New Two-Parameter Estimators for the Logistic Regression Model with Multicollinearity

FUAD A. AWWAD[1], KEHINDE A. ODENIYI[2], ISSAM DAWOUD[3], ZAKARIYA YAHYA ALGAMAL[4], MOHAMED R. ABONAZEL[5], B. M. GOLAM KIBRIA[6], ELSAYED TAG ELDIN[7]

[1]Department of Quantitative Analysis, College of Business Administration, King Saud University, Riyadh, SAUDI ARABIA

[2]Department of Agricultural Economics and Agribusiness Management, Osun State University, Osogbo, NIGERIA

[3]Department of Mathematics, Al-Aqsa University, Gaza, PALESTINE

[4]Department of Statistics and Informatics, University of Mosul, Mosul, IRAQ

[5]Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, EGYPT

[6]Department of Mathematics and Statistics, Florida International University, Miami, Florida, USA

[7]Electrical Engineering Department, Faculty of Engineering & Technology, Future University in Egypt, New Cairo, EGYPT

Abstract: - We proposed new two-parameter estimators to solve the problem called multicollinearity for the logistic regression model in this paper. We have derived these estimators' properties and using the mean squared error (MSE) criterion; we compare theoretically with some of existing estimators, namely the maximum likelihood, ridge, Liu estimator, Kibria-Lukman, and Huang estimators. Furthermore, we obtain the estimators for *k* and *d*. A simulation is conducted in order to compare the estimators' performances. For illustration purposes, two real-life applications have been analyzed, that supported both theoretical and a simulation. We found that the proposed estimator, which combines the Liu estimator and the Kibria-Lukman estimator, has the best performance.

Key-Words: - Logistic regression model; Multicollinearity; Ridge regression, LLKL estimator; Simulation Study; Real-life applications.

Received: August 13, 2021. Revised: May 5, 2022. Accepted: May 23, 2022. Published: June 16, 2022.

1 Introduction

The regression model called binary logistic is considered to obtain a model for getting the relationship between variable with a binary response and one or group of regressor variables. The usage of this model are in many areas, as biostatistics, finance, and medical sciences, among others. The maximum likelihood estimator (MLE) is considered for estimating the logistic model coefficients. In practice, we are assuming that the regressor variables are orthogonal. However, in real-life situations, the regressor variables are often in correlation, and this causes a multicollinearity. So, in this case, the MLE has unduly large variance and hence, it becomes inefficient. Therefore, Hoerl and Kennard [1] proposed a different estimation method which is ridge regression (RR) in the linear model. Many authors have studied and made some improvements in RR for the linear model, to mention a few, [1, 2, 3, 4, 5, 6, 7, 8] among others.

Then, Schaeffer et al. [9] have extended the RR to logistic model for solving the multicollinearity in this model. In addition, Kibria et al. [10] have verified some biasing parameters estimators' performance in RR for the logistic model. Also, there are different studies of the few biased estimators in the logistic model as: Inan and Erdogan [11], Nagarajah and Wijekoon [12], Asar et al. [13], Asar and Genc [14], and Varathan and Wijekoon [15]. Recently, Lukman et al. [16] have developed the modified version of ridge-type for the logistic model. Also, Abonazel and Farghali [17] have developed a new estimator with two-parameter for the multinomial logistic model. Then, Farghali et al. [18] have proposed two generalized estimators with two-parameter for the multinomial logistic model. As well as, Yang and Chang [19] have

proposed a new estimator of two-parameter based on the Liu [20] estimator and RR estimator.
This paper focuses on extending the estimator proposed by [19] and proposing a new one for the binary logistic model.
Then, the paper is like that: In Section 2, we present the model and the proposed estimators. We made the theoretical comparison among the estimators in Section 3. The results of simulation are given in Section 4 and that of real-life are illustrated in Section 5. In Section 6, the conclusions are stated.

2 Statistical Methodology

The logistic regression model, where the distribution of the response (y) is Bernoulli: $y_i \sim Ber(\omega_i)$ such that

$$\omega_i = \frac{e^{x_i\beta}}{1+e^{x_i\beta}}, \tag{1}$$

where x_i is given as the ith row of X matrix with the dimension of $n \times p$ and β is unknown coefficients vector with the dimension of $p \times 1$. The transformation of logit is

$$f(x_i) = \ln\left(\frac{\omega_i}{1-\omega_i}\right) = x_i\beta \tag{2}$$

The MLE is used widely in parameter estimation for this model. The function of log likelihood is

$$L = \sum_{i=1}^{n} y_i \log(\omega_i) + \sum_{i=1}^{n}(1-y_i)\log(1-\omega_i) \tag{3}$$

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^{n}(y_i - \omega_i)\, x_i = 0\,. \tag{4}$$

With the iteratively reweighted least squares (IRLS), equation (4) is solved. Since it is nonlinear in parameter. So, the MLE for the logistic model is defined as

$$\hat{\beta}_{MLE} = S^{-1}X'\hat{G}\,\hat{z} \tag{5}$$

where $S = X'\hat{G}\,X$, $\hat{G} = diag\,(\hat{\omega}_i(1-\hat{\omega}_i))$ and $\hat{z}_i = \log(\hat{\omega}_i) + \frac{y_i - \hat{\omega}_i}{\hat{\omega}_i\,(1-\hat{\omega}_i)}$.

In the presence of multicollinearity, Schaeffer et al. [9] introduced ridge regression for logistic (LRR) as a different method. The LRR is given by:

$$\hat{\beta}_{LRR} = M\,\hat{\beta}_{MLE}, \qquad k > 0 \tag{6}$$

where $M = (\mathrm{I}_p + k\,S^{-1})^{-1}$ is weight matrix and k is parameter of ridge or biasing.
Then, Mansson et al. [21] suggested Liu estimator for logistic (LLE) as:

$$\hat{\beta}_{LLE} = F_d\,\hat{\beta}_{MLE}, \qquad 0 < d < 1 \tag{7}$$

where $F_d = (S + \mathrm{I}_p)^{-1}(S + d\mathrm{I}_p)$ is the weight matrix and d is the biasing parameter.
Huang [22] proposed the logistic two parameter estimator (LTPE) and is given by

$$\hat{\beta}_{LTPE} = R_{kd}\,\hat{\beta}_{MLE},\; k > 0,\; 0 < d < 1 \tag{8}$$

where $R_{kd} = (S + k\,\mathrm{I}_p)^{-1}(S + k\,d\,\mathrm{I}_p)$.

Following [23], the Kibria-Lukman estimator for the logistic (LKL) is defined as

$$\hat{\beta}_{LKL} = MW\,\hat{\beta}_{MLE}, \qquad k > 0 \tag{9}$$

where $W = (\mathrm{I}_p - k\,S^{-1})$.
Kibria and Lukman [23] proved that Kibria-Lukman is more efficient than ridge estimator. Recently, the Kibria-Lukman estimator is extended in gamma and beta regression models by [24, 25], respectively.

2.1 New Two-Parameter Estimators

Yang and Chang [19] have developed estimator of a two parameter as

$$\hat{\beta}_{YC} = (X'X + I)^{-1}(X'X + d)(1 + k(X'X)^{-1})^{-1}\,\hat{\beta}, \qquad k > 0,\; 0 < d < 1 \tag{10}$$

Yang and Chang [19] proved that their estimator is more efficient than Liu estimator as well as ridge estimator, meaning that combining them with Liu and ridge gives an effective estimator. Therefore, following [19], we are going to propose a new modified estimator of two-parameter based on Liu estimator and the Kibria-Lukman [23] estimator, as

$$\hat{\beta}_{LKL} = (X'X + I)^{-1}(X'X + d)(1 + k(X'X)^{-1})^{-1}(1 - k(X'X)^{-1})\hat{\beta}, \quad k > 0,\ 0 < d < 1. \tag{11}$$

Therefore, we develop the logistic version of the estimator of Yang and Chang [19] and the new estimator of two-parameter and defined as follows:

- The logistic version of the Yang and Chang [19] (LYC) estimator is defined as

$$\hat{\beta}_{LYC} = F_d M \hat{\beta}_{MLE}, \qquad k > 0, 0 < d < 1. \tag{12}$$

- The logistic version of the new estimator (LLKL) estimator is defined as

$$\hat{\beta}_{LLKL} = F_d M W \hat{\beta}_{MLE}, \qquad k > 0, 0 < d < 1. \tag{13}$$

2.2 MSEM and MSE Properties of the Estimators

The matrix form of mean squared error (MSEM) and the mean squared error (MSE) for an estimator $(\tilde{\beta})$ are defined respectively as follows:

$$MSEM(\tilde{\beta}) = Cov(\tilde{\beta}) + \left(Bias(\tilde{\beta})\right)\left(Bias(\tilde{\beta})\right)' \tag{14}$$

and

$$MSE(\tilde{\beta}) = trace\left(MSEM(\tilde{\beta})\right). \tag{15}$$

By matrix spectral decomposition, $S = LHL'$ where L matrix columns and H are the eigenvectors and eigenvalues of S. The estimators MSEMs are respectively as follows:

$$MSEM(\hat{\alpha}_{MLE}) = LH^{-1}L' \tag{16}$$

and

$$MSEM(\hat{\alpha}_{LRR}) = LMH^{-1}M'L' + (M - I_p)\alpha\alpha'(M - I_p)' \tag{17}$$

where $M = (\mathrm{I}_p + kH^{-1})^{-1}$ and $\alpha = L'\beta$.

$$MSEM(\hat{\alpha}_{LLE}) = LF_d H^{-1}F_d'L' + (1-d)^2(H + I_p)^{-1}\alpha\alpha'(H + I_p)^{-1}. \tag{18}$$

where $F_d = (H + \mathrm{I}_p)^{-1}(H + d\mathrm{I}_p)$

$$MSEM(\hat{\alpha}_{LTPE}) = LR_{kd}H^{-1}R_{kd}'L' + [R_{kd} - I_p]\alpha\alpha'[R_{kd} - I_p]'. \tag{19}$$

where $R_{kd} = [H + k\mathrm{I}_p]^{-1}[H + kd\,\mathrm{I}_p]$.

$$MSEM(\hat{\alpha}_{LKL}) = LMWH^{-1}W'M'L' + [MW - I_p]\alpha\alpha'[MW - I_p]'. \tag{20}$$

where $W = [I_p - kH^{-1}]$.

$$MSEM(\hat{\alpha}_{LYC}) = LF_d MH^{-1}M'F_d'L' + [F_d M - I_p]\alpha\alpha'[F_d M - I_p]'. \tag{21}$$

and finally.

$$MSEM(\hat{\alpha}_{LLKL}) = LF_d MWH^{-1}W'M'F_d'L' + [F_d MW - I_p]\alpha\alpha'[F_d MW - I_p]' \tag{22}$$

Lemma 2.1: [26], let V is a positive definite matrix with $n \times n$ dimension, i.e. $V > 0$, and α is a vector; then, $V - \alpha\alpha' > 0$ iff $\alpha'V^{-1}\alpha < 1$,

Lemma 2.2: [27], suppose that $\alpha_i = Q_i\, y$, $i = 1, 2$ are any α two linear estimators. Assume $D = C(\hat{\alpha}_1) - C(\hat{\alpha}_2) > 0$ where $C(\hat{\alpha}_i)$, $i = 1, 2$ is $\hat{\alpha}_i$ covariance matrix and $m_i = bias(\hat{\alpha}_i) = (Q_i X - I)\,\alpha$, $i = 1, 2$. Consequently,

$$MSEM(\hat{\alpha}_1) - MSEM(\hat{\alpha}_2) = \sigma^2 D + m_1 m_1' - m_2 m_2' > 0 \qquad (23)$$

iff $m_2'(\sigma^2 D + m_1' m_1) m_2 < 1$ where $MSEM(\hat{\alpha}_i) = C(\hat{\alpha}_i) + m_i m_i'$.

3 Comparison of Estimators

3.1. Comparison between $\hat{\alpha}_{MLE}$ and $\hat{\alpha}_{LLKL}$.

Theorem 3.1

$MSEM(\hat{\alpha}_{MLE}) - MSEM(\hat{\alpha}_{LLKL}) > 0$ if and only if

$$\alpha' [F_d\, MW - I_p]' \, [L(H^{-1} - F_d\, MW\, H^{-1} W' M' F_d')\, L'] [F_d\, MW - I_p]\, \alpha < 1 \qquad (24)$$

***Proof*:**

$$D(\hat{\alpha}_{MLE}) - D(\hat{\alpha}_{LLU}) = L\,(H^{-1} - F_d\, MW\, H^{-1} W' M' F_d')\, L'$$

$$= L\, diag \left\{ \frac{1}{h_i} - \frac{(h_i + d)^2\, (h_i - k)^2}{h_i\, (h_i + 1)^2 (h_i + k)^2} \right\}_{i=1}^{p} L' \qquad (25)$$

where h_i is the ith eigenvalue of the matrix H and $H^{-1} - F_d\, W\, M\, H^{-1}\, M' W' F_d'$ will be positive definite (pd) if and only if $(h_i + 1)^2 (h_i + k)^2 - (h_i + d)^2\ (h_i - k)^2 > 0$ or $(h_i + 1)(h_i + k) - (h_i + d)(h_i - k) > 0$. We observed that, for $k > 0$ and $0 < d < 1$, $(h_i + 1)(h_i + k) - (h_i + d)(h_i - k) = h_i\,(2k + 1 - d) + k\,(1 + d) > 0$. By Lemma 2.2, the proof is completed.

3.2. Comparison between $\hat{\alpha}_{LRR}$ and $\hat{\alpha}_{LLKL}$.

Theorem 3.2

$MSEM(\hat{\alpha}_{LRR}) - MSEM(\hat{\alpha}_{LLKL}) > 0$ if and only if

$$\alpha' [F_d\, MW - I_p]' \, [D_1 + [M - I_p]\, \alpha \alpha' [M - I_p]'] [F_d\, MW - I_p]\, \alpha < 1 \qquad (26)$$

where

$D_1 = L(M\, H^{-1} M' - F_d\, MW\, H^{-1} W' M' F_d')\, L'$

***Proof*:**

$$D_1 = L\,(M\, H^{-1} M' - F_d\, MW\, H^{-1}\, W' M' F_d')\, L'$$

$$= L\, diag \left\{ \frac{h_i}{(h_i + k)^2} - \frac{(h_i + d)^2\, (h_i - k)^2}{h_i\, (h_i + 1)^2 (h_i + k)^2} \right\}_{i=1}^{p} L' \qquad (27)$$

where $M\, H^{-1} M' - F_d\, MW\, H^{-1} W' M' F_d'$ will be pd if and only if $(h_i + 1)^2 h_i^2 - (h_i + d)^2\ (h_i - k)^2 > 0$ or $(h_i + 1)\, h_i - (h_i + d)(h_i - k) > 0$. We observed that, for $k > 0$ and $0 < d < 1$, $(h_i + 1)\, h_i - (h_i + d)(h_i - k) = h_i\,(k + 1 - d) + kd > 0$. By Lemma 2.2, the proof is completed.

3.3. Comparison between $\hat{\alpha}_{LLE}$ and $\hat{\alpha}_{LLKL}$.

Theorem 3.3

$MSEM(\hat{\alpha}_{LLE}) - MSEM(\hat{\alpha}_{LLKL}) > 0$ if and only if

$$\alpha' [F_d\, MW - I_p]' \, [D_2 + [F_d - I_p]\, \alpha \alpha' [F_d - I_p]'] [F_d\, MW - I_p]\, \alpha < 1 \qquad (28)$$

where $D_2 = L(F_d\, H^{-1} F_d' - F_d\, MW\, H^{-1} W'\, M' F_d')\, L'$

***Proof*:**

$$\begin{aligned} D_2 &= L\,(F_d\ H^{-1}F_d' - F_d\ MW\ H^{-1}\ W'M'\ F_d'\,)\, L' \\ &= L\, diag\left\{ \frac{(h_i+d)^2}{h_i\,(h_i+1)^2} - \frac{(h_i+d)^2\,(h_i-k)^2}{h_i\,(h_i+1)^2 (h_i+k)^2} \right\}_{i=1}^{p} L' \end{aligned} \tag{29}$$

where $F_d\, H^{-1} F_d' - F_d\, MW\, H^{-1}\, W'M'\, F'$ will be pd iff $(h_i+k)^2 - (h_i-k)^2 > 0$ or $(h_i+k) - (h_i-k) > 0$. We had that, for $k>0$, $(h_i+k)-(h_i-k) = 2k > 0$. By Lemma 2.2, the proof is completed.

3.4. Comparison between $\hat{\alpha}_{LKL}$ and $\hat{\alpha}_{LLKL}$.

Theorem 3.4

$MSEM\ (\hat{\alpha}_{LKL}) - MSEM\ (\hat{\alpha}_{LLKL}) > 0$ if and only if

$$\alpha'[F_d\, MW - I_p]'\, [D_3 + [MW - I_p]\,\alpha\,\alpha'[MW - I_p]']\,[F_d\, MW - I_p]\,\alpha < 1 \tag{30}$$

where $D_3 = L(MW\, H^{-1} W'\, M' - F_d\, MW\, H^{-1}\, W'\, M' F_d')\, L'$

***Proof*:**

$$\begin{aligned} D_3 &= L\,(MW\ H^{-1}\ W'M' - F_d\ MW\ H^{-1}\ W'\ M'F_d'\,)\, L' \\ &= L\, diag\left\{ \frac{(h_i-k)^2}{h_i\,(h_i+k)^2} - \frac{(h_i+d)^2\,(h_i-k)^2}{h_i\,(h_i+1)^2 (h_i+k)^2} \right\}_{i=1}^{p} L' \end{aligned} \tag{31}$$

where $MW\, H^{-1} W'M' - F_d\, MW\, H^{-1}\, W'M'\, F_d'$ will be pd iff $(h_i+1)^2 - (h_i+d)^2 > 0$ or $(h_i+1)-(h_i+d) > 0$. We had that, for $k>0$, and $0<d<1$, $(h_i+1)-(h_i+d) = 1-d > 0$. By Lemma 2.2, the proof is completed.

3.5. Comparison between $\hat{\alpha}_{LTPE}$ and $\hat{\alpha}_{LLKL}$.

Theorem 3.5

$MSEM\ (\hat{\alpha}_{LTPE}) - MSEM\ (\hat{\alpha}_{LLKL}) > 0$ if and only if

$$\alpha'[F_d\, MW - I_p]'\, [D_4 + [R_{kd} - I_p]\,\alpha\,\alpha'[R_{kd} - I_p]']\,[F_d\, MW - I_p]\,\alpha < 1 \tag{32}$$

where $D_4 = L(R_{kd}\, H^{-1} R_{kd}' - F_d\, MW\, H^{-1}\, W'M'\, F_d')\, L'$

***Proof*:**

$$\begin{aligned} D_4 &= L\,(R_{kd}\ H^{-1}R_{kd}' - F_d\ MW\ H^{-1}\ W'M'\ F_d'\,)\, L' \\ &= L\, diag\left\{ \frac{(h_i+kd)^2}{h_i\,(h_i+k)^2} - \frac{(h_i+d)^2\,(h_i-k)^2}{h_i\,(h_i+1)^2 (h_i+k)^2} \right\}_{i=1}^{p} L' \end{aligned} \tag{33}$$

where $R_{kd}\, H^{-1} R_{kd}' - F_d\, MW\, H^{-1}\, W'M'\, F_d'$ will be pd iff $(h_i+1)^2\,(h_i+kd)^2 - (h_i+d)^2 (h_i-k)^2 > 0$ or $(h_i+1)(h_i+kd) - (h_i+d)(h_i-k) > 0$. We had that, for $k > 0$, and $0 < d < 1$, $(h_i + 1)(h_i + kd) - (h_i + d)(h_i - k) =$ $h_i(kd + k + 1 - d) + 2kd > 0$. By Lemma 2.2, the proof is completed.

3.6. Comparison between $\hat{\alpha}_{LYC}$ and $\hat{\alpha}_{LLKL}$.

Theorem 3.6

$MSEM\,(\hat{\alpha}_{LYC}) - MSEM\,(\hat{\alpha}_{LLKL}) > 0$ if and only if

$$\alpha'[F_d MW - I_p]'[D_5 + [F_d M - I_p]\alpha\alpha'[F_d M - I_p]'][F_d MW - I_p]\alpha < 1 \tag{34}$$

where $D_5 = L(F_d MH^{-1}M'F_d' - F_d MWH^{-1}W'M'F_d')L'$

Proof:

$$\begin{aligned} D_5 &= L(F_d M H^{-1}M'F_d' - F_d MW H^{-1}W'M'F_d')L' \\ &= L\,diag\left\{ \frac{h_i(h_i + d)^2}{(h_i + 1)^2(h_i + k)^2} - \frac{(h_i + d)^2(h_i - k)^2}{h_i(h_i + 1)^2(h_i + k)^2} \right\}_{i=1}^{p} L' \end{aligned} \tag{35}$$

where $F_d M H^{-1}M'F_d' - F_d MW H^{-1}W'M'F_d'$ will be pd if and only if $h_i^2 - (h_i - k)^2 > 0$ or $h_i - (h_i - k) > 0$. We had that, for $k > 0$, $\lambda_i - (\lambda_i - k) = k > 0$. By Lemma 2.2, the proof is completed.

3.7 Determination of the Parameters k and d for the estimators

In this section, we suggest the biasing parameters for the existing and the proposed estimators.

- The estimator of the biasing parameter k of LRR, LYC, LKL, and LTPE estimators is obtained as follows [4]:

$$\hat{k} = \frac{p}{\sum_{i=1}^{p} \alpha_i^2}. \tag{36}$$

- The estimator of biasing parameter d for the Liu estimator and likewise the second biasing parameter for LYC and LTPE is obtained as follows [20]:

$$\hat{d} = \max\left[0, \min\left(\frac{\alpha_i^2 - 1}{(1/h_i) + \alpha_i^2}\right)\right]. \tag{37}$$

- For the proposed LLKL estimator, the optimal value of k can be obtained by choosing k that minimize

$$MSEM(\hat{\alpha}_{LLKL}) = E((\hat{\alpha}_{LLKL} - \alpha)'(\hat{\alpha}_{LLKL} - \alpha)),$$
$$m(k,d) = tr(MSEM(\hat{\alpha}_{LLKL})),$$

$$m(k,d) = \sum_{i=1}^{p} \frac{(h_i + d)^2(h_i - k)^2}{h_i(h_i + 1)^2(h_i + k)^2} + \sum_{i=1}^{p} \frac{(h_i(2k - d + 1) + k(d + 1))^2 \alpha_i^2}{(h_i + 1)^2(h_i + k)^2} \tag{38}$$

Differentiate equation (38) with respect to k, then

$$\hat{k} = \frac{h_i(h_i + d) + h_i^2(d - 1)\alpha_i^2}{(h_i + d) + h_i\alpha_i^2(2h_i + d + 1)}. \tag{39}$$

Differentiate equation (38) with respect to d gives

$$\hat{d} = \frac{h_i(h_i(2k + 1) + k)\alpha_i^2 - h_i(h_i - k)}{(h_i - k) + h_i(k - h_i)\alpha_i^2}. \tag{40}$$

For the purpose of simplifying the selection of the biasing parameters k and d we carry out the following: From equation (40),

$$\hat{d} = \frac{h_i(h_i(2k + 1) + k)\alpha_i^2 - h_i(h_i - k)}{(h_i - k) + h_i(k - h_i)\alpha_i^2} > 0.$$

It implies that
$h_i\left(h_i(2k+1)+k\right)\alpha_i^2-h_i\left(h_i-k\right)>0;$ parameter k is obtained as follows:

$$k>\frac{h_i\left(1-\alpha_i^2\right)}{2h_i\alpha_i^2+\alpha_i^2+1}. \quad (41)$$

In this study, we take the absolute value of the minimum value of equation (41), the absolute value is taken to ensure the estimate return a positive value:

$$k_1=\left|\min\left(\frac{h_i\left(1-\alpha_i^2\right)}{2h_i\alpha_i^2+\alpha_i^2+1}\right)\right| \quad (42)$$

The steps for the practical selection of the biasing parameters are as follows:

1. Obtain $\hat{k}_1$ by replacing σ^2 and α_i^2 with their unbiased estimates.
2. Substitute $\hat{k}_1$ into equation (40), and obtain the absolute value of the minimum value of equation (40):

$$\hat{d}_1=\left|\min\left(\frac{h_i\left(h_i\left(2\hat{k}_1+1\right)+\hat{k}_1\right)\hat{\alpha}_i^2-h_i\left(h_i-\hat{k}_1\right)}{\left(h_i-\hat{k}_1\right)+h_i\left(\hat{k}_1-h_i\right)\alpha_i^2}\right)\right|. \quad (43)$$

Then, we suggest the following basing parameters k and d for the proposed LLKL estimator as follows:

a. LLKL1: $\hat{k}$, $\hat{d}$

b. LLKL2: $\hat{k}_2$, $\hat{d}_1$

where

$$\hat{k}_2=\max\left(\frac{h_i(1-\hat{\alpha}_i^2)}{(2h_i\hat{\alpha}_i^2+\hat{\alpha}_i^2+1)}\right) \quad (44)$$

c. LLKL3: $\hat{k}_3$, $\hat{d}$

where

$$\hat{k}_3=\min\left(\frac{1}{(2\hat{\alpha}_i^2+(1/h_i))}\right) \quad (45)$$

d. LLKL4: $\hat{k}_4=\sqrt{\hat{k}}$, $\hat{d}_2=\sqrt{\hat{d}_1}$.

4 Monte Carlo Simulation

A simulation study has been conducted to compare the performance of the estimators under the condition of multicollinearity. Literature on the linear regression model includes [4, 28, 29, 30, 31]. A few available studies on the logistic regression model includes [14, 15, 16, 17, 18, 21, 32], among others. The regression coefficient is constrained such that $\beta'\beta=1$ [25, 31, 33, 34, 35, 3]. The explanatory variables can be obtained using the following simulation procedure [28, 37]:

$$x_{ij}=\left(1-\rho^2\right)^{1/2}w_{ij}+\rho w_{ip},$$
$$i=1,2,...,n;\ j=1,2,...p, \quad (46)$$

where ρ is considered by many authors as the correlation of regressor variables. The ρ values are 0.80, 0.90, 0.99, and 0.999. While the response is generated with the distribution of Bernoulli $Be(\pi_i)$ where $\pi_i=\frac{e^{X\beta}}{1+e^{X\beta}}$. The sample size n is taken to be 50, 100 and 200. The estimated MSE is

$$MSE\left(\hat{\beta}\right)=\frac{1}{1000}\sum_{l=1}^{1000}\left(\hat{\beta}_{il}-\beta_i\right)'\left(\hat{\beta}_{il}-\beta_i\right) \quad (47)$$

where $\hat{\beta}_{il}$ is estimate of i^th^ parameter in l^th^ replication and β_i ($i=1,2,\ ...,\ p$) is the true parameter values (p is taken to be 3 and 8). The experiment is repeated 1000 times. The simulation results are presented in Tables 1 and 2. The results showed that increasing the sample size results in a decrease in estimated MSE values of estimators. However, the MSE values of estimators increase as correlation values and regressor variables number are increased. Furthermore, from Tables 1 and 2, it appears that the two proposed estimators (LYC and LLKL) are generally preferred to other estimators. The MLE performs least when there is multicollinearity in the data. Among the single parameter estimators, the LKL estimator performs better the LRR and the LLE estimator, especially when ρ=0.8-0.99. The considered two-parameter estimators in this study are the LYC, LTPE and LLKL. The LLKL performance is best followed by LYC estimator, and the LYC performs better the LRR, LLE, LTPE, and LKL estimators, especially when ρ > 0.9. Generally, the most preferred estimator is LLKL. Although, the estimator performance is a function of the biasing parameter. The LLKL estimator works well with the biasing parameter $\hat{k}_4$.

Table 1. Estimated MSE for different estimator when $p = 3$

n	ρ	MLE	LRR	LLE	LTPE	LKL	LYC	LLKL1	LLKL2	LLKL3	LLKL4
50	**0.8**	0.7794	0.4900	0.4683	0.4920	0.4855	0.3951	0.4891	0.5654	0.3999	0.3828
	0.9	1.3581	0.6563	0.6128	0.6807	0.4953	0.4276	0.4679	0.5656	0.4583	0.3902
	0.99	11.2305	3.2760	1.3486	4.0765	1.2697	0.6886	0.4605	1.5829	0.8124	0.4431
	0.999	110.0943	29.6252	11.9849	43.7412	9.7408	4.1582	0.9935	19.5853	6.0294	0.5825
100	**0.8**	0.3731	0.2654	0.2798	0.2655	0.2699	0.2362	0.2757	0.4092	0.2469	0.2233
	0.9	0.7071	0.3975	0.4210	0.3988	0.3317	0.2994	0.3182	0.4096	0.3333	0.2663
	0.99	6.4264	2.0980	0.8664	2.3859	0.9703	0.4884	0.3981	0.8770	0.5629	0.3912
	0.999	67.4281	20.9737	7.8557	28.6695	8.3799	3.5395	0.9135	11.4231	4.6993	0.3589
200	**0.8**	0.1851	0.1705	0.1703	0.1705	0.1873	0.1675	0.1921	0.3354	0.1668	0.1626
	0.9	0.3172	0.2397	0.2588	0.2397	0.2371	0.2195	0.2377	0.3135	0.2325	0.2058
	0.99	2.9718	1.0669	0.6862	1.1323	0.5465	0.3860	0.3379	0.5623	0.4521	0.3163
	0.999	29.7578	8.6876	3.0615	11.4073	3.3366	1.4092	0.6317	0.9534	1.8151	0.3547

Table 2. Estimated MSE for different estimator when $p = 8$

n	ρ	MLE	LRR	LLE	LTPE	LKL	LYC	LLKL1	LLKL2	LLKL3	LLKL4
50	**0.8**	3.8735	1.8766	2.2775	1.8766	1.5930	1.5529	1.5460	1.9549	1.7948	1.4297
	0.9	5.9781	2.2989	2.4691	2.2989	1.6288	1.5643	1.4795	1.9162	1.8782	1.4026
	0.99	43.5254	10.1309	2.2022	10.4985	3.6542	1.4956	1.3583	2.0035	1.7285	1.4557
	0.999	429.4912	91.8397	6.1904	102.1948	24.8774	2.4958	1.4318	17.7479	3.6082	1.5450
100	**0.8**	2.9662	1.5984	2.1144	1.5984	1.1738	1.3505	1.1485	1.4098	1.6780	1.1572
	0.9	4.7487	2.0107	2.4012	2.0107	1.1989	1.3654	1.1025	1.4359	1.7900	1.2564
	0.99	35.6395	9.5175	2.9062	9.6573	2.8575	1.3884	1.3109	1.4746	1.8667	1.3089
	0.999	343.8151	84.8848	3.6548	91.0053	19.8508	1.8033	1.4761	7.6382	2.5354	2.0172
200	**0.8**	1.8607	1.3546	1.6996	1.3546	1.1463	1.2964	1.1384	1.2466	1.4909	1.2621
	0.9	2.4909	1.4925	2.0163	1.4925	1.2133	1.3369	1.2904	1.2715	1.6554	1.2959
	0.99	13.1361	4.0705	2.4493	4.0707	1.5544	1.3382	1.3319	1.3202	1.7924	1.3148
	0.999	121.4520	30.9715	4.7055	31.9513	6.9551	1.4571	1.9756	1.8326	1.8762	1.4766

5 Applications

5.1. Pena Data

The dataset was originally adopted by [38, 39]. Pena et al. [38] employed logistic model to examine the regressors of temperature effect, pH, as well soluble solids concentration with the nisin concentration on the response of Alicyclobacillus growth probability for apple Juice. The eigenvalues of the $X'\hat{G}X$ matrix are 12373.8, 1313.949, 46.54678, 3.4102, and 0.0475. Consequently, the condition number evaluates as 260293.8 which revealed presence of multicollinearity in the model. The estimated values of regression coefficient from each of the estimators and their corresponding mean squared error are available in Table 3.

From Table 3, we note that the estimated coefficients of some variables differ from one estimator to another, so we can use the MSE as a good criterion for judging the efficiency of the estimation. The MLE is not giving any good performance as known. The efficiency of bias estimators depends on the selected values of *k* and *d*. The estimator has the least estimated MSE is LLKL4. This result as same as simulation result.

Through this application, we verify the theoretical conditions of theorems 3.1 to 3.6 as follows:

Table 3. Regression coefficients with MSE for Pena data

Coef.	MLE	LRR	LLE	LTPE	LKL	LYC	LLKL1	LLKL2	LLKL3	LLKL4
$\hat{\alpha}_1$	-7.246	-2.449	-0.244	-0.029	2.348	-2.449	0.187	0.257	-0.133	0.318
$\hat{\alpha}_2$	1.886	1.267	0.790	0.744	0.648	1.267	0.697	-0.674	0.772	0.595
$\hat{\alpha}_3$	-0.066	-0.051	-0.042	-0.041	-0.037	-0.051	-0.040	-0.018	-0.041	-0.038
$\hat{\alpha}_4$	0.110	0.065	0.046	0.044	0.020	0.065	0.042	0.024	0.045	0.041
$\hat{\alpha}_5$	-0.312	-0.349	-0.310	-0.306	-0.386	-0.349	-0.303	0.162	-0.310	-0.279
MSE	21.3515	2.892	0.329	0.288	2.3913	2.892	0.284	0.312	0.306	0.272

- For theorem 3.1, since the condition $\alpha'[F_d W M - I_p]' [L(H^{-1} - F_d W M H^{-1} M' W' F_d')L'][F_d W M - I_p]\alpha = 0.001 < 1$ is satisfied, then the LLKL estimator is better than the MLE estimator.

- For theorem 3.2, since the condition $\alpha'[F_d W M - I_p]' [D_1 + [W - I_p]\alpha\alpha'[W - I_p]'][F_d W M - I_p]\alpha = 0.376 < 1$ is satisfied, then the LLKL estimator is better than the LRR estimator.

- For theorem 3.3, since the condition $\alpha'[F_d W M - I_p]' [D_2 + [F_d - I_p]\alpha\alpha'[F_d - I_p]'][F_d W M - I_p]\alpha = 0.009 < 1$ is satisfied, then the LLKL estimator is better than the LLE estimator.

- For theorem 3.4, since the condition $\alpha'[F_d W M - I_p]' [D_3 + [W M - I_p]\alpha\alpha'[W M - I_p]'][F_d W M - I_p]\alpha = 0.017 < 1$ is satisfied, then the LLKL estimator is better than the LKL estimator.

- For theorem 3.5, since the condition $\alpha'[F_d W M - I_p]' [D_4 + [R_{kd} - I_p]\alpha\alpha'[R_{kd} - I_p]'][F_d W M - I_p]\alpha = 0.005 < 1$ is satisfied, then the LLKL estimator is better than the LYC estimator.

- For theorem 3.6, since the condition $\alpha'[F_d W M - I_p]'[D_5 + [F_d W - I_p]\alpha\alpha'[F_d W - I_p]'][F_d W M - I_p]\alpha = 0.003 < 1$ is satisfied, then the LLKL estimator is better than the LTPE estimator.

5.2 Cancer Data

The theoretical results was analyzed using the cancer remission data. This data was originally adopted by [40] and recently employed by [32]. The response y_i has a value 1 when the patient a remission of complete cancer and the value of zero elsewhere. The regressor variables are: the index of cell (x_1), the index of smear (x_2), the index of infil (x_3), the index of blast (x_4) and the values of temperature (x_5). There are 27 patients in number where 9 are a remission complete cancer. The regressor variables are standardized. The $X'\hat{G}X$ matrix eigenvalues are $\lambda_1 = 9.2979$, $\lambda_2 = 3.8070$, $\lambda_3 = 3.0692$, $\lambda_4 = 2.2713$ and $\lambda_5 = 0.0314$. Consequently, the condition number was computed as max(λ)/min(λ) = 295.703. The results of the eigenvalue and the condition number means the multicollinearity exists. The estimated regression coefficients and the corresponding MSE values are given in Table 4. The results indicate that proposed LLKL3 estimator is preferred corresponding to possessing smallest MSE. Also, we verified the theoretical conditions to the cancer data.

As in the first application, we found that all conditions of theorems 3.1-3.6 are met, i.e., all the theorems inequalities are less than one.

Table 4. Regression coefficients and MSE for cancer data

Coef.	MLE	LRR	LLE	LTPE	LKL	LYC	LLKL1	LLKL2	LLKL3	LLKL4
$\hat{\alpha}_1$	-0.197	0.3591	0.3344	0.280	0.4696	0.350	0.226	-0.040	0.425	0.211
$\hat{\alpha}_2$	-1.5957	-0.1205	-.0724	0.021	-0.1212	-0.099	0.114	0.214	-0.085	0.123
$\hat{\alpha}_3$	1.8139	0.1564	0.1378	0.106	0.0576	0.147	0.074	0.073	0.068	0.076
$\hat{\alpha}_4$	1.3073	1.0211	0.9247	0.723	1.2297	0.984	0.522	-0.252	1.109	0.481
$\hat{\alpha}_5$	-0.4208	-0.3019	-0.2672	-0.195	-0.3815	-0.288	-0.123	0.105	-0.336	-0.109
MSE	32.9393	1.242	1.315	1.817	1.269	1.254	2.776	10.821	1.171	3.022

6 Some Concluding Remarks

The logistic model is used popularly for building model with a binary response with one or group of regressor variables. It is known, MLE is used for estimating the parameters of the logistic model. However, this estimator performance in the multicollinearity occurrence is not good. The logistic of ridge, Liu, KL, and estimator with two-parameter by [22] have been developed in replace of MLE. Here, we have proposed a new estimator called the LLKL estimator and the extended of Yang and Chang [19] estimator to handle multicollinearity in the logistic model. Theoretically, we have observed that the LLKL outperforms other considered in this study. We have evaluated and have compared these estimators through a simulation and two real-life data. Generally, the proposed LLKL with $\hat{k}_4$ is the best.

In future work, for example, we can provide a robust biased estimation of the logistic regression as an extension of [41, 42].

Acknowledgments:

The authors appreciate the Deanship of Scientific Research at King Saud University represented by the Research Center at CBA for supporting this research financially.

References:

[1] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems. *Technometrics,* 12(1):55–67 (1970).

[2] A.E. Hoerl, R.W. Kennard, K.F. Baldwin, Ridge regression: some simulation. *Commun Stat Theory Methods*, 4:105–123 (1975).

[3] J.F. Lawless, P. Wang, A simulation study of ridge and other regression estimators. *Commun Stat Theory Methods,* 5:307–323 (1976).
[4] B.M.G. Kibria, Performance of some new ridge regression estimators. *Commun Stat Theory Methods,* 32:419–435 (2003).
[5] G. Muniz, B.M.G. Kibria, On some ridge regression estimators: an empirical comparisons. *Commun Stat Simul Comput,* 38:621–630 (2009).
[6] A.F. Lukman, K. Ayinde, Review and classifications of the ridge parameter estimation techniques. *Hacet J Math Stat*, 46(5):953–967 (2017).
[7] A.F. Lukman, K. Ayinde, S.A. Ajiboye, Monte-Carlo study of some classification-based ridge parameter estimators. *J Mod Appl Stat Methods*, 16(1):428–451 (2017).
[8] A.F. Lukman, O.T. Arowolo, Newly proposed biased ridge estimator: an application to the Nigerian economy. *Pak J Stat*, 34(2):91–98 (2018).
[9] R.L. Schaeffer, L.D. Roi, R.A. Wolfe, A ridge logistic estimator. *Commun Stat Theory Methods,* 13:99–113 (1984).
[10] B.M.G. Kibria, K. Mansson, G. Shukur, Performance of some logistic ridge regression estimators. *Comput Economics,* 40(4):401–414 (2012).
[11] D. Inan, B.E. Erdogan, Liu-type logistic estimator. *Commun Stat Simul Comput,* 42(7):1578–1586 (2013).
[12] V. Nagarajah, P. Wijekoon, Stochastic restricted maximum likelihood estimator in logistic regression model. *Open J Stat,* 5:837–851 (2015).
[13] Y. Asar, M. Arashi, J. Wu, Restricted ridge estimator in the logistic regression model. *Commun Stat Simul Comput.* 46(8): 6538-6544 (2017).
[14] Y. Asar, A. Genc, Two-parameter ridge estimator in the binary logistic regression. *Commun Stat Simul Comput*. 46:9, 7088-7099 (2017).
[15] N., Varathan, P. Wijekoon, Optimal generalized logistic estimator. *Commun Stat Theory Methods*, 47(2):463–474 (2018).
[16] A.F. Lukman, E. Adewuyi, A.C. Onate, K. Ayinde, A Modified Ridge-Type Logistic Estimator. *Iran J Sci Technol Trans Sci.,* 44(3): 437-443 (2020).
[17] M.R. Abonazel, R.A. Farghali, Liu-type multinomial logistic estimator. *Sankhya B* 81(2): 203-225 (2019).
[18] R.A. Farghali, M. Qasim, B.M. Kibria, M.R. Abonazel, (2021). Generalized two-parameter estimators in the multinomial logit regression model: methods, simulation and application, *Commun Stat Simul Comput*. 1-16. https://doi.org/10.1080/03610918.2021.1934023
[19] H. Yang, X. Chang, A new two-parameter estimator in linear regression. *Commun Stat Theory Methods,* 39(6): 923–934 (2010).
[20] K. Liu, A new class of biased estimate in linear regression. *Communication in Statistics- Theory and Methods,* 22: 393–402 (1993).
[21] K. Mansson, B.M.G. Kibria, G. Shukur, On Liu estimators for the logit regression model. *Econ. Model*, 29(4):1483-1488 (2012).
[22] J. Huang, A simulation research on a biased estimator in logistic regression model. In Z. Li, X. Li, Y. Liu, & Z. Cai (Eds.), *Computational Intelligence and Intelligent Systems*, 389-395 (2012). Berlin, Germany: Springer.
[23] B.M.G. Kibria, A.F. Lukman, A New Ridge-Type Estimator for the Linear Regression Model: Simulations and Applications. *Scientifica* Article ID 9758378, 1-16 (2020).
[24] M.N. Akram, B.G. Kibria, M.R. Abonazel, N. Afzal, On the performance of some biased estimators in the gamma regression model: simulation and applications. *Journal of Statistical Computation and Simulation*, 1-23 (2022). DOI: 10.1080/00949655.2022.2032059
[25] M.R. Abonazel, I. Dawoud, F.A. Awwad, A.F. Lukman, Dawoud–Kibria Estimator for Beta Regression Model: Simulation and Application. *Front. Appl. Math. Stat. 7: 775068* (2022). *doi: 10.3389/fams.2022.*
[26] R.W. Farebrother, Further results on the mean square error of ridge regression. *J R Stat Soc,* B38:248–250 (1976).
[27] G. Trenkler, H. Toutenburg, Mean squared error matrix comparisons between biased estimators an overview of recent results. *Stat Pap* 31:165–179 (1990).
[28] D.G. Gibbons, A simulation study of some ridge estimators. *J. Amer. Statist. Assoc*. 76:131–139 (1981).
[29] A.F. Lukman, K. Ayinde, S. Binuomote, O.A. Clement, Modified ridge- type estimator to combat multicollinearity: Application to chemical data. *Journal of Chemometrics*, e3125 (2019).

[30] A.F. Lukman, K. Ayinde, S.K. Sek, E. Adewuyi, A modified new two-parameter estimator in a linear regression model. *Modelling and Simulation in Engineering*, 2019:6342702 (2019).

[31] I. Dawoud, M.R. Abonazel, F.A. Awwad, Generalized Kibria-Lukman Estimator: Method, Simulation, and Application. *Frontiers in Applied Mathematics and Statistics*, 8:880086 (2022).

[32] M.R. Ozkale, E. Arıcan, A new biased estimator in logistic regression model. *Statistics* 50(2): 233-253 (2016).

[33] J.P. Newhouse, S.D. Oman, An evaluation of ridge estimators. *Rand Corporation*. P-716-PR, 1–28 (1971).

[34] M.R. Abonazel, Z.Y. Algamal, F.A. Awwad, I.M. Taha, A New Two-Parameter Estimator for Beta Regression Model: Method, Simulation, and Application. *Front. Appl. Math. Stat. 7: 780322* (2022).

[35] Z.Y. Algamal, M.R. Abonazel, Developing a Liu- type estimator in beta regression model. *Concurrency and Computation: Practice and Experience*, *34*(5):e6685 (2022).

[36] M.R. Abonazel, I. Dawoud, Developing robust ridge estimators for Poisson regression model. *Concurrency and Computation: Practice and Experience*, e6979 (2022). https://doi.org/10.1002/cpe.6979.

[37] G. McDonald, D.I. Galarneau, A Monte Carlo evaluation of some ridge-type estimators. *J. Amer. Statist. Assoc*., 70(350):407–416 (1975).

[38] W.E.L., Pena, P.R.De., Massaguer, A.D.G., Zuniga, S.H.Saraiva, Modeling the growth limit of Alicyclobacillus acidoterrestris CRA7152 in apple juice: effect of pH, Brix, temperature and nisin concentration. *J Food Process Preserv* 35: 509-517 (2011).

[39] M.R. Ozkale, Iterative algorithms of biased estimation methods in binary logistic regression. *Stat. Pap*. 1-41 (2016).

[40] E. Lesaffre, B.D. Marx Collinearity in generalized linear regression. *Commun Stat Theory Methods*, 22(7):1933–1952 (1993).

[41] I. Dawoud, M.R. Abonazel, Robust Dawoud–Kibria estimator for handling multicollinearity and outliers in the linear regression model. *J Stat Comput Simul*. 91:3678–92 (2021).

[42] F.A. Awwad, I. Dawoud, M.R. Abonazel, Development of robust Özkale-Kaçiranlar and Yang-Chang estimators for regression models in the presence of multicollinearity and outliers. *Concurr Computat Pract Exp*. 34(6):e6779 (2022).