Iraqi Journal for Computer Science and Mathematics

Manuscript 1248

A New Ridge-Type Estimator in the Zero-Inflated Bell Regression Model

Nawal Mahmood Hammood

Nadwa Khazaal Rashad

Zakariya Yahya Algamal

ORIGINAL STUDY

A New Ridge-Type Estimator in the Zero-Inflated Bell Regression Model

Nawal Mahmood Hammood [a], **Nadwa Khazaal Rashad** [a], **Zakariya Yahya Algamal** [b,*]

[a] Department of Management Information Systems, University of Mosul, Mosul, Iraq
[b] Department of Statistics and Informatics, University of Mosul, Mosul, Iraq

ABSTRACT

Count data modeling requires usage of the Poisson regression model as a primary analytic method. Excess dispersion in variables makes the model unfit to use when the Poisson distribution mean value differs from its variance value. Data fits well with the results obtained by using the Bell regression model. Excess zeros occur frequently in the observed count data records. The Zero-Inflated Bell regression model is a substitute for the Bell regression model in this situation. The approach of maximum likelihood is mostly used to estimate the Zero-Inflated Bell regression model's parameters. When modeling the link between the response variable and two or more explanatory variables in an extended linear model, such as the Zero-Inflated Bell regression model, linear dependency poses a risk in a real-life application. It decreased the greatest likelihood estimator's effectiveness. To address this problem, we proposed a new ridge estimator for the Zero-Inflated Bell regression model. The results of the simulations and implementations validate the suggested approaches' superiority to the traditional maximum likelihood estimator.

Keywords: Bell regression, Ridge estimator, Liu estimator, Over-dispersion, Poisson regression, Zero-Inflated Bell

1. Introduction

When the response variable does not have a gaussian (normal) distribution, a generalized linear model is used [1–4]. Data in the form of counts are typically prevalent in modeling, especially in the fields of economics and medical. The Poisson regression model is unquestionably the most widely used model for count data in practice [5, 6]. The distribution frequently makes the assumption that the variance and mean of the distribution are the same. Over-dispersion, or variation that is greater than the mean, is a significant flaw in the Poisson regression model that typically arises in count data. To model count data with over-dispersion, On the Bell distribution and its associated regression model for count data Castellares, et al. [7] developed an alternative discrete distribution model known as the Bell regression model. The Ridge estimator and the Liu estimator were recently presented for the parameter estimation of the Bell regression model with multicollinearity by Amin, et al. [8], Majid, et al. [9–14], and [15]. The presence of excess zeros in the count data is another drawback of the Bell regression model. Numerous fields, including medicine, public health, environmental sciences, agriculture, and manufacturing applications, have a tendency for the count to have a lot of zeros (or zero-inflation) An alternative that offers a better match for this kind of count data is the Zero-Inflated Bell regression model.

Actually, the computation of a MLE in real applications shows the following consequence of high multicollinearity of the independent variables: since the $\mathbf{X}^T\mathbf{X}$ is near to singularity, its inverse leads to a high variance for the MLE. That is why commonly used methods for estimation such as MLE

Received 12 September 2024; revised 14 February 2025; accepted 10 March 2025.
Available online 26 April 2025

* Corresponding author.
E-mail addresses: nawal_almamary@uomosul.edu.iq (N. M. Hammood), nadwa_khazaal@uomosul.edu.iq (N. K. Rashad), zakariya.algamal@uomosul.edu.iq (Z. Y. Algamal).

https://doi.org/10.52866/2788-7421.1248

often provide poor outcomes. Due to the problem of multicollinearity in the linear regression model, many writers have advised on the use of Ridge, Liu, Liu-type, among others [20–29]. Additionally, robust estimators have been suggested to address the multicollinearity and outlier value issues concurrently.

The Bell distribution is a discrete probability distribution that counts the number of partitions of a set. The zero-inflated Bell distribution (ZIBRM) expands on the idea of zero-inflation. When modeling count data with a significant percentage of zeros, this distribution can be very helpful because it provides greater flexibility for data fitting than typical count models. Multicollinearity is a prevalent problem among continuous explanatory variables, as the literature makes apparent. The Zero-Inflated Bell regression model with multicollinearity taken into account has not been used in any research. Ecological research investigations employ The Zero-Inflated Bell regression model as a model for analyzing species counts in different habitats. Observed species absent from specific locations causes numerous ecological data sets to contain numerous zero counts. Researchers understand species distribution and abundance with better precision by using the ZIBell model since it addresses both zero occurrences and definite count observations.

The primary goal of this work is to create a novel ridge ZIBRM estimator for over-dispersion count data modeling. Compared to several of the current estimators in GLM, the suggested estimator will effectively outperform them. A real-world application and simulated examples will demonstrate the superiority of the suggested estimators.

2. Zero-inflated Bell regression model

Let $(t_i, z_i),\ i = 1, 2, \ldots, n$ is independent observed data with the predictor vector $z_i \in R^{p+1}$ and the response variable $t_i \in R$ which follows a distribution that belongs to the Bell distribution. Then, the density function of t_i can be expressed as

$$P(T = t) = \frac{\tau^t e^{-e^{\tau}+1} B_t}{t!},\quad t = 0, 1, 2, \ldots, \tag{1}$$

where $\tau > 0$ and $B_t = (1/e)\sum_{q=0}^{\infty}(q^t/q!)$ is the Bell numbers. Then

$$E(t) = \tau e^{\tau}, \tag{2}$$

$$Var(t) = \tau(1+\tau)e^{\tau}. \tag{3}$$

Assuming $\phi = \tau e^{\tau}$ and $\tau = D_{\circ}(\phi)$ where $D_{\circ}(.)$ is the Lambert function. Then Eq. (1) can be written in the new parameterization as

$$P(T = t) = \exp\left(1 - e^{D_{\circ}(\phi)}\right)\frac{D_{\circ}(\phi)^t B_t}{t!},\quad t = 0, 1, 2, \ldots, \tag{4}$$

The linear function is $\eta_i = \sum_{j=1}^{p} z_{ij}\alpha_j = z_i^T\alpha$ with $z_i^T = (z_{i1}, z_{i2}, \ldots, z_{ip})$ and $\alpha = (\alpha_1, \ldots, \alpha_p)^T$. The link function is $\mu_i = g^{-1}(\eta_i) = g^{-1}(z_i^T\alpha)$. The Bell regression model (BRM) can be modeled by assuming $\phi_i = \exp(z_i^T\alpha)\ \exp(\exp(z_i^T\alpha))$ and $\log\phi_i = z_i^T\alpha\ \exp(z_i^T\alpha)$ as $t_i \sim \text{Bell}(D_{\circ}(\phi_i))$. The parameter estimation in the BRM is achieved through using the MLE based on the iteratively reweighted least-squares algorithm. The log-likelihood is defined

$$\ell(\alpha,\phi) = \sum_{i=1}^{n} t_i \log\left(\exp\left(z_i^T\alpha\right)\exp\left(e^{\left(z_i^T\alpha\right)}\right)\right) + \sum_{i=1}^{n}\left(1 - e^{e^{\left(z_i^T\alpha\right)}e^{e^{\left(z_i^T\alpha\right)}}}\right) + \log B_t - \log\left(\prod_{i=1}^{n} t_i!\right). \tag{5}$$

Then, the MLE is derived by equaling the first derivative of Eq. (5) to zero. After solving the first derivative iteratively, the estimated coefficients are defined as

$$\hat{\alpha}_{\text{MLE}} = \left(Z^T\hat{W}Z\right)^{-1}Z^T\hat{W}\hat{v}, \tag{6}$$

where $\hat{W} = \text{diag}[(\partial\mu_i/\partial\eta_i)^2/V(t_i)]$ and $\hat{v}$ is a vector where i^{th} element equals to $\hat{v}_i = \log\hat{\phi}_i + [(t_i - \hat{\mu}_i)/\sqrt{\text{var}(\hat{\phi}_i)}]$. The MLE is distributed asymptotically normal with a covariance matrix as

$$\text{cov}(\hat{\alpha}_{\text{MLE}}) = \left[-E\left(\frac{\partial^2\ell(\alpha,\phi)}{\partial\alpha\,\partial\alpha^T}\right)\right]^{-1} = \left(Z^T\hat{W}Z\right)^{-1}. \tag{7}$$

It is well knowledge that count data frequently have more zeros than expected, or mean zero counts, than would be expected. Numerous zeros in a count, or zero inflation, are frequently seen in many practical applications.

When there are too many zeroes in the sample, the BRM is insufficient. To model counts data with an excess of zeroes, we introduced the Zero-inflated Bell regression model (ZIBRM) in this paper. Thus, the following formulation of the ZIBRM is as:

$$p(T = t) = \begin{cases} \sigma + (1-\sigma)\exp\left(1 - e^{D(\tau)}\right), & t = 0, \\ (1-\sigma)\exp\left(1 - e^{D(\tau)}\right)\frac{D(\tau)^t B_t}{t!}, & t > 0, \end{cases} \tag{8}$$

where $\sigma \in (0, 1)$. Then, according to Eq. (8), $E(t) = (1-\sigma)\tau,\quad \text{var}(t) = (1-\sigma)\tau[1 + D(\tau) + \tau\sigma]$.

In zero-inflated regression modeling, there are two link functions used as:

$$\log(\mu_i) = \eta_{1i} = z_i^T \alpha, \quad \log\left(\frac{\sigma_i}{1-\sigma_i}\right) = \eta_{2i} = \mathbf{s}_i^T \vartheta, \tag{9}$$

where $\vartheta = (\vartheta_1, \ldots, \vartheta_q)^T$ are vectors of unknown regression coefficients which are assumed to be functionally independent, and $\mathbf{s}_i^T = (s_{i1}, \ldots, s_{iq})$are observation on q known explanatory variables. The log-likelihood function is defined as

$$\begin{aligned}\ell(\alpha, \vartheta) = &\sum_{t_i:t_i=0} \log\left[e^{\eta_{2i}} + \exp\left(1 - e^{D(\mu_i)}\right)\right] \\ &- \sum_{i=1}^{n} \log\left(1 - e^{\eta_{2i}}\right) + \sum_{t_i:t_i>0} t_i \log\left[D(\mu_i)\right] \\ &- \sum_{t_i:t_i>0} e^{D(\mu_i)},\end{aligned} \tag{10}$$

Then, the MLE estimator is $\hat{\alpha}_{MLE}$ and $\hat{\vartheta}_{MLE}$ [12].

3. The proposed estimator

In the presence of multicollinearity, the $rank(Z^T\hat{W}Z) \leq rank(Z)$, and, therefore, the near singularity of $Z^T\hat{W}Z$ makes the estimation unstable and enlarges the variance [16, 17]. When multicollinearity is present, it has been repeatedly shown that the ridge estimator (RE) and Liu estimator (LE) are appealing alternatives to the MLE. Amin, et al. [8] and Majid, et al. [9] have proposed the ridge estimator and Liu estimator in the Bell regression model, respectively. To extend the RE and LE for ZIBRM and according to Asar, et al. [18], Kibria, et al. [19], Algamal, et al. [20] defined these estimators as, respectively,

$$\hat{\alpha}_{\text{Ridge}} = \left(Z^T\hat{W}Z + c\mathbf{I}\right)^{-1} Z^T\hat{W}Z\hat{\alpha}_{\mathbf{MLE}}, \tag{11}$$

$$\hat{\alpha}_{\text{Liu}} = \left(Z^T\hat{W}Z + \mathbf{I}\right)^{-1}\left(Z^T\hat{W}Z + d\,\mathbf{I}\right)\hat{\alpha}_{\mathbf{MLE}}, \tag{12}$$

where $c > 0$ and $0 < d < 1$. The scalar mean squared error (MSE) of the RE and LE are defined, respectively, as follows:

$$\text{MSE}\left(\hat{\alpha}_{\text{Ridge}}\right) = \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + c)^2} + c^2 \sum_{j=1}^{p} \frac{\gamma_j^2}{(\lambda_j + c)^2}, \tag{13}$$

$$\text{MSE}\left(\hat{\alpha}_{\text{Liu}}\right) = \sum_{j=1}^{p} \frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} + (d-1)^2 \sum_{j=1}^{p} \frac{\gamma_j^2}{(\lambda_j + 1)^2}, \tag{14}$$

where λ_j is the eigenvalue of the $Z^T\hat{W}Z$ matrix and γ_j is defined as the j$^{\text{th}}$ element of $\delta^T\hat{\alpha}_{\text{MLE}}$ and δ is the eigenvector of the $Z^T\hat{W}Z$.

According to [3], the new estimator is introduced and derived. Let $F = (f_1, f_2, \ldots, f_p)$ and $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$, respectively, be the matrices of eigenvectors and eigenvalues of the $Z^T\hat{W}Z$ matrix, such that $F^T Z^T \hat{W} Z F = M^T \hat{W} M = \boldsymbol{\Lambda}$, where $M = ZF$. Consequently, the ZIBRM, $\hat{\alpha}_{ZIBRM}$, can be written as

$$\begin{aligned}\hat{\vartheta}_{ZIBRM} &= \boldsymbol{\Lambda}^{-1} M^T \hat{\mathbf{W}} \hat{u} \\ \hat{\alpha}_{ZIBRM} &= F\,\hat{\vartheta}_{ZIBRM}.\end{aligned} \tag{15}$$

Accordingly, the ZIBRM ridge estimator, $\hat{\alpha}_{ZIBRMR}$, is rewritten as

$$\begin{aligned}\hat{\alpha}_{ZIBRMR} &= (\boldsymbol{\Lambda} + C)^{-1} M^T \hat{\mathbf{W}} u \\ &= (\mathbf{I} - CK^{-1})\hat{\alpha}_{ZIBRMR},\end{aligned} \tag{16}$$

where $K = \boldsymbol{\Lambda} + C$ and $C = \text{diag}(c_1, c_2, \ldots, c_p)$; $c_i \geq 0$, $i = 1, 2, \ldots, p$[23-25]. In Eq. (16), the Jackknifing approach was used [24, 26, 27].

In this paper, following the study of Batah, et al. [23], the new estimator (NRZB) is derived. In ZIBRM, the Jackknife estimator (JE) and the modified Jackknife estimator (MJE) are defined as

$$\hat{\alpha}_{JE} = (\mathbf{I} - C^2K^{-2})\hat{\alpha}_{ZIBRM}, \tag{17}$$

$$\hat{\alpha}_{MJE} = (\mathbf{I} - CK^{-1})(\mathbf{I} - C^2K^{-2})\hat{\alpha}_{ZIBRM}. \tag{18}$$

Accordingly, our proposed estimator is an improvement of Eq. (18) by multiplying it with the amount $[(\mathbf{I} - C^3K^{-3})/(\mathbf{I} - C^2K^{-2})]$. This is done in an attempt to obtain a diagonal matrix with tiny diagonal element values, which will lower the shrinkage parameter and improve the final estimator with less bias. The new estimator is defined as

$$\hat{\alpha}_{NRZB} = (\mathbf{I} - CK^{-1})(\mathbf{I} - C^2K^{-2})\frac{(\mathbf{I} - C^3K^{-3})}{(\mathbf{I} - C^2K^{-2})}\hat{\alpha}_{ZIBRM}, \tag{19}$$

4. Bias, variance, and MSE of the new estimator

The MSE of the new estimator can be obtained as

$$\text{MSE}(\hat{\alpha}_{NRZB}) = \text{var}(\hat{\alpha}_{NRZB}) + [\text{bias}(\hat{\alpha}_{NRZB})]^2 \tag{20}$$

According to Eq. (19),

$$\begin{aligned}\text{bias}(\hat{\alpha}_{NRZB}) &= E[\hat{\alpha}_{NRZB}] - \alpha \\ &= (\mathbf{I} - CK^{-1})(\mathbf{I} - C^3K^{-3})E[\hat{\alpha}_{NRZB}] - \alpha \\ &= -C\left[(CK^{-1})^{-1} - (CK^{-1})^{-1}(\mathbf{I} - CK^{-1})\right. \\ &\quad \left. + C^2K^{-2}(\mathbf{I} - CK^{-1})\right]C^{-1}\alpha,\end{aligned} \tag{21}$$

$$\begin{aligned}\operatorname{var}(\hat{\alpha}_{NRZB}) &= (\mathbf{I}-CK^{-1})(\mathbf{I}-C^3K^{-3}) \\ &\quad \times \operatorname{var}(\hat{\alpha}_{NRZB})(\mathbf{I}-C^3K^{-3})^T(\mathbf{I}-CK^{-1})^T \\ &= (\mathbf{I}-CK^{-1})(\mathbf{I}-C^3K^{-3}) \\ &\quad \times \Lambda^{-1}(\mathbf{I}-C^3K^{-3})^T(\mathbf{I}-CK^{-1})^T. \quad (22)\end{aligned}$$

Then,

$$\begin{aligned}&\mathrm{MSE}(\hat{\alpha}_{NRZB}) \\ &= (\mathbf{I}-CK^{-1})(\mathbf{I}-C^3K^{-3})\Lambda^{-1}(\mathbf{I}-C^3K^{-3})^T(\mathbf{I}-CK^{-1})^T \\ &\quad + \left[-C\left[(CK^{-1})^{-1}-(CK^{-1})^{-1}(\mathbf{I}-CK^{-1})\right.\right. \\ &\quad \left.\left. + C^2K^{-2}(\mathbf{I}-CK^{-1})\right]C^{-1}\alpha\right] \qquad (23) \\ &\quad \left[-C\left[(CK^{-1})^{-1}-(CK^{-1})^{-1}(\mathbf{I}-CK^{-1})\right.\right. \\ &\quad \left.\left. + C^2K^{-2}(\mathbf{I}-CK^{-1})\right]C^{-1}\alpha\right]^T\end{aligned}$$

When including NRZB as a shrinkage method in ZIBRM the procedure performs better by using penalty parameters to control coefficient sizes especially in situations of correlated predictor variables. The process of coefficient penalization produces more reliable estimates together with lower variance levels without significant bias increases.

5. Estimating c

The shrinkage parameters, c, which regulate the amount of shrinkage, are the only factors that affect Ridge's efficiency. These two shrinkage parameters can be estimated using a variety of techniques, particularly in linear regression. The following approach is examined in this work for the Zero-Inflated Bell Ridge estimator.

$$C = \frac{1}{\hat{\delta}^2_{\max}}, \; j = 1, 2, \ldots, p, \qquad (24)$$

6. Simulation study

We produced collinear explanatory variables in this part, along with a zero-inflated bell-shaped response variable (y). The explanatory factors are determined in accordance with the subsequent research findings:

$$z_{ij} = \sqrt{(1-\rho^2)}a_{ij} + \rho a_{ip}, \; i = 1, \ldots, n; \; j = 2, \ldots p \qquad (25)$$

where a_{ij} are independent standard uniform pseudo-random numbers, ρ denotes the correlation between the explanatory variables such that $\rho = 0.9, \; 0.95,$ and 0.99, $n = 50, 100,$ and 200, and p = 4 and 8. The n, p and ρ have great influence on the shrinkage estimators, in general.

We assumed that $y_i \sim ZIBell(\alpha_i, \sigma)$, where $\log(\alpha_i) = \alpha_1 z_{i1} + \cdots + \alpha_p z_{ip}$. The percentages of zeros values of the model are chosen such that $\sigma = 0.2, \; 0.4, \; 0.6,$ and 0.8. The experiment was replicated 1000 times and the mean squared error (MSE) was employed to evaluate the estimators' performance.

$$MSE(\alpha^*) = \frac{1}{1000}\sum_{l=1}^{1000}(\alpha_l^* - \alpha)^T(\alpha_l^* - \alpha) \qquad (26)$$

where α_l^* denotes the estimated vector of the true parameter vector α in l^{th} replication.

Under various simulation circumstances, the MSE of the simulated data is given in Tables 1 to 8. Because of multicollinearity, MLE performance is not acceptable. In Table 2, for example, for sample size 100, $\rho = 0.95$, p = 4, and $\sigma = 0.4$, the mean square error (MSE) for MLE is 4.319, whereas the MSEs for the other estimators are negligible in comparison. This is consistent with the literature suggesting that when there is a linear dependency between the explanatory factors, MLE experiences a setback.

Additionally, we noticed that at a given sample size, the MSE of each estimator rises as the degree of multicollinearity does. Additionally, when all other conditions remain constant, the MSE of each estimator falls as sample sizes rise. It is evident that

Table 1. Averaged MSE values of when p = 4 and $\sigma = 0.2$.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	4.228	3.584	3.447	3.258
	0.95	4.514	3.687	3.593	3.337
	0.99	4.832	3.822	3.728	3.507
100	0.90	3.911	3.267	3.13	2.941
	0.95	4.197	3.37	3.276	3.02
	0.99	4.515	3.505	3.411	3.19
250	0.90	3.903	3.259	3.122	2.933
	0.95	4.189	3.362	3.268	3.012
	0.99	4.507	3.497	3.403	3.182

Table 2. Averaged MSE values of when p = 4 and $\sigma = 0.4$.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	4.35	3.706	3.569	3.38
	0.95	4.636	3.809	3.715	3.459
	0.99	4.954	3.944	3.85	3.629
100	0.90	4.033	3.389	3.252	3.063
	0.95	4.319	3.492	3.398	3.142
	0.99	4.637	3.627	3.533	3.312
250	0.90	4.025	3.381	3.244	3.055
	0.95	4.311	3.484	3.39	3.134
	0.99	4.629	3.619	3.525	3.304

Table 3. Averaged MSE values of when p = 4 and σ = 0.6.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	4.529	3.885	3.748	3.559
	0.95	4.815	3.988	3.894	3.638
	0.99	5.133	4.123	4.029	3.808
100	0.90	4.212	3.568	3.431	3.242
	0.95	4.498	3.671	3.577	3.321
	0.99	4.816	3.806	3.712	3.491
250	0.90	4.204	3.56	3.423	3.234
	0.95	4.49	3.663	3.569	3.313
	0.99	4.808	3.798	3.704	3.483

Table 4. Averaged MSE values of when p = 4 and σ = 0.8.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	4.583	3.939	3.802	3.613
	0.95	4.869	4.042	3.948	3.692
	0.99	5.187	4.177	4.083	3.862
100	0.90	4.266	3.622	3.485	3.296
	0.95	4.552	3.725	3.631	3.375
	0.99	4.87	3.86	3.766	3.545
250	0.90	4.258	3.614	3.477	3.288
	0.95	4.544	3.717	3.623	3.367
	0.99	4.862	3.852	3.758	3.537

Table 5. Averaged MSE values of when p = 8 and σ = 0.2.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	4.899	4.255	4.118	3.929
	0.95	5.185	4.358	4.264	4.008
	0.99	5.503	4.493	4.399	4.178
100	0.90	4.582	3.938	3.801	3.612
	0.95	4.868	4.041	3.947	3.691
	0.99	5.186	4.176	4.082	3.861
250	0.90	4.574	3.93	3.793	3.604
	0.95	4.86	4.033	3.939	3.683
	0.99	5.178	4.168	4.074	3.853

Table 6. Averaged MSE values of when p = 8 and σ = 0.4.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	5.021	4.377	4.24	4.051
	0.95	5.307	4.48	4.386	4.13
	0.99	5.625	4.615	4.521	4.3
100	0.90	4.704	4.06	3.923	3.734
	0.95	4.99	4.163	4.069	3.813
	0.99	5.308	4.298	4.204	3.983
250	0.90	4.696	4.052	3.915	3.726
	0.95	4.982	4.155	4.061	3.805
	0.99	5.3	4.29	4.196	3.975

Table 7. Averaged MSE values of when p = 8 and σ = 0.6.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	5.2	4.556	4.419	4.23
	0.95	5.486	4.659	4.565	4.309
	0.99	5.804	4.794	4.7	4.479
100	0.90	4.883	4.239	4.102	3.913
	0.95	5.169	4.342	4.248	3.992
	0.99	5.487	4.477	4.383	4.162
250	0.90	4.875	4.231	4.094	3.905
	0.95	5.161	4.334	4.24	3.984
	0.99	5.479	4.469	4.375	4.154

Table 8. Averaged MSE values of when p = 8 and σ = 0.8.

n	r	MLE	Ridge	Liu	NRZB
50	0.90	5.254	4.61	4.473	4.284
	0.95	5.54	4.713	4.619	4.363
	0.99	5.858	4.848	4.754	4.533
100	0.90	4.937	4.293	4.156	3.967
	0.95	5.223	4.396	4.302	4.046
	0.99	5.541	4.531	4.437	4.216
250	0.90	4.929	4.285	4.148	3.959
	0.95	5.215	4.388	4.294	4.038
	0.99	5.533	4.523	4.429	4.208

the MSE increases with an increase in σ% or the number of explanatory factors. When compared to the ridge and Liu estimators, the suggested estimator, NRZB, performs the best.

7. Applications

A fish dataset was used to forecast how many fish would be taken by 250 groups visiting a state park. The response variable is the number of fish caught, and the predictors are whether or not live bait was used, whether or not the fishermen brought a camper to the park, how many people were in the group, and how many children were in the group (0 if no, 1 if yes- x1, 0 if no, 1 if yes- x2, 0 if no) [28, 29]. There is multicollinearity, as indicated by the condition index of 181.76. The Poisson regression model utilizing the Vuong test does not fit as well as the zero inflated Poisson regression model. The sufficiency test using AIC and log-likelihood in Table 15 further supports this. With a z-value of 2.2357 and a p-value of 0.0000, the over-dispersion test reveals that the data are over-dispersed. This demonstrates why the Poisson regression model fails to provide a good fit to the data. Despite being superior to the Poisson regression model, the zero-inflated Poisson regression model (ZIPRM) performs poorly when compared to other fitted models. Al-Taweel and Algamal [30] recently used the zero-inflated negative binomial regression model (ZNBRM) to model the data. The results are listed in Table 9.

The result in Table 9 shows that the NRZB produced the most preferred estimating using the MSE comparing with Ridge and Liu estimators. Again, the MLE is the worst among them.

8. Conclusion

Because of its simplicity, the Poisson regression model is used to model count data. The Poisson

Table 9. ZIBRM estimates using MLE, Ridge, Liu, and NRZB.

Coef.	MLE	Ridge	Liu	NRZB
α_0	1.565	0.375	1.565	1.321
α_1	–1.632	–0.448	–1.632	–1.508
α_2	–0.338	0.235	–0.338	–0.307
α_3	–1.017	–0.606	–1.017	–1.114
α_4	–0.828	–0.542	–0.828	–0.738
α_5	2.052	1.430	2.052	2.031
ϑ_0	–1.794	–1.294	–1.794	–1.628
ϑ_1	–0.744	–0.770	–0.744	–0.538
ϑ_2	1.176	0.798	1.176	1.014
ϑ_3	0.464	0.461	0.464	0.422
ϑ_4	0.880	0.841	0.880	0.738
ϑ_5	–1.175	–1.141	–1.175	–1.108
c		1.808	1.020	0.871
MSE	9.996	2.512	2.496	**2.105**

regression model, however, clearly yields a poor match for count data with over-dispersion. Count data modeling over-dispersion is efficiently accounted for by alternative models including the Bell regression model and others. Additionally, this work has demonstrated how some of these models are affected by excess zeros. In this study, a new ridge estimator, NRZB, was proposed as alternatives to the maximum likelihood estimator, ridge, and Liu estimators which has limitations when there is linear dependency among the X's. A simulated study and an empirical application are used to highlight the methods developed in this work. In the application research, we demonstrated that the ZIBRM offers a better match when there is excess zero dispersion compared to some existing models. Additionally, both theoretically and practically, the proposed estimator in this work beat the maximum likelihood method. NRZB, while effective in addressing multicollinearity, does have limitations when it comes to outliers. Future work can be focusing on the extension of the NRZB to other generalized linear model family.

References

1. M. Amin, M. Qasim, M. Amanullah, and S. J. S. p. Afzal, "Performance of some ridge estimators for the gamma regression model," vol. 61, pp. 997–1026, 2020.
2. A. A. Hamad, Z. Y. J. I. J. o. N. A. Algamal, and Applications, "Jackknifing KL estimator in generalized linear models," vol. 12, no. Special Issue, pp. 2093–2104, 2021.
3. N. K. Rashad and Z. Y. Algamal, "A new ridge estimator for the poisson regression model," *Iranian Journal of Science and Technology, Transactions A: Science*, vol. 43, no. 6, pp. 2921–2928, 2019, doi: 10.1007/s40995-019-00769-3.
4. Z. Yahya Algamal, "Performance of ridge estimator in inverse gaussian regression model," *Communications in Statistics-Theory and Methods*, vol. 48, no. 15, pp. 3836–3849, 2019.
5. Z. Y. Algamal, "Diagnostic in poisson regression models," *Electronic Journal of Applied Statistical Analysis*, vol. 5, no. 2, pp. 178–186, 2012.
6. Z. Y. Algamal and M. M. Alanaz, "Proposed methods in estimating the ridge regression parameter in poisson regression model," *Electronic Journal of Applied Statistical Analysis*, vol. 11, no. 2, pp. 506–515, 2018.
7. F. Castellares, S. L. Ferrari, and A. J. J. A. M. M. Lemonte, "On the bell distribution and its associated regression model for count data," vol. 56, pp. 172–185, 2018.
8. M. Amin, M. N. Akram, and A. Majid, "On the estimation of bell regression model using ridge estimator," *Communications in Statistics - Simulation and Computation*, pp. 1–14, 2021, doi: 10.1080/03610918.2020.1870694.
9. A. Majid, M. Amin, and M. N. Akram, "On the liu estimation of bell regression model in the presence of multicollinearity," *Journal of Statistical Computation and Simulation*, pp. 1–21, 2021, doi: 10.1080/00949655.2021.1955886.
10. Z. Y. Algamal, A. F. Lukman, M. R. Abonazel, and F. A. Awwad, "Performance of the ridge and liu estimators in the zero-inflated bell regression model," *Journal of Mathematics*, vol. 2022, no. 1, p. 9503460, 2022.
11. Z. Algamal, A. Lukman, B. K. Golam, and A. Taofik, "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.
12. S. Seifollahi, 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, doi: 10.1007/s42519-024-00415-1.
13. Y. M. Bulut, A. F. Lukman, M. Işılar, E. T. Adewuyi, and Z. Y. Algamal, "Modified ridge estimator in the bell regression model," *Journal of Inverse and Ill-posed Problems*, vol. 32, no. 6, pp. 1081–1091, 2024, doi: 10.1515/jiip-2022-0069.
14. S. Seifollahi, H. Bevrani, and Z. Y. Algamal, "Improved estimators in bell regression model with application," *Journal of Statistical Computation and Simulation*, vol. 94, no. 12, pp. 2710–2726, 2024, doi: 10.1080/00949655.2024.2350553.
15. E. Ertan, Z. Y. Algamal, A. Erkoç, and K. U. Akay, "A new improvement liu-type estimator for the bell regression model," *Communications in Statistics - Simulation and Computation*, pp. 1–12, 2023, doi: 10.1080/03610918.2023.2252624.
16. A. M. Salih, Z. Algamal, and M. A. Khaleel, "A new ridge-type estimator for the gamma regression model," *Iraqi Journal For Computer Science and Mathematics*, vol. 5, no. 1, pp. 85–98, 2024.
17. Z. Y. Algamal, "A modified ridge estimator in cox regression model," *Communications in Statistics-Simulation and Computation*, pp. 1–8, 2025.
18. Y. Asar, S. E. Ahmed, and B. Yüzbaşı, "Efficient and improved estimation strategy in zero-inflated poisson regression models," *Cham, 2019: Springer International Publishing, in Proceedings of the Twelfth International Conference on Management Science and Engineering Management*, pp. 329–342.
19. B. M. G. Kibria, K. Månsson, and G. Shukur, "Some ridge regression estimators for the zero-inflated poisson model," *Journal of Applied Statistics*, vol. 40, no. 4, pp. 721–735, 2013, doi: 10.1080/02664763.2012.752448.
20. Z. Algamal, A. Lukman, B. K. Golam, A. J. I. J. F. C. S. Taofik, and Mathematics, "Modified jackknifed ridge estimator in bell regression model: Theory, simulation and applications," vol. 4, no. 1, pp. 146–154, 2023.
21. Z. Y. Algamal and N. Hammood, "A new jackknifing ridge estimator for logistic regression model," *Pakistan Journal of Statistics and Operation Research* pp. 955–961, 2022.

22. N. N. Alobaidi, R. E. Shamany, and Z. Y. Algamal, "A new ridge estimator for the negative binomial regression model," *Thailand Statistician*, vol. 19, no. 1, pp. 116–125, 2021.
23. F. S. M. Batah, T. V. Ramanathan, and S. D. Gore, "The efficiency of modefied jackknife and ridge type regression estimators - a comparison," *Surveys in Mathematics and its Applications*, vol. 3, pp. 111–122, 2008.
24. M. Khurana, Y. P. Chaubey, and S. Chandra, "Jackknifing the ridge regression estimator: A revisit," *Communications in Statistics-Theory and Methods*, vol. 43, no. 24, pp. 5249–5262, 2014.
25. M. R. Özkale, "A jackknifed ridge estimator in the linear regression model with heteroscedastic or correlated errors," *Statistics & Probability Letters*, vol. 78, no. 18, pp. 3159–3169, 2008, doi: 10.1016/j.spl.2008.05.039.
26. H. Nyquist, "Applications of the jackknife procedure in ridge regression," *Computational Statistics & Data Analysis*, vol. 6, no. 2, pp. 177–183, 1988.
27. B. Singh, Y. Chaubey, and T. Dwivedi, "An almost unbiased ridge estimator," *Sankhyā: The Indian Journal of Statistics, Series B*, vol. 13, pp. 342–346, 1986.
28. Z. Y. Algamal, M. H. Lee, and A. M. Al-Fakih, "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, pp. 50–57, 2016.
29. Z. Algamal and M. Lee, "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, pp. 75–90, 2017.
30. Y. Al-Taweel and Z. Algamal, "Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model," *Periodicals of Engineering Natural Sciences*, vol. 8, no. 1, pp. 248–255, 2020.