Forest Sci., Vol. 27, No. 2, 1981, pp. 339–348

Applications of Ridge Regression in Forestry

B. BRUCE BARE
DAVID W. HANN

ABSTRACT. Describes the use of ridge regression for dealing with multicollinearity in multiple linear regression. Ridge regression is reviewed and three criteria for selecting the "best" ridge estimator—ridge trace, variance inflation factor, and determinant of the correlation matrix—are discussed. The first application demonstrates the use of ridge regression for selecting independent variables during the development of a ponderosa pine basal area growth model. This use of ridge regression produced a meaningful predictive model with interpretable coefficients. The second application uses ridge regression to develop a descriptive model for estimating bare land values in the Douglas-fir region. The objective was to produce precise and stable estimates of model parameters and not to predict the dependent variable. The resulting bare land value estimates fall in the range of values produced by other techniques. FOREST SCI. 27:339–348.

ADDITIONAL KEY WORDS. Biased estimators, multiple linear regression, land valuation, multicollinearity.

CONSIDERABLE ATTENTION has focused on the use of biased estimation procedures during the 1970's. Much of this is due to the pioneering work of Hoerl and Kennard (1970a,b) who introduced ridge regression (RR) as a biased estimation procedure which avoids most of the pitfalls of ordinary least squares (OLS) in the presence of multicollinearity. Other biased estimation techniques recently introduced include principal components regression, James-Stein estimation, fractional rank, and generalized ridge regression (Marquardt 1970, Hocking 1976, Hocking and others 1976, Andrews 1974, Webster and others 1974, Vinod 1978, Hemmerle and Brantle 1978, Draper and Van Nostrand 1979, Marquardt and Snee 1975). Partly because of its relationship to OLS, RR has become one of the most common biased estimation techniques.

EFFECTS OF MULTICOLLINEARITY

One of the basic assumptions of OLS is that perfect correlation between a linear combination of one or more independent variables does not exist. If partial correlation exists then the regression is said to contain multicollinearity between the independent variables. Problems can arise depending upon the degree of multicollinearity that the regression model exhibits (Kmenta 1971).

While a model with a high degree of multicollinearity still provides unbiased estimates of the model parameters, the adverse effect of high multicollinearity is the production of very imprecise estimates of the regression coefficients (Kmenta

The authors are, respectively, Associate Professor, College of Forest Resources, University of Washington, Seattle; and Assistant Professor, School of Forestry, Oregon State University, Corvallis. An earlier version of this article was presented at the SAF/IUFRO 1979 Workshop on Forest Resources Inventories, Fort Collins, Colorado, July 23. Manuscript received 4 September 1979 and in revised form 15 December 1980.

1971). This impreciseness can cause some of the regression coefficients to be of the wrong sign from what is expected (Hoerl and Kennard 1970a,b; Brown and Beattie 1975). Further problems which result from multicollinearity are regression coefficients which change drastically when a new independent variable is added (or deleted) or which are sensitive to the addition (or deletion) of new data points (Chatterjee and Price 1977).

Concerns over the degree of multicollinearity are important whether the regression equation is to be used for purposes of prediction or for descriptive (structural) model building. In developing a predictive model, the primary objective is to select those variables which lead to the minimization of mean square error of prediction (Chatterjee and Price 1977). However, variables considered for selection should be based on theoretical grounds. A descriptive model is used to clarify or describe the nature of complex interactions present in the system being modeled. The primary objective is to select as few variables as possible which explain the largest amount of variation (Chatterjee and Price 1977). Generally a descriptive model requires a solid theoretical foundation which explains the behavior of the system under study. In a descriptive model, multicollinearity is a concern because interpretation of the regression coefficients is desired. Multicollinearity is a concern in a predictive model because future prediction errors can occur unless relationships present in the set of collinear data used for estimating the regression coefficients remain fixed in future sets of data.

Multicollinearity can also seriously affect the selection of variables in a stepwise regression analysis (Hocking 1976). This is a result of insignificant t-ratios on individual regression coefficients brought about by inflated variances of the coefficients. Thus, many relevant variables can be omitted from a stepwise regression equation and problems with model specification can arise.

Ridge Regression

RR sacrifices unbiasedness to obtain parameter estimates that, when compared to their unbiased OLS counterparts, have a smaller mean square error of the parameter estimates. In terms of solving for the standardized regression coefficients, the method consists of adding a small constant value, K, to the diagonal elements of the correlation matrix and then solving in the usual manner for the regression coefficients. When K is zero, OLS estimates result. While K can be any positive value, it usually lies between zero and one. Also, the larger the value of K, the larger the bias.

When using RR there are two persistent problems. First, if the true least squares population parameters are unknown, the amount of bias introduced is unknown. Hoerl and Kennard (1970a,b) have shown that there is a value of $K > 0$ such that the mean square error of the ridge estimator is less than the mean square error of the OLS estimator.

The second problem arises when determining the "best" value of K for a particular problem. There have been numerous methods proposed for determining this value. One common technique, first proposed by Hoerl and Kennard (1970a,b), is the ridge trace. This is a plot of all standardized regression coefficients over a range of K values. Hoerl and Kennard (1970a) suggest four criteria to consider when deciding upon a value of K from the ridge trace: (1) at a certain value of K, the ridge trace stabilizes, (2) coefficients will not have unreasonable absolute values in terms of *a priori* knowledge, (3) coefficients with theoretically improper signs at $K = 0$ will have proper signs, and (4) the residual sum of squares will not be significantly inflated. Although these guidelines are intuitively appealing, there is a considerable amount of subjectivity involved in their use in actual situations.

Another method is the maximum variance inflation factor (VIF) criterion proposed by Marquardt (1970). VIF's are the diagonal elements of the inverse of the correlation (standardized) matrix. In OLS the precision of each coefficient is measured by its variance which is proportional to σ^2, the variance of the error term. The VIF is the constant of proportionality (Chatterjee and Price 1977). Marquardt (1970) proposes a K value such that the maximum VIF is between ten and one, and closer to one if possible. Under perfectly orthogonal conditions, all VIF's are equal to one whereas under perfectly collinear conditions one or more of the VIF's tend toward infinity.

There are two advantages to the maximum VIF approach. First, the method is more objective than the ridge trace approach. Second, the resulting K value is nonstochastic (Obenchain 1975). This property is required if the equations for the expectation and covariance of the ridge estimators, developed by Hoerl and Kennard (1970a), are to remain valid.

A third criterion to use when selecting K is the determinant of the correlation matrix (Farrar and Glauber 1967). The determinant provides a measure of multicollinearity being close to zero when the degree of multicollinearity is high and close to one when the degree of multicollinearity is low. Thus, to provide additional insight, the determinant is calculated for each K value. Under conditions that the independent variables are normally distributed, Bartlett's statistic can be used to test the degree of multicollinearity as indicated by the determinant (Bartlett 1950, Haitovsky 1969, Farrar and Glauber 1967).

Other approaches for selecting a K value include those of Hoerl and Kennard (1970a, 1976), Hemmerle (1975), Hoerl and others (1975), and Mallows (1973). After reviewing these proposals, Mitchell and Hann (1979) conclude that the most promising method for selecting K involves the use of Marquardt's (1970) VIF.

Ridge Regression as a Screening Tool

The first application concerns the use of RR as a screening tool to help select independent variables during the course of model development. In presenting this application, our primary objective is to demonstrate the use of RR during the development of a predictive model and not to discuss all of the background details associated with the model. Readers interested in these details should consult Hann (1980). In using RR as a screening tool, variables having either small standardized regression coefficients or coefficients which are unstable as K increases can be identified on the ridge trace. Such variables can be eliminated from further screening runs unless other considerations warrant their retention.

This use of the ridge trace was suggested by Hoerl and Kennard (1970b). However, Marquardt and Snee (1975) caution against stabilizing a model through elimination of independent variables because doing so can introduce model specification problems. This concern is reasonable if the exact model form is known. However, if the exact form of the model is unknown, or if the objective of the analysis is to develop a meaningful predictive model with interpretable regression coefficients, the use of RR to help select independent variables appears justified.

Because an exact model form could not be prespecified, the above screening process was applied during the development of a model to predict basal area growth for a given diameter class for blackjack pine on the Fort Valley Experimental Forest (Hann 1980). Permanent plot information dating from 1920 were used in the analysis. The objectives of the regression analysis were to develop a predictive model whose coefficients were interpretable without unduly increasing the mean square error of residuals.

Prior to application of RR, a common model form was developed and separate OLS estimates were obtained for each of two data sets. Data set one included virgin uneven-aged blackjack stands and data set two included managed uneven-aged stands. The common equation form for each data set selected during this phase of the analysis was

$$\ln(\mathrm{BAG}/S) = a + b_1\ln(D) + b_2D + b_3\mathrm{LBA}_2 + b_4\mathrm{MBA}_2 + b_5\mathrm{UBA}_2$$

where

$\ln(\mathrm{BAG}/S)$ = Logarithm of 5-year basal area growth divided by site index
D = diameter class size
MBA_2 = total basal area in a given diameter class and the adjoining larger diameter class and smaller diameter class
LBA_2 = total basal area in all diameter classes less than the smallest diameter class in MBA_2
UBA_2 = total basal area in all diameter classes greater than the largest diameter class in MBA_2

This common model form agreed with results previously reported in the literature. Further, it satisfied our *a priori* assumptions as dictated by silvicultural and mensurational research. Thus, all variables were considered as appropriate and were fixed in further regressions (Hann 1980).

During the next phase of the analysis the differences in regression coefficients between the virgin and managed data sets were modeled as a function of time since cutting. This was done by introducing a number of new independent variables, which were products of the original, fixed set of independent variables from the common model form plus three sigmoidal transforms of time since cutting. The three sigmoidal transforms of time since cutting were of the form: $A_i = a_i + b_i\mathrm{EXP}(c_iT^{d_i})$ where T represents time since cutting. These sigmoidal equations were picked from the set of MATCH-A-CURVE sigmoidal models (Jensen and Homeyer 1970) to represent a range in sigmoidal forms. Therefore, the equations and parameters were prespecified in this analysis. A sigmoidal transform of time since cutting was used instead of a linear transform because the effect of cutting was expected to be asymptotic over the time interval considered. The new "time since cutting" independent variables selected during this phase of the analysis were: $A_3 \cdot D$, $A_1 \cdot \ln(D)$, $A_3 \cdot \mathrm{MBA}_2$, $A_2 \cdot \mathrm{UBA}_2$, and A_3.

With introduction of the new independent variables involving time since cutting, a high degree of multicollinearity was observed. To examine the severity of this, a ridge trace involving the new and original, fixed independent variables was prepared (Fig. 1). An examination of this ridge trace indicated that all of the variables being screened (except $A_3 \cdot \mathrm{MBA}_2$ whose standardized regression coefficient was almost zero) were unstable. The instability between D and $\ln(D)$ was expected, but was acceptable because the two variables were necessary to provide the desired effect over diameter class size. The problem that faced us, therefore, was which of the new variables to eliminate to both (1) minimize multicollinearity and (2) to do so while not significantly increasing the mean square error of prediction.[1] To do this, we eliminated the weakest predictor ($A_3 \cdot \mathrm{MBA}_2$) and then examined the ridge trace of the resulting model to see if acceptable stability resulted. This process was repeated until we obtained a satisfactory compromise to our dual objective. The order of elimination was (1) $A_3 \cdot \mathrm{MBA}_2$, (2) $A_2 \cdot \mathrm{UBA}_2$, and (3) $A_3 \cdot D$.

The ridge trace of the resulting final model is shown in Figure 2. As expected,

[1] For the large sample size involved in this problem the parameter estimates which minimize the mean square error of residuals also closely approximate minimization of the mean square error of prediction (Neter and Wasserman 1974).

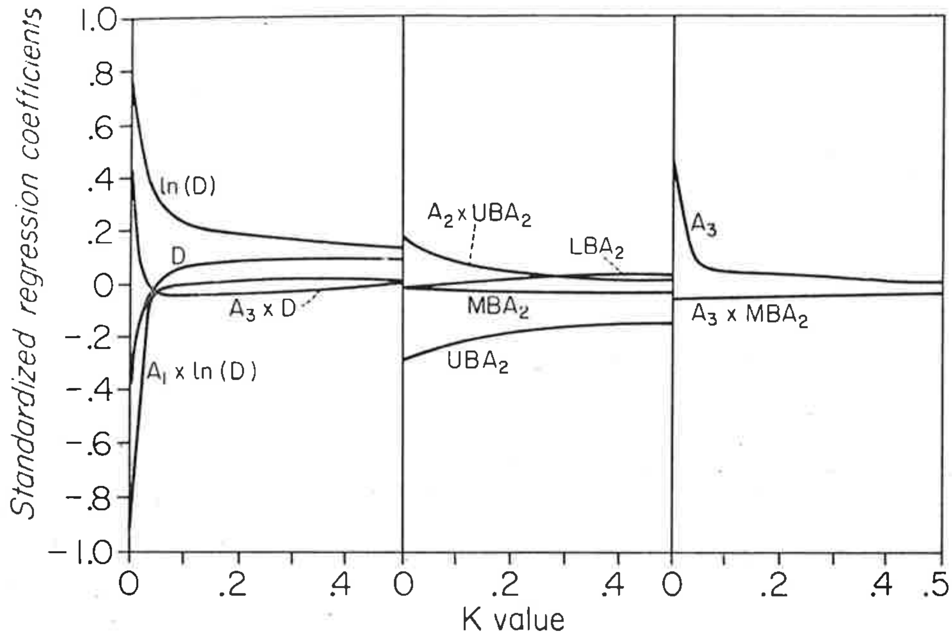

FIGURE 1. Ridge trace for blackjack pine before elimination of variables.

the system appears much more stable in this ridge trace. To verify the gain in stability for the OLS estimators we computed the maximum VIF for the original and final models. The reduction of the maximum VIF for these two models was from 260.9 to 23.02. This gain was obtained at the small cost of increasing the mean square error of residuals from 1.2679 to 1.2746. We could have reduced the

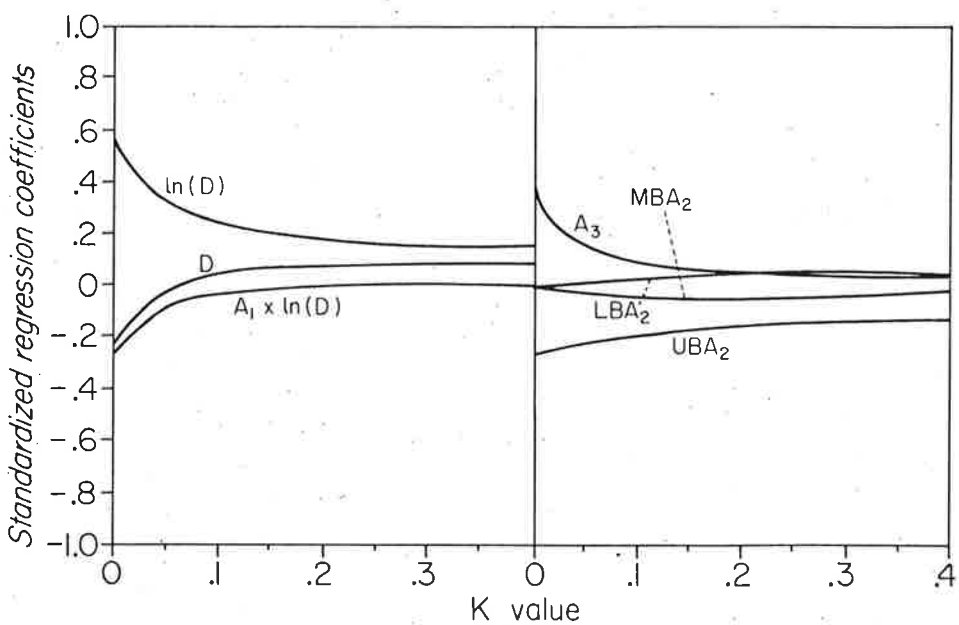

FIGURE 2. Ridge trace for blackjack pine after elimination of variables.

maximum VIF even further if RR estimators were used in place of their OLS counterparts. However, this was not done because our primary objective was to select variables for the final model and not to obtain precise estimates of the regression coefficients.

Our first example has shown how RR can be used to aid in the development of a predictive model. Admittedly, the use of the ridge trace introduces an element of subjectiveness. However, we believe that this is not necessarily objectionable nor often unavoidable in developing a regression model. During the model building process trade-offs must be weighed between *a priori* considerations of model form, the explanatory power of the model, the number of independent variables and the expense of measuring them. In this example our use of the ridge trace aided us in finding a meaningful model with stable and interpretable coefficients.

Ridge Regression as a Valuation Tool

Our second application demonstrates a different use of RR. In this case we work with a previously specified model and use RR to estimate a set of biased regression coefficients which possess small standard errors and appear stable when plotted over increasing values of K. This is important because in this application our primary objective is to interpret the regression coefficients. Thus, our objective is not to develop a prediction equation but to estimate the parameters of a previously specified model.

Bare forest land in western Oregon and Washington is currently subject to an annual ad valorem tax where the value is based on "current forest use" and not "highest and best use." An administering agency in each state is responsible for determining the bare land value for different site, location, and topographic classes. The data for the analysis are taken from actual sales of forest land. Although the objective of the valuation analysis is to estimate the true and fair value of bare forest land, most of the market evidence collected by the two agencies involves mature and immature timber value elements in addition to the land itself. Thus, the problem becomes one of allocating the gross sales price of a property to each of the separate value elements consisting of land, immature trees, and merchantable timber. It is analogous to determining the price of each item in a grocery basket when only the total sales price and the number of like items in the basket are known.

Instead of predicting the sales price of bare forest land as a function of site quality, location, and topography, we must estimate the per acre contribution of each value element present on a sale so that the total sales price is properly allocated. Site quality, location, and topography are used as determinants of this distribution but are not considered as separate value elements. The coefficient on the bare land value is of primary importance in this application.

Information collected for each valid forest land sale includes: (1) gross sales price observed in the market, (2) average site index, (3) total number of forest land acres, (4) number of acres of bare forest land, (5) number of acres of land stocked with immature conifer trees, (6) number of acres of land covered with brush, and (7) the volume of mature timber on the parcel.

Based on the data collected by the two administering agencies the following model is proposed:

$$\mathrm{GSP}_i = \beta_1[f(S)\mathrm{TA}_i] + \beta_2\mathrm{CON}_i + \beta_3\mathrm{BR}_i + \beta_4\mathrm{VOL}_i + E_i$$

where

GSP_i = Gross sales price of i^{th} forest land sale composed of bare land, immature conifer, brush and mature timber values.

TA_i = total number of forest acres on i^{th} sale.

TABLE 1. Weighted regression results of land valuation model (through origin).

Coefficient	Least squares		Ridge (k = 0.09)	
	Estimate	Standard error	Estimate	Standard error
β_1	0.87407	0.2456	0.97396	0.1204
β_2	319.03	60.8685	297.84	36.7988
β_3	−134.32	165.4315	−75.75	130.1441
β_4	192.00	14.5496	170.53	10.2135
	R^2 (Uncorrected for mean) = 0.92			
	$\bar{R}^2$ (Corrected for mean and degrees of freedom) = 0.85		$\bar{R}^2$ (Corrected for mean and degrees of freedom) = 0.84	
	MSE of residuals = 1.8426*10^8		MSE of residuals = 1.8985*10^8	
	Max VIF = 7.8291		Max VIF = 1.882	
	Determinant of correlation matrix = 0.0853		Determinant of correlation matrix = 0.2439	

CON_i = number of acres of immature conifer on i^{th} sale.
BR_i = number of acres of brush on i^{th} sale.
VOL_i = thousands of board feet (MBF) of mature timber on i^{th} sale.
E_i = error term.
β_2 = per acre value of immature conifer exclusive of land.
β_3 = per acre value of brush exclusive of land.
β_4 = per MBF value of mature timber.

The $f(S)$ term in the model quantifies the relationship of volume (or value) across different site classes. We have used two functional forms for this purpose: maximum mean annual cubic foot increment (MAI) and maximum soil expectation value. Both functions facilitate the derivation of a bare land value which is a function of site index. In this paper we present results only for the maximum MAI case. The parameter β_1 represents the value of bare land per unit of MAI for a given site class. The function used is

$$\text{Maximum MAI (CF)} = 0.244\ S^{1.33}$$

where S represents the 100-year base Douglas-fir site index. This function was derived from the Douglas-fir Managed Yield Simulator (Bruce and others 1977) using the optimization model proposed by Brodie and others (1978).

Previous experience with the model revealed that the variance of the error term increased with increasing parcel size. Thus, weighted least-squares was required to stabilize the variance. In the following weighting function:

$$W_i = 1/TA_i^c,$$

c was estimated to be 1.44 by the procedure described by Hann and McKinney (1975). A plot of residuals confirmed that this weighting function stabilized the variance. It further revealed that the model was correctly specified.

Using the above model, a weighted least-squares regression (through the origin) was performed on 123 forest land sales taken from the Douglas-fir zone. The results shown in Table 1 indicate that multicollinearity may be a problem because the determinant of the correlation matrix is not significantly different from zero at the 0.01 level of significance. While the maximum VIF is below Marquardt's (1970) critical value of ten, it is still large enough to cause some concern.

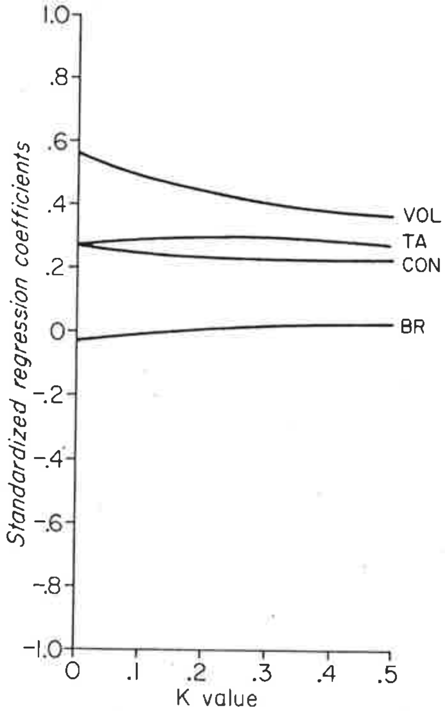

FIGURE 3. Ridge trace for forest valuation analysis.

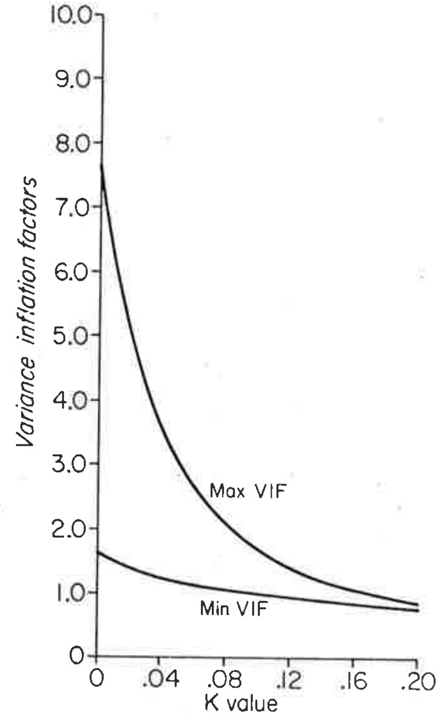

FIGURE 4. Variance inflation factors for forest valuation ridge regression.

To test the stability of the coefficients a weighted RR analysis was run. The ridge trace from this run is displayed in Figure 3. Although the regression coefficients are relatively stable it appears that a K value of approximately 0.10 would be appropriate. To help make the final decision the maximum and minimum VIF's for a range of K values were computed (Fig. 4). Because OLS VIF's range between one and infinity, we do not wish to drive the RR VIF below one during the ridge analysis. From Figure 4 the minimum VIF drops below one when K increases beyond 0.09. Thus, a K value of 0.09 was selected.

The weighted RR results for $K = 0.09$ are shown in Table 1. The determinant of 0.24 is now significantly different from zero at the 0.01 level of significance. Furthermore, the mean square error of residuals has only increased 3 percent.

As expected, the estimated per acre value of brush is negative (8–75.75/A). However, the large degree of variability associated with this estimate results in a near zero standardized regression estimate (Fig. 3). Two things contributing to this are (1) few sales include brush covered acres and (2) a high degree of variability in the condition and estimated cost of treating the brush exists in the market. Both factors result in a high degree of variability in the estimate of β_3. However, this variable was retained in the model because our objective is to allocate the gross sales price across the value elements present on each sale. Thus, the model would be misspecified if the variable was omitted.

An important result of the ridge analysis is the reduction in the standard error of the estimate of β_1. As shown in Table 1, the standard error of the ridged estimate of β_1 is about half the size of the weighted least-squares estimate. Accompanying this compaction of the distribution of error has been an increase in the ridge estimate of β_1 of about 11 percent. Unfortunately, no way exists to

TABLE 2. Bare land value estimates produced by weighted ridge regression model.

Site index (100 years)	200	170	140	110	85
Bare land value ($/A)	266	215	166	121	86

estimate the amount of expected bias unless the true population parameters are known. Brown and Beattie (1975) report results of simulation studies which indicate that the expected bias will be relatively small if (1) the explanatory variables are positively correlated, (2) the true β values have the same sign, and (3) the β's are about equal in magnitude. Conditions one and three were satisfied, but β_3, the per acre value of brush, was of opposite sign than the other β's. However, since β_3 is stable and close to zero, we believe that the amount of bias introduced into the analysis is relatively small. Supporting this conclusion is the small value of K needed to stabilize all regression coefficients.

Interpretation of Results

The main objective of the valuation problem discussed above was to estimate the per acre value of bare forest land. From the RR results shown in Table 1, a schedule of bare land values can be developed. These values, shown in Table 2, are obtained by multiplying the estimate of β_1 by $f(S)$. The resulting bare land values reflect average prices being paid for bare forest land in the Douglas-fir zone as of 1977.

The establishment of bare forest land values has been the subject of two recent court cases in Oregon and Washington. Consequently, considerable controversy surrounds the various methods which have been used for this purpose. Results produced by these methods have ranged from $65 to $240 per acre for site class 140. Most of these differences are the result of differing valuation techniques and/or assumptions. We are hopeful that the approach discussed in this paper will shed additional light on this subject and lead to an equitable resolution of the conflict.

Literature Cited

Andrews, D. F. 1974. A robust method for multiple linear regression. Technometrics 16:523–531.

Bartlett, M. S. 1950. Tests of significance in factor analysis. Br J Psychol, Stat Sect, 3:83.

Brodie, J. D., D. M. Adams, and C. Kao. 1978. Analysis of economic impacts on thinning and rotation for Douglas-fir, using dynamic programming. Forest Sci 24:513–522.

Brown, W. G., and B. R. Beattie. 1975. Improving estimates of economic parameters by use of ridge regression with production function applications. Am J Agric Econ 57:21–32.

Bruce, D., D. J. Demars, and D. L. Reukema. 1977. Douglas-fir managed yield simulator—DFIT User's Guide. USDA Forest Serv Gen Tech Rep PNW-57, 26 p. Pac Northwest Forest and Range Exp Stn, Portland, Oreg.

Chatterjee, S., and B. Price. 1977. Regression analysis by example. John Wiley and Sons, New York. 228 p.

Draper, N. R., and R. C. Van Nostrand. 1979. Ridge regression and James-Stein estimation: review and comments. Technometrics 21:451–466.

Farrar, D. E. and R. R. Glauber. 1967. Multicollinearity in regression analysis: the problem revisited. Rev econ and stat 49(1):92–107.

Haitovsky, Y. 1969. Multicollinearity in regression analysis: comment. Rev econ and stat 51(4):486–489.

Hann, D. W. 1980. Development and evaluation of an even- and uneven-aged ponderosa pine/Arizona fescue stand simulator. USDA Forest Serv Res Pap INT-267, 95 p. Intermt Forest and Range Exp Stn, Ogden, Utah.

HANN, D. W., and R. K. MCKINNEY. 1975. Stem surface area equations for four tree species of New Mexico and Arizona. USDA Forest Serv Res Note INT-190, 7 p. Intermt Forest and Range Exp Stn, Ogden, Utah.

HEMMERLE, W. J. 1975. An explicit solution for generalized ridge regression. Technometrics 17(3):309–314.

HEMMERLE, W. J., and T. F. BRANTLE. 1978. Explicit and constrained generalized ridge estimation. Technometrics 20(2):109–120.

HOCKING, R. R. 1976. The analysis and selection of variables in linear regression. Biometrics 32:1–49.

HOCKING, R. R., F. M. SPEED, and M. I. LYNN. 1976. A class of biased estimators in linear regression. Technometrics 18(4):425–437.

HOERL, A. E., and R. W. KENNARD. 1970a. Ridge regression: biased estimation for nonorthogonal problems. Technometrics 12(1):55–67.

HOERL, A. E., and R. W. KENNARD. 1970b. Ridge regression: applications to nonorthogonal problems. Technometrics 12(1):69–82.

HOERL, A. E., and R. W. KENNARD. 1976. Ridge regression: iterative estimation of the biasing parameter. Commun in Stat A5:77–88.

HOERL, A. E., R. W. KENNARD, and K. F. BALDWIN. 1975. Ridge regression: some simulations. Commun in Stat 4(2):105–123.

JENSEN, C. E., and J. W. HOMEYER. 1970. MATCH A CURVE—1 For algebraic transforms to describe sigmoid or bell shaped curves. USDA Forest Serv, Intermt Forest and Range Exp Stn, Ogden, Utah. 22 p.

KMENTA, J. 1971. Elements of econometrics. MacMillan, New York. 655 p.

MALLOWS, C. L. 1973. Some comments on Cp. Technometrics 15(4):661–675.

MARQUARDT, D. W. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12(3):591–612.

MARQUARDT, D. W., and R. D. SNEE. 1975. Ridge regression in practice. Am Stat 29(1):3–20.

MITCHELL, B. R., AND D. W. HANN. 1979. A computer program for applying ridge regression techniques to multiple linear regression. USDA Forest Serv Gen Tech Rep INT-51, 25 p. Intermt Forest and Range Exp Stn, Ogden, Utah.

NETER, J., and W. WASSERMAN. 1974. Applied linear statistical models. Irwin Inc., Homewood, Ill. 842 p.

OBENCHAIN, R. L. 1975. Ridge analysis following a preliminary test of the shrunken hypothesis. Technometrics 17(4):431–441.

VINOD, H. D. 1978. A survey of ridge regression and related techniques for improvements over ordinary least squares. Rev econ and stat 60(1):121–131.

WEBSTER, J. T., R. F. GUNST, and R. L. MASON. 1974. Latent root regression analysis. Technometrics 16:513–522.