Pakistan Journal of Statistics and Operation Research

Jackknifing Ridge Estimator for Logistic Regression Model

Nawal Mahmood Hammood[1] , Zakariya Yahya Algamal[2*]

* Corresponding Author

1. Department of management Information systems, College of Administration and Economics, University of Mosul, Mosul, Iraq, nawal_almamary@uomosul.edu.iq.
2. Department of Statistics and Informatics, University of Mosul, Mosul, Iraq, zakariya.algamal@uomosul.edu.iq

Abstract

It has been repeatedly shown that the ridge regression model is a desirable shrinking technique to lessen the consequences of multicollinearity. When the response variable involves binary data, the logistic regression model is a well-known model in use. However, it is well known that multicollinearity has a detrimental impact on the variance of the logistic regression coefficients' maximum likelihood estimate. Numerous scholars have suggested a logistic ridge estimator as a solution to this issue. The jackknifing logistic ridge estimator (JLRE) is suggested and derived in this study. The goal of the JLRE is to obtain a diagonal matrix with low diagonal element values, which will reduce the shrinkage parameter and enable a better, less biassed estimator to be produced. According on the results of our Monte Carlo simulation, the JLRE estimator can significantly outperform other available estimators. Additionally, the JLRE estimator surpasses the logistic ridge estimator and the maximum likelihood estimator in terms of predictive performance, according to the practical application findings.

Key Words: Multicollinearity; ridge estimator; logistic regression model; Jackknife estimator; Monte Carlo simulation.

1. Introduction

Binary classification using a logistic regression model has often been adopted in several real data applications, “such as cancer classification. Various studies have attempted to apply the logistic regression model as a base to build a classification model.

In the presence of multicollinearity, when estimating the regression coefficients for logistic regression model using the maximum likelihood (ML) method, the estimated coefficients are usually become unstable with a high variance, and therefore low statistical significance (Kibria, Månsson, & Shukur, 2015). Numerous remedial methods have been proposed to overcome the problem of multicollinearity. The ridge regression method (Hoerl & Kennard, 1970) has been consistently demonstrated to be an attractive and alternative to the ML estimation method.

Ridge regression is a shrinkage method that shrinks all regression coefficients toward zero to reduce the large variance (Asar & Genç, 2015). This done by adding a positive amount to the diagonal of $\boldsymbol{X}^T\boldsymbol{X}$. As a result, it is guaranteed that for some range of values of $k$the ridge estimator has a smaller MSE than the ML, but for sufficiently large $k$the MSE will typically be larger than that of the ML.

In linear regression, the ridge estimator is defined as

$$\hat{\boldsymbol{\beta}}_{Ridge} = (\boldsymbol{X}^T\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}, \qquad (1)$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of observations of the response variable, $\boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_p)$ is an $n \times p$ known design matrix of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is a $p \times 1$ vector of unknown regression coefficients, $\boldsymbol{I}$ is the identity matrix with dimension $p \times p$, and $k \geq 0$ represents the ridge parameter (shrinkage parameter). The ridge parameter, k, controls the shrinkage of $\boldsymbol{\beta}$ toward zero. The OLS estimator can be considered as a special estimator from Eq. (1) with $k = 0$. For larger value of k, the $\hat{\boldsymbol{\beta}}_{Ridge}$ estimator yields greater shrinkage approaching zero (Algamal & Lee, 2015; Hoerl & Kennard, 1970).

2. Logistic ridge model

Logistic regression is a statistical method to model a binary classification problem. The regression function has a nonlinear relation with the linear combination of the explanatory variables. In classification, the response variable of the logistic regression has two values either 1 for the positive class or 0 for the normal class. Assume that we have n observations and p explanatory variables. Let $y_i \in \{0,1\}$ be the response variable value for observation i, $i = 1,2,\ldots,n$ and $\boldsymbol{x}_i = (x_{i1}, x_{i2},\ldots, x_{in})^T$ be the i^{th} explanatory variable vector of the design matrix $\boldsymbol{X}$. Then, the response variable is related to explanatory variables by

$$\pi_i = p(y_i = 1|\boldsymbol{x}_i) = \frac{exp(x_i^T\beta)}{1+exp(x_i^T\beta)},\ i = 1,2,\ldots,n \qquad (1)$$

where $\beta = (\beta_0, \beta_1,\ldots,\beta_p)^T$ is a $(p+1)\times 1$ vector of unknown explanatory variables coefficients. The log-likelihood function of the logit transformation of Eq. (1) is defined as

$$\ell(\beta) = \sum_{i=1}^{n}\{y_i\, log(\pi_i) + (1-y_i)\, log(1-\pi_i)\}. \qquad (2)$$

The ML estimator is then obtained by computing the first derivative of the Eq. (3) and setting it equal to zero, as

$$\frac{\partial\ell(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} = \sum_{i=1}^{n}[y_i - \pi_i]\,\boldsymbol{x}_i = 0. \qquad (3)$$

Because Eq. (4) is nonlinear in $\boldsymbol{\beta}$, the iteratively weighted least squares (IWLS) algorithm can be used to obtain the ML estimators of the logistic regression parameters (LR) as

$$\widehat{\boldsymbol{\beta}}_{LR} = (\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{W}}\widehat{\boldsymbol{v}}, (4)$$

where $\widehat{\boldsymbol{W}} = \text{diag}(\hat{\pi}_i(1-\hat{\pi}_i))$ and $\widehat{\boldsymbol{v}}$ is a vector where ith element equals to $\hat{v}_i = ln(\hat{\pi}_i) + ((y_i - \hat{\pi}_i)/\hat{\pi}_i(1-\hat{\pi}_i))$. The ML estimator is asymptotically normally distributed with a covariance matrix that corresponds to the inverse of the Hessian matrix

$$cov(\widehat{\boldsymbol{\beta}}_{LR}) = \left[-E\left(\frac{\partial^2\ell(\boldsymbol{\beta})}{\partial\beta_i\,\partial\beta_k}\right)\right]^{-1} = (\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}. \qquad (5)$$

The mean squared error (MSE) of Eq. (5) can be obtained as

$$\begin{aligned}\text{MSE}(\widehat{\boldsymbol{\beta}}_{LR}) &= E(\widehat{\boldsymbol{\beta}}_{LR} - \boldsymbol{\beta})^T(\widehat{\boldsymbol{\beta}}_{LR} - \boldsymbol{\beta}) \\ &= tr[(\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X})^{-1}] \\ &= \sum_{j=1}^{p}\frac{1}{\lambda_j}, \qquad (6)\end{aligned}$$

where λ_j is the eigenvalue of the $\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X}$ matrix.

In the presence of multicollinearity, the matrix $\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X}$ becomes ill-conditioned leading to high variance and instability of the ML estimator of the logistic regression parameters. As a remedy, logistic ridge estimator (LRE) (Le Cessie & Van Houwelingen, 1992; Lee & Silvapulle, 1988; Schaefer, Roi, & Wolfe, 1984) as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{LRE} &= (\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{LR} \\ &= (\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X} + k\boldsymbol{I})^{-1}\boldsymbol{X}^T\widehat{\boldsymbol{W}}\widehat{\boldsymbol{v}}, \quad (7)\end{aligned}$$

where $k \geq 0$. The ML estimator can be considered as a special estimator from Eq. (8) with $k = 0$. Regardless of $k$value, the MSE of the $\widehat{\boldsymbol{\beta}}_{LRE}$is smaller than that of $\widehat{\boldsymbol{\beta}}_{LR}$because the MSE of $\widehat{\boldsymbol{\beta}}_{LRE}$is equal to (Kibria et al., 2015; Rashad & Algamal, 2019)

$$\text{MSE}(\widehat{\boldsymbol{\beta}}_{LRE}) = \sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j+k)^2} + k^2\sum_{j=1}^{p}\frac{\alpha_j}{(\lambda_j+k)^2}, \qquad (8)$$

where α_j is defined as the jth element of $\gamma\widehat{\boldsymbol{\beta}}_{LR}$and γ is the eigenvector of the $\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X}$ matrix. Comparing with the MSE of Eq. (7), $\text{MSE}(\widehat{\boldsymbol{\beta}}_{LRE})$ is always small for $k > 0$.

3. The proposed estimator

In this section, the new estimator is introduced and derived. Let $\boldsymbol{D} = (d_1, d_2,\ldots, d_p)$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2,\ldots,\lambda_p)$, respectively, be the matrices of eigenvectors and eigenvalues of the $\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X}$ matrix, such that $\boldsymbol{M}^T\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X}\boldsymbol{D} = \boldsymbol{S}^T\widehat{\boldsymbol{W}}\boldsymbol{S} = \boldsymbol{\Lambda}$, where $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{D}$. Consequently, the logistic regression estimator of Eq. (5), $\widehat{\boldsymbol{\beta}}_{LR}$, can be written as

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}_{LR} &= \boldsymbol{\Lambda}^{-1}\boldsymbol{S}^T\widehat{\boldsymbol{W}}\widehat{\boldsymbol{v}} \\ \widehat{\boldsymbol{\beta}}_{LR} &= \boldsymbol{D}\widehat{\boldsymbol{\gamma}}_{LR}. \qquad (9)\end{aligned}$$

Accordingly, the logistic ridge estimator, $\widehat{\boldsymbol{\beta}}_{LRE}$, is rewritten as

$$\begin{aligned}\widehat{\boldsymbol{\gamma}}_{LRE} &= (\boldsymbol{\Lambda} + \boldsymbol{K})^{-1}\boldsymbol{S}^T\widehat{\boldsymbol{W}}\boldsymbol{v} \\ &= (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{B}^{-1})\widehat{\boldsymbol{\gamma}}_{LR}, \qquad (10)\end{aligned}$$

where $\boldsymbol{B}=\boldsymbol{\Lambda}+\boldsymbol{K}$ and $\boldsymbol{K}=\text{diag}(k_1,k_2,\dots,k_p)$; $k_i\geq 0,\ i=1,2,\dots,p$. Equation (11) represents the generalized ridge logistic regression estimator (Batah, Ramanathan, & Gore, 2008; Khurana, Chaubey, & Chandra, 2014; Özkale, 2008).

In generalized ridge estimator, the Jackknifing approach was used (Khurana et al., 2014; Nyquist, 1988; Singh, Chaubey, & Dwivedi, 1986). Batah et al. (2008) proposed a modified Jackknifed ridge regression estimator in linear regression model. Related to Poisson regression model, Türkan and Özel (2015) proposed a modified Jackknifed Poisson ridge estimator depending on the study of Singh et al. (1986).

In this paper, the new estimator (JLRE) is derived by following the study of Batah et al. (2008). Let the Jackknife estimator (JE), in logistic regression, is defined as

$$\widehat{\boldsymbol{\gamma}}_{JE}=(\boldsymbol{I}-\boldsymbol{K}^2\boldsymbol{B}^{-2})\widehat{\boldsymbol{\gamma}}_{LR}, \quad (11)$$

and the modified Jackknife estimator (MJE) of Batah et al. (2008), in logistic regression model, is defined as

$$\widehat{\boldsymbol{\gamma}}_{MJE}=(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})(\boldsymbol{I}-\boldsymbol{K}^2\boldsymbol{B}^{-2})\widehat{\boldsymbol{\gamma}}_{LR}. \qquad (12)$$

Consequently, our new estimator is an improvement of Eq. (13) by multiplying it with the amount $[(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})/(\boldsymbol{I}-\boldsymbol{K}^2\boldsymbol{B}^{-2})]$. The idea behind this is to get diagonal matrix with small values of diagonal elements which leading to decrease the shrinkage parameter, and, therefore, the resultant estimator can be better with small amount of bias. The new estimator is defined as

$$\widehat{\boldsymbol{\gamma}}_{NLRE}=(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})(\boldsymbol{I}-\boldsymbol{K}^2\boldsymbol{B}^{-2})\frac{(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})}{(\boldsymbol{I}-\boldsymbol{K}^2\boldsymbol{B}^{-2})}\widehat{\boldsymbol{\gamma}}_{LR}, \qquad (13)$$

and

$$\widehat{\boldsymbol{\beta}}_{NLRE}=\boldsymbol{D}^T\widehat{\boldsymbol{\gamma}}_{NLRE}. \qquad (14)$$

3.1 Bias, Variance, and MSE of the new estimator

The MSE of the new estimator can be obtained as

$$\text{MSE}(\widehat{\boldsymbol{\gamma}}_{NLRE})=var(\widehat{\boldsymbol{\gamma}}_{NLRE})+[\text{bias}(\widehat{\boldsymbol{\gamma}}_{NLRE})]^2 \qquad (15)$$

According to Eq. (16), the bias and variance of $\widehat{\boldsymbol{\gamma}}_{NLRE}$ can be obtained as, respectively,

$$\begin{aligned}\text{bias}(\widehat{\boldsymbol{\gamma}}_{NLRE})&=E[\widehat{\boldsymbol{\gamma}}_{NLRE}]-\boldsymbol{\gamma}\\&=(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})\,E[\widehat{\boldsymbol{\gamma}}_{LR}]-\boldsymbol{\gamma}\\&=-\boldsymbol{K}[(\boldsymbol{K}\boldsymbol{B}^{-1})^{-1}-(\boldsymbol{K}\boldsymbol{B}^{-1})^{-1}(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})+\boldsymbol{K}^2\boldsymbol{B}^{-2}(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})]\boldsymbol{B}^{-1}\boldsymbol{\gamma}, \qquad (16)\\ \text{var}(\widehat{\boldsymbol{\gamma}}_{NLRE})&=(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})\,var(\widehat{\boldsymbol{\gamma}}_{LR})(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})^T(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})^T\\&=(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})\Lambda^{-1}(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})^T(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})^T. \qquad (17)\end{aligned}$$

Then,

$$\begin{aligned}\text{MSE}(\widehat{\boldsymbol{\gamma}}_{NLRE})&=(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})\Lambda^{-1}(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})^T(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})^T+\\&\quad[-\boldsymbol{K}[(\boldsymbol{K}\boldsymbol{B}^{-1})^{-1}-(\boldsymbol{K}\boldsymbol{B}^{-1})^{-1}(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})+\boldsymbol{K}^2\boldsymbol{B}^{-2}(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})]\boldsymbol{B}^{-1}\boldsymbol{\gamma}]\\&\quad[-\boldsymbol{K}[(\boldsymbol{K}\boldsymbol{B}^{-1})^{-1}-(\boldsymbol{K}\boldsymbol{B}^{-1})^{-1}(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})+\boldsymbol{K}^2\boldsymbol{B}^{-2}(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})]\boldsymbol{B}^{-1}\boldsymbol{\gamma}]^T\\&=\Phi\Lambda^{-1}\Phi^T+\boldsymbol{K}\Psi\boldsymbol{B}^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^T\boldsymbol{B}^{-1}\Psi^T\boldsymbol{K}, \quad (18)\end{aligned}$$

where $\Phi=(\boldsymbol{I}-\boldsymbol{K}^3\boldsymbol{B}^{-3})^T(\boldsymbol{I}-\boldsymbol{K}\boldsymbol{B}^{-1})$ and $\Psi=[\boldsymbol{I}+\boldsymbol{K}\boldsymbol{B}^{-1}-\boldsymbol{K}\boldsymbol{B}^{-3}\boldsymbol{K}]$.

3.2 Selection of parameter k

The efficiency of ridge estimator strongly depends on appropriately choosing the k parameter. To estimate the values of $k$for our new estimator, the most well-known used estimation methods are employed and are given below (Kibria et al., 2015)

Hoerl and Kennard (1970) (HK), which is defined as

$$k_j(HK)=\frac{1}{\hat{\alpha}^2_{max}} \quad (19)$$

Kibria et al. (2015) (KMS1), which is defined as

$$k_j(\text{KMS1})=\text{Median}\left\{\left[\sqrt{\frac{1}{\hat{\alpha}_j^2}}\right]^2\right\},\ j=1,2,\dots,p, \qquad (20)$$

Kibria et al. (2015) (KMS2), which is defined as (21)

In literature, there are many estimators available to estimate the ridge parameter k for various models. We have considered only three estimators to show the benefits of the proposed estimator. However, for more on the estimation of k for various models, we refer our readers to Hoerl and Kennard (1970), Le Cessie and Van Houwelingen (1992), Kibria (2003), Saleh and Kibria (2012), Firinguettia, Kibria, and Araya (2017), Arashi, Kibria, and Valizade (2017) and very recently Williams, Kibria, and Mansson (2019) and Saleh, Arashi, and Kibria (2019) among others".

4. Simulation study

In this section, a Monte Carlo simulation experiment is used to examine the performance of the new estimator with different degrees of multicollinearity.

4.1 Simulation design

The response variable of n observations is generated from logistic regression model by Eq. (1) with $\sum_{j=1}^{p}\beta_j^2 = 1$ and $\beta_1 = \beta_2 = \ldots = \beta_p$ (Kibria, 2003; Månsson & Shukur, 2011). In addition, the explanatory variables $\boldsymbol{x}_i^T = (x_{i1}, x_{i2}, \ldots, x_{in})$ have been generated from the following formula

$$x_{ij} = (1-\rho^2)^{1l2} w_{ij} + \rho w_{i,p+1}, \quad i = 1,2,\ldots,n, \quad j = 1,2,\ldots,p, \qquad (22)$$

where ρ represents the correlation between the explanatory variables and w_{ij}'s are independent standard normal pseudo-random numbers. "Because the sample size has direct impact on the prediction accuracy, three representative values of the sample size are considered: 30, 100 and 150. In addition, the number of the explanatory variables is considered as $p = 4$ and $p = 8$ because increasing the number of explanatory variables can lead to increase the MSE. Further, because we are interested in the effect of multicollinearity, in which the degrees of correlation considered more important, three values of the pairwise correlation are considered with $\rho = \{0.90, 0.95, 0.99\}$. For a combination of these different values of n, p, and ρ the generated data is repeated 1000 times and the averaged mean squared errors (MSE) is calculated as

$$\text{MSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{1000}\sum_{i=1}^{1000}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}), \qquad (23)$$

where $\hat{\boldsymbol{\beta}}$ is the estimated coefficients for the used estimator.

4.2 Simulation results

The estimated MSE of Eq. (24) for MLE, LRE, and JLRE, for all the different selection methods of k and the combination of n, p, and ρ, are respectively summarized in Tables 1 and 2. Several observations can be made.

First, in terms of ρ values, there is increasing in the MSE values when the correlation degree increases regardless the value of n, p. However, JLRE performs better than LRE and MLE for all the different selection methods of kFor instance, in Table 1, when $p = 4$, $n = 100$, and $\rho = 0.95$, the MSE of NRLR was about 51.64%, 35.81%, and 20.81% lower than that of LRE for KH, KMS1 and KMS2, respectively. In addition, the MSE of JLRE was about 94.73% lower than that of MLE.

Second, regarding the number of explanatory variables, it is easily seen that there is increasing in the MSE values when the p increasing from four variables to eight variables. Although this increasing can affected the quality of an estimator, JLRE is achieved the lowest MSE comparing with MLE and LRE, for different n, ρ and different selection methods of k.

Third, with respect to the value of n, The MSE values decreases when n increases, regardless the value of ρ, p, and the value of k. However, JLRE still consistently outperforms LRE and MLE by providing the lowest MSE.

Finally, for the different selection methods of k, the performance of all methods suggesting that the JLRE estimator is better than the other used two estimators. The KMS1 efficiently provides less MSE comparing with the KMS1 and KH for both JLRE and LRE estimators. Besides, KH is more efficient for providing less MSE than KMS2 or both NRLE and RLE estimators.

To summary, all the considered values of n, ρ, p, and the value of k, JLRE is superior to LRE, clearly indicating that the new proposed estimator is more efficient".

Table 1: MSE values when $p = 4$

			KH		KMS1		KMS2	
		MLE	LRE	JLRE	LRE	JLRE	LRE	JLRE
	ρ							
$n = 30$	0.90	5.728	1.767	1.614	1.407	1.306	2.152	2.052
	0.95	6.356	1.998	1.847	1.856	1.755	2.313	2.21
	0.99	6.754	2.648	2.496	2.388	2.287	2.657	2.556
$n = 100$	0.90	4.099	1.4	1.247	1.318	1.217	1.693	1.592
	0.95	5.174	1.672	1.519	1.427	1.326	1.789	1.688
	0.99	5.366	1.989	1.836	2.317	2.216	2.002	1.902
$n = 150$	0.90	3.942	1.202	1.049	1.289	1.188	1.55	1.449
	0.95	4.152	1.326	1.173	1.316	1.215	1.566	1.465

	0.99	4.907	2.352	2.199	2.092	1.992	1.684	1.583

Table 2: MSE values when $p = 8$

		MLE	KH LRE	JLRE	KMS1 LRE	JLRE	KMS2 LRE	JLRE
	ρ							
$n = 30$	0.90	6.232	2.27	2.117	1.91	1.809	2.655	2.554
	0.95	6.859	2.502	2.35	2.359	2.258	2.814	2.713
	0.99	7.257	3.153	2.999	2.892	2.79	3.16	3.059
$n = 100$	0.90	4.602	1.903	1.75	1.822	1.72	2.196	2.095
	0.95	5.677	2.175	2.022	1.93	1.829	2.292	2.193
	0.99	5.869	2.492	2.339	2.82	2.719	2.505	2.404
$n = 150$	0.90	4.445	1.705	1.552	1.792	1.692	2.053	1.952
	0.95	4.655	1.829	1.676	1.819	1.718	2.069	1.968
	0.99	5.41	2.855	2.702	2.595	2.494	2.187	2.086

5. Real data application

A dataset of 121 molecules of anti-hepatitis C virus activity of thiourea derivatives was used for constructing quantitative structure-activity relationship (QSAR) model. “The molecular structures and their experimental EC50 (the concentration of a drug that gives a half-maximal response) were obtained from the literature (Kang, Wang, Hsu, et al., 2009; Kang, Wang, Lee, et al., 2009; Kang et al., 2010; Khatri, Lather, & Madan, 2015). The molecules were divided into two categories by the threshold value of 0.1μM: actives (EC50 $< 0.1\mu$M) and inactive (EC50 $\geq 0.1\mu$M). For the classification purpose, the two categories were labeled as 1 for the active and 0 for the inactive. First, the deviance test (Montgomery, Peck, & Vining, 2015) is used to check whether the logistic regression model is fit well to this data or not. The result of the residual deviance test is equal to 8.027 with 120 degrees of freedom and the p-value is 0.837. It is indicated form this result that the logistic regression model fits very well to this data.

Second, to check whether there are relationships between the explanatory variables or not, Table 3 displays the correlation matrix among the five explanatory variables. It is obviously seen that there are correlations greater than 0.90 among several variables.

Third, to test the existence of multicollinearity, the eigenvalues of the matrix $\boldsymbol{X}^T\widehat{\boldsymbol{W}}\boldsymbol{X}$ are obtained as 941.295, 201.332, 71.385, 36.588, 20.602, and 1.324. The determined condition number $\mathrm{CN} = \sqrt{\lambda_{\max} / \lambda_{\min}}$ of the data is 29.9026.663 indicating that the multicollinearity issue is existing.

The estimated logistic regression coefficients, standard errors which are computed by using bootstrap with 1000 replications, and MSE values for the MLE, LRE, and JLRE estimators are listed in Table 4. According to Table 4, it is clearly seen that the JLRE estimator shrinkages the value of the estimated coefficients efficiently. Additionally, in terms of the calculated standard errors, the LRE and JLRE show substantial decreasing comparing with MLE, regardless of the selection method of k. Furthermore, in terms of the selection method of k, JLRE shows the superiority results of both coefficient estimation and standard error using KMS1. In terms of MSE, the JLRE using KMS1 achieves the lowest MSE”.

Table 3: The correlation matrix among the five explanatory variables.

	Mor02u	RDF015u	Mor25v	PJI3
CIC3	0.912	0.102	0.889	0.957
Mor02u		0.875	0.947	0.624
RDF015u			0.913	0.806
Mor25v				0.962

Table 4: The estimated coefficients and MSE values for the MLE, LRE, and JLRE estimators. The number in parenthesis is the standard error.

		KH		KMS1		KMS2	
	MLE	LRE	JLRE	LRE	JLRE	LRE	JLRE
$\hat{\beta}_1$	-3.041	-2.105	-1.516	-1.604	-1.415	-1.213	-0.284
	(0.111)	(0.103)	(0.097)	(0.088)	(0.078)	(0.077)	(0.073)
$\hat{\beta}_2$	2.329	2.035	2.004	2.032	1.924	1.440	1.603
	(0.123)	(0.113)	(0.101)	(0.111)	(0.102)	(0.089)	(0.086)
$\hat{\beta}_3$	1.561	1.107	1.016	0.986	0.911	0.546	0.329
	(0.124)	(0.124)	(0.118)	(0.114)	(0.098)	(0.108)	(0.085)
$\hat{\beta}_4$	-3.168	-2.046	-1.934	-1.863	-1.521	1.604	1.121
	(0.214)	(0.204)	(0.188)	(0.124)	(0.117)	(0.102)	(0.080)
$\hat{\beta}_5$	2.0431	1.017	1.008	1.014	0.984	0.919	0.508
	(0.127)	(0.110)	(0.104)	(0.111)	(0.103)	(0.103)	(0.092)
MSE	4.102	3.557	2.397	1.981	1.761	1.242	0.961

6. Conclusion

To solve the multicollinearity issue in the logistic regression model, a new estimator of logistic ridge regression is suggested in this study. Studies using Monte Carlo simulation show that the novel estimator performs better in terms of MSE than both the maximum likelihood estimator and the conventional logistic ridge estimator. To further demonstrate the advantages of employing the novel estimator in the context of the logistic regression model, a real data application is also taken into consideration. The effectiveness of the new estimator based on the resulting MSE was noted, and it was demonstrated that the findings are in line with those of Monte Carlo simulations. Finally, it is advised to employ the new estimator when the logistic regression model has multicollinearity.

References

1. Algamal, Z. Y., & Lee, M. H. (2015). Penalized Poisson Regression Model using adaptive modified Elastic Net Penalty. Electronic Journal of Applied Statistical Analysis, 8(2), 236-245.
2. Arashi, M., Kibria, B. M. G., & Valizade, T. (2017). On ridge parameter estimators under stochastic subspace hypothesis. Journal of Statistical Computation and Simulation, 87(5), 966–983.
3. Asar, Y., & Genç, A. (2015). New shrinkage parameters for the Liu-type logistic estimators. Communications in Statistics - Simulation and Computation, 45(3), 1094-1103. doi:10.1080/03610918.2014.995815.
4. Batah, F. S. M., Ramanathan, T. V., & Gore, S. D. (2008). The efficiency of modefied jackknife and ridge type regression estimators - A comparison. Surveys in Mathematics and its Applications, 3, 111 – 122.
5. Firinguettia, L., Kibria, B. M. G., & Araya, R. (2017). Study of Partial Least Squares and Ridge Regression Methods. Communications in Statistics - Simulation and Computation, 46(8), 6631-6644.
6. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55-67.
7. Kang, I. J., Wang, L. W., Hsu, S. J., Lee, C. C., Lee, Y. C., Wu, Y. S., . . . Chern, J. H. (2009). Design and efficient synthesis of novel arylthiourea derivatives as potent hepatitis C virus inhibitors. Bioorganic & Medicinal Chemistry Letters, 19(21), 6063-6068. doi:10.1016/j.bmcl.2009.09.037.
8. Kang, I. J., Wang, L. W., Lee, C. C., Lee, Y. C., Chao, Y. S., Hsu, T. A., & Chern, J. H. (2009). Design, synthesis, and anti-HCV activity of thiourea compounds. Bioorganic & Medicinal Chemistry Letters, 19(7), 1950-1955. doi:10.1016/j.bmcl.2009.02.048.
9. Kang, I. J., Wang, L. W., Yeh, T. K., Lee, C. C., Lee, Y. C., Hsu, S. J., . . . Chern, J. H. (2010). Synthesis, activity, and pharmacokinetic properties of a series of conformationally-restricted thiourea analogs as novel hepatitis C virus inhibitors. Bioorganic & Medicinal Chemistry, 18(17), 6414-6421. doi:10.1016/j.bmc.2010.07.002.

10. Khatri, N., Lather, V., & Madan, A. K. (2015). Diverse classification models for anti-hepatitis C virus activity of thiourea derivatives. Chemometrics and Intelligent Laboratory Systems 140, 13-21. doi:10.1016/j.chemolab.2014.10.007.
11. Khurana, M., Chaubey, Y. P., & Chandra, S. (2014). Jackknifing the ridge regression estimator: A revisit. Communications in Statistics-Theory and Methods, 43(24), 5249-5262.
12. Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. Communications in Statistics - Simulation and Computation, 32(2), 419-435. doi:10.1081/SAC-120017499.
13. Kibria, B. M. G., Månsson, K., & Shukur, G. (2015). A Simulation Study of Some Biasing Parameters for the Ridge Type Estimation of Poisson Regression. Communications in Statistics - Simulation and Computation, 44(4), 943-957. doi:10.1080/03610918.2013.796981.
14. Le Cessie, S., & Van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. Journal of the Royal Statistical Society: Series C (Applied Statistics), 41(1), 191-201.
15. Lee, A., & Silvapulle, M. (1988). Ridge estimation in logistic regression. Communications in Statistics-Simulation and Computation, 17(4), 1231-1257.
16. Månsson, K., & Shukur, G. (2011). A Poisson ridge regression estimator. Economic Modelling, 28(4), 1475-1481. doi:10.1016/j.econmod.2011.02.030.
17. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). Introduction to linear regression analysis. New York: John Wiley & Sons.
18. Nyquist, H. (1988). Applications of the jackknife procedure in ridge regression. Computational Statistics & Data Analysis, 6(2), 177-183.
19. Özkale, M. R. (2008). A jackknifed ridge estimator in the linear regression model with heteroscedastic or correlated errors. Statistics & Probability Letters, 78(18), 3159-3169. doi:10.1016/j.spl.2008.05.039.
20. Rashad, N. K., & Algamal, Z. Y. (2019). A New Ridge Estimator for the Poisson Regression Model. Iranian Journal of Science and Technology, Transactions A: Science. doi:10.1007/s40995-019-00769-3.
21. Saleh, A. K. M. E., Arashi, M., & Kibria, B. M. G. (2019). Theory of Ridge Regression Estimation with Applications. New York: Wiley.
22. Saleh, A. K. M. E., & Kibria, B. M. G. (2012). Improved ridge regression estimators for the logistic regression model. Computational Statistics, 28(6), 2519-2558.
23. Schaefer, R., Roi, L., & Wolfe, R. (1984). A ridge logistic estimator. Communications in Statistics-Theory and Methods, 13(1), 99-113.
24. Singh, B., Chaubey, Y., & Dwivedi, T. (1986). An almost unbiased ridge estimator. Sankhyā: The Indian Journal of Statistics, Series B, 13, 342-346.
25. Türkan, S., & Özel, G. (2015). A new modified Jackknifed estimator for the Poisson regression model. Journal of Applied Statistics, 43(10), 1892-1905. doi:10.1080/02664763.2015.1125861.
26. Williams, U., Kibria, B. M. G., & Mansson, K. (2019). Performance of some ridge regression estimators for the Logistic Regression Model: An Empirical Comparison. International Journal of Statistics and Reliability Engineering, 6(1), 1-12.