Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 4 (2022) 713

Journal of the Nigerian Society of Physical Sciences

Combating the Multicollinearity in Bell Regression Model: Simulation and Application

G. A. Shewa[a,*], F. I. Ugwuowo[b]

[a] *Department of Mathematical Sciences, Taraba State University, Jalingo/Taraba, Nigeria*
[b] *Department of Statistics, University of Nigeria, Nsukka/Enugu, Nigeria*

Abstract

Poisson regression model has been popularly used to model count data. However, over-dispersion is a threat to the performance of the Poisson regression model. The Bell Regression Model (BRM) is an alternative means of modelling count data with over-dispersion. Conventionally, the parameters in BRM is popularly estimated using the Method of Maximum Likelihood (MML). Multicollinearity posed challenge on the efficiency of MML. In this study, we developed a new estimator to overcome the problem of multicollinearity. The theoretical, simulation and application results were in favor of this new method.

DOI:10.46481/jnsps.2022.713

Keywords: Bell regression, Liu, Multicollinearity, Poisson regression, Ridge

Article History :
Received: 17 March 2022
Received in revised form: 07 June 2022
Accepted for publication: 08 June 2022
Published: 15 August 2022

Communicated by: T. Latunde

1. Introduction

Regression modeling is crucial in describing the outcome (response) variable of interest as a function of predictor variable(s). The outcome variable usually assumed to follow a normal distribution in the linear regression model. However, in practice this may not hold. The Generalized Linear Model (GLM) is employed when the response variable fails to follow a normal distribution. GLM is the generalization of the linear regression model that allow the linear model to be related to the response variable via a *link function*. Examples include the Poisson regression, negative binomial, the Bell regression, the Beta regression models and others. It becomes inappropriate to adopt the linear regression model when the response variable is a count data. Example of count data includes the number of thunderstorms occurrences, the number of accidents, the number of insurance claims, the number of species in a habitat among others.

The Poisson regression model is popularly employed to model the count data. However, the major drawback of the model is that the model restricts the variance to be equal to the mean and when the variance exceeds the mean [1]. The bell regression model was introduced as an alternative to the Poisson regression model to model count data with over-dispersion [1]. The properties of the bell regression model were discussed in detail by [1]. The parameters of the model is determined using the Method of Maximum Likelihood (MML), but the efficiency of MML suffers drawback in the presence of multicollinearity. The Ridge and the Liu estimators were proposed for the

*Corresponding author tel. no:
Email address: gladysshewa@yahoo.com (G. A. Shewa)

parameter estimation of the Bell regression model with multicollinearity [2,3].

The main objective given in this study is to propose a new estimator to account for multicollinearity in the Bell regression model, derive its property. We illustrate the proposed estimator using a real life data and compare with the popular Poisson regression model. The rest of this article is organized as follows. In Section 2, we introduce the Bell regression model and the parameter estimation. Also, we discuss the new method of estimation and its property. The simulation and the practical illustration are in Section 3 and 4, respectively. The concluding remark is given in the last section.

2. Existing Estimators in Bell Regression Model

Assuming the probability distribution of the response variable y_i is as follows:

$$f(y) = \frac{\theta^y e^{1-e^\theta} B_y}{y!}, \quad y = 0, 1, 2, \ldots, \tag{1}$$

where $\theta > 0$ and $B_y = (1/e)\sum_{d=0}^{\infty}(d^y/d!)$ denotes the Bell numbers [1,4,5]. The Bell distribution in (1) has the following properties:

$$E(y) = \theta e^\theta, \tag{2}$$

$$Var(y) = \theta(1+\theta)e^\theta, \tag{3}$$

The model is expressed as a function of the mean response. Assume there exists a function, $\varphi = \theta e^\theta$ and $\theta = W_o(\varphi)$, where W_o represent the Lambert function [1]. Therefore, equation (1) can be re-parameterize as follows:

$$f(y) = \frac{e^{1-e^{W_o(\varphi)}} W_o(\varphi)^y B_y}{y!}, \quad y = 0, 1, 2, \ldots, \tag{4}$$

Where $\varphi > 0$ is the mean response. Consequently,

$$E(y) = \varphi, \tag{5}$$

$$Var(y) = \varphi(1 + W_o(\varphi)), \tag{6}$$

The Probability Mass Function (pmf) in equation (4) is an example of the one-parameter exponential family. The Bell distribution is fit for modelling over-dispersed data because $Var(y) > E(y)$. Assume y_i follows a Bell distribution with mean φ_i, $y_i \sim Bell(W_o(\varphi_i))$, such that

$$\mathrm{g}(\varphi_i) = \eta_i = x_i^T\beta, \; i = 1, 2, \ldots n, \tag{7}$$

where $\beta = \left(\beta_1, \beta_2, \ldots, \beta_p\right)^T$ is the vector of the regression parameters, η_i is the linear predictor, and $x_i^T = \left(x_{i1}, x_{i2}, \ldots, x_{ip}\right)$ denotes the p-known predictors. The Bell regression model (BRM) can be modeled by assuming that $\varphi_i = e^{x_i^T\beta}e^{e^{x_i^T\beta}}$ and $ln(\varphi_i) = x_i^T\beta e^{x_i^T\beta}$ as $y_i \sim Bell(ln(\varphi_i))$. The log-likelihood function becomes

$$l(\beta, \varphi_i) = \sum_{i=1}^{n} y_i log\left(e^{x_i^T\beta}e^{e^{x_i^T\beta}}\right) + \sum_{i=1}^{n}\left(1 - e^{e^{x_i^T\beta}e^{e^{x_i^T\beta}}}\right) + logB_y - log\left(\prod_{i=1}^{n} y_i!\right) \tag{8}$$

Thus, the Method of Maximum Likelihood (MML) is obtained by equating the first derivative of equation (8) to zero. The first derivative of equation (8) cannot be solved analytically since it is nonlinear in β. So, β is obtained iteratively using the Fisher-scoring algorithm [6] defined as follows:

$$\beta^{(r+1)} = \beta^{(r)} + I^{-1}\beta^{(r)}S(\beta^{(r)}) \tag{9}$$

where $I^{-1}(\beta) = \left(-E\left(\partial^2 l(\beta, \varphi_i)/\partial\beta\partial\beta^T\right)\right)^{-1}$.

Consequently, the MML of β is defined as

$$\widehat{\beta}_{MML} = H^{-1}X^T\hat{W}\hat{u} \tag{10}$$

where

$H = X^T\hat{W}X$ and $\hat{u} = log(\widehat{\varphi_i}) + \frac{y_i - \widehat{\varphi_i}}{\sqrt{var(\widehat{\varphi_i})}}$, and

$$\hat{W} = \mathrm{diag}\left[(\partial\varphi_i/\partial\eta_i)^2/V(y_i)\right].$$

The asymptotic covariance matrix is given by:

$$\mathrm{Cov}(\widehat{\beta}_{MML}) = \left(X^T\hat{W}X\right)^{-1}. \tag{11}$$

Multicollinearity threatens the efficiency of the MML. Multicollinearity occurs when the predictors are correlated and makes the MML estimate unstable. Also, multicollinearity inflates the covariance matrix of MML [7-9].

[10] developed the ridge estimator for the linear regression model while [2] developed the Bell ridge regression model and defined it as follows:

$$\widehat{\beta}_{k-B} = (H + kI)^{-1}X^T\hat{W}\hat{u}, \tag{12}$$

where the tuning parameter k>0. The bias, variance and Matrix Mean Squared Error (MSEM) of Bell ridge estimator are shown as follows:

$$\mathrm{Bias}\left(\widehat{\upbeta}_{\mathrm{k-B}}\right) = -kQ(B + k\mathrm{I})^{-1}\gamma \tag{13}$$

$$\mathrm{Variance}\left(\widehat{\upbeta}_{\mathrm{k-B}}\right) = \mathrm{Q}(H + k\mathrm{I})^{-1}\mathrm{H}(\mathrm{H} + k\mathrm{I})^{-1}Q^T, \tag{14}$$

$$\mathrm{MSEM}\left(\widehat{\upbeta}_{\mathrm{k-B}}\right) = \mathrm{Q}(H + k\mathrm{I})^{-1}\mathrm{H}(\mathrm{H} + k\mathrm{I})^{-1}Q^T + \mathrm{k}^2(H + k\mathrm{I})^{-2}\gamma\gamma^T, \tag{15}$$

where Q denotes the eigenvectors of the $\mathrm{X}^T\widehat{\mathrm{W}}\mathrm{X}$ matrix, and $\gamma = Q^T\beta$.

[11] developed the Liu estimator for linear regression model while [3] developed the Bell Liu estimator and defined as follows:

$$\widehat{\beta}_{d-B} = (H + I)^{-1}(H + dI)\widehat{\beta}_{MML}. \tag{16}$$

where the tuning parameter $d > 0$.

The bias, variance and Matrix Mean Squared Error (MSEM) of Bell Liu estimator are as follows:

$$\text{Bias}\left(\widehat{\beta}_{\text{d-B}}\right) = -(1-\text{d})\text{Q}\,(H+\text{I})^{-1}\gamma \tag{17}$$

$$\text{Variance}\left(\widehat{\beta}_{\text{d-B}}\right) = \text{Q}\,(H+\text{I})^{-1}\,(H+d\text{I})\,H^{-1}\,(H+d\text{I})\,(\text{H}+\text{I})^{-1}Q^{T}, \tag{18}$$

$$\text{MSEM}\left(\widehat{\beta}_{\text{d-B}}\right) = \text{Q}\,(H+\text{I})^{-1}\,(H+d\text{I})\,H^{-1}\,(H+d\text{I})\,(\text{H}+\text{I})^{-1}Q^{T} + (1-\text{d})^{2}(H+\text{I})^{-2}\beta\beta^{T}, \tag{19}$$

where Q denotes the eigenvectors of the $\text{X}^T\widehat{\text{W}}\text{X}$ matrix.

Let $Q^T X^T \hat{W} X Q = E = diag\left(e_1, ..., e_p\right), e_1 \geq e_2 \geq ... \geq e_p$, E is the matrix of eigenvalues of $X^T\hat{W}X$and Q is the matrix whose columns are the eigenvectors of $X^T\hat{W}X$.Thus, the canonical model can be expressed in terms of $M = XQ, \gamma = Q^T\beta$and $M^T\hat{W}M = E$. Consequently, the MML in equation (10) can be re-written as

$$\widehat{\gamma}_{MML} = E^{-1}M^T\hat{W}\hat{u}, \tag{20}$$

$$Cov(\widehat{\gamma}_{MML}) = E^{-1}. \tag{21}$$

Thus, the Scalar Mean Squared Error (SMSE) is as follows:

$$SMSE(\widehat{\gamma}_{MML}) = \sum_{j=1}^{p} e_j^{-1}, \tag{22}$$

The ridge estimator in canonical form is as follows:

$$\widehat{\gamma}_{k-B} = (E+kI)^{-1}M^T\hat{W}\hat{u}, \tag{23}$$

The MSEM and SMSE of the ridge estimator in canonical form is calculated as

$$\text{MSEM}\left(\widehat{\gamma}_{\text{k-B}}\right) = \text{Q}\,(E+k\text{I})^{-1}E(\text{E}+k\text{I})^{-1}Q^{T} + \text{k}^{2}(E+k\text{I})^{-2}\gamma\gamma^{T}, \tag{24}$$

$$SMSE\left(\widehat{\gamma}_{k-B}\right) = \sum_{j=1}^{p}\frac{e_j}{\left(e_j+k\right)^2} + k^2\sum_{j=1}^{p}\frac{\gamma_j^2}{\left(e_j+k\right)^2} \tag{25}$$

The Liu estimator and its MSEM and SMSE in canonical form are given by:

$$\widehat{\gamma}_{d-B} = (E+I)^{-1}(E+dI)\widehat{\gamma}_{MML}. \tag{26}$$

$$\text{MSEM}\left(\widehat{\gamma}_{\text{d-B}}\right) = \text{Q}\,(E+\text{I})^{-1}\,(E+d\text{I})\,E^{-1}\,(E+d\text{I})\,(\text{E}+\text{I})^{-1}Q^{T} + (1-\text{d})^{2}(E+\text{I})^{-2}\gamma\gamma^{T}, \tag{27}$$

$$SMSE\left(\widehat{\gamma}_{d-B}\right) = \sum_{j=1}^{p}\frac{\left(e_j+d\right)^2}{e_j\left(e_j+1\right)^2} + (1-d)^2\sum_{j=1}^{p}\frac{\gamma_j^2}{\left(e_j+1\right)^2} \tag{28}$$

2.1. The Proposed Estimator

The Kibria-Lukman estimator for the linear regression model is defined as follows:

$$\widehat{\gamma}_{KL} = \left(X^TX+kI\right)^{-1}(X^TX-kI)\widehat{\gamma}_{MML}. \tag{29}$$

Hence, the Kibria-Lukman estimator for the Bell regression model will be as follows:

$$\widehat{\gamma}_{KL} = (E+kI)^{-1}(E-kI)\widehat{\gamma}_{MML}. \tag{30}$$

$$\text{MSEM}\left(\widehat{\gamma}_{\text{kl-B}}\right) = \text{Q}\,(E+kI)^{-1}\,(E-k\text{I})\,E^{-1}\,(E-k\text{I})\,(\text{E}+kI)^{-1}Q^{T} + (2\text{k})^{2}(E+k\text{I})^{-2}\gamma\gamma^{T}, \tag{31}$$

$$SMSE\left(\widehat{\gamma}_{kl-B}\right) = \sum_{j=1}^{p}\frac{\left(e_j-k\right)^2}{e_j\left(e_j+k\right)^2} + (2k)^2\sum_{j=1}^{p}\frac{\gamma_j^2}{\left(e_j+k\right)^2} \tag{32}$$

Also, [12] developed the Modified Ridge Estimator for the linear regression model which is given by:

$$\widehat{\gamma}_{MRT} = \left(X^TX+k(1+d)I\right)^{-1}X^Ty \tag{33}$$

The proposed estimator in this study is motivated by replacing $\widehat{\gamma}_{MML}$ in equation (30) with $\widehat{\gamma}_{MRT}$. Hence, the Modified Kibria-Lukman estimator can be defined as follows:

$$\widehat{\gamma}_{kd-B} = (X^TX-kI)\left(X^TX+kI\right)^{-1}\left(X^TX+k(1+d)I\right)^{-1}X^T\hat{W}\hat{u}, \quad k=0. \tag{34}$$

The proposed can be written in canonical form as follows:

$$\widehat{\gamma}_{kd-B} = (E-kI)\,(E+kI)^{-1}(E+k(1+d)I)^{-1}M^T\hat{W}\hat{u}, \quad k=0. \tag{35}$$

The statistical properties of $\widehat{\gamma}_{kd-B}$ are as follows:

$$Bias\left(\widehat{\gamma}_{kd-B}\right) = -k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}\gamma \tag{36}$$

$$Var\left(\widehat{\gamma}_{kd-B}\right) = (E-kI)^2\,(E+kI)^{-2}(E+k(1+d)I)^{-2} \tag{37}$$

$$MSEM\left(\widehat{\gamma}_{kd-B}\right) = (E+kI)^{-2}(E-kI)^2\,(E+k(1+d)I)^{-2} + k^2((3+d)\,E+k(1+d))^2\,(E+kI)^{-2}(E+k(1+d)I)^{-2}\gamma^T\gamma \tag{38}$$

$$SMSE\left(\widehat{\gamma}_{kd-B}\right) = \sum_{j=1}^{p}\frac{\left(\lambda_j-k\right)^2}{\left(\lambda_j+k\right)^2\left(\lambda_j+k(1+d)\right)^2} + k^2\sum_{j=1}^{p}\frac{\left(3\lambda_j+\lambda_j d+k(1+d)\right)^2\gamma_j^2}{\left(\lambda_j+k\right)^2\left(\lambda_j+k(1+d)\right)^2} \tag{39}$$

2.2. Theoretical Comparison based on MSEM and MSE

Lemma 2.1. Given a positive definite (p.d) matrix M*, and θ^* be some vector, then $M^* - \theta^*\theta^{*T} \geq 0$ if and only if $\theta^{*'}M^{*-1}\theta^* \leq 1$[13].

Lemma 2.2. $\widehat{\theta}_1 = C_1y$ and $\widehat{\theta}_2 = C_2y$ are two estimators of θ with covariance matrix $\mathrm{Cov}(\widehat{\theta}_1)$ and $\mathrm{Cov}\left(\widehat{\theta}_2\right)$, respectively. Suppose that $\mathrm{Cov}(\widehat{\theta}_1) > Cov\left(\widehat{\theta}_2\right)$, and the bias $f_i = (C_iX - I)\theta, \quad i = 1,\ 2$ then, $\mathrm{MSEM}\left(\widehat{\theta}_1\right) - \mathrm{MSEM}\left(\widehat{\theta}_2\right) > 0$ if and only if $f_2'^T\left[\phi V + f_1 f_1^T{}_1\right]^{-1} f_2 < 1$ where $MSEM\left(\hat{\theta}_i\right) = Cov\left(\hat{\theta}_i\right) + f_i f_i^T$ [14].

Theorem 1. Under the Bell regression model, if $k>0$ and $d>0$, then the proposed estimator $\widehat{\gamma}_{\mathrm{kd-B}}$ is preferred to $\widehat{\gamma}_{\mathrm{k-B}}$ if and only if,

$$\gamma^T b\left[(E+k\mathrm{I})^{-1}E(E+k\mathrm{I})^{-1} + \ k^2(E+k\mathrm{I})^{-2}\gamma\gamma^T - \ (E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}\ \right]^{-1} b^T\gamma < 1$$

where $b = -k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}$.

Proof: We show that the difference in the bias is positive definite.

$$\mathrm{Bias}(\widehat{\gamma}_{\mathrm{k-B}}) - \mathrm{Bias}\left(\widehat{\gamma}_{\mathrm{kd-B}}\right) =$$

$$-k(E+k\mathrm{I})^{-1}\gamma + k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}\gamma = -k\mathrm{diag}\left\{\frac{1}{e_j+k} - \frac{3e_j+de_j+k(1+d)}{(e_j+k)(e_j+k(1+d))}\right\}_{j=1}^{p}\gamma \quad (40)$$

$$-k\left[\left(e_j+k(1+d)\right) - \left(3e_j+de_j+k(1+d)\right)\right]\gamma = k(2e_j+e_jd)\gamma > 0.$$

Consequently, Bias $(\widehat{\gamma}_{\mathrm{k-B}})$-Bias $(\widehat{\gamma}_{\mathrm{kd-B}}) > 0$.

We show that the variance difference of the two estimators is Positive Definite (pd).

$$\mathrm{Var}(\widehat{\gamma}_{\mathrm{k-B}}) - \mathrm{Var}\left(\widehat{\gamma}_{\mathrm{kd-B}}\right) = (E+k\mathrm{I})^{-2}E(E+k\mathrm{I})^{-2} - (E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}$$

$$= \mathrm{diag}\left\{\frac{e_j}{(e_j+k)^2} - \frac{(e_j-k)^2}{(e_j+k)^2(e_j+k(1+d))^2}\right\}_{j=1}^{p} \quad (41)$$

$(E+k\mathrm{I})^{-2}E(E+k\mathrm{I})^{-2}$-$(E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}$ is pd since $e_j\left(e_j+k(1+d)\right)^2 - \left(e_j-k\right)^2 > 0$ for $k,d>0$.

Consequently, the proposed estimator is preferred.

Theorem 2. Under the bell regression model, if k>0 and d>0, then the proposed estimator $\widehat{\gamma}_{\mathrm{kd-B}}$ is preferred to $\widehat{\gamma}_{\mathrm{d-B}}$ if and only if

$$\gamma^T b\left[(E+\mathrm{I})^{-2}E^{-1}(E+d\mathrm{I})^2 + \ (1-\mathrm{d})^2(E+\mathrm{I})^{-2}\gamma\gamma^T - \ (E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}\ \right]^{-1} b^T\gamma < 1$$

where $b = -k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}$.

Proof: We show that the difference in the bias is positive definite.

$$\mathrm{Bias}(\widehat{\gamma}_{\mathrm{k-B}})\text{-}\mathrm{Bias}\left(\widehat{\gamma}_{\mathrm{kd-B}}\right) = -(1+d)(E+\mathrm{I})^{-1}\gamma + k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}\gamma$$

$$= -k\mathrm{diag}\left\{\frac{1+d}{e_j+1} - \frac{3e_j+de_j+k(1+d)}{(e_j+k)(e_j+k(1+d))}\right\}_{j=1}^{p}\gamma \quad (42)$$

$-(1+d)(E+\mathrm{I})^{-1}\gamma + k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}\gamma = k(1+d)\left[k+e_jd+dk-1\right] + e_j(k+dk-3-d) > 0$ for k,d>0

Consequently, Bias $(\widehat{\gamma}_{\mathrm{d-B}})$-Bias $(\widehat{\gamma}_{\mathrm{kd-B}}) > 0$.

We show that the variance difference of the two estimators is positive definite (pd).

$\mathrm{Var}(\widehat{\gamma}_{\mathrm{k-B}})$-$\mathrm{Var}\left(\widehat{\gamma}_{\mathrm{kd-B}}\right)$=$(E+\mathrm{I})^{-2}(E+d\mathrm{I})^2$-$(E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}$

$$= \mathrm{diag}\left\{\frac{\left(e_j+d\right)^2}{e_j\left(e_j+1\right)^2} - \frac{(e_j-k)^2}{(e_j+k)^2(e_j+k(1+d))^2}\right\}_{j=1}^{p} \quad (43)$$

$(E+k\mathrm{I})^{-2}E(E+k\mathrm{I})^{-2}$-$(E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}$ is pd since $\left(e_j+d\right)^2\left(e_j+k(1+d)\right)^2(e_j+k)^2 > e_j\left(e_j+1\right)^2\left(e_j-k\right)^2 > 0$ for $k,d>0$.

Consequently, the proposed estimator is preferred.

Theorem 3. Under the bell regression model, if k>0 and d>0, then the proposed estimator $\widehat{\gamma}_{\mathrm{kd-B}}$ is preferred to $\widehat{\gamma}_{\mathrm{kl-B}}$ if and only if,

$$\gamma^T b\left[(E+kI)^{-1}(E-k\mathrm{I})\,E^{-1}(E-k\mathrm{I})\,(\mathrm{E}+kI)^{-1}+(2\mathrm{k})^2(E+k\mathrm{I})^{-2}\gamma\gamma^T-\ (E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}\right]^{-1}b^T\gamma<1$$

$$\text{where } b=-k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}.$$

Proof: We show that the difference in the bias is positive definite.

$$\mathrm{Bias}(\widehat{\gamma}_{\mathrm{kl-B}})-\mathrm{Bias}\,(\widehat{\gamma}_{\mathrm{kd-B}})=-2k(E+\mathrm{I})^{-1}\gamma+k\left((3E+dE)+k(1+d)\right)(E+kI)^{-1}(E+k(1+d)I)^{-1}\gamma$$

$$=-k\mathrm{diag}\left\{\frac{2}{e_j+k}-\frac{3e_j+de_j+k\,(1+d)}{(e_j+k)(e_j+k\,(1+d))}\right\}_{j=1}^{p}\gamma \quad (44)$$

$-2k(E+\mathrm{kI})^{-1}\gamma+k\left((3E+dE)+k\,(1+d)\right)(E+kI)^{-1}(E+k\,(1+d)\,I)^{-1}\gamma>0$ for k,d>0

Consequently, Bias $(\widehat{\gamma}_{\mathrm{kl-B}})$-Bias $(\widehat{\gamma}_{\mathrm{kd-B}})>0$.

We show that the variance difference of the two estimators is positive definite (pd).

Var$(\widehat{\gamma}_{\mathrm{k-B}})$-Var$(\widehat{\gamma}_{\mathrm{kd-B}})$=$(E-k\mathrm{I})^2E^{-1}(E+k\mathrm{I})^{-2}$-$(E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}$

$$=\mathrm{diag}\left\{\frac{\left(e_j-k\right)^2}{e_j\left(e_j+k\right)^2}-\frac{(e_j-k)^2}{(e_j+k)^2(e_j+k(1+d))^2}\right\}_{j=1}^{p} \quad (45)$$

$(E-k\mathrm{I})^2E^{-1}(E+k\mathrm{I})^{-2}$-$(E+k\mathrm{I})^{-2}(E-kI)^2(E+k(1+d)\mathrm{I})^{-2}$ is pd since $\left(e_j-k\right)^2\left(e_j+k\,(1+d)\right)^2(e_j+k)^2>e_j\left(e_j+k\right)^2\left(e_j-k\right)^2>0$ for k,d>0. Consequently, the proposed estimator is preferred.

2.3. Estimation of Shrinkage Parameters k and d

The shrinkage parameter for the proposed is obtained by differentiating its mean squared error. We adopted the Maple software for the simplification and simplest form of the calculated shrinkage parameter. Hence, the ridge parameter k is defined as

$$\hat{k}=\left(\prod_{j=1}^{p}\frac{\left(d+1+\sqrt{2d^2+6d+4}\right)e_j}{1+d}\right)^{1/p} \quad (46)$$

$$\hat{d}=min\left(\frac{\widehat{\gamma}_j^2}{\frac{1}{e_j}+\widehat{\gamma}_j^2}\right)[15] \quad (47)$$

where e_j is the eigenvalue of $M'\hat{W}M$. $\widehat{\mathrm{k}}$ in (46) produced optimum performance for the proposed and Kibria-Lukman estimator. Also, d in (47) is adopted for the proposed and the Liu estimator.

The shrinkage parameter for the ridge estimator is defined as:

$$\hat{k}=\frac{1}{\max(\widehat{\gamma}_j^2)} \quad (48)$$

3. Simulation Study

In this study, we simulate using R software with the help of bellreg-package [1,16]. The predictors are generated in accordance to [7,8,9,17,18,19,20,21,22,23,24,25]:

$$x_{ij}=\sqrt{(1-\rho^2)}m_{ij}+\rho m_{i(j+1)},\ i=1,\ldots,n;\ j=1,\ldots p, \quad (49)$$

where m_{ij} are independent standard normal pseudo-random numbers and ρ^2 denotes the correlation between the explanatory variables such that $\rho=0.8,\ 0.9,\ 0.99$ and 0.999. It is assumed that $y_i\sim bell\,(W_o\,(\mu_i))$, such that

$$\log(\mu_i)=\eta_i=\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_px_{ip} \quad (50)$$

The sample sizes n are 50, 100, 200 and 500 while p is taken to be 3, 8 and 12. We choose the true regression parameters β such that $\sum_{i=1}^{p}\hat{\beta}_i^2=1$ [26].

Table 1. Simulated result in terms of MSE when p= 4

coef	50	100	200	500
		0.8		
$\widehat{\gamma}_{\text{MML}}$	5.1368	1.9656	1.1792	1.0398
$\widehat{\gamma}_{\text{k-B}}$	4.3841	1.8570	1.1404	1.0386
$\widehat{\gamma}_{\text{d-B}}$	3.3761	1.9492	1.1699	1.0392
$\widehat{\gamma}_{\text{kl-B}}$	1.9286	1.5390	1.1393	1.0234
$\widehat{\gamma}_{\text{kd-B}}$	1.4073	1.3149	1.2070	1.0183
		0.9		
$\widehat{\gamma}_{\text{MML}}$	5.6531	1.9926	1.3388	1.1347
$\widehat{\gamma}_{\text{k-B}}$	4.8511	1.9512	1.3192	1.1335
$\widehat{\gamma}_{\text{d-B}}$	4.7792	1.9629	1.3342	1.1338
$\widehat{\gamma}_{\text{kl-B}}$	3.8790	1.7835	1.3128	1.1234
$\widehat{\gamma}_{\text{kd-B}}$	2.5191	1.5851	1.3093	1.0191
		0.99		
$\widehat{\gamma}_{\text{MML}}$	6.8980	2.9614	1.8329	1.2415
$\widehat{\gamma}_{\text{k-B}}$	5.5356	2.2753	1.3614	1.2360
$\widehat{\gamma}_{\text{d-B}}$	5.2118	2.4122	1.4868	1.2081
$\widehat{\gamma}_{\text{kl-B}}$	4.2356	2.1345	1.3545	1.2076
$\widehat{\gamma}_{\text{kd-B}}$	3.4242	1.6709	1.4037	1.2001
		0.999		
$\widehat{\gamma}_{\text{MML}}$	15.4196	8.3391	8.2790	3.4418
$\widehat{\gamma}_{\text{k-B}}$	6.1161	3.7683	7.2700	2.0326
$\widehat{\gamma}_{\text{d-B}}$	5.6982	2.9212	4.2932	1.7886
$\widehat{\gamma}_{\text{kl-B}}$	4.8923	2.7634	2.4512	1.5531
$\widehat{\gamma}_{\text{kd-B}}$	3.4986	2.0303	1.6574	1.2747

The simulation study was conducted by adopting the RStudio programming language. The experiment was replicated 1000 times and the Mean Squared Error (MSE) was employed to evaluate the performance of the estimators'.

$$MSE(\beta^*) = \frac{1}{1000}\sum_{j=1}^{1000}(\beta^*_{ij} - \beta_i)^T(\beta^*_{ij} - \beta_i) \tag{51}$$

where β^*_{ij} is the estimator and β_i is the parameter. The estimator with the minimum MSE is preferred to the estimator with maximum MSE. The simulation result is presented in Tables 1-3.

The following observations were made from Tables 1-3. The Mean Squared Errors for each of the estimators were computed at different specifications. The Method of Maximum Likelihood has the least performance in this study. This is in line with the literature as expected. We observed that its performance drops as the level of multicollinearity changes. The new estimator produces a better performance in terms of minimum Mean Squared Error than the ridge and the Liu estimator. The other noticeable trend from Tables 1-3 are as follows:

1. The MSE rises as the level of multicollinearity rises, keeping other factors constant.
2. Also, MSE rises as the predictor variable increases keeping other factors constant.
3. Increasing the sample size n results in a decrease in the MSE for all the estimators' keeping other factors constant
4. These results agreed with the theoretical section.

Table 2. Simulated result in terms of MSE when p= 8

coef	50	100	200	500
		0.8		
$\widehat{\gamma}_{\text{MML}}$	5.5490	5.2430	4.9503	1.4723
$\widehat{\gamma}_{\text{k-B}}$	5.5467	4.9460	4.1475	1.4700
$\widehat{\gamma}_{\text{d-B}}$	5.5375	4.9272	3.1725	1.4710
$\widehat{\gamma}_{\text{kl-B}}$	4.7675	4.3211	3.2415	1.3980
$\widehat{\gamma}_{\text{kd-B}}$	3.8533	3.1229	1.2796	1.3476
		0.9		
$\widehat{\gamma}_{\text{MML}}$	6.1409	6.0730	5.2045	2.0049
$\widehat{\gamma}_{\text{k-B}}$	6.1372	6.0708	4.8290	2.0023
$\widehat{\gamma}_{\text{d-B}}$	6.1169	6.0617	4.1508	2.0024
$\widehat{\gamma}_{\text{kl-B}}$	5.9890	5.3678	4.0234	1.9087
$\widehat{\gamma}_{\text{kd-B}}$	4.3141	4.4619	3.4109	1.7746
		0.99		
$\widehat{\gamma}_{\text{MML}}$	7.3489	7.7949	6.8973	3.0801
$\widehat{\gamma}_{\text{k-B}}$	7.3374	6.7782	5.8251	3.0555
$\widehat{\gamma}_{\text{d-B}}$	7.2683	6.7231	5.0666	3.0412
$\widehat{\gamma}_{\text{kl-B}}$	6.9764	6.0123	5.1034	2.9807
$\widehat{\gamma}_{\text{kd-B}}$	6.1818	5.5338	4.0534	2.4082
		0.999		
$\widehat{\gamma}_{\text{MML}}$	17.7282	10.7036	10.0083	5.9225
$\widehat{\gamma}_{\text{k-B}}$	7.5111	6.9166	5.5853	4.7222
$\widehat{\gamma}_{\text{d-B}}$	8.0643	7.0671	6.7115	4.1167
$\widehat{\gamma}_{\text{kl-B}}$	8.0423	7.0654	6.0980	4.0088
$\widehat{\gamma}_{\text{kd-B}}$	7.0007	6.9086	4.4547	2.5738

Table 3. Simulated result in terms of MSE when p= 12

coef	50	100	200	500
		0.8		
$\widehat{\gamma}_{\text{MML}}$	8.3479	6.6716	5.6898	2.5792
$\widehat{\gamma}_{\text{k-B}}$	8.3446	5.6696	5.4483	2.1417
$\widehat{\gamma}_{\text{d-B}}$	8.3357	5.6696	4.6284	2.2176
$\widehat{\gamma}_{\text{kl-B}}$	6.7845	5.3222	3.9908	2.2214
$\widehat{\gamma}_{\text{kd-B}}$	5.9400	4.1370	3.8202	1.8786
		0.9		
$\widehat{\gamma}_{\text{MML}}$	9.7388	7.6736	6.0451	3.0170
$\widehat{\gamma}_{\text{k-B}}$	9.7362	6.8409	5.0426	2.6473
$\widehat{\gamma}_{\text{d-B}}$	9.7256	5.8649	5.0422	2.8627
$\widehat{\gamma}_{\text{kl-B}}$	8.7656	5.0119	4.8765	2.9080
$\widehat{\gamma}_{\text{kd-B}}$	7.8950	4.4621	4.2318	2.5720
		0.99		
$\widehat{\gamma}_{\text{MML}}$	15.5725	9.5476	5.6815	4.7363
$\widehat{\gamma}_{\text{k-B}}$	10.6565	7.4973	5.6736	4.6573
$\widehat{\gamma}_{\text{d-B}}$	10.6124	7.3872	5.1511	4.6358
$\widehat{\gamma}_{\text{kl-B}}$	9.6787	7.0577	5.0335	4.4631
$\widehat{\gamma}_{\text{kd-B}}$	9.2586	6.7495	4.9946	2.9945
		0.999		
$\widehat{\gamma}_{\text{MML}}$	31.7659	21.8870	11.1505	8.6753
$\widehat{\gamma}_{\text{k-B}}$	16.2341	10.0161	9.0282	6.1041
$\widehat{\gamma}_{\text{d-B}}$	12.1141	9.0713	7.0907	5.7521
$\widehat{\gamma}_{\text{kl-B}}$	11.0886	8.0989	6.5656	4.8989
$\widehat{\gamma}_{\text{kd-B}}$	11.1061	7.0443	5.0506	3.2270

Table 4. Bell regression estimates

Coef.	$\widehat{\gamma}_{MML}$	$\widehat{\gamma}_{k-B}$	$\widehat{\gamma}_{d-B}$	$\widehat{\gamma}_{kl-B}$	$\widehat{\gamma}_{kd-B}$
Intercept	-0.5423	-0.1484	-0.3006	0.0375	0.0322
x_1	0.5992	0.3396	-0.0034	0.3286	-0.0459
x_2	0.1630	0.1667	0.0119	0.1605	0.1685
x_3	-0.0117	-0.0146	0.0023	-0.0161	-0.0136
k/d		2.78554	0.1108	1.1333	18.0445 (0.1108)
MSE	1.7447	0.3914	0.5327	0.1493	0.0181

4. Application

In this section, we adopt the aircraft data to evaluate the performance of the existing estimators and the proposed. This dataset is originally assumed to follow the Poisson regression model [8, 9, 27]), among others. There is one response variable and three predictors (see [8, 9] for the details). Poisson distribution fits well to the outcome variable [8, 9, 27]). The model suffers from multicollinearity because the condition number is 219.3654 [8, 9]. However, the variance of the number of locations with damage on the aircraft is more than twice the mean (2.0569). With this, it is evident that the data exhibit over-dispersion. We fit the Bell regression model as alternative to account for the over-dispersion in the data. Table 4 provides the regression estimates and the Mean Squared Error of each of the adopted estimators in this study.

It is obvious from the result in Table 4, the proposed estimator produced the lowest MSE and dominates the MML, the ridge and the Liu estimators. The MML has the highest mean squared error. Thus, not recommended when there is multicollinearity.

5. Conclusion

The Poisson regression is often employed to model count data. However, it is certain that the Poisson regression model gives poor fit for count data with over-dispersion. Recently, the Bell regression model was introduced as alternative to the Poisson regression model for the purpose of accounting for over-dispersion in count data modelling. The conventional Method of Maximum Likelihood (MML) is employed to estimate the regression parameters. The estimator flop when the regressors are correlated. The ridge and the Liu estimator were developed to combat correlated regressors in Bell regression model. In this study, we developed a new method of parameter estimation in Bell regression model to compete with the existing ones. We compared the performance of the new method with some methods. The new method dominates the existing methods by giving a minimum Mean Squared Error.

Acknowledgments

The authors are very grateful to anonymous referees for their careful and diligent reading of the paper and helpful suggestions. This research has not received any sponsorship.

References

[1] F. Castellares, S. L. P. Ferrari & A. J. Lemonte, "On the Bell distribution and its associated regression model for count data", Applied Mathematical Modelling **56** (2017) 172, https://doi.org/10.1016/j.apm.2017.12.014

[2] M. Amin, M. N. Akram, & A. Majid, "On the estimation of Bell regression model using ridge estimator", Communications in Statistics - Simulation and Computation **2021** (2021), https://doi.org/10.1080/03610918.2020.1870694

[3] A. Majid, M. Amin, & M. N. Akram, "On the Liu estimation of Bell regression model in the presence of multicollinearity", Journal of Statistical Computation and Simulation **2021** (2021) 21, https://doi.org/10.1080/00949655.2021.1955886

[4] E. T. Bell, "Exponential numbers", The American Mathematical Monthly **41** (1934a) 419.

[5] E. T. Bell, "Exponential polynomials", Annals of Mathematics **35** (1934b) 258.

[6] N. T. Longford, "A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects", Biometrika **74** (1987) 817.

[7] B. M. G. Kibria, "Performance of some new ridge regression estimators", Communications in Statistics-Simulation and Computation **32** (2003) 419.

[8] A. F. Lukman, B. Aladeitan, K. Ayinde & M. R. Abonazel, "Modified ridge – type for the Poisson regression model**:** Simulation and Application", Journal of Applied Statistics **2021** (2021a). 10.1080/02664763.2021.1889998

[9] A. F. Lukman, E. Adewuyi, K. Månsson & B. M. G. Kibria, "A new estimator for the multicollinear Poisson regression model: Simulation and Application, Scientific Reports **11** (2021b) 3732. https://doi.org/10.1038/s41598-021-82582-w.

[10] A. E. Hoerl & R. W. Kennard "Ridge regression: biased estimation for nonorthogonal problems", Technometrics, **12** (1970) 55.

[11] K. Liu, "A new class of biased estimate in linear regression", Commun Stat. **22** (1993) 393.

[12] A. F. Lukman, K. Ayinde, S. Binuomote & A. C. Onate, "Modified ridge - type estimator to combat multicollinearity: application to chemical data", Journal of Chemometrics **2019** (2019) e3125. https://doi.org/10.1002/cem.3125

[13] R. Farebrother, "Further results on the mean square error of ridge regression", Journal of the Royal Statistical Society, Series B (Methodological) **38** (1976) 248.

[14] G. Trenkler & H. Toutenburg, "Mean squared error matrix comparisons between biased estimators an overview of recent results", Statistical Papers **31** (1990) 179.

[15] M. R. Ozkale & S. Kaciranlar, The restricted and unrestricted two-parameter estimators" Commun. Statist. Theory, Meth. **36** (2007) 2707.

[16] Stan Development Team, RStan: *The R interface to Stan. R package version 2.19.3* (2020), https://mc-stan.org

[17] M. Arashi, M. Roozbeh, N. A. Hamzah, & M. Gasparini, "Ridge regression and its applications in genetic studies" PloS One **16** (2021a) 4, e0245376.

[18] M. Arashi, M. Norouzirad, M. Roozbeh, & N. M. Khan, "A high-dimensional counterpart for the ridge estimator in multicollinear situations", Mathematics **9** (2021b) 3057, https://doi.org/10.3390/math9233057

[19] Y. M. Bulut, "Performance of the Liu-type estimator in the Bell regression model", 9th International Conference on Applied Analysis and Mathematical Modeling (ICAAMM21), Istanbul/TURKEY (2021).

[20] O. G. Obadina, A. F. Adedotuun & O. A. Odusanya, "Ridge estimation's effectiveness for multiple linear regression with multicollinearity: an Investigation Using Monte-Carlo Simulations", Journal of the Nigerian Society of Physical Sciences **3** (2021) 278.

[21] M. Qasim, K. Månsson, P. Sjolander & B. M. G. Kibria, "A new class of efficient and debiased two – step shrinkage estimators: method and application", Journal of Applied Statistics **2021** (2021), https://doi.org/10.1080/02664763.2021.1973389

[22] A. K. M. E. Saleh, M. Arashi, & B. M. G. Kibria, *Theory of ridge regression estimation with applications* John Wiley, USA (2019).

[23] M. Suhail, S. Chand & B. M. G. Kibria, "Quantile-based robust ridge M-estimator for linear regression model in presence of multicollinearity and

outliers", Communications in Statistics-Simulation and Computation **50** (2021) 3194.

[24] N. K. Rashad & Z. Y. Algamal, "A new ridge estimator for the Poisson regression model", Iranian Journal of Science and Technology, Transactions A: Science, **43** (2019), https://doi.org/10.1007/s40995-019-00769-3

[25] Z. Y. Algamal & Y. Asar, "Liu-type estimator for the Gamma regression model", Communications in Statistics-Simulation and Computation **8** (2018) 2035.

[26] B. M. G. Kibria & A. F. Lukman, "A new ridge – type estimator for the linear regression model: simulations and applications", Scientific **2020** (2020) 9758378, https://doi.org/10.1155/2020/9758378

[27] R. H. Myers, D. C. Montgomery, G. G. Vining & T. J. Robinson, *Generalized linear models: with applications in engineering and the sciences*, Wiley, New York (2012) 791