Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 8** : 1–27.
ISSN-0037-8712
doi:10.5269/bspm.79632

On the Estimation of the Wald Distribution Parameters with Diverse Applications

Chinyere P. Okechukwu, Manzoor A. Khanday, Okechukwu J. Obulezi*, Mohamed A. F. Elbarkawy, Ehab M. Almetwally and Mohammed Elgarhy

ABSTRACT: This paper evaluates various parameter estimation methods for the Wald distribution using simulations and real-world datasets. Simulation results confirm that increased sample sizes improve estimates, reducing bias and Root Mean-Squared Error (RMSE). The Maximum Likelihood Estimator (MLE) is generally the most robust method for large samples but unstable and biased for smaller ones, particularly in estimating λ. The Maximum product of spacing estimation (MPS) method performs well asymptotically, with bias and RMSE decreasing as sample size increases. Least Squares (LS) and Weighted Least Squares (WLS) are suitable alternatives to MLE for small-to-moderate samples, showing similar estimates and reduced bias with larger samples. The Cramer-von Mises Estimator (CvM) displayed the worst efficiency due to high RMSE. Bayesian Estimators (BE) showed greater bias and lower efficiency than their frequentist counterparts, especially for λ, with performance strongly dependent on prior selection. Application to real datasets (HIV/AIDS mortality, COVID-19 death rates, and industrial gauge measurements) demonstrates the Wald distribution's feasibility in diverse health data analysis and reliability studies. The study concludes that MLE and MPSE are efficient estimation methods, suggesting the Wald distribution is a strong candidate for applied parameter estimation. Future work could focus on improving Bayesian methods via better prior selection.

Keywords: Goodness of fit, COVID-19 mortality rate, estimation, HIV/AIDS mortality rate, reliability analysis, Wald distribution.

Contents

1. Introduction

The Wald distribution, also known as the inverse Gaussian distribution, is a continuous probability distribution that is widely used in various scientific fields such as finance, survival analysis, reliability engineering, and environmental modeling. The inverse Gaussian is a skewed, two-parameter continuous distribution whose density is similar to the Gamma distribution with greater skewness and a sharper

* Corresponding author.
2020 *Mathematics Subject Classification:* 62F10.
Submitted October 18, 2025. Published March 12, 2026

peak. It was introduced by Schrödinger in 1915 and further developed by Wald in 1944, the distribution is particularly useful for modeling positive data with a right-skewed distribution, making it ideal for applications where the variable of interest cannot take negative values ([23,20]). The significance of this distribution lies in its ability to model first-passage time processes, particularly in Brownian motion with positive drift. This property makes it essential in stochastic processes, particularly in modeling survival times, stock market returns, and equipment failure rates. Additionally, the inverse Gaussian distribution has proven effective in handling over-dispersed data in count models and positively skewed data sets. Weighted versions of the inverse Gaussian distribution have further enhanced its application by improving robustness against outliers, overdispersion, and data heterogeneity ([9]). These weighted models assign different weights to observations, allowing for better adaptability in complex data structures. Furthermore, parameter estimation techniques like Maximum Likelihood Estimation (MLE), the Method of Moments, and Bayesian methods have improved the accuracy and efficiency of using this distribution in real-world scenarios ([20]). [1] considered the Inverse Gaussian distribution whose variance is proportional to the mean and compared the length of bias for MVUE and MLE in terms of their variances and mean square errors while maintain that the data were simulated from IGD(μ, c, μ^2) type of environment. [18] proposed a new family of generalized distributions based on logistic-x transformation and showed how useful the new family is in the presence of big data and extreme values. [7] studied the presentation of the statistical properties, methods and applications of two parameters IGD and concluded that the distribution is a valuable tool to both statistician and non-statistician alike. [19] asserted that wrong specification can lead to invalid inference giving rise to a third kind of error and he introduced the concept of weighted distributions as a method of adjustment applicable to many situations. In his paper, he studied the relationship between the weighted distributions and the parent distributions in the context of reliability and life testing. [19] confirmed that these relationships depended on the nature of the weight function and give rise to interesting connections between the different aging criteria of the two distributions.

[11] carried out a Monte Carlo study to examine the efficiency and bias of parameter estimates. His results indicated that samples of at least 400 are necessary to obtain adequate estimates of the Ex-Wald, and that for some parameter ranges, much larger samples may be required. He observed that for shifted Wald estimation, smaller samples of around 100 were adequate, at least when fits identified by the software as having an ill-conditioned maximum were excluded. This study he carried out using S-PLUS functions, and Schwarz's data. [21] affirmed that the inverse Gaussian Distribution is known for its flexible shape and that it is widely used across various applications and as such existing confidence intervals for the mean parameter, such as profile likelihood, reparametrized profile likelihood, and Wald-type reparametrized profile likelihood with observed Fisher information intervals, are generally effective, but their simulation study identifies scenarios where the coverage probability falls below the nominal confidence level. They mathematically derive the Wald-type profile likelihood (WPL) interval and the Wald-type reparametrized profile likelihood with expected Fisher information (WRPLE) interval and compared their performance to existing methods. Their results indicated that the WRPLE interval outperforms others in terms of coverage probability, while the WPL typically yields the shortest interval. In addition, they applied these proposed intervals to a real dataset, demonstrating their potential applicability to other datasets that follow the IG distribution. [14] examined the conjugate Bayesian framework applied to the Wald model, particularly focusing on deriving an exact posterior distribution for the drift-rate parameter. He noted that simple response times are commonly modeled using a one-dimensional Wiener process with drift, where the first-passage time is governed by the Wald distribution—a specialized form of the inverse Gaussian. Meyer-Grant observed that although the Gaussian-Gamma distribution acts as a conjugate prior to an inverse-Gaussian likelihood, this relationship holds only under a different parameterization from that used in the Wald model. Consequently, the resulting posterior distribution does not align directly with the core parameters of the Wiener process, namely the drift-rate and the threshold. While the marginal posterior of the threshold parameter is relatively straightforward and conforms to a known distribution, deriving the drift-rate's marginal posterior proved more complex. To address this gap, Meyer-Grant derived its exact form under the Gaussian-Gamma prior. However, because the probability density function of this posterior cannot be expressed in elementary terms, he implemented various approximation techniques to ensure feasibility in time-sensitive applications. [4] introduced the

Shifted Wald (SW) distribution as a valuable tool for analyzing response time (RT) data in psychological research. They proposed that the SW distribution could serve both as a practical measurement model and as a process model to capture intra-individual differences in response behavior. Their methodological contribution emphasized accessibility, showing that the fitting procedures could be widely applied across diverse experimental settings. To highlight the generalizability of the model, they applied it to three distinct datasets involving manual, vocal, and oculomotor response modalities. Their findings underscored the adaptability and robustness of the Shifted Wald distribution across varying cognitive tasks and response mechanisms. Other interesting literature relevant for further studies include but not limited to those presented in Table (1) below

Table 1: Other relevant distributions

Author	Distributions
[18]	New family of generalized distributions based on logistic-x transformation
[2]	T-X family of distributions
[12]	Type II general inverse exponential family of distributions
[3]	T-normal family of distributions
[8]	The Kumaraswamy generalized half-normal distribution
[10]	Beta-normal distribution
[16]	Type-I heavy-tailed Rayleigh distribution
[5]	Burr-type XII distributions
[24]	Modified Burr III Odds Ratio–G Distribution
[26]	`Obulezi` distribution
[27]	A new bivariate trigonometric Gaussian distribution
[28]	Power Mira distribution
[29]	Three-parameter quadratic hazard rate distribution
[30]	New hybrid Weibull-inverse Weibull distribution
[31]	Updated Lindley distribution
[32]	A new odd reparameterized exponential Transformed-X family of distributions
[17]	Weibull Sine Generalized distribution family

Although the Wald distribution has been analyzed and investigated extensively, it is still a relevant topic of research, owing to the unsolved challenges in parameter estimation, especially concerning small sample sizes and skewness of the data. If some work could support its use and elucidation, this study intends to fine-tune the estimation procedures for comparison in the empirical sense against methods such as Maximum Likelihood Estimation (MLE), Maximum product of spacing (MPS), Least Squares Estimation (LSE), Weighted Least Squares (WLS), Cramér-von Mises (CvM), Anderson-Darling (AD), right-tailed Anderson-Darling (RTAD) and Bayesian Estimators (BE). In this manner, the methods are characterized concerning bias, RMSE, and efficiency concerning their strengths and weaknesses, particularly for cases where MLE is less stable. Further, through its application in case studies in various fields, from health data analysis and epidemiology to reliability study, the work emphasizes the practical applicability of the Wald distribution. Thus, through the tackling of the estimation and application issues, this contribution guarantees the continued relevance and usefulness of the Wald distribution in applied statistics.

The remaining sections of this study are presented in the sequence presented in the flow chart in Figure (1) below.

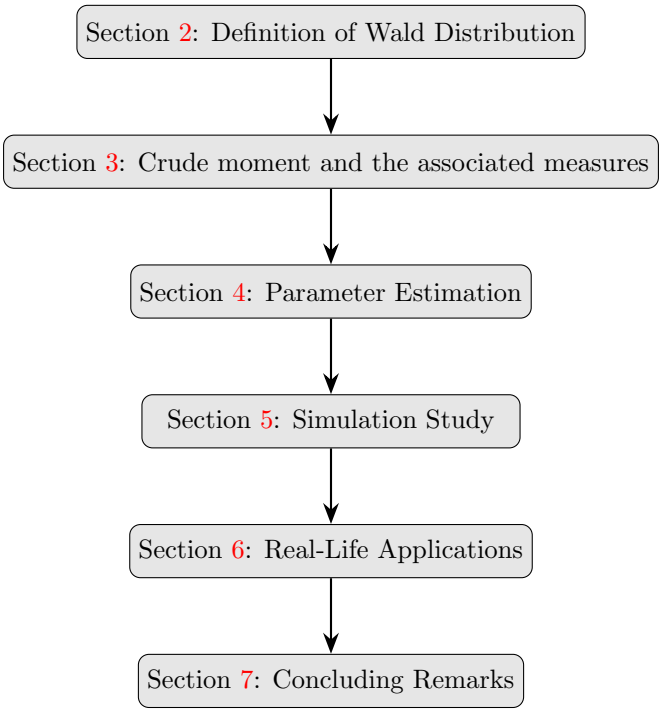


Figure 1: Organization of the study

2. Definition of Wald Distribution

The cumulative distribution function (CDF) of the Inverse Gaussian (Wald) distribution due to [23], which was extensively studied by [25] is given by

$$F(x;\mu,\lambda) = \Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right) + e^{\frac{2\lambda}{\mu}}\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right), \quad x>0, \tag{2.1}$$

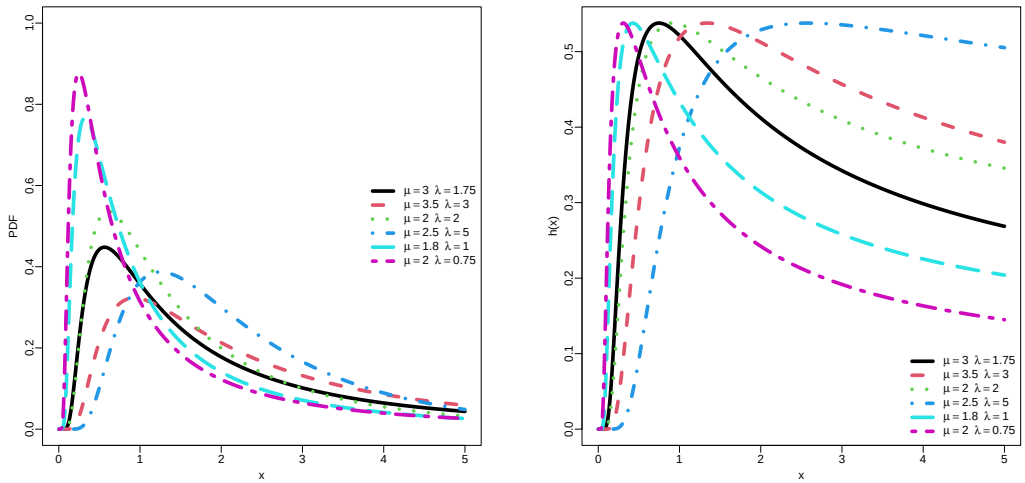
where μ is the mean (location parameter), λ is the shape parameter, $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The Wald distribution is commonly used in reliability analysis, survival analysis, and financial data modeling. It is widely used in modeling first-passage times, reliability engineering, and financial data analysis. The corresponding probability density function (PDF) of the Wald distribution is written as

$$f(x;\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right). \tag{2.2}$$

The hazard function is

$$h(x;\mu,\lambda) = \frac{\sqrt{\frac{\lambda}{2\pi x^3}}\exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right)}{1-\left[\Phi\left(\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right)\right) + e^{\frac{2\lambda}{\mu}}\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right)\right)\right]}, \tag{2.3}$$

Figure (2) contains the plots of the PDF and hazard function of the Wald distribution. This classical distribution obviously exhibits reverse bath-tub shape as can be seen from the hazard function.



(a) (b)

Figure 2: Plots of the Wald dist. (a) PDF plot (b) Hazard function plot

3. Crude moment and the associated measures

The rth moment of $X \sim$ Wald (μ, λ) is

$$E[X^r] = \int_0^\infty x^r f(x;\mu,\lambda)\, dx = \int_0^\infty x^r \sqrt{\frac{\lambda}{2\pi x^3}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) dx.$$

Rewriting:

$$E[X^r] = \sqrt{\frac{\lambda}{2\pi}} \int_0^\infty x^{r-\frac{3}{2}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right) dx.$$

Define a new variable $t = \frac{\lambda}{2\mu^2}\frac{(x-\mu)^2}{x}$. Then:

$$E[X^r] = \sqrt{\frac{\lambda}{2\pi}} \int_0^\infty x^{r-\frac{3}{2}} \exp(-t) dx.$$

Using a standard integral result from the Inverse Gaussian distribution moments:

$$E[X^r] = \mu^r \Gamma\left(r+\frac{1}{2}\right)\left(\frac{2}{\lambda}\right)^{r/2}; \quad r = 1, 2, \cdots \tag{3.1}$$

The mean of the Wald distribution is obtained by replacing $r = 1$ in equation (3.1), hence

$$E[X] = \mu\Gamma\left(\frac{3}{2}\right)\left(\frac{2}{\lambda}\right)^{1/2} = \mu\sqrt{\frac{\pi}{2\lambda}}. \tag{3.2}$$

The second crude moment is obtained by replacing $r = 2$ in equation (3.1), hence

$$E[X^2] = \mu^2\Gamma\left(\frac{5}{2}\right)\left(\frac{2}{\lambda}\right) = \frac{3\mu^2\sqrt{\pi}}{2\lambda}. \tag{3.3}$$

Thus, the variance is expressed as

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{\mu^2}{2\lambda}(3\sqrt{\pi} - \pi).$$

The third and fourth crude moments are obtained by replacing $r = 3$ and $r = 4$ in equation (3.1) respectively, hence

$$E[X^3] = \mu^3 \Gamma\left(\frac{7}{2}\right)\left(\frac{2}{\lambda}\right)^{3/2} = \frac{15\mu^3\sqrt{\pi}}{8}\left(\frac{2}{\lambda}\right)^{3/2}, \tag{3.4}$$

and

$$E[X^4] = \mu^4 \Gamma\left(\frac{9}{2}\right)\left(\frac{2}{\lambda}\right)^{2} = \frac{105\mu^4\sqrt{\pi}}{16}\left(\frac{2}{\lambda}\right)^{2}. \tag{3.5}$$

The central moments $\mu_k = E[(X - E[X])^k]$ are used to compute skewness and kurtosis, and they can be derived from the crude moments. The third central moment μ_3 is given as $\mu_3 = E[X^3] - 3E[X^2]E[X] + 2(E[X])^3$, hence the skewness ($\gamma_1$):

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\lambda^{3/2}\left(-1.73205\pi + 0.3849\pi^{3/2} + 2.88675\sqrt{\pi}\right)}{\left(-0.33333\pi\lambda + \sqrt{\pi}\lambda\right)^{3/2}}$$

The fourth central moment μ_4 is $\mu_4 = E[X^4] - 4E[X^3]E[X] + 6E[X^2](E[X])^2 - 3(E[X])^4$, so kurtosis ($\gamma_2$) is expressed as

$$\gamma_2 = \frac{\mu_4}{\sigma^4} = \frac{\lambda^2\left(-60\pi - 3\pi^2 + 18\pi^{3/2} + 105\sqrt{\pi}\right)}{9\left(0.33333\pi\lambda - \sqrt{\pi}\lambda\right)^2}.$$

$$\text{Excess Kurtosis} = \gamma_2 - 3 = \frac{\lambda^2\left(-60\pi - 3\pi^2 + 18\pi^{3/2} + 105\sqrt{\pi}\right) - 27\left(0.33333\pi\lambda - \sqrt{\pi}\lambda\right)^2}{9\left(0.33333\pi\lambda - \sqrt{\pi}\lambda\right)^2}.$$

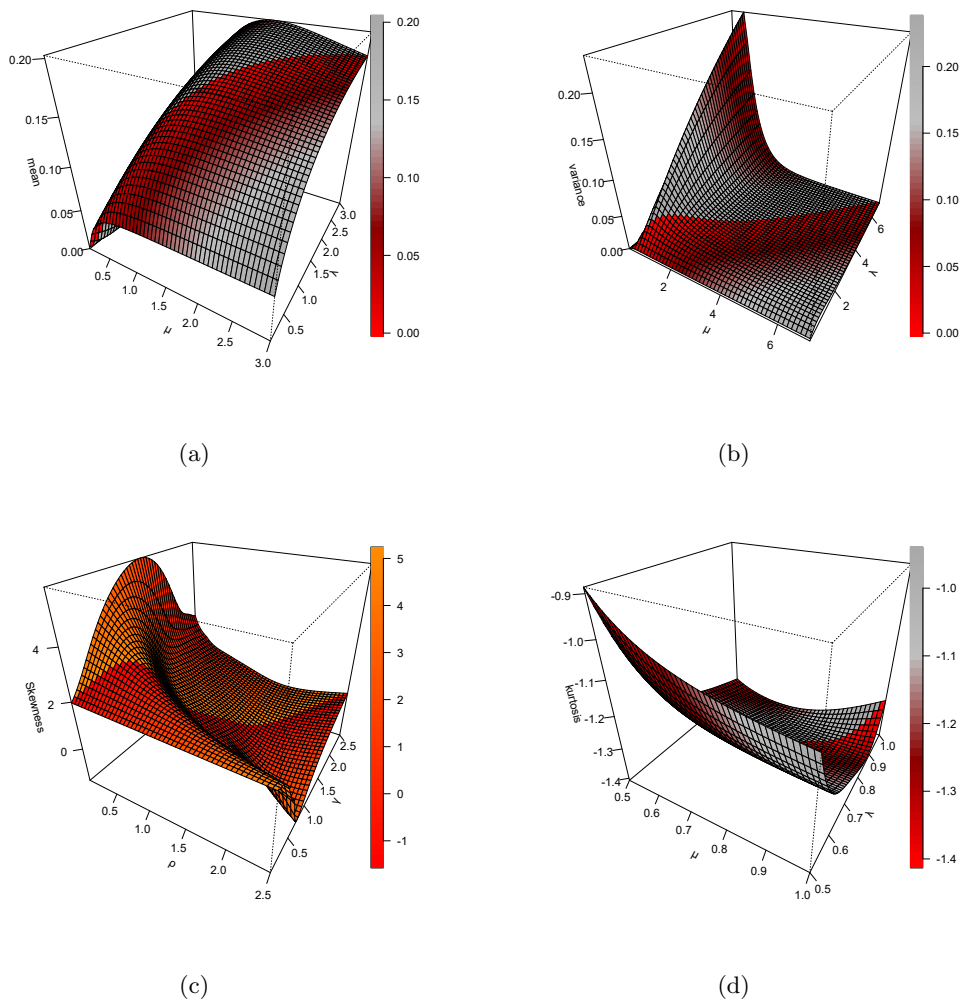


Figure 3: Moments (a) Mean (b) Variance (c) Skewness (d) Kurtosis

The Mean, Variance, Skewness and Kurtosis plots of the Wald distribution are presented in 3D, see Figure (3).

4. Parameter Estimation

In this section, a number of estimation procedures are adopted for the study. In the frequentist school of thought, the maximum likelihood, least squares, weighted least squares, maximum product of spacing, Cramer-von Mises, Anderson-Darling, and right-tailed Anderson-Darling methods are suggested. In Bayesian inference, the squared error loss function, the linear-exponential loss function, and the generalized entropy loss function under gamma informative priors are studied.

4.1. Maximum Likelihood Estimation (MLE)

To evaluate the MLE for the parameters μ and λ for the PDF of the Wald distribution given in Eq. (2.2), given the observed sample $x_1, x_2, \ldots, x_n$, the likelihood function is

$$L(\mu,\lambda)=\prod_{i=1}^{n}\sqrt{\frac{\lambda}{2\pi x_i^3}}\exp\left(-\frac{\lambda(x_i-\mu)^2}{2\mu^2x_i}\right).$$

The log-likelihood function is:

$$\ell(\mu,\lambda)=\frac{n}{2}\ln\lambda-\frac{n}{2}\ln(2\pi)-\frac{3}{2}\sum_{i=1}^{n}\ln x_i-\frac{\lambda}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\mu^2x_i}.$$

So that;

$$\frac{\partial\ell}{\partial\mu}=-\frac{\lambda}{2}\sum_{i=1}^{n}\frac{\partial}{\partial\mu}\left(x_i\mu^{-2}-2\mu^{-1}+x_i^{-1}\right)=-\frac{\lambda}{2}\sum_{i=1}^{n}\left(x_i(-2\mu^{-3})-2(-1\mu^{-2})+0\right)=\frac{\lambda}{\mu^3}\sum_{i=1}^{n}(x_i-\mu).$$

Hence,

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$$

The MLE for μ is the sample mean ($\bar{x}$). Similarly,

$$\hat{\lambda}(\mu)=\frac{n}{\sum\limits_{i=1}^{n}\frac{(x_i-\mu)^2}{\mu^2x_i}}.$$

To find the MLE for λ, $\hat{\lambda}$, we substitute the MLE for μ, which is $\hat{\mu}=\bar{x}$:

$$\hat{\lambda}=\frac{n}{\sum\limits_{i=1}^{n}\frac{(x_i-\bar{x})^2}{\bar{x}^2x_i}}=\frac{n\bar{x}^2}{\sum\limits_{i=1}^{n}\frac{(x_i-\bar{x})^2}{x_i}}.$$

This is simplified by expanding the numerator of the sum:

$$\sum_{i=1}^{n}\frac{x_i^2-2\bar{x}x_i+\bar{x}^2}{x_i}=\sum_{i=1}^{n}\left(x_i-2\bar{x}+\frac{\bar{x}^2}{x_i}\right)=\sum x_i-2\bar{x}\sum 1+\bar{x}^2\sum\frac{1}{x_i}=\bar{x}^2\sum\frac{1}{x_i}-n\bar{x}.$$

So, the MLE for λ can be written as:

$$\hat{\lambda}=\frac{n}{\frac{1}{\bar{x}^2}\left(\bar{x}^2\sum\frac{1}{x_i}-n\bar{x}\right)}=\frac{n}{\sum\frac{1}{x_i}-\frac{n}{\bar{x}}}.$$

4.2. The Least Squares Estimation (LS)

This study adopts the Least Squares technique, which was first introduced by Swain et al. for estimating parameters of the beta distribution, for parameter estimation. We then determined parameter values by minimizing squared differences between observed and model-predicted values according to that method. It produces an excellent fit between the model and empirical data. Method of LS thus increases the accuracy of parameter estimates in the proposed model. Hence, LS is a very important method in parameter estimation.

$$E\left[F\left(x_{j:n}|\mu,\lambda\right)\right]=\frac{j}{n+1};\quad\text{and}\quad Var\left[F\left(x_{j:n}|\mu,\lambda\right)\right]=\frac{j\left(n-j+1\right)}{\left(n+1\right)^2\left(n+2\right)}.$$

By the method of Least Squares (LS), the parameters μ and λ are estimated with estimates $\hat{\mu}_{\mathrm{LS}}$ and $\hat{\lambda}_{\mathrm{LS}}$. These estimates are obtained from minimizing the objective function $L(\mu,\lambda)$ that measures the total squared deviation of the observed values from their corresponding expected values. Hence, the optimization finds the parameter values in terms of μ and λ that best fit the given observed data.

$$L(\mu,\lambda)=\arg\min_{(\mu,\lambda)}\sum_{j=1}^{n}\left[F\left(x_{j:n}|\mu,\lambda\right)-\frac{j}{n+1}\right]^2.$$

By setting into zero the partial derivatives of the objective function $L(\delta,\alpha,\beta)$, estimates for μ and λ can be obtained and this would give a system of non-linear equations. The solution of which gives the least-squares estimators $\hat{\mu}_{\mathrm{LSE}}$ and $\hat{\lambda}_{\mathrm{LSE}}$.

$$\sum_{j=1}^{n}\left[F\left(x_{j:n}|\mu,\lambda\right)-\frac{j}{n+1}\right]^2\Delta_1\left(x_{j:n}|\mu,\lambda\right)=0, \tag{4.1}$$

and

$$\sum_{j=1}^{n}\left[F\left(x_{j:n}|\mu,\lambda\right)-\frac{j}{n+1}\right]^2\Delta_2\left(x_{j:n}|\mu,\lambda\right)=0 \tag{4.2}$$

where

$$\Delta_1\left(x_{j:n}|\mu,\lambda\right)=\phi(Z_1)\sqrt{\frac{\lambda}{x}}\cdot\frac{x}{\mu^2}+e^{\frac{2\lambda}{\mu}}\left[\phi(Z_2)\sqrt{\frac{\lambda}{x}}\cdot\frac{x}{\mu^2}-\frac{2\lambda}{\mu^2}\Phi(Z_2)\right], \tag{4.3}$$

and

$$\Delta_2\left(x_{j:n}|\mu,\lambda\right)=\phi(Z_1)\cdot\frac{1}{2}\frac{1}{\sqrt{\lambda x}}\left(\frac{x}{\mu}-1\right)+e^{\frac{2\lambda}{\mu}}\left[-\phi(Z_2)\frac{1}{2}\frac{1}{\sqrt{\lambda x}}\left(\frac{x}{\mu}+1\right)+\frac{2}{\mu}\Phi(Z_2)\right], \tag{4.4}$$

where $\Phi(z)$ is standard normal CDF with derivative:

$$\frac{d\Phi(z)}{dz}=\phi(z),$$

$\phi(z)$ is the standard normal distribution or simply:

$$\phi(z)=\frac{1}{\sqrt{2\pi}}e^{-z^2/2}.$$

hence

$$Z_1=\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}-1\right),\quad Z_2=-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu}+1\right).$$

Naturally, here, equations (4.3) and (4.4) arise from the differentiation of the distribution function of the Wald model as shown by equation (2.1) with respect to the parameters δ, α, and β. These derivatives thus measure the sensitivity of the cumulative distribution function (CDF) with respect to each parameter entering the least-squares estimations through the resulting system of equations.

4.3. The Weighted Least Squares Estimation (WLS)

The approach of Weighted Least Squares (WLS) is employed to estimate the parameters μ and λ of the Wald distribution. WLS differs from the ordinary least squares method in that, in addition to the usual observations, weights w_j are assigned to each observation in order to incorporate their relative reliability or precision. In this way, WLS gives more weight to those points which are deemed to carry more information or precision into the parameter estimation. The resulting estimators are denoted by $\hat{\mu}_{WLS}$ and $\hat{\lambda}_{WLS}$, obtained by minimizing the weighted least squares objective function denoted as $W(\mu,\lambda)$.

The WLS function is defined as:

$$W(\mu,\lambda)=\arg\min_{(\mu,\lambda)}\sum_{j=1}^{n}w_j\left[F\left(x_{j:n}|\mu,\lambda\right)-\frac{j}{n+1}\right]^2 \tag{4.5}$$

The weight assigned to each observation w_j is defined as follows:

$$w_j = \frac{(n+1)^2 (n+2)}{j (n-j+1)}.$$

Such weights become even more artificially enforced on a more general setting of the linear model.

Parameter estimation is done and achieved by solving a system of equations generated by making the weighted sums of squared deviations equal to zero:

$$\sum_{j=1}^{n} w_j \left[G(x_{j:n}|\delta, \alpha, \beta) - \frac{j}{n+1} \right]^2 \Delta_1 (x_{j:n}|\mu, \lambda) = 0, \tag{4.6}$$

and

$$\sum_{j=1}^{n} w_j \left[F(x_{j:n}|\mu, \lambda) - \frac{j}{n+1} \right]^2 \Delta_2 (x_{j:n}|\mu, \lambda) = 0. \tag{4.7}$$

The terms $\Delta_1(x_{j:n}|\mu, \lambda)$ and $\Delta_2(x_{j:n}|\mu, \lambda)$ in equations (4.6) and (4.7) are related to partial derivatives of the CDF of the Wald distribution with respect to parameters μ and λ, respectively. These expressions will be crucial for deriving weighted least squares estimates, as they will state how the CDF would change according to variations in each parameter.

4.4. The Maximum Product of Spacing Estimation (MPS)

The MPS, a 1979 conception by [6], provides a robust alternative to MLE. Whereas MLE maximizes the likelihood function, MPS maximizes the product of the spacings between successive CDF values corresponding to the ordered data, exploiting Kullback-Leibler information. Hence, the MPS generally yields estimators that are less biased and more robust than the MLE in the case of difficult estimation.

Product of spacings: $P_s(X|\mu, \lambda)$

$$P_s (X|\mu, \lambda) = \left[\prod_{j=1}^{n+1} D_k (x_{j:n}|\mu, \lambda) \right]^{\frac{1}{n+1}},$$

The equation $D_k(x_{j:n}|\mu, \lambda)$ defines spacing between consecutive values of this CDF:

$$D_k (x_{j:n}|\mu, \lambda) = G(x_j|\mu, \lambda) - G(x_{j-1}|\mu, \lambda)), j = 1, 2, 3, \ldots, n.$$

Under MPS framework, parameter estimation involves maximizing the log of the product of these spacings, which is expressed in prescribed M-function $M_s(\mu, \lambda)$:

$$M_s (\mu, \lambda) = \frac{1}{n+1} \sum_{j=1}^{n+1} \ln (D_k (x_{j:n}|\mu, \lambda)).$$

Estimates are obtained by calculating partial derivatives of $M_s(\mu, \lambda)$ for each parameter and then setting them equal to zero. Thus:

$$\frac{\partial M_s(\mu, \lambda)}{\partial \delta} = 0, \quad \frac{\partial M_s(\mu, \lambda)}{\partial \alpha} = 0.$$

Then the solutions of this nonlinear system will yield the so-called MPS estimates, which will maximize the product of spacings and give a very good fit of the empirical data to the theoretical distribution.

4.5. The Cramér-von Mises Estimation (CvM)

It is the CvM estimation method that is used to identify the unknown parameters μ and λ for the TII-HTW distribution. The corresponding estimators $\hat{\mu}_{\text{CvM}}$ and $\hat{\lambda}_{\text{CvM}}$ were obtained by minimizing the CvM criterion function $W(\mu, \lambda)$. Such estimators minimize the squared deviations between the empirical distribution function and the theoretical cumulative distribution function, thereby providing parameter estimates with a close fit against the observed data.

This CvM criterion function is defined by:

$$W(\mu,\lambda) = \arg\min_{(\mu,\lambda)} \left\{ \frac{1}{12n} + \sum_{j=1}^{n} \left[F(x_{j:n}|\mu,\lambda) - \frac{2j-1}{2n} \right]^2 \right\}.$$

The parameter estimates are determined by solving the system of nonlinear equations:

$$\sum_{j=1}^{n} \left[G(x_{j:n}|\mu,\lambda) - \frac{2j-1}{2n} \right] \Delta_1(x_{j:n}|\mu,\lambda) = 0, \tag{4.8}$$

and

$$\sum_{j=1}^{n} \left[G(x_{j:n}|\mu,\lambda) - \frac{2j-1}{2n} \right] \Delta_2(x_{j:n}|\mu,\lambda) = 0, \tag{4.9}$$

The quantities $\Delta_1(x_{j:n} \mid \mu,\lambda)$ and $\Delta_2(x_{j:n} \mid \mu,\lambda)$ represent the partial derivatives as defined in equations (4.3) and (4.4), respectively. The solutions for these equations yield the optimal estimates that best fit the model to both the theoretical distribution and the empirical distribution associated with the observed data.

4.6. The Anderson-Darling Estimation (AD)

The parameters μ and λ of the Wald distribution are estimated by the AD approach. The parameter estimates, $\hat{\mu}_{\text{AD}}$ and $\hat{\lambda}_{\text{AD}}$, are obtained by minimizing the criterion function $T(\mu,\lambda)$ as indicated in the Anderson-Darling procedure.

The function $T(\delta,\alpha,\beta)$ is given as follows:

$$T(\mu,\lambda) = \arg\min_{(\mu,\lambda)} \sum_{j=1}^{n} (2j-1) \left\{ \ln\left(F(x_{j:n}|\mu,\lambda)\right) + \ln\left[1 - F(x_{n+1-j:n}|\mu,\lambda)\right] \right\}.$$

The parameter estimates are obtained by solving the following system of nonlinear equations:

$$\sum_{j=1}^{n} (2j-1) \left[\frac{\Delta_1(x_{j:n}|\mu,\lambda)}{F(x_{j:n}|\delta,\alpha,\beta)} - \frac{\Delta_1(x_{n+1-j:n}|\mu,\lambda)}{1 - F(x_{n+1-j:n}|\mu,\lambda)} \right] = 0, \tag{4.10}$$

and

$$\sum_{j=1}^{n} (2j-1) \left[\frac{\Delta_2(x_{j:n}|\mu,\lambda)}{F(x_{j:n}|\mu,\lambda)} - \frac{\Delta_2(x_{n+1-j:n}|\mu,\lambda)}{1 - F(x_{n+1-j:n}|\mu,\lambda)} \right] = 0, \tag{4.11}$$

$\Delta_1(x_{j:n} \mid \mu,\lambda)$ and $\Delta_2(x_{j:n} \mid \mu,\lambda)$ are the two partial derivatives defined in equations (4.3) and (4.4), respectively. Their solution yields better estimates because the empirical and the theoretical distributions then match more closely.

4.7. The Right-Tailed Anderson-Darling (RTAD) Estimation

The optimal procedure under the RTAD approach finds out μ and λ such that the value of the function, $T_r(\mu,\lambda)$, is minimized. The result is the corresponding estimates, $\hat{\mu}_{\text{RTAD}}$ and $\hat{\lambda}_{\text{RTAD}}$.

The formula for the function $T_r(\mu,\lambda)$ is defined as such-:

$$T_r(\mu,\lambda) = \arg\min_{(\mu,\lambda)} \left\{ \frac{n}{2} - 2\sum_{j=1}^{n} F(x_{j:n}|\mu,\lambda) - \frac{1}{n}\sum_{j=1}^{n} (2j-1) \ln\left[1 - F(x_{n+1-j:n}|\mu,\lambda)\right] \right\}.$$

Parameter Estimation from the Following Nonlinear Equations.

$$-2\sum_{j=1}^{n} \frac{1}{F(x_{j:n}|\mu,\lambda)} \Delta_1(x_{j:n}|\mu,\lambda) + \frac{1}{n}\sum_{j=1}^{n} (2j-1) \left[\frac{\Delta_1(x_{n+1-j:n}|\mu,\lambda)}{1 - F(x_{n+1-j:n}|\mu,\lambda)} \right] = 0, \tag{4.12}$$

and

$$-2\sum_{j=1}^{n}\frac{\Delta_2\left(x_{j:n}|\mu,\lambda\right)}{F\left(x_{j:n}|\delta,\alpha,\beta\right)}+\frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\frac{\Delta_2\left(x_{n+1-j:n}|\mu,\lambda\right)}{1-F\left(x_{n+1-j:n}|\mu,\lambda\right)}\right]=0, \tag{4.13}$$

The quantities $\Delta_1(x_{j:n}|\mu,\lambda)$ and $\Delta_2(x_{j:n}|\mu,\lambda)$ are defined in (4.3) and (4.4), respectively. The estimates given in (4.1)-(4.4) and (4.6)-(4.13) were obtained using R's *optim()* function based on the Newton-Raphson iterative method for optimal parameter estimation.

4.8. Bayesian Estimation (BE)

This section deals with deriving Bayesian estimates for the unknown parameters of a Wald distribution. Besides the classical methods of estimation, Bayesian estimation provides the freedom to use other loss functions in parameter estimation: for example, squared error, LINEX, etc. We assume that μ and λ are independent gamma random variables whose probability density functions are being considered as prior distributions.

Therefore, the prior distributions for μ and λ should be as follows:

$$\pi_1(\mu)\propto\mu^{p_1-1}e^{-q_1\mu},\quad \mu>0,\ p_1>0,\ q_1>0,$$
$$\pi_2(\lambda)\propto\lambda^{p_2-1}e^{-q_2\lambda},\quad \lambda>0,\ p_2>0,\ q_2>0.$$

The hyperparameters p_j and q_j $(j=1,2)$ are based on some prior knowledge. Therefore, the joint prior for the parameter vector $\Theta=(\mu,\lambda)$ will simply be the product of both individual priors.

$$\pi(\Theta)=\pi_1(\mu)\pi_2(\lambda)\propto\mu^{p_1-1}\lambda^{p_2-1}e^{-q_1\mu-q_2\lambda}.$$

The posterior distribution of the parameters $\Theta=(\mu,\lambda)$ given the observed data $X=(x_1,x_2,\ldots,x_n)$:

$$\pi(\Theta\mid X)=\frac{\pi(\Theta)\iota(\Theta)}{\int_\Theta\pi(\Theta)\iota(\Theta)d\Theta}.$$

This makes the posterior density look like:

$$\pi(\Theta\mid X)\propto\mu^{p_1-1}\lambda^{p_2-1}e^{-q_1\mu-q_2\lambda}\left(\frac{\lambda}{2\pi}\right)^{\frac{n}{2}}\exp\left(-\sum_{i=1}^{n}\left[\frac{3}{2}\ln x_i+\frac{\lambda(x_i-\mu)^2}{2\mu^2x_i}\right]\right).$$

For a given function $\vartheta(\Theta)$, the Bayes estimator under squared error loss (SEL) criterion is:

$$\hat{\Theta}_{\text{BE\_SEL}}=E[\vartheta(\Theta)\mid X]=\int_\Theta\vartheta(\Theta)\pi(\Theta\mid X)\,d\Theta.$$

The Bayes estimates with different loss functions can then be derived using posterior samples generated via MCMC. They are computed as follows:

1. Squared Error Loss (SEL)

$$\hat{\Theta}_{\text{BE\_SEL}}=\frac{1}{N-\vartheta_b}\sum_{j=\vartheta_b}^{N}\Theta(j),$$

2. LINEX Loss

$$\hat{\Theta}_{\text{BE\_LINEX}}=-\frac{1}{\kappa}\log\left(\frac{1}{N-\vartheta_b}\sum_{j=\vartheta_b}^{N}e^{-\kappa\Theta(j)}\right),$$

and

3. Generalized Error Loss (GEL)

$$\hat{\Theta}_{\text{BE\_GEL}}=\left(\frac{1}{N-\vartheta_b}\sum_{j=\vartheta_b}^{N}\Theta(j)^{-\beta}\right)^{-\frac{1}{\beta}}.$$

Here, ϑ_b denotes the number of burn-in samples discarded during the MCMC simulation process.

5. Simulation Study

Non-Bayesian and Bayesian estimation methods were subjected to 25-200 instances of simulation to obtain the average estimated bias and RMSE value of Wald parameter estimation under the following cases:

Case I: $\mu = 1.75$ and $\lambda = 2.75$

Case II: $\mu = 0.75$ and $\lambda = 1.25$

Case III: $\mu = 1.25$ and $\lambda = 2.0$;

Case IV: $\mu = 0.5$ and $\lambda = 1.75$.

The results are presented in Tables (2) - (5), respectively.

Table 2: Simulation Results for Case I

Method	Parameter	$n = 25$		$n = 75$		$n = 150$		$n = 200$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	μ	0.0047	0.0751	0.0040	0.0250	0.0052	0.0128	0.0010	0.0100
	λ	0.3443	1.1267	0.1032	0.2468	0.0644	0.1155	0.0535	0.0820
MPS	μ	0.1146	0.1105	0.0535	0.0307	0.0222	0.0141	0.0209	0.0109
	λ	0.2485	0.6977	0.1509	0.2176	0.0837	0.1066	0.0644	0.0765
LS	μ	0.0776	0.1205	0.0283	0.0343	0.0060	0.0166	0.0091	0.0129
	λ	0.1262	1.6664	0.0178	0.2981	0.0216	0.1627	0.0134	0.1168
WLS	μ	0.0549	0.1016	0.0178	0.0290	0.0006	0.0144	0.0042	0.0112
	λ	0.1452	1.3993	0.0438	0.2534	0.0401	0.1341	0.0290	0.0938
CvM	μ	0.0159	0.1009	0.0092	0.0324	0.0033	0.0163	0.0021	0.0127
	λ	0.4771	2.4206	0.1207	0.3399	0.0725	0.1748	0.0513	0.1231
AD	μ	0.0376	0.0916	0.0162	0.0285	0.0013	0.0143	0.0045	0.0112
	λ	0.1728	1.1323	0.0482	0.2455	0.0330	0.1257	0.0265	0.0909
RTAD	μ	0.0344	0.0865	0.0152	0.0273	0.0014	0.0138	0.0044	0.0108
	λ	0.3478	1.9086	0.0915	0.3424	0.0544	0.1878	0.0389	0.1208
BE_{SEL}	μ	0.1671	0.0722	0.1639	0.0390	0.1628	0.0320	0.1627	0.0305
	λ	0.5332	0.9601	1.0013	1.2709	1.2320	1.6675	1.2917	1.7887
BE_{Linex1}	μ	0.1563	0.0710	0.1604	0.0380	0.1611	0.0315	0.1614	0.0301
	λ	0.6485	1.1967	1.0746	1.4456	1.2788	1.7926	1.3289	1.8913
BE_{Linex2}	μ	0.1774	0.0737	0.1674	0.0400	0.1645	0.0325	0.1639	0.0309
	λ	0.4298	0.7902	0.9319	1.1192	1.1867	1.5512	1.2553	1.6918
BE_{GEL1}	μ	0.1733	0.0734	0.1661	0.0396	0.1638	0.0323	0.1634	0.0307
	λ	0.5017	0.9168	0.9826	1.2314	1.2206	1.6385	1.2827	1.7649
BE_{GEL2}	μ	0.1857	0.0763	0.1703	0.0409	0.1659	0.0330	0.1650	0.0312
	λ	0.4398	0.8399	0.9454	1.1553	1.1977	1.5815	1.2646	1.7178

The simulation results for Case I contained in Table (2) provide valuable information concerning each of the estimation techniques in terms of bias and the root mean squared error (RMSE) calculated for two parameters of interest, μ and λ, for varying sample sizes. As expected, generally larger sample sizes result in lower bias and RMSE values, suggesting better accuracy and consistency for the corresponding estimates.

In the case of MLE, the bias for μ remains appreciably low for all sample sizes, while the RMSE keeps decreasing with increased n. In contrast, the bias for λ is significantly different for $n = 25$ with a relatively large value, but improves consistently with increasing sample sizes. This indicates that while MLE would provide reliable estimates for larger sample sizes, it becomes unstable for the smaller ones. The MPS proved better performing as an estimator; this estimator shows higher bias for μ than MLE, particularly for small sample sizes, but decreases steadily with sample size. The RMSE of MPS demonstrates a parallel trend in decreasing values with increased n, augmenting its asymptotic efficiency.

Similarly, for LS and WLS, bias and RMSE decrease as sample sizes increase. Note that the bias related to μ for LS and WLS is somewhat higher than that for MLE, while RMSE is fairly competitive, especially at large sample sizes. This shows that with improvement, these methods can serve as alternative methods to MLE, especially for small to moderate sample sizes.

According to MLE, the bias for μ stays constant for all other sample sizes in CvM, with relatively high values for the bias in λ with respect to small n, decreasing as n increases. Still, the RMSE values do not seem to be in favor of CvME and invariably exceed those of MLE and MPS, suggesting that, in terms of estimation, CvM may not be wrong but, in efficiency, is far from perfect.

The AD and its robust version, the restricted-trimmed absolute deviation estimator, are now in the forefront, showing a parallel behavior such as that of WLS, with bias and RMSE reducing as n increases. Thus, the results show that these methods give stable estimates throughout sample size variations and can compete against ones based on least squares.

With respect to BE_{SEL}, $\text{BE}_{\text{Linex1}}$, $\text{BE}_{\text{Linex2}}$, BE_{GEL1}, and BE_{GEL2}, these Bayesian estimates were derived using vague Gamma priors (as specified in Section 4.8). This choice of non-informative prior likely contributes to the results observed. Specifically, a consistently higher bias for μ is observed compared to the frequentist estimators across all sample sizes, remaining relatively stable. The RMSE for λ is also significantly larger than that of the frequentist methods, suggesting that the non-informative prior introduces substantial uncertainty or is poorly matched to the data likelihood, resulting in sub-optimal performance in terms of RMSE. It must be noted that these results reflect the performance only under the condition of vague priors, and a more informative prior may lead to substantially improved estimation efficiency for the Bayesian methods.

Table 3: Simulation Results for Case II

Method	Parameter	$n = 25$		$n = 75$		$n = 150$		$n = 200$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	μ	0.0034	0.0137	0.0017	0.0041	0.0033	0.0022	0.0003	0.0016
	λ	0.1818	0.2444	0.0630	0.0563	0.0196	0.0226	0.0188	0.0162
MPS	μ	0.0518	0.0202	0.0177	0.0049	0.0080	0.0024	0.0086	0.0017
	λ	0.0911	0.1435	0.0543	0.0457	0.0477	0.0221	0.0348	0.0158
LS	μ	0.0361	0.0212	0.0113	0.0059	0.0023	0.0029	0.0037	0.0022
	λ	0.0567	0.2976	0.0254	0.0756	0.0006	0.0334	0.0013	0.0218
WLS	μ	0.0269	0.0179	0.0062	0.0050	0.0005	0.0025	0.0017	0.0019
	λ	0.0759	0.2695	0.0377	0.0649	0.0086	0.0265	0.0086	0.0179
CvM	μ	0.0109	0.0177	0.0036	0.0056	0.0015	0.0029	0.0009	0.0022
	λ	0.2150	0.4349	0.0731	0.0872	0.0224	0.0355	0.0185	0.0228
AD	μ	0.0197	0.0163	0.0054	0.0049	0.0003	0.0025	0.0019	0.0018
	λ	0.0893	0.2158	0.0392	0.0608	0.0067	0.0255	0.0074	0.0175
RTAD	μ	0.0192	0.0157	0.0044	0.0047	0.0004	0.0024	0.0018	0.0018
	λ	0.1613	0.3886	0.0671	0.0795	0.0174	0.0362	0.0146	0.0255
BE_{SEL}	μ	0.0507	0.0113	0.0621	0.0060	0.0642	0.0051	0.0648	0.0049
	λ	0.5502	0.5579	0.6066	0.4419	0.6453	0.4529	0.6527	0.4543
$\text{BE}_{\text{Linex1}}$	μ	0.0488	0.0113	0.0615	0.0060	0.0639	0.0051	0.0646	0.0049
	λ	0.5831	0.6144	0.6244	0.4667	0.6558	0.4674	0.6609	0.4656
$\text{BE}_{\text{Linex2}}$	μ	0.0526	0.0113	0.0627	0.0061	0.0645	0.0051	0.0650	0.0049
	λ	0.5191	0.5084	0.5893	0.4186	0.6349	0.4388	0.6446	0.4433
BE_{GEL1}	μ	0.0533	0.0114	0.0629	0.0061	0.0646	0.0052	0.0651	0.0050
	λ	0.5337	0.5359	0.5973	0.4301	0.6399	0.4457	0.6485	0.4487
BE_{GEL2}	μ	0.0583	0.0116	0.0646	0.0063	0.0655	0.0053	0.0657	0.0050
	λ	0.5012	0.4947	0.5789	0.4073	0.6290	0.4314	0.6400	0.4375

Results from simulation for Case II in Table (3) indicate performance of different estimation techniques with respect to bias and RMSE (root mean squared error) at changing sample size. The bias and RMSE low, with respect to accuracy and consistency of estimators, decrease with increasing sample size.

With low bias and RMSE of μ for maximum likelihood estimator (MLE) across all sample sizes and improvement with the increment of n: the bias for λ starts higher at $n = 25$, and with increased sample size, there is a huge decrement, thus demonstrating the classical asymptotic behavior of MLE. However, when compared to a few other estimators, the error of MLE for λ remains moderate for the lower sample sizes.

The MPS is similar to the MLE, except it is likely to exhibit higher bias in estimating μ for small sample sizes. The MPSE viability increases as a function of sample size, whereupon the RMSE for MPS begins to match that of MLE. The least squares estimator (LS) and weighted least squares estimator (WLS) generally have lower bias estimates for μ than MPS for larger sample sizes but tend to have higher RMSE estimates for λ, which indicates their inferior efficiency in parameter estimation.

The CvM performs badly, especially in small sample sizes, in estimating λ. In classical methods, CvM has the highest RMSE for λ. Despite an obvious improvement with an increasing n, the conditions under which, and the extent to which, the CvM will perform well in the present instance, remain doubtful. Both approximate density estimators AD and robust counterparts RTAD present some stiff competition regarding bias and RMSE of μ and are good alternatives among conventional.

Bayesian estimators under different loss functions (BE_{SEL}, $\text{BE}_{\text{Linex1}}$, $\text{BE}_{\text{Linex2}}$, BE_{GEL1}, BE_{GEL2}) demonstrate an atypical situation, showing relatively high bias and RMSE for λ. These results suggest that Bayesian estimators in this context may not be suitable for λ. At the same time, Bayesian estimators, regarding RMSE of μ, reveal consistent values across various sample sizes, which confirms this parameter's estimation is robust.

In sum, MLE and AD appeared to offer the best trade-offs between bias and RMSE according to the classical methods. On the other hand, LS and WLS might be good for μ estimations but not for λ, while CvM suffers the worst for this latter one. Bayesian estimators are consistently biased for μ and generally have poor performance for λ. Most of the methods improve in accuracy as the sample sizes increase, thereby confirming the asymptotic properties of these estimators.

Table 4: Simulation Results for Case III

Method	Parameter	$n = 25$		$n = 75$		$n = 150$		$n = 200$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	μ	0.0072	0.0405	0.0072	0.0135	0.0011	0.0066	0.0008	0.0049
	λ	0.2523	0.5767	0.0882	0.1341	0.0355	0.0600	0.0378	0.0425
MPS	μ	0.0928	0.0610	0.0415	0.0166	0.0208	0.0074	0.0145	0.0053
	λ	0.1776	0.3612	0.0962	0.1139	0.0723	0.0578	0.0480	0.0399
LS	μ	0.0650	0.0643	0.0251	0.0196	0.0102	0.0091	0.0074	0.0067
	λ	0.0563	0.6612	0.0231	0.1826	0.0064	0.0864	0.0098	0.0569
WLS	μ	0.0501	0.0553	0.0189	0.0166	0.0055	0.0078	0.0036	0.0057
	λ	0.0807	0.5659	0.0396	0.1503	0.0209	0.0722	0.0216	0.0471
CvM	μ	0.0207	0.0533	0.0115	0.0184	0.0036	0.0088	0.0025	0.0065
	λ	0.3056	0.9747	0.0992	0.2097	0.0432	0.0920	0.0374	0.0601
AD	μ	0.0361	0.0496	0.0174	0.0163	0.0058	0.0077	0.0040	0.0056
	λ	0.1121	0.4995	0.0397	0.1366	0.0171	0.0681	0.0192	0.0458
RTAD	μ	0.0342	0.0469	0.0174	0.0157	0.0055	0.0074	0.0036	0.0054
	λ	0.2166	0.8342	0.0729	0.2282	0.0360	0.1013	0.0320	0.0611
BE_{SEL}	μ	0.1020	0.0341	0.1111	0.0185	0.1126	0.0155	0.1130	0.0148
	λ	0.6091	0.8423	0.8436	0.8751	0.9613	1.0097	0.9905	1.0491
$\text{BE}_{\text{Linex1}}$	μ	0.0965	0.0339	0.1094	0.0182	0.1117	0.0153	0.1124	0.0147
	λ	0.6798	0.9862	0.8855	0.9575	0.9870	1.0632	1.0108	1.0919
$\text{BE}_{\text{Linex2}}$	μ	0.1073	0.0344	0.1128	0.0189	0.1134	0.0157	0.1136	0.0149
	λ	0.5441	0.7268	0.8034	0.8006	0.9361	0.9591	0.9705	1.0080
BE_{GEL1}	μ	0.1064	0.0346	0.1126	0.0188	0.1133	0.0156	0.1135	0.0149
	λ	0.5847	0.8058	0.8294	0.8500	0.9528	0.9929	0.9838	1.0356
BE_{GEL2}	μ	0.1151	0.0356	0.1156	0.0195	0.1147	0.0159	0.1146	0.0152
	λ	0.5367	0.7387	0.8012	0.8014	0.9357	0.9599	0.9704	1.0088

In Case III in Table (4), simulation results provide an eye-opener comparison of various estimation techniques on both bias and RMSE (Root Mean Square Error) across different parametric estimations on status parameters μ and λ over varied data sizes, namely $n = 25, 75, 150, 200$.

For mean parameter μ, MLE proved effective across almost all sample sizes as the bias and RMSE values reduced less for larger sample sizes. For instance, n=25 has a bias of 0.0072, whereas RMSE is 0.0405, which, as n increases to 200, reduces to a final bias of 0.0008 and RMSE of 0.0049. Thus, MLE gives reasonably accurate estimates for μ, with its performance improving with sample size.

For example, at n=25, the bias is equal to 0.0928, and the RMSE is measured at 0.0610, which is indeed higher than that by MLE. However, the two values minimize as n grows up, with bias reaching the value of 0.0145 and RMSE being equal to 0.0053 at n=200. It thus suggests that MPS has issues with smaller dataset sizes but is better off given a larger sample size than MLE.

Outcome mixed for μ with outcome for LS. Bias and RMSE figures showed values less than those of MLE and MPSE for small samples, but a gain in their improvement on the larger data size. At n=25, bias is equal to 0.0650, having an RMSE of 0.0643. However, for n=200, both improved lowering bias and RMSE to 0.0074 and 0.0067, respectively. So although LSE improves with an increase in sample size, LSE remains lower than MLE in terms of precision and accuracy.

Weighted values have lower bias and RMSE figures for μ as compared to the simple Least Squares method; for that, they are gradually decreasing as the sample size increases. For example, at n = 25, bias is equal to 0.0501, and RMSE equals 0.0553; for both values, there is a high tendency to improve with increasing n up to 200- showing that WLS is robust.

In terms of CvM, the bias in μ is quite small as a relative measure in comparison with bias MPSE but very much larger for RMSE, especially within small sample sizes. At n = 25, the bias is measured at 0.0207, while the RMSE equates to 0.0533. As n increases, bias and RMSE values improve further, but not to comparable levels with MLE and WLS estimates.

About the λ, MLE again produces the estimates across all sample sizes that are most consistent and reliable. The value of bias and RMSE for $n = 25$ is highest (bias=0.2523, RMSE=0.5767) but improves significantly with increasing sample sizes until bias reaches as low as 0.0378 and RMSE drops to 0.0425 at $n = 200$. This reflection shows that even for the scale parameter, MLE presents very strong performance.

For MPS, the bias and RMSE values for λ follow a similar trend to those observed for μ, with relatively high values for small sample sizes, particularly at $n = 25$, where the bias is 0.1776 and the RMSE is 0.3612. These values decrease as the sample size increases, although they still do not match the accuracy of MLE.

The performance of LS for λ is somewhat less reliable compared to MLE, particularly at smaller sample sizes. At $n = 25$, the bias is 0.0563 with an RMSE of 0.6612, which decreases as the sample size increases, but does not reach the level of MLE's precision.

For WLSE, similar trends are observed with a reduction in bias and RMSE values as the sample size increases, but the values are consistently higher than those of MLE.

Methods such as BESEL, BELinex1, BELinex2, BEGEL1, and BE_{GEL2} show relatively poor performance for both μ and λ. The bias and RMSE values are much higher compared to the other methods, particularly for the λ parameter. This suggests that these methods struggle to provide reliable estimates, especially for the scale parameter, where bias and RMSE remain substantially large even for larger sample sizes.

In conclusion, MLE consistently outperforms all other methods in terms of both bias and RMSE for both μ and λ, showing robust performance across different sample sizes. While other methods, such as MPS, WLS, and LS, also show improved results with increasing sample size, they do not match MLE's accuracy. Methods based on bias estimation (BE) exhibit poor performance, particularly for λ, and should be used with caution.

The fourth case's simulation results in Table (5) put forth efficient patterns in various estimations at sample sizes. The bias and RMSE values corresponding to the parameter μ remain very low for all the methods, suggesting that all estimators provide good-quality estimates for μ with a minimum possible deviation. The behaviour of the estimators is quite different, though, when it comes to the case of λ.

The MLE approach under this design shows high bias and RMSE in terms of smaller sample sizes, especially for λ, among frequentist approaches. Yet sample sizes are increased; both bias and RMSE

Table 5: Simulation Results for Case IV

Method	Parameter	$n = 25$		$n = 75$		$n = 150$		$n = 200$	
		Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
MLE	μ	0.0003	0.0027	0.0017	0.0010	0.0002	0.0005	0.0005	0.0004
	λ	0.2552	0.4779	0.0620	0.0953	0.0312	0.0420	0.0271	0.0334
MPS	μ	0.0164	0.0034	0.0050	0.0010	0.0036	0.0005	0.0035	0.0004
	λ	0.1405	0.2811	0.1046	0.0854	0.0658	0.0409	0.0503	0.0325
LS	μ	0.0115	0.0036	0.0017	0.0012	0.0018	0.0006	0.0016	0.0004
	λ	0.0643	0.5473	0.0006	0.1379	0.0004	0.0599	0.0012	0.0439
WLS	μ	0.0087	0.0032	0.0004	0.0011	0.0009	0.0005	0.0010	0.0004
	λ	0.0854	0.4698	0.0171	0.1121	0.0127	0.0483	0.0100	0.0361
CvM	μ	0.0032	0.0033	0.0009	0.0012	0.0005	0.0006	0.0006	0.0004
	λ	0.2937	0.8073	0.0678	0.1558	0.0339	0.0638	0.0239	0.0460
AD	μ	0.0061	0.0030	0.0001	0.0011	0.0009	0.0005	0.0011	0.0004
	λ	0.1182	0.4070	0.0205	0.1058	0.0109	0.0460	0.0083	0.0352
RTAD	μ	0.0063	0.0030	0.0002	0.0011	0.0009	0.0005	0.0012	0.0004
	λ	0.2423	0.8166	0.0432	0.1439	0.0266	0.0628	0.0127	0.0452
BE_{SEL}	μ	0.0166	0.0022	0.0218	0.0010	0.0227	0.0008	0.0230	0.0007
	λ	0.6348	0.8013	0.8110	0.7878	0.9012	0.8800	0.9223	0.9040
BE_{Linex1}	μ	0.0161	0.0022	0.0217	0.0010	0.0227	0.0008	0.0229	0.0007
	λ	0.6934	0.9193	0.8447	0.8507	0.9217	0.9196	0.9384	0.9354
BE_{Linex2}	μ	0.0171	0.0022	0.0220	0.0010	0.0228	0.0008	0.0230	0.0007
	λ	0.5805	0.7037	0.7786	0.7302	0.8812	0.8423	0.9064	0.8737
BE_{GEL1}	μ	0.0176	0.0022	0.0221	0.0010	0.0229	0.0008	0.0231	0.0007
	λ	0.6126	0.7673	0.7984	0.7662	0.8937	0.8661	0.9164	0.8928
BE_{GEL2}	μ	0.0196	0.0022	0.0227	0.0010	0.0232	0.0008	0.0233	0.0007
	λ	0.5689	0.7046	0.7732	0.7246	0.8786	0.8386	0.9045	0.8708

decrease drastically, as raised by the well-known asymptotic efficiency of MLE. The MPS and WLS exhibit lower bias and RMSE compared to MLE for smaller sample sizes, but closely follow it as n increases. The LS, competes favorably for μ but not so good for λ whereby the RMSE remains high compared to others especially when the sample size is reduced. In general, CvM and RTAD tend to have higher RMSE values, more so when working with small sample sizes.

For example, these Bayesian estimators differ mostly in the fact that it is BE_{SEL}, BE_{Linex}, and BE_{GEL}. They behave in such a manner that, while their bias concerning μ remains constant among sample sizes, their bias and RMSE for λ are hugely, comparatively higher than that of frequentists' methods, as n grows beyond a point. This indicates that the Bayesian approaches may not work optimally in this scenario, probably due to prior assumptions influencing the estimates.

In summary, all methods work very well in estimating μ with very low bias and low RMSE. For λ, frequentist methods instead, including MLE, MPS, or WLS, perform better, especially at larger sample sizes. Regarding μ, the Bayesian estimators were stable, while for λ they keep being biased and have RMSE high and above others; hence, their use here might need further polishing or other prior specifications.

6. Real-Life Applications

The first data to be utilized to demonstrate the viability of the Wald distribution is the mortality rate due to HIV/AIDS in Germany from 2000 to 2020, obtained from https://platform.who.int/mortality/themes/theme-details/topics/indicator-groups/indicator-group-details/MDB/hiv-aids and presented in Table (6) below.

Table 6: Mortality rate of HIV/AIDs patients in Germany (Data-I)

0.70570244	0.65946256	0.62801346	0.6143952	0.61453596	0.59540885	0.61190438
0.56040019	0.53945593	0.52641369	0.55652406	0.56615856	0.50050448	0.49723726
0.47911586	0.3270769	0.34298934	0.35461942	0.37625366	0.41652161	0.45295061

The second dataset represent weekly death rate due to COVID-19 from 22/3/2020 to 20/12/2020 in Dominica. It was accessed from https://data.who.int/dashboards/covid19/data?n=c recorded in Table (7) below.

Table 7: Weekly death rate due to COVID-19 in Dominica (Data-II)

0.029850746	0.035363458	0.05292172	0.051236749	0.053061224	0.035197989	0.032082324
0.025607639	0.01929982	0.012975779	0.016090105	0.016615654	0.012140954	0.024329382
0.01301384	0.012219227	0.013564214	0.010826889	0.008958089	0.010932598	0.016161891
0.02201862	0.023125997	0.037598736	0.033069307	0.02618165	0.019748264	0.015044519
0.011788481	0.009540117	0.009618688	0.008082768	0.008333333	0.006851922	0.005307263
0.005720572	0.004095843	0.003398641	0.002529511	0.004015331		

The third dataset 74 measurements of gauge lengths (20 mm) studied by [13] reported in Table (8) below.

Table 8: Measurements of gauge lengths (20 mm) (Data-III)

1.312	1.314	1.479	1.552	1.700	1.803	1.861	1.865	1.944	1.958	1.966	1.997	2.006
2.021	2.027	2.055	2.063	2.098	2.140	2.179	2.224	2.240	2.253	2.270	2.272	2.274
2.301	2.301	2.359	2.382	2.382	2.426	2.434	2.435	2.478	2.490	2.511	2.514	2.535
2.554	2.566	2.570	2.586	2.629	2.633	2.642	2.648	2.684	2.697	2.726	2.770	2.773
2.800	2.809	2.818	2.821	2.848	2.880	2.809	2.818	2.821	2.848	2.880	2.954	3.012
3.067	3.084	3.090	3.096	3.128	3.233	3.433	3.585	3.585				

Data-IV contains 46 observations of repair times in hours for an airborne communication transceiver. The data was studied by [22] and reported in Table (9) below.

Table 9: Duration of active repairs for airborne communication transceivers (Data-IV)

0.50	0.60	0.60	0.70	0.70	0.70	0.80	0.80	1.00	1.00
1.00	1.00	1.10	1.30	1.50	1.50	1.50	1.50	2.00	2.00
2.20	2.50	2.70	3.00	3.00	3.30	4.00	4.00	4.50	4.70
5.00	5.40	5.40	7.00	7.50	8.80	9.00	10.20	22.00	24.50

Table 10: Basic Statistics for the Datasets

	Data-I	Data-II	Data-III	Data-IV
n	21	40	74	40
Q_1	0.4530	0.0094	2.1498	1.0000
Q_3	0.6119	0.0258	2.8158	4.7750
IQR	0.1590	0.0164	0.6660	3.7750
Mean	0.5203	0.0190	2.4773	4.0125
Median	0.5395	0.0143	2.5125	2.1000
Var	0.0119	0.0002	0.2378	26.6801
Std	0.1090	0.0136	0.4877	5.1653
Range	0.3786	0.0505	2.2730	24.0000
Skewness	-0.3335	1.0777	-0.1542	2.7171
Kurtosis	2.1101	3.4169	2.9512	10.5433
Outlier	None	0.05292172 0.05123675 0.05306122	None	None

Table (10) summarizes the statistics of the four data sets, which are labeled Data-I, Data-II, Data-III, and Data-IV. The data sets vary in sample size, from 21 observations (Data-I) up to 74 observations (Data-III). The first quartile and third quartile indicate how dispersed the data set might be. Data-I and Data-II have relatively low interquartile ranges (IQR), showing smaller spreads, while Data-IV has a notably higher IQR value of 3.7750, indicating a high degree of variability.

Looking at the central tendency, for all four datasets, mean values and median values are close, indicating that Data-I and Data-III are uniformly distributed, whereas Data-II and IV reveal their respective skewness. It states that Data-II is said to be positively skewed (1.0777) since the left-hand tail of the distribution is shorter causing the bulk of the data to lie on the left. Data-IV has a long and strong positive skew (2.7171), indicating that many large values are present, thus pulling the mean higher. Then again, Data-I and Data-III have negative values of skewness that were relatively small, suggesting minor hefts on the left.

The variance and standard deviation elucidate the degree of variability present in these datasets. Data-I and II showed less variability with standard deviations of 0.1090 and 0.0136, respectively. On the opposite, very high variance (26.6801) and standard deviation (5.1653) for Data-IV reinforce the notion of its great spread. An additional piece of evidence comes from the range, which finds Data-IV with a range of 24.0000 units as far as conceivable, much higher than the other datasets.

The kurtosis estimates indicate that Data-I and Data-III come close to appearing champions of the normal distribution in that their kurtosis estimates hover near the actual value of 3. Data-II has a kind of moderately peaked distribution (3.4169), while Data-IV, with a peakedness value of 10.5433, strongly indicates a dataset with extreme values or heavy tails.

Data-II is the only dataset with three outliers reported and, hence, can be regarded as containing extreme values that must be contributing to its positive skewness and raised kurtosis. The other datasets seem to be less contaminated in such a way.

Generally, Data-I and Data-III suggest relative symmetry and low dispersion, while Data-II has slight positive skewness and some outliers; Data-IV suggests high dispersion associated with strong positive skewness that could suggest possible extreme values or non-normal distribution.

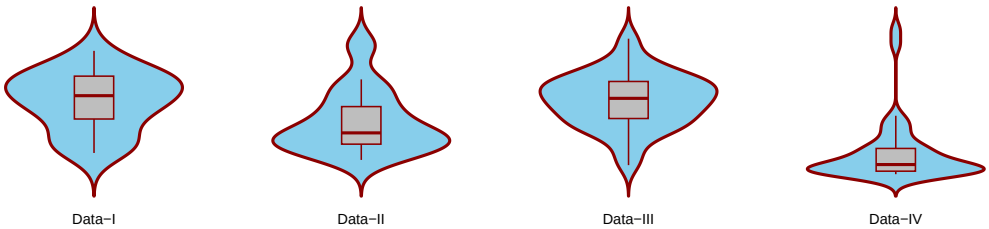


Figure 4: Boxplots superimposed on Violin plots

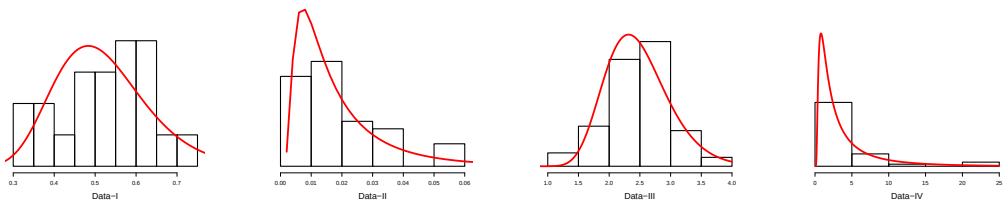


Figure 5: Density plots superimposed on Histogram

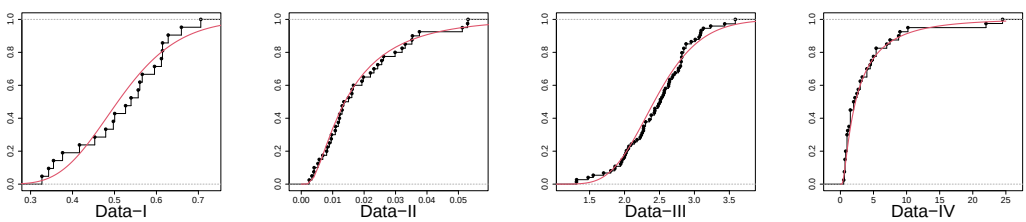


Figure 6: Empirical CDF with superimposed Wald CDF

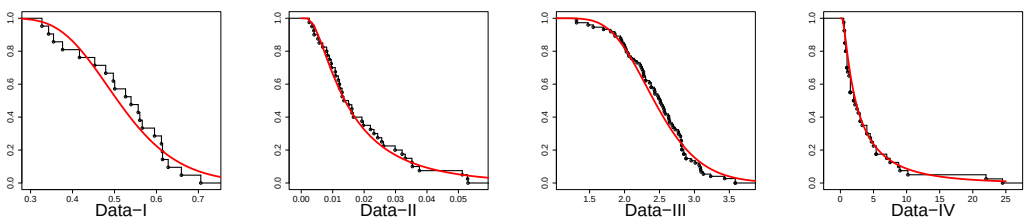


Figure 7: Empirical S(x) with superimposed Wald S(x)

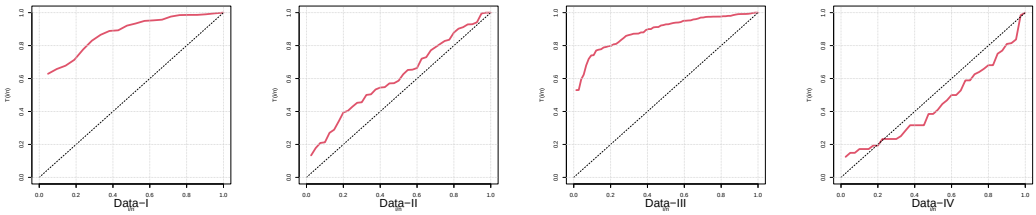


Figure 8: TTT plots

From the TTT plots shown in Figure (8), Data-IV, which is the TTT curve, exhibits a concave-down behavior that may indicate early failure occurrence. Such a situation generally concerns defective components failing early on in their lifecycle. On the other hand, Data-II shows a straight diagonal line on the TTT diagram. This line points toward constant failure times and rates of failure. This is typical for systems exhibiting an exponential distribution-the failures are random and unaffected by age. Once again Data-I and II lie above the diagonal (concave down) suggesting later failure, a classic signature for 'wear-out' failures where some degradation will have happened during later life.

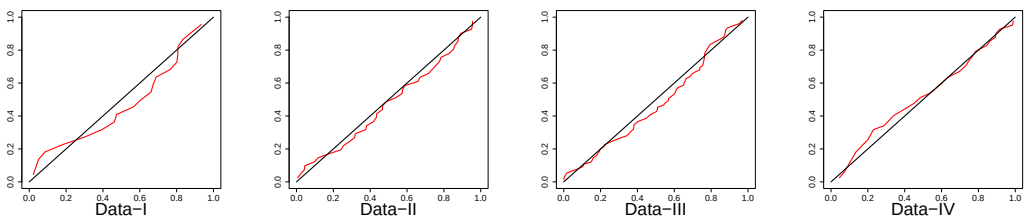


Figure 9: PP plots

Figure (9), most points lie below the diagonal for Data-I, II and II, which indicate that the fitted model underestimates the probability of extreme values (longer tails). This suggest that the datasets might have heavier tails than the Wald distribution. For Data-IV, the points closely follow the diagonal line ($y = x$), it suggests that the Wald model fits the data well. The deviation is minimal, hence its a well-fitted model.

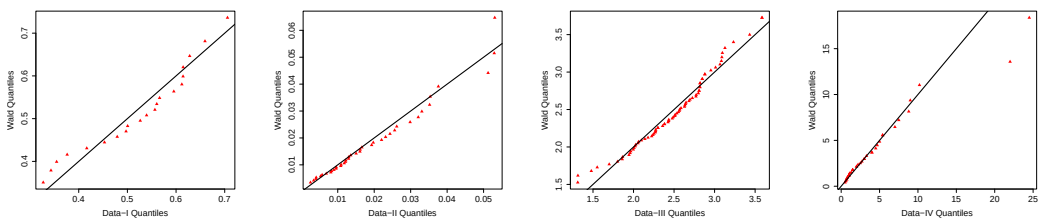


Figure 10: QQ plots

In Figure (10), Data-I and III points form a curve that bends upward (S-shape with points above the line at the right tail), the Wald distribution has heavier tails than the data. The plots suggest that the datasets follow the Wald distribution well. Minor deviations at the ends are acceptable.

Table 11: Model Performance Metrics

Distribution	Data	LL	AIC	CAIC	BIC	HQIC	W	A	KS	p-value
Wald		**16.26**	**-28.539**	**-27.862**	**-26.439**	**-28.075**	**0.102**	**0.640**	**0.137**	**0.7747**
Gumbel		15.4	-26.801	-26.134	-24.712	-26.348	0.130	0.799	0.144	0.7232
EAD	Data-I	15.55	-26.098	-26.432	-25.009	-26.645	0.125	0.079	0.141	0.7439
TI-HTR		10.76	-17.528	-16.862	-15.439	-17.075	0.084	0.543	0.197	0.3405
TI-HTE		-1.33	6.667	7.334	8.7566	7.121	0.091	0.091	0.343	0.0104
Wald		**123.19**	**-242.373**	**-242.049**	**-238.995**	**-241.152**	**0.035**	**0.278**	**0.076**	**0.9912**
Gumbel		121.26	-238.525	-238.200	-235.147	-237.303	0.075	0.479	0.111	0.6664
EAD	Data-II	123.51	-243.022	-242.698	-239.645	-241.801	0.034	0.232	0.085	0.9123
TI-HTR		121.88	-239.759	-239.435	-236.381	-238.538	0.053	0.337	0.148	0.3152
TI-HTE		117.71	-233.051	-232.727	-229.673	-231.830	0.034	0.227	0.172	0.1693
Wald		**-55.01**	**114.029**	**114.198**	**118.637**	**115.867**	**0.133**	**0.872**	**0.083**	**0.6820**
Gumbel		-58.59	121.172	121.341	125.780	123.011	0.211	1.364	0.096	0.5083
EAD	Data-III	-57.95	119.909	120.078	124.517	121.748	0.197	1.278	0.090	0.5800
TI-HTR		-76.51	157.021	157.191	161.630	158.860	0.082	0.524	0.224	0.0012
TI-HTE		-120.07	244.136	244.305	248.744	245.974	0.105	0.686	0.341	6.891×10^{-8}
Wald		**-89.24**	**182.489**	**182.813**	**185.867**	**183.710**	**0.047**	**0.344**	**0.110**	**0.7206**
Gumbel		-105.16	214.320	214.644	217.697	215.541	0.315	2.015	0.188	0.1182
EAD	Data-IV	-96.22	196.44	196.764	199.818	197.661	0.164	1.177	0.160	0.2597
TI-HTR		-113.18	230.36	230.684	233.738	231.581	0.202	1.386	3.853	0.0001
TI-HTE		-94.54	193.080	193.404	196.458	194.301	0.115	0.824	0.142	0.3971

The results summarized in Table (11) intend the comparative study of the performance of various probability distributions on four datasets, and various statistical metrics were used to assess their goodness of fit. For the model adequacy comparison of the likelihood-based criteria, the log-likelihood (LL), Akaike Information Criterion (AIC), Consistent AIC (CAIC), Bayesian Information Criterion (BIC), and the Hannan-Quinn Information Criterion (HQIC) were taken here. The results indicate that across all datasets, the Wald distribution has demonstrated a consistent capability of outperforming all other distributions. The Wald had higher LL values, hence providing the highest likelihood fit, while lower AIC, CAIC, BIC, and HQIC values reflect that it had the best model fit in relation to its complexity.

Goodness-of-fit tests were also conducted on the empirical data using alternative criteria Shapiro-Wilk statistic (W), the Anderson-Darling statistic (A), and the Kolmogorov-Smirnov (KS) statistic determine the adequacy of distributional fits. The KS statistic, which quantifies the maximum distance discrepancies between the empirical and theoretical cumulative distributions, is found to be comparatively lower for the Wald distribution as opposed to the other models. Further, the huge supporting evidence from the KS test on p-values yields that the Wald distribution fits the data well; they are kept well above the conventional significance level (0.05). Contrarily, the TI-HTR and TI-HTE distributions set very low p-values, indicating a poor fit.

When ranking the distributions, the Gumbel and EAD provide a moderate fit. Though they have lower LL values than the Wald distribution, their performance exceeds that of TI-HTR and TI-HTE. Their relatively high KS statistics also indicate that they fit the empirical data less well than the Wald distribution. For all datasets under consideration, the TI-HTR and TI-HTE distributions perform the worst, as shown by their extremely low LL values, high AIC and BIC results, and poor KS test results. In particular, the TI-HTE distribution fits the worst, showcasing the highest test statistics and the lowest p-values.

The general inferences indicate that among them all, the Wald distribution exhibits the most confident fit to the datasets under examination. The Wald distribution appears to consistently achieve the best compromise between fit quality and model complexity, hence the superiority in the performance of all criteria examined. The Gumbel and EAD distributions might serve as fair alternatives; however, merit-wise, they fall behind. TI-HTR and TI-HTE distributions are not to be considered under any circumstance since they are proved invalid. These results are an indication that the Wald distribution is the most compliant model for the respective datasets and supports its selection adequately by means of various statistical proofs.

Table 12: MLEs for the Datasets

Data	Parameter	Distribution Wald		Gumbel		EAD		TI-HTR		TI-HTE	
		μ location	λ shape	location	scale	shape	scale	shape	scale	shape	scale
Data-I	MLE	0.5203	10.5679	0.4656	0.1049	29.0335	5.5868	0.1875	22.5224	0.2174	10.2041
	Standard Error	0.0252	3.2613	0.0243	0.0171	18.3855	0.8410	0.1154	12.5954	0.0958	3.9770
Data-II	MLE	0.0190	0.0280	0.0130	0.0095	1.0320	53.0900	2.4847	402.4492	116.2300	0.0039
	Standard Error	0.0026	0.0056	0.0016	0.0012	0.2409	8.7326	1.3669	442.2347	24.1405	0.0015
Data-III	MLE	2.4770	55.1450	2.2310	0.4930	32.3180	1.1940	0.2020	0.9210	0.2200	2.1200
	Standard Error	0.0610	9.0658	0.0610	0.0410	10.7090	0.0920	0.0570	0.2350	0.0510	0.4340
Data-IV	MLE	4.0130	2.5190	2.2240	2.4760	0.5120	0.1660	3.8530	0.0030	3.3590	0.0270
	Standard Error	0.8010	0.5630	0.4030	0.3480	0.1080	0.0330	0.6270	0.0009	3.4310	0.0580

Table (12) shows the estimation of maximum likelihood in terms of parameter values and standard errors under different distributions fitted for each of the four datasets. The distributions considered were Wald, Gumbel, exponentiated Ailamujia distribution (EAD) by [33], type-I heavy-tailed Rayleigh (TI-HTR) by [16] and type-I heavy-tailed exponential (TI-HTE) by [15], while parameters such as location, shape and scale show variability between the distributions.

For Data-I, the Wald distribution shows an estimate for the location parameter of 0.5203 with a shape parameter of 10.5679; the Gumbel distribution approximates a location estimate of 0.4656 with a scale of 0.1049. The EAD distribution estimates a shape parameter of 29.0335 and scale 5.5868, while the TI-HTR and TI-HTE distributions are lower in values. The Wald and Gumbel distributions have lower standard errors, thus suggesting stable estimates, while the EAD distribution contains a higher standard error, especially for the shape parameter.

In Data-II, the parameter estimates from the Wald distribution were found to be distinctly smaller, at the location parameter 0.0190 and the shape parameter of 0.0280. The parameter estimates increase slightly for the Gumbel. For the EAD distribution, a large shape parameter of 1.0320 and a scale of 53.0900 were observed. TI-HTR and TI-HTE distributions displayed extreme parameter values, the TI-HTR shape and scale parameters being 2.4847 and 402.4492 and TI-HTE shape and scale parameters being 116.2300 and 0.0039. The standard errors for the TI-HTR and TI-HTE distributions are very high, indicating that the estimates might be unstable.

For Data-III, the Wald distribution was characterized by a location parameter of 2.4770 and a shape parameter of 55.1450, which was also found to have relatively high standard errors. The estimates for the Gumbel distribution are lower with a location of 2.2310 and a scale of 0.4930. High shape parameters of 32.3180 and lower scale of 1.1940 indicate the EAD is once again a good fit. The parameter estimates for the TI-HTR and TI-HTE have decreased in comparison to Data-II, though scale parameters decreased as well. The standard errors are moderate for the shape parameters.

The Wald distribution presents overall the highest location parameter for Data-IV among all datasets with a value of 4.0130, while a value of 2.5190 was estimated as the shape parameter. The Gumbel distribution estimates thus again a location of 2.2240 and a scale of 2.4760, consistent over all preceding datasets. The shape and scale for the EAD distribution are much smaller than for previous datasets, while the TI-HTR and TI-HTE distributions show diverse behavior in regard to their parameters. Standard errors in this dataset show great variability, while certain parameters from the Wald and TI-HTE distributions showed relatively high standard errors.

The overall differences in the estimated MLEs for the different datasets emphasize the ability of each distribution to perfectly model the data. The Wald and Gumbel distributions retain relatively stable estimates, while the EAD, TI-HTR, and TI-HTE distributions show more variation, particularly in their shape parameters. These results suggest that dataset characteristics would be paramount for parameter estimation, with some distributions fitting certain datasets more reliably than others.

7. Concluding Remarks

The current study evaluates the estimation procedures of the parameters of the Wald distribution in great length and detail, with respect to numerous applications studied with different datasets. As expected from statistical theory, the simulated results showed that for all estimation techniques, a greater sample

size generally results in improved estimations, reduced bias, and RMSE. Pertinent features included the fact that, generally speaking, MLE gave efficient estimations for larger sample sizes, but was found to be highly erratic for small sample sizes, particularly in estimating parameter λ. In this regard, on the other hand, MPS demonstrated good asymptotic efficiency with bias and RMSE steadily decreasing with an increase in sample size, although that in itself is not to say that it was favorable for parameter μ in relation to that of MLE for small sample sizes.

The LS and WLS methods had a fairly similar performance, as bias tended to decrease with increasing sample sizes; thus, LS and WLSE could be possible alternatives for any application involving a range of small and moderate sample sizes. In contrast, the CvM method yielded relatively higher RMSE values and did not perform efficiently in most settings, indicating its weaknesses. Regarding the Bayesian estimators (BE), they tended to have more bias than their frequentist counterparts and to be less efficient, especially concerning parameter λ. There could be some merit in Bayesian strategies if the prior information could be assessed credibly; however, the greater uncertainty stems from prior settings that inflate the RMSE. The Wald distribution is further substantiated by the application to real-world datasets. With the HIV/AIDS mortality data in Germany, the weekly COVID-19 death rates in Dominica, and the gauge length measurements in the industry, the Wald distribution denotes feasible applicability in various settings, from health data analysis to reliability studies. The various statistical summaries for these datasets underscore the variability and underlying structure in the data that justify choosing the Wald distribution to model these real-world phenomena.

This study strongly proves that the Wald distribution with the efficient estimation methods, i.e., MLE and MPS, can be a potent instrument for parameter estimation in applied statistics and shows great promise in solving other complex real-life problems in various disciplines. Further research may consider the combination of advanced Bayesian methods with carefully selected priors for increased efficiency and stability in estimating Wald distribution parameters.

Conclusions drawn from the parameter estimates over the four datasets provide hints of the difference in performance between the distributions. The Wald and Gumbel distributions yield estimates that are most commonly stable and, in particular, characterized by lower standard errors, especially concerning the location and shape parameters. These distributions may be regarded as robust over all datasets, and the Wald distribution is especially criticized for having fewer degrees of freedom from the location parameter when considering all datasets.

Conversely, none was as stable in its estimates as the EAD, by which it displays a thorough amount of variability, particularly for its shape parameter. In other words, the shape parameter of the EAD distribution can very well fit some datasets but may not provide stable estimates. The EAD distribution would therefore be sensitive to data variability and hence its practical application may be limited.

The TI-HTR and TI-HTE distributions show the highest amount of variation between datasets, where on some occasions they induce very high parameter estimates and yet on others present extreme parameter estimates, with correspondingly high standard errors on both occasions, leading to potential instability in their parameter estimation; hence these distributions may not seem to provide a reliable parameter estimate consistently, especially within those datasets when the data characteristics result in high variability.

In summary, based on found operations, we can stipulate that the choice of distribution is important and should be dictated by the peculiarity of the dataset considered. Thus, while Wald and Gumbel distributions seem to be the most reliable for most of the datasets studied, the EAD, TI-HTR, and TI-HTE distributions may be more suitable in those contexts where their larger variability can be accounted for. The consistency of the results reiterates the decisive impact data nature holds in making a judgment of a particular model for parameter estimation.

References

1. Akman, O., and Gupta, R. C., *A comparison of various estimators of the mean of an inverse Gaussian distribution*, Journal of Statistical Computation and Simulation, 40, 1-2, 71–81, (1992). DOI: https://doi.org/10.1080/00949659208811366

2. Alzaatreh, A., Lee, C., and Famoye, F., *A new method for generating families of continuous distributions*, Metron, 71, 1, 63–79, (2013). DOI: https://doi.org/10.1007/s40300-013-0007-y

3. Alzaatreh, A., Lee, C., and Famoye, F., *T-normal family of distributions: a new approach to generalize the normal distribution*, Journal of Statistical Distributions and Applications, 1, 1–18, (2014). DOI: https://doi.org/10.1186/2195-5832-1-16

4. Anders, R., Alario, F., Van Maanen, L., et al., *The shifted Wald distribution for response time data analysis*, Psychological methods, 21, 3, 309, (2016).

5. Aslam, M., Jun, C. H., Lio, Y. L., Ahmad, M., and Rasool, M., *Group acceptance sampling plans for resubmitted lots under Burr-type XII distributions*, Journal of the Chinese institute of Industrial Engineers, 28, 8, 606–615, (2011). DOI: https://doi.org/10.1080/10170669.2011.651165

6. Cheng, R., and Amin, N., *Maximum product of spacings estimation with application to the lognormal distribution (Mathematical Report 79-1)*, Cardiff: University of Wales IST, (1979).

7. Chhikara, R., and Folks, J. L., *The inverse Gaussian distribution: theory: methodology, and applications*, CRC Press, (2024).

8. Cordeiro, G. M., Pescim, R. R., and Ortega, E. M. M., *The Kumaraswamy generalized half-normal distribution for skewed positive data*, Journal of Data Science, 10, 2, 195–224, (2012).

9. Ducharme, G. R., *Goodness-of-fit tests for the inverse Gaussian and related distributions*, Test, 10, 271–290, (2001). DOI: https://doi.org/10.1007/BF02595697

10. Eugene, N., Lee, C., and Famoye, F., *Beta-normal distribution and its applications*, Communications in Statistics-Theory and methods, 31, 4, 497–512, (2002). DOI: https://doi.org/10.1081/STA-120003130

11. Heathcote, A., *Fitting Wald and ex-Wald distributions to response time data: An example using functions for the S-PLUS package*, Behavior Research Methods, Instruments, & Computers, 36, 678–694, (2004). DOI: https://doi.org/10.3758/BF03206550

12. Jamal, F., Chesneau, C., and Elgarhy, M., *Type II general inverse exponential family of distributions*, Journal of Statistics and Management Systems, 23, 3, 617–641, (2020). DOI: https://doi.org/10.1080/09720510.2019.1668159

13. Kundu, D., and Raqab, M. Z., *Estimation of R= P (Y¡ X) for three-parameter Weibull distribution*, Statistics & Probability Letters, 79, 17, 1839–1846, (2009). DOI: https://doi.org/10.1016/j.spl.2009.05.026

14. Meyer-Grant, C. G., *Conjugate Bayesian analysis of the Wald model: On an exact drift-rate posterior*, Journal of Mathematical Psychology, 124, 102904, (2025). DOI: https://doi.org/10.1016/j.jmp.2025.102904

15. Nwankwo, B. C., Obiora-Ilouno, H. O., Almulhim, F. A., SidAhmed Mustafa, M., and Obulezi, O. J., *Group acceptance sampling plans for type-I heavy-tailed exponential distribution based on truncated life tests*, AIP Advances, 14, 3, (2024). DOI: https://doi.org/10.1063/5.0194258

16. Nwankwo, M. P., Alsadat, N., Kumar, A., Bahloul, M. M., and Obulezi, O. J., *Group acceptance sampling plan based on truncated life tests for Type-I heavy-tailed Rayleigh distribution*, Heliyon, 10, 19, (2024). DOI: https://doi.org/10.1016/j.heliyon.2024.e38150

17. Obulezi, O. J., Obiora-Ilouno, H. O., Osuji, G. A., Kayid, M., and Balogun, O. S., *Weibull Sine Generalized Distribution Family: Fundamental Properties, Sub-model, Simulations, with Biomedical Applications*, Electronic Journal of Applied Statistical Analysis, 18, 01, 183–212, (2025). DOI: https://doi.org/10.1285/i20705948v18n1p183

18. Obulezi, O. J., Obiora-Ilouno, H. O., Osuji, G. A., Kayid, M., and Balogun, O. S., *A new family of generalized distributions based on logistic-x transformation: sub-model, properties and useful applications*, Research in Statistics, 3, 1, 2477232, (2025). DOI: https://doi.org/10.1080/27684520.2025.2477232

19. Rao, C. R., *A natural example of a weighted binomial distribution*, The American Statistician, 31, 1, 24–26, (1977). DOI: https://doi.org/10.1080/00031305.1977.10479187

20. Seshadri, V., and Seshadri, V., *Reliability and Survival Analysis*, The Inverse Gaussian Distribution: Statistical Theory and Applications, 92–113, (1999). DOI: https://doi.org/10.1007/978-1-4612-1456-4_5

21. Srisuradetchai, P., Niyomdecha, A., and Phaphan, W., *Wald intervals via profile likelihood for the mean of the inverse gaussian distribution*, Symmetry, 16, 1, 93, (2024). DOI: https://doi.org/10.3390/sym16010093

22. Von Alven, W. H., et al., *Reliability engineering*, (No Title), (1964).

23. Wald, A., *On cumulative sums of random variables*, The Annals of Mathematical Statistics, 15, 3, 283–296, (1944). URL: https://www.jstor.org/stable/2236250

24. Yang, H., Huang, M., Chen, X., He, Z., and Pu, S., *Enhanced Real-Life Data Modeling with the Modified Burr III Odds Ratio–G Distribution*, Axioms, 13, 6, 401, (2024). DOI: https://doi.org/10.3390/axioms13060401

25. Zigangirov, K. S., *Expression for the Wald distribution in terms of normal distribution*, Radio Engineering and Electronic Physics, 7, 145–48, (1962).

26. Obulezi, O. J., *Obulezi distribution: a novel one-parameter distribution for lifetime data modeling*, Modern Journal of Statistics, 2, 1, 32–74, (2026). DOI: https://doi.org/10.64389/mjs.2026.02140

27. Chesneau, C., *Theory on a new bivariate trigonometric Gaussian distribution*, Innovation in Statistics and Probability, 1, 2, 1–17, (2025). DOI: https://doi.org/10.64389/isp.2025.01223

28. Gemeay, A. M., Moakofi, T., Balogun, O. S., Ozkan, E., and Hossain, M. M., *Analyzing real data by a new heavy-tailed statistical model*, Modern Journal of Statistics, 1, 1, 1–24, (2025). DOI: https://doi.org/10.64389/mjs.2025.01108

29. Mousa, M. N., Moshref, M. E., Youns, N., and Mansour, M. M. M., *Inference under Hybrid Censoring for the Quadratic Hazard Rate Model: Simulation and Applications to COVID-19 Mortality*, Modern Journal of Statistics, 2, 1, 1–31, (2026). DOI: https://doi.org/10.64389/mjs.2026.02113

30. Noori, N. A., Abdullah, K. N., et al., *Development and applications of a new hybrid Weibull-inverse Weibull distribution*, Modern Journal of Statistics, 1, 1, 80–103, (2025). DOI: https://doi.org/10.64389/mjs.2025.01112

31. Onyekwere, C. K., Aguwa, O. C., and Obulezi, O. J., *An updated lindley distribution: Properties, estimation, acceptance sampling, actuarial risk assessment and applications*, Innovation in Statistics and Probability, 1, 1, 1–27, (2025). DOI: https://doi.org/10.64389/isp.2025.01103

32. Orji, G. O., Etaga, H. O., Almetwally, E. M., Igbokwe, C. P., Aguwa, O. C., and Obulezi, O. J., *A new odd reparameterized exponential transformed-x family of distributions with applications to public health data*, Innovation in Statistics and Probability, 1, 1, 88–118, (2025). DOI: https://doi.org/10.64389/isp.2025.01107

33. Rather, A. A., Subramanian, C., Al-Omari, A. I., and Alanzi, A. R. A., *Exponentiated Ailamujia distribution with statistical inference and applications of medical data*, Journal of Statistics and Management Systems, 25, 4, 907–925, (2022). DOI: https://doi.org/10.1080/09720510.2021.1966206

Chinyere P. Okechukwu,
Department of Statistics,
School of Chemical Engineering and Physical Sciences,
Lovely Professional University, Phagwara 144411, Punjab,
India.
E-mail address: chinyere.12404612@lpu.in

and

Manzoor A. Khanday,
Department of Statistics,
School of Chemical Engineering and Physical Sciences,
Lovely Professional University, Phagwara 144411, Punjab,
India.
E-mail address: manzoorstat@gmail.com

and

Okechukwu J. Obulezi,
Department of Statistics,
Faculty of Physical Sciences,
Nnamdi Azikiwe University, P. O. Box 5025, Awka,
Nigeria.
E-mail address: oj.obulezi@unizik.edu.ng

and

Mohamed A. F. Elbarkawy,
Department of Insurance and Risk Management
College of Business, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432,
Saudi Arabia.
E-mail address: maalbarqawi@imamu.edu.sa

and

Ehab M. Almetwally,
Department of Mathematics and Statistics,
College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432,
Saudi Arabia.

E-mail address: emalmetwally@imamu.edu.sa

and

Mohammed Elgarhy,
Faculty of Computers and Information Systems,
Egyptian Chinese University,
Egypt;
Department of Computer Engineering,
Biruni University, 34010, Istanbul,
Turkey.
E-mail address: dr.moelgarhy@gmail.com