Necmettin Erbakan Üniversitesi
Fen ve Mühendislik Bilimleri Dergisi
Journal of Science and Engineering (NEJSE)
DOI: 10.47112/neufmbd.2023.11

Cilt: 5 Sayı:2 Yıl: 2023 | Araştırma Makelesi / Research Article | E-ISSN: 2667- 7989

Beta Regresyon Modelinde Yeni Bir Yanı Düzeltilmiş Tahmin Edicinin Performansı: Bir Monte Carlo Çalışması

Yasin ASAR [1]

[1] Necmettin Erbakan University, Faculty of Science, Department of Mathematics and Computer Sciences, Konya, Türkiye, yasar@erbakan.edu.tr *(Corresponding Author)*

Makale Bilgileri

Makale Geçmişi
Geliş: 07.06.2023
Kabul: 31.08.2023
Yayın: 31.12.2023

Anahtar Kelimeler:
Beta regresyon, çoklu bağlantı, yanı düzeltilmiş tahmin edici, maksimum olabilirlik

ÖZ

Beta regresyon modelinde regresyon parametrelerini elde etmek için birincil yaklaşım, maksimum olabilirlik tahmin tekniğinin kullanılmasıdır. Bununla birlikte, beta regresyon modelinde çoklu bağlantının maksimum olabilirlik tahmin edicisinin varyansı üzerinde negatif bir etkiye sahip olduğu, yani maksimum olabilirlik tahmin edicisinin varyansının şişirildiği kabul edilmektedir. Bu konuyu ele almak için, çoklu bağlantı sorununu çözmek için yeni bir yanı düzeltilmiş tahmin edici tanıtılmıştır. Bu yeni tahmin edicinin etkinliği, bir Monte Carlo simülasyon deneyi kullanılarak sayısal bir araştırma yoluyla değerlendirilmiştir. Sonuçlar, önerilen tahmin edicinin diğer rakip tahmin edicilere kıyasla hem hata kareler ortalaması hem de karesel yan değerleri bakımından önemli iyileştirmeler sağladığını göstermektedir.

Performance of A New Bias Corrected Estimator in Beta Regression Model: A Monte Carlo Study

Article Info

Article History
Received: 07.06.2023
Accepted: 31.08.2023
Published: 31.12.2023

Keywords:
Multicollinearity, bias-adjusted estimator, maximum likelihood

ABSTRACT

The primary approach for obtaining regression parameters in the beta regression model is the utilization of the maximum likelihood estimation technique. However, it is acknowledged that multicollinearity has a detrimental effect on the variance of the maximum likelihood estimator in the beta regression model, namely, the variance of the maximum likelihood estimator is inflated. To address this issue, a novel bias-adjusted estimator is introduced to tackle the problem of multicollinearity. The effectiveness of these new estimator is assessed through a numerical investigation using a Monte Carlo simulation experiment. The results indicate that the proposed estimators yield substantial improvements compared to other competing estimators in terms of both the mean squared error and squared bias values.

Atıf/Citation: Asar, Y. (2023). Performance of a new bias corrected estimator in beta regression model: a Monte Carlo study, *Necmettin Erbakan University Journal of Science and Engineering*, *5*(2), 75-87. https://doi.org/10.47112/neufmbd.2023.11

INTRODUCTION

Regression analysis is used to model data such that the explanatory variables are used to explain the dependent or response variable. The type of regression model depends on the distribution of the response variable. The beta regression model was firstly proposed in [1] to model the proportions data in other words the response variable is restricted to the interval $(0,1)$ and its distribution is beta. As it is the case in generalized linear models, the mean of the response variable is related to several explanatory variables linearly via a link function.

The parameters of the beta regression model are generally obtained by the maximum likelihood estimation technique. Although, it is well known that the explanatory variables are assumed not to be linearly correlated, this is not the case in many real-life situations. This problem is known as the multicollinearity in the literature [2]. Multicollinear settings affect the performance of the maximum likelihood estimator (MLE) negatively. For example, the variance of MLE becomes inflated and therefore the inference may not be reliable.

To overcome this problem, many methods have been proposed in the linear regression model and most of these methods have been generalized to beta regression model. The well-known ridge estimator [3] was proposed for the beta regression model in [4] using a penalized likelihood approach and obtained in [2]. Another method is called Liu estimator [5] which is extended to the beta regression by [6]. Moreover, the Liu-type estimator [7] has been generalized to beta regression model by [8]. These methods are called biased estimators.

On the other hand, Ospina et al. [9] introduced the second order biases of maximum likelihood estimator and proposed bias-adjusted estimators based on them in beta regression. Simas et al. [10] proposed three bias corrected estimators such that one is based on an analytical method and the other two of them are based on bootstrap methods where they let the regression structure to be nonlinear and they also allow a regression model for the dispersion parameter. Ospina and Ferrari [11] proposed asymptotically unbiased estimators based on the second order bias of maximum likelihood estimators in the zero-or-one inflated beta regression models.

Moreover, Kadiyala [12] and Ohtani [13] proposed almost unbiased versions of ridge estimator using different methods. Although almost unbiased estimators are also biased estimators, they are expected to have less bias than their biased versions. Almost unbiased estimators are well studied in generalized linear models. We refer to the following papers: [14–17].

Although the Liu-type estimator [7] is a biased estimator, the bias or squared bias of this estimator has not been investigated in the literature. Therefore, the purpose of this study is to define a bias corrected or almost unbiased version of the Liu-type estimator in beta regression models and study its mean squared error (MSE) and squared bias (SB) properties and compare the performance of the new estimator to the ridge estimator and the Liu-type estimator by designing an extensive Monte Carlo simulation study. The advantage of using the new estimator is that it does not only overcome the collinearity problem but also reduces the bias of the Liu-type estimator in beta regression.

Thus, the organization of this paper is as follows: beta regression fundamentals are given briefly in the next section. The new proposed method which is called bias corrected beta Liu-type estimator or almost unbiased beta Liu-type estimator is introduced and its theoretical properties are derived afterwards. In Monte Carlo simulation section, a Monte Carlo simulation experiment is provided to compare the performances of the existing estimators to the new bias corrected beta Liu-type estimator. Finally, some conclusive remarks are stated in conclusion section.

BETA REGRESSION MODEL

To introduce the beta regression model, we assume that the dependent variable is beta distributed. Thus, the observations of the dependent variable $y_1, y_2, \ldots, y_n$ are assumed to follow a beta distribution which has the following beta density function

$$f(y;\mu,\phi)=\frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma\big((1-\mu)\phi\big)}y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1} \tag{1}$$

where $0<y<1$, $E(y)=\mu$ and $\mathrm{var}(y)=\frac{V(\mu)}{1+\phi}$ where $V(\mu)=\mu(1-\mu)$ and $\Gamma(.)$ is the gamma function. Here μ is the mean of the response and ϕ is called the precision parameter which directly affects the variance of y. Using this representation of beta distribution Ferrari and Cribari-Neto [1] proposed the following beta regression model by relating the mean of y_i with the linear predictors as

$$g(\mu_i)=\sum_{i=1}^{n}x_{ij}\beta_j=\boldsymbol{x}_i^T\boldsymbol{\beta}=\eta_i \tag{2}$$

where $\boldsymbol{\beta}=\left(\beta_1,\beta_2,\ldots,\beta_p\right)^T$ is the vector of regression coefficients and $X=[\boldsymbol{x}_1^T,\ldots,\boldsymbol{x}_n^T]$ is an $n\times p$ data matrix with p explanatory variables. The function $g(.)$ is called the link function similar to the generalized linear models. There are several possible candidates for the link function, however, we use the logit link function given as $g(\mu)=\log(\mu/(1-\mu))$. Using the logit link function, one can define

$$\mu_i=\frac{e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i^T\boldsymbol{\beta}}}. \tag{3}$$

Thus, the corresponding log-likelihood function of the beta regression model can be written as follows

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta})=\sum_{i=1}^{n}&\log\big(\Gamma(\phi)\big)-\log\big(\Gamma(\phi\mu_i)\big)-\log\Big(\Gamma\big(\phi(1-\mu_i)\big)\Big)+\phi(1-\mu_i)\log(y_i)\\&+\{\phi(1-\mu_i)-1\}\log(1-y_i).\end{aligned} \tag{4}$$

Since the log-likelihood function given above is nonlinear in $\boldsymbol{\beta}$, one should use an iterative algorithm to obtain the maximum likelihood estimator (MLE). One possible option is to use Fisher's scoring method. In order to do so, we need to obtain the score function and the Fisher's information matrix. The theoretical derivations of these functions are given in [1] in details. Therefore, we just provide the maximum likelihood estimator in matrix notation as follows

$$\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}=\left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\hat{\boldsymbol{z}} \tag{5}$$

where $\widehat{\mathbf{W}}$ is the weight matrix whose diagonal elements are given by

$$w_i=\phi\{\psi'(\phi\mu_i)+\psi'\big(\phi(1-\mu_i)\big)\}\frac{1}{\big(g'(\mu_i)\big)^2}$$

and $\hat{\boldsymbol{z}}=\hat{\boldsymbol{\eta}}+\widehat{\mathbf{W}}^{-\mathbf{1}}\widehat{\mathbf{T}}(\boldsymbol{y}^*-\hat{\boldsymbol{\mu}})$ is called the working response such that $y_i^*=\log(y_i/(1-y_i))$ and $\widehat{\mathbf{T}}$ is a diagonal matrix such that $\widehat{\mathbf{T}}=\mathrm{diag}\big(1/g'(\mu_1),\ldots,1/g'(\mu_n)\big)$ and $\hat{\eta}_i=\boldsymbol{x}_i^T\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}$ which is called the linear predictor.

The theoretical properties of the estimators can be investigated using variance-covariance matrices and MSE functions. The asymptotic variance-covariance matrix of MLE is given by [18]

$$\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}\big)=\frac{1}{\phi}\left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}\right)^{-1}. \tag{6}$$

Thus, the MSE function of MLE is obtained by taking the trace of $\mathrm{Cov}\big(\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}\big)$ as

$$\mathrm{MSE}(\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}}) = \frac{1}{\phi}\sum_{j=1}^{p}\frac{1}{\lambda_j} \tag{7}$$

where $\lambda_1, \lambda_2, \dots, \lambda_p$ are the eigenvalues of $\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}$. It is obvious from Equation (7) that if the matrix $\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}$ is ill conditioned then the variances of the estimators become inflated since the inverses of the eigenvalues of the matrix $\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}$ becomes close to zero and this situation results in a very large MSE values.

Due to the high variance of MLE, many biased estimators are defined to be used in the presence of ill-conditioned matrix of cross-products $\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}$. One of them is the beta-ridge estimator (RE) proposed and studied by [2,4,19]. RE is defined as

$$\widehat{\boldsymbol{\beta}}_{\mathrm{RE}} = \left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\widehat{\mathbf{z}} \tag{8}$$

where $k > 0$ is the biasing ridge parameter. The MSE function of RE is a complicated function of the parameter k. One other estimator is the beta-Liu estimator (LE) introduced by [6] as follows

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LE}} = \left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} + \mathbf{I}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} + d\mathbf{I}\right)\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} \tag{9}$$

where $0 < d < 1$ is the biasing Liu parameter. The advantage of using LE is that its MSE function is just a second-degree function of the parameter d.

Although RE and LE have both their unique advantages and disadvantages, Algamal and Abonazel [8] proposed the following beta-Liu-type estimator (LT) combining the advantages of both RE and LE

$$\widehat{\boldsymbol{\beta}}_{\mathrm{LT}} = \left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} - d\mathbf{I}\right)\widehat{\boldsymbol{\beta}}_{\mathrm{MLE}} \tag{10}$$

where $k > 0$ and $-\infty < d < \infty$. The bias vector, variance-covariance matrix and the MSE function or LT are given respectively by

$$\mathrm{bias}(\widehat{\boldsymbol{\beta}}_{\mathrm{LT}}) = -(k+d)\left(\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\boldsymbol{\beta}, \tag{11}$$

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{\mathrm{LT}}) = \frac{1}{\phi}\mathbf{C}_k^{-1}\mathbf{C}_d\mathbf{C}^{-1}\mathbf{C}_d\mathbf{C}_k^{-1}, \tag{12}$$

$$\mathrm{MSE}(\widehat{\boldsymbol{\beta}}_{\mathrm{LT}}) = \frac{1}{\phi}\sum_{j=1}^{p}\left\{\frac{(\lambda_j - d)^2}{\lambda_j(\lambda_j + k)^2} + \frac{(d+k)^2}{(\lambda_j + k)^2}\alpha_j^2\right\} \tag{13}$$

where $\mathbf{C}_k = \mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}$, $\mathbf{C}_d = \mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X} - d\mathbf{I}$, $\mathbf{C} = \mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}$, α_j is the jth element of $\boldsymbol{\alpha} = \mathbf{Q}\boldsymbol{\beta}$ where the matrix $\mathbf{Q}$ is composed of the normalized eigenvectors of $\mathbf{C}$ as its columns. Moreover, we have $\mathbf{Q}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\widehat{\mathbf{W}}\mathbf{X}\mathbf{Q} = \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$.

It is clear from Equations (12-13) that LT is a biased estimator, and the aim of this study is to reduce this bias by applying the bias correction procedure introduced by Kadiyala [12] and Ohtani [13].

A NEW BIAS ADJUSTED LT ESTIMATOR

The estimator LT was proposed to have the advantages of both RE and LE. However, the squared bias properties of LT have not been studied so far. In this section, we propose an almost unbiased or bias corrected version of LT (AULT) so that it has even less squared bias than that of LT under certain scenarios. The following definition will be used to define AULT.

Definition 1: [20] Suppose $\widetilde{\boldsymbol{\beta}}$ is a biased estimator of parameter vector $\boldsymbol{\beta}$, and if the bias vector of $\widetilde{\boldsymbol{\beta}}$ is given by $\text{bias}(\widetilde{\boldsymbol{\beta}}) = \text{E}(\widetilde{\boldsymbol{\beta}}) - \boldsymbol{\beta} = \mathbf{R}\boldsymbol{\beta}$, which shows that $\text{E}(\widetilde{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}) = \boldsymbol{\beta}$, then the estimator $\widetilde{\widetilde{\boldsymbol{\beta}}} = \widetilde{\boldsymbol{\beta}} - \mathbf{R}\widetilde{\boldsymbol{\beta}}$ is called the almost unbiased estimator based on the biased estimator $\widetilde{\boldsymbol{\beta}}$.

Now, using Definition 1, the almost unbiased beta LT estimator (AULT) can be defined as

$$\begin{aligned}\widehat{\boldsymbol{\beta}}_{\text{AULT}} &= \widehat{\boldsymbol{\beta}}_{\text{LT}} + (k+d)\left(\mathbf{X}^{\text{T}}\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\widehat{\boldsymbol{\beta}}_{\text{LT}} \\ &= \left(\mathbf{I} - (k+d)^2\left(\mathbf{X}^{\text{T}}\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-2}\right)\widehat{\boldsymbol{\beta}}_{\text{MLE}}\end{aligned} \tag{14}$$

where $k > 0$ and $-\infty < d < \infty$. To study further the properties of AULT, we provide the bias vector, variance-covariance matrix and MSE function of AULT respectively as follows

$$\text{bias}\left(\widehat{\boldsymbol{\beta}}_{\text{AULT}}\right) = -(k+d)\left(\mathbf{X}^{\text{T}}\widehat{\mathbf{W}}\mathbf{X} + k\mathbf{I}\right)^{-1}\boldsymbol{\beta}, \tag{15}$$

$$\text{Cov}\left(\widehat{\boldsymbol{\beta}}_{\text{AULT}}\right) = \frac{1}{\phi}\boldsymbol{\Psi}\mathbf{C}^{-1}\boldsymbol{\Psi}^{\text{T}}, \tag{16}$$

$$\text{MSE}\left(\widehat{\boldsymbol{\beta}}_{\text{AULT}}\right) = \frac{1}{\phi}\sum_{j=1}^{p}\left\{\frac{\left(\lambda_j - d\right)^2\left(\lambda_j + d + 2k\right)^2}{\lambda_j\left(\lambda_j + k\right)^4} + \phi\frac{(d+k)^4}{\left(\lambda_j + k\right)^4}\alpha_j^2\right\} \tag{17}$$

where $\boldsymbol{\Psi} = \mathbf{I} - (k+d)^2\mathbf{C}_k^{-2}$.

Estimation of the parameter d in AULT

Our aim in this subsection is to find an estimator of the parameter d for a given estimator of k. Thus, we start by minimizing the MSE of AULT with respect to d. The derivative of $\text{MSE}\left(\widehat{\boldsymbol{\beta}}_{\text{AULT}}\right)$ is computed as

$$\frac{\partial \text{MSE}\left(\widehat{\boldsymbol{\beta}}_{\text{AULT}}\right)}{\partial d} = \sum_{j=1}^{p}\frac{(d+k)}{\lambda_j\left(\lambda_j + k\right)^4}\left(-\varphi\left(\lambda_j - d\right)\left(\lambda_j + d + 2k\right) + (d+k)^2\lambda_j\alpha_j^2\right) \tag{18}$$

where $\varphi = 1/\phi$. Now, we want to equate Equation (18) to zero and solve for the parameter d. If we let $f(d) = -\varphi\left(\lambda_j - d\right)\left(\lambda_j + d + 2k\right) + (d+k)^2\lambda_j\alpha_j^2$ and find the roots of $f(d)$ which is a squared function d. Thus, we can rewrite $f(d)$ as

$$f(d) = d^2\left(\varphi + \lambda_j\alpha_j^2\right) + d\left(2k\varphi + 2k\lambda_j\alpha_j^2\right) + \left(-\varphi\lambda_j^2 - 2k\varphi\lambda_j + k^2\lambda_j\alpha_j^2\right). \tag{19}$$

The discriminant of $f(d)$ given in Equation (19) is obtained as $\Delta_j = 4\varphi\left(\varphi + \lambda_j\alpha_j^2\right)\left(\lambda_j + k\right)^2 > 0$. Thus Δ_j has two roots for each $j = 1,2,\ldots,p$. One can obtain the roots as follows

$$d_{1j,2j} = -k \mp \frac{\left(\lambda_j + k\right)\sqrt{\varphi\left(\varphi + \lambda_j\alpha_j^2\right)}}{\varphi + \lambda_j\alpha_j^2} \tag{20}$$

Although it is possible to propose many estimators of the parameter d using $d_{1j,2j}$'s, we only propose the following estimator and use it in the Monte Carlo simulation study

$$d_{opt} = \text{median}\left(-k + \frac{(\lambda_j + k)\sqrt{\varphi(\varphi + \lambda_j \alpha_j^2)}}{\varphi + \lambda_j \alpha_j^2}\right). \quad (21)$$

MONTE CARLO SIMULATION STUDY

In this section, we present the details of the Monte Carlo simulation experiment, which is designed to compare the performances of MLE, RE, LT, and AULT. The dependent variable $\boldsymbol{y}$ is generated using the probability density function given in Equation (1) which is the beta density function such that $y_i \sim \text{Beta}(\phi\mu_i, \phi(1-\mu_i))$ where $\mu_i = e^{x_i^T\boldsymbol{\beta}} / \left(1 + e^{x_i^T\boldsymbol{\beta}}\right)$ such that the regression parameters are chosen to be the corresponding eigenvector of the maximum eigen value of the matrix $\mathbf{X}^\mathbf{T}\mathbf{X}$, so that $\sum_{j=1}^{p} \beta_j^2 = 1$ which is a very common restriction, see [8,17,18] and [21]. Moreover, following [8] the data matrix **X** is generated via

$$x_{ij} = (1-\rho^2)^{1/2} m_{ij} + \rho m_{i(p+1)} \quad (22)$$

where m_{ij}'s are independent random number standard normal distribution such that $i = 1, \dots, n$ and $j = 1, \dots, p$. In Equation (19), ρ controls the correlation between the predictors. The sample size n is taken to be 50, 100 and 200. ρ changes as $0.90, 0.95$ and 0.99. The number of predictors p are selected as $4, 8, 12$. Also, the precision parameter ϕ is chosen as 5 and 10.

To compare the simulation performances of the estimators, the following equations are respectively used to obtain the simulated MSE and square bias (SB) values for each estimator

$$\text{SMSE}(\widetilde{\boldsymbol{\beta}}) = \frac{1}{1000} \sum_{l=1}^{1000} (\widetilde{\boldsymbol{\beta}}_l - \boldsymbol{\beta})^{\mathrm{T}} (\widetilde{\boldsymbol{\beta}}_l - \boldsymbol{\beta}) \quad (23)$$

$$\text{SB}(\widetilde{\boldsymbol{\beta}}) = \left(\overline{\widetilde{\boldsymbol{\beta}}} - \boldsymbol{\beta}\right)^{\mathrm{T}} \left(\overline{\widetilde{\boldsymbol{\beta}}} - \boldsymbol{\beta}\right), \overline{\widetilde{\boldsymbol{\beta}}} = \frac{1}{1000} \sum_{l=1}^{1000} \widetilde{\boldsymbol{\beta}}_l \quad (24)$$

where $\widetilde{\boldsymbol{\beta}}_l$ represents each estimator at the rth repetition of the simulation in the study. The number of replications in the simulation is taken as 1000 and only the dependent variable is generated for each repetition while the data matrix **X** is fixed.

We estimate the biasing parameters of each method using the best estimators of that method proposed by the related paper as follows:

- For RE, the parameter k is estimated by $\hat{k}_{RE} = \frac{\lambda_{\min}}{\hat{\phi}\hat{\alpha}_{\min}^2}$ [19]
- For LT, $\hat{k}_{\mathrm{LT}} = \frac{1}{\hat{\phi}\sum_{j=1}^{p}\hat{\alpha}_j^2}$ and $\hat{d}_{\mathrm{LT}} = \frac{\sum_{j=1}^{p}\left(\frac{1}{\hat{\phi}} - k\hat{\alpha}_j^2\right)/(\lambda_j + k)^2}{\sum_{j=1}^{p}\left(\frac{1}{\hat{\phi}} + \lambda_j\hat{\alpha}_j^2\right)/\lambda_j(\lambda_j + k)^2}$ [8]
- For AULT, $\hat{k}_{\mathrm{AULT}} = \frac{1}{\hat{\phi}\sum_{j=1}^{p}\hat{\alpha}_j^2}$ and $\hat{d}_{\mathrm{AULT}} = \text{median}\left(-\hat{k}_{\mathrm{AULT}} + \frac{(\lambda_j + \hat{k}_{\mathrm{AULT}})\sqrt{\hat{\varphi}(\hat{\varphi} + \lambda_j\hat{\alpha}_j^2)}}{\hat{\varphi} + \lambda_j\hat{\alpha}_j^2}\right)$

The results of the Monte Carlo simulation experiment are summarized in Tables 1-2 and Figures 1-4. In Tables 1-2, the squared biases of the estimators are provided. According to the tables, it is observed that the SB of AULT is the least in most of the situations considered. However, when the sample size is 50 and the number of variables is 12, SB of AULT becomes greater than SB of LT. An increase in the sample size affects the SBs

Table 1. *Squared bias values of the estimators when $\phi = 5$ for different values of p and ρ*

	ρ		0.90			0.95			0.99	
p	n	50	100	200	50	100	200	50	100	200
	RE	0.0084	0.1257	0.1702	0.0318	0.0625	0.0769	0.0174	0.0426	0.0479
4	LT	0.0075	0.0123	0.0184	0.0051	0.0117	0.0186	0.0072	0.0192	0.0264
	AULT	0.0025	0.0058	0.0113	0.0027	0.0078	0.0125	0.0052	0.0125	0.0161
	RE	0.0404	0.0706	0.0833	0.0668	0.0760	0.0623	0.3170	0.3801	0.1453
8	LT	0.0376	0.0469	0.0445	0.0515	0.0609	0.0526	0.1409	0.1801	0.0894
	AULT	0.0310	0.0556	0.0390	0.0399	0.0803	0.0480	0.0858	0.2219	0.0738
	RE	0.1320	0.0666	0.0857	0.3023	0.0888	0.0909	5.3136	0.3680	0.3142
12	LT	0.0961	0.0585	0.0667	0.1890	0.0750	0.0865	2.2530	0.1640	0.1719
	AULT	0.1533	0.0624	0.0590	0.3691	0.0858	0.0747	32.1480	0.2885	0.1120

Table 2. *Squared bias values of the estimators when $\phi = 10$ for different values of p and ρ*

	ρ		0.90			0.95			0.99	
p	n	50	100	200	50	100	200	50	100	200
	RE	0.0038	0.1140	0.1599	0.0173	0.0422	0.0589	0.0173	0.0121	0.0264
4	LT	0.0039	0.0051	0.0068	0.0014	0.0032	0.0055	0.0015	0.0031	0.0099
	AULT	0.0005	0.0009	0.0024	0.0003	0.0012	0.0027	0.0006	0.0020	0.0041
	RE	0.0128	0.0366	0.0435	0.0208	0.0306	0.0252	0.1149	0.1589	0.0675
8	LT	0.0121	0.0143	0.0149	0.0156	0.0209	0.0176	0.0587	0.0722	0.0353
	AULT	0.0095	0.0115	0.0111	0.0160	0.0266	0.0147	0.0503	0.1013	0.0261
	RE	0.0500	0.0239	0.0365	0.1525	0.0343	0.0344	0.9578	0.1491	0.1360
12	LT	0.0339	0.0195	0.0227	0.0865	0.0264	0.0307	0.4486	0.0648	0.0779
	AULT	0.1224	0.0223	0.0179	0.4596	0.0369	0.0238	2.3292	0.1094	0.0462

of the estimators negatively. Similarly, an increase in the degree of correlation ρ results in an increase in the SB values. Moreover, an increase in the number of explanatory variables causes an increase in the SB values as well.

According to the tables, one can conclude the followings:

- It is observed from figures that AULT has a best performance in almost all of the situations.
- The MLE becomes the worst estimator, in other words, it produces the highest MSE values in most of the cases, except for some of the cases when $p = 4$, in these cases RE becomes the worst.
- Increasing the sample size has a positive effect on the estimators generally meaning that the MSEs decrease.
- The MSE values increase if the number of explanatory variables increase.
- In general, if the degree of correlation is increased, the MSEs of the estimators are affected negatively.
- There is no monotonic behavior in the distributions of the performances of the estimators, i.e., in some situations RE has more variability than the others. However, AULT has more variability especially when the sample size is low, and the number of explanatory variables is high.
- It is observed that LT is more stable generally.

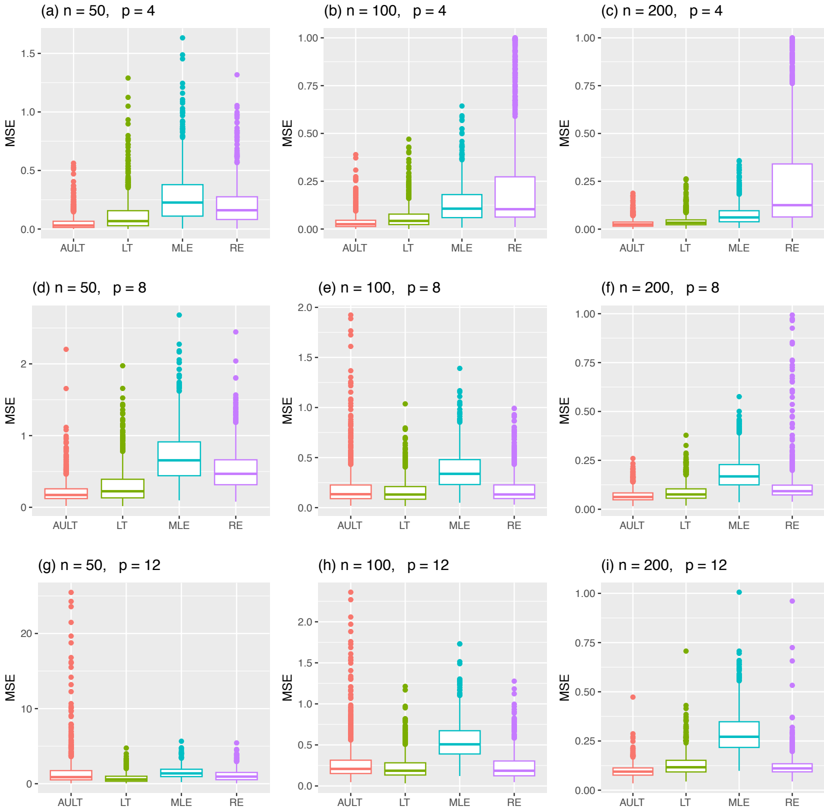

Figure 1. *Boxplots of the simulated MSE values of the estimators when* $\phi = 5$ *and* $\rho = 0.90$

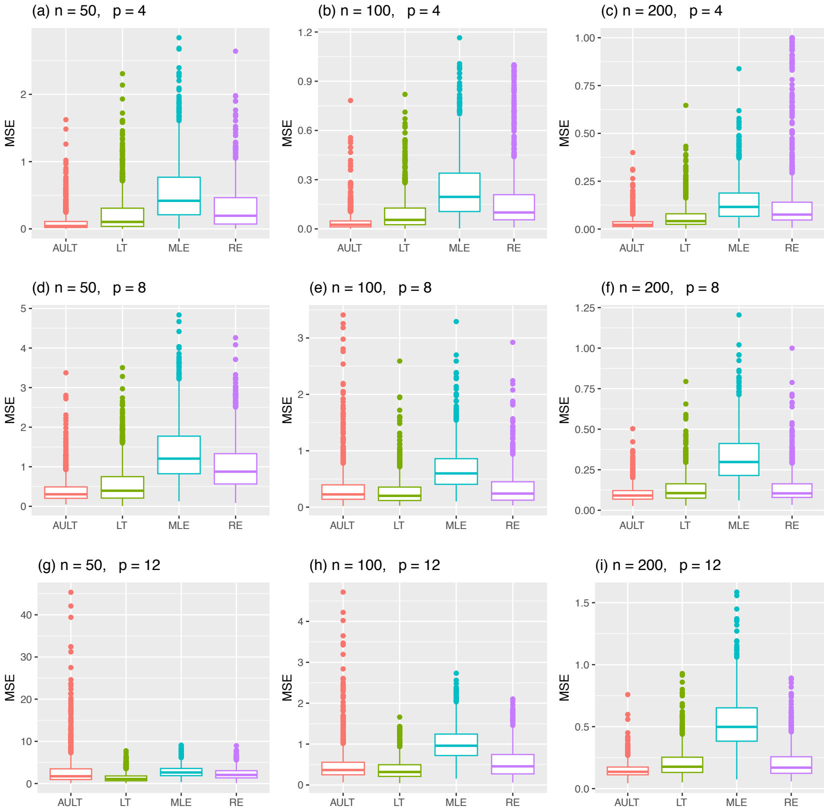

Figure 2. *Boxplots of the simulated MSE values of the estimators when* $\phi = 5$ *and* $\rho = 0.95$

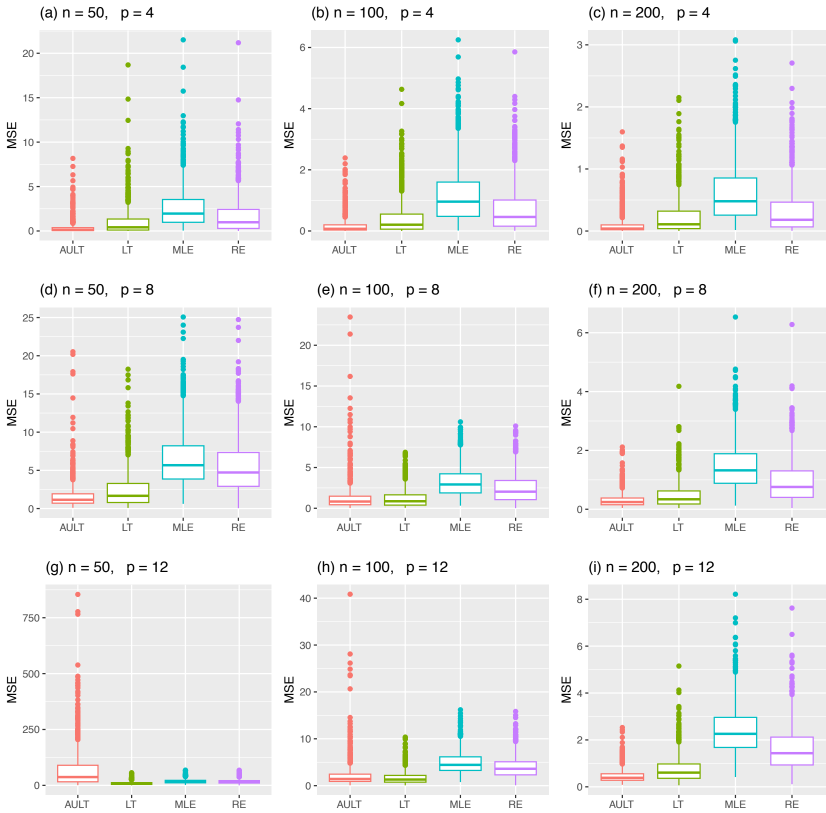

Figure 3. *Boxplots of the simulated MSE values of the estimators when* $\phi = 5$ *and* $\rho = 0.99$

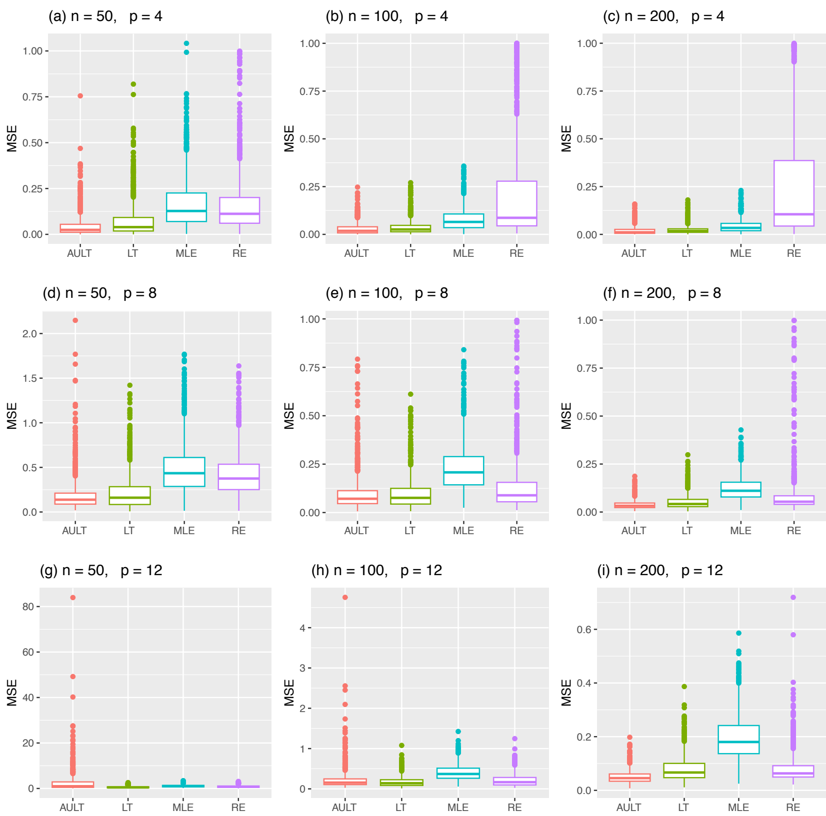

Figure 4. *Boxplots of the simulated MSE values of the estimators when* $\phi = 10$ *and* $\rho = 0.90$

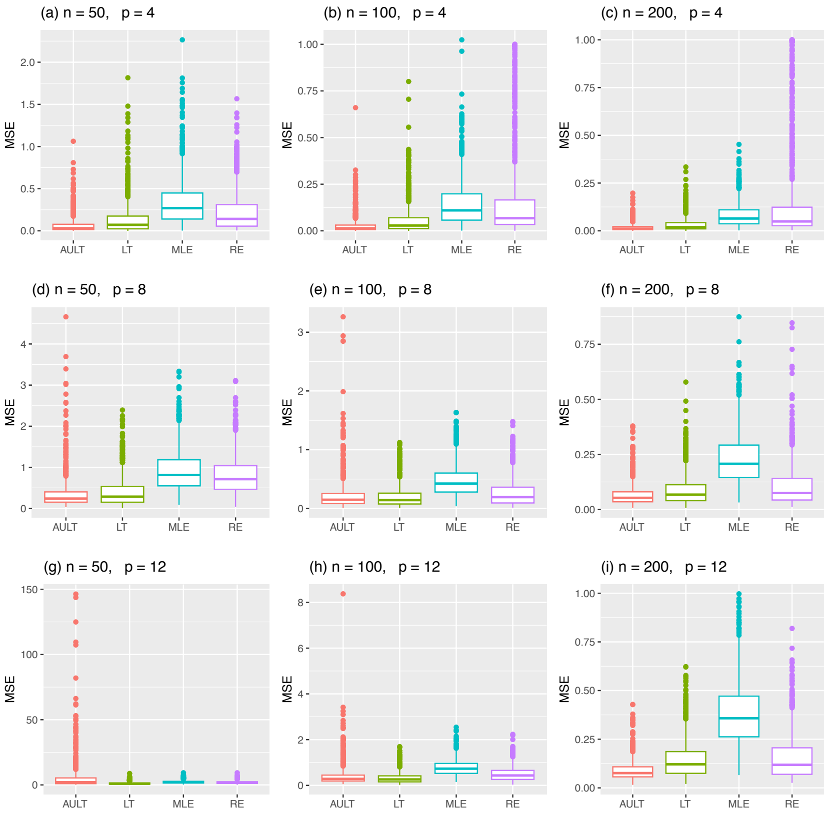

Figure 5. *Boxplots of the simulated MSE values of the estimators when* $\phi = 10$ *and* $\rho = 0.95$

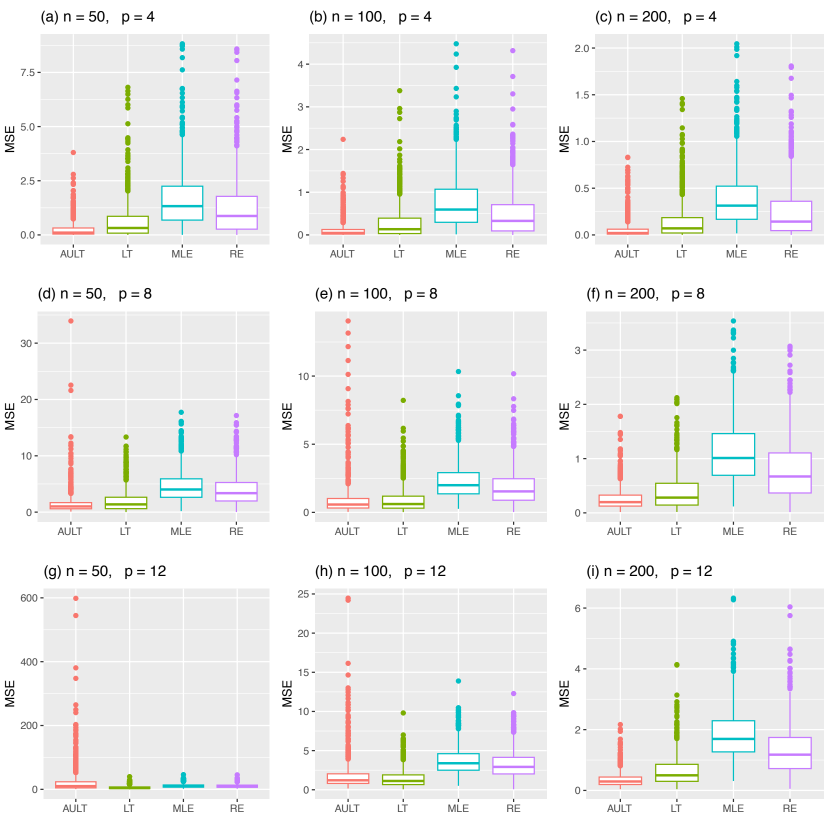

Figure 6. *Boxplots of the simulated MSE values of the estimators when* $\phi = 10$ *and* $\rho = 0.99$

CONCLUSIONS

In this paper, a new almost unbiased estimator based on the beta Liu-type estimator in the beta regression models is proposed. The MSE function, bias vector and variance-covariance matrix of the new estimator are derived. Since the new almost unbiased estimator has two parameters k and d, we also propose an estimator to estimate d to be used in the numerical studies. The squared bias performance of the beta Liu-type estimator is also studied. The new estimator AULT is compared to LT, RE and MLE by conducting an extensive Monte Carlo simulation using the mean squared error and squared bias criteria. According to the results of the simulation, AULT can be an alternative method to the existing estimators in the presence of ill-conditioned data metrices when beta regression is used by the researchers. For future work, it is possible to introduce a generalized version of AULT in terms of the generalized Liu-type estimator and study its performance numerically.

REFERENCES

[1] S.L.P. Ferrari, F. Cribari-Neto, Beta regression for modelling rates and proportions, *Journal of Applied Statistics*. 31 (2004) 799-815. doi:10.1080/0266476042000214501.

[2] M. Qasim, K. Månsson, B.M. Golam Kibria, On some beta ridge regression estimators: method, simulation and application, *Journal of Statistical Computation and Simulation*. 91 (2021) 1699-1712. doi:10.1080/00949655.2020.1867549.

[3] A.E. Hoerl, R.W. Kennard, Ridge Regression: Biased Estimation for Nonorthogonal Problems, *Technometrics*. 12 (1970) 55-67. doi:10.1080/00401706.1970.10488634.

[4] E.O. Ogundimu, G.S. Collins, Predictive performance of penalized beta regression model for continuous bounded outcomes, *Journal of Applied Statistics*. 45 (2018) 1030-1040. doi:10.1080/02664763.2017.1339024.

[5] K. Liu, A new class of blased estimate in linear regression, *Communications in Statistics - Theory and Methods*. 22 (1993) 393-402. doi:10.1080/03610929308831027.

[6] P. Karlsson, K. Månsson, B.M.G. Kibria, A Liu estimator for the beta regression model and its application to chemical data, *Journal of Chemometrics*. 34 (2020) e3300. doi:10.1002/cem.3300.

[7] K. Liu, Using Liu-Type Estimator to Combat Collinearity, *Communications in Statistics - Theory and Methods*. 32 (2003) 1009-1020. doi:10.1081/STA-120019959.

[8] Z. Yahya Algamal, Abonazel, Developing a Liu-type estimator in beta regression model, *Concurrency and Computation: Practice and Experience*. 34 (2022) e6685. doi:10.1002/CPE.6685.

[9] R. Ospina, F. Cribari-Neto, K.L.P. Vasconcellos, Improved point and interval estimation for a beta regression model, *Computational Statistics & Data Analysis*. 51 (2006) 960-981. doi:10.1016/j.csda.2005.10.002.

[10] A.B. Simas, W. Barreto-Souza, A. V Rocha, Improved estimators for a general class of beta regression models, *Computational Statistics and Data Analysis*. 54 (2010) 348-366. doi:10.1016/j.csda.2009.08.017.

[11] R. Ospina, S.L.P. Ferrari, On Bias Correction in a Class of Inflated Beta Regression Models, *International Journal of Statistics and Probability*. 1 (2012) p269. doi:10.5539/IJSP.V1N2P269.

[12] K. Kadiyala, A class of almost unbiased and efficient estimators of regression coefficients, *Economics Letters*. 16 (1984) 293-296. doi:10.1016/0165-1765(84)90178-2.

[13] K. Ohtani, On small sample properties of the almost unbiased generalized ridge estimator, *Communications in Statistics - Theory and Methods*. 15 (1986) 1571-1578. doi:10.1080/03610928608829203.

[14] Y. Asar, M. Korkmaz, Almost unbiased Liu-type estimators in gamma regression model, *Journal of Computational and Applied Mathematics*. 403 (2022) 113819. doi:10.1016/J.CAM.2021.113819.

[15] M. Amin, M. Qasim, A. Yasin, M. Amanullah, Almost unbiased ridge estimator in the gamma regression model, *Communications in Statistics - Simulation and Computation*. 51 (2022) 3830-3850. doi:10.1080/03610918.2020.1722837.

[16] M. Qasim, K. Månsson, M. Amin, B.M. Golam Kibria, P. Sjölander, Biased adjusted Poisson ridge estimators-method and application, *Iran. J. Sci. Technol. Trans. A Sci.* 44 (2020) 1775-1789. doi:10.1007/s40995-020-00974-5.

[17] M.I. Alheety, M. Qasim, K. Månsson, B.M.G. Kibria, Modified almost unbiased two-parameter estimator for the Poisson regression model with an application to accident data, *SORT*. 45 (2021) 121-142.

[18] A. Erkoç, E. Ertan, Z. Yahya Algamal, K. Ulaş Akay, The beta Liu-type estimator: simulation and

application, *Hacettepe Journal of Mathematics and Statistics*. 52 (2023) 828-840. doi:10.15672/HUJMS.1145607.

[19] M.R. Abonazel, I.M. Taha, Beta ridge regression estimators: simulation and application, *Communications in Statistics - Simulation and Computation*. (2021) 1-13. doi:10.1080/03610918.2021.1960373.

[20] J. Xu, H. Yang, More on the Bias and Variance Comparisons of the Restricted Almost Unbiased Estimators, *Communications in Statistics - Theory and Methods*. 40 (2011) 4053-4064. doi:10.1080/03610926.2010.505693.

[21] M. Erişoğlu, N. Yaman, Ridge Tahminine Dayalı Kantil Regresyon Analizinde Yanlılık Parametresi Tahminlerinin Performanslarının Karşılaştırılması, *Necmettin Erbakan University Journal of Science and Engineering*. 1 (2019) 103-111.