A Modified Ridge Estimator for the Logistic Regression Model

Mazin M. Alanaz
Department of Operation Research and Intelligence Techniques, University of Mosul, Iraq.

Nada Nazar Alobaidi
Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

Zakariya Yahya Algamal*
Department of Statistics and Informatics University of Mosul, Iraq

The ridge estimator has been consistently demonstrated to be an attractive shrinkage method to reduce the effects of multicollinearity. The logistic regression model is a well-known model in application when the response variable is binary data. However, it is known that multicollinearity negatively affects the variance of maximum likelihood estimator of the logistic regression coefficients. To address this problem, a logistic ridge regression model has been proposed by numerous researchers. In this paper, a modified logistic ridge estimator (MLRE) is proposed and derived. The idea behind the MLRE is to get diagonal matrix with small values of diagonal elements that leading to decrease the shrinkage parameter and, therefore, the resultant estimator can be better with small amount of bias. Our Monte Carlo simulation results suggest that the MLRE estimator can bring significant improvement relative to other existing estimators.

Keywords: *multicollinearity, ridge estimator, logistic regression model, shrinkage, Monte Carlo simulation*

I. Introduction

Logistic regression model is widely applied for studying several real data problems, such as in medicine (Algamal and Lee 2015a). In dealing with the

* *Corresponding author: zakariya.algamal@uomosul.edu.iq*

logistic regression model, it is assumed that there is no correlation among the explanatory variables. In practice, however, this assumption often not holds, which leads to the problem of multicollinearity. 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 et al. 2015). Numerous remedial methods have been proposed to overcome the problem of multicollinearity. The ridge regression method (Hoerl and 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 and Genç 2015; Rashad and Algamal 2019). This is done by adding a positive amount to the diagonal of $\boldsymbol{X}^T\boldsymbol{X}$. As a result, the ridge estimator is biased but it guaranties a smaller mean squared error than the ML estimator.

In linear regression, the ridge estimator is defined as

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

where $\boldsymbol{y}$ is an n x 1 vector of observations of the response variable, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_p)$ is an n x p known design matrix of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is a p x 1 vector of unknown regression coefficients, $\mathbf{I}$ is the identity matrix with dimension p x 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 and Lee 2015b; Hoerl and Kennard 1970).

2. Logistic Ridge Regression 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 variables. In binary classification, the response variable of the logistic regression has two values either 1 for the tumor class, or 0 for the normal class. Let $\mathbf{y}_i \in \{0,1\}$ be a vector of size n x 1 of tissues, and let $\mathbf{x}_j$ be a p x 1 vector of variables. The logistic transformation of the vector of probability estimates $\pi_i = p(y_i = 1|\mathbf{x}_j)$ is modeled by a linear function, logit transformation,

$$\ln\left[\pi_i/1-\pi_i\right] = \beta_0 + \sum_{j=1}^{P}\mathbf{x}_j^T\beta_j, i = 1,2,\ldots,n, \tag{2}$$

where β_0 is the intercept, and β_j is a p x 1 vector of unknown variable coefficients. The log-likelihood function of Eq. (1) is defined as

$$\ell(\beta_0,\boldsymbol{\beta}) = \sum_{i=1}^{n}\left\{\mathbf{y}_i \ln \pi(\mathbf{x}_{ij}) + (1-\mathbf{y}_i)\ln(1-\pi(\mathbf{x}_{ij})\right\}. \tag{3}$$

Logistic regression offers the advantage of simultaneously estimating the probabilities $\pi(\mathbf{x}_{ij})$ and $1-\pi(\mathbf{x}_{ij})$ for each class and classifying subjects. The probability of classifying the i^{th} sample in class 1 is estimated by $\hat{\pi}_i = \exp\left(\beta_0 + \sum_{j=1}^{p}\mathbf{x}_j^T\beta_j\right) \Big/ 1+\exp\left(\beta_0 + \sum_{j=1}^{p}\mathbf{x}_j^T\beta_j\right)$ (Algamal and Lee 2017; Algamal and Lee 2018; Algamal et al. 2017). The predicted class is then obtained by $I\{\hat{\pi}_i > 0.5\}$, where $I(\bullet)$ is an indicator function. The ML estimator is then obtained by computing the first derivative of the Eq. (2) and setting it equal to zero. Then, ML estimators of the logistic regression parameters (LRM) as

$$\hat{\boldsymbol{\beta}}_{LRM} = (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\hat{\mathbf{v}}, \tag{4}$$

where $\hat{\mathbf{W}} = \text{diag}(\hat{\theta}_i)$ and $\hat{\mathbf{v}}$ is a vector where i^{th} element equals to logit link function. The ML estimator is asymptotically normally distributed with a covariance matrix that corresponds to the inverse of the Hessian matrix

$$\text{cov}(\hat{\boldsymbol{\beta}}_{LRM}) = \left[-E\left(\frac{\partial^2\ell(\boldsymbol{\beta})}{\partial\beta_i\partial\beta_k}\right)\right]^{-1} = (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}. \tag{5}$$

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

$$\begin{aligned}\text{MSE}(\hat{\boldsymbol{\beta}}_{LRM}) &= E(\hat{\boldsymbol{\beta}}_{LRM} - \hat{\boldsymbol{\beta}})^T(\hat{\boldsymbol{\beta}}_{LRM} - \hat{\boldsymbol{\beta}}) \\ &= tr\left[(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\right] \\ &= \sum_{j=1}^{p}\frac{1}{\lambda_j},\end{aligned} \tag{6}$$

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

In the presence of multicollinearity, the matrix $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ becomes ill-conditioned leading to high variance and instability of the ML estimator of the Poisson regression parameters (Algamal 2018a; Algamal 2018b; Algamal and Alanaz 2018; Algamal and Asar 2018; Alkhateeb and Algamal 2020; Yahya Algamal 2018). As a remedy, Schaefer et al. (1984) proposed the logistic ridge regression model (LRRM) as

$$\hat{\boldsymbol{\beta}}_{LRRM} = (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}+k\mathbf{I})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}\hat{\boldsymbol{\beta}}_{LRM}$$
$$= (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}+k\mathbf{I})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\hat{\mathbf{v}}, \tag{7}$$

where $k \geq 0$. The ML estimator can be considered as a special estimator from Eq. (7) with $k = 0$. Regardless of k value, the MSE of the $\hat{\boldsymbol{\beta}}_{LRRM}$ is smaller than that of $\hat{\boldsymbol{\beta}}_{LRM}$ because the MSE of $\hat{\boldsymbol{\beta}}_{LRRM}$ is equal to (Asar et al. 2017; Asar and Genç 2015; Kibria et al. 2012; Lukman et al. 2020; Månsson et al. 2011; Schaefer et al. 1984; Wu et al. 2016)

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

where α_j is defined as the j^{th} element of $\gamma\hat{\boldsymbol{\beta}}_{LRM}$ and γ is the eigenvector of the $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ matrix. Comparing with the MSE of Eq. (6), $\text{MSE}(\hat{\boldsymbol{\beta}}_{LRRM})$ is always small for $k > 0$.

3. The New Estimator

In this section, the new estimator is introduced and derived. Let $\mathbf{M} = (m_1, m_2, \ldots, m_p)$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, respectively, "be the matrices of eigenvectors and eigenvalues of the $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$ matrix, such that $\mathbf{M}^T\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}\mathbf{M} = \mathbf{S}^T\hat{\mathbf{W}}\mathbf{S} = \Lambda$, where $\mathbf{S} = \mathbf{XM}$. Consequently, the logistic regression estimator of Eq. (4), $\hat{\boldsymbol{\beta}}_{LRM}$, can be written as

$$\hat{\boldsymbol{\gamma}}_{LRM} = \Lambda^{-1}S^T\hat{\mathbf{W}}\hat{\mathbf{v}}$$
$$\hat{\boldsymbol{\beta}}_{LRM} = \mathbf{M}\hat{\boldsymbol{\gamma}}_{LRM}. \tag{9}$$

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

$$\hat{\gamma}_{LRRM} = (\boldsymbol{\Lambda}+\mathbf{K})^{-1}\mathbf{S}^T\hat{\mathbf{W}}\mathbf{v}$$
$$= (\mathbf{I}-\mathbf{K}\mathbf{D}^{-1})\hat{\gamma}_{LRM}, \tag{10}$$

where $\mathbf{D} = \Lambda + \mathbf{K}$ and $\mathbf{K} = \text{diag}(k_1, k_2, \ldots, k_p)$; $k_i \geq 0$, $i = 1,2, \ldots, p$.

In generalized ridge estimator, the Jackknifing approach was used (Khurana et al. 2014; Nyquist 1988; Singh et al. 1986). Batah et al. (2008) proposed a modified Jackknifed ridge regression estimator in linear regression model.

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

$$\hat{\boldsymbol{\gamma}}_{JE} = (\mathbf{I} - \mathbf{K}^2\mathbf{D}^{-2})\hat{\boldsymbol{\gamma}}_{LRM}, \tag{11}$$

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

$$\hat{\boldsymbol{\gamma}}_{MJE} = (\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})(\mathbf{I} - \mathbf{K}^2\mathbf{D}^{-2})\hat{\boldsymbol{\gamma}}_{LRM}. \tag{12}$$

Consequently, our modified estimator is an improvement of Eq. (12) by multiplying it with the amount $[(\mathbf{I}-\mathbf{K}^3\mathbf{D}^{-3}) / (\mathbf{I}-\mathbf{K}^2\mathbf{D}^{-2})$. 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

$$\hat{\boldsymbol{\gamma}}MLRE = (\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})(\mathbf{I} - \mathbf{K}^2\mathbf{D}^{-2})\frac{(\mathbf{I} - \mathbf{K}^3\mathbf{D}^{-3})}{(\mathbf{I} - \mathbf{K}^2\mathbf{D}^{-2})}\hat{\boldsymbol{\gamma}}LRM, \tag{13}$$

and

$$\hat{\boldsymbol{\beta}}_{MLRE} = \mathbf{M}^T\hat{\boldsymbol{\gamma}}_{MLRE}. \tag{14}$$

4. Bias, Variance, and MSE of the New Estimator

The MSE of the new estimator can be obtained as

$$\mathrm{MSE}(\hat{\boldsymbol{\gamma}}_{MLRE}) = \mathrm{var}(\hat{\boldsymbol{\gamma}}_{MLRE}) + \left[\mathrm{bias}(\hat{\boldsymbol{\gamma}}_{MLRE})\right]^2 \tag{15}$$

According to Eq. (15), the bias and variance of $\hat{\boldsymbol{\gamma}}_{MLRE}$ can be obtained as, respectively,

$$\begin{aligned}\mathrm{bias}(\hat{\gamma}_{MLRE}) &= E\left[\hat{\gamma}_{MLRE}\right] - \boldsymbol{\gamma} \\ &= (\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})(\mathrm{I} - \mathbf{K}^3\mathbf{D}^{-3})E\left[\hat{\gamma}_{MLRE}\right] - \gamma \\ &= -\mathbf{K}\left[(\mathbf{K}\mathbf{D}^{-1})^{-1} - (\mathbf{K}\mathbf{D}^{-1})^{-1}(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1}) + \mathbf{K}^2\mathbf{D}^{-2}(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})\right]\mathbf{D}^{-1}\gamma,\end{aligned} \tag{16}$$

$$\begin{aligned}\mathrm{var}(\hat{\boldsymbol{\gamma}}_{MLRE}) &= (\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})(\mathbf{I} - \mathbf{K}^3\mathbf{D}^{-3})\mathrm{var}(\hat{\boldsymbol{\gamma}}_{MLRE})(\mathbf{I} - \mathbf{K}^3\mathbf{D}^{-3})^T(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})(\mathbf{I} - \mathbf{K}^3\mathbf{D}^{-3})\Lambda^{-1}(\mathbf{I} - \mathbf{K}^3\mathbf{D}^{-3})^T(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})^T.\end{aligned} \tag{17}$$

Then,

$$\begin{aligned}\mathrm{MSE}(\hat{\boldsymbol{\gamma}}_{MLRE}) &= (\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})(\mathbf{I} - \mathbf{K}^3\mathbf{D}^{-3})\Lambda^{-1}(\mathbf{I} - \mathbf{K}^3\mathbf{D}^{-3})^T(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})^T + \\ &\quad \left[-\mathbf{K}\left[(\mathbf{K}\mathbf{D}^{-1})^{-1} - (\mathbf{K}\mathbf{D}^{-1})^{-1}(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1}) + \mathbf{K}^2\mathbf{D}^{-2}(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})\right]\mathbf{D}^{-1}\boldsymbol{\gamma}\right] \\ &\quad \left[-\mathbf{K}\left[(\mathbf{K}\mathbf{D}^{-1})^{-1} - (\mathbf{K}\mathbf{D}^{-1})^{-1}(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1}) + \mathbf{K}^2\mathbf{D}^{-2}(\mathbf{I} - \mathbf{K}\mathbf{D}^{-1})\right]\mathbf{D}^{-1}\boldsymbol{\gamma}\right]^T \\ &= \boldsymbol{\Phi}\Lambda^{-1}\boldsymbol{\Phi}^T + \mathbf{K}\boldsymbol{\Psi}\mathbf{D}^{-1}\boldsymbol{\gamma}\boldsymbol{\gamma}^T\mathbf{D}^{-1}\boldsymbol{\Psi}^T\mathbf{K},\end{aligned} \tag{18}$$

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

2.7. 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).

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

$$k_j(\mathrm{HK}) = \frac{\hat{\sigma}^2}{\alpha_{\max}^2}, j = 1, 2, ..., p, \tag{19}$$

where $\hat{\sigma}^2 = \sum_{i=1}^{n} (y_i - \hat{\theta}_i)^2 / n - p - 1$.

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

$$k_j(\mathrm{KMSl}) = \mathrm{Median}\left\{\left[\sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}}\right]^2\right\}, j = 1, 2, ...p, \tag{20}$$

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

$$k_j(\mathrm{KMS2}) = \mathrm{Median}\left\{\frac{\lambda_{\max}}{(n-p)\hat{\sigma}^2 + \lambda_{\max}\hat{\alpha}_j^2}\right\}, j = 1, 2, ...p, \tag{21}$$

5. 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.

The response variable of n observations is generated from Bernoulli distribution regression model by

$$\pi_i = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})}, \tag{22}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$ with $\sum_{j=1}^{p} \beta_j^2 = 1$ and $\beta_1 = \beta_2 = \ldots, = \beta_p$ (Kibria 2003; Månsson and Shukur 2011).

The explanatory variables $x_i^T = (x_{i1}, x_{i2},\ldots, x_{in})$, have been generated from the following formula

$$x_{ij} = (1-\rho^2)^{1/2}\, w_{ij} + \rho w_{ip},\ i = 1, 2,\ldots, n,\ \ j = 1,2,\ldots, p, \tag{23}$$

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, 50 and 100. 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 are 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

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

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

6. Simulation Results

The estimated MSE of Eq. (24) for MLE, LRM, and MLRE, for all the different selection methods of k and the combination of n, p, and ρ, are summarized in Tables 1, 2, and 3, respectively. Several observations can be made.

First, in terms of ρ values, there is increasing in the MSE values when the correlation degree increases regardless of the value of n, p. However, MLRE performs better than LRM and MLE for all the different selection methods of k. For instance, in Table 1, when $p = 8$ and $\rho = 0.99$, the MSE of MLRE was about 4.38%, 3.13%, and 2.86% lower than that of LRM for KH, KMS1 and KMS2, respectively. In addition, the MSE of MLRE was about 53.51% 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 affect the quality of an estimator, MLRE is achieved the lowest MSE comparing with MLE and LRM, for different n, p and different selection methods of k.

Third, with respect to the value of n, the MSE values decrease when n increases, regardless of the value of ρ, p, and the value of k. However, MLRE still consistently outperforms LRM and MLE by providing the lowest MSE.

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

To summarize, all the considered values of n, p, ρ, and the value of k, MLRE is superior to LRM, clearly indicating that the new proposed estimator is more efficient.

Table 1. MSE values when n = 30

			KH		KMS1		KMS2	
	ρ	MLE	LRM	MLRE	LRM	MLRE	LRM	MLRE
$p = 4$	0.90	6.367	2.406	2.253	2.046	1.945	2.791	2.691
	0.95	6.995	2.637	2.486	2.495	2.394	2.952	2.849
	0.99	7.393	3.287	3.135	3.027	2.926	3.296	3.195
$p = 8$	0.90	6.472	2.608	2.455	2.238	2.137	2.986	2.885
	0.95	7.091	2.839	2.686	2.687	2.586	3.145	3.044
	0.99	7.506	3.489	3.336	3.219	3.118	3.491	3.391

Table 2. MSE values when *n* = 50

			KH		KMS1		KMS2	
	ρ	MLE	LRM	MLRE	LRM	MLRE	LRM	MLRE
$p = 4$	0.90	6.04	2.079	1.926	1.719	1.618	2.464	2.363
	0.95	6.668	2.312	2.159	2.168	2.067	2.623	2.522
	0.99	7.066	2.962	2.808	2.711	2.599	2.969	2.868
$p = 8$	0.90	6.145	2.281	2.128	1.911	1.811	2.659	2.558
	0.95	6.764	2.512	2.359	2.362	2.259	2.818	2.717
	0.99	7.179	3.162	3.009	2.892	2.791	3.164	3.063

Table 3. MSE values when *n* = 100

			KH		KMS1		KMS2	
	ρ	MLE	LRM	MLRE	LRM	MLRE	LRM	MLRE
$p = 4$	0.90	5.628	1.667	1.514	1.307	1.206	2.052	1.951
	0.95	6.256	1.898	1.747	1.756	1.655	2.211	2.112
	0.99	6.654	2.548	2.396	2.288	2.187	2.557	2.456
$p = 8$	0.90	5.733	1.869	1.716	1.499	1.398	2.247	2.146
	0.95	6.352	2.141	1.947	1.948	1.847	2.406	2.305
	0.99	6.767	2.751	2.597	2.481	2.379	2.752	2.651

7. Conclusion

In this paper, a modified estimator of logistic ridge regression is proposed to overcome the multicollinearity problem in the logistic regression model. According to Monte Carlo simulation studies, the modified estimator has a better performance than the maximum likelihood estimator and ordinary logistic ridge estimator, in terms of MSE. In conclusion, the use of the modified estimator is recommended when multicollinearity is present in the logistic regression model.

References

Algamal ZY. 2018a. "Developing a Ridge Estimator for the Gamma Regression Model." *Journal of Chemometrics* 32. doi:10.1002/cem.3054.

Algamal ZY. 2018b. "A New Method for Choosing the Biasing Parameter in Ridge Estimator for Generalized Linear Model." *Chemometrics and Intelligent Laboratory Systems* 183: 96-101.

Algamal ZY, Alanaz MM. 2018. "Proposed Methods in Estimating the Ridge Regression Parameter in Poisson Regression Model." *Electronic Journal of Applied Statistical Analysis* 11: 506-515.

Algamal ZY, Asar Y. 2018. "Liu-type Estimator for the Gamma Regression Model." *Communications in Statistics - Simulation and Computation* 49:2035-2048 doi:10.1080/03610918.2018.1510525.

Algamal ZY, Lee MH. 2015a. "High Dimensional Logistic Regression Model Using Adjusted Elastic Net Penalty." *Pakistan Journal of Statistics and Operation Research* 11: 667-676.

Algamal ZY, Lee MH. 2015b. "Penalized Poisson Regression Model Using Adaptive Modified." *Elastic Net Penalty Electronic Journal of Applied Statistical Analysis* 8: 236-245.

Algamal ZY, Lee MH. 2017. "A Novel Molecular Descriptor Selection Method in QSAR Classification Model Based on Weighted Penalized Logistic Regression." *Journal of Chemometrics* 31 doi:10.1002/cem.2915.

Algamal ZY, Lee MH. 2018. "A Two-stage Sparse Logistic Regression for Optimal Gene Selection in High-dimensional microarray data classification." *Advances in Data Analysis and Classification* 13: 753-771 doi:10.1007/s11634-018-0334-1.

Algamal ZY, Lee MH, Al-Fakih AM, Aziz M. 2017. "High-dimensional QSAR Classification Model for Anti-hepatitis C Virus Activity of Thiourea Derivatives Based on the Sparse Logistic Regression Model with a Bridge Penalty." *Journal of Chemometrics* 31: e2889.

Alkhateeb A, Algamal Z. 2020. "Jackknifed Liu-type Estimator in Poisson Regression Model." *Journal of the Iranian Statistical Society* 19:21-37. doi:10.29252/jirss.19.1.21

Asar Y, Arashi M, Wu JJ. 2017. "CiS-S." *Computation Restricted Ridge Estimator in the Logistic Regression Model* 46: 6538-6544.

Asar Y, Genç A. 2015. "New Shrinkage Parameters for the Liu-type Logistic Estimators Communications." *Statistics - Simulation and Computation* 45: 1094-1103 doi:10.1080/03610918.2014.995815.

Batah FSM, Ramanathan TV, Gore SD. 2008. "The Efficiency of Modified Jackknife and Ridge Type Regression Estimators." *A comparison Surveys in Mathematics and its Applications* 3:111 – 122.

Hoerl AE, Kennard RW. 1970. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics* 12:55-67.

Khurana M, Chaubey YP, Chandra S. 2014. "Jackknifing the Ridge Regression Estimator: A Revisit Communications." *Statistics-Theory and Methods* 43: 5249-5262.

Kibria BG, Månsson K, Shukur GJCE. 2012. *Performance of Some Logistic Ridge Regression Estimators* 40: 401-414.

Kibria BMG. 2003. "Performance of Some New Ridge Regression Estimators Communications." *Statistics - Simulation and Computation* 32: 419-435. doi:10.1081/SAC-120017499

Kibria BMG, Månsson K, Shukur G. 2015. "A Simulation Study of Some Biasing Parameters for the Ridge Type Estimation of Poisson Regression Communications." *Statistics - Simulation and Computation* 44: 943-957. doi:10.1080/03610918.2013.796981

Lukman AF, Emmanuel A, Clement OA, Ayinde K. 2020. "A Modified Ridge-type Logistic Estimator." *Iranian Journal of Science and Technology, Transactions A: Science.* 44:437-443 doi:10.1007/s40995-020-00845-z.

Månsson K, Shukur G. 2011. "A Poisson Ridge Regression Estimator." *Economic Modelling* 28: 1475-1481. doi:10.1016/j.econmod.2011.02.030.

Månsson K, Shukur GJCiS-T, Methods (2011). *On Ridge Parameters in Logistic Regression* 40:3366-3381

Nyquist H. 1988. "Applications of the Jackknife Procedure in Ridge Regression." *Computational Statistics & Data Analysis* 6: 177-183.

Rashad NK, Algamal ZY. 2019. "A New Ridge Estimator for the Poisson Regression Model." *Iranian Journal of Science and Technology, Transactions A: Science* 43: 2921-2928. doi:10.1007/s40995-019-00769-3.

Schaefer R, Roi L, Wolfe RJCiS-T, Methods (1984) A ridge logistic estimator 13:99-113.

Singh B, Chaubey Y, Dwivedi T. 1986. "An Almost Unbiased Ridge Estimator." *Sankhyā: The Indian Journal of Statistics*, Series B 13:342-346

Wu J, Asar YJHJoM, Statistics, 2016. "On Almost Unbiased Ridge Logistic Estimator for the Logistic Regression Model 45:989-998

Yahya Algamal Z. 2018. "Performance of Ridge Estimator in Inverse Gaussian Regression Model." *Communications in Statistics - Theory and Methods* 48: 3836-3849 doi:10.1080/03610926.2018.1481977.