

COMMUNICATIONS

DE LA FACULTÉ DES SCIENCES
DE L'UNIVERSITÉ D'ANKARA

Série: A Mathématiques

TOME : 33

ANNÉE : 1984

**A DISCRETE SHRINKING METHOD AS ALTERNATIVE TO
LEAST SQUARES**

by

F. ÖZTÜRK

22

Faculté des Sciences de l'Université d'Ankara
Ankara, Turquie

Communications de la Faculté des Sciences de l'Université d'Ankara

Comité de Redaction de la Série A₁

H. Hacısalihođlu – C. Kart – M. Balcı

Secrétaire de Publication

Ö. Çakar

La Revue "Communications de la Faculté des Sciences de l'Université d'Ankara" est un organe de publication englobant toutes les disciplines scientifique représentées à la Faculté des Sciences de l'Université d'Ankara.

La Revue, jusqu'à 1975 à l'exception des tomes I, II, III etait composé de trois séries

Série A: Mathématiques, Physique et Astronomie,

Série B: Chimie,

Série C: Sciences Naturelles.

A partir de 1975 la Revue comprend sept séries:

Série A₁: Mathématiques,

Série A₂: Physique,

Série A₃: Astronomie,

Série B: Chimie,

Série C₁: Géologie,

Série C₂: Botanique,

Série C₃: Zoologie.

A partir de 1983 les séries de C₂ Botanique et C₃ Zoologie on été réunies sous la seule série Biologie C et les numéros de Tome commencerons par le numéro 1.

En principe, la Revue est réservée aux mémoires originaux des membres de la Faculté des Sciences de l'Université d'Ankara. Elle accepte cependant, dans la mesure de la place disponible les communications des auteurs étrangers. Les langues Allemande, Anglaise et Française seraient acceptées indifféremment. Tout article doit être accompagné d'un résumé.

Les article soumis pour publications doivent être remis en trois exemplaires dactylographiés et ne pas dépasser 25 pages des Communications, les dessins et figuers portés sur les feuilles séparées devant pouvoir être reproduits sans modifications.

Les auteurs reçoivent 25 extrais sans couverture.

l'Adresse : Dergi Yayın Sekreteri,
Ankara Üniversitesi,
Fen Fakültesi,
Beşevler—Ankara
TURQUIE

A DISCRETE SHRINKING METHOD AS ALTERNATIVE TO LEAST SQUARES

F. ÖZTÜRK

University of Ankara, Faculty of Science, Department of Mathematics.

(Received September 17, 1984 and December 28, 1984)

SUMMARY

For the classical linear regression problem, a number of estimators alternative to least squares have been proposed for situations in which multicollinearity is a problem. This paper investigates mean square error properties of biased regression estimators and presents a procedure for obtaining improved estimators.

INTRODUCTION

One of the problems that can occur in regression analysis is multicollinearity among the independent variables. In the presence of multicollinearity the least squares estimates of regression coefficients are unstable in the sense that a different sample can produce dramatically different parameter estimates; it is possible that some will even have a wrong sign. As a result, estimation is not precise and determination of the relative influences of the individual variables is difficult. Also, the proper specification of the model is difficult to achieve when multicollinearity is present.

Hoerl and Kennard in their papers [3], [4] gave the theoretical basis for this difficulties and they presented a new estimation method—ridge regression. Ridge regression, based on adding a small quantity to the diagonal elements of a correlation matrix can reduce the variance of estimators, but at the expense of introducing bias. Existence of bias calls for mean square error as an adequate reliability measure.

In the next section we examine the effects of multicollinearity. Mean square error properties of biased estimators are given in Section

3. Section 4 presents a procedure for obtaining improved estimators. Section 5 contains the concluding remarks.

THE MODEL AND STATISTICAL CONSEQUENCES OF MULTICOLLINEARITY

The standard linear regression model may be written,

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\epsilon}, E(\underline{\epsilon}) = \underline{0}, E(\underline{\epsilon}\underline{\epsilon}') = \sigma^2 I_n \quad (1)$$

where, \underline{Y} is an $n \times 1$ vector of observations on a dependent variable, \underline{X} is an $n \times p$ matrix of nonstochastic regressors, $\text{rank}(\underline{X}) = p$, ($n > p$), $\underline{\beta}$ is a $p \times 1$ vector of unknown regression coefficients and $\underline{\epsilon}$ is an $n \times 1$ vector of unobservable disturbances. It is assumed that $\underline{X}'\underline{X}$ is in the form of a correlation matrix. The objective is to estimate $\underline{\beta}$. The parameter space is E^p — p dimensional Euclidean space.

The model (1) can be reduced to an orthogonal form by using $\underline{X} = \underline{U}\underline{\Lambda}^{1/2}\underline{V}'$, the singular value decomposition of \underline{X} , [8], [10]. Let $\underline{Z} = \underline{U}\underline{\Lambda}^{1/2} = \underline{X}\underline{V}$ and $\underline{\alpha} = \underline{V}'\underline{\beta}$. Then (1) becomes

$$\underline{Y} = \underline{Z}\underline{\alpha} + \underline{\epsilon}. \quad (2)$$

The type of estimators we select, depend upon the criterion we adopt to assess their performance. The least squares estimator of $\underline{\alpha}$ is,

$$\hat{\underline{\alpha}} = (\underline{Z}'\underline{Z})^{-1}\underline{Z}'\underline{Y} = \underline{\Lambda}^{-1}\underline{Z}'\underline{Y}. \quad (3)$$

In general,

$$\hat{\underline{\alpha}} = \underline{Z}^+\underline{Y} \quad (4)$$

where, $\underline{Z}^+ = \lim_{k \rightarrow 0^+} (\underline{Z}'\underline{Z} + kI_p)^{-1}\underline{Z}'$ is the Moore-Penrose pseudoinverse matrix of \underline{Z} , [1].

As is known $\hat{\underline{\alpha}}$ is minimum variance unbiased in the class of estimators $\{\tilde{\underline{\alpha}} : \tilde{\underline{\alpha}} = \underline{A}\underline{Y}, \underline{A}$ is a $p \times p$ matrix $\}$.

$$E(\hat{\underline{\alpha}}) = \underline{\alpha}$$

$$\text{Cov}(\hat{\underline{\alpha}}) = \sigma^2 \underline{\Lambda}^{-1}$$

$$\text{MSE}(\hat{\underline{\alpha}}) = E(\|\hat{\underline{\alpha}} - \underline{\alpha}\|^2) = E(\hat{\underline{\alpha}} - \underline{\alpha})' (\hat{\underline{\alpha}} - \underline{\alpha}) = \sigma^2 \sum_{i=1}^p (1/\lambda_i)$$

$$E(\|\hat{\underline{\alpha}}\|^2) = E(\|\hat{\underline{\beta}}\|^2) = \|\underline{\alpha}\|^2 + \sigma^2 \sum_{i=1}^p (1/\lambda_i)$$

However, in the presence of multicollinearity, that is in the presence of small eigenvalues λ_i , some of the least squares estimates have large variances, implying that least squares produces unreliable point estimates in the sense that the estimate may be far from the true value. In repeated sampling situations the estimates may vary substantially from one sample to another. Also $\text{MSE}(\hat{\underline{\alpha}})$ is going to be large, so $\|\hat{\underline{\alpha}}\|$ is going to be larger than $\|\underline{\alpha}\|$.

LINEAR BIASED ESTIMATORS

One approach to improve the least squares estimation procedure centers at finding biased estimators, which have smaller MSE's than the least squares estimators. The most popular of these is the ridge estimator proposed by Hoerl and Kennard [3], [4]. The shrunken least squares estimator and estimators based on principal components also fall into this category. These estimators are members of the class of estimators,

$$\hat{\underline{\alpha}}(B) = \{B\hat{\underline{\alpha}} : B = \text{diag}(b_1, b_2, \dots, b_p), 0 \leq b_i \leq 1, i=1, 2, \dots, p\} \quad (5)$$

The class so defined is all those estimators obtained by shrinking one or more of the components of $\hat{\underline{\alpha}}$. For these estimators,

$$E(B\hat{\underline{\alpha}}) = B\underline{\alpha}$$

$$\text{Cov}(B\hat{\underline{\alpha}}) = \sigma^2 B \Lambda^{-1} B'$$

$$\text{MSE}(B\hat{\underline{\alpha}}) = \sigma^2 \sum_{i=1}^p b_i^2 / \lambda_i + \sum_{i=1}^p (b_i - 1)^2 \alpha_i^2$$

$$E(\|B\hat{\underline{\alpha}}\|^2) = \|\underline{\alpha}\|^2 + \text{MSE}(B\hat{\underline{\alpha}})$$

Setting $\partial \text{MSE}(B\hat{\underline{\alpha}}) / \partial b_i = 0, i=1, 2, \dots, p$ we obtain the values b_i , that give minimum MSE,

$$b_i = \lambda_i / (\lambda_i + \sigma^2 / \alpha_i^2), i=1, 2, \dots, p \quad (6)$$

Steine-like shrunken estimators: For these estimators all the shrinkage factors are equal to a single constant, i.e.

$$B = bI, 0 \leq b \leq 1$$

Use of a constant shrinkage factor is most appropriate when the independent variables are orthogonal or very nearly so, [6], [11].

Principal component estimators: In principal component estimation the shrinkage factor for an individual component is equal to either zero or one. If $b_i = 0$ the component is deleted and if $b_i = 1$ the component is retained. There are numerous suggestions in the literature concerning appropriate criteria for retention or deletion of components, [2], [7], [9].

Ridge regression estimators: The generalized ridge estimator is defined by

$$\hat{\underline{\beta}}(\mathbf{K}) = (\mathbf{\Lambda} + \mathbf{K})^{-1}\mathbf{Z}'\underline{\mathbf{Y}} \quad (7)$$

where $\mathbf{K} = \text{diag}(k_1, k_2, \dots, k_p)$, $k_i \geq 0$, $i=1, 2, \dots, p$. If $\mathbf{K} = k\mathbf{I}$, $k \geq 0$ then the estimator is defined to be the simple ridge estimator or just ridge estimator.

(7) can be written

$$\hat{\underline{\beta}}(\mathbf{K}) = \mathbf{B}(\hat{\underline{\beta}}) \quad (8)$$

where $\mathbf{B} = \mathbf{\Lambda}(\mathbf{\Lambda} + \mathbf{K})^{-1}$. The values which give minimum MSE for generalized ridge are $k_i = \sigma^2/\alpha_i^2$, $i=1, 2, \dots, p$, [2], [3]. The properties, optimality conditions and problems associated with ridge estimators have been investigated in more than a hundred papers. Hua and Gunst, [5] in their paper extend generalized ridge regression to include negative values of the ridge parameter.

A DISCRETE SHRINKING METHOD AS ALTERNATIVE TO LEAST SQUARES

Let $\underline{\beta}^{(0)}$ be a fixed point in the parameter space E^p . Consider the sequence of estimators of $\underline{\beta}$, $(\hat{\underline{\beta}}^{(n)})$, which is defined as follows

$$\hat{\underline{\beta}}^{(n)} = (\mathbf{I} - \mathbf{h}\mathbf{X}'\mathbf{X})\hat{\underline{\beta}}^{(n-1)} + \mathbf{h}\mathbf{X}'\underline{\mathbf{Y}}, n=1, 2, \dots \quad (9)$$

where $\mathbf{h} \in \mathbf{R}$. Then

$$V' \hat{\beta}^{(n)} = (I - hV'X'XV)V' \hat{\beta}^{(n-1)} + hV'X'Y \tag{10}$$

Write $V' \hat{\beta}^{(0)} = \underline{\hat{\alpha}}^{(0)}$, $V' \hat{\beta}^{(n)} = \underline{\hat{\alpha}}^{(n)}$, $n=1,2,\dots$. Thus it follows that

$$\begin{aligned} \underline{\hat{\alpha}}^{(n)} &= (I - h \wedge) \underline{\hat{\alpha}}^{(n-1)} + h \wedge \underline{\hat{\alpha}} \\ &= (I - h \wedge)^2 \underline{\hat{\alpha}}^{(n-2)} + (I - h \wedge) h \wedge \underline{\hat{\alpha}} + h \wedge \underline{\hat{\alpha}} \\ &\dots \\ &= [I - (I - h \wedge)^n] \underline{\hat{\alpha}} + (I - h \wedge)^n \underline{\alpha}^{(0)}, n=1,2,\dots \end{aligned} \tag{11}$$

For $0 < h < 2/\lambda_{\max}$, $\|I - h \wedge\| < 1$ and $\lim_{n \rightarrow \infty} \underline{\hat{\alpha}}_i^{(n)} = \alpha_i, i=1,2,\dots,p$.

We may take any point in the parameter space as initial solution $\underline{\alpha}^{(0)}$. When doing this, our aim is to provide, that the estimators which we are going to select from the sequence $(\underline{\hat{\alpha}}^{(n)})$, to overcome the problems related to least squares estimator. Individual parameter estimator $\hat{\alpha}_i^{(n)}$, for $0 < h < 1/\lambda_{\max}$ can be considered as weighted average and for some n might be an effective compromise between $\underline{\alpha}^{(0)}$ and $\underline{\hat{\alpha}}$. In this paper we take as initial solution $\underline{\alpha}^{(0)} = 0$. Then, for $0 < h < 1/\lambda_{\max}$ we obtain the sequence of estimators $(\underline{\hat{\alpha}}^{(n)}(0))$, where

$$\underline{\hat{\alpha}}^{(n)}(0) = [I - (I - h \wedge)^n] \underline{\hat{\alpha}} \tag{12}$$

The estimator $\underline{\hat{\alpha}}^{(n)}(0)$ is a linear transformation of $\underline{\hat{\alpha}}$. In addition $\{\underline{\hat{\alpha}}^{(n)}(0), n=1,2,\dots\} \subset \underline{\hat{\alpha}}(B)$ and thus

$$\begin{aligned} E[\underline{\hat{\alpha}}^{(n)}(0)] &= \underline{\alpha} - (I - h \wedge)^n \underline{\alpha} \\ \text{Cov}[\underline{\hat{\alpha}}^{(n)}(0)] &= \sigma^2 [I - (I - h \wedge)^n]^2 \wedge^{-1} \end{aligned}$$

$$\text{MSE}[\underline{\hat{\alpha}}^{(n)}(0)] = \sigma^2 \sum_{i=1}^p [1 - (1 - h\lambda_i)^n]^2 / \lambda_i + \sum_{i=1}^p (1 - h\lambda_i)^{2n} \alpha_i^2$$

$$\|\underline{\hat{\alpha}}^{(n)}(0)\|^2 = \underline{\hat{\alpha}}' [I - (I - h \wedge)^n]^2 \underline{\hat{\alpha}} \leq \|\underline{\hat{\alpha}}\|^2$$

The first term in the expression for MSE is an increasing function of n and the second term is a decreasing function of n . A sufficient condition for MSE $[\underline{\hat{\alpha}}^{(n)}(0)]$ to be less than $\text{MSE}(\underline{\hat{\alpha}})$ is

$$n > \max_i \left\{ \left(\log \frac{2}{1 + \lambda_i \alpha_i^2 / \sigma^2} \right) / \log(1 - h\lambda_i), i=1,2,\dots,p \right\} \tag{13}$$

Therefore if $\|\underline{\alpha}\|$ is bounded there are values of n providing,

$$\text{MSE}[\hat{\underline{\alpha}}^{(n)}(0)] < \text{MSE}(\hat{\underline{\alpha}}).$$

For individual parameters we obtain,

$$\text{Bias}(\hat{\alpha}_i^{(n)}(0)) = E(\hat{\alpha}_i^{(n)}(0)) - \alpha_i = -(1-h\lambda_i)^n \alpha_i, \quad i=1,2,\dots,p$$

$$\lim_{n \rightarrow \infty} \text{Bias}(\hat{\alpha}_i^{(n)}(0)) = 0$$

$$|\text{Bias}(\hat{\alpha}_i^{(n)}(0))| > |\text{Bias}(\hat{\alpha}_i^{(n+1)}(0))|, \quad n=1,2,\dots$$

$$|\text{Bias}(\hat{\alpha}_i^{(n)}(0))| \leq |\text{Bias}(\hat{\alpha}_{i+1}^{(n)}(0))|, \quad i=1,2,\dots,p-1$$

$$\text{Var}(\hat{\alpha}_i^{(n)}(0)) = \sigma^2(1-(1-h\lambda_i)^n)^2 / \lambda_i, \quad i=1,2,\dots,p$$

$$\lim_{n \rightarrow \infty} \text{Var}(\hat{\alpha}_i^{(n)}(0)) = \text{Var}(\hat{\alpha}_i)$$

$$\text{Var}(\hat{\alpha}_i^{(n)}(0)) < \text{Var}(\hat{\alpha}_i^{(n+1)}(0)), \quad n=1,2,\dots$$

To summarize, the given procedure produces estimates $\hat{\underline{\alpha}}^{(n)}(0)$ which are biased, shorter than the least squares estimates, with smaller variance and which are closer to the true value of the coefficients, for a suitable chosen value of h , $h \in (0, 1/\lambda_{\max})$. However this procedure obeys the same difficulties associated with other biased estimation procedures, the choice of biasing parameters.

CONCLUSIONS

We have seen that biased, shrunken estimates can be used in the cases of multicollinearity to obtain smaller mean squared errors than least squares. It is necessary to provide some assurance that the benefit of reduced sample total variance of estimates is not likely to be offset by a large squared bias. The MSE's of the biased estimators depend not only on biasing parameters and the eigenvalues of $X'X$ but also on the unknown parameters of the model, so the optimal estimates cannot be obtained in practice. The members from the class of estimators $\hat{\underline{\alpha}}(B)$ considered, are more precise and offer alternatives for those who feel that multicollinearity has made their least squares results an unreliable basis for decision making purposes. However they do not offer complete solutions.

REFERENCES

- 1- **Albert, A.**, (1972) *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York.
- 2- **Hocking, R.R., Speed, F.M. and Lynn, M.J.**, (1976) "A Class of Biased Estimators in Linear Regression", *Technometrics*, 18, 425-437.
- 3- **Hoerl, A.E. and Kennard, R.W.**, (1970) "Ridge Regression: Biased Estimation of Nonorthogonal Problems", *Technometrics*, 12, 55-67.
- 4- **Hoerl, A.E. and Kennard, R.W.**, (1970) "Ridge Regression: Applications to Nonorthogonal Problems", *Technometrics*, 12, 69-82.
- 5- **Hua, T.A. and Gunst, R.F.**, (1983) "Generalised Ridge Regression: A Note on Negative Ridge Parameters", *Comm. Statist. -Theor. Meth.*, 12 (1), 37-45.
- 6- **James, W. and Stein, C.**, (1961) "Estimation with Quadratic Loss", *Proceedings of the Fourth Berkeley Symposium on Math. Statist. and Prob.*, Vol. 1, Berkeley: University of California Press, pp. 361-397.
- 7- **Lott, W.S.**, (1973) "The Optimal Set of Principal Component Restrictions on a Least Squares Regression", *Comm. Statist.*, 2, 449-464.
- 8- **Mandel, J.**, (1982) "Use of the Singular Value Decomposition in Regression Analysis", *The American Statistician*, 36, 1, 15-24.
- 9- **Massey, W.S.**, (1965) "Principal Components Regression in Exploratory Statistical Research", *JASA*, 70, 407-416.
- 10- **Rao, C.R.**, (1973) *Linear Statistical Inference and its Applications*, Wiley, New York.
- 11- **Sclove, S.L., Morris, C. and Radhakrishnan, R.**, (1972) "Non Optimality of Pre-Test Estimators for the Multinormal Mean", *Annals of Mathematical Statistics*, 43, 1481-1490.

ÖZET

İç ilişkinin sorun olduğu durumlarda, lineer regresyon problemi için en küçük kareler kestiricisi yerine birçok kestirici önerilmiştir. Bu çalışma yanlış kestiricilerin hata kareleri ortalaması özelliklerini incelemekte ve iyileştirilmiş kestiriciler elde etmek için bir yöntem sunmaktadır.