A Metaheuristic Approach to Parameter Estimation of a Flexible Parametric Mixture Cure Rate Model with Interval-Censored Data

P. BORGES[1] and M. CAMPOS[2]

ABSTRACT. A flexible parametric mixture cure model, called bi-lognormal cure rate model or simply BLN model, is defined and studied. The BLN model can be effectively used to analyze survival dataset in the presence of long-term survivors, especially when the dataset presents the underlying phenomenon of latent competing risks or when there is evidence that a bimodal hazard function is appropriated to described it, which are advantages over other cure rate models found in the literature. We discuss the maximum likelihood estimation for the model parameters considering interval-censored data through the differential evolution algorithm that is a nature-inspired computing metaheuristic used for global optimization of functions defined in multidimensional spaces. This approach is also used because the likelihood function of the model is multimodal and the direct application of gradient methods in this case is not ideal, since such methods are local search methods with a high chance of getting stuck at a local maximum when the starting point is chosen outside the basin of attraction of a global maximum. In addition, a simulation study was implemented to compare the performance of differential evolution algorithm with the performance of the Newton-Raphson algorithm in terms of bias, root mean square error, and the coverage probability of the asymptotic confidence intervals for the parameters. Finally, an application of the BLN model to real data is presented to illustrate that it can provide a better fit than other mixture cure rate models.

Keywords: polyhazard model, lognormal distribution, differential evolution algorithm, Newton-Raphson algorithm, generalized Turnbull's nonparametric estimator.

1 INTRODUCTION

Survival models that take into account a cure fraction, known as cure rate models or long-term survival models, are frequent topics considered in many theoretical and applied statistics articles. The relevance of these models has increased in the last decade, mainly due to medical advances,

---

*Corresponding author: Patrick Borges – E-mail: patrick.borges@ufes.br

[1] Departamento de Estatística, Universidade Federal do Espiríto Santo, Vitória, ES, Brazil – E-mail: patrick.borges@ufes.br https://orcid.org/0000-0001-6143-4353

[2] Departamento de Estatística, Universidade Federal do Espiríto Santo, Vitória, ES, Brazil – E-mail: mauro.campos@ufes.br https://orcid.org/0009-0004-0382-6702

in which a considerable fraction of patients has been cured of certain pathologies, for example, in some types of cancer. The most common approach to the cure rate model is the standard mixture model approached by Boag [7], advanced by Berkson and Gage [6], and later extensively studied by Farewell [23, 24], Maller and Zhou [33], Roman et al. [42], and Mazucheli et al. [35], among others. The standard mixture model assumes that the survival function for the entire population, $S_p(t)$, is a mixture of the immunes and susceptibles individuals (or components) to the cause of failure under study, i.e., $S_p(t) = p_0 + (1 - p_0)S(t)$ is an improper survival function (with $\lim_{t\to\infty} S_p(t) = S_p(\infty) = p_0$), where $p_0 \in [0, 1]$ is the cure fraction and $S(t)$ is the proper survival function (with $S(\infty) = 0$) of susceptibles patients, in which the frequent choices for $S(t)$ have been the exponential, gamma, log-logistic, lognormal, Weibull and Gompertz distributions. We emphasize that, although the mixture cure rate model is the most widely used in the literature, other forms of population modeling containing cured individuals have been studied, such as: (i) models based on the latent competing risks structure [16], see for example [48], [13], [9, 10, 14, 41], and [8], among others; and (ii) models based on defective distributions, referred to as distributions that are not normalized to one for some values of its parameters, a concept introduced by Balka et al. [3].

As noted and studied by Mazucheli et al. [35], we can find in practice datasets a fraction $(1 - p_0)$ of units which are subject to failures of various competing causes. In addition, the exact cause of failure may be unknown, leading to the latent competing risks problem [31]. Thus, in order to accommodate such a situation, the survival function, $S(t)$, of the lifetime of susceptible individuals is modeled by using polysurvival models. The main advantage of such models compared to single hazard models, such as the Weibull and log-logistic models, is the flexibility to represent hazard rate functions with unusual shapes, as bathtub and multimodal curves. There are many applied examples of these models in the literature, for example: Kalbeisch and Prentice [26] proposed the poly-log-logistic model for log-logistic competing risks; Berger and Sun [5] proposed the poly-Weibull model for Weibull's competing risks; Louzada-Neto [31] proposed a generalized polyhazard model encompassing the poly-Weibull, poly-log-logistic and generalized-poly-gamma models; Basu et al. [4] and Kuo and Yang [28] used the poly-Weibull model to model masked systems in which the cause of the failure may be unknown or partially known; Mazucheli et al. [34] presented a Bayesian inference procedure for polyhazard models with covariates; Louzada-Neto et al. [32] analyzed the non-identifiability problem arising on the poly-Weibull model; Fachini et al. [22] presented several diagnostic methods for the poly-survival model; Tsai and Hotta [46, 47] generalized the traditional polyhazard model by allowing the latent causes of failure to be dependent by using copula functions; and Demiris et al. [20] presented methods for survival extrapolation using the poly-Weibull model.

In some situations, however, the population of interest having a component of cure and the form of the data to be interval-censored, rather than the more common form of censoring found in many practical problems, i.e. right-censored. Interval censoring occurs when the subjects under study are not continuously monitored over time, but are inspected at some specific time points only. Hence, if the event is observed, the time-to-event is known to occur between two consec-

utive inspection times. Therefore, the exact lifetimes are not recorded, but it is known that the lifetimes belong to an interval $[L, R]$, where L is the latest inspection time before the occurrence of the event and R is the first inspection after the occurrence of event [45]. $[L, \infty]$ represents the case where the event of interest is not observed by the end of the last inspection time in which case the lifetime is considered to be right-censored. Although interval-censored data have been extensively studied in the literature, there are not many references dealing with interval-censored data in a cure rate setup. The following references [11,25,37,43] and the references therein could be mentioned as few examples.

As an illustration of the aforementioned problems, that is, the scenario of competing risks, long-term survivors, and interval-censored, we consider data from a field trial of 4,992 circuit boards extracted from [12]. The data consists on the lifetimes of the circuit boards observed during a period of 10,000 hours of operation or until failure. A total of 95 circuit board failures are observed. For the circuit boards that fail before the end of the experiment, engineering judgment indicates that failure can occur due to infant failure or wearout, but the exact cause of failure is unknown. Besides, the lifetimes of these failed units are interval-censored between two inspections, and the remaining lifetimes are right-censored. There is a great deal of censorship (there were 4,897 censored lifetimes) in the data, evidencing a possible adequacy of the cure rate model approach. Furthermore, the cumulative hazard plot for these dataset was shown in [35] and it was observed that two possible causes of failure are competing. Given this, at least in principle, these behaviors indicate that models that ignore the possibility of cure and competing risks will not be adequate for these dataset.

Therefore, in this paper, we consider the situation where the objective is to analyse a survival dataset in the presence of long-term survivors, in particular, when this dataset presents the underlying phenomenon of latent competing risks or when there is evidence that a bimodal hazard function is appropriated to described it. For this situation, we propose a mixture cure rate model based on the structure proposed by Mazucheli et al. [35] with two causes of failure latent, assuming in our approach the hypothesis that the lifetime associated with a particular cause of failure follows a lognormal distribution in presence of interval-censoring. For proposed model, the explicit expression of the maximum likelihood estimators of the parameters by considering interval-censored data cannot be obtained explicitly by analytical methods, since equating the first-order log-likelihood derivatives to zero leads us to a complicated system of non-linear equations. In this case, we need of a iterative computer method to solve this problem. Gradient methods (such as Newton-Raphson and Fisher scoring) require finding the scoring vector and the inverse of the Hessian matrix (or a good approximation of this matrix) associated with the log-likelihood function, but these operations are hard to be realized analytically. In addition, gradient methods are local search procedures that usually are trapped by a local optima when applied to optimize multimodal functions, even when strategies of multiple runs are used. For such reasons, we have proposed to use a well-known evolutionary algorithm, called differential evolution (DE) to find a good approximation of a global maximum of the log-likelihood function. DE is a nature-inspired metaheuristic developed by Storn and Prince [44] for global optimization

of multimodal functions defined in multidimensional spaces. It uses the principles of biological evolution and natural selection to simulate the evolution of living organisms subject to biological mechanisms, which lead to fittest descendants to their environment. DE does not use derivatives of the objective function in its search procedure. Instead, it uses an ensemble of feasible solutions of the optimization problem as a population of organisms that evolve by application of operators simulating biological mechanisms such as mutation, crossover, and selection. In other words, DE is a derivative-free algorithm that can be considered as an efficient method for the general problem of global optimization, with many successful applications in various scientific field [17, 21, 27, 30, 36, 38]. In our Monte Carlo simulation study, it was observed that DE converges to a good approximation of a global maximum of the log-likelihood function in almost all runs. That is, given a priori approximation error, DE converges with high success rate. Besides, DE is robust and does not require any other method to choose good starting points to achieve results near of a global maximum.

The remainder of this paper is structured as follows: Section 2 presents the bi-lognormal cure rate model and describes the maximum likelihood estimation based on DE algorithm to estimate the model parameters for interval-censored data. Section 3 presents a simulation study to compare the performance of DE algorithm with the performance of the Newton-Raphson algorithm in terms of bias, root mean square error, and coverage probability of the asymptotic confidence intervals for the model parameters. An application of our model to a real dataset is presented in Section 4. Some final comments in Section 5 conclude the paper.

2   METHODS

2.1   Bi-lognormal mixture cure rate model

The lognormal distribution assumes that lifetime T has density function given by

$$f(t;\mu,\sigma) = \frac{1}{\sqrt{2\pi}t\sigma}\exp\left\{-\frac{1}{2}\left(\frac{\log(t)-\mu}{\sigma}\right)^2\right\}, \quad t > 0, \tag{2.1}$$

where $\mu \in \mathbb{R}$ is the mean lifetime logarithm ($\mu = E[\log(T)]$), just as $\sigma > 0$ is the standard deviation of $\log(T)$. The corresponding survival and hazard functions are represented respectively by

$$S(t;\mu,\sigma) = \Phi\left(\frac{-\log(t)+\mu}{\sigma}\right) \quad \text{and} \quad h(t;\mu,\sigma) = \frac{f(t;\mu,\sigma)}{S(t;\mu,\sigma)}, \tag{2.2}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. The shape of the hazard function is unimodal, but it does not have a closed analytical form. A similar comment holds for the survival function. It is important to highlight that the normal and lognormal distributions have certain equivalences, because if a random variable has a lognormal distribution, its logarithm has a normal distribution. For this reason, there are many mathematical properties between the two distributions, as discussed in [2].

Now, in building our model, let T be the lifetime of an individual (or component). The mixture cure rate model performs a T lifetime dissection as

$$T = \eta\infty + (1-\eta)T^*, \tag{2.3}$$

where T^* denotes the lifetime of a susceptible (who is not a long-term survivor or immune the cause of failure under study) and η indicates, by the values 1 or 0, whether the sampled subject is a susceptible or long-term survivor, independently of T^*. Let $p_0 = Pr(\eta = 1) \in [0,1]$ be the proportion of immunes, i.e., the cure fraction. Then, the survival function of T, $S_p(t)$, is given by

$$S_p(t) = p_0 + (1-p_0)S(t), \tag{2.4}$$

where $S(t) = \Pr(T^* > t)$ is a proper survival function (with $S(\infty) = 0$).

Following Mazucheli et al. [35], we suppose that an individual (or component) is subject to two independent sources of failure cause indexed by $j = 1,2$. Assume further that the distribution of lifetime related to the jth cause, X_j, may be sufficiently described by a lognormal form with density defined in Eq.(2.1). Then, the observed event time, $T^* = \min\{X_1, X_2\}$, is said to follow a bi-lognormal distribution if its survival and hazard functions are respectively

$$S(t; \mu_1, \mu_2, \sigma_1, \sigma_2) = \prod_{j=1}^{2} S_j(t; \mu_j, \sigma_j) \tag{2.5}$$

and

$$h(t; \mu_1, \mu_2, \sigma_1, \sigma_2) = \sum_{j=1}^{2} h_j(t; \mu_j, \sigma_j), \tag{2.6}$$

where $S_j(\cdot)$ and $h_j(\cdot)$ are the lognormal survival and hazard functions defined in Eq.(2.2). The shape of hazard function (2.6) is unimodal or bimodal.

Remark 1. *Although there are probability distributions that accommodate unimodal hazard functions, such as inverse Weibull, log-logistic, and exponentiated Weibull distributions, the literature in this area is lacking distributions that support bimodal hazard functions. In this sense, the bi-lognormal distribution is flexible enough to accommodate this type of hazard, which is another motivation for its formulation.*

Under these ponderations, based on bi-lognormal survival function (2.5), the survival function (2.4) is given by

$$S_p(t; \boldsymbol{\vartheta}) = p_0 + (1-p_0)\prod_{j=1}^{2} S_j(t; \mu_j, \sigma_j) \tag{2.7}$$

where $\boldsymbol{\vartheta} = (p_0, \mu_1, \mu_2, \sigma_1, \sigma_2)^T \in \boldsymbol{\Theta} = [0,1] \times \mathbb{R}^2 \times \mathbb{R}_+^2$. From here on, we refer to the model in Eq. (2.7) as the bi-lognormal mixture cure rate model, or simply, the BLN model.

2.2   Maximum likelihood estimation for the BLN model

Let us consider the situation when the lifetime T in Eq. (2.3) is not completely observed and is subject to interval-censoring, i.e., T belongs to an interval $[L, R]$, where L is the latest inspection

time before the occurrence of the event of interest and R is the first inspection after the occurrence of event of interest. Note that $[L,\infty]$ refers to the situation where the event of interest is not observed before the last inspection time in which case T is considered to be right censored. The censoring indicator is thus defined as $\delta = I(R < \infty)$, which takes the value 0 if $[L,\infty]$, meaning that the event is not observed for a subject before the last inspection time, and takes the value 1 if $R < \infty$, meaning that the event took place but its exact time is not known and is only known to belong to the interval $[L,R]$.

Based on n observations of times and censoring indicators $(L_1 = l_1, R_1 = r_1, \delta_1), \ldots, (L_n = l_n, R_n = r_n, \delta_n)$, Hashimoto et al. [25] shown that the likelihood function under uninformative censoring can be expressed as

$$L(\boldsymbol{\vartheta}; \mathscr{D}) \propto \prod_{i=1}^{n} [S_p(l_i; \boldsymbol{\vartheta}) - S_p(r_i; \boldsymbol{\vartheta})]^{\delta_i} [S_p(l_i; \boldsymbol{\vartheta})]^{1-\delta_i} \tag{2.8}$$

where $\mathscr{D} = (\mathbf{l}, \mathbf{r}, \boldsymbol{\delta})$, $\mathbf{l} = (l_1, \ldots, l_n)^T$, $\mathbf{r} = (r_1, \ldots, r_n)^T$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)^T$. Using the survival function of the BLN model, given in Eq. (2.7), the likelihood function in Eq. (2.8) can also be expressed as

$$L(\boldsymbol{\vartheta}; \mathscr{D}) \propto \prod_{i=1}^{n} \left(q_0^{\delta_i} [\Delta S(l_i, r_i)]^{\delta_i} [p_0 + q_0 S(l_i)]^{1-\delta_i} \right) \tag{2.9}$$

where $q_0 = 1 - p_0$, $\Delta S(l_i, r_i) = S(l_i) - S(r_i)$, and $S(\cdot) = S(\cdot; \mu_1, \mu_2, \sigma_1, \sigma_2)$ is the bi-lognormal survival function defined in Eq.(2.5). From the likelihood function in Eq. (2.9), the maximum likelihood estimator (MLE) $\widehat{\boldsymbol{\vartheta}}$ of the parameter $\boldsymbol{\vartheta}$ can be carried out. Since the MLE $\widehat{\boldsymbol{\vartheta}}$ is not available in closed form, we need of a iterative computer method to maximize the log-likelihood function $l(\boldsymbol{\vartheta}; \mathscr{D}) = \log L(\boldsymbol{\vartheta}; \mathscr{D})$.

Remark 2. *The likelihood function (2.9) is exchangeable, i.e., considering the BNL model, the value of (2.9) will be the same if the MLEs of μ_1 and σ_1 are exchanged for the MLEs of μ_2 and σ_2 and vice versa. Therefore, to avoid this identifiability problem, we consider the constraint $\mu_1 < \mu_2$.*

Remark 3. *We consider the R programming language [39] through the* `DEoptim` *function, which is fully documented in the* `DEoptim` *package in R [1], to compute the MLE $\widehat{\boldsymbol{\vartheta}}$ of the parameter $\boldsymbol{\vartheta}$. The* `DEoptim` *function implements the DE algorithm that is discussed in detail in Subsection 2.3.*

Asymptotic properties of MLE $\widehat{\boldsymbol{\vartheta}}$ are required to make inferences about the unknown parameter $\boldsymbol{\vartheta}$ of the BLN model. Following [15, 29], under certain conditions of regularity, the MLE $\widehat{\boldsymbol{\vartheta}}$ is asymptotically unbiased and efficient. Moreover, its distribution converges to normal with the variance-covariance matrix given by the inverse of the Fisher information matrix. Given this, we present an approximate method for constructing confidence intervals (CIs) for $\boldsymbol{\vartheta}$ using the asymptotic properties of $\widehat{\boldsymbol{\vartheta}}$. Thus, let us first denote the Fisher information matrix of $\boldsymbol{\vartheta}$ by

$$\mathbf{I}_{\mathrm{E}}(\boldsymbol{\vartheta}) = E\left[-\frac{\partial^2 \ell(\boldsymbol{\vartheta}; \mathscr{D})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T} \right]. \tag{2.10}$$

Since the Fisher information matrix is very complicated to be obtained due to the censored observations (censoring is random and noninformative), we resort to the observed Fisher information matrix of $\boldsymbol{\vartheta}$ obtained in the form

$$\mathbf{I}_0(\boldsymbol{\vartheta}) = -\frac{\partial^2 \ell(\boldsymbol{\vartheta}; \mathscr{D})}{\partial \boldsymbol{\vartheta} \partial \boldsymbol{\vartheta}^T}, \tag{2.11}$$

evaluated at the maximum likelihood estimator $\boldsymbol{\vartheta} = \widehat{\boldsymbol{\vartheta}}$, which is a consistent estimator of I_E. The required second derivatives are computed numerically.

The multivariate normal $\mathcal{N}_5(\boldsymbol{\vartheta}, I_0^{-1}(\widehat{\boldsymbol{\vartheta}}))$ distribution can be used to construct asymptotic CIs for the parameters. In fact, an $100(1-\gamma)\%$ asymptotic CI for each parameter κ is given by

$$CI(\kappa, 100(1-\gamma)\%) = \left(\hat{\kappa} - z_{\gamma/2}\sqrt{\widehat{\mathbb{V}ar(\hat{\kappa})}}; \hat{\kappa} + z_{\gamma/2}\sqrt{\widehat{\mathbb{V}ar(\hat{\kappa})}} \right), \tag{2.12}$$

where $\widehat{\mathbb{V}ar}(\cdot)$ is the diagonal element of $I_0^{-1}(\widehat{\boldsymbol{\vartheta}})$ corresponding to each parameter ($\kappa = \mu_1, \mu_2, \sigma_1, \sigma_2, p_0$) and $z_{\gamma/2}$ is the $(1-\gamma/2)$ quantile of the standard normal distribution.

Remark 4. *As mentioned above the asymptotic properties of MLE $\widehat{\boldsymbol{\vartheta}}$ are valid only under certain regularity conditions, which are not easy to verify analytically in our BLN model. Therefore, in the next section, we investigate the asymptotic properties of the MLEs by means of a simulation study. It is worth noting that many authors have performed simulations to assess the asymptotic behavior of MLEs, especially when the analytical investigation is not trivial [11, 40].*

2.3 MLE based on differential evolution algorithm

In general, differential evolution (DE) is a nature-inspired algorithm used for global optimization of multimodal functions defined in multidimensional spaces [17, 38, 44]. In the context of the maximum likelihood estimation, DE is described as follows. Let $l(\boldsymbol{\vartheta}; \mathscr{D})$ be a log-likelihood function to be maximized over a given parametric space $\boldsymbol{\Theta} \subseteq \mathbb{R}^D$. For this optimization problem, consider a population $\mathscr{P}^g$ of K feasible solutions in $\boldsymbol{\Theta}$. A feasible solution is a D-dimensional vector $\boldsymbol{\vartheta}_k^g = (\vartheta_{k1}^g, \ldots, \vartheta_{kD}^g)$ in $\boldsymbol{\Theta}$. The index k, where $k = 1, \ldots, K$, labels the kth solution in $\mathscr{P}^g$ and the index g represents the generation counter of the algorithm, which in turn acts to evolve the population $\mathscr{P}^g$ in order to find a good approximation for an optimal solution of the problem.

DE consists of three main steps: mutation, crossover, and selection. Mutation produces a donor vector $\boldsymbol{v}_k^{g+1}$ according to the following scheme: for each solution $\boldsymbol{\vartheta}_k^g \in \mathscr{P}^g$, three distinct solutions

$$\boldsymbol{\vartheta}_{k_1}^g, \boldsymbol{\vartheta}_{k_2}^g, \boldsymbol{\vartheta}_{k_3}^g \in \mathscr{P}^g \tag{2.13}$$

are randomly chosen to generate $\boldsymbol{v}_k^{g+1}$ as follows

$$\boldsymbol{v}_k^{g+1} = \boldsymbol{\vartheta}_{k_1}^g + F \cdot (\boldsymbol{\vartheta}_{k_2}^g - \boldsymbol{\vartheta}_{k_3}^g) \tag{2.14}$$

where $k \neq k_1 \neq k_2 \neq k_3$ and $F \in [0, 1]$ is a parameter controlling the amplitude of the differential variation. Crossover implements a discrete recombination between the donor vector $\boldsymbol{v}_k^{g+1}$ and

the current solution $\boldsymbol{\vartheta}_k^g$ to produce the offspring $\boldsymbol{u}_k^{g+1}$. Crossover is controlled by a crossover probability $CR \in [0,1]$ and it is implemented as

$$u_{kd}^{g+1} = \begin{cases} v_{kd}^{g+1} & \text{with probability } CR \\ \vartheta_{kd}^g & \text{otherwise} \end{cases} \tag{2.15}$$

for all $d = 1, \ldots, D$. After the mutation step, if an element of $\boldsymbol{v}_k^{g+1}$ is found to violate the bounds of the parametric space Θ, it is reset in such a way that the bounds are respected. An analogous procedure is done for $\boldsymbol{u}_k^{g+1}$ after the crossover step. Finally, the selection step is used to select the best solution between $\boldsymbol{u}_k^{g+1}$ and $\boldsymbol{\vartheta}_k^g$ to go to the next generation. Therefore, we have

$$\boldsymbol{\vartheta}_k^{g+1} = \begin{cases} \boldsymbol{u}_k^{g+1} & \text{if } l(\boldsymbol{u}_k^{g+1}) \geq l(\boldsymbol{\vartheta}_k^g) \\ \boldsymbol{\vartheta}_k^g & \text{otherwise.} \end{cases} \tag{2.16}$$

The population $\mathscr{P}^{g+1}$ in generation $g+1$ consists of the set of solutions $\{\boldsymbol{\vartheta}_k^{g+1} : k = 1, \ldots, K\}$. The process is repeated until some stopping criterion is met. At the end of the evolution process, DE returns the best solution $\boldsymbol{\vartheta}_*$ found in the last population (corresponding to last generation of the algorithm) and its objective value $l_* = l(\boldsymbol{\vartheta}_*; \mathscr{D})$, which is the value of the log-likelihood function on $\boldsymbol{\vartheta}_*$.

The essential steps of DE algorithm are summarized as the pseudo code shown in Algorithm 1. The parameter F is commonly chosen in the range $[0.4, 1]$, CR is a probability in the range $[0, 1]$, and K is commonly chosen between $5D$ and $10D$, where D is the dimensionality of the parametric space $\Theta \subseteq \mathbb{R}^D$. Variants of DE algorithm are discussed in the DE literature. These different DE strategies basically differ in the way that the donor vector is created and the way that the offspring is generated. In order to characterize these strategies, a general notation was adopted in the literature, namely: $DE/x/y/z$, where x refers to the mutation scheme to create the donor vector, y indicates the number of difference vectors used in the mutation scheme, and z indicates the crossover scheme.

The DE strategy discussed in this subsection is referred to as $DE/rand/1/bin$ (see Algorithm 1), indicating that the donor vector was created from three solutions randomly chosen and with only one difference vector, and that the offspring was generated using the binomial crossover scheme. Readers may be referred to [38] and [17] for more details.

Finally, it is important to note that the `DEoptim` package in R [1] implements several DE strategies, including the strategy $DE/rand/1/bin$. The structure of the `DEoptim` function in the `DEoptim` package is familiar to readers who are accustomed to the R syntax.

Algorithm 1 Differential Evolution - DE/rand/1/bin

Require: $D, K, F \in [0,1]$ and $CR \in [0,1]$

1:   $g = 0$
2:   Initialize the population $\mathscr{P}^0$ with K solutions
3:   **repeat**
4:     **for** each solution $\boldsymbol{\vartheta}_k^g$ in $\mathscr{P}^g$ **do**
5:       Select $k_1, k_2,$ and k_3 such that $k \neq k_1 \neq k_2 \neq k_3$
6:       Generate $\boldsymbol{v}_k^{g+1} = \boldsymbol{\vartheta}_{k_1}^g + F \cdot (\boldsymbol{\vartheta}_{k_2}^g - \boldsymbol{\vartheta}_{k_3}^g)$
7:       **for** $d = 1$ **to** D **do**
8:         $r \sim U(0,1)$
9:         **if** $r < CR$ **then**
10:           $u_{kd}^{g+1} = v_{kd}^{g+1}$
11:         **else**
12:           $u_{kd}^{g+1} = \vartheta_{kd}^g$
13:         **end if**
14:       **end for**
15:       **if** $l(\boldsymbol{u}_k^{g+1}) \geq l(\boldsymbol{\vartheta}_k^g)$ **then**
16:         $\boldsymbol{\vartheta}_k^{g+1} = \boldsymbol{u}_k^{g+1}$
17:       **else**
18:         $\boldsymbol{\vartheta}_k^{g+1} = \boldsymbol{\vartheta}_k^g$
19:       **end if**
20:     **end for**
21:     $\mathscr{P}^{g+1} = \{\boldsymbol{\vartheta}_k^{g+1} : k = 1, \ldots, K\}$
22:     $\boldsymbol{\vartheta}_* =$ Best solution in $\mathscr{P}^{g+1}$ and $l_* = l(\boldsymbol{\vartheta}_*)$
23:     $g = g + 1$
24:   **until** some termination condition is met
25:   **return** $\boldsymbol{\vartheta}_*$ and l_*

---

3   SIMULATION RESULTS

As noted by Mazucheli et al. [35], in lifetime studies, it is common to find datasets with a small or moderate amount of observed lifetimes. In addition, in the dataset in Section 4, we can observe the presence of a large amount of censoring. Therefore, a Monte Carlo simulation study was conducted to assess the performance of the MLEs via DE algorithm and Newton-Raphson method in small and moderate samples when censoring is observed. To this end, we examine three measures, namely: the bias, root mean square error, and probability of coverage of the asymptotic CIs for the parameters of BLN model. The results were obtained from 1,000 Monte Carlo repetitions and the simulations were carried out in the R programming language. In each replication, we consider the BLN model in Eq. (2.7) with parameters $\mu_1 = 4.0$, $\mu_2 = 12.0$, $\sigma_1 = 6.5$, $\sigma_2 = 0.2$, and the cured fraction fixed in $p_0 = 0.3$, 0.6 and 0.9 when sample size $n = 50, 100, 200, 300, 400$ and 500. To introduce random censoring, the distribution of censoring times is assumed to

be exponential with rate α, which is set to control the proportion of right-censored observations. Interval-censored data (L_i, R_i, δ_i) are generated by using the following steps, according to the same strategy described in [37]:

Algorithm 2 Interval-censored data generated according to the BLN model

Require: $\mu_1, \mu_2, \sigma_1, \sigma_2, p_0, \alpha$, and n

1: Define values for $\mu_1, \mu_2, \sigma_1, \sigma_2, p_0$, and α.

2: For the ith observation, draw $u_i \sim \text{Uniform}(0,1)$ and $C_i \sim \text{Exponential}(\alpha)$, where α is set to control the proportion of right-censored observations.

3: If $u_i < p_0$, set $L_i = C_i$, $R_i = \infty$ and $\delta_i = 0$; otherwise, generate X_j lognormal random variable with parameters μ_j and σ_j, $j = 1,2$ and set $T_i^* = \min\{X_1, X_2\}$.

4: If $\min\{T_i^*, C_i\} = C_i$, set $L_i = C_i$, $R_i = \infty$ and $\delta_i = 0$.

5: If $\min\{T_i^*, C_i\} = T_i^*$, $\delta_i = 1$ and generate len_i from Uniform(0.2,0.7) and l_i from Uniform(0,1). Then, create intervals $(0, l_i], (l_i, l_i + len_i], \ldots, (l_i + k \times len_i, \infty)$, $k = 1, 2, \ldots$, and choose (L_i, R_i) that satisfies $L_i < T_i^* \le R_i$.

6: Repeat steps 2 to 5 for all $i = 1, \ldots, n$. The data set for the ith observation is (L_i, R_i, δ_i), $i = 1, \ldots, n$.

7: **return** $\{(L_i, R_i, \delta_i) : i = 1, \ldots, n\}$

The generated data have approximately $100(p_0 + 0.05)$ of censoring data, which fits in the context of models with a cure fraction p_0. Figures 1-3 list the bias of the five parameters of the BLN model and the corresponding root mean squared errors (RMSEs) and the empirical coverage probabilities (CPs) of 95% of the asymptotic CIs. The results indicate that: (i) with respect to the performance of bias, RMSEs for parameters, the DE algorithm is slightly better than the Newton-Raphson method in most simulation combinations; (ii) as expected and independent of the estimation algorithm, bias and RMSEs decrease as sample size increases, while the amount of censorship induces increases, especially for small samples; (iii) the empirical CPs of the parameters μ_1, μ_2 and p_0 are closer to nominal levels when n increases, while σ_1 and σ_2 are always less than nominal level, worsening as the proportion of censorship increases. Although the CPs for the σ_1 and σ_2 parameters improves for a large sample size, we still find it unsatisfactory even for n as large as 500, especially when the censorship proportion is very large. Therefore, the confidence interval based on the asymptotic normality of maximum likelihood estimators should not be used unless n is considerably large, which is the situation studied in our application of Section 4. Other parameter adjustments were also used, however, the performance of the parameter estimation procedure did not present significant differences for the results of Figures 1-3.
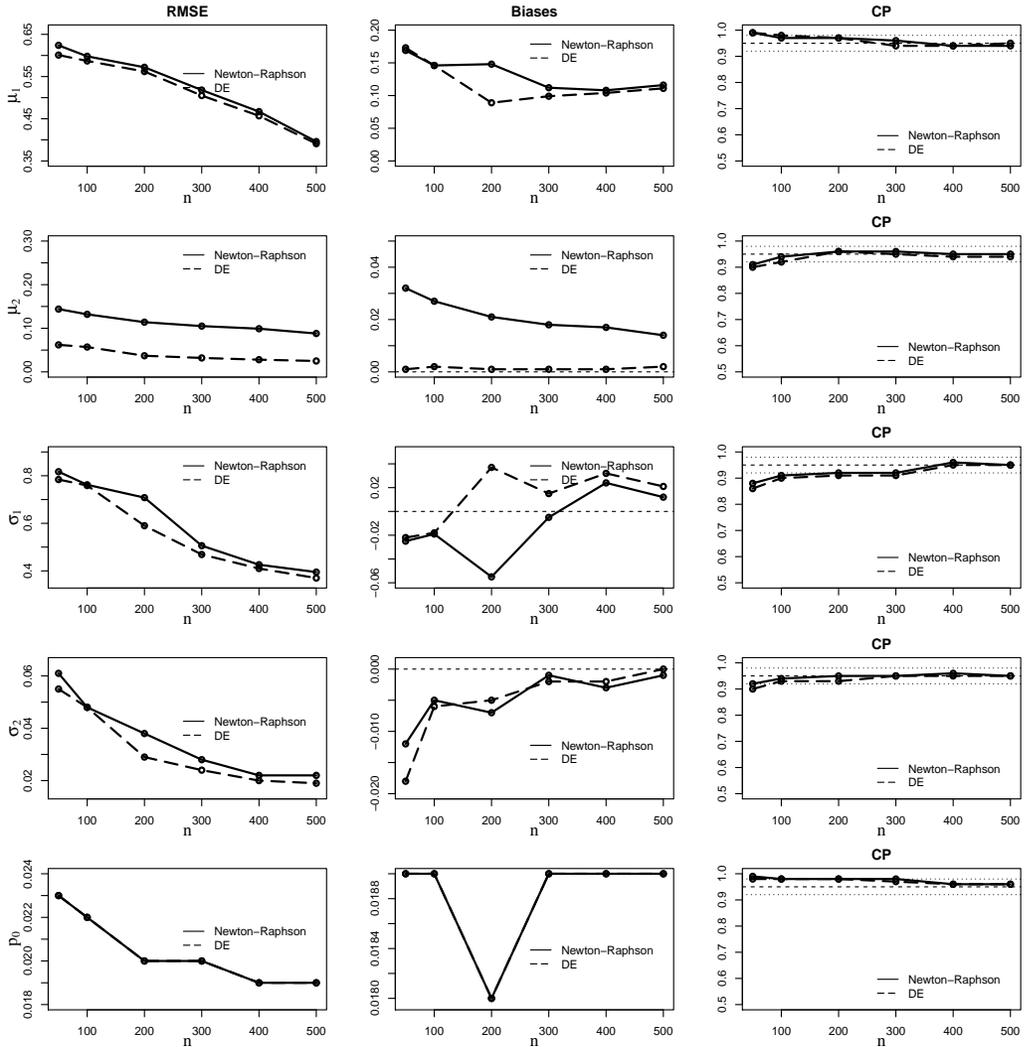
Figure 1: Bias, RMSEs of the MLEs of BLN model parameters with censoring percentages $100(p_0 = 0.3 + 0.05)$.
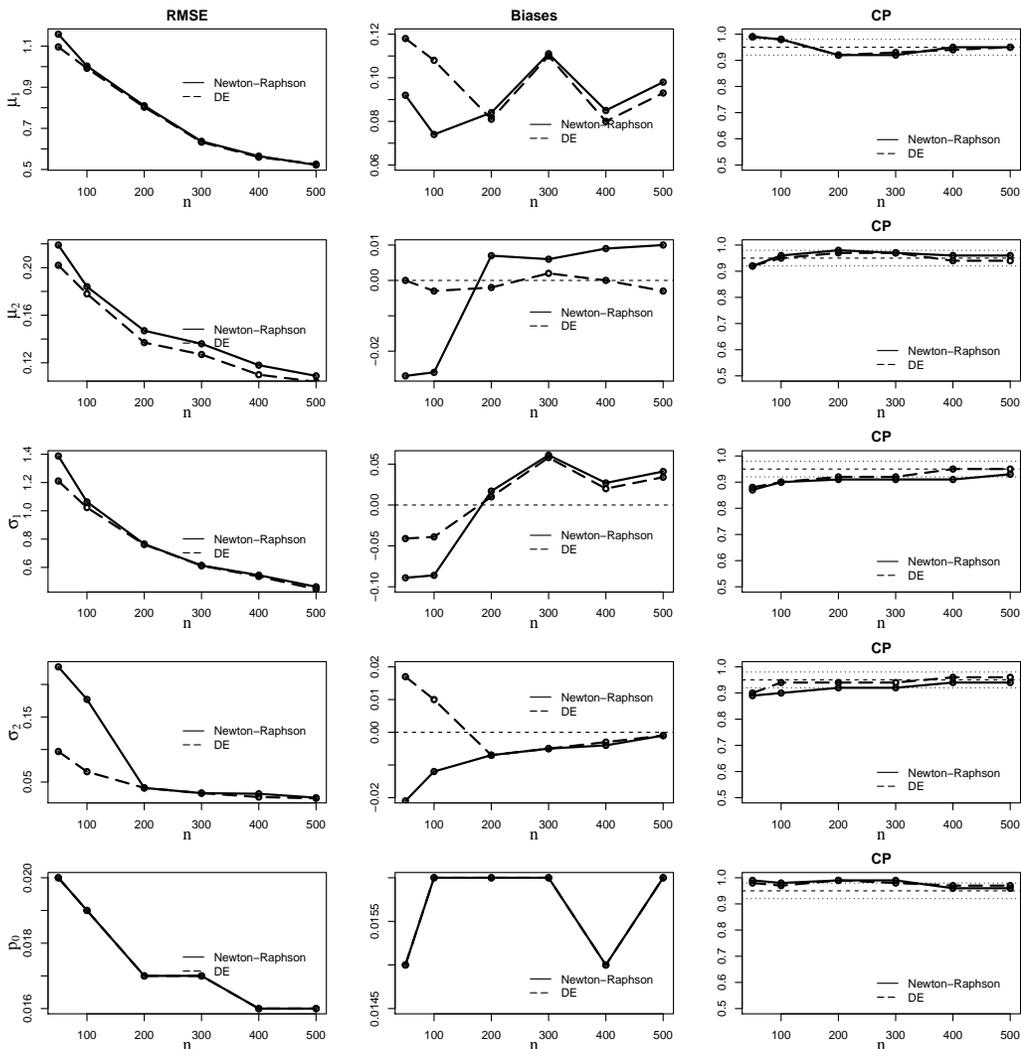
Figure 2: Bias, RMSEs of the MLEs of BLN model parameters with censoring percentages $100(p_0 = 0.6 + 0.05)$.
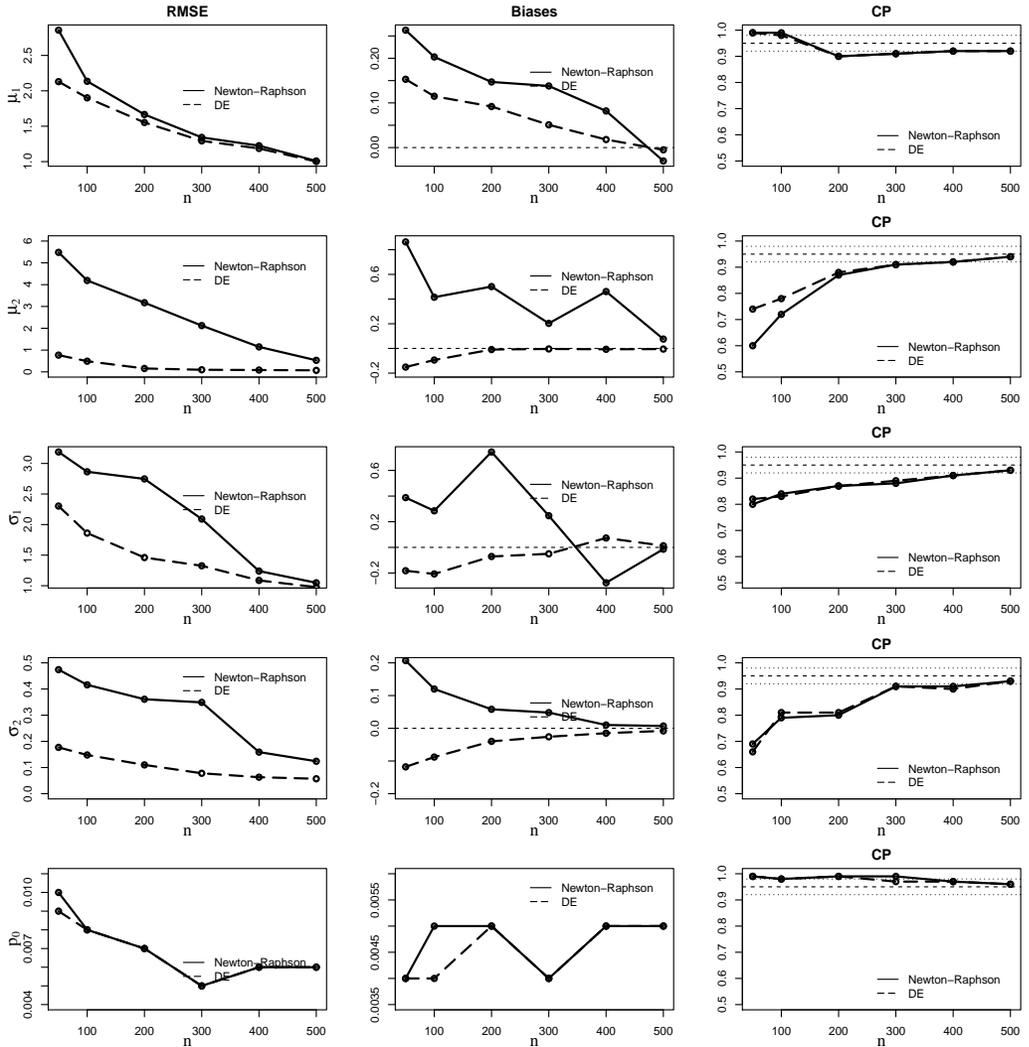
Figure 3: Bias, RMSEs of the MLEs of BLN model parameters with censoring percentages $100(p_0 = 0.9 + 0.05)$.

Further, we give the plots of the empirical distributions of $\widehat{\mu_1}$, $\widehat{\mu_2}$, $\widehat{\sigma_1}$, $\widehat{\sigma_2}$ and $\hat{p}_0$ when $n = 500$ and $100(p_0 = 0.9 + 0.05)$. The plots displayed in Figure 4 indicate that the normal distributions provides reasonable approximations to the distributions of these estimators.
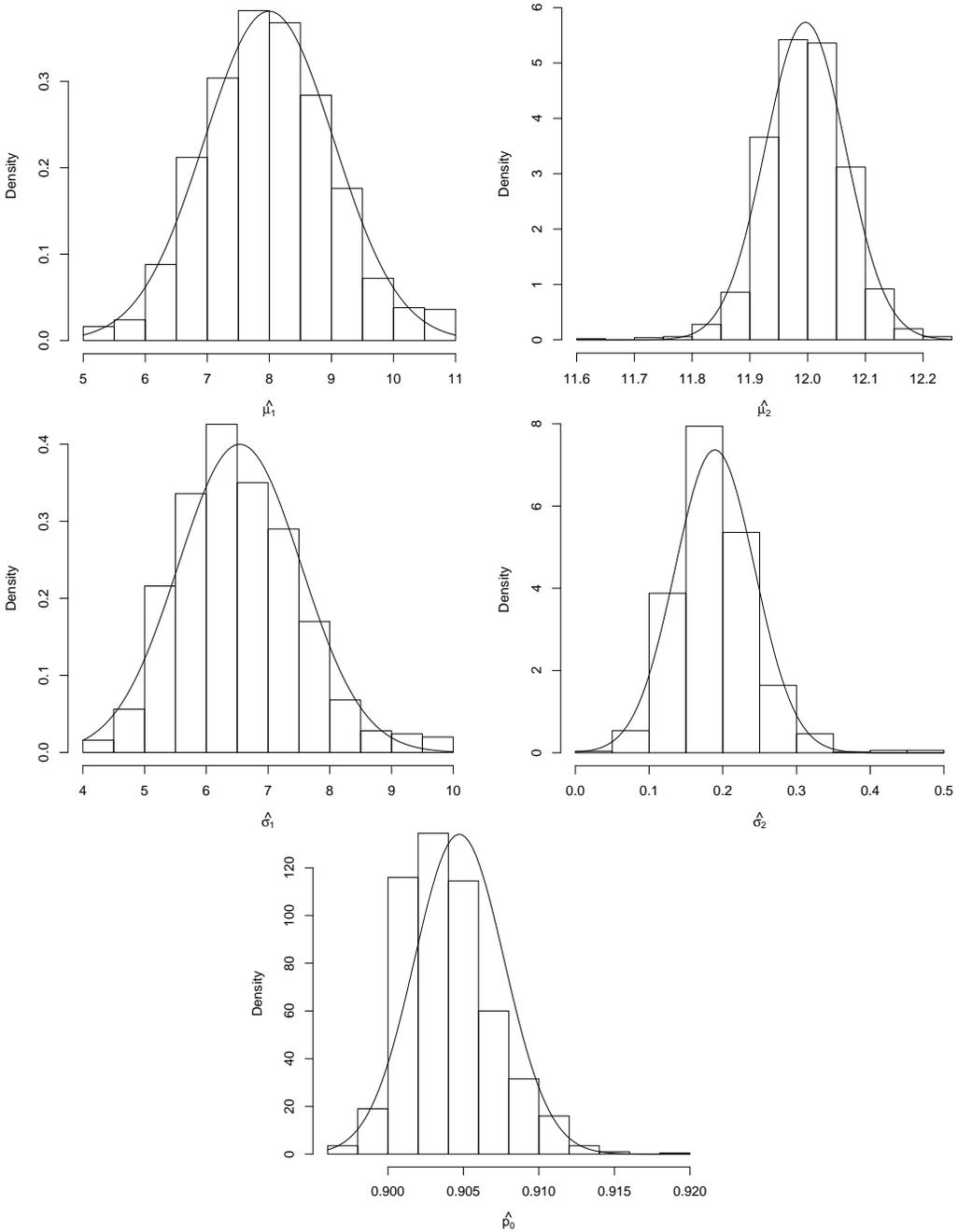
Figure 4: Histogram for the MLEs of μ_1, μ_2, σ_1, σ_2 and p_0 when $n = 500$ and $100(p_0 = 0.9 + 0.05)$.

4 APPLICATION TO THE CIRCUIT BOARD DATA

To verify the performance of the BLN model on real-world problems, we conducted an analysis of the circuit board dataset (initially described in Section 1), using the BLN model as a parametric model appropriated to describe this dataset. In this application, the performance of the BLN model was compared with the performances obtained by other models proposed in the literature to deal with interval-censored data in a cure rate setup, namely: (1) the negative binomial Weibull (NBW) model [25], (2) the COM-Poisson (CP) model [37], and (3) the extended generalized gamma accelerated failure time (EGG-AFT) model [43], whose the respective survival functions are given by follows:

$$S_p(t;\alpha,\theta,\tau,\lambda) = \left\{ 1 + \alpha\theta \left[1 - \exp\{-\left(\frac{t}{\tau}\right)^\lambda\} \right] \right\}^{-\frac{1}{\alpha}}, \ t,\theta,\tau,\lambda > 0, \alpha > -\frac{1}{\theta}, \quad (4.1)$$

$$S_p(t;\eta,\phi,\gamma_1,\gamma_2) = \frac{\sum_{j=0}^{\infty} \frac{\left(\eta \exp\{-(\gamma_2 t)^{\frac{1}{\gamma_1}}\} \right)^j}{(j!)^\phi}}{\sum_{j=0}^{\infty} \frac{\eta^j}{(j!)^\phi}}, \ t,\eta,\phi,\gamma_1,\gamma_2 > 0, \quad (4.2)$$

$$S_p(t;\mu,\sigma,q,p_0) = p_0 + (1-p_0)S_{EGG}(t;\mu,\sigma,q), \ t > 0, \mu,q \in \mathbb{R}, \sigma > 0, \quad (4.3)$$

where $p_0 \in [0,1]$,

$$S_{EGG}(t;\mu,\sigma,q) = \begin{cases} 1 - \frac{1}{\Gamma(q^{-2})} \int_0^{q^{-2}\exp(qv)} x^{q^{-2}-1} \exp(-x)dx & \text{if } q > 0 \\ \frac{1}{\Gamma(q^{-2})} \int_0^{q^{-2}\exp(qv)} x^{q^{-2}-1} \exp(-x)dx & \text{if } q < 0 \\ \int_v^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx & \text{if } q = 0 \end{cases} \quad (4.4)$$

and $v = \frac{\log(t)-\mu}{\sigma}$.

To compare these models, we consider the $\max \log L(\cdot)$ value, the Akaike Information Criterion (AIC), and the Bayesian Information Criterion (BIC). AIC and BIC are respectively defined by $-2\log L(\widehat{\boldsymbol{\vartheta}_g}) + 2\kappa$ and $-2\log L(\widehat{\boldsymbol{\vartheta}_g}) + \kappa\log(n)$, where $\widehat{\boldsymbol{\vartheta}_g}$ is the MLE of the g-model, κ is the number of estimated parameters in the g-model, and n is the sample size. The best model corresponds to the lowest values of $\max \log L(\cdot)$, AIC, and BIC.

Given the above considerations, we have adjusted the BLN, NBW, COM-Poisson and EGG-AFT models using the DE algorithm for parameter estimation. Table 1 presents the $\max \log L(\cdot)$, AIC, and BIC values for all models considered. Based on these results, we can conclude that the BLN model outperforms all the other models in comparison.

Table 1: The $\max \log L(\cdot)$ values and AIC/BIC criteria for all fitted models.

Criterion	BLN	NBW	COM-Poisson	EGG-AFT
$\max \log L(\cdot)$	-727.952	-759.473	-760.005	-759.315
AIC	1465.903	1526.946	1528.010	1526.631
BIC	1498.482	1553.009	1554.073	1552.694

Table 2: MLE and 95% CIs for the parameters of the BLN model.

Parameter	Estimate	95% CI
μ_1	8.679	[6.737 , 10.621]
μ_2	9.018	[8.912 , 9.124]
σ_1	6.399	[4.653 , 8.145]
σ_2	0.181	[0.089 , 0.273
p_0	0.979	[0.975 , 0.983]

Furthermore, in order to assess if the BLN model is appropriate, Figure 5 shows plots of the empirical survival function estimates using the generalized Turnbull's nonparametric estimator proposed by Dehghan and Duchesne [18] against the respective predicted values obtained from the BLN, NBW, COM-Poisson and EGG-AFT models. Clearly, we observe from Figure 5, that the predicted values obtained from the BLN model are the closest to the values of the non-parametric estimator, suggesting that this model give a satisfactory fit to the data, thus going against the results from the Table 1. In a way, this result corroborates our research hypothesis, since the BLN model was proposed to have properties that can adequately describe specific characteristics of the dataset considered in our analysis, namely: interval censoring, presence of long-term survivors and latent competitive risks. We consider the R programming language through the `gte` function of the `gte` package in R [19] to compute the survival function estimates using the generalized Turnbull.
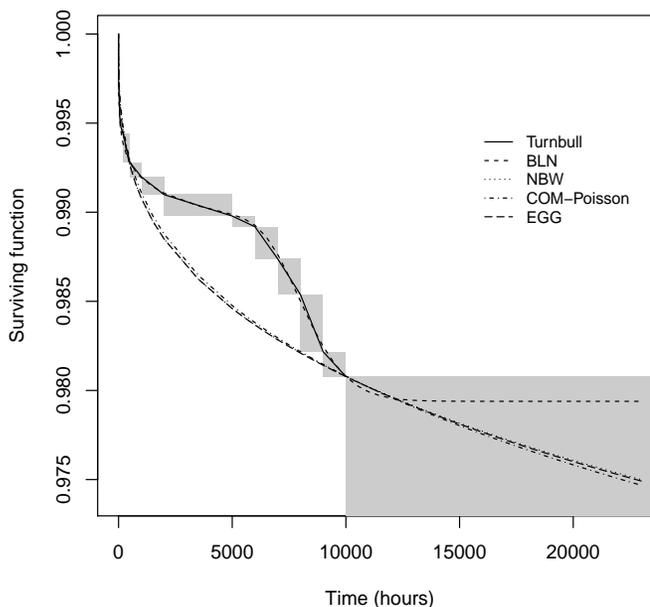

Figure 5: Turnbull estimate against the survival functions estimated by the BLN, NBW, COM-Poisson and EGG-AFT models.

5   CONCLUSIONS

In this paper, based on interval-censored data, we consider maximum likelihood estimation via DE algorithm to estimate the parameters of a parametric mixture cure model, referred to as the bi-lognormal mixture rate model, or simply, the BLN model. The BLN model can be used when the dataset particularly presents the underlying phenomenon of latent competing risks or when there is evidence that a bimodal hazard function is appropriated to described it. This feature makes the BLN model quite flexible in terms of its use to describe real datasets and represents its main advantage over other cure rate models found in the literature. A Monte Carlo simulation study was developed, indicating that the MLE via DE algorithm outperforms the Newton-Raphson method in most cases in terms of bias, root mean square error, and the coverage probability of confidence intervals. Besides, DE algorithm is robust and does not require any other method to choose starting points to obtain good approximations of MLE. The practical importance of the proposed methodology was demonstrated in a real-world interval-censored dataset, where it provided a good fit.

Acknowledgments

REFERENCES

[1] D. Ardia, K. Boudt, P. Carl, K. Mullen & B. Peterson. Differential evolution with DEoptim. *The R Journal*, **3** (2011), 27–34.

[2] N. Balakrishnan & W. Chen. "Lognormal Distributions and Properties. In: Handbook of Tables for Order Statistics from Lognormal Distributions with Applications". Boston, Springer (1999).

[3] J. Balka, A.F. Desmond & P.D. McNicholas. Review and implementation of cure models based on first hitting times for wiener processes. *Lifetime Data Analysis*, **15** (2009), 147–176.

[4] S. Basu, A. Basu & C. Mukhopadhyay. Bayesian analysis for masked system failure data using non-identical Weibull models. *Statistics and Probability Letters*, **78** (1999), 255–275.

[5] J. Berger & D. Sun. Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, **88** (1993), 1412–1418.

[6] J. Berkson & R. Gage. Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, **42** (1952), 501–515.

[7] J. Boag. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *Journal of the Royal Statistical Society B*, **11** (1949), 15–53.

[8] P. Borges. EM algorithm-based likelihood estimation for a generalized Gompertz regression model in presence of survival data with long-term survivors: an application to uterine cervical cancer data. *Journal of Statistical Computation and Simulation*, **87** (2017), 1–11.

[9] P. Borges, J. Rodrigues & N. Balakrishnan. Correlated destructive generalized power series cure rate models and associated inference with an application to a cutaneous melanoma data. *Computational Statistics and Data Analysis*, **56** (2012), 1703–1713.

[10] P. Borges, J. Rodrigues, F. Louzada-Neto & N. Balakrishnan. A cure rate survival model under a hybrid latent activation scheme. *Statistical Methods in Medical Research*, **25** (2016), 838–856.

[11] V. Calsavara, A. Rodrigues, R. Rocha, V. Tomazella & F. Louzada-Neto. Defective regression models for cure rate modeling with interval-censored data. *Biometrical Journal*, **61** (2019), 841–859.

[12] V. Chan & W. Meeker. A failure-time model for infant mortality and wearout failure models. *IEEE Transactions on Reliability*, **48** (1999), 377–387.

[13] M. Chen, J. Ibrahim & D. Sinha. A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association*, **94** (1999), 909–914.

[14] F. Cooner, S. Banerjee, B. Carlin & D. Sinha. Flexible cure rate modelling under latent activation schemes. *Journal of the American Statistical Association*, **102** (2007), 560–572.

[15] D. Cox & D. Hinkley. "Theoretical Statistics". London, Chapman and Hall (1974).

[16] D. Cox & D. Oakes. "Analysis of Survival Data". London, Chapman and Hall (1984).

[17] S. Das & P. Suganthan. Differential evolution: a survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, **3** (2011), 4–31.

[18] M. Dehghan & T. Duchesne. A generalization of turnbull's estimator for nonparametric estimation of the conditional survival function with interval-censored data. *Lifetime Data Analysis*, **17** (2011), 234–255.

[19] M. Dehghan, T. Duchesne & S. Baillargeon. "Package gte: Generalized Turnbull's Estimator" (2015). URL `http://CRAN.R-project.org/package=gte`. R package version 1.2.2, R CRAN package repository.

[20] N. Demiris, D. Lunn & L. Sharples. Survival extrapolation using the poly-Weibull model. *Statistical Methods in Medical Research*, **24** (2015), 287–301.

[21] S. Elsayed, R. Sarker & D. Essam. A self-adaptive combined strategies algorithm for constrained optimization using differential evolution. *Applied Mathematics and Computation*, **241** (2014), 267–282.

[22] J. Fachini, E. Ortega & F. Louzada-Neto. Influence diagnostics for polyhazard models in the presence of covariates. *Statistical Methods and Applications*, **17** (2008), 413–433.

[23] V. Farewell. A model for a binary variable with time censored observations. *Biometrics*, **64** (1977), 43–46.

[24] V. Farewell. The use mixture models for the analysis of survival data with long term survivors. *Biometrics*, **38** (1982), 1041–1046.

[25] E. Hashimoto, E. Ortega & G. Cordeiro. A new long-term survival model with interval-censored data. *Sankhya*, **77** (2015), 207–239.

[26] J. Kalbeisch & R. Prentice. "The Statistical Analysis of Failure Time Data". New York, Wiley (1980).

[27] T. Koutny. Using meta-differential evolution to enhance a calculation of a continuous blood glucose level. *Computer Methods and Programs in Biomedicine*, **133** (2016), 45–54.

[28] L. Kuo & T. Yang. Bayesian reliability modeling for masked system lifetime. *Statistics and Probability Letters*, **47** (2000), 229–241.

[29] E. Lee & J. Wang. "Statistical Methods for Survival Data Analysis, 3rd ed.". New Jersey, John Wiley and Sons (2003).

[30] F. Lobato, V. Machado & V. Steffen Jr. Determination of an optimal control strategy for drug administration in tumor treatment using multi-objective optimization differential evolution. *Computer Methods and Programs in Biomedicine*, **131** (2016), 51–61.

[31] F. Louzada-Neto. Polyhazard regression models for lifetime data. *Biometrics*, **55** (1999), 1281–1285.

[32] F. Louzada-Neto, C. Andrade & F. Almeida. On the non-identifiability problem arising on the poly-Weibull model. *Communications in Statistics - Simulation and Computation*, **33** (2004), 541–552.

[33] R. Maller & X. Zhou. Testing for the presence of immune or cured individuals in censored survival data. *Biometrics*, **51** (1995), 1197–1205.

[34] J. Mazucheli, F. Louzada-Neto & J. Achcar. Bayesian inference for polyhazard models in the presence of covariates. *Computational Statistics and Data Analysis*, **38** (2001), 1–14.

[35] J. Mazucheli, F. Louzada-Neto & J. Achcar. The polysurvival model with long-term survivors. *Brazilian Journal of Probability and Statistics*, **26** (2012), 313–324.

[36] H. Orkcu, E. Aksoy & M. Dogan. Estimating the parameters of 3-p Weibull distribution through differential evolution. *Applied Mathematics and Computation*, **251** (2015), 211–224.

[37] S. Pal & N. Balakrishnan. Likelihood inference for COM-Poisson cure rate model with interval-censored data and Weibull lifetimes. *Statistical Methods in Medical Research*, **26** (2017), 2093–2113.

[38] K. Price, R. Storn & J. Lampinen. "Differential Evolution: A Practical Approach to Global Optimization, 2nd ed.". Berlin, Springer-Verlag (2006).

[39] R Core Team. "R: A Language and Environment for Statistical Computing". R Foundation for Statistical Computing, Vienna, Austria (2022). URL http://www.R-project.org/.

[40] R. Rocha, S. Nadarajah, V. Tomazella & F. Louzada-Neto. A new class of defective models based on the Marshall-Olkin family of distributions for cure rate modeling. *Computational Statistics and Data Analysis*, **107** (2017), 48–63.

[41] J. Rodrigues, V. Cancho, M. Castro & F. Louzada-Neto. On the unification of the long-term survival models. *Statistics and Probability Letters*, **79** (2009), 753–759.

[42] M. Roman, F. Louzada-Neto, V. Cancho & J. Leite. A new long-term survival distribution for cancer data. *Journal of Data Science*, **10** (2012), 241–258.

[43] S. Scolas, A. EL Ghouch, C. Lengrand & A. Oulhaj. Variable selection in a flexible parametric mixture cure model with interval-censored data. *Statistics in Medicine*, **35** (2016), 1210–1225.

[44] R. Storn & K. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, **11** (1997), 341–359.

[45] J. Sun. "The Statistical Analysis of Interval-Censored Failure Time Data". New York, Chapman and Hall (2006).

[46] R. Tsai & L. Hotta. Polyhazard models with dependent causes. *Brazilian Journal of Probability and Statistics*, **27** (2013), 357–376.

[47] R. Tsai & L. Hotta. Polyhazard models with dependent causes, Fitting distributions with the poly-hazard model with dependence. *Communications in Statistics - Theory and Methods*, **44** (2015), 1886–1895.

[48] A. Yakovlev & A. Tsodikov. "Stochastic Models of Tumor Latency and Their Biostatistical Applications". Singapore: World Scientific Publishers (1996).

APPENDIX: R CODE

The following R code can be used to obtain the ML estimates for parameters of the BLN model.

```
library(DEoptim)
##-----------------------------------------------------------------
# Circuit board data
# V. Chan and W.Q. Meeker, A failure-time model for infant mortality
# and wearout failure models,IEEE Trans. Reliab. 35 (2000),
# pp. 550-555
##-----------------------------------------------------------------
y.left <- c(rep(0,10),rep(1,1),rep(2,3),rep(5,1),rep(10,2),rep(20,6),
rep(50,3),rep(100,2),rep(200,8),rep(500,4),rep(1000,5),
rep(2000,6),rep(5000,3),rep(6000,9),rep(7000,10),
rep(8000,16),rep(9000,7),rep(10000,4897))
y.right <- c(rep(1,10),rep(2,1),rep(5,3),rep(10,1),rep(20,2),rep(50,6),
rep(100,3),rep(200,2),rep(500,8),rep(1000,4),rep(2000,5),
rep(5000,6),rep(6000,3),rep(7000,9),rep(8000,10),
rep(9000,16),rep(10000,7),rep(Inf,4897))
status  <- c(rep(1,96),rep(0,4897))
##-----------------------------------------------------------------
# Log-likelihood function
##-----------------------------------------------------------------
log.vero <- function(theta1){
mu2      <- theta1[1]
mu1      <- theta1[2]
sigma2   <- theta1[3]
sigma1   <- theta1[4]
p        <- theta1[5]

lvero    <- sum(ifelse(status==1,log((1-p)*pnorm((-log(y.left)+mu1)/sigma1)
           *pnorm((-log(y.left)+mu2)/sigma2)-(1-p)*pnorm((-log(y.right)+mu1)/
           sigma1)*pnorm((-log(y.right)+mu2)/sigma2)),log(p+(1-p)*pnorm((
           -log(y.left)+mu1)/sigma1)*pnorm((-log(y.left)+mu2)/sigma2))))
return(-lvero)
}


##-----------------------------------------------------------------
# Obtaining the ML estimates

l        <- c(0,0,0,0,0)
u        <- c(20,20,20,20,1)
mle      <- DEoptim(log.vero, lower=l, upper=u, DEoptim.control(NP = 50,
           itermax = 500,trace=FALSE))
summary(mle)
##-----------------------------------------------------------------
```