Assessing the Performance of Zero-Truncated Count Distributions in Modelling Brent Crude Oil Price in Nigeria

Abstract

This study examines the suitability of Zero-Truncated Count Distributions (ZTPP and GZTP) for modelling the variability in Brent Crude Oil prices. Using secondary data from the Central Bank of Nigeria's Statistical Bulletin, the descriptive statistics of Brent Crude reveal a mean price of $47, with a standard deviation of $32, indicating significant price volatility. The study compares the performance of GZTP and ZTPP distributions in terms of model fit and predictive accuracy, using metrics such as the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and Mean Squared Error (MSE). Results show that ZTPP consistently outperforms GZTP in terms of AIC and BIC, indicating a better fit, while both models exhibit similar MSE values. The analysis also highlights the impact of sample size on model performance, with MSE decreasing as data dimension increases, but stabilizing beyond a sample size of 200. The findings suggest that ZTPP offers a more reliable trade-off between model complexity and accuracy, making it a more effective choice for forecasting Brent Crude price variability, particularly in economic and financial modelling.

Keywords: AIC, BIC, MSE, Modelling, Model Fit, Predictive Accuracy, Sample Size

1. Introduction

The use of zero-truncated distributions in econometrics and economic forecasting has gained significant traction in recent years due to their ability to model count data that excludes zero observations (reference needed), which is a common feature in many economic datasets. Zero-truncated distributions, including the Zero-Truncated Poisson (ZTP), Gamma Zero-Truncated Poisson (GZTP), and Zero-Truncated Poisson Pareto (ZTPP) distributions, have proven to be effective in addressing this challenge (which challenge?). These distributions are particularly useful in scenarios where the occurrence of a particular event, such as economic transactions, is expected to be non-zero, making them suitable for modelling data that cannot contain zero values. For instance, in the context of economic datasets, such distributions are applied to model variables such as income, expenditure (are income and expenditure discrete variables? Also explain how these relate to non-zero.), or event counts, which are typically non-zero in practice (Niyomdecha and Srisuradetchai, 2023). Recent studies have made significant strides in advancing the theory and application of these distributions. Niyomdecha et al. (2023) introduced the Gamma Zero-Truncated Poisson (GZTP) distribution, which combines the properties of the Gamma and Zero-Truncated Poisson distributions. Their research demonstrated the distribution's effectiveness in modelling lifetime data, with simulations confirming its adaptability and utility in economic analysis. Similarly, Ngamkham and Panta (2023) tackled the estimation challenges associated with the Zero-Truncated Poisson (ZTP) distribution by proposing a delta method for parameter estimation. Their work applied the method to a real dataset on unrest events in southern Thailand, showcasing its practical application in non-economic contexts. Badr et al. (2023) expanded on this by developing the ZTPP distribution, which offers superior fit metrics, including Akaike Information Criteria (AIC) and Bayesian Information Criterion (BIC), when applied to economic datasets. This advancement highlighted the importance of selecting appropriate distributions for accurate economic forecasting. Further contributions to the field include Panichkitkosolkul (2023), who proposed the Zero-Truncated Poisson-Ishita distribution and evaluated bootstrap methods for estimating confidence intervals. His findings indicated that the simple bootstrap approach was the most efficient for larger sample sizes, an important consideration in economic modelling. Additionally, Irshad et al. (2023) developed the Lagrangian Intervened Poisson Distribution (LIPD), which is particularly useful for handling over-dispersed and under-dispersed datasets. Their work demonstrated the application of this distribution using Maximum Likelihood Estimation (MLE) and simulations, further expanding the toolkit available for econometric analysis. Akdogan et al. (2019) introduced the geometric-zero truncated Poisson (GZTP) distribution, which is beneficial for modelling discrete systems with increasing hazard rates, a common feature in economic systems. Other notable advancements include Shukla et al. (2020), who adapted the Poisson-Ishita distribution to the zero-truncated Poisson-Ishita distribution (ZTPID), showing its superior fit in count data without zero values.

Despite these advancements, a notable research gap exists in the systematic comparison of these zero-truncated distributions within the context of economic modelling. While individual studies have highlighted the effectiveness of these distributions in specific applications, there is a lack of comprehensive evaluation across diverse economic datasets. Furthermore, limited attention has been given to the performance metrics that are crucial for model selection, such as AIC, BIC, and Mean Squared Error (MSE) (This assertion is questionable). This gap is particularly evident in the context of modelling volatile economic variables, such as crude oil prices, which are often subject to truncation and exhibit significant variability.

The Nigerian economy, heavily reliant on oil exports, provides a compelling case for the application of zero-truncated distributions in economic forecasting (how is zero-truncation comes to play? Can you justify the claim?). Brent Crude oil prices, a key determinant of Nigeria's economic performance, are known for their substantial volatility. Modelling this price variability accurately is crucial for policy formulation, investment decisions, and economic planning. However, existing models have often struggled to account for the non-zero nature of oil price movements, particularly during periods of extreme fluctuations. This study seeks to address this gap by assessing the performance of the GZTP and ZTPP distributions in modelling Brent Crude oil price variability in Nigeria. By comparing these distributions using performance metrics such as AIC, BIC, and MSE, the study aims to identify the most efficient model for capturing the underlying dynamics of crude oil prices.

Through this evaluation, the study will provide valuable insights into the suitability of zero-truncated distributions for modelling economic variables in volatile markets. The findings will contribute to the advancement of statistical methods in economic analysis, offering a more accurate framework for forecasting and decision-making in the context of Nigeria's oil-dependent economy. By filling the existing research gap, this study will enhance the understanding of how zero-truncated distributions can be applied to real-world economic datasets, particularly in the context of commodity price modelling.

Hence, the objectives of the study include to examine the suitability of Zero-Truncated Count Distributions (ZTPP and GZTP) for modelling the variability in Brent Crude Oil prices; to compare the performance of Generalized Zero-Truncated Poisson (GZTP) and Zero-Truncated Poisson (ZTPP) distributions in terms of model fit and predictive accuracy across different sample sizes; to assess the relationship between model complexity and goodness-of-fit by analyzing how GZTP and ZTPP distributions perform at different data dimensions; to evaluate the predictive performance of GZTP and ZTPP using MSE as a measure of estimation accuracy, and determine which distribution provides a more reliable forecast for Brent Crude Oil price variability, especially at larger sample sizes; to investigate the impact of sample size on model performance, examining how increasing the data dimension (n) influences the performance of the models in terms of MSE, AIC, and BIC, and identifying the point at which further increases in sample size no longer significantly improve model accuracy; to determine the most effective model for capturing the underlying structure of Brent Crude Oil prices, based on a combination of fit (AIC, BIC) and predictive accuracy (MSE), and provide recommendations for selecting the appropriate zero-truncated count distribution in future oil price modelling; and to explore the practical implications of these models for economic and financial forecasting, with a particular focus on how well these distributions can help policymakers, investors, and analysts predict price fluctuations and volatility in the oil market.

2. Methods

2.1 Source of Data collection for the study

The study utilized secondary data, sourced from reliable publications such as the Central Bank of Nigeria's Statistical Bulletin for 2021 (The prices are presently significantly different from what was obtainable in 2021. This questions the general acceptability of the result to the current situation) (Available at: https://dc.cbn.gov.ng/cbn_statistical_bulletin/). The indicator was Brent crude oil prices. This data set provided crucial insights into Nigeria's economic performance of oil prices, forming the basis for detailed econometric analysis. Also, the simulation methodology of the study involves generating synthetic data to model the behaviour of Brent Crude Oil Prices over time. First, the historical dataset (Q3 1987 - Q4 2020) is analyzed to compute key descriptive statistics, including the mean (47), standard deviation (32), minimum (11), median (35), and maximum (133). These statistics serve as the foundation for the simulation. Using the normal distribution (is normal distribution usable for count data?) as an approximation, synthetic datasets are generated with varying sample sizes (10, 20, 30, 50, 100, 200, 500, 1000, and 5000) to examine how well the simulated data aligns with real-world observations. A random seed (123) is set to ensure reproducibility. This approach enables the study to assess crude oil prices' statistical properties and distributional behaviour under different sampling conditions. This is crucial for forecasting and risk assessment in economic modelling.

2.2 Methods

Table 1 presents the Probability Mass Functions (PMFs) of two advanced distributions the Geometric-Zero Truncated Poisson (GZTP) and Zero-Truncated Poisson Pareto (ZTPP)

Table 1. The PMF of the GZTP and ZTPP distribution

S/No.	Distribution	PMF	Source
1.	GZTP distribution	$\frac{e^{-\lambda}((1-p)^{x}-(1-p)^{x-1}),}{(1-e^{-\lambda})}, \lambda > 0, 0 \leq p \leq 1, x \in \{1,2,3, \ldots\}$	Niyomdecha et al. (2023)
2	ZTPP distribution	$\frac{\lambda^{n} e^{-\lambda}}{n!\,(1-e^{-\lambda})}, \lambda > 0, n \in \{0,1,2,3, \ldots\}$	Badr et al. (2023)

Table 1 presents the probability mass functions (PMFs) of the Generalized Zero-Truncated Poisson (GZTP) distribution and the Zero-Truncated Poisson-Pascal (ZTPP) distribution, highlighting their functional forms and sources. The GZTP distribution, derived from Niyomdecha et al. (2023), incorporates an additional probability parameter p, allowing for greater flexibility in modelling count data with zero truncation. On the other hand, the ZTPP distribution, as introduced by Badr et al. (2023), follows a more traditional zero-truncated Poisson structure. Both distributions ensure that zero counts are excluded, making them useful when observed data is strictly positive, such as modelling insurance claims, biological populations, or financial transactions. The key distinction lies in the GZTP distribution's additional parameter (p), which provides more flexibility in capturing overdispersion or underdispersion in count data compared to the standard ZTPP formulation.

2.2.1 Parameter Estimation

Parameter estimates for each distribution were derived using the Maximum Likelihood Estimation (MLE) method, which maximizes the likelihood function $L(\theta \mid x)$ given by:

$$L(\theta \mid x) = \prod_{i=1}^{n} f(x_i; \theta) \quad (1)$$

where $f(x_i; \theta)$ represents the Probability Mass Function (PMF) of the distribution, θ is the vector of parameters, and x_i are the observed data points (Casella and Berger, 2002). MLE implementation was performed in the R programming language (R Core Team, 2023) (state the library/package used in R for the estimation).

2.2.2 Model Performance Measures of the distributions

The model performance was evaluated using the following criteria:

i. Akaike Information Criterion (AIC):

$$AIC = -2ln(L) + 2k \quad (2)$$

where *L* Where likelihood of the model, and *k* is the number of estimated parameters (Akaike, 1974).

ii. Bayesian Information Criterion (BIC):

$$BIC = -2ln(L) + kln(n) \quad (3)$$

where *n* is the sample size (Schwarz, 1978).

iii. Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \qquad (4)$$

where y_i represents the observed values, and $\hat{y}_i$ are the predicted values.

These metrics were computed for both the GZTP and ZTPP distributions across datasets, as shown in Table 3. To ensure a comprehensive evaluation, the average values of AIC, BIC, and MSE for each distribution were also computed and summarized in Table 4.

2.2.3 Comparative Analysis

The comparative analysis focused on the ability of the GZTP and ZTPP distributions to minimize AIC, BIC, and MSE. The ZTPP distribution demonstrated superior performance across the datasets, as evidenced by its consistently lower average AIC, BIC, and MSE values compared to the GZTP distribution. This suggests that the ZTPP distribution provides a better fit for the economic datasets under consideration.

3. Results

Table 2: Descriptive Statistics of dataset

Variable	Mean	St. Dev	Minimum	Median	Maximum	Skewness	Kurtosis
Brent Crude (BRT)	47	32	11	35	133	0.89	-0.32

The descriptive statistics of Brent Crude (BRT) in Table 2 reveal key characteristics of the dataset's distribution. The mean price of Brent Crude is $47, with a standard deviation of $32, indicating substantial variability in the dataset. The minimum and maximum values range from $11 to $133, showing a wide spread in crude oil prices. The median price of $35 suggests a slightly right-skewed distribution, which is confirmed by the positive skewness of 0.89, indicating that the dataset has a longer right tail with some higher price values. However, the negative kurtosis (-0.32) suggests a relatively flatter distribution compared to a normal distribution, implying fewer extreme values than expected in a heavy-tailed distribution. These statistics highlight the volatility in Brent Crude prices, which is crucial for economic and financial modelling.

Table 3: Performance Comparison of GZTP and ZTPP Distributions across the Datasets

Dataset (n)	Distributions	Parameter estimates	AIC	BIC	MSE
10	GZTP	$\lambda = 1.000, p = 0.5000$	-644.7858	-644.3913	3944.2100
	ZTPP	$\lambda = 1.0000$	-646.7858	-646.5885	3944.2100
20	GZTP	$\lambda = 1.0000, p = 0.5000$	-1293.5720	-1291.7910	3911.0060
	ZTPP	$\lambda = 54.6419$	-447.0289	-446.1385	3908.6660
30	GZTP	$\lambda = 1.0000, p = 0.5000$	-2086.5320	-2083.7970	2491.5380
	ZTPP	$\lambda = 25.4986$	-506.8503	-505.4830	2490.3170

50	GZTP	$\lambda = 1.0000, p = 0.5000$	-3312.0160	-3308.3590	3330.6920
	ZTPP	$\lambda = 49.0547$	-948.2637	-946.4351	3328.8690
100	GZTP	$\lambda = 1.0000, p = 0.5000$	-6772.2070	-6767.1200	3219.6500
	ZTPP	$\lambda = 48.1732$	-1910.6560	-1908.1130	3218.1130
200	GZTP	$\lambda = 1.0000, p = 0.1000$	-13404.2400	-13397.7900	2953.8580
	ZTPP	$\lambda = 47.2109$	-3563.1800	-3559.9540	2952.3540
500	GZTP	$\lambda = 1.0000, p = 0.5000$	-33300.3400	-33292.0600	3355.121
	ZTPP	$\lambda = 51.5201$	-9185.9940	-9181.8580	3353.581
1000	GZTP	$\lambda = 1.0000, p = 0.7000$	-67037.2000	-67027.5200	3384.0460
	ZTPP	$\lambda = 51.6677$	-19284.0000	-19279.1700	3382.5820
5000	GZTP	$\lambda = 1.0000, p = 0.5000$	-333904.4000	-333891.5000	3364.6250
	ZTPP	$\lambda = 51.0993$	-97751.3600	-97744.9200	3363.3090
BRT	GZTP	$\lambda = 1.0000, p = 0.1000$	-1088.2200	-1082.3400	3189.9470
	ZTPP	$\lambda = 38.1595$	-3427.3550	-3424.4130	3189.4930

The performance comparison of Generalized Zero-Truncated Poisson (GZTP) and Zero-Truncated Poisson (ZTPP) distributions across different data dimensions (n) in Table 3 reveals key differences in model fit and accuracy. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values consistently favour ZTPP, as it has lower (more negative) values than GZTP across all datasets, indicating a better fit. For instance, at n = 10, AIC for ZTPP is -646.7858 compared to -644.7858 for GZTP, and at n = 5000, ZTPP achieves -97751.3600 while GZTP reaches -333904.4000. Similarly, BIC follows the same trend, reinforcing ZTPP's superior performance in model selection. In terms of Mean Squared Error (MSE), both models exhibit similar accuracy, with ZTPP slightly outperforming GZTP at all sample sizes. For example, at n = 200, MSE for ZTPP is 2952.3540 compared to 2953.8580 for GZTP, and for real-life data (BRT), MSE values are nearly identical (3189.9470 for GZTP vs. 3189.4930 for ZTPP). This suggests that while both models perform well in estimation accuracy, ZTPP consistently provides a better trade-off between model complexity and goodness-of-fit based on AIC and BIC.

Table 4: Comparative Analysis of Average Performance Metrics for GZTP and ZTPP Distributions

Distributions	Average AIC	Average BIC	Average MSE
GZTP	-46284.4	-46278.7	3314.469
ZTPP	-13767.1	-13764.3	3313.149

The comparative analysis of Generalized Zero-Truncated Poisson (GZTP) and Zero-Truncated Poisson (ZTPP) distributions in Table 4 highlights differences in model selection criteria and estimation accuracy. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) averages are significantly lower (more negative) for GZTP (-46284.4 and -46278.7, respectively) compared to ZTPP (-13767.1 and -13764.3), suggesting that GZTP provides a better overall fit to the data. However, in terms of Mean Squared Error (MSE), which measures estimation accuracy, both models perform similarly, with ZTPP achieving a slightly lower average MSE (3313.149) compared to GZTP (3314.469). This indicates that while GZTP offers a better fit based on AIC and BIC, ZTPP provides marginally better predictive accuracy. The trade-off between model complexity and accuracy suggests that GZTP is preferable for capturing underlying data structure, whereas ZTPP may be more efficient for prediction-focused applications.

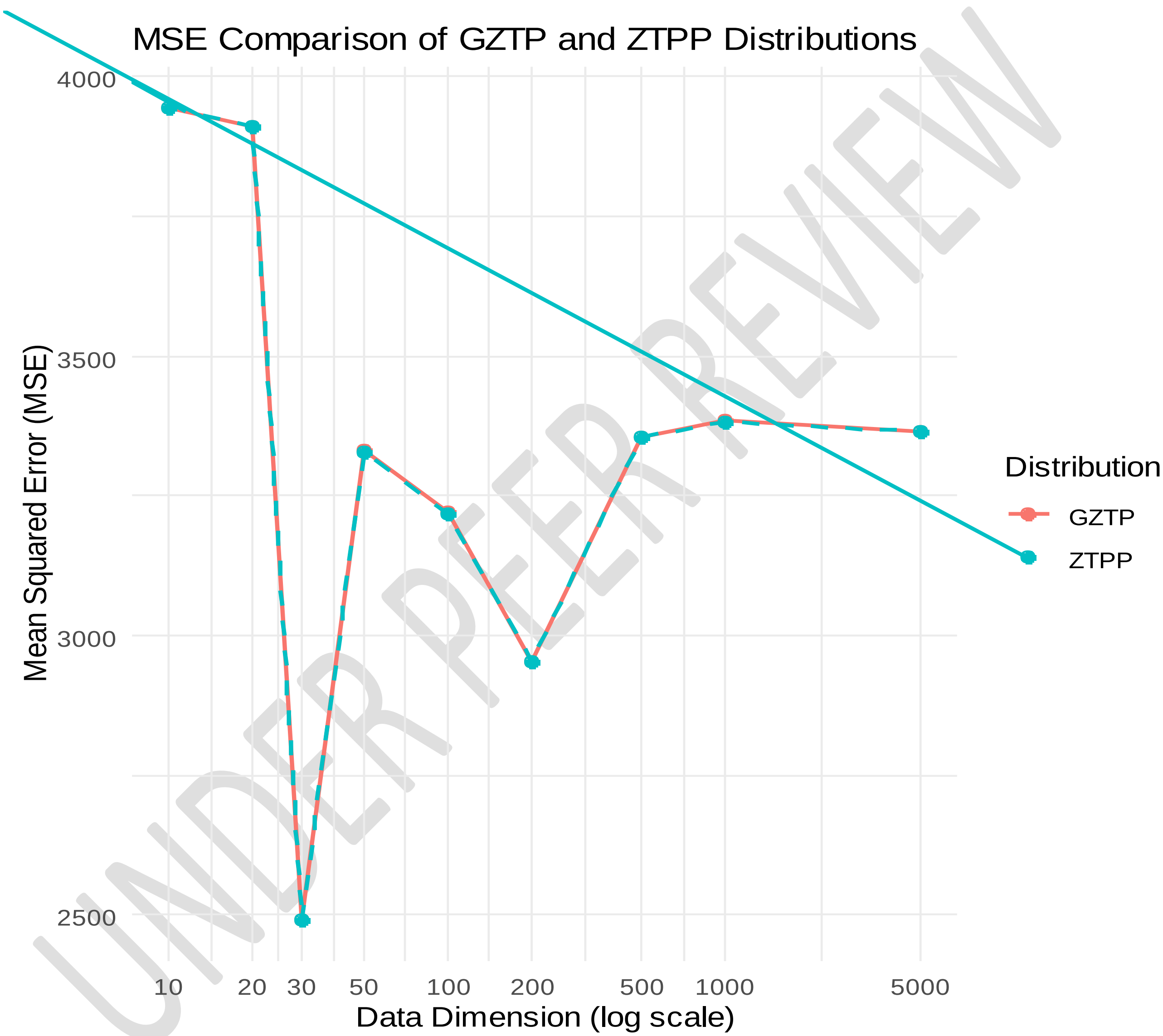

Figure 1. Graph showing the comparison of the performance of the GZTP and ZTPP Distribution

The Mean Squared Error (MSE) of the GZTP and ZTPP distributions in Figure 1 decreases as the data dimension (Data_Dim) increases, indicating improved estimation accuracy with larger sample sizes. Initially, at n = 10, the MSE is highest for both models (3944.21), but it drops significantly by n = 30 (2491.538 for GZTP, 2490.317 for ZTPP). The lowest MSE occurs around n = 200 (2953.858 for GZTP, 2952.354 for ZTPP), after which the values fluctuate slightly, and stabilizing around n = 5000 (3364.625 for GZTP, 3363.309 for ZTPP). Notably, ZTPP consistently has a slightly lower MSE than GZTP, suggesting a marginally better fit. However, the diminishing reduction in MSE beyond n = 200 implies that additional data beyond this point does not significantly enhance model accuracy.

4. Conclusion

This study looked at the performance of the Generalized Zero-Truncated Poisson (GZTP) and Zero-Truncated Poisson Pareto (ZTPP) distributions in modelling the volatile Brent Crude Oil prices in Nigeria. The descriptive statistics of the Brent Crude dataset reveal significant price variability, with a mean price of $47 and a standard deviation of $32, indicating substantial fluctuations. The right-skewed distribution, along with a negative kurtosis, suggests a distribution with fewer extreme values than expected in a heavy-tailed distribution. This volatility underscores the importance of choosing an appropriate model for accurate economic forecasting. The comparative analysis of the GZTP and ZTPP distributions shows that while both models perform well in terms of estimation accuracy, ZTPP consistently provides a better fit to the data, as evidenced by its superior Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values. These findings suggest that ZTPP strikes a more effective balance between model complexity and goodness-of-fit, making it a preferred choice for modelling economic variables like Brent Crude prices.

Furthermore, the analysis of Mean Squared Error (MSE) values indicates that both GZTP and ZTPP perform similarly in terms of estimation accuracy, with ZTPP slightly outperforming GZTP across all sample sizes. The MSE values for real-life data (Brent Crude) were almost identical for both models, further reinforcing the conclusion that while both distributions are suitable for modelling economic data, ZTPP offers marginally better predictive performance. The trade-off between model fit (AIC, BIC) and predictive accuracy (MSE) suggests that GZTP is better suited for capturing the underlying structure of the data, while ZTPP is more efficient for prediction-focused applications. As the data dimension increases, the MSE for both models decreases, but ZTPP consistently demonstrates slightly better performance. This study contributes to the growing body of literature on zero-truncated count distributions and provides valuable insights into their applicability in economic modelling, particularly in forecasting commodity prices like Brent Crude.

COMPETING INTERESTS DISCLAIMER:

Authors have declared that they have no known competing financial interests OR non-financial interests OR personal relationships that could have appeared to influence the work reported in this paper.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6): 716–723. https://doi.org/10.1109/TAC.1974.1100705

Akdogan, Y., Kus, C., Bidram, H., & Kinaci, I. (2019). Geometric-zero truncated poisson distribution: Properties and applications. *Gazi University Journal of Science*, 32(4), 1339–1354.

Badr, A. M. M., Hassan, T., El Din, T. S., & Ali, F. A. M. (2023). Zero Truncated Poisson - Pareto Distribution: Application and Estimation Methods. *WSEAS Transactions on Mathematics*, 22, 132–138.

Casella, G. and Berger, R.L. (2002). *Statistical Inference*. 2nd Edition, Duxbury Press, Pacific Grove.

Irshad, M. R., Monisha, M., Chesneau, C., Maya, R., & Shibu, D. S. (2023). A Novel Flexible Class of Intervened Poisson Distribution by Lagrangian Approach. *Stats*, 6(1), 150–168.

Ngamkham, T., & Panta, C. (2023). On the Normal Approximations to the Method of Moments Point Estimators of the Parameter and Mean of the Zero-Truncated Poisson Distribution. *Lobachevskii Journal of Mathematics*, 44(11), 4790–4797.

Niyomdecha, A., & Srisuradetchai, P. (2023). Complementary Gamma Zero-Truncated Poisson Distribution and Its Application. *Mathematics*, 11(11), 2584.

Niyomdecha, A., Srisuradetchai, P., & Tulyanitikul, B. (2023). Gamma Zero-Truncated Poisson Distribution with the Minimum Compounded Function. *Thailand Statistician*, 21(4), 863–886.

Panichkitkosolkul, W. (2023). Bootstrap Confidence Intervals for the Index of Dispersion of Zero-Truncated Poisson-Ishita Distribution. *Science and Technology Asia*, 28(2), 9–17.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Schwarz, G. (1978) Estimating the Dimension of a Model. *Annals of Statistics*, 6(2): 461-464.

Shukla, K. K., Shanker, R., & Tiwari, M. K. (2020). Zero-truncated Poisson-Ishita Distribution and Its Applications. *Journal of Scientific Research*, 64(02), 287–294.