Journal of Quality Measurement and Analysis *JQMA* 18(3) 2022, 1-15
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*

BATHTUB HAZARD MODEL WITH COVARIATE AND RIGHT CENSORED DATA

(Model *Bathtub Hazard* Berkovariat dengan Data Tertapis Kanan)

IDARI ISMAIL*, JAYANTHI ARASAN, MOHD SHAFIE MUSTAFA & MUHAMMAD ASLAM MOHD SAFARI

ABSTRACT

Lifetime distributions that present bathtub-shaped failure rates are becoming increasingly important especially when dealing with medical data. In this research, a two-parameter bathtub hazard model was extended to incorporate covariates in the presence of right censored data. The parameter estimates were computed based on maximum likelihood estimation (MLE) method. A simulation study was then executed to assess the performance of parameter estimates based on their bias, standard error (SE) and root mean square error (RMSE) at various censoring proportions and sample sizes. The results suggested that the performance of the estimator is better at larger sample size as it lower the standard error and root mean square error (RMSE). A decrease in censoring proportions yield smaller SE and RMSE values, whereas the values of bias decrease as the sample sizes and censoring proportions increase. Finally, the extended model was applied to a real medical data. Also, three confidence interval estimation (Wald, bootstrap-p, bootstrap-t) were obtained for each of the parameters of the model. The results suggested that the real data fitted the bathtub hazard model well.

Keywords: bathtub hazard; covariates; right censored

ABSTRAK

Taburan jangka hayat yang yang mempunyai bentuk fungsi kadar kegagalan seperti tab mandi menjadi kian penting terutamanya apabila melibatkan data perubatan. Dalam kajian ini, model *bathtub hazard* dengan dua parameter dikembangkan dengan menambah kovariat bersama data tertapis kanan. Anggaran parameter diperoleh melalui kaedah penganggaran kebolehjadian maksimum (MLE). Seterusnya, kajian simulasi dibuat untuk menilai prestasi parameter berdasarkan nilai kecondongan, ralat piawai(SE), dan punca min kuasa ralat(RMSE) pada nilai perkadaran tapisan dan saiz sampel yang berbeza. Hasil analisis menunjukkan bahawa prestasi penganggar adalah lebih baik pada saiz sampel yang lebih besar, kerana ianya menghasilkan nilai SE dan RMSE yang lebih rendah. Penyusutan dalam nilai perkadaran tapisan menghasilkan nilai SE dan RMSE yang lebih rendah, namun nilai kecondongan didapati semakin menurun dengan peningkatan saiz sampel dan nilai perkadaran tapisan. Akhir sekali, model yang telah dikembangkan ini diaplikasikan kepada data perubatan yang sebenar. Tiga kaedah ukuran kebolehpercayaan anggaran (Wald, bootstrap-p, bootstrap-t) juga telah didapatkan untuk setiap parameter dalam model tersebut. Keputusan analisis mencadangkan model *bathtub hazard* sesuai diaplikasikan kepada data sebenar.

Kata kunci: *bathtub hazard*; kovariat; tertapis kanan

1. Introduction

Throughout the last decades, a number of research was devoted to the construction of lifetime distribution in modelling real-life data with more than traditional increasing and decreasing failure rates. There are several existing statistical distributions for modelling lifetime data where the distributions are named as “lifetime probability distributions” or simply known as "life distribution" (Yousof *et al*. 2021). Early studies have proposed some parametric probability distribution to analyze real life data with bathtub-shaped rates. Smith and Bain (1975) suggested exponential power distribution and the distribution was further studied by Leemis (1986). Another distribution with four parameter was proposed by Gaver and Acar (1979). Hjorth (1980) investigated similar distribution with increasing, decreasing, or bathtub-shaped failure rate function. Meanwhile, Mudholkar and Srivastava (1993) suggested an exponentiated-Weibull distribution. Furthermore, some researchers have studied a new three-parameter distribution that exhibit bathtub-shape, exponential-type family of distributions (Lemonte 2013); lognormal-power distribution with flexible behavior (Reed 2011); and two-parameter modified Weibull extension (Xie *et al*. 2002).

In this paper, by adding covariate with the presence of right-censored data, we extend a two-parameter bathtub hazard model that was initially proposed by Chen (2000). The cumulative distribution function (cdf) of the distribution is given by

$$F(t) = 1 - e^{\lambda\left(1-e^{t^{\alpha}}\right)}, \; t \geq 0 \tag{1}$$

where $\lambda > 0$ is a parameter that does not affect the shape of the failure rate function and $\alpha > 0$ is defined as the shape parameter. The corresponding survival function can be written as

$$S(t;\lambda,\alpha) = e^{\lambda\left(1-e^{t^{\alpha}}\right)} \tag{2}$$

The probability density function (pdf) is

$$f(t;\lambda,\alpha) = \lambda\alpha t^{\alpha-1} e^{t^{\alpha}} e^{\lambda\left(1-e^{t^{\alpha}}\right)} \tag{3}$$

The corresponding failure rate function is defined as

$$h(t;\lambda,\alpha) = \lambda\alpha t^{\alpha-1} e^{t^{\alpha}} \tag{4}$$

The first derivative of $h(t)$ is given by

$$h'(t;\lambda,\alpha) = \lambda\alpha t^{\alpha-2} e^{t^{\alpha}} \left(\alpha - 1 + \alpha t^{\alpha}\right) \tag{5}$$

There have been some prior works that studied and discussed this bathtub hazard model (Sarhan *et al*. 2012; Wu *et al*. 2004; Wu 2008) among others. The distribution exhibits an increasing failure rate when $\alpha \geq 1$. Also, the model displays a bathtub-shaped failure rates when $\alpha < 1$ which makes it as suitable distribution in modelling real lifetime datasets. In particular, bathtub hazard model provides a good alternative to other existing lifetime distribution due to its flexibility that can model both monotonic and non-monotonic failure rates. Failure rate function is an important characteristic in modelling lifetime data. Increasing failure rate (IFR) and bathtub-shaped are among of interest in many lifetime studies. Unimodal failure rates can be commonly observed in medical studies such as heart or kidney transplantation where the patients tend to have an IFR during an adaptation period while a decreasing failure rate (DFR) can be seen afterward (Lemonte 2013). Bathtub curve can be classified into three phases of hazard: early failure, random failure, and wear-out region (Lemonte 2013; Wang *et al*. 2015). A study of human mortality provides an example of bathtub-shaped failure rate, with a high infant mortality rate at the start, which rapidly fall to a low point and remains steady until accelerating again (Lemonte 2013). Also, bathtub-shaped hazard can be seen in study of animal survival from birth and in the field of reliability such as failure of equipment (Xie *et al*. 2002).

Wu *et al*. (2004) suggested a simple and exact statistical test for the shape parameter in bathtub hazard model due to its vital contribution. Furthermore, the exact confidence interval for the parameter was also discussed. Monte Carlo simulation was applied in order to find the critical values for the statistical test and to construct the confidence interval of the shape parameter. The results obtained from Monte Carlo simulation shows that the average lengths for the confidence intervals for the parameter are shorter compared to the intervals presented by the method in Chen (2000). The new statistical test proposed by Wu *et al*. (2004) was also suggested to be better and more powerful. In another study by Wu (2008), the estimation problem from the two parameters in bathtub hazard model was investigated by analyzing a progressively type-II censored sample using maximum likelihood estimation. In addition, Wu *et al*. (2004) proposed a method for constructing the exact confidence interval and joint confidence region for the parameters. Sarhan *et al*. (2012) discussed and studied the parameter estimation for the distribution. They used maximum likelihood and Bayes method to estimate the unknown parameters of this model. By analyzing real data set from (Lawless 2003), the comparison between the point estimates obtained by both methods was made. Additional work by Chen and Gui (2020) have presented maximum likelihood estimations and confidence intervals for parameters in the model. Similarly, another recent work by Zhang and Gui (2022) discussed maximum likelihood estimations and presented asymptotic confidence interval for the parameters. Overall, there are few works related to bathtub hazard model as discussed above. Yet, these previous works do not extend the model such as incorporating any fixed or time-varying covariates in the model. There were several works involving different models which extended the model to accommodate covariates in the presence of censored data such as Gompertz model with right-censored data (Kiani *et al*. 2012; Maarof *et al*. 2021); Gompertz model with right-, left- and interval-censored data (Kiani & Arasan 2013); log-logistic model with right-censored data (Loh *et al*. 2015); log-logistic model with right- and interval-censored data (Lai & Arasan 2020) and generalized exponential model with interval-censored data (Al-Hakeem *et al*. 2022). To the best of our knowledge, no works has been done on bathtub hazard model specifically to estimate covariates with the presence of censored data such as right- or interval-censored data.

As such, this study aims to extend the bathtub hazard model by incorporating covariate in the context of right-censored data. A simulation study was then conducted to assess the performance of the model by examining the value of bias, standard error (SE), and root mean square error (RMSE) at various sample sizes and censoring proportions(cp). Moreover, an application to a real data set using the extended model was also illustrated. Also, we computed confidence interval estimates for the parameters using Wald and bootstrap method.

2. Methodology

2.1. *Bathtub hazard model with covariate*

The effects of covariates on survival time are incorporated in hazard function of the bathtub hazard model by allowing the parameter λ as a function of covariates (Kiani *et al*., 2012; Maarof *et al*., 2021). The function can be written as

$$\lambda_i = e^{-\beta_0-\beta_1 x_i} \tag{6}$$

Thus, based on hazard function specified in Eq. (4), the hazard function for a data set with a fixed covariate x_i where $i=1,2,...,n$ can be expressed as,

$$h(t_i) = \lambda_i \alpha t^{\alpha-1} e^{t_i^{\alpha}} = e^{-\beta_0-\beta_1 x_i} \alpha t_i^{\alpha-1} e^{t_i^{\alpha}} \tag{7}$$

The vector of parameters for this model is $\theta = (\alpha, \beta_0, \beta_1)$. The estimation of the parameters was done via the method of maximum likelihood.

2.2. *Maximum likelihood estimation*

The idea of construction of actual likelihood function can be found in Hosmer and Lemeshow (1999). It is based on the contribution of the triplet (t,c,x) where the lower case letters in the triplet denote the actual observed values of variables T, C and X respectively. T represents the actual observed time and C is the censoring indicator. In the case of the triplet $(t,1,x)$, $c=1$ indicates that the observed value of t measures the subject's actual survival time (i.e. death from the disease). For this triplet, it is known that the survival time was exactly t. Therefore, the contribution to the likelihood for this triplet is the probability that a subject with covariate value x dies from the disease at time t where this is given by the value of density function. Hence, the overall likelihood of the bathtub hazard model for a sample of n uncensored observations, , $i=1,2,...,n$ can be written as

$$L(\theta) = \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{n} \left\{ e^{(-\beta_0-\beta_1 x_i)} \alpha t_i^{\alpha-1} e^{t_i^{\alpha}} e^{e^{(-\beta_0-\beta_1 x_i)}\left(1-e^{t_i^{\alpha}}\right)} \right\} \tag{8}$$

where t_i is the observed survival time for the ith subject.

To incorporate right censored data to the likelihood function given above, as mentioned earlier, a censoring indicator need to be defined where indicator 0 indicates observation that is right censored, 1 for observations that is not censored. Henceforth, the censoring indicator will be denoted by S. For ith observation, the censoring indicator is described as follows:

$$s_i = \begin{cases} 1, & \text{observation is not censored} \\ 0, & \text{observation is right censored} \end{cases}$$

In the case of the triplet $(t,0,x)$, on the other hand, we know that the survival time was at least t. Consequently, the contribution to the likelihood function of this triplet is the probability that a subject with covariate value x survives for at least t time units where this probability is given by the survival function. Under the independent observation assumptions, to get the full likelihood function, we therefore multiply the contribution of the triplets, a value of density function for an uncensored observation and a value of survival function for censored observation as given below,

$$\{f(t_i)\}^{S_i}\{S(t_i)\}^{1-S_i} \tag{9}$$

Since the observations are assumed to be independent, thus the likelihood function for a sample of n observations, $i = 1, 2, ..., n$, is the product of the expression given in Eq. (9),

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n}\{f(t_i)\}^{S_i}\{S(t_i)\}^{1-S_i} \\ &= \prod_{i=1}^{n}\left\{\left[e^{(-\beta_0-\beta_1 x_i)}\alpha t_i^{\alpha-1}e^{t_i^{\alpha}}e^{e^{(-\beta_0-\beta_1 x_i)}(1-e^{t_i^{\alpha}})}\right]\right\}^{S_i} \\ &\quad \mathrm{X}\left\{\left[e^{e^{(-\beta_0-\beta_1 x_i)}(1-e^{t_i^{\alpha}})}\right]\right\}^{1-S_i} \end{aligned} \tag{10}$$

and the log-likelihood function is,

$$\begin{aligned} \ell(\theta) &= \sum_{i=1}^{n} S_i \ln\left(e^{(-\beta_0-\beta_1 x_i)}\right) + \ln(\alpha) + \ln(t_i^{\alpha-1}) + \ln(e^{t_i^{\alpha}}) + \ln\left(e^{e^{(-\beta_0-\beta_1 x_i)(1-e^{t_i^{\alpha}})}}\right) \\ &\quad + (1-S_i)\ln\left(e^{e^{(-\beta_0-\beta_1 x_i)(1-e^{t_i^{\alpha}})}}\right) \\ &= \sum_{i=1}^{n} S_i\left[(-\beta_0-\beta_1 x_i) + \ln(\alpha) + (\alpha-1)\ln(t_i) + t_i^{\alpha} + e^{(-\beta_0-\beta_1 x_i)}(1-e^{t_i^{\alpha}})\right] \\ &\quad + (1-S_i)\left[e^{(-\beta_0-\beta_1 x_i)}(1-e^{t_i^{\alpha}})\right] \end{aligned} \tag{11}$$

The observed information matrix $[i(\hat{\beta}_0, \hat{\beta}_1, \hat{\alpha})]^{-1}$ can be found from second partial derivative of the log-likelihood function in Eq. (10) evaluated at maximum likelihood estimates where it gives us the estimated for variance and covariance. The Newton-Raphson iterative approach was used to obtain the MLE of the parameters.

2.3. *Confidence interval estimates*

This section provides a brief discussion on three methods of confidence interval estimates that were used in this study. The first method is asymptotic normality confidence interval or known as Wald interval. The Wald confidence interval is the most widely used interval technique in survival analysis. Bootstrapping techniques, on the other hand, have been proven to be more robust to model assumption violations. Bootstrap resampling methods have known as powerful tools in constructing inferential procedures (Léger *et al*. 1992). Resampling is the process of repeated sampling from the original data set. In this research, two common bootstrap confidence interval estimation known as bootstrap-p (B-p) and bootstrap-t (B-t) were employed.

2.3.1. *Asymptotic confidence interval (Wald)*

Let $\hat{\theta}$ be the maximum likelihood estimator for the vector of the parameters θ and the log-likelihood function of θ is denoted by $l(\theta)$. $\hat{\theta}$ is known to be asymptotically normally distributed with mean θ and covariance matrix $I^{-1}(\theta)$ where $l(\theta)$ is the Fisher information matrix that can be estimated by the observed information matrix $I(\hat{\theta})$. Hence, the 100(1-α)% confidence interval for θ is given as follows

$$\hat{\theta}_j \pm z_{(1-\alpha/2)}\sqrt{i^{-1}(\hat{\theta}_{jj})}$$

2.3.2. *Bootstrap-p confidence interval*

Given a data set $x=(x_1,x_2,...,x_n)$ and let $\hat{\theta}$ be the MLE of θ. To begin with, B bootstrap samples, x^b for $b=1,2,...,B$ will be generated. Then, we obtain the bootstrap version of MLE, $\hat{\theta}^*_B$ for each of the bootstrap samples. The 100(1-α) % confidence interval for θ is given by

$$\left\{\hat{\theta}^*_{[l]},\hat{\theta}^*_{[u]}\right\} \quad (13)$$

where $l = B$ x $\left(1\text{-}\alpha/2\right)$ and $u = B$ x $\alpha/2$.

The B-p method is a simple method that generates confidence intervals directly from the percentiles of the bootstrap distribution of estimated parameters (Arasan & Adam 2014).

2.3.3. *Bootstrap-t confidence interval*

This method involves more work than the *B-p* interval as it requires the standard error of an estimate. Suppose we have *B* bootstrap samples $x^*_1,...,x^*_B$, and $\hat{\theta}$ is the MLE of θ . For each bootstrap sample, the bootstrap version of its MLE, $\hat{\theta}^*_b$ is computed. Following that, we compute z^*_b for each bootstrap samples.

$$Z^*_b = \frac{\hat{\theta}^*_b - \hat{\theta}}{\hat{\sigma}_b}$$

where $\hat{\sigma}_b$ is the standard error for bth bootstrap sample.

Then, we sort Z^*_b in the form of ascending order to obtain $Z^*_{[K]}$ where K=1,2,...,B. The 100(1-α)% confidence interval for θ can be obtained by

$$\left\{\hat{\theta}-Z^{*}_{[l]}se\left(\hat{\theta}\right),\hat{\theta}-Z^{*}_{[u]}se\left(\hat{\theta}\right)\right\}$$

where $l = B$ x $\left(1\text{-}\alpha/2\right)$ and $u = B$ x $\alpha/2$.

3. Results

3.1 *Simulation study*

A simulation study was conducted to assess the performance of bathtub hazard model described in previous section. The simulation was done using 1000 replications with various sample sizes $n = 50, 100, 250$. The covariate values were simulated from the standard normal distribution. We mimicked survival data by setting the parameters β_0, β_1, α to 3.3, 0.9, and 0.4 respectively. In order to generate lifetimes t_i, a sequence of random numbers, u_i were drawn from uniform distribution. Furthermore, the censoring time denoted as c_i was drawn from exponential distribution with parameter μ in which the value of μ would be set and adjusted to produce the desired approximate cp. The survival time is classified as censored when $t_i > c_i$ while $t_i \le c_i$ indicates uncensored. The survival time, t_i was obtained by

$$t_i = \left(\ln\left(1 - \frac{\ln\left(1-U_i\right)}{e^{\left(-\beta_0-\beta_1 x_i\right)}} \right) \right)^{1/\alpha} \tag{16}$$

To evaluate the performance of the estimators at various sample sizes and censoring proportions, bias, SE and root RMSE were obtained by

$$bias = E(\hat{\theta}) - \theta, \tag{17}$$

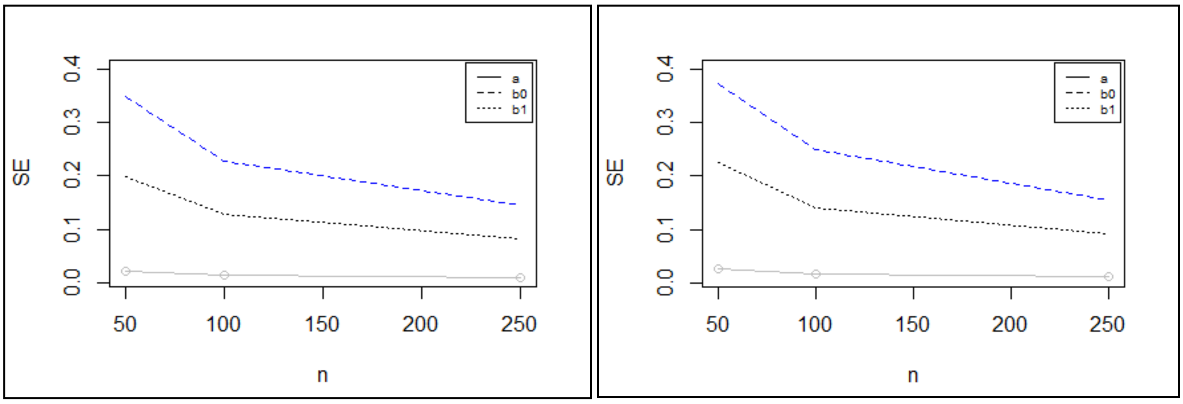
$$SE = \sqrt{E\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2}, \tag{18}$$

$$RMSE = \sqrt{\left(SE^2 + bias^2\right)} \tag{19}$$

Table 1 presents the results of the estimated bias, SE and RMSE based on the simulated data. It is clear from Table 1 that the value of bias, SE and RMSE increases as the CP and sample size increase. This demonstrates that higher sample sizes and smaller censoring proportions resulted in better estimates. These bias, SE ad RMSE values from the simulation study are small enough to show that the simulated data are generated by a stable simulation process. Likewise, as seen in Figure 1 and Figure 3, the values of SE and RMSE for each estimate decrease as the sample size increase. The SE and RMSE values, on the other hand, are shown to increase as the censoring proportion increases (see Figure 2 and Figure 4).

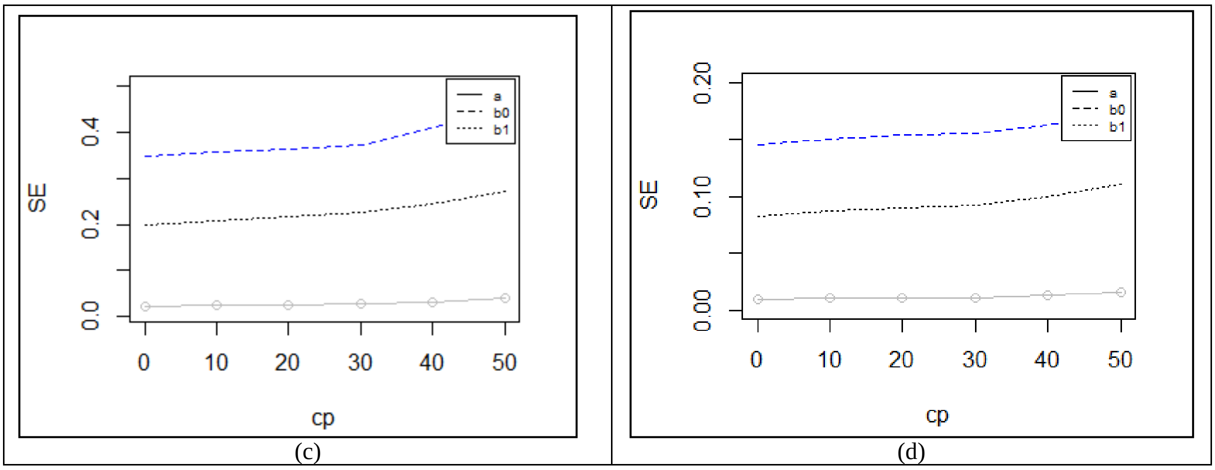
Table 1: Bias, SE, and RMSE of the estimates for bathtub hazard model with covariate

		$\hat{\alpha}$			$\hat{\beta}_0$			$\hat{\beta}_1$		
	CP	n=50	n=100	n=250	n=50	n=100	n=250	n=50	n=100	n=250
Bias	0	0.00706	0.00411	0.00133	0.07889	0.05278	0.01321	0.04488	0.02244	0.00750
	10	0.00701	0.00416	0.00147	0.07716	0.05333	0.01430	0.04620	0.02382	0.00947
	20	0.00751	0.00416	0.00181	0.08046	0.05406	0.01815	0.05072	0.02721	0.01130
	30	0.00773	0.00431	0.00190	0.08194	0.05527	0.01830	0.05206	0.02900	0.01216
	40	0.00981	0.00491	0.00220	0.09911	0.05981	0.02194	0.05883	0.03350	0.01463
	50	0.01121	0.00512	0.00305	0.11585	0.06275	0.02872	0.07012	0.03524	0.01792
SE	0	0.02350	0.01562	0.00996	0.34866	0.22819	0.14521	0.19834	0.12862	0.08212
	10	0.02462	0.01637	0.01063	0.35562	0.23396	0.15009	0.20766	0.13393	0.08693
	20	0.02605	0.01755	0.01126	0.36310	0.24244	0.15428	0.21765	0.13936	0.09051
	30	0.02733	0.01855	0.01154	0.37202	0.24842	0.15575	0.22573	0.14186	0.09270
	40	0.03249	0.02011	0.01306	0.40683	0.25839	0.16319	0.24369	0.15065	0.09924
	50	0.03948	0.02243	0.01596	0.44718	0.27184	0.17649	0.27053	0.16362	0.11075
RMSE	0	0.02454	0.01615	0.01005	0.35747	0.23421	0.14581	0.20335	0.13056	0.08246
	10	0.02560	0.01689	0.01073	0.36390	0.23996	0.15077	0.21273	0.13603	0.08745
	20	0.02711	0.01804	0.01140	0.37191	0.24839	0.15534	0.22348	0.14199	0.09122
	30	0.02840	0.01904	0.01170	0.38904	0.25449	0.15682	0.23165	0.14480	0.09350
	40	0.03394	0.02070	0.01324	0.41873	0.26522	0.16466	0.25069	0.15433	0.10031
	50	0.04104	0.02301	0.01625	0.46194	0.27899	0.17882	0.27947	0.16737	0.11219



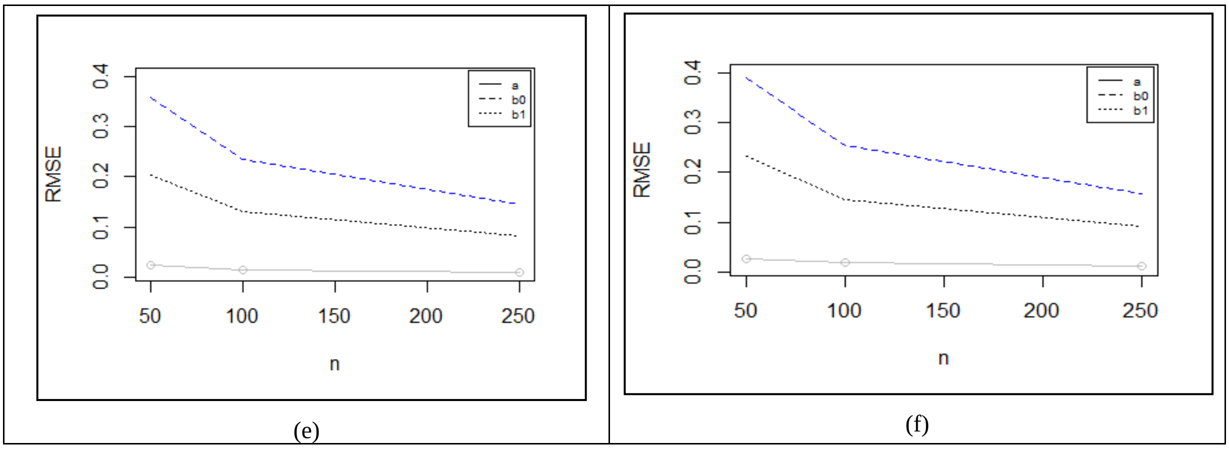
(a) (b)

Figure 1: Line plot of SE at various sample sizes for (a) cp=0 and (b) cp=30



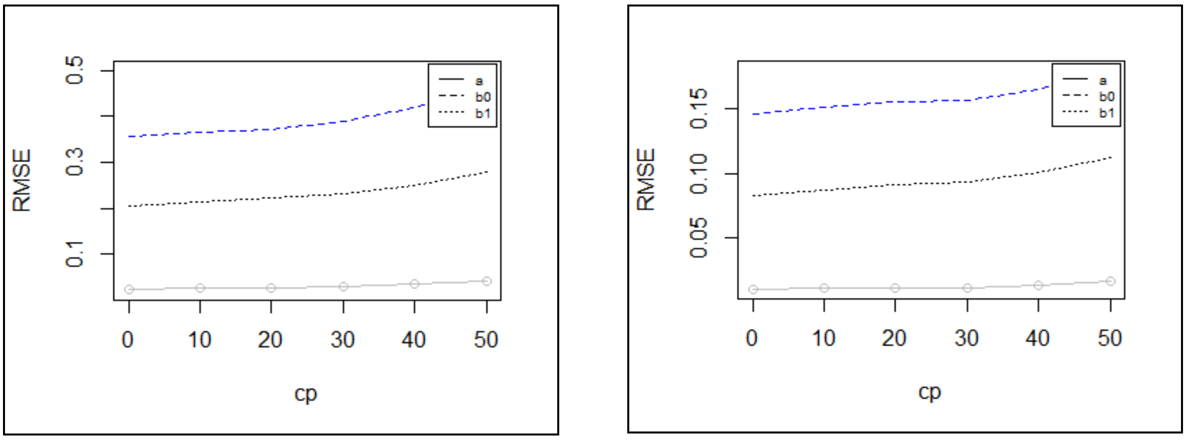
(c) (d)

Figure 2: Line plot of SE at different censoring proportions for (c) *n*=50 and (d) *n*=250



(e) (f)

Figure 3: Line plot of RMSE at various sample sizes for (e) CP=0 and (f) CP=30



(g) (h)

Figure 4: Line plot of RMSE at different censoring proportions for (g) n=50 and (h) n=250

3.2 *Real data analysis*

For illustration purpose, we analyze a real data set to show the application of the bathtub hazard model. The real data set represents the survival time (in days) of a random sample of 227 advanced lung cancer patient where the event of interest is time to death with patients censored if the event did not occur. There were 137 male and 90 female patients, with censoring indicators of 0 and 1 indicating that the patient was censored and dead respectively (27.8% of the patients was right censored). Previously, the range of simulation study included a sample size of 250 and censoring proportion of 30% to ensure the reliability of estimates obtained in this real data analysis. A non-parametric Kaplan-Meier survival curve and $S(t)$ based on the bathtub hazard model were plotted on the same graph. As depicted in the Figure 5, it can be said that the bathtub hazard model is appropriate for the advanced lung cancer data.

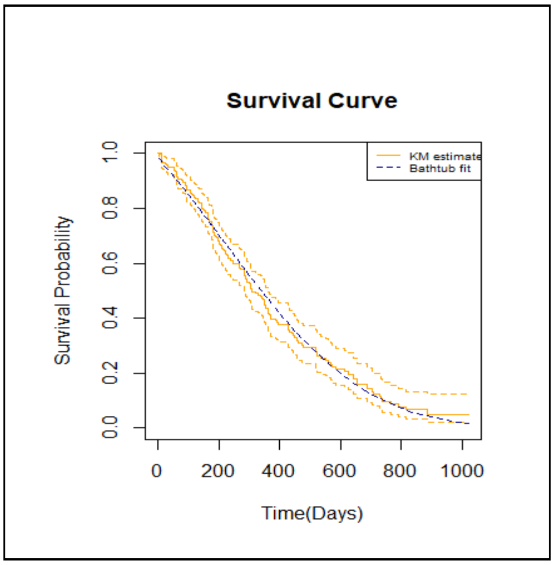


Figure 5: Survival Curve

Table 2 summarizes the descriptive statistics for uncensored observation where the mean value is greater than the median value, indicating that the survival time distribution for advanced lung cancer patients is positively skewed. This can also be seen from the histogram in Figure 6, which shows that the distribution is skewed to the right, which is in accordance with the conclusion drawn based on the mean and median values.

Table 2: Descriptive statistics of survival time(uncensored observation)

Mean	284.29
Median	227.5
Standard deviation	202.74
Standard error	15.83
Skewness	0.86
Kurtosis	-0.01

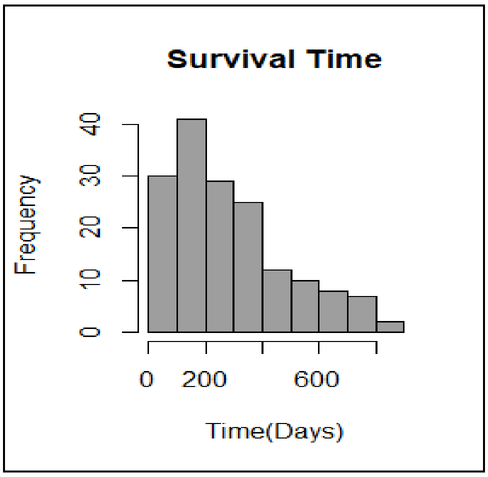


Figure 6: Histogram of survival time

Bathtub hazard model without covariate was then generated and the maximum likelihood estimates with the corresponding standard error given in parentheses for the parameters of the model are summarized in Table 3.

Table 3: Summary of maximum likelihood estimates without covariates

Parameter	Estimates (Standard Error)
λ	0.0048(0.0011)
α	0.2751(0.0066)

Further, bathtub hazard model with covariate was obtained since the present study interested in incorporating covariates. Descriptive statistics of survival time for uncensored observation according to gender was obtained as given in Table 4. It is clear from the table that female patients have a longer mean survival time than male patients, implying that females with advanced lung cancer can survive longer than males.

Table 4: Descriptive statistics of survival time by gender

Gender	n	Mean	Standard error	Standard deviation	Median
Male	111	264.3	18.8	198.1	210
Female	53	326.1	28.6	207.9	293

In a similar manner, as shown in Figure 8, males exhibited higher hazard rates compared with females.

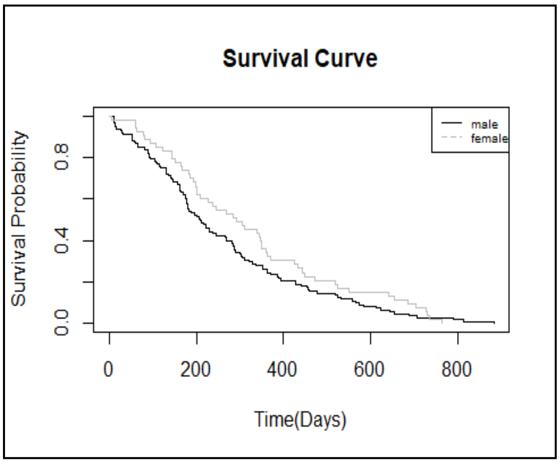


Figure 7: Survival Curve (gender)

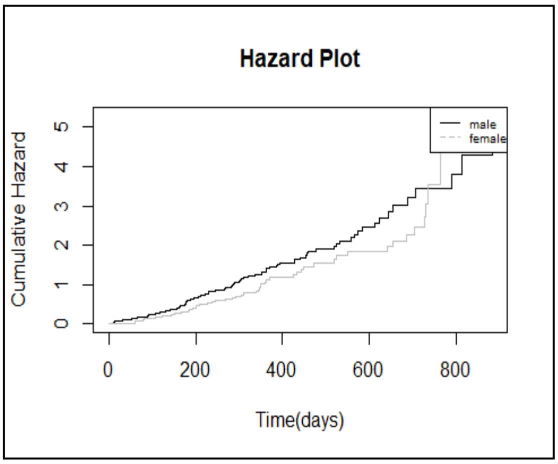


Figure 8: Hazard plot (gender)

The summary of maximum likelihood estimates for the parameters in bathtub hazard model with covariates is shown in Table 5.

Table 5: Summary of maximum likelihood estimates with covariates

Parameters	Estimate (Standard error)	t value	Pr(>t)
β_0	4.6466(0.3194)	14.55	0.000
β_1	0.5027(0.1665)	3.02	0.0025
α	0.2755(0.0066)	41.88	0.000

Table 6 presents the confidence interval estimates for each of the parameter in the model which were computed using three estimation methods.

Table 6: 95% confidence interval of β_0, β_1 and α

Parameters	Wald	B-p	B-t
β_0	(4.0207,5.2726)	(4.1150,5.2623)	(4.1150,5.2623)
β_1	(0.1764, 0.8291)	(0.1553,0.8566)	(0.1625,0.8529)
α	(0.2626,0.2883)	(0.2632, 0.2903)	(0.2604, 0.2873)

The effect of covariate (gender) on survival time for advanced lung cancer patients can be determined based on the confidence interval estimates. It is noteworthy that the all the interval estimates for parameter β_1 do not include 0 indicating that gender has a significant effect on the survival time of advanced lung cancer patients. The significance of the covariate can also be concluded based on the p-value. Here, we test the null hypothesis $H_0 : \beta_1 = 0$ against the alternative hypothesis $H_1 : \beta_1 \neq 0$. Referring to Table 5, the p-value for β_1 is less than alpha value of 0.05 which lead to the decision of rejecting the null hypothesis, implying that gender does give a significant impact on the survival time. The findings are in line with those of Siddiqui *et al*. (2010), who found that gender is the most important factor in influencing survival time among nonoperative non-small cell lung cancer patients . In addition, Elkbuli *et al*. (2020) indicated that women had better survival for lung cancer after controlling for other covariates compared to men which is in line with this study as seen in Figure 8.

4. Conclusion

This study extended a two-parameter bathtub hazard model by incorporating covariates in the presence of right-censored data. The estimation of the parameters is approached by the MLE method. Also, this study set out to assess the performance of parameter estimates based on their bias, SE and RMSE at various censoring proportions and sample sizes. For this purpose, a simulation study was executed. The results of the simulation study, which are presented in a summary table and illustrated in line plots, show that as the censoring proportion increases, the SE and RMSE values increase. Furthermore, the results revealed that that the larger the sample size, the lower the SE and RMSE. The SE and RMSE values increase as the censoring proportion increase and sample size decrease, demonstrating that smaller censoring proportions and larger sample sizes yield better estimates. For the real data analysis, the bathtub hazard distribution provides a good fit to real data sets of advanced lung cancer. Based on the preceding discussion of the results, it was also concluded that the gender has a significant effect on survival time of advanced lung cancer patient.

In this study, three confidence interval estimation methods were obtained. Thus, other confidence interval estimation methods such as likelihood ratio, jackknife or other bootstrap methods could be constructed in future study and the performance of the estimation methods could be assessed using a coverage probability study. This study only focused on extending the model in the presence of right-censored data. Therefore, future study could consider estimating the parameters in the context of interval-censored data. Moreover, the proposed model could be extended to include more covariates as well as time-varying covariates. Also, an application to a real data set from different fields other than medical might be examined to further investigate the flexibility of bathtub hazard model.

Acknowledgments

We would like to thank the Putra Grant, Vot 9595300, Universiti Putra Malaysia for supporting this research project.

References

Al-Hakeem H.A., Arasan J., Mustafa M.S. & Peng L.F. 2022. Generalized exponential distribution with interval-censored data and time dependent covariate. *Communications in Statistics - Simulation and Computation*: 1–11.

Arasan J. & Adam M.B. 2014. Double bootstrap confidence interval estimates with censored and truncated data. *Journal of Modern Applied Statistical Methods* **13**(2): 399–419.

Chen S. & Gui W. 2020. Statistical analysis of a lifetime distribution with a bathtub-shaped failure rate function under adaptive progressive type-II censoring. *Mathematics* **8**(5): 670.

Chen Z. 2000. A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics and Probability Letters* **49**(2): 155–161.

Elkbuli A., Byrne M.M., Zhao W., Sutherland M., McKenney M., Godinez Y., Dave D.J., Bouzoubaa L. & Koru-Sengul T. 2020. Gender disparities in lung cancer survival from an enriched Florida population-based cancer registry. *Annals of Medicine and Surgery* **60**: 680–685.

Gaver D.P. & Acar M. 1979. Analytical hazard representations for use in reliability, mortality, and simulation studies. *Communications in Statistics - Simulation and Computation* **8**(2): 91–111.

Hjorth U. 1980. A reliability distribution with increasing, decreasing, constant and bathtub-shaped failure rates. *Technometrics* **22**(1): 99–107.

Hosmer D.W & Lemeshow S. 1999. *Applied Survival Analysis, Regression Modeling of Time-To-Event Data*. NY: John Wiley & Sons.

Kiani K. & Arasan J. 2013. Gompertz model with time-dependent covariate in the presence of interval-, right- and left-censored data. *Journal of Statistical Computation and Simulation* **83**(8): 1472–1490.

Kiani K., Arasan J. & Midi H. 2012. Interval estimations for parameters of Gompertz model with time-dependent covariate and right censored data. *Sains Malaysiana* **41**(4): 471–480.

Lai M. & Arasan J. 2020. Single covariate log-logistic model adequacy with right and interval censored data. *Journal of Quality Measurement and Analysis* **16**(2): 131–140.

Lawless J.F. 2003. *Statistical Models and Methods for Lifetime Data*. 2nd Ed. NY: John Wiley & Sons, Inc.

Leemis L.M. 1986. Relationships among common univariate distributions. *American Statistician* **40**(2): 143–146.

Léger C., Politis D. N. & Romano J. P. 1992. Bootstrap Technology and Applications. *Technometrics* **34**(4): 378–398.

Lemonte A.J. 2013. A new exponential-type distribution with constant, decreasing, increasing, upside-down bathtub and bathtub-shaped failure rate function. *Computational Statistics and Data Analysis* **62**: 149–170.

Loh Y.F., Arasan J., Midi H. & Abu Bakar M.R. 2015. Inferential procedures based on the double bootstrap for log logistic regression model with censored data. *Malaysian Journal of Science* **34**(2): 199–207.

Maarof N.N.N.A., Arasan J., Zulkafli H.S. & Adam M.B. 2021. Assessing the adequacy of the gompertz regression model in the presence of right censored data. *Journal of Physics: Conference Series* **1988**(1): 1–8.

Mudholkar G.S. & Srivastava D.K. 1993. Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability* **42**(2): 299–302.

Reed W.J. 2011. A flexible parametric survival model which allows a bathtub-shaped hazard rate function. *Journal of Applied Statistics* **38**(8): 1665–1680.

Sarhan A.M., Hamilton D.C. & Smith B. 2012. Parameter estimation for a two-parameter bathtub-shaped lifetime distribution. *Applied Mathematical Modelling* **36**(11): 5380–5392.

Siddiqui F., Bae K., Langer C.J., Coyne J.C., Gamerman V., Komaki R., Choy H., Curran W.J., Watkins-Bruner D. & Movsas B. 2010. The influence of gender, race, and marital status on survival in lung cancer patients: Analysis of radiation therapy oncology group trials. *Journal of Thoracic Oncology* **5**(5): 631–639.

Smith R.M. & Bain L.J. 1975. An exponential power life-testing distribution. *Communications in Statistics* **4**(5): 469–481.

Wang X., Yu C., & Li Y. 2015. A new finite interval lifetime distribution model for fitting bathtub-shaped failure rate curve. *Mathematical Problems in Engineering* **2015**(1).

Wu J.W., Lu H.L., Chen C.H. & Wu C.H. 2004. Statistical inference about the shape parameter of the new two-parameter bathtub-shaped lifetime distribution. *Quality and Reliability Engineering International* **20**(6): 607–616.

Wu S.J. 2008. Estimation of the two-parameter bathtub-shaped lifetime distribution with progressive censoring. *Journal of Applied Statistics* **35**(10): 1139–1150.

Xie M., Tang Y. & Goh T.N. 2002. A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering and System Safety* **76**(3): 279–285.

Yousof H.M., Korkmaz M. & Sen S. 2021. A new two-parameter lifetime model. *Annals of Data Science* **8**(1): 91–106.

Zhang Z. & Gui W. 2022. Statistical analysis of the lifetime distribution with bathtub-shaped hazard function under lagged-effect step-stress model. *Mathematics* **10**(5).

Department of Mathematics and Statistics
Faculty of Science
Universiti Putra Malaysia
43400 UPM Serdang
Selangor DE, MALAYSIA
E-mail: idarismail@gmail.com, jayanthi@upm.edu.my, mshafie@upm.edu.my, aslam.safari@upm.edu.my*

Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA Cawangan Kelantan
Kampus Machang, Bukit Ilmu,
18500 Machang
Kelantan DN, MALAYSIA
*E-mail : idarismail@gmail.com**

Received: 22 August 2022
Accepted: 2 November 2022

*Corresponding author