Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v14n1p44

Almost unbiased ridge estimator in the count data regression models

By Abdulahad, Algamal

Published: 20 May 2021

Electronic Journal of Applied Statistical Analysis
Vol. 14, Issue 01, May 2021, 44-57
DOI: 10.1285/i20705948v14n1p44

Almost unbiased ridge estimator in the count data regression models

Fedaa Noeel Abdulahad[a] and Zakariya Yahya Algamal[*b]

[a]*Department of Mathematics, University of Al-Hamdaniya, Mosul, Iraq*
[c]*Department of Statistics and Informatics, University of Mosul, Mosul, Iraq*

Published: 20 May 2021

The ridge estimator has been consistently demonstrated to be an attractive shrinkage method to reduce the effects of multicollinearity. The Poisson regression negative binomial regression models are well-known model in application when the response variable is count data. However, it is known that multicollinearity negatively affects the variance of maximum likelihood estimator of the count regression coefficients. To address this problem, a count data ridge estimator has been proposed by numerous researchers. In this paper, an almost unbiased regression estimator is proposed and derived. Our Monte Carlo simulation results suggest that the proposed estimator can bring significant improvement relative to other existing estimators. In addition, the real application results demonstrate that the proposed estimator outperforms both negative binomial ridge regression and maximum likelihood estimators in terms of predictive performance.

keywords: Multicollinearity, ridge estimator, almost unbiased estimator, negative binomial regression model, Poisson regression model, Monte Carlo simulation.

1 Introduction

In regression modeling, data in the form of counts are usually common. Count data regression modeling has received much attention in medicine, behavioral sciences, psychology, and econometrics (Algamal, 2012; Asar and Genç, 2017; Coxe et al., 2009). The Poisson and negative binomial regression models are the most basic models under

*Corresponding author: zakariya.algamal@uomosul.edu.iq

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

count data regression models (Wang et al., 2014). The problem of overdispersion usually occurs in count data. Unlike Poisson regression model, negative binomial regression can handle the overdispersion issue (Cameron and Trivedi, 2013; Hilbe, 2011).

In dealing with the count data regression model, it is assumed that there is no correlation among the explanatory variables. In practice, however, this assumption often not holds, which leads to the problem of multicollinearity. In the presence of multicollinearity, when estimating the regression coefficients using the maximum likelihood (ML) method, the estimated coefficients are usually become unstable with a high variance, and therefore low statistical significance (Kibria et al., 2014; Alkhateeb and Algamal, 2020; Rashad and Algamal, 2019; Alobaidi et al., 2021; Al-Taweel and Algamal, 2020; Algamal, 2019). Numerous remedial methods have been proposed to overcome the problem of multicollinearity. The ridge regression method (Hoerl and Kennard, 1970) has been consistently demonstrated to be an attractive and alternative to the ML estimation method.

Ridge regression is a shrinkage method that shrinks all regression coefficients toward zero to reduce the large variance (Asar and Genç, 2015). This done by adding a positive amount to the diagonal of X^TX. As a result, the ridge estimator is biased but it guaranties a smaller mean squared error than the ML estimator.

In linear regression, the ridge estimator is defined as

$$\hat{\beta}_{Ridge} = (X^TX + kI)^{-1}X^Ty, \tag{1}$$

where y is an $n \times 1$ vector of observations of the response variable, $X = (x_1, ..., x_p)$ is an $n \times p$ known design matrix of explanatory variables, $\beta = (\beta_1, ..., \beta_p)$ is a $p \times 1$ vector of unknown regression coefficients, I is the identity matrix with dimension $p \times p$, and $k \geq 0$ represents the ridge parameter (shrinkage parameter). The ridge parameter, k, controls the shrinkage of β toward zero. The OLS estimator can be considered as a special estimator from Eq. (1) with $k = 0$. For larger value of k, the $\hat{\beta}_{Ridge}$ estimator yields greater shrinkage approaching zero (Algamal and Lee, 2015; Hoerl and Kennard, 1970).

2 Almost unbiased Poisson ridge regression estimator

The Poisson regression model is a popular tool when the dependent count data. There is a widespread usage in micro econometric dependent variable y_i is Poisson distributed, where $_i$=exp($x_i\beta$). Here, x_i is the i^{th} row of X which is an n*p data matrix with p independent variable and β can be estimated by maximizing the log-likelihood given by:

$$l\,(\beta) = \sum_{i=1}^{n} [-exp\,(x_i\beta)\ + (x_i\beta)\,y_i - log\,(y_i!)\ ] \tag{2}$$

The vector of coefficients using the ML is then estimated by solving the following equation

$$S\,(\beta) = \frac{\partial l(\beta)}{?\beta} = \sum_{i=1}^{n} [y_i - exp\,(x_i\beta)\ ]\,x_i = 0 \quad \ldots \quad (2) \tag{3}$$

Since eq.(2) is nonlinear in β, the solution of the score vector $S(\beta)$ is found by the iterative weighted least-square algorithm

$$\widehat{\beta}_{ML} = (X^T\hat{W}X)^{-1}X^T\hat{W}\hat{s} \tag{4}$$

where $\hat{W}$=diag[$\widehat{\mu}_i$] and $\hat{s}$ is a vector where the ith element equals to $\hat{s}_i = log\left(\widehat{\mu}_i\right) + \left(\frac{y_i-\widehat{\mu}_i}{\widehat{\mu}_i}\right)$. The maximum likelihood estimator asymptotically normally distributed with a covariance matrix that corresponds to the inverse of the matrix of the second derivative[1].

$$Cov\left(\widehat{\beta}_{ML}\right) = (X^T\hat{W}X)^{-1} \tag{5}$$

Furthermore, the MSE of Equation eq.(4) can be written as

$$MSE(\widehat{\beta}_{ML}) = E(\widehat{\beta}_{ML} - \beta)^T(\widehat{\beta}_{ML} - \beta) = tr[(X^T\hat{W}X)^{-1}] \tag{6}$$

$$= \sum_{j=1}^{J} \frac{1}{\lambda_j(\widehat{\beta}_{ML})} \tag{7}$$

where λ_j is the jtheigenvalue of the $(X^T\hat{W}X)$ matrix [2]. When the explanatory variables are highly correlated the weighted matrix of cross-products, $X^T\hat{W}X$ is ill-conditioned which leads to instability and high variance of the maximum likelihood estimator. In that situation it is very hard to interpret the estimated parameters since the vector of estimated coefficients is on average too long.

As a remedy to the problem caused by multicollinearty the Poisson ridge regression (PRR) method was proposed. The derivation of this method starts by noting that the ML method approximately minimizes the weighted sum of square error WSSE. Hence, $\widehat{\beta}_{ML}$ can be seen as the optimal estimator in a WSSE sense. If we choose another estimator, $\widehat{\beta}_{RR}$ of the parameter vector β we can write the WSSE of this estimator as

$$\Phi = (y - \widehat{\beta}_{RR})^T\left(y - \widehat{\beta}_{RR}\right)$$

$$= (y - X\widehat{\beta}_{ML})^T\left(y - X\widehat{\beta}_{ML}\right) + (\widehat{\beta}_{RR} - \widehat{\beta}_{ML})^T X^T\hat{W}X(\widehat{\beta}_{RR} - \widehat{\beta}_{ML})$$

$$= \Phi_{min} + \Phi(\widehat{\beta}_{RR}) \tag{8}$$

Where $\Phi(\widehat{\beta}_{RR})$ is the increase of WSSE when $\widehat{\beta}_{ML}$ is replaced by $\widehat{\beta}_{RR}$. To find the PRR estimator the length of $\widehat{\beta}_{RR}^T\widehat{\beta}_{RR}$ Should be minimized subject to the constraint $\Phi\left(\widehat{\beta}_{RR}\right) = \Phi_0$. As a Lagrangian problem this may be stated as:

$$minimize\ F = \widehat{\beta}_{RR}^T\widehat{\beta}_{RR} + (1/k)(\widehat{\beta}_{RR} - \widehat{\beta}_{ML})^T X^T\hat{W}X(\widehat{\beta}_{RR} - \widehat{\beta}_{ML} - \Phi_0) \tag{9}$$

Differentiating the above expression with respect to $\widehat{\beta}_{RR}$ and setting the result equal to zero and by solving the equation with respect to $\widehat{\beta}_{RR}$ we obtain the PRR estimator

$$\widehat{\beta}_{RR} = (X^T\hat{W}X + kI)^{-1}X^T\hat{W}X\widehat{\beta}_{ML} = Z\widehat{\beta}_{ML} \tag{10}$$

The MSE of this new estimator is:

$$E(\widehat{\beta}_{RR} - \beta)^T \left(\widehat{\beta}_{RR} - \beta\right) = \sum_{j=1}^{J} \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^{J} \frac{\alpha_j^2}{(\lambda_j + k)^2}$$

$$= \gamma_1(k) + \gamma_2(k) \quad (11)$$

Where α_j^2 defined as the jth element of $\gamma^T \widehat{\beta}_{ML}$ and $X^T \hat{W} X = \gamma^T \Lambda \gamma$, where $\Lambda = diag(\lambda_j)$ and γ be the eigenvector.

The AURE (almost unbiased regression estimator) in linear model is given by

$$\widehat{\beta}_{AURE} = (I - X^T X + kI)^{-2} k^{-2})\widehat{\beta}_{ML} \quad (12)$$

So, we proposed the AURE for the PRM as

$$\widehat{\beta}_{AUPRRE} = (I - X^T \hat{W} X + kI)^{-2} k^{-2})\widehat{\beta}_{ML} \quad (13)$$

Taking expectation of equation (13) on both sides we have

$$E\left(\widehat{\beta}_{AUPRRE}\right) = \left(I - X^T \hat{W} X + kI)^{-2} k^{-2}\right) E\left(\widehat{\beta}_{ML}\right)$$

$$= (I - X^T \hat{W} X + kI)^{-2} k^{-2})(X^T \hat{W} X)^{-1} X^T \hat{W} E(Y)$$

$$= (I - X^T \hat{W} X + kI)^{-2} k^{-2})(X^T \hat{W} X)^{-1} X^T \hat{W} X \beta$$

$$= \left(I - X^T \hat{W} X + kI)^{-2} k^{-2}\right) \beta \quad (14)$$

The bias of the AUPRRE can be found as

$$Bias\left(\widehat{\beta}_{AUPRRE}\right) = E\left(\widehat{\beta}_{AUPRRE}\right) - \beta$$

$$= \left(I - X^T \hat{W} X + kI)^{-2} k^{-2}\right) \beta - \beta$$

$$= -k^{-2}(X^T \hat{W} X + kI)^{-2} \beta$$

$$= -k^{-2} \sum_{j=1}^{J} \frac{\alpha_j}{(\lambda_j + k)^2} \quad (15)$$

The variance of the AUPRRE can be defined as

$$V\left(\widehat{\beta}_{AUPRRE}\right) = \left(I - X^T \hat{W} X + kI)^{-2} k^{-2}\right) V(\widehat{\beta})(I - X^T \hat{W} X + kI)^{-2} k^{-2})^T$$

$$= \left(I - X^T \hat{W} X + kI)^{-2} k^{-2}\right) (X^T \hat{W} X)^{-1} (I - X^T \hat{W} X + kI)^{-2} k^{-2})^T$$

$$= \frac{1}{\lambda_j} \sum_{j=1}^{J} (1 - \frac{k^2}{(\lambda_j + k)^2})^2 \quad (16)$$

Using Eq.(15) and Eq.(16), the EMSE of AUPRRE can be simplified as

$$EMSE\left(\widehat{\beta}_{AUPRRE}\right) = V\left(\widehat{\beta}_{AUPRRE}\right) + (Bias(\widehat{\beta}_{AUPRRE}))^2$$

$$= \frac{1}{\lambda_j}\sum_{j=1}^{J}(1-\frac{k^2}{(\lambda_j+k)^2})^2 + (-k^{-2}\sum_{j=1}^{J}\frac{\alpha_j}{(\lambda_j+k)^2})^2$$

$$= \frac{1}{\lambda_j}\sum_{j=1}^{J}\frac{(\lambda_j^2+2\lambda_j k)^2}{(\lambda_j+k)^4} + k^4\sum_{j=1}^{J}\frac{\alpha_j^2}{(\lambda_j+k)^4} \quad (17)$$

3 Almost unbiased negative Binomial ridge regression estimator

The negative binomial regression model is very popular in applied research when the dependent variable y_i becomes non-negative integers or counts distributed as NB($\mu_i, \mu_i + \delta\mu_i^2$) where $\mu_i = exp?(x_i\beta)$ such that x_i is the i^{th} row of the data matrix X which is a nxp data matrix with p explanatory variables, β is the coefficient vector with intercept and z_i is a random variable following the gamma distribution such that z_i (δ,δ), i=1, 2, 3, ..., n. The density function of the dependent variable y_i is

$$pr\left(y_i \mid x_i\right) = \frac{\Gamma\left(\alpha^{-1}+y_i\right)}{\Gamma\left(\alpha^{-1}\right)\Gamma\left(1+y_i\right)}(\frac{\alpha^{-1}}{\alpha^{-1}+\mu_i})^{\alpha^{-1}}(\frac{\mu_i}{\alpha^{-1}+\mu_i})^{y_i} \quad (18)$$

where the over dispersion parameter α is define as α=1/δ. The conditional mean and variance are given as follows

$$E\left(y_i \mid x_i\right) = \mu_i \qquad , \qquad V\left(y_i \mid x_i\right) = \mu_i(1+\alpha\mu_i) \quad (19)$$

This is the most commonly applied NB model and the estimate of β is usually found by maximizing the log likelihood

$$l\left(\alpha,\beta\right) = \sum_{i=1}^{n}\{\sum_{j=0}^{y-1} log\left(j+\alpha^{-1}\right) - log\left(y_i!\right) - \left(y_i-\alpha^{-1}\right)log\left(1+\alpha exp\left(x_i\beta\right)\right) +$$

$$+y_i log\left(\alpha\right) + y_i log?(exp\left(x_i\beta\right)\} \quad (20)$$

Since $ln\left[\frac{\Gamma(\alpha^{-1}+y_{i})}{\Gamma(\alpha^{-1})}\right] = \sum_{i=1}^{n}\left(j+\alpha^{-1}\right)$. The vector of coefficients using maximum likelihood estimation by solving the equation

$$S\left(\beta\right) = \frac{\partial l(\mu,y)}{?\beta} = \sum_{i=1}^{n}\frac{(y_i-\mu_i)}{1+\alpha\mu_i}x_i = 0 \quad (21)$$

Since the eq.(19) is nonlinear in β the solution of S(β) equal to zero is found by using the method of scoring.

$$\beta_r = \beta_{r-1} + I^{-1}\left(\beta_{r-1}\right)S(\beta_{r-1}) \quad (22)$$

where $S(\beta_{r-1})$ is the first derivative of the log likelihood evaluated at β_{r-1} and

$$I^{-1}\left(\beta_{r-1}\right) = E\left(\frac{\partial^2 l\left(X,\beta\right)}{?\beta?\beta'}\right) = X^T W X \tag{23}$$

where $W = diag\left[\frac{(\mu_i(\beta_{r-1}))}{1+\alpha_{r-1}\mu_i(\beta_{r-1})}\right]$. The final part of eq.(22) may be written as

$$X^T W X \beta_r = X^T W z(\beta_{r-1})$$

By define $z(\beta_{r-1})$ as a vector where the ith value equals $log\left(\mu_i\left(\beta_{r-1}\right)\right) + \frac{y_i-\mu_i(\beta_{r-1})}{\mu_i(\beta_{r-1})}$. Hence, the method of scoring may be written by

$$\beta_r = (X^T W X)^{-1} X^T W z\left(\beta_{r-1}\right) \tag{24}$$

which has the same form as the weighted least squares regression with the weighted matrix defined as W. However, both W and z depend on β_{r-1} , the current estimate of the parameter vector so the solution of eq.(23) has to be found iterative. This method is known as iteratively weighted least squares IWLS and in the final the maximum likelihood estimate of β denoted as β_{ML} is obtained

The covariance matrix of this estimator given by

$$Cov\left(\widehat{\beta}_{ML}\right) = (X^T W X)^{-1} \tag{25}$$

and the MSE is given by

$$\begin{aligned} MSE\left(\widehat{\beta}_{ML}\right) &= E(\widehat{\beta}_{ML} - \beta)^T(\widehat{\beta}_{ML} - \beta) \\ &= tr[(X^T \hat{W} X)^{-1}] \\ &= \sum_{j=1}^{J} \frac{1}{\lambda_j(\widehat{\beta}_{ML})} \end{aligned} \tag{26}$$

where λ_j is the jtheigenvalue of the $(X^T \hat{W} X)$ matrix. When the explanatory variables are highly correlated the weighted matrix of cross-products, $X^T \hat{W} X$ is ill-conditioned which leads to instability and high variance of the maximum likelihood estimator. In that situation it is very hard to interpret the estimated parameters since the vector of estimated coefficients is on average too long.

The maximum likelihood estimator of β is found by using IWLS method and it can therefore be seen as the optimal estimator in a weighted sum of squares error WSSE sense. If another estimator, $\widehat{\beta}$, of the parameter vector β is chosen, the WSSE of this estimator can be written as

$$\varphi = (y - \widehat{\beta})^T \left(y - \widehat{\beta}\right) = (y - X\beta_{ML})^T \left(y - X\beta_{ML}\right) + (\widehat{\beta} - \beta_{ML})^T X^T W X \left(\widehat{\beta} - \beta_{ML}\right) = \varphi_{min} + \varphi(\widehat{\beta}) \tag{27}$$

where $\varphi(\widehat{\beta})$ is the increase of the WSSE when β_{ML} is replaced by $\widehat{\beta}$. The negative binomial ridge regression NBRR estimator is found by minimizing the length of $\widehat{\beta}^T\widehat{\beta}$ subject to the constraint $\delta\left(\widehat{\beta}\right) = \delta_0$. This may be stated as a Lagrangian problem

$$Min.\ F = \widehat{\beta}^T\widehat{\beta} + (1/k)(\widehat{\beta} - \beta_{ML})^T X^T W X(\widehat{\beta} - \beta_{ML} - \delta_0) \tag{28}$$

Where k is the Lagrange multiplier. If the Lagrange is differentiated with respect to $\widehat{\beta}$ and if the result is set to zero the following equation is obtained

$$\frac{\partial F}{\partial \widehat{\beta}} = 2\ \widehat{\beta} + \left(\frac{1}{k}\right)\left(2X^T W X\ \widehat{\beta} - 2X^T W X\ \beta_{ML}\right) = 0 \tag{29}$$

Then the NBRR estimator can be obtained by solving Eq.(28)

$$\beta_{RR} = (X^T W X + kI)^{-1}\left(X^T W X \beta_{ML}\right) = Z\beta_{ML} \tag{30}$$

The MSE of this new estimator is:

$$E(\widehat{\beta}_{RR} - \beta)^T\left(\widehat{\beta}_{RR} - \beta\right) = \sum_{j=1}^{J}\frac{\lambda_j}{(\lambda_j + k)^2} + k^2\sum_{j=1}^{J}\frac{\alpha_j^2}{(_j + k)^2}$$

$$= \gamma_1(k) + \gamma_2(k) \tag{31}$$

Where α_j^2 defined as the jth element of $\gamma^T\widehat{\beta}_{ML}$ and $X^T\hat{W}X = \gamma^T\Lambda\gamma$, where $\Lambda = diag(\lambda_j)$ and γ be the eigenvector.

The AURE for the NBRM is defined as

$$\widehat{\beta}_{AUNBRRE} = (I - X^T\hat{W}X + kI)^{-2}k^{-2})\widehat{\beta}_{ML} \tag{32}$$

Taking expectation of equation (11) on both sides we have

$$E\left(\widehat{\beta}_{AUNBRRE}\right) = \left(I - X^T\hat{W}X + kI)^{-2}k^{-2}\right)E\left(\widehat{\beta}_{ML}\right)$$

$$= (I - X^T\hat{W}X + kI)^{-2}k^{-2})(X^T\hat{W}X)^{-1}X^T\hat{W}E(Y)$$

$$= (I - X^T\hat{W}X + kI)^{-2}k^{-2})(X^T\hat{W}X)^{-1}X^T\hat{W}X\beta$$

$$= \left(I - X^T\hat{W}X + kI)^{-2}k^{-2}\right)\beta \tag{33}$$

The bias of the AUNBRRE can be found as

$$Bias\left(\widehat{\beta}_{AUNBRRE}\right) = E\left(\widehat{\beta}_{AUNBRRE}\right) - \beta$$

$$= \left(I - X^T\hat{W}X + kI)^{-2}k^{-2}\right)\beta - \beta$$

$$= -k^{-2}(X^T\hat{W}X + kI)^{-2}\beta$$

$$= -k^{-2}\sum_{j=1}^{J}\frac{\alpha_j}{(\lambda_j+k)^2} \quad (34)$$

The variance of the AUNBRRE can be defined as

$$V\left(\widehat{\beta}_{AUNBRRE}\right) = \left(I - X^T\hat{W}X + kI)^{-2}k^{-2}\right)V(\widehat{\beta})(I - X^T\hat{W}X + kI)^{-2}k^{-2})^T$$

$$= \left(I - X^T\hat{W}X + kI)^{-2}k^{-2}\right)(X^T\hat{W}X)^{-1}(I - X^T\hat{W}X + kI)^{-2}k^{-2})^T$$

$$= \frac{1}{\lambda_j}\sum_{j=1}^{J}(1 - \frac{k^2}{(\lambda_j+k)^2})^2 \quad (35)$$

Using Eq.(34) and Eq.(35), the EMSE of AUNBRRE can be simplified as

$$EMSE\left(\widehat{\beta}_{AUNBRRE}\right) = V\left(\widehat{\beta}_{AUNBRRE}\right) + (Bias(\widehat{\beta}_{AUNBRRE}))^2$$

$$= \frac{1}{\lambda_j}\sum_{j=1}^{J}(1 - \frac{k^2}{(\lambda_j+k)^2})^2 + (-k^{-2}\sum_{j=1}^{J}\frac{\alpha_j}{(\lambda_j+k)^2})^2$$

$$= \frac{1}{\lambda_j}\sum_{j=1}^{J}\frac{\left(\lambda_j^2 + 2\lambda_j k\right)^2}{(\lambda_j+k)^4} + k^4\sum_{j=1}^{J}\frac{\alpha_j^2}{(\lambda_j+k)^4} \quad (36)$$

4 Simulation results

In this section, a Monte Carlo simulation experiment is used to examine the performance of the new estimator with different degrees of multicollinearity. The response variable of n observations is generated from Poisson regression model by

$$\theta_i = \exp(x_i^T\beta), \quad (37)$$

where $\beta = (\beta_0, \beta_1, ..., \beta_p)$ with $\sum_{j=1}^{p}\beta_j^2 = 1$ and $\beta_1 = \beta_2 = ... = \beta_p$ (Kibria, 2003). The response variable of n observations is generated from negative binomial regression model by using Eq.(37) when the value of α is chosen as 2.

The explanatory variables $x_i^T = (x_{i1}, x_{i2}, ..., x_{in})$ have been generated from the following formula

$$x_{ij} = (1-\rho^2)^{1l2}w_{ij} + \rho w_{ip}, \quad i = 1, 2, ..., n, \quad j = 1, 2, ..., p, \quad (38)$$

where ρ represents the correlation between the explanatory variables and w_{ij}'s are independent standard normal pseudo-random numbers. Because the sample size has direct impact on the prediction accuracy, three representative values of the sample size are considered: 30, 50 and 100. In addition, the number of the explanatory variables is considered as $p = 4$ and $p = 8$ because increasing the number of explanatory variables can lead to increase the MSE. Further, because we are interested in the effect of multi-collinearity, in which the degrees of correlation considered more important, three values of the pairwise correlation are considered with $\rho = \{0.90, 0.95, 0.99\}$. For a combination of these different values of n, p, and ρ the generated data is repeated 1000 times and the averaged mean squared errors (MSE) is calculated as

$$\mathbf{MSE}(\hat{\beta}) = \frac{1}{1000} \sum_{i=1}^{1000} (\hat{\beta} - \beta)^T (\hat{\beta} - \beta), \tag{39}$$

where $\hat{\beta}$ is the estimated coefficients for the used estimator. The efficiency of ridge estimator strongly depends on appropriately choosing the k parameter. To estimate the value of k for our new estimator, the most well-known used estimation methods is Hoerl and Kennard (1970)(HK), which is defined as

$$HK = \frac{1}{\hat{\alpha}_{\max}^2}, \quad j = 1, 2, ..., p, \tag{40}$$

The estimated MSE of Eq. (26) for ML, RR, AUPRE, NBRE, and AUNBRRE, for all the combination of n, p, and ρ, are respectively summarized in Tables 1 and 2. Several observations can be made. First, in terms of ρ values, there is increasing in the MSE values when the correlation degree increases regardless the value of n, p. However, the proposed estimators AUPRE and AUNBRRE perform better than ML, RR, and NBRE. For instance, in Table 1, when $p = 8$, $n = 50$, and $\rho = 0.99$, the MSE of AUPRE was about 77.022% and 11.897% lower than that of ML and RR, respectively.

Second, regarding the number of explanatory variables, it is easily seen that there is increasing in the MSE values when the p increasing from four variables to eight variables. Although this increasing can affected the quality of an estimator, AUPRE and AUNBRRE are achieved the lowest MSE comparing with ML, RR and NBRE, for different n, ρ.

Third, with respect to the value of n, The MSE values decreases when n increases, regardless the value of ρ, p. However, AUPRE and AUNBRRE still consistently outperforms RR, NBRE, and ML by providing the lowest MSE.

Table 1: MSE values for the Poisson regression model

		ρ	ML	PR	AUPRE
$p = 4$	$n = 30$	0.90	4.823	0.862	0.709
		0.95	5.451	1.093	0.942
		0.99	5.849	1.743	1.591
	$n = 50$	0.90	3.194	0.495	0.342
		0.95	4.269	0.767	0.614
		0.99	4.461	1.084	0.931
	$n = 100$	0.90	3.037	0.297	0.144
		0.95	3.247	0.421	0.268
		0.99	4.002	1.447	1.294
$p = 8$	$n = 30$	0.90	4.928	1.064	0.911
		0.95	5.547	1.295	1.142
		0.99	5.962	1.945	1.792
	$n = 50$	0.90	3.463	0.697	0.544
		0.95	4.606	0.969	0.816
		0.99	4.931	1.286	1.133
	$n = 100$	0.90	3.373	0.489	0.336
		0.95	3.648	0.613	0.461
		0.99	4.206	1.639	1.486

Table 2: MSE values for the negative binomial regression model

			ML	NBPR	AUNBRRE
		ρ			
$p = 4$	$n = 30$	0.90	5.094	1.133	0.981
		0.95	5.722	1.364	1.213
		0.99	6.12	2.014	1.862
	$n = 50$	0.90	3.465	0.766	0.613
		0.95	4.54	1.038	0.885
		0.99	4.732	1.355	1.202
	$n = 100$	0.90	3.308	0.568	0.415
		0.95	3.518	0.692	0.539
		0.99	4.273	1.718	1.565
$p = 8$	$n = 30$	0.90	5.199	1.335	1.182
		0.95	5.818	1.566	1.413
		0.99	6.233	2.216	2.063
	$n = 50$	0.90	3.734	0.968	0.815
		0.95	4.877	1.241	1.087
		0.99	5.202	1.557	1.404
	$n = 100$	0.90	3.644	0.76	0.607
		0.95	3.919	0.884	0.732
		0.99	4.477	1.91	1.757

5 Real application

To further investigate the usefulness of our new estimator, we apply the proposed estimator to the football Spanish La Liga, season 2016-2017. This data contains 20 teams. The response variable represents the number of won matches. The six considerable explanatory variables included the number of yellow cards (x_1), the number of red cards (x_2), the total number of substitutions (x_3), the number of matches with 2.5 goals on average (x_4), the number of matches that ended with goals (x_5), and the ratio of the goal scores to the number of matches (x_6).

First, the deviance test (Montgomery et al., 2015) is used to check whether the negative binomial regression model is fit well to this data or not. The result of the residual deviance test is equal to 8.651 with 14 degrees of freedom and the p-value is 0.822. It is indicated form this result that the negative binomial regression model fits very well to this data.

Second, to check whether there are relationships between the explanatory variables or not, it is obviously seen that there are correlations greater than 0.82 between x_1 and x_6, x_1 and x_4, x_2 and x_4, and, x_4 and x_6.

Third, to test the existence of multicollinearity, the eigenvalues of the matrix $X^T\hat{W}X$ are obtained as 997.247, 321.922, 170.541, 41.386, 22.694, and 2.054. The determined condition number $\mathbf{CN} = \sqrt{\lambda_{\max}/\lambda_{\min}}$ of the data is 22.034 indicating that the multicollinearity issue is exist.

The estimated negative binomial regression coefficients and MSE values for the ML, NBRE, and AUNBRRE estimators are listed in Table 3. According to Table 3, it is clearly seen that the AUNBRRE estimator shrinkages the value of the estimated coefficients efficiently". In terms of MSE, the AUNBRRE achieves the lowest MSE.

Table 3: The estimated coefficients and MSE values for the ML, NBRE, and AUNBRRE estimators

	ML	NBRE	AUNBRRE
$\hat{\beta}_1$	-1.219	-1.057	-0.616
$\hat{\beta}_2$	0.441	0.135	0.184
$\hat{\beta}_3$	0.575	0.127	0.106
$\hat{\beta}_4$	-3.476	-1.158	-0.122
$\hat{\beta}_5$	-2.432	-1.118	-0.108
$\hat{\beta}_6$	5.121	2.173	1.217
MSE	4.148	2.102	1.157

6 Conclusion

In this paper, a new estimator of ridge regression is proposed to overcome the multicollinearity problem in the Poisson regression and negative binomial regression models. According to Monte Carlo simulation studies, the proposed estimator has better performance than maximum likelihood estimator and ordinary ridge estimator, in terms of MSE. Additionally, a real data application is also considered to illustrate benefits of using the new estimator in the context of negative binomial regression model. The superiority of the new estimator based on the resulting MSE was observed and it was shown that the results are consistent with Monte Carlo simulation results.

7 Acknowledgment

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

References

Al-Taweel, Y. and Algamal, Z. (2020). Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model. *Periodicals of Engineering and Natural Sciences*, 8(1):248–255.

Algamal, Z. (2019). Variable selection in count data regression model based on firefly algorithm. *Statistics, Optimization & Information Computing*, 7(2):520–529.

Algamal, Z. Y. (2012). Diagnostic in poisson regression models. *Electronic Journal of Applied Statistical Analysis*, 5(2):178–186.

Algamal, Z. Y. and Lee, M. H. (2015). Adjusted adaptive lasso in high-dimensional poisson regression model. *Modern Applied Science*, 9(4):170–177.

Alkhateeb, A. and Algamal, Z. (2020). Jackknifed liu-type estimator in poisson regression model. *Journal of The Iranian Statistical Society*, 19(1):21–37.

Alobaidi, N. N., Shamany, R. E., and Algamal, Z. Y. (2021). A new ridge estimator for the negative binomial regression model. *Thailand Statistician*, 19(1):116–125.

Asar, Y. and Genç, A. (2015). New shrinkage parameters for the liu-type logistic estimators. *Communications in Statistics - Simulation and Computation*, 45(3):1094–1103.

Asar, Y. and Genç, A. (2017). A new two-parameter estimator for the poisson regression model. *Iranian Journal of Science and Technology, Transactions A: Science*.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*. Cambridge university press, 2nd edition.

Coxe, S., West, S. G., and Aiken, L. S. (2009). The analysis of count data: a gentle introduction to poisson regression and its alternatives. *J Pers Assess*, 91(2):121–36.

Hilbe, J. M. (2011). *Negative binomial regression*. Cambridge University Press.

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

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

Kibria, B. M. G., Månsson, K., and Shukur, G. (2014). A simulation study of some biasing parameters for the ridge type estimation of poisson regression. *Communications in Statistics - Simulation and Computation*, 44(4):943–957.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2015). *Introduction to linear regression analysis*. John Wiley & Sons, New York.

Rashad, N. K. and Algamal, Z. Y. (2019). A new ridge estimator for the poisson regression model. *Iranian Journal of Science and Technology, Transactions A: Science*, 43(6):2921–2928.

Wang, Z., Ma, S., Zappitelli, M., Parikh, C., Wang, C. Y., and Devarajan, P. (2014). Penalized count data regression with application to hospital stay after pediatric cardiac surgery. *Stat. Meth. Med. Res.*, In press.