P- ISSN 1991-8941 E-ISSN 2706-6703 Journal of University of Anbar for Pure Science (JUAPS) Open Access
2012,(6), (1) :67-70

Liu -Type Estimator and Selection of Variables

Mustafa Ismaeel. Niaf.

University of Anbar –College of Education for pure science.

ARTICLE INFO

Received: 14 / 9 /2011
Accepted: 2 / 4 /2012
Available online: 30/10/2012
DOI: 10.37652/juaps.2012.63154

Keywords:
Cp statistic;
generalized Liu-type estimator.

ABSTRACT

In this paper we consider the generalized Liu-type estimator and combine it into subset selection criterion using Cp statistic. Our proposed method can be derived via natural extension of two well-known techniques: one is shrinkage estimators and the other select the best subset.

Introduction

Consider the following multiple linear regression model

$$Y = X\beta + \epsilon, \qquad (1)$$

where Y is an **n** ×1 vector of observations, X is an **n ×p** matrix , β is a **p ×1** vector of unknown parameters, and ϵis an **n ×1** vector of non observable errors which distributed as normal pdf with $E(\epsilon) = 0$ and $Cov(\epsilon) = \sigma^2 I_n$.

The most common method used for estimating the regression coefficients in (1) is the ordinary least squares (OLS) method. The estimator of β by using OLS method is:

$$\hat{\beta} = (XX)^{-1}\acute{X}Y, \qquad (2)$$

Both the OLS estimator and its covariance matrix heavily depend on the characteristics of the X'X matrix. If X'X is ill-conditioned, i.e., the column vectors of X are linearly dependent; the OLS estimators are sensitive to a number of errors. For example, some of the regression coefficients may be statistically insignificant or have the wrong sign, and they may result in wide confidence intervals for individual parameters. With ill-conditioned X'X matrix, it is difficult to make valid statistical inferences about the regression parameters. One of the most popular estimator dealing with multicollinearity is the ordinary ridge regression (ORR) estimator proposed by Hoerl and Kennard [1][2] and defined as:

$$\tilde{\beta}_k = (XX + kI)^{-1}\acute{X}Y$$
$$= [I + k(X'X)^{-1}]^{-1}\hat{\beta}\,, \quad (3)$$

Where k > 0 is a constant. Walker and Page [3] considered a generalization of ridge regression and demonstrated advantages over ridge regression and they determined the ridge constant using a method based on a generalization of the C_p statistic where this provide an automatic variable selection procedure for the canonical variables. Liu [4] introduced a new type of estimators by combining the ORR estimator with any other estimator. The Liu proposed estimator is defined as:

$$\hat{\beta}_{k,d} = (XX + kI)^{-1}(\acute{X}Y - d\hat{\beta}), \quad (4)$$

where $k > 0$, $-\infty < d < \infty$ are the two parameters. He called it as the Liu-type estimator. This estimator is claimed to have advantages over ridge regression estimator. Other solution to this problem is that of subset selection of variables (see Miller [5]).

In this paper, we consider the generalized Liu-type estimator and combine it into the subset selection criterion using C_p statistic. Under some conditions for selection the shrinkage parameter in the generalized Liu-type estimator and the generalized contraction estimator, we don't select the variable by putting some values of this estimator as a zero.

Section 2 reviews the Liu-type estimator while Section 3 introduces the generalized Liu-type estimator. Section 4 gives a generalization of the subset selection criterion C_p. We demonstrate how a generalization of C_p leads to estimate the shrinkage parameter of the

* Corresponding author at: University of Anbar –College of Education for pure science.E-mail address: **mustafaalheety1974@gmail.com**

generalized Liu-type estimator and use this shrinkage estimates for selection of variables.

2 Liu-type estimator

The method of ordinary ridge regression is the most commonly used technique to overcome the multicollinearity. Liu [3] found that when there exists severe collinearity, the shrinkage parameter selected by existing methods for ridge regression may not fully address the ill-conditioning problem. Therefore, he proposed a new estimator, which involves two shrinkage parameters. He claimed that such an estimator will have less mean squares error (MSE) and it can fully address the ill conditioning problem.

Since X'X is symmetric, there exists a $p \times p$ orthogonal matrix P such that $P'X'XP = \Lambda$, Λ is a $p \times p$ diagonal matrix, where the elements of it are the eigenvalues of X'X. So, model (1) can be written in the canonical form as:

$$Y = Z\alpha + \epsilon, \qquad (5)$$

where $Z = XP$ and $\alpha = P'\beta$. The OLS and Liu-type estimators for (5) are respectively:

$$\hat{\alpha} = \Lambda^{-1}Z'Y \, , \qquad (6)$$

and

$$\hat{\alpha}_{k,d} = (\Lambda + kI)^{-1}(Z'Y - d\hat{\alpha}) \quad . \qquad (7)$$

3 The generalized Liu-type estimator

We can generalize the Liu-type estimator for the original model as follows:

$$\hat{\alpha}_{K,D} = (\Lambda + K)^{-1}(Z'Y - D\hat{\alpha}) \quad , \qquad (8)$$

Where $K = diag\{k_1, \dots, k_p\}$ and $= diag\{d_1, \dots, d_p\}$.

So, we can get $\hat{\alpha}_{(K,D)i}$ as:

$$\hat{\alpha}_{(K,D)i} = \frac{(\lambda_i - d_i)\hat{\alpha}_i}{(\lambda_i + k_i)} \qquad (9)$$

Using the generalized Liu-type estimator, it is easy to find the optimal value of k_i.

For that we have to give a theorem.

Theorem 2.1: *For fixed d_i, the mean squares error (MSE) of $\hat{\alpha}_{(K,D)i}$ is minimized with respect to k_i* at

$$k_i = \frac{\sigma^2\lambda_i - d_i(\sigma^2 + \lambda_i\hat{\alpha}_i{}^2)}{\lambda_i\hat{\alpha}_i{}^2}$$

Proof:

The mean squares error of $\hat{\alpha}_{(K,D)i}$ is defined as:

$$MSE(\hat{\alpha}_{(K,D)i}) = \left\|E(\hat{\alpha}_{(K,D)i} - \alpha_i)\right\|^2 + Cov(\hat{\alpha}_{(K,D)i})$$

$$= \frac{(d_i + k_i)^2\alpha_i^2}{(\lambda_i + k_i)^2} + \frac{(d_i - \lambda_i)^2\sigma^2}{\lambda_i(\lambda_i + k_i)^2}$$

to minimize the MSE of $\hat{\alpha}_{(K,D)i}$ with respect to k_i , we take the derivative of the MSE of $\hat{\alpha}_{(K,D)i}$ with respect to k_i and equal it by zero to obtain

$$k_i = \frac{\sigma^2\lambda_i - d_i(\sigma^2 + \lambda_i\hat{\alpha}_i{}^2)}{\lambda_i\hat{\alpha}_i{}^2},$$

The proof is completed.

Since $k_i > 0$, and $-\infty < d_i < \infty$ for all $i = 1, \dots, p$, we suggest to select k_i as follows:

$$k_i = \begin{cases} \dfrac{\sigma^2\lambda_i - d_i(\sigma^2 + \lambda_i\hat{\alpha}_i{}^2)}{\lambda_i\hat{\alpha}_i{}^2} & if\ \sigma^2\lambda_i > \ d_i(\sigma^2 + \lambda_i\hat{\alpha}_i{}^2) \\ 0 & if\ \sigma^2\lambda_i < \ d_i(\sigma^2 + \lambda_i\hat{\alpha}_i{}^2) \end{cases}$$

Now, we can see that there is a good reason for using generalized Liu-type estimator where, it is easier to find the optimal values of k_i, i.e., values for which the mean squares error of Liu-type estimator is a minimum.

Generalization of C_p

The C_p statistic is related to the mean squares error of a fitted model. A common method used for the variable selection is the one based on the C_p statistic, an estimate of the expected value of Γ_p, where

$$\Gamma_p = \frac{1}{\sigma^2}(\hat{\beta}_p - \beta)'X'X(\hat{\beta}_p - \beta), \qquad (10)$$

is the standardized total mean square error (See Montgomery D. C. et al. [6]) and $\hat{\beta}_p$ is the subset least squares estimator. The C_p Statistic is defined as:

$$C_p = \frac{(Y - X\hat{\beta}_p)'(Y - X\hat{\beta}_p)}{\sigma^2} - n + 2p. \qquad (11)$$

Usually, we choose the subset p by minimizing the C_p . For more information on C_p and how it can be used in subset selection of variables, see Mallows [7][8].

Mallows (1973) modified the C_p statistic to a C_k statistic that can be used to determine the shrinkage parameter of ridge regression estimator k by considering

$$C_k = \frac{(Y - X\hat{\beta}_k)'(Y - X\hat{\beta}_k)}{\sigma^2} - n + 2 + 2tr(XL),$$

Where $L = (S + kI)^{-1}X'$ and *tr* stand for the trace. Liu [9] had given some estimates of d by analogy with the estimates of k. Accordingly, the C_p statistic for Liu estimator is:

$$C_d = \frac{(Y - X\hat{\beta}_d)'(Y - X\hat{\beta}_d)}{\sigma^2} - n + 2 + 2tr(XL_d),$$

Where $L_d = (S+I)^{-1}(S+dI)S^{-1}X'$. Here we consider a further generalization of the C_p statistic which can be applied to the generalized Liu-type estimator. We can rewrite (8) as follows:

$$\hat{\alpha}_{K,D} = (\Lambda+K)^{-1}(Z'Y - D\hat{\alpha})$$
$$= (\Lambda+K)^{-1}(Z'Y - D\Lambda^{-1}Z'Y)$$
$$= L_{K,D}Y,$$

Wher $L_{K,D} = (\Lambda+K)^{-1}(\Lambda - D)\Lambda^{-1}Z'$.

Let

$$\Gamma_{L_{K,D}} = \frac{1}{\sigma^2}(\hat{\alpha}_{K,D} - \alpha)'\Lambda(\hat{\alpha}_{K,D} - \alpha), \quad (12)$$

be the scaled sum of squares errors . If we can find a statistic $C_{L_{K,D}}$ which is an estimate of $(\Gamma_{L_{K,D}})$, then we can choose $L_{K,D}$ such that $C_{L_{K,D}}$ is minimized.

Lemma 1: **(See Walker and Page [3])**

$$E(\Gamma_{L_{K,D}}) = W_{L_{K,D}} + \frac{1}{\sigma^2}J_{L_{K,D}},$$

Where $W_{L_{K,D}} = tr(L'_{K,D}\Lambda L_{K,D})$ and $J_{L_{K,D}} = \alpha'(L_{K,D}Z - I_p)'\Lambda(L_{K,D}Z - I_p)\alpha$.

Lemma 2: **(See Walker and Page [3])**

The expectation of the residual sum of squares $RSS_{L_{K,D}}$ is given by

$$E(RSS_{L_{K,D}}) = \sigma^2 W^*_{L_{K,D}} + J_{L_{K,D}},$$

Where $RSS_{L_{K,D}} = (Y - Z\hat{\alpha}_{K,D})'(Y - Z\hat{\alpha}_{K,D}), W^*_{L_{K,D}}$

$$= n - 2tr(ZL_{K,D}) + tr(L'_{K,D}\Lambda L_{K,D}).$$

Now, we can see that

$$C_{L_{K,D}} = 2tr(ZL_{K,D}) - n + \frac{RSS_{L_{K,D}}}{\sigma^2} \quad (13)$$

is an estimate of $E(\Gamma_{L_{K,D}})$.

Lemma 3: **(See Walker and Page [3])**

In terms of the canonical model,

$$C_{L_{K,D}} = 2\sum_{i=1}^{p}\frac{(\lambda_i - d_i)}{(\lambda_i + k_i)} - n + \frac{1}{\sigma^2}\sum_{i=1}^{p}\left(z_i - \sqrt{\lambda_i}\hat{\alpha}_i\right)^2$$

Theorem 1:

1) If $d_i < 0$, then $C_{L_{K,D}}$ is minimized when

$$k_i = \begin{cases} \dfrac{\sigma^2\lambda_i - d_i z_i^2}{z_i^2 - \sigma^2} & if\ z_i^2 > \sigma^2 \\ \infty & if\ z_i^2 \le \sigma^2 \end{cases}$$

2) If $d_i > 0$, then $C_{L_{K,D}}$ is minimized when

$$k_i = \begin{cases} \dfrac{\sigma^2\lambda_i - d_i z_i^2}{z_i^2 - \sigma^2} & if\ \sigma^2\lambda_i > d_i z_i^2\ and\ z_i^2 > \sigma^2\ or \\ & if\ \sigma^2\lambda_i < d_i z_i^2\ and\ z_i^2 < \sigma^2 \\ \infty & if\ \sigma^2\lambda_i \le d_i z_i^2\ or\ z_i^2 \le \sigma^2 \end{cases}$$

Proof:

The $C_{L_{K,D}}$ which is given in Lemma 3 is minimized when

$$G(k_i) = \frac{1}{\sigma^2}\left(z_i - \sqrt{\lambda_i}\hat{\alpha}_{K,D_i}\right)^2 + 2\frac{(\lambda_i - d_i)}{(\lambda_i + k_i)}$$

$$= \frac{z_i^2}{\sigma^2}[1 - 2(\lambda_i + k_i)^{-1}(\lambda_i - d_i) + (\lambda_i + k_i)^{-2}(\lambda_i - d_i)^2] + 2(\lambda_i + k_i)^{-1}$$

is minimized for all i. So,

$$\frac{\partial G(k_i)}{\partial k_i} = \frac{2(\lambda_i - d_i)z_i^2}{\sigma^2(\lambda_i + k_i)^2} - \frac{2(\lambda_i - d_i)^2 z_i^2}{\sigma^2(\lambda_i + k_i)^3} - \frac{2(\lambda_i - d_i)}{(\lambda_i + k_i)^2}.$$

Hence solving $\frac{\partial G(k_i)}{\partial k_i} = 0$ gives $k_i = \frac{\sigma^2\lambda_i - d_i z_i^2}{z_i^2 - \sigma^2}$ which is a stationary point for $G(k_i)$. It is easy to see that $G(k_i)$ is continuous everywhere except at the point $k_i = -\lambda_i$. Since the discontinuity occurs at a negative value of k_i and we require a positive choice of k_i for Liu-type estimator, if $\frac{\sigma^2\lambda_i - d_i z_i^2}{z_i^2 - \sigma^2} > 0$.

That is, if $d_i > 0$ and $z_i^2 > \sigma^2$, then we minimize $G(k_i)$ by choosing $k_i = \frac{\sigma^2\lambda_i - d_i z_i^2}{z_i^2 - \sigma^2}$.

So, we can minimize $C_{L_{K,D}}$ by choosing this k_i which proves 1. Also if $d_i > 0$ and if $if\ \sigma^2\lambda_i > d_i z_i^2\ and\ z_i^2 > \sigma^2\ or\ if\ \sigma^2\lambda_i < d_i z_i^2\ and\ z_i^2 < \sigma^2$, then we minimize $G(k_i)$ by choosing

$$k_i = \frac{\sigma^2\lambda_i - d_i z_i^2}{z_i^2 - \sigma^2}.$$

So, we can minimize $C_{L_{K,D}}$ by choosing this k_i which proves 2, otherwise we will minimize $G(k_i)$ by letting $k_i \to \infty$. The proof is completed.

Now, by using theorem 1 part 1 we can obtain that

$$\hat{\alpha}_{(K,D)_i} = \begin{cases} \dfrac{z_i\sqrt{\lambda_i} - d_i\hat{\alpha}_i}{\lambda_i + k_i} & if\ z_i^2 > \sigma^2 \\ 0 & if\ z_i^2 \le \sigma^2 \end{cases}$$

Also, by using theorem 3 part 2 we can obtain that

$$\hat{\alpha}_{(K,D)_i} = \begin{cases} \dfrac{z_i\sqrt{\lambda_i} - d_i\hat{\alpha}_i}{\lambda_i + k_i} & if\ z_i^2 > \ \sigma^2\ and\ \sigma^2\lambda_i > \ d_i z_i^2\ \ or \\ & if\ \sigma^2\lambda_i < \ d_i z_i^2\ \ and\ \ \ z_i^2 < \sigma^2 \\ 0 & f\ \ \sigma^2\lambda_i \leq \ d_i z_i^2\ \ \ \ or\ \ \ \ z_i^2 \leq \sigma^2 \end{cases}$$

In this paper, we proposed to use the generalized Liu-type estimator for selection of variable at the same time shrinking the remaining estimates. The proposed work is related to the work given by Breiman [10] and Walker and Page [3]. The advantage of this work is that the Liu-type estimator and generalized Liu-type estimator give us more precision over ridge regression estimator and the generalized ridge estimator.

References

[1] Hoerl, A. E. and Kennard, R. W. (1970a). Ridge Regression : Biased estimation for non orthogonal problem. Technometrics, 12, 55-67.

[2] Hoerl, A. E. and Kennard, R. W. (1970b). Ridge Regression : Application for non orthogonal problem. Technometrics, 12, 69-82.

[3] Walker S. G.and Page C. J. (2001). Generalized ridge regression and a generalization of the Cp statistic. Journal of Applied Statistics, 28(7), 911-922.

[4] Liu, K. (2003). Using Liu-type estimator to combat collinearity. Communications in Statistics – Theory and Methods, 32(5), 1009- 1020.

[5] Miller A. J.(2002).Subset selection in regression,2nd edition, Chapman and Hall, USA.

[6] Montgomery, D. C., Peck E. A. and Vining, G. G. (2003). Introduction to linear regression analysis, third edition,John Wiley and Sons .

[7] Mallows, C.L. (1973). Some omment on Cp, Technometrics,15, 661-675.

[8] Mallows, C.L. (1995). More comment on Cp, Technometrics,37, 362-372.

[9] Liu, K. (1993). A new class of biased estimate in linear regression. Communications in Statistics-Theory and Methods, 22, 393-402.

[10] Breiman, L. (1995), Better subset regression using the nonnegative garrote, Technometrics,37, 373-384.

مقدر ليو واختيار المتغيرات

مصطفى اسماعيل نايف

E.mail: mustafaalheety1974@gmail.com

الخلاصة:

في هذا البحث ،سوف نقدم مقدر ليو العام وثم سوف نضمه الى داخل موضوع اختيار المجاميع الجزيئية المختارة من النموذج الخطي المتعدد باستخدام احصائية Cp . الطريقة المقترحة في هذا البحث يمكن ان تشتق باستخدام تقنيتين واسلوبين معرفين في موضوع الانحدار الخطي هما المقدرات المقلصة واختيار افضل مجموعة جزئية من متغيرات النموذج التي من الممكن ان تمثل النموذج.