# Estimation methods of logistic regression in context of multicollinearity (Comparative study)

**Presented by**


Prof. Dr .**Mohamed Abdel Samie Enany**

**Professor of Applied Statistics**

**Faculty of Commerce**

**Zagazig University**

Prof. Dr . **Hassan Mohamed Ali**

**Emeritus Professor of Applied Statistics**

**Faculty of Commerce,**

**Zagazig University**

**Mahmoud Arif Awad Allah**

**A Demonstrator at the Department of Statistics and Insurance**

**Faculty of Commerce – Zagazig University**



**Journal of Business Research**

**Faculty of Commerce -Zagazig University**

**Volume 4 5- Issue 4 October 2023**

**link: https://zcom.journals.ekb.eg/**

## Abstract

The binary logistic regression (BLR) model is used as an alternative to the commonly used linear regression model when the response variable is binary. As in the linear regression model, there can be a relationship between the predictor variables in a BLR, especially when they are continuous, thus giving rise to the problem of multicollinearity. The efficiency of maximum likelihood estimator (MLE) is low in estimating the parameters of BLR when there is multicollinearity alternatively, the ridge estimator (RR), the Liu estimator (LE), the Liu-type estimator (LTE) and The Modified Ridge-Type estimator (MRTE) were developed to replace MLE. However, in this study, we compared all estimators by the mean squares errors (MSE) to get the best estimator that mitigates the effect of multicollinearity. Finally, a simulation study was conducted to illustrate the theoretical results. The result shows that the modified Ridge type estimator outperforms all other estimators followed by Liu estimator.

**Keywords:** Logistic regression, Multicollinearity, Maximum likelihood estimator, Ridge estimator, Liu estimator, Liu-type estimator, Modified Ridge-Type estimator.

## 1. Introduction

To explain the link between a binary response variable and one or more predictor variables, a binary logistic regression model is frequently utilized. In the fields of applied sciences including biostatistics, criminology, business and finance, engineering, biology, and medical research, we frequently used this model. The maximum likelihood estimator is one technique for estimating the parameters of a logistic model (ML). If the predictor variables are continuous, multicollinearity may result from their correlation. When multicollinearity is present and we use MLE, the estimation model's conclusion could not be accurate. The ridge regression (LRE) estimator credited to Hoerl and Kennard is a different approach to estimating (1970a). For the (ML), they added a biasing parameter called k to lessen the impact of multicollinearity. The LRE estimator in the linear regression model has been improved by a number of writers. Hoerl and Kennard (1970b), Hoerl et al. (1975), Lawless and Wang (1976), Kibria (2003), Khalaf and Shukur (2005), Muniz and Kibria (2009), Lukman and Ayinde (2017), Lukman et al. (2017), Lukman and Arowolo (2018), and additional studies are among them. To address the issue of multicollinearity, Schaeffer et al. (1984) modified the ridge estimator for the logistic regression model. Furthermore, Kibria et al. (2012) assess how well different biassing parameters perform in logistic ridge regression. For addressing the multicollinearity problem, Liu (1993) proposed the Liu estimator (LE), which is a biassed estimator. Urgan and Tez (2008) proposed an improved LE. The shrinkage parameter (d) has five estimators, according to Mnsson et al. (2012). Based on the works of Hoerl and Kennard (1970), Kibria (2003), and Khalaf and

Shukur, these estimators (2005). (2019) saw the introduction of two shrinkage parameter estimators from Sudjai and Duangsaphon. The work of Hoerl and Kennard (1970), Hoerl and Kennard, served as the foundation for the first estimator (1970). Based on the work of Mnsson et al., the second estimator was developed (2012).Inan and Erdogan added the Liu-type estimator to the logistic regression model (2013). For the logistic regression model, Nagarajah and Wijekoon (2015) introduced a stochastic restricted maximum likelihood estimator with stochastic linear limitations. Recently, a modified ridge-type estimator was included to a linear regression model by Lukman et al. (2019). This paper's main objective is to integrate this approach with the binary logistic regression model. The mean squared error (MSE) criterion was used to determine this new estimator's properties and compare them to those of the previous ones.

The rest of the article is organized as follows. In Section 2, Logistic Regression Model. In Section 3, we describe the Estimation methods. In Section 4, the simulation results. Finally, we conclude the study in Section 5.

## 2. Logistic Regression Model

Consider the logistic regression model with a Bernoulli distribution for the response variable (y).

$$(y;\pi) = \prod_{i=1}^{n} {\pi_i}^{y_i} (1-\pi_i)^{1-y_i} \tag{1}$$

Bernoulli can be redistributed as an exponential distribution as follows:

$$\mathrm{f}(y;\pi) = \exp\left\{y \log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right\} \tag{2}$$

From the previous equation to estimate $\pi$ given y to get likelihood function:

$$\mathrm{L}(\pi; y) = \prod_{i=1}^{n} \exp\left\{y \log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right\} \tag{3}$$

The mean and variance functions of the Bernoulli distribution follow as:

Mean = $\pi$ , Variance = $\pi\,(1-\pi)$

Through the LR model, the dependent variable is binary, which takes two values of 1 and 0 with probabilities $\pi$ and $(1-\pi)$ , y is a random variable that follows the Bernoulli distribution with parameter $E(Y) = \pi$. Since the error term $\varepsilon_i$ follows the Bernoulli distribution of the dependent variable y, therefore the regression model can be coordinated as follows:

$$y_i = E(y_i|x_i) + \varepsilon_i \quad , \quad i = 1,2,3, \dots \dots .., n \tag{4}$$

To simplify to represent Conditional mean of y given x when logistic distribution can be used quantity$\pi(x) = E(y_i|x_i)$. The general form of the LR model is:

$$\pi_i = \frac{e^{x_i\beta}}{1 + e^{x_i\beta}} \tag{5}$$

Where $x_i$ is the $i$th row of a $n \times p$ matrix X and $\beta$ is a $p \times 1$ vector of regression coefficients.

The LR model can be converted into a linear form as follows:

$$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = x_i\beta \tag{6}$$

This transformation is called the logit transformation of probability $\pi$. the ratio $(\frac{\pi}{1-\pi})$ in this transformation is called the odds. The logarithm of the transformation is often called the log-odds. The logit link is natural form of the binomial distribution. Maximum likelihood estimation the main method used to estimate the parameters of the LR model, other methods can be used to estimate model parameters. The sum of the log-likelihood value on each note in the form is mathematically easier than multiplying different probabilities. The log-likelihood function is as follows:

$$L(\beta; y) = \sum_{i=1}^{n} y_i x_i \beta - \sum_{i=1}^{n} \log(1 + \exp(x_i \beta)) \tag{7}$$

Take the derivative of the log -likelihood with respect to $\beta$ and set the result the equations are called likelihood equations to zero. The ML estimates are found by solving the subsequent equation:

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n} x_i (y_i - \pi_i) = 0 \tag{8}$$

Since Equation (8) has nonlinear parameters, the iteratively reweighted least square (IRLS) method is used to solve it. Consequently, the logistic regression model's maximum likelihood estimator is

$$\hat{\beta}_{ML} = (\grave{X} W X)^{-1} \grave{X} \widehat{W} \hat{z} \tag{9}$$

Where $\hat{\beta}_{ML}$ stands for the vector of estimated parameters for logistic model using maximum likelihood method, $\widehat{W} = diag\big(\pi_i(1 - \pi_i)\big)$ is the weighting matrix for the logistic model $and\ \hat{z} = \log(\pi_i) + \frac{Y_i-\pi_i}{\pi_i(1-\pi_i)}$ is the $i^{th}$ th element of the vector $\hat{z}$.

## 3. Estimation methods

The regression estimate generated by machine learning produces a high mean square error when multicollinearity is present (Schaeffer et al. 1984). When the predictor variables are correlated, Schaeffer et al. (1984) recommended the logistic ridge regression (LRE) as an alternative to ML. It is said that the logistic ridge regression estimator (LRE) is:

$$\hat{\beta}_{LRE} = (\grave{X}WX + kI)^{-1}\grave{X}\widehat{W}X\hat{\beta}_{ML} = V\hat{\beta}_{ML}\,, k > 0 \qquad (10)$$

It is observed that $\hat{\beta}_{ML} = \hat{\beta}_{LRE}$ when $k = 0$

The logistic Liu estimator was a version of the logistic regression model that was introduced by Mansson et al. (2012). (LE) According to its definition, this estimator:

$$\hat{\beta}_{LE} = (\grave{X}WX + \mathrm{I})^{-1}\left(\grave{X}WX + d\mathrm{I}\right)\hat{\beta}_{ML} = Z\hat{\beta}_{ML} \qquad (11)$$

The logistic version of the Liu-type estimator was proposed by Inan and Erdogan (2013) as:

$$\hat{\beta}_{LTE} = (\grave{X}WX + kI)^{-1}\left(\grave{X}WX + dI\right)\hat{\beta}_{ML} \qquad (12)$$

Recently, the modified ridge-type (MRTE) estimator in a logistic regression model was proposed by Lukman et al. (2020). A definition of this estimator is:

$$\hat{\beta}_{MRTE} = (\grave{X}WX + k(1+d)I)^{-1}\grave{X}WX\hat{\beta}_{ML} \qquad (13)$$

Where $k > 0\ and\ 0 < d < 1$

## 4. Monte Carlo Simulation

The simulation is the same as it was in the prior chapter. The explanatory variables, according to McDonald and Galarneau (1975), Gibbons (1981), Kibria (2003), and Lukman et al. (2019b), were produced from a

$$x_{ij} = (1-\rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{ip} \tag{14}$$

Where $i = 1,2,\cdots\cdots,n$ ; $j = 1,2,\cdots\cdots,$p, and $\rho^2$ represents the correlation between the explanatory variables, and $z_{ij}{}'s$ are independent random numbers obtained from the standard normal distribution.

The Bernoulli distribution (Be $(\pi_i)$ ) is used to create the response variable, where $\pi_i = \frac{e^{x_i\beta}}{1+e^{x_i\beta}}$, since $x_i$ is the ith row of the matrix X.

The parameter values were set with the restriction that $\acute{\beta}\beta$ = 1, which is typical for simulation research of this kind (Newhouse and Oman 1971).

3-The number of explanatory variables p where (p = 3, 5, 7, 9 and 13).

4- The sample size n where (n = 30, 50, 75, 100, 200 and 300).

5- The correlation among the explanatory variables $\rho$ was chosen as effective parameters in our simulation analysis ($\rho =$ 0.85, 0.9, 0.95, and 0.99).

The simulation investigation, we use MSE to compare the performance of the suggested methods. MSE is a familiar criteria as it depends on the mean which has good statistical properties. The MSE have been used in several studies such as Abonazel (2018), Asar et al. (2017) and Lukman et al. (2020), the MSE is calculated as follows:

$$MSE = (\hat{\beta}) = \frac{1}{L}\sum_{l=1}^{L}(\hat{\beta}_l - \beta)'(\hat{\beta}_l - \beta) \quad (15)$$

Where $\hat{\beta}_l$ indicates the vector of estimated parameters at lth iteration and $\beta$ is the vector of real parameter values. The experiment was repeated 2000 times. We present the estimated mean squared errors for each of the estimators in Tables 1 through 6 using the relevant bias parameters for both of which have been evaluated.

We compare the performance of five methods for the logistic model:

ML in equation (9).

LRE which defined in equation (10), and depending on small positive constant quantity LRE.

$$k = \max\left(\frac{1}{\sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}}\right) \text{(Muniz \& Kibria, 2009).}$$

LE defined in equation (11), and depending on shrinkage Parameter $d$ :

$$d_{opt} = 1 - \left[\frac{\sum_{j=1}^{p}\frac{1}{\lambda_i(\lambda_i + 1)}}{\sum_{j=1}^{p}\frac{\hat{\alpha}_j^2}{(\lambda_i + 1)^2}}\right]$$

If $d_{opt}$ is negative, we use Ozkale and Kaciranlar's biassing parameter (2007). This is defined as $\hat{d}_{alt} = \min[\frac{\hat{\alpha}_j^2}{\frac{1}{\lambda_j + \hat{\alpha}_j^2}}]$ .

LTE which defined in equation (12), and depending on the shrinkage factor used

$$k = \max(\frac{1}{\sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}})\text{(Muniz \& Kibria, 2009)}.$$

$$d_{opt} = \left[\frac{\sum_{j=1}^{p}(1 - k\alpha_j^2)/(\lambda_j + k)^2}{\sum_{j=1}^{p}(\lambda_j\alpha_j^2 + 1)/\lambda_j(\lambda_j + k)^2}\right]$$

MRTE logistic in equation (13) depending on

$$k = (\prod_{j=1}^{p}\frac{1}{(1 + d)\alpha_j^2})^{1/p}.$$

$$d_p = \left(\frac{1}{k\alpha_j^2} - 1\right)$$

## Table 1: Simulated result when p = 3

| | | MSE | | | | |
|---|---|---|---|---|---|---|
| ***n*** | ***ρ*** | **MLE** | **LRE** | **LE** | **LTE** | **MRTE** |
| 30 | 0.85 | 2.943125147 | 1.174025668 | 0.795993616 | 1.091999205 | 0.85254178 |
| | 0.9 | 4.09276339 | 1.557247136 | 0.969128032 | 1.194567641 | 0.921481467 |
| | 0.95 | 9.235789379 | 3.029998319 | 1.055393613 | 1.322146479 | 0.947221352 |
| | 0.99 | 54.69982474 | 15.68934509 | 3.128698316 | 1.381615547 | 1.191256679 |
| 50 | 0.85 | 1.612416939 | 0.857573256 | 0.657319608 | 0.796455635 | 0.854694102 |
| | 0.9 | 1.82384732 | 0.920540217 | 0.75680351 | 0.991537643 | 0.863655081 |
| | 0.95 | 5.166609418 | 1.914797853 | 1.209793888 | 1.292415333 | 0.934555072 |
| | 0.99 | 19.60832557 | 6.146044393 | 1.906778 | 1.370021504 | 1.020401814 |
| 75 | 0.85 | 0.94615015 | 0.638028011 | 0.528553939 | 0.574976117 | 0.818892136 |
| | 0.9 | 1.199821937 | 0.728579132 | 0.610275242 | 0.671393577 | 0.826978192 |
| | 0.95 | 2.926669094 | 1.283263169 | 0.975864771 | 1.054314282 | 0.891907776 |
| | 0.99 | 15.03840815 | 4.899707005 | 1.956754493 | 1.345240113 | 1.026270667 |
| 100 | 0.85 | 0.643867305 | 0.502464059 | 0.397049773 | 0.352706996 | 0.805052322 |
| | 0.9 | 0.901184381 | 0.616757784 | 0.515479534 | 0.545465722 | 0.824021005 |
| | 0.95 | 1.867702773 | 1.000908837 | 0.860540119 | 1.012469519 | 0.889385893 |
| | 0.99 | 9.815338984 | 3.370415346 | 1.771613068 | 1.292171946 | 0.998129807 |
| 200 | 0.85 | 0.294659472 | 0.305747675 | 0.2363579 | 0.165343768 | 0.730933122 |
| | 0.9 | 0.490842068 | 0.437639337 | 0.357663396 | 0.304004697 | 0.832231134 |
| | 0.95 | 0.93071839 | 0.635964734 | 0.590304044 | 0.60934311 | 0.821134518 |
| | 0.99 | 4.458416215 | 1.764496725 | 1.235270742 | 1.15092055 | 0.924148232 |
| 300 | 0.85 | 0.197347906 | 0.225173051 | 0.170137304 | 0.107774141 | 0.70581404 |
| | 0.9 | 0.317714344 | 0.317420819 | 0.254667082 | 0.19423693 | 0.762796386 |
| | 0.95 | 0.650772826 | 0.537051122 | 0.46113749 | 0.413518661 | 0.809968786 |
| | 0.99 | 2.688728625 | 1.203975949 | 1.024980009 | 1.011869142 | 0.892116673 |

## Table 2: Simulated result when p = 5

| | | | | MSE | | |
|---|---|---|---|---|---|---|
| ***n*** | ***ρ*** | **MLE** | **LRE** | **LE** | **LTE** | **MRTE** |
| 30 | 0.85 | 9.432006357 | 2.655010627 | 1.312411808 | 1.363821434 | 0.977343705 |
| | 0.9 | 15.86692291 | 3.966067949 | 1.460005671 | 1.419357883 | 1.041811407 |
| | 0.95 | 27.41333006 | 6.272570989 | 1.425981539 | 1.428116239 | 1.045088519 |
| | 0.99 | 2528.740742 | 29.01099533 | 2.216261328 | 1.422905506 | 1.41429214 |
| 50 | 0.85 | 4.045062723 | 1.467072376 | 1.170540958 | 1.214027258 | 0.927138437 |
| | 0.9 | 4.200301237 | 1.552115916 | 1.277035271 | 1.330450029 | 0.938618231 |
| | 0.95 | 9.083813454 | 2.799509708 | 1.390269002 | 1.211015029 | 0.979609146 |
| | 0.99 | 46.60760041 | 12.29991036 | 1.890229647 | 1.401156026 | 1.31961834 |
| 75 | 0.85 | 1.880083835 | 0.897044876 | 0.88760729 | 0.937146939 | 0.87891218 |
| | 0.9 | 2.67154224 | 1.163922757 | 1.137399311 | 1.155826444 | 0.899806006 |
| | 0.95 | 5.26039555 | 1.860033113 | 1.429002376 | 1.257038887 | 0.961620943 |
| | 0.99 | 28.49053191 | 8.06240609 | 1.934463784 | 1.387066357 | 1.280933725 |
| 100 | 0.85 | 1.503068821 | 0.811620254 | 0.829882574 | 0.854546846 | 0.879900287 |
| | 0.9 | 2.234349645 | 1.050138636 | 1.034546201 | 1.051277685 | 0.88385908 |
| | 0.95 | 3.75163386 | 1.434294574 | 1.207603588 | 1.079676191 | 0.902305873 |
| | 0.99 | 18.1325588 | 5.216496065 | 1.637332872 | 1.277297187 | 1.067422016 |
| 200 | 0.85 | 0.55932895 | 0.442652843 | 0.431100183 | 0.425149232 | 0.805563092 |
| | 0.9 | 0.854154743 | 0.571869912 | 0.602277869 | 0.641198443 | 0.841969372 |
| | 0.95 | 1.923637438 | 0.950905801 | 1.014893269 | 1.050748224 | 0.868314982 |
| | 0.99 | 8.580338058 | 2.8274513 | 1.666459094 | 1.190985649 | 0.99938304 |
| 300 | 0.85 | 0.386697439 | 0.331424169 | 0.321113515 | 0.312773953 | 0.821773817 |
| | 0.9 | 0.58276612 | 0.448594936 | 0.453666502 | 0.485234581 | 0.789139607 |
| | 0.95 | 1.143926689 | 0.695197095 | 0.751609584 | 0.814819576 | 0.835676391 |
| | 0.99 | 6.538653141 | 2.267521461 | 1.591812659 | 1.121206394 | 0.953350897 |

**Table 3: Simulated result when p = 7**

| | | | | MSE | | |
|---|---|---|---|---|---|---|
| ***n*** | ***ρ*** | **MLE** | **LRE** | **LE** | **LTE** | **MRTE** |
| 30 | 0.85 | 22418.66089 | 3.700878574 | 1.522956288 | 1.463461812 | 0.991295173 |
| | 0.9 | 120311.5305 | 6.305755861 | 1.50699216 | 1.416278835 | 1.03244754 |
| | 0.95 | 39027.88813 | 12.80414663 | 1.54271134 | 1.450749732 | 1.142812596 |
| | 0.99 | 286.1433774 | 39.33388371 | 1.433071835 | 1.439476811 | 1.334456546 |
| 50 | 0.85 | 6.059395649 | 1.858505938 | 1.534457468 | 1.35710633 | 0.961717904 |
| | 0.9 | 8.064580905 | 2.331414686 | 1.581149242 | 1.286253885 | 0.965465472 |
| | 0.95 | 17.54256404 | 4.603148081 | 1.753218261 | 1.318250454 | 1.058522644 |
| | 0.99 | 82.04419329 | 19.62713573 | 1.572891542 | 1.403320755 | 1.397583969 |
| 75 | 0.85 | 3.414465499 | 1.298165378 | 1.383091694 | 1.33920769 | 0.925135719 |
| | 0.9 | 4.71781738 | 1.630164766 | 1.527169793 | 1.244149393 | 0.938411869 |
| | 0.95 | 10.04693818 | 2.974401241 | 1.830938564 | 1.293890289 | 1.018512411 |
| | 0.99 | 45.416161 | 12.00514006 | 1.706207741 | 1.357033452 | 1.299379399 |
| 100 | 0.85 | 1.863208921 | 0.872726282 | 1.03362391 | 0.984292794 | 0.866889933 |
| | 0.9 | 3.051395497 | 1.234237631 | 1.385340376 | 1.273295725 | 0.907033774 |
| | 0.95 | 7.254813221 | 2.356557472 | 1.864666698 | 1.316003967 | 0.988044584 |
| | 0.99 | 33.74731784 | 9.058222887 | 1.950596988 | 1.365287698 | 1.343355626 |
| 200 | 0.85 | 0.902695768 | 0.565612566 | 0.67451354 | 0.752426597 | 0.865248668 |
| | 0.9 | 1.48389953 | 0.785391494 | 0.969479434 | 1.068744793 | 0.859800668 |
| | 0.95 | 3.046817459 | 1.275277612 | 1.470842073 | 1.350697039 | 0.892176654 |
| | 0.99 | 14.28394793 | 4.201080005 | 2.060353377 | 1.165431138 | 1.12911808 |
| 300 | 0.85 | 0.605039681 | 0.448326824 | 0.495012067 | 0.553437242 | 0.81406682 |
| | 0.9 | 0.918125028 | 0.585154961 | 0.698663138 | 0.83218815 | 0.821706798 |
| | 0.95 | 1.735582383 | 0.866585633 | 1.088998047 | 1.212471164 | 0.86376238 |
| | 0.99 | 8.553887898 | 2.755212279 | 1.980362775 | 1.131973099 | 1.009890721 |

**Table 4: Simulated result when p = 9**

| | | | MSE | | | |
|---|---|---|---|---|---|---|
| ***n*** | ***ρ*** | **MLE** | **LRE** | **LE** | **LTE** | **MRTE** |
| 50 | 0.85 | 9.694229 | 2.487900644 | 1.8568624 | 1.472461718 | 0.984378433 |
| | 0.9 | 13.83942478 | 3.394214177 | 1.918298982 | 1.354028904 | 1.00859322 |
| | 0.95 | 25.00856612 | 5.834771162 | 1.863861101 | 1.312729112 | 1.046805505 |
| | 0.99 | 165.3009457 | 35.74193345 | 1.693506819 | 1.492317875 | 1.84397882 |
| 75 | 0.85 | 5.15324346 | 1.666509203 | 1.75722269 | 1.429313714 | 0.947349367 |
| | 0.9 | 7.198625141 | 2.188379158 | 2.025205299 | 1.459604591 | 0.978512344 |
| | 0.95 | 13.97199806 | 3.837214187 | 2.176290819 | 1.30801768 | 1.038114286 |
| | 0.99 | 66.99111255 | 15.88194483 | 1.815379751 | 1.413819376 | 1.46253858 |
| 100 | 0.85 | 3.08378938 | 1.216654654 | 1.522306052 | 1.442822523 | 0.911666672 |
| | 0.9 | 4.416925685 | 1.54440972 | 1.780748102 | 1.512726946 | 0.93430819 |
| | 0.95 | 8.831974324 | 2.654236528 | 2.173241137 | 1.30836885 | 0.998570356 |
| | 0.99 | 47.23966829 | 11.57771145 | 1.999807265 | 1.391455297 | 1.45009978 |
| 200 | 0.85 | 1.251667369 | 0.704361872 | 0.916981168 | 1.066111174 | 0.851763708 |
| | 0.9 | 1.861037065 | 0.899137458 | 1.217500959 | 1.346409134 | 0.868881556 |
| | 0.95 | 3.793706855 | 1.43592567 | 1.807837778 | 1.585842562 | 0.917966514 |
| | 0.99 | 19.06025579 | 5.369396182 | 2.357234717 | 1.182140774 | 1.178729547 |
| 300 | 0.85 | 0.841686072 | 0.54563428 | 0.677308605 | 0.798107521 | 0.854523803 |
| | 0.9 | 1.180644015 | 0.675454222 | 0.88557476 | 1.089964648 | 0.843475615 |
| | 0.95 | 2.542728751 | 1.118063874 | 1.517882427 | 1.628963912 | 0.87081655 |
| | 0.99 | 11.33237039 | 3.431347422 | 2.390235799 | 1.158790559 | 1.062591647 |

**Table 5: Simulated result when p = 11**

| | | | | MSE | | |
|---|---|---|---|---|---|---|
| ***n*** | ***ρ*** | **MLE** | **LRE** | **LE** | **LTE** | **MRTE** |
| 50 | 0.85 | 15.62248882 | 3.360151504 | 2.136558092 | 1.538140331 | 1.014294647 |
| | 0.9 | 19.69216589 | 4.349115917 | 2.227719841 | 1.37975586 | 1.017730679 |
| | 0.95 | 42.11609908 | 8.57400015 | 2.085414071 | 1.414166621 | 1.130259348 |
| | 0.99 | 220.5709789 | 42.58890958 | 1.5511824 | 1.534085627 | 1.670881377 |
| 75 | 0.85 | 7.045362827 | 2.021799075 | 2.152184908 | 1.542734023 | 0.970693448 |
| | 0.9 | 8.762489659 | 2.428445922 | 2.319711733 | 1.502064006 | 0.991345701 |
| | 0.95 | 19.31454681 | 4.828934178 | 2.407789925 | 1.282225532 | 1.051209773 |
| | 0.99 | 87.0682104 | 20.00550452 | 1.924638544 | 1.431408672 | 1.589297792 |
| 100 | 0.85 | 4.084547307 | 1.465047736 | 1.905235909 | 1.714048791 | 0.942480124 |
| | 0.9 | 6.171541569 | 1.95473513 | 2.251601338 | 1.566334482 | 0.946929536 |
| | 0.95 | 11.94603707 | 3.285987657 | 2.552439629 | 1.377097541 | 1.033680994 |
| | 0.99 | 62.2427181 | 14.71829477 | 2.117384719 | 1.446936182 | 1.548666637 |
| 200 | 0.85 | 1.60109178 | 0.802449099 | 1.144481482 | 1.334543454 | 0.87106372 |
| | 0.9 | 2.644235236 | 1.103197492 | 1.622603454 | 1.72806387 | 0.889121394 |
| | 0.95 | 5.198299028 | 1.797138779 | 2.297212738 | 1.769512859 | 0.939415841 |
| | 0.99 | 23.71040775 | 6.45816965 | 2.690510631 | 1.183064983 | 1.243882729 |
| 300 | 0.85 | 1.038238979 | 0.608769193 | 0.830899524 | 1.00435901 | 0.84304741 |
| | 0.9 | 1.588214963 | 0.793616982 | 1.164444284 | 1.395794248 | 0.856952221 |
| | 0.95 | 3.248648893 | 1.304996018 | 1.880437211 | 1.963565551 | 0.897447549 |
| | 0.99 | 14.91179488 | 4.312952614 | 2.875091392 | 1.19630965 | 1.155107687 |

**Table 6: Simulated result when p = 13**

| | | | | MSE | | |
|---|---|---|---|---|---|---|
| ***n*** | ***ρ*** | **MLE** | **LRE** | **LE** | **LTE** | **MRTE** |
| 50 | 0.85 | 21.58602372 | 4.042722851 | 2.25087649 | 1.479418798 | 0.998016362 |
| | 0.9 | 33.4420806 | 5.975710635 | 2.266709197 | 1.427155261 | 1.036979891 |
| | 0.95 | 59.83143694 | 10.37434984 | 2.074453498 | 1.405233222 | 1.086736965 |
| | 0.99 | 392.0816392 | 62.30674651 | 1.498023085 | 1.524284614 | 1.73387796 |
| 75 | 0.85 | 9.356946584 | 2.439343837 | 2.518697399 | 1.534857312 | 0.977827166 |
| | 0.9 | 12.93841699 | 3.241147108 | 2.657999481 | 1.433311839 | 1.001578618 |
| | 0.95 | 28.43241157 | 6.461438958 | 2.629394766 | 1.326861553 | 1.109542561 |
| | 0.99 | 120.8220344 | 26.12698631 | 1.823265349 | 1.466373938 | 1.538855333 |
| 100 | 0.85 | 5.107245108 | 1.635955341 | 2.227491189 | 1.773500748 | 0.953318229 |
| | 0.9 | 7.985165604 | 2.290127219 | 2.643476788 | 1.663320141 | 0.977309431 |
| | 0.95 | 17.58326063 | 4.461535908 | 2.980437718 | 1.400815021 | 1.090302543 |
| | 0.99 | 79.97751887 | 18.46711277 | 2.120996485 | 1.433606293 | 1.51764211 |
| 200 | 0.85 | 2.103579571 | 0.955445058 | 1.461260901 | 1.656535141 | 0.889099001 |
| | 0.9 | 3.266398796 | 1.259167254 | 1.946184869 | 1.931269616 | 0.909615449 |
| | 0.95 | 5.935080799 | 1.947347818 | 2.629337145 | 2.013660326 | 0.956603372 |
| | 0.99 | 31.3911278 | 8.175046258 | 2.933127861 | 1.17337897 | 1.279323178 |
| 300 | 0.85 | 1.283256035 | 0.718493975 | 1.024497652 | 1.303291782 | 0.855333343 |
| | 0.9 | 1.848098967 | 0.889001328 | 1.35598625 | 1.699199082 | 0.869734751 |
| | 0.95 | 3.997041133 | 1.50974873 | 2.259041828 | 2.245593103 | 0.929670003 |
| | 0.99 | 19.16171984 | 5.289506136 | 3.25249318 | 1.189965926 | 1.207164139 |

## 5. Conclusion

In this article, all the estimators that mitigate the problem of multicollinearity in Logistic regression model are compared through the MSE standard. Finally, a simulation study was conducted to clarify the theoretical results. The result shows that the performance the MRTE estimator outperforms all other estimators in most cases. We recommend this method to choose the desired bias factor it was taken into account in future studies.

## References

**A. E. Hoerl, R. W. Kennard, (1970a)** "Ridge regression: Biased estimation for non-orthogonal Problems", Technometrics, 12, 55–67.

**A. E. Hoerl, R. W. Kennard, (1970b)** "Ridge regression: Application to non-orthogonal problems", Technometrics, 12(1), 69–82.

**A. K. M. Saleh, B. M. G. Kibria, (1993)** "Performance of some new preliminary test ridge regression and their properties", Communications in Statistics - Theory and Methods, 22(10), 2747–2764.

**A. F. Lukman, Adewuyi Emmanuel, Onate A. Clement, Kayode Ayinde, (2020) "** A Modified Ridge-Type Logistic Estimator", Iranian Journal of Science and Technology, Transactions A: Science.

**B. M. G. Kibria, (2003) "**Performance of some new ridge regression estimators", Communications in Statistics - Theory and Methods, 32(2), 419–435.

**B. M. G. Kibria, (2004) "**Performance of the shrinkage preliminary test ridge regressionestimators based on the conflicting of W, LR and LM tests", Journal of Statistical Computation and Simulation, 74(11), 793–810

**Deniz Inan, Birsen E. Erdogan, (2013)** " Liu-Type Logistic Estimator", Communications in Statistics - Simulation and Computation, 42(7), 1578-1586.

**G. C. McDonald, D. I. Galarneau, (1975)** "A Monte Carlo evaluation of some ridge-type estimators", Journal of the American Statistical Association, 70 (350), 407–416.

**G. Khalaf, G. Shukur, (2005) "**Choosing ridge parameters for regression problems", Communications in Statistics - Theory and Methods, 34(5), 1177–1182.

**G. Muniz, B. M. G. Kibria, (2009) "**On some ridge regression estimators: An empirical comparison", Communication in Statistics- Simulation and Computation , 38(3), 621–630.

**Gibbons DG, (1981) "**A simulation study of some ridge estimators**"**,J Am Stat Assoc, 76:131–139

**Gülesen Üstündağ Şiray, Selma Toker , Selahattin Kaçiranlar, (2015) "**On the Restricted Liu Estimator in the Logistic Regression Model**",** Communications in Statistics - Simulation and Computation, 44(1), 217-232.

**Guoping Zeng, Emily Zeng, (2019) "**On the relationship between multicollinearity and separation in logistic regression", Communications in Statistics - Simulation and Computation

**H. Midi, S. K. Sarkar, S. Rana, (2013)** "Collinearity diagnostics of binary logistic regression model", Journal of Interdisciplinary Mathematics, 13 (3), 253–67.

**Habshah Midi, S.K. Sarkar, Sohel Rana, (2010) "**Collinearity diagnostics of binary logistic regression model ", Journal of Interdisciplinary Mathematics, 13(3), 253-267.

**Hoerl AE, Kennard RW (1970) "**Ridge regression: biased estimation for nonorthogonal problems**"**, Technometrics, 12(1):55–67.

**Hoerl AE, Kennard RW, Baldwin KF, (1975) "** Ridge regression: some

Simulation. Commun Stat Theory Methods 4:105–123

**J. F. Lawless, P.Wang, (1976)** "A simulation study of ridge and other regression estimators", Communications in Statistics - Theory and Methods, 5(4), 307–323.

**J. Shen, S. Gao., (2008)** "A solution to separation and multicollinearity in multiple logistic regression", Journal of Data Science, 6 (4), 515–31.

**Kibria, B. M. G., Mansson, K. and Shukur, G. (2012)** "Performance of some logistic ridge regression estimators", Comp. Econ, 40(4), 401-414 .

**Mansson, K. and Shukur, G. (2011) "**On Ridge Parameters in Logistic Regression", Communications in Statistics - Theory and Methods, 40(18), 3366-3381.

**Lee, A. H., & Silvapulle, M. J. (1988) "**Ridge estimation in logistic regression**"**, Communications in Statistics-Simulation and Computation, 17(4), 1231-1257.

**Lukman AF, Ayinde K, Sek SK, Adewuyi E (2019b) "**A modified new two

Parameter estimator in a linear regression model. Model Simul Eng.doi.org/10.1155/2019/6342702

**M. A. Alkhamisi, G. Khalaf, G. Shukur, (2006) "**Some modifications for choosing ridge parameter", Communications in Statistics - Theory and Methods, 35(11), 2005-2020.

**M. S. Haq, B. M. G. Kibria, (1996) "**A shrinkage estimator for the restricted linear regression model: Ridge regression approach", Journal of Applied Statistical Science, 3, 301–316.

**Mansson K, Kibria B, Shukur G (2012) "**On Liu estimators for the logit regression model**"**, Econ Model ,29(4):1483–1488.

**Mansson, K. and Shukur, G (2011) "** On ridge parameters in logistic regression**"**, Comm. Statist. Theo. Meth, 40(18), 3366-3381

**McCullagh, P. (1980) "** Regression models for ordinal data**"**, Journal of the Royal Statistical Society: Series B. (Methodological)), 42(2), 109–142

**Mohamed R. Abonazel, (2018) "** Liu-Type Multinomial Logistic Estimator", the Indian Journal of Statistics, 81-B (2), 203-225.

**N. A. M. R. Senaviratna, T. M. J. A. Cooray, (2019) "**Diagnosing Multicollinearity of Logistic Regression Model", Asian Journal of Probability and Statistics, 5(2), 1-9.

**Newhouse JP, Oman SD (1971) "**An evaluation of ridge estimators**"**, Rand Corporation, P-716-PR, 1–28

**Ozkale MR, Kaciranlar S (2007) "**The restricted and unrestricted two-parameter estimators. Commun Stat Theory Methods,36:2707–2725

**Pena, W. E. L., De Massaguer, P. R., Zuniga, A. D. G., & Saraiva, S. H. (2011) "** Modeling the growth limit of Alicyclobacillus acidoterrestris CRA7152 in apple juice: effect of pH, Brix, temperature and nisin concentration**"**, Journal of Food Processing and Preservation, 35(4), 509-517.

**Schaeffer RL, Roi LD, Wolfe RA (1984) "**A ridge logistic estimator**"**, Commun Stat Theory Methods, 13:99–113.

**Senaviratna N. and Cooray T., (2019) "**Diagnosing Multicollinearity of Logistic Regression Model", Asian Journal of Probability and Statistics, 5(2), 1-9.

**Siray, G.U., Toker, S., Ka¸cıranlar, S. (2015) "**On the Restricted Liu Estimator in the Logistic Regression Model**"**, Comm. Statist. Sim Comp, 44:217-232.

**Varathan Nagarajah and Wijekoon Pushpakanthie, (2019) "**Modified almost unbiased Liu estimator in logistic regression", Communications in Statistics-SimulationandComputation,1-17**.**doi. /10.1080/03610918.2019.1626888

**Yasin Asar, (2015)** " Some New Methods to Solve Multicollinearity in Logistic Regression", Communications in Statistics - Simulation and Computation, 46 (4), 2576-2586.

**Yasin Asar, Aşır Genç, (2017) "**Two-parameter ridge estimator in the binary logistic regression", Communications in Statistics - Simulation and Computation, 46(9), 7088-7099.

**Yasin Asar, Jibo Wu, (2019) "**An improved and efficient biased estimation technique in logistic regression model", Communications in Statistics - Theory and Methods, 49(9), 2237-2252.

**Yasin Asar, Murat Erişoğlu, Mohammad Arashi, (2016) "** Developing a Restricted Two-Parameter Liu-Type Estimator: A Comparison of Restricted Estimators in the Binary Logistic Regression Model", Communications in Statistics - Theory and Methods, 46(16), 6864-6873.

**Yasin Asar, Aşir Genç, (2015) "**New Shrinkage Parameters for the Liu-type Logistic Estimator", Communications in Statistics - Simulation and Computation, 45(3), 1094-1103.

**Yasin Asar, Mohammad Arashi, Jibo Wu, (2017)"** Restricted ridge estimator in the logistic regression model", Communications in Statistics - Simulation and Computation, 46(8), 6538-6544.

**Zeng Guoping and Zeng Emily, (2019)"** On the relationship between multicollinearity and separation in logistic regression", Communications in Statistics–Simulation and Computation, 1-9. doi./10.1080/03610918.2019.158951

**الملخص:**

يتم استخدام نموذج الانحدار اللوجستي الثنائي (BLR) كبديل لنموذج الانحدار الخطي الشائع الاستخدام عندما يكون متغير الاستجابة ثنائيًا. كما هو الحال في نموذج الانحدار الخطي ، يمكن أن تكون هناك علاقة بين متغيرات التوقع في BLR ، خاصة عندما تكون مستمرة ، مما يؤدي إلى ظهور مشكلة الخطية المتعددة. كفاءة مقدر الاحتمالية القصوى (MLE) منخفضة في تقدير معلمات BLR عندما يكون هناك علاقة خطية متعددة بدلاً من ذلك ، تم تطوير مقدر التلال ومقدر Liu ومقدر نوع Liu ومقدر نوع ريدج المعدل ليحل محل MLE. ومع ذلك ، في هذه الدراسة ، قمنا بمقارنة جميع المقدرات بواسطة MSE للحصول على أفضل مقدر يخفف من تأثير العلاقة الخطية المتعددة. أخيرًا ، تم إجراء دراسة محاكاة لتوضيح النتائج النظرية. تظهر النتيجة أن مقدر نوع ريدج المعدل يتفوق في الأداء على مقدرات أخرى.