R- 105

Center for Research on Economic and Social Theory
Research Seminar in Quantitative Economics

Discussion Paper

[redacted]

DEPARTMENT OF ECONOMICS

University of Michigan

Ann Arbor, Michigan 48109

[redacted]

SOME NEW RESULTS ON RIDGE REGRESSION ESTIMATION

Karl Lin
Jan Kmenta*

The University of Michigan

R-105

SOME NEW RESULTS ON RIDGE REGRESSION ESTIMATION

Abstract

In this paper we consider various interpretations of the ordinary ridge regression estimator with a given shrinkage factor k, and report the results of an extensive Monte Carlo of several ridge regression estimators involving sample-based rules for selecting k. A major distinguishing feature of the study is the use of a general loss structure, the p-norm, in the evaluation process. Other factors taken into consideration include different degree of ill-conditioning of data, different number of explanatory variables, and different shape and non-centrality of the regression coefficients. The main results are:

(i) With minor exceptions, all the ridge regression estimators considered yield a smaller average loss regardless of the loss function used.

(ii) The reduction in the average loss of the ridge regression estimators increases when the degree ill-conditioning of data increases. The reduction reaches a substantial level when the degree of ill-conditioning is only moderate.

(iii) On the basis of our experiment it is possible to make a recommendation concerning the rule of selecting k.

SOME NEW RESULTS ON RIDGE REGRESSION ESTIMATION

1. Introduction

The introduction by Hoerl and Kennard (1970a,1970b) of a ridge regression estimator to deal with the problem of multicollinearity in regression has been followed by a large number of papers in the statistical literature. In the area of econometrics, though, the method of ridge regression has only recently been given some attention.[1)] One of the reasons for the lack of interest in ridge regression on the part of the econometricians may be the fact that Hoerl and Kennard have justified their method on pragmatic grounds without providing any interpretation. Other reasons for the reluctant reception of ridge regression by econometricians are likely to include the difficulty in selecting a suitable value of the shrinking factor, which is important in securing a dominance over least squares, and the restrictive nature of the mean-square-error criterion, on which the claim of this dominance rests.

In this paper we address all of the above mentioned issues. The plan is as follows. The ridge regression method and its properties are described in the remainder of this section. In section 2 we provide several interpretations of the ridge regression method and discuss the meaning of the shrinking factor. In section 3 we consider the case where the value of the shrinking factor is not given *a priori* and describe a number of rules for choosing its value on the basis of sample observations. Section 4 contains a description of an extensive Monte Carlo experiment designed to check the dominance of the ridge regression estimation over least squares under various loss structures in a situation when the value of the shrinking factor is not known a priori. The results of the experiment are evaluated in section 5 and concluding remarks are presented in section 6.

1.1 Ordinary Ridge Regression

Throughout this paper we consider the problem of estimating the coefficients of the standard linear regression model

$$y = X\beta+\varepsilon \tag{1}$$

where y is a $n\times 1$ vector of observed values of the dependent variable, X is a $n\times p$ matrix of the nonstochastic values of the explanatory variables, β is a $p\times 1$ vector of the coefficients to be estimated, and ε a $n\times 1$ vector of stochastic disturbances assumed to be distributed $N(0,\sigma^2 I_n)$.

Following Hoerl and Kennard (1970a) we define the ordinary ridge regression estimator (ORR) as follows:

$$\begin{aligned}\hat{\beta}(k) &= (X'X+kI)^{-1}X'y \\ &= (X'X+kI)^{-1}X'X\hat{\beta} \\ &= [I+k(X'X)^{-1}]^{-1}\hat{\beta}\end{aligned} \tag{2}$$

where k is a positive scalar and $\hat{\beta}$ is an ordinary least squares (OLS) estimator of β. Note that $\hat{\beta}(k)$ shrinks $\hat{\beta}$ in the sense that $\hat{\beta}(k)'\hat{\beta}(k) < \hat{\beta}'\hat{\beta}$. For a given k, $\hat{\beta}(k)$ is biased but consistent provided that $\text{plim}(X'X)/n$ exists.

The main attractive feature of the ORR estimator, established by Hoerl and Kennard (1970a,1970b), is that there exists a k>0 such that

$$E[\hat{\beta}(k)-\beta]'[\hat{\beta}(k)-\beta] < E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) \qquad (3)$$

i.e., that

$$\text{tr MSE}[\hat{\beta}(k)] < \text{tr MSE}(\hat{\beta}).$$

An important extension of the above result was provided by Theobald (1974) who proved that

(a) the statement in (3) above also holds if we use an arbitrary non-negative definite weighting matrix W, i.e.,

$$E[\hat{\beta}(k)-\beta]'W[\hat{\beta}(k)-\beta] < E(\hat{\beta}-\beta)'W(\hat{\beta}-\beta) \qquad (4)$$

(b) the condition in (4) above is equivalent to the condition that

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

is non-negative definite;

(c) a sufficient condition for (4) to hold (i.e., for the mean-square-error dominance of ORR over OLS) is that

$$k < 2\sigma^2/\beta^{*\prime}\beta^{*} \qquad (5)$$

where β^* is the coefficient vector in (1) with each of the explanatory variable normalized so that its sample sum of squares is unity.

2. Interpretation of ORR

2.1. ORR as a Mixed Estimator

We note that the ORR estimator of β can be obtained by an application of the least squares method to the following:

$$\begin{bmatrix} y \\ \underline{0} \end{bmatrix} = \begin{bmatrix} X \\ \sqrt{k} I_p \end{bmatrix} \beta + \begin{bmatrix} \varepsilon \\ v \end{bmatrix} \tag{6}$$

where $\underline{0}$ a $p \times 1$ vector of zeros. Let us compare this with the mixed estimator of β of the model in (1) estimated with the restriction that very likely

$$a \leq \beta_j \leq b \qquad (j = 1,2,\ldots,p) \tag{7}$$

where a and b are constants to be determined in such a way that the application of OLS to (6) yields $\hat{\beta}(k)$. Following Theil and Goldberger (1961) we write

$$\beta_j = \frac{a+b}{2} + u_j \tag{8}$$

where $u_j \sim N[0,(b-a)^2/16]$. The p-pieces of information about each of the p-regressors can then be represented as

$$\frac{a+b}{2} = \beta_1 \times 0 + \beta_2 \times 0 + \ldots + \beta_j \times 1 + \beta_{j+1} \times 0 + \ldots + \beta_p \times 0 + (-u_j) \ . \tag{9}$$

But since $Var(u_j) = (b-a)^2/16$ whereas $Var(\varepsilon_i) = \sigma^2$ $(i = 1,2,\ldots,n)$, we remove the resulting heteroskedasticity by re-writing (9) as

$$\left(\frac{a+b}{2}\right)\left(\frac{4\sigma}{b-a}\right) = \beta_1 \times 0 + \ldots + \beta_j \times \left(\frac{4\sigma}{b-a}\right) + \ldots + \beta_p \times 0 + v_j \tag{10}$$

where $v_j = \left(\frac{-4\sigma}{b-a}\right) u_j$.

Comparing (10) with (6) we have

$$\left(\frac{a+b}{2}\right)\left(\frac{4\sigma}{b-a}\right) = 0 \tag{11}$$

$$\left(\frac{4\sigma}{b-a}\right) = \sqrt{k} \tag{12}$$

which, for b>a, gives

$$a = -b$$

$$b = \frac{2\sigma}{\sqrt{k}} \quad .$$

Thus ORR can be viewed as a mixed estimator with the prior restriction that very likely

$$-\frac{2\sigma}{\sqrt{k}} \leq \beta_j \leq +\frac{2\sigma}{\sqrt{k}} \tag{13}$$

for $j = 1,2,\ldots,p$. Note that if the value of k is very small relative to σ, the restriction is not very binding and ORR is close to OLS. If, on the other hand, the value of k is large relative to σ, the interval in (13) becomes rather tight and the difference between ORR and OLS becomes larger.

2.2 ORR as a Result of Restricted Minimization

Consider the problem of obtaining an estimator of β by minimizing $(y-X\beta)'(y-X\beta)$ subject to the restriction that $\beta'\beta = r$ where r is positive and given. Setting up the Lagrange multiplier function

$$H = (y-X\beta)'(y-X\beta)-\lambda(\beta'\beta-r), \tag{14}$$

differentiating H with respect to β and equating the result to zero, we obtain

$$\tilde{\beta} = (X'X+\lambda I)^{-1}X'y \tag{15}$$

where the value of λ is to be chosen so that $\tilde{\beta}'\tilde{\beta} = r$, i.e., so that

$$y'X(X'X+\lambda I)^{-2}X'y = r. \qquad (16)$$

It is clear, of course, that $\tilde{\beta}$ is then an ORR estimator of β with λ=k. A small value of r results in a large value of k and <u>vice</u> <u>versa</u>.

2.3 ORR as a Bayesian Estimator

If $y/\beta \sim N(X\beta, \sigma^2 I_n)$ and the prior distribution of β is specified as $\beta \sim N(0, \omega^2 I_p)$, then β has the following posterior distribution:

$$\beta \sim N[(X'X + \frac{\sigma^2}{\omega^2} I)^{-1} X'y,\ \sigma^2 (X'X + \frac{\sigma^2}{\omega^2} I)^{-1}] . \qquad (17)$$

Thus the ORR estimator with $k=\sigma^2/\omega^2$ can be represented as the mean of the posterior distribution of β given that the mean of the prior distribution of β is zero. If ω^2 is relatively large, i.e., if the prior distribution of β is relatively flat, then ORR and OLS are relatively close to each other. A tight prior distribution of β, on the other hand, leads to a more substantial departure of ORR from OLS.

3. Rules for Selecting k

In most cases the value of k is not given *a priori* but has to be determined on the basis of available sample observations. A large number of suggestions for calculating k by various authors is presented in Dempster *et al*.(1977). In this study we consider only those rules for which a reasonable rationalization can be provided and which can be implemented without a high computational cost. Since most of the rules are developed by reference to a principal component form of (1), we precede the discussion of the rules for selecting k by a description of the prefered transformation.

The regression model in (1) can be re-written as follows:

$$\begin{aligned} y &= X\beta+\varepsilon \\ &= XPP'\beta+\varepsilon \\ &= X^*\alpha+\varepsilon \end{aligned} \tag{18}$$

where $X^* = XP$, $\alpha = P'\beta$, and P is an orthonormal matrix whose columns are eigenvectors of X'X, that is,

$$PP' = I \tag{19}$$

$$P'X'XP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & & \lambda_p \end{bmatrix} = \Lambda \tag{20}$$

and $\lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_p$. The OLS estimator of α then is

$$\hat{\alpha} = (P'X'XP)^{-1}P'X'y \tag{21}$$
$$= (P'X'XP)^{-1}P'X'X\hat{\beta}$$
$$= (P'X'XP)^{-1}P'X'XPP'\hat{\beta}$$
$$= P'\hat{\beta}$$

It is now possible to define an ORR estimator of α in two different ways. Firstly, in analogy with $\alpha = P'\beta$ we can set

$$\hat{\alpha}(k) = P'\hat{\beta}(k) \;. \tag{22}$$

Alternatively, following (2) we can write

$$\hat{\alpha}(k) = [I+k(X*'X*)^{-1}]^{-1}\hat{\alpha} \tag{23}$$
$$= [I+k(P'X'XP)^{-1}]^{-1}\hat{\alpha}$$

which, with the use of (2) and (21), becomes

$$\hat{\alpha}(k) = [I+k(P'X'XP)^{-1}]^{-1}P'[I+k(X'X)^{-1}]\hat{\beta}(k). \tag{24}$$

It is not difficult to show that the right-hand-sides of (22) and (24) are equal, that is, that the two definitions of $\hat{\alpha}(k)$ are equivalent. Further, from (23) and the diagonality of (P'X'XP) it follows that

$$\hat{\alpha}_j(k) = \left(\frac{\lambda_j}{\lambda_j+k}\right)\hat{\alpha}_j \tag{25}$$

$$j = 1,2,\ldots,p.$$

3.1 Hoerl, Kennard and Baldwin Rule (HKB)

Hoerl, Kennard and Baldwin (1975) have suggested that the value of k be determined as

$$k_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \tag{26}$$

where $\hat{\sigma}^2 = (y-X\hat{\beta})'(y-X\hat{\beta})/(n-p)$. This suggestion is justified minimizes the sum of the mean square errors is equal to $p\sigma^2/\beta'\beta$. By replacing the unknown parameters by their least squares estimates, we obtain (26).

3.2 Thisted's Modification of the HKB Rule (HKBM)

Thisted (1976) finds that in some subsets of the parameter space, and particularly in the case where there is a high degree of multicollinearity, the OLS estimator tends to have a smaller mean square error more frequently than the HKB estimator because the latter seems to over-shrink the OLS estimator toward the origin. For this reason he suggested modifying the HKB estimator by using

$$k_{HKMB} = \frac{(p-2)\sigma^2}{\hat{\beta}'\hat{\beta}} \quad . \tag{27}$$

Thisted argues that k_{HKBM} is likely to do better for small p because it does not shrink so greatly.

3.3 Wermuth Rule

Wermuth (1972) notes that

$$\text{tr MSE}[\hat{\alpha}(k)] = \sigma^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^2} + k^2 \sum_{i=1}^{p} \frac{\alpha_i^2}{(\lambda_i+k)^2} \quad . \tag{28}$$

By setting the first derivate of the above expression with respect to k equal to zero we get

$$\hat{\sigma}^2 \sum_{i=1}^{p} \frac{\lambda_i}{(\lambda_i+k)^3} = k \sum_{i=1}^{p} \frac{\lambda_i\hat{\alpha}_i^2}{(\lambda_i+k)^3} \quad . \tag{29}$$

The value of k that solves the above equation, say k_W, is then to be used in ORR.

3.4 Dempster Rule

An empirical Bayes estimator proposed by Dempster (1973) and Dempster *et al*. (1977) can be developed as follows. For a prior distribution of β given as $\beta \sim N(0,\omega^2 I)$, and consequently that of α given as $\alpha \sim N(0,\omega^2 I)$, the marginal distribution of $\hat{\alpha}_j$ $(j = 1,2,...,p)$ is given as

$$\hat{\alpha}_j \sim N\{0,\sigma^2(\frac{\omega^2}{\sigma^2} + \frac{1}{\lambda_j})\} \qquad (30)$$

from which it follows that

$$\sum_{i=1}^{p} \frac{\hat{\alpha}_i^2}{\sigma^2(\frac{1}{k}+\frac{1}{\lambda_i})} \sim \chi_p^2 \qquad (31)$$

where $k = \sigma^2/\omega^2$. Dempster suggests replacing σ^2 by $\hat{\sigma}^2$ and, using the fact that $E(\chi_p^2) = p$, setting

$$\sum_{i=1}^{p} \frac{\hat{\alpha}_i^2}{\hat{\sigma}^2(\frac{1}{k}+\frac{1}{\lambda_i})} = p \quad . \qquad (32)$$

The suggested value of k, say k_D, is then obtained by solving (32).

3.5. Sclove Rule

Another empirical Bayesian estimator proposed by Sclove (1973) is based on the idea that since the left-hand side of (31) and $(n-p)\hat{\sigma}^2$ are independent and are distributed as χ_p^2 and $\sigma^2\chi_{n-p}^2$ respectively, it follows that the quantity

$$\sum_{i}^{p} \frac{\hat{\alpha}_i^2}{(\frac{1}{k}+\frac{1}{\lambda_i})} /(n-p)\hat{\sigma}^2 \tag{33}$$

is distributed as $F_{p,n-p}$. By noting that $E(F_{p,n-p}) = p/(n-p-2)$, Sclove suggests calculating k, say k_s, by solving the following equation:

$$\sum_{i}^{p} \frac{\hat{\alpha}_i^2}{\frac{1}{k}+\frac{1}{\lambda_i}} = p\hat{\sigma}^2(\frac{n-p}{n-p-2}) \tag{34}$$

3.6. Criteria for Comparing Estimators

In the past all of the simulation studies of the ridge regression estimators have used square-error loss (either of estimation or of prediction) as the criterion for comparing estimators. Clearly, square-error loss cannot represent all of the loss structures in the decision making problems. Therefore we use a more general measurement of loss, the p'-norm, defined as

$$L_{\beta}^{p'} = \{\sum_j |\hat{\beta}_j(k)-\beta_j|^{p'}\}^{1/p'} \tag{35}$$

and

$$L_{\alpha}^{p'} = \{\sum_j |\hat{\alpha}_j(k)-\alpha_j|^{p'}\}^{1/p'} . \tag{36}$$

We take p = 1,2, and ∞ so that the loss functions considered are

$$L_{\beta} = \sum_j |\hat{\beta}_j(k)-\beta| \qquad (p = 1) \tag{37}$$

$$L_{\beta} = \{\sum_j \text{MSE}[\hat{\beta}_j(k)]\}^{1/2} \qquad (p = 2) \tag{38}$$

$$L_{\beta} = \max\{\hat{\beta}_1(k),\ldots,\hat{\beta}_p(k)\} \qquad (p = \infty) \tag{39}$$

and similarly for α.

Note that for p=2 we have $L_\alpha = L_\beta$, otherwise the values of the loss functions differ.

4. Design of the Monte Carlo Experiment

Unlike in the case of a known k, the small sample properties of an ORR estimators based on a sample-determined value of k are not known. In particular, it is not clear to what extent, if at all, is the mean-square-error dominance of ORR over OLS preserved under these circumstances. Further, it is also not clear what the small-sample performance of the ORR estimators relative to the OLS estimator would be under a loss criterion other than that of the mean-square-error. Finally, it would be instructive to compare the performance of the different ORR estimators discussed in Section 3 above so that there is a basis for making a choice in practical applications. The Monte Carlo experiment whose design is presented below is intended to provide at least tentative answers to these questions.

The performance of a ridge regression estimator based on a given value of k depends on (i) the number and the values of the regression coefficients, (ii) the degree of multicollinearity, and (iii) the value of the variance of the disturbances, σ^2. It can be expected that the same factors would also be relevant for the ORR estimation with unknown k. In the Monte Carlo experiment at hand we take the factors (i) and (ii) into consideration but, following Thisted (1976), leave the value of σ^2 constant (equal to unity) throughout the experiment in order to keep the computer costs down.

4.1 Construction of the Data Sets

In constructing the data sets (and in determining the values of the regression coefficients discussed in the next subsection) we follow, with some modifications, the approach of Dempster *et al*. (1977). Two models, one with 4 explanatory variables and 20 observations and one with 8 explanatory variables and 40 observations, were used in this study.

The values of the explanatory variables have been generated from a standard normal distribution, modified to reflect a low, a medium, and a high degree of multicollinearity, and standardized to be used in a correlation matrix form.[2)] The resulting matrices are denoted by XpX11, XpX21, and XpX31 to represent a low, a medium, and a high degree of multicollinearity, respectively. The values of the determinants of these matrices are presented in Table 1.

To see the degree of multicollinearity among the explanatory variables more clearly, the multiple correlation coefficients of each individual explanatory variable on all of the other explanatory variables have been calculated. For each data set the highest of these multiple correlation coefficients can serve as a convenient measure of multicollinearity.[3)] The results of the calculations are presented in Table 2.

Given a model, a design matrix X'X, a true coefficient vector β, and $\sigma^2=1$, the values of ε should be generated from $N(0,\sigma^2 I_n)$, and the values of the dependent variable y should be obtained through the relation $y=X\beta+\varepsilon$. But since the distribution of $\hat{\beta}$, the OLS estimator of β, is well known, the values of y need not be actually calculated; instead, the values of $\hat{\beta}$ can be generated directly from $N(\beta,(X'X)^{-1})$. The values of $\hat{\beta}(k)$ have been calculated using equation (2).[4)]

4.2 Determination of the Values of the Regression Coefficients

The sets of the true regression coefficients to be used in this study are determined by two factors, the shape and the noncentrality of the coefficients. The first factor determines the patterns of the coefficient vector while the second factor determines the size of the vector. Two shapes of coefficients are used.

Shape 1. The coefficients are in the following pattern:

[1 1 1 1] for the 4-variable model

[1 1 1 1 1 1 1 1] for the 8-variable model.

Shape 2. The coefficients are in the following pattern:

[0 0 1 0] for the 4-variable model

[0 0 0 0 1 0 0 0] for the 8-variable model.

The second factor, the noncentrality pattern δ, is defined as

$$\delta = \frac{\beta'\beta}{tr(X'X)} \quad . \tag{40}$$

To see the sensitivity of the estimation results to the variation in δ, we use the values $\delta=5$, $\delta=20$, and $\delta=35$.

The shape and noncentrality parameter jointly determine the following sets of coefficients used in this study.

For the 4-variable model

B11 = (2.2361, 2.2361, 2.2361, 2.2361)

B12 = (0 , 0 , 4.4721, 0)

B21 = (4.4721, 4.4721, 4.4721, 4.4721)

B22 = (0 , 0 , 8.9443, 0)

B31 = (5.9161, 5.9161, 5.9161, 5.9161)

B32 = (0 , 0 ,11.8332, 0).

For the 8-variable model

B11 = (2.2361, 2.2361, 2.2361, 2.2361, 2.2361, 2.2361, 2.2361, 2.2361)

B12 = (0 , 0 , 0 , 0 , 6.3246, 0 , 0 ,)

B21 = (4.4721, 4.4721, 4.4721, 4.4721, 4.4721, 4.4721, 4.4721, 4.4721)

B22 = (0 , 0 , 0 , 0 ,12.6491, 0 , 0 , 0)

B31 = (5.9161, 5.9161, 5.9161, 5.9161, 5.9161, 5.9161, 5.9161, 5.9161)

B32 = (0, 0 , 0 , 0 ,16.7332, 0 , 0 , 0)

In the last subsection 3 X'X matrices with different degrees of ill-conditioning were constructed for each of the two models. In this subsection 6 different sets of coefficients with different combinations of shapes and values of δ were determined for each model. The combination of all the factors yield 3×6 = 18 different designs for each model. Each design is identified by a four-digit number (for example, Design 1221) in which the right-most digit is always 1. The second digit from the right identifies the degree of ill-conditioning. It takes values 1,2, and 3 for XpX11, XpX21, and XpX31, respectively. The third digit from the right identifies the shape of the coefficient, 1 for shape 1, and 2 for shape 2. The fourth digit identifies the values of the noncentrality. It takes a value 1 when $\delta=5$, 2 when $\delta=20$, and 3 when $\delta=35$.

4.3 Determination of the Number of Replications.

The performance of the estimators considered in this study is to be judged by the size of the average loss.

Since the properties of the distribution of the losses of the ORR estimators are not known, the number of replications is based on the distribution of square error loss of OLS. Let $L_1, \ldots L_n$ be the square error losses of the OLS estimator with n replication. It is known that, if the error ε is normally distributed,

$$E(L_i) = \sigma^2 tr(X'X)^{-1} = \sigma^2 \sum_{i=1}^{p} \frac{1}{\lambda_i} \tag{41}$$

and

$$V(L_i) = 2\sigma^4 tr(X'X)^{-2} = 2\sigma^4 \sum_{i=1}^{p} \left(\frac{1}{\lambda_i}\right)^2 \tag{42}$$

for i=1,...,n. Therefore it follows from the basic sampling theory that

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

$$V(\bar{L}) = \frac{1}{n}V(L_i) = \frac{2\sigma^4}{n} \sum_{i=1}^{p} (\frac{1}{\lambda_i})^2$$

and the coefficient of variation of $\bar{L}$ is

$$CV(\bar{L}) = \frac{\sqrt{V(\bar{L})}}{E(\bar{L})} = \sqrt{\frac{2}{n} \sum_{1}^{p} (\frac{1}{\lambda_i})^2} / \Sigma(\frac{1}{\lambda_i}) \quad . \qquad (43)$$

If $X'X = I$, i.e., $\lambda_1 = \ldots = \lambda_p = 1$, then

$$CV(\bar{L}) = \sqrt{\frac{2}{np}} \quad .$$

If $X'X$ is extremely collinear, then

$$CV(\bar{L}) = \sqrt{\frac{2}{n}} \quad \text{since} \quad \sqrt{\Sigma(\frac{1}{\lambda_i})^2} / \Sigma\frac{1}{\lambda_i} \to 1$$

as some $\lambda_i \to 0$. Therefore we have

$$\sqrt{\frac{2}{np}} \leq CV(I) \leq \sqrt{\frac{2}{n}} \; .$$

If the coefficient of variation of $\bar{L}$ (the average square loss of the least squares estimator) is to be less than 5%, that is, if

$$CV(\bar{L}) \leq 0.05$$

the number of the replications should be as follows.

For the 4-variable model:

Data Set Used	Number of Replication Needed
XpX11	346
XpX21	747
XpX31	783

For the 8-variable model:

Data Set Used	Number of Replication Needed
XpX11	189
XpX21	617
XpX31	702

From the above calculations it appears that up to 800 replications should be used. Because of cost considerations, we used 500 replications for each model and each data set. With this number of replications 6.3% accuracy is achieved in the 4-variable model when the most ill-conditioned data set XpX31 is used, and 5.5% in the 8-variable model when XpX31 is used.

5. Evaluation of Results

5.1 Presentation of Results

In each of the different regression problems of each model the following estimators of the regression coefficients have been computed:

(i) The least squares estimators $\hat{\beta}$ and $\hat{\alpha}$.

(ii) The various types of ridge estimators in both original and principal component forms, that is, $\hat{\beta}(k_\ell)$ and $\hat{\alpha}(k_\ell)$ for ℓ = HKB, HKBM, D, W, and S.

The following statistics based on the 500 replications have been computed for all the estimators, for three different loss structures, and for both original and principal component forms:

(i) The average loss.

(ii) The standard deviation of loss.

(iii) The number of times that $L_{\beta}^{P'}(0) < L_{\beta}^{P'}(k_{\ell})$ (and $L_{\alpha}^{P'}(0) < L_{\alpha}^{P'}(k_{\ell})$) for $P' = \infty, 1, 2$, and ℓ = HKB, HKBM, D, W, and S.

These statistics are presented in Tables 3 through 20.

5.2 Summary of the Main Results

Regardless of the loss structure used in the experiment the following results are apparent.

(a) The ORR estimators never perform significantly worse than OLS, and they perform very much better in many regressions.

(b) The advantage of the ORR estimators over OLS is the greater

(i) the higher the degree of multicollinearity;

(ii) the lower the value of the noncentrality parameter;

(iii) to a lesser extent, the higher the number of explanatory variables.[5]

(c) The shape of the regression coefficients affects the performance of the ridge estimators. In both models, other things unchanged, the improvement the ridge estimators can achieve is greater when the shape is [0 0 1 0] or [0 0 0 0 1 0 0 0] than when the shape is [1 1 1 1] or [1 1 1 1 1 1 1 1].

(d) With a very few exceptions, the HKBM estimator is dominated by the HKB estimator. The sums of the simple ranks for each loss structure over the 36 regressions used in the experiment are as follows.[6)]

<u>Mean Square Error Loss</u>:	Sclove	74
	Dempster	78
	HKB	108
	Wermuth	138
	HKBM	142
<u>Mean Absolute Error Loss</u>:	Sclove	75
	Dempster	84
	HKB	109
	Wermuth	132
	HKBM	140
<u>Maximum Absolute Error Loss</u>:	Dempster	76
	Sclove	77
	HKB	101
	HKBM	135
	Wermuth	151

Although approximately the same results were used regardless of the loss structure used, the magnitude of the improvement of ORR over OLS is notably smaller when the absolute error (average or maximum) rather than the mean square error criterion is used. This is, of course, to be expected since the ORR estimators are especially designed to reduce the mean square error relative to OLS.

6. Concluding Remarks

The ORR estimator with a given k is a linear estimator which is biased but which, for values of k in a certain interval, has a smaller mean square error than the OLS estimator. Since the interval of dominance of ORR over OLS depends on the true values of the regression parameters, the advantage of ORR (of this type) over OLS is for practical purposes illusory. The various interpretations of the ORR estimator offered in Section 2 above, however, indicate that if we do have some prior knowledge about the parameter space of β, and if this knowledge is sufficiently sharp, the ORR estimation provides a convenient and simple way of incorporating such knowledge in estimation and of reducing the size of the mean square error.

When the value of k is not given *a priori* and has to be determined from sample observations, the resulting ORR estimators are no longer linear andcan compete with OLS on equal grounds of the same prior information. The results of our Monte Carlo experiment indicate that, in general, the ORR estimators do outperform the OLS estimator very substantially when the degree of multicollinearity is medium or high, even when a loss criterion other than that of mean square error is used.

In examining the performance of the various ORR estimators considered in this study, it is apparent that the empirical Bayes estimators (i.e., those proposed by Dempster and by Sclove) lead the pack. The disadvantage of these estimators, though, is the difficulty and the messiness of computation. It may thus be reasonable in practical applications to use the estimator proposed by Hoerl, Kennard, and Baldwin (1975) which is simple to calculate and which performs also very well relative to OLS. The modification of this estimator proposed by Thisted (1976) has not worked out too well, and neither has the estimator of Wermuth (1972) which, in addition, is hard to compute. On the basis of our experiment neither of the two last-mentioned estimators can be recommended.

In drawing our conclusions we should be reminded of the fact that the assessment of the ORR and OLS estimators is based entirely on the loss in estimation. Since the small sample properties of the (nonlinear) ORR estimators are not known, the ORR procedure is not suited for testing hypotheses. This makes ORR uninteresting for many econometric problems. It would seem, though, that ORR may well become a powerful tool in forecasting, particularly in situations where a high degree of multicollinearity makes the OLS forecasts unstable.

TABLE 1

Determinant Values of the XpX Matrices

4-Variable Model

det (XpX11) = 0.39454

det (XpX21) = 0.01594

det (XpX31) = 0.004954

8-Variable Model

det(XpX11) = 0.11827

det(XpX21) = 0.00119

det(XpX31) = 0.00003

TABLE 2

Calculations of Multiple Correlation Coefficients of Independent Variables

Data	Regression	Regression Coefficients	R^2
		4-Variable Model	
XpX11	$X_1/X_2,X_3,X_4$	-0.698, 0.232, -0.294	0.291
	$X_2/X_1,X_3,X_4$	-0.396, 0.478, -0.407	0.597
	$X_3/X_1,X_2,X_4$	0.206, 0.748, 0.306	0.370
	$X_4/X_1,X_2,X_3$	-0.294, -0.716, 0.345	0.291
XpX21	$X_1/X_2,X_3,X_4$	-2.199, 1.085, -1.111	0.893
	$X_2/X_1,X_3,X_4$	-0.397, 0.515, -0.109	0.980
	$X_3/X_1,X_2,X_4$	0.680, 1.787, 0.900	0.933
	$X_4/X_1,X_2,X_3$	-0.719, -1.823, 0.928	0.931
XpX31	$X_1/X_2,X_3,X_4$	-2.436, 1.198, -1.234	0.963
	$X_2/X_1,X_3,X_4$	-0.393, 0.497, -0.509	0.994
	$X_3/X_1,X_2,X_4$	0.761, 1.956; 0.989	0.976
	$X_4/X_1,X_2,X_3$	-0.751, -1.918, 0.948	0.9772
		8-Variable Model	
XpX11	X_1/others	Omitted	0.592
	X_2/others		0.547
	X_3/others		0.561
	X_4/others		0.306

Table 2 (Continued)

Data	Regression	Regression Coefficients	R^2
		8-Variable Model	
XpX11	X_5/others	Omitted	0.248
	X_6/others		0.446
	X_7/others		0.371
	X_8/others		0.542
XpX21	X_1/others		0.977
	X_2/others		0.969
	X_3/others		0.954
	X_4/others		0.670
	X_5/others		0.655
	X_6/others		0.953
	X_7/others		0.902
	X_8/others		0.960
XpX31	X_1/others		0.991
	X_2/others		0.988
	X_3/others		0.983
	X_4/others		0.767
	X_5/others		0.779
	X_6/others		0.986
	X_7/others		0.960
	X_8/others		0.979

TABLE 3

4 Variables, 20 Observations, Design 1111

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	$P' = \infty$	1.97696	1.00720		
	A	$P' = 1$	3.83375	1.51744		
	A,B	$P' = 2$	6.73029	6.02392		
	B	$P' = \infty$	1.81186	0.75321		
	B	$P' = 1$	4.10552	1.94685		
HKB	A	$P' = \infty$	1.73360	0.73095	0.87690	196
	A	$P' = 1$	3.62562	1.31289	0.94571	203
	A,B	$P' = 2$	5.38974	3.87112	0.80082	210
	B	$P' = \infty$	1.64997	0.58333	0.91065	206
	B	$P' = 1$	3.77358	1.52757	0.91915	208
HKBM	A	$P' = \infty$	1.79035	0.81969	0.90561	187
	A	$P' = 1$	3.62690	1.35215	0.94604	172
	A,B	$P' = 2$	5.62652	4.50050	0.83600	188
	B	$P' = \infty$	1.68300	0.64146	0.92888	180
	B	$P' = 1$	3.79841	1.64395	0.92519	178
Dempster	A	$P' = \infty$	1.76725	0.69821	0.89392	213
	A	$P' = 1$	3.79550	1.37892	0.99002	235
	A,B	$P' = 2$	5.74768	3.86075	0.85400	228
	B	$P' = \infty$	1.67003	0.56288	0.92173	228
	B	$P' = 1$	3.96813	1.58639	0.96654	225
Wermuth	A	$P' = \infty$	1.81929	0.78928	0.92025	198
	A	$P' = 1$	3.72688	1.34143	0.97212	203
	A,B	$P' = 2$	5.83925	4.37954	0.86761	210
	B	$P' = \infty$	1.71213	0.61287	0.94496	202
	B	$P' = 1$	3.91538	6.61033	0.95369	197
Sclove	A	$P' = \infty$	1.77619	0.67680	0.89844	222
	A	$P' = 1$	3.86099	1.40389	1.00711	254
	A,B	$P' = 2$	5.88509	3.82794	0.87442	241
	B	$P' = \infty$	1.67510	0.55517	0.92452	237
	B	$P' = 1$	4.04781	1.60230	0.98594	240

Factors used: 1) Coordinate system #1. 2) EI1 = [10 8 5 1]
3) $\delta = \beta'\beta/tr(X'X) = 5$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX11 2) Parameter vector B11

A: principal component form B: original form

TABLE 4

4 Variables, 20 Observation, Design 1121

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(O) Out of 500
OLS	A	$P' = \infty$	7.50875	5.45809		
	A	$P' = 1$	9.73959	5.64434		
	A,B	$P' = 2$	88.9236	117.381		
	B	$P' = \infty$	5.87520	4.12414		
	B	$P' = 1$	14.1968	10.0631		
HKB	A	$P' = \infty$	3.34691	2.68892	0.44574	87
	A	$P' = 1$	5.52808	2.91111	0.56759	105
	A,B	$P' = 2$	21.0191	39.5339	0.23637	90
	B	$P' = \infty$	2.82988	1.99122	0.48167	87
	B	$P' = 1$	6.68027	4.88294	0.47055	89
HKBM	A	$P' = \infty$	4.12447	3.62961	0.54929	78
	A	$P' = 1$	6.29570	3.82753	0.64640	85
	A,B	$P' = 2$	32.7538	61.8084	0.36834	84
	B	$P' = \infty$	3.40078	2.70307	0.57884	80
	B	$P' = 1$	8.04792	6.62166	0.56688	80
Dempster	A	$P' = \infty$	2.59282	1.36283	0.34531	93
	A	$P' = 1$	5.09211	1.70442	0.52283	122
	A,B	$P' = 2$	11.9878	18.3533	0.13481	102
	B	$P' = \infty$	2.28045	1.01916	0.38815	96
	B	$P' = 1$	5.74137	2.62389	0.40441	112
Wermuth	A	$P' = \infty$	2.45183	0.40858	0.32653	84
	A	$P' = 1$	5.26325	1.02268	0.54040	109
	A,B	$P' = 2$	10.4172	3.72918	0.11715	96
	B	$P' = \infty$	2.19769	0.36034	0.37406	91
	B	$P' = 1$	5.84381	1.20561	0.41163	102
Sclove	A	$P' = \infty$	2.53540	1.12367	0.33766	93
	A	$P' = 1$	5.10914	1.53612	0.52458	130
	A,B	$P' = 2$	11.319	14.4443	0.12729	103
	B	$P' = \infty$	2.24202	0.84230	0.38161	87
	B	$P' = 1$	5.72476	2.26039	0.40324	119

Factors used: 1) Coordinate system #1   2) EI2 = [30 15 6 0.1]
3) $\delta = \beta'\beta/tr(X'X) = 5$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX21   2) Parameter vector B11

A: principal component form   B: original form

TABLE 5

4 Variables, 20 Observation, Design 1131

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	P′ = ∞	13.2312	9.84503		
	A	P′ = 1	15.5176	9.94256		
	A,B	P′ = 2	274.879	371.344		
	B	P′ = ∞	10.3129	7.58909		
	B	P′ = 1	24.7517	18.2217		
HKB	A	P′ = ∞	4.80502	5.05833	0.36316	51
	A	P′ = 1	6.99956	5.18712	0.45107	53
	A,B	P′ = 2	51.2816	123.819	0.18656	55
	B	P′ = ∞	3.93423	3.84347	0.38149	57
	B	P′ = 1	9.28257	9.26437	0.37503	57
HKBM	A	P′ = ∞	6.30677	6.77884	0.47666	51
	A	P′ = 1	8.51672	6.89508	0.54884	53
	A,B	P′ = 2	88.3901	194.303	0.32156	54
	B	P′ = ∞	5.07679	5.16986	0.49228	55
	B	P′ = 1	12.0332	12.4568	0.48616	57
Dempster	A	P′ = ∞	2.84529	2.33016	0.21505	52
	A	P′ = 1	5.30203	2.45236	0.34168	66
	A,B	P′ = 2	16.9249	63.3557	0.06157	59
	B	P′ = ∞	2.46812	1.76895	0.23932	58
	B	P′ = 1	6.15169	4.33448	0.24854	64
Wermuth	A	P′ = ∞	2.52789	0.51794	0.19106	52
	A	P′ = 1	5.35044	0.94069	0.34480	64
	A,B	P′ = 2	11.1427	6.70392	0.04054	61
	B	P′ = ∞	2.24069	0.40819	0.21727	56
	B	P′ = 1	6.09543	1.28564	0.24626	66
Sclove	A	P′ = ∞	2.7186	1.72171	0.20547	55
	A	P′ = 1	5.23653	1.91565	0.33746	67
	A,B	P′ = 2	13.9587	36.8649	0.05078	61
	B	P′ = ∞	2.37539	1.30561	0.23033	58
	B	P′ = 1	6.00202	3.267	0.24249	65

Factors used: 1) Coordinate system #1 2) EI3 = [50 20 10 0.05]
3) $\delta = \beta'\beta/tr(X'X) = 5$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX31 2) Parameter vector B11

A: principal component form B: original form

TABLE 6

4 Variables, 20 Observation, Design 1211

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	P´ = ∞	1.97696	1.0072		
	A	P´ = 1	3.83375	1.51744		
	A,B	P´ = 2	6.73029	6.02392		
	B	P´ = ∞	1.81186	0.75321		
	B	P´ = 1	4.10552	1.94685		
HKB	A	P´ = ∞	1.66727	0.70714	0.84335	170
	A	P´ = 1	3.54674	1.30482	0.92514	193
	A,B	P´ = 2	5.08039	3.73144	0.75485	174
	B	P´ = ∞	1.65646	0.64741	0.91423	165
	B	P´ = 1	3.57704	1.40375	0.87128	183
HKBM	A	P´ = ∞	1.75506	0.81754	0.88776	153
	A	P´ = 1	3.57978	1.37111	0.93376	153
	A,B	P´ = 2	5.46381	4.48126	0.81183	137
	B	P´ = ∞	1.67031	0.67695	0.92188	132
	B	P´ = 1	3.71064	1.61526	0.90382	154
Dempster	A	P´ = ∞	1.66994	0.66515	0.84470	185
	A	P´ = 1	3.66291	1.34501	0.95544	210
	A,B	P´ = 2	5.23017	3.62608	0.77711	200
	B	P´ = ∞	1.72204	0.68537	0.95043	195
	B	P´ = 1	3.57923	1.31867	0.87181	211
Wermuth	A	P´ = ∞	1.7484	0.76696	0.88439	175
	A	P´ = 1	3.67322	1.33332	0.95813	188
	A,B	P´ = 2	5.52122	4.24091	0.82035	180
	B	P´ = ∞	1.712	0.67430	0.94489	168
	B	P´ = 1	3.73471	1.49982	0.90968	189
Sclove	A	P´ = ∞	1.66686	0.63684	0.84314	195
	A	P´ = 1	3.71974	1.35687	0.97026	221
	A,B	P´ = 2	5.28638	3.52731	0.78546	216
	B	P´ = ∞	1.75303	0.70071	0.96753	209
	B	P´ = 1	3.57202	1.25509	0.87005	215

Factors used: 1) Coordinate system #1 2) EI1 = [10 8 5 1]
3) $\delta = \beta'\beta/tr(X'X) = 5$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX11 2) Parameter vector B12

A: principal component form B: original form

TABLE 7

4 Variables, 20 Observations, Design 1221

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)> L(0) Out of 500
OLS	A	$P' = \infty$	7.50875	5.45809		
	A	$P' = 1$	9.73959	5.64434		
	A,B	$P' = 2$	88.9236	117.381		
	B	$P' = \infty$	5.87520	4.12414		
	B	$P' = 1$	14.1968	10.0631		
HKB	A	$P' = \infty$	3.13077	2.76830	0.41695	61
	A	$P' = 1$	5.25333	3.01297	0.53938	70
	A,B	$P' = 2$	19.8525	40.0885	0.22325	69
	B	$P' = \infty$	2.70245	2.04486	0.45998	72
	B	$P' = 1$	6.26951	5.01894	0.44147	74
HKBM	A	$P' = \infty$	3.95631	3.72633	0.52689	54
	A	$P' = 1$	6.08859	3.95064	0.62514	70
	A,B	$P' = 2$	31.9849	62.3575	0.35969	67
	B	$P' = \infty$	3.29381	2.76611	0.56063	60
	B	$P' = 1$	7.75833	6.7842	0.54649	67
Dempster	A	$P' = \infty$	2.18604	1.44862	0.29113	64
	A	$P' = 1$	4.48875	1.78897	0.46088	77
	A,B	$P' = 2$	9.56378	20.3617	0.10755	76
	B	$P' = \infty$	2.15363	1.14979	0.36656	83
	B	$P' = 1$	4.64847	2.66154	0.32743	72
Wermuth	A	$P' = \infty$	2.01942	0.44315	0.26894	60
	A	$P' = 1$	4.7831	1.08916	0.49110	80
	A,B	$P' = 2$	7.79519	3.13401	0.08766	69
	B	$P' = \infty$	2.35667	0.61844	0.40112	82
	B	$P' = 1$	4.3145	0.84083	0.30391	69
Sclove	A	$P' = \infty$	2.09389	1.18701	0.27886	64
	A	$P' = 1$	4.45463	1.59956	0.45737	87
	A,B	$P' = 2$	8.57504	15.5224	0.09643	75
	B	$P' = \infty$	2.12574	0.99580	0.36182	88
	B	$P' = 1$	4.47557	2.20046	0.31525	70

Factors used: 1) Coordinate system #1 2) EI2 = [30 15 6 0.1]
3) $\delta = \beta'\beta/tr(X'X) = 5$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX21 2) Parameter vector B12

A: principal component form B: original form

TABLE 8

4 Variables, 20 Observations, Design 1231

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(O) Out of 500
OLS	A	P´ = ∞	13.2312	9.84503		
	A	P´ = 1	15.5176	9.94256		
	A,B	P´ = 2	274.879	371.344		
	B	P´ = ∞	10.3129	7.58909		
	B	P´ = 1	24.7517	18.2217		
HKB	A	P´ = ∞	4.55621	5.17622	0.34435	39
	A	P´ = 1	6.71843	5.33061	0.43296	40
	A,B	P´ = 2	50.0478	124.63	0.18207	40
	B	P´ = ∞	3.77921	3.91657	0.36646	41
	B	P´ = 1	8.88687	9.44767	0.35904	39
HKBM	A	P´ = ∞	6.11257	6.90027	0.46198	37
	A	P´ = 1	8.28736	7.04901	0.53406	39
	A,B	P´ = 2	87.5499	195.05	0.31850	43
	B	P´ = ∞	4.93789	5.25758	0.47881	35
	B	P´ = 1	11.7263	12.6517	0.47376	41
Dempster	A	P´ = ∞	2.31643	2.38596	0.17507	42
	A	P´ = 1	4.6115	2.54452	0.29718	45
	A,B	P´ = 2	13.7259	59.1812	0.04993	40
	B	P´ = ∞	2.25938	1.83653	0.21908	49
	B	P´ = 1	4.91127	4.37501	0.19842	41
Wermuth	A	P´ = ∞	2.04423	0.37133	0.15450	38
	A	P´ = 1	4.84041	0.92322	0.31193	41
	A,B	P´ = 2	7.91314	2.48927	0.02879	39
	B	P´ = ∞	2.44731	0.55461	0.23731	42
	B	P´ = 1	4.28263	0.65151	0.17302	40
Sclove	A	P´ = ∞	2.18103	2.05906	0.16484	42
	A	P´ = 1	4.51623	2.24332	0.29104	48
	A,B	P´ = 2	11.727	55.6343	0.04266	40
	B	P´ = ∞	2.1912	1.61356	0.21247	49
	B	P´ = 1	4.6473	3.78633	0.18776	40

Factors used: 1) Coordinate system #1 2) EI3 = [50 20 10 0.05]
3) $\delta = \beta'\beta/tr(X'X) = 5$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX31 2) Parameter vector B12

A: principal component form B: original form

TABLE 9

4 Variables, 20 Observations, Design 2111

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	$P' = \infty$	1.97696	1.0072		
	A	$P' = 1$	3.83375	1.51774		
	A,B	$P' = 2$	6.73029	6.02392		
	B	$P' = \infty$	1.81186	0.75321		
	B	$P' = 1$	4.10552	1.94685		
HKB	A	$P' = \infty$	1.93723	0.94044	0.97990	276
	A	$P' = 1$	3.83387	1.47664	1.00003	251
	A,B	$P' = 2$	6.51543	5.48678	0.96808	268
	B	$P' = \infty$	1.79896	0.71843	0.99288	252
	B	$P' = 1$	4.05982	1.83243	0.98887	269
HKBM	A	$P' = \infty$	1.93374	0.95469	0.97814	266
	A	$P' = 1$	3.7879	1.47421	0.98804	227
	A,B	$P' = 2$	6.44532	5.8949	0.95766	251
	B	$P' = \infty$	1.78712	0.72228	0.98635	231
	B	$P' = 1$	4.02312	1.85771	0.97993	244
Dempster	A	$P' = \infty$	1.94595	0.93944	0.98431	276
	A	$P' = 1$	3.86146	1.4898	1.00723	255
	A,B	$P' = 2$	6.59174	5.52311	0.97942	272
	B	$P' = \infty$	1.81028	0.71731	0.99913	256
	B	$P' = 1$	4.09022	1.83734	0.99627	277
Wermuth	A	$P' = \infty$	2.03829	0.99994	1.03102	276
	A	$P' = 1$	3.96299	1.5468	1.03371	251
	A,B	$P' = 2$	7.10665	5.97493	1.05592	260
	B	$P' = \infty$	1.86553	0.75377	1.02962	247
	B	$P' = 1$	4.23171	1.93399	1.03074	269
Sclove	A	$P' = \infty$	1.9583	0.94224	0.99056	281
	A	$P' = 1$	3.89177	1.50278	1.01513	265
	A,B	$P' = 2$	6.68726	5.53819	0.99361	279
	B	$P' = \infty$	1.82072	0.72003	1.00489	265
	B	$P' = 1$	4.1268	1.84077	1.00518	286

Factors used: 1) Coordinate system #1 2) EI1 = [10 8 5 1]
3) $\delta = \beta'\beta/tr(X'X) = 20$ 4) Coefficient shape [1 1 1 1]

Data input: 1) Design matrix XpX11 2) Parameter vector B21

A: principal component form B: original form

TABLE 10

4 Variables, 20 Observation, Design 2121

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K) >L(0) Out of 500
OLS	A	P´ = ∞	7.50875	5.45809		
	A	P´ = 1	9.73959	5.64434		
	A,B	P´ = 2	88.9236	117.381		
	B	P´ = ∞	5.8752	4.12414		
	B	P´ = 1	14.1968	10.0631		
HKB	A	P´ = ∞	4.57091	2.76715	0.60875	170
	A	P´ = 1	6.84102	3.0185	0.70239	171
	A,B	P´ = 2	31.423	42.0576	0.35337	169
	B	P´ = ∞	3.67535	2.06741	0.62557	169
	B	P´ = 1	8.78642	5.05215	0.61890	170
HKBM	A	P´ = ∞	5.153	3.63381	0.68627	152
	A	P´ = 1	7.38174	3.83796	0.75791	153
	A,B	P´ = 2	42.5101	64.0309	0.47805	153
	B	P´ = ∞	4.11319	2.7277	0.70009	155
	B	P´ = 1	9.85388	6.65282	0.69409	151
Dempster	A	P´ = ∞	4.2573	1.99359	0.56698	179
	A	P´ = 1	6.64912	2.34312	0.68269	181
	A,B	P´ = 2	25.3484	26.278	0.28506	179
	B	P´ = ∞	3.43954	1.48956	0.58543	180
	B	P´ = 1	8.26368	3.65369	0.58208	181
Wermuth	A	P´ = ∞	4.67897	0.87838	0.62314	192
	A	P´ = 1	9.20477	2.32666	0.94509	247
	A,B	P´ = 2	35.8725	13.2017	0.40341	214
	B	P´ = ∞	3.87957	0.73771	0.66033	199
	B	P´ = 1	10.9817	2.55127	0.77354	218
Sclove	A	P´ = ∞	4.23369	1.7493	0.56384	181
	A	P´ = 1	6.67845	2.14654	0.68570	183
	A,B	P´ = 2	24.4054	20.7779	0.27445	182
	B	P´ = ∞	3.42492	1.3033	0.58295	182
	B	P´ = 1	8.23177	3.23037	0.57983	183

Factors used: 1) Coordinate system #1 2) EI2 = [30 15 6 0.1]
3) $\delta = \beta'\beta/tr(X'X) = 20$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX21 2) Parameter vector B21

A: principal component form B: original form

TABLE 11

4 Variables, 20 Observation, Design 2131

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
	A	$P' = \infty$	13.2312	9.84503		
	A	$P' = 1$	15.5176	9.94256		
OLS	A,B	$P' = 2$	274.879	371.344		
	B	$P' = \infty$	10.3129	7.58909		
	B	$P' = 1$	24.7517	18.2217		
	A	$P' = \infty$	6.09794	4.83074	0.46088	106
	A	$P' = 1$	8.38509	4.94935	0.54036	105
HKB	A,B	$P' = 2$	63.4131	124.96	0.23070	105
	B	$P' = \infty$	4.84131	3.70275	0.46944	108
	B	$P' = 1$	11.526	8.89425	0.46567	103
	A	$P' = \infty$	7.4165	6.50686	0.56053	100
	A	$P' = 1$	9.68579	6.60208	0.62418	99
HKBM	A,B	$P' = 2$	100.183	195.305	0.36446	98
	B	$P' = \infty$	5.85945	4.99352	0.56817	101
	B	$P' = 1$	13.9739	12.0024	0.56456	97
	A	$P' = \infty$	4.904	2.38995	0.37064	112
	A	$P' = 1$	7.33294	2.58684	0.47256	115
Dempster	A,B	$P' = 2$	33.0985	50.8505	0.12041	111
	B	$P' = \infty$	3.92621	1.82372	0.38071	115
	B	$P' = 1$	9.3331	4.42644	0.37707	109
	A	$P' = \infty$	4.858	0.41573	0.36716	112
	A	$P' = 1$	9.55912	1.79953	0.61602	142
Wermuth	A,B	$P' = 2$	38.8612	9.90598	0.14138	125
	B	$P' = \infty$	4.01987	0.42438	0.38979	116
	B	$P' = 1$	11.6384	1.9912	0.47021	127
	A	$P' = \infty$	4.80871	2.07469	0.36344	114
	A	$P' = 1$	7.28546	2.31217	0.46950	116
Sclove	A,B	$P' = 2$	30.9347	44.8917	0.11254	115
	B	$P' = \infty$	3.85276	1.59159	0.37359	116
	B	$P' = 1$	9.18055	3.84689	0.37091	110

Factors used: 1) Coordinate system #1 2) EI3 = [50 20 10 0.05]
3) $\delta = \beta'\beta/tr(X'X) = 20$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX31 2) Parameter vector B21

A: principal component form B: original form

TABLE 12

4 Variables, 20 Observation, Design 2211

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	P´ = ∞	1.97696	1.0072		
	A	P´ = 1	3.83375	1.51744		
	A,B	P´ = 2	6.73029	6.02392		
	B	P´ = ∞	1.81186	0.75321		
	B	P´ = 1	4.10552	1.94685		
HKB	A	P´ = ∞	1.8961	0.91364	0.95910	265
	A	P´ = 1	3.78838	1.47938	0.98817	234
	A,B	P´ = 2	6.29414	5.23561	0.93520	250
	B	P´ = ∞	1.76668	0.72775	0.97506	223
	B	P´ = 1	3.99554	1.77468	0.97321	247
HKBM	A	P´ = ∞	1.91253	0.94916	0.96741	248
	A	P´ = 1	3.7731	1.48012	0.98418	219
	A,B	P´ = 2	6.3602	5.49406	0.94501	233
	B	P´ = ∞	1.76865	0.73178	0.97615	207
	B	P´ = 1	4.00314	1.84425	0.97506	227
Dempster	A	P´ = ∞	1.89786	0.90994	0.95999	267
	A	P´ = 1	3.80248	1.47978	0.99185	237
	A,B	P´ = 2	6.31633	5.22718	0.93849	255
	B	P´ = ∞	1.77376	0.72883	0.97897	227
	B	P´ = 1	4.00589	1.76269	0.97573	251
Wermuth	A	P´ = ∞	2.00756	0.95997	1.01548	276
	A	P´ = 1	3.97794	1.57056	1.03761	233
	A,B	P´ = 2	6.98476	5.69546	1.03781	259
	B	P´ = ∞	1.85454	0.77791	1.02355	227
	B	P´ = 1	4.21108	1.84813	1.02571	256
Sclove	A	P´ = ∞	1.89952	0.90353	0.96083	272
	A	P´ = 1	3.82123	1.49126	0.99673	242
	A,B	P´ = 2	6.35136	5.18492	0.94370	265
	B	P´ = ∞	1.78238	0.73300	0.98373	232
	B	P´ = 1	4.01442	1.74932	0.97781	255

Factors used: 1) Coordinate system #1 2) EI1 = [10 8 5 1]
3) $\delta = \beta'\beta/tr(X'X) = 5$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX11 2) Parameter vector B22

A: principal component form B: original form

TABLE 13

4 Variables, 20 Observations, Design 2221

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	$P' = \infty$	7.50875	5.45809		
	A	$P' = 1$	9.73959	5.64434		
	A,B	$P' = 2$	88.9236	117.381		
	B	$P' = \infty$	5.8752	4.12414		
	B	$P' = 1$	14.1968	10.0631		
HKB	A	$P' = \infty$	4.10834	2.82058	0.54714	115
	A	$P' = 1$	6.34545	3.07799	0.65151	116
	A,B	$P' = 2$	27.5633	42.949	0.30997	118
	B	$P' = \infty$	3.3291	2.10604	0.56664	112
	B	$P' = 1$	7.95861	5.14542	0.56059	117
HKBM	A	$P' = \infty$	4.83256	3.75143	0.64359	103
	A	$P' = 1$	7.03397	3.98244	0.72220	108
	A,B	$P' = 2$	40.0807	65.2216	0.45073	107
	B	$P' = \infty$	3.87435	2.80919	0.65944	106
	B	$P' = 1$	9.2772	6.87122	0.65347	110
Dempster	A	$P' = \infty$	3.57022	2.03753	0.47547	122
	A	$P' = 1$	5.89682	2.3493	0.60545	128
	A,B	$P' = 2$	19.8371	28.8658	0.22308	125
	B	$P' = \infty$	2.95146	1.52034	0.50236	118
	B	$P' = 1$	7.03659	3.69273	0.49565	121
Wermuth	A	$P' = \infty$	3.76584	0.88234	0.50153	143
	A	$P' = 1$	8.20624	2.5385	0.84257	207
	A,B	$P' = 2$	25.0808	11.3569	0.28205	165
	B	$P' = \infty$	4.1525	1.40389	0.70679	179
	B	$P' = 1$	7.52448	1.64142	0.53001	149
Sclove	A	$P' = \infty$	3.46261	1.77004	0.46114	124
	A	$P' = 1$	5.81649	2.12342	0.59720	134
	A,B	$P' = 2$	18.1304	22.2712	0.20389	130
	B	$P' = \infty$	2.87441	1.32599	0.48925	121
	B	$P' = 1$	6.85392	3.19653	0.48278	123

Factors used: 1) Coordinate system #1 2) EI2 = [30 15 6 0.1]
2) $\delta = \beta'\beta/tr(X'X) = 20$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX21 2) Coefficient vector B22

A: principal component form B: original form

TABLE 14

4 Variables, 20 Observations, Design 2231

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	$P' = \infty$	13.2312	9.84503		
	A	$P' = 1$	15.5176	9.94256		
	A,B	$P' = 2$	274.879	371.344		
	B	$P' = \infty$	10.3129	7.58909		
	B	$P' = 1$	24.7517	18.2217		
HKB	A	$P' = \infty$	5.50288	5.02217	0.41590	77
	A	$P' = 1$	7.7673	5.15278	0.50055	81
	A,B	$P' = 2$	58.3073	127.137	0.21212	78
	B	$P' = \infty$	4.40015	3.83933	0.42667	74
	B	$P' = 1$	10.5018	9.22235	0.42429	78
HKBM	A	$P' = \infty$	6.96587	6.73774	0.52647	74
	A	$P' = 1$	9.2223	6.85285	0.59431	78
	A,B	$P' = 2$	96.7182	197.6	0.35186	75
	B	$P' = \infty$	5.52704	5.16157	0.53594	71
	B	$P' = 1$	13.2028	12.4051	0.53341	76
Dempster	A	$P' = \infty$	3.85186	2.65602	0.29112	78
	A	$P' = 1$	6.21263	2.8361	0.40036	83
	A,B	$P' = 2$	24.9306	61.4718	0.09070	78
	B	$P' = \infty$	3.15869	2.03111	0.30629	76
	B	$P' = 1$	7.53338	4.85544	0.30436	79
Wermuth	A	$P' = \infty$	3.89365	0.61691	0.29428	84
	A	$P' = 1$	8.58925	2.15303	0.55352	129
	A,B	$P' = 2$	26.6273	9.06662	0.09687	97
	B	$P' = \infty$	4.47345	1.23782	0.43377	112
	B	$P' = 1$	7.55982	1.07224	0.30543	83
Sclove	A	$P' = \infty$	3.68158	2.28428	0.27825	78
	A	$P' = 1$	6.06367	2.49639	0.39076	84
	A,B	$P' = 2$	21.8739	55.6435	0.07958	79
	B	$P' = \infty$	3.03145	1.75494	0.29395	75
	B	$P' = 1$	7.22997	4.17399	0.2921	79

Factors used: 1) Coordinate system #1 2) EI3 = [50 20 10 0.05]
3) $\delta = \beta'\beta/tr(X'X) = 20$ 4) Coefficient shape [0 0 1 0]

Input Data: 1) Design matrix XpX31 2) Coefficient vector B22

A: principal component form B: original form

TABLE 15

4 Variables, 20 Observations, Design 3111

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	P´ = ∞	1.97696	1.0072		
	A	P´ = 1	3.83375	1.51744		
	A,B	P´ = 2	6.73029	6.02392		
	B	P´ = ∞	1.81186	0.75321		
	B	P´ = 1	4.10552	1.94685		
HKB	A	P´ = ∞	1.96188	0.97969	0.99237	279
	A	P´ = 1	3.84641	1.49325	1.0033	259
	A,B	P´ = 2	6.65518	5.81239	0.98884	266
	B	P´ = ∞	1.81256	0.74041	1.00039	250
	B	P´ = 1	4.08568	1.89353	0.99517	264
HKBM	A	P´ = ∞	1.95519	0.97889	0.98899	267
	A	P´ = 1	3.80769	1.49396	0.99320	242
	A,B	P´ = 2	6.57341	5.80321	0.97669	255
	B	P´ = ∞	1.80065	0.73631	0.99382	233
	B	P´ = 1	4.05748	1.89674	0.98830	249
Dempster	A	P´ = ∞	1.9702	0.98067	0.99658	280
	A	P´ = 1	3.86472	1.4981	1.00808	266
	A,B	P´ = 2	6.70997	5.82878	0.99698	269
	B	P´ = ∞	1.82018	0.74061	1.0046	250
	B	P´ = 1	4.1067	1.89582	1.00029	271
Wermuth	A	P´ = ∞	2.0521	1.05098	1.03801	278
	A	P´ = 1	3.94728	1.57206	1.02961	249
	A,B	P´ = 2	7.2023	6.41352	1.07013	260
	B	P´ = ∞	1.87299	0.78102	1.03374	245
	B	P´ = 1	4.238	2.0102	1.03227	265
Sclove	A	P´ = ∞	1.97377	0.99027	0.99838	282
	A	P´ = 1	3.88415	1.51603	1.01315	270
	A,B	P´ = 2	6.77556	5.90524	1.00673	272
	B	P´ = ∞	1.82612	0.75008	1.00787	253
	B	P´ = 1	4.12485	1.90983	1.00471	279

Factors used: 1) Coordinate system #1 2) EI1 = [10 8 5 1]
3) $\delta = \beta'\beta/tr(X'X) = 35$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX11 2) Coefficient vector B31

A: principal component form B: original form

TABLE 16

4 Variables, 20 Observations, Design 3121

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	P´ = ∞	7.50875	5.45809		
	A	P´ = 1	9.73959	5.64434		
	A,B	P´ = 2	88.9236	117.381		
	B	P´ = ∞	5.8752	4.12414		
	B	P´ = 1	14.1968	10.0631		
HKB	A	P´ = ∞	5.25195	3.02493	0.69945	193
	A	P´ = 1	7.51885	3.27289	0.77199	196
	A,B	P´ = 2	39.5842	46.0541	0.44515	193
	B	P´ = ∞	4.17017	2.26434	0.70979	194
	B	P´ = 1	10.0146	5.51882	0.70541	194
HKBM	A	P´ = ∞	5.6952	3.81803	0.75848	173
	A	P´ = 1	7.92924	4.02003	0.81413	173
	A,B	P´ = 2	49.771	67.4455	0.55971	173
	B	P´ = ∞	4.51292	2.86787	0.76813	174
	B	P´ = 1	10.8451	6.98618	0.76391	171
Dempster	A	P´ = ∞	5.06019	2.53087	0.67391	202
	A	P´ = 1	7.38858	2.84016	0.75861	204
	A,B	P´ = 2	35.055	33.1512	0.39421	203
	B	P´ = ∞	4.03118	1.88029	0.68614	203
	B	P´ = 1	9.66345	4.61812	0.68068	203
Wermuth	A	P´ = ∞	6.10226	0.99493	0.81269	246
	A	P´ = 1	11.7742	3.28387	1.20891	332
	A,B	P´ = 2	60.3507	21.1468	0.67868	278
	B	P´ = ∞	4.98151	0.84801	0.84789	254
	B	P´ = 1	14.341	3.34052	1.01016	286
Sclove	A	P´ = ∞	5.07801	2.35469	0.67628	205
	A	P´ = 1	7.4387	2.68783	0.76376	211
	A,B	P´ = 2	34.4723	29.797	0.38766	206
	B	P´ = ∞	4.03615	1.75765	0.68698	210
	B	P´ = 1	9.69918	4.30227	0.68320	208

Factors used: 1) Coordinate system #1 2) EI2 = [30 15 6 0.1]
3) $\delta = \beta'\beta/tr(X'X) = 35$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX21 2) Coefficient vector B31

A: principal component form B: original form

TABLE 17

4 Variables, 20 Observations, Design 3131

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	$P' = \infty$	13.2312	9.84503		
	A	$P' = 1$	15.5176	9.94256		
	A,B	$P' = 2$	274.879	371.344		
	B	$P' = \infty$	10.3129	7.58909		
	B	$P' = 1$	24.7517	18.2217		
HKB	A	$P' = \infty$	6.94577	4.83452	0.52496	136
	A	$P' = 1$	9.24429	4.95593	0.59573	135
	A,B	$P' = 2$	24.5318	126.942	0.27114	136
	B	$P' = \infty$	5.47491	3.70899	0.53088	137
	B	$P' = 1$	13.0613	8.90438	0.52769	136
HKBM	A	$P' = \infty$	8.13493	6.46275	0.61483	125
	A	$P' = 1$	10.4114	6.55887	0.67094	124
	A,B	$P' = 2$	110.801	197.066	0.40309	124
	B	$P' = \infty$	6.39601	4.96424	0.62020	126
	B	$P' = 1$	15.279	11.9262	0.61729	124
Dempster	A	$P' = \infty$	6.08317	2.89662	0.47976	145
	A	$P' = 1$	8.47008	3.06204	0.54584	148
	A,B	$P' = 2$	48.5558	64.8965	0.17664	146
	B	$P' = \infty$	4.81002	2.20221	0.46641	144
	B	$P' = 1$	11.4446	5.32566	0.46238	144
Wermuth	A	$P' = \infty$	6.392	0.45076	0.48310	152
	A	$P' = 1$	12.2729	2.59896	0.79091	209
	A,B	$P' = 2$	66.3237	17.1107	0.24128	176
	B	$P' = \infty$	5.21901	0.48253	0.50607	158
	B	$P' = 1$	15.2809	2.59162	0.61737	176
Sclove	A	$P' = \infty$	6.02981	2.44517	0.45573	147
	A	$P' = 1$	8.44827	2.64236	0.54443	151
	A,B	$P' = 2$	45.6066	49.7868	0.16592	147
	B	$P' = \infty$	4.76437	1.86716	0.46198	147
	B	$P' = 1$	11.3401	4.49651	0.45816	145

Factors used: 1) Coordinate system #1 2) EI3 = [50 20 10 0.05]
3) $\delta = \beta'\beta/tr(X'X) = 35$ 4) Coefficient shape [1 1 1 1]

Input data: 1) Design matrix XpX31 2) Coefficient vector B31

A: principal component form B: original form

TABLE 18

4 Variables, 20 Observations, Design 3211

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K) >L(0) Out of 500
OLS	A	$P' = \infty$	1.97696	1.0072		
	A	$P' = 1$	3.83375	1.51744		
	A,B	$P' = 2$	6.73029	6.02392		
	B	$P' = \infty$	1.81186	0.75321		
	B	$P' = 1$	4.10552	1.94685		
HKB	A	$P' = \infty$	1.93507	0.95905	0.97881	276
	A	$P' = 1$	3.81795	1.50497	0.99588	253
	A,B	$P' = 2$	6.51345	5.60331	0.96778	257
	B	$P' = \infty$	1.78501	0.74253	0.98518	248
	B	$P' = 1$	4.05744	1.85787	0.98829	256
HKBM	A	$P' = \infty$	1.94013	0.97607	0.98137	260
	A	$P' = 1$	3.8033	1.49804	0.99206	239
	A,B	$P' = 2$	6.52562	5.72365	0.96959	246
	B	$P' = \infty$	1.78725	0.74221	0.98642	239
	B	$P' = 1$	4.05038	1.89177	0.98657	238
Dempster	A	$P' = \infty$	1.93869	0.96304	0.98064	279
	A	$P' = 1$	3.82379	1.51069	0.99740	255
	A,B	$P' = 2$	6.53882	5.61431	0.97155	262
	B	$P' = \infty$	1.78722	0.74619	0.98640	249
	B	$P' = 1$	4.0661	1.86436	0.99040	263
Wermuth	A	$P' = \infty$	2.04663	1.0271	1.03524	286
	A	$P' = 1$	3.98437	1.60449	1.03929	252
	A,B	$P' = 2$	7.21478	6.21929	1.07199	266
	B	$P' = \infty$	1.86386	0.79098	1.0287	252
	B	$P' = 1$	4.26889	1.9662	1.03979	267
Sclove	A	$P' = \infty$	1.93767	0.96108	0.98012	282
	A	$P' = 1$	3.83957	1.51775	1.00152	256
	A,B	$P' = 2$	6.56476	5.61451	0.97541	269
	B	$P' = \infty$	1.79448	0.74632	0.99041	256
	B	$P' = 1$	4.06755	1.86026	0.99075	262

Factors used: 1) Coordinate system #1 2) EI1 = [10 8 5 1]
3) $\delta = \beta'\beta/tr(X'X) = 35$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX11 2) Coefficient vector B32

A: principal component form B: original form

TABLE 19

4 Variables, 20 Observations, Design 3221

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	P´ = ∞	7.50875	5.45809		
	A	P´ = 1	9.73959	5.64434		
	A,B	P´ = 2	88.9236	117.381		
	B	P´ = ∞	5.8752	4.12414		
	B	P´ = 1	14.1968	10.0631		
HKB	A	P´ = ∞	4.69114	3.10437	0.62476	145
	A	P´ = 1	6.94086	3.27743	0.71264	139
	A,B	P´ = 2	33.8554	46.4166	0.38072	142
	B	P´ = ∞	3.73291	2.24409	0.63537	143
	B	P´ = 1	8.99235	5.50618	0.63341	145
HKBM	A	P´ = ∞	5.33431	3.90105	0.71041	128
	A	P´ = 1	7.54096	4.14523	0.77426	127
	A,B	P´ = 2	46.3527	68.5224	0.52126	129
	B	P´ = ∞	4.22899	2.92445	0.71980	131
	B	P´ = 1	10.1778	7.15837	0.71691	134
Dempster	A	P´ = ∞	4.31527	2.48446	0.57470	150
	A	P´ = 1	6.62647	2.77188	0.68036	153
	A,B	P´ = 2	27.6969	34.1758	0.31147	150
	B	P´ = ∞	3.45675	1.83431	0.58836	151
	B	P´ = 1	8.34296	4.50877	0.58767	152
Wermuth	A	P´ = ∞	4.82861	1.0633	0.64306	186
	A	P´ = 1	10.2237	3.63264	1.04971	274
	A,B	P´ = 2	40.4334	18.6292	0.45470	213
	B	P´ = ∞	5.24381	1.92614	0.89253	235
	B	P´ = 1	9.48193	2.03986	0.66789	193
Sclove	A	P´ = ∞	4.23228	2.28582	0.56365	155
	A	P´ = 1	6.57599	2.56796	0.67518	159
	A,B	P´ = 2	26.1122	31.0835	0.29365	156
	B	P´ = ∞	3.39289	1.68054	0.57749	158
	B	P´ = 1	8.19908	4.13536	0.57753	155

Factors used: 1) Coordinate system #1 2) EI2 = [30 15 6 0.1]
3) $\delta = \beta'\beta/tr(X'X) = 35$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX21 2) Coefficient vector B32

A: principal component form B: original form

TABLE 20

4 Variables, 20 Observations, Design 3231

Estimator	Model Form	Structure Loss	Average Loss	S.D. of Loss	Ratio of Average Loss to OLS	Frequency L(K)>L(0) Out of 500
OLS	A	P´ = ∞	13.2313	9.84503		
	A	P´ = 1	15.5176	9.94256		
	A,B	P´ = 2	274.879	371.344		
	B	P´ = ∞	10.3129	7.58909		
	B	P´ = 1	24.7517	18.2217		
HKB	A	P´ = ∞	6.16576	5.01348	0.46600	95
	A	P´ = 1	8.44044	5.15407	0.54393	94
	A,B	P´ = 2	65.9921	129.869	0.24008	94
	B	P´ = ∞	4.86331	3.84513	0.47158	96
	B	P´ = 1	11.6629	9.23114	0.47120	93
HKBM	A	P´ = ∞	7.56178	6.71777	0.57151	91
	A	P´ = 1	9.82786	6.83724	0.63334	90
	A,B	P´ = 2	105.138	200.386	0.38249	91
	B	P´ = ∞	5.95569	5.15544	0.5775	87
	B	P´ = 1	14.2589	12.3893	0.57608	91
Dempster	A	P´ = ∞	4.82286	3.09178	0.36451	99
	A	P´ = 1	7.16243	3.2724	0.46157	100
	A,B	P´ = 2	35.8296	70.6102	0.13035	98
	B	P´ = ∞	3.83249	2.35601	0.37162	100
	B	P´ = 1	9.22771	5.65327	0.37281	99
Wermuth	A	P´ = ∞	5.10748	0.76024	0.38602	111
	A	P´ = 1	10.9984	3.07317	0.70877	176
	A,B	P´ = 2	44.8498	16.0987	0.16316	135
	B	P´ = ∞	5.78921	1.69102	0.56136	155
	B	P´ = 1	9.73134	1.34396	0.39316	112
Sclove	A	P´ = ∞	4.65638	2.61661	0.35193	100
	A	P´ = 1	7.00693	2.84438	0.45155	101
	A,B	P´ = 2	31.5676	58.0213	0.11484	100
	B	P´ = ∞	3.6891	2.00191	0.35772	100
	B	P´ = 1	8.91269	4.78297	0.36008	101

Factors used: 1) Coordinate system #1   2) EI3 = [50 20 10 0.05]
3) $\delta = \beta'\beta/tr(X'X) = 35$ 4) Coefficient shape [0 0 1 0]

Input data: 1) Design matrix XpX31   2) Coefficient vector B32

A: principal component form   B: original form

FOOTNOTES

*The Monte Carlo results presented in this study have been obtained as a part of Karl Lin's dissertation at the University of Michigan. The work of Jan Kmenta at the University of Bonn has been supported by the Alexander von Humboldt Foundation.

1) For a recent survey of the literature see Vinod (1978).

2) The details of the construction of the data sets and the values of the variables are available on request.

3) For a discussion of this measure see Kmenta (1971).

4) A detailed description is available on request. For the 500 replications used the Monte Carlo experiment the means and the standard deviations of the $\hat{\beta}$'s were very close to their theoretical values. The distributions of the $\hat{\beta}$'s were also found to be very close to normal.

5) Since the results for the 8-variable model have been essentially similar to those for the 4-variable model, their presentation has been omitted to save space. Interested readers may obtain the appropriate tables on request.

6) The ranks refer to the results for the original form (B) of the model.

Dempster, A.P.,(1973). "Alternatives to Least Squares in Multiple Regression", in D. Kabe and R.P. Gupta, eds. Multivariate Statistical Inference, North-Holland, Amsterdam, pp. 25-40.

Dempster, A.P., M. Schatzoff, and N. Wermuth (1977). "A Simulation Study of Alternatives to Ordinary Least Squares" Journal of American Statistical Association, 72, pp. 77-104.

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

Hoerl, A.E. and R.W. Kennard (1970b), "Ridge Regression: Applications to Nonorthogonal Problems", Technometrics, 12, pp. 67-82.

Hoerl, A.E., R.W. Kennard, and K.F. Baldwin (1975), "Ridge Regression: Some Simulation", Communications in Statistics, 4, pp. 105-123.

Kmenta, J. (1971). Elements of Econometrics, Macmillan, New York.

Sclove, S. (1973). "Least Squares with Random Regression Coefficient", Technical Report 87, Economic Series, Stanford University.

Theil, H. and A.S. Goldberger (1961). "On Pure and Mixed Statistical Estimation in Economics", International Economic Review, 2, pp. 65-78.

Theobald, C.M. (1974). "Generalizations of Mean Square Error Applied to Ridge Regression", Journal of the Royal Statistical Society, Series B, 36, pp. 103-106.

Thisted, R.A. (1976). "Ridge Regression, Minimax Estimation, and Empirical Bayes Methods", Technical Report No. 28, Division of Biostatistics, Stanford University.

Vinod, H.D. (1977). "A Survey of Ridge Regression and Related Techniques for Improvments Over Ordinary Least Squares", Review of Economics and Statistics, 60, pp. 121-131.

Wermuth, N. (1972). "An Empirical Comparison of Regression Methods Unpublished Ph.D. thesis,Dep. of Statistics, Harvard University, Cambridge, MA.