Iraqi Journal of Statistical Sciences

www.stats.mosuljournals.com

Performance of Some Yang and Chang estimators in Logistic Regression Model.

O.J. Oladapo[1] O.O Alabi[2] and K. Ayinde[3]

[1]Department of Statistics, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, Nigeria, [2,3]Department of Statistics, Federal University of Technology, Akure, Nigeria

Article information

Article history:
Received November 06, 2023
Accepted December 17, 2023
Available online June 1, 2024

Keywords:
Logistic regression
Multicollinearity
Biased estimators
Maximum likelihood
Simulation
MSE

Correspondence:
O.J Oladapo
ojoladapo@lautech.edu.ng.

Abstract

In logistic regression models, the maximum likelihood (ML) method is always one of the commonly used to estimate the model parameters. However, unstable parameter estimates are obtained as a result due to the problem of multicollinearity and the mean square error (MSE) gotten cannot also be relied on. Several biased estimators has been proposed to handle the issue of multicollinearity and the logistic Yang and Chang estimator (LYC) is one of them. Likewise research has also made us to understand that the biasing parameter has effect too on the value of the MSE. In this paper we proposed seven LYC biasing estimators and they were all subjected to Monte Carlo simulation studies and Pena data set was also used too. The result from the simulation study shows that LYC estimators outperforms the Logistic Ridge Regression (LRR) and the ML approach. Furthermore, application to Pena real data set also conform to the simulation results.

DOI: 10.33899/iqjoss.2024.183228 ,

Introduction

In multiple regression models, multicollinearity occurs when the explanatory variables are collinear, which was first introduced by Frisch [1]. This is a common occurrence in applied research. Using ordinary least squares (OLS) or maximum likelihood (ML) to estimate linear regression models and logit regression models leads to high variance and unstable parameter estimates. As a result, multicollinearity may lead to questions about the validity of the regression analysis conclusions. In the past few decades, ridge regression (RR) has been widely used as a corrective measure for linear regressions, with a lot of research focusing on estimating the ridge parameter k. Hoerl and Kennard [2,3] initially proposed ridge regression. Several studies have been conducted on the subject, including Gibbons [4], Lawless and Wang [5], Dempster et al. [6], Hoerl and Kennard [2], Hoerl et al. [7], McDonald and Galarneau [8], Alkhamisi et al. [9], Alkhamisi and Shukur [10], Muniz and Kibria [11], Muniz et al. [12], and Månsson et al. [13], Lukman and Ayinde [14], Ayinde et al [15] and others too. However, much attention has not been given into the logit model and those that worked on it are researchers like Schaeffer et al. [16], Schaeffer [17], Månsson and Shukur [18], Kibria et al [19], and few others. These researchers only focused on Ridge regression and no or little attention is given to other biasing parameter k emanating from other estimators too.

The main focus of this paper is to propose some Logistic Yang and Chang (LYC) estimators based on the work of Kibra [20] and Kibra et al [19]. Since it is anticipated that these estimators will have lower mean squared error (MSE) than that of the logistic ridge regression (LRR) and ML. The MSE is computed in order to assess the estimators' performance.

The work is structured as follows: we provide a description of the materials and statistical methods in Section. 2. Section 3 discusses the simulation and numerical findings. Section 4 provides a succinct overview and conclusions.

Materials and Methodology

On the basis of the research of Kibria [20] and Kibra et al. [19], we suggest a few LYC estimators in this section for estimating the biasing parameter k.

Logit Regression

Logit regression has always been one of the commonest method in statistic that is applied whenever the ith value of our so called dependent variable (y) of the regression model follows a Be (πi) distribution with the following parameter value:

$$\pi_i = \frac{\exp\left(x_i^{'}\beta\right)}{1+\exp\left(x_i^{'}\beta\right)} \tag{1}$$

Where $x_i^{'}$ is the ith row of X, and is a $n\times(p+1)$ data matrix, having p explanatory variables and such that β is a $(p+1)\times 1$ vector of coefficients. Using the Maximum Likelihood technique, which maximizes the following log likelihood, is one of the most used ways to estimate β and can be expressed as:

$$\text{L}=\sum_{i=1}^{n} y_i \log(\pi_i)+\sum_{i=1}^{n}(1-y_i)\log(1-\pi_i) \tag{2}$$

This can be achieved by setting the first derivative of the above expression to be equal to zero. Hence, the ML estimates are found by solving the subsequent equation:

Setting the first derivative of the aforementioned equation (2) to zero will do this. Therefore, the following equation below must be solved in order to find the ML estimates:

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

The iterative weighted least square (IWLS) algorithm is used:

$$\hat{\beta}_{MLE}=\left(X'\hat{W}X\right)^{-1}X'\hat{W}z \tag{4}$$

Where the following are the expression of $\hat{W}$ and $\hat{z}$ respectively

$\hat{W}=\pi_i(1-\pi_i)$ and $\hat{z}$ is known to be a vector where the ith element equals

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

Since equation (3) is nonlinear in β , we can express the MSE of the ML estimator as:

$$E\left(L_{ML}^2\right) = E\left(\beta_{ML} - \beta\right)' E\left(\beta_{ML} - \beta\right) = tr\left(X'\hat{W}X\right)^{-1} = \sum_{i=1}^{j} \frac{1}{\lambda_j} \tag{5}$$

λ_j is said to be the jth eigenvalues of the $X'\hat{W}X$ matrix.

Logistic Ridge Regression Estimator

Schaeffer et al. [16], proposed LRR estimator, as a substitute to the ML estimate that mitigates the problem of multicollinearity, Instead of directly estimating the coefficients of the regression model, the LRR estimator focuses on estimating the inverse of the covariance matrix. By doing so, the LRR estimator effectively reduces the impact of small eigenvalues caused by multicollinearity, resulting in a more reliable and robust estimation of the regression coefficients. The LRR estimator is defined as:

$$\hat{\beta}_{LRR} = \left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\hat{\beta}_{MLE} \tag{6}$$

k is the biasing parameter, $\hat{W}$ and $\hat{\beta}_{MLE}$ is the $\hat{\beta}_{MLE}$ estimates derived by equation(4). The LRR estimator MSE can be expressed as:

$$E\left(L_{LRR}^2\right) = E\left(\beta_{LRR} - \beta\right)' E\left(\beta_{LRR} - \beta\right) = \sum_{j=1}^{j} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + k^2 \sum_{j=1}^{j} \frac{\alpha_j}{\left(\lambda_j + k\right)^2} \tag{7}$$

Logistic Yang and Chang Estimator

The Logistic Yang and Chang (LYC) estimator, which is a special estimator of Liu and ridge estimator combined was proposed by Awwad et al [21] and it also handle the problem of multicollinearity effectively too. The estimator LYC is defined as:

$$\hat{\beta}_{LYC} = \left(X'\hat{W}X + I\right)^{-1}\left(X'\hat{W}X + dI\right)\left(X'\hat{W}X + kI\right)^{-1} X'\hat{W}X\hat{\beta}_{MLE} \tag{8}$$

k and d is the biasing parameters, $\hat{W}$ and $\hat{\beta}_{MLE}$ is the $\hat{\beta}_{MLE}$ estimates derived by equation(4). The LYC estimator MSE can be expressed as:

$$E\left(L_{LYC}^2\right) = E\left(\beta_{LYC} - \beta\right)' E\left(\beta_{LYC} - \beta\right) = \sum_{i=1}^{p} \left[\frac{\lambda_i\left(\lambda_i + d\right)^2}{\left(\lambda_i + I\right)^2\left(\lambda_i + k\right)^2}\right] + \sum_{i=1}^{p} \left[\frac{\left(\left(k + 1 - d\right)\lambda_i + k\right)^2}{\left(\lambda_i + I\right)^2\left(\lambda_i + k\right)^2}\right]\alpha_j^2 \tag{9}$$

where α_j^2 is expressed as the jth element of $\gamma\beta$ and γ is known to be the eigenvector expressed as $X'\hat{W}X = \gamma'\Lambda\gamma$, where $\Lambda = diag\left(\lambda_j\right)$.

The Proposed Estimators.

A ridge parameter can be chosen in many different ways, however, a number of approaches have been put out for the linear RR model, and these have been extended to the logistic ridge regression model. In the classical RR a biasing parameter k from the works of Hoerl and Kennard [2,3] is as expressed as follows:

$$\hat{k}_{HK1} = \frac{\sigma^2}{\alpha_{\max}^2} \tag{10}$$

The above biasing parameter was also adopted to have the following biasing parameter proposed by Kibria [20] as stated below

$$\hat{k}_{GM} = \frac{\sigma^2}{\left(\prod_{i=1}^{l} \hat{\alpha}_j^2\right)^{\frac{1}{l}}} \tag{11}$$

Later on, both equation (10 & 11) was adopted into the LRR by Schaeffer et al. [16 and Kibra et al [19] respectively.

However the biasing parameter k for LYC can be gotten from the MSE

$$MSE(\beta_{LYC}) = \sum_{i=1}^{p}\left[\frac{\lambda_i(\lambda_i+d)^2}{(\lambda_i+I)^2(\lambda_i+k)^2}\right] + \sum_{i=1}^{p}\left[\frac{((k+1-d)\lambda_i+k)^2}{(\lambda_i+I)^2(\lambda_i+k)^2}\right]\alpha_j^2 \tag{12}$$

Given that d is fixed, an ideal value of k is the value that will be to minimize $MSE(\hat{\beta}_{LYC})$.

Then, by differentiating $MSE(\hat{\beta}_{LYC})$ w.r.t. k and equating to 0, we have k as below:

$$k = \frac{(\lambda_i+d)-(1-d)\lambda_i\alpha_j^2}{d(\lambda_i+1)\alpha_j^2} \tag{13}$$

However, k depends on the unknown α_j. For practical purposes, it will be replaced by its unbiased estimator $\hat{\alpha}_j^2$. Hence, this will be obtained as:

$$\hat{k} = \frac{(\lambda_i+d)-(1-d)\lambda_i\hat{\alpha}_j^2}{(\lambda_i+1)\hat{\alpha}_j^2} \tag{14}$$

As an operational estimator for k. Furthermore, when d = 1, the above equation returns back to the k proposed by Schaeffer et al. [17] which is expressed as:

$$\hat{k} = \frac{1}{\alpha_i^2}$$

Following the works of Schaeffer et al. [16], Kibra [20] and Kibra et al [19] the following biasing parameter k for Yang and Chang are proposed as:

$$\hat{k}_{MAX} = Maximum\left(\frac{(\lambda_i+d)-(1-d)\lambda_i\hat{\alpha}_i^2}{(\lambda_i+1)\hat{\alpha}_i^2}\right) \tag{15}$$

$$\hat{k}_{MIN} = Minimum\left(\frac{(\lambda_i+d)-(1-d)\lambda_i\hat{\alpha}_i^2}{(\lambda_i+1)\hat{\alpha}_i^2}\right) \tag{16}$$

$$\hat{k}_{MED} = Median\left(\frac{(\lambda_i+d)-(1-d)\lambda_i\hat{\alpha}_i^2}{(\lambda_i+1)\hat{\alpha}_i^2}\right) \tag{17}$$

$$\hat{k}_{AM} = \frac{1}{p}\sum_{i=1}^{p}\left(\frac{(\lambda_i+d)-(1-d)\lambda_i\hat{\alpha}_i^2}{(\lambda_i+1)\hat{\alpha}_i^2}\right) \tag{18}$$

$$\hat{k}_{HM} = p\sum_{i=1}^{p} \left(\frac{(\lambda_i + d) - (1-d)\lambda_i \hat{\alpha}_i^2}{(\lambda_i + 1)\hat{\alpha}_i^2} \right) \tag{19}$$

$$\hat{k}_{MR} = \left(\hat{k}_{MAX} + \hat{k}_{MIN}\right)/2 \tag{20}$$

Furthermore, the optimal value for d can also be derived by minimizing $MSE\left(\hat{\beta}_{LYC}\right)$ in equation (12)

$$MSE(\hat{\alpha}_{YC}) = \sum_{i=1}^{p}\left[\frac{\lambda_i(\lambda_i + d)^2}{(\lambda_i + I)^2(\lambda_i + k)^2}\right] + \sum_{i=1}^{p}\left[\frac{((k+1-d)\lambda_i + k)^2}{(\lambda_i + I)^2(\lambda_i + k)^2}\right]\alpha_i^2$$

Then, by differentiating $MSE\left(\hat{\beta}_{LYC}\right)$ w.r.t. d and equating to 0, we have d as below:

$$\hat{d}_{opt} = \frac{\sum_{i=1}^{p}\left[\left(\left(\hat{k}+1\right)\lambda_i + k\right)\lambda_i \alpha_i^2 - \lambda_i^2\right]}{\sum_{i=1}^{p}\left(1+\lambda_i\alpha_i^2\right)\lambda_i} \tag{21}$$

Noting that when k =0 the above equation leads to the Liu biasing estimator, the above estimator would be

$$\hat{d}_{opt} = \frac{\sum_{i=1}^{p}\left[\left(\left(\hat{\alpha}_i^2 - 1\right)\lambda_i\right)\right]}{\sum_{i=1}^{p}\left(1+\lambda_i\alpha_i^2\right)} \tag{22}$$

The Monte Carlo Simulation

This paper's major goal is to determine how multicollinearity impacts ML, LRR and LYC Estimators. Therefore, the most important variable in the experiment is the degree to which the regressors were correlated. Hence, we generate the explanatory variables by using the following formula, which lets us adjust the correlation's strength:

$$x_{ij} = \left(1-\rho^2\right)^{1/2} z_{ij} + \rho z_{ip} \quad \text{i=1, 2,…, n, j=1,2,…,p} \tag{23}$$

The degree of correlation between the explanatory variables is referred to as ρ^2 and z_{ij} is also the pseudo random numbers from the standard normal distribution. The four various levels of ρ considered are said to be 0.8, 0.9, 0.95 and 0.99. Likewise the dependent variable is also derived from the $Be(\pi_i)$ distribution where

$$\pi_i = \frac{\exp\left(x_i^{'}\beta\right)}{1+\exp\left(x_i^{'}\beta\right)} \tag{24}$$

We set $\beta'\beta = 1$, according to Newhouse and Oman [22] statement that if our MSE is a function of β, σ^2 and k if all the explanatory variables used are fixed, we can then say that the MSE is minimized when this coefficient are choose. Models consisting p=2 and 3 explanatory variables, sample sizes n=20, 30 and 40 are used, models, p=5 and 6 explanatory variables, sample sizes n=200,250 and 300 and models consists of p as 9 and 10 explanatory variables, sample sizes n=800,900 and 1000 are used. With this experimental design we will be able to examine which of the Yang and Chang biasing k parameters that will give an optimal result for the designs. From the works of Kibria [20], Lukman et al [23,24], Oladapo et al [25], Muniz and Kibria [11], Idowu et al [26], Owolabi et al [27,28], Månsson and Shukur [18], among others we can get more detailed information on simulation procedures.

Results and Discussion based on Simulation

Findings from Monte Carlo investigation are all presented. The MSE values of all the estimators for Monte Carlo study are shown in Table 1- Table 3 and that of real life is in Table 4. Also, the outcome of varying the different factors we used in this study on the ML, LRR and LYC estimators are discussed too.

Table 1 Estimated MSE for different estimator when p=2 and 4

							LYCAM	LYCHM	LYCMAX	LYCMIN	LYCMR
20	0.8	P2	15.7012	0.8669	1.2025	0.9676	0.9676	**0.7260**	0.8100	5.1743	0.9676
		P3	24.7418	1.2913	2.0600	1.5576	0.9536	**0.7188**	0.7497	7.5635	0.9519
	0.9	P2	18.9648	1.5806	2.1055	1.7638	1.7638	**0.9226**	1.2596	6.3612	1.7638
		P3	45.9319	2.5406	3.8055	3.4579	1.8574	**0.6825**	1.1125	2.3387	1.7788
	0.95	P2	33.0799	2.8735	3.6370	3.5172	3.5172	**1.3650**	2.2381	15.9088	3.5172
		P3	118.2331	5.4285	7.3578	8.2295	4.2553	**0.7580**	2.1636	10.9160	3.9622
	0.99	P2	41.2311	14.2273	15.2330	20.8501	20.8501	**6.1897**	12.5252	27.1860	20.8501
		P3	323.0345	28.4999	37.3550	47.8129	28.9098	**2.5166**	14.6529	43.8784	27.3260
30	0.8	P2	852.0	0.8921	1.2706	1.6479	1.6479	**0.6506**	0.8444	85.1341	1.6479
		P3	40.0	1.3235	2.0820	1.1955	0.7060	0.7367	**0.6579**	2.7215	0.7172
	0.9	P2	3.3957	1.1102	1.6140	0.9092	0.9092	**0.6925**	0.7729	2.0927	0.9092
		P3	57.2895	1.8766	2.8766	1.8023	0.9592	**0.6693**	0.7402	5.0297	0.9333
	0.95	P2	6.8001	2.0606	2.7707	2.1703	2.1703	**1.0832**	1.5673	4.9591	2.1703
		P3	72.3784	3.9599	5.3176	4.6710	2.3247	**0.6986**	1.4299	23.8521	2.2156
	0.99	P2	33.6745	9.4476	10.0045	13.7039	13.7039	**4.2315**	8.3740	30.7374	13.7039
		P3	114.9015	20.6633	24.1214	29.3686	13.9822	**1.4808**	6.9275	154.2833	12.6276
40	0.8	P2	1.0500	0.4647	0.7297	0.4130	0.4130	0.4453	**0.4117**	0.5853	0.4130
		P3	1.9671	0.7233	1.0681	0.6170	**0.5642**	0.7113	0.5957	1.2121	0.5819
	0.9	P2	2.0654	0.7919	1.1649	0.6758	0.6758	**0.6164**	0.6307	1.0988	0.6758
		P3	4.0375	1.2530	1.7422	1.0173	0.7921	**0.6553**	0.6757	2.4140	0.7871
	0.95	P2	3.9378	1.3514	1.8909	1.3221	1.3221	**0.9298**	1.1088	2.4407	1.3221
		P3	7.8012	2.2170	3.0394	2.1573	1.4233	**0.6109**	0.9856	4.9249	1.3904
	0.99	P2	19.4569	5.8467	6.4598	8.1060	8.1060	**3.0853**	5.4174	16.7178	8.1060
		P3	39.6774	10.3239	13.0468	13.9107	7.1394	**1.0158**	3.8367	31.8973	6.5902

Bold values show the smallest MSE

In Table 1 is the estimated MSE values when n is 20, 30 and 40 for explanatory variables, p=2 and 3 it can be observed that the LYC with the biasing parameter k of the Harmonic version gives the lowest MSE in all cases except when n is 40, p=2 and ρ= 0.8 likewise for when n is 40, p=3 and ρ= 0.8.

Table 2: Estimated MSE for different estimator when p=5 and 6

							LYCAM	LYCHM	LYCMAX	LYCMIN	LYCMR
200	0.8	P5	0.6437	0.3082	0.6918	**0.2117**	0.3041	0.7669	0.4480	0.4711	0.3707
		P6	0.8897	0.3619	0.7920	**0.2103**	0.3533	0.8563	0.5380	0.6147	0.4492
	0.9	P5	1.3254	0.5094	0.9017	**0.2870**	0.3427	0.7419	0.4584	0.8895	0.4007

		P6	1.8711	0.6049	1.0433	**0.3291**	0.3711	0.8144	0.5079	1.1751	0.4407
	0.95	P5	2.7185	0.8755	1.4135	0.5020	**0.3785**	0.6872	0.4349	1.6734	0.4054
		P6	3.8387	1.0614	1.6751	0.5593	**0.3860**	0.7530	0.4601	2.2541	0.4239
	0.99	P5	14.3758	3.8963	5.4482	2.8161	0.7774	0.4529	**0.4332**	8.8957	0.6269
		P6	21.4196	5.2274	7.4239	2.6308	0.9041	0.5061	**0.4381**	13.2737	0.6521
250	0.8	P5	0.4684	0.2425	0.6515	**0.1740**	0.2409	0.6978	0.3713	0.3484	0.2990
		P6	0.6067	0.2957	0.9077	**0.1922**	0.3024	0.8365	0.4808	0.4507	0.3900
	0,9	P5	0.9953	0.4108	0.8095	**0.2422**	0.2806	0.7148	0.4039	0.6817	0.3368
		P6	1.2397	0.4749	0.8914	**0.2594**	0.3415	0.8217	0.4937	0.8356	0.4199
	0.95	P5	1.9988	0.6814	1.1652	0.3829	**0.3210**	0.6593	0.3888	1.2577	0.3525
		P6	2.5455	0.8113	1.3474	0.4300	**0.3285**	0.7744	0.4480	1.5805	0.3879
	0.99	P5	10.6428	2.8765	4.0386	1.8822	0.5850	0.4465	**0.3608**	6.0543	0.4921
		P6	14.0437	3.6615	5.2863	1.8841	0.6025	0.5543	**0.3940**	8.1619	0.4920
300	0.8	P5	0.3884	0.2125	0.9185	**0.1581**	0.2271	0.6639	0.3415	0.2973	0.2780
		P6	0.5511	0.2701	1.4256	**0.1767**	0.2823	0.7966	0.4474	0.4114	0.3659
	0.9	P5	0.8046	0.3458	0.8087	**0.2116**	0.2628	0.7029	0.3900	0.5597	0.3203
		P6	1.1421	0.4333	0.7841	**0.2337**	0.3050	0.7887	0.4499	0.7660	0.3782
	0.95	P5	1.6083	0.5546	0.9556	0.3175	**0.3013**	0.6497	0.3789	1.0052	0.3375
		P6	2.4319	0.7680	1.3087	0.3739	**0.3077**	0.7335	0.4052	1.5117	0.3548
	0.99	P5	8.5604	2.2965	3.2985	1.6024	0.4759	0.4506	**0.3249**	4.6604	0.4046
		P6	13.3230	3.4459	5.0080	1.6875	0.4859	0.5273	**0.3344**	7.7822	0.3986

Bold values show the smallest MSE

In Table 2 is the estimated MSE values when n is 200, 250 and 300 for explanatory variables, p=5 and 6 it can be observed that the LYC with the biasing parameter k of the median version gives the lowest MSE when the multicollinearity level ρ= 0.8 and 0.9, likewise for when ρ= 0.95 its seen that the LYC with the biasing parameter k of the arithmetic mean version gives the lowest MSE and when ρ= 0.99 its seen that the LYC with the biasing parameter k of the Maximum version gives the lowest MSE.

Table3: Estimated MSE for different estimator when p=9 and 10

							LYCAM	LYCHM	LYCMAX	LYCMIN	LYCMR
800	0.8	P9	0.3417	0.1878	0.1821	**0.1397**	0.2014	0.8273	0.3675	0.2725	0.2937
		P10	0.3792	0.1991	0.2160	**0.1432**	0.2098	0.8773	0.4158	0.2961	0.3280
	0.9	P9	0.7358	0.3092	0.1933	**0.1907**	0.2378	0.8591	0.4361	0.5263	0.3497
		P10	0.8318	0.3295	0.6946	**0.1929**	0.2710	0.8986	0.4932	0.5817	0.4019
	0.95	P9	1.5571	0.5153	0.9953	**0.2796**	0.2500	0.8421	0.4326	1.0094	0.3522
		P10	1.8180	0.5708	1.1537	0.2883	**0.2861**	0.8821	0.4858	1.1771	0.4064
	0.99	P9	8.5245	2.0855	3.2954	0.9461	**0.2717**	0.7163	0.3344	5.0472	0.2955
		P10	9.9631	2.3190	3.6313	0.9558	**0.2882**	0.7667	0.3714	5.8876	0.3271
900	0.8	P9	0.2841	0.1687	0.8738	**0.1342**	0.1727	0.7936	0.3180	0.2320	0.2514
		P10	0.3258	0.1833	0.3565	**0.1411**	0.1981	0.8583	0.3804	0.2599	0.3015
	0.9	P9	0.6200	0.2810	0.2457	**0.1879**	0.2377	0.8452	0.4257	0.4541	0.3442
		P10	0.6912	0.2909	0.1923	**0.1835**	0.2294	0.8890	0.4614	0.4912	0.3642
	0.95	P9	1.2866	0.4574	1.7676	0.2558	**0.2494**	0.8369	0.4351	0.8609	0.3567
		P10	1.4883	0.4886	1.7130	0.2607	**0.2592**	0.8797	0.4716	0.9635	0.3838

	0.99	P9	7.1335	1.7859	2.9241	0.8173	**0.2464**	0.7161	0.3262	4.2664	0.2836
		P10	8.3093	1.9836	3.1485	0.8233	**0.2613**	0.7709	0.3639	4.9146	0.3124
1000	0.8	P9	0.2530	0.1541	131.8577	**0.1250**	0.1581	0.7633	0.2838	0.2073	0.2267
		P10	0.2973	0.1722	90.8400	**0.1343**	0.1764	0.8380	0.3434	0.2409	0.2697
	0.9	P9	0.5497	0.2548	38.5145	**0.1719**	0.2241	0.8338	0.4064	0.4043	0.3278
		P10	0.6513	0.2848	31.8431	**0.1812**	0.2458	0.8691	0.4460	0.4712	0.3645
	0.95	P9	0.2530	0.1541	131.8577	**0.1250**	0.1581	0.7633	0.2838	0.2073	0.2267
		P10	0.2973	0.1722	90.8400	**0.1343**	0.1764	0.8380	0.3434	0.2409	0.2697
	0.99	P9	0.5497	0.2548	38.5145	**0.1719**	0.2241	0.8338	0.4064	0.4043	0.3278
		P10	0.6513	0.2848	31.8431	**0.1812**	0.2458	0.8691	0.4460	0.4712	0.3645

Bold values show the smallest MSE

In Table 3 is the MSE values for n at 800, 900 and 1000 for explanatory variables, p=9 and 10 it can be observed that the LYC with the biasing parameter k of the median version gives the lowest MSE when the multicollinearity level ρ= 0.8 and 0.9, likewise for when ρ= 0.95 its seen that the LYC with the biasing parameter k of the arithmetic mean version gives the lowest MSE for both sample sizes 800 and 900. But when n =1000 its seen that the LYC with the biasing parameter k of the Median version gives the lowest MSE.

Real Life Data

Pena et al [29] used a logistic model to investigate the effects of temperature, pH, and soluble solids content on the response of Alicyclobacillus growth likelihood in apple juice. The matrix's eigenvalues are 13464.7990, 1715.9257, 56.5515, and 3.5445. As a result, the condition index (C.I) is 61.6342, indicating that multicollinearity exists in the model. Table 4 shows the estimated regression coefficient values from each estimator, as well as the accompanying mean squared error.

When there is multicollinearity, the ML estimator performs the least well, as expected. The selection of k and d (as shrinkage parameters) determines the efficiency of biased estimators. All of the proposed estimators performed admirably, and one of them has the minimum mean square error, which corresponds to the simulation outcome.

Table 4: Regression coefficients and MSE

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	SMSE
$\hat{\beta}_{ML}$	-7.24633	1.885951	-0.06628	0.110422	-0.31173	21.3513884
$\hat{\beta}_{LRE}$	-2.4E-06	0.008038	-0.02442	0.015783	-0.01186	0.28340673
$\hat{\beta}_{LYC}$	8.57E-05	0.004176	-0.02095	0.01126	-0.0053	0.28280132
$\hat{\beta}_{LYCMED}$	-0.00629	0.2343	-0.03332	0.042609	-0.1491	*0.13778692*
$\hat{\beta}_{LYCMIN}$	-7.11297	1.86508	-0.06581	0.109188	-0.3117	20.5909074
$\hat{\beta}_{LYCMAX}$	-6.2E-05	0.01115	-0.02575	0.018235	-0.01707	0.20795783
$\hat{\beta}_{LYCMR}$	-0.00029	0.022397	-0.0278	0.023823	-0.03289	*0.14054511*
$\hat{\beta}_{LYCHM}$	8.33E-05	0.002112	-0.01649	0.006978	-0.00223	1.43890463
$\hat{\beta}_{LYCAM}$	-0.00055	0.0406	-0.02908	0.029073	-0.05205	*0.12388644*

Conclusion

In this paper we were able to recommend some LYC estimators in estimating the biasing parameter k, in which both Monte Carlo study and real life data was used to investigate the performance of these estimators. The MSE criterion was used in evaluating the performances of the estimators to know the best among them. In the simulation study it was seen that at small samples sizes (20, 30 and 40) and number of explanatory variable say, p= 2 and 3 the biasing parameter k with the Harmonic version has the lowest MSE at almost all degree of correlation. Likewise for sample size (200, 250 and 300) with number of explanatory variable say, p= 5 and 6 and the degree of correlation is low the biasing parameter k with the median version has the lowest and when ρ is high biasing parameter k with the Arithmetic mean and maximum version is the best option to use. Also when sample size (800, 900 and 1000) with number of explanatory variable say, p= 9 and 10 and the degree of correlation is low the biasing parameter k with the median version has the lowest and when ρ is high biasing parameter k with the Arithmetic mean is the one with lowest MSE. In addition, from the numerical example the biasing parameter k with the arithmetic mean and the median version has the two lowest MSE respectively.

Recommendations

Hence, based from our findings both in simulation and numerical examples we thereby recommend to researchers and scholars, when having the issue of multicollinearity in logistic model, let the LYC (biasing k with the version of Arithmetic mean, Harmonic mean and the Median version) estimator be used.

Acknowledgment

The authors are very grateful to Federal University of Technology, Akure and Ladoke Akintola University of Technology for their support, love and tutoring which also helped to improve this work's quality.

Conflict of interest

The author has no conflict of interest.

References

1. Frisch, R. . Statistical confluence analysis by means ofcomplete regression systems. Publication 5. Oslo: University Institute of Economics.
2. Hoerl, A.E., Kennard, R.W. Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, 1970a;12(1): 55–67. https://doi.org/doi:10.1080/00401706.1970.10488634.
3. Hoerl, A. E., Kennard, R. W. . Ridge regression: Application to non-orthogonal problems. Technometrics, 1970,12(1): 69–82.
4. Gibbons, D. G. . A simulation study of some ridge estimators. *Journal of the American Statistical Association,1981;* 76:131–139.
5. Lawless, J. F., Wang, P. A simulation study of ridge and other regression estimators. Com- munications in Statistics: Theory and Methods, 1976;5, 307–323.
6. Dempster, A. P., Schatzoff, M., Wermuth, N. . A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association*, 1977;72: 77–91.
7. Hoerl, A.E., Kennard, R.W., Baldwin, K.F. Ridge regression: Some simulation. Commun. Stat. Theory Methods. 1975;4: 105–123. https://doi.org/doi:10.1080/03610927508827232
8. McDonald, G. C., & Galarneau, D. I. . A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*,1975; 70:407–416.
9. Alkhamisi, M. A., Khalaf, G., Shukur, G. . Some modifications for Choosing Ridge Parameter. *Communications in Statistics: Theory and Methods,* 2006;35: 1–16.
10. Alkhamisi, M., Shukur, G. . Developing ridge parameters for SUR model. *Communications in Statistics: Theory and Methods*, 2008;37,:544–564.
11. Muniz, G., Kibria, B. M. G. . On some ridge regression estimators: An empirical compari sons. *Communications in Statistics: Simulation and Computation*, 2009;38: 621–630
12. Muniz G., Golam Kibria B. M., Shukur, G. (On developing ridge regression parameters: A graphical investigation. Submitted for Publication in SORT 2011.
13. Månsson, K., Shukur, G., & Kibria, B. M. G. . A simulation study of some ridge regression estimators under different distributional assumptions. *Communications in Statistics: Simulation and Computation*, 2010;39: 1639–1670.
14. Lukman A.F., Ayinde K., Review and classifications of the ridge parameter estimation techniques. Hacet J Math Stat ,2017;46(5):953–967

15. Ayinde, K., Lukman, A.F., Samuel, O.O., Ajiboye, S.A.Some new adjusted ridge estimators of linear regression model. International Journal Civil Engineerint Technology.,2018; 9, 2838–2852.
16. Schaefer R.L., Roi L.D., Wolfe R.A., A ridge logistic estimator, Commu- nications in Statistics — Theory and Methods,1984; 13(1): 99–113.
17. Schaefer R. L.Alternative regression collinear estimators in logistic when the data are col- linear. Journal of Statistical Computation and Simulation ,1986;25 (1–2):75–91.
18. Månsson, K., Shukur, G. On ridge parameters in logistic regression. To appear in Communications in Statistics: Theory and Methods.
19. Kibria, B. M. G., Månsson,K., Shukur,G. . Performance of some logistic ridge regres- sion estimators. *Computational Economics* 2012;40 (4):401–14. https://doi.org/10.1007/s10614-011-9275-x.
20. Kibria, B.M.G. Performance of some new ridge regression estimators, Communications in Statistics — *Simulation and Computation, 2003;*32(2), 419–435. https://doi.org/10.1081/S.AC-120017499.
21. Awwad, F. A., Odeniyi, K.A, Dawoud, I, Yahya,Z., Abonazel,M.R, Kibria,B.M, Eldin, E.T. . New Two-Parameter Estimators for the Logistic Regression Model with Multicollinearity. *WSEAS TRANSACTIONS on MATHEMATICS.2022;* 21:403–414. https://doi.org/10.37394/23206.2022.21.48
22. Newhouse J. P., Oman S. D. An evaluation ofridge estimators. Rand Corporation,1971;(P-716-PR), 1-16.
23. Lukman AF, Ayinde K, Binuomote S, & Clement OA Modified ridge-type estimator to combat multicollinearity: Application to chemical data. *Journal of Chemometrics 2019;*33: e3125. https://doi.org/ 10.1002/cem.3125
24. Lukman, A. F., Emmanuel, A., Clement, O. A., Ayinde, K. A Modified Ridge-Type Logistic Estimator. *Iranian Journal of Science and Technology, Transactions A: Science*, 2020;*44*(2), 437–443. https://doi.org/10.1007/s40995-020-00845-z
25. Oladapo, O. J., Owolabi,A.T, Idowu,J.I, Ayinde K.. A New Modified Liu Ridge-Type Estimator for the Linear Regression Model : Simulation and Application. *Clinical Biostatistics and Biometrics.* 2022;8(2):1–14. https://doi.org/10.23937/2469-5831/1510048
26. Idowu,J.I.,Oladapo,O.J.,Owolabi,A.T., Ayinde,K.. . On the Biased Two-Parameter Estimator to Combat Multicollinearity in Linear Regression Model. *African Scientific Reports* 2022;1:188–204. https://doi.org/10.46481/asr.2022.1.3.57
27. Owolabi, A. T., Ayinde, K., & Alabi, O. O. . *African Scientific Reports A Modified Two Parameter Estimator with Di ff erent Forms of Biasing Parameters in the Linear Regression*. 2022;*1*, 212–228.
28. Owolabi, A. T.,Ayinde,K., Idowu,J.I., Oladapo,O.J., Adewale,A.F. . A New Two-Parameter Estimator in the Linear Regression Model with Correlated Regressors. *Journal of Statistics. Application and Probability.* 2022;11(2):499–512. http://dx.doi.org/10.18576/jsap/110211
29. Pena, W, Massaguer P.D, Zuniga, A, Saraiva S.H .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* ,2011;35 (4):509–17. https://doi.org/10.1111/j.1745-4549.2010.00496.x

أداء بعض مقدرات يانغ وتشانغ في نموذج الانحدار اللوجستي.

أو. ج. أولادابو[1] أو .أو ألابي[2] و ك .أينديي[3]

[1]قسم الإحصاء، جامعة لادوك أكينتولا للتكنولوجيا، أوغبوموسو، ولاية أويو، نيجيريا، [2،3]قسم الإحصاء، الجامعة الفيدرالية للتكنولوجيا، أكوري، نيجيريا

الخلاصة: في نماذج الانحدار اللوجستي، تكون طريقة الاحتمالية القصوى (ML) دائمًا إحدى الطرق الشائعة الاستخدام لتقدير معلمات النموذج. ومع ذلك، يتم الحصول على تقديرات المعلمات غير المستقرة نتيجة لمشكلة العلاقة الخطية المتعددة ولا يمكن أيضًا الاعتماد على متوسط مربع الخطأ (MSE) الذي تم الحصول عليه. لقد تم اقتراح العديد من المقدرين المتحيزين للتعامل مع مسألة العلاقة الخطية المتعددة، ومقدر يانغ وتشانغ اللوجستي (LYC) هو واحد منهم. وبالمثل، فقد جعلنا البحث نفهم أن معلمة الانحياز لها تأثير أيضًا على قيمة المشروعات متناهية الصغر والصغيرة. في هذا البحث اقترحنا سبعة مقدرات متحيزة لـ LYC وتم إخضاعهم جميعًا لدراسات محاكاة مونت كارلو وتم استخدام مجموعة بيانات Pena أيضًا. تظهر نتيجة دراسة المحاكاة أن مقدرات LYC تتفوق في الأداء على انحدار اللوجستي ريدج (LRR) ونهج التعلم الآلي. علاوة على ذلك، فإن التطبيق على مجموعة بيانات Pena الحقيقية يتوافق أيضًا مع نتائج المحاكاة.

الكلمات المفتاحية: الانحدار اللوجستي، متعددة الخطية ،المقدرون المتحيزون ،أقصى احتمال ،محاكاة ،MSE.