Rural Sociology 45(3), 1980, pp. 483–500

Biased Estimation in Policy Research: an Illustrative Example of Ridge Regression in a Health System Model[1]

Michael K. Miller and Ken R. Smith
Department of Agricultural Economics and Rural Sociology, University of Arkansas, Fayetteville, Arkansas, and Department of Sociology, Cornell University, Ithaca, New York

ABSTRACT The paper develops an argument for the necessity of examining individual coefficients in policy models. As a result of this need, it is posited that something other than OLS estimators should be used since they are inflated and have extremely large variances when multicollinearity is present. Further, it is argued that policy models are by definition theoretically nonorthogonal. Ridge regression as one of a class of biased estimators is offered as one possible approach to dealing with the nonorthogonality problem in policy research. The logic of the approach is articulated and an empirical model of a health system is estimated with ordinary least squares and ridge estimators. The models are compared and implications discussed.

Introduction

Most rural sociologists do not engage in practical policy-oriented research (Nolan *et al.,* 1975; Nolan and Galliher, 1973; Stokes and Miller, 1975). While such neglect is philosophically at odds with the sanctioned role for the discipline (Bealer, 1969; Ford, 1973; Loomis and Loomis, 1967), more "practical" reasons for conducting such research have captured the attention of many academically-based sociologists. First, it has become clear to academicians that there is a real need for non-academic job markets for their products (see the ongoing series in the footnotes). The existence of such a market is at least partly contingent upon the ability of the graduates to engage in applied policy-relevant research. Second is the realization that the era of *carte blanche* in public science is rapidly being replaced by an era of tough-minded accountability (Bozeman, 1976). The gatekeepers of the monetary faucet have begun to demand that publicly supported research have some demonstrable utility. Relevance both to the solution of complex societal problems and/or the attainment of national ob-

[1] The authors acknowledge helpful suggestions provided by James E. Martin, Donald E. Voth, and reviewers for the Journal, and programming assistance provided by Patricia Firey and Randy White. The paper is published with the approval of the director of the Arkansas Agricultural Experiment Station. Work on the paper was supported by CSRS Project NE129 and Arkansas Experiment Station Project 894.

jectives is the definite front runner when it comes to criteria for assessing that utility (Spilerman, 1975).

With the renewed push for applied policy research has come a set of coterminous issues ranging from philosophical value neutrality problems to very practical questions of how academically-based research can organize to respond to policy problems in the restricted time frame required. One set of concerns that is ever present in the debate centers around methodological issues. It is the intent of this paper to contribute to that dialogue. Specifically, we examine the theoretically desirable but methodologically arduous presence of multicollinearity in substantive policy research models. Subsequently, we examine the efficacy of using a biased estimation technique—ridge regression—as one possible solution to the multicollinearity problem. The parameters of a health systems model are estimated by ordinary least squares and ridge regression, and the results are presented for comparison. Finally, implications for health care policy are discussed.

The problem of multicollinearity in policy models

The literature is replete with statements addressed to the differences (or lack thereof) between basic academic research and applied research, such as the hows of problem definition, variable selection, model specification, time frame, and so forth (Coleman, 1972; Fennessey, 1972; MacRae, 1971; Merton, 1949; Rossi and Wright, 1977; Scott and Shore, 1974). Although there are distinctive features of policy research that set broad parameters within which the researcher must operate, the fundamental goal is still the same: "making valid and reliable inferences about the effects of one set of variables on another" (Rossi and Wright, 1977).

Rossi and Wright (1977) have argued that if there is a difference, it is that the methodological requirements for policy research are (or should be) more stringent because so much depends on the correctness of casual inference. In such cases, the issue of discerning "net" effects of policy tractable variables becomes of paramount concern. Inherent in these "internal validity" issues is the problem of multicollinearity. In the absence of the ability to randomly assign policy treatments to randomized experimental and control groups, such as handling multicollinearity issues in the design stage, the problem must be dealt with in the analysis phase. One such strategy is the estimation of structural parameters by ridge regression. The problems of multicollinearity and how those problems are tackled by ridge estimation is the concern of the remainder of this paper.[2]

[2] It is critical to point out that neither ridge regression nor any other "mechanical fix" can substitute for well-constructed measures and attention to precise theoretical detail explicating the nature of the relationship among independent variables. We are

From a strictly methodological point of view and from the point of view of those people making policy, it would be ideal to have a set of tractable variables strongly related to the target variable and completely unrelated to each other and to other variables not included in the model. This situation would allow the researcher to estimate parameters of the relationships between any given policy variable and the target or dependent variable without being concerned about controlling the effects of other variables. In such instances, the gross effect would be the net effect and the policy maker would have a very clear picture of how to proceed to acquire the desired level of the target variable. However, completely orthogonal systems are very unsatisfying theoretically and nonexistent empirically. Some degree of intercorrelation among the set of germane independent variables is always present. The question is: How does one go about offering a structural interpretation, such as discussing "net effects" of individual independent variables, where there is some degree of multicollinearity present? It will be useful to examine the problem somewhat formally.

Probably the most widely-used approach to obtaining coefficients of net impact is through the use of the basic linear regression model:

$$Y = XB + U$$

where:

B is a vector of regression coefficients which summarize, quantitatively, the effect of a change in any given independent variable on the value of the dependent or target variable.

Since B provides information that allows the researcher to analyze the separate effects of each of the independent variable's influence on the dependent variable, it is this coefficient that is of central concern. In order for B to provide valid information on net impacts, however, certain assumptions about the nature of the data must be met (Farrar and Glauber, 1967; Johnston, 1972).

Although there are several assumptions—disturbance assumption, homoscedasticity, absence of serial correlation,[3] independence of explanatory variables and error terms—one which is often neglected relates directly to the multicollinearity problem. Formally, the assumption is $\rho(X) = K$, a rank assumption. The assumption dictates that the rank of the matrix X be equal to the number of independent

only suggesting that, subsequent to these efforts, if it is still necessary to undertake some steps to deal with multicollinearity, ridge regression has some characteristics to recommend its use.

[3] It should be pointed out that if serial correlation is present, as when the assumption is violated, the variance estimates will be affected in the same way as when the independent variables in the model are correlated.

or exogenous variables (including the "dummy" variable of the intercept). In general this means that no independent variable in the matrix is an exact linear combination of any or all of the other independent variables. This guarantees that each column of X is an independent vector and hence there can be a parameter estimated for each variable (again including the intercept).[4] If the assumption is not met, there are two levels at which difficulties can arise: a purely mathematical level and a statistical analysis level (Rockwell, 1975). In the first instance the problem is rather straight-forward. If one column of X is a perfect multiple of another (or several others in combination), then the determinant of the matrix is zero, such as in $|X'X| = 0$. This implies that X'X is a singular matrix mathematically impossible to invert. Since no inversion is possible, the vector $\hat{B}$ cannot be estimated and further analysis is unachievable.[5] The more typical situation and the one of greatest concern arises where rather extensive, but not perfect, linear association exists between two or more variables. This results in a *multicollinearity problem* which renders interpretation of individual regression coefficients inappropriate. The problem is that as the interdependency among the predictor variables increases, the magnitude of the determinant of the X'X matrix will approach zero. If $|X'X| \approx 0$, then its inverse, $(X'X)^{-1}$, will tend to inherit large diagonal elements. The condition will typically result in least squares regression estimates that have large standard errors (implying lack of precision in the estimators) and small t ratios, (Mason and Brown, 1975). This will be the case because the estimated standard errors are directly proportional to the square roots of the main diagonal elements in $(X'X)^{-1}$. In other words, the variances of the respective regression coefficients are inflated by an amount directly proportional to the size of the diagonal elements of $(X'X)^{-1}$ and the increase is due solely to the correlation among the independent variables.[6] Additionally, the estimated parameters, $\hat{B}$, are susceptible to incorrect algebraic signs, have absolute values that

[4] It should be realized that the rank assumption $\rho(X) = K$ implies a degrees-of-freedom assumption, $d.f. = N - K > 0$. This is true since if $N < K$ it would be impossible for X to have rank K. As a result, the major assumption is one of rank, not of degrees of freedom (see Intriligator, 1978).

[5] If perfect multicollinearity is present, it is usually caused by a problem that can be corrected fairly easily. Examples include situations where: (1) one of the explanatory variables is a constant over the range of the entire sample (if this is the case, then that particular column is a multiple of the unity variable included in the equation to account for the intercept); and (2) the investigator has included all categories of a dummy variable in the equation as well as an intercept term. In both instances the solution is simply to eliminate the offending explanatory variable (or category of the dummy) with no loss of information.

[6] The reason that the diagonal elements of $(X'X)^{-1}$ are referred to as "parameter variance inflation factors" (Marquardt, 1970) should be clear by noting that the standard error of the regression coefficient is defined algebraically as $\hat{s}_j = \hat{s}^2(x'x)^{-1}$.

are too large and are plagued by the "bouncing beta" syndrome (Marquardt, 1970). This last condition implies that the regression estimates are highly sensitive to even minor changes in the data base. Consequently, it becomes impractical to attempt replication or validation through the use of another sample (Churchill, 1975). Finally, if multicollinearity is a problem, it is common to get relatively large R^2 values and significant F values for the entire model while simultaneously getting small t values indicating that no individual predictor variable is statistically important. Hence, one is left with exactly the situation that a policy researcher does not want. You know that the set of explanatory variables does affect the target variable, but you can't be certain what the *separate net effect* of any given variable is.

What does one do if multicollinearity is judged to be an issue? Maddala (1977) suggests six possible solutions to the problem: (1) dropping variables, (2) using extraneous estimates, (3) using ratios or first differences, (4) using principal components, (5) getting more data, and (6) ridge regression. All of the possible solutions have benefits and costs associated with them. Ultimately, the circumstances surrounding a particular research problem must dictate which solution is most desirable. If the model to be estimated is based on a sound theoretical framework and/or if policy makers have dictated that the model include a certain set of variables, then respecification, such as dropping variables, is an unacceptable solution. Similarly, when working with cross-sectional data, as is often the case, using first differences is not an available alternative. Further, the use of ratios and first differences adversely affects the properties of residuals by introducing heteroscedasticity and autocorrelation respectively (Maddala, 1977). While the use of principal components has one very ingratiating methodological feature, producing a completely orthogonal set of explanatory variables, its utility for policy research appears almost nonexistent. For example, how is a useful policy interpretation given to a regression coefficient attached to a complex linear com-

Hence, if you have a 2 × 2 correlation matrix, $(X'X) = \begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix}$ and

$$(X'X)^{-1} = \begin{bmatrix} \dfrac{1}{1 - r_{12}^2} & \dfrac{-r_{12}}{1 - r_{12}^2} \\ \dfrac{-r_{12}}{1 - r_{12}^2} & \dfrac{1}{1 - r_{12}^2} \end{bmatrix}$$

Suppose the multicollinearity is evidenced by a correlation between X_1 and X_2 equal to $r_{12} = .98$. Then the variance inflation factors will be equal to $\dfrac{1}{1 - .9604} = 25.25$. In a 2 × 2 matrix, the inflation factor will be the same for both parameters. For larger matrices this will not usually be true since the value of the diagonal elements depends on the magnitude of the partial correlation between each x and all other x's. The same logic does, however, apply.

bination defined by 2 (race) + 3 (differentiation) + 4 (percent urban)? It is not overstating the case to suggest that such a variable has little meaningful interpretation from a policy perspective.

The use of ridge regression is not the be-all-end-all solution to multicollinearity problems. That is not the claim. It does, however, seem particularly well suited (among the less than perfect alternatives) for use in policy research, because it allows structural interpretation of individual coefficients with more confidence. The logic of the procedure is presented below.

Ridge regression

As pointed out above, if there is no intercorrelation among the set of independent variables (and the other relevant assumptions are met), ordinary least squares does a yeoman's job of providing estimates of the slope coefficients. But, as multicollinearity increases, the variances of the OLS estimates also increase. The resulting increase renders the estimated values for the coefficients, $\hat{B}$, produced from any one sample questionable. Small changes in the data, such as adding or deleting one or two observations, can result in very significant changes in the magnitude of the coefficients and even in the direction of their impact. Hence, although OLS estimates still retain their BLUE properties in the presence of multicollinearity, the results from any particular sample are likely to exhibit a host of undesirable characteristics. Ridge regression (Hoerl and Kennard, 1970a, 1970b) is an estimation technique producing estimates of the coefficients that are closer, on the average, to the true population parameter, B, than are the OLS estimates (Feig, 1978). Somewhat more formally, ridge regression is one of a whole class of biased estimation techniques that, in the face of multicollinearity, result in a total mean square error that is significantly smaller than the total variance resulting from an OLS solution (Feldstein, 1973; McDonald and Galarneau, 1975). Hence, the chances of getting an estimate of B that is a "good" approximation is increased.[7] The ridge technique accomplishes this result by a very straightforward procédure. Since the problem stems from inflated values in the diagonal of the inverse, the ridge proce-

[7] It should be noted at this point that the decision to employ minimum mean square error (MSE) as an evaluation criteria is not inconsistant with the "least squares" criteria normally encountered in OLS regression. In fact, least squares is a special case of the more general mean square error. In particular, it is true that: mean square error = bias^2 + variance. Since OLS estimates are unbiased, as when $E(\hat{B}) = B$, MSE turns out to be the traditional variance criteria normally encountered. Since MSE is a summary measure of the accuracy of estimated coefficients, that is the smaller the value of MSE the closer the estimates, $\hat{B}$, cluster around the true parameter, and takes variance and bias into account it is a most appropriate measure to use where unbiased and biased estimators need to be compared.

dure counters the tendency by adding small positive quantities, K, to the diagonal of X'X such that $\hat{B}\rho = [X'X + K]^{-1}X'Y$. This mechanically decreases the size of the diagonal entries in the inverse matrix.[8] The reason this procedure produces estimates closer to the true value of the parameter has to do with the relationship between the bias, variance and mean square error of an estimator. In particular, it turns out that if the criterion of least mean square error is to be used to evaluate an estimator, the following relationship must be considered: Mean square error = Bias^2 + Variance. Mathematically, the nature of the relationship is such that to facilitate a reduction in the variance means you will, at the same time, increase the bias. That is exactly what happens when the positive quantity, K, is added to produce the ridge estimates, $\hat{B}^r$. Both the variance and the squared bias are functions of K (Hoerl and Kennard, 1970a). Specifically, the total variance is a monotonic decreasing function of K while the squared bias is a monotonic increasing function. Hence, there is a trade-off necessary. Fortunately, the nature of the relationships between K and the variance and bias is such that there are admissable values of K for which the mean square error of the ridge estimator is less than the OLS estimator. This is possible because the aforementioned monotone functions do not change at the same rate (Churchill, 1975).

[8] The reason for the decrease in the diagonal elements of the inverse can best be understood by noting what happens to the determinant of the 2 × 2 matrix presented in Footnote 3 when small positive quantities are added to the diagonal of X'X.

If $r_{12} = .98$ then:

$$(X'X) = \begin{bmatrix} 1 & .98 \\ .98 & 1 \end{bmatrix}$$

The determinant of that matrix is given by:

$$|X'X| = (1)(1) - (.98)(.98) = .0396$$

$$\text{so:} \quad X'X^{-1} = \begin{bmatrix} \frac{1}{.0396} & \frac{-.98}{.0396} \\ \frac{-.98}{.0396} & \frac{1}{.0396} \end{bmatrix}$$

Hence, in this case the main diagonal elements, as the variance inflation factors, equal 25.25. Now suppose a positive quantity, K equal to .40 is added to the main diagonal of X'X. It then follows that:

$$(X'X + K) = \begin{bmatrix} 1.4 & .98 \\ .98 & 1.4 \end{bmatrix}$$

and $|X'X + K| = (1.4)(1.4) - (.98)(.98) = .9996$

$$\text{Then:} \quad [X'X + K]^{-1} = \begin{bmatrix} \frac{1.4}{.9996} & \frac{-.98}{.9996} \\ \frac{-.98}{.9996} & \frac{1.4}{.9996} \end{bmatrix}$$

The main diagonal elements of the inverse have now been reduced from 25.25 to 1.30.

As K increases away from zero, such as when small positive values are added to the diagonal, the variance decreases rapidly. Conversely, the squared bias remains almost zero at first and then begins to increase more quickly (Churchill, 1975; Hoerl and Kennard, 1970a). Hence, as Hoerl and Kennard (1970a) point out, it is possible to substantially reduce the total mean square error of estimation by allowing a small amount of bias, but at the same time substantially reducing the variance. As a result, when multicollinearity is a problem, "ridge estimates can be produced which tend to be closer to the true parameter value, on the average, than the corresponding least square estimates" (Churchill, 1975). Further, the magnitude of the improvement of using the ridge estimation increases rapidly as X'X becomes less well conditioned, as when multicollinearity becomes more severe, and as the model fit decreases, as when R^2 is low (Deegan, 1975; Feig, 1978; Hoerl, Kennard and Baldwin, 1975).

Since ridge regression provides more efficient estimation in the presence of multicollinearity (for a certain range of K), the problem becomes one of determining what value of the biasing parameter, K, to employ. Although Farebrother (1975), Hoerl, Kennard and Baldwin (1975), and Kasarda and Shih (1977) provide algorithms for the automatic selection of the "optimum" K value, for pedagogical purposes we will discuss the use of the more subjective *ridge trace.* An examination of the trace provides a method of selecting a reasonable range of K values which will, in any given instance, provide practical results (Marquardt, 1970).

A ridge trace is simply a graphical procedure for estimating an appropriate value for K. The strategy suggested by Hoerl and Kennard (1970b) is to estimate $\hat{B}^r = (X'X + K)^{-1}X'Y$ for a series of K values from 0 to 1 (as in $0 \leq k \leq 1$). Subsequently, the estimates $\hat{B}_i^r$ are plotted as a function of K. Finally, using guidelines based on (1) "stability" of the trace, (2) magnitudes, and (3) sign reversals of estimated coefficients, and (4) increase in residual sum of squares,[9] the ridge trace is examined and a specific K value for the given model is selected. In the section that follows, the ridge procedure is applied to an empirical health systems model presented by Miller and Stokes (1978).

[9] Hoerl and Kennard (1970) offer the following specific guidelines for selecting the value of K from an examination of the ridge trace: (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; and (4) the residual sum of squares will not have been inflated to an unreasonable value. The reader is urged to examine the article by Conniffe and Stone (1973) for a critical assessment of using the ridge trace as a means of selecting K values.

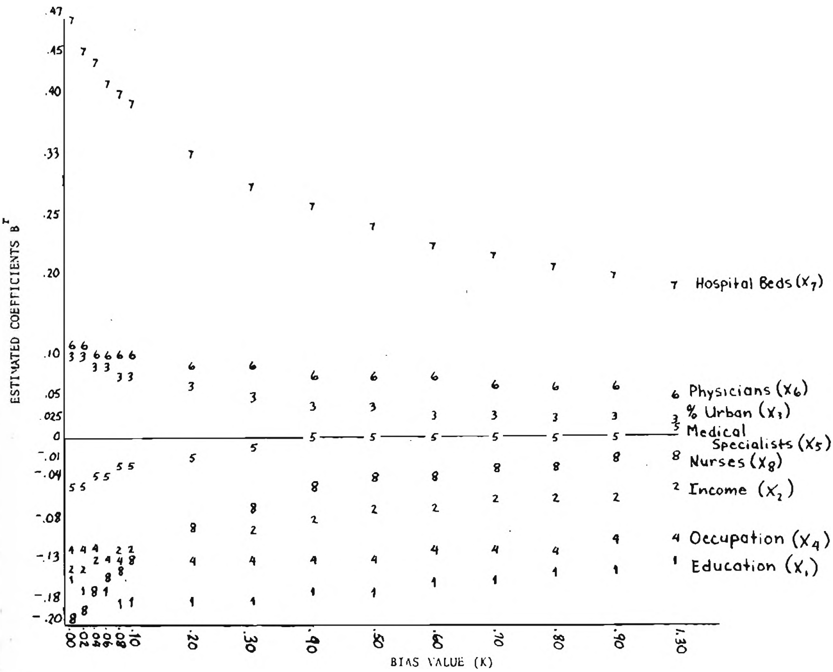

Figure 1. Ridge trace: infant mortality rate 1970.

Ridge estimates for a health systems model

Many governmental efforts designed to improve health status are guided by the "Great Equation." To wit: more resources = better health. While this basic formula is consistent with conventional wisdom, there is a growing literature (Fuchs, 1974; Illich, 1976; Miller and Stokes, 1978; Somers, 1973) to suggest that empirically it does not work. In fact, several recent studies (Illich, 1976; Miller and Stokes, 1978) have argued that the reverse is true; increased concentrations of manpower and facilities result in higher death rates. To empirically evaluate the two opposing claims, it is necessary to construct a model of the health care system and then observe the system's impact on health. However, the internal components of the health system are themselves interrelated both theoretically and empirically (Anderson, 1972, 1973; Field, 1973; Rushing, 1975). Further, the levels and types of manpower and facilities available to a population are determined in large part by the socioeconomic structure of the particular community (Miller and Stokes, 1978; Rushing, 1971, 1975). In short, interdependency within the health system itself and between the system and the structure of the supporting community is ever present. Hence, if the goal is to contribute to informed public health care policy via estimation and structural interpretation of empirical models, the difficulties inherent in collinear systems must somehow be dealt with.

Miller and Stokes (1978) used OLS to estimate an eight variable model designed to assess the net impact of health resources on physical health status. Briefly, their conclusions were that increased concentrations of manpower (with the single exception of nurses) and facilities (such as hospital beds) are associated with higher death rates once the impacts of socioeconomic structure of the community have been removed. Conversely, the higher the concentration of nurses, the lower the death rates. Throughout their analysis the authors make reference to and interpret as significant (substantively) both the direction and magnitude of individual coefficients. The problem is that the degree of multicollinearity among the set of predictor variables was, by most all criteria (Johnston, 1972; Kmenta, 1971; Maddala, 1977; Rockwell, 1975) at a problematic level.[10] Hence, the assessment of individual effects was somewhat tenuous. Because of the nature of the results and the implications for public health care policy, and because the extant level of multicollinearity could have had a marked effect on the estimates of the parameters, we have reestimated the model via ridge regression. The ridge traces are presented in Figures 1 and 2. A summary of the reestimation results are contained in Table 2.[11]

Turn first to Figures 1 and 2. An examination of the ridge traces allows several conclusions to be drawn:

(1) The set of coefficients at $K = 0$ (the ordinary least squares solution) is collectively unstable. This is evidenced by the fact that very small increases in K result in widely different estimates of $\hat{B}$. This suggests that the estimates, B, are very sample specific and that another sample would, most likely, produce substantially different estimates of the parameters.

[10] There are any number of ways to assess whether multicollinearity is in fact at a problematic level. For the data analyzed in the present study (and by Miller and Stokes, 1973), the following information summarizes the degree of difficulty:

(1) Determinant of $X'X = .0202$

(2) Condition number of matrix = 20.366. This is the ratio of the largest eigenvalue to the smallest eigenvalue. Thus, when the matrix is orthogonal the condition number is one. Any value beyond one indicates increasing problems of multicollinearity. See Von Neumann and Goldstine (1947) for a formal discussion.

(3) Haitovsky's heuristic chi square statistic = 2.982 indicating a very high probability of severe multicollinearity (see Rockwell, 1975 for a discussion of this statistic).

[11] Because of space limitations we have not included the models for 1950 or 1960. The results of those reestimated models are consistent with the 1970 data presented in this paper. Readers who are interested in the specific results can obtain a copy of the ridge traces and the coefficient estimates in the range $0 \leq K \leq 1$ from the authors.

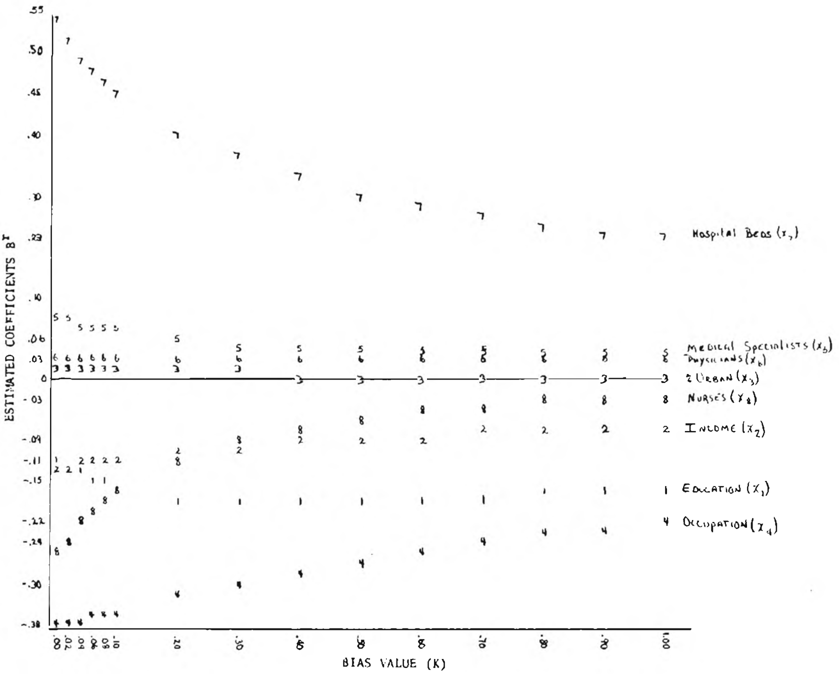

Figure 2. Ridge trace: age-sex standardized death rate 1970.

(2) The coefficients for variables 7 and 8 (hospital beds and nurses, respectively) are decidedly unstable at the least squares point; their respective coefficients demonstrate very rapid declines in magnitude as K becomes positive. The instability is caused by the degree of multicollinearity or correlation between these two variables and others in the system.

(3) The coefficients for variables 1 (education), 2 (income), 3 (percent urban), 4 (occupation) and 6 (physicians) maintain relative consistency over the entire range of K values considered. The trace also indicates that degree of urbanization (variable 3) exerts virtually no impact on the two measures of health; the coefficient moves very close to zero with the introduction of small values of the biasing parameter.

(4) Medical specialists (variable 5) exhibits opposite signs at K = 0 for infant mortality (−) and age-sex standardized death rate (+). In both cases, however, the coefficient moves toward zero and the negative sign changes to positive as K approaches .40.

(5) Finally, an examination of the two ridge traces indicates that both systems stabilize around K = .30. It should be noted that there has been some increase in error sum of squares at this value; notice

Table 1. Estimates of Mean Square Error (MSE) for OLS and Ridge Estimators*

(Dependent variable)	Estimator	Variance	+	$Bias^2$	= MSE
Infant mortality	OLS (k = 0)	19.68 (12.007)	+	0	236.30
	Ridge (k = .3)	5.04 (12.007)	+	.05679	60.57
Standardized death rate	OLS (k = 0)	19.68 (.2981)	+	0	5.866
	Ridge (k = .3)	5.04 (.2981)	+	.0894	1.592

* The two components of the MSE (variance and $Bias^2$) are defined algebraically as:

$$MSE = \sigma^2 \Sigma \lambda i/(\lambda i + k)^2 + k^2 B'(X'X + kI)^{-2} B$$

$$[\text{Variance}] + [\text{Bias}^2]$$

the R^2 for infant mortality decreased from .32 to .30. For standardized death rate, the reduction was from .59 to .56. Hopefully, however, the small increases in error sum of squares will be more than offset by a set of coefficients closer to the true parameter values. The degree to which that is true in the present case is evidenced by the mean square error values presented in Table 1. Clearly, the desired increase in accuracy has been achieved by employing the ridge technique. For infant mortality, the OLS equation has an accompanying mean square error (MSE) of 236.3. The MSE for the ridge estimate is 60.5. Hence, the addition of a small amount of bias has resulted in virtually a fourfold reduction in MSE. The gain in accuracy is also apparent for the standardized death rate; the MSE reduction is from 5.8 to 1.5.[12]

Assuming the ridge estimates $\hat{B}^r$ at K = .30 are close to the true parameter values, what can be said about the substantive findings of the model? First, it should be pointed out that all of the coefficients have the same sign at K = .3 as for the OLS estimates. Hence, the

[12] The reader should be aware that the reported MSE for the ridge estimates are based on the *estimated* b coefficients and not the *true* b's. The equation for estimating the MSE is given by Hoerl and Kennard, (1970a:60) as:

$$MSE = \sigma^2 \Sigma \lambda i/(\lambda i + k)^2 + k^2 B'(X'X + kI)^{-2} B = \text{variance} + \text{bias}^2$$

Where σ = standard error of the OLS equations,
λ = eigenvalues of the X'X matrix,
k = positive quantities added to the main diagonal of the X'X matrix, where $0 \leq k \leq 1$,
B = vector of the true parameter values, and
X'X = correlation matrix,

The equation shows that the bias squared term is partly a function of the true B's. However, by using the "inflated" estimates of the parameters, it is obvious that the bias squared term will be biased upwards. The consequence of this problem is simply that the gain in MSE for the ridge regressions is underestimated.

Note that the use of the estimated B's does not affect the MSA estimates for the OLS equations. Since k = 0, the bias squared term reduces to zero.

Table 2. OLS and Ridge Regressions of Health Status on Structural Characteristics and Health System Resources[1]

	Infant mortality 1970		Age/sex standardized death rate 1970	
	K = 0 (OLS)	K = .30	K = 0 (OLS)	K = .30
(X_1) Education	−.162	−.177	−.117	−.188
(X_2) Income	−.150	−.094	−.137	−.099
(X_3) % Urban	.099	.049	.027	.016
(X_4) Occupation	−.122	−.136	−.383	−.308
(X_5) Medical specialists	−.056	−.005	.099	.059
(X_6) Physicians	.106	.083	.039	.042
(X_7) Hospital beds	.475	.294	.552	.352
(X_8) Nurses	−.211	−.075	−.266	−.092
R^2	.32	.30	.59	.56

[1] Appendix A contains the ridge estimates for the entire range of K values used to produce the ridge trace.

nature of the conclusions offered by Miller and Stokes (1978) has not changed. "Increases in health resources give no assurance of decreasing mortality rates" (Miller and Stokes, 1978:275). It is also important to note, however, that although the nature of the structural interpretation has not changed, the absolute values of the coefficients have tended to decrease. In other words, the OLS model overestimated the impact of most of the eight variables on health status. (The exception is education which was slightly underestimated at the least squares point.) The overestimate was particularly pronounced for two of the health resource variables, hospital beds and nurses. The fact that the reestimated model produced results which were consistent with those presented by Miller and Stokes (1978) gives support to the policy implications that "if medical care is going to have little or no impact on physical health except insofar as poor care can be worse than no care, quality becomes a central concern" (Miller and Stokes, 1978:275). Clearly, additional research designed to further clarify the relationship between concentration of resources and health status is needed before policy is implemented which uses the "Great Equation" as its principal rationale.

Summary

The paper argues that rural sociologists need to engage in applied, policy-relevant research. To facilitate such research, it is posited that individual coefficients in a policy model must be estimated and interpreted. But, interdependency among germane variables is an integral part of virtually any theoretically-based model. As a result, the individual coefficients generated by an ordinary least squares estimation procedure are unstable, have large variances, and are often plagued by erroneous algebraic signs and absolute values that overestimate the true impact of individual predictors. Hence, interpretation of individual coefficients is tenuous. Ridge regression, as one of a class of biased estimators is suggested as one approach to dealing with the multicollinearity problem in policy research. The logic of the approach was articulated and the procedure was applied to an eight variable health systems model reported by Miller and Stokes (1978). An examination of the ridge trace indicated that the OLS coefficients were collectively unstable. Further, the coefficients produced by OLS tended to overestimate the impact of the predictor variables on health status. The overestimate was most pronounced for hospital beds and nurses. The system stabilized at approximately K = .30. An examination of the ridge coefficients, $\hat{B}^r$, at that point indicated that, although the magnitude of the coefficients was considerably reduced, the nature of the relationship between health resources and physical health status was consistent with those reported by Miller and Stokes. Nurses tended to have a positive impact on health while concentration of physicians and physical facilities exhibited a negative impact. The results indicate the need to conduct additional research before implementing policy based on the intuitively appealing formula: more resources = better health.

References

Anderson, J. G.
1972 "Causal model of a health service system." Health Services Research 7 (Spring):23–42.
1973 "Causal models and social indicators: toward the development of social systems models." American Sociological Review 38 (June):285–301.

Bealer, Robert C.
1969 "Identity and the future of rural sociology." Rural Sociology 34 (June): 229–33.

Boulding, Kenneth
1966 The Impact of the Social Sciences. New Brunswick, New Jersey: Rutgers University Press.

Bozeman, Barry
1976 "Reflections on the end of carte blanche: the inevitability of conflict between congress and the scientific community." Policy Studies Journal (Winter): 175–80.

Churchill, Gilbert A.
1975 "A regression estimation method for collinear predictors." Decision Sciences, 6 (October):670–87.

Coleman, James S.
1972 Policy Research in the Social Sciences. Morristown, New Jersey: General Learning Press.

Conniffe, D., and J. Stone
1973 "A critical view of ridge regression." The Statistician 22 (2):181–87.

Deegan, John, Jr.
1975 "The process of political development: an illustrative use of a strategy for

regression in the presence of multicollinearity." Sociological Methods and Research 3, (May):384–415.
Farebrother, R. W.
1975 "The minimum mean square error linear estimator and ridge regression." Technometrics, 17 (February):127–28.
Farrar, Donald E., and Robert R. Glauber
1967 "Multicollinearity in regression analysis: the problem revisited." Review of Economics and Statistics, 49 (February):92–107.
Feig, Douglas G.
1978 "Ridge regression: when biased estimation is better." Social Science Quarterly 58, (March):708–16.
Feldstein, Martin
1973 "Multicollinearity and the mean square error of alternative estimators." Econometrica 41 (March):337–45.
Fennessey, James
1972 "Some problems and possibilities in policy-related social research." Social Science Research 1 (December):359–93.
Field, M. G.
1973 "The concept of the 'health system' at the macrosociological level." Social Science and Medicine 7 (October):763–85.
Ford, Thomas R.
1973 "Toward meeting the social responsibilities of rural sociology." Rural Sociology 38 (Winter):372–90.
Fuchs, Victor R.
1974 Who Shall Live? Health, Economics and Social Choice. New York: Basic Books.
Hoerl, Arthur E., and Robert W. Kennard
1970a "Ridge regression: biased estimation for nonorthogonal problems." Technometrics 12 (February):55–67.
1970b "Ridge regression: application to nonorthogonal problems." Technometrics 12 (February):69–82.
Hoerl, A. E., R. W. Kennard, and K. F. Baldwin
1975 "Ridge regression: some simulations." Communication in Statistics 4 (February):105–23.
Illich, Ivan
1976 Medical Nemesis: The Expropriation of Health. New York: Pantheon.
Intriligator, Michael D.
1978 Econometric Models, Techniques, and Applications. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.
Johnston, J.
1972 Econometric Methods. New York: McGraw-Hill Book Company.
Kasarda, John D., and Wen-Fu P. Shih
1977 "Optimal bias in ridge regression approaches to multicollinearity." Sociological Methods and Research 5 (May):461–69.
Kmenta, Jan
1971 Elements of Econometrics. New York: Macmillian Company.
Loomis, Charles P., and Zona Kemp Loomis
1967 "Rural sociology." Pp. 655–91 in Paul F. Lazarsfeld, William H. Sewell, and Harold L. Wilensky (eds.), The Uses of Sociology. New York: Basic Books.
MacRae, Duncan, Jr.
1971 "The dilemma of sociology: science versus policy." The American Sociologist 6 (Supplementary Issue, June):2–7.
Maddala, G. S.
1977 Econometrics. New York: McGraw-Hill.
Marquardt, Donald W.
1970 "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation." Technometrics 12 (August):591–612.

Mason, Robert, and William G. Brown
1975 "Multicollinearity problems and ridge regression in sociological models." Social Science Research 4 (June):135–49.
McDonald, Gary C., and Diane Galarneau.
1975 "A Monte Carlo evaluation of some ridge-type estimators." Journal of the American Statistical Association 70 (June):407–16.
Merton, Robert K.
1949 "The role of applied social science in the formulation of policy: a research memorandum." Philosophy of Science 16 (July):161–81.
Miller, Michael K., and C. Shannon Stokes
1978 "Health status, health resources, and consolidated structural parameters: implications for public health care policy." Journal of Health and Social Behavior 19 (September):263–79.
Nolan, Michael F., and John F. Galliher
1973 "Rural sociological research and social policy: hard data, hard times." Rural Sociology 38 (Winter):491–99.
Nolan, Michael F., Robert A. Hagan, and Mary S. Hoekstra
1975 "Rural sociological research, 1966–1974: implications for social policy." Rural Sociology 40 (Winter):435–54.
Rockwell, Richard G.
1975 "Assessment of multicollinearity: the Haitovsky test of the determinant." Sociological Methods and Research 3 (February):308–20.
Rossi, Peter H., and Sonia Wright
1977 "Evaluation research: an assessment of theory, practice, and politics." Evaluation Quarterly 1 (February):5–52.
Rushing, William A.
1971 "Public policy, community constraints, and the distribution of medical resources." Social Problems 19 (Summer):21–36.
1975 Community, Physicians, and Inequality. Lexington, Kentucky: Lexington Books.
Scott, Robert A., and Arnold Shore
1974 "Sociology and policy analysis." The American Sociologist 9 (May):51–9.
Somers, A. R.
1973 "Some basic determinants of medical care and health policy—an overview of trends and issues." Pp. 3–22 in John B. McKinlay (ed.), Politics and Law in Health Care Policy. New York: Prodist.
Spilerman, Seymour
1975 Forecasting Social Events. Chapter 15 in K. C. Land and S. Spilerman (eds.), Social Indicator Models. New York: Russel Sage.
Stokes, C. Shannon, and Michael K. Miller
1975 "A methodological review of research in rural sociology since 1965." Rural Sociology 40 (Winter):411–34.
Von Neumann, J., and H. Goldstine
1947 "Numerical inverting of matrices of high order." Bulletin of the American Mathematical Society 53:1021–99.

	Value of K:	0.0	0.02	0.04	0.06	0.08
Education	Beta 1	-.117	-.131	-.141	-.150	-.158
Income	Beta 2	-.137	-.132	-.128	-.124	-.121
% Urban	Beta 3	.027	.027	.027	.027	.026
Occupation	Beta 4	-.383	-.377	-.371	-.365	-.359
No. Specialists	Beta 5	.099	.093	.089	.085	.081
Physicians	Beta 6	.039	.041	.042	.043	.430
Hospital Beds	Beta 7	.553	.526	.504	.484	.467
Nurses	Beta 8	-.267	-.241	-.220	-.202	-.186
	SS Error	.409	.409	.410	.412	.414
	R^2	.591	.590	.589	.588	.586

	Value of K:	0.10	0.20	0.30	0.40	0.50
Education	Beta 1	-.164	-.182	-.188	-.189	-.188
Income	Beta 2	-.118	-.107	-.099	-.093	-.088
% Urban	Beta 3	.025	.021	.016	.013	.009
Occupation	Beta 4	-.354	-.329	-.308	-.291	-.275
No. Specialists	Beta 5	.078	.067	.060	.054	.049
Physicians	Beta 6	.043	.043	.042	.040	.039
Hospital Beds	Beta 7	.451	.392	.352	.322	.298
Nurses	Beta 8	-.172	-.122	-.092	-.071	-.057
	SS Error	.416	.427	.440	.452	.464
	R^2	.584	.573	.560	.548	.535

	Value of K:	0.60	0.70	0.80	0.90	1.00
Education	Beta 1	-.184	-.180	-.176	-.171	-.166
Income	Beta 2	-.084	-.080	-.077	-.074	-.071
% Urban	Beta 3	.007	.005	.003	.001	.000
Occupation	Beta 4	-.262	-.249	-.239	-.229	-.219
No. Specialists	Beta 5	.045	.041	.038	.036	.033
Physicians	Beta 6	.037	.035	.034	.033	.031
Hospital Beds	Beta 7	.278	.262	.247	.235	.223
Nurses	Beta 8	-.046	-.038	-.031	-.026	-.022
	SS Error	.477	.489	.500	.512	.523
	R^2	.523	.511	.499	.488	.477

Appendix A. Ridge estimates for age-sex standardized death rates 1970: $0 \leqq K \leqq 1$.

	Value of K:	0.0	0.02	0.04	0.06	0.08
Education	Beta 1	-.162	-.167	-.171	-.174	-.177
Income	Beta 2	-.150	-.143	-.137	-.132	-.127
% Urban	Beta 3	.099	.093	.087	.083	.078
Occupation	Beta 4	-.122	-.126	-.129	-.131	-.132
No. Specialists	Beta 5	-.057	-.049	-.043	-.038	-.033
Physicians	Beta 6	.106	.104	.103	.101	.099
Hospital Beds	Beta 7	.475	.452	.432	.414	.398
Nurses	Beta 8	-.210	-.191	-.175	-.161	-.149
	SS Error	.676	.676	.677	.678	.679
	R^2	.324	.324	.323	.322	.321

	Value of K:	0.10	0.20	0.30	0.40	0.50
Education	Beta 1	-.178	-.180	-.177	-.172	-.167
Income	Beta 2	-.122	-.106	-.094	-.086	-.079
% Urban	Beta 3	.074	.059	.049	.042	.036
Occupation	Beta 4	-.134	-.136	-.136	-.133	-.130
No. Specialists	Beta 5	-.029	-.014	-.005	.001	.005
Physicians	Beta 6	.098	.090	.084	.078	.073
Hospital Beds	Beta 7	.384	.330	.294	.268	.247
Nurses	Beta 8	-.138	-.099	-.075	-.058	-.046
	SS Error	.681	.689	.697	.705	.712
	R^2	.319	.311	.303	.295	.288

	Value of K:	0.60	0.70	0.80	0.90	1.00
Education	Beta 1	-.161	-.155	-.149	-.144	-.139
Income	Beta 2	-.073	-.069	-.065	-.061	-.058
% Urban	Beta 3	.032	.028	.025	.023	.021
Occupation	Beta 4	-.127	-.123	-.119	-.116	-.112
No. Specialists	Beta 5	.007	.009	.010	.010	.011
Physicians	Beta 6	.069	.065	.062	.059	.057
Hospital Beds	Beta 7	.229	.215	.203	.192	.183
Nurses	Beta 8	-.038	-.031	-.026	-.021	-.018
	SS Error	.720	.727	.733	.740	.746
	R^2	.280	.273	.266	.260	.254

Appendix B. Ridge estimates for 1970 infant mortality: $0 \leqq K \leqq 1$.