the coefficient vector at k=.000 is .184.

Figure 27 illustrates the monotonic increase in the error sum of squares as k>.000. Reduction in R^2 with increasing k as reported in Table 41 shows an obtained decrease of .0003 in the variance accounted for at k=.400 as compared with k=.000. Observation of the ridge trace depicted in Figure 28 shows little variation among estimated coefficients. Coefficients converge as k>.000. No predictor variable has a negative sign. Negative signs which change as k>0 are noted at N=25 and N=50. No coefficients have negative signs for N=391.

Table 41

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=150 Set B

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.6918	4.081	.1795	.033	.184
Observation	.050	.6918	3.238	.2295	.029	.178
Hoerl & Kennard (1976)	.060	.6918	3.107	.2398	.028	.177
McDonald & Galarneau (1975)	.400	.6915	1.354	.5795	.017	.148
Hoerl (1962)	.400	.6915	1.354	.5795	.017	.148

Table 42 reports the results of cross-validation of the regression equation obtained for N=150 Set B. The regression equation was cross-validated with N=150 Set A. At every level of k including k=.000, the cross-validated R^2 is larger than the R^2 obtained on the estimation group, i.e., N=150 Set B. The increase in R^2 is 2.33% at k=.000 and 2.32% at k=.400.

Table 42

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=150 Set B

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.6918	.7151	-.0233
Observation	.050	.6918	.7151	-.0233
Hoerl & Kennard (1976)	.060	.6918	.7151	-.0233
McDonald & Galarneau (1975)	.400	.6915	.7147	-.0232
Hoerl (1962)	.400	.6915	.7147	-.0232

Table 43 reports the estimated beta coefficients for each predictor variable by selection procedure together with the zero order correlation between sets of estimated beta coefficients by selection procedure. Obtained correlations at the k selected by each of the ridge k selection procedures are higher than the correlation between the two sets of estimated coefficients at k=.000. When coefficients are ordered by magnitude three changes in order occur across Sets A and B for k=.000 and at the k selected by observation. Two changes in order occur across Sets A and B at the k's selected by the Hoerl (1962), McDonald and Galarneau (1975) and Hoerl and Kennard (1976) selection procedures.

Table 43

Estimated Beta Coefficients ($\hat{\beta}^*_i$) for each Predictor Variable, k and the Correlation Between Set A and Set B Coefficients (r) Reported by Selection Procedure for N=150 Sets A and B

Selection Procedure	Selected k	Set	$\hat{\beta}^*_1$	$\hat{\beta}^*_2$	$\hat{\beta}^*_3$	$\hat{\beta}^*_4$	$\hat{\beta}^*_5$	r
Least Squares Analysis	.000	A	.081	.348	.072	.244	.195	
	.000	B	.153	.176	.160	.225	.231	.4387
Hoerl & Kennard (1976)	.046	A	.102	.315	.084	.235	.196	
	.060	B	.155	.178	.160	.217	.221	.5108
McDonald & Galarneau (1975)	.090	A	.115	.293	.092	.228	.195	
	.400	B	.154	.172	.152	.188	.192	.6547
Observation	.600	A	.143	.205	.118	.182	.173	
	.050	B	.155	.178	.160	.218	.223	.5569
Hoerl (1962)	.700	A	.142	.197	.119	.176	.169	
	.400	B	.154	.172	.152	.188	.192	.6991

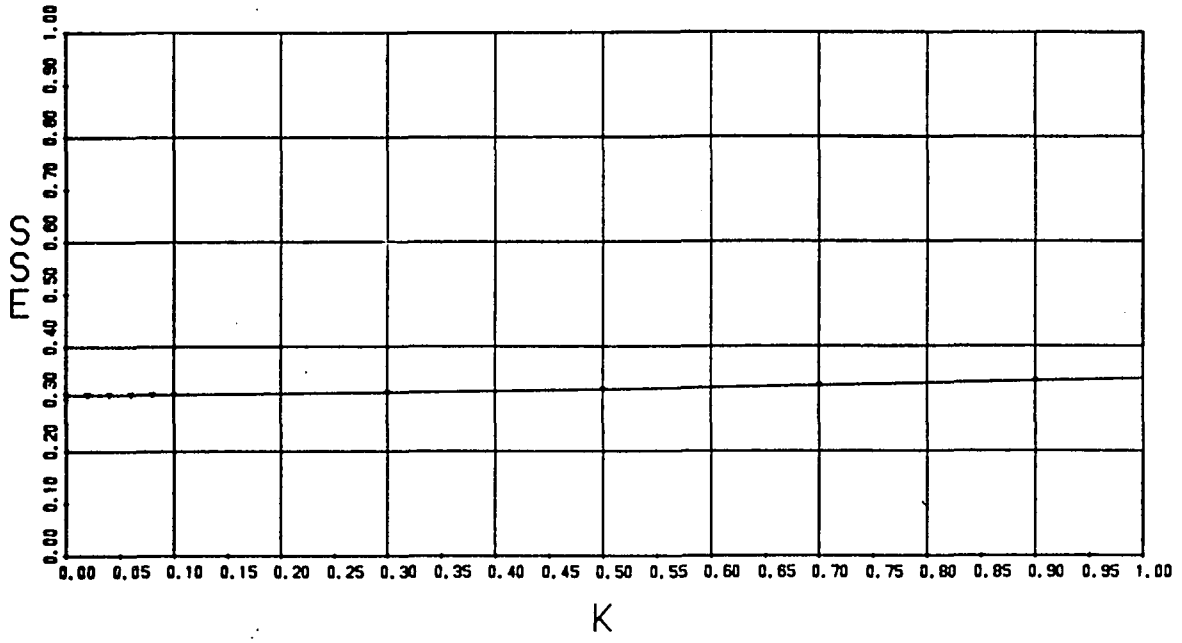

FIGURE 27
ERROR SUM OF SQUARES

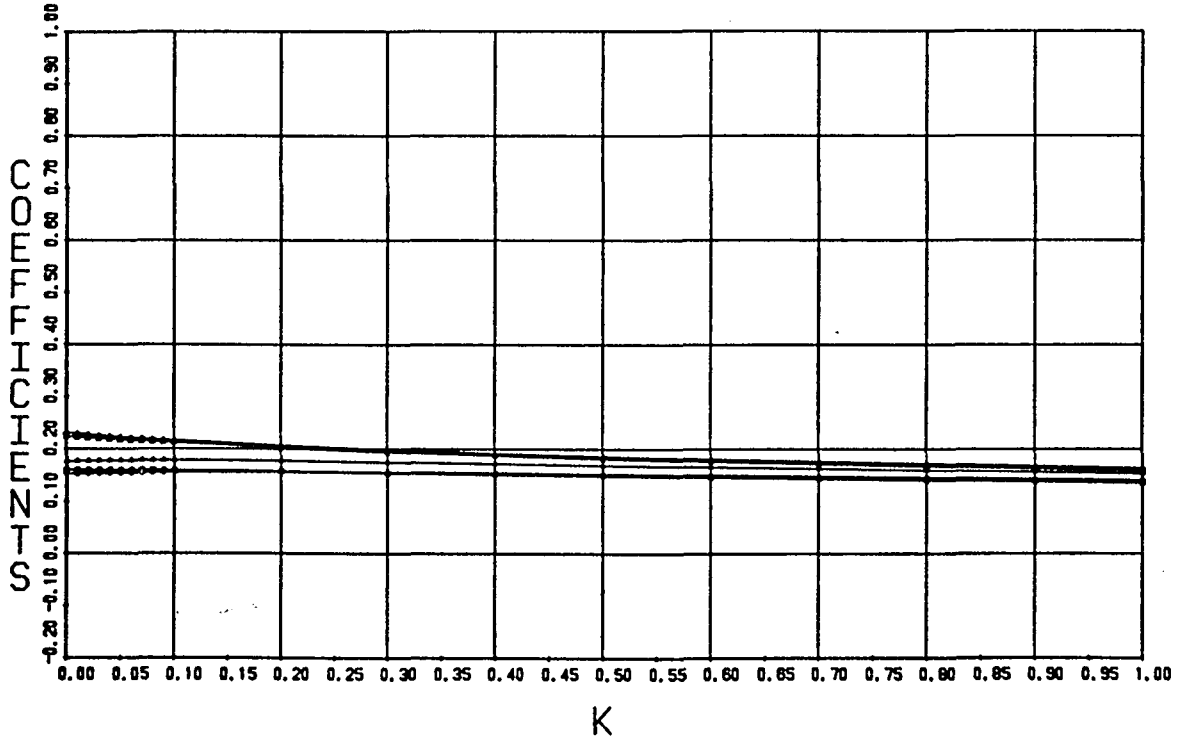

FIGURE 28
SAMPLE SIZE=150 SET B

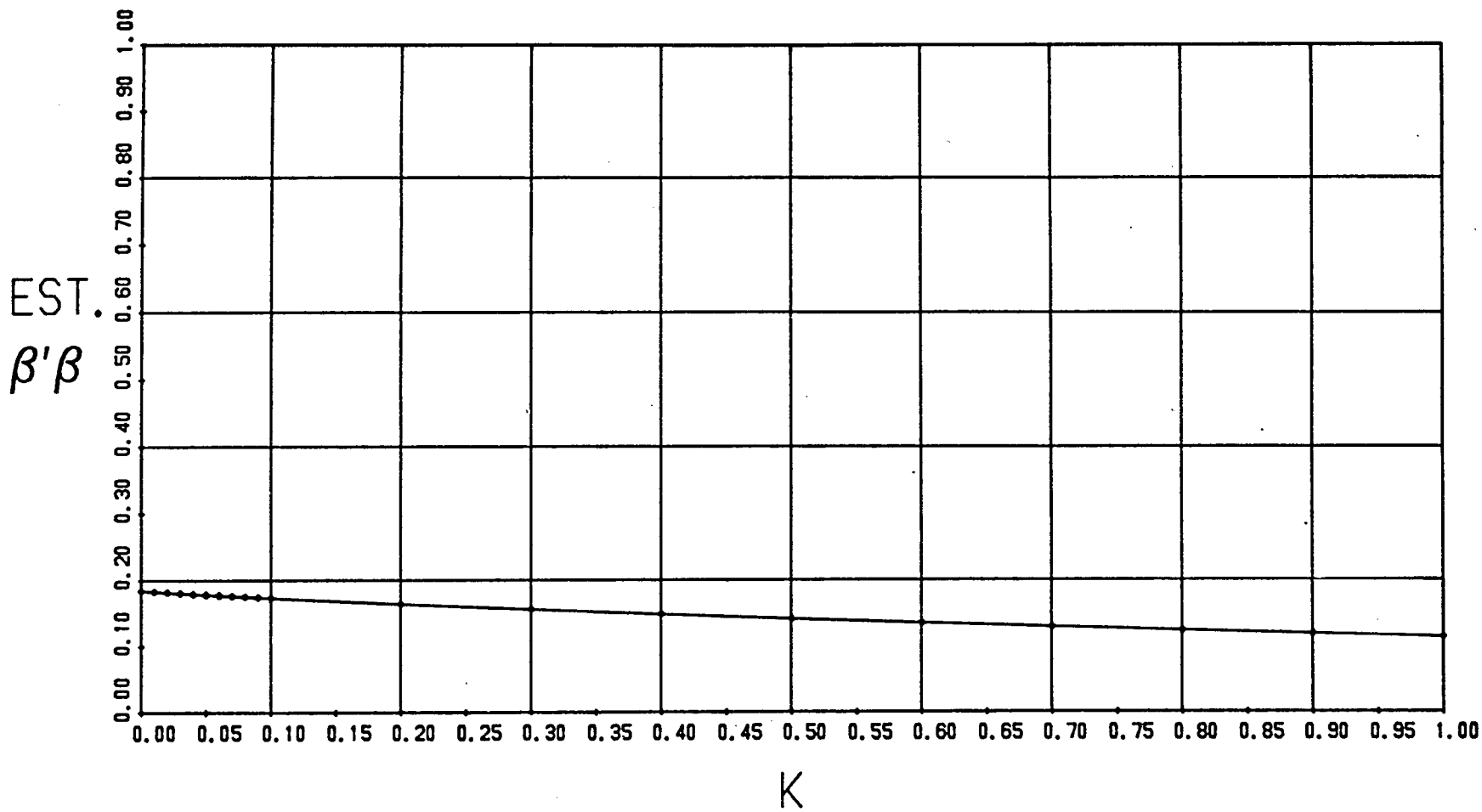

FIGURE 29
SAMPLE SIZE=150 SET B

Table 44 reports the mean and standard deviation for each raw score variable while Table 45 contains the matrix of correlations among predictors and between each predictor and the criterion for data set A which consists of 200 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 44

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=200 Set A

Variable	Mean	Standard Deviation
Mathematics Concepts*	15.780	5.914
Reading Vocabulary	21.180	9.482
Reading Comprehension	24.860	11.492
Language Spelling	29.345	8.806
Language Mechanics	11.035	5.056
Language Expression	17.130	7.214

* Criterion

Table 45

Matrix of Correlations Among Predictors and
Between Each Predictor and the Criterion for N=200 Set A

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts*	1.000	.739	.762	.669	.722	.785
Reading Vocabulary	.739	1.000	.798	.691	.621	.794
Reading Comprehension	.762	.798	1.000	.695	.707	.818
Language Spelling	.669	.691	.695	1.000	.608	.712
Language Mechanics	.722	.621	.707	.608	1.000	.757
Language Expression	.785	.794	.818	.712	.757	1.000

* Criterion

Table 46 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .042 and .500. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 32 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector at k=0 is .190 as compared with .184 for N=391. Estimated $\beta'\beta$ for N=200 Set A is .184 at k=.042.

Figure 30 illustrates the monotonic increase in the error sum of squares as k>0. Reduction in R^2 with increasing k as reported in Table 46 results in an obtained decrease of .0011 in the variance accounted for at k=.500 as compared with k=.000. Figure 31 shows little variation among estimated coefficients. No coefficients change sign as k>0 and no predictor variable has a negative coefficient as can be seen in Figure 31. Negative signs which change as k>0 are noted at N=25 and N=50. No coefficients have negative signs for N=391.

Table 46

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λ min), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=200 Set A

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.7009	4.591	.1672	.050	.190
Hoerl & Kennard (1976)	.042	.7009	3.695	.2089	.045	.184
McDonald & Galarneau (1975)	.200	.7005	2.139	.3672	.033	.167
Observation	.200	.7005	2.139	.3672	.033	.167
Hoerl (1962)	.500	.6998	1.204	.6672	.023	.143

Table 47 reports the results of cross-validation of the regression equation obtained for N=200 Set A. The regression equation was cross-validated with N=200 Set B. At every level of k including k=.000, the cross-validated R^2 is larger than the R^2 obtained on the estimation group, i.e., N=200 Set A. The increase in R^2 is 2.76% at k=.000 and 2.71% at k=.400.

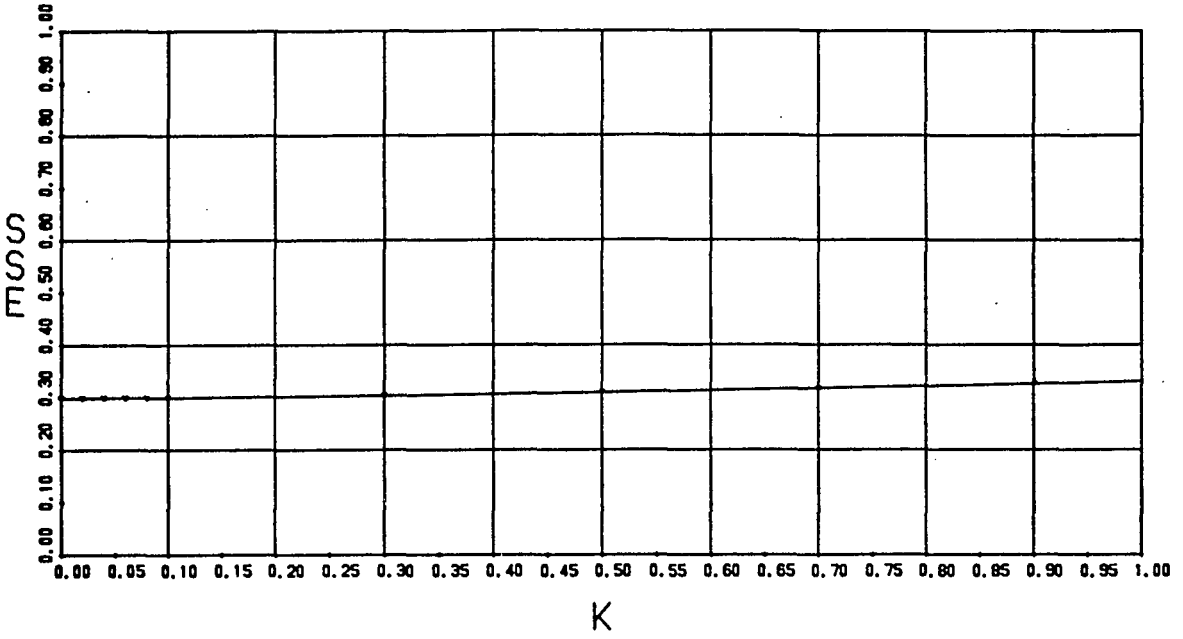

FIGURE 30
ERROR SUM OF SQUARES

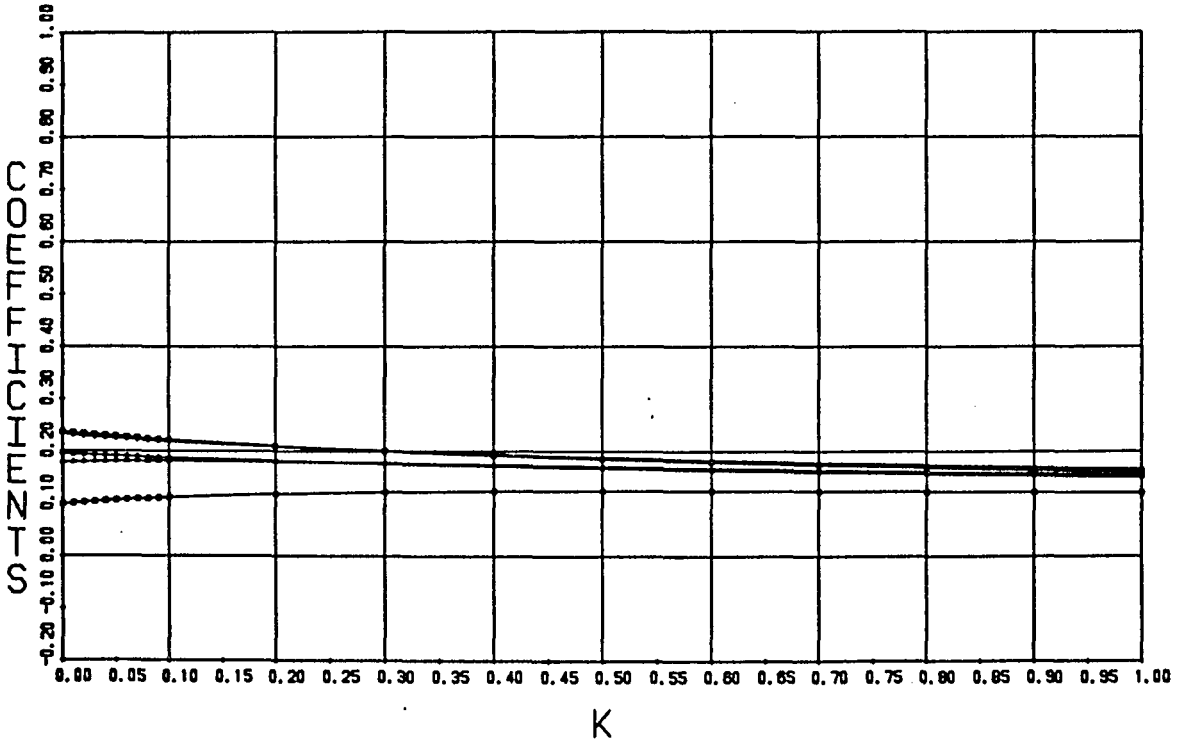

FIGURE 31
SAMPLE SIZE=200 SET A

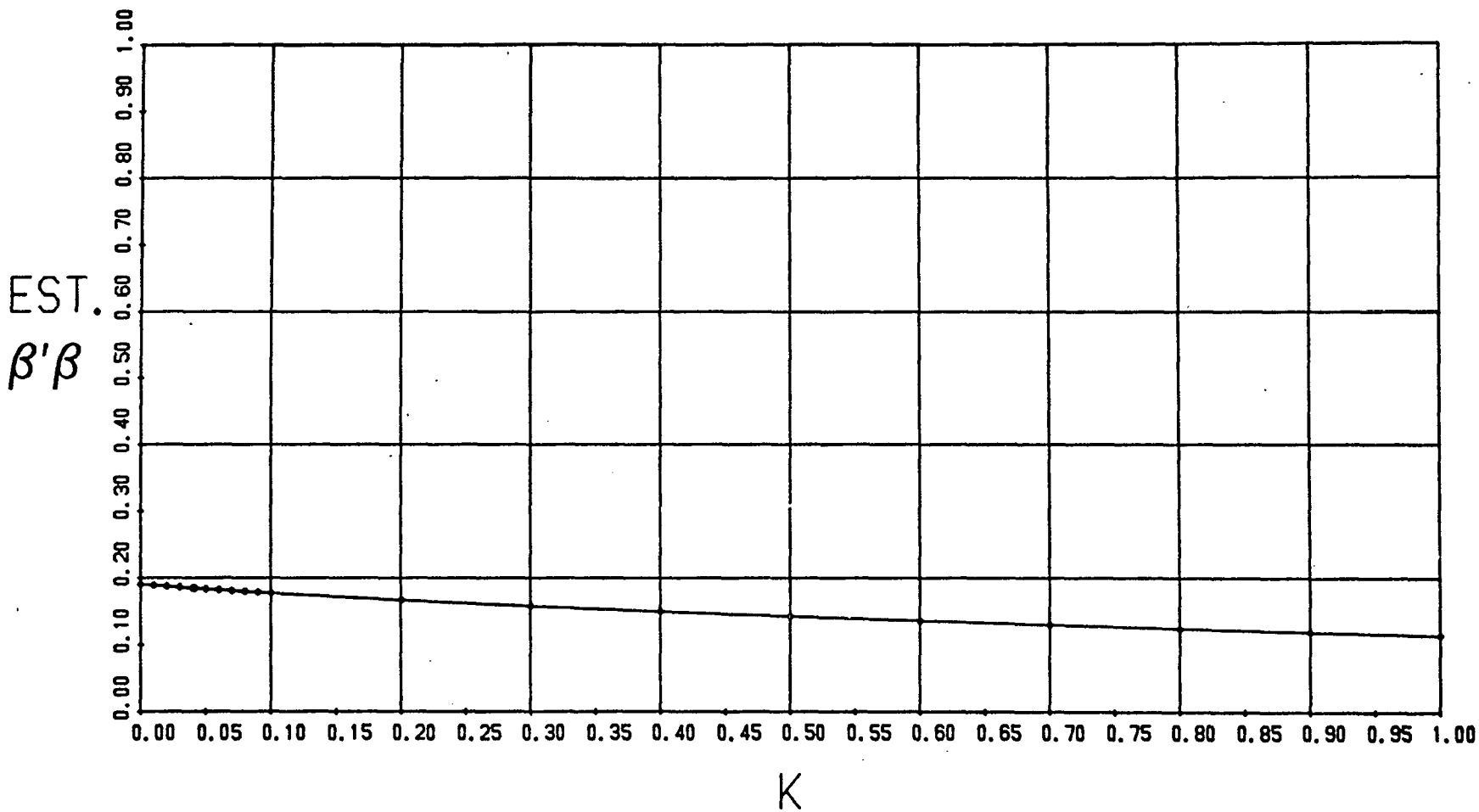

FIGURE 32
SAMPLE SIZE=200 SET A

Table 47

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=200 Set A

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.7009	.7285	-.0276
Hoerl & Kennard (1976)	.042	.7009	.7283	-.0276
McDonald & Galarneau (1975)	.200	.7005	.7277	-.0272
Observation	.200	.7005	.7277	-.0272
Hoerl (1962)	.500	.6998	.7269	-.0271

Table 48 reports the mean and standard deviation for each raw score variable while Table 49 contains the matrix of correlations among predictors and between each predictor and the criterion for data set B which consists of 200 subjects' criterion and predictor scores. Predictor variables exhibit high positive intercorrelations, an indicator of multicollinearity identified by Mason, Gunst, and Webster (1975).

Table 48

Raw Score Means and Standard Deviations of Criterion and Predictor Variables for N=200 Set B

Variable	Mean	Standard Deviation
Mathematics Concepts *	15.225	6.007
Reading Vocabulary	21.115	9.791
Reading Comprehension	24.180	11.268
Language Spelling	28.790	8.894
Language Mechanics	10.625	5.201
Language Expression	16.275	7.562

* Criterion

Table 49

Matrix of Correlations Among Predictors and
Between Each Predictor and the Criterion for N=200 Set B

	Mathematics Concepts	Reading Vocabulary	Reading Comprehension	Language Spelling	Language Mechanics	Language Expression
Mathematics Concepts*	1.000	.756	.776	.703	.767	.820
Reading Vocabulary	.756	1.000	.850	.687	.709	.818
Reading Comprehension	.776	.850	1.000	.728	.743	.838
Language Spelling	.703	.687	.728	1.000	.695	.745
Language Mechanics	.767	.709	.743	.695	1.000	.797
Language Expression	.820	.818	.838	.745	.797	1.000

* Criterion

Table 50 reports the k selected by each selection procedure including least squares analysis, i.e., k=.000. Selected k's chosen through application of the four ridge k selection procedures show greater variability than the k's selected by the same procedures for N=391 and range between .034 and .700. The obtained range of k for N=391 is .279 as compared with an obtained range of .458 for N=200 Set A and .666 for N=200 Set B. As k increases, the VIF decreases and the λmin increases as expected (Hoerl & Kennard, 1970a, 1970b). Additionally, the standard deviation of the p sample beta weights and the estimated squared length of the coefficient vector decrease as k increases. Figure 35 depicts the decrease in estimated $\beta'\beta$ with increasing k. The estimated squared length of the coefficient vector is .216 at k=.000. For N=391, the estimated squared length of the coefficient vector at k=.000 is .184. Estimated $\beta'\beta$ for N=200 Set B approaches the estimated length of $\beta'\beta$ at k=0 for N=391 with $.100<k<.200$.

Figure 33 illustrates the monotonic increase in the error sum of squares as $k>.000$. Reduction in R^2 with increasing k as reported in Table 50 shows an obtained decrease of .0041 in the variance accounted for at k=.700 as compared with k=.000. The ridge trace depicted in Figure 34 shows that no predictor variable has a negative sign. Negative signs which change as $k>0$ are noted at N=25 and N=50. No coefficients have negative signs for N=391.

Table 50

Variance Accounted for (R^2), Maximum Variance Inflation Factor (VIF), Minimum Eigenvalue (λmin), Standard Deviation of the Estimated Betas ($\hat{s}$), and Estimated Squared Length of the Coefficient Vector (est $\beta'\beta$) Reported by the k Selected Through Each Selection Procedure for N=200 Set B

Selection Procedure	Selected k	R^2	VIF	λmin	$\hat{s}$	est. $\beta'\beta$
Least Squares Analysis	.000	.7317	5.032	.1456	.090	.216
Hoerl & Kennard (1976)	.034	.7315	4.141	.1796	.077	.203
McDonald & Galarneau (1975)	.090	.7310	3.213	.2356	.063	.189
Observation	.400	.7287	1.451	.5456	.033	.153
Hoerl (1962)	.700	.7276	.958	.8456	.023	.132

Table 51 reports the results of cross-validation of the regression equation obtained for N=200 Set B. The regression equation was cross-validated with N=200 Set A. Shrinkage decreases with increasing k. The obtained difference between shrinkage at k=.000 and shrinkage at k=.700 is .0058.

Table 51

Variance Accounted for (R^2), Cross-Validated Variance Accounted for (R^2c), and Shrinkage Reported by the k Selected Through Each Selection Procedure for N=200 Set B

Selection Procedure	Selected k	R^2	R^2c	Shrinkage
Least Squares Analysis	.000	.7317	.6976	.0341
Hoerl & Kennard (1976)	.034	.7315	.6987	.0328
McDonald & Galarneau (1975)	.090	.7310	.6996	.0314
Observation	.400	.7287	.6997	.0290
Hoerl (1962)	.700	.7276	.6993	.0283

Table 52 reports the estimated beta coefficients for each predictor variable by selection procedure together with the zero order correlation between sets of estimated beta coefficients by selection procedure. The obtained correlation between the two sets of estimated coefficients at the k selected by each ridge k selection procedure is higher than the correlation between the two sets of estimated coefficients at k=.000. When coefficients are ordered by magnitude four changes in order occur across Sets A and B for k=.000 and at the k selected by each of the following ridge k selection procedures: observation, McDonald and Galarneau (1975), and Hoerl and Kennard (1976). No changes in order by magnitude occur at the k selected by the Hoerl (1962) procedure. Due to the convergence of estimated coefficients as $k>0$, slight changes in estimated β produce significant differences in the order by magnitude of estimated β when sets A and B are compared.

Table 52

Estimated Beta Coefficients (β^*_i) for each Predictor Variable, k and the Correlation Between Set A and Set B Coefficients (r) Reported by Selection Procedure For N=200 Sets A and B

Selection Procedure	Selected k	Set	$\hat{\beta}^*_1$	$\hat{\beta}^*_2$	$\hat{\beta}^*_3$	$\hat{\beta}^*_4$	$\hat{\beta}^*_5$	r
Lease Squares Analysis	.000	A	.196	.179	.100	.237	.233	
	.000	B	.126	.135	.101	.236	.341	.7611
Hoerl & Kennard (1976)	.042	A	.192	.181	.105	.230	.226	
	.034	B	.133	.143	.108	.233	.315	.7754
McDonald & Galarneau (1975)	.200	A	.181	.180	.118	.210	.208	
	.090	B	.141	.150	.116	.226	.286	.7924
Observation	.200	A	.181	.180	.118	.210	.208	
	.400	B	.151	.158	.131	.197	.221	.8515
Hoerl (1962)	.500	A	.168	.171	.125	.185	.188	
	.700	B	.147	.153	.130	.178	.196	.8708

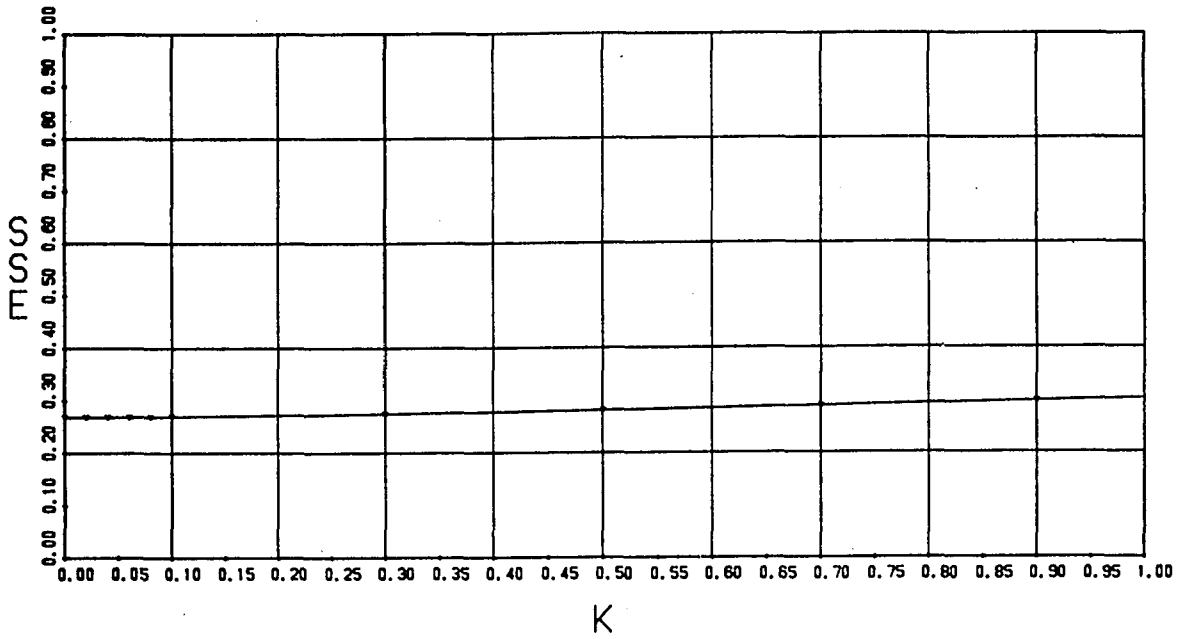

FIGURE 33
ERROR SUM OF SQUARES

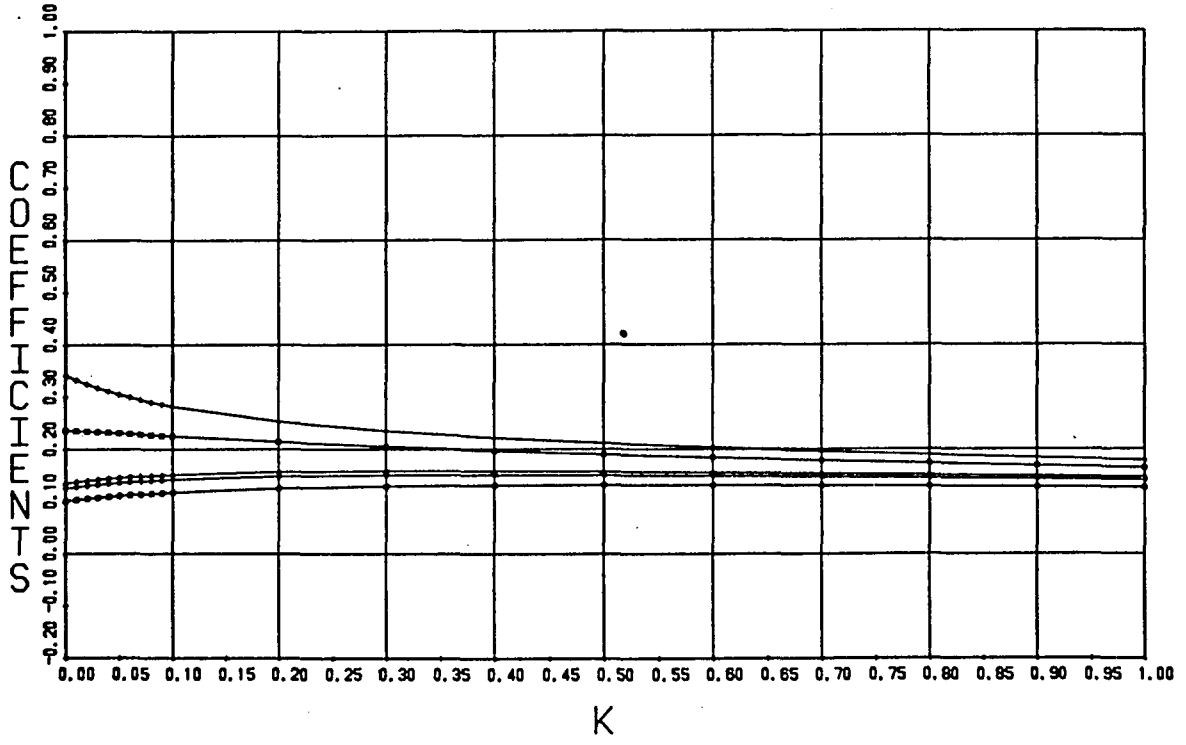

FIGURE 34
SAMPLE SIZE=200 SET B

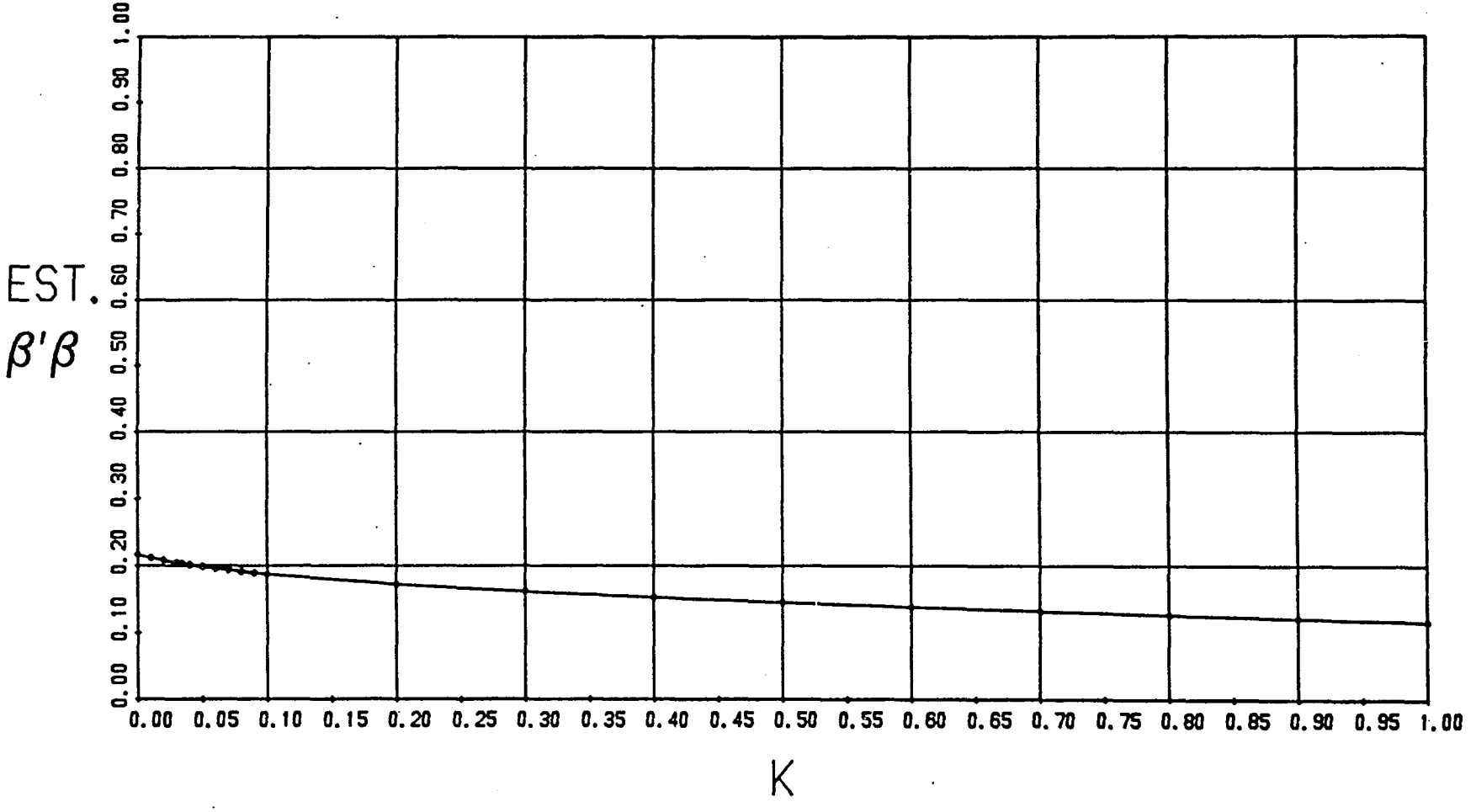

FIGURE 35
SAMPLE SIZE=200 SET B

Summary

Results are summarized across samples by research question. Descriptive comparisons across sample sizes are made.

Question 1: Is the maximum variance inflation factor obtained through inversion of the input correlation matrix with each of the selected values of k added to the diagonal smaller than the maximum variance inflation factor obtained through inversion of the original input correlation matrix at the following sample sizes: 25, 50, 100, 150, 200, 391?

Table 53 reports the minimum eigenvalue at k=0 by sample. The λmin obtained at k=0 range between .1795 and .1140. A minimum eigenvalue of <1.00 indicates multicollinearity. The lower the λmin, the greater the multicollinearity (Marquardt & Snee, 1975). No pattern is noted with respect to the relative magnitudes of λmin at k=0 as sample size varies. For every sample size and data set at every selected level of k (Tables 6, 10, 14, 19, 23, 28, 32, 37, 41, 46, 50), the obtained λmin is larger than the obtained λmin at k=0 for the same data set.

Table 53

Minimum Eigenvalue at k=.000 Reported by Sample

Sample Size	Set	Minimum Eigenvalue
391	A	.1620
25	A	.1621
25	B	.1482
50	A	.1140
50	B	.1391
100	A	.1545
100	B	.1737
150	A	.1703
150	B	.1795
200	A	.1672
200	B	.1456

Question 2: Is the minimum eigenvalue obtained through analysis of the input correlation matrix with each of the selected values of k added to the diagonal larger than the minimum eigenvalue obtained through analysis of the original input correlation matrix at the following sample sizes: 25, 50, 100, 150, 200, 391?

Table 54 reports the maximum variance inflation factor at k=0 by sample. The VIF obtained at k=0 range between 6.474 and 4.003. The greater the VIF, the greater the multicollinearity. VIF can range between 1.00 and infinity (Marquardt & Snee, 1975). No pattern is noted with respect to the relative magnitudes of VIF at k=.000 as sample size varies.

For every sample size and data set at every selected level of k (Tables 6, 10, 14, 19, 23, 28, 32, 37, 41, 46, 50), the obtained VIF is smaller than the obtained VIF at k=0 for the same data set.

Table 54

Maximum Variance Inflation Factor
at k=.000 Reported by Sample

Sample Size	Set	Maximum Variance Inflation Factor
391	A	4.649
25	A	4.003
25	B	4.632
50	A	6.474
50	B	4.975
100	A	4.107
100	B	4.022
150	A	4.323
150	B	4.081
200	A	4.591
200	B	5.032

Question 3: Is the standard deviation of the five sample beta weights obtained through ridge regression at the selected values of k less than the standard deviation of the five sample beta weights obtained through ordinary least squares analysis at the following sample sizes: 25, 50, 100, 150, 200, 391?

Table 55 reports the standard deviation of each set of five estimated beta weights by sample. Claudy (1972) contends that the variability of the p sample beta weights at k=.000 is inflated and the inflation increases with decreasing sample size. The standard deviation of the five sample beta weights at N=25 and N=50 is greater than the standard deviation of the five sample beta weights at the larger sample sizes. When compared with the standard deviation obtained with all available data, i.e., at N=391, greater variability among the estimated coefficients is noted at N=25 and N=50.

Table 56 reports the range of k by sample size and data set in which the standard deviation of the five sample beta weights approaches .029, i.e., the standard deviation at N=391. Greater k must be added to the diagonal of X'X at N=25 and N=50 for the standard deviation to approximate .029 as compared with the larger sample sizes. For every sample size and data set at every selected level of k (Tables 6, 10, 14, 19, 23, 28, 32, 37, 41, 46, 50), the obtained standard deviation is less than the obtained standard deviation at k=0 for the same data set.

Table 55

Standard Deviation of the Estimated Beta Coefficients Reported by Sample

Sample Size	Set	Standard Deviation
391	A	.029
25	A	.187
25	B	.284
50	A	.177
50	B	.208
100	A	.083
100	B	.028
150	A	.104
150	B	.033
200	A	.050
200	B	.090

Table 56

Range of k Within Which the Standard Deviation of Estimated Coefficients Approaches the Standard Deviation of Estimated Coefficients at k=.000 for N=391 Reported by Sample

Sample Size	Set	k Range
25	A	.300<k<.400
25	B	.400<k<.500
50	A	.400<k<.500
50	B	.300<k<.400
100	A	.100<k<.200
100	B	.010<k<.020
150	A	.100<k<.200
150	B	.000<k<.010
200	A	.042<k<.050
200	B	.100<k<.200

Question 4: Is the estimated squared length of the coefficient vector which is calculated based upon ridge regression coefficients at the selected values of k less than the estimated squared length of the coefficient vector calculated based upon ordinary least squares regression coefficients at the following sample sizes: 25, 50, 100, 150, 200, 391?

Table 57 reports the estimated length of $\beta'\beta$ at k=0 by sample. Hoerl and Kennard (1970a) contend that $\beta'\beta$ is overestimated when data are multicollinear. Estimated $\beta'\beta$ at N=25 and N=50 is greater than estimated $\beta'\beta$ at the larger

sample sizes. When compared with estimated $\beta'\beta$ obtained with all available data, i.e, at N=391, greater estimated $\beta'\beta$ is noted at N=25 and N=50. Table 58 reports the range of k by sample size and data set in which estimated $\beta'\beta$ approaches .184, i.e., estimated $\beta'\beta$ at N=391. Greater k must be added to the diagonal of X'X at N=25 and N=50 for estimated $\beta'\beta$ to approximate .184 as compared with the larger sample sizes. For every sample size and data set at every selected level of k (Tables 6, 10, 14, 19, 23, 28, 32, 37, 41, 46, 50), estimated $\beta'\beta$ is less than estimated $\beta'\beta$ at k=0 for the same data set.

Table 57

Estimated Length of $\beta'\beta$ at k=.000 Reported by Sample

Sample Size	Set	Est. $\beta'\beta$
391	A	.184
25	A	.381
25	B	.603
50	A	.354
50	B	.401
100	A	.216
100	B	.186
150	A	.231
150	B	.184
200	A	.190
200	B	.216

Table 58

Range of k Within Which Estimated $\beta'\beta$ Approaches Estimated $\beta'\beta$ at k=.000 for N=391 Reported by Sample

Sample Size	Set	k Range*
25	A	.900<k<1.000
25	B	1.000<k<Unknown
50	A	1.000<k<Unknown
50	B	1.000<k<Unknown
100	A	.300<k<.400
100	B	Unknown<k<.000
150	A	.600<k<.700
150	B	.040<k<.050
200	A	.400<k<.500
200	B	.200<k<.300

*The word unknown is entered when k is out of the range of k for which estimated $\beta'\beta$ were computed.

Question 5: Is the amount of shrinkage in the R^2 obtained through ridge regression upon cross-validation with a prediction group at the selected values of k less than the amount of shrinkage in the R^2 obtained through ordinary least squares analysis upon cross-validation with a prediction group at the following sample sizes: 25, 50, 100, 150, 200?

Table 59 reports the variance accounted for (R^2) at k=.000 and k=1.00 by sample. Claudy (1972) states that, given multicollineary predictors, sample R^2 at k=.000 is

inflated with respect to population R^2 and the inflation increases as sample size decreases. The R^2 at N=25 and N=50 are greater than the R^2 at the larger sample sizes. A greater difference between the R^2 obtained at N=25 and N=50 at k=.000 and the R^2 at k=.000 for N=391 is noted than for the larger sample sizes. The difference between R^2 at k=.000 and R^2 at k=1.000 for N=25 and N=50 is greater than for the larger sample sizes, ranging between .0209 and .0476 as compared with a range of .0005 to .0066 for the larger sample sizes. Shrinkage upon cross-validation decreases with increasing k for the following data sets: N=25 Set A, N=25 Set B, N=50 Set A, N=50 Set B, N=100 Set A, N=100 Set B, N=150 Set A, N=200 Set B. Shrinkage upon cross-validation increases with increasing k for the following data sets: N=150 Set B, N=200 Set A. Results are reported in Tables 11, 15, 20, 24, 29, 33, 38, 42, 47 and 51. A difference in shrinkage upon cross-validation of greater than one percent is noted between the maximum selected k and k=.000 for the following data sets: N=25 Set A, N=25 Set B, N=50 Set A, N=50 Set B, N=100 Set A. A difference in shrinkage upon cross-validation of less than one percent is noted between the maximum selected k and k=.000 for the following data sets: N=100 Set B, N=150 Set A, N=200 Set A, N=200 Set B. Since the maximum selected k is greater for N=25 and N=50 than for the larger sample sizes and a different set of k is selected for each sample, data necessary to compare differences in shrinkage with varying sample size at a given k as compared to k=.000 are not available. For every data set except N=150 Set B and N=200 Set A, shrinkage upon cross-validation at all ridge selected values of k is less than shrinkage upon cross-validation at k=.000 for the same data set.

Table 59

Variance Accounted for (R^2) at k=.000 and k=1.000
Reported by Sample

Sample Size	Set	R^2 k=.000	R^2 k=1.000
391		.7094	.7089
25	A	.8010	.7801
25	B	.7625	.7149
50	A	.8307	.8091
50	B	.7772	.7518
100	A	.7221	.7184
100	B	.6925	.6919
150	A	.7245	.7179
150	B	.6918	.6911
200	A	.7009	.6991
200	B	.7317	.7269

Question 6: Are the estimated beta coefficients obtained through ridge regression at each selected level of k more similar to a second set of estimated beta coefficients at the level of k selected by the same selection procedure in the other sample of the same size than estimated coefficients obtained through ordinary least squares analysis at the following sample sizes: 25, 50, 100, 150, 200?

Hoerl and Kennard (1970a, 1970b) contend that estimated beta coefficients obtained through ridge analysis are more stable than estimated beta coefficients obtained through least squares analysis when data are multicollinear. Stabil-

ity is operationally measured through calculation of the zero order correlation between the sets of coefficients obtained at k=0 and at the k selected by each ridge k selection procedure for each pair of samples of the same size, as well as calculation of the number of changes in order by magnitude of each obtained β^* across data sets at the same sample size. Results are reported in Tables 16, 25, 34, 43, and 52. Estimated coefficients at a given k are said to be more similar to a second set at the level of k selected by the same procedure in the other data set of the same sample size than for k=.000 if the obtained correlation is greater than for k=.000 and if the number of changes in order by magnitude are fewer than for k=.000.

At N=25, estimated beta coefficients obtained through least squares analysis are more similar across sets A and B than estimated coefficients obtained at the k selected by any of the ridge k selection procedures. At N=50, estimated beta coefficients obtained at the k selected through the McDonald and Galarneau (1975) and Hoerl and Kennard (1976) procedures are more similar across sets A and B than estimated coefficients obtained at the k selected through at k=.000 although differences in the correlation coefficient are small. At N=50, pairs of estimated beta coefficients obtained through least squares analysis are more similar across sets A and B than pairs of estimated coefficients obtained at the k selected by observation and the Hoerl (1962) procedure although differences in the correlation coefficient are small. At N=100, the correlation between sets of pairs of estimated beta

is negative at k=.000 and at the k selected by the ridge k selection procedures. At N=150 and N=200, pairs of estimated beta coefficients obtained at the k selected by each of the ridge procedures are more similar across sets A and B than pairs of estimated coefficients at k=.000.

Table 60

Correlation Between Pairs of Estimated Coefficients and Number of Changes in Order by Magnitude of Estimated Coefficients Across Data Sets of the Same Sample Size Reported by Selection Procedure

	N=25		N=50		N=100		N=150		N=200	
Selection Procedure	r	Changes In Order By Magnitude	r	Changes	r	Changes	r	Changes	r	Changes
Least Squares Ana- Analysis	.5436	3	.5495	2	-.4038	4	.4387	3	.7611	4
Observation	.3000	4	.5331	2	-.4819	5	.5569	3	.8515	4
Hoerl (1962)	.2772	4	.5256	4	-.2141	4	.6991	2	.8708	0
McDonald & Galarneau (1975)	.4889	4	.5552	2	-.0588	5	.6547	2	.7924	4
Hoerl & Kennard (1976)	.4030	4	.5559	2	-.2700	4	.5108	2	.7754	4

CHAPTER V

Conclusions

Summary

Hoerl (1962) proposes ridge regression as an alternative to ordinary least squares analysis when data are multicollinear. Criteria for comparing the results of ridge analysis to the results of least squares analysis are reported in the ridge regression literature (Hoerl, 1962; Hoerl & Kennard, 1970a, 1970b; Marquardt & Snee, 1975; Walton, Newman, and Fraas, 1978) and include the following: stability of estimated beta coefficients across samples from the same population, value of the obtained VIF and λmin among procedures, amount of reduction in R^2 from the least squares analysis estimate of R^2 obtained when ridge regression is applied both before and after cross-validation, and subjective evaluation of how reasonable the signs and absolute values of the estimated beta coefficients are.

Claudy (1972) contends that the problems which result from the application of ordinary least squares analysis to multicollinear data are increased as sample size decreases. Overestimation of R^2 at k=.000 and inflated variability of the p sample beta weights at k=.000 are noted (Claudy, 1972).

Research questions which are the basis for comparison of ordinary least squares analysis with ridge regression are stated in Chapter 3. Results are reported in Chapter 4.

Data are scores on the CTBS for 391 fourth grade students. Two samples at each of the following sample sizes as well as N=391 are analyzed: 25, 50, 100, 150, 200. The following ridge k selection procedures are employed: Observation, Hoerl (1962), McDonald and Galarneau (1975), and Hoerl and Kennard (1976). Each selection procedure is described in Chapter 2.

Conclusions based upon obtained results as reported in Chapter 4 are made. Recommendations for further research are stated.

Conclusions

Ridge regression is effective in increasing the minimum eigenvalue and reducing the maximum variance inflation factor from the levels obtained with k=.000, i.e., ordinary least squares analysis. The ridge results behave more like results obtained given orthogonal input data in that the standard deviation of the five sample beta weights and the estimated squared length of the coefficient vector are both reduced. If, for discussion purposes, the entire available data set, i.e., N=391, is treated as the population, the addition of $k>0$ to the diagonal for every data set except N=100 Set B is needed for the standard deviation and estimated $\beta'\beta$ to approach the population values. A larger k is needed for N=25 and N=50 than for N=100, N=150, and N=200. Inflation of the standard deviation of the five sample beta weights and estimated $\beta'\beta$ is greater at N=25 and N=50 than for the larger sample sizes so that larger k must be employed to obtain comparable results.

As Claudy (1972) contends, the problems concomitant with application of ordinary least squares analysis to multicollinear data are heightened at the smaller sample sizes. Ridge analysis is effective in the reduction of the inflated variability of the p sample beta weights as well as in lessening the inflation of estimated $\beta'\beta$. Given smaller sample size, a larger k is needed to accomplish results comparable to those obtained with relatively smaller k in a larger sample. Once a subjects to predictor variable ratio of 20 to 1 is obtained, i.e., N=100, no consistent pattern in obtained reduction of the inflation of the standard deviation of the p sample beta weights and estimated $\beta'\beta$ is noted as the subjects to predictor variable ratio increases. Given a subjects to predictor variable ratio of 20 to 1 or greater, obtained standard deviations and estimated $\beta'\beta$ more closely approximate the population values than with lesser subjects to predictor variable ratios, indicating the greater need for ridge analysis at the smaller sample sizes if the sample statistics are to approach the population parameters, i.e., σ of the p sample beta weights and $\beta'\beta$.

Claudy's (1972) statement that, given multicollinear predictors, sample R^2 at k=.000 is inflated with respect to population R^2 and the inflation increases as sample size decreases is supported. If, for discussion purposes, the entire available data set, i.e., N=391, is treated as the population, R^2 at k=0 exhibits greater inflation at N=25 and N=50 than at the larger sample sizes. The addition of $k>0$ to the diagonal of X'X given N=25 and N=50 results in a greater reduction of R^2

from R^2 at k=.000 than for the larger sample sizes. Shrinkage upon cross-validation is less for the ridge selected k's than for k=.000, i.e., least squares analysis, for all data sets except N=150 Set B and N=200 Set B. The difference in shrinkage exceeds 1% given N=25, N=50, and N=100 Set A, an indication of the greater efficacy of ridge regression as compared to ordinary least squares analysis at the smaller sample sizes.

Coefficients change sign as k>0 given N=25 and N=50. No sign changes are noted for N=100, N=150, N=200, or N=391. Coefficients change to the expected, i.e., positive, sign as k>0 for N=25 and N=50. Hoerl and Kennard's (1970a, 1970b) contention that ridge analysis results in more meaningful signs is, thus, supported at N=25 and N=50.

Obtained results with respect to the stability of estimated beta coefficients are mixed. For N=25, estimated β are more similar across sets A and B for k=.000, i.e., ordinary least squares analysis. For N=50, estimated β are more similar across sets A and B for k=.000 than at the k selected by observation and the Hoerl (1962) procedure but less similar across sets A and B than at the k selected by the McDonald and Galarneau (1975) and Hoerl and Kennard (1976) procedures. Negative correlations between pairs of estimated β at every selected k are obtained for N=100. For N=150 and N=200, estimated β are less similar across sets A and B for ordinary least squares analysis than for the ridge k selection procedures. Ridge regression does not consistently produce more stable, i.e., more similar, obtained coefficients across samples from the same population, than ordinary least squares analysis.

Ridge regression is a useful alternative to ordinary least squares analysis given multicollinear predictors. Particularly with smaller sample sizes, ridge regression lessens some of the problems associated with ordinary least squares analysis when X'X is ill conditioned. However, interpretation of the relative magnitudes of $\hat{\beta}^*$ as indicators of the relative importance of the predictors in the equation is not recommended since obtained $\hat{\beta}^*$ across samples from the same population are not consistently more stable than obtained $\hat{\beta}$.

Recommendations

Further research is needed in order to establish a reliable procedure for the point estimation of k. Procedures suggested by Hoerl (1962), McDonald and Galarneau (1975), Hoerl and Kennard (1976) as well as observation of the ridge trace result in the selection of k's which fluctuate widely across samples from the same population. In addition, the need for further research into the effect of sample size on the relative efficacy of ridge procedures and least squares analysis is noted.

APPENDIX A

Filer

```
FILER:PROC OPTIONS(MAIN);
 /*     THIS PROGRAM TAKES A FILE OF A SAMPLE POPULATION AND
    RANDOMLY SELECTS LINES FROM THAT FILE WITHOUT REPLACEMENT.
    IT PLACES THESE LINES INTO THE OUTPUT FILE .                    */
   DCL REP BIT(1),
       (RFSIZE,WFSIZE) FIXED BIN(31,0),
       BUFF CHAR(100) VARYING,
       (RFNAME,WFNAME) CHAR(10) VARYING,
       (DSET,SFILE,SUBSET) FILE,
       (WNUM,LNUM) FIXED DEC(9,3),
       REPLACE(1000) BIT(1),
       MOD BIT(32),
       IHERITE ENTRY( ,BIT(32),FIXED DEC(9,3),FILE),
       IHEREAD ENTRY( ,BIT(32),FIXED DEC(9,3),FILE),
       RAND ENTRY(FIXED BIN(31)) RETURNS(FLOAT BINARY),
       LINKSTR CHAR(20) VARYING;
   MOD='00000000000000000000000000000010'B;
   DO I=1 TO 1000;
     REPLACE(I)='0'B;
   END;
   /*     SFILE IS A FILE THAT CONTAINS THE NAME AND SIZE OF BOTH
       THE FILE CHOSEN TO RECIEVE THE SUBSET OF THE POPULATION
       AND THE FILE THAT CONTAINS THE ENTIRE POPULATION DATA.    */
   OPEN FILE(SFILE);
   GET FILE(SFILE) EDIT(WFNAME,WFSIZE,RFNAME,RFSIZE)
       (A(5),X(1),F(3,0),SKIP,A(4),X(1),F(3,0));
   LINKSTR='DSET='||RFNAME;
   CALL ATTACH(LINKSTR);
   LINKSTR='SUBSET='||WFNAME;
   CALL ATTACH(LINKSTR);
   OPEN FILE (DSET) UPDATE;
   OPEN FILE(SUBSET) UPDATE;
   WNUM=1;
   /*     RANDOMLY SELECT A LINENUMBER AND THEN IF IT HAS NOT BEEN
        PREVIOUSLY SELECTED AND IT IS WITHIN THE RANGE OF THE
        POPULATION LINE NUMBERS, THEN PLACE THAT DATA SET INTO
        THE SUBSET BEING FORMED.                                    */
   DO WHILE(WNUM<=WFSIZE);
       LNUM=CEIL(RAND(0)*1000);
       IF LNUM<=RFSIZE & (~REPLACE(LNUM)) THEN DO;
           REPLACE(LNUM)='1'B;
           CALL IHEREAD(BUFF,MOD,LNUM,DSET);
           CALL IHERITE(BUFF,MOD,WNUM,SUBSET);
           WNUM=WNUM+1;
        END;
    END;
   CLOSE FILE(DSET);
   CLOSE FILE(SUBSET);
   END;
```

APPENDIX B

Correlate

```
C     THIS PROGRAM TAKES AN ARRAY OF RAW SCORES OF A POPULATION OF
C     SIZE=ROWS AND NUMBER OF VARIABLES=COL.  IT OUTPUTS THE COL
C     BY COL CORRELATION MATRIX AND ALL INFORMATION NEEDED TO
C     DRIVE THE PROGRAM LABELED REGRESS.
      REAL*4 Z(400,20),C(20,20),FMTIN(20),FMTOUT(20)
      REAL*4 D(400,20),AVG(20),SD(20),DSQ(20),SDC(20)
      INTEGER*4 ROWS,TYPE,COL,ORDER /1/,ZERO /0/
C     Z IS THE ARRAY OF Z SCORES.  C IS THE CORRELATION MATRIX.
C     FMTIN AND FMTOUT ARE THE FORMATS TO READ IN D AND PRINT OUT C
C     TYPE IS THE LETTER DESIGNATING THE SUBSET.
C     D IS THE ARRAY THAT CONTAINS THE RAW SCORES.  AVG AND SD
C     CONTAIN THE AVERAGES AND STANDARD DEVIATIONS OF EACH
C     VARIABLE OF THE RAW DATA.
      READ(9,100) ROWS,TYPE,COL,FMTIN,FMTOUT
 100  FORMAT(I3,A1,/,I3,2(/,20A4))
      DO 55 J=1,COL
      DO 50 I=1,COL
  50  C(I,J)=0.0
      AVG(J)=0.0
  55  DSQ(J)=0.0
C     READ IN THE RAW SCORES.
      DO 10 I=1,ROWS
  10  READ(5,FMTIN)(D(I,J),J=1,COL)
C     COMPUTE THE AVERAGES AND STANDARD DEVIATIONS OF THE J VARIABLES.
      DO 80 J=1,COL
      DO 85 I=1,ROWS
  85  AVG(J)=AVG(J)+D(I,J)
  80  AVG(J)=AVG(J)/ROWS
      DO 12 J=1,COL
      DO 14 I=1,ROWS
  14  DSQ(J)=DSQ(J)+(D(I,J)-AVG(J))**2
  12  SD(J)=SQRT(DSQ(J)/(ROWS))
      DO 94 I=1,ROWS
      DO 98 J=1,COL
  98  Z(I,J)=(D(I,J)-AVG(J))/SD(J)
  94  CONTINUE
C     PRINT OUT THE STANDARD DEVIATIONS OF THE J VARIABLES ON ONE
C     LINE AND THEN PRINT THE AVERAGES ON THE NEXT LINE.
      WRITE(8,FMTOUT)(SD(I),I=1,COL)
      WRITE(8,FMTOUT)(AVG(I),I=1,COL)
C     DERIVE THE CORRELATION MATRIX.
      DO 20 I=1,COL
      DO 30 J=1,COL
      DO 40 K=1,ROWS
  40  C(I,J)=C(I,J)+Z(K,I)*Z(K,J)
  30  C(I,J)=C(I,J)/(ROWS)
  20  CONTINUE
C     WRITE OUT THE TITLE,NUMBER OF VARIABLES,1,0,AND FMTOUT
      WRITE(6,110) ROWS,TYPE
 110  FORMAT('SAMPLE SIZE=',I3,3X,'SUBSET ',A1)
C     PRINT OUT THE CORRELATION MATRIX
      DO 60  I=1,COL
  60  WRITE(6,FMTOUT)(C(I,J),J=1,COL)
C     PRINT OUT THE REDUCED CORRELATION MATRIX
      DO 70 I=2,COL
   70 WRITE (7,FMTOUT)(C(I,J),J=2,COL)
      STOP
      END
```

APPENDIX C

Regress

```
      REAL*4 COR(6,6),X(5,5),XINV(5,5),WK(54),OUT(20,9),BSQ
      REAL*4 B(10),D(400,20),C(400,2),AVG,SD,DSD(20),DAVG(20),A
      REAL*4 BETA(20,20),RSQ(20,3),BIAS(20)
      INTEGER*4 FMTOUT(20),FMT(20),NUM,TYPE
      INTEGER*4 FMTIN(20),IDGT,IER,CNT,IA/5/,N/5/,GBG(20),ROWS
      INTEGER*4 COLS,WIDE
C     DESCRIPTION OF VARIABLES:  COR IS THE CORRELATION MATRIX, X IS
C     THE MATRIX OF CORRELATIONS BETWEEN PREDICTOR VARIABLES.  XINV
C     IS THE INVERSE OF X.  D IS THE ARRAY OF RAW DATA SCORES.  BETA
C     IS THE ARRAY OF STANDARDIZED REGRESSION WEIGHTS.  B IS THE
C     ARRAY OF UNSTANDARDIZED REGRESSION WEIGHTS.  RSQ IS THE ARRAY
C     CONTAINING RSS(ERROR SUM OF SQUARES) AND R SQUARED.  FMT'S
C     ARE THE VARIABLE FORMATS USED TO READ IN AND PRINT OUT DATA.
      READ(5,100) GBG
 100  FORMAT(20A4)
      READ(9,120) ROWS,TYPE,COLS,FMTIN,FMTOUT,FMT
 120  FORMAT(I3,A1,/,I3,3(/,20A4))
      DO 5 I=1,ROWS
   5  READ(7,FMTIN)(D(I,J),J=1,COLS)
      NUM=COLS-1
C     READ IN THE STANDARD DEVIATIONS AND AVERAGES FOR THE VARIABLES.
C     THEN READ IN THE CORRELATION MATRIX.
      READ(4,FMTOUT)(DSD(I),I=1,COLS)
      READ(4,FMTOUT)(DAVG(I),I=1,COLS)
      DO 10 I=1,COLS
  10  READ(5,FMTOUT)(COR(I,J),J=1,COLS)
C     CREATE A BIAS WITH INCREMENTS OF 0.01 BETWEEN 0.0 AND 0.1,
C     AND INCREMENTS OF .1 BETWEEN 0.1 AND 1.0
      DO 20 CNT=1,20
      IF(CNT-11)30,30,40
  30  BIAS(CNT)=0.01*(CNT-1)
      GOTO 50
  40  BIAS(CNT)=0.1*(CNT-10)
  50  DO 60 I=1,NUM
      II=I+1
      DO 70 J=1,NUM
      JJ=J+1
  70  X(I,J)=COR(II,JJ)
  60  X(I,I)=X(I,I)+BIAS(CNT)
      IDGT=0
C     FIND THE INVERSE OF X, THEN CALCULATE THE BETA VALUES FOR
C     THAT LEVEL OF BIAS.
      CALL LINV2F(X,N,IA,XINV,IDGT,WK,IER)
      DO 80 I=1,NUM
      BETA(CNT,I)=0.0
      DO 90 J=1,NUM
      JC=J+1
```

```
      DO 90 J=1,NUM
      JC=J+1
   90 BETA(CNT,I)=BETA(CNT,I)+XINV(I,J)*COR(1,JC)
   80 CONTINUE
      BSQ=0.0
C     FIND RESIDUAL SUM OF SQUARES AND THE CORRELATION SQUARED.
      RSQ(CNT,1)=0.0
      RSQ(CNT,2)=1.0
      RSQ(CNT,3)=0.0
      DO 210 J=2,COLS
      JO=J-1
      RSQ(CNT,1)=RSQ(CNT,1)+BETA(CNT,JO)*COR(1,J)
  210 BSQ=BSQ+BETA(CNT,JO)**2
      RSQ(CNT,2)=RSQ(CNT,2)-RSQ(CNT,1)-BIAS(CNT)*BSQ
      DO 200 I=1,NUM
      II=I+1
C     CALCULATE THE B VALUES FOR THIS LEVEL OF K.
  200 B(I)=BETA(CNT,I)*DSD(1)/DSD(II)
      AVG=0.0
      SD=0.0
      A=DAVG(1)
      DO 220 J=1,NUM
      JJ=J+1
C     CALCULATE THE CONSTANT TERM A.  CALCULATE THE PREDICTORS
C     TIMES THE B VALUES MINUS A.  THIS WILL PRODUCE Y'.  Y IS
C     THE RAW SCORE ON THE DEPENDENT VARIABLE.  Y AND Y' ARE
C      CORRELATED TO PRODUCE THE VALUE OF R.
  220 A=A-B(J)*DAVG(JJ)
      DO 230 I=1,ROWS
      C(I,1)=D(I,1)
      C(I,2)=0.0
      DO 240 J=1,NUM
      JJ=J+1
  240 C(I,2)=C(I,2)+B(J)*D(I,JJ)
      C(I,2)=C(I,2)+A
  230 AVG=AVG+C(I,2)
      AVG=AVG/ROWS
      DO 250 I=1,ROWS
  250 SD=SD+(C(I,2)-AVG)**2
      SD=SQRT(SD/ROWS)
      DO 260 I=1,ROWS
  260 RSQ(CNT,3)=RSQ(CNT,3)+C(I,1)*C(I,2)
      RSQ(CNT,3)=RSQ(CNT,3)/ROWS
      RSQ(CNT,3)=(RSQ(CNT,3)-AVG*DAVG(1))/(SD*DSD(1))
      RSQ(CNT,3)=RSQ(CNT,3)**2
C     PRINT OUT THE VALUES OF THE BIAS,RSS,R SQUARED, AND THE BETA'S.
      WRITE(6,FMTOUT) BIAS(CNT),(RSQ(CNT,J),J=2,3),
     C  (BETA(CNT,J),J=1,NUM)
   20 CONTINUE
      STOP
      END
```

APPENDIX D

Plotted

```
C     THE VARIABLES REPRESENT THE FOLLOWING:
C     V1 - V5 ARE THE VARIABLES TESTED,   X IS THE VALUES OF K
C     R2 IS THE VALUE OF R SQUARED,   RS =1-R**2
C     XMIN,YMIN ARE THE MINIMUM VALUES OF K AND THE VALUE OF THE
C     SMALLEST VARIABLE RESPECTIVELY.   XORG,YORG ARE THE VALUE
C     OF THE (0,0) POINT OF THE GRAPH IN ABSOLUTE INCHES AS
C     MEASURED FROM THE LOWER LEFT CORNER OF THE PAPER.
C     XFACT,YFACT ARE THE FACTORS THAT THE VARIABLES ARE DIVIDED
C     BY TO PRODUCE THE RELATIVE COORDINATES.  VHGHT,THGHT ARE
C     THE HEIGHT OF THE NUMBERS AND LETTERS USED TO TITLE THE
C     AXES.  AXLTHX,AXLTHY ARE THE LENGTHS OF THE X AND Y AXES
C      IN INCHES.  XTIC,YTIC,XSPACE,YSPACE, ARE THE SPACINGS IN
C     INCHES OF THE TICK MARKS AND GRID LINES.  TXM,TYM,TXL,TYL
C     ARE THE ABSOLUTE COORDINATES OF THE FIRST LETTER OF EACH
C     OF THE TITLES..  NXDIV,NYDIV ARE THE NUMBER OF GRID DIVISIONS
C     MNT,LNT ARE THE NUMBER OF LETTERS IN THE TITLES.  MTITLE,
C     LTITLE ARE THE TITLES.
      REAL V1(20),V2(20),V3(20),V4(20),V5(20),X(20),R2(20),RS(20)
      REAL XMIN,XFACT,YMIN,YFACT,XORG,YORG,VHGHT,THGHT,AXLTHX,AXLTHY
      REAL XTICD,YTICD,XSPACE,YSPACE,TXM,TYM,TXL,TYL,FIGX,FIGY,LHGHT
      INTEGER*4 NXDIV,NYDIV,ITYP,MNT,LNT,FNT
      INTEGER*4 HOLD,FIGURE(3)/'FIGU','RE  ','  '/,
     C COEF(12)/'C','O','E','F','F','I','C','I','E','N','T','S'/
      INTEGER*4 SSE(3) /'S','S','E'/
      INTEGER*4 MTITLE(6)/'SAMP','LE S','IZE=',' ','SET ',' '/
      INTEGER*4 LTITLE(5) /'ERRO','R SU','M OF',' SQU','ARES'/
      XORG=6.0
      YORG=6.0
      VHGHT=0.20
      THGHT=0.5
      LHGHT=0.75
      MNT=21
      FNT=10
      LNT=20
      ITYP=1
      AXLTHX=20.0
      XMIN=0.0
      XFACT=0.05
      XTICD=1.0
      YTICD=1.0
      NXDIV=10
      XSPACE=2.0
      YSPACE=2.0
C     READ IN THE VALUES OF THE VARIABLES AND R SQUARED FOR EACH
C     VALUE OF K.  DO THIS FOR K=0.0 TO K=1.
      DO 10 I=1,20
        READ(5,20 ) X(I),RS(I),R2(I),V1(I),V2(I),V3(I),V4(I),V5(I)
 20   FORMAT(8F12.7)
 10   CONTINUE
C     READ IN THE MIN Y VALUE,THE FACTOR FOR THE Y AXIS,THE LENGTH
C     OF THE Y AXIS,THE NUMBER OF GRID DIVISIONS ALONG TH Y AXIS,
C     THE NUMBER OF CHARACTERS IN THE TITLE, AND THE TITLE.
      READ(7,30) YMIN,YFACT,AXLTHY
 30   FORMAT (F6.0,/,F6.0,/,F6.0)
      READ(7,40)NYDIV,MTITLE(4),MTITLE(6)
  40  FORMAT(I1,/,A4,/,A1)
      READ(7,45) FIGURE(3),HOLD
```

```
   45 FORMAT(A2,/,A2)
C     DRAW THE X AND Y AXES SUCH THAT K LIES ALONG THE X AND THE
C     COEFFICIENTS LIE ALONG THE Y AXIS.  DRAW A GRID THAT IS
C     SPACED AT ALTERNATE TICK MARKS.
      CALL PLTOFS(XMIN,XFACT,YMIN,YFACT,XORG,YORG)
      CALL PAXVAL(VHGHT)
      CALL PAXTTL(THGHT)
      CALL PAXIS(XORG,YORG,'K',-1,AXLTHX,0.0,XMIN,XFACT,XTICD)
      CALL PAXIS(XORG,YORG,' ',+1,AXLTHY,90.0,YMIN,YFACT,YTICD)
      CALL PGRID(XORG,YORG,XSPACE,YSPACE,NXDIV,NYDIV)
      TXM=5.0
      TYM=15.0
      DO 60 I=1,12
      CALL PSYM(TXM,TYM,THGHT,COEF(I),0.0,1,0)
      TYM=TYM-THGHT-0.2
   60 CONTINUE
C     DRAW IN THE CURVES WITH A DIFFERENT SYMBOL DRAWN FOR EACH
C     VARIABLE'S POINTS.
      CALL PLINE(X,V1,20,1,ITYP,5,1)
      CALL PLINE(X,V2,20,1,ITYP,2,1)
      CALL PLINE(X,V3,20,1,ITYP,1,1)
      CALL PLINE(X,V4,20,1,ITYP,11,1)
      CALL PLINE(X,V5,20,1,ITYP,3,1)
C     CHANGE THE PEN COLOR BACK TO BLACK.  COMPUTE THE VALUES OF
C     THE COORDINATES OF THE (0,0) POINT OF THE SUM OF SQUARES
C     OF ERROR GRAPH.  THESE COORDINATE VALUES ARE IN ABSOLUTE
C     TERMS RELATIVE TO THE LOWER LEFT EDGE OF THE PAPER WITH
C     ALL MEASUREMENTS GIVEN IN INCHES.
      YORG=YORG+17.0
      AXLTHY=10.0
      YFACT=0.1
      YMIN=0.0
      NYDIV=5
C     CREATE THE X AND Y AXES FOR THE SUM OF SQUARES OF ERRORS
C     GRAPH.  DRAW IN A GRID AND THEN DRAW IN THE CURVE.
C     THEN DRAW THE LINE.
      CALL PLTOFS(XMIN,XFACT,YMIN,YFACT,XORG,YORG)
      CALL PAXIS(XORG,YORG,'K',-1,AXLTHX,0.0,XMIN,XFACT,XTICD)
      CALL PAXIS(XORG,YORG,' ',+1,AXLTHY,90.0,YMIN,YFACT,YTICD)
      CALL PGRID(XORG,YORG,XSPACE,YSPACE,NXDIV,NYDIV)
      TXL=5.0
      TYL=29.0
      DO 70 I=1,3
      CALL PSYM(TXL,TYL,THGHT,SSE(I),0.0,1,0)
      TYL=TYL-THGHT-0.2
   70 CONTINUE
      CALL PLINE(X,RS,20,1,2,9,1)
      CALL PALPHA('ITALIC.3 ',0)
      CALL PSYM(8.0,2.0,LHGHT,MTITLE,0.0,MNT,0)
      CALL PSYM(12.0,3.0,LHGHT,FIGURE,0.0,FNT,0)
      CALL PSYM(8.0,19.0,LHGHT,LTITLE,0.0,LNT,0)
      FIGURE(3)=HOLD
      CALL PSYM(12.0,20.0,LHGHT,FIGURE,0.0,FNT,0)
      CALL PLTEND
      STOP
      END
```

APPENDIX E

Select

```
      REAL*4 BIAS(20),COR(6,6),X(5,5),ESS(2),KVAL(8),DSD(20),DAVG(20)
      REAL*4 D(400,20),C(400,2),XINV(5,5),WK(54),BSQ(30),RSQ(20,2)
      REAL*4 R,M(5),BETA(20,5),AVG,SD,SQ,S,K(31),B(30,5),A,Q
      REAL*4 H(20),MG(20),EV(5),LINV,TU,TK,STORE(15)
      INTEGER*4 FMTIN(20),FMTOUT(20),FMT(20),NUM,TYPE,ROWS,COLS
      INTEGER*4  GBG(20),IDGT,IER,IA/5/,N/5/,CNT,ADV,IDEN,HIDEN
      INTEGER*4 MGIDEN,IJOB/0/,FLAG/1/,MGID,HID
      INTEGER*4 OBS(5)/'OBSE','RVAT','ION ',' ',' '/
      INTEGER*4 HOERL(5)/'HOER','L   ',' ',' ',' '/
      INTEGER*4 MCG(5)/'MCDO','NALD',' - G','ALAR','NEAU'/
      INTEGER*4 KEN(5)/'KENN','ARD ',' ',' ',' '/,USED(8)
C     DESCRIPTION OF VARIABLES:  COR IS THE CORRELATION MATRIX, X IS
C     THE MATRIX OF CORRELATIONS BETWEEN PREDICTOR VARIABLES.  XINV
C     IS THE INVERSE OF X.  D IS THE ARRAY OF RAW DATA SCORES.  BETA
C     IS THE ARRAY OF STANDARDIZED REGRESSION WEIGHTS.  B IS THE
C     ARRAY OF UNSTANDARDIZED REGRESSION WEIGHTS.  RSQ IS THE ARRAY
C     CONTAINING RSS(ERROR SUM OF SQUARES) AND R SQUARED.  FMT'S
C     ARE THE VARIABLE FORMATS USED TO READ IN AND PRINT OUT DATA.
      READ(5,100) GBG
 100  FORMAT(20A4)
      READ(9,110) ROWS,TYPE,COLS,FMTIN,FMTOUT,FMT
 110  FORMAT(I3,A1,/,I3,3(/,20A4))
      NUM=COLS-1
      DO 10 I=1,20
  10  READ(7,FMTOUT) BIAS(I),(RSQ(I,J),J=1,2),(BETA(I,J),J=1,NUM)
      DO 20 I=1,COLS
  20  READ(5,FMTOUT)(COR(I,J),J=1,COLS)
      SQ=RSQ(1,1)/(ROWS-NUM-1)
      LINV=0.0
C     ADD THE BIAS TO THE DIAGONAL OF THE CORRELATION MATRIX X'X.
      DO 200 I=1,NUM
      II=I+1
      DO 210 J=1,NUM
      JJ=J+1
 210  X(I,J)=COR(II,JJ)
 200  EV(I)=0.0
C     PUT X'X IN SYMMETRIC STORAGE MODE, THEN FIND THE EIGENVALUES.
      CALL VCVTFS(X,IA,IA,STORE)
      IJOB=0
      CALL EIGRS(STORE,N,IJOB,EV,XINV,N,WK,IER)
C     FIND THE SUM OF THE INVERSES OF THE EIGENVALUES AND THE SUM OF
C     THE SQUARES OF THE BETAS.  Q IS A CONSTANT TERM BASED ON THE
C     BETAS AT K=0.
      DO 220 I=1,NUM
      LINV=LINV+1.0/EV(I)
 220  BSQ(1)=BSQ(1)+BETA(1,I)**2
      Q=BSQ(1)-SQ*LINV
C     FOR EACH OF THE TWENTY LEVELS OF BIAS(K), DO THE FOLLOWING:
C      ADD THE BIAS TO THE  DIAGONAL OF X'X.  FIND THE INVERSE OF
C     THIS NEW MATRIX.  FIND THE SUM OF THE SQUARES OF BETA.  COMPUTE
C     THE HOERL AND THE MCDONALD AND GALARNEAU FORMULAS.
      DO 600 CNT=1,20
      BSQ(CNT)=0.0
      DO 610 I=1,NUM
      II=I+1
      DO 620 J=1,NUM
      JJ=J+1
 620  X(I,J)=COR(II,JJ)
 610  X(I,I)=X(I,I)+BIAS(CNT)
      IDGT=0
      CALL LINV2F(X,N,IA,XINV,IDGT,WK,IER)
      DO 630 J=1,NUM
 630  BSQ(CNT)=BSQ(CNT)+BETA(CNT,J)**2
      R=SQRT(BSQ(CNT))
      MG(CNT)=ABS(BSQ(CNT)-Q)
```

```
      H(CNT)=1.0/SQRT(RSQ(CNT,1))
      DO 640 I=1,NUM
      M(I)=0.0
      DO 650 J=1,NUM
 650  M(I)=M(I)+XINV(I,J)*BETA(CNT,J)
 640  CONTINUE
      S=0.0
      DO 660 I=1,NUM
 660  S=S+BETA(CNT,I)*M(I)
      H(CNT)=H(CNT)*(R**2/S-BIAS(CNT)-(BIAS(CNT)*R)**2/RSQ(CNT,1))
 600  CONTINUE
      DO 30 I=1,30
      DO 40 J=1,NUM
  40  B(I,J)=0.0
  30  BSQ(I)=0.0
      K(1)=0.0
C     FOR THIRTY ITERATIONS(MAXIMUM), CALL THE ITERATIVE PROCEDURE TO
C     PREDICT THE BEST FIT K ACCORDING TO THE FORMULA OF HOERL AND
C     KENNARD.  IF THE NEXT ITERATION PRODUCES A K LESS THAN 0.0001
C     FROM THE LAST K, THEN STOP THE ITERATIVE PROCEDURE.
      DO 500 CNT=1,30
      ADV=CNT+1
      DO 50 I=1,NUM
      II=I+1
      DO 60 J=1,NUM
      JJ=J+1
  60  X(I,J)=COR(II,JJ)
  50  X(I,I)=X(I,I)+K(CNT)
      IDGT=0
      CALL LINV2F(X,N,IA,XINV,IDGT,WK,IER)
      DO 70 I=1,NUM
      DO 80 J=1,NUM
      JJ=J+1
  80  B(CNT,I)=B(CNT,I)+XINV(I,J)*COR(1,JJ)
  70  BSQ(CNT)=BSQ(CNT)+B(CNT,I)**2
      K(ADV)=NUM*SQ/BSQ(CNT)
      IDEN=CNT
      A=ABS(K(ADV)-K(CNT))
      IF(A.LE.0.0001) GOTO 800
      IF(BSQ(CNT).LE.0.000002) GOTO 800
      IF(K(ADV).GT.1.0) GOTO 800
 500  CONTINUE
C     COMPUTE THE BETAS,THE VALUE OF R SQUARED AND RSS FOR THE BEST
C     FIT K PRODUCED BY THE HOERL AND KENNARD PROCEDURE.
 800  ESS(1)=0.0
      ESS(2)=0.0
      DO 810 J=2,COLS
      JO=J-1
 810  ESS(1)=ESS(1)+B(IDEN,JO)*COR(1,J)
      ESS(1)=1.0-ESS(1)-K(IDEN)*BSQ(IDEN)
      READ(4,FMTOUT)(DSD(J),J=1,COLS)
      READ(4,FMTOUT)(DAVG(J),J=1,COLS)
      DO 820 I=1,ROWS
 820  READ(8,FMTIN)(D(I,J),J=1,COLS)
      SD=0.0
      AVG=0.0
      A=DAVG(1)
      DO 830 I=1,NUM
      II=I+1
 830  M(I)=B(IDEN,I)*DSD(1)/DSD(II)
      DO 840 I=1,NUM
      II=I+1
 840  A=A-M(I)*DAVG(II)
      DO 850 I=1,ROWS
      C(I,1)=D(I,1)
      C(I,2)=0.0
```

```
      DO 860 J=1,NUM
      JJ=J+1
 860  C(I,2)=C(I,2)+M(J)*D(I,JJ)
      C(I,2)=C(I,2)+A
 850  AVG=AVG+C(I,2)
      AVG=AVG/ROWS
      DO 870 I=1,ROWS
 870  SD=SD+(C(I,2)-AVG)**2
      SD=SQRT(SD/ROWS)
      DO 880 I=1,ROWS
 880  ESS(2)=ESS(2)+C(I,1)*C(I,2)
      ESS(2)=ESS(2)/ROWS
      ESS(2)=(ESS(2)-AVG*DAVG(1))/(SD*DSD(1))
      ESS(2)=ESS(2)**2
C     WRITE OUT THE 21 SETS OF K,RSS,R SQUARED, AND BETAS.  THESE
C     ARE TO BE ORDERED ACCORDING TO THE VALUE OF K.
      DO 900 I=1,20
      IF(K(IDEN).LT.BIAS(I).AND.FLAG.EQ.1) GOTO 910
 930  WRITE(11,FMTOUT)BIAS(I),RSQ(I,1),RSQ(I,2),(BETA(I,J),J=1,NUM)
      GOTO 900
 910  FLAG=0
      WRITE(11,FMTOUT)K(IDEN),ESS(1),ESS(2),(B(IDEN,J),J=1,NUM)
      GOTO 930
 900  CONTINUE
      DO 740 I=1,8
 740  USED(I)=1
      MGID=1
      HID=1
C     FIND THE MAXIMUM HOERL VALUE AND THE MINIMUM MCDONALD AND
C     GALARNEAU VALUE.CORRESPONDING K'S REPRESENT THE 'BEST FIT K'
C     ACCORDING TO THESE THEORIES.
      DO 700 I=1,20
      IF(MG(I).LT.MG(MGID)) MGID=I
 700  IF(H(I).GT.H(HID))HID=I
      IF(Q.LT.0.0) MGID=1
      CNT=0
C     THIS NEXT SECTION IS USED TO ORDER THE K'S SELECTED BY ALL THE
C     ABOVE PROCESSES.  IT ALSO READS IN K'S SELECTED BY OBSERVATION
C     OF THE RIDGE TRACE FOR THIS SAMPLE.  THE K'S ARE THEN PUT INTO
C     AN OUTPUT FILE ORDERED BY INCREASING VALUE.  EACH K HAS A
C     CORRESPONDING STATEMENT TELLING WHAT PROCEDURE WAS USED TO
C     SELECT THAT PARTICULAR VALUE OF K.  ALSO, THE ITERATIVE PROCEDURE
C     HAS THE ITERATION STEPS USED TO SELECT THAT K BEFORE DEFAULT
C     HAS OCCURED.
      DO 710 I=1,NUM
      READ(6,FMTOUT,END=720)KVAL(I)
 710  CNT=CNT+1
 720  CNT=CNT+1
      KVAL(CNT)=BIAS(HID)
      USED(CNT)=2
      CNT=CNT+1
      KVAL(CNT)=BIAS(MGID)
      USED(CNT)=3
      CNT=CNT+1
      KVAL(CNT)=K(IDEN)
      USED(CNT)=4
      DO 750 I=1,CNT
      FLAG=I
      DO 760 J=I,CNT
      IF(KVAL(J).GE.KVAL(FLAG)) GOTO 760
      FLAG=J
 760  CONTINUE
      TK=KVAL(I)
      TU=USED(I)
      KVAL(I)=KVAL(FLAG)
```

```
      USED(I)=USED(FLAG)
      USED(FLAG)=TU
750   KVAL(FLAG)=TK
      DO 770 I=1,CNT
      J=USED(I)
      GOTO(300,310,320,330),J
300   WRITE(10,380)KVAL(I),OBS
      GOTO 770
310   WRITE(10,380) KVAL(I),HOERL
      GOTO 770
320   WRITE(10,380) KVAL(I),MCG
      GOTO 770
330   WRITE(10,380) KVAL(I),HOERL,KEN,IDEN
380   FORMAT(F12.7,12X,2(5A4,1X),I4)
770   CONTINUE
      STOP
      END
```

APPENDIX F

Bias. Inv.

```
C     PROGRAM DESCRIPTION
C     ADDITION OF A UNIT TO THE DIAGONAL OF A MATRIX.
C     COMPUTES THE INVERSES & EIGENVALUES.
      REAL*4 A(5,5),B(5,5),C(5,5),WKA(54),SB(15),Z(5,5),D(5),WK(50),N
      INTEGER*4 ROWS/5/,COLS/5/
C     DESCRIPTION OF VARIABLES:
C     A - INITIAL MATRIX, B - DIAG BIAS ADDED TO A, C - INVERSE OF B
C     WKA - WORKAREA FOR INV., SB - SYMMETRIC STORAGE OF B,
C     D - EIGENVALUES
C     INPUT OF MATRIX A
      DO 2 I=1,ROWS
    2 READ(5,100) (A(I,J),J=1,COLS)
  100 FORMAT(6F12.7)
C     READ IN THE VALUE OF THE BIAS
 50   READ(7,60,END=200) N
 60   FORMAT(F12.7)
C     PREPARING B FOR ADDITION OF BIAS
      DO 4 J=1,COLS
      DO 4 I=1,ROWS
    4 B(I,J)=A(I,J)
C     ADDING BIAS TO DIAGONAL
      DO 6 K=1,ROWS
    6 B(K,K)=B(K,K)+N
C     INVERSE OF B BECOMES C (FULL STORAGE MODE)
      IDGT=0
      CALL LINV2F(B,ROWS,5,C,IDGT,WKA,IER)
C     CONVERTING B TO SYMMETRIC STORAGE
      CALL VCVTFS(B,ROWS,5,SB)
C     DETERMINIG EIGENVALUES
      IJOB=2
      CALL EIGRS(SB,ROWS,IJOB,D,Z,5,WK,IER)
C     OUTPUT OF MATRICES B,C,D
      WRITE(6,101)
  101 FORMAT('1',10X,'ADDITION TO DIAGONAL OF ORIGINAL CORR MATRIX',/)
      IOPT=-1
      CALL USWTFM(IOPT,0,0,ROWS,COLS,5,B)
      WRITE(6,102)
  102 FORMAT(//'0',10X,'INVERSE MATRIX',/)
      CALL USWTFM(IOPT,0,0,ROWS,COLS,5,C)
      WRITE(6,103)
  103 FORMAT(//'0',10X,'EIGENVALUES',/)
      CALL USWTFM(-1,0,0,ROWS,1,5,D)
      GOTO 50
 200  CONTINUE
      STOP
      END
```

APPENDIX G

Doublex

```
      REAL*4 D(400,20),C(400,2),BETA(20),CCSQ,K,KOLD,COR,AVG,SD,BSQ
      REAL*4 BSDC(20),BBSDC(20), BAVG(20),BSD(20),B(20),A,BBAVG(20),BBSD(20)
      INTEGER*4 COLS,ROWS,TYPE,FMT(20),FMTIN(20),FMTOUT(20),FLAG /0/
C     DESCRIPTION OF VARIABLES:
C     D IS THE ARRAY CONTAINING THE RAW SCORES.  C IS AN ARRAY
C     THAT WILL CONTAIN Y AND Y'.  BETA CONTAINS THE BETA WEIGHTS
C     FOR ANY ONE VALUE OF THE BIAS.  K IS THE BIAS ADDED.
C     AVG AND SD ARE THE AVERAGE AND STANDARD DEVIATIONS.
C     BAVG AND BSD ARE THE AVERAGES AND STANDARD DEVIATIONS OF THE
C     RAW SCORE VARIABLES.  B IS AN ARRAY THAT HOLD THE B VALUES
C     WHICH ARE COMPUTED FROM THE BETA VALUES.TYPE IS THE LETTER
C     DESIGNATION OF THE SET.  FMT IS A FORMAT VARIABLE.
      WRITE(6,100)
  100 FORMAT('1',30X,'DOUBLE CROSS-VALIDATION OF D SCORES')
C     READ NFILE TO GET THE SAMPLE SIZE AND LETTER DESIGNATION
C     ALSO READ IN THE FORMATS USED FOR INPUT AND OUTPUT OF DATA.
      READ (9,130) ROWS,TYPE,COLS,FMTIN,FMTOUT,FMT
 130  FORMAT(I3,A1,/,I3,3(/,20A4))
C     INPUT OF RAW SCORES AND THEIR STANDARD DEVIATIONS AND AVERAGES.
      DO 4 I=1,ROWS
    4 READ(5,FMTIN) (D(I,J),J=1,COLS)
      READ(11,FMTOUT)(BSD(I),I=1,COLS)
      READ(11,FMTOUT)(BAVG(I),I=1,COLS)
      READ (4,FMTOUT)(BBSD(I),I=1,COLS)
      READ(4,FMTOUT)(BBAVG(I),I=1,COLS)
C     STORING COL 1 OF D (THE VALUES OF Y)  IN C.
      DO 6 I=1,ROWS
    6 C(I,1)=D(I,1)
C     READ IN A VALUE OF K FROM KFILE WHICH CONTAINS THE SELECTED
C     VALUES OF BIAS TO BE USED TO COMPUTE DOUBLE CROSS-VALIDATION.
C     READ IN A LINE FROM RFILE.  RFILE CONTAINS ALL VALUES OF K
C     BETWEEN 0 AND 1 ALONG WITH THEIR CORRESPONDING BETA WEIGHTS.
      K=100.0
 210  KOLD=K
      READ(7,FMTOUT,END=500) K
      IF(ABS(K-KOLD).LT.0.000004) GOTO 210
 200  READ (8,FMT,END=600) (BETA(I),I=1,COLS)
      SD=0.0
      BSQ=0.0
      AVG=0.0
C     COMPUTE THE SD,AVG,AND SUM OF SQUARED VALUES OF THE BETA
C     WEIGHTS FOR THIS VALUE OF K.  WRITE THESE OUT TO UNIT 10.
      DO 20 I=2,COLS
   20 AVG=AVG+BETA(I)
      AVG=AVG/(COLS-1)
      DO 30 I=2,COLS
      BSQ=BSQ+BETA(I)**2
   30 SD=SD+(BETA(I)-AVG)**2
      SD=SQRT(SD/(COLS-1))
      WRITE(10,FMTOUT) BSQ,SD,BETA(1)
C     IF THE K VALUE SELECTED IS EQUAL TO THE K READ IN FROM RFILE,
C     THEN GO ON.  OTHERWISE, RETURN TO STATEMENT 200 AND READ IN
C     A NEW SET OF BETAS AND THEIR CORRESPONDING K VALUE.
      IF(ABS(BETA(1)-K).GT.0.000004) GOTO 200
      DO 14 I=2,COLS
```

```
C     COMPUTE B VALUES FROM THE BETAS READ IN.
  14  B(I)=BETA(I)*BSD(1)/BSD(I)
      AVG=0.0
      SD=0.0
      A=BAVG(1)
      DO 16 J=2,COLS
  16  A=A-B(J)*BAVG(J)
C     MULTIPLY THE VALUE OF B BY THE RAW SCORE AND SUM ACROSS DEPENDENT
C     VARIABLES.  PLACE THE RESULT IN C.  THIS IS THE VALUE OF Y'.
      DO 11 I=1,ROWS
      C(I,2)=0.0
      DO 10 J=2,COLS
  10  C(I,2)=C(I,2)+B(J)*D(I,J)
      C(I,2)=C(I,2)+A
  11  AVG=AVG+C(I,2)
      AVG=AVG/ROWS
      DO 13 I=1,ROWS
  13  SD=SD+(C(I,2)-AVG)**2
      SD=SQRT(SD/ROWS)
      COR=0.0
C     COMPUTE THE CORRELATION BETWEEN Y AND Y' WHICH ARE STORED IN C.
C     PRINT OUT THE VALUES OF THE CORRELATION AND ITS SQUARED VALUE.
      DO 40 I=1,ROWS
   40 COR=COR+C(I,1)*C(I,2)
      COR=COR/ROWS
      COR=(COR-AVG*BBAVG(1))/(SD*BBSD(1))
      CCSQ=COR**2
      WRITE(6,FMTOUT) BETA(1),CCSQ
      IF (FLAG) 600,210,200
 500  CONTINUE
      FLAG=1
      K=0.0
      GOTO 200
 600  CONTINUE
      STOP
      END
```

APPENDIX H

Plotter

```
C     THE VARIABLES REPRESENT THE FOLLOWING:
C     V1 - V5 ARE THE VARIABLES TESTED,   X IS THE VALUES OF K
C     R2 IS THE VALUE OF R SQUARED,   RS =1-R**2
C     XMIN,YMIN ARE THE MINIMUM VALUES OF K AND THE VALUE OF THE
C     SMALLEST VARIABLE RESPECTIVELY.   XORG,YORG ARE THE VALUE
C     OF THE (0,0) POINT OF THE GRAPH IN ABSOLUTE INCHES AS
C     MEASURED FROM THE LOWER LEFT CORNER OF THE PAPER.
C     XFACT,YFACT ARE THE FACTORS THAT THE VARIABLES ARE DIVIDED
C     BY TO PRODUCE THE RELATIVE COORDINATES.  VHGHT,THGHT ARE
C     THE HEIGHT OF THE NUMBERS AND LETTERS USED TO TITLE THE
C     AXES.  AXLTHX,AXLTHY ARE THE LENGTHS OF THE X AND Y AXES
C      IN INCHES.  XTIC,YTIC,XSPACE,YSPACE, ARE THE SPACINGS IN
C     INCHES OF THE TICK MARKS AND GRID LINES.  TXM,TYM,TXL,TYL
C     ARE THE ABSOLUTE COORDINATES OF THE FIRST LETTER OF EACH
C     OF THE TITLES..  NXDIV,NYDIV ARE THE NUMBER OF GRID DIVISIONS
C     MNT,LNT ARE THE NUMBER OF LETTERS IN THE TITLES.  MTITLE,
C     LTITLE ARE THE TITLES.
      REAL V1(21),V2(21),V3(21),V4(21),V5(21),X(21),R2(21),RS(21)
      REAL XMIN,XFACT,YMIN,YFACT,XORG,YORG,VHGHT,THGHT,AXLTHX,AXLTHY
      REAL XTICD,YTICD,XSPACE,YSPACE,TXM,TYM,TXL,TYL,FIGX,FIGY,LHGHT
      INTEGER*4 NXDIV,NYDIV,ITYP,MNT,LNT,FNT
      INTEGER*4 HOLD,FIGURE(3)/'FIGU','RE  ','  '/
      INTEGER*4 EST /'EST.'/, HEX/Z827D8240/
      INTEGER*4 MTITLE(6)/'SAMP','LE S','IZE=',' ','SET ',' '/
      INTEGER*4 LTITLE(5) /'ERRO','R SU','M OF',' SQU','ARES'/
      XORG=6.0
      YORG=6.0
      VHGHT=0.20
      THGHT=0.5
      LHGHT=0.75
      MNT=21
      FNT=10
      LNT=20
      ITYP=1
      AXLTHX=20.0
      XMIN=0.0
      XFACT=0.05
      XTICD=1.0
      YTICD=1.0
      NXDIV=10
      XSPACE=2.0
      YSPACE=2.0
C     READ IN THE VALUES OF THE VARIABLES AND R SQUARED FOR EACH
C     VALUE OF K.  DO THIS FOR K=0.0 TO K=1.
      DO 10 I=1,21
      READ(5,20) V1(I),V2(I),X(I)
   20 FORMAT(3F12.7)
  10  CONTINUE
C     READ IN THE MIN Y VALUE,THE FACTOR FOR THE Y AXIS,THE LENGTH
C     OF THE Y AXIS,THE NUMBER OF GRID DIVISIONS ALONG TH Y AXIS,
C     THE NUMBER OF CHARACTERS IN THE TITLE, AND THE TITLE.
```

```
      READ(7,30) YMIN,YFACT,AXLTHY
 30   FORMAT (F6.0,/,F6.0,/,F6.0)
      READ(7,40)NYDIV,MTITLE(4),MTITLE(6)
  40  FORMAT(I1,/,A4,/,A1)
      READ(7,45) FIGURE(3),HOLD
  45  FORMAT(A2,/,A2)
C     DRAW THE X AND Y AXES SUCH THAT K LIES ALONG THE X AND THE
C     COEFFICIENTS LIE ALONG THE Y AXIS.  DRAW A GRID THAT IS
C     SPACED AT ALTERNATE TICK MARKS.
      CALL PLTOFS(XMIN,XFACT,YMIN,YFACT,XORG,YORG)
      CALL PAXVAL(VHGHT)
      CALL PAXTTL(THGHT)
      CALL PAXIS(XORG,YORG,'K',-1,AXLTHX,0.0,XMIN,XFACT,XTICD)
      CALL PAXIS(XORG,YORG,' ',+1,AXLTHY,90.0,YMIN,YFACT,YTICD)
      CALL PGRID(XORG,YORG,XSPACE,YSPACE,NXDIV,NYDIV)
      TXM=4.0
      TYM=12.0
      CALL PSYM(TXM,TYM,THGHT,EST,0.0,4,0)
C     DRAW IN THE CURVES WITH A DIFFERENT SYMBOL DRAWN FOR EACH
C     VARIABLE'S POINTS.
      CALL PLINE(X,V1,21,1,ITYP,5,1)
      CALL PALPHA('ITALIC.3 ',0)
      CALL PSYM(8.0,2.0,LHGHT,MTITLE,0.0,MNT,0)
      CALL PSYM(12.0,3.0,LHGHT,FIGURE,0.0,FNT,0)
      CALL PALPHA('GREEK.2 ',0)
      TYM=11.0
      CALL PSYM(TXM,TYM,THGHT,HEX,0.0,3,0)
      CALL PLTEND
      STOP
      END
```

REFERENCES

Banerjee, K. S. & Carr, R. N. A comment on ridge regression. Biased estimation for non-orthogonal problems. Technometrics, 1971, 13(4), 895-898.

Barcikowski, R. S., & Stevens, J. P. Studying canonical analysis: A reply to Thorndike's comments. Multivariate Behavioral Research, 1976, 11(2), 255-258.

Blalock, H. M. Jr. Correlated independent variables: The problem of multicollinearity. Social Forces, 1963, 42, 233-237.

Bolding, J. T., & Houston, S. R. A fortran computer program for computation of ridge regression coefficients. Educational and Psychological Measurement, 1974, 34, 151-152.

Conniffe, D. & Stone, J. A critical view of ridge regression. The Statistician, 1974, 22(2), 181-187.

Claudy, J. A comparison of five variable weighting procedures. Educational and Psychological Measurement, 1972, 32(2), 311-327.

Cohen, J. Multiple regression as a general data - analytic system. Psychological Bulletin, 1968, 70(6), 426-443.

Cohen, J., & Cohen, P. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. New York: John Wiley & Sons, 1975.

Comprehensive Tests of Basic Skills Form S All Levels, Technical Bulletin No. 1. Monterey, California, CTB/McGraw-Hill, 1974.

Comprehensive Tests of Basic Skills Forms S and T All Levels, Technical Bulletin No. 2. Monterey, California, CTB/McGraw-Hill, 1977.

Cramer, E. M. Brief report: The use of highly correlated predictors in regression analysis, Multivariate Behavioral Research, 1974, 241-243.

Cureton, E. E. Multivariate psychological statistics. Unpublished manuscript, The University of Tennessee, Knoxville, 1962.

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

Dorans, N., Drasgow, F., and Tucker, L. R. Raw Score Weights and Standard Score Weights in Cross-Validation. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1978.

Douglass, H. J. The effects of sample size on prediction with three test weighing methods. Unpublished Doctoral Dissertation, Purdue University, Lafayette, Indiana, 1958.

Farebrother, R. W. The minimum mean square error linear estimator and ridge regression. Technometrics, 1975, 17(1), 127-128.

Farrar, D. E. and Glauber, R. R. Multicollinearity in regression analysis: The problem revisited. The Review of Economics and Statistics, 1967, 49, 92-107.

Fisher, R. A. The influence of rainfall on the yield of wheat at Rothemsted. Philosophical Transactions of the Royal Society of London, 1924, 213B, 89-142.

Gibbons, D. G. A Simulation Study of Some Ridge Estimators (GMR-2659). Warren, Michigan: General Motors Research Laboratories, 1978.

Golub, G. H., Heath, M., and Wahba, G. Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter. (Technical Report No. 491). Madison, Wisconsin: Department of Statistics, University of Wisconsin, 1977.

Gordon, R. A. Issues in multiple regression. The American Journal of Sociology, 1968, 73, 592-616.

Guilkey, D. K., & Murphy, J. L. Directed ridge regression techniques in cases of multicollinearity. Journal of The American Statistical Association, 1975, 70(352), 769-775.

Halinski, R., & Feldt, L. The selection of variables in multiple regression analysis. Journal of Educational Measurement, 1970, 7(2), 151-157.

Hawkins, D. M. Relations between ridge regression and eigenanalysis of the augmented correlation matrix. Technometrics, 1975, 17(4), 477-480.

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

Hoerl, A. E., & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970a, 12(1), 55-67.

Hoerl, A. E., & Kennard, R. W. Ridge regression: Applications to nonorthogonal problems. Technometrics, 1970b, 12(1), 69-82.

Hoerl, A. E., & Kennard, R. W. Ridge regression iterative estimation of the biasing parameter. Communications in Statistics -Theoretical Methods, 1976, 5, 77-88.

Hoerl, A. E. Personal communication, August 1979.

Kerlinger, F. N., & Pedhazur, E. J. Multiple Regression in Behavioral Research, New York: Holt, Rinehart, & Winston, 1973.

Klein, L. R., and Nakamura, M. Singularity in the equation systems of econometrics. Some aspects of the problem of multicollinearity. International Economic Review, 1962, 3(3), 274-299.

Kunce, J. Prediction and statistical overkill. Measurement and Evaluation in Guidance, 1971, 4(1), 38-42.

Lawless, J. F., & Wang, P. A simulation study of ridge and other regression estimators. Communications in Statistics - Theoretical Methods, 1976, 5, 307-323.

Lindley, D. V., & Smith, A. F. M. Bayes estimates for the linear model. Journal of the Royal Statistical Society, Series B, 1972, 34, 1-41.

Mallows, C. L. Some comments on Cp. Technometrics, 1973, 15, 661-675.

Marquardt, D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics, 1970, 12(3), 591-612.

Marquardt, D. W., & Snee, R. D. Ridge regression in practice. The American Statistician, 1975, 29(1), 3-20.

Marquette, J. F., & Dufala, M. M. An interactive approach to ridge regression. Multiple Linear Regression Viewpoints, 1978, 8(2), 1-7.

Mason, R. L., Gunst, R. F., & Webster, J. T. Regression analysis and problems of multicollinearity. Communications in Statistics, 1975, 4(3), 277-292.

Mayer, L. S., & Willke, T. A. On biased estimation in linear models. Technometrics, 1973, 15(3), 497-508.

McDonald, G. C., & Galarneau, D. I. A Monte Carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association, 1975, 70, 407-416.

McDonald, G. C. Discussion of ridge analysis following a preliminary test of the shrunken hypothesis. Technometrics, 1975, 17(4), 443-445.

McNeil, K., & Lewis, E. A rejoinder to prediction and statistical overkill. Measurement and Evaluation in Guidance, 1972, 5(2), 360-378.

McNeil, K. A., & Spaner, S. D. Brief report: Highly correlated predictor variables in multiple regression models. Multivariate Behavioral Research, 1971, 6, 117-125.

Newhouse, J. P., & Oman, S. D. An Evaluation of Ridge Estimators (R-716-PR). Santa Monica, Ca.: RAND, April, 1971.

Obenchain, R. L. Ridge analysis following a preliminary test of the shrunken hypothesis. Technometrics, 1975, 17(4), 431-441.

Pedhazur, E. Analytic methods in studies of educational effects in Kerlinger, F. (Ed.), Review of Research in Education (Vol. 3). Itasca, Illinois: F. E. Peacock Publishers, Inc., 1975.

Price, B. Ridge regression: Application to nonexperimental data. Psychological Bulletin, 1977, 84(3), 759-766.

Simon, C. W. Methods for Improving Information from "Undesigned" Human Factor Experiments. Arlington, Virginia: Air Force Office of Scientific Research, 1975. (ERIC Document Reproduction Service No. ED 113 377).

Theobald, C. M. Generalizations of mean square error applied to ridge regression. Journal of the Royal Statistical Society, Series B, 1974, 36, 103-106.

Walton, J. M., Newman, I., & Fraas, J. W. Ridge regression: A panacea? Multiple Linear Regression Viewpoints, 1978, 8(2), 8-15.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill, Inc., 1971.

Zaitzeff, L. P. Target Background Scaling and Its Impact on the Prediction of Aircrew Target Acquisition Performance. Seattle: Boeing Company, Aerospace Group, Attack Subsystems. Technical Report No. D180-14156-1, December 1971.

AUTOBIOGRAPHICAL STATEMENT

I graduated third in my class from Kimball High School in Royal Oak, Michigan in 1966. Graduation from Wayne State University, Detroit, Michigan in December 1969 with a major in Sociology and minor in advanced Spanish was with high distinction. I became a member of Phi Beta Kappa and Alpha Kappa Delta, the National Honor Society in Sociology, Beta Chapter of Michigan.

During my four years as a fifth-grade teacher for Utica Community Schools in Michigan, I co-authored "Mini-guide: Living On Spaceship Earth" which was published in the Elementary Teacher's Edition of Scholastic Teacher. I studied anthropology in Guadalajara, Mexico and comparative education in Copenhagen, Denmark. While residing in California in 1975, I assisted in writing and editing a criterion referenced instructional program in the language arts for grades k-2.

I was awarded a Master of Education in Educational Evaluation and Research from Wayne State University in June 1975 and enrolled in the doctoral program in 1976. My graduate education was, in part, supported by Graduate Professional Scholarships and a University Graduate Fellowship awarded by Wayne State University.

Since 1976, I have worked as a Research Assistant in substance abuse for the Detroit Hospital Drug Treatment Program, Evaluation Specialist for the Center for Professional Growth and Development, and served as an Evaluation Consultant to the Detroit Health Department and the Oakland County Health Department. I am currently Associate Director of the Training Institute for Desegregated Education at Wayne State University.