AN ABSTRACT OF THE THESIS OF

Ali Ahmed Rahuma for the degree of Doctor of Philosophy

in Resource Economics presented on December 3, 1980

Title: SOME SMALL SAMPLE PROPERTIES OBTAINED BY

SIMULATION FOR A SQUARED ERROR LOSS ESTIMATOR

FOR RIDGE REGRESSION WITH APPLICATION TO AN

EMPIRICAL ECONOMIC MODEL

Abstract approved: 

Redacted for Privacy

William G. Brown

When multicollinearity is severe, the ordinary least square (OLS) estimate, although the best linear unbiased estimate (BLUE), is unreliable and imprecise because of greatly inflated variances. Under these circumstances, individual coefficients may be statistically insignificant and even take the wrong signs, even though the overall regression measured by $R^2$ may be highly significant.

Ridge regression was found to be more effective than OLS in the Monte Carlo experiments in this thesis under a wide range of regression situations. Four applied ridge estimators were compared in the Monte Carlo simulations. The ridge estimators outperformed OLS significantly for situations with relatively high multicollinearity and where the smallest true coefficients corresponded to the
smallest eigenvalues of \( (X'X) \).

The Monte Carlo experiments indicated that the Lawless-Wang ridge estimator tended to be more accurate and stable than the other ridge estimators, as measured by observed squared error loss. Performance of the other ridge estimators compared was more volatile and unreliable.

A serious limitation of the empirical use of biased linear estimators has been the lack of measures of reliability of these estimates. A technique for estimating observed squared error loss is proposed in this thesis, and this technique worked fairly well under certain conditions.

Monte Carlo results indicated that the proposed method for estimating squared error loss was most accurate for the Lawless-Wang ridge estimator for models with relatively lower multicollinearity. However, simulation results indicated that the proposed squared error loss estimator was not suitable for estimating squared error loss for the other ridge estimates (RIDGN, \( k_{HKB} \), and KCRR) because it is consistently underestimated the loss for these estimators. This underestimation could be dangerously misleading to the researcher depending upon the analysis of the sample data for measures of reliability, such as variance, t tests, etc. Results of this thesis also revealed that when the regression conditions are 'favorable', namely that the ratio of the estimated bias
squared to the total squared error loss is small (less than 10 percent), the Lawless-Wang ridge coefficient appeared to be distributed approximately normal. If so, the t* statistic defined by
\[ t^* = \frac{m_i \hat{\alpha}_i}{\sqrt{m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \hat{\sigma}_i^2}} \]
could be used to test hypotheses about the significance of the ridge coefficient \( \hat{\alpha}_i^* \). It should be emphasized, however, that the t* statistic should not be used if the ratio of estimated bias squared to total squared error loss is more than 10 percent. (When the ratio was more than 10 percent, the t* statistic departed from the normal distribution, due to the lack of independence between \( \hat{\alpha}_i^* \) and the estimated loss, \( L_i^2 (\hat{\alpha}_i^*) \).)

Results of the simulation also indicated that the use of the t* statistic is possible in empirical research by estimating
\[ \frac{B^2}{L^2} = \frac{(1-m_i)^2 \hat{\sigma}_i^2}{[m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \hat{\sigma}_i^2]} \]
For models with relatively low multicollinearity, the proposed method gave an excellent performance for all the models with a probability of making the correct decision of greater than 0.80. However, this probability was only greater than 0.50 for models with relatively high multicollinearity and \( \sigma^2 \) value greater than one.
Some Small Sample Properties Obtained by Simulation for a Squared Error Loss Estimator for Ridge Regression with Application to an Empirical Economic Model

by

Ali Ahmed Rahuma

A THESIS

submitted to

Oregon State University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

June, 1981
Date thesis is presented  December 3, 1980

Typed by Lesa Harris for Ali Ahmed Rahuma
ACKNOWLEDGMENT

I wish to express my most sincere gratitude and appreciation to Dr. William Brown, my major professor, for his contribution to my graduate study and to this thesis. Without Dr. Brown's time and patience, this research could not have been possible.

I also extend a special thanks to the members of my graduate committee, Dr. Donald Pierce, Dr. Ronald Oliviera, Dr. Doug Brodie, Dr. James Whittaker, and Dr. Darrel Hueth for serving on my graduate committee, and acknowledge the contribution of many other persons to this study.

Finally, I wish to thank my wife for her patience and encouragement during my graduate study.
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Multicollinearity</td>
<td>1</td>
</tr>
<tr>
<td>Definitions</td>
<td>2</td>
</tr>
<tr>
<td>Sources of Multicollinearity</td>
<td>2</td>
</tr>
<tr>
<td>Detection of Multicollinearity</td>
<td>5</td>
</tr>
<tr>
<td>Suggested Solutions to Multicollinearity</td>
<td>10</td>
</tr>
<tr>
<td>Theil-Goldberger Mixed Model</td>
<td>12</td>
</tr>
<tr>
<td>Biased Linear Estimators</td>
<td>14</td>
</tr>
<tr>
<td>The Loss Function</td>
<td>26</td>
</tr>
<tr>
<td>Proposed Method for estimating the loss function</td>
<td>30</td>
</tr>
<tr>
<td>II</td>
<td></td>
</tr>
<tr>
<td>THE MONTE CARLO EXPERIMENTS</td>
<td>33</td>
</tr>
<tr>
<td>Description</td>
<td>33</td>
</tr>
<tr>
<td>Estimators to be Studied</td>
<td>35</td>
</tr>
<tr>
<td>Models Considered</td>
<td>36</td>
</tr>
<tr>
<td>III</td>
<td></td>
</tr>
<tr>
<td>EXPERIMENTAL RESULTS VIA MONTE CARLO METHODS</td>
<td>40</td>
</tr>
<tr>
<td>The Distribution of the Lawless-Wang Ridge Estimator</td>
<td>79</td>
</tr>
<tr>
<td>IV</td>
<td></td>
</tr>
<tr>
<td>APPLICATION TO AN EMPIRICAL ECONOMIC MODEL</td>
<td>104</td>
</tr>
<tr>
<td>Nature of Economic Data</td>
<td>104</td>
</tr>
<tr>
<td>Estimating the Marginal Value Productivity of Irrigation Water</td>
<td>104</td>
</tr>
<tr>
<td>The Model</td>
<td>106</td>
</tr>
<tr>
<td>Multicollinearity in the Model</td>
<td>106</td>
</tr>
<tr>
<td>Ridge Regression Analysis</td>
<td>112</td>
</tr>
<tr>
<td>Regression Results from OLS</td>
<td>114</td>
</tr>
<tr>
<td>Ridge Regression Results</td>
<td>115</td>
</tr>
<tr>
<td>V</td>
<td></td>
</tr>
<tr>
<td>SUMMARY AND CONCLUSIONS</td>
<td>119</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>124</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>127</td>
</tr>
</tbody>
</table>
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Signal-to-noise ratio, $M_1$, $\tilde{M}<em>1$ and the coefficients of loss variation for the eight models $(a_1-a_8)$, with $\lambda = (.01,1,1.99)$ and $\sigma^2 = 4$. OLS vs. ridge estimators = $K</em>{LW}^r$, $K_{HKB}^r$, KGRR and ROP.</td>
<td>42</td>
</tr>
<tr>
<td>2</td>
<td>Signal-to-noise ratio, $M_1$, estimated $M_1$ and the coefficients of loss variation for the eight models $(a_1-a_8)$ with $\lambda = (.01,1,1.99)$, $\sigma^2 = 2$. OLS vs. ridge estimators = $K_{LW}^r$, $K_{HKB}^r$, KGRR and ROP.</td>
<td>52</td>
</tr>
<tr>
<td>3</td>
<td>Signal-to-noise ratio, $M_1$, estimated $M_1$, and the coefficients of loss variation for the eight models $(a_1-a_8)$ with $\lambda = (.01,1,1.99)$, $\sigma^2 = 1$. OLS vs. ridge estimators: $K_{LW}^r$, $K_{HKB}^r$, KGRR and ROP.</td>
<td>60</td>
</tr>
<tr>
<td>4</td>
<td>Signal-to-noise, $M_1$, $\tilde{M}<em>1$ and the coefficients of loss variations for the eight models $(a_1-a_8)$ with $\lambda = (.01,1,1.99)$, $\sigma^2 = 0.50$, OLS vs. ridge estimators $K</em>{LW}^r$, RIDGN, $K_{HKB}^r$, KGRR and ROP.</td>
<td>63</td>
</tr>
<tr>
<td>5</td>
<td>Signal-to-noise $M_1$, estimated $M_1$ and the coefficients of loss variation for the eight models $(a_1-a_8)$ with $\lambda = (.01,1,1.99)$, $\sigma^2 = 0.25$ OLS vs. ridge estimators: $K_{LW}^r$, RIDGN, $K_{HKB}^r$, KGRR and ROP.</td>
<td>66</td>
</tr>
<tr>
<td>6</td>
<td>Signal-to-noise, $M_1$, estimated $M_1$, and the coefficients of loss variation for the eight models $(a_1-a_8)$ with $\lambda = (0.1,1,1.9)$, $\sigma^2 = 4$. OLS vs. ridge estimators: $K_{LW}^r$, RIDGN, $K_{HKB}^r$, KGRR and ROP.</td>
<td>70</td>
</tr>
<tr>
<td>7</td>
<td>Signal-to-noise, $M_1$, estimated $M_1$, and the coefficients of loss variation for the eight models $(a_1-a_8)$ with $\lambda = (.1,1,1.9)$, $\sigma^2 = 2$ OLS vs. ridge estimators: $K_{LW}^r$, RIDGN, $K_{HKB}^r$, KGRR and ROP.</td>
<td>72</td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>Signal-to-noise, Ml, estimated Ml, and the coefficients of loss variation for the eight models ((a_1-a_8)) with (\lambda = (.1,1,1.9)), (\sigma^2 = 1) OLS vs. ridge estimators: (K_{LW}'), RIDGN, (K_{HKB}'), KGRR and ROP.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Signal-to-noise, Ml, estimated Ml, and the coefficients of loss variation for the eight models ((a_1-a_8)) with (\lambda = (.1,1,1.9)), (\sigma^2 = 0.50) OLS vs. ridge estimators: (K_{LW}'), RIDGN, (K_{HKB}'), KGRR and ROP.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>78</td>
<td></td>
</tr>
<tr>
<td>Signal-to-noise, Ml, estimated Ml, and the coefficients of loss variation for the eight models ((a_1-a_8)) with (\lambda = (.1,1,1.9)), (\sigma^2 = 0.25) OLS vs. ridge estimators: (K_{LW}'), RIDGN, (K_{HKB}'), KGRR and ROP.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>The cumulative distribution function of (\hat{\alpha}) for the eight models ((a_1-a_8)) with (\lambda = (.1,1,1.9)), (\sigma^2 = 4), ((\text{Bias})^2/\text{total loss ratio} ) (B^2/L^2) and the Standard t distribution with (N = \infty).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>87</td>
<td></td>
</tr>
<tr>
<td>The cumulative distribution function of (\hat{\alpha}) for the eight models ((a_1-a_8)) with (\lambda = (.1,1,1.9)), (\sigma^2 = 2), ((\text{Bias})^2/\text{total loss ratio} ) and the Standard t distribution with (N = \infty).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>The cumulative distribution function of (\hat{\alpha}) for the eight models ((a_1-a_8)) with (\sigma^2 = 1), (\lambda = (.1,1,1.9)), ((\text{Bias})^2/\text{total loss ratio} ) and the Standard t distribution with (N = \infty).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>92</td>
<td></td>
</tr>
<tr>
<td>The cumulative distribution function of (\hat{\alpha}) for the eight models ((a_1-a_8)), with (\sigma^2 = 0.50), (\lambda = (.1,1,1.9)), ((\text{Bias})^2/\text{total loss ratio} ) and the Standard t distribution with (N = \infty).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>94</td>
<td></td>
</tr>
<tr>
<td>The cumulative distribution function of (\hat{\alpha}) for the eight models ((a_1-a_8)), with (\lambda = (.1,1,1.9)), (\sigma^2 = 0.25), ((\text{Bias})^2/\text{total loss ratio} ) and the Standard t distribution with (N = \infty).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Table</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>16 Summary of tables Al-A5, showing the Kolmogorov-Smirnov calculated D statistics, bias squared to total squared loss, $B^2/L^2$, the critical D value and the decision about the stated hypothesis.</td>
<td>96</td>
<td></td>
</tr>
<tr>
<td>17 The true $B^2/L^2$, the estimated $B^2/L^2$, and the number of observations with estimated $B^2/L^2 &lt; 10$ percent for the eight models $(\alpha_1-\alpha_9)$ with $\lambda = (0.1,1,1.9)$ and $\sigma^2 = 4, 2, 1, 0.5, \text{ and } 0.25$.</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>18 The true $B^2/L^2$, the estimated $B^2/L^2$, and the number of observations with estimated $B^2/L^2 &lt; 10$ percent for the eight models $(\alpha_1-\alpha_9)$ with $\lambda = (0.1,1,1.9)$ and $\sigma^2 = 4, 2, 1, 0.5, \text{ and } 0.25$.</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>19 Estimated values for regression of county values of all farm products sold as a Cobb-Douglas function of inputs, OLS estimates for 25 Central Pacific counties, 1954.</td>
<td>107</td>
<td></td>
</tr>
<tr>
<td>20 Simple correlation coefficients and Eigenvalues for regression of county values of all farm products sold as a Cobb-Douglas function of inputs.</td>
<td>109</td>
<td></td>
</tr>
<tr>
<td>21 Estimates of the coefficients, square root of the squared error loss, $\hat{L}<em>1$, and the $B^2/L^2$ ratio for the OLS, the linearly restricted LS, and the ridge regression estimators: $K</em>{Lw}$, $K_{HKB}$, and KGR for regression of county values of all form products as a Cobb-Douglas function of inputs.</td>
<td>116</td>
<td></td>
</tr>
</tbody>
</table>
Economists, unlike physical scientists, don't have the opportunity to experiment, they instead obtain data by observing processes which cannot be repeated or controlled most of the time. As a result, when the data have undesirable properties, economists have no chance of designing experiments to remedy the problem. One common problem in economic data is the correlation of explanatory variables. Intercorrelation among explanatory variables in economic data is unavoidable since economic variables tend to move in the same direction over time and due to the use of aggregate data over time or over some geographical location (cities and counties). Consequently, at least some intercorrelation among economic variables is expected due to the nature of the economic data.

**Multicollinearity**

Multicollinearity is a sample or data problem in which one cannot single out or separate the effect of two or more explanatory variables in the model, because they tend to move together in sample data. It is the degree of interdependency between two or more theoretically exogeneous variables. Because of the nature of economic
models which measure the relationships among variables, multicollinearity is usually then an unavoidable problem. Some degree of collinearity is expected to appear in the estimation of most economic relationships.

Definitions

Farrar and Glauber (1967) defined multicollinearity as "a departure from orthogonality in an independent variable set." It is perhaps very important to separate two factors:

1. The nature of the problem: the lack of independence among explanatory variables in the regression model.

2. The effect and degree of multicollinearity. It is important for the researcher to find to what degree and in what locations variables in the model are affected by multicollinearity.

Sources of multicollinearity

There are many sources of multicollinearity discussed in the literature. Perhaps the best known sources are:

1. Data aggregation and averaging
2. Model specifications
3. Sampling procedures and techniques
4. Physical and non-physical limitations on the model
Data aggregation. Due to the bulkiness of the information and the difficulty and sometimes impossibility of handling each individual factor of information separately, economists and researchers very often depend in their analyses on aggregated data based on available sources, such as agricultural and/or commercial censuses. Aggregation will often result in loss of information and will create dependency among explanatory factors and hence multicollinearity.

Model specifications. A problem usually associated with overdefined models in cases where there are nearly as many variables as there are observations. A serious multicollinearity problem is more likely to occur in such researches, such as medical and industrial research.

Sampling techniques. When a sample is taken over only a subset of the experimental space, a loss of information will result and a problem of multicollinearity may occur.

Physical limitations on the model. Some limitations on the data, such as the presence of fixed proportionality or simply that some information is not available to the researcher.

It is very important to recognize the different sources of multicollinearity and its effect in order to take some corrective steps to partially remove its effect.

When estimating the parameters of a linear model in
the form:
\[ y = X\beta + u \]

\( y \) is \((n \times 1)\) vector of dependent variables; \( X \) is the \((n \times p)\) matrix of theoretically independent explanatory variables, \( \beta \) is a \((p \times 1)\) vector of coefficients, and \( u \) is \((n \times 1)\) vector of error terms where \( u \sim n(0, \sigma^2 I_n) \).

The OLS estimate of \( \beta \) is given by
\[ \hat{\beta} = (X'X)^{-1}X'y \]
and
\[ \text{var-cov } (\hat{\beta}) = \sigma^2 (X'X)^{-1}. \]

As far as the problem of multicollinearity is concerned, one can easily recognize two extreme cases:

1. An orthogonal case where there is no multicollinearity and the correlation matrix has one's in all the diagonal elements and zeros in all off diagonal elements; i.e.,
\[ r_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j. \end{cases} \]

2. Perfect multicollinearity is the most drastic case where explanatory variables are perfectly dependent on each other and the determinant of \((X'X)\) is zero; hence, the inverse of \((X'X)\) doesn't exist. The OLS estimate of \( \beta \) are not possible to obtain in this case, and the variance-covariance matrix of \( \hat{\beta} \) tends to \( \infty \) at the limit.

Between these two extreme cases there exist degrees of multicollinearity. Only when the degree of multi-
collinearity causes estimation and interpretation problems
does it become of some concern to the researcher.

Detection of multicollinearity

1. In a simple two-variable model the detection of multicollinearity is an easy task. A simple correlation
coefficient $r_{ij} = 0$ indicates no multicollinearity,
$r_{ij} = \pm 1$ indicates perfect collinearity, and $-1 < r_{ij} < 1.0$
measures the degree of association between variables $i$
and $j$ in the model. This simple interpretation is not
possible, however, when more than two variables are in
the model.

2. Another method widely practiced in research is
to regress $y$ on all $X$ (explanatory variables) to obtain
$R$ and then to regress $y$ on all $X$ except $X_j$ and obtain
$R_j$. $R$ will be very close to $R_j$ when multicollinearity is
severe. This method has been criticized on the ground
that it will reflect the importance of $X_j$ as a predictor
but not that $X_j$ is collinear with other variables in the
model.

3. Suppose as before: $y = X\beta + u$ is the general
linear model $\hat{\beta} = (X'X)^{-1}X'y$ and var–cov. ($\hat{\beta}$) = $\sigma^2(X'X)^{-1}$.
Following Glauber and Farrar (1967):
The variance inflation factor (VIF) of the diagonal elements of the inverted correlation matrix \((X'X)\) is a good measure of the degree and location of multicollinearity, Glauber and Farrar (1967).

Let
\[
\hat{c}_{ii} = \frac{|(X'X)_{ii}|}{|(X'X)|} \quad \text{where} \quad (XX)_{ii} \text{ denotes the correlation matrix excluding the } i\text{th variable and } (X'X) \text{ is the whole correlation matrix.}
\]

This ratio lies between 1 and \(\infty\), i.e.,
\[
1 \leq \hat{c}_{ii} \leq \infty
\]

\(\hat{c}_{ii} = \text{unity indicates that } X_i \text{ is orthogonal to the other explanatory variables.}\)

On the other hand, if \(\hat{c}_{ii}\) tends to \(\infty\), \((X'X)\) is sing-
ular and there is perfect multicollinearity, and at least one explanatory variables in the variable set is a linear combination of the other \((p-l)\) explanatory variables.

However, \(\text{VIF}_j = 10\), for example, indicates that the variance associated with the \(j\)th variable in \((X'X)\) is ten times larger than its value would be if \(X_i\) were orthogonal to the other \((p-l)\) explanatory variables.

**Thisted Method**

Thisted (1976, 1978) and Thisted and Morris (1979) suggested that the information content of the data may be different for different purposes. They pointed out the inadequacy of \((\text{VIF})\) measure of multicollinearity and introduced indexes based on MSE and predictive mean squared error \((\text{PMSE})\) where:

\[
\text{PMSE} \left( \hat{\beta}' \beta \right) = (\hat{\beta} - \beta)'(X'X)(\hat{\beta} - \beta)
\]

These indexes are calculated for overall model estimation \((\text{Mci})\) and for model predictability \(\text{Pmci}\).

\[
\text{Mci} = \sum_{i=1}^{P} \frac{\lambda_i^2}{\lambda_i^2} \quad \text{and } \text{Pmci} \text{ is given by}
\]

\[
\text{Pmci} = \sum_{i=1}^{P} \frac{\lambda_i}{\lambda_i} \quad \text{where } \lambda_i \text{ is the smallest eigenvalue of } (X'X).
\]
Values of Mci and Pmci close to one indicate a high multicollinearity problem, and values greater than two indicate relatively low multilinearity.

Thisted and Morris claimed some advantages for the Thisted multicollinearity indexes over other methods. However, these advantages are not very clear: For example, they claim that the indexes show that the effect of multicollinearity is worse for estimation than for prediction. However, this same result can also easily be shown by the following:

\[
V(y) = V(X\hat{\beta}) = \text{tr}E[(X\hat{\beta} - EX\hat{\beta})(X\hat{\beta} - EX\hat{\beta})']
\]

\[
= \text{tr}E[(X(\hat{\beta} - \beta))[X(\hat{\beta} - \beta)']']
\]

\[
= \text{tr}E[(X(\hat{\beta} - \beta)(\hat{\beta} - \beta)X')]
\]

\[
= \text{tr}[X[E(\hat{\beta} - \beta)(\hat{\beta} - \beta)']X'] \quad \text{since } X \text{ is fixed}
\]

\[
= \text{tr}[X[\sigma^2(X'X)^{-1}]X'] = \sigma^2 \text{tr}[[X(X'X)X']] = \sigma^2 \text{tr}(I) = \sigma^2.
\]

Consequences of multicollinearity

The most serious consequences of multicollinearity can be summarized as follows:

1. Fall in the precision of the estimates.
2. OLS will result in very sensitive and unstable estimates.
3. OLS will result in estimates that are very sensitive to alternative model specifications.
I will try to elaborate on each point of the consequences of multicollinearity.

1. Fall in the precision of the OLS estimates:
Due to the lack of independence among some or all explanatory variables the separation of individual effect will be difficult and the precision of OLS estimates will fall. The following example could be used to illustrate this point:

Let \((X'X) = R = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}\)

and \(\text{Var-Cov} (\hat{\beta}) = \sigma^2 (X'X)^{-1} = \frac{\sigma^2}{1-r_{12}^2} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix}\) (1.1)

From (1.1) one can see how \(r_{12}\) will affect the variance of \(\hat{\beta}_1\). Also high intercorrelation will result in deleting important variables based on t and F statistics.

2. Due to multicollinearity OLS estimates are very sensitive to changes in data. An inclusion of more data (more information) might have the effect of decreasing the variance and changing the magnitude of the coefficients.

3. Estimates obtained by OLS when multicollinearity is present are very sensitive to model specification. The addition of another variable to the model specification will usually increase the variance, as would be expected from \(\text{V}(\hat{\beta}) = \sigma^2 \text{tr}(X'X)^{-1}\). \(\text{V}(\hat{\beta})\) will always increase as the specification of the model is expanded to include
another explanatory variable, although by only $\sigma^2$ if the added variable is uncorrelated with the original set of explanatory variables.

**Suggested solutions to multicollinearity**

As pointed out by Sass, (1979), there is no complete remedy to the multicollinearity problem but there are some widely recommended solutions to be discussed in this section:

1. Data disaggregation
2. Use of prior information (including Theil-Goldberger)
3. Biased linear estimators
   a. Variable deletion
   b. Use of principal components
   c. Use of ridge regressions

**Data disaggregation:** the solution to multicollinearity is highly dependent on the sources of multicollinearity and the nature of data used. As indicated by Brown and Nawas (1973), when multicollinearity is caused by data aggregation they found that data disaggregation tends to greatly alleviate the problem of multicollinearity. Of course this solution might not be possible under most situations, but when possible it results in more stable OLS estimates.

**Use of prior information:** use of prior information
is also a common practice for breaking the pattern of multicollinearity. It is limited "only" by the availability of this prior information, which is supposed to be unbiased. Prior information about some coefficients in a regression model might be obtained from similar studies and from theoretical and practical considerations. For example, the model:

\[ y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e \]

could be estimated even though \( x_2 \) and \( x_3 \) were highly correlated if the researcher had some information about the value of \( \beta_3 \). For example, he/she could fit the model \((y - \beta_3 x_3) = \beta_0 + \beta_1 x_1 + \beta_2 x_2\) to estimate \( \beta_0, \beta_1 \) and \( \beta_2 \). Depending on how accurate the information is about \( \beta_3 \), the estimates obtained by this method may be biased and perhaps should be included in the class of biased estimators, even though they are not usually considered to be biased.

Smith and Campbell (1980) argued that if prior beliefs or informations are available, then they ought to be incorporated directly into the estimation procedure and that this incorporation should be done 'accurately and efficiently.' Others, Thisted (1979), Thisted and Morris (1979) made the point that one's beliefs are not always easily and correctly reduced to prior distributions and hence, the investigator does better to incorporate imprecise information approximately through the use of
other methods, such as ridge regression which serve as a good approximation, rather than to insist rigidly on formulating a prior distribution with imprecise information.

Theil-Goldberger mixed model

One specific method of incorporating prior information into a regression model is the method introduced by Theil and Goldberger (1964) which could be summarized as follows:

Assume the general linear regression model
\[ y = X\beta + u \]
where \( y \) is \((n \times 1)\) vector of dependent variable, \( X \) is the \((n \times p)\) matrix of explanatory variables, \( \beta \) is the \((p \times 1)\) vector of coefficients and \( u \) is the \((n \times 1)\) vector of disturbances, OLS estimate \( \hat{\beta} = (X'X)^{-1}X'y \) and var-cov \( (\hat{\beta}) = \sigma^2(X'X)^{-1} \). Assume also that prior information about the coefficients are available in the form:
\[ \phi = R\beta + v \]
where \( \phi \) is a \((k \times 1)\) vector of estimates for \( R\beta \), \( R \) is a \((k \times r)\) matrix of known fixed elements determining which elements have prior information and how this information is weighted, \( \beta \) is the \((r \times 1)\) vector of fixed unknown parameters and \( v \) is the \((k \times 1)\) vector of error term for the prior information. It is also assumed that \( v \) and \( u \) are independent and that \( E(v) = 0 \) and \( E(vv') = \psi \).

The general model with prior information becomes:
\[
\begin{bmatrix}
  y \\
  \phi 
\end{bmatrix} =
\begin{bmatrix}
  X \\
  R 
\end{bmatrix} \beta +
\begin{bmatrix}
  u \\
  v 
\end{bmatrix}
\]
where
\[
E\begin{bmatrix}
  u \\
  v 
\end{bmatrix} = 0 \text{ and } E\begin{bmatrix}
  u \\
  v 
\end{bmatrix}(uv) =
\begin{bmatrix}
  \sigma^2 I & 0 \\
  0 & \psi 
\end{bmatrix}
\]

\(\beta\), the generalized least square (GLS) estimators

becomes:
\[
\tilde{\beta}' = \left( (X'X)^{-1} + (X'R')^{-1} \right) \left( X'X \right)^{-1} \left( X'R' \right)^{-1} \begin{bmatrix}
  \sigma^2 I & 0 \\
  0 & \psi 
\end{bmatrix}
\begin{bmatrix}
  y \\
  \phi 
\end{bmatrix}
\]

when \(\theta = \frac{1}{\sigma^2}\) \(\tilde{\beta}'\) is reduced to
\[
\tilde{\beta} = \left( \theta X'X + R'\psi^{-1}R \right)^{-1} \left( \theta X'y + R'\psi^{-1}\phi \right)
\]

and the variance = covariance \((\tilde{\beta})\) is
\[
\text{var-cov} (\tilde{\beta}) = \left( \theta X'X + R'\psi^{-1}R \right)^{-1}.
\]

The application of the mixed estimation method to incorporate quantitative prior information is not without its limitations, one of which is the prerequisite that the sample and prior information be compatible, for which Theil (1972) suggested the \(\chi^2\) test for the following hypothesis: \(H_0: \text{The two sources of informations are compatible vs. } H_a: \text{The two sources of information are not compatible.}\)

Another limitation of the Theil-Goldberger mixed model is the assumption that the prior information vector \(\phi\) is unbiased.

Johnston (1972) pointed out that there are some limitations to the use of the model(1.2) because of the
effect of the prior $\beta$ information on obtaining sample information and on the reliability of this information. 

(2) As can be seen in (1.2), the expression for $\tilde{\beta}$ contains the unknown parameters $\sigma^2$ which has to be estimated from the sample information, which together with the dominancy of the prior information might undermine the whole purpose of regression analysis in some cases. It is important to point out that the use of prior information is not limited to information in an equality form, an inequality form of prior information also could be incorporated. A different procedure has to be used for inequality constraints Judge and Takayama (1966).

**Biased linear estimators:** in multiple linear regression, ordinary least square estimates (OLS) are best linear unbiased estimators (BLUE). Although these properties enjoyed by OLS are theoretically satisfying, in practice they may have inflated variances caused by the problem of multicollinearity.

A number of alternatives to ordinary least squares have been introduced and recommended. They are estimates that are biased, but may be desirable on other counts. These estimates are included in a class of biased estimators which may include any estimators that yield a biased estimate. Only a subgroup of this class will
be discussed in this section.

a. Variable Deletion:

Deletion of variable is a common practice among researchers and results from the effect of multicollinearity when no other satisfactory solution appears to exist. It is very important to recognize that when deleting a relevant variable from the specification of a model a "specification bias" is incurred (Johnston (1972). The following discussion is offered for the inclusion of variable deletion in the class of biased estimators. Suppose the true multiple regression model is

$$y = X\beta + Z\alpha + e$$  \hspace{1cm} (1.3)

where $y$ is the $(n \times 1)$ dependent variable vector, $X$ and $Z$ are matrices of independent variables of $(n \times p-r)$ and $(n \times r)$ dimensions, $\beta$ and $\alpha$ are two vectors of coefficients of $(p-r \times 1)$ and $(r \times 1)$ dimensions, respectively, and $e$ is the $(n \times 1)$ vector of random numbers, $e \sim N(0,1)$, etc.

Suppose, instead of fitting the general true model in (1) the researcher fits by OLS
\[ y = X\beta + e \]

implies \[ \hat{\beta} = (X'X)^{-1}X'y \]

\[ \hat{\beta} = (X'X)^{-1}X'(X\beta + Z\alpha + e) \]

\[ E(\hat{\beta}) = E\{(X'X)^{-1}X'X\beta + (X'X)^{-1}X'Z\alpha + (X'X)^{-1}X'e\} \]

\[ = \beta + (X'X)^{-1}X'Z E(\alpha) + 0 \] since \( E(e) = 0 \) implies \( E(\hat{\beta}) - \beta = (X'X)^{-1}X'Z\alpha \). Note that the expression \((X'X)^{-1}X'Z\)

defines the OLS estimators that results from regressing \( Z \) on \( X \). Then,

\[ [E(\hat{\beta}) - \beta] = \hat{\Gamma}\alpha \] where \( \hat{\Gamma} = (X'X)^{-1}X'Z \).

This bias could be large depending on the correlation between the omitted variables and the retained ones and depends on the (coefficients)\( \alpha \). From this exposition, a serious "specification bias" could be incurred when a relevant explanatory variable is omitted from the model.

b. Principal component regression:

The occurrence of small eigenvalues of \((X'X)\) is a warning of the presence of multicollinearity. It is clear that if there are \((S)\) zero eigenvalues, the number of input variables can be reduced by \((S)\), but if the small eigenvalues are near zero the situation is not so clear. Near zero eigenvalues may represent near actual linear dependencies and the departure from zero may be due to measurement and computational errors.

Principal components analysis, like other biased linear estimation, is suitable for coping with some problems of multicollinearity and has been introduced
as a suitable technique to partially cope with the problem of multicollinearity in the data. Principal components regression is like variable deletion, but it involves deleting components rather than deleting the whole variable. Consider an orthogonal transformation of the general model \( y = X\beta + u \) such that

\[
Q'X'XQ = \Lambda \quad \text{and} \quad Q'Q = I = QQ'
\]

where \( \Lambda \) denotes the diagonal matrix of eigenvalues (\( \lambda \)) and \( Q \) denotes the orthogonal matrix of eigenvectors.

Letting \( Z = XQ \) and \( \beta = Q\alpha \) the linear model becomes

\[
y = Z\alpha + e \quad \text{where} \quad \hat{\alpha} \text{ the (OLS) estimate of} \ \alpha \text{ is obtained from the equation}
\]

\[
\Lambda \hat{\alpha} = Z'y \quad \text{which could be translated to the original parameters as}
\]

\[
\hat{\beta} = Q\hat{\alpha}.
\]

Now, if there are \( S \) zero eigenvalues it follows that the corresponding columns of \( Z \) are zero and these variables drop out of the orthogonal model and the linear equation becomes of dimension \( g = p-s \). Following Hocking (1970) these equations could be written in partitioned forms as:

\[
Q_1 = (Q_g: \ Q_s)
\]

\[
\alpha' = (\alpha'_g: \ \alpha'_s) \quad \text{and}
\]

\[
\Lambda = \text{Diag} (\Lambda_g, \ \Lambda_s) \quad \text{then}
\]

\[
\Lambda_g \hat{\alpha}_g = Q'X'y \quad \text{and interms of the untransformed parameters:}
\]

\[
\hat{\beta}_g = Q_g' \hat{\alpha}_g
\]
The techniques of 'principal components' regression are based on this analysis. That is, if X is actually of rank p, but s of the eigenvalues are judged to be "sufficiently small" they are then set equal to zero and the true parameter \( \beta \) is estimated by \( \hat{\beta} \) as shown by Hocking (1970).

It is interesting to compare Principal Component Regression with Ridge Regression and other biased linear estimators. Hocking also showed that the above problem could be formulated as the solution to the problem:

Minimize \( F(X) = (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \)

\( \hat{\beta} \)

subject to \( Q_s'\beta = 0 \) which enables us to write \( \hat{\beta} = (I - Q_sQ_s') \hat{\beta} \)

which is similar but different from Ridge Estimator \( \hat{\beta}^* = (X'X + kI)^{-1} \hat{\beta} \). To illustrate for the case \( p = 2 \) and \( q = p - s = 1 \):

\[
X'X = \begin{bmatrix}
1 & r_{12} \\
r_{12} & 1
\end{bmatrix}
\]

, the eigenvalues are computed as

\( \lambda_1 = 1 + r_{12} \) and \( \lambda_2 = 1 - r_{12} \) and the eigenvectors are given by

\[
Q = \frac{1}{\sqrt{2}} \begin{bmatrix}
1 & 1 \\
1 & -1
\end{bmatrix}
\]

If \( r_{12} \) is close to one and hence \( \lambda_2 \) is close to zero the problem becomes

\[
\min \frac{1}{\hat{\beta}} (\beta - \hat{\beta})'X'X(\beta - \hat{\beta})
\]
subject to \( \beta_1 = \beta_2 \) which yields

\[
\beta_+ = (I - Q_{sp}'Q_s)\hat{\beta}, \quad \text{Hocking (1970)}.
\]

Note that when all the components are to be retained the Principal component estimate has the same total variance as the OLS estimate. Thus, Principal component Regression is essentially a technique to choose a subset of orthogonalized input variables to include in the regression analysis. It is clear that reducing the number of principal components could result in a more stable estimate than OLS which can be seen by noticing that the total variation in the standardized explanatory variables is

\[
\sum_{i=1}^{P} x_i^2 = \sum_{i=1}^{P} 1 = p,
\]

which is the trace of the correlation matrix \((X'X)\).

But \( \text{tr} \Lambda \) is also \( p \) for the orthogonal model for principal components:

\[
\text{tr}(X'X) = \text{tr}(X'XI). \quad \text{Since } QQ' = I \text{ and } Q \text{ is an orthogonal } (p \times p) \text{ matrix.}
\]

\[
\text{tr}(X'X) = \text{tr}(X'XQQ') \quad \text{by substitution}
\]

\[
= \text{tr}(Q'X'XQ) \quad \text{since } \text{tr}(AB) = \text{tr}(BA). \quad \text{Since}
\]

\[Q'X' = Z' \text{ and } XQ = Z, \text{ by substitution}
\]

\[
\text{tr}(X'X) = \text{tr}(Z'Z)
\]

\[
= \sum_{i=1}^{P} z_i^2
\]

\[
= \sum_{i=1}^{P} \lambda_i \quad \text{since } Z'Z = \Lambda = \text{diag } (\lambda_1 \ldots \lambda_P)
\]

\[
= p \text{ number of explanatory variables.}
\]

For the case where some \( \lambda_i \) are near zero but not
zero, a smaller number of $\lambda_i$ will usually account for most of the variation in the explanatory variables and, equivalently, the corresponding smaller number of component $Z_i$ will usually also account for most of the variation in the explanatory variables.

There are many principal component estimators in the literature, each differentiated by the method used to select principal components:

1. Delete component corresponding to the smallest eigenvalues.

2. Test the significance of each component coefficient $\alpha_i$ by t or F test to determine whether the corresponding component should be deleted.

3. Operational mean-square error related criterion for deleting principal components introduced by Sass (1979). Based upon the results of this thesis, another squared-error related criterion would be to estimate the true squared coefficient $\alpha_i^2$ by the unbiased estimator,

$$\hat{\theta}_i = (\alpha_i^2 - \sigma^2 / \lambda_i) \text{ where } \hat{\theta}_i \geq 0.$$

Then the decision criterion could be to delete components $\alpha_i$ if $\text{Var} (\hat{\alpha}_i) \geq \hat{\theta}_i$ and use $\hat{\theta}_i$ for computing the estimated loss $\hat{L}_1(\hat{\alpha}_i)$. This criterion is similar in principle to that employed by Sass where the criterion was to delete the component $Z_i$ if $\text{Var} (\hat{\alpha}_i) \geq c^2 = (\sum_{i=1}^{p} \lambda_i \hat{\theta}_i^2) / p$. 
c. Ridge Regression:

The concept of ridge regression as described by Hoerl and Kennard (1970a, 1970b) was motivated by the fact that models with "ill-conditioned" correlation matrices can lead to estimates with very high variance and far removed from the true parameters. To see this assume the general linear model

\[ y = X\beta + u \]

where \( y \) denotes the \((n \times 1)\) dependent variable vector, \( X \) denotes the \((n \times p)\) matrix of fixed explanatory variables, \( \beta \) denotes the \((p \times 1)\) vector of parameters and \( u \) is the \((n \times 1)\) vector of disturbances with the assumed properties that \( E(u) = 0 \), and \( E uu' = \sigma^2 I_n \). Let \((X'X)\) denote the correlation matrix of the explanatory variables. \( \hat{\beta}^* \), the standardized ridge regression estimates of the coefficient is

\[
\hat{\beta}^* = (X'X + kI)^{-1}X'y
= (X'X + kI)^{-1}X'(X\beta + u)
= (X'X + kI)^{-1}X'X\beta + (X'X + kI)^{-1}X'u
\]

which implies \( E(\hat{\beta}^*) = (X'X + kI)^{-1}X'X\hat{\beta} \) since \( E(u) = 0 \)

\[
= C\hat{\beta}
\]

where \( C = (X'X + kI)^{-1}X'X \). Note that \( \hat{\beta}^* \) is also the solution to the problem.

\[
\min_{\beta} (\hat{\beta}^* - \beta)'X'X(\hat{\beta}^* - \beta)
\]
subject to $\beta^T \beta \leq d^2$, which depends on the value of $k$ in the model, as shown by Hocking (1970). The variance-covariance of $\hat{\beta}^\top$ is $\text{var-cov}(\hat{\beta}^\top) = \sigma^2 (X'X + kI)^{-1}X'X(X'X + kI)^{-1}$.

Marquardt (1970) showed that for OLS the mean squared error is $E(L_1^2) = \sigma^2 \sum_{i=1}^{p} 1/\lambda_i$ in terms of the orthogonal model and $\text{var}(L_1^2) = 2\sigma^4 \sum_{i=1}^{p} 1/\lambda_i^2$. If one or more of the $\lambda_i$ are very small it is clear that $\hat{\alpha}_i$, although unbiased, may be far removed from the true parameter $\alpha_i$. The ridge regression techniques have generated considerable interest since being introduced into the literature of statistics by Hoerl and Kennard (1970a).

Much of the discussion of ridge regression centers around the choice of the constant term $k$. Hoerl and Kennard (1970a) established the general notion of the existence of a constant $k$ which yields an estimator with smaller mean square error (MSE), and they recommended the inspection of the "ridge trace" as a means for estimating $k$.

Other authors have recommended alternative schemes for estimating $k$ based on the data, some of which will be discussed in the following.

**Hoerl, Baldwin and Kennard (\text{k}_{\text{HBK}}) Ridge Estimator**

Hoerl, Baldwin and Kennard (HKB), in 1975 proposed a technique for choosing $k$ which will be referred to as
As before, assume a general linear model $y = \mathbf{X}\beta + u$, which could be orthogonally transformed into $y = \mathbf{Z}\alpha + u$ where $\mathbf{Z} = \mathbf{X}\mathbf{Q}$ and $\alpha = \mathbf{Q}'\beta$, as discussed earlier. HBK show that minimum MSE corresponds to an optimal value of $k = p \sigma^2/\beta'\beta$ where $p$ is the number of parameters in the model (rank of $(\mathbf{X}'\mathbf{X})$), $\beta$ is the $(p \times 1)$ vector of true coefficients, and $\sigma^2$ is the true variance. Since $\sigma^2$ and $\beta$ are unknown to the researcher, HKB recommended that $k$ be estimated using

$$k_{HBK} = \frac{p \hat{\sigma}^2}{\hat{\beta}'\hat{\beta}},$$

where $\hat{\sigma}^2$ is the OLS estimate of the variance and $\hat{\beta}$ is the $(p \times 1)$ vector of OLS estimates.

Lawless and Wang (1976) introduced another criterion for choosing the biasing coefficient $k$ which will be referred to as $k_{LW}$. $k_{LW}$ is different from $k_{HBK}$ by the fact that the coefficients are weighted according to the eigenvalues of $(\mathbf{X}'\mathbf{X})$. The optimal value of $k$ is

$$k = \frac{p\sigma^2}{\sum_{i=1}^{p} \lambda_i \hat{\sigma}_i^2},$$

which can be estimated from the sample data by

$$k_{LW} = \frac{p\hat{\sigma}^2}{\sum_{i=1}^{p} \lambda_i \hat{\sigma}_i^2},$$
where $\lambda_i$ denotes the $i$th eigenvalue, $\hat{\sigma}^2$ is the OLS estimate of the variance, and $\hat{\alpha}_i$ is the OLS estimate of the $i$th coefficient, $\alpha_i$.

A simulation study by Lawless and Wang (1976) showed that the ridge estimate based upon $k_{LW}$ usually averaged lower squared error than the ridge estimate based upon $k_{HKB}$.

Dempster, Shatzoff and Wermuth ($k_{DSW}$) Ridge Regression

Dempster, Shatzoff and Wermuth (1977), motivated by a Bayesian interpretation of ridge, introduced another method for the choice of $k$, called ridgm, which will be referred to as $k_{DSW}$.

$$k_{DSW} = \frac{p\hat{\sigma}^2}{\sum \lambda_i \hat{\alpha}_i^2 - p\hat{\sigma}^2},$$

where $\hat{\alpha}_i$ is the OLS estimate of $\alpha_i$, $\hat{\sigma}^2$ is the OLS estimate of $\sigma^2$, $\lambda_i$ is the $i$th eigenvalue of $(X'X)$. This specific choice of $k$ yields an estimator with MSE very similar to $k_{LW}$.

Lawless (1978) concluded that there is actually very little gain over $k_{LW}$ by using $k_{DSW}$.

Generalized Ridge Regression

Consider the standard general regression model
\[ y = X\beta + u \]

where \( X \) is the \((n \times p)\) matrix of explanatory variables, \( y \) is the \((n \times 1)\) dependent variable vector, \( \beta \) is a \((p \times 1)\) vector of coefficients to be estimated, and \( u \) is the \((n \times 1)\) vector of disturbances where \( E(u) = 0 \) and \( E(uu') = \sigma^2 I_n \). The ridge regression estimate of \((\hat{\beta}^*)\) for a fixed \( k \) is \( \hat{\beta}^* = (X'X + kI)^{-1}X'y \), which is the form discussed in earlier sections.

The generalized ridge estimator of \( \beta \) is given by

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

where \( K \) is a diagonal matrix with non-negative diagonal elements \((k_1, k_2, \ldots, k_p)\). The \( k_i \) \((i=1, 2, \ldots, p)\) need not be equal.

Choice of \( k_i \)

Optimal value for the \( k_i \) for the generalized ridge regression are those \( k_i \)'s which minimize MSE \((\hat{\beta}^*)\), which also minimize MSE \((\hat{\alpha}^*)\) which expressed as

\[
\psi = E(\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta) = E(\hat{\alpha}^* - \alpha)'(\hat{\alpha}^* - \alpha),
\]

Since

\[
E(\hat{\alpha}^* - \alpha)'(\hat{\alpha}^* - \alpha) = E(Q'^{\prime}\hat{\beta}^* - Q\beta)'(Q^{\prime}\hat{\beta}^* - Q\beta)
= E[Q'(\hat{\beta}^* - \beta)]'[Q'(\hat{\beta}^* - \beta)]
= E[(\hat{\beta}^* - \beta)'QQ'(\hat{\beta}^* - \beta)]
= E(\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta) \text{ since } QQ' = I = Q'Q.
\]

Since the \( Z_i \) explanatory variables are orthogonal, we can look at each one separately without loss of generality (wlog).
\[ \psi_i = E[(\hat{\alpha}_i^* - \alpha_i)'(\hat{\alpha}_i^* - \alpha_i)] \]

\[ \psi_i = E[\hat{\alpha}_i^{*2} - 2\hat{\alpha}_i^*\alpha_i + \alpha_i^2] \]

\[ = E[m_i^2 - 2m_i\alpha_i\alpha_i + \alpha_i^2] \]

where \( m_i = \frac{\lambda_i}{\lambda_i + k_i} \) which implies that

\[ \psi_i = m_i^2(\alpha_i^2 + \sigma^2/\lambda_i) - 2m_i\alpha_i^2 + \alpha_i^2. \]

Therefore,

\[ \frac{\partial \psi_i}{\partial m_i} = 2m_i(\alpha_i^2 + \sigma^2/\lambda_i) - 2\alpha_i^2 = 0, \]

which implies that optimal \( m_i = \frac{\alpha_i^2}{\alpha_i^2 + \sigma^2/\lambda_i} \). But substituting

\[ \frac{\lambda_i}{\lambda_i + k_i} = m_i \]

and rearranging terms gives

\[ (\lambda_i + k_i)\alpha_i^2 = \lambda_i(\alpha_i^2 + \sigma^2/\lambda_i) = (\lambda_i\alpha_i^2 + \sigma^2) \]

or \((\lambda_i + k_i) = (\lambda_i\alpha_i^2 + \sigma^2)/\alpha_i^2 = \lambda_i + \sigma^2/\alpha_i^2\)

which implies that optimal \( k_i = \sigma^2/\alpha_i^2 (i=1,2,..p) \).

Since \( \sigma^2 \) and \( \alpha_i^2 \) are generally unknown, the \( k_i \) need to be estimated. Different estimation procedures using iteration techniques are suggested by HBK (1975) and by Hemmerie (1975).

**The loss function**

One method of judging the accuracy of the estimates in regression analysis is the loss function, defined as the squared difference between the estimate and its true value,
\[ L_1^2 = \sum_{i=1}^{p} (\hat{\beta}_i - \beta_i)^2, \] 
where \( \hat{\beta}_i \) is the coefficient estimate and \( \beta_i \) is the true parameter. Ideally, the researcher's objective is to minimize the loss function, \( L_1^2 \), for each sample which can be done if one knows the contribution of each estimated coefficient to the loss function. With a relatively unstable OLS estimate, \( \hat{\beta}_i \), the loss function might be minimized by setting \( \hat{\beta}_i = 0 \), i.e., by deleting the corresponding explanatory variable, \( X_i \). The loss function is then equal to \( \beta_i^2 \). On the other hand, when \( (\hat{\beta}_i - \beta_i)^2 < \beta_i^2 \), there would be a net gain by not deleting \( X_i \). In an applied regression analysis, the true coefficients, \( \beta_i \)'s, are usually unknown which makes the estimation of the loss function not a straightforward procedure.

**Estimation of the loss function**

**for biased linear estimators**

Unfortunately, the true loss function as described above is of no help in applied situations where the true regression parameters are unknown and the only information that is available is sample information. Practical use of the loss function is only possible by estimation. Following Sass, (1979), one could simplify the earlier expression for the loss function by estimating its two components the variance and the square of the bias.
Note that in the case of unbiased estimators the above expression is simplified to only one component, the variance. In contrast, estimation of the squared error, $L_1^2$, is more difficult in the case of biased estimators due to the bias squared component which depends on the true coefficients, $\beta$, which are in practice unknown.

Brown (1978) and Sass (1979) suggested a procedure for estimating the loss function for ridge estimators and some other biased linear estimators.

Recall that from the orthogonal general regression model $y = Z\alpha + u$, the ridge regression estimate of $\alpha$ is given by

$$\hat{\alpha}^* = (Z'Z + kI)^{-1}Z'y$$

and that its mean square error, \(MSE(\hat{\alpha}^*)\), is given by

$$MSE(\hat{\alpha}^*) = \sigma^2 \sum_{i=1}^{p} \lambda_i (\lambda_i + k)^{-2} + k^2 \sum_{i=1}^{p} \alpha_i^2 (\lambda_i + k)^{-2}$$

To simplify the above expression further, recall that

$$\hat{\alpha}^* = (Z'Z + kI)^{-1}Z'y$$

$$= (Z'Z + kI)^{-1}(Z'Z)(Z'Z)^{-1}Z'y$$

$$= M\hat{\alpha}$$

where $M = (Z'Z + kI)^{-1}Z'Z$ and

$$\hat{\alpha} = (Z'Z)^{-1}Z'y$$

and an individual ridge coefficient could be expressed as a function of the OLS estimate, i.e.: \(\hat{\alpha}_i^* = m_i \hat{\alpha}_i\) where:

$$m_i = \lambda_i (\lambda_i + k)^{-1}$$

since the diagonal elements of $(Z'Z + kI)^{-1}$ and $(Z'Z)$ are $(\lambda_i + k)^{-1}$ and $\lambda_i$ respectively.
Hence, the MSE ($\hat{\alpha}^*$) could be expressed as

$$\text{MSE} (\hat{\alpha}^*) = \sum_{i=1}^{p} E(\hat{\alpha}^*_i - \alpha_i)^2$$

$$= \sum_{i=1}^{p} E(m_i \hat{\alpha}_i - \alpha_i)^2$$

$$= \sum_{i=1}^{p} E(m_i^2 \hat{\alpha}_i^2 - 2m_i \alpha_i \hat{\alpha}_i + \alpha_i^2)$$

$$= \sum_{i=1}^{p} (m_i^2 (\alpha_i^2 + \sigma^2 \lambda_i^{-1}) - 2m_i \alpha_i^2 + \alpha_i^2)$$

since $E(\hat{\alpha}_i^2) = \alpha_i^2 + \sigma^2 / \lambda_i$ and $E(\alpha_i \hat{\alpha}_i) = \alpha_i^2$

$$= \sum_{i=1}^{p} [\sigma^2 m_i^2 \lambda_i^{-1} + m_i^2 \alpha_i^2 - 2m_i \alpha_i^2 + \alpha_i^2]$$

$$= \sum_{i=1}^{p} (m_i^2 \sigma^2 \lambda_i^{-1} + (1 - m_i)^2 \alpha_i^2).$$

which reduces to $\sigma^2 / \lambda_i$ for OLS since when $k = 0$, $m_i = 1$.

As indicated earlier the variance portion of MSE($\hat{\alpha}^*$) could be estimated using the sample information only; however, the bias square component depends on the unknown true parameter $\alpha_i$. In this section an estimation procedure for estimating the true coefficient $\alpha_i$ will be introduced.
Proposed method for estimating the loss function

Simulation results by Brown (1978) and Sass (1979) showed that using $C^2 = 1/p\sum \hat{\alpha}_i^2$ in the expression for estimating the loss function performed well in the experiments. However, the design of the Monte Carlo experiments was similar to that employed by Hoerl, Kennard, and Baldwin and by Lawless and Wang where the true $\alpha_i$ were essentially selected randomly and the variances of the true $\alpha_i$ were all nearly equal. Although the results of such simulations might indicate how such an estimate of squared error might perform across many different empirical situations, it can be argued that researchers usually have only one particular empirical situation and model. Consequently, they need a reliable estimate of squared error, $\hat{L}_1^2$, for their particular situation. As a result, there has been a recent trend toward simulations with various fixed vectors of $\alpha_i$ values, some favorable for ridge regression and some unfavorable, e.g., Lawless (1978).

Because of the preceding considerations, the earlier simulations with random $\alpha_i$, using $\hat{C}^2 = 1/p\sum \hat{\alpha}_i$ for estimating $\alpha_i^2$ in computing bias squared, have to be discounted somewhat. These considerations have motivated the search for an estimate of $\alpha_i^2$ for computing bias squared that would be more reliable for detecting those
particular situations unfavorable for ridge regression, namely, where the $\alpha_i$ value(s) associated with the smallest eigenvalue(s) is(are) considerably larger than the other $\alpha_i$ values.

Recall that for the general regression model, $y = X\beta + u$, the OLS estimate of $\beta$ is $\hat{\beta}$, given by

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

Substituting $X\beta + u$ for $y$,

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

$$= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$$

$$= \beta + (X'X)^{-1}X'u,$$

which implies that

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

$$= (\beta' + u'(X'X)^{-1})(\beta + (X'X)^{-1}X'u).$$

$$E(\hat{\beta}'\hat{\beta}) = \beta'\beta + \beta'(X'X)^{-1}X'u + u'X(X'X)^{-1}\beta +$$

$$u'X(X'X)^{-1}(X'X)^{-1}X'u$$

$$= \beta'\beta + E(u'X(X'X)^{-1}(X'X)^{-1}X'u)$$

by taking the transpose and since $X$ is assumed to be fixed and $E(u_i) = 0$

$$E(u_i) = 0$$

$$= \beta'\beta + E(u'u(X'X)(X'X)^{-1})(X'X)^{-1}$$

$$= \beta'\beta + \sigma^2 \text{tr}(X'X)^{-1}$$

implies that

$$\beta'\beta = E(\hat{\beta}'\hat{\beta}) - \sigma^2 \text{tr}(X'X)^{-1}$$

is an unbiased estimate of $(\beta'\beta)$, to be used instead of $\hat{C}^2$ to estimate the loss function. Recall that using the same derivation one could derive an unbiased estimate for the true coefficients square ($\alpha'\alpha$) as:

$$\alpha'\alpha = E(\hat{\alpha}'\hat{\alpha}) - \sigma^2 \Lambda^{-1}$$

where $\Lambda$ is the $(p \times p)$ matrix of eigenvalues and for the individual coefficient $\alpha_i^2$ is
\[ \alpha_i^2 = (E(\hat{\alpha}_i^2) - \sigma^2 / \lambda_i). \]

The mean square error expression could be estimated by substituting \( \hat{\theta}_i = (\hat{\alpha}_i^2 - \sigma^2 / \lambda_i) \), restricting \( \hat{\theta}_i > 0 \) in place of \( \hat{c}^2 \) to become:

\[
\text{MSE}(\hat{\alpha}_i^2) = \sum_{i=1}^{P} \left[ m_i^2 \sigma^2 \lambda_i^{-1} + (1-m_i)^2 \hat{\theta}_i \right].
\]
II. THE MONTE CARLO EXPERIMENTS

Description:

The Monte Carlo Experiments were based on the three explanatory variable models used by Lawless (1978). Throughout this simulation the standard multiple linear regression model

\[ y = X\beta + u \]

\( y \) is a \((n \times 1)\) vector, a dependent variable, \( X \) is a known \((n \times p)\) matrix of explanatory variables of rank \( p \), \( \beta \) is a \((p \times 1)\) known vector of coefficients, \( u \) is the \((n \times 1)\) vector of disturbances, and \( u \) is normally distributed with mean 0 and variance-covariance matrix \( \sigma^2 I_n \). The simulation assumes that \( X \) is standardized so that \((X'X)\) is a correlation matrix.

To show clearly the effects of multicollinearity the model \( y = X\beta + u \) is transformed into the canonical form where there exists an orthogonal \((p \times p)\) matrix, \( Q \), such that

\[ Q'X'XQ = \Lambda \]

where \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) \) contains the eigenvalues of \((X'X)\), let \( Z = XQ \) and

\[ \alpha = Q'\beta. \]

Then the general linear regression model can be written as:

\[ y = Z\alpha + u. \]

The OLS estimate of \( \beta \) is \( \hat{\beta} = (X'X)^{-1}X'y \), but for \( y = Zd + u \), \((Z'Z) = \Lambda\) and the OLS estimator of \( \alpha \) is...
\[ \hat{\alpha} = \Lambda^{-1}z'y = Q'\hat{\beta}. \]

It is well known that \( \hat{\beta} \) is distributed multivariate normal with mean \( \beta \) and variance-covariance matrix \( \sigma^2(X'X)^{-1} \).

It follows that \( \hat{\alpha} \) is also distributed multivariate normal with mean \( \alpha \) and variance-covariance matrix \( \sigma^2\Lambda^{-1} \). An indication of some degree of multicollinearity is when one or more \( \lambda_i \) is close to zero. The effect of small \( \lambda_i \) on \( \hat{\alpha}_i \) is clear since \( \text{var}(\hat{\alpha}_i) = \sigma^2/\lambda_i \) since \( \hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_p \) are independent.

Hoerl and Kennard (1970) showed that the effect of small \( \lambda_i \) on \( \hat{\beta}_i \) is not obvious, depending on the orthogonal matrix \( Q \), but what is clear is that when some \( \lambda_i \) are small, quantities such as
\[
E[(\hat{\beta} - \beta)'(\hat{\beta} - \beta)] = E[(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)]
\]
will be large.

In this simulation the MSE, the most often used measure taken in scaled form is to be used to judge the performance of the ridge biased estimators

\[
M_1(\beta) = 1/\sigma^2 E(\hat{\beta}^* - \beta)'(\hat{\beta}^* - \beta)
\]

\[ = 1/\sigma^2 E(\alpha^* - \alpha)'(\alpha^* - \alpha) \quad \text{as shown in the preceding chapter} \]

\[ = 1/\sigma^2 \sum_{i=1}^{p} E(\hat{\alpha}_i^* - \alpha_i)^2. \]

Since \( M_1(\beta) = \sum_{i=1}^{p} 1/\lambda_i \), when one or more \( \lambda_i \) are small

\( M_1(\hat{\beta}_i) \) will be large.

Also, \( M(\hat{\alpha}^*) = 1/\sigma^2 \sum_{i=1}^{p} \left( m_i^2 \sigma^2/\lambda_i + (1 - m_i)^2 \hat{\beta}_i \right) \)
where $\tilde{M}_l(\hat{\alpha})$ is the estimate (based on sample information),
\[ m_i = \frac{\lambda_i}{\lambda_i + k}, \quad \text{and} \quad \hat{\theta}_i = (\hat{\alpha}_i^2 - \sigma^2/\lambda_i) > 0 \]
is an estimate of $\alpha_i^2$.

For the OLS and $k_{LW}$ ridge estimator, a measure somewhat analogous to the coefficient of variation (C/V) will be computed and defined in a later chapter.

Estimators to be studied in this simulation

To cope with the problem of multicollinearity, the ridge class of biased linear estimators have been suggested by many authors; e.g., Hocking et al. (1976), Dempster et al. (1977), Lawless (1978) and others, as discussed in the preceding chapter. The following ridge estimators are considered here:

$k_{LW}$ ordinary ridge estimator with $k$ value computed by
\[ k_{LW} = P \frac{\hat{\sigma}^2}{\sum_{i=1}^{P} \lambda_i \hat{\alpha}_i^2} \]
and

$k_{HKB}$ ordinary ridge estimator with $k$ computed by
\[ k_{HKB} = P \frac{\hat{\sigma}^2}{\sum \hat{\alpha}_i^2}. \]

Estimators $k_{LW}$ and $k_{HKB}$ have been suggested by Hoerl, Kennard and Baldwin (1975) and by Lawless and Wang (1976), and these ridge estimators showed good properties in several simulation models.

RIDGN is an estimator similar to KGRR (below) based on the idea of using a shrunken value of $\hat{\alpha}_i$ by precomputing
\( \tau_i^2 = b(\lambda_i \sigma_i^2) \), estimating \( \tau_i^2 \) for each \( \lambda_i \). Forcing \( \tau_i^2 \) through the origin showed some advantages over \( \tau_{\text{HKB}} \) in terms of estimating \( \tau_i \).

**KGRR**: a generalized ridge estimator where

\[ k_i = \frac{\sigma^2}{\alpha_i^2} \]

This estimator is suggested by the fact that \( k_i = \frac{\sigma^2}{\alpha_i^2} \) is the value of \( k_i \) which minimize \( \text{MSE}(\hat{\tau}_i^2) \) if \( \sigma^2 \) and \( \alpha_i^2 \) are known as discussed in a previous chapter.

**ROP**: a theoretical estimator, unusable in practice, since it is based on the assumption of known \( \sigma^2 \) and \( \alpha_i^2 \) and choosing value of \( k \) that minimizes \( \text{MSE}(\hat{\tau}_i^2) \) for all \( \hat{\tau}_i^2 \).

**OLS**: finally, the ordinary least squares estimator will be presented for comparison purposes.

**Models considered**: following Lawless (1973) the following 3-explanatory variables are to be used in the simulations. These models give a wide range of signal-to-noise ratios and simulate realistic regression problems. These models were constructed with two sets of \( \lambda \) values:

a. \( \lambda_1 = .01 \quad \lambda_2 = 1.00 \quad \lambda_3 = 1.99 \)

b. \( \lambda_1 = .10 \quad \lambda_2 = 1.00 \quad \lambda_3 = 1.90 \)

Eight \( \alpha \) vectors: Seven of the eight \( \alpha \) vectors were suggested by Lawless (1979), and the eighth vector was added to this simulation as a more extreme case. These \( \alpha \) vectors are all constructed so that \( \sum_{i=1}^{3} \alpha_i^2 = 300 \):
\[ a_1 = (10 \ 10 \ 10), a_2 = (15 \ \sqrt{37.5} \ \sqrt{37.5}), a_3 = (\sqrt{37.5} \ 15 \ \sqrt{37.5}), \]
\[ a_4 = (\sqrt{37.5} \ \sqrt{37.5} \ 15), a_5 = (12 \ 12 \ \sqrt{12}), a_6 = (12 \ \sqrt{12} \ 12), \]
\[ a_7 = (\sqrt{12} \ 12 \ \sqrt{12}) \text{ and } a_8 = (16 \ \sqrt{22} \ \sqrt{22}). \]

There were a total of 16 \( \lambda, \alpha \) combinations. For each \( \lambda, \alpha \) value, 5 different \( \sigma^2 \) values were used. These were \( \sigma^2 = 4, 2, 1, 0.5, \) and 0.25.

With \( p = 3 \), a fairly wide range of values for \( \frac{3}{\sum \alpha_i^2/\sigma^2} \) are realized, and \( a_1 \ldots a_8 \) give models with different sizes of \( \alpha_i \) corresponding to the smallest eigenvalue \( \lambda_i \).

Based on samples of size 1000, the following are summary of the statistics collected.

For OLS:

a. \( M1(\hat{\alpha}) \)
b. \( \tilde{M1}(\hat{\alpha}) \)
c. coefficient of loss variation \( C*/V \), to be defined in chapter III.

For \( k_{\text{LW}} \):

a. \( M1(\hat{\alpha}^*) \)
b. \( \tilde{M1}(\hat{\alpha}^*) \)
c. coefficient of loss variation, \( C*/V \).

For \( k_{\text{HKB}}, \text{RIDGN}, \) and \( \text{KGRR} \):

a. \( M1(\hat{\alpha}^*) \)
b. \( \tilde{M1}(\hat{\alpha}^*) \)
Also for the distribution of $\hat{\alpha}^*$ by $k_{LW}$ the following intervals are computed:

a. frequencies of $\alpha_1$ falling beyond $-3\hat{\phi}$, i.e.,

$$\alpha_1 \leq -3\hat{\phi}$$

where $\hat{\phi} = \left[ m_1^2 \sigma^2 / \lambda_1 + (1 - m_1)^2 \hat{\phi}_1 \right]^{1/2}$, where $\hat{\phi} = \sqrt{L_1^*(\alpha^*)}$

$$\hat{\phi}_1 = (\alpha_1^2 - m_1^2 / \lambda_1)^{1/2} \geq 0.$$

b. frequencies of $-3\hat{\phi} < \alpha_1 \leq -2\hat{\phi}$

c. $-2\hat{\phi} < \alpha_1 \leq -\hat{\phi}$

d. $-\hat{\phi} < \alpha_1 \leq -0.5\hat{\phi}$

e. $-0.5\hat{\phi} < \alpha_1 \leq 0.5\hat{\phi}$

f. $0.5\hat{\phi} < \alpha_1 \leq \hat{\phi}$

g. $\hat{\phi} < \alpha_1 \leq 2\hat{\phi}$

h. $-2\hat{\phi} < \alpha_1 \leq 3\hat{\phi}$

i. $\alpha_1 \geq 3\hat{\phi}$

This information will be used to test hypotheses about the distribution of $\hat{\alpha}^*$.

For each 1000 experiments for each model with a given $\sigma^2$, $\lambda$, and $\alpha$ values, $\hat{\alpha}$ is first generated via

$$\hat{\alpha}_i \sim INN(\alpha_i, \sigma^2 / \lambda_i), i = 1, 2, \ldots, p.$$ 

Estimates are then generated for $\hat{\alpha}^*$, $M_1(\hat{\alpha}^*)$ and $M_1(\hat{\alpha}^*)$. The indicated statistics are collected, summed, and averaged over the 1000 experiments.

The uniform pseudo random numbers were generated via the multiplicative congruential method discussed and presented by Knuth (1969) and Fishman (1973). Then they are transformed to normal distribution via the method.
\[ E_i = (-2 \cdot (\log (G))^{0.5} \cdot \cos(2 \pi)) \cdot (S) \] where \( G \) and \( S \) are two uniform \((0,1)\) deviates.

\( \hat{\alpha}_i \) is generated by

\[ \hat{\alpha}_i = \alpha_i + E_i \cdot \sigma^2/\lambda_i. \]

The ridge regression coefficients, \( \hat{\alpha}^*_i \), is computed from \( \hat{\alpha}_i \) by

\[ \hat{\alpha}^*_i = \frac{\lambda_i}{\lambda_i + k} \hat{\alpha}_i, \]

where \( k \) takes different values for different ridge estimates. Note that \( \hat{\alpha}_i = \hat{\alpha}^*_i \) for \( k = 0 \) (OLS).
III. EXPERIMENTAL RESULTS VIA MONTE CARLO METHOD

The effect of multicollinearity on the variance of ordinary least squares (OLS) estimates was discussed in chapter 1 and can be observed from the eight models \((\alpha_1-\alpha_8)\) presented in table 1. There are six regression estimators to be discussed and analysed in chapter III. These estimators are the following:

1. Ordinary least square estimator, OLS, is an unbiased estimator and is included in this study for two main reasons:

   a. To compare the performance of some ridge regression estimators with OLS in terms of the observed loss, \(M_1\), defined as

   \[
   M_1 = \sum_{i=1}^{3} \frac{(\hat{\alpha}_i - \alpha_i)^2}{\sigma^2},
   \]

   where \(\hat{\alpha}_i\) is the OLS estimate of the true coefficient \(\alpha_i\), and \(\sigma^2\) is the variance which is assumed to be fixed and known.

   b. To compare the stability and accuracy of the estimate of \(M_1\) by \(\tilde{M}_1\) defined as

   \[
   \tilde{M}_1 = \sum_{i=1}^{P} \frac{m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \hat{\theta}}{\lambda_i + k}
   \]

   where \(m_i = \frac{\lambda_i}{\lambda_i + k}\), \(\hat{\theta}\) is the proposed method of estimating \(\alpha_i\) defined by

   \[
   \hat{\theta} = (\hat{\alpha}_i^2 - V(\hat{\alpha}_i)) \geq 0, \text{ and } \lambda_i \text{ is the } i^{th} \text{ eigenvalue of } (X'X).\]
The measure of the accuracy and stability of the estimated loss function, \( \tilde{M}_1 \), in this chapter is defined by the coefficient of loss variation defined as

\[
C^*/V = \left[ \frac{1}{P} \sum_{i=1}^{P} \left( \tilde{L}^2_1(\hat{\alpha}^*) - L^2_1(\hat{\alpha}^*) \right)^2/n - \frac{1}{P} \sum_{i=1}^{P} \frac{L^2_1(\hat{\alpha}^*)}{n} \right]^{.5}
\]

where \( \tilde{L}^2_1(\hat{\alpha}^*) \) is the estimated squared loss, \( L^2_1(\hat{\alpha}^*) \) is the true observed squared loss, and \( n \) is the number of experiments.

2. Lawless-Wang ridge estimator, \( k_{LW} \), introduced and discussed in earlier chapters.

3. Hoerl, Kennard and Baldwin (HKB) ridge estimators, \( k_{HKB} \).

4. Lawless generalized ridge estimator, \( K_{GRR} \).

5. RIDGN generalized ridge estimator, introduced in chapter II and thought to have some advantages over \( k_{HKB} \) in terms of measuring its estimated \( \tilde{M}_1 \).

6. ROP is a generalized ridge estimator, included in the analysis as a theoretical reference, since it assumes that the values of \( \alpha^2_i \) and \( \sigma^2 \) are known as far as computing the optimal value of the \( k_i \).

Results in table 1 indicate with respect to model \( \alpha_1 = (10 \ 10 \ 10) \) a significant improvement with ridge regression in \( M_1 \) compared to OLS. This improvement should be expected from the use of ridge regression for \( \alpha_1 \) since the known true coefficients are all of the same magnitude. It is well known that simulations with nearly equal magnitudes for the true coefficients
Table 1. Signal-to-noise ratio, HI, estimated ML, Hl, and the coefficient of loss variation for the eight models (a1-a8) with λ = (0.01, 1, 1.99) and \( \sigma^2 = 4 \). OLS vs. ridge estimators: \( k_{LM} \), RIDGN, \( k_{HKB} \), KGB, and ROP.

<table>
<thead>
<tr>
<th>Signal-Noise</th>
<th>S/N^a</th>
<th>( a_1^b )</th>
<th>( a_2^b )</th>
<th>( a_3^b )</th>
<th>( a_4^b )</th>
<th>( a_5^b )</th>
<th>( a_6^b )</th>
<th>( a_7^b )</th>
<th>( a_8^b )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10,10,10)</td>
<td>75.00</td>
<td>100.367</td>
<td>105.330</td>
<td>101.626</td>
<td>99.251</td>
<td>99.727</td>
<td>102.222</td>
<td>103.314</td>
<td>96.977</td>
</tr>
<tr>
<td>(V7.5/3/3/3)</td>
<td>75.00</td>
<td>121.41</td>
<td>42.33</td>
<td>75.00</td>
<td>107.67</td>
<td>14.84</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(( V7.5/33/3/15 ))</td>
<td>46.197</td>
<td>77.908</td>
<td>42.192</td>
<td>37.050</td>
<td>55.710</td>
<td>51.251</td>
<td>36.306</td>
<td>87.198</td>
<td></td>
</tr>
<tr>
<td>(( V7.5/33/3/15 ))</td>
<td>57.084</td>
<td>69.316</td>
<td>51.594</td>
<td>48.126</td>
<td>50.773</td>
<td>60.106</td>
<td>49.159</td>
<td>67.297</td>
<td></td>
</tr>
<tr>
<td>(( V7.5/33/3/15 ))</td>
<td>37.759</td>
<td>43.106</td>
<td>36.023</td>
<td>37.280</td>
<td>40.640</td>
<td>41.124</td>
<td>36.763</td>
<td>44.847</td>
<td></td>
</tr>
<tr>
<td>(( V7.5/33/3/15 ))</td>
<td>39.968</td>
<td>53.347</td>
<td>33.304</td>
<td>30.946</td>
<td>42.525</td>
<td>43.213</td>
<td>31.208</td>
<td>53.430</td>
<td></td>
</tr>
<tr>
<td>(( V7.5/33/3/15 ))</td>
<td>50.224</td>
<td>70.286</td>
<td>52.908</td>
<td>49.415</td>
<td>59.030</td>
<td>61.291</td>
<td>50.555</td>
<td>68.137</td>
<td></td>
</tr>
<tr>
<td>(( V7.5/33/3/15 ))</td>
<td>22.239</td>
<td>42.575</td>
<td>11.335</td>
<td>11.136</td>
<td>29.430</td>
<td>29.239</td>
<td>7.499</td>
<td>45.811</td>
<td></td>
</tr>
</tbody>
</table>

a. Signal-to-noise S/N = \( \frac{1}{n} \sum_{i=1}^{n} a_i^2 / \sigma^2 \).

b. \( H_{LM} = \frac{1}{n} \sum_{i=1}^{n} (a_i - \bar{a})^2 / \sigma^2 \) averaged over 1000 experiments.

c. \( \bar{H}_{LM} = \frac{1}{n} \sum_{i=1}^{n} (a_i^2 - \bar{a}^2) / \sigma^2 \) where \( \bar{a} = (\sum_{i=1}^{n} a_i^2) / n \) averaged over 1000 experiments.

do. \( C^2/V \) - the coefficient of loss variation = \( \left( \frac{1}{n} \sum_{i=1}^{n} (\frac{\bar{a}^2}{\bar{a}^2} - 1)^2 / n - 1 \right) \left( \frac{1}{n} \sum_{i=1}^{n} (\frac{\bar{a}}{\bar{a}}) / n \right)^{0.5} \).
will favor ridge regression (Thisted and Morris, 1977). Results in table 1 reveals that $M_1(\text{OLS})$ is about 4 to 5 times the value of $M_1(k_{LW})$ since the expected reduction in the variance component of $M_1$ more than offsets the squared bias component of $k_{LW}$, which could be approximated as follows:

An approximate $k_{LW}$ value $= \frac{\rho \sigma^2}{\sum \lambda_i \alpha_i^2} = \frac{(3)(4)}{300} = 0.040$,

and the approximate $m_1$ value $= \frac{\lambda_i}{\lambda_1+k} = 0.01 = 0.200$.

Therefore, the variance component of

$$M_1 = \frac{m_1^2 \sigma^2}{\sigma^2 \lambda_1} = \frac{(0.2)^2(4.04)}{(4.04)(0.01)} = 0.16 = 4.04$$

of $\hat{\alpha}_1$. The approximate bias squared component of $M_1$ is equal to

$$(1-m_1)^2 \sigma^2 / \sigma^2 = \frac{(1-0.2)^2}{4} = \frac{0.16}{4} = 0.04.$$

The above calculations illustrate that, although the $k_{LW}$ ridge estimator is a biased estimator, the approximate reduction in $M_1$ more than offsets the increase in $M_1$ due to the bias squared component. Other ridge estimators also showed a substantial reduction in $M_1$ compared to OLS, with $k_{HKB}$ second only to $k_{LW}$, followed by RIDGN and KGRR. $M_1(\text{OLS})$ was estimated by

$$\tilde{M}_1(\text{OLS}) = \frac{3}{\sum \sigma^2 / \lambda_i} / \sigma^2$$

and did not vary among models, since $\sigma^2$ and $\lambda_i$ are fixed known parameters. $M_1$ for
ridge estimators, however, is expected to vary among models and among estimators since the proposed method for estimating \( M_1 \) is

\[
\tilde{M}_1(\hat{\alpha}^*) = \left( \sum_{i=1}^{3} \left( m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \check{\theta} \right) / \sigma^2 \right),
\]

where all the variables are as previously defined. Results in table 1 showed that the proposed method overestimated \( M_1 \) for \( k_{LW} \). However, this overestimation of \( M_1 \) for the Lawless-Wang ridge estimator should not be viewed as being as serious as the problem of underestimating \( M_1 \) for the other ridge estimators as shown in table 1. The results of table 1 showed that the estimate of \( \tilde{M}_1(k_{LW}) \) was fairly good compared to \( \tilde{M}_1(\text{OLS}) \) where the true fixed variance of OLS was known. Before leaving the results for model 1 it should be noted that the \( k_{LW} \) value was usually very close to the optimal value of \( k \) since

\[
k_{LW} = \frac{p \sigma^2}{\sum \lambda_i \hat{\alpha}_i} = \sigma^2 / \alpha_i^2 \text{ when } \alpha_1^2 = \alpha_2^2 = \alpha_3^2.
\]

Therefore, \( M_1(k_{LW}) \) was about the same magnitude as \( M_1(\text{ROP}) \), which demonstrates a high performance by the \( k_{LW} \) estimator for the case of \( \alpha_1 = (10, 10, 10) \).

For model two, \( \alpha_2 = (15 \sqrt{37.5} \sqrt{37.5}) \). Results in table 1 revealed that \( M_1(\text{OLS}) \) was more than 2 times the value of \( M_1(k_{LW}) \), and this significant improvement was possible because of the greater reduction in the variance component of \( M_1(k_{LW}) \) compared to the OLS variance. Results in table 1 also revealed a substantial
improvement compared to OLS for all the other ridge estimators with $k_{HKB}$ performing next best to $k_{LW}$, followed by RIDGN and KGRR. $\tilde{M}(OLS)$, as expected, was more stable in terms of $C*/V$ than $\tilde{M}(k_{LW})$. $\tilde{M}(k_{LW})$ was overestimated, $M(k_{LW})$ being about 63% of estimated $\tilde{M}(k_{LW})$. However, these overestimates of $\tilde{M}$ are not considered by the author to be very serious for applied ridge regression analysis. On the other hand, the underestimation of $M$ by the proposed estimated loss function for the other ridge estimators is considered to be a serious problem, and prior information would be needed before applying it to these other ridge estimators. It should be noted that a theoretical improvement, although not realizable in practice since the true regression coefficients would not be known, of up to 250% over OLS was shown by the ROP performance. However, the Lawless-Wang ridge estimator gave results that were surprisingly close to ROP in table 1.

For model 3 with $\alpha_3 = (\sqrt{37.5}, 15, \sqrt{37.5})$, results in table 1 were favorable for all the ridge estimators. This favorable result was expected since the smallest coefficient ($\alpha = \sqrt{37.5}$) is associated with the smallest eigenvalue $\lambda_1 (\lambda_1 = .01)$. Significant improvement over OLS can be expected from the use of ridge regression in such cases, and from the Lawless-Wang ridge estimator in particular (Lawless, 1978), as can be seen from
$$M_1 = \sum_{i=1}^{3} \left[ \frac{\lambda_i \sigma^2}{(\lambda_i + k)^2} + \left( \frac{k}{\lambda_i + k} \right)^2 \hat{\sigma}_1^2 \right].$$  As noted earlier, the $k_{LW}$ estimator for $\alpha_3$ had an estimated $k$ value almost equal to the optimal $k$ value given by ROP, resulting in an observed $M_1(k_{LW})$ nearly equal to $M_1(ROP)$, $(M_1(k_{LW}) = 11.84, M_1(ROP) = 11.34)$.

Results in table 1 for this model also show that $M_1(OLS)$ is about 8.6 times the value of $M_1(k_{LW})$. Other ridge estimators in this model performed well, although not as well as $k_{LW}$ because the estimated values of $k$ by these estimators were not very close to the optimal $k$ (ROP), which minimizes $L_1(\alpha^*)$. The proposed method of estimating $M_1$ overestimated $M_1(k_{LW})$. This overestimation might be because of the following:

1. Overestimation of the value of $\alpha_1^2$ in the expression for $\tilde{M}_1$,

$$\tilde{M}_1 = \sum_{i=1}^{3} \left( \frac{m_i \sigma^2}{\lambda_i} + (1-m_i)^2 \hat{\theta}_i \right) / \sigma^2,$$

where $\hat{\theta}_i = (\hat{\alpha}_i^2 - V(\hat{\alpha}_i)) \geq 0$ is an estimate of $\alpha_1^2$.

2. The value of $M_1(k_{LW})$ in this model was very small, and any overestimation, although small in magnitude, is large in proportion to the observed $M_1(k_{LW})$. The proposed method for estimating squared error loss underestimated $M_1$ for all the other ridge estimators. In this Monte Carlo experiments, RIDGN was the least underestimated and KGRR was the most underestimated. It should be emphasized that the proposed estimate of
squared error, by underestimating $\tilde{M}_1(k_{\text{HKB}})$, $\tilde{M}_1(\text{RIDGN})$, and $\tilde{M}_1(\text{KGRR})$, may be dangerously misleading to the researcher in some critical cases where the true $M_1$ may be much greater than $M_1(\text{OLS})$, as pointed out earlier. Note that although the observed $M_1(\text{RIDGN})$ was about the same magnitude as $M_1(\text{KGRR})$, $\tilde{M}_1(\text{RIDGN})$ was closer to the observed $M_1(\text{RIDGN})$ than $\tilde{M}_1(\text{KGRR})$ to the observed $M_1(\text{KGRR})$, thereby giving RIDGN some advantage over KGRR for the analysis of sample data. (However, the Lawless-Wang ridge estimator had even more advantage in this respect.)

Model 4, $\alpha_4 = (\sqrt{37.5}, \sqrt{37.5}, 15)$, was also an appropriate model for ridge regression in general, and for the $k_{\text{LW}}$ ridge estimator in particular, for the same reasons given for model 3. The value of $k$ given by the $k_{\text{LW}}$ estimator was closer to the optimal $k$ than for any of the other ridge estimators for this model. $M_1(\text{OLS})$ was about 6.9 times the value of $M_1(k_{\text{LW}})$ in table 1, and $k_{\text{LW}}$ outperformed all the other ridge estimators for all models studied in this simulation in table 1, except for $\alpha_8$. The HKB ridge estimator had the second lowest observed squared error, followed by RIDGN and KGRR.

Again, as for $\alpha_1$, $\alpha_2$, and $\alpha_3$, the proposed method for estimating squared error, $M_1(\hat{\alpha}_1^*)$, overestimated $M_1$ for $k_{\text{LW}}$ and gave unstable and less accurate estimates
than OLS, as measured by C*/V (C*/V for OLS = .083 while C*/V for k\textsubscript{LW} = .283). This apparent instability appears to be due to the overestimation of \( \alpha_1^2 \) and the unusually low squared error from k\textsubscript{LW} for the \( \alpha_4 \) vector. It is shown in table 1 that M1 for all the other ridge estimators were underestimated, but not as much so as for models 2 or 3. RIDGN in this model, and for all the \( \alpha \) vectors in table 1, out performed KGRRR in terms of both M1 and estimated value for M1. Since the value of \( \alpha_1 \) (the most effected by multicollinearity since \( \lambda_1 = .01 \)) did not change from model 3 to model 4, ROP gave a value of M1 similar to M1 given by model 3.

Model 5, \( \alpha_5 = (12, 12, \sqrt{12}) \), is generally less suited for ridge regression since one of the largest \( \alpha_i \) (\( \alpha_1 = 12 \)) corresponds to the smallest eigenvalue (\( \lambda_1 = .01 \)). However, for this model, all ridge estimators in table 1 had M1 values significantly less than M1(OLS), and an improvement of up to 337% was possible by the use of the optimal value for k, the one given by ROP. A substantial gain in M1 compared to OLS, was given by k\textsubscript{LW} followed by k\textsubscript{HKB}, RIDGN, and KGRR. The proposed method for estimating squared error overestimated M1(k\textsubscript{LW}) for the same reasons given earlier, but in some respects it is safe to use since it overestimates M1(k\textsubscript{LW}) rather than under-estimating it. On the other hand, the proposed
approximation of squared error understates method Ml for k_{HKB}, RIDGN, and KGRR, making them seem misleadingly better than their actual performance as measured by their observed squared error, Ml. Therefore, for this model, a great deal of caution should be exercised with regard to the use of the squared error estimator, except for k_{LW} type estimator.

Model 6, α_5 = (12, √12, 12) is somewhat similar to model 5, characterized by the fact that one of the largest α_i (α_1 = 12) is associated with the smallest eigenvalue (λ_1 = .01). Table 1 showed results for model 6 to be very similar to the results of model 5 except for some improvement in the performance of the ridge estimators as compared to α_5 = (12, 12, √12), since the smallest coefficient (α_2 = √12) is now associated with the second smallest eigenvalue of (X'X), (λ_2 = 1.), rather than with the largest eigenvalue, as for α_5. This fact seems to cause improvement in both Ml and the estimated Ml for all ridge estimators. There were no further major differences worth mentioning between the results of model 5 and model 6, except for the observations just noted.

Model 7, α_7 = (√12, 12, 12), is very appropriate for ridge analysis in general, and for k_{LW} type ridge in particular since the smallest coefficient (α_1 = √12) corresponds to the smallest eigenvalue of (X'X).
\( \lambda_1 = .01 \). ROP, although unusable in practice, showed a remarkable improvement in \( M_l \) compared to OLS, and although this improvement cannot be fully realized in practice, some of this improvement is possible by the use of ridge regression analysis. The \( k_{LW} \) estimator gave a \( k \) value which is very close to the optimal ROP \( k \) value, resulting in an observed \( M_l(k_{LW}) \) that is very close to \( M_l(ROP) \). Next to \( k_{LW} \) in performance was \( k_{HKB} \), followed by RIDGN and KGRR. For the same reasons given earlier, the proposed method for predicting square error overestimated \( M_l \) for \( k_{LW} \) and underestimated it for all the other ridge estimators (\( k_{HKB} \), RIDGN, and KGRR).

Model 8, \( \alpha_8 = (16, \sqrt{22}, \sqrt{22}) \), is similar to model 2 except that it is an even more unfavorable case. The largest coefficient (\( \alpha_1 = 16 \)) is associated with the smallest eigenvalue (\( \lambda_1 = .01 \)). Results in table 1 show that with respect to \( M_l \), \( k_{LW} \) gave the most gain compared to OLS, followed by \( k_{HKB} \), RIDGN, and KGRR. The proposed \( M_l \) method overestimated \( M_l(k_{LW}) \) but the estimate, \( \tilde{M}_l(k_{LW}) \), was more stable for \( \alpha_8 \) than for the other \( \alpha_i \)’s in table 1 (stability measured by \( C^*/V \)). \( \tilde{M}_l \) seriously underestimated \( M_l \) for the other ridge estimators studied in this experiment.

Results of the Monte Carlo experiments for the same models and the same estimators presented in table 1
<table>
<thead>
<tr>
<th></th>
<th>(a_1^m)</th>
<th>(a_2^m)</th>
<th>(a_3^m)</th>
<th>(a_4^m)</th>
<th>(a_5^m)</th>
<th>(a_6^m)</th>
<th>(a_7^m)</th>
<th>(a_8^m)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(10,10,10)</td>
<td>(15,37.5,37.5)</td>
<td>(25,37.5,37.5,15)</td>
<td>(12,12,12)</td>
<td>(12,12,12)</td>
<td>(12,12,12)</td>
<td>(16,22,22)</td>
<td></td>
</tr>
<tr>
<td><strong>Signal-Noise</strong></td>
<td><strong>S/N</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>OLS</strong></td>
<td>M1^b</td>
<td>100.367</td>
<td>105.330</td>
<td>101.626</td>
<td>99.251</td>
<td>99.727</td>
<td>102.222</td>
<td>103.314</td>
</tr>
<tr>
<td></td>
<td>C*/V^d</td>
<td>.064</td>
<td>.061</td>
<td>.062</td>
<td>.062</td>
<td>.061</td>
<td>.063</td>
<td>.063</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>49.142</td>
<td>100.831</td>
<td>41.785</td>
<td>40.620</td>
<td>64.753</td>
<td>57.698</td>
<td>37.416</td>
</tr>
<tr>
<td></td>
<td>C*/V</td>
<td>.112</td>
<td>.090</td>
<td>.139</td>
<td>.076</td>
<td>.109</td>
<td>.109</td>
<td>.070</td>
</tr>
<tr>
<td><strong>RIDG</strong></td>
<td>M1</td>
<td>66.048</td>
<td>85.605</td>
<td>55.218</td>
<td>51.695</td>
<td>70.036</td>
<td>71.424</td>
<td>50.283</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>41.213</td>
<td>49.589</td>
<td>38.265</td>
<td>38.500</td>
<td>45.355</td>
<td>45.355</td>
<td>37.232</td>
</tr>
<tr>
<td><strong>k^HKR</strong></td>
<td>M1</td>
<td>51.847</td>
<td>75.280</td>
<td>40.759</td>
<td>38.633</td>
<td>57.130</td>
<td>57.852</td>
<td>37.373</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>30.365</td>
<td>38.093</td>
<td>29.560</td>
<td>29.093</td>
<td>33.136</td>
<td>33.990</td>
<td>28.677</td>
</tr>
<tr>
<td><strong>KGR</strong></td>
<td>M1</td>
<td>66.972</td>
<td>86.162</td>
<td>56.453</td>
<td>52.902</td>
<td>70.813</td>
<td>72.254</td>
<td>51.653</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>29.094</td>
<td>38.453</td>
<td>26.032</td>
<td>25.914</td>
<td>33.208</td>
<td>33.865</td>
<td>24.575</td>
</tr>
<tr>
<td><strong>ROP</strong></td>
<td>M1</td>
<td>36.348</td>
<td>64.869</td>
<td>20.630</td>
<td>20.300</td>
<td>46.321</td>
<td>46.097</td>
<td>15.472</td>
</tr>
</tbody>
</table>

a. Signal-to-noise S/N = \(\sum_{i=1}^{p} \frac{\lambda_i a_i^2}{\sigma^2}\).

b. M1 = \(\frac{\sum_{i=1}^{3} (a_i - \hat{a}_i)^2}{\sigma^2}\) averaged over 1000 experiments.

c. \(\hat{M}_1 = \frac{\sum_{i=1}^{3} (M_i^2 - \lambda_1 + (1 - \lambda_1)^2 \hat{\lambda}_4)}{\sigma^2}\) where \(\hat{\lambda}_4 = (\hat{\sigma}^2 - m_1) \geq 0\), averaged over 1000 experiments.

d. C*/V = the coefficients of variation = \(\sqrt{\frac{\sum_{i=1}^{3} (\frac{L_i}{L_i^2})}{\sum_{i=1}^{3} L_i L_i^0}}\).
are also presented in table 2, with the only difference being that $\sigma^2 = 2$ in table 2 versus $\sigma^2 = 4$ in table 1. $\text{ML}(k_{LW})$ for model $\alpha_1 = (10, 10, 10)$ still shows a substantial improvement compared to $\text{ML}(\text{OLS})$ as the signal-to-noise ratio, defined as $S/N = \sum \alpha_i^2 \lambda_i / \sigma^2$, is doubled. However, the value of $\text{ML}(k_{LW})$ is increased due to the decrease in the value of $\sigma^2$. This increase in the value of $\text{ML}(k_{LW})$ was expected and can be shown by the formula for computing $\text{ML}(k_{LW})$. Since $\text{ML}(k_{LW}) = \sum_{i=1}^{p} \left[ m_i \sigma^2 / \sigma^2 \lambda_i + (1 - m_i) \alpha_i^2 / \sigma^2 \right]$, the variance component of $\text{ML}$ will not change since $\sigma^2$ in the denominator and the numerator will cancel out, but the bias squared component of $\text{ML}$ will increase for a given value of $k$ as $\sigma^2$ decreases from 4 (in table 1) to 2. The above reasoning is also true for the other ridge estimators ($k_{HKB}$, RIDGN, and KGRR). The second observation from the results of model $\alpha_1$ in table 2 is that an improvement can be observed in the accuracy and the stability of the proposed method for estimating $\text{ML}$ for $k_{LW}$. $\text{ML}(k_{LW})$ in table 2 was closer to $\text{ML}(k_{LW})$ than that observed from table 1. The coefficient of loss variation for $k_{LW}, C*/V$, was .112 compared to $C*/V$ of .278 presented in table 1. Despite an overstatement of $\text{ML}(k_{LW})$ by $\text{ML}(k_{LW})$, some improvement relative to OLS is evident from table 2. The estimate of $\text{ML}$ for the other ridge estimators was again understated, but
the value of $\bar{M}_1$ for these estimators was somewhat closer to the observed squared loss $M_1(\hat{\alpha}^*)$ than observed in table 1. Before leaving model $\alpha_1$, the remarkable performance of the $k_{LW}$ ridge estimator should be noted since observed $M_1(k_{LW})$ was almost equal to $M_1(\text{ROP})$.

Model $\alpha_2 = (15, \sqrt{37.5}, \sqrt{37.5})$ in table 2 also showed a significant increase in $M_1(k_{LW})$ compared to $M_1(k_{LW})$ of table 1, due to the same reasons just given for $\alpha_1$. However, $M_1(k_{LW})$ was still somewhat smaller than $M_1(\text{OLS})$ for this rather unfavorable model for ridge regression. The ridge estimator, $k_{HKB}$, gave an observed squared error loss $M_1(k_{HKB})$ which was somewhat less than $M_1(k_{LW})$. RIDGN and KGRR both gave an observed $M_1$ almost the same as $M_1(k_{LW})$. The proposed method of estimating $M_1$ for model $\alpha_2$ also improved since $M_1(k_{LW})$ was closer to the observed $M_1(k_{LW})$ as compared to table 1, with only a slight overestimation. The stability and accuracy of $\bar{M}_1(k_{LW})$ in estimating $M_1$ also improved considerably, comparing table 1 to table 2, since $C*/V = .098$ compared to $C*/V$ for OLS of .061. The proposed method of estimating $M_1$ understated $M_1$ for all the other ridge estimators, with RIDGN being the least underestimated and KGRR was the most underestimated. The $M_1(k_{LW})$ for this model was not as close to $M_1(\text{ROP})$ as for model $\alpha_1$ indicating that the $k_{LW}$ value
of k may be far from the optimal k value, as could be expected since \( a_2 \) was the second most unfavorable vector for ridge regression.

Model \( a_3 \) of table 2 is one of the favorable models for ridge analysis, as noted earlier, since the smallest coefficient \( a_1 \) is associated with the smallest eigenvalue of \((X'X)\). Results in table 2 indicated that \( M_1(\text{OLS}) \) is about 5 times larger than \( M_1(k_{\text{LW}}) \). \( M_1(k_{\text{LW}}) \), although increased due to a smaller value of \( \sigma^2 \), still outperformed \( M_1(\text{OLS}) \) by a ratio of 5:1. \( M_1(k_{\text{LW}}) \) is also smaller than the \( M_1 \) for all the other ridge estimators (\( k_{\text{HKB}}, \text{RIDGN}, \) and \( \text{KGRR} \)). The proposed method for estimating \( M_1(k_{\text{LW}}) \) showed an improvement compared to the results of model \( a_3 \) in table 1. The measure of accuracy and stability, \( C*/V \) as defined earlier, showed an improvement since \( C*/V \) in table 1 was about 3 times larger than \( C*/V \) for \( k_{\text{LW}} \) in table 2. The proposed method of estimating the observed squared error underestimated \( M_1 \) for all the other ridge estimators in table 2. However, \( M_1 \) for RIDGN was the closest to the observed squared error loss, \( M_1(\text{RIDGN}) \), followed by \( k_{\text{HKB}} \) and \( \text{KGRR} \). The value of k obtained by the Lawless-Wang method was almost equal to the value of k obtained by the theoretical estimator, ROP; therefore, \( M_1(k_{\text{LW}}) \) is almost equal to \( M_1(\text{ROP}) \).

Model \( a_4 = (\sqrt{37.5}, \sqrt{37.5}, 15) \) in table 2 is a more
favorable model for ridge analysis than model $a_3$, since $a_2 = \sqrt{37.5}$ is associated with the second smallest eigenvalue ($\lambda_2 = 1.0$). However, the observed squared error loss, $M_1(k_{LW})$, increased due to the smaller value of $\sigma^2$ in table 2 as compared to $a_4$ in table 1. This increase in the value of $M_1(k_{LW})$ is attributed mainly to the increase in the bias squared component of $M_1(k_{LW})$, as was explained for model $a_3$. Also, $M_1(k_{LW})$ was more stable and accurate in terms of $C*/V$ as compared to the value of $C*/V$ for model $a_3$. On the other hand $M_1$ for the other ridge estimators was underestimated with $\tilde{M}_1(k_{HKB})$ second only to $\tilde{M}_1(k_{LW})$ in performance, followed by RIDGN and KGRR. It should be noted that the value of $k$ for the $k_{LW}$ ridge estimator was not very close to the value of optimal $k$; therefore, $M_1(k_{LW})$ was larger in value than $M_1(ROP)$.

Model $a_5 = (12 \ 12 \ \sqrt{12})$ in table 2 is generally less appropriate for ridge analysis as compared to model $a_1$ or model $a_3$. Nevertheless, for this model, $M_1(OLS)$ was still 196 percent that of $M_1(k_{LW})$, since the variance component of $M_1(k_{LW})$ was only about 4.8% of the variance of $a_1$, which more than offsets the increase in $M_1$ of about 43.8% due to the bias squared component of $M_1(k_{LW})$. $M_1(k_{LW})$ was significantly lower than the other ridge estimators, ($k_{HKB}$, RIDGN, and KGRR). The proposed method of estimating $M_1(k_{LW})$ again showed some
improvement compared to the results in table 1, giving an overestimation of only about 20%. The $\tilde{M}_1$ estimates were fairly stable since the coefficient of loss variation $C^*/V$ was about .109 compared to $C^*/V$ of .234 for the same model in table 1. $\tilde{M}_1$(RIDGN) was closer to $M_1$(RIDGN) than $M_1$(HKB) to $M_1$(HKB) or $M_1$(KGR) to $M_1$(KGR). It is clear from the results of the Monte Carlo experiments in table 2 for $\alpha_5$ that the Lawless-Wang estimator had an estimated $k$ value closer to the optimal $k$ value than did the other adaptive ridge estimators. Hence, the $M_1(k_{LW})$ value was closer to $M_1$(ROP) than for any of the other ridge estimators.

Model $\alpha_6 = (12, \sqrt{12}, 12)$ in table 2 like $\alpha_5$ was less appropriate for ridge analysis than models $\alpha_1$, $\alpha_3$ or $\alpha_4$, since the largest $\alpha_i$ is associated with the smallest eigenvalues of $(X'X)$. However, under this rather unfavorable condition for the Lawless-Wang ridge estimator, a remarkable improvement over OLS is reported for $k_{LW}$ and the other ridge estimators in this Monte Carlo experiment. The proposed method of estimating $M_1(k_{LW})$ showed an improvement over the results for model $\alpha_6$ reported in table 1, although still showing an overestimation of about 1.2% of the observed $M_1(k_{LW})$.

The proposed method of estimating the observed squared error loss again understated $M_1$ for the other ridge estimators ($k_{HKB}$, RIDGN, and KGR). $\tilde{M}_1$(RIDGN)
was closer to the observed squared error loss for RIDGN, M1(RIDGN), than both M1(k_{HKB}) to M1(k_{HKB}) or M1(KGRR) to the observed M1(KGRR). For model a_6, the estimated k_{LW} value by the Lawless-Wang method was very close to the optimal k value; therefore, M1(k_{LW}) was almost equal to M1(ROP) with M1(k_{LW}) = 46.45 and M1(ROP) = 46.1.

As was the case in table 1, model a_7 = (\sqrt{12}, 12, 12) in table 2 was the most favorable model for ridge regression analysis as indicated by the results for the theoretical ROP ridge estimator. It should be noted that in model a_7, M1(ROP) was the smallest of any of the models. M1(OLS) was about 5 times the value of M1(k_{LW}), but all the ridge estimators for model a_7 outperformed OLS in terms of the observed squared error loss, M1, due to the substantial reduction in the variance of the ridge estimates, V(\hat{\alpha}_i^*), defined earlier as, 

\[ V(\hat{\alpha}_i^*) = \sigma^2/\lambda_i/(\lambda_i+k)^2 = m_i^2\sigma^2/\lambda_i. \]

Therefore, M1 for the ridge estimators are lower than M1(OLS) as long as the reduction in the variance of \( \hat{\alpha}_i^* \) more than offsets the increase in M1 due to the bias squared component of M1. The proposed method of estimating M1(k_{LW}) showed some improvement in stability as compared to the results for a_7 in table 1. The coefficient of loss variation, C*/V, for M1(k_{LW}) was 0.07 compared to C*/V = .061 for M1(OLS). The proposed method of estimating M1 underestimated M1 for all the other ridge estimators;
However, the estimated $M_l(k_{HKB})$ was very close to $M_l(k_{HKB})$ despite the slight underestimation shown in table 2. It should be noted here that the observed squared error loss for $M_l(ROP)$ was considerably lower than $M_l(k_{LW})$, implying that the value of $k$ obtained by the Lawless-Wang method was somewhat different from the optimal $k$ value.

Reducing $\sigma^2$ from 4 in table 1 to $\sigma^2 = 2$ in table 2 seems to be crucial for model $a_8$ because $a_8$ is the most unfavorable case for ridge analysis in these experiments. Results in table 2 for model $a_8$ showed $M_l(OLS)$ to be about 94% of the value of $M_l(k_{LW})$. This result is not surprising because the reduction in the variance component of $M_l$ to about .81 was associated with an approximate increase in the bias squared component of $M_l(k_{LW})$ to about 105.9. Therefore, $M_l(k_{LW})$ did not show any improvement over OLS; in fact the Lawless-Wang estimator performed slightly worse than OLS.

The results in table 2 also showed that the HKB method of estimating $k$ by $k_{HKB} = \hat{\sigma}^2 / \hat{\beta}'\hat{\beta}$ always outperformed OLS. For the case of $a_8$, the HKB method also outperformed all the other ridge estimators, including $k_{LW}$. RIDGN and KGRR both gave an observed squared error loss which was significantly smaller than $M_l(OLS)$. The value of $M_l(k_{HKB})$ was also not too much larger than $M_l(ROP)$ for $a_8$. Note that for
model $\alpha_8$, as well as for the other models in table 2, the proposed method for estimating $M_l$ showed an improvement despite an overestimation of $M_l(k_{LW}')$ by about 19%. The stability and the accuracy of the estimated $M_l(k_{LW}')$ measured by $C*/V$ was fairly stable since $C*/V$ for $k_{LW}' = 0.09$ compares favorably with $C*/V = 0.07$ for OLS if one takes into consideration that $M_l(OLS)$ was computed with a known fixed variance, $\sigma^2$. 

Note that the proposed method of estimating squared error loss underestimated $M_l$ for all the other ridge estimator. However, this underestimation was not quite as bad as was shown in table 1 for model $\alpha_8$, especially for $k_{HKB}$ and RIDGN.

The Monte Carlo results for the same 8 models ($\alpha_1-\alpha_8$) with $\sigma^2 = 1$ are shown in table 3. Since the $\sigma^2$ value in table 3 is equal to 1, the Monte Carlo results showed an increase in the value of the observed squared error loss for the ridge estimators due to the increase in the bias squared component of $M_l$, as discussed earlier. For models $\alpha_1$, $\alpha_3$, $\alpha_4$, $\alpha_5$, $\alpha_6$ and $\alpha_7$, the ridge estimators still performed substantially better than OLS in terms of $M_l$. $M_l(k_{LW}')$ was smaller than $M_l$ for all the other ridge estimators for these six models. The value of $k$ obtained by the Lawless-Wang method was very close to the optimal $k$ for models $\alpha_1$, $\alpha_3$ and $\alpha_6$; therefore $M_l(k_{LW}')$ was very close to
Table 3. Signal-to-noise ratio, $\text{S/N}$, estimated $M_l$, $\tilde{M}_l$ and the coefficients of loss variation for the eight models $(a_1^*=a_8^*)$, with $\Lambda = (0.01, \ldots, 1.99)$, $\sigma^2 = 1$. OLS vs. ridge estimators: $k_{LM}$, RIDGN, $k_{HKB}$, KGRR and ROP.

<table>
<thead>
<tr>
<th></th>
<th>$a_1^*$</th>
<th>$a_2^*$</th>
<th>$a_3^*$</th>
<th>$a_4^*$</th>
<th>$a_5^*$</th>
<th>$a_6^*$</th>
<th>$a_7^*$</th>
<th>$a_8^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(10,10,10)</td>
<td>(15/37.5/37.5)</td>
<td>(37.5/15/37.5)</td>
<td>(37.5/37.5/15)</td>
<td>(12,12/12)</td>
<td>(12/12,12)</td>
<td>(12/12,12)</td>
<td>(16/37/37)</td>
</tr>
<tr>
<td>Signal-Noise $\text{S/N}^a$</td>
<td>300.00</td>
<td>114.375</td>
<td>300.00</td>
<td>485.625</td>
<td>169.320</td>
<td>300.00</td>
<td>430.680</td>
<td>59.34</td>
</tr>
<tr>
<td>OLS</td>
<td>$M_l^b$</td>
<td>100.367</td>
<td>105.330</td>
<td>101.626</td>
<td>99.251</td>
<td>99.727</td>
<td>102.222</td>
<td>103.314</td>
</tr>
<tr>
<td></td>
<td>$C^{*}/V$</td>
<td>0.045</td>
<td>0.045</td>
<td>0.044</td>
<td>0.044</td>
<td>0.043</td>
<td>0.043</td>
<td>0.043</td>
</tr>
<tr>
<td>$k_{LM}$</td>
<td>$M_l$</td>
<td>54.265</td>
<td>129.045</td>
<td>35.755</td>
<td>44.066</td>
<td>76.068</td>
<td>65.366</td>
<td>39.260</td>
</tr>
<tr>
<td></td>
<td>$M_l$</td>
<td>56.725</td>
<td>131.057</td>
<td>46.662</td>
<td>51.250</td>
<td>78.277</td>
<td>66.707</td>
<td>46.318</td>
</tr>
<tr>
<td></td>
<td>$C^{*}/V$</td>
<td>0.054</td>
<td>0.052</td>
<td>0.051</td>
<td>0.036</td>
<td>0.058</td>
<td>0.055</td>
<td>0.033</td>
</tr>
<tr>
<td>RIDGN</td>
<td>$M_l$</td>
<td>81.154</td>
<td>108.722</td>
<td>62.073</td>
<td>58.583</td>
<td>88.102</td>
<td>89.071</td>
<td>52.565</td>
</tr>
<tr>
<td></td>
<td>$M_l$</td>
<td>47.655</td>
<td>60.216</td>
<td>41.068</td>
<td>40.073</td>
<td>53.427</td>
<td>53.985</td>
<td>38.132</td>
</tr>
<tr>
<td>$k_{HKB}$</td>
<td>$M_l$</td>
<td>67.983</td>
<td>104.803</td>
<td>51.926</td>
<td>50.152</td>
<td>76.716</td>
<td>77.087</td>
<td>47.321</td>
</tr>
<tr>
<td>KGRR</td>
<td>$M_l$</td>
<td>41.866</td>
<td>52.700</td>
<td>40.053</td>
<td>39.726</td>
<td>46.045</td>
<td>46.640</td>
<td>38.781</td>
</tr>
<tr>
<td>ROP</td>
<td>$M_l$</td>
<td>81.683</td>
<td>108.806</td>
<td>63.157</td>
<td>59.635</td>
<td>88.394</td>
<td>89.417</td>
<td>53.884</td>
</tr>
<tr>
<td></td>
<td>$M_l$</td>
<td>35.661</td>
<td>50.558</td>
<td>28.979</td>
<td>28.734</td>
<td>42.519</td>
<td>43.086</td>
<td>25.337</td>
</tr>
</tbody>
</table>

a. Signal-to-noise $\text{S/N} = \frac{\sum_{i=1}^{p} \sigma_i^2}{n}^{1/2}/n^{2}$.

b. $M_l = \sum_{i=1}^{3} \frac{(\bar{a}_i - a_i)^2}{\sigma^2}$ averaged over 1000 experiments.

c. $M_l = \sum_{i=1}^{3} \frac{(m_{i1}^2 - \hat{m}_{i1}^2)^2/\sigma_i^2 (1-m_{i1}^2)^2}{\hat{\theta}_i}$ where $\hat{\theta}_i = (\hat{m}_i^2 - V(\hat{m}_i))/\sigma_i$, averaged over 1000 experiments.

d. $C^{*}/V$ = the coefficients of variation $= \frac{\sum_{i=1}^{3} (\bar{a}_i - a_i)^2}{n-1/\sum_{i=1}^{3} \bar{a}_i^2/n} 0.5$. 

M1(ROP) for these models. Results in table 3 also showed that ridge regression analysis was not very appropriate for models \( a_2 = (15, \sqrt{37.5}, \sqrt{37.5}) \), and model \( a_8 = (16, \sqrt{22}, \sqrt{22}) \). For these two models and \( \sigma^2 = 1 \), the ridge regression estimators gave an observed squared error loss significantly larger than M1(OLS) due to the significant increase in the bias squared component of M1. Although ROP performance for these models (\( a_2 \) and \( a_8 \)) showed a possible theoretical improvement compared to M1(OLS), none of the adaptive ridge regression estimators showed any improvement relative to OLS. For these models, however, \( k_{\text{HKB}} \) showed the best performance of the ridge estimators, followed by RIDGN, KGRR, and \( k_{\text{LW}} \). Results in table 3 indicate that the proposed method of estimating observed squared error loss, M1, performed very well, overestimating the observed M1(\( k_{\text{LW}} \)) by only about 2 to 8% for all the models, including \( a_2 \) and \( a_8 \). The estimated M1 for all the other ridge estimators were not very close to the observed M1, due to the underestimation problem, especially for the KGRR ridge estimator and the RIDGN estimator. The accuracy and stability of \( \tilde{M}_1(k_{\text{LW}}) \), measured by the coefficients of loss variation \( C^*/V \) for models \( a_4 \) and \( a_7 \) were smaller than \( C^*/V \) reported for OLS. Judging
from the performance of ROP, a significant improvement compared to M1 (OLS) was always possible, although this potential improvement was not realized by any of the ridge estimators for models \( a_2 \) and \( a_8 \), indicating a need for further improvement in the methods for selecting the biasing parameter, \( k \), for such unfavorable situations for ridge regressions.

The results of the Monte Carlo experiments for the 8 models, \( a_1 - a_8 \) with \( \lambda = (0.01, 1, 1.99) \) and \( \sigma^2 = .5 \), are shown in table 4. Results in this table reveal some increase in the value of M1 for all the ridge estimators because of an increase in the bias squared component of M1, \( 1/\sigma^2 (1-m_1)^2 \alpha_1^2 \), resulting from a decrease in the variance, \( \sigma^2 \). The \( k_{LW} \) and \( k_{HKB} \) ridge estimators showed some improvement in M1 compared to OLS for models \( a_1, a_3, a_4, a_5, a_6, \) and \( a_7 \) with the most improvement shown by the most favorable models for ridge analysis: \( a_1, a_3, a_4, \) and \( a_7 \). On the other hand, RIDGN and KGRR showed favorable performances only for the models \( a_3 \) and \( a_7 \) because these estimators apparently gave \( k \) values which deviated from the optimal \( k \) value for the other models (\( a_1, a_2, a_4, a_5, a_6 \) and \( a_8 \)). It should be noted from the results in table 4 that the ridge estimators performed rather poorly compared to OLS for models \( a_2 = (15, \sqrt{37.5}, \sqrt{37.5}) \) and model \( a_8 = (16, \sqrt{22}, \sqrt{22}) \). At \( \sigma^2 \) value of .5 for these
Table 4. Signal-to-noise MI, estimated MI, \( \bar{M}_1 \) and the coefficients of loan variation for the eight models with \( \lambda = (0.01, 1.0, 1.99) \), \( \sigma^2 = .5 \). OLS vs. ridge estimators: \( k_{LM}, RIGDN, k_{KRB}, KGRR \) and ROP.

<table>
<thead>
<tr>
<th>Signal-Noise</th>
<th>S/N</th>
<th>( a_1 )</th>
<th>( a_2 )</th>
<th>( a_3 )</th>
<th>( a_4 )</th>
<th>( a_5 )</th>
<th>( a_6 )</th>
<th>( a_7 )</th>
<th>( a_8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \bar{M}_1 )</td>
<td>100.367</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
</tr>
<tr>
<td>( \bar{M}_2 )</td>
<td>101.626</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
<td>101.503</td>
</tr>
<tr>
<td>( \bar{M}_3 )</td>
<td>60.000</td>
<td>228.75</td>
<td>600.00</td>
<td>971.250</td>
<td>338.640</td>
<td>600.00</td>
<td>861.36</td>
<td>118.68</td>
<td>96.977</td>
</tr>
<tr>
<td>( \bar{M}_4 )</td>
<td>338.640</td>
<td>600.00</td>
<td>971.250</td>
<td>338.640</td>
<td>600.00</td>
<td>861.36</td>
<td>118.68</td>
<td>96.977</td>
<td>103.314</td>
</tr>
<tr>
<td>( \bar{M}_5 )</td>
<td>600.00</td>
<td>971.250</td>
<td>338.640</td>
<td>600.00</td>
<td>861.36</td>
<td>118.68</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
</tr>
<tr>
<td>( \bar{M}_6 )</td>
<td>861.36</td>
<td>118.68</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
</tr>
<tr>
<td>( \bar{M}_7 )</td>
<td>118.68</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
</tr>
<tr>
<td>( \bar{M}_8 )</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
<td>103.314</td>
<td>96.977</td>
</tr>
</tbody>
</table>

- **a.** Signal-to-noise \( S/N = \frac{P}{\lambda} \sum_{i=1}^{3} a_i^2 / \sigma^2 \).
- **b.** \( \bar{M}_1 = \sum_{i=1}^{3} (a_i - q) / \sigma^2 \) averaged over 1000 experiments.
- **c.** \( \bar{M}_2 = \sum_{i=1}^{3} (m_i - q)^2 / \lambda_i + (1-m_i)^2 / \sigma^2 \) where \( \bar{M}_2 = (m_i - q)^2 / \lambda_i + (1-m_i)^2 / \sigma^2 \) averaged over 1000 experiments.
- **d.** \( C^* / V = \text{the coefficients of loan variation} = \left( \sum_{i=1}^{3} (1/ \lambda_i) \right)^{0.5} \).
models \((\alpha_2, \alpha_8)\), the \(k_{\text{LW}}\) ridge estimator gave the worse performance among all ridge estimators studied in this chapter, and RIDGN and KGRR values of \(M_l\) were the closest to \(M_l(\text{OLS})\). However, judging from the performance of ROP, the theoretical ridge estimator for \(\alpha_1\) known, for the 8 models, ridge regression would not be outperformed by OLS for any model if the optimal value of \(k\) were to be computed from the true \(\alpha_1\). Unfortunately, the \(k_{\text{LW}}\) ridge estimator apparently gave a \(k\) value which was the closest to the optimal \(k\) only for models \(\alpha_1, \alpha_3\) and \(\alpha_7\).

Results in table 4 also show that the proposed method of estimating the observed squared loss, \(M_l(k_{\text{LW}})\), gave an excellent performance since it overestimated \(M_l(k_{\text{LW}})\) by less than 8% for all the models except for models \(\alpha_5\) and \(\alpha_6\) where it underestimated \(M_l(k_{\text{LW}})\) by about 6%, 5%, respectively. The stability and accuracy of the estimated \(\tilde{M}_l(k_{\text{LW}})\), measured by the coefficients of loss variation \(C*/V\), were very comparable to OLS for all the models, and it even had smaller \(C*/V\) than OLS for models \(\alpha_3, \alpha_4, \alpha_7,\) and \(\alpha_8\), indicating an excellent performance by the proposed method. \(M_l\) for the other ridge estimators were again seriously underestimated. \(\tilde{M}_l(k_{\text{HKB}})\) was the closest to the observed \(M_l(k_{\text{HKB}})\), second only to \(k_{\text{LW}}\) ridge estimator. However, \(M_l(k_{\text{HKB}})\) was seriously under-
estimated for \( \alpha_2 \) and \( \alpha_8 \), the two situations where it would be the most desirable not to underestimate.

Results of the simulations for the same 8 models \((\alpha_1-\alpha_8)\) with \( \sigma^2 = 0.25 \) are reported in table 5. Relatively high signal-to-noise ratio \( S/N \) are evident for the 8 models where \( S/N \) ranges from 273 to about 1,942. The trend of the increase in the value of \( M_1 \) for the ridge estimators due to the decrease in the value of \( \sigma^2 \) continues to be true for table 5. \( M_1(k_{LW}) \) was significantly higher than \( M_1(OLS) \) for models less "favorable" for ridge regression analysis, namely \( \alpha_2, \alpha_5 \) and \( \alpha_8 \) since the smallest \( \lambda_i \) for these three vectors is associated with the largest \( \alpha_i \) value. The \( k_{LW} \) estimator tends to over shrink the value of \( \alpha_1 \) for these three models, resulting in a substantial bias squared component, as pointed out earlier. In fact, all ridge estimators in this simulation performed poorly compared to OLS for models \( \alpha_2, \alpha_5 \) and \( \alpha_8 \). However, it is encouraging that the proposed squared error loss estimator performed very well in predicting \( M_1(k_{LW}) \), especially important for the least favorable cases, \( \alpha_2 \) and \( \alpha_8 \), since it tends to alert the researcher of the potential loss that might result from the use of ridge regression for these unfavorable cases. The stability of the estimated \( \tilde{M}_1(k_{LW}) \) measured by \( C*/V \) was also excellent, giving a \( C*/V \) value even slightly smaller.
Table 5. Signal-to-noise $M_l$, estimated $M_l$, $\hat{M}_l$ and the coefficients of loss variation for the eight models $(a_1-a_8)$, with $\lambda = (0.01, 1.0, 1.99)$, $\sigma^2 = 0.25$. OLS vs. ridge estimators: $k_{LM}$, RIDGN, $k_{HKB}$, KGRR and ROP.

<table>
<thead>
<tr>
<th>Signal-To-Noise</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$a_5$</th>
<th>$a_6$</th>
<th>$a_7$</th>
<th>$a_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>S/N$^a$</td>
<td>1200.00</td>
<td>475.40</td>
<td>1200.00</td>
<td>1942.5</td>
<td>677.20</td>
<td>1200.00</td>
<td>1722.72</td>
<td>273.36</td>
</tr>
<tr>
<td>OLS $^{b,c}$</td>
<td>$M_l$</td>
<td>100.367</td>
<td>105.329</td>
<td>101.626</td>
<td>99.251</td>
<td>99.727</td>
<td>102.222</td>
<td>103.314</td>
</tr>
<tr>
<td>$C^*/V$</td>
<td>.023</td>
<td>.021</td>
<td>.022</td>
<td>.021</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
<td>.021</td>
</tr>
<tr>
<td>$k_{LM}$</td>
<td>$M_l$</td>
<td>84.406</td>
<td>166.722</td>
<td>71.055</td>
<td>77.154</td>
<td>108.355</td>
<td>92.180</td>
<td>76.437</td>
</tr>
<tr>
<td>$C^*/V$</td>
<td>.022</td>
<td>.021</td>
<td>.021</td>
<td>.020</td>
<td>.023</td>
<td>.024</td>
<td>.021</td>
<td>.017</td>
</tr>
<tr>
<td>RIDGN $^{d}$</td>
<td>$M_l$</td>
<td>123.040</td>
<td>134.280</td>
<td>92.179</td>
<td>90.660</td>
<td>127.179</td>
<td>127.970</td>
<td>65.174</td>
</tr>
<tr>
<td>$k_{HKB}$</td>
<td>$M_l$</td>
<td>96.700</td>
<td>141.765</td>
<td>78.639</td>
<td>77.184</td>
<td>107.571</td>
<td>108.558</td>
<td>74.478</td>
</tr>
<tr>
<td>$C^*/V$</td>
<td>.023</td>
<td>.023</td>
<td>.023</td>
<td>.023</td>
<td>.023</td>
<td>.023</td>
<td>.023</td>
<td>.023</td>
</tr>
<tr>
<td>KGRR $^{e}$</td>
<td>$M_l$</td>
<td>122.295</td>
<td>132.703</td>
<td>92.474</td>
<td>91.241</td>
<td>125.883</td>
<td>126.813</td>
<td>66.199</td>
</tr>
<tr>
<td>$C^*/V$</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
<td>.022</td>
</tr>
<tr>
<td>ROP $^{f}$</td>
<td>$M_l$</td>
<td>84.203</td>
<td>107.610</td>
<td>70.937</td>
<td>69.395</td>
<td>91.237</td>
<td>91.905</td>
<td>68.488</td>
</tr>
</tbody>
</table>

a. Signal-to-noise $S/N = \frac{P}{\sum_{i=1}^{p} \lambda i a_i^2 / \sigma^2}$.

b. $M_l = \sum_{i=1}^{p} (\hat{a}_i - a_i)^2 / \sigma^2$ averaged over 1000 experiments.

c. $M_l = \sum_{i=1}^{p} (m_i^2 / \lambda_i + (1-m_i)^2 \hat{a}_i^2) / \sigma^2$ where $\hat{a}_i = (\hat{a}_i - V(\hat{a}_i)) > 0$ averaged over 1000 experiments.

d. $C^*/V = \text{coefficients of loss variation} = \frac{3}{\sum_{i=1}^{p} (1-i_1-i_2)^2 / n-1} \left( \frac{3}{\sum_{i=1}^{p} (1-i_1-i_2)^2 / n} \right)^{0.5}$. 

than C*/V for OLS for models $a_1$, $a_3$, $a_4$ and $a_8$. However, $M_1$ was again understated by the proposed loss estimator for all the other ridge estimators, with $M_1(k_{HKB})$ being predicted second best after $k_{LW}$. The ROP performance in table 5 again showed that ordinary (one value of $k$) ridge regression usually outperforms OLS (except for $a_2$ and $a_8$) if the true $a$ vector is known.

At this point the reader may wonder why the ROP estimator had a higher squared error loss divided by $\sigma^2(M_1)$ than did OLS for $a_2$ and $a_8$ in table 5, given that the true $a$ is known for the ROP estimator. The reason for this paradox is because ROP is an ordinary (one value of $k$) ridge estimator. The single $k$ value for ROP was computed by $k = \sigma^2(\sum_{i=1}^{3} a_i^2/3)$. This single value of $k$ works well if the $a_i$ value associated with the smallest eigenvalue is not too much larger than the other $a_i$ values. However, for $a_2$ and $a_8$, $a_1$, which has the smallest eigenvalue of 0.01, is from about 2.45 to 4.69 times larger than $a_2$ and $a_3$. Thus, for $a_2$ and $a_8$ with $\sigma^2 = 0.25$ in table 5, using the simple average of the $a_i^2$ values does not result in a good single value of $k$. If ROP were converted to a generalized ridge estimator, it would never be outperformed by OLS.

Results in tables 1-5 illustrate the fact that none of the presently proposed methods for selecting $k$ will
always give a low squared error loss for all the models. However, the Lawless-Wang ridge estimator gave the smallest squared error loss for the majority of the models studied in this simulation, followed in performance by the \( k_{HKB} \) estimator.

One final result worth mentioning before leaving this section is that the KGRR and RIDGN ridge estimators always gave an observed squared error loss (\( M_l \)) less than \( M_l(OLS) \) as long as \( \xi = \lambda_i a_i^2/\sigma^2 \) was less than or equal to 2, where \( \lambda_i \) is the smallest eigenvalue and \( a_i \) is associated with \( \lambda_i \). To illustrate this result, the following two examples are given:

Example 1: Model \( \tilde{\alpha}_4 = (\sqrt{3.5}, \sqrt{3.5}, 15) \), \( \vec{\lambda} = (.01, 1., 1.99) \), and \( \sigma^2 = 0.25 \) in table 5. For this model, \( \xi = ((.01)(37.5)/.25) = 1.5 \leq 2 \) and \( M_l(\text{RIDGN}) = 90.96 \ M_l(\text{OLS}) \) and \( M_l(\text{KGRR}) = 91.24 \ M_l(\text{OLS}) \).

Example 2: Model \( \tilde{\alpha}_3 = (10 \ 10 \ 10) \), \( \vec{\lambda} = (.01, 1., 1.99) \), and \( \sigma^2 = 0.25 \) in table 5. For this model, \( \xi = (.01)(100)/.25 = 4 > 2 \), \( M_l(\text{RIDGN}) = 123.23 \ M_l(\text{OLS}) \), and \( M_l(\text{KGRR}) = 122.29 > M_l(\text{OLS}) \).

The above results held for all the models of the simulation.

The results of the Monte Carlo experiments for the 8 models \( (\alpha_1 - \alpha_8) \) with \( \vec{\lambda} = (.1, 1, 1.9) \) and 5 different values of \( \sigma^2 \), \( \sigma^2 = 4, 2, 1, 0.5 \) and \( .25 \), are given in tables 6-10. The multicollinearity problem for these 8
models is less severe since the smallest eigenvalue is \( \lambda_1 = .1 \). Therefore, \( 1/\lambda_1 \), the variance inflation factor for \( \alpha_1 \), VIF\(_1\), is only 10 compared to VIF\(_1\) = 100 for the models discussed in tables 1-5. Results in table 6 show that M1(OLS) = \( \frac{1}{3} \sum (\hat{\alpha}_i - \alpha_i)^2 / \sigma^2 \) did not vary much among the models and averaged about 11.5 across all models. The Lawless-Wang ridge estimator, \( k_{LW} \), performed somewhat better than OLS in terms of M1 for models \( \alpha_1, \alpha_3, \alpha_4, \alpha_6 \), and \( \alpha_7 \) in table 6. The greatest reduction in M1(\( k_{LW} \)) was shown for model \( \alpha_3 \) with M1(\( k_{LW} \)) = 66% of M1(OLS), followed by models \( \alpha_7, \alpha_4, \alpha_1 \), and \( \alpha_6 \).

The other ridge estimators (\( k_{HKB}, \) RIDGN, and KGRR) showed improvement in M1 compared to M1(OLS) only for models \( \alpha_4, \alpha_3 \), and \( \alpha_7 \). All ridge estimators including \( k_{LW} \) had M1 values substantially larger than M1(OLS) for models \( \alpha_2 \) and \( \alpha_8 \) since the largest coefficient, \( \alpha_1 \), is associated with the smallest eigenvalue of \( \lambda_1 = .1 \) for these two models. It should be noted that for the "unfavorable" models (\( \alpha_2 \) and \( \alpha_8 \)), \( k_{HKB}, \) RIDGN and KGRR gave an observed M1 smaller than M1(\( k_{LW} \)); however, these ridge estimators are outperformed by M1(OLS) for these two models. M1(\( k_{LW} \)) was very close to M1(ROP) for models \( \alpha_1, \alpha_3, \alpha_4, \) and \( \alpha_6 \). The potential performance of ordinary ridge regression as judged by M1(ROP) showed some improvement compared to M1(OLS) for all the eight models except for \( \alpha_8 \). However, this
Table 6. Signal-to-noise Ml, estimated Ml, \( \hat{M} \) and the coefficients of loss variation for the eight models with \( \lambda = (1.1, 1.9) \), \( \sigma^2 = 4 \). OLS vs. ridge estimators: \( k_{LM} \), RIDGN, \( k_{HK} \), KGRR and ROP.

<table>
<thead>
<tr>
<th>Signal-Noise S/N²</th>
<th>( u_1^a )</th>
<th>( u_2^a )</th>
<th>( u_3^a )</th>
<th>( u_4^a )</th>
<th>( u_5^a )</th>
<th>( u_6^a )</th>
<th>( u_7^a )</th>
<th>( u_8^a )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( (10,10,10) )</td>
<td>75.00</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
</tr>
<tr>
<td>( (15,37.5,37.5) )</td>
<td>32.81</td>
<td>0.0803</td>
<td>0.0775</td>
<td>0.0775</td>
<td>0.0733</td>
<td>0.0765</td>
<td>0.0765</td>
<td>0.0765</td>
</tr>
<tr>
<td>( (37.5,15,37.5) )</td>
<td>117.19</td>
<td>11.3865</td>
<td>11.2482</td>
<td>11.4152</td>
<td>11.6187</td>
<td>11.7895</td>
<td>11.0984</td>
<td>11.0984</td>
</tr>
<tr>
<td>( (57.5,57.5,15) )</td>
<td>45.30</td>
<td>11.5126</td>
<td>11.6187</td>
<td>11.7215</td>
<td>10.0602</td>
<td>8.1994</td>
<td>24.3493</td>
<td>24.3493</td>
</tr>
<tr>
<td>( (12,12,12) )</td>
<td>75.00</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
</tr>
<tr>
<td>( (12,12,12) )</td>
<td>104.70</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
</tr>
</tbody>
</table>

**a.** Signal-to-noise \( S/N = \sum_{i=1}^{3} \lambda_i a_i^2 / \sigma^2 \).  

**b.** \( Ml = \sum_{i=1}^{3} (\hat{a}_i - a_i)^2 / \sigma^2 \) averaged over 1000 experiments.  

**c.** \( \hat{M} = \sum_{i=1}^{3} (m^2 \sigma^2 / \lambda_i) \hat{v}(i-1)/a_i^2 \) where \( \hat{a}_i = (\hat{a_i}^2 - V(\hat{a}_i)) > 0 \) averaged over 1000 experiments.  

**d.** \( C^*/V = \) the coefficients of loss variation \( = \sum_{i=1}^{3} (l_i^2 / (l_i^2 - 1)) / (l_i^2 - 1 / l_i^2) )0.5 \).
potential improvement was not completely realized, even by the $k_{LW}$ estimator, because of the inherent difficulty in estimating the optimal value of $k$ from the OLS estimates of the $\alpha_i$.

The proposed method of estimating the squared error loss, $\tilde{M}_l(k_{LW})$ was surprisingly accurate since in all the eight models $\tilde{M}_l(k_{LW})$ was different from $M_l(k_{LW})$ by less than 3%. The accuracy and stability of $\tilde{M}_l(k_{LW})$ measured by the coefficients of loss variation, $C*/V$, were good since $C*/V$ for $k_{LW}$ were smaller than or equal to $C*/V$ for OLS in all the models in table 6, except for $\alpha_2$, $\alpha_5$, and $\alpha_6$. The proposed method for estimating the squared error loss did underestimate $M_l$ for the other ridge estimator; however, $\tilde{M}_l(k_{HKB})$ and $\tilde{M}_l(RIDGN)$ were fairly close to their average observed squared error loss, $M_l$, for some cases, such as for models $\alpha_1$, $\alpha_3$, $\alpha_4$, and $\alpha_7$.

Results of the Monte Carlo experiments for the 8 models $\alpha_1-\alpha_8$ with $\sigma^2 = 2$ are shown in table 7. $\tilde{M}_l(OLS)$ did not change, as expected since $M_l(OLS) = \sum_{i=1}^{p} \sigma^2 / \sigma^2 \lambda_i$, the $\sigma^2$ will cancel out, and $M_l(OLS)$ is just equal to the sum of the VIF, i.e., $M_l(OLS) = \sum_{i=1}^{3} 1 / \lambda_i$. The Lawless-Wang ridge estimator, $k_{LW}$ had $M_l(k_{LW})$ smaller than $M_l(OLS)$ for models "favorable" to ridge analysis, namely $\alpha_1$, $\alpha_3$, $\alpha_4$, $\alpha_6$, and $\alpha_7$. The other ridge estimators varied in performance among models with the $k_{HKB}$
Table 7. Signal-to-noise $M_l$, estimated $M_l$, $M_l$ and the coefficients of loss variation for the eight models 
$(\alpha_i-\alpha_j)$, with $\lambda = (1, 1, 1.9)$, $\sigma^2 = 2$. OLS vs. ridge estimators: $k_{LM}$, RIDGN, $k_{HKB}$, KGRR and ROP.

| Signal-Noise S/N & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 & a_8 |
|-----------------|-----|-----|-----|-----|-----|-----|-----|-----|
| $(10,10,10)$    | 150.00 | 150.00 | 234.375 | 90.60 | 150.00 | 209.40 | 44.70 |
| $(15/37.5/37.5)$ | 65.625 | 11.387 | 11.248 | 11.415 | 11.619 | 11.789 | 11.098 |
| $(12/12/12)$    | | .057 | .055 | .052 | .054 | .055 | .056 |
| $k_{LM}$        | | | | | | | |
| $k_{HKB}$       | | | | | | | |
| KGRR            | | | | | | | |
| ROP             | | | | | | | |

a. Signal-to-noise $S/N = \sum_{i=1}^{N} \frac{1}{\sigma^2} \lambda_i \eta_i^2 / \sigma^2$.

b. $M_l = \sum_{i=1}^{N} \frac{(\eta_i - \eta_{i-1})^2}{\sigma^2}$ averaged over 1000 experiments.

c. $M_l = \sum_{i=1}^{N} \frac{(\nu_i^2 - \nu_{i-1})^2}{\sigma^2} \lambda_i + (1 - \lambda_i)^2 \eta_i^2 / \sigma^2$ where $\eta_i = (\nu_i^2 - \nu_{i-1})^2 \theta_i > 0$ averaged over 1000 experiments.

d. $C^*/V = \text{coefficients of loss variation} = \sum_{i=1}^{N} \frac{1}{\lambda_i^2 - L_{i-1}^2} \eta_i^2 / \left( \sum_{i=1}^{N} \frac{L_i^2}{L_i^2} \right)^{0.5}$.
estimator seeming to slightly outperform both RIDGN
and KGRR in most cases. The $k_{HKB}$ estimator had an $M_l$
smaller than $M_l(OLS)$ only for models $a_1$, $a_3$, $a_4$, and $a_7$.
It should be noted that in terms of the observed squared
error loss, $M_l$, RIDGN estimator outperformed all ridge
estimators including $k_{LW}$ for models $a_2$, $a_7$, and $a_8$, which
would generally be regarded as "unfavorable" models,
extcept for model $a_7$. $M_l(k_{LW})$ was very close to $M_l(ROP)$
for models $a_1$, $a_3$, $a_4$, $a_6$, and $a_7$ for the same reasons
given earlier. The excellent performance of the pro-
posed method of estimating the observed squared error
loss, $M_l$, should be noted, since $M_l(k_{LW})$ was different
from $M_l(k_{LW})$ for all the models by less than 7%. The
stability and the accuracy of $M_l(k_{LW})$ measured by $C*/V$
was also very good compared to $M_l(OLS)$ since $C*/V$ for
$M_l(k_{LW})$ was smaller than $C*/V$ for $M_l(OLS)$ for all the
models except models $a_5$ and $a_6$. The underestimation of
$M_l$ for the other ridge estimators by the proposed
method for predicting $M_l$ was usually less serious for
the ridge estimator $k_{HKB}$ than for RIDGN and KGRR.

The results in table 8 show the performance of
the ridge estimators and the proposed method of esti-
mating the observed squared error loss $M_l$ for the same
8 models ($a_1-a_8$) with $\sigma^2 = 1$. The Lawless-Wang ridge
estimator, $k_{LW'}$, showed some marginal improvement in $M_l$
compared to $M_l(OLS)$ for models "favorable" to ridge
Table 8. Signal-to-noise $M_l$, estimated $M_l$, $M_l$ and the coefficients of loss variation for the eight models 
($u_i - \mu_i$), with $\lambda = (1.1, 1.1, 1.9)$, $\sigma^2 = 1$. OLS vs. ridge estimators: $k_{LM}$, $K_{RIGG}$, $K_{HKB}$, $K_{BAR}$ and $R_{OP}$.

<table>
<thead>
<tr>
<th>Signal-Noise</th>
<th>$u_1^{\ast}$</th>
<th>$u_2^{\ast}$</th>
<th>$u_3^{\ast}$</th>
<th>$u_4^{\ast}$</th>
<th>$u_5^{\ast}$</th>
<th>$u_6^{\ast}$</th>
<th>$u_7^{\ast}$</th>
<th>$u_8^{\ast}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>300.00</td>
<td>300.00</td>
<td>468.75</td>
<td>151.20</td>
<td>300.00</td>
<td>410.80</td>
<td>89.40</td>
</tr>
<tr>
<td>$S/N^i$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C*/V</td>
<td>.0402</td>
<td>.0378</td>
<td>.0367</td>
<td>.0366</td>
<td>.0383</td>
<td>.0392</td>
<td>.0381</td>
</tr>
<tr>
<td>k$_{LM}$</td>
<td>$M_l$</td>
<td>10.8053</td>
<td>17.5922</td>
<td>10.1329</td>
<td>10.2327</td>
<td>12.5013</td>
<td>11.4289</td>
<td>10.5361</td>
</tr>
<tr>
<td></td>
<td>C*/V</td>
<td>.0388</td>
<td>.0372</td>
<td>.0378</td>
<td>.0366</td>
<td>.0383</td>
<td>.0405</td>
<td>.0381</td>
</tr>
</tbody>
</table>

- a. Signal-to-noise $S/N = \frac{\sigma}{\sum_{i=1}^{3} \chi_i^2 / 3}$
- b. $M_l = \sum_{i=1}^{3} \frac{(\mu_i - g_i)^2}{\sigma^2}$ averaged over 1000 experiments.
- c. $M_l = \sum_{i=1}^{3} \frac{(\mu_i^2 - \mu_1^2) / 3}{\sigma^2}$ averaged over 1000 experiments.
- d. $C*/V = \text{the coefficients of loss variation} = \sqrt{\left( \sum_{i=1}^{3} \frac{(\mu_i^2 - \mu_1^2)^2}{n-1} / (\sum_{i=1}^{3} \mu_i^2 / n) \right)^{0.5}}$. 
regression analysis, $\alpha_1$, $\alpha_3$, $\alpha_4$, and $\alpha_7$. However, $k_{LW}$ performed worse than OLS for models "unfavorable" to ridge analysis, such as $\alpha_2$ and $\alpha_8$. For these "unfavorable" models, RIDGN and KGRR outperformed all the other ridge estimators, including $k_{LW}$, where they gave an observed $M_1$ substantially smaller than $M_1(k_{LW})$. Nearly excellent prediction of $M_1(k_{LW})$ by $\tilde{M}_1(k_{LW})$ can be seen in table 8 for all the models, including the "unfavorable" ones, giving some promise that the researcher could evaluate the performance of $k_{LW}$, based only on the sample information.

The coefficient of loss variation, $C*/V$, for $\tilde{M}_1(k_{LW})$ was equal to or smaller than $C*/V$ for $M_1(\text{OLS})$ for all the models in table 8, except for $\alpha_6$, indicating an accurate and stable performance of $\tilde{M}_1(k_{LW})$ by the proposed method. The proposed method also improved its performance in predicting $M_1(k_{HKB})$ for all the models, despite some underestimation, usually less than 12%. $M_1(k_{LW})$ was again very close to $M_1(\text{ROP})$ for models $\alpha_1$, $\alpha_3$, $\alpha_4$, $\alpha_6$, and $\alpha_7$. The performance of the ordinary ridge estimator based upon the true $\alpha$ vector, ROP, showed a potential marginal improvement over OLS for all the models, except model $\alpha_2$ and $\alpha_8$, which are the most "unfavorable" models for ridge analysis.

The Monte Carlo results for the same 8 models presented in the earlier tables, but with $\sigma^2$ equal to
Table 9. Signal-to-noise $M_l$, estimated $M_l$, $\hat{M}_l$ and the coefficients of loss variation for the eight models $(a_1^{10-18})$, with $\lambda = (1.1.1.1.9)$, $\sigma^2 = .5$. OLS vs. ridge estimators: $k_{LM}$, RIDGN, $k_{HKB}$, KGRR and ROP.

<table>
<thead>
<tr>
<th>Signal-Noise $S/N^a$</th>
<th>$\alpha_1^a$ ($10,10,10$)</th>
<th>$\alpha_2^a$ ($15/37.5/37.5$)</th>
<th>$\alpha_3^a$ ($37.5/15/37.5$)</th>
<th>$\alpha_4^a$ ($57.5/37.5/15$)</th>
<th>$\alpha_5^a$ ($12,12/12$)</th>
<th>$\alpha_6^a$ ($12/12,12$)</th>
<th>$\alpha_7^a$ ($1/12,12,12$)</th>
<th>$\alpha_8^a$ ($16/12/22$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C^*/V^d$</td>
<td>.028</td>
<td>.026</td>
<td>.027</td>
<td>.026</td>
<td>.027</td>
<td>.027</td>
<td>.027</td>
</tr>
<tr>
<td>$k_{LM}$</td>
<td>$M_1^e$</td>
<td>11.196</td>
<td>15.709</td>
<td>10.775</td>
<td>10.709</td>
<td>12.217</td>
<td>11.593</td>
<td>11.125</td>
</tr>
<tr>
<td></td>
<td>$C^*/V$</td>
<td>.028</td>
<td>.026</td>
<td>.027</td>
<td>.025</td>
<td>.028</td>
<td>.029</td>
<td>.027</td>
</tr>
<tr>
<td></td>
<td>$C^*/V$</td>
<td>.028</td>
<td>.026</td>
<td>.027</td>
<td>.025</td>
<td>.028</td>
<td>.029</td>
<td>.027</td>
</tr>
<tr>
<td>$k_{HKB}$</td>
<td>$M_1^i$</td>
<td>10.988</td>
<td>11.547</td>
<td>10.754</td>
<td>10.746</td>
<td>11.180</td>
<td>11.175</td>
<td>10.657</td>
</tr>
</tbody>
</table>

a. Signal-to-noise $S/N = \sum_{i=1}^{p} A_i^2 / \sigma^2$.
b. $M_1 = \sum_{i=1}^{3} (a_i - \bar{a}_i) / \sigma^2$ averaged over 1000 experiments.
c. $\hat{M}_1 = \sum_{i=1}^{3} \left( \frac{m_i^2}{\sigma^2} \right) + \frac{1}{2} \frac{\sigma^2}{\bar{a}_i} \frac{1}{\sigma^2}$ where $\bar{a}_i = (\sum_{i=1}^{p} V(a_i)) / 9$ averaged over 1000 experiments.
d. $C^*/V$ = the coefficients of loss variation $= \left( \sum_{i=1}^{3} \left( \frac{C_i^2}{C_{i-1}} \right)^2 / n - 1 / \left( \sum_{i=1}^{3} \frac{C_i^2}{n} \right) \right)^{0.5}$. 

.5, are presented in table 9. For this value of $\sigma^2$, it should be noted that only a slight marginal improvement compared to OLS is realized for models $x_1$, $x_3$, $x_4$, $x_6$, and $x_7$. Certainly there were no significant improvements from the use of ridge regression, even when the optimal value of $k$ was known, as can be seen from the ROP performance. It should be noted that under the "unfavorable" conditions, the ridge estimators could be ranked in terms of Ml performance as follows: $k_{HKB'}$, RIDGN, KGRR, and $k_{LW}$, respectively. The proposed method of estimating $M_l(k_{LW})$ gave an excellent prediction of the value of $M_l(k_{LW})$ for all the models. This result was very encouraging since it could prevent misleading conclusions based upon the sample observations. The coefficient of variation, $C^*/V$, for $\tilde{M}_l(k_{LW})$ was almost identical to $C^*/V$ for $\tilde{M}_l(OLS)$ for all the models shown in table 9. The Lawless-Wang ridge estimator, although outperformed by $k_{HKB'}$ and RIDGN for the "unfavorable" models, gave $M_l(k_{LW})$ which is very close to $M_l(ROP)$ for all the "favorable" models. The results of the 8 models with $\sigma^2 = 0.25$ reported in table 10 showed no gain from the use of ridge regression compared to OLS, as indicated by the performance of ROP. The results in table 10 show that the Lawless-Wang ridge estimator, $k_{LW}$, provided only a non-significant marginal improvement in $M_l(k_{LW})$ for the most "favorable" models,
Table 10. Signal-to-noise \( \text{S/N} \), estimated \( \text{Ml} \) and the coefficients of loss variation for the eight models

\( \{a_1-a_8\} \), with \( \lambda = \{0.1, 1, 1.9\} \), \( \sigma^2 = 25 \). OLS vs. ridge estimators: \( k_{\text{LM}} \), RIDGN, \( k_{\text{HKB}} \), KGR and ROP.

<table>
<thead>
<tr>
<th></th>
<th>( a_1 = (10, 10, 10) )</th>
<th>( a_2 = (15, 15, 15, 17.5) )</th>
<th>( a_3 = (17.5, 17.5, 15) )</th>
<th>( a_4 = (12, 12, 12, 12) )</th>
<th>( a_5 = (12, 12, 12, 12) )</th>
<th>( a_6 = (12, 12, 12, 12) )</th>
<th>( a_7 = (12, 12, 12, 12) )</th>
<th>( a_8 = (16, 22, 22) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal-Noise S/N^b</td>
<td>1200.00</td>
<td>525.00</td>
<td>1200.00</td>
<td>1075.00</td>
<td>724.8</td>
<td>1200.00</td>
<td>1675.2</td>
<td>357.6</td>
</tr>
<tr>
<td>OLS</td>
<td>( \text{Ml} )</td>
<td>11.3865</td>
<td>12.0013</td>
<td>11.5382</td>
<td>11.2482</td>
<td>11.4152</td>
<td>11.6187</td>
<td>11.7885</td>
</tr>
<tr>
<td></td>
<td>( \text{MlC} )</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
<td>11.5263</td>
</tr>
<tr>
<td></td>
<td>( \text{C}*/\text{V}^d )</td>
<td>0.0200</td>
<td>0.0189</td>
<td>0.0193</td>
<td>0.0183</td>
<td>0.0191</td>
<td>0.0196</td>
<td>0.0191</td>
</tr>
<tr>
<td>( k_{\text{LM}} )</td>
<td>( \text{Ml} )</td>
<td>11.3392</td>
<td>14.2125</td>
<td>11.1344</td>
<td>10.9665</td>
<td>11.9405</td>
<td>11.6518</td>
<td>11.4545</td>
</tr>
<tr>
<td></td>
<td>( \text{C}*/\text{V}^d )</td>
<td>0.0197</td>
<td>0.0190</td>
<td>0.0193</td>
<td>0.0180</td>
<td>0.0197</td>
<td>0.0200</td>
<td>0.0189</td>
</tr>
<tr>
<td></td>
<td>( \text{C}*/\text{V}^d )</td>
<td>0.0197</td>
<td>0.0190</td>
<td>0.0193</td>
<td>0.0180</td>
<td>0.0197</td>
<td>0.0200</td>
<td>0.0189</td>
</tr>
<tr>
<td>( k_{\text{HKB}} )</td>
<td>( \text{Ml} )</td>
<td>11.6022</td>
<td>12.9243</td>
<td>11.2381</td>
<td>10.9569</td>
<td>11.8756</td>
<td>12.0359</td>
<td>11.3672</td>
</tr>
<tr>
<td></td>
<td>( \text{MlC} )</td>
<td>11.2630</td>
<td>11.5597</td>
<td>11.1300</td>
<td>11.1266</td>
<td>11.3662</td>
<td>11.3640</td>
<td>11.0734</td>
</tr>
<tr>
<td></td>
<td>( \text{C}*/\text{V}^d )</td>
<td>0.0197</td>
<td>0.0190</td>
<td>0.0193</td>
<td>0.0180</td>
<td>0.0197</td>
<td>0.0200</td>
<td>0.0189</td>
</tr>
<tr>
<td></td>
<td>( \text{C}*/\text{V}^d )</td>
<td>0.0197</td>
<td>0.0190</td>
<td>0.0193</td>
<td>0.0180</td>
<td>0.0197</td>
<td>0.0200</td>
<td>0.0189</td>
</tr>
</tbody>
</table>

a. Signal-to-noise S/N = \( \frac{p}{\sum_{i=1}^{3} \lambda_i q_i^2/n^2} \).

b. \( \text{Ml} = \frac{3}{\sum_{i=1}^{3} (\tilde{a}_i - a_i)^2/n^2} \) averaged over 1000 experiments.

c. \( \tilde{Ml} = \frac{3}{\sum_{i=1}^{3} (\text{Ml}_i - a_i)^2/n} \) averaged over 1000 experiments.

d. \( \text{C}*/\text{V} = \text{the coefficients of loss variation} = \left( \frac{\sum_{i=1}^{3} (\tilde{a}_i - a_i)^2/n-1/\sum_{i=1}^{3} (\tilde{a}_i - a_i)^2/n) \right)^{0.5} \).
The other observation that could be made is that \( k_{HKB} \), \( RIDGN \), and \( KGRR \), all outperformed \( k_{LW} \) for the "unfavorable" models by a substantial margin, making them a better alternative to use if ridge regression is to be used at all for these kinds of situations. The proposed method of estimating the observed squared error loss for ridge estimators performed well in predicting \( M_1 \) for most of the models in table 10.

The Distribution of the Lawless-Wang Ridge Estimator

Consider the general linear regression model,
\[
y = X\beta + u
\]
where \( y \) is \((n \times 1)\) vector of \( n \) observation on the variable to be explained, \( X \) is an \((n \times p)\) matrix with full rank of \( n \) observation on \( p \) explanatory variables, \( \beta \) is a \((p \times 1)\) vector of coefficients and \( u \) is a \((n \times 1)\) vector of unobservable disturbances with the following assumptions: \( u \) is distributed normally with
\[
E(u) = 0 \text{ and } E(uu') = \sigma^2 I_n.
\]
Following Hoerl and Kennard (1970, a,b), Lawless and Wang (1976), and Hoerl, Baldwin and Kennard (1976), it is assumed that \( X'X = \Lambda \) where \( \Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p) \) is a \((p \times p)\) diagonal matrix. The ordinary ridge regression estimator is given by the following:
\[
\hat{\alpha} = (Z'Y + kI)^{-1}Z'y = (A + kI)^{-1}Z'y
\]
\[
= M\hat{\alpha}
\]

where \( k \) is a scalar non-negative parameter, \( \hat{\alpha} = (Z'Z)^{-1}Z'y = A^{-1}Z'y \) is the OLS estimate of \( \alpha \) and
\[
M = (A + kI)^{-1}A
\]
such that for each \( \hat{\alpha}_i \), \( m_i = \frac{\lambda_i}{\lambda_i + k} \) where \( \lambda_i \) is the \( i \)th eigenvalue of \( (X'X) \).

The squared error loss of ridge estimators, \( L_1(\hat{\alpha}_i) \), can be written as:
\[
L_1(\hat{\alpha}_i) = E((\hat{\alpha}_i - \alpha_i)^2) = E(m_i \hat{\alpha}_i - \alpha_i)^2 = E(m_i^2 \hat{\alpha}_i^2 - 2m_i \hat{\alpha}_i \alpha_i + \alpha_i^2)
\]
\[
= m_i^2 E\hat{\alpha}_i^2 - 2m_i \alpha_i E\hat{\alpha}_i + \alpha_i^2
\]
\[
= m_i^2 (\alpha_i^2 + \sigma^2 / \lambda_i) - 2m_i \alpha_i^2 + \alpha_i^2
\]
\[
= m_i^2 \alpha_i^2 + m_i^2 \sigma^2 / \lambda_i - 2m_i \alpha_i^2 + \alpha_i^2
\]
\[
= m_i^2 \sigma^2 / \lambda_i + \alpha_i^2 (1 - 2m_i + m_i^2)
\]
\[
= m_i^2 \sigma^2 / \lambda_i + (1 - m_i)^2 \alpha_i^2
\]
\[
= \text{Variance}(\hat{\alpha}_i) + (\text{Bias}(\hat{\alpha}_i))^2.
\]

Provided that \( k \) is nonstochastic, the above result was written in Chapter I as
\[
E(\hat{\alpha}_i - \alpha_i) (\hat{\alpha}_i - \alpha_i) = \sum_{i=1}^{p} \frac{\sigma^2 \lambda_i + \alpha_i^2 k_i}{(\lambda_i + k_i)^2}
\]

Minimizing the just indicated expression for the squared error loss with respect to \( k \) yields the optimal \( k \) value, \( k_i = \frac{\alpha_i^2}{\lambda_i} \). Of course, in a real situation both \( \sigma^2 \) and \( \alpha_i^2 \) are unknown.

Different methods for the selection of the \( k \) value have been discussed in Chapter I. In this chapter, the Lawless-Wang ridge estimator, suggested by Lawless
and Wang (1976) will be referred to throughout this section where, \( k_{LW} \) is estimated by

\[
k_{LW} = \frac{p \hat{\sigma}^2}{\sum_{i=1}^{p} \hat{a}_i^2 \lambda_i}
\]

where \( \hat{\sigma}^2 \) is the OLS estimate of \( \sigma^2 \) which could be obtained from the residual squared error,

\[
\hat{\sigma}^2 = \frac{1}{n-v} (y - X\hat{\beta})' (y - X\hat{\beta})
\]

where

\[
v = \begin{cases} 
  n-p-1 & \text{if observations are deviations from their corresponding means.} \\
  n-p & \text{Otherwise.}
\end{cases}
\]

The properties of an unbiased estimators like OLS are often expressed in terms of the expected value and the variance of the estimates. The expected value of an OLS estimate, \( \hat{\beta} \), is its true value, \( E(\hat{\beta}) = \beta \), since by assumption the normally distributed \( u \) has \( E(u) = 0 \). \( \hat{\beta} \) is an unbiased estimate of \( \beta \), and is distributed normally with mean \( \beta \) and \( \text{Var-Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1} \) which could be written as

\[
\hat{\beta} \sim \text{INN}(\beta, \sigma^2 (X'X)^{-1}).
\]

The OLS estimate, \( \hat{\beta} \), is the best linear unbiased estimator, BLUE, \( \hat{\beta} = (X'X)^{-1}X'y \) is independent from its variance, \( \sigma^2 (X'X)^{-1} \), which makes it possible to test hypotheses about the significance of the coefficients of the general linear regression model, \( y = X\beta + u \), using the standard t and F statistics.

Unlike the OLS estimates, \( \hat{\alpha} \), the Lawless-Wang ridge regression estimates, \( \hat{\alpha}^* \), given by
\[ \hat{\alpha}^*_i = (Z'Z + kI)^{-1}Z'y \] where all the variables are as defined in chapter I. The \( \hat{\alpha}^*_i \), the \( i \)th ridge regression coefficient, \( \hat{\alpha}^*_i = \frac{\lambda_i}{\lambda_i + k} \hat{\alpha}_i = m_i \hat{\alpha}_i \) is a biased estimator since the expected value of \( \hat{\alpha}^*_i \) is not equal to \( \alpha_i \) and \( \text{MSE} (\hat{\alpha}^*_i) \neq V (\hat{\alpha}^*_i) \), but it is equal to the variance (\( \hat{\alpha}^*_i \)) plus some bias squared component. As noted earlier, \( L_1^2(\hat{\alpha}^*_i) = m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \alpha_i^2 \).

Information about the distribution of the Lawless-Wang ridge estimator is given in tables 11-15 and Appendix A. Before discussing the results in tables 11-15, it should be noted that it was hypothesized that tests of significance based upon the estimated squared error loss might be possible, provided that the bias squared component was not too large relative to the total loss. That is, if the bias squared component of the observed squared error loss, \( (1-m_i)^2 \alpha_i^2 \), to the total squared error loss, \( m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \alpha_i^2 \), is less than some specified percent, the Lawless-Wang ridge estimate, \( \hat{\alpha}^*_i \), is distributed approximately normally, the same as the OLS estimate, \( \hat{\alpha}_i \). Thus, under the above stated conditions, \( \hat{\alpha}^*_i \) is expected to follow the normal distribution with mean \( m_i \alpha_i \) and squared error loss = \( m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \alpha_i^2 \).

The alternative hypothesis could be stated as \( H_a \) the distribution of \( \hat{\alpha}^*_i \) does not
follow the normal distribution when bias squared is equal or less than the specified percent of the total loss. Data for testing these hypotheses are collected and presented in tables 11-15, and appendix A, tables A1-A5. Test statistics: To test the aforementioned hypotheses, the following statistical test is presented in some detail. The Kolmogorov-Smirnov Test: The Kolmogorov-Smirnov Test is a test of goodness of fit between observed sample values and some specified theoretical known distribution. This test has been explained in detail by Sigel (1956) and is known to be a more powerful test than the alternative $\chi^2$ test, defined as follows:

$$\chi^2 = \frac{n}{n-1d.f.} \sum_{i=1}^{n} \left( \frac{S_i - F_i}{F_i} \right)^2$$

where $S_i$ is the $i$th sample observation, $F_i$ is the $i$th theoretically expected observation for all $i = 1, 2, \ldots, n$. Testing procedures for the Kolmogorov-Smirnov Test, as outlined by Sigel (1956), are summarized in the following steps:

1. The cumulative frequency distribution function expected under the specified null hypothesis is completely constructed. The theoretical distribution assumed here is the normal distribution, i.e., specify $F_0(X) = P_r(X < x)$ for all $x$.

2. The cumulative frequency distribution function, $S_n = P_r(X < x)$ for the observed sample is arranged,
pairing each interval of $S_n(X)$ with its equivalent interval of $F_0(X)$.

3. Calculate the D value which is defined as follows:

$$D = \text{maximum} \left| F_0(X) - S_n(X) \right|,$$

which is the maximum absolute difference between the observed and the theoretically specified frequency distributions.

4. From the table provided by Sigel (1956), appendix 5, find the critical value of D for the prespecified level of significance, say 5%.

5. Reject the null hypothesis, $H_0$, if the critical value of D is less than or equal to the calculated D, or, fail to reject $H_0$ at the specified level of significance.

Table 11 contains information about the distribution of $\hat{\alpha}_1^*$ as compared to the standard t distribution for a very large number of observations ($n=\infty$) for the same 8 models ($\alpha_1 - \alpha_8$) used earlier with $\lambda = (.1, 1, 1.9)$ and $\sigma^2 = 4$. The results in table 11 show that only models $\alpha_7$ and $\alpha_4$ had the bias squared component $B^2$, equal or less than 10 percent of the total squared error loss, $L^2$, which could be specified as the null hypothesis to be tested. All the other models had a $B^2/L^2$ ratio greater than 10 percent. Models $\alpha_7$ and $\alpha_4$, being the most favorable models for ridge analysis, had D statistics of 0.0507 and 0.0512, respectively, as compared
Table 11. The cumulative distribution function of $\hat{a}_i^2$ for the eight models ($a_1-a_8$) with $\lambda (1., 1., 1.9)$, $\sigma^2 = 4$, $(\text{Bias})^2$

to total loss ratio ($B^2/L^2$) and the standard $t$ distribution with $N = m$.

<table>
<thead>
<tr>
<th>Models:</th>
<th>$a_1^2$</th>
<th>$a_2^2$</th>
<th>$a_3^2$</th>
<th>$a_4^2$</th>
<th>$a_5^2$</th>
<th>$a_6^2$</th>
<th>$a_7^2$</th>
<th>$a_8^2$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^2/L^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10,10,10)</td>
<td>.286</td>
<td>.219</td>
<td>.130</td>
<td>.094</td>
<td>.612</td>
<td>.365</td>
<td>.024</td>
<td>.920</td>
<td>0.000</td>
</tr>
<tr>
<td>(15/37.5,15/37.5)</td>
<td>.002</td>
<td>.000</td>
<td>.006</td>
<td>.023</td>
<td>.000</td>
<td>.028</td>
<td>.000</td>
<td>.000</td>
<td>.0013</td>
</tr>
<tr>
<td>(17.5,15/37.5)</td>
<td>.002</td>
<td>.024</td>
<td>.000</td>
<td>.000</td>
<td>.006</td>
<td>.000</td>
<td>.028</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>(17.5/37.5,15)</td>
<td>.085</td>
<td>.000</td>
<td>.004</td>
<td>.000</td>
<td>.000</td>
<td>.006</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>(17.5/37.5,15,15)</td>
<td>.056</td>
<td>.074</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.006</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>(12/12,12)</td>
<td>.130</td>
<td>.000</td>
<td>.116</td>
<td>.019</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>(12/12,12,12)</td>
<td>.023</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>(16/22/22)</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

a. $B^2/L^2$ is computed by $(1-m_1)^2a_1^2/(a_1^2m_1^2(1-m_1)^2a_1^2)$ where $m_1 = \frac{1}{2}$.

b. The cumulative distribution of $\hat{a}_i^2$ for the specified range, where $\phi$ is defined by

$$\hat{\phi} = (m_1^2a_1^2)^{1/2}(1-m_1)^2\hat{\sigma}_1^2$$

and $\hat{a}_i^2 = (a_i^2 - \text{Bias}(a_i^2)) > 0$. 


to a tabulated value of D of 0.0520 at the one percent
level of significance. Therefore, the test indicates
that we fail to reject the null hypothesis, concluding
that $\hat{\alpha}_1^*$ in models $\alpha_7$ and $\alpha_4$ are distributed approximately
normal with mean equal to $m_1\alpha_1$ and squared error loss
of $L_1^2$, as defined earlier. Inspection of table 11 for
the other models indicate that the distribution of $\hat{\alpha}_1^*$
was positively skewed with most of the observations
falling on the positive side of the distribution, unlike
the normal distribution which is symmetric about the
mean. This skewness of the $\hat{\alpha}_1^*$ distribution indicates
that $\hat{\alpha}_1^*$ has some unknown distribution differing from
the normal distribution.

In table 12, only models $\alpha_7$, $\alpha_4$ and $\alpha_3$ had the
ratio of bias squared to total squared error loss
($B^2/L^2$) less than 10 percent, which was chosen as the
hypothesized crucial percentage. The model $\alpha_3$ in table
11 has a $B^2/L^2$ ratio of 13 percent and $B^2/L^2$ in table
12 for the same model is 7 percent. Applying the
Kolmogorov-Smirnov Test to the data in table 12, cal-
culated D statistics for the models $\alpha_7$, $\alpha_4$ and $\alpha_3$ were
.0412, .0427 and .0515, respectively. These statistics
compared to the critical D value of 0.0520 for the one
percent level indicate that one would fail to reject
the null hypothesis that $\hat{\alpha}_1^*$ is distributed normally with
mean $m_1\alpha_1$ and squared error loss given by
Table 12. The cumulative distribution function of $\tilde{a}_1^+$ for the eight models, $(a_1^--a_8^+)$, with $a^2 = 2$, $\lambda = (1, 1.1, 1.2, 1.3, 1.4) = (10.1, 12.5/13.5, 15/16.5)$, and the standard $t$ distribution with $N = \infty$.  

<table>
<thead>
<tr>
<th>Models: $\lambda$</th>
<th>$\tilde{a}_1^+$</th>
<th>$\tilde{a}_2^+$</th>
<th>$\tilde{a}_3^+$</th>
<th>$\tilde{a}_4^+$</th>
<th>$\tilde{a}_5^+$</th>
<th>$\tilde{a}_6^+$</th>
<th>$\tilde{a}_7^+$</th>
<th>$\tilde{a}_8^+$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(10,10,10)$</td>
<td>$^{18}_1$</td>
<td>$^{18}_2$</td>
<td>$^{18}_3$</td>
<td>$^{18}_4$</td>
<td>$^{18}_5$</td>
<td>$^{18}_6$</td>
<td>$^{18}_7$</td>
<td>$^{18}_8$</td>
<td></td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>.167</td>
<td>.702</td>
<td>.070</td>
<td>.010</td>
<td>.044</td>
<td>.022</td>
<td>.012</td>
<td>.052</td>
<td>.000</td>
</tr>
<tr>
<td>$u_1 &lt; -3\hat{\phi}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
<td>0.000</td>
<td>0.0013</td>
</tr>
<tr>
<td>$-3\hat{\phi} &lt; u_1 &lt; -2\hat{\phi}$</td>
<td>0.017</td>
<td>0.000</td>
<td>0.036</td>
<td>0.021</td>
<td>0.001</td>
<td>0.007</td>
<td>0.027</td>
<td>0.001</td>
<td>0.0228</td>
</tr>
<tr>
<td>$-2\hat{\phi} &lt; u_1 &lt; -1\hat{\phi}$</td>
<td>0.086</td>
<td>0.006</td>
<td>0.147</td>
<td>0.116</td>
<td>0.021</td>
<td>0.074</td>
<td>0.147</td>
<td>0.003</td>
<td>0.1507</td>
</tr>
<tr>
<td>$-1\hat{\phi} &lt; u_1 &lt; -0.5\hat{\phi}$</td>
<td>0.187</td>
<td>0.085</td>
<td>0.259</td>
<td>0.305</td>
<td>0.079</td>
<td>0.185</td>
<td>0.302</td>
<td>0.122</td>
<td>0.3085</td>
</tr>
<tr>
<td>$-0.5\hat{\phi} &lt; u_1 &lt; 0.5\hat{\phi}$</td>
<td>0.432</td>
<td>0.235</td>
<td>0.640</td>
<td>0.689</td>
<td>0.324</td>
<td>0.435</td>
<td>0.650</td>
<td>0.166</td>
<td>0.6915</td>
</tr>
<tr>
<td>$0.5\hat{\phi} &lt; u_1 &lt; 1\hat{\phi}$</td>
<td>0.506</td>
<td>0.403</td>
<td>0.845</td>
<td>0.847</td>
<td>0.482</td>
<td>0.512</td>
<td>0.838</td>
<td>0.352</td>
<td>0.8413</td>
</tr>
<tr>
<td>$1\hat{\phi} &lt; u_1 &lt; 2\hat{\phi}$</td>
<td>0.812</td>
<td>0.712</td>
<td>0.966</td>
<td>0.979</td>
<td>0.747</td>
<td>0.798</td>
<td>0.973</td>
<td>0.749</td>
<td>0.9772</td>
</tr>
<tr>
<td>$2\hat{\phi} &lt; u_1 &lt; 3\hat{\phi}$</td>
<td>0.949</td>
<td>0.879</td>
<td>0.999</td>
<td>0.993</td>
<td>0.892</td>
<td>0.934</td>
<td>0.991</td>
<td>0.907</td>
<td>0.9987</td>
</tr>
<tr>
<td>$u_1 &gt; 3\hat{\phi}$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a. $B^2/L^2$ is computed by: $(1-m_1^2)^2n_1^2/(0.2^2m_1^2/\lambda_1^2(1-m_1^2)^2n_1^2)$, where, $m_1 = \frac{1}{\lambda_1^2k}$.  

b. The cumulative distribution of $\tilde{a}_1^+$ for specified ranges, where:  

$\hat{\phi} = (m_1^2/\lambda_1^2+1-M_1^2)^{0.5}$, and $\tilde{\phi}_1 = (a^2-\sqrt{\mu}) > 0$. 


The cumulative frequency distribution of $\hat{a}_1$ in the same 8 models presented earlier, with $\sigma^2 = 1$, and $\lambda = (0.1, 1, 1.9)$ are shown in table 13. Models $a_7$, $a_4$, $a_3$ and $a_1$ had a bias squared to total squared loss ratio less than the hypothesized crucial ratio of 10 percent. The D statistics (computed as the maximum absolute difference between the observed frequency distribution and the hypothesized theoretical t distribution) for models $a_7$, $a_4$, $a_3$ and $a_1$ are, 0.0245, 0.0385, 0.0505 and .0517, respectively. The critical tabulated Kolmogorov-Smirnov D statistics at the one percent level is .0520, which indicates that one would fail to reject the null hypothesis that $\hat{a}_1$ is distributed approximately normal, with mean $m_1a_1$, and squared error loss $[m_1^2\sigma^2/\lambda_1+(1-m_1)^2a_1^2]$. Note that using the same test statistics one would reject the hypothesis that $\hat{a}_1$ in models with $B^2/L^2 < 10\%$ is distributed normal, since by inspection of table 13 it is clear that the $\hat{a}_1$ distribution is positively skewed with almost no observations on the left side of the distributions in models $a_9$ with $B^2/L^2 = .75$, and $a_2$ with $B^2/L^2 = .54$. It should be emphasized that when $B^2/L^2$ is relatively large, the distribution of $\hat{a}_1$ in table 13 indicates an extreme lack of fit between the t distribution and the observed distributions for models with very large bias squared.
components.

The information in table 14 reveals that all the models "favorable" to ridge analysis, namely, $\alpha_7$, $\alpha_4$, $\alpha_3$, and $\alpha_1$ and even some of the "unfavorable" models, namely, $\alpha_6$, had $B^2/L^2$ values which are less than the hypothesized value ($B^2/L^2 < 10\%$). The Kolmogorov-Smirnov D statistics, calculated for models $\alpha_7$, $\alpha_6$, $\alpha_4$, $\alpha_3$, and $\alpha_1$ are 0.024, 0.0305, 0.0273, 0.0513, and 0.0207, respectively. Using the Kolmogrov-Smirnov D Test, one would fail to reject the hypothesis that $\hat{\alpha}_1$ was drawn from an approximately normal distribution. Note that as the $B^2/L^2$ ratio gets smaller, inspection of table 14 indicates an improved goodness of fit between the $\hat{\alpha}_1$ distribution and the $t$ distribution, and a worse fit when $B^2/L^2$ increases, because more and more observations are falling on the positive side of the distribution, indicating a significant lack of fit between the $\hat{\alpha}_1$ distribution and the standard $t$ distribution, using 1,000 sample observations and with $n$ tending to $\infty$ for the $t$ distribution.

Comparisons of the cumulative distribution function of $\hat{\alpha}_1$ (for the same 8 models ($\alpha_1$-$\alpha_8$) given earlier) with the standard $t$ distribution, can again be made from the information given in table 15. In all the models, except the most "unfavorable" models ($\alpha_8$ and $\alpha_2$) $\hat{\alpha}_1$ had a bias squared component to total loss ratio less than or
Table 13. The cumulative distribution function of $a_1^2$ for the eight models, $(x_1-x_8)$ with $\sigma^2 = 1, \bar{x} = (1, 1, 1.9)$, the $(\text{Bias})^2$/total loss ratio and the standard t distribution with $N = n$.

<table>
<thead>
<tr>
<th>Models:</th>
<th>$a_1^2$</th>
<th>$a_2^2$</th>
<th>$a_3^2$</th>
<th>$a_4^2$</th>
<th>$a_5^2$</th>
<th>$a_6^2$</th>
<th>$a_7^2$</th>
<th>$a_8^2$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_2^2/L^2$</td>
<td>.091</td>
<td>.540</td>
<td>.036</td>
<td>.015</td>
<td>.283</td>
<td>.126</td>
<td>.006</td>
<td>.747</td>
<td>.000</td>
</tr>
<tr>
<td>$a_1 &lt; 3^\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>.0013</td>
</tr>
<tr>
<td>$-3^\hat{a} &lt; a_1 &lt; -2^\hat{a}$</td>
<td>0.016</td>
<td>0.000</td>
<td>0.009</td>
<td>0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.019</td>
<td>0.000</td>
<td>.0228</td>
</tr>
<tr>
<td>$-2^\hat{a} &lt; a_1 &lt; -1^\hat{a}$</td>
<td>0.107</td>
<td>0.070</td>
<td>0.117</td>
<td>0.133</td>
<td>0.022</td>
<td>0.061</td>
<td>0.140</td>
<td>0.001</td>
<td>.1597</td>
</tr>
<tr>
<td>$-1^\hat{a} &lt; a_1 &lt; -.5^\hat{a}$</td>
<td>0.250</td>
<td>0.377</td>
<td>0.258</td>
<td>0.270</td>
<td>0.103</td>
<td>0.199</td>
<td>0.289</td>
<td>0.004</td>
<td>.3085</td>
</tr>
<tr>
<td>$-.5^\hat{a} &lt; a_1 &lt; .5^\hat{a}$</td>
<td>0.640</td>
<td>0.737</td>
<td>0.649</td>
<td>0.643</td>
<td>0.588</td>
<td>0.659</td>
<td>0.667</td>
<td>.285</td>
<td>.6915</td>
</tr>
<tr>
<td>$-.5^\hat{a} &lt; a_1 &lt; .1^\hat{a}$</td>
<td>0.788</td>
<td>0.969</td>
<td>0.892</td>
<td>0.827</td>
<td>0.892</td>
<td>0.866</td>
<td>0.812</td>
<td>.574</td>
<td>.8413</td>
</tr>
<tr>
<td>$1^\hat{a} &lt; a_1 &lt; 2^\hat{a}$</td>
<td>0.951</td>
<td>0.968</td>
<td>0.961</td>
<td>0.951</td>
<td>0.965</td>
<td>0.913</td>
<td>0.982</td>
<td>.929</td>
<td>.9772</td>
</tr>
<tr>
<td>$2^\hat{a} &lt; a_1 &lt; 3^\hat{a}$</td>
<td>0.980</td>
<td>0.982</td>
<td>0.970</td>
<td>0.982</td>
<td>0.977</td>
<td>0.990</td>
<td>0.999</td>
<td>.988</td>
<td>.9987</td>
</tr>
<tr>
<td>$a_1 &gt; 3^\hat{a}$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a. $B_2^2/L^2$ is computed by $(1-m_1)^2 u_1^2/(a_1^2 m_1^2/L_1 + (1-m_1)^2 u_1^2)$ where $m_1 = \frac{1}{1+\alpha}$.

b. The cumulative distribution of $\hat{a}_1$ for specified ranges where:

$\hat{a}_1 = \left[m_1^2 a_1^2/L_1 + (1-H)^2 \theta_1 \right]^{1.5}$ and $\hat{a}_1 = (a_1^2 - \hat{a}_1) \geq 0$. 

05
equal to the previously hypothesized 10 percent. The 
Kolmogorov-Smirnov D statistics for models \( a_1, a_3, a_4, \\
a_5, a_6, \) and \( a_7 \) are 0.0310, 0.038, 0.0518, 0.028, 0.0420, 
and 0.0510, respectively. The crucial tabulated D 
statistic at the one percent confidence level is equal to 
0.0520, which indicates that the \( \hat{\alpha}_1 \) distribution is not 
significantly different from the normal distribution for 
these models. It should be noted from table 15 that the 
goodness of fit between the \( \hat{\alpha}_1 \) distribution and the 
theoretical t distribution improves as \( B^2/L^2 \) decreases. 
Of course, \( \hat{\alpha}_1 \) is equal to \( \hat{\alpha}_1 \) when \( B^2/L^2 = 0 \), and would 
be distributed normally, the same as \( \hat{\alpha}_1 \) in this case.

The cumulative frequency distribution of \( \hat{\alpha}_1 \) in 
models \( (\alpha_1-\alpha_8) \) with \( \lambda = (0.01, 1, 1.99) \) and \( \sigma^2 = 4, 2, 1, \\
0.5 \) and 0.25 are given in tables A1-A5 in appendix A. 
The multicollinearity problem in tables A1-A5 is much 
more severe than for the models in tables 11-15 since 
\( VIF_1 \) for the models in tables A1-A5 is approximately 
equal to 100, compared to \( VIF_1 \) of 10 for the same models 
in tables 11-15. Table 16 summarizes the results of the 
hypothesis testing from tables A1-A5 in appendix A, 
again utilizing the Kolmogorov-Smirnov statistical testing 
procedure. It should be noted that \( \hat{\alpha}_1 \) in all the 8 
models \( (\alpha_1-\alpha_8) \) with \( \lambda = (.01, 1, 1.99) \) and \( \sigma^2 = 4 \) and 
2 had a biased squared to total squared loss ratio greater 
than ten percent, and that for these models the
Table 14. The cumulative distribution function of $\hat{a}_1^*$ for the eight models $(\sigma_1^2, \lambda_1)$, with $\sigma^2 = .5$, $\lambda = (1, 1.9)$, the (Bias)$^2$/total loss ratio and the standard t distribution with $N = n$.

<table>
<thead>
<tr>
<th>Models: $\sigma^2/\lambda^2$</th>
<th>$a_{1}^*$</th>
<th>$a_{2}^*$</th>
<th>$a_{3}^*$</th>
<th>$a_{4}^*$</th>
<th>$a_{5}^*$</th>
<th>$a_{6}^*$</th>
<th>$a_{7}^*$</th>
<th>$a_{8}^*$</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10, 10, 10)</td>
<td>0.048</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.013</td>
</tr>
<tr>
<td>(15, 37.5, 37.5)</td>
<td>0.047</td>
<td>0.095</td>
<td>0.198</td>
<td>0.049</td>
<td>0.013</td>
<td>0.027</td>
<td>0.069</td>
<td>0.000</td>
<td>0.0228</td>
</tr>
<tr>
<td>(12.5, 15, 15)</td>
<td>0.138</td>
<td>0.015</td>
<td>0.351</td>
<td>0.323</td>
<td>0.198</td>
<td>0.298</td>
<td>0.315</td>
<td>0.008</td>
<td>0.1587</td>
</tr>
<tr>
<td>(12, 12, 12)</td>
<td>0.678</td>
<td>0.524</td>
<td>0.718</td>
<td>0.616</td>
<td>0.661</td>
<td>0.703</td>
<td>0.470</td>
<td>0.6915</td>
<td></td>
</tr>
<tr>
<td>(12, 12, 12)</td>
<td>0.817</td>
<td>0.790</td>
<td>0.839</td>
<td>0.845</td>
<td>0.833</td>
<td>0.827</td>
<td>0.816</td>
<td>0.8413</td>
<td></td>
</tr>
<tr>
<td>(12, 12, 12)</td>
<td>0.992</td>
<td>0.994</td>
<td>0.994</td>
<td>0.990</td>
<td>0.990</td>
<td>0.987</td>
<td>0.991</td>
<td>0.9722</td>
<td></td>
</tr>
<tr>
<td>(12, 12, 12)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.998</td>
<td>1.000</td>
<td>0.999</td>
<td>1.000</td>
<td>0.9987</td>
<td></td>
</tr>
<tr>
<td>(12, 12, 12)</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

a. $B^2/L^2$ is computed by $(1-m_1)\sigma^2/\lambda_1^2+\lambda_1(1-m_1^0)\sigma_1^2$ where $m_1 = \lambda_1^{1-k}$.

b. The cumulative distribution of $\hat{a}_1^*$ for specified ranges, where

$\dot{c} = m_1^0\sigma^2/\lambda_1(1-M_1)^{1.5}$ and $\ddot{a}_1 = \sigma_1^2\mathcal{V}(\hat{a}_1)$.
Kolmogorov-Smirnov Test indicated that one would reject the hypothesis that \( \hat{\alpha}_1^* \) was distributed normally at the one percent significance level. At \( \sigma^2 = 1.0 \), only \( \alpha_7 \), the most "favorable" model for ridge analysis, had a \( B^2/L^2 \) ratio less than the hypothesized percentage. The calculated D statistics for model \( \alpha_7 \) in table A3 is 0.040, indicating that one fails to reject the hypothesis that \( \hat{\alpha}_1^* \) is distributed approximately normal, with mean \( \mu_1 \alpha_1 \) and squared error loss, \( [m_1^2 \sigma^2/\lambda_1 + (1-m_1)^2 \alpha_1^2] \). At \( \sigma^2 = 0.5 \) only \( \alpha_7 \) and \( \alpha_4 \) had \( B^2/L^2 \) less than the ten percent hypothesized crucial value, and these two models had calculated D statistics of 0.022 and 0.043, respectively. Using the Kolmogrov-Smirnov statistical test, one would fail to reject the null hypothesis for these two models (\( \alpha_7 \) and \( \alpha_4 \)), concluding that \( \hat{\alpha}_1^* \) was distributed approximately normal. At \( \sigma^2 = 0.25 \), three models, \( \alpha_7 \), \( \alpha_4 \) and \( \alpha_3 \), characterized by being the three most "favorable" models for ridge analysis, had \( B^2/L^2 \) less than ten percent. The calculated Kolmogorov-Smirnov Test statistics for these three models are 0.042, 0.020, and 0.028, respectively. The Kolmogorov-Smirnov test results indicate that one would also fail to reject the null hypothesis for these three models at the one percent level of significance.

Unlike the OLS estimates, it is very difficult to make a general statement about the distribution of the
Table 15. The cumulative distribution function of $\hat{\alpha}_1^*$ for the eight models ($a_1-a_8$), with $\lambda (0.1, 1, 1.9)$, $\sigma^2 = .25$, $(\text{Bias})^2$/
total loss ratio and the $t$ distribution with $N = n$.

<table>
<thead>
<tr>
<th>Models:</th>
<th>$u_1^*$</th>
<th>$u_2^*$</th>
<th>$u_3^*$</th>
<th>$u_4^*$</th>
<th>$u_5^*$</th>
<th>$u_6^*$</th>
<th>$u_7^*$</th>
<th>$u_8^*$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^2/L^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.024</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.0013</td>
</tr>
<tr>
<td>0.002</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.0228</td>
</tr>
<tr>
<td>0.048</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.1587</td>
</tr>
<tr>
<td>0.117</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.3085</td>
</tr>
<tr>
<td>0.187</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.6915</td>
</tr>
<tr>
<td>0.115</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.8413</td>
</tr>
<tr>
<td>0.120</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.9772</td>
</tr>
<tr>
<td>0.001</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.035</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.419</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a. $B^2/L^2$ is computed by $(1-m_1)^2a^2_1/(0^2m_1^2/1+\lambda_{1}(1-m_1)^2a^2)$ where $m_1 = \frac{1}{\lambda_{1}+k}$.

b. The cumulative distribution for $\hat{\alpha}_1^*$ for the specified ranges, where:

$\hat{\phi} = \sqrt{\lambda_{1}^{2}} [1-(1-M)]^{2}0_{1}^{4}1.5$ and $\hat{\alpha}_1^* = (\hat{\alpha}_1^* - V(\hat{\alpha}_1^*)) > 0$. 


Lawless-Wang ridge estimate, $\hat{\alpha}_i^*$, because of the following two facts:

1. The ridge estimates $\hat{\alpha}^*$ are biased estimates of $\alpha$ since, as pointed out earlier, $E(\hat{\alpha}_i^*) = m_1 \alpha_i$ which is equal to $\alpha_i$ only when $k = 0 (m_1 = 1)$.

2. The ridge estimate $\hat{\alpha}^*$ and its squared error loss, $L_1^2(\hat{\alpha}^*)$, are not generally independent, contrary to the OLS estimate, $\hat{\alpha}$, and its variance, $V(\hat{\alpha})$, which are independent. The summary of the results of tables 11-15 and A1-A5 indicates that under some conditions a $t^*$ statistic could be used, where

$$ t^* = m_1 \hat{\alpha}_i / (m_1^2 \sigma^2 / \lambda_i + (1 - m_1)^2 \theta) $$

where all the variables are as defined earlier. The previous results indicate that $t^*$ is distributed approximately as $t$ with $v$ degrees of freedom. However, such use of $t^*$ requires that the bias squared component to total squared loss, $B^2/L^2$, defined as

$$ B^2/L^2 = (1 - m_1)^2 \alpha_i / (m_1^2 \sigma^2 / \lambda_i + (1 - m_1)^2 \alpha_i^2), $$

is not large, say, less than 10 percent. But this condition is usually satisfied only if the coefficient $\alpha_i$ corresponding to the smallest eigenvalue, $\lambda_i$, is not too large relative to the other true $\alpha_i$ values. That is, the condition is usually satisfied only for models generally "favorable" for ridge analysis. It should be emphasized that the above statement is only true for the Lawless-Wang ridge estimators and for models where the ratio $B^2/L^2$ is equal to or less than 10 percent. Generally speaking, $m_1 \hat{\alpha}_i$ is
Table 16. Summary of Tables A1-A5, showing the Kolmogorov-Smirnov calculated $D$ statistics, bias squared to total squared loss, $B^2/L^2$, the critical $D$ value, and the decision about the stated hypothesis.

<table>
<thead>
<tr>
<th>Models</th>
<th>$B^2/L^2$</th>
<th>D statistics$^b$</th>
<th>D value at one percent $N = 1000$</th>
<th>Decision about $H_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_7$ ($\sigma^2 = 1$)</td>
<td>0.055</td>
<td>0.040</td>
<td>0.0520</td>
<td>Fail to reject $H_0$.</td>
</tr>
<tr>
<td>$\alpha_7$ ($\sigma^2 = .5$)</td>
<td>0.028</td>
<td>0.022</td>
<td>0.0520</td>
<td>$\approx$</td>
</tr>
<tr>
<td>$\alpha_4$ ($\sigma^2 = .5$)</td>
<td>0.067</td>
<td>0.043</td>
<td>0.0520</td>
<td>$\approx$</td>
</tr>
<tr>
<td>$\alpha_7$ ($\sigma^2 = .25$)</td>
<td>0.014</td>
<td>0.042</td>
<td>0.0520</td>
<td>$\approx$</td>
</tr>
<tr>
<td>$\alpha_4$ ($\sigma^2 = .25$)</td>
<td>0.035</td>
<td>0.020</td>
<td>0.0520</td>
<td>$\approx$</td>
</tr>
<tr>
<td>$\alpha_3$ ($\sigma^2 = .25$)</td>
<td>0.086</td>
<td>0.028</td>
<td>0.0520</td>
<td>$\approx$</td>
</tr>
</tbody>
</table>

*a. $B^2/L^2 = \frac{(1-m_1)^2\sigma_1^2}{\{m_1\sigma_1^2/\lambda_1 + (1-m_1)^2\sigma_1^2\}}$ where $m_1 = \frac{\lambda_1}{\lambda_1^* + k}$.*

*b. Kolmogorov-Smirnov $D$ statistics $D = \max\{|F_0(X) - S_n(X)|\}$ where $F_0(X)$ is the cumulative distribution hypothesized and $S_n(X)$ is the observed cumulative distribution.
not independent of $L_1(\hat{\theta}_1^*)$, and $t^*$ would not be distributed as $t$ with $v$ degrees of freedom.

One final important problem that needs to be addressed in this chapter is how well the researcher would be able to predict the value of the ratio of the squared bias to the total squared error loss, in order to decide whether or not to use the suggested $t^*$ statistic to test the significance of the Lawless-Wang ridge coefficients. In light of the results of tables (1-10), I propose that the ratio of the squared bias to the total squared loss be estimated using the sample data as follows:

Let estimated $B^2/L^2 = \frac{(1-m_i)^2\hat{\theta}_i}{(m_i^2\hat{\sigma}^2/\lambda_i + (1-m_i)^2\hat{\theta}_i)}$ where all the variables are as defined earlier in this chapter. Table 17 contains information about the true $B^2/L^2$ ratio, the estimated $B^2/L^2$ ratio, and the number of observations in 1000 experiments with an estimated $B^2/L^2$ ratio less than the critical hypothesized ratio (10 percent) for the use of the $t^*$ statistic. (Recall that the $t^*$ statistic was only for testing the significance of the Lawless-Wang ridge coefficients.) Results for the same 8 models ($\alpha_1 - \alpha_8$) analyzed earlier in this chapter, with $\lambda = (0.10, 1, 1.90)$ and $\sigma^2$ values at 0.25, 0.50, 1, 2, and 4, are given in Table 17.

Results in Table 17 show that the proposed estimation of the bias squared to total squared error loss ratio from
Table 17. The true $B'^2/L'^2$, the estimated $B'^2/L'^2$ and the number of observations with estimated $B'^2/L'^2$ less than 10 percent for the 8 models $(a_i - a_k)$ with $\lambda = (0.1, 1, 1.9)$ and $\sigma^2 = 4, 2, 1, .5, 2$.

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$a_5$</th>
<th>$a_6$</th>
<th>$a_7$</th>
<th>$a_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(10, 10, 10)$</td>
<td>$(15, \sqrt{7.5}, \sqrt{7.5})$</td>
<td>$(\sqrt{7.5}, 15, \sqrt{7.5})$</td>
<td>$(\sqrt{7.5}, \sqrt{7.5}, 15)$</td>
<td>$(12, 12, \sqrt{72})$</td>
<td>$(12, \sqrt{72}, 12)$</td>
<td>$(\sqrt{72}, 12, 12)$</td>
<td>$(16, \sqrt{72}, \sqrt{72})$</td>
</tr>
<tr>
<td>$\sigma^2 = 0.25$</td>
<td>$B'^2/L'^2_a$</td>
<td>0.025</td>
<td>0.227</td>
<td>0.009</td>
<td>0.004</td>
<td>0.089</td>
<td>0.035</td>
</tr>
<tr>
<td>obs. $&lt; 10%$</td>
<td>1000</td>
<td>000</td>
<td>1000</td>
<td>721</td>
<td>1000</td>
<td>000</td>
<td>1000</td>
</tr>
<tr>
<td>$\sigma^2 = 0.50$</td>
<td>$B'^2/L'^2_b$</td>
<td>0.048</td>
<td>0.370</td>
<td>0.018</td>
<td>0.008</td>
<td>0.165</td>
<td>0.067</td>
</tr>
<tr>
<td>obs. $&lt; 10%$</td>
<td>991</td>
<td>000</td>
<td>1000</td>
<td>125</td>
<td>925</td>
<td>1000</td>
<td>000</td>
</tr>
<tr>
<td>$\sigma^2 = 1.0$</td>
<td>$B'^2/L'^2_c$</td>
<td>0.084</td>
<td>0.509</td>
<td>0.036</td>
<td>0.015</td>
<td>0.283</td>
<td>0.126</td>
</tr>
<tr>
<td>obs. $&lt; 10%$</td>
<td>762</td>
<td>000</td>
<td>934</td>
<td>60</td>
<td>384</td>
<td>1000</td>
<td>000</td>
</tr>
<tr>
<td>$\sigma^2 = 2.0$</td>
<td>$B'^2/L'^2_d$</td>
<td>0.167</td>
<td>0.702</td>
<td>0.070</td>
<td>0.029</td>
<td>0.441</td>
<td>0.223</td>
</tr>
<tr>
<td>obs. $&lt; 10%$</td>
<td>321</td>
<td>11</td>
<td>727</td>
<td>927</td>
<td>113</td>
<td>230</td>
<td>957</td>
</tr>
<tr>
<td>$\sigma^2 = 4.0$</td>
<td>$B'^2/L'^2_e$</td>
<td>0.286</td>
<td>0.825</td>
<td>0.130</td>
<td>0.058</td>
<td>0.612</td>
<td>0.365</td>
</tr>
<tr>
<td>obs. $&lt; 10%$</td>
<td>319</td>
<td>100</td>
<td>623</td>
<td>849</td>
<td>234</td>
<td>181</td>
<td>033</td>
</tr>
</tbody>
</table>

a. $B'^2/L'^2 = \frac{(1-m_1)^2 a_i^2}{[m_1^2 \sigma^2/\lambda_1 + (1-m_1)^2 a_i^2]}$

b. estimated $B'^2/L'^2 = \frac{(1-m_i)^2 \hat{a}}{[m_i^2 \sigma^2/\lambda_1 + (1-m_i)^2 \hat{a}]}$

c. observations with est. $B'^2/L'^2$ less than 10 percent.
the sample data performed very well in almost all the cases. The average estimated $B^2/L^2$ ratio and the true $B^2/L^2$ ratio were identical for four out of 8 models at $\sigma^2 = .25$, indicating an excellent performance by the proposed method for estimating $B^2/L^2$. Generally speaking, the estimated $B^2/L^2$ and the true $B^2/L^2$ for all the 8 models in Table 17 differed from the true $B^2/L^2$ by only a few percentage points, and this difference increased as $\sigma^2$ increased. However, this difference does not appear to be important for practical purposes.

Another criterion, besides the closeness of the average estimated $B^2/L^2$ to the true $B^2/L^2$, is the number of successes where success is defined as making the right decision about whether or not to use $t^*$. For $\sigma^2 = 0.25$ in Table 17, decisions were 100 percent correct in all the models except for model $a_5$ where there would have been only 721 correct decisions out of 1000 experiments to use the $t^*$ statistic. This result occurred for $a_5$ because the true $B^2/L^2$ ratio was marginal in this case.

For $\sigma^2 = 0.5$, the decision about the use of the $t^*$ statistic was 100 percent correct for models $a_2, a_3, a_4, a_7,$ and $a_8$, and the success rate for models $a_1, a_5,$ and $a_6$ were 99.1, 87.5, and 97.5 percent, respectively.

It can be observed from Table 17 that the success rate for models $a_1 - a_8$ at the larger values of $\sigma^2$ ($\sigma^2 = 1, 2,$ and 4) decreased compared to the results with $\sigma^2 = 0.50$ or
0.25. However, these performances are still very good considering the variation due to individual sample observations. It should be emphasized that the excellent performance of the proposed method for estimating the ratio of the bias squared to the total squared error loss of the Lawless-Wang ridge coefficients is closely related to the results of Tables 6-10 where the squared error loss from the sample data were accurately estimated.

The observed $B^2/L^2$, the estimated $B^2/L^2$, and the number of observations (out of 1000 experiments) with estimated $B^2/L^2 \leq 10$ percent for models $\alpha_1 - \alpha_8$ with $\sigma^2$ values of 0.25, 0.50, 1.0, 2.0 and 4.0 are shown in Table 18. At the $\sigma^2$ value of 0.25, the estimates of bias squared to total squared error loss were fairly good for all the models. Except for the somewhat marginal cases of models $\alpha_1$ and $\alpha_3$, the researcher would have been able to make the correct decision about the use of the $t^*$ statistic better than 75 percent of the time. The performance of the proposed method was fairly accurate for models $\alpha_7, \alpha_4$, and $\alpha_3$ for $\sigma^2 = 0.25$ where the use of the $t^*$ statistic would be expected, and for these three models the correct decision rate was relatively high, ranging from about 68 to 92 percent.

It should also be noted that for $\sigma^2 = 0.25$ and at relatively high $B^2/L^2$ ratios, the results of Table 18
Table 10. The true $B^2/L^2$, the estimated $B^2/L^2$ and the number of observations with estimated $B^2/L^2$ less than or equal to 10 percent for the eight models $(a_i - a_s)$ with $\lambda^2 = (0.01, 0.1, 0.9)$

<table>
<thead>
<tr>
<th>$a_i = (10, 10, 10)$</th>
<th>$a_i = (15, 15, 15)$</th>
<th>$a_i = (7.5, 7.5, 7.5)$</th>
<th>$a_i = (12, 12, 12)$</th>
<th>$a_i = (12, 12, 12)$</th>
<th>$a_i = (12, 12, 12)$</th>
<th>$a_i = (16, 16, 16)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B^2/L^2$</td>
<td>$0.2000$</td>
<td>$0.0947$</td>
<td>$0.0857$</td>
<td>$0.0345$</td>
<td>$0.5305$</td>
<td>$0.2647$</td>
</tr>
<tr>
<td>$\sigma^2 B^2/L^2$</td>
<td>$0.16585$</td>
<td>$0.0704$</td>
<td>$0.0846$</td>
<td>$0.0375$</td>
<td>$0.4205$</td>
<td>$0.2202$</td>
</tr>
<tr>
<td>$\text{obs. with }$</td>
<td>$392$</td>
<td>$33$</td>
<td>$679$</td>
<td>$872$</td>
<td>$133$</td>
<td>$248$</td>
</tr>
<tr>
<td>$B^2/L^2 &lt; 10%$</td>
<td>$0.333$</td>
<td>$0.886$</td>
<td>$0.158$</td>
<td>$0.067$</td>
<td>$0.693$</td>
<td>$0.419$</td>
</tr>
<tr>
<td>$\sigma^2 B^2/L^2$</td>
<td>$0.231$</td>
<td>$0.692$</td>
<td>$0.145$</td>
<td>$0.076$</td>
<td>$0.471$</td>
<td>$0.309$</td>
</tr>
<tr>
<td>$\text{obs. } &lt; 10%$</td>
<td>$326$</td>
<td>$140$</td>
<td>$513$</td>
<td>$722$</td>
<td>$274$</td>
<td>$210$</td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>$0.500$</td>
<td>$0.939$</td>
<td>$0.273$</td>
<td>$0.125$</td>
<td>$0.819$</td>
<td>$0.590$</td>
</tr>
<tr>
<td>$\sigma^2 B^2/L^2$</td>
<td>$0.275$</td>
<td>$0.596$</td>
<td>$0.210$</td>
<td>$0.136$</td>
<td>$0.450$</td>
<td>$0.356$</td>
</tr>
<tr>
<td>$\text{obs. } &lt; 10%$</td>
<td>$439$</td>
<td>$231$</td>
<td>$533$</td>
<td>$722$</td>
<td>$210$</td>
<td>$375$</td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>$0.667$</td>
<td>$0.969$</td>
<td>$0.429$</td>
<td>$0.223$</td>
<td>$0.900$</td>
<td>$0.742$</td>
</tr>
<tr>
<td>$\sigma^2 B^2/L^2$</td>
<td>$0.298$</td>
<td>$0.470$</td>
<td>$0.262$</td>
<td>$0.205$</td>
<td>$0.405$</td>
<td>$0.462$</td>
</tr>
<tr>
<td>$\text{obs. } &lt; 10%$</td>
<td>$519$</td>
<td>$502$</td>
<td>$561$</td>
<td>$556$</td>
<td>$443$</td>
<td>$439$</td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>$0.800$</td>
<td>$0.984$</td>
<td>$0.600$</td>
<td>$0.364$</td>
<td>$0.948$</td>
<td>$0.852$</td>
</tr>
<tr>
<td>$\sigma^2 B^2/L^2$</td>
<td>$0.416$</td>
<td>$0.491$</td>
<td>$0.392$</td>
<td>$0.258$</td>
<td>$0.458$</td>
<td>$0.466$</td>
</tr>
<tr>
<td>$\text{obs. } &lt; 10%$</td>
<td>$544$</td>
<td>$540$</td>
<td>$569$</td>
<td>$573$</td>
<td>$524$</td>
<td>$487$</td>
</tr>
</tbody>
</table>

a. $B^2/L^2 = \frac{(1-m_1)^2 \sigma^2}{[m_1 \sigma^2 / \lambda + (1-m_1) \sigma^2]}$

b. estimated $B^2/L^2 = \frac{(1-m_1)^2 \theta}{[m_1 \sigma^2 / \lambda + (1-m_1) \theta]}$

c. number of observations with est. $B^2/L^2 \leq 10 \%$.
indicate that there would have been a fairly low probability of using the $t^*$ statistic when it was inappropriate since this error rate would have been only about .018 for model $\alpha_8$ and about 0.033 for model $\alpha_2$, the two most unfavorable models for ridge regression in these experiments. At $\sigma^2 = 0.50$, the estimated $B^2/L^2$ also was relatively close to the observed $B^2/L^2$; however, for model $\alpha_3$ there was about a 51 percent probability of making the wrong decision (to use $t^*$) due to the fact that $\alpha_3$ had a marginal value of $B^2/L^2$ of about 0.158. Except for the marginal case of model $\alpha_3$, the proposed criteria performed fairly well, based on the averages as well as based on individual experiments. At $\sigma^2$ value of 1.0, the proposed method gave an underestimate of the $B^2/L^2$ ratio for all the models, except for models $\alpha_7$ and $\alpha_4$ where it tended to overestimate $B^2/L^2$. As far as making the decision about the use of the $t^*$ statistic based on the sample data, the performance of the proposed method was fairly satisfactory with rates of success ranging between 56 percent to about 78 percent for all the models, except models $\alpha_3$ and $\alpha_4$ which were two marginal cases. It should be noted that the proposed method performed poorly for almost all the models at $\sigma^2 = 2$ and 4 which may be attributed to the same fact noted earlier from Tables 1-5, that is, the proposed method has a tendency to overestimate the observed squared
error loss $M_1$. Despite its limitations, namely, not performing well in all the models in Table 18 with $\sigma^2$ values of 2 and 4, the proposed method results were encouraging for all the models in Table 17 and Table 18, except at the two $\sigma^2$ values of 2 and 4. It should be noted that at $\sigma^2 = 2$, the variance of $\hat{\alpha}_1$ in Table 18 is approximately equal to $\sigma^2/\lambda_1 = 2/.01 = 200$ and at $\sigma^2 = 4$, $v(\alpha_1)$ is equal to $\sigma^2/\lambda_1 = 4/.01 = 400$, which indicates a very high degree of multicollinearity. It should also be mentioned that the probability that one would be making the correct decision about the use of $t^*$ statistic is greater than 0.50 for all the 'favorable' models in Table 18 at $\sigma^2 = 2$ and $\sigma^2 = 4$. 
IV. APPLICATION TO AN EMPIRICAL ECONOMIC MODEL

Nature of Economic Data

Unfortunately, good economic data from primary sources are seldom available. Instead, economists must often turn to secondary data sources, such as the U.S. Census of Agriculture and economic and business survey reports. These are almost always available only on an aggregate basis, such as by counties and states, and one has to use these data in spite of the fact that the aggregation may result in loss of information and cause multicollinearity problems. To solve such multicollinearity problems, it would be better to return to the original data whenever possible. When such a return is not possible or practical, different techniques to cope with multicollinearity have been suggested and tested via simulations, as in the preceding chapter.

The main objective of this chapter is to apply some of the results of the previous chapter to demonstrate their usefulness in coping with a practical economic problem.

Estimating the Marginal Value Productivity of Irrigation Water

Economists and others are interested in the concept of the marginal value productivity of economic resources on efficiency grounds, as well as other reasons. Since the
water resource used in agriculture is a very important economic variable from the regional as well as the national interest, numerous attempts have been made to arrive at a value for water used in irrigated agriculture. The marginal value productivity, MVP, for a factor of production, $x_i$, is defined by first partial derivative of the total value product function, $y$, with respect to $x_i$. That is,

$$MVP_{x_i} = \frac{\partial y(x_1, x_2, \ldots, x_p)}{\partial x_i}$$

for $i = 1, 2, \ldots, p$.

When the total value product function is from a Cobb-Douglas production function, the estimated coefficients are a direct measure of the product elasticities, and the estimated marginal value productivity of $x_i$ is simply given by

$$MVP_{x_i} = \left(\frac{\hat{y}}{\hat{x}_i}\right) \hat{\beta}_i$$

for $i = 1, 2, \ldots, p$.

The importance of obtaining a reliable estimate of $\hat{\beta}$ motivated the discussion in this chapter.

The pioneering work by Ruttan (1965) will be discussed in this chapter from the viewpoint of the empirical estimation problem involved and the possible improvement in these estimates by the use of ridge regression analysis. Ruttan used U.S. Census agricultural data as the main source of data in his formulation.
The Model

A production function model with six important explanatory variables was originally formulated. In functional form the model was:

\[ Y = f(x_1, x_2, x_3, x_4, x_5, x_6) \]

where \( Y \) denotes the total value of all farm products sold;

- \( x_1 \) is the number of units of family and hired labor;
- \( x_2 \) is the number of tractors on the farm;
- \( x_3 \) is value of livestock investment in the farm;
- \( x_4 \) is acres of irrigated cropland;
- \( x_5 \) is acres of non-irrigated cropland; and
- \( x_6 \) is the annual amount of current operating expenses.

The Cobb-Douglas production function was fitted to the model. Means, OLS estimates, standard deviations, and t-values are presented in Table 19.

Multicollinearity in the model:

Inspection of the correlation matrix given in Table 20 reveals that some multicollinearity is present in the model. For instance, the simple correlation coefficient for \( x_1^* \) and

\[ ^1x_1^* \text{ indicates the natural logarithm of } x_1. \]
Table 19. Estimated values for regression of county values of all farm products sold as a Cobb-Douglas function of inputs, OLS estimates for 25 Central Pacific counties, 1954

<table>
<thead>
<tr>
<th></th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>$x_4$</th>
<th>$x_5$</th>
<th>$x_6$</th>
<th>$Y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Means</td>
<td>4.03550944</td>
<td>3.589889</td>
<td>7.0501228</td>
<td>5.249523</td>
<td>3.168056</td>
<td>7.0311667</td>
<td>7.7910822</td>
</tr>
<tr>
<td>Betas OLS</td>
<td>0.26250</td>
<td>-0.0754</td>
<td>0.02470</td>
<td>0.3892</td>
<td>0.0222</td>
<td>0.41120</td>
<td></td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.2093</td>
<td>0.3355</td>
<td>0.1587</td>
<td>0.1354</td>
<td>0.0129</td>
<td>0.1193</td>
<td></td>
</tr>
<tr>
<td>t-values</td>
<td>1.25413</td>
<td>-2.24758</td>
<td>0.155847</td>
<td>2.87553</td>
<td>1.72522</td>
<td>3.44745</td>
<td></td>
</tr>
</tbody>
</table>

$x_1$ = number of family and hired workers

$x_2$ = number of tractors

$x_3$ = value of livestock investment

$x_4$ = acres of irrigated cropland

$x_5$ = acres of non-irrigated cropland

$x_6$ = current operating expenses

$Y$ = farm products sold

$R^2 = 0.925$
$x_2^*$ is $r_{12} = .9299$. Similarly, $r_{16} = .8188$ and $r_{13} = .7624$, which may indicate some degree of collinearity between $x_1^*$ and the remaining variables in the model. Also, from Table 20, $x_2^*$ is highly correlated with at least some of the variables in the model since

$$
\begin{align*}
    r_{23} & = 0.8048 \\
    r_{24} & = 0.7685 \text{ and} \\
    r_{25} & = 0.8043.
\end{align*}
$$

The main diagonal elements of the inverted correlation matrix (which are good measures of the location and degree of multicollinearity) were as follows:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Variance Inflation Factor (VIF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1^*$</td>
<td>11.2</td>
</tr>
<tr>
<td>$x_2^*$</td>
<td>18.9</td>
</tr>
<tr>
<td>$x_3^*$</td>
<td>5.2</td>
</tr>
<tr>
<td>$x_4^*$</td>
<td>5.2</td>
</tr>
<tr>
<td>$x_5^*$</td>
<td>2.0</td>
</tr>
<tr>
<td>$x_6^*$</td>
<td>5.5</td>
</tr>
</tbody>
</table>

The preceding VIF show exactly how much the variances of the regression coefficients for each variable have been increased because of multicollinearity. For example, the variance of the coefficient for $x_1^*$ is 11.2 times larger
Table 20. Simple correlation coefficient and eigenvalues for regression of county values of all farm products sold as a Cobb-Douglas function of inputs

<table>
<thead>
<tr>
<th></th>
<th>x_1</th>
<th>x_2</th>
<th>x_3</th>
<th>x_4</th>
<th>x_5</th>
<th>x_6</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>x_1</td>
<td>1.00000000</td>
<td>0.929918569</td>
<td>0.762442321</td>
<td>0.59685312</td>
<td>-.293808879</td>
<td>0.811853556</td>
<td>0.857723175</td>
</tr>
<tr>
<td>x_2</td>
<td>1.00000000</td>
<td>0.804854828</td>
<td>0.768523668</td>
<td>-.343767007</td>
<td>0.804250815</td>
<td>0.895857998</td>
<td></td>
</tr>
<tr>
<td>x_3</td>
<td>1.00000000</td>
<td>0.652256285</td>
<td>-.394812482</td>
<td>0.849321658</td>
<td>0.834051825</td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_4</td>
<td>1.00000000</td>
<td>-.598540061</td>
<td>0.507591052</td>
<td>0.714936</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_5</td>
<td>1.00000000</td>
<td>-.170285996</td>
<td>-.17028596</td>
<td>-.251882809</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>x_6</td>
<td>1.00000000</td>
<td>0.896652600</td>
<td>1.00000000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td>1.00000000</td>
<td>0.03248987</td>
<td>.10684128</td>
<td>.26112004</td>
<td>.35373836</td>
<td>1.03665134</td>
<td>4.2091592</td>
</tr>
</tbody>
</table>

Eigenvalues: 0.03248987 0.10684128 0.26112004 0.35373836 1.03665134 4.2091592

\[
\sum_{i=1}^{6} \frac{1}{\lambda_i} = 47.99360
\]

\[
\sum_{i=1}^{6} \lambda_i = 5.999987
\]
than it would have been if $x_2$ were uncorrelated with $x_2^*, x_3^*, x_4^*, x_5^*$, and $x_6^*$.

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

$$R_2^2: 1, 3, 4, 5, 6 = 1 - (1/18.9) = .9470,$$

if $x_2^*$ were regressed on the other explanatory variables. Similarly,

$$R_1^2: 2, 3, 4, 5, 6 - 1-(1/11.2) = .9107.$$

Multicollinearity, causing a real difficulty in separating the effect of $x_1^*$ (family and hired farm labor) from that of $x_2^*$ (number of tractors on the farm) on the value of farm products sold, leads to some serious problems in the model when ordinary least square is fitted. Some of the problems are:

1. $\hat{\beta}_2$, the marginal productivity of tractors, took an unexplainable negative sign (-.075). Economists usually attach considerable importance to the estimated coefficients, and do not consider them as just numbers. The interpretation of the negative sign for the coefficient $x_2$ (number of tractors) has some
implications that violate some economic assumptions, namely, that if \( \hat{\beta}_2 \) is negative, then the VMP_{x_2} = \( \frac{\hat{\gamma}}{x_2} (\hat{\beta}_2) \) < 0 for all nonnegative value of \( x_2 \) and \( Y \). A negative VMP would violate the important assumption that producers are rational and wish to maximize profit. Thus, they would not knowingly commit resources to the production process when the value of marginal product was negative. For that reason, it was impossible for Ruttan to retain \( x_2 \) in the final model because of the wrong sign obtained for \( \hat{\beta}_2 \), although the importance of \( x_2 \) is justifiable by economic theory.

2. Four out of six explanatory variables in the model, namely \( x_1^* \), \( x_2^* \), \( x_3^* \), and \( x_5^* \) were not statistically significant at the 5 percent probability level, which is a theoretically unexpected result.

3. Only \( x_4^* \) and \( x_6^* \) were statistically significant at the 5 percent probability level.

To avoid such serious problems as the preceding, economic researchers have employed the practice of deleting variables that were highly affected by multicollinearity, and only a subset of the explanatory variables have been used in the formulation of the final model. However, as explained
earlier in Chapter I, deletion of relevant variables can cause a serious omitted-variable specification bias.

To avoid omitted-variable specification bias, the use of biased linear estimation of the full model has become more popular during the last 10 years, following the pioneering articles by Hoerl and Kennard in 1970.

In this section the technique of ridge type estimators will be given and some of the results of the Monte Carlo simulation will be applied to compare the use of ridge estimators versus the normal practice of using OLS.

RIDGE REGRESSION ANALYSIS

Three ridge regression estimators:

The following three ridge estimators used in the Monte Carlo experiments were used to estimate the Model 1 presented earlier, equation (4.1).

1. $k_{LW}$ Lawless-Wang ridge type estimator; the $k$ value for this sample was calculated as

$$k_{LW} = \frac{\hat{\sigma}^2}{\sum_{i=1}^{6} \lambda_i \hat{\alpha}_i^2} = \frac{\hat{\sigma}^2}{\hat{\beta}' (X'X) \hat{\beta}} = 0.0267$$

The $k$ value for this estimator may be obtained using $\hat{\alpha}$ values from the orthogonalized model
\[ y = Z\alpha + u, \quad \hat{\alpha} = \Lambda^{-1}z'Y \]

as discussed in Chapter II.

Or, equivalently, using the standardized \( \beta \) estimates presented in Table 1, can be used, as shown above.

2. \( k_{\text{HKB}} \) estimator: following Hoerl, Baldwin and Kennard, the \( k \) value for this estimator is calculated by

\[ k_{\text{HKB}} = p\sigma^2 / \sum_{i=1}^{6} \alpha_i^2. \]

Or, equivalently,

\[ k = p\sigma^2 / \hat{\beta}'\hat{\beta} = .045. \]

Similarly, \( \hat{\alpha} \) from the orthogonalized model or the standardized coefficients can be used.

3. KGRR, the generalized ridge estimator, presented by Lawless (1978) was used, the \( k_i \) values for the estimator were calculated by

\[ k_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} (i = 1,2,\ldots,6). \]

An IMSL Routin Program-EIGRS available at OSU Computer Center was used to calculate the eigenvalues and the eigenvectors of the model, then a program written for a TI-59 programmable calculator was used to do the analysis. The
results of ridge regression analysis are presented in Table 21.

Regression results from OLS:

OLS regression results of the model as shown in Table 3, as noted earlier, were the following:

1. Except for variables $x_4$ and $x_6$, no variables were statistically significant at the 5 percent probability level.

2. $\hat{\beta}_2$, the coefficient of the number of tractors in the farm took an illogical negative value of (-.0754).

3. The overall goodness of fit of the model, measured by $R^2$, is supposedly good since $R^2 = .925$.

Linear restriction on OLS coefficients:

To alleviate the problem caused by multicollinearity when the OLS procedure was used to obtain estimates of the coefficients for model (4.1), some economists have suggested the use of linear restrictions on the coefficients (Johnston, 1972). A common restriction for the Cobb-Douglass production function is that all the coefficients should sum to unity.

The constrained OLS problem becomes

Minimize $(Y - x\beta)^\prime (Y - x\beta)$ subject to the constraint

$\beta^*$
that \( \sum_{i=1}^{6} \beta_i^* = 1 \).

The results in Table 21 indicate clearly that this technique suffers from the same problem as the unconstrained OLS due to the following facts:

(1) \( \beta_2^* \), the coefficient of the number of tractors variable took an even bigger negative unexpected sign, which has rather serious implications, as mentioned earlier.

(2) \( \beta_3^* \), the coefficient for the value of livestock investment becomes very small (24.7 times smaller than its value obtained by unrestricted OLS).

(3) No significant improvement in the precision of OLS estimates was realized by this method of linear restriction.

**Ridge regression results:**

In contrast to OLS, all three ridge estimators appear to be very effective in coping with the problem of multicollinearity since, as shown in Table 21, they all have estimated \( L_1(\hat{\beta}^*) \) less than the variance of OLS. \( \hat{\beta}_2^* \), the coefficient of the number of tractors in the farm, for all three ridge regressions took the expected positive sign:
Table 21. Estimates of the coefficients, square root of the squared error loss, $\hat{L}_1$, and the $B^2/L^2$ ratio for the OLS, the linearly restricted OLS, and the ridge regression estimators: $k_{LW}$, $k_{HKB}$ and KGRR for regression of county values of all farm products as a Cobb-Douglas function of inputs.

<table>
<thead>
<tr>
<th></th>
<th>K=0 OLS</th>
<th>OLS with linear restrictions</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
<th>$\hat{L}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>0.2625</td>
<td>0.2684</td>
<td>0.1973</td>
<td>0.1836</td>
<td>0.2035</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>0.2093</td>
<td>0.2091</td>
<td>0.1943</td>
<td>0.1495</td>
<td>0.1719</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.0754</td>
<td>-0.1039</td>
<td>0.0702</td>
<td>0.1133</td>
<td>0.0793</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>0.3355</td>
<td>0.3250</td>
<td>0.1698</td>
<td>0.1681</td>
<td>0.1692</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.1400</td>
<td>0.3220</td>
<td>0.1663</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.0247</td>
<td>0.0010</td>
<td>0.0628</td>
<td>0.0815</td>
<td>0.0669</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>0.1587</td>
<td>0.1418</td>
<td>0.1147</td>
<td>0.1091</td>
<td>0.0964</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0100</td>
<td>0.0340</td>
<td>0.1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.3892</td>
<td>0.3864</td>
<td>0.3216</td>
<td>0.2976</td>
<td>0.2767</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>0.1354</td>
<td>0.1354</td>
<td>0.1147</td>
<td>0.1092</td>
<td>0.0622</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0123</td>
<td>0.0030</td>
<td>0.1219</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_5$</td>
<td>0.0222</td>
<td>0.0200</td>
<td>0.0194</td>
<td>0.0184</td>
<td>0.0132</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>0.0129</td>
<td>0.0112</td>
<td>0.0062</td>
<td>0.0061</td>
<td>0.0060</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0060</td>
<td>0.0180</td>
<td>0.0770</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_6$</td>
<td>0.4112</td>
<td>0.4281</td>
<td>0.4336</td>
<td>0.3447</td>
<td>0.3686</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{K}_1$</td>
<td>0.1193</td>
<td>0.1080</td>
<td>0.0313</td>
<td>0.0313</td>
<td>0.0313</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0080</td>
<td>0.0230</td>
<td>0.0050</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$ $\hat{L}_1$ was computed by $\hat{L}_1 = [m_i \sigma^2/\lambda_i + (1-m_i)^2 (\hat{\alpha}_1^2 - V(\hat{\alpha}_1)) \geq 0]^{1/2}$.

$^b$ was set to zero since $(\hat{\alpha}_1^2 - V(\hat{\alpha}_1))$ is negative and is set to zero.
\[ \hat{\beta}_2^{*}(\text{KLW}) = 0.0702 \]
\[ \hat{\beta}_2^{*}(k_{HKB}) = 0.1133 \]
\[ \hat{\beta}_2^{*}(\text{KGRR}) = 0.0793 \]

The results in Table 21 also showed that the three ridge estimators' performance was very close in terms of estimated squared error.

Another interesting result in Table 21 is that \( B^2/L^2 \) (the ratio of estimated bias squared to the estimated total loss) for \( k_{\text{LW}} \) and \( k_{\text{HKB}} \) were (except for \( \hat{\beta}_2 \)) within the range at which the simulation results showed \( \hat{\alpha}_1^* \) to follow the normal distribution, indicating that for all these \( k_{\text{LW}} \) coefficients, except for \( \beta_2 \), it might be feasible to test the hypothesis about the significance of the coefficients using

\[ t^* = \frac{\hat{\alpha}_1^*(L_1(\hat{\alpha}_1^*))^{.5}}{L_1(\hat{\alpha}_1^*)} \]

The evaluation of the Lawless-Wang ridge estimates in Table 21 can be done by testing the hypothesis that the ridge estimates are lower in weak MSE following Wallace (1971). Computing

\[ u = \frac{(\text{SSE}(\hat{\beta}^*) - \text{SSE}(\hat{\beta}))}{6\hat{\sigma}^2} \]

Where \( \text{SSE}(\hat{\beta}^*) \) is the error sum of squares for the Lawless-Wang ridge estimates, \( \text{SSE}(\hat{\beta}) \) is the error sum of squares for
the OLS estimates, $\sigma^2$ is the OLS estimate of the variance, and $u = 0.2597$ with six and eighteen degrees of freedom. This value of $u$, however, is not significant even for central $F$. Thus the $u$ test indicated that the reduction in $R^2$ from the use of the Lawless-Wang ridge regression is not significant, and the problem is not unsuited for ridge regression analysis.

In general, a significant improvement over OLS appeared to be realized by the use of these three ridge estimators. When orthogonalization of the model is not possible due to lack of computational facilities, the analysis showed that one can obtain the same estimate of $L^2 \_1$ using the non-orthogonalized data:

Let $A = (X'X + kI)^{-1}$ where $(X'X)$ is the correlation matrix,

$$L_{1i} = [V(\hat{\beta}^*)]^{-1} + (-ka_i\hat{\beta}_i)^2$$

where $a_i$ is the $i$th row of $A$ and

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 - V(\hat{\beta}_1) \geq 0 \\ \vdots \\ \hat{\beta}_{p} - V(\hat{\beta}_{p}) \geq 0 \end{bmatrix}$$

This method enables one to compute the $(\text{bias})^2$ components of $L_1(\hat{\beta}^*)$ using only standardized $\hat{\beta}$ values and their variances.
V. SUMMARY AND CONCLUSIONS

When multicollinearity is severe, the ordinary least square (OLS) estimate, although the best linear unbiased estimate (BLUE), may be unreliable and imprecise because of the greatly inflated variances. Under these circumstances, the individual coefficients may be statistically insignificant and even take the wrong signs, even though the overall regression measured by $R^2$ may be highly significant. Since multicollinearity itself is a data problem related to the underlying experimental design, the first logical solution to the problem is to attempt to acquire more observations, or to disaggregate the data, or to focus on obtaining better model specification. In some cases redefinition of the variables or the use of prior information about the values of the coefficients, can improve OLS estimation and partially remove the effect of multicollinearity. When these solutions are not possible, variable deletion has been a common practice to solve the problem of multicollinearity, which may not be too bad a practice when the omitted variable specification bias is not too large. But if the specification bias is large, variable deletion could be a very poor practice for reducing the degree of multicollinearity. One alternative to variable
deletion is the use of biased linear estimators. If the added bias is more than offset by the reduction in variance, biased linear estimation could give better estimates of the coefficients.

Ridge regression is a method of biased linear estimation and was found to be more effective than OLS in the Monte Carlo experiments of this thesis under a wide range of regression situations. Four applied ridge estimators were compared in the Monte Carlo simulations, namely, $k_{\text{LW}}$, RIDGN, $k_{\text{HKB}}$, and KGRR. It was shown in Chapter III of this thesis that the ridge regression estimators outperformed OLS significantly for the situations where there was a relatively high degree of multicollinearity and where the smallest true coefficients corresponded to the smallest eigenvalues of $(X'X)$.

A comparison via the Monte Carlo experiments in Chapter III indicated that the Lawless-Wang ordinary ridge estimator tended to be more accurate and more stable in performance than the other ridge estimators, as measured by the observed squared error loss over a wide range of situations. The performance of the other ridge estimators compared were more volatile, depending on the degree of multicollinearity and the individual models.

Perhaps the most cited serious limitation of the use of biased linear estimators for empirical problems has been the inability to assess the reliability of these estimates. Prior information about the coefficients to be estimated
(relative sizes, variances, etc.) offers some basis for the
evaluation of the expected quality of the ridge estimates.
However, such information may not be available and diff-
cult to obtain. A technique for estimating observed
squared error loss was proposed in Chapter I:
\[ L_1(\hat{\alpha}^*) = \sum_{i=1}^{P} \left( m_i^2 \sigma^2 / \lambda_i + (1-m_i)^2 \theta \right) \]
where \( m_i = \frac{\lambda_i}{\lambda_i + k} \), \( \theta \) is the
proposed method of estimating \( \alpha_i^2 \) defined by
\[ \theta = (\hat{\alpha}_i^2 - V(\hat{\alpha}_i)) > 0, \]
and \( \lambda_i \) is the \( i \)th eigenvalue of \((X'X)\).

Results of the Monte Carlo experiments in Chapter III
indicated that the proposed method of estimating the
squared error loss appeared to be fairly accurate for the
Lawless-Wang ridge estimator for models with relatively
high multicollinearity, despite some tendency toward over-
estimation. However, the proposed method usually gave
excellent estimates of the observed squared error loss for
models with relatively lower multicollinearity.

On the other hand, the results of Chapter III indi-
cated that the proposed method of estimating squared error
loss did not appear to be a suitable measure for estimating
the squared error loss for the other ridge estimates (RIDGN,
k_{HKB}, and KGRR) due to the fact that it consistently under-
estimated the loss. This underestimation could be danger-
ously misleading to the researcher who usually depends on
the analysis of the sample data for measures of reliabili-
ty, such as variance, t tests, etc. The results in Chapter
III of this thesis also revealed that when the regression conditions are 'favorable', namely that the ratio of the estimated bias squared to the total squared error loss is small (less than 10 percent), the ridge coefficient obtained by the Lawless-Wang estimator appeared to be distributed approximately normal. If so, the $t^*$ statistic defined by

$$t^* = \frac{m_i \hat{\alpha}_i}{[m_i \sigma^2 / \lambda_i + (1-m_i)^2 \theta_i]^{1/2}}$$

could be used to test hypotheses about the significance of the ridge coefficient $\hat{\alpha}_i^*$, as discussed in Chapter III. It should be emphasized, however, that the $t^*$ statistic should not be used for the Lawless-Wang estimator if the ratio of estimated bias squared to the total squared error loss is more than 10 percent. (If the ratio was more than 10 percent, the $t^*$ statistic departed from the normal distribution, due to the lack of independence between $\hat{\alpha}_i^*$ and the estimated loss, $L_1(\hat{\alpha}_i^*)$.)

The results of Chapter III also indicated that the use of the $t^*$ statistic is possible in empirical research by the use of the estimated

$$\frac{B_i^2}{L_i^2} = \frac{(1-m_i)^2 \theta_i}{[m_i \sigma^2 / \lambda_i + (1-m_i)^2 \theta_i]}$$

where all the variables are as defined earlier. For models with relatively low multicollinearity, the proposed method gave an excellent performance for all the models with a probability of making the correct decision of greater than 0.80. However, this probability is only
greater than 0.50 for models with relatively high multi-
collinearity and $\sigma^2$ value greater than one.
BIBLIOGRAPHY


Appendix A
Table A1. The cumulative frequency distribution of $k_{LM}$ estimate, $\hat{\alpha}_1^+$, for the 8 models $\{a_1-a_8\}$, with $\lambda = (.01, 1, 1.99)$, $\sigma^2 = 4$ and the theoretical t distribution ($N = \infty$).

<table>
<thead>
<tr>
<th>$B^2/L^2$</th>
<th>$a_1^+$</th>
<th>$a_2^+$</th>
<th>$a_3^+$</th>
<th>$a_4^+$</th>
<th>$a_5^+$</th>
<th>$a_6^+$</th>
<th>$a_7^+$</th>
<th>$a_8^+$</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&lt;0$</td>
<td>.800</td>
<td>.984</td>
<td>.600</td>
<td>.364</td>
<td>.948</td>
<td>.852</td>
<td>.189</td>
<td>.995</td>
<td>0.000</td>
</tr>
<tr>
<td>$0$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>.0013</td>
</tr>
<tr>
<td>$&lt;.5$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>.0228</td>
</tr>
<tr>
<td>$&lt;1$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>.1587</td>
</tr>
<tr>
<td>$&lt;.5$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.003</td>
<td>0.057</td>
<td>0.000</td>
<td>0.001</td>
<td>0.000</td>
<td>.3085</td>
</tr>
<tr>
<td>$&lt;1$</td>
<td>0.166</td>
<td>0.049</td>
<td>0.230</td>
<td>0.114</td>
<td>0.089</td>
<td>0.149</td>
<td>0.363</td>
<td>0.020</td>
<td>.6915</td>
</tr>
<tr>
<td>$&lt;.5$</td>
<td>0.229</td>
<td>0.184</td>
<td>0.280</td>
<td>0.391</td>
<td>0.208</td>
<td>0.243</td>
<td>0.506</td>
<td>0.172</td>
<td>.8413</td>
</tr>
<tr>
<td>$&lt;1$</td>
<td>0.314</td>
<td>0.331</td>
<td>0.477</td>
<td>0.620</td>
<td>0.315</td>
<td>0.328</td>
<td>0.794</td>
<td>0.431</td>
<td>.9712</td>
</tr>
<tr>
<td>$&lt;.5$</td>
<td>0.399</td>
<td>0.375</td>
<td>0.674</td>
<td>0.827</td>
<td>0.349</td>
<td>0.395</td>
<td>0.960</td>
<td>0.490</td>
<td>.9987</td>
</tr>
<tr>
<td>$&lt;1$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a. Bias squared to total squared loss, $B^2/L^2 = \frac{(1-m_1)^2}{m_1^2}$ where: $m_1 = \lambda_1^{1/\lambda_1} \cdot \frac{\lambda_1^{2/\lambda_1}}{(1-m_1)^2}$

b. The cumulative distribution of $\hat{\alpha}_1^+$ for some specified ranges where, $\hat{\phi} = [m_1^2\lambda_1^{2/\lambda_1}((1-m_1)^2)]^{1/2}$ and $\hat{\theta}_1 = (\alpha_1^2-V(\hat{\alpha}_1)) \geq 0$. 


Table A2. The cumulative frequency distribution of $k_{LM}$ estimate, $\hat{\alpha}_1$, for the 8 models ($a_1$-$a_8$), with $\lambda = (0.01, 1, 1.99)$, $\sigma^2 = 2$, and the theoretical t distribution ($N = \infty$).

<table>
<thead>
<tr>
<th></th>
<th>$a_1^m$</th>
<th>$a_2^m$</th>
<th>$a_3^m$</th>
<th>$a_4^m$</th>
<th>$a_5^m$</th>
<th>$a_6^m$</th>
<th>$a_7^m$</th>
<th>$a_8^m$</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>$b^2/L^2a$</td>
<td>(10,10,10)</td>
<td>(15,37.5,15)</td>
<td>(37.5,15,15)</td>
<td>(37.5,37.5,15)</td>
<td>(12,12,12)</td>
<td>(12,12,12)</td>
<td>(16,22,22)</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>$a_1 &gt; 3^b$</td>
<td>.667</td>
<td>.969</td>
<td>.429</td>
<td>.223</td>
<td>.500</td>
<td>.742</td>
<td>.104</td>
<td>.990</td>
<td>.000</td>
</tr>
<tr>
<td>$-3^b &lt; a_1 &lt; -2^b$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.013</td>
</tr>
<tr>
<td>$-2^b &lt; a_1 &lt; -1^b$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.013</td>
<td>0.050</td>
<td>0.000</td>
<td>0.000</td>
<td>0.040</td>
<td>0.000</td>
<td>0.1587</td>
</tr>
<tr>
<td>$-1^b &lt; a_1 &lt; -0.5^b$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.026</td>
<td>0.206</td>
<td>0.000</td>
<td>0.000</td>
<td>0.382</td>
<td>0.000</td>
<td>0.0985</td>
</tr>
<tr>
<td>$-0.5^b &lt; a_1 &lt; 0.5^b$</td>
<td>0.230</td>
<td>0.083</td>
<td>0.309</td>
<td>0.458</td>
<td>0.145</td>
<td>0.217</td>
<td>0.507</td>
<td>0.052</td>
<td>0.6915</td>
</tr>
<tr>
<td>$-0.5^b &lt; a_1 &lt; 1^b$</td>
<td>0.311</td>
<td>0.279</td>
<td>0.440</td>
<td>0.687</td>
<td>0.278</td>
<td>0.342</td>
<td>0.836</td>
<td>0.251</td>
<td>0.8413</td>
</tr>
<tr>
<td>$0 &lt; a_1 &lt; 1^b$</td>
<td>0.473</td>
<td>0.429</td>
<td>0.743</td>
<td>0.822</td>
<td>0.394</td>
<td>0.445</td>
<td>0.953</td>
<td>0.456</td>
<td>0.9772</td>
</tr>
<tr>
<td>$2^b &lt; a_1 &lt; 3^b$</td>
<td>0.752</td>
<td>0.484</td>
<td>0.970</td>
<td>1.000</td>
<td>0.446</td>
<td>0.656</td>
<td>1.000</td>
<td>0.518</td>
<td>0.9987</td>
</tr>
<tr>
<td>$a_1 &gt; 3^b$</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

a. Bias squared to total squared loss, $B^2/L^2 = \frac{(1-m_1)^2}{\lambda_1^2}$ where; $m_1 = \frac{\lambda_1}{\lambda_1 + K}$.

b. The cumulative distribution of $\hat{\alpha}_1$ for some specified range, where,

$\hat{\phi} = [m_1^2 \frac{B^2}{\lambda_1^2} (1-m_1)^2 \hat{\theta} \hat{\theta}]^{0.5}$ and $\hat{\theta}_1 = (\hat{\alpha}_1^2 + \hat{\alpha}_1) > 0$. 

---

Page 12
Table A3. The cumulative frequency distribution of \( \hat{a}_{11}^2 \), for the 8 models \( (a_1-a_8) \) with \( \lambda = (.01, 1, 1.99) \), \( \sigma^2 = 1 \) and the theoretical t distribution \( (N = \infty) \).

<table>
<thead>
<tr>
<th>( a_1^2 )</th>
<th>( a_2^2 )</th>
<th>( a_3^2 )</th>
<th>( a_4^2 )</th>
<th>( a_5^2 )</th>
<th>( a_6^2 )</th>
<th>( a_7^2 )</th>
<th>( a_8^2 )</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10,10,10)</td>
<td>(15,37.5,37.5)</td>
<td>(37.5,15,37.5)</td>
<td>(37.5,37.5,15)</td>
<td>(12.12,12)</td>
<td>(12/12.12)</td>
<td>(12/12,12)</td>
<td>(16/22,22) distribution</td>
<td></td>
</tr>
<tr>
<td>( B^2/L^2a )</td>
<td>.500</td>
<td>.939</td>
<td>.273</td>
<td>.125</td>
<td>.819</td>
<td>.590</td>
<td>.055</td>
<td>.980</td>
</tr>
<tr>
<td>( a_1 &gt; 3\hat{\phi}b )</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.016</td>
<td>0.000</td>
<td>0.027</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>-3\hat{\phi} &lt; a_1 &lt; -2\hat{\phi}</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>-2\hat{\phi} &lt; a_1 &lt; -1\hat{\phi}</td>
<td>0.05</td>
<td>0.000</td>
<td>0.110</td>
<td>0.156</td>
<td>0.000</td>
<td>0.192</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>-1\hat{\phi} &lt; a_1 &lt; -0.5\hat{\phi}</td>
<td>.101</td>
<td>0.000</td>
<td>0.265</td>
<td>0.298</td>
<td>0.000</td>
<td>0.269</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>-0.5\hat{\phi} &lt; a_1 &lt; 0.5\hat{\phi}</td>
<td>.338</td>
<td>0.167</td>
<td>0.560</td>
<td>0.544</td>
<td>0.242</td>
<td>0.345</td>
<td>0.660</td>
<td>.120</td>
</tr>
<tr>
<td>0.5\hat{\phi} &lt; a_1 &lt; 1\hat{\phi}</td>
<td>.738</td>
<td>0.415</td>
<td>0.850</td>
<td>0.792</td>
<td>0.418</td>
<td>0.693</td>
<td>0.864</td>
<td>.412</td>
</tr>
<tr>
<td>1\hat{\phi} &lt; a_1 &lt; 2\hat{\phi}</td>
<td>.992</td>
<td>0.583</td>
<td>0.954</td>
<td>0.967</td>
<td>0.568</td>
<td>0.868</td>
<td>0.990</td>
<td>.639</td>
</tr>
<tr>
<td>2\hat{\phi} &lt; a_1 &lt; 3\hat{\phi}</td>
<td>.996</td>
<td>0.652</td>
<td>1.000</td>
<td>1.000</td>
<td>0.796</td>
<td>0.960</td>
<td>1.000</td>
<td>.704</td>
</tr>
<tr>
<td>( a_1 &gt; 3\hat{\phi} )</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

a. Bias squared to total squared loss, \( B^2/L^2 = \frac{(1-m_1)^2}{m_1^2+1} \), where \( m_1 = \frac{1}{\lambda_{11} k} \).

b. The cumulative distribution of \( \hat{a}_{11}^2 \) for some specified ranges, where,

\[ \begin{align*}
\hat{\phi} &= \left[ m_1^2/(1-m_1) \right]^{1/2} \hat{\theta}^5 \\
\hat{\theta} &= \left[ (\hat{a}_{11}^2 - \nu(\hat{a}_{11})) \right]^{1/2} \end{align*} \]
Table A4. The cumulative frequency distribution of $\hat{a}_1$ for the 8 models ($a_1$-$a_8$) with $\lambda = (0.01, 1, 1.99)$, $\sigma^2 = 0.5$ and the theoretical t distribution ($N = m$).

<table>
<thead>
<tr>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$a_5$</th>
<th>$a_6$</th>
<th>$a_7$</th>
<th>$a_8$</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>(10,10,10)</td>
<td>(15,17.5</td>
<td>(17.5,19.5</td>
<td>(17.5,37.5,17.5)</td>
<td>(37.5,17.5,37.5)</td>
<td>(12,12,12)</td>
<td>(12,12,12)</td>
<td>(12,12,12)</td>
<td>(16,17,22)</td>
</tr>
<tr>
<td>$B^2/L^2$</td>
<td>0.333</td>
<td>0.886</td>
<td>0.158</td>
<td>0.067</td>
<td>0.693</td>
<td>0.419</td>
<td>0.028</td>
<td>0.961</td>
</tr>
<tr>
<td>$a_1 &gt; 3\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$-3\hat{a} &lt; a_1 &lt; -2\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.019</td>
<td>0.000</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$-2\hat{a} &lt; a_1 &lt; -1\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.128</td>
<td>0.000</td>
<td>0.014</td>
<td>0.140</td>
<td>0.000</td>
</tr>
<tr>
<td>$-1\hat{a} &lt; a_1 &lt; -0.5\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.265</td>
<td>0.000</td>
<td>0.036</td>
<td>0.324</td>
<td>0.000</td>
</tr>
<tr>
<td>$-0.5\hat{a} &lt; a_1 &lt; 0.5\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.609</td>
<td>0.000</td>
<td>0.387</td>
<td>0.495</td>
<td>0.000</td>
</tr>
<tr>
<td>$0.5\hat{a} &lt; a_1 &lt; 1\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.849</td>
<td>0.000</td>
<td>0.818</td>
<td>0.741</td>
<td>0.000</td>
</tr>
<tr>
<td>$1\hat{a} &lt; a_1 &lt; 2\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.954</td>
<td>0.000</td>
<td>0.994</td>
<td>0.974</td>
<td>0.000</td>
</tr>
<tr>
<td>$2\hat{a} &lt; a_1 &lt; 3\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.985</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$a_1 &gt; 3\hat{a}$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>1.000</td>
<td>0.000</td>
<td>1.000</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

a. Bias squared to total squared loss, $B^2/L^2 = \frac{(1-m_1)^2}{\lambda_1^2 \times (1-n_1)^2}$ where $m_1 = \frac{1}{\lambda_1 + k}$.

b. The cumulative frequency distribution of $\hat{a}_1$, for some specified ranges, where

\[ \hat{\delta} = [m_1^2 + \lambda_1^2 (1-n_1)^2]^{-1} \times \hat{a}_1 \text{ and } \hat{a}_1 = (\hat{a}_1^2) \times (\hat{a}_1) \geq 0. \]
Table A5. The cumulative frequency distribution of $\hat{a}_1^*$ for the 8 models ($a_1^* - a_8^*$) with $\lambda = (0.01, 1, 1.99)$, $\sigma^2 = 0.25$ and the theoretical $t$ distribution ($N = n$).

<table>
<thead>
<tr>
<th>$a_1^*$</th>
<th>$a_2^*$</th>
<th>$a_3^*$</th>
<th>$a_4^*$</th>
<th>$a_5^*$</th>
<th>$a_6^*$</th>
<th>$a_7^*$</th>
<th>$a_8^*$</th>
<th>$t$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(B^2/L^2) &amp; .200 &amp; .795 &amp; .086 &amp; .035 &amp; .531 &amp; .265 &amp; .014 &amp; .925 &amp; 0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_1^* &gt; 3\hat{\delta}$ &amp; .000 &amp; .000 &amp; .000 &amp; .001 &amp; .000 &amp; .000 &amp; .000 &amp; .000 &amp; .000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-3\hat{\delta} &lt; a_1^* &lt; -2\hat{\delta}$ &amp; .000 &amp; .000 &amp; .011 &amp; .024 &amp; .000 &amp; .000 &amp; .010 &amp; .000 &amp; .0228</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-2\hat{\delta} &lt; a_1^* &lt; -1\hat{\delta}$ &amp; .034 &amp; .000 &amp; .168 &amp; .164 &amp; .000 &amp; .032 &amp; .123 &amp; .000 &amp; .1587</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-1\hat{\delta} &lt; a_1^* &lt; -0.5\hat{\delta}$ &amp; .354 &amp; .000 &amp; .316 &amp; .318 &amp; .000 &amp; .241 &amp; .266 &amp; .000 &amp; .3085</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$-0.5\hat{\delta} &lt; a_1^* &lt; 0.5\hat{\delta}$ &amp; .689 &amp; .514 &amp; .693 &amp; .683 &amp; .590 &amp; .679 &amp; .689 &amp; .503 &amp; .6915</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$0.5\hat{\delta} &lt; a_1^* &lt; 1\hat{\delta}$ &amp; .925 &amp; .831 &amp; .861 &amp; .860 &amp; .854 &amp; .906 &amp; .848 &amp; .879 &amp; .8413</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$1\hat{\delta} &lt; a_1^* &lt; 2\hat{\delta}$ &amp; .998 &amp; .977 &amp; .956 &amp; .961 &amp; .986 &amp; 1.000 &amp; .957 &amp; .973 &amp; .9772</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$2\hat{\delta} &lt; a_1^* &lt; 3\hat{\delta}$ &amp; 1.000 &amp; 1.000 &amp; 1.000 &amp; .999 &amp; .999 &amp; 1.000 &amp; 1.000 &amp; .993 &amp; .9987</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_1^* &gt; 3\hat{\delta}$ &amp; 1.000 &amp; 1.000 &amp; 1.000 &amp; 1.000 &amp; 1.000 &amp; 1.000 &amp; 1.000 &amp; 1.000 &amp; 1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. $B^2/L^2$ is computed as $B^2/L^2 = \frac{(1-m_1)^2 a_1^2}{(m_1^2 a_1^2 / \lambda_1 + (1-m_1)^2 \sigma^2 \hat{a_1}^5)}$ where $m_1 = \frac{\lambda_1}{\lambda_1 + k}$.

b. The cumulative frequency distribution of $\hat{a}_1^*$ for some specified ranges where,

\[
\hat{\delta}_1 = \left|m_1^2 a_1^2 / \lambda_1 + (1-H1)\right|^{0.5} \text{ and } \hat{\delta}_1 = (a_1^2 \cdot V(a_1)) > 0.
\]