

A new Liu-type estimator in inverse Gaussian regression model

Muhammed Jumaa Azzawi Al-Sahan

Department of Mathematics, College of Education for pure Sciences, Tikrit University, Tikrit, Iraq.

Email: muhammed.g.azzawi@st.tu.edu.iq

Zakariya Yahya Algamal *

Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

E-mail: zakariya.algamal@uomosul.edu.iq

ORCID: 0000-0002-0229-7958

Qasim Nasir Husain

Department of Mathematics, College of Education for pure Sciences, Tikrit University, Tikrit, Iraq.

Email: qasim11@tu.edu.iq

*Corresponding Author

Zakariya Yahya Algamal

E-mail: zakariya.algamal@uomosul.edu.iq

Telephone number: +964 7701640834

Article Info

Page Number: 291-302

Publication Issue:

Vol 71 No. 3s2 (2022)

Abstract

The ridge regression model has been shown to be an effective shrinking strategy for reducing the impacts of multicollinearity on a number of occasions. When the response variable is positively skewed, the inverse Gaussian regression model (IGR) is a popular model to use. Multicollinearity, on the other hand, is known to reduce the variance of the maximum likelihood estimator of inverse Gaussian regression coefficients. A inverse Gaussian ridge regression model (IGRR) has been presented as a solution to this problem. A novel estimator is proposed in this paper by presenting a generalization of the Liu-type estimator using IGR. In terms of absolute bias and mean squared error, our Monte Carlo simulation findings and real-world application demonstrate that the suggested estimator can provide significant improvements over other competing estimators.

Keywords: Multicollinearity; ridge estimator; inverse Gaussian regression model; Liu-type estimator; Monte Carlo simulation.

Article History

Article Received: 28 April 2022

Revised: 15 May 2022

Accepted: 20 June 2022

Publication: 21 July 2022

1. Introduction

The inverse Gaussian regression (IGR) has been widely used in industrial engineering, life testing, reliability, marketing, and social sciences (Bhattacharyya & Fries, 1982; Ducharme, 2001; Folks & Davis, 1981; Fries & Bhattacharyya, 1986; Heinzl & Mittlböck, 2002; Lemeshko, Lemeshko, Akushkina, Nikulin, & Saaidia, 2010; Malehi, Pourmotahari, & Angali, 2015). Specifically, IGR model is used when the response variable under the study is positively skewed (Babu & Chaubey, 1996; Chaubey, 2002; Wu & Li, 2011). When the response variable is extremely skewness, the IGR is preferable than gamma regression model (De Jong & Heller, 2008).

In dealing with the inverse Gaussian regression model, it is assumed that there is no correlation among the regressors. In practice, however, this assumption often not holds, which leads to the problem of multicollinearity. In the presence of multicollinearity, when estimating the regression

coefficients for inverse Gaussian regression model using the maximum likelihood (ML) method, the estimated coefficients are usually become unstable with a high variance, and therefore low statistical significance (Asar & Genç, 2015; Kurtoğlu & Özkale, 2016). Numerous remedial methods have been proposed to overcome the problem of multicollinearity (Al-Taweel, Algamal, & Sciences, 2020; Al-Thanoon, Qasim, & Algamal, 2020; Z. Y. Algamal, 2011; Zakariya Y Algamal, 2012; Z. Y. Algamal, 2018a, 2018b; Zakariya Yahya Algamal & Abonazel, 2021; Zakariya Yahya Algamal & Alanaz, 2018; Z. Y. J. E. J. o. A. S. A. Algamal, 2018; A. Alkhateeb & Algamal, 2020; A. N. Alkhateeb & Algamal, 2021; Alobaidi, Shamany, & Algamal, 2021; Hamad, Algamal, & Applications, 2021; Adewale F. Lukman, Algamal, Kibria, & Ayinde, 2021; A. F. Lukman, Dawoud, Kibria, Algamal, & Aladeitan, 2021; Mahmood & Algamal, 2021; Mohammed & Algamal, 2021; Noel & Algamal, 2021; Rashad & Algamal, 2019; Shamany, Alobaidi, & Algamal, 2019; Yahya Algamal, 2018). The ridge regression method (Hoerl & Kennard, 1970) has been consistently demonstrated to be an attractive and alternative to the ML estimation method.

In classical linear regression models the following relationship is usually adopted

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where \mathbf{y} is an $n \times 1$ vector of observations of the response variable, $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_p)$ is an $n \times p$ known design matrix of explanatory variables, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ is a $p \times 1$ vector of unknown regression coefficients, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors with mean 0 and variance σ^2 .

Ridge regression is a shrinkage method that shrinks all regression coefficients toward zero to reduce the large variance (Asar & Genç, 2015; Batah, Ramanathan, & Gore, 2008). This done by adding a positive amount to the diagonal of $\mathbf{X}^T \mathbf{X}$. As a result, the ridge estimator is biased, but it guaranties a smaller mean squared error than the ML estimator.

In linear regression, the ridge estimator is defined as

$$\hat{\boldsymbol{\beta}}_{Ridge} = (\mathbf{X}^T \mathbf{X} + k \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}, \quad (2)$$

where \mathbf{I} is the identity matrix with dimension $p \times p$ and $k \geq 0$ represents the ridge parameter (shrinkage parameter). The ridge parameter, k , controls the shrinkage of $\boldsymbol{\beta}$ toward zero. For larger value of k , the $\hat{\boldsymbol{\beta}}_{Ridge}$ estimator yields greater shrinkage approaching zero (Hoerl & Kennard, 1970).

2. Statistical methodology

2.1. Inverse Gaussian ridge regression model

Positively skewed data often arise in epidemiology, social, and economic studies. This type of data consists of nonnegative values. Inverse Gaussian distribution is a well-known distribution that fits to such type of data. Inverse Gaussian regression model (IGR) is used to model the relationship between the positively skewed response variable and potentially regressors (Uusipaikka, 2009).

The inverse Gaussian distribution is a continuous distribution with two positive parameters: location parameter, μ , and scale parameter, τ , denoted as $IG(\mu, \tau)$. Its probability density function is defined as

$$f(y, \mu, \tau) = \frac{1}{\sqrt{2\pi y^3 \tau}} \exp \left[-\frac{1}{2y} \left(\frac{y - \mu}{\mu \sqrt{\tau}} \right)^2 \right], \quad y > 0. \quad (3)$$

The mean and variance of this distribution are, respectively, $E(y) = \mu$ and $\text{var}(y) = \tau \mu^3$.

Inverse Gaussian regression model is considered a member of the generalized linear models (GLM) family, extending the ideas of linear regression to the situation where the response variable is following the inverse Gaussian distribution. Following the GLM methodology, Eq. (1) can re-write in terms of exponential family function as

$$f(y, \mu, \tau) = \frac{1}{\tau} \left\{ -\frac{y}{2\mu^2} + \frac{1}{\mu} \right\} + \left\{ -\frac{1}{2} \ln(2\pi y^3) - \frac{1}{2} \ln(\tau) \right\}, \quad (4)$$

where $C(y, \tau) = -(1/2) \ln(2\pi y^3) - (1/2) \ln(\tau)$ and $\frac{y\theta - a(\theta)}{\phi} = \frac{1}{\tau} \left\{ -\frac{y}{2\mu^2} + \frac{1}{\mu} \right\}$. Here, τ represents the dispersion parameter and $1/\mu^2$ represents the canonical link function.

In GLM, a monotonic and differentiable link function connects the mean of the response variable with the linear predictor $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, where \mathbf{x}_i is the i^{th} row of \mathbf{X} and $\boldsymbol{\beta}$ is a $(p+1) \times 1$ vector of unknown regression coefficients. Because η_i depends on $\boldsymbol{\beta}$ and the mean of the response variable is a function of η_i , then $E(y_i) = \mu_i = g^{-1}(\eta_i) = g^{-1}(\mathbf{x}_i^T \boldsymbol{\beta})$. Related to the IGR, the $\mu = 1/\sqrt{\mathbf{x}_i^T \boldsymbol{\beta}}$. Another possible link function for the IGR is log link function, $\mu = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$.

The model estimation of the IGR is based on the maximum likelihood method (ML). The log likelihood function of the IGR under the canonical link function is defined as

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^n \left\{ \frac{1}{\tau} \left[\frac{y_i \mathbf{x}_i^T \boldsymbol{\beta}}{2} - \sqrt{\mathbf{x}_i^T \boldsymbol{\beta}} \right] - \frac{1}{2\tau y_i} - \frac{\ln \tau}{2} - \ln(2\pi y_i^3) \right\}. \quad (5)$$

The ML estimator is then obtained by computing the first derivative of the Eq. (3) and setting it equal to zero, as

$$\frac{\partial \ell(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \frac{1}{2\tau} \left[y_i - \frac{1}{\sqrt{\mathbf{x}_i^T \boldsymbol{\beta}}} \right] \mathbf{x}_i = 0. \quad (6)$$

Unfortunately, the first derivative cannot be solved analytically because Eq. (4) is nonlinear in β . The iteratively weighted least squares (IWLS) algorithm or Fisher-scoring algorithm can be used to obtain the ML estimators of the IGR parameters. In each iteration, the parameters are updated by

$$\beta^{(r+1)} = \beta^{(r)} + I^{-1}(\beta^{(r)})S(\beta^{(r)}), \quad (7)$$

where $S(\beta^{(r)})$ and $I^{-1}(\beta^{(r)})$ are $S(\beta) = \partial \ell(\beta) / \partial \beta$ and $I^{-1}(\beta) = \left(-E \left(\partial^2 \ell(\beta) / \partial \beta \partial \beta^T \right) \right)^{-1}$ evaluated at $\beta^{(r)}$, respectively. The final step of the estimated coefficients is defined as

$$\hat{\beta}_{IGR} = (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{W}} \hat{\mathbf{u}}, \quad (8)$$

where $\hat{\mathbf{W}} = \text{diag}(\hat{\mu}_i^3)$, $\hat{\mathbf{u}}$ is a vector where i^{th} element equals to $\hat{u}_i = (1 / \hat{\mu}_i^2) + ((y_i - \hat{\mu}_i) / \hat{\mu}_i^3)$, and $\hat{\mu} = 1 / \sqrt{\mathbf{x}_i^T \hat{\beta}}$. The covariance matrix of $\hat{\beta}_{IGR}$ equals

$$\text{cov}(\hat{\beta}_{IGR}) = \left[-E \left(\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^T} \right) \right]^{-1} = \tau (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1}, \quad (9)$$

and the mean squared error (MSE) equals

$$\begin{aligned} \text{MSE}(\hat{\beta}_{IGR}) &= E(\hat{\beta}_{IGR} - \beta)^T (\hat{\beta}_{IGR} - \beta) \\ &= \tau \text{tr}[(\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X})^{-1}] \\ &= \tau \sum_{j=1}^p \frac{1}{\lambda_j}, \end{aligned} \quad (10)$$

where λ_j is the eigenvalue of the $\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}$ matrix and the dispersion parameter, τ , is estimated by (Uusipaikka, 2009)

$$\hat{\tau} = \frac{1}{(n-p)} \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^3}. \quad (11)$$

In the presence of multicollinearity, the matrix $\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}$ becomes ill-conditioned leading to high variance and instability of the ML estimator of the inverse Gaussian regression parameters. As a remedy, the inverse Gaussian ridge regression estimator (IGRR) can be defined as

$$\begin{aligned} \hat{\beta}_{IGRR} &= (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^T \hat{\mathbf{W}} \mathbf{X} \hat{\beta}_{IGR} \\ &= (\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X} + k\mathbf{I})^{-1} \mathbf{X}^T \hat{\mathbf{W}} \hat{\mathbf{u}}, \end{aligned} \quad (12)$$

where $k \geq 0$. The ML estimator can be considered as a special estimator from Eq. (11) with $k = 0$.

2.2. The proposed estimator

K. Liu (1993) proposed an estimator, which is called Liu estimator, combining the Stein estimator with the ridge estimator. Comparing with ridge estimator, the Liu estimator is a linear function of the shrinkage parameter, therefore it is easy to choose the shrinkage parameter than to choose ridge parameter.

Consequently, the Liu estimator in IGRM, the inverse Gaussian Liu estimator (IGLE), is defined as

$$\hat{\beta}_{IGLE} = (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + \mathbf{I})^{-1} (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + d \mathbf{I}) \hat{\beta}_{IGR} \quad (13)$$

where d is a shrinkage parameter, $0 < d < 1$. For $d = 1$, $\hat{\beta}_{IGLE} = \hat{\beta}_{IGR}$ and for $0 < d < 1$, $\|\hat{\beta}_{IGLE}\| < \|\hat{\beta}_{IGR}\|$.

Liu estimator is upgraded by proposing Liu-type estimator to overcome the problem of severe multicollinearity, Liu-type estimator is defined as follows (Zakariya Yahya Algamal & Asar, 2020)

$$\hat{\beta}_{IGLTE} = (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + k \mathbf{I})^{-1} (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} - d \mathbf{I}) \hat{\beta}_{IGR} \quad (14)$$

where $-\infty < d < \infty$ and $k \geq 0$. Liu-type estimator has superior over ridge estimator (Liu, 2003). The MSE of $\hat{\beta}_{IGLTE}$ is

$$MSE(\hat{\beta}_{IGLTE}) = \nu^{-1} \sum_{j=1}^J \frac{(\lambda_j - d)^2}{\lambda_j (\lambda_j + k)^2} + (d + k)^2 \sum_{j=1}^J \frac{\alpha_j^2}{(\lambda_j + k)^2} \quad (15)$$

In the context of the linear regression model, Kurnaz and Akay (2014), proposed the a new generalized Liu-type estimator to alleviate the problem of multicollinearity in linear regression model. The theoretical of the generalized Liu-type estimator have been studied by Kurnaz and Akay (2014) and Ertan and Akay (2020).

According to Kurnaz and Akay (2014) and Ertan and Akay (2020), our proposed new Liu-type estimator for inverse Gaussian regression model (NGLTE) is defined as:

$$\hat{\beta}_{NGLTE} = (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + k \mathbf{I})^{-1} (\mathbf{X}^T \hat{\mathbf{W}}\mathbf{X} + f(k) \mathbf{I}) \hat{\beta}_{IGR} \quad (16)$$

where $f(k)$ is a continuous function of the k . Usually, $f(k)$ is selected as a linear function of the biasing parameter such as $f(k) = ak + b$, where $a, b \in R$. As a result, the NGLTE becomes a general estimator which includes the other biased estimators (Ertan & Akay, 2020; Kurnaz & Akay, 2014).

The MSE of $\hat{\beta}_{NGLTE}$ is

$$MSE(\hat{\beta}_{NGLTE}) = \nu^{-1} \sum_{j=1}^{p+1} \frac{(\lambda_j + f(k))^2}{\lambda_j (\lambda_j + k)^2} + \sum_{j=1}^{p+1} \frac{(f(k) - k)^2 \alpha_j^2}{(\lambda_j + k)^2} \quad (17)$$

3. Simulation study

In this section, a Monte Carlo simulation experiment is used to examine the performance of our proposed estimator under different degrees of multicollinearity.

3.1. Simulation design

The response variable of n observations from inverse Gaussian regression model is generated as $y_i \sim IG(\theta_i, \nu)$, where $\nu \in \{0.50, 1.5\}$ and $\theta_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ with $\sum_{j=1}^p \beta_j^2 = 1$ and $\beta_1 = \beta_2 = \dots = \beta_p$ (Kibria, 2003). The explanatory variables $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{in})$ have been generated from the following formula

$$x_{ij} = (1 - \rho^2)^{1/2} w_{ij} + \rho w_{ip}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, p, \quad (18)$$

where ρ represents the correlation between the explanatory variables and w_{ij} 's are independent standard normal pseudo-random numbers. Because the sample size has direct impact on the prediction accuracy, three representative values of the sample size are considered: 50, 100, and 150. In addition, the number of the explanatory variables is considered as $p = 4$ and $p = 8$ because increasing the number of explanatory variables can lead to increase the MSE. Further, because we are interested in the effect of multicollinearity, in which the degrees of correlation considered more important, three values of the pairwise correlation are considered with $\rho = \{0.90, 0.95, 0.99\}$. The optimum value of k can be obtained by using Hoerl, Kannard, and Baldwin (1975) formula as

$$\hat{k} = \frac{\hat{\psi} p}{\hat{\alpha}^T \hat{\alpha}}, \quad (19)$$

where $\hat{\psi}$ is the estimated dispersion parameter which is calculated by Eq. (9). For a combination of these different values of n, ν, p , and ρ the generated data is repeated 1000 times and the average absolute bias and average MSE are determined as

$$\text{MSE}(\hat{\boldsymbol{\beta}}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}). \quad (20)$$

3.2. Simulation results

The averaged MSE all the combination of n, ν, p , and ρ , are respectively summarized in Tables 1 and 2. The best value of the averaged bias and MSE is highlighted in bold. As Table 1 shows, the proposed method, NIGLTE, gives low bias comparing with IGRR and IGLTE. On other hand, NIGLTE performances better than GRR. It is noted from Table 2 that NIGLTE ranks first with respect to MSE. In the second rank, IGLTE estimator performs better than both IGRR and IGR estimators. Additionally, IGR estimator has the worst performance among IGRR, IGLTE, and NIGLTE which is significantly impacted by the multicollinearity.

Furthermore, with respect to ρ , there is increasing in the MSE values when the correlation degree increases regardless the value of n, ν and p . Regarding the number of explanatory variables, it is easily seen that there is a negative impact on MSE, where there are increasing in their values when the p increasing from four variables to eight variables. In Addition, in terms of the sample size n , the MSE values decrease when n increases, regardless the value of ρ, ν and p . Clearly, in terms of the dispersion parameter ν , MSE values are decreasing when ν increasing.

Table 1: Averaged MSE values for the four estimators when $\nu = 0.5$

n	p	ρ	IGR	IGRR	IGLTE	NIGLTE
50	4	0.90	4.785	4.544	4.205	4.091
		0.95	4.829	4.594	4.255	4.141
		0.99	5.095	4.86	4.521	4.407
	8	0.90	4.899	4.664	4.325	4.211
		0.95	4.949	4.714	4.375	4.261
		0.99	5.215	4.98	4.641	4.527
100	4	0.90	4.537	4.302	3.963	3.849
		0.95	4.587	4.352	4.013	3.899
		0.99	4.853	4.618	4.279	4.165
	8	0.90	4.663	4.422	4.083	3.969
		0.95	4.707	4.472	4.133	4.019
		0.99	4.973	4.738	4.399	4.285
150	4	0.90	4.486	4.251	3.912	3.798
		0.95	4.536	4.301	3.962	3.849
		0.99	4.802	4.567	4.228	4.114
	8	0.90	4.606	4.371	4.032	3.919
		0.95	4.656	4.421	4.082	3.968
		0.99	4.922	4.687	4.348	4.234

Table 2: Averaged MSE values for the four estimators when $\nu = 1.5$

n	p	ρ	IGR	IGRR	IGLTE	NIGLTE
50	4	0.90	4.676	4.441	4.102	3.988
		0.95	4.725	4.49	4.151	4.037
		0.99	4.992	4.757	4.418	4.304
	8	0.90	4.796	4.561	4.222	4.108
		0.95	4.845	4.61	4.271	4.156
		0.99	5.112	4.877	4.538	4.424
100	4	0.90	4.434	4.199	3.86	3.746
		0.95	4.484	4.248	3.909	3.795
		0.99	4.75	4.515	4.176	4.062
	8	0.90	4.554	4.319	3.98	3.866
		0.95	4.604	4.369	4.03	3.916
		0.99	4.87	4.635	4.296	4.182
150	4	0.90	4.383	4.148	3.809	3.695
		0.95	4.432	4.197	3.859	3.744
		0.99	4.699	4.464	4.125	4.011
	8	0.90	4.503	4.268	3.929	3.815
		0.95	4.552	4.317	3.979	3.864
		0.99	4.819	4.584	4.245	4.131

4. Real Data Application

To demonstrate the usefulness of the NIGLTE in real application, we present here a chemistry dataset with $(n, p) = (65, 15)$, where n represents the number of imidazo[4,5-b]pyridine derivatives, which are used as anticancer compounds. While p denotes the number of molecular descriptors, which are treated as explanatory variables (Zakariya Yahya Algamal, Lee, Al-Fakih, & Aziz, 2015). The response of interest is the biological activities (IC_{50}). Quantitative structure-activity relationship (QSAR) study has become a great deal of importance in chemometrics. The principle of QSAR is to model several biological activities over a collection of chemical compounds in terms of their structural properties (Zakariya Yahya Algamal & Lee, 2017). Consequently, using of regression model is one of the most important tools for constructing the QSAR model. A description of the used explanatory variables is provided in Table 7. All the variables are numerical. First, to check whether the response variable belongs to the inverse Gaussian distribution, a Chi-square test is used. The result of the test equals to 5.2762 with p-value equals to 0.2601. It is indicated from this result that the inverse Gaussian distribution fits very well to this response variable. That is, the following model is set

$$\hat{y}_{IC_{50}} = \exp\left(\sum_{j=1}^{15} \mathbf{x}_j \hat{\beta}_j\right). \quad (21)$$

Second, to check whether there is a relationship among the explanatory variables or not, Figure 1 displays the correlation matrix among the 15 explanatory variables. It is obviously seen that there are correlations greater than 0.90 among MW, SpMaxA_D, and ATS8v ($r=0.96$), between SpMax3_Bh(s) and ATS8v ($r=0.92$), and between Mor21v with Mor21e ($r=0.93$).

Third, to test the existence of multicollinearity after fitting the inverse Gaussian regression model using log link function and the estimated dispersion parameter is 0.00103, the eigenvalues of the matrix $\mathbf{X}^T \hat{\mathbf{W}} \mathbf{X}$ are obtained as 1.884×10^9 , 3.445×10^6 , 2.163×10^5 , 2.388×10^4 , 1.290×10^3 , 9.120×10^2 , 4.431×10^2 , 1.839×10^2 , 1.056×10^2 , 5525, 3231, 2631, 1654, 1008, and 1.115. The determined condition number $CN = \sqrt{\lambda_{\max} / \lambda_{\min}}$ of the data is 40383.035 indicating that the severe multicollinearity issue is exist.

The estimated inverse Gaussian regression coefficients and MSE values for the IGR, IGRR, IGLTE, and NIGLTE estimators are listed in Table 4. According to Table 4, it is clearly seen that the NIGLTE shrinkages the value of the estimated coefficients efficiently. Additionally, in terms of the MSE, there is an important reduction in favor of the NIGLTE. Specifically, it can be seen that the MSE of the NIGLTE estimator was about 45.11%, 36.71%, and 19.84% lower than that of IGR, IGRR, and IGLTE estimators, respectively.

Table 3: Description of the used explanatory variables

Variable name's	description
MW	molecular weight
IC3	Information Content index (neighborhood symmetry of 3-order)
SpMaxA_D	normalized leading eigenvalue from topological distance matrix
ATS8v	Broto-Moreau autocorrelation of lag 8 (log function) weighted by van der

	Waals volume
MATS7v	Moran autocorrelation of lag 7 weighted by van der Waals volume
MATS2s	Moran autocorrelation of lag 2 weighted by I-state
GATS4p	Geary autocorrelation of lag 4 weighted by polarizability
SpMax8_Bh(p)	largest eigenvalue n. 8 of Burden matrix weighted by polarizability
SpMax3_Bh(s)	largest eigenvalue n. 3 of Burden matrix weighted by I-state
P_VSA_e_3	P_VSA-like on Sanderson electronegativity, bin 3
TDB08m	3D Topological distance based descriptors - lag 8 weighted by mass
RDF100m	Radial Distribution Function - 100 / weighted by mass
Mor21v	signal 21 / weighted by van der Waals volume
Mor21e	signal 21 / weighted by Sanderson electronegativity
HATS6v	leverage-weighted autocorrelation of lag 6 / weighted by van der Waals volume"

Table 4: The estimated coefficients and MSE values for the four used estimators.

	Estimators			
	IGR	IGRR	IGLTE	NIGLTE
$\hat{\beta}_{MW}$	1.002	0.8839	0.5797	0.3397
$\hat{\beta}_{IC3}$	1.237	0.0816	0.0676	0.0244
$\hat{\beta}_{SpMaxA_D}$	-1.102	1.1159	0.8421	0.6021
$\hat{\beta}_{ATS8v}$	-1.379	0.0816	0.0676	0.0244
$\hat{\beta}_{MATS7v}$	-1.219	-0.5142	-0.6252	-0.8652
$\hat{\beta}_{MATS2s}$	-1.215	0.0816	0.0676	0.08084
$\hat{\beta}_{GATS4p}$	-1.237	-0.8834	-1.0011	-1.2411
$\hat{\beta}_{SpMax8_Bh(p)}$	2.506	0.0816	0.0676	0.11684
$\hat{\beta}_{SpMax3_Bh(s)}$	2.069	-0.7311	-0.8427	-1.0827
$\hat{\beta}_{P_VSA_e_3}$	2.001	0.0816	0.0676	0.0931
$\hat{\beta}_{TDB08m}$	-2.103	-0.7423	-0.8568	-1.0968
$\hat{\beta}_{RDF100m}$	1.571	0.0816	0.0676	0.0593
$\hat{\beta}_{Mor21v}$	-2.434	-0.7732	-0.8859	-1.1259
$\hat{\beta}_{Mor21e}$	-2.352	0.0816	0.0676	0.04284
$\hat{\beta}_{HATS6v}$	2.211	1.4145	1.101	0.8611
MSE	3.2951	2.7058	2.5482	1.9366

5. Conclusions

In this paper, a new estimator was developed by generalizing the Liu-type estimator with inverse Gaussian regression model. According to Monte Carlo simulation studies, it has been seen that some estimator can bring significant improvement relative to others, in terms of MSE. Moreover, in real data application, compared to IGR, IGRR, and IGLTE estimators, the developed estimator can efficiently reduce the MSE.

REFERENCES

1. Al-Taweel, Y., Algamal, Z. J. P. o. E., & Sciences, N. (2020). Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model. 8(1), 248-255.
2. Al-Thanoon, N. A., Qasim, O. S., & Algamal, Z. Y. (2020). Variable selection in Gamma regression model using binary gray Wolf optimization algorithm. *Journal of Physics: Conference Series*, 1591(1). doi:10.1088/1742-6596/1591/1/012036
3. Algamal, Z. Y. (2011). Paired Bootstrapping procedure in Gamma Regression Model using R. *Journal of Basrah Researches*, 37(4), 201-211.
4. Algamal, Z. Y. (2012). Diagnostic in Poisson regression models. *Electronic Journal of Applied Statistical Analysis*, 5(2), 178-186.
5. Algamal, Z. Y. (2018a). A new method for choosing the biasing parameter in ridge estimator for generalized linear model. *Chemometrics and Intelligent Laboratory Systems*, 183, 96-101. doi:10.1016/j.chemolab.2018.10.014
6. Algamal, Z. Y. (2018b). Shrinkage parameter selection via modified cross-validation approach for ridge regression model. *Communications in Statistics - Simulation and Computation*, 49(7), 1922-1930. doi:10.1080/03610918.2018.1508704
7. Algamal, Z. Y., & Abonazel, M. R. (2021). Developing a Liu-type estimator in beta regression model. *Concurrency and Computation: Practice and Experience*. doi:10.1002/cpe.6685
8. Algamal, Z. Y., & Alanaz, M. M. J. E. J. o. A. S. A. (2018). Proposed methods in estimating the ridge regression parameter in Poisson regression model. 11(2), 506-515.
9. Algamal, Z. Y., & Asar, Y. (2020). Liu-type estimator for the gamma regression model. *Communications in Statistics - Simulation and Computation*, 49(8), 2035-2048. doi:10.1080/03610918.2018.1510525
10. Algamal, Z. Y., & Lee, M. H. (2017). A novel molecular descriptor selection method in QSAR classification model based on weighted penalized logistic regression. *Journal of Chemometrics*, 31(10), e2915. doi:10.1002/cem.2915
11. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. (2015). High-dimensional QSAR prediction of anticancer potency of imidazo[4,5-b]pyridine derivatives using adjusted adaptive LASSO. *Journal of Chemometrics*, 29(10), 547-556. doi:10.1002/cem.2741
12. Algamal, Z. Y. J. E. J. o. A. S. A. (2018). Shrinkage estimators for gamma regression model. 11(1), 253-268.
13. Alkhateeb, A., & Algamal, Z. (2020). Jackknifed Liu-type Estimator in Poisson Regression Model. *Journal of the Iranian Statistical Society*, 19(1), 21-37. doi:10.29252/jirss.19.1.21

14. Alkhateeb, A. N., & Algamal, Z. Y. J. E. J. o. A. S. A. (2021). Variable selection in gamma regression model using chaotic firefly algorithm with application in chemometrics. 14(1), 266-276.
15. Alobaidi, N. N., Shamany, R. E., & Algamal, Z. Y. J. T. S. (2021). A New Ridge Estimator for the Negative Binomial Regression Model. 19(1), 116-125.
16. Asar, Y., & Genç, A. (2015). New shrinkage parameters for the Liu-type logistic estimators. *Communications in Statistics - Simulation and Computation*, 45(3), 1094-1103. doi:10.1080/03610918.2014.995815
17. Babu, G. J., & Chaubey, Y. P. (1996). Asymptotics and bootstrap for inverse Gaussian regression. *Annals of the Institute of Statistical Mathematics*, 48(1), 75-88. doi:10.1007/bf00049290
18. Batah, F. S. M., Ramanathan, T. V., & Gore, S. D. (2008). The efficiency of modified jackknife and ridge type regression estimators - A comparison. *Surveys in Mathematics and its Applications*, 3, 111 – 122.
19. Bhattacharyya, G. K., & Fries, A. (1982). Inverse Gaussian regression and accelerated life tests. *Lecture Notes-Monograph Series*, 2, 101-117.
20. Chaubey, Y. P. (2002). Estimation in inverse Gaussian regression: comparison of asymptotic and bootstrap distributions. *Journal of Statistical Planning and Inference*, 100(2), 135-143. doi:[https://doi.org/10.1016/S0378-3758\(01\)00128-8](https://doi.org/10.1016/S0378-3758(01)00128-8)
21. De Jong, P., & Heller, G. Z. (2008). *Generalized linear models for insurance data* (Vol. 10): Cambridge University Press Cambridge.
22. Ducharme, G. R. (2001). Goodness-of-fit tests for the inverse Gaussian and related distributions. *Test*, 10(2), 271-290. doi:10.1007/bf02595697
23. Ertan, E., & Akay, K. U. (2020). A new Liu-type estimator in binary logistic regression models. *Communications in Statistics - Theory and Methods*, 1-25. doi:10.1080/03610926.2020.1813777
24. Folks, J. L., & Davis, A. S. (1981). Regression models for the inverse Gaussian distribution. *Statistical Distributions in Scientific Work*, 4(1), 91-97.
25. Fries, A., & Bhattacharyya, G. K. (1986). Optimal design for an inverse Gaussian regression model. *Statistics & probability letters*, 4(6), 291-294.
26. Hamad, A. A., Algamal, Z. Y. J. I. J. o. N. A., & Applications. (2021). Jackknifing KL estimator in generalized linear models. 12(Special Issue), 2093-2104.
27. Heinzl, H., & Mittlböck, M. (2002). Adjusted R² Measures for the Inverse Gaussian Regression Model. *Computational Statistics*, 17(4), 525-544. doi:10.1007/s001800200125
28. Hoerl, A. E., Kannard, R. W., & Baldwin, K. F. (1975). Ridge regression: Some simulations. *Communications in Statistics-Theory and Methods*, 4(2), 105-123.
29. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
30. Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics - Simulation and Computation*, 32(2), 419-435. doi:10.1081/SAC-120017499
31. Kurnaz, F. S., & Akay, K. U. (2014). A new Liu-type estimator. *Statistical Papers*, 56(2), 495-517. doi:10.1007/s00362-014-0594-6

32. Kurtoğlu, F., & Özkale, M. R. (2016). Liu estimation in generalized linear models: application on gamma distributed response variable. *Statistical Papers*, 57(4), 911-928. doi:10.1007/s00362-016-0814-3
33. Lemeshko, B. Y., Lemeshko, S. B., Akushkina, K. A., Nikulin, M. S., & Saaidia, N. (2010). Inverse Gaussian model and its applications in reliability and survival analysis *Mathematical and Statistical Models and Methods in Reliability* (pp. 433-453): Springer.
34. Liu, K. (2003). Using Liu-type estimator to combat collinearity. *Communications in Statistics-Theory and Methods*, 32(5), 1009-1020.
35. Lukman, A. F., Algamal, Z. Y., Kibria, B. M. G., & Ayinde, K. (2021). The KL estimator for the inverse Gaussian regression model. *Concurrency and Computation: Practice and Experience*, 33(13). doi:10.1002/cpe.6222
36. Lukman, A. F., Dawoud, I., Kibria, B. M. G., Algamal, Z. Y., & Aladeitan, B. (2021). A New Ridge-Type Estimator for the Gamma Regression Model. *Scientifica (Cairo)*, 2021, 5545356. doi:10.1155/2021/5545356
37. Mahmood, S. W., & Algamal, Z. Y. J. T. S. (2021). Reliability Estimation of Three Parameters Gamma Distribution via Particle Swarm Optimization. 19(2), 308-316.
38. Malehi, A. S., Pourmotahari, F., & Angali, K. A. (2015). Statistical models for the analysis of skewed healthcare cost data: A simulation study. *Health Economics Review*, 5, 1-11. doi:10.1186/s13561-015-0045-7
39. Mohammed, H. S., & Algamal, Z. Y. (2021). Shrinkage estimators for semiparametric regression model. *Journal of Physics: Conference Series*, 1897(1). doi:10.1088/1742-6596/1897/1/012012
40. Noeel, F., & Algamal, Z. Y. (2021). Almost unbiased ridge estimator in the count data regression models. *Electronic Journal of Applied Statistical Analysis*, 14(1), 44-57.
41. Rashad, N. K., & Algamal, Z. Y. (2019). A New Ridge Estimator for the Poisson Regression Model. *Iranian Journal of Science and Technology, Transactions A: Science*, 43(6), 2921-2928. doi:10.1007/s40995-019-00769-3
42. Shamany, R., Alobaidi, N. N., & Algamal, Z. Y. J. E. J. o. A. S. A. (2019). A new two-parameter estimator for the inverse Gaussian regression model with application in chemometrics. 12(2), 453-464.
43. Uusipaikka, E. (2009). Confidence intervals in generalized regression models. NW: Chapman & Hall/CRC Press.
44. Wu, L., & Li, H. (2011). Variable selection for joint mean and dispersion models of the inverse Gaussian distribution. *Metrika*, 75(6), 795-808. doi:10.1007/s00184-011-0352-x
45. Yahya Algamal, Z. (2018). Performance of ridge estimator in inverse Gaussian regression model. *Communications in Statistics - Theory and Methods*, 48(15), 3836-3849. doi:10.1080/03610926.2018.1481977