AN ABSTRACT OF THE THESIS OF

<u>Ali Ahmed Rahuma</u>   for the degree of   <u>Master of Science</u>

in <u>Agricultural and Resource Economics</u> presented on <u>November 17, 1977</u>

Title: <u>APPLICATION AND EVALUATION OF RIDGE REGRESSION</u>

       <u>TO SELECTED EMPIRICAL ECONOMIC MODELS</u>

Abstract approved <u>Redacted for privacy</u>
<div align="center">W. G. Brown</div>

When the assumption of independency between the independent variables in a regression equation is violated, and in spite of Ordinary Least Square (OLS) procedure yielding estimates that are best linear unbiased (BLUE), the presence of multicollinearity can have severe effects on the estimation of the coefficients and on variable selection techniques.

In an attempt to cope with the problem of multicollinearity in a regression analysis, a number of techniques have been proposed. Among these techniques, the deletion of variables most affected by multicollinearity, the use of prior information models and the use of ridge regression are described and commented upon. Advantages and disadvantages are stated in this study. The sources of multicollinearity, its harmful effects and several methods of its detection are presented and discussed.

The purpose of this study is to evaluate ridge regression as a

possible method in obtaining stable and reliable estimates when multi-collinearity effects are present in economic models. In order to do that, four different economic models suffering from various degrees of multicollinearity have been chosen, and three versions of ridge regression estimators have been applied to these economic models.

Results of these studies indicate that the three ordinary ridge estimators, one proposed by Horel, Kennard and Baldwin (K_A), the second proposed by Lawless and Wang (K_B) and the third introduced by Dempster, Schatzoff and Wermuth (RIDGM) appear to be effective in obtaining stable and reliable estimates when multicollinearity is a serious problem in a regression analysis.

Although the estimates obtained by ridge regression in this study are biased, they appear to have desirable characteristics, are all stable and have the correct sign, and lack the symptoms of nonsense regression that was observed when OLS was used to estimate these models with highly multicollinear data.

Application and Evaluation of Ridge Regression
to Selected Empirical Economic Models

by

Ali Ahmed Rahuma

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

June 1978

APPROVED:

Redacted for privacy

Professor of Agricultural and Resource Economics
in charge of major

Redacted for privacy

Head of Department of Agricultural and Resource Economics

Redacted for privacy

Dean of Graduate School

Date thesis presented   November 17, 1977

Typed by Susie Avery for Ali Ahmed Rahuma

ACKNOWLEDGMENT

TABLE OF CONTENTS

LIST OF TABLES

(List of Tables Continued)

APPLICATION AND EVALUATION OF RIDGE REGRESSION
TO SELECTED EMPIRICAL ECONOMIC MODELS

I. INTRODUCTION

A serious problem that can occur in a regression analysis is the presence of multicollinearity among the independent variables in a regression equation.

The problem is unavoidable in most economic relationships (1), due to the nature of the economic data available. These data are usually aggregated over time, or over geographical location (cities and counties). In any case, a researcher is usually interested in cases of dependency or interrelationships among variables rather than in states of independency.

When the assumption of independency[1] among explanatory variables in a regression equation is violated, and in spite of the Ordinary Least Square (OLS) procedure yielding estimates that are best linear unbiased estimates (BLUE), the presence of milticollinearity will destroy the quality of these estimates. This happens because the variances of the estimates are unreasonably large. It also results in rejecting variables for lack of statistical significance, or estimates of the wrong sign are obtained.

---

[1] Vectors $\underline{x}_1, \underline{x}_2, \ldots \underline{x}_p$ are linearly dependent if there exist non-zero constants, $a_1, a_2, \ldots a_p$ such that $\sum_{j=1}^{p} a_j x_j = 0$

Many researchers and economists concern themselves with the problem of multicollinearity in order to improve the situation. There are two ways that multicollinearity can be handled in a regression analysis. One way is to delete the variable most affected in an attempt to obtain more accurate and reliable estimates. There are some advantages and many disadvantages to this technique. Few economists have applied the ridge regression techniques to economic data (2) since most of the ridge studies are based on non-economic data.

The purpose of this study is to apply some ridge regression models which are shown to have promising results for improving estimates with multicollinearity present (10). These ridge regression models are different only in the choice of K, but they are taken from Hoerl, Baldwin and Kennard (11), Lawless and Wang (15), and Dempster, Shatzoff and Wermuth (5).

The plan of this paper is as follows: multicollinearity will be discussed in chapter one; in chapter two, some suggested remedies for multicollinearity will be presented; in chapter three, some models of ridge regression will be discussed; and chapter four will apply the ridge regression models to four different economic models discussed in the literature (2).

The use of alternatives to variable deletion is both possible, and reliable in some cases, and economists are coming to accept the technique more readily. It is being discovered increasingly how to

break the multicollinearity deadlock. Different techniques are discussed in past studies by Theil (19), Goldberger (8), and Brown (2), by which prior information can be incorporated into regression analysis.

Ridge regression, introduced by Hoerl and Kennard (1968), is among the techniques benefitting from prior information. Since then, economists and statisticians have been developing the technique: Lawless and Wang (16), Dempster, Schatzoff and Wermuth (5). When multicollinearity is present, ridge regression techniques result in estimates of the coefficients that are biased, but have smaller variances than that of Ordinary Least Square (OLS), as a result of introducing some bias into the ($X'X$) data matrix. Hoerl and Kennard (10), Lawless and Wang (6), Dempster (5), Brown (1) and others have considered ridge regression as a possible remedy for improving estimates when multicollinearity is a problem in regression analysis.

Multicollinearity

Multicollinearity is a situation in which one cannot single out or separate the effects of two or more explanatory variables, because they tend to move together in sample data. It is the degree of interdependency between two or more theoretically independent variables.

Because of the nature of the economic and social relationships whose data tend to move together, we measure parameters of dependency and not independency. Multicollinearity is then an unavoidable case.

Definition

Multicollinearity is defined by Farrar and Glauber (7) as "a departure from Orthogonality in an independent variable set." With this definition in hand, two things can now be separated:

(1) The nature of the problem of multicollinearity: the lack of independence between explanatory variables in the model.

(2) The effects and locations of multicollinearity: to what degree variables are affected by multicollinearity.

Sources of Multicollinearity

There are many sources of multicollinearity. Only four are well known in the literature (see 1, 2, 6, 17, 19 and 20).

(1) Aggregation of data.

(2) Specification of the model.

(3) Sampling technique.

(4) Some physical limitations on the model.

Each of these sources presents a problem in the detection of multicollinearity and its effects.

Aggregation of data: a common practice, resulting from the bulkiness of the information and the difficulty of handling each individual factor of this information. Economists and researchers very often depend in their analyses on aggregated data based on

available sources, such as an Agricultural or Commercial census. Aggregation will result in loss of information, and it will create dependency between economic variables with multicollinearity arising as a result.

Specification of the model: in many cases, especially in medical and industrial sections, a model may be specified in such a way that it becomes over-defined, a situation in which one has more variables in the model than observations. There is then a serious problem.

Sampling technique: this technique results in multicollinearity when the investigator only samples a subspace of the experiment space. An omission of important information is caused, often leading to dependency between one or more of the explanatory variables in the model.

Physical limitation of the model: the experimenter is confronted with some limitation in the data such as in the presence of fixed proportionality, or simply that some information is not available.

It is important to recognize the different sources of multicollinearity in an attempt to remove it or to assist in the discovery of some workable solutions to the problem.

In estimating parameters of an economic model in the form

$$y = x\beta + u$$

where y is (n x 1) vector of dependent variables, x is (n x p) matrix of theoretically independent explanatory variables, β is (p x 1) vector

of coefficients and u is (n x 1) vector of terms, where $u \sim N(0, 2)$.

OLS estimates

$$\hat{\beta} = (x'x)^{-1} x'y$$

and

$$\text{Var-Cov}(\hat{\beta}) = \sigma(xx)^{-1}.$$

Two extreme cases may be recognized.

(1) The orthogonal case where (xx) is an orthogonal matrix with zeros in all the off diagonal elements because of perfect independence between each explanatory variable. Since

$$r_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

the OLS estimates are BLUE with minimum bias. $V(\hat{\beta}_i)$ is small and $\hat{\beta}_i$ is not sensitive to change in data or the specification of the model.

(2) The most drastic cases of multicollinearity where $(x'x)^{-1}$ does not exist, or is hard to obtain using the existing computational techniques. This is a case of perfect multicollinearity between two or more explanatory variables: the determinant of (x'x) is zero and (x'x) is not invertable. OLS estimates $(\hat{\beta})$ are not possible and the variance $V(\hat{\beta}_i)$ explodes at the limit.

Between these two extremes on the multicollinearity scale, there are degrees of multicollinearity. 'Harmful multicollinearity,' a term defined as "one that causes a wrong sign or symptoms of nonsense

regression" (7), concerns researchers the most, because of difficulties it creates in estimation and model specification. The problem is of estimation because the coefficients are sensitive to change in data, in specification. Wrong signs of some coefficients may mislead the researcher to a wrong decision of deleting these variables.

Detection of Multicollinearity

(1) A simple measure for the detection of multicollinearity is through the elements of the correlation matrix $(x'x)$. r_{ij}[2] is near unity when there is perfect collinearity and it is equal to zero when the two variables x_i and x_j are not collinear. Unfortunately, this measure is good only in the two factor case, and it is not useful when more than two factors are in the model.

(2) Another method which is widely practiced for the detection of multicollinearity is by regressing y on all the explanatory variables in the model except x_j. Compare R^2[3] with R_j^2 where x_j is excluded from the model. R^2 will be close to R_j^2 in cases of severe multicollinearity. Unfortunately this technique is not very effective in the detection of multicollinearity, because it will reflect

---

[2] $r_{ij} = \dfrac{\Sigma x_i x_j}{\sqrt{x_i^2}\,\sqrt{x_j^2}}$ where x_i and x_j are in deviation form from their respective means.

[3] R^2 = regression sum of squares/total sum of squares.

only the importance of x_j as a predictor and not necessarily that x_j is collinear with other variables in the model.

(3) Another simple measure for the detection of multicollinearity when (xx) is standardized, is that

$$\text{Det } (x'x) = |x'x| = 0$$

when exact relationships are present between two or more columns of $(x'x)$ and $|x'x| = 1$ indicates orthogonal columns of the data matrix.

(4) Different methods for the detection of multicollinearity are suggested by Webster, Grunst and Mason (17). The idea of this test is to utilize the smallest Eigenvalues (λ_i min) and the corresponding Eigenvector V.

The closer λ_i is to zero, the stronger the linear relationship among the columns or rows of (x) matrix.

The following is a more recent approach for the detection of multicollinearity.

(5) To decide how to detect multicollinearity and how to determine its location:

As before assume $y = x\beta + u$ is the general linear model with $\hat{\beta} = (x'x)^{-1} x'y$ and $\text{var} - \text{cov} (\hat{\beta}) = \sigma^2 (x'x)^{-1}$. Following Glauber and Farrar (7)

$(x'x)$ is the correlation matrix

$$R = (x'x) = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

$$r_{ij} = \frac{\sum\limits_{i=1}^{p} x_i x_j}{\sqrt{x_i^2} \sqrt{x_j^2}}$$

(x_i) is measured in a deviation form from its mean. The "variance inflation factor" or the diagonal elements of the inverted correlation matrix (x'x) is a good measure of the degree and location of multicollinearity. Let it be

$$C^{ii} = \frac{\left| (x'x)_{ii} \right|}{\left| (x'x) \right|}$$

where $(x'x)_{ii}$ denotes the correlation matrix excluding the ith variable x_i and (x'x) is the whole correlation matrix

$$1 \le C^{ii} \le \infty$$

C^{ii} the diagonal element or "variance inflation factor", equal to unity, when there is no multicollinearity. Since in this case, $(x'x)_{ii}$ will be the same as (x'x) and the ratio of

$$\frac{\left| (x'x)_{ii} \right|}{\left| (x'x) \right|} = 1.0$$

$c^{ii} = \infty$ when (x'x) is singular, and there is perfect collinearity between two or more explanatory variables in the variable set.

The bigger c^{ii} is, the more serious the problem of multicollinearity because $V(\hat{\beta}_i)$ will be inflated by this factor (c^{ii}) and the usual (t)[4/] test will reject more variables (x_i) as statistically insignificant due to the small t values. As $V(\hat{\beta}_i)$ increases t becomes small and may not reflect the true importance of x_i as an explanatory variable.

As seen from this test, the partial correlation matrix R is not a sufficient test for the detection of multicollinearity, because the latter is a question of degree relative to the overall level of association.

Consequences of Multicollinearity

(1) Because of the lack of independence among the variables in the set of explanatory variables in the model, the precision of the estimates falls due to the difficulty in separating individual variable effects from other variable effects. This difficulty will create large estimate errors and the variance of $\hat{\beta}_i$, $V(\hat{\beta}_i)$ will be unreasonably large. The following example will illustrate this point.

Assume the following: A model with two explanatory variables x_1, x_2

[4/]
$$t = \frac{\hat{\beta}_i - E(\hat{\beta}_i)}{S \cdot E(\hat{\beta}_i)}$$

$$(x'x) = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix} \qquad \text{and}$$

$$V(\hat{\beta}_i) = \sigma_u^2 (x'x)^{-1} = \frac{\sigma_u^2}{1-r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix}$$

in a situation near orthogonality with $r_{12} = 0.20$

$$V(\hat{\beta}_i) = \sigma_u^2 \begin{bmatrix} 1.04 & -.208 \\ -.208 & 1.04 \end{bmatrix}$$

Now assume another case where $r_{12} = 0.90$. Due to multi-collinearity, $V(\hat{\beta}_i)$ will increase to

$$V(\hat{\beta}_i) = \sigma_u^2 \begin{bmatrix} 10 & -9 \\ -9 & 10 \end{bmatrix}$$

i.e. $V(\hat{\beta}_i)$ increased by a multiple of about (10).

(2) Estimates obtained by OLS will be very sensitive to change in data, i.e., an addition of more observations may result for instance in a change in variance of the estimates. This can be seen by

$$V(\hat{\beta}_i) = \sigma_u^2 (x'x)^{-1}$$

in addition to the effect of multicollinearity on $(x'x)^{-1}$, σ_u^2 also will be affected,

$$\sigma^2_u = \frac{1 - R^2}{N - p}$$

a change in data, as in the case of adding more observations, N will be increased and σ^2_u will be more stable.

(3) Estimates obtained by OLS when multicollinearity is present will be sensitive to model specification. For example, let the model be specified to include another explanatory variable x_3, and let one assume (for simplicity) that $r_{12} = r_{13} = r_{23} = r$. Then

$$(x'x) = \begin{bmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{bmatrix}$$

$$V(\hat{\beta}_i) = \sigma^2_u (x'x)^{-1} = \frac{\sigma^2_u}{2r^3 - 3r^2 + 1} \begin{bmatrix} 1-r^2 & r^2-r & r^2-r \\ r^2-r & 1-r^2 & r^2-r \\ r^2-r & r^2-r & 1-r^2 \end{bmatrix}$$

$V(\hat{\beta}_1)$ for example $= V(\beta_i) = \sigma^2_u \dfrac{1-r^2}{2r^3-3r^2+1}$ instead of

$$V(\beta_i) = \sigma^2_u \left(\frac{1}{1-r^2}\right) \quad \text{when } p = 2.$$

When $r = 0.9$

$$V(\hat{\beta}_i) = \sigma^2 (5.263) \text{ for } p = 2$$

$$V(\hat{\beta}_i) = \sigma^2 (6.786) \text{ for } p = 3,$$

i.e. $V(\hat{\beta}_i)$ increased by about 29% when changing the specification

of the model to include another explanatory variable.   Although

$$(x'x) = R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ r_{p1} & r_{12} & \cdots & r_{pp} \end{bmatrix}_{n,n}$$

r_{ij} may be small, but $(x'x)^{-1}$ may not exist and $\left| (x'x) \right| = 0$.

Glauber writes, "As viewed here, multicollinearity is a property of the independent variable set alone. No account whatever is taken of the extent, or even the existence, of dependence between y and x."

II. SUGGESTED SOLUTIONS TO MULTICOLLINEARITY

In this section, some suggested solutions for the removal of multicollinearity will be discussed. Some of these options are:

(1) Deletion of variables proven to be highly affected by multicollinearity.

(2) Use of prior information to estimate regression coefficients.

(3) The Theil-Goldberger mixed model.

(4) Use of Ridge Regression.

Deletion of Variables

As a consequence of multicollinearity, some researchers may choose to drop one or more explanatory variables and specify the model in terms of the remaining variables that were not affected by multicollinearity. This practice is useful either to obtain OLS estimates which are otherwise not possible because of the singularity[5] of (x'x), or to obtain more precision in the estimates by reducing the error created by interdependency between a set of explanatory variables in the model, i. e., to obtain small $(\hat{\beta}_i)$. Although this solution has been popular, many econometricians are aware of the bias it introduces to the model, or what is known as "specification errors."

---

[5] A matrix (x), t is singular if the Det(x) = $|x|$ = 0

Since many types of error can be categorised as specification error, only one type that results from the omission of one or more relevant variable, will be discussed.

Let a true linear model be estimated in the form

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

Where (y) is a dependent variable, x_1, x_2, and x_3 are theoretically independent variables. u is an error term. (All variables are measured in a deviation from their respective means.)

Let one assume that instead of estimating the whole model, the researcher chooses to drop x_2 and x_3 and to fit the model in terms of x_1 only.

$$y = \beta_1 x_1 + u$$

The ordinary least square estimates of the coefficient of (x) will be

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_1 y}{\sum x_1^2}$$

Contrary to the fact that this estimate $(\hat{\beta}_1)$ ignores completely the effect of x_2 in the estimate of the coefficient of x_1, the true estimate of (β_1), as expected under the full model will be

$$E(\hat{\beta}_1) = \beta_1 + \hat{\beta}_2 \frac{\sum x_1 x_2}{\sum x_1^2} + \hat{\beta}_3 \frac{\sum x_1 x_3}{\sum x_1^2}$$

if $y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ is the true model and not the fitted

$$y = \beta_1 x_1 + u$$

$(\hat{\beta}_1)$ estimates will not be $\dfrac{\Sigma x_1 y}{\Sigma x_1^2}$ but a bias is introduced.

This bias is in the form of

$$\hat{\beta}_2 \frac{\Sigma x_1 x_2}{\Sigma x_1^2} + \hat{\beta}_3 \frac{\Sigma x_1 x_3}{\Sigma x_1^2}$$

and will depend on

(1) the coefficients of the dropped explanatory variables $(\hat{\beta}_2)$ and $(\hat{\beta}_3)$,

(2) the quantities:

$$\frac{\Sigma x_1 x_2}{\Sigma x_1^2} \quad \text{and} \quad \frac{\Sigma x_1 x_3}{\Sigma x_1^2} \ .$$

The bias will not be zero unless $\hat{\beta}_2$ and $\hat{\beta}_3$ are zeros because $\Sigma x_1 x_2$ or $\Sigma x_1 x_3$ will not be zero, and this is also the case for Σx_1^2.

This case can be generalised to K independent variables, as in the following

$$y = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u +$$

$$E(\beta_1) = \beta_1 + \beta_2 \frac{\Sigma x_1 x_2}{\Sigma x_1^2} + \beta_3 \frac{\Sigma x_1 x_3}{\Sigma x_1^2} + \ldots + \beta_k \frac{\Sigma x_1 x_k}{\Sigma x_1^2} \ .$$

So whenever an important variable is deleted, OLS estimates will be

biased and the degree of the bias will depend upon the relative share

of the deleted variable in explaining the regression relationships, i.e.,

on $\hat{\beta}_1$, $\hat{\beta}_3$, ... etc. For more discussion on the general case see

Johnston (14).

There are only trivial cases where no bias is introduced from

deleting the variable. In such a case, the deleted variables are irrele-

vant, i.e., where $\hat{\beta}_i$ of the x_i omitted, is zero. But this case

is unlikely because if x_i is irrelevant, it will not be in the specifica-

tion of the model in the first place.

Although OLS is a biased procedure in cases of deleted variables,

it still produces a minimum variance

$$V(\beta_i) \ + \frac{\sigma_x^2}{\Sigma x_i^2}$$

But the bias created in the estimates can sometimes outweigh the

reduction in variance of $\hat{\beta}_i$, $V(\hat{\beta}_i)$

if $\qquad\qquad$ MSE (β_i)

is defined as

$$MSE\,(\hat{\beta}_i) \ = \ var\,(\hat{\beta}_i) + Bias^2\,(\hat{\beta}_i) \quad [6/]$$

MSE $(\hat{\beta}_i)$ of the full model < MSE $(\hat{\beta}_i)$ with incomplete model due to

the bias^2 factor introduced to the system.

Roa and Miller (18) pointed out that researchers should not be

---

[6/] $Bias\,(\hat{\beta}_i) \ = \ E(\hat{\beta}_i - \beta)$

discouraged by the reduction in variances of the estimates they obtain

from deleting variables because this reduction should be viewed in

terms of MSE $(\hat{\beta}_i)$ and not just in terms of the $V(\hat{\beta}_i)$.

The Use of Prior Information

Most econometricians and applied researchers agree that addi-

tional information is needed to lessen difficulties of multicollinearity.

Various methods of adding prior information into regression

analysis to help account for multicollinearity may be used.   Prior

information could be in the form of:

(1) Exact prior information where the investigator knows

exactly some or all of the coefficients and their standard errors, either

from previous studies or on a theoretical basis.   This information

can be incorporated into the regression analysis to ameliorate the

difficulties of multicollinearity.

(2) Subjective probability estimates of the coefficients and

their variances may be used similarly to solve multicollinearity.

Information in the form of inequality is based on the researcher's

knowledge about the coefficients from previous experience.   It is in

the form of probabilities and inequalities such as

$$0 \leq \beta_i \leq 1.0$$

All this information can be used to help solve multicollinearity

problems.

The Theil-Goldberger Mixed Model

Statistical prior information can be incorporated into a regression analysis using a method suggested by Theil and Goldberger (20) and summarized as follows: Let the usual linear model expressed as $y = x\beta + u$ where y is (nxl) vector of the dependent variable, x is (n x p) matrix of independent variables, (β) is (n x 1) vector of parameters and (u) is (n x 1) vector of disturbances.

$$E(u) = 0, \quad E(u'u) = \sigma^2 In \ ,$$

$$\hat{\beta} = (x'x)^{-1} x'y \ , \qquad var(\hat{\beta}_1) = \sigma^2 (x'x)^{-1}$$

and the usual assumptions apply.

Let the statistical information given in the form

$$r = R\beta + V$$

apply, where (r) is a known (kxl) vector of estimates for $R\beta$, (R) is a (kxr) matrix of known fixed elements determining which parameters have prior information, and how this information is weighted, and (β) is (rxl) vector of fixed unknown parameters and V is (kxl) vector of prior information errors.

Also assume that (V) is distributed independent from (u),

$$E(y) = 0 \quad \text{and} \quad E(V'V) = \psi \ .$$

When incorporating prior information into the basic linear regression model, one obtains the mixed model in the form

$$\begin{bmatrix} y \\ r \end{bmatrix} = \begin{bmatrix} x \\ R \end{bmatrix} B + \begin{bmatrix} u \\ v \end{bmatrix}$$

where

$$E\begin{bmatrix} u \\ v \end{bmatrix} = 0 \quad \text{and} \quad E\begin{bmatrix} u \\ v \end{bmatrix}[u'v'] = \begin{bmatrix} \sigma^2 I n & 0 \\ 0 & \psi \end{bmatrix}$$

(The off diagonal elements were zeros because of the assumption of independency of the distribution of v from the distribution of u.)

Instead of OLS estimates $\hat{\beta} = (x'x)^{-1}x'y$

$$\hat{b} = (x'R')\begin{bmatrix} \sigma^2 I n & 0 \\ 0 & \psi \end{bmatrix}^{-1}\begin{bmatrix} x' \\ R \end{bmatrix}(x'R')\begin{bmatrix} \sigma^2 I n & 0 \\ 0 & \psi \end{bmatrix}^{-1}\begin{bmatrix} y \\ r \end{bmatrix}$$

When $\varphi = \dfrac{1}{\sigma^2}$, $\hat{b}$ becomes

$$\hat{b} = (\varphi x'x + R'\psi^{-1}R)^{-1}(\varphi x'y + R'\psi^{-1}r)$$

and

$$\text{var - cov}(\hat{b}) = (\varphi x'x + R'\psi^{-1}R)^{-1}.$$

Johnston (14) points out that the suggested Theil-Goldberger mixed model raises several questions.

(1) Is $\hat{b}$ BLUE only in terms of the combined information, namely the prior information and the sample information (or in terms of the model V and x from the sample, and r and R from prior knowledge)?

(2) The Theil-Goldberger method was designed to cope with apparent conflict between sample and prior information and as a

result, $\hat{b}$ will be a weighted average of the two sets of information, and it makes it difficult to separate just how much both sample and prior information each contribute to the estimate.

(3) The formula for estimating $\hat{b}$ contains the term $(\varphi - \frac{1}{\sigma^2})$ and σ^2 is not usually known from the sample information.

The use of $\overset{\wedge}{\sigma}{}^2$ as an estimate for σ^2 or as recommended by Theil

$$\overset{\wedge}{\sigma}{}^2 = \frac{(y'y - yx((x'x)^{-1}x'y)}{n - k}$$

will have only asymptotic distribution and has many limitations in generalization.

Sometimes the estimates are affected by dominancy of prior information and the sample information contribution is very limited. When this happens, the whole purpose of the regression analysis is undermined.

Prior Information in the Form of Inequality Restraints

G. G. Judge and T. Takayama (15) proposed a method to incorporate prior information in the form of inequality restraints that cannot be utilized with the Theil-Goldberger mixed model technique. This method can be summarized using their notation as follows

$$y = x\beta + u, \quad E(u) = 0 \quad \text{and} \quad E(u'u) = \sigma^2 In$$

all as designed before

$$\hat{\beta} = (x'x)^{-1} x'y \quad \text{and} \quad V(\hat{\beta}) = \sigma^2 (x'x)^{-1}.$$

When prior information is in the form of equality restraints, exact linear relationship is in the form $r = R\beta$ where R is (jxk) constant known coefficient matrix, r is (jx1) known vector. To find a vector $\hat{b}$ which minimizes $u'u = (y-x\beta)'(y-x\beta)$ and is subject to $r - RB = 0$ a solution by the lagrangean method yields

$$\hat{b} = \beta + (x'x)^{-1} R' [R(x'x)^{-1} R']^{-1} (r-R\beta)$$

$\hat{b}$, the restricted estimate, is the best linear unbiased estimator (BLUE), and a linear function of y and r.

When the prior information is not exact, then $r = RB + V$ which is the case discussed by the Theil and Goldberger mixed model.

Assume that the following constraints prevail

$$\begin{bmatrix} I_1 \\ -I_1 \end{bmatrix} (\beta_1) \leq \begin{bmatrix} r_1^u \\ -r_1^s \end{bmatrix} \text{ for } \begin{bmatrix} r_1^u \\ r_1^s \end{bmatrix} \begin{matrix} > 0 \\ \geq 0 \end{matrix} \text{ or } 0 \leq r_1^3 \leq \beta_1 \, r_1^u \qquad (1)$$

$$\begin{bmatrix} I_2 \\ -I_2 \end{bmatrix} (\beta_2) \leq \begin{bmatrix} r_2^u \\ -r_2^s \end{bmatrix} \text{ for } \quad 0 \leq r_2^s \leq \beta_2 \leq r_2^u \qquad (2)$$

$$\begin{bmatrix} I_3 \\ -I_3 \end{bmatrix} (\beta_3) \leq \begin{bmatrix} r_3^u \\ r_3^s \end{bmatrix} \text{ for } \quad 0 \leq r_3^s \leq \beta_3 \leq r_3^u \qquad (3)$$

Where r_i^u and r_i^s, $i = 1, 2, 3$ are known vectors of upper and lower bound constraints on the unknown coefficients in the ith set β_i.

I_i for $i = 1, 2, 3$ is the identity matrix with a rank equal to the number of elements included in the parameter vector β_i. To find a $\hat{b}$ that minimize $u'u = (y - x\beta)'(y - x\beta)$ subject to the given constraints 1, 2, 3. The mathematical solution using Lagragean multipliers can be seen in Judge and Takayama (15).

Use of Ridge Regression

A different way of incorporating prior information into the regression analysis is through the use of "ridge regression." The basic idea of ridge regression is that multicollinearity is moderated or alleviated by augmenting the main diagonal elements of the correlation matrix by small positive quantities.

Following Hoerl and Kennard's (10) notation, assume a linear model in a deviation form $y = x\beta + u$

where

y denotes (nxl) vector of dependent variable,

x is (nxp) matrix of explanatory variables. β is (pxl) vector of parameters, and u is (nxl) vector of disturbances.

Assume u has $E(u) = 0$

$$E(u'u) = \sigma^2 (In) = V$$

is positive definite. [7]

---

[7] V is a positive definite matrix if it is non-singular, has positive eigenvalues and a positive determinant.

Let $(x'x)$ denote the correlation matrix of the explanatory variables. $\hat{\beta}^*$, the standardized ridge regression estimate of the coefficients is

$$\hat{\beta}^* = (x'x + KI)^{-1} x'y$$

$$\hat{\beta}^* = \beta + \gamma + (x'x + KI)^{-1} x'u$$

and variance-covariance of $\hat{\beta}^*$ is

$$\text{var - cov}(\hat{\beta}^*) = (x'x + KI)^{-1} x'x (x'x + KI)^{-1}$$

The non-standardized $\hat{\beta}^*$ estimate in terms of the corrected sums and cross-products is

$$\hat{\beta}^* = (x'x + K\lambda)^{-1} x'y$$

where $(x'x)$ is the $(p \times p)$ matrix of mean corrected sum of squares, and cross-products, and

$$\lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$$

the diagonal matrix of order (p) consisting of the sums of squares.

$$\text{var - cov}(\hat{\beta}^*)$$

non-standardized is

$$\text{var - cov}(\hat{\beta}^*) = \sigma^2 (x'x + K\lambda)^{-1} x'x (x'x + K\lambda)^{-1}$$

Since ridge regression is a biased estimator, the expected bias of it is obtained from

$$E(\hat{\beta}^*) = (x'x + K\lambda)^{-1}(x'x)\beta = (x'x + K\lambda)^{-1}[(x'x + K\lambda) - K\lambda]\beta$$

$$= (x'x + K\lambda)^{-1}(x'x + K\lambda)^{-1}\beta - (x'x + K\lambda)^{-1}(K\lambda)\beta$$

$$= \beta - K(x'x + K\lambda)^{-1}\lambda\beta$$

where the bias of (β^*) is $-K(x'x + K\lambda)^{-1}\lambda\beta$. For more details,

see Hoerl and Kennard (10) or Brown (1).

III. BIASED LINEAR ESTIMATIONS

In multiple linear regression models, ordinary least squares estimates (OLS) in the form $\hat{\beta} = (x'x)^{-1} x'y$ are in the class of unbiased linear estimators (BLUE). But as discussed earlier, this class of estimators can have very undesirable and unacceptable characteristics. This is because of the large variance inflations of the estimates of the coefficients (β_i) that can occur because of multicollinearity.

In this section, a class of linear biased estimators will be briefly discussed.

Stein Estimators

The Shrunken Least Square Estimators

Assume a standard linear model in the form $y = x\beta + u$ where as before y is (nxl) vector of dependent variables, x is (nxp) matrix of explanatory variables, β is (pxl) vector of parameters, and U is (nxl) vector of disturbances

$$u \sim N(0, \sigma^2 I) \quad i.e., \quad E(u) = 0, \quad E(u'u) = \sigma^2 In.$$

Let $(x'x)$ denote the correlation matrix, and $\hat{\beta} = (x'x)^{-1} x'y$ the OLS estimates of the coefficients as before.

$$\text{var - cov } (\hat{\beta}) = \sigma^2 (x'x)^{-1}$$

$\hat{\beta}$ has the distribution independent, normal, with a mean β and $\sigma^2(x'x)^{-1}$ variance, $\hat{\beta} \sim 1N \; N \; (\beta, \sigma^2(x'x)^{-1})$

Assume the following transformation for simplicity: Q is an Orthogonal symmetric positive definite matrix, such that

$$Q'Q \; = \; QQ' \; = \; I$$

$$Q'(x'x)Q \; = \; \Lambda \; = \; diag \; (\lambda_1, \; \lambda_2, \; \lambda_p)$$

where $(x'x)$ = correlation matrix and λ_i is the eigenvalue of $(x'x)$.

Let $Z \; = \; Qx$ and $\alpha \; = \; Q\beta$. Then the linear model is transformed to $y \; = \; Z\alpha + e$. As before

$$e \sim N(0, \; \sigma^2 In)$$

$$\alpha \sim N(\alpha_i, \; \sigma^2/\lambda_i)$$

$$\hat{\beta}^* \; = \; (x'x + KT)^{-1}x'y$$

T is a positive definite symmetric matrix and K is a non-negative scalar.

Stein estimators are obtained by choosing $T = (x'x)$ and $\hat{\alpha}^*_s$ Stein estimates will be equal to $\hat{\alpha}^*_s = f\alpha_i$ where $f = (1/1+K)$ that gives all the coefficients an equal weight, which is a function of (K).

Stein M[8], the most promising in the Stein class, in which

$$\alpha^*_{si} \quad N(f\alpha_i, \; f^2\sigma^2/\lambda_i)$$

K is chosen so that $\Sigma \lambda_i \hat{\alpha}_i^2/(w^2 + \sigma^2)$ is equal to its marginal

---

[8] Stein M was first introduced and evaluated by Dempster, Scharzoff and Wermuth in 1977.

expectation p where $\hat{\sigma}^2$ is substituted for σ^2 and $w^2 = \hat{\sigma}^2/K$.

Because of the fact that $f = (1/1+K)$ in Stein estimates is not a function of the sample in formation $(x'x)$, Stein estimators have very little to recommend them for alleviating the multicollinearity problem.

Ridge Regression

Based on the choice of K in ridge regression, a class of ridge estimators is defined, i.e., each choice of (K) determines a ridge estimator.

Three of this class of biased estimators will be discussed in this section.

Hoerl-Baldwin and Kennerd Ridge Regression

Hoerl-Baldwin and Kennerd (11) proposed a technique for choosing K. In this paper, K will be referred to as K_A.

As before, assume a general linear model $y = x\beta + u$ where x is (nxp) matrix of variables, y is $(nx1)$ vector of observations, β is $(px1)$ vector of parameters, and u is $(nx1)$ vector of disturbances. $U \sim N(0, \sigma^2 I)$ and $(x'x)$ is a correlation matrix.

Assume the same transformation applies

$$\hat{\beta}^* = (x'x + Q'KQ)^{-1} x'y$$

where K is a non-negative diagonal matrix and $(x'x) = Q'\lambda Q$. λ is the diagonal matrix of eigenvalues.

If $\alpha = Q\beta$ minimum MSE[9] (Mean Squared Error) is obtained

when $K_i = \sigma^2/\alpha_1^2$ and the harmonic mean of these individual is

$$\left(\frac{1}{K_A}\right) = \left(\frac{1}{\sigma^2}\right) \Sigma \alpha^2 = \frac{\alpha'\alpha}{p\sigma^2}$$

$$= \beta'\beta/p\sigma^2$$

$$K_A = p\sigma^2/\beta'\beta$$

and since σ^2 and β are unknown, then K_a can be estimated as

$$K_A = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}}$$

where $\hat{\beta}$ is OLS estimate for β, and $\hat{\sigma}^2$ is also an OLS estimate for σ^2.

K_A defines HBK ridge regression.

Lawless and Wang Ridge Estimators

This estimator is the same as HBK K_A estimators, except

that the coefficients are weighted according to the eigenvalues of

$(x'x)$. K_B is chosen in such a way that

$$K_B = \frac{p\sigma^2}{\sum\limits_{i=1}^{p} \lambda_i \alpha_i^2}$$

where p denotes the number of explanatory variables, λ_i is the

i^{th} eigenvalue of $(x'x)$, α_i is the OLS estimates of the coefficients

and $\hat{\sigma}^2$ is the OLS estimate of σ^2.

[9] $MSE(\hat{\beta}) = V(\hat{\beta}_i) + [E(\hat{\beta}_i - \beta_i)]^2$

In a study by Lawless and Wang (16) this choice of K_B proved to be more effective and produced better results than under the use of K_A.

RIDGM

RIDGM developed by Dempster, Shatzoff and Wermuth (5) is motivated by the Baysian interpretation of ridge. The observable least square estimate $\hat{\alpha}_i$ is marginally independently distributed $N(0, \sigma^2/\lambda_i)$.

K_M is chosen in such a way that the prior expectation of $\sum \hat{\alpha}_i^2/w^2 + \sigma^2/\lambda_i)$ is p where $\hat{\alpha}_i$ is the OLS estimated of the coefficients, $\hat{\sigma}^2$ is substituted for σ^2, λ_i is the ith eigenvalue of $(x'x)$, and w^2 is σ^2/K_M and p is the number of parameters. This specific choice of K yields a ridge regression referred to as RIDGM.

IV. APPLICATION OF SOME RIDGE REGRESSION MODELS FOR THE ESTIMATION OF ECONOMIC RELATIONSHIPS

Nature of Economic Data

Unfortunately, economists seldom have good quality data of primary source. Instead, one often turns to some secondary sources, such as the U.S. Census Agricultural and Business Survey reports. These provide a somewhat poor indication of many important inputs in our economic models.

Data are always available on an aggregate basis, such as counties and states, and one has to use these data in spite of the fact that aggregation results in a loss of information and a higher degree of multicollinearity.

To solve the data problems, it is better to return to the original data whenever possible. When such a return is not possible, different techniques to cope with the problem are tried.

In this section, several economic problems are analyzed using ridge models. These economic models are chosen mainly because:

(1) They suffer from multicollinearity, i.e., they lack independence in one or more of their explanatory variables.

(2) These models have some symptoms of nonsense regression or harmful multicollinearity, i.e., the coefficients may take wrong signs or have low t values.

(3) These models cannot be handled very well by ordinary least square regression, and one or more of their important explanatory variables is not statistically different from zero when OLS is used.

The analysis results of ridge regression will be viewed in terms of their performances on a theoretical basis and expectation, and how they compare with ordinary least square estimates.

Ridge Regression in Estimating the Marginal Value Productivity of Irrigation Water

Background

Using U.S. Census Agricultural data, Ruttan pioneered the estimation of total value product functions to derive an economic value for the water used in irrigating agriculture. The original formulation of the production function model included six important explanatory variables (1).

In functional relationship, the model was

$$y = g(x_1, x_2, x_3, x_4, x_5, x_6)$$

where

y denotes farm product sold;

x_1 is number of family and hired labor;

x_2 is number of tractors on the farm;

x_3 is value of livestock investment in the farm;

x_4 is acres of irrigated cropland;

Table 1. Estimated values for regression of county values of all farm products sold as a Cobb-Douglas function of inputs, OLS estimates for 25 Central Pacific counties, 1954.

	X_1	X_2	X_3	X_4	X_5	X_6	y
Means	4.03550944	3.589889	7.0501228	5.249523	3.168056	7.0311667	7.7910822
Betas (OLS)	0.269805	-.0628507	0.0227421	0.420649	0.155389	0.521785	
Std. Dev.	0.215133	0.279637	0.145926	0.145286	0.0900693	0.151354	
t values	1.25413	-.224758	.155847	2.87553	1.72522	3.44745	

X_1 = number of family and hired workers

X_2 = number of tractors

X_3 = value of livestock investment

X_4 = acres of irrigated cropland

X_5 = acres of non-irrigated cropland

X_6 = current operating expenses

y = farm products sold

R^2 = 0.925

x_5 is acres of non-irrigated cropland, and

x_6 is current operating expenditure.

The Cobb-Douglas production function was fitted to the model. Means, OLS estimates, standard deviation and t values are presented in Table 1.

Multicollinearity in the Model

The correlation matrix given in Table 2 shows that

$$r_{12} = 0.9299$$

$$r_{13} = 0.7624$$

$$r_{14} = 0.5968$$

$$r_{16} = 0.8188$$

which indicates a degree of collinearity between x_1 and the remaining variable set.

Also from Table 2, x_2 is highly collinear with some variable sets in the model.

$$r_{23} = 0.8048$$

$$r_{24} = 0.7685$$

$$r_{26} = 0.80425$$

The main diagonal elements of the inverted correlation matrix were:

Table 2. Simple correlation coefficient for regression of county values of all farm products sold as a Cobb-Douglas function of inputs.

	X_1	X_2	X_3	X_4	X_5	X_6	y
X_1	1.00000000	0.929918569	0.762442321	0.59685312	-.293808879	0.811853556	0.857723175
X_2		1.00000000	0.804854828	0.768523668	-.343767007	0.804250815	0.895857998
X_3			1.00000000	0.652256285	-.394812482	0.849321658	0.834051825
X_4				1.00000000	-.598540061	0.507591052	-.251882809
X_5					1.00000000	-.170285996	-.251882809
X_6						1.00000000	0.896652600
y							1.00000000

Variable	Main diagonal or VIF
x_1	11.2
x_2	18.9
x_3	05.2
x_4	05.2
x_5	02.0
x_6	05.5

These diagonal elements show that x_2 is the most likely to be affected by multicollinearity and x_1 is in the second place. x_1 and x_2 were highly correlated with the other variables.

Since

$$r_{213456} = 1 - (1 / 18.9) = 0.947$$

and

$$r_{123456} = 1 - (1 / 11.2) = 0.9107,$$

then multicollinearity, brought about by the practical problem of separating the effects of x_1 (labor) from (x_2) tractors on productivity, leads to some serious problems in the model when ordinary least square is fitted. Some of the problems are:

(1) Symptoms of harmful multicollinearity, since $\hat{\beta}_2$ took an illogical negative sign.

(2) Four out of six explantory variables in the model, namely x_1, x_2, x_3 and x_5, were statistically insignificant at the 5% confidence level, due to the high $V(\hat{\beta}_i)$.

(3) Only x_4 and x_6 were statistically significant at the 5% probability level.

Due to these serious problems caused by multicollinearity, many econometricians advocate the idea of deletion of the variables that were highly affected by the problem, and only a subset of these variables was used in the formulation of the final model.

To avoid the possibly serious bias introduced by the specification error which results from the deletion of important variables, the suggested ridge regression (with K_A, K_B and K_M) was fitted to the basic data using the same correlation matrix.

The coefficients, variances and their estimated MSE $(\hat{\beta}_1)$ for OLS, K_A, K_B and K_M, are given in Table 3. Very important results are noticed in the table and these are:

(1) $\hat{\beta}_2$ has changed from an unexpected negative sign $(-.06285)$ to an expected positive sign.

$$\hat{\beta}_2^* = 0.1056 \quad \text{when} \quad K_A \quad \text{is used}$$

$$\hat{\beta}_1^* = 0.1275 \quad \text{when} \quad K_M \quad \text{is used; and,}$$

$$\hat{\beta}_2^* = 0.0699 \quad \text{using} \quad K_B \quad \text{ridge regression.}$$

(2) All the coefficients in the model were possibly statistically significant at 5% (if the problem of bias is temporarily neglected) due to the fact that more precision was obtained and the variances of $V(\hat{\beta}_i)$ were reduced (see Table 4).

Table 3. Estimates of the betas, variances and estimates MSE, OLS vs. K_β, K_M and K_A for regression of county values of all farm products sold as a Cobb-Douglas function of inputs.

	K = 0 OLS	Lawless-Wang K	RIDGM K_M	HBK K_A
$\hat{\beta}_1$	0.2698	0.2015	0.1821	0.1882
$V(\hat{\beta}_1)*$	0.0462824	0.0167051	0.00874386	0.0114868
$V(\hat{\beta}_1) + E(\hat{\beta}_1* - \hat{\beta}_1)^2$	0.0462824	0.0213712	0.0164372	0.0181405
$\hat{\beta}_2$	-.06285	0.0699	0.1275	0.1056
$V(\hat{\beta}_2)*$	0.0891969	0.0226519	0.00939931	0.0137192
$V(\hat{\beta}_2) + E(\hat{\beta}_2* - \hat{\beta}_2)^2$	0.0781969	0.040276	0.0456276	0.0420975
$\hat{\beta}_3$	0.02274	0.0613	0.0935	0.07952
$V(\hat{\beta}_3)*$	0.0212945	0.0141309	0.00939659	0.0113570
$V(\hat{\beta}_3) + E(\hat{\beta}_3* - \hat{\beta}_3)$	0.0212945	0.0156170	0.0144069	0.0145812
$\hat{\beta}_4$	0.42065	0.3435	0.2969	0.3165
$V(\hat{\beta}_4)*$	0.0213995	0.019185	0.007310094	0.00867318
$V(\hat{\beta}_4) + E(\hat{\beta}_4* - \hat{\beta}_4)^2$	0.0213995	0.168770	0.0226222	0.0195232
$\hat{\beta}_5$	0.15539	0.1353	0.1211	0.127550
$V(\hat{\beta}_5)*$	.00811247	0.00618171	0.00504492	0.00553259
$V(\hat{\beta}_5) + E(\hat{\beta}_5* - \hat{\beta}_5)^2$	0.00811247	0.00658359	0.00621976	0.00630762
$\hat{\beta}_6$	0.52179	0.4597	0.4082	0.4309
$V(\hat{\beta}_6)$	0.0229081	0.0146759	0.00938154	0.0115460
$V(\hat{\beta}_6) + E(\hat{\beta}_6* - \hat{\beta}_6)^2$	0.0229081	0.0185291	0.0222841	0.0198085

* Ordinary least square estimates $\hat{\beta}$ were used to estimate $MSE(\hat{\beta}_i*)$.

Table 4. Variances and estimated MSE of coefficients OLS vs. K_A, K_B and K_M (OLS = 100.00) for regression of county values of all farm products sold as a Cobb-Douglas function of inputs for 25 Central Pacific counties, 1954.

	OLS $K = 0$	K_β	K_M	K_A
$V(\hat{\beta}_1)$	100.00	36.09%	18.89%	24.82
$V(\hat{\beta}_1) + E(\hat{\beta}_1{}^* - \hat{\beta}_1)^2$	100.00	46.17	35.5	39.2
$V(\hat{\beta}_2)$	100.00	28.96	12.02	17.54
$V(\hat{\beta}_2) + E(\hat{\beta}_2{}^* - \hat{\beta}_2)^2$	100.00	51.5	58.34	53.84
$V(\hat{\beta}_3)$	100.00	66.35	44.13	53.33
$V(\hat{\beta}_3) + E(\hat{\beta}_3{}^* - \hat{\beta}_3)^2$	100.00	73.34	67.66	68.47
$V(\hat{\beta}_4)$	100.00	51.02	34.16	40.53
$V(\hat{\beta}_4) + E(\hat{\beta}_4{}^* - \hat{\beta}_4)$	100.00	78.87	105.7	91.23
$V(\hat{\beta}_5)$	100.00	76.20	62.19	68.19
$V(\hat{\beta}_5) + E(\hat{\beta}_5{}^* - \hat{\beta}_5)$	100.00	81.15	76.67	77.75
$V(\hat{\beta}_6)$	100.00	64.06)	40.95	50.40
$V(\hat{\beta}_6) + E(\hat{\beta}_6{}^* - \hat{\beta}_6)$	100.00	80.88	97.27	86.47
Over all $V(\hat{\beta}_i) + E(\hat{\beta}_i{}^* - \hat{\beta}_i)^2$	100.00	60.17	64.38	60.77

* $V(\beta^*) + [E(\hat{\beta} - \hat{\beta})]$ was used to estimate MSE $(\hat{\beta}^*)$ because the true β values are unknown.

$V(\hat{\beta}_1)$ was reduced 50-80% from the original under OLS. K_M yielded the highest reduction in variance, $V(\beta_1)$ reduced 70-88%. Again, K_M performs the best if possible bias is ignored. In $V(\hat{\beta}_4)$ there was 50-60% reduction variance, and $V(\hat{\beta}_6)$, $V(\hat{\beta}_5)$ was down by 30-60%.

Among the three ridge regressions used in this analysis, all performed well. The overall performance of the three ridge regression models were obviously superior to that of ordinary least square procedure in terms of the coefficient signs and the precision of the estimates. This type of problem seems to be better tackled by ridge regression rather than OLS.

A Production Function Analysis of Water Resource Productivity in Pacific Northwest Agriculture

Background

The basic data for this analysis was taken from M. H. Holloway (13) (Appendix Table IV). Agriculture has been a major consumer of water in the Pacific Northwest, and the basic objective of this study is to determine the contribution of agricultural water resource development to recent agricultural production.

Cobb-Douglas production function was selected and ordinary least square procedure was primarily chosen to determine the initial estimates.

Table 5. Ordinary least square estimates, standard errors, means and t-values for analysis of water resource productivity in Pacific Northwest (Area C) 1964.

	(X_1)	(X_2)	(X_3)	(X_4)	(X_5)	(X_6)	(X_7)	(X_8)	y
OLS $(\hat{\beta}_i)$	-5.0295	2.7431	1.2982	8.4439	-2.1119	9.1729	16.2737	-1.9993	
S E$(\hat{\beta}_i)$	0.2369	0.1422	0.3119	0.1341	0.0781	0.1174	0.08408	0.1016	
means $\overline{X}_i$	1241.10	5236.29	3889.34	466.678	180.191	225.613	10.7535	84.468	18416.1
T values	-1.21953	5.22212	0.7946	2.2225	-.368790	2.463536	2.29539	-.392828	

Area of Analysis

Only area 'C' which consists of 20 counties producing mostly field crops will be analyzed.

The Model

Cobb-Douglas production function (linear form) was fitted.

$$y = F(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$$

where

y denotes values of farm products sold and value of home consumption.

x_1 is hired and family labor;

x_2 is value of current operating expenses;

x_3 is flow of capital on farm;

x_4 is cropland quantity adjusted for quality;

x_5 is animal unit months;

x_6 is irrigation water application;

x_7 is service flow of form investment in drainage;

x_8 is service flow of farm investment in water conservation practice.

Ordinary least square estimates, standard deviations, t values, and means of the variables are given in Table 5.

Table 6. Water resource productivity in Pacific Northwest agriculture. Simple correlation between variables.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y
X_1	1.0000000								
X_2	.9357189	1.0000000							
X_3	.9550682	.9138426	1.0000000						
X_4	.4227123	.5606500	.5220780	1.0000000					
X_5	.5271189	.4865410	.4759668	.0192848	1.0000000				
X_6	.4992776	.3576612	.4362811	-0.2761547	.2773230	1.0000000			
X_7	.3263360	.4369850	.4825480	.7575409	.1787583	.3354978	1.0000000		
X_8	.3316825	.3137947	.4439894	.1110539	-0.0948184	.1058459	.2659048	1.0000000	
Y	.8908467	.9510433	.9101004	.6511569	.3913535	.4263892	.4818734	.2175557	1.0000000

Multicollinearity

With reference to the sample correlation matrix, (Table 6) and to the diagonal elements of the inverted correlation matrix (VIF), Table 7, the indication is that x_1, x_2, and x_3 are the most affected by multicollinearity since x_3 has the highest value of 59.487, x_1 is 32.584 and x_2 has a value of 11.732. Also, one notices that $r_{12} = 0.9357$ and $r_{13} = 0.955$, which points to the high degree of dependency between these three variables.

Multicollinearity is brought about by the dependency between some of these variables in the variable sets and it is due to the loss of information because of the aggregation in measuring the variables. Multicollinearity here is unavoidable.

Consequences of Multicollinearity in the Model

(1) Ordinary least square procedure yields estimates which are unstable because the variances of the estimates are unreasonably high.

(2) $\hat{\beta}_1$, $\hat{\beta}_5$ and $\hat{\beta}_8$ took illogical, negative signs. These signs were not expected theoretically, and were caused only by the presence of multicollinearity. This is especially true of $\hat{\beta}_1$ and $\hat{\beta}_5$, because x_1 and x_5 are highly collinear with other variables in the variable set.

(3) Most of the variables in the model were statistically

Table 7. Variance inflation factors in the diagonal elements of the inverted correlation matrix (XX) Pacific Northwest water resource productivity data (Area C), 1964.

Explanatory Variables	Main diagonal elements of inverted correlation matrix (VIF)
(X_1)	32. 584
(X_2)	11. 732
(X_3)	56. 487
(X_4)	10. 436
(X_5)	03. 538
(X_6)	08. 006
(X_7)	04. 105
(X_8)	05. 997

Table 8. Estimated coefficients and standard error of OLS, Theil-Goldberger mixed model and ridge regression for the analysis of water resource productivity in Pacific Northwest (Area C) 1964.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	σ^2
OLS $\hat{\beta}_i$	-5.029	2.743	1.298	8.443	-2.112	9.173	16.273	-1.999	4,760.308
S·E $(\hat{\beta}_i)$	(0.2369)	(0.1422)	(0.3119)	(0.1341)	(0.0781)	(0.1174)	(0.08408)	(0.1016)	
PM-1 $\hat{\beta}_i^*$	0.917	1.562	0.785	10.904	1.539	8.674	0.994	0.894	
S·E $(\hat{\beta}_i^*)$	0.917	(0.328)	(0.318)	(1.637)	(0.892)	(1.634)	(0.556)	(0.541)	
ridgm $\hat{\beta}_i^*$	3.113	1.007	0.9923	5.906	0.5353	4.9657	6.348	-1.2459	
S·E (β_i^*)	(0.0168)	(0.0202)	.01416)	(0.0229)	(.0266)	(.02475)	(0.0255)	(0.02613)	
K_B $\hat{\beta}_i^*$	2.9638	1.15944	0.9954	6.6344	-.6738	5.68397	59.5734	-1.86549	
S·E $(\hat{\beta}_i^*)$	(0.02334)	(0.02849)	(.01911)	(.02933)	(.03216)	(.03085)	(0.0328)	(.030821)	
K_A $\hat{\beta}_i^*$	0.904	1.7828	0.8305	8.2599	-2.3398	7.8145	44.9101	-2.12267	
S·E $(\hat{\beta}_i^*)$	(0.0661)	(.07118)	(.05824)	(.06026)	(.04855)	(.05344)	(.05679)	(.04589)	

insignificant and not different from zero at the 5% level (Table 8). This results from the fact that the $V(\hat{\beta}_i)$ are high and resulting t values are small.

Prior Information Model

Due to the problem of multicollinearity, Holloway (13) fitted another model based on incorporating prior information into the regression analysis. The Theil-Goldberger approach (14) was chosen among serveral alternatives because the model allows the specification of the prior parameter estimates with variances (13).

In this analysis, a comparison between ordinary least square, the Theil-Goldberger prior information model (discussed earlier) and ridge regression, will be made.

The Use of Ridge Regression

Although ridge regression can be thought of as another way of incorporating prior information into the regression analysis, it has some advantages over both OLS and the Theil-Goldberger prior information mixed model, as shown in Table 8.

Using the basic data given by Holloway (13), ridge regression in terms of K_B and K_M were fitted, and the results are given in Table 8.

Comparison Between Ordinary Least Square, The Theil-
Goldberger Prior Information Model and Ridge Regression Models

With reference to Table 8, it was noted that

(1) $\hat{\beta}_1$ estimate changed from a wrong negative sign to a posi-
tive sign for the Theil-Goldberger prior information (PM-1)[11/] model.
All ridge procedures (K_A, K_B and K_M) seem to also correct for this
problem, although the magnitude of the coefficient in K_A and Theil-
Goldberger PM-1 were close (0.917 for PM-1 and 0.904 for K_A).

$\hat{\beta}_5$ also changed from a wrong negative sign under OLS to a
positive sign in PM-1 and K_M. For $(\hat{\beta}_8)$, the sign did not change for
any of the ridge regression models K_A and K_B.

Overall results indicate that the use of ridge m worked as well
as the use of the Theil-Goldberger mixed model in correcting the wrong
sign caused by multicollinearity, except for β_8.

$(\hat{\beta}_8)$ changed into positive sign in PM-1, as a result of the prior
information dominancy over the sample information and in this case
the sample contribution to the estimate was minimum.

(2) The variances of the coefficients $V(\hat{\beta}_i)$ were reduced under
ridge regression with ridge performing the best, while PM-1 $V(\hat{\beta}_i)$
was higher than that obtained by OLS. For example:

$SE(\hat{\beta}_1) = 0.2369$ in OLS $= 0.628$ in PM-1 and 0.0168 in ridge m.

[11/] Theil-Goldberger model will be referred to as PM-1
throughout this section.

the same relativity applies to all the coefficients in Table 8.

(3) All the coefficients were significant under ridge m at 5% confidence level, if one ignores the bias of the ridge m estimates.

This problem suffers from a severe case of multicollinearity, worsened by some degree of aggregation in the data. Ridge m seems to perform as well as the prior information model utilized by Holloway (13) to account for multicollinearity, and in fact it has very good merits as an effective tool in this type of problem.

Prior information about the coefficients and their variances is not always available in a practical situation. If they are available almost always there is a bias in utilizing this information because of the different circumstances under which it is obtained. Thus, the assumption of an unbiased prior estimate of one or more β values would often be violated in using the Theil-Goldberger mixed model.

The use of prior information as in this case was associated with the dominance of this prior information over the sample information, which raises questions about the results. Use of the Theil-Goldberger model under these conditions tends to reflect only the prior information effect, and the sample information contribution tends to be obscured by the dominancy of the possibly biased prior information.

Salmon Steelhead Fishing Demand Relationship

Basic data for this analysis was given by Brown, Singh and Castle (3) and by Brown (2).

A demand relationship was proposed in such a way that it differed from past studies. The difference was in its separation of two types of cost involved: a direct cost measured in a monetary form, and another type of cost, time cost, measured by the distance of travel in miles.

The demand functional relationships were

$$DYS = F(INC_j, MLS_j, CST_j)$$

DYS denotes Salmon steelhead (S-S) days of fishing taken per unit of population of subzone j.

INC is average family income of subzone j

MLS is average miles per (S-S) trip of subzone j

and CST is average (S-S) variable cost per day of subzone j.

The demand function fitted by Ordinary Least Square (OLS) yields the following estimate:

$$In \ DYS = 0.7054 \ INC_j + 0.2948 \ MLS_j - 1.1691 \ CST_j$$
$$(.1297) \qquad (.2313) \qquad (.2551)$$

$$\text{number of observations} = 35$$

$$R^2 = 0.6534$$

(Standard deviations of the coefficients are reported in parentheses below the corresponding regression coefficients.)

Table 9. Simple correlation coefficient matrix for salmon steelhead fishing demand relationship.

	INC	MLS	CST	DYS
INC	1.000000			
MLS	0.282220	1.000000		
CST	0.493300	0.872860	1.000000	
DYS	0.211790	-.526645	-.563864	1.000000

DYS = Salmon days of fishing
INC = Average family income
MLS = Average miles per S-S trip
CST = Average variable cost S-S per day

Multicollinearity

Multicollinearity brought about by the difficulty in separating the effect of the monetary cost of travel versus the time cost, is discussed by Brown (3). The use of zone average and the loss of information through the process of aggregation are some of the reasons to suspect multicollinearity.

Test for Multicollinearity

Examining the simple correlation matrix shown in Table 9, it is seen that the highest simple correlation is between the variables, MLS and CST $= r_{23} = 0.87286$.

A good indication of the degree of multicollinearity is given by the main diagonal elements of the inverted correlation matrix $(x'x)$. These were

Explanatory Variable	Variance Inflation Factor (VIF)
INC,	1.5057
MLS	4.7844
CST	5.8197

The highest variance inflation factors (VIF) corresponded to the two variables most affected by multicollinearity, namely MLS and CST.

As a consequence of multicollinearity, the following took place.

(1) Ordinary least square estimates $(\hat{\beta}_2)$ the coefficient of MLS took an insignificant, wrong and theoretically unexpected positive sign.

(2) OLS coefficients, except for INC and CST, were statistically insignificant due to the relatively high variance associated with the coefficients.

Improving the Estimation of the Demand Relationship

To improve the estimation of the demand relationships, the three types of ridge regression suggested to cope with the problem of multi-collinearity were applied to the data. The results are given in Tables 10 and 11. With reference to these tables, the following results are evident.

(1) Hoerl-Baldwin and Kennard Basic K_A Performance

$\hat{\beta}_1$ changed from a large positive number to a relatively smaller positive number.

β_2 is still taking an illogical positive sign although some improvement in the $V(\hat{\beta}_2)$ is noticed.

(2) Lawless and Wang K_B Performances

(i) $\hat{\beta}_2$ changed from illogical positive to the theoretically expected negative sign. $\hat{\beta}_2$ changed from 0.29468 to -.038079 so the sign has been corrected.

Table 10. Salmon steelhead demand relationship coefficients, variances and estimated MSE, OLS[a]
vs. ridge regression.

	K=0 OLS	HGK K_A	LW K_B	RIDGM K_M
β_1^*	0.705293	0.591099	0.533813	0.521988
$V(\beta_1^*)$	0.0168358	0.0128648	0.011627	0.0108254
$V(\beta_1^*)$ % OLS	100.00%	76.4%	69.06%	64.2%
$V(\beta_1^*) + E(\beta_1^* - \beta_1)^2$	0.0168358	0.0259050	0.0405679	0.0444289
% OLS	100.00	153.86	240.96	263.89
β_2^*	0.29468	0.05447	-.038079	-.0054505
$V(\beta_2^*)$	0.0534989	0.0256152	0.0172359	0.0158737
% OLS	100.00	47.879	32.217	29.67
$V(\beta_2^*) + E(\beta_2^* - \beta_2)^2$	0.0534989	0.083317	0.127965	0.137804
% OLS	100.00	153.73	239.13	257.58
β_3^*	-1.1690	-.85999	-.72840	-.703438
$V(\beta_3^*)$	0.065073	0.0296403	0.0191532	0.0174674
% OLS	100.00	45.549	29.433	26.84
$V(\beta_3^*) + E(\beta_3^* - \beta_2)$	0.065073	0.125123	0.213281	0.234215
% OLS	100.00	192.28	327.756	359.9

[a] MSE (β_i) is estimated using OLS β_i as the true (β) since the true β are unknown.

Table 11. Salmon steelhead fishing demand relationship, coefficients, variances and estimated MSE(β_i) OLS vs. ridge regression.[a]

	OLS K = 0	K_A	K_B	K_M	% OLS K_A	% OLS K_B	% OLS K_M
β_1	0.705293	0.591099	0.533813	0.521988			
$V(\beta_1)$	0.0168358	0.0128648	0.011627	0.0108254	.76	0.69	.64
$V(\beta_1) + E(\beta_1{}^* - \beta^*)^2$	0.57002	0.0177410	0.0113200	0.0108259	.34	.22	.21
β_2	0.29468	0.05447	-.038079	-.054505			
$V(\beta_2)$	0.05344989	0.0256152	0.0172359	0.0158737	.47	.32	.29
$V(\beta_2) + E(\beta_2{}^* - \beta^*)^2$	0.534774	0.231303	0.147543	0.0108259	.43	.27	.02
β_3	-.11690	-.859996	-.7284	-.703438			
$V(\beta_3)$	0.065073	0.0296403	0.0191532	0.0174674	.45	.29	.26
$V(\beta_3) + E(\beta_3{}^* - \beta_3{}^*)^2$	0.657880	0.242103	0.127618	0.1101130	.36	.19	.16

[a] *Restriction has been made as to the true coefficient as follows:

$\beta_1{}^* = 0.52127$

$\beta_2{}^* = -.39906$

$\beta_3{}^* = -.39906$

(ii) Again the precision of the estimates is recognized and the estimates of $\hat{\beta}_2$ changed from statistically insignificant to statistically significant at a 5% level.

(3) RIDGM Performances

(i) $\hat{\beta}_2$ also changed from 0.2948 insignificant wrong positive sign to -.054505.

(ii) All estimates were possibly statistically significant at a 5% level.

(iii) Gain in precision was obtained since $V(\hat{\beta}_1)$ decreased by about 70% and $V(\hat{\beta}_3)$ decreased by about 73%.

Some Significant Advantages Over OLS

Two important advantages to ridge regression applied over OLS can be noted from Table 10.

(1) The coefficient $(\hat{\beta}_2)$ of the variable MLS no longer carries a wrong sign under RIDGM and K_B. The symptoms of nonsense regression caused by multicollinearity have been corrected.

(2) Precision has been improved under the suggested ridge, and as a result, coefficients are now different from zero at a 5% level, if possible bias effects are ignored.

To conclude this analysis, RIDGM and Lawless-Wang ridge have very important advantages over OLS in improving the estimate in a model which suffers from multicollinearity.

Estimating the Demand Function for
Commercially Caught Salmon in the Columbia River

Background

Objectives of the study as stated in Brown et al. (3) can be summarized as follows.

(1) A necessary step in estimating benefits to consumers from commercially caught salmon is to estimate their demand for salmon.

(2) The demand function can be used to compute the prices consumers will be willing to pay for specified quantities of the salmon.

The Demand Function

Basic data is given in Brown et al. (3) (Appendix Tables 1-2) and a demand function is specified in the form

$$Pf_t = f(INC_t, \ PR_t, \ QF_t)$$

where Pf_t denotes the wholesale price of fresh and frozen Chinook salmon in New York for the ith year, deflated by the wholesale price index; INC is U.S. per capita disposable personal income deleted by the consumer price index; PR_t is the price of round steak deflated by the consumer price index; and QF_t is U.S. per capita consumption of fresh and frozen salmon.

Multicollinearity

Fortunately, multicollinearity is not a problem in the model since the diagonal elements of the inverted correlation matrix (VIF) showed no severe signs of multicollinearity. VIF were

Variables	VIF
INC	2.07
PR_t	1.04
QF_t	2.07

The simple correlation matrix shown in Table 13 shows low correlation except for $r_{13} = -.716686$. Other correlations were $r_{12} = -.1965$ and $r_{23} = 0.18710$. No harmful multicollinearity is detected in the model.

Ordinary Least Square Estimates

This model is a case where ordinary least square performance is expected to be effective and accurate because there is no multi-collinearity. As shown in Table 12

(1) All coefficients have the correct sign.

(2) All variables were highly significant at the 5% level.

(3) $V(\hat{\beta}_i)$ are not unreasonably large.

Ridge Regressions

Ridge regression models are not expected to do much better than

Table 12. Coefficients, means and standard deviations for commercially caught salmon in Columbia River demand relationship.

	INC.	PR_t	Qf_t	Pf_t
β_i	0.000422	0.21262	-1.45308	
mean $\bar{x}$	2362.21	1.18336	0.199357	0.84878
Standard deviation	46.0965	0.0851157	0.0504147	0.25827

Table 13. Simple correlation coefficients between variables for commercially caught salmon in Columbia River.

	Pf_t	INC	PR_t	Qf_t
Pf_t	1.000000			
INC	0.042967	1.000000		
PR_t	-.1310547	-.196504	1.000000	
Qf_t	0.810522	-.716686	0.187101	1.000000

---

ordinary least square models because of the fact that the data is close

to Orthogonality and only a small improvement can be expected.

Comparison Between OLS and Ridge Regression Selected Models

As shown from Table 14 all ridge models (K_A, K_B, K_M) perform

as well as ordinary least square models with ridge regression carrying

a slight advantage over OLS in terms of

(1) $V(\hat{\beta}_i)$ slightly reduced in ridge models

(2) The measurement of $MSE(\hat{\beta}_i)$ in terms of $MSE(\hat{\beta}_i) =$

$V(\hat{\beta}_i^*) + E(\hat{\beta}_i^* - \hat{\beta}_i)^2$ indicating that ridge models show slight improve-

ment over OLS.

As stated earlier, it is not surprising that OLS performs so

well because of the nature of (x) data that tends toward Orthogonality.

Unfortunately economists do not always have access to such high

quality data.

V. SUMMARY AND CONCLUSIONS

When there is a relatively severe multicollinearity problem in economic data, $x'x$ will be ill-conditioned[12/] and ordinary least square estimates will be both unstable and unreliable. The variances of the OLS coefficients will be unreasonably high, and t and F tests will often result in the rejection of more important variables considered statistically the same as zero at any reasonable significance level.

In an attempt to solve the problem, many methods have been suggested and practiced. Some require that data be collected outside the multicollinearity range, while for others, variables must be either combined or deleted from the economic model.

Frequently, these practices result in inadequate prediction equations giving poor and sometimes erroneous information about the population under investigation. Yet at the same time, there are more promising techniques demanding more information for the improvement of regression estimates in the presence of multicollinearity. The method of Theil and Goldberger (14) is one way of incorporating prior information into the regression problem to lessen the difficulties created by multicollinearity. But this method, also, is not without drawbacks caused by dominance of the prior information over

---

[12/] Generally speaking, $(x'x)$ is ill conditioned when one or more of the eigenvalues of $(x'x)$ are close to zero.

the sample information, and sometimes because of the biased nature of the prior information.

Ridge regression can be considered as a method of incorporating prior information into the regression analysis to solve the multicollinearity problem, where the prior β vector is the null vector. Ridge is among a class of biased linear estimators. It yields regression coefficients that are biased in terms of the true β vector. But these coefficients have a very desirable property, which is the relatively small variance associated with the estimates. The gain in precision obtained by the use of ridge regression overweighs the bias introduced to the system, as measured by $MSE(\hat{\beta})$. The bias is much smaller than the variances of ordinary least square as indicated by the Monte Carlo studies by Hoerl, Kennard and Baldwin (11) and by Lawless and Wang (16). RIDGM, a version of ridge regression, is also a promising method for improving estimates highly affected by multicollinearity.

When RIDGM was compared to ordinary least square, it appeared to have important advantages.

(1) RIDGM estimates were biased but theoretically they were more stable and reliable, while ordinary least square yielded estimates of the wrong sign and offered no significant meaning when interpreted. RIDGM yielded statistically significant estimates of the right sign, if possible bias is ignored.

(2) The variances of the coefficients obtained by RIDGM were relatively smaller compared to those obtained by ordinary least square.

The other two versions of ridge regression, K_A and K_B also yielded estimates with variances much smaller than that obtained by OLS.

In conclusion, significant improvement is not expected from the use of ridge regression over ordinary least squares, when the data are not ill-conditioned or near Orthogonality. In such cases, ordinary least square will yield estimates that are unbiased with variances not much larger than for ridge regression.

Limitations and Additional Research Needed

Although ridge regression performed well for the empirical models considered, additional questions raised by this research are the following:

(1) Which method for estimating the biasing parameter K is best? For the models considered, RIDGM appeared to do well. However, an extensive Monte Carlo study would be needed, where RIDGM and the other methods for selecting K could be compared.

(2) How would ridge regression compare with OLS when there are other defects in the data, in addition to multicollinearity? For example, errors in measurement of the explanatory variables are

known to result in bias when OLS is used. Would such a problem be made worse or better when ridge regression is employed? Again, an extensive Monte Carlo experiment under carefully specified conditions would seem to be a promising method for answering such questions.

In summary, much additional research is still needed to adequately evaluate the role of biased linear estimation in economic research. However, for the four economic models considered in this thesis, the use of ridge regression appears to be very promising given the recent advances in the selection of the biasing parameter, K.

BIBLIOGRAPHY

1. Brown, W. G. "Effect of Omitting Relevant Variables Versus Use of Ridge Regression in Economic Research." Agricultural Experiment Station, Special report #394, Oregon State University, Corvallis, 1973.

2. Brown, W. G. et al. "Improving Economic Evaluation of Commercially and Sport Caught Salmon and Steelhead of the Columbia River." Agricultural Experiment Station, Special report #463, Oregon State University, Corvallis, 1976.

3. Brown, W. G. "Economic Implications of Allocations." Chapter in Marine Recreational Fisheries, Sport Fishing Institute, Washington, D. C., 1976.

4. Brown, W. G. and B. R. Beattie. "Improving Estimates of Economic Parameters by Use of Ridge Regression with Production Function Application." American Journal of Agricultural Economics, vol. 57, No. 1, pp. 29, 30, 1975.

5. Dempster, A. P. et al. "A Simulation Study of Alternatives to Ordinary Least Square." Journal of the American Statistical Association, vol. 72, pp. 78-81, 1977.

6. Draper, N. R., and H. Smith. "Methods for Selecting Variables from a Given Set of Variables for Regression Analysis." Bulletin of the International Statistical Institute 43:7-15, 1975.

7. Farrar, D. E., and R. R. Glauber. "Multicollinearity in Regression Analysis : The Problem Revisited". Review of Economic and Statistics, vol. 49, pp. 92-107, 1967.

8. Goldberger, A. S. Econometric Theory, John Wiley & Sons, Inc., New York, 1964.

9. Hocking, R. R. et al. "A Class of Biased Estimators in Linear Regression." Technometrics, vol. 18, No. 4, pp. 426-429, 1976.

10. Hoerl, A. E., and R. W. Kennard. "Ridge Regression : Biased Estimation for Non-Orthogonal Problems." Technometrics, vol. 12, No. 1, pp. 55-67, 1970.

11. Hoerl, A.E. et al. "Ridge Regression : Some Simulations." Communications in Statistics, 4(2), pp. 106-108, 1975.

12. Hoerl, A.E., and R.W. Kennard. "Ridge Regression: Applications to Non-Orthogonal Problems." Technometrics, vol. 12, No. 1, pp. 69-82, 1970.

13. Holloway, M. "A Production Function Analysis of Water Resource Productivity in Pacific Northwest Agriculture." A thesis submitted to Oregon State University, June 1972, pp. 92-179.

14. Johnston, J. Econometric Methods. 2nd Ed. McGraw-Hill Book Co., Inc., New York, N.Y. 1972.

15. Judge, G.G. and T. Takayama. "Inequality Constraints in Regression Analysis." Journal of the American Statistical Association, vol. 61, 1966, pp. 166-181.

16. Lawless, J.F. and P. Wang. "A Simulation Study of Ridge and Other Regression Estimators." Communications in Statistics, A5(4), 307-323, 1976.

17. Mason, R.L., et al. "Regression Analysis and Problems of Multicollinearity." Communications in Statistics, 4(3), pp. 280-285, 1975.

18. Roa, P. and R.L. Miller. Applied Econometrics, Wadsworth Publishing Co., Inc., Belmont, California, 1971.

19. Theil, H. "Specification Errors and the Estimation of Economic Relationships." Rev. International Stat. Inst. 25:1/3 (1957) pp. 41-51.

20. Theil, H. "On the Use of Incomplete Prior Information in Regression Analysis." Journal of the American Statistical Association, vol. 58, pp. 401-414, 1963.