Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
`http://siba-ese.unisalento.it/index.php/ejasa/index`
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v15n3p510

Almost Unbiased Ridge Estimator in the Inverse Gaussian Regression Model

By Al-Taweel, Algamal

Published: November 20, 2022

Electronic Journal of Applied Statistical Analysis
Vol. 15, Issue 03, November 2022, 510-526
DOI: 10.1285/i20705948v15n3p510

Almost Unbiased Ridge Estimator in the Inverse Gaussian Regression Model

Younus Hazim Al-Taweel*[a] and Zakariya Yahya Algamal †[b]

[a] *Department of Mathematics, College of Education for Pure Science, University of Mosul, Iraq*
[b] *Department of Statistics and Informatics, College of Computers Sciences and Mathematics, University of Mosul, Iraq*

Published: November 20, 2022

The inverse Gaussian regression (IGR) model is a very common model when the shape of the response variable is positively skewed. The traditional maximum likelihood estimator (MLE) is used to estimate the IGR model parameters. However, when multicollinearity exists among the explanatory variables, the MLE becomes not an efficient estimator as the mean squared error (MSE) becomes inflated. In order to remedy this problem, the ridge estimator (RE) is used. In this paper, we present an almost unbiased ridge estimator for the IGR model in order to overcome multicollinearity problem. We also investigate the performance of the almost unbiased ridge estimator using a Monte Carlo simulation. The results of the almost unbiased ridge estimator are compared with those of the MLE and of the RE in terms of the MSE measure. In addition, a real example of dataset is used and the results show that the performance of the suggested estimator is superior when the multicollinearity is presented among the explanatory variables in the IGR model.

keywords: Inverse Gaussian regression, multicollinearity, almost unbiased ridge estimator, Monte Carlo simulation.

1 Introduction

It is common in regression model applications to see that the explanatory variables are correlated. If the correlation among the explanatory variables is high, the maximum

*Corresponding authors: younus.altaweel@uomosul.edu.iq.
†Corresponding authors: zakariya.algamal@uomosul.edu.iq.

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

likelihood estimator (MLE) will not be efficient (Algamal, 2018; Månsson and Shukur, 2011).

The use of the inverse Gaussian regression (IGR) model is common in many science and technology fields like marketing, life testing and industrial engineering (Malehi et al., 2015; Akram et al., 2020; Lukman et al., 2021). The IGR model is specifically used when y, the response variable, is positively skewed (Babu and Chaubey, 1996). In this case, the IGR is preferable to the gamma regression model (De Jong and Heller, 2008).

The probability density function of the inverse Gaussian (IG) distribution is given by

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

where μ is the location parameter and τ is the scale parameter. The mean of the IG distribution is $E(y)=\mu$ and the variance is $Var(y)=\tau\mu^3$.

The IGR model is a family member of the generalized linear models (GLMs). Thus, equation (1) can be rewritten as

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

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

The mean of the response variable is connected by a link function with the linear predictor $\eta_i=\mathbf{x}_i^T\boldsymbol{\beta}$ where $\boldsymbol{\beta}$ is a $(p+1)\times 1$ vector of unknown regression coefficient parameters and $\mathbf{x}_i$ is the ith row of the design matrix $\mathbf{X}$. The mean of y is a function of η_i and so it depends on $\boldsymbol{\beta}$. Hence, we have $E(y_i)=\mu_i=g^{-1}(\eta_i)=g^{-1}(\mathbf{x}_i^T\boldsymbol{\beta})$ where $\mu=1/\sqrt{\mathbf{x}_i^T\boldsymbol{\beta}}$ is the inverse link function. The log link function is an alternative link function for the IGR model, $\mu=\exp(\mathbf{x}_i^T\boldsymbol{\beta})$.

The coefficient parameters of the IGR model are estimated using the maximum likelihood estimation method. For the inverse link function, the log-likelihood of the IGR is given by

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

Thus, the MLE is obtained by differentiating equation (3) with respect to $\boldsymbol{\beta}$ and equaling it to zero

$$\frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}=\sum_{i=1}^{n}\frac{1}{2\tau}\left[y_i-\frac{1}{\sqrt{\mathbf{x}_i^T\boldsymbol{\beta}}}\right]\mathbf{x}_i=0. \tag{4}$$

Equation (4) cannot be solved analytically as it is nonlinear in $\boldsymbol{\beta}$. The coefficient parameters can be estimated using the MLE by the Fisher-scoring algorithm or the iteratively weighted least squares (IWLS) algorithm.

$$\boldsymbol{\beta}^{(r+1)}=\boldsymbol{\beta}^{(r)}+I^{-1}\left(\boldsymbol{\beta}^{(r)}\right)S\left(\boldsymbol{\beta}^{(r)}\right), \tag{5}$$

where $I^{-1}(\boldsymbol{\beta}) = \left(-E\left(\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right)\right)^{-1}$ and $S(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}$ that are evaluated at $\boldsymbol{\beta}^{(r)}$. Thus, the estimated values of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{\mathrm{ML}} = (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\hat{\mathbf{u}}, \tag{6}$$

where $\hat{\mathbf{W}} = \mathrm{diag}(\mu_1^3, \ldots, \mu_n^3)$, $\hat{\mathbf{u}}$ is a vector with the ith element is $\hat{u}_i = (1/\hat{\mu}_i^2) + \left((y_i - \hat{\mu}_i)/\hat{\mu}_i^3\right)$ and $\hat{\mu}_i = 1/\sqrt{\mathbf{x}_i^T\hat{\boldsymbol{\beta}}}$. The covariance matrix of $\hat{\boldsymbol{\beta}}_{\mathrm{ML}}$ is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{\mathrm{ML}}) = \left(-E\left(\frac{\partial^2 l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\right)\right)^{-1} = \tau(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}, \tag{7}$$

whereas the estimated mean squared error (MSE) is given by

$$\begin{aligned} \mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{ML}}) &= \mathrm{E}(\hat{\boldsymbol{\beta}}_{\mathrm{ML}} - \hat{\boldsymbol{\beta}})^T(\hat{\boldsymbol{\beta}}_{\mathrm{ML}} - \hat{\boldsymbol{\beta}}) \\ &= \tau tr\left[(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\right] \\ &= \tau\sum_{j=1}^{p}\frac{1}{\lambda_j}, \end{aligned} \tag{8}$$

where λ_j is the eigenvalue of the matrix $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$. The parameter τ is estimated by (Uusipaikka, 2008)

$$\hat{\tau} = \frac{1}{(n-p)}\sum_{j=1}^{p}\frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i^3}. \tag{9}$$

This work is organized as follows. In Section 2, we review the ridge estimator for the inverse Gaussian regression model. Section 3 presents the almost unbiased ridge estimator for the inverse Gaussian regression model (AUIGRE) and methods for estimating the almost unbiased ridge parameter. In addition, several properties for the AUIGRE are given. In section 4, we investigate the performance of the AUIGRE using Monte Carlo simulation. The proposed AUIGRE is applied to a real dataset in Section 5. Finally, the conclusion is given in Section 6.

2 Inverse Gaussian ridge estimator

The MLE is not an efficient estimator when multicollinearity exists among the explanatory variables. When a high correlation exists among the explanatory variables, the MSE is inflated as the eigenvalues of the highly correlated explanatory variables are small (Mackinnon and Puterman, 1989; Segerstedt, 1992; Liu and Piantadosi, 2017). In order to tackle this problem, the ridge estimator (RE), proposed by Hoerl and Kennard (1970), is used by adding a positive amount to the diagonal of $\mathbf{X}^T\mathbf{X}$.

Yahya Algamal (2018) used the ridge estimator for the inverse Gaussian regression model. The inverse Gaussian ridge estimator (IGRE) is defined as

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{\mathrm{IGRE}} &= (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\hat{\boldsymbol{\beta}}_{\mathrm{ML}} \\ &= (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\hat{\mathbf{u}}, \end{aligned} \tag{10}$$

where $k \geq 0$. When $k = 0$ we have $\hat{\boldsymbol{\beta}}_{\text{IGRE}} = \hat{\boldsymbol{\beta}}_{\text{ML}}$ and when $k > 0$ we have $\|\hat{\boldsymbol{\beta}}_{\text{IGRE}}\| < \|\hat{\boldsymbol{\beta}}_{\text{ML}}\|$. The MSE of the IGRE is given by

$$\begin{aligned} \text{MSE}(\hat{\boldsymbol{\beta}}_{\text{IGRE}}) &= \text{E}(\hat{\boldsymbol{\beta}}_{\text{IGRE}} - \hat{\boldsymbol{\beta}})^T(\hat{\boldsymbol{\beta}}_{\text{IGRE}} - \hat{\boldsymbol{\beta}}) \\ &= Q_1 + Q_2, \end{aligned} \tag{11}$$

where $Q_1 = \hat{\tau}\sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j+k)^2}$, $Q_2 = k^2\sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j+k)^2}$, α_j is the jth element of $\gamma\hat{\boldsymbol{\beta}}_{\text{ML}}$ and γ is the eigenvector of the matrix $\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$. We can notice from equation (11) that Q_1 is the asymptotic variance and Q_2 is the squared bias. The choice of α_j can make the decrease in Q_1 greater than the increase in Q_2.

3 The almost unbiased inverse Gaussian ridge estimator

The ridge estimator for overcoming multicollinearity problem has a large bias. In order to solve this problem, the almost unbiased ridge estimator (AURE) was proposed by Singh et al. (1986) for linear regression models. Therefore, in this work, we propose the AURE for the inverse Gaussian regression model.

The almost unbiased ridge estimator for the inverse Gaussian regression (AUIGRE) is able to tackle multicollinearity problem and decreases the bias of the IGRE. The AUIGRE model is given by

$$\hat{\boldsymbol{\beta}}_{\text{AUIGRE}} = (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)\hat{\boldsymbol{\beta}}_{\text{ML}}. \tag{12}$$

By having the expectation of equation (12), we obtain

$$\begin{aligned} \text{E}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}}) &= (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)\text{E}(\hat{\boldsymbol{\beta}}_{\text{ML}}) \\ &= (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\text{E}(\mathbf{y}) \\ &= (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}\boldsymbol{\beta} \\ &= (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)\boldsymbol{\beta}. \end{aligned} \tag{13}$$

The bias of the AUIGRE is obtained by

$$\begin{aligned} \text{bias}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}}) &= \text{E}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}}) - \boldsymbol{\beta} \\ &= (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)\boldsymbol{\beta} - \boldsymbol{\beta} \\ &= -k^2(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}\boldsymbol{\beta} \\ &= -k^2\sum_{j=1}^{p}\frac{\alpha_j}{(\lambda_j + k)^2}. \end{aligned} \tag{14}$$

The variance of the AUIGRE is obtained by

$$\begin{aligned} \text{Var}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}}) &= (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)\text{Var}(\hat{\boldsymbol{\beta}})(\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)^T \\ &= (\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X})^{-1}\hat{\tau}(\mathbf{I} - (\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X} + k\mathbf{I})^{-2}k^2)^T \\ &= \frac{\hat{\tau}}{\lambda_j}\sum_{j=1}^{p}\left(1 - \frac{k^2}{(\lambda_j + k)^2}\right). \end{aligned} \tag{15}$$

The estimated MSE of the AUIGRE can be obtained using equations (14) and (15)

$$\begin{aligned}
\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{AUIGRE}}) &= \mathrm{Var}(\hat{\boldsymbol{\beta}}_{\mathrm{AUIGRE}}) + \Big(\mathrm{bias}(\hat{\boldsymbol{\beta}}_{\mathrm{AUIGRE}})\Big)^2 \\
&= \frac{\hat{\tau}}{\lambda_j}\sum_{j=1}^{p}\left(1-\frac{k^2}{(\lambda_j+k)^2}\right) + \left(-k^2\sum_{j=1}^{p}\frac{\alpha^2}{(\lambda_j+k)^2}\right) \\
&= \frac{\hat{\tau}}{\lambda_j}\sum_{j=1}^{p}\frac{(\lambda_j^2+2\lambda_j k)^2}{(\lambda_j+k)^2} + k^4\sum_{j=1}^{p}\frac{\alpha^2}{(\lambda_j+k)^2}. \qquad (16)
\end{aligned}$$

3.1 Obtaining the value of the parameter k

The value of the parameter k can be obtained by first differentiating equation (16) with respect to k and then equating it to zero

$$\begin{aligned}
\frac{\partial}{\partial k}(\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{\mathrm{AUIGRE}})) \qquad & \frac{-4\hat{\tau}\lambda_j k(\lambda_j+2k)}{(\lambda_j+k)^5} + \frac{4k^3\lambda_j\alpha_j^2}{(\lambda_j+k)^5} = 0 \\
& \frac{-4\hat{\tau}\lambda_j k(\lambda_j+2k)+4k^3\lambda_j\alpha_j^2}{(\lambda_j+k)^5} = 0 \\
& -4\hat{\tau}\lambda_j k(\lambda_j+2k)+4k^3\lambda_j\alpha_j^2 = 0 \\
& k^2\alpha_j^2 - \hat{\tau}k - \hat{\tau}\lambda_j = 0.
\end{aligned}$$

Hence, the value of k can be found by

$$k = \frac{\hat{\tau}\left(1 \pm \sqrt{\left(1+\frac{\lambda_j\alpha_j^2}{\hat{\tau}}\right)}\right)}{\alpha_j^2}.$$

In this work, we propose values of k for the AUIGRE in the IGR model from the work of Kibria (2003) and Khalaf and Shukur (2005)

$$k_1 = \mathrm{mean}(k_j), k_2 = \mathrm{diag}(k_j),$$

where

$$k_j = \frac{\hat{\tau}\left(1 \pm \sqrt{\left(1+\frac{\lambda_j\alpha_j^2}{\hat{\tau}}\right)}\right)}{\alpha_j^2}.$$

Theorem 1: In the IGR model, we have $\|\mathrm{bias}(\hat{\boldsymbol{\beta}}_{\mathrm{AUIGRE}})\|^2 < \|\mathrm{bias}(\hat{\boldsymbol{\beta}}_{\mathrm{IGRE}})\|^2$ for $k > 0$.

Proof: Let $D_1 = \|\text{bias}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}})\|^2 - \|\text{bias}(\hat{\boldsymbol{\beta}}_{\text{IGRE}})\|^2$. Hence, we have

$$\begin{aligned} D_1 &= \frac{k^2\alpha_j^2}{(\lambda_j+k)^2} - \frac{k^4\alpha_j^2}{(\lambda_j+k)^4} \\ &= \frac{\lambda_j^2 k^2\alpha_j^2 + 2k^3\lambda_j\alpha_j^2}{(\lambda_j+k)^4} \\ &= \sum_{j=1}^{n} k^2 \left\{ \frac{\lambda_j\alpha_j^2(\lambda_j+2k)}{(\lambda_j+k)^4} \right\}. \end{aligned}$$

Hence, for $k > 0$, the proof is completed.

Theorem 2: For the IGR model, if $\left(k > 3\hat{\tau} - \lambda_j\alpha_j^2 + \sqrt{\lambda_j^2\alpha_j^4 + 9\hat{\tau}^4 + 10\lambda_j\alpha_j^2\hat{\tau}}\right)/4\alpha^2$, for $j = 1, \ldots, p$, then the AUIGRE is superior to the IGRE in terms of the MSE.
Proof: Let $D_2 = \text{MSE}(\hat{\boldsymbol{\beta}}_{\text{IGRE}}) - \text{MSE}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}})$. Hence, we have

$$\begin{aligned} D_2 &= \frac{\hat{\tau}\lambda_j}{(\lambda_j+k)^2} + \frac{k^2\alpha_j^2}{(\lambda_j+k)^4} - \frac{\hat{\tau}(\lambda_j^2+2\lambda_j k)^2}{\lambda_j(\lambda_j+k)^4} - \frac{k^4\alpha_j^2}{(\lambda_j+k)^4} \\ &= \sum_{j=1}^{n} \left(\frac{\lambda_j \left\{(2\alpha_j^2)k^2 + (\lambda_j\alpha_j^2 - 3\hat{\tau})k - 2\hat{\tau}\lambda_j\right\} k}{(\lambda_j+k)^4} \right). \end{aligned}$$

The D_2 is a positive definite for $k > 0$, if and only if $\left\{(2\alpha_j^2)k^2 + (\lambda_j\alpha_j^2 - 3\hat{\tau})k - 2\hat{\tau}\lambda_j\right\} > 0$. Thus, this function is quadratic of k and has the following root

$$k = \frac{\left(3\hat{\tau} - \lambda_j\alpha_j^2 + \sqrt{\lambda_j^2\alpha_j^4 + 9\hat{\tau}^4 + 10\lambda_j\alpha_j^2}\right)}{4\alpha_j^2\hat{\tau}}.$$

Hence, the AUIGRE is superior to the IGRE in terms of the MSE for the IGR model, the proof is completed.

Theorem 3: For the IGR model, the AUIGRE is superior to the ML estimator.
Proof: Let $D_3 = \text{MSE}(\hat{\boldsymbol{\beta}}_{\text{ML}}) - \text{MSE}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}})$. Hence, we have

$$\begin{aligned} D_3 &= \frac{\hat{\tau}}{\lambda_j} - \frac{\hat{\tau}\lambda_j^2(\lambda_j+2k)^2}{\lambda_j(\lambda_j+k)^4} - \frac{k^4\alpha_j^2}{(\lambda_j+k)^4} \\ &= \sum_{j=1}^{n} k^2 \frac{\left\{(\hat{\tau} - \lambda_j\alpha_j^2)k^2 + 4\hat{\tau}\lambda_j k + 2\hat{\tau}\lambda_j^2\right\}}{\lambda_j(\lambda_j+k)^4}. \end{aligned} \tag{17}$$

From equation (17), it can be shown that D_3 is a positive definite if and only if $\left\{(\hat{\tau} - \lambda_j\alpha_j^2)k^2 + 4\hat{\tau}\lambda_j k + 2\hat{\tau}\lambda_j^2\right\} > 0$. Hence, the AUIGRE is superior to the MLE in terms of the MSE for the IGR model, the proof is completed.

4 Monte Carlo Simulation study

In order to compare the estimated MSE of AUIGRE with the MLE and IGRE, a Monte Carlo simulation experiment is conducted to investigate the performance of AUIGRE with several different levels of multicollinearity for the inverse link function.

The MSE is considered a measure for our comparison which is given by

$$\text{MSE}(\hat{\boldsymbol{\beta}}_{\text{AUIGRE}}) = \sum_{i=1}^{R} \frac{(\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}})^T (\hat{\boldsymbol{\beta}}_i - \hat{\boldsymbol{\beta}})}{R}. \tag{18}$$

where $R = 1000$ is the number of the Monte Carlo simulations and $\hat{\boldsymbol{\beta}}_i$ is the ith simulated value of $\hat{\boldsymbol{\beta}}$.

4.1 Simulation design of experiment

The explanatory variables $\mathbf{x}_i^T = (x_{i1}, \ldots, x_{in})$ were generated using the following equation

$$x_{ij} = (1-\rho^2)^{1/2} w_{ij} + \rho w_{ij}, \qquad i = 1, \ldots, n, \qquad j = 1, \ldots, p. \tag{19}$$

where ρ represents coefficient of correlation between the explanatory variables and w_{ij}'s are independent pseudo-random numbers. The w_{ij} were generated from the standard uniform distribution. The number of the explanatory variables was set to be 3, 6 and 9, and three different values of ρ are considered, 0.85, 0.90 and 0.99. The dependent variable, y, was generated from IG distribution with different number of sample sizes $n = 50, 100, 150$ and 200 respectively. The values of the dispersion parameter were chosen to be $0.25, 0.5$ and 0.75.

The IGR model was then fitted using the inverse link function that is given by

$$\mu_i = \frac{1}{\sqrt{\mathbf{x}_i \boldsymbol{\beta}}}, \qquad i = 1, \ldots, n.$$

The sum of the coefficient regression parameters $\boldsymbol{\beta}$ was assumed to be 1 to ensure that the values of $\mathbf{x}_i \boldsymbol{\beta}$ are positive. This is because the results can be generalized to any parameter Kibria (2003).

4.2 Simulation results

In this section, we present the Monte Carlo simulation results of the MSE, equation (18), for different selection methods of k under different combinations of n, p, τ and ρ. The results are shown in Tables 1-9.

4.2.1 The performance as a function of n

From Tables 1-9, the following points can be concluded

1. As the sample size, n, increases, the estimated MSE is decreased.

2. When the number of explanatory variables, p, dispersion parameter, τ, and the level of multicollinearity, ρ, are kept constant, then the estimated MSE of AUIGRE, MLE and IGRE decrease with the increase in sample size.

3. The results show that the AUIGRE provides a smaller estimated MSE than those of the MLE and the IGRE. Overall, the performance of AUIGRE is quite well as opposed to both the MLE and IGRE.

4.2.2 The performance as a function of ρ

Based on different levels of multicollinearity, ρ, it can be shown from Tables 1-9 that

1. For a fixed number of n, p and τ, as the level of multicollinearity, ρ, increases, the estimated MSE of the AUIGRE, IGRE, and MLE increases.

2. As a general trend, the estimated MSE of the AUIGRE is always smaller than those of the MLE and the IGRE.

4.2.3 The performance as a function of p

Based on different set sizes of explanatory variables, the following points can be noticed from Tables 1-9:

1. The simulated results indicated that the estimated MSE of the estimators increases with the increase of explanatory variables.

2. The estimated MSE of AUIGRE is smaller as compared to those of the MLE and the IGRE.

4.2.4 The performance as a function of τ

Based on different values of the dispersion parameter, the following points are seen in Tables 1-9:

1. For fixed values of n, p and ρ, the estimated MSE values of the estimators decrease as the value of the dispersion parameter increases.

2. Again, we can see that the estimated MSE of AUIGRE is always smaller as compared to those of the MLE and the IGRE.

It can be concluded that the MSE of AUIGRE is generally smaller than those of the MLE and the IGRE. Furthermore, in terms of MSE, the AUIGRE with the k_2 improved the performance of the AUIGRE in comparison with the MLE and the IGRE in most of the cases. All the selection methods of k are superior to the ML estimator in terms of MSE. Furthermore, k_2 is the best estimation method for k of the AUIGRE. On the contrast, the MLE estimator provides poor MSE in comparison with the other estimators.

Table 1: Estimated MSE when τ=0.25 and p=3 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	89.04	1.521	1.628	**1.327**	1.347
	0.90	126.2	1.485	1.633	**1.298**	1.348
	0.99	1022	1.49	1.678	**1.398**	1.436
100	0.85	27.34	1.555	1.304	1.214	**0.928**
	0.90	36.93	1.515	1.309	1.201	**0.934**
	0.99	264	1.217	1.356	1.996	**0.982**
150	0.85	18.44	1.465	1.088	1.138	**0.729**
	0.90	24.91	1.402	1.106	1.091	**0.749**
	0.99	143.9	1.125	1.167	1.884	**0.799**
200	0.85	10.32	1.501	1.091	1.82	**0.777**
	0.90	13.37	1.481	1.124	1.176	**0.807**
	0.99	67.64	1.303	1.202	1.038	**0.868**

The best values are in bold font.

Table 2: Estimated MSE when τ=0.25 and p=6 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	658.2	3.603	3.807	3.163	**3.099**
	0.90	926.8	3.553	3.833	**3.105**	3.124
	0.99	7200	3.333	3.955	**3.007**	3.301
100	0.85	186	3.412	3.058	2.929	**2.309**
	0.90	254.2	3.294	3.058	2.835	**2.307**
	0.99	1697	2.933	3.213	2.611	**2.474**
150	0.85	124.8	3.157	2.477	2.722	**1.744**
	0.90	170.2	3.105	2.534	2.689	**1.795**
	0.99	1100	2.661	2.745	2.413	**1.99**
200	0.85	90.8	3.09	2.132	2.603	**1.488**
	0.90	120.6	3.065	2.207	2.61	**1.552**
	0.99	686.8	2.681	2.37	2.439	**1.687**

The best values are in bold font.

Table 3: Estimated MSE when τ=0.25 and p=9 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	1621	6.142	6.425	5.494	**5.423**
	0.90	2226	5.921	6.435	**5.362**	5.440
	0.99	15970	5.521	6.621	**5.074**	5.676
100	0.85	666.3	5.024	4.781	4.547	**3.601**
	0.90	924.5	4.867	4.791	4.403	**3.613**
	0.99	6484	4.397	5.033	4.146	**3.884**
150	0.85	378.5	4.746	3.871	4.185	**2.745**
	0.90	516.6	4.707	3.93	4.203	**2.796**
	0.99	3319	3.998	4.238	3.936	**3.071**
200	0.85	285.7	4.683	3.313	4.185	**2.351**
	0.90	385.7	4.608	3.398	4.208	**2.413**
	0.99	2483	3.805	3.781	3.922	**2.723**

The best values are in bold font.

Table 4: Estimated MSE when τ=0.50 and p=3 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	38.32	0.976	0.890	0.976	**0.658**
	0.90	53.73	0.941	0.898	0.932	**0.659**
	0.99	397.7	0.852	0.949	0.785	**0.692**
100	0.85	13.11	0.982	0.586	0.883	**0.437**
	0.90	17.28	0.956	0.597	0.867	**0.434**
	0.99	105	0.717	0.65	0.727	**0.439**
150	0.85	8.573	0.96	0.423	0.850	**0.335**
	0.90	11.38	0.902	0.44	0.837	**0.339**
	0.99	62.65	0.722	0.541	0.715	**0.382**
200	0.85	4.985	1.015	0.459	0.881	**0.397**
	0.90	6.252	0.999	0.48	0.878	**0.406**
	0.99	29.23	0.811	0.552	0.785	**0.433**

The best values are in bold font.

Table 5: Estimated MSE when τ=0.50 and p=6 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	276	2.556	2.073	2.632	**1.505**
	0.90	383.4	2.503	2.091	2.636	**1.505**
	0.99	2779	1.994	2.213	2.15	**1.572**
100	0.85	92.69	2.447	1.402	2.563	**1.106**
	0.90	124.6	2.345	1.452	2.504	**1.12**
	0.99	748	2.049	1.615	2.473	**1.178**
150	0.85	65.08	2.324	1.004	2.428	**0.925**
	0.90	86.45	2.271	1.046	2.39	**0.928**
	0.99	493.6	1.949	1.277	2.665	**0.986**
200	0.85	45.28	2.296	**0.828**	2.155	0.921
	0.90	58.92	2.172	**0.866**	2.165	0.922
	0.99	299.1	1.948	1.065	2.502	**0.961**

The best values are in bold font.

Table 6: Estimated MSE when τ=0.50 and p=9 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	675	4.455	3.717	4.729	**2.795**
	0.90	905.7	4.29	3.758	4.695	**2.806**
	0.99	6033	3.765	4.08	4.309	**3.073**
100	0.85	330.1	3.793	2.073	4.555	**1.658**
	0.90	446.6	3.711	2.141	4.556	**1.665**
	0.99	2791	3.322	2.466	5.011	**1.759**
150	0.85	198.9	3.463	1.459	3.978	**1.43**
	0.90	266	3.415	1.514	4.199	**1.409**
	0.99	1571	3.002	1.818	5.212	**1.349**
200	0.85	145.4	3.388	**1.258**	3.689	1.517
	0.90	192.7	3.577	**1.312**	4.164	1.487
	0.99	1123	3.416	1.602	5.955	**1.393**

The best values are in bold font.

Table 7: Estimated MSE when τ=0.75 and p=3 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	25.25	0.796	0.552	1.011	**0.465**
	0.90	34.35	0.759	0.562	0.951	**0.465**
	0.99	231	0.609	0.599	0.669	**0.4428**
100	0.85	8.657	0.77	**0.358**	0.826	0.3894
	0.90	11.28	0.762	**0.363**	0.842	0.377
	0.99	63.08	0.551	0.409	0.787	**0.338**
150	0.85	5.476	0.804	**0.257**	0.802	0.328
	0.90	7.045	0.765	**0.271**	0.806	0.320
	0.99	35.63	0.576	0.324	0.711	**0.309**
200	0.85	3.329	0.865	**0.285**	0.815	0.370
	0.90	4.182	0.813	**0.294**	0.789	0.366
	0.99	17.58	0.684	**0.333**	0.778	0.358

The best values are in bold font.

Table 8: Estimated MSE when τ=0.75 and p=6 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	183	2.157	1.308	2.98	**1.104**
	0.90	253.7	2.108	1.308	2.975	**1.061**
	0.99	1764	1.665	1.406	2.505	**0.999**
100	0.85	63.28	2.049	**0.826**	2.649	1.079
	0.90	83.43	2.042	**0.848**	2.656	1.041
	0.99	467.5	1.872	0.984	3.087	**0.927**
150	0.85	46.16	1.975	**0.622**	2.298	1.03
	0.90	60.24	1.985	**0.64**	2.452	0.998
	0.99	320.3	1.867	**0.75**	2.299	0.8467
200	0.85	30.4	1.96	**0.557**	2.051	1.1
	0.90	39.2	1.839	**0.574**	2.001	1.083
	0.99	185.4	1.799	**0.696**	2.782	1.029

The best values are in bold font.

Table 9: Estimated MSE when τ=0.75 and p=9 for the AUIGRE, IGRE and MLE.

			IGRE		AUIGRE	
n	ρ	MLE	k_1	k_2	k_1	k_2
50	0.85	435.3	3.82	2.383	5.413	**2.043**
	0.90	584.5	3.87	2.434	5.639	**2.017**
	0.99	3612	3.543	2.716	6.023	**2.004**
100	0.85	224.7	3.432	**1.242**	4.732	1.658
	0.90	300.6	3.364	**1.261**	4.855	1.567
	0.99	1794	3.403	1.398	6.627	**1.277**
150	0.85	140.2	2.821	**0.96**	3.653	1.731
	0.90	185.9	2.96	**0.971**	4.073	1.647
	0.99	1048	3.432	**1.058**	7.095	1.21
200	0.85	100.7	3.024	**0.868**	3.801	1.895
	0.90	133.5	2.865	**0.874**	3.958	1.804
	0.99	731.7	3.292	**0.966**	6.738	1.406

The best values are in bold font.

5 Real data application

In this section, we apply the AUIGRE to real data. Derivatives of $n = 65$ imidazo[4,5-b] pyridine as anticancer compounds are used as explanatory variables (Algamal, 2018). The response variable is represented by the activity of the explanatory variables that are expressed as biological activities (IC50). In this work, the explanatory variables are 15 molecular descriptors (Algamal et al., 2015). In chemometrics, the quantitative structure-activity relationship (QSAR) study is a very commonly used model. The QSAR's principle is to model many chemical activities in terms of their structural characteristics over a set of chemical compounds (Algamal and Lee, 2017). Therefore, one of the most significant approaches to constructing the QSAR model is using regression models.

Algamal (2018) showed that the response variable y follows the IG distribution using a χ^2 test. Moreover, Algamal (2018) showed that the IGR model with log link function fits very well using the residual deviance test. Therefore, we fitted the IGR model with the log link function.

In order to see the correlation among the explanatory variables or not, we calculated the correlation matrix among the explanatory variables as shown in Figure 1. It can be shown that a high correlation exists among the variables SpMaxA_D, ATS8v, and MW($r = 0.96$). Also, there is a high correlation between Mor21e and Mor21v ($r = 0.93$), and a high correlation exist between ATS8v and SpMax3_Bh(s)($r = 0.92$). Therefore, multicollinearity exists among the explanatory variables.

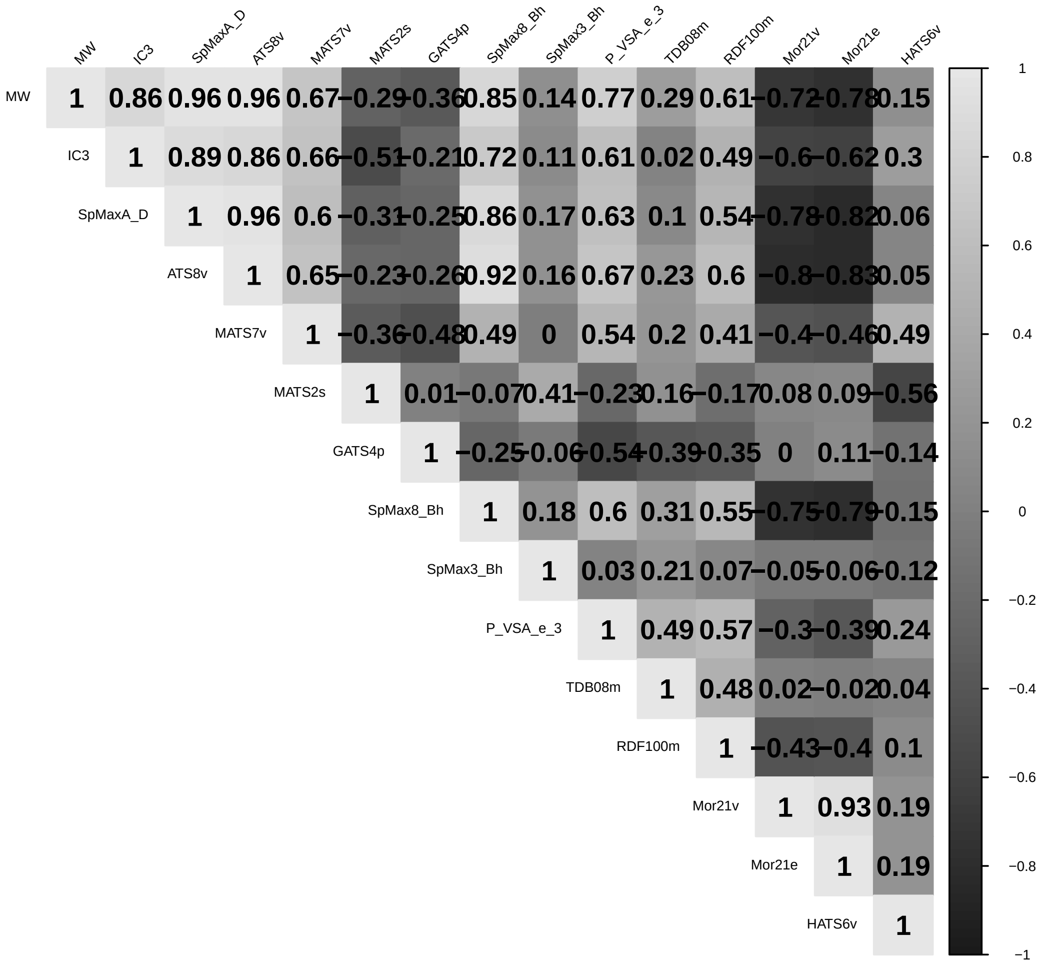

Figure 1: Correlation matrix among the explanatory variables of the real dataset.

Table 10 shows the estimated values of the IGR model coefficient parameters and the estimated values of the MSE of the real dataset for different estimators. It can be shown that the AUIGRE has the smallest MSE value in comparison with the IGRE and the MLE. In addition, the k_1 parameter has the best performance as compared to the k_2 parameter.

Table 10: The estimated coefficient parameters and the estimated MSE for the AUIGRE, IGRE and MLE.

	MLE	IGRE		AUIGRE	
		IGRE.k_1	IGRE.k_2	AUIGRE.k_1	AUIGRE.k_2
MW	1.002	0.0023	0.0041	0.0012	0.0039
IC3	1.237	0.0554	0.0079	0.0843	0.0150
SpMaxA_D	-1.102	0.0164	0.0026	0.0239	0.0048
ATS8v	-1.379	0.0218	0.0032	0.0328	0.0060
MATS7v	-1.219	-0.0064	-0.0009	-0.0100	-0.0017
MATS2s	-1.215	0.0022	0.0003	0.0031	0.0007
GATS4p	-1.237	0.0156	0.0022	0.0238	0.0043
SpMax8_Bh	2.506	0.0298	0.0042	0.0457	0.0079
SpMax3_Bh	2.069	0.0650	0.0096	0.0962	0.0183
P_VSA_e_3	2.001	0.0012	0.0001	0.0017	0.0004
TDB08m	-2.103	0.0034	0.0004	0.0052	0.0009
RDF100m	1.571	-0.0092	-0.0076	-0.0053	-0.0117
Mor21v	-2.434	-0.0003	-0.0001	-0.0007	-0.0001
Mor21e	-2.352	0.0013	0.0002	0.0015	0.0004
HATS6v	2.211	0.0012	0.0002	0.0019	0.0003
MLE	3.295	1.5087	3.8283	**0.7373**	3.5272

The best values are in bold font.

6 Conclusions

In this article, we proposed an almost unbiased ridge estimator based on the ridge estimator for the inverse Gaussian regression model. In addition, an optimal value of the ridge parameter was proposed for the AUIGRE. Furthermore, several theorems were presented that demonstrate the superiority of the proposed estimator. The performance of the proposed estimator was investigated using the Monte Carlo simulation experiment and a real dataset. Based on our results, the performance of the AUIGRE is better than the MLE and IGRE as it has a smaller MSE than the other estimators for the IGR model when multicollinearity exists.

Acknowledgements

The authors are very grateful to the University of Mosul/ College of Education for Pure Science and College of Computers Sciences and Mathematics for their provided facilities, which helped to improve the quality of this work.

References

Akram, M. N., Amin, M., and Qasim, M. (2020). A new Liu-type estimator for the inverse Gaussian regression model. *Journal of Statistical Computation and Simulation*, 90(7):1153–1172.

Algamal, Z. Y. (2018). Shrinkage estimators for gamma regression model. *Electronic Journal of Applied Statistical Analysis*, 11(1):253–268.

Algamal, Z. Y. and 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.

Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., and 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.

Babu, G. J. and Chaubey, Y. P. (1996). Asymptotics and bootstrap for inverse Gaussian regression. *Annals of the Institute of Statistical Mathematics*, 48(1):75–88.

De Jong, P. and Heller, G. Z. (2008). Generalized linear models for insurance data. Technical report, Cambridge University Press.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

Khalaf, G. and Shukur, G. (2005). Choosing ridge parameter for regression problems.

Kibria, B. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics-Simulation and Computation*, 32(2):419–435.

Liu, G. and Piantadosi, S. (2017). Ridge estimation in generalized linear models and proportional hazards regressions. *Communications in Statistics-Theory and Methods*, 46(23):11466–11479.

Lukman, A. F., Algamal, Z. Y., Kibria, B. G., and Ayinde, K. (2021). The KL estimator for the inverse Gaussian regression model. *Concurrency and Computation: Practice and Experience*, 33(13):e6222.

Mackinnon, M. J. and Puterman, M. L. (1989). Collinearity in generalized linear models. *Communications in statistics-theory and methods*, 18(9):3463–3472.

Malehi, A. S., Pourmotahari, F., and Angali, K. A. (2015). Statistical models for the analysis of skewed healthcare cost data: a simulation study. *Health economics review*, 5(1):11.

Månsson, K. and Shukur, G. (2011). A Poisson ridge regression estimator. *Economic Modelling*, 28(4):1475–1481.

Segerstedt, B. (1992). On ordinary ridge regression in generalized linear models. *Communications in Statistics-Theory and Methods*, 21(8):2227–2246.

Singh, B., Chaubey, Y., and Dwivedi, T. (1986). An almost unbiased ridge estimator. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 342–346.

Uusipaikka, E. (2008). *Confidence intervals in generalized regression models*. Chapman and Hall/CRC.

Yahya Algamal, Z. (2018). Performance of ridge estimator in inverse Gaussian regression model. *Communications in Statistics-Theory and Methods*, pages 1–14.