Eksploatacja i Niezawodnosc – Maintenance and Reliability

Volume 28 (2026), Issue 4

journal homepage: http://www.ein.org.pl

Article citation info:
Zhang Y, Han F, Yang F, Sun M, Yang H, Liu J, Research on parameters estimation method of three-parameters log-normal distribution for automotive batteries under small sample data, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2026: 28(4) http://doi.org/10.17531/ein/219912

Research on parameters estimation method of three-parameters log-normal distribution for automotive batteries under small sample data

Indexed by:

Yingzhi Zhang[a,b], Feng Han[a,b], Fang Yang[c], Mo Sun[a,b], Haipeng Yang[a,b], Jialin Liu[a,b,*]

[a] School of Mechanical and Aerospace Engineering, Jilin University, China
[b] Key Laboratory of CNC Equipment Reliability, Ministry of Education, China
[c] NEV Inst, China FAW Group Co., Ltd., China

Highlights

- A three-parameter estimation method for lognormal distribution considering global error.
- Optimization of empirical distribution formulas using linear correlation coefficients.
- The parameter estimation accuracy is evaluated using the RMSE.

Abstract

To address the issue that the classical LSE and MLE neglect global errors when estimating the two parameters of the log-normal distribution with small samples, a three-parameter estimation method integrating empirical distribution function model optimization, threshold parameter estimation, and cumulative sum of squared errors minimization is proposed. The initial dataset is derived via inverse transformation of the empirical distribution, and the empirical distribution model is optimized using the linear correlation coefficient. The interpolation method is employed to estimate the threshold parameter corresponding to the maximum linear correlation coefficient. The particle swarm optimization（PSO）algorithm optimizes the location and scale parameters to minimize the cumulative sum of squared errors. The K-S test is applied to evaluate the fitting performance, and the RMSE is used as an indicator to verify the accuracy of the proposed method. An empirical study is conducted by combining 13 sets of actual lifetime data and simulated data of a certain automotive battery.

Keywords

small sample size, automotive batteries, three-parameter log-normal distribution, parameter estimation, cumulative sum of squared errors

1. Introduction

With the continuous deepening of research on the reliability of automotive batteries, the reliability level of such products has been steadily improved, which makes it difficult to obtain life data in a short period of time. Therefore, how to make full use of limited information to achieve more accurate parameter estimation for the life reliability model of automotive batteries has become an important problem that urgently needs to be solved at present [1,2].

Under small sample conditions, an effective approach to improve the accuracy of reliability modeling is to increase information volume. Currently, methods such as the Bootstrap method [3–5], Maximum Likelihood Estimation [6,7], Bayesian method [8], Least Squares Estimation, and machine learning algorithms have been applied to reliability modeling of products with small sample sizes. Among these methods, the Bootstrap method is a resampling technique relying on computer technology. It only requires actual observed data without the need to predefine the distribution type of the data, thus offering convenience in practical data processing. The modified Maximum Likelihood Estimation [9,10] is mainly designed for

(*) Corresponding author.
E-mail addresses: Y. Zhang (ORCID: 0000-0002-2856-5156) zhangyz@jlu.edu.cn, F. Han (ORCID: 0009-0002-0900-7922) hanfeng23@mails.jlu.edu.cn, F. Yang, yangfang1@faw.com.cn, M. Sun, 605402283@qq.com, H. Yang, 584029697@qq.com, J. Liu, 1094267138@qq.com

scenarios with partial or complete zero-failure data under time-censored tests; notably, the higher the censoring degree, the more pronounced the modification effect. The key challenge of the Bayesian method lies in determining an appropriate prior distribution, which is generally not easy to achieve in practical applications. Although the classical Least Squares Estimation satisfies the unbiasedness criterion, the presence of heteroscedasticity renders the regression model statistically insignificant [11]. To address this issue, relevant scholars have proposed the weighted least squares estimation, which not only meets the unbiasedness requirement but also achieves the property of the Best Linear Unbiased Estimation. However, Yu [12] mathematically proved that even though this method minimizes the error variance when the weight is set as the reciprocal of the variance of the disturbance term at each sample point, it still fails to satisfy the efficiency criterion. With the development of computer science and statistics, machine learning algorithms have found increasingly extensive applications. In 1998, Vapnik V [13] proposed Support Vector Machines, which have been widely adopted due to their functional versatility and excellent performance in handling nonlinear and small-sample problems. Zheng [14,15] proposed a decision-making model, which is of great significance for maintenance decision-making. In 1999, Suykens et al. [16] proposed the Least Squares Support Vector Machine, which replaces the inequality constraints in the standard Support Vector Machine with equality constraints and modifies the loss function to the least squares loss. This modification converts the original quadratic programming problem into solving a system of linear equations, thereby reducing computational complexity and improving computational speed. As a result, LSSVM has been widely applied in parameter estimation [17–20]. Nevertheless, this model has inherent drawbacks: it cannot guarantee the optimal solution when solving linear equations and exhibits weak robustness against outliers, which to a certain extent limits its effectiveness in practical applications [21].

Current research on life distribution modeling is mainly based on life data. It assumes the distribution that the model follows, and then constructs the reliability model through parameter estimation and hypothesis testing [22]. Different parameter estimation methods yield varying levels of parameter estimation accuracy. A three-parameter estimation method for the lognormal distribution considering the global error is proposed for skewed data. The linear correlation coefficient is applied to optimize the empirical distribution model, and the interpolation method is adopted to estimate the threshold parameter based on the linear correlation coefficient. Taking the minimum sum of squared cumulative errors as the objective, the PSO algorithm is used to optimize the location and scale parameters. The Kolmogorov-Smirnov (K-S) test is employed to verify the model fitting effect, and the root mean square error (RMSE) is used as an index to evaluate the parameter estimation performance. This method reduces the empirical distribution model error caused by rank-transformed probability, takes the impact of the threshold parameter on the model into account, and improves the modeling accuracy. Finally, the proposed method is applied in detail with the automotive battery lifetime data and simulated data from Reference [8] as examples. Its effectiveness is verified by comparing the proposed parameter estimation method with the maximum likelihood estimation, least squares method and other conventional methods.

2. Parameter estimation of log-normal distribution with minimum cumulative sum of squared errors

The parameter estimation method with minimum cumulative sum of squared errors uses the univariate linear regression method to fit and obtain the initial value of the failure distribution function $F_0(t)$. Then, it calculates the sum of squared cumulative errors between $F_0(t)$ and F_n derived from the empirical distribution function. With the goal of minimizing the sum of squared cumulative errors, the particle swarm optimization algorithm is adopted to determine the optimal parameter estimates, thereby deriving the failure distribution function $F(t)$. The K-S test is applied to conduct goodness-of-fit verification, and the RMSE is calculated to evaluate the performance of the proposed method.

The three-parameter estimation procedure of the lognormal distribution based on minimizing the sum of squared cumulative errors is shown in Figure 1.

2.1. Initial dataset determined and normality test

2.1.1. Initial dataset determined

Suppose n samples be used for the life test, obtaining n life data $t_1 \le t_2 \le \cdots \le t_n$. If the life data follows a three-parameter log-normal distribution $LN(\mu, \sigma, \theta)$ according to the Shapiro-Wilk test (referred to as the S-W test), then its distribution function is:

$$F(t;\mu,\sigma,\theta)=\int_{-\infty}^{ln(t-\theta)}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx, t>\theta, \sigma>0, -\infty<\mu<+\infty \ =\Phi\left\{\frac{ln(t-\theta)-\mu}{\sigma}\right\}, t>\theta \tag{1}$$

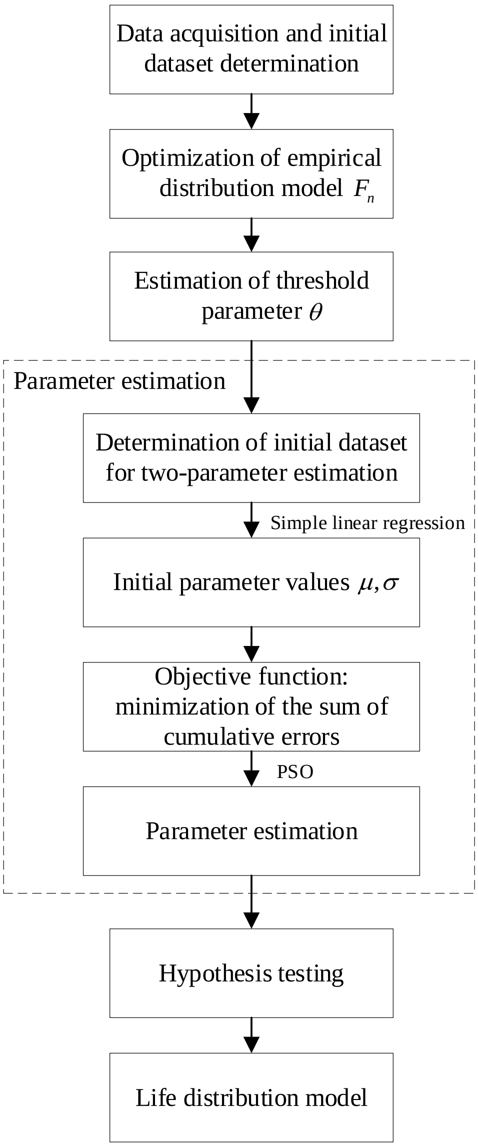


Figure 1. Flowchart of the three-parameter estimation of log-normal distribution based on the minimum cumulative squared error.

Where, μ is called the location parameter, which is the logarithmic mean; $\sigma>0$ is called the scale parameter, which is the logarithmic standard deviation; $\theta>0$ is called the threshold parameter, used to limit the range of values of the distribution and control the shift and shape of the distribution. When the threshold parameter $\theta=0$, the three-parameter log-normal distribution simplifies to the two-parameter log-normal distribution $F(t;\mu,\sigma)=\Phi\left\{\frac{ln\, t-\mu}{\sigma}\right\}$.

Under the assumption of a two-parameter log-normal distribution, a linearized dataset $S=\{(\boldsymbol{x}_i,y_i),\ldots,(\boldsymbol{x}_n,y_n)\}$ can be obtained by setting $x=ln\, t$ and $y=\Phi^{-1}[F(t)]$.

2.1.2. Shapiro-Wilk test

The Shapiro-Wilk test was proposed by Samuel Shapiro and Martin Wilk in 1965 as a method to verify whether a random sample of data is derived from a normal distribution. Its principle is to compare the degree of agreement between the order statistics of the sample and the expected order statistics under the normal distribution. If a sample strictly follows a normal distribution, the ordered values of the sample (values sorted in ascending order) will be highly linearly correlated with the corresponding expected ordered values of the normal distribution; if the sample deviates from the normal distribution, such linear correlation will decrease significantly.

The Shapiro-Wilk test calculates a test statistic based on the order statistics $x_1\le x_2\le\cdots\le x_n$ of n sample data points.

$$W_n=\frac{\left[\sum_{i=1}^{\left[\frac{n}{2}\right]}a_i(x_{n+1-i}-x_i)\right]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \tag{2}$$

Where, $\bar{x}$ denotes the mean of the sample observed values, $[n/2]$ represents the integer part of $n/2$, and the weight coefficient a_i is obtained from the quantile table of the test statistic.

According to the given significance level α and sample size n, query the p -quantile table of the test statistic W to determine the critical value of the α -quantile, denoted as $W_{n,\alpha}$. If $W_n<W_{n,\alpha}$, the normality hypothesis is rejected, and the data is considered to not follow a normal distribution; otherwise, the normality hypothesis is accepted. Alternatively, at a significance level of α, if $p=P\left(W_n\ge W_{n,\alpha}\right)>\alpha$, the normality hypothesis is deemed valid, meaning the product lifetime conforms to the hypothesized normal distribution.

2.2. Experience distribution model optimization

The rank method is used to process experimental data, sorting the data in ascending order as $t_1\le t_2\le\cdots\le t_n$, and then approximating the cumulative frequency of event occurrences as its statistical probability $F(t_i)$. There are various different approximate methods for transforming ranks into probabilities in various materials. For small samples ($n \leq 20$), several commonly used approximate formulas are as follows:

Hansen formula:

$$F_n(t_i) = \frac{i-0.5}{n} \quad (3)$$

Mathematical expectation formula:

$$F_n(t_i) = \frac{i}{n+1} \quad (4)$$

Approximate median rank formula:

$$F_n(t_i) = \frac{i-0.3}{n+0.4} \quad (5)$$

Equations (3) to (5) yield different errors in the parameter estimation for different distribution models.

Assuming the lifetime data follows a two-parameter log-normal distribution, let $x = \ln t$ and $y = \Phi^{-1}[F(t)]$, which can be linearized to $y = -\frac{\mu}{\sigma} + \frac{1}{\sigma}x$, where $\mu = \frac{1}{n}\sum_{i=1}^{n} \ln t_i$, $\sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\ln t_i - \hat{\mu})^2}$; $F(t)$ is calculated based on the approximate formulas (3), (4), and (5). However, whether there is a truly linear correlation between the variables x and y requires a linear correlation test.

The linear correlation coefficient is calculated using formula (6).

$$\hat{\rho} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2)}} \quad (6)$$

Where, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$,$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

When $|\hat{\rho}| > \rho_\alpha$, It is widely recognized that a linear correlation exists between x and y. Here, ρ_α is the minimum value of the correlation coefficient at the significance level α, which can be referenced in the literature [23] and obtained from the table of minimum value of the correlation coefficient. When the significance level $\alpha = 0.1$, the approximate formula for the minimum value of the correlation coefficient is $\rho_\alpha = \frac{1.645}{(n-1)^{\frac{1}{2}}}$.

Based on three approximate formulas (3), (4), and (5), calculate $y = \Phi^{-1}[F_n(t)]$, and use the linear correlation coefficient calculated from $y = \Phi^{-1}[F_n(t)]$ and $x = \ln t$ as the criterion for evaluating the quality of the empirical distribution function model.

2.3. Threshold parameter estimation

There are many different criteria to measure how well a set of data fits a theoretical curve. Since the three-parameter log-normal distribution can be linearized as $y = -\frac{\mu}{\sigma} + \frac{1}{\sigma}x$ with $x = \ln(t-\theta)$ and $y = \Phi^{-1}[F_n(t)]$, where $F_n(t)$ is the formula for the preferred empirical distribution function

Therefore, this article uses interpolation methods to estimate the threshold parameters of the log-normal distribution under the maximum linear correlation coefficient.

The specific steps are:

(1) Determine the threshold parameter θ and calculation precision Δ based on the actual situation and calculation needs.

(2) It can be concluded from Reference [24] that when the interpolation method is used to estimate the value of θ, the accuracy is in the order of $l, \frac{2}{3}l, \frac{2}{3}\times\frac{1}{2}l, \cdots, \frac{2}{3}\left(\frac{1}{2}\right)^{k-2}l, \cdots$, and the condition $\lim_{t\to t_0}\rho(x,y)=1$ holds. Given that the length l satisfies $[\theta_{00},\theta_{01}]$, $\theta_{00} = 0, \theta_{01} = x_1$, it follows that $l = x_1$. Suppose that the required precision Δ is achieved after k iterative steps, thus yielding $\frac{2}{3}\left(\frac{1}{2}\right)^{k-2}l = \Delta$. Taking the logarithm of both sides of this equation, we obtain $k = -\frac{(\ln 3)\times\Delta - \ln(2l)}{\ln 2} + 2$.

(3) Taking the endpoint values within the interval $[0, x_1]$ as the initial values, calculate the corresponding correlation coefficients ρ_{00} and ρ_{01} between x and y using Equation (6).

(4) Insert $\frac{1}{3}x_1$ and $\frac{2}{3}x_1$ into the interval $[0, x_1]$ to divide it into three equal subintervals. Calculate the correlation coefficients ρ_{011} and ρ_{012} between x and y corresponding to the interpolation points using Equation (6). Let $\rho_{02} = max(|\rho_{011}|, |\rho_{012}|)$, and take the interpolation point $\frac{1}{3}x_1$ or $\frac{2}{3}x_1$ corresponding to the larger correlation coefficient as the secondary value of θ. Taking ρ_{011} as an example, $\frac{1}{3}x_1$ is then the secondary value of θ, and the new interval will be $\left[0, \frac{1}{3}x_1\right]$ in this case; otherwise, the new interval will be $\left[\frac{2}{3}x_1, x_1\right]$.

(5) Take the secondary value of θ as the endpoint and bisect the new interval. Calculate the correlation coefficient ρ_{021} between x and y corresponding to the interpolation point using Equation (6). If the new interval is $\left[0, \frac{1}{3}x_1\right]$, compare $\rho_{021}\left(\frac{x_1}{6}\right)$, $\rho(0)$ and $\rho\left(\frac{x_1}{3}\right)$, assign $\rho_{03} = max\left(|\rho_{021}|, |\rho(0)|, \left|\rho\left(\frac{x_1}{3}\right)\right|\right)$, and take the interpolation point corresponding to the largest correlation coefficient as the tertiary value of θ.

(6) By analogy, select the θ_{0i} corresponding to the maximum value of ρ_{0i}. Under this condition, the fitting effect of the determined straight line to the data points is optimal, and θ_{0i} is the desired location parameter. That is, at this point, $\hat{\theta} = \theta_{0i}$ and the correlation coefficient $\rho = max(\theta_{0i})$.

2.4. Parameter estimation method based on minimizing the sum of squared cumulative errors

2.4.1. Determination of the objective function

Taking the minimization of the sum of squared cumulative errors as the objective, a parameter estimation model for the lognormal distribution is established. The goal function is expressed as follows:

$$min\, f\,(\mu,\sigma^2) = min \sum_{i=1}^{n}[F(\tau_i,\mu,\sigma) - F_n(\tau_i)]^2 \quad (7)$$

Where, $F(\tau_i,\mu,\sigma)$ is the failure distribution function following the two-parameter lognormal distribution. It can be derived from Equation (1) that $F(\tau;\mu,\sigma) = \Phi\left\{\frac{ln\,\tau-\mu}{\sigma}\right\}$. Let $\tau = (t - \hat{\theta})$ and $x = ln\,\tau$, and set $y = \frac{ln\,\tau-\mu}{\sigma}$; then $\Phi(y)$ follows the standard normal distribution function. $F_n(t_i)$ denotes the empirical distribution function.

2.4.2. Calculation procedure for lognormal distribution parameters

a. Calculation of the initial value $(\hat{\mu}_0, \hat{\sigma}_0)$ of the parameter

Calculate the empirical distribution function values $F_n(\tau_i)$ with the life data, and obtain the initial value $(\hat{\mu}_0, \hat{\sigma}_0)$ of the parameters of $F(\tau_i,\mu,\sigma)$ by using the univariate linear regression method.

$$\begin{cases} \hat{\mu}_0 = \frac{1}{n}\sum_{i=1}^{n} ln\,\tau_i \\ \hat{\sigma}_0 = \frac{1}{n}\sum_{i=1}^{n}(ln\,\tau_i - \hat{\mu}_0)^2 \end{cases} \quad (8)$$

b. Optimal parameter estimation suing the PSO algorithm

To minimize the sum of squared cumulative errors, we leverage the PSO algorithm to ascertain the optimal parameter estimates, and the lognormal distribution function is consequently derived.

PSO draws inspiration from the foraging activities of bird flocks and serves to address optimization tasks. Within this framework, each particle stands for a feasible solution to the target problem, and is associated with a fitness score calculated via a dedicated fitness function. For each particle in PSO, its velocity is adjusted dynamically based on both its personal movement history and the experiences of other particles in the swarm, which in turn enables the optimization of individual solutions within the solution space.

In this paper, the formula $F(t) = \Phi\left(\frac{ln\,\tau-\mu}{\sigma}\right)$ is adopted as the objective function of the PSO algorithm. The parameter values obtained by univariate linear regression are used as the initial values of particles. The life data and empirical distribution are substituted into the fitness function of the algorithm to calculate the particle fitness. Within this algorithmic framework, each particle stands for a feasible candidate solution to the target problem. The number of particles is generally set to a range of 20 to 60, and 40 particles are selected in this paper. According to Reference [25], the maximum allowable velocity of the particles is set to 0.8, and the initial velocity of particles is set to 0.8×rand (1,2), where rand (1,2) denotes that each particle randomly generates a set of 2-dimensional initial velocity values. The velocity and position of particles are updated in accordance with Equations (9) and (10). The stopping condition of the algorithm is set as the number of iterations exceeding a specific value (MaxNum).

$$V_{hd}^{k+1} = V_{hd}^{k} + c_1 r_1 \left(P_{hd}^{k} - X_{hd}^{k}\right) + c_2 r_2 \left(P_{gd}^{k} - X_{hd}^{k}\right) \quad (9)$$

$$X_{hd}^{k+1} = X_{hd}^{k} + V_{hd}^{k+1} \quad (10)$$

Where, V denotes velocity, X denotes position, and h represents the h-th particle with $h = 1,2,\cdots,40$. c_1 and c_2 are learning factors, which represent the particle's individual learning ability and global learning ability respectively. They are generally non-negative constants, usually ranging from 0 to 2. According to Reference [26], $c_1 = c_2 = 1.494$ is adopted. k is the current number of iterations with $1 \leq k \leq MaxNum$; r_1 and r_2 are random numbers within the interval [0,1]. d represents the d-th estimated parameter of each particle. Since the number of estimated parameters is 2, $d = 1,2$, which correspond to the location parameter μ and the scale parameter σ respectively. P_i and P_g are the individual extremum and the global extremum respectively.

The steps of the Particle Swarm Optimization algorithm are as follows:

Step 1: Initialize the particle count of the swarm, and set the values of learning factors, as well as the positions and velocities of particles.

Step 2: Calculate the individual fitness values according to the formula $F(t) = \Phi\left(\frac{ln\,\tau-\mu}{\sigma}\right)$, and set the individual optimal values (p_ibest) as the global optimal values (p_gbest).

Step 3: Adjust the particle velocity and position based on the formulations presented in Equations (9) and (10), and calculate the individual fitness values at the new positions.

Step 4: Take the minimum value between the current fitness of each particle and the fitness of its individual extremum p_ibest, assign this minimum value to p_ibest, and record the current parameter values.

Step 5: Compare the individual extremum p_ibest with the global optimal value p_gbest. If the individual fitness value is smaller, assign this value to p_gbest; otherwise, keep p_gbest unchanged.

Step 6: If the current iteration count is below the preset threshold, loop back to Step 3; otherwise, terminate the algorithm and output the optimal parameter values concurrently.

2.5. Fitting test and model optimization

Let the product lifetime data $t_1 \le t_2 \le \cdots \le t_n$ obtained from experiments follow a two-parameter lognormal distribution $LN(\mu,\sigma)$. Based on the optimally selected empirical distribution values $F_n(t)$, the fitting values $\hat{F}(t)$ of the lognormal distribution are derived using the relevant parameter estimation method.

2.5.1. Fitting test

The K-S test is a non-parametric statistical test method designed for continuous distributions. It is often used to verify whether a single sample conforms to a known distribution. The advantage of the K-S test is that it does not require data grouping, which enables more comprehensive utilization of data and has a wide scope of application. Its disadvantage is that when the potential distribution of the test sample is relatively clear, its testing performance is inferior to that of the corresponding distribution-specific tests.

The test statistic is expressed as follows:

$$D_n = \sup_{-\infty<t<+\infty} \left|F_n(t) - \hat{F}(t)\right| = max\{d_i\} \le D_{n,\alpha} \qquad (11)$$

Where, $D_{n,\alpha}$ is the critical value, which can be found in Reference [27].

For a given significance level α, the rejection region is expressed as follows: $r = \{D_n > D_{n,\alpha}\}$,$p = P\{D_n > D_{n,\alpha}\} = \alpha$.

At a significance level of α, if $D_n \le D_{n,\alpha}$, the null hypothesis is considered valid, meaning that the product life follows the assumed distribution; otherwise, the null hypothesis is rejected.

2.5.2. Model optimization

Assume that the product life data $t_1 \le t_2 \le \cdots \le t_l$ obtained from the test follows a two-parameter lognormal distribution denoted as $LN(\mu,\sigma)$. Calculate the empirical distribution function $F_n(t)$ based on the optimally selected empirical distribution model. Adopt the relevant parameter estimation method to derive the fitting value $\hat{F}_n(t)$ of the lognormal distribution model.

Calculate the root mean square error (RMSE) according to Equation (12).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}[F_n(t_i)-\hat{F}_n(t_i)]^2}{n}} \qquad (12)$$

The RMSE serves as the evaluation criterion for assessing the performance of parameter estimation methods. A higher RMSE value denotes inferior accuracy of the parameter estimation results.

3. Parameter estimation for the two-parameter lognormal distribution

3.1. Common parameter estimation methods

3.1.1. Maximum likelihood estimate

The maximum likelihood estimates of parameters μ and σ for the lognormal population are calculated according to Equation (13).

$$\begin{cases} \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} ln\, t_i \\ \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ln\, t_i - \hat{\mu})^2} \end{cases} \qquad (13)$$

3.1.2. Least Squares Estimate

The principle of parameter estimation using the LSE is to obtain parameter estimates by minimizing the squared residual sum. In other words, the parameter estimates of the linear model are obtained by minimizing the SSR. Least squares estimation falls into two categories that correspond to the x-direction and y-direction, respectively. As indicated in Reference [22], when the sample size $n < 14$, the deviation obtained by the y-direction parameter estimation is smaller than that obtained by the x-direction parameter estimation.

The two-parameter lognormal distribution is linearized to the form $y = A + Bx$ by setting $x = ln\, t$ and $y = \Phi^{-1}[F(t)]$. The Least Squares Estimate is adopted to ascrtain the model parameters, where $\hat{B} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$ and $\hat{A} = \bar{y} - \hat{B}\bar{x}$, with $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ and $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$.

The parameters of the lognormal population are calculated according to Equation (14).

$$\begin{cases} \hat{\mu} = -\frac{A}{B} \\ \hat{\sigma} = \frac{1}{B} \end{cases} \quad (14)$$

3.1.3. Total Least Squares Estimate

In view of the non-unique results when estimating reliability distribution parameters using the Ordinary Least Squares estimate, a parameter estimation method based on the Total Least Squares criterion is proposed [28].

The Total Least Squares Estimate aims to minimize the sum of squared vertical residuals min $min \frac{1}{1+a^2} \sum_{i=1}^{n} [y_i - (ax_i + b)]^2$ between a point (x_i,y_i)(i=1,2,⋯,n) and the fitted line y=ax+b, so as to obtain the total least squares parameter estimates aˆand bˆ of the linear model. With this method, the regression coefficients can be estimated.

$$\begin{cases} \hat{a} = -c \pm \sqrt{1 + c^2} \\ \hat{b} = \bar{y} - \hat{a}\bar{x} \end{cases} \quad (15)$$

Where, $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$, $c = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2 - \sum_{i=1}^{n}(y_i - \overline{y})^2}{2\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}$; the determination of "$\pm$" depends on the distribution of the sample points $(x_i, y_i)(i = 1,2, \cdots, n)$. If the sample covariance exceeds 0, the plus sign "+" is selected; if the sample covariance is less than 0, the minus sign "-" is adopted.

Since the fitting relationship between x and y on the plane is uniquely determined under the Total Least Squares criterion, the parameter estimates are also unique. When the reliability distribution function follows a lognormal distribution, the inverse normal transformation is performed on $F(t; \mu, \sigma) = \Phi\left\{\frac{ln\, t - \mu}{\sigma}\right\}$, and by setting $x = ln\, t$ and $y = \Phi^{-1}[F(t)]$, the linear model $y = ax + b$ is derived. The slope of the line corresponds to the logarithmic standard deviation $\hat{\sigma} = 1/\hat{a}$, and the intercept of the line corresponds to the logarithmic mean $\hat{\mu} = -\frac{\hat{b}}{\hat{a}}$. For a set of product lifetime values $\{t_i\}, i = 1,2, \dots, n$, the initial dataset $S = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ can be obtained through this linear transformation. From a practical perspective, there exists a positive correlation between the two variables x_i and y_i, and their covariance value is greater than 0. Therefore, in Equation (15), the parameters of the lognormal distribution satisfy this condition, so $\hat{a}_2 = -c + \sqrt{1 + c^2}$. The parameter estimates of the lognormal distribution are expressed as:

$$\begin{cases} \hat{\mu} = -\hat{b}/\hat{a} = -\hat{b}/\left(-c + \sqrt{1 + c^2}\right) \\ \hat{\sigma} = 1/\hat{a} = 1/\left(-c + \sqrt{1 + c^2}\right) \end{cases} \quad (16)$$

3.2. Parameter estimation based on LSSVM

The idea of regression using the Least Squares Support Vector Machine is to map a low-dimensional space to a high-dimensional space through a nonlinear function transformation, find a linearly separable hyperplane, and minimize the "distance" from the fitted points to the farthest sample points on the hyperplane. The difference between LSSVM regression and traditional Support Vector Machine regression lies in that LSSVM employs squared error terms instead of the epsilon-insensitive loss function used in SVM for the optimization problem. This reduces the complexity of the model optimization process and effectively shortens the model training time.

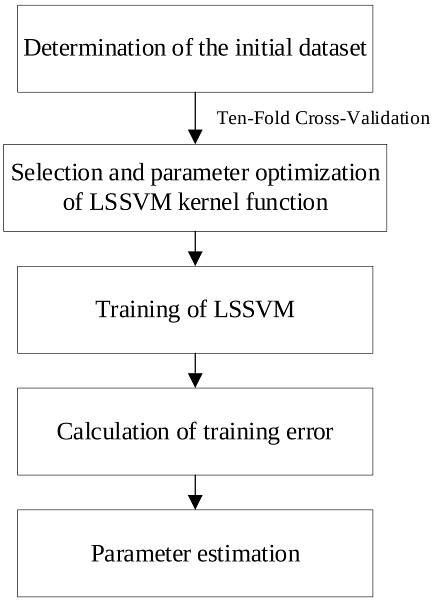


Figure 2. Flow chart of two-parameter estimation for lognormal distribution based on LSSVM.

The parameter estimation process under the Least Squares Support Vector Machine consists of the following steps: first, based on the collected lifetime data, the optimal lifetime distribution is selected according to the probability distribution scatter plot; then, the threshold parameters are estimated with reference to the preferred empirical distribution values, and the initial sample set of LSSVM is obtained through the inverse transformation of the empirical distribution. Next, parameter estimation is performed via LSSVM, and the parameters of LSSVM are optimized by adopting the cross-validation method, thus deriving the unknown parameters in the hypothetical distribution function. Finally, a hypothesis test is conducted on the estimated model. By selecting the empirical distribution function that passes the test and its corresponding parameter estimates, the lifetime distribution model can be established.

The process is illustrated in Figure 2.

3.2.1. Kernel function selection for LSSVM

The kernel functions applicable to the LSSVM encompass the linear kernel, polynomial kernel, Gaussian radial basis function, as well as the sigmoid kernel. As a kernel function featured by prominent locality, the Gaussian RBF is capable of projecting sample data into a higher-dimensional feature space. It stands as the most commonly utilized kernel among these options, and it delivers favorable performance regardless of whether the sample size is large or small. Additionally, in contrast to the polynomial kernel, it involves a smaller number of parameters. For these reasons, the Gaussian radial basis function is employed as the kernel function in the present study.

$$k(\boldsymbol{x},\boldsymbol{x}_i)=exp(-\|\boldsymbol{x}-\boldsymbol{x}_i\|^2/2\delta^2),\gamma>0 \quad (17)$$

Where, $\boldsymbol{x}_i\in\chi$, $\chi\in R^n$, R denotes the set of real numbers, and R^n denotes the n-dimensional real number space. That is, $\boldsymbol{x}_i$ is an n-dimensional vector, serving as the center of the function; δ is the width parameter of the function, which controls the radial range of action of the function.

3.2.2. Parameter optimization

The selection of kernel functions and their parameters is also a parameter optimization process. It is essential to use a large proportion of the dataset for model training; otherwise, the training process may fail, and a model with substantial bias will be generated in the end. Based on historical experience, the ten-fold cross-validation method is adopted herein. The ten-fold cross-validation method randomly divides the initial dataset of the Least Squares Support Vector Machine model into 10 subsets. Nine of these subsets serve as the training dataset for training the LSSVM model, with the remaining one assigned as the validation set. For each iteration of the validation process, the Mean Squared Error between the predicted results of the validation set and its corresponding actual target values is logged. After completing ten rounds of iteration, the average value of the recorded MSEs is defined as the Cross-Validation Error. The set of kernel function parameters corresponding to the minimum cross-validation error is selected as the optimal internal parameters of the LSSVM.

The steps are as follows:

(1) Randomly divide the initial dataset of the LSSVM model into 10 subsets;

(2) Use 9 of these subsets as the training set to train the LSSVM model, and take the 10th subset as the validation set;

(3) Record the MSE obtained from each prediction result.

$$MSE=\frac{1}{n}\sum_{i-1}^{n}(y_i-\hat{y}_i)^2 \quad (18)$$

Where, y_i denotes the actual target value of the validation set, and $\hat{y}_i$ denotes the estimated target value of the validation set.

(4) Repeat this process until every subset has served as the validation set;

(5) The average value of the 10 recorded MSEs is defined as the Cross-Validation Error, which is used as the criterion for evaluating model performance. The set of internal LSSVM parameters corresponding to the minimum Cross-Validation Error is selected as the optimal internal parameters.

3.2.3. Parameter estimation

The LS-SVM toolbox in MATLAB is used to perform parameter optimization for the LSSVM model, yielding the optimal regression line. Subsequently, combined with the graphical estimation method, parameter estimation is conducted for the underlying distribution.

4. Empirical analysis

4.1. Application examples

A total of 13 lifetime data points of a certain type of automotive battery were collected through a full-life test [8], and the lifetime data are presented in Table 1.

Table 1. Battery lifetime.

Sample number	Lifetime (h)	Sample number	Lifetime (h)
1	687	8	1240
2	1103	9	1251
3	1115	10	1723
4	1131	11	1884
5	1149	12	2006
6	1151	13	2625
7	1173		

4.1.1. Determination of the initial dataset

Assuming that the lifetime data in Table 1 follows a two-parameter lognormal distribution, a linearly transformed dataset $S=\{(x_1,y_1),(x_2,y_2),\dots,(x_n,y_n)\}$ can be obtained, take the Hellinger distance as an example, as presented in Table 2.

Table 2. Sample dataset and empirical distribution function values.

i	t_i (h)	y_i	x_i	$F_n(t_i)=\frac{(i-0.5)}{n}$
1	687	-1.7688	6.5323	0.0385
2	1103	-1.1984	7.0057	0.1154
3	1115	-0.8694	7.0166	0.1923
4	1131	-0.6151	7.0308	0.2692
5	1149	-0.3957	7.0466	0.3462
6	1151	-0.1940	7.0483	0.4231
7	1173	0.0000	7.0673	0.5000
8	1240	0.1940	7.1228	0.5769
9	1251	0.3957	7.1316	0.6538
10	1723	0.6151	7.4518	0.7308
11	1884	0.8694	7.5406	0.8077
12	2006	1.1984	7.6038	0.8846
13	2625	1.7688	7.8728	0.9615

4.1.2. Selection of empirical distribution function model

(1) Model selection based on the Shapiro-Wilk test

The optimal distribution model is selected based on the lifetime distribution scatter plot in Table 1, which is shown in Figure 3. Meanwhile, the p-value is greater than 0.05 via the Shapiro-Wilk test, indicating acceptance of the null hypothesis. It can be concluded from Figure 3 and the Shapiro-Wilk test that the lifetime of the automotive battery follows the lognormal distribution preferentially within the 95% confidence interval.

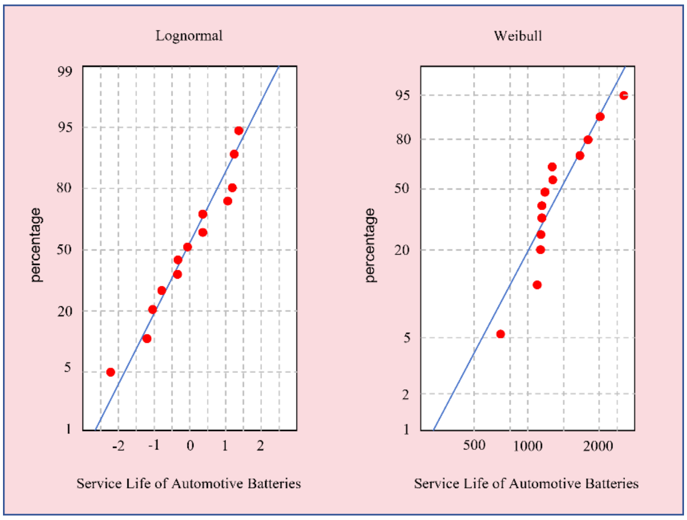


Figure 3. Scatter plot of automotive battery lifetime.

Therefore, it is assumed that the lifetime of the automotive battery follows a lognormal distribution, i.e., $T_i \sim LogNormal(\mu, \sigma^2), i = 1,2,\ldots,13$.

(2) Selection of the Empirical Distribution Function Model

For the data in Table 1, the linear correlation coefficients between $x = \ln t$ and $y = \Phi^{-1}(F(t))$ are calculated according to Equations (3), (4), (5) and (6), as detailed in Table 3.

Table 3. Linear correlation coefficients under three empirical distribution function estimation models.

Empirical Distribution Function $F_n(t_i)$	$\frac{i-0.5}{n}$	$\frac{i}{n+1}$	$\frac{i-0.3}{n+0.4}$
linear correlation coefficient ρ	0.943	0.935	0.939

It can be seen from Table 3 that the optimal empirical distribution function model is the Hellinger distance formula, namely $F_n(t_i) = \frac{i-0.5}{n}$.

4.1.3. Modeling of the two-parameter lognormal distribution

(1) Estimation of location and scale parameters

The estimation results of the location and scale parameters for the two-parameter lognormal distribution were obtained based on the criteria of minimum sum of squared cumulative errors, least squares support vector machine, maximum likelihood method, least squares method, and total least squares method, as shown in Table 4.

Table 4. Estimated values of location and scale parameters for five lognormal distribution.

Parameter estimation method	**Location parameter**	**Scale parameter**
Based on the minimum sum of squared cumulative errors	7.1661	0.3418
Least squares support vector machine	7.1902	0.3437
Maximum likelihood estimate	7.1901	0.3299
Least squares estimate	7.1902	0.3674
Total least squares estimate	7.1907	0.3307

(2) Model fitting test and optimization

The K-S test statistics and RMSE test values are shown in Table 5.

Table 5. K-S test statistic and RMSE Value.

Parameter estimation method	**K-S Test Statistic**	**RMSE Value**
Based on the minimum sum of squared cumulative errors	0.204143	0.105807
Least squares support vector machine	0.219730	0.108033
Maximum likelihood estimate	0.224103	0.108802
Least squares estimate	0.217103	0.107663
Total least squares estimate	0.224647	0.108872

When $\alpha = 0.10$, the K-S test was used to verify the fitting performance. The critical value $D_{n,\alpha} = 0.214$. As can be seen from Table 5, for this set of small-sample data, only the K-S test statistic derived from the principle of minimizing the sum of squared cumulative errors in parameter estimation was less than the critical value, thus passing the test.

It can be seen from the RMSE that the criterion of minimum sum of squared cumulative errors yields the smallest RMSE value and thus achieves the highest accuracy of parameter estimation.

4.1.4. Modeling of the three-parameter lognormal distribution

(1) Threshold Parameter Estimation

As can be seen from Table 1, the minimum lifetime is 687 hours, so $l = 686$. In engineering practice, it is sufficient for the lifetime values to be rounded to integers. A calculation accuracy of $\Delta = 1$ hour is adopted; thus, according to Formula (6), we obtain $k = -\frac{(ln\,3)\times 1 - ln(2\times 686)}{ln\,2} + 2 = 10$, meaning that 10 interpolation calculations are required.

The interval $[0,686]$ was divided into three equal subintervals, and the corresponding values were calculated as follows: $\rho_{00}(0) = 0.943017$, $\rho_{01}(686) = 0.730300$, $\rho_{011}(228.7) = 0.942994$, $\rho_{012}(457.4) = 0.934232$. By comparison, it was found that $\theta_{02} = 228.7$, and the new interval was determined as $[0,228.7]$. The value of $\rho_{021}(114.35) = 0.943391$ was then calculated, and the rest of the values were obtained by analogy. Finally, the maximum value was derived as $\rho = max(127.5) = 0.94339655$. Therefore, the estimated threshold parameter of the lognormal distribution was determined to be $\hat{\theta} = 127.5$ hours.

(2) Estimation of location and scale parameters

When the threshold value set at 127.5, the estimation results of the location and scale parameters were derived in accordance with the criterion of minimum cumulative sum of squared errors, least squares support vector machine, maximum likelihood method, least squares method, and total least squares method, as shown in Table 6.

(3) Model fitting test and optimization

The test statistics and RMS values are shown in Table 7.

When $\alpha = 0.10$, the K-S test was adopted to verify the fitting performance. By referring to the critical value table, the critical value was found to be $D_{n,\alpha} = 0.214$. As can be seen from Table 7, for this set of small-sample data, only the K-S test statistics derived from the principle of minimizing the cumulative sum of squared errors and the least squares method were both less than the critical threshold value, thus passing the test; in addition, the K-S test statistic in accordance with the minimum sum of squared cumulative errors was the smallest among them. It can be concluded from the RMSE that the minimum sum of squared cumulative errors yielded the lowest RMSE value, corresponding to the highest accuracy of parameter estimation. Therefore, the lifetime data population follows the three-parameter lognormal distribution, with the parameters $\hat{\theta} = 127.5$, $\hat{\mu} = 7.0601$ and $\hat{\sigma} = 0.3746$.

Table 6. Estimated values of location and scale parameters for five lognormal distribution.

Parameter estimation method	Location parameter	Scale parameter
Based on the minimum sum of squared cumulative errors	7.0601	0.3746
Least squares support vector machine	7.0827	0.3681
Maximum likelihood estimate	7.0827	0.3657
Least squares estimate	7.0827	0.4072
Total least squares estimate	7.0827	0.3676

Table 7. K-S test statistic and RMSE value.

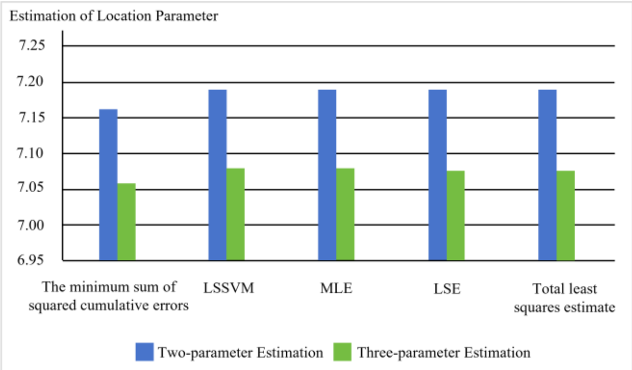
Parameter estimation method	K-S Test Statistic	RMSE Value
Based on the minimum sum of squared cumulative errors	0.202756	0.105080
Least squares support vector machine	0.216977	0.107053
Maximum likelihood estimate	0.217388	0.107171
Least squares estimate	0.210959	0.106576
Total least squares estimate	0.217063	0.107077

4.1.5. Parameter estimation comparison of lognormal distribution

(1) Comparison of Parameter Estimation Stability

From Tables 4 and 6, the comparison diagrams of the location parameters and scale parameters of the two-parameter and three-parameter models obtained by the five parameter estimation methods can be derived, as shown in Figure 4.

As can be seen from Figure 4, under the same parameter estimation method, the location parameter derived from the two-parameter model is larger while the scale parameter is smaller than those from the three-parameter model. For the two-parameter model, the location parameters obtained by the five estimation methods vary from 7.1661 to 7.1902, with a variation range of 0.0246. In contrast, for the three-parameter model, the location parameters obtained by the five estimation methods range from 7.0601 to 7.0827, corresponding to a variation range of 0.0226. This indicates that the location parameters derived from the five estimation methods have better stability under the three-parameter model. Similarly, the scale parameters obtained by the five estimation methods exhibit better stability under the two-parameter model.



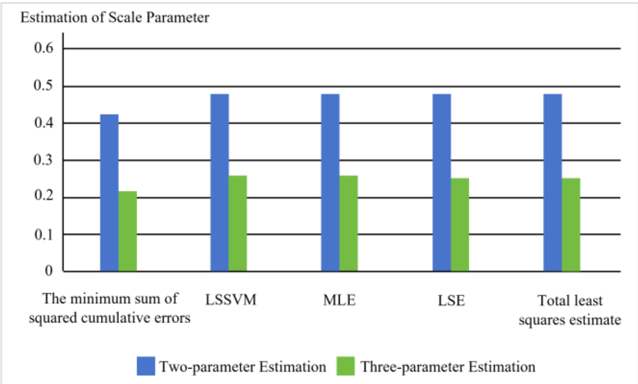
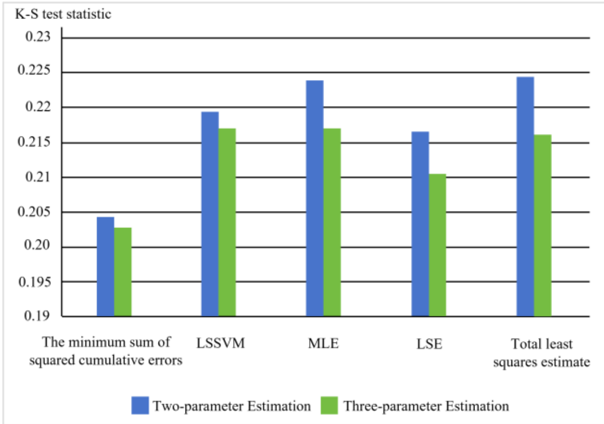


Figure 4. Comparison of two-parameter and three-parameter estimations for lognormal distribution.

(2) Comparison of K-S test statistics and RMSE values for parameter estimation

From Tables 5 and 7, the comparison diagrams of K-S test statistics and RMSE values for the two-parameter and three-parameter models obtained by the five parameter estimation methods can be derived, as shown in Figure 5.

As can be seen from Figure 5, for the same parameter estimation method, the K-S test statistics and RMSE values derived from the three-parameter model are smaller than those from the two-parameter model. Among the K-S test values obtained by the five estimation methods under the two-parameter framework, only one is less than the critical value and thus passes the test; in contrast, under the three-parameter framework, two out of the five K-S test values are below the critical value and pass the test. This indicates that the location parameters and scale parameters estimated by the five parameter estimation methods have smaller deviations under the three-parameter framework.



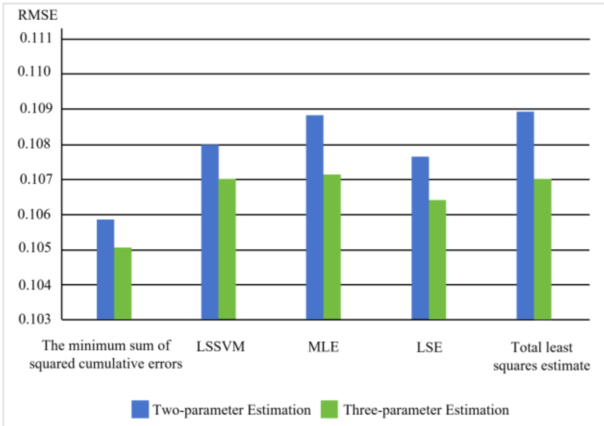


Figure 5. Comparison of K-S test statistics and RMSE values for two-parameter and three-parameter estimations of lognormal distribution.

4.2. Numerical Simulation

To verify the effectiveness of the proposed method, a numerical simulation approach is adopted. Generate random numbers by invoking numpy in Python based on the CDF value and three parameters, and the method presented in this paper is applied for parameter estimation and verification.

4.2.1. Parameter estimation for the two-parameter lognormal distribution

In this subsection, simulated data for the two-parameter lognormal distribution are generated with $\mu_0 = 7.1900$ and $\sigma_0 = 0.1088$, as shown in Table 8. The parameter estimates and test values derived using the method presented in this paper are listed in Table 9.

Table 8. Simulated data columns of the two-parameter lognormal distribution.

Serial number i	t_i	Serial number i	t_i
1	1006.1592	8	1258.6121
2	1121.5250	9	1276.6445
3	1147.6716	10	1326.0750
4	1213.7325	11	1363.5649
5	1224.9972	12	1447.9812
6	1226.5929	13	1589.5803
7	1245.4011		

Table 9. Parameter estimation and test values of the two-parameter lognormal distribution for simulated data.

Parameter estimation method	Location parameter	Scale parameter	K-S Test Statistic	RMSE Value
Based on the minimum sum of squared cumulative errors	7.132837	0.091705	0.096869	0.046326
Least squares support vector machine	7.136918	0.113655	0.108282	0.055445
Maximum likelihood estimate	7.136918	0.114833	0.1095	0.05617
Least squares estimate	7.136918	0.113751	0.108383	0.055503
Total least squares estimate	7.136918	0.113806	0.10844	0.055536

It can be seen from Table 9 that five methods are applied for parameter estimation of this dataset, all of which pass the tests. Based on the RMSE results, the parameter estimation with the minimum sum of squared cumulative errors yields the optimal performance. The relative errors calculated from the parameter estimates with the minimum sum of squared cumulative errors and the true values are presented in Table 10.

Table 10. Relative errors of two-parameter estimation for lognormal distribution of simulated data based on the minimum sum of squared cumulative errors.

Parameter	True Value	Estimated Value	Relative Error
Location parameter	7.1900	7.1328	0.796%
Scale parameter	0.1088	0.0917	15.717%

It can be seen from Table 10 that the parameter estimation for this dataset conducted by minimizing the sum of squared cumulative errors yields relatively stable values.

4.2.2. Modeling of the Three-Parameter Lognormal Distribution

In this subsection, the simulated data of the three-parameter lognormal distribution generated with $\theta_0 = 127.5$, $\mu_0 = 7.0601$ and $\sigma_0 = 0.3746$ are presented in Table 11. For the generated data, the threshold parameter is solved as P=129 by the method presented in this paper, and the other parameter estimates as well as the test results are shown in Table 12.

Table 11. Simulated data columns of the three-parameter lognormal distribution.

Serial number i	t_i	Serial number i	t_i
1	727.8412	8	1379.8574
2	870.8648	9	1478.1466
3	968.3505	10	1593.8504
4	1052.3838	11	1740.3953
5	1131.6145	12	1951.9120
6	1210.4207	13	2386.5550
7	1292.0616		

Table 12. Three-parameter estimation and test of lognormal distribution (threshold parameter = 129).

Parameter estimation method	Location parameter	Scale parameter	K-S Test Statistic	RMSE Value
Based on the minimum sum of squared cumulative errors	7.058780	0.375096	0.000068	0.000037
Least squares support vector machine	7.058727	0.374791	0.000217	0.000154
Maximum likelihood estimate	7.058727	0.371761	0.002134	0.001571
Least squares estimate	7.058727	0.375109	0.000090	0.000057
Total least squares estimate	7.058727	0.375109	0.000090	0.000057

It can be seen from Table 12 that five methods are applied for parameter estimation of this dataset, all of which pass the tests. Based on the RMSE results, the parameter estimation with the minimum sum of squared cumulative errors achieves the optimal performance. The relative errors calculated from the parameter estimates with the minimum sum of squared cumulative errors and the true values are presented in Table 13.

Table 13. Relative errors of three-parameter estimation for the lognormal distribution of simulated data based on the minimum sum of squared cumulative errors.

Parameter	True Value	Estimated Value	Relative Error
Location parameter	7.0601	7.05878	0.02%
Scale parameter	0.3746	0.375096	0.13%
Threshold Parameter	127.5	129	1.176%

It can be seen from Table 13 that the parameter estimates obtained by the method of minimizing the sum of squared cumulative errors proposed in this paper are relatively stable.

It can be concluded from Tables 10 and 13 that the parameter estimation based on the minimum sum of squared cumulative errors yields reasonable and stable modeling results under different raw data conditions. Therefore, this subsection verifies the effectiveness and stability of the parameter estimation method proposed in this paper through numerical simulation.

5. Conclusion

(1) The proposed empirical distribution model selection in this paper, which takes the maximization of the linear correlation coefficient as the objective, is applicable to any linearizable model. It is designed to compensate for the reliability modeling error caused by the model itself in the process of converting ranks into probabilities when experimental data are processed via the rank method.

(2) The threshold parameter estimation method presented in this paper, which adopts the interpolation method with the maximization of the linear correlation coefficient as the objective, gradually approaches the optimal value through linear interpolation. This method reduces the model deviation arising from the assumption of the two-parameter lognormal distribution and improves the fitting accuracy of the lognormal distribution model.

(3) Aiming at the characteristics of small sample size and high reliability of automotive battery life data, this paper proposes a lognormal distribution parameter estimation method that takes the minimization of the sum of squared cumulative errors as the objective. By accounting for the errors between the sample-fitted distribution function and the empirical distribution values, this method can make up for the deficiency of the classical least square method and maximum likelihood estimation in neglecting global errors. The particle swarm optimization algorithm is introduced for parameter estimation, which avoids tedious mathematical derivation processes, features a clear and understandable principle, and is convenient for popularization and application. Taking the root mean square error as the model evaluation index, this paper compares the proposed parameter estimation method with other methods, verifying the effectiveness of the proposed method.

(4) In this paper, the research method is applied in combination with simulated data. From the relative errors of the estimated values, it can be concluded that the lognormal distribution parameter estimation method based on the minimum sum of squared cumulative errors yields stable and valid parameter estimates. The method proposed in this paper resolves the contradiction between high modeling accuracy and the characteristics of small samples with insufficient information, and provides guidance for the construction of similar models.

Acknowledgment

This research was supported by the Key Science and Technology Special Project of Jilin Province: "Research on Key Technologies of High-voltage Electrical Systems for Electric Vehicles" (Project Number: 20220301015GX).

References

1. Luo X, Wei C, Wang Y. Comparison of reliability parameters estimation methods for exponential distribution under small samples. Statistics & Decision, 2018, 34(01): 14–17. https://doi.org/10.13546/j.cnki.tjyjc.2018.01.003.
2. Kong X, Liu J, Wang J, Tang S. Comparative study on estimation methods of aircraft IDG reliability parameters. Mechanical Science and Technology for Aerospace Engineering, 2022, 41(06): 977–984. https://doi.org/10.13433/j.cnki.1003-8728.20200397.
3. Wang Z, Yang J. Bootstrap method of parameter estimation in reliability model. Journal of Harbin Institute of Technology, 2010, 42(05): 820–823. https://doi.org/10.11918/j.issn.0367-6234.2010.05.033.
4. Wei S, Li N. Bootstrap estimation of Weibull distribution parameters based on small sample and censored cases. Statistics & Decision, 2019, 35(23): 34–37. https://doi.org/10.13546/j.cnki.tjyjc.2019.23.006.
5. Liu Y, Wang Z, Li X, Zhao X, Liao J. Reliability evaluation of electronic residual current operated circuit breakers based on improved Bootstrap-Bayes. Transactions of China Electrotechnical Society, 2022, 37(16): 4250–4258. https://doi.org/10.19595/j.cnki.1000-6753.tces.210229.
6. Huang L, Wu L. Maximum likelihood estimation based on lognormal distribution combined logarithmic mean and logarithmic variance models. Statistics & Decision, 2011(21): 8–11. https://doi.org/10.13546/j.cnki.tjyjc.2011.21.001.

7. Gu B, Xu X, Wang R. Approximate maximum likelihood estimation of two-parameter lognormal distribution and Weibull distribution. Statistics & Decision, 2019, 35(03): 34–39. https://doi.org/10.13546/j.cnki.tjyjc.2019.03.008.
8. Tao H, Liu W, Fu J. Lognormal distribution reliability model and its application. Statistics & Decision, 2019, 35(03): 89–92. https://doi.org/10.13546/j.cnki.tjyjc.2019.03.020.
9. Li H. Study on least squares support vector machine theory and its application. Lanzhou University, 2022.
10. Wang J, Tang Y, Lu J, Zhao W. Estimation method of reliability characteristic parameter of electrical apparatus products for a Few of failure Data. Systems Engineering and Electronics, 2000, 22(10): 94–96. https://doi.org/10.3321/j.issn:1000-6753.2000.04.007.
11. Wang C, Zheng Q, Li Q, Lin S. Inspection and elimination of heteroscedasticity based on SAS. Journal of Changchun University of Technology, 2016, 37(3): 209–216. https://doi.org/10.15923/j.cnki.cn22-1382/t.2016.3.01.
12. Yu X. Research on Weighted Least Squares with Multiplicative Conditional Variance. Chengdu: University of Electronic Science and Technology of China, 2018.
13. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. Machine learning, 2002, 46(1-3): 389–422. https://doi.org/10.1023/A:1012487302797.
14. Zheng, R. Structured replacement policies for a system subject to random mission types. Naval Research Logistics, 2024, Vol. 71(7): 1. https://doi.org/10.1002/nav.22201.
15. Zheng, R. Optimal Replacement for a Partially Observable System Subject To Dependent Failure Modes. Naval Research Logistics, 2025. https://doi.org/10.1002/nav.70024.
16. Suykens J, Vandewalle J. Least squares support vector machine classifiers. Neural Processing Letters, 1999, 9(3): 293–300. https://doi.org/10.1023/A:1018628609742.
17. Cherkassky V, Ma Yunqian. Practical selection of SVM parameters and noise estimation for SVM regression. Neural Networks, 2004, 17(1): 113–126. https://doi.org/10.1016/S0893-6080(03)00169-2.
18. Peng Y, Cheng J, Liu Y, Li X, Peng Z. An adaptive data-driven method for accurate prediction of remaining useful life of rolling bearings. Frontiers of Mechanical Engineering, 2018, 13(2). 301–310. https://doi.org/10.1007/s11465-017-0449-7.
19. Zhang X, Zhao Y. Weibull reliability analysis in small samples based on LSSVM. China Mechanical Engineering, 2012, 23(16): 1967–1971. https://doi.org/10.3969/j.issn.1004-132X.2012.16.017.
20. He X, Jiang X, Zhang P, Gao X, Li Z. SVM based parameter estimation of relay protection reliability with small samples. Power System Technology, 2015, 39(05): 1432–1437. https://doi.org/10.13335/j.1000-3673.pst.2015.05.039.
21. Aghaei A, Nejad G , Yousefi S , Parand K. A new Chebyshev operational matrix formulation of least-squares support vector regression for solving fractional integro-differential equations. Journal of Computational and Applied Mathematics, 2026, 474: 116816–116816. https://doi.org/10.1016/j.cam.2025.116816.
22. Wang X, Shen G, Zhang Y, Chen B, Zheng S. Comparison of Parameter Estimation Methods for Reliability Model. Journal of South China University of Technology (Natural Science Edition), 2011, 39(06): 47–52. https://doi.org/10.3969/j.issn.1000-565X.2011.06.009.
23. Jiang R, Zuo M. Reliability Models and Applications. Beijing: China Machine Press, 1999.
24. Yang Z, Liu R. Improved Method for Parameter Estimation of Three-parameter Weibull Distribution. Chinese Journal of Engineering Mathematics, 2004, (02): 281–284+267. https://doi.org/10.3969/j.issn.1005-3085.2004.02.025.
25. Zhuo J. Applications of MATLAB in Mathematical Modeling. 2nd ed. Beijing: Beihang University Press, 2014.
26. Parvasideh P, Rezghi M. A novel dictionary learning method based on total least squares approach with application in high dimensional biological data. Advances in Data Analysis and Classification, 2021, 15(3): 575–597. https://doi.org/10.1007/s11634-020-00417-4.
27. Wang C, Wang J. Mechanical Reliability Engineering. Beijing: Metallurgical Industry Press, 1992.
28. Wang J, Sun Z. Reliability Evaluation of Relay Protection Devices Based on Weibull Distribution and Iterative Maximum Likelihood Estimation Method. Electrical Appliances and Energy Efficiency Management Technology, 2022, (11): 63–68+73. https://doi.org/10.16628/j.cnki.2095-8188.2022.11.011.