

S105  
ESS  
no. 394  
Dep. 2



# Effect of Omitting Relevant Variables Versus Use of Ridge Regression in Economic Research



Special Report 394    October 1973  
Agricultural Experiment Station  
Oregon State University, Corvallis

EFFECT OF OMITTING RELEVANT VARIABLES VERSUS  
USE OF RIDGE REGRESSION IN ECONOMIC RESEARCH

William G. Brown\*

Foreword

An earlier draft of the first part of this paper, The Effect of Omitting Relevant Variables, was prepared in 1968. Since that time, most econometrics textbooks have improved their treatment of this important topic, e.g., Johnston's 1972 second edition [17, pp. 168-169]. Thus, econometricians and experienced applied economic researchers are sufficiently aware of the pitfalls of specification bias. However, less experienced researchers (even though they may know, mathematically, that bias exists whenever their estimated model does not include the same variables as the true model) could be helped by observing the great magnitude of bias which can actually occur in certain empirical problems. With increased awareness of the dangerous consequences of specification bias, researchers should use more care in specifying their models, and should also be encouraged to employ various methods of incorporating prior information into their regression analysis [18, 32] to mitigate problems of multicollinearity, rather than to merely delete variables to overcome this problem.

Having outlined the general problem involved in omitting relevant variables, a relatively new method for coping with multicollinearity, "ridge" regression [15, 16, 26], is explored. At the time of this writing, I have found no application of this technique to economic research, even though most economic data are highly aggregated, resulting in increased multicollinearity. Therefore, the potential usefulness and limitations of ridge regression for economic research are quite important to know, and are explored in the second part of this paper. Because biased linear estimation is in a rapidly evolving stage of development, this second part is in the nature of a progress report.

---

\* Research conducted under Oregon Agricultural Experiment Station Research Project No. 128. The author is greatly indebted to David Fawcett for writing the ridge regression computer programs for this research, and to Doug Young, Richard Johnston, Don Pierce, Jim Fitch, Lyle Calvin, and Norman Johnson for helpful comments. Encouragement at several stages of this research, from Emery Castle, is also appreciated. Of course, only the author is responsible for errors or deficiencies.

## TABLE OF CONTENTS

	<u>Page</u>
FOREWORD. . . . .	i
BACKGROUND. . . . .	1
OMITTED-VARIABLE BIAS IN PRODUCTION FUNCTION MODELS. . . . .	2
Knight's Empirical Model . . . . .	2
Ruttan's Production Function Estimates . . . . .	5
Effect of Omitting Variables from the Knight Model . . . . .	6
OMISSION OF DISTANCE FROM AN OUTDOOR RECREATION DEMAND MODEL . . . . .	9
SUGGESTIONS FOR ALLEVIATING PROBLEMS OF MULTICOLLINEARITY AND OMITTED-VARIABLE BIAS . . . . .	11
DEFINITION AND VARIANCE OF THE RIDGE REGRESSION ESTIMATOR . . . . .	12
EXPECTED BIAS OF THE RIDGE REGRESSION ESTIMATOR IN TERMS OF THE TRUE $\beta$ VALUES . . . . .	14
EXPERIMENTAL RESULTS FOR TWO-EXPLANATORY- VARIABLES . . . . .	16
Selection of Optimal k Values. . . . .	27
Implications for Larger Models . . . . .	30
SUMMARY AND CONCLUSIONS . . . . .	35
REFERENCES. . . . .	38

EFFECT OF OMITTING RELEVANT VARIABLES VERSUS  
USE OF RIDGE REGRESSION IN ECONOMIC RESEARCH

Background

Almost every type of estimation difficulty can be categorized as a "specification" problem of one kind or another. However, some of the other sources of specification error will be neglected here to concentrate attention on the problem of omission of important variables. The mathematical consequences of this problem have long been known (Theil [31]). Theil's results were applied by Griliches [10], primarily to ascertain the effect of specification bias on returns to scale for Cobb-Douglas production functions.

Despite this earlier research, the dangerous consequences of "omission-of-relevant-variables" bias has been somewhat neglected, perhaps because in the past there have been no very satisfactory solutions to the problem of multicollinearity, which has often forced the deletion of one or more explanatory variables. Mathematically, the problem of omitted variables can be seen by looking at the well-known formula for estimating an ordinary least squares regression coefficient, where there are two explanatory variables,

$$(1) \quad \hat{\beta}_1 = \left( \frac{r_{y1} - r_{12}r_{y2}}{1 - r_{12}^2} \right) \sqrt{\frac{\sum y_i^2}{\sum x_{1i}^2}}.$$

In (1) above, lower case x and y indicate mean-corrected variables and  $r_{12}$  refers to the correlation between the two explanatory variables,  $X_1$  and  $X_2$ . As can be easily seen from (1), the unbiased estimate of  $\hat{\beta}_1$  must also include the effect of  $X_2$ . If  $X_2$  is omitted, then the resulting estimate of  $\hat{\beta}_1$  will be biased, except for the trivial cases where  $X_2$  does not affect Y, or where  $X_1$  and  $X_2$  are uncorrelated. (These are trivial cases, since if  $X_2$  and Y were unrelated,  $X_2$  should not have been included in the model in the first place. Similarly, zero correlation between  $X_1$  and  $X_2$  is unusual for economic data.)

The bias for the two-variable case extends easily to the case for  $k$  explanatory variables, as first shown by Theil [31]. A good discussion of the importance of the omitted-variables problem is given by Malinvaud [25, pp. 263-266]. Some other texts also show the bias resulting from omitted variables but, unfortunately, give little indication of its importance in applied research - for example, Goldberger [9, pp. 195-197].

My contention is that the possibly devastating effect on parameter estimation caused by omission of significant variables in economic research has been too often ignored. At this point, some may object that we should be willing to pay the price of a small bias, if the estimates are fairly "good" in other respects.

This objection is difficult to refute mathematically, since there seems to be little that can be deduced in general about the magnitude of bias of a particular regression coefficient when one or more unknown but significant variables are omitted. As with many questions involving bias, the actual magnitude of omitted-variable bias depends upon the particular empirical problem. Therefore, it should be informative to economists to examine the impact of deleting certain commonly omitted variables from some actual models, where the original models appear to be satisfactorily specified. Then, in the latter part of this paper, the possibilities for correcting the conditions leading to omission of relevant variables is explored more extensively.

#### Omitted-Variable Bias in Production Function Models

The specification problem appears to have received the most attention in empirical work by those researchers, such as Griliches [10], Mundlak [27], Hoch [12], and Paris [29], attempting to estimate statistical production functions. Much of this earlier work concentrated on the problem of accounting for the effect of management in estimating the production function.

#### Knight's Empirical Model

In some recent remarkable studies, Knight [22, 23, 24] has examined the relation of farm size and efficiency to economic outcomes on Eastern, Central,

and Western Kansas farms. In terms of the relatively complete specification of model,<sup>1/</sup> accuracy and stability of data, and plausibility of results, Knight's empirical findings are surprisingly good, in comparison to other such studies where coefficients often have the "wrong" signs and problems of multicollinearity force the combination or deletion of various inputs.

Knight's success can be attributed partly to the excellent data that he analyzed. For his Central Kansas study, he had 394 farm management association farms, with price, input, and output data for each farm averaged over a five-year period, 1960-1964 [23, p. 1]. Knight was thus able to average out most of the year-to-year variability in yields and prices for crops and livestock. Also, the use of only farm management association farms was an important factor in maintaining high (and therefore fairly uniform) levels of management. If one were to randomly sample all farms, the large variation in the management input would be expected to seriously bias the other parameter estimates, Griliches [10]. Then, one would need to use other procedures, as did Hoch [14] and Mundlak [27], or else one should try to measure the managerial input, as did Paris [29].

Knight's Central Kansas results are summarized in Table 1. The functional form was a first-degree polynomial. Only two coefficients (Betas) seem somewhat lower than might first be expected. These two low coefficients are for  $X_2$ , feed purchased and veterinary and livestock expense, and for  $X_6$ , value of livestock inventory. However, cattle prices during the five-year period 1960-1964 were somewhat below average,<sup>2/</sup> and could be one reason for lower coefficients for  $X_2$  and  $X_6$ . Persons familiar with the Central Kansas Farm Management Association area

---

<sup>1/</sup> Of course, in practice an empirical production function can never be completely specified, since input categories could always be more finely subdivided or better measurements obtained. Also, questions of model formulation relating to identification, Klein [19, pp. 193-194] and Hoch [12, 14], or control for management by Mundlak [27] and Paris [29] can be raised. Other algebraic forms of the production function have also been proposed, Halter, et al. [11], Arrow, et al. [1].

<sup>2/</sup> During the post-World War II period, 1947-1967, average price per 100 pounds for all grades of slaughter cattle averaged \$26.32. During the five years 1960-1964, the average price was approximately 94.4 percent of the 1947-1967 period [34, 35].

Table 1. Values Obtained in Regression Analysis of 394 Central Kansas Farm Management Association Farms, 1960 to 1964. (Taken from Knight [23, p. 45].)

	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	Y
Means.....	1.6182	5600.8	2318.3	6596.2	9964.5	15494	38939	56150	26497
Betas.....	4498.2	.98264	1.6932	1.1691	.19121	.03333	.04496	.01793	
Standard deviation....	496.30	.04500	.17182	.12693	.02857	.02150	.00936	.00699	13583
t values.....	9.0634	21.838	9.8544	9.2107	5.6432	1.5505	4.8029	2.5657	

X<sub>1</sub> = Number of men.

X<sub>2</sub> = Dollars feed purchased and veterinary and livestock expense.

X<sub>3</sub> = Dollars seed and crop and fertilizer and lime expense.

X<sub>4</sub> = Dollars machinery expense.

X<sub>5</sub> = Dollars operating capital.

X<sub>6</sub> = Dollars livestock inventory.

X<sub>7</sub> = Dollar value of real estate owned.

X<sub>8</sub> = Dollar value of real estate rented.

Y = Gross farm income.

Y intercept = -\$2,699.30.

R<sup>2</sup> = 0.90639.

should be much impressed by Knight's regression coefficients (Betas) shown in Table 1.

### Ruttan's Production Function Estimates

Unfortunately, economists seldom have data of the quality used in the Knight study. Instead, we must usually resort to secondary sources, such as the U.S. Census of Agriculture. These data may be available only on a county basis for all the variables that we want to incorporate into our model. Also, census data provide only very crude indicators of many important inputs, such as for labor or for machinery and equipment investment. Thus, the economist is usually immediately faced with a serious specification problem which is compounded by aggregation and errors-in-variables difficulties.

As an example of these difficulties, consider the pioneering research of Ruttan [30] in developing new methodology for estimating the economic demand for irrigated land acreage. In his prize-winning book for outstanding published research in 1966, Ruttan developed new methods for projecting future irrigated acreage, based upon the economic productivity of irrigated land rather than previous crude procedures of merely extrapolating past trends in irrigated acreages or of projecting the quantity of inputs "required" to produce some projected level of output. Thus, Ruttan's study was particularly significant as a framework for analysis.<sup>3/</sup>

Ruttan's basic model for analysis was a Cobb-Douglas type production function estimated from county data for the major water resource regions. County value of all farm products sold was assumed to be a linear function in logarithms of the following inputs:

---

<sup>3/</sup> Problems of empirical estimation encountered by Ruttan are used for illustrative purposes, and this use is not intended to disparage the significant contribution of his study. Ruttan was careful to note that the grossness of the secondary data in some instances put a considerable strain on the methodology [30, p. 3]. His empirical estimation problems are especially interesting because they are so typical of specification problems generally encountered by economists using secondary data sources.



$X_1$  = Number of family and hired workers.

$X_2$  = Value of implements and machinery.

$X_3$  = Value of livestock investment.

$X_4$  = Irrigated land.

$X_5$  = Non-irrigated cropland harvested.

$X_6$  = Current operating expenses.

Due to problems of multicollinearity and a limited number of observations, Ruttan used only a subset of the above variables for several regions. In the most extreme case, for seven counties making up the South Pacific region, only  $X_4$  and  $X_6$  were retained in the regression model. As pointed out by Hoch [13], quite a serious omission-of-variable specification bias would likely result from omitting the other variables. Hence, Hoch hypothesized that the marginal value products for irrigated land were probably overstated in those cases where variables were deleted. This hypothesis led me to make similar omissions from Knight's fairly complete model, to examine the effects on parameter estimates.

#### Effect of Omitting Variables from the Knight Model

To recast Knight's model into Ruttan's abbreviated South Pacific region model, Knight's variable  $X_2$ , dollars feed purchased and veterinary livestock expense, needs to be combined with his  $X_3$  variable, dollars seed, crop, fertilizer, and lime expense. Then the combined  $X_2$ - $X_3$  variable from Knight's model is roughly equivalent to Ruttan's  $X_6$ , current operating expenses.

Ruttan's remaining variable,  $X_4$ , irrigated land acreage, is not entirely comparable to Knight's land variables,  $X_7$  and  $X_8$ . Knight's variables are defined in terms of dollar value of real estate owned and real estate rented, respectively, in contrast to Ruttan's irrigated acres. However, in either case the land resource is the primary variable, with the land input being weighted by its current market value in the Knight model.

After combining variables  $X_2$  and  $X_3$  and deleting all other variables

except  $X_7$  and  $X_8$ , we have the following equation, with standard errors in parentheses below their respective coefficients:

$$(2) \quad \hat{Y} = 2604.6 + 1.5366 (X_2 + X_3) + 0.1751 X_7 + 0.0874 X_8$$

$$\quad \quad \quad (.0508) \quad \quad \quad (.0102) \quad \quad \quad (.00899)$$

$$R^2 = 0.79.$$

Note that the above model, with omitted variables, is still very highly significant overall by standard classical statistical tests. But a very drastic change has resulted in the coefficients of the two land variables,  $X_7$  and  $X_8$ , with almost a four-fold increase in the coefficient for owned land, and more than a four-fold increase for rented land! Most misleading of all is the apparent, but not real, precision of estimation of the parameters in the poorly specified model in (2). By ordinary statistical procedures, the 99 percent confidence limits of the coefficient for owned land represent a return per dollar of land of approximately 14.9 to 20.1 percent of the market value of the land. On the other hand, for the original well-specified Knight model, the 99 percent confidence interval is 2.1 percent to 6.9 percent, certainly a more plausible range. It is interesting to note that the 99 percent confidence intervals of the owned land coefficients for the two models do not overlap. A similar result holds for the rented land coefficients.

One would infer from the preceding analysis of Knight's data that Ruttan's estimates of the marginal value product of irrigated land in his more incomplete models would be positively biased. How large this bias would be is uncertain, since there were several important differences between the two models and the data. Ruttan's model was of the Cobb-Douglas type, whereas the results from the Knight data were for a linear model. Also, the Knight data were individual farm observations, averaged over a five-year period, whereas Ruttan utilized county data. Furthermore, the land inputs were measured differently and represented a different type of land input, irrigated versus dryland acreage. In fact, where Ruttan deleted variables, it had much less effect on his land coefficient than was the case for the Knight model. Whether this difference was due to the difference between the original models or due to data differences is uncertain at this point.

At this point it should be noted that I have implicitly assumed that Knight's model in Table 1 was the "complete" or "true" model, and that the smaller model in (2) is therefore incomplete or "improperly specified". Actually, of course, we never know with certainty what all the variables in the "true" model should be, or even the functional form of the "true" model.

This point is especially relevant from a non-Bayesian point of view which might consider Knight's model in Table 1 and the abbreviated model in (2) as both being equally valid before fitting, even though the two models later give markedly different coefficient estimates. However, from the point of view of economic theory, one should relate value of output to all "economic" inputs, at least insofar as it is possible to measure them. ("Economic" inputs refer to those inputs costing money to use, such as those in Table 1, as contrasted to those free or uncontrollable inputs affecting production, such as sunshine or rain. Of course, one might even wish to include the free or "non-economic" inputs, such as precipitation, if the inclusion of such variables would improve the accuracy of estimation for the coefficients of the "economic" inputs.)

Therefore, from the economic point of view, it is contended that the specification of Knight's model in Table 1 is better than for the abbreviated model in Equation (2), since Knight's model does, at least, include all measurable "economic" factors of production. Thus, it seems reasonable to expect that the positive contributions to gross farm income from the deleted inputs would be erroneously attributed to the retained variables of Equation (2)). Hence, the author believes, although he cannot prove, that the coefficients (called Betas) in Table 1 are much closer to the true parameter values than those of Equation (2). (Some statistical evidence in favor of Knight's complete model is also provided by the t values in Table 1 and by the highly significant reduction in  $R^2$  for Equation (2), as indicated by F test.)

The preceding results do show that omission of relevant variables from a linear regression model can cause great changes in the individual estimates of the remaining coefficients. However, some readers may object that the preceding example was an extreme case, and that possible omitted-variable bias may be important only when estimating production functions. To counter this argument, another example is presented for an outdoor recreation demand function.

### Omission of Distance from an Outdoor Recreation Demand Model

The most commonly employed approach for estimating outdoor recreation demand is based upon the pioneering research of Clawson [4]. This approach attempts to measure recreational benefits by first estimating the inverse relation between transfer costs and participation rates. If it is assumed that the population of consumers is essentially homogeneous from one distance zone to another, then it can be argued that consumption-transfer cost combinations trace out a demand relationship, since geographical location forces the more distant consumers to bear heavier travel and related transfer costs.

Although some have argued that the Clawson approach is inappropriate or needs modification for one or more reasons [7, 21, 36], the Clawson approach still remains as the basis for most estimations of outdoor recreational benefits. However, one serious difficulty with the traditional Clawson approach was the omission of the non-monetary effects of distance, which tends to underestimate the value of a particular outdoor attraction, as noted by Knetsch in 1963 [20], and recently reiterated by Cesario and Knetsch [3, p. 702].

If distance does have an important negative effect on participation rates, then the inclusion of a separate variable, such as travel time, would be needed. Thus, visit rate could be expressed as a function of both money and time cost. However, as pointed out by Cesario and Knetsch [3], the difficulty has been that travel costs in monetary outlay and time are usually highly correlated, making it very difficult to separate the effect of one from the other. Also, including average distance or time traveled has tended to inflate the variance of the transfer cost variable. Therefore, most researchers have simply omitted variables such as hours or miles traveled, perhaps not being sufficiently concerned about the resulting specification bias.

In a recent study of the Oregon big game resource [2, 28], we wanted to estimate the net economic value of the resource to hunters. For the Northeast region of Oregon, a participation rate-transfer cost relationship was estimated from 31 distance zone averages:

$$(3) \quad \hat{Y}_j = 2.4141 - 0.008712 X_{1j} - 0.007943 X_{2j}$$

$$\quad \quad \quad (-1.252) \quad \quad (-3.749)$$

$$R^2 = 0.604.$$

In the preceding equation,  $\hat{Y}_j$  denotes the predicted average number of hunter unit trips to be taken per hunting family in the  $j^{\text{th}}$  distance zone,  $X_{1j}$  denotes the average variable transfer cost per hunter trip in the  $j^{\text{th}}$  distance zone, and  $X_{2j}$  denotes the distance in miles of the  $j^{\text{th}}$  distance zone to the Northeast Oregon hunting region. Numbers in parentheses below the coefficients are t values.

In Equation (3), the coefficient for the important transfer cost variable is not significant, even at the 20 percent level. Given this situation, that is, an unreliable statistical estimate of the important transfer cost variable, in the past many researchers have deleted the distance variable in order to try to "better" estimate the effect of the important transfer costs. If one follows this procedure, one obtains an estimate of  $\beta_1$  which appears to be more accurate, but which might be dangerously misleading, as in (4):

$$(4) \quad \hat{Y}_j = 2.1458 - 0.026836 X_{1j}$$

$$\quad \quad \quad (-4.452)$$

$$R^2 = 0.406.$$

Despite the apparent (but not real) accuracy of the estimation of the effect of transfer costs, the estimate of  $\beta_1$  is now increased by a factor of about three! Since the decline in participation rate in (4) is now all attributed to the transfer cost variable rather than to both distance and transfer costs, one should suspect a serious upward bias, due to the omission of the important distance variable. In fact, based upon a more thorough analysis utilizing all the individual observations, an estimate of  $\beta_1$  equal to about -0.009 was obtained [2, p. 71]. Thus, if the distance variable were omitted, the absolute value of  $\beta_1$  would be overestimated by a factor of about three. Similarly, net economic value based upon (4) would be underestimated by a factor of 3 [28, pp. 121-124].

The two preceding empirical examples from production and demand function parameter estimation illustrate the danger of omitted-variable bias.<sup>4/</sup> No doubt many readers could provide other instances where possible omitted-variable bias was equal to or greater than for the examples given here. However, if the seriousness of the potential bias is accepted, how can the researcher best avoid omitted-variable bias?

Suggestions for Alleviating Problems  
of Multicollinearity and Omitted-Variable Bias

Of course, relevant variables are not usually deliberately omitted from economic models. One usually starts with a model that is as completely specified as possible. However, when attempting to estimate the complete model by ordinary least squares (OLS), or closely related simultaneous equation methods or generalized least squares (GLS), one usually runs into severe problems of estimation, due to multicollinearity.

High correlation among explanatory variables in economic models should be expected because of the aggregated or averaged nature of economic data. The underlying logic for increased correlation between grouped or aggregated variables was shown in 1964 by Cramer [5]. However, research on this topic was primarily motivated by an interest in measuring the efficiency of Engel curve estimation where observations were grouped by various methods. Possibly for this reason, the important implications of increased correlation for grouped variables for the case of estimation of relationships involving many variables seems to have been either overlooked or ignored, until noted by Brown and Nawas [2a, pp. 344-359].

At any rate, given aggregated data, what should be done? Ideally, one would like to go back to the original individual observations, which would reduce intercorrelation and give more accurate estimates [2, 28]. However, this procedure is often impossible when working with secondary data, such as census

---

<sup>4/</sup> Of course, if the objective is to merely predict the dependent variable, accurate estimation of individual parameters is of less concern. However, note that intelligent decisions in economic policy almost always require good estimates of the relevant economic parameters.

data, where the anonymity of the respondents is protected. In other cases, the way the data are collected and recorded may also preclude the use of individual observations.

Farrar and Glauber [8] suggest an analysis to detect those explanatory variables most affected by multicollinearity by computing the  $r^{ii}$ , which correspond to the main diagonal elements of the inverted correlation matrix of the explanatory variables. Inspection of the  $r^{ii}$  will indicate the variables for which additional data are most needed to break the multicollinearity deadlock.

Various methods for incorporating prior information into regression analysis can be used to ameliorate the difficulties of multicollinearity. Even if one has only a subjective probability estimate for given parameter(s), this information can often be used to reduce multicollinearity by following the approach suggested by Theil [32] in 1963. Similarly, prior knowledge about the permissible bounds for a coefficient (e.g., that a probability must take a value,  $0 \leq p_i \leq 1$ ) can be used to form inequality constraints which can significantly lessen difficulties of multicollinearity [18].

Another promising approach for the estimation of models with high inter-correlation among the explanatory variables is through the use of so-called "ridge regression", which is, in effect, another way of incorporating prior information into the estimating procedure [15, 16, 26]. Potential gains from, and limitations of, ridge regression for estimation of economic models are explored mathematically and by means of some simple Monte Carlo experiments in the following sections.

#### Definition and Variance of the Ridge Regression Estimator

The basic idea of ridge regression is quite simple. Multicollinearity is mitigated by augmenting the main diagonal elements of the correlation matrix by small positive quantities [15, 16] or, as to be shown, one can increase all the main diagonal elements of the mean-corrected sums of squares and cross-product matrix by the same percentage to obtain the same result in terms of the regular (non-standardized) ridge regression coefficients. However, first following Hoerl

and Kennard [15], let  $X'X$  represent the correlation matrix of the explanatory variables. Then, assume the linear model,

$$(5) \quad Y = X\beta + u,$$

where  $Y$  is  $n \times 1$ , fixed  $X$  is  $n \times p$ ,  $\beta$  is  $p \times 1$ ,  $u$  is  $n \times 1$ ,  $E u = 0$ , and  $E u u' = V$  is positive definite. The ridge estimator,  $\hat{\beta}^*$ , is defined as

$$(6) \quad \hat{\beta}^* = (X'X + kI)^{-1} X'Y.$$

Thus,

$$(7) \quad \hat{\beta}^* = \beta + \gamma + (X'X + kI)^{-1} X'u$$

by substituting for  $Y$  from (5) where  $\gamma$  represents the bias in  $\hat{\beta}^*$ . By definition [17, p. 125], the variance-covariance matrix for  $\hat{\beta}^*$  is:

$$(8) \quad \begin{aligned} \text{Var-Cov } (\hat{\beta}^*) &= E[(\hat{\beta}^* - E\hat{\beta}^*)(\hat{\beta}^* - E\hat{\beta}^*)'] \\ &= E\{[(X'X + kI)^{-1} X'u][(X'X + kI)^{-1} X'u]'\} \\ &= (X'X + kI)^{-1} X'VX (X'X + kI)^{-1}. \end{aligned}$$

If we further assume that we have a homogeneous, nonautocorrelated error term, then

$$(9) \quad \text{Var-Cov } (\hat{\beta}^*) = \sigma^2 (X'X + kI)^{-1} X'X (X'X + kI)^{-1}.$$

For the experiments presented later, it was much easier to interpret and summarize the results by working with the mean-corrected sums of squares and cross-products, rather than the correlations. In terms of the corrected sums of squares and cross-products, one can estimate the non-standardized ridge regression coefficients as

$$(10) \quad \hat{\beta}^* = (X'X + k\lambda)^{-1} X'Y$$

where  $X'X$  represents the  $p \times p$  matrix of mean-corrected sums of squares and cross-products, and  $\lambda$  is a diagonal matrix of order  $p$  consisting of the sums



of squares.<sup>5/</sup> Then, the equivalent of (9) for the regular (non-standardized) ridge regression coefficients is:

$$(9a) \quad \text{Var-Cov } (\hat{\beta}^*) = \sigma^2 (X'X + k\lambda)^{-1} X'X (X'X + k\lambda)^{-1}.$$

The variances as defined by (9a) are the appropriate ones for the simple experiments to be reported in terms of the regular ridge regression coefficients. However, before the experimental results can be properly interpreted, the expected bias of the ridge estimator should be reviewed.

Expected Bias of the Ridge Regression Estimator in  
Terms of the True  $\beta$  Values

Although Hoerl and Kennard [12, p. 60] derive the proper expression for the bias squared of the standardized ridge estimator in terms of  $k$  and the true standardized  $\beta$  values, a more elementary derivation, in terms of the non-standardized  $\beta$  values, is presented here.<sup>6/</sup> By assuming fixed  $X$  values in (10), substituting  $X\beta + u$  for  $Y$ , and taking expected values, we obtain

$$(11) \quad \begin{aligned} E\hat{\beta}^* &= (X'X + k\lambda)^{-1} (X'X) \beta = (X'X + k\lambda)^{-1} (X'X + k\lambda - k\lambda) \beta \quad \text{since } Eu_1=0 \\ &= (X'X + k\lambda)^{-1} (X'X + k\lambda) \beta - (X'X + k\lambda)^{-1} (k\lambda) \beta \\ &= \beta - k (X'X + k\lambda)^{-1} \lambda \beta. \end{aligned}$$

Letting  $A^* = (X'X + k\lambda)$ ,

$$(12) \quad E(\hat{\beta}^* - \beta) = \frac{-k}{|A^*|} (\text{adj } A^*) \lambda \beta.$$

For any particular  $\hat{\beta}_i^*$ , and denoting the  $j, i^{\text{th}}$  element of the adjoint of  $A^*$  by  $c_{ji}$ ,

<sup>5/</sup> It should be emphasized that (10) gives exactly the same estimate of the non-standardized ridge coefficients as one obtains by multiplying the standardized estimates from (6) by the corresponding  $\sqrt{\sum y^2} / \sqrt{\sum x_j^2}$ .

<sup>6/</sup> The  $i^{\text{th}}$  observation of the standardized variable,  $X$ , is defined as

$$X_i = \frac{x_i - \bar{X}}{\sqrt{\sum x_i^2}}.$$

$$\begin{aligned}
 (13) \quad E(\hat{\beta}_j^* - \beta_j) &= \frac{-k}{|A^*|} [c_{j1} X_1' X_1 \beta_1 + c_{j2} X_2' X_2 \beta_2 \\
 &+ \dots + c_{jj} X_j' X_j \beta_j + \dots + c_{jp} X_p' X_p \beta_p] \\
 &= \frac{-k}{|A^*|} \sum_{i=1}^p c_{ji} X_i' X_i \beta_i.
 \end{aligned}$$

Of course, for standardized variables,  $X_i' X_i = 1$  for all  $i$ , and  $X'X$  can be ignored in (13). The significance of (13) can be more easily seen if we apply it to the simple case of two standardized explanatory variables. Then, for  $\hat{\beta}_1^*$ , Equation (13) reduces to

$$(14) \quad E(\hat{\beta}_1^* - \beta_1) = \frac{-k}{(1+k)^2 - r_{12}^2} [(1+k) \beta_1 - r_{12} \beta_2]$$

where  $r_{12}$  denotes the correlation between the two explanatory variables,  $X_1$  and  $X_2$ , and  $k$  denotes the amount of increment of the main diagonal elements of the correlation matrix.

Note that in Equation (14), for the usual case of high positive correlation between economic variables, the expected bias in  $\hat{\beta}_1^*$  will be lessened if the true  $\beta$  values have the same sign, and even more so if they are also of about equal magnitude. Conversely, if  $\beta_1$  and  $\beta_2$  are of opposite signs (and  $X_1$  and  $X_2$  are positively correlated), then the expected bias in  $\hat{\beta}_1^*$  will be greatly increased, as indicated by Equation (14). This result can be generalized to cases of more than two explanatory variables, as shown later, and is of great importance in deciding whether or not to use ridge regression for a particular problem. The crucial importance of similar sign and similar (standardized) magnitude for models with positively correlated explanatory variables is next illustrated by some simple experiments. (Of course, if  $r_{12}$  in (14) were negative, then bias squared and mean square error will be smallest when  $\beta_1$  and  $\beta_2$  are about equal in absolute magnitude but differ in sign.)<sup>7/</sup>

<sup>7/</sup> By definition, the mean square error (MSE) of an estimated parameter,  $\hat{\theta}$ , is  $E(\hat{\theta} - \theta)^2$ , which implies  $MSE(\hat{\theta}) = \text{Variance}(\hat{\theta}) + \text{square of bias}(\hat{\theta})$ .

Experimental Results for Two-Explanatory-Variables

The surprising estimating power of ridge regression for nonorthogonal data can be illustrated by the experimental results from the simple two-explanatory-variable model,

$$(15) \quad Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

where  $E(u_i) = 0.5(4) + 0.5(-4)$  and  $E(u_i u_{i+s}) = 0$  for all  $s \neq 0$ . The five fixed values for  $X_{1i}$  and  $X_{2i}$  were, respectively, (1,1.2), (2,1.67335), (3,3), (4,4.32665), and (5,4.8). By taking all possible samples ( $2^5 = 32$ ) generated by the simple binomial error term, the distribution of the OLS estimates of  $\beta_1$  and  $\beta_2$  were computed and the results summarized in the third column,  $k=0$ , of Table 2. The OLS estimates are quite variable, ranging from -10.657 to 18.657 for the OLS estimates of  $\beta_1$ , and with a similar range for the OLS estimates of  $\beta_2$ . Although exactly the true value of  $\beta_1 = 4$  is obtained by OLS 4/16 of the time, 5/16 of the time one does not even obtain the proper sign for  $\hat{\beta}_1$ . Similarly, the true value of  $\beta_2 = 4$  is also obtained 4/16 of the time by OLS, but negative values occur 6/16 = 37.5 percent of the time. This erratic behavior of the OLS estimated coefficients is reflected in the high variances for  $\hat{\beta}_1$  and  $\hat{\beta}_2$  in the lower part of Table 2. Since the OLS estimates are unbiased, the mean square error in the last two lines of Table 2 is equal to the variance for OLS.

The main reason for the large variances of the OLS estimates in Table 2 is, of course, the high intercorrelation between  $X_1$  and  $X_2$ ,  $r_{12} = 0.98533$ . If  $X_1$  and  $X_2$  were orthogonal, the variance of  $\hat{\beta}_1$  would be  $\sigma^2 \div \sum x_{1i}^2 = 16/10$ . Since  $X_1$  and  $X_2$  are correlated, the variance is increased by a factor of  $1/(1 - r_{12}^2) \doteq 34.3$ . Similarly, the variance of  $\hat{\beta}_2$  is increased by the same amount.

The ridge regression estimates in Table 2 show a dramatic reduction in variability, compared to OLS. For  $k = 0.1$ , the variance of  $\hat{\beta}_1^*$  is 1.258, only slightly over 2 percent of that for  $\hat{\beta}_1$ ; the variance of  $\hat{\beta}_2^*$  is similarly reduced. However, more important than the size of the variance itself is the dramatic reduction in the mean square error (MSE) for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$ . For  $k = 0.25$ , MSE for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  is less than 1.25 percent that for OLS. With such great improvement in accuracy

Table 2. Distribution of Estimated  $\beta_1$  and  $\beta_2$  Coefficients for OLS Versus Ridge Regression  
 Where  $Y_i = \alpha + 4X_{1i} + 4X_{2i} + u_i$ ;  $E(u_i) = 0.5(4) + 0.5(-4)$ ; and Fixed  $X_1$  and  $X_2$   
 Take the 5 Values (1,1.2), (2,1.67335), (3,3), (4,4.32665), and (5,4.8)

Probability	Variable number	Distribution of $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ for several levels of k						
		k=0 (OLS)	k=0.05	k=0.10	k=0.15	k=0.20	k=0.25	k=0.30
1/16	1	-10.657	0.355	1.688	2.164	2.383	2.493	2.548
	2	18.063	6.870	5.363	4.722	4.346	4.085	3.886
2/16	1	-4.438	2.339	3.115	3.361	3.451	3.475	3.467
	2	13.375	6.379	5.394	4.948	4.668	4.463	4.297
2/16	1	-2.219	1.918	2.382	2.521	2.566	2.570	2.556
	2	8.688	4.392	3.777	3.493	3.311	3.175	3.064
1/16	1	1.781	4.322	4.542	4.559	4.519	4.458	4.386
	2	8.688	5.889	5.425	5.174	4.991	4.840	4.708
4/16	1	4.000	3.902	3.808	3.719	3.634	3.553	3.475
	2	4.000	3.902	3.808	3.719	3.634	3.553	3.475
1/16	1	6.219	3.481	3.075	2.879	2.748	2.648	2.564
	2	-0.688	1.914	2.191	2.264	2.277	2.265	2.242
2/16	1	10.219	5.886	5.235	4.917	4.702	4.535	4.394
	2	-0.688	3.411	3.839	3.945	3.957	3.930	3.886
2/16	1	12.438	5.465	4.501	4.077	3.817	3.630	3.483
	2	-5.375	1.424	2.222	2.490	2.599	2.643	2.652
1/16	1	18.657	7.449	5.928	5.274	4.885	4.612	4.403
	2	-10.063	0.937	2.254	2.716	2.922	3.020	3.063
$E(\hat{\beta}_1^*)$		4.000	3.902	3.808	3.719	3.634	3.553	3.475
$E(\hat{\beta}_2^*)$		4.000	3.902	3.808	3.719	3.634	3.553	3.475
$E \text{ Var}(\hat{\beta}_1^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ Var}(\hat{\beta}_2^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ MSE}(\hat{\beta}_1^*)$		54.936	3.199	1.295	0.860	0.721	0.686	0.698
$E \text{ MSE}(\hat{\beta}_2^*)$		54.936	3.199	1.295	0.860	0.721	0.686	0.698

of estimation from ridge regression in Table 2, one may well wonder if OLS regression is not obsolete for most economic research where one must work with non-orthogonal data. However, it should be remembered from the earlier discussion of Equations (13) and (14) that for high positive correlation between two explanatory variables, bias is increased if the true  $\beta$  values are of unequal magnitude or of different sign.

Before leaving the results in Table 2, it should be noted that the use of the binomial error term in combination with only five observations can be criticized, since it permits only nine different values of  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  in Table 2 and the other experiments. Thus, the use of a normally distributed error term might have been preferable for generating a greater range of values for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$ . Alternatively, the number of outcomes with the binomial error could have been increased by increasing the number of observations from five to, say, seven. However, despite these limitations, it is thought that a fairly good range of experimental values were obtained, as shown in Table 2. Some idea of the distributional pattern for  $\hat{\beta}_1$  can be obtained from Figure 1, where the cumulative probability for the various observed OLS estimates of  $\beta_1$  are presented. As would be expected, the distribution of  $\hat{\beta}_1$  in Figure 1 is not too far from the cumulative normal although, no doubt, a closer approximation could have been obtained by using a normally distributed error term with many samples. However, the expected mean square error for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  would have been exactly the same, since the variance and bias of the ridge estimates do not depend upon how the error term is distributed, as can be seen from Equations (9a) and (13).<sup>8/</sup> Also, the binomial error term was extremely convenient, since only a few samples needed to be considered per experiment.

It should also be noted that the estimates of  $\alpha$  are not included in Tables 2-5. The ridge estimates of  $\beta_1$  and  $\beta_2$  were obtained by solving the following normal equations in terms of the mean-corrected sums of squares and cross-products:

<sup>8/</sup> Note also from (9) or (9a) and (13) that the sample size,  $n=5$ , does not affect the ratio of mean square error of OLS to ridge regression.

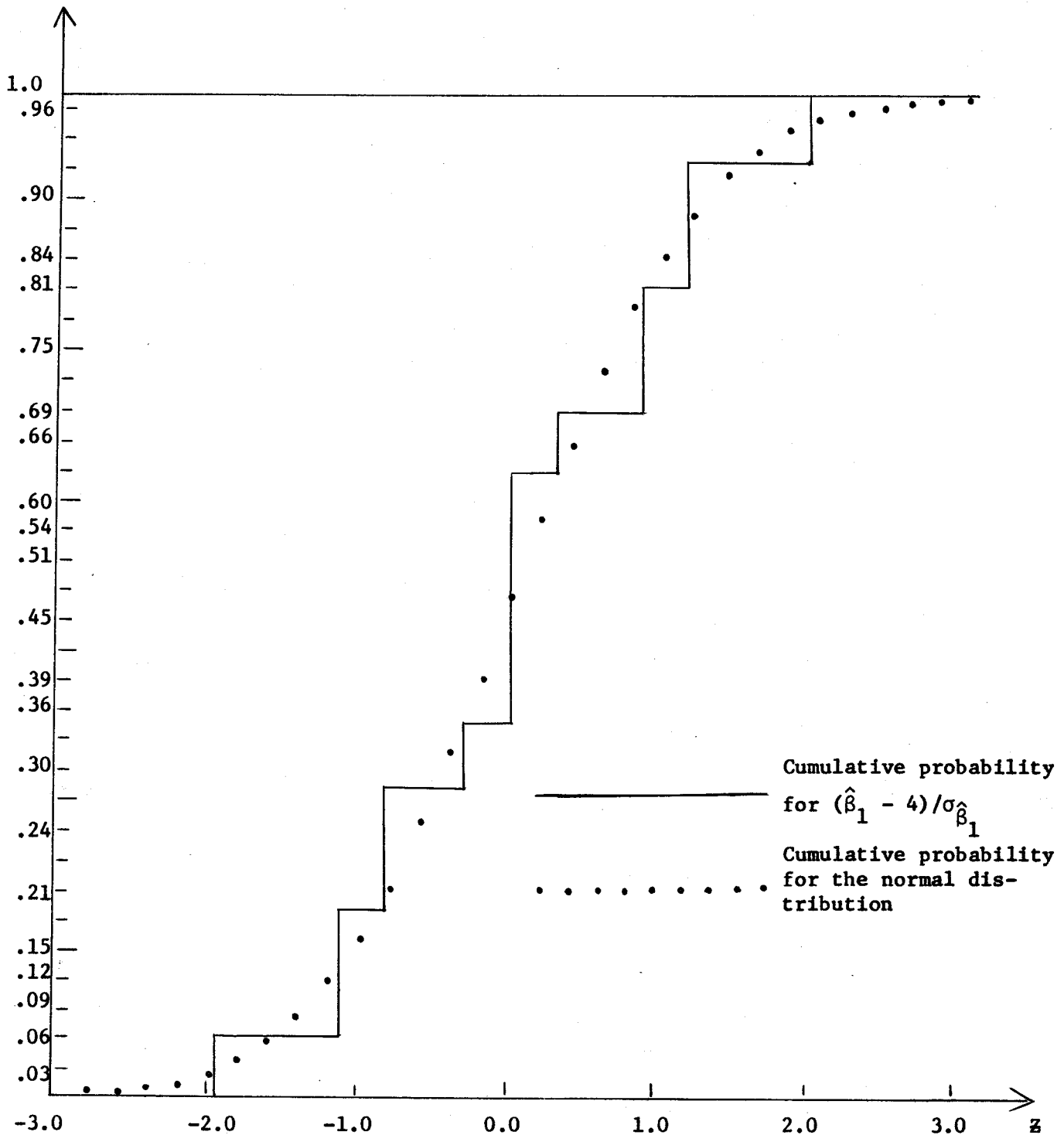


Figure 1. Cumulative probability of  $Z = (\hat{\beta}_1 - 4)/\hat{\sigma}_{\hat{\beta}_1}$  from the model,

$Y_i = \alpha + 4X_{1i} + 4X_{2i} + u_i$ ;  $E(u_i) = 0.5(4) + 0.5(-4)$ ; and fixed  $X_1$  and  $X_2$  take the 5 values (1,1.2), (2,1.67335), (3,3), (4,4.32665), and (5,4.8).

$$(1+k) \sum x_1^2 \hat{\beta}_1^* + \sum x_1 x_2 \hat{\beta}_2^* = \sum x_1 y$$

$$\sum x_1 x_2 \hat{\beta}_1^* + (1+k) \sum x_2^2 \hat{\beta}_2^* = \sum x_2 y$$

Although apparently not treated in the literature, one could define  $\alpha^* = \bar{Y} - \hat{\beta}_1^* \bar{X}_1 - \hat{\beta}_2^* \bar{X}_2$ . Then, one can derive

$$\text{var}(\alpha^*) = \frac{\sigma^2}{n} + \bar{X}_1^2 \text{var}(\hat{\beta}_1^*) + \bar{X}_2^2 \text{var}(\hat{\beta}_2^*) + 2\bar{X}_1 \bar{X}_2 \text{cov}(\hat{\beta}_1^*, \hat{\beta}_2^*).$$

Using the above formula for  $k=0$  in Table 2, we obtain:

$$\text{var}(\hat{\alpha}) = 16/5 + 9(54.936) + 9(54.936) + 2(9)(-54.130) \doteq 17.71.$$

However, for  $k=0.25$ , we obtain

$$\text{var}(\alpha^*) = 16/5 + 9(0.4854) + 9(0.4854) + 2(9)(0.150325) \doteq 14.64.$$

Although we obtain a lower variance for the preceding estimate of  $\alpha$  at  $k=0.25$ , it should be noted that mean square error for  $\alpha^*$  is not reduced. For  $k=0.25$ , we have  $E(\alpha^*) = [E(Y) - E(\hat{\beta}_1^*) \bar{X}_1 - E(\hat{\beta}_2^*) \bar{X}_2] = [18 - 3.553(3) - 3.553(3)] \doteq 3.318$ . Thus, the bias of  $\alpha^*$  is  $-3.318 - (-6) \doteq 2.682$ . Therefore,  $\text{MSE}(\alpha^*) \doteq 14.64 + 7.19 \doteq 21.8$ . Thus, mean square error of  $\alpha^*$  at  $k=0.25$  is greater than for OLS. Even for  $k=0.05$ , only a slight reduction in mean square error relative to OLS is obtained. The relative lack of improvement in  $\alpha^*$  as compared to OLS  $\hat{\alpha}$  seems puzzling until one remembers that  $\alpha$  is the coefficient for a variable consisting entirely of 1.0's. Hence, its correlation with any of the other variables is identically zero and, therefore, perhaps little improvement over OLS should really be expected. However, further investigation of various methods of estimating  $\alpha$  appears to be needed.

In Table 3, results from another model are presented, where the true  $\beta$  values are  $\beta_1 = 5.33$  and  $\beta_2 = 2.665$ . These particular numerical values were selected so as to have  $\beta_1 = 2\beta_2$ , and to have approximately the same  $R^2$  as for the case of equal  $\beta_1$  and  $\beta_2$  in Table 2. Thus, the overall significance of the regression equation was maintained at about the same level for all the experiments, with

Table 3. Distribution of Estimated  $\beta_1$  and  $\beta_2$  Coefficients for OLS Versus Ridge Regression  
 Where  $Y_1 = \alpha + 5.33X_{11} + 2.665X_{21} + u_1$ ;  $E(u_1) = 0.5(4) + 0.5(-4)$ ; and Fixed  $X_1$  and  
 $X_2$  Take the 5 Values (1,1.2), (2,1.67335), (3,3), (4,4.32665), and (5,4.8)

Probability	Variable number	Distribution of $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ for several levels of k						
		k=0 (OLS)	k=0.05	k=0.10	k=0.15	k=0.20	k=0.25	k=0.30
1/16	1	19.987	7.749	6.096	5.391	4.974	4.684	4.462
	2	-11.398	0.624	2.081	2.595	2.829	2.944	2.999
2/16	1	13.768	5.765	4.669	4.193	3.906	3.702	3.543
	2	-6.710	1.119	2.050	2.369	2.506	2.567	2.588
2/16	1	11.549	6.185	5.403	5.033	4.791	4.607	4.454
	2	-2.023	3.107	3.667	3.824	3.863	3.854	3.822
1/16	1	7.549	3.781	3.243	2.995	2.837	2.719	2.624
	2	-2.023	1.610	2.018	2.143	2.183	2.189	2.177
4/16	1	5.330	4.202	3.976	3.835	3.723	3.624	3.535
	2	2.665	3.597	3.635	3.598	3.541	3.477	3.411
1/16	1	3.111	4.622	4.710	4.675	4.608	4.529	4.446
	2	7.353	5.584	5.252	5.053	4.898	4.764	4.644
2/16	1	-0.889	2.218	2.550	2.638	2.654	2.642	2.616
	2	7.353	4.087	3.604	3.372	3.218	3.099	3.000
2/16	1	-3.108	2.638	3.283	3.478	3.540	3.547	3.527
	2	12.040	6.075	5.221	4.827	4.575	4.387	4.233
1/16	1	-9.327	0.655	1.857	2.280	2.472	2.565	2.608
	2	16.728	6.565	5.190	4.601	4.252	4.009	3.822
$E(\hat{\beta}_1^*)$		5.330	4.202	3.976	3.835	3.723	3.624	3.535
$E(\hat{\beta}_2^*)$		2.665	3.597	3.635	3.598	3.541	3.477	3.411
$E \text{ Var}(\hat{\beta}_1^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ Var}(\hat{\beta}_2^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ MSE}(\hat{\beta}_1^*)$		54.936	4.463	3.090	3.015	3.171	3.395	3.645
$E \text{ MSE}(\hat{\beta}_2^*)$		54.936	4.058	2.199	1.652	1.354	1.144	0.979



the true  $\beta$  values always selected to give an expected sum of squares for the dependent variable of about 699.3056, by using

$$E(\sum y_i^2) = [\sum (E y_i)^2 + \sum E(u_i - \bar{u})^2] = [\beta_1^2 \sum x_1^2 + 2\beta_1\beta_2 \sum x_1 x_2 + \beta_2^2 \sum x_2^2 + (n-1)\sigma^2] \\ \doteq (635.3056 + 64) = 699.3056.$$

As should be expected from earlier Equations (13) and (14), the reduction in the total mean square error for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  is somewhat less in Table 3 than for Table 2. The sum of MSE ( $\hat{\beta}_1^*$ ) + MSE ( $\hat{\beta}_2^*$ ) = 4.525 for  $k = 0.20$  in Table 3, about 4.12 percent that for OLS. Although this reduction in MSE is still considerable, it is less than for Table 2, where MSE for  $k = 0.25$  was less than 1.25 percent that for OLS.

Although mean square error is still much reduced by ridge regression in Table 3, the bias is fairly substantial. For example, for  $k = 0.20$ , the expected value of  $\hat{\beta}_1^* = 3.723$  in the lower part of Table 3 implies a percentage bias of  $(3.723 - 5.33) / 5.33 \doteq 30$  percent. For  $\hat{\beta}_2^*$  and  $k = 0.20$ , the percentage bias is even worse, being equal to over 70 percent of the true  $\beta_2$  value of 2.665. Thus, it can be seen that for even substantial reduction in mean square error, as in Table 3, the problem of bias can become serious, at least for economic analysis where parameter estimates are to be used for economic policy.

The third experiment performed followed exactly the same format as for those shown in Tables 2 and 3. The true  $\beta$  values selected were  $\beta_1 = 6.6550$  and  $\beta_2 = 1.3310$ . These values again yield approximately the same significance of the overall regression and the same OLS variance estimates. For the sake of brevity, the results for this third experiment are not presented in table form. The results, however, followed the trend started in Table 3; that is, ridge regression considerably underestimated  $\beta_1$  and overestimated  $\beta_2$  in all cases for  $k \geq 0.10$ . The sum of mean square error for  $\beta_1$  and  $\beta_2$  was about 13.365 at its lowest point for  $k = 0.10$  (using increments of 0.05 for  $k$ ). Therefore, MSE ( $\hat{\beta}_1^*$ ) + MSE ( $\hat{\beta}_2^*$ ) was about 12.2 percent that for OLS at  $k = 0.10$ . However, accompanying this gratifying reduction in mean square error from ridge reduction was an underestimate of  $\beta_1$ ,  $\hat{\beta}_1^* \doteq 4.142$ , only about 62 percent of the

true value. Even worse was the overestimate of  $\beta_2$ ,  $\hat{\beta}_2^* \doteq 3.461$ , about 2.6 times the true value. However, one must remember that the OLS estimates were also very poor.

The substantial bias possible from ridge regression is illustrated even more strongly by the results from the fourth experiment, presented in Table 4. For the model in Table 4, exactly the same explanatory variables and error terms were used as for the models in Tables 2 and 3, but the values  $\beta_1 = 7.9706$  and  $\beta_2 = 0.0$  were used.

Danger of being misled by ridge regression, if one does not have prior information regarding the true  $\beta$  values, is illustrated by the results of Table 4. Suppose, for example, that one thought that the expected mean square error was only 0.721 at  $k = 0.20$ , similar to that for the first model of Table 2. Then, one would be misled into thinking that  $\beta_2$  was significantly different from zero since  $\hat{\beta}_2^*$  ranges from 1.99 to over 4.0 in Table 4 for  $k = 0.20$ . In fact, of course, the estimated variance for  $\hat{\beta}_2^*$  varies in Table 4, depending upon the estimate of  $\sigma^2$ , which changes over the 16 distinct samples. If one computes the ratio of  $\hat{\beta}_2^*$  to the square root of its variance at  $k = 0.20$  for each of the 16 distinct samples making up Table 4, then these ratios range from 2.55, 2.82, 3.33, ..., 9.71 (excluding the one distinct sample where the estimate of  $\sigma^2$  is zero).

It is clear from the results of Table 4 that one can be in danger of making a Type I error when using ridge regression for the type of model of Table 4. That is, one would tend to reject the null hypothesis that  $\beta_2 = 0$  when, in fact,  $\beta_2$  were equal to zero - unless one had good prior information about the true  $\beta_1$  and  $\beta_2$  values from which one could estimate the bias likely to be involved. This serious problem involved in trying to make inferences about the true parameter values will be examined in more detail in the next section, Rules for Selection of Optimal k Values. However, before proceeding with that section, results from a fifth experiment, shown in Table 5, should be noted.

For the fifth experiment, values of  $\beta_1 = 9.918$  and  $\beta_2 = -1.9836$  were assigned. Again, the same explanatory variables and error terms were used as

Table 4. Distribution of Estimated  $\beta_1$  and  $\beta_2$  Coefficients for OLS Versus Ridge Regression

Where  $Y_1 = \alpha + 7.9706X_{11} + 0.0X_{21}$ ;  $E(u_1) = 0.5(4) + 0.5(-4)$ ; and Fixed  $X_1$  and  $X_2$  Take the 5 Values (1,1.2), (2,1.67335), (3,3), (4,4.32665), and (5,4.8)

Probability	Variable number	Distribution of $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ for several levels of k						
		k=0 (OLS)	k=0.05	k=0.10	k=0.15	k=0.20	k=0.25	k=0.30
1/16	1	22.627	8.338	6.424	5.616	5.144	4.820	4.575
	2	-14.063	0.015	1.730	2.347	2.637	2.786	2.865
2/16	1	16.408	6.355	4.997	4.418	4.076	3.838	3.656
	2	-9.375	0.506	1.699	2.121	2.314	2.409	2.454
2/16	1	14.190	6.775	5.731	5.258	4.961	4.743	4.567
	2	-4.688	2.493	3.316	3.576	3.671	3.696	3.687
1/16	1	10.190	4.371	3.571	3.220	3.007	2.856	2.737
	2	-4.688	0.996	1.667	1.895	1.991	2.031	2.043
4/16	1	7.971	4.791	4.304	4.060	3.893	3.760	3.648
	2	0.000	2.983	3.284	3.350	3.348	3.319	3.276
1/16	1	5.752	5.212	5.037	4.900	4.778	4.665	4.559
	2	4.688	4.971	4.901	4.806	4.706	4.606	4.510
2/16	1	1.752	2.808	2.877	2.863	2.825	2.778	2.729
	2	4.688	3.474	3.253	3.124	3.025	2.941	2.865
2/16	1	-0.467	3.228	3.611	3.703	3.710	3.683	3.640
	2	9.375	5.461	4.870	4.580	4.383	4.229	4.099
1/16	1	-6.686	1.244	2.184	2.505	2.642	2.701	2.721
	2	14.063	5.951	4.839	4.354	4.060	3.851	3.688
$E(\hat{\beta}_1^*)$		7.971	4.791	4.304	4.060	3.893	3.760	3.648
$E(\hat{\beta}_2^*)$		0.000	2.983	3.284	3.350	3.348	3.319	3.271
$E \text{ Var}(\hat{\beta}_1^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ Var}(\hat{\beta}_2^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ MSE}(\hat{\beta}_1^*)$		54.936	13.297	14.701	16.071	17.215	18.211	19.108
$E \text{ MSE}(\hat{\beta}_2^*)$		54.936	12.090	12.045	12.006	11.798	11.499	11.095

Table 5. Distribution of Estimated  $\beta_1$  and  $\beta_2$  Coefficients for OLS Versus Ridge Regression  
 Where  $Y_i = \alpha + 9.918X_{1i} - 1.9836X_{2i} + u_i$ ;  $E(u_i) = 0.5(4) + 0.5(-4)$ ; and Fixed  $X_1$  and  
 $X_2$  Take the 5 Values (1,1.2), (2,1.67335), (3,3), (4,4.32665), and (5,4.8)

Probability	Variable number	Distribution of $\hat{\beta}_1^*$ and $\hat{\beta}_2^*$ for several levels of k						
		k=0 (OLS)	k=0.05	k=0.10	k=0.15	k=0.20	k=0.25	k=0.30
1/16	1	24.575	8.767	6.658	5.774	5.262	4.913	4.651
	2	-16.047	-0.448	1.461	2.155	2.486	2.661	2.757
2/16	1	18.356	6.783	5.231	4.576	4.194	3.931	3.732
	2	-11.359	0.042	1.430	1.929	2.163	2.284	2.347
2/16	1	16.137	7.204	5.965	5.416	5.079	4.836	4.643
	2	-6.671	2.030	3.047	3.384	3.520	3.571	3.580
1/16	1	12.137	4.799	3.805	3.379	3.125	2.948	2.812
	2	-6.671	0.532	1.399	1.703	1.840	1.906	1.936
4/16	1	9.918	5.220	4.538	4.219	4.011	3.853	3.724
	2	-1.984	2.520	3.016	3.158	3.197	3.194	3.169
1/16	1	7.699	5.640	5.272	5.059	4.896	4.758	4.635
	2	2.704	4.507	4.633	4.614	4.555	4.481	4.402
2/16	1	3.699	3.236	3.112	3.021	2.943	2.871	2.804
	2	2.704	3.010	2.984	2.932	2.875	2.816	2.758
2/16	1	1.480	3.656	3.845	3.861	3.828	3.776	3.716
	2	7.392	4.998	4.601	4.388	4.232	4.104	3.991
1/16	1	-4.739	1.673	2.419	2.663	2.760	2.794	2.796
	2	12.079	5.488	4.570	4.162	3.909	3.726	3.581
$E(\hat{\beta}_1^*)$		9.918	5.220	4.538	4.219	4.011	3.853	3.724
$E(\hat{\beta}_2^*)$		-1.984	2.520	3.016	3.158	3.197	3.194	3.169
$E \text{ Var}(\hat{\beta}_1^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ Var}(\hat{\beta}_2^*)$		54.936	3.190	1.258	0.781	0.587	0.485	0.423
$E \text{ MSE}(\hat{\beta}_1^*)$		54.936	25.264	30.199	33.264	35.482	37.265	38.790
$E \text{ MSE}(\hat{\beta}_2^*)$		54.936	23.471	26.250	27.221	27.431	27.290	26.972

preceding experiments. Also,  $\beta_1$  and  $\beta_2$  were selected such that  $\beta_1 = 5(-\beta_2)$  and such that the expected sum of squares for the dependent variable was again approximately equal to 699.3056.

Ridge regression results continue to worsen for the model presented in Table 5, as would be expected from Equation (14) for explanatory variables with high positive correlation and for true  $\beta$  coefficients with unlike signs. For  $k = 0.05$ , the sum of mean square error of  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  was reduced to about 44 percent that for the OLS estimates. However, this reduction was accompanied by a large bias. The estimated value of  $\beta_1$  was 5.220, only about 53 percent of the true value of 9.918. Even worse was the bias for  $\hat{\beta}_2^* = 2.520$  versus the true value of  $\beta_2 = -1.9836$ .

At this point it should be noted that the increments of the  $k$  values were too large for the model in Table 5, since the smallest mean square error may have occurred for some  $k$  value less than 0.05. To check this hypothesis, solutions were obtained for  $k = 0.00, 0.01, 0.02, \dots, 0.10$ . Smallest sum of mean square error for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  was observed at  $k = 0.02$ , where  $\text{MSE}(\hat{\beta}_1^*) + \text{MSE}(\hat{\beta}_2^*) = 43.89$ . Expected values at  $k = 0.02$  were  $E(\hat{\beta}_1^*) = 6.446$  and  $E(\hat{\beta}_2^*) = 1.410$ . Thus, a rather substantial bias is still encountered, the relative bias in  $\hat{\beta}_1^*$  being  $(6.446 - 9.918)/9.918$ , equal to about 35 percent. Of course, the relative bias for the ridge estimate of  $\beta_2$  is even worse, being about  $(1.410 + 1.9836)/1.9836$ , equal to about 171 percent of the true value. (Perhaps it should have been mentioned even earlier that the mean square error function for the ridge estimator will always have a unique minimum for some  $k > 0$ , due to the properties of the variance and squared bias functions, as shown by Hoerl and Kennard [15, pp. 61-63].)

Although the greatest danger in using ridge regression (when one does not have fairly good information about the true  $\beta$  values) would be that of making Type I statistical errors (that is, of concluding that some  $\beta$  values were statistically different from zero when, in fact, some of the  $\beta$  values were equal to zero), the possibility of making the opposite type of error (Type II) cannot be ruled out. Suppose, for example, that we have exactly the same explanatory variables and error term as before, but we select  $\beta_2 = -0.86008798 \beta_1$  and  $\beta_1, \beta_2$

such that there is the same overall level of significance for the total regression as for the preceding models with  $E(\sum y_i^2) = 699.3056$ . Solving, we obtain  $\beta_1 = 37.6532092$  and  $\beta_2 = -32.38507275$ . For this model, if we set  $k = 0.20$ , we obtain  $E(\hat{\beta}_2^*) = -0.0001$ . Thus, for  $k \geq 0.20$ , one could mistakenly conclude that  $\beta_2 = 0$ .

Actually, such a mistaken conclusion would be quite unlikely in the preceding case, since the OLS estimates of  $\beta_1$  and  $\beta_2$  would almost always have large t ratios. Therefore, the researcher would usually automatically reject k values large enough to cause the Type II error, at least when working with only two explanatory variables. Even for models with more explanatory variables, treated in a later section, Type II errors should be less likely than Type I errors.

One question that can be raised about the preceding two-explanatory-variable experiments is, "What would have happened if the  $\beta$  values had been in the same ratio, but larger?" The answer can be inferred from the earlier equations for the variance, (9a), and the bias, (13), for the regular (non-standardized)  $\beta$ 's. With increased values for the true  $\beta$  values in (13), the bias would be increased proportionally. At the same time, assuming the same error term, the variance for  $\hat{\beta}^*$  (and OLS) would remain unchanged. Thus, the advantage in mean square error from ridge regression compared to OLS would decrease, since the bias squared of  $\hat{\beta}^*$  would increase with the square of the true  $\beta$  values. Of course, if one had increased (decreased) correlation between  $X_1$  and  $X_2$  for the equations used in Tables 2-5, the relative advantage of ridge regression would increase (decrease).

Before considering implications for larger regression models, the question of how the k values should be selected is briefly considered (since the level of k has a marked influence on the size of the estimated parameters, as illustrated in Tables 2-5.)

#### Selection of Optimal k Values

Although it has been proven that there always exists a  $k > 0$  such that a smaller mean square error can be obtained from ridge regression than from ordinary least squares [15], the best method for selecting a particular value of

$k$  is not obvious. Although other approaches are mentioned, Hoerl and Kennard seem to place the most confidence in the use of the "Ridge Trace" [15, p. 65]:<sup>9/</sup>

"Based on experience, the best method for achieving a better estimate of  $\hat{\beta}^*$  is to use  $k_i = k$  for all  $i$  and use the Ridge Trace to select a single value of  $k$  and a unique  $\hat{\beta}^*$ ."

Hoerl and Kennard then indicate several considerations that can be used to guide one to a choice of a particular  $k$  value:

- (1) At a certain value of  $k$  the system will stabilize and have the general characteristics of an orthogonal system.
- (2) Coefficients will not have unreasonable absolute values with respect to the factors for which they represent rates of change.
- (3) Coefficients with apparently incorrect signs at  $k = 0$  will have changed to have the proper sign.
- (4) The residual sum of squares will not have been increased to an unreasonable value. It will not be large relative to the minimum residual sum of squares, or large relative to what would be a reasonable variance for the process generating the data.

In their second article [16], Hoerl and Kennard illustrate the use of the Ridge Trace by analyzing an empirical ten-factor regression model and a thirteen-factor model. Because of their heavy reliance upon the stabilization of the Ridge Trace (preceding consideration No. 1), the question arises as to whether one could formulate some decision rule as it relates to the stability of the ridge estimates.

One such rule could be defined as the following:

---

<sup>9/</sup> "Ridge Trace" denotes a simple graph of the values of the ridge estimates on the vertical axis plotted against the corresponding values of  $k$  on the horizontal axis.

RULE: Select a particular value of  $k$  at that point where the last ridge estimate attains its maximum absolute magnitude after having attained its "ultimate" sign, where "ultimate" sign is defined as being the sign at, say,  $k = 0.9$ .

To illustrate the use of the preceding rule, consider the results for the model in Table 2. For this model, the "ultimate" sign for both coefficients is positive for  $k = 0.9$  over all the possible samples presented in Table 2. Therefore, the rule reduces to finding the value of  $k$  for each sample such that the last ridge estimate has attained its maximum positive value. For the first sample in the top part of Table 2,  $\hat{\beta}_2^*$  declines throughout; therefore, attention should be focused on  $\hat{\beta}_1^*$ . The value of  $\hat{\beta}_1^*$  continues to increase in Table 2, reaching 2.548 at  $k = 0.30$ . However, the maximum value (by increments of 0.05) of  $\hat{\beta}_2^*$  is 3.59019 at  $k = 0.40$ . Using these values, the sum of mean square error for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  is  $(2.57415 - 4)^2 + (3.59019 - 4)^2 \doteq 2.2010$ . Similarly, for the second set of estimates from the top in Table 2,  $\hat{\beta}_1^*$  reaches its highest value at  $\hat{\beta}_1^* = 3.475$  at  $k = 0.25$ . Therefore, sum of mean square error for this sample situation (which occurs 2/16 of the time) would be  $(3.475 - 4)^2 + (4.463 - 4)^2$  which adds to about 0.489.

Following the above procedure for each sample situation, and weighting by the probabilities given in the left hand column of Table 2, an expected sum of mean square error for  $\hat{\beta}_1^*$  and  $\hat{\beta}_2^*$  of 1.3076 is obtained from following the preceding selection rule. Surprisingly, the selection rule gives a slightly better result than any single value of  $k$  listed in Table 2. Best result for any single  $k$  value given in Table 2 is for  $k = 0.25$ , with an expected sum of mean square error of  $2(0.686) \doteq 1.372$ .

Using the same selection rule for the model of Table 3, the ridge estimate of  $\beta_2$  for the first sample in the top part of Table 3 attains its highest positive value (by  $k$  increments of 0.05) at  $k = 0.35$ . For  $k = 0.35$ ,  $\hat{\beta}_1^* = 4.28202$  and  $\hat{\beta}_2^* = 3.01947$ , yielding a total mean square error of  $(4.28202 - 5.33)^2 + (3.01947 - 2.665)^2 \doteq 1.2239$ . Computing the mean square error in this way for the other samples, and weighting by the probabilities in the left hand column of Table 3, the expected mean square errors were as follows:



$$\text{MSE } (\hat{\beta}_1^*) \doteq 3.1629$$

$$\text{MSE } (\hat{\beta}_2^*) \doteq 1.3279$$

---


$$\text{Sum} \quad 4.4908$$

Again, as for Table 2, the  $k$  selection rule gave a surprisingly good result, with the sum of mean square error being slightly less than for any of the  $k$  values listed in Table 3. (The smallest sum of mean square error in Table 3 is 4.525 for  $k = 0.20$ ). Considering the fact that the selection rule utilized no prior information, the results of its use on the models of Tables 2 and 3 were encouraging. However, it is interesting that for models with greater bias resulting from ridge regression, such as Model 4 of Table 4, the results from using the preceding selection rule were also less satisfactory. Use of the selection rule for the model of Table 4 gave a sum of mean square error of around 30.1, about 18.6 percent higher than for the sum of mean square error of 25.4 for  $k = 0.05$  in Table 4.

In summary, the use of the preceding selection rule gives surprisingly good results for models which are "well-suited" to ridge regression, as discussed earlier with regard to expected bias of the ridge estimator in terms of the true  $\beta$  values. Similarly, use of the Ridge Trace also might give fairly good results for the models of Tables 2 and 3, although the ambiguity of the Ridge Trace for selecting  $k$  is a disadvantage. However, an important finding of this study is the fact that both the Ridge Trace and the preceding selection rule would give unreliable or poor results for models similar to those of Tables 4 and 5. Therefore, fairly good prior information about the true  $\beta$  values of the model appears necessary to be able to benefit from ridge regression.

#### Implications for Larger Models

The preceding conclusion about the necessity of good prior information for the use of ridge regression on the simple two-explanatory-variable model can be extended to the general linear model with  $p$  explanatory variables. As indicated by earlier Equation (13), the expected bias of the ridge estimates can be expressed in terms of cofactors and the true  $\beta$  values. Then, for the case of two explanatory variables, a very simple interpretation was possible, as discussed

earlier for (14). However, the direct interpretation of (13) for three or more explanatory variables is much less obvious. An illuminating interpretation follows for the general model by means of the following theorem which is, so far as I can find, original with this paper.

**THEOREM 1.** The bias of the ridge estimate of the  $j^{\text{th}}$  standardized regression coefficient can be expressed as

$$E(\hat{\beta}_j^*) - \beta_j = \frac{k c_{jj}}{|A^*|} \sum_{i=1}^p \hat{b}_{ji}^* \beta_i, \text{ where } \hat{b}_{ji}^* = -1.0 \text{ if } i = j$$

and if  $i \neq j$ ,  $\hat{b}_{ji}^*$  denotes the ridge estimate of the  $i^{\text{th}}$  regression coefficient of the model where  $X_j$  has been regressed on the remaining  $(p-1)$  explanatory variables. (Of course,  $\beta_i$  and  $\beta_j$  in Theorem 1 represent the true regression coefficients in the original model,  $Y = X\beta + u$ .)

**PROOF:** In terms of the standardized variables, and for  $j=p$ , (13) is

$$\begin{aligned} (13a) \quad E(\hat{\beta}_p^*) - \beta_p &= \frac{-k}{|A^*|} \sum_{i=1}^p c_{pi} \beta_i \\ &= \frac{-k}{|A^*|} [c_{p1} \beta_1 + c_{p2} \beta_2 + \dots + c_{pp} \beta_p] \\ &= \frac{-k c_{pp}}{|A^*|} \left[ \frac{c_{p1}}{c_{pp}} \beta_1 + \frac{c_{p2}}{c_{pp}} \beta_2 + \dots + 1.0 \beta_p \right]. \end{aligned}$$

To evaluate  $c_{pi}/c_{pp}$ , let  $i=1$  and observe that

$$(15) \quad \frac{c_{p1}}{c_{pp}} = \frac{(-1)^{p+1}}{c_{pp}} \begin{vmatrix} r_{12} & r_{13} & \dots & r_{1p} \\ (1+k) & r_{23} & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p-1,2} & r_{p-1,3} & \dots & r_{p-1,p} \end{vmatrix}$$

$$(16) \quad = \frac{(-1)^{2p-1}}{c_{pp}} \begin{vmatrix} r_{1p} & r_{12} & \dots & r_{1,p-1} \\ r_{2p} & (1+k) & \dots & r_{2,p-1} \\ \dots & \dots & \dots & \dots \\ r_{p-1,p} & r_{p-1,2} & \dots & (1+k) \end{vmatrix} = -\hat{b}_{p1}^*, \text{ by Cramer's rule.}$$

Note that in (16), the value of  $c_{p1}/c_{pp}$  will always be a negative number,  $(-1)^{2p-1}/c_{pp}$ , times the minor given in (16), since the number of column interchanges needed to transfer the last column of the minor of (15) to the first column position (and to leave the relative order of the other columns unchanged), will always be  $(p-2)$ . Since the sign of  $|A_{p1}^*|$ , the minor in (15), is  $(-1)^{p+1}$ , the sign of the rearranged minor in (16) must always be  $(-1)^{p+1}(-1)^{p-2} = (-1)^{2p-1} = -1.0$ . In the same way, for any  $i < p$ ,  $(-1)^{p+i}(-1)^{p-i-1} = (-1)^{2p-1} = (-1)$  implies that  $c_{pi}/c_{pp}$  must be equal to  $-\hat{b}_{pi}^*$ . Since the order of an explanatory variable in the equation has no effect on its estimated coefficient in ridge or OLS regression, the theorem is proved.

The earlier conclusions based upon Equation (13) and the experiments for the two-explanatory-variable case can now be extended to the general case for  $p$  explanatory variables. The only difference from the simple two-variable case is that, instead of talking about bias as it relates to the simple correlation between the two-explanatory-variables and the true  $\beta$  values, we now need to talk about the bias of a particular ridge estimate,  $\hat{\beta}_j^*$ , as it is affected by the functional relationship of  $X_j$  to the other  $p-1$  explanatory variables and the true  $\beta$  values.

For illustration, consider a three-explanatory-variable model. From Theorem 1, the bias of the ridge estimate of the coefficient for the first standardized explanatory variable can be written as

$$(17) \quad E(\hat{\beta}_1^*) - \beta_1 = \frac{k c_{11}}{|A^*|} [(-1) \beta_1 + \hat{b}_{12 \cdot 3}^* \beta_2 + \hat{b}_{13 \cdot 2}^* \beta_2]$$

where the notation is the same as for Theorem 1 except that the longer, but more conventional, notation is used to denote the ridge estimates of the coefficients of the equation where  $X_1$  is regressed on  $X_2$  and  $X_3$ ; that is, where

$$(18) \quad \hat{X}_1^* = \hat{b}_{12 \cdot 3}^* X_2 + \hat{b}_{13 \cdot 2}^* X_3.$$

If  $X_1$  is a positive function of  $X_2$  and  $X_3$  in (18), as will usually be the case for economic variables which tend to go up and down together, then the bias

in (17) will tend to be small if  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$  are all of the same sign and about the same magnitude. Suppose, for example, that  $\beta_1 = \beta_2 = \beta_3$  and  $R_{y \cdot 123}^2 = 0.99$ . Further, suppose that  $r_{12} = r_{13} = r_{23} = 0.99$ . Then,  $r_{y1} = r_{y2} = r_{y3} \doteq \pm 0.9916652658$  and  $\beta_1 = \beta_2 = \beta_3 \doteq \pm 0.332773579$ . Thus, for  $k=0.2$  and positive  $\beta$ 's, the bias in  $\hat{\beta}_1^*$  is

$$(19) \quad E(\hat{\beta}_1^*) - \beta_1 = \frac{0.2(.4599)}{0.140238} [-0.33277358 + 2(.45205479)(0.33277358)] \\ \doteq -0.02092916.$$

For the preceding model, variance for estimated  $\beta_1$  by OLS was 0.06677852, if we assume that there were 14 observations. The corresponding mean square error for  $\hat{\beta}_1^*$  at  $k=0.2$  would be the variance plus the bias squared, or  $MSE(\hat{\beta}_1^*) = 0.00024940 + 0.00043803 \doteq 0.00068743$ . Thus, mean square error of  $\hat{\beta}_1^*$  at  $k=0.2$  is only about one percent that for OLS.

Suppose, however, that for the same explanatory variables as before, the dependent variable is such that  $\beta_1 = -\beta_2 = -\beta_3$ . Then, for  $R_{y \cdot 123}^2 = 0.99$ , we would have  $\beta_1 \doteq -0.98518436$  and  $\beta_2 = \beta_3 \doteq 0.98518436$ . Substituting into (17), we obtain

$$(20) \quad E(\hat{\beta}_1^*) - \beta_1 = \frac{0.2(.4599)}{0.140238} [.98518436 + 2(.45205479)(.98518436)] \\ \doteq 1.230374.$$

Thus, mean square error for  $\hat{\beta}_1^*$  at  $k=0.2$  is the variance, 0.00024940, plus  $(1.230374)^2$ , or  $MSE(\hat{\beta}_1^*) \doteq 1.51407$ , which is over 22 times the mean square error for OLS! Therefore, some caution in the use of ridge regression on a model with the preceding structure would be advisable. But to be able to know whether one can expect the small bias and small mean square error from ridge regression, such as that for the model giving the small bias in (19), or whether one can expect a large bias and mean square error as in (20), one must have good prior information about the true  $\beta$  values for  $Y = X\beta + u$  and good information or data concerning the nature of the interrelationships among the explanatory variables. Only if one has this good information does it appear possible to evaluate one's results from ridge regression.

Fortunately, at least for economic research, there are important estimating problems which should lend themselves very well to ridge regression. For example, in the estimation of production functions we would usually expect the explanatory variables (the factor inputs) to all be positively related to each other. Furthermore, we would expect each productive input to contribute a positive amount to total value product; otherwise, the manager or operator is being irrational in his use of resources. Therefore, from Theorem 1 we can conclude that ridge regression may provide a powerful new tool to the economist for estimation of production functions. (In fact, our preliminary results in estimating production functions appear very promising. These empirical results will be presented in a later paper.)

On the other hand, from Theorem 1 we can deduce that there are some estimating problems in economics where ridge regression should be used only with extreme caution, if at all. For example, consider an oversimplified hypothesized demand model

$$(21) \quad Q = \beta_0 + \beta_1 I + \beta_2 P + u.$$

In (21),  $Q$  denotes the product quantity demanded,  $I$  denotes per capita income, and  $P$  may denote either actual or predicted product price. In (21), both income and price may have trended upward over time; hence, in many cases they would be positively related to each other. If the product is not a so-called "inferior" good, the consumption may have tended to increase because of increased incomes. If so, we would expect  $\beta_1$  to be positive. However, from economic theory, we would expect the price coefficient,  $\beta_2$ , to be negative. Therefore, for at least some products, we would expect "poor" results from the use of ridge regression for the estimation of (21).

Economic models intermediate between production functions and the demand function of (21) would need to be examined on an individual basis. One advantage of Theorem 1 is that it allows one to examine the possible bias from ridge regression for each coefficient. (The economist may often be more interested in the mean square error of a particular coefficient than for the sum of mean square error for all coefficients.) Therefore, at this point in time, the following procedure appears reasonable:

1. For one or more coefficients of particular interest, say  $\beta_j$ , ridge regress  $X_j$  as a function of the other  $p-1$  explanatory variables at various plausible  $k$  values.
2. Hypothesize one or more sets of  $\beta$  values for the main model,  $Y = X\beta + u$ .
3. Substitute these hypothesized  $\beta$  values and the ridge estimates from Step #1 into Theorem 1 to provide some idea of expected bias at various values of  $k$ .
4. If the expected bias does not appear to be too serious, use ridge regression to estimate the main model,  $Y = X\beta + u$ , and select a value of  $k$  giving "fairly low" variance and "fairly low" bias (from Step #3).

In practice, of course, the above procedure can be criticized as being rather vague and subjective. Nevertheless, it appears to be a reasonable alternative to the present all-too-common practice of merely deleting important variables, as discussed in the early sections of this paper. Other alternatives should also be considered by economists, such as incorporating prior information by means of the Theil-Goldberger mixed model [32].

#### SUMMARY AND CONCLUSIONS

Two empirical economic models were analyzed to measure the effect of omitting relevant variables from these models. Effect of omitting variables from an originally well-specified production function was to overestimate the coefficients of interest by a factor of about four. Similarly, deleting the commonly omitted distance or travel time variable from an outdoor recreation model resulted in overestimating the transfer cost coefficient by about three. In both cases, small variances for the coefficients in the omitted-variables model could lull the inexperienced researcher into a false sense of security, if he were not properly aware of the dangerous consequences of omitted-variable bias.

Relevant variables have usually been omitted from economic models because of multicollinearity problems encountered with the complete models. One fairly recent alternative to the deletion of relevant variables is ridge regression [15, 16, 26], developed by engineering and physical science researchers. The second part of this paper explores the potential of ridge regression for estimating economic models.

Although always effective for reducing high variances caused by multicollinearity, a few experiments with simple two-explanatory-variable models showed that ridge regression estimates were sometimes very good and, at other times, rather poor and misleading. Poor results were caused by large biases in the ridge estimates. Examination of an expression for the bias showed that for high positive intercorrelation in the two-explanatory-variable model, low bias results if the true coefficients are of the same sign and of about equal magnitude.

The preceding result was extended to the general case of  $p$  explanatory variables by means of a mathematical theorem. In essence, this theorem states that the expected bias for the ridge estimate of the coefficient of the  $j^{\text{th}}$  variable in the model,  $Y = X\beta + u$ , is the following:

$$E(\hat{\beta}_j^*) - \beta_j = \frac{k c_{jj}}{|A^*|} [\hat{b}_{j1}^* \beta_1 + \hat{b}_{j2}^* \beta_2 + \dots + \hat{b}_{j,j-1}^* \beta_{j-1} - 1.0 \beta_j + \hat{b}_{j,j+1}^* \beta_{j+1} + \dots + \hat{b}_{jp}^* \beta_p].$$

In the above theorem,  $\hat{b}_{ji}^*$  denotes the ridge estimate of the coefficient for the  $i^{\text{th}}$  variable, where  $X_j$  has been regressed on the  $p-1$  remaining explanatory variables;  $A^* = (X'X + kI)$ ; and  $c_{jj}$  is a cofactor of  $A^*$ . From the theorem, one can deduce that if the sum of the  $\hat{b}_{ji}^*$  values sum to around 1.0, then the bias and mean square error for the ridge estimate of the  $j^{\text{th}}$  coefficient will be relatively small if the true  $\beta$  values all have the same sign. Thus, some economic models, such as production functions, may be very well suited for estimation by means of ridge regression. On the other hand, ridge estimation of some other economic models, such as certain demand functions, could give poor results.

Further experimentation with various decision rules for selecting a particular  $k$  value would likely be of value for the empirical estimation of economic models by ridge regression.



## REFERENCES

- [1] Arrow, K. J., H. B. Chenery, S. B. Minhas, and R. M. Solow. "Capital-Labor Substitution and Economic Efficiency." The Review of Economics and Statistics, XLIII (Aug. 1961) pp. 225-250.
- [2] Brown, W. G. and Farid Nawas. The Oregon Big Game Resource: An Economic Evaluation. Corvallis, Oregon Agricultural Experiment Station Special Report 379, March 1973, 110 pages.
- [2a] Brown, W. G. and Farid Nawas, "Effect of Aggregation Upon the Estimation and Specification of Outdoor Recreation Demand Functions," Proceedings 1972, 45th Annual Meeting of the Western Agricultural Economics Association at Logan, Utah, July 23, 24, 25, 1972, pp. 344-350.
- [3] Cesario, Frank J., and Jack L. Knetsch. "Time Bias in Recreation Benefit Estimates." Water Resources Research, Vol. 6, No. 3, 700-704, June 1970.
- [4] Clawson, Marion. Methods of Measuring the Demand For, and Value Of, Outdoor Recreation. Reprint No. 10 (Resources for the Future, Inc., Washington, D.C., Feb. 1959).
- [5] Cramer, J. S. "Efficient Grouping, Regression, and Correlation in Engel Curve Analysis." Journal of the American Statistical Association, March 1964, pp. 233-250.
- [6] Edwards, Clark. "Non-linear Programming and Non-linear Regression Procedures." Journal of Farm Economics, Feb. 1962, pp. 100-114.
- [7] Edwards, J. A., K. C. Gibbs, L. J. Guedry, and H. H. Stoevener. "The Demand for Outdoor Recreation in the Bend Ranger District, Deschutes National Forest, Oregon." Presented at the Natural Environments Workshop sponsored by Resources for the Future, Inc., at Missoula, Mont., August 5, 6, 1971. In process of publication by Oregon Agricultural Experiment Station, Corvallis.
- [8] Farrar, D. E., and R. R. Glauber. "Multicollinearity in Regression Analysis: The Problem Revisited." Review of Economics and Statistics, Vol. 49, 1967, pp. 92-107.
- [9] Goldberger, A. S. Econometric Theory. John Wiley & Sons, Inc., New York, N.Y. 1964.
- [10] Griliches, Z. "Specification Bias in Estimates of Production Functions." Journal of Farm Economics, 39, Feb. 1957, pp. 8-20.
- [11] Halter, A. N., H. O. Carter, and J. G. Hocking. "A Note on the Transcendental Production Function." Journal of Farm Economics, 39, Nov. 1957, pp. 966-974.
- [12] Hoch, I. "Estimation of Production Function Parameters Combining Time-Series and Cross-Section Data." Econometrica, 30:34-53, Jan. 1962.

- [13] \_\_\_\_\_ . Book Reviews Section, Natural Resources Journal, Vol. 7, July 1967, pp. 466-471.
- [14] \_\_\_\_\_ . "Simultaneous Equation Bias in the Context of the Cobb-Douglas Production Function." Econometrica, 26, Oct. 1958, pp. 566-578.
- [15] Hoerl, A. E., and R. W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." Technometrics, Vol. 12, No. 1, Feb. 1970, pp. 55-67.
- [16] \_\_\_\_\_ . "Ridge Regression: Applications to Nonorthogonal Problems." Technometrics, Vol. 12, No. 1, Feb. 1970, pp. 69-82.
- [17] Johnston, J. Econometric Methods. McGraw-Hill Book Co., Inc., New York, N.Y. 1963.
- [18] Judge, G. C., and T. Takayama. "Inequality Constraints in Regression Analysis." Journal of the American Statistical Association, Vol. 61, 1966, pp. 166-181.
- [19] Klein, L. R. A Textbook of Econometrics. Row, Peterson & Co., Evanston, Ill. 1953.
- [20] Knetsch, Jack L. "Outdoor Recreation Demands and Benefits." Land Economics, 39(4):387-396, November 1963.
- [21] Knetsch, Jack L., and Robert K. Davis. "Comparisons on Methods for Recreation Evaluation." Water Research, ed. by A. V. Kneese and S. C. Smith, pp. 125-142, Johns Hopkins, Baltimore, Md. 1966.
- [22] Knight, D. A. Farm Size and Resource Efficiency on Eastern Kansas Farms. Kansas Agr. Exp. Sta. Tech. Bull. 155, Manhattan, Kansas, October 1967.
- [23] \_\_\_\_\_ . The Relation of Farm Size and Efficiency to Economic Outcomes on Central Kansas Farms. Kansas Agr. Exp. Sta. Tech. Bull. 157, Manhattan, Kansas, May 1968.
- [24] \_\_\_\_\_ . The Relation of Farm Size and Efficiency to Economic Outcomes on Western Kansas Farms. Kansas Agr. Exp. Sta. Tech. Bull. 158, Manhattan, Kansas, June 1968.
- [25] Malinvaud, E. Statistical Methods of Econometrics. Rand McNally & Co., Chicago, Ill. 1966.
- [26] Marquardt, D. W. "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation." Technometrics, Vol. 12, No. 3, Aug. 1970, pp. 591-612.
- [27] Mundlak, Y. "Empirical Production Function Free of Management Bias." Journal of Farm Economics, Vol. 43, Feb. 1961, pp. 44-56.

- [28] Nawas, Farid. The Oregon Big Game Resource: An Economic Evaluation. Ph.D. thesis. Corvallis, Oregon State University, 166 numb. leaves. 1972.
- [29] Paris, D. G. Predicting Firm Behavior from Estimates of Input Productivity for a Sample of Western Kentucky Farms. Ph.D. thesis, University of Kentucky, Lexington, Ky., 1960, 134 pp.
- [30] Ruttan, V. W. The Economic Demand for Irrigated Acreage - New Methodology and some Preliminary Projections, 1954-1980. Published for Resources for the Future, Inc., by The Johns Hopkins Press, Baltimore, Md., 1965.
- [31] Theil, H. "Specification Errors and the Estimation of Economic Relationships." Rev. Internal Stat. Inst., 25:1/3(1957), pp. 41-51.
- [32] \_\_\_\_\_. "On the Use of Incomplete Prior Information in Regression Analysis." Journal of the American Statistical Association, Vol. 58, 1963, pp. 401-414.
- [33] U.S. Bureau of the Census, United States Census of Agriculture: 1954, U.S. Government Printing Office, Washington, D.C.
- [34] U.S. Department of Agriculture. Agricultural Statistics 1962, Table 467, U.S. Government Printing Office, Washington, D.C., p. 381.
- [35] \_\_\_\_\_. Agricultural Statistics 1968, Table 458, U.S. Government Printing Office, Washington, D.C., p. 313.
- [36] Wilson, Robert R. "Demand Theory: Time Allocation and Outdoor Recreation." Southern Journal of Agricultural Economics, 3:103-108, Dec. 1971.