ΣΠΟΥΔΑΙ / SPOUDAI

ΕΤΟΣ 1992 ΙΑΝ. - ΜΑΡΤΙΟΣ ΤΟΜΟΣ 42 ΤΕΥΧ. 1

YEAR 1992 JAN. - MARCH VOL. 42 No 1

A SIMULATION STUDY OF LEAST SQUARES AND RIDGE ESTIMATORS FOR SMALL SAMPLES

By

George S. Donatos and *George C. Michailidis*

Dept. of Economics University of Athens

Abstract

In this paper we consider the Least Squares (LS) estimator (predictor) and various ridge estimators (predictors) and report on a Monte Carlo study their small sample properties. The Monte Carlo experiment is applied to a residential electricity demand function with data from the Greek economy. On the basis of 2,500 replications of sample size 24 for normal disturbances we find that for the measures of dispersion the HKB estimator appears to be superior to the rest of the examined estimators. On the other hand the choice of alternative predictors for several measures of bias and dispersion is not clear. Furthermore, it should be noted that the small sample properties of the ridge estimators turn out to be different from the small sample properties of their respective predictors.

1. Introduction

Ridge regression is an alternative to Least Squares estimation in the multiple linear regression model, primarily dealing with the problem of multicollinearity. Ridge regression defines a class of estimators indexed by a biasing parameter k. Several algorithms have been proposed for k and tested through Monte Carolo simulations.

The purpose of this paper is to compare LS and five well known ridge estimators (predictors) according to measures of bias and dispersion and report their small sample properties. The Monte Carlo simulation is applied to a residential electricity demand function with real data, on the basis of four alternative forms of normal disturbances. In Section 2, the five ridge algorithms are defined. The design of the Monte Carlo experiment is outlined in Section 3, followed by the simulation results in Section 4. Some concluding remarks comprise the final section of the paper.

2. Ridge Estimators

Consider the linear regression model $y= X\beta+\varepsilon$ with $E(\varepsilon)= 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I$. The LS estimator is defined by $\hat{\beta}= (X'X)^{-1}X'y$. However, if the explanatory variables of the design matrix X are collinear then X'X is nearly singular and LS suffer from variance inflation, instability and incorrect signs of the estimates of the coefficients, that turn out to be critical in a real life application.

2. Ridge Estimators

Consider the linear regression model $y= X\beta+\varepsilon$ with $E(\varepsilon)= 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I$. The LS estimator is defined by $\hat{\beta}= (X'X) - 1X'y$. However, if the explanatory variables of the design matrix X are collinear then X' X is nearly singular and LS suffer from variance inflation, instability and incorrect signs of the estimates of the coefficients, that turn out to be critical in a real life application.

Hoerl and Kennard (1970) propose the class of estimators defined by

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'y, \ k>0 \qquad (1)$$

For suitable k, the ridge estimator dominates LS in mean square error (MSE) and improves the ill-contitioning of the matric X'X. The problem with this class of estimators is that the optimal value of k depends on the unknown parameter β, as well as the error variance σ^2.

Several algorithms for the biasing paramater k have been proposed in the literature. Five o these algorithms are evaluated in this paper and can be divided in the following categories: a) the MSE approaches, b) the Bayesian approaches and c) the constrained LS approaches.

In the first category, consider the more general ridge estimator suggested by Hoerl and Kennard (1970) and Goldstein and Smith (1974).

$$\hat{\beta}(k) = (X'X + G'KG)^{-1}X'y \qquad (2)$$

where K= diag (k_1. ..., k_p), p the number of explanatory variables. The MSE function is minimized at $k_i= s^2/\gamma_i^2$, where $\gamma = G'\beta$, $G'X'XG= \Lambda$ with Λ= diag (λ_i,..., λ_p), λ_i being the eigenvalues of X'X. Hoerl, Kennard and Baldwin (1975) propose the HKB estimator, using the harmonic mean of these k_i to obtain a single value of k. It is defined by

$$k = ps^2/\hat{\beta}'\hat{\beta} \tag{3}$$

where $$s^2 = (y-X\hat{\beta})'(y-X\hat{\beta})/(n-p) \tag{4}$$

Hoerl and Kennard (1976) suggest an iterative procedure of the HKB algorithm that corresponds to the HK estimator.

The second category represents Bayes ridge estimators provided that some conditions are satisfied. Lawless and Wang (1976) adopt a Bayes approach, which is known as the L & W estimator and estimate k (which is also the variance ratio of the samplin and prior distributions) by $ps^2/\Sigma\lambda_i\hat{\gamma}_i^2$, where $\hat{\gamma} = G'\hat{\beta}$ and s^2 is defined by (4). Dempster, Schatzoff and Wermuth (1977) propose to choose k in a way that

$$\Sigma\hat{\gamma}_i^2/(\sigma_\beta^2 + s^2/\lambda_i) = p \tag{5}$$

where $\sigma_\beta^2 = s^2/k$, and s^2 is defined by (4). This approach, that gives the RIDGM estimator, is implemented by evaluating (5) for a mesh of k values and choosing the one that most closely satisfies the equality.

Finally, in the third category, McDonald and Galarneau (1975) suggest to estimate

$$Q = \hat{\beta}'\hat{\beta} - s^2\Sigma\lambda_i^{-1} \tag{6}$$

If Q is positive this criterion is implemented by evaluating $\hat{\beta}'(k)\hat{\beta}(k)$ for a mesh of k values and choosing the one that minimizes the quantity $|\hat{\beta}'(k)\hat{\beta}(k) - Q|$. This procedure gives the Mc & G estimator. If Q is negative the procedure defaults to LS.

3. Design of the Experiment

The experiment involves a Monte Carlo simulation of a demand function for residential electricity in Greece. The demand function in matrix notation is given by:

$$C = Z\delta + \varepsilon \tag{7}$$

where $Z = [\iota\ Y\ P\ H\ N\ D\ E\ T_1\ T_2]$, C being the vector of per capital consumption of electricity, δ the coefficient vector to be estimated and ε the error term. Also, i is a vector of unit elements, Y the per capita personal disponsable income, P the marginal price of electricity, H an index of heating degree days with respect to consumption, D the average price of diesel fuel, N the number of electricity consumers, E the sales of electrical appliances and T_1, T_2 indicator variables for the two oil shocks of 1973-74 and 1978-79 respectively. The sample consists of 24 observations covering the period 1961-84.

The numerical values assigned to the elements of the vector $\delta = \delta^o$ are defined to be the "true" coefficients. The elements of δ^0 are obtained from the LS estimates of (7). Furthermore, 2,500 samples (size 24) of disturbances are obtained from independent normal numbers with zero mean and standard deviation given equal to 24.7 (standard error of the regression), 75, 150 and 225. Using (7) the values of the dependent variable C are computed. Having the generated observations of C and the values of the design matrix Z, LS and the five ridge algorithms of section 2 are applied. The predicted values $\hat{C}$ of C are then obtained by:

$$\hat{C} = Z \hat{\delta} \tag{8}$$

where $\hat{\delta}$ is the vector of the estimated coefficients of δ. Finally, given that the "true" parameters are known, since they are defined at the first stage of the experiment, the bias, MSE, and variance (Var) of each algorithm are computed (Donatos, 1989).

4. Comparison of Estimators

Given the small sample summary statistics of the bias, MSE, and variance for the LS and the five ridge estimators of the parameter coefficients (or the predictors of the mean of the dependent variable) of the demand function for residential electricity, we examined the small sample rankings of the estimators (predictors) for disturbances following forms of normal distribution. The results are given in Tables 1 and 2.

The LS estimator turns out to have the smallest bias and to be the least efficient in MSE, a finding expected and verified in other studies (Hoerl, Kennard and Baldwin (1975), Hoerl and Kennard (1976), Gibbons (1981), R. Hoerl, Schuenemeyer and A Hoerl (1986), Fomby (1987)).

The overall performance of the HKB estimator is good with respect to MSE, despite its high bias. Other studies, such as Hoerl, Kennard and Baldwin (1975) and Gibbons (1981), comment on its good performance as well. On the

Table 1
Small Sample Ranking of the Estimates of the Parameter Coefficients

Criterion of Ranking	Standard Deviation of Disturbance Distribution			
	24.7	75	150	225
Bias	1. LS	1. Mc&G	1. HK	1. LS
	2. Mc&G~	2. LS~	2. LS~	2. HK
	3. L&W	3. L&W	3. RIDGM~	3. L&W
	4. RIDGM	4. RIDGM	4. L&W	4. RIDGM
	5. HKB	5. HKB	5. Mc&G	5. Mc&G
	6. HK	6. HK	6. HKB	6. HKB
MSE	1. HKB	1. HKB	1. HK~	1. HK
	2. RIDGM	2. RIDGM	2. HKB	2. HKB~
	3. Mc&G	3. HK	3. RIDGM	3. L&W
	4. HK~	4. L&W	4. L&W	4. RIDGM
	5. L&W	5. Mc&G	5. Mc&G	5. Mc&G
	6. LS	6. LS	6. LS~	6. LS~
Var	1. HKB	1. HKB	1. HKB	1. HK
	2. HK	2. RIDGM	2. HK	2. HKB~
	3. RIDGM	3. HK	3. RIDGM	3. L&W
	4. Mc&G	4. L&W	4. L&W	4. RIDGM
	5. L&W	5. Mc&G	5. Mc&G	5. Mc&G
	6. LS	6. LS	6. LS	6. LS

Note: The ~ sign denotes that the "linked" estimators yield almost identical results.

other hand Lawless (1978) and Wichern and Churchill (1978) cast doubts on its use without further simulation studies.

The HK estimator is the most biased for small variances of the disturbances and the least biased for large variances of the disturbances. A similar pattern characterizes this estimator with respect to MSE. Hence, the present study casts doubts on the superiority of the HK estimator over HKB, especially for small variances of the disturbance term (Gibbons (1981)).

Table 2
Small Sample Ranking of the Predictors of the Mean of the Depended Variable

Criterion of Ranking	Standard Deviation of Disturbance Distribution			
	24.7	75	150	225
Bias	1. LS	1. LS	1. LS	1. LS
	2. HK	2. L&W	2. Mc&G	2. Mc&G
	3. RIDGM	3. RIDGM	3. HK	3. HKB
	4. Mc&G	4. Mc&G	4. L&W	4. L&W
	5. L&W	5. HKB	5. RIDGM	5. RIDGM
	6. HKB	6. HK	6. HKB	6. HK
MSE	1. Mc&G	1. HKB	1. HK	1. HK
	2. RIDGM	2. RIDGM	2. Mc&G	2. HKB~
	3. L&W	3. L&W	3. L&W	3. L&W
	4. HKB	4. Mc&G	4. HKB	4. Mc&G
	5. HK	5. HK	5. RIDGM	5. RIDGM
	6. LS	6. LS	6. LS	6. LS
Var	1. Mc&G	1. HKB	1. HK	1. HK
	2. RIDGM	2. RIDGM	2. Mc&G	2. HKB~
	3. L&W	3. L&W	3. L&W	3. L&W
	4. HKB	4. Mc&G	4. HKB	4. Mc&G
	5. HK	5. HK	5. RIDGM	5. RIDGM
	6. LS	6. LS	6. LS	6. LS

Note: The ~ sign denotes that the "linked" estimators yield almost identical results.

The RIDGM estimator has a satisfactory performance in both criteria, although it turns out to be rather inefficient for large variances of the disturbances. However, it is not superior to the rest of the examined estimators as reported by Dempster, Schatzoff and Wermuth (1977) and Gibbons (1981) in their respective extensive studies.

The L & W estimator does not take extreme positions with respect to the criteria of bias and dispersion. However, the L & W estimator exhibits a poorer performance than the one reported in the studies of Lawless and Wang (1976) and Lawless (1978).

Finally, the Mc & G estimator has a rather poor overall performance despite its good ranking with respect to the bias for small variances of the disturbances. A similar result is reported in Wichern and Churchill (1978).

Table 2 provides the basis for the following observations regarding the predictors of the mean values of the consumption variable. None of the ridge predictors can be considered better than the others in every case, with respect to the chosen criteria of bias and dispersion. That is, changing variances of the disturbances cause an instability in the ranking of the ridge predictors. However, the L&W and the Mc&G predictors perform relatively well, both for large and small variances of the disturbances. The HK predictor turns out to be the most efficient one for large variances of the disturbances and the least efficient for small variances. The opposite can be said regarding the efficiency of the RIDGM predictor. Ridge predictors have not been included in previous simulation studies and therefore it is impossible to compare our results. Finally, the LS predictor is inferior to all ridge predictors with respect to the MSE and variance criteria. On the other hand it is the least biased for all variances of the disturbances.

5. Concluding Remarks

Several extensive simulation studies have been conducted to compare ridge estimators along with the LS estimator and resulted in providing substantial evidence in favor of ridge estimators. However, lack of common elements in the design of these simulation studies makes it difficult to compare them with one another and with this study and reach a universal conclusion regarding the small sample properties of the ridge estimators. In view of these shortcomings it seems desirable to examine whether any ambiguities in previous studies can be cleared up. Specifically, it needs to be seen whether the findings of other studies are specific to the chosen structures or also hold for different structures and data sets.

In our simulation study we single out the HKB estimator for its overall good performance. The LS estimator (predictor) is dominated in MSE by the ridge estimators (predictors), as expected. No ridge predictor turns out to be superior than the other ones. Finally, the small sample properties of the ridge estimators are different from the small sample properties of the predictors, a result worth investigating in alternative experimental designs.

References

Dempster, A. Schatzoff, M., and Wermuth, N. (1977), "A Simulation Study of Alternatives to Least Squares", Journal of the American Statistical Association, 72, 77-106.

Donatos, G. (1989), "A Monte Carlo Study of k-class Estimators for Small Samples with Normal and Non-normal Disturbances". The Statistician, 38, 11-20.

Fomby, T. (1987), "The Ridge Regression Monte Carlo Controversy: Where Do We Stand?", in Advances in Econometrics, JAI Press, Inc. 17-49.

Gibbons, D(1981), "A Simulation Study of Some Ridge Estimators" Journal of the American Statistical Association, 76, 131-139.

Goldstein, M., and Smith, A. (1974), "Ridge Type Estimators for Regression Analysis", Journal of the Royal Statistical Society, Series B, 36, 284-291.

Hoerl, A. and Kennard, R, (1970), "Ridge Regression: Biased Estimation for Non-Orthogonal problems", Technometrics, 12, 56-67.

Hoerl, A., Kennard, R., and Baldwin K. (1975), "Ridge Regression: Some Simulations", Communications in Statistics, Theory and Methods,4, 105-123.

Hoerl, A., and Kennard, R. (1976), "Ridge Regression: Iterative Estimation of the Biasing Parameter", Communications in Statistics, Theory and Methods, 5, 77-88.

Hoerl, R., Schuenemeyer, J., and Hoerl, A, (1986), "A Simulation of Biased Estimation and Subset Regression Techniques", Technometrics, 28, 369-380.

Lawless, J. (1978), "Ridge and Related Estimation Procedures: Theory and Practice", Communications in Statistics, Theory and Methods, 7, 139-163.

Lawless, J. and Wang, P. (1976). "A Simulation Study of Ridge and Other Regression Estimators", Communications in Statistics, Theory and Methods, 5, 307-323.

Lindley, D., and Smith, A. (1972), "Bayes Estimates for the Linear Model", Journal of the Royal Statistical Society, Series B, 34, 1-41.

McDonald, G. and Galarneau, D. (1975), "A Monte Carlo Evaluation of Some Ridge Type Estimators", Journal of the American Statistical Association, 70, 407-416.

Wichern, D., and Churchill, G. (1978), "A Comparison of Ridge Estimators", Technometrics, 20,301-311.