

On the Almost Unbiased Ridge and Liu Estimator in the Logistic Regression Model

Xinfeng CHANG

Department of Statistics, Jiangsu University, Zhenjiang, China

ABSTRACT: This paper is concerned with the parameter estimation in logistic regression model. To overcome the multicollinearity problem, Schaefer et al. (1984) and Urgan and Tez (2008), respectively, proposed the logistic ridge regression estimator and logistic Liu estimator for the logistic regression model. In this article, the almost unbiased Ridge and Liu estimator are proposed by applying the almost unbiased method. Necessary and sufficient conditions for the superiority of the new estimators over the logistic ridge regression estimator and logistic Liu estimator in the mean squared error sense are derived.

KEYWORD: Logistic regression; Maximum likelihood estimator; Ridge regression estimator

1 INTRODUCTION

Consider the following logistic regression model

$$y_i = \pi_i + \varepsilon_i$$

where $\varepsilon_i (i=1, 2, \dots, n)$ are disturbance assumed to be distributed with mean 0 and variance $\omega_i = \pi_i(1-\pi_i)$, π_i is the expectation of y_i when the i th value of the response variable is $Be(\pi_i)$ as

$$\pi_i = \frac{\exp(x_i' \beta)}{1 + \exp(x_i' \beta)}$$

where β is a $(p+1) \times 1$ vector of coefficients, x_i is the i th row of X which is an $n \times (p+1)$ data matrix.

The maximum likelihood estimator (MLE), which is commonly used to estimate the coefficient β , can be obtained as follows:

$$\hat{\beta}_{ML} = (X^T \hat{W} X)^{-1} (X^T \hat{W} z)$$

where $\hat{W} = \text{diag}[\hat{\pi}_i(1-\hat{\pi}_i)]$ and z is a vector

with the i th element equals $\log(\hat{\pi}_i) + \frac{y_i - \hat{\pi}_i}{\hat{\pi}_i(1-\hat{\pi}_i)}$.

The MLE has been treated as the best estimator for a long time. However, the variance of the MLE becomes inflated when the multicollinearity is present. To remedy the effect of multicollinearity, many new biased estimators have been proposed. Schaefer et al. (1984) proposed the logistic ridge regression estimator (RRE) in the Logistic regression model. By following Liu (1993), Urgan and Tez (2008) and Mansson et al. (2011) obtained the

logistic Liu estimator (LE) for the logistic regression model. The properties of the above estimators were studied by Mansson and Shukur (2011), Kibria et al. (2012), Siray et al. (2015) and so on.

However, the alternative estimators proposed to combat multicollinearity are all biased estimators. In the multiple linear regression model, in order to find an estimator with a smaller bias and the some of the worthwhile features of the original estimator, Singh et al. (1986) introduced the almost unbiased ridge estimator (AURE) by using the jack-knife procedure. Akdeniz and Kaciranlar (1995) proposed the almost unbiased generalized Liu estimator for the linear regression model.

Our primary aim in this paper is to introduce the almost unbiased estimators by using the almost unbiased method to the RRE and LE of the logistic regression model, and provide an alternative method to overcome multicollinearity in logistic regression model.

The paper is organized as follows. In Section 2, the new estimator is introduced, and some properties of the new estimator are discussed in Section 3. Finally, some conclusions are given in Section 4.

2 THE PROPOSED ESTIMATORS

The logistic ridge regression estimator (RRE) by Schaefer et al. (1984) is defined as

$$\hat{\beta}_{RR} = (X^T \hat{W} X + kI)^{-1} X^T \hat{W} X \hat{\beta}_{ML}$$

where $k > 0$.

By following Liu (1993), Urgan and Tez (2008) and Mansson et al. (2011) examined the LE for the logistic regression model which is defined as

$$\hat{\beta}_{LE} = (X^T \hat{W}X + I)^{-1} (X^T \hat{W}X + dI) \hat{\beta}_{ML}$$

where $0 < d < 1$.

Definition 2.1. (Kadiyala, 1984) Assume $\hat{\beta}$ is some biased estimator of β and if the bias vector of $\hat{\beta}$ is given by $Bias(\hat{\beta}) = E(\hat{\beta}) - \beta = A\beta$, which means that $E(\hat{\beta}) - A\beta = \beta$, then the estimator $\tilde{\beta} = \hat{\beta} - A\hat{\beta} = (I - A)\hat{\beta}$ is called the almost unbiased estimator based on the biased estimator $\hat{\beta}$.

Now, we are to obtain the almost unbiased Ridge regression and Liu estimator based on the logistic RR by Schaefer et al. (1984) and the LE by Mansson et al. (2011), respectively. We can firstly compute the bias of the RRE and LE as:

$$Bias(\hat{\beta}_{RR}) = -k(X^T \hat{W}X + kI)^{-1} \beta$$

$$Bias(\hat{\beta}_{LE}) = (d - 1)(X^T \hat{W}X + I)^{-1} \beta$$

Therefore, according to Definition 2.1, we may obtain the following almost unbiased Ridge regression estimator (AURE) proposed by Schaefer et al. (1984) and the almost unbiased Liu estimator (AULE) based on the LE by Mansson et al. (2011):

$$\tilde{\beta}_{RR} = [I - k^2(X^T \hat{W}X + kI)^{-2}] \hat{\beta}_{ML} \quad (2.1)$$

$$\tilde{\beta}_{LE} = [I - (1 - d)^2(X^T \hat{W}X + I)^{-2}] \hat{\beta}_{ML} \quad (2.2)$$

3 PROPERTIES OF THE NEW ESTIMATORS

In this section, we will discuss the superiority of our new estimators over some other competing estimators in the literature by the MSE criterion.

The mean squared error (MSE) of an estimator $\hat{\beta}$ is defined as

$$MSE(\hat{\beta}) = tr(Var(\hat{\beta})) + Bias(\hat{\beta})' Bias(\hat{\beta}) \quad (3.1)$$

For the convenience of the following discussions, we denote $\alpha = Q'\beta$, and Q is the orthogonal matrix whose columns constitute the eigenvectors of $X'WX$. Then

$Z'Z = Q'X'WXQ = \Lambda = diag(\lambda_1, \dots, \lambda_p)$, where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$ are the ordered eigenvalues of $X'WX$. α_j is the j th elements of $Q'\beta$, $j = 1, 2, \dots, p$.

It is easy to compute that

$$MSE(\tilde{\beta}_{RR}) = \sum_{j=1}^p \frac{(\lambda_j + 2k)^2 \lambda_j}{(\lambda_j + k)^4} + \sum_{j=1}^p \frac{k^4 \alpha_j^2}{(\lambda_j + k)^4} \quad (3.2)$$

$$MSE(\tilde{\beta}_{LE}) = \sum_{j=1}^p \frac{(\lambda_j + d)^2 (2 + \lambda_j - d)^2}{\lambda_j (\lambda_j + 1)^4} + \sum_{j=1}^p \frac{(1 - d)^4 \alpha_j^2}{(\lambda_j + 1)^4} \quad (3.3)$$

3.1 Comparison of the AURE to the RRE

We can obtain the MSE of $\hat{\beta}_{RR}$

$$MSE(\hat{\beta}_{RR}) = \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + \sum_{j=1}^p \frac{k^2 \alpha_j^2}{(\lambda_j + k)^2} \quad (3.4)$$

Therefore, we have

$$MSE(\tilde{\beta}_{RR}) - MSE(\hat{\beta}_{RR}) = \sum_{j=1}^p \frac{k \lambda_j [-2\alpha_j^2 k^2 + (3 - \lambda_j \alpha_j^2)k + 2\lambda_j]}{(\lambda_j + k)^4}$$

Let $f(k) = -2\alpha_j^2 k^2 + (3 - \lambda_j \alpha_j^2)k + 2\lambda_j$, we know that $MSE(\tilde{\beta}_{RR}) - MSE(\hat{\beta}_{RR}) \leq 0$ if and only if $f(k) \leq 0$. It is easy to see that $-2\alpha_j^2 k^2 + (3 - \lambda_j \alpha_j^2)k + 2\lambda_j = 0$ has one positive

root $k_j = \frac{(3 - \lambda_j \alpha_j^2) + \sqrt{\Delta_1}}{4\alpha_j^2} \geq 0$, where

$$\Delta_1 = (3 - \lambda_j \alpha_j^2)^2 + 16\lambda_j \alpha_j^2 \geq 0.$$

Let $k^* = \max_{1 \leq j \leq p} \{k_j\}$, then for $k \geq k^* > 0$

$$MSE(\tilde{\beta}_{RR}) - MSE(\hat{\beta}_{RR}) \leq 0$$

Theorem 3.1 Let $k \geq k^* > 0$, then the almost unbiased Ridge regression estimator $\tilde{\beta}_{RR}$ is superior to the logistic ridge regression estimator $\hat{\beta}_{RR}$ in the MSE sense, namely $MSE(\tilde{\beta}_{RR}) - MSE(\hat{\beta}_{RR}) \leq 0$, where $k^* = \max_{1 \leq j \leq p} \{k_j\}$.

3.2 Comparison of the AULE to the LE

We can obtain the MSE of $\hat{\beta}_{LE}$

$$MSE(\hat{\beta}_{LE}) = \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)^2} + \sum_{j=1}^p \frac{(1 - d)^2 \alpha_j^2}{(\lambda_j + 1)^2}$$

Therefore, we have

$$MSE(\hat{\beta}_{LE}) - MSE(\tilde{\beta}_{LE}) = \sum_{j=1}^p \frac{\delta_j (1 - \delta_j) \alpha_j^2}{\lambda_j} [\lambda_j \delta_j (1 + \delta_j) - \gamma_j (1 - \delta_j) (2 + \delta_j)]$$

where $\delta_j = \frac{1 - d}{\lambda_j + 1}$, $\gamma_j = \frac{1}{\alpha_j^2}$.

We know that $0 < \delta_j = \frac{1-d}{\lambda_j+1} < 1$, $\lambda_j > 0$, then

$MSE(\beta_{LE}) - MSE(\tilde{\beta}_{LE}) \geq 0$ if and only if

$$\lambda_j \delta_j (1 + \delta_j) - \gamma_j (1 - \delta_j) (2 + \delta_j) \geq 0$$

Or

$$(\lambda_j + \gamma_j)m^2 + (\lambda_j + \gamma_j)(\lambda_j + 1)m - 2\gamma_j(\lambda_j + 1)^2 \geq 0$$

where $m = 1 - d$.

$$(\lambda_j + \gamma_j)m^2 + (\lambda_j + \gamma_j)(\lambda_j + 1)m - 2\gamma_j(\lambda_j + 1)^2 = 0$$

has two roots, respectively

$$m_{1j} = \frac{(\lambda_j + 1)}{2} \left(-1 - \sqrt{\frac{\lambda_j + 9\gamma_j}{\lambda_j + \gamma_j}} \right), m_{2j} = \frac{(\lambda_j + 1)}{2} \left(-1 + \sqrt{\frac{\lambda_j + 9\gamma_j}{\lambda_j + \gamma_j}} \right)$$

where $\Delta_2 = (\lambda_j + 1)^2 (\lambda_j + \gamma_j) (\lambda_j + 9\gamma_j) \geq 0$. It is obvious that $m_{1j} < 0$ and $m_{1j} < m_{2j}$. Since

$0 < m < 1$, we obtain $\alpha_j^2 > \frac{\lambda_j + 2}{2\lambda_j + 3}$, then for

$$0 < \min_{1 \leq j \leq p} \{m_{2j}\} < \max_{1 \leq j \leq p} \{m_{2j}\} < m < 1$$

$$MSE(\beta_{LE}) - MSE(\tilde{\beta}_{LE}) \geq 0.$$

Theorem 3.2 If $\alpha_j^2 > \frac{\lambda_j + 2}{2\lambda_j + 3}$,

$$0 < \min_{1 \leq j \leq p} \{m_{2j}\} < \max_{1 \leq j \leq p} \{m_{2j}\} < m < 1, \quad j = 1, 2, \dots, p,$$

then the almost unbiased Liu estimator $\tilde{\beta}_{LE}$ is superior to the logistic Liu estimator $\hat{\beta}_{LE}$ in the MSE sense, namely $MSE(\tilde{\beta}_{LE}) - MSE(\hat{\beta}_{LE}) \leq 0$.

4 CONCLUSION

In this article, to overcome the multicollinearity problem, we proposed the almost unbiased Ridge and Liu estimator a logistic regression model. Furthermore, the necessary and sufficient conditions for the superiority of the new estimators over the logistic ridge regression estimator and logistic Liu

estimator in the mean squared error sense are derived.

5 ACKNOWLEDGMENTS

This work was supported by the Natural Science Foundation of Jiangsu Province of Chain (Grant No. BK20140521), the Scientific Research Foundation of Jiangsu University (Grant No. 11JDG179) and Application of statistical research base in Jiangsu Province (Grant No.2014JD06).

REFERENCES

- [1] Schaefer, R. L., Roi, L. D., Wolfe, R. A. (1984). A ridge logistic estimator. *Commun. Statist.Theor. Meth.* 13:99–113.
- [2] Liu, K. (1993). A new class of biased estimate in linear regression. *Commun. Statist. Theor. Meth.* 22:393–402.
- [3] Urgan, N. N., Tez, M. (2008). Liu estimator in logistic regression when the data are collinear. In:20th EURO Mini Conference “Continuous Optimization and Knowledge Based Technologies”Lithuania, Selected papers, Vilnius: 323–327.
- [4] Mansson, K., Kibria, B. M. G., Shukur, G. (2011). On Liu estimators for the logistic regression model. *The Royal Institute of Technology, Centre of Excellence for Science and Innovation Studies (CES`IS)*, Sweden, Paper No. 259.
- [5] Kibria, B.M.G., Mansson, K., Shukur, G. (2012). Performance of some logistic ridge regression estimators. *Comput Econ* 40:401–414.
- [6] Siray, G.U., Toker, S., Kaciranlar, S. (2015). On the restricted Liu estimator in the logistic regression model. *Commun. Statist. Theor. Meth.* 44:217–232.
- [7] Mansson, K., Shukur, G. (2011). On ridge parameters in logistic regression. *Commun. Statist. Theor. Meth.* 40:3366–3381.
- [8] Mansson, K. (2012). On Ridge Estimators for the Negative Binomial Regression Model. *Economic Modelling.* 29:178–184.
- [9] Singh, B., Chaubey, Y.P., Dwivedi, T.D. (1986). An almost unbiased ridge estimator. *Sankhya.* 48: 342-346.
- [10] Akdeniz, F., Kaciranlar, S. (1995). On the almost unbiased generalized Liu estimator and unbiased estimation of the bias and MSE. *Commun. Statist. Theor. Meth.* 24:1789-1797.
- [11] Kadiyala, K. (1984). A class of almost unbiased and efficient estimators of regression coefficients. *Economic Letters.* 16:293-296.