AIMS Mathematics

AIMS Mathematics, 10(10): 23169–23186.
DOI: 10.3934/math.20251028
Received: 21 June 2025
Revised: 10 September 2025
Accepted: 29 September 2025
Published: 13 October 2025

https://www.aimspress.com/journal/Math

Research article

Mitigating multicollinearity in zero-inflated negative binomial regression using the modified Kibria-Lukman estimator

Masad A. Alrasheedi[1], Adewale F. Lukman[2], Rasha A. Farghali[3] and Asamh Saleh M. Al Luhayb[4,*]

[1] Department of Management Information Systems, College of Business Administration, Taibah University, Madinah, Saudi Arabia
[2] Department of Mathematics and Statistics, University of North Dakota, Grand Forks, North Dakota 58202, USA
[3] Department of Mathematics, Insurance and Applied Statistics, Helwan University, Cairo 11795, Egypt
[4] Department of Mathematics, College of Science, Qassim University, P.O. Box 6644, Buraydah 51452, Saudi Arabia

* **Correspondence:** Email: a.alluhayb@qu.edu.sa.

Abstract: Multicollinearity presents a significant challenge in zero-inflated negative binomial (ZINB) regression, leading to unstable maximum likelihood estimates (MLEs) and inflated prediction errors. To address this issue, we investigated the performance of the Kibria-Lukman estimator (ZINB-KLE) and proposed a modified Kibria-Lukman estimator (ZINB-MKLE) that introduces an enhanced bias-adjustment mechanism for improved coefficient stability. Using extensive Monte Carlo simulations under varying degrees of multicollinearity and overdispersion, we demonstrated that the ZINB-MKLE consistently achieves substantially lower scalar mean squared error (SMSE) than MLEs, ZINB-KLEs, and other competing estimators. Application to the Blood Transfusion dataset further confirmed the practical advantages of the ZINB-MKLE, yielding an SMSE of 1.8568 compared to 14,638.75 for the MLE and 685.81 for the ZINB-KLE, highlighting dramatic improvements in predictive accuracy. These findings establish the ZINB-MKLE as a robust and efficient alternative for handling multicollinearity in zero-inflated regression models, with broad implications for statistical modeling in biomedical, epidemiological, and other applied data settings.

Keywords: zero-inflated, negative binomial, multicollinearity, ridge regression, modified

Kibria-Lukman estimator, simulation
Mathematics Subject Classification: 62J05, 62J07, 62J12

1. Introduction

Modeling count data has traditionally relied on Poisson and negative binomial (NB) regression frameworks, particularly in applications involving non-negative integer outcomes such as the number of hospital visits, traffic accidents, or manufacturing defects [1,2]. The Poisson model provides efficient estimation under the equidispersion assumption, where the conditional mean and variance are approximately equal. The NB model extends this framework by introducing a dispersion parameter, making it better suited for data exhibiting overdispersion—situations where the variance exceeds the mean. Despite their widespread adoption, both models perform poorly when confronted with excessive zero counts, a phenomenon frequently observed in healthcare, ecology, and transportation studies [3,4]. This limitation has led to the development of zero-inflated regression models, which explicitly account for structural zeros while maintaining the interpretability of conventional count frameworks.

When datasets exhibit a disproportionately high number of zeros, standard Poisson and NB models often yield biased estimates and fail to capture the underlying data-generating process. To address this issue, zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) models have been proposed [5–7]. These models assume a two-component structure: a binary process that governs the occurrence of structural zeros and a secondary count-generating process, typically modeled via Poisson or NB regression. Among these, the ZINB model has become especially prominent for analyzing overdispersed count data where zero inflation is prevalent, demonstrating effectiveness in fields such as healthcare utilization, highway safety, and public health surveillance [7–9].

Despite their flexibility, ZINB models are highly sensitive to multicollinearity among predictors. Strong correlations among covariates inflate the variances of the estimated parameters, resulting in unstable inference and unreliable predictions. The conventional maximum likelihood estimator (MLE), though commonly employed for ZINB models, deteriorates under severe multicollinearity, necessitating the development of alternative estimation techniques [10–12]. In recent years, several biased estimators have been proposed to improve estimation stability in the presence of multicollinearity within ZINB regression. These include the ridge estimator, Stein-type shrinkage estimators, Kibria-Lukman estimator (KLE), and various ridge-type extensions, as demonstrated by Al-Taweel and Algamal [5] and Akram et al. [7–9]. While these approaches have shown promise, there remains significant room for improving estimation accuracy and predictive performance.

To address these limitations, we propose the modified Kibria-Lukman estimator (MKLE) within the ZINB framework. This estimator leverages recent advances in shrinkage-based estimation while incorporating additional bias-adjustment mechanisms to enhance robustness under multicollinearity, zero inflation, and overdispersion. The MKLE effectively combines variance stabilization with improved interpretability, offering a theoretically sound and computationally efficient alternative to existing estimators. The performance of MKLE is evaluated through an extensive Monte Carlo simulation study across varying degrees of multicollinearity and zero inflation. Additionally, we demonstrate its practical utility via an empirical application to the Blood Transfusion dataset, where

MKLE consistently achieves lower mean squared error (MSE) compared to competing estimators, establishing its value for modeling complex count data.

The remainder of this article is organized as follows. Section 2 introduces the ZINB model, develops the MKLE, and provides theoretical comparisons to existing estimators. Section 3 presents the results of simulation studies assessing estimator performance under different conditions. Section 4 illustrates the application of MKLE using real-world data. Section 5 concludes with a discussion of the findings, practical implications, and potential directions for future research.

2. Zero-inflated negative binomial regression model (ZINBRM)

The zero-inflated negative binomial (ZINB) regression model is a flexible framework commonly used for modeling count data exhibiting both over-dispersion and a prevalence of zero counts. The model assumes that the data arise from two underlying processes: one that structurally generates zeros and another that produces counts following a negative binomial distribution. This dual-process formulation allows the ZINB model to effectively accommodate heterogeneous zero-inflation while capturing variability in the nonzero counts.

2.1. Model specifications

Consider a response variable $y_i, i = 1,2,3, \dots, n,$ that follows a zero-inflated negative binomial distribution. The probability density function is defined as:

$$P(y_i; \phi_i) = \begin{cases} \phi_i + (1-\phi_i) f(y_i = 0), & \text{if } y_i = 0, \\ (1-\phi_i) f(y_i), & \text{if } y_i > 0, \end{cases} \tag{2.1}$$

where ϕ_i is the probability of an excess zero, and $f(y_i)$ is the negative binomial distribution given by:

$$f(y_i) = P(y_i; \mu_i, \theta) = \frac{\Gamma(y_i + \theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_i + 1)} \left(\frac{1}{1+\theta\mu_i}\right)^{\theta^{-1}} \left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)^{y_i},$$

such that $\mu_i > 0$ is the mean parameter of the negative binomial distribution, $\theta > 0$ is the dispersion parameter, and $\Gamma(.)$ denotes the gamma function. The expected value and variance of y_i in Eq (2.1) are given by

$$E(y_i) = (1-\phi_i)\mu_i, \tag{2.2}$$

$$V(y_i) = \mu_i(1-\phi_i)\big(1 + \mu_i(\phi_i + \theta)\big). \tag{2.3}$$

The ZINB model links ϕ_i and μ_i to explanatory variables through appropriate link functions: $\log(\mu_i) = x_i'\beta$, $logit(\phi_i) = z_i'\gamma$, where x_i and z_i are vectors of predictors for the count and zero-inflation components, respectively. β and γ are the corresponding regression vectors. Note that the *z*'s and the *x*'s may or may not include terms in common. Maximum likelihood estimation (MLE) is commonly used to estimate the regression parameters. The log-likelihood function for a sample $\{y_1, y_2, \dots, y_n\}$ is expressed as:

$$\ell(\beta,\gamma,\theta) = \sum_{i=1}^{n}\left[\begin{array}{c} D_i log\left(\phi_i + (1-\phi_i)(1+\theta\mu_i)^{-\theta^{-1}}\right) + (1-D_i) \\ \left\{\log(1-\phi_i) + \log\left[\frac{\Gamma(y_i+\theta^{-1})}{\Gamma(\theta^{-1})\Gamma(y_i+1)}\right] + \theta^{-1}log(1+\theta\mu_i) + y_i log\left(\frac{\theta\mu_i}{1+\theta\mu_i}\right)\right\}\end{array}\right], \quad (2.4)$$

where $D_i = 1$ if $y_i = 0$ and $D_i = 0$ otherwise. The parameters are iteratively updated using the expectation-maximization (EM) algorithm to ensure numerical stability and convergence [13–15]. We compute the partial derivatives of the log-likelihood function to each of the parameters of interest. The EM algorithm maximizes this likelihood iteratively in two steps: the E-step and the M-step.

E-step. The expected value of D_i given the observed data and current parameter $\left(\beta^{(t)},\gamma^{(t)},\theta^{(t)}\right)$ is computed as

$$E\left[D_i \middle| y_i,\beta^{(t)},\gamma^{(t)},\theta^{(t)}\right] = \begin{cases} \dfrac{\phi_i^{(t)}}{\phi_i^{(t)} + (1-\phi_i^{(t)})(1+\theta^{(t)}\mu_i^{(t)})^{\frac{-1}{\theta^{(t)}}}}, & y_i = 0, \\ 0, & y_i > 0. \end{cases}$$

M-step. Using the expected values from the E-step, the parameters (β, γ, θ) are updated by maximizing the complete data log-likelihood. Specifically, we update β by solving

$$\frac{\partial\ell}{\partial\beta} = \sum_{i=1}^{n}(1-E[D_i])\frac{y_i-\mu_i}{\mu_i(1+\theta\mu_i)}.\frac{\partial\mu_i}{\partial\beta} = 0.$$

The EM algorithm iterates between the E-step and M-step until convergence, defined as: $\left\|\beta^{(t+1)} - \beta^{(t)}\right\| < \epsilon$, where ϵ is a small positive threshold. By iteratively applying these steps, the parameter estimates, β converge to values that maximize the likelihood for the ZINB model. Thus, the MLE of β is defined as

$$\hat{\beta}^{ZINB-MLE} = \left(X'\widehat{W}X\right)^{-1}X'\widehat{W}z, \quad (2.5)$$

where X is the design matrix for the count component of the ZINB model, $\hat{\widehat{W}}$ = diag($\mu_i(1+\theta\mu_i)$) is a diagonal weight matrix, with entries dependent on the conditional expectations from the E-step such that $\mu_i = \exp(x_i'\beta)$ and θ is the dispersion parameter. *z* is a vector of transformed responses: $z_i = \frac{y_i-\mu_i}{\mu_i(1+\theta\mu_i)} + x_i'\beta$.

Let $\alpha = L'\beta$ and $\Lambda = diag(\lambda_1,\lambda_2,\ldots,\lambda_q)$, where L is an orthogonal matrix whose columns are the eigenvectors of $X'\widehat{W}X$, and $\lambda_1 > \lambda_2 > \cdots > \lambda_q \geq 0$ are the eigenvalues of $X'\hat{\widehat{W}}X$. Here, α_j denotes the *j*-th component of $L'\hat{\beta}^{ZINB-MLE}$. The matrix mean squared error (MMSE) and scalar mean squared error (SMSE) can then be expressed as:

$$MMSE(\hat{\beta}^{ZINB-MLE}) = \hat{\theta}\left(X^T\widehat{W}X\right)^{-1}. \quad (2.6)$$

Hence, the SMSE is defined as:

$$SMSE(\hat{\alpha}^{ZINB-MLE}) = \hat{\theta}\sum_{j=1}^{q}\frac{1}{\lambda_j}. \quad (2.7)$$

2.2. ZINB-RRE

Ridge regression, also called L_2-penalized regression, was one of the earliest penalization techniques introduced for linear models. It was proposed by Hoerl and Kennard [16] and involves applying a penalty based on the L_2 norm of the regression parameters. The ridge regression estimator was extended to the (ZINB) regression model by Kibria et al. [6] and defined as follows:

$$\hat{\beta}^{ZINB-RRE} = \left(X'\widehat{W}X + kI\right)^{-1}\left(X'\widehat{W}X\right)\hat{\beta}^{ZINB-MLE},\ k > 0, \tag{2.8}$$

$$cov(\hat{\beta}^{ZINB-RRE}) = \hat{\theta}\left(X'\widehat{W}X + kI\right)^{-1}\left(X'\widehat{W}X\right)\left(X'\widehat{W}X + kI\right)^{-1},\ k > 0, \tag{2.9}$$

$$Bias\left(\hat{\beta}^{ZINB-RRE}\right) = \left[\left(X'\widehat{W}X + kI\right)^{-1}\left(X'\widehat{W}X\right) - I\right]\beta. \tag{2.10}$$

Let $\Lambda_{k^+}^{-1} = \left(X'\widehat{W}X + kI\right)^{-1}$and then

$$MMSE(\hat{\alpha}^{ZINB-RRE}) = \hat{\theta}\Lambda_{k^+}^{-1}\Lambda\Lambda_{k^+}^{-1} + k^2\Lambda_{k^+}^{-1}\alpha'\alpha\Lambda_{k^+}^{-1}, \tag{2.11}$$

$$SMSE(\hat{\alpha}^{ZINB-RRE}) = \hat{\theta}\sum_{j=1}^{q}\frac{\lambda_j}{(\lambda_j+k)^2} + k^2\sum_{j=1}^{q}\frac{\alpha_j^2}{(\lambda_j+k)^2}. \tag{2.12}$$

2.3. ZINB-JSE

The James-Stein estimator was extended to the ZINB regression model by Akram et al. [9] following the work of Amin et al. [17].

$$\hat{\beta}^{ZINB-JSE} = c\hat{\beta}^{ZINB-MLE}, 0 < c < 1, \tag{2.13}$$

where $c = \frac{\hat{\beta}^{ZINB-MLE\prime}\hat{\beta}^{ZINB-MLE}}{\hat{\beta}^{ZINB-MLE\prime}\hat{\beta}^{ZINB-MLE} + \hat{\phi}\, tr\left(cov\left(\hat{\beta}^{ZINB-MLE}\right)\right)}$,

$$cov\left(\hat{\beta}^{ZINB-JSE}\right) = c^2\, Cov\left(\hat{\beta}^{ZINB-MLE}\right) = c^2\,\hat{\theta}\left(X'\widehat{W}X\right)^{-1}, \tag{2.14}$$

$$Bias\left(\hat{\beta}^{ZINB-JSE}\right) = E\left(c\hat{\beta}^{ZINB-MLE}\right) - \beta = (c-1)\beta, \tag{2.15}$$

$$MMSE(\,\hat{\alpha}^{ZINB-JSE}) = c^2\hat{\theta}\Lambda^{-1} + (c-1)\alpha'\alpha(c-1), \tag{2.16}$$

$$SMSE(\,\hat{\alpha}^{ZINB-JSE}) = c^2\hat{\theta}\sum_{j=1}^{q}\frac{1}{\lambda_j} + (c-1)^2\sum_{j=1}^{q}\alpha_j^2. \tag{2.17}$$

2.4. ZINB-KLE

Kibria and Lukman [18] introduced the Kibria-Lukman estimator specifically for the linear regression model. Akram et al. [8] expanded it to the (ZINB) regression model in the following way:

$$\hat{\beta}^{ZINB-KLE} = \left(X'\widehat{W}X + kI\right)^{-1}(X'\widehat{W}X - kI)\hat{\beta}^{ZINB-MLE},\ k > 0, \tag{2.18}$$

$$cov(\hat{\beta}^{ZINB-KLE}) = \hat{\theta}\left(X'\widehat{W}X + kI\right)^{-1}(X'\widehat{W}X - kI)\left(X'\widehat{W}X\right)^{-1}(X'\widehat{W}X - kI)\left(X'\widehat{W}X + kI\right)^{-1}, \quad (2.19)$$

$$Bias\left(\hat{\beta}^{ZINB-KLE}\right) = \left[\left(X'\widehat{W}X + kI\right)^{-1}(X'\widehat{W}X - kI) - I\right]\beta. \quad (2.20)$$

Let $\Lambda_{k^-}^{-1} = \left(X'\widehat{W}X - kI\right)^{-1}$ and $\Lambda_{k^+}^{-1}$ is as defined before. Then:

$$MMSE(\hat{\alpha}^{ZINB-KLE}) = \hat{\theta}\Lambda_{k^+}^{-1}\Lambda_{k^-}^{-1}\Lambda^{-1}\Lambda_{k^-}^{-1}\Lambda_{k^+}^{-1} + 4k^2\Lambda_{k^+}^{-1}\alpha'\alpha\Lambda_{k^+}^{-1}, \quad (2.21)$$

$$SMSE(\hat{\alpha}^{ZINB-KLE}) = \hat{\theta}\sum_{j=1}^{q}\frac{(\lambda_j - k)^2}{(\lambda_j + k)^2\lambda_j} + 4k^2\sum_{j=1}^{q}\frac{\alpha_j^2}{(\lambda_j + k)^2}. \quad (2.22)$$

2.5. ZINB-LE

Liu [19] proposed a biased estimator for the linear regression model that competes effectively with the ridge regression estimator in addressing the effects of multicollinearity, building on the work of Hoerl and Kennard [16]. The Liu estimator was expanded to the ZINB regression model by Akram et al. [8] in the manner described below:

$$\hat{\beta}^{ZINB-LE} = \left(X'\widehat{W}X + I\right)^{-1}(X'\widehat{W}X + dI)\hat{\beta}^{ZINB-MLE}, \ 0 < d < 1, \quad (2.23)$$

$$cov(\hat{\beta}^{ZINB-LE}) = \hat{\theta}\left(X'\widehat{W}X + I\right)^{-1}(X'\widehat{W}X + dI)\left(X'\widehat{W}X\right)^{-1}(X'\widehat{W}X + dI)\left(X'\widehat{W}X + I\right)^{-1}, (2.24)$$

$$Bias\left(\hat{\beta}^{ZINB-LE}\right) = \left[\left(X'\widehat{W}X + I\right)^{-1}(X'\widehat{W}X + dI) - I\right]\beta. \quad (2.25)$$

Let $\Lambda_I^{-1} = \left(X'\widehat{W}X + I\right)^{-1}$ and $\Lambda_{d^+} = (X'\widehat{W}X + dI)$. Then

$$MMSE(\hat{\alpha}^{ZINB-LE}) = \hat{\theta}\Lambda_I^{-1}\Lambda_{d^+}\Lambda^{-1}\Lambda_{d^+}\Lambda_I^{-1} + (d-1)^2\Lambda_I^{-1}\alpha'\alpha\Lambda_I^{-1}, \quad (2.26)$$

$$SMSE(\hat{\alpha}^{ZINB-LE}) = \hat{\theta}\sum_{j=1}^{q}\frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2\lambda_j} + (d-1)^2\sum_{j=1}^{q}\frac{\alpha_j^2}{(\lambda_j + 1)^2}. \quad (2.27)$$

3. Proposed estimator

The modified Kibria-Lukman (MKL) estimator was first presented by Aladeitan et al. [20] for the Poisson regression model. Despite having only one biasing parameter, the (MKL) estimator outperforms most of the shrinkage estimators. In this paper, we propose the modified Kibria-Lukman (MKL) estimator by replacing the ZINB-MLE in the RHS of Eq (2.18) with the ridge estimator $\left(X'\widehat{W}X + kI\right)^{-1}\left(X'\widehat{W}X\right)\hat{\beta}^{ZINB-MLE}$. The ZINB-MKLE improves upon the ZINB-KLE by introducing a ridge-based preconditioning step, which applies stronger and more adaptive shrinkage to unstable directions caused by multicollinearity, thereby reducing variance inflation and improving estimation stability. Hence, the proposed method is defined as follows:

$$\hat{\beta}^{ZINB-MKLE} = \left(X'\widehat{W}X + kI\right)^{-1}(X'\widehat{W}X - kI)\left(X'\widehat{W}X + kI\right)^{-1}\left(X'\widehat{W}X\right)\hat{\beta}^{ZINB-MLE}, \ k > 0. (2.28)$$

The variance-covariance property of the proposed estimator is as follows:

$$cov(\hat{\beta}^{ZINB-\mathrm{MKLE}}) = \hat{\theta}(X'^{\widehat{W}}X + kI)^{-1}(X'^{\widehat{W}}X - kI)(X'^{\widehat{W}}X + kI)^{-1}$$

$$(X'\widehat{W}X)(X'^{\widehat{W}}X + kI)^{-1}(X'\widehat{W}X - kI)(X'\widehat{W}X + kI)^{-1}. \quad (2.29)$$

The bias of the proposed estimator is

$$Bias(\hat{\beta}^{ZINB-\mathrm{MKLE}}) = \left[(X'\widehat{W}X + kI)^{-1}(X'\widehat{W}X - kI)(X'\widehat{W}X + kI)^{-1}(X'\widehat{W}X) - I\right]\beta. \quad (2.30)$$

Let $\Lambda_{k^-}^{-1} = (X'\widehat{W}X - kI)^{-1}$ and $\Lambda_{k^+}^{-1}$ is as defined before. Then the MMSE is defined as follows:

$$MMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}}) = \hat{\phi}\Lambda_{k^+}^{-1}\Lambda_{k^-}\Lambda_{k^+}^{-1}\Lambda\Lambda_{k^+}^{-1}\Lambda_{k^-}\Lambda_{k^+}^{-1} + k^2\Lambda_{k^+}^{-2}(3X'\widehat{W}X + kI)\alpha'\alpha(3X'\widehat{W}X + kI)\Lambda_{k^+}^{-2}, \quad (2.31)$$

$$SMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}}) = \hat{\phi}\sum_{j=1}^{q}\frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4} + k^2\sum_{j=1}^{q}\frac{(3\lambda_j + k)^2\alpha_j^2}{(\lambda_j + k)^4}. \quad (2.32)$$

3.1. Theoretical comparisons between estimators

Lemma 1. Let $\hat{\pi}_1$ and $\hat{\pi}_2$ be two estimators of π; the estimator $\hat{\pi}_2$ is said to be better than the estimator $\hat{\pi}_1$ in the implication of MMSE criterion, if and only if:

$$diff(\hat{\pi}_1, \hat{\pi}_2) = MMSE(\hat{\pi}_1) - MMSE(\hat{\pi}_2) > 0.$$

Lemma 2. Let D be a positive definite matrix, ϖ be a vector of nonzero constants, and b be a positive constant. Then $bD - \varpi\varpi' > 0$ if and only if $\varpi' D\, \varpi < b$ [21].

Lemma3. Let $\hat{\pi}_1$ and $\hat{\pi}_2$ be two estimators of π, $\hat{\pi}_1 = b_1 y$ and $\hat{\pi}_2 = b_2 y$, and the covariance matrices of the estimators $\hat{\pi}_1, \hat{\pi}_2$, respectively, are $cov(\hat{\pi}_1), cov(\hat{\pi}_2)$.

$$diff(\hat{\pi}_1, \hat{\pi}_2) = MMSE(\hat{\pi}_1) - MMSE(\hat{\pi}_2) = cov(\hat{\pi}_1) + b_1 b_1' - cov(\hat{\pi}_2) - b_2 b_2' > 0,$$

if and only if $b_2'(cov(\hat{\pi}_1) - cov(\hat{\pi}_2) + b_1 b_1')b_2 < 1$, where b_1, b_2 denote the bias vectors of the estimators $\hat{\pi}_1, \hat{\pi}_2$, respectively. For more details, see [21,22].

Theorem 1. The estimator $\hat{\alpha}^{ZINB-\mathrm{MKLE}}$ is superior to the estimator $\hat{\alpha}^{ZINB-\mathrm{MLE}}$ in the sense of MMSE criterion, i.e., $MMSE(\hat{\alpha}^{ZINB-\mathrm{MLE}}) - MMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}}) > 0$ if and only if

$$bias(\hat{\alpha}^{ZINB-\mathrm{MKLE}})'^{\left[\hat{\theta}\Lambda^{-1} - \hat{\theta}\Lambda_{k^+}^{-1}\Lambda_{k^-}\Lambda_{k^+}^{-1}\Lambda\Lambda_{k^+}^{-1}\Lambda_{k^-}\Lambda_{k^+}^{-1}\right]}bias(\hat{\alpha}^{ZINB-\mathrm{MKLE}}) < 1.$$

Proof. Let $diff_1 = MMSE(\hat{\alpha}^{ZINB-\mathrm{MLE}}) - MMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}})$, and then

$$diff_1 = \hat{\theta}\Lambda^{-1} - \hat{\theta}\Lambda_{k^+}^{-1}\Lambda_{k^-}\Lambda_{k^+}^{-1}\Lambda\Lambda_{k^+}^{-1}\Lambda_{k^-}\Lambda_{k^+}^{-1} - bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})'bias(\hat{\beta}^{ZINB-\mathrm{MKLE}}),$$

$$diff_1 = \psi diag\left\{\frac{1}{\lambda_j} - \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4}\right\}_{j=1}^{q}\psi' - bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})'bias(\hat{\beta}^{ZINB-\mathrm{MKLE}}). \quad (3.1)$$

Since $bias(\hat{\alpha}^{ZINB-MKLE})bias(\hat{\alpha}^{ZINB-MKLE})'$ in Eq (2.30) is nonnegative definite, then the difference matrix $diff_1$ is positive definite if and only if $(\lambda_j + k)^4 > \lambda_j^2(\lambda_j - k)^2$, for $k > 0$.

Thus, $\hat{\alpha}^{ZINB-MKLE}$ is better than $\hat{\alpha}^{ZINB-\mathrm{MLE}}$ and the proof is completed by Lemmas 1–3.

Theorem 2. The estimator $\hat{\alpha}^{ZINB-MKLE}$ is superior to the estimator $\hat{\alpha}^{ZINB-RRE}$ in the sense of MMSE criterion, i.e., $MMSE(\hat{\alpha}^{ZINB-RRE}) - MMSE(\hat{\alpha}^{ZINB-MKLE}) > 0$ if and only if

$$bias(\hat{\alpha}^{ZINB-MKLE})'\left[\hat{\theta}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1} - \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\right]bias(\hat{\alpha}^{ZINB-MKLE}) < 1.$$

Proof. Let $diff_2 = MMSE(\hat{\alpha}^{ZINB-RRE}) - MMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}})$, and then

$$diff_2 = \hat{\theta}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1} - \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1} + bias\left(\hat{\beta}^{ZINB-RRE}\right)'bias\left(\hat{\beta}^{ZINB-RRE}\right) - bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)'\, bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right),$$

$$diff_2 = \psi\ diag\left\{\frac{\lambda_j}{\left(\lambda_j+k\right)^2} - \frac{\lambda_j\left(\lambda_j-k\right)^2}{\left(\lambda_j+k\right)^4}\right\}_{j=1}^{q}\psi' + bias\left(\hat{\beta}^{ZINB-RRE}\right)'bias\left(\hat{\beta}^{ZINB-RRE}\right) - bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)'bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right). \tag{3.2}$$

From Eqs (2.17) and (2.30), it is obvious that $bias\left(\hat{\beta}^{ZINB-RRE}\right)'bias\left(\hat{\beta}^{ZINB-RRE}\right)$ and $bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)'bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)$ are nonnegative definite, and then the difference matrix $diff_2$ is positive definite if and only if $\left(\lambda_j+k\right)^4 > \lambda_j^2(\lambda_j-k)^2$, for $k > 0$. Thus, $\hat{\alpha}^{ZINB-MKLE}$ is better than $\hat{\alpha}^{ZINB-RRE}$ and the proof is completed by Lemmas 1–3.

Theorem 3. The estimator $\hat{\alpha}^{ZINB-\mathrm{MKLE}}$ is superior to the estimator $\hat{\alpha}^{ZINB-JSE}$ in the sense of MMSE criterion, i.e., $MMSE(\hat{\alpha}^{ZINB-JSE}) - MMSE(\hat{\alpha}^{ZINB-MKLE}) > 0$ if and only if

$$bias(\hat{\alpha}^{ZINB-MKLE})'\left[c^2\hat{\theta}\Lambda^{-1} - \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\right]bias(\hat{\alpha}^{ZINB-MKLE}) < 1.$$

Proof. Let $diff_3 = MMSE(\hat{\alpha}^{ZINB-JSE}) - MMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}})$, and then

$$diff_3 = c^2\hat{\theta}\Lambda^{-1} - \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1} + bias\left(\hat{\beta}^{ZINB-JSE}\right)'bias\left(\hat{\beta}^{ZINB-JSE}\right) - bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)'bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right),$$

$$diff_2 = \psi\ diag\left\{\frac{c^2}{\lambda_j} - \frac{\lambda_j\left(\lambda_j-k\right)^2}{\left(\lambda_j+k\right)^4}\right\}_{j=1}^{q}\psi' + bias\left(\hat{\beta}^{ZINB-JSE}\right)'bias\left(\hat{\beta}^{ZINB-JSE}\right) - bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)'bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right). \tag{3.3}$$

From Eqs (2.15) and (2.30), it is obvious that $bias\left(\hat{\beta}^{ZINB-JSE}\right)'bias\left(\hat{\beta}^{ZINB-JSE}\right)$ and $bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)'bias\left(\hat{\beta}^{ZINB-\mathrm{MKLE}}\right)$ are nonnegative definite. Then the difference matrix $diff_3$ is positive definite if and only if $c^2\left(\lambda_j+k\right)^4 > \lambda_j^2(\lambda_j-k)^2$, for $k > 0\ and\ 0 < c < 1$. Thus, $\hat{\alpha}^{ZINB-MKLE}$ is better than $\hat{\alpha}^{ZINB-JSE}$ and the proof is completed by Lemmas 1–3.

Theorem 4. The estimator $\hat{\alpha}^{ZINB-MKLE}$ is superior to the estimator $\hat{\alpha}^{ZINB-KLE}$ in the sense of MMSE criterion, i.e., $MMSE(\hat{\alpha}^{ZINB-KLE}) - MMSE(\hat{\alpha}^{ZINB-MKLE}) > 0$ if and only if

$$bias(\hat{\alpha}^{ZINB-MKLE})'^{\left[\hat{\phi}\Lambda_{k+}^{-1}\Lambda_{k-}^{-1}\Lambda^{-1}\Lambda_{k-}^{-1}\Lambda_{k+}^{-1} - \hat{\phi}\Lambda_{k+}^{-1}\Lambda_{k}-\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}-\Lambda_{k+}^{-1}\right]}bias(\hat{\alpha}^{ZINB-MKLE}) < 1.$$

Proof. Let $diff_4 = MMSE(\hat{\alpha}^{ZINB-KLE}) - MMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}})$, and then

$$diff_4 = \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k-}^{-1}\Lambda^{-1}\Lambda_{k-}^{-1}\Lambda_{k+}^{-1} - \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1} + bias(\hat{\beta}^{ZINB-KLE})' bias(\hat{\beta}^{ZINB-KLE}) - bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})' \,\mathrm{b}ias(\hat{\beta}^{ZINB-\mathrm{MKLE}}),$$

$$diff_4 = \psi \; diag\left\{\frac{(\lambda_j - k)^2}{(\lambda_j + k)^2 \lambda_j} - \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4}\right\}_{j=1}^{q} \psi' + bias(\hat{\beta}^{ZINB-KLE})' \mathrm{b}ias(\hat{\beta}^{ZINB-KLE}) - bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})' \, bias(\hat{\beta}^{ZINB-\mathrm{MKLE}}). \quad (3.4)$$

From Eqs (2.20) and (2.30), it is obvious that $bias(\hat{\beta}^{ZINB-KLE})' \mathrm{b}ias(\hat{\beta}^{ZINB-KLE})$ and $bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})' \mathrm{b}ias(\hat{\beta}^{ZINB-\mathrm{MKLE}})$ are nonnegative definite. Then the difference matrix $diff_4$ is positive definite if and only if $(\lambda_j - k)^2 (\lambda_j + k)^2 > \lambda_j^2 (\lambda_j - k)^2$, for $k > 0$. Thus, $\hat{\alpha}^{ZINB-MKLE}$ is better than $\hat{\alpha}^{ZINB-KLE}$ and the proof is completed by Lemmas 1–3.

Theorem 5. The estimator $\hat{\alpha}^{ZINB-MKLE}$ is superior to the estimator $\hat{\alpha}^{ZINB-LE}$ in the sense of MMSE criterion, i.e., $MMSE(\hat{\alpha}^{ZINB-LE}) - MMSE(\hat{\alpha}^{ZINB-MKLE}) > 0$ if and only if

$$bias(\hat{\alpha}^{ZINB-MKLE})'^{\left[\hat{\theta}\Lambda_I^{-1}\Lambda_{d}+\Lambda^{-1}\Lambda_{d}+\Lambda_I^{-1} - \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k}-\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}-\Lambda_{k+}^{-1}\right]} bias(\hat{\alpha}^{ZINB-MKLE}) < 1.$$

Proof. Let $diff_5 = MMSE(\hat{\alpha}^{ZINB-LE}) - MMSE(\hat{\alpha}^{ZINB-\mathrm{MKLE}})$, and then

$$diff_5 = \hat{\theta}\Lambda_I^{-1}\Lambda_{d+}\Lambda^{-1}\Lambda_{d+}\Lambda_I^{-1} - \hat{\theta}\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1}\Lambda\Lambda_{k+}^{-1}\Lambda_{k}\text{-}\Lambda_{k+}^{-1} + bias(\hat{\beta}^{ZINB-LE})' bias(\hat{\beta}^{ZINB-LE}) - bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})' \mathrm{b}ias(\hat{\beta}^{ZINB-\mathrm{MKLE}}),$$

$$diff_5 = \psi \; diag\left\{\frac{(\lambda_j + d)^2}{(\lambda_j + 1)^2 \lambda_j} - \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4}\right\}_{j=1}^{q} \psi' + bias(\hat{\beta}^{ZINB-LE})' bias(\hat{\beta}^{ZINB-LE}) - \mathrm{b}ias(\hat{\beta}^{ZINB-\mathrm{MKLE}})' \, \mathrm{b}ias(\hat{\beta}^{ZINB-\mathrm{MKLE}}). \quad (3.5)$$

From Eqs (2.25) and (2.30), it is obvious that $Bias(\hat{\beta}^{ZINB-LE})' Bias(\hat{\beta}^{ZINB-LE})$ and $Bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})' Bias(\hat{\beta}^{ZINB-\mathrm{MKLE}})$ are nonnegative definite. Then the difference matrix $diff_5$ is positive definite if and only if $(\lambda_j + k)^4 (\lambda_j + d)^2 > \lambda_j^2 (\lambda_j - k)^2 (\lambda_j + 1)^2$, for $k > 0$ and $0 < d < 1$. Thus, $\hat{\alpha}^{ZINB-MKLE}$ is better than $\hat{\alpha}^{ZINB-LE}$ and the proof is completed by Lemmas 1–3.

4. Simulation study

In this section, we design a Monte Carlo simulation to assess the estimators' performance under varying conditions of multicollinearity. The predictors are generated using the following equation:

$$x_{ij} = (1 - \rho^2)^{1/2} n_{ij} + \rho n_{i,q+1}, i = 1,2, \dots, n, \; j = 1,2,3, \dots, q, \quad (4.1)$$

where n_{ij} denotes independent standard normal random numbers, q is the number of predictors, and γ controls the degree of correlation between predictors. We investigate different levels of multicollinearity with $\rho = 0.8, 0.9, 0.95, \text{and } 0.99$ generated from the zero-inflated negative binomial (ZINB) distribution, which accounts for excess zeros in the data. The model is specified as:

$$y_i \sim ZINB(\mu_i, \theta, \pi), \quad (4.2)$$

where $\mu_i = \exp(x_i^T \beta)$ represents the mean of the negative binomial component, θ is the dispersion parameter, and π takes the zero-inflation probability of 10% and 20% in this study. The sample sizes are set to n = 40, 50, 100, and 200, and two scenarios for the number of predictors are considered: q=4,8. The true regression coefficients β are chosen such that $\beta'\beta = 1$, following the approach of similar studies in the literature [23,24].

The primary criterion for evaluating the estimators' performances is the estimated mean squared error (MSE) which is computed as follows:

$$\mathrm{MSE} = \frac{1}{1000}\sum_{r=1}^{1000}\sum_{j=1}^{p}(\hat{\beta}_{rj} - \beta_j)^2, \tag{4.3}$$

where $\hat{\beta}_{rj}$ denotes the estimate obtained in the r^{th} replication of the simulation, and β_j is the true vector of regression parameters. Each simulation is repeated 1000 times to ensure robust and reliable results. The **pscl** package in R is used to generate data and estimate the parameters of the ZINB regression model [25]. The model fitting is performed using the zeroinfl function from the package, which provides maximum likelihood estimates for ZINB regression models. The simulation results are summarized in Tables 1 and 2.

Table 1. Estimated MSE values for p = 4.

π	0.1				0.2			
n	40	50	100	200	40	50	100	200
	$\rho = 0.8$				$\rho = 0.8$			
$\hat{\beta}_{MLE}$	0.6572	0.6270	0.2143	0.1299	0.8312	0.8225	0.2541	0.1585
$\hat{\beta}_{ZINB-RRE}$	0.4786	0.4285	0.1804	0.1143	0.5874	0.5264	0.2046	0.1357
$\beta_{ZINB-LE}$	0.5578	0.4326	0.2012	0.1260	0.6826	0.6341	0.2331	0.1513
$\hat{\beta}_{ZINB-JSE}$	0.5516	0.5256	0.1885	0.1182	0.6911	0.6861	0.2212	0.1423
$\hat{\beta}_{ZINB-KLE}$	0.3437	0.3009	0.1507	0.1000	0.4111	0.3407	0.1629	0.1154
$\hat{\beta}_{ZINB-MKLE}$	0.2789	0.2356	0.1300	0.0888	0.3334	0.2629	0.1366	0.1004
	$\rho = 0.9$				$\rho = 0.9$			
$\hat{\beta}_{MLE}$	1.1425	1.0287	0.3559	0.2329	1.4484	1.3572	0.4604	0.2903
$\hat{\beta}_{ZINB-RRE}$	0.6805	0.4614	0.2683	0.1853	0.7772	0.6726	0.3258	0.2194
$\beta_{ZINB-LE}$	0.8272	0.4718	0.3120	0.2142	0.9297	0.6564	0.3870	0.2577
$\hat{\beta}_{ZINB-JSE}$	0.9429	0.8945	0.3027	0.2014	1.1929	1.1890	0.3874	0.2477
$\hat{\beta}_{ZINB-KLE}$	0.4132	0.3760	0.1957	0.1440	0.5020	0.3862	0.2203	0.1602
$\hat{\beta}_{ZINB-MKLE}$	0.3116	0.2696	0.1532	0.1165	0.3732	0.2730	0.1660	0.1248
	$\rho = 0.95$				$\rho = 0.95$			
$\hat{\beta}_{MLE}$	2.2349	1.6314	0.6672	0.4539	2.8017	2.6556	0.7854	0.5630
$\hat{\beta}_{ZINB-RRE}$	1.0726	0.5060	0.4292	0.3092	1.3585	1.1815	0.4651	0.3587
$\beta_{ZINB-LE}$	1.2878	0.5257	0.4982	0.3740	1.5918	1.3922	0.5748	0.4430
$\hat{\beta}_{ZINB-JSE}$	1.8279	1.3275	0.5552	0.3803	2.2877	2.1678	0.6514	0.4689
$\hat{\beta}_{ZINB-KLE}$	0.4423	0.4320	0.2509	0.1949	0.6392	0.5035	0.2446	0.2061
$\hat{\beta}_{ZINB-MKLE}$	0.3544	0.2925	0.1727	0.1376	0.4408	0.3014	0.1975	0.1595
	$\rho = 0.99$				$\rho = 0.99$			
$\hat{\beta}_{MLE}$	10.4219	7.7560	2.8698	2.2134	13.2806	12.3154	3.7021	2.6627
$\hat{\beta}_{ZINB-RRE}$	2.8234	2.6931	1.0246	0.8208	3.9114	2.9044	1.1795	0.8834
$\beta_{ZINB-LE}$	3.0238	2.6610	1.2126	1.0106	4.1697	3.0176	1.3755	1.0708
$\hat{\beta}_{ZINB-JSE}$	8.4607	6.2890	2.3382	1.8053	10.7767	9.9931	3.0135	2.1692
$\hat{\beta}_{ZINB-KLE}$	1.2294	1.1319	0.2118	0.1997	1.9924	1.5557	0.2457	0.2148
$\hat{\beta}_{ZINB-MKLE}$	0.5424	0.4395	0.2047	0.1766	0.8459	0.4942	0.2314	0.2064

Table 2. Estimated MSE values for p = 8.

π	0.1				0.2			
n	40	50	100	200	40	50	100	200
	$\rho = 0.8$				$\rho = 0.8$			
$\hat{\beta}_{MLE}$	1.4143	0.8280	0.4108	0.2247	1.8412	1.1804	0.6667	0.3317
$\hat{\beta}_{ZINB-RRE}$	0.6772	0.4407	0.3054	0.1895	1.0085	0.5380	0.4730	0.2684
$\beta_{ZINB-LE}$	0.6750	0.5200	0.3594	0.2123	1.1256	0.6437	0.5656	0.3066
$\hat{\beta}_{ZINB-JSE}$	1.1584	0.6810	0.3416	0.1922	1.5063	0.9671	0.5523	0.2811
$\hat{\beta}_{ZINB-KLE}$	0.7391	0.4670	0.2194	0.1578	0.8033	0.6208	0.3220	0.2134
$\hat{\beta}_{ZINB-MKLE}$	0.4682	0.4249	0.1705	0.1344	0.5796	0.5115	0.2449	0.1763
	$\rho = 0.9$				$\rho = 0.9$			
$\hat{\beta}_{MLE}$	1.8661	1.0815	0.7236	0.4397	2.8712	1.4467	1.2718	0.6191
$\hat{\beta}_{ZINB-RRE}$	0.7417	0.5819	0.4600	0.3308	0.8574	0.7717	0.7561	0.4431
$\beta_{ZINB-LE}$	0.9160	0.6053	0.5623	0.3897	1.1139	0.7618	0.9238	0.5327
$\hat{\beta}_{ZINB-JSE}$	1.5207	0.9413	0.5941	0.3661	2.3375	1.4331	1.0411	0.5135
$\hat{\beta}_{ZINB-KLE}$	0.7672	0.4786	0.2698	0.2395	0.8346	0.8061	0.4091	0.3013
$\hat{\beta}_{ZINB-MKLE}$	0.5814	0.4426	0.1897	0.1845	0.7016	0.5484	0.2814	0.2243
	$\rho = 0.95$				$\rho = 0.95$			
$\hat{\beta}_{MLE}$	3.5822	2.1220	1.4741	0.8592	5.3309	2.6962	2.3959	1.1956
$\hat{\beta}_{ZINB-RRE}$	1.1089	1.0728	0.7646	0.5461	2.1819	1.2520	1.1327	0.7121
$\beta_{ZINB-LE}$	1.3670	1.2611	0.9339	0.6796	1.8392	1.7943	1.3621	0.8963
$\hat{\beta}_{ZINB-JSE}$	2.9108	1.7369	1.2024	0.7064	4.3284	2.1914	1.9518	0.9804
$\hat{\beta}_{ZINB-KLE}$	2.6636	0.5095	0.3315	0.3123	3.6741	0.9253	0.4489	0.3721
$\hat{\beta}_{ZINB-MKLE}$	0.6431	0.4726	0.2084	0.2105	1.2646	0.6158	0.3715	0.2434
	$\rho = 0.99$				$\rho = 0.99$			
$\hat{\beta}_{MLE}$	20.3974	9.5257	7.8039	4.2661	34.3705	16.0418	13.6993	5.7252
$\hat{\beta}_{ZINB-RRE}$	14.5601	4.4487	2.1208	1.5188	26.0909	10.3974	3.2849	1.8511
$\beta_{ZINB-LE}$	11.2872	3.3563	2.0281	1.7302	20.2901	8.1482	2.5652	2.0153
$\hat{\beta}_{ZINB-JSE}$	16.5331	7.7329	6.3296	3.4662	27.8534	13.0043	11.1094	4.6497
$\hat{\beta}_{ZINB-KLE}$	13.1185	4.0672	0.5607	0.3932	23.7760	9.7297	1.1210	0.4521
$\hat{\beta}_{ZINB-MKLE}$	9.4647	2.1048	0.3201	0.2385	18.7320	6.2457	1.0721	0.2531

The simulation results in Tables 1 and 2 evaluates the performance of six estimators—MLE, ZINB-RRE, ZINB-LE, ZINB-JSE, ZINB-KLE, and ZINB-MKLE under varying conditions of multicollinearity (ρ), sample size (n), and zero-inflation probability (π). The primary evaluation metric is the mean squared error (MSE), where lower values indicate better estimator performance. The general observations are as follows: As the multicollinearity level (ρ) increases, the mean squared error (MSE) of most estimators demonstrates a noticeable upward trend, reflecting their sensitivity to stronger correlations among predictors. This pattern, illustrated in Figure 1, highlights the growing challenge of maintaining estimation accuracy under severe multicollinearity conditions. The ZINB-MKLE consistently achieves the lowest MSE across almost all scenarios, making it the best-performing estimator under severe multicollinearity (ρ = 0.99). The MLE performs poorly as multicollinearity intensifies, with its MSE increasing substantially, especially for higher ρ. MSE decreases for all estimators as n increases.

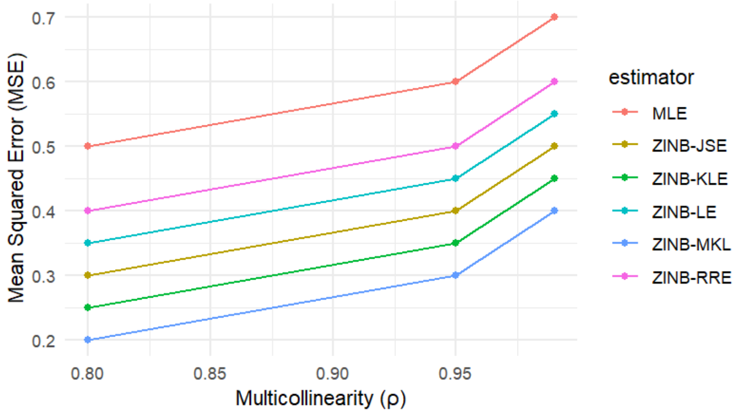

Figure 1. Graph of MSE against the level of multicollinearity.

Larger sample sizes (n) mitigate the adverse effects of multicollinearity, leading to improved precision for all estimators. This trend is evident in Figure 2, where the mean squared error (MSE) decreases as n increases, even under high multicollinearity. Among the estimators, ZINB-KLE and ZINB-MKLE demonstrate significant improvements with increasing n, maintaining relatively low MSE values. Higher levels of zero-inflation (π = 0.2) generally exacerbate the MSE across all estimators, reflecting the additional complexity introduced by excess zeros in the data. However, as shown in Figure 2, the ZINB-MKLE remains robust to these changes, consistently outperforming other estimators regardless of the degree of zero inflation. This robustness underscores the effectiveness of the ZINB-MKLE in handling challenging data scenarios.

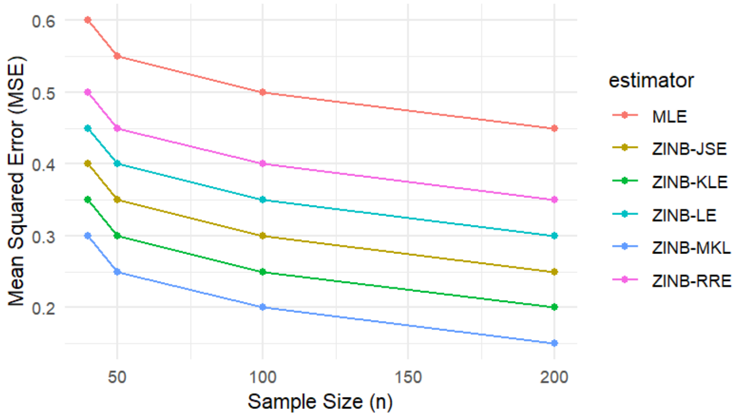

Figure 2. Graph of MSE against the sample size.

The ZINB-MKLE consistently outperforms other estimators across all scenarios, making it the most robust estimator under multicollinearity and varying sample sizes. This estimator's design effectively handles multicollinearity while maintaining precision in parameter estimation. The MLE is highly sensitive to multicollinearity, with its MSE escalating significantly as ρ increases. This underscores the necessity of using alternative estimators like the ZINB-MKLE or ZINB-KLE in the presence of multicollinearity. The ZINB-KLE is a strong alternative to the ZINB-MKLE, performing well across moderate and severe multicollinearity scenarios. While slightly less effective than the

ZINB-MKLE, its MSE values are consistently lower than other estimators except the ZINB-MKLE. Increasing sample sizes consistently reduces MSE across all estimators, highlighting the importance of larger datasets in mitigating multicollinearity's effects. Higher zero-inflation increases MSE for all estimators, but the ZINB-MKLE exhibits greater resilience compared to others.

In conclusion, the simulation results emphasize the superiority of the ZINB-MKLE across all levels of multicollinearity, sample sizes, and zero-inflation probabilities. It is recommended as the best choice for modeling zero-inflated data with high multicollinearity. The ZINB-KLE is a reliable alternative, particularly in large sample settings. The MLE's performance is inadequate under multicollinearity, reinforcing the need for specialized estimators like the ZINB-MKLE in such scenarios.

5. Application

This dataset comprises information on blood transfusions received by 150 thalassemia patients in Mosul, Iraq, as documented by Algamal et al. [3]. The recorded explanatory variables for each patient include the following: x_1 (age in months), x_2 (duration of thalassemia in months), x_3 (hemoglobin concentration), x_4 (packed cell volume), x_5 (number of blood units), and x_6 (age at the onset of blood transfusion in months). A zero-inflation ratio of 0.52 is observed in the dataset, as illustrated in Figure 3, indicating a considerable presence of zeros in the response variable. This characteristic supports the application of zero-inflated models, such as the zero-inflated negative binomial (ZINB) regression model, for analyzing the data. The correlation heatmap (Figure 4) reveals notable relationships between certain predictors, suggesting the presence of multicollinearity. To quantify the extent of multicollinearity, a variance inflation factor (VIF) analysis was conducted. The VIF was calculated, and the results indicate high multicollinearity for x_1 (VIF = 37.51) and x_5 (VIF = 35.12), and moderate multicollinearity for x_3 (VIF = 14.25) and x_4 (VIF = 11.13). Given the severity of multicollinearity among some predictors, regularization techniques, such as ridge regression and the proposed method, are employed in the ZINB regression model to address this issue effectively.

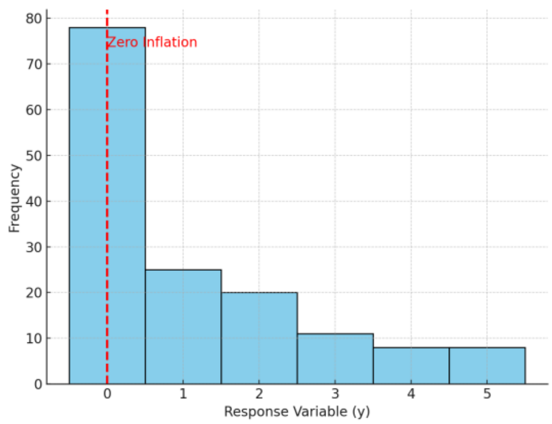

Figure 3. Histogram of the Blood Transfusion dataset.

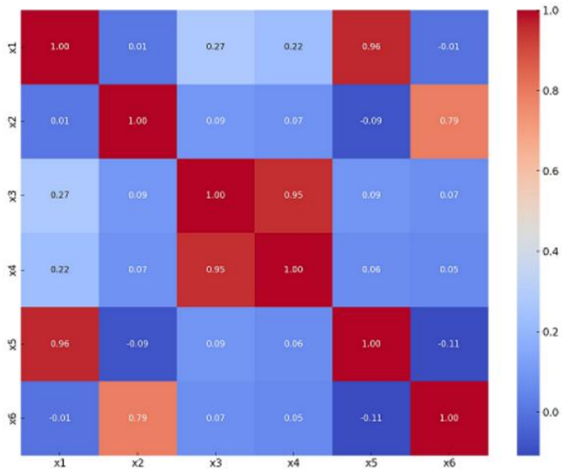

Figure 4. Correlation heatmap of predictors.

Table 3 presents the estimated regression coefficients for six different estimators applied to the Blood Transfusion dataset, including the MLE, and five shrinkage-based methods designed for zero-inflated negative binomial (ZINB) models. These estimators, ZINB-RRE, ZINB-LE, ZINB-JSE, ZINB-KLE, and ZINB-MKLE, are specifically introduced to mitigate multicollinearity and enhance predictive stability.

Table 3. Estimated regression coefficients for the Blood Transfusion dataset.

Coef.	Estimators					
	$\hat{\beta}_{MLE}$	$\hat{\beta}_{ZINB-RRE}$	$\hat{\beta}_{ZINB-LE}$	$\hat{\beta}_{ZINB-JSE}$	$\hat{\beta}_{ZINB-KLE}$	$\hat{\beta}_{ZINB-MKLE}$
x_1	0.3259	0.0423	0.2532	0.0019	-0.4271	-0.0349
x_2	-0.0338	0.0007	0.0040	-0.0002	-0.0196	-0.0006
x_3	-0.8024	-0.0460	-0.3280	-0.0046	-0.7822	0.0401
x_4	0.7701	0.0393	0.2886	0.0044	0.7330	-0.0345
x_5	-0.4668	-0.0396	-0.2349	-0.0027	-0.3559	0.0326
x_6	0.0274	-0.0032	-0.0170	0.0002	0.0027	0.0026
SMSE	14638.7500	1.8943	21.2755	2.0211	685.8135	1.8568

The MLE coefficients exhibit the largest absolute magnitudes across predictors, suggesting potential instability due to multicollinearity among the explanatory variables. In contrast, all five biased estimators induce varying degrees of shrinkage toward zero, thereby improving coefficient stability and interpretability. Among these, the ZINB-RRE and ZINB-MKLE demonstrate the most substantial shrinkage, effectively reducing the variance of the estimates.

Notably, the ZINB-MKLE yields coefficients of smaller magnitude relative to the ZINB-KLE, indicating that the MKLE's additional bias-adjustment mechanism provides stronger regularization and enhances robustness against overfitting. The ZINB-LE and ZINB-JSE produce moderately shrunken estimates, achieving a more balanced bias–variance trade-off. Predictors x_2 and x_6 consistently show near-zero coefficients across all estimators, suggesting that these variables may have limited predictive contribution to the response.

The SMSE provides a comparative assessment of predictive performance. The MLE performs poorly, with an SMSE of 14,638.75, highlighting its sensitivity to multicollinearity and model misspecification. In contrast, the shrinkage-based methods achieve substantial improvements:

- The ZINB-MKLE attains the lowest SMSE (1.8568), demonstrating superior predictive accuracy.
- The ZINB-RRE performs comparably well with an SMSE of 1.8943, suggesting that ridge-type penalization is highly effective for this dataset.
- The ZINB-KLE improves prediction over the MLE but remains less competitive relative to the MKLE due to its comparatively weaker regularization.
- The ZINB-LE and ZINB-JSE achieve intermediate performance, balancing stability with moderate predictive gains.

The results underscore the limitations of the MLE in multicollinear settings, where variance inflation degrades both coefficient stability and predictive accuracy. Among the biased estimators, the ZINB-MKLE emerges as the most reliable, combining effective shrinkage, stable coefficient estimation, and superior predictive performance. The ZINB-RRE offers a strong alternative, whereas the ZINB-LE and ZINB-JSE may be preferable when moderate shrinkage is desirable. Collectively, these findings highlight the importance of incorporating regularization into ZINB regression for improved inference and prediction.

6. Concluding remarks

This study investigated the challenges posed by multicollinearity in zero-inflated negative binomial (ZINB) regression models and evaluated several biased estimators, with particular emphasis on the Kibria-Lukman estimator (ZINB-KLE) and the proposed modified Kibria-Lukman estimator (ZINB-MKLE). Through extensive Monte Carlo simulations and an empirical application to the Blood Transfusion dataset, we systematically assessed estimator performance across varying sample sizes, degrees of multicollinearity, and model complexities.

The findings demonstrate that biased estimators consistently outperform the conventional maximum likelihood estimator (MLE), particularly under moderate to high multicollinearity. Among them, the ZINB-MKLE achieved the lowest mean squared error (MSE) across most settings, reflecting its superior predictive accuracy and effective bias–variance trade-off. The modifications introduced in the ZINB-MKLE, relative to the ZINB-KLE, provided stronger shrinkage and enhanced estimator stability, making the MKLE the most robust option under challenging correlation structures.

Our results further highlight the role of sample size and model dimensionality in estimator performance. While larger sample sizes consistently reduced estimation error across all methods, scenarios with smaller samples and higher predictor counts (e.g., $q = 8$) increased estimation variance and bias. Even under these conditions, the ZINB-MKLE maintained comparatively stable performance, underscoring its robustness in complex data environments. Although the ZINB-MKLE shows clear advantages under high multicollinearity, its performance under sparse data structures and scenarios of extreme overdispersion warrants further investigation. Future work may explore optimizing the algorithm to enhance scalability for high-dimensional or large-scale zero-inflated datasets.

In conclusion, this study establishes the ZINB-MKLE as a robust and efficient alternative to traditional and existing biased estimators for modeling zero-inflated count data under multicollinearity. By combining theoretical advancements with practical performance gains, this research contributes a flexible framework that can be extended to broader applications in biostatistics, epidemiology, transportation safety, and other fields where zero-inflated, overdispersed counts are prevalent.

Author contributions

Conceptualization: Masad Alrasheedi and Adewale Lukman; Methodology: All authors; Formal analysis: All authors; Resources: All authors; Writing-original draft preparation: All authors; Writing-review and editing: All authors; Visualization: Adewale Lukman.

Use of Generative-AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The researchers would like to thank the Deanship of Graduate Studies and Scientific Research at Qassim University for financial support (QU-APC-2025).

Conflict of interest

There is no conflict of interest to declare in this study.

References

1. A. Lukman, B. Aladeitan, K. Ayinde, M. Abonazel, Modified ridge-type for the Poisson regression model: simulation and application, *J. Appl. Stat.*, **49** (2022), 2124–2136. https://doi.org/10.1080/02664763.2021.1889998
2. A. Lukman, O. Albalawi, M. Arashi, J. Allohibi, A. Alharbi, R. Farghali, Robust negative binomial regression via the Kibria-Lukman strategy: methodology and application, *Mathematics*, **12** (2024), 2929. https://doi.org/10.3390/math12182929
3. Z. Algamal, A. Lukman, M. Abonazel, F. Awwad, Performance of the ridge and Liu estimators in the zero-inflated Bell regression model, *J. Math.*, **2022** (2022), 9503460. https://doi.org/10.1155/2022/9503460
4. S. Seifollahi, H. Bevrani, Z. Algamal, Shrinkage estimators in zero-inflated Bell regression model with application, *J. Stat. Theory Pract.*, **19** (2025), 1. https://doi.org/10.1007/s42519-024-00415-1
5. Y. Al-Taweel, Z. Algamal, Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model, *Periodicals of Engineering and Natural Sciences*, **8** (2020), 248–255.
6. B. Kibria, K. Månsson, G. Shukur, Some ridge regression estimators for the zero-inflated Poisson model, *J. Appl. Stat.*, **40** (2013), 721–735. https://doi.org/10.1080/02664763.2012.752448
7. M. Akram, N. Afzal, M. Amin, A. Batool, Modified ridge-type estimator for the zero-inflated negative binomial regression model, *Commun. Stat.-Simul. C.*, **53** (2024), 5305–5322. https://doi.org/10.1080/03610918.2023.2179070

8. M. Akram, M. Amin, N. Afzal, B. Kibria, Kibria-Lukman estimator for the zero-inflated negative binomial regression model: theory, simulation and applications, *Commun. Stat.-Simul. C.*, **54** (2025), 1464–1480. https://doi.org/10.1080/03610918.2023.2286436
9. M. Akram, M. Abonazel, M. Amin, B. Golam Kibria, N. Afzal, A new Stein estimator for the zero-inflated negative binomial regression model, *Concurr. Comp.-Pract. Expe.*, **34** (2022), e7045. https://doi.org/10.1002/cpe.7045
10. T. Omer, P. Sjölander, K. Månsson, B. Kibria, Improved estimators for the zero-inflated Poisson regression model in the presence of multicollinearity: simulation and application of maternal death data, *Communications in Statistics: Case Studies, Data Analysis and Applications*, **7** (2021), 394–412. https://doi.org/10.1080/23737484.2021.1952493
11. M. Zeeshana, A. Khana, M. Amanullaha, M. Bakrb, A. Alshangitib, O. Balogun, et al., A new modified biased estimator for Zero inflated Poisson regression model, *Heliyon*, **10** (2024), e24225. https://doi.org/10.1016/j.heliyon.2024.e24225
12. M. Amin, B. Ashraf, S. Siddiqa, Liu estimation method in the zero-inflated Conway Maxwell Poisson regression model, *J. Stat. Theory Appl.*, **24** (2025), 71–90. https://doi.org/10.1007/s44199-024-00101-y
13. A. Dempster, M. Schatzoff, N. Wermuth, A simulation study of alternatives to ordinary least squares, *J. Am. Stat. Assoc.*, **72** (1977), 77–91.
14. D. Lambert, Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics*, **34** (1992), 1–14.
15. G. Nanjundan, An EM algorithmic approach to maximum likelihood estimation in a mixture model, *Vignana Bharathi*, **18** (2006), 7–13.
16. A. Hoerl, R. Kennard, Ridge regression: biased estimation for nonorthogonal problems, *Technometrics*, **12** (1970), 55–67.
17. M. Amin, M. Akram, M. Amanullah, On the James-Stein estimator for the Poisson regression model, *Commun. Stat.-Simul. C.*, **51** (2022), 5596–5608. https://doi.org/10.1080/03610918.2020.1775851
18. B. Kibria, A. Lukman, A new ridge-type estimator for the linear regression model: simulations and applications, *Scientifica*, **2020** (2020), 9758378. https://doi.org/10.1155/2020/9758378
19. L. Kejian, A new class of blased estimate in linear regression, *Commun. Stat.-Theor. M.*, **22** (1993), 393–402. https://doi.org/10.1080/03610929308831027
20. B. Aladeitan, O. Adebimpe, A. Lukman, O. Oludoun, O. Abiodun, Modified Kibria-Lukman (MKL) estimator for the Poisson regression model: application and simulation, *F1000Res.*, **10** (2021), 548. https://doi.org/10.12688/f1000research.53987.2
21. R. Farebrother, Further results on the mean square error of ridge regression, *J. R. Stat. Soc. B*, **38** (1976), 248–250.
22. G. Trenkler, H. Toutenburg, Mean squared error matrix comparisons between biased estimators—an overview of recent results, *Stat. Pap.*, **31** (1990), 165–179. https://doi.org/10.1007/BF02924687
23. R. Farghali, A. Lukman, A. Ogunleye, Enhancing model predictions through the fusion of Stein estimator and principal component regression, *J. Stat. Comput. Sim.*, **94** (2024), 1760–1775. https://doi.org/10.1080/00949655.2024.2302011

24. A. Lukman, E. Adewuyi, K. Månsson, B. Kibria, A new estimator for the multicollinear Poisson regression model: simulation and application, *Sci. Rep.*, **11** (2021), 3732. https://doi.org/10.1038/s41598-021-82582-w
25. A. Zeileis, C. Kleiber, S. Jackman, Regression models for count data in R, *J. Stat. Softw.*, **27** (2008), 1–25. https://doi.org/10.18637/jss.v027.i08