On A New Extended Log-Normal Distribution: Properties, Regression, Bayesian Regression, and Data Analysis

Soman Latha Nitin[1] · Damodaran Santhamani Shibu[1] · Radhakumari Maya[1] · Christophe Chesneau[2] · Muhammed Rasheed Irshad[3] · Naushad Mamode Khan[4]

Abstract

This article introduces and investigates the Marshall-Olkin Topp-Leone log-normal (MOTLLN) distribution, a novel extension of the log-normal distribution. It can be presented as a new four-parameter continuous distribution designed to analyze a wide range of versatile positive-valued data. In a brief first part, we explore its main aspects, including the quantile function and the hazard rate function. We then focus on its applied aspect from a statistical perspective. Parameter estimation is performed using both maximum likelihood and Bayesian methods. Furthermore, we employ the MOTLLN distribution to develop a parametric regression model and a Bayesian regression model, demonstrating its versatility. A simulation study supports the practical performance of the maximum likelihood estimation procedure. Real datasets are used to demonstrate the applicability of our methodology. The effectiveness of the additional parameter in the MOTLLN model is assessed by a likelihood ratio test. In addition, the parametric bootstrap method is used to evaluate the suitability of the MOTLLN model for the datasets. All the results obtained confirm the great potential of the proposed model in all aspects.

Keywords Marshall-Olkin family · Topp-Leone family · Bayesian estimation · Infrared astronomy data · Strength of glass fiber data · Cancer data

1 Introduction

Various fields of lifetime data analysis, including survival analysis, astronomical observations, financial modelling, risk assessment, insurance, rare event analysis and biology, require the introduction of novel distributions with different hazard rate characteristics. In fact, the effectiveness of statistical analysis depends largely on the selection of an appropriate distribution adapted to the dataset under consideration, highlighting the importance of model adaptability. However, traditional distributions

---

often prove inadequate to accurately capture the complexity of real-world datasets arising from concrete scenarios. For this reason, there has been a notable resurgence of interest in refining established distributions through the addition of parameter(s), resulting in a number of hazard rate functions (hrfs) suitable for the analysis of skewed data with varying kurtosis. A review of some of these extended models is given in Pham and Lai (2007).

As a fundamental concept, the log-normal (LN) distribution describes the distribution of a continuous random variable whose logarithm follows a normal distribution. It has significant implications for the analysis of lifetime data, particularly when dealing with asymmetric datasets. This is the case in many disciplines within the life sciences, such as biology, geology, ecology and meteorology, as well as economics, finance and risk analysis (see Jobe et al. (1989)). It is also gaining attention in environmental sciences, physics, astrophysics and cosmology (see Bernardeau and Kofman (1994), Blasi et al. (1999), Parravano et al. (2012)). As a key mathematical point, the probability density function (pdf) of the LN distribution is given by

$$q(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right], \tag{1.1}$$

where $x > 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$. The generalizations of the LN distribution have been extensively explored in the statistical literature by various authors, enhancing flexibility (see Chen (1995), Singh et al. (2012), Gui (2013), Kleiber (2014), and Toulias and Kitsos (2013)). In this article, we propose a new generalized version of the LN distribution, achieved through a flexible generalization technique.

The subsequent sections of this article are organized as follows: Sect. 2 elucidates the construction method of this generalization. Section 3 presents its main mathematical definition. Section 4 provides the quantile function (qf) and related measures. Section 5 explores the hrf and its graphical point of view via some examples of plots. Section 6 introduces the maximum likelihood (ML) and Bayesian methods for estimating the unknown parameters of the model. In addition, Sect. 7 outlines a parametric bootstrap method using the ML estimation in a simulation context. Section 8 defines a parametric regression model associated with the new distribution, while Sect. 9 introduces a Bayesian regression method. Section 10 presents a simulation study evaluating the performance of the ML estimates of the parameters involved. Finally, to demonstrate the effectiveness of the new distribution compared to other distributions, Sect. 11 analyzes three univariate uncensored real datasets and one censored real dataset for regression analysis. Section 12 provides concluding remarks.

2 Construction

To understand the construction of the new distribution, a retrospective analysis of various distributions and generator distribution schemes is essential. The Topp-Leone (TL) distribution, first introduced by Topp and Leone (1955), provides

a precisely bounded J-shaped distribution that serves as an alternative to the uniform(0,1) and beta distributions. It has gained particular attention among statisticians in recent years. The cumulative distribution function (cdf) of the TL distribution is

$$F_{TL}(x) = \{x(2-x)\}^{\alpha}, \tag{2.1}$$

where $0 < x < 1$ and $\alpha > 0$. From this simple expression we can immediately derive the corresponding pdf and hrf. It is important to note that the TL distribution has a bathtub shaped hrf for all $\alpha < 1$. A notable study in the literature using the TL distribution as a generator was the TL generalized exponential (TLGE) distribution introduced by Sangsanit and Bodhisuwan (2016). Convincing applications to datasets in materials science and engineering have been obtained.

On the other hand, Marshall and Olkin (1997) proposed a new scheme for generalizing distributions by introducing an additional parameter to a family of distributions called the Marshall-Olkin-G family. The authors demonstrate its applicability by comparing it with the famous exponential and Weibull families. Later, Al-Shomrani et al. (2016) developed another family of distributions, the so-called TL-G family, and discussed the inferential aspects of certain important members. By combining the Marshall-Olkin-G and TL-G families, Khaleel et al. (2020) established the Marshall-Olkin TL-G (MOTL-G) family and studied some of its general properties. The cdf of the MOTL-G family is given by

$$F_{MOTL-G}(x) = \frac{\{F(x)[2-F(x)]\}^{\alpha}}{\beta + (1-\beta)\{F(x)[2-F(x)]\}^{\alpha}}, \tag{2.2}$$

which is defined for any fixed baseline cdf $F(x)$ and $\beta > 0$.

The usefulness of the proposed family is also emphasised by an important member of this family, the so-called Marshall-Olkin TL Weibull (MOTLWe) distribution. The flexibility of the corresponding model compared to various competitors is very convincing and inspires more work in this direction. It is to be hoped that other members of this family will also perform well relative to their respective sub-models. So far, there hasn't been much work done on other members of this family in the literature, which can be seen as a research gap.

Therefore, in this article, we consider an important member of this family by selecting the LN distribution as the baseline distribution, motivated in part by a high potential for applicability. We call the resulting distribution the Marshall-Olkin TL log-normal (MOTLLN) distribution. Our aim is to uncover various statistical properties of this distribution and to use them in reliability analysis. Our motivations are as follows:

(i)   To introduce a novel and versatile continuous distribution capable of effectively modelling lifetime data across a wider range of reliability challenges.
(ii) To extend both LN and TL distributions.
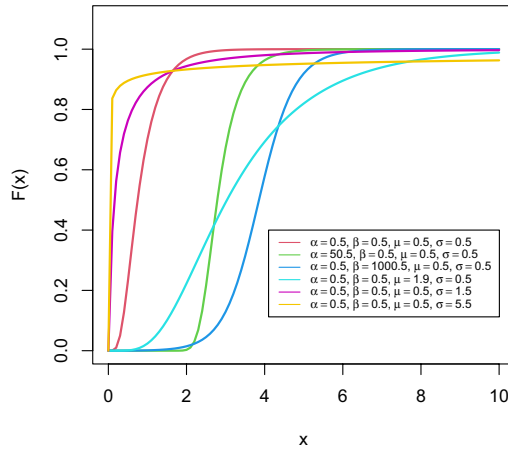(iii) To include additional forms of hrfs.

Fig. 1 Examples of plots of the cdf of the MOTLLN distribution

3 Definition

This part introduces the definition and key features of the MOTLLN distribution.

Definition 3.1 Let X be a continuous random variable. Then we say that X follows the MOTLLN distribution with parameters α, β, μ and σ, if its cdf is expressed as follows:

$$F(x) = \frac{\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha}}{\beta + (1-\beta)\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha}}, \quad (3.1)$$

where $x > 0$, $\mu \in \mathbb{R}$, α, β, $\sigma > 0$, and $\Phi(x)$ and $\phi(x)$ are the cdf and pdf of the standard normal distribution, respectively.

Under this setting, the corresponding pdf is indicated as

$$f(x) = \frac{2\alpha\beta\,\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[1 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\sigma x\left(\beta + (1-\beta)\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha}\right)^{2}}.$$

$$(3.2)$$

To get a visual idea of the functional possibilities of the MOTLLN distribution, Figs. 1 and 2 show some examples of plots of the corresponding cdf and pdf, respectively. The pdf may show a decreasing trend or a single peak characterising unimodality. Some flexibility is observed both in the mode and in the tails. However, it tends to be right-skewed or nearly symmetric.

Remark 3.1 For $\beta = 1$, the cdf of the MOTLLN distribution in Eq. (3.1) reduces to the cdf of the Topp-Leone log-normal (TLLN) distribution, which was developed by Chesneau et al. (2022).
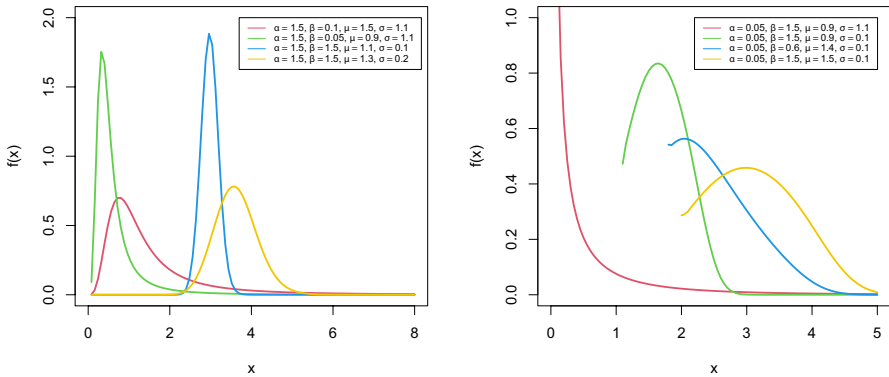
Fig. 2 Examples of plots of the pdf of the MOTLLN distribution

Remark 3.2 The cdf of the MOTLLN distribution reduces to that of the LN distribution if x is fixed, $\beta = 1$ and

$$\alpha = \frac{\log\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]}{\log\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}}.$$

The proof is straightforward and omitted for the sake of brevity.

4 Quantile Function and Measures

In this part, we develop an explicit equation for the qf of the MOTLLN distribution, as well as several of its associated measures.

Theorem 4.1 *Let $p \in (0, 1)$. The pth quantile of the MOTLLN distribution is given by*

$$Q_p = F^{-1}(p) = \exp\left[\mu + \sigma\,\Phi^{-1}\left\{1 - \sqrt{1 - \left(\frac{p\beta}{1 - p(1 - \beta)}\right)^{1/\alpha}}\right\}\right], \quad (4.1)$$

where $F(x)$ is cdf specified in Eq. (3.1) and $\Phi^{-1}(x)$ is the qf associated with the standard normal distribution.

Proof By the definition, the pth quantile Q_p is the solution of the equation $F(Q_p) = p$, so

$$\frac{\left\{1-\left[1-\Phi\left(\frac{\log(Q_p)-\mu}{\sigma}\right)\right]^2\right\}^\alpha}{1-(1-\beta)\left[1-\left\{1-\left[1-\Phi\left(\frac{\log(Q_p)-\mu}{\sigma}\right)\right]^2\right\}^\alpha\right]} = p. \tag{4.2}$$

Let us solve it with a step-by-step approach. On simplification, Eq. (4.2) reduces to

$$\left\{1-\left[1-\Phi\left(\frac{\log(Q_p)-\mu}{\sigma}\right)\right]^2\right\}^\alpha (1-p(1-\beta)) = p\beta$$

$$\Rightarrow \Phi\left(\frac{\log(Q_p)-\mu}{\sigma}\right) = 1-\left[1-\left(\frac{p\beta}{1-p(1-\beta)}\right)^{1/\alpha}\right]^{1/2}$$

$$\Rightarrow \frac{\log(Q_p)-\mu}{\sigma} = \Phi^{-1}\left\{1-\left[1-\left(\frac{p\beta}{1-p(1-\beta)}\right)^{1/\alpha}\right]^{1/2}\right\}$$

$$\Rightarrow Q_p = \exp\left[\mu+\sigma\,\Phi^{-1}\left\{1-\sqrt{1-\left(\frac{p\beta}{1-p(1-\beta)}\right)^{1/\alpha}}\right\}\right].$$

The desired expression is obtained, ending the proof. □

The comprehensive analytical expression of the qf is an advantage of the MOTLLN distribution.

Remark 4.1 An alternative expression of Q_p in Eq. (4.1) is

$$Q_p = \exp\left[\mu+\sigma\sqrt{2}\,\mathrm{erf}^{-1}\left\{1-2\sqrt{1-\left(\frac{p\beta}{1-p(1-\beta)}\right)^{1/\alpha}}\right\}\right], \tag{4.3}$$

where $\mathrm{erf}^{-1}(x)$ is the standard inverse error function.

The median of the MOTLLN distribution is calculated as

$$M = Q_{0.5} = \exp\left[\mu+\sigma\sqrt{2}\,\mathrm{erf}^{-1}\left\{1-2\sqrt{1-\left(\frac{\beta}{1+\beta}\right)^{1/\alpha}}\right\}\right]. \tag{4.4}$$

Similarly, we derive the first and third quartiles of the distribution by taking $p=1/4$ and $p=3/4$ into Q_p, respectively.

5 Hazard Rate Function

The hrf of the MOTLLN distribution is given by the following ratio:

$$h(x) = \frac{f(x)}{S(x)},$$

where $S(x) = 1 - F(x)$ is the survival function (sf) of the MOTLLN distribution. Since

$$S(x) = \frac{\beta \left[1 - \left\{ \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \left[2 - \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \right] \right\}^{\alpha} \right]}{\beta + (1 - \beta) \left\{ \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \left[2 - \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \right] \right\}^{\alpha}},$$

the desired hrf gets the form

$$h(x) = \frac{2\alpha\phi\left(\frac{\log(x) - \mu}{\sigma} \right) \left[1 - \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \right]}{\sigma x \left[1 - \left\{ \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \left[2 - \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \right] \right\}^{\alpha} \right]} \times$$

$$\frac{\left\{ \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \left[2 - \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \right] \right\}^{\alpha-1}}{\beta + (1 - \beta) \left\{ \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \left[2 - \Phi\left(\frac{\log(x) - \mu}{\sigma} \right) \right] \right\}^{\alpha}}.$$

We examine this from a graphical point of view with some sample plots in Fig. 3.

We can also see from this figure that the MOTLLN distribution has increasing, decreasing, bathtub and upside-down bathtub shapes for its hrf. It is therefore complete in terms of hrf modelling.

6 Parametric Estimation

In this part, we examine the MOTLLN distribution from a statistical perspective, focusing on the estimation of its parameters. We explore two famous methods known for their efficiency: ML and Bayesian methods.

6.1 Maximum Likelihood Estimation

First, we consider the ML estimation for the parameters α, β, μ, and σ. Let n be a positive integer, $X_1, X_2, \ldots, X_n$ denote a random sample drawn from the MOTLLN distribution with these unknown parameters, and $x_1, x_2, \ldots, x_n$ represent the corresponding observed values, constituting the data. The log-likelihood function is then be formulated as follows:
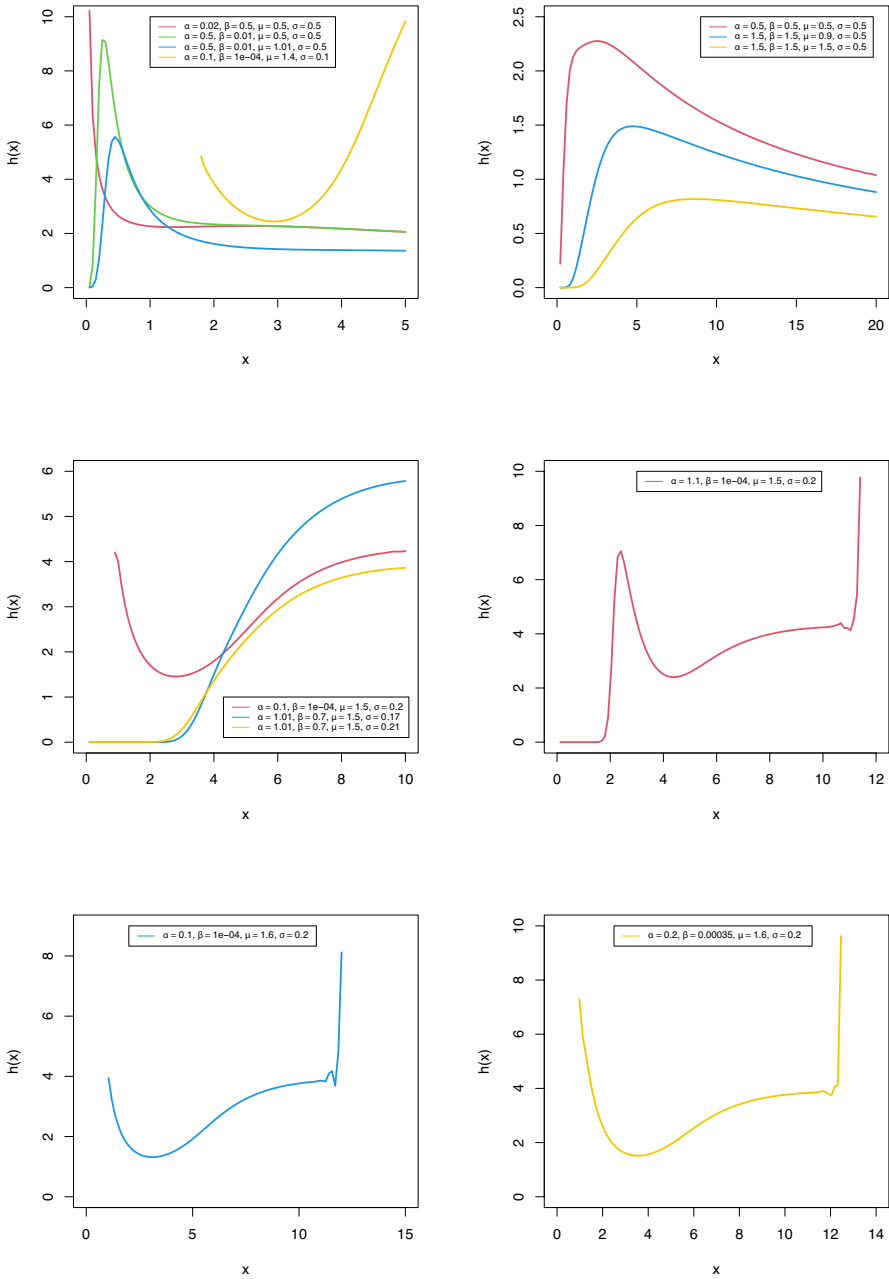
Fig. 3 Examples of plots of the hrf of the MOTLLN distribution

$$\mathcal{L}_n = n\log(2) + n\log(\alpha) + n\log(\beta) - n\log(\sigma)$$

$$- \sum_{i=1}^{n} \log(x_i) + \sum_{i=1}^{n} \log\left[\phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\right] + \sum_{i=1}^{n} \log\left[1 - \Phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\right]$$

$$+ (\alpha - 1) \sum_{i=1}^{n} \log\left[\Phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\right] + (\alpha - 1) \sum_{i=1}^{n} \log\left[2 - \Phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\right]$$

$$- 2 \sum_{i=1}^{n} \log\left(\beta + (1 - \beta)\left\{\Phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\right]\right\}\right).$$

(6.1)

Ideally, we obtain the ML estimates (MLEs) $(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\sigma})$ of $(\alpha, \beta, \mu, \sigma)$ by maximizing this function. From a mathematical point of view, a differentiation can be made. For this purpose, the score function associated with the log-likelihood function is

$$\mathbf{U} = \left(\frac{\partial \mathcal{L}_n}{\partial \alpha}, \frac{\partial \mathcal{L}_n}{\partial \beta}, \frac{\partial \mathcal{L}_n}{\partial \mu}, \frac{\partial \mathcal{L}_n}{\partial \sigma}\right)^T.$$

We now investigate the nonlinear equations $\partial \mathcal{L}_n/\partial \alpha = 0$, $\partial \mathcal{L}_n/\partial \beta = 0$, $\partial \mathcal{L}_n/\partial \mu = 0$, and $\partial \mathcal{L}_n/\partial \sigma = 0$, which are equivalent to

$$\frac{n}{\alpha} + \sum_{i=1}^{n} \log\left[\Phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\right] + \sum_{i=1}^{n} \log\left[2 - \Phi\left(\frac{\log(x_i) - \mu}{\sigma}\right)\right] = 0, \quad (6.2)$$

$$\frac{n}{\beta} + 2 \sum_{i=1}^{n} \frac{\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]}{\beta + (1-\beta)\left\{\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]\right\}} = 0, \quad (6.3)$$

$$\sum_{i=1}^{n} \frac{\log(x_i) - \mu}{\sigma^2} + \frac{1}{\sigma} \sum_{i=1}^{n} \left[\frac{\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}\right]$$

$$- \frac{\alpha - 1}{\sigma} \sum_{i=1}^{n} \left[\frac{\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}\right] + \frac{\alpha - 1}{\sigma} \sum_{i=1}^{n} \left[\frac{\phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}{2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)}\right] \quad (6.4)$$

$$+ \frac{4(1-\beta)}{\sigma} \sum_{i=1}^{n} \left[\frac{\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]}{\beta + (1-\beta)\left\{\Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma}\right)\right]\right\}}\right] = 0,$$

and

$$\frac{-n}{\sigma} + \sum_{i=1}^{n} \frac{(\log(x_i) - \mu)^2}{\sigma^3} + \sum_{i=1}^{n} \left[\frac{\left(\frac{\log(x_i)-\mu}{\sigma^2} \right) \phi\left(\frac{\log(x_i)-\mu}{\sigma} \right)}{1 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right)} \right]$$

$$- \frac{\alpha - 1}{\sigma^2} \sum_{i=1}^{n} \left[\frac{(\log(x_i) - \mu) \, \phi\left(\frac{\log(x_i)-\mu}{\sigma} \right)}{\Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right)} \right] + \frac{\alpha - 1}{\sigma^2} \sum_{i=1}^{n} \left[\frac{(\log(x_i) - \mu) \, \phi\left(\frac{\log(x_i)-\mu}{\sigma} \right)}{2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right)} \right]$$

$$+ \frac{4(1-\beta)}{\sigma^2} \sum_{i=1}^{n} \left[\frac{(\log(x_i) - \mu) \, \Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right) \left[2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right) \right]}{\beta + (1-\beta) \left\{ \Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right) \left[2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right) \right] \right\}} \right] = 0,$$

$$(6.5)$$

respectively. Solving the nonlinear equations given in Eqs. (6.2), (6.3), (6.4) and (6.5) synergistically, we obtain the MLEs $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\mu}$ and $\widehat{\sigma}$. Let us notice that, for known μ and σ, the MLE of α is simply calculated as

$$\widehat{\alpha} = \frac{-n}{\sum_{i=1}^{n} \log\left[\Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right) \right] + \sum_{i=1}^{n} \log\left[2 - \Phi\left(\frac{\log(x_i)-\mu}{\sigma} \right) \right]}.$$

Asymptotic confidence intervals (CIs) for the parameters can be determined. By computing the second partial derivatives of $\mathcal{L}_n$, we can obtain the Hessian matrix for the MOTLLN distribution, which is expressed as

$$H(\Theta) = \begin{pmatrix} \frac{\partial^2 \mathcal{L}_n}{\partial \alpha^2} & \frac{\partial^2 \mathcal{L}_n}{\partial \alpha \, \partial \beta} & \frac{\partial^2 \mathcal{L}_n}{\partial \alpha \, \partial \mu} & \frac{\partial^2 \mathcal{L}_n}{\partial \alpha \, \partial \sigma} \\[2mm] \frac{\partial^2 \mathcal{L}_n}{\partial \beta \, \partial \alpha} & \frac{\partial^2 \mathcal{L}_n}{\partial \beta^2} & \frac{\partial^2 \mathcal{L}_n}{\partial \beta \, \partial \mu} & \frac{\partial^2 \mathcal{L}_n}{\partial \beta \, \partial \sigma} \\[2mm] \frac{\partial^2 \mathcal{L}_n}{\partial \mu \, \partial \alpha} & \frac{\partial^2 \mathcal{L}_n}{\partial \mu \, \partial \beta} & \frac{\partial^2 \mathcal{L}_n}{\partial \mu^2} & \frac{\partial^2 \mathcal{L}_n}{\partial \mu \, \partial \sigma} \\[2mm] \frac{\partial^2 \mathcal{L}_n}{\partial \sigma \, \partial \alpha} & \frac{\partial^2 \mathcal{L}_n}{\partial \sigma \, \partial \beta} & \frac{\partial^2 \mathcal{L}_n}{\partial \sigma \, \partial \mu} & \frac{\partial^2 \mathcal{L}_n}{\partial \sigma^2} \end{pmatrix},$$

where $\Theta = (\alpha, \beta, \mu, \sigma)$. The observed Fisher's information matrix is given by $J(\Theta) = -H(\Theta)$. The inverse of this matrix yields the variance-covariance matrix of the MLEs. We define it via general matrix coefficients as

$$\Sigma = J^{-1}(\Theta) = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} & \Sigma_{34} \\ \Sigma_{41} & \Sigma_{42} & \Sigma_{43} & \Sigma_{44} \end{pmatrix}$$

and $\Sigma_{ij} = \Sigma_{ji}$ for $i \neq j = 1, 2, 3, 4$. Furthermore, it is widely acknowledged that the MLEs exhibit asymptotic normality. That is, $\sqrt{n}(\Theta - \widehat{\Theta}) \sim N_4(0, \Sigma)$, where n is the sample size, $\sim$ means distribution approximation, and $\widehat{\Theta} = (\widehat{\alpha}, \widehat{\beta}, \widehat{\mu}, \widehat{\sigma})$ is the ML vector estimate of Θ.

For a fixed $\delta \in (0, 1)$, we determine $100 \times (1 - \delta)\%$ asymptotic CIs of the parameters by

$$\alpha \in \left[\widehat{\alpha} \mp u_{\delta/2} \sqrt{\Sigma_{11}}\right], \ \beta \in \left[\widehat{\beta} \mp u_{\delta/2} \sqrt{\Sigma_{22}}\right],$$
$$\mu \in \left[\widehat{\mu} \mp u_{\delta/2} \sqrt{\Sigma_{33}}\right] \text{ and } \sigma \in \left[\widehat{\sigma} \mp u_{\delta/2} \sqrt{\Sigma_{44}}\right],$$

where u_δ is the upper δ^{th} percentile of the standard normal distribution.

6.2 Bayesian Estimation

We now perform a Bayesian analysis on the parameters involved in the MOTLLN model. To do this, we need to adopt a prior distribution for each parameter. For this purpose, we use two types of priors, the half-Cauchy (HC) prior and the normal (N) prior. Specifically, the pdf of the HC distribution with scale parameter a is defined as

$$f_{HC}(x_*) = \frac{2a}{\pi(x_*^2 + a^2)}, \tag{6.6}$$

where $x_* > 0$ and $a > 0$. As is well known, the HC distribution lacks a defined mean and variance, while its mode is 0. Although its pdf is almost flat, with $a = 25$, it provides sufficient information for the numerical approximation algorithm to continue exploring the target posterior pdf. The HC distribution thus configured is recommended as a non-informative prior. We also refer to the study by Gelman and Hill (2006), where the HC distribution is shown to be a superior alternative to the uniform distribution and other informative priors. With this in mind, we specify the prior distributions of the parameters as follows:

$$\mu \text{ is assumed to follow the distribution}: N(0, 1000),$$
$$\text{and} \quad \alpha, \beta, \sigma \text{ are assumed to follow the distribution}: HC(25). \tag{6.7}$$

Now, using Eqs (6.7), the joint posterior pdf is obtained as

$$d(\mu, \alpha, \beta, \sigma) \propto \ell_n d(\mu) d(\alpha) d(\beta) d(\sigma), \tag{6.8}$$

where $\ell_n = \exp(\mathcal{L}_n)$ is the likelihood function of the MOTLLN distribution and $\propto$ means "directly proportional to". From Eq. (6.8), it is clear that the Bayesian estimates have no analytical expression. Therefore, we employ the Metropolis-Hastings algorithm, a notable simulation method within the framework of Markov Chain

Monte Carlo (MCMC) techniques. The detailed description of the MCMC method can be found in Upadhyay et al. (2001).

7 Bootstrap CIs

In this part, we apply the parametric bootstrap method to approximate the distribution of the MLEs for the parameters of the MOTLLN model. We then use the bootstrap distribution to estimate the CIs for each parameter of the fitted MOTLLN distribution. Let $\hat{v}$ be an ML estimate of $v \in \{\alpha, \beta, \mu, \sigma\}$ using data, say $x_1, x_2, \ldots, x_n$. In short, the bootstrap is a method of estimating the distribution of the underlying distribution of $\hat{v}$ by taking a random sample for v based on B random samples $v_1^*, v_2^*, \ldots, v_B^*$, which are drawn with replacement from the original data. Further details can be found in Wasserman (2006). The resulting bootstrap sample $v_1^*, v_2^*, \ldots, v_B^*$ can then be used to construct bootstrap CIs for v.

Thus, we obtain $100 \times (1 - \delta)\%$ bootstrap CIs of the parameters using the following formulas:

$$\alpha \in \left[\hat{\alpha} \mp v_{\delta/2} \, \hat{se}_{\alpha,boot}\right], \; \beta \in \left[\hat{\beta} \mp v_{\delta/2} \, \hat{se}_{\beta,boot}\right],$$

$$\mu \in \left[\hat{\mu} \mp v_{\delta/2} \, \hat{se}_{\mu,boot}\right] \text{ and } \sigma \in \left[\hat{\sigma} \mp v_{\delta/2} \, \hat{se}_{\sigma,boot}\right],$$

where v_δ denotes the δth percentile of the bootstrap sample and, for $v \in \{\alpha, \beta, \mu, \sigma\}$,

$$\hat{se}_{v,boot} = \sqrt{\frac{1}{B} \sum_{b=1}^{B} \left(v_b^* - \frac{1}{B} \sum_{b=1}^{B} v_b^*\right)^2}.$$

8 MOTLLN Regression Model

This part develops a regression model based on the MOTLLN distribution. To do this, let X be a random variable following the MOTLLN distribution. We recall that the corresponding pdf is specified in Eq. (3.2). Then the random variable $Y = \log(X)$ has the following pdf:

$$f_Y(y) = \frac{2\alpha\beta \, \phi\left(\frac{y-\mu}{\sigma}\right)\left[1 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right]\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\sigma\left(\beta + (1-\beta)\left\{\Phi\left(\frac{y-\mu}{\sigma}\right)\left[2 - \Phi\left(\frac{y-\mu}{\sigma}\right)\right]\right\}^{\alpha}\right)^2}, \quad (8.1)$$

where $y \in \mathbb{R}$, α, $\beta > 0$ are the shape parameters, $\mu \in \mathbb{R}$ is the location parameter, and $\sigma > 0$ is the scale parameter. We refer to the distribution defined by the pdf in Eq. (8.1) as the log-MOTLLN distribution. In this setting, the standardized random variable $Z = (Y - \mu)/\sigma$ has the pdf indicated as

$$f_Z(z) = \frac{2\alpha\beta \ \phi(z)[1 - \Phi(z)]\{\Phi(z)[2 - \Phi(z)]\}^{\alpha-1}}{\left(\beta + (1 - \beta)\{\Phi(z)[2 - \Phi(z)]\}^{\alpha}\right)^2}, \tag{8.2}$$

where $z \in \mathbb{R}$. Now, from n independent observations, for any $i = 1, \dots, n$, the linear location-scale regression model linking the ith value of the response variable Y, say y_i, and the explanatory variable vector $\mathrm{v}_i^T = (v_{i1}, v_{i2}, \dots, v_{ip})$ is given by

$$y_i = \mu_i + \sigma z_i, \tag{8.3}$$

where z_i is the random error component that has the pdf in Eq. (8.2), $\mu_i = \mathrm{v}_i^T \tau$ is the location parameter of y_i, where $\tau = (\tau_1, \tau_2, \dots, \tau_p)^T$, and α, β and σ are unknown parameters. Thus, the location parameter vector $\mu = (\mu_1, \mu_2, \dots, \mu_n)^T$ is represented by a linear model $\mu = V\tau$, where $V = (V_1, V_2, \dots, V_n)^T$ is a known model matrix.

In this setting, we propose the MOTLLN regression model from Eq. (8.3). It is governed by the following formula:

$$x_i = \exp(y_i) = \exp(\mu_i + \sigma z_i), \qquad i = 1, 2, \dots, n. \tag{8.4}$$

Let us consider a sample $(x_1, \mathrm{v}_1), (x_2, \mathrm{v}_2), \dots, (x_n, \mathrm{v}_n)$ of n independent observations and the vector of parameters $\boldsymbol{\psi} = (\tau^T, \alpha, \beta, \sigma)^T$ from model Eq. (8.4). Then the total log-likelihood function for right censored data has the form

$$\begin{aligned} l(\boldsymbol{\psi}) &= \log \left\{ \prod_{i=1}^{n} \left[f(x_i)\right]^{\delta_i} \left[S(x_i)\right]^{1-\delta_i} \right\} \\ &= \sum_{i=1}^{n} \delta_i \log \left[f(x_i)\right] + \sum_{i=1}^{n} (1 - \delta_i)\left[\log[S(x_i)]\right], \end{aligned} \tag{8.5}$$

where, for $i = 1, \dots, n$, $\delta_i = 1$ if survival (uncensored), and $\delta_i = 0$ otherwise (censored), and $f(x)$ and $S(x)$ are the pdf and sf of the MOTLLN distribution, respectively.

9 Bayesian Regression Model

The Bayesian analysis technique is known to be efficient for analyzing survival patterns in various real-world scenarios. In this study, we investigate its usefulness for fitting a regression model based on the MOTLLN distribution while incorporating information on prior parameters. We then use a simulation approach to facilitate Bayesian analysis of the model.

To perform a Bayesian analysis, it is essential to specify prior probability distributions for the model parameters. In this context, similarly to Subsection 6.2, we use two different priors: HC and N priors. We recall that the pdf of the HC distribution with scale parameter a is described in Eq. (6.6). Now, we can write the likelihood function for right censored data as

$$L = \prod_{i=1}^{n} \left[f(x_i)\right]^{\delta_i} \left[S(x_i)\right]^{1-\delta_i}, \tag{9.1}$$

where, for $i = 1, \ldots, n$, $\delta_i = 1$, if survival (uncensored) and $\delta_i = 0$ otherwise (censored), and $f(x)$ and $S(x)$ are the pdf and sf of the MOTLLN distribution, respectively. We use the following link function:

$$\mu = V\tau, \tag{9.2}$$

which operates a linear combination of explanatory variables. Here,

$$\tau_j \text{ is assumed to follow the distribution}: N(0, 1000); \quad j = 1, 2, \ldots, J$$

and α, β, σ are assumed to follow the distribution : $HC(25)$.

$$(9.3)$$

Now, using Equations (9.1), (9.2) and (9.3), the joint posterior pdf is obtained as

$$d(\tau, \alpha, \beta, \sigma) \propto L(V, \tau, \alpha, \beta, \sigma) d(\tau) d(\alpha) d(\beta) d(\sigma). \tag{9.4}$$

From Equation (9.4), it is evident that obtaining analytical solutions for the Bayesian estimates is not feasible. Hence, we employ the simulation approach, specifically the Metropolis Hastings algorithm of the MCMC method, as discussed in Subsection 6.2.

10 Performance of the Estimates Using Simulation Study

This part presents simulation experiments aimed at evaluating the performance of the MLEs for the parameters of the MOTLLN distribution across various finite sample sizes. We have simulated datasets of sizes $n = 50, 250, 500$, and 1000 from the MOTLLN distribution for the parameter values $\alpha = 0.5, \beta = 5.5, \mu = 0.9, \sigma = 0.6$ and iterated each sample 500 times. The average absolute biases, mean squared errors (MSEs), coverage probabilities (CPs) and average lengths (ALs) are then calculated over all replications within the respective sample sizes. The analysis computes their values using the following formulas:

- Average absolute bias of the simulated estimates $= \dfrac{1}{500} \sum\limits_{i=1}^{500} |\hat{v}_i - v|$,

- Average MSE of the simulated estimates $= \dfrac{1}{500} \sum\limits_{i=1}^{500} (\hat{v}_i - v)^2$,

where $\hat{v} \in \{\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\sigma}\}$ is the estimate of the corresponding parameter $v \in \{\alpha, \beta, \mu, \sigma\}$. Table 1 contains the results. We draw the conclusion that as sample size increases, the MSE and AL of each estimate decrease. Also, the average absolute bias often appears to decrease as sample size increases. Also, the CPs of the confidence intervals (CIs) for each parameter are relatively close

Table 1 The MLE simulation results

Parameters	n	MLE	Absolute bias	MSE	CP	AL
α	50	0.5569	0.0569	0.0926	0.940	3.4074
	250	0.5799	0.0799	0.0588	0.968	1.7770
	500	0.5658	0.0658	0.0444	0.976	1.2498
	1000	0.5448	0.0448	0.0264	0.970	0.8418
β	50	6.1530	0.6530	0.7804	0.944	38.4724
	250	6.2569	0.7569	0.6037	0.996	22.3307
	500	6.1219	0.6219	0.4844	0.999	16.0512
	1000	6.0575	0.5575	0.3557	0.999	11.6241
μ	50	0.8753	0.0247	0.0447	0.932	3.3381
	250	0.8356	0.0644	0.0312	0.990	1.9255
	500	0.8451	0.0549	0.0252	0.994	1.4052
	1000	0.8571	0.0429	0.0189	0.998	1.0092
σ	50	0.5908	0.0092	0.0150	0.940	1.3215
	250	0.6186	0.0186	0.0074	0.984	0.7465
	500	0.6169	0.0169	0.0053	0.994	0.5458
	1000	0.6147	0.0147	0.0037	0.996	0.3959

to the nominal 95 percent level. This demonstrates how consistent the subjacent estimators are.

11 Applications and Empirical Study

In this part, we illustrate the empirical relevance of the MOTLLN distribution. In order to highlight its data modelling ability over several competing distributions, we study four real datasets. The first is a dataset related to astronomy, while the second and third are datasets based on the strength of glass fibers of two different lengths. In the fourth, we use a real censored dataset based on the prognosis of women with breast cancer for the regression study. To analyze these datasets numerically, we use the **R** software.

11.1 Real Illustration on Univariate Real Datasets

The astronomy dataset (Dataset 1) consists of the near-infrared K-band magnitudes of the 360 globular clusters in Messier 31 (M31), our neighbouring Andromeda galaxy, taken from Nantais et al. (2006). It is given in Appendix C.3 of Feigelson and Babu (2012), as well as in the R package astrodatR. A first statistical analysis of this dataset, focusing on a specific distribution, was carried out by Chesneau et al. (2022). The K-band represents a crucial atmospheric transmission window in infrared astronomy, indicating a region within the infrared spectrum

where terrestrial heat radiation is minimally absorbed by atmospheric gases. Globular clusters, characterized by densely packed groups of old stars ranging from 10^4 to 10^6, have a distinctly dense, roughly spherical structure that sets them apart from the wider stellar population. Astronomers use these clusters to deduce the age of the Universe or to pinpoint the Galactic Centre through meticulous study.

In addition, the second and third datasets (Datasets 2 and 3) represent the experimental data on the strength of glass fibers of two lengths, 1.5 cm and 15 cm, from the National Physical Laboratory in England. Both of them are available in Smith and Naylor (1987).

By using standard measures, the descriptive statistics of these three datasets are given in Table 2.

We also investigate the corresponding empirical hrf using the idea of a total time on test (TTT) plot. This plot is a graph formed by the points $(i/n, w_{i,n})$, $i = 1, 2, \ldots, n$, where

$$
w_{i,n} = \frac{(n-i)x_{i:n} + \sum_{r=1}^{i} x_{r:n}}{\sum_{r=1}^{n} x_{r:n}},
$$

where $x_{r:n}$, $r = 1, 2, \ldots, n$ are the data ordered in an increasing order. It is used as a visual indicator to distinguish between several types of ageing as displayed by the hrf shapes. For more about the construction and interpretation of the TTT plot, see Aarset (1987).

Thus, based on the TTT plot approach, Fig. 4 indicates that the datasets have an increasing shape for their empirical hrf. Therefore, the MOTLLN distribution provides a credible model for analyzing these data.

The following distributions are taken into consideration for comparison purposes and highlight the tangible benefits of the MOTLLN distribution.

- The two-parameter log-normal (LN) distribution with the pdf indicated as

$$
f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} \exp\left[-\frac{(\log(x) - \mu)^2}{2\sigma^2}\right],
$$

where $x > 0$, $\mu \in \mathbb{R}$ and $\sigma > 0$.
- The TLLN distribution already mentioned with the pdf given by

Table 2 Descriptive statistics of the three datasets

	n	M	Md	Var	Sk	Ku	min	max
Dataset 1	360	14.458	14.54	1.427	−0.395	0.344	10.749	18.052
Dataset 2	63	1.507	1.59	0.105	−0.899	0.924	0.55	2.24
Dataset 3	46	1.129	1.16	0.073	−0.802	0.599	0.37	1.6

Astronomical data set

Strength of glass fibre of length 1.5 cm

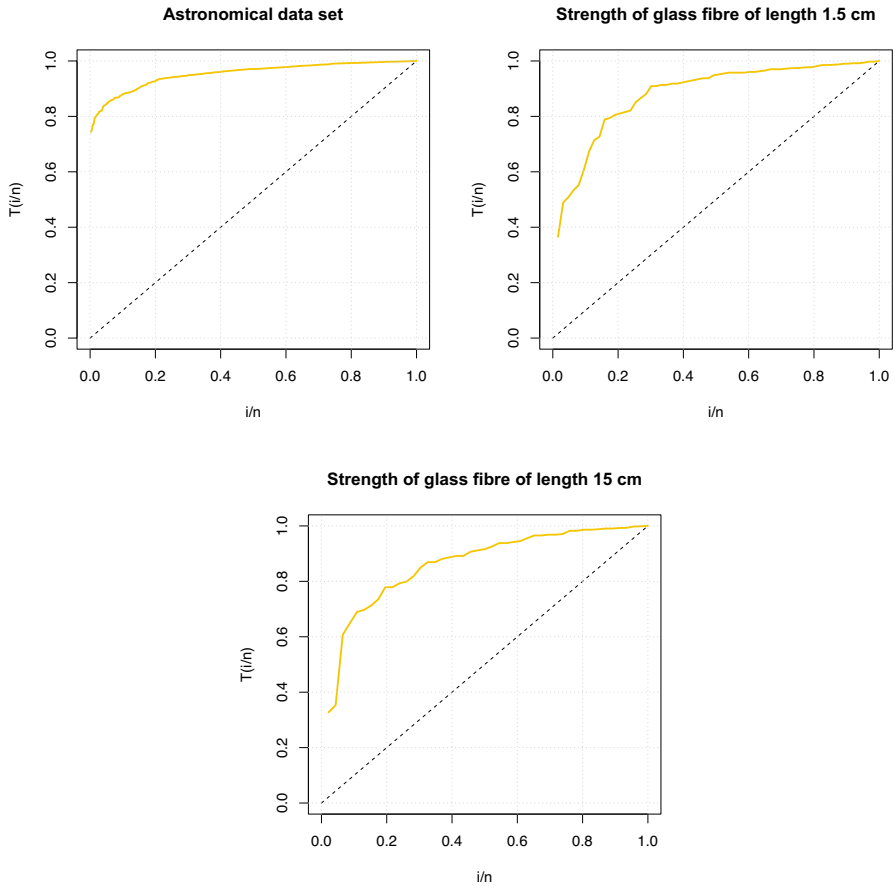**Strength of glass fibre of length 15 cm**

Fig. 4 The TTT plots of the three data sets

$$f(x) = \frac{2\alpha}{\sigma x} \phi\left(\frac{\log(x) - \mu}{\sigma}\right)\left[1 - \Phi\left(\frac{\log(x) - \mu}{\sigma}\right)\right] \times$$

$$\left\{\Phi\left(\frac{\log(x) - \mu}{\sigma}\right)\left[2 - \Phi\left(\frac{\log(x) - \mu}{\sigma}\right)\right]\right\}^{\alpha - 1},$$

where $x > 0$, $\mu \in \mathbb{R}$ and $\sigma, \alpha > 0$. See Chesneau et al. (2022).

- The exponentiated log-normal distribution (ELN) with the pdf specified by

$$f(x) = \frac{\alpha}{x\sigma} \phi\left(\frac{\log(x) - \mu}{\sigma}\right)\left[\Phi\left(\frac{\log(x) - \mu}{\sigma}\right)\right]^{\alpha - 1},$$

where $x > 0$, $\mu \in \mathbb{R}$ and $\alpha, \sigma > 0$.

- The odd log-logistic log-normal (OLL-LN) distribution with the pdf indicated by

$$f(x) = \frac{\alpha\phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left\{\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\left[1-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]\right\}^{\alpha-1}}{\sigma x\left\{\left[\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^{\alpha}+\left[1-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right]^{\alpha}\right\}^{2}},$$

where $x > 0$, $\mu \in \mathbb{R}$, and $\alpha, \sigma > 0$. See Ozel Kadilar et al. (2018).

- The beta log-normal distribution (BLN) with the pdf given by

$$f(x) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{\log(x)-\mu}{\sigma}\right)^{2}\right\}}{x\sigma\sqrt{2\pi}\,B(\alpha,\beta)}\,\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)^{\alpha-1}\left\{1-\Phi\left(\frac{\log(x)-\mu}{\sigma}\right)\right\}^{\beta-1},$$

where $x > 0$, $\mu \in \mathbb{R}$, $\alpha, \beta, \sigma > 0$ and $B(\alpha,\beta)$ is the standard beta function. See Castellares et al. (2011).

- The exponentiated exponential (EE) distribution with the pdf indicated as

$$f(x) = \alpha\sigma(1 - e^{-\sigma x})^{\alpha-1}e^{-\sigma x},$$

where $x, \alpha, \sigma > 0$.

- The standard gamma distribution with the pdf specified by

$$f(x) = \frac{1}{\sigma^{\alpha}\Gamma(\alpha)}x^{\alpha-1}\exp\left(-\frac{x}{\sigma}\right),$$

where $x, \alpha, \sigma > 0$ and $\Gamma(\alpha)$ is the standard gamma function.

We use a variety of statistical techniques to assess the goodness-of-fit of distributions to real datasets, including the log-likelihood (*LL*), Kolmogorov-Smirnov (*KS*), Cramér-von Mises (*W**), Anderson-Darling (*A**) statistics, alongside the Akaike information criterion (AIC) and Bayesian information criterion (BIC) values.

Furthermore, Tables 3, 4, and 5 provide the MLEs and goodness-of-fit statistics for the corresponding three datasets, respectively. Notably, the values of *KS*, *W**, *A**, AIC, and BIC associated with the MOTLLN distribution are consistently smaller compared to other distributions. Addition, we present essential graphical representations, including empirical pdf plots, empirical cdf plots, quantile-quantile (Q-Q) and probability-probability (P-P) plots for the three datasets in Figs. 5, 6, and 7, respectively. These plots show superimposed curves of the fitted and empirical functions. Thus, based on these comprehensive analyses, we assert that the MOTLLN distribution emerges as the most appropriate choice for all three given datasets when compared to alternative distributions.

To consolidate the previous analyses, we utilize the likelihood ratio (LR) test to compare the adequacy of the MOTLLN distribution with that of the TLLN, LN, and ELN distributions. Thus, we test the null hypotheses H_0 : TLLN distribution, H_0 : LN distribution and H_0 : ELN distribution against H_1 : MOTLLN distribution.

Tables 6, 7, and 8 show the LR statistics and corresponding *p*-values for all three datasets, respectively. Given the values of the test statistics and their

Table 3 Dataset 1: MLEs and goodness-of-fit statistics results

Model	ML estimates	-log L	AIC	BIC	KS	W^*	A^*
MOTLLN	$\hat{\alpha} = 0.8343$	567.2737	1142.547	1158.092	0.0347	0.0611	0.3902
	$\hat{\beta} = 7.2339$						
	$\hat{\mu} = 2.6445$						
	$\hat{\sigma} = 0.1086$						
TLLN	$\hat{\alpha} = 0.2694$	571.7256	1149.451	1161.110	0.0621	0.2083	1.2120
	$\hat{\mu} = 2.7796$						
	$\hat{\sigma} = 0.0601$						
BLN	$\hat{\alpha} = 0.6271$	569.6364	1147.273	1162.817	0.0507	0.1184	0.7092
	$\hat{\beta} = 13.7742$						
	$\hat{\mu} = 2.9161$						
	$\hat{\sigma} = 0.1221$						
OLL-LN	$\hat{\alpha} = 4.1863$	578.3391	1162.678	1174.337	0.0467	0.1809	1.9124
	$\hat{\mu} = 2.6731$						
	$\hat{\sigma} = 0.3147$						
ELN	$\hat{\alpha} = 0.1070$	573.2549	1152.510	1164.168	0.0657	0.2517	1.4640
	$\hat{\mu} = 2.7826$						
	$\hat{\sigma} = 0.0371$						
LN	$\hat{\mu} = 2.6677$	582.5224	1169.045	1176.817	0.0774	0.5396	3.2814
	$\hat{\sigma} = 0.0847$						
EE	$\hat{\alpha} = 44847.58$	624.776	1253.552	1261.324	0.1293	1.7963	10.8203
	$\hat{\sigma} = 0.7740$						
Gamma	$\hat{\alpha} = 142.2090$	579.3487	1162.697	1170.470	0.0722	0.4406	2.7088
	$\hat{\sigma} = 9.8355$						

associated *p*-values, we reject the null hypotheses in all cases. Therefore, we conclude that the MOTLLN distribution provides a significantly better representation than the compared distributions.

Now, the Hessian matrix corresponding to Datasets 1, 2, and 3 are, respectively, obtained as

$$H_1(\Theta) = \begin{pmatrix} 240.1912 & 22.0588 & 3419.8414 & 441.0525 \\ 22.0588 & 2.2954 & 363.4472 & 125.6094 \\ 3419.8414 & 363.4472 & 63104.8204 & 36568.9384 \\ 441.0525 & 125.6094 & 36568.9384 & 69063.4630 \end{pmatrix},$$

$$H_2(\Theta) = \begin{pmatrix} 3241.6238 & 22.8931 & 2627.8527 & -2301.7318 \\ 22.8931 & 0.1841 & 19.8020 & -12.2441 \\ 2627.8527 & 19.8020 & 158784.2252 & -92225.4781 \\ -2301.7318 & -12.2441 & -92225.4781 & 325745.0176 \end{pmatrix},$$

and

Table 4 Dataset 2: MLEs and goodness-of-fit statistics results

Model	ML estimates	-log L	AIC	BIC	KS	W^*	A^*
MOTLLN	$\hat{\alpha} = 0.0934$	11.5938	31.1876	39.7602	0.1044	0.0901	0.6015
	$\hat{\beta} = 10.8277$						
	$\hat{\mu} = 0.5538$						
	$\hat{\sigma} = 0.1411$						
TLLN	$\hat{\alpha} = 0.0733$	-20.6278	47.2555	53.6849	0.2196	0.5605	2.8067
	$\hat{\mu} = 0.7507$						
	$\hat{\sigma} = 0.0916$						
BLN	$\hat{\alpha} = 0.3258$	15.8890	39.7780	48.3504	0.1743	0.2999	1.5543
	$\hat{\beta} = 101.8488$						
	$\hat{\mu} = 1.2917$						
	$\hat{\sigma} = 0.2808$						
OLL-LN	$\hat{\alpha} = 34.4571$	22.7971	51.5942	58.0236	0.1539	0.3066	2.3773
	$\hat{\mu} = 0.4227$						
	$\hat{\sigma} = 6.9423$						
ELN	$\hat{\alpha} = 0.0111$	20.6062	47.2124	53.6418	0.2175	0.5531	2.7723
	$\hat{\mu} = 0.7944$						
	$\hat{\sigma} = 0.0371$						
LN	$\hat{\mu} = 0.3811$	28.0049	60.0099	64.2961	0.2313	0.6918	3.7736
	$\hat{\sigma} = 0.2578$						
EE	$\hat{\alpha} = 31.3533$	31.3835	66.7669	71.0532	0.2291	0.7985	4.3372
	$\hat{\sigma} = 2.6117$						
Gamma	$\hat{\alpha} = 17.4410$	23.9515	51.9031	56.1894	0.2164	0.5658	3.0870
	$\hat{\sigma} = 11.5746$						

$$H_3(\Theta) = \begin{pmatrix} 3211.0658 & 22.4359 & 1772.6019 & -1893.3673 \\ 22.4359 & 0.1917 & 13.2660 & -11.1864 \\ 1772.6019 & 13.2660 & 163236.1780 & -303.3624 \\ -1893.3673 & -11.1864 & -303.3624 & 263621.1053 \end{pmatrix}.$$

The 95% asymptotic CIs of the unknown parameters of the MOTLLN distribution for the three datasets are given in Table 9.

We then focus on the application of the Bayesian procedure to estimate the parameters of the MOTLLN distribution. The analysis was carried out using the Metropolis-Hastings algorithm of the MCMC method with 1000 iterations in the context of Bayesian estimation. Table 10 shows the MLEs and Bayes estimates of the parameters of the MOTLLN model for each of the three datasets. The Bayesian estimation calculations were also performed in the **R** software.

We now calculate the 95% bootstrap CIs for the parameters α, β, μ, and σ using the obtained MLEs. Based on the MOTLLN distribution, we simulate 1001 samples of the corresponding sizes found in Datasets 1, 2, and 3. The true

Table 5 Dataset 3: MLEs and goodness-of-fit statistics results

Model	ML estimates	-log L	AIC	BIC	KS	W^*	A^*
MOTLLN	$\hat{\alpha} = 0.0814$	1.8334	11.6668	18.9814	0.0639	0.0209	0.1863
	$\hat{\beta} = 9.1544$						
	$\hat{\mu} = 0.3140$						
	$\hat{\sigma} = 0.1603$						
TLLN	$\hat{\alpha} = 0.0111$	6.7521	19.5042	24.9901	0.1487	0.2311	1.3159
	$\hat{\mu} = 0.5356$						
	$\hat{\sigma} = 0.0408$						
BLN	$\hat{\alpha} = 0.1230$	4.7707	17.5413	24.8559	0.1209	0.1309	0.8077
	$\hat{\beta} = 11.0099$						
	$\hat{\mu} = 0.6514$						
	$\hat{\sigma} = 0.1535$						
OLL-LN	$\hat{\alpha} = 173.6657$	9.0974	24.1947	29.6806	0.1053	0.0911	0.9267
	$\hat{\mu} = 0.1256$						
	$\hat{\sigma} = 40.6434$						
ELN	$\hat{\alpha} = 0.0115$	7.1653	20.3306	25.8165	0.1549	0.2719	1.5147
	$\hat{\mu} = 0.5135$						
	$\hat{\sigma} = 0.0402$						
LN	$\hat{\mu} = 0.0849$	13.2199	30.4399	34.0973	0.1598	0.2773	1.8512
	$\hat{\sigma} = 0.2963$						
EE	$\hat{\alpha} = 20.4425$	14.5772	33.1545	36.8117	0.1603	0.3358	2.2139
	$\hat{\sigma} = 3.1153$						
Gamma	$\hat{\alpha} = 13.6452$	9.6319	23.2637	26.9210	0.1391	0.1916	1.3027
	$\hat{\sigma} = 12.0778$						

values of the parameters are assumed to be the MLEs of the parameters. The MLEs $\hat{\alpha}_b^*$, $\hat{\beta}_b^*$, $\hat{\mu}_b^*$, and $\hat{\sigma}_b^*$ were estimated for each sample that was obtained, with $b \in \{1, 2, \ldots, 1001\}$. Table 11 displays the median and 95% bootstrap CIs for the parameters α, β, μ, and σ for the three datasets.

11.2 Real Illustration on Censored Dataset Using Regression

In this part, we use an actual censored dataset derived from the prognosis of breast cancer patients in women. As described by Collett (2015) and documented in Leathem and Brooks (1987), it contains the number of months that women who underwent either a simple or radical mastectomy to treat a grade II, III or IV tumour between January 1969 and December 1971 survived.

The MOTLLN regression model summary resulting from the censored dataset is shown in Table 12. It includes the estimates of all parameters, the negative log-likelihood, and the value of the AIC.
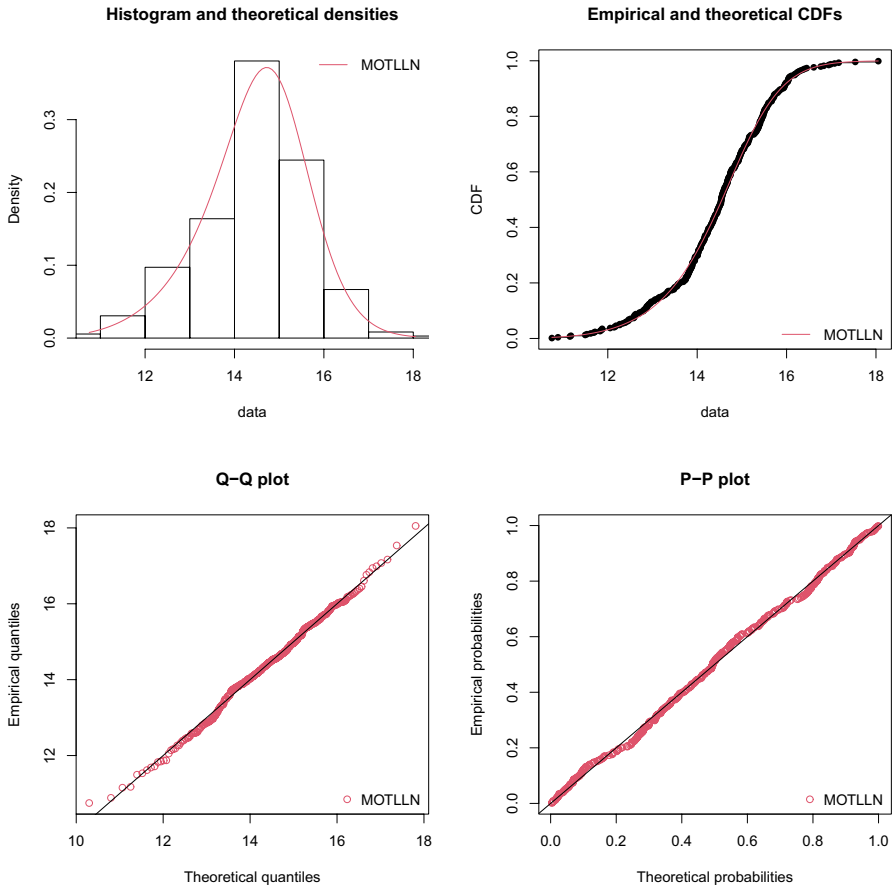
Fig. 5 Various empirical plots of Dataset 1

When using the Metropolis-Hastings algorithm of the MCMC method to solve the Bayesian regression problem, Table 13 provides a summary of 1000 times the iterated simulated results due to the censored dataset. This summary includes the posterior mean, standard deviation (SD), Monte Carlo standard error (MCSE), 95% CIs, and posterior median.

12 Concluding Remarks

We propose a new distribution that is a generalization of the LN distribution, which we call the Marshall-Olkin Topp-Leone log-normal (MOTLLN) distribution. In the first part, we study its mathematical and statistical aspects. In particular, the associated quantile function and the hazard rate function are given as explicit formulations. In terms of applicability, maximum likelihood and Bayesian
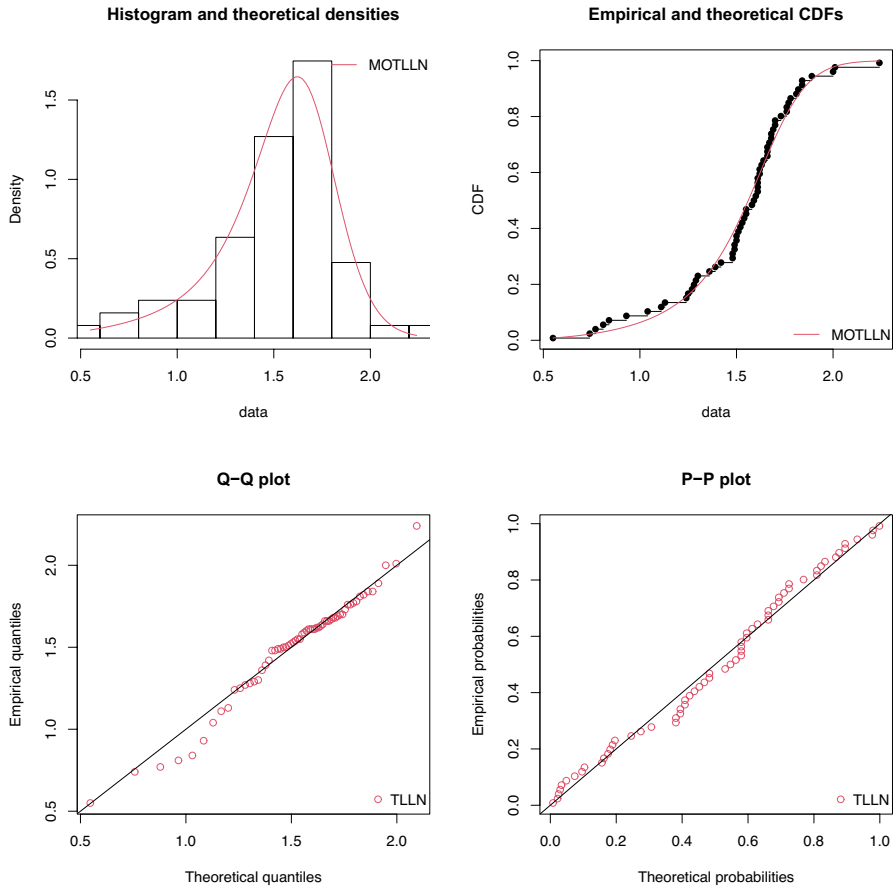
Fig. 6 Various empirical plots of Dataset 2

estimation are used to estimate the model parameters. The parametric bootstrap technique is applied to obtain confidence intervals for the model parameters. More importantly, based on the new distribution, we offer a parametric regression model and a Bayesian regression approach. In this study, four applications of the new model to real datasets using goodness-of-fit tests demonstrate its utility. In particular, it consistently outperforms previous models in the literature on valuable goodness-of-fit criteria. We expect that the proposed model will be used to analyze positive real datasets in a variety of fields, including physics, astronomy, engineering, survival analysis, hydrology, economics, and others.
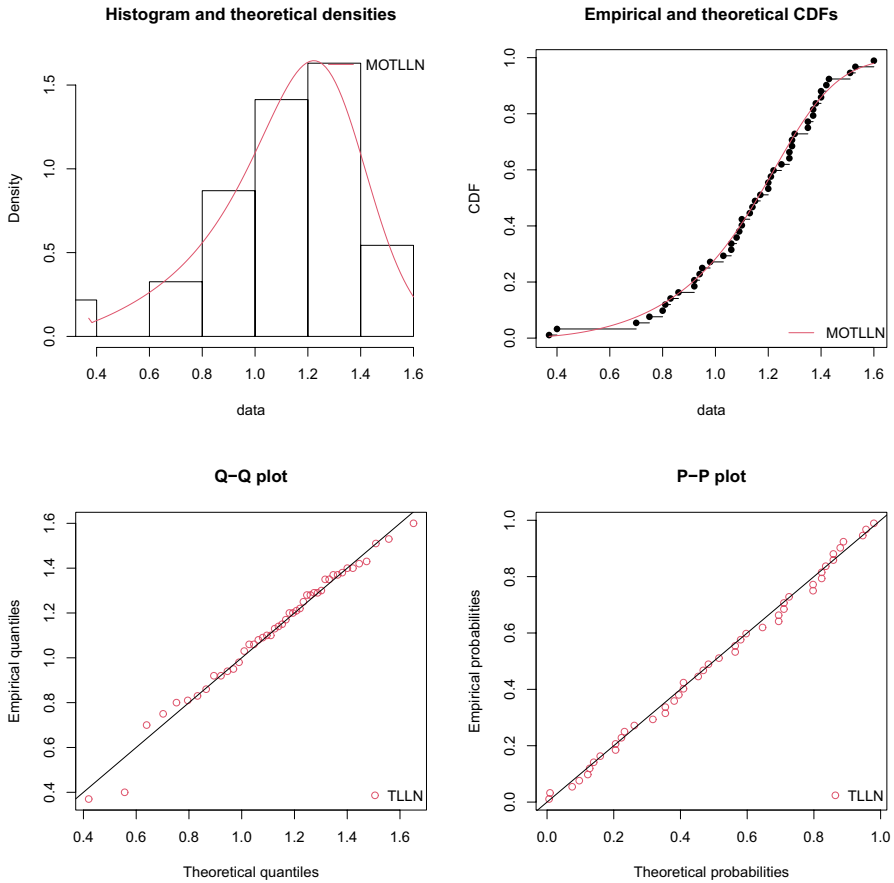
Histogram and theoretical densities

Empirical and theoretical CDFs

Q–Q plot

P–P plot

Fig. 7 Various empirical plots of Dataset 3

		Null hypothesis	LR	*p*-value
Table 6 The LR test results for Dataset 1 (~ means "follows the distribution")	MOTLLN versus TLLN	$H_0 : X \sim$ TLLN	8.9038	0.002846
	MOTLLN versus LN	$H_0 : X \sim$ LN	30.497	2.385e−07
	MOTLLN versus ELN	$H_0 : X \sim$ ELN	11.962	0.0005428

		Null hypothesis	LR	*p*-value
Table 7 The LR test results for Dataset 2	MOTLLN versus TLLN	$H_0 : X \sim$ TLLN	17.856	2.383e−05
	MOTLLN versus LN	$H_0 : X \sim$ LN	32.822	0.0075e−05
	MOTLLN versus ELN	$H_0 : X \sim$ ELN	18.025	0.0022e−02

Table 8 The LR test results for Dataset 3

	Null hypothesis	LR	p-value
MOTLLN versus TLLN	$H_0 : X \sim$ TLLN	9.8374	0.00171
MOTLLN versus LN	$H_0 : X \sim$ LN	22.773	1.135e-05
MOTLLN versus ELN	$H_0 : X \sim$ ELN	10.664	0.001093

Table 9 The 95% asymptotic CIs of the parameters of the MOTLLN distribution for the three datasets

Parameter	Dataset 1		Dataset 2		Dataset 3	
	Lower	Upper	Lower	Upper	Lower	Upper
α	−0.3930	2.0615	−0.0054	0.1922	0.00019	0.1625
β	−5.2171	19.6848	−2.2783	23.9337	−1.3393	19.6482
μ	2.4830	2.8060	0.5484	0.5592	0.3091	0.3188
σ	0.0504	0.1668	0.1374	0.1449	0.1565	0.1641

Table 10 MLEs and Bayes estimates of the parameters of the MOTLLN distribution for the three datasets

Parameter	Dataset 1		Dataset 2		Dataset 3	
	ML	Bayes	ML	Bayes	ML	Bayes
α	0.8343	0.7004	0.0934	0.1042	0.0814	0.0793
β	7.2339	9.8815	10.8277	12.4572	9.1544	7.0191
μ	2.6445	2.6445	0.5538	0.5323	0.3140	0.3645
σ	0.1086	0.1089	0.1411	0.1564	0.1603	0.1859

Table 11 The median and 95% bootstrap CIs for the parameters of the MOTLLN distribution for the three datasets

	Parameter	Median	Bootstrap CI
Dataset 1	α	0.9443	(0.3138, 3.4431)
	β	7.2339	(3.0246, 51.0255)
	μ	2.6334	(2.4840, 2.7236)
	σ	0.1158	(0.0771, 0.1624)
Dataset 2	α	0.1131	(0.0476, 1.3003)
	β	10.2651	(1.5259, 65.6214)
	μ	0.5562	(0.1136, 0.6591)
	σ	0.1375	(0.0927, 0.3071)
Dataset 3	α	0.0968	(0.0404, 1.9244)
	β	8.7691	(1.0524, 54.7217)
	μ	0.3191	(-0.2333, 0.4401)
	σ	0.1517	(0.0993, 0.4059)

Table 12 Summaries for the MOTLLN regression model for the breast cancer dataset

Parameter	τ_0	τ_1	α	β	σ	$-l(\psi)$	AIC
Estimates	5.6335	−1.2242	5.8946	0.1942	3.6718	154.3322	318.6644

Table 13 Summaries for the MOTLLN Bayesian regression model for the breast cancer dataset

Parameter	Mean	SD	MCSE	95% CI	Median
τ_0	5.6860	0.8617	0.5257	(3.7324, 6.8760)	5.8141
τ_1	−1.3559	0.5937	0.2304	(−2.7898, −0.4668)	−1.1944
α	5.8497	1.1934	0.5045	(3.2849, 7.8314)	6.2198
β	0.4406	0.6036	0.3287	(0.0469, 2.1098)	0.1931
σ	3.8589	0.7689	0.2764	(2.2360, 5.1247)	3.9895

Conflict of interest The authors declare no conflict of interest.

References

Aarset MV (1987) How to identify a bathtub hazard rate. IEEE Trans Reliab 36(1):106–108

Al-Shomrani A, Arif O, Ibrahim S, Hanif S, Shahbaz M (2016) Topp-Leone family of distributions: some properties and application. Pak J Stat Oper Res 12:443

Bernardeau F, Kofman L (1994) Properties of the cosmological density distribution function. Astrophys J 443:479

Blasi P, Burles S, Olinto A (1999) Cosmological magnetic field limits in an inhomogeneous universe. Astrophys J Lett 514:L79–L82

Castellares F, Montenegro L, Cordeiro G (2011) The beta log-normal distribution. J Stat Comput Simul 83:203–228

Chen G (1995) Generalized log-normal distributions with reliability application. Comput Stat Data Anal 19:309–319

Chesneau C, Irshad MR, Shibu DS, Nitin SL, Maya R (2022) On the Topp-Leone log-normal distribution: properties, modeling and applications in astronomical and cancer data. Chil J Stat 13(1):67–90. https://doi.org/10.32372/chjs.13-01-04

Collett D (2015) Modelling survival data in medical research. Chapman Hall CRC texts in statistical science. CRC Press, New York

Feigelson E, Babu GJ (2012) In modern statistical methods for astronomy: with R applications, Cambridge University Press

Gelman A, Hill J (2006) Data analysis using regression and multilevel/hierarchical models (Analytical methods for social research). Cambridge University Press, Cambridge

Gui W (2013) A Marshall-Olkin power log-normal distribution and its applications to survival data. Int J Stat Probab 2:63

Jobe J, Crow E, Shimizu K (1989) Lognormal distributions: theory and applications. Technometrics 31:392

Khaleel MA, Oguntunde PE, Abbasi JNA, Ibrahim NA, AbuJarad MH (2020) The Marshall-Olkin Topp Leone-G family of distributions: a family for generalizing probability models. Sci Afr 8:e00470

Kleiber C (2014) The generalized lognormal distribution and the Stieltjes moment problem. J Theor Prob 27:1167–1177

Leathem A, Brooks S (1987) Predictive value of lectin binding on breast-cancer recurrence and survival. The Lancet 329(8541):1054–1056

Marshall AW, Olkin I (1997) A new method for adding a parameter to a family of distributions with application to the exponential and Weibull families. Biometrika 84(3):641–652

Nantais JB, Huchra JP, Barmby P, Olsen KAG, Jarrett TH (2006) Nearby spiral globular cluster systems. I. Luminosity functions. Astronomical J 131(3):1416–1425

Ozel Kadilar G, Altun E, Alizadeh M, Mozafari M (2018) The odd log-logistic log-normal distribution with theory and applications. Adv Data Sci Adapt Anal 10:1850009

Parravano A, Sánchez N, Alfaro EJ (2012) The dependence of Prestellar core mass distributions on the structure of the parental cloud. Astrophys J 754:150

Pham A, Lai C-D (2007) On recent generalizations of the Weibull distribution. IEEE Trans Reliab 56:454–458

Sangsanit Y, Bodhisuwan W (2016) The Topp-Leone generator of distributions: properties and inferences. Songklanakarin J Sci Technol 38:537–548

Singh B, Sharma K, Rathi S, Singh G (2012) A generalized log-normal distribution and its goodness of fit to censored data. Comput Stat 27:51–67

Smith RL, Naylor JC (1987) A comparison of maximum likelihood and Bayesian estimators for the three-parameter Weibull distribution. J R Stat Soc Ser C Appl Stat 36(3):358–369

Topp C, Leone F (1955) A family of J-shaped frequency functions. J Am Stat Assoc 50:209–219

Toulias T, Kitsos C (2013) On the generalized lognormal distribution. J Prob Stat 1–15(07):2013

Upadhyay SK, Vasishta N, Smith AFM (2001) Bayes inference in life testing and reliability via Markov chain monte Carlo simulation. Sankhya Indian J Stat Ser A 63(1):15–40

Wasserman L (2006) All of nonparametric statistics. Springer texts in statistics. Springer, New York

Authors and Affiliations

Soman Latha Nitin[1] · Damodaran Santhamani Shibu[1] · Radhakumari Maya[1] · Christophe Chesneau[2] · Muhammed Rasheed Irshad[3] · Naushad Mamode Khan[4]

✉ Christophe Chesneau
christophe.chesneau@unicaen.fr

Soman Latha Nitin
nitinstat24@gmail.com

Damodaran Santhamani Shibu
statshibu@gmail.com

Radhakumari Maya
publicationsofmaya@gmail.com

Muhammed Rasheed Irshad
irshadmr@cusat.ac.in

Naushad Mamode Khan
n.mamodekhan@uom.ac.mu

[1] Department of Statistics, University College, Thiruvananthapuram, Kerala 695 034, India

2   Department of Mathematics, Université de Caen Basse-Normandie, UFR de Sciences,
    14032 Caen, France

3   Department of Statistics, Cochin University of Science and Technology, Cochin,
    Kerala 682 022, India

4   Department of Economics and Statistics, University of Mauritius, Moka, Mauritius