frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 22 February 2022
doi: 10.3389/fams.2022.775068

Dawoud–Kibria Estimator for Beta Regression Model: Simulation and Application

Mohamed R. Abonazel[1]*, *Issam Dawoud*[2], *Fuad A. Awwad*[3] and *Adewale F. Lukman*[4]

[1] Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt, [2] Department of Mathematics, Al-Aqsa University, Gaza City, Palestine, [3] Department of Quantitative Analysis, College of Business Administration, King Saud University, Riyadh, Saudi Arabia, [4] Biostatistics and Epidemiology, University of Medical Sciences, Ondo City, Nigeria

The linear regression model becomes unsuitable when the response variable is expressed as percentages, proportions, and rates. The beta regression (BR) model is more appropriate for the variable of this form. The BR model uses the conventional maximum likelihood estimator (BML), and this estimator may not be efficient when the regressors are linearly dependent. The beta ridge estimator was suggested as an alternative to BML in the literature. In this study, we developed the Dawoud–Kibria estimator to handle multicollinearity in the BR model. The properties of the new estimator are derived. We compared the performance of the estimator with the existing estimators theoretically using the mean squared error criterion. A Monte Carlo simulation and a real-life application were carried out to show the benefits of the proposed estimator. The theoretical comparison, simulation, and real-life application results revealed the superiority of the proposed estimator.

Keywords: beta Kibria–Lukman estimator, beta Özkale–Kaçiranlar estimator, beta ridge estimator, maximum likelihood, mean square

OPEN ACCESS

Edited by:
Lixin Shen,
Syracuse University, United States

Reviewed by:
Xueying Zeng,
Ocean University of China, China
Erol Egrioglu,
Giresun University, Turkey
Zakariya Yahya Algamal,
University of Mosul, Iraq
Mahdi Roozbeh,
Semnan University, Iran

Correspondence:
Mohamed R. Abonazel
mabonazel@cu.edu.eg

Specialty section:
This article was submitted to Optimization,
a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *13 September 2021*
Accepted: *17 January 2022*
Published: *22 February 2022*

Citation:
Abonazel MR, Dawoud I, Awwad FA and Lukman AF (2022) Dawoud–Kibria Estimator for Beta Regression Model: Simulation and Application. Front. Appl. Math. Stat. 8:775068. doi: 10.3389/fams.2022.775068

INTRODUCTION

The linear regression (LR)model is used if the dependent variable follows a normal distribution. The assumption of the normality of the dependent variable may be violated and then it will fit some of the exponential family distributions as a negative binomial, Poisson, gamma, inverse Gaussian, and beta, so in this case, we use the generalized linear (GL) model instead of the LR model. The beta regression (BR) model is applied in many different fields such as engineering, medical sciences, physical sciences, social sciences, environment, and business if the dependent variable observations are between (0, 1). To estimate the BR model parameters, we use the maximum likelihood (ML) estimator which is more convenient than the ordinary least squares (OLS) estimator for describing and investigating different phenomena.

In the LR model, the explanatory variables may be correlated and this causes a problem called multicollinearity in which this problem may arise in the BR model. The ML estimator is the most popular used method for estimating the unknown regression parameters in the BR model. But also, in the existence of multicollinearity problems, the regression parameters' variances and standard errors are very large. To reduce the multicollinearity effect, different biased estimation methods are proposed and the most popular method is the ordinary ridge regression (ORR) estimation method which was proposed by Hoerl and Kennard [1, 2]. Another recent one parameter estimator proposed by Kibria and Lukman [3] to solve the multicollinearity is the Kibria and Lukman estimator. Also, in the case of an estimator with two parameters, Özkale and Kaçiranlar [4] proposed a two-parameter estimator. Very recently, Dawoud and Kibria [5] proposed a new kind of two-parameter estimator called the Dawoud–Kibria (DK) estimator. There are other recent studies regarding the one parameter and two-parameter estimators in LR and GL models, such as Roozbeh et al. [6], Lukman et al. [7], Arashi et al. [8], Farghali et al. [9], Lukman et al. [10, 11], Algamal and Abonazel [12], Akram et al. [13], and Abonazel et al. [14]. In this article, we drive the Dawoud–Kibria estimator for the BR model in the presence of the multicollinearity problem. Then, the properties of the Dawoud–Kibria estimator for the BR model are investigated.

This article is organized as follows. The methodology and the proposed estimator are given in section methodology. In section the superiority of the proposed estimator, the theoretical comparisons among the estimators are conducted. Section selection of biasing parameters k and d gives the proposed biasing parameters for the estimators. In sections Monte Carlo simulation study and real data application, the Monte Carlo simulation and the real-life dataset results are presented. Finally, in section conclusion, some conclusions of this article are given.

METHODOLOGY

In this section, we discuss the BR model. Then, the ridge, Kibria–Lukman, and Özkale–Kaçiranlar estimators are stated to the BR model. After that, we introduce the Dawoud–Kibria estimator for the BR model. Finally, the biasing parameters of the Dawoud–Kibria estimator for the BR model are proposed.

The BR Model

The BR model is popularly used in many different fields such as economics and medical studies. The BR model is used to show the effect of explanatory variables on a non-normal response variable as any generalized LR model. However, the response variable for the BR model is restricted to the interval (0, 1) as rates, proportions, and fractions. The BR model was given firstly by the authors Ferrari and Cribari-Neto [15] with relating the response variable mean function to linear predictors set through a link function. The BR model has a precision parameter where its reciprocal is determined as a dispersion measure [16, 17].

Let y be a continuous random variable having a beta distribution, then the probability density function of y is given as:

$$\mathrm{f}\left(y;\mu,\phi\right)=\frac{\Gamma\left(\phi\right)}{\Gamma\left(\mu\phi\right)\Gamma\left(\left(1-\mu\right)\phi\right)}y^{\mu\phi-1}\left(1-y\right)^{\left(1-\mu\right)\phi-1};\quad 0<y<1, 0<\mu<1, \phi>0, \tag{1}$$

where $\Gamma(\cdot)$ is called as the gamma function and ϕ is called as the precision parameter. The beta probability distribution mean and variance are as follows:

$$\mathrm{E}\left(y\right)=\mu, \mathrm{Var}\left(y\right)=\frac{Var\left(\mu\right)}{1+\phi}=\frac{\mu\left(1-\mu\right)}{1+\phi}.$$

Let $y_1,\ldots,y_n$ be independent random variables, where each y_i; $i=1,\ldots,n$ follows the density in Equation (1) with mean μ_i and unknown precision ϕ. The model is obtained by assuming that the mean of y_i can be written as:

$$\mathrm{g}\left(\mu_i\right)=\log\left(\frac{\mu_i}{1-\mu_i}\right)={x_i}'\beta=\eta_i, \tag{2}$$

where $\mathrm{g}(\cdot)$ is the used link function, $\beta=\left(\beta_1,\ldots,\beta_p\right)'$ is an $\left(p\times 1\right)$ unknown parameters vector, $x_i=\left(x_{i1},\ldots,x_{ip}\right)'$ is the vector of p regressors, and η_i is the linear predictor.

Beta Maximum Likelihood Estimator

The BR parameters estimation is done using the beta maximum likelihood (BML) method [18]. The BR log-likelihood function is given as:

$$\mathcal{L}\left(\beta\right)=\sum_{i=1}^{n}\left\{\log\Gamma\left(\phi\right)-\log\Gamma\left(\mu_i\phi\right)-\log\Gamma\left(\left(1-\mu_i\right)\phi\right)+\left(\mu_i\phi-1\right)\log\left(y_i\right)+\left(\left(1-\mu_i\right)\phi-1\right)\log\left(1-y_i\right)\right\}. \tag{3}$$

Differentiating the log-likelihood given in Equation (3) with respect to the parameter β provides us the score function of the parameter β that is given as:

$$U\left(\beta\right)=\phi X'T\left(y^*-\mu^*\right), \tag{4}$$

where $T=\mathrm{diag}\left(\frac{1}{\mathrm{g}'(\mu_1)},\ldots,\frac{1}{\mathrm{g}'(\mu_n)}\right)$; with $\mathrm{g}'(\cdot)$ is the first derivative of $\mathrm{g}(\cdot)$; with $y_i^*=\log\left(\frac{y_i}{1-y_i}\right)$, and $\mu^*=\left(\mu_1^*,\ldots\mu_n^*\right)'$; with $\mu_i^*=\psi\left(\mu_i\phi\right)-\psi\left(\left(1-\mu_i\right)\phi\right)$, such that $\psi(\cdot)$ denoting the digamma function. The iterative reweighted least-squares (IRLS) algorithm or the Fisher scoring algorithm are used for estimating the parameter β [19, 20]. This algorithm form is given as:

$$\beta^{(r+1)}=\beta^{(r)}+\left(I_{\beta\beta}^{(r)}\right)^{-1}U_{\beta}^{(r)}\left(\beta\right), \tag{5}$$

where $U_{\beta}^{(r)}$ is called the score function, and $I_{\beta\beta}^{(r)}$ is called the information matrix for β, for more details, see Espinheira et al. [20]. With the use of the IRLS algorithm with initial values of β and ϕ as in Ferrari and Cribari-Neto [15] and Espinheira et al. [20], the BML estimator of the parameter β is provided as:

$$\hat{\beta}_{\mathrm{BML}}=\left(X'\hat{W}X\right)^{-1}X'\hat{W}z, \tag{6}$$

where X is an $\left(n\times p\right)$ design matrix, $z=\hat{\eta}+\hat{W}^{-1}\hat{T}\left(y^*-\hat{\mu}^*\right)$, and $\hat{W}=\mathrm{diag}\left(\hat{w}_1,\ldots,\hat{w}_n\right)$; with

$$\hat{w}_i=\hat{\phi}\left\{\psi'\left(\hat{\mu}_i\hat{\phi}\right)+\psi'\left(\left(1-\hat{\mu}_i\right)\hat{\phi}\right)\right\}\frac{1}{\left[\mathrm{g}'\left(\hat{\mu}_i\right)\right]^2}.$$

Here, $\hat{W}$, $\hat{T}$, $\hat{\mu}_i$, and $\hat{\mu}^*$ are the estimates of W, T, μ_i, and μ^*, respectively, evaluated at the ML estimator of β and ϕ [15].

Now, let $\Gamma=\mathrm{diag}\left(\gamma_1,\ldots,\gamma_p\right)=Q'X'\hat{W}XQ$, and $\alpha=\left(\alpha_1,\ldots,\alpha_p\right)'=Q'\beta$; where $\gamma_1\geq\ldots\geq\gamma_p\geq 0$ and Q is the matrix whose columns are the eigenvectors of the $\left(X'\hat{W}X\right)$ matrix. Then, the mean squared error matrix (MSEM) and the mean squared error (MSE) of an estimator $\tilde{\beta}$ are defined as follows:

$$MSEM(\tilde{\beta}) = Var(\tilde{\beta}) + \left(Bias(\tilde{\beta})\right)\left(Bias(\tilde{\beta})\right)', \quad (7)$$

$$MSE(\tilde{\beta}) = trace\left(MSEM(\tilde{\beta})\right). \quad (8)$$

Then the MSEM and MSE of $\hat{\beta}_{BML}$ are.

$$MSEM(\hat{\beta}_{BML}) = \frac{1}{\phi}\Gamma^{-1}, \quad (9)$$

$$MSE(\hat{\beta}_{BML}) = \frac{1}{\phi}\sum_{j=1}^{p}\frac{1}{\gamma_j}. \quad (10)$$

Beta Ridge Regression (BRR) Estimator

To reduce the effects of multicollinearity in the BR model, Abonazel and Taha [21] and Qasim et al. [22] introduced the BRR estimator as an alternative to the BML estimator and is given as:

$$\hat{\beta}_{BRR} = (X'\hat{W}X + k\,\mathrm{I}_p)^{-1}X'\hat{W}z, \ldots k > 0. \quad (11)$$

The MSEM and MSE of $\hat{\beta}_{BRR}$ are

$$MSEM(\hat{\beta}_{BRR}) = \frac{1}{\phi}UL^{-1}\Gamma L^{-1}U' + (UL^{-1}\Gamma U' - I_p)\alpha\alpha'(UL^{-1}\Gamma U' - I_p)', \quad (12)$$

$$MSE(\hat{\beta}_{BRR}) = \frac{1}{\phi}\sum_{j=1}^{p}\frac{\gamma_j}{L_j^2} + k^2\sum_{j=1}^{p}\frac{\alpha_j^2}{L_j^2} \quad (13)$$

where $L = (\Gamma + k\,I_p)$ and $L_j = (\gamma_j + k)$.

Beta Kibria–Lukman (BKL) Estimator

The BKL estimator is defined as follows:

$$\hat{\beta}_{BKL} = (X'\hat{W}X + k\,\mathrm{I}_p)^{-1}(X'\hat{W}X - k\,\mathrm{I}_p)\,\hat{\beta}_{BML}, \quad k > 0. \quad (14)$$

The MSEM and MSE of $\hat{\beta}_{BKL}$ are

$$MSEM(\hat{\beta}_{BKL}) = \frac{1}{\phi}UL^{-1}N\Gamma^{-1}NL^{-1}U' + (UL^{-1}NU' - I_p)\alpha\alpha'(UL^{-1}NU' - I_p)', \quad (15)$$

$$MSE(\hat{\beta}_{BKL}) = \frac{1}{\phi}\sum_{j=1}^{p}\frac{N_j^2}{\gamma_j L_j^2} + 4k^2\sum_{j=1}^{p}\frac{\alpha_j^2}{L_j^2} \quad (16)$$

where $N = (\Gamma - k\,I_p)$ and $N_j = (\gamma_j - k)$.

TABLE 1 | Simulated mean square error (SMSE) values of different estimators when $p = 2$ and $\phi = 2$.

n	ρ	BML	BRR	BKL	BOK	BDK(1)	BDK(2)
50	0.80	6.053	5.233	5.202	4.891	3.394	3.677
	0.85	7.303	6.118	6.069	5.503	3.395	3.838
	0.90	13.692	11.901	11.848	10.107	4.656	6.070
	0.95	31.207	26.815	26.691	19.919	4.923	8.081
	0.99	67.136	47.857	46.909	24.971	15.684	7.492
75	0.80	5.348	4.832	4.816	4.652	3.656	3.785
	0.85	7.008	6.210	6.185	5.792	3.933	4.325
	0.90	10.863	9.611	9.577	8.573	4.808	5.736
	0.95	18.291	14.916	14.788	11.584	3.846	5.274
	0.99	68.451	53.912	53.360	31.408	8.463	5.586
100	0.80	3.933	3.621	3.609	3.566	3.152	3.153
	0.85	9.107	8.289	8.271	7.700	5.047	5.730
	0.90	9.991	8.846	8.815	8.019	4.732	5.544
	0.95	15.744	13.466	13.396	11.168	4.514	6.142
	0.99	115.376	102.758	102.521	65.853	5.542	14.038
150	0.80	6.437	6.100	6.095	5.940	4.889	5.030
	0.85	6.972	6.518	6.510	6.286	4.929	5.129
	0.90	10.034	9.210	9.195	8.569	5.607	6.456
	0.95	18.945	17.151	17.119	14.781	6.782	9.339
	0.99	115.789	106.225	106.100	73.332	8.145	23.265
200	0.80	5.511	5.243	5.239	5.150	4.441	4.421
	0.85	6.501	6.162	6.157	5.999	4.966	5.051
	0.90	8.751	8.133	8.123	7.712	5.521	6.094
	0.95	16.097	14.810	14.791	13.240	7.235	9.253
	0.99	146.696	138.772	138.709	102.489	19.521	45.366

Beta Özkale–Kaçiranlar (BOK) Estimator

Recently, Abonazel et al. [14] proposed the BOK estimator as an extension of the Özkale and Kaçiranlar [4] estimator in the BR model and is defined as follows:

$$\hat{\beta}_{BOK} = (X'\,\hat{W}X + k\,\mathrm{I}_p)^{-1}(X'\hat{W}\,X + kd\mathrm{I}_p)\,\hat{\beta}_{BML},\quad k>0,\quad 0<d<1. \tag{17}$$

The MSEM and MSE of $\hat{\beta}_{BOK}$ are

$$MSEM(\hat{\beta}_{BOK}) = \frac{1}{\phi}UL^{-1}G\Gamma^{-1}GL^{-1}U' + (UL^{-1}GU'-I_p)\alpha\alpha'(UL^{-1}GU'-I_p)', \tag{18}$$

$$MSE(\hat{\beta}_{BOK}) = \frac{1}{\phi}\sum_{j=1}^{p}\frac{G_j^2}{\gamma_j L_j^2} + (1-d)^2k^2\sum_{j=1}^{p}\frac{\alpha_j^2}{L_j^2} \tag{19}$$

where $G=(\Gamma+kd\,I_p)$ and $G_j=(\gamma_j+kd)$.

The Proposed Estimator

Extensions of the two-parameter estimators to the area of GLMs have been recently developed; such as Qasim et al. [22], Farghali et al. [9], Lukman et al. [23], Algamal and Abonazel [12], and Abonazel et al. [14]. Following the previous works, we introduced the beta version of the two-parameter estimator of Dawoud and Kibria [5] (BDK) as follows:

$$\hat{\beta}_{BDK} = (X'\,\hat{W}\,X + k(1+d)\mathrm{I}_p)^{-1}(X'\,\hat{W}X - k\,(1+d)\mathrm{I}_p)\hat{\beta}_{BML},\quad k>0,\quad 0<d<1. \tag{20}$$

We give the MSEM of the proposed $\hat{\beta}_{BDK}$ as follows:

$$MSEM(\hat{\beta}_{BDK}) = \frac{1}{\phi}UM^{-1}R\Gamma^{-1}RM^{-1}U' + (UM^{-1}RU'-\mathrm{I}_p)\alpha\alpha'(UM^{-1}RU'-I_p)', \tag{21}$$

$$MSE(\hat{\beta}_{BDK}) = \frac{1}{\phi}\sum_{j=1}^{p}\frac{R_j^2}{\gamma_j M_j^2} + 4k^2(1+d)^2\sum_{j=1}^{p}\frac{\alpha_j^2}{M_j^2}, \tag{22}$$

where $M=(\Gamma+k(1+d)\,I_p)$, $R=(\Gamma-k(1+d)\,I_p)$, $M_j=(\gamma_j+k(1+d))$ and $R_j=(\gamma_j-k(1+d))$.

THE SUPERIORITY OF THE PROPOSED ESTIMATOR

Theorem 1: If $4k^2(1+d)^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2 < \sum_{j=1}^{p}\left(M_j^2-R_j^2\right)$, then $MSE\,(\hat{\beta}_{BDK}) < MSE\,(\hat{\beta}_{BML})$.

TABLE 2 | SMSE values of different estimators when $p=2$ and $\phi=6$.

n	ρ	BML	BRR	BKL	BOK	BDK(1)	BDK(2)
50	0.80	11.491	10.645	10.627	9.735	7.130	6.903
	0.85	13.247	12.228	12.206	11.035	7.716	7.654
	0.90	18.158	16.463	16.425	14.042	7.790	8.644
	0.95	35.625	32.457	32.400	25.909	10.109	13.399
	0.99	192.874	168.676	168.149	93.746	17.848	12.119
75	0.80	15.037	14.280	14.271	13.177	9.543	9.539
	0.85	17.553	16.759	16.751	15.364	10.642	11.255
	0.90	27.598	26.653	26.646	24.351	16.533	18.294
	0.95	59.231	56.845	56.829	47.709	21.969	30.398
	0.99	214.373	186.262	185.683	103.517	16.966	10.659
100	0.80	11.001	10.576	10.571	10.034	8.496	7.578
	0.85	14.828	14.055	14.046	13.017	9.301	9.558
	0.90	19.577	18.642	18.633	17.090	11.662	12.394
	0.95	43.621	41.238	41.215	34.604	15.749	21.251
	0.99	265.201	252.353	252.261	175.379	22.736	61.884
150	0.80	10.789	10.545	10.544	10.227	9.133	8.370
	0.85	12.325	11.800	11.794	11.136	8.899	8.413
	0.90	17.602	16.783	16.775	15.534	10.948	11.554
	0.95	46.605	45.329	45.324	40.763	25.761	31.100
	0.99	252.047	245.921	245.900	195.177	67.404	114.900
200	0.80	9.358	9.133	9.131	8.897	8.099	7.276
	0.85	15.565	15.263	15.262	14.763	12.608	12.707
	0.90	21.999	21.571	21.569	20.598	16.733	17.630
	0.95	28.486	27.156	27.146	24.343	14.704	17.431
	0.99	213.727	207.448	207.422	163.383	53.122	92.802

Proof: The MSE difference between the BML and the BDK estimators is written as

$$\Delta_1 = MSE\left(\hat{\beta}_{BDK}\right) - MSE\left(\hat{\beta}_{BML}\right) = \frac{1}{\phi}\sum_{j=1}^{p}\left[\frac{R_j^2 - M_j^2 + 4k^2(1+d)^2\gamma_j\phi\alpha_j^2}{\gamma_j M_j^2}\right]. \quad (23)$$

In the case of $R_j^2 - M_j^2 + 4k^2(1+d)^2\gamma_j\phi\alpha_j^2 < 0$ in the equation (23), it implies that $4k^2(1+d)^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2 < \sum_{j=1}^{p}\left(M_j^2 - R_j^2\right)$, then $MSE\left(\hat{\beta}_{BDK}\right) < MSE\left(\hat{\beta}_{BML}\right)$. That means the BDK estimator is better than the BML estimator if $4k^2(1+d)^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2 < \sum_{j=1}^{p}\left(M_j^2 - R_j^2\right)$.

Theorem 2: If $\sum_{j=1}^{p}\left(R_j^2L_j^2 - \gamma_j^2M_j^2\right) < k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2(M_j^2 - 4(1+d)^2L_j^2)$,
then $MSE\left(\hat{\beta}_{BDK}\right) < MSE\left(\hat{\beta}_{BRR}\right)$.

Proof: The MSE difference between the BRR and the BDK estimators is written as

$$\Delta_2 = MSE\left(\hat{\beta}_{BDK}\right) - MSE\left(\hat{\beta}_{BRR}\right) = \frac{1}{\phi}\sum_{j=1}^{p}\left[\frac{R_j^2L_j^2 - \gamma_j^2M_j^2 - k^2\phi\gamma_j\alpha_j^2(M_j^2 - 4(1+d)^2L_j^2)}{\gamma_j L_j^2 M_j^2}\right]. \quad (24)$$

In the case of $R_j^2L_j^2 - \gamma_j^2M_j^2 - k^2\phi\gamma_j\alpha_j^2(M_j^2 - 4(1+d)^2L_j^2) < 0$ in the Equation (24), it implies that $\sum_{j=1}^{p}\left(R_j^2L_j^2 - \gamma_j^2M_j^2\right) < k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2(M_j^2 - 4(1+d)^2L_j^2)$, then $MSE\left(\hat{\beta}_{BDK}\right) < MSE\left(\hat{\beta}_{BRR}\right)$. That means the BDK estimator is better than the BRR estimator if $\sum_{j=1}^{p}\left(R_j^2L_j^2 - \gamma_j^2M_j^2\right) < k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2(M_j^2 - 4(1+d)^2L_j^2)$.

Theorem 3: If $\sum_{j=1}^{p}\left(R_j^2L_j^2 - N_j^2M_j^2\right) < 4k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2(M_j^2 - (1+d)^2L_j^2)$.
then $MSE\left(\hat{\beta}_{BDK}\right) < MSE\left(\hat{\beta}_{BKL}\right)$.

TABLE 3 | SMSE values of different estimators when $p = 4$ and $\phi = 2$.

n	ρ	BML	BRR	BKL	BOK	BDK(1)	BDK(2)
50	0.80	8.618	7.044	6.999	6.275	3.469	3.367
	0.85	9.465	7.472	7.431	6.421	3.357	3.368
	0.90	16.331	13.295	13.254	10.640	4.493	4.428
	0.95	35.235	29.336	29.247	20.055	5.494	5.071
	0.99	271.828	224.818	223.712	92.905	25.607	41.613
75	0.80	8.023	7.017	7.003	6.611	4.308	4.009
	0.85	10.160	8.556	8.535	7.630	4.303	4.124
	0.90	17.399	14.616	14.573	12.058	5.093	4.947
	0.95	31.134	25.589	25.472	18.173	4.529	4.793
	0.99	187.813	161.921	161.693	81.603	6.541	17.808
100	0.80	7.523	6.513	6.494	6.149	4.096	4.124
	0.85	9.167	8.000	7.983	7.366	4.665	4.648
	0.90	19.593	17.076	17.026	14.578	6.429	6.697
	0.95	31.651	27.423	27.365	21.103	6.951	6.707
	0.99	217.675	194.858	194.677	112.730	6.734	13.765
150	0.80	7.066	6.454	6.448	6.285	4.672	4.599
	0.85	9.429	8.514	8.504	8.044	5.535	5.626
	0.90	14.144	12.871	12.861	11.751	7.393	7.420
	0.95	33.151	29.960	29.924	24.822	10.332	10.585
	0.99	178.793	161.444	161.275	101.082	9.202	10.705
200	0.80	7.135	6.584	6.578	6.434	4.908	5.053
	0.85	7.726	7.044	7.036	6.819	4.988	5.190
	0.90	13.439	12.248	12.236	11.275	7.226	7.363
	0.95	30.354	28.056	28.043	24.286	12.687	12.634
	0.99	201.543	188.253	188.153	133.641	21.328	24.457

Proof: The MSE difference between the BKL and the BDK estimators is written as

$$\Delta_3 = MSE\left(\hat{\beta}_{BDK}\right) - MSE\left(\hat{\beta}_{BKL}\right) = \frac{1}{\phi}\sum_{j=1}^{p}\left[\frac{R_j^2L_j^2 - N_j^2M_j^2 - 4k^2\phi\gamma_j\alpha_j^2\left(M_j^2-(1+d)^2L_j^2\right)}{\gamma_jL_j^2M_j^2}\right]. \quad (25)$$

In the case of $R_j^2L_j^2 - N_j^2M_j^2 - 4k^2\phi\gamma_j\alpha_j^2(M_j^2-(1+d)^2L_j^2) < 0$ in the Equation (25), it implies that $\sum_{j=1}^{p}\left(R_j^2L_j^2 - N_j^2M_j^2\right) < 4k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2(M_j^2-(1+d)^2L_j^2)$, then $MSE(\hat{\beta}_{BDK}) < MSE(\hat{\beta}_{BKL})$. That means the BDK estimator is better than the BKL estimator if $\sum_{j=1}^{p}\left(R_j^2L_j^2 - N_j^2M_j^2\right) < 4k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2(M_j^2-(1+d)^2L_j^2)$.

Theorem 4: If $\sum_{j=1}^{p}\left(R_j^2L_j^2 - G_j^2M_j^2\right) < k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2((1-d)^2M_j^2-4(1+d)^2L_j^2)$, then $MSE(\hat{\beta}_{BDK}) < MSE(\hat{\beta}_{BOK})$.

Proof: The MSE difference between the BOK and the BDK estimators is written as

$$\Delta_4 = MSE\left(\hat{\beta}_{BDK}\right) - MSE\left(\hat{\beta}_{BOK}\right) = \frac{1}{\phi}\sum_{j=1}^{p}\left[\frac{R_j^2L_j^2 - G_j^2M_j^2 - k^2\phi\gamma_j\alpha_j^2\left((1-d)^2M_j^2-4(1+d)^2L_j^2\right)}{\gamma_jL_j^2M_j^2}\right]. \quad (26)$$

In the case of $R_j^2L_j^2 - G_j^2M_j^2 - k^2\phi\gamma_j\alpha_j^2((1-d)^2M_j^2 - 4(1+d)^2L_j^2) < 0$ in the Equation (26), it implies that $\sum_{j=1}^{p}\left(R_j^2L_j^2 - G_j^2M_j^2\right) < k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2((1-d)^2M_j^2-4(1+d)^2L_j^2)$, then $MSE(\hat{\beta}_{BDK}) < MSE(\hat{\beta}_{BOK})$. That means the BDK estimator is better than the BOK estimator if $\sum_{j=1}^{p}\left(R_j^2L_j^2 - G_j^2M_j^2\right) < k^2\phi\sum_{j=1}^{p}\gamma_j\alpha_j^2((1-d)^2M_j^2-4(1+d)^2L_j^2)$.

SELECTION OF BIASING PARAMETERS *k* and *d*

We will suggest the following biasing parameters' estimators for the mentioned estimators.

TABLE 4 | SMSE values of different estimators when $p = 4$ and $\phi = 6$.

n	ρ	BML	BRR	BKL	BOK	BDK(1)	BDK(2)
50	0.80	20.276	19.341	19.333	17.555	11.620	11.975
	0.85	18.047	17.189	17.185	15.592	10.837	10.841
	0.90	27.687	25.923	25.917	22.257	12.368	12.520
	0.95	87.147	83.888	83.879	69.421	30.856	31.332
	0.99	603.570	577.427	577.263	357.585	36.519	43.205
75	0.80	14.407	13.659	13.653	12.829	9.262	9.390
	0.85	25.218	23.940	23.934	21.759	13.710	13.419
	0.90	38.729	37.702	37.702	34.643	24.505	26.369
	0.95	84.082	80.781	80.767	67.356	30.970	35.892
	0.99	472.368	460.962	460.940	338.777	90.885	87.179
100	0.80	16.825	16.274	16.272	15.473	11.865	13.190
	0.85	24.909	23.521	23.510	21.343	12.895	15.321
	0.90	30.072	28.630	28.622	25.641	15.512	17.823
	0.95	67.417	64.185	64.170	53.419	24.947	27.241
	0.99	658.305	644.268	644.236	479.563	123.607	135.271
150	0.80	13.969	13.573	13.572	13.099	10.726	11.658
	0.85	20.345	19.764	19.762	18.816	14.670	16.154
	0.90	33.991	33.263	33.261	31.344	23.972	26.417
	0.95	86.685	85.104	85.101	77.562	52.626	58.964
	0.99	535.740	526.202	526.187	422.009	160.031	192.949
200	0.80	17.816	17.532	17.532	17.071	14.878	15.815
	0.85	21.343	20.971	20.971	20.283	17.201	18.443
	0.90	36.550	35.967	35.966	34.245	27.566	29.896
	0.95	98.462	96.739	96.736	88.492	60.933	68.280
	0.99	516.039	507.722	507.706	413.974	171.268	199.438

Following Hoerl et al. [24] and Qasim et al. [22], $\hat{k}$ of the BRR estimator is written as

$$\hat{k}_{BRR} = \frac{p}{\hat{\phi}\sum_{j=1}^{p}\hat{\alpha}_j^2}, \quad (27)$$

where $\hat{\alpha}_j$ is the jth element of $\hat{\alpha} = Q'\hat{\beta}_{\text{BML}}$ vector and $\hat{\phi}$ is the ML estimate of ϕ [15].

- Following Lukman et al. [25], $\hat{k}_{BKL}$ of the BKL estimator is written as

$$\hat{k}_{BKL} = \frac{p}{\hat{\phi}\sum_{j=1}^{p}\left(\frac{1}{\hat{\phi}\gamma_j} + 2\hat{\alpha}_j^2\right)} \quad (28)$$

- Following Özkale and Kaçiranlar [4] and Abonazel et al. [14], $\hat{k}_{BOK}$ and $\hat{d}_{BOK}$ of the BOK estimator are written as

$$\hat{d}_{BOK} = \min\left(\frac{\hat{\alpha}_j^2}{\frac{1}{\hat{\phi}\gamma_j} + \hat{\alpha}_j^2}\right)_{j=1}^{p} \quad (29)$$

$$\hat{k}_{BOK} = \left(\frac{p}{\hat{\phi}\sum_{j=1}^{p}\left(\hat{\alpha}_j^2 - \hat{d}_{BOK}\left(\frac{1}{\hat{\phi}\gamma_j} + \hat{\alpha}_j^2\right)\right)}\right)^{1/2} \quad (30)$$

- Following Dawoud and Kibria [5], we suggest two different $\hat{k}$ of the proposed BDK estimator as follows:

$$\hat{k}_{BDK(1)} = \left(\hat{k}_{BRR}\right)^{1/p} \quad (31)$$

$$\hat{k}_{BDK(2)} = \left(\frac{1}{p}\sum_{j=1}^{p}\frac{1}{\hat{\phi}\left(1 + \hat{d}_{BOK}\right)\left(\frac{1}{\hat{\phi}\gamma_j} + 2\hat{\alpha}_j^2\right)}\right)^{1/p} \quad (32)$$

MONTE CARLO SIMULATION STUDY

In this section, a Monte Carlo simulation study has been conducted to compare the performances of BML, BRR, BKL, and

TABLE 5 | SMSE values of different estimators when $p = 6$ and $\phi = 2$.

n	ρ	BML	BRR	BKL	BOK	BDK(1)	BDK(2)
50	0.80	11.105	9.078	9.057	7.810	4.655	3.876
	0.85	13.514	10.799	10.772	8.788	4.607	3.753
	0.90	25.321	20.354	20.247	14.685	5.540	4.364
	0.95	47.497	37.538	37.399	22.139	5.574	5.085
	0.99	264.803	213.947	212.957	76.134	20.147	57.276
75	0.80	7.189	5.922	5.914	5.500	3.811	3.354
	0.85	13.325	10.910	10.870	9.195	5.001	3.725
	0.90	19.350	16.098	16.061	12.725	6.394	4.503
	0.95	48.120	38.549	38.286	23.752	6.202	5.768
	0.99	243.478	204.310	203.653	83.874	10.836	41.688
100	0.80	9.013	7.461	7.427	6.761	4.419	3.543
	0.85	10.948	9.292	9.270	8.258	5.295	4.066
	0.90	16.251	13.651	13.624	11.390	6.264	4.555
	0.95	30.617	25.870	25.833	19.041	8.188	5.088
	0.99	231.060	201.644	201.293	96.261	8.733	20.457
150	0.80	7.345	6.380	6.365	6.074	4.492	3.810
	0.85	9.407	8.216	8.201	7.613	5.405	4.359
	0.90	15.455	13.587	13.566	11.983	7.540	5.237
	0.95	36.480	32.601	32.573	25.971	13.137	7.642
	0.99	228.210	207.108	206.955	119.693	20.897	11.495
200	0.80	7.730	6.957	6.950	6.700	5.232	4.412
	0.85	9.778	8.752	8.741	8.222	6.035	4.953
	0.90	12.742	11.311	11.302	10.195	7.161	5.536
	0.95	33.667	30.140	30.105	24.579	12.452	7.475
	0.99	219.889	204.152	204.074	131.491	37.952	10.424

BOK with the suggested estimator (BDK). The program of the simulation study is written in R programming language based on the *betareg* package.

The Design of the Experiment

We simulated the datasets with the following settings:

1) The response variable y_i is generated from the beta distribution as $Beta\ (\mu_i,\ \phi)$, where $\mu_i = \exp\left(x_i{}'\beta\right) / \left(1+\exp\left(x_i{}'\beta\right)\right)$; $i = 1, \ldots, n$, and x_i is the ith row of X. The precision parameter ϕ chosen in the simulation is $\phi = 2$ and 6.
2) Sample size: $n = 50, 75, 100, 150$, and 200.
3) Explanatory variables are generated with a degree of multicollinearity as in Kibria [26]: $x_{ij} = u_{ij}\sqrt{1-\rho^2} + \rho u_{ip}$, where u_{ij} are the independent standard uniform pseudorandom numbers, and ρ is defined as the correlation between the explanatory variables, $\rho = 0.80, 0.85, 0.90, 0.95$, and 0.99.
4) The number of explanatory variables is $p = 2$, 4, and 6; with $\beta'\beta = 1$ and $\beta_1 = \ldots = \beta_p$, as per Kaçiranlar and Dawoud [27], Rady et al. [28], Abonazel and Farghali [29], Farghali et al. [9], Dawoud and Abonazel [30], and Awwad et al. [31].
5) We used the simulated MSE (SMSE) criterion for verification, which are computed as

$$SMSE\left(\hat{\beta}\right) = \frac{1}{5000}\sum_{l=1}^{5000}\left(\hat{\beta}_l - \beta\right)'\left(\hat{\beta}_l - \beta\right), \quad (33)$$

where $\hat{\beta}_l$ is the estimated value vector at the lth experiment of the simulation, β is the true parameter vector. The number of replications is 5,000.

Simulation Results

We have the following comments according to the simulation results in **Tables 1–6**: Obviously, from **Tables 1–6**, the proposed estimator possesses a smaller SMSE than the BML estimator and other estimators understudy for all sample sizes. For instance, from **Table 3**, when $\rho = 0.9$, $n = 50$, the SMSE of BML is 16.331 while the SMSE for other estimators is as follows: 13.295 (BRR), 13.254 (BKL), 10.640 (BOK), 4.493 (BDK(1)), and 4.428 (BDK(2)), respectively. Similarly, when the values of ϕ increase the SMSE also increases: from **Table 1**, when $\phi = 2$, $n = 100$ and $\rho = 0.99$, and **Table 2**, when $\phi = 6$, $n = 100$ and $\rho = 0.99$, the SMSE of BRR rises from 102.758 to 252.353. Also, it is evident that the SMSE values of all the estimators increased as the number of explanatory p increased. For the one-parameter shrinkage estimator, the BKL estimator consistently dominates the BRR estimator. For two-parameter shrinkage estimators, the BDK estimator dominates

TABLE 6 | SMSE values of different estimators when $p = 6$ and $\phi = 6$.

n	ρ	BML	BRR	BKL	BOK	BDK(1)	BDK(2)
50	0.80	24.308	23.396	23.396	21.226	16.249	12.890
	0.85	32.188	30.588	30.577	26.475	17.289	12.522
	0.90	49.485	46.799	46.789	38.084	21.315	13.021
	0.95	101.291	97.077	97.075	76.197	39.597	22.003
	0.99	693.505	670.736	670.667	410.533	90.039	37.843
75	0.80	24.844	23.949	23.944	21.907	16.472	13.086
	0.85	34.987	33.539	33.532	29.804	20.256	16.453
	0.90	47.139	44.266	44.241	36.283	19.387	12.870
	0.95	93.080	89.724	89.717	73.289	41.415	24.254
	0.99	754.197	735.126	735.082	496.580	123.219	36.798
100	0.80	17.825	17.177	17.175	16.127	12.728	11.537
	0.85	21.441	20.510	20.506	18.863	13.433	11.505
	0.90	49.162	47.478	47.473	42.140	27.865	22.938
	0.95	103.039	99.905	99.896	85.178	47.090	35.859
	0.99	634.956	619.106	619.078	444.692	140.691	59.963
150	0.80	17.301	16.801	16.800	16.101	13.051	13.019
	0.85	32.077	31.463	31.463	30.007	24.389	23.897
	0.90	42.623	41.614	41.612	38.705	28.606	27.190
	0.95	98.059	95.272	95.263	83.046	49.071	42.380
	0.99	724.422	714.279	714.269	583.521	270.776	188.774
200	0.80	19.214	18.795	18.794	18.181	15.345	15.624
	0.85	26.161	25.717	25.716	24.805	21.034	21.107
	0.90	47.018	46.298	46.297	44.058	35.750	35.533
	0.95	108.860	107.407	107.405	99.800	75.129	71.876
	0.99	652.981	645.213	645.206	547.813	293.726	233.134

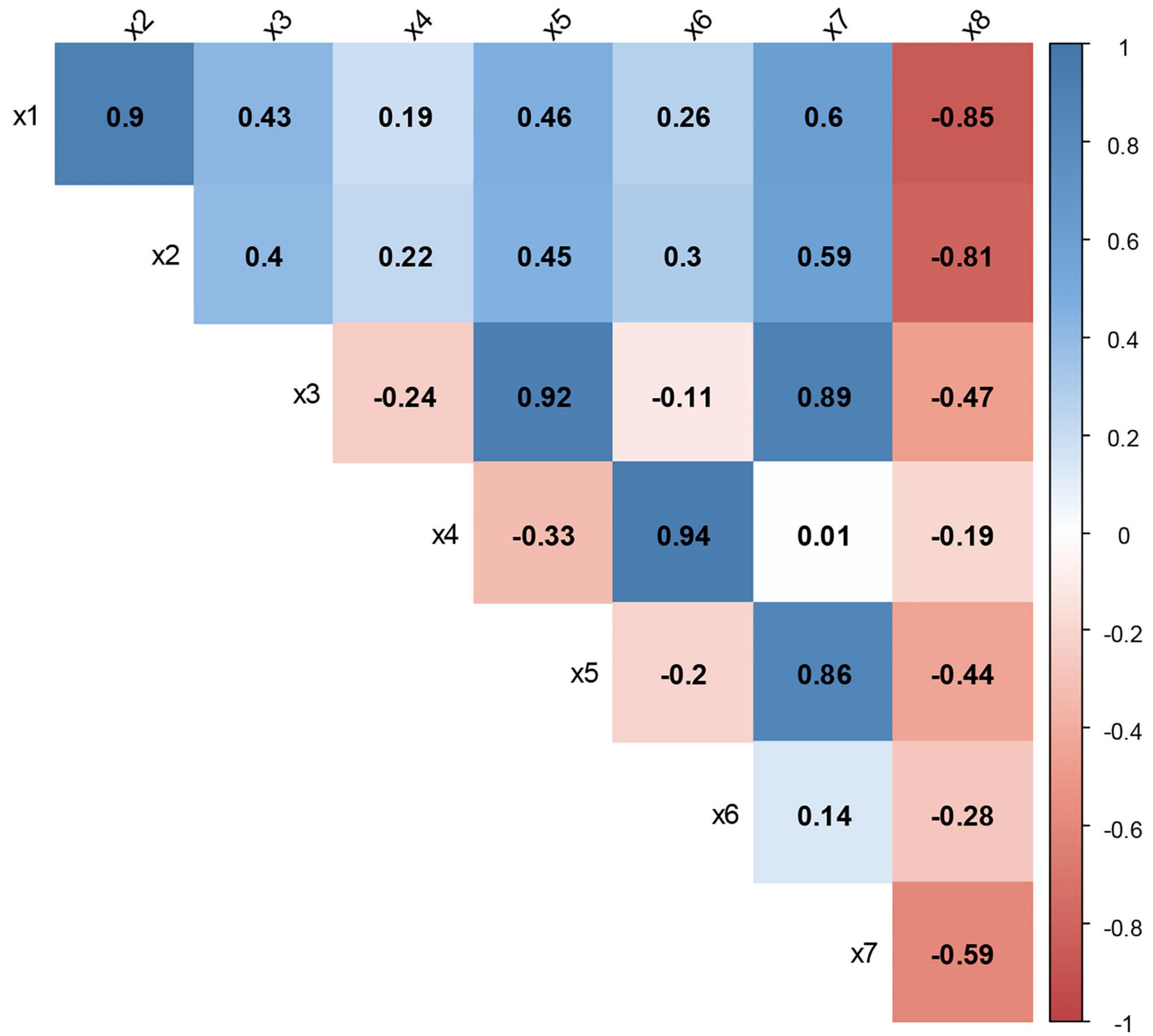

FIGURE 1 | Visualization of the correlation matrix.

the BOK estimator. Overall, the BDK dominates both the one-parameter and the two-parameter estimators. However, the performance of each estimator is a function of the employed shrinkage parameter.

REAL DATA APPLICATION

The implementation of the proposed estimator is illustrated by a study applied to the well-being index of Turkey in 2015 [32]. The index involves the aspects of accommodation, jobs, income and wealth, health, education, climate, protection, public engagement and access to community resources and social life. As the life satisfaction index is between 0 and 1. The values close to 1 refer to a better standard of living. The data are obtained from the Turkish Statistics Association. The original dataset consists of some dimensions that are represented by 41 indicators. Here, we are interested in only nine indicators used by Abonazel and Taha [21] and the number of observations is 50. The response variable is the level of happiness and eight explanatory variables are x1: Number of rooms per person, x2: Average point of necessary placement scores of the system for transition to secondary education from basic education, x3: Satisfaction rate with public education services, x4: Percentage of the population receiving waste services, x5: Satisfaction rate with public safety services, x6: The access rate of the population to sewerage and pipe system, x7: Satisfaction rate with public health services, and x8: Percentage of households declaring to fail on meeting basic needs.

TABLE 7 | Estimation results for the used estimators.

	BML	BRR	BKL	BOK	BDK(1)	BDK(2)
x1	−0.4269	−0.4028	−0.4022	−0.3942	−0.3584	0.2425
x2	0.0014	0.0012	0.0012	0.0012	0.0010	−0.0022
x3	0.0017	0.0018	0.0018	0.0018	0.0020	0.0048
x4	−0.0019	−0.0019	−0.0019	−0.0019	−0.0019	−0.0021
x5	−0.0076	−0.0076	−0.0076	−0.0077	−0.0077	−0.0085
x6	−0.0044	−0.0044	−0.0044	−0.0044	−0.0044	−0.0038
x7	0.0270	0.0269	0.0269	0.0269	0.0266	0.0229
x8	−0.0095	−0.0093	−0.0093	−0.0092	−0.0088	−0.0033
k	–	0.4997	0.2489	0.7182	0.7069	29.3690
d	–	–	–	0.0308	0.0308	0.0308
MSE	0.00138	0.00123	0.00122	0.00117	0.00097	0.00047
R^2	0.752	0.779	0.780	0.789	0.825	0.915
GCV	–	74.714	74.707	74.617	73.795	73.250

To investigate the multicollinearity through correlation coefficients between the explanatory variables, a visualization of the correlation matrix of the variables is constructed with the corresponding coefficients reported in **Figure 1**. The correlation coefficients indicate that there are strong relationships (more than 0.8) between some explanatory variables. This denotes the severe multicollinearity presence. Moreover, this conclusion is confirmed by the variance inflation factor (VIF) and the condition number $\left(\mathrm{CN} = \sqrt{\max(\gamma_j)/\min(\gamma_j)}\right)$ [33]; where the VIFs of the eight explanatory variables are 7.5, 6.1, 10.8, 10.1, 9.1, 9.8, 9.7, and 4.3, respectively, and the CN is 3,936.055.

Table 7 provides the regression parameter estimates for the BR model using BML, BRR, BKL, BOK, and BDK. From **Table 7**, it can note that the estimated regression parameters of all estimators have the same signs (except x1 and x2 in BDK(2) only); this means that the type of relationship between each explanatory variable and the response variable is not changed from what it was in the BML. The estimated MSE of the five estimators were obtained by Equations (10), (13), (16), (19), and (22), respectively. The results of **Table 7** indicate that the estimated MSE value of BML is greater than the estimated MSE values of BRR, BKL, BOK, and BDK estimators. Moreover, the MSE values of BDK(1) and BDK(2) estimators are lower than other estimators, which means that the BDK estimator achieves the best performance. Furthermore, in terms of the prediction, the R^2 value of the proposed estimator (BDK) is the greatest among all the used estimators. To further highlight the performance of the BDK estimator, generalized cross-validation (GCV) criterion is used in comparison [8, 34, 35]. Regarding GCV values, it can note that the BDK yielded the least value compared with other estimators.

Through this application, we verify the theoretical results as follows:

1. Since the condition $4\hat{k}^2_{BDK(2)}(1+\hat{d}_{BOK})^2\hat{\phi}\sum_{j=1}^{p}\gamma_j\hat{\alpha}_j^2 = 7.26e+7 < \sum_{j=1}^{p}\left(\hat{M}_j^2-\hat{R}_j^2\right) = 1.58e+10$ is satisfied, then the BDK estimator is better than the BML estimator.
2. Since the condition $\sum_{j=1}^{p}\left(\hat{R}_j^2\hat{L}_j^2-\gamma_j^2\hat{M}_j^2\right) = -1.35e+26 < \hat{k}^2_{BDK(2)}\hat{\phi}\sum_{j=1}^{p}\gamma_j\hat{\alpha}_j^2(\hat{M}_j^2-4(1+\hat{d}_{BOK})^2\hat{L}_j^2) = -7.83e+23$ is satisfied, then the BDK estimator is better than the BRR estimator.
3. Since the condition $\sum_{j=1}^{p}\left(\hat{R}_j^2\hat{L}_j^2-\hat{N}_j^2\hat{M}_j^2\right) = -7.84e+24 < 4\,\hat{k}^2_{BDK(2)}\,\hat{\phi}\sum_{j=1}^{p}\gamma_j\hat{\alpha}_j^2(\hat{M}_j^2-(1+\hat{d}_{BOK})^2\hat{L}_j^2) = -6.03e+22$ is satisfied, then the BDK estimator is better than the BKL estimator.
4. Since the condition $\sum_{j=1}^{p}\left(\hat{R}_j^2\hat{L}_j^2-\hat{G}_j^2\hat{M}_j^2\right) = -1.39e+26 < \hat{k}^2_{BDK(2)}\hat{\phi}\sum_{j=1}^{p}\gamma_j\hat{\alpha}_j^2((1-\hat{d}_{BOK})^2\hat{M}_j^2-4(1+\hat{d}_{BOK})^2\hat{L}_j^2) = -7.98e+23$ is satisfied, then the BDK estimator is better than the BOK estimator.

CONCLUSION

Regression modeling describes the relationship that exists between a dependent variable and one or more explanatory variables. Linear dependency, a situation called multicollinearity, is a common problem with two or more explanatory variables. Multicollinearity is a threat to the efficiency of the maximum likelihood estimator in both the linear and generalized linear models, such as the BR model. The ridge regression estimator serves as an alternative to the maximum likelihood estimator for parameter estimation in the beta regression model. In this article, we developed the BDK estimator and compared its performance theoretically with some other estimators. A simulation study has been conducted to compare the performance of the estimators. Real-life data have been analyzed to illustrate the findings of the article. We concluded that the BDK estimator proposed in this articles generally preferred when there is multicollinearity in the beta regression model. For future work, for example, one can use new methods to select the shrinkage parameters as an extension to Uslu et al. [36] and Inan et al. [37] in the BR model, or provide robust biased estimators for handling multicollinearity and outliers together in the beta regression model as an extension to Awwad et al. [31] and Dawoud and Abonazel [30].

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

MA, ID, and FA contributed to conception and structural design of the manuscript. MA performed the simulation and application sections. AL wrote the abstract and conclusion sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

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

REFERENCES

1. Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics.* (1970) 12:55–67. doi: 10.1080/00401706.1970.10488634
2. Hoerl AE, Kennard RW. Ridge regression: applications to non-orthogonal problems. *Technometrics.* (1970) 12:69–82. doi: 10.1080/00401706.1970.10488635
3. Kibria BMG, Lukman AFA. New ridge-type estimator for the linear regression model: simulations and applications. *Scientifica.* (2020) 2020:9758378. doi: 10.1155/2020/9758378
4. Özkale MR, Kaçiranlar S. The restricted, and unrestricted two-parameter estimators. *Commun Stat Theory Methods.* (2007) 36:2707–25. doi: 10.1080/03610920701386877
5. Dawoud I, Kibria BMG. A new biased estimator to combat the multicollinearity of the gaussian linear regression model. *Stat J.* (2020) 3:526–41. doi: 10.3390/stats3040033
6. Roozbeh M, Arashi M, Hamzah NA. Generalized cross-validation for simultaneous optimization of tuning parameters in ridge regression. *Iran J Sci Technol Trans A Sci.* (2020) 44:473–85. doi: 10.1007/s40995-020-00851-1
7. Lukman AF, Ayinde K, Kibria GBM, Adewuyi E. Modified ridge-type estimator for the gamma regression model. *Commun Stat Simul Comput.* (2020). doi: 10.1080/03610918.2020.1752720
8. Arashi M, Roozbeh M, Hamzah NA, Gasparini M. Ridge regression and its applications in genetic studies. *PloS One.* (2021) 16:e0245376. doi: 10.1371/journal.pone.0245376
9. Farghali RA, Qasim M, Kibria BG, Abonazel MR. Generalized two-parameter estimators in the multinomial logit regression model: methods, simulation and application. *Commun Stat Simul Comput.* (2021) 1−16. doi: 10.1080/03610918.2021.1934023
10. Lukman AF, Aladeitan B , Ayinde K, Abonazel MR. Modified ridge-type for the Poisson regression model: simulation and application. *J Appl Stat.* (2021). 1−13. doi: 10.1080/02664763.2021.1889998
11. Lukman AF, Issam D, Kibria GBM, Zakariya A, Aladeitan B. A new ridge-type estimator for the gamma regression model. *Scientifica.* (2021) 2021:1–8. doi: 10.1155/2021/5545356
12. Algamal ZY, Abonazel MR. Developing a Liu-type estimator in beta regression model. *Concurrency Comput Pract Exp.* (2021) 34:e6685. doi: 10.1002/cpe.6685
13. Akram MN, Amin M, Elhassanein A, Ullah MA. A new modified ridge-type estimator for the beta regression model: simulation and application. *AIMS Math.* (2022) 7:1035–57. doi: 10.3934/math.2022062
14. Abonazel MR, Algamal ZY, Awwad FA, Taha IM. A new two-parameter estimator for beta regression model: method, simulation, and application. *Front Appl Math Stat.* (2022) 7:780322. doi: 10.3389/fams.2021.780322
15. Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. *J Appl Stat.* (2004) 31:799–815. doi: 10.1080/0266476042000214501
16. Algamal ZY. A particle swarm optimization method for variable selection in beta regression model. *Electron J Appl Stat Anal.* (2019) 12:508–19.
17. Mahmood SW, Seyala NN, Algamal ZY. Adjusted R2-type measures for beta regression model. *Electron J Appl Stat Anal.* (2020) 13:350–7. doi: 10.1285/i20705948v13n2p350
18. Espinheira PL, Ferrari SL, Cribari-Neto F. On beta regression residuals. *J Appl Stat.* (2008) 35:407–19. doi: 10.1080/02664760701834931
19. Espinheira PL, da Silva LCM, Silva ADO. *Prediction Measures in Beta Regression Models*. arXiv *preprint arXiv:*1501.04830 (2015).
20. Espinheira PL, da Silva LCM, Silva ADO, Ospina R. Model selection criteria on beta regression for machine learning. *Mach Learn Knowl Extraction.* (2019) 1:427–49. doi: 10.3390/make1010026
21. Abonazel MR, Taha IM. Beta ridge regression estimators: simulation and application. *Commun Stat Simul Comput.* (2021) 1−13. doi: 10.1080/03610918.2021.1960373
22. Qasim M, Månsson K, Golam Kibria BM. On some beta ridge regression estimators: method, simulation and application. *J Stat Comput Simul.* (2021) 91:1699–712. doi: 10.1080/00949655.2020.1867549
23. Lukman AF, Adewuyi E, Månsson K, Kibria GBM. A new estimator for the multicollinear Poisson regression model: simulation and application. *Sci Rep.* (2021) 11:3732. doi: 10.1038/s41598-021-82582-w
24. Hoerl AE, Kennard RW, Baldwin KF. Ridge regression: some simulations. *Commun Stat Theory Methods.* (1975) 4:105–23. doi: 10.1080/03610917508548342
25. Lukman AF, Ayinde K, Binuomote S, Onate AC. Modified ridge-type estimator to combat multicollinearity: application to chemical data. *J Chemometr.* (2019) 33:e3125. doi: 10.1002/cem.3125
26. Kibria BG. Performance of some new ridge regression estimators. *Commun Stat Simul Comput.* (2003) 32:419–35. doi: 10.1081/SAC-120017499
27. Kaçiranlar S, Dawoud I. On the performance of the Poisson and the negative binomial ridge predictors. *Commun Stat Simul Comput.* (2018) 47:1751–70. doi: 10.1080/03610918.2017.1324978
28. Rady EA, Abonazel MR, Taha IM. A new biased estimator for zero-inflated count regression models. *J Mod Appl Stat Methods.* (2019). Available online at: https://www.researchgate.net/publication/337155202_A_New_Biased_Estimator_for_Zero-Inflated_Count_Regression_Models
29. Abonazel MR, Farghali RA. Liu-Type multinomial logistic estimator. *Sankhya B.* (2019) 81:203–25. doi: 10.1007/s13571-018-0171-4
30. Dawoud I, Abonazel MR. Robust Dawoud–Kibria estimator for handling multicollinearity and outliers in the linear regression model. *J Stat Comput Simul.* (2021) 91:3678–92. doi: 10.1080/00949655.2021.1945063
31. Awwad FA, Dawoud I, Abonazel MR. Development of robust Özkale–Kaçiranlar and Yang–Chang estimators for regression models in the presence of multicollinearity and outliers. *Concurrency Comput Pract Exp.* (2021) e6779. doi: 10.1002/cpe.6779

32. Akta,ş S, Unlu H. Beta regression for the indicator values of well-being index for provinces in Turkey. *J Eng Technol Appl Sci.* (2017) 2:101–11. doi: 10.30931/jetas.321165
33. Kim JH. Multicollinearity and misleading statistical results. *Korean J Anesthesiol.* (2019) 72:558–69. doi: 10.4097/kja.19087
34. Amini M, Roozbeh M. Optimal partial ridge estimation in restricted semiparametric regression models. *J Multivariate Anal.* (2015) 136:26–40. doi: 10.1016/j.jmva.2015.01.005
35. Roozbeh M. Optimal QR-based estimation in partially linear regression models with correlated errors using GCV criterion. *Comput Stat Data Anal.* (2018) 117:45–61. doi: 10.1016/j.csda.2017.08.002
36. Uslu VR, Egrioglu E, Bas E. Finding optimal value for the shrinkage parameter in ridge regression via particle swarm optimization. *Am J Intell Syst.* (2014) 4:142–7. doi: 10.5923/j.ajis.20140404.03
37. Inan D, Egrioglu E, Sarica B, Askin OE, Tez M. Particle swarm optimization based Liu-type estimator. *Commun Stat Theory Methods.* (2017) 46:11358–69. doi: 10.1080/03610926.2016.1267759

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Abonazel, Dawoud, Awwad and Lukman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.