Asia Pac. J. Math. 2026 **13**:7

# MODELING THE DENGUE EFFECTIVE REPRODUCTIVE NUMBER IN THE CARAGA REGION, PHILIPPINES USING THE MIXTURE OF GENERALIZED GAMMA AND LOMAX (MIGGLO) DISTRIBUTION

JOE VINCENT B. DELUAO

Department of Mathematics, University of Southeastern Philippines, Philippines

joe.deluao@usep.edu.ph

Received Nov. 5, 2025

Abstract. Modeling the effective reproductive number ($R_{\text{eff}}$) of dengue transmission is inherently challenging due to the stochastic and heterogeneous nature of epidemic dynamics. This study introduces the Mixture of Generalized Gamma and Lomax (MiGGLo) distribution as a novel probabilistic model capable of capturing both the skewed central tendency and heavy-tailed behavior observed in epidemiological data. Theoretical properties, including the $k$th raw moment, were derived to establish its analytical tractability. A Monte Carlo simulation assessed the performance of Maximum Likelihood Estimation (MLE), Ordinary Least Squares (OLS), and Partial Expectation–Maximization (EM) methods across varying sample sizes. Results revealed that MLE consistently achieved the lowest mean squared error and bias, affirming its superior parameter recovery. Application to dengue $R_{\text{eff}}$ data in the Caraga Region, Philippines, demonstrated excellent model fit, with the Generalized Gamma component ($\alpha_g = 0.7609$, $\theta_g = 1.2119$, $\tau = 4.1938$) representing endemic transmission, and the Lomax component ($\alpha_l = 19.328$, $\theta_l = 10.0$, $p = 0.05$) capturing sporadic outbreak surges. The MiGGLo model effectively characterizes both the central and extreme behaviors of dengue transmission, offering a robust and flexible framework for epidemiological modeling and disease forecasting.

2020 Mathematics Subject Classification: Primary 62E15; Secondary 62F10, 62P12.

Key words and phrases. dengue epidemiological modeling; mixture distribution frameworks; heavy-tailed epidemic behavior; maximum likelihood estimation; disease outbreak characterization.

## 1. Introduction

The escalating complexity and heterogeneity of data across diverse scientific and practical domains necessitate the development of increasingly flexible and robust statistical models. Traditional unimodal and symmetrically distributed models often prove inadequate for characterizing phenomena exhibiting diverse tail behaviours, multimodality, or pronounced asymmetry, which are increasingly prevalent in fields such as econometrics, biostatistics, environmental science, and public health [10,14]. The failure

DOI: 10.28924/APJM/13-7

to employ appropriate probabilistic frameworks can lead to biased analyses, flawed inferences, and sub-optimal decision-making. Consequently, there is a persistent and growing demand for advanced statistical distributions capable of capturing these intricate data patterns with greater fidelity [2,9,11,13]. This is particularly true for heavy-tailed distributions, which are crucial for modelling extreme events and outliers often observed in real-world data [6,21].

Among the foundational distributions frequently employed in statistical modelling are the Generalized Gamma (GG) and Lomax (Lo) distributions, each possessing unique characteristics suited for specific data behaviours. The Generalized Gamma distribution, a highly versatile three-parameter model, encompasses several well-known distributions as special cases, including the Weibull, Gamma, and Exponential distributions [15]. Its flexibility stems from its two shape parameters and one scale parameter, allowing it to model a wide range of monotonic and non-monotonic hazard functions, making it valuable in survival analysis and reliability engineering [5,8,12]. However, while powerful for various applications, the Generalized Gamma distribution may not always adequately capture data with extremely heavy tails or pronounced multimodality arising from heterogeneous underlying processes [4].

Conversely, the Lomax distribution, also known as the Pareto Type II distribution, is a two-parameter heavy-tailed distribution widely recognised for its ability to model extreme values and phenomena with power-law decay in its tail [15]. It finds extensive application in areas such as actuarial science, queueing theory, and income distribution, where the occurrence of rare but significant events is a critical characteristic [1,20]. Despite its strength in modelling heavy-tailed data, the standard Lomax distribution may lack the flexibility to accurately represent the central tendency or finer structural details of datasets that also contain a substantial component best described by a lighter-tailed or more adaptable distribution such as the Generalized Gamma. This limitation often necessitates the use of more complex models when dealing with mixed data behaviours [17,18].

The limitations of these individual distributions highlight a significant gap in the current statistical toolkit: the absence of a single, parsimonious model that can seamlessly integrate the flexibility of the Generalized Gamma distribution for varied shape and scale behaviours with the robust heavy-tail modelling capability of the Lomax distribution. Data from many real-world phenomena often manifest a blend of these characteristics, exhibiting a primary cluster of observations alongside a smaller, yet impactful, proportion of extreme values [3]. Existing mixture models sometimes address multimodality, but a specific combination of Generalized Gamma and Lomax components has the potential to offer a unique advantage by comprehensively characterising both the central dynamics and the extreme events within a single framework.

This paper introduces the Mixture of Generalized Gamma and Lomax (MiGGLo) distribution, a novel probabilistic model designed to bridge this gap. By amalgamating the Generalized Gamma and

Lomax distributions through a mixing parameter, the MiGGLo distribution provides a highly flexible framework capable of capturing the nuanced complexities of heterogeneous data that concurrently exhibit diverse tail behaviours and potential multimodality. The objective of this study is two-fold: first, to thoroughly investigate the parameter estimation of the MiGGLo distribution using Maximum Likelihood Estimation (MLE), Ordinary Least Squares (OLS), and Partial Expectation-Maximization (EM) algorithms through extensive Monte Carlo simulations; and second, to demonstrate its practical applicability by fitting the model to real-world epidemiological data, specifically focusing on the dengue effective reproductive number ($R_{\text{eff}}$). The subsequent sections of this paper are structured as follows: Sections 2 and 3 details the mathematical derivations and statistical properties of the MiGGLo distribution, respectively. Section 4 outlines the methodology for parameter estimation and the simulation study while in Section 5 presents the applicability of MiGGLo distribution in modeling dengue effective reproductive number in Caraga Region, Philippines. Section 6 presents the results from both the simulation study and the real-world data application. Finally, Section 7 provides the conclusion and recommendations for future research.

## 2. The MiGGLo Distribution

The Generalized Gamma (GG) distribution was originally introduced by [19] as a flexible extension of the standard Gamma distribution. This continuous probability model accommodates a wide range of distributional shapes through the inclusion of an additional parameter, thereby encompassing several well-known distributions as special cases. The probability density function (pdf) of the Generalized Gamma distribution is expressed as

$$f_g(x;\alpha_g,\tau,\theta_g)=\frac{\tau x^{\tau\alpha_g-1}e^{-\left(\frac{x}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}\Gamma(\alpha_g)},\quad x>0, \tag{1}$$

where $\alpha_g>0$, $\tau>0$, and $\theta_g>0$ denote the shape, family, and scale parameters, respectively. The corresponding cumulative distribution function (cdf) is given by

$$F_g(x;\alpha_g,\tau,\theta_g)=\Gamma\left(\alpha_g;\left(\frac{x}{\theta_g}\right)^{\tau}\right), \tag{2}$$

where $\Gamma(\cdot\,;\,\cdot)$ denotes the upper incomplete gamma function. The $k$-th raw moment of the Generalized Gamma distribution is expressed as

$$\mu'_{g_k}=\frac{\theta_g^k\Gamma\left(\alpha_g+\frac{k}{\tau}\right)}{\Gamma(\alpha_g)},\quad \text{for } k>-\alpha_g\tau. \tag{3}$$

The Lomax distribution, also known as the Pareto Type II distribution, is another heavy-tailed continuous probability distribution widely applied in fields such as reliability engineering, business risk modeling, and survival analysis [15]. Its probability density function is given by

$$f_l(x;\alpha_l,\theta_l)=\frac{\alpha_l\theta_l^{\alpha_l}}{(x+\theta_l)^{\alpha_l+1}},\quad x\geq 0, \tag{4}$$

where $\alpha_l > 0$ and $\theta_l > 0$ are the shape and scale parameters, respectively. The corresponding cumulative distribution function is defined as

$$F_l(x;\alpha_l,\theta_l) = 1 - \left(\frac{\theta_l}{x+\theta_l}\right)^{\alpha_l}. \tag{5}$$

The $k$-th raw moment of a Lomax-distributed random variable is given by

$$\mu'_{l_k} = \frac{\theta_l^k \Gamma(k+1)\Gamma(\alpha_l - k)}{\Gamma(\alpha_l)}, \quad -1 < k < \alpha_l. \tag{6}$$

2.1. **Definition.** Let $p$ denote the mixing proportion such that $0 \leq p \leq 1$. The Mixture of Generalized Gamma and Lomax (MiGGLo) distribution is defined as a two–component continuous mixture that integrates the Generalized Gamma (GG) and Lomax (Lo) distributions. This formulation provides a flexible model capable of capturing a wide range of data behaviors, including light-tailed and heavy-tailed phenomena. The probability density function (pdf) of the MiGGLo distribution is given by

$$f(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) = (1-p)f_g(x;\alpha_g,\tau,\theta_g) + pf_l(x;\alpha_l,\theta_l), \tag{7}$$

where $f_g(\cdot)$ and $f_l(\cdot)$ correspond to the pdfs of the Generalized Gamma and Lomax distributions, respectively. Substituting their explicit functional forms yields

$$f(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) = (1-p)\frac{\tau x^{\tau\alpha_g-1}e^{-\left(\frac{x}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}\Gamma(\alpha_g)} + p\frac{\alpha_l\theta_l^{\alpha_l}}{(x+\theta_l)^{\alpha_l+1}}, \tag{8}$$

for $x > 0$ and positive parameters $\alpha_g, \alpha_l, \tau, \theta_g, \theta_l$. The mixture parameter $p$ governs the relative contribution of each component, where $p = 0$ reduces the model to the Generalized Gamma distribution and $p = 1$ reduces it to the Lomax distribution.

2.2. **Cumulative Distribution Function (CDF).** The cumulative distribution function (cdf) of a continuous random variable $X$ with pdf $f(x)$ is defined as

$$F(x) = \int_{-\infty}^{x} f(t)\,dt. \tag{9}$$

Accordingly, the cdf of the MiGGLo distribution is expressed as

$$\begin{aligned} F(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) &= \int_{-\infty}^{x} f(t;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p)\,dt \\ &= (1-p)F_g(x;\alpha_g,\tau,\theta_g) + pF_l(x;\alpha_l,\theta_l), \end{aligned} \tag{10}$$

where $F_g(\cdot)$ and $F_l(\cdot)$ denote the cdfs of the Generalized Gamma and Lomax components, respectively. Substituting their explicit expressions yields

$$F(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) = (1-p)\Gamma\left(\alpha_g;\left(\frac{x}{\theta_g}\right)^{\tau}\right) + p\left[1-\left(\frac{\theta_l}{x+\theta_l}\right)^{\alpha_l}\right], \tag{11}$$

for $x \geq 0$. This mixture formulation ensures that the resulting model inherits both the flexibility of the Generalized Gamma distribution and the heavy-tailed characteristics of the Lomax distribution,

making it particularly suitable for modeling heterogeneous datasets exhibiting multimodality or tail asymmetry.

2.3. **Survival and Hazard Functions.** The survival function, also known as the reliability function, represents the probability that a random variable $X$ exceeds a particular value $x$. For a random variable $X$ following the MiGGLo distribution, the survival function is defined as

$$S(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) = 1 - F(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p), \tag{12}$$

where $F(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p)$ denotes the corresponding cumulative distribution function. Substituting from Equation (11), the explicit form of the survival function becomes

$$S(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) = 1 - \left\{(1-p)\Gamma\left(\alpha_g;\left(\frac{x}{\theta_g}\right)^{\tau}\right) + p\left[1-\left(\frac{\theta_l}{x+\theta_l}\right)^{\alpha_l}\right]\right\}, \tag{13}$$

for $x \geq 0$ and positive parameters $\alpha_g, \alpha_l, \tau, \theta_g, \theta_l$.

The hazard rate function, or failure rate function, describes the instantaneous rate of failure at time $x$, given that the event of interest has not yet occurred. It is defined as the ratio of the probability density function (pdf) to the survival function:

$$h(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) = \frac{f(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p)}{S(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p)}. \tag{14}$$

Substituting Equations (8) and (13) to Equation (14), the hazard rate function of the MiGGLo distribution can be expressed as

$$h(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p) = \frac{(1-p)\frac{\tau x^{\tau\alpha_g-1}e^{-\left(\frac{x}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}\Gamma(\alpha_g)} + p\frac{\alpha_l\theta_l^{\alpha_l}}{(x+\theta_l)^{\alpha_l+1}}}{1-\left\{(1-p)\Gamma\left(\alpha_g;\left(\frac{x}{\theta_g}\right)^{\tau}\right) + p\left[1-\left(\frac{\theta_l}{x+\theta_l}\right)^{\alpha_l}\right]\right\}}. \tag{15}$$

The functional form of $h(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p)$ enables the MiGGLo distribution to accommodate diverse reliability behaviors, including increasing, decreasing, or bathtub-shaped hazard rates, depending on the parameter configuration. This flexibility enhances its applicability in survival analysis, reliability modeling, and risk assessment contexts where heterogeneous or multimodal lifetimes are observed.

## 3. Statistical Properties of the MiGGLo Distribution

This section presents the key statistical characteristics of the MiGGLo distribution to elucidate its theoretical behavior and modeling flexibility. The $k$-th raw moments are derived, from which the mean, variance, skewness, and kurtosis are obtained to describe its central tendency, dispersion, and shape. Moreover, several special cases are identified by imposing specific parameter constraints, demonstrating that the MiGGLo distribution encompasses the Generalized Gamma, Lomax, and other related distributions as special forms.

3.1. **$k$-th Raw Moment and Central Moments.** The $k$-th raw moment of a random variable $X$ that follows the MiGGLo distribution provides a fundamental characterization of its distributional behavior. It is defined as

$$\begin{aligned}\mu'_k = E(X^k) &= \int_{-\infty}^{\infty} x^k f(x;\alpha_g,\alpha_l,\tau,\theta_g,\theta_l,p)\,dx \\ &= (1-p)\int_0^{\infty} x^k f_g(x;\alpha_g,\tau,\theta_g)\,dx + p\int_0^{\infty} x^k f_l(x;\alpha_l,\theta_l)\,dx \\ &= (1-p)\mu'_{g_k} + p\mu'_{l_k},\end{aligned}$$

where $\mu'_{g_k}$ and $\mu'_{l_k}$ denote the $k$-th raw moments of the Generalized Gamma and Lomax components, respectively. Substituting the corresponding expressions from Equations (3) and (6), the $k$-th raw moment of $X$ is obtained as

$$\mu'_k = (1-p)\frac{\theta_g^k\Gamma\left(\alpha_g+\frac{k}{\tau}\right)}{\Gamma(\alpha_g)} + p\frac{\theta_l^k\Gamma(k+1)\Gamma(\alpha_l-k)}{\Gamma(\alpha_l)}. \tag{16}$$

The first raw moment corresponds to the mean of the distribution, given by

$$\mu = \mu'_1 = (1-p)\frac{\theta_g\Gamma\left(\alpha_g+\frac{1}{\tau}\right)}{\Gamma(\alpha_g)} + p\frac{\theta_l}{\alpha_l-1}. \tag{17}$$

The variance is derived from the second central moment, $\sigma^2 = E[(X-\mu)^2] = \mu'_2 - \mu^2$, yielding

$$\sigma^2 = (1-p)\frac{\theta_g^2\Gamma\left(\alpha_g+\frac{2}{\tau}\right)}{\Gamma(\alpha_g)} + p\frac{\theta_l^2\Gamma(3)\Gamma(\alpha_l-2)}{\Gamma(\alpha_l)} - \mu^2. \tag{18}$$

The third and fourth central moments are used to describe the skewness and kurtosis of the MiGGLo distribution, which measure the asymmetry and tail heaviness, respectively. The skewness is defined as

$$\gamma_1 = \frac{E[(X-\mu)^3]}{\sigma^3} = \frac{\mu'_3 - 3\mu'_2\mu + 2\mu^3}{\sigma^3}, \tag{19}$$

where $\mu'_2 = (1-p)\frac{\theta_g^2\Gamma\left(\alpha_g+\frac{2}{\tau}\right)}{\Gamma(\alpha_g)} + p\frac{2\theta_l^2\Gamma(\alpha_l-2)}{\Gamma(\alpha_l)}$ and $\mu'_3 = (1-p)\frac{\theta_g^3\Gamma\left(\alpha_g+\frac{3}{\tau}\right)}{\Gamma(\alpha_g)} + p\frac{6\theta_l^3\Gamma(\alpha_l-3)}{\Gamma(\alpha_l)}$.

Similarly, the kurtosis, which quantifies the degree of peakedness or flatness of the distribution relative to the normal distribution, is given by

$$\gamma_2 = \frac{E[(X-\mu)^4]}{\sigma^4} = \frac{\mu'_4 - 4\mu'_3\mu + 6\mu'_2\mu^2 - 3\mu^4}{(\mu'_2-\mu^2)^2}, \tag{20}$$

where $\mu'_4 = (1-p)\frac{\theta_g^4\Gamma\left(\alpha_g+\frac{4}{\tau}\right)}{\Gamma(\alpha_g)} + p\frac{24\theta_l^4\Gamma(\alpha_l-4)}{\Gamma(\alpha_l)}$.

These moments provide a comprehensive summary of the MiGGLo distribution's behavior, revealing its ability to model a broad range of data patterns with varying degrees of skewness and tail thickness arising from the mixture of the Generalized Gamma and Lomax components.

3.2. **Special Cases.** The proposed Mixture of Generalized Gamma and Lomax (MiGGLo) distribution encompasses several well-known distributions as its special forms, depending on specific parameter configurations. These cases demonstrate the flexibility and generality of the MiGGLo framework in modeling diverse data behaviors.

(1) **Lomax and Generalized Gamma Distributions**

When the mixing proportion is set to $p = 1$, the MiGGLo distribution reduces entirely to the Lomax distribution, with probability density function (PDF)

$$f(x; \alpha_l, \theta_l) = \frac{\alpha_l \theta_l^{\alpha_l}}{(x + \theta_l)^{\alpha_l + 1}}, \quad x > 0.$$

Conversely, when $p = 0$, the MiGGLo distribution simplifies to the Generalized Gamma distribution, expressed as

$$f(x; \alpha_g, \tau, \theta_g) = \frac{\tau x^{\tau \alpha_g - 1} e^{-\left(\frac{x}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau} \Gamma(\alpha_g)}, \quad x > 0.$$

(2) **Mixture of Gamma and Lomax Distributions**

Setting the shape parameter $\tau = 1$ transforms the generalized gamma component into a standard gamma distribution, yielding the *Mixture of Gamma and Lomax* (*MiGaLo*) distribution:

$$f(x; \alpha_g, \alpha_l, \theta_g, \theta_l, p) = (1 - p) \frac{x^{\alpha_g - 1} e^{-\frac{x}{\theta_g}}}{\theta_g^{\alpha_g} \Gamma(\alpha_g)} + p \frac{\alpha_l \theta_l^{\alpha_l}}{(x + \theta_l)^{\alpha_l + 1}}, \quad x > 0.$$

(3) **Mixture of Weibull and Lomax Distributions**

By setting $\alpha_g = 1$, the generalized gamma component reduces to a Weibull distribution. Thus, the MiGGLo distribution becomes a *Mixture of Weibull and Lomax* (*MiWeiLo*) distribution, defined as

$$f(x; \tau, \alpha_l, \theta_g, \theta_l, p) = (1 - p) \frac{\tau x^{\tau - 1} e^{-\left(\frac{x}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}} + p \frac{\alpha_l \theta_l^{\alpha_l}}{(x + \theta_l)^{\alpha_l + 1}}, \quad x > 0.$$

(4) **Mixture of Exponential and Lomax Distributions**

Further, when both $\alpha_g = 1$ and $\tau = 1$, the generalized gamma component reduces to an exponential distribution. In this case, the MiGGLo distribution becomes a *Mixture of Exponential and Lomax* (*MiExLo*) distribution, given by

$$f(x; \alpha_l, \theta_g, \theta_l, p) = (1 - p) \frac{1}{\theta_g} e^{-\frac{x}{\theta_g}} + p \frac{\alpha_l \theta_l^{\alpha_l}}{(x + \theta_l)^{\alpha_l + 1}}, \quad x > 0.$$

These special cases underscore the hierarchical structure and adaptability of the MiGGLo distribution. By encompassing the Generalized Gamma, Weibull, Gamma, Exponential, and Lomax models as limiting forms, the MiGGLo serves as a unifying probabilistic framework capable of modeling both light-tailed and heavy-tailed phenomena across a broad spectrum of applications.

## 4. Parameter Estimation Methods

In this section, we present the parameter estimation methods employed for the MiGGLo distribution. Specifically, the maximum likelihood estimation (MLE), ordinary least squares (OLS), and partial expectation–maximization (EM) algorithm are discussed in detail. Furthermore, a Monte Carlo simulation study is conducted to evaluate and compare the consistency, efficiency, and robustness of these estimation techniques under varying parameter settings.

### 4.1. Maximum Likelihood Estimation (MLE).

The maximum likelihood estimation (MLE) method is one of the most widely used techniques for estimating the parameters of probability distributions due to its desirable asymptotic properties, such as consistency, efficiency, and invariance. For a random sample $x_1, x_2, \ldots, x_n$ drawn from the MiGGLo distribution, the likelihood function is defined as

$$l(\alpha_g, \alpha_l, \tau, \theta_g, \theta_l, p) = \prod_{i=1}^{n} \left[ (1-p) \frac{\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}\Gamma(\alpha_g)} + p \frac{\alpha_l \theta_l^{\alpha_l}}{(x_i + \theta_l)^{\alpha_l + 1}} \right].$$

Taking the natural logarithm of the likelihood function, the corresponding log-likelihood function is expressed as

$$L(\alpha_g, \alpha_l, \tau, \theta_g, \theta_l, p) = \sum_{i=1}^{n} \ln \left[ (1-p) \frac{\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}\Gamma(\alpha_g)} + p \frac{\alpha_l \theta_l^{\alpha_l}}{(x_i + \theta_l)^{\alpha_l + 1}} \right]. \tag{21}$$

To obtain the maximum likelihood estimates of the parameters $\alpha_g, \alpha_l, \tau, \theta_g, \theta_l$, and $p$, the partial derivatives of Equation (21) with respect to each parameter are taken and equated to zero, resulting in the following system of nonlinear equations:

$$\begin{cases}
\frac{\partial L}{\partial \alpha_g} = \sum_{i=1}^{n} \frac{\frac{(1-p)\tau^2 x_i^{\tau\alpha_g - 1} \ln(x_i) e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}\Gamma(\alpha_g)} - \frac{(1-p)\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} \Psi(\alpha_g)}{\theta_g^{\tau}\Gamma(\alpha_g)}}{\frac{(1-p)\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}}}{\theta_g^{\tau}\Gamma(\alpha_g)} + \frac{p\alpha_l \theta_l^{\alpha_l}}{(x_i+\theta_l)^{\alpha_l+1}}} = 0 \\
\frac{\partial L}{\partial \alpha_l} = \sum_{i=1}^{n} \frac{p\theta_l^{\alpha_l} \left[\alpha_l \ln(\theta_l) - \alpha_l \ln(x_i + \theta_l) + 1\right] \theta_g^{\tau}\Gamma(\alpha_g)}{p\alpha_l \theta_l^{\alpha_l} \theta_g^{\tau}\Gamma(\alpha_g) - (1-p)\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} (x_i + \theta_l)^{\alpha_l + 1}} = 0 \\
\frac{\partial L}{\partial \tau} = -\sum_{i=1}^{n} \frac{(1-p)\, x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} \left(\ln(x_i)\tau\alpha_g - \tau \left(\frac{x_i}{\theta_g}\right)^{\tau} \ln\left(\frac{x_i}{\theta_g}\right) - \tau \ln(\theta_g) + 1\right) (x_i + \theta_l)^{\alpha_l + 1}}{p\alpha_l \theta_l^{\alpha_l} \theta_g^{\tau}\Gamma(\alpha_g) - (1-p)\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} (x_i + \theta_l)^{\alpha_l + 1}} = 0 \\
\frac{\partial L}{\partial \theta_g} = -\sum_{i=1}^{n} \frac{(1-p)\tau^2 x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} \left[\left(\frac{x_i}{\theta_g}\right)^{\tau} - 1\right] (x_i + \theta_l)^{\alpha_l + 1}}{\theta_g \left[p\alpha_l \theta_l^{\alpha_l} \theta_g^{\tau}\Gamma(\alpha_g) - (1-p)\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} (x_i + \theta_l)^{\alpha_l + 1}\right]} = 0 \\
\frac{\partial L}{\partial \theta_l} = \sum_{i=1}^{n} \frac{p\alpha_l \theta_l^{\alpha_l} (\alpha_l x_i - \theta_l) \theta_g^{\tau}\Gamma(\alpha_g)}{\theta_l(x_i + \theta_l) \left[p\alpha_l \theta_l^{\alpha_l} \theta_g^{\tau}\Gamma(\alpha_g) - (1-p)\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} (x_i + \theta_l)^{\alpha_l + 1}\right]} = 0 \\
\frac{\partial L}{\partial p} = \sum_{i=1}^{n} \frac{\theta_g^{\tau}\Gamma(\alpha_g)\theta_l^{\alpha_l}\alpha_l - \tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} (x_i + \theta_l)^{\alpha_l + 1}}{p\alpha_l \theta_l^{\alpha_l} \theta_g^{\tau}\Gamma(\alpha_g) - (1-p)\tau x_i^{\tau\alpha_g - 1} e^{-\left(\frac{x_i}{\theta_g}\right)^{\tau}} (x_i + \theta_l)^{\alpha_l + 1}} = 0,
\end{cases} \tag{22}$$

where $\Psi(\cdot)$ denotes the digamma function.

Since these likelihood equations are highly nonlinear and do not admit a closed-form analytical solution, numerical optimization techniques are required to obtain the parameter estimates. In this study, numerical maximization of the log-likelihood function is performed using the `optimize.minimize()` function in Python, which implements quasi-Newton and derivative-based algorithms suitable for high-dimensional likelihood surfaces. The resulting estimates serve as the MLEs of the MiGGLo parameters, which can subsequently be used for inferential and simulation purposes.

4.2. **Ordinary Least Squares (OLS).** The Ordinary Least Squares (OLS) method provides an alternative approach to estimating the parameters of the MiGGLo distribution by minimizing the discrepancy between the theoretical and empirical cumulative distribution functions (CDFs). Specifically, the OLS estimator minimizes the sum of squared deviations between the empirical CDF, denoted by $\hat{F}(x_i)$, and the theoretical CDF $F(x_i;\boldsymbol{\Theta})$, where $\boldsymbol{\Theta} = (\alpha_g, \alpha_l, \tau, \theta_g, \theta_l, p)$ represents the parameter vector of the MiGGLo distribution. The OLS objective function is given by

$$Q(\boldsymbol{\Theta}) = \sum_{i=1}^{n} \left[\hat{F}(x_i) - F(x_i;\boldsymbol{\Theta})\right]^2. \tag{23}$$

Minimizing $Q(\boldsymbol{\Theta})$ yields parameter estimates that best align the theoretical distribution with the observed data in a least-squares sense. Compared to the maximum likelihood estimation, the OLS approach is computationally simpler and less sensitive to numerical instabilities in the likelihood function, particularly for mixture distributions such as the MiGGLo. However, it may exhibit reduced statistical efficiency in small samples. In this study, the OLS method serves as a complementary estimation technique, providing an additional benchmark for evaluating the performance of the MLE and EM algorithms through Monte Carlo simulation.

4.3. **Partial Expectation–Maximization (EM) Algorithm.** The Expectation–Maximization (EM) algorithm is a well-established iterative framework for parameter estimation in mixture models, particularly when direct maximization of the likelihood function is analytically intractable. In the context of the Mixture of Generalized Gamma and Lomax (MiGGLo) distribution, the EM algorithm is adapted as a *Partial EM Algorithm* to enhance computational stability and convergence efficiency. Specifically, only the mixing proportion parameter is iteratively updated, whereas the component parameters are fixed at their Ordinary Least Squares (OLS) estimates obtained from preliminary estimation.

Let $X = \{x_1, x_2, \ldots, x_n\}$ denote a random sample from the MiGGLo distribution, and let $Z_i$ represent a latent indicator variable such that $Z_i = 1$ if $x_i$ originates from the Lomax component and $Z_i = 0$ otherwise. The complete-data likelihood function can thus be expressed as

$$L_c(\boldsymbol{\Theta} \mid X, Z) = \prod_{i=1}^{n} \left[(1-p) f_g(x_i;\alpha_g,\theta_g,\tau)\right]^{1-Z_i} \left[p f_l(x_i;\alpha_l,\theta_l)\right]^{Z_i}, \tag{24}$$

where $f_g(x;\alpha_g,\theta_g,\tau)$ and $f_l(x;\alpha_l,\theta_l)$ denote the Generalized Gamma and Lomax probability density functions defined in Equations (2) and (4), respectively. The full parameter vector is given by $\boldsymbol{\Theta} = (\alpha_g,\theta_g,\tau,\alpha_l,\theta_l,p)$.

The algorithm alternates between two principal steps until convergence is achieved. In the **E-step**, the conditional expectation of the latent variable $Z_i$ is computed using the current parameter estimates. This yields the posterior probability that observation $x_i$ arises from the Lomax component:

$$\gamma_i^{(t)} = \frac{p^{(t)} f_l(x_i;\alpha_l,\theta_l)}{(1-p^{(t)}) f_g(x_i;\alpha_g,\theta_g,\tau) + p^{(t)} f_l(x_i;\alpha_l,\theta_l)}, \tag{25}$$

where the superscript $(t)$ denotes the iteration index.

In the **M-step**, only the mixing proportion $p$ is updated, while the remaining parameters $(\alpha_g,\theta_g,\tau,\alpha_l,\theta_l)$ are held fixed at their corresponding OLS estimates, denoted by $(\hat{\alpha}_g,\hat{\theta}_g,\hat{\tau},\hat{\alpha}_l,\hat{\theta}_l)$. The updated estimate of the mixing proportion is then computed as

$$p^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} \gamma_i^{(t)}, \tag{26}$$

which ensures that $0 < p^{(t+1)} < 1$. The iterative procedure terminates when the relative change in $p$ satisfies the convergence criterion

$$\left| p^{(t+1)} - p^{(t)} \right| < \varepsilon, \tag{27}$$

where $\varepsilon$ is a pre-specified small tolerance level.

By adopting this partial EM approach, the estimation process achieves numerical stability while maintaining the efficiency of parameter refinement. This design prevents divergence issues that commonly arise when estimating all parameters simultaneously in highly nonlinear likelihood surfaces. Consequently, the Partial EM algorithm provides reliable and consistent estimates of the mixing proportion for the MiGGLo model, particularly when validated through extensive Monte Carlo simulations.

4.4. **Monte Carlo Simulation Study.** To evaluate the performance and consistency of the proposed parameter estimation methods for the Mixture of Generalized Gamma and Lomax (MiGGLo) distribution, a comprehensive Monte Carlo simulation study was conducted. The primary objective of this simulation is to assess and compare the statistical accuracy, efficiency, and computational stability of the Maximum Likelihood Estimation (MLE), Ordinary Least Squares (OLS), and Expectation–Maximization (EM) algorithms under various sample sizes.

A series of random samples of sizes $n = 100, 500, 1000, 3000, 5000$ were generated independently from the MiGGLo distribution using the true parameter vector

$$\boldsymbol{\Theta}_0 = (\alpha_g,\theta_g,\tau,\alpha_l,\theta_l,p) = (2.0,\ 1.5,\ 1.2,\ 1.2,\ 1.0,\ 0.3),$$

where $(1-p)$ governs the contribution from the Generalized Gamma component and $p$ from the Lomax component. For each sample size, $10{,}000$ replications were performed to obtain robust estimates of the average parameter values, mean squared errors (MSE), and biases for each estimation technique.

The MLE method was implemented through numerical optimization of the negative log-likelihood function using the L-BFGS-B algorithm within the `scipy.optimize.minimize` routine. The OLS method estimated parameters by minimizing the sum of squared deviations between the empirical cumulative distribution function (ECDF) and the theoretical CDF of the MiGGLo model. Finally, the EM algorithm refined the OLS-based initial estimates through iterative expectation and maximization steps, updating the posterior weights and mixture proportion $p$ until convergence.

For each replication and estimation method, the average estimates $\bar{\boldsymbol{\Theta}}$, mean squared error (MSE), and bias were computed as

$$\text{MSE}(\hat{\theta}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta_0)^2, \qquad \text{Bias}(\hat{\theta}) = \frac{1}{R}\sum_{r=1}^{R}(\hat{\theta}_r - \theta_0),$$

where $R = 10{,}000$ denotes the number of replications, $\hat{\theta}_r$ is the $r$-th estimate of the parameter $\theta \in \Theta$, and $\theta_0$ is the corresponding true parameter value. In addition, the computational time for each estimation method was recorded to assess relative efficiency.

The results were summarized in tabular form and visually examined through bias and MSE plots across sample sizes. The simulation findings will reveal the convergence behavior of each method as the sample size increases, providing empirical evidence on the consistency and relative accuracy of the MLE, OLS, and EM estimators for the MiGGLo distribution. The Monte Carlo framework offers a rigorous numerical validation of the parameter estimation procedures and demonstrates the robustness of the proposed methods under different data-generating conditions.

Tables 1–9 summarize the performance of the Maximum Likelihood Estimation (MLE), Ordinary Least Squares (OLS), and Expectation–Maximization (EM) methods. The MLE consistently produced the most accurate and stable parameter estimates, with bias and mean squared error (MSE) decreasing sharply as sample size increased, confirming its asymptotic efficiency. In contrast, OLS and EM yielded higher bias and MSE for small samples, though both improved with larger $n$. While EM achieved better precision than OLS at high $n$, it required substantially longer computation times. Overall, the MLE demonstrated superior reliability and computational efficiency, providing the most robust parameter recovery for the proposed mixture model.

TABLE 1. Average Maximum Likelihood Estimates (MLE) of Parameters

| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ | Time (sec) |
|---|---|---|---|---|---|---|---|
| 100 | 4.95549 | 89.3317 | 472.979 | 1166.43 | 2787.95 | 0.275437 | 443.196 |
| 500 | 3.26976 | 1.75722 | 1.44081 | 3.42937 | 24.4141 | 0.301392 | 408.911 |
| 1000 | 2.64966 | 1.60480 | 1.31757 | 1.45217 | 3.29997 | 0.304246 | 381.135 |
| 3000 | 2.16619 | 1.52663 | 1.21661 | 1.23915 | 1.08675 | 0.306004 | 371.759 |
| 5000 | 2.10233 | 1.50761 | 1.20671 | 1.22541 | 1.05857 | 0.309241 | 434.834 |

TABLE 2. Mean Squared Error (MSE) of MLE Parameter Estimates

| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ |
|---|---|---|---|---|---|---|
| 100 | 92.1512 | 5.28778e+07 | 1.64259e+09 | 1.06183e+10 | 4.09283e+10 | 0.0395652 |
| 500 | 18.6413 | 1.66005 | 3.72573 | 667.759 | 224415 | 0.0122254 |
| 1000 | 6.27147 | 0.969944 | 6.60668 | 31.2273 | 6775.86 | 0.00570145 |
| 3000 | 0.815959 | 0.325756 | 0.0495992 | 0.0382777 | 0.127263 | 0.00223302 |
| 5000 | 0.358811 | 0.192916 | 0.0248899 | 0.0201176 | 0.059526 | 0.00212509 |

TABLE 3. Bias of MLE Parameter Estimates

| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ |
|---|---|---|---|---|---|---|
| 100 | 2.95549 | 87.8317 | 471.779 | 1165.23 | 2786.95 | -0.024563 |
| 500 | 1.26976 | 0.25722 | 0.24081 | 2.22937 | 23.4141 | 0.001392 |
| 1000 | 0.64966 | 0.10480 | 0.11758 | 0.25217 | 2.29997 | 0.004246 |
| 3000 | 0.16619 | 0.02663 | 0.01661 | 0.03915 | 0.08675 | 0.006004 |
| 5000 | 0.10234 | 0.00761 | 0.00671 | 0.02541 | 0.05857 | 0.009241 |

TABLE 4. Average Ordinary Least Squares (OLS) Estimates of Parameters

| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ | Time (sec) |
|---|---|---|---|---|---|---|---|
| 100 | 10.3589 | 1.71584 | 12.6415 | 56.9327 | 75.0406 | 0.364079 | 1394.68 |
| 500 | 5.86601 | 1.61730 | 1.74728 | 49.6542 | 61.1774 | 0.347119 | 2632.96 |
| 1000 | 4.63760 | 1.57661 | 1.33262 | 31.5645 | 42.7905 | 0.350544 | 3665.68 |
| 3000 | 3.23186 | 1.48879 | 1.21019 | 5.18949 | 9.56764 | 0.330542 | 6417.46 |
| 5000 | 2.75331 | 1.46337 | 1.19436 | 2.12430 | 3.08468 | 0.322058 | 8426.86 |

TABLE 5. Mean Squared Error (MSE) of OLS Parameter Estimates

| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ |
|---|---|---|---|---|---|---|
| 100 | 494.841 | 2.95772 | 5268.96 | 87176.2 | 52571.5 | 0.0711151 |
| 500 | 75.2203 | 2.04279 | 90.3189 | 67792.6 | 44932.3 | 0.0452763 |
| 1000 | 41.3591 | 1.54934 | 0.679436 | 51930.8 | 25657.0 | 0.0367024 |
| 3000 | 13.1258 | 0.843574 | 0.155076 | 1346.87 | 4510.02 | 0.0192134 |
| 5000 | 6.05886 | 0.594871 | 0.087717 | 106.606 | 576.01 | 0.0125738 |

TABLE 6. Bias of OLS Parameter Estimates

| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ |
|---|---|---|---|---|---|---|
| 100 | 8.35891 | 0.215842 | 11.4415 | 55.7327 | 74.0406 | 0.064079 |
| 500 | 3.86601 | 0.117300 | 0.547283 | 48.4542 | 60.1774 | 0.047119 |
| 1000 | 2.63760 | 0.076611 | 0.132618 | 30.3645 | 41.7905 | 0.050544 |
| 3000 | 1.23186 | -0.011211 | 0.010190 | 3.98949 | 8.56764 | 0.030542 |
| 5000 | 0.753315 | -0.036626 | -0.005643 | 0.924299 | 2.08468 | 0.022058 |

TABLE 7. Partial EM Estimates — Parameter Averages

| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ | Time (sec) |
|---|---|---|---|---|---|---|---|
| 100 | 10.4027 | 1.68935 | 12.302 | 51.5922 | 71.0318 | 0.314224 | 1410.55 |
| 500 | 5.89018 | 1.63349 | 1.74408 | 46.3259 | 64.2155 | 0.341894 | 2610.44 |
| 1000 | 4.72463 | 1.56087 | 1.34137 | 26.4966 | 41.4412 | 0.346332 | 3650.26 |
| 3000 | 3.27330 | 1.46978 | 1.20287 | 4.98078 | 7.93113 | 0.331319 | 6497.09 |
| 5000 | 2.68632 | 1.48534 | 1.20062 | 2.07001 | 2.98436 | 0.318146 | 8302.65 |

TABLE 8. Partial EM Estimates — Mean Squared Errors (MSE)

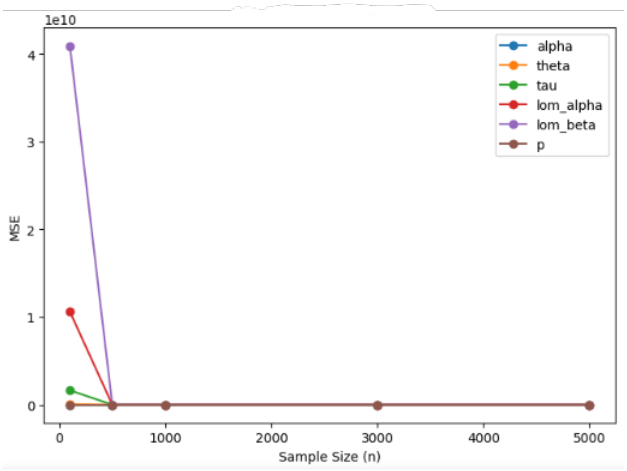
| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ |
|---|---|---|---|---|---|---|
| 100 | 435.237 | 2.87922 | 5017.91 | 82655.1 | 45176.0 | 0.0564066 |
| 500 | 78.5331 | 2.07117 | 72.3497 | 49148.6 | 43204.5 | 0.0428069 |
| 1000 | 43.0940 | 1.55273 | 6.96979 | 27859.6 | 26308.4 | 0.0339802 |
| 3000 | 13.2719 | 0.85285 | 0.15352 | 2737.24 | 2833.75 | 0.0185690 |
| 5000 | 5.50637 | 0.58994 | 0.08516 | 140.153 | 771.318 | 0.0116995 |

TABLE 9. Partial EM Estimates — Biases

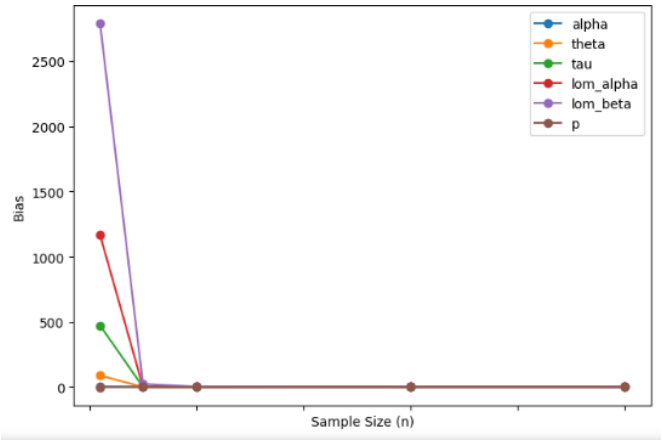
| $n$ | $\alpha_g$ | $\theta_g$ | $\tau$ | $\alpha_l$ | $\theta_l$ | $p$ |
|---|---|---|---|---|---|---|
| 100 | 8.40273 | 0.18935 | 11.1020 | 50.3922 | 70.0318 | 0.01422 |
| 500 | 3.89018 | 0.13349 | 0.54408 | 45.1259 | 63.2155 | 0.04189 |
| 1000 | 2.72463 | 0.06087 | 0.14137 | 25.2966 | 40.4412 | 0.04633 |
| 3000 | 1.27330 | -0.03022 | 0.00287 | 3.78078 | 6.93113 | 0.03132 |
| 5000 | 0.68632 | -0.01466 | 0.00062 | 0.87001 | 1.98436 | 0.01815 |

As shown in Figures 1a and 1b, the MSE and bias values under the MLE approach demonstrate a consistent decline as the sample size increases. For smaller samples ($n = 100$), the parameters associated with the Lomax component ($\alpha_L$ and $\theta_L$) exhibit relatively higher MSE and bias, suggesting sensitivity to tail behavior and sample variability. However, for larger samples ($n \geq 1000$), both error measures substantially decrease, converging toward zero. This outcome indicates the consistency and asymptotic efficiency of the MLE estimator for the MiGGLo distribution. At $n = 5000$, all parameters yield stable estimates with minimal MSE, confirming the robustness of MLE for large-sample estimation. Figures 1c and 1d display the MSE and bias for the OLS method. The results reveal a downward trend in both measures with increasing sample size, similar to the MLE pattern. However, OLS consistently produces higher MSE values than MLE, particularly for the Lomax-related parameters. This finding suggests that OLS is less effective when modeling heavy-tailed distributions. Although estimation accuracy improves with larger samples, convergence remains slower than that observed for MLE. By $n = 5000$, OLS achieves reasonably precise estimates, but they remain slightly less accurate compared to MLE results. Figures 1e and 1f illustrate the MSE and bias patterns obtained under the partial EM algorithm. The results indicate that EM exhibits similar performance behavior to OLS, with both metrics decreasing as sample size increases. The partial EM algorithm shows moderate bias and MSE at smaller samples ($n = 100$), which gradually diminish as $n$ increases. The iterative nature of partial EM, coupled with its dependency on initialization values, contributes to slower convergence in small samples. Nevertheless, as $n$ approaches 5000, the estimates become nearly unbiased, reflecting the stability of partial EM in larger datasets. Figure 2 presents the average computation time (in seconds) required for parameter estimation using the three methods. The MLE method consistently demonstrates superior computational efficiency, completing estimations considerably faster than OLS and partial EM across all sample sizes. Both OLS and partial EM exhibit exponential increases in computation time beyond $n = 1000$, exceeding 8000 seconds at $n = 5000$, while MLE remains below 500 seconds. This result highlights a clear trade-off between computational cost and estimation precision, where

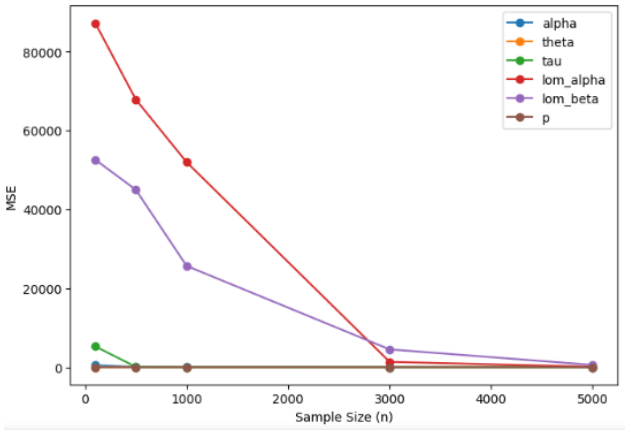
MLE achieves both efficiency and accuracy, whereas OLS and EM require greater time resources for comparable precision.



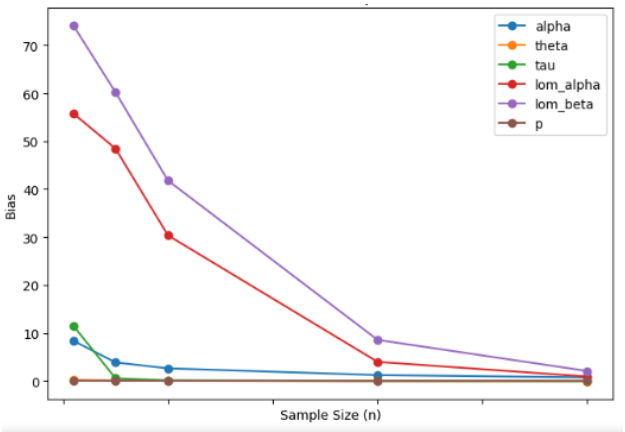

(A) MSE under MLE Method



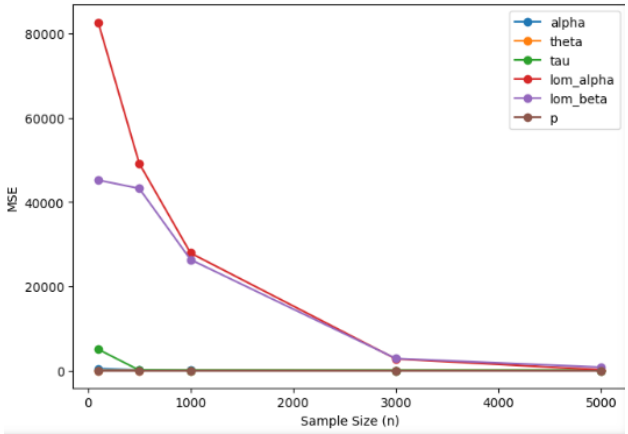

(B) Bias under MLE Method



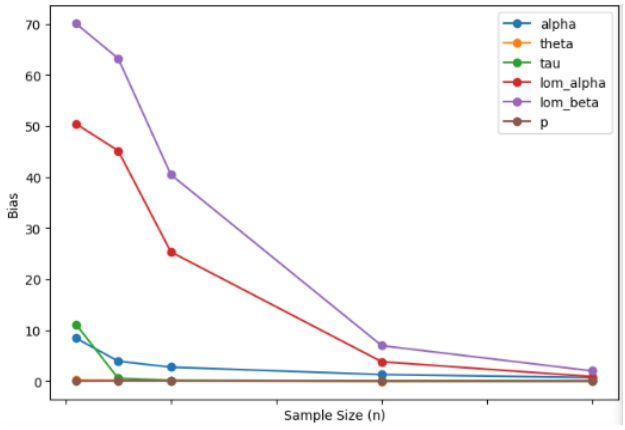

(C) MSE under OLS Method




(D) Bias under OLS Method




(E) MSE under Partial EM Algorithm




(F) Bias under Partial EM Algorithm

FIGURE 1. Comparison of MSE and Bias under MLE, OLS, and Partial EM Methods

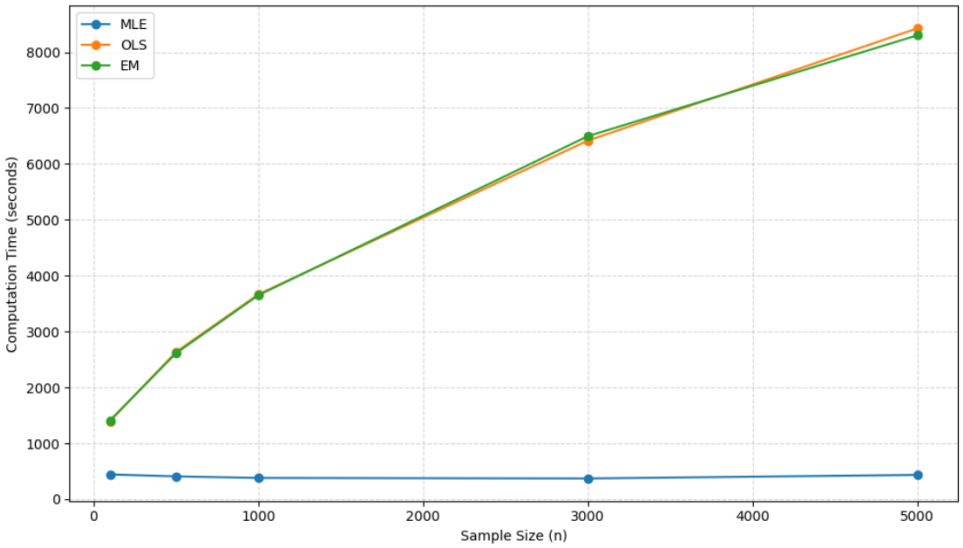



FIGURE 2. Average Time Required (in seconds) to Estimate the parameters of the MiGGLo distribution using the Maximum Likelihood Estimation (MLE), Ordinary Least Squares (OLS), and Expectation–Maximization (EM) methods

## 5. Fitting the Dengue Effective Reproductive Number $R_{\text{EFF}}$

This section demonstrates the practical applicability of the Mixture of Generalized Gamma and Lomax (MiGGLo) distribution by fitting it to real-world epidemiological data. Specifically, the dengue effective reproductive number ($R_{\text{eff}}$) for the Caraga Region, Philippines, is modeled to examine the capability of the MiGGLo distribution in capturing complex, heavy-tailed, and asymmetric patterns inherent in infectious disease dynamics. The subsequent subsections present the dataset utilized in this analysis and the procedures employed for parameter estimation and model evaluation.

5.1. **Data Description.** The data employed in this study consist of daily dengue incidence in the Caraga Region, Philippines, spanning the period from January 2015 to December 2020. These data were utilized to compute the dengue effective reproductive number ($R_{\text{eff}}$), which quantifies the average number of secondary infections generated by an infectious individual. The computation of $R_{\text{eff}}$ follows the SIR-UV modeling framework described in [16], which accounts for environmental factors influencing disease transmission dynamics. Descriptive statistics of the computed $R_{\text{eff}}$ values, including measures of central tendency and dispersion, are summarized and illustrated in Table 10, as adapted from [7]. These descriptive summaries provide essential insights into the empirical characteristics of the data prior to model fitting.

TABLE 10. Descriptive Statistic of the Dengue $R_{\text{eff}}$, Precipitation and Temperature in Caraga Region January 2015 to December 2020

| Variables | Mean | SD | Minimum | Maximum | Skewness | Kurtosis |
|---|---|---|---|---|---|---|
| $R_{\text{eff}}$ | 0.964 | 0.357 | 0.00 | 1.95 | -0.381 | 0.0889 |

*Note.* Summary statistics for the daily dengue effective reproductive number ($R_{\text{eff}}$) computed from dengue incidence in the Caraga Region for the period January 2015 – December 2020.

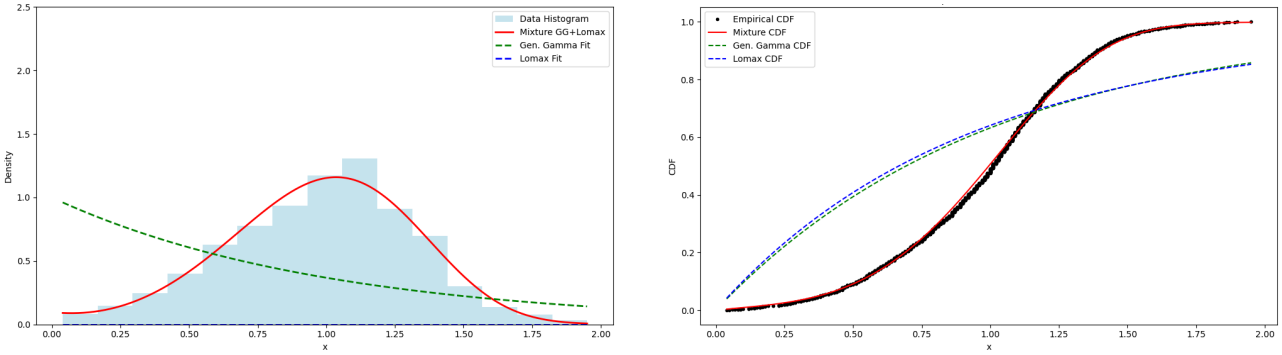
**Source:** Adapted from [7].

5.2. **Model Fitting and Selection.** To evaluate the suitability of candidate distributions for modeling the bladder cancer remission times dataset, the parameters of each model were estimated using the Maximum Likelihood Estimation (MLE) method. MLE identifies parameter values that maximize the likelihood function, thereby producing estimators with desirable asymptotic properties such as consistency and efficiency. The optimization procedure was performed using the `scipy.optimize.minimize` function in Python, which implements numerical maximization of the log-likelihood function under convergence tolerance criteria.

The quality of model fit was subsequently assessed using the root mean square error (RMSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). Lower values of AIC and BIC indicate a better balance between model goodness-of-fit and parsimony. Furthermore, the Kolmogorov–Smirnov (K–S) test was employed to evaluate the null hypothesis that the observed data follow the fitted distribution, where higher $p$-values suggest a stronger fit to the empirical data.

Table 11 presents the estimated parameters, log-likelihood, AIC, BIC, and K–S test results for the fitted distributions. Based on these criteria, the distribution yielding the minimum AIC and BIC values, along with a non-significant K–S test ($p > 0.01$), was identified as the most appropriate model for describing the data.

FIGURE 3. Graphs of Fitted PDF and CDF




(A) Fitted Densities

(B) Fitted Cumulative

TABLE 11. Goodness-of-Fit Results for Different Distributions

| Distribution | Parameter Estimates | KS Statistic | p-value | RMSE | AIC / BIC |
|---|---|---|---|---|---|
| MiGGLo | $\alpha_g = 0.7609$<br>$\theta_g = 1.2119$<br>$\tau = 4.1938$<br>$\alpha_l = 19.328$<br>$\theta_l = 10.0$<br>$p = 0.05$ | 0.0323 | 0.0221 | 0.0622 | 1416.02 / 1450.07 |
| General Gamma | $\alpha_g = 1$<br>$\theta_g = 1$<br>$\tau = 1$ | 0.3116 | $3.5119 \times 10^{-186}$ | 0.5119 | 4229.50 / 4246.52 |
| Lomax | $\alpha_l = 10.7419$<br>$\theta_l = 10$ | 0.3259 | $4.31 \times 10^{-204}$ | 0.6652 | 4402.69 / 4414.04 |

## 6. RESULTS AND DISCUSSION

6.1. **Simulation Study Results.** The results in Tables 1–3 demonstrate that the MLE exhibited superior accuracy and stability across all sample sizes. The parameter estimates converged rapidly to their true values as $n$ increased, with substantial reductions in both bias and MSE. Specifically, the MSE values for all parameters decreased sharply from small to large samples, indicating the consistency and asymptotic efficiency of the MLE. For instance, the MSE of $\alpha_g$ decreased from 92.15 at $n = 100$ to 0.36 at $n = 5000$, while the bias fell from 2.96 to 0.10 over the same range. Similarly, the estimated mixing proportion $p$ stabilized near 0.30 with minimal deviation and negligible bias. Although computation time ranged from 371 to 443 seconds, the MLE remained computationally efficient compared to iterative methods. These findings confirm that MLE offers the most robust estimation performance and convergence stability for the MiGGLo Model.

The OLS results, summarized in Tables 4–6, exhibited lower estimation precision relative to MLE. The OLS estimates showed noticeable deviations from the true parameter values at smaller sample sizes, particularly for the Lomax shape and scale parameters $(\alpha_l, \theta_l)$. Although both bias and MSE decreased with increasing $n$, their magnitudes remained consistently higher than those of MLE. For instance, at $n = 100$, the MSE of $\alpha_g$ and $\alpha_l$ reached 494.84 and 87,176.20, respectively, reflecting poor stability and high sampling variability. Even at $n = 5000$, OLS estimates remained slightly biased, particularly for the mixture proportion parameter $p$. Moreover, the computational time required by OLS was substantially greater, increasing from 1,394.68 seconds at $n = 100$ to 8,426.86 seconds at $n = 5000$. These results suggest that while OLS may achieve acceptable convergence with very large samples, it is less efficient and computationally burdensome for complex mixture models such as the MiGGLo distribution.

The partial EM estimation results (Tables 7–9) demonstrated a performance intermediate between MLE and OLS. Similar to MLE, partial EM estimates improved with increasing sample size, showing reduced bias and MSE across all parameters. For example, the MSE of $\alpha_g$ declined from 435.24 at $n = 100$ to 5.51 at $n = 5000$, while the bias decreased from 8.40 to 0.69. However, partial EM required substantially longer computation times, with processing duration exceeding 8,000 seconds at $n = 5000$, indicating slower convergence. The partial EM algorithm also exhibited a mild sensitivity to initialization values, occasionally converging to local maxima in small samples. Despite these limitations, partial EM achieved relatively smoother convergence trends than OLS and offered reasonable precision for large $n$.

A comparative examination of the three estimation techniques reveals that MLE consistently outperformed OLS and partial EM in both statistical efficiency and computational economy. Across all sample sizes, MLE yielded the smallest MSE and bias, confirming its robustness for estimating the parameters of the MiGGLo distribution. The OLS method exhibited significant estimation variability and required excessive computation time, while the partial EM algorithm, although more precise than OLS, incurred a high computational cost due to its iterative nature. Taken together, these findings affirm that MLE provides the most efficient, accurate, and computationally feasible approach for parameter estimation in the MiGGLo framework, making it the preferred method for practical implementation and inferential analysis.

6.2. **Application Results.** The comparative results summarized in Table 11 demonstrate substantial differences in the performance of the three fitted distributions. The Mixture of Generalized Gamma and Lomax (MiGGLo) model exhibited the most favorable goodness-of-fit indices, with a Kolmogorov–Smirnov (K–S) statistic of $0.0323$ and an associated $p$-value of $0.0221$. Since the $p$-value exceeds the $0.01$ significance threshold, the null hypothesis that the data follow the MiGGLo distribution cannot be rejected, thereby confirming that the MiGGLo model provides an adequate and superior representation of the empirical distribution compared to the competing models. In contrast, the General Gamma and Lomax models yielded extremely small $p$-values ($3.51 \times 10^{-186}$ and $4.31 \times 10^{-204}$, respectively), indicating a significant deviation from the observed data.

In terms of error-based metrics, the MiGGLo distribution achieved the lowest root mean square error (RMSE = 0.0622), implying a closer alignment between the predicted and observed values compared to the General Gamma (RMSE = 0.5119) and Lomax (RMSE = 0.6652) distributions. Furthermore, the information-theoretic criteria reinforce this conclusion, as the MiGGLo model attained substantially smaller AIC (1416.02) and BIC (1450.07) values relative to the General Gamma (AIC = 4229.50; BIC = 4246.52) and Lomax (AIC = 4402.69; BIC = 4414.04) alternatives. Since both AIC and BIC penalize model complexity, these results indicate that the MiGGLo distribution attains an optimal balance between parsimony and goodness-of-fit.

The Generalized Gamma component ($\alpha_g = 0.7609$, $\theta_g = 1.2119$, $\tau = 4.1938$) characterizes the main body of the effective reproductive number ($R_{\text{eff}}$) distribution, corresponding to the prevailing endemic transmission pattern of dengue in the Caraga Region. Its moderate right-skewness indicates that most $R_{\text{eff}}$ values cluster near unity, consistent with sustained but manageable disease spread. The Lomax component ($\alpha_l = 19.328$, $\theta_l = 10.0$) contributes to the tail behavior, capturing the occasional yet epidemiologically significant surge periods where $R_{\text{eff}}$ markedly exceeds 1, representing local epidemic amplification. The small mixing weight ($p = 0.05$) empirically supports the rarity of these high-transmission episodes. Collectively, these parameter estimates and fit indices (lowest AIC/BIC, non-significant K–S test) affirm that the MiGGLo model effectively encapsulates both the central tendency and tail risk of dengue transmission. From an epidemiological standpoint, the coexistence of these two components implies that dengue transmission in Caraga operates under a heterogeneous regime, predominantly endemic, with intermittent outbreak potential triggered by environmental or behavioral perturbations.

Taken together, the results clearly establish that the MiGGLo distribution provides the most adequate statistical representation of the dataset among the candidate models. Its superior performance across all fit indices underscores its potential as a robust alternative for modeling data exhibiting mixed Gamma–Lomax behavior, particularly in contexts where tail heaviness and heterogeneous variance structures are empirically observed.

## 7. Conclusions and Recommendations

The present study established the Mixture of Generalized Gamma and Lomax (MiGGLo) distribution as a robust and adaptable probabilistic framework capable of modeling complex datasets characterized by heterogeneity, multimodality, and varying tail behaviors. Through an extensive Monte Carlo simulation analysis, the Maximum Likelihood Estimation (MLE) method consistently demonstrated superior performance in parameter estimation. Specifically, MLE yielded the most accurate and stable parameter estimates, exhibiting the lowest bias and mean squared error (MSE) across all sample sizes while maintaining computational efficiency relative to the Ordinary Least Squares (OLS) and partial Expectation–Maximization (EM) algorithms. The rapid convergence of MLE estimates toward the true parameter values with increasing sample size confirmed its asymptotic efficiency, underscoring its suitability as the preferred estimation technique for the MiGGLo distribution.

In contrast, although the OLS and partial EM algorithms showed improvement with larger samples, they consistently produced higher bias and MSE values and required substantially longer computation times. These findings indicate that, for complex mixture models such as MiGGLo, MLE remains the most efficient and reliable estimation approach.

The practical utility of the MiGGLo distribution was further substantiated through its successful application to the effective reproductive number ($R_{\text{eff}}$) of dengue in the Caraga Region, Philippines. Among the candidate distributions examined, the MiGGLo model achieved the most favorable fit indices, including the lowest root mean square error (RMSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values. Moreover, a non-significant Kolmogorov–Smirnov (K-S) test result ($p = 0.0221 > 0.01$) confirmed that the MiGGLo distribution provides a statistically adequate and superior representation of the empirical data. Epidemiologically, the fitted model effectively captured both the central tendency of dengue transmission, represented by the Generalized Gamma component, and the heavy-tailed, high-transmission episodes modeled by the Lomax component. The estimated small mixing weight ($p = 0.05$) suggests that high-transmission periods are relatively infrequent, implying a predominantly endemic regime punctuated by sporadic outbreak events. Collectively, these findings demonstrate that the MiGGLo distribution not only offers statistical rigor but also provides meaningful interpretability in contexts characterized by dynamic and heterogeneous processes, such as infectious disease transmission. Its strong empirical and simulation-based performance positions it as a valuable tool for modeling complex stochastic phenomena across a range of disciplines.

Based on the empirical and simulation evidence presented, several recommendations are proposed for future research and practical implementation. To enhance the generalizability and applicability of the MiGGLo distribution, future studies should explore its performance using datasets from various disciplines, such as reliability engineering, financial risk assessment, and survival analysis. These fields frequently exhibit data with heavy-tailed and mixture characteristics similar to those captured by the MiGGLo model. Although MLE demonstrated superior performance, the EM algorithm remains a promising approach due to its iterative suitability for mixture models. Future research should focus on optimizing EM algorithm efficiency by developing improved initialization techniques and hybrid optimization methods that combine EM with other numerical strategies to enhance convergence stability and computational speed, especially for large-scale datasets. Beyond frequentist estimation, Bayesian approaches offer potential for integrating prior information and handling parameter uncertainty, particularly in contexts with limited data. Incorporating Bayesian inference could yield richer probabilistic insights into the MiGGLo distribution's parameters and enhance model interpretability in applied research. Future investigations should also employ additional goodness-of-fit tests, such as the Anderson–Darling and Cramér–von Mises tests, to provide a more comprehensive evaluation of model adequacy. These methods may reveal subtle discrepancies between observed and fitted distributions that complement traditional K-S test results. Lastly, extending the MiGGLo framework to include covariate-dependent parameters could yield deeper insights into how environmental, demographic, or intervention-related factors influence distributional properties. Such an extension would be particularly

valuable in epidemiological modeling, where understanding the contextual drivers of transmission variability can inform targeted disease control and prevention strategies.

**Acknowledgements.** The author acknowledges the University of Southeastern Philippines for providing an enabling research environment and institutional support that made this study possible.

**Conflicts of Interest.** The author declares that there are no conflicts of interest regarding the publication of this paper.

## References

[1] W.S. Abu El Azm, E.M. Almetwally, S. Naji AL-Aziz, A.A.H. El-Bagoury, R. Alharbi, et al., A New Transmuted Generalized Lomax Distribution: Properties and Applications to COVID-19 Data, Comput. Intell. Neurosci. 2021 (2021), 5918511. https://doi.org/10.1155/2021/5918511.

[2] Z. Ahmad, G.G. Hamedani, N.S. Butt, Recent Developments in Distribution Theory: A Brief Survey and Some New Generalized Classes of Distributions, Pak. J. Stat. Oper. Res. 15 (2019), 87–110. https://doi.org/10.18187/pjsor.v15i1.2803.

[3] M. Bee, A Parsimonious Dynamic Mixture for Heavy-Tailed Distributions, Math. Comput. Simul. 230 (2025), 193–206. https://doi.org/10.1016/j.matcom.2024.11.011.

[4] C.V. Cavalcante, K.C.M. Gonçalves, Mixture Models Applied to Heterogeneous Populations, Braz. J. Probab. Stat. 32 (2018), 320–345. https://doi.org/10.1214/16-bjps345.

[5] C. Cox, H. Chu, M.F. Schneider, A. Muñoz, Parametric Survival Analysis and Taxonomy of Hazard Functions for the Generalized Gamma Distribution, Stat. Med. 26 (2007), 4352–4374. https://doi.org/10.1002/sim.2836.

[6] E. D'Innocenzo, A. Lucas, B. Schwaab, X. Zhang, Modeling Extreme Events: Time-Varying Extreme Tail Shape, J. Bus. Econ. Stat. 42 (2023), 903–917. https://doi.org/10.1080/07350015.2023.2260439.

[7] J.V.B. Deluao, M.O. Macalos, J.P. Arcede, Construction of RLUF-Type Copula Generated by Rüschendorf Method to Model Time-Varying Dependency for Dengue $R_{\text{eff}}$, Asia Pac. J. Math. 11 (2024), 105. https://doi.org/10.28924/APJM/11-105.

[8] J.C. Dunic, J. Conner, S.C. Anderson, J.T. Thorson, The Generalized Gamma Is a Flexible Distribution That Outperforms Alternatives When Modelling Catch Rate Data, ICES J. Mar. Sci. 82 (2025), fsaf040. https://doi.org/10.1093/icesjms/fsaf040.

[9] I. Elbatal, A New Lifetime Family of Distributions: Theoretical Developments and Analysis of COVID 19 Data, Results Phys. 31 (2021), 104979. https://doi.org/10.1016/j.rinp.2021.104979.

[10] S. Frühwirth-Schnatter, S. Pyne, Bayesian Inference for Finite Mixtures of Univariate and Multivariate Skew-Normal and Skew-T Distributions, Biostatistics 11 (2010), 317–336. https://doi.org/10.1093/biostatistics/kxp062.

[11] A.M. Gemeay, I. Elbatal, E.M. Almetwally, S.O. Bashiru, I.A. Husseiny, Statistical Modeling with a Novel Distribution: Inference, Information Measures, and Applications to Inflation Rates and Mechanical Failure Data, AIMS Math. 10 (2025), 19357–19394. https://doi.org/10.3934/math.2025865.

[12] R.C. Gupta, S. Lvin, Monotonicity of Failure Rate and Mean Residual Life Function of a Gamma-Type Model, Appl. Math. Comput. 165 (2005), 623–633. https://doi.org/10.1016/j.amc.2004.04.121.

[13] C. Ley, Flexible Modelling in Statistics: Past, Present and Future, J. Soc. Française Stat. 156 (2015), 76–96. https://www.numdam.org/item/JSFS_2015__156_1_76_0.

[14] A. Nandy, A. Basu, A. Ghosh, Robust Inference for Skewed Data in Health Sciences, J. Appl. Stat. 49 (2021), 2093–2123. https://doi.org/10.1080/02664763.2021.1891527.

[15] H.H. Panjer, Operational Risk: Modeling Analytics, John Wiley & Sons, (2006).

[16] J.W. Puspita, M. Fakhruddin, N. Nuraini, E. Soewono, Time-Dependent Force of Infection and Effective Reproduction Ratio in an Age-Structure Dengue Transmission Model in Bandung City, Indonesia, Infect. Dis. Model. 7 (2022), 430–447. https://doi.org/10.1016/j.idm.2022.07.001.

[17] E.A. Rady, W.A. Hassanein, T.A. Elhaddad, The Power Lomax Distribution with an Application to Bladder Cancer Data, SpringerPlus 5 (2016), 1838. https://doi.org/10.1186/s40064-016-3464-y.

[18] M. Sakthivel, P. Pandiyan, A New Model of Mixture Distribution Using a Survival Analysis of Cancer Patients, Braz. J. Biom. 43 (2025), e43733. https://doi.org/10.28951/bjb.v43i1.733.

[19] E.W. Stacy, A Generalization of the Gamma Distribution, Ann. Math. Stat. 33 (1962), 1187–1192. https://doi.org/10.1214/aoms/1177704481.

[20] M.H. Tahir, G.M. Cordeiro, M. Mansoor, M. Zubair, Weibull-Lomax Distribution: Properties and Applications, Hacet. J. Math. Stat. 44 (2014), 1. https://doi.org/10.15672/hjms.2014147465.

[21] W. Zhao, S.K. Khosa, Z. Ahmad, M. Aslam, A.Z. Afify, Type-I Heavy Tailed Family with Applications in Medicine, Engineering and Insurance, PLOS ONE 15 (2020), e0237462. https://doi.org/10.1371/journal.pone.0237462.