International Journal of Advanced Science and Research
ISSN: 2455-4227
Impact Factor: RJIF 5.12
www.allsciencejournal.com
Volume 2; Issue 6; November 2017; Page No. 60-63

Logistic shrinkage partial ridge regression for the liu type estimators

H Bagyalakshmi, MR Srinivasan
Department of Statistics, University of Madras, Chennai, Tamil Nadu, India

Abstract
In this paper an approach of finding the partial ridge regression is applied to the logistic ridge regression for the liu type estimators and a study has been analyzed. The new method called Logistic Shrinkage Partial Ridge Regression for the liu type estimators of estimating coefficients and mean squared error in logistic ridge regression. The method is evaluated through a Simulation study. Different bias values of *k* are considered for the analysis of the method together with the maximum likelihood approach. Simulation study results shows that the proposed methods generally outperforms maximum likelihood, Logistic ridge regression for the liu type estimators, Generalized Logistic Ridge Regression for the liu type estimators and Directed Logistic Ridge Regression for the liu type estimator methods.

Keywords: Linear regression, multicollinearity, ridge estimator, liu type logistic estimators, logit, maximum likelihood estimator, mean squared error, simulation study

1. Introduction

Frisch (1934), who introduced the concept of Multicollinearity to a situation where the linear dependency have high correlation among the variables. Hoerl and Kennard (1970 a, b) [1, 2] first introduced the concept of ridge regression to overcome the problem of Multicollinearity. They proved that there is a non zero value of such ridge parameter for which the mean squared error (MSE) for the slope parameter using the ridge regression is smaller than the variance of the Least Square Estimator (LSE) of the respective parameter. Other authors have worked with this area and developed and proposed different estimates for the ridge regression parameter. Various ridge estimators are suggested by many researchers in ridge regression viz., McDonald and Galarneau (1975), Lawless and Wang (1976), Saleh and Kibria (1993), Haq and Kibria (1996), Kibria (2003, 2004), Kibria and Saleh (2004), Khalaf and Shukur (2005), Alkhamisi *et al.* (2006). These methods have been applied on linear regression models for which the dependent variable is quantitative with the objective to estimate its expected value given the values of the explanatory variable. The disadvantage of the ridge estimator is that the estimated parameter is nonlinear functions of the ridge parameter and that the small *k* chosen in the process may not be large enough to overcome multicollinearity (Asar and Genc, 2016) [18].

In ridge regression, the condition numbers of the matrix $X'X+kI$ becomes a decreasing function of k (Liu, 2003).Thus large values of k has to be used to get small condition number. However a small value of k is applied in $X'X+kI$, which is still ill conditioned. Liu estimator having the advantage of being a linear function of the shrinkage parameter was proposed by Liu (2003). Liu estimator has the advantage of ridge and stein estimator, Stein (1956). Liu type estimator is a two parameter estimator which shows superiority of other estimators in ridge regression proposed by Liu (2003).

Logistic regression is one of the extensively used method exists in generalized linear models. Schaefer *et al.* (1984) proposed logit ridge regression estimator and some methods of estimating *k* were suggested and evaluated by means of Monte Carlo simulations. It is shown that logistic ridge regression outperforms ML when the explanatory variables are collinear. Liu and Liu-type logistic estimators were proposed by Kaciranlar (2003) Mansson and Shukur (2011), Mansson *et al.*, (2012), Kibria *et al.*, (2012),Inan and Endogan (2013). Logistic ridge estimators were proposed in Kibria *et al.* (2012). Other authors proposed different methods for the estimation of the shrinkage parameter *d* used in Liu and Liu-type estimators.

Huang (2012) suggested new methods to estimate the shrinkage parameter *d* used in Liu-type estimators in order to combat multicollinearity in logistic regression model.Asar and Genc, 2016) [18] used the same estimator with optimal shrinkage parameter and proposed some shrinkage parameters *d* used in Liu-type logistic estimators.

When the explanatory variables are correlated Liu-type logistic estimators with the new shrinkage estimators are assumed to perform better than the ML estimators. In the present article a new method called Logistic Shrinkage Partial Ridge Regression for the Liu type estimators (LSPRR-L) is introduced to estimate Liu type logistic estimators.

Hoerl and Kennard (1970) [1] proposed an extension of the ridge regression procedure that allows separate biasing parameters k for each regressor known as Generalized Ridge Regression (GRR). The discussion on generalized approach of estimation is simplified by a suitable orthogonal transformation of regressors. Shrinking the elements of the parameter vector is restricted to only those coefficients corresponding to small eigenvalues and the shrinkage is carried out using generalized ridge. Guilkey and Murphy (1975) suggested a method called Directed Ridge Regression

(DRR) that Xi be defined as small if corresponding $\lambda_i < 10^{-c}\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $X'X$ and c is some arbitrary constant. The process is still computationally intensive as in GRR but to lesser extent as still a partial set of vectors $(k_1, k_2, .., k_p)$ need to be obtained. A method called Partial Ridge Regression (PRR) (Chandrasekhar *et al.*, (2016) which involves selectively adjusting the ridge constants associated with highly collinear variables to control instability in the variances of coefficient estimates is discussed.

The present paper details out the methodology of logistic ridge regression in section 2.Other methods such as Partial logistic ridge regression, Generalized logistic ridge regression, Directed logistic ridge regression are also discussed in Section 2. A detailed description of Monte Carlo simulation is given in section 3. The simulation results are discussed in section 4. Finally the conclusion is given in section 5and results in Appendix.

2. Model and Methodology

Logit regression is a widely used statistical method when the ith value of the dependent variable y of the regression model is $Be(\pi_i)$ with the following parameter value:

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

where xi is the ith row of X which is an $n \times (p+1)$ data matrix with p explanatory variables and β is a $(p+1)\times 1$ vector of coefficients. The most common method of estimating β is to apply the ML method .The log likelihood should be maximized is given by,

$$l = \sum_{i=1}^{N} y_i \log(\pi_i) + \sum_{i=1}^{N} (1-y_i)\log(1-\pi_i) \tag{2}$$

Solving the above equation by taking the first derivative and equals the expression to zero,the ML estimates are obtained.

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{N} (y_i - \pi_i)x_i = 0 \tag{3}$$

The following iterative weighted least square (IWLS) algorithm

$$\beta_{\text{ML}} = (X'\hat{W}X)^{-1} X'\hat{W}\hat{z} \tag{4}$$

where, $\hat{W} = \hat{\pi}_i(1-\hat{\pi}_i)$ and $\hat{z}$ is a vector where the ith element equals to

$$z_i = \log(\hat{\pi}_i) + \frac{(y_i - \hat{\pi}_i)}{\hat{\pi}_i(1-\hat{\pi}_i)} \tag{5}$$

Since (3) is nonlinear in β .

The MSE of the ML estimator equal to

$$E(L^2_{\text{ML}}) = E(\beta_{\text{ML}} - \beta)' E(\beta_{\text{ML}} - \beta)$$

$$= tr(X'\hat{W}X)^{-1}$$

$$= \sum_{j=1}^{J} \frac{1}{\lambda_j} \tag{6}$$

$$\text{MMSE}(\hat{\beta}\ \text{MLE}) = E(\hat{\beta}_{\text{MLE}} - \beta)' E(\beta_{\text{MLE}} - \beta)$$

$$= (X'\hat{W}X)^{-1} \tag{7}$$

Where λ_j is the j th eigenvalue of the $X'\hat{W}X$ matrix. When the independent variables are highly correlated the MSE of the ML estimate becomes inflated. Huang (2012) proposed the Liu type logistic estimator, is given below

$$\beta_{k,d} = (X'\hat{W}X + kI)^{-1}(X'\hat{W}X + kdI)\hat{\beta}_{MLE} \tag{8}$$

Where $k > 0$ and $0 < d < 1$.When $k = 1$, $\beta_{k,d} = \beta_d$ which is the Liu estimator in logistic regression defined by Mansson *et al.*(2012) being a generalization of the Liu estimator such that $\beta_d = (X'\hat{W}X + I)^{-1}(X'\hat{W}X + dI)\hat{\beta}_{MLE}$ where $0 < d < 1$.When $d = 1$ in β_d, then it becomes $\beta_d = \hat{\beta}_{MLE}$.The MSE and MMSE of the estimator $\beta_{k,d}$ are given as,

$$MSE(\hat{\beta}_{k,d}) = \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k_j)^2} + \sum_{j=1}^{p} \frac{k_j^2 \alpha_j^2}{(\lambda_j + k_j)^2} \tag{9}$$

$$MMSE(\hat{\beta}_{k,d}) = E(\beta_{k,d} - \beta)' E(\beta_{k,d} - \beta) \tag{10}$$

Where α_j^2 is defined as the jth element of $\gamma'\beta$ and γ is the eigen vector of $(X'\hat{W}X)$ such that $(X'\hat{W}X) = (\gamma'\Lambda\gamma)$. Where Λ equals $diag(\lambda_1, \lambda_2, \lambda_3.......\lambda_{p+1})$, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq .... \geq \lambda_{p+1}$.

Logistic Shrinkage Partial Ridge Regression for the Liu type estimators Method

Chandrasekar et.al.,(2016) proposed a PRR approach, which involves selectively adjusting the ridge constants associated with highly collinear variables to control instability in the variances of coefficient estimates. Applying the Singular value decomposition to $X'WX$, we get,

$$X'WX = U\Lambda U'$$

The eigenvalues of $U\Lambda U'$ are used to decide cutoff based criteria. A cutoff criteria method has been adapted to first identify q (q<p) collinear regressors and biasing constant k is added only to the q regressors by selectively choosing those regressor variables exhibiting high degree of collinearity.

The $X'WX$ matrix of order $p \times p$ may be decomposed as $U\Lambda U'$ where Λ is a $p \times p$ diagonal matrix whose main diagonal elements are the eigen values λ_j (j = 1,2,..p) of $X'WX$. If the eigen value λ_j is close to zero it indicates near linear dependence. The diagonal elements of Λ are the eigen values of $X'WX$ which are arranged in descending order in value from $\lambda_1 to \lambda p$ with $\lambda_1 \geq \lambda_2 \geq ..... \geq \lambda_p$.Highly collinear variables are associated with low eigen values. The ratio of jth eigen value to the largest eigen value say $\delta = \lambda_j / \lambda_1$ is a useful measure of degree of near linear dependence. There is a general consensus that a value of the order of 0.001 (one thousandth) may be considered as indicating high collinearity and used in deciding the number of linear dependence in $X'X$.Thus applying the above

procedure to shrinkage logistic ridge regression and a logistic shrinkage partial ridge estimator is obtained. The new method is known as Logistic Shrinkage Partial Ridge Regression of Liu type estimator (LSPRR_L) is given by,

$$\beta_{LSPRR\_L} = (U\Lambda U' + kI)^{-1}(U\Lambda U' + kdI)\beta_{ML} \qquad (11)$$

Where the constant *k* is added only to the q regressors by selectively choosing those regressor variables exhibiting high degree of collinearity

The MSE of the LSPRR_L estimator equals

$$MSE_{LSPRR\_L} = E(\beta_{LSPRR\_L} - \beta)' E(\beta_{LSPRR\_L} - \beta) \qquad (12)$$

The bias values of *k* has been considered in the range between 0.001 to 0.512 (Montgomery, 2012). In order to attain stability in MSE, the k values are further increased to 0.55 to 11.55. The study has also shown the regression coefficients obtained stability for a partial set of regressors and attains minimum MSE compared to ML and other methods.

Hoerl and Kennard (1970) [1] suggested to replace σ^2 and α_i, by their corresponding unbiased estimators. That is,

$$\hat{K}_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \qquad (13)$$

Where, $\hat{\sigma}^2 = \frac{\sum e_i^2}{(n-p)}$ is the residual mean square estimate, which is an unbiased estimator of σ^2 and α_i is the ith element of $\hat{\alpha}$, which is an unbiased estimator of α.The method of adopting the above biasing values of K where $K = K_1, K_2,.., K_p$ are based on generalized ridge regression, which is applied in logistic ridge regression for the liu type estimators and is called Generalized Logistic Ridge Regression for the liu type estimators(GLRR_L). The method GLRR_L is a computationally intensive approach of dealing with multicollinearity. Guilkey and Murphy (1975) suggested a selective approach of biasing the parameters may be considered, than biasing all the parameters $(K_1, K_2,.., K_p)$ which is named as directed ridge regression method. The directed ridge regression method of biasing the ridge parameter is adopted in logistic ridge regression and is called as Directed Logistic Ridge Regression off or the liu type estimators (DLRR_L).

However, the efficiency of LSPRR_L approach needs to be evaluated in comparison with ML, GLRR_L _L and DLRR_L methods discussed earlier. In the next section, to know which method show better performance, all the logistic ridge regression methods discussed above are evaluated under simulation study.

3. The Monte Carlo Simulation

The aim of this study is to evaluate the performance of different logistic shrinkage ridge regression methods and find a good one. A simulation study has been conducted using R programming language. Because degree of collinearity among explanatory variables (X) is of central importance, Kibria (2011) is followed for generating X variables using the following equation

$$x_{ij} = (1-\rho^2)^{(1/2)} z_{ij} + \rho z_{ip} \quad i = 1,2,....n, \; j = 1,2,....p \qquad (14)$$

Where z_{ij} are independent standard normal pseudo random numbers and ρ represents correlation between any two X variables. The n observations of the dependent variable are obtained from Be (π_i) distribution where

$$\pi_i = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} \qquad (15)$$

The parameter values of β are chosen so that $\beta'\beta = 1$. Newhouse and Oman (1971) stated that if the mean squared error is a function of $\beta, \sigma^2 and$ k, if the explanatory variables are fixed, then the MSE is minimized when β is the normalized eigen vectors corresponding to the largest eigen value of $X'\hat{W}X$. The sample size n are varied from 100,150 and 200 with the explanatory variables p is considered as 4 and 8. In this study, four values of ρ are investigated which are given by 0.75, 0.85, 0.95 and 0.99.Under these criterions for different combinations of n and p the procedure of LSPRR_L is followed to find which method is optimal and stable with respect to biased values of k.

4. Results and discussion

The results of the simulation have been provided in this section. The simulated MSE of the different methods of logistic regression are provided in the Appendix (Tables 1 and 8). The degree of correlation ρ, the number of observations *n* and the number of explanatory variables *p* are the factors that influence the Mean square error. For all of the cases and for all of the methods MSE have the better performance than ML estimator. For all the methods, when the data size increases, the estimated MSE decreases. It is observed from the tables, that when the degree of correlation increases, the values of MSE are enlarged. The increase is larger for GLRR_L and DLRR_L than LSPRR_L. The MSE attains stability for LSPRR_L when the biased values are increased. The values are generally less than ML, LRR_L, GLRR_L and DLRR_L. It has been observed from the tables when ρ increases the mean square values are increasing with respect to the sample size values. However for LSPRR_L method, the increase in the MSE are smaller than the other methods. It is observed that, in general MSE of LSPRR_L is less than other methods and outperforms for both the explanatory variables 4 and 8. It has been noted that the MSE values of LRR_L is considerably decreasing for an increasing set of the biased values. However, using the method LSPRR_L, the obtained MSE values are decreasing moderately for the given biased values and attains stability for a particular value of *k.* After obtaining the stability, the MSE is increasing even for a higher biased value of *k*. Hence the limit of the biased value can be fixed for the sample size, correlation and explanatory variables. Where as in GLRR_L and DLRR_L the MSE values are constant for the respective correlation, sample size and explanatory variables. Hence, results obtained by LSPRR_L method has a potential improvement over the other methods viz., LRR_L, GLRR_L and DLRR_L. The stability of the MSE for the *k* values in LSPRR_L is consistent irrespective of the sample size.

5. Summary and Conclusion

In this study, a new application of logistic shrinkage partial ridge regression for the Liu type estimators method have been analyzed to combat the problem of multicollinearity. A Monte Carlo simulation study has been used by generating random numbers for the explanatory variables and the dependent variables. Different degrees of correlation, sample size and the number of explanatory variables are considered. The MSE is used as a performance criterion. The new suggested method is compared with ML, GLRR_L and DLRR_L. The results generally show that LSPRR_L results are better than the other methods. Using a single *k* gives efficient results in terms of stability and the MSE remains considerably less compared to other methods. The study based on the results obtained by simulation, has shown that LSPRR_L is the best method because it overcomes the shortcomings of other logistic shrinkage ridge regression for Liu type estimator methods.

6. References

1. Arthur, Hoerl, Robert W. Kennard. Ridge Regression: Biased estimation for non orthogonal problems, Technometrics. 1970a; 12:55-67.
2. Arthur. Hoerl and Robert W.kennard. Ridge Regression: Applications to Non orthogonal problems, Technometrics. 1970b; 12:69-82.
3. Chandrasekhar CK, Bagyalakshmi H, Srinivasan MR, Gallo M. Partial Ridge Regression under Multicollinearity, Journal of Applied Statistics, 2016, 1-12.
4. Frisch R. Statistical confluence analysis by means of complete regression systems. Publication 5. Oslo: University Institute of Economics.
5. Golam Kibria BM. Performance of some new ridge regression, Communication in Statistics–Simulation and Computation. 2003; 32(2):419-435
6. Golam Kibria BM, Mansson K, Shukur G. Performance of some logistic ridge regression estimators. Computational Economics. 2011; 40(4):401-414.
7. Huang J. A simulation research on a biased estimator in logistic regression model ComputationalIntelligence and Intelligent Systems: Springer, 2012, 389-395.
8. Inan D, Erdogan BE. Liu-type logistic estimator. Communications in Statistics-Simulation and Computation. 2013; 42(7):1578-1586.
9. Kaciranlar S. Liu estimator in the general linear regression model, Journal of Applied Statistical Sciences, 2003, 12(3).
10. Kibria BMG. Performance of some new ridge regression estimators. Communications in Statistics: Theory and Methods. 2003; 32:419-435.
11. Kibria BMG. Performance of some logistic ridge regression estimators, Computational Economics. 2012, 40:12.
12. Liu K. A new class of blased estimate in linear regression. Communications in Statistics-Theory and Methods. 1993; 22(2):393-402.
13. Liu K. Using Liu-type estimator to combat collinearity. Communications inStatistics-Theory and Methods. 2003; 32(5):1009-1020.
14. .Månsson K, Shukur G. On ridge parameters in logistic regression. To appear in Communications in Statistics: Theory and Methods, 2011.
15. Mansson K, Kibria B, Shukur G. On Liu estimators for the logit regression model. Economic Modelling. 2012; 29(4):1483-1488.
16. Newhouse JP, Oman SD. An evaluation of ridge estimators. Rand Corporation. 1971; 716-PR, 1-28. 17.
17. Schaeffer RL, Roi LD, Wolfe RA. A ridge logistic estimator. Communications in Statistics: Theory and Methods. 1984; 13:99-113.
18. Yasin Asar, Asir Genc. New Shrinkage Parameters for the Liu-Type Logistic Estimators, Communication in Statistics- Simulation and Computation. 2016; 45:1094-1103.