International Journal of Multidisciplinary Trends

E-ISSN: 2709-9369
P-ISSN: 2709-9350
www.multisubjectjournal.com
IJMT 2025; 7(3): 40-45
Received: 18-01-2025
Accepted: 22-01-2025

Fanny Winda Aini
Department of Mathematics, University of Lampung, Lampung, Indonesia

Netti Herawati
Department of Mathematics, University of Lampung, Lampung, Indonesia

Misgiyati
Department of Mathematics, University of Lampung, Lampung, Indonesia

Nusyirwan
Department of Mathematics, University of Lampung, Lampung, Indonesia

Corresponding Author:
Fanny Winda Aini
Department of Mathematics, University of Lampung, Lampung, Indonesia

Differentiation of various estimators in handling multicollinearity in poisson regression model: Case study of infant mortality rate in East Nusa Tenggara Province, Indonesia

Fanny Winda Aini, Netti Herawati, Misgiyati and Nusyirwan

DOI: https://www.doi.org/10.22271/multi.2025.v7.i3a.626

Abstract
Poisson regression is used to model Poisson-distributed count data as a dependent variable with one or more independent variables. Poisson regression uses the Maximum Likelihood Estimation method by considering the assumptions that must be met, including the absence of multicollinearity. The assumption of multicollinearity in Poisson regression can cause large variance in the data. In this study, Poisson regression with three different estimator methods will be compared, namely the Poisson James Stein Estimator, the Poisson Ridge Regression Estimator and the Poisson Kibria Lukman estimator, on infant mortality rates in East Nusa Tenggara Province, Indonesia, where there is a multicollinearity problem is based on the MSE value. The estimator with the best MSE will be declared as the best estimator. The results show the Poisson Ridge Regression Estimator method, with a parameter ridge k_2 of 0.00007, is the most effective in handling multicollinearity, because it produces the smallest Mean Squared Error.

Keywords: Infant mortality rate, Poisson regression, multicollinearity, Poisson James-stein estimator, Poisson ridge regression estimator, Poisson Kibria Lukman estimator.

Introduction
Poisson regression analyzes count variables assumed to follow a Poisson distribution, with the expectation depending on predictor variables through a logarithmic link function (Agresti, 2015) [1]. Common issue in regression analysis is multicollinearity, where strong linear relationships between independent variables make accurate coefficient estimation difficult (Wooldridge, 2012) [17]. The presence of multicollinearity can be assessed via correlation coefficients.

Methods to address multicollinearity include the Poisson James-Stein Estimator, which reduces coefficient variance through shrinkage, especially with many predictors according to Judge *et al*. (1985) [8]. Poisson Ridge Regression Estimator is combining Poisson regression with ridge regression, is effective in handling multicollinearity. Poisson Modified Kibria Lukman Estimator is an extension of the Kibria Lukman Estimator to address multicollinearity, particularly in Poisson regression models.

Previous research shows that Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Modified Kibria Lukman Estimator are effective in addressing multicollinearity in Poisson regression. Amin *et al*. (2020) [3] demonstrated that the Poisson James Stein Estimator outperforms classical MLE in terms of MSE. Oghenekevwe *et al*. (2021) [13] found that Poisson Ridge Regression Estimator is effective in handling multicollinearity, while Aladeitan *et al*. (2021) [2] showed that Poisson Modified Kibria Lukman Estimator is more efficient compared to other methods.

Indonesia has a high child mortality rate, with Nusa Tenggara Timur, Indonesia recording 16, 85 deaths per 1.000 live births. To identify influencing factors, Poisson regression analysis is used, addressing multicollinearity with methods like Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Modified Kibria Lukman Estimator. This study aims to determine the most effective method for reducing multicollinearity and minimizing MSE.

Material and Methods
Poisson James Stein Estimator: This method was first proposed by Charles Stein in 1956.

The origin and development of the Poisson James-Stein Estimator are based on the concept of "shrinkage" and the theory of minimizing the Mean Squared Error (MSE). This method applies centering and scaling transformations. In the Poisson James-Stein Estimator, shrinkage means that the estimate is slightly pulled toward a specific point, usually zero. Poisson James-Stein Estimator method is defined as:

$$\hat{\beta}_{PJSE} = c\hat{\beta}_{MLE}$$

Where,

(0 < c < 1) is the shrinkage factor used to reduce the MLE estimate, defined as:

$$c = \frac{(\hat{\beta}'_{MLE}\hat{\beta}_{MLE})}{(\hat{\beta}'_{MLE}\hat{\beta}_{MLE} + \text{trace}(S)^{-1})}$$

Where,

$S = X^T\widehat{W}X, \widehat{W} = diag[\mu_i]$ and $\hat{\beta}_{MLE}$ is the unbiased of β.

MSE value of the Poisson James Stein Estimator is as follows:

$$MSE(\hat{\beta}_{PJSE}) = \left(\frac{(\hat{\beta}'_{MLE}\hat{\beta}_{MLE})}{(\hat{\beta}'_{MLE}\hat{\beta}_{MLE} + trace(S)^{-1})}\right)'$$

$$(trace(S)^{-1})\left(\frac{(\hat{\beta}'_{MLE}\hat{\beta}_{MLE})}{(\hat{\beta}'_{MLE}\hat{\beta}_{MLE} + trace(S)^{-1})}\right) + \left(-\frac{trace(S)^{-1}}{\beta'\beta + trace(S)^{-1}}\beta\right)$$

Simplified, the scalar $MSE(\hat{\beta}_{PJSE})$ is formulated as:

$$MSE(\hat{\beta}_{PJSE}) = \sum_{j=1}^{r}\frac{\alpha_j^4\lambda_j}{(\alpha_j^2\lambda_j+1)^2} + \sum_{j=1}^{r}\frac{\alpha_j^2}{\alpha_j^2\lambda_j{}^2+1}$$

Poisson Ridge Regression Estimator

In the Poisson Ridge Regression Estimator (PRRE), the multicollinearity issue is addressed by adding a bias constant k to reduce the variance of the estimator compared to the variance of the Poisson regression estimator. This method applies centering and scaling transformations. It was introduced by Hoerl and Kennard in 1970 and later developed by Mansso and Shukur in 2011 by exploring the PRR method, which demonstrated effectiveness in handling multicollinearity. The Poisson Ridge Regression Estimator equation is as follows:

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

MSE value of the Poisson Ridge Regression Estimator is as follows:

$$MSE(\hat{\beta}_{PRRE}) = \sum_{j=1}^{p}\frac{\lambda_j}{(\lambda_j+k)^2} + k^2\sum_{j=1}^{p}\frac{\hat{\alpha}_j^2}{(\lambda_j+k)^2}$$

Ridge parameter k plays a role in reducing multicollinearity and preventing overfitting. This parameter is applied to balance bias and variance in the model. In a Poisson regression model with multicollinearity, the ridge parameter is defined as follows:

1. In the Poisson regression model, the optimal value of k used is as follows:

$$\hat{k}_1 = \frac{1}{\hat{\alpha}^2_{max}}$$

Where $\hat{\alpha}^2_{max}$ is the maximum value of $\delta^T\hat{\beta}_{ML}$ with δ^T being the eigenvector element of the matrix $X^T\widehat{W}X$.

2. Optimal k value used in the Poisson regression model is as follows:

$$\hat{k}_2 = \text{median}(q_i)$$

Where,

$q_i = \frac{\lambda_{max}}{(n-p)\hat{\sigma}^2+\lambda_{max}\hat{\alpha}_i^2}$.where λ_{max} is the maximum eigen value of $X^T\widehat{W}X$.

$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2}{(n-p-1)}$ where p is the number of independent variables.

Poisson Modified Kibria Lukman Estimator

Poisson Modified Kibria-Lukman Estimator is a further development of the Poisson Kibria-Lukman Estimator. This method aims to improve parameter estimation accuracy compared to the ridge regression method. It applies centering and scaling transformations. In the Poisson Modified Kibria-Lukman Estimator, the parameter k is used to determine the parameter values in order to balance bias and variance, making it more efficient in addressing multicollinearity issues. The equation for this method is as follows:

$$\hat{\beta}_{PMKL} = (X'\widehat{W}X + k)^{-1}(X'\widehat{W}X + k)$$

$$(X'\widehat{W}X + k)^{-1}X'\widehat{W}X\hat{\beta}_{MLE}$$

MSE value of the Poisson Modified Kibria-Lukman Estimator is as follows:

$$MSE(\hat{\beta}_{PMKLE}) = \sum_{j=1}^{p}\frac{\lambda_j(\lambda_j-k)^2}{(\lambda_j+k)^4} + k^2\sum_{j=1}^{p}\frac{(3\lambda_j+k)^2\alpha_j^2}{(\lambda_j+k)^4}$$

The shrinkage parameter estimated by Mansson and Shukur (2011) as well as Kibria and Lukman (2020) is also adopted for this study as stated below:

$$k_1 = \frac{1}{\max(\alpha_j^2)}$$

$$k_2 = \frac{p}{\Sigma(2\alpha_j^2 + \frac{1}{e})}$$

$$k_3 = min\frac{\lambda_i}{2\lambda_j\alpha_j^2+1}$$

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

To compare the best method among the Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Kibria Lukman Estimator, this study analyzes the number of infant mortality rate in East Nusa Tenggara Province, Indoensia in 2023 (Y) using five independent variables and

22 observations from the Central Bureau of Statistics (BPS) of East Nusa Tenggara Province, Indonesia (https://ntt.bps.go.id/). The independent variables include the number of immunized children (X_1), number of stunted children (X_2), the number of underweight children (X_3), the number of overweight children (X_4), and the number of severely malnourished children (X_5).

Analysis begins with data exploration, Poisson distribution testing using histograms and the Kolmogorov-Smirnov test, and the development of a Poisson regression model through parameter estimation with Maximum Likelihood Estimation (MLE) use:

$$\hat{\beta}_{ML} = [X^T W X]^{-1} X^T \widehat{W}_s$$

After that, significance testing is conducted using the likelihood ratio test., and deviance testing to assess model accuracy. Multicollinearity is analyzed using the Variance Inflation Factor (VIF), followed by centering and scaling transformations on independent variables. Poisson regression parameters are then estimated using the Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Modified Kibria Lukman Estimator by computing the matrix $X^T \widehat{W} X$, where $\widehat{W} = diag[\mu_i]$, determining the shrinkage factor or ridge parameter, and calculating the estimation values for each method along with the Mean Squared Error (MSE). The estimation results are compared based on the minimum MSE value to determine the best estimator and construct the optimal model equation.

Model Accuracy Measure

Mean Square Error (MSE) measures the average squared difference between the predicted and actual values. A lower MSE indicates a better model fit. The formula for MSE is:

$$MSE = \sum_{i=1}^{n} \frac{(\hat{Y}_t - Y_t)^2}{n}$$

Where Y_t is observed data value, $\hat{Y}_t$ is predicted value, and n is number of data.

Results and Discussion

To compare the Performance of Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Kibria Lukman Estimator in overcoming multicollinearity in Poisson Regression Model, a descriptive analysis was first conducted on infant mortality data in East Nusa Tenggara in 2023 in Kupang Regency, Indonesia to obtain a clear data picture. The results of the descriptive data analysis can be seen in Table 1.

Table 1: Descriptive Data Analysis

	Y	X_1	X_2	X_3	X_4	X_5
Mean	51.136	3412.818	2900.181	4099.727	132.636	175.545
Std. Deviasi	31.720	2169.190	2376.731	2381.422	78.869	153.515
Min	9.000	1378.000	549.000	1094.000	28.000	10.000
Q_1	26.750	2233.250	1253.500	2502.250	70.000	77.500
Median	43.500	2722.500	2302.000	3764.000	118.000	130.500
Q_3	68.000	3481.500	3440.250	4678.000	188.500	254.250
Max	127.000	9437.000	9762.000	11463.000	330.000	564.000

In Table 1, it can be seen that the average value of the number of infant mortality cases variable is 51.136, with a standard deviation of 31.720. The minimum value is found in Sumba Tengah Regency, while the highest value is in Kupang Regency. This indicates that the highest number of infant mortality cases in East Nusa Tenggara, Indonesia in 2023 was in Kupang Regency. After getting an overview of the data distribution, the next step is to see whether the data is distributed Poisson or not using a histogram. The histogram results can be seen in Figure 1.

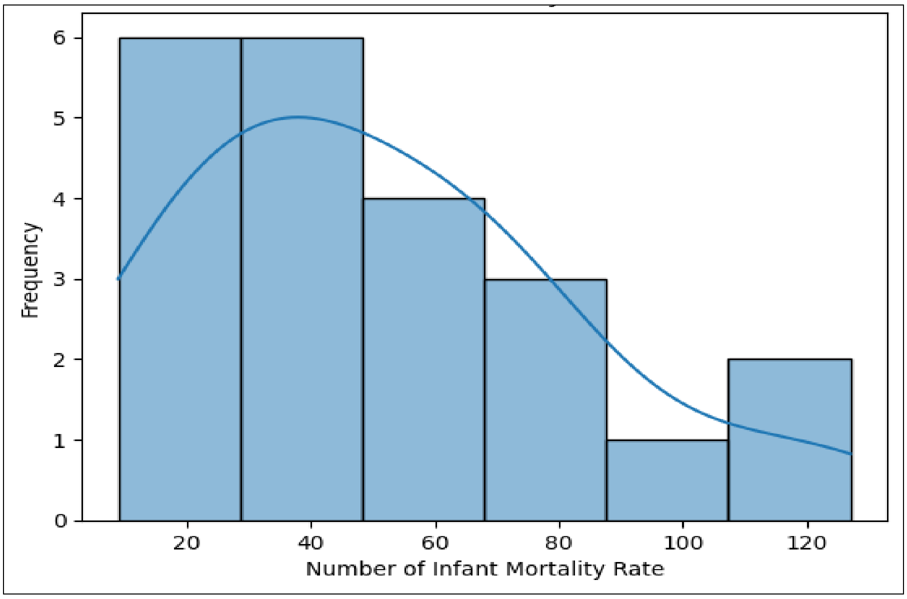

Fig 1: Histogram of infant mortality data

The data follows a Poisson distribution, with a right-skewed histogram in Figure 1 shows that the infant mortality curve where the mode of 23 is lower than the mean of 51.136. In addition, to further confirm the results, the Kolmogorov-Smirnov test was also performed on the data and the test results can be seen in Table 2 below.

Table 2: Kolmogorov-Smirnov Test

Kolmogorov-Smirnov	
D_{hit}	0.0951
P-Value	0.9774

The test results in Table 2 show that the significance value of $D_{hit} = 0.0951 < D_{(0.05,22)} = 0.290$ or the p-value $= 0.9774 > \alpha = 0.05$. This means that H_0 is failed to be rejected and it can be concluded that the data follows a Poisson distribution.

After getting the results that the data is distributed Poisson, the next step is to estimate the parameters using the maximum likelihood method for the Poisson regression model without considering whether the data contains multicollinearity or not. The results of parameter estimation can be seen in Table 3.

Table 3: MLE Estimator

Variable	Parameter	B
Constant	β_0	3.0072
X_1	β_1	-2.016×10^{-6}
X_2	β_2	-0.0001
X_3	β_3	0.0002
X_4	β_4	-0.0035
X_5	β_5	0.0008

It can be seen from Table 3 that the Poisson regression model is as follows:

$$\log(\mu) = 3.0072 - 2.016 \times 10^{-6} X_{i1} - 0.0001 X_{i2} + 0.0002 X_{i3} - 0.0035 X_{i4} + 0.0008 X_{i5}$$

Afterwards, a VIF test is conducted to test whether the data East Nusa Tenggara Province, Indonesia contains multicollinearity or not. The test results are presented in Table 4.

Table 4: Multicollinearity Test

Variable	Variance Inflation Factor (VIF)
X_1	3.8591
X_2	15.5611
X_3	17.3595
X_4	1.8226
X_5	3.3668

Table 4 shows that the VIF values for variables X_2 and X_3 are greater than 10, indicating multicollinearity in the infant mortality data from East Nusa Tenggara, Indonesia. Therefore, further analysis using the Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Modified Kibria Lukman Estimator is required to address this issue in the Poisson regression model.

Next, to address the issue of multicollinearity in the data, the analysis will be conducted using the *Poisson James-Stein Estimator* method.

Table 5: Shrinkage factor and parameter of poisson james-stein estimator

Factor Shrinkage	0.2319
Variable	Estimate
X_1	-0.0046
X_2	-0.2732
X_3	0.5280
X_4	0.2992
X_5	-0.1462

Based on Table 5, which shows the James Stein parameter estimates for Poisson regression, the resulting Poisson regression model is as follows:

$$\log(\mu) = -0.0046 X_{i1}^{*} - 0.2732 X_{i2}^{*} + 0.5280 X_{i3}^{*} + 0.2992 X_{i4}^{*} - 0.1462 X_{i5}^{*}$$

Table 6: Ridge Parameter

Parameter (k_i)	Parameter Estimation
k_1	0.27056
k_2	0.00007

Next, the analysis will be conducted using the Poisson Ridge Regression Estimator method by applying the ridge *k* parameter to the table 6.

Table 7: Estimation of the Poisson Ridge Regression Estimator

Variable	Estimate ridge k_1	Estimate ridge k_2
X_1	0.42573	-0.01913
X_2	0.11698	-1.17579
X_3	0.68356	2.27426
X_4	0.85390	1.28986
X_5	-0.35235	-0.63090

Based on the Table 7, which shows the estimates for all the ridge *k* parameters in Poisson regression, the resulting Poisson regression model is as follows:

$$\log(\mu) = 0.42573 X_{i1}^{*} + 0.11698 X_{i2}^{*} + 0.68356 X_{i3}^{*} + 0.85390 X_{i4}^{*} - 0.35235 X_{i5}^{*}$$

$$\log(\mu) = -0.01913 X_{i1}^{*} - 1.17579 X_{i2}^{*} + 2.27426 X_{i3}^{*} + 1.28986 X_{i4}^{*} - 0.63090 X_{i5}^{*}$$

The next step will be analyzed using the Poisson Modified Kibria Lukman Estimator method, with the use of parameter *k*, and the parameter estimates will be determined.

Table 8: Parameter *k* in PMKLE

Parameter (k_i)	Parameter Estimation
k_1	0.2705
k_2	0.0526
k_3	0.0399

Table 9: Estimation of the poisson modified Kibria Lukman Estimator for all k parameters

Variable	Estimate k_1	Estimate k_2	Estimate k_3
X_1	-0.24244	-0.07930	-0.06601
X_2	-1.54033	-1.27479	-1.25316
X_3	2.65850	2.37862	2.35582
X_4	0.93338	1.19524	1.21657
X_5	-1.04195	-0.74032	-0.71575

Based on the Table 9, which shows the estimates for all k parameters of the MKLE in Poisson regression, the resulting Poisson regression model is as follows:

$$\log(\mu) = -0.24244X_{i1}^* - 1.54033X_{i2}^* + 2.65850X_{i3}^* + 0.93338X_{i4}^* - 1.04195X_{i5}^*$$

$$\log(\mu) = -0.07930X_{i1}^* - 1.27479X_{i2}^* + 2.37862X_{i3}^* + 1.19524X_{i4}^* - 0.74032X_{i5}^*$$

$$\log(\mu) = -0.06601X_{i1}^* - 1.25316X_{i2}^* + 2.35582X_{i3}^* + 1.21657X_{i4}^* - 0.71575X_{i5}^*$$

Next, the Mean Square Error (MSE) values for the Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Modified Kibria Lukman Estimator methods will be calculated. Below are the performance results of the three methods based on the Mean Square Error (MSE) values:

Table 10: Mean square error for all estimators

Method	MSE
MLE	24.9508
PJSE (c)	0.04631
PRRE (k_1)	0.21288
PRRE (k_2)	6.66747×10^{-7}
PMKLE (k_1)	0.00040
PMKLE (k_2)	0.00024
PMKLE (k_3)	0.21288

It can be seen from the output in Table 10 that the smallest MSE value is found in the Poisson Ridge Regression Estimator (PRRE) method with parameter ridge k_2, which is 0.00007. Thus, it can be concluded that using the Poisson Ridge Regression Estimator (PRRE) is more effective in handling the multicollinearity issue in the case study of infant mortality in East Nusa Tenggara, Indonesia in 2023. The regression model obtained using the Poisson Ridge Regression Estimator (PRRE) method is as follows:

$$\log(\mu) = -0.01913X_{i1}^* - 1.17579X_{i2}^* + 2.27426X_{i3}^* + 1.28986X_{i4}^* - 0.63090X_{i5}^*$$

Where X^* is the transformed data of the independent variables.

The data can be transformed into the original PRRE form by multiplying the PRRE parameters with matrix G. The resulting model is as follows:

$$\log(\mu) = -0.05679 + 0.00003X_{i1} - 0.00002X_{i2} + 6.276 \times 10^{-6}X_{i3} - 0.00115X_{i4} + 0.00034X_{i5}$$

Conclusions

The analysis indicates that the observed data follows a Poisson distribution, and the independent variables significantly impact infant mortality in NTT in 2023. The Kolmogorov-Smirnov test and other tests support this finding. Among the Poisson James Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Modified Kibria Lukman Estimator methods, the Poisson Ridge Regression Estimator is the best for handling multicollinearity, with a minimum MSE of 6.66747×10^{-7} and a parameter ridge k_2 of 0.00007. The obtained regression model is:

$$\log(\mu) = -0.01913X_{i1}^* - 1.17579X_{i2}^* + 2.27426X_{i3}^* + 1.28986X_{i4}^* - 0.63090X_{i5}^*$$

This is transformed into the original MKL form:

$$\log(\mu) = -0.05679 + 0.00003X_{i1} - 0.00002X_{i2} + 6.276 \times 10^{-6}X_{i3} - 0.00115X_{i4} + 0.00034X_{i5}$$

References

1. Agresti A. Foundations of Linear and Generalized Linear Models. 1st ed. John Wiley & Sons, Canada, 2015.
2. Aladeitan BB, Adebimpe O, Lukman AF, Oludoun O, Abiodun OE. Modified Kibria-Lukman estimator for the Poisson regression model. F1000 Research. 2021;10(548):1-19.
3. Amin M, Akram MN, Amanullah M. On the James-Stein estimator for the Poisson regression model. Communications in Statistics-Simulation and Computation. 2020;51(10):5596-5608.
4. Statistika BPNTT. Mortalitas di Indonesia: Hasil Long Form Sensus Penduduk 2020. Badan Pusat Statistika, Nusa Tenggara Timur; 2023.
5. Gujarati DN, Porter DC. Basic Econometrics. 5th ed. McGraw-Hill Irwan, USA, 2009.
6. Hamdaoui A, Terbeche M, Benkhaled A. On shrinkage estimator under balanced loss function. Pakistan Journal of Statistics and Operation Research. 2021;17(10):711-727.
7. Herawati N, Sukma H, Nisa K, Nusyirwan N, Saidi S, Tiryono T, *et al.* Poisson ridge regression for multicollinearity data: Case study of the number of maternal deaths in Lampung Province, Indonesia. International Journal of Innovative Research in Science, Engineering and Technology. 2024;13(1):1-8.
8. Judge GG, Griffiths WE, Hill RC, Lutkepohl H, Lee TC. The Theory and Practice of Econometrics. 2nd ed. John Wiley & Sons, USA, 1985.
9. Khoshsirat S, Kambhamettu C. Improving normalization with the James-Stein estimator. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2023;42(12):2980-2992.
10. Kennedy P. A Guide to Econometrics. 6th Ed. Blackwell, UK, 2008.
11. Kutner MH, Natchsheim CJ, Neter J, Li W. Applied Linear Statistical Models. 5th ed. McGraw-Hill Irwan, USA, 2004.
12. Montgomery DC, Peck EA, Vining GG. Introduction to Linear Regression Analysis. 5th ed. John Wiley & Sons, Canada, 2012.
13. Oghenekevwe EH, Florence AK, Abidemi AK. Poisson Poisson ridge regression estimators: A performance test.

American Journal of Theoretical and Applied Statistics. 2021;10(2):111-121.
14. Park K. Preventive and Social Medicine. 27th Ed. M/s B. I. Churchill Livingstone, India, 2023.
15. Sami F, Amin M, Aljeddani SMA. Stein estimation in the Conway-Maxwell Poisson model with correlated regressors. International Journal of Advanced and Applied Sciences. 2024;11(7):49-56.
16. Wooldridge JM. Econometric Analysis of Cross Section and Panel Data. 2nd Ed. MIT Press, Cambridge, 2010.
17. Wooldridge JM. Introductory Econometrics: A Modern Approach. 5th Ed. Cengage Learning, USA, 2012.
18. Yehia EG. On the restricted Poisson Poisson ridge regression estimator. Science Journal of Applied Mathematics and Statistics. 2021;9(4):106-112.