Mathematics

AIMS Mathematics, 11(1): 543–557.
DOI: 10.3934/math.2026023
Received: 14 October 2025
Revised: 30 November 2025
Accepted: 10 December 2025
Published: 08 January 2026

https://www.aimspress.com/journal/Math

Research article

Modified two-parameter ridge estimator for the Beta logistic model to mitigate multicollinearity

Asma Ahmad Alzahrani*

Department of Mathematics, Faculty of Science, Al-Baha University, Al-Baha, Saudi Arabia

* **Correspondence:** Email: asmaalzahrani@bu.edu.sa.

Abstract: Multicollinearity poses significant challenges to parameter estimation in regression models, often undermining the reliability of traditional methods like maximum likelihood estimation (MLE). This study addresses the issue by evaluating and enhancing regularization techniques, specifically the ridge-based regression (RBR), the Liu regression estimator (LRE), and a modified two-parameter ridge estimator (MTPRE) within the context of the Beta regression model (BRM). However, the selection of appropriate shrinkage parameters remains a persistent challenge. To address this limitation, we propose an MTPRE that eliminates the need for shrinkage parameter tuning, thereby improving estimation stability and accuracy. Through extensive simulation studies, the MTPRE consistently outperformed the MLE, RBR, and LRE under severe multicollinearity based on the mean squared error (MSE). The effectiveness of proposed estimators was further validated using a real-world gasoline yield dataset having multicollinearity issues, where the MTPRE demonstrated superior predictive accuracy and estimation precision. These results highlight the potential of the MTPRE as a practical and efficient method for handling multicollinearity in regression analysis.

Keywords: multicollinearity; Beta regression model; ridge estimator; maximum likelihood method; monte Carlo simulation
Mathematics Subject Classification: 62F10, 62G05, 62J07

1. Introduction

The Beta regression model (BRM) is a powerful statistical tool, particularly well-suited for modeling data within the continuous interval (0,1). It has become widely used in disciplines such as economics, ecology, and epidemiology, where the response variable typically represents proportions, rates, or percentages [1,2]. However, putting it into practice can be tricky, especially when choosing the right regularization method, like ridge regression, to handle problems like strong multicollinearity and overfitting [3].

In Beta regression, parameter estimation can be done using either maximum likelihood estimation (MLE) or a Bayesian approach. The MLE works by maximizing the likelihood of the observed data, providing efficient and consistent estimates. On the other hand, the Bayesian approach incorporates prior knowledge, combining it with the data to form a posterior distribution that integrates both sources of information [4–6]. For the large sample, the MLE provides efficient estimates as compared to Bayesian paradigm, which is often preferred for small samples parameter estimates. While linear and logistic regression models each have their unique strengths, they also come with limitations. The drawback of the BRM is its sensitivity to outliers, which can produce significantly unreliable parameter estimates and affect the overall model fit [7]. Moreover, the model's performance is contingent upon the assumption that the predictor variables are independent. However, in practical applications, this assumption is often violated, as most predictor variables are dependent, so this results in multicollinearity issues. When predictor variables are strongly correlated, it can cause standard errors to increase and lead to unstable, and unreliable coefficient estimates, as highlighted in recent studies [8–10].

To address these challenges, ridge estimators are widely used to mitigate multicollinearity issues in regression models. In ridge regression, a penalty term or shrinkage parameter is controlled by a biasing parameter (k), which helps reduce the risk of overfitting and obtains more reliable parameter estimates of the regression model in the presence of severe multicollinearity [11]. Amin et al. [12] highlighted the significant challenge that multicollinearity poses in logistic regression models, which can undermine the predictive accuracy of the model.

To mitigate these issues, numerous studies have proposed strategies for selecting the optimal values of k and q to obtain reliable estimates and minimize the estimated mean squared error (MSE) (see [13–15]). Most recently, Hammad et al. [16] modified the two-parameter ridge estimators for severe multicollinearity data. However, the literature indicates that no single estimator consistently performs well across all levels of collinearity. Choosing the best shrinkage parameter is still a tough challenge and an active area of research, especially in Beta regression models. These models are more complex because their likelihood structure is different from that of basic linear models.

To address this gap, we introduced a new modified two-parameter ridge estimator (MTPRE) for the BRM to improve handling of multicoolinearity issues. This new estimator builds on existing regularization methods by adding extra shrinkage parameters. This helps to better handle problems of severe multicollinearity. The MTPRE improves both the flexibility and accuracy of parameter estimation, making Beta regression more stable and reliable, particularly when multicollinearity is a significant issue. This paper is structured as follows: Section 2 introduces the proposed estimator and explains the shrinkage parameter selection; Section 3 presents Monte Carlo simulation results; Section 4 applies the method to real data; and finally, Section 5 concludes the study.

2. Materials and methods

In this section, we discuss the BRM, the ridge and Liu estimators, two-parameter estimators, and the newly proposed estimators $(k$ and $q)$.

2.1. Beta regression model

The BRM relies on MLE for parameter estimation, which, while offering good asymptotic properties, is highly sensitive to multicollinearity. When predictor variables are highly correlated, the model's likelihood function flattens, leading to unstable parameter estimates with large standard errors. Furthermore, due to the nonlinear link function and the bounded nature of the dependent variable in Beta regression, multicollinearity poses even more severe challenges within this framework. Within the framework of the BRM, a functional relationship between the mean of the dependent variable and a set of covariates can be established by specifying an appropriate link function [2]. The model also introduces a precision parameter, which represents the inverse of the variance term.

Suppose Y is a random variable with independent observations $y_1, y_2, \dots, y_n$, and it is assumed to follow a Beta distribution with two parameters $\mathrm{g} > 0$ and $h > 0$, denoted as $\mathrm{Beta}(\mathrm{g}, h)$:

$$f(y; g, h) = \frac{\Gamma(\mathrm{g+h})}{\Gamma(\mathrm{g})\Gamma(\mathrm{h})} y^{g-1}(1-y)^{h-1}, \quad 0 < y < 1. \tag{2.1}$$

The Gamma function denoted by $\Gamma(.)$ with the mean and variance of Gamma distribution is given by

$$\text{Mean} = \frac{\mathrm{g}}{\mathrm{g+h}} \text{ and } \text{Variance} = \frac{\mathrm{g\,h}}{(\mathrm{g+h})^2(\mathrm{g}+h+1)}.$$

The provided reparameterization approach for constructing a model with Beta distribution responses based on Eq (2.1) is as follows:

$$\theta = \frac{\mathrm{g}}{\mathrm{g+h}} \text{ and } \mu(\mathrm{g} + h) = \mathrm{g}.$$

By re-parameterizing the Beta distribution parameters, where $\theta z = \mathrm{g}$ and $z - \theta z = h$, the probability distribution of Beta can be written as

$$f(y; \theta, z) = \frac{\Gamma(\mathrm{z})}{\Gamma(\theta z)\Gamma(1-\theta)} y^{\theta z-1}(1-y)^{z-\theta z-1}. \tag{2.2}$$

In Beta regression, the outcome variable y is restricted to lie within the range $0 < y < 1$. The model incorporates two key parameters: a scaling factor μ, constrained to the interval $(0,1)$, and a precision parameter z, which must be positive. The value of z is derived using the Gamma function $\Gamma(\cdot)$. The precision parameter z plays a central role in adjusting model variability and is defined as

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

The Beta distribution has the following mean and variance:

$$\text{Mean} = \theta, \text{ Variance} = \theta(1-\theta) = \sigma^2.$$

The regression coefficient μ, which depends on the covariates, and the logit link function are commonly used. It can be written as

$$\mathrm{g}(\mu_i) = \log\left(\frac{\theta_i}{1-\theta_i}\right) = \boldsymbol{X}_i^T \boldsymbol{\mu} = \eta_{\mathrm{I}}, \tag{2.3}$$

where $\boldsymbol{\mu} = (\mu_{i1}, \mu_{i2}, \dots, \mu_{ip})^T$, $\boldsymbol{\mu} \in R^{1\times p}$ is the vector of unknown parameters, and $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ represents the vector of explanatory variables for the i^{th} observation. The linear component of the regression model is η_i. The log-likelihood function of the model is given below:

$$L(\theta_i, z_i; y_i) = \sum_{i=1}^{n}\{\log\Gamma(z) - log\Gamma(\theta_i(z)) - \log((1-\theta_i)\,(z)) + (\theta_i(z)-1)\log(y_i) + (1-\theta_i)(z) + ((1-\theta_i)(z)-1)\log(1-y_i)\}. \quad (2.4)$$

Parameter estimates differentiating w.r.t. μ, as shown in Eq (2.3):

$$S(\mu) = \boldsymbol{Z}\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{\Lambda}\,(\mathbf{y}^* - \boldsymbol{\theta}^*), \quad (2.5)$$

where $\boldsymbol{\Lambda} = \left(\frac{1}{g^{T}(\theta_1)}, \frac{1}{g^{T}(\theta_2)}, \dots, \frac{1}{g^{T}(\theta_n)}\right)$, $\boldsymbol{y}^* = (y_1^*, y_2^*, \dots, y_n^*)$, $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*, \dots, \theta_n^*)$ is a diagonal matrix, $\boldsymbol{y}^*$ is the vector of transformed response variables, and θ^* is the transformed vector of predicted values. y_i^* is the logit of y_i, and θ_i^* represents the transformed predictions:

$$y_i^* = log\left(\frac{y_i}{1-y_i}\right)$$

and

$$\theta_i^* = \psi(\theta_{iz}) - \psi(1-\theta_i)z.$$

Here, $\psi(\cdot)$ is the Digamma function. The μ is calculated using optimization techniques, such as weighted regression updates or the Fisher algorithm. The update rule is as follows:

$$\mu^{(r+1)} = \mu^{r} + \left(I_{\mu\mu}^{r}\right)^{-1} + S_{\mu}^{(r)}(\mu).$$

Here, $S_{\mu}^{(r)}$ represents the score function at iteration r, and $I_{\mu\mu}^{r}$ is the information matrix at the same iteration, as defined in Eq (2.4). The least-squares method is usually applied to compute the initial values for the parameter μ, while the precision parameters are initialized as follows:

$$\hat{z}_i = \frac{\hat{\theta}_\iota(1-\hat{\theta}_\iota)}{\hat{\sigma}_i^2}. \quad (2.6)$$

The estimates $\widehat{\theta}_{\iota i}$ and $\hat{\sigma}_i^2$ are obtained through linear regression. The iteration process continues until the changes between consecutive estimates are smaller than the small predefined threshold, indicating that the algorithm has converged, with $\boldsymbol{X} \in R^{n\times p}$. The final step is to calculate the MLE of parameter μ.

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{MLE}} = \left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\hat{\boldsymbol{z}}, \quad (2.7)$$

$$\widehat{\boldsymbol{W}} = diag(\widehat{W}_1, \widehat{W}_2, \dots, \widehat{W}_n),$$

$$\hat{\boldsymbol{z}} = \widehat{\boldsymbol{\eta}} + \widehat{\mathbf{W}}^{-1}\widehat{\boldsymbol{\Lambda}}(\boldsymbol{y}^* - \boldsymbol{\theta}^*),$$

where

$$\widehat{W}_i = \frac{(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\left\{\Psi^{\mathrm{T}}\left(\frac{(\theta_i^T(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right) + \Psi^{\mathrm{T}}\left(\frac{(1-\theta_i)(1-\hat{\sigma}^2)}{\hat{\sigma}^2}\right)\right\}\frac{1}{\{g^T(\hat{\theta}_i)\}^2}.$$

The MLE for μ follows an asymptotically normal distribution. Under regularity conditions, the expected value of the estimator satisfies

$$E(\hat{\mu}_{MLE}) \to \mu \text{ as } n \to \infty.$$

Furthermore, the asymptotic covariance matrix of μ_{MLE} is expressed as

$$\mathrm{Cov}(\hat{\boldsymbol{\mu}}_{\mathrm{MLE}}) = \frac{1}{z}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-\mathbf{1}}. \quad (2.8)$$

Accordingly, the asymptotic trace of the MSE μ_{MLE} is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\mu}}_{\mathrm{MLE}}) = tr\frac{1}{z}\left[\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\right)^{-\mathbf{1}}\right]. \quad (2.9)$$

2.2. Ridge-type biased regression estimators

In the BRM, explanatory variables are assumed to be uncorrelated, similar to linear regression. When this assumption is violated, multicollinearity occurs, indicating strong dependencies among predictors. Although multicollinearity does not bias coefficient estimates, it inflates their variances, widens confidence intervals, and can cause statistically significant variables to appear non-significant. This study builds on the foundational work of [11], who introduced penalty term for linear models. Recently, the researchers in [8] and [17] modified the ridge-based regression (RBR) estimator to a ridge-type biased estimator used to address multicollinearity and improve estimation in weighted linear regression models. It is given by

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{RBR}} = \left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{kI}\right)^{-\mathbf{1}}\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\hat{\boldsymbol{z}},\ k > 0. \quad (2.10)$$

The bias introduced due to the shrinkage parameter k in the RBR is given by

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{RBR}} = -\boldsymbol{k}\left(\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{kI}\right)^{-\mathbf{1}}\hat{\boldsymbol{\mu}}_{\boldsymbol{MLE}}. \quad (2.11)$$

Suppose the matrix $\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}$ is symmetric and positive semi-definite. Let $\boldsymbol{Q} =$ matrix of orthonormal eigenvectors and $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \dots, \lambda_p)$ diagonal matrix of eigenvalues, ordered $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Then

$$\boldsymbol{Q}^{\boldsymbol{T}}\boldsymbol{X}^{\boldsymbol{T}}\widehat{\boldsymbol{W}}\boldsymbol{X}\boldsymbol{Q} = \boldsymbol{\Lambda} \text{ and } \boldsymbol{\mu} = \boldsymbol{Q}^{\top}\boldsymbol{\gamma}.$$

The MSE of the RBR estimator, incorporated with both variance and bias term, is given by

$$\begin{aligned} MSE(\hat{\mu}_{RBR}) &= Cov(\hat{\mu}_{RBR}) + \left(Bias(\hat{\mu}_{RBR})\right)^T(\hat{\mu}_{RBR}) \\ &= \frac{\mathbf{1}}{\boldsymbol{z}}\left(\boldsymbol{Q}\boldsymbol{\Lambda}_{\boldsymbol{k}}^{-\mathbf{1}}\boldsymbol{\Lambda}\,\boldsymbol{\Lambda}_{\boldsymbol{k}}^{-\mathbf{1}}\boldsymbol{Q}^{\boldsymbol{T}} + \boldsymbol{k}^{\mathbf{2}}\boldsymbol{Q}\,\boldsymbol{\Lambda}_{\boldsymbol{k}}^{-\mathbf{1}}\boldsymbol{\mu}\,\boldsymbol{\mu}^{\boldsymbol{T}}\boldsymbol{\Lambda}_{\boldsymbol{k}}^{-\mathbf{1}}\boldsymbol{Q}^{\boldsymbol{T}}\right). \end{aligned} \quad (2.12)$$

Construct $(\boldsymbol{\Lambda_k})$ with $(\boldsymbol{\Lambda_k})_{\boldsymbol{ii}} = \lambda_i + k$, and $(\boldsymbol{\Lambda_k})_{\boldsymbol{ij}}$ for $i \neq j$.

Trace-based calculation for the MSE is provided in the following expression:

$$TMSE(\widehat{\boldsymbol{\mu}}_{RBR}) = tr(MSE(\widehat{\boldsymbol{\mu}}_{RBR})),$$
$$tr(MSE(\widehat{\boldsymbol{\mu}}_{RBR})) = \frac{1}{z}\sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j+k)} + k^2 \frac{\mu_j^2}{((\lambda_j+k))^2}. \tag{2.13}$$

In Eq (2.11), the first term reflects the asymptotic variance component, whereas the second term reflects bias contribution.

2.3. Liu regression estimator

Liu [18] introduced a Liu estimator to obtain stable parameter estimates in the presence of severe multicollinear predictors within linear regression.

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{LRE}} = \left(\boldsymbol{X^T}\widehat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{I}\right)^{-\boldsymbol{1}}\left(\boldsymbol{X^T}\widehat{\boldsymbol{W}}\boldsymbol{X} + q\boldsymbol{I}\right)^{-\boldsymbol{1}}\widehat{\boldsymbol{\mu}}_{\boldsymbol{MLE}}, \quad 0 < q < 1, \tag{2.14}$$

where q is the Liu parameter. Additionally, the biasing parameter q, as mathematically defined in Eq (2.15), is adopted from the approach proposed by Månsson et al. [19]. The Liu estimator is expressed as follows:

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{LRE}} = \left(\boldsymbol{X^T}\widehat{\boldsymbol{W}}\boldsymbol{X} + \boldsymbol{I}\right)^{-\boldsymbol{1}}(q-1)\widehat{\boldsymbol{\mu}}_{MLE}. \tag{2.15}$$

The MSE is expressed as follows:

$$MSE(\hat{\mu}_{LRE}) = Cov(\widehat{\boldsymbol{\mu}}_{LRE}) + (\widehat{\boldsymbol{\mu}}_{LRE})Bias(\widehat{\boldsymbol{\mu}}_{LRE})^T$$
$$= \frac{1}{z}[\boldsymbol{Q}\boldsymbol{\Lambda}_{\boldsymbol{L}}^{-\boldsymbol{1}}\,\boldsymbol{\Lambda}_{\boldsymbol{q}}\boldsymbol{\Lambda}^{-\boldsymbol{1}}\boldsymbol{\Lambda}_{\mathbf{q}}\,\boldsymbol{\Lambda}^{-\boldsymbol{1}}\boldsymbol{Q^T}](1-q)^{\boldsymbol{2}}\,\boldsymbol{Q}\,\boldsymbol{\Lambda}_{\boldsymbol{L}}^{-\boldsymbol{1}}\boldsymbol{\mu}\,\boldsymbol{\mu^T}\boldsymbol{\Lambda}_{\boldsymbol{L}}^{-\boldsymbol{1}}\boldsymbol{Q^T}, \tag{2.16}$$

where $\boldsymbol{\Lambda_L} = diag(\lambda_1 + 1, \lambda_2 + 1, \ldots, \lambda_p + 1)$ and $\boldsymbol{\Lambda_q} = diag(\lambda_1 + q, \lambda_2 + q, \ldots, \lambda_p + q)$.

The MSE corresponds to the trace of Liu regression estimator, as shown below:

$$TMSE(\widehat{\boldsymbol{\mu}}_{LRE}) = tr(MSE(\widehat{\boldsymbol{\mu}}_{LRE})),$$
$$tr(MSE(\hat{\mu}_{LRE})) = \frac{1}{z}\sum_{j=1}^{p}\frac{(\lambda_j+q)^2}{\lambda_j\,(\lambda_j+1)^2} + \sum_{j=1}^{p}\frac{\mu_j^2\,z\,(1-q)^2}{((\lambda_j+1))^2} = \frac{1}{z}\sum_{j=1}^{p}\left(\frac{(\lambda_j+q)^2}{\lambda_j\,(\lambda_j+1)^2} + \frac{\mu_j^2\,z\,(1-q)^2}{((\lambda_j+1))^2}\right). \tag{2.17}$$

2.4. Two-parameter Beta regression estimator

In this section, we present an extended discussion of the two-parameter estimator initially proposed by Algamal and Abonazel [20]. The estimator is modified and reformulated to suit the framework of the BRM. This modification is designed to mitigate the problem of multicollinearity and to improve both robustness and estimation accuracy. The resulting approach is referred to as the two-parameter regression estimator (TPRE), and the expression is

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{TPRE}} = \left(\boldsymbol{X^T\widehat{W}X} + \boldsymbol{kI}\right)^{-\mathbf{1}}\left(\boldsymbol{X^T\widehat{W}X} - kq\boldsymbol{I}\right)^{-\mathbf{1}}\hat{\boldsymbol{\mu}}_{\boldsymbol{MLE}},\ k > 0\ \ and\ 0 < q < 1. \tag{2.18}$$

The TPRE bias is

$$\boldsymbol{Bias}(\hat{\boldsymbol{\mu}}_{\boldsymbol{TPRE}}) = k(q-1)\left(\boldsymbol{X^T\widehat{W}X} + \boldsymbol{kI}\right)^{-\mathbf{1}}\hat{\boldsymbol{\mu}}_{\boldsymbol{MLE}}. \tag{2.19}$$

The MSE of the TPRE can be derived as

$$\begin{aligned} MSE(\hat{\boldsymbol{\mu}}_{TPRE}) &= Cov(\hat{\boldsymbol{\mu}}_{TPRE}) + Bias(\hat{\boldsymbol{\mu}}_{TPRE})Bias(\hat{\boldsymbol{\mu}}_{TPRE}) \\ &= \frac{1}{z}\left[\boldsymbol{Q\Lambda_k^{-1}}\ \boldsymbol{\Lambda^{-1}\Lambda_{kd}\Lambda_k^{-1}Q^T}\right] + k^2\ \boldsymbol{Q}\ \boldsymbol{\Lambda_k^{-1}}\boldsymbol{\mu}\ \boldsymbol{\mu^T\Lambda_k^{-1}Q^T}, \end{aligned} \tag{2.20}$$

where $\boldsymbol{\Lambda_{kq}} = \operatorname{diag}(\lambda_1 + kq, \lambda_2 + kq, \ldots, \lambda_p + kq)$.

The MSE of the TPRE is given below:

$$TMSE(\hat{\boldsymbol{\mu}}_{TPRE}) = tr(MSE(\hat{\boldsymbol{\mu}}_{TPRE})),$$

$$MSE(\hat{\mu}_{TPRE}) = \frac{1}{z}\sum_{j=1}^{p}\left(\frac{(\lambda_j+kd)^2}{\lambda_j\,(\lambda_j+k)^2} + \frac{\mu_j^2\,zk^2\,(1-q)^2}{((\lambda_j+k)^2}\right). \tag{2.21}$$

2.5. Two-parameter ridge estimator

Although prior studies have introduced various basic shrinkage-based techniques, a critical limitation remains in the selection of an optimal ridge parameter. Hoerl and Kennard [11] introduced shrinkage parameter k for enhancing the severe multicollinearity and increasing the precision of the estimator. It is mathematically defined as

$$k_{opt} = \frac{\hat{\sigma}^2}{\hat{\mu}_{\max}^2}, \tag{2.22}$$

where $\hat{\mu}_{\max}^2 = max(\hat{\mu}_1^2,\ \hat{\mu}_2^2, \ldots, \hat{\mu}_p^2)$.

Equation (2.22) defines k_{opt} as a penalty term, chosen to balance bias and variance, thereby improving the accuracy of the estimator.

Lukman et al. [21] introduced shrinkage parameter for Poisson distribution, mathematically defined as below:

$$k_1 = \max\left(0, min\left(\frac{\lambda_j}{1+\lambda_j\hat{\mu}_j^2}\right)\right., \tag{2.23}$$

$$k_2 = \sqrt{k_1}, \tag{2.24}$$

$$q = \left(0, max\left(\frac{\hat{\mu}_{max}^2 - 1}{\frac{1}{\hat{\lambda}_{max}} + \hat{\mu}_{max}^2}\right)\right). \tag{2.25}$$

Equation (2.23) defines k_1 to mitigate the impact of extreme eigenvalues, ensuring the estimator remains stable in the presence of multicollinearity. In Eq (2.25), the second bias estimator reduces the effects of severe multicollinearity while improving the estimator's overall accuracy.

The selection of k and q, plays a crucial role in reducing the adverse effects of multicollinearity while enhancing the precision and accuracy of the estimator.

We introduce a modified two-parameter ridge estimator for the TPRE, defined by the biasing parameters k and q, within the BRM framework, referred to in this study as the MTPRE. The selection of k follows the formulas in Eqs (2.26)–(2.28), while q is determined using Eq (2.25):

$$k_3 = \frac{\min(\lambda_j) + \frac{1}{p}\sum_{i=1}^{p}\lambda_j}{1+\exp(\lambda_j \hat{\mu}_j^2)}, \tag{2.26}$$

$$k_4 = (0, k_3), \tag{2.27}$$

$$k_5 = \sqrt{k_4}. \tag{2.28}$$

Equation (2.26) defines k_3 based on minimum and average of the eigenvalues, applies balanced shrinkage, and ensures reliable estimates. Equation (2.27) ensures the shrinkage and keeps within a reasonable range to maintaining the stability of the models. In Eq (2.28), square root of k_4 fine-tunes the shrinkage, helping to improve the precision of the estimates.

3. Simulation study

This section presents computational experiments carried out using random sampling techniques to assess how the new proposed method compares with existing approaches. The evaluation of both the maximum likelihood method and competing techniques was carried out using the MSE as the criterion.

3.1. Simulation technique

Predictor variables x_{ij} were formulated using the correlation-based structure in Eq (3.1):

$$X_{ij} = \sqrt{(1-\rho^2)} \cdot w_{ij} + \rho \cdot w_{ip}, \quad i = 1,2,\ldots,n, and\ j = 1,2,\ldots,p. \tag{3.1}$$

This method was used by researchers [22] and [23] to generate the predictor variables, where $\rho = (0.8, 0.9, 0.95, 0.99)$ represents different levels of correlation between the predictors, and w_{ij} are independently drawn from a standard normal distribution $N(0,1)$. The response variable y_i follows a Beta distribution:

$$y_i \sim \text{Beta}(\theta_i, z),$$

where θ_i is defined by the logistic transformation of the linear predictor:

$$\theta_i = \frac{\exp(x_i^T \mu)}{1+\exp(x_i^T \mu)},$$

where z represents the precision parameter of the Beta distribution. For performance comparison, we considered two values of z: 0.5 and 1.5. The true coefficient vector $\boldsymbol{\mu} = (\mu_1, \mu_2 \ldots, \mu_p)^{\mathrm{T}}$ was constrained such that $\boldsymbol{\mu}^T\boldsymbol{\mu} = \mathbf{1}$. Various sample sizes and numbers of predictors were evaluated with $n = 20, 50, 80, 100$ and $p = 4,6,10$, where n represents the sample size and p represents the number of predictors. The MSE of the estimators was computed as

$$\mathbf{MSE}(\widehat{\boldsymbol{\mu}}) = \frac{\mathbf{1}}{M}\sum_{j=1}^{M}(\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu})^{\mathbf{T}}(\widehat{\boldsymbol{\mu}}_j - \boldsymbol{\mu}). \tag{3.2}$$

Monte Carlo simulations with $M = 1{,}000$ replications were carried out in R to assess the MSE across different values of ρ, n, and p. The results are summarized in Tables 1–4, which report the MSEs for the proposed estimator alongside competing methods under these settings. All computations were performed using R version 4.1.0. A detailed discussion of the findings is provided in the subsequent section.

Table 1. The MSE of proposed, MLE, and other estimators ($n = 20$; varying p and ρ).

p	ρ	z	MLE	RBR	LRE	$MTPRE$
4	0.8	0.5	1.89337927	0.68874723	0.61263832	0.403636
	0.9		6.60476849	1.6506358	1.17118708	0.43076697
	0.95		5.4875001	2.0688279	0.9716932	0.36646374
	0.99		27.8779952	35.0360272	1.93396569	0.41355193
	0.8	1.5	0.82360435	0.45041869	0.42965759	0.36360744
	0.9		1.32971048	0.57949174	0.49230068	0.35198521
	0.95		2.89378272	1.31766411	0.70087545	0.35020277
	0.99		13.4932132	12.4768994	0.97676757	0.36144394
6	0.8	0.5	6.05544531	1.1654949	0.90649226	0.47748835
	0.9		11.7621962	4.13828861	2.14314128	0.4743637
	0.95		25.2808742	9.41112987	2.62505304	0.4091923
	0.99		140.816216	209.390747	7.69295661	0.72743087
	0.8	1.5	1.11371561	0.49784088	0.45495709	0.35731557
	0.9		2.18292544	0.92230567	0.64376386	0.37356781
	0.95		4.77941975	2.04961351	0.85484047	0.35471643
	0.99		28.6380008	36.6854303	1.50273537	0.37911
10	0.8	0.5	119969.198	178633.102	19449.90	14124.00
	0.9		14366291.2	3960328.8	1576728	29807.02
	0.95		5568129.67	2656129.34	461092.8	127682.7
	0.99		18414302.2	36023310	670961.13	26299.27
	0.8	1.5	278272.3	16038.4	124912.6	33345.25
	0.9		151992.3	53072.27	24836.27	24943.26
	0.95		6916927.23	1096728.1	269726.40	43486.24
	0.99		194712902	1007830	675742.137	22998.26

Table 2. The MSE of proposed, MLE, and other estimators ($n = 50$; varying p and ρ).

p	ρ	z	MLE	RBR	LRE	$MTPRE$
4	0.8	0.5	0.37530089	0.39796272	0.39641556	0.37040818
	0.9		0.66600608	0.42991829	0.4206935	0.34325317
	0.95		1.39138414	0.610477	0.54818297	0.33976427
	0.99		6.75443344	3.49531244	1.12329186	0.36065323
	0.8	1.5	0.20200623	0.37717701	0.37631954	0.36125222
	0.9		0.41656468	0.42155962	0.41512719	0.35632064
	0.95		0.77260661	0.44895677	0.42508437	0.34138058
	0.99		3.67338019	1.64883831	0.64287705	0.33840556
6	0.8	0.5	0.57995449	0.39857752	0.39616526	0.36281025
	0.9		0.96186938	0.4819198	0.46627117	0.35076195
	0.95		1.98699333	0.79581885	0.68030995	0.34744211
	0.99		9.36501563	5.46653292	1.42244086	0.34434936
	0.8	1.5	0.28161166	0.37000208	0.36875896	0.35188949
	0.9		0.49561963	0.40362481	0.39603953	0.34122005
	0.95		1.01214215	0.52992831	0.48178349	0.33848651
	0.99		5.00873657	2.6296709	0.7971724	0.34983953
10	0.8	0.5	2.69036589	0.51721547	0.50899907	0.42534169
	0.9		3.09378133	0.68623015	0.63226182	0.37590749
	0.95		4.99771177	1.17937098	0.91632546	0.36191942
	0.99		23.3610857	14.1262585	2.59223146	0.35620088
	0.8	1.5	0.46391046	0.3886185	0.38585499	0.35968497
	0.9		0.808654	0.44221776	0.42657054	0.34547417
	0.95		1.6010262	0.61448474	0.51995144	0.33415129
	0.99		7.7745318	4.28764085	1.00605329	0.33688014

Table 3. The MSE of proposed, MLE, and other stimators ($n = 80$; varying p and ρ).

p	ρ	z	MLE	RBR	LRE	$MTPRE$
4	0.8	0.5	0.22260502	0.37476892	0.37440858	0.36146085
	0.9		0.37137154	0.38471557	0.38301261	0.34812563
	0.95		0.7359284	0.43739645	0.42634099	0.33649565
	0.99		3.96333983	1.83887135	0.96429088	0.36322992
	0.8	1.5	0.10502935	0.35341575	0.35323885	0.34645556
	0.9		0.22579774	0.37071801	0.36954543	0.34440079
	0.95		0.44653584	0.42090666	0.41273033	0.34521269
	0.99		2.08362177	0.89627699	0.57193671	0.34388531
6	0.8	0.5	0.26186357	0.36653408	0.36613102	0.35329268
	0.9		0.46544069	0.39765232	0.39510771	0.35071086
	0.95		0.99057811	0.5084249	0.48929845	0.35040366
	0.99		4.54539868	1.95563911	0.97321437	0.34992413

Continued on next page

p	ρ	z	MLE	RBR	LRE	$MTPRE$
6	0.8	1.5	0.14112583	0.35459398	0.35433144	0.3460298
	0.9		0.27715309	0.38407685	0.38235334	0.35267154
	0.95		0.54222944	0.42797836	0.41744324	0.34283269
	0.99		2.75865622	1.16028156	0.66232048	0.34657035
10	0.8	0.5	0.41718717	0.37891103	0.37814471	0.35994101
	0.9		0.83910061	0.43283971	0.42761476	0.35669988
	0.95		1.51546087	0.55679443	0.52198017	0.3385074
	0.99		8.2165454	3.7231466	1.47922649	0.33786217
	0.8	1.5	0.20414798	0.36330504	0.36288702	0.35313477
	0.9		0.40840003	0.38757241	0.38486254	0.3481658
	0.95		0.79909187	0.45359083	0.43574222	0.336994
	0.99		3.8474832	1.59030405	0.74786225	0.33639902

Table 4. The MSE of proposed, MLE, and other estimators ($n = 100$; varying p and ρ).

p	ρ	z	MLE	RBR	LRE	$MTPRE$
4	0.8	0.5	0.15030191	0.35109827	0.35096469	0.34399569
	0.9		0.30607083	0.39835357	0.39730275	0.36552374
	0.95		0.59080558	0.43370148	0.42738169	0.34732612
	0.99		2.64919779	1.0835372	0.72239963	0.33785071
	0.8	1.5	0.08304166	0.35463953	0.35454403	0.34969135
	0.9		0.18368698	0.36575454	0.3652278	0.34934324
	0.95		0.34019675	0.39248335	0.38869088	0.34100898
	0.99		1.65934166	0.69480589	0.52189793	0.34053362
6	0.8	0.5	0.18991808	0.35881425	0.35860212	0.34922323
	0.9		0.39316502	0.40061095	0.39922576	0.36312785
	0.95		0.74982912	0.46410825	0.45424003	0.34786402
	0.99		3.70107802	1.63102225	0.95825838	0.34156198
	0.8	1.5	0.11376439	0.3580375	0.35789435	0.35146892
	0.9		0.21457824	0.36747128	0.36665259	0.34508626
	0.95		0.43843102	0.40794434	0.4026904	0.34636058
	0.99		2.06567986	0.84193828	0.57563338	0.33134479
10	0.8	0.5	0.2852229	0.36214458	0.36184414	0.35119063
	0.9		0.55900374	0.39738078	0.39543622	0.35452892
	0.95		1.07864837	0.49410747	0.4779846	0.34328554
	0.99		5.41336219	2.10637724	1.07889904	0.33944774
	0.8	1.5	0.14558119	0.34798569	0.34780402	0.3412932
	0.9		0.2859369	0.37825692	0.37696302	0.34923878
	0.95		0.5826818	0.41894999	0.41036542	0.33827206
	0.99		2.87696505	1.1413311	0.68494021	0.34441979

3.2. Discussion of simulation results

The estimated MSEs of our proposed estimators MTPRE and MLE and other biased estimators RBR and LRE are shown in Tables 1–4. The conclusion based on Monte Carlo simulation results are as follows:

i. The proposed MTPRE consistently demonstrates superior performance compared to the other methods (MLE, RBR, and LRE) across different scenarios. The proposed estimator has minimum MSE across all simulation scenarios.
ii. To assess the impact of multicollinearity, we varied ρ from 0.8 to 0.99. It was observed that the MSE values for the MLE, RBR, and LRE increased significantly with higher ρ. It was also observed that the MTPRE estimator remained stable, maintaining the lowest MSE across all levels of multicollinearity.
iii. To assess the effect of sample size, we evaluated performance across $n = 20, 50, 80$, and 100. It was found that increasing sample size led to improved performance for all estimators. However, it was particularly evident that MTPRE achieved the lowest the MSE across all sample sizes as compared to the MLE and other methods.
iv. To assess the effect of the number of predictors p, we compared results for $p = 4, 6,$ and 10. As the number of predictors increased, the newly proposed estimator continued to perform better than the MLE and other methods. So, we consistently observed that the MTPRE outperformed all other estimators, even in high-dimensional settings.
v. Finally, the simulation results clearly show that the new MTPRE method performed better than the MLE, RBR, and LRE under varying levels of collinearity, predictor dimensions, and sample sizes.

4. Gasoline yield dataset

The gasoline yield dataset, which contains a total of 32 observations, was used by [24] and shows the proportion of crude oil remaining after the distillation process. The dependent variable is proportion, and the predictors include API gravity, vapor pressure, the temperature at which 10% of the crude vaporizes (temp10), and the temperature at which all gasoline components vaporize. The BRM offers a more appropriate fit, particularly due to the bounded nature of the response variable.

To evaluate potential multicollinearity among the variables, we examined the correlation matrix and visualized it using a heatmap. Additionally, we calculated the condition number (CN) for multicollinearity measure in the gasoline dataset.

Figure 1 shows pairwise correlations, indicating strong multicollinearity; the two independent variables, pressure and temp10, are highly negative correlated. The presence of multicollinearity is further supported by the high CN of approximately 11,280.5, which exceeds the commonly accepted threshold, indicating severe multicollinearity dataset.

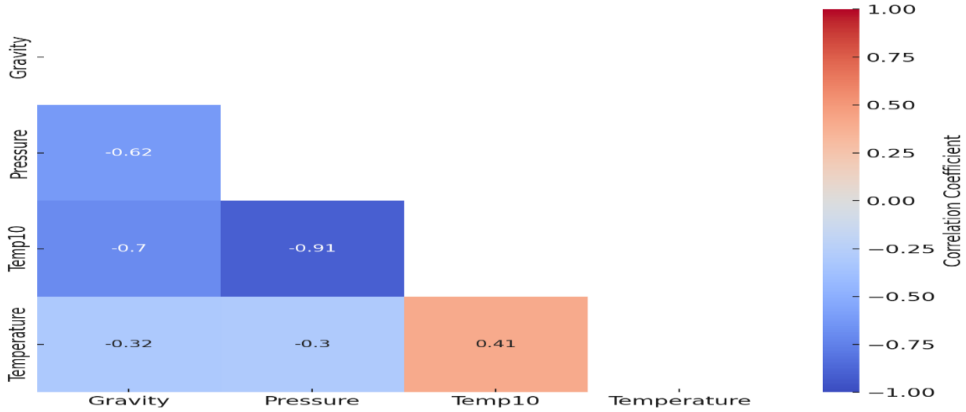

Figure 1. Pairwise correlations display of dataset.

We present the Beta regression estimates and their corresponding MSE for proposed and existing methods in Table 5.

Table 5. Estimated MSE and coefficients.

Coefficients	MLE	RBR	LRE	MTPRE
MSE	123.9608	67.8025	23.2334	12.8740
$\hat{\mu}_0$	-2.8646	-1.8696	-0.0941	-0.0971
$\hat{\mu}_1$	2.047	-0.0149	-0.0313	-0.0119
$\hat{\mu}_2$	1.0319	1.0044	-0.0122	-0.0423
$\hat{\mu}_3$	-0.0112	-0.0147	-0.0192	-0.0193
$\hat{\mu}_4$	0.3114	0.2114	0.1209	0.0110

Table 5 shows the estimated MSE and coefficients for each estimator as well as the findings from the gasoline dataset with the simulation results.

Figure 2 presents a comparison of MSE values, showing that the proposed MTPRE yields the best performance.

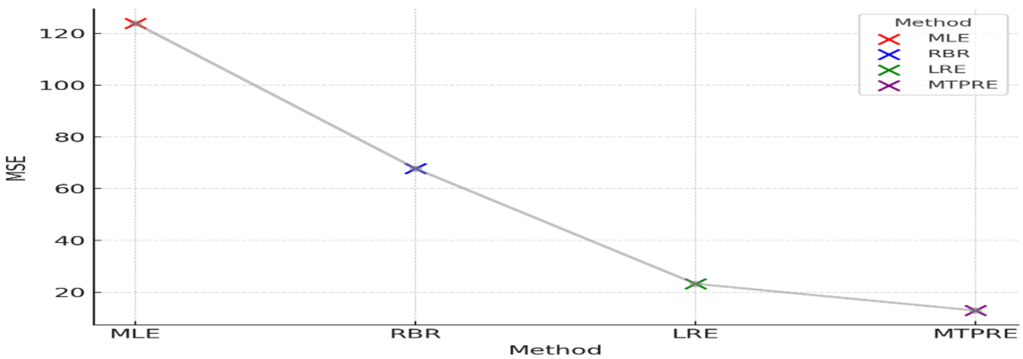

Figure 2. MSE comparison of the MLE, RBR, LRE, and MTPRE.

5. Conclusions

This study presents the MTPRE for the BRM as the alternative to the MLE to better handle the severe multicollinearity issues. The performance of the proposed estimator compares to the MLE and other biased estimators, including the RBR and LRE, through the Monte Carlo simulation results based on mean squared criterion. The findings show that multiple factors, sample size, number of predictors, and different level of correlations among the predictor variables the affect the effectiveness of the MTPRE and other estimators. The simulation results clearly show that the MTPRE consistently outperformed the MLE, RBR, and LRE, with lower MSE, particularly in cases of significant multicollinearity. Furthermore, practical applications demonstrated that the MTPRE has superior performance over the MLE and other biased estimators, highlighting its effectiveness and robustness in practical, real-world scenarios.

Use of Generative-AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in the creation of this article.

Conflict of interest

The author declares no conflict of interest.

References

1. L. Fahrmeir, G. Tutz, *Multivariate statistical modelling based on generalized linear models*, 2 Eds., New York: Springer, 2001.
2. S. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, *J. Appl. Stat.*, **31** (2004), 799–815. https://doi.org/10.1080/0266476042000214501
3. Z. W. Xiao, F. X. Zhu, L. F. Wang, R. K. Liu, F. Yu, An evaluation method of construction reliability of cable system of cable-stayed bridge based on Bayesian network, *Int. J. Struct. Integrity*, **15** (2024), 1027–1050. https://doi.org/10.1108/IJSI-06-2024-0079
4. H. Ma, S. P. Zhu, Y. Q. Guo, L. D. Pan, S. Y. Yang, D. B. Meng, Residual useful life prediction of the vehicle isolator based on Bayesian inference, *Structures*, **58** (2023), 105518. https://doi.org/10.1016/j.istruc.2023.105518
5. D. B. Meng, S. P. Zhu, *Multidisciplinary design optimization of complex structures under uncertainty*, Boca Raton: CRC Press, 2024. https://doi.org/10.1201/9781003464792
6. A. Zeileis, C. Kleiber, S. Jackman, Regression models for count data in R, *J. Stat. Softw.*, **27** (2008), 1–25. https://doi.org/10.18637/jss.v027.i08
7. N. R. Draper, H. Smith, *Applied regression analysis*, John Wiley & Sons, 1998.
8. M. Qasim, K. Månsson, B. M. G. Kibria, On some Beta ridge regression estimators: method, simulation and application, *J. Stat. Comput. Simul.*, **91** (2021), 1699–1712. https://doi.org/10.1080/00949655.2020.1867549
9. S. Seifollahi, H. Bevrani, James-Stein type estimators in Beta regression model: simulation and application, *Hacet. J. Math. Stat.*, **52** (2023), 1046–1065.

10. A. Erkoç, E. Ertan, Z. Y. Algamal, K. U. Akay, The Beta Liu-type estimator: simulation and application, *Hacet. J. Math. Stat.*, **52** (2023), 828–840.
11. A. E. Hoerl, R. W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12** (1970), 55–67. https://doi.org/10.1080/00401706.1970.10488634
12. M. Amin, H. Ashraf, H. S. Bakouch, N. Qarmalah, James Stein estimator for the Beta regression model with application to heat-treating test and body fat datasets, *Axioms*, **12** (2023), 526. https://doi.org/10.3390/axioms12060526
13. Y. Asar, A. Genç, Two-parameter ridge estimator in the binary logistic regression, *Commun. Stat. Simul. Comput.*, **46** (2017), 7088–7099. https://doi.org/10.1080/03610918.2016.1224348
14. D. Y. Dai, D. Wang, A generalized Liu-type estimator for logistic partial linear regression model with multicollinearity, *AIMS Math.*, **8** (2023), 11851–11874. https://doi.org/10.3934/math.2023600
15. W. B. Altukhaes, M. Roozbeh, N. A. Mohamed, Feasible robust Liu estimator to combat outliers and multicollinearity effects in restricted semiparametric regression model, *AIMS Math.*, **9** (2024), 31581–31606. https://doi.org/10.3934/math.20241519
16. A. T. Hammad, I. Elbatal, E. M. Almetwally, M. M. A. El-Raouf, M. A. El-Qurashi, A. M. Gemeay, A novel robust estimator for addressing multicollinearity and outliers in Beta regression: simulation and application, *AIMS Math.*, **10** (2025), 21549–21580. https://doi.org/10.3934/math.2025958
17. M. R. Abonazel, I. M. Taha, Beta ridge regression estimators: simulation and application, *Commun. Stat. Simul. Comput.*, **52** (2023), 4280–4292. https://doi.org/10.1080/03610918.2021.1960373
18. K. J. Liu, A new class of blased estimate in linear regression, *Commun. Stat. Theory Methods*, **22** (1993), 393–402. https://doi.org/10.1080/03610929308831027
19. K. Månsson, B. M. G. Kibria, G. Shukur, On Liu estimators for the logit regression model, E*con. Model.*, **29** (2012), 1483–1488. https://doi.org/10.1016/j.econmod.2011.11.015
20. Z. Y. Algamal, M. R. Abonazel, Developing a Liu-type estimator in Beta regression model, *Concurr. Comput. Pract. Exp.*, **34** (2022), e6685. https://doi.org/10.1002/cpe.6685
21. A. F. Lukman, B. Aladeitan, K. Ayinde, M. R. Abonazel, Modified ridge-type for the Poisson regression model: simulation and application, *J. Appl. Stat.*, **49** (2022), 2124–2136. https://doi.org/10.1080/02664763.2021.1889998
22. N. Akhtar, M. F. Alharthi, Enhancing accuracy in modelling highly multicollinear data using alternative shrinkage parameters for ridge regression methods, *Sci. Rep.*, **15** (2025), 10774. https://doi.org/10.1038/s41598-025-94857-7
23. S. Chand, B. M. G. Kibria, A new ridge type estimator and its performance for the linear regression model: simulation and application, *Hacet. J. Math. Stat.*, **53** (2024), 837–850. https://doi.org/10.15672/hujms.1359446
24. N. H. Prater, Estimate gasoline yields from crudes, *Petroleum Refiner*, **35** (1956), 236–238.