F1000Research

Check for updates

RESEARCH ARTICLE

REVISED Modified Kibria-Lukman (MKL) estimator for the Poisson Regression Model: application and simulation [version 2; peer review: 2 approved, 1 approved with reservations]

Benedicta B. Aladeitan [iD][1,2], Olukayode Adebimpe[1,2], Adewale F. Lukman [iD][1,2], Olajumoke Oludoun [iD][1,2], Oluwakemi E. Abiodun [iD][1,2]

[1]Department of Physical Sciences, (Mathematics) Landmark University, Omu-Aran, Kwara State, +234, Nigeria
[2]SDG 13 (Climate Action), Landmark University, Omu-Aran, Kwara State, Nigeria

Abstract

Background: Multicollinearity greatly affects the Maximum Likelihood Estimator (MLE) efficiency in both the linear regression model and the generalized linear model. Alternative estimators to the MLE include the ridge estimator, the Liu estimator and the Kibria-Lukman (KL) estimator, though literature shows that the KL estimator is preferred. Therefore, this study sought to modify the KL estimator to mitigate the Poisson Regression Model with multicollinearity.
Methods: A simulation study and a real-life study was carried out and the performance of the new estimator was compared with some of the existing estimators.
Results: The simulation result showed the new estimator performed more efficiently than the MLE, Poisson Ridge Regression Estimator (PRE), Poisson Liu Estimator (PLE) and the Poisson KL (PKL) estimators. The real-life application also agreed with the simulation result.
Conclusions: In general, the new estimator performed more efficiently than the MLE, PRE, PLE and the PKL when multicollinearity was present.

Keywords

Linear regression model, generalized regression model, Ridge estimator, Liu estimator, KL estimator.

Open Peer Review

Approval Status ✔ ✔ ?

	1	2	3
version 2 (revision) 14 Dec 2021	✔ view	✔ view	
version 1 08 Jul 2021	? view	? view	? view

1. **Mohammad Arashi** [iD], Ferdowsi University of Mashhad, Mashhad, Iran

2. **Muhammad Amin** [iD], University of Sargodha, Sargodha, Pakistan

3. **Nimet Özbay** [iD], Çukurova University, Adana, Turkey

Any reports and responses or comments on the article can be found at the end of the article.

Corresponding author: Benedicta B. Aladeitan (aladeitan.benedicta@lmu.edu.ng)

Author roles: Aladeitan BB: Conceptualization, Formal Analysis, Methodology, Writing – Original Draft Preparation; **Adebimpe O**: Supervision, Writing – Review & Editing; **Lukman AF**: Resources, Writing – Review & Editing; **Oludoun O**: Writing – Review & Editing; **Abiodun OE**: Writing – Review & Editing

> **REVISED** **Amendments from Version 1**
>
> The difference between this version and the first is that all corrections that were raised by the three reviewers were effected. The new version included more equations to simplify methods earlier discussed as raised by the reviewers.
>
> **Any further responses from the reviewers can be found at the end of the article**

Introduction

A special case of the Generalized Linear Models (GLM) is the Poisson Regression Model (PRM) which is generally applied for count or frequency data modelling. Other count data models include: Bell regression model, Negative binomial regression model, zero inflated bell regression model, zero inflated regression model (Amin *et al.*, 2020, 2021; Sami *et al.*, 2021; Rashad and Algamal, 2019; Majid *et al.*, 2021). The PRM is employed to model the relationship between a response variable and one or more explanatory variable where the response variable denotes a rare event or count data. The response variable also takes the form of a non-negative variable, and it is applicable in the following fields: economics, health, social and physical sciences. The Maximum Likelihood Estimation (MLE) method is popularly used to estimate the regression coefficient in a PRM. In both a Linear Regression Model (LRM) and Generalized Linear Model (GLM), MLE suffers a setback when the explanatory variables are correlated, which implies multicollinearity. Multicollinearity effects include large variance and regression coefficient covariances, negligible t-ratio and a high coefficient of determination (R-square) values. Alternative estimators to the MLE in the linear regression model include the ridge regression estimator by Hoerl and Kennard (1970), Liu estimator by Liu (1993), Liu-type estimator by Liu (2003), two-parameter estimator by Özkale and Kaciranlar (2007), *r-d* class estimator Kaçiranlar and Sakallioğlu (2007), *k-d* class estimator Sakallioglu and Kaciranlar (2008), a two-parameter estimator by Yang and Chang (2010), modified two-parameter estimator by Dorugade (2014), modified ridge-type estimator by Lukman *et al.* (2019), modified Liu estimator by Lukman *et al.* (2020), Kibria-Lukman (KL) estimator by Kibria and Lukman (2020), modified new two-parameter estimator by Ahmad and Aslam (2020), the modified Liu ridge type estimator by Aslam and Ahmad (2020) and the DK estimator by Dawoud and Kibria (2020) among others. Researchers have extended some of these existing estimators in LRM to the PRM. Mansson *et al.* (2012) introduced the Liu estimator into the PRM. The modified **jackknifed ridge estimator** for the PRM was introduced by Türkan and Özel (2016). The ridge estimator was introduced into the PRM by Månsson and Shukur (2011). A new two-parameter for PRM was developed by Asar and Genç (2017). Recently, Poisson KL estimator was developed by Lukman *et al.* (2021) for combating multicollinearity in the PRM.

In this study, we propose the Modified Kibria-Lukman estimator to handle multicollinearity in PRM. The estimator is a single parameter estimator which makes it less computationally intensive as compared with the two-parameter estimators. Also, since the Kibria-Lukman estimator is found to outperform the Ridge and the Liu estimators, it is expected that the modification in this study will enhance the performance of the Kibria-Lukman estimator. Furthermore, we compared the performance of the estimator with the Poisson Maximum Likelihood Estimator (PMLE), Poisson Ridge Regression Estimator (PRE), Poisson Liu Estimator (PLE) and the Poisson KL estimator (PKLE).

Methods

Given that the response variable, y_i is in the form of count data, then it is assumed to follow a Poisson distribution as $P_o(\mu_i)$ where $\mu_i = e^{(x_i\beta)}$, and $\ln \mu_i = (x_i\beta)$, x_i is the ith row of matrix X which is a $n \times (p+1)$ data matrix with p explanatory variables and β is a $(p+1) \times 1$ vector of coefficients. The log likelihood of the model is given as:

$$l(\mu;y) = \sum_{i=1}^{n} \gamma_i \log(e^{(xi\beta)}) - \left(\sum_{i=1}^{n} e^{(xi\beta)}\right)_i - \log\left(\prod_{i=1}^{n} y_i!\right) \tag{2.1}$$

The most common method of maximizing the likelihood function is to use the iterated weighted least squares (IWLS) algorithm which results to:

$$\widehat{\beta}_{MLE} = \left(X'\widehat{L}X\right)^{-1}\left(X'\widehat{L}\hat{z}\right) \tag{2.2}$$

where $\widehat{L} = diag[\widehat{\mu}_i]$ and $\hat{z}$ is a vector while the ith element equals $\hat{z}_i = \log(\widehat{\mu}_i) + \frac{y_i - \widehat{\mu}_i}{\widehat{\mu}_i}$.

The MLE is normally distributed with a covariance matrix that is equivalent to the inverse of the second derivative as:

$$Cov\left(\widehat{\beta}_{MLE}\right) = \left(-E\left(\frac{\partial^2 l}{\partial\beta_j\partial\beta'_k}\right)\right)^{-1} = \left(X'\widehat{L}X\right)^{-1} \tag{2.3}$$

and the mean squared error is given as:

$$E\left(\widehat{\beta}_{MLE}\right) = E\left(\widehat{\beta}_{MLE} - \beta\right)'\left(\widehat{\beta}_{MLE} - \beta\right) = \mathrm{tr}\left(X'\widehat{L}X\right)^{-1} = \sum_{j=1}^{P}\frac{1}{\lambda_j} \tag{2.4}$$

where λ_j is the jth eigen value of the $\left(X'\widehat{V}X\right)$ matrix.

The Ridge estimator was adopted by Månsson and Shukur (2011) to solve multicollinearity problem in count data. The estimator is defined as follows:

$$\widehat{\beta}_{PRE} = \left(X'\widehat{L}X + kI\right)^{-1}X'\widehat{L}X\widehat{\beta}_{MLE} \tag{2.5}$$

where $k = \left(\frac{1}{max(\alpha_i^2)}\right)$ and $(k > 0)$.

The mean squared error is:

$$MSE\left(\widehat{\beta}_{PRE}\right) = \sum_{j=1}^{p}\frac{\lambda_j}{\left(\lambda_j+k\right)^2} + k^2\sum_{j=1}^{p}\frac{\widehat{\alpha}_j^2}{\left(\lambda_j+k\right)^2} \tag{2.6}$$

β_{PRE} is effective in practice but it is a complicated function of the biasing parameter k (Liu, 1993).

Mansson *et al.* (2012) developed the Liu estimator to the Poisson regression model as:

$$\widehat{\beta}_{PLE} = \left(X'\widehat{L}X + I\right)^{-1}\left(X'\widehat{L}X + d\widehat{\beta}\right)\widehat{\beta}_{ML} \tag{2.7}$$

where

$$\hat{d} = max\left(0\frac{\widehat{\alpha}_j^2 - 1}{\widehat{\alpha}_j^2 + \frac{1}{\lambda_j}}\right), 0 \le d \le 1. \tag{2.8}$$

The MSE for the Liu estimator is defined as:

$$MSE\left(\widehat{\beta}_{PLE}\right) = \sum_{j=1}^{P}\frac{\left(\lambda_j+d\right)^2}{\lambda_j\left(\lambda_j+1\right)^2} + (d-1)^2\sum_{j-1}^{p}\frac{\alpha_j^2}{\left(\lambda_j+1\right)^2} \tag{2.9}$$

where λ_j is the jth eigenvalue of $X'\widehat{L}X$ and α_j is the jth element of α.

The KL estimator was proposed by Kibria and Lukman (2020) as a means of mitigating the effect of multicollinearity on parameter estimation. The estimator is defined as

$$\widehat{\beta}_{KL} = (X'X + k)^{-1}(X'X - k)\widehat{\beta}_{MLE} \tag{2.10}$$

By means of extension, the Poisson K-L estimator was proposed by Lukman *et al.* (2021) as follows:

$$\widehat{\beta}_{PKL} = \left(X'\widehat{L}X + k\right)^{-1}\left(X'\widehat{L}X - k\right)\widehat{\beta}_{MLE} \tag{2.11}$$

$$MSE\left(\widehat{\beta}_{PKL}\right) = \sum_{j=1}^{p}\left(\frac{\left(\lambda_j-k\right)^2}{\lambda_j\left(\lambda_j+k\right)^2}\right) + 4k^2\sum_{j=1}^{p}\left(\frac{\alpha_j^2}{\left(\lambda_j+k\right)^2}\right) \tag{2.12}$$

where $k = min\left(\frac{\alpha_j^2}{2\alpha_j^2 + \frac{1}{\lambda_j}}\right)$ and $k > 0$.

The Poisson Modified KL estimator (PMKL)

The proposed estimator is obtained as follows: $\widehat{\beta}_{MLE}$ in equation (2.11) is replaced with the ridge estimator. Thus, we have:

$$\widehat{\beta}_{MKL} = (X'X+k)^{-1}(X'X-k)(X'X+k)^{-1}X'y \tag{2.13}$$

The properties of the new estimator include:

$$E\left(\widehat{\beta}_{MKL}\right) = (X'X+kI)^{-1}(X'X-kI)(X'X+kI)^{-1}X'X\beta \tag{2.14}$$

$$Bias\left(\widehat{\beta}_{MKL}\right) = (X'X+kI)^{-1}(X'X-kI)(X'X+kI)^{-1}X'X\beta - \beta$$
$$= (X'X+kI)^{-2}k[-3X'X-kI]\beta \tag{2.15}$$

The bias can be written in scalar form as:

$$Bias\left(\widehat{\beta}_{MKL}\right) = k\sum_{j=1}^{p}\frac{(-3\lambda_j-k)\beta}{(\lambda_j+k)^2} \tag{2.16}$$

$$V\left(\widehat{\beta}_{MKL}\right) = \sigma^2(X'X+kI)^{-1}(X'X-kI)(X'X+kI)^{-1}X'X(X'X+kI)^{-1}(X'X-kI)(X'X+kI)^{-1} \tag{2.17}$$

$V\left(\widehat{\beta}_{MKL}\right)$ can be represented in scalar form as follows:

$$V\left(\widehat{\beta}_{MKL}\right) = \sum_{j=1}^{p}\frac{\lambda_j(\lambda_j-k)^2}{(\lambda_j+k)^4} \tag{2.18}$$

Thus, the MSE is obtained as:

$$MSE\left(\widehat{\beta}_{MKL}\right) = \sigma^2\sum_{j=1}^{p}\frac{\lambda_j(\lambda_j-k)^2}{(\lambda_j+k)^4} + k^2\sum_{j=1}^{p}\frac{(3\lambda_j+k)^2\beta^2}{(\lambda_j+k)^4} \tag{2.19}$$

The proposed estimator in (2.14) is extended to the PRM. It is referred to as the Poisson modified KL (PMKL) estimator and defined as:

$$\widehat{\beta}_{PMKL} = \left(X'\widehat{L}X+k\right)^{-1}\left(X'\widehat{L}X-k\right)\left(X'\widehat{L}X+k\right)^{-1}X'\widehat{L}X\widehat{\beta}_{MLE} \tag{2.20}$$

The mean squared error of the PMKL is defined as:

$$MSE\left(\widehat{\beta}_{PMKL}\right) = \sum_{j=1}^{p}\frac{\lambda_j(\lambda_j-k)^2}{(\lambda_j+k)^4} + k^2\sum_{j=1}^{p}\frac{(3\lambda_j+k)^2\alpha_j^2}{(\lambda_j+k)^4} \tag{2.21}$$

where $k = min\left[\frac{\sqrt{(3\lambda_i\alpha^2+\sigma^2)^2 4\alpha^2\sigma^2\lambda_i-3\lambda_i\alpha^2+\sigma^2}}{2\beta^2} - (3\lambda_i\alpha^2+\sigma^2)\right]$ and $k > 0$.

Suppose $\alpha = Q'\beta$ and $Q'X'\widehat{L}XQ = \Lambda = diag(\lambda_1,\lambda_2,..,\lambda_p)$. Where $\lambda_1 \geq \lambda_2,...,\geq \lambda_p$, Λ is the matrix of eigen-values of $X'\widehat{L}X$ and Q is the matrix whose columns are the eigenvectors of $X'\widehat{L}X$.

The mean squared error (MSEM) and the following lemmas are adopted for theoretical comparisons among the estimators.

Lemma 2.1 Let A be a positive definite (pd) matrix, that is, $A > 0$, and a be some vector, then $A - aa' \geq 0$ if and only if (iff) $a'A^{-1}a \leq 1$ (Farebrother, 1976).

Lemma 2.2 $MSEM\widehat{\beta}_1 - MSEM\widehat{\beta}_2 = \delta^2 D + b_1 b'_1 - b_2 b'_2 > 0$, if and only if $b'_2 [\sigma^2 D + b_1 b'_1]^{-1} b_2 < 1$ where $MSE(\widehat{\beta}_j) = V(\widehat{\beta}_j) + b'_j b_j$, $b_1 = bias(\widehat{\beta}_1)$ and $b_2 = bias(\widehat{\beta}_2)$ (Trenker and Toutenburg, 1990).

Theorem 2.1: $\widehat{\alpha}_{PMKL}$ is preferred to $\widehat{\alpha}_{PMLE}$ iff, $MSEM(\widehat{\alpha}_{PMLE}) - MSEM(\widehat{\alpha}_{PMKL}) > 0$ provided $k > 0$.

Proof

$$V(\widehat{\alpha}_{PMLE}) - V(\widehat{\alpha}_{PMKL}) = Q diag\left\{\frac{1}{\lambda_j} - \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4}\right\}_{j=1}^{P} Q'$$

It is observed that $(\lambda_j + k)^4 - \lambda_j^2(\lambda_j - k)^2 > 0$ such that the expression above is non-negative for k > 0

Theorem 2.2: $\widehat{\alpha}_{PMKL}$ is preferred to $\widehat{\alpha}_{PRE}$ iff, $MSEM(\widehat{\alpha}_{PRE}) - MSEM(\widehat{\alpha}_{PMKL}) > 0$ provided $k > 0$.

Proof

$$V(\widehat{\alpha}_{PRE}) - V(\widehat{\alpha}_{PMKL}) = Q diag\left\{\frac{\lambda_j}{(\lambda_j + k)^2} - \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4}\right\}_{j=1}^{p} Q'$$

We can observe that the difference of the variance of the estimator is non-negative since $(\lambda_j + k)^2 - (\lambda_j^2 - k^2) > 0$ for $k > 0$.

Theorem 2.3: $\widehat{\alpha}_{PMKL}$ is preferred to $\widehat{\alpha}_{PLE}$ iff, $MSEM(\widehat{\alpha}_{PLE}) - MSEM(\widehat{\alpha}_{PMKL}) > 0$ provided $k > 0$ and $0 < d < 1$.

Proof

$$V(\widehat{\alpha}_{PLE}) - Cov(\widehat{\alpha}_{PMKL}) = Q diag\left\{\frac{(\lambda_j + d)^2}{\lambda_j(\lambda_j + 1)^2} - \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4}\right\}_{j=1}^{p} Q'$$

The difference of the variance is non-negative since

$(\lambda_j + k)(\lambda_j + d) - \lambda_j(\lambda_j + 1)(\lambda_j - k) > 0$ for $0 < d < 1$ and $k > 0$.

Theorem 2.4: $\widehat{\alpha}_{PMKL}$ is preferred to $\widehat{\alpha}_{PKL}$ iff, $MSEM(\widehat{\alpha}_{PKL}) - MSEM(\widehat{\alpha}_{PMKL}) > 0$ provided $k > 0$.

Proof

$$V(\widehat{\alpha}_{PKL}) - Cov(\widehat{\alpha}_{PMKL}) = Q diag\left\{\frac{(\lambda_j - k)^2}{\lambda_j(\lambda_j + k)^2} - \frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^4}\right\}_{j=1}^{p} Q'$$

The difference of the variance is non-negative since $(\lambda_j + k)(\lambda_j - k) - \lambda_j(\lambda_j - k) > 0$ for $k > 0$.

Selection of biasing parameter

The biasing parameter *k* for the estimator is obtained by differentiating the MSE in equation (2.21) with respect to *k* as follows:

$$\partial MSE(\widehat{\beta}_{MKL}) = -\left[2\sigma^2\sum_{j=1}^{p}\frac{\lambda_j(\lambda_j - k)}{(\lambda_j + k)^4} + 4\sigma^2\sum_{j=1}^{p}\frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^5}\right] + \left[2k\beta^2\sum_{j=1}^{p}\frac{(3\lambda_j + k)[(3\lambda_j + k) + K]}{(\lambda_j + k)^4} - 4k^2\sum_{j=1}^{p}\frac{(3\lambda_j + k)^2\beta^2}{(\lambda_j + k)^5}\right]$$

(2.22)

By equating to 0 and dividing through by 2 we have the resulting equation as:

$$-\sum_{j=1}^{p}\left[\sigma^2\frac{(\lambda_j(\lambda_j - k)}{(\lambda_j + k)^4} + 2\sigma^2\frac{\lambda_j(\lambda_j - k)^2}{(\lambda_j + k)^5}\right] + \sum_{j=1}^{p}\left[k\beta^2\frac{(3\lambda_j + k)(3\lambda_j + 2k)}{(\lambda_j + k)^4} - 2k^2\frac{(3\lambda_j + k)^2\beta^2}{(\lambda_j + k)^5}\right] = 0 \qquad (2.23)$$

$$-\sigma^2\lambda_j(\lambda_j - k)[\lambda_j + k + 2\lambda_j - 2k)] = k(3\lambda_j + k)\beta^2[6k\lambda_j + 2k^2 - 3\lambda_j^2 - 5k\lambda_j - 2k^2]$$
$$-\sigma^2\lambda_j(\lambda_j - k)[3\lambda_j - k)] = k(3\lambda_j + k)\beta^2[k\lambda_j - 3\lambda_j^2] \qquad (2.24)$$

Solving the equation above for k yields the biasing parameter k given below as:

$$k_{MKL} = min \left[\frac{\sqrt{(3\lambda_i\alpha^2 + \sigma^2)^2 + 4\alpha^2\sigma^2\lambda_i - 3\lambda_i\alpha^2 + \sigma^2}}{2\beta^2} - (3\lambda_i\alpha^2 + \sigma^2) \right] \quad (2.25)$$

The shrinkage parameter estimated by Mansson and Shukur, (2011) and Kibria and Lukman (2020) was also adopted for this study as listed:

$$k_1 = \frac{1}{\max\left(\alpha_j^2\right)} \quad (2.26)$$

$$k_2 = \frac{p}{\sum\left(2\alpha_j^2 + \frac{1}{\lambda_j e}\right)} \quad (2.27)$$

$$k_3 = k_{MKL} = k_3 = min \left[\frac{\sqrt{(3\lambda_j\alpha^2 + \sigma^2)^2 + 4\alpha^2\sigma^2\lambda_j - 3\lambda_j\alpha^2 + \alpha^2}}{2\beta^2} - (3\lambda_j\alpha^2 + \sigma^2) \right] \quad (2.28)$$

k_1 and k_2 is the biasing parameter for PMKL1 and PMKL2, while k_3 is the biasing parameters for PMKL3.

Simulation Design and Real-Life Application
Simulation study and result
In this section, a simulation study is carried out to compare the performance of the different estimators. The generation of the dependent variables are done using pseudo-random numbers from $P_o(\mu_i)$ where $\mu_i = e^{x_i\beta} i = 1, 2, ..., n$ and X_i is the ith row of the design matrix with $\beta = (\beta_0, \beta_1, ..., \beta_p)$ being the coefficient vector. The generation of the explanatory variables with different levels of correlation is obtained using

$$x_{ij} = \left(1 - \rho^2\right)^{\frac{1}{2}} z_{ij} + \rho z_{ip+1}; i = 1, 2, ..., n \text{ and } j = 1, 2, ..., p. \quad (3.1)$$

where ρ is the level of multicollinearity between the explanatory variables (Kibria et al. 2015; Kibria and Banik, 2016; Lukman *et al.*, 2019b, Lukman *et al.* 2020b). z_{ij} are pseudo-random numbers generated using the standard normal distribution such that i ranges from 1 to n and j from 1 to p. As a common restriction used in simulation studies, it is assumed that $\sum_{j=1}^{p}\beta_j^2 = 1$ and $\beta_1 = \beta_2 = ... = \beta_p$. Also, the effect of the intercept value is also being investigated as values are taken to be 1, 0 and -1 (Kibria *et al.* 2014). The different levels of correlation taken are 0.8, 0.9, 0.95, 0.99 and 0.999. The other factors varied in the simulation study are the sample size n and the number of explanatory variable p. We assume $n = 50$, 100 and 200 observations and $p = 4$ and 8 explanatory variables.

The simulation results in Tables 1 to 6 that for each of the estimators, the simulated MSE values increase as the multicollinearity level increases, keeping other factors constant. There is also an increase in the mean squared error as the sample size increases for all estimators compared while other factors were kept constant. As the intercept values varied from -1 to +1, the values of the mean squared error reduced for all estimators. Result shows that the PMKL1 performed best with minimum MSE at varying sample sizes. It was closely followed by PMKL2. They are both considered more suitable for estimation of parameters in the Poisson regression model than the MLE as it performed worst when multicollinearity is a challenge. In general, the PMKL1 estimator consistently performed more efficiently than the MLE, PRE, PLE and the PKL estimators.

Real Life Application
Having carried out a simulation study, the efficacy of the proposed estimator needs to be further investigated by considering a real-life application. The Poisson regression model has been applied to the aircraft damage dataset initially by Myers *et al.* (2012) and subsequently by other researchers such as Asar and Genc (2017) and Amin *et al.* (2020) among others. By following the Pearson chi-square goodness of fit test, Amin *et al.* (2020) was able to ascertain that the data fits a Poisson regression model. The test confirms the suitability of the response variable to Poisson distribution with P-value of 6.898122 (0.07521). The dataset provides some detail on two separate aircrafts: The McDonnell Douglas A-4 Skyhawk and the A-6 Grumman Itruder. The dependent variable denotes the number of locations with damage on the aircraft and this follows a Poisson distribution (Asar and Genc, 2017; Amin *et al.*, 2020). The data set has three explanatory variables,

Table 1. Simulation result for mean squared error (MSE) when P = 4 and intercept = 1.

β_0	N	ρ	MLE	PRE	PLE	PKL	PMKL1	PMKL2	PMKL3
1	50	0.8	0.0389	0.0376	0.0384	0.0383	0.0366	0.0366	0.0383
		0.9	0.0534	0.0494	0.0520	0.0515	0.0440	0.0446	0.0515
		0.95	0.0852	0.0729	0.0808	0.0791	0.0553	0.0574	0.0806
		0.99	0.3548	0.2013	0.2800	0.2435	0.0696	0.0750	0.2835
		0.999	3.4244	0.8757	1.8191	0.2302	0.1187	0.1041	1.4952
	100	0.8	0.0107	0.0108	0.0107	0.0107	0.0113	0.0111	0.0107
		0.9	0.0125	0.0124	0.0125	0.0125	0.0125	0.0125	0.0125
		0.95	0.0182	0.0177	0.0181	0.0181	0.0171	0.0172	0.0181
		0.99	0.0691	0.0595	0.0670	0.0665	0.0451	0.0480	0.0682
		0.999	0.6465	0.2995	0.5061	0.4404	0.0852	0.0979	0.6098
	200	0.8	0.0057	0.0056	0.0056	0.0056	0.0060	0.0059	0.0056
		0.9	0.0068	0.0068	0.0068	0.0068	0.0067	0.0068	0.0067
		0.95	0.0105	0.0104	0.0105	0.0105	0.0103	0.0104	0.0105
		0.99	0.0422	0.0394	0.0416	0.0415	0.0347	0.0357	0.0419
		0.999	0.5234	0.1897	0.4355	0.2322	0.0419	0.0324	0.5211

MLE = Maximum Likelihood Estimator; PRE=Poisson Ridge Estimator; PLE = Poisson Liu Estimator; PKL = Poisson Kibria-Lukman Estimator; PMKL1 = Poisson Modified Kibria Lukman Estimator 1; PMKL2 = Poisson Modified Kibria Lukman Estimator 2; PMKL3 = Poisson Modified Kibria Lukman Estimator 3.

Table 2. Simulation result for mean squared error (MSE) when P = 4 and intercept = 0.

β_0	n	ρ	MLE	PRE	PLE	PKL	PMKL1	PMKL2	PMKL3
0	50	0.8	0.1091	0.0821	0.1003	0.1000	0.0474	0.0532	0.1036
		0.9	0.1479	0.0978	0.1303	0.1295	0.0505	0.0556	0.1386
		0.95	0.2356	0.1287	0.1910	0.1866	0.0536	0.0588	0.2158
		0.99	0.9393	0.2847	0.5410	0.3425	0.0565	0.0593	0.6883
		0.999	9.4757	1.9349	4.8030	2.2184	0.2562	0.1817	2.9303
	100	0.8	0.0295	0.0266	0.0291	0.0291	0.0230	0.0238	0.0294
		0.9	0.0340	0.0301	0.0335	0.0335	0.0243	0.0258	0.0339
		0.95	0.0500	0.0416	0.0488	0.0488	0.0297	0.0325	0.0497
		0.99	0.1896	0.1088	0.1712	0.1706	0.0420	0.0503	0.1867
		0.999	3.5624	1.1897	1.5432	0.7168	0.0945	0.0991	1.5706
	200	0.8	0.0154	0.0153	0.0153	0.0153	0.0138	0.0139	0.0154
		0.9	0.0178	0.0187	0.0187	0.0187	0.0161	0.0166	0.0188
		0.95	0.0262	0.0284	0.0284	0.0284	0.0223	0.0234	0.0286
		0.99	0.8292	0.1083	0.1083	0.1082	0.0454	0.0529	0.1126
		0.999	1.5185	0.3222	0.8548	0.1183	0.0543	0.0743	0.9527

MLE = Maximum Likelihood Estimator; PRE = Poisson Ridge Estimator; PLE = Poisson Liu Estimator; PKL = Poisson Kibria-Lukman Estimator; PMKL1 = Poisson Modified Kibria Lukman Estimator 1; PMKL2 = Poisson Modified Kibria Lukman Estimator 2; PMKL3 = Poisson Modified Kibria Lukman Estimator 3.

Table 3. Simulation result for mean squared error (MSE) when P = 4 and intercept = -1.

β_0	n	ρ	MLE	PRE	PLE	PKL	PMKL1	PMKL2	PMKL3
-1	50	0.8	0.3089	0.2230	0.2569	0.2211	0.2093	0.2146	0.2478
		0.9	0.4295	0.2604	0.3327	0.2702	0.2299	0.2366	0.3205
		0.95	0.6924	0.3372	0.4890	0.3364	0.2562	0.2755	0.4573
		0.99	2.7802	0.7723	1.5114	0.3302	0.3020	0.3734	1.0038
		0.999	26.9726	5.4928	14.1059	9.2016	0.7486	1.9380	4.1156
	100	0.8	0.0809	0.0772	0.0775	0.0764	0.0533	0.0739	0.0778
		0.9	0.0935	0.0834	0.0891	0.0886	0.1043	0.1004	0.0812
		0.95	0.1389	0.1116	0.1290	0.1287	0.1175	0.1099	0.1296
		0.99	0.5161	0.2666	0.4024	0.3878	0.1281	0.1231	0.4790
		0.999	4.6805	1.1257	2.6775	0.7517	0.1994	0.2043	3.5989
	200	0.8	0.0421	0.0426	0.0417	0.0409	0.0404	0.0423	0.0409
		0.9	0.0511	0.0498	0.0501	0.0498	0.0446	0.0535	0.0499
		0.95	0.0767	0.0704	0.0741	0.0741	0.0672	0.0633	0.0741
		0.99	0.3107	0.2115	0.2766	0.2731	0.1226	0.1309	0.3016
		0.999	4.1247	0.8450	2.0275	0.7102	0.1369	0.1528	2.6541

MLE = Maximum Likelihood Estimator; PRE = Poisson Ridge Estimator; PLE = Poisson Liu Estimator; PKL = Poisson Kibria-Lukman Estimator; PMKL1 = Poisson Modified Kibria Lukman Estimator 1; PMKL2 = Poisson Modified Kibria Lukman Estimator 2; PMKL3 = Poisson Modified Kibria Lukman Estimator 3.

Table 4. Simulation result for mean squared error (MSE) when P = 8 and intercept = 1.

β_0	n	ρ	MLE	PRE	PLE	PKL	PMKL1	PMKL2	PMKL3
1	50	0.8	0.0980	0.0883	0.0960	0.0951	0.0796	0.0804	0.0969
		0.9	0.1404	0.1163	0.1355	0.1329	0.0906	0.0938	0.1369
		0.95	0.2256	0.1631	0.2113	0.2036	0.1004	0.1072	0.2201
		0.99	0.9255	0.3956	0.7190	0.5837	0.1133	0.1256	0.8598
		0.999	8.4713	1.9414	4.8816	1.0453	0.3070	0.2788	5.8787
	100	0.8	0.0232	0.0229	0.0231	0.0231	0.0227	0.0227	0.0231
		0.9	0.0340	0.0329	0.0337	0.0335	0.0314	0.0316	0.0336
		0.95	0.0534	0.0499	0.0525	0.0518	0.0440	0.0449	0.0526
		0.99	0.2226	0.1634	0.2026	0.1891	0.0853	0.0792	0.2081
		0.999	2.1185	0.7284	1.2950	0.4499	0.0882	0.0951	1.2901
	200	0.8	0.0057	0.0057	0.0057	0.0057	0.0056	0.0057	0.0057
		0.9	0.0076	0.0076	0.0076	0.0076	0.0075	0.0074	0.0076
		0.95	0.0117	0.0115	0.0116	0.0116	0.01130	0.01134	0.0116
		0.99	0.0443	0.0412	0.0434	0.0430	0.0356	0.0365	0.0436
		0.999	1.8722	0.2071	0.5491	0.2671	0.0566	0.0562	0.8730

MLE = Maximum Likelihood Estimator; PRE = Poisson Ridge Estimator; PLE=Poisson Liu Estimator; PKL = Poisson Kibria-Lukman Estimator; PMKL1 = Poisson Modified Kibria Lukman Estimator 1; PMKL2 = Poisson Modified Kibria Lukman Estimator 2; PMKL3 = Poisson Modified Kibria Lukman Estimator 3.

Table 5. Simulation result for mean squared error (MSE) when P = 8 and intercept = 0.

β_0	n	ρ	MLE	PRE	PLE	PKL	PMKL1	PMKL2	PMKL3
0	50	0.8	0.2738	0.1473	0.2377	0.2352	0.0808	0.0879	0.2682
		0.9	0.3927	0.1834	0.3258	0.3201	0.0829	0.0899	0.3825
		0.95	0.6114	0.2382	0.4677	0.4448	0.0942	0.0952	0.5888
		0.99	2.4858	0.6218	1.4882	0.7865	0.1205	0.1428	2.1778
		0.999	23.3548	4.6860	13.2573	6.5291	0.7055	0.4636	12.7807
	100	0.8	0.0646	0.0554	0.0617	0.0617	0.0491	0.0491	0.0635
		0.9	0.0934	0.0750	0.0879	0.0878	0.0526	0.0578	0.0907
		0.95	0.1462	0.1041	0.1327	0.1324	0.0553	0.0593	0.1395
		0.99	0.6068	0.2589	0.4234	0.2577	0.0587	0.0643	0.5154
		0.999	5.6854	1.4178	2.9951	0.3878	0.1433	0.1070	2.2332
	200	0.8	0.0159	0.0151	0.0157	0.0157	0.0141	0.0143	0.0158
		0.9	0.0207	0.0196	0.0205	0.0205	0.0176	0.0181	0.0206
		0.95	0.0319	0.0290	0.0312	0.0312	0.0242	0.0254	0.0315
		0.99	0.1185	0.0857	0.1089	0.1087	0.0441	0.0506	0.1141
		0.999	1.7596	0.4612	1.0889	0.1810	0.0794	0.0921	1.0001

MLE = Maximum Likelihood Estimator; PRE = Poisson Ridge Estimator; PLE = Poisson Liu Estimator; PKL=Poisson Kibria-Lukman Estimator; PMKL1 = Poisson Modified Kibria Lukman Estimator 1; PMKL2 = Poisson Modified Kibria Lukman Estimator 2; PMKL3 = Poisson Modified Kibria Lukman Estimator 3.

Table 6. Simulation result for mean squared error (MSE) when P = 8 and intercept = -1.

β_0	n	ρ	MLE	PRE	PLE	PKL	PMKL1	PMKL2	PMKL3
-1	50	0.8	0.8248	0.259	0.6469	0.4512	0.2159	0.6700	0.7314
		0.9	1.1355	0.4945	0.8253	0.5699	0.4693	0.5526	0.9927
		0.95	1.7701	0.6264	1.1745	0.6921	0.5848	0.5614	1.4256
		0.99	7.1964	1.6865	4.2093	1.0061	0.4735	0.4493	4.8258
		0.999	65.8760	12.8251	38.0726	37.7305	2.0409	0.9017	24.8129
	100	0.8	0.1800	0.1542	0.1653	0.1493	0.1422	0.1616	0.1547
		0.9	0.2575	0.2001	0.2269	0.2073	0.2020	0.1858	0.2131
		0.95	0.4105	0.2729	0.3392	0.3019	0.2354	0.2115	0.3419
		0.99	1.6914	0.6536	1.0983	0.5039	0.2420	0.2180	1.1012
		0.999	15.5667	3.7188	8.3620	4.0565	0.3710	0.2221	3.8870
	200	0.8	0.0436	0.0422	0.0425	0.0419	0.0303	0.0475	0.0419
		0.9	0.0568	0.0535	0.0546	0.0544	0.0512	0.0545	0.0542
		0.95	0.0860	0.0766	0.0810	0.0808	0.0693	0.0691	0.0813
		0.99	0.3260	0.2168	0.2753	0.2562	0.1022	0.1087	0.2890
		0.999	8.7594	1.5639	3.3266	2.0065	0.2433	0.1621	1.7855

MLE = Maximum Likelihood Estimator; PRE = Poisson Ridge Estimator; PLE = Poisson Liu Estimator; PKL = Poisson Kibria-Lukman Estimator; PMKL1 = Poisson Modified Kibria Lukman Estimator for k1; PMKL2 = Poisson Modified Kibria Lukman Estimator for k_2; PMKL3 = Poisson Modified Kibria Lukman Estimator for k_3.

Table 7. Regression coefficients and MSE.

coef.	MLE	PRE	PLE	PKL	PMKL1 (k1)	PMKL2 (k2)	PMKL3 (k3)
$\hat{\alpha}_0$	-0.406	-0.167	-0.255	-0.107	-0.019	-0.002	-0.077
$\hat{\alpha}_1$	0.569	0.380	0.479	0.391	0.120	0.179	0.322
$\hat{\alpha}_2$	0.165	0.171	0.167	0.168	0.183	0.179	0.172
$\hat{\alpha}_3$	-0.014	-0.015	-0.015	-0.016	-0.017	-0.017	-0.016
MSE	1.029	0.273	0.432	0.225	0.083	0.095	0.092
k		2.5444	0.1120	1.9409	2.5444	1.9409	0.9905

MLE = Maximum Likelihood Estimator; PRE = Poisson Ridge Estimator; PLE = Poisson Liu Estimator; PKL = Poisson Kibria-Lukman Estimator; PMKL1 = Poisson Modified Kibria Lukman Estimator 1; PMKL2 = Poisson Modified Kibria Lukman Estimator 2; PMKL3 = Poisson Modified Kibria Lukman Estimator 3.

X_1 shows the type of aircraft which makes the outcome binary (A-4 is coded as 0 and A-6 is coded as 1). X2 is the bomb load in tons and X3 is the number of months of aircrew experience. Meyers *et al*. (2012) was able to ascertain that the data set is greatly affected by multicollinearity. The eigenvalues of the matrix X were obtained as 4.3333, 374.8961 and 2085.2251. The condition number of 219.3654 was also obtained which is an indication of the problem of multicollinearity since it is greater than 30 (Asar and Genc, 2017). The performance of the estimators is judged based on the mean squared error of each of the estimators.

From Table 7, it is evident that all of the regression coefficients had identical signs. The estimator with the highest mean squared error is the MLE due to the presence of multicollinearity. The suggested estimator (PMKL1, PMKL2, PMKL3) has the lowest MSE that has established its dominance. We also observed that the performance of the estimator is highly dependent on the biasing parameter k. The expressions for the biasing parameters are defined in equation (2.26)-(2.28).

Conclusion
The parameters in the PRM are commonly estimated using the Maximum Likelihood Estimator. However, literature had shown that the estimator suffers a setback when the explanatory variables are correlated. This problem led to the implementation of alternative estimators with single shrinkage parameters such as the Poisson Ridge Regression Estimator (PRE), Poisson Liu Estimator (PLE) and the Poisson KL Estimator (PKLE). The KL estimator was generally preferred to the ridge regression and Liu estimator in the linear regression model. According to Lukman *et al*. (2021), the Poisson KL estimator outperforms PRE and PLE. This study modified the KL estimator to propose a new estimator called the Poisson Modified KL estimator (PMKL). The new estimator falls in the same class with the ridge, Liu and KL estimators since they possessed a single shrinkage parameter. We investigated the performance of the estimators with a simulation study and a real-life application. From the results, we observed that the new estimator consistently performed well in the presence of multicollinearity with the lowest MSE. Finally, the new estimator is more suitable to combat multicollinearity in the PRM.

Data availability
All data underlying the results are available as part of the article and no additional source data are required.

References

Ahmad S, Aslam M: **Another proposal about the new two-parameter estimator for linear regression model with correlated regressors.** *Communications in Statistics - Simulation and Computation.* 2020.
Publisher Full Text

Amin M, Akram M, Amanullah M: **On the James-Stein estimator for the poisson regression model.** *Communications in Statistics - Simulation and Computation.* 2020; 1–13.

Amin M, Akram M, Majid A: **On the estimation of Bell regression model using ridge estimator.** *Communications in Statistics - Simulation and Computation.* 2021; 1–14.

Amin M, Qasim M, Yasin A, *et al*.: **Almost unbiased ridge estimator in the gamma regression model.** *Communications in Statistics: Simulation and Computation.* 2020.
Publisher Full Text

Aslam M, Ahmad S: **The modified Liu-ridge-type estimator: a new class of biased estimators to address multicollinearity.** *Communications in Statistics - Simulation and Computation.* 2020; **0**(0): 1–20.
Publisher Full Text

Asar Y, Genç A: **A New Two-Parameter Estimator for the Poisson Regression Model.** *Iranian J Science Technology, Transaction A: Science.* 2017; **42**(2): 793–803.
Publisher Full Text

Dawoud I, Kibria BMG: *A new biased estimator to combat multicollineatity in the Gaussian linear regression model.*
PublisherFull Text

Dorugade AV: **Modified two parameter estimator in linear regression.** *J Stat Trans New Ser.* 2014; **15**(1): 23–36.
Publisher Full Text

Farebrother RW: **Further results on the mean square error of ridge regression.** *J. Roy. Statist. Soc., B.* 1976; **38**: 248–250.
Publisher Full Text

Hoerl AE, Kennard RW: **Ridge Regression: Applications to Nonorthogonal Problems.** *Technometrics.* 1970; **12**(1): 69–82.

Kaçiranlar S, Sakallioğlu S: **Combining the liu estimator and the principal component regression estimator.** *Communications in Statistics - Theory and Methods.* 2007; **30**(12): 2699–2705.
Publisher Full Text

Kibria BMG (2014): **Performance of some new ridge regression estimators.** *Commun Stat Simul Comput.* 2003; **32**(2): 419–435.
Publisher Full Text

Kibria BMG, Banik S: **Some ridge regression estimators and their performances.** *J Modern Applied Statistical Methods.* 2016; **15**(1): 206–238.

Kibria BMG, Lukman AF: **A new ridge-type estimator for the linear regression model: Simulations and applications.** *Scientifica.* 2020; **2020**.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Kibria BMG, *et al.*: **A simulation study of some biasing parameters for the ridge type estimation of Poisson regression.** *Commun. Stat.-Simul. Comput. I.* 2015; **44**: 943–957.
Publisher Full Text

Liu K: **A new class of biased estimate in linear regression.** *Communications in Statistics - Theory and Methods.* 1993; **22**(2): 393–402.
Publisher Full Text

Liu K: **Using Liu-Type estimator to combat collinearity.** *Communications in Statistics Theory and Methods.* 2003; **32**(5): 1009–1020.
Publisher Full Text

Lukman AF, Ayinde K, Binuomote S, *et al.*: **Modified ridge-type estimator to combat multicollinearity: Application to chemical data.** *J Chemometrics.* 2019a; **33**(5): 1–12.
Publisher Full Text

Lukman AF, Ayinde K, Sek SK, *et al.*: **A modified new two-parameter estimator in a linear regression model.** *Modelling and Simulation in Engineering.* 2019b.
Publisher Full Text

Lukman AF, Ayinde K, Aladeitan BB, *et al.*: **An Unbiased Estimator with Prior information.** *Arab Journal of Basic and Applied Sciences.* 2020a; **27**(1): 45–55.
Publisher Full Text

Lukman AF, Ayinde K, Kibria BMG, *et al.*: **Modified ridge-type estimator for the gamma regression model.** *Communications in Statistics: Simulation and Computation.* 2020b; **0**(0): 1–15.
Publisher Full Text

Lukman AF, Adewuyi MK, Kibria BMG: **A new estimator for the multicollinear Poisson regression model: simulation and application.** *Sci Rep.* 2021; **11**: 3732.
Publisher Full Text

Majid A, Amin M, Akram M: **On the Liu estimation of Bell regression model in the presence of multicollinearity.** *J. Stat. Comput. Simul.* 2021; 1–21.

Månsson K, Shukur G: **A Poisson ridge regression estimator.** *Econ. Model.* 2011; **28**: 1475–1481.
Publisher Full Text

Mansson K, Kibria BMG, Sjolander P, *et al.*: **Improved Liu Estimators for the Poisson Regression Model.** *Int J Statistics Probability.* 2012; **1**(1): 2–6.
Publisher Full Text

Myers RH, *et al.*: *Generalized linear models: with applications in engineering and the sciences.* New York: Wiley; 2012; **791**.

Ozkale MR, Kaciranlar S: **The restricted and unrestricted two-parameter estimators.** *Commun. Statist. Theor. Meth.* 2007; **36**: 2707–2725.
Publisher Full Text

Qasim M, Kibria BMG, Månsson K, *et al.*: **A new Poisson Liu Regression Estimator: method and application.** *J Applied Statistics.* 2019; 4763.
Publisher Full Text

Rashad N, Algamal Z: **A New Ridge Estimator for the Poisson Regression Model.** *Iran. J. Sci. Technol. Trans. A: Sci.* 2019; **43**(6): 2921–2928.

Sakallıoğlu S, Kaçıranlar S: **A new biased estimator based on ridge estimation.** *Statist. Papers.* 2008; **49**(4): 669–689.
Publisher Full Text

Sami F, Amin M, Butt M: **On the ridge estimation of the Conway-Maxwell Poisson regression model with multicollinearity: Methods and applications.** *Concurr. Comput. Pract. Exp.* 2021.

Trenkler G, Toutenburg H: **Mean squared error matrix comparisons between biased estimators—an overview of recent results.** *Stat Pap.* 1990; **31**(1): 165–179.
Publisher Full Text

Türkan S, Özel G: **A new modified Jackknifed estimator for the Poisson regression model.** *J Applied Statistics.* 2016; **43**: 1892–1905.
Publisher Full Text

Üstündağ Şiray G, Toker S, Özbay N: **Defining a two-parameter estimator: a mathematical programming evidence.** *J. Stat. Comput. Simul.* 2021; **91**(11): 2133–2152.

Yang H, Chang X: **A New Two-Parameter Estimator in Linear Regression.** *Communications in Statistics - Theory and Methods.* 2010; **39**(6): 923–934.
Publisher Full Text

Open Peer Review

Current Peer Review Status: ✔ ✔ ?

Version 2

Reviewer Report 21 December 2021

https://doi.org/10.5256/f1000research.78664.r113239

✔ **Muhammad Amin** (iD)

Department of Statistics, University of Sargodha, Sargodha, Pakistan

This paper can be accepted for indexing after incorporation of the following minor points:
1. Check equations 2.1 and 3.1.

2. Change "PMKL" to "PMKLE" in the whole manuscript.

3. In the application section, change "eigenvalues of the matrix X were" to "eigenvalues of the matrix X'X were".

4. In the application section, report the estimated values of the biasing parameters of the considered biased estimators with proper equation citations.

5. Change " Sami et al., 2021" to "Sami et al., 2022".

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Regression Analysis, Biased Estimation Methods

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 15 December 2021

https://doi.org/10.5256/f1000research.78664.r113241

✔ **Mohammad Arashi** iD

Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

The authors covered all the raised comments and I am happy with the current version.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: High-dimensional modeling; shrinkage estimation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 09 August 2021

https://doi.org/10.5256/f1000research.57427.r89263

? **Nimet Özbay** iD

Department of Statistics, Faculty of Science and Letters, Çukurova University, Adana, Turkey

This article focuses on proposing a modified KL estimator to mitigate the Poisson Regression Model with multicollinearity. Some theoretical properties of the new estimator are examined. A numerical example is conducted to show the performance of the new estimator. I think the authors should give the definition of the modified KL estimator in more detail and explain its statistical necessity. The organization of the paper, grammatical mistakes, and punctuation errors should also be controlled. This article may be indexed after the major comments below are applied.

Comments:

The main document of the article does not contain line numbers, thus it has been quite difficult to pinpoint the location of the comments.
1. There are lots of language and punctuation errors throughout the whole article, so the authors should recheck the writing of the manuscript. For example:

   ○ In the Abstract, "the simulation result showed…" should be changed to "the simulation result showed that…".

   ○ On page 3, a dot is required before equation (2.1).

- ○ On page 3, a comma is required before equation (2.2).

- ○ On page 5, line 1, "estimator" should be "estimators".

- ○ Section number is required for the sections, etc.

- ○ The basic punctuation marks are missing throughout the paper.

2. On page 3, the abbreviation "PRE" is repeated.

3. In the whole article, there are some inappropriate uses of the abbreviations. The authors should rearrange the use of abbreviations. In some places, previously made abbreviations are repeated.

4. The use of "hat" is missing while presenting some estimators.

5. In the Introduction section, the manuscript *Defining a two-parameter estimator: a mathematical programming evidence* by Üstündağ Şiray *et al.* (2021)[1] may be mentioned since this is a more recent article in which a new biased estimator is proposed to mitigate multicollinearity.

6. On page 4, the authors should explain what lambdas are.

7. On page 6, before equation (2.8), "means square error" should be "MSE".

8. There is no explanation for equation (2.11).

9. I think, "equation (2.10)" should be "equation (2.9)" before equation (2.12) on page 4.

10. On page 5, the authors should explain what lambdas are. Do the authors use "V" to Show variance? If so, some explanations should be added about it.

11. On page 5, there is the incorrect use of "MSEM". This abbreviation does not exist, although it is used while representing the lemmas and theorems.

12. I think the authors employ the canonical form in the proof of the theorems. Unfortunately, I did not find some information about the canonical model.

13. The selection of the biasing parameter section is insufficient. A detailed derivation and more information should be given.

14. In the simulation section, on page 7, why does the mean square error increase as the sample size increases?

15. It would be better if no abbreviations were used in the title.

References
1. Üstündağ Şiray G, Toker S, Özbay N: Defining a two-parameter estimator: a mathematical programming evidence. *Journal of Statistical Computation and Simulation*. 2021; **91** (11): 2133-2152
Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Monte Carlo simulation, Linear regression model, Econometric models, Applied statistics, Biased estimation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 03 August 2021

https://doi.org/10.5256/f1000research.57427.r89258

? **Muhammad Amin** (iD)

Department of Statistics, University of Sargodha, Sargodha, Pakistan

In this paper, the authors introduced a new estimator by modified KL estimator for the Poisson regression model to overcome the effect of multicollinearity. The paper is original and deals with a topic of interest. This paper could be accepted for indexing after incorporating the following

points.

- ○ Write one paragraph on count data models and their importance at the start of the Introduction and include some citations that demonstrate the importance of count data models, for example: Amin *et al*., 2020[1]; Amin *et al*., 2021[2]; Sami *et al.,* 2021[3]; Amin *et al*., 2021[4]; Majid *et al*., 2021[5]; Rashad *et al*., 2019[6]; Algamal *et al*., 2015[7]; Algamal *et al*., 2021[8]; Alanaz *et al*., 2018[9].

- ○ Write the reason for your proposed estimator over other estimators in the last paragraph of the Introduction section.

- ○ Change independent variables to explanatory variables in the whole study.

- ○ Write the first paragraph clearly and correct equation 2.1 of the Methods section by following Amin *et al*., 2020 [1].

- ○ Change "mean square error" to "mean squared error" in the whole manuscript.

- ○ On page 3, write the reason for adapting the Poisson ridge estimator.

- ○ On page 4, line 1, write the range of ridge parameter k.

- ○ Write the limitations of the ridge estimator after equation (2.5).

- ○ Write the range of Liu parameter d after equation (2.6).

- ○ Write different notations of the ridge parameter k, for ridge, KL, and MKL estimators and also mention the ranges of these biasing parameters.

- ○ Write the expressions for MSEs of ridge, Liu, and KL estimators.

- ○ In Lemma 2.2, define b1, b2,

- ○ The statement of Theorem 2.2 is wrong, I suggest the authors correct this.

- ○ Define e in equation (2.23).

- ○ In equation (2.24), change λi^{λ_i} to λj^{λ_j}

- ○ Correct expressions above equation (3.1).

- ○ The interpretations of simulation results need more detailed discussion.

- ○ In real application, report the estimated values of each biasing parameter with proper citation of equations. Moreover, cite equation to compute MSE of the consider estimators.

- ○ There are some grammatical issues that should be corrected.

References

1. Amin M, Akram M, Amanullah M: On the James-Stein estimator for the poisson regression model. *Communications in Statistics - Simulation and Computation*. 2020. 1-13 Publisher Full Text

2. Amin M, Akram M, Majid A: On the estimation of Bell regression model using ridge estimator. *Communications in Statistics - Simulation and Computation*. 2021. 1-14 Publisher Full Text

3. Sami F, Amin M, Butt M: On the ridge estimation of theConway-Maxwell Poisson regression model with multicollinearity: Methods and applications. *Concurrency and Computation: Practice and Experience*. 2021. Publisher Full Text

4. Amin M, Akram M, Kibria B: A new adjusted Liu estimator for the Poisson regression model. *Concurrency and Computation: Practice and Experience*. 2021. Publisher Full Text

5. Majid A, Amin M, Akram M: On the Liu estimation of Bell regression model in the presence of multicollinearity. *Journal of Statistical Computation and Simulation*. 2021. 1-21 Publisher Full Text

6. Rashad N, Algamal Z: A New Ridge Estimator for the Poisson Regression Model. *Iranian Journal of Science and Technology, Transactions A: Science*. 2019; **43** (6): 2921-2928 Publisher Full Text

7. Algamal Z, Lee M: Adjusted Adaptive LASSO in High-dimensional Poisson Regression Model. *Modern Applied Science*. 2015; **9** (4). Publisher Full Text

8. Algamal ZY: DIAGNOSTIC IN POISSON REGRESSION MODELS. *Electronic Journal of Applied Statistical Analysis*. 2012; **5** (2): 178-186

9. Alanaz MM, Algamal ZY: Proposed methods in estimating the ridge regression parameter in Poisson regression model. *Electronic Journal of Applied Statistical Analysis*. 2018; **11** (2): 505-515 Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Partly

Are all the source data underlying the results available to ensure full reproducibility?

Partly

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Regression Analysis, Biased Estimation Methods

I confirm that I have read this submission and believe that I have an appropriate level of

expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 21 July 2021

? **Mohammad Arashi** (iD)

Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

The paper extends the Liu estimator in generalized linear modeling. Specifically, the authors propose a new biased estimator for the estimation of regression coefficients in the discrete Poisson regression.

The results are interesting and the topic is eye-catching. The theoretical results are well supported by extensive numerical analysis.

I suggest the authors make minor revisions to improve the presentation before indexing.
1. Check the notation entirely to be consistent. For instance, in equation (2.10), "hat" must be added for the estimator. It happens also for (2.12).

2. Use another notation for diagonal matrices in equations. For example, you may use "L" and then define the elements.

3. Explain equation (2.21) more. Is the minimization over i?

4. In the simulation study, for the design generation, I suggest using another notation for "z_{ip}" since it is not the last element of the series of generated independent normals.

5. Provide a reference for the accessibility of the real data use in the real-life application.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?
Partly

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: High-dimensional modeling; shrinkage estimation

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 27 Jul 2021
BENEDICTA Aladeitan, Landmark University, Omu-Aran, Nigeria

Thanks for your observations and corrections. All will be duely implemented.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

• Your article is published within days, with no editorial bias

• You can publish traditional articles, null/negative results, case reports, data notes and more

• The peer review process is transparent and collaborative

• Your article is indexed in PubMed after passing peer review

• Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com