استخدام أسلوب تحليل القيمة الشاذة

في تقدير معلمة الحرف

د. حازم منصور كوركيس

جامعة بغداد- كلية التربية- ابن الهيثم

الملخص

في هذا البحث تم استخدام طريقة تحليل القيمة الشاذة في اقتراح أسلوب جديد لتقدير معلمة الحرف في مقدر انحدار الحرف والذي يعتبر البديل المناسب لمقدر المربعات الصغرى الاعتيادية عند ظهور مشكلة التعدد الخطي غير التام في نموذج الانحدار الخطي العام.

Abstract:

In this paper the method of singular value decomposition is used to estimate the ridge parameter of ridge regression estimator which is an alternative to ordinary least squares estimator when the general linear regression model suffer from near multicollinearity.

1- Introduction:

The linear regression model is one of the most widely used statistical models. It is used to model relations of one or more dependent variables to one or more explanatory variables. Although the ordinary least squares estimator (OLS) is the uniformly best unbiased estimator for the regression vector when the errors are iid normally distributed , there are situations when there are better estimators for the given problem even if the errors are iid normal. This may happen when multicollinearity occur. In this situation least squares estimators are unbiased but their variances are large so they may be far from the true value. One possibility to improve the OLS is to perform a ridge regression (RR) . In a ridge regression an additional parameter "the ridge parameter" is introduced . However there exist a number of different methods for choosing the ridge parameter. The one we tried is to estimate the ridge parameter by using the technique of singular value decomposition .

2- Singular Value Decomposition:

Given an (n×p) matrix of regressors X and an (n×1) vector of the corresponding responses Y, assume that sample means have been removed from the data and write the standard multiple linear regression model as:

Y=Xβ+u …….(1)

Where β is a (p×1) vector of unknown regression coefficients , u is the (n×1) vector of errors. The singular value decomposition is a fundamental way of studying the X data in the above regression model to get a deepoer understanding of our data which are often collinear. According to this way we decompose the matrix X into three matrices: X = H$D^{1/2}$G′ …….(2)

Where H is (n×p) semi orthogonal matrix of standardized "principal coordinates" of X $D^{1/2}$ is (p×p) diagonal matrix contains ordered singular values of X.

$d_1^{1/2} \geq d_2^{1/2} \geq \ldots \geq d_p^{1/2} > 0$ **……..(3)**

G is a (p×p) orthogonal matrix containing columns g_1 to g_p .The column g_i is a direction cosine vector which orient the i_th principal axis of X with respect to the given original axis of the X data. It is interesting that G also comes from eigenvalue eigenvector decomposition for the information matrix X′X it is given by:

X′X=G$D^{1/2}$H′H$D^{1/2}$G′=GDG′……….(4)

Thus the eigenvalues in the diagonal matrix D are squares of singular values. Assuming that b_{LS} is the ordinary least squares estimator, it is well known that

$b_{LS}=(x'x)^{-1}$ **x′y and also** b_{LS}~N[β,$\sigma^2(x'x)^{-1}$], **with the new notations:**

b_{LS}=$(GDG')^{-1}$G$D^{1/2}$H′y=GD^{-1}G′G$D^{1/2}$H′y or:

b_{LS}=G$D^{-1/2}$H′y=GC………(5)

Where C=$D^{-1/2}$H′y is the vector of uncorrelated components of the ordinary least squares estimator b_{LS}. From (5) we can write C=G′b_{LS} thus:

E(C)=E(G′b_{LS})=G′β =γ say…….(6)

The variance covariance matrix of C is given by:

Var (C) = var(G′b_{LS})=G′var(b_{LS})G=σ^2 G′$(GDG')^{-1}$G=σ^2G′GD^{-1}G′G = $\sigma^2 D^{-1}$……(7)

Since the off diagonal elements (covariances) are all zero we call them uncorrelated components.

3- Generalized Shrinkage Estimators:

The regression estimators of more interest to our exposition here are known as generalized shrinkage estimators. The vector of estimators for all p of the elements of the β coefficient vector in a linear model such as that in equation (1) will be denoted here by b_{SH} and will be of the form: b_{SH}=GΔC= $\sum_{j=1}^{p} g_j \delta_j c_j$ ……….(8)

Where g_j is the j_th column of the principal axis direction cosines matrix G , δ_j is the j_th diagonal element of the shrinkage factors matrix Δ and c_j is the j_th element of the uncorrelated components vector, C . The range of shrinkage factors in equation (8) δ_1,….,δ_p is usually restricted as: 0≤δ_i≤1,i=1,2,….,p. The generalized shrinkage estimator corresponding to δ_1=,….,=δ_p=1 (i.e Δ=I) coincides with the ordinary least squares estimator.

4- The Case Of Multicollinearity:

The problem of multicollinearity exists when there exist a linear relationship or an approximate linear relationship among two or more explanatory variables. The problem of multicollinearity is among the most intractable of the problems that plague regression analysts.It can be thought of as a situation where two or more explanatory variables in the data set move together. As a consequence it is impossible to use this data set to decide which of the explanatory variables is producing the observed change in the response variable. Thus multicollinearity is a problem in which the available data is inadequate to give us the desired answer.

5- Ridge Regression Estimators:

Hoerl and Kennard (1970) proposed the use of ridge regression to estimate β when the explanatory variables are highly correlated. The basic idea is to reduce the variance by shrinking the estimator so that the mean squares error MSE can be reduced. To achieve that in a ridge regression,an additional parameter k is added to the OLS estimation problem.

$$b_{RR}=(x'x+kI)^{-1}x'y \ , \ k\geq 0 \ ....(9)$$

If k=0 the resulting estimator is the OLS estimator for β. It can be shown that the ridge regression estimator given in (9) is special case of generalized shrinkage estimator given in (8) as follows: By using matrix algebra and singular value decomposition of matrix X we obtain: $b_{RR}=[G(D+kI)G']^{-1}GD^{1/2}H'y$

$$=G(D+kI)^{-1}\,G'GD^{1/2}H'y$$

$$=G(D+kI)^{-1}D^{1/2}H'y$$

$$=G[(D+kI)^{-1}D]D^{-1/2}H'y = G\Delta C \quad ....(10)$$

Where $\Delta = (D+kI)^{-1}D$. Equivalently,the shrinkage factors δ_j of ridge estimator have the form:

$$\delta_j=\frac{d_j}{d_j+k} \;, \quad 1\le j\le p \quad ....(11)$$

6- Proposed Method For Estimating The Ridge Parameter:

By using the MSE for generalized shrinkage estimator we can derive a new method for estimating the ridge parameter as follows:

$$MSE(b_{SH})=MSE(G\Delta C)=GMSE(\Delta C)G'$$

$$\text{Where: } MSE(\Delta C)=\sigma^2\Delta^2D^{-1}+(I-\Delta)\gamma\gamma'(I-\Delta) \quad ....(12)$$

Is the mean squared error matrix of ΔC . It is the sum of two matrices, namely:

1- The diagonal variance matrix , $\sigma^2\Delta^2D^{-1}$ which is void of covariance terms, because the components of ΔC are uncorrelated.

2- The matrix $(I-\Delta)\gamma\gamma'(I-\Delta)$,with squares bias terms along its main diagonal and bias cross product terms off that diagonal.

Let us focus upon any single diagonal element,say the i_th of the mean squared error matrix of ΔC.

$$MSE(\delta_i c_i)=\sigma^2\delta_i^2/d_i+(1-\delta_i)^2\gamma_i^2 \quad ....(13)$$

Now MSE($\delta_i c_i$) of (13) will clearly change as the i_th δ factor changes. In fact the partial derivative of MSE($\delta_i c_i$) with respect to δ_i is:

$$\frac{\partial MSE\,(\delta_i c_i)}{\partial \delta_i} = 2\sigma^2 \delta_i/d_i - 2(1-\delta_i)\gamma_i^2 \quad \text{………..(14)}$$

While the second partial derivative is a nonnegative constant.

$$\frac{\partial^2 MSE(\delta_i c_i)}{\partial \delta_i^2} = \frac{2\sigma^2}{d_i} + 2\delta_i^2 \quad \text{………..(15)}$$

It follows from (15) that equating $\frac{\partial MSE\,(\delta_I C_I)}{\partial \delta_i}$ in (14) to zero will yield a minimum value for MSE($\delta_i c_i$) as long as either σ^2>0 or δ_i^2>o, this optimal amount of shrinkage for the i_th uncorrelated component c_i is:

$$\delta_i^{MSE} = \frac{\gamma_I^2}{\gamma_I^2 + (\frac{\sigma^2}{d_i})} = \frac{d_i}{d_i + \frac{\sigma^2}{\gamma_i^2}} \quad \text{…………….(16)}$$

Our proposed method for estimating the ridge parameter $\hat{k}$ is summarized by comparing the shrinkage factor of ridge regression estimator given in equation (11) with δ_i^{MSE} given in equation (16) .Accordingly,we conclude that the value of k must equal to σ^2/γ_i^2 .Since each of σ^2 and γ_i^2 is unknown, we can use their estimated values. This yield:

$$\hat{k} = \frac{s^2}{b_{LS}' G' G b_{LS}/p} = \frac{ps^2}{b'_{LS} b_{LS}} \quad \text{…………..(17)}$$

Where s^2 is the residual mean square in the analysis of variance table obtaind from the standard least squares fit. Hoerl and Kennard (1976) proposed an iterative method for selecting the ridge parameter k the method is based on the formula:

$$k = \frac{ps^2}{b'_{RR} b_{RR}} \quad \text{……………(18)}$$

To obtain the first value of k we use the least squares coefficients.In this case the iterative method in (18) coincides with our proposed method in (17).

7- Numerical Example:

An example of data appropriate for this procedure is shown below.These data were concocted to have a high degree of multicollinearity as follows.We put a sequence of numbers in x_1.Next we put another series of numbers in x_3 that were selected to be unrelated to x_1. We created x_2 by adding x_1 and x_3 . We made a few changes in x_2 so that there was not perfect correlation. Finally, we added all three variables and some random error to form y.

x_1	x_2	x_3	y
1	2	1	3
2	4	2	9
3	6	4	11
4	7	3	15
5	7	2	13
13	6	7	1
17	7	8	1
8	10	2	21
9	12	4	25
10	13	3	27
11	13	2	25

12	13	1	27
13	14	1	29
14	16	2	33
15	18	4	35
16	19	3	37
17	19	2	37
18	19	1	39

Pearson correlation coefficients given in the correlation matrix below show which independent variables are highly correlated with the dependent variable and with each other. Independent variables that are highly correlated with one another may cause multicollinearity problems.

	x_1	x_2	x_3	y
x_1	1.000000	0.987841	-0.015051	0.985544
x_2	0.987841	1.000000	0.133813	0.995574
x_3	-0.015051	0.133813	1.000000	0.116539
Y	0.985544	0.995574	0.116539	1.000000

A quick summary of the various statistics that might go into the choice of the ridge parameter k is given below:

K	R^2	s	$b_{RR}'b_{RR}$	Ave VIF	Max VIF
0.000000	0.9915	1.1028	1.4905	324.9567	485.8581
0.010000	0.9857	1.4349	0.5002	3.3071	4.4637
0.020000	0.9807	1.6639	0.4891	1.2575	1.4055

0.030000	0.9759	1.8619	0.4830	0.8309	0.9372
0.040000	0.9711	2.0389	0.4778	0.6711	0.9146
0.050000	0.9663	2.2000	0.4729	0.5918	0.8951
0.060000	0.9616	2.3487	0.4682	0.5450	0.8771
0.066237	0.9587	2.4361	0.4653	0.5244	0.8664
0.070000	0.9570	2.4871	.4636	0.5140	0.8602
0.080000	0.9523	2.6170	0.4591	0.4914	0.8439
0.090000	0.9478	2.7396	0.4547	0.4739	0.8283
0.100000	0.9432	2.8558	0.4503	0.4597	.8131

The values of k in column (1) are the actual values. The value found by the analytic search (0.066237) sticks out because it does not end to zeros. The values of coefficient of determination R^2 are given in column (2). Since the least squares solution maximizes R^2 so

the largest value of R^2 occurs when k=0. We want to select a value of k that does not stray very much from this value. S in column (3) is the square root of the mean squared error. Leasat squares minimizes this value, so we want to select a value of k that does not stray very much from the least squares value. In column (4) the sum of the squared standardized ridge regression coefficients are given. We want to find a value of k at which this value has stabilized. The average and maximum variance inflation factors are given in columns (5) and (6) respectively. The variance inflation factor (VIF) is a measure of multicollinearity, it is the reciprocal of 1-R^2 . We treat any VIF in excess of 10 as evidence of multicollinearity.

8- Conclusions:

In this paper the ridge estimators have been viwed as a subclass of the class of generalized shrinkage estimators. A new method for estimating the ridge parameter was proposed by using the singular value decomposition approach. The proposed method depend on the fact that the shrinkage factor can be choosen which will guarantee the ridge estimator has mean square error smaller than the variance of the least squares estimator.

References:

1- Chatterjee,S.and Price,P.(1977) :Regression analysis by example,john wiley and sons, New York.

2- Drapper,N.R.and Smith,H.(1981) :Applied regression analysis,second edition.john wiley and sons,New York.

3-George,E.I.(1986): minimax multiple shrinkage stimation.Ann.Statist.14,188_205.

4- Goldstein ,M.and Smith,A.F.M.(1974):Ridge type estimators for regression analysis.J.Roy. Statist.B.36,284_291.

5- Hoerl,A.E.and Kennard,R.W.(1970): Ridge regression,biased estimation for non orthogonal problems.Technometrics,12, 55_67.

6- Obenchain,R.L.(1978): Good and optimal ridge estimators.Ann.Statist. 6, 1111_1121.

7- Rao,C.R.(1965): Linear statistical inference and its applications.New York,Wiley.

8- Smith,G.and Campbell,F.(1980): A critique of some ridge regression methods, J.Amer.Statist.Assoc.75,74_81