Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN: 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.50
https://jatuc.uoturath.edu.iq/

A Comparative Study of Estimation Methods for the Bell Regression Model under Multicollinearity

Sarah S. Sachat [1*], Eman M. Abdallah [2]

[1,2]Department of Statistics, College of Administration and Economics, University of Baghdad, Baghdad, Iraq.

[*]*Correspondence email*: sarah.salem2201@coadec.uobaghdad.edu.iq

KEYWORDS

Bell regression; Multicollinearity; Ridge regression, K-L estimator; Jackknifed K-L.

ABSTRACT

The Bell regression model is a contemporary statistical framework for the analysis of count data, especially in instances of overdispersion, in contrast to the Poisson model and the Negative binomial regression model. However, the existence of multicollinearity among explanatory variables can cause instability in parameter estimations when employing the Maximum Likelihood Estimator (MLE), leading to diminished estimation accuracy.

This paper examines the issue of multicollinearity in the Bell regression model by evaluating various estimating techniques, including the Maximum Likelihood Estimator (MLE), Bell Ridge Regression (BRR), Bell Liu Regression (BLR), the K-L estimator, and the Jackknifed K-L estimator. The efficacy of these estimators was assessed utilizing the Mean Squared Error (MSE) criterion. A Monte Carlo simulation analysis was performed utilizing sample sizes (n = 50, 100, 150, 250), five explanatory variables (p = 5), and varying correlation levels (0.90, 0.95, 0.99). The simulation results indicated that the Bell Ridge regression estimator at the shrinkage parameter (k_1) achieved the lowest mean squared error (MSE) values in most instances relative to other estimators, indicating its superior performance in the presence of multicollinearity. Furthermore, real data from prostate cancer patients were analyzed. The results indicated that the Bell Ridge regression estimator outperformed the maximum likelihood estimator (MLE) and other estimators by attaining the lowest mean squared error (MSE), thereby demonstrating its capacity to deliver more stable and precise estimates, enhance the reliability of predicting disease-related outcomes, and effectively address issues of multicollinearity and overdispersion in the count data.

1. INTRODUCTION

Regression models are essential statistical techniques for examining the relationship between the response variable and the predictor variables. In the conventional linear regression model, the response variable is presumed to adhere to a normal distribution. Nevertheless, this presumption is rarely consistently fulfilled in practical applications, particularly when handling numerical data. Consequently, the Generalized Linear Model (GLMs) is employed, facilitating the modeling of this data type by associating the response variable with a linear combination of the explanatory factors through an appropriate correlation function.

Poisson regression is one of the most prevalent models in this domain. This model presupposes that the mean is equivalent to the variance, a condition that may be unattainable in numerous practical scenarios marked by significant

dispersion. Consequently, Bell regression has been suggested as a more adaptable option for modeling quantitative data.

Model parameters are typically calculated using Maximum Likelihood Estimation (MLE), but the existence of multicollinearity among the explanatory variables can impair the effectiveness of the estimation. This study seeks to compare various shrinkage estimators within the Bell regression model (BRM), specifically the maximum likelihood estimator, the Bell Ridge Regression estimator, the Liu estimator, the K-L estimator, and the Jackknifed K-L estimator, to mitigate multicollinearity and enhance estimation accuracy.

2. MATERIAL AND METHODS

2.1. Bell Distribution

The Bell distribution is a contemporary counting distribution employed to describe discrete random variables that assume non-negative integer values. It is distinguished by its superior flexibility relative to the Poisson distribution, particularly in instances with significant dispersion. This distribution was initially presented by [3].

Let y_i denote a random variable that signifies the frequency of a specific occurrence occurring within a designated time frame. y_i Is presumed to adhere to a Bell distribution characterized by a parameter $\theta > 0$. The Probability Mass Function (PMF) of the Bell distribution is articulated as follows:

$$Y \sim Bell(\theta)\,, \qquad \mathrm{pr(Y = y)} = \frac{\theta^{y}\, e^{-e^{\theta}+1} B_y}{y!} \qquad y = 0,1,2,..n \tag{1}$$

θ :The distribution parameter is represented by a value greater than zero .

B_y : Bell number $B_n = \frac{1}{e}\sum_{k=0}^{\infty}\frac{k^n}{k!}$.

2.2. Derivation of the mean and Variance of the Bell Distribution

The following detailed derivation demonstrates the verification of the mean and variance of the Bell distribution, as described in reference [3].

can be derived utilizing the probability-generating function as follows:

$$\mathrm{G_Y(s)} = \exp\left(e^{\theta s} - e^{\theta}\right) \tag{2}$$

$$\begin{gathered} E[Y] = G_Y\grave{}(s) \\ \grave{G}_Y(s) = \frac{d}{ds}[\exp\left(e^{\theta s} - e^{\theta}\right)] = \exp\left(e^{\theta s} - e^{\theta}\right)\,.\left(\theta e^{\theta s}\right) \\ \text{s=1 :} \\ \grave{G}_Y(1) = \exp\left(e^{\theta} - e^{\theta}\right)\,.\left(\theta e^{\theta}\right) \\ \grave{G}_Y(1) = 1.\theta \mathrm{e}^{\theta} \\ E[Y] = \theta e^{\theta} \end{gathered} \tag{3}$$

Variance:

$$\mathrm{Var[Y]} = \mathrm{G''_Y}\,(1) + \grave{\mathrm{G}}_\mathrm{Y}(1) - \left[\grave{\mathrm{G}}_\mathrm{Y}(1)\right]^2$$

$$\begin{gathered} G_Y''(s) = \frac{d}{ds}\left[\grave{G}_Y(s).\left(\theta e^{\theta s}\right)\right] = \grave{G}_Y(s).\frac{d}{ds}\left(\theta e^{\theta s}\right) + .\left(\theta e^{\theta s}\right)\,\frac{d}{ds}\,\grave{G}_Y(s) \\ G_Y''(s) = \exp\left(e^{\theta s} - e^{\theta}\right)\,.\frac{d}{ds}\left(\theta e^{\theta s}\right) + \left(\theta e^{\theta s}\right).\frac{d}{ds}\exp\left(e^{\theta s} - e^{\theta}\right) \\ G_Y''(s) = \exp\left(e^{\theta s} - e^{\theta}\right)\,.\theta^2 e^{\theta s} + \exp\left(e^{\theta s} - e^{\theta}\right)\,.\theta^2 e^{2\theta s} \\ = \exp\left(e^{\theta s} - e^{\theta}\right).\theta^2 e^{\theta s}(1 + e^{\theta s}) \end{gathered} \tag{4}$$

$$G_Y''(s=1) = 1.\ \theta^2 e^{\theta}(1+e^{\theta})$$

$$G_Y''(1) = \theta^2 e^{\theta}(1+e^{\theta})$$

$$\mathrm{Var(Y)} = \theta^2 e^{\theta}(1+e^{\theta}) + \theta\, e^{\theta} - (\theta e^{\theta})^2$$

$$= \theta^2 e^{\theta} + \theta^2 e^{2\theta} + \theta e^{\theta} - \theta^2 e^{2\theta} = \theta^2 e^{\theta} + \theta e^{\theta}$$

$$\mathrm{Var(Y)} = \theta e^{\theta}(1+\theta)$$

2.3. Bell Regression Model

Bell regression model is a count data model employed to analyze discrete numerical data that signifies the frequency of an event, such as the count of doctor visits, hospital days, or accidents during a specified timeframe, within the context of generalized linear models [4].It is used as a flexible alternative to Poisson regression when overdispersion is present.

To design a regression framework for the Bell distribution, we adjust its parameter using the response mean **μ**, defined as $\boldsymbol{\mu = \theta e^{\theta}}$, where $\theta = W_0(\mu)$. As previously defined, $W_0(.)$is an The Lambert function is used to obtain a tractable form for the Bell regression model by expressing the parameter θ in terms of the mean μ, thus facilitating model c formulation for count data exhibiting overdispersion[5]. Hence, the mean, variance, and probability mass function may be expressed using the new parameter for the Bell distributio

$$\Pr(Y = y_i) = \exp\left(1 - e^{W_0(\mu)}\right)\frac{W_0(\mu)^y B_y}{y!}\ , \qquad y = 0,1,2\ldots. \tag{5}$$

$$E(Y) = \mu \quad , \qquad VAR(Y) = \mu[1 + W_0(\mu)] \tag{6}$$

The Bell regression model can be articulated using a logarithmic linking function as follows *:*

$$\begin{aligned} Y_i &\sim Bell\left(W_0(\mu_i)\right) && i = 1,2,\ldots n \\ g(\mu_i) &= \log(\mu) = \eta_i = x_i^{\mathrm{T}}\beta && i = 1,2,\ldots n \end{aligned} \tag{7}$$

$g(.)$: The link function

η **:** Vector of linear predictors and its dimensions $(n * 1)$.

Let $\beta = (\beta_1, \beta_2, .., \beta_p) \in \mathbb{R}^p$ denote the vector of regression coefficients of dimension p, where p < n. The quantity η_i represents the linear predictor, and $x_i = (x_{i1}, x_{i2}\ldots, x_{ip})$ denotes the observed values of the p explanatory variables associated with the i-th observation.

It is worth noting that the variance of Y_i is expressed as a function of its mean μ_i, and therefore depends implicitly on the covariates. This characteristic enables the model to naturally account for heteroscedasticity in the response variable[1].

Furthermore, the link function $g : (0, \infty) \to \mathbb{R}$ is assumed to be strictly monotonic and twice continuously differentiable. Different specifications of the link function may be adopted according to the modeling framework. Commonly used choices include the logarithmic link $g(\mu) = \log(\mu)$, the square-root link $g(\mu) = \sqrt{\mu}$ and the identity link $g(\mu) = \mu$, although the latter requires careful handling to ensure that the estimated mean values remain positive. For further discussion on link functions [1].

$$\mu_i = \exp\left(x_i^{\mathrm{T}}\beta\right) \qquad > 0 \tag{8}$$

$g(.)$: The link function

η **:** Vector of linear predictors and its dimensions $(n * 1)$.

2.4. Multicollinearity

The linear regression model assumes the absence of significant linear dependence among the independent variables; however, empirical data often exhibit varying levels of linear correlation, either direct or indirect[6]. Multicollinearity is a prevalent issue in regression models, occurring when there is a perfect or nearly perfect linear correlation among the explanatory variables. The severity of this issue depends on the extent of linear dependency among the independent variables, specifically whether the correlation is perfect or nearly perfect. Moreover, to guarantee the estimation of model parameters, the number of parameters must be less than the sample size.

$$rank(x) = (p+1) < n \tag{9}$$

The correlation coefficient between any pair of explanatory variables should not be equal to : $r_{x_i x_j} \neq \pm 1$

In the presence of a perfect linear relationship among the explanatory factors, it becomes impossible to distinguish the effects of these variables within the model.

2.4.1 Near-Perfect Multicollinearity

Near-perfect multicollinearity arises when certain explanatory variables are linear combinations of other variables, The determinant of the information matrix is exceedingly very small and approaches zero, resulting in unstable estimates of the model's coefficients.

2.4.2 Detection of Multicollinearity

1-Correlation Matrix: The correlation coefficient among the explanatory variables is computed. A high correlation coefficient between two variables may indicate multicollinearity. This method is simple to use but does not provide an accurate measure multicollinearity levels.

2- Condition Index (CI) :The conditional indices serve as a prevalent diagnostic tool for identifying multicollinearity issues. They depend on the characteristic roots (eigenvalues) of the information matrix, which represent the variation structure the independent variables if characteristic roots equals zero signifies perfect multicollinearity. values of characteristic roots close to zero indicate high multicollinearity .If the characteristic root values are equal to one, it indicates the absence of multicollinearity.

$$Cl = \sqrt{\frac{Maximum\ eigenvalue}{Minimum\ eigenvalue}} \quad >1 \tag{10}$$

Explanation of Cl

If it is between 10 and 30 → there is moderate to strong multicollinearity.

If it is greater than 30 → there is strong multicollinearity.

2.5. Methods for Estimating the Parameters of a Bell Regression Model

2.5.1. The maximum likelihood estimator method for Bell regression model

The parameters of a Bell regression model are estimated using the maximum likelihood approach. let the response

variable as counting data Y_i follow a Bell distribution, $Y_i \sim Bell\ (W_0(\mu))$, where the mean μ>0.

The likelihood function for a sample consisting of n observations is written as follows:

$$\ell(\beta) = \sum_{i=1}^{n} \left[\sum_{i=1}^{n} y_i \log(W_0(\mu_i)) - e^{W_0(\mu_i)}\right] \dots (11)$$

The score function U(β) is obtained by taking the first partial derivative of the log-likelihood function with respect to the parameter vector β

$$\boldsymbol{U(\beta)} = X^T\, W^{1/2}\ \ V^{-1/2}\,(y-\mu)$$

The model matrix $X = (x_1 x_2 \dots x_n)^T$, $\mathrm{W} = \mathrm{diag}\{\mathrm{w}_1, w_2, \dots, w_n\}$, $\mathbf{V} = diag\{v_1, v_2, \dots, v_n\}$, $\mathbf{y} = (y_1, y_2, \dots, y_n)^T$, $\boldsymbol{\mu} = (\mu_1 \mu_2, \dots, \mu_n)^T$ $\mathrm{w_i} = \frac{(d\mu_i/\mathrm{d}\eta_1)^2}{V_i}$, $\mathrm{V_i} = \mu_i\,[1 + \mathrm{W_0}\,(\mu_i)]$, $\mathrm{i} = 1, 2, \dots, \mathrm{n}$

Note that V i is the variance function of Y_i. The Fisher information matrix for β is given by $\mathrm{K}(\beta) = \mathrm{X^T\,W\,X}$.

A closed-form solution cannot be derived after computing the first derivative $\boldsymbol{U(\hat{\beta})} = \mathbf{0}$ because of the nonlinear characteristics of the possibility function. Consequently, numerical techniques such as the Newton–Raphson iterative

approach and the Fisher scoring method are employed. Fisher's scoring system is employed to estimate parameters with in probability models, particularly in generalized linear models (GLMs) [3].

$$\beta^{(m+1)} = (X\,W^m X)^{-1} X\,W^m z^m \tag{12}$$

The frequency number represents: $m = 0,1 \ldots$

W: Weight matrix

$z = (z_1, z_2, \ldots, z_n) = \eta + W^{\frac{1}{2}} V^{\frac{-1}{2}} (y - \mu)$: A contributing variable called the working response and its dimensions $(n * 1)$.

Estimated can be written MLE

$$\hat{\beta}_{MLE} = \left(X^T \widehat{W} X\right)^{-1} X' \widehat{W} \hat{z} \tag{13}$$

$$Cov\left(\hat{\beta}_{MLE}\right) = D^{-1} \tag{14}$$

It can be written individually as a scalar using : $SMSE = tr(D)^{-1} = \sum_{j=1}^{p} \frac{1}{\lambda_j}$ (15)

λ_j: eigenvalue (j) For the matrix D [7] .

2.5.2. The Bell Ridge Regression Estimator

The Bell ridge regression estimator is a technique employed to estimate regression model parameters in the presence of multicollinearity among the explanatory variables.[8] initially introduced the ridge regression technique, which incorporates a positive bias parameter $k > 0$ into the principal diagonal elements of the information matrix, applicable across many model types[9]. And others expanded ridge regression to incorporate Bell regression model, resulting in the Bell Ridge Regression Estimator (BRR), using Bell regression as a logarithmic-linear model. This estimator seeks to diminish the sensitivity of the maximum likelihood estimator to multicollinearity while preserving its appropriateness for modeling count data.

$$\hat{\beta}_{BRR} = (D + kI)^{-1} D \hat{\beta}_{MLE} \tag{16}$$

$$Bias\left(\hat{\beta}_{BRR}\right) = \quad [-k\left(D + kI_p\right)^{-1} \beta] \tag{17}$$

$$Var\left(\hat{\beta}_{BRR}\right) = \quad D_k^{-1} D D_k^{-1} \tag{18}$$

$$MMSE\left(\hat{\beta}_{BRR}\right) = D_k^{-1} D D_k^{-1} + k^2 \mathrm{D_k}^{-1} \beta\beta^T D_k^{-1} \tag{19}$$

$$MSE\left(\hat{\beta}_{BRR}\right) = \sum_{j=1}^{p} \frac{\lambda_j}{\left(\lambda_j + k\right)^2} + k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{\left(\lambda_j + k\right)^2} \tag{20}$$

Shrinkage parameter of the Bell Ridge Regression

This work employs the methodologies defined in [10] to estimate the character regression parameter. The parameter is derived by computing the first derivative of the mean squared error equation about the character parameter[9], resulting in:

$$\hat{k}_j = \frac{1}{\hat{\alpha}_j^2} \qquad\qquad j = 1,2, \ldots n$$

$\alpha_j = Q'\hat{\beta}_{\mathrm{MLE}}$ Where

α_j: It refers to intrinsic values(j) from the vector $\alpha_j = (\alpha_1, \alpha_2 \ldots., \alpha_{p+1})$

And based on $\hat{k}_j$ ،using k_j the character regression parameters are determined:

$$k_1 = \frac{p}{\sum_{j=1}^{p}(1/\hat{k}_j)} \tag{21}$$

$$k_2 = \min(\hat{k}_j) \tag{22}$$

2.5.3. The Bell Liu Regression Estimator

Kejian (1993) [11] established the theoretical framework for an estimating method designed to mitigate variance inflation in estimated parameters, frequently caused by multicollinearity among explanatory factors. This strategy depends on creating an estimator with favorable statistical characteristics for efficiency and reliability by integrating a shrinkage technique into the estimating process. In the framework of the Bell regression model, the philosophy put forward by Liu was embraced, acknowledging the model's nonlinear nature, which requires the estimation method to be tailored to the specific attributes of such models. To resolve this issue, [7]introduced the Bell-Liu regression estimator, a variant of a shrinkage estimator, which modifies the Fisher information matrix to mitigate the impact of multicollinearity and enhance estimation accuracy. It is provided by:

$$\hat{\beta}_{BLR} = (X^t WX + I)^{-1}(X^t WX + dI)\hat{\beta}_{MLE} \tag{23}$$

d This represents the shrinkage parameter (bias) and its value is between $0 < d < 1$

If it was $d = 1$ we achieve the identical maximal potential.$\hat{\beta}_{MLE}$

$$Bias(\hat{\beta}_{BLR}) = D^{-1}\beta(d-1) \tag{24}$$

$$Cov(\hat{\beta}_{BLR}) = \mathrm{D}^{-1}\, D_d D^{-1} D_d D^{-1} \tag{25}$$

$$MMSE(\hat{\beta}_{BLR}) = Cov(\hat{\beta}_{BLR}) + Bias(\hat{\beta}_{BLR})Bias(\hat{\beta}_{BLR})^t$$

$$= \mathrm{D}^{-1}\, D_d D^{-1} D_d D^{-1} + (d-1)^2 D^{-1}\beta\beta^t D^{-1} \tag{26}$$

The mean squared error of the Liu estimator in the Bell regression model[7] :

$$SMSE(\hat{\beta}_{BLR}) = \sum_{j=1}^{p} \frac{(\lambda+d)^2}{\lambda(\lambda+1)^2} + (1-d)^2 \sum_{j=1}^{p} \frac{\alpha^2}{(\lambda+1)^2} \tag{27}$$

Shrinkage parameter of the Bell Liu estimator

The efficacy of BLR estimation is contingent upon the appropriate value of the bias parameter d. To achieve the ideal value of the BLR estimation, we compute the first derivative of the mean squared error equation and determine d by equating it to zero. This provides the ideal Lieu parameter.

$$d_{opt} = \sum_{j=1}^{p} \frac{\hat{\alpha}_j^2 - 1}{(1/\lambda_j) + \alpha_j^2}$$

The range of d depends between 0 and1 [11] :

The shrinkage factor will be provided d

$$d = max\left(0, min\left(\frac{\hat{\alpha}_j^2 - 1}{(1/\lambda_j) + \alpha_j^2}\right)\right) \tag{28}$$

2.5.4. K-L Eestimator in Bell regression model

Introduced a Ridge-type estimator termed the Kibria-Lockman (K-L) estimator[12], which combines the properties of the Ridge and Liu estimators to enhance statistical performance amid significant linear correlation among explanatory variables.

$$\hat{\beta}_{kl} = (((X^T X + KI)^{-1}((X^T X - KI)^{-1})\ \hat{\beta}_{MLE} \tag{29}$$

Consequently, the researchers [13] introduced an estimator (K-L) for the Bell regression model to address the multicollinearity issue and estimate the parameters. The formula for estimation is as follows:

$$\hat{\beta}_{KLB} = \left(\left(X^T \widehat{W} X + KI\right)^{-1}\left(X^T \widehat{W} X - KI\right)^{-1}\right)\ \hat{\beta}_{MLE} \tag{30}$$

$\hat{\beta}_{kl}$To calculate the bias and variance of the estimate

$$Bias\left(\hat{\beta}_{KLB}\right) = -2kQ\left(X\widehat{W}X + kI\right)^{-1}\alpha \tag{31}$$

$$\text{Variance}\left(\hat{\beta}_{KLB}\right) = \text{Q}\left(X^T \widehat{W} X + kI\right)^{-1}\left(X^T \widehat{W} X - kI\right)^{-1}\left(X^T \widehat{W} X\right)^{-1}\left(X^T \widehat{W} X + kI\right)^{-1}\left(X^T \widehat{W} X - kI\right)^{-1} Q^T \tag{32}$$

The mean squared error of the Kiberia-Lockman estimator within the Bell regression model as a singular entity:

$$SMSE\left(\hat{\beta}_{KLB}\right) = \sum_{j=1}^{p} \frac{(\lambda-k)^2}{\lambda(\lambda+k)^2} + 4k^2 \sum_{j=1}^{p} \frac{\alpha^2}{(\lambda+k)^2} \tag{33}$$

Shrinkage parameter of the K–L estimator

The shrinkage is determined by calculating the first derivative of the mean squared error concerning k [12]:

$$k = min\left(\frac{\lambda}{1+2\lambda\,\alpha^2}\right) \tag{34}$$

λ_j :It is the eigenvalue j of $X^T \widehat{W} X$

α_j : The element j represents a vector $\alpha = (\alpha_1, \alpha_2, .. \alpha_{p+1})$

2.5.5 Jackknifeing K-L estimator in Bell regression model

The K-L estimator is a contraction estimator designed to mitigate the issue of multicollinearity, as previously mentioned. This estimator is intrinsically biased. To mitigate this bias, the integration of the Jackknife approach with the K-L estimator has been suggested, yielding the Jackknife K-L estimator. The Jackknife method entails removing one observation from the sample sequentially, estimating the parameters, and subsequently producing a new estimate using pseudo-values. This mitigates bias while maintaining the stability characteristics provided by contraction[14]. To systematically develop the Jackknife K-L estimator, the maximum likelihood estimator and the K-L estimator will initially be reformulated within the spectral decomposition of the $X^T \widehat{W} X$ matrix. This will enable the examination of the mathematical properties and the exact formulation of the Jackknife formula.

Whereas $\Lambda = Diag(\lambda_1, \lambda_2, .., \lambda_p)$ the eigenvalues of a matrix $X^T \widehat{W} X$ Whereas $M = XQ$ The design matrix is represented in dimensions $(n * p)$

$$\acute{Q}(X^T \widehat{W} X)Q' = M^T \widehat{W} M = \Lambda \tag{35}$$

Therefore, the MLE can be expressed as $\hat{\beta}_{MLE} = Q\hat{v}_{MLE}$ Where as

$$\hat{v}_{MLE} = \Lambda^{-1} M^T \widehat{W} \hat{\mu} \tag{36}$$

As a result, K-L estimator can be written in the following form:

$$\hat{v}_{KL-GLM} = (\Lambda + \mathrm{kI})^{-1}(\Lambda - \mathrm{kI})^{-1}\mathrm{M}^{\mathrm{T}}\widehat{W}\hat{\mu} \tag{37}$$

In accordance with row i was sequentially removed from both the response vector μ_{-i} and the estimated weights matrix W_{-i}, as well as from the design matrix M_{-i} [15] .

The formula for the leave-one-out method can be written as follows:

$$\hat{v}_{KL-GLM(-i)} = \left(M_{-i}^{T}\widehat{W}_{-i}M_{-i} + kI\right)^{-1}\left(M_{-i}^{T}\widehat{W}_{-i}M_{-i} - kI\right)^{-1}M_{-i}^{T}\, W_{-i}\, M_{-i} \tag{38}$$

The inverse $\left(M_{-i}^{T}\widehat{W}_{-i}M_{-i} \pm kI\right)^{-1}$ is computed utilizing the Sherman–Morrison–Woodbury theorem. To prevent the need for recalculating the full inverse following each deletion, the estimator for the removal of the -one can be reformulated:

$$\hat{v}_{KL-GLM(-i)} = \hat{v}_{KL-GLM} - \frac{\left(M^{T}\widehat{W}M+kI\right)^{-1}\left(M^{T}\widehat{W}M+kI\right)^{-1}m_i^{T}(\widehat{u_i}-m_i^{T}\hat{v}_{KL-GLM})}{1-m_i^{T}\left(M^{T}\widehat{W}M+kI\right)^{-1}\left(M^{T}\widehat{W}M-kI\right)^{-1}m_i} \tag{39}$$

Where it represents m_i^{T} the row vector i of the design matrix M and its dimensions $(1 * p)$

Based on the resulting estimates and according to [15], weighted pseudo-values are generated and calculated as follows:

$$T_i = \hat{v}_{KL-GLM} + n(1 - m_i^{T}\left(M^{T}\widehat{W}M + kI\right)^{-1}\left(M^{T}\widehat{W}M - kI\right)^{-1}m_i(\hat{v}_{KL-GLM} - \hat{v}_{KL-GLM(-i)}) \tag{40}$$

The Jackknife K-L estimator can now be defined by the following spectral formula:

$$\hat{v}_{JKL-BRM} = \hat{v}_{KL-BRM} + \left(M^{T}\widehat{W}M + kI\right)^{-1}\left(M^{T}\widehat{W}M - kI\right)^{-1}\sum_{i=1}^{n} m_i^{T}(\hat{\mu} - m_i^{T}\,\hat{v}_{KL-BRM}) \tag{41}$$

Returning from spectral space to the original model parameters according to the formula:

$$\hat{\beta}_{JKL-BRM} = Q\hat{v}_{JKL-BRM} \tag{42}$$

The efficacy of the $\hat{\beta}_{JKL-BRM}$ estimator can be evaluated using the Mean Squared Error (MSE) criterion, a holistic metric of bias and variance, as outlined below*:*

$$\mathrm{Bias}(\hat{v}_{JKL-BRM}) = \left[I - 2k\left(M^{T}\widehat{W}M + kI\right)^{-1}\right]^{2}\nu$$

$$\mathrm{Var}(\hat{v}_{JKL-BRM}) = \left[I - \left(2k\left(M^{T}\widehat{W}M + kI\right)^{-1}\right)^{2}\right]^{2}\left(M^{T}\widehat{W}M + kI\right)^{-1}\left[I - 2k\left(M^{T}\widehat{W}M + kI\right)^{-1}\right]^{2}$$

$$MSE\left(\hat{\beta}_{jkl-BRM}\right) = \sum_{j=1}^{p}\frac{\left((\lambda_j+k)^{2}-4k^{2}\right)^{2}(\lambda_j-k)^{2}}{\lambda_j(\lambda_j+k)^{6}} + \sum_{j=1}^{p}\frac{\left((\lambda_j-k)^{2}(\lambda_j+3k)-(\lambda_j+k)^{3}\right)^{2}\alpha_j^{2}}{(\lambda_j+k)^{6}} \tag{43}$$

The Jackknife estimator utilizes the same shrinkage parameter k as the K-L contraction estimator, as it does not introduce a new shrinkage parameter; instead, it applies the single deletion method (Jackknife) to the K-L contraction estimator without adding an additional shrinkage parameter.

3. RESULTS

3.1. Simulation Design

The experiment was conducted utilizing the R-Studio, wherein virtual data was generated based on the response variable (y) distributed by $Bell(\mathrm{W}_0\ (\mu_i))$, and a regression model was developed as follows:

$\log(\mu) = \eta = \beta + \sum_1^n \beta\ X_{ij}$, $i = 1,2,..,n$

The explanatory variables follow a conventional normal distribution[9].

Data were produced under three correlation circumstances ($\rho = 0.9,0.95,0.99$),

Different sample sizes($n = 50,100,150,250$), were used to examine the behavior and stability of the estimators under small, moderate, and large samples, and to evaluate their performance under finite-sample conditions.

And a range of variables $p = 5$.

$x_{ij} = \sqrt{1-\rho^2}\, z_{ij} + \rho z_{i(j+1)}$, $i = 1,2 \dots n$, $j = 1,2, \dots \rho$

z_{ij}:pseudo-random numbers which are generated from the standard normal distribution .

Default parameter values : $(\beta_1 = 0.45\ ,\ \beta_2 = -0.35\ ,\ \beta_3 = 0.30\ ,\ \beta_4 = 0.60,\ \beta_5 = -0.40)$

Whereas$\sum_{j=1}^{p} \beta_j^2 = 1$

The experiment $(R = 1000)$ times is repeated to calculate the MSE simulation criterion as follows:

$$MSE(\hat{\beta}) = \frac{\sum_{i=1}^{R} (\hat{\beta}_i - \beta)^T (\hat{\beta}_i - \beta)}{R}$$

Examination of simulation outcomes with five explanatory variables (p=5):

- The mean squared error values diminish as the sample size grows, but the number of variables remains constant (p=5) across all correlation coefficients, signifying that the estimated value converges towards the true value with a larger sample size.
- The Bell Ridge Regression estimator with k_1 generally outperformed the other estimators across most sample sizes. However, at very high correlation levels, the Bell Liu Regression estimator showed better performance for smaller sample sizes, while the Ridge estimator regained superiority as the sample size increased.

Table 1. Mean Squared Error Values for All Methods ($p = 5$)

n	50	100	150	250	50	100	150	250	50	100	150	250
ρ	0.90				0.95				0.99			
MLE	2.145	1.447	1.314	1.164	3.259	2.081	1.686	1.410	14.605	7.311	5.095	3.465
BRR-k_1	1.286	1.079	1.060	1.013	1.605	1.258	1.160	1.075	4.827	2.829	2.136	1.718
BRR-k_2	1.630	1.225	1.163	1.074	2.237	1.583	1.368	1.206	8.687	4.589	3.344	2.414
BLR	1.601	1.314	1.248	1.139	1.813	1.578	1.447	1.301	1.933	1.053	2.922	1.858
K-L	1.640	1.229	1.166	1.076	2.270	1.592	1.376	1.210	8.880	4.670	3.396	2.448
JKL	1.399	1.125	1.094	1.033	1.821	1.364	1.228	1.116	6.260	3.453	2.591	1.974

Source: Author's calculations based on simulated data

The Fig.1 illustrates a comparison between different estimators at a sample size of (n = 250) and under varying correlation levels, based on the mean squared error (MSE). This sample size was chosen as a representative case to avoid redundancy, as similar patterns were observed across most sample sizes with only minor differences.

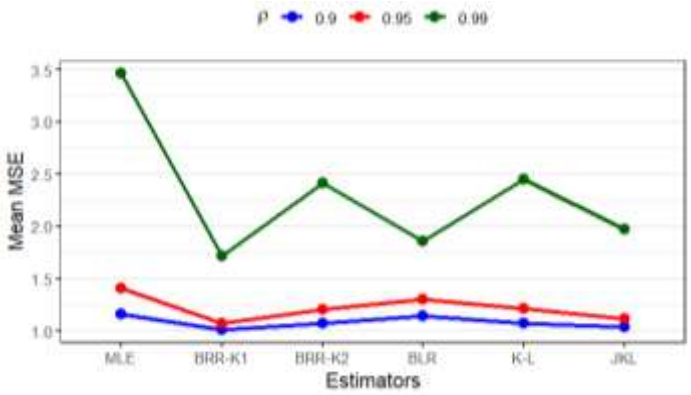

Fig.1 Performance comparison of the estimators based on (MSE) with n=250

3.2. Application Regarding Prostate Cancer

In practical terms, statistics pertaining to prostate cancer, a prevalent malignancy among males, are utilized. This study is to examine the quantity of lymph nodes impacted by cancer as a response variable indicative of quantifiable data. The dissemination of cancer cells to the lymph nodes serves as a significant marker of disease advancement and metastasis beyond the prostate gland. The data encompass multiple significant explanatory variables, including patient age, Prostate-Specific Antigen (PSA) test level, and tumor grade as determined by the Gleason score [1]. Due to the response variable representing quantified data and potentially exhibiting excessive dispersion, Bell regression was employed to examine the connection between the explanatory variables and the count of afflicted lymph nodes. Various techniques for estimating model coefficients were employed, and their efficacy was assessed to identify the most effective way in the context of multicollinearity among the explanatory variables. Patient data were acquired from the medical records of prostate cancer patients at a private medical facility in Baghdad. This data was utilized to develop a Bell regression model with a sample size of n = 87, whereby the dependent variable comprised:

Y: The number of cancerous nodules present in people with prostate cancer.

X1: Patient's age.

X2: Prostate-Specific Antigen level.

X3: Result of lymph node biopsy.

X4: Tumor size.

X5:The extent of the disease.

The descriptive statistics for the dependent variable indicated that it exhibited overdispersion.

$$I_d = Var(Y)/E(Y) = \frac{2.570}{1.344} = 1.911 \quad > 1$$

This supports the implementation of a more adaptable counting paradigm, such as Bell regression.

Table 2. The correlation matrix values among the explanatory factors**.**

Correlation matrix of X	X1	X2	X3	X4	X5
X1	1.000	-0.049	-0.014	-0.035	-0.042
X2		1.000	-0.101	-0.063	0.604
X3			1.000	0.211	0.097
X4				1.000	0.211
X5					1.000

The correlation matrix indicates a robust association between variables (x5, x2), evidenced by a correlation coefficient of (0.604). This signifies a robust positive association between PSA levels and disease prevalence. The correlations among the remaining pairs were weak and nearly negligible, thus exerting minimal practical influence on the stability of the estimate.

Eigenvalues and the conditional index (CI)

Table 3. represents the eigenvalues and the conditional index (CI) of the explanatory variables:

Eigenvalues	Number of conditions	CI
583984.7	1	1
23380	24	4.89
21	27808.7	166.7
13	44,921.9	211.9
7	83426.3	288.8

The eigenvalues of the $(\acute{X}X)$ matrix exhibit a wide range, from 7 to 583984.7, indicating significant variability and suggesting a possible near-perfect linear dependence among some explanatory variables. Furthermore, the Condition Index (CI) values are notably high, reaching a maximum of (288.8) exceeding the generally accepted threshold of 30, thus providing evidence of multicollinearity.

Applied experiment results

Table 4. Comparison of methods for estimating Bell regression model parameters and mean squared error (MSE).

	$\hat{\beta}_o$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	MSE
MLE	-0.1690	0.0090	-0.0001	0.1619	-0.0785	-0.2857	2.0472
BRR-k_1	-0.0003	0.0085	-0.0028	0.0324	-0.0313	-0.0487	0.1168
BRR-k_2	-0.0020	0.0081	-0.0023	0.0656	-0.0543	-0.1004	0.1203
BLR	-0.0534	0.0076	-0.0006	0.1463	-0.0810	-0.2511	0.4475
K-L	0.0082	0.0067	-0.0007	0.1434	-0.0831	-0.2445	0.2362
JKL	0.0099	0.0070	-0.0010	0.1301	-0.0816	-0.2154	0.2329

The results indicate a distinct disparity in MSE values among the estimators, highlighting their differing efficacy in the context of multicollinearity. The Bell Ridge Regression (k_1) estimator attained the minimal mean squared error, evidencing its superiority in diminishing variance and enhancing estimation accuracy relative to the other estimators. The Bell Ridge Regression (k_2) estimator demonstrated a comparable value, showing its relative efficiency. Conversely, the maximum likelihood estimator (MLE) exhibited a substantial mean squared error (MSE), indicating its vulnerability to multicollinearity and inherent instability.

The regression model was developed via the Bell Ridge Regression estimator method because of its minimal mean squared error (MSE) of 0.1168.

$$\log(\hat{\mu}_i) = \widehat{\beta_o} + \sum_1^n \widehat{\beta}_\iota \; X_{ij}$$

$$\log(\hat{\mu}_i) = \; -0.0003 + \; 0.0085\, x_{1i} - 0.0028 x_{2i} + 0.0324 x_{3i} - 0.0313 x_{4i} - 0.0487 x_{5i}$$

5. CONCLUSIONS

The study's principal conclusions, both theoretical and practical, are as follows:

1. The findings indicated that Bell regression model is appropriate for assessing count data characterized by overdispersion, since it offers more flexibility relative to conventional models, rendering it suitable for datasets where variance exceeds the arithmetic mean.
2. The comparison of estimating approaches revealed that shrinkage estimators significantly surpassed the maximum likelihood estimator based on the mean squared error (MSE) criterion, hence validating their efficacy in mitigating the impacts of multicollinearity.
3. The simulation study of the Bell Ridge Regression k_1 demonstrated superior performance compared to the other estimators, albeit marginally, as the Liu and Jacknifed estimators also yielded commendable results
4. The results of applying this to prostate cancer data indicate that the Bell Ridge regression estimator at the shrinkage parameter (k_1) achieved the lowest mean squared error (MSE) value (0.1168)compared to the maximum likelihood estimator (MLE)(2.0472) and the other estimators, demonstrating its superiority in handling multicollinearity and providing more stable parameter estimates. The low MSE value reflects an effective balance between minimizing variance and controlling for bias. Consequently, the model

provides better predictive accuracy for the number of affected lymph nodes, supporting a more reliable clinical assessment of disease progression.

Conflict of interest

The author declares that there is no conflict of interest regarding the publication of this research.

Consent for publications

The author confirms that the final version of the manuscript has been read and approved for publication.

Availability of data and material

All data used in this study are included within the manuscript, and no additional data are required.

Authors' contributions

The author was solely responsible for the conception and design of the study, data analysis, interpretation of results, and writing of the manuscript.

The author conducted the research, while the supervisor contributed to reviewing, guiding, and revising the manuscript. All authors should write their part in designing the idea, doing, analyzing and writing the article.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgement:

The author gratefully acknowledges all those who provided support and encouragement throughout this study.

6. REFERENCES

[1] Nelder, J. A., Wedderburn, R. W. M. (1972) Generalized linear models.*Journal of the Royal Statistical Society*: Series A, 135(3), 370–384. https://doi.org/10.2307/2344614

[2] Harris, T., Hilbe, J. M., Hardin, J. W. (2014) Modeling count data with generalized distributions . *The Stata Journal*, 14(3), 562–579. doi: 10.1177/1536867X1401400306.

[3] Castellares, F., Ferrari, S. L. P., Lemonte, A. J. (2018) On the Bell distribution and its associated regression model for count data . Applied Mathematical Modelling, 56, 172–185. doi: 10.1016/j.apm.2017.12.014.

[4] Myers, R. H., Montgomery, D. C., Vining, G. G., et al. (2012)Generalized linear models: With applications in engineering and the sciences, John Wiley & Sons. DOI:10.1002/9780470556986

[5] Corless, R. M., Gonnet, G. H., Hare, D. E. G., et al. (1996) On the Lambert W function *Advances in Computational Mathematics*, 5(1), 329–359. doi: 10.1007/BF02124750.

[6] Kyriazos, T. and Poga, M. (2023) Dealing with Multicolli nearity in Factor Analysis: The Problem, Detections, and Solutions. *Open Journal of Statistics*, 13, 404-424. https://doi.org/10.4236/ojs.2023.133020

[7] Majid, A., Amin, M., Akram, M. N. (2022) On the Liu estimation of Bell regression model in the presence of multicollinearity. *Journal of Statistical Computation and Simulation*, 92(2), 262–282. doi: 10.1080/00949655.2021.1955886.

[8] Hoerl, A. E., Kennard, R. W. (1970) Ridge regression: Biased estimation for nonorthogonal. problems *Technometrics*, 12(1), 55–67. doi: 10.1080/00401706.1970.10488634.

[9] Amin, M., Akram, M. N., Majid, A. (2023) On the estimation of Bell regression model using ridge estimator. *Communications in Statistics – Simulation and Computation*, 52(3), 854–867. doi: 10.1080/03610918.2020.1870694.

[10] Muniz, G., Kibria, B. M. G. (2009) On some ridge regression estimators: An empirical comparison, *Communications in Statistics – Simulation and Computation*, 38(3), 621–630. doi: 10.1080/03610910802592838.

[11] Kejian, L. (1993) A new class of biased estimate in linear regression. *Communications in Statistics – Theory and Methods*, 22(2), 393–402. doi: 10.1080/03610929308831027.

[12] Kibria, B. M. G., & Lukman, A. F. (2020). A New Ridge-Type Estimator for the Linear Regression Model: Simulations and Applications. Scientifica, 2020, 9758378. doi:10.1155/2020/9758378

[13] Shewa, G. A., Ugwuowo, F. I. (2022) Combating multicollinearity in Bell regression model. *Journal of the Nigerian Society of Physical Sciences*. doi: 10.46481/jnsps.2022.713.

[14] Abduljabbar, L. A. (2022) Jackknifed K–L estimator in Bell regression model.*Mathematics and Statistics Engineering Applications*, 71(3) https://share.google/v5B6RNJB1xhryXs75.

[15] Hinkley, D. V. (1977) Jackknifing in unbalanced. situations *Technometrics*, 19(3), 285–292.

[16] Zhu, Y., Wang, H. K., Zhang, H. L., Yao, X. D., Zhang, S. L., & Dai, B. (2015). Prostate cancer epidemiology and risk factors in Asia *Asian. Journal of Andrology*, 17(1), 48–52. https://doi.org/10.4103/1008-682X.13278

Journal of Al-Turath University College (JATUC)
Open Access Journal
ISSN: 2074-5621
DOI: https://doi.org/10.63964/JATUC.43.2.2026.50
https://jatuc.uoturath.edu.iq/

دراسة مقارنة لطرائق التقدير لأنموذج انحدار بيل بوجود التعدد الخطي

ساره سالم ساجت [1]*، ايمان محمد عبدالله [2]

[2,1] قسم الاحصاء، كلية الادارة والاقتصاد، جامعة بغداد، بغداد، العراق.

* *البريد الإلكتروني للتواصل:* sarah.salem2201@coadec.uobaghdad.edu.iq

الملخص

يُعد نموذج انحدار بيل إطارًا إحصائيًا حديثًا لتحليل بيانات العد، لا سيما في حالات فرط التشتت مقارنةً بنموذج أنحدار بواسون وانموذج انحدار ذي الحدين السالب. إلا أن وجود التعدد الخطي بين المتغيرات التفسيرية قد يؤدي إلى عدم استقرار تقديرات المعلمات عند استخدام مقدر الإمكان الأعظم (MLE)، مما ينعكس سلبًا على دقة التقدير. تناولت هذه الدراسة مشكلة التعدد الخطي في أنموذج انحدار بيل من خلال مقارنة عدد من طرائق التقدير، وهي: مقدر الإمكان الأعظم (MLE)، ومقدر انحدار بيل الحرفي (BRR)، ومقدر انحدار بيل ليو (BLR)، ومقدر K-L، ومقدر K-L المعتمد على الجكنايف. وقد تم تقييم كفاءة هذه المقدرات باستخدام معيار متوسط مربعات الخطأ (MSE).

أُجريت دراسة محاكاة بأسلوب مونت كارلو باستخدام أحجام عينات مختلفة ، وخمسة متغيرات تفسيرية ، ومستويات ارتباط مختلفة . وأظهرت نتائج المحاكاة أن مقدر انحدار بيل الحرف عند معلمة الانكماش k_1 حقق أقل قيم لمتوسط مربعات الخطأ (MSE) في أغلب الحالات مقارنةً ببقية المقدرات، مما يدل على تفوقه في ظل وجود التعدد الخطي. علاوةً على ذلك، تم تحليل بيانات حقيقية لمرضى سرطان البروستات، حيث أظهرت النتائج أن مقدر انحدار بيل الحرف تفوق على مقدر الإمكان الأعظم (MLE) وبقية المقدرات الأخرى من خلال تحقيق أقل قيمة لمتوسط مربعات الخطأ (MSE). ويعكس ذلك قدرته على توفير تقديرات أكثر استقرارًا ودقة، وتحسين موثوقية التنبؤ بالنتائج المرتبطة بالمرض، فضلًا عن كفاءته في معالجة مشكلتي التعدد الخطي وفرط التشتت في بيانات العد.

الكلمات المفتاحية

انحدار بيل؛ التعدد الخطي؛ انحدار ريدج؛ مقدّر K-L ؛ مقدّر K-L بطريقة جاك نايف.