Performance of Some Dawoud-Kibra Estimators for Logistic Regression Model: Application to Pena data set

OLASUNKANMI JAMES OLADAPO[1,*], OLUSEGUN O.ALABI[2], KAYODE AYINDE[2,3]

[1]Department of Statistics, Ladoke Akintola University of Technology, Ogbomoso, Oyo State, NIGERIA
[2]Department of Statistics, Federal University of Technology, Akure, Ondo state, NIGERIA
[3]Department of Mathematics and Statistics, Northwest Missouri state university, Maryville, Missouri, USA
[1*]Corresponding author

*Abstract:*A logistic regression model's parameters are usually estimated using the maximum likelihood (ML) method. As a consequence of the problem of multicollinearity, unstable parameter estimates are obtained, and the mean square error (MSE) obtained cannot also be relied upon. There have been several biased estimators proposed to deal with multicollinearity, and the logistic Dawoud-Kibra (LDK) estimator is one of them, and research has shown that biasing parameters have an effect on MSE, too. Our study proposed seven LDK biasing estimators and all of them were subjected to Monte Carlo simulations, as well as using Pena data sets. According to the simulation study, LDK estimators outperform Logistic Ridge Regression (LRR) and ML methods. Furthermore, application to Pena real data set also align with the simulation results.

*Key-words***:** Logistic regression, Multicollinearity, Biased estimators, Maximum likelihood, Simulation, MSE.

Received: January 5, 2023. Revised: October 15, 2023. Accepted: November 17, 2023. Published: December 31, 2023.

1 Introduction

It was Frisch [1] who first introduced the concept of multicollinearity in multiple regression models. It is a common occurrence in applied research. For linear regression models and logit regression models, the use of ordinary least squares (OLS) and maximum likelihood (ML) leads to high variance and unstable parameter estimates. Due to multicollinearity, regression analysis's conclusions may be questioned. As a correction measure for linear regression, ridge regression (RR) has become increasingly popular in recent decades [2]. Many studies have been conducted to estimate the ridge parameter k, originally proposed by Hoerl and Kennard [2, 3]. Such studies, performed on ridge regression, include the works of Gibbons [4], Lawless and Wang [5], and 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, logit models have not received much attention, and only a few researchers have tried to work on them. Those who have studied them are the likes of Schaeffer *et al*. [16, 17], Månsson and Shukur [18], Kibria *et al.* [19], and a few others. Almost all the researchers working on the logit models focused only on the RR and paid little or no attention to the biasing parameter of all other estimators proposed by other researchers to handle the problem of multicollinearity in the logit models.

This research paper will focus on proposing some Logistic Dawoud and Kibra (LDK) estimators, following the works of Kibra *et al.* [19, 20]. Research has shown that the biasing parameters of an estimator have an effect on the value of the mean square error (MSE). Therefore, its anticipated LDK estimators will have MSE values lower than those of the LRR and ML. The MSE has been one of the criteria used in judging the performances of estimators.

This research paper is structured as follows: Section 2 entails the materials and methodology adopted for the study. Section 3 includes the results and discussion of simulation and numerical results. Section 4 provides a succinct overview and conclusions.

2. Materials and Methodology.

Adopting the concepts in the research works of Kibra *et al.* and Kibra [19, 20], we will be proposing some LDK estimators in this section for estimating the biasing parameter k.

2.1 Logit Regression

The logit regression has always been one of the most common statistical methods used whenever the i-th value of the dependent variable (y) follows a Be (πi) distribution with the following parameter value:

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

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:

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

To get this we set the first derivative of equation (2) to be equal to zero, the ML estimates can now be derived by solving the equation below

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

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

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

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:

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

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

The LRR estimator, proposed by Schaeffer *et al*. [16], is a substitute for ML estimates that mitigates multicollinearity problems. Instead of estimating the regression model coefficients directly, it estimates the inverse of the covariance matrix. In this way, the LRR estimator effectively reduces small eigenvalues caused by multicollinearity; therefore, the regression coefficients are more reliable and robust.

The LRR estimator is express as:

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

With k has the biasing parameter, $\hat{W}$ and $\hat{\beta}_{MLE}$ is the $\hat{\beta}_{MLE}$ estimates derived from equation (4). The LRR estimator MSE is shown to be:

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

The Logistic Dawoud and Kibra (LDK) estimator, which is a special two parameter estimator of Kibra-Lukman (KL) estimator that was proposed by Afzal *et al* [21] and it also handle the problem of multicollinearity effectively too. The estimator LDK is defined as:

$$\hat{\beta}_{LDK} = \left(X'\hat{W}X + k(1+d)I\right)^{-1}\left(X'\hat{W}X - k(1+d)I\right)\hat{\beta}_{MLE} \quad (8)$$

With k and d, the biasing parameters, $\hat{W}$ and $\hat{\beta}_{MLE}$ is the $\hat{\beta}_{MLE}$ estimates derived from equation (4).

The MSE of the LDK estimator is express to be:

$$MSE\left(\hat{\beta}_{LDK}\right) = \sum_{i=1}^{p}\frac{\left(\lambda_i - k(1+d)\right)^2}{\lambda_i\left(\lambda_i + k(1+d)\right)^2} + 4k^2(1+d)\sum_{i=1}^{p}\frac{\alpha_i^2}{\left(\lambda_i + k(1+d)\right)^2} \quad (9)$$

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 $\Lambda = diag(\lambda_j)$.

2.2 The Dawoud-Kibra Estimators.

There are numerous approaches that have been developed for the linear regression model and then transferred to the logistic ridge regression model for selecting a ridge parameter. A biasing parameter k from Hoerl and Kennard's [2, 3] research in the classical RR is represented as follows:

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

As shown below, the aforementioned biasing parameter was also used to obtain the biasing parameter that Kibria [19] suggested.

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

Later on, equation (10) was adopted into the LRR by Schaeffer *et al*. [16] as:

$$\hat{k}_{SRW} = \frac{1}{\alpha_{\max}^2} \quad (12)$$

However the biasing parameters k and d for LDK can be gotten from the MSE:

$$MSE(\hat{\beta}_{LDK}) = \sum_{i=1}^{p}\frac{(\lambda_i - k(1+d))^2}{\lambda_i(\lambda_i + k(1+d))^2} + 4k^2(1+d)\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i + k(1+d))^2}$$

Then, by differentiating $MSE(\hat{\alpha}_{LDK})$ w.r.t. d and equating to 0, we have

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

However, d depends on the unknown α_i. For practical purposes, it will be replaced by its unbiased estimator $\hat{\alpha}_i$. Hence, this will be expressed as:

$$\hat{d} = \sum_{i=1}^{p}\left[\frac{\lambda_i}{\hat{k}(1+2\lambda_i\alpha_i^2)}\right] - 1 \quad (13)$$

Then, by differentiating $MSE(\hat{\alpha}_{LDK})$ w.r.t. k and equating to 0, we have

$$k = \frac{1}{(1+d)\left(\frac{1}{\lambda_i} + 2\alpha_i^2\right)} \quad (14)$$

However, k depends on the unknown α_i. For practical purposes, it will be replaced by its unbiased estimator $\hat{\alpha}_i$. Hence, this will be expressed as:

$$\hat{k} = \frac{1}{(1+d)\left(\frac{1}{\lambda_i} + 2\hat{\alpha}_i^2\right)} \quad (15)$$

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

$$\hat{k}_{AM} = \frac{1}{p}\sum_{i=1}^{p}\frac{1}{(1+d)\left(\frac{1}{\lambda_i} + 2\hat{\alpha}_i^2\right)} \quad (16)$$

$$\hat{k}_{HM} = p\sum_{i=1}^{p}\frac{1}{(1+d)\left(\frac{1}{\lambda_i} + 2\hat{\alpha}_i^2\right)} \quad (17)$$

$$\hat{k}_{MAX} = Maximum\left(\frac{1}{(1+d)\left(\frac{1}{\lambda_i} + 2\hat{\alpha}_i^2\right)}\right) \quad (18)$$

$$\hat{k}_{MIN} = Minimum\left(\frac{1}{(1+d)\left(\frac{1}{\lambda_i} + 2\hat{\alpha}_i^2\right)}\right) \quad (19)$$

$$\hat{k}_{MED} = Median\left(\frac{1}{(1+d)\left(\frac{1}{\lambda_i} + 2\hat{\alpha}_i^2\right)}\right) \quad (20)$$

$$\hat{k}_{MR} = \frac{(\hat{k}_{MAX} + \hat{k}_{MIN})}{2} \quad (21)$$

2.3 The Monte Carlo Simulation

As the main objective of this paper is to ascertain the effects of multicollinearity on ML, LRR, and LDK Estimators, the degree of correlation between the regressors is the most significant variable in the experiment. Accordingly, we generate the explanatory variables using the following formula, which allows us to adjust the correlation's strength:

$$x_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho z_{ip} \quad i=1, 2,\ldots, \; , j=1,2,\ldots,p \quad (22)$$

The term ρ^2 describes the level of correlation between the explanatory factors and z_{ij} is the usual normal distribution's pseudorandom numbers as well. The four different levels of correlation that are being evaluated are 0.8, 0.9, 0.95, and 0.99 respectively.

Similarly, the dependent variable comes from $Be(\pi_i)$ distribution where

$$\pi_i = \frac{\exp\left(x_i'\beta\right)}{1+\exp\left(x_i'\beta\right)} \quad (23)$$

We set $\beta'\beta = 1$, then the MSE is minimized when this coefficient is chosen, in accordance with Newhouse and Oman's [22] assertion that if our MSE is a function of β, σ^2 and k if all the explanatory variables utilized are fixed. Sample sizes used are 50, 60,70,80,90,100,150,200 and 250, likewise the number of explanatory variable considered are p=3, 4,5 and 6 .We will be able to determine which of the Dawoud-Kibra biasing k parameters will be more efficient using this experimental design.

We may find additional details on simulation processes from the works of Kibria [20] Lukman *et al* [23]; Oladapo *et al* [24,25]; Muniz and Kibria [11]; Idowu *et al* [26]; Owolabi *et al* [27]; Månsson and Shukur [18] and others too.

3. Results and Discussion

3.1 Simulation Results

Results from both real-world data and Monte Carlo simulations are presented in this section. Tables 1 through 4 display the MSE values of all the estimators used in the Monte Carlo study, and Table 5 displays the MSE values from a real-life data set. Additionally, the effects of varying the many factors we employed in this investigation on the ML, LRR, and LDK estimators are also covered.

Table 1 shows the estimated MSE values when n = 50, 60, and 70 for explanatory variables, p = 3, 4, and 6. It can be observed that the LDK with the biasing parameter k of the Median (MED) version gives the lowest MSE in all cases, with few exceptions.

Table 1: Estimated MSE for different estimator when p=3, 4, 5 and 6 when n=50, 60 and 70

n	ρ	P	MLE	RIDGE	DWD	DWDMED	DWDAM	DWDHM	DWDMAX	DWDMIN	DWDMR
50	0.8	P3	3.0134	1.1814	1.7391	1.0676	**1.0149**	5.3236	1.2515	1.6028	1.1761
		P4	7.9549	2.2022	3.1145	**1.8250**	2.1445	11.5557	2.9825	3.5858	2.7601
		P5	9.4676	2.4363	3.3850	**1.7441**	2.2337	15.0712	3.6191	4.1488	3.2496
	0.9	P3	5.6308	1.8903	2.5500	**1.4500**	1.5099	5.9809	1.7870	2.5973	1.8724
		P4	14.8361	3.7137	4.9707	**2.7193**	3.4177	16.0040	4.6564	6.0935	4.6380
		P5	18.8392	4.4338	5.7219	**3.3098**	4.2719	22.0825	6.1699	7.7862	5.8966
	0.95	P3	11.2610	3.4218	4.1419	**2.6086**	2.6929	8.4299	3.0232	4.7187	3.3710
		P4	30.0187	7.1853	9.0706	**4.7665**	6.5271	27.4225	8.9713	12.0734	8.8419
		P5	47.2517	9.3427	10.6448	**6.1189**	8.7187	46.7525	17.6080	19.4230	11.9529
	0.99	P3	59.2260	16.5227	16.8080	14.1993	**12.2145**	34.6057	12.6600	23.1891	14.1815
		P4	158.611	36.4383	41.0370	**20.6063**	34.2339	130.699	45.2742	64.5093	38.5315
		P5	205.860	**44.972**	202.364	203.297	204.133	205.666	202.475	186.491	202.413
60	0.8	P3	2.4854	1.0930	1.4374	0.8925	**0.7803**	4.6463	0.9476	1.3806	0.9419
		P4	5.2939	1.6122	2.3796	1.1421	**1.3809**	8.7297	1.9256	2.5126	1.9483
		P5	5.8827	1.6936	2.4839	**1.2193**	1.3175	11.4159	2.1385	2.7495	1.9077
	0.9	P3	4.9228	1.8428	2.3971	**1.3138**	1.3603	5.0615	1.5324	2.4047	1.7521
		P4	10.4370	2.7783	3.8214	**2.0090**	2.6639	11.8563	3.1496	4.5404	3.7450
		P5	13.4614	3.3901	4.6813	**2.4497**	2.8022	16.9669	4.0399	5.9529	4.2871
	0.95	P3	9.7720	3.2493	3.9695	**2.3307**	2.5091	6.8777	2.5684	4.2936	3.3139
		P4	21.5940	5.5589	6.9320	**3.8352**	5.9078	19.9931	6.0877	9.4089	7.6073
		P5	28.7990	6.8667	8.9802	**5.2633**	6.9909	29.4111	8.4786	12.4831	9.8113
	0.99	P3	51.1616	14.8469	16.0517	13.1154	13.5494	29.1057	**10.4073**	20.4184	14.7440
		P4	112.583	27.4557	31.4111	**18.2555**	32.4423	93.4984	30.8212	49.4065	34.3072
		P5	139.995	30.1438	37.5924	**22.3143**	36.3197	125.311	35.8197	59.6002	39.1094

		P3	2.2642	0.9261	1.4152	**0.8510**	0.8589	4.4711	0.9870	1.2523	1.0312
		P4	3.0729	1.1828	1.7313	**0.7969**	0.8413	7.0600	1.2098	1.6311	1.0258
		P5	5.0440	1.6044	2.3690	**1.1760**	1.2152	10.1655	1.9391	2.5519	1.7429
	0.8	P6	7.2446	1.9445	3.0431	**1.2420**	1.5566	13.5424	2.7503	3.4616	2.4517
		P3	4.1361	1.3787	1.9861	**1.0633**	1.3002	4.8238	1.4420	1.9447	1.6783
		P4	6.2659	2.0261	2.8236	**1.3620**	1.4329	8.2506	1.8382	2.9689	1.9375
		P5	10.2050	2.7862	3.9426	**2.2180**	2.3544	13.5358	3.2415	4.7539	3.6131
70	0.9	P6	14.6311	3.4994	5.0923	**2.1444**	2.8953	19.0866	4.2545	6.6134	4.5341
		P3	8.6273	2.5826	3.3307	**1.9202**	2.3896	6.8469	2.5479	3.6586	3.0848
		P4	13.2264	3.8898	5.0430	**2.6938**	2.8508	12.4850	3.4892	5.8005	3.9922
		P5	21.4089	5.3822	7.1031	**4.2611**	5.3253	22.2454	6.3741	9.8633	7.4715
	0.95	P6	38.3056	7.7671	9.7615	**5.0326**	10.6950	40.2202	14.6099	13.9551	14.7786
		P3	45.6046	11.9755	13.1080	**10.5775**	12.6906	28.5947	10.2081	17.2596	13.8085
		P4	71.2235	18.9457	22.1628	**12.6500**	15.1726	55.6095	16.6507	30.9732	17.9173
		P5	116.284	26.8902	32.2734	**20.5964**	37.8054	105.429	32.5831	54.4174	38.0870
	0.99	P6	167.394	36.2466	46.0637	**24.4990**	42.8970	156.495	50.1987	80.0121	48.6794

Bold values show the smallest MSE

Table 2 shows the estimated MSE values when n is 80 and 90 for explanatory variables, p = 3, 4, 5 and 6. It can be observed that the LDK with the biasing parameter k of the Median (MED) version gives the lowest MSE values in almost all the designs used. In the only six cases where the biasing parameter k of the MED version is not the lowest MSE, the Arithmetic Mean (AM) or Maximum (MAX) version takes the lowest.

Table 2: Estimated MSE for different estimator when p=3, 4, 5 and 6 when n= 80 and 90

n	ρ	P	MLE	RIDGE	DWD	DWDMED	DWDAM	DWDHM	DWDMAX	DWDMIN	DWDMR
		P3	1.67084	0.76002	1.17827	0.68708	**0.65797**	3.92575	0.84528	0.98687	0.77798
		P4	2.37561	0.94544	1.42684	**0.61893**	0.64444	6.42256	1.14272	1.30217	0.86454
		P5	4.56249	1.48556	2.30453	1.13672	**1.0484**	9.54568	1.80814	2.3741	1.5589
	0.8	P6	6.10864	1.7288	2.69452	**1.23503**	1.29881	12.0151	2.45743	2.9857	1.97318
		P3	3.28215	1.25322	1.8403	**0.90998**	1.0815	4.1846	1.37606	1.70816	1.43604
		P4	4.92192	1.6199	2.32145	**1.01588**	1.1329	7.20872	1.61942	2.39717	1.63462
		P5	9.5979	2.78954	4.0644	2.12481	**2.02427**	12.8622	3.18842	4.77824	3.14244
80	0.9	P6	12.8205	3.17826	4.71354	**2.28173**	2.78894	17.0417	4.27894	6.1909	4.19413
		P3	6.70392	2.24802	3.00094	**1.49985**	1.93836	5.57572	2.33735	3.13699	2.69762
		P4	10.4078	3.01482	4.02146	**2.03269**	2.33279	10.3204	2.74061	4.65749	3.30158
		P5	19.7217	5.24449	7.30642	**4.02324**	4.42273	20.4511	5.88366	9.55326	6.37405
	0.95	P6	25.501	5.78902	8.27669	**4.30208**	6.32236	27.6567	8.2777	12.1418	9.05972
		P3	34.9579	10.2086	11.3736	8.98314	9.06533	21.5403	**8.74356**	14.1798	11.4622
		P4	59.8215	15.456	18.3358	**9.92249**	13.7838	46.8038	12.8531	25.2849	16.3161
	0.99	P5	107.684	26.3674	33.759	**19.8468**	27.6923	96.1537	31.3508	52.3113	32.5601

		P6	142.28	30.1585	39.2711	**24.6125**	43.358	133.81	46.9836	70.5871	45.2701
		P3	1.49818	0.70953	1.09467	0.63742	**0.61847**	3.85657	0.79869	0.92377	0.71497
	0.8	P4	2.26256	0.88341	1.40847	**0.64595**	0.69797	5.90918	1.01898	1.23439	0.84665
		P5	3.47151	1.19339	1.88429	**0.91177**	0.84869	8.56464	1.43314	1.8101	1.25234
		P6	4.93237	1.46581	2.30555	**1.03609**	1.1637	10.9625	1.99174	2.3975	1.64359
		P3	2.95288	1.14763	1.70848	**0.84454**	0.94364	4.01652	1.19596	1.53553	1.22847
	0.9	P4	4.68027	1.48584	2.23311	**1.05472**	1.21078	6.76912	1.63985	2.18753	1.61355
		P5	7.20916	2.08718	3.16389	**1.64908**	1.73209	10.5566	2.25829	3.4854	2.50726
90		P6	10.3385	2.66288	3.98108	**1.77071**	2.0069	14.5726	3.31374	4.8147	3.23471
		P3	5.93709	1.97881	2.66166	**1.34554**	1.61218	5.1497	1.95366	2.71272	2.23391
	0.95	P4	9.93969	2.80057	3.79945	**1.96502**	2.34297	9.96151	2.87208	4.31996	3.23358
		P5	15.2227	3.93143	5.61191	**3.35878**	3.75497	16.3295	4.21628	7.03583	5.11609
		P6	22.2637	5.35722	7.58602	**3.70923**	4.75993	24.255	6.58373	10.2753	7.04778
		P3	31.2156	9.06655	10.0829	7.87813	7.87485	18.6596	**7.00095**	12.5249	9.53534
	0.99	P4	56.133	14.7771	16.8065	**10.3468**	13.4994	44.3914	14.1579	24.9554	15.3007
		P5	86.3689	20.8267	26.1321	**16.6688**	24.3271	76.622	22.8095	41.2024	26.5591
		P6	127.411	28.0873	36.6302	**20.5124**	33.104	119.176	40.3713	61.642	39.953

Bold values show the smallest MSE

Table 3 shows the estimated MSE values when n is 100 and 150 for explanatory variables, p = 3, 4, 5, and 6. It can be observed that the LDK with the biasing parameter k of the Median (MED) version gives the lowest MSE values in almost all the designs used. In the only seven cases where the biasing parameter k of the MED version does not have the lowest MSE, the Arithmetic Mean (AM) or Maximum (MAX) version takes the lowest MSE values.

Table 3: Estimated MSE for different estimator when p=3, 4, 5 and 6 when n=100 and 150

n	ρ	P	MLE	RIDGE	DWD	DWDMED	DWDAM	DWDHM	DWDMAX	DWDMIN	DWDMR
		P3	1.10667	0.57641	0.92229	0.59703	**0.51189**	3.47516	0.67127	0.7264	0.54209
		P4	2.04769	0.85039	1.36836	**0.56686**	0.65772	5.63264	0.99689	1.14864	0.84464
		P5	3.18344	1.14312	1.82142	0.7652	**0.73325**	7.77339	1.24737	1.69627	1.16373
	0.8	P6	3.38548	1.20059	1.93054	0.71406	**0.68376**	9.40583	1.46008	1.81556	1.05236
		P3	2.13936	0.8779	1.39011	**0.71609**	0.71637	3.44448	0.91839	1.19915	0.89358
100		P4	4.04429	1.38248	2.09127	**0.86116**	1.00724	6.10775	1.46213	2.01459	1.4555
		P5	6.42167	1.93073	2.97477	**1.30502**	1.36996	9.382	1.99169	3.15922	2.34011
	0.9	P6	6.7475	1.99696	3.19688	**1.17718**	1.28795	11.2886	2.21201	3.45088	2.06967
		P3	4.5131	1.58155	2.19985	**1.13836**	1.27518	4.14274	1.52954	2.17144	1.6557
		P4	8.27104	2.51889	3.53271	**1.73128**	1.97484	8.40246	2.38231	3.86988	2.80587
		P5	13.8134	3.77986	5.41429	**2.86362**	3.03591	14.515	3.59937	6.50331	4.75187
	0.95	P6	14.1998	3.73794	5.65766	**2.4425**	2.83459	16.5814	4.01551	6.95609	4.42935

		P3	23.4325	6.65367	7.43936	5.76364	**5.00067**	13.8058	5.23017	9.05736	6.18347
		P4	43.6483	11.4788	14.0823	**8.17882**	10.6092	34.006	9.82906	19.3253	12.4966
		P5	75.7412	19.0236	24.6698	**14.576**	19.3033	66.097	18.3414	35.6377	23.7271
	0.99	P6	83.9213	20.5012	28.3234	**14.2543**	20.0539	78.0145	23.5789	42.9646	26.0933
		P3	0.82151	0.47392	0.7799	0.48763	**0.42367**	2.74849	0.53576	0.59676	0.45538
	0.8	P4	1.15912	0.59309	0.97813	**0.40149**	0.43629	4.52427	0.6575	0.75689	0.54479
		P5	1.58944	0.66214	1.20284	**0.59098**	0.55942	6.25743	0.91058	0.91821	0.68144
		P6	2.12487	0.79073	1.50726	**0.6085**	0.63646	7.61557	1.1329	1.16051	0.86887
		P3	1.53221	0.68931	1.17824	0.61379	**0.60818**	2.75719	0.75969	0.92177	0.70542
	0.9	P4	2.34108	0.94283	1.56592	**0.58713**	0.70447	4.55324	0.9752	1.30587	0.98549
		P5	3.39555	1.1314	1.944	**0.91522**	0.92921	6.68573	1.3572	1.72528	1.28829
150		P6	4.5645	1.38477	2.46174	**1.03214**	1.20082	8.79154	1.83033	2.27762	1.76696
		P3	3.04433	1.11379	1.70868	**0.85217**	0.96199	3.15954	1.13891	1.51311	1.21765
	0.95	P4	4.72117	1.56457	2.49307	**1.02429**	1.23206	5.55093	1.6333	2.35545	1.78623
		P5	7.15526	2.07029	3.26311	**1.75067**	1.80586	8.86719	2.3168	3.41488	2.53422
		P6	9.47384	2.53167	4.03438	**1.94517**	2.32271	11.9861	3.07958	4.6558	3.38184
		P3	16.6319	4.83053	5.59182	4.2063	**3.57767**	9.91765	3.8345	6.50948	4.31519
	0.99	P4	26.9865	7.52902	9.9394	**5.14879**	6.64162	21.3561	6.87033	12.0961	8.34484
		P5	41.1357	10.3477	13.7712	**8.29434**	10.3638	36.7184	11.9716	19.8814	12.6076
		P6	51.6935	12.2463	17.4252	**9.66809**	14.0596	48.3635	15.0132	26.1452	17.3009

Bold values show the smallest MSE

Table 4 shows the estimated MSE values when n is 200 and 250 for explanatory variables, p = 3, 4, 5, and 6. it can be observed that as there is increase in sample sizes with a low multicollinearity strength of 0.8, the LDK with the biasing parameter k of the Arithmetic mean (AM) has the minimum MSE values compared with the rest of the estimators proposed and compared with. Also, at other multicollinearity levels, the median (MED) version gives the lowest MSE values in almost all the designs used. Except in five cases, the biasing parameter k of the arithmetic mean (AM) has the lowest MSE values again.

Table 4: Estimated MSE for different estimator when p=3, 4, 5 and 6 when n=200 and 250

n	ρ	P	MLE	RIDGE	DWD	DWDMED	DWDAM	DWDHM	DWDMAX	DWDMIN	DWDMR
		P3	0.5562	0.3562	0.5861	0.3475	**0.3104**	2.2064	0.4326	0.4236	0.3491
		P4	0.7832	0.4369	0.7596	0.3465	**0.3402**	3.9929	0.5193	0.5475	0.4012
		P5	1.2874	0.6164	1.076	0.4883	**0.4447**	5.6277	0.7891	0.8014	0.6237
	0.8	P6	1.7794	0.7238	1.332	**0.5272**	0.5634	6.8464	1.0012	0.9999	0.799
200		P3	1.0757	0.5552	0.9481	0.4746	**0.4417**	2.1763	0.608	0.7094	0.5339
		P4	1.5963	0.7012	1.2151	**0.456**	0.51	3.8229	0.7905	0.9297	0.6893
		P5	2.6508	1.0189	1.7239	0.7305	**0.7049**	5.7017	1.1385	1.4642	1.0186
	0.9	P6	3.7422	1.2098	2.1293	**0.8633**	0.9656	7.5765	1.5594	1.8758	1.4429
	0.95	P3	2.2299	0.9329	1.4868	**0.6802**	0.7709	2.3972	0.9502	1.242	1.0196

		P4	3.2122	1.1137	1.8735	**0.7343**	0.9427	4.4117	1.3039	1.5946	1.3526
		P5	5.4369	1.7511	2.7792	1.3962	**1.3156**	7.0486	1.7782	2.7187	1.8951
		P6	7.6774	2.1229	3.4816	**1.5849**	1.991	10.013	2.6419	3.7625	2.9002
		P3	11.981	3.6901	4.5526	**2.8014**	3.1929	7.3608	3.2207	5.0733	3.9902
		P4	17.616	4.7227	6.4307	**3.2996**	4.5012	14.479	4.4991	7.8549	5.6858
		P5	28.752	7.7927	10.618	**6.461**	6.6636	25.17	7.0923	13.952	8.2015
	0.99	P6	42.839	10.455	14.348	**8.5635**	11.77	39.935	12.521	21.937	14.331
		P3	0.3876	0.2656	0.4527	0.2951	**0.2438**	1.823	0.337	0.3099	0.2717
	0.8	P4	0.6778	0.387	0.7016	0.334	**0.3228**	3.4348	0.4507	0.4739	0.3794
		P5	0.9369	0.4849	0.8739	0.4232	**0.3702**	4.8709	0.5763	0.6142	0.4546
		P6	1.2133	0.5914	1.0345	0.4018	**0.3943**	6.4796	0.7983	0.752	0.5615
		P3	0.7641	0.417	0.7802	0.4361	**0.3629**	1.9291	0.508	0.5275	0.4376
	0.9	P4	1.3725	0.6266	1.126	**0.4675**	0.5291	3.4313	0.7332	0.8176	0.6747
		P5	1.9906	0.8217	1.4847	0.6062	**0.576**	4.8904	0.9479	1.1346	0.8067
250		P6	2.4794	0.9498	1.6524	**0.602**	0.653	6.6238	1.146	1.3607	1.0132
		P3	1.548	0.6611	1.1944	**0.562**	0.5761	2.0985	0.7935	0.8633	0.7454
	0.95	P4	2.771	1.0205	1.7289	**0.7352**	0.8293	3.8792	1.1399	1.4375	1.1565
		P5	3.9976	1.3629	2.3283	**1.0193**	1.0252	5.702	1.5567	2.0644	1.6168
		P6	5.091	1.6225	2.6839	**1.0383**	1.1809	7.8837	1.8657	2.6139	1.9725
		P3	8.5376	2.5856	3.4005	**2.012**	2.2543	5.5918	2.3966	3.5737	2.8497
	0.99	P4	14	3.8519	5.3812	**2.8961**	3.9074	11.503	3.8124	6.2268	4.9171
		P5	21.286	5.7531	8.2829	**4.3694**	5.525	19.132	6.1618	10.417	7.7491
		P6	28.087	7.323	10.763	**5.4302**	7.6465	26.466	7.9873	14.237	10.057

Bold values show the smallest MSE

3.2 Numerical example

Pena *et al*. [28] examined the impact of temperature, pH, and soluble solids concentration on the chance of Alicyclobacillus development in apple juice using a logistic model. The eigenvalues of the matrix are 13464.7990, 1715.9257, 56.5515, and 3.5445. Consequently, multicollinearity is present in the model, as shown by the condition index (C.I.) of 61.6342.

When there is multicollinearity, the ML estimator performs the least well, as expected. The choice of the biasing parameters k and d 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.

Table5: 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	21.35138842
$\hat{\beta}_{LRR}$	-2.4E-06	0.008038	-0.02442	0.015783	-0.01186	0.28340673
$\hat{\beta}_{LDK}$	7.244206	-1.74152	0.005895	-0.042	0.160006	21.57368811

$\hat{\beta}_{LDKMED}$	3.801875	0.438059	-0.03186	0.006407	-0.39032	5.782491876
$\hat{\beta}_{LDKMN}$	-4.75037	1.571575	-0.05868	0.086835	-0.33312	9.513392712
$\hat{\beta}_{LDKMX}$	7.244195	-1.74077	0.005875	-0.04191	0.159524	21.57256675
$\hat{\beta}_{LDKMR}$	7.240503	-1.57812	0.002386	-0.03006	0.076265	21.35343462
$\hat{\beta}_{LDKHM}$	7.245876	-1.85107	0.01198	-0.06703	0.259096	21.89852536
$\hat{\beta}_{LDKAM}$	7.222206	-1.28265	-0.00251	-0.02318	-0.03271	20.95427864

4. Conclusion

Based on the work of Kibra *et al*. [20], where real-world data and Monte Carlo simulation studies were utilized to examine the estimator performance, we were able to suggest a few LDK estimators in this paper for estimating the biasing parameter k. The estimators' performances were assessed using the Mean Squared Error (MSE) criteria. In the simulation study, it was observed that nearly all sample sizes that were taken into account and that the biasing parameter k with the Arithmetic mean version exhibits the lowest MSE values when the strength of multicollinearity is at 0.8. Additionally, at the remaining design used in this paper, the biasing parameter k with the Median (MED) version has the least, with the exception in some cases. In addition, from the numerical example, the biasing parameter k with the Median (MED) and the LRR have the two lowest MSEs, respectively. Hence, based on our findings, both in simulation and numerical examples, we thereby recommend to practitioners, researchers, and scientists that when faced with multicollinearity issues in using the logistic model, they should use the LDK estimator with the biasing k of the Arithmetic version when the multicollinearity level is not severe, but in severe multicollinearity cases, they should go with the Median (MED) version.

References

[1] Frisch, R. Statistical Confluence Analysis by Means of Complete Regression Systems; University Institute of Economics: Oslo, Norway, 1934.

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

[3] Hoerl, A. E., & Kennard, R. W. Ridge regression: Application to non-orthogonal problems. Technometrics, 12(1) 1970:69–82.

[4] Gibbons, D. G. A simulation study of some ridge estimators. *Journal of the American Statistical Association,* 76 ,1981:131–139.

[5] Lawless, J. F., & Wang, P. A simulation study of ridge and other regression estimators. *Communications in Statistics: Theory and Methods*, 5, 1976:307–323

[6] Dempster, A. P., Schatzoff, M., & Wermuth, N. A simulation study of alternatives to ordinary least squares. *Journal of the American Statistical Association,* 72,1977: 77–91.

[7] Hoerl, A.E., Kennard, R.W., & Baldwin, K.F. Ridge regression: Some simulation. *Communications in Statistics: Theory and Methods.* 4: 1975:105–123.

[8] McDonald, G. C., & Galarneau, D. I. A Monte Carlo evaluation of some ridge-type estimators. *Journal of the American Statistical Association*, 70,1975: 407–416.

[9] Alkhamisi, M. A., Khalaf, G., & Shukur, G. Some modifications for Choosing Ridge Parameter. *Communications in Statistics: Theory and Methods*, 35, 2006:1–16.

[10] Alkhamisi, M., & Shukur, G. Developing ridge parameters for SUR model. *Communications in Statistics: Theory and Methods*, 37, 2008:544–564.

[11] Muniz, G., & Kibria, B. M. G.. On some ridge regression estimators: An empirical compari

sons. *Communications in Statistics: Simulation and Computation*, 38,2009: 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*, 39, 2010:1639–1670

[14] Lukman A.F.&Ayinde K. Review and classifications of the ridge parameter estimation techniques. Hacet J Math Stat 46(5) 2017:953–967

[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*. 9, 2018:2838–2852.

[16] Schaefer R.L., Roi L.D., &Wolfe R.A. A ridge logistic estimator, *Communications in Statistics — Theory and Methods*, 13(1) 1984: 99–113.

[17] Schaefer R. L. Alternative regression collinear estimators in logistic when the data are col linear. *Journal of Statistical Computation and Simulation* 25, 1986:75–91.

[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 regression estimators. *Computational Economics* 40 (4) 2012:401–14.

[20] Kibria, B.M.G. (2003). Performance of some new ridge regression estimators, Communications in Statistics — *Simulation and Computation,* 32(2)2003:419–435.

[21] Afzal, N. & Amanullah, M. Dawoud–Kibria Estimator for the Logistic Regression Model: Method, Simulation and Application. *Iranian Journal of Science and Technology Transition Science,* 46, 2022:1483–1493

[22] Newhouse J. P., & Oman S. D. An evaluation ofridge estimators. *Rand Corporation.* P-716-PR, 1971: 1–28.

[23] 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*, *44*(2),2020: 437–443.

[24] Oladapo, O. J., Owolabi,A.T, Idowu,J.I, & Ayinde K. (2022). A New Modified Liu Ridge-Type Estimator for the Linear Regression Model : Simulation and Application. *Clinical Biostatistics and Biometrics.* 8(2):1–14

[25] Oladapo, O.J., & Idowu, J.I., Owolabi, A.T., & Ayinde, Kayode. A New Biased Two-Parameter Estimator in Linear Regression Model. EQUATIONS. 3, 2023: 73-92.

[26] Idowu, J. I., Oladapo, O. J., Owolabi, A.T., Ayinde, K. and Akinmoju, O. Combating multicollinearity: A new two-parameter approach, *Nicel Bilimler Dergisi*, 5(2), 2023: 90-116.

[27] 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.* 11(2) 2022:499–512.

[28] 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* 35 (4)2011:509–17.

Contribution of Individual Authors to the Creation of a Scientific Article

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article