Proceedings of the 1st International Conference on Statistical Sciences and Data Analytics

LENGTH-BIASED TEISSIER DISTRIBUTION WITH PROPERTIES, ESTIMATION AND APPLICATION TO WATER QUALITY ANALYSIS

Musaddiq Sirajo[1], Abubakar Yahaya[1], Yahaya Zakari[1] and Abubakar Usman[1]

[1]Department of Statistics, Ahmadu Bello University, Zaria

Corresponding Author Email: musaddiqsirajo@gmail.com

ABSTRACT

In this paper, a new version of scaled Teissier distribution namely length biased scaled Teissier distribution is proposed and studied. The density function and its behavior, mode and non-central moments have been discussed and studied. The parameters of this distribution are estimated by the methods of moments and maximum likelihood. Finally, an application of the new model to some real-life and synthetic datasets is presented, and the performance compared with that of the traditional scaled Teissier distribution.

Keywords: Generalized integro-exponential function, Golden ratio, Likelihood method, Length-biased distributions, Water quality.

1. INTRODUCTION

One class of probability distributions that stood the test of time is the length-biased (LB) distributions; see Fisher (1934), Rao (1965), and Cox (1969). The LB distributions are a special case of weighted distributions, which have been derived from a length-biased sampling procedure (Patil, 2002). Traditional probability distributions assume independence and identical distribution of observations, which may not hold true with some real-life datasets. LB distributions address this limitation by incorporating the observation period or "length" into the distribution. LB probability distributions offer a valuable tool for improving the accuracy and uncertainty quantification in modeling, and research effort is geared towards exploring their full potential and integrating them with emerging methodologies.

The LB distributions have found applications in diverse areas including reliability analysis, environmental science and ecology. Patil and Rao (1978) gave a review of various classical distributions and their LB versions, while Khattree (1989) offered a tabular presentation of relationships existing among variables of some probability distributions and their LB versions. A comprehensive review of different LB distributions and their applications has been provided by Olyede and George (2002). Specifically, Leiva, Sanhueza and Angulo (2009) studied the LB version of Birnbaum-Saunders distribution and illustrated its application in water quality analysis. Reyad *et al*. (2013) discussed on the length-biased weighted Frechet distribution with properties and estimation. Seenoi *et al*. (2014) discussed on the length-biased exponentiated inverted Weibull distribution with various properties and applications. Modi and Gill (2015) obtained the length biased version of weighted Maxwell distribution with various statistical properties. Reyad *et al*. (2017) obtained the length biased weighted Frechet

distribution with properties and estimation. Karimi and Nasiri (2018) discussed the length biased weighted Lomax distribution in the presence of outliers. Rather and Subramanian (2018), discussed on length biased Sushila distribution with properties and applications. In recent times, Subramanian and Rather (2020) studied a new extension of Shanker distribution with real-life data.

The scaled Teissier (ST) distribution is an important positively skewed probability model with attractive properties; introduced by Jodra *et al*. (2015). Due to the interesting theoretical arguments established in its genesis, the ST model is appropriate for describing precipitation patterns. The arguments behind its inception have transformed the traditional Teissier distribution into a model appropriate to studying lifetimes. However, the ST model has been used beyond the lifetime analysis. Biçer *et al*. (2021) constructed a geometric process for modeling geometric count data when the distribution of the parameter of first occurrence time follows the ST distribution. Recently, Nascimento *et al*. (2023) in their analysis of financial market time series introduced a new autoregressive moving average (ARMA) process called the ST-ARMA model, which has the ST law as the marginal distribution that has its scale parameter serve as a proportion that can control amodal and unimodal behavior. The ST-ARMA model was used in capturing fluctuations in cryptocurrencies, which has played an increasingly important role in the global economy.

2. LENGTH BIASED ST (LBST) DISTRIBUTION

Length-biased sampling arises when the probability of selecting a unit is proportional to its size or duration, a situation common in reliability, environmental, and medical studies. In such contexts, classical models like the ST distribution may become inadequate due to the influence of bias introduced by the sampling mechanism. The ST distribution with shape κ and scale σ parameters has a probability density function (pdf) given by

$$f\left(x;\kappa,\sigma\right)=\frac{1}{\sigma}\left[\exp\left(\frac{\kappa}{\sigma}x\right)-\kappa\right]\exp\left\{\frac{\kappa}{\sigma}x-\frac{1}{\kappa}\left[\exp\left(\frac{\kappa}{\sigma}x\right)-1\right]\right\}, \quad (1)$$

$$\kappa\in\left(0,1\right],\ \sigma>0,\ \ x>0.$$

The corresponding cumulative distribution function (cdf) is given by

$$F\left(x;\kappa,\sigma\right)=1-\exp\left\{\frac{\kappa}{\sigma}x-\frac{1}{\kappa}\left[\exp\left(\frac{\kappa}{\sigma}x\right)-1\right]\right\},\quad x>0. \quad (2)$$

Let X be a non-negative random variable with pdf $f_X\ x$. The weighted version of X, say X_w, with weight function $w\ x$, has a probability distribution called the weighted distribution with pdf defined as

$$f_{X_w}\ x\ =\frac{w\ x\ f_X\ x}{E\left[w\ X\ \right]},\qquad x>0, \quad (3)$$

where $E\left[w\ X\ \right]<\infty.$

The LB distribution is obtained using equation (3) by setting $w\ x\ = x,$ making it a special case of the weighted distribution. The LB version of any distribution consists of the same number of parameters as in the original distribution. Let $f_X\left(x\right) = f\left(x;\kappa,\sigma\right)$ be the density of ST distribution and $E\left[w\ X\ \right] = \sigma$ in equation (3), the pdf of Y can be expressed, following Balakrishnan *et al*. (2011), as

$$f_Y\left(y;\kappa,\sigma\ \right) = \frac{y\ f\left(y;\kappa,\sigma\right)}{\sigma}, \quad y > 0, \tag{4}$$

where $w\ y\ = y,$ and $\sigma = E\ y\ ,$ so that Y is the LB version of X.

Let $X \sim \text{ST}\ \kappa,\sigma$. Then, the LB version of X, say Y, has an LB distribution, denoted by $LBST\ y;\kappa,\sigma$, with pdf expressed as

$$f_Y\left(y;\kappa,\sigma\right) = \frac{y}{\sigma^2}\left[\exp\left(\frac{\kappa}{\sigma}y\right) - \kappa\right]\exp\left\{\frac{\kappa}{\sigma}y - \frac{1}{\kappa}\left[\exp\left(\frac{\kappa}{\sigma}y\right) - 1\right]\right\}, \tag{5}$$
$$\kappa \in \left(0,1\right],\ \sigma > 0,\ \ y > 0.$$

The cdf of $LBST\ y;\kappa,\sigma$ distribution can be written in terms of the incomplete generalized integro-exponential function, which is defined by

$$E_r^q\ t;w\ = \frac{1}{\Gamma\ q+1}\int_1^w \log u\ ^q \exp\ -tu\ u^{-r}du \tag{6}$$

where $r \in\ -\infty,\infty\ ,\ q > -1$ and $\Gamma\ \cdot\ $ is the gamma function.

The generalized integro-exponential function is obtained as a limiting value of the incomplete generalized integro-exponential function

$$\lim_{w\to\infty} E_r^q\ t;w\ = E_r^q\ t\ .$$

Proposition 1. Let Y be a random variable having a $LBST\ y;\kappa,\sigma$ distribution. Then, Y has a cdf defined by

$$F_Y\ y;\kappa,\sigma\ = A_\kappa\left[E_{-1}^1\left(\frac{1}{\kappa};u\right) - \kappa E_0^1\left(\frac{1}{\kappa};u\right)\right], \quad y > 0, \tag{7}$$

where $A_\kappa = \frac{\exp\ 1/\kappa}{\kappa^2}$ and $u = \exp\left(\frac{\kappa}{\sigma}y\right).$

Proof. From (5), the pdf can be rewritten as

$$f_Y\left(y;\kappa,\sigma\right)=\frac{\exp\left(1/\kappa\right)}{\sigma^2}y\left[\exp\left(\frac{\kappa}{\sigma}y\right)-\kappa\right]\exp\left\{\frac{\kappa}{\sigma}y-\frac{1}{\kappa}\left[\exp\left(\frac{\kappa}{\sigma}y\right)\right]\right\}.$$

The cdf defined by the integral representation $F_Y\ y;\kappa,\sigma\ =\int_0^u f\ u\ du,$ can be derived by making change of variable $u=\exp\left(\frac{\kappa}{\sigma}y\right).$ Thus,

$$F_Y\left(y;\kappa,\sigma\right)=A_k\left[\int_1^u \ln u\exp\left(-\frac{1}{\kappa}u\right)udu-\kappa\int_1^u \ln u\exp\left(-\frac{1}{\kappa}u\right)du\right].$$

Using (6), the cdf may be expressed as

$$F_Y\ y;\kappa,\sigma\ =A_k\left[\Gamma\ 2\ E_{-1}^1\left(\frac{1}{\kappa};u\right)-\kappa\Gamma\ 2\ E_0^1\left(\frac{1}{\kappa};u\right)\right].$$

taking into account $\Gamma\ 2\ =\ 2-1\ !=1,$ the result follows.

The hazard rate function of Y is easily obtained from equations (5) and (7) as

$$h_Y\ y;\kappa,\sigma\ =\frac{f_Y\ y;\kappa,\sigma}{1-F_Y\ ly;\kappa,\sigma},\qquad y>0. \tag{8}$$

When $y\to 0,$ the numerator in (8) is small because the function is close to zero at very low values of y; i.e. $f_Y\ y;\kappa,\sigma\ \to 0,$ and the denominator $1-F_Y\ y;\kappa,\sigma$ is close to 1. It then follows that $h_Y\ y;\kappa,\sigma\ \sim f_Y\ y;\kappa,\sigma\ \to 0,$ which starts small. When $y\to\infty,$ the pdf $f_Y\ y;\kappa,\sigma$ increases to a peak and the cdf $F_Y\ y;\kappa,\sigma$ approaches 1, making $1-F_Y\ y;\kappa,\sigma$ very small. It then follows that $h_Y\ y;\kappa,\sigma$ is a strictly increasing function of y.

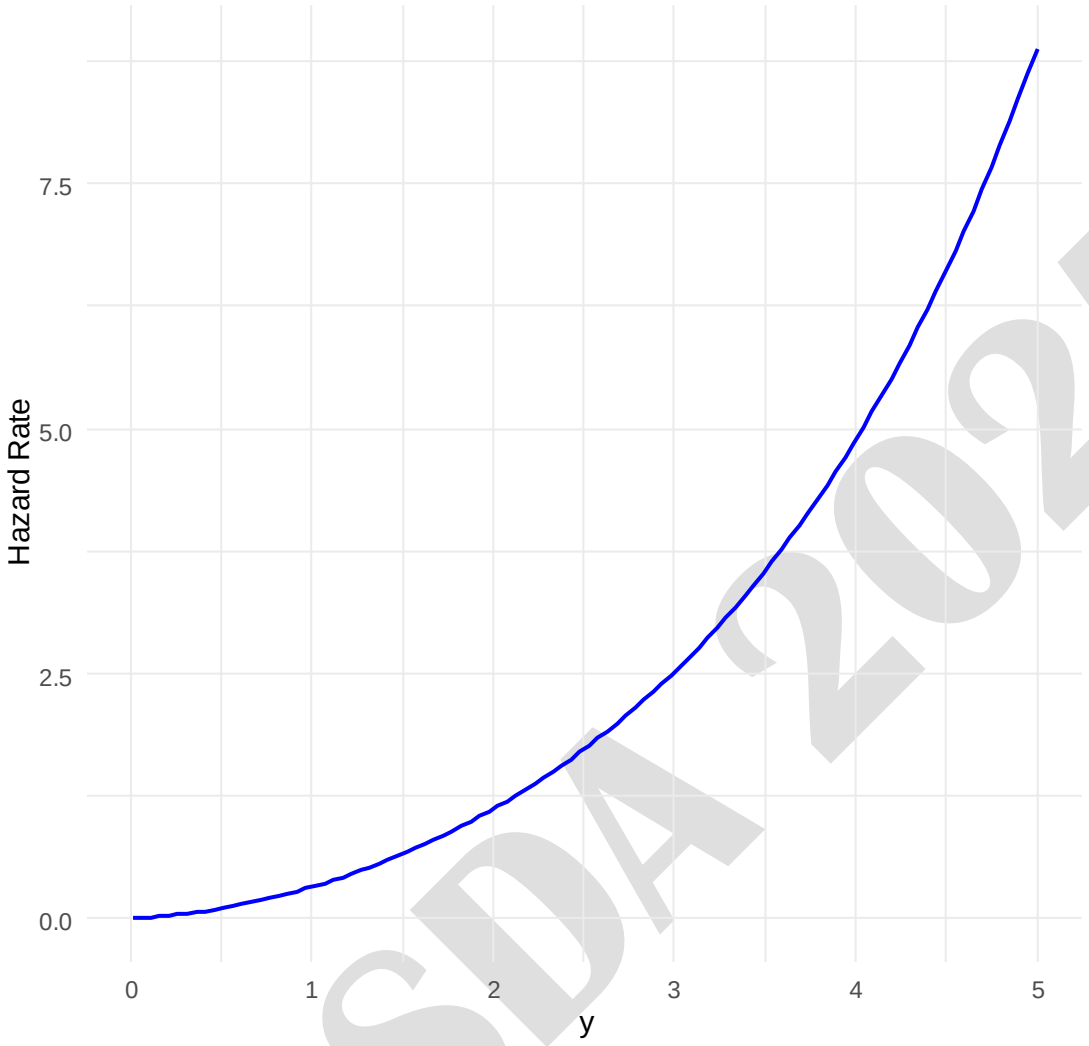

Figure 1: LBST hazard rate

Figure 1 shows that $LBST\ \ y;\kappa,\sigma$ behaves similarly to an exponential-type distribution. In this case, $h_Y\ \ y;\kappa,\sigma$ increases with y, meaning the likelihood of event occurrence grows over time. This suggests an aging process, where older entities are more likely to experience failure (e.g., machine wear-out, human mortality). If the LBST distribution models rainfall data, an increasing hazard rate implies that as a rainy period progress, the chance of continued rain increases.

The remainder of this paper is organized as follows. Section 3 studies some statistical properties of the LBST distribution. More precisely, the mode is derived in terms of golden ratio and the non-central moments expressed in terms of the generalized integro-exponential function. The quantile function is written in terms of Lambert W function; a function that has found nice application in statistics for the expression of quantile functions of some probability laws. The parameter estimation problem is considered in Section 4 and a simulation study carried out to assess the performance of maximum likelihood method as compared to method of moments. Section 5 investigates a Likelihood Ratio Test

(LRT) that compares the ST distribution against its Length-Biased Scaled Teissier (LBST) counterpart. In Section 6, application to real datasets is presented to illustrate the practical usefulness of the proposed distribution and Section 7 concludes the paper.

3. STATISTICAL PROPERTIES

3.1 Mode

The mode of the LBST distribution is the value of y that maximizes $f_Y(y;\kappa,\sigma)$, which can be found by solving $f_Y'(y;\kappa,\sigma)=0$. The next result shows that the $LBST\ \ \kappa,\sigma$ distribution is unimodal and that golden ratio φ is embedded in the structure of the mode-defining equation.

Proposition 2. Let Y be a random variable having a $LBST\ \ \kappa,\sigma$ distribution with $\sigma>0$ and $\kappa\in\ 0,1$. Then Y has a unique mode

$$\text{Mode } Y\ =\frac{\sigma}{\kappa}\ \ln\kappa+2\ln\varphi\ , \qquad (9)$$

where φ is the positive solution of the equation $x^2-x-1=0$, given by $\varphi=\frac{1+\sqrt{5}}{2}\approx 1.618034$.

Proof. In order to obtain the mode of Y, the equation $\partial/\partial y\ \ f_Y\ \ y;\kappa,\sigma\ =0$ must be solved with respect to y. This is tantamount to solving with respect to x the equation $e^x-\kappa\ ^2=\kappa e^x$, with $x=\frac{\kappa y}{\sigma}$. Solving the quadratic equation in exponential x and choosing the positive root, the result follows.

Figure 2 shows that the LBST distribution is unimodal and the probability of observing larger values compared to the standard ST distribution is increased, making it suitable for modeling data with over-representation of larger values. The exact peak and spread depend on the shape and scale parameters. Larger shape parameters shift the peak rightward and make the tail heavier. The LBST distribution would be particularly useful in real-life applications where larger values have an increased probability of being observed, such as rainfall intensities, biological growth, and reliability modeling.

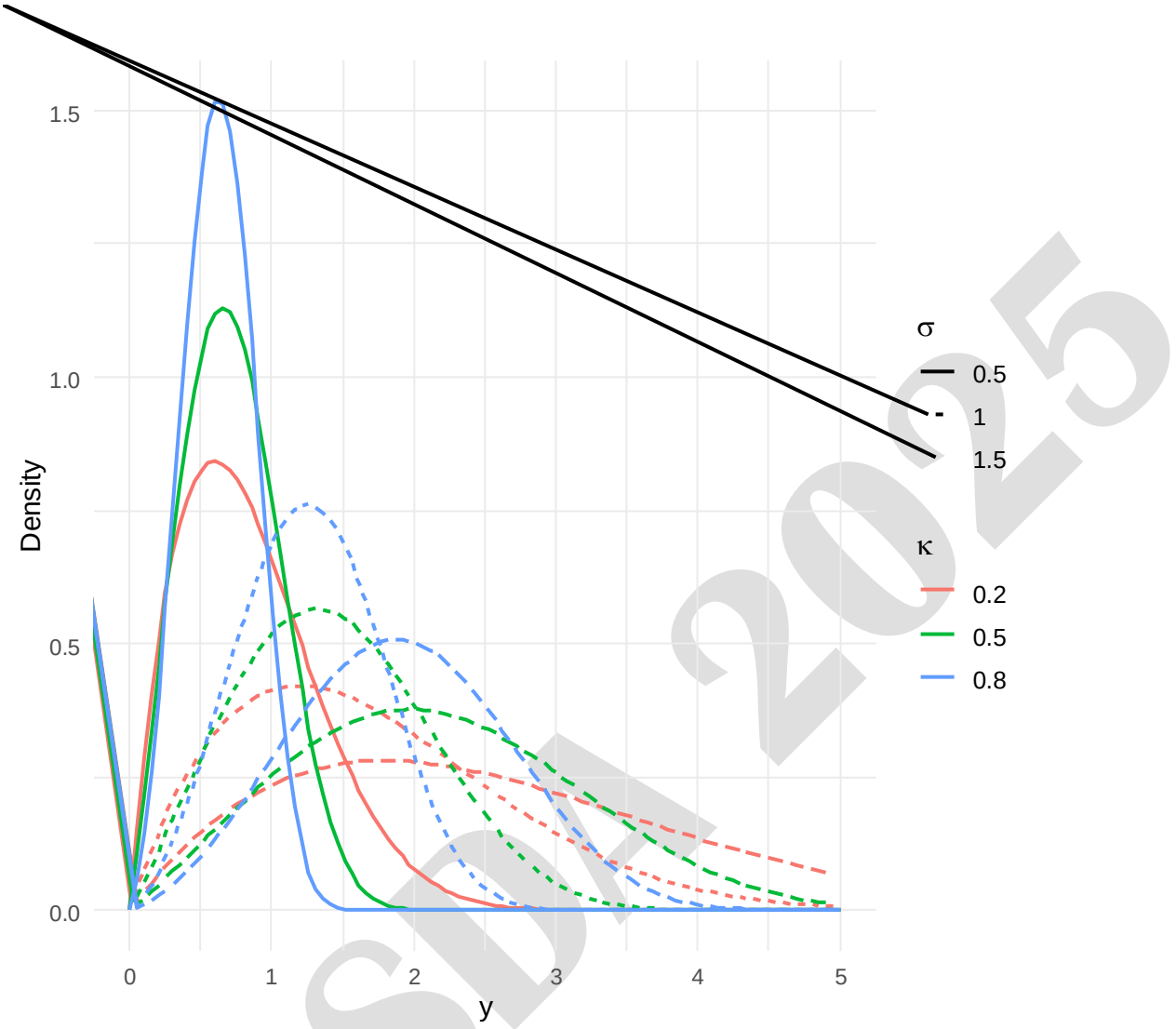

Figure 2: Length-Biased Scaled Teissier distribution

3.2 Moments

The moment properties of the LBST distribution are discussed in terms of the upper incomplete gamma function. The moments $E\left[Y^r\right]$ can be related through integration to the moments of the original ST distribution $E\left[X^r\right]$. The moments of Y may be obtained as follows

$$E\left[Y^r\right]=\frac{E\left[X^{r+1}\right]}{E\ X}=\frac{1}{\sigma}E\left[X^{r+1}\right]. \qquad (10)$$

Proposition 3. Let Y have the $LBST\ \ \kappa,\sigma$ distribution, then

(i) $E\ Y\ = \frac{2\sigma \exp\ 1/\kappa}{\kappa}\Gamma\ 0,1/\kappa\ ,$ (ii) $E\left[Y^2\right] = \frac{2\sigma^2 \exp\ 1/\kappa}{\kappa}\Gamma\ 0,1/\kappa\ .$

Proof. (i) The result follows from (10) because $E_0^0\ 1/\kappa\ = \kappa \exp\ -1/\kappa\ ,\ E_0^1\ 1/\kappa\ = \kappa\Gamma\ 0,1/\kappa$ and

$$E\left[X^r\right] = \frac{\sigma^r \exp\ 1/\kappa\ \Gamma\ r+1}{\kappa^r} E_0^{r-1}\ 1/\kappa\ . \tag{11}$$

(ii) From (11), since $E\left[X^3\right] = \frac{2\sigma^3 \exp\ 1/\kappa}{\kappa}\Gamma\ 0,1/\kappa\ ,$ the result follows by virtue of (10).

3.3 Quantile Function

It is well known that for length-biased distributions, the cdf $F_Y\ y$ can be inverted numerically using the inverse transform sampling to obtain the quantile function $Q_Y\ u\ ,$ which is defined as the value of y such that $F_Y\ y\ = u.$ Then, the quantile function of the LB distribution is

$$Q_Y\ u\ = Q_X\ \nu\ , \tag{12}$$

where $\nu = H^{-1}\ u\ ,$ and $H\ u$ is the size-biased transform of the uniform, defined as

$$H\ u\ = \frac{1}{\mu}\int_0^u Q_X\ s\ ds. \tag{13}$$

For LBST distribution, the base quantile is

$$Q_X\ u;\kappa,\sigma\ = \sigma Q\ u;\kappa\ , \tag{14}$$

where $Q\ u;\kappa\ = \frac{1}{\kappa}\log\ 1-u\ - \frac{1}{\kappa}W_{-1}\left(\frac{u-1}{\kappa \exp\ 1/\kappa}\right) - \frac{1}{\kappa^2}, \qquad 0 < u < 1.$

Proposition 4. The quantile function of the $LBST\ \ \kappa,\sigma$ distribution is

$$Q_Y\ u;\kappa,\sigma\ = \sigma Q\ H^{-1}\ u\ ;\kappa\ , \qquad 0 < u < 1. \tag{15}$$

where W_{-1} denotes the negative branch of the Lambert W function.

Proof. For any $0<u<1$, the size-biased transform of the uniform is as defined in (13). Using $\mu=\sigma$ and the base quantile (14) in (13), $H\ u;\kappa\ =\int_0^u Q\ s;\kappa\ ds \Rightarrow F_Y\ y\ =H\ F_X\ y\ ;\kappa$, so that $H^{-1}\ u$ is the inverse of the integral function $H\ u;\kappa$. Now, using (12), the result follows.

3.4. PARAMETER ESTIMATION OF LBST

In this section, the following methods to estimate the parameters were discussed; maximum likelihood and moments, which are presented in subsections 4.1 and 4.2 respectively. As it will be seen, the estimators cannot be obtained in closed form so their performance must be assessed via a Monte Carlo simulation study, which is given in subsection 4.3.

3.4.1 Method of Maximum Likelihood

The maximum likelihood estimates (MLEs) of the parameters of the LBST distribution can be obtained based on a random sample $Y_1, Y_2, ..., Y_n$; where $Y_i \sim LBST\ \kappa,\sigma$ for $1,2,...,n.$ The log-likelihood function is given by

$$\ell\ \kappa,\sigma\ =\sum_{i=1}^{n}\log y_i-2n\log\sigma+\sum_{i=1}^{n}\log\left(\exp\left(\frac{\kappa}{\sigma}y_i\right)-\kappa\right)+\frac{\kappa}{\sigma}\sum_{i=1}^{n}y_i -\frac{1}{\kappa}\sum_{i=1}^{n}\left(\exp\left(\frac{\kappa}{\sigma}y_i\right)-1\right). \tag{16}$$

To find the MLEs, it is required to differentiate $\ell\ \kappa,\sigma$ with respect to κ and σ and set the equations to zero.

$$\frac{\partial\ell\ \kappa,\sigma}{\partial\sigma}=-\frac{2n}{\sigma}-\frac{\kappa}{\sigma^2}\sum_{i=1}^{n}y_i\exp\left(\frac{\kappa}{\sigma}y_i\right)\cdot\frac{1}{\exp\left(\frac{\kappa}{\sigma}y_i\right)-\kappa}-\frac{\kappa}{\sigma^2}\sum_{i=1}^{n}y_i +\frac{1}{\sigma^2}\sum_{i=1}^{n}y_i\exp\left(\frac{\kappa}{\sigma}y_i\right), \tag{17}$$

$$\frac{\partial\ell\ \kappa,\sigma}{\partial\kappa}=\sum_{i=1}^{n}\frac{1}{\exp\left(\frac{\kappa}{\sigma}y_i\right)-\kappa}\cdot\left(\frac{y_i}{\sigma}\exp\left(\frac{\kappa}{\sigma}y_i\right)-1\right)+\frac{1}{\sigma}\sum_{i=1}^{n}y_i -\frac{1}{\kappa\sigma}\sum_{i=1}^{n}y_i\exp\left(\frac{\kappa}{\sigma}y_i\right)+\frac{1}{\kappa^2}\sum_{i=1}^{n}\exp\left(\frac{\kappa}{\sigma}y_i\right)-\frac{n}{\kappa^2}, \tag{18}$$

The two equations (17) and (18) can be solved iteratively to obtain the MLEs $\hat{\kappa}$ and $\hat{\sigma}$.

3.4.2 Method of moments

Let $C = \exp\ 1/\kappa\ \Gamma\ 0,1/\kappa\ ,$ then the first and second moments of LBST can be rewritten respectively as $E\ Y\ = \frac{2\sigma C}{\kappa}$ and $E\left[Y^2\right] = \frac{2\sigma^2 C}{\kappa}$. The sample mean and variance are respectively defined as $\bar{Y} = \frac{1}{n}\sum Y_i$ and $S_Y^2 = \frac{1}{n}\sum\ Y_i - \bar{Y}\ ^2$. The Method of Moments equates the theoretical moments with sample moments.

$\bar{Y} = \frac{\sigma C}{\kappa},$ so that $\sigma = \frac{\kappa \bar{Y}}{C}\cdot$

Solving for κ using the variance equation,

$S_Y^2 = \frac{2\sigma^2 C}{\kappa} - \left(\frac{\sigma C}{\kappa}\right)^2.$

This leads to quadratic equation in κ,

$S_Y^2\kappa^2 - 2\sigma^2 C\kappa + 4\sigma^2 C^2 = 0.$

Using the quadratic formula,

$\kappa = \frac{2\sigma^2 C \pm \sqrt{4\sigma^4 C^2 - 16\sigma^2 C^2 S_Y^2}}{2S_Y^2}$

or equivalently,

$\kappa = \frac{\sigma^2 C \pm \sigma C\sqrt{\sigma^2 - 4S_Y^2}}{S_Y^2} \Rightarrow \kappa = \frac{\kappa^2\bar{Y}^2/C\ \pm \kappa\bar{Y}\sqrt{\kappa^2\bar{Y}^2/C^2\ - 4S_Y^2}}{S_Y^2}.$

Since $\kappa \in\ 0,1\ ,$ the valid root must be chosen and finally $\hat{\kappa}$ be substituted into $\sigma = \frac{\kappa\bar{Y}}{2C}$ to get $\hat{\sigma}$.

4. Simulation Study

Tables 1 and 2 summarized the parameter estimation performance of the method of moments (MOM) and maximum likelihood estimation (MLE) for the LBST model across varying sample sizes. The performance is evaluated based on statistics like bias, variance**,** and mean squared error (MSE).

Table 1: MOM Estimates

	$\kappa = 0.8$			$\sigma = 1.5$		
n	*Bias* $\hat{\kappa}$	*Var* $\hat{\kappa}$	*MSE* $\hat{\kappa}$	*Bias* $\hat{\sigma}$	*Var* $\hat{\sigma}$	*MSE* $\hat{\sigma}$

50	0.1999	0.0000	0.0400	0.6918	0.0119	0.4905
100	0.1999	0.0000	0.0400	0.6994	0.0055	0.4947
200	0.1999	0.0000	0.0400	0.7017	0.0029	0.4953
500	0.1999	0.0000	0.0400	0.7038	0.0011	0.4964
1000	0.1999	0.0000	0.0400	0.7028	0.0005	0.4945

From Table 1, all the statistics for estimating κ remain constant for all sample sizes, indicating that the MOM estimator is consistently biased for this parameter. Surprisingly, the bias increases slightly with sample size for σ estimates. This contradicts the expected behavior, though the variance decrease as sample size increases, suggesting a limitation in the efficiency of the MOM estimator for this model. The MSE of σ estimates remain relatively stable (approximately 0.49), showing no significant improvement in accuracy as the sample size increases. These results suggest that the MOM approach is not consistent or efficient for estimating parameters of the LB-ST distribution. It fails to leverage the benefits of larger sample sizes, making it less suitable for practical inference in this context.

Table 2: MLE Estimates

	$\kappa = 0.8$			$\sigma = 1.5$		
n	*Bias* $\hat{\kappa}$	*Var* $\hat{\kappa}$	*MSE* $\hat{\kappa}$	*Bias* $\hat{\sigma}$	*Var* $\hat{\sigma}$	*MSE* $\hat{\sigma}$
50	0.0150	0.0237	0.0239	-0.0013	0.0180	0.0180
100	0.0148	0.0158	0.0160	0.0056	0.0111	0.0112
200	0.0092	0.0088	0.0088	0.0039	0.0060	0.0061
500	0.0026	0.0033	0.0033	0.0012	0.0023	0.0022
1000	0.0004	0.0018	0.0018	-0.0009	0.0013	0.0013

From Table 2, the bias decreases steadily as the sample size increases, approaching zero. This shows that the MLE estimator is asymptotically unbiased for both parameters. The variance also decreases significantly with sample size, consistent with theoretical expectations. The MSE drops sharply, indicating improved estimation accuracy with more data. For example, the MSE for estimating the scale parameter drops from 0.0180 (n = 50) to 0.0013 (n = 1000). MLE demonstrates superior statistical properties: it is consistent, asymptotically normal, and efficient. The declining MSE confirms that estimation improves reliably with larger samples.

5. LIKELIHOOD RATIO TEST

Let $x_1, x_2, ..., x_n$ be a random sample of size n from either the ST or LB-ST distribution. The hypotheses for the LRT are

$H_0 : x_i \sim ST\ \kappa, \sigma$,

$H_1 : x_i \sim LB-ST\ \kappa, \sigma$,

where $\kappa \in$ 0,1 and $\sigma > 0$. The LRT statistic is given by

$$\Lambda = -2\log\left(\frac{\sup_{\kappa,\sigma} L_0\ \kappa,\sigma}{\sup_{\kappa,\sigma} L_1\ \kappa,\sigma}\right) = -2\ \ell_0\ \hat{\kappa}_0,\hat{\sigma}_0\ -\ell_1\ \hat{\kappa}_1,\hat{\sigma}_1\ \ , \tag{19}$$

where $\hat{\kappa}_0,\hat{\sigma}_0$ and $\hat{\kappa}_1,\hat{\sigma}_1$ are the MLEs under H_0 and H_1 respectively, obtained using the respective log-likelihoods

$$\ell_0\ \kappa,\sigma\ = -n\log\sigma + \sum_{i=1}^{n}\log\left(\exp\left(\frac{\kappa}{\sigma}x_i\right)-\kappa\right) + \frac{\kappa}{\sigma}\sum_{i=1}^{n}x_i - \frac{1}{\kappa}\sum_{i=1}^{n}\left(\exp\left(\frac{\kappa}{\sigma}y_i\right)-1\right),$$

and

$$\ell_1\ \kappa,\sigma\ = -2n\log\sigma + \sum_{i=1}^{n}\log x_i + \sum_{i=1}^{n}\log\left(\exp\left(\frac{\kappa}{\sigma}x_i\right)-\kappa\right) + \sum_{i=1}^{n}\left[\frac{\kappa}{\sigma}x_i - \frac{1}{\kappa}\left(\exp\left(\frac{\kappa}{\sigma}y_i\right)-1\right)\right].$$

If Λ is large (i.e., log-likelihood under LBST is much higher), H_0 is rejected in favor of H_1, suggesting data is better described by the length-biased model.

Under standard regularity conditions, the LRT statistic in (19) approximately follows a chi-square distribution with degrees of freedom equal to the difference in number of parameters under H_0 and H_1. Since both models have the same number of parameters, the test is not a nested model in the usual sense**.** The LBST distribution is a weighted version of ST, so standard chi-squared approximation may not be valid**.** Instead, parametric bootstrap is therefore employed for computing the p-value of the LRT.

A Monte Carlo simulation study was conducted to evaluate the performance of the LRT under both the null and alternative hypotheses. The number of simulation replicates is 500. Samples were generated from the ST distribution and the LB-ST distribution. For each simulated sample, the ST and LB-ST models were fitted using MLE, and the LRT statistic was computed across replications. The empirical size and power of the test were then estimated at 5% level of significance. The size was defined as the proportion of rejections when the true model was the ST distribution, while the power was defined as the proportion of rejections when the true model was the LB-ST distribution.

Table 3 summarizes the empirical performance of the LRT. The empirical size matches the nominal level exactly, suggesting that the LRT correctly controls the Type I error under the null. However, the empirical power is relatively low at small sample sizes, indicating a limited ability to detect length bias when present, at these sample sizes. A substantial increase in power is observed at *n = 150* (power = 0.59), and the test becomes highly effective at *n = 200*, where the power reaches 0.973, indicating near-certain rejection of the null hypothesis when the LB-ST model is correct.

Table 3: Empirical power and size

Sample size (n)	Empirical power	Empirical type I error (size)
30	0.11	0.050
50	0.083	0.050
70	0.273	0.050
100	0.173	0.050

150	0.59	0.050
200	0.973	0.050

The power curve in Figure 3 illustrates that the LRT is underpowered in small samples, but becomes highly reliable in larger samples, specifically $n \geq 150$**.** The horizontal dashed line marks the conventional target power level of 0.8. The curve intersects this threshold just before n = 200, implying that a sample size of at least 180–200 may be required for the LRT to achieve adequate power in distinguishing LB-ST from ST.

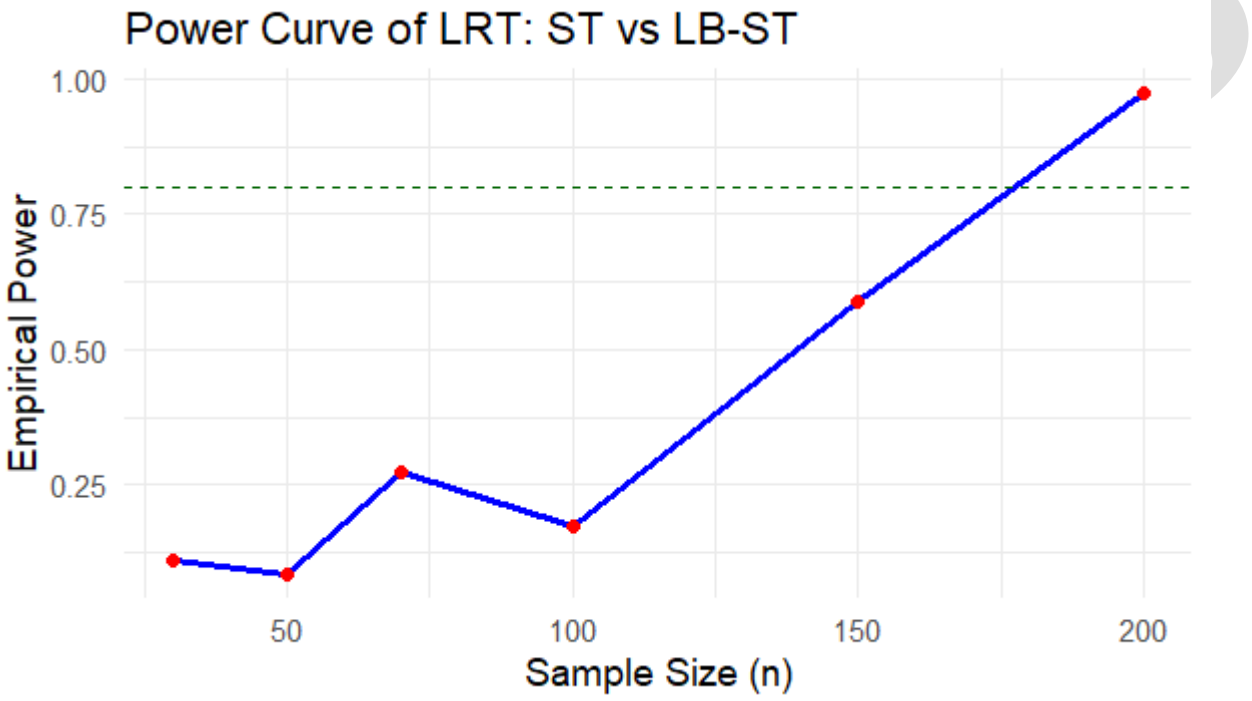

Figure 3: Detecting LB-ST Distribution

6. APPLICATION

To illustrate its practical usefulness, the LBST distribution is applied to water quality data involving dissolved oxygen (DO) concentrations. DO is an important indicator of water quality and ecosystem health, as it affects the survival of aquatic organisms. The dataset used in this application was obtained from the publicly available U.S. Geological Survey (USGS) National Water Information System (NWIS) database (https://waterdata.usgs.gov). The dataset in Table 4 consist of DO concentrations (in mg/L) recorded at various river monitoring stations.

Table 4: DO concentrations (in mg/L)

8.2	7.5	6.9	7.8	9.1	6.5
8.6	7.2	6.8	8.0	9.5	7.1
6.6	7.4	8.3	7.0	6.7	8.4
7.3	8.1	9.0	6.4	7.9	6.3
8.8	7.6	8.7	6.2	9.2	7.7

Both the ST and LBST models were fitted to the data using maximum likelihood estimation. Model performance was compared using the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). The formulae for calculating AIC and BIC are

$$AIC = 2k - 2\ell, \qquad BIC = k \cdot \ln\ n\ - 2\ell, \tag{20}$$

where k is the number of parameters (which is 2 for both ST and LBST models) and n is the sample size (in this case, 30). From Table 5, it has been observed that the LBST distribution have the lesser AIC and BIC values as compared to ST distribution.

Table 5: Estimates, AIC and BIC (real-world dataset)

Model	κ **(Shape)**	σ **(Scale)**	ℓ **(log-likelihood)**	**AIC**	**BIC**
ST	0.1502	7.0031	-39.267	82.534	85.344
LBST	0.2089	7.3987	-37.010	78.020	80.830

The results in Table 5 show that LBST model has a higher log-likelihood, indicating a better fit. Because the LBST model yields lower AIC and BIC values compared to the standard ST model, it suggests that the length-biased structure of the LBST model captures the data's distribution more effectively, accounting for possible overrepresentation of higher DO values. This evidence confirms that the LBST model is more suitable for analyzing DO data, where observational biases such as preferential recording of larger values may occur. The improved fit also supports its relevance in ecological and hydrological studies. Hence it can be concluded that the LBST distribution leads to a better fit than the ST distribution.

To further illustrate the application of the LBST model, we analyze a relatively large dataset (n = 160) of DO concentrations. The dataset used here is synthetically generated to mimic potential patterns observed in real-world DO measurements. Both the ST and LBST models were fitted to the data using maximum likelihood estimation.

Table 6: Estimates, AIC and BIC (synthetic dataset)

Model	κ **(Shape)**	σ **(Scale)**	ℓ **(log-likelihood)**	**AIC**	**BIC**
ST	0.0960	6.0011	-534.9380	1073.8761	1080.0264
LBST	0.0960	6.0011	-571.6730	1147.3460	1153.4963

In this dataset, the ST model outperformed the LBST model, as indicated by a higher log-likelihood and lower AIC and BIC values. This result contrasts with the small-sample example, highlighting that the advantage of the LBST model is context-dependent and more pronounced when there is significant length bias in the data.

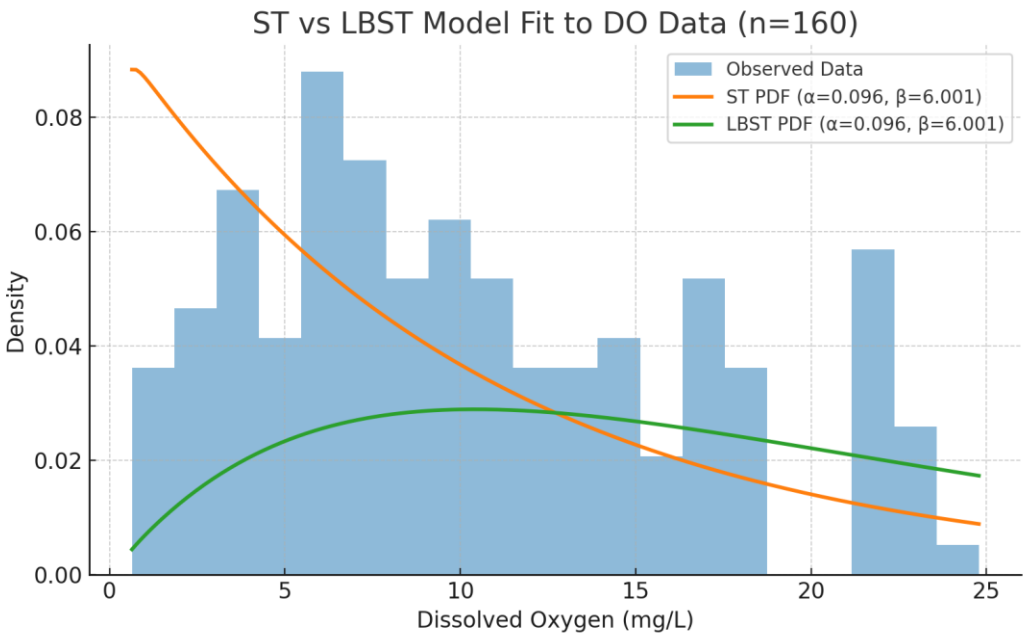

Figure 4: Histogram of observed dissolved oxygen data (n=160) with overlaid fitted PDFs for the ST and LBST models.

7. CONCLUSION

The LB distributions provide a unifying approach for confronting problems of model specification and data interpretation in traditional environmetric theory and practice. Failure to make such adjustments can lead to incorrect conclusions. In this article, the LB version of the ST distribution was developed. This new model turns out to be quite flexible for modeling water quality. The density, distribution function, hazard function, moments and some plots were provided. Also, some properties of this new model were obtained. By using likelihood method, statistical inference was carried out on the parameters of the new model, and likelihood ratio test discussed. An application to real data showed that this new model is a flexible alternative to the ST distribution. Here, it is postulated that the LBST distribution is appropriate for modeling water quality. In the synthetic example, possible dependencies within the observations have been taken into account. This represents an important aspect to be considered in water quality analysis.

REFERENCES

Biçer, C., Bakouch, H.S. & Biçer, H.D. (2021). Inference on parameters of a geometric process with scaled Muth distribution. Fluct. Noise Lett., 20, 2150006.

Cox D. R. (1969). Some sampling problems in technology, In New Development in Survey Sampling, Johnson, N. L. and Smith, H., Jr .(eds.) New York Wiley- Interscience, 506-527.

Fisher, R. A. (1934). The effects of methods of ascertainment upon the estimation of frequencies. Annals of Eugenics, 6, 13-25.

Jodra, P., Jiménez-Gamero, M. D., & Alba-Fernandez, M. V. (2015). On the Muth distribution. *Mathematical Modelling and Analysis*, *20*(3), 291-310.

Khattree, R., 1989. Characterization of inverse Gaussian and gamma distributions through their length biased versions. IEEE Trans. Rel. 38, 610-611.

Karimi, H. and Nasiri P. (2018), Estimation Parameter of R = P(Y < X) for Length-Biased Weighted Lomax Distributions in the Presence of Outliers, Mathematical and Computational Applications,; doi:10.3390/mca23010009

Leiva, V., Sanhueza, A. & Angulo, J.M. (2009). A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stoch. Environ. Res. Risk Anal*. 23, 299-307.

Nascimento, A. D., Lima, M. C., Bakouch, H., & Qarmalah, N. (2023). Scaled Muth–ARMA Process Applied to Finance Market. *Mathematics*, *11*(8), 1908.

Olyede, B.O. & George, E.O. (2002). On stochastic inequalities and comparisons of reliability measures for weighted distributions. Math. Prob. Eng. 8, 1-13.

Patil, G.P. (2002). Weighted distributions. In Encyclopedia of Environmetrics, 4, 2369 - 2377. Wiley, New York..

Patil, G.P. & Rao, C.R. (1978). Weighted distributions and size-biased sampling with applications to wildlife populations and human families. Biometrics 34, 179-189.

Rao, C.R. (1965). On discrete distributions arising out of methods of ascertainment. In Classical and Contagious Discrete Distributions (Patil, G.P. ed.). Pergamon Press and Statistical Publishing Society, Calcutta, pp. 320-332. (Also in Sankhya A, 27, 311-324)

Rather, A. A. & Subramanian, C. (2018). Length Biased Sushila Distribution, Universal Review, volume 7, issue XII, pp 1010-1023.

Reyad, H. M., Hashish, A. M., Othman, S. A. & Allam, S. A. (2013). The Length-Biased Weighted frechet Distribution: properties and estimation, International Journal of Statistics and Applied Mathematics, 3(1): 189-200.

Reyad, H. M., Othman, S. A. & Moussa, A. A. (2017). The Length-Biased Weighted Erlang Distribution, Asian Research Journal of Mathematics, 6(3): 1-15.

Seenoi, P., Supapakorn, T. & Bodhisuwan, W., (2014), The Length-Biased Exponentiated Inverted Weibull Distribution, International Journal of Pure and Applied Mathematics, Volume 92, No. 2, 191-206.

Shanker, R. & Shukla, K. K. (2018). OM Distribution with Properties and Applications. *RT&A, No. 4 (51)*, Volume 13, pp 27-41.

Subramanian, C. & Rather, A. A. (2020). A New Extension of Shanker distribution with Real Life Data, Journal of Xi'an University of Architecture & Technology, volume XII, Issue III, pp2001-2009.