

When the multicollinearity among the independent variables in a regression model is due to the high correlations of a multiplicative function with its constituent variables, the multicollinearity can be greatly reduced by centering these variables around minimizing constants before forming the multiplicative function. The values of these constants that minimize the multicollinearity are derived, and the conditions are identified under which centering the variables about their means will reduce the multicollinearity. Among the advantages of this procedure are that the mean square error remains at its minimum, that the coefficients for other variables in the model are unaffected by it, and that the OLS estimates for the original model can be calculated from those for the modified model. Thus, even when estimates of the original model are desired, the procedure can be used to reduce numerical error.

DECREASING MULTICOLLINEARITY

A Method for Models with Multiplicative Functions

KENT W. SMITH

Northwestern University

M. S. SASAKI

University of Michigan

the problems created by multicollinearity among the independent variables in least-squares regression are by now well-known, and there is no need to go into a lengthy litany of them. Briefly, as multicollinearity increases, the least-squares estimates of the regression coefficients remain unbiased, but the determinants of the independent variables' covariance and correlation matrices approach zero, and the standard errors of the coefficients increase. Also, the expected distance between the vector of least-squares coefficients and the vector of true regression coefficients increase with some estimates frequently

AUTHORS' NOTE: We thank Chu Sato and Gary A. Simon for their advice and assistance. We also thank John W. Tukey and Donald E. Rubin for their helpful comments on an earlier version, and Robert W. Jackman for his advice on our numerical example. Naturally, we are solely responsible for any errors.

having either unreasonably large values or unreasonable signs. Moreover, slight sampling fluctuations in the estimates of the zero-order covariances can cause large swings in the values and signs of least-squares estimates of the coefficients in the presence of multicollinearity—a phenomenon someone once called the problem of the “bouncing betas.” Apart from these statistical problems of estimation, multicollinearity also creates a numerical one: as the determinant of the covariance matrix decreases, the rounding error in computing the inverse of the matrix, which is needed for the least-squares estimates, increases. (For further discussion of these problems, see the articles on ridge regression cited below and, for example, Blalock, 1963; Gordon, 1968; Althausen, 1971; and Rockwell, 1975.)

Perhaps the most frequently cited symptom of multicollinearity, large standard errors, is actually a function of both multicollinearity, in the sense of high intercorrelations among the independent variables, and large residual variance. We are concerned here primarily with the former. The size of the standard errors can often be reduced by using larger or more efficient samples and by increasing the numerical accuracy in one's data; however, reasonable applications of these tactics frequently do not overcome the problems posed by the multicollinearity inherent in the models with multiplicative terms which we shall be discussing.

Hoerl (1962) and Hoerl and Kennard (1970a, 1970b) have proposed *ridge regression* as one way of overcoming the problems of multicollinearity. (For practically oriented didactic articles on ridge regression, see Marquardt and Snee, 1975; and Deegan, 1975. See also Henry, 1976.) The improvements suggested recently by Guilkey and Murphy (1975) and Kasarda and Shih (1977) make ridge regression even more attractive; however, it is not necessarily the final answer to all problems of multicollinearity. A few limitations and disadvantages, in addition to those discussed by Deegan (1975) and Henry (1975), are worth mentioning. First, ridge regression may require considerable computational time if there are a large number of independent variables. Even with Kasarda and Shih's iterative procedure (which eliminates the necessity of calculating a set of complete

ridge-trace plots), the correlation matrix, with its diagonal elements augmented by the constant, must be inverted separately for several different values of the constant. Second, ridge regression does not provide a minimum mean-square error for the model. While the proportion of variance explained may not be the primary goal of regression analysis, one may not want to give up the minimum mean-square error provided by least-squares regression for the sake of more efficient estimates of the partial regression coefficients. Finally, there is a characteristic that is admittedly only occasionally a problem. The vector of least-squares regression coefficients is a linear transformation of the vector of the ridge coefficients; however, the transformation involves the inverse of the observed correlation matrix. Thus, if one wanted to calculate the unbiased least-squares coefficients indirectly to circumvent the numerical rounding error in calculating this matrix inverse, one could not do so through ridge regression.

In this article, we shall suggest an alternative means of overcoming multicollinearity that does not have these disadvantages of ridge regression. This alternative applies only to a limited, but frequently encountered, set of models with an inherent multicollinearity problem (Althausen, 1971)—namely, models with both “main” effects and multiplicative interaction terms. The simplest of these models is the three-term one of the general form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon \quad [1]$$

in which X_3 is the product of the two variables X_1 and X_2 :

$$X_3 = X_1 X_2. \quad [2]$$

This product is often so highly correlated with its constituent variables that it is difficult to separate the multiplicative effect from the additive ones. Alone, the effect of X_3 may be significant; but it frequently becomes insignificant when the additive effects β_1 and β_2 are added, as in model 1. Likewise, the additive effects alone may be jointly significant but decrease the residual sum of squares by an insignificant amount when added to a model

containing the constant and X_3 . (The statistical problems of these models are explored also by Allison, 1977; and Southwood, 1978. Their discussions complement ours.)

SOME ALTERNATIVE MODELS WITH MULTIPLICATIVE FUNCTIONS

Multiplicative functions such as X_3 have been frequently used to represent interaction effects, wherein the effect of one independent variable on the dependent variable depends upon the value of the other independent variable. The product of the two variables, however, is only one of an infinite number of continuous two-variable functions representing interactions, not to mention the infinite number of discontinuous functions as well. For instance, the term $\beta(\alpha_0 + \alpha_1 X_1)e^{-X_2}$ in a regression model would also represent an interaction between X_1 and X_2 . Our theories and theory-like descriptions of processes are seldom developed enough to suggest the specific form of an interaction; indeed, we are usually very fortunate if our theorizing unambiguously explains or predicts merely the presence, or absence, of an interaction between variables. (For a discussion of the linkage of verbal formulations of interactions and algebraic models, see Southwood, 1978.) Probably among the primary reasons why the multiplicative function 2 has been widely employed are that it is very easy to calculate and that it is at least somewhat sensitive to the presence of interactions of a wide range of functional forms.

Even among models with multiplicative functions, model 1 with X_3 is but one special case of the more general model

$$Y = \beta_0^* + \beta_1^* X_1 + \beta_2^* X_2 + \beta_4^* X_4 + \epsilon, \quad [3]$$

with

$$X_4 = (X_1 - c)(X_2 - d) \quad [4]$$

in which c and d are constants. In the commonly used model 1, c and d are zero; but there usually is again no clear theoretical

reason why they should be zero as opposed to some other values. Indeed, one can at times make a good argument for their being nonzero. Such an argument is in fact made at least implicitly by some proponents of consistency effects. Centering the variables around any values within the range of sample observations before forming their product means not only that the absolute value of the effect increases with the deviations of both variables from these central values, but also that the function will be positive if both variables deviate in the same direction and negative if they deviate in opposite directions. In some instances, one might substantively argue that c and d should be the means of X_1 and X_2 respectively so that the interaction effect increases with the deviations of the variables from their expected values; in others, one might argue that c and d should be the modes so that the effect increases with deviations from the most frequent values. No doubt there are cases in which arguments could be made for other nonzero values as well, although usually the choice will probably not be clearly dictated by substantive or theoretical considerations.

In general, we suspect the function 4 with nonzero values of c and d is more often substantively interpretable as an interaction effect than is function 2. There are, however, at least two cases in which function 2 with zero c and d does have a clear interpretation. The first is when both X_1 and X_2 are dummy or indicator variables for different categorized variables. In this case, the coefficient β_3 corresponds to the interaction effect in the analysis of variance for factorial designs. The second is when one variable, say X_1 , is a dummy variable and the other, X_2 , is an interval-level variable. In this case, β_3 represents the difference between the regression coefficient for X_2 in the group identified by the dummy X_1 and that in the base-line group. But even in cases such as these where one substantively wants estimates of the parameters in model 1, numerical error can be minimized by, first, estimating the parameters in model 3 using the minimizing values of c and d that we shall develop, and then calculating the estimates for model 1 from those for model 3. Using the procedure this way is analogous to using orthogonal polynomials for testing and estimating polynomial models.

The estimates of the parameters for model 1 can be calculated from those for model 3 because the two models are exact linear functions of one another. Expressing model 1 in terms of model 3,

$$\beta_0 = \beta_0^* + \beta_4^* cd, \quad [5]$$

$$\beta_1 = \beta_1^* - \beta_4^* d, \quad [6]$$

$$\beta_2 = \beta_2^* - \beta_4^* c, \quad [7]$$

$$\beta_3 = \beta_4^*. \quad [8]$$

Notice that the coefficients for the multiplicative function are unaffected by the values of c and d (equation 8). What are affected are the interpretations and the consequent values of the coefficients for X_1 and X_2 . More generally, estimates based on any given pair of values for c and d can be used to calculate the estimates for any other values of c and d .

Moreover, both models are alternative ways of stating the model

$$Y = \beta_0'' + \beta_3''(X_1 - \alpha)(X_2 - \gamma) + \epsilon, \quad [9]$$

in which

$$\beta_0'' = \beta_0^* + \beta_1^* c + \beta_2^* d - \beta_1^* \beta_2^* / \beta_4^* = \beta_0 - \beta_1 \beta_2 / \beta_3, \quad [10]$$

$$\beta_3'' = \beta_4^* = \beta_3, \quad [11]$$

$$\alpha = c - \beta_2^* / \beta_4^* = -\beta_2 / \beta_3, \quad [12]$$

$$\gamma = d - \beta_1^* / \beta_4^* = -\beta_1 / \beta_3. \quad [13]$$

The equivalencies after the first equal signs are for model 3; those after the second are for model 1. In other words, no matter what the values of c and d are, adding function 4 (or function 2) to a two-variable additive model is equivalent to using model 9. Stated in this form, the model has no additive effects of X_1 and X_2 ; it has instead only an effect for the product of their

deviations from α and γ respectively. Stated this way, the model can be interpreted as a "threshold" one. The parameter α is the threshold level for X_1 , above which X_1 contributes positively to Y and below which it contributes negatively; and the parameter γ is the threshold level for X_2 .

Two general points should be made about models 1, 3, and 9. First, because they are linear transformations of each other, they are in one respect the very same model. But, perhaps equally important, the three models represent very different ways of describing this same dependency of Y on X_1 and X_2 . They lend themselves to different kinds of substantive interpretations, and they vary in the ease and numerical accuracy with which they can be estimated. Depending upon the values of c and d , models 1 and 3 can also lead to different conclusions about the statistical significance of effects because the models can have different levels of multicollinearity. (For an insightful discussion of the problems of describing and interpreting coefficients more generally in multiple regression, see Mosteller and Tukey, 1977: 299-303, 419-422.)

The threshold model 9 is not directly amenable to linear estimation procedures; and it must be restated as either model 3 or its special case, model 1, before ordinary least-squares regression can be employed. Since our substantive theories are usually not precise enough to suggest specific values of c and d , it seems appropriate—at least for estimation—to fall back on numerical considerations and to select the values which will minimize the collinearity of the multiplicative function X_4 with X_1 and X_2 for the given sample, rather than to assume a priori that these constants should be zero.¹

In this article, these minimizing values of c and d are derived, and the effects that they and other possible values can have on multicollinearity in model 3 are explored. In what follows, asterisks will be used to denote statistics and parameters relating to model 3 with nonzero values of c and d ; symbols without asterisks refer to model 1 with X_3 rather than X_4 .

This procedure for decreasing multicollinearity has several attractive characteristics, some of which have already been alluded to. First, as we shall see, the values of c and d that mini-

mize the collinearity also make X_4 uncorrelated in the sample with X_1 and X_2 . Thus, the estimates of β_1 and β_2 that are obtained using model 3 and the minimizing constants are the same as those obtained using a model containing only the additive effects of X_1 and X_2 (just the first three terms on the right of models 1 and 3). In effect, the procedure establishes an estimating and testing hierarchy using the criterion of parsimony: additive effects first followed by multiplicative ones. Again, the parallel with orthogonal polynomials is clear.

Second, since the ordinary least-squares (OLS) estimates b_i of the β_i in model 1 can be expressed as linear functions of the OLS estimates b_i^* for model 3 and the constants c and d (equations 5-8), one can obtain the estimates b_i without the numerical errors caused by multicollinearity by calculating them indirectly through the estimates b_i^* and the minimizing values of c and d . Third, for the same reason the variances and covariances of one set of coefficients can be calculated from the variances and covariances of the other set and the values of c and d . Fourth, because models 1 and 3 are exact linear functions of one another, OLS regression for model 3 with X_4 produces exactly the same coefficient of determination as does OLS regression of model 1 and exactly the same minimum value of the mean-square error, assuming of course that the standard assumptions for OLS are correct for model 1. Thus, there is no loss in predictive or explanatory power when the procedure discussed here is used. Fifth, the procedure is not limited to three-term models such as models 1 and 3. If there are other variables in the model, the procedure can be used to decrease the collinearity between the multiplicative term and the variables comprising it without affecting the estimates of the coefficients for the other variables or the variances of these estimates.

AN INDEX OF MULTICOLLINEARITY

Before we can proceed, we need to adopt a quantitative measure of the degree of multicollinearity in a covariance or correlation matrix. The collinearity of any two variables is clearly

related to the size of their covariance or correlation, but we need a single summary measure for the multicollinearity among all the independent variables and functions as a set, not taken two at a time. Following Rockwell (1975), we shall use the determinant of the correlation matrix: the smaller it is, the greater is the multicollinearity. Since the determinant of a matrix is equal to the product of its eigenvalues, this criterion is closely associated with another common one, namely, the ratio of the matrix's largest to smallest eigenvalues. We shall use the determinant of the correlation matrix rather than of the covariance matrix, since the latter is greatly affected by the scales on which the variables are measured and multicollinearity is something that should be independent of the measurement scales. Measuring annual income in thousands of dollars rather than hundreds of dollars should not, for instance, affect the index of collinearity between it and other variables. The correlation matrix is also more relevant since the preferred numerical procedure for regression analysis, used by almost all computerized regression routines, works with the correlation matrix (see, for instance, Draper and Smith, 1966: 142-149). The determinant of the correlation matrix of just the independent variables and functions has the further advantage that it distinguishes between the amount of multicollinearity and the amount of the residual variance in the model. For a given set of independent variables and functions, the multicollinearity is the same regardless of the dependent variable. We should note that our entire discussion is predicated on the acceptance of this index of multicollinearity.

The following discussion is divided into four sections. After setting forth some notation and establishing some basic relationships, we shall see in part three that model 3 with X_4 will have less multicollinearity than model 1 only if the variance of X_4 is less than that of X_3 . In part four we shall derive the values of c and d that will minimize the index of multicollinearity for a given sample. Also in this section, we shall consider the conditions under which setting c and d equal to the sample means of X_1 and X_2 will decrease the level of multicollinearity relative to that with c and d set equal to zero. In part five we shall establish the linear relationship between the variances and covariances of

the OLS coefficients b_i and those of the b_i^* and show that this procedure for reducing the multicollinearity related to multiplicative functions does not affect the mean-square error, the OLS coefficients for other variables, or their variances. Finally, we shall close with an empirical example.

SOME BASIC RELATIONSHIPS

Since our ultimate interest is in the determinants of the correlation matrices, let us begin with the variances and covariances of our variables (and functions). One advantage of working with variances and covariances rather than with uncentered sums of squares and crossproducts is that the intercept terms β_0 and β_0^* are independent of the centered independent variables and can be omitted from most of our discussion. Once the other OLS estimates are obtained, b_0 and b_0^* can easily be calculated by the familiar formulas.

We shall use the notation s_{ij} to refer to the unbiased sample covariance between X_i and X_j and s_{ii} to refer to the unbiased sample variance of X_i . At several points, we shall need to express the variances and covariances involving X_4 in terms of those involving X_1 , X_2 , and X_3 . Expanding equation 4 for X_4 and substituting from equation 2 for X_3 ,

$$X_4 = X_3 - dX_1 - cX_2 + cd. \quad [14]$$

Using equation 14 and the standard formula for the covariance of a variable with a composite, the covariance of X_1 and X_4 is

$$s_{14} = s_{13} - ds_{11} - cs_{12}. \quad [15]$$

Similarly,

$$s_{24} = s_{23} - ds_{12} - cs_{22}. \quad [16]$$

The standard formula for the variance of a composite applied to equation 14 gives us

$$s_{44} = s_{33} + d^2s_{11} + c^2s_{22} - 2ds_{13} - 2cs_{23} + 2cds_{12}. \quad [17]$$

Let \underline{S} be the 3-by-3 covariance matrix for X_1 , X_2 , and X_3 and \underline{S}^* be the covariance matrix for X_1 , X_2 , and X_4 . Also define the matrix

$$\underline{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -d & -c & 1 \end{bmatrix} \quad [18]$$

From equations 15 to 17, it follows that $\underline{S}^* = \underline{A}\underline{S}\underline{A}'$. The determinant of \underline{S}^* , $|\underline{S}^*|$, consequently equals $|\underline{A}| |\underline{S}| |\underline{A}'|$. Since \underline{A} is lower triangular, $|\underline{A}|$ equals the product of its diagonal elements, or 1.0; and $|\underline{S}^*| = |\underline{S}|$.²

Thus, using X_4 rather than X_3 will not alter the determinant of the covariance matrix. But will it alter the determinant of the correlation matrix, our index of multicollinearity? The correlation matrix \underline{R} for X_1 , X_2 , and X_3 is simply the covariance matrix \underline{S} with every row and column divided through by the corresponding standard deviation $s_{ii}^{1/2}$. A basic theorem in linear algebra is that multiplying any row or column of a square non-singular matrix \underline{F} by a constant k results in a determinant of $k|\underline{F}|$. Given the relationship between \underline{S} and \underline{R} , it follows that

$$|\underline{R}| = |\underline{S}| / s_{11} s_{22} s_{33}. \quad [19]$$

Similarly, if \underline{R}^* is the correlation matrix for X_1 , X_2 , and X_4 , then

$$\begin{aligned} |\underline{R}^*| &= |\underline{S}^*| / s_{11} s_{22} s_{44} \\ &= |\underline{S}| / s_{11} s_{22} s_{44} \end{aligned} \quad [20]$$

since $|\underline{S}^*| = |\underline{S}|$. From equations 19 and 20 we can see that using X_4 instead of X_3 will decrease the multicollinearity if and only if $s_{44} < s_{33}$.³

REDUCING THE VARIANCE OF THE MULTIPLICATIVE FUNCTION

To minimize the multicollinearity due to the multiplicative function, we want to minimize the variance of X_4 . Treating the variances and covariances in \underline{S} as given constants for a particular sample of observations, we can see from equation 17 that s_{44} is a continuous function of c and d . Differentiating equation 17 with respect to c and d , we obtain the partial derivatives

$$\frac{\partial s_{44}}{\partial d} = 2ds_{11} - 2s_{13} + 2cs_{12}. \tag{21}$$

$$\frac{\partial s_{44}}{\partial c} = 2cs_{22} - 2s_{23} + 2ds_{12}. \tag{22}$$

Setting the partial derivatives 21 and 22 to zero and solving for c and d , we obtain as the values of c and d which minimize s_{44} :

$$c = \frac{s_{11}s_{23} - s_{12}s_{13}}{s_{11}s_{22} - s_{12}^2} = \frac{(r_{23} - r_{12}r_{13})s_{33}^{1/2}}{(1 - r_{12}^2)s_{22}^{1/2}} \tag{23}$$

$$d = \frac{s_{13}s_{22} - s_{12}s_{23}}{s_{11}s_{22} - s_{12}^2} = \frac{(r_{13} - r_{12}r_{23})s_{33}^{1/2}}{(1 - r_{12}^2)s_{11}^{1/2}}. \tag{24}$$

The right-hand expressions in terms of the correlations r_{ij} are more easily used with the standard output of the popular statistical packages. An examination of the second partial derivatives confirm that these values of c and d do indeed define a minimum for s_{44} .

Interestingly, we can see from equations 15 and 16 that $\partial s_{44} / \partial d = -2s_{14}$ and $s_{44} / \partial c = -2s_{24}$. Thus, by setting the partial derivatives to zero, we are also setting s_{14} and s_{24} to zero.⁴ Put somewhat figuratively, ridge regression deals with multicollinearity by increasing the height of the main diagonal, or ridge, whereas the procedure described here deals with it by lowering the height of the off-diagonal elements. One consequence of setting s_{14} and

s_{24} to zero is that the OLS estimates b_1^* and b_2^* will be exactly the same as they would be if the multiplicative function were omitted from the model. Thus, setting c and d to their minimizing values means that X_1 and X_2 are allowed to account additively for all the variance in Y that they can; the multiplicative function X_4 merely works with the residual variance left over. Consequently, this procedure does not find multiplicative interaction effects at the expense of main effects.

As noted earlier, substantive considerations at times may suggest that the multiplicative term should involve deviations from means, that is, that c and d in equation 4 should be the sample means \bar{X}_1 and \bar{X}_2 respectively. Under what conditions will a multiplicative function based on deviations from means result in less multicollinearity than the function 2 based on raw values of X_1 and X_2 ? Since s_{33} is simply s_{44} with $c = 0$ and $d = 0$, the problem can be stated more tractably as that of finding the conditions under which s_{44} with $c = \bar{X}_1$ and $d = \bar{X}_2$ is less than s_{44} with $c = 0$ and $d = 0$.

From equation 17 for s_{44} , we can see that the surface of the function s_{44} is symmetrical with respect to the planes $c = 0$ and $d = 0$. As a result of this symmetry, there is an easily calculated set of such conditions. The value of s_{44} with c and d equal to the means will be less if the absolute values of the partial derivatives of s_{44} with respect to c and d are less at the point on the surface defined by $c = \bar{X}_1$ and $d = \bar{X}_2$ than are the absolute values of the partials at the point defined by $c = 0$ and $d = 0$. At this latter point where $s_{44} = s_{33}$, $\partial s_{44} / \partial d = -2s_{13}$ and $\partial s_{44} / \partial c = -2s_{23}$ (see equations 21 and 22); similarly, at the point where c and d equal the means, $\partial s_{44} / \partial d = 2\bar{X}_2 s_{11} - 2s_{13} + 2\bar{X}_1 s_{12}$ and $\partial s_{44} / \partial c = 2\bar{X}_1 s_{22} - 2s_{23} + 2\bar{X}_2 s_{12}$. Thus, the value of s_{44} with X_1 and X_2 centered about their means will be less than the value of s_{33} if and only if the following two conditions are true:

$$(1) \text{ If } s_{13} > 0, \text{ then } 0 < \bar{X}_2 s_{11} + \bar{X}_1 s_{12} < 2s_{13}; \quad [25]$$

$$\text{if } s_{13} < 0, \text{ then } 0 > \bar{X}_2 s_{11} + \bar{X}_1 s_{12} > 2s_{13}.$$

$$(2) \text{ If } s_{23} > 0, \text{ then } 0 < \bar{X}_1 s_{22} + \bar{X}_2 s_{12} < 2s_{23};$$

$$\text{if } s_{23} < 0, \text{ then } 0 > \bar{X}_1 s_{22} + \bar{X}_2 s_{12} > 2s_{23}. \quad [26]$$

For most sets of data, these conditions are met; indeed, the minimizing values of c and d will often be close to \bar{X}_1 and \bar{X}_2 . However, there are some unusual combinations of means, variances, and covariances for which they are not true. In these unusual cases, centering the variables about their means before forming the multiplicative term will not, consequently, decrease the multicollinearity.

Conditions 25 and 26 can also be used to determine the effect on multicollinearity of any other set of values for c and d , by merely replacing \bar{X}_1 and \bar{X}_2 in conditions 25 and 26 with these other values. Of course, one can always estimate the parameters for model 3 using the minimizing values of c and d and then, through linear transformations, obtain estimates of the parameters associated with any other values of c and d .

SOME IMPLICATIONS OF THE LINEAR RELATIONSHIP OF THE b_1 AND b_1^*

As we saw in equations 5 to 8, model 1 with X_1 , X_2 , and X_3 is a linear transformation of model 3 with X_1 , X_2 , and X_4 . Models 1 and 3 are simply alternative ways of stating the same functional relationship of Y with X_1 and X_2 . As a consequence, the sums of squares explained by the two models are equal; and since they both have the same degrees of freedom, the mean-square errors for the two models are also equal. Furthermore, the variances and covariances of one set of OLS estimates can be calculated from those of the other set. For instance, substituting the OLS estimates into equivalencies 5 to 8 and using the standard formula for the variance of a sum,

$$\text{var}(b_0) = \text{var}(b_0^*) + c^2 d^2 \text{var}(b_4^*) + 2cd \text{cov}(b_0^*, b_4^*) \quad [27]$$

$$\text{var}(b_1) = \text{var}(b_1^*) + d^2 \text{var}(b_4^*) - 2d \text{cov}(b_1^*, b_4^*) \quad [28]$$

$$\text{var}(b_2) = \text{var}(b_2^*) + c^2 \text{var}(b_4^*) - 2c \text{cov}(b_2^*, b_4^*) \quad [29]$$

$$\text{var}(b_3) = \text{var}(b_4^*). \quad [30]$$

Finally, there is the fact that using X_4 rather than X_3 has no effect on the OLS estimates of the regression coefficients for other variables in an expanded model and none on the variances of these estimates. For instance, let us say that we add a term for variable X_5 to both models 1 and 3, creating the expanded models whose OLS estimates are respectively

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_5 X_5 + e, \quad [31]$$

and

$$Y = b_0^* + b_1^* X_1 + b_2^* X_2 + b_4^* X_4 + b_5^* X_5 + e. \quad [32]$$

Substituting the values of X_4 from equation 4 into model 32 and gathering terms,

$$Y = (b_0^* + b_4^* cd) + (b_1^* - b_4^* d)X_1 + (b_2^* - b_4^* d)X_2 + b_4^* X_1 X_2 + b_5^* X_5 + e. \quad [33]$$

Since equations 31 and 33 are both OLS solutions, it is immediately evident from a comparison of the two that b_5^* does indeed equal b_5 and that their variances are equal.

AN EXAMPLE

For an empirical example, we can use a model which Jackman (1974: 38-39) suggested, but which he did not analyze because of the high multicollinearity in his data. The data are aggregate for 60 countries in 1960. The dependent variable, Y , is the Schutz coefficient of income equality, reversed from its usual form so that a high value indicates a high level of income equality (Jackman, 1975: 64-65; 1974: 37-38). The first independent variable, X_1 , is the normal logarithm of the energy consumption per capita in 1960, measured in kilograms. The second independent variable, X_2 , is Jackman's (1975: 64-65; 1974: 37-38) measure of a country's democratic performance in the period around

1960. The values for Y and X_2 were obtained from Jackman (1975: 210-211, 216-218); those for X_1 , from Taylor and Hudson (1972: 326-328).

In the literature on the determinants of social equality, there are several at least implicit suggestions that there may be an interaction effect between the level of democracy in a country and its level of economic development. Accordingly, Jackman suggested using our model 1, including the multiplicative X_3 , with the log of per capita energy consumption and the democratic performance index. However, the high collinearity of X_3 with X_1 and X_2 make the OLS estimates for the model both unreasonable and unstable.

In the first four rows and columns of Table 1 are the correlations and covariances of the three variables and the function X_3 . On the diagonal of the matrix are the variances, below the diagonal are the correlations, and above the diagonal are the covariances; to the right of the matrix are the means. Note especially the very high values of 0.852 and 0.907 for r_{13} and r_{23} . The determinant of the correlation matrix for X_1 , X_2 , and X_3 is 0.00727; Haitovsky's heuristic chi-square statistic for this determinant (Rockwell, 1975) has a value of 0.417. We have here an extreme case of multicollinearity.

In the first two columns of Table 2 are the OLS coefficients and their standard errors from the regression of Y on X_1 , X_2 , and X_3 . At least two aspects of these results are of interest. First, the partial coefficients for both log energy consumption per capita and democratic performance indicate that these variables have negative effects on income equality; and yet the zero-order correlations, the estimates for the additive model containing only β_0 , β_1 , and β_2 , other research, and common sense all strongly suggest that their effects are positive. One suspects that we have an instance of "bouncing betas." Second, the standard errors of the coefficients are all quite large relative to the coefficients, even though the regression as a whole is significant well beyond the 0.001 level ($F = 14.750$). The coefficient for the function X_3 is just significant at the 0.05 level ($F = 4.016$); that for X_2 has a descriptive significance of 0.065 ($F = 3.555$). The absolute value

TABLE 1
Means, Variances (on Diagonal), Covariances (above Diagonal), and
Correlations (below Diagonal) for (Y) Schutz Coefficient, (X_1)
Log Energy Consumption, (X_2) Democratic Performance, (X_3)
Raw Product of X_1 and X_2 , (X_4) Product with Minimum Variance,
and (X_4') Product Using Means

Var.	(Y)	(X_1)	(X_2)	(X_3)	(X_4)	(X_4')	Mean
(Y)	<u>141.743</u>	10.212	89.515	1311.983	46.938	62.460	80.852
(X_1)	0.632	<u>1.840</u>	14.783	219.645	0.000	3.127	6.430
(X_2)	0.392	0.568	<u>367.999</u>	3306.117	0.000	-35.753	66.000
(X_3)	0.580	0.852	0.907	<u>36139.113</u>	-----	-----	438.902
(X_4)	0.200	0.000	0.000	-----	<u>388.003</u>	-----	13.643
(X_4')	0.260	0.114	-0.092	-----	-----	<u>408.181</u>	14.537

SOURCES: The data for these calculations were taken from Jackman (1975: 210-211, 216-218) and Taylor and Hudson (1972: 326-328).

N = 60 countries.

of the coefficient for log energy consumption per capita is only about three-fourths the value of its standard error ($p = .475$).

We know that using X_4 with some central values for c and d , rather than X_3 , would keep the estimates for β_1 and β_2 (now β_1^* and β_2^*) positive, as they are in the simple additive model. Furthermore, in the literature discussed by Jackman, there is at least partial support for basing the multiplicative function on deviations from some central values. For instance, he quotes Cutright (1965: 547; quote from Jackman, 1974: 39) as saying: "Again, it is seen that nations with the lowest level of economic development are not introducing social insurance programs even when they have a positive PRI [Political Representativeness Index] change." This sentence implies that countries with negative deviations for energy consumption and positive ones for democratic performance are likely to have less income equality than would be expected from a simple additive model, and the function X_4 in such a case would be negative. It should be noted, however, that other consequences of incorporating X_4 into the model are not suggested by Jackman's discussion. Do a positive deviation for energy consumption and a negative one for democratic performance also tend to deflate income equality relative

TABLE 2
Regressions of Schutz Coefficient of Income Equality, 1960, on
Log Energy Consumption, Democratic Performance, and
Three Alternative Multiplicative Functions

Independent Variable	Product of Raw X_1 and X_2		Product with Minimum s_{e1}		Product with $c = X_1; d = X_2$	
	Coef.	S. E.	Coef.	S. E.	Coef.	S. E.
Ln energy consumption per capita, 1960	-3.120	4.338	5.307	1.065	4.865	1.088
Democratic perform- ance, c. 1960	-0.718	0.381	0.030	0.075	0.060	0.077
Multiplicative function	0.121	0.060	0.121	0.060	0.121	0.060
(Intercept)	95.219		43.094		43.061	
Multiple R-square	0.441		0.441		0.441	
Mean square error	83.421		83.421		83.421	

N = 60.

to that expected from an additive model; do two negative deviations have the same positive interaction effect as do two positive ones (they will, of course, have opposite additive effects in the model)? Clearly, a serious substantive investigation using X_4 to represent an interaction effect should include a careful analysis of the residuals. Such an analysis, however, is beyond the purview of this article.

First, let us examine the regression results when the constants c and d are set to the values that will minimize the multicollinearity. Using equations 23 and 24 and the figures in Table 1, these values are $c = 6.186$ and $d = 69.661$. (These and the following calculations were made with more significant figures than are reported here and then rounded.) The variance, correlations, and covariances for X_4 using these values of c and d are given in the fifth row and column of Table 1. The variance of X_4 (388.003) is considerably less than that of X_3 (36,139.113); and, as expected, the correlations of X_4 with X_1 and X_2 are both zero. The determinant of the correlation matrix of X_1 , X_2 , and X_4 is 0.677, and the value of Haitovsky's chi-square statistic is 64.657, a vast improvement over the value of 0.417 for the matrix with X_3 .

The results from the regression including this minimizing X_4 are given in the second pair of columns in Table 2. First, note that the multiple R-square and the mean-square error are indeed the same as in the first regression. Second, the coefficients for log energy consumption per capita and for democratic performance now have the expected signs. Also, their standard errors have been reduced by more than a factor of four. The coefficient for log energy consumption per capita is now very significant with an F value of 24.833. The coefficient for democratic performance, however, has an F value of only 0.159, which is consistent with Jackman's findings. As noted briefly in the section entitled "Reducing the Variance of the Multiplicative Function," the coefficients for these two variables are exactly the same as those obtained in the regression of Y on just X_1 and X_2 , without any multiplicative function.⁵ Their standard errors in Table 2, however, are less since the addition of X_4 has decreased the mean-square error.

Turning to the coefficient for X_4 , we see that the effect of X_4 is larger than the additive effect of X_2 (the standardized regression coefficient for X_4 is 0.200; that for X_2 is 0.048). This is particularly interesting in light of the fact that X_4 is, in effect, working with only the residual variance in Y left after X_1 and X_2 have accounted for all they can. The coefficient for X_4 is again just significant at the 0.05 level, suggesting that the possibility of an interaction between log energy consumption per capita and democratic performance should be explored more thoroughly, starting probably with a careful analysis of residuals. Is there indeed an interaction of a functional form that is only being partially captured by this multiplicative function? It may be of interest to note that adding X_4 to an additive model provides a better fit with the observed values of Y than does a logarithmic interaction model (Jackman, 1974: 39-40).

Alternatively, since the coefficient for X_2 is clearly insignificant in the model with X_4 (with c and d equal to either their minimizing value or their sample means, as discussed next), one might instead wish to explore the residuals from the model including just X_1 and X_4 .⁶

We can also set c and d equal to the means of X_1 and X_2 respectively. Not much needs to be said about this exercise since these data happen to be one of those instances in which the means of X_1 and X_2 (6.430 and 66.000 from Table 1) are very close to the minimizing values of c and d (6.186 and 69.661). To confirm that basing X_4 on deviations from the means will reduce the multicollinearity, we can check conditions 25 and 26. Using the figures in Table 1, $\bar{X}_2s_{11} + \bar{X}_1s_{12} = 216.519$ and $X_1s_{22} + X_2s_{12} = 3341.870$, which are very close to the values of s_{13} and s_{23} respectively.

The product of deviations from means is identified in Table 1 as X_4' . Its variance, correlations, and covariances are given in the sixth row and column of that table. Although its correlations with X_1 and X_2 are not zero, they are close to it and are considerably less than those for X_3 . Note also that the variance of X_4' is only fractionally greater than that of X_4 . The results of the regression including this X_4' are given in the last pair of columns in Table 2. Not surprisingly, these results give essentially the same picture as the regression with the minimizing values of c and d . The fact that the coefficient for the multiplicative function is the same in all three regressions, as is its standard error, merely illustrates the algebraic equivalences developed in equations 8 and 30.

Finally, we can express all three sets of estimates in Table 2 in terms of the general model 9. Substituting the OLS estimates into the equivalencies 10 to 13, our estimate of this general model is

$$Y = 76.698 + 0.121 (X_1 - 5.937) (X_2 - 25.787) + e.$$

Thus, in this restatement of our three four-term models, we estimate 5.937 to be the "threshold" for log normal energy consumption. Above this level, it contributes positively to the multiplicative effect of it and democratic performance, and below this level it contributes negatively. Likewise, 25.787 is our estimate of the threshold for the democratic performance index.

NOTES

1. Mosteller and Tukey (1977: 285-287) recommend a similar centering transformation to decrease collinearity in polynomial models.

2. We are indebted to an anonymous reviewer for this proof, which is simpler and more elegant than the one in our original manuscript.

3. Of course, in the case where X_1 is an exact linear function of X_2 and \underline{S} is singular, then $|\underline{R}| = |\underline{R}^*| = 0$ regardless of the relative values of s_{13} and s_{44} . But there is probably something wrong with the conceptualization of the analysis if two variables that are exact linear functions of each other are in the same regression model, and it is not reasonable to expect any numerical procedure to rescue one from the muddle. We are concerned here with high multicollinearity and very small determinants, not with singularity and algebraically zero determinants.

4. John W. Tukey (personal communication) has suggested that one could decrease the multicollinearity even more by using, instead of X_2 , $X_2' = X_2 + fX_1$, where f is chosen to make zero for correlation of X_2' and X_1 , and then to use $X_4 = (X_1 - c)(X_2' - d)$. We would then have a complete hierarchical ordering of effects; the estimate of the effect for X_1 would be unaffected by the addition of X_2' to the model, and the estimates for both X_1 and X_2' would be unaffected by the addition of X_4 . Since the substitution of X_2' for X_2 is simply a further linear transformation of model 3, one could readily calculate estimates of the parameters for models 1, 3, or 9 from this one. If X_1 and X_2 were highly correlated, this extension of the procedure would be useful. Since the extension is straightforward, there is no need to explore it further here.

5. Our results do not agree exactly with Jackman's (1974: 38; compare also 1975: 08) because of rounding in the published data.

6. Using X_4' with c and d equal to the means, the fitted equation for just X_1 and X_4' is $Y = 44.771 + 5.358X_1 + 0.112X_4' + e$. Using equations 10 to 13 with $\beta_2^* = 0$, this can be restated in the threshold form of model 9: $Y = 79.224 + 0.112(X_1 - 6.430)(X_2 - 18.144) + e$. For this model, the values of c and d that one uses do make a difference because, in expanded form, the coefficient for X_2 is constrained to be $-c$ times the coefficient for the product X_1X_2 . The OLS function being fitted is, following the notation for model 3, $b_0^* + b_1^*X_1 + b_4^*(X_1 - c)(X_2 - d)$. Expanded, this becomes $(b_0^* + b_4^*cd) + (b_1^* - b_4^*d)X_1 - b_4^*cX_2 + b_4^*X_1X_2$. The constraint is in the last two terms.

REFERENCES

- ALLISON, P. D. (1977) "Testing for interaction in multiple regression." *Amer. J. of Sociology* 83: 144-153.
- ALTHAUSER, R. P. (1971) "Multicollinearity and non-additive regression models," pp. 453-472 in H. M. Blalock, Jr. (ed.) *Causal Models in the Social Sciences*. Chicago: Aldine-Atherton.
- BLALOCK, H. M., Jr. (1963) "Correlated independent variables: the problem of multicollinearity." *Social Forces* 62: 233-238.
- CUTRIGHT, P. (1965) "Political structure, economic development, and national social security programs." *Amer. J. of Sociology* 70: 537-550.

- DEEGAN, J., Jr. (1975) "The process of political development: an illustrative use of a strategy for regression in the presence of multicollinearity." *Soc. Methods & Research* 3: 384-415.
- DRAPER, N. R. and H. SMITH (1966) *Applied Regression Analysis*. New York: John Wiley.
- GORDON, R. A. (1968) "Issues in multiple regression." *Amer. J. of Sociology* 73: 592-616.
- GUILKEY, D. K. and J. L. MURPHY (1975) "Directed ridge regression techniques in cases of multicollinearity." *J. of the Amer. Statistical Assn.* 70: 769-775.
- HENRY, N. W. (1976) "A note on ridge regression." *Soc. Methods & Research* 4: 495-500.
- HOERL, A. E. (1962) "Application of ridge analysis to regression problems." *Chemical Engineering Progress* 58: 54-59.
- and R. W. KENNARD (1970a) "Ridge regression: biased estimation for non-orthogonal problems." *Technometrics* 12: 55-67.
- (1970b) "Ridge regression: applications to nonorthogonal problems." *Technometrics* 12: 69-82.
- JACKMAN, R. W. (1975) *Politics and Social Equality: A Comparative Analysis*. New York: Wiley-Interscience.
- (1974) "Political democracy and social equality: a comparative analysis." *Amer. Soc. Rev.* 39 (February): 20-45.
- KASARDA, J. D. and W-F. P. SHIH (1977) "Optimal bias in ridge regression approaches to multicollinearity." *Soc. Methods & Research* 5: 461-470.
- MARQUARDT, D. W. and R. D. SNEE (1975) "Ridge regression in practice." *Amer. Statistician* 29 (February): 3-20.
- MOSTELLER, F. and J. W. TUKEY (1977) *Data Analysis and Regression: A Second Course in Statistics*. Reading, MA: Addison-Wesley.
- ROCKWELL, R. C. (1975) "Assessment of multicollinearity: the Haitovsky test of the determinant." *Soc. Methods & Research* 3 (February): 308-320.
- SOUTHWOOD, K. E. (1978) "Substantive theory and statistical interaction: five models." *Amer. J. of Sociology* 83: 1154-1203.
- TAYLOR, C. L. and M. C. HUDSON (1972) *World Handbook of Political and Social Indicators*. New Haven: Yale Univ. Press.

Kent W. Smith is Assistant Professor of Sociology at Northwestern University. His primary interests are in quantitative methods, especially the analysis of contingency tables and structural-equation models; research design and theory construction; professional socialization; and higher education.

M. S. Sasaki, a Ph.D. candidate at Princeton University, has been a visiting lecturer at the University of Michigan. He is currently completing his dissertation, "Network Analysis of Industrial Organizations in Japan: Group Structure and Productivity." His substantive interests are, in addition to small groups, complex organizations, Japanese social structure, and comparative societies.