Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 1, March 2020, pp.248-255

Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model

Younus Al-Taweel[1], Zakariya Algamal [2]
[1] Department of Mathematics, College of Education for Pure Science, University of Mosul
[2] Department of Statistics and Informatics, College of Computers Sciences and Mathematics, University of Mosul

ABSTRACT

Zero-inflated negative binomial regression (ZINB) models are commonly used for count data that show overdispersion and extra zeros. The correlation among variables of the count data leads to the presence of a multicollinearity problem. In this case, the maximum likelihood estimator (MLE) will not be an efficient estimator as the value of the mean squared error (MSE) will be large. Several alternative estimators, such as ridge estimators, have been proposed to solve the multicollinearity problem. In this paper, we propose an estimator called an almost unbiased ridge estimator for the ZINB model (AUZINBRE) to solve the multicollinearity problem in the correlated count data. The performance of the AUZINBRE is investigated using a Monte Carlo simulation study. The MSE is used as a measure to compare the results of the proposed estimators with those of the ridge estimators and the MLE. In addition, the AUZINBRE is applied to a real dataset.

Keywords: Multicollinearity, Zero-inflated Negative Binomial regression, Ridge estimator, almost Unbiased ridge estimator, Monte Carlo simulation.

Corresponding Author:

Younus Al-Taweel
Department of Mathematics
University of Mosul
Address. Albaker quarter, House no. 40, Mosul, Iraq
E-mail: younus.altaweel@uomosul.edu.iq

1. Introduction

In regression models for count data, the binomial model is very popular when the counts are bounded whereas the Poisson model is very popular when the counts are unbounded. The binomial model can involve explanatory variables that lead to a binomial regression model. This occurs when the counts exhibit more variability than the binomial model and so the overdispersion is modeled by supposing that the model parameter itself has a distribution. The negative binomial regression model is one of the most popular models for accounting the overdispersion when the Poisson mean has a gamma distribution. However, when extra zeros exist in the count data, the zero-inflated negative binomial (ZINB) regression model is used to account for the inflation of the extra zeros. The ZINB model can be seen as a mixture of a negative binomial distribution and a degenerate distribution at zero [1-3].

Suppose we have count data, $Y_i, i = 1, \dots, n$. Hence, Y_i are random variables that have a negative binomial distribution.

$$f(Y = y) = \frac{\Gamma(y+\tau)}{y!\Gamma(\tau)} \left(\tau + \frac{\tau}{\mu+\tau}\right)^{\tau} \left(\lambda + \frac{\tau}{\mu+\tau}\right)^{y}, y = 0,1, \dots; \mu, \tau > 0 \tag{1}$$

where μ is the expected value $\mu = \mathrm{E}(Y)$, τ represents the parameter that quantifies the overdispersion amount, and Y is the dependent variable. The variance of Y is $\mu + \mu^2/\tau$. The negative binomial distribution approaches to a Poisson distribution when $\tau \longrightarrow \infty$.

In some situations, however, a huge number of zeros may be included in the count data. Hence, a zero-inflated negative binomial (ZINB) regression model is used for such kind of count data [1, 2, 3]. In the presence of multicollinearity among the explanatory variables, the variance of the maximum likelihood estimator (MLE) will be inflated [3]. In order to overcome this multicollinearity problem, authors have proposed several other estimators. For example, [4] used a ridge regression (RE) for the negative binomial regression models. [3] proposed a Liu-type estimator for the negative binomial regression models. However, these estimators may have a large bias. In order to solve this problem, the almost unbiased ridge estimator (AURE) was proposed by [5] for linear regression models. Hence, in this paper, we propose the AURE for the ZINB (AUZINBRE) model in order to tackle the multicollinearity problem and decrease the variance of the MLE to obtain a reliable estimator. We demonstrate the performance of the AUZINBRE by a Monte Carlo simulation study with different sample sizes and different combinations of the correlation levels. Moreover, we also use a real dataset to show the performance of the AURE and compare it with those of other estimators.

This work is organized as follows. Section 2 reviews the methodology of the zero-inflated negative binomial regression model. In Section 3, the ridge estimator is presented for the ZINB model. In Section 4, we propose the almost unbiased ridge regression estimator for the ZINB model (AUZINBRE). In addition, several estimation methods are presented for obtaining the values of the almost unbiased ridge parameter. Section 5 investigates the behavior of the AUZINBRE using a simulation experiment. In Section 6, the performance of the AUZINBRE is also investigated using a real dataset. Finally, in Section 7, the conclusion is given.

2. Zero-inflated negative binomial regression models

Suppose we have observations that have two possible cases for each observation. If the first case occurs with probability π_i, the count is zero, whereas if the second case occurs with probability $1-\pi_i$, the counts will follow the negative binomial model. The ZINB distribution composes the logit distribution and the negative binomial distribution. Thus, the values of Y are 0, 1, 2, 3, and so on. Hence, for a ZINB random variable y, the probability function is written as

$$f(Y=y)=\begin{cases}\pi_i+(1-\pi_i)\left(1+\frac{\mu}{\tau}\right)^{-\tau} & if\ y=0\\ (1-\pi_i)\,\frac{\Gamma(y+\tau)}{y!\Gamma(\tau)}\left(1+\frac{\mu}{\tau}\right)^{-\tau}\left(1+\frac{\tau}{\mu}\right)^{y} & if\ y=1,2,\ldots\end{cases} \tag{2}$$

The ZINB distribution can be seen as a mixture distribution that assigns a probability π_i for extra zeroes and a probability $1-\pi_i$ for a negative binomial distribution. The mean of the ZINB distribution is $\mathrm{E}(Y)=(1-\pi)\mu$ and the variance is $\mathrm{Var}(Y)=(1-\pi)\mu(1+\pi\mu+\mu/\tau)$. This ZINB distribution approaches to the zero-inflated Poisson (ZIP) distribution when $\tau\longrightarrow\infty$ whereas it will be reduced to the negative binomial distribution when $\pi\longrightarrow\infty$. Moreover, the ZINB distribution will be reduced to the Poisson distribution when both $1/\tau\approx 0$ and $\pi_i\approx 0$ [6].

The ZINB regression model relates π and μ to explanatory variables such that

$$\log(\mu_i)=\mathbf{x}_i^T\boldsymbol{\beta} \qquad \mathrm{logit}(\pi)=\mathbf{z}_i^T\boldsymbol{\gamma}. \tag{3}$$

where $\mathbf{x}_i$ and $\mathbf{z}_i$ are d- and q-dimensional vectors of explanatory variables of the ith observation. The $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the corresponding vectors of regression coefficients, respectively. The MLE of the ZINB model parameters can be obtained using the iteratively weighted least squares (IWLS) algorithm [7].

3. Zero-inflated negative binomial ridge estimator

In the regression models, multicollinearity among the explanatory variables is an important problem that may lead to undesirable results. Hence, the MLE estimator may not be reliable as the eigenvalues will be small for the highly correlated explanatory variables [8, 9, 10]. Alternative estimators have been proposed by authors to solve the multicollinearity problem. For example, [4] proposed the ridge estimator (RE) to tackle the multicollinearity problem for linear regression. In the RE, a positive amount is added to the diagonal of $\mathbf{X}^T\mathbf{X}$. The RE was proposed by [11] for the ZIP model. The zero-inflated negative binomial ridge regression estimator (ZINBRE) is defined as

$$\widehat{\boldsymbol{\beta}}_{IGRE}=\left(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}+k\mathbf{I}\right)^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\widehat{\boldsymbol{\beta}}_{MLE} \tag{4}$$

where $\widehat{\boldsymbol{\beta}}_{MLE}$ is the MLE. The parameter $k \geq 0$ is called the ridge parameter. When the ridge parameter is zero, we have $\widehat{\boldsymbol{\beta}}_{\text{ZINBRE}} = \widehat{\boldsymbol{\beta}}_{MLE}$. However, we have $\left\|\widehat{\boldsymbol{\beta}}_{\text{ZINBRE}}\right\| < \left\|\widehat{\boldsymbol{\beta}}_{MLE}\right\|$ when $k > 0$ [12]. The matrix $\widehat{\mathbf{W}} = \operatorname{diag}(\hat{\mu}_i)$ where $\hat{\mu}_i = \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}$.

The mean squared error (MSE) of the MLE for the ZINB is defined as

$$\begin{aligned}\operatorname{MSE}(\widehat{\boldsymbol{\beta}}_{MLE}) &= \operatorname{E}(\widehat{\boldsymbol{\beta}}_{MLE} - \widehat{\boldsymbol{\beta}})^T(\widehat{\boldsymbol{\beta}}_{MLE} - \widehat{\boldsymbol{\beta}}) \\ &= \tau \textstyle\sum_{j=1}^{p} \frac{1}{\lambda_j} \end{aligned} \tag{5}$$

where $\lambda_j$is the eigenvalue of the matrix $\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}$. The MSE of the ZINBRE is given by

$$\begin{aligned}\operatorname{MSE}(\widehat{\boldsymbol{\beta}}_{ZINBRE}) &= \operatorname{E}(\widehat{\boldsymbol{\beta}}_{ZINBRE} - \widehat{\boldsymbol{\beta}})^T(\widehat{\boldsymbol{\beta}}_{ZINBRE} - \widehat{\boldsymbol{\beta}}) \\ &= \tau \textstyle\sum_{j=1}^{p} \frac{\lambda_j^2}{(\lambda_j+k)^2} + k^2 \sum_{j=1}^{p} \frac{\alpha_j^2}{(\lambda_j+k)^2} \end{aligned} \tag{6}$$

where $\alpha_i = \boldsymbol{\psi}^T\boldsymbol{\beta}$ and $\boldsymbol{\psi}$ is the eigenvector of the matrix $\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}$.

4. The almost unbiased zero-inflated negative binomial ridge estimator

In the ridge regression models, the RE may have a large bias when the k value is large. [5] proposed an alternative estimator, called almost unbiased ridge estimator (AURE), to tackle the multicollinearity problem in linear regression models. Hence, in this work, we present the almost unbiased ridge estimator for the zero-inflated negative binomial (AUZINBRE) model. The AUZINBRE can overcome the multicollinearity problem and is able to decrease the bias of the ZINBRE. The AUZINBRE is defined by

$$\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}} = \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)\widehat{\boldsymbol{\beta}}_{MLE}. \tag{7}$$

By having the expectation of equation (7), we have

$$\begin{aligned} E(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}}) &= \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)\operatorname{E}(\widehat{\boldsymbol{\beta}}_{MLE}) \\ &= \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{W}}E(\boldsymbol{y}) \\ &= \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}\boldsymbol{\beta} \\ &= \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-\mathbf{2}}k^2\right)\boldsymbol{\beta}. \end{aligned} \tag{8}$$

The bias of the AUZINBRE is obtained by

$$\begin{aligned} \operatorname{Bias}(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}}) &= \operatorname{E}(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}}) - \boldsymbol{\beta} \\ &= \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)\boldsymbol{\beta} - \boldsymbol{\beta} \\ &= -k^2(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}\boldsymbol{\beta} \\ &= -k^2 \textstyle\sum_{j=1}^{p} \frac{\alpha_j}{(\lambda_j+k)^2} \end{aligned} \tag{9}$$

The variance of the AUZINBRE is obtained by

$$\begin{aligned} Var(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}}) &= \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)\operatorname{Var}(\widehat{\boldsymbol{\beta}})\left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)^T \\ &= \left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)(\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X})^{-1}\hat{\tau}\left(\boldsymbol{I} - (\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X} + k\boldsymbol{I})^{-2}k^2\right)^T \\ &= \frac{\hat{\tau}}{\lambda_j}\textstyle\sum_{j=1}^{p}\left(1 - \frac{k^2}{(\lambda_j+k)^2}\right)^2 \end{aligned} \tag{10}$$

where $\hat{\tau}^2 = \sum_{j=1}^{p}(y_i - \hat{\mu}_i)^2 / (n - p - 1)$ [11]. The MSE of the AUZINBRE is calculated using equations (9) and (10)

$$\begin{aligned} \operatorname{MSE}(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}}) &= \operatorname{Var}(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}}) + \operatorname{Bias}(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}})^2 \\ &= \frac{\hat{\tau}}{\lambda_j}\textstyle\sum_{j=1}^{p}\left(1 - \frac{k^2}{(\lambda_j+k)^2}\right)^2 + \left(-k^2 \sum_{j=1}^{p} \frac{\alpha_j}{(\lambda_j+k)^2}\right)^2 \end{aligned}$$

$$= \frac{\hat{\tau}}{\lambda_j} \sum_{j=1}^{p} \frac{\left(\lambda_i^2 + 2\lambda_i k\right)^2}{\left(\lambda_j + k\right)^2} + k^4 \sum_{j=1}^{p} \frac{\alpha_i^2}{\left(\lambda_j + k\right)^2} \tag{11}$$

4.1 Obtaining the value of the parameter k

Several methods have been proposed to obtain estimated values of the ridge estimator, k, because no specific method is available for estimating k. In this paper, values of k for the AUZINBRE in the ZINB regression model were suggested from [13] and [14]. The estimated values of k are as follows

$$k_1 = \frac{\hat{\tau}^2}{\left(\prod_{i=1}^{p} \alpha_i^2\right)^{\frac{1}{p}}}, \quad k_2 = \text{median}\left(m_i^2\right), \quad k_3 = \frac{p\hat{\tau}^2}{\hat{\alpha}^T\hat{\alpha}} + \frac{1}{\left(\lambda_{max}\hat{\alpha}^T\hat{\alpha}\right)}, \quad k_4 = \frac{p\hat{\tau}^2}{\hat{\alpha}^T\hat{\alpha}} + \frac{1}{2\sqrt{\frac{\lambda_{max}}{\lambda_{min}}}},$$

where $\hat{\alpha}_i$ is the *i*th element of $\boldsymbol{\psi}^T\widehat{\boldsymbol{\beta}}_{MLE}$, $\boldsymbol{\psi}$ is the eigenvector of the $\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}$ matrix, $m_i = \sqrt{\frac{\hat{\tau}^2}{\alpha_i^2}}$, λ_{max} and λ_{min} are the maximum and minimum eigenvalues of the $\mathbf{X}^T\widehat{\mathbf{W}}\mathbf{X}$ matrix. [13] proposed k_1 and k_2 for the ridge parameter in the linear regression model. [14] proposed k_3 and k_4 as estimators for k in the linear regression model. Hence, we will investigate the performance of these four estimators for the AUZINBRE using the MSE as a measure and compare the results with those of the ZINBRE and MLE.

5. Monte Carlo Simulation study

In order to investigate the behavior of the AUZINBRE, a Monte Carlo (MC) simulation study is conducted. Using the MSE as a measure, the MSE values of the AUZINBRE are compared with those of the ZINBRE and MLE with different multicollinearity levels. The MSE measure is obtained by

$$\text{MSE}\left(\widehat{\boldsymbol{\beta}}_{\text{AUZINBRE}}\right) = \sum_{i=1}^{R} \frac{\left(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}\right)^T\left(\widehat{\boldsymbol{\beta}}_i - \widehat{\boldsymbol{\beta}}\right)}{R} \tag{12}$$

where $\widehat{\boldsymbol{\beta}}_i$ represents the *i*th simulated value of $\widehat{\boldsymbol{\beta}}$. We use $R = 1000$ to be the simulation number of the MC experiment.

5.1 The simulation design

The design of the explanatory variables, $\mathbf{x}_i^T = (x_{i1}, x_{i2}, \dots, x_{in})$ in the MC experiment was generated by the following formula

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}}\vartheta_{ij} + \rho\vartheta_{ip}, \quad i = 1, \dots, n, \quad j = 1, \dots, p \tag{13}$$

where ρ represents the correlation level between the explanatory variables and ϑ's are independent random variables that were simulated from the uniform distribution. We set the size of the explanatory variables to be $p = 2$ and $p = 4$. The ρ is the key point in the MC experiment, so it was set to be 0.85, 0.95 and 0.99.

A binary variable was then generated from the binomial distribution using $\theta_i = \frac{\exp(q_i\delta)}{1+\exp(q_i\delta)}$. The value of q_i was set to be 1 and δ contains the intercept term only. The value of the intercept δ was set to be 0, 1 and 2 as it affects the probability of having zeros and ones [11]. Then, we obtain the binary variables that have values of one from the Poisson distribution with $\mu_i = \exp\left(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p\right)$. The sum of the coefficient regression parameters $\boldsymbol{\beta}$ was assumed to be 1 and the intercept of the Poisson model was always set to be zero. The response variable, y, of the AUZINBRE model was generated using equation (2) with different sample sizes $n = 100, 150$ and 200.

5.2 The simulation results

In this section, the simulated results of the MSE, calculated using equation (13), are presented for the AUZINBRE, ZINBRE and MLE. The values of the MSE for the AUZINBRE, ZINBRE and MLE in the ZINB regression model are shown in Tables 1- 6 for the different estimators, k_1, k_2, k_3 and k_4 under various combinations of n, p and ρ.

Table 1. Estimated MSE when $p = 2$ and the intercept of logit $= 0$ for the estimators. The best values are in bold font

n	ρ	MLE	ZINBRE				AUZINBRE			
			k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
100	0.85	3.128	1.711	1.258	1.078	1.079	1.534	0.877	**0.668**	0.669
	0.90	4.336	1.756	1.385	1.160	1.161	1.594	1.029	**0.745**	0.746
	0.99	37.261	1.957	1.914	1.794	1.791	1.924	1.854	1.643	**1.639**
150	0.85	3.792	1.792	1.373	1.198	1.199	1.638	0.990	**0.769**	**0.769**
	0.90	5.263	1.833	1.518	1.299	1.299	1.701	1.189	**0.882**	0.883
	0.99	45.188	1.987	1.972	1.910	1.910	1.974	1.947	**1.831**	**1.831**
200	0.85	3.617	1.814	1.360	1.196	1.197	1.670	0.963	**0.755**	**0.755**
	0.90	4.995	1.852	1.511	1.303	1.303	1.731	1.171	**0.878**	0.879
	0.99	41.532	1.988	1.972	1.910	1.910	1.977	1.947	**1.830**	**1.830**

Table 2. Estimated MSE when $p = 4$ and the intercept of logit $= 0$ for the estimators. The best values are in bold font

n	ρ	MLE	ZINBRE				AUZINBRE			
			k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
100	0.85	4.510	3.852	3.748	3.389	3.389	3.739	3.547	**2.956**	**2.956**
	0.90	6.308	3.857	3.742	3.372	3.367	3.740	3.528	2.917	**2.914**
	0.99	54.893	3.840	3.723	3.123	3.124	3.704	3.490	**2.531**	**2.531**
150	0.85	4.985	3.903	3.756	3.486	3.486	3.816	3.544	**3.070**	**3.070**
	0.90	7.106	3.906	3.763	3.453	3.453	3.820	3.555	**3.013**	**3.013**
	0.99	65.437	3.926	3.871	3.271	3.271	3.856	3.750	**2.733**	**2.733**
200	0.85	5.275	3.916	3.786	3.594	3.594	3.837	3.592	**3.248**	**3.248**
	0.90	7.409	3.911	3.774	3.554	3.554	3.827	3.571	**3.176**	3.177
	0.99	64.578	3.915	3.774	3.334	3.334	3.834	3.698	**2.811**	2.812

Table 3. Estimated MSE when $p = 2$ and the intercept of logit $= 1$ for the estimators. The best values are in bold font

n	ρ	MLE	ZINBRE				AUZINBRE			
			k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
100	0.85	4.356	1.945	1.951	1.917	1.917	1.892	1.904	**1.839**	**1.839**
	0.90	5.767	1.946	1.956	1.917	1.917	1.894	1.913	**1.839**	**1.839**
	0.99	43.535	1.971	1.982	1.951	1.951	1.947	1.966	**1.911**	**1.911**
150	0.85	4.961	1.946	1.956	1.923	1.923	1.894	1.913	**1.850**	**1.850**
	0.90	6.589	1.949	1.962	1.924	1.924	1.899	1.925	**1.853**	**1.853**
	0.99	50.200	1.987	1.992	1.977	1.977	1.975	1.985	**1.956**	**1.956**
200	0.85	4.685	1.946	1.955	1.924	1.924	1.894	1.911	**1.852**	**1.852**
	0.90	6.122	1.949	1.961	1.926	1.926	1.900	1.923	**1.856**	**1.856**
	0.99	45.357	1.988	1.993	1.979	1.979	1.976	1.986	**1.958**	**1.958**

Table 4. Estimated MSE when $p = 4$ and the intercept of logit $= 1$ for the estimators. The best values are in bold font

n	ρ	MLE	ZINBRE				AUZINBRE			
			k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
100	0.85	8.287	3.987	3.984	3.972	3.972	3.974	3.969	**3.945**	**3.945**
	0.90	11.524	3.985	3.982	3.967	3.967	3.969	3.964	**3.934**	**3.934**
	0.99	101.106	3.973	3.980	3.916	3.916	3.946	3.960	3.836	**3.83**
150	0.85	6.639	3.991	3.989	3.980	3.980	3.983	3.979	**3.960**	**3.960**
	0.90	8.961	3.990	3.987	3.976	3.976	3.979	3.973	**3.953**	**3.953**
	0.99	76.004	3.974	3.982	3.923	3.923	3.949	3.965	**3.849**	**3.849**
200	0.85	7.820	3.993	3.991	3.979	3.979	3.986	3.982	**3.958**	**3.958**
	0.90	10.105	3.992	3.990	3.977	3.977	3.984	3.980	**3.953**	3.953
	0.99	86.143	3.983	3.989	3.938	3.939	3.968	3.978	**3.880**	3.881

Table 5: Estimated MSE when $p = 2$ and the intercept of logit $= 2$ for the estimators. The best values are in bold font.

n	ρ	MLE	ZINBRE				AUZINBRE			
			k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
100	0.85	9.491	2.032	2.025	2.034	2.007	2.042	2.036	2.046	**2.008**
	0.90	11.893	2.032	2.022	2.035	2.006	2.042	2.032	2.044	**2.007**
	0.99	72.709	2.060	2.063	2.065	1.990	2.067	2.072	2.098	**1.986**
150	0.85	10.008	2.003	2.010	2.020	2.006	2.002	2.015	2.023	**2.005**
	0.90	11.500	2.014	2.014	2.025	2.005	2.016	2.019	2.031	**2.005**
	0.99	71.568	2.047	2.043	2.072	2.006	2.065	2.052	2.135	**2.009**
200	0.85	8.373	1.997	2.001	1.997	1.995	1.990	1.997	1.991	**1.988**
	0.90	9.937	1.993	1.997	1.994	1.993	1.995	1.993	**1.986**	**1.986**
	0.99	56.805	2.000	2.002	2.001	1.996	1.996	2.002	1.996	**1.991**

Table 6: Estimated MSE when $p = 4$ and the intercept of logit $= 2$ for the estimators. The best values are in bold font

n	ρ	MLE	ZINBRE				AUZINBRE			
			k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
100	0.85	41.720	3.993	3.992	3.965	3.974	3.986	3.985	**3.943**	3.953
	0.90	42.297	4.113	4.022	4.346	3.983	4.265	4.052	4.605	**3.973**
	0.99	72.709	4.377	4.572	5.389	3.970	4.816	5.106	6.419	**3.943**
150	0.85	11.216	3.998	3.997	3.994	3.994	3.996	3.995	**3.989**	**3.989**
	0.90	13.910	3.998	3.997	3.996	3.995	3.997	3.995	**3.991**	**3.991**
	0.99	95.220	3.991	3.993	3.984	3.985	3.983	3.988	**3.970**	3.971
200	0.85	12.066	3.998	3.998	3.990	3.991	3.996	3.995	**3.982**	3.983
	0.90	15.637	3.997	3.997	3.985	3.987	3.994	3.994	**3.974**	3.976
	0.99	84.436	3.986	3.987	3.977	3.983	3.976	3.988	**3.961**	3.968

We can see from Tables 1- 6 that as the level of the multicollinearity, ρ, increases and fixed number of n and p, the values of the MLE increase. Thus, it has a negative impact on the MLE estimator. On the other hand, this is not always true for the AUZINBRE and ZINBRE where it can be shown in many cases that increasing the level of multicollinearity has a positive impact on the AUZINBRE and ZINBRE. For the number of explanatory variables p, we can see that when p increases, the MSE of all estimators increases with fixed values of ρ and n.

All the k estimators of the AUZINBRE are better than the corresponded k estimators of the ZINBRE as they have smaller values of the MSE. In contrast, the MLE has the largest values of the MSE. Among the k

estimators of the AUZINBRE, the estimators k_3 and k_4 outperform the estimators k_1 and k_2 as they have smaller MSE values.

It can be concluded from the simulation study that the MSE of AUZINBRE is always smaller than those of the ZINBRE and the MLE. Moreover, the AUZINBRE with the k_3 and k_4 improved the performance of the AUZINBRE compared with the ZINBRE and the MLE in all of the cases. All the selection methods of k are superior to the MLE in terms of MSE. Furthermore, k_3 and k_4 are the optimal estimation methods for k of the AUZINBRE. In contrast, the values of the MLE estimator are the poorest compared with the other estimators.

6. Real data application

In this section, we consider the wildlife fish dataset [15]. The biologists of the state wildlife want to model the number of fish caught by fishermen at a state park. Some visitors did not catch any fish so there are excess zeros in the data. Some visitors do not fish, but there is no data on whether a person fished or not.

The wildlife fish dataset consists of 250 groups that went to a park. The people in each group were asked the following questions: how many fish did they catch (count), how many children were in the group (child), how many people were in the group (persons) and whether they brought a camper to the park or not (camper). The response variable is the number of fish that were caught and it depends on 5 variables as described in Table 7.

Table 7: The description of the explanatory variables of the wildlife fish data.

Variable names	Description
nofish	represents whether the trip was not just for fishing, 0 if no 1 and if yes.
livebait	represents whether live bait was used or not, 0 if no and 1 if yes.
camper	represents whether or not they brought a camper.
persons	represents how many total persons on the trip.
child	represents how many children present

We fitted the ZINB regression model to the wildlife fish dataset. Then, we calculated the estimators AUZINBRE, ZINBRE and MLE. The MSE values and the coefficient estimated values of the ZINB model for the wildlife fish dataset for different estimators are presented in Table 8. It can be seen that the value of the MSE of the AUZINBRE is the smallest in comparison with the ZINBRE and the MLE. Furthermore, the k_3 and then the k_4 estimators of the ridge parameter have the best performance among the other AUZINBRE estimators, k_1 and k_2 as they have small MSE values.

Table 8: The estimated coefficient parameters and the estimated MSE for the AUZINBRE, ZINBRE and MLE.

	MLE	ZINBRE				AUZINBRE			
		k_1	k_2	k_3	k_4	k_1	k_2	k_3	k_4
Intercept	0.502	0.001	0.001	0.001	0.001	0.001	0.002	0.001	0.002
nofish	-1.691	-0.055	-1.687	-0.128	-0.127	-0.110	-1.691	-0.248	-0.246
livebait	-0.087	0.014	-0.083	0.032	0.032	0.028	-0.087	0.061	0.061
camper	0.090	0.040	0.090	0.067	0.067	0.068	0.090	0.100	0.100
persons	-0.015	-0.090	-0.015	-0.115	-0.115	-0.129	-0.015	-0.128	-0.128
child	0.005	-0.114	0.004	-0.174	-0.173	-0.182	0.005	-0.236	-0.235
MSE	332.897	138.653	134.252	121.318	121.495	131.067	119.828	94.422	94.680

7. Conclusions

This study proposed an almost unbiased ridge regression estimator for the zero-inflated negative binomial regression model to tackle the multicollinearity problem. The proposed estimator was able to overcome the inflation problem of the maximum likelihood estimation method for estimating the parameters of the ZINB model. A Monte Carlo simulation experiment was conducted to investigate the behavior of the proposed estimator using the MSE measure. Furthermore, a real dataset was used to see the performance of the proposed estimator. The results showed that the performance of the AUZINBRE is better than that of the MLE and ZINBRE as the MSE values of the AUZINBRE were smaller than those of the other estimators when multicollinearity exists.

For the AUZINBRE, the performance of the k_3 and k_4 estimators were better than that of the k_1 and k_2 as their values of the MSE were smaller than those of the other ridge estimators. Therefore, we recommended the k_3 and k_4 for estimating the ridge regression parameter for the AUZINBRE model.

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

References

[1] A. C. Cameron and P. K. Trivedi, "Regression analysis of count data", vol. 53. Cambridge university press, 2013.

[2] A. M. Garay, E. M. Hashimoto, E. M. Ortega, and V. H. Lachos, "On estimation and influence diagnostics for zero-inflated negative binomial regression models", *Computational Statistics & Data Analysis*, vol. 55, no. 3, pp. 1304–1318, 2011.

[3] J. M. Hilbe, "Modeling Count Data". Cambridge University Press, 2014.

[4] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

[5] B. Singh, Y. Chaubey, and T. Dwivedi, "An almost unbiased ridge estimator", *Sankhyā: The Indian Journal of Statistics, Series B*, Vol. 48, No. 3, pp. 342–346, 1986.

[6] S. M. Mwalili, E. Lesaffre, and D. Declerck, "The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research", *Statistical methods in medical research*, vol. 17, no. 2, pp. 123–139, 2008.

[7] Y. Asar, "Liu-type negative binomial regression: A comparison of recent estimators and applications", *in Trends and Perspectives in Linear Statistical Inference*, Springer, 2018, pp. 23–39.

[8] A. Dorugade, "Adjusted ridge estimator and comparison with Kibria's method in linear regression", *Journal of the Association of Arab Universities for Basic and Applied Sciences*, vol. 21, no. 1, pp. 96–102, 2016.

[9] G. Liu and S. Piantadosi, "Ridge estimation in generalized linear models and proportional hazards regressions", *Communications in Statistics-Theory and Methods*, vol. 46, no. 23, pp. 11466–11479, 2017.

[10] M. J. Mackinnon and M. L. Puterman, "Collinearity in generalized linear models", *Communications in statistics-theory and methods*, vol. 18, no. 9, pp. 3463–3472, 1989.

[11] B. G. Kibria, K. Månsson, and G. Shukur, "Some ridge regression estimators for the zero-inflated Poisson model", *Journal of Applied Statistics*, vol. 40, no. 4, pp. 721–735, 2013.

[12] Z. Yahya Algamal, "Performance of ridge estimator in inverse Gaussian regression model", *Communications in Statistics-Theory and Methods*, vol. 48, no. 15, pp. 1–14, 2018.

[13] B. G. Kibria, "Performance of some new ridge regression estimators", *Communications in Statistics-Simulation and Computation*, vol. 32, no. 2, pp. 419–435, 2003.

[14] S. S. Bhat, "A comparative study on the performance of new ridge estimators", *Pakistan Journal of Statistics and Operation Research*, vol. 12, no. 2, pp. 317–325, 2016.

[15] S. E. Saffari and R. Adnan, "Parameter estimation on zero-inflated negative binomial regression with right truncated data", Sains Malaysiana. 41. 1483-1487, 2012.