Florida Journal of
Educational Research
Fall, 1988, Vol. 30
No. 1, Pp. 15-33

Ridge Regression for Interactive Models

Richard L. Tate
Florida State University

ABSTRACT. An exploratory study of the value of ridge regression for interactive models is reported. Assuming that the linear terms in a simple interactive model are centered to eliminate nonessential multicollinearity, a variety of common models, representing both ordinal and disordinal interactions, are shown to have "orientations" which are favorable to ridge regression. Comparisons of the potential efficiency of ridge regression to that of ordinary least squares across a wide range of conditions clearly suggest the value of ridge procedures for many centered interactive models.

Interactive models have long been used to represent the complexity of human behavior in the social sciences because they allow the description of the effect of one independent variable on an outcome as a function of one or more other variables. A relatively common formulation for a simple interactive model with two independent variables, X_1 and X_2, is

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \qquad (1)$$

where E(Y) is the expected value of the outcome, Y; and the coefficient, β_3, of the product term represents the interaction effect. The effect of X_1 on Y (denoted X_1E) and the effect of X_2 on Y (X_2E) are defined as the partial derivatives of E(Y) with respect to X_1 and X_2 respectively; i.e.,

$$X_1E = \beta_1 + \beta_3 X_2$$

$$X_2E = \beta_2 + \beta_3 X_1 \qquad (2)$$

The "main effect" of each independent variable is defined as the effect of that variable when the other independent variable is equal to its own mean. It will be assumed here that X_1 and X_2 are standardized so that β_1 and β_2 represent the main effects of X_1 and X_2 respectively. Interactive models for multiple regression in nonexperimental inquiry are of primary interest here, although model (1) can also be used to represent an interactive 2X2 ANOVA or an aptitude-treatment-interaction analysis in experimental studies.

The question addressed here is whether it is possible to significantly improve the efficiency of estimation for this important class of models with the use of ridge regression techniques rather than the usual ordinary least squares (OLS) estimation. One possible response to such a question is "of course." We know that there is an apparent multicollinearity problem often found for interactive models, especially in nonexperimental studies because of quite large correlations associated with the product term. These large correlations can result in very large standard errors for the regression coefficients with the attendant lower statistical power and precision. We also know that ridge regression was originally presented by Hoerl and Kennard (1970a,b) as a possible solution to the problem of multicollinearity. Thus, interactive models and ridge regression might appear to be a logical match, and, in fact, a relatively early illustration of ridge regression was based on a model with expansion terms (powers and products) of the independent variables (Marquardt & Snee, 1975). Some of the studies and critiques of ridge regression in the years since its introduction include Bingham amd Larentz (1977); Darlington (1978); Dempster, Schatzoff and Wermuth (1977); Draper and Van Nostrand (1979); Gibbons (1981); Gunst and Mason (1977); Hoerl, Kennard and Baldwin (1975); McDonald and Galarneau (1975); Pagel and Lunneborg (1985); Smith and Campbell (1980); Vinod and Ullah (1981); and Wichern and Churchill (1978).

The answer to the question of the value of ridge regression for interactive models is not so apparent, however, when one recognizes that the problem of multicollinearity associated with many interactive models usually disappears when recommendations about centering of terms are followed. Marquardt and Snee (1975) suggested a linear transformation of the original model in which the linear model components are first standardized, followed by the formation of any expansion terms from the standardized components (see however Smith & Campbell, 1980). This transformation accomplishes two purposes. First, it establishes a convenient scale for the interpretation of study results. Second, "nonessential" collinearity is removed. The formation of expansion terms with centered linear terms often reduces the correlations associated with the expansion terms to near zero (see Smith & Sasaki, 1979, for a more detailed discussion of this). Thus, if we assume that the correlation between the linear components is less than roughly 0.8 (a reasonable upper limit if X_1 and X_2 are supposed to be conceptually different, considering typical reliabilities in the social sciences), the removal of the collinearity associated with the product term produces a model which has no significant problem of multicollinearity.

Why consider the possible use of ridge regression for model (1) if centering eliminates any problem of multicollinearity? Many of the studies cited above have demonstrated that the orientation of the model coefficient vector is also critical in determining the possible gain associated with the use of ridge regression (e.g., Bingham & Larentz, 1977; Darlington, 1978; Gibbons, 1981; McDonald & Galarneau, 1975; Pagel & Lunneborg, 1985; and Wichern & Churchill, 1978). Briefly, the relative superiority of ridge regression over OLS tends to be greater for a "favorable" orientation in which the coefficient vector is nearly orthogonal to the minor principal axis of the data. This current view of the critical importance of orientation has led to conclusions such as, "Routine use of ridge regression without prior knowledge of the predictor space is not recommended" (Pagel & Lunneborg, 1985). Thus, despite the absence of any significant multicollinearity problem for a centered interactive model, ridge regression may still offer an attractive alternative to OLS if common interactive models tend to have favorable orientations.

One goal of this article, then, is to characterize the orientation of centered interactive models. The following results indicate that many interactive models do, in fact, have favorable orientations for ridge regression. The remainder of the article describes the efficiency of ridge regression relative to OLS for a systematic variation of degree of collinearity, sample size, strength of effect, and coefficient orientation. Both "ordinary" and "generalized" versions of ridge regression are considered.

The concern here is with the maximum potential of ridge regression for centered interactive models. All calculations have been conducted for population values, using the biasing parameters which are known to be "MSE optional" for generalized ridge regression; i.e., the results reflect the minimum MSE for ridge regression possible under the specified conditions. Of course, in practice it is necessary to estimate the unknown population biasing parameters and, in general, only some fraction of the potential gain will be realized. Thus, if the potential gain described here is impressive enough, follow-up computer simulations will be required to determine what part of the potential can be achieved in practice.

Methods

The model and estimation procedures will first be briefly summarized. The canonical formulation used here closely follows that described in more detail in Vinod and Ullah (1981). The model considered is that in (1), assuming Y, X_1, and X_2 are standardized. The product term is not, in general, a standard score. For purposes of computation, Marquardt and Snee (1975) recommend a final normalization, resulting in the model, now expressed in matrix form,

$$E(\underset{\sim}{Y}) = \underset{\sim}{X}\underset{\sim}{\beta} \qquad (3)$$

in which $\underset{\sim}{X}'\underset{\sim}{X}$ is equal to the correlation matrix for the independent variables $\underset{\sim}{R}_{xx}$. Two aspects of the coefficient vector, $\underset{\sim}{\beta}$, are of interest: the orientation of the vector, represented with the vector of direction cosines of β ($\underset{\sim}{v}$), and the length $|\underset{\sim}{\beta}|$ of the vector which is an indicator of the magnitudes of the effects in the model (thus, $\underset{\sim}{\beta} = |\underset{\sim}{\beta}|\underset{\sim}{v}$).

As discussed in the introduction, because of the centering of the linear components of the model, the population correlation matrix is assumed to be

$$\underset{\sim}{R}_{xx} = \begin{bmatrix} 1 & & \\ r_{12} & 1 & \\ 0 & 0 & 1 \end{bmatrix} .$$

where r_{12} is the correlation between x_1 and x_2. (It is this simple form of the assumed population correlation matrix which allows the relatively simple formulation and conclusions given below.) The specification of $\underset{\sim}{\beta}$ and r_{12} implies other parameters. The vector of correlations between Y and the X's is $\underset{\sim}{R}_{xy} = \underset{\sim}{R}_{xx}\underset{\sim}{\beta}$, the coefficient of determination is $R^2 = \underset{\sim}{\beta}'\underset{\sim}{R}_{xy}$ and the residual variance is $\sigma^2 = 1 - R^2$. The specification of $\underset{\sim}{\beta}$ is of course constrained by the maximum R^2 value of one. In addition, it was assumed that each resulting effect represented in (2) should be less than roughly one for any value of the other variable.

<u>Canonical form</u>. The eigenvalues of $\underset{\sim}{R}_{xx}$ are $\lambda_1 = 1+r_{12}$, $\lambda_2 = 1$, and $\lambda_3 = 1-r_{12}$. Assuming an upper limit for r_{12} of 0.8 for this study, the maximum value of the ratio λ_1/λ_3, a common indicator of the presence of multicollinearity, is only nine, indicating the absence of any serious problem. The eigenvectors associated with these eigenvalues are given in the columns of

$$G = \begin{bmatrix} .707 & 0 & .707 \\ .707 & 0 & -.707 \\ 0 & 1 & 0 \end{bmatrix}$$

Based on these vectors the canonical form of the model is

$$E(\underset{\sim}{Y}) = \underset{\sim}{W}\underset{\sim}{\gamma} \tag{4}$$

where $\underset{\sim}{W} = \underset{\sim}{X}\underset{\sim}{G}$ is a matrix of principal component scores and

$$\begin{aligned} \underset{\sim}{\gamma} &= \underset{\sim}{G}'\underset{\sim}{\beta} \\ &= |\underset{\sim}{\beta}|\underset{\sim}{G}'\underset{\sim}{v} \end{aligned}$$

is the coefficient vector of the canonical model. The eigenvalues indicate the variances of the transformed variables (i.e., the principal components). Since the principal components are uncorrelated, the correlation between the ith transformed variable and Y is equal to

$\lambda_i^{1/2}\gamma_i$, where γ_i is the ith element of $\underset{\sim}{\gamma}$.

OLS estimation. The covariance matrix for the OLS estimates of $\underset{\sim}{b} = (\underset{\sim}{X}'\underset{\sim}{X})^{-1}(\underset{\sim}{X}'\underset{\sim}{Y})$ can be expressed in terms of the eigenvalues and vectors of $\underset{\sim}{R}_{xx}$ as

$$\underset{\sim}{V}(b) = [\sigma^2/(n-1)]\ \underset{\sim}{G}\underset{\sim}{Dia}(1/\lambda_i)\underset{\sim}{G}', \tag{5}$$

where $\underset{\sim}{Dia}(1/\lambda_i)$ is a diagonal matrix with $1/\lambda_i$ in the ith diagonal position. The OLS global MSE(b) is the trace of $\underset{\sim}{V}(b)$, which is

$$MSE(b) = (\sigma^2/[n-1])\ \Sigma\ (1/\lambda_i). \tag{6}$$

The MSE for the interaction effect is $MSE(b_3)$ while the MSE's for the effects in (2) are obtained by defining each effect as a linear combination of the β's, $C = \underset{\sim}{c}'\underset{\sim}{\beta}$, with the associated MSE then being $\underset{\sim}{c}'\underset{\sim}{V}(b)\underset{\sim}{c}$.

Ridge estimation. The estimator for generalized ridge regression (Hoerl & Kennard, 1970a) is

$$\underset{\sim}{b}_k = (\underset{\sim}{X}'\underset{\sim}{X} + \underset{\sim}{G}\underset{\sim}{K}\underset{\sim}{G}')^{-1}(\underset{\sim}{X}'\underset{\sim}{Y}), \tag{7}$$

where the diagonal matrix $\underset{\sim}{K}$ contains the biasing parameters, k_i (i=1,2,3). The MSE optimal values of the biasing parameters are

$$k_i = \frac{\sigma^2}{(n-1)\gamma_i^2} \tag{8}$$

For the purposes here, it is useful to consider the canonical form of the model. The ridge estimate (c_{ik}) of a canonical parameter (γ_i) can be obtained by "shrinking" the corresponding OLS estimate of the same parameter (c_i) towards zero; that is, $c_{ik} = \delta_i c_i$ where the shrinkage parameter δ_i is obtained with

$$\delta_i = \frac{\lambda_i}{(k_i + \lambda_i)} \tag{9}$$

The bias of the ridge estimation of β is

$$\underset{\sim}{\text{Bias}}(b_k) = \underset{\sim}{G}\underset{\sim}{\text{Dia}}(\delta_i - 1)\underset{\sim}{G}'\underset{\sim}{\beta} \tag{10}$$

and the covariance matrix is

$$\underset{\sim}{V}(b_k)=[\sigma^2/(n-1)]\underset{\sim}{G}\underset{\sim}{\text{Dia}}(\delta_i^2/\lambda_i)\underset{\sim}{G}'. \tag{11}$$

The MSE matrix is then determined with

$$\underset{\sim}{\text{MtxMSE}}(b_k) = \underset{\sim}{V}(b_k) + \underset{\sim}{\text{Bias}}(b_k)\ \underset{\sim}{\text{Bias}}(b_k)'. \tag{12}$$

The global $MSE(b_k)$ is the trace of this matrix, which is equal to

$$MSE(b_k) = (\sigma^2/[n-1])\Sigma\ \ (\delta_i^2/\lambda_i) + \Sigma\ \ (\delta_i-1)^2\gamma_i^2. \tag{13}$$

The first term of this expression is the variance component while the second is that for the bias. For any linear combination, $C = \underset{\sim}{c}'\underset{\sim}{\beta}$, the bias is $\text{Bias}(C) = \underset{\sim}{c}'\underset{\sim}{\text{Bias}}(b_k)$, the variance is $V(C) = \underset{\sim}{c}'\underset{\sim}{V}(b_k)\underset{\sim}{c}$, and the MSE is $MSE(c) = \underset{\sim}{c}'\underset{\sim}{MSE}(b_k)\underset{\sim}{c}$.

The inefficiency ratio (IR) for ridge regression is defined, following Pagel & Lunneborg (1985), as the ratio of the MSE for ridge regression to that for OLS ($IR = MSE(b_k)/MSE(b)$). Thus, IR values less than one indicate relatively superior efficiency for GRR. The global IR can be expressed in terms of the IR's for the individual canonical components as

$$IR = \frac{\Sigma\ (1/\lambda_i)IR_i}{\Sigma\ (1/\lambda_i)} \tag{14}$$

When MSE optimal parameters are used, the individual IR_i is equal to the ith shrinkage parameter, δ_i, so that

$$IR = \frac{\Sigma\ (\delta_i/\lambda_i)}{\Sigma\ (1/\lambda_i)} \qquad (15)$$

Generalized ridge regression (GRR) allows for different k_i values. Ordinary ridge regression (ORR), on the other hand, sets all k's to a single constant value (Hoerl & Kennard, 1970a). There is no analytical expression for the MSE optimal k value for ordinary ridge regression; for the purpose of this study, the single k has been set equal to the harmonic mean of the optimal k_i's, following Hoerl, Kennard, and Baldwin (1975).

The determinants of IR which were varied systematically in this study were: orientation of $\tilde{\beta}$ (represented with the direction cosine vector $\tilde{v}$), the strength of the model effects (represented with the length of $\tilde{\beta}$, $|\tilde{\beta}|$), the correlation between the two independent variables (r_{12}), and the sample size (n). The general direction of the effect of each of these determinants on IR is easy to see given previous equations. Decreasing n, decreasing $|\beta|$, and increasing r_{12} result in decreasing IR (i.e., result in increasing relative efficiency of ridge regression). (The effect of orientation is discussed in a later section.) These directions are consistent with those found in various simulation studies reported in the literature. The relationship between IR and the four determinants is, however, highly nonlinear and interactive. Thus, even though we know the general direction of the effect of each factor, the magnitude of that effect varies strongly with the levels of the other factors. One of the goals here is to describe the nature of this interaction.

Before reporting the general results of the study, ridge regression and OLS will first be compared in some detail for a specific set of conditions.

Example

To illustrate the potential gain associated with the use of generalized ridge regression (GRR), consider an interactive population model with coefficients $\tilde{\beta}' = [.4\ .3\ .2]$. The length of this $\tilde{\beta}$ is 0.539 and the orientation is $\tilde{v}' = (.743\ .557\ .371)$. The correlation between X_1 and X_2 (r_{12}) is assumed to be 0.4. As indicated in the preceding section, it is assumed that Y, X_1 and X_2 are standardized and that, because of the centering of X_1 and X_2, the population correlations between these two variables and the interaction product term are zero. The resulting vector of correlations between the independent variables and the dependent variable ($\tilde{R}_{xy}$) is (.520 .460 .214) and R^2 is 0.392. The population effect of X_1 ($X_1E = .4 + .2X_2$) is represented with the straight line in the middle of the band in Figure 1. The effect is positive across the range of X_2, with a relatively strong value of 0.8 at $X_2 = 2$. The large variation of the magnitude of the effect across the range of X_2 reflects a strong interaction. (The description of the effect of X_2 is similar in nature and not given here.)

The eigenanalysis of $\underset{\sim}{R}_{xx}$ results in eigenvalues of (1.4 1.0 0.6) and the associated vectors given in the preceding section. The parameters in the canonical form of the model are $\underset{\sim}{\gamma}' = [.495\ .214\ .0707]$ and the correlations between the dependent variable and the transformed independent variables are [.693 .214 .042]. The relatively small value of γ_3 suggests a favorable orientation, as discussed in the next section.

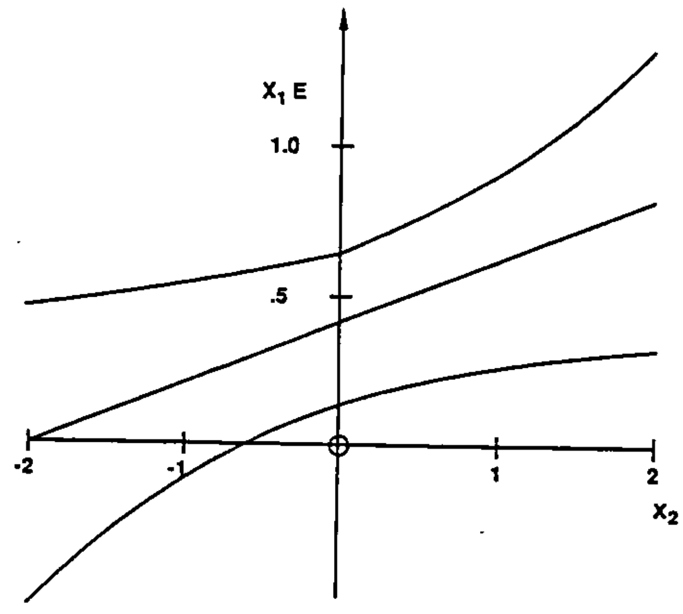

Figure 1 True Value and Confidence Band for the Effect of X_1

We now compare the OLS and GRR sampling characteristics associated with this model for a sample size of 50. The coefficient covariance matrix for the OLS estimates is

$$\underset{\sim}{V}(b) = \begin{bmatrix} .0148 & & \\ -.0059 & .0148 & \\ .0 & .0 & .0108 \end{bmatrix}$$

with, summing over the diagonal, a global MSE(b) of 0.0404. Computation of standard errors for the effect of X_1 at different levels of X_2 results in the continuous confidence band shown in Figure 1.

The GRR solution for the same model results in the MSE optimal biasing parameters (k_i) of (.0507 .271 2.482) and associated shrinkage parameters (δ_1) of (.965 .787 .195). The expected value vector for the GRR estimates with these biasing parameters is $\underset{\sim}{E}(b_k)$ = (.348 .328 .157) and the MSE matrix is

$$Mt\underset{\sim}{x}MSE(b_k) = \begin{bmatrix} .0073 & & \\ .0023 & .0053 & \\ .0022 & -.0012 & .0085 \end{bmatrix}$$

with an associated global MSE(b_k) = 0.0211

The inefficiency ratio, IR is thus MSE(b_k)/MSE(b) = 0.523. There is a relative lack of sensitivity of this ratio to modest departures of the k_i from their MSE optimal values. For example, when the assumed k_i are one half of the optimal values, the resulting relative efficiency is 0.547, and assumed k_i which are twice their optimal values result in a ratio of 0.556.

The relative efficiencies for specific effects vary in both directions from the global value. The IR for the interaction effect β_3 is 0.787, and the relative efficiency associated with X_1E is shown in panel a of Figure 2 to vary from approximately 0.40 to 0.87, with the smallest ratios found for the "central" effects (i.e. X_1E's for small values of X_2). The difference between the expected values of $X_1\hat{}E$ and the corresponding true values shown in panel b of Figure 2 indicate the estimation bias associated with the MSE reduction for this example.

To summarize, for this example in which there was a relatively strong relationship and a sample size of 50, there was a significant reduction in global MSE resulting from the use of GRR rather than OLS. There was variation in the gain in efficiency when considering individual effects. We now consider the results from systematic variation of some of the critical parameters to examine how these results hold up in other circumstances.

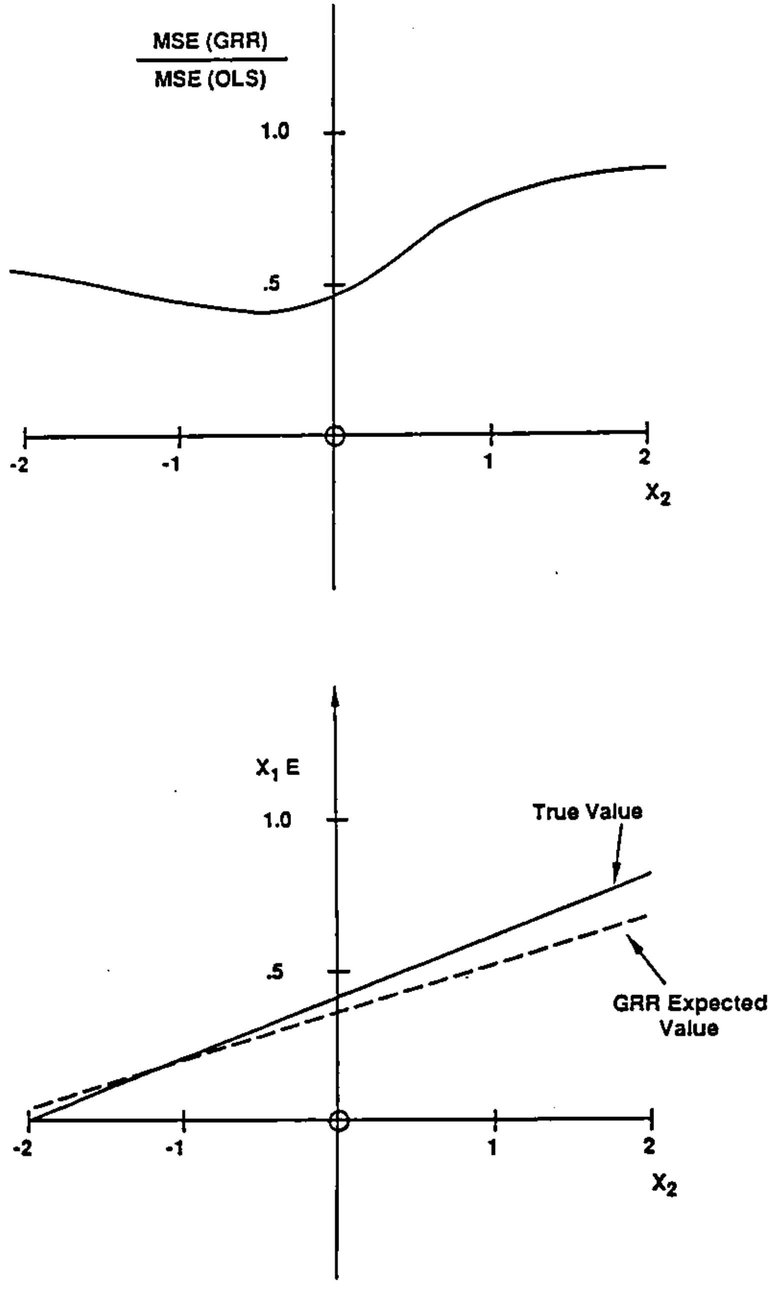

Figure 2. Comparison of Ridge and OLS for the Effect of X_1

Results

Orientation

The orientation of β for a given interactive model, represented by $v' = [v_1 v_2 v_3]$, can be shown as a point on the v_1 versus v_2 plot of Figure 3. Each v_1, v_2 combination implies the v_3 value; constant v_3 contours would be represented by circles centered at the origin, with v_3 equal to 1 at the origin and equal to zero at the outer circle. The IR results in the third and fourth quadrants are identical to those in the first and second quadrants, respectively. The limiting additive case is represented by points on the outer circle where $v_3 = \beta_3 - 0$, while the origin ($v_3=1$, $v_1=v_2=0$) represents a "pure" interactive model in which the main effects are both equal to zero. The shaded area near the outer circle indicates all possible "ordinal" interactions in which there is no reversal of the direction of either effect across the approximate range of the other variable. (This region is comprised of all points for which $\beta_1 \geq 2\beta_3$ and $\beta_2 \geq 2\beta_3$.) All points in the unshaded area represent "disordinal" interactive models in which there is a reversal of direction of one or both of the effects somewhere within the range of the other variable.

To aid in the assessment of the orientation of a β relative to the principal axes of the data, the first and third principal axes (labeled γ_1 and γ_3) have been superimposed on the v_1 versus v_2 coordinate system of Figure 3. The second principal axis, γ_2, is identical to the v_3 axis. The β vectors which are orthogonal to the third axis are found along the γ_1 axis, those orthogonal to the second axis are along the outer circle, and those orthogonal to the first axis are along the γ_3 axis. Points 1, 2, and 3 in the figure represent the β's which are orthogonal to two of the three principal axes; i.e., these points represent vectors which are aligned with principal axes 1, 2, and 3, respectively.

As noted in the introduction, a favorable orientation for ridge regression is identified in the literature as one in which β is orthogonal to the last principal axis. Such characterization is easily understood by inspection of Equation 15. Note that the $1/\lambda_i$ weighting factors in the equation determine the relative importance of the δ's in determining IR. When r_{12} is relatively large, the δ_3 has the largest weight (recall $\lambda_1 = 1 + r_{12}$, $\lambda_2 = 1$, and $\lambda_3 = 1 - r_{12}$), and therefore is the most important. Thus, favorable orientations are those in which β is nearly orthogonal to the third principal axis (i.e., when γ_3 and δ_3 are nearly equal to zero). These orientations are found along the γ_1 axis in Figure 3. The most favorable of all these orientations would be the two where β is orthogonal to two principal axes, including the third (i.e., when β is aligned with the first or the second principal axis). In contrast, alignment of $\underset{\sim}{\beta}$ with the third axis (resulting in $\gamma_3 = 1$ and maximization of δ_3) produces an unfavorable orientation with large IR. This latter situation corresponds to points along the γ_3 axis in Figure 3. To illustrate, when $r_{12} = 0.8$, n = 50, and $|\underset{\sim}{\beta}| = 0.539$, the IR values for $\underset{\sim}{\beta}$'s aligned with the first, second, and third principal axes are 0.08, 0.15 and 0.61, respectively. Thus, for these conditions the effect of orientation is quite strong, with the ratio of the IR for the most favorable orientation to that for the least favorable being 0.08/0.61 = 0.13.

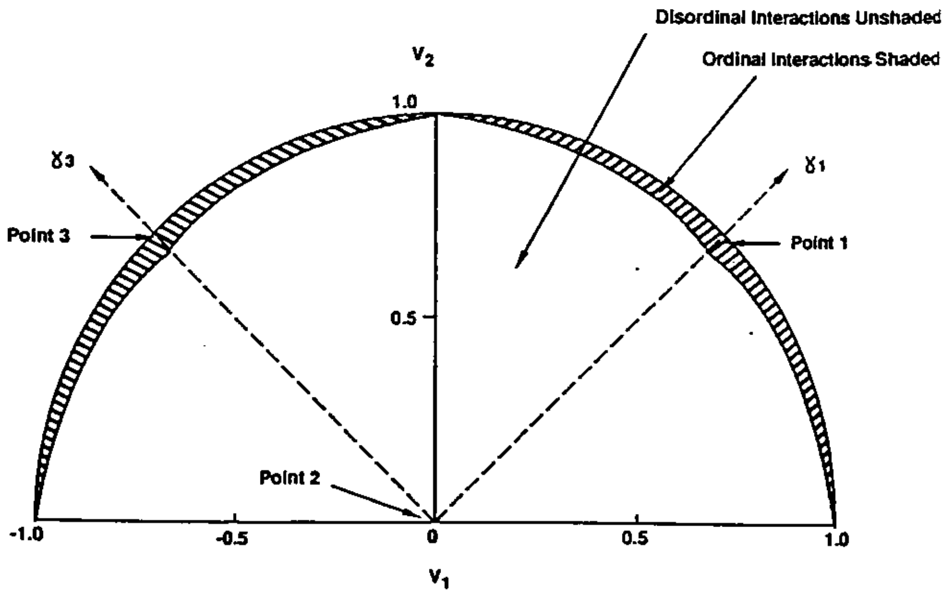

Figure 3 Coefficient Vector Orientation

As r_{12} decreases and approaches zero, additional orientations become relatively favorable. In the limit, when $r_{12} = 0$, all δ's are equal and are weighted equally. The most favorable orientations would then be when β is orthogonal to any two of the axes, i.e., when it is aligned with any of the axes. To illustrate, the IR value is equal to 0.32 for β's aligned with all three of the axes when $r_{12} = 0$, n = 50, and $|\beta| = 0.539$.

The above IR values plus results for other orientations are shown in Figure 4. For $r_{12} = 0.8$, the IR values shown range from 0.08 to 0.72, with the lowest values found along the γ_1 axis. Referring back to the different types of interactive models, it is seen that the most favorable orientations are associated with models in which the two main effects are of the same sign and roughly equal in magnitude. The relative size of the interaction effect in this category can range from zero in the limiting additive model to that for a purely interactive model, corresponding to different positions along the γ_1 axis. In contrast, unfavorable orientations correspond to interactive models in which (a) one main effect is very large compared to the other or (b) the main effects have opposite signs. When r_{12} decreases to zero (panel b of the figure), the models with the smallest IR of 0.32 and favorable orientations are the pure interactive model and the limiting additive model with main effects of equal or opposite signs. The IR values increase rapidly from this minimum value with departure from these orientations reaching 0.86.

Effects of Other IR Determinants

Inefficiency ratios for a systematic variation of the four determinants (orientation, r_{12}, $|\beta|$, and n) are shown in Table 1. Two values of each parameter are considered. For coefficient orientation, a direction cosine vector of $v' = (.70\ .70\ .141)$ was defined as "favorable" and one of $v' = (.70\ 0\ .70)$ was defined as "unfavorable" (see Figure 4). The limiting values on the other parameters were 0 and 0.8 for r_{12}, 0.1 and 0.6 for $|\beta|$, and 10 and 100 for n. In addition to IR, the overall strength of relationship (R^2) and MSE for OLS estimation are also given for each combination of study parameters. As expected, holding other factors constant, the MSE for OLS increases with decreasing $|\beta|$, decreasing n, and increasing r_{12}.

Consider first the inefficiency ratio for generalized ridge regression. The IR values are relatively small (equal to or less than, say, 0.5) for a very wide range of conditions. For favorable orientations, for example, IR is equal to or less than about 0.6 for all conditions. For the most unfavorable orientation, the IR is significantly larger than 0.6 only for the combination of $|\beta| = 0.6$ and n = 100. The general directions of the

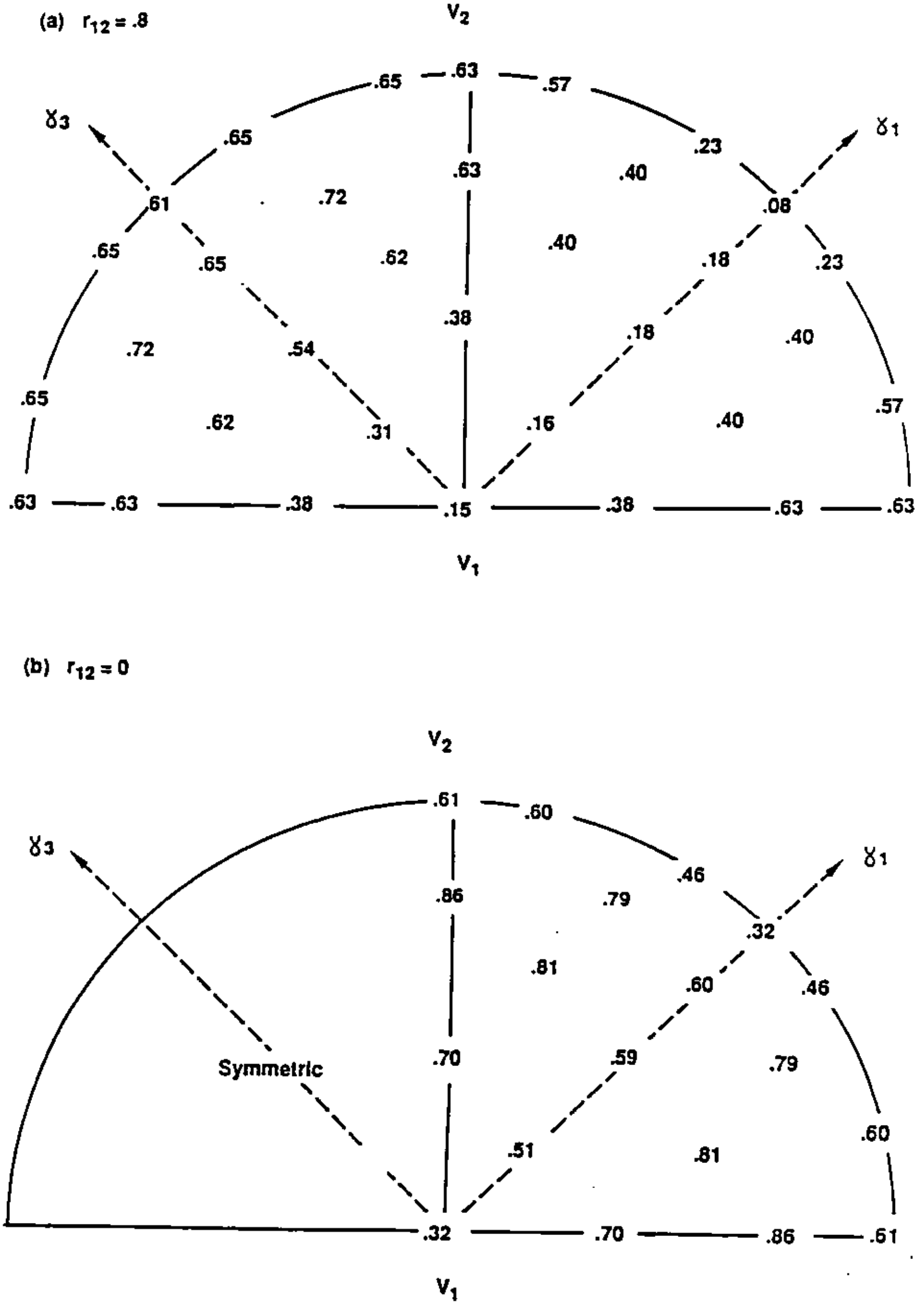

Figure 4. Effect of Orientation on Inefficiency Ratio (IR)

effects of each parameter are consistent with the expectations outlined in the methods section. Holding other factors constant, IR decreases with decreasing $|\beta|$, decreasing n, increasing r_{12}, and improving orientation. Thus, for a given orientation, the potential gain from the use of GRR rather than OLS increases for exactly those conditions that produce the larger MSE values for OLS.

The magnitudes of the effects of each parameter on IR vary widely, ranging from near zero to quite large values, reflecting a highly interactive relationship. For the orientation effect, the IR differences between the favorable and unfavorable orientations range from zero to 0.65, with the strongest effects found for the combination of large $|\beta|$ and large n. In contrast, the IR differences were all less than 0.1 for the small $|\beta|$ condition. Thus, with respect to the $|\beta|$ and n parameters, the orientation effect is weakest for those conditions where the potential gain associated with GRR is the greatest.

The IR differences for the $|\beta|$ effect range from 0.1 to 0.7, with the combination of unfavorable orientation and large n resulting in the largest differences. For the r_{12} effect, the IR differences vary from 0.02 to 0.34, with very weak effects (less than 0.2) for most of the conditions. (In considering this characterization of the strength of the r_{12} effect on IR, remember that the largest value of r_{12} considered here is 0.8, a value much smaller than those typically used in studying the multicollinearity problem.) Finally, the effect of n ranges from 0.05 to 0.47. All of these differences due to n were equal to or less than approximately 0.2 except for those found for the combination of unfavorable orientation and large $|\beta|$.

The comparison between the generalized ridge regression and ordinary ridge regression results in Table 1 indicates that the improvement potentially provided by the more complex approach varies widely with study and model conditions. Relatively large gains due to the use of GRR are found primarily for the combination of favorable orientation and large $|\beta|$, with relatively small gains for all other conditions.

Summary

A variety of centered interactive models have orientations which are relatively favorable to ridge regression. The models which always exhibited favorable orientation across the range of correlation consisted of those with both ordinal and disordinal interactions of widely varying strength in which the main effects have the same sign and roughly equal magnitudes. These are the models which have orientations which are approximately orthogonal to the last principal axis of the data.

Table 1 Inefficiency Ratio (IR) for Generalized and Ordinary Ridge Regression

Favorable orientation[a]

$\|\beta\|$	r_{12}	n	OLS R^2	OLS MSE	GRR IR	ORR IR
.1	0	10	.010	.330	.028	.029
.1	0	100	.010	.030	.172	.250
.1	.8	10	.018	.671	.013	.014
.1	.8	100	.018	.061	.061	.086
.6	0	10	.360	.213	.308	.628
.6	0	100	.360	.019	.503	.949
.6	.8	10	.647	.241	.108	.275
.6	.8	100	.647	.022	.164	.802

Unfavorable orientation[a]

$\|\beta\|$	r_{12}	n	OLS R^2	OLS MSE	GRR IR	ORR IR
.1	0	10	.010	.330	.029	.029
.1	0	100	.010	.030	.244	.250
.1	.8	10	.013	.674	.014	.014
.1	.8	100	.013	.061	.111	.119
.6	0	10	.360	.213	.611	.628
.6	0	100	.360	.019	.944	.949
.6	.8	10	.484	.352	.342	.367
.6	.8	100	.484	.032	.813	.841

[a]For favorable orientation v'= [.70 .70 .141] and for unfavorable v'= [.70 0 .70]

As the correlation r_{12} decreases to zero, additional orientations also become relatively favorable. These results suggest the potential value of ridge regression for such favorably oriented interactive models, despite the absence of any significant multicollinearity problem due to the assumed centering.

The theoretical efficiency of ridge regression relative to that of OLS was determined for a wide range of conditions based on systematic variation of the orientation of the model coefficient vector, the strength of the effects represented by the length of the vector ($|\underset{\sim}{\beta}|$), the correlation between the two independent variables, and the sample size. The general directions of the effects of these determinants of the relative efficiency were consistent with expectations from the literature. That is, the efficiency of ridge regression relative to OLS increased with more favorable orientations, increasing correlation between the two independent variables, decreasing strength of effect, and decreasing sample size. Moreover, as expected, the generalized version of ridge regression was always potentially superior to ordinary ridge regression, although the degree of superiority was unimpressive for many conditions.

Such general statements about the direction of various effects are, however, very misleading given the degree of interaction among the determinants of the relative efficiency. For the wide range of conditions considered, the magnitude of the effect of each determinant (defined as the change in relative efficiency, holding other determinants constant) varied from near zero up to large values of approximately 0.7. For example, the effect of orientation was strongest for those conditions usually thought of as being least favorable for ridge regression, i.e., for large $|\underset{\sim}{\beta}|$ and large sample size. In contrast, the effect of orientation was relatively weak for small $|\underset{\sim}{\beta}|$ and small n. The strength of these various interactions indicates that any precise description of how each determinant affects relative efficiency would have to be quite complex.

Fortunately, when one shifts from an interest in the effects of the various determinants to the final question of relative efficiency across a wide range of conditions, the picture is much simpler. For a favorable orientation, the theoretical relative efficiency was less than 0.6 for all conditions considered. Moreover, even for the unfavorable orientation it was significantly larger than 0.6 for only two of eight conditions. Again the conclusion is that ridge regression appears to have potential value for interactive models.

It is important to remember that these results show the maximum possible gain which could be achieved by using ridge regression rather than OLS. Of course there is ample evidence in the literature that the performance of ridge regression in practice, when the optimal shrinkage parameters must be estimated, is often inferior to that which is theoretically possible. Therefore, follow-up computer simulation study is required to determine how much of any apparent advantage of ridge regression found in this initial study can be realized in practice.

References

Bingham, C. & Lametz, K. (1977). Comment. Journal of the American Statistical Association, 72, 97-102.

Darlington, R. B. (1978). Reduced-variance regression. Psychological Bulletin, 85, 1238-1255.

Dempster, A. P., Schatzoff, M., & Wermuth, N. (1977). A simulation study of alternatives to Ordinary Least Squares. Journal of the American Statistical Association, 72, 77-91.

Draper, N. R., & Van Nostrand, R. C. (1979). Ridge regression and James-Stein estimation: Review and comments. Technometrics, 21, 451-456.

Gibbons, D. (1981). A simulation study of some ridge estimators. Journal of the American Statistical Association, 76, 131-139.

Gunst, R. F., & Mason, R. L. (1977). Biased estimation in regression: An evaluation using mean square error. Journal of the American Statistical Association, 72, 616-628.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 55-67.

Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Applications to nonorthogonal problems. Technometrics, 12, 69-82.

Hoerl, A. E., Kennard, R. W., & Baldwin, K. F. (1975). Ridge regression: Some simulations. Communications in Statistics, 5, 105-123.

Marquardt, D. W., & Snee, R. D. (1975). Ridge regression in practice. The American Statistician, 29, 3-20.

McDonald, G. C., & Galarneau, D. I. (1975). A monte carlo evaluation of some ridge-type estimators. Journal of the American Statistical Association, 70, 407-416.

Pagel, M. D., & Lunneborg, C. E. (1985). Empirical evaluation of ridge regression. Psychological Bulletin, 97, 342-355.

Smith, A., & Campbell, F. (1980). A critique of some ridge regression methods. Journal of the American Statistical Association, 75, 74-81.

Vinod, H. D., & Ullah, A. (1981). Recent advances in regression methods. New York: Marcel Dekker.

Wichern, D. W., & Churchill, G. A. (1978). A comparison of ridge estimators. Technometrics, 20, 301-311.

AUTHOR

RICHARD L. TATE, Associate Professor, Department of Educational Research, 307 Stone Bldg., Florida State University, Tallahassee, FL 32306

Acknowledgment: The author gratefully acknowledges the programming assistance of Sea Im and comments on an early version of the paper by Dwayne Meeter.