



A New Two-Parameter Estimator for Beta Regression Model: Method, Simulation, and Application

Mohamed R. Abonazel^{1*}, Zakariya Yahya Algamal², Fuad A. Awwad³ and Ibrahim M. Taha⁴

¹ Department of Applied Statistics and Econometrics, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza, Egypt, ² Department of Statistics and Informatics, University of Mosul, Mosul, Iraq, ³ Department of Quantitative Analysis, College of Business Administration, King Saud University, Riyadh, Saudi Arabia, ⁴ Department of Mathematics, Statistics, and Insurance, Sadat Academy for Management Sciences, Tanta, Egypt

OPEN ACCESS

Edited by:

Gareth William Peters,
University of California, Santa Barbara,
United States

Reviewed by:

Mingjun Zhong,
University of Aberdeen,
United Kingdom
Mohammad Arashi,
Ferdowsi University of Mashhad, Iran

*Correspondence:

Mohamed R. Abonazel
mabonazel@cu.edu.eg

Specialty section:

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

Received: 20 September 2021

Accepted: 17 December 2021

Published: 26 January 2022

Citation:

Abonazel MR, Algamal ZY, Awwad FA
and Taha IM (2022) A New
Two-Parameter Estimator for Beta
Regression Model: Method,
Simulation, and Application.
Front. Appl. Math. Stat. 7:780322.
doi: 10.3389/fams.2021.780322

The beta regression is a widely known statistical model when the response (or the dependent) variable has the form of fractions or percentages. In most of the situations in beta regression, the explanatory variables are related to each other which is commonly known as the multicollinearity problem. It is well-known that the multicollinearity problem affects severely the variance of maximum likelihood (ML) estimates. In this article, we developed a new biased estimator (called a two-parameter estimator) for the beta regression model to handle this problem and decrease the variance of the estimation. The properties of the proposed estimator are derived. Furthermore, the performance of the proposed estimator is compared with the ML estimator and other common biased (ridge, Liu, and Liu-type) estimators depending on the mean squared error criterion by making a Monte Carlo simulation study and through two real data applications. The results of the simulation and applications indicated that the proposed estimator outperformed ML, ridge, Liu, and Liu-type estimators.

Keywords: biased estimation, Fisher's scoring, mean squared error (MSE), multicollinearity, Liu beta regression, relative efficiency, ridge beta regression, two-parameter estimator

INTRODUCTION

The beta regression model has been common in many areas, primarily economic and medical research, such as income share, unemployment rates in certain nations, the Gini index for each region, graduation rates in major universities, or the percentage of body fat in medical subjects. Beta regression model, such as any regression model in the context of generalized linear models (GLMs) is used to examine the effect of certain explanatory variables on a non-normal response variable. However, in the case of beta regression, the response component is restricted to an interval (0, 1), such as proportions, percentages, and fractions.

Multicollinearity is a popular issue in econometric modeling. It indicates that there is a strong association between the explanatory variables. It is well-established that the covariance matrix of the maximum likelihood (ML) estimator is ill-conditioned in the case of severing multicollinearity. One of the negative consequences of this issue is that the variance of the regression coefficients gets inflated. As a consequence, the significance and the magnitude of the coefficients are affected. Many of the conventional approaches used to address this issue include: gathering additional data, re-specifying the model, or removing the correlated variable/s.

During the last years, shrinkage methods have become a commonly recognized and more effective methodology for solving the impact of the multicollinearity problem in several regression models. To solve this problem, Hoerl and Kennard [1, 2] proposed the ridge estimator. The concept of the ridge estimator is to add a small definite amount (k) to the diagonal entries of the covariance matrix to increase the conditioning of this matrix, reduce the mean squared error (MSE), and achieve consistent coefficients. For a review of the ridge estimator in both linear and GLMs, e.g., as shown in References Rady et al. [3], Abonazel and Taha [4], Qasim et al. [5], Alobaidi et al. [6], and Sami et al. [7].

One of the drawbacks of the ridge estimator is that estimated parameters are non-linear functions of the ridge parameter and the small k selected might not be high enough to solve multicollinearity. As a solution to this problem, Liu [8] developed the Liu estimator which is a linear function of the shrinkage parameter. The Liu estimator is a combination of the ridge estimator and the Stein estimator suggested by Stein [9]. For a review of the Liu estimator in both linear and GLMs, e.g., as shown in References. Liu [8], Karlsson et al. [10], Qasim et al. [11], and Naveed et al. [12]. Furthermore, Liu [13] proved the supremacy of the Liu-type estimator over the ridge and Liu estimators. Details about Liu-type estimator, properties, and applications in regression models are shown in References Liu [14], Özkale and Kaciranlar [15], Li and Yang [16], Kurnaz and Akay [17], Sahriman and Koerniawan [18], and Algalal and Abonazel [19]. As a good alternative for the Liu-type estimator, Özkale and Kaciranlar [15] proposed the two-parameter estimator, and they proved that the two-parameter estimator utilizes the power of both the ridge estimator and the Liu estimator. Extensions of two-parameter estimator in GLMs include Huang and Yang [20], Algalal [21], Asar and Genç [22], Rady et al. [23, 24], Çetinkaya and Kaçiranlar [25], Abonazel and Farghali [26], Akram et al. [27], and Lukman et al. [28].

The rest of the article is arranged as follows: Section Methodology presents an introduction about the beta regression model, its estimation using the ML method, and the proposed two-parameter estimator; Section Choosing the Shrinkage Parameters provides suggested shrinkage parameters for our estimator; Sections Simulation Study and Real Data Applications provide a numerical evaluation using both Monte Carlo simulation and two empirical data applications, respectively; and Section Conclusion offers some concluding remarks.

METHODOLOGY

Beta Regression Model

Practitioners usually use linear regression modeling to investigate the relationship and effect of some selected explanatory variables on the normal response variable. However, this is not suitable for circumstances where the response variable is constrained to the interval (0, 1) because it may give fitted values for the variable of concern that surpass its lower and upper limits. Therefore, inference based on the normality assumption can be deceptive. The beta regression model was first developed by Ferrari and Cribari-Neto [29] by connecting the mean function

of its response variable to a set of linear predictors *via* a monotone differentiable function called the link function. This model contains a precision parameter, the inverse of which is called a dispersion scale. In the basic type of a beta regression model, the precision parameter is believed to be constant through observations. Nevertheless, the precision parameter might not be constant through findings such as those of Smithson and Verkuilen [30] and Cribari-Neto and Zeileis [31].

Let y is a continuous random variable that follows a beta distribution with the following probability density function:

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

where $\Gamma(\cdot)$ is the gamma function and ϕ is the precision parameter [32]:

$$\phi = \frac{1 - \sigma^2}{\sigma^2}.$$

The mean and variance of the beta probability distribution are: $E(y) = \mu$, $\text{var}(y) = \mu(1-\mu)\sigma^2$. Using the logit link function, the model allows μ_i , depending on covariates as follows:

$$g(\mu_i) = \log\left(\frac{\mu_i}{1-\mu_i}\right) = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i, \quad (2)$$

where $g(\cdot)$ be a monotonic differentiable link function used to relate the systematic component with the random component, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ is a $p \times 1$ vector of unknown parameters, $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$ is the vector of p regressors, and η_i is a linear predictor.

Estimation of the beta regression parameters is done by using the ML method [33]. The log-likelihood function of the beta regression model is given by:

$$\begin{aligned} \mathcal{L}(\mu_i, \phi; y_i) &= \sum_{i=1}^n \{ \log \Gamma(\phi) - \log \Gamma(\mu_i(\phi)) \\ &\quad - \log \Gamma((1-\mu_i)(\phi)) + (\mu_i(\phi) - 1) \log(y_i) \\ &\quad + ((1-\mu_i)(\phi) - 1) \log(1-y_i) \} \end{aligned} \quad (3)$$

By differentiating the log-likelihood function in Eq. (3) with respect to $\boldsymbol{\beta}$, gives us the score function for $\boldsymbol{\beta}$:

$$S(\boldsymbol{\beta}) = \boldsymbol{\phi} X^T A (y^* - \mu^*), \quad (4)$$

where $A = \text{diag}\left(\frac{1}{g'(\mu_1)}, \dots, \frac{1}{g'(\mu_n)}\right)$, $y^* = (y_1^*, \dots, y_n^*)^T$, $\mu^* = (\mu_1^*, \dots, \mu_n^*)^T$, $y_i^* = \log\left(\frac{y_i}{1-y_i}\right)$, and $\mu_i^* = \psi(\mu_i\phi) - \psi((1-\mu_i)\phi)$, such that $\psi(\cdot)$ denoting the digamma function, and $g'(\cdot)$ is the first derivative of $g(\cdot)$. The iterative reweighted least-squares (IWLS) algorithm or Fisher's scoring algorithm was used for estimating $\boldsymbol{\beta}$ [34, 35]. The form of this algorithm can be written as:

$$\boldsymbol{\beta}^{(r+1)} = \boldsymbol{\beta}^{(r)} + \left(\mathbf{I}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{(r)}\right)^{-1} \mathbf{S}_{\boldsymbol{\beta}}^{(r)}(\boldsymbol{\beta}),$$

where $S_{\beta}^{(r)}$ is the score function defined in Eq. (4), and $I_{\beta\beta}^{(r)}$ is the information matrix for β , as shown in References Espinheira et al. [35] for more details. The initial value of β can be obtained by the least-squares estimation, while the initial value for each precision parameter is:

$$\hat{\phi}_i = \frac{\hat{\mu}_i(1 - \hat{\mu}_i)}{\hat{\sigma}_i^2}, \tag{5}$$

where $\hat{\mu}$ and $\hat{\sigma}_i^2$ values are obtained from linear regression. Given $r = 0, 1, 2, \dots$ is the number of iterations that are performed, convergence occurs when the difference between successive estimates becomes smaller than a given small constant. At the final step, the ML estimator of β is obtained as:

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

where X is an $n \times p$ matrix of regressors, $\hat{z} = \hat{\eta} + \hat{W}^{-1} \hat{A} (y^* - \mu^*)$, and $\hat{W} = \text{diag}(\hat{w}_1, \dots, \hat{w}_n)$;

$$\hat{w}_i = \frac{(1 - \hat{\sigma}^2)}{\hat{\sigma}^2} \left\{ \psi' \left(\frac{\hat{\mu}_i(1 - \hat{\sigma}^2)}{\hat{\sigma}^2} \right) + \psi' \left(\frac{(1 - \hat{\mu}_i)(1 - \hat{\sigma}^2)}{\hat{\sigma}^2} \right) \right\} \frac{1}{\{g'(\hat{\mu}_i)\}^2},$$

where \hat{W} and \hat{A} are the estimated ML matrices of W and A , respectively. The ML estimator of β is normally distributed with asymptotic mean vectors $E(\hat{\beta}_{BR}) = \beta$ and asymptotic covariance matrix:

$$\text{Cov}(\hat{\beta}_{BML}) = \frac{1}{\phi} (X^T \hat{W} X)^{-1} \tag{7}$$

Hence, the asymptotic trace mean squared error (TMSE) of $\hat{\beta}_{BML}$ is

$$\text{TMSE}(\hat{\beta}_{BML}) = \text{tr} \left[\frac{1}{\phi} (X^T \hat{W} X)^{-1} \right] \tag{8}$$

Ridge and Liu Estimators

Recently, Abonazel and Taha [4] and Qasim et al. [5] introduced the ridge beta regression (RBR) estimator as follows:

$$\hat{\beta}_{RBR} = (X^T \hat{W} X + kI)^{-1} X^T \hat{W} \hat{z}; k > 0 \tag{9}$$

It can note that if $k = 0$, then $\hat{\beta}_{RBR} = \hat{\beta}_{BML}$. The bias vector of the RBR estimator is

$$\text{Bias}(\hat{\beta}_{RBR}) = -k(X^T \hat{W} X + kI)^{-1} \beta \tag{10}$$

Suppose that $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ are the ordered eigenvalues of $X^T \hat{W} X$ matrix and Q is the matrix whose columns are the eigenvectors of $X^T \hat{W} X$ matrix. Then $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p) =$

$Q^T X^T \hat{W} X Q$ and $\alpha = Q^T \gamma$. Then, the matrix mean squared error (MMSE) of the RBR estimator is:

$$\begin{aligned} \text{MMSE}(\hat{\beta}_{RBR}) &= \text{Cov}(\hat{\beta}_{RBR}) + \text{Bias}(\hat{\beta}_{RBR}) \text{Bias}(\hat{\beta}_{RBR})^T \\ &= \frac{1}{\phi} (Q \Lambda_k^{-1} \Lambda \Lambda_k^{-1} Q^T) + k^2 Q \Lambda_k^{-1} \alpha \alpha^T \Lambda_k^{-1} Q^T, \end{aligned} \tag{11}$$

where $\Lambda_k = \text{diag}(\lambda_1 + k, \dots, \lambda_p + k)$, and the TMSE of the RBR estimator is

$$\begin{aligned} \text{TMSE}(\hat{\beta}_{RBR}) &= \text{tr}(\text{MMSE}(\hat{\beta}_{RBR})) \\ &= \frac{1}{\phi} \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^2} \end{aligned} \tag{12}$$

The first term in Eq. (12) is an asymptotic variance, and the second term is a square bias. Abonazel and Taha [4] and Qasim et al. [5] showed the derivation of the MSE properties of the RBR estimator.

The Liu estimator can be extended to the beta regression model, the Liu beta regression (LBR) estimator is given by Karlsson et al. [10] as:

$$\begin{aligned} \hat{\beta}_{LBR} &= (X^T \hat{W} X + I)^{-1} (X^T \hat{W} X + dI) \hat{\beta}_{BML}; \\ 0 < d < 1, \end{aligned} \tag{13}$$

where d is the Liu parameter, the bias vector of the LBR estimator is:

$$\text{Bias}(\hat{\beta}_{LBR}) = (X^T \hat{W} X + I)^{-1} (d - 1) \beta \tag{14}$$

The MMSE for the LBR estimator can be derived as:

$$\begin{aligned} \text{MMSE}(\hat{\beta}_{LBR}) &= \text{Cov}(\hat{\beta}_{LBR}) + \text{Bias}(\hat{\beta}_{LBR}) \text{Bias}(\hat{\beta}_{LBR})^T \\ &= \frac{1}{\phi} (Q \Lambda_1^{-1} \Lambda_d \Lambda^{-1} \Lambda_d \Lambda_1^{-1} Q^T) \\ &\quad + (d - 1)^2 Q \Lambda_1^{-1} \alpha \alpha^T \Lambda_1^{-1} Q^T, \end{aligned} \tag{15}$$

where $\Lambda_1 = \text{diag}(\lambda_1 + 1, \dots, \lambda_p + 1)$ and $\Lambda_d = \text{diag}(\lambda_1 + d, \dots, \lambda_p + d)$. The TMSE of the LBR estimator is:

$$\begin{aligned} \text{TMSE}(\hat{\beta}_{LBR}) &= \text{tr}(\text{MMSE}(\hat{\beta}_{LBR})) \\ &= \frac{1}{\phi} \sum_{j=1}^p \left\{ \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + \frac{(d - 1)^2 \alpha_j^2 \phi}{(\lambda_j + 1)^2} \right\} \end{aligned} \tag{16}$$

Recently, Algamal and Abonazel [19] developed the Liu-type beta regression (LTBR) estimator:

$$\begin{aligned} \hat{\beta}_{LTBR} &= (X^T \hat{W} X + kI)^{-1} (X^T \hat{W} X - dI) \hat{\beta}_{BML}; \\ k > 0, -\infty < d < \infty \end{aligned} \tag{17}$$

The bias vector of the LTBR estimator is:

$$\text{Bias}(\hat{\beta}_{LTBR}) = -(d+k)(X^T \hat{W}X + kI)^{-1} \beta \tag{18}$$

The MMSE of the LTBR estimator is:

$$\begin{aligned} \text{MMSE}(\hat{\beta}_{LTBR}) &= \text{Cov}(\hat{\beta}_{LTBR}) + \text{Bias}(\hat{\beta}_{LTBR}) \text{Bias}(\hat{\beta}_{LTBR})^T \\ &= \frac{1}{\phi} \left(Q \Lambda_k^{-1} \Lambda_{-d} \Lambda^{-1} \Lambda_{-d} \Lambda_k^{-1} Q^T \right) \\ &\quad + (d+k)^2 Q \Lambda_k^{-1} \alpha \alpha^T \Lambda_k^{-1} Q^T \end{aligned} \tag{19}$$

where $\Lambda_{-d} = \text{diag}(\lambda_1 - d, \dots, \lambda_p - d)$. Then, the MSE of the LTBR estimator is

$$\begin{aligned} \text{MSE}(\hat{\beta}_{LTBR}) &= \text{tr}(\text{MMSE}(\hat{\beta}_{LTBR})) \\ &= \frac{1}{\phi} \sum_{j=1}^p \left\{ \frac{(\lambda_j - d)^2}{\lambda_j(\lambda_j + k)^2} + \frac{(d+k)^2 \alpha_j^2 \phi}{(\lambda_j + k)^2} \right\} \end{aligned} \tag{20}$$

The Proposed Estimator

In this section, we extend the two-parameter estimator introduced by Özkale and Kaçiranlar [15] to the beta regression model to combat multicollinearity and obtain more stable and accurate results. The two-parameter beta regression (TPBR) estimator can be written as follows:

$$\begin{aligned} \hat{\beta}_{TPBR} &= (X^T \hat{W}X + kI)^{-1} (X^T \hat{W}X + kdI) \hat{\beta}_{BML}; \\ k > 0; 0 < d < 1 \end{aligned} \tag{21}$$

It is worth noting that the TPBR estimator is a general class that has some estimators as special cases. These estimators are the LBR, RBR, and beta maximum likelihood (BML) estimators, which can be given, respectively, as follows:

$$\begin{aligned} \lim_{k \rightarrow 1} \hat{\beta}_{TPBR} &= \hat{\beta}_{LBR} = (X^T \hat{W}X + I)^{-1} (X^T \hat{W}X + dI) \hat{\beta}_{BML}, \\ \lim_{d \rightarrow 0} \hat{\beta}_{TPBR} &= \hat{\beta}_{RBR} = (X^T \hat{W}X + kI)^{-1} (X^T \hat{W}X) \hat{\beta}_{BML}, \\ \lim_{k \rightarrow 0} \hat{\beta}_{TPBR} &= \hat{\beta}_{BML} = (X^T \hat{W}X)^{-1} (X^T \hat{W}z). \end{aligned}$$

The bias vector of the TPBR estimator is

$$\text{Bias}(\hat{\beta}_{TPBR}) = k(d-1)(X^T \hat{W}X + kI)^{-1} \beta \tag{22}$$

The MMSE for TPBR estimator can be derived as:

$$\begin{aligned} \text{MMSE}(\hat{\beta}_{TPBR}) &= \text{Cov}(\hat{\beta}_{TPBR}) + \text{Bias}(\hat{\beta}_{TPBR}) \text{Bias}(\hat{\beta}_{TPBR})^T \\ &= \frac{1}{\phi} \left(Q \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1} Q^T \right) \\ &\quad + k^2(d-1)^2 Q \Lambda_k^{-1} \alpha \alpha^T \Lambda_k^{-1} Q^T, \end{aligned} \tag{23}$$

where $\Lambda_{kd} = \text{diag}(\lambda_1 + kd, \lambda_2 + kd, \dots, \lambda_p + kd)$, the TMSE of the TPBR estimator is:

$$\begin{aligned} \text{TMSE}(\hat{\beta}_{TPBR}) &= \text{tr}(\text{MMSE}(\hat{\beta}_{TPBR})) \\ &= \frac{1}{\phi} \sum_{j=1}^p \left\{ \frac{(\lambda_j + kd)^2}{\lambda_j(\lambda_j + k)^2} + \frac{\alpha_j^2 \phi k^2 (d-1)^2}{(\lambda_j + k)^2} \right\} \end{aligned} \tag{24}$$

The Superiority of the New Estimator

The following lemmas prove the superiority of the two-parameter beta estimator over the other estimators.

Lemma 1. Farebrother [36]: Let M be a positive definite matrix, δ be a vector of non-zero constants, and c be a positive constant. Then, $cM - \delta\delta^T > 0$ if and only if (iff) $\delta M \delta^T < c$.

Two-Parameter Beta Estimator vs. ML Estimator

The following lemma gives the condition that the TPBR estimator is superior to the ML estimator:

Lemma 2. under the beta regression model, let $k > 0$, $0 < d < 1$, and $b_{TPBR} = \text{Bias}(\hat{\beta}_{TPBR})$. Then, $\text{MMSE}(\hat{\beta}_{BML}) - \text{MMSE}(\hat{\beta}_{TPBR}) > 0$ iff $k(1-d)(2\lambda_j + k(1+d)) > 0$.

Proof: the difference between the MMSE functions of the ML estimator and the TPBR estimator is obtained by:

$$\begin{aligned} &\text{MMSE}(\hat{\beta}_{BML}) - \text{MMSE}(\hat{\beta}_{TPBR}) \\ &= \frac{1}{\phi} Q (\phi \Lambda^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1}) Q^T - b_{TPBR} b_{TPBR}^T, \end{aligned}$$

The matrix $(\phi \Lambda^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1})$ is positive definite, if $\phi(\lambda_j + k)^2 - (\lambda_j + kd)^2 > 0$, which is equivalent to $[(\lambda_j + k) + (\lambda_j + kd)][(\lambda_j + k) - (\lambda_j + kd)] > 0$. Simplifying the last inequality, one gets $k(1-d)(2\lambda_j + k(1+d)) > 0$. The proof is finished by Lemma 1.

Two-Parameter Estimator vs. Ridge Estimator

The following lemma gives that the TPBR estimator is superior to the RBR estimator:

Lemma 3. under the beta regression model, consider $k > 0$, $0 < d < 1$, and $b_{RBR} = \text{Bias}(\hat{\beta}_{RBR})$. Then, $\text{MMSE}(\hat{\beta}_{RBR}) - \text{MMSE}(\hat{\beta}_{TPBR}) > 0$ iff $kd(2\lambda_j + kd) > 0$.

Proof: the difference between the MMSE functions of the RBR estimator and the TPBR estimator is obtained by:

$$\begin{aligned} &\text{MMSE}(\hat{\beta}_{RBR}) - \text{MMSE}(\hat{\beta}_{TPBR}) \\ &= \frac{1}{\phi} Q (\Lambda_k^{-1} \Lambda_k^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1}) \\ &\quad Q^T + b_{RBR} b_{RBR}^T - b_{TPBR} b_{TPBR}^T, \end{aligned}$$

This can be rewritten as:

$$\begin{aligned} & \text{MMSE}(\hat{\beta}_{\text{RBR}}) - \text{MMSE}(\hat{\beta}_{\text{TPBR}}) \\ &= \frac{1}{\phi} Q \text{diag} \left\{ \frac{\lambda_j}{(\lambda_j + k)^2} - \frac{(\lambda_j + kd)^2}{\lambda_j(\lambda_j + k)^2} \right\}_{j=1}^p \\ & \quad Q^T + b_{\text{RBR}} b_{\text{RBR}}^T - b_{\text{TPBR}} b_{\text{TPBR}}^T, \end{aligned}$$

The matrix $\Lambda_k^{-1} \Lambda \Lambda_k^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1}$ is positive definite if $\lambda_j^2 - (\lambda_j + kd)^2 > 0$ which is equivalent to $[\lambda_j - (\lambda_j + kd)][\lambda_j + (\lambda_j + kd)] > 0$. Simplifying the last inequality, one gets $kd(2\lambda_j + kd) > 0$. Then, using Lemma 1, the proof is finished.

Two-Parameter Estimator vs. Liu Estimator

The following lemma gives the condition that the TPBR estimator is superior to the LBR estimator:

Lemma 4. under the beta regression model, consider $k > 0$, $0 < d < 1$, and $b_{\text{LBR}} = \text{Bias}(\hat{\beta}_{\text{LBR}})$. Then, $\text{MMSE}(\hat{\beta}_{\text{LBR}}) - \text{MMSE}(\hat{\beta}_{\text{TPBR}}) > 0$ iff $(\lambda_j + d)^2(\lambda_j + k)^2 - (\lambda_j + kd)^2(\lambda_j + I)^2 > 0$.

Proof: the difference between the MMSE functions of $\hat{\beta}_{\text{LBR}}$ and $\hat{\beta}_{\text{TPBR}}$ is obtained by:

$$\begin{aligned} & \text{MMSE}(\hat{\beta}_{\text{LBR}}) - \text{MMSE}(\hat{\beta}_{\text{TPBR}}) \\ &= \frac{1}{\phi} Q (\Lambda_1^{-1} \Lambda_d \Lambda^{-1} \Lambda_d \Lambda_1^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1}) \\ & \quad Q^T + b_{\text{LBR}} b_{\text{LBR}}^T - b_{\text{TPBR}} b_{\text{TPBR}}^T, \end{aligned}$$

This can be rewritten as:

$$\begin{aligned} & \text{MMSE}(\hat{\beta}_{\text{LBR}}) - \text{MMSE}(\hat{\beta}_{\text{TPBR}}) \\ &= \frac{1}{\phi} Q \text{diag} \left\{ \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + I)^2} - \frac{(\lambda_j + kd)^2}{\lambda_j(\lambda_j + k)^2} \right\}_{j=1}^p \\ & \quad Q^T + b_{\text{LBR}} b_{\text{LBR}}^T - b_{\text{TPBR}} b_{\text{TPBR}}^T, \end{aligned}$$

The matrix $\Lambda_1^{-1} \Lambda_d \Lambda^{-1} \Lambda_d \Lambda_1^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1}$ is positive definite if $(\lambda_j + d)^2(\lambda_j + k)^2 - (\lambda_j + kd)^2(\lambda_j + I)^2 > 0$, which is equivalent to $[(\lambda_j + d)^2(\lambda_j + k)^2 > (\lambda_j + kd)^2(\lambda_j + I)^2]$. For $k > 0, 0 < d < 1$, it can be observed that $(\lambda_j + d)^2(\lambda_j + k)^2 - (\lambda_j + kd)^2(\lambda_j + I)^2 > 0$. The proof is finished by Lemma 1.

Two-Parameter Estimator vs. Liu-Type Estimator

The following lemma gives the condition that the TPBR estimator is superior to the LTBR estimator:

Lemma 5. under the beta regression model, consider $k > 0$, $-\infty < d_1 < \infty$, $0 < d_2 < 1$, and $b_{\text{LTBR}} = \text{Bias}(\hat{\beta}_{\text{LTBR}})$,

where d_1 and d_2 are the d values of LTBR and TPBR estimators, respectively. Then, $\text{MMSE}(\hat{\beta}_{\text{LTBR}}) - \text{MMSE}(\hat{\beta}_{\text{TPBR}}) > 0$ iff $d_1(d_1 - 2\lambda_j) - kd_2(kd_2 + 2\lambda_j) > 0$.

Proof: the difference between the MMSE functions of $\hat{\beta}_{\text{LTBR}}$ and $\hat{\beta}_{\text{TPBR}}$ is obtained by:

$$\begin{aligned} & \text{MMSE}(\hat{\beta}_{\text{LTBR}}) - \text{MMSE}(\hat{\beta}_{\text{TPBR}}) \\ &= \frac{1}{\phi} Q (\Lambda_k^{-1} \Lambda_{-d} \Lambda^{-1} \Lambda_{-d} \Lambda_k^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1}) \\ & \quad Q^T + b_{\text{LTBR}} b_{\text{LTBR}}^T - b_{\text{TPBR}} b_{\text{TPBR}}^T. \end{aligned}$$

This can be rewritten as:

$$\begin{aligned} & \text{MMSE}(\hat{\beta}_{\text{LTBR}}) - \text{MMSE}(\hat{\beta}_{\text{TPBR}}) \\ &= \frac{1}{\phi} Q \text{diag} \left\{ \frac{(\lambda_j - d_1)^2}{\lambda_j(\lambda_j + k)^2} - \frac{(\lambda_j + kd_2)^2}{\lambda_j(\lambda_j + k)^2} \right\}_{j=1}^p \\ & \quad Q^T + b_{\text{LTBR}} b_{\text{LTBR}}^T - b_{\text{TPBR}} b_{\text{TPBR}}^T, \end{aligned}$$

The matrix $\Lambda_k^{-1} \Lambda_{-d} \Lambda^{-1} \Lambda_{-d} \Lambda_k^{-1} - \Lambda_k^{-1} \Lambda_{kd} \Lambda^{-1} \Lambda_{kd} \Lambda_k^{-1}$ is positive definite if $(\lambda_j - d_1)^2 - (\lambda_j + kd_2)^2 > 0$, which is equivalent to $[(\lambda_j - d_1)^2 > (\lambda_j + kd_2)^2]$. For $k > 0, -\infty < d_1 < \infty, 0 < d_2 < 1$, it can be observed that $d_1(d_1 - 2\lambda_j) - kd_2(kd_2 + 2\lambda_j) > 0$. The proof is finished by Lemma 1.

CHOOSING THE SHRINKAGE PARAMETERS

There is no definite rule for estimating the shrinkage parameters (k and d). However, we propose some methods based on the work of Hoerl et al. [37] and Kibria [38]. For the RBR estimator, we can use the k parameter of Hoerl and Kennard [1] after modifying their formula based on the optimal k of the beta regression model [5]:

$$k = \frac{1}{\phi \sum_{j=1}^p \hat{\alpha}_j^2}, \tag{25}$$

where $\hat{\alpha}_j$ is the j th element of the vector $\hat{\alpha} = Q^T \hat{\beta}_{\text{BML}}$.

For the LBR estimator, we can use the optimal d parameter proposed by Karlsson et al. [10]:

$$d = \frac{\sum_{j=1}^p \left[\left(\hat{\alpha}_j^2 - \frac{1}{\phi} \right) / (\lambda_j + 1)^2 \right]}{\sum_{j=1}^p \left[\left(\frac{1}{\phi \lambda_j} + \hat{\alpha}_j^2 \right) / (\lambda_j + 1)^2 \right]} \tag{26}$$

For the LTBR estimator, we can use the optimal d parameter of the LTBR estimator that was proposed by Algamal and Abonazel [19]:

$$d_{\text{LTBR}} = \frac{\sum_{j=1}^p \left[\left(\frac{1}{\phi} - k \hat{\alpha}_j^2 \right) / (\lambda_j + k)^2 \right]}{\sum_{j=1}^p \left[\left(\frac{1}{\phi} + \lambda_j \hat{\alpha}_j^2 \right) / \lambda_j (\lambda_j + k)^2 \right]} \tag{27}$$

Since d_{LTBR} depends on k , we suggest using the k parameter in Eq. (25).

For the proposed estimator (TPBR), we start by taking the derivative of MSE function given in Eq. (24) with respect to k and equating the resulting function to zero and by solving for the parameter k , we obtain the following individual parameters:

$$k_j = \frac{\lambda_j}{\phi(\lambda_j \hat{\alpha}_j^2(1-d) - (d/\phi))}; j = 1, \dots, p. \tag{28}$$

Since each individual parameter k_j should be positive, we obtain the following upper bound for the k_j parameter d 's so that $k_j > 0$:

$$d < \min \left(\frac{\lambda_j \hat{\alpha}_j^2}{\frac{1}{\phi} + \lambda_j \hat{\alpha}_j^2} \right)_{j=1}^p, \tag{29}$$

where $\min(\cdot)$ is the minimum function such that $0 < d < 1$ and $\hat{\alpha}_j$ is the j th element of the vector $\hat{\alpha}$. Therefore, we propose the following shrinkage parameters for the

TPBR estimator:

$$d_{TPBR} = \frac{1}{2} \min \left(\frac{\lambda_j \hat{\alpha}_j^2}{\frac{1}{\phi} + \lambda_j \hat{\alpha}_j^2} \right)_{j=1}^p; \tag{30}$$

$$k_{TPBR} = \frac{1}{p} \sum_{j=1}^p \left(\frac{\lambda_j}{\phi(\lambda_j \hat{\alpha}_j^2(1-d_{TPBR}) - d_{TPBR}/\phi)} \right) \tag{31}$$

Note that d_{TPBR} in Eq. (30) is always < 1 and bigger than zero, and k_{TPBR} in Eq. (31) is always positive [15].

SIMULATION STUDY

A Monte Carlo simulation study has been conducted to compare the performances of BML, RBR, LBR, and LTBR estimators with the proposed estimator (TPBR estimator). Our simulation study is computed based on R-software, using the “betareg” package.

Simulation Design

The response variable y_i is generated as $y_i \sim \text{Beta}(\mu_i, \phi)$, with $\phi \in \{0.5, 1, 1.5\}$ and $\mu_i = \exp(x_i^T \beta) / (1 + \exp(x_i^T \beta))$ for $i =$

TABLE 1 | Mean squared error (MSE) values for different estimators when $n = 50$.

ϕ	ρ	ρ	BML	RBR	LBR	LTBR	TPBR
0.5	4	0.90	2.259	1.906	1.828	1.8026	1.419
		0.95	3.356	2.391	2.117	2.0916	2.033
		0.99	4.399	2.908	2.175	2.1496	2.134
	8	0.90	2.376	1.736	1.487	1.4616	1.312
		0.95	4.421	2.567	2.011	1.9856	1.862
		0.99	5.353	3.292	2.518	2.4926	1.669
	12	0.90	3.613	1.364	1.189	1.1636	1.013
		0.95	6.836	2.715	1.987	1.9616	1.107
		0.99	9.375	3.042	1.814	1.7886	1.078
1	4	0.90	1.951	1.598	1.520	1.4946	1.111
		0.95	3.048	2.083	1.809	1.7836	1.725
		0.99	4.091	2.601	1.867	1.8416	1.826
	8	0.90	2.068	1.428	1.179	1.1536	1.004
		0.95	4.112	2.259	1.703	1.6776	1.554
		0.99	5.045	2.984	2.210	2.1846	1.361
	12	0.90	3.305	1.056	0.881	0.8556	0.705
		0.95	6.528	2.407	1.679	1.6536	0.799
		0.99	9.067	2.734	1.506	1.4806	0.771
1.5	4	0.90	1.829	1.476	1.398	1.3726	0.989
		0.95	2.926	1.961	1.687	1.6616	1.603
		0.99	3.969	2.478	1.745	1.7196	1.704
	8	0.90	1.946	1.306	1.057	1.0316	0.882
		0.95	3.990	2.137	1.581	1.5556	1.432
		0.99	4.923	2.862	2.088	2.0626	1.239
	12	0.90	3.183	0.934	0.759	0.7336	0.583
		0.95	6.406	2.285	1.557	1.5316	0.677
		0.99	8.945	2.612	1.384	1.3586	0.648

TABLE 2 | Mean squared error values for different estimators when $n = 100$.

ϕ	ρ	ρ	BML	RBR	LBR	LTBR	TPBR
0.5	4	0.90	2.212	1.859	1.782	1.7566	1.372
		0.95	3.309	2.343	2.071	2.0456	1.985
		0.99	4.352	2.861	2.128	2.1026	2.087
	8	0.90	2.328	1.689	1.439	1.4136	1.265
		0.95	4.372	2.519	1.964	1.9386	1.814
		0.99	5.306	3.244	2.471	2.4456	1.621
	12	0.90	3.565	1.317	1.142	1.1166	0.965
		0.95	6.788	2.668	1.941	1.9156	1.063
		0.99	9.327	2.994	1.766	1.7406	1.031
1	4	0.90	1.904	1.551	1.472	1.4466	1.064
		0.95	3.001	2.035	1.762	1.7366	1.677
		0.99	4.044	2.553	1.822	1.7966	1.779
	8	0.90	2.022	1.381	1.131	1.1056	0.957
		0.95	4.064	2.211	1.656	1.6306	1.506
		0.99	4.998	2.936	2.162	2.1366	1.313
	12	0.90	3.257	1.009	0.834	0.8086	0.657
		0.95	6.483	2.364	1.632	1.6066	0.752
		0.99	9.019	2.686	1.458	1.4326	0.723
1.5	4	0.90	1.782	1.429	1.355	1.3296	0.942
		0.95	2.879	1.913	1.647	1.6216	1.555
		0.99	3.922	2.431	1.698	1.6726	1.657
	8	0.90	1.898	1.259	1.009	0.9836	0.835
		0.95	3.942	2.089	1.534	1.5086	1.384
		0.99	4.876	2.814	2.041	2.0156	1.191
	12	0.90	3.135	0.887	0.712	0.6866	0.535
		0.95	6.358	2.238	1.513	1.4876	0.631
		0.99	8.897	2.564	1.336	1.3106	0.601

1, 2, ..., n, and $\beta = (\beta_1, \dots, \beta_p)^T$ with $\sum_{j=1}^p \beta_j^2 = 1$ and $\beta_1 = \dots = \beta_p$ [19, 26, 39–41].

The explanatory variables $x_i = (x_{i1}, \dots, x_{ip})^T$ are generated from the following:

$$x_{ij} = (1 - \rho^2)^{0.5} w_{ij} + \rho w_{ip}, \quad i = 1, 2, \dots, n, \\ j = 1, 2, \dots, p, \quad (32)$$

where ρ is the coefficient of the correlation between the explanatory variables and w_{ij} are independent standard normal pseudo-random numbers.

It is well-known that the sample size (n), the number of explanatory variables (p), and the pairwise correlation (ρ) between the explanatory variables have a direct impact on the prediction accuracy. Therefore, four values of n are considered: 50, 100, 250, and 400. In addition, three values of p are considered: 4, 8, and 12. Further, three values of ρ are considered: 0.90, 0.95, and 0.99. For a combination of these different values of $n, \phi, p,$ and ρ , the generated data are repeated $L = 1,000$ times and the average MSE is calculated as:

$$MSE(\hat{\beta}) = \frac{1}{L} \sum_{l=1}^L (\hat{\beta}_l - \beta)^T (\hat{\beta}_l - \beta), \quad (33)$$

where $\hat{\beta}_l$ is the estimated vector of β .

Simulation Results

The averaged MSE for all the combinations of $n, \phi, p,$ and ρ are summarized in **Tables 1–4**. According to the simulation results, we conclude the following:

1. The TPBR estimator has the best performance in all the situations considered. Moreover, the performance of the TPBR estimator is better for larger values of ρ .
2. It is noted from **Tables 1–4** that the TPBR estimator ranks first with respect to MSE. In the second rank is the LTBR estimator, as it performs better than BML, RBR, and LBR estimators. Additionally, the BML estimator has the worst performance among RBR, LBR, and TPBR estimators which is significantly impacted by the multicollinearity.
3. Regarding the number of explanatory variables, it is easily seen that there is a negative impact on MSE, where there are increases in their values when the p increase from four variables to eight and twelve variables. In addition, in terms of the sample size, the MSE values decrease when n increases, regardless of the value of $\rho, \phi,$ and p .
4. Clearly, the MSE values are decreasing when ϕ is increasing.

TABLE 3 | Mean squared error values for different estimators when $n = 250$.

ϕ	p	ρ	BML	RBR	LBR	LTBR	TPBR
0.5	4	0.90	2.151	1.798	1.719	1.6936	1.311
		0.95	3.248	2.282	2.009	1.9836	1.924
		0.99	4.291	2.801	2.067	2.0416	2.026
	8	0.90	2.267	1.628	1.378	1.3526	1.204
		0.95	4.311	2.458	1.903	1.8776	1.753
		0.99	5.245	3.183	2.409	2.3836	1.561
	12	0.90	3.504	1.256	1.081	1.0556	0.904
		0.95	6.727	2.607	1.879	1.8536	0.999
		0.99	9.266	2.933	1.705	1.6796	0.972
1	4	0.90	1.843	1.491	1.411	1.3856	1.003
		0.95	2.942	1.974	1.701	1.6756	1.616
		0.99	3.983	2.492	1.759	1.7336	1.718
	8	0.90	1.959	1.325	1.073	1.0476	0.896
		0.95	4.003	2.154	1.595	1.5696	1.445
		0.99	4.937	2.875	2.101	2.0756	1.252
	12	0.90	3.196	0.948	0.773	0.7476	0.596
		0.95	6.419	2.299	1.571	1.5456	0.691
		0.99	8.958	2.625	1.397	1.3716	0.662
1.5	4	0.90	1.721	1.368	1.289	1.2636	0.881
		0.95	2.818	1.852	1.579	1.5536	1.494
		0.99	3.861	2.372	1.637	1.6116	1.596
	8	0.90	1.837	1.198	0.948	0.9226	0.774
		0.95	3.881	2.028	1.473	1.4476	1.323
		0.99	4.815	2.753	1.979	1.9536	1.131
	12	0.90	3.074	0.826	0.651	0.6256	0.474
		0.95	6.297	2.177	1.449	1.4236	0.569
		0.99	8.836	2.503	1.275	1.2496	0.547

TABLE 4 | Mean squared error values for different estimators when $n = 400$.

ϕ	p	ρ	BML	RBR	LBR	LTBR	TPBR
0.5	4	0.90	2.117	1.764	1.685	1.2596	1.277
		0.95	3.214	2.248	1.975	1.7496	1.891
		0.99	4.257	2.767	2.033	2.0076	1.992
	8	0.90	2.233	1.594	1.344	1.3186	1.172
		0.95	4.277	2.424	1.869	1.3436	1.719
		0.99	5.211	3.149	2.375	2.3496	1.527
	12	0.90	3.47	1.222	1.047	1.0216	0.871
		0.95	6.693	2.573	1.845	1.8196	0.965
		0.99	9.232	2.899	1.671	1.6456	0.938
1	4	0.90	1.809	1.457	1.377	1.3516	0.969
		0.95	2.908	1.94	1.667	1.6416	1.582
		0.99	3.949	2.458	1.725	1.6996	1.684
	8	0.90	1.925	1.291	1.039	1.0136	0.862
		0.95	3.969	2.12	1.561	1.5356	1.411
		0.99	4.903	2.841	2.067	2.0416	1.218
	12	0.90	3.162	0.914	0.739	0.7136	0.562
		0.95	6.385	2.265	1.537	1.5116	0.657
		0.99	8.924	2.591	1.363	1.3376	0.628
1.5	4	0.90	1.687	1.334	1.255	1.2296	0.847
		0.95	2.784	1.818	1.545	1.5196	1.462
		0.99	3.827	2.338	1.603	1.5776	1.562
	8	0.90	1.803	1.164	0.914	0.8886	0.742
		0.95	3.847	1.994	1.439	1.4136	1.289
		0.99	4.781	2.719	1.945	1.9196	1.097
	12	0.90	3.04	0.792	0.617	0.5916	0.448
		0.95	6.263	2.143	1.415	1.3896	0.535
		0.99	8.802	2.469	1.241	1.2156	0.513

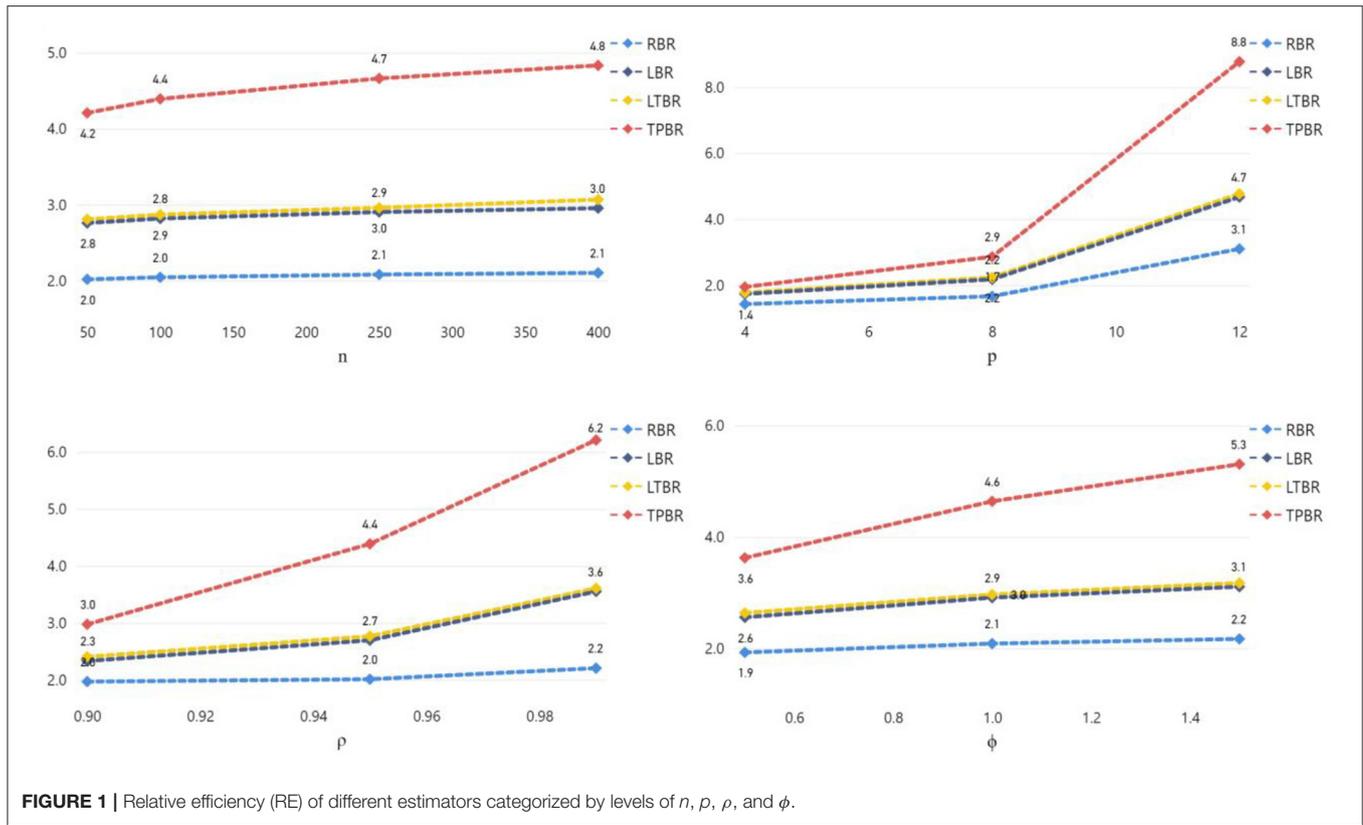


TABLE 5 | The estimated coefficients and MSE values for the used estimators (football data).

	BML	RBR	LBR	LTBR	TPBR
x_1	-0.01749	-0.01761	-0.026786	-0.002165	-0.003165
x_2	0.026057	0.026399	0.052970	0.000354	0.000254
x_3	0.030190	0.030276	0.036945	0.000701	-0.000807
x_4	-0.032857	-0.031889	0.043208	-0.000355	0.000323
x_5	-0.129230	-0.128710	-0.088372	-0.002704	-0.002633
x_6	1.643973	1.629061	0.472132	0.021849	0.028245
MSE	0.04345	0.018314	0.006583	0.005208	0.005085

TABLE 6 | The estimated coefficients and MSE values for the used estimators (gasoline yield data).

	BML	RBR	LBR	LTBR	TPBR
Gravity	-0.01749	-0.01761	-0.026786	-0.00571	-0.003165
Pressure	0.026057	0.026399	0.052970	0.04212	0.000254
Temp10	0.030190	0.030276	0.036945	0.04284	-0.000807
Temp	-0.032857	-0.031889	0.043208	0.00235	0.000323
MSE	0.04345	0.018314	0.006583	0.00537	0.005085

Relative Efficiency

Another comparative performance called relative efficiency (RE) can be utilized, it is calculated based on the MSE in Eq. (33) as follows [4, 39]:

$$RE(\hat{\beta}_S) = \frac{MSE(\hat{\beta}_{BML})}{MSE(\hat{\beta}_S)}, \quad (34)$$

where $\hat{\beta}_S$ denotes the estimators of RBR, LBR, LTBR, or TPBR. The RE results are shown in Figure 1.

Figure 1 shows that the RE of the four biased (RBR, LBR, LTBR, and TPBR) estimators were increased if the sample size (n), the number of explanatory variables (p), the degree

of correlation between explanatory variables (ρ), and/or the precision parameter value (ϕ) are increased. Moreover, we can observe that the TPBR estimator has higher RE values than the other estimators.

REAL DATA APPLICATIONS

In this section, we used two real data applications to investigate the advantage of our proposed (TPBR) estimator in different fields.

Football Spanish Data

We apply the proposed estimator to the football Spanish La Liga, season 2016–2017 [19]. The data contain 20 teams. The response variable is the proportion of won matches. The six considerable explanatory variables are: x_1 is the number of yellow cards, x_2 is

the number of red cards, x_3 is the total number of substitutions, x_4 is the number of matches with 2.5 goals on average, x_5 is the number of matches that ended with goals, and x_6 is the ratio of the goal scores to the number of matches.

First, to check whether there is a multicollinearity problem or not, the correlation matrix and condition number (CN) are used. Based on the correlation matrix among the six explanatory variables that are presented, displayed by Algamal and Abonazel [19]. It is obviously seen that there are correlations >0.82 between x_1 and x_6 , x_1 and x_4 , x_2 and x_4 , and x_4 and x_6 . Second, the condition number, $CN = \sqrt{\lambda_{\max}/\lambda_{\min}}$ of the data is 806.63 indicating the existence of multicollinearity. The estimated beta regression coefficients and MSE values for the BML, RBR, LBR, LTBR, and TPBR estimators are recorded in **Table 5**. From **Table 5**, it can be noted that the estimated coefficients of all estimators have the same signs; 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 estimator. But MSE values of RBR, LBR, LTBR, and TPBR estimators are lower than the BML estimator. Whereas, the MSE value of the TPBR estimator is the lowest.

Gasoline Yield Data

To further investigate the advantage of our proposed estimator (TPBR), we apply the TPBR estimator to the chemical dataset (gasoline yield data) which was originally obtained by Prater [42], and later used by the following authors: Ospina et al. [43] and Karlsson et al. [10]. The dataset contains 32 observations on the response and four explanatory variables. The variables in the study are described as follows: the dependent variable y is the proportion of crude oil after distillation and fractionation while the explanatory variables are crude oil gravity (Gravity), vapor pressure of crude oil (Pressure), temperature at which 10% of the crude oil has vaporized (Temp10), and temperature at which all petrol in the amount of crude oil vaporizes (Temp). Atkinson [44] analyzed this dataset using the linear regression model and observed some anomalies in the distribution of the error. Recently, Karlsson et al. [10] showed that the beta regression model is more suitable to model the data.

The CN for the dataset under study is 11,281.4, which signals severe multicollinearity. The estimated beta regression coefficients and MSE values for the used estimators are recorded in **Table 6**. From **Table 6**, it can be noted that the estimated coefficients of all estimators have the same signs. In addition, MSE values of RBR, LBR, LTBR, and TPBR estimators are lower

than the BML estimator. Whereas, the MSE value of the TPBR estimator is the lowest.

CONCLUSION

This article provided a two-parameter (TPBR) estimator for the beta regression model as a remedy for a multicollinearity problem. We proved, theoretically, that our proposed estimator is efficient than other biased estimators (ridge, Liu, and Liu-type estimators) suggested in the literature. Furthermore, a Monte Carlo simulation study was conducted to study the performance of the proposed estimator and ML, ridge, Liu, and Liu-type estimators. The results indicated that the proposed estimator outperforms these estimators, especially when there is a high-to-strong correlation between the explanatory variables. Finally, the benefit is shown by two empirical applications where the TPBR estimator performed well by decreasing the MSE compared with the ML, ridge, Liu, and Liu-type estimators.

For future work, for example, one can study the high-dimensional case in beta regression as an extension to Arashi et al. [45] or provide a robust biased estimation of beta regression as an extension to Awwad et al. [41] and Dawoud and Abonazel [40].

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: For first application: <https://www.laliga.es>. For second application: the R-package betareg.

AUTHOR CONTRIBUTIONS

MA: methodology, relative efficiency, interpreting the results, abstract, conclusions, writing—original draft, and final revision. ZA: simulation study and applications, interpreting the results, and revision. FA: methodology, interpreting the results, and revision. IT: introduction, methodology, and writing—original draft. All authors contributed to the article 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

- Hoerl AE, Kennard RW. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics*. (1970) 12:55–67. doi: 10.1080/00401706.1970.10488634
- Hoerl AE, Kennard RW. Ridge regression: applications to non-orthogonal problems. *Technometrics*. (1970) 12:69–82. doi: 10.1080/00401706.1970.10488635
- Rady EA, Abonazel MR, Taha IM. Ridge estimators for the negative binomial regression model with application. In: *The 53rd Annual Conference on Statistics, Computer Science, and Operation Research 3-5 Dec, 2018*. (2018). Giza: ISSR, Cairo University.
- Abonazel MR, Taha IM. Beta ridge regression estimators: simulation and application. *Commun Statist Simul Computat*. (2021) 1–13. doi: 10.1080/03610918.2021.1960373
- 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
- Alobaidi NN, Shamany RE, Algamal ZY. A new ridge estimator for the negative binomial regression model. *Thailand Statistician*. (2021) 19:116–25.
- Sami F Amin M, Butt MM. On the ridge estimation of the Conway-Maxwell Poisson regression model with multicollinearity: Methods and applications. *Concurr Computat Pract Exp*. (2021) 2021:6477. doi: 10.1002/cpe.6477

8. Liu K. A new class of biased estimate in linear regression. *Commun Statist Theor Method.* (1993) 22:393–402. doi: 10.1080/03610929308831027
9. Stein C. Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics.* Berkeley, CA: University of California Press. (1956). p. 197–206. doi: 10.1525/9780520313880-018
10. Karlsson P, Månsson K, Kibria BG. A Liu estimator for the beta regression model and its application to chemical data. *J Chemom.* (2020) 34:e3300. doi: 10.1002/cem.3300
11. Qasim M, Kibria BMG, Månsson K, Sjölander P. A new Poisson Liu regression estimator: method and application. *J Appl Stat.* (2020) 47:2258–71. doi: 10.1080/02664763.2019.1707485
12. Naveed K, Amin M, Afzal S, Qasim M. New shrinkage parameters for the inverse Gaussian Liu regression. *Commun Statist Theor Method.* (2020) 2020:1–21. doi: 10.1080/03610926.2020.1791339
13. Liu K. Using Liu-type estimator to combat collinearity. *Commun Statist Theor Method.* (2003) 32:1009–20. doi: 10.1081/STA-120019959
14. Liu K. More on Liu-type estimator in linear regression. *Commun Statist Theor Method.* (2004) 33:2723–33. doi: 10.1081/STA-200037930
15. Özkale MR, Kaciranlar S. The restricted and unrestricted two-parameter estimators. *Commun Stat Theor Method.* (2007) 36:2707–25. doi: 10.1080/03610920701386877
16. Li Y, Yang H. A new Liu-type estimator in linear regression model. *Stat Pap.* (2012) 53:427–37. doi: 10.1007/s00362-010-0349-y
17. Kurnaz FS, Akay KU. A new Liu-type estimator. *Stat Pap.* (2015) 56:495–517. doi: 10.1007/s00362-014-0594-6
18. Sahriman S, Koerniawan V. Liu-type regression in statistical downscaling models for forecasting monthly rainfall salt as producer regions in Pangkep regency. *J Phys Conf Ser.* (2019) 1341:092021. doi: 10.1088/1742-6596/1341/9/092021
19. Algamal ZY, Abonazel MR. Developing a Liu-type estimator in beta regression model. *Concurr Computat Pract Exp.* (2021) 2021:e6685. doi: 10.1002/cpe.6685
20. Huang J, Yang H. A two-parameter estimator in the negative binomial regression model. *J Stat Comput Simul.* (2014) 84:124–34. doi: 10.1080/00949655.2012.696648
21. Algamal ZY. Shrinkage estimators for gamma regression model. *Electr J Appl Stat Anal.* (2018) 11:253–68. doi: 10.1285/i20705948v11n1p253
22. Asar Y, Genç A. A new two-parameter estimator for the Poisson regression model. *Iran J Sci Technol Trans A Sci.* (2018) 42:793–803. doi: 10.1007/s40995-017-0174-4
23. Rady EA, Abonazel MR, Taha IM. A new biased estimator for zero-inflated count regression models. *J Modern Appl Stat Method.* (2019). Available online at: https://www.researchgate.net/publication/337155202_A_New_Biased_Estimator_for_Zero-Inflated_Count_Regression_Models
24. Rady EA, Abonazel MR, Taha IM. New shrinkage parameters for Liu-type zero inflated negative binomial estimator. In: *The 54th Annual Conference on Statistics, Computer Science, and Operation Research 3-5 Dec, 2019.* Giza: FGSSR, Cairo University (2019).
25. Çetinkaya M, Kaçiranlar S. Improved two-parameter estimators for the negative binomial and Poisson regression models. *J Stat Comput Simul.* (2019) 89:2645–60. doi: 10.1080/00949655.2019.1628235
26. Abonazel MR, Farghali RA. Liu-type multinomial logistic estimator. *Sankhya B.* (2019) 81:203–25. doi: 10.1007/s13571-018-0171-4
27. Akram MN, Amin M, Qasim M. A new Liu-type estimator for the Inverse Gaussian Regression Model. *J Stat Comput Simul.* (2020) 90:1153–72. doi: 10.1080/00949655.2020.1718150
28. Lukman AF, Aladeitan B, Ayinde K, Abonazel MR. Modified ridge-type for the Poisson regression model: simulation and application. *J Appl Stat.* (2021) 2021:1–13. doi: 10.1080/02664763.2021.1889998
29. Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. *J Appl Stat.* (2004) 31:799–815. doi: 10.1080/0266476042000214501
30. Smithson M, Verkuilen J. A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol Method.* (2006) 11:54–71. doi: 10.1037/1082-989X.11.1.54
31. Cribari-Neto F, Zeileis A. Beta regression in R. *J Stat Softw.* (2010) 34:1–24. doi: 10.18637/jss.v034.i02
32. Bayer FM, Cribari-Neto F. Model selection criteria in beta regression with varying dispersion. *Commun Stat Simul Computat.* (2017) 46:729–46. doi: 10.1080/03610918.2014.977918
33. Espinheira PL, Ferrari SL, Cribari-Neto F. On beta regression residuals. *J Appl Stat.* (2008) 35:407–19. doi: 10.1080/02664760701834931
34. Espinheira PL, da Silva LCM, Silva ADO. Prediction measures in beta regression models. *arXiv preprint.* (2015). arXiv: 1501.04830.
35. Espinheira PL, da Silva LCM, Silva ADO, Ospina R. Model selection criteria on beta regression for machine learning. *Machine Learn Knowl Extr.* (2019) 1:427–49. doi: 10.3390/make1010026
36. Farebrother RW. Further results on the mean square error of ridge regression. *J Royal Stat Soc Ser B.* (1976) 38:248–50. doi: 10.1111/j.2517-6161.1976.tb01588.x
37. Hoerl AE, Kennard RW, Baldwin KF. Ridge regression: some simulations. *Commun Stat Theor Method.* (1975) 4:105–23. doi: 10.1080/03610917508548342
38. Kibria BMG. Performance of some new ridge regression estimators. *Commun Stat Simul Computat.* (2003) 32:419–35. doi: 10.1081/SAC-120017499
39. 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 Computat.* (2021) 1–16. doi: 10.1080/03610918.2021.1934023
40. 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
41. 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. *Concurr Computat Pract Exp.* (2021) 2021:e6779. doi: 10.1002/cpe.6779
42. Prater NH. Estimate gasoline yields from crude. *PetroLium Refiner.* (1956) 35:236–8.
43. Ospina R, Cribari-Neto F, Vasconcellos KL. Improved point and interval estimation for a beta regression model. *Comput Stat Data Anal.* (2006) 51:960–81. doi: 10.1016/j.csda.2005.10.002
44. Atkinson AC. *Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis.* New York, NY: Oxford University Press (1985).
45. Arashi M, Norouzirad M, Mamode Khan N. A high-dimensional counterpart for the ridge estimator in multicollinear situations. *Mathematics.* (2021) 9:3057. doi: 10.3390/math9233057

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, Algamal, Awwad and Taha. 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.