行政院國家科學委員會補助專題研究

模糊分析在脊迴歸上的應用
(An Application of Fuzzy Analysis Method on Ridge Regression Model)

計畫類別：個別型計畫

計畫編號：NSC 89－2118－M－032－003－

執行期間：88 年 8 月 1 日至 89 年 7 月 31 日

計畫主持人：蔡宗儒

本成果報告包括以下應繳交之附件：

- □赴國外出差或研習心得報告一份
- □赴大陸地區出差或研習心得報告一份
- □出席國際學術會議心得報告及發表之論文各一份
- □國際合作研究計畫國外研究報告書一份

執行單位：淡江大學統計學系

中 華 民 國 89 年 9 月 22 日

行政院國家科學委員會專題研究計畫成果報告

模糊分析在脊迴歸上的應用

An Application of Fuzzy Analysis Method on Ridge Regression Model

計畫編號：NSC 89-2118-M-032-003

執行期限：88 年 08 月 01 日至 89 年 07 月 31 日

主持人：蔡宗儒　執行機構及單位名稱:淡江大學統計學系

一、中文摘要

當共線性及離群值同時存在一個迴歸模型中時，吾人提出了一個模擬脊迴歸估計法，並證明在模糊權數被給定下，模糊脊迴歸估計量的均方誤會比加權最小迴歸估計量的均方誤來的小。

Abstract

When a data set contains outliers and the explanatory variables in a linear regression model are highly correlated, a fuzzy-weighted ridge regression (FWRR) estimation method is provided to combat multicollinearity and reduce the influence of outliers. Moreover, I prove that the FWRR estimator has smaller mean squared error than the weighted least squares estimator. Some Monte Carlo simulations and an example are provided to demonstrate the application of FWRR estimator.

Keywords: Ridge Regression; Jackknife Method; Multicollinearity; Optimal Fuzzy Clustering Analysis Method; Degree of Membership; Outlier.

1. Introduction

Consider the linear regression model:

$$Y = \mathbf{1}\beta_0 + X\beta + \varepsilon, \qquad (1)$$

where $Y^T = (Y_1, Y_2, \cdots, Y_n)$ is an $n \times 1$ vector of response variables; X is an $n \times p$ centered matrix of explanatory variables; $\mathbf{1}$ is an n-dimensional column vector of ones; β_0 is an unknown parameter; β is a $p \times 1$ vector of unknown parameters; ε is a column vector of random error terms with $E(\varepsilon) = \mathbf{0}$ and $Cov(\varepsilon) = \sigma^2\mathbf{I}$. We can use the sample mean $\bar{Y}$ to estimate the unknown parameter β_0. When the explanatory variables in model (1) are highly correlated, the ordinary least squares (OLS) estimator $\hat{\beta} = (X^T X)^{-1} X^T Y$ becomes unstable and is inadequate to interpret the relationship between response variable and explanatory variables. Hoerl and Kennard (1970) proposed a ridge regression (RR) estimation method that can generate a smaller total variance than the OLS estimator. Usually, the RR estimators are biased. Liu (1993) provided a new RR estimator to estimate the regression coefficients. The RR estimator proposed by Liu (1993) is a linear function of an unknown biasing constant. Thus, the computation of RR estimator proposed by Liu (1993) is easier than the RR estimator proposed by Hoerl and Kennard (1970). Shia and Chow (1998) used jackknife method to reduce the bias of RR estimator. According to their simulations, the proposed method exhibited a better forecasting feature than the RR estimation proposed by Hoerl and Kennard (1970) both in stability and precision. These methods proposed by Hoerl and Kennard (1970), Liu (1993) and Shia and Chow (1998) are discussed when a data set does not contain outliers. If the data set contains outliers and the explanatory variables in model (1) are highly correlated, then the precision of RR estimator will be affected. Walker and Birch (1988) discussed some influence measures in RR estimation. Van Cutsem and Gath (1993) and Wu *et al.* (1996) proposed an optimal fuzzy clustering analysis (OFCA) method which can identify outliers efficiently in multivariate data sets. In order to reduce the influence of outliers, they suggested that the normalized degrees of membership of optimal fuzzy clustering analysis method could be added into the estimation procedure. Moreover, Wu *et al.* (1996) provided a fuzzy-weighted regression coefficient estimator in semi-parametric model, which can reduce the influence of outliers in parameter estimation efficiently. Hence, a FWRR estimation procedure combining the optimal fuzzy clustering analysis method and the RR estimation method proposed by Hoerl and Kennard (1970) is provided. The new estimation procedure can improve the precision and stability of RR estimator when a data set contains outliers.

2. FWRR Estimation

Assume that X_i is the ith row of X in model (1) and the observation $Z_i = (X_i, Y_i)$ is the ith (p+1)-dimensional data vector. If one observation, say Z_l, is an outlier, $1 \le l \le n$, then the RR estimator will produce large fluctuation after Z_l is dropped from the data set. Hence, a hold-out procedure can be used to help with identifying outliers. In each time, we delete one observation from the data set, and then to compute the RR estimator based on the remaining n-1 observations. Let $\hat{\beta}_R^{(l)}$ be the RR estimator when lth observation Z_l is dropped, $l = 1,2,\cdots,n$. Assume that there are some outliers in the original data set and $\hat{\beta}_R^{(1)}, \hat{\beta}_R^{(2)}, \cdots, \hat{\beta}_R^{(n)}$ can be grouped into q clusters, $1 \le q \le n$. Without loss of generality, suppose that the number of observations of first cluster is the largest one among all clusters and is denoted by n_1. Suppose the condition $n_j << n_1$ holds, where n_j is the number of observations in the jth cluster, $j = 2,\cdots,q$. That is, I treat the cluster 1 as the main group and the other groups are the outlier groups.

When the optimal fuzzy clustering method is performed, the degrees of membership in the main classified cluster $u^T = (u_{11}, u_{12}, \cdots, u_{1n})$ can be normalized to be the fuzzy weights w_i, $i = 1,2,\cdots,n$ such that $\sum_{i=1}^{n} w_i = 1$. Then, we will use these fuzzy weights to do the FWRR estimation

Assume that model (1) holds and the data set contains some outliers. The variance-covariance matrix of ε is revised as $Cov(\varepsilon) = \Phi\sigma^2$, where Φ is a diagonal matrix. When RR estimation is used in model (1), we often use correlation transformation to help with controlling roundoff errors such that the regression coefficients have the same unit. Let

$$s_y = \sqrt{\frac{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}{n-1}}$$ and

$$s_j = \sqrt{\frac{\sum_{i=1}^{n}(X_{ij}-\overline{X}_j)^2}{n-1}}, \qquad j=1,2,\cdots,p.$$

The correlation transformations of Y and X are given as follows,

$$\widetilde{Y}_i = \frac{1}{\sqrt{n-1}}\left(\frac{Y_i-\overline{Y}}{s_y}\right) \qquad (2)$$

and

$$\widetilde{X}_{ij} = \frac{1}{\sqrt{n-1}}\left(\frac{X_{ij}-\overline{X}}{s_j}\right), \qquad (3)$$

$j=1,2,\cdots,p$, $i=1,2,\cdots n$. Using the correlation transformation, we can rewrite model (1) as

$$\widetilde{Y} = \widetilde{X}\beta' + \varepsilon', \qquad (4)$$

where $\widetilde{Y}$ is an $n\times 1$ vector with entries $\widetilde{Y}_i$, $i=1,2,\cdots,n$; $\widetilde{X}$ is an $n\times p$ matrix with entries $\widetilde{X}_{ik}$, $i=1,2,\cdots,n$, $k=1,2,\cdots,p$; β' is a $p\times 1$ vector of unknown parameters and ε' is an $n\times 1$ vector of error terms. The OLS estimators in model (10) is denoted by $\widetilde{\beta}=(\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^T\widetilde{Y}$ $=(\widetilde{\beta}_1,\widetilde{\beta}_2,\cdots,\widetilde{\beta}_p)^T$ and the OLS estimators in model (1) also can be computed by using $\hat{\beta}_j=\frac{s_y}{s_j}\widetilde{\beta}_j$, $j=1,2,\cdots,p$ and $\hat{\beta}_0=\overline{Y}$ $-\hat{\beta}_1\overline{X}_1-\cdots-\hat{\beta}_p\overline{X}_p$. Suppose that the data set contains outliers and the fuzzy weights w_i, $i=1,2,\cdots,n$ are computed by using optimal fuzzy clustering analysis method. Let W be a $n\times n$ diagonal weighted-matrix with fuzzy weights w_i, $i=1,2,\cdots,n$ in the main diagonal. Wu *et al.* (1996) showed that the fuzzy-weighted least squares (FWLS) estimator $\widetilde{\beta}_w=(\widetilde{X}^TW\widetilde{X})^{-1}\widetilde{X}^TW\widetilde{Y}$ is more stable than the OLS estimator $\widetilde{\beta}$ when the data set contains outliers. In addition, if the design matrix $\widetilde{X}^T\widetilde{X}$ is close to singular, Hoerl and Kennard (1970) suggested that we could add a positive number k to the main diagonal of matrix $\widetilde{X}^T\widetilde{X}$ when the OLS estimators are computed. This estimator $\widetilde{\beta}_R=(\widetilde{X}^T\widetilde{X}+kI)^{-1}\widetilde{X}^T\widetilde{Y}$ is called the RR estimator. Mathematically speaking, the RR estimator is biased. In fact, when we use the RR estimation, we get a stable but biased estimator. Many computing procedures are available for choosing the value of biasing constant k (see, e.g. Horel and Kennard 1976; Myers 1986, pp. 247-262; Neter *et al.* 1996, pp. 412-416).

When the data set contains outliers and the design matrix $\widetilde{X}^T\widetilde{X}$ is close to singular, I revise the RR estimator $\widetilde{\beta}_R$ and provide a FWRR estimator $\widetilde{\beta}_{wR}=(\widetilde{X}^TW\widetilde{X}+kI)^{-1}\widetilde{X}^TW\widetilde{Y}$, where W is a diagonal matrix with fuzzy weights on the main diagonal. Moreover, we can use the prediction-oriented criterion proposed by Mayers (1986) to find the biasing constant k. This consists of selecting the value of k such that

$$C_k = \frac{RSS_k}{\hat{\sigma}_w^2} - n + 2tr(H_k^*) \qquad (5)$$

is minimal, where $RSS_k=(\widetilde{Y}-\widetilde{X}\widetilde{\beta}_{wR})^TW(\widetilde{Y}-\widetilde{X}\widetilde{\beta}_{wR})$ is the weighted sum of squares of residuals by using FWRR estimation and $\hat{\sigma}_w^2=\frac{1}{n-p}(\widetilde{Y}-\widetilde{X}\widetilde{\beta}_w)^TW(\widetilde{Y}-\widetilde{X}\widetilde{\beta}_w)$ is the weighted estimator of σ^2 by using FWLS estimation, $H_k^*=\widetilde{X}(\widetilde{X}^TW\widetilde{X}+kI)^{-1}X^TW$ and $tr(H_k^*)$ is the trace of matrix H_k^*. The matrix H_k^* plays the same roles as the hat matrix in OLS estimation, however, the matrix H_k^* is not a projection matrix. When the fuzzy weights w_i, $i=1,2,\cdots,n$ in matrix W are fixed, we can get the following theorem.

Theorem 1: Let $E(\varepsilon)=0$ and $Cov(\varepsilon) = \Phi\sigma^2$ in the linear regression model $Y = \mathbf{1}\beta_0 + X\beta + \varepsilon$, where Φ is a diagonal matrix. If the weights w_i, $i=1,2,\cdots,n$ are given, there always exists a positive k such that $MSE(\hat{\beta}_{wR}) \le MSE(\hat{\beta}_w)$.

The proof of Theorem 1 is given in Tsai (1999). Monte Carlo simulations are provided to assess the performance of FWRR estimator.

3. Monte Carlo Simulations and Example

3.1 Monte Carlo Simulations

In order to evaluate the performance of FWRR estimator, 100 regression data sets are simulated at different level of multicollinearity with outliers. The sample size in each simulated data set is 20 and each data set contains 4 outliers. Assume that the regression model in the main group is

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i, \quad i=1,2,\cdots,16, \qquad (6)$$

We use the simulation procedure suggested by McDonald and Galarneau (1975) to generate the explanatory variables. Let $z_{i1}, z_{i2}, \cdots, z_{i6}$ are normally distributed with mean 0 and standard deviation 1, and let

$$x_{ij} = (1-r^2)^{1/2} z_{ij} + r z_{i6}, \quad i=1,2,\cdots,16, \quad j=1,2,\cdots,5. \qquad (7)$$

It is easy to show that the correlations of any two explanatory variables equal to r^2. For $i=1,2,\cdots,16$, the data set is simulated according to model (6) and equation (7). Moreover, the error terms ε_i, $i=1,2,\cdots,16$ are generated from standard normal distribution. For the remaining four observations (outlier group), the data set is simulated from model (8) as follows:

$$Y_i + 3 = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \varepsilon_i, \quad i=17,18,19,20. \qquad (8)$$

In model (8), the equation (7) is used to simulate the explanatory variables. However, the common distribution of $z_{i1}, z_{i2}, \cdots, z_{i6}$ is normal with mean 3 and standard deviation 1 and the error terms ε_i's are generated from normal distribution with mean 0 and standard deviation 0.5. That is, the last 4 observations are outliers. In this paper, the values of r^2 are considered as 0.9 and 0.95. Moreover, let $\beta_1 = \beta_2 = \cdots = \beta_5$ and the standardizable parameters $\beta_{i*} = \frac{\beta_i}{\sqrt{\beta^T\beta}} = \frac{1}{\sqrt{5}}$, $i=1,2, \cdots,5$ are used. Set $\beta_*^T = (\beta_{1*}, \beta_{2*}, \beta_{3*}, \beta_{4*}, \beta_{5*})$. Two criteria are used to evaluate the performances of the FWRR estimator $\hat{\beta}_{wR}$ and the RR estimator $\hat{\beta}_R$. The first criterion is the mean squared error (*MSE*). The *MSE* of FWRR estimator $\hat{\beta}_{wR}$ is

$$MSE(\hat{\beta}_{wR}) = \frac{1}{100}\sum_{j=1}^{100} (\hat{\beta}_{wR}(j) - \beta_*)^T (\hat{\beta}_{wR}(j) - \beta_*),$$

where $\hat{\beta}_{wR}(j)$ is the FWRR estimator in j simulated sample, $j=1,2,\cdots,100$. The *MSE* of RR estimator $\hat{\beta}_R$ is the same as $MSE(\hat{\beta}_{wR})$ except the estimator $\hat{\beta}_{wR}$ is replaced by $\hat{\beta}_R$. The second criterion is the sum of squares of biases (*SSB*).

$$SSB(\hat{\beta}_{wR}) = \left(Bias(\hat{\beta}_{wR})\right)^T \cdot Bias(\hat{\beta}_{wR}),$$

where $Bias(\hat{\beta}_{wR}) = \left[\frac{1}{100} \sum_{j=1}^{100} \hat{\beta}_{wR}(j) \right] - \beta_*$.

The SSB of RR estimator is the same as $SSB(\hat{\beta}_{wR})$ except the estimator $\hat{\beta}_{wR}$ is replaced by $\hat{\beta}_R$. When the correlation transformation and RR estimation are used, we always take values between 0 and 1 as the possible values of biasing constant k. Hence, I also consider the possible values of k between 0 and 1 in the simulation procedure. All simulated results are displayed in Figure 1 - Figure 4. (see Tsai, 1999)

From Figure 1 and Figure 2, we can see that the values of $MSE(\hat{\beta}_{wR})$ are smaller than $MSE(\hat{\beta}_R)$ for different values of k. Moreover, in Figure 3 and Figure 4, the values of $SSB(\hat{\beta}_{wR})$ are also smaller than $SSB(\hat{\beta}_R)$ for different values of k. That is, the estimator $\hat{\beta}_{wR}$ is more stable and less biased than the RR estimator $\hat{\beta}_R$ when the explanatory variables are highly correlated and data set contains outliers.

3.2 Example

In this section, a real data set is used to demonstrate the application of FWRR estimator. The data set is related to the performance of a computerized system for processing military personnel action forms. The data are listed in Table 1 (see Tsai, 1999). Hill (1977) and Walker and Birch (1988) discussed this data set. In this example, the condition number is 57.14. Walker and Birch (1988) showed that some influential cases exist when the RR estimation is used. When the correlation transformation is used, we can compute the RR estimator $\tilde{\beta}_R$ by using criterion (11). The optimal value of basing constant $k = 0.02$. Then, we can perform the hold-out procedure and optimal fuzzy clustering analysis method to identify outliers. The RR estimators $\tilde{\beta}_R^{(1)}, \tilde{\beta}_R^{(2)}, \cdots, \tilde{\beta}_R^{(20)}$ and the fuzzy weights are given in Table 2 (see Tsai, 1999). The fuzzy weights of observation 1 and 8 are very small. These two observations are identified as outliers. Walker and Birch (1988) also identified them as the most two influential cases. Using the fuzzy weights in Table 2, the FWRR estimator $\tilde{\beta}_{wR}$ can be computed and the value of biasing constant k=0.0006. Moreover, we can get $\hat{\beta}_{Ri} = \frac{s_y}{s_i} \tilde{\beta}_{Ri}$ and

$\hat{\beta}_{wRi} = \frac{s_y}{s_i} \tilde{\beta}_{wRi}$, $i = 1, 2, \cdots, 5$. The estimates of OLS, FWLS, RR and FWRR estimation are displayed in Table 3 (see Tsai, 1999).

4. Conclusion

If the data set contains outliers and explanatory variables are highly correlated, the FWRR estimator is more adequate than RR estimator both in stability and precision. According to the simulation results, the FWRR estimator has smaller MSE and SSB than the RR estimator. Hence, the optimal fuzzy clustering analysis procedure can be used to identify outliers and the FWRR estimation is suggested if outliers appear.

References

[1] Gath, I and Geva, A. (1989a) Unsupervised optimal fuzzy clustering. *IEEE Trans. PAMI* 11, 773-781.

[2] Hoerl, A. E. and Kennard, R. W.(1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, **12**, 55-67.

[3] Hoerl, A. E. and Kennard, R. W.(1976) Ridge regression: Iterative estimation of the biased prarmeter. *Commum, Statist.* A, **5**, 77-88.

[4] Liu, K. (1993) A new class of biased estimate in linear regression, *Commun. Statist.-Theory Meth.*, 22(2), 393-402.

[5] Mayers, R. H. (1986) *Classical and Modern Regression with Applications*, Boston: Duxbury Press.

[6] McDonald, G. C. and Galarneau, D. I. (1975) A Monte Carlo evaluation of some ridge type estimators, *JASA*, 70, 407-416.

[7] Neter, J. Kutner, M. H. and Wasserman, W. (1996) *Applied Linear Statistical Models*, 4/e, IRWIN.

[8] Shia, B-C and Chow, M.-F (1998) Improved ridge regression analysis, *Journal of the Chinese Statistical Association*, Vol. 36, No.3, 259-277.

[9] T. R. Tsai (1999) Fuzzy-weighted Estimation in Ridge Regression Analysis, *Technical Report,* No.1, 1999, Department of Statistics, Tamkang University.

[10] Van Cutsem, B. and Gath, I. (1993) Detection of outliers and robust estimation using fuzzy clustering, *Comput. Statist. Data Anal.*, 15, 47-61.

[11] Walker, E. and Birch, J. B. (1988) Influence measures in ridge regression, *Technometrics*, Vol. 30, No.2, 221-227.

[12] Wu, J-W, Jang, J-B and Tsai, T-R (1996) Fuzzy weighted scaled coefficients in semi-parametric model, *Ann. Inst. Statist. Math.*, Vol. 48, No.1, 97-110.