STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–7.
Published online in International Academic Press (www.IAPress.org)

Linearized ridge estimator in Poisson-inverse Gaussian regression model

Israa Abduljawad Saleh, Nada Nazar Alobaidi, Zakariya Yahya Algamal*

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

Abstract This paper suggests a new linearized ridge estimation of the Poisson-Inverse Gaussian Regression Model (PIGRM) which is expected to resolve the problem of multicollinearity and overdispersion in the analyses based on the count data modeling. Using asymptotics of the maximum likelihood estimator (MLE), an advantage of the suggested estimator will be harnessing the possibility of adding a parameter associated with shrinkage in order to stabilize coefficient estimates as well as decrease their variance. Wide-ranging Monte-Carlo analyses were induced to assess selection of the linearized ridge estimator in comparison to conventional MLE and various types of biased estimator in an assortment of circumstances of precinct size, engagingness of predictors and variability. It is showed consistently that linearized ridge estimator exhibits lower values of mean squared error (MSE) which implies its better accuracy in estimation and robustness. In addition, the practical utility of the approach is confirmed by the application to real-world data that demonstrate better model fit and predictive performance. These results indicate that the linearized ridge estimator should be of great use to the reliable inference about PIGRM, particularly with multicollinearity and overdispersion.

Keywords Multicollinearity, overdispersion, Poisson-inverse Gaussian regression model, ridge estimator

AMS 2010 subject classifications 62J07, 62J12

DOI: 10.19139/soic-2310-5070-2730

1. Introduction

Data modeling through counting is a way to study data on outcomes that cannot be lower than 0. They are needed in economics, epidemiology and social sciences when the dependent variable is a count, rather than having continuous values [1]. Data modeling for counts uses robust models, depending on the situation, like the Poisson, negative binomial, zero-inflated and hurdle models. The choice of a model is affected by the spread of data and how many zeros are in it [2]. Being familiar with these models is very important to make proper decisions and evaluate policies for outcomes that involve numbers.

Poisson regression model, is a form of regression analysis used to model counts; that is, when the dependent variable counts the number of times an event occurs in some unit of measurement [3]. Overdispersion happens when the actual spread of data in a set is wider than what was expected by applying a specific model [**?**, 20, 21, 22]. It usually happens when the actual data spread is further than what is expected by the model. The mean and variance of the Poisson distribution are considered to be the same for count data. When variances in the observations are greatly higher than the expected average, this assumption is broken and overdispersion takes place [5, 6, 23, 24, 25]. Since real-world populations often differ, overdispersion appears often in applied analysis as simple parametric models have trouble handling this type of variation. It may happen because of unseen differences among individuals, wrong measurements, inaccurate models or too many count data entries that equal zero. In ecological studies, data on seedling numbers might be more spread out than the Poisson model predicts, because there are more seedlings in some areas [26, 27, 28].

*Correspondence to: Zakariya Yahya Algamal (Email: zakariya.algamal@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Mosul, Iraq.

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

The problem with multicollinearity occurs when independent variables in a regression model are strongly correlated which makes it hard to estimate and understand the effects of each variable. It does not usually affect how well the model predicts, but it lowers the dependability of the estimates of coefficients and results from hypothesis testing [29, 30, 31, 32]. Builder must be able to detect and manage multicollinearity to build effective regression models [7, 8, 33, 34, 35, 36]. The maximum likelihood (ML) estimator is typically used to calculate the unknown regression coefficients when modeling the count data. The issue of multicollinearity in the data occurs when several of the independent variables are very closely connected. The covariance of the ML estimator rises as a result of the existence of multicollinearity and may provide false indications of the estimates, resulting in unstable or declining effects of the independent variables. As a result, it is difficult to draw statistical conclusions using the ML estimator in the presence of multicollinearity [9, 10, 37, 38, 39, 40, 41].

2. Poisson–inverse Gaussian regression model (PIGRM)

Poisson-inverse Gaussian (PIG) is a discrete probability distribution, based on a Poisson process, that is used to model overdispersed counts, where the mean is less than the variance. It occurs as an incessant infusion of a Poisson with an inversion Gaussian mixing distribution [11, 12, 13]. The distribution is especially useful in analysis of count data that have fat tails or too many zeros. The probability density function of the PIG distribution is defined as:

$$f(y) = \frac{\mu^y}{y!}\sqrt{\left(\frac{2\varphi}{\pi}\right)} exp(\varphi)(1+2\frac{\mu}{\varphi})^{(-s/2)} K_S(\alpha) \quad , y = 0, 1, 2, \ldots \tag{1}$$

where $s = y - \frac{1}{2}$; $\alpha = \varphi\sqrt{1+2\frac{\mu}{\varphi}}$ $and\, K_s\,(\alpha)$is modified Bessel function of the second kind. Both the parameters of the PIG distributionμ and φ have always non-negative values. The log-likelihood function is often used to obtain the maximum likelihood estimators of parameters easily. Let $l(\mu,\ \varphi)$ be the log-likelihood function, then it is for the given model is given by

$$\begin{aligned} l(\mu_i,\varphi) &= \textstyle\sum_{i=1}^{n}[y_i log(\mu_i) - log(y_i!) - \frac{1}{2}log\left(\frac{1}{\varphi}\right) + \varphi - \frac{(2y_i-1)}{4}log(1+2\frac{\mu_i}{\varphi})] \\ &+ \textstyle\sum_{i=1}^{n} logK_{(yi-\ \frac{1}{2})}\, \varphi(1+2\frac{\mu_i}{\varphi}) \end{aligned} \tag{2}$$

Let $\mu_i = exp(xi^T\beta)$, $i = 1, 2, ..., n$,where x is the matrix of covariates of order$r \times r$ where r= p+1, $\beta$is the vector of regression coefficients.
Now, the partial derivative of the log-likelihood function is given as

$$\frac{\partial l(\beta,\varphi)}{\partial \beta_i} = \sum_{i=1}^{n}\left[x_i(y_i - \frac{K_{yi-\frac{1}{2}}}{\sqrt{}(1+\frac{2}{\varphi}\exp(x_i^T\beta))}\exp(x_i^T\beta))\right] = 0 \ \ ; i = 0, 1, 2, ..., n \ and \ \ x_{i0} \ = 1 \tag{3}$$

The MLE is generally used to estimate the unknown parameters of the model. The following iterative reweighted least square (IRLS) algorithm is used to maximize the log -likelihood function [14].

$$\hat{\beta}_{MLE} = (X^T\hat{V}X)^{-1}X^T\hat{V}Z^* \tag{3}$$

represents the adjusted response variable. $\hat{V}\ and\ \ z^*$both are evaluated by using the fisher scoring procedure. The matrix MSE (MMSE) and MSE can be calculated by assuming $\alpha = \mathbb{Q}^T\hat{\beta}_{MLE}\ \ and\ \ \Lambda = diag(\lambda_1, \lambda_2, ..., \lambda_r)$ which is equal to the value of $\mathbb{Q}(X^T\hat{V}X)^{-1}\mathbb{Q}^T$where $\mathbb{Q}$ represents the orthogonal matrix whose columns are the eigenvector s of$X^T\hat{V}X$.Then the covariance and the MMSE of the $\hat{\beta}_{MLE}$are defined respectively as

$$Cov(\hat{\beta}_{MLE}) = \hat{\varphi}(X^T\hat{V}X)^{-1} \tag{4}$$

where $\hat{\varphi}$ is the estimated dispersion parameter computed as

$$\hat{\varphi} = \frac{1}{n-p}\sum_{i=1}^{n}\frac{y_i - \hat{\mu}_i}{Var(\hat{\mu}_i)} \tag{5}$$

$MMSE(\hat{\beta}_{MLE}) = \hat{\varphi}\mathbb{Q}\Lambda^{-1}\mathbb{Q}^T$. Therefore, the scalar MSE of the $\hat{\beta}_{MLE}$ is given as

$$\begin{array}{l} MSE(\hat{\beta}_{MLE}) = E(\hat{\beta}_{MLE} - \hat{\beta})^T(\hat{\beta}_{MLE} - \hat{\beta}) \\ = \hat{\varphi} \;\; tr(\mathbb{Q}\Lambda^{-1}\mathbb{Q}^T) = \hat{\varphi} \sum_{j=1}^{r} \frac{1}{\lambda_j} \end{array} \tag{6}$$

where λ_j is the j^{th}eigenvalue$X^T\hat{V}X$ of matrix.

3. The proposed estimator

To address the issue of multicollinearity in PIGRM, Liu and Gao [15] presented the linearized ridge regression estimator (LRRE). This LRRE is a more versatile estimator that incorporates the least squares, ridge estimator, Liu estimator, and numerous other shrinkage estimators. Gao and Liu [15] investigated the LRRE characteristics and demonstrated that it achieves the lower bound of MSE in the category of generalized shrinkage estimators. Further, Gao and Liu [15] have demonstrated, based on theoretical and simulation studies, that the LRRE is the best estimator in the class of generalized shrinkage estimators according to the MSE criterion.
Hence, by motivating the work done by Mansson et al. [15] and Mansson and Shukur [16], we define new estimator designated as linearized ridge PIGRM (LRPIGRM) estimator using the LRR estimator as

$$\hat{\beta}_{LRPIGRM} = \left(X'\hat{W}X + I\right)^{-1}\left(X'\hat{W}X + \gamma D\gamma'\right)\hat{\beta}_{MLE} \tag{7}$$

Where $D = diag\left(d_1, d_2, ..., d_{p+1}\right),\ d_j \in (-\infty, \infty)\ \forall_j = 1, 2, ..., p+1$,
$\gamma = (\gamma_1, \gamma_2, ..., \gamma_{p+1})$ is an orthogonal matrix such that $\gamma' X'\hat{W}X\gamma = \Lambda = diag\left(\lambda_1, \lambda_2, ..., \lambda_{p+1}\right),\ \lambda_j \geq 0\ \forall_j = 1, 2, ..., p+1$ are the eigenvalues of $X'\hat{W}X$ and $\gamma_1, \gamma_2, ..., \gamma_{p+1}$ are the corresponding eigenvectors.
The shrinkage matrix $\left(X'\hat{W}X + I\right)^{-1}\left(X'\hat{W}X + \gamma D\gamma'\right)$ in the proposed LRPIGRM estimator will reduce the inflated MSE of the MLE estimator and will become a stable estimator in Poisson regression when data suffers from the problem of multicollinearity.
Based on the asymptotic properties of MLE estimator, we derive the expressions for asymptotic bias, covariance and MSE of the LRPIGRM estimator as follows.

$$Bias\left(\hat{\beta}_{LRPIGRM}\right) = E\left(\hat{\beta}_{LRPIGRM}\right) - \beta = -\left(X'\hat{W}X + I\right)^{-1}\left(I - \gamma D\gamma'\right)\beta \tag{8}$$

$$\begin{aligned} Cov = \left(\hat{\beta}_{LRPIGRM}\right) &= Cov\left(\left(X'\hat{W}X + I\right)^{-1}\left(X'\hat{W}X + \gamma D\gamma'\right)\beta_{ML}\right) \\ &= \left(X'\hat{W}X + I\right)^{-1}\left(X'\hat{W}X + \gamma D\gamma'\right)\left(X'\hat{W}X\right)^{-1} \\ &\quad \times \left(X'\hat{W}X + \gamma D\gamma'\right)\left(X'\hat{W}X + I\right)^{-1} \end{aligned} \tag{9}$$

$$\begin{aligned} MSE\left(\hat{\beta}_{LRPIGRM}\right) &= tr\left(Cov\left(\hat{\beta}_{LRPIGRM}\right)\right) + \left[Bias\left(\hat{\beta}_{LRPIGRM}\right)\right]^T\left[Bias\left(\hat{\beta}_{LRPIGRM}\right)\right] \\ &= \hat{\varphi}\sum_{j=1}^{p+1}\frac{\left(\lambda_j + d_j\right)^2}{\lambda_j\left(\lambda_j + 1\right)^2} + \sum_{j=1}^{p+1}\frac{\left(1 - d_j\right)^2}{\left(\lambda_j + 1\right)^2}\beta_j^2 \end{aligned} \tag{10}$$

4. Simulation study

In this section, a discussion is presented about data generation with different factors. Also, the assessment criterion is discussed to assess the performance of our propose ridge estimator for the PIGRM. The $PIG(\mu, \varphi)$ distribution is used to generate the response variable of the PIGRM, where

$$\mu_i = \exp(\beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}) \ , i - 1, 2, ..., n. \tag{11}$$

where β_p represents the parameters of the PIGRM. These regression parameters are selected by taking into view the condition of $\sum_{j=1}^{p} \beta_j^2 = 1$. The correlated explanatory variables are generated as follows

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{i(j+1)} \ , i = 1, ..., n \ ; j = 1, ..., p, \tag{12}$$

Then n=30, 50, 100, and 200, p is taken to be 4, 8 and 12, and φ is 2 and 6. The experiment was replicated 1000 times and the mean squared error (MSE) was employed to evaluate the estimators' performance.

$$MSE(\beta^*) = \frac{1}{1000} \sum_{j=1}^{100} (\beta_{ij}^* - \beta_i)' (\beta_{ij}^* - \beta_i) \tag{13}$$

where β_{ij}^* is the estimator and β_i is the parameter.

The Tables 1-4 present the MSE values for four estimators: MLE, ridge estimator, Liu estimator, and LRPIGRM across different numbers of predictors (p = 2, 6, 8, 12) and correlation levels (r = 0.90, 0.95, 0.99). MSE measures the average squared difference between estimated and true parameter values, combining bias and variance; lower MSE indicates better estimator accuracy.

For each fixed number of predictors, MSE increases as the correlation among predictors increases from 0.90 to 0.99. This is expected because higher multicollinearity inflates estimation variance, making parameter estimation more challenging. For a fixed correlation, increasing the number of predictors generally leads to higher MSE values, reflecting increased model complexity and estimation difficulty.

MLE consistently has the highest MSE across all scenarios, indicating it is the least efficient estimator when multicollinearity is present. Ridge estimator reduces MSE compared to MLE by introducing shrinkage, which stabilizes estimates. Further, Liu estimator further improves upon ridge by balancing bias and variance more effectively, resulting in lower MSE.

LRPIGRM achieves the lowest MSE values in all cases, demonstrating superior estimation accuracy and robustness, especially under high correlation and larger model sizes. For example, when p=12, r=0.99, n=30, and $\varphi = 2$, MSE decreases from 13.396 (MLE) to 5.0441 (LRPIGRM), showing a substantial gain in estimation precision. The results highlight that shrinkage-based estimators (Ridge, Liu, LRPIGRM) outperform the classical MLE in terms of MSE, particularly as predictor correlation and dimensionality increase. Among these, the LRPIGRM estimator provides the best performance by achieving the lowest MSE, making it the preferred choice for regression problems with multicollinearity and higher predictor counts. Lower MSE values indicate more reliable and precise parameter estimates, which is crucial for accurate prediction and inference.

5. Real data results

In this application, we carried out a careful analysis of the dataset of the nut. The documentation the two authors include [17, 18]. The That data set can be found in the COUNT library of the R programming environment. The data gathering, that comprises of fifty-two observations and four predictors. Specifically, this number the response variable is the number of cones which are salvaged by the squirrels. The determinants are the number of DBH (diameter at breast height) per plot (x1). mean tree height per plot (x2). canopy closure% (x3) and number of trees plot-1 (x4). To ascertain the model that fitted best on the dataset we fitted the PIGRM, negative binomial regression model (NBRM) and Poisson regression model (PRM). The findings have been presented in Table 5. The PIGRM gives the best fit in terms of residual variance since it has considerably lower errors. As AIC is concerned, NBRM, compared with PIGRM, is a very slight improvement on, whereas in residual variance, there was a significant drop. It is reasonable to suggest that PIGRM can provide a more appropriate model even though AIC is a little bit worse. PRM has the lowest scores on these two metrics, which is why it is the undesirable one. Furthermore, A dispersion

Table 1. Values of the MSE when n=30

φ	*p*	*r*	*MLE*	*Ridge*	*Liu*	*LRPIGRM*
2	*4*	*0.90*	*6.28*	*5.927*	*5.849*	***5.3859***
		0.95	*7.377*	*6.412*	*6.138*	***5.9998***
		0.99	*8.42*	*6.929*	*6.196*	***6.1009***
	8	*0.90*	*6.397*	*5.757*	*5.529*	***5.2789***
		0.95	*8.442*	*6.588*	*6.032*	***5.8287***
		0.99	*9.374*	*7.313*	*6.539*	***5.6359***
	12	*0.90*	*7.634*	*5.385*	*5.21*	***4.9792***
		0.95	*10.857*	*6.736*	*6.019*	***5.0733***
		0.99	*13.396*	*7.063*	*5.835*	***5.0441***
6	*4*	*0.90*	*6.391*	*6.038*	*5.96*	***5.4969***
		0.95	*7.488*	*6.523*	*6.249*	***6.1108***
		0.99	*8.531*	*7.04*	*6.307*	***6.2119***
	8	*0.90*	*6.508*	*5.868*	*5.64*	***5.3899***
		0.95	*8.553*	*6.699*	*6.143*	***5.9397***
		0.99	*9.485*	*7.424*	*6.65*	***5.7469***
	12	*0.90*	*7.745*	*5.496*	*5.321*	***5.0902***
		0.95	*10.968*	*6.847*	*6.13*	***5.1843***
		0.99	*13.507*	*7.174*	*5.946*	***5.1551***

Table 2. Values of the MSE when n=50

φ	*p*	*r*	*MLE*	*Ridge*	*Liu*	*LRPIGRM*
2	*4*	*0.90*	*6.069*	*5.716*	*5.638*	***5.1749***
		0.95	*7.166*	*6.201*	*5.927*	***5.7888***
		0.99	*8.209*	*6.718*	*5.985*	***5.8899***
	8	*0.90*	*6.186*	*5.546*	*5.318*	***5.0679***
		0.95	*8.231*	*6.377*	*5.821*	***5.6177***
		0.99	*9.163*	*7.102*	*6.328*	***5.4249***
	12	*0.90*	*7.423*	*5.174*	*4.999*	***4.7682***
		0.95	*10.646*	*6.525*	*5.808*	***4.8623***
		0.99	*13.185*	*6.852*	*5.624*	***4.8331***
6	*4*	*0.90*	*6.086*	*5.733*	*5.655*	***5.1919***
		0.95	*7.183*	*6.218*	*5.944*	***5.8058***
		0.99	*8.226*	*6.735*	*6.002*	***5.9069***
	8	*0.90*	*6.203*	*5.563*	*5.335*	***5.0849***
		0.95	*8.248*	*6.394*	*5.838*	***5.6347***
		0.99	*9.18*	*7.119*	*6.345*	***5.4419***
	12	*0.90*	*7.44*	*5.191*	*5.016*	***4.7852***
		0.95	*10.663*	*6.542*	*5.825*	***4.8793***
		0.99	*13.202*	*6.869*	*5.641*	***4.8501***

test revealed over-dispersion in the model hence the need to correct it. There is inappropriateness of the Poisson regression model in this dataset.

From Table 6, the MSE value of 125.221 is the highest among all estimators, indicating that MLE produces the least accurate estimates in this context. This is often due to issues like multicollinearity or small sample sizes, where MLE can have high variance and unstable estimates. On the other hand, ridge estimator reduces the MSE to 101.783, showing a significant improvement over MLE. By introducing a penalty term that shrinks coefficients, ridge estimator reduces variance at the cost of some bias, resulting in more stable and reliable estimates. While, the Liu estimator further improves the MSE to 89.308, outperforming ridge estimator. This indicates that Liu's approach to shrinkage provides a better balance between bias and variance, leading to more accurate parameter estimates.

The LRPIGRM estimator achieves the lowest MSE of 71.242, demonstrating the best performance among the compared methods. This suggests that LRPIGRM is the most effective at handling the data characteristics (such as multicollinearity or model complexity) in this study, providing the most precise and reliable estimates. The decreasing trend in MSE values from MLE to LRPIGRM clearly illustrates the benefit of applying shrinkage and regularization techniques in regression estimation. While MLE suffers from high variance, especially under

Table 3. Values of the MSE when n=100

φ	p	r	*MLE*	*Ridge*	*Liu*	*LRPIGRM*
2	4	0.90	5.858	5.505	5.427	**4.9639**
		0.95	6.955	5.99	5.716	**5.5778**
		0.99	7.998	6.507	5.774	**5.6789**
	8	0.90	5.975	5.335	5.107	**4.8569**
		0.95	8.02	6.166	5.61	**5.4067**
		0.99	8.952	6.891	6.117	**5.2139**
	12	0.90	7.212	4.963	4.788	**4.5572**
		0.95	10.435	6.314	5.597	**4.6513**
		0.99	12.974	6.641	5.413	**4.6221**
6	4	0.90	5.781	5.428	5.35	**4.8869**
		0.95	6.878	5.913	5.639	**5.5008**
		0.99	7.921	6.43	5.697	**5.6019**
	8	0.90	5.898	5.258	5.03	**4.7799**
		0.95	7.943	6.089	5.533	**5.3297**
		0.99	8.875	6.814	6.04	**5.1369**
	12	0.90	7.135	4.886	4.711	**4.4802**
		0.95	10.358	6.237	5.52	**4.5743**
		0.99	12.897	6.564	5.336	**4.5451**

Table 4. Values of the MSE when n=200

φ	p	r	*MLE*	*Ridge*	*Liu*	*LRPIGRM*
2	4	0.90	5.647	5.294	5.216	**4.7529**
		0.95	6.744	5.779	5.505	**5.3668**
		0.99	7.787	6.296	5.563	**5.4679**
	8	0.90	5.764	5.124	4.896	**4.6459**
		0.95	7.809	5.955	5.399	**5.1957**
		0.99	8.741	6.68	5.906	**5.0029**
	12	0.90	7.001	4.752	4.577	**4.3462**
		0.95	10.224	6.103	5.386	**4.4403**
		0.99	12.763	6.43	5.202	**4.4111**
6	4	0.90	5.476	5.123	5.045	**4.5819**
		0.95	6.573	5.608	5.334	**5.1958**
		0.99	7.616	6.125	5.392	**5.2969**
	8	0.90	5.593	4.953	4.725	**4.4749**
		0.95	7.638	5.784	5.228	**5.0247**
		0.99	8.57	6.509	5.735	**4.8319**
	12	0.90	6.83	4.581	4.406	**4.1752**
		0.95	10.053	5.932	5.215	**4.2693**
		0.99	12.592	6.259	5.031	**4.2401**

Table 5. Modelling fitting

criteria	PIGRM	PRM	NBRM
AIC	397.9	873.1	406.0
Residual Variance	0.5498	661.2	661.2

challenging data conditions, ridge and Liu estimators improve estimation by introducing bias to reduce variance. The LRPIGRM estimator, with the lowest MSE, offers the best trade-off, yielding the most accurate and stable parameter estimates in this comparison.

Table 6. Performance of the used estimators

criteria	MLE	Ridge	Liu	LRPIGRM
MSE	125.221	101.783	89.308	71.242

This study introduced and evaluated a linearized ridge estimator within the PIGRM framework to address the challenges posed by multicollinearity among explanatory variables in count data modeling with overdispersion.

Through extensive Monte Carlo simulations, the proposed estimator demonstrated superior performance compared to the MLE and other biased estimators, notably achieving lower MSE across various sample sizes and correlation structures. The application to real-world dataset further confirmed the practical advantages of the linearized ridge estimator, providing more stable and reliable parameter estimates while effectively reducing estimation variance. These findings underscore the estimator's potential as a robust alternative for practitioners dealing with correlated covariates in overdispersed count data contexts. Future research may focus on extending this approach to zero-inflated and multivariate count models, as well as exploring adaptive biasing parameter selection techniques to further enhance estimation accuracy.

REFERENCES

1. Hammood, N. M. and N. K. Rashad and Z. Y. Algamal, *A New Ridge-Type Estimator in the Zero-Inflated Bell Regression Model*, Iraqi Journal for Computer Science and Mathematics, vol. 6, no. 2, 2025.
2. Seifollahi, S. and H. Bevrani and Z. Y. Algamal, *Shrinkage Estimators in Zero-Inflated Bell Regression Model with Application*, Journal of Statistical Theory and Practice, vol. 19, no. 1, 2024.
3. Alangood, Heyaa Nadhim Ahmed and Algamal, Zakariya Yahya and Khaleel, Mundher Abdullah, *Variable selection in Poisson regression model based on chaotic meta-heuristic search algorithm*, BIO Web of Conferences, 2024.
4. Hawa, Noora Suhail and Mustafa, Marwah Yahya and Kibria, BM Golam and Algamal, Zakariya Yahya, *Bootstrap Liu-type estimator for Conway-Maxwell-Poisson regression model*, Communications in Statistics-Simulation and Computation, pp. 1–12, 2025.
5. Alreshidi, Nasser A and Alrasheedi, Masad A and Lukman, Adewale F and Alrweili, Hleil and Farghali, Rasha A, *Modified Kibria–Lukman Estimator for the Conway–Maxwell–Poisson Regression Model: Simulation and Application*, Mathematics, vol. 13, no. 5, pp. 794, 2025.
6. Özkale, M. R. and U. Mammadova, *Performance analysis of shrinkage estimators in Conway-Maxwell-Poisson regression model*, Communications in Statistics-Theory and Methods, pp. 1–29, 2025.
7. Algamal, Z. Y., *Shrinkage parameter selection via modified cross-validation approach for ridge regression model*, Communications in Statistics - Simulation and Computation, vol. 49,no. 7, pp. 1922–1930, 2018.
8. Lukman, Adewale F and Algamal, Zakariya Y and Kibria, BM Golam and Ayinde, Kayode, *The KL estimator for the inverse Gaussian regression model*, Concurrency and Computation: Practice and Experience, vol. 33,no. 13, 2021.
9. Algamal, Zakariya and Lukman, Adewale and Golam, BM Kibria and Taofik, Arowolo, *Modified Jackknifed Ridge Estimator in Bell Regression Model: Theory, Simulation and Applications*, Iraqi Journal For Computer Science and Mathematics, vol. 4, no. 1, pp. 146-154, 2023.
10. Algamal, Zakariya Yahya and Lukman, Adewale F and Abonazel, Mohamed R and Awwad, Fuad A, *Performance of the Ridge and Liu Estimators in the zero-inflated Bell Regression Model*, Journal of Mathematics,vol. 2022, pp. 1–15, 2022.
11. Holla, M., *On a Poisson-inverse Gaussian distribution*, Metrika, vol. 11, no. 1, pp. 115–121, 1967.
12. Farghali, Rasha A and Lukman, Adewale F and Algamal, Zakariya and Genc, Murat and Attia, Hend, *An Alternative Estimator for Poisson–Inverse-Gaussian Regression: The Modified Kibria–Lukman Estimator*, Algorithms, vol. 18, no. 3, 2025.
13. Alrweili, H., *Liu-Type Estimator for the Poisson-Inverse Gaussian Regression Model: Simulation and Practical Applications*, Statistics, Optimization and Information Computing, vol. 12, no. 4, pp. 982–1003, 2024.
14. Alkhateeb, Ahmed and Algamal, Zakariya, *Jackknifed Liu-type estimator in Poisson regression model*, Journal of the Iranian Statistical Society, vol. 19, no. 1, pp. 21–37, 2022.
15. Liu, X. and F. Gao, *Linearized Ridge Regression Estimator in Linear Regression*, Communications in Statistics - Theory and Methods, vol. 40, no. 12, pp. 2182–2192, 2011.
16. Gao, F. and X.-Q. Liu, *Linearized Ridge Regression Estimator Under the Mean Squared Error Criterion in a Linear Regression Model*, Communications in Statistics - Simulation and Computation, vol. 40, no. 9, pp. 1434–1443, 2011.
17. Hardin, J. W. and J. M. Hilbe, *Generalized linear models and extensions*, Stata press, 2007.
18. Lukman, Adewale F and Albalawi, Olayan and Arashi, Mohammad and Allohibi, Jeza and Alharbi, Abdulmajeed Atiah and Farghali, Rasha A, *Robust Negative Binomial Regression via the Kibria–Lukman Strategy: Methodology and Application*, Mathematics, vol. 12, no. 18, pp. 2929, 2024.
19. Algamal, Z. Y., & Asar, Y. *Liu-type estimator for the gamma regression model* , Communications in Statistics-Simulation and Computation, vol. 49, no . 8, p. 2035-2048, 2020.
20. Algamal, Z. Y., & Lee, M. H. *A new adaptive L1-norm for optimal descriptor selection of high-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives* , SAR and QSAR in Environmental Research, vol. 28, no. 1 , p. 75-90, 2017.
21. Kahya, M. A., Altamir, S. A., & Algamal, Z. Y. *Improving whale optimization algorithm for feature selection with a time-varying transfer function* , Numerical Algebra, Control and Optimization, vol. 11, no. 1 , p. 87-98 , 2020.
22. Algamal, Z. Y., Qasim, M. K., & Ali, H. T. M. *A QSAR classification model for neuraminidase inhibitors of influenza A viruses (H1N1) based on weighted penalized support vector machine* , SAR and QSAR in Environmental Research, vol. 28 , no. 5, p. 415-426 , 2017.
23. Algamal, Z. Y., Lee, M. H., & Al-Fakih, A. M. *High-dimensional quantitative structure–activity relationship modeling of influenza neuraminidase a/PR/8/34 (H1N1) inhibitors based on a two-stage adaptive penalized rank regression* , Journal of Chemometrics, vol. 30 , no.2 m p. 50-57 , 2016.
24. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. *High-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives based on the sparse logistic regression model with a bridge penalty* , Journal of Chemometrics, vol.

31 , p.6, p. e2889 , 2017.
25. Algamal, Z. Y., Qasim, M. K., Lee, M. H., & Ali, H. T. M. *High-dimensional QSAR/QSPR classification modeling based on improving pigeon optimization algorithm* , Chemometrics and Intelligent Laboratory Systems, vol. 206, p. 104170 , 2020.
26. Ismael, O. M., Qasim, O. S., & Algamal, Z. Y. *Improving Harris hawks optimization algorithm for hyperparameters estimation and feature selection in v-support vector regression based on opposition-based learning* , Journal of Chemometrics, vol 34 , no. 11, e3311, 2020.
27. Abonazel, M. R., Algamal, Z. Y., Awwad, F. A., & Taha, I. M. *A new two-parameter estimator for beta regression model: method, simulation, and application* , Frontiers in Applied Mathematics and Statistics, vol. 7,p. 780322 ,2022.
28. Algamal, Z. Y., & Abonazel, M. R. *Developing a Liu-type estimator in beta regression model* , Concurrency and Computation: Practice and Experience, vol.34 ,no. 5 , p. e6685.
29. Algamal, Z., & Ali, H. M. *An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression* , Electronic Journal of Applied Statistical Analysis, vol . 10 , no . 1 , 242-256 , 2017.
30. Salih, A. M., Algamal, Z., & Khaleel, M. A. *A new ridge-type estimator for the gamma regression model* , Iraqi Journal for Computer Science and Mathematics, vol . 5 , no .1 , p. 85-98.
31. Mahmood, S. W., Basheer, G. T., & Algamal, Z. Y. *Quantitative Structure–Activity Relationship Modeling Based on Improving Kernel Ridge Regression* , Journal of Chemometrics, vol. 39, no. 5, p. e70027, 2025.
32. Mahmood, S. W., Basheer, G. T., & Algamal, Z. Y. *Improving kernel ridge regression for medical data classification based on meta-heuristic algorithms* , Kuwait Journal of Science, vol. 52, no. 3 , p. 100408 , 2025
33. Algamal, Z. Y., Alhamzawi, R., Ali, H. T. M. *Gene selection for microarray gene expression classification using Bayesian Lasso quantile regression*, Computers in biology and medicine, vol. 97, p. 145-152, 2018
34. Algamal, Z. Y., Lee, M. H. *A novel molecular descriptor selection method in QSAR classification model based on weighted penalized logistic regression*, Journal of Chemometrics, vol. 31, no.(10), e2915, 2017
35. Qasim, M. K., Algamal, Z. Y., Ali, H. M. *A binary QSAR model for classifying neuraminidase inhibitors of influenza A viruses (H1N1) using the combined minimum redundancy maximum relevancy criterion with the sparse support vector machine*, SAR and QSAR in Environmental Research, vol. 29, no.(7), p.517-527, 2018
36. Algamal, Z. Y., Lee, M. H. *Applying penalized binary logistic regression with correlation based elastic net for variables selection*, Journal of Modern Applied Statistical Methods, vol. 14, no.(1), p.15, 2015
37. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., Aziz, M. *High-dimensional QSAR modelling using penalized linear regression model with L 1/2-norm*, SAR and QSAR in Environmental Research, vol. 27, no.(9), p.703-719, 2016
38. Al-Taweel, Y., Algamal, Z. *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model*, Periodicals of Engineering and Natural Sciences, vol. 8, no.(1), p.248-255, 2020
39. Ewees, A. A., Algamal, Z. Y., Abualigah, L., Al-Qaness, M. A., Yousri, D., Ghoniem, R. M., Abd Elaziz, M. *A cox proportional-hazards model based on an improved aquila optimizer with whale optimization algorithm operators*, Mathematics, vol. 10, no.(8), p.1273, 2022
40. Shamany, R., Alobaidi, N. N., Algamal, Z. Y. *A new two-parameter estimator for the inverse Gaussian regression model with application in chemometrics*, Electronic Journal of Applied Statistical Analysis, vol. 12, no.(2), p.453-464, 2019
41. Awwad, F. A., Odeniyi, K. A., Dawoud, I., Algamal, Z. Y., Abonazel, M. R., Kibria, B. G., Eldin, E. T. *New two-parameter estimators for the logistic regression model with multicollinearity*, WSEAS Trans. Math, vol. 21, p.403-414, 2022