International Journal of Applied Science and Engineering Review

ISSN: 2582-6271

Vol. 6, Issue.2, Mar-Apr 2025, page no. 30-40

To cite this article: M. Fikri Alyasa Zam Zami, Netti Herawati, Misgiyati, and Nusyirwan (2025). SIMULATION OF COMPARATIVE STUDY OF JAMES-STEIN ESTIMATOR, RIDGE REGRESSION ESTIMATOR, AND MODIFIED KIBRIA LUKMAN ESTIMATOR IN HANDLING MULTICOLLINEARITY IN POISSON REGRESSION, International Journal of Applied Science and Engineering Review (IJASER) 6 (2): 30-40 Article No. 219 Sub Id 336

SIMULATION OF COMPARATIVE STUDY OF JAMES-STEIN ESTIMATOR, RIDGE REGRESSION ESTIMATOR, AND MODIFIED KIBRIA LUKMAN ESTIMATOR IN HANDLING MULTICOLLINEARITY IN POISSON REGRESSION

M. Fikri Alyasa Zam Zami, Netti Herawati, Misgiyati, and Nusyirwan

Department of Mathematics, University of Lampung, Lampung, Indonesia

*Corresponding Author: Netti Herawati
Department of Mathematics, University of Lampung, Lampung, Indonesia

DOI : https://doi.org/10.52267/IJASER.2025.6203

ABSTRACT

Poisson regression is a statistical method used to analyze data with a response in the form of a count variable. The purpose of this study is to compare the performance of the Poisson James-Stein Estimator, Poisson Ridge Regression Estimator, and Poisson Modified Kibria-Lukman Estimator methods in dealing with multicollinearity using simulated data with n = 20, 40, 60 and 80 in poisson model (p=6) with ρ = 0.3 and 0.99. The best model was compered based on the MSE value. The results showed that in the partial correlation, PRRE method of k2 parameters better in overcoming multicollinearity at n = 20 and PMKLE parameters k2 better in overcoming multicollinearity at n = 40, 60, and 80 and in the full correlation data, PRRE method of k2 parameters better in overcoming multicollinearity at n = 20 and 40, and PMKLE parameters k2 were better in overcoming multicollinearity at n = 60 and 80.

KEYWORDS: James-Stein Estimator, Ridge Regression Estimator, Modified Kibria-Lukman Estimator, Multicollinearity.

1. INTRODUCTION

Poisson regression is a statistical method commonly used to analyze data with response variables in the form of counts. However, its application often faces the challenge of multicollinearity, which is a condition when two or more independent variables have a high correlation. High levels of multicollinearity can cause large variance in the least squares estimates of beta coefficients in the regression model, and can produce biased results (Lavery et al., 2019).

To overcome multicollinearity in Poisson regression, various alternative estimators have been developed. Among them are the Poisson Ridge Regression Estimator (PRRE) by Månsson and Shukur (2011), the Poisson Modified Kibria-Lukman Estimator (PMKLE) by Aladeitan et al. (2021), and the Poisson James-Stein Estimator (PJSE) by Amin et al. (2020). PRRE integrates the Poisson regression approach with ridge regression to stabilize parameter estimates, while PMKLE, a modification of the Kibria-Lukman Estimator, is designed to improve the efficiency of parameter estimation under high multicollinearity conditions. PJSE, on the other hand, utilizes a shrinkage technique to reduce the variance of the estimates and shows better performance than the Maximum Likelihood Estimation (MLE) method.

Previous research has examined the Poisson James-Stein Estimator to address multicollinearity issues in Poisson regression models, as conducted by Amin et al. (2020) using Monte Carlo simulations and aircraft damage data to evaluate the estimator's performance. The results indicated that the Poisson James-Stein Estimator yielded a lower Mean Square Error (MSE) compared to Maximum Likelihood Estimation (MLE) and other methods. Additionally, Oghenekevwe et al. (2021) studied the Poisson Ridge Regression Estimator using simulated data, finding it effective in handling multicollinearity based on the ridge parameter k used. Furthermore, Aladeitan et al. (2021) investigated the Poisson Modified Kibria-Lukman Estimator (PMKLE) through simulation data and case studies, demonstrating its efficiency in addressing multicollinearity compared to other estimators.

Previous studies have evaluated the performance of each estimator under various conditions using simulation data and case studies. The results show that these three methods are effective in overcoming multicollinearity with different advantages. However, there has been no study that directly compares the performance of the three.

Therefore, this study aims to compare PJSE, PRRE, and MKLE in overcoming multicollinearity in Poisson regression using a simulation approach. This study is expected to provide in-depth insight into the advantages and limitations of each estimator under various conditions, thus helping in selecting the right method.

2. LITERATURE REVIEW

a. Poisson James-Stein Estimator

Poisson James-Stein Estimator (PJSE) is an estimation method proposed to address the problem of multicollinearity in Poisson regression models. PJSE was developed as a solution by utilizing the concept of shrinkage estimator. This estimator is designed to reduce the variance inflation produced by MLE in situations where the explanatory variables are highly correlated. PJSE is defined as follows:

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

where (0 < c < 1) is a multiplier determined to reduce the MLE estimate which is defined as follows:

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

where S = $X^t\widehat{W}X$, $\widehat{W}$ = diag $\{\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \dots, \hat{\mu}_i\}$ and $\hat{\beta}_{MLE}$ is the unbiased estimate of β.

b. Poisson Ridge Regression Estimator

Poisson Ridge Regression Estimator (PRRE) is a modified approach to handle the problem of multicollinearity or high correlation among independent variables, by applying the ridge method in Poisson regression. This Ridge Regression method was first introduced by Hoerl & Kennard in 1970, and then modified by Mansson & Shukur in 2011. PRRE is defined as follows:

$$\hat{\beta}_{PRRE} = \left(kI + X^t\widehat{W}X\right)^{-1}X^tWX\hat{\beta}_{MLE}$$

where k (k > 0) is the ridge parameter, In dealing with multicollinearity problems in Poisson regression, it is necessary to determine the ridge parameter k that will be used in the Poisson Ridge Regression estimator model. Several methods have been proposed by previous studies, so in this study the following ridge parameters will be used:

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

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

Where $\hat{\alpha}^2_{max}$ is defined as the maximum value of $\gamma\hat{\beta}_{ML}$ and γ is an eigenvector element of the matrix $X^t\widehat{W}X$, $q_i = \frac{\lambda max}{(n-p)\hat{\sigma}^2+\lambda_{max}\hat{\alpha}_i^2}$ where λmax is the maximum value of the eigenvalues $X^t\widehat{W}X$ and $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i-\hat{\mu}_i)^2}{n-p-1}$

2.3 Poisson Modified Kibria-Lukman Estimator

Poisson Modified Kibria-Lukman Estimator (PMKLE) is a development of the KL estimator used to handle multicollinearity in the Poisson regression model. The Poisson regression model is usually applied to count data, where the dependent variable indicates a rare event. PMKLE is created by replacing the

International Journal of Applied Science and Engineering Review
IJASER
ISSN: 2582-6271
Vol. 6, Issue.2, Mar-Apr 2025, page no. 30-40

initial estimated component of the KL estimator using the Ridge estimator. Parameter estimates in PMKLE are as follows:

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

The selection of the value of k is usually done based on an approach that minimizes the Mean Squared Error (MSE) of the estimator. Several methods are used to determine the value of k:

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

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

$$k_3 = min\left(\frac{\lambda_i}{2\lambda_j\alpha_j^2 + 1}\right)$$

3. METHODOLOGY

In this study, we generated data with sample sizes of n = 20, 40, 60, and 80, containing full multicollinearity in 6 explanatory variables and partial multicollinearity in 3 explanatory variables, using software R with 1000 iterations. The explanatory variables were produced through Monte Carlo simulations:

$$X_p = \sqrt{(1-\rho^2)}Z_{ij} + \rho Z_{i(p+1)}$$

where ρ is set to ensure a high correlation among the 6 explanatory variables, and Z_{ij} refers to the pseudo-random numbers generated from the standard normal distribution.

The dependent variable of the Poisson regression model is generated using pseudo-random numbers from the $Po(\mu_i)$ where:

$$\mu_i = exp\left(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip}\right)$$

To evaluate the performance of the estimator using MSE as follows:

$$MSE = \frac{\sum_{i=1}^{R}\left(\hat{\beta}_i - \beta\right)'\left(\hat{\beta}_i - \beta\right)}{R}$$

ISSN: 2582-6271
Vol. 6, Issue.2, Mar-Apr 2025, page no. 30-40

where $\hat{\beta}_i$ is the β estimator obtained from PJSE, PRRE, and PMKLE, and R is the number of iterations.

4. RESULTS AND DISCUSSION

The simulation data in this study has 6 independent variables, and there are 2 variations, namely partial correlation data and full correlation data. Multicollinearity is examined through VIF. The results of the simulation data analysis with n = 20, 40, 60, and 80, and ρ = 0.3 and 0.99 cause the simulation data to contain high multicollinearity. The VIF results are shown in Table 1 and Table 2.

Table 1. Partial Correlation VIF Value.

n	X_1	X_2	X_3	X_4	X_5	X_6
20	70.54	28.89	69.47	1.35	1.18	1.24
40	43.04	37.15	47.34	1.52	1.34	1.39
60	20.59	26.42	20.27	1.06	1.24	1.07
80	45.10	39.57	47.79	1.26	1.17	1.15

Table 2. Full Correlation VIF Value.

n	X_1	X_2	X_3	X_4	X_5	X_6
20	66.36	83.38	53.57	63.05	155.98	82.97
40	44.19	78.69	105.59	83.58	64.45	105.55
60	39.64	25.08	34.01	40.14	38.10	45.90
80	26.88	25.82	26.72	29.45	29.78	35.14

Based in Table 1, it can be seen that the independent variables $X_{1,}, X_2$ and X_3 have VIF values of more than 10, indicates multicollinearity in the three variables. Meanwhile, the independent variables $X_{4,}, X_5$ and X_6 have VIF values of less than 10, indicating that there is no multicollinearity.

And based in Table 2, it can be seen that the independent variables $X_{1,}, X_2$, $X_3, X_{4,}, X_5$ and X_6 have VIF values of more than 10, which indicates multicollinearity in the six variables.

Next is to calculate the MSE values for PJSE, PRRE and PMKLE at $n = 20, 40, 60, 80$ to find which one is better at handling the multicollinearity present in the model.

ISSN: 2582-6271

Vol. 6, Issue.2, Mar-Apr 2025, page no. 30-40

Table 3: MSE Value on Partial Correlation Data

MSE		n			
		20	40	60	80
MLE		7.076	4.823	3.641	3.277
PJSE	c	6.565	4.738	3.599	3.251
PRRE	k_1	6.535	4.732	3.596	3.248
PRRE	k_2	5.082	4.549	3.561	3.239
PMKLE	k_1	5.761	4.562	3.510	3.193
PMKLE	k_2	5.428	4.465	3.460	3.161
PMKLE	k_3	6.302	4.687	3.574	3.234

The results of the analysis in Tables 3 show that at small sample sizes (n = 20), PRRE k_2 shows a smaller MSE value compared to PJSE and PMKLE, indicate that PRRE k_2 is more effective in overcoming multicollinearity in data with small sample sizes. Meanwhile, at larger sample sizes (n = 40, 60 and 80), PMKLE shows better performance with a lower MSE value compared to other estimators. This indicates that PMKLE is more effective in overcoming multicollinearity when the sample size increases.

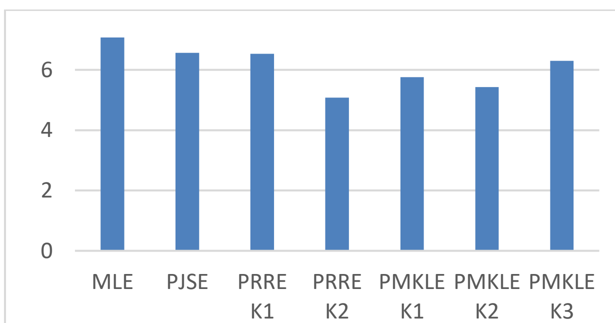

Figure 1: Graph of MSE values for n = 20

Figure 1 shows a graphical comparison of MSE values for each estimator, indicate that PRRE k_2 tends to have better performance with a lower MSE value than other estimators at n = 20.

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 6, Issue.2, Mar-Apr 2025, page no. 30-40

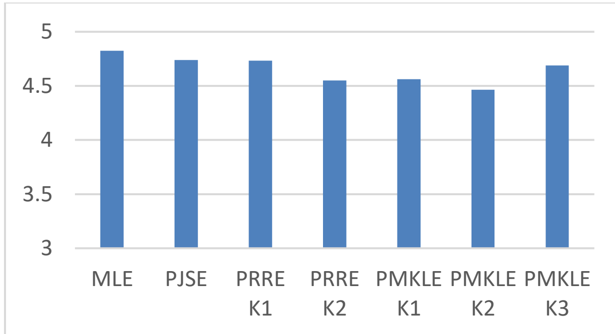

Figure 2: Graph of MSE values for n = 40

Figure 2 shows a graphical comparison of MSE values for each estimator, indicate that PMKLE k_2 tends to have better performance with a lower MSE value than other estimators at n = 40.

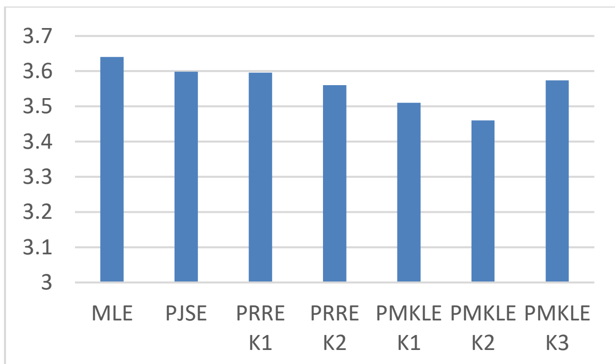

Figure 3: Graph of MSE values for n = 60

Figure 3 shows a graphical comparison of MSE values for each estimator, indicate that PMKLE k_2 tends to have better performance with a lower MSE value than other estimators at n = 60

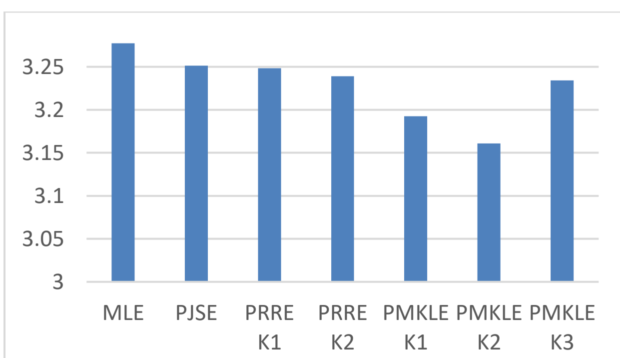

Figure 4: Graph of MSE values for n = 80

Figure 4 shows a graphical comparison of MSE values for each estimator, indicate that PMKLE k_2 tends to have better performance with a lower MSE value than other estimators at n = 80

Table 4: MSE Values on Full Correlation Data

MSE		n			
		20	40	60	80
MLE		43.355	19.343	13.823	10.831
PJSE	c	42.225	19.149	13.719	10.760
PRRE	k_1	42.687	19.249	13.776	10.801
PRRE	k_2	39.535	18.950	13.685	10.760
PMKLE	k_1	41.469	19.064	13.685	10.741
PMKLE	k_2	40.809	18.957	13.628	10.704
PMKLE	k_3	42.365	19.202	13.754	10.786

The results of the analysis in Tables 4 show that at sample sizes n = 20 and 40, PRRE k_2 shows a smaller MSE value compared to PJSE and PMKLE, indicating that PPRE k_2 is more effective in overcoming multicollinearity in data with sample sizes n = 20 and 40. Meanwhile, at sample sizes n = 60 and 80, PMKLE shows better performance with a lower MSE value compared to the other two estimators. This indicates that PMKLE is effective in overcoming multicollinearity when the sample size increases.

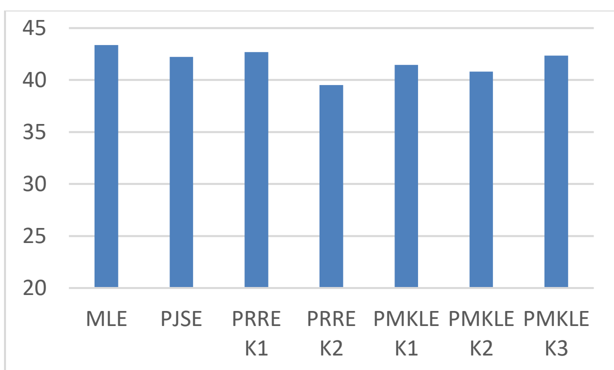

Figure 5: Graph of MSE values for n = 20

Figure 5 shows a graphical comparison of MSE values for each estimator, indicate that PRRE k_2 tends to have better performance with a lower MSE value than other estimators at n = 20.

International Journal of Applied Science and Engineering Review

IJASER

ISSN: 2582-6271

Vol. 6, Issue.2, Mar-Apr 2025, page no. 30-40

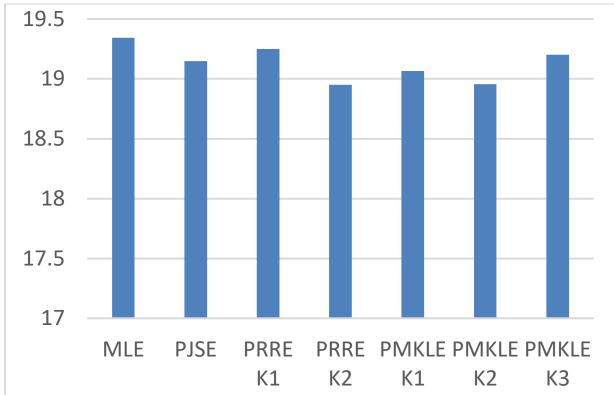

Figure 6: Graph of MSE values for n = 40

Figure 6 shows a graphical comparison of MSE values for each estimator, indicate that PRRE k_2 tends to have better performance with a lower MSE value than other estimators at n = 40.

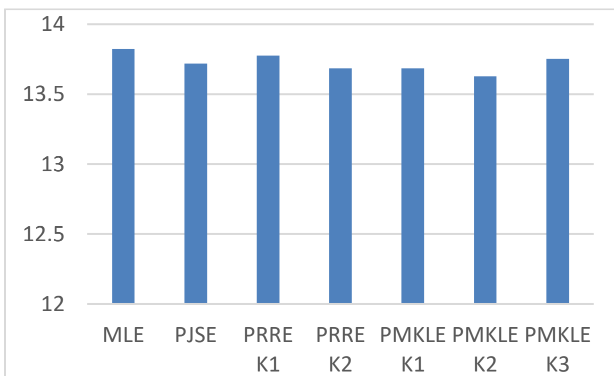

Figure 7: Graph of MSE values for n = 60

Figure 7 shows a graphical comparison of MSE values for each estimator, indicate that PMKLE k_2 tends to have better performance with a lower MSE value than other estimators at n = 60.

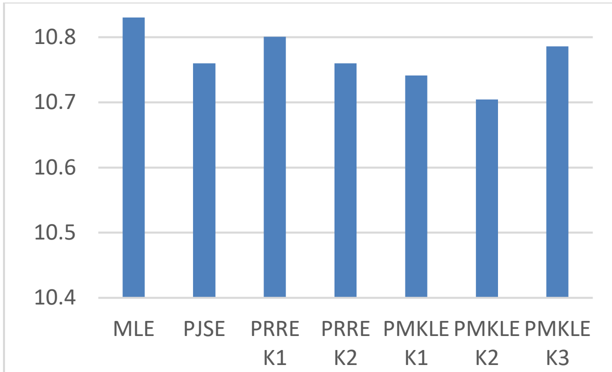

Figure 8: Graph of MSE values for n = 80

Figure 8 shows a graphical comparison of MSE values for each estimator, indicate that PMKLE k_2 tends to have better performance with a lower MSE value than other estimators at n = 80.

5. CONCLUSIONS

Based on the results obtained, the following are conclusions regarding the use of the PJSE, PRRE and PMKLE methods in Poisson regression analysis:

1. At partial correlation data simulation for sample size n = 20, the performance of the Poisson Ridge Regression Estimator method using the parameter value k_2 in overcoming multicollinearity is better than the Poisson James-Stein Estimator and Poisson Modified Kibria-Lukman Estimator methods because it produces a smaller MSE value. Meanwhile, for sample sizes n = 40, 60 and 80, the performance of the Poisson Modified Kibria-Lukman Estimator method using the parameter value k_2 in overcoming multicollinearity is better than the other methods because it produces a smaller MSE value.
2. At full correlation data simulation for sample sizes n = 20 and 40, the performance of the Poisson Ridge Regression Estimator method using the parameter value k_2 in overcoming multicollinearity is better than the Poisson James-Stein Estimator and Poisson Modified Kibria-Lukman Estimator methods because it produces a smaller MSE value. Meanwhile, at sample sizes n = 60 and 80, the performance of the Poisson Modified Kibria-Lukman Estimator method using the parameter value k_2 in overcoming multicollinearity is better than the other methods because it produces a smaller MSE value.

REFERENCES

1. Abonazel, M. R. 2023. New modified two-parameter Liu estimator for the Conway—Maxwell Poisson regression model. *Journal of Statistical Computation and Simulation.* **93**(12): 1976-1996.

2. Aladeitan, B. B., Adebimpe, O., Lukman, A. F., Oludoun, O., & Abiodun, O.E. 2021. Modified Kibria-Lukman (MKL) Estimator for the Poisson Regression Model, *F1000Research*. **10**(548): 1-19.
3. Amin, M., Akram, M. N., & Amanullah, M. 2020 . On the James-Stein Estimator for the Poisson Regression Model. *Communications in Statistics: Simulation and Computation*. **51**(10): 5596-5608.
4. Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., & Lee, T. C. 1985. *The Theory and Practice of Econometrics.* 2nd Edition. John Wiley & Sons. USA.
5. Khoshsirat, S., & Kambhamettu, C. 2023 . Improving Normalization with the James-Stein Estimator. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **42**(12): 2980-2992.
6. Lavery, M. R., Acharya, P., Sivo, S. A. & Xu, L. 2019. Number of predictors and Multicollinearity: What are Their Effects on Error and Bias in Regression?,.*Communications in Statistics-Simulation and Computation*. **48**(1): 27-38.
7. Lukman, A. F., Allohibi, J., Jegede, S. L., Adewuyi, E. T., Oke, S., & Alharbi, A. A. 2023. Kibria–Lukman-Type Estimator for Regularization and Variable Selection with Application to Cancer Data. *Mathematics.* **11**(23): 4795.
8. Månsson, K., & Shukur, G. 2011. A Poisson Ridge Regression Estimator. *Econimic Modelling.* **28**(4): 1475–1481.
9. Montgomery, D. C., Peck, E. A., & Vining, G. G. 2012. *Introduction to Linear Regression Analysis*. 4th Edition. John Willey and Sons, New York.
10. Myers, R. H. 1990. Classical and Modern Regression with Applications. 2nd Edition. Duxbury Press. California.
11. Oghenekevwe, E. H., Florence, A. K., & Abidemi, A. K., 2021. Poisson Ridge Regression Estimators: A Performance Test. *American Journal of Theoretical and Applied Statistics*. **10**(2): 111-121.
12. Pudjoatmodjo, B. Hendayun, M. 2016. Kehandalan Software Berdasarkan Data Sekunder Menggunakan Distribusi Poisson dan Kualifikasi Cronbach's Alpha. *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* (JNTETI). **5**(2): 53-62.
13. Sami, F., Amin, M., & Aljeddani, S. M. A., 2024 . Stein Estimation in the Conway-Maxwell Poisson Model with Correlated Regressors. *International Journal of Advanced and Applied Sciences*. **11**(7): 49- 56.
14. Wardah, T. L., Lubis, R. S., & Aprilia, R. 2022. Analisis Poisson Ridge Regression (PRR) pada Faktor yang Mempengaruhi Kecelakaan Lalu Lintas di Sumatera Utara. *Jurnal Matematika dan Pendidikan Matematika*. **5**(2): 154-160.