STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–12.
Published online in International Academic Press (www.IAPress.org)

A Modified Jackknife Liu-Type Estimator for the Gamma Regression

Ahmed Mutlag Algboory [1], Ahmed Naziyah Alkhateeb [2,*], Zakariya Yahya Algamal [3]

[1] *College of Physical Education and Sport Sciences, University of Samarra, Iraq*
[2] *Department of Operation Research and Intelligent Techniques, University of Mosul, Iraq*
[3] *Department of Statistical Information, University of Mosul, Iraq*

Abstract Additional methods were suggested to enhance the biased estimation in the multiple linear regression model. The jackknife-biased estimate approach is essential for addressing high variance and multicollinearity issues. Reduce the effects of multicollinearity with the Liu estimator: This shrinkage method is attractive on several occasions. This document aims to derive a Jackknifed Liu-type Gamma estimator (JGLTE) and a Modified Jackknifed Liu-type Gamma estimator (MJGLTE) when multicollinearity exists. Based on Monte Carlo simulations, the proposed estimate outperforms the maximum likelihood estimator (MLE) in terms of mean square error (MSE). Finally, we illustrate the performance of this estimator using real-world data.

Keywords Jackknife Liu-type, Gamma Regression, Multicollinearity, Monte Carlo Simulation

DOI: 10.19139/soic-2310-5070-2304

1. Introduction

The explanatory variables in many regression model applications include a built-in correlation. When correlations are significant, the estimation of the regression parameters becomes unstable, making it challenging to interpret the estimates of the regression coefficients, and the maximum likelihood (ML) approach becomes more sensitive. It is difficult to assess the distinct impacts of each explanatory variable in the regression model when multicollinearity is present. Furthermore, the regression coefficients sample variance will increase, impacting prediction and inference. The literature has presented a wide range of solutions to the multicollinearity problem.

One of the important application statistics is the gamma regression model (GR), a specific kind of generalized linear model (GLM). The iterative reweighted least square (IRLS) algorithm should be used fundamental method Maximum likelihood Estimator (MLE) to estimate the regression coefficients of the GR model [8] [18]. It is tough to construct a meaningful statistical inference in the presence of multicollinearity since the ML method's estimated variance gets inflated. Hoerl and Kennard [10] suggested the ordinary ridge regression approach for the linear regression model (LRM) in order to address the multicollinearity issue by adding the biasing parameter k that has the range 0 to ∞. Several ridge parameters were suggested by [17] [18] for the Poisson regression, [25] presented a novel Jackknife estimator For the Poisson regression [17].

Ridge regression's disadvantage, according to Liu [11], In order to solve the multicollinearity issue, the Liu estimator is taken into consideration as a substitute for the ML estimator. Liu estimator's benefit is that a linear function of the biased parameter d with a range between 0 to 1. The Liu estimator is useful method for mitigating bias when dealing with the problem of collinearity among the explanatory variables. Due to the linear nature of

*Correspondence to: Ahmed Naziyah alkhateeb (Email: ahmed.alkhateeb@uomosul.edu.iq). Department of Operation Research and Intelligent Techniques, College of Computer Sciences and Mathematics, University of Mosul, Nineveh, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

the variable d, many researchers have chosen to employ the Liu estimator instead of ridge regression [17]. These researchers have proposed the optimal Liu value for logit regression and negative binomial models. Yang and Chang [27] proposed a two-parameter estimator that combines the ridge and Liu estimators. There is very little material available on the Liu estimator for the GLM. [4] suggested a new estimator that blends the $(r-k)$ class estimator and the Liu estimator. [16] suggested a few biased Liu parameters for negative binomial regression. [26] have conducted further research on the restricted Liu estimator for the logistic regression model in order to address the issue of multicollinearity.[3] Proposed a novel estimate that combines the estimator of the (r-(k-d)) class and the Gamma regression. Liu-type estimator refers to a certain method of estimation. In a Liu-type estimator, employing a substantial shrinkage value is permissible, as there exists an additional parameter to optimize the estimator for a good fit. Despite the Liu-type estimator's commendable characteristics, it exhibits a smaller bias. This bias can be mitigated by subjecting a biased estimator to a jackknife procedure. The jackknife procedure involves processing experimental data to generate a statistical estimator for unknown parameters. A truncated sample is employed to compute a specific function of the estimators. The advantage of the jackknife procedure lies in its ability to yield an estimator with minimal bias while capitalizing on the advantageous properties of large samples, as highlighted by [25].

The available literature shows that no researcher recommends the jackknife technique to the gamma Liu-type estimator. In this article, we applied the JGLTE and MJGLTE. The concept of our proposed estimator is to decrease the shrinkage parameter, thereby improving the resulting estimator with a minimal amount of bias.

2. Gamma regression

One way to think about the gamma distribution is as an extension of the exponential distribution, where $\lambda > 0$ and mean $1/\lambda$. The duration until the occurrence of the initial event is denoted by an exponential random variable with an average of $1/\lambda$. where a Poisson process with mean λ generates events, On the other hand, the waiting period until the occurrence of the $\alpha - th$ event is represented by the gamma random variable X, such that [12]:

$$X = \sum_{i}^{\alpha} Y_i \tag{1}$$

The Gamma regression (GR) model is used when the dependent variable is positively skewed with its mean proportional to the dispersion parameter. Assume that $Y_1, Y_2, \ldots, Y_n$ are explanatory random variables, and $y_1, y_1, \ldots y_n$ that the probability density function for the matching observations from the gamma distribution is displayed below:

$$\mathrm{f}\left(\mathrm{y_i};\alpha,\phi\right) = \frac{\mathrm{y_i}^{\alpha-1}\,\mathrm{e}^{-\mathrm{y_i}/\phi}}{\Gamma\left(\alpha\right)\left(\phi\right)^{\alpha}} \qquad \mathrm{y_i} \geq 0 \text{ and } \alpha,\phi > 0 \tag{2}$$

Where (ϕ) represents a scale parameter and (α) represents a shape parameter that is greater than or equal to zero, such that:

$$\mathrm{E}\left(\mathrm{Y_i}\right) = \mu\mathrm{i} = \alpha\phi = \theta\mathrm{i} \tag{3}$$

This is referred to as the canonical parameter as well and:

$$\mathrm{var}\left(\mathrm{Yi}\right) = \alpha\phi^2 = 1/\left(\mathrm{k}\theta_{\mathrm{i}}^{\ 2}\right) \qquad \text{where} \quad \theta\mathrm{i} = \mathrm{e}^{\mathrm{x_i}^{\mathrm{T}}\beta} \tag{4}$$

Such that:

$$\mathrm{X_i} = \left(\mathrm{X_{i1}},\mathrm{X_{i2}},\ldots,\ \mathrm{X_{ip}}\right)' \ \ \mathrm{i} = 1,2,\ldots,\mathrm{n} \quad \mathrm{j} = 1,2,\ldots,\mathrm{p} \qquad \mathrm{n} > \mathrm{p}$$

Here, n represents the size of the sample, where p denotes the count of explanatory variables. Generally, to determine the parameters, the maximum likelihood estimate is used. The dependent variable of a random component in the GLM is frequently the probability density function, which is a member of the distribution's exponential family. The gamma density function is a member of the exponential family of distributions, the general density of the exponential family may be found using the following formula:

$$f(y,\theta,\phi)=\left\{\frac{y\,\theta-b(\theta)}{\alpha(\phi)}+C(y,\phi)\right\} \tag{5}$$

The symbols ϕ represent the dispersion parameter, represents the location parameter, and b () represents the cumulant function. The average function y_i for the generalized regression model (GR) is defined as:

$$\mathrm{g}(\mu \mathrm{i})=\ \ln(\mu \mathrm{i})=\ \mathrm{X}_{\mathrm{i}}^{'}\beta \qquad \mathrm{i}\ =\ 1,\ 2,\ \ldots\ ,\ \mathrm{n}$$

$$\eta \mathrm{i}\ =\ \ln(\mu \mathrm{i})\ =\ \mathrm{X}_{\mathrm{i}}^{'}\beta$$

Let X be an $n\times p$ information matrix with p explanatory variables. Each row x_i of X is denoted as $X=[X_{i1},X_{i2},...,X_{ip}]$. The slope coefficients are represented by the p×1 vector β [6]. The function $g(.)$ is the log link function used for the GR model, where $\eta i\ =\ g(\mu i)$. The log likelihood expression for the GR model can be represented for Eq :

$$\begin{aligned} l(y;\mu,\varphi)=\sum_{i=1}^{n}\Big[\Big\{\tfrac{yi}{\mu i}-(-ln(\mu i))\Big\}(-\varphi)^{-1}+(1-\varphi)(\varphi)^{-1}ln(yi)\\ -ln(\varphi)(\varphi)^{-1}-ln\Gamma(\varphi)^{-1}\Big]\end{aligned} \tag{6}$$

Assume that the estimates $\hat{\beta}$, $\hat{\mu}$, and $\hat{\varphi}$ may be determined via the ML technique by utilizing the Newton–Raphson iterative method to maximize Equation . The ML estimates can be found by solving:

$$S(\beta)\ =\ \frac{\partial l(\mu i,\ \varphi)}{\partial\beta}\ =\ \frac{1}{\alpha(\varphi)}\left(yi\ -\ exp\left(X_i^{'}\beta\right)\right)X_i^{'}\ =\ 0 \tag{7}$$

where $\alpha(\varphi)=\varphi$ and $S(\beta)$ stands for the vector score. Equation is non-linear; hence the iterative Fisher scoring method must be used to estimate the unknown parameter. Let $\beta^{(m)}$ represent the predicted ML of β after m iterations which can be expressed as:

$$\beta^{(m+1)}\ =\ \beta^{(m)}\ +\ \left\{I\left(\beta^{(m)}\right)\right\}^{-1}\ S\left(\beta^{(m)}\right) \tag{8}$$

Where:

$$S(\beta)\ =\ \frac{\partial\ell(\beta)}{\partial\beta}$$

and:

$$\{I(\beta\ )\}^{-1}=(-E\left(\frac{\partial^2\ \ell(\beta)\ )}{\partial\beta\partial\beta'}\right)^{-1}$$

Here $I\left(\beta^{(m)}\right)$ is a $p\ \times\ p$ Fisher information matrix, $S\left(\beta^{(m)}\right)$ and $I\left(\beta^{(m)}\right)$ are evaluated at $\beta^{(m)}$. Using the IRLS approach, the ML method is discovered at convergence in deviance for Equation :

$$\hat{\beta}_{MLE}=(\mathrm{X}^{\mathrm{T}}\hat{\mathrm{W}}\mathrm{X})^{-1}\ \mathrm{X}^{\mathrm{T}}\hat{\mathrm{W}}\hat{\mathrm{z}} \tag{9}$$

In this case $\hat{W}=\ diag\left[\hat{\mu}_1^2\ ,\hat{\mu}_2^2,\ \ldots\ ,\hat{\mu}_n^2\right]$ where the adjusted dependent variable is $\hat{z}=\ \hat{\eta}i\ +\ (yi\ -\ \hat{\mu}_i)/\hat{\mu}_i^2$ and the estimated mean with a log link function is $\hat{\mu}_i\ =\ exp\left(x_i^{'}\hat{\beta}_{MLE}\right)$. It is commonly known that $\hat{\beta}_{MLE}$ covariance matrix is:

$$\mathrm{cov}\left(\hat{\beta}_{MLE}\right)=\left(-E\left(\frac{\partial^2\ \ell(\beta)}{\partial\beta_i\partial\beta_k}\right)\right)^{-1}=\varphi\ (\mathrm{X}^{\mathrm{T}}\hat{\mathrm{W}}\mathrm{X})^{-1}$$

The (MSE) of (MLE) is given by:

$$\mathrm{MSE}\left(\hat{\beta}_{\mathrm{MLE}}\right)=\mathrm{E}\left(\hat{\beta}_{MLE}-\beta\right)^{\mathrm{T}}\left(\hat{\beta}_{MLE}-\beta\right)$$

$$\therefore \mathrm{MSE}\left(\hat{\beta}_{\mathrm{MLE}}\right) = \mathrm{tr}\left[\varphi\ (\mathrm{X^T\hat{W}X})^{-1}\right] \quad = \varphi \sum_{\mathrm{j=1}}^{\mathrm{p}} \frac{1}{\lambda_{\mathrm{j}}} \tag{10}$$

where λ_{j} is the eigenvalue of the $\mathcal{D} = \mathcal{X}^{\mathrm{T}}\hat{\mathrm{W}}\mathcal{X}$ and tr (.) is the trace of a $\mathcal{D}$. Furthermore, the following is also taken into consideration for the matrix $\mathcal{M}$ eigenvalue decomposition:

$$\mathcal{D} = \mathcal{Q}^{\mathrm{T}}\Lambda\mathcal{Q}$$

such that $\Lambda = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots,\ \lambda_{\mathrm{p}}\right)$ and $\mathcal{Q}$ is the orthogonal matrix made up of the eigenvectors corresponding to the eigenvalues of ($\mathcal{D}$). The MSE of the MLE inflates when one or more of the eigenvalues approaches 0, as can be easily observed, and this has a negative impact on the regression coefficients.

3. Gamma Liu estimator

The popular estimator Liu has gained recognition and is used in generalized linear models. The gamma dependent variable was taken into consideration by [13] the authors when examining the Liu gamma estimator's performance using both real data applications and Monte Carlo simulations. The gamma Liu estimator (GLE) is defined as:

$$\hat{\beta}_{Liu} = (\mathcal{D} + \mathcal{I})^{-1}\,(\mathcal{D} + d\mathcal{I}) \quad \hat{\beta}_{MLE} \tag{11}$$

Where $\boldsymbol{\mathcal{T}}_{Liu} = (\mathcal{D} + \mathcal{I})^{-1}\,(\mathcal{D} + d\mathcal{I})$ and biased Liu $0 < d < 1$. The GLE's bias vector and covariance matrix can be obtained respectively:

$$\mathbf{b}_{GLE} = bias\left(\hat{\beta}_{Liu}\right) = -\left(1-d\right)(\mathcal{D} + \mathcal{I})^{-1} \ \ \beta\mathbf{b}_{GLE} = bias\left(\hat{\beta}_{Liu}\right) = -\left(1-d\right)(\mathcal{D} + \mathcal{I})^{-1} \ \ \beta$$

$$cov\left(\hat{\beta}_{Liu}\right) = \varphi\,\boldsymbol{\mathcal{T}}_{Liu}\mathcal{D}^{-1}\mathcal{T}_{Liu}{}^{T}$$

The following MMSE and MSE functions can be obtained, respectively, by using the covariance and bias of GLE:
$\mathrm{MMSE}\left(\hat{\beta}_{Liu}\right) = cov\left(\hat{\beta}_{Liu}\right) + \mathbf{b}_{GLE}\mathbf{b}_{GLE}{}^{T} = \varphi\,\boldsymbol{\mathcal{T}}_{Liu}\mathcal{D}^{-1}\mathcal{T}_{Liu}{}^{T} + (1-?)^2\,(\mathcal{D} + \mathcal{I})^{-1}\ \ \beta\beta\,(\mathcal{D} + \mathcal{I})^{-1}$
$\mathrm{MSE}\left(\hat{\beta}_{Liu}\right) = \mathrm{tr}\ \left[\mathrm{MMSE}\left(\hat{\beta}_{Liu}\right)\right] = \varphi\ \sum_{j=1}^{p} \frac{(\lambda_{\mathrm{j}}+d)^2}{\lambda_{\mathrm{j}}(\lambda_{\mathrm{j}}+1)^2} + (d-1)^2\ \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_{\mathrm{j}}+1)^2}$

4. Gamma Liu-type estimator

Introduced a novel biased estimator known as the Liu-type estimators (LTE) [14], The following gamma Liu-type estimator (GLTE) is given:

$$\hat{\beta}_{GLTE} = \boldsymbol{\mathcal{T}}_{k}^{-1}\boldsymbol{\mathcal{T}}_{d}\hat{\beta}_{MLE} \tag{12}$$

Where $-\infty < d < \infty$, $\mathcal{K} > 0$, $\boldsymbol{\mathcal{T}}_{k}^{-1} = (\mathcal{D} + \mathcal{K}\mathcal{I})$, $\boldsymbol{\mathcal{T}}_{d} = (\mathcal{D} - d\mathcal{I})$, we obtain the bias vector and covariance of GLTE as:

$$bias\left(\hat{\beta}_{GLTE}\right) = -\left(d+K\right)\Upsilon_{K}^{-1}\beta$$

$$cov\left(\hat{\beta}_{GLTE}\right) = \varphi\,\boldsymbol{\mathcal{T}}_{LTE}\mathcal{D}^{-1}\mathcal{T}_{LTE}{}^{T}$$

As a result, the MMSE and MSE functions of GLTE are calculated as follows:

$$\mathrm{MMSE}\left(\hat{\beta}_{GLTE}\right) = cov\left(\hat{\beta}_{GLTE}\right) + \mathbf{b}_{GLTE}\mathbf{b}_{GLTE}{}^{T} \tag{13}$$

$$\mathrm{MSE}\left(\hat{\beta}_{GLTE}\right) = \mathrm{tr}\ \left[\mathrm{MMSE}\left(\hat{\beta}_{GLTE}\right)\right] \tag{14}$$

$$= \varphi \sum_{j=1}^{p} \frac{(\lambda_j - d)^2}{\lambda_j (\lambda_j + \mathcal{K})^2} + (d + \mathcal{K})^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + \mathcal{K})^2} \tag{15}$$

The optimal value of d_j can be found by setting equation (15) to zero and solving for d_j, which may then be shown as:

$$d_j = \frac{\phi \sum_{j=1}^{p} \frac{1}{(\lambda_j + \kappa)^2} - \kappa \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + \kappa)^2}}{\phi \sum_{j=1}^{p} \frac{1}{\lambda_j (\lambda_j + \kappa)^2} + \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j + \kappa)^2}}$$

5. First Proposed Estimator: Jackknifed Liu-type Gamma Estimator (JGLTE)

To solve the bias in the ridge estimator, [24] proposed the Jackknife technique for the linear regression model. Many articles, including those by [1], [25], [2], and [7], have proposed many jackknife estimators. In this section, the jackknife Liu-type Estimator of the JGLTE model is proposed. Let $\Lambda = \ diag\,(\lambda 1, \ ..., \lambda p)$ and $Q \ = \ (q1, \ ..., \ qp)$, respectively, be the matrices of eigenvalues and eigenvectors of $\mathcal{D} = \mathrm{X}'\hat{\mathrm{W}}\mathrm{X}$, such that $\mathcal{D} = \mathcal{T}\Lambda\mathcal{T}'$, where $\mathcal{T}$ is an orthogonal matrix and Λ is a diagonal matrix. Consequently, the GMLE model in Eq.(5) can be re-written as:

$$\hat{\tau}_{GMLE} = \Lambda^{-1}\, \mathcal{S}^{\mathrm{T}}\hat{\mathrm{V}}\hat{\mathrm{z}} \tag{16}$$

$$\hat{\beta}_{GMLE} = Q\hat{\tau}_{GMLE} \tag{17}$$

In [14], a new estimator is proposed for τ. This estimator is biased and is called Liu-type estimator (LTE), and is defined as follows:

$$\hat{\tau}_{GLTE}\,(\mathcal{K}, d) = (\Lambda + \mathcal{K}\mathcal{I})^{-1} \,(Q'y - d\hat{\tau}_{GMLE})$$

$$\hat{\tau}_{GLTE}\,(\mathcal{K}, d) = (\Lambda + \mathcal{K}\mathcal{I})^{-1} \left(Q'y - d\Lambda^{-1}Q'y\right)$$

$$= \left[\mathcal{I} - (\Lambda + \mathcal{K}\mathcal{I})^{-1} (\mathcal{K} + y)\right] \hat{\tau}_{GMLE}$$

$$= \mathcal{L}\,(\mathcal{K}, d)\, \hat{\tau}_{GMLE}$$

Where $\mathcal{L}\,(\mathcal{K}, d) = (\Lambda + \mathcal{K}\mathcal{I})^{-1}\,(\Lambda - d\mathcal{I})$, $\hat{\tau}_{GLTE}$ has a bias vector defined as bias:

$$bias\,(\hat{\tau}_{GLTE}) \ = \ (\mathcal{L}\,(\mathcal{K}, d) \ - \ \mathcal{I})\,\tau$$

and a covariance matrix

$$cov\,(\hat{\tau}_{GLTE}) \ = \sigma^2\, \mathcal{L}\,(\mathcal{K}, d)\,\Lambda^{-1}\mathcal{L}\,(\mathcal{K}, d)$$

A jackknifed form of $\hat{\tau}_{GLTE}$ can be proposed by utilizing the works of [9], [24], [21] and [15]. The jackknife approach was established in [23] and [23] to lower the bias value. With a few notable exceptions, balanced models can be fitted using the jackknife approach, according to [9]. Following several algebraic operations, the jackknifed estimator is obtained by removing the $i - th$ observation $(q_i\ , y_i)$:

$$\hat{\tau}_{GLTE}\,(\mathcal{K}, d) = \left(Q'_{-i}\,\hat{\mathrm{W}}_{-i}Q_{-i} + \mathcal{K}\mathcal{I}\right)^{-1} \left(Q'_{-i}\,\hat{\mathrm{W}}_{-i}Q_{-i} - d\mathcal{I}\right)\left(Q'_{-i}\,\hat{\mathrm{W}}_{-i}Q_{-i}\right)^{-1} Q'_{-i}y_{-i}$$

$$= \left(\mathcal{A} - Q'_{-i}\,\hat{\mathrm{W}}_{-i}Q_{-i} + \mathcal{K}\mathcal{I}\right)^{-1} (Q'y - q_i y_i)$$

$$= \left(\mathcal{A}^{-1} + \frac{\mathcal{A}^{-1}q_i w_i q_i' \mathcal{A}^{-1}}{1 - q_i\mathcal{A}^{-1}q_i}\right)$$

$$= \mathcal{A}^{-1}\,Q'y - \mathcal{A}^{-1}q_i y_i \left[\frac{\mathcal{A}^{-1}q_i w_i q_i'\mathcal{A}^{-1}}{1 - q_i'\mathcal{A}^{-1}q_i}\,Q'y - \frac{\mathcal{A}^{-1}q_i w_i q_i'\mathcal{A}^{-1}}{1 - q_i'\mathcal{A}^{-1}q_i} q_i y_i\right]$$

$$= \hat{\tau}_{GLTE}(\mathcal{K},d) + \mathcal{A}^{-1}q_i y_i \left[1 - \frac{q_i^{'}\mathcal{A}^{-1}q_i}{1-q_i^{'}\mathcal{A}^{-1}q_i}\right] + \frac{\mathcal{A}^{-1}q_i w_i q_i^{'}}{1-q_i^{'}\mathcal{A}^{-1}q_i}\,\hat{\tau}_{GLTE}(\mathcal{K},d)$$

$$= \hat{\tau}_{GLTE}(\mathcal{K},d) - \mathcal{A}^{-1}q_i \frac{\mathcal{A}^{-1}q_i\left(y_i - q_i^{'}\hat{\tau}_{GLTE}(\mathcal{K},d)\right)}{1-q_i^{'}\mathcal{A}^{-1}q_i}$$

$$= \hat{\tau}_{GLTE}(\mathcal{K},d) - \frac{\mathcal{A}^{-1}q_i e_i}{1-\xi_i} \tag{18}$$

where $q_i^{'}$ is the i-th row of the matrix Q, $e_i = y_i - q_i^{'}\hat{\tau}_{GLTE}(\mathcal{K},d)$ is the Liu-type residual, $Q_{-i}^{'}\,\hat{\mathrm{W}}_{-i}Q_{-i} = Q'\hat{\mathrm{W}}Q - q_i w_i q_i^{'}$, $Q_{-i}^{'}y_{-i} = Q'y - q_i y_i$, $\xi_i = q_i^{'}\mathcal{A}^{-1}q_i$ is the distance factor and$\mathcal{A}^{-1} = (\Lambda + \mathcal{K}\mathcal{I})^{-1}\left(\mathcal{I} - d\Lambda^{-1}\right) = \mathcal{L}(\mathcal{K},d)\,\Lambda^{-1}$. Due to the non-zero value of ξ_i, which indicates an imbalance in the model, we employ the weighted jackknife approach. Weighted pseudo values are defined as follows:

$$\mathcal{Z}_i = \hat{\tau}_{GLTE}(\mathcal{K},d) + \mathrm{n}\ (1-\xi_i)\left[(\hat{\tau}_{GLTE}(\mathcal{K},d)) - (\hat{\tau}_{GLTE}(\mathcal{K},d))\right]$$

The weighted jackknifed estimator of τ is obtained as:

$$\hat{\tau}_{JGLTE}(\mathcal{K},d) = \frac{1}{n}\sum_{i=1}^{n}\mathcal{Z}_i = \hat{\tau}_{GLTE}(\mathcal{K},d) + \mathcal{A}^{-1}\sum_{i=1}^{n}q_i e_i$$

Since:

$$\sum_{i=1}^{n}q_i e_i = \sum_{i=1}^{n}q_i\left(y_i - q_i^{'}\hat{\tau}_{GLTE}(\mathcal{K},d)\right) = \left(\mathcal{I} - \mathcal{A}^{-1}\right)Q'y$$

it follows that:

$$\hat{\tau}_{JGLTE}(\mathcal{K},\mathrm{d}) = \hat{\tau}_{GLTE}(\mathcal{K},\mathrm{d}) + \mathcal{A}^{-1}Q'y - \mathcal{A}^{-1}\Lambda\mathcal{A}^{-1}Q'y$$
$$= \left(2\mathcal{I} - \mathcal{A}^{-1}\Lambda\right)\hat{\tau}_{GLTE}(\mathcal{K},\mathrm{d})$$

However, since $\mathcal{I} - \mathcal{A}^{-1}\Lambda = \mathcal{I} - (\Lambda + \mathcal{K}\mathcal{I})^{-1}(\Lambda - \mathrm{d}\mathcal{I}) = \mathcal{I} - \mathcal{L}(\mathcal{K},\mathrm{d})$
we obtain:

$$\hat{\tau}_{JGLTE}(\mathcal{K},\mathrm{d}) = (2\mathcal{I} - \mathcal{L}(\mathcal{K},\mathrm{d}))\,\hat{\tau}_{GLTE}(\mathcal{K},\mathrm{d})$$
$$= (2\mathcal{I} - \mathcal{L}(\mathcal{K},\mathrm{d}))\,\mathcal{L}(\mathcal{K},\mathrm{d})\,\hat{\tau}_{MLE}$$

The bias part and the variance of $\hat{\tau}_{JGLTE}(\mathcal{K},\mathrm{d})$ are obtained as, respectively:

$$bias\,(\hat{\tau}_{JGLTE}(\mathcal{K},\mathrm{d})) = -(\mathcal{I} - \mathcal{L}(\mathcal{K},\mathrm{d}))^2\ \tau,$$

and a covariance matrix

$$cov\,(\hat{\tau}_{JGLTE}(\mathcal{K},\mathrm{d})) = \sigma^2\left[(2\mathcal{I} - \mathcal{L}(\mathcal{K},\mathrm{d}))\,\mathcal{L}(\mathcal{K},\mathrm{d})\,\Lambda^{-1}\mathcal{L}(\mathcal{K},\mathrm{d})^{'}\,(2\mathcal{I} - \mathcal{L}(\mathcal{K},\mathrm{d}))\right]^{'}$$

The MSEMs of the jackknifed Gamma Liu-type estimator (JGLTE) and the Gamma Liu-type estimator (GLTE) are given as follows [5]:

$$MSEM\,(\hat{\tau}_{JGLTE}(\mathcal{K},d)) = cov\,(\hat{\tau}_{JGLTE}(\mathcal{K},d)) + bias\,(\hat{\tau}_{JGLTE}(\mathcal{K},d))\ bias\,(\hat{\tau}_{JGLTE}(\mathcal{K},d))^{\ '}$$
$$= \sigma^2\,(2\mathcal{I} - \mathcal{L}(\mathcal{K},d))\,\mathcal{L}(\mathcal{K},d)\,\Lambda^{-1}\mathcal{L} \tag{19}$$

$$MSEM\,(\hat{\tau}_{JGLTE}(\mathcal{K},d)) = \sigma^2\mathcal{L}(\mathcal{K},d)\,\Lambda^{-1}\mathcal{L}(\mathcal{K},d)^{'} + (\mathcal{L}(\mathcal{K},d)\ -\ \mathcal{I})\ \tau\tau'(\mathcal{L}(\mathcal{K},d-\ \mathcal{I})$$

6. Second Proposed Estimator: Modified Jackknifed Liu-type Gamma Estimator (MJGLTE)

In this section, we propose a new estimator for τ. The proposed estimator is designated as the modified jackknifed Gamma Liu-type estimator (MJGLTE) denoted by $\hat{\tau}_{MJGLTE}(\mathcal{K}, d)$ [20]:

$$\hat{\tau}_{MJGLTE}(\mathcal{K}, \mathrm{d}) = \left[\mathcal{I} - (\mathcal{K}+\mathrm{d})^2 (\Lambda + \mathcal{K}\mathcal{I})^{-2}\right] \left[\mathcal{I} - (\mathcal{K}+\mathrm{d})(\Lambda + \mathcal{K}\mathcal{I})^{-1}\right] \hat{\tau}_{MLE}$$

The expressions of bias are:

$$bias\left(\hat{\tau}_{MJGLTE}(\mathcal{K}, \mathrm{d})\right) = -(\mathcal{K}+\mathrm{d})(\Lambda + \mathcal{K}\mathcal{I})^{-1} \mathcal{H} \ (\Lambda + \mathcal{K}\mathcal{I})^{-1} \tau$$

and a covariance matrix:

$$cov\left(\hat{\tau}_{MJGLTE}(\mathcal{K}, \mathrm{d})\right) = \sigma^2 \Psi \Lambda^{-1} \Psi'$$

The mean squared error matrix (MSEM) of the Modified jackknifed Gamma Liu-type estimator (MJGLTE) are given as follows:

$$MSEM\left(\hat{\tau}_{MJGLTE}(\mathcal{K}, \mathrm{d})\right) = \sigma^2 \Psi \Lambda^{-1} \Psi' (\mathcal{K}+\mathrm{d})^2 (\Lambda + \mathcal{K}\mathcal{I})^{-1} \mathcal{H} \ (\Lambda + \mathcal{K}\mathcal{I})^{-1} \tau\tau' \left[(\Lambda + \mathcal{K}\mathcal{I})^{-1} \mathcal{H} \ (\Lambda + \mathcal{K}\mathcal{I})^{-1}\right]'$$

Such that:

$$\mathcal{H} = \mathcal{I} + (\mathcal{K}+\mathrm{d})(\Lambda + \mathcal{K}\mathcal{I})^{-1} - (\mathcal{K}+\mathrm{d})^2 (\Lambda + \mathcal{K}\mathcal{I})^{-2} = \mathcal{I} + \mathcal{L}(\mathcal{K}, \mathrm{d}) - \mathcal{L}(\mathcal{K}, \mathrm{d})^2$$

And

$$\Psi = (2\mathcal{I} - \mathcal{L}(\mathcal{K}, \mathrm{d}))\, \mathcal{L}(\mathcal{K}, \mathrm{d})^2$$

7. Monte Carlo Simulation

A Monte Carlo simulation establishes the performance evaluations of GLTE under varying multicollinearity levels [28, 29] by using RStudio 2024.12.1 Build 563.

7.1. Simulation Design

The response variable of n observations follow gamma regression model [13, 30], as following:

$$y \sim Gam\left(\frac{\vartheta^2}{\nu}, \frac{\nu}{\vartheta}\right)$$

Here $\vartheta = \exp(X'\beta)$, νdenotes ϑ, the parameter vector $\beta = [\beta_1, \beta_2, ..., \beta_p]$ is equal $\sum_{j=1}^{p} \beta_j^2 = 1$ and off course the explanatory variables is defined as:

$$X_{ij} = \left(1 - \ell^2\right)^{1/2} w_{ij} + \ell w_{ip}$$

where ℓ is the relation between the explanatory variables and $w'_{ij}s$ are independent standard normal pseudo-random numbers.
Where $\hat{\beta}$ is the estimated coefficient for the estimator used. For the value of κ, the best method was used as:

$$\mathcal{K} = \max\left(\frac{1}{\omega_j}\right) \quad , \quad j = 1, 2, ..., p$$

Where $\omega_j = \sqrt{\hat{\sigma} / \hat{\alpha}_j^2}$

Table 1. when n=100

	MSE Method	$\rho = 0.90$	$\rho = 0.95$	$\rho = 0.99$
$p = 4$	*MLE*	*0.08516720*	*0.16412397*	*0.83767855*
	LIU	*0.07756584*	*0.13800186*	*0.52542411*
	GLTE	*0.06755732*	*0.10765406*	*0.30920398*
	JGLTE	*0.07981816*	*0.13473082*	*0.38039059*
	MJGLTE	*0.06629650*	*0.10474486*	*0.29282898*
$p = 8$	*MLE*	*0.10238572*	*0.21984955*	*1.00893942*
	LIU	*0.08719720*	*0.16742328*	*0.55313540*
	GLTE	*0.06596976*	*0.10556239*	*0.25002168*
	JGLTE	*0.08521492*	*0.14050221*	*0.47008751*
	MJGLTE	*0.06216450*	*0.09920032*	*0.19574131*

Table 2. when n=200

	MSE Method	$\rho = 0.90$	$\rho = 0.95$	$\rho = 0.99$
$p = 4$	*MLE*	*0.03849833*	*0.08054200*	*0.41612926*
	LIU	*0.03667397*	*0.07353502*	*0.30235040*
	GLTE	*0.03414572*	*0.06434660*	*0.19760315*
	JGLTE	*0.03780903*	*0.07571560*	*0.24362274*
	MJGLTE	*0.03390618*	*0.06320418*	*0.19585618*
$p = 8$	*MLE*	*0.04580515*	*0.09693123*	*0.46570838*
	LIU	*0.04213151*	*0.08271436*	*0.29893149*
	GLTE	*0.03634023*	*0.06270486*	*0.15593178*
	JGLTE	*0.04338668*	*0.08102833*	*0.21866854*
	MJGLTE	*0.03529622*	*0.05901512*	*0.14167689*

Table 3. when n=500

	MSE Method	$\rho = 0.90$	$\rho = 0.95$	$\rho = 0.99$
$p = 4$	*MLE*	*0.01531791*	*0.02920502*	*0.16033022*
	LIU	*0.01503117*	*0.02814826*	*0.13525437*
	GLTE	*0.01461999*	*0.02671485*	*0.10536677*
	JGLTE	*0.01527313*	*0.02887990*	*0.13292834*
	MJGLTE	*0.01460074*	*0.02660643*	*0.10202614*
$p = 8$	*MLE*	*0.01764312*	*0.03650899*	*0.17693787*
	LIU	*0.01703358*	*0.03405376*	*0.13726153*
	GLTE	*0.01600047*	*0.03007160*	*0.09073022*
	JGLTE	*0.01746795*	*0.03515101*	*0.11852901*
	MJGLTE	*0.01589131*	*0.02939902*	*0.08635068*

As the sample size has direct effect on the prediction accuracy, three representative values of the sample size are considered: Which are 100, 200 and 500. Besides, two numbers of the explanatory variables are taken as $p = 4$ and $p = 8$ because it can be seen that as the number of the explanatory variables rises, the MSE increases. Also because of interesting in the effect of multicollinearity, in which the degrees of correlation considered to be more *Table* 1*: MSE when n=100*

MSE is calculated as $MSE = (y_i - \bar{y})^2$ Where y_i is the $i - th$ observation and $\bar{y}$ is the mean of the observed values. The table below shows the MSE of each replication after running the simulation for 50 cycles.

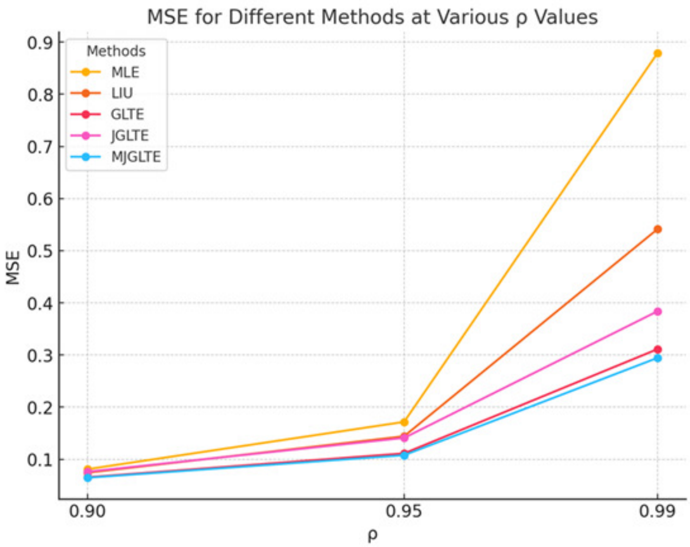

Figure 1. comparison between the methods when p=4.

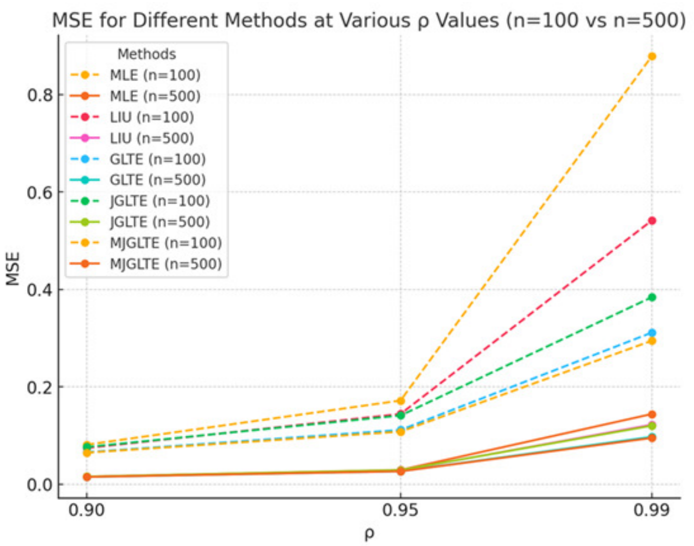

Figure 2. comparison between different sample space when n=100 and n=500 and p=4.

7.2. *Simulation Results*

The simulation data in Tables 1–3 demonstrate that MJGLTE estimator achieves the lowest MSE among the presented estimators. MJGLTE estimator demonstrates lower MSE than all other presented methods throughout its applications. The MLE produces the highest amount of MSE among all evaluated estimators.

The MSE for all the estimators grows when p and q increase while the estimators benefit from larger n. The best performance emerges from gamma Liu-type estimator while gamma Liu estimator and gamma ridge estimator and MLE estimator rank successively behind it.

The study results in Tables 1 to 3 show that gamma Liu-type estimator achieves superior MSE performance compared with other considered estimators. The simulation outcomes accorded with the mathematical proofs from the theoretical section.

8. A real application: dataset of hydrocarbon escape

In this section, the performance of the suggested estimator is assessed on a real application. For this purpose, we examine a hydrocarbon dataset, obtained from [22]. During the pumping of petrol, hydrocarbons get released into the atmosphere once they are in the tanks. To prevent the emission causing pollution in the atmosphere various equipments are fitted for the absorption of vapours. To assessing the efficacy of the strategy, 32 laboratory experiments were performed without the devices. There were four explanatory variables (x_1= hydrocarbon escaping (in ^{o}F), x_2=temperature (in ^{o}F), x_3=the initial pressure (in pounds per square inch) and x_4=the petrol pumped (in pounds per square inch)) that are involved in this laboratory experiment and the response variable (y=the quantity (in grams)).

The test of sufficiency of the gamma distribution on the response variable (y=the quantity (in grams)) was done with the help of the Kolmogorov Smirnov goodness-of-fit test. The value of test statistic D was about 0.101, whereas p-value was about 0.864, and it was greater than accepted level of significance (0.05). In response, to the next null hypothesis:

H_0 : The response variable follows Gamma dist.

H_1 : The response variable doesn't follow Gamma dist.

From above, the data follow a gamma distribution cannot be rejected, supporting the validity of the assumption of using the gamma regression model in this research.

Table 4. Kolmogorov–Smirnov test of the gamma distribution on the response variable

Test (D)	**Statistic**	**p-value**	**Decision**
0.101		0.864	Fail to reject the null hypothesis (Fits Gamma)

The issue of multicollinearity between the explanatory variables present in real data was diagnosed by computing the variance inflation factors (VIF). The findings were as follows: the VIF values of the variables x1, x3 and x4 were large (which is larger than 10) and indicate the multicollinearity issue. The application of the suggested estimator, MJGLTE is hence applicable in rectifying this issue and the approach is useful in enhancing the precision of the statistical estimation of the model. The value of the condition number for the design matrix containing the explanatory variables (x1 to x4), the CN=$\sqrt{\lambda_{max}/\lambda_{min}}$ is 660.51. Since this value significantly exceeds the acceptable limit of 30, it is a strong indication multicollinearity problem.

Table 6 compares three well-known statistical estimators, MLE, LIU and GLTE, alongside the proposed estimators JGLTE and MJGLTE, focusing on the mean squared error and the estimated regression coefficient values. It is noted that the proposed method MJGLTE, outperformed the other four methods due to its lower mean squared error, making it more accurate than the other estimators. Meanwhile, the MLE method performed the

Table 5. Variance Inflation Factors (VIF) for the explanatory variables

variables	*X1*	*X2*	*X3*	*X4*
VIF	*12.99*	*4.72*	*71.30*	*61.93*

Table 6. The study provides estimates for parameters and relative efficiency rates between various estimation methods

Estimators	*MLE*	*LIU*	*GLTE*	*JGLTE*	*MJGLTE*
MSE	*0.331602*	*0.267362*	*0.192884*	*0.207127*	***0.186817***
$\widehat{\beta}_1$	*-0.0176951*	*-0.017695*	*-0.01769*	*-0.017695*	*-0.0176950*
$\widehat{\beta}_2$	*0.0498453*	*0.0498328*	*0.049814*	*0.049845*	*0.0498142*
$\widehat{\beta}_3$	*0.7581212*	*0.7205976*	*0.6646805*	*0.746604*	*0.65458316*
$\widehat{\beta}_4$	*-0.3883334*	*-0.241706*	*-0.023204*	*-0.0450221*	*-0.0026902*

worst, indicating that it is the least accurate in the presence of multicollinearity. It was also found that the b1 and b2 estimators were very close across all estimators, as they were least affected by multicollinearity, and their effect remained constant regardless of the estimation method used. However, we observed significant variations in both the b3 and b4 values, with b3 yielding the highest value (0.75812) when using MLE, while the lowest value (0.6488) was obtained when using MJGLTE. This indicates that b3 is very sensitive to the estimator's method, indicating a high correlation between the estimators. From the b4 value, it was noted that the MLE method yielded a positive value (0.3883), while LIU yielded a negative value (0.2604). The GLTE, JGLTE, and MJGLTE methods reduced the coefficient to near zero, indicating the unimportance of this variable in the model and that the MLE method overestimates the coefficients when multicollinearity is present. The high correlation between the independent variables makes their estimates unstable when using the MLE method and has a high MSE compared to other modified methods that gave the lowest MSE at the expense of adding a small bias to stabilize these variables. Both MJGLTE and GLTE achieved the best performance because they combine Liu-type and Ridge techniques to balance bias and variance.

Table 7. Performance Analysis Based on MSE

Estimators	*MSE*	*Rank*	*Improvement Over MLE*
MLE	*0.331602*	*5*	*-*
LIU	*0.267362*	*4*	*19.3726%*
GLTE	*0.192884*	*2*	*41.8326%*
JGLTE	*0.207127*	*3*	*37.5374%*
MJGLTE	*0.186817*	*1*	***43.98%***

Table 7 also shows that the GLTE, JGLTE, and MJGLTE methods achieved a significant 41%, 37% and 43% respectively improvement in accuracy compared to the MLE method, while the LIU estimator achieved only a 19% improvement, making it the least efficient of the other modified estimators.

9. Conclusion

In this paper, we address the problem of multicollinearity in the Gamma regression model by combining the Liu-type and Jackknife estimators and propose a new estimator. Theoretically, we observed that the estimator MJGLTE outperform the other estimators. In addition, simulation and real-world application were conducted to examine the performance of the estimators. According to the simulation results, it was concluded that the proposed estimators MJGLTE have better performance than MLE, LIU, GLTE and JGLTE. Moreover, in the application of real data, it can be observed that the proposed estimator outperforms and by 43.98% for the MLE. Therefore, we recommend using the estimator MJGLTE in case of high or severe multicollinearity problem among the explanatory variables in the Gamma regression model.

REFERENCES

1. F. Akdeniz and E. A. Duran, *Liu-type estimator in semiparametric regression models*, Journal of Statistical Computation and Simulation, vol. 80, no. 8, pp. 853–871, 2010.
2. Z. Y. Algamal, *Developing a ridge estimator for the gamma regression model*, Journal of Chemometrics, vol. 32, no. 10, pp. e3054, 2018.
3. A. M. Algboory and Z. Y. Algamal, *Liu-Type estimator in gamma regression model based on (r-(kd)) class estimator*, International Journal of Nonlinear Analysis and Applications, vol. 13, no. 1, pp. 2455–2465, 2022.
4. M. I. Alheety and B. Golam Kibria, *Modified Liu-type estimator based on (rk) class estimator*, Communications in Statistics-Theory and Methods, vol. 42, no. 2, pp. 304–319, 2013.
5. A. Alkhateeb and Z. Algamal, *Jackknifed Liu-type estimator in Poisson regression model*, Journal of the Iranian Statistical Society, vol. 19, no. 1, pp. 21–37, 2022.
6. A. N. Alkhateeb and Q. N. N. Al-Qazaz, *Variable Selection in Weibull Accelerated Survival Model Based on Chaotic Sand Cat Swarm Algorithm*, IEStatistics, Optimization & Information Computing, vol. 13, no. 5, pp. 2105-2118, 2025.
7. A. N. Alkhateeb and Z. Y. Algamal, *Variable selection in gamma regression model using chaotic firefly algorithm with application in chemometrics*, Electronic Journal of Applied Statistical Analysis, vol. 14, no. 1, pp. 266–276, 2021.
8. M. Amin, M. Qasim, M. Amanullah and S. Afzal, *Performance of some ridge estimators for the gamma regression model*, Statistical papers, vol. 61, pp. 997–1026, 2020.
9. D. V. Hinkley, *Jackknifing in unbalanced situations*, Technometrics, vol. 19, no. 3, pp. 285–292, 1977.
10. A. E. Hoerl and R. W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems*, Technometrics, vol. 12, no. 1, pp. 55–67, 1970.
11. L. Kejian, *A new class of blased estimate in linear regression*, Communications in Statistics-Theory and Methods, vol. 22, no. 2, pp. 393–402, 1993.
12. K. Krishnamoorthy, *Handbook of statistical distributions with applications*, Chapman and Hall/CRC, 2006.
13. F. Kurtoğlu and M. R. Özkale, *Liu estimation in generalized linear models: application on gamma distributed response variable*, Statistical Papers, vol. 57, pp. 911–928, 2016.
14. K. Liu, *Using Liu-type estimator to combat collinearity*, Communications in Statistics-Theory and Methods, vol. 32, no. 5, pp. 1009–1020, 2003.
15. F. S. Mahmood Batah, S. D. Gore and M. R. Verma, *Effect of jackknifing on various ridge type estimators*, Model Assisted Statistics and Applications, vol. 3, no. 3, pp. 201–210, 2008.
16. K. Månsson, *Developing a Liu estimator for the negative binomial regression model: method and application*, Journal of Statistical Computation and Simulation , vol. 83, no. 9, pp. 1773–1780, 2013.
17. K. Månsson, B. G. Kibria and G. Shukur, *On Liu estimators for the logit regression model*, Economic Modelling, vol. 29, no. 4, pp. 1483–1488, 2012.
18. K. Månsson, B. G. Kibria and G. Shukur, *Performance of some weighted Liu estimators for logit regression model: an application to Swedish accident data*, Communications in Statistics-Theory and Methods, vol. 44, no. 2, pp. 363–375, 2015.
19. K. Månsson and G. Shukur, *A Poisson ridge regression estimator*, Economic Modelling, vol. 28, no. 4, pp. 1475–1481, 2011.
20. A. A. Mohammed and F. S. M. Batah , *Modified Jackknife Estimator in Linear Regression Model*, Journal of Survey in Fisheries Sciences , vol. 10, no. 3s, pp. 4570-4581, 2023.
21. H. Nyquist, *Applications of the jackknife procedure in ridge regression*, Computational Statistics & Data Analysis, vol. 6, no. 2, pp. 177–183, 1988.
22. M. Qasim, M. N. Akram, M. Amin and K. Månsson , *A restricted gamma ridge regression estimator combining the gamma ridge regression and the restricted maximum likelihood methods of estimation*, Journal of Statistical Computation and Simulation, vol. 92, no. 8, pp. 1696–1713, 2022.
23. M. H. Quenouille, *Notes on bias in estimation*, Biometrika, vol. 43, no. 3/4, pp. 353–360, 1956.
24. B. Singh, Y. Chaubey and T. Dwivedi, *An almost unbiased ridge estimator*, Sankhyā: The Indian Journal of Statisticsm Series B , pp. 342–346, 1986.
25. S. Türkan and G. Özel, *A new modified Jackknifed estimator for the Poisson regression model*, Journal of Applied Statistics, vol. 43, no. 10, pp. 1892–1905, 2016.
26. J. Wu and Y. Asar, *More on the restricted Liu estimator in the logistic regression model*, Communications in Statistics-Simulation and Computation, vol. 46, no. 5, pp. 3680–3689, 2017.

27. H. Yang and X. Chang, *A new two-parameter estimator in linear regression*, Communications in Statistics—Theory and Methods, vol. 39, no. 6, pp. 923–934, 2010.
28. Algamal, Zakariya Yahya and Asar, Yasin, *Liu-type estimator for the gamma regression model*, Communications in Statistics-Simulation and Computation, vol. 49, no. 8, pp. 2035–2048, 2020.
29. Abonazel, Mohamed R and Algamal, Zakariya Yahya and Awwad, Fuad A and Taha, Ibrahim M, *A new two-parameter estimator for beta regression model: method, simulation, and application*, Frontiers in Applied Mathematics and Statistics, vol. 7, pp. 780322, 2022.
30. Lukman, Adewale F and Dawoud, Issam and Kibria, BM Golam and Algamal, Zakariya Y and Aladeitan, Benedicta, *A new ridge-type estimator for the gamma regression model*, Scientifica, vol. 2021,no. 1, pp. 5545356, 2021.