STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–12.
Published online in International Academic Press (www.IAPress.org)

Linearized Ridge-Type Estimator for the Poisson Modification of the Quasi-Lindley Regression Model

Noora Suhail Hawa [1], Zeina Ameer Hadied [2], Oday Esam Al-Saqal [3], Zakariya Yahya Algamal [4,*]

[1] *Department of Statistics and Informatics, University of Mosul, Iraq*
[2] *University of Mosul, Mosul, Iraq*
[3] *Department of Sharia, University of Mosul, Mosul, Iraq*
[4] *Department of Statistics and Informatics, University of Mosul, Mosul, Iraq*

Abstract This paper develops a linearized ridge-type estimator for the Poisson-modification of the quasi-Lindley regression model (PMQL-RM), which is designed for overdispersed count responses frequently encountered in epidemiology, insurance, and related fields. The PMQL-RM extends classical Poisson regression by incorporating a flexible mixed Poisson distribution with two dispersion parameters, thereby accommodating substantial overdispersion while retaining a generalized linear model structure and maximum likelihood estimation via iteratively weighted least squares. However, in many practical applications, the explanatory variables are highly correlated, leading to multicollinearity that inflates the variance of the PMQL maximum likelihood estimator (PMQL-MLE) and yields unstable inference. Existing shrinkage estimators for this model, including the PMQL ridge estimator (PMQL-RRE), the PMQL Liu estimator (PMQL-LE), and the PMQL Liu-type estimator (PMQL-LTE), partially address this issue by introducing biasing parameters, but there remains room for further reduction in mean squared error (MSE). Motivated by the superior MSE properties of the linearized ridge regression estimator in linear and generalized linear models, we propose its extension to the PMQL-RM, termed the PMQL linearized ridge estimator (PMQL-LRE). We derive the bias, variance–covariance matrix, and matrix MSE of the PMQL-LRE and establish theoretical conditions under which it dominates PMQL-MLE, PMQL-RRE, PMQL-LE, and PMQL-LTE in both matrix and scalar MSE senses. A Monte Carlo simulation study, conducted under varying levels of multicollinearity, sample sizes, numbers of predictors, and dispersion parameter settings, demonstrates that the PMQL-LRE generally achieves the smallest MSE, especially under severe multicollinearity. An empirical application to real overdispersed count data further illustrates the practical advantages of the proposed estimator in terms of improved estimation accuracy and more stable coefficient estimates.

Keywords Shrinkage estimation, Ridge estimator; overdispersed count data, multicollinearity, Poisson-modification of quasi-Lindley regression

DOI: 10.19139/soic-2310-5070-3848

1. Introduction

Count responses are common in a variety of sciences, including epidemiology, ecology, insurance, and public health, where the outcome is the number of events in a particular unit of time, ”space or exposure [1, 2, 3, 44, 46, 47]. These are discrete, non-negative, frequently right skewed data, which may have a large percentage of zero values, and cannot be analyzed using classical linear regression models that may produce negative fitted values and assume that the variance is constant [4, 5, 6, 42, 43]. Generalized linear models that use a Poisson response to model these characteristics are a common choice, with the conditional mean of the count variable being a function of a set of exploratory variables via an appropriate link function [48, 49, 50, 51, 52].

*Correspondence to: Zakariya Yahya Algamal (Email: zakariya.algamal@uomosul.edu.iq). Department of Statistics and Informatics, University of Mosul, Iraq.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

One of the major assumptions of the standard Poisson model is equidispersion, in which the conditional variance is the same as the conditional mean. In reality, however, most real data sets are overdispersed, the variance greater than the mean, because of unobserved heterogeneity, missing covariates, excess zeros, or clustering effects [7, 8, 9, 24, 25]. Ignoring overdispersion typically leads to underestimation of standard errors, inflated test statistics, and misleading inference on covariate effects [10, 11, 26, 27, 28]. To address this problem, several overdispersed Poisson-type models have been proposed, including the negative binomial, Poisson–Lindley, and quasi-Lindley families, which introduce additional dispersion parameters and thus provide greater flexibility for modeling count data [9, 11, 12, 29, 30].

Within this class, the Poisson-modification of the quasi-Lindley (PMQL) distribution has attracted attention because it accommodates overdispersion through two dispersion parameters while retaining a tractable probability mass function [13, 14, 31, 32]. By incorporating the PMQL distribution into the generalized linear model framework, the Poisson modification of the quasi-Lindley regression model (PMQL-RM) is obtained whereby the conditional mean of the response is modeled by a set of predictors through a log link function and the regression coefficients are estimated using maximum likelihood. The iteratively weighted least squares algorithm can be used to compute the resulting PMQL maximum likelihood estimator (MLE) and has good large sample properties [33, 34, 35].

Count regression models, however, in most applications have highly correlated predictors, which leads to multicollinearity. Dependent multicollinearity severely blows up the variance of the MLEs, makes the information matrix ill conditioned and can result in unstable estimates with large confidence intervals and counter intuitive signs. To mitigate these issues in the PMQL-RM setting, shrinkage estimators such as the Poisson-modification of the quasi-Lindley ridge regression estimator (PMQL-RRE) [15, 36, 37, 38] and the Poisson-modification of the quasi-Lindley Liu estimator (PMQL-LE) [16] have been proposed, introducing one or more biasing parameters to trade a small amount of bias for a substantial reduction in variance. More recently, Liu-type estimators combining the advantages of ridge and Liu estimators have been developed for generalized linear models", leading to the Poisson-modification of the quasi-Lindley Liu-type estimator (PMQL-LTE) with two shrinkage parameters [17].

Although these estimators improve the performance of PMQL-MLE in the presence of multicollinearity, there remains scope to construct estimators that achieve further reductions in mean squared error by exploiting additional structure in the biasing mechanism.

In order to address the issue of multicollinearity in PMQL-RM, this paper proposes a new estimator known as the linearized ridge estimator. "The goal of the proposed estimate is to outperform the current estimators by reducing the inflated mean squared error of the PMQL-MLE caused by the existence of multicollinearity.

The main objectives of this study are threefold. First, we define the linearized ridge estimator for the PMQL-RM and derive its bias, variance-covariance matrix, and matrix mean squared error. Second, we perform theoretical comparisons based on matrix and scalar mean squared error between the proposed estimator and competing estimators under suitable regularity conditions. Third, we discuss data-driven strategies for selecting the biasing parameters and evaluate the finite-sample performance of the estimator through simulation studies and an empirical application.

2. PMQL-RM

The PMQL-RM is a generalized linear model designed for over-dispersed count data, where the variance of the response exceeds its mean [18, 45]. In many practical applications such as epidemiology, insurance, reliability, and environmental studies, counts are non-negative, right-skewed, and often more variable than assumed by the classical Poisson regression model (PRM), which imposes the equidispersion restriction [7, 10, 19, 20]. Ignoring over-dispersion leads to underestimated standard errors, inflated test statistics, and misleading inference, motivating the search for more flexible Poisson-type models.

An appropriate accommodation to over dispersion is to build up mixed Poisson distributions by making the Poisson mean parameter take a convenient continuous distribution. The quasi Lindley family gives a flexible mixing distribution that produces the PMQL distribution when compounded with a Poisson, giving a count model with a variance that is always larger than its mean and whose dispersion is governed by two over dispersion parameters [10, 13]. This additional flexibility allows the PMQL distribution to capture a wide range of over-dispersed patterns better than classical alternatives.

To use this distribution in regression, the PMQL model is re-parametrized in terms of its mean and dispersion parameters and embedded in the generalized linear model framework [21]. The resulting PMQL regression equation associates the conditional mean of the count response with a set of covariates with the help of a log link, and the PMQL distribution controls the conditional distribution of the response conditional on the covariates. Maximum likelihood may be used to estimate the regression and dispersion parameters, usually through iteratively weighted least squares or numerical optimization, and asymptotic properties like consistency and normality may be proved under standard regularity conditions.

As in Tharshan and Wijekoon [18], the probability density function (pdf) of PMQL distribution is given by

$$f\,(y,\gamma,\phi,\eta)=\frac{\gamma\left(\Gamma(\eta)\Gamma(y+1)\phi^3(\gamma+1)^{\eta-1}+\gamma^{\eta-1}\Gamma(y+\eta)\right)}{y!(\phi^3+1)(1+\gamma)^{y+\eta}\Gamma(\eta)}\qquad y=0\,,1\,,2\,,.... \tag{1}$$

where $\gamma>0,\eta>0$ and $\phi>-1$. The mean and variance of the Eq. (1) are expressed as

$$E\,(Y)=\ \mu=\frac{\phi+\eta}{(\phi^3+1)^{\gamma}}, \tag{2}$$

$$Var\,(Y)=\frac{\phi^6(1+\gamma)+\phi^3(2+\gamma+\eta(\gamma+\eta-1))+\eta(1+\gamma)}{(\phi^3+1)^2\gamma^2}. \tag{3}$$

This distribution is highly flexible, capable of modeling count data with a mean, μ, and two over-dispersion parameters, ϕ and η, making it suitable for a wide range of applications.

In order to describe the PMQL distribution using the GLM model, the relationship between μ and γ is reparametrized by substituting $\gamma=(\phi^3+\eta/(\phi^3+1)\mu)$into Eq. (1) which leads to a pdf defined in terms of (μ,ϕ,η)as

$$f\,(y;\mu,\phi,\eta)=\frac{\left[((\phi^3+1)\mu)^y(\phi^3+\eta)\right]\left(\Gamma(\eta)\Gamma(y+1)\phi^3\mathrm{M}^{\eta-1}+(\phi^3+\eta)^{\eta-1}\Gamma(y+\eta)\right)}{y!\ (\phi^3+1)\mathrm{M}^{y+\eta}\Gamma(\eta)}, \tag{4}$$

where $y=0,\ 1,\ 2\,,.....,\,\mu>0,\,\eta>0,\,\phi>-1,$ and $\mathrm{M}=(\phi^3+1)\mu+\phi^3+\eta$. Then, the conditional mean and variance of the regression model are given, respectively,

$$E(y_i\,|z_i\,)=\ \exp(z_i^T\beta), \tag{5}$$

and

$$Var(y_i\,|z_i\,)=\ \exp(z_i^T\beta)+(\exp{(z_i^T\beta)})^2\left(\frac{\phi^3(\phi^3+2+\eta(\eta-1))+\eta}{(\gamma^3+\eta)^2}\right)^2. \tag{6}$$

Let $y_1,y_2,...,y_n$be a random sample from the PMQL distribution. The relationship between the *p*-dimensional explanatory variables and the mean response *y* is given by the link function

$$\eta_i=g(\mu_i)=\log(\mu_i)=z_i^T\beta\Rightarrow\mu_i=\exp{(z_i^T\beta)},\quad i=1,2,....,n \tag{7}$$

where $z_i^T=(1,z_{i1},z_{i2},.....,z_{i(p-1)})$ is the *i*-th row of the design matrix, including an intercept term. The vector $\beta^T=(\beta_0,\beta_1,.....,\beta_{p-1})$ contains the unknown regression coefficients, while $\phi$and η are the overdispersion parameters. The log-likelihood function for the regression model coefficients is expressed as follows:

$$\begin{aligned}\ell(\beta,\phi,\eta|y)=&\textstyle\sum_{i=1}^n\log((\phi^3+1)\exp(z_i^T\beta))+n\log(\phi^3+\eta)-\sum_{i=1}^n\log(y_i!)-n\log(\phi^3+1)\\&-n\log\Gamma(\eta)+\textstyle\sum_{i=1}^n(\Gamma(\eta)\Gamma(y_i+1)\phi^3\mathrm{M}_i^{\eta-1}+(\phi^3+\eta)^{\eta-1}\Gamma(y_i+\eta))\\&-\textstyle\sum_{i=1}^n(y_i+\eta)\log(M_i)\end{aligned} \tag{8}$$

where $\mathrm{M} = (\phi^3 + 1)\exp(z_i^T\beta) + \phi^3 + \eta$. The score function associated with the regression coefficients vector β is derived as follows:

$$S(\beta) = \frac{\partial\ \ell(\beta,\phi,\eta|y)}{\partial\beta} = \sum_{i=1}^{n} y_i z_i - \sum_{i=1}^{n}(y_i+\eta)(\phi^3+1)\frac{\exp(z_i^T\beta)z_i}{M_i} + \sum \frac{\left(\Gamma(\eta)\Gamma(y_i+1)\phi^3(\eta-1)\right)M_i^{\delta-2}(\phi^3+1)\exp(z_i^T\beta)z_i}{\Gamma(\eta)\Gamma(y_i+1)\phi^3 M_i^{\eta-1}+(\phi^3+\eta)^{\eta-1}\Gamma(y_i+\eta)} \tag{9}$$

Given that Eq. (9) exhibits non-linearity in β, the IWLS approach (Fisher scoring method) is employed to derive PMQL-MLE. Then, the PMQL-MLE estimator is defined as

$$\hat{\beta}_{\mathrm{PMQL-MLE}} = B^{-1}Z^T\hat{\mathrm{Q}}\ \hat{\mathrm{e}} \tag{10}$$

where$\mathrm{B} = \mathrm{Z}^{\mathrm{T}}\ \hat{\mathrm{Q}}\ \mathrm{Z}$, $\hat{\mathrm{Q}}$ is a diagonal matrix, and$\hat{\mathrm{e}}$is a vector where the i^{th} element is given by

$$\hat{\mathrm{Q}}{=}diag\left(\frac{(\phi^3{+}1)^2\mu_{\mathrm{i}}}{(\phi^3{+}\eta)^2{+}\phi^3(\phi^3{+}2\eta{-}1)(\phi^3{+}1)-1}\right), \tag{11}$$

and

$$\hat{e}_i = \log(\hat{\mu}_i) + \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \tag{12}$$

respectively. The asymptotic variance-covariance matrix, the matrix mean squared error (MMSE), and the MSE of this estimator are given by

$$\mathrm{Var}-\mathrm{Cov}(\hat{\beta}_{\mathrm{PMQL-MLE}}) = B^{-1}, \tag{13}$$

$$\mathrm{MMSE}(\hat{\beta}_{\mathrm{PMQL-MLE}}) = \Gamma B^{-1}\Gamma^T, \tag{14}$$

$$\mathrm{MSE}(\hat{\beta}_{\mathrm{PMQL-MLE}}) = \mathrm{trace}\ (\mathrm{MMSE}(\hat{\beta}_{\mathrm{PMQL-MLE}})) = \sum_{j=1}^{p}\frac{1}{\lambda_j}, \tag{15}$$

where λ_j denotes the j^{th} eigenvalue of the matrix B and Γ represents the eigenvectors of matrix B.

In cases of strong multicollinearity among predictors, the matrix B may be ill-conditioned, yielding small eigenvalues and inflating the MSE, thereby distorting the inferential interpretation of the response-covariate relationship. To address the issue of multicollinearity within the PMQL-RM framework, Tharshan and Wijekoon [15] proposed the ridge estimator for the PMQL-RM (PMQL-RRE) as

$$\hat{\beta}_{\mathrm{PMQL-RRE}} = (B + kI_p)^{-1}B\hat{\beta}_{\mathrm{PMQL-MLE}}, \tag{16}$$

where $k > 0$is the ridge parameter. The bias vector and variance-covariance matrix of the $\hat{\beta}_{\mathrm{PMQL-RRE}}$ are given by

$$\mathrm{Bias}(\hat{\beta}_{\mathrm{PMQL-RRE}}) = -k(\Lambda + kI_a)^{-1}\beta, \tag{17}$$

$$\mathrm{Var}-\mathrm{Cov}(\hat{\beta}_{\mathrm{PMQL-RRE}}) = \Gamma(\Lambda + kI_a)^{-1}\Lambda(\Lambda + kI_a)^{-1}\Gamma^T \tag{18}$$

Subsequently, the MMSE and MSE for the PMQL-RRE estimator are derived as follows:

$$\begin{aligned}\mathrm{MSE}(\hat{\beta}_{\mathrm{PMQLRRE}}) &= \mathrm{Var}-\mathrm{Cov}(\hat{\beta}_{\mathrm{PMQL-RRE}}) + \mathrm{Bias}(\hat{\beta}_{\mathrm{PMQL-RRE}}) + \mathrm{Bias}(\hat{\beta}_{\mathrm{PMQL-RRE}})^{\mathrm{T}} \\ &= \Gamma(\Lambda + kI_p)^{-1}\Lambda(\Lambda + kI_p)^{-1}\Gamma^T + \mathrm{k}^2(\Lambda + kI_p)^{-1}\beta\beta^T(\Lambda + kI_p)^{-1},\end{aligned} \tag{19}$$

$$\begin{aligned}\mathrm{MSE}(\hat{\beta}_{\mathrm{PMQL-RRE}}) &= \mathrm{trace}(\mathrm{MSE}(\hat{\beta}_{\mathrm{PMQL-RRE}})) \\ &= \sum_{j=1}^{p}\frac{\lambda_i}{(\lambda_j+k)^2} + k^2\sum_{j=1}^{p}\frac{\phi_j^2}{(\lambda_j+k)^2}\end{aligned}$$

where Γ represents the spectral decomposition matrix, whose columns are the normalized eigenvectors of B. The diagonal matrix $\Lambda = \Gamma B\Gamma = diag(\lambda_1, \lambda_2, ..., \lambda_p)$contains the corresponding eigenvalues, and $\alpha_j$denotes the j^{th} element of the transformed coefficient vector $\Gamma^T\beta$.

Later, [16] introduced Liu estimator for PMQL-RM (PMQL-LE) as an alternative to the PMQL-RRE. The PMQL-LE is defined as:

$$\hat{\beta}_{\mathrm{PMQL-LE}} = (B + I_p)^{-1}(B + dI_p)\hat{\beta}_{\mathrm{PMQL-MLE}} \tag{20}$$

where $0 < d < 1$is the Liu shrinkage parameter. The Bias vector and variance-covariance matrix of PMQL-LE are given by:

$$\text{Bias }(\hat{\beta}_{\mathrm{PMQL-LE}}) = (d-1)(\Lambda + I_p)^{-1}\beta, \tag{21}$$

$$\mathrm{Var-Cov}(\hat{\beta}_{\mathrm{PMQL-LE}}) = \Gamma(\Lambda + I_p)^{-1}(\Lambda + dI_p)^{-1}\Lambda^{-1}(\Lambda + dI_p)^{-1}(\Lambda + I_p)^{-1}\Gamma^T \tag{22}$$

Then, the MMSE and MSE of $\hat{\beta}_{\mathrm{PMQL-LE}}$ are

$$\begin{aligned}\mathrm{MMSE}(\hat{\beta}_{\mathrm{PMQL-LE}}) &= \mathrm{Var-Cov}(\hat{\beta}_{\mathrm{PMQL-LE}}) + Bias(\hat{\beta}_{\mathrm{PMQL-LE}})Bias(\hat{\beta}_{\mathrm{PMQL-LE}})^T\\ &= \Gamma(\Lambda + I_p)^{-1}(\Lambda + dI_p)^{-1}\Lambda^{-1}(\Lambda + dI_p)^{-1}(\Lambda + I_p)^{-1}\Gamma^T\\ &\quad + (d-1)^2(\Lambda + I_p)^{-1}\beta\beta^T(\Lambda + I_p)^{-1},\\ \mathrm{MSE}(\hat{\beta}_{\mathrm{PMQL-LE}}) &= \mathrm{trace}(\mathrm{MSE}(\hat{\beta}_{\mathrm{PMQL-LE}}))\\ &\sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j+d)^2} + (d-1)^2\sum_{j=1}^{p}\frac{\phi_j^2}{(\lambda_j+d)^2}.\end{aligned}$$

In 2025, Alqasem, Hammad [17] presented the Liu-type estimator for PMQL-RM (PMQL-LTE) which is defined as

$$\hat{\beta}_{\mathrm{PMQL-LTE}} = (B + kI_p)^{-1}(B + kdI_p)\hat{\beta}_{\mathrm{PMQL-MLE}} \tag{23}$$

Then, the bias and variance-covariance matrix of the PMQL-LTE are

$$\text{Bias }(\hat{\beta}_{\mathrm{PMQL-LTE}}) = \mathrm{E}(\hat{\beta}_{\mathrm{PMQL-LTE}}) - \beta = k(d-1)(\Lambda + kI_p)^{-1}\beta, \tag{24}$$

$$\begin{aligned}\mathrm{Var-Cov}(\hat{\beta}_{\mathrm{PMQL-LTE}}) &= \mathrm{E}\left(\left[\hat{\beta}_{\mathrm{PMQL-LTE}} - \mathrm{E}(\hat{\beta}_{\mathrm{PMQL-LTE}})\right]\left[\hat{\beta}_{\mathrm{PMQL-LTE}} - \mathrm{E}(\hat{\beta}_{\mathrm{PMQL-LTE}})\right]^T\right)\\ &= \Gamma(\Lambda + kI_p)^{-1}(\Lambda + kdI_p)\Lambda^{-1}(\Lambda + kdI_p)(\Lambda + kdI_p)^{-1}\Gamma^T.\end{aligned} \tag{25}$$

After that, the MSE and the MMSE of the PMQL-LTE are defined as:

$$\begin{aligned}\mathrm{MMSE}(\hat{\beta}_{\mathrm{PMQL-LTE}}) &= \mathrm{Cov}(\hat{\beta}_{\mathrm{PMQL-LTE}}) + Bias(\hat{\beta}_{\mathrm{PMQL-LTE}})Bias(\hat{\beta}_{\mathrm{PMQL-LTE}})^T\\ &= \Gamma(\Lambda + kI_p)^{-1}(\Lambda + kdI_p)^{-1}\Lambda^{-1}(\Lambda + kdI_p)^{-1}(\Lambda + kI_p)^{-1}\Gamma^T\\ &\quad + (k(d-1))^2(\Lambda + kI_p)^{-1}\beta\beta^T(\Lambda + kI_p)^{-1},\end{aligned} \tag{26}$$

$$\begin{aligned}\mathrm{MSE}(\hat{\beta}_{\mathrm{PMQL-LTE}}) &= \mathrm{tr}(\mathrm{MMSE}(\hat{\beta}_{\mathrm{PMQL-LTE}}))\\ &= \sum_{j=1}^{p}\frac{(\lambda_i+kd)^2}{\lambda_i(\lambda_i+k)^2} + (d-1)^2k^2\sum_{j=1}^{p}\frac{\phi_j^2}{(\lambda_j+k)^2}\end{aligned} \tag{27}$$

3. The proposed estimator

To address the issue of multicollinearity in linear regression, Liu and Gao [22] presented the linearized ridge regression estimator (LRRE). This LRRE is a more versatile estimator that incorporates the least squares, ridge estimator, Liu estimator, and numerous other shrinkage estimators. Gao and Liu [23] investigated the LRRE characteristics and demonstrated that it achieves the lower bound of MSE in the category of generalized shrinkage estimators. Further, Gao and Liu [23] have demonstrated, based on theoretical and simulation studies, that the LRRE is the best estimator in the class of generalized shrinkage estimators according to the MSE criterion.

Having the success of the LRRE estimator in other GLMs, it is anticipated that this LRRE estimator will have a better estimation accuracy, in terms of the MSE, as compared to the current PMQL-RRE, PMQL-LE, and PMQL-LTE estimators, particularly under extreme multicollinearity in practice in count-data models [39, 40, 41].

The LRRE estimator for the PMQL-RM, PMQL-LRE, is defined as:

$$\hat{\beta}_{\mathrm{PMQL-LRE}} = (\mathrm{B} + I)^{-1}\left(\mathrm{B} + \gamma D\gamma^T\right)\,\hat{\beta}_{\mathrm{PMQL-MLE}} \tag{28}$$

where $D = diag\left(d_1, d_2, ..., d_{p+1}\right),\ d_j \in (-\infty, \infty)\ \forall_j = 1, 2, ..., p+1$ and $\gamma = \left(\gamma_1, \gamma_2, ..., \gamma_{p+1}\right)$ are the corresponding eigenvectors.

According to Eq. (31), bias and variance-covariance matrix of $\hat{\beta}_{\mathrm{PMQL-LRE}}$ are computed as

$$\mathrm{Bias}(\hat{\beta}_{\mathrm{PMQL-LRE}}) = \mathrm{E}\left[\hat{\beta}_{\mathrm{PMQL-LRE}}\right] - \beta = \left(\mathrm{B} + I\right)^{-1}\left(I - \gamma D \gamma'\right)\beta \tag{29}$$

$$\mathrm{Var-Cov}(\hat{\beta}_{\mathrm{PMQL-LRE}}) = \left(\mathrm{B} + I\right)^{-1}\left(\mathrm{B} + \gamma D \gamma'\right)\left(\mathrm{B}\right)^{-1}\left(\mathrm{B} + \gamma D \gamma'\right)\left(\mathrm{B} + I\right)^{-1} \tag{30}$$

Then, the MSE of the $\hat{\beta}_{\mathrm{PMQL-LRE}}$ is defined as

$$\mathrm{MSE}\left[\hat{\beta}_{\mathrm{PMQL-LRE}}\right] = \sum_{j=1}^{p} \frac{(\lambda_j + d_j)^2}{\lambda_j\left(\lambda_j + 1\right)^2} + \sum_{j=1}^{p} \frac{\left(1 - d_j\right)^2 \phi_j^2}{\left(\lambda_j + 1\right)^2} \tag{31}$$

4. Simulation study

A Monte Carlo simulation study will be conducted to compare the performance of the PMQL-MLE, PMQL-RRE, PMQL-LE, PMQL-LTE, and the PMQL-LRE estimators. The with a number of degrees of multicollinearity has been obtained McDonald and Galarneau (1975) as:

$$x_{ij} = (1 - r^2)^{1/2} m_{ij} + r m_{ij} + r m_{i,p+1,},\ i = 1, 2....., n, j = 1, 2, ..., p \tag{32}$$

where m_{ij} are independent standard normal pseudo-random numbers and r represents the correlation between the covariates. The response variable, y, of the PMQL-RM is generated from the PMQL distribution with (μ_i, ϕ, η) by using the inverse transform method, where $\mu_i = \exp(x_i^T \beta)$ with $\sum_{i=1}^{p} \beta_j^2 = 1$ and $\beta_1 = \beta_2 = ... = \beta_p$, and (ϕ, η) as dispersion parameters. The simulation study was designed based on the following factors: The correlation coefficient $r \in \{0.90, 0.95, 0.99\}$, the sample size $n \in \{30, 50, 200\}$, the number of the explanatory variables $p \in \{3, 7\}$, the value of the β_0 is -1 and 1. Finally, the values of the dispersion parameters were as $(\phi = 0.05, \eta = 0.4)$, $(\phi = 0.15, \eta = 1)$, and $(\phi = 0.8, \eta = 1.7)$. The simulation is repeated 5000 times" and the averaged MSE is obtained as

$$\mathrm{MSE}(\hat{\beta}) = \frac{\sum_{rr=1}^{5000} (\hat{\beta}_{rr} - \beta)^T (\hat{\beta}_{rr} - \beta)}{5000}, \tag{33}$$

where $\hat{\beta}_i$ is an estimator of β at the rr^{th} replication.

The simulation results are presented in Tables 1 - 6. In all these tables, smaller MSE means better estimator performance, so the focus is on how each factor p, n, r, ϕ, η affects the relative MSE of PMQL-MLE, PMQL-RRE, PMQL-LE, PMQL-LTE, and PMQL-LRE.

Tables 1 and 2 correspond to two different choices of the true regression vector and dispersion parameters, but their patterns are similar. PMQL-MLE always has the largest MSE among the five estimators, confirming its sensitivity to multicollinearity in PMQL-RM. For example, in Table 1 with $p = 3$, $n = 30$, $r = 0.90$, MSE is 5.548 for MLE versus 3.487 of RRE, 3.329 of LE, 3.199 of LTE, and 2.947 of LRE. Shrinkage estimators, RRE, LE, and LTE, systematically reduce MSE relative to MLE, but none of them dominates uniformly. In the same cell, LTE improves upon LE and RRE, but LRE still provides the smallest MSE. Further, As n increases from 30 to 150 or 200, MSE values decrease for all estimators, with the reduction most marked for MLE; however, LRE retains the best performance for each p and r, showing that its bias penalty remains small even in larger samples. Moreover, when correlation increases from 0.90 to 0.99, all MSE values increase, but the increase is most severe for MLE and mildest for LRE, indicating that the proposed estimator is the most robust to worsening multicollinearity. For $p = 7$, the same ordering holds, but MSEs are larger overall, reflecting the harder estimation problem with more highly correlated regressors. Even then, PMQL-LRE always has the minimal MSE in the reported cells, demonstrating that its advantage extends to moderately high dimensional settings.

Tables 3 and 4 use different combinations of regression and dispersion parameters, yet the ranking of estimators remains stable. The MSE levels are somewhat lower than in Tables 1-2 because of the specific parameter choices, but MLE still performs worst, and the proposed LRE remains best in nearly every cell. For instance, in Table 4 with $p = 3$, $n = 50$, $r = 0.90$, the MSEs are 3.245 of MLE, 1.184 of RRE, 1.027 of LE, 0.898 of LTE, and 0.647 of LRE. The gap between LTE and LRE is now particularly striking in some cells, showing that even relative to an already improved Liu-type estimator, the linearized ridge mechanism yields substantial extra MSE reduction. Further, increasing $p$from 3 to 7 again inflates MSE for all estimators, but LRE's increase is less severe, which suggests that the eigen-structure-based shrinkage in LRE scales more gracefully with dimension than the simpler ridge or Liu structures. These two tables demonstrate that the superiority of LRE does not depend on a specific configuration of true parameters; it is robust across a broad spectrum of scenarios for the PMQL distribution.

Tables 5 and 6 correspond to the last two parameter settings and further reinforce the earlier conclusions. Again, PMQL-MLE is uniformly worst, while shrinkage estimators offer marked improvements, with PMQL-LRE consistently delivering the smallest MSE. For example, in Table 5 with $p = 3$, $n = 50$, $r = 0.90$, MSE values are 2.858 of MLE, 0.797 of RRE, 0.639 of LE, 0.509 of LTE, and 0.263 of LRE. The relative gain of LRE is especially large when n is moderate and r is high, underscoring that the proposed estimator is particularly well suited to the most practically problematic situations (moderate sample size with strong multicollinearity). In Table 6, for $p = 3$,$n = 200$,$r = 0.90$, LRE attains an MSE as small as 0.017, much lower than the competing estimators (0.538 for RRE, 0.378 for LE, 0.248 for LTE), showing that the linearized ridge structure remains beneficial even in larger samples. Further, For $p = 7$, although all MSEs are higher, LRE still provides the best performance across all n and r, reinforcing its robustness.

Taken together, Tables 1–6 show a clear and consistent hierarchy across all examined configurations. PMQL-MLE suffers from inflated MSE in the presence of multicollinearity and always performs worst among the considered estimators. Further, the existing PMQL-RRE, PMQL-LE, and PMQL-LTE estimators provide steady MSE reductions relative to MLE, confirming the value of shrinkage in the PMQL-RM. In addition, the proposed PMQL-LRE dominates its competitors in terms of MSE for nearly all combinations of p, n, r and parameter settings, with particularly pronounced gains under severe multicollinearity and smaller sample sizes. These findings empirically support the theoretical mean squared error comparisons derived in the paper and confirm that the proposed linearized ridge-type estimator is a strong candidate for practical use in overdispersed count regression with multicollinear covariates. Moreover, Figures 1-3 show the performance of the used method over several correlation values.

Table 1. MSE values of the used estimators when $\beta_0 = -1$, $\phi = 0.05$ and $\eta = 0.4$

p	n	r	PMQL-MLE	PMQL-RRE	PMQL-LE	PMQL-LTE	PMQL-LRE
3	30	0.90	5.548	3.487	3.329	3.199	2.947
		0.95	5.644	3.537	3.347	3.224	3.088
		0.99	5.974	3.782	3.587	3.255	3.132
	50	0.90	4.687	2.626	2.468	2.338	2.089
		0.95	4.783	2.679	2.488	2.363	2.227
		0.99	5.113	2.921	2.727	2.394	2.271
	200	0.90	4.523	2.462	2.304	2.174	1.922
		0.95	4.619	2.512	2.322	2.199	2.063
		0.99	4.949	2.757	2.562	2.23	2.107
7	30	0.90	6.232	4.171	4.013	3.883	3.631
		0.95	6.328	4.221	4.031	3.908	3.772
		0.99	6.658	4.468	4.271	3.939	3.817
	50	0.90	5.371	3.312	3.152	3.022	2.773
		0.95	5.467	3.363	3.172	3.047	2.911
		0.99	5.797	3.605	3.411	3.078	2.955
	200	0.90	5.207	3.148	2.988	2.858	2.607
		0.95	5.303	3.199	3.008	2.883	2.747
		0.99	5.633	3.441	3.247	2.914	2.791

Table 2. MSE values of the used estimators when $\beta_0 = 1$, $\phi = 0.05$and $\eta = 0.4$

p	n	r	PMQL-MLE	PMQL-RRE	PMQL-LE	PMQL-LTE	PMQL-LRE
3	30	0.90	4.961	2.901	2.742	2.612	2.364
		0.95	5.057	2.952	2.763	2.637	2.501
		0.99	5.387	3.195	3.004	2.668	2.545
	50	0.90	4.104	2.039	1.881	1.751	1.502
		0.95	4.198	2.092	1.901	1.778	1.642
		0.99	4.527	2.334	2.143	1.807	1.684
	200	0.90	3.947	1.875	1.717	1.587	1.335
		0.95	4.032	1.925	1.735	1.612	1.477
		0.99	4.362	2.174	1.975	1.643	1.523
7	20	0.90	5.645	3.584	3.427	3.298	3.044
		0.95	5.741	3.634	3.444	3.321	3.185
		0.99	6.071	3.881	3.684	3.352	3.234
	50	0.90	4.784	2.723	2.565	2.435	2.188
		0.95	4.882	2.776	2.585	2.463	2.324
		0.99	5.213	3.018	2.824	2.491	2.368
	200	0.90	4.622	2.561	2.401	2.271	2.022
		0.95	4.718	2.612	2.421	2.297	2.164
		0.99	5.048	2.854	2.664	2.327	2.204

Table 3. MSE values of the used estimators when $\beta_0 = -1$, $\phi = 0.15$and $\eta = 1$

p	n	r	PMQL-MLE	PMQL-RRE	PMQL-LE	PMQL-LTE	PMQL-LRE
3	30	0.90	4.693	2.632	2.474	2.344	2.092
		0.95	4.789	2.682	2.492	2.369	2.233
		0.99	5.119	2.927	2.732	2.402	2.277
	50	0.90	3.832	1.771	1.613	1.483	1.234
		0.95	3.928	1.824	1.633	1.508	1.372
		0.99	4.258	2.068	1.872	1.539	1.418
	200	0.90	3.668	1.607	1.449	1.319	1.067
		0.95	3.764	1.657	1.467	1.344	1.208
		0.99	4.094	1.902	1.707	1.375	1.252
7	30	0.90	5.377	3.317	3.158	3.028	2.777
		0.95	5.473	3.368	3.177	3.053	2.917
		0.99	5.803	3.613	3.418	3.084	2.962
	50	0.90	4.517	2.455	2.297	2.167	1.918
		0.95	4.612	2.508	2.317	2.192	2.057
		0.99	4.942	2.753	2.558	2.223	2.102
	200	0.90	4.352	2.293	2.133	2.003	1.752
		0.95	4.448	2.344	2.153	2.028	1.892
		0.99	4.778	2.587	2.392	2.059	1.948

Table 4. MSE values of the used estimators when $\beta_0 = 1$, $\phi = 0.15$and $\eta = 1$

p	n	r	PMQL-MLE	PMQL-RRE	PMQL-LE	PMQL-LTE	PMQL-LRE
3	30	0.90	4.108	2.045	1.887	1.757	1.505
		0.95	4.202	2.095	1.905	1.782	1.647
		0.99	4.532	2.342	2.145	1.813	1.691
	50	0.90	3.245	1.184	1.027	0.898	0.647
		0.95	3.343	1.237	1.048	0.923	0.785
		0.99	3.672	1.479	1.285	0.952	0.829
	200	0.90	3.092	1.025	0.862	0.732	0.482
		0.95	3.177	1.074	0.883	0.757	0.622
		0.99	3.507	1.315	1.122	0.788	0.665
7	30	0.90	4.792	2.729	2.572	2.443	2.189
		0.95	4.888	2.779	2.589	2.468	2.332
		0.99	5.218	3.027	2.829	2.497	2.375
	50	0.90	3.929	1.868	1.713	1.583	1.333
		0.95	4.025	1.921	1.734	1.605	1.469
		0.99	4.355	2.163	1.969	1.647	1.513
	200	0.90	3.765	1.707	1.548	1.418	1.165
		0.95	3.863	1.757	1.567	1.442	1.305
		0.99	4.193	1.999	1.805	1.472	1.349

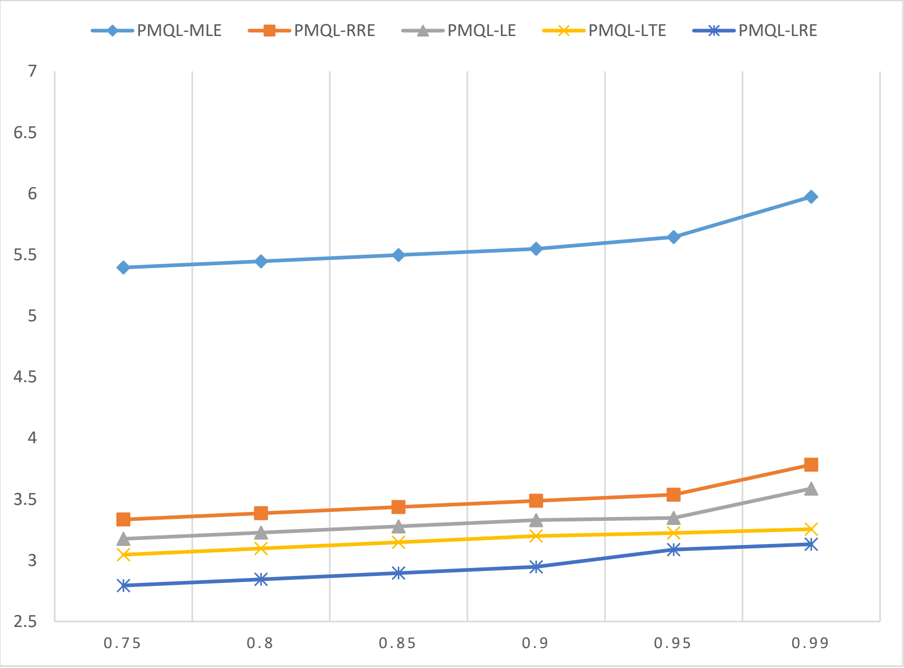

Figure 1. The performance of used methods when $p = 3$, $n = 30$, $\beta_0 = -1$, $\phi = 0.05$, and $\eta = 0.4$

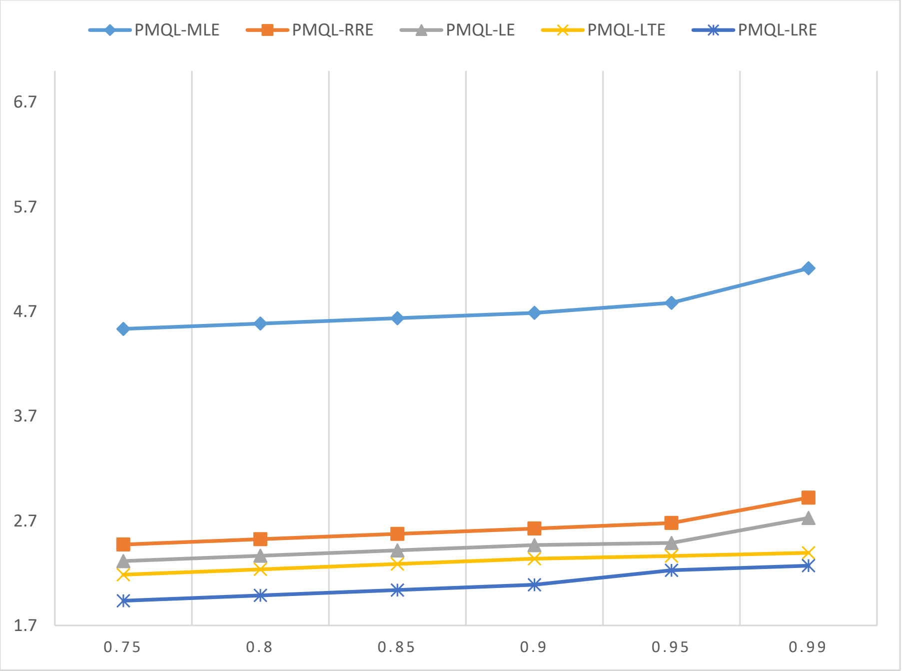

Figure 2. The performance of used methods when $p = 3$, $n = 50$, $\beta_0 = -1$, $\phi = 0.05$, and $\eta = 0.4$

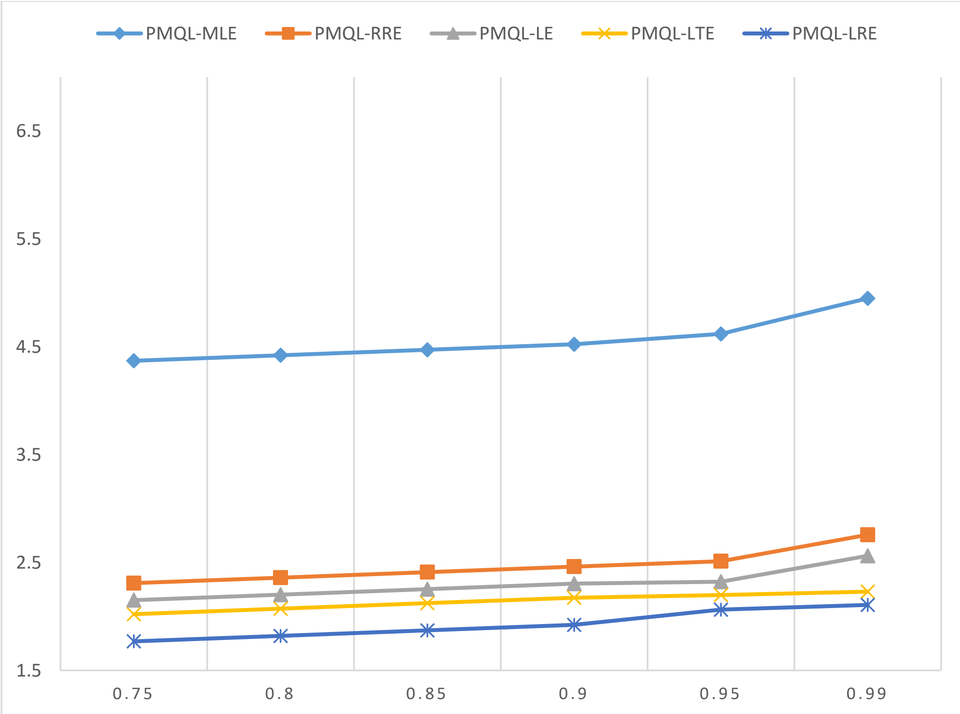

Figure 3. The performance of used methods when $p = 3$, $n = 200$, $\beta_0 = -1$, $\phi = 0.05$, and $\eta = 0.4$

Table 5. MSE values of the used estimators when $\beta_0 = -1$, $\phi = 0.8$and $\eta = 1.7$

p	n	r	PMQL-MLE	PMQL-RRE	PMQL-LE	PMQL-LTE	PMQL-LRE
3	30	0.90	3.719	1.658	1.502	1.373	1.118
		0.95	3.815	1.708	1.518	1.395	1.259
		0.99	4.145	1.953	1.758	1.428	1.303
	50	0.90	2.858	0.797	0.639	0.509	0.263
		0.95	2.954	0.854	0.659	0.534	0.398
		0.99	3.284	1.094	0.898	0.565	0.444
	200	0.90	2.694	0.633	0.475	0.345	0.093
		0.95	2.792	0.683	0.493	0.375	0.234
		0.99	3.123	0.928	0.733	0.401	0.278
7	20	0.90	4.403	2.343	2.184	2.054	1.803
		0.95	4.499	2.394	2.203	2.079	1.943
		0.99	4.829	2.639	2.444	2.115	1.988
	50	0.90	3.543	1.481	1.323	1.193	0.944
		0.95	3.638	1.534	1.343	1.218	1.083
		0.99	3.968	1.777	1.584	1.249	1.127
	200	0.90	3.378	1.319	1.159	1.029	0.778
		0.95	3.474	1.374	1.179	1.054	0.918
		0.99	3.804	1.613	1.418	1.085	0.974

5. Real data application

In this section, the applicability of the PMQL-LRE estimator is demonstrated by using the Swedish football data set which includes the Swedish football teams' performance in the top Swedish league (Allsvenskan) during the year 2012. The data set is publicly available ”at http://www.football-data.co.uk/sweden.php. The number of full-time home team goals is consedered as the response variable with six covariates by fitting the PMQL-RM model. This data set contains 242 observations. The covariates are the pinnacle home win odds (X1), pinnacle away win odds

Table 6. MSE values of the used estimators when $\beta_0 = 1$, $\phi = 0.8$and $\eta = 1.7$

p	n	r	PMQL-MLE	PMQL-RRE	PMQL-LE	PMQL-LTE	PMQL-LRE
3	30	0.90	3.624	1.561	1.403	1.273	1.021
		0.95	3.718	1.611	1.421	1.298	1.163
		0.99	4.048	1.858	1.661	1.329	1.207
	50	0.90	2.761	0.701	0.543	0.414	0.163
		0.95	2.859	0.753	0.564	0.439	0.301
		0.99	3.188	0.995	0.801	0.468	0.345
	200	0.90	2.608	0.538	0.378	0.248	0.017
		0.95	2.693	0.587	0.398	0.273	0.138
		0.99	3.023	0.831	0.639	0.304	0.181
7	30	0.90	4.307	2.245	2.088	1.959	1.705
		0.95	4.404	2.295	2.105	1.984	1.849
		0.99	4.734	2.543	2.345	2.013	1.891
	50	0.90	3.445	1.384	1.227	1.098	0.849
		0.95	3.541	1.437	1.247	1.121	0.985
		0.99	3.871	1.679	1.485	1.163	1.029
	200	0.90	3.281	1.223	1.064	0.934	0.681
		0.95	3.379	1.273	1.083	0.958	0.821
		0.99	3.709	1.515	1.321	0.988	0.865

(X2), maximum odds portal home win (X3), maximum odds portal away win (X4), average odds portal home win (X5), and average odds portal away win (X6). According to Tharshan and Wijekoon [15], the response variable is over-dispersed, where the variance to mean ratio equals 1:201 > 1.

To assess the performance of the PMQL-LRE estimator relative to other estimators, Table 7 presents the MSE values and regression coefficients values. All estimators yield the same sign pattern for the regression coefficients, indicating a consistent qualitative interpretation of the effects of the odds variables on home-team scoring, while the shrinkage estimators produce noticeably smaller coefficient magnitudes than the MLE. Among them, the PMQL-LRE provides the most regularized coefficients and achieves the smallest MSE (0.127), compared with 0.371 for the MLE, 0.160 for the ridge estimator, 0.148 for the Liu estimator, and 0.139 for the Liu-type estimator. These results corroborate the simulation findings and demonstrate that the linearized ridge-type estimator attains a substantial improvement in estimation accuracy and stability for real overdispersed and multicollinear count data", without altering the substantive interpretation of the covariate effects.

Table 7. The estimated MSE and regression coefficient of the used estimators for real data application

Parameter	PMQL-MLE	PMQL-RRE	PMQL-LE	PMQL-LTE	PMQL-LRE
$\hat{\beta}_0$	-1.046	-0.939	-0.927	-0.921	-0.915
$\hat{\beta}_1$	0.408	0.377	0.365	0.359	0.355
$\hat{\beta}_2$	0.408	0.413	0.401	0.397	0.389
$\hat{\beta}_3$	0.407	0.482	0.472	0.464	0.458
$\hat{\beta}_4$	0.409	0.427	0.415	0.410	0.405
$\hat{\beta}_5$	0.408	0.335	0.323	0.317	0.311
$\hat{\beta}_6$	0.406	0.367	0.355	0.349	0.348
MSE	0.371	0.160	0.148	0.139	0.127

6. Conclusion

The study introduces a new linearized ridge-type estimator for PMQL-RM to handle multicollinearity in overdispersed count data. Building on the structure of mixed Poisson models, the proposed PMQL-LRE incorporates an additional biasing mechanism that systematically reduces the inflated MSE of the classical PMQL-MLE. In the theoretical investigation, we derived the bias, variance–covariance matrix, and matrix MSE of the PMQL-LRE, and established conditions under which it dominates PMQL-MLE, as well as existing shrinkage estimators such as PMQL-RRE, PMQL-LE, and PMQL-LTE, with respect to both matrix and scalar MSE criteria. These results confirm that the PMQL-LRE is a more efficient member of the shrinkage family for PMQL-RM, particularly when the regressors exhibit strong multicollinearity. The Monte Carlo simulation study, conducted across a wide range of correlation levels, sample sizes, numbers of covariates, and dispersion parameter settings, consistently shows that the PMQL-LRE achieves the smallest average MSE, with its relative superiority becoming more pronounced under severe multicollinearity. Moreover, the empirical application to real overdispersed count data demonstrates that the PMQL-LRE yields more stable coefficient estimates and improved predictive performance compared with PMQL-MLE and its ridge- and Liu-type competitors. Overall, the findings indicate that the proposed estimator provides a practical and powerful alternative for analyzing multicollinear overdispersed count data within the PMQL-RM framework, and they motivate further extensions to other mixed Poisson and complex count regression models.

REFERENCES

1. F. Al-Hussein, et al., *Developing count regression techniques for predicting the number of new type 2 diabetes cases in Saudi Arabia*, PLoS One, vol. 21, no. 1, p. e0341436, 2026.
2. A.B. Asnake, et al., *Non-reassuring fetal heart rate and associated factors among laboring mothers at southern public hospitals in Ethiopia: a poisson regression model*, Scientific Reports, 2026.
3. I. Siramaneerat, F. Agushybana, and P. Sunthrayuth, *A mixed-effects multilevel Poisson regression analysis of diarrhea prevalence among under-five children in Indonesia*, Journal of Health, Population, and Nutrition, 2026.
4. Y. Al-Taweel and Z. Algamal, *Almost unbiased ridge estimator in the zero-inflated Poisson regression model*, TWMS Journal Of Applied And Engineering Mathematics, vol. 12, no. 1, pp. 235–246, 2022.
5. Z.Y. Algamal, *Diagnostic in Poisson regression models*, Electronic Journal of Applied Statistical Analysis, vol. 5, no. 2, pp. 178–186, 2012.
6. F. Noeel and Z.Y. Algamal, *Almost unbiased ridge estimator in the count data regression models*, Electronic Journal of Applied Statistical Analysis, vol. 14, no. 1, pp. 44–57, 2021.
7. Z.Y. Algamal, et al., *Modified jackknife ridge estimator for the Conway-Maxwell-Poisson model*, Scientific African, vol. 19, 2023.
8. N.S. Hawa, et al., *Bootstrap Liu-type estimator for Conway-Maxwell-Poisson regression model*, Communications in Statistics - Simulation and Computation, pp. 1–12, 2025.
9. M.N. Newaz, et al., *An overdispersed count regression model for analyzing road accident fatalities and injuries in Bangladesh*, PLoS One, vol. 21, no. 2, p. e0341775, 2026.
10. E. Altun, *A new model for over-dispersed count data: Poisson quasi-Lindley regression model*, Mathematical Sciences, vol. 13, no. 3, pp. 241–247, 2019.
11. M.J. Olmo-Jiménez, et al., *Enhanced regression modelling for both under-and overdispersed count data*, International Journal of Data Science and Analytics, vol. 21, no. 1, p. 63, 2026.
12. C.S. Kumar and P. Sathyan, *The weighted hyper-Poisson model for overdispersed and underdispersed situations*, Journal of Statistical Computation and Simulation, pp. 1–29, 2026.
13. M. Vardaki and H. Papageorgiou, *Poisson quasi-Lindley and Poisson-new XLindley univariate and bivariate models: derivation techniques and automobile insurance applications*, AIMS Mathematics, vol. 11, no. 2, pp. 3772–3810, 2026.
14. F.M. Alghamdi, et al., *Discrete Poisson Quasi-XLindley distribution with mathematical properties, regression model, and data analysis*, Journal of Radiation Research and Applied Sciences, vol. 17, no. 2, 2024.
15. R. Tharshan and P. Wijekoon, *Ridge estimator in a mixed Poisson regression model*, Communications in Statistics - Simulation and Computation, vol. 53, no. 7, pp. 3253–3270, 2024.
16. R. Tharshan and P. Wijekoon, *On a mixed Poisson Liu regression estimator for overdispersed and multicollinear count data*, The Scientific World Journal, vol. 2022, no. 1, p. 8171461, 2022.
17. O.A. Alqasem, et al., *A new Liu-type estimator in a mixed Poisson regression model*, Scientific Reports, vol. 15, no. 1, p. 19042, 2025.
18. R. Tharshan and P. Wijekoon, *Poisson-Modification of Quasi Lindley regression model for over-dispersed count responses*, Communications in Statistics - Simulation and Computation, vol. 53, no. 3, pp. 1258–1273, 2022.
19. Y.M. Bulut, et al., *Modified ridge estimator in the Bell regression model*, Journal of Inverse and Ill-posed Problems, vol. 32, no. 6, pp. 1081–1091, 2024.
20. S. Seifollahi, H. Bevrani, and Z.Y. Algamal, *Improved estimators in bell regression model with application*, Journal of Statistical Computation and Simulation, vol. 94, no. 12, pp. 2710–2726, 2024.
21. R. Tharshan and P. Wijekoon, *A new mixed Poisson distribution for over-dispersed count data: Theory and applications*, Reliability: Theory & Applications, vol. 17, no. 1 (67), pp. 33–51, 2022.
22. X.-Q. Liu and F. Gao, *Linearized Ridge Regression Estimator in Linear Regression*, Communications in Statistics - Theory and Methods, vol. 40, no. 12, pp. 2182–2192, 2011.
23. F. Gao and X.-Q. Liu, *Linearized Ridge Regression Estimator Under the Mean Squared Error Criterion in a Linear Regression Model*, Communications in Statistics - Simulation and Computation, vol. 40, no. 9, pp. 1434–1443, 2011.
24. Algamal, Z. Y., & Asar, Y. *Liu-type estimator for the gamma regression model* , Communications in Statistics-Simulation and Computation, vol. 49, no . 8, p. 2035-2048, 2020.

25. Algamal, Z. Y., & Lee, M. H. *A new adaptive L1-norm for optimal descriptor selection of high-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives* , SAR and QSAR in Environmental Research, vol. 28, no. 1 , p. 75-90, 2017.
26. Algamal, Z. Y., Qasim, M. K., & Ali, H. T. M. *A QSAR classification model for neuraminidase inhibitors of influenza A viruses (H1N1) based on weighted penalized support vector machine* , SAR and QSAR in Environmental Research, vol. 28 , no. 5, p. 415-426 , 2017.
27. Algamal, Z. Y., Lee, M. H., & Al-Fakih, A. M. *High-dimensional quantitative structure–activity relationship modeling of influenza neuraminidase a/PR/8/34 (H1N1) inhibitors based on a two-stage adaptive penalized rank regression* , Journal of Chemometrics, vol. 30 , no.2 m p. 50-57 , 2016.
28. Algamal, Z. Y., Lee, M. H., Al-Fakih, A. M., & Aziz, M. *High-dimensional QSAR classification model for anti-hepatitis C virus activity of thiourea derivatives based on the sparse logistic regression model with a bridge penalty* , Journal of Chemometrics, vol. 31 , p.6, p. e2889 , 2017.
29. Algamal, Z. Y., Qasim, M. K., Lee, M. H., & Ali, H. T. M. *High-dimensional QSAR/QSPR classification modeling based on improving pigeon optimization algorithm* , Chemometrics and Intelligent Laboratory Systems, vol. 206, p. 104170 , 2020.
30. Abonazel, M. R., Algamal, Z. Y., Awwad, F. A., & Taha, I. M. *A new two-parameter estimator for beta regression model: method, simulation, and application* , Frontiers in Applied Mathematics and Statistics, vol. 7,p. 780322 ,2022.
31. Algamal, Z. Y., & Abonazel, M. R. *Developing a Liu-type estimator in beta regression model* , Concurrency and Computation: Practice and Experience, vol.34 ,no. 5 , p. e6685.
32. Algamal, Z., & Ali, H. M. *An efficient gene selection method for high-dimensional microarray data based on sparse logistic regression* , Electronic Journal of Applied Statistical Analysis, vol . 10 , no . 1 , 242-256 , 2017.
33. Salih, A. M., Algamal, Z., & Khaleel, M. A. *A new ridge-type estimator for the gamma regression model*, Iraqi Journal for Computer Science and Mathematics, vol . 5 , no .1 , p. 85-98, 2023.
34. Alharthi, A. M., Kadir, D. H., Al-Fakih, A. M., Algamal, Z. Y., Al-Thanoon, N. A., & Qasim, M. K. *Quantitative structure-property relationship modelling for predicting retention indices of essential oils based on an improved horse herd optimization algorithm*, SAR and QSAR in Environmental Research, vol . 34 , no .10 , p. 831-846, 2023.
35. Algamal, Z. Y., Alhamzawi, R., & Ali, H. T. M. *Gene selection for microarray gene expression classification using Bayesian Lasso quantile regression*, Computers in biology and medicine, vol. 97, p. 145-152, 2018
36. Algamal, Z. Y., & Lee, M. H. *A novel molecular descriptor selection method in QSAR classification model based on weighted penalized logistic regression*, Journal of Chemometrics, vol. 31, no.(10), e2915, 2017
37. Qasim, M. K., Algamal, Z. Y., & Ali, H. M. *A binary QSAR model for classifying neuraminidase inhibitors of influenza A viruses (H1N1) using the combined minimum redundancy maximum relevancy criterion with the sparse support vector machine*, SAR and QSAR in Environmental Research, vol. 29, no.(7), p.517-527, 2018
38. Al-Taweel, Y., & Algamal, Z. *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model*, Periodicals of Engineering and Natural Sciences, vol. 8, no.(1), p.248-255, 2020
39. Shamany, R., Alobaidi, N. N., & Algamal, Z. Y. *A new two-parameter estimator for the inverse Gaussian regression model with application in chemometrics*, Electronic Journal of Applied Statistical Analysis, vol. 12, no.(2), p.453-464, 2019
40. Awwad, F. A., Odeniyi, K. A., Dawoud, I., Algamal, Z. Y., Abonazel, M. R., Kibria, B. G., & Eldin, E. T. *New two-parameter estimators for the logistic regression model with multicollinearity*, WSEAS Trans. Math, vol. 21, p.403-414, 2022
41. Algamal, Z. Y., Qasim, M. K., Lee, M. H., & Ali, H. T. M. *QSAR model for predicting neuraminidase inhibitors of influenza A viruses (H1N1) based on adaptive grasshopper optimization algorithm*, SAR and QSAR in Environmental Research, vol.31, no.11, p.803-814, 2020.
42. Al-Fakih, A. M., Algamal, Z. Y., Lee, M. H., Aziz, M., & Ali, H. T. M. *QSAR classification model for diverse series of antifungal agents based on improved binary differential search algorithm*, SAR and QSAR in Environmental Research, vol.30, no.2, p.131-143, 2019.
43. Alharthi, A. M., Kadir, D. H., Al-Fakih, A. M., Algamal, Z. Y., Al-Thanoon, N. A., & Qasim, M. K. *Improving golden jackel optimization algorithm: An application of chemical data classification*, Chemometrics and Intelligent Laboratory Systems, vol.250, 105149, 2024.
44. Qasim, O. S., & Algamal, Z. Y. *A gray wolf algorithm for feature and parameter selection of support vector classification* , International Journal of Computing Science and Mathematics, vol. 13, 93-102, 2021.
45. Ahmed, R. A., Algamal, Z. Y., & Mohamud, L. A. *A Poisson-Modified Quasi-Lindley Regression Approach for Multicollinear Data using Kibria–Lukman Estimator*, Scientific African, vol. 33, e03465, 2026.
46. Al-Fakih, A. M., Qasim, M. K., Algamal, Z. Y., Alharthi, A. M., & Zainal-Abidin, M. H. *QSAR classification model for diverse series of antifungal agents based on binary coyote optimization algorithm* SAR and QSAR in Environmental Research, vol.34, no.4, p. 285-298, 2023.
47. Al-Fakih, A. M., Algamal, Z. Y., & Qasim, M. K. *An improved opposition-based crow search algorithm for biodegradable material classification* SAR and QSAR in Environmental Research, vol.33, no.5, p. 403-415, 2022.
48. Hawa, N. S., Mustafa, M. Y., Kibria, B. G., & Algamal, Z. Y. *Bootstrap Liu-type estimator for Conway-Maxwell-Poisson regression model* Communications in Statistics-Simulation and Computation, p. 1-12, 2025.
49. Naziyah, A. A., & Algamal, Z. Y. *Jackknifed Liu-type estimator in Poisson regression model* Journal of the Iranian Statistical Society, vol.19, no.1, p. 21-37, 2020.
50. Al-Taweel, Y., & Algamal, Z. Y. *Some almost unbiased ridge regression estimators for the zero-inflated negative binomial regression model* Periodicals of Engineering and Natural Sciences, vol.8, no.1, p. 248-255, 2020.
51. Al-Fakih, A. M., Qasim, M. K., Algamal, Z. Y., Alharthi, A. M., & Zainal-Abidin, M. H. *QSAR classification model for diverse series of antifungal agents based on binary coyote optimization algorithm* SAR and QSAR in Environmental Research, vol.34, no.4, p. 285-298, 2023.
52. Z. Y. Algamal, A. F., Lukman, M. R., Abonazel, & F. A. Awwad *Performance of the ridge and Liu estimators in the zero-inflated Bell regression model* Journal of Mathematics, vol. 22, no. 1, p. 9503460, 2022.