RIDGE REGRESSION UNDER ALTERNATIVE LOSS CRITERIA

Karl Lin and Jan Kmenta*

I. Introduction

THE introduction by Hoerl and Kennard (1970) 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 received little 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 these issues.

The basic problem is that of estimating the coefficients of the standard linear regression model

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

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

The ordinary ridge regression estimator (ORR) introduced by Hoerl and Kennard (1970) is defined as

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

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[2] is that if k is such that

$$0 < k < 2\sigma^2/\beta^{*\prime}\beta^* \tag{3}$$

where β^* is the coefficient vector in (1) with each of the explanatory variables normalized so that its sample sum of squares is unity, then ORR dominates OLS in the sense that

$$\text{tr MSE}[\hat{\beta}(k)] < \text{tr MSE}(\hat{\beta}). \tag{4}$$

II. Interpretation of ORR

When k is *fixed* (*nonstochastic*), its role can be interpreted as that of conveying prior information about β. The nature of this information can be assessed in several ways. One of the earliest interpretations of ORR from the Bayesian viewpoint was provided by Lindley and Smith (1972) who noted that if the prior distribution of β is specified as $\beta \sim N(0, \omega^2 I_p)$, then β has the following posterior distribution:

$$\begin{aligned}\beta \sim N\{&[X'X + (\sigma^2/\omega^2)I]^{-1}X'y, \\ &\sigma^2[X'X + (\sigma^2/\omega^2)I]^{-1}\}.\end{aligned} \tag{5}$$

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.

Another interpretation of ORR has been of-

Received for publication August 18, 1981. Revision accepted for publication December 1, 1981.

* Lockheed/EMSCO (Houston) and University of Michigan, respectively.

The Monte Carlo results presented in this study were obtained as a part of Karl Lin's dissertation at the University of Michigan. The work of Jan Kmenta was in part carried out at the University of Bonn and supported by the Alexander von Humboldt Foundation. An earlier version of the paper was presented at the Econometric Society World Congress 1980 in Aix-en-Provence.

[1] In a recent bibliographical survey by Hoerl and Kennard (1981) of ridge regression, the authors list some 211 entries of which only 20 represent articles published in econometric (or economic) journals.

[2] See Theobald (1974).

fered by Newhouse and Oman (1971) who found that $\hat{\beta}(k)$ can be obtained by minimizing $(y - X\beta)'(y - X\beta)$ subject to the restriction that $\beta'\beta = r$, where r is positive and given. The constant k is then identified as the Lagrange multiplier which is related to r by the constraint $\hat{\beta}(k)'\hat{\beta}(k) = r$. A small value of r results in a large value of k and *vice versa*.

Yet another interpretation of ORR can be provided by reference to the mixed estimation method of Theil and Goldberger (1961).[3] Note that the ORR estimator of β can be obtained by application of the least squares method to the following:

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

where 0 is 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

$$\beta_j = (a + b)/2 + u_j \tag{8}$$

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

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

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

$$\begin{aligned}[(a + b)/2][4\sigma/(b - a)] = {} & \beta_1 \times 0 \\ & + \ldots + \beta_j \\ & \times 4\sigma/(b - a) \\ & + \ldots + \beta_p \times 0 \\ & + v_j\end{aligned} \tag{10}$$

where $v_j = [-4\sigma/(b - a)]u_j$. Comparing (10) with (6) we have

$$[(a + b)/2][4\sigma/(b - a)] = 0 \tag{11}$$

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

which, for $b > a$, gives

$$a = -b$$

$$b = 2\sigma/\sqrt{k}.$$

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

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

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.

III. Sample-based Selection of k

If we have prior information that enables us to determine k, and if the value of k falls within the limits specified in (3)—which is always uncertain since these limits involve unknown parameters—then ORR dominates OLS in the mean square error (MSE) sense. In most cases, however, the value of k is not given a priori but is determined on the basis of available sample observations. Under these circumstances the ORR estimator is no longer linear in observations and its properties are unknown. It incorporates *no* prior information but provides a convenient way for trading bias for a reduction in variance. In our study we consider several rules for calculating k suggested by various authors and supported by a reasonable rationalization. Our aim is to compare these estimators with each other and with the OLS estimator under alternative loss structures by means of a Monte Carlo experiment.

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

$$\begin{aligned} y &= X\beta + \epsilon \\ &= XPP'\beta + \epsilon \\ &= X^*\alpha + \epsilon \end{aligned} \tag{14}$$

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

$$PP' = I \tag{15}$$

[3] This interpretation was also suggested—in a general way—by Smith (1976).

$$P'X'XP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{bmatrix} \tag{16}$$

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

$$\begin{aligned} \hat{\alpha} &= (P'X'XP)^{-1}P'X'y \\ &= (P'X'XP)^{-1}P'X'X\hat{\beta} \\ &= (P'X'XP)^{-1}P'X'XPP'\hat{\beta} \\ &= P'\hat{\beta}. \end{aligned} \tag{17}$$

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

$$\hat{\alpha}(k) = P'\hat{\beta}(k). \tag{18}$$

Alternatively, following (2) we can write

$$\begin{aligned} \hat{\alpha}(k) &= [I + k(X^{*\prime}X^{*})^{-1}]^{-1}\hat{\alpha} \\ &= [I + k(P'X'XP)^{-1}]^{-1}\hat{\alpha} \end{aligned} \tag{19}$$

which, with the use of (2) and (17), becomes

$$\begin{aligned} \hat{\alpha}(k) = [I + k(P'X'XP)^{-1}]^{-1}P' \\ \times [I + k(X'X)^{-1}]\hat{\beta}(k). \end{aligned} \tag{20}$$

It is not difficult to show that the right-hand sides of (18) and (20) are equal, that is, that the two definitions of $\hat{\alpha}(k)$ are equivalent. Further, from (19) and from the diagonality of $(P'X'XP)$ it follows that

$$\begin{aligned} \hat{\alpha}_j(k) = [\lambda_j/(\lambda_j + k)]\hat{\alpha}_j; \\ (j = 1, 2, \ldots, p). \end{aligned} \tag{21}$$

We considered the following rules for selecting k.[4]

HKB-rule: Hoerl, Kennard and Baldwin (1975) suggested that the value of k be determined as

$$k_{\text{HKB}} = p\hat{\sigma}^2/\hat{\beta}'\hat{\beta} \tag{22}$$

where $\hat{\sigma}^2 = (y - X\hat{\beta})'(y - X\hat{\beta})/(n - p)$. This suggestion is justified on the grounds that when $X'X = I$, the value of k that minimizes the sum of the mean square errors is equal to $p\sigma^2/\beta'\beta$.

HKBM-rule: Thisted (1976) found that the HKB estimator seems to overshrink towards zero and suggested modifying it by using

$$k_{\text{HKBM}} = (p - 2)\hat{\sigma}^2/\hat{\beta}'\hat{\beta}. \tag{23}$$

Wermuth-rule: Wermuth (1972) noted that the necessary condition for minimizing tr MSE$[\hat{\alpha}(k)]$ with respect to k is that

$$\sigma^2 \sum_{i=1}^{p} \lambda_i/(\lambda_i + k)^3 = k \sum_{i=1}^{p} \lambda_i\alpha_i^2/(\lambda_i + k)^3 \tag{24}$$

and suggested replacing σ^2 by $\hat{\sigma}^2$ and α_i^2 by $\hat{\alpha}_i^2$ and solving for k.

Dempster-rule: Dempster (1973) developed an empirical Bayes estimator for a prior distribution of α given as $\alpha \sim N(0, \omega^2 I)$ from which it follows that

$$\sum_{i=1}^{p} \hat{\alpha}_i^2/\sigma^2[(1/k) + (1/\lambda_i)] \sim \chi_p^2 \tag{25}$$

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

$$\sum_{i=1}^{p} \hat{\alpha}_i^2/\hat{\sigma}^2[(1/k) + (1/\lambda_i)] = p. \tag{26}$$

The suggested value of k is then obtained by solving (26).

Sclove-rule: Another empirical Bayesian estimator proposed by Sclove (1973) is based on the idea that the left-hand side of (25) and $(n - p)\hat{\sigma}^2$ are independent and are distributed as χ_p^2 and $\sigma^2\chi_{n-p}^2$, respectively, so that their ratio is distributed as $F_{p,n-p}$. By noting that $E(F_{p,n-p}) = p/(n - p - 2)$, Sclove suggests calculating k by solving the following equation:

$$\begin{aligned} \sum_{i=1}^{p} \hat{\alpha}_i^2/[(1/k) + (1/\lambda_i)] \\ = p\hat{\sigma}^2[(n - p)/(n - p - 2)]. \end{aligned} \tag{27}$$

To compare the above estimators we use a general measure of loss, the p'-norm, of which the MSE criterion is a special case. The p'-norm is defined as

$$L_\beta^{p'} = \{\Sigma_j|\hat{\beta}_j(k) - \beta_j|^{p'}\}^{1/p'}. \tag{28}$$

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

$$\begin{aligned} L_\beta &= \Sigma_j|\hat{\beta}_j(k) - \beta_j| && (p' = 1) \\ L_\beta &= \{\Sigma_j \text{MSE}[\hat{\beta}_j(k)]\}^{1/2} && (p' = 2) \\ L_\beta &= \max\{|\hat{\beta}_1(k) - \beta_1|, \ldots, |\hat{\beta}_p(k) - \hat{\beta}_p|\} \\ & && (p' = \infty). \end{aligned}$$

[4] These rules are presented in Dempster et al. (1977).

IV. Design of the Monte Carlo Experiment

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.

In constructing the data sets and in determining the values of the regression coefficients we followed, 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.[5] For the 4-variable model the values of the determinants of this correlation matrix were 0.394, 0.016, and 0.005. The corresponding values of the highest R^2 obtained by regressing each explanatory variable on the remaining explanatory variables were 0.597, 0.980, and 0.994.[6]

The selection of the values of the regression coefficients was based on the consideration of two factors, shape (pattern of similarity and dissimilarity) and noncentrality (relative distance from the origin). A measure of noncentrality, δ, is defined as

$$\delta = \beta'\beta/\mathrm{tr}(X'X). \tag{29}$$

In the experiment we used two shapes of coefficients (all coefficients equal, and all coefficients but one equal to zero) and three values of the noncentrality measure δ (5, 20 and 35). The values of the regression coefficients for the 4-variable model were

$B11 = (2.2361, 2.2361, 2.2361, 2.2361)$
$B12 = (0, 0, 4.4721, 0)$
$B21 = (4.4721, 4.4721, 4.4721, 4.4721)$
$B22 = (0, 0, 8.9443, 0)$
$B31 = (5.9161, 5.9161, 5.9161, 5.9161)$
$B32 = (0, 0, 11.8332, 0)$.

These values correspond to the standardized values of the explanatory variables to avoid problems of units of measurement.[7]

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 was based on the distribution of square error loss of OLS. If accuracy is measured by the coefficient of variation of the average square error loss of the least squares estimator, then with 500 replications we achieve at worst 6.3% accuracy in the 4-variable model and 5.5% accuracy in the 8-variable model.

V. Evaluation of Results

In our experiment we considered three different degrees of multicollinearity and six different sets of values of the regression coefficients, giving rise to 18 different designs for each of the two models (the 4-variable and the 8-variable model). The following statistics based on the 500 replications have been computed for all the estimators for each of the three different loss structures:

(i) the average loss;
(ii) the standard deviation of loss;
(iii) the number of times that the ORR loss exceeded the OLS loss (in 500 samples);
(iv) the ratio of ORR average loss to OLS average loss.

The above statistics were recorded in 36 tables which are not presented here to save space but are available on request from the authors. Instead, we present a condensation of the results by showing—for the 4-variable model only—the average ratio of ORR loss to OLS loss

(i) for different degrees of multicollinearity over all shapes and all noncentralities of the coefficients (table 1);

[5] The details of the construction of the data sets and the values of the variables are available on request.

[6] For a discussion of this measure see Kmenta (1971, p. 390).

[7] The values of the explanatory variables and of the regression coefficients for the 8-variable model are similar to those for the 4-variable model.

(ii) for different shapes of the coefficients over all degrees of multicollinearity and all noncentralities (table 2);
(iii) for different noncentralities of the coefficients over all degrees of multicollinearity and all shapes of the coefficients (table 3).

We also present the average ratio of ORR loss to OLS loss for different numbers of explanatory variables over all degrees of multicollinearity and all shapes and all noncentralities of the coefficients (table 4).

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) the higher the number of explanatory variables.

(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 smaller when all the coefficients are equal (shape 1) than when all coefficients but one are equal to zero (shape 2).

(d) With a very few exceptions, the HKBM estimator is dominated by the HKB estimator.

To assess the relative performance of the five ORR estimators considered in our study, we obtained the sums of the simple ranks for each loss structure over the 36 regressions used in the experiment. The results are as follows:

TABLE 1.—EFFECT OF MULTICOLLINEARITY (4-VARIABLE MODEL)

Estimator	Loss Structure	Average ORR Loss/OLS Loss: Multicollinearity Low	Medium	High
HKB	$p' = 2$	0.90259	0.32480	0.22044
	$p' = \infty$	0.96306	0.57984	0.44109
	$p' = 1$	0.95599	0.57172	0.43715
HKBM	$p' = 2$	0.91613	0.45630	0.35699
	$p' = \infty$	0.96558	0.66449	0.54548
	$p' = 1$	0.95981	0.65696	0.54188
Dempster	$p' = 2$	0.91959	0.24270	0.10493
	$p' = \infty$	0.97354	0.51950	0.33057
	$p' = 1$	0.96684	0.51299	0.32726
Wermuth	$p' = 2$	0.98730	0.33727	0.11867
	$p' = \infty$	1.00091	0.64712	0.39094
	$p' = 1$	0.99865	0.61619	0.36757
Sclove	$p' = 2$	0.92989	0.23056	0.09349
	$p' = \infty$	0.97982	0.51335	0.32167
	$p' = 1$	0.97241	0.50697	0.31858

TABLE 2.—EFFECT OF THE SHAPE OF COEFFICIENTS (4-VARIABLE MODEL)

Estimator	Loss Structure	Average ORR Loss/OLS Loss: Shape 1	2
HKB	$p' = 2$	.49789	.46734
	$p' = \infty$	.67808	.64457
	$p' = 1$	.67405	.63586
HKBM	$p' = 2$	.58506	.56788
	$p' = \infty$	.73742	.71295
	$p' = 1$	.73181	.70728
Dempster	$p' = 2$	.44479	.40002
	$p' = \infty$	.63018	.58556
	$p' = 1$	.63536	.56936
Wermuth	$p' = 2$	.51281	.44927
	$p' = \infty$	.69222	.66708
	$p' = 1$	.72732	.59429
Sclove	$p' = 2$	.44371	.39285
	$p' = \infty$	.62830	.58160
	$p' = 1$	.63707	.56157

Note: Shape 1: All coefficients are equal.
Shape 2: All coefficients but one are zero.

TABLE 3.—EFFECT OF THE NONCENTRALITY OF COEFFICIENTS (4-VARIABLE MODEL)

Estimator	Loss Structure	Average ORR Loss/OLS Loss: Noncentrality Low	Medium	High
HKB	$p' = 2$	.39732	.50157	.54895
	$p' = \infty$	.58575	.67604	.72220
	$p' = 1$	.57275	.67192	.72019
HKBM	$p' = 2$	.50265	.59129	.63547
	$p' = \infty$	.66022	.73769	.77764
	$p' = 1$	.65038	.73342	.77484
Dempster	$p' = 2$	.33083	.43953	.49687
	$p' = \infty$	.51421	.62548	.68392
	$p' = 1$	.50286	.62186	.68237
Wermuth	$p' = 2$	.32702	.50291	.61322
	$p' = \infty$	.51994	.70731	.81171
	$p' = 1$	.49970	.68927	.79344
Sclove	$p' = 2$	.32951	.43463	.49070
	$p' = \infty$	.51305	.62139	.68041
	$p' = 1$	.50079	.61810	.67907

TABLE 4.—EFFECT OF THE NUMBER OF VARIABLES

Estimator	Loss Structure	Average ORR Loss/OLS Loss: Number of Variables 4	8
HKB	$p' = 2$	.48261	.40687
	$p' = \infty$	.66132	.61586
	$p' = 1$	.65455	.59579
HKBM	$p' = 2$	.57647	.43209
	$p' = \infty$	.72518	.63551
	$p' = 1$	.71954	.61709
Dempster	$p' = 2$	.42240	.37195
	$p' = \infty$	.60787	.58228
	$p' = 1$	.60236	.55753
Wermuth	$p' = 2$	.48108	.47739
	$p' = \infty$	.73957	.67965
	$p' = 1$	.66080	.66013
Sclove	$p' = 2$	.41828	.37284
	$p' = \infty$	.60495	.58394
	$p' = 1$	.59932	.55801

Mean Square Error Loss	
Sclove	74
Dempster	78
HKB	108
Wermuth	138
HKBM	142

Mean Absolute Error Loss	
Sclove	75
Dempster	84
HKB	109
Wermuth	132
HKBM	140

Maximum Absolute Error Loss	
Dempster	76
Sclove	77
HKB	101
HKBM	135
Wermuth	151

Although approximately the same results were obtained 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.

VI. Conclusion

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 II 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 and can compete with OLS on equal terms of the same prior information. The results of our Monte Carlo experiment indicate that, in general, the ORR estimators do out-perform 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 also performs 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. We hope that this paper might convince economic researchers to pay more attention to ridge regression than has so far been the case.

REFERENCES

Dempster, Arthur P., "Alternatives to Least Squares in Multiple Regression," in D. Kabe and R. P. Gupta (eds.), *Multivariate Statistical Inference* (Amsterdam: North-Holland Publishing Co., 1973), 25–40.

Dempster, Arthur P., Martin Schatzoff, and Nanny Wermuth, "A Simulation Study of Alternatives to Ordinary Least Squares," *Journal of the American Statistical Association* 72 (Mar. 1977), 77–104.

Hoerl, Arthur E., and Robert W. Kennard, "Ridge Regression: Biased Estimation for Non-Orthogonal Problems," *Technometrics* 12 (Feb. 1970), 55–67.

———, "Ridge Regression 1980: Advances, Algorithms, and Applications," *American Journal of Mathematical and Management Sciences* 1 (1) (1981), 5–83.

Hoerl, Arthur E., Robert W. Kennard, and Kent F. Baldwin, "Ridge Regression: Some Simulations," *Communications in Statistics* 4 (2) (1975), 105–123.

Kmenta, Jan, *Elements of Econometrics* (New York: Macmillan, 1971).

Lindley, Dennis V., and Adrian F. M. Smith, "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society, Series B,* 34 (1) (1972), 1–41.

Newhouse, Joseph P., and Samuel D. Oman, "An Evaluation of Ridge Estimators," Rand Corporation Report R-716 PR, Santa Monica, California (Apr. 1971).

Sclove, Stanley, "Least Squares with Random Regression Coefficient," Technical Report 87, Department of Economics, Stanford University (1973).

Smith, V. Kerry, "A Note on Ridge Regression," *Decision Sciences* 7 (July 1976), 562–566.

Theil, Henri, and Arthur S. Goldberger, "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review* 2 (Jan. 1961), 65–78.

Theobald, C. M., "Generalizations of Mean Square Error Applied to Ridge Regression," *Journal of the Royal Statistical Society, Series B,* 36 (1) (1974), 103–106.

Thisted, Ronald A., "Ridge Regression, Minimax Estimation, and Empirical Bayes Methods," Technical Report 87, Division of Biostatistics, Stanford University (1976).

Wermuth, Nanny, "An Empirical Comparison of Regression Methods," unpublished doctoral dissertation, Harvard University, 1972.