45003

NAME OF AUTHOR/*NOM DE L'AUTEUR* RICHARD J. ZARZECZNY

TITLE OF THESIS/*TITRE DE LA THÈSE* AN EVALUATION OF VARIOUS ALTERNATIVES TO ORDINARY LEAST SQUARES ESTIMATION IN THE PRESENCE OF MULTICOLLINEARITY

UNIVERSITY/*UNIVERSITÉ* SIMON FRASER UNIVERSITY

DEGREE FOR WHICH THESIS WAS PRESENTED/ *GRADE POUR LEQUEL CETTE THÈSE FUT PRÉSENTÉE* MASTER OF ARTS

YEAR THIS DEGREE CONFERRED/*ANNÉE D'OBTENTION DE CE GRADE* 1979

NAME OF SUPERVISOR/*NOM DU DIRECTEUR DE THÈSE* DR. PETER KENNEDY

DATED/*DATÉ* April 18, 1979 SIGNED/*SIGNÉ*

PERMANENT ADDRESS/*RÉSIDENCE FIXÉ* [redacted]

NL-91 (3-74)

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us a poor photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30. Please read the authorization forms which accompany this thesis.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de mauvaise qualité.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30. Veuillez prendre connaissance des formules d'autorisation qui accompagnent cette thèse.

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

AN EVALUATION OF VARIOUS ALTERNATIVES TO ORDINARY LEAST SQUARES ESTIMATION IN THE PRESENCE OF MULTICOLLINEARITY

by

Richard Zarzeczny

B.A. (HONOURS), University of Regina, 1974

M.A., University of Regina, 1975

A THESIS SUBMITTED IN PARTIAL

FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Economics and Commerce

SIMON FRASER UNIVERSITY

December, 1976

Title of Thesis/Dissertation:

AN EVALUATION OF VARIOUS ALTERNATIVES TO ORDINARY LEAST SQUARES ESTIMATION IN THE PRESENCE OF MULTICOLLINEARITY

Author:

(signature)

RICHARD J. ZARZECZNY

(name)

April 18, 1979

(date)

APPROVAL

Name: Richard J. Zarzéczny

Degree: Master of Arts

Title of Thesis: An Evaluation of Various Alternatives to Ordinary Least Squares Estimation in the Presence of Multicollinearity

Examining Committee:

Chairman: Zane A. Spindler

Peter Kennedy
Senior Supervisor

Dennis Maki

Professor Ken White
Dept. of Economics
University of British Columbia

Date Approved: December 10, 1976

Abstract

Dissatisfaction with ordinary least squares estimation (OLS) in the presence of multicollinearity has resulted in the development of various alternative estimators which claim to be better than OLS. These claims have not been resolved. This thesis evaluates a number of alternative estimators to determine their performance relative to OLS and each other. The performance of each estimator is evaluated under a variety of data conditions. In addition, the subjective nature of the ridge trace estimation method is determined and evaluated.

The method used to evaluate the estimators is the Monte Carlo simulation method. The performance criterion is total mean square error.

The simulation study revealed that the ridge trace estimator, even when adjusted for its subjectivity, ranked first in overall performance relative to OLS and relative to the other estimators considered. The overall performance of the alternative estimators (except one) was better than OLS. The absolute performance of the alternative estimators, and thus their performance relative to OLS and each other, varied with the degree of multicollinearity, the variance of the error term, and the direction of the true coefficient vector.

It is concluded that one of the alternative estimators should be used in place of OLS, with the specific choise dependent on the condition of the data. In the absence of reliable knowledge of certain data conditions, the ridge trace estimator should be used in place of OLS in the presence of multicollinearity.

Acknowledgements

I wish to express my sincere gratitude to my senior supervisor, Dr. Peter Kennedy, who provided me with valuable criticisms, guidance and comments through the several drafts of the thesis. I am also indebted to Ernie Dainow of the Simon Fraser Computing Centre who aided me extensively in programming the Monte Carlo simulation. In this regard, my gratitude extends to my fellow graduate students, Kent Sproule and Brian Pratt, who removed many bugs from my program. I would like to thank Dr. Ken White of Rice University for serving as external examiner and for reading and commenting on the thesis. I appreciated the help of those in the Simon Fraser Economics department who eyeballed ridge trace graphs for me. Any errors contained in the thesis are my own.

Table of Contents

		Page
Approval		ii
Abstract		iii
Acknowledgements		v
Table of Contents		vi
List of Tables		vii
List of Fiqures		ix
Chapter 1	Introduction	1
Chapter 2	The Classical Linear Regression Model and OLS in the Presence of Multicollinearity	4
Chapter 3	Alternatives to OLS	11
Chapter 4	Results of Previous Monte Carlo Studies	25
Chapter 5	Design of the Experiment	34
Chapter 6	Results and Conclusions	44
Chapter 7	Suggestion for Future Research	68
Appendix 1		71
Appendix 2		75
Appendix 3		80
Bibliography		95

List of Tables

Table		Page
1.	Comparison of TMSE of ORR. Ridge Trace vs. Mathematical Methods of Selecting k	53
2.	M for All Estimators	56
3.	Aggregated TMSE	62
4.	Ratio of Aggregated TMSE of Estimators for β_S	63
5.	Ratio of Aggregated TMSE of Estimators for β_L	64
6.	Ratio of Aggregated TMSE of Estimators to Aggregated TMSE of OLS	65
7.	Overall Rank of Estimators Based on Average Ratio of the Particular Estimator to All Estimators	66
8.	Values of $V_L'\hat{\beta}_L$ and $V_L'\hat{\beta}_S$	70
A1.	Correlation Matrices by Degree of Multicollinearity	80
A2.	Eigenvalues	82
A3.	True Coefficient Vectors	82
A4.	Average k Value and Average δ_i Value	83
A5.	Average Q Value and Number of Negative Q Values Out of 50 Samples	84
A6.	Average Squared Length	85
A7.	Total Bias	87

List of Tables - continued

Table		Page
A8.	Total Variance	89
A9.	TMSE	91
A10.	$\sigma_\varepsilon^2/\alpha^2 max$	93
A11.	Number of k Values in $\left(0, \frac{\sigma_\varepsilon^2}{\alpha^2 max}\right)$	94

List of Figures

Figure		Page
1.	Ridge Trace for $\mu = .9$, $\sigma_\varepsilon = .5$, $\beta=\beta_S$ for ORR	50
2.	Ridge Trace for $\mu = .9$, $\sigma_\varepsilon = .5$, $\beta=\beta_S$ for SSE1	51

CHAPTER 1

INTRODUCTION

Dissatisfaction with ordinary least squares (OLS) estimation in the presence of multicollinearity has led to the development of several alternative estimators. The most popular of these is the ridge estimator of Hoerl and Kennard. (Hoerl and Kennard 1970a and 1970b) It has received considerable attention in the literature and has been the subject of several empirical evaluations (Newhouse and Oman 1971, McDonald and Galarneau 1975, Guilkey and Murphy 1975, Deegan 1975, Hoerl, Kennard and Baldwin 1975, Lawless and Wang 1976 and Hoerl and Kennard 1976).

Although a theoretically promising alternative, there is a practical problem with ridge regression. For any given problem the performance (in terms of total mean square error) of the ridge estimator depends on an elusive parameter, which must be estimated from the sample data. If a "correct" value of the parameter is chosen a reduction in the total mean square error relative to OLS is achieved. If an incorrect value is chosen the ridge estimator will perform worse than OLS. Most current research has attempted to derive an explicit mathematical formula for determining a value of the parameter conserving a good performance from the ridge estimator.

In addition to the mathematical means of locating this parameter value, there exists an intuitive method for choosing it. The technique, known as the ridge trace method, was originally proposed by Hoerl. (Hoerl 1962, Hoerl and Kennard 1970a and 1970b). Although satisfactory results have been obtained with this method in several specific applications (Hoerl and Kennard (1970b), Watson and White (1975), Mason and Brown (1975) and McDonald and Schwing (1973) it has never been objectively evaluated.

The primary purpose of this thesis is to attempt to provide, by means of a Monte Carlo simulation study, an evaluation of the ridge trace method of choosing a value of the parameter in question. This method is compared to OLS and to several methods using mathematical rules for determining the parameter value.

Although the ridge estimator has been receiving the bulk of attention, the literature contains a number of estimators which also claim to be better than OLS (and in some cases better than ridge estimation itself) under conditions of multicollinearity. Some are similar to the ridge estimator; in fact some have arisen out of investigations of the ridge estimator. However, none have as yet been directly evaluated. The thesis proposes to do this. By comparing these estimators with OLS and with the various versions of the ridge estimator, it is hoped that the characteristics of these alternative estimators can be specified sufficiently accurately to allow a researcher to choose that estimator which best suits his purposes and data conditions.

The thesis is structured as follows. Chapter 2 presents the classical linear regression model and the problems associated with OLS when the assumption of uncorrelated independent variables is dropped. All the estimators under investigation are summarily introduced. Chapter 3 provides a detailed description of these estimators. Chapter 4 summarizes the results of existing empirical studies on the topic of ridge regression. In Chapter 5, the design of the Monte Carlo experiment and the criterion used to evaluate the estimators is presented. The method used to evaluate the ridge trace parameter selection in ordinary ridge regression is also discussed. Chapter 6 contains the results and conclusions drawn from the experiment and finally, Chapter 7 provides a suggestion for future research.

CHAPTER 2

THE CLASSICAL LINEAR REGRESSION MODEL AND OLS IN THE PRESENCE OF MULTICOLLINEARITY

Consider the classical linear regression model (CLRM)

$$y = X\beta + \varepsilon \qquad (1)$$

where β is the $N\times 1$ vector of population values of the parameters; ε is a $T\times 1$ vector of error terms having the properties; $E(\varepsilon)=0$ and $E(\varepsilon\varepsilon')=\sigma_\varepsilon^2 I$ (I being the $T\times T$ identity matrix). The vector y is $T\times 1$ and contains the observed values of the dependent variable at the T data points. X is a $T\times N$ matrix having rank N and contains the values of N explanatory variables at each of T data points. X is fixed in repeated samples which implies that the regressors are nonstochastic and that X and ε are independent. Unless otherwise mentioned, the independent variables in the CLRM are assumed to be standardized so that X'X is in the form of a correlation matrix.

The conventional estimator of β is the ordinary least squares (OLS) estimator defined by

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

with variance - covariance matrix

$$\Sigma_{\hat{\beta}\hat{\beta}} = \sigma_\varepsilon^2 (X'X)^{-1}$$

Its essential features are that it is unbiased and has minimum variance among all linear unbiased estimators of β.

In practical estimation problems a researcher is usually faced with a small sample in which there is some degree of intercorrelation among the explanatory variables. The term multicollinearity (ill-conditioned or non-orthogonal) is used to denote the presence of approximate linear relationships among the explanatory variables as reflected by a high degree of intercorrelation.

Although the OLS estimator remains an unbiased and minimum variance estimator in the presence of multicollinearity, these properties are based on averages taken over a large number of samples (or in a large sample) but in any particular sample which is small, the OLS estimator may exhibit large errors. How the presence of multicollinearity in a particular sample may impair the accuracy and stability of the OLS estimator of β will be revealed in the following discussion.

The presence of multicollinearity in a particular sample implies that the vectors of X deviate from orthogonality resulting in an X'X matrix which is "nearly" singular in the sense of possessing one or more small eigenvalues. If $\lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda_N \geq 0$ are the ordered eigenvalues of X'X then as X deviates further from orthogonality, λ_N becomes smaller. If we let $TV(\hat{\beta})$ denote the total variance of $\hat{\beta}$, then

$$TV(\hat{\beta}) = tr\textstyle\sum_{\hat{\beta}\hat{\beta}}$$
$$= \sigma_\varepsilon^2 tr(X'X)^{-1}$$
$$= \sigma_\varepsilon^2 \sum_{n=1}^{N} 1/\lambda_n \qquad (2)$$

(where tr denotes trace). If one or more λ_n are small (as they are certain to be in the presence of multicollinearity) equation (2) shows that the total variance of $\hat{\beta}$ will be large (ie. the distance between $\hat{\beta}$ and β will be large). Thus in any particular sample, β can be expected to be far from the true value β (in the sense of Euclidean distance).

For any individual coefficient estimate, $\hat{\beta}_i$, of the true parameter β_i, the presence of multicollinearity will produce OLS estimates which on average are larger in absolute value than the true values. To see this, Hoerl and Kennard (1970a) have shown that

$$E(\hat{\beta}'\hat{\beta}) = E\left(\sqrt{\sum_{i=1}^{N} \hat{\beta}_i^2}\right) = \sqrt{\sum_{i=1}^{N} \beta_i^2 + TV(\hat{\beta})}, \qquad (3)$$

implying that one or more terms $(\hat{\beta}_i)$ appearing in the left hand summation are expected to be larger than the true values β_i in the right hand summation. It is in this sense that the individual OLS estimates are deemed to be inaccurate.

of multicollinearity has led to a search for alternative estimators. As seen above in (2) and (3), the OLS estimates are imprecise due to the inflation of $TV(\hat{\beta})$. Consequently, the search for alternatives to OLS have focused on deflating the total variance. Considering the total mean square error (TMSE) of any estimator B, defined as

$$\begin{aligned} TMSE(B) &= E[(B-\beta)'(B-\beta)]^{1} \\ &= (\text{Total bias } B)^2 + TV(B), \end{aligned}$$

it would be desirable to obtain an estimator which by allowing a little bias, substantially reduces the total variance and hence circumvents the undesirable effects of multicollinearity. The criterion for a good estimator would then be small TMSE.

The above considerations have led to the development of several biased estimators each of which claim to have smaller TMSE than the OLS estimator. For reference, we now list these estimators; they are described in greater detail in Chapter 3.

---

1. If B_1 and B_2 are two estimators of the true parameter vector β, the conventional definition of MSE is $M_i = E(B_i-\beta)(B_i-\beta)'$, $i=1,2$. Then B_2 has a "smaller" MSE than B_1 if M_1-M_2 is non-negative definite. Now since $TMSE(B_i)=trM_i$, an equivalent criterion is obtained by comparing the sizes of $TMSE(B_1)$ and $TMSE(B_2)$. (See Appendix 1).

1. The ordinary ridge estimator (ORE) defined as

$$\hat{\beta}(k) = (X'X+kI)^{-1}X'y,$$

where $k \geq 0$ is fixed. The following methods of choosing k are considered:

(a) ORE 1; The ridge trace method;

(b) ORE 2; Choose $k = \frac{\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}}$, where

$\hat{\sigma}_\varepsilon^2 = \frac{y'y-\hat{\beta}'X'y}{T-N}$ and $\hat{\beta}$ is the OLS estimate of β;

(c) ORE 3; Choose $k = \frac{N\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}}$, where

N is the number of independent variables,

(d) Let $Q = \hat{\beta}'\hat{\beta}-\hat{\sigma}_\varepsilon^2 \sum_{n=1}^{N} 1/\lambda_n$

then the following rules determine k:

ORE 4a; If $Q>0$, choose k such that $(\hat{\beta}(k))'\hat{\beta}(k)=Q$.

If $Q \leq 0$, choose $k=0$;

ORE 4b; If $Q>0$, choose k such that $(\hat{\beta}(k))'\hat{\beta}(k)=Q$.

If $Q=0$, choose $k=\infty$.

Note that ORE 4a and ORE 4b are identical if $Q>0$, however if $Q \leq 0$, ORE 4a specifies that $\hat{\beta}$ be used to estimate β whereas ORE 4b specifies that the zero vector estimates β.

2. The generalized ridge estimator (GRE) is defined as

$$\hat{\alpha}(K) = [(X^*)'X^*+K]^{-1}(X^*)'y$$

where $X^*=XP'$ (P being the N×N orthogonal transformation such that $P'P=I$) and K is an N×N diagonal matrix with diagonal elements $k_1,\ldots,k_N\geq 0$. The following method leads to a criterion which is equivalent to choosing values for the $k_i (1\leq i\leq N)$:

GRE 1; i. For the model $y=X^*\alpha+\varepsilon$ ($\alpha=P\beta$) obtain the OLS estimates $\hat{\alpha}_i$ (i=1,...,N) and the estimate $\hat{\sigma}_\varepsilon^2$ of σ_ε^2.

ii. Calculate

$$e_{i(0)} = \frac{\hat{\sigma}_\varepsilon^2}{\lambda_i \hat{\alpha}_i^2} \text{ for } i=1,\ldots,N.$$

iii. If $e_{i(0)}>1/4$, let $\hat{\alpha}_i(k_i)=0$.

If $0<e_{i(0)}\leq 1/4$, let $\hat{\alpha}_i(k_i)=\frac{\hat{\alpha}_i}{1+e_i}$

where $\hat{\alpha}_i(k_i)$ is the ith GRE and

$$e_i = \frac{1-2e_{i(0)}-\sqrt{1-4e_{i(0)}}}{2e_{i(0)}}.$$

3. The Minimum Mean Square Error Linear Estimator (MMSELE) defined as

$$\text{MMSELE} = \left(\frac{\hat{\beta}'X'y}{\hat{\sigma}_\varepsilon^{\,2} + \hat{\beta}'X'X\hat{\beta}} \right) \hat{\beta}.$$

4. Stochastically Shrunken Estimators defined as

$$\text{SSE1} = \delta\hat{\beta}\hat{\beta}'(I+\delta\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}$$

where $\delta\varepsilon[0,\infty)$. A value of δ is chosen by means of the ridge trace method.

$$\text{SSE2} = [1+\gamma S^2(\hat{\beta}'\hat{\beta})^{-1}]\hat{\beta} \quad \text{where}$$

$$S^2 = y'y-\hat{\beta}'X'X\hat{\beta} \text{ and } \gamma = \frac{N-2}{T-N+2} .$$

5. Generalized Inverse Estimator defined as

$$\text{GIE} = A_r^+ X'y$$

where $A_r^+ = (X'X_r)^{-1}$, r being a suitably chosen rank of $X'X$.

The performance of the above estimators (in terms of TMSE) relative to OLS will be evaluated under a variety of data conditions. First, however, the next chapter discusses, in some detail, the properties of these estimators and the claims made regarding their performance.

CHAPTER 3

ALTERNATIVES TO OLS

Ordinary Ridge Estimator (ORE)

The ordinary ridge estimator is defined by

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

where $k\geq 0$ is fixed. Note that $\hat{\beta}=\hat{\beta}(0)$ and $\hat{\beta}(k)$ is a biased estimator if $k>0$. A comprehensive theory concerning the properties of $\hat{\beta}(k)$ has been developed by Hoerl and Kennard (1970a). In particular, the ORE minimizes the sum of squared residuals subject to a constraint on the squared Euclidean length $(\hat{\beta}(k))'\hat{\beta}(k)$ of the estimator. Thus $\hat{\beta}(k)$, for $k>0$ is "shorter" than $\hat{\beta}$, i.e., $(\hat{\beta}(k))'\hat{\beta}(k)<\hat{\beta}'\hat{\beta}$. In addition, for any given problem there exists a range of values of k (admissable values) for which $TMSE[\hat{\beta}(k)]<TMSE(\hat{\beta})$. However these admissable values depend on the unknown coefficient vector β, on σ_ε^2 and the X matrix through the eigenvalues of X'X. Thus no constant value of k can be assured to yield an ORE which is better than OLS (in terms of TMSE) for all β, σ_ε^2 and X. Consequently, various "rules" have been developed for choosing an admissable k value. There are essentially two types of rules - subjective and mathematical.

The subjective method of choosing an admissable k value (ORE 1) consists of examining the ridge trace which is a two-dimensional plot of $\hat{\beta}_i(k)$

and the residual sum of squares, $\phi(k)$, for values of k in the interval [0,1]. Hoerl and Kennard deem this to be the best method for selecting an admissable value of k and thus a unique $\hat{\beta}(k)$. They give the following criteria to quide one to a good choice of k when examining the ridge trace:

- At a certain value of k the system will stabilize and have the general characteristics of an orthogonal system;
- Coefficients will not have unreasonable absolute values;
- Coefficients with apparently incorrect signs at k=0 will have changed to have the proper sign;
- The residual sum of squares will not have been inflated to an unreasonable value, i.e., not be large relative to the minimum residual sum of squares (at k=0).

The problem with the ridge trace method of selecting k is that different investigators are likely to choose somewhat different values of k. (This problem is examined in more detail in Chapter 6). This problem with the ridge trace method has led to the development of various mathematical rules for determining an admissable value of k.

In their study of ridge estimation Newhouse and Oman (1971) proposed the following choice of k (defining ORE 2):

$$k = \frac{\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}} .$$

The rationale behind choosing this value of k is that

$$TMSE(\hat{\beta}(k)) < TMSE(\hat{\beta}) \text{ when}$$

$k < \dfrac{\sigma_\varepsilon^2}{\beta'\beta}$. Since, in the presence of multicollinearity, $\hat{\beta}'\hat{\beta} > \beta'\beta$, it will be likely that

$$\frac{\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}} < \frac{\sigma_\varepsilon^2}{\beta'\beta} .$$

A slight variation of ORE 2 gives us another choice of k, namely

$$k = \frac{N\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}} .$$

This choice of k (defining ORE 3) comes from an article by Farebrother (1975). He observed that if X'X=I, then $TMSE(\hat{\beta}(k))$ is minimized at this value of k. In various examples cited by Farebrother, $\dfrac{N\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}}$ generated approximately the minimal value of $TMSE(\hat{\beta}(k))$ even though X'X≠I. Also Hoerl, Kennard, and Baldwin (1975) have shown that if $\dfrac{N\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}}$ is used as an estimate of k, then significant improvement (in terms of TMSE) over OLS is obtained.

McDonald and Galarneau (1975) base their choice of k on an unbiased estimator of $\beta'\beta$, the squared length of the true coefficient vector. This unbiased estimator is given by

$$Q = \hat{\beta}'\hat{\beta} - \hat{\sigma}_\varepsilon^2 \sum_{n=1}^{N} 1/\lambda_n$$

where $\hat{\sigma}_\varepsilon^2 = \frac{\emptyset(0)}{T-N}$ is an unbiased estimator of σ_ε^2, $\emptyset(0)$ being the residual sum of squares for the OLS estimate. The mathematical rules for choosing k are then given as follows:

ORE 4a. If $Q>0$, choose k such that $(\hat{\beta}(k))'\hat{\beta}(k)=Q$

If $Q\leq 0$, choose $k=0$.

ORE 4b. If $Q>0$, choose k such that $(\hat{\beta}(k))'\hat{\beta}(k)=Q$

If $Q=0$, choose $k=\infty$.

ORE 4a and 4b are identical if $Q>0$, however if $Q\leq 0$, ORE4a specifies that $\hat{\beta}$ be used to estimate β whereas ORE4b specifies that the zero vector estimates β. This follows from the fact that as k approaches infinity, $(\hat{\beta}(k))'\hat{\beta}(k)$ approaches zero. The reason for basing the choice of k on Q is that in the presence of multicollinearity (3) shows that the squared Euclidean length of $\hat{\beta}$ is expected to be too large. So, choosing k such that

$$(\hat{\beta}(k))'\hat{\beta}(k) = \hat{\beta}'\hat{\beta} - \hat{\sigma}_\varepsilon^2 \sum_{n=1}^{N} 1/\lambda_n \quad (4)$$

is insuring that the squared length of $\hat{\beta}(k)$ is expected to equal the squared length of the true coefficient vector. That is from (4)

$$E[(\hat{\beta}(k))'\hat{\beta}(k)] = \beta'\beta \ .$$

Thus unlike the OLS estimate $\hat{\beta}$, the elements of $\hat{\beta}(k)$ are not expected to be large in the presence of multicollinearity.

Generalized Ridge Estimator. (GRE)

In their 1970a paper, Hoerl and Kennard proposed a general form of the ridge estimator. It is obtained as follows:

From the classical linear regression model (1) consider $X'X$ which is an $N\times N$ symmetric matrix. Since $X'X$ is symmetric there exists an orthogonal transformation $P(P'P=I)$ where P is $N\times N$, such that

$$P(X'X)P' = \Omega$$

where Ω is a diagonal matrix with diagonal elements $\lambda_1,\ldots,\lambda_N$, the eigenvalues of the $X'X$ matrix. We can then write (1) as

$$Y = X^*\alpha+\varepsilon \qquad (5)$$

where $X^*=XP'$ and $\alpha=P\beta$.

The generalized ridge estimate (GRE) is defined as

$$\hat{\alpha}(K) = [(X^*)'X^*+K]^{-1}(X^*)'y$$

$$= (\Omega+K)^{-1}(X^*)'y$$

where K is an N×N diagonal matrix with diagonal elements $k_1,\ldots,k_N \geq 0$. Note that $\hat{\alpha}(K)$ is an estimate of $\alpha = P\beta$. To transform $\hat{\alpha}(K)$ into estimates of β we use the following relationship:

$$\hat{\beta}(K) = P'\hat{\alpha}(K).$$

Instead of adding the same quantity k to each eigenvalue of X'X as in ordinary ridge regression (ORR), the GRE is obtained by augmenting the eigenvalues of X'X by differing positive quantities k_i $(i=1,\ldots,N)$.

Clearly if $k_i=k$ for all i, then $\hat{\alpha}(K)=\hat{\beta}(k)$. In this case, if $k>0$, Hoerl and Kennard (1970) have shown that TMSE $[\hat{\beta}(k)] < \text{TMSE}(\hat{\beta})$ provided that

$$k < \frac{\sigma_\varepsilon^2}{\alpha_{max}^2}$$

where $\alpha_{max} = \max\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$.

Similarly, when the k_i are not equal, TMSE$[\hat{\alpha}(K)] <$ TMSE$(\hat{\alpha})$ ($\hat{\alpha} = \Omega^{-1}(X^*)'y$ is the OLS estimate of α in model (5)) if

$$k_i = \frac{\sigma_\varepsilon^2}{\alpha_i^2},\ i=1,\ldots,N.$$

As with ORR, the values of k_i are functions of the unknown coefficient vector β, σ_ε^2 and the eigenvalues of X'X. Furthermore, for GRR, there is no equivalent to the ridge trace method of ORR for determining the optimal value of k_i.

Hoerl and Kennard (1970a) suggest an iterative procedure to determine the optimal k_i values. By examining the convergence properties of this iterative procedure, Hemmerle (1975) was able to derive an explicit solution for GRR. This constitutes our selection rule for GRR.

GRE1: The following steps lead to a criterion which is equivalent to choosing the optimal k values for GRR.

1. For the model

$$y = X^{*}\alpha + \varepsilon$$

obtain the OLS estimates $\hat{\alpha}_i$ (i=1,...,N) and the estimate $\hat{\sigma}_\varepsilon^2$ of σ_ε^2 where

$$\hat{\sigma}_\varepsilon^2 = \frac{y'y - \hat{\alpha}'(X^{*})'y}{T-N}.$$

2. Calculate

$$e_{i(0)} = \frac{\hat{\sigma}_\varepsilon^2}{\lambda_i \hat{\alpha}_i^2} \quad \text{for } i=1,\ldots,N.$$

3. If $e_{i(0)} > 1/4$, let $\hat{\alpha}_i(k_i) = 0$.

If $0 < e_{i(0)} \leq 1/4$, let $\hat{\alpha}_i(k_i) = \frac{\hat{\alpha}_i}{1+e_i}$

where $\hat{\alpha}_i(k_i)$ is the ith GRR estimate and

$$e_i = \frac{1-2e_{i(0)} - \sqrt{1-4e_{i(0)}}}{2e_{i(0)}}.$$

The study of ORR and GRR has led to the development of other biased estimators which attempt to alleviate the ill effects of multicollinearity.

Minimum Mean Square Error Linear Estimator (MMSELE)

The details of the MMSELE are given by Farebrother (1975). It is defined as

$$\beta^* = \left(\frac{\beta'X'y}{\sigma_\varepsilon^2+\beta'X'X\beta}\right)\beta .$$

The TMSE (β^*) is less than the TMSE of any other estimator which is linear in y. Since it depends on β and σ_ε^2 it is not operational. Farebrother suggests using the operational variant of β^* given by

$$b^* = \left(\frac{b'X'y}{s_\varepsilon^2+b'X'Xb}\right)b$$

where b and s_ε^2 are any consistent estimators of β and σ_ε^2 respectively. In particular he suggests via an example the use of $b=\hat{\beta}$ and $s_\varepsilon^2=\hat{\sigma}_\varepsilon^2$.

Farebrother states, "Our study of Hoerl and Kennard's worked examples suggests that the subjective method of the ridge trace should be replaced by

$$\text{MMSELE} = b^* = \left(\frac{\hat{\beta}'X'y}{\hat{\sigma}_\varepsilon^2+\hat{\beta}'X'X\hat{\beta}}\right)\hat{\beta} .$$

Shrunken Estimators

Mayer and Wilkie (1973) have proposed the following estimators as possible alternatives to OLS estimation in the presence of multicollinearity. For a detailed discussion of these estimators, the reader is referred to the above article.

In general, Mayer and Wilke define a stochastically shrunken estimator (SSE) as

$$SSE = f\hat{\beta}$$

where $f=f(\hat{\beta}'\hat{\beta})$ is a scalar function of $\hat{\beta}'\hat{\beta}$.

In particular for

$$f = \delta[\hat{\beta}'\hat{\beta}+(1+\hat{\beta}'\hat{\beta})^{-1}\delta(\hat{\beta}'\hat{\beta})^{2}]$$

and

$$f = [1+\gamma S^{2}\hat{\beta}'\hat{\beta}]^{-1}$$

where $S^2=y'y-\hat{\beta}'X'X\hat{\beta}$ and δ, $\gamma\varepsilon[0,\infty)$ are fixed, we obtain two shrunken estimators which (depending on the values of δ and γ chosen) will have TMSE less than that of the corresponding OLS estimate. The term "shrunken" is used to describe the fact that the squared length of these estimators is smaller than the squared length of $\hat{\beta}$. As was shown earlier, this is a desirable asset for an estimator in the presence of multicollinearity as the squared length of $\hat{\beta}$ tends to exceed that of the true coefficient vector.

The following shrunken estimators are included as alternatives to OLS in the presence of multicollinearity:

$$SSE1 = \delta[\hat{\beta}'\hat{\beta}+(1+\hat{\beta}'\hat{\beta})^{-1}\delta(\hat{\beta}'\hat{\beta})^{2}]\hat{\beta}$$

$$= \delta\hat{\beta}\hat{\beta}'(I+\delta\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}, \quad \delta\varepsilon[0,\infty)$$

and

$$SSE2 = [1+\gamma S^{2}(\hat{\beta}'\hat{\beta})^{-1}]\hat{\beta}, \quad \gamma\varepsilon[0,\infty).$$

As with ORR, the problem arises of how to choose the proper shrinkage factor i.e., a δ and γ which will ensure that SSE1 and SSE2 have a smaller TMSE than that of $\hat{\beta}$. Mayer and Wilke (1973) suggest that δ can be chosen by plotting $(SSE1)_i$ against a range of values of δ in [0,1] and to then use the same stability criteria as used by Hoerl and Kennard for choosing k in ORR. This method will be used in this paper.

For SSE2 the value

$$\gamma = \frac{N-2}{T-N+2}$$

is suggested by Mayer and Wilke. This value of γ will guarantee that SSE2 has a smaller weighted total mean square error (WTMSE) than $\hat{\beta}$.[1] This value of γ is used in this study.

---

1. The WTMSE of an estimator B is defined by Mayer and Wilke (1973) as

$$WTMSE(B) = E(B-\beta)'X'X(B-\beta).$$

Generalized Inverse Estimator (GIE)

One of the assumptions of the CLRM (1) is that the X matrix is of full rank N. This ensures that X'X is non-singular and hence that $(X'X)^{-1}$ exists. If the rank of X is less than N, the OLS estimating procedure breaks down mathematically. However, in the presence of multicollinearity, X'X may be "nearly" singular in the sense of possessing near zero eigenvalues. These eigenvalues are made even smaller in practice due to computer rounding error. As seen earlier, it is the presence of these small eigenvalues which leads to inaccurate and unstable OLS estimates of β. This suggests that better estimates of β may be obtained by effectively discarding these small eigenvalues from the estimating procedure. The GIE attempts to do just that. By considering a range of "effective" ranks of X'X obtained by discarding its near zero eigenvalues the GIE obtains a TMSE less than TMSE($\hat{\beta}$). The GIE and its properties as outlined by Marquardt (1970) are discussed below.

For the CLRM (1) consider the OLS estimate $\hat{\beta}=(X'X)^{-1}X'y$. Suppose the rank of X'X is r, where $r<N$, then $(X'X)^{-1}$ does not exist. The "solution" of $y=X\beta+\varepsilon$ can be written in the form

$$\beta^{+} = A^{+}X'y$$

where A^{+} is a suitably defined matrix of rank r. To see this, consider the transformed model

$$y = X^*\alpha+\varepsilon.$$

as defined in (5). Now $(X'X)^{-1}=P\Omega^{-1}P'$ if $X'X$ is of full rank N. But rank $(X'X)=r$ implies that the last $(N-r)$ ordered elements of Ω are zero (or nearly so if $X'X$ is "nearly" singular).

Partition P as follows:

$$P = (P_r | P_{N-r})$$

where P is $N\times r$, P_{N-r} is $N\times(N-r)$.

Partitition Ω similarly;

$$\Omega = \left(\begin{array}{c|c} \Omega_r & 0 \\ \hline 0 & \Omega_{N-r} \end{array}\right)$$

where Ω_r is $r\times r$, Ω_{N-r} is $(N-r)\times(N-r)$.

Now by supposition, $\Omega_{N-r}=0$. So,

$$X'X_r = (P_r|P_{N-r}) \left(\begin{array}{c|c} \Omega_r & 0 \\ \hline 0 & 0 \end{array}\right) \left(\begin{array}{c} P_r' \\ \hline P_{N-r}' \end{array}\right)$$

$$= (P_r\Omega_r|0) \left(\begin{array}{c} P_r' \\ \hline P_{N-r}' \end{array}\right)$$

$$= P_r\Omega_r P_r'$$

and $(X'X_r)^{-1}=P_r\Omega_r^{-1}P_r'$.

Let

$$A_r^+ = (X'X_r)^{-1}$$

then A_r^+ is a generalized inverse, and

$$\hat{\beta}_r^+ = A_r^+ X'y$$

is defined to be the generalized inverse estimator of β.

A_r^+ is a generalized inverse of rank r only if X'X is of rank r. In order for the GIE to be valid, the effective rank (r) of X'X must be determined. In general, there is an "optimum" value of r for any problem but it is desirable to examine the GIE for a range of admissable values of r for the following reason.

In practice, X'X has eigenvalues falling into two groups (note that the rank of X'X is the number of non-zero eigenvalues of X');

1. substantially greater than zero
2. slightly greater than zero.

Those eigenvalues in group 2 can further be differentiated as; (a) zero but for rounding error; (b) genuinely non-zero. In practice, the distinction between (a) and (b) may be unclear and there will be uncertainty regarding the actual rank of X'X. In fact there may be a range of acceptable ranks that would be reasonable choices. The GIE however, has the ability to determine the generalized inverse for any assigned rank in this reasonable range. The ranks and hence their corresponding GIE are deemed admissable (optimal) by the criterion of TMSE. That is, $\hat{\beta}_r^+$ is an admissable estimator if TMSE $\hat{\beta}_r^+ <$ TMSE $\hat{\beta}$.

For $\hat{\beta}_r^+$, the assigned rank may be either an integer, r, ($1 \leq r \leq N$) representing the number of terms retained in

$$A_r^+ = \sum_{j=1}^{r} 1/\lambda_j S_j S_j'$$

(where S_j is the normalized eigenvector corresponding to λ_j) or a continuous variable ($0 < r \leq N$) obtained by including all terms in A_r^+ for which j is less than or equal to the integer part of r, plus the fraction in the next term by which r exceeds its integer part. The choice of the range of values of r is based on inspection of the spectrum of eigenvalues of X'X.

The above estimators are those that will be evaluated in this study. The following Chapter summarizes and discusses the findings of other empirical studies which have evaluated various alternatives to OLS in the presence of multicollinearity.

CHAPTER 4

RESULTS OF PREVIOUS MONTE CARLO STUDIES

Newhouse and Oman (1971) conducted a series of Monte Carlo experiments comparing various ridge estimators with each other and with OLS. Their model consisted of two explanatory variables with 100 observations on each. The values of the correlation between the two variables considered were .90 and .99. One value of σ_ε^2 and seven different true coefficient vectors were considered.

The following ridge estimators were evaluated. They pertain to the class of estimators defined by

$$\hat{\beta}(k) = (X'X+K)^{-1}X'y.$$

1. The first estimator (identified as R1A by the authors) set K = KI. Also X'X was in the form of a correlation matrix. The value of k was chosen by considering the sum of squared errors,

$$f[\hat{\beta}(k)] = [y-X\hat{\beta}(k)]'[y-X\hat{\beta}(k)]$$

as a function of $\rho = ||\hat{\beta}(k)||$, the length of the estimated coefficient vector. The value of k corresponding to the value of ρ for which

$$g = \frac{d^2(f^{1/2})}{d\rho^2}$$

is maximized is the one chosen. The estimator is claimed by the authors to be the mathematical equivalent of the ridge trace method of choosing k. However, as the authors recognize, it has not been shown that a maximum for g always exists. In fact, g consistently failed to achieve a maximum for the models under consideration. In such cases, k was assigned the arbitrary value of .5. It turned out that R1A performed worse than OLS for most values of β.

2. R1B was identical to R1A except that the variables were not first standardized. R1A and R1B gave considerably different estimates for the same value of β, but their performance averaged over all the β vectors was similar.

3. R1C was the iterative procedure for determining K in the orthogonal model presented by Hoerl and Kennard (1970a). Its performance, in terms of TMSE, was markedly worse than OLS.

4. R3A was the same as R1C except that

$$k_i = k = \frac{\hat{\sigma}_\varepsilon^2}{||\hat{\beta}||^2}, \qquad \text{for all } i,$$

were the diagonal elements chosen for K.

5. R3B was the same as R3A except that the variables were not standardized.

6. R3C appears as ORE2 in this paper.

All three of these estimators were approximately equal to OLS for r=.9. For r=.99 they were better or worse than OLS depending on the value of β.

Newhouse and Oman concluded that, "All the ridge estimators ... did worse than OLS for at least some choices of the true coefficient vector. Moreover, in our opinion, the ridge estimators failed by a sufficient number of cases to preclude their use".

A study by McDonald and Galarneau (1975) uses a model with 3 explanatory variables with 100 observations on each. They consider four different sets of correlation (.64, .81, .90, .98) between the variables as well as 7 values of σ_ε and 2 values for β.

They evaluate ORE4a and ORE4b presented in this paper with OLS by means of TMSE as a function μ, the degree of correlation, β and σ_ε. They also compare these rules with the estimators proposed by Newhouse and Oman for the two variable case.

ORE4a and ORE4b proved to be better than OLS for only some values of μ, σ_ε and β, however the reduction in TMSE achieved in some cases led the authors to recommend their use.

Guilkey and Murphy (1975) introduce and evaluate a modification of GRR. The authors refer to it as directed ridge regression (DRR). It was developed on the basis of the following discussion:

It is known that GRR may substantially reduce the total variance relative to OLS. However this is done at the expense of introducing bias into the estimating procedure. DRR is a modification of GRR which attempts to reduce the bias associated with GRR while maintaining the reduction in total variance afforded by GRR. As was discussed previously, the individual OLS estimates which are most affected by the presence of multicollinearity are those which correspond to small eigenvalues of the X'X matrix. The magnitude of an eigenvalue dictates whether or not the variance of an OLS estimate will be unduly large - the smaller the magnitude, the larger the variance of the individual OLS estimate. This is evident from the formula for the variance of an OLS estimate, $\hat{\beta}_i$, which is

$$\frac{\sigma_\varepsilon^2}{\lambda_i}.$$

DRR alters only those diagonal elements of X'X which correspond to relatively small eigenvalues. Consequently, only the individual estimates corresponding to small eigenvalues will be biased while those corresponding to large eigenvalues (which do not contribute much to the imprecision of the OLS estimates) remain unbiased. The end result should be a reduction in total bias over GRR, a reduction in total variance over OLS and finally a reduction in TMSE over both OLS and GRR.

Specifically, Guilkey and Murphy introduce and evaluate the following DRR estimators:

Recall that the GRE is

$$\hat{\alpha}(K) = (\Omega+K)^{-1}(X^*)'y$$

where Ω is the diagonal matrix whose diagonal elements are the eigenvalues, λ_i, of X'X. In order to apply DRR it must be established which of these eigenvalues are to be considered as "small". Guilkey and Murphy consider an eigenvalue, λ_i, small if

$$\lambda_i < 10^{-c}\lambda_{max},$$

where λ_{max} is the largest eigenvalue of X'X and c is some arbitrary constant. (The authors considered 3 values of c in their evaluation, namely 1, 2 and 3).

Once the small eigenvalues have been identified the authors construct an estimator called DRE1, which is based on an iterative procedure for selecting values of

k_i as suggested by Hoerl and Kennard (1970a). Here, however, only those values of k_i corresponding to the small eigenvalues of X'X were added to the diagonal elements of Ω.

A second variation given by Guilkey and Murphy, called DRE2, allows $k_i=k$ (where i is selected as above) to increase until the unexplained variance has increased from $\hat{\sigma}_\varepsilon$ to $(\hat{\sigma}_\varepsilon^2+q\hat{\sigma}_\varepsilon^2)$ where q is some arbitrary constant (the authors used q=.1, .05 and .01 in their study).

The evaluation used a sample of size 30. Different degrees of multicollinearity and values of the true coefficient vector were considered in the experiment. The number of regressors used was 2 and 5. DRE1 and DRE2 were compared with GRR and OLS, the comparative criterion being TMSE.

The results obtained by the authors for the 2 regressor model showed that DRE1 and GRR usually provided substantial reductions in TMSE relative to OLS with DRE1 generally slightly better than GRR. DRE2 consistently provided a reduction in TMSE, though not as substantial as DRE1 and GRR. DRE2 was found to be less biased than DRE1 and GRR.

For the 5 regressor model, GRR and DRE2 (with c=1 and 2) consistently provided an improvement over OLS. In general, DRE1 (c=1 and 2) resulted in a dramatic improvement over OLS in terms of TMSE. Also DRE1 proved better than GRR and DRE2 however DRE2 was less biased.

Hoerl, Kennard and Baldwin (1975) evaluate ORR where k is chosen to equal

$$\frac{N\hat{\sigma}_{\varepsilon}^{2}}{\hat{\beta}'\hat{\beta}},$$

where N is the number of independent variables. They found that significant improvement over OLS in terms of TMSE is achieved by using this estimate of k.

In a later paper, Hoerl and Kennard (1976) note that in general, $\hat{\beta}'\hat{\beta}$ will be larger than $\beta'\beta$ particularly in the presence of multicollinearity. This fact may then make $\frac{N\hat{\sigma}_{\varepsilon}^{2}}{\hat{\beta}'\hat{\beta}}$ too small. To overcome this difficulty the authors suggest an iterative method. They consider a sequence of estimates of β and k given as follows:

$$\hat{\beta}, k_1 = \frac{N\hat{\sigma}_{\varepsilon}^{2}}{\hat{\beta}'\hat{\beta}};\ \hat{\beta}(k_1), k_2 = \frac{N\hat{\sigma}_{\varepsilon}^{2}}{(\hat{\beta}(k_1))'\hat{\beta}(k_1)};$$

$$\hat{\beta}(k_2), k_3 = \frac{N\hat{\sigma}_{\varepsilon}^{2}}{(\hat{\beta}(k_2))'\hat{\beta}(k_2)};\ \ldots;\ \hat{\beta}(k_t).$$

Here $\hat{\beta}(k_1)$ is used in place of $\hat{\beta}$ (the iterative method using $\hat{\beta}$ is suggested in Hoerl and Kennard (1970a)) since $(\hat{\beta}(k_1))'\hat{\beta}(k_1)$ is smaller than $\hat{\beta}'\hat{\beta}$. Several methods for terminating the iteration are discussed in Hoerl and Kennard (1976).

The conclusions of an empirical evaluation of the above procedure were that a significant reduction in TMSE was achieved and that the improvement increases as the degree of multicollinearity increases.

Lawless and Wang (1976) evaluate a number of ridge and principal components estimators. The ridge estimators considered in this paper are of the following type;

$$\tilde{\alpha}_i = \left(\frac{\lambda_i}{\lambda_i+k_i}\right)\hat{\alpha}_i,$$

where $\hat{\alpha}_i$ is the OLS estimate of α_i in the canonical model (5). The estimators vary as to the method of choosing k_i, the parameter in GRR. A detailed discussion of the estimators is given in the above paper.

In addition to comparing the estimators in terms of TMSE error, the authors use a criterion called the total mean square error of prediction given by

$$M = \sum_{i=1}^{N} \lambda_i E\{(\tilde{\alpha}_i-\alpha_i)^2\}.$$

The major conclusion of an empirical evaluation was that, "the ridge estimators considered have much to recommend them on the basis of the mean square error criteria considered, whereas the principal component estimators do not offer a great improvement over OLS".

In general, the various ridge estimators evaluated in the above studies perform better than OLS in the presence of multicollinearity (with the exception of the ridge estimators evaluated by Newhouse and Oman (1971)). However, there is no clear indication of which ridge estimator is best as the performance of the various ridge estimators depends on such factors as σ_ε, β and X'X. The remainder of the thesis will attempt to clarify the relative merits of the alternatives to OLS under a variety of data conditions.

CHAPTER 5

DESIGN OF THE EXPERIMENT

The Monte Carlo Evaluation Study

The criterion for comparing estimators will be TMSE. In particular for ORR (as well as for the other estimators considered) the TMSE of $\hat{\beta}(k)$ depends upon the true coefficient vector, β; the error variance, σ_ε^2; and the sample data matrix X (through the eigenvalues of X'X). Thus different values of β, σ_ε^2 and the degree of multicollinearity (as reflected in the elements of X'X when in correlation form) will be used in the experiment. To determine the appropriate values it is necessary to investigate TMSE $\hat{\beta}(k)$ as a function of β, σ_ε^2 and μ (the degree of multicollinearity) respectively.

1. TMSE $[\hat{\beta}(k)]$ as a function of β

Newhouse and Oman (1971) have observed that if TMSE $[\hat{\beta}(k)]$ is regarded as a function of β, with σ_ε^2, k and X fixed, then subject to the constraint that the Euclidean length of β, $||\beta||$, be one, the TMSE $[\hat{\beta}(k)]$ is minimized when β is the normalized eigenvector (V_L) corresponding to the largest eigenvalue of the X'X matrix (note $||V_L||=1$). Similarly, TMSE $[\hat{\beta}(k)]$ is maximized when $\beta = V_s$, where V_s is the normalized eigenvector corresponding to the smallest eigenvalue of X'X.

It should be noted that the above result requires that k be fixed. For any given X however, the absence of a criterion for explicitly determining the optimal value of k means that k is not fixed.

Despite this, the above result concerning TMSE $[\hat{\beta}(k)]$ and β presents a rationale for choosing values of β. The two values of β, V_L and V_s will be used throughout for all experiments.

To obtain V_L and V_s explicitly for a given X matrix, the ordered eigenvalues,

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N$ of X'X are determined. For λ_1 and λ_N we solve the equations

$$(X'X-\lambda_1)Z_L = 0 \quad \text{and} \quad (X'X-\lambda_N)Z_s = 0$$

to obtain the eigenvectors Z_L and Z_s corresponding to λ_1 and λ_N respectively. We then normalize Z_L and Z_s to obtain V_L and V_s i.e.

$$V_L = \frac{Z_L}{\sqrt{\sum_{i=1}^{N} Z_{Li}^2}} \quad \text{and} \quad V_s = \frac{Z_s}{\sqrt{\sum_{i=1}^{N} Z_{si}^2}} .$$

2. TMSE$[\hat{\beta}(k)]$ as a function of σ_ε.

The TMSE$[\hat{\beta}(k)]$ is the sum of the total variance and squared total bias of the parameter estimates. The total variance TV$[\hat{\beta}(k)]$ is given by

$$TV[\hat{\beta}(k)] = \sigma_\varepsilon^2 \sum_{i=1}^{N} \frac{\lambda_i}{(\lambda_i+k)^2} .$$

If $\sigma_\varepsilon^2 \varepsilon [0,1]$, then TV$\hat{\beta}(k)$ (and hence TMSE$[\hat{\beta}(k)]$) is a strictly increasing function of σ_ε^2 approaching o as $\sigma_\varepsilon^2 \to o$ and being maximized when $\sigma_\varepsilon^2=1$. Many of the rules considered here for choosing a value of k in ORR (as well as some other estimators) are functions of $\hat{\sigma}_\varepsilon^2$, an estimate of σ_ε^2. By allowing σ_ε^2 to take on different values its effect on the choice of k can be determined.

Values for ε_i, where ε_i, i=1,...,T are independent $N(o,\sigma_\varepsilon^2)$ pseudo-random numbers are generated. In view of the above discussion the following values of σ_ε will be used for all experiments; 0.1, 0.5 and 1.0.

3. TMSE$[\hat{\beta}(k)]$ as a function of μ, the degree of multicollinearity.

As mentioned in the introduction, as the degree of multicollinearity increases (or alternatively, as the vectors of X deviate further from orthogonality), λ_N, the smallest eigenvalue of the X'X matrix, approaches zero. The TMSE$[\hat{\beta}(k)]$ (as well as the TMSE of all other estimators considered in this paper) depends on the data matrix X through the eigenvalues of the X'X matrix. So the TMSE of our estimators is a function of the degree of multi-collinearity.

If $X=\{x_1,...,x_N\}$ is a set of explanatory variables, then the degree of multicollinearity of the set X depends on the correlation within all possible subsets of X (with more than one element) and the number of these subsets which are highly correlated. For our purposes we will only be concerned with the correlation of subsets of size two.

The desired degree of multicollinearity may be obtained by generating the observations on the explanatory variables as follows:

Let

$$x_{tn} = \sqrt{1-\mu^2}\, z_{tn}+\mu z_{t,N+1}$$

for $t=1,...,T$ and $n=1,...,N$ where z_{tn}, $n=1,...,N+1$ are independent $N(o,1)$ pseudo - random numbers and μ is specified so that the correlation between any two x_{tn} is μ^2. In this study $\mu=.6,.8,.9$ and $.95$. These variables are then standardized so that X'X is in correlation form. The role of the standardization is presented in Appendix I. Observations on the dependent variable are determined by $y_t=\beta_1 x_{t1}+...+\beta_N x_{tN}+\varepsilon_t$, $t=1,...,T$.

It is common in practice to have models with a relatively large number of variables and a small number of observations. Thus we set N=6 and T=20. Fifty samples are generated by allowing the random error ε (and hence the dependent variable) to change while the explanatory variables (X) and the true coefficient vector (β) remain fixed.

For each sample so constructed the following calculations are made (for a given β, σ_ε^2 and μ):

1. $\hat{\beta}(k) = (X'X+kI)^{-1}X'y$

   (a) For the ridge trace method, ORE 1, $\hat{\beta}(k)$ is evaluated for 15 values of k (0,.01,...,.05,.1,..., .9). Note that $\hat{\beta}(0)=\hat{\beta}$.

   (b) For ORE2 and ORE3 we calculate $\frac{\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}}$ and $\frac{N\hat{\sigma}_\varepsilon^2}{\hat{\beta}'\hat{\beta}}$ respectively.

   (c) For ORE4a and ORE4b, $(\hat{\beta}(k))'\hat{\beta}(k)$ is evaluated for 201 values of k (0(.0005),.05(.001),.1(.01), .5(.05),1.0). k is chosen so that $(\hat{\beta}(k))'\hat{\beta}(k) = \hat{\beta}'\hat{\beta}-\hat{\sigma}_\varepsilon^2 \sum_{n=1}^{N} 1/\lambda_n$.

2. For GRR, we generate the model $Y=X\beta+\varepsilon$ as above. For each X matrix we must determine XP' and $P\beta$ where P is the orthogonal transformation such that $P'X'XP=\Omega$.

Now $(X'X)P=P\Omega$. Let $P=(V_1,\ldots,V_N)$ and let $\lambda_n (1\le n\le N)$ denote the nth diagonal element of Ω, then

$$(X'X)V_n = \lambda_n V_n, \quad n=1,\ldots,N$$

implying that V_n is an eigenvector corresponding to the eigenvalue λ_n. Thus the columns of P are the eigenvectors corresponding to the eigenvalues of X'X. We then form the model

$$y = (XP')P\beta+\varepsilon$$
$$= X^*\alpha+\varepsilon$$

and employ GRE1 to determine the optimal k_i values and $\hat{\alpha}(K)$. $\hat{\alpha}(K)$ is then transformed to $\hat{\beta}(K)$ by

$$\hat{\beta}(K) = P'\hat{\alpha}(K).$$

3. $b^* = \left(\frac{\hat{\beta}'X'y}{\hat{\sigma}_\varepsilon^2+\hat{\beta}'X'X\hat{\beta}}\right)\hat{\beta}$, the MMSELE of β.

4. (a) $SSE1 = \delta\hat{\beta}\hat{\beta}'(I+\delta\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}$ for 15 values of δ (same as 15 k values). An admissable δ value is obtained by the method of the ridge trace.

(b) $SSE_2 = [1+\gamma s^2(\hat{\beta}'\hat{\beta})^{-1}]\hat{\beta}$ for

$$\gamma = \frac{N-2}{T-N+2} = 0.25$$

5. $GIE = A_r^+X'y$ where r is chosen after inspection of the eigenvalue spectrum of $X'X$.

Throughout the above discussion we have assumed that $X'X$ is in correlation form and $X'y$ represents the vector of correlations between the dependent variable and each explanatory variable. The coefficient estimates obtained from this standardized model are the "standardized" coefficient estimates. Appendix 1 provides an interpretation of the standardized coefficient estimates as well as the method for transforming the standardized coefficient estimates to the non-standardized form. It should be noted that the properties of the estimators considered in this paper have been developed for the standardized model. In particular, the TMSE properties are developed in the context of this model. That these TMSE properties hold in the non-standardized model is discussed in Appendix I. This result is important since if $TMSE(\beta^*) < TMSE(\hat{\beta})$ holds in the standardized model we wish this result to be preserved when the coefficient estimates are in the original units.

The experiment presented above is repeated for all values of μ, σ_ε^2 and β.

Criterion For Evaluating Estimators

The estimators (and rules) presented here will be compared with each other and with OLS through the Monte Carlo simulation method. To evaluate their relative performance, a criterion which is deemed desirable or optimal for estimators to possess must be chosen.

When confronted with a small sample size, the desirable aspects of an estimator are that it have a small total bias and total variance. Ideally, we would prefer an estimator which is unbiased and has a smaller total variance than any other estimator, including biased ones. Unfortunately, in the presence of multicollinearity, the acceptance of an unbiased estimator (OLS) is made at the expense of a large total variance. Given that there is a trade-off between the size of the total variance and the amount of total bias of an estimator, the alternatives to OLS presented here hope that by allowing a little bias, the total variance can be greatly reduced. The ideal of unbiasedness and minimum total variance is then replaced by the next best thing - an estimator is deemed desirable if its total bias and total variance combined is smaller than any other estimator. This property is manifested in the TMSE criterion.

Being forced to choose among biased estimators there may still be those who prefer a small total bias (despite the size of the total variance) or vice-versa. Thus in addition to TMSE, the total bias and total variance of each estimator is reported separately.

In summary, the following criteria will be used to evaluate the proposed estimators. (Note that the total bias, total variance and total mean square error have been estimated on the basis of 50 samples).

1. Total bias $\beta^* = \sum_{i=1}^{6} (\bar{\beta}_i^* - \beta_i)$

where $\bar{\beta}_i^* = \dfrac{\sum_{j=1}^{50} (\beta_i^*)_j}{50}$ is an estimate of $E\beta_i^*$.

2. Total variance β^*

$$\sum_{i=1}^{6} \left(\frac{\sum_{j=1}^{50} (\beta_i^* - \bar{\beta}_i^*)^2{}_j}{50} \right) .$$

3. $TMSE\beta^* = (\text{total bias } \beta^*)^2 + \text{total variance } \beta^*$.

In addition to the above, the following quantities were recorded.

1. The average squared length of the estimator.
2. The average k value of the ORE's.
3. The average Q value for ORE4a and ORE4b.

Evaluation of the Ridge Trace Method of Choosing k in ORR

The performance of the ridge trace method of choosing k, as proposed and accredited by Hoerl and Kennard (1970a, p.65) and several other investigators, has never been evaluated save in several specific applications. This is due to its subjective nature; the choice of k depends on who is examining the ridge trace. Despite this drawback, the ridge trace method is being used in practice (eg. Watson and White (1976)) and is claimed by Hoerl and Kennard to be the method for achieving a better estimate of β. Consequently, an attempt will be made here to evaluate this technique.

To do this, 14 individuals with general experience in the area of econometrics were each asked to choose their k values for 50 graphical representations of the ridge trace. The standard deviation of the k's so chosen was recorded to determine the subjective error associated with the ridge trace method. An interval around the k value chosen will be constructed and the ORE evaluated at the endpoints (as well as at the mean value) to determine the sensitivity of the TMSE to the k value chosen. The above evaluation will provide an indication of the sensitivity associated with the ridge trace selection of k in ORR.

CHAPTER 6

RESULTS AND CONCLUSIONS

As discussed previously, the use of ridge regression as a viable alternative to OLS under conditions of multicollinearity is dependent on finding a method which consistently chooses a k value such that $TMSE[\hat{\beta}(k)] < TMSE(\hat{\beta})$. Consequently, research has focused for some time on finding methods of selecting a k value which would provide ORR with such consistency.

One such method of selecting k values, the ridge trace method, was claimed by Hoerl (1962) to be the best method. However, the subjective nature of this method resulted in research being directed at developing mathematical methods for selecting an admissable k value. Several such mathematical methods subsequently appeared in the literature as possibly being better alternatives to the subjective ridge trace method. The subjectivity of the ridge trace method, of course, lies in the fact that different individuals are likely to choose different values of k through the visual interpretation of a given ridge trace.

In addition to the various ridge estimators, several other estimators appear in the literature which claim to be superior to OLS (and in some cases to ORR) in the presence of multicollinearity.

This thesis investigates the claims made above by evaluating the relative performance of these estimators (in terms of TMSE) under a variety of data conditions. In particular, the purpose of this thesis is to:

1. Estimate the "subjective error" associated with the ridge trace method of selecting k for ORR;
2. Compare the performance of ORR using the subjective ridge trace method of selecting k with the performance of ORR when the mathematical methods are used to select k;
3. Compare the performance of all ridge estimators and all other estimators considered with OLS and with each other under a variety of data conditions;
4. Investigate and compare the performance of all estimators as functions of β, the true coefficient vector; σ^2, the variance of the error term; and μ, the degree of multicollinearity.

CONCLUSIONS

In summary, the following are the conclusions obtained (a detailed discussion of these conclusions is presented throughout the remainder of this chapter):

1. The "subjective error" of the ridge trace method of selecting k was estimated to be 0.062. That is, on the average, the value of k chosen by using the ridge trace method will vary by 0.062 from individual to individual. The subjective error of 0.062 amounts

to 44% of the average value of k (0.14) chosen by fourteen individuals (over a sample of 50).

2. The ridge trace method of selecting k is a consistent method providing improvement in TMSE (or in a few cases, no significant increase) relative to ORR using a variety of mathematical means to select k. The ridge trace method is considered to be the best method for choosing k (relative to the other methods considered).

3. The ridge trace estimator ranks first in overall performance relative to OLS and relative to all other estimators considered. Furthermore, in the presence of multicollinearity one of the alternative estimators should be used in place of OLS. In particular, the ridge trace estimator or ORE3 can be used to ensure a better performance in most cases.

4. The TMSE of the estimators considered (and consequently, the performance of these estimators relative to OLS) was demonstrated to vary according to the true value specified for σ_ε, μ and β. It was found that the performance of all estimators relative to OLS improves as μ, the degree of multicollinearity increases. When β_S is the true coefficient vector, the relative TMSE of the alternative estimators and OLS appears to be a decreasing function of σ_ε for values of μ. When β_L was the true coefficient vector, the TMSE

of SSE2, ORE1, SSE1, G1E and GRR relative to TMSE(OLS) decreased as σ_ε increased. No definite relationship was found for ORE2 and ORE3. The relative performance of the estimators was found to be particularly sensitive to the true value of β. Estimators which performed well relative to OLS when β_S was the true coefficient vector, saw the situation reversed when β_L was the true coefficient vector. (This result is discussed further in this Chapter and in Chapter 7).

RESULTS

Appendix 3 contains the results of the Monte Carlo evaluation of the estimators considered. Table A1 contains the X'X matrices (correlation matrices) for μ=0.6, 0.8, 0.9 and 0.95. Table A2 contains the eigenvalues for each X'X matrix. The true coefficient vectors, β_S and β_L (being the eigenvectors corresponding to the smallest and largest eigenvalues of each X'X matrix respectively) are presented in Table A3. Table A4 contains the average (over 50 samples) k values for each ridge estimator and the average δ value for SSE1. Table A5 gives the average Q value and the number of negative Q values (out of 50 samples) for ORE4a and 4b. The average (over 50 samples) of each estimator is presented in Table A6. Tables A7, A8 and A9 contain the average total bias, average total variance and average total mean square error of each estimator respectively. The value of $\frac{\sigma_\varepsilon^2}{\alpha^2_{max}}$ for each estimator, degree of multicollinearity and each value of β, appears in Table A10. Finally, Table A11 gives the number of k values (out of 50) in the interval $\left(o, \frac{\sigma_\varepsilon^2}{\alpha^2_{max}}\right)$ for each ridge estimator.

1. Subjective Error of the Ridge Trace Method

As discussed at the beginning of the Chapter, the ridge trace method of selecting an admissable k value in ORR was claimed by Hoerl and Kennard (1970a, p.65) to be the best method for obtaining a better estimate of β under conditions of multicollinearity. However, the ridge trace method is subjective - given any graphical representation of the ridge trace, different individuals will choose different values of k no matter how stringently they follow the guidelines given by Hoerl and Kennard. This section proposes to estimate how subjective the ridge trace method is and how the subjective error affects the performance of ORR (in terms of TMSE) relative to the performance of ORR when various mathematical means are used to select k.

The subjective error associated with the ridge trace method was estimated in the following manner: Fifty samples were generated with σ_ε=0.5, μ=0.9 and $\beta=\beta_S$. For each sample, $\hat{\beta}(k)=(X'X+kI)^{-1}X'y$ was evaluated for 15 values of k between 0 and 1. A ridge trace was generated for each of the 50 samples. Figure 1 displays the ridge trace for one such sample.

Although the graphical representation of the ridge trace changes from sample to sample (and for different values of μ, σ and β) it is assumed that the subjective error will be constant across samples. This assumption is based on the premise that the subjective error is not a function of the ridge trace method per se, but rather a function of who is examining it. (For comparative purposes, an example of the ridge trace for SSE1 with $\mu=0.9$, $\sigma_\varepsilon=0.5$ and $\beta=\beta_S$ appears in Figure 2).

To estimate the subjective error, fourteen individuals with experience in the area of econometrics were each provided with copies of the 50 ridge trace graphs as outlined above. Each individual was requested to choose a value of k for each ridge trace according to the guidelines given by Hoerl and Kennard. The average standard deviation (over 50 samples) of the fourteen sets of k values so chosen was taken to be the estimate of the subjective error associated with the ridge trace selection of k. The subjective error was estimated to be 0.062. The average k value (over 50 samples) of the fourteen sets of k values chosen was 0.14.

Thus, the value of k chosen via the ridge trace method can be expected to differ by an average of 0.062 from individual to individual due to the visual interpretation

Figure 1 Ridge Trace for $\mu=.9$, $\sigma_{\varepsilon}=.5$, $\beta=\beta_S$ for ORR

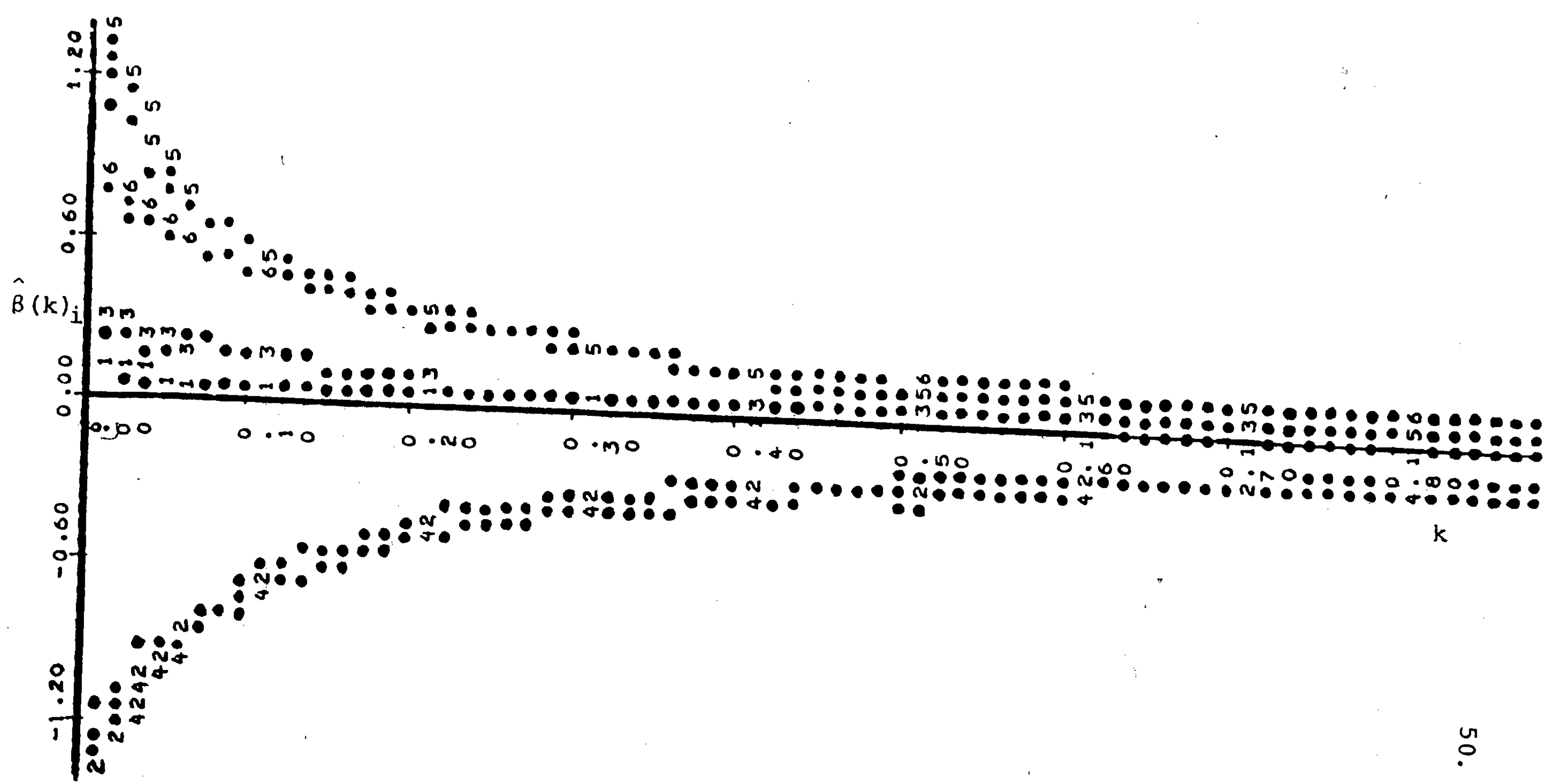
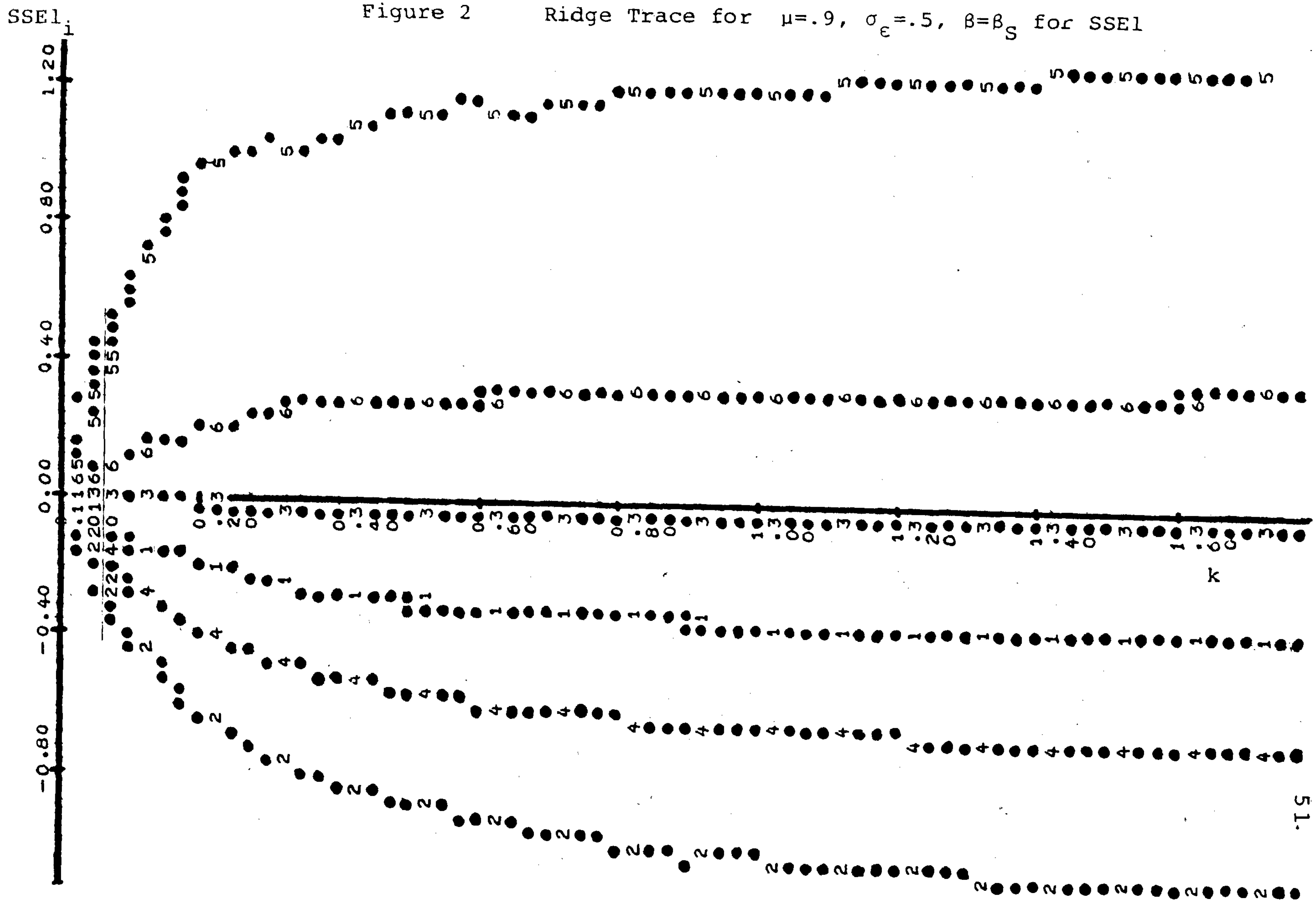


Figure 2 Ridge Trace for $\mu=.9$, $\sigma_{\varepsilon}=.5$, $\beta=\beta_S$ for SSE1



required by the ridge trace method. This subjective error of 0.062 amounts to 44% of the average k value (0.14).

The concern caused by the subjective error of the ridge trace method of selecting k is that it may be inconsistent with respect to other means of selecting k. That is, on average, due to the subjective error, TMSE$\hat{\beta}(k)$ where k is chosen from the ridge trace may exceed TMSE$\hat{\beta}(k)$ with k selected by some mathematically consistent method. If in fact this were the case, a researcher would be wise to opt for the mathematical method of choosing k. To determine if this indeed is the case, the following experiment was conducted.

The average k value was obtained from a sample of 50 using the ridge trace method for each value of σ_ε, μ and β. This average k value was adjusted by the average variation of 0.062 obtained from the subjective error evaluation. That is, if k_A is the average k value chosen (over 50 samples) for a given value of μ, σ_ε and β, then TMSE$[\hat{\beta}(k_A)]$, TMSE$[\hat{\beta}(k_{min})]$ and TMSE$[\hat{\beta}(k_{max})]$, where $k_{min}=k_A-0.062$ and $k_{max}=k_A+0.062$ are averaged and compared to the TMSE of ORE2, 3 and 4a, b, the ridge estimators where mathematical methods are used to select k. Table 1 below reports the results of this comparison.

When β_S is the true coefficient vector, the TMSE of the ridge trace estimator, ORE1 is less than the TMSE of

Table 1 Comparison of TMSE of ORR. Ridge trace vs mathematical methods of selecting k.

TMSE

μ	σ	β_S					β_L				
		ORE1	ORE2	ORE3	ORE4 a	ORE4 b	ORE1	ORE2	ORE3	ORE4 a	ORE4 b
	0.1	.017	.011	.011	.011	.011	1.207	1.174	1.176	1.180	1.180
0.6	0.5	.144	.194	.183	.194	.194	1.346	1.292	1.329	1.393	1.393
	1.0	.284	.358	.261	.347	.328	1.655	1.612	1.712	1.908	1.908
	0.1	.047	.035	.035	.035	.035	1.685	1.663	1.664	1.672	1.672
0.8	0.5	.219	.408	.373	.407	.407	1.789	1.767	1.786	1.901	1.901
	1.0	.405	1632	.435	.603	.587	2.053	2.043	2.072	2.404	2.272
	0.1	.087	.114	.114	.114	.114	1.903	1.887	1.888	1.902	1.902
0.9	0.5	.462	.979	.795	.972	.972	2.011	2.004	2.077	2.184	2.184
	1.0	.731	1.234	.735	1.008	1.098	2.238	2.307	2.268	3.223	2.376
	0.1	.099	.387	.383	.386	.386	2.018	1.998	1.999	2.019	2.019
0.95	0.5	.587	2.171	1.499	2.072	2.017	2.123	2.155	2.130	2.394	2.330
	1.0	.740	2.374	1.272	1.628	2.166	2.325	2.553	2.411	3.856	2.508

ORE2, 3 and 4a, b for almost all values of σ_ε and μ (for μ=0.6 and 0.8 and σ=0.1, TMSE of ORE2, 3, 4a and b is slightly less than the average TMSE of ORE1). Furthermore, as the degree of multicollinearity increases, the performance of the ridge trace estimator improves relative to the other estimators. On the basis of these results, it is concluded that, on average, the ridge trace method of selecting k, in spite of its subjective nature, is consistent, and hence is a better method than the mathematical methods considered (in terms of providing an average improvement in TMSE) when β_S is the true coefficient vector.

The results are not as apparent when β_L is the true coefficient vector. In the majority of cases, TMSE(ORE2)<TMSE(ORE1), however only marginally so. ORE3 exhibits slightly better results than ORE1 for small values of σ_ε with the results reversed for σ_ε=1 and for μ=0.95. Both ORE4a and b in general have a greater TMSE than ORE1.

It is concluded that the ridge trace method, in general, is superior to the mathematical methods considered for ORR. This conclusion is based on the consistent performance and significant improvement in TMSE of ORE1 when β_S is the true coefficient vector. When β_L is the true coefficient only one estimator, ORE2, performs better (in most cases) than ORE1 and only by a slight margin. Furthermore, as μ increases the performance of ORE1 improves (and even exceeds the performance of all other estimators).

In general then, it is concluded that the ridge trace method of selecting k is a consistent method and that improvement in TMSE (or in a few cases, no significant increase) results relative to ORR using mathematical means to select k.

2. Relative Performance of the Estimators

This section compares the performance of each estimator considered relative to OLS and then relative to each other. The criterion of performance will be TMSE. In particular, the quantity M, defined by the ratio,

$$M = \frac{TMSE(\beta_1)}{TMSE(\beta_2)}$$

is used to compare any two estimators, β_1 and β_2.

A. Performance of Estimators relative to OLS

This section compares the performance of ORE1, 2, 3, 4a, b, MMSELE, GRR, G1E, SSE1, and SSE2 with OLS. The ratio, $M = \frac{TMSE(\beta^*)}{TNSE(\hat{\beta})}$, where β^* is an estimator other than OLS, is used in making the comparison. Table 2 contains the value of M for each estimator as a function of

Table 2 M For All Estimators

Estimator	μ=.6			μ=.8			μ=.9			μ=.95		
β_S	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
ORE1	1.55	.73	.74	1.34	.53	.59	.76	.45	.51	.26	.25	.25
ORE2	1.0	.99	.93	1.0	.98	.91	1.0	.96	.86	1.0	.91	.81
ORE3	1.0	.93	.68	1.0	.90	.63	.99	.78	.51	.99	.63	.44
ORE4a	1.0	.98	.90	1.0	.98	.87	1.0	.95	.70	1.0	.87	.56
ORE4b	1.0	.98	.85	1.0	.98	.85	1.0	.95	.76	1.0	.85	.74
MMSELE	1.0	.96	.89	1.0	.95	.87	.99	.92	.84	.99	.88	.81
GRR	.48	.15	.05	.60	.26	.14	.55	.21	.12	.53	.20	.12
GIE	1.0	1.0	1.0	1.0	1.0	1.0	.99	.86	.78	.35	.28	.27
SSE1	1.01	.32	.25	.48	.29	.27	.33	.25	.24	.21	.21	.23
SSE2	1.0	1.05	1.18	1.0	1.04	1.13	1.0	1.0	1.11	1.0	1.03	1.07

Table 2 - continued

Estimator	μ=.6			μ=.8			μ=.9			μ=.95		
β_L	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
ORE1	1.03	1.05	1.04	1.01	1.01	1.0	1.01	1.0	.96	1.01	.98	.88
ORE2	1.0	1.01	1.01	1.0	1.0	1.0	1.0	1.0	.99	1.0	.99	.97
ORE3	1.0	1.03	1.07	1.0	1.01	1.01	1.0	1.0	.97	1.0	.98	.91
ORE4a	1.01	1.08	1.20	1.01	1.08	1.18	1.01	1.09	1.38	1.01	1.10	1.46
ORE4b	1.01	1.08	1.20	1.01	1.08	1.11	1.01	1.09	1.02	1.01	1.08	.95
MMSELE	1.0	1.01	1.02	1.0	1.0	1.01	1.0	1.0	1.01	1.0	1.0	1.01
GRR	4.28	4.16	3.52	3.12	3.10	2.79	2.78	2.78	2.51	2.70	2.58	2.26
GIE	1.0	1.0	1.0	1.0	1.0	1.0	1.0	.99	.97	1.0	.96	.87
SSE1	3.34	3.14	2.62	2.87	2.73	2.06	2.63	2.10	1.84	2.37	2.08	1.59
SSE2	1.0	.93	.82	1.0	.95	.89	1.0	.96	.92	1.0	.97	.96

β, σ_ε and μ. The results contained in this table are discussed below:

a. M as a function of β

In general, the performance of all estimators (with the exception of SSE2) is much worse when β is the eigenvector corresponding to the largest eigenvalue of X'X. For this β, M is in the neighbourhood of 1 meaning that most estimators are slightly better or worse (depending on the value of σ_ε and μ) than OLS.

The estimators which are most sensitive to the value of the true coefficient vector are GRR and SSE1. For β_S, GRR puts in an astounding performance with $.05 \le M \le .60$ for all values of μ and σ_ε. However, for β_L, $2.26 \le M \le 4.28$ for GRR. Similarly, for SSE1, $.21 \le M \le .48$ when β_S is the true coefficient vector and $1.59 \le M \le 3.34$ when β_L is the true coefficient vector.

When β_S is the true coefficient vector, the average ridge trace estimator, ORE1, performs well relative to OLS, with $.25 \le M \le 1.55$. For β_L, $.88 \le M \le 1.05$ for ORE1. The remaining ridge estimators perform consistently better (or at least as well as) OLS. ORE3 performs the best in this group with $.44 \le M \le 1.00$. For ORE2, ORE4a and ORE4b the respective ranges of M are [.81, 1.00], [.56, 1.00] and [.74, 1.00]. For β_L, these estimators, on the whole, perform worse than OLS in varying degrees. For ORE2, ORE3, ORE4a and ORE4b, M is confined to the following intervals respectively, [.97, 1.01], [.91, 1.07], [1.01, 1.46] and [.95, 1.20].

For both β_S and β_L, GIE performs as well as or better than OLS. For β_S, $.27 \leq M \leq 1.00$ and for β_L, $.87 \leq M \leq 1.00$. It should be noted that for $\mu = .6$ and $.8$ the GIE is in fact $\hat{\beta}$. This is so since the value of r choosen was 6, the rank of X'X. Had we not restricted ourselves to integer values of r, the performance of the GIE may have been improved in these cases. (For a complete discussion of the selection of r the reader is referred to Marquardt (1970)).

The performance of SSE2 is better for β_L than it is for β_S; a reversal of the situation experienced by all other estimators. For β_S, $1.0 \leq M \leq 1.18$ and for β_L, $.82 \leq M \leq 1.0$. In fact for β_S, SSE2 consistently performs worse than all other estimators however, for β_L it performs better than all other estimators in every case save $\mu = .95$, $\sigma_\varepsilon = 1$.

b. M as a function of σ_ε

When β_S is the true coefficient vector, M appears to be a decreasing function of σ_ε (for all values of μ) and for all estimators except SSE2. When β_L is the true coefficient vector, and for SSE2, ORE1, SSE1, GIE and GRR, M appears to be a decreasing function of σ_ε for all values of μ. ORE4a and MMSELE are increasing functions of σ_ε for all values of μ. For ORE2 and ORE3, M increases as σ_ε increases for μ=.6 and .8 and decreases for μ=.9 and .95.

c. M as a function of μ

Regardless of the true coefficient vector, for all estimators, M decreases as μ increases i.e. the performance of all estimators relative to OLS improves as the degree of multicollinearity increases.

As seen above, the performance of the estimators considered depends on σ_ε, the standard deviation of the error term; μ, the degree of multicollinearity; and most critically on β, the true coefficient vector. Since none of these parameters are known in practice, a researcher would be in doubt as to which estimator to use given a particular sample. To aid in making such a decision, we provide an overall ranking of the estimators which takes into consideration the following:

- the overall performance of each estimator (over all values of σ_ε, μ and β),
- the degree of improvement of each estimator over OLS and relative to each other.

Table 3 contains the aggregated TMSE of each estimator which is obtained by summing the TMSE of an estimator over all values of σ_ε and μ (for a given β). Tables 4 and 5 contain the ratio of the aggregated TMSE of the estimators for β_S and β_L respectively. For example, the entry in row 2, column 6 (2.09) in Table 4, is the ratio of the aggregated TMSE(ORE4a) to the aggregated TMSE(ORE1) meaning that the aggregated TMSE(ORE4a) is 2.09 times greater than the aggregated TMSE(ORE1). Tables 4 and 5, then provide the overall relative performance of the estimators.

Table 6 ranks the estimators in terms of the magnitude of the ratio of their aggregated TMSE to the aggregated TMSE(OLS) for β_S and β_L individually, and for β_S and β_L together. For example, the aggregated TMSE(ORE1) (for β_S and β_L) is .68 times that for OLS. This table provides the overall performance of the estimators relative to OLS in order of improvement.

Finally, Table 7 presents an overall ranking of the estimators for β_S and β_L individually and taken together. For example, for β_S and β_L together, the average ratio for ORE1 is .80. This was obtained by summing the ratio of the aggregated TMSE(ORE1) to the aggregated TMSE of each estimator and then dividing by 10, the number of estimators other than ORE1.

B. Overall Performance of the Estimators

The ridge trace estimator, ORE1, ranks first in overall performance relative to OLS and relative to all other estimators considered. However, this is based on an average performance over all values of σ_ε, μ and β. When

Table 3 Aggregated TMSE

Estimator	β_S	β_L
ORE1	3.719	22.345
ORE2	8.898	22.415
ORE3	6.096	22.442
ORE4a	7.777	26.108
ORE4b	8.315	23.645
MMSELE	8.686	22.682
GRR	1.732	65.951
GIE	8.419	22.336
SSE1	2.361	53.461
SSE2	10.658	21.410
OLS	10.232	22.564

Table 4 Ratio of Aggregated TMSE of Estimators for β_S

	ORE1	ORE2	ORE3	ORE4a	ORE4b	MMSELE	GRR	GIE	SSE1	SSE2	OLS
ORE1		2.39	1.64	2.09	2.24	2.34	.47	2.26	.63	2.87	2.75
ORE2	.42		.69	.87	.94	.97	.20	.95	.27	1.20	1.15
ORE3	.61	1.46		1.28	1.36	1.43	.28	1.38	.39	1.75	1.68
ORE4a	.48	1.14	.78		1.07	1.12	.22	1.08	.30	1.37	1.32
ORE4b	.45	1.07	.73	.94		1.05	.21	1.01	.28	1.28	1.23
MMSELE	.43	1.02	.70	.90	.96		.20	.97	.27	1.23	1.18
GRR	2.15	5.14	3.52	4.49	4.80	5.02		4.86	1.36	6.15	5.91
GIE	.44	1.06	.72	.92	.99	1.03	.21		.28	1.27	1.22
SSE1	1.58	3.77	2.58	3.29	3.52	3.68	.73	3.57		4.51	4.33
SSE2	.35	.84	.57	.73	.78	.82	.16	.79	.22		.96
OLS	.36	.87	.60	.76	.81	.85	.17	.82	.23	1.04	

Table 5 Ratio of Aggregated TMSE of Estimators for β_L

	ORE1	ORE2	ORE3	ORE4a	ORE4b	MMSELE	GRR	GIE	SSE1	SSE2	OLS
ORE1		1.0	1.0	1.17	1.06	1.02	2.95	1.0	2.39	.96	1.01
ORE2	1.0		1.0	1.17	1.06	1.01	2.94	1.0	2.39	.96	1.01
ORE3	1.0	1.0		1.16	1.05	1.01	2.94	1.0	2.38	.95	1.01
ORE4a	.86	.86	.86		.91	.87	2.53	.86	2.05	.82	.86
ORE4b	.95	.95	.95	1.10		.96	2.79	.95	2.26	.91	.95
MMSELE	.99	.99	.99	1.15	1.04		2.91	.99	2.36	.94	1.0
GRR	.34	.34	.34	.40	.36	.34		.34	.81	.33	.34
GIE	1.0	1.0	1.01	1.17	1.06	1.02	2.95		2.39	.96	1.01
SSE1	.42	.42	.42	.49	.44	.42	1.23	.42		.40	.42
SSE2	1.04	1.05	1.05	1.22	1.10	1.06	3.08	1.04	2.50		1.05
OLS	.99	.99	1.0	1.16	1.05	1.01	2.92	.99	2.37	.95	

Table 6 Ratio of Aggregated TMSE of Estimators to Aggregated TMSE of OLS

β_S		β_L		β_L and β_S	
Rank	Ratio	Rank	Ratio	Rank	Ratio
GRR	.17	SSE2	.95	ORE1	.68
SSE1	.23	ORE1	.99	ORE3	.80
ORE1	.36	ORE2	.99	GIE	.91
ORE3	.60	GIE	.99	ORE2	.93
ORE4a	.76	ORE3	1.00	ORE4b	.93
ORE4b	.81	MMSELE	1.01	MMSELE	.93
GIE	.82	ORE4b	1.06	ORE4a	.96
MMSELE	.85	ORE4a	1.16	SSE2	1.0
ORE2	.87	SSE1	2.37	SSE1	1.3
SSE2	1.04	GRR	2.92	GRR	1.55

Table 7 Overall Rank of Estimators Based on Average Ratio of Aggregated TMSE of the Particular Estimator to all Estimators

β_S		β_L		β_S and β_L	
Rank	Average Ratio	Rank	Average Ratio	Rank	Average Ratio
GRR	.29	SSE2	.82	ORE1	.80
SSE1	.42	ORE1	.86	ORE3	1.06
ORE1	.73	GIE	.86	SSE1	1.31
ORE3	1.25	ORE2	.86	GIE	1.32
ORE4a	1.63	ORE3	.86	ORE4a	1.33
ORE4b	1.75	OLS	.87	ORE4b	1.33
GIE	1.77	MMSELE	.87	MMSELE	1.35
MMSELE	1.83	ORE4b	.91	ORE2	1.37
ORE2	1.88	ORE4a	1.02	GRR	1.50
OLS	2.17	SSE1	2.19	OLS	1.52
SSE2	2.27	GRR	2.72	SSE2	1.55

β_S is the true coefficient vector, both GRR and SSE1 perform extremely well (and better than ORE1). For β_L however, these two estimators do not perform well. Since β is not known in practice, it would be unwise to use GRR or SSE1 due to their sensitivity to the true value of β. Rather, it is recommended that ORE1 or 3 be used as, on average, they provide a substantial improvement over OLS for β_S (and for extreme multicollinearity). For β_L ORE1 and 3 do not perform significantly worse than OLS.

The general conclusion drawn on the basis of the overall performance of the estimators is that one of the alternative estimators should be used in place of OLS when the data exhibits multicollinearity. In particular the ridge trace estimators, ORE1, or ORE3 can be used to ensure a good performance in all cases. If a particular value of σ_ε, μ or β is suspected, the estimator which provides the best performance for these values can be used. Tables 2, 4 and 5 can be used to make the decision as to which estimator to use.

CHAPTER 7

SUGGESTION FOR FUTURE RESEARCH

This thesis has concluded that there are several estimators (eg. various versions of ridge regression) which, in most cases, perform better than OLS in the presence of multi-collinearity on the basis of the TMSE criterion. However, the relative performance of these estimators depended most critically on the direction of the true coefficient vector, β. That is, most estimators performed particularly well relative to OLS when β was in the direction of the eigenvector corresponding to the smallest eigenvalue of the X'X matrix. However, when β was in the direction of the eigenvector corresponding to the largest eigenvalue of X'X, most of the estimators performed worse than OLS. The sensitivity of the estimators' performance to the direction of β casts doubt on the use of ridge regression and other alternatives in practical situations as the direction of β is not known.

Given this situation, the use of ridge regression or some other alternative to OLS in practice is predicated on being able to comment on the direction of β based on sample data. Consequently, one important direction for future research is to develop some means of obtaining a reasonable estimate of the direction of β based on the single sample on hand. The following discussion suggests one possible method of doing this.

Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_N$ be the ordered eigenvalues of X'X. Let $V_1, \ldots, V_N$ be the eigenvectors corresponding to $\lambda_1, \ldots, \lambda_N$ respectively. Then $V_i' \beta\mu$ (where $\beta\mu = \frac{\beta}{\sqrt{\beta'\beta}}$ is the unit vector in the direction of β) is the cosine of the angle between V_i and $\beta\mu$ (and hence β). It is easily shown that $V_i'\hat{\beta}\mu$ is the best linear unbiased estimator of $V_i'\beta\mu$ (where $\hat{\beta}\mu = \frac{\hat{\beta}}{\sqrt{\hat{\beta}'\hat{\beta}}}$ and $\hat{\beta}$ is the OLS estimate of β). Furthermore, $\text{var}(V_i'\hat{\beta}\mu) = \frac{\sigma_\varepsilon^2}{\lambda_i}$. This result shows that $V_i'\hat{\beta}\mu$ will be a precise estimate (in terms of having a small variance) of the direction of β relative to the V_i corresponding to large eigenvalues; and an imprecise estimate of the direction of β relative to the V_i corresponding to small eigenvalues. Thus the spectrum of eigenvalues can be used to reasonably comment on the direction of β relative to some eigenvectors. In most practical situations it is expected that there will be at least one "large" eigenvalue (ie. the degree of multicollinearity would unlikely be so severe as to reduce all eigenvalues to 0).

Suppose $\lambda_j, \ldots, \lambda_N$, $j \leq N$ are the large eigenvalues (ie. significantly different from 0) of X'X for a given X. Then $V_i'\hat{\beta}\mu$, $j \leq i \leq N$, will be precise estimates of $V_i'\beta\mu$. If $V_i'\hat{\beta}\mu$, for some i, is large (ie. close to 1) then β is in the general direction of V_i. In this case, the results of the last chapter indicate that OLS estimation is relatively precise and could then be used to estimate β. If, on the other hand, $V_i'\hat{\beta}\mu$ is small (close to zero) for all i, β will be in the direction of some V_ℓ $(1 \leq \ell \leq j)$ corresponding to a small eigenvalue, in which case our results have shown that considerable improvement in TMSE can be achieved by using one of the alternatives to OLS.

As an example, suppose the true coefficient vector was β_L and $\mu=0.9$, $\sigma_\varepsilon=0.1$. Now $\beta_L=V_L$, where V_L is the eigenvector corresponding to the largest eigenvalue of X'X (see Table A2 for the eigenvalue spectrum of X'X corresponding to $\mu=0.9$). In this case, $V_L'\hat{\beta}\mu=.99884$ suggesting that β is in the direction of V_L and consequently OLS should be used to estimate β. On the other hand, if β_S were the true coefficient vector $\beta_S=V_S$, where V_S is the eigenvector corresponding to the smallest eigenvalue of X'X, then $V_L'\hat{\beta}\mu=.00712$ implying that β is in the direction of V_S and consequently that some alternative to OLS be used to improve TMSE. Table 8 below provides the values of $V_L'\hat{\beta}_L$ and $V_L'\hat{\beta}_S$ (where $\hat{\beta}_L$ and $\hat{\beta}_S$ are the unit OLS estimates of β_L and β_S respectively) for all μ and σ_ε.

Table 8. Values of $V_L'\hat{\beta}_L$ and $V_L'\hat{\beta}_S$

μ	0.6			0.8			0.9			0.95		
σ_ε	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0	0.1	0.5	1.0
$V_L'\hat{\beta}_L$	1.00	0.95	0.83	1.00	0.91	0.75	0.99	0.85	0.63	0.99	0.76	0.51
$V_L'\hat{\beta}_S$	-.003	-.005	-.004	.007	.009	.009	.007	.008	.008	.006	.006	.006

APPENDIX 1

Throughout our discussion of the model $y=X\beta+\varepsilon$ we have been assuming that y and the variables of X have been standardized so that X'X is in the form of a correlation matrix and X'y is the vector of correlations of the dependent variable with each explanatory variable. We now discuss the standardization procedure as well as the transformation used to convert the standardized coefficient estimates to the non-standardized coefficient estimates of β in the model $y=X\beta+\varepsilon$ where y and X are in deviation units.

Let X be the T×N matrix of observations, x_{tn}, on the explanatory variables expressed in deviation units, that is $x_{tn}=X_{tn}-\bar{X}_n$ where X_{tn} is the t-th observation on the nth variable in original units and $\bar{X}_n$ is the mean of the t observations on the nth variable. Let S be the N×N diagonal matrix whose nth diagonal element is

$$\frac{1}{\sqrt{\sum_{t=1}^{T} x_{tn}^2}} .$$

Then $X_s=XS$ is the matrix of standardized variables and $X_s'X_s=S'X'XS$ is the correlation matrix.

Let

$$w = \frac{1}{\sqrt{\sum_{t=1}^{T} y_t^2}}$$

and $y_s=wy$

then $X_s'Y_s$ is the vector of correlations of the dependent variable with each explanatory variable.

If $y=X\beta+\varepsilon$ is the model with the variables in the original units we can write it as

$$w^{-1}(wy) = (XS)S^{-1}\beta+\varepsilon$$

$$w^{-1}y_s = X_s\beta_s+\varepsilon$$

where the variables are now standardized. Thus the coefficient estimates obtained will be estimates of $\beta_S=S^{-1}\beta$.

The estimated coefficients in the non-standardized model describe what happens at the origin which is usually far from the data. The role of the standardization is that the standardized coefficient estimates describe the effects of the variables in the region of the data. Also the estimated coefficients are easy to interpret; $\hat{\beta}_i$ is the predicted change in the dependent variable when x_i is changed by one standard deviation.

Probably the most significant role of the standardization is that it puts the estimators on a common scale. This scaling also applies to the eigenvalues and eigenvectors corresponding to the $X_s'X_s$ matrix and as we have seen, eigenanalysis of the $X_s'X_s$ matrix is crucial for determining the admissable values of k, δ and r.

In this study the standardized model is used to obtain admissable values of k, δ and r. The coefficient estimates so obtained are not transformed into original units since as will be shown the ranking of the estimators (in terms of TMSE) is invariant under the transformation.

We have seen that the standardized coefficient estimates are estimates of $S^{-1}\beta$. If b and b* are the non-standardized and standardized coefficient estimates (of β and $S^{-1}\beta$) respectively then

$$b = W^{-1}Sb^*$$

is the transformation which changes the standardized coefficient estimates into the non-standardized estimates.

Suppose b_1 and b_2 are any two estimates of β, then $b_i^*=S^{-1}b_i$, i=1,2 is an estimate of $S^{-1}\beta=\beta^*$. Then

$$\begin{aligned} TMSE(b_i^*) &= E(b_i^*-\beta^*)'(b_i^*-\beta^*) \\ &= E(b_i-\beta)'S^{-2}(b_i-\beta),\ i=1,2 \end{aligned}$$

where S^{-2} is non-negative definite (n.n.d.)

Now Theobald (1973), page 104) has shown that if

$$M_i = E(b_i-\beta)(b_i-\beta)',\ i=1,2$$

and $m_i=E(b_i-\beta)'B(b_i-\beta)$ where B is any n.n.d. matrix then the following are equivalent

(a) M_1-M_2 is n.n.d.

(b) $m_1-m_2\geq 0$ for any n.n.d. B.

The implication of this result for us is that if $TMSE(b_1^*)\geq TMSE(b_2^*)$ then M_1-M_2 is n.n.d.. Now M_1-M_2 is n.n.d.

if and only if $tr(M_1-M_2) \geq 0$, that is, if an only if $TMSE(b_1) \geq TMSE(b_2)$.

This shows that the TMSE is invariant under the transformation from standardized to non-standardized coefficient estimates.

APPENDIX 2

This section provides the formulae for the standard errors of the estimators considered. For all the estimators considered, the standard errors are estimates of their respective variance - covariance matrices.

1. The variance-covariance matrix for the OLS estimate is given by

$$\Sigma_{\hat{\beta}\hat{\beta}} = \sigma_{\varepsilon}^{2}(X'X)^{-1} .$$

It is estimated by

$$S_{\hat{\beta}\hat{\beta}} = \hat{\sigma}_{\varepsilon}^{2}(X'X)^{-1}$$

where

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{(y-X\hat{\beta})'(y-X\hat{\beta})}{T-N} .$$

The standard error of the ith OLS coefficient estimate is given by

$$S_{\hat{\beta}i} = \hat{\sigma}_{\varepsilon}\sqrt{C_{ii}}$$

where C_{ii} is the ith diagonal element of $(X'X)^{-1}$.

2. Hoerl and Kennard (1970, page 60) have shown that

$$\Sigma_{\hat{\beta}(k)\hat{\beta}(k)} = \sigma_{\varepsilon}^{2}Z(X'X)^{-1}Z$$

where

$$Z = [I+k(X'X)^{-1}]^{-1} .$$

Then we let

$$S_{\hat{\beta}(k)\hat{\beta}(k)} = \hat{\sigma}_{\varepsilon}^{2} Z(X'X)^{-1} Z$$

where

$$\hat{\sigma}_{\varepsilon}^{2} = \frac{[y-X\hat{\beta}(k)]'[y-X\hat{\beta}(k)]}{T-N} .$$

Then the standard error of the ith ORE is given by

$$S_{\hat{\beta}_i(k)} = \hat{\sigma}_{\varepsilon}\sqrt{C_{ii}}$$

where

C_{ii} = ith diagonal element of $Z(X'X)^{-1}Z$.

3. For GRR,

$$\sum_{\hat{\alpha}(K),\hat{\alpha}(K)} = \sigma_{\varepsilon}^{2} W[(X^*)'X^*]^{-1} W$$

where

$$W = [I+K[(X^*)'X^*]^{-1}]^{-1} .$$

Then

$$\sum_{\hat{\beta}(K)\hat{\beta}(K)} = P\sum_{\hat{\alpha}(K)\hat{\alpha}(K)} P'$$

$$= \sigma_{\varepsilon}^{2} PW[(X^*)'X^*]^{-1}(PW)' .$$

Hence we let

$$S_{\hat{\beta}_i(K)} = \hat{\sigma}_\varepsilon \sqrt{C_{ii}}$$

where

$$\hat{\sigma}_\delta^2 = \frac{[y-X\hat{\beta}(K)]'[y-X\hat{\beta}(K)]}{T-N}$$

and C_{ii} is the ith diagonal element of

$$PW[(X^*)'X^*]^{-1}(PW)'$$

4. For the MMSELE

$$b^* = \left(\frac{\hat{\beta}'X'y}{\hat{\sigma}_\varepsilon^2+\hat{\beta}X'X\hat{\beta}}\right)\hat{\beta}$$

$$= a\hat{\beta}$$

we have

$$\sum_{b^*b^*} = a^2\sigma_\varepsilon^2(X'X)^{-1}$$

which is estimated by

$$S_{b^*b^*} = a^2\hat{\sigma}_\varepsilon^2(X'X)^{-1}$$

where

$$\hat{\sigma}_\varepsilon^2 = \frac{(y-Xb^*)'(y-Xb^*)}{T-N}$$

then the standard error of the ith MMSELE is given by

$$S_{b_i^*} = a\hat{\sigma}_\varepsilon\sqrt{C_{ii}}$$

where C_{ii} is the ith diagonal element of $(X'X)^{-1}$.

(a) For SSE_1 we have

$$SSE_1 = \delta\hat{\beta}\hat{\beta}(I+\delta\hat{\beta}\hat{\beta}')^{-1}\hat{\beta}$$
$$= \delta A\hat{\beta} \quad .$$

Then

$$\sum_{SSE_1,SSE_1} = \delta^2\sigma_\varepsilon^2 A'(X'X)^{-1}A$$

and

$$S_{SSE_1,SSE_1} = \delta^2\hat{\sigma}_\varepsilon^2 A'(X'X)^{-1}A$$

where

$$\hat{\sigma}_\varepsilon^2 = \frac{[y-X(SSE_1)]'[y-X(SSE_1)]}{T-N} \quad .$$

So

$$S_{(SSE_1)_i} = \delta\hat{\sigma}_\varepsilon\sqrt{C_{ii}}$$

is the standard error of the ith SSE_1 where C_{ii} is the ith diagonal element of $A'(X'X)^{-1}A$.

(b) For SSE_2 we have in an analagous fashion to SSE_1

$$S_{(SSE_2)_i} = a\hat{\sigma}_\varepsilon\sqrt{C_{ii}} \quad .$$

where

$$a = 1+\gamma s^2(\hat{\beta}'\hat{\beta})^{-1}$$

and

$$\hat{\sigma}_\varepsilon^2 = \frac{[y-X(SSE_2)]'[y-X(SSE_2)]}{T-N}$$

and C_{ii} is the ith diagonal element of $(X'X)^{-1}$.

For the GIE we have

$$\Sigma_{\hat{\beta}_r^+\hat{\beta}_r^+} = \sigma_\varepsilon^{\,2} S_r \Omega^{-1} S_r'$$

$$S_{\hat{\beta}_r^+\hat{\beta}_r^+} = \hat{\sigma}_\varepsilon^{\,2} S_r \Omega^{-1} S_r' = \hat{\sigma}_\varepsilon^{\,2} A_r^+$$

where

$$\hat{\sigma}_\varepsilon^{\,2} = \frac{(y-X\hat{\beta}_r^+)'(y-X\hat{\beta}_r^+)}{T-N} .$$

So

$$S_{(\hat{\beta}_r^+)} = \hat{\sigma}_\varepsilon \sqrt{C_{ii}}$$

where C_{ii}=ith diagonal element of $S_r\Omega^{-1}S_r'=A_r^+$.

APPENDIX III

The following table contains the X'X matrices for μ=.6,.8,.9 and .95. The elements of each matrix are the various correlations between pairs of independent variables.

Table A1. Correlation Matrices by Degree of Multicollinearity

μ=.6

	1	2	3	4	5	6
1	1.00					
2	.45	1.00				
3	.20	.35	1.00			
4	.44	.40	.33	1.00		
5	.48	.70	.28	.54	1.00	
6	.53	.46	.37	.51	.50	1.00

μ=.8

	1	2	3	4	5	6
1	1.00					
2	.70	1.00				
3	.55	.63	1.00			
4	.71	.67	.62	1.00		
5	.72	.83	.58	.75	1.00	
6	.75	.70	.64	.74	.72	1.00

Table A1 (continued)

μ=.9

	1	2	3	4	5	6
1	1.00					
2	.84	1.00				
3	.76	.80	1.00			
4	.85	.82	.80	1.00		
5	.85	.91	.77	.87	1.00	
6	.87	.84	.80	.86	.85	1.00

μ=.95

	1	2	3	4	5	6
1	1.00					
2	.92	1.00				
3	.88	.89	1.00			
4	.92	.91	.89	1.00		
5	.92	.95	.88	.93	1.00	
6	.93	.91	.90	.93	.92	1.00

Table A2 Eigenvalues

μ=.6	μ=.8	μ=.9	μ=.95
.26087	.14715	.08094	.04349
.43315	.23497	.12441	.06492
.56413	.29679	.15498	.08039
.68889	.38418	.20948	.11189
.83011	.49357	.26831	.13945
3.22283	4.44332	5.16186	5.55985

Table A3 True Coefficient Vectors

	μ=.6	μ=.8	μ=.9	μ=.95
	-.03494	.05024	.06092	.06935
	.62725	-.61588	-.60552	-.59697
	-.14624	.14317	.14733	.15447
β_S	.26651	-.28876	-.31179	-.33106
	-.71616	.71713	.71463	.71085
	.00521	-.00368	-.00439	-.00617
	-.39622	-.40819	-.40929	-.40903
	-.43652	-.41746	-.41145	-.40934
	-.29873	-.36435	-.38901	-.39953
β_L	-.41037	-.41268	-.41092	-.40966
	-.45784	-.42549	-.41536	-.41139
	-.43034	-.41833	-.41289	-.41042

Table A4 Average k Value and Average δ Value

μ	σ	ORE1min	ORE1	ORE1max	ORE2	ORE3	ORE4a/b	SSE1
	.1	.045	.107	.169	0.0	.002	.001	.21
.6	.5	.042	.104	.166	.010	.058	.016	.51
	1.0	.025	.082	.144	.038	.230	.091	.64
	.1	.101	.163	.225	0.0	.002	.001	.10
.8	.5	.069	.130	.192	.008	.050	.015	.28
	1.0	.030	.090	.152	.028	.165	.081	.42
	.1	.056	.118	.180	0.0	.002	.001	.04
.9	.5	.032	.094	.156	.008	.049	.021	.20
	1.0	.012	.069	.131	.022	.134	.060	.30
	.1	.078	.140	.202	0.0	.002	.001	.02
.95	.5	.044	.105	.167	.008	.046	.024	.17
	1.0	.033	.094	.156	.016	.096	.034	.27
	.1	0.0	.026	.088	0.0	.003	.007	1.61
.6	.5	.019	.067	.129	.990	.057	.135	1.47
	1.0	.053	.113	.175	.031	.185	.459	1.34
	.1	0.0	.021	.083	0.0	.002	.015	1.29
.8	.5	.010	.059	.121	.009	.053	.253	1.12
	1.0	.100	.158	.220	.026	.156	.538	.95
	.1	0.0	.015	.077	0.0	.003	.028	1.32
.9	.5	.018	.074	.136	.008	.048	.426	2.43
	1.0	.107	.168	.230	.021	.124	.335	1.63
	.1	0.0	.033	.100	0.0	.003	.004	1.96
.95	.5	.041	.102	.164	.007	.042	.512	1.67
	1.0	.106	.168	.230	.015	.091	.163	1.52

Table A5 Average Q Value and Number of Negative Q Values Out of 50 Samples

	μ	σ	Average Q Value	Number of Negative Q Values
β_S	.6	.1	3.715	0
		.5	2.135	0
		1.0	.859	3
	.8	.1	6.455	0
		.5	3.038	0
		1.0	1.164	7
	.9	.1	11.242	0
		.5	3.704	0
		1.0	1.252	15
	.95	.1	19.318	0
		.5	4.204	6
		1.0	1.330	22
β_L	.6	.1	.311	0
		.5	.293	0
		1.0	.246	0
	.8	.1	.225	0
		.5	.216	0
		1.0	.190	2
	.9	.1	.194	0
		.5	.187	0
		1.0	.167	13
	.95	.1	.180	0
		.5	.175	1
		1.0	.158	21

Table A6 Average Squared Length

Estimator	μ=.6			μ=.8			μ=.9			μ=.95		
β_S	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
OLS	3.730	2.371	1.275	6.500	3.563	1.953	11.393	4.929	2.829	19.831	6.907	4.451
ORElmin	2.755	1.806	1.087	2.453	1.818	1.444	4.305	2.892	2.234	2.780	2.409	1.865
OREl	1.898	1.264	.794	1.534	1.110	.870	1.977	1.254	1.036	1.187	.853	.719
ORElmax	1.388	.936	.598	1.054	.756	.589	1.138	.718	.598	.660	.455	.395
ORE2	3.719	2.221	1.069	6.468	3.244	1.589	11.290	4.265	2.207	19.489	5.657	3.377
ORE3	3.668	1.674	.593	6.314	2.214	.838	10.795	2.535	1.100	17.915	3.011	1.619
ORE4a	3.714	2.135	.867	6.454	3.039	1.185	11.244	3.702	1.448	19.307	4.260	2.002
ORE4b	3.714	2.135	.896	6.454	3.039	1.308	11.244	3.702	1.854	19.307	4.628	3.007
MMSELE	3.719	2.230	1.092	6.468	3.270	1.637	11.291	4.339	2.295	19.495	5.822	3.533
GRR	.120	.051	.022	.382	.171	.116	.806	.284	.193	2.632	.628	.393
GIE	3.730	2.371	1.275	6.500	3.563	1.953	.117	.888	1.140	.384	1.979	2.282
SSEl	.720	.698	.275	.958	.878	.446	1.187	1.123	.619	1.960	1.444	.998
SSE2	3.739	2.525	1.563	6.516	3.757	2.255	11.423	5.175	3.153	19.884	7.193	4.787

Table A6 continued

Estimator	$\mu=.6$			$\mu=.8$			$\mu=.9$			$\mu=.95$		
β_L	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$
OLS	.312	.326	.359	.227	.259	.341	.197	.257	.417	.185	.300	.612
ORE1min	.312	.318	.322	.227	.252	.253	.197	.236	.235	.195	.218	.226
ORE1	.307	.304	.298	.225	.236	.230	.195	.209	.206	.180	.188	.189
ORE1max	.295	.289	.278	.218	.223	.214	.189	.194	.189	.175	.176	.172
ORE2	.312	.323	.342	.227	.255	.318	.197	.250	.372	.185	.283	.515
ORE3	.312	.309	.287	.227	.242	.254	.196	.226	.268	.185	.237	.328
ORE4a	.311	.293	.251	.225	.216	.210	.194	.191	.202	.180	.193	.252
ORE4b	.311	.293	.251	.225	.216	.218	.194	.191	.270	.180	.197	.397
MMSELE	.312	.324	.350	.227	.257	.335	.197	.256	.411	.185	.299	.603
GRR	.022	.017	.013	.011	.010	.013	.011	.011	.018	.015	.019	.037
GIE	.312	.326	.359	.227	.259	.341	.196	.238	.350	.184	.267	.494
SSE1	.035	.035	.042	.012	.014	.024	.009	.040	.077	.013	.037	.166
SSE2	.313	.347	.435	.228	.274	.399	.197	.270	.469	.186	.313	.660

Table A7 Total Bias

Estimator	μ=.6			μ=.8			μ=.9			μ=.95		
β_S	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
ORE1	.016	-.007	-.018	-.046	-.042	-.035	-.059	-.043	-.033	-.068	-.041	-.029
ORE2	.017	-.007	-.019	-.045	-.042	-.035	-.059	-.044	-.033	-.068	-.041	-.030
ORE3	.017	-.007	-.018	-.045	-.042	-.036	-.059	-.044	-.034	-.068	-.042	-.030
ORE4a	.017	-.007	-.019	-.045	-.042	-.046	-.059	-.044	-.033	-.068	-.051	-.027
ORE4b	.017	-.007	-.017	-.045	-.042	-.035	-.059	-.044	-.031	-.068	-.041	-.029
MMSELE	.017	-.007	-.020	-.045	-.041	-.035	-.058	-.042	-.034	-.068	-.040	-.031
GRR	-.049	-.025	.008	.111	.003	-.016	.182	.006	-.012	-.068	-.004	-.029
GIE	.017	-.007	-.019	-.045	-.042	-.035	-.063	-.046	-.035	.337	-.042	-.031
SSE1	.007	-.002	-.007	-.019	-.023	-.019	-.020	-.024	-.017	-.070	-.020	-.014
SSE2	.017	-.007	-.018	-.045	-.042	-.035	-.059	-.044	-.033	-.027	-.042	-.030
OLS	.017	-.007	-.019	-.045	-.042	-.035	-.059	-.044	-.033	-.068	-.041	-.030

Table A7 - continued

Estimator	$\mu=.6$			$\mu=.8$			$\mu=.9$			$\mu=.95$		
β_L	$\sigma=.01$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$
ORE1	1.098	1.147	1.261	1.298	1.328	1.414	1.379	1.407	1.479	1.420	1.448	1.509
ORE2	1.083	1.123	1.231	1.289	1.314	1.384	1.373	1.394	1.451	1.412	1.431	1.483
ORE3	1.084	1.142	1.284	1.289	1.325	1.414	1.373	1.402	1.470	1.412	1.437	1.496
ORE4a	1.085	1.171	1.361	1.293	1.371	1.529	1.378	1.468	1.762	1.420	1.531	1.913
ORE4b	1.085	1.171	1.361	1.293	1.371	1.487	1.378	1.468	1.503	1.420	1.510	1.506
MMSELE	1.083	1.124	1.345	1.288	1.315	1.388	1.373	1.394	1.455	1.412	1.432	1.489
GRR	2.240	2.310	2.367	2.278	2.337	2.388	2.291	2.362	2.418	2.314	2.361	2.437
GIE	1.083	1.120	1.220	1.289	1.312	1.378	1.373	1.392	1.448	1.412	1.430	1.481
SSE1	1.981	2.008	2.042	2.185	2.193	2.192	2.229	2.048	2.063	2.174	2.118	2.017
SSE2	1.081	1.076	1.090	1.287	1.277	1.282	1.371	1.362	1.375	1.410	1.405	1.429
OLS	1.083	1.120	1.220	1.289	1.312	1.378	1.373	1.392	1.447	1.412	1.430	1.481

Table A8 Total Variance

Estimator	$\mu=.6$			$\mu=.8$			$\mu=.9$			$\mu=.95$		
β_S	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$
ORE1	.017	.144	.284	.045	.217	.404	.083	.460	.730	.095	.586	.739
ORE2	.011	.194	.358	.033	.406	.631	.111	.977	1.233	.382	2.169	2.374
ORE3	.011	.183	.261	.033	.371	.433	.110	.793	.734	.378	1.497	1.271
ORE4a	.011	.194	.346	.033	.406	.601	.111	.970	1.007	.381	2.070	1.628
ORE4b	.011	.194	.328	.033	.406	.586	.111	.970	1.097	.381	2.015	2.165
MMSELE	.011	.189	.343	.033	.392	.604	.110	.935	1.207	.379	2.084	2.382
GRR	.003	.028	.021	.009	.107	.095	.029	.209	.65	.092	.467	.343
GIE	.011	.197	.384	.033	.414	.691	.110	.876	1.124	.375	1.951	2.250
SSE1	.012	.063	.096	.016	.118	.183	.037	.256	.340	.079	.500	.657
SSE2	.011	.206	.454	.033	.429	.781	.111	1.058	1.587	.384	2.454	3.135
OLS	.011	.197	.384	.033	.414	.691	.111	1.021	1.438	.383	2.375	2.926

Table A8 continued

Estimator	μ=.6			μ=.8			μ=.9			μ=.95		
β_L	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
ORE1	.001	.023	.065	.001	.027	.053	.002	.031	.052	.003	.026	.047
ORE2	.001	.031	.096	.002	.039	.126	.003	.062	.201	.005	.108	.353
ORE3	.001	.026	.063	.002	.030	.074	.003	.042	.106	.005	.064	.172
ORE4a	.001	.022	.056	.002	.022	.067	.002	.029	.120	.003	.051	.198
ORE4b	.001	.022	.056	.002	.022	.062	.002	.029	.116	.003	.048	.242
MMSELE	.001	.032	.105	.002	.042	.145	.003	.068	.241	.005	.124	.440
GRR	.0003	.006	.010	.0004	.003	.010	.001	.005	.016	.001	.011	.036
GIE	.001	.032	.107	.002	.042	.147	.002	.051	.181	.004	.093	.333
SSE1	.0005	.005	.016	.0002	.003	.013	.0004	.013	.051	.001	.018	.132
SSE2	.001	.034	.127	.002	.044	.168	.003	.072	.271	.005	.129	.477
OLS	.001	.032	.107	.002	.042	.147	.003	.068	.245	.005	.124	.446

Table A9 TMSE

Estimator	$\mu=.6$			$\mu=.8$			$\mu=.9$			$\mu=.95$		
β_S	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$	$\sigma=.1$	$\sigma=.5$	$\sigma=1$
ORE1	.017	.144	.284	.047	.219	.405	.087	.462	.731	.099	.587	.740
ORE2	.011	.194	.358	.035	.408	.632	.114	.979	1.234	.387	2.171	2.375
ORE3	.011	.183	.261	.035	.373	.435	.114	.795	.735	.383	1.499	1.272
ORE4a	.011	.194	.347	.035	.407	.603	.114	.972	1.008	.386	2.072	1.628
ORE4b	.011	.194	.328	.035	.407	.587	.114	.972	1.098	.386	2.017	2.166
MMSELE	.011	.189	.343	.035	.393	.605	.114	.936	1.208	.384	2.085	2.383
GRR	.006	.029	.021	.021	.107	.095	.063	.209	.165	.205	.467	.344
GIE	.011	.197	.384	.035	.416	.693	.114	.878	1.125	.380	1.952	2.251
SSE1	.012	.063	.096	.017	.119	.184	.037	.257	.340	.079	.499	.658
SSE2	.012	.206	.454	.035	.431	.782	.115	1.060	1.588	.389	2.456	3.136
OLS	.011	.197	.384	.035	.146	.692	.115	1.023	1.439	.388	2.377	2.927

Table A9 continued.

Estimator	μ=.6			μ=.8			μ=.9			μ=.95		
β_L	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
ORE1	1.207	1.339	1.655	1.685	1.789	2.053	1.903	2.011	2.238	2.018	2.123	2.325
ORE2	1.174	1.292	1.612	1.663	1.767	2.043	1.887	2.004	2.307	1.998	2.155	2.553
ORE3	1.176	1.329	1.712	1.664	1.786	2.072	1.888	2.007	2.268	1.999	2.130	2.411
ORE4a	1.180	1.393	1.908	1.672	1.901	2.404	1.902	2.184	3.223	2.019	2.394	3.856
ORE4a	1.180	1.393	1.908	1.672	1.901	2.272	1.902	2.184	2.376	2.019	2.330	2.508
MMSELE	1.174	1.294	1.629	1.663	1.770	2.070	1.887	2.012	2.358	1.998	2.173	2.654
GRR	5.019	5.340	5.613	5.192	5.466	5.713	5.251	5.584	5.863	5.353	5.584	5.973
GIE	1.174	1.285	1.596	1.663	1.763	2.047	1.886	1.988	2.275	1.997	2.136	2.526
SSE1	3.924	4.037	4.187	4.774	4.811	4.819	4.970	4.207	4.306	4.726	4.502	4.198
SSE2	1.169	1.192	1.316	1.658	1.676	1.811	1.883	1.928	2.161	1.994	2.102	2.520
OLS	1.173	1.285	1.596	1.663	1.763	2.047	1.887	2.006	2.339	1.998	2.168	2.639

Table A10 $\frac{\sigma_\varepsilon^2}{\alpha^2_{max}}$

	μ=.6			μ=.8			μ=.9			μ=.95		
	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
β_S	.027	.677	2.707	.026	.655	2.621	.026	.655	2.619	.020	.509	2.036
β_L	.018	.447	1.787	.023	.563	2.254	.0120	.500	2.000	.015	.364	1.454

Table All Number of k Values in $\left(0, \frac{\sigma_\varepsilon^2}{\alpha^2_{max}}\right)$

Estimator	μ=.6			μ=.8			μ=.9			μ=.95		
β_S	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1	σ=.1	σ=.5	σ=1
ORE1min	10	50	50	1	50	50	1	50	50	0	50	50
ORE1	0	50	50	0	50	50	0	50	50	0	50	50
ORE1max	0	50	50	0	50	50	0	50	50	0	50	50
ORE2	21	50	50	30	50	50	36	50	50	0	50	50
ORE3	0	50	50	0	50	50	1	50	50	34	50	50
ORE4a,b	0	34	50	3	26	50	16	35	50	3	50	50
										27	43	50
β_L												
ORE1min	0	50	50	50	50	50	50	50	50	50	50	50
ORE1	2	50	50	36	50	50	35	50	50	13	50	50
ORE1max	2	50	50	0	50	50	0	50	50	0	50	50
ORE2	8	50	50	23	50	50	26	50	50	27	50	50
ORE3	0	50	50	0	50	50	0	50	50	1	50	50
ORE4a,b	0	27	50	3	26	50	14	35	50	25	43	50

BIBLIOGRAPHY

1. Cragg, J.G. "On the Relative Small Sample Properties of Several Structural Equation Estimators", Econometrics, 35, 1, 1967, 89-110.

2. Deegan, J. "Optimal k in Ordinary Ridge Regression", Unpublished, 1975.

3. Farebrother, R.W. "The Minimum Mean Square Error Linear Estimator and Ridge Regression", Technometrics, 17, 1, 1975, 127-28.

4. Guilkey, D.K. and Murphy, J.L. "Directed Ridge Regression Techniques in Cases of Multicollinearity", Journal of the American Statistical Association, 70, 352, 1975, 769-775.

5. Hemmerle, W.J. "An Explicit Solution for Generalized Ridge Regression", Technometrics, 17, 3, 1975, 309-314.

6. Hoerl, A.E. and Kennard, R.W. "Ridge Regression: Biased Estimation for Nonorthogonal Problems", Technometrics, 12, 1970a, 55-67.

7. Hoerl, A.E. and Kennard, R.W. "Ridge Regression: Applications to Nonorthogonal Problems", Technometrics, 12, 1, 1970b, 69-82.

8. Hoerl, A.E. and Kennard, R.W. "Ridge Regression: Iterative Estimation of the Biasing Parameter", Communications in Statistics. Theory and Methods, A5(1), 1976, 77-88.

9. Hoerl, A.E., Kennard, R.W. and Baldwin, K.F. "Ridge Regression: Some Simulations", Communications in Statistics. Theory and Methods, 4, 1975, 105-123.

10. Hoerl, A.E. "Application of Ridge Analysis to Regression Problems", Chemical Engineering Progress, 58, 1962, 54-59.

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

12. Marquardt, D.W. "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation", Technometrics, 12, 3, 1970, 591-612.

13. Mason, R. and Brown, W.G. "Multicollinearity Problems and Ridge Regression in Sociological Models", Social Science Research, 4, 1975, 135-149.

14. Mayer, L.S. and Willke, T.A. "On Estimation in Linear Models", Technometrics, 15, 3, 1973, 497-508.

15. McDonald, G.C. and Schwing, R.C. "Instabilities of Regression Estimates Relating Air Pollution to Mortality", Technometrics, 15, 3, 1973.

16. Newhouse, J.P. and Oman, S.D. An Evalution of Ridge Estimators, Santa Monica, Rand, 1971.

17. Silvey, S.D. "Multicollinearity and Imprecise Estimation", Technometrics, 3, 1969, 539-552.

18. Theobald, C.M. "Generalizations of Mean Square Error Applied to Ridge Regression", Technometrics, 1974, 103-106.

19. Watson, D.E. and White, K.J. "Forecasting the Demand for Money Under Changing Term Structure of Interest Rates: An Application of Ridge Regression", Southern Economic Journal, 43, 2, 1976, 1096-1105.