

An Improved Beta Ridge-Type Estimator For Regression Problem

¹Dawodu G. A., ²Oyelere O. M., ³Ariyo O. S., and ⁴Ogunsola I. A.*

^{1,2,3,4}Department of Statistics, Federal University of Agriculture, Abeokuta, Nigeria

Corresponding Author's Email: ogunsolaia@funaab.edu.ng

Received on: April, 2024

Revised and Accepted on: May, 2024

Published on: July, 2024

ABSTRACT

The shrinkage-based estimators have been shown to be effective in regression problems such as multicollinearity which voids the independently identically distributed (IID) assumption on which most regression analysis is based. Estimation based on minimizing the sum of squares is not satisfactory because it is practically impossible to interpret the regression coefficient estimates optimally. The ridge parameter and other shrinkage parameters can be applied to improve the efficiency of the estimators when the explanatory variables are too related. We proposed a modified estimator for the beta regression by augmenting the ridge parameter to optimize industrial and traditional processes, modelling the dependence of a continuous random variable that assumes values in the standard unit interval $[0,1]$, called the beta modified ridge-type estimator (BMRT) to cushion the effect of multicollinearity. Finally, simulation and real-life data are used to show the advantages of the proposed estimator over similar existing ones.

Keywords: Shrinkage-based, Multicollinearity, Beta regression; Ridge parameter and Modified ridge-type regression.

1.0 INTRODUCTION

The Beta ridge regression (BRR) model is a type of shrinkage-based estimator, and it is used when practitioners find the need to model the beta distributed continuous variable that is affected by the problem of collinearity among the explanatory variables. Qasim *et al.*, (2021) introduced a BRR to combat multicollinearity when the dependent variable is explained as such that assumes values in the interval $(0,1)$. The modelling of events with this unique property in the presence of multicollinearity will void the independently identically distributed (IID) assumption but it is more powerful in real-life contexts. If multicollinearity occurs in a given dataset, the obtained maximum likelihood (ML) estimators are usually unstable with a high variance resulting in low statistical significance. Hence, Multicollinearity can cause a great deal of misunderstanding in studies with two or more explanatory variables if not properly addressed. In addition, most design matrices tends towards singularity in presence of multicollinearity. As such, X and the respective $X'X$ matrix will not be of full rank. In cases like this, it is impossible to optimally identify the ordinary least squares (OLS) estimate. Therefore, the presence of multicollinearity among explanatory variable needs to be taken into consideration since extracting the greatest possible amount of information from the dataset at hand is the optimum goal of carrying out analysis. As such, it is required from the onset to test for multicollinearity amongst explanatory variables since relationships among independent variables is

unavoidable in most real-life cases. Traditional methods are credibly limited in predicting the compressive strength of concrete, as minimum effect that has no or negligible impact on the coefficients when performing RR as investigated by Chopra *et al.* (2013).

A lot of methods have been proposed to fix the problem of multicollinearity as it has received increased attention over the years, recently some estimators were introduced to combat the problem of multicollinearity in beta regression, estimators such as beta ridge regression estimator (Qasim *et al.*, 2021) and Liu estimator for beta regression (Karlsson *et al.*, 2020). Jeremia *et al.* (2020), obtained a Robust Ridge estimator by combining the Ridge method and robust regression. Yahya (2019) examined the performance of inverse Gaussian ridge regression estimators. Cribari-Neto (2023) adopted beta regression n analys COVID-19 death cases. However, this study follow the improvements proposed by Lukman *et al.*, 2019 in obtaining an estimator to better cushion the effects of multicollinearity in beta regression. Based on the study and review of existing estimators by Kibria and Lukman (2020), Muniz *et al.*, 2012 proposed new estimators for obtaining the optimum shrinkage parameter k , particularly recommended when dealing with models with high residual variances, the researchers studied the investment decisions making, considering the presence of financial limitations which are mostly correlated. This study made an improvement on the existing estimator in handling multicollinearity in regression.

The organization of this paper follows: proposal and derivation of properties of the beta modified ridge-type estimator in Section 2. Theoretical comparison in Section 3 and discussion of the shrinkage parameter in Section 4. Simulation study to investigate the performance of the proposed estimator in Section 5. Real-life data was evaluated in Section 6 and we gave a conclusion in Section 7.

2.0 MATERIALS AND METHODS

2.1. The Beta model

Suppose y_1, y_2, \dots, y_n are outcome of the random variable, $Y \sim \beta_e(a, b)$ with parameters $a, b > 0$ and with probability density function;

$$f(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}, y \in (0,1) \quad (1)$$

where $\Gamma(\cdot)$ is the gamma function and $a, b > 0$, with $\mu = \frac{a}{(a+b)}$ and $V(Y) = \frac{ab}{\{(a+b)^2(a+b+1)\}}$.

Ferrari and Cribari-Neto, (2004) then proposed the new parametrization of the density function as given as follows.

$$\ell(\beta) = \sum_{i=1}^n \ell_i(\mu_i, \vartheta) = \sum_{i=1}^n \left[\log \Gamma(\vartheta) - \log \Gamma(\mu_i \vartheta) - \log \Gamma(\vartheta - \mu_i \vartheta) + (\vartheta \mu_i - 1) \log y_i + (\vartheta - \vartheta \mu_i - 1) \log(1 - y_i) \right] \quad (4)$$

Fisher's scoring iterative procedure within the Newton Raphson methods for estimating parameter vector β as:

$$\beta_{m+1} = \beta_m + \left\{ \frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_k} (\beta_m) \right\}^{-1} \delta(\beta_m), \quad (5)$$

where $\delta(\beta_m)$ is the first derivative and $m = 0, 1, 2, \dots$ are iterations which are carried out until convergence. The second derivative w.r.t β_k becomes:

$$\frac{\partial^2 \ell(\beta)}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^n \vartheta^{-1} \left(\frac{\partial \ell(\beta)}{\partial \beta_k} \right) \left(\frac{y_i - \mu_i}{V(\mu_i)} \frac{d\mu_i}{d\eta_i} x_{ji} \right) = \sum_{i=1}^n \vartheta^{-1} \left[\left\{ V(\mu_i)^{-1} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \right\} - (\mu_i - y_i) \times \left\{ V(\mu_i)^{-2} \left(\frac{d\mu_i}{d\eta_i} \right)^2 \frac{\partial V(\mu_i)}{\partial \mu_i} - V(\mu_i)^{-1} \left(\frac{\partial^2 V(\mu_i)}{\partial \eta_i^2} \right) \right\} \right] x_{ji} x_{ki}. \quad (6)$$

With the final estimation algorithm defined as:

$$\beta_m = (X^t W_m X)^{-1} X^t W_m z_m, \quad (7)$$

where $W_m = \text{diag}[\{g(\mu_i)\}^2 V(\mu_i) \vartheta]^{-1}$, $g(\cdot)$ is the logit link function and $z_m = \log(\mu_i) + \frac{y_i - \mu_i}{\mu_i(1 - \mu_i)}$. Since β_m converges to β_{ML} as $m \rightarrow \infty$, giving a MLE of the form;

$$\hat{\beta}_{ML} = (X^t W X)^{-1} X^t W z \quad (8)$$

where W and z are estimated at the final iteration. The variance of the MLE equals:

$$V(\hat{\beta}_{ML}) = \phi(X^t W X)^{-1} \quad (9)$$

Suppose $\mu = a/(a+b)$ and $\vartheta = a+b$, then the pdf becomes;

$$f(y; \mu, \vartheta) =$$

$$\frac{\Gamma(\vartheta)}{\Gamma(\mu\vartheta)\Gamma(\vartheta-\mu\vartheta)} y^{\mu\vartheta-1} (1-y)^{\vartheta-\mu\vartheta-1}, y \in (0,1), \quad (2)$$

where $Y \sim \beta_e(\mu, \vartheta)$ in the interval $(0,1)$, $\mu(0 < \mu < 1)$ and precision parameter $\vartheta(\vartheta > 0)$, with reparametrized mean and variance $E(Y) = \mu$ and $V(Y) = \mu(1-\mu)/(1+\phi)$ ϕ is called the dispersion parameter ($\phi = \vartheta^{-1}$). The BRM is obtained using the link function: $g(\mu_i) = \eta_i = x_i^t \beta$ where x_i is the i^{th} row $n \times p$ matrix, $\beta = (\beta_1, \beta_2, \dots, \beta_p)^t$ is a $p \times 1$ vector of the unknown regression coefficients, the link predictor η_i and $g(\cdot)$ is the link function of the BRM.

Using the Logit link function:

$$g(\mu_i) = \log \left(\frac{\mu_i}{1-\mu_i} \right),$$

$$\mu_i = \frac{\exp(x_i^t \beta)}{1 + \exp(x_i^t \beta)}. \quad (3)$$

where μ_i is the mean response function. Since η depends on β and the mean response μ is a function of η , the means $\mu_1, \mu_2, \dots, \mu_n$ are functions of β . The log-likelihood function is expressed as:

The MSE:

$$MSE(\hat{\beta}_{ML}) = E(\hat{\beta}_{ML} - \beta)^t(\hat{\beta}_{ML} - \beta) = \text{trace}\{\varphi(X^tWX)^{-1}\} = \varphi \sum_{j=1}^p \frac{1}{\lambda_j} \quad (10)$$

2.2 The BRR Estimator

The beta ridge regression estimator $\hat{\beta}_{BRR}$ was defined by Qasim et al. (2020) in generalized form as;

$$\hat{\beta}_{BRR} = (X^tWX + KI)^{-1}X^tWX \hat{\beta}_{ML} \quad (11)$$

$$V(\hat{\beta}_{BRR}) = \varphi[(X^tWX + KI)^{-1}X^tWX(X^tWX + KI)^{-1}] \quad (12)$$

$$\text{Bias}(\hat{\beta}_{BRR}) = E(\hat{\beta}_{BRR}) - \beta = -k(X^tWX + KI)^{-1}\beta \quad (13)$$

$$\begin{aligned} MSE(\hat{\beta}_{BRR}) &= \text{trace}[\varphi\{X^tWX + KI)^{-1}X^tWX(X^tWX + KI)^{-1}\}] + K^2\beta^t(X^tWX + KI)^{-2}\beta \\ &= \varphi \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j+k)^2} + \sum_{j=1}^p \frac{k^2\gamma_j^2}{(\lambda_j+k)^2} \end{aligned} \quad (14)$$

where $\lambda_j(\lambda_1, \dots, \lambda_p) > 0$ is the eigen values of the matrix X^tWX γ_j^2 is the j th element of $\xi^t\beta_{ML}$ and ξ is the matrix of eigenvectors X^tWX , i.e., $\xi = (\xi_1, \dots, \xi_p)$ and ξ_j is the j th eigenvector of X^tWX , such that $X^tWX = \xi\Lambda\xi^t$ and $\Lambda = \text{diag}(\lambda_j)$

2.3 The Proposed Estimator

In this section, we adopted the modified ridge type estimator (MRT) as developed by Lukman et al. (2019), into the beta regression model. The proposed beta modified ridge-type estimator is defined as follows;

$$\hat{\beta}_{BMRT}(Kd) = (X^tWX + K(1+d)I)^{-1}X^tWX\hat{\beta}_{ML} = \mathfrak{R}_{Kd}\hat{\beta}_{ML} \quad (15)$$

where $\mathfrak{R}_{Kd} = X^tWX + K(1+d)I)^{-1}X^tWX$.

$$\begin{aligned} \text{Bias}(\hat{\beta}_{BMRT}) &= E(\hat{\beta}_{BMRT}) - \hat{\beta}_{ML} \\ &= [(X^tWX + K(1+d)I)^{-1}X^tWX - I]\hat{\beta}_{ML} \\ &= (X^tWX + K(1+d)I)^{-1}[X^tWX - (X^tWX + K(1+d)I)]\hat{\beta}_{ML} \\ &= -K(1+d)G_K\hat{\beta}_{ML}. \end{aligned} \quad (16)$$

where $G_K = (X^tWX + K(1+d)I)^{-1}$.

Therefore,

$$\begin{aligned} MSE(\hat{\beta}_{BMRT}) &= \text{tr}[V(\hat{\beta}_{BMRT}) + \text{Bias}(\hat{\beta}_{BMRT})\text{Bias}(\hat{\beta}_{BMRT})'] \\ &= \varphi \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i+(K(1+d)))^2} + K(1+d)^2 \sum_{i=1}^p \frac{\beta_i^2}{(\lambda_i+(K(1+d)))^2} \end{aligned}$$

2.4 MSE Comparison

2.4.1 BMRT vs BML

Considering biased estimators using the MSE criterion as a measure of performance, the $\hat{\beta}_{ML}$ and $\hat{\beta}_{BMRT}(K, d)$. If $K > 0, 0 < d < 1$, the $\hat{\beta}_{BMRT}(K, d)$ performs better than its corresponding $\hat{\beta}_{ML}$ if;

$$\begin{aligned} MSE[\hat{\beta}_{ML}] - MSE[\hat{\beta}_{BMRT}] &> 0, \text{ i.e.,} \\ &= \varphi \sum_{i=1}^p \frac{1}{\lambda_i} - \varphi \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i+(K(1+d)))^2} + (K(1+d))^2 \sum_{i=1}^p \frac{\beta_i^2}{(\lambda_i+(K(1+d)))^2} > 0 \end{aligned} \quad (17)$$

For all $K > 0, 0 < d < 1$.

$$\frac{\partial MSE(\hat{\beta}_{BMRT})}{\partial K(1+d)} = -2 \sum_{i=1}^p \frac{\varphi\lambda_i + K\lambda_i\beta_i^2}{(\lambda_i+(K(1+d)))^3} > 0 \quad (18)$$

2.4.2 BMRT vs BRR

The proposed estimator is being compared to the beta ridge regression (BRR) estimator. The MSE of the BRR is defined as;

$$MSE(\hat{\beta}_{BRR}) = \varphi \sum_{i=1}^p \frac{\lambda_i}{(\lambda_i + K)^2} + K^2 \sum_{i=1}^p \frac{\beta_i^2}{(\lambda_i + K)^2} \quad (19)$$

The dispersion matrix $D(\hat{\beta}_{BRR})$ is expressed as;

$$D(\hat{\beta}_{BRR}) = \varphi^2 (\Lambda + KI)^{-1} \Lambda (\Lambda + KI)^{-1} \quad (20)$$

The superiority of $\hat{\beta}_{BMRT}(K, d)$ over $\hat{\beta}_{BRR}$ is justified as follows;

$$D[\hat{\beta}_{BRR}] - D[\hat{\beta}_{BMRT}] > 0 \quad (21)$$

$$= \varphi [\Lambda (\Lambda + KdI)^{-1} \Lambda (\Lambda + KdI)^{-1} - (\Lambda + K(1+d)I)^{-1} \Lambda (\Lambda + K(1+d)I)^{-1}]$$

$$D = \varphi (\Lambda + KI)^{-1} (\Lambda + K(1+d)I)^{-1} [Kd(2\Lambda + 2K + Kd)] (\Lambda + K(1+d)I)^{-1} (\Lambda + KI)^{-1} \quad (22)$$

It can be observed that for $K > 0, 0 < d < 1$ D is positive finite

2.5 Selection of parameters k and d

The biasing parameters for the proposed BMRT estimator was obtained following the work of Lukman et al., (2019) k and d can be obtained iteratively as follows;

- Obtain an initial estimate of d using

$$\hat{d} = \min \left(\frac{\varphi}{\hat{v}_i^2} \right), \text{ where, } \varphi = (1/(n-p)) \sum_{i=1}^n ((y_i - \hat{\theta}_i)^2 / \theta_i^2).$$

- Obtain K_{LWP} using from 1 : $K_{LWP} = \frac{p\varphi}{\sum \lambda_j(1+d_{mrt}) - \beta_i^2}$

- Estimate $\hat{d}_{BMRT} = \frac{\hat{\beta}_{\max} - 1}{\lambda_{\max}^{-1} + \hat{\beta}_{\max}^2}$

- In case \hat{d}_{MRT} is not between 0 and 1, use $\hat{d}_{MRT} = \hat{d}$

2.6 Numerical Evaluation

2.6.1 Performance Criteria

This section evaluates numerically, the behaviour of the proposed estimator when the explanatory variables are highly linearly inter-correlated. the performance of the proposed BMRT is evaluated under various conditions during simulation and real-life data using the MSE criteria.

2.6.2 Monte Carlo Simulation

The numerical evaluations were obtained following the beta regression model;

$$\log \left(\frac{\mu_i}{1-\mu_i} \right) = (\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip})^{-1}, i = 1, 2, \dots, n, j = 1, 2, \dots, p. \quad (23)$$

where x_{ij} , β_j are the covariates and regression coefficients respectively. $\sum_{j=1}^q \beta_j^2 = 1$ (Kibria and Lukman (2020)) when the intercept is considered to be zero. The beta regression model is defined using the logit link function and the response variable are generated as $y_i \sim \beta_e(\mu_i, \varphi)$, the correlated independent variables were generated as follows;

$$\rho^2)^{(1/2)} \omega_{ij} + \rho \omega_{i(j+1)} \quad (24)$$

where, ρ is the level of correlation between the explanatory variables expressed as ρ^2 , ω_{ij} are the independent standard normal pseudorandom numbers. The initial estimates for β_j were obtained from the OLS estimate i.e. $(X^t X)^{-1} X^t y^*$, where $y^* = \{g(y_1), g(y_2), \dots, g(y_n)\}^t$

The behaviour of the proposed estimator was observed under varying correlation coefficients among the explanatory variables ($\rho^2 = 0.70, 0.80, 0.90$), the number of independent variables ($p = 4, 8$) and sample sizes ($n = 20, 50$).

3.0 RESULTS AND DISCUSSION

3.1 Simulation Study

The results from simulation is provided in Tables 1 and 2, $n = 20, 50$ respectively. The results show that the proposed BMRT estimator outperforms existing estimators. It was observed that the higher the correlation coefficient and sample size, the higher the mean square error supporting theoretical claims. Furthermore, it was observed generally that the shrinkage-based estimators outperform the MLE in the presence of multicollinearity.

Table 1: Estimated MSEs of the estimators when n=20

$n=20$		$p=4$				$n=20$	$p=8$	
k	d	φ	OLS	RE	MRT	OLS	RE	MRT
				$\rho = 0.7$			$\rho = 0.7$	
0.3	0.2	1	61.782	2.445	2.1	19.053	5.013	4.354
		5	308.912	12.226	10.498	95.265	25.067	21.768
		15	926.735	36.678	31.494	285.795	75.2	65.303
		25	1544.56	61.131	52.49	476.325	125.333	08.838
0.7	0.5	1	61.782	1.38	1.177	19.053	2.484	1.756
		5	308.912	6.898	5.886	95.265	12.422	8.78
		15	926.735	20.694	17.659	285.795	37.264	26.34
		25	1544.56	34.49	29.431	476.325	62.107	43.9
0.9	0.8	1	61.782	1.241	1.057	19.053	1.999	1.269
		5	308.912	6.206	5.284	95.265	9.993	6.344
		15	926.735	18.617	15.853	285.795	29.98	19.031
		25	1544.56	31.028	26.421	476.325	49.966	31.718
				$\rho = 0.8$			$\rho = 0.8$	
0.3	0.2	1	89.413	2.192	1.902	27.261	5.315	4.531
		5	447.067	10.961	9.508	136.305	26.578	22.654
		15	1341.2	32.884	28.524	408.914	79.732	67.961
		25	2235.33	54.807	47.539	681.523	132.887	113.269
0.7	0.5	1	89.413	1.31	1.143	27.261	2.442	1.698
		5	447.067	6.549	5.715	136.305	12.208	8.488
		15	1341.2	19.646	17.144	408.914	36.625	25.465
		25	2235.33	32.743	28.573	681.523	61.042	42.442
0.9	0.8	1	89.413	1.196	1.041	27.261	1.94	1.232
		5	447.067	5.98	5.204	136.305	9.7	6.162
		15	1341.2	17.94	15.612	408.914	29.099	18.485
		25	2235.33	29.899	26.02	681.523	48.499	30.809
				$\rho = 0.9$			$\rho = 0.9$	
0.3	0.2	1	175.219	1.824	1.627	52.441	5.299	4.373
		5	876.097	9.12	8.135	262.205	26.495	21.868
		15	2628.29	27.361	24.406	786.616	79.486	65.599
		25	4380.49	45.602	40.677	1311.03	132.477	109.332
0.7	0.5	1	175.219	1.224	1.102	52.441	2.178	1.513
		5	876.097	6.121	5.509	262.205	10.888	7.563
		15	2628.29	18.362	16.526	786.616	32.665	22.689
		25	4380.49	30.603	27.543	1311.03	54.441	37.814
0.9	0.8	1	175.219	1.142	1.021	52.441	1.721	1.14
		5	876.097	5.709	5.105	262.205	8.604	5.701
		15	2628.29	17.126	15.314	786.616	25.812	17.104
		25	4380.49	28.543	25.523	1311.03	43.019	28.507

Table 2: Estimated Mean SEs of the estimators when n=50

$n=50$		$p=4$				$n=50$	$p=8$	
K	D	φ	OLS	RE	MRT	OLS	RE	MRT
				$\rho = 0.7$			$\rho = 0.7$	
0.3	0.2	1	20.718	2.752	2.305	3.862	2.805	2.654
		5	103.59	13.761	11.523	19.311	14.025	13.271

		15	310.77	41.284	34.57	57.933	42.075	39.813
		25	517.96	68.806	57.616	96.554	70.125	66.355
0.7	0.5	1	20.718	1.324	1.064	3.862	2.025	1.627
		5	103.59	6.621	5.318	19.311	10.124	8.137
		15	310.77	19.862	15.954	57.933	30.371	24.411
		25	517.96	33.103	26.59	96.554	50.618	40.685
0.9	0.8	1	20.718	1.143	0.93	3.862	1.776	1.253
		5	103.59	5.713	4.649	19.311	8.879	6.264
		15	310.77	17.138	13.948	57.933	26.636	18.792
		25	517.96	28.564	23.247	96.554	44.393	31.321
				$\rho = 0.8$			$\rho = 0.8$	
0.3	0.2	1	29.911	2.518	2.095	5.524	3.551	3.305
		5	149.55	12.589	10.477	27.62	17.776	16.525
		15	448.66	37.768	31.431	82.86	53.327	49.574
		25	747.77	62.946	52.386	138.1	88.878	82.624
0.7	0.5	1	29.911	1.234	1.022	5.524	2.339	1.793
		5	149.55	6.169	5.112	27.62	11.695	8.961
		15	448.66	18.506	15.337	82.86	35.085	26.907
		25	747.77	30.843	25.562	96.554	58.474	44.846
0.9	0.8	1	29.911	1.086	0.917	5.524	1.991	1.328
		5	149.55	5.428	4.585	27.62	9.957	6.638
		15	448.66	16.284	13.756	82.86	29.871	19.914
		25	747.77	27.14	22.927	96.554	49.786	33.19
				$\rho = 0.9$			$\rho = 0.9$	
0.3	0.2	1	58.36	2.022	1.702	10.625	4.945	4.414
		5	291.8	10.108	8.511	53.125	24.726	22.071
		15	875.41	30.323	25.534	159.37	74.177	66.213
		25	1459	50.538	42.556	265.62	123.63	110.36
0.7	0.5	1	58.36	1.107	0.972	10.625	2.677	1.897
		5	291.8	5.533	4.86	53.125	13.384	9.483
		15	875.41	16.598	14.581	159.37	40.152	28.45
		25	1459	27.664	24.301	265.62	66.92	47.417
0.9	0.8	1	58.36	1.012	0.905	10.625	2.163	1.34
		5	291.8	5.06	4.523	53.125	10.817	6.697
		15	875.41	15.179	13.568	159.37	32.452	20.09
		25	1459	28.299	22.613	265.62	54.087	33.484

3.2 An Empirical Application

An investigation of the effect of some explanatory variables on the response variable using a dataset consisting of 32 observations on the proportion of crude oil converted to gasoline after fractional distillation was recently used by Karlsson, Mansson and Kibria [5]. The independent variables, firstly, oil gravity (x_1), as measured by the American Petroleum Institute Index by evaluating the density of a liquid. Secondly, the vapour pressure (x_2) exerted by the crude oil using the Reid vapour pressure defined as the pressure required to suppress a liquid from vaporizing at $100^\circ F$. finally, the temperature $^\circ F$ at which 10 percent of the crude oil evaporated (x_3).

The presence of multicollinearity among the explanatory variables was confirmed in Table 3 with condition index of the independent variables equals 83.3 and the eigen values ($\lambda_1 = 1957272$, $\lambda_2 = 3709.846$, $\lambda_3 = 75.255$).

Table 3: The correlation matrix of the explanatory variable of the gasoline yield data

Variables	X1	X2	X3
X1	1		
X2	0.62	1	
X3	-0.7	-0.91	1

The results of the mean square error of the regression

coefficient are in Table 4.

The estimated values of K and d are 0.2385617, 0.9999774 respectively. The proposed estimator BMRT has been shown to outperform existing estimators using MSE criteria.

Table 4: Estimated MSE of Regression coefficients

Variable Name	ML	BRE	BMRT
Gravity	-0.0021	-0.0021	-0.0021
Pressure	0.1505	0.15046	0.15041
Tempten	0.00416	0.00416	0.00416
MSE	0.20039	0.19915	0.19792

4.0 CONCLUSION

We introduced a beta ridge regression (BMRT) estimator in this study. Analytically, we demonstrate that the novel estimator outperforms the industry standard MLE and BRR. The shrinkage parameters were then determined using the derived MSE and earlier studies for the linear model. To compare the performance of the suggested estimators under different circumstances, including the quantity of explanatory variables, degree of correlation, and sample size, a simulation study has been carried out. It was observed that our proposed estimator performed better than some existing estimators using the MSE criteria and therefore recommended for practitioners

ACKNOWLEDGEMENTS

The authors thanked the Management of FUNAAB for providing an enabling environment for carrying out this research.

REFERENCES

- Alobaidi, N. N., Shamany, R. E., and Algama, Z. Y. (2021). A new ridge estimator for the negative binomial regression model. *Thailand Statistician*, 19(1), 115-124. <https://ph02.tci-thaijo.org/index.php/thaistat/article/view/242818>.
- Chopra, P., Sharma, R. K., and Kumar, M. (2013). Ridge regression for the prediction of compressive strength of concrete. *Int. J. Innov. Eng. Technol*, 2, 106-111. <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=727579f5d70a5a79384e966d5c842b0b8bd1e761>.
- Ferrari, S., and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7), 799-815. <https://doi.org/10.1080/0266476042000214501>.
- Cribari-Neto, F. (2023). A beta regression analysis of COVID-19 mortality in Brazil. *Infectious Disease Modelling*, 8(2), 309-317. <https://doi.org/10.1016/j.idm.2023.02.005>.

- Jeremia, N. E., Nurrohmah, S., and Fithriani, I. (2020). Robust Ridge regression to solve a multicollinearity and outlier. In *Journal of Physics: Conference Series*. 1442(1), p. 012030. IOP Publishing. <https://doi.org/10.1088/1742-6596/1442/1/012030>.
- Karlsson, P., Månsson, K., and Kibria, B. G. (2020). A Liu estimator for the beta regression model and its application to chemical data. *Journal of Chemometrics*, 34(10), e3300. <https://doi.org/10.1002/cem.3300>.
- Kibria, B. G., and Lukman, A. F. (2020). A new ridge-type estimator for the linear regression model: simulations and applications. *Scientifica*, 2020(1), 9758378. <https://doi.org/10.1155/2020/9758378>.
- Lukman, A. F., Ayinde, K., Binuomote, S., and Clement, O. A. (2019). Modified ridge-type estimator to combat multicollinearity: Application to chemical data. *Journal of Chemometrics*, 33(5), e3125. <https://doi.org/10.1002/cem.3125>.
- Muniz, G., Kibria, B. M., and Shukur, G. (2012). On developing ridge regression parameters: a graphical investigation. https://digitalcommons.fiu.edu/math_fac/10.
- Qasim, M., Kibria, B. M. G., Månsson, K., and Sjölander, P. (2020). A new Poisson Liu regression estimator: method and application. *Journal of Applied Statistics*, 47(12), 2258-2271. <https://doi.org/10.1080/02664763.2019.1707485>
- Qasim, M., Månsson, K., and Golam Kibria, B. M. (2021). On some beta ridge regression estimators: method, simulation and application. *Journal of Statistical Computation and Simulation*, 91(9), 1699-1712. <https://doi.org/10.1080/00949655.2020.1867549>.
- Yahya Algama, Z. (2019). Performance of ridge estimator in inverse Gaussian regression model. *Communications in Statistics-Theory and Methods*, 48(15), 3836-3849. <https://doi.org/10.1080/03610926.2018.1481977>.