Fitting Characteristics of the Two-Parameter Estimator in Linear Regression

Xin-feng Chang
Department of Statistics, Jiangsu University, Zhenjiang 212013, China
cxf@ujs.edu.cn

Abstract - Ozkale and Kaciranlar (2007) introduced the two-parameter (TP) estimator for the vector of parameters in a linear regression model. In this paper, several better fitting characteristics of the TP estimator than those of the ordinary least squares and ridge estimators are considered. Furthermore, we propose and investigate the prediction error sum of squares (PRESS) statistics for selecting the parameter d of the TP estimator.

Index Terms - Ridge estimator; Two-parameter estimator; PRESS statistic; Multicollinearity

1. Introduction

Regression analysis is one of the main tools of statistical modelling which is widely used for the prediction of the dependent variable from a set of independent variables. In the presence of multicollinearity, the ordinary least squares estimator is unstable and often gives misleading information, which leads to the development of biased estimators, such as Stein estimator (Stein, 1956) [1], principal components estimator（Massy, 1965）[2], ridge regression estimator (Hoerl and Kennard, 1970) [3], contraction estimator (Mayer and Willke, 1973) [4] and the Liu estimator (Liu, 1993) [5]. Considering the shrinkage selected by existing methods for ridge regression may not fully address the ill conditioning problem when exists severe multicollinearity, Liu (2003) proposed the Liu-Type estimator, which can fully deal with the ill conditioning problem [6]. Ozkale and Kaciranlar (2007) introduced the two-parameter (TP) estimator for the vector of parameters in a linear regression model [7]. Yang and Chang (2010) proposed a new two-parameter estimator which includes the ordinary least squares (OLS) estimator, the ridge regression (RR) estimator and the Liu estimator as special case [8].

The user of biased estimators must choose a biasing parameter so that the improvements in the estimates are being experienced. These improvements often take the form of evidence that the estimates are more stable or that prediction is improved. A usual approach to improve the quality of the prediction performance of the model is to minimize the prediction error sum of squares (PRESS) statistic. Because small value of PRESS statistic represents a model with smaller total mean square error and this implies that the regression model will provide good predictions of new observations. Allen (1971, 1974) [9] [10] considered the PRESS statistic for the OLS estimator. Montgomery and Friedman (1993) obtained the PRESS statistic for the RR estimator and stated that examining the plot the value of PRESS statistic for the RR estimator versus k is helpful [11]. Ozkale and Kaciranlar (2007) investigated the PRESS statistic for selecting d in Liu estimator [12]. Yang and Xu (2008) considered some fitting characteristics for the Liu-Type estimator in linear regression[13].

The main purpose of this paper is to study some other good fitting characteristics of the TP estimator, such as the coefficient of multiple determination and the PRESS statistic. Furthermore, we derived one method to determine the parameter d in TP estimator.

2. Fitting Characteristics of the TP Estimator

Ozkale and Kaciranlar(2007) [7] introduced the TP estimator

$$\hat{\beta}(k,d)=(X'X+kI)^{-1}(X'y+kd\hat{\beta})=\hat{\beta}(k)+C_k^{-1}kd\hat{\beta}$$

where $k>0, 0<d<1$. For this estimator, we can choose a large k because we can adjust another parameter d. So we expect that the TP estimator has advantages over the ridge estimator.

A. R^2 Statistic for TP Estimator

Now we consider the coefficient of multiple determination of the TP estimator. Substituting $\hat{\beta}$ with $\hat{\beta}(k,d)$, we can get

$$\begin{aligned}\hat{R}_{k,d}^2 &= 2\hat{\beta}'(k,d)r-\hat{\beta}'(k,d)C\hat{\beta}(k,d)\\ &=\hat{R}_k^2+2k^2dr'C_k^{-1}C^{-1}C_k^{-1}r-k^2d^2r'C_k^{-1}C^{-1}C_k^{-1}r\end{aligned}$$

To compare the quality of regression, we consider the following difference:

$$\Delta=\hat{R}_{k,d}^2-\hat{R}_k^2=2k^2dr'C_k^{-1}C^{-1}C_k^{-1}r-k^2d^2r'C_k^{-1}C^{-1}C_k^{-1}r$$

Let $k>0$ fixed, we can see that Δ is a quadratic function of d. Noting that

$$\begin{aligned}\frac{\partial\Delta}{\partial d}&=2k^2r'C_k^{-1}C^{-1}C_k^{-1}r-2k^2dr'C_k^{-1}C^{-1}C_k^{-1}r\\ &=2k^2r'C_k^{-1}C^{-1}C_k^{-1}r(1-d)\end{aligned}$$

We can get that $\frac{\partial \Delta}{\partial d} > 0$ for $0 < d < 1$, which means that Δ is a monotone increasing function for $0 < d < 1$. However, $\Delta = 0$ for $d = 0$. Now, we can give the following theorem.

Theorem 2.1. For any fixed parameter $k > 0$, the two-parameter estimator $\hat{\beta}(k,d)$ is superior to the ridge regression estimator $\hat{\beta}(k)$ in the coefficient of multiple determination sense , namely $\hat{R}^2_{k,d} > \hat{R}^2_k$, for $0 < d < 1$.

B. PRESS Statistic for Tp Estimator

Let us now consider PRESS statistic to the TP estimator. The estimator $\hat{\beta}(k,d)$ can also be expressed as $\hat{\beta}(k,d) = (X'X + kI)^{-1}(X'X + kdI)\hat{\beta}$. Let $\hat{\beta}(k,d)_{-i} = (X'_{-i}X_{-i} + kI)^{-1}(X'_{-i}X_{-i} + kdI)\hat{\beta}_{-i}$ be the TP estimator when i th observation is deleted from the sample and $\hat{\beta}_{-i} = (X'_{-i}X_{-i})^{-1}X'_{-i}y_{-i}$ be the OLS estimator when i th observation is deleted from the sample, where y_{-i} and X_{-i} denote the vector y with its i th observation deleted and the matrix X with its i th row deleted. Let x_i be the i th row vector of X and y_i be the i th coordinate observation of y, we can write that $X_{-i}{}'X_{-i} = X'X - x_i x_i{}'$ and $X'_{-i}y_{-i} = X'y - x_i y_i$, $\hat{\beta}(k,d)_{-i}$ equals to

$$\hat{\beta}(k,d)_{-i} = (X'X - x_i x_i' + kI)^{-1}(X'X - x_i y_i + kdI)\hat{\beta}_{-i} \tag{2.1}$$

If we replace $B = X'X$ and then $B = X'X + kI$, $c = x_i$ in $(B - cc')^{-1} = B^{-1} + \frac{B^{-1}cc'B^{-1}}{1 - c'B^{-1}c}$, we get

$$(X'_{-i}X_{-i})^{-1} = (X'X - x_i x_i')^{-1}$$
$$= (X'X)^{-1} + \frac{(X'X)^{-1}x_i x_i'(X'X)^{-1}}{1 - h_{ii}} \tag{2.2}$$

and $(X'_{-i}X_{-i} + kI)^{-1} = (X'X + kI - x_i x_i')^{-1}$

$$= (X'X + kI)^{-1} + \frac{(X'X + kI)^{-1}x_i x_i'(X'X + kI)^{-1}}{1 - h_{k,ii}} \tag{2.3}$$

where $h_{ii} = x_i'(X'X)^{-1}x_i$ and $h_{k,ii} = x_i'(X'X + kI)^{-1}x_i$. Using (2.2) and after some algebraic manipulation, we have

$$\hat{\beta}_{-i} = \hat{\beta} - \frac{\hat{e}_i (X'X)^{-1}x_i}{1 - h_{ii}} \tag{2.4}$$

where $\hat{e}_i = y_i - x_i'\hat{\beta}$.

So we have

$$\hat{\beta}(k,d)_{-i} = \hat{\beta}(k,d) - \frac{(X'X + kI)^{-1}x_i\hat{y}_i}{1 - h_{k,ii}}$$
$$+ \frac{(X'X + kI)^{-1}x_i\hat{y}_{(k,d)i}}{1 - h_{k,ii}}$$
$$+ \frac{(X'X + kI)^{-1}x_i\hat{y}_{(k,d)i}}{1 - h_{k,ii}}$$
$$- \frac{\hat{e}_i}{1 - h_{ii}}(X'X + kI)^{-1}(X'X + kdI)(X'X)^{-1}x_i$$
$$+ \frac{\hat{e}_i h_{ii}}{1 - h_{ii}}\frac{(X'X + kI)^{-1}x_i}{1 - h_{k,ii}} - \frac{\hat{e}_i h_{(k,d)ii}}{1 - h_{ii}}\frac{(X'X + kI)^{-1}x_i}{1 - h_{k,ii}} \tag{2.5}$$

where $\hat{y}_i = x_i'\hat{\beta}$, $\hat{y}_{(k,d)i} = x_i'\hat{\beta}(k,d)$ and $h_{(k,d)ii}$ is the i th diagonal element of the two-parameter matrix $\hat{H}_{(k,d)} = X(X'X + kI)^{-1}(X'X + kdI)(X'X)^{-1}X'$.

Multiplying the above equation by x_i' and taking algebraic manipulation, we have

$$x_i'\hat{\beta}(k,d)_{-i} = \frac{\hat{y}_{(k,d)i}}{1 - h_{k,ii}} - \frac{h_{k,ii}}{1 - h_{k,ii}}\hat{y}_i$$
$$+ \frac{\hat{e}_i}{(1 - h_{ii})(1 - h_{k,ii})}(h_{ii}h_{k,ii} - h_{(k,d)ii})$$

It can be easily checked $\hat{e}_{(k,d)-i} = y_i - x_i'\hat{\beta}(k,d)_{-i}$ equals

$$\hat{e}_{(k,d)-i} = \frac{\hat{e}_{(k,d)i}}{1 - h_{k,ii}} - \frac{\hat{e}_i}{(1 - h_{ii})(1 - h_{k,ii})}(h_{k,ii} - h_{(k,d)ii}) \tag{2.6}$$

where $\hat{e}_{(k,d)i} = y_i - x_i'\hat{\beta}(k,d)$.

Noting that $k(X'X + kI)^{-1}(X'X)^{-1} = (X'X)^{-1} - (X'X + kI)^{-1}$, we have

$$h_{(k,d)ii} = x_i'(X'X + kI)^{-1}(X'X + kdI)(X'X)^{-1}x_i$$
$$= x_i'(X'X + kI)^{-1}(I + kd(X'X)^{-1})x_i$$
$$= (1 - d)h_{k,ii} + dh_{ii} \tag{2.7}$$

We also have

$$\hat{e}_{(k,d)i} - d\hat{e}_i = y_i - x_i'\hat{\beta}(k,d) - d(y_i - x_i'\hat{\beta})$$
$$= (1 - d)y_i + x_i'(d\hat{\beta} - \hat{\beta}(k,d))$$

$$= (1-d)(y_i - \hat{y}(k)) = (1-d)\hat{e}_{ki} \tag{2.8}$$

So we obtain

$$\hat{e}_{(k,d)i} = (1-d)\hat{e}_{ki} + d\hat{e}_i \tag{2.9}$$

Using (3.7) and (3.8), we can write

$$\hat{e}_{(k,d)-i} = \frac{(1-d)\hat{e}_{ki}}{1-h_{k,ii}} + \frac{\hat{e}_i}{(1-h_{ii})} d = (1-d)m + dn \tag{2.10}$$

where $m = \dfrac{\hat{e}_{ki}}{1-h_{k,ii}}$ and $n = \dfrac{\hat{e}_i}{(1-h_{ii})}$.

So we have the PRESS statistic of two-parameter

$$PRESS_{(k,d)} = \sum_{i=1}^{n} (\hat{e}_{(k,d)-i})^2 = \sum_{i=1}^{n} ((1-d)m + dn)^2 \tag{2.11}$$

For k fixed, $PRESS_{(k,d)}$ is a quadratic function of d. The value of d which minimizes $PRESS_{(k,d)}$ can be obtained by differentiating $PRESS_{(k,d)}$ with respect to d

$$\frac{\partial PRESS_{(k,d)}}{\partial d} = 2\sum_{i=1}^{n} d(m-n)^2 - 2\sum_{i=1}^{n} m(m-n) \tag{2.12}$$

Equating to zero we get

$$\hat{d}_{PRESS} = \frac{\sum_{i=1}^{n} m(m-n)}{\sum_{i=1}^{n} (m-n)^2} \tag{2.13}$$

We know that when $k=1$, $\hat{\beta}(k,d)$ leads to the Liu estimator. Therefore, when $k=1$, d in (2.13) reduces to the estimate of d given by Ozkale and Kaciranlar (2007) [12].

Summarizing the above derivations, we can now state the following theorem

Theorem 2.2 The PRESS statistics for the TP estimator is

$$PRESS_{(k,d)} = \sum_{i=1}^{n} (\hat{e}_{(k,d)-i})^2 = \sum_{i=1}^{n} ((1-d)m + dn)^2$$

And it is minimized at

$$\hat{d}_{PRESS} = \frac{\sum_{i=1}^{n} m(m-n)}{\sum_{i=1}^{n} (m-n)^2},$$

where $m = \dfrac{\hat{e}_{ki}}{1-h_{k,ii}}$ and $n = \dfrac{\hat{e}_i}{(1-h_{ii})}$.

3. conclusion

Ozkale and Kaciranlar (2007) introduced the two-parameter (TP) estimator for the vector of parameters in a linear regression model. In this paper, several better fitting characteristics of the TP estimator than those of the ordinary least squares and ridge estimators are considered. Furthermore, we propose and investigate the prediction error sum of squares (PRESS) statistics for selecting the parameter d of the TP estimator.

Acknowledgments

This work was supported by the Application of statistical research base in Jiangsu Province (Grant No.2014JD06) and the Natural Science Foundation of Jiangsu Province of Chain(Grant No. BK20140521).

References

[1] Stein, C. (1956). Inadmissibility of the usual estimator for mean of multivariate normal distribution. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probality, vol 1:197-206.

[2] Massy, W. F. (1965). Principal components regression in exploratory statistical research. JASA 60:234-266.

[3] Hoerl, A. E., Kennard, R. W. (1970). Ridge regression: biased estimation for non-orthogonal problems. Technometrics 12:55-67.

[4] Mayer L. S., Willke T. A. (1973) On biased estimation in linear models. Technometrics 15:497-508.

[5] Liu, K. (1993). A new class of biased estimate in linear regression. Commun Stat Theory Methods 22:393-402.

[6] Liu, K. (2003). Using Liu-Type estimator to combat collinearity. Commun Stat Theory Methods 32:1009-1020.

[7] Ozkale, M.R., Kaciranlar, S. (2007). The restricted and unrestricted two-parameter estimators. Commun Stat Theory Methods 36:2707-2705.

[8] Yang, H., Chang, X. F.. (2010). A new two-parameter estimator in linear regression. Commun Stat Theory Methods 39(6):923-934

[9] Allen, D. M. (1971). Mean square error of prediction as a criterion for selection of variables. Technometrics 13:469-475.

[10] Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. Technometrics 16:125-127.

[11] Montgomery, D. C., Friedman, D. J. (1993). Prediction using regression models with multicollinear predictor variables. IIE Trans. 25:73-85.

[12] Ozkale, M. R., Kaciranlar, S. (2007). A prediction-oriented criterion for choosing the biasing parameter in Liu estimator. Commun Stat Theory Methods 36:1889-1903.

[13] Yang, H., Xu, J. (2008). Tuning parameter selection and various good fitting characteristics for the Liu-Type estimator in linear regression. Commun Stat Theory Methods 37 (20): 3204-3215.