

Article

On Modeling Concrete Compressive Strength Data Using Laplace Birnbaum-Saunders Distribution Assuming Contaminated Information

Farouq Mohammad A. Alam ^{1,*} and Mazen Nassar ^{1,2}

¹ Department of Statistics, Faculty of Science, King Abdulaziz University, P.O. Box 80200, Jeddah 201589, Saudi Arabia; mezo10011@gmail.com

² Department of Statistics, Faculty of Commerce, Zagazig University, Zagazig 44519, Egypt

* Correspondence: fmalam@kau.edu.sa

Abstract: Compressive strength is a well-known measurement to evaluate the endurance of a given concrete mixture to stress factors, such as compressive loads. A suggested approach to assess compressive strength of concrete is to assume that it follows a probability model from which its reliability is calculated. In reliability analysis, a probability distribution's reliability function is used to calculate the probability of a specimen surviving to a certain threshold without damage. To approximate the reliability of a subject of interest, one must estimate the corresponding parameters of the probability model. Researchers typically formulate an optimization problem, which is often nonlinear, based on the maximum likelihood theory to obtain estimates for the targeted parameters and then estimate the reliability. Nevertheless, there are additional nonlinear optimization problems in practice from which different estimators for the model parameters are obtained once they are solved numerically. Under normal circumstances, these estimators may perform similarly. However, some might become more robust under irregular situations, such as in the case of data contamination. In this paper, nine frequentist estimators are derived for the parameters of the Laplace Birnbaum-Saunders distribution and then applied to a simulated data set and a real data set. Afterwards, they are compared numerically via Monte Carlo comparative simulation study. The resulting estimates for the reliability based on these estimators are also assessed in the latter study.



Citation: Alam, F.M.A.; Nassar, M. On Modeling Concrete Compressive Strength Data Using Laplace Birnbaum-Saunders Distribution Assuming Contaminated Information. *Crystals* **2021**, *11*, 830. <https://doi.org/10.3390/cryst11070830>

Academic Editors: Yang Yu, Weiqiang Wang, Rafael Shehu and Beatrice Pomaro

Received: 30 June 2021

Accepted: 15 July 2021

Published: 17 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In reality, concrete is the most widely used construction material in the world. Concrete compressive strength is a measure used in determining the amount of resistance a structural element can offer to deformation. Compressive strength is a widely used measure to access the performance of a given concrete mixture. Considering this approach of concrete is vital due to it is the main measure deciding how well concrete can withstand loads that influence its measure. It precisely tells us whether a particular mix is suitable to meet the necessities of a particular venture. Concrete can astoundingly stand up to compressive loading. This is often why it is reasonable for constructing arches, columns, dams, foundations, and tunnel linings among other structures.

Researchers from different fields of science may attempt to describe phenomena of interest, such as concrete compressive strength, using either mathematical or probabilistic models. For instance, scientists describe the life of an object using a probabilistic model called a lifetime probability distribution. Such practice is common in the scientific community, especially in many science fields that involve reliability and reliability analyses. For example, in material science, the two-parameter Birnbaum-Saunders (BS) lifetime distribution can be used by analysts to model fatigue of materials due to periodic cyclic loading [1,2]. This non-negative lifetime model has unimodal skewed probability density and hazard rate curves.

Unimodal hazard rate curves are common in practice; see, for example, Langlands et al. [3]. A non-negative continuous random variable T is said to follow the BS distribution if the corresponding cumulative distribution function (CDF) is given by:

$$F(t; \alpha, \beta) = \Phi\left(\frac{1}{\alpha}\left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right), \quad t > 0, \alpha, \beta > 0, \quad (1)$$

where α is the shape parameter, β is the scale parameter, and $\Phi(\cdot)$ is the CDF of the standard normal distribution. Desmond [4] provided a more general derivation for the BS distribution assuming a biological model which reinforced the physical rationalization for the use of the BS distribution by relaxing the original presumptions of [1,2]. The BS distribution has desirable aspects and a close relationship to the normal distribution. Consequently, at least a couple of hundred papers and a single research monograph have already appeared describing all properties and developments of this distribution; see, for example, the comprehensive review by Balakrishnan and Kundu [5] in this connection. Examples of recent applications of the BS distribution are such as Bourguignon et al. [6], Hassani et al. [7], and Kannan et al. [8], among others. The BS distribution belongs to a generalized family of distributions called the generalized BS distribution [9]. The generalization is obtained by replacing the Gaussian kernel in Equation (1) with kernels of symmetrical distributions such as the Laplace and logistic distributions. Sampling plans from truncated life tests assuming the generalized BS distribution were developed in [10], while the generalized BS distribution was used to analyze air pollutant concentration in [11]. For further details in this connection, see [12]. The Laplace BS (LBS) distribution is a BS distribution based on Laplace kernel. Its properties and some associated estimation methods were studied by Zhu and Balakrishnan [13]. This paper expands upon their work by discussing additional methods to estimate the parameter of the LBS distribution assuming data contamination. A positive random variable is said to follow the LBS distribution if the corresponding CDF, reliability function, and the probability density functions are given by:

$$F(t; \alpha, \beta) = 0.5 + 0.5 \operatorname{sgn}(t - \beta) \left[1 - \exp\left(-\operatorname{sgn}(t - \beta) \frac{A(t; \beta)}{\alpha}\right) \right], \quad (2)$$

$$S(t; \alpha, \beta) = 0.5 - 0.5 \operatorname{sgn}(t - \beta) \left[1 - \exp\left(-\operatorname{sgn}(t - \beta) \frac{A(t; \beta)}{\alpha}\right) \right], \quad (3)$$

and

$$f(t; \alpha, \beta) = \frac{a(t; \beta)}{4\alpha\sqrt{\beta t}} \exp\left(-\operatorname{sgn}(t - \beta) \frac{A(t; \beta)}{\alpha}\right), \quad x, \alpha, \beta > 0, \quad (4)$$

respectively, such that α is a positive shape parameter, while β is a positive scale parameter, where:

$$A(t; \beta) = \frac{t - \beta}{\sqrt{\beta t}} \quad (5)$$

and

$$a(t; \beta) = \frac{t + \beta}{\sqrt{\beta t}}. \quad (6)$$

The LBS distribution and other lifetime models are characterized by essential statistical properties such as hazard (failure) rates and reliability. The latter concept represents the probability of a specimen continuing to exist for a certain amount of time without loss. Approximating the aspects of the life of objects of interest (e.g., reliability) is associated with model parameters estimation. Estimating the probability distribution parameters has been of great interest to scientists and has received much attention in the statistical literature. In practice, one can obtain various estimators for the model parameters. Therefore, several researchers conducted comparative Monte Carlo simulation studies to numerically assess the estimators from different statistical and computation perspectives; see, for example,

Gupta and Kundu [14], Alkasasbeh and Raqab [15], Mazucheli et al. [16], do Espirito Santo and Mazucheli [17], and Balakrishnan and Alam [18], among other research contributions.

Due to the advancement of human civilization, researchers currently deal with various types and large amounts of data originated from the targeted phenomena. In practice, there is no guarantee that part of the data may be missing or contaminated with unusual information values called outliers or extremes. In both cases, estimation efficiency is negatively impacted since the quality of data became questionable. Under data contamination, the maximum likelihood method is not robust since the deviations caused by existing outliers can negatively affect the likelihood. Data contamination motivated many researchers to propose alternative estimators to those obtained by the maximum likelihood theory for various distributions; see, for example, Lawson et al. [19], Boudt et al. [20], Agostinelli et al. [21], Wang et al. [22], among other papers.

The aim of this paper is twofold. Firstly, the LBS distribution is fitted to a real concrete compressive strength data using nine estimation methods. Secondly, the performances of the resulting estimators from the considered estimation methods; namely, modified moments estimators (MMEs), maximum likelihood estimators (MLEs), the maximum product of spacings estimators (MPSEs), the least-squares estimators (LSEs), the weighted least-squares estimators (WLSEs), the percentile estimators (PCEs), the Cramér-von Mises estimators (CVMEs), the Anderson-Darling estimators (ADEs), and the right-tailed Anderson-Darling estimators (RADEs) are investigated using numerical applications and Monte Carlo simulations. The remaining parts of this article are organized as follows. Section 2 discusses the nine estimation methods of interest. Section 3 illustrates the practical application of the discussed estimators using a simulated data set and a real data set. Section 4 reports the outcomes of extensive Monte Carlo simulation study to compare the performance of each estimator under different settings. Finally, remarks and future research are used to conclude this paper in Section 5.

2. Reliability and Model Parameters Estimation

In this section, eight nonlinear optimization problems are formulated to obtain eight estimators for the α and β parameters mentioned in the introductory section. The ninth estimators are the MMEs which have closed-form expression [13]. Before establishing the targeted optimization problems, one must discuss some computational considerations to solve them.

2.1. Computational Considerations for the Optimization Process

The first important computational consideration is finding suitable starting values to solve the optimization problems. One of the important and practical aspects of the LBS distribution is that the scale parameter β is actually the median of the population. Hence, a reasonable starting value for this parameter is the sample median, i.e., $\hat{\beta}^{(0)} = \text{median}(t_1, \dots, t_n)$. Regarding the starting value for the shape parameter α , one can acquire such value by using the relationship between the LBS distribution and the standard Laplace distribution. In fact, if T follows the LBS distribution with model parameters α and β , then:

$$\alpha Z = \sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}}$$

follows the Laplace distribution with the location parameter equals to 0, and the scale parameter equals to α . Hence, a starting value for α is obtained by determining the sample Z values; say, z_1, \dots, z_n , based on the observed data t_1, \dots, t_n and the corresponding sample median $\hat{\beta}^{(0)}$, and then finding the mean absolute deviation (MAD) of z_1, \dots, z_n as a starting value for α . That is,

$$\hat{\alpha}^{(0)} = \frac{1}{n} \sum_{i=1}^n |z_i - \text{median}(z_1, \dots, z_n)|,$$

such that

$$z_i = \sqrt{\frac{t_i}{\hat{\beta}^{(0)}}} - \sqrt{\frac{\hat{\beta}^{(0)}}{t_i}}.$$

The second computational consideration is the choice of optimization algorithm from which one can obtain the estimators for the LBS distribution. Zhu and Balakrishnan [13] showed that the probability density function of the LBS distribution Equation (4) is continuous, but it is not differentiable at β . Consequently, estimators like the MLEs require an optimization method free of derivatives to be obtained, such as the Nelder-Mead algorithm [23]. Note that the latter algorithm is utilized to acquire the remaining estimators to avoid any potential bias in the competitive computations.

2.2. Maximum Likelihood Estimation

Zhu and Balakrishnan [13] have obtained MLEs for α and β and prove their existence and uniqueness. One approach to obtain such estimators is by solving the following maximization algorithm:

$$\begin{aligned} & \text{maximize } -n \log \alpha + \sum_{i=1}^n \log a(t_i, \beta) - \frac{1}{\alpha} \sum_{i=1}^n |A(t_i, \beta)| \\ & \text{s.t. } \alpha, \beta \in \mathbb{R}_+^2, \end{aligned}$$

such that t_1, \dots, t_n are the observed random sample.

2.3. Least-Squares-Based Estimations

Given an observed random sample t_1, \dots, t_n from the LBS distribution with parameters α and β . Suppose that $t_{1:n} < \dots < t_{n:n}$ are the corresponding observed sample order statistics, and consider the following minimization problem:

$$\begin{aligned} & \text{minimize } \sum_{i=1}^n w_i \left(0.5 + 0.5 \operatorname{sgn}(t_{i:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{i:n} - \beta) \frac{A(t_{i:n}, \beta)}{\alpha}} \right] - \frac{i}{n+1} \right)^2 \\ & \text{s.t. } \alpha, \beta \in \mathbb{R}_+^2, \end{aligned}$$

Hence, the solutions of the above optimization problem are the LSEs for α and β given that $w_1 = \dots = w_n = 1$. However, if $w_i = [(n+1)^2(n+1)]/[i(n-i+1)]$, then the solutions of the above minimization problem are the WLSEs of α and β [24].

2.4. Percentile Estimation

PCEs of α and β are acquired by fitting a linear model to the theoretical percentiles and the sample percentiles [25,26]. This method requires closed-form cumulative distribution and quantile functions. In the case of the LBS distribution, the cumulative distribution function was given by Equation (2), while the quantile function is defined as:

$$Q(u; \alpha, \beta) = \frac{\beta}{4} \left[\alpha Q_L(u) + \sqrt{4 + \{\alpha Q_L(u)\}^2} \right]^2, \quad (7)$$

such that $0 < u < 1$, $\alpha, \beta > 0$, and $Q_L(u)$ is the quantile function of the standard Laplace distribution, i.e., $Q_L(u) = -\operatorname{sgn}(u - 0.5) \log(1 - 2|u - 0.5|)$. The PCEs of α and β are the solutions for the following minimization problem:

$$\begin{aligned} & \text{minimize } \frac{1}{n} \sum_{i=1}^n \left(t_{i:n} - \frac{\beta}{4} \left[\alpha Q_L(p_{i:n}) + \sqrt{4 + \{\alpha Q_L(p_{i:n})\}^2} \right]^2 \right)^2 \\ & \text{s.t. } \alpha, \beta \in \mathbb{R}_+^2, \end{aligned}$$

where $t_{1:n}, \dots, t_{n:n}$ are the observed sample order statistics, and $p_{i:n} = i/(n+1)$.

2.5. Maximum Product of Spacing Estimation

Another estimators of α and β that depends on solving a maximization problem are the MPSEs. Recent research indicates that such estimates compete with MLEs in terms of estimation efficiency and asymptotic properties [27–29]. Given the observed sample order statistics $t_{1:n} < \dots < t_{n:n}$, then the MPSEs for the model parameters are determined numerically by solving the following maximization problem:

$$\begin{aligned} & \text{maximize } \frac{1}{n+1} \sum_{i=1}^{n+1} \log \Delta_i \\ & \text{s.t. } \alpha, \beta \in \mathbb{R}_+^2, \end{aligned}$$

such that

$$\Delta_i = \begin{cases} 0.5 + 0.5 \operatorname{sgn}(t_{1:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{1:n} - \beta) \frac{A(t_{1:n}; \beta)}{\alpha}} \right] & \text{if } i = 1 \\ 0.5 \left\{ \operatorname{sgn}(t_{i:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{i:n} - \beta) \frac{A(t_{i:n}; \beta)}{\alpha}} \right] - \operatorname{sgn}(t_{i-1:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{i-1:n} - \beta) \frac{A(t_{i-1:n}; \beta)}{\alpha}} \right] \right\} & \text{if } 1 < i \leq n \\ 0.5 - 0.5 \operatorname{sgn}(t_{n:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{n:n} - \beta) \frac{A(t_{n:n}; \beta)}{\alpha}} \right] & \text{if } i = n+1 \end{cases}.$$

2.6. Goodness-of-Fit Estimations

The remaining three estimators are based on the idea of minimizing goodness-of-fit statistics, i.e., minimizing the difference between the estimated cumulative distribution function and an empirical counterpart. Examples of such statistics are the Cramér-von Mises, the Anderson-Darling, and the right-tailed Anderson-Darling statistics. The CVMEs of α and β are obtained by evaluating the following minimization problem:

$$\begin{aligned} & \text{minimize } \frac{1}{12n} + \sum_{i=1}^n \left(0.5 + 0.5 \operatorname{sgn}(t_{i:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{i:n} - \beta) \frac{A(t_{i:n}; \beta)}{\alpha}} \right] - \frac{2i-1}{2n} \right)^2 \\ & \text{s.t. } \alpha, \beta \in \mathbb{R}_+^2, \end{aligned}$$

On the other hand, ADEs and RADEs of α and β are acquired as solutions for the following minimization problems:

$$\begin{aligned} & \text{minimize } -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \log \left(0.5 + 0.5 \operatorname{sgn}(t_{i:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{i:n} - \beta) \frac{A(t_{i:n}; \beta)}{\alpha}} \right] \right) \\ & \quad - \frac{1}{n} \sum_{i=1}^n (2i-1) \log \left(0.5 - 0.5 \operatorname{sgn}(t_{n-i+1:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{n-i+1:n} - \beta) \frac{A(t_{n-i+1:n}; \beta)}{\alpha}} \right] \right) \\ & \text{s.t. } \alpha, \beta \in \mathbb{R}_+^2, \end{aligned}$$

and

$$\begin{aligned} & \text{minimize } \frac{n}{2} - 2 \sum_{i=1}^n \left(0.5 + 0.5 \operatorname{sgn}(t_{i:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{i:n} - \beta) \frac{A(t_{i:n}; \beta)}{\alpha}} \right] \right) \\ & \quad - \frac{1}{n} \sum_{i=1}^n (2i-1) \log \left(0.5 - 0.5 \operatorname{sgn}(t_{n-i+1:n} - \beta) \left[1 - e^{-\operatorname{sgn}(t_{n-i+1:n} - \beta) \frac{A(t_{n-i+1:n}; \beta)}{\alpha}} \right] \right) \\ & \text{s.t. } \alpha, \beta \in \mathbb{R}_+^2, \end{aligned}$$

respectively.

3. Numerical Applications

In this section, a simulated data is analyzed; afterwards, a real data is analyzed for the sake of illustration.

3.1. Simulated Data Analysis

Suppose 15 random observations are generated from the LBS distribution with $\alpha = \beta = 1$ as shown in Table 1.

Table 1. Simulated data from LBS distribution with $\alpha = \beta = 1$.

0.5790902	2.3047977	0.8182292	3.8502755	6.3717917
0.1314455	1.0595503	4.1175531	1.1146064	0.9132579
7.8732325	0.9067038	1.5453950	1.1697677	0.2345327

The simulated data in Table 1 are obtained according to the following algorithm:

1. Generate a random sample U_1, \dots, U_n from the standard uniform distribution (i.e., $U_i \sim \text{Uniform}(0, 1), \forall i$)
2. For $i = 1, \dots, n$, obtain the desired simulated random sample X_1, \dots, X_n from the LBS distribution with model parameters α and β by using Equation (7), i.e.

$$X_i = \frac{\beta}{4} \left[\alpha Q_L(U_i) + \sqrt{4 + \{\alpha Q_L(U_i)\}^2} \right]^2,$$

such that $Q_L(U_i) = -\text{sgn}(U_i - 0.5) \log(1 - 2|U_i - 0.5|)$.

Before obtaining the estimates for α and β , one must check their existence and uniqueness. Mathematically proving these requirements is beyond the scope of this study; nevertheless, one may prove them using graphical devices. Using extensive Monte Carlo simulations, a three-dimensional (3D) profile plot for each objective function in the preceding section is established, as shown in Figure 1. The 3D charts clearly indicate that there are areas in which global extrema exist and are expected to be unique for each objective function. The MMEs are obtained first for α and β as shown in [13], and then the remaining estimators are obtained successively as shown in Table 2. The outcomes of the latter table are obtained assuming no contamination in the data, and assuming contamination in the upper 20% order statistics. On the other hand, based on the obtained estimates under both assumptions, Table 3 provides the actual reliability probabilities calculated from the complement of Equation (2) vs. their approximated counterparts using different estimators. Both true and approximated reliability probabilities are evaluated at the sample minimum, the sample three quartiles (Q_1, Q_2, Q_3), and the sample maximum. From the previous tables, one can easily observe that the PCEs for α and β provided the farthest approximations for the model parameters and the reliability probabilities. The performances of the remaining estimates are further assessed later.

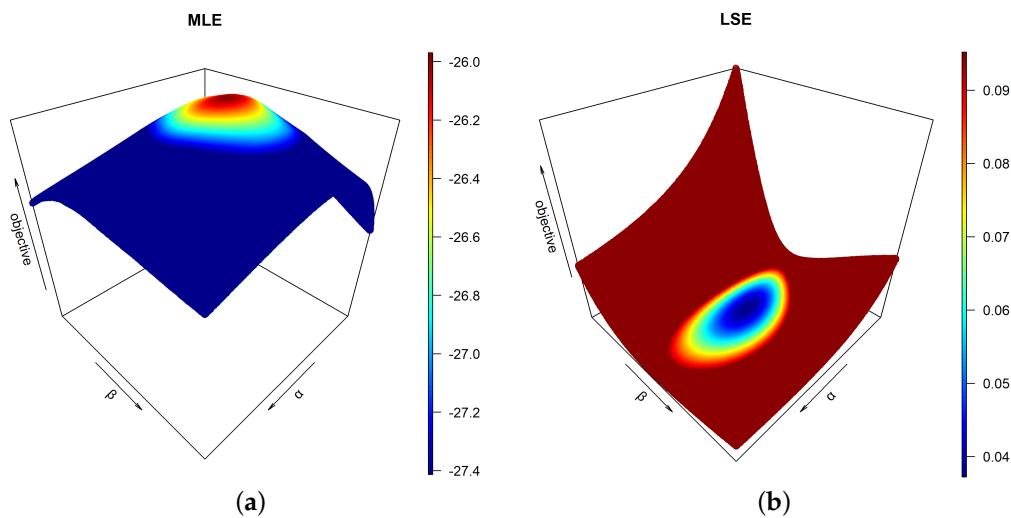
Table 2. Estimators of α and β based on data in Table 1.

Method	No Data Contamination		Upper 20% Data Contamination	
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
MME	0.882537	1.236373	1.471021	2.242812
MLE	0.921295	1.114606	1.533757	1.169768
LSE	1.148499	1.265467	1.306802	1.302364
WLSE	1.156698	1.280952	1.566834	1.431192
PCE	0.944854	1.522263	1.962323	2.392981
MPSE	1.144715	1.109365	1.888477	1.348811
CVME	0.999214	1.252308	1.047122	1.262961
ADE	1.03128	1.270303	1.748756	1.553576
RADE	1.064402	1.260156	2.120223	1.493293

Table 3. True reliability vs. approximated reliability based on estimates in Table 2.

No Data Contamination					
Source	Min.	Q ₁	Q ₂	Q ₃	Max.
α, β	0.954444	0.568825	0.448565	0.152986	0.043167
MME	0.977602	0.668269	0.555444	0.171589	0.044892
MLE	0.969228	0.62178	0.5	0.158257	0.042022
LSE	0.955583	0.642644	0.552352	0.224777	0.080792
WLSE	0.955623	0.64561	0.556697	0.228682	0.083095
PCE	0.981386	0.728172	0.640978	0.233718	0.071739
MPSE	0.946619	0.598942	0.497945	0.197174	0.067713
CVME	0.968506	0.656495	0.555049	0.197184	0.060605
ADE	0.966481	0.657323	0.55958	0.206088	0.066026
RADE	0.963067	0.6506	0.554487	0.210102	0.069557

Upper 10% Data Contamination					
Source	Min.	Q ₁	Q ₂	Q ₃	Max.
α, β	0.954444	0.568825	0.448565	0.152986	0.043167
MME	0.964443	0.745342	0.692176	0.402873	0.034084
MLE	0.911042	0.590419	0.515503	0.259574	0.012739
LSE	0.942659	0.636059	0.556207	0.253684	0.008556
WLSE	0.926149	0.639353	0.573918	0.303064	0.019867
PCE	0.935947	0.709472	0.664469	0.439685	0.071761
MPSE	0.891835	0.606203	0.548099	0.319037	0.031561
CVME	0.96475	0.65341	0.556279	0.207602	0.002869
ADE	0.917316	0.644617	0.586833	0.335648	0.031491
RADE	0.882689	0.61531	0.564643	0.352855	0.048659

**Figure 1.** *Cont.*

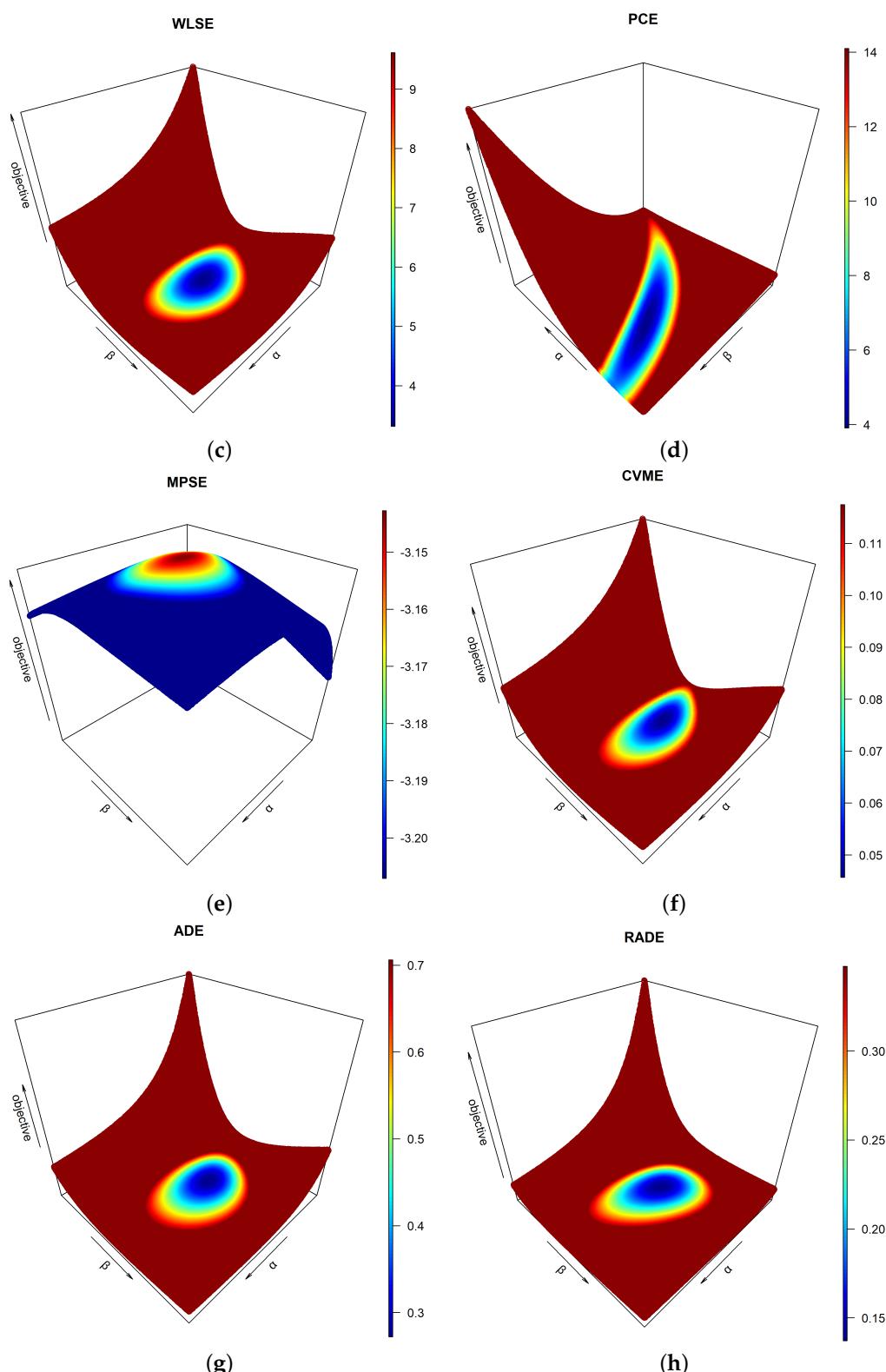


Figure 1. 3D plots of the objective functions based on data in Table 1 for (a) MLE, (b) LSE, (c) WLSE, (d) PCE, (e) MPSE, (f) CVME, (g) ADE and (h) RADE. (cont.)

3.2. Real Data Analysis

To illustrate the application of the considered estimation methods in practice, the concrete compressive strength data of [30] is considered for analysis. This data set established from 17 different sources to check the reliability of a suggested strength model. The data

gathered concrete comprising cement alongside fly ash, blast furnace slag, and superplasticizer. The data set consisted of a single response variable; namely, the compressive strength of concrete (in MPa), and 8 covariates. Using this data, various estimated models are obtained and compared by the means of Kolmogorov-Smirnov (KS) test. The latter one-sample testing procedure is used to test the null that the distribution function of a given data set is that of the probability distribution of interest. To obtain the KS statistic, one must consider the following steps:

1. Obtain the estimates of the parameters α and β , denoted by $\hat{\alpha}$ and $\hat{\beta}$.
2. Compute $u_{i:n} = F(x_{i:n}, \hat{\alpha}, \hat{\beta})$, such that $x_{i:n}$ is the observed i th sample order statistics, where $i = 1, \dots, n$. Here, $F(\cdot, \cdot, \cdot)$ is given by Equation (2).
3. Calculate the value of KS statistic as follows:

$$\text{KS} = \max_{1 \leq i \leq n} \left\{ u_{i:n} - \frac{i-1:n}{n}, \frac{i}{n} - u_{i:n} \right\},$$

and accordingly calculate the p -value to make a decision about the hypotheses.

Furthermore, since ties exist and the model parameters were estimated, the p -values for the KS statistics were obtained using $B = 1,000$ parametric bootstrapping samples. The steps to obtain the bootstrapping p -value for KS test are as follow:

1. For each method, obtain the estimates of the model parameters α and β ; say, $\hat{\alpha}$ and $\hat{\beta}$.
2. Use the estimates in the previous step and the algorithm in the preceding section to generate a random sample X_1^*, \dots, X_n^* from the LBS distribution with shape parameter $\hat{\alpha}$ and scale parameter $\hat{\beta}$.
3. Compute the KS statistics for each bootstrap sample as discussed before, i.e., repeat Steps 2 and 3, B times to obtain $\text{KS}^{(1)}, \dots, \text{KS}^{(B)}$.
4. Calculate the p -value as follows:

$$p\text{-value}^* = \frac{1}{B} \sum_{j=1}^B I_j,$$

where I_j is an indicator function, such that $I_j = 1$ if $\text{KS}^{(j)} > \text{KS}$, and zero otherwise, for $j = 1, \dots, B$ and KS is the KS statistic obtained from the original data set.

Finally, it is important to mention that since ties were observed in the data, the MPSEs cannot be acquired directly. When ties exist, one may use a generalization of the maximum product of spacings method to obtain the required estimators; see Murage et al. [31] for additional details. Alongside the estimated parameters and the goodness-of-fit statistics, the reliability probability is calculated at 17 MPa, 28 MPa, and 70 MPa by substituting these values in Equation (3) and replacing the model parameters with the corresponding estimates. In practice, concrete compressive strength can fluctuate between 17 MPa and 28 MPa for residential concrete, while it can be higher as 70 MPa in the case of commercial constructions [32].

Tables 4–7 respectively summarize the analyses of the considered data set assuming no contamination in the data, 20% of upper data contamination (i.e., the upper 20% of order statistics are multiplied by 5), 20% of lower data contamination (i.e., the lower 20% of order statistics are divided by 5), and 40% two-tailed data contamination which is a mixture of the previous data contamination cases. From the latter tables, one can note the following observations based on the values of the KS test statistics:

- When there is no data contamination, both MLEs and MPSEs performed well in terms of goodness-of-fit.
- In the case of upper data contamination, ADEs outperformed both MLEs and MPSEs which took second and third place, respectively.
- On the other hand, both MLEs and MPSEs maintained their performance followed by ADEs in the case of lower data contamination.

- In contrast, WLSEs have performed better than MPSEs and MLEs when two-tailed data contamination exists.
- Overall, MLEs and MPSEs provided the best results in terms of goodness-of-fit, and they both have endured data contamination unlike their counterparts. This is most likely due to the fact that the sample size is large (1000+ units). Furthermore, PCEs and MMEs did not perform well among compared to their counterparts in all considered settings.
- Finally, according to the reliability proportions estimated by MLEs and MPSEs, one can conclude that the sampled specimens of [30] were suitable for residential buildings.

Table 4. Statistical analysis assuming no data contamination.

Source	α	β	KS	p-Value	$\Pr(X > 17)$	$\Pr(X > 28)$	$\Pr(X > 70)$
MME	0.409	30.679	0.139	0.517	0.884	0.600	0.063
MLE	0.437	33.943	0.064	0.602	0.900	0.678	0.092
LSE	0.468	33.788	0.072	0.527	0.888	0.666	0.102
WLSE	0.394	33.261	0.074	0.588	0.912	0.677	0.072
PCE	0.262	34.057	0.139	0.353	0.967	0.764	0.030
MPSE	0.444	33.719	0.065	0.658	0.896	0.671	0.093
CVME	0.467	33.789	0.072	0.508	0.888	0.666	0.101
ADE	0.462	33.369	0.069	0.682	0.887	0.658	0.097
RADE	0.390	34.211	0.075	0.517	0.920	0.701	0.077

Table 5. Statistical analysis assuming 20% of upper data contamination.

Source	α	β	KS	p-Value	$\Pr(X > 17)$	$\Pr(X > 28)$	$\Pr(X > 70)$
MME	0.855	48.919	0.287	0.490	0.873	0.654	0.116
MLE	0.846	36.446	0.165	0.549	0.877	0.666	0.123
LSE	0.541	34.110	0.194	0.478	0.877	0.659	0.114
WLSE	0.548	34.091	0.193	0.490	0.877	0.654	0.111
PCE	0.555	70.314	0.528	0.364	0.925	0.745	0.109
MPSE	0.842	36.242	0.166	0.449	0.871	0.654	0.119
CVME	0.540	34.110	0.194	0.497	0.878	0.659	0.114
ADE	0.959	36.881	0.160	0.502	0.869	0.651	0.121
RADE	1.246	35.897	0.180	0.476	0.868	0.650	0.120

Table 6. Statistical analysis assuming 20% of lower data contamination.

Source	α	β	KS	p-Value	$\Pr(X > 17)$	$\Pr(X > 28)$	$\Pr(X > 70)$
MME	0.932	17.976	0.383	0.516	0.783	0.466	0.084
MLE	0.906	32.245	0.175	0.482	0.866	0.647	0.122
LSE	0.450	33.947	0.198	0.461	0.888	0.666	0.102
WLSE	0.366	33.806	0.199	0.487	0.915	0.682	0.071
PCE	0.293	31.851	0.200	0.354	0.962	0.745	0.031
MPSE	0.923	31.872	0.176	0.546	0.859	0.638	0.124
CVME	0.450	33.948	0.198	0.481	0.888	0.666	0.102
ADE	1.050	29.162	0.181	0.604	0.864	0.636	0.113
RADE	0.379	34.413	0.199	0.491	0.920	0.702	0.076

Table 7. Statistical analysis assuming 40% of two-tailed data contamination.

Source	α	β	KS	p-Value	$\Pr(X > 17)$	$\Pr(X > 28)$	$\Pr(X > 70)$
MME	1.375	28.531	0.199	0.503	0.790	0.542	0.127
MLE	1.325	33.400	0.165	0.598	0.841	0.632	0.147
LSE	0.549	34.119	0.195	0.461	0.876	0.658	0.115
WLSE	1.024	33.683	0.156	0.525	0.873	0.651	0.114
PCE	0.565	68.760	0.514	0.363	0.921	0.736	0.108
MPSE	1.331	33.062	0.163	0.747	0.837	0.627	0.148
CVME	0.547	34.119	0.195	0.477	0.876	0.658	0.115
ADE	1.563	32.972	0.180	0.793	0.845	0.630	0.138
RADE	1.436	33.489	0.175	0.609	0.867	0.649	0.120

4. Simulation Outcomes

This section presents the outcomes of Monte Carlo simulation experiments based on 1000 random samples from the LBS distribution with different combinations of values for the shape parameter and sample sizes assuming the following scenarios:

- Model 1: A model with no contamination.
- Model 2: A model with 10% of severe upper contamination, i.e., the upper 10% of order statistics are multiplied by 5.
- Model 3: A model with 10% of severe lower contamination, i.e., the lower 10% of order statistics are multiplied by 1/5.
- Model 4: A model with 20% of severe two-tailed contamination, i.e., the upper 10% of order statistics are multiplied by 5, while the lower 10% of order statistics are multiplied by 1/5.

For each scenario, the simulation study assumes $n = 10(10)100$, $\alpha = 0.5(0.5)2.0$, and $\beta = 1$, without loss of any generality. To measure estimation efficiency, the simulated bias and simulated root mean-squared-error (RMSE) are calculated as

$$\text{Bias}(\alpha) = \frac{1}{N} \sum_{i=1}^N (\hat{\alpha}_i - \alpha), \quad \text{Bias}(\beta) = \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_i - \beta),$$

$$\text{RMSE}(\alpha) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\alpha}_i - \alpha)^2} \quad \text{and} \quad \text{RMSE}(\beta) = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{\beta}_i - \beta)^2},$$

such that $N = 1000$, while $\hat{\alpha}_i$ ($\hat{\beta}_i$) is an estimate of the model parameter α (β) based on simulation repetition i . Furthermore, to measure the goodness-of-fit of the fitted model parameters based on the nine estimators, the average absolute difference between the true and estimated reliability function (D_{abs}), and the maximum absolute difference between the true and estimated reliability function (D_{max}) are determined as

$$D_{\text{abs}} = \frac{1}{n \times N} \sum_{i=1}^N \sum_{j=1}^n |S(t_j; \alpha, \beta) - S(t_j; \hat{\alpha}_i, \hat{\beta}_i)|,$$

and

$$D_{\text{max}} = \frac{1}{N} \sum_{i=1}^N \max_{j=1,\dots,n} |S(t_j; \alpha, \beta) - S(t_j; \hat{\alpha}_i, \hat{\beta}_i)|,$$

respectively, such that $S(t; \alpha, \beta) = 1 - F(t; \alpha, \beta)$, while $\hat{\alpha}_i$ ($\hat{\beta}_i$) is an estimate of the model parameter α (β) based on simulation repetition i .

From a statistical perspective, an estimator is computationally consistent when its simulated bias tends to 0 as the sample size increases. Furthermore, when the simulated bias neither increases nor decreases when data contamination exists, then one can conclude that the estimator is computationally robust. Here, Figures 2 and 3 clearly indicate that the

most consistent and robust estimators for the model parameters α and β are the MPSEs and CVMEs regardless of the sample size and the true value of α .

In computational statistics, an estimator is computationally efficient when the simulated RMSE tends to 0 as the sample size increases regardless of the existence of data contamination. Figures 4 and 5 suggest that the MPSEs, CVMEs, and the LSEs of α and β are the most efficient estimators compared to the other ones regardless of the simulation settings.

In goodness-of-fit analysis, whenever a pair of estimators yields D_{abs} and D_{max} that tends to 0 as the sample size increases, and are not negatively affected by data contamination, then this pair of estimates provides the least difference between the true and estimated reliability function. This is very important in practice since the aim is to find the best approximation for the reliability. Figures 6 and 7 again indicate that MPSEs, CVMEs, and the LSEs of α and β are the estimators that performed well in terms of goodness-of-fit.

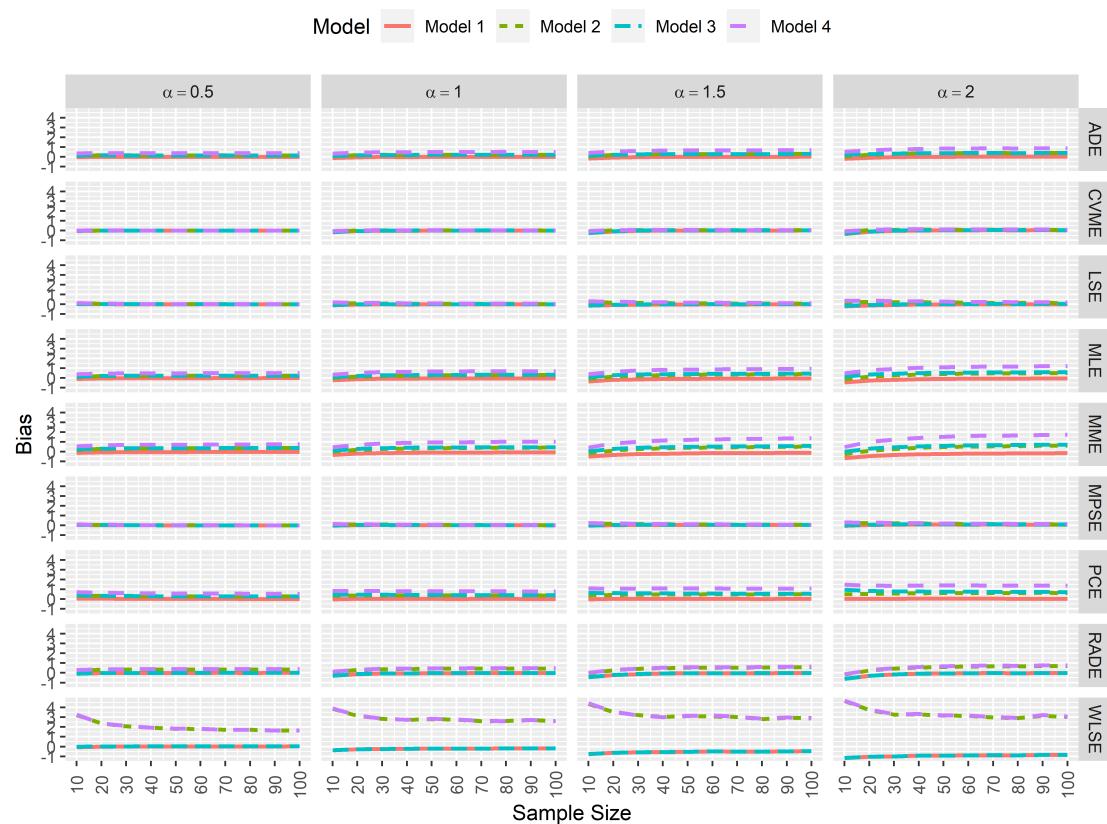
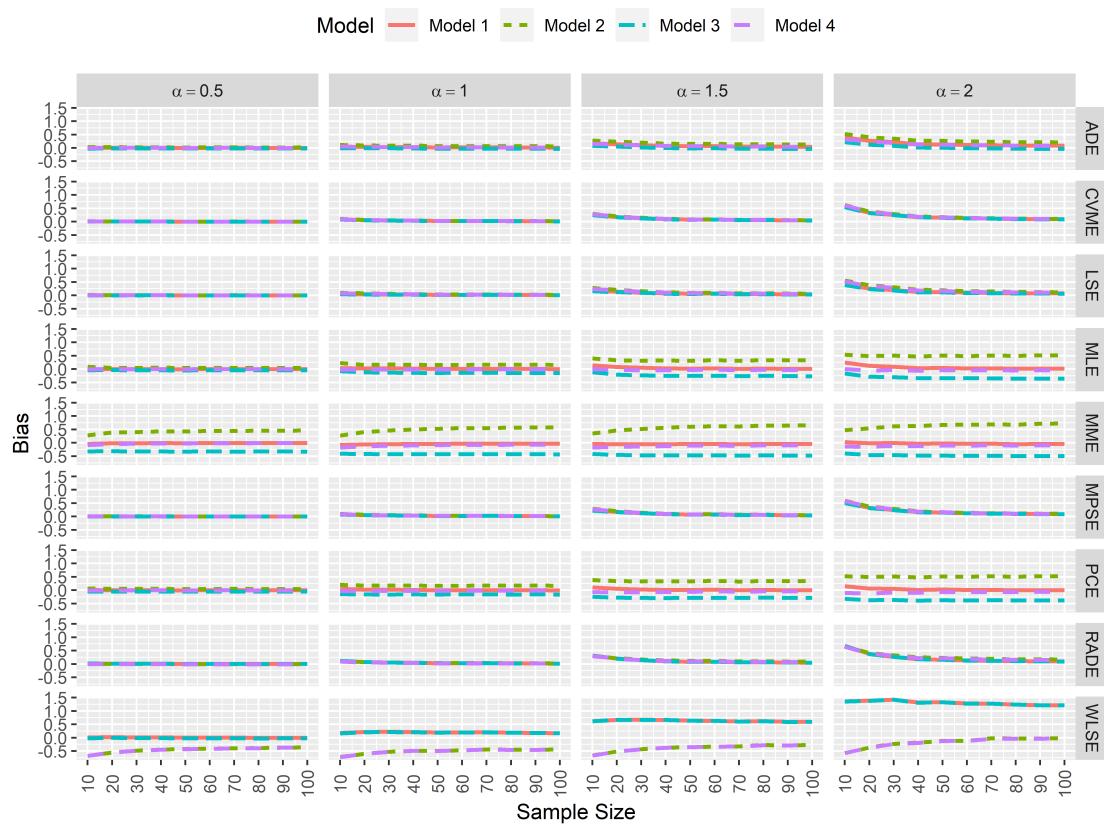
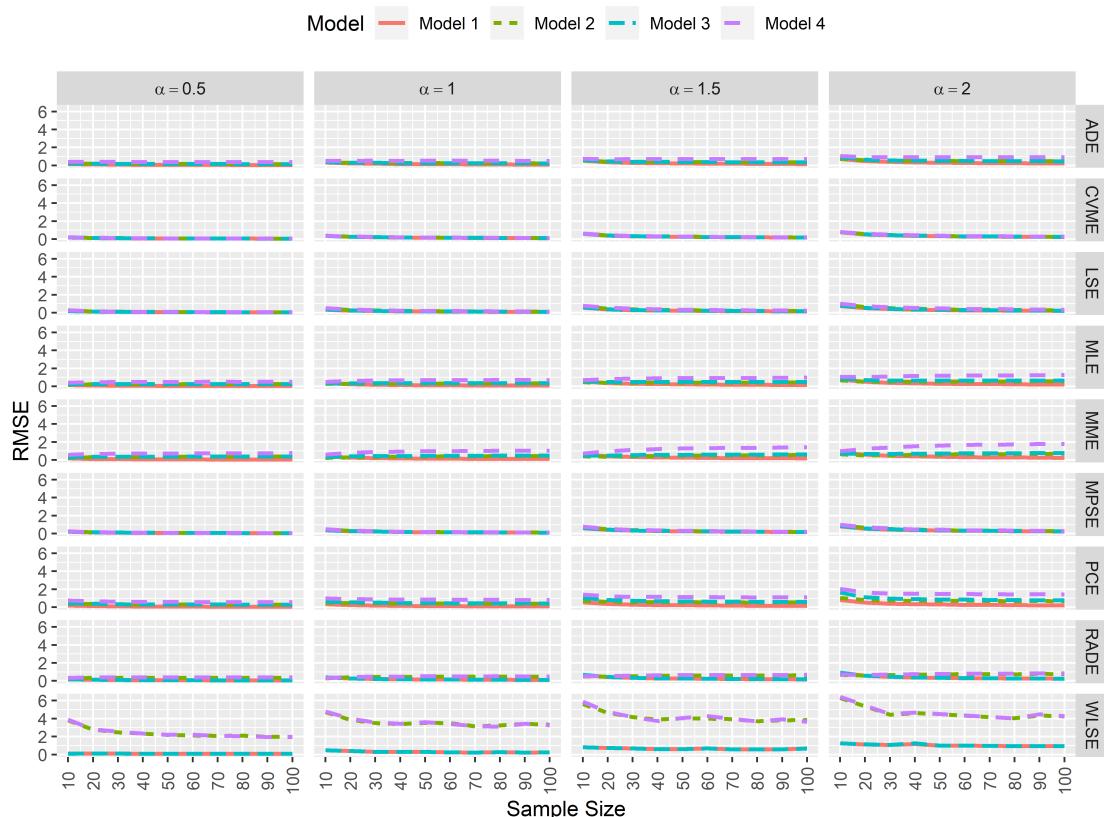


Figure 2. Simulated biases for the estimators of α .

Figure 3. Simulated biases for the estimators of β .Figure 4. Simulated RMSEs for the estimators of α .

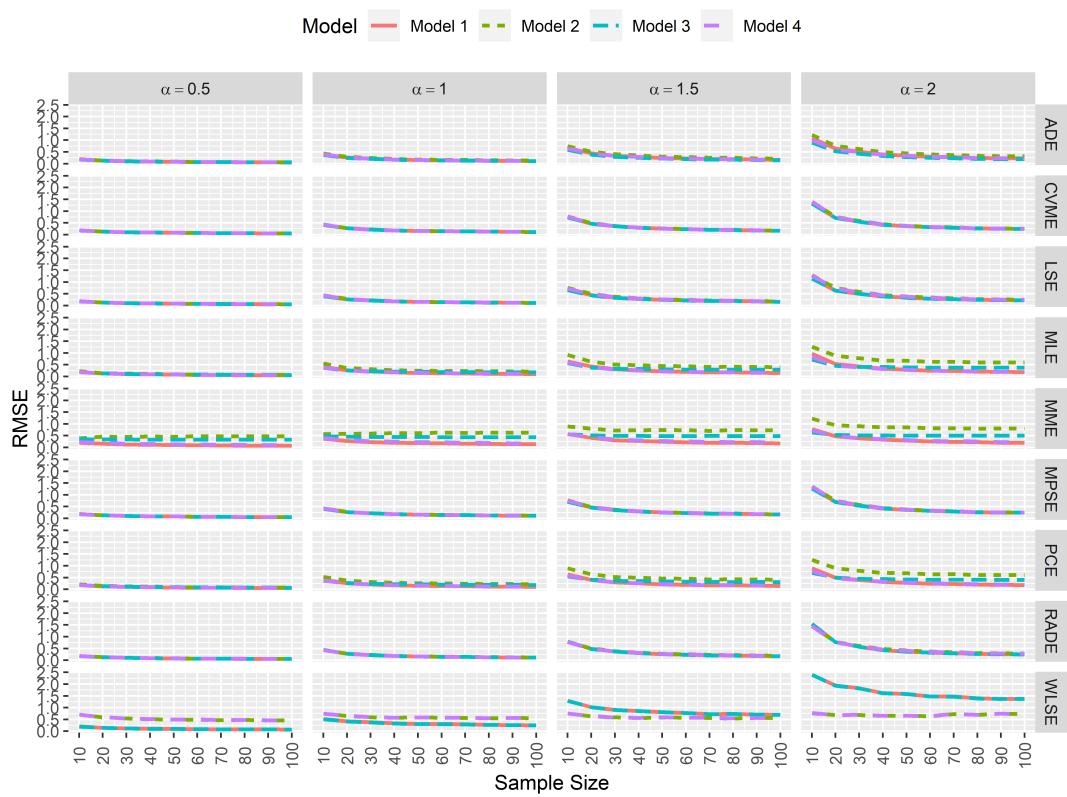


Figure 5. Simulated RMSEs for the estimators of β .

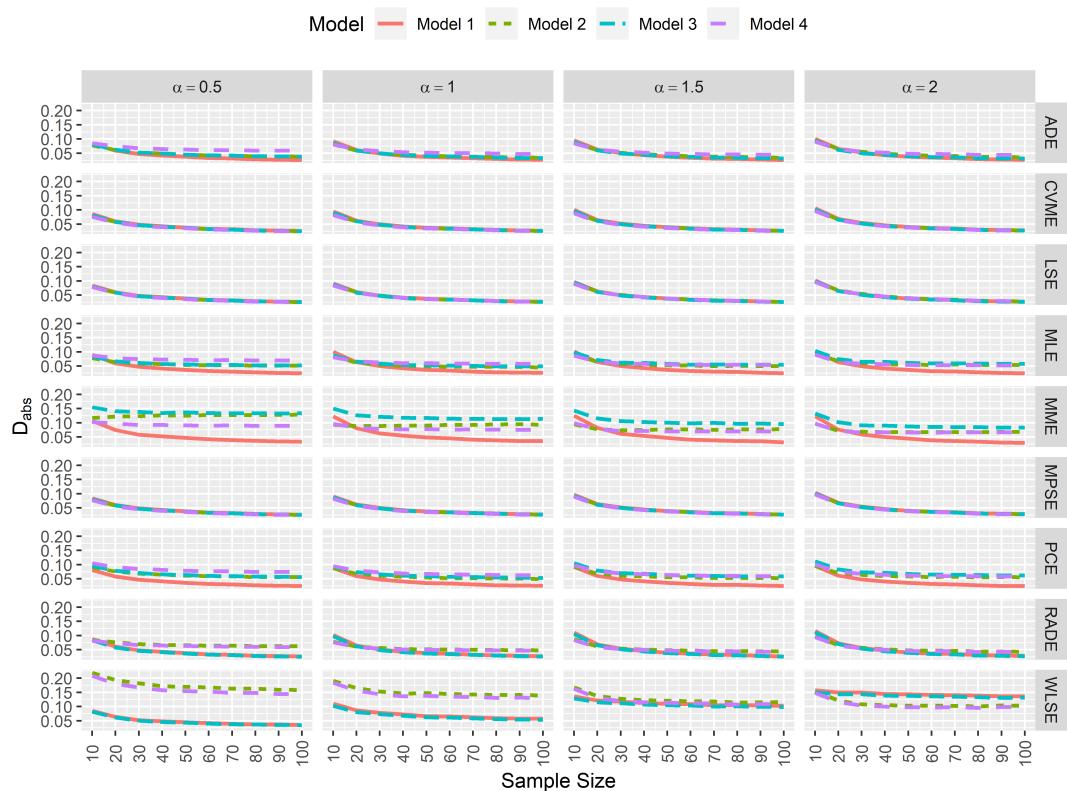


Figure 6. Simulated average absolute difference between the true and estimated reliability function.

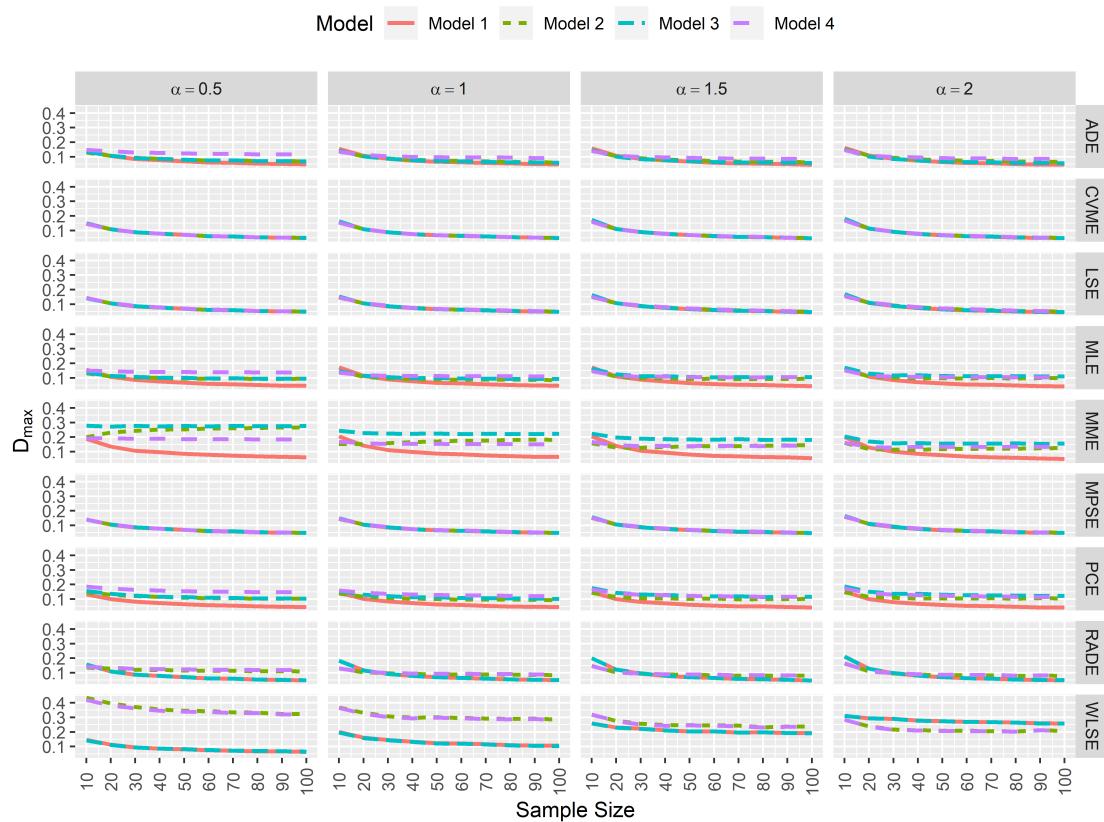


Figure 7. Simulated maximum absolute difference between the true and estimated reliability function.

5. Conclusions

In this paper, the estimation problem of the parameters and the reliability function of the Laplace Birnbaum-Saunders lifetime distribution is considered. Besides the method of maximum likelihood, eight classical frequentist estimation methods have been discussed for this purpose; namely, modified moments, maximum product of spacings, least-squares, weighted least-squares, percentile, Cramér-von Mises, Anderson-Darling and right-tailed Anderson-Darling estimation methods. Based on the assumption that the invariance property is exist for the different estimation methods, the reliability function is also estimated using the different estimation methods. To compare the performance of the different estimators a Monte Carlo simulation study is conducted. The practical application of the estimators is illustrated by analyzing a simulated data set and one real data set belongs to compressive strength of concrete. Both data analyses and the Monte Carlo simulation study indicated that all methods perform well when there is no contamination in the data. Once there is some contamination in the data, maximum product of spacings, least-squares, and Cramér-von Mises estimates are notably robust compared to the other estimators and the performance of the other method improve as the sample size increases. Data contamination is not the only problem that faces researchers in practice. Data censoring is another practical challenge that needs to be addressed in future research since it negatively impacts estimation efficiency and robustness. Another important research direction is to compare the studied frequentist estimators to Bayesian estimation in terms of performance based on additional real experimental results which are available in literature.

Author Contributions: F.M.A.A. and M.N. contributed equally to this work. Both the authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: The authors would like to thank the Editorial Board and two anonymous reviewers for their constructive feedback that improved the final version of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Birnbaum, Z.; Saunders, S. A new family of life distributions. *J. Appl. Probab.* **1969**, *6*, 319–327. [[CrossRef](#)]
2. Birnbaum, Z.W.; Saunders, S.C. Estimation for a Family of Life Distributions with Applications to Fatigue. *J. Appl. Probab.* **1969**, *6*, 328–347. [[CrossRef](#)]
3. Langlands, A.O.; Pocock, S.J.; Kerr, G.R.; Gore, S.M. Long-term survival of patients with breast cancer: A study of the curability of the disease. *BMJ* **1979**, *2*, 1247–1251. [[CrossRef](#)]
4. Desmond, A. Stochastic models of failure in random environments. *Can. J. Stat.* **1985**, *13*, 171–183. [[CrossRef](#)]
5. Balakrishnan, N.; Kundu, D. Birnbaum-Saunders distribution: A review of models, analysis, and applications. *Appl. Stoch. Model. Bus. Ind.* **2019**, *35*, 4–49. [[CrossRef](#)]
6. Bourguignon, M.; Ho, L.L.; Fernandes, F.H. Control charts for monitoring the median parameter of Birnbaum-Saunders distribution. *Qual. Reliab. Eng. Int.* **2020**, *36*, 1333–1363. [[CrossRef](#)]
7. Hassani, H.; Kalantari, M.; Entezarian, M.R. A new five-parameter Birnbaum-Saunders distribution for modeling bicoid gene expression data. *Math. Biosci.* **2020**, *319*, 108275. [[CrossRef](#)] [[PubMed](#)]
8. Kannan, G.; Jeyadurga, P.; Balamurali, S. Economic design of repetitive group sampling plan based on truncated life test under Birnbaum-Saunders distribution. *Commun. Stat. Simul. Comput.* **2020**, *1*–17. [[CrossRef](#)]
9. Diáz-García, J.A.; Leiva-Sánchez, V. A new family of life distributions based on the elliptically contoured distributions. *J. Stat. Plan. Inference* **2005**, *128*, 445–457 [[CrossRef](#)]
10. Balakrishnan, N.; Leiva, V.; López, J. Acceptance Sampling Plans from Truncated Life Tests Based on the Generalized Birnbaum-Saunders Distribution. *Commun. Stat. Simul. Comput.* **2007**, *36*, 643–656. [[CrossRef](#)]
11. Leiva, V.; Barros, M.; Paula, G.A.; Sanhueza, A. Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics* **2008**, *19*, 235–249. [[CrossRef](#)]
12. Sanhueza, A.; Leiva, V.; Balakrishnan, N. The Generalized Birnbaum-Saunders Distribution and Its Theory, Methodology, and Application. *Commun. Stat. Theory Methods* **2008**, *37*, 645–670. [[CrossRef](#)]
13. Zhu, X.; Balakrishnan, N. Birnbaum-Saunders distribution based on Laplace kernel and some properties and inferential issues. *Stat. Probab. Lett.* **2015**, *101*, 1–10. [[CrossRef](#)]
14. Gupta, R.D.; Kundu, D. Generalized exponential distribution: Different method of estimations. *J. Stat. Comput. Simul.* **2001**, *69*, 315–337. [[CrossRef](#)]
15. Alkasasbeh, M.R.; Raqab, M.Z. Estimation of the generalized logistic distribution parameters: Comparative study. *Stat. Methodol.* **2009**, *6*, 262–279. [[CrossRef](#)]
16. Mazucheli, J.; Louzada, F.; Ghitany, M. Comparison of estimation methods for the parameters of the weighted Lindley distribution. *Appl. Math. Comput.* **2013**, *220*, 463–471. [[CrossRef](#)]
17. do Espírito Santo, A.; Mazucheli, J. Comparison of estimation methods for the Marshall-Olkin extended Lindley distribution. *J. Stat. Comput. Simul.* **2014**, *85*, 3437–3450. [[CrossRef](#)]
18. Balakrishnan, N.; Alam, F.M.A. Maximum likelihood estimation of the parameters of Student's *t* Birnbaum-Saunders distribution: A comparative study. *Commun. Stat. Simul. Comput.* **2019**, *1*–30. [[CrossRef](#)]
19. Lawson, C.; Keats, J.; Montgomery, D. Comparison of robust and least-squares regression in computer-generated probability plots. *IEEE Trans. Reliab.* **1997**, *46*, 108–115. [[CrossRef](#)]
20. Boudt, K.; Caliskan, D.; Croux, C. Robust explicit estimators of Weibull parameters. *Metrika* **2009**, *73*, 187–209. [[CrossRef](#)]
21. Agostinelli, C.; Marazzi, A.; Yohai, V.J. Robust Estimators of the Generalized Log-Gamma Distribution. *Technometrics* **2014**, *56*, 92–101. [[CrossRef](#)]
22. Wang, M.; Park, C.; Sun, X. Simple robust parameter estimation for the Birnbaum-Saunders distribution. *J. Stat. Distrib. Appl.* **2015**, *2*. [[CrossRef](#)]
23. Nelder, J.A.; Mead, R. A Simplex Method for Function Minimization. *Comput. J.* **1965**, *7*, 308–313. [[CrossRef](#)]
24. Swain, J.J.; Venkatraman, S.; Wilson, J.R. Least-squares estimation of distribution functions in Johnson's translation system. *J. Stat. Comput. Simul.* **1988**, *29*, 271–297. [[CrossRef](#)]
25. Kao, J.H.K. Computer Methods for Estimating Weibull Parameters in Reliability Studies. *IRE Trans. Reliab. Qual. Control* **1958**, *PGRQC-13*, 15–22. [[CrossRef](#)]
26. Kao, J.H.K. A Graphical Estimation of Mixed Weibull Parameters in Life-Testing of Electron Tubes. *Technometrics* **1959**, *1*, 389–407. [[CrossRef](#)]
27. Cheng, R.C.H.; Amin, N.A.K. *Maximum Product-of-Spacings Estimation with Applications to the Lognormal Distribution*; Technical Report 1; Department of Mathematics, University of Wales IST: Cardiff, UK, 1979.

28. Cheng, R.C.H.; Amin, N.A.K. Estimating Parameters in Continuous Univariate Distributions with a Shifted Origin. *J. R. Stat. Soc. Ser. B* **1983**, *45*, 394–403. [[CrossRef](#)]
29. Ranneby, B. The Maximum Spacing Method. An Estimation Method Related to the Maximum Likelihood Method. *Scand. J. Stat.* **1984**, *11*, 93–112.
30. Yeh, I.C. Modeling of strength of high performance concrete using artificial neural networks. *Cem. Concr. Res.* **1998**, *28*, 1797–1808. [[CrossRef](#)]
31. Murage, P.; Mungátu, J.; Odero, E. Optimal Threshold Determination for the Maximum Product of Spacing Methodology with Ties for Extreme Events. *Open J. Model. Simul.* **2019**, *7*, 149–168. [[CrossRef](#)]
32. NRMCA. *Concrete in Practice: What, Why and How? CIP 35—Testing Compressive Strength of Concrete*; National Ready Mixed Concrete Association: Silver Spring, MD, USA, 2020.