Journal of Quality Measurement and Analysis *JQMA* 20(2) 2024, 89-103
e-ISSN: 2600-8602 *http://www.ukm.my/jqma*
https://doi.org/10.17576/jqma.2002.2024.07

INTERVAL ESTIMATION FOR PARAMETERS OF A BATHTUB HAZARD MODEL

(Penganggaran Selang untuk Parameter Model Hazad Tab Mandi)

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

ABSTRACT

In this study, a two-parameter lifetime model has been extended to incorporate covariate in the presence of right-censored data. The model has bathtub-shaped or increasing failure rate function which enables it to fit real lifetime data set. The method of maximum likelihood was used to estimate the parameters in the model and a simulation study was then conducted to evaluate the performance of parameter estimates at various sample sizes and censoring proportion levels. The results from simulation study show that larger sample sizes and smaller censoring proportion give better estimates. Further, two interval estimation methods: Wald and likelihood ratio were constructed, and the performance of these methods was evaluated based on a coverage probability study. Both Wald and likelihood ratio techniques appear to have better performance when the sample size is larger. Also, a real right-censored lifetime data on patients with multiple myeloma was employed to illustrate the practical application of the extended model.

Keywords: bathtub-shaped; interval estimation; likelihood ratio; Wald; coverage probability study.

ABSTRAK

Dalam kajian ini, model jangka hayat yang mempunyai dua parameter telah dikembangkan dengan memasukkan kovariat dengan data yang tertapis kanan. Model tersebut mempunyai fungsi kadar kegagalan yang menaik ataupun berbentuk seperti tab mandi yang membolehkan model ini disesuaikan dengan data jangka hayat yang sebenar. Kaedah kebolehjadian maksimum digunakan untuk membuat anggaran parameter di dalam model ini dan seterusnya kajian simulasi dijalankan untuk menilai prestasi anggaran parameter pada beberapa saiz sampel dan kadar tapisan yang berlainan. Hasil kajian simulasi itu menunjukkan saiz sampel yang besar dan kadar tapisan yang kecil menghasilkan anggaran yang lebih baik. Selanjutnya, dua kaedah anggaran selang: Wald dan likelihood ratio telah dibina dan prestasi setiap kaedah tersebut dinilai melalui kajian kebarangkalian liputan. Kedua-dua kaedah Wald dan likelihood ratio dilihat mempunyai prestasi yang lebih baik apabila saiz sampel lebih besar. Data sebenar iaitu data jangka hayat tertapis kanan yang melibatkan pesakit myeloma pelbagai juga digunakan untuk menunjukkan aplikasi model yang telah dikembangkan ini secara praktiknya.

Kata kunci: berbentuk tab mandi; selang keyakinan; butstrap; likelihood ratio; Wald; kaedah kebarangkalian liputan.

1. Introduction

In reliability analysis, failure rate or hazard function is crucial in studying the lifetime of a product. Lifetime distribution can be classified into five categories based on its failure rates; constant, decreasing, increasing, bathtub and upside-down bathtub shaped (Maurya *et al*.

2020). The well-known lifetime models such as Gamma or Weibull are those that can accommodate monotone failure rates, however, there are several situations in engineering or medical fields where the failure rates firstly decrease, then stagnant at a constant level and eventually increase. Such failure rate resembles a bathtub, and thus, it is known as bathtub-shaped failure rate. It is possible to observe the bathtub-shaped failure rate when studying the lifespan of an industrial product or the lifetime of a biological entity (Dimitrakopoulou *et al*. 2007). In recent decades, there were several studies that examined the distribution with bathtub-shaped failure rates, including those by Mudholkar and Srivastava (1993) and Smith and Bain (1975) among others. The Smith-Bain model was found to provide a satisfactory fit to the dataset on 500 MW generators discussed in Dhillon (1981), as demonstrated by Paranjape *et al*. (1985). Subsequently, Paranjpe and Rajarshi (1986) applied the Smith-Bain exponential power model to dataset on bird populations discussed in Deevey (1947). This illustrates the applicability of the lifetime distributions with bathtub-shaped failure rates across various fields.

In this paper, the distribution that was proposed by Chen (2000) for which the failure rate exhibits a bathtub-shape depending on its parameter (written as bathtub hazard model throughout this paper). Additionally, the failure rate can be monotonically increasing, demonstrating the flexibility of this distribution. Chen (2000) also discussed exact confidence intervals and joint confidence regions for the parameters in the model based on type-II censoring. Chen (2000) further stated that there is no two-parameter distribution that the failure rate exhibit bathtub-shaped. Given its useful properties, a number of studies have investigated and studied the bathtub hazard model. For example, Wu *et al*. (2004) proposed a simple method for conducting statistical test with regards to the shape parameter where the method can be applied for a type-II right-censored data. Wu (2008) discussed exact confidence interval and exact join confidence region for the parameters in the bathtub hazard model under progressive type-II censored sample. Based on type-II censored sample, Wang *et al*. (2014) discussed interval estimations for the parameters in bathtub hazard model. Sarhan *et al*. (2012) examined parameter estimation of the bathtub hazard model by using maximum likelihood and Bayes method. Recently, additional work by Sarhan and Mustafa (2022) developed a new lifetime distribution based on bathtub hazard model and generalized exponential distribution. They discussed the parameter estimation using maximum likelihood method and Bayesian procedures. Another recent work by Zhang and Gui (2022) and Chen and Gui (2020) also discussed parameter estimation of bathtub hazard model and presented confidence intervals for the model's parameters.

In summary, extensive research has been undertaken to study the bathtub hazard model. However, the research to date has not focused on investigating the bathtub hazard model by expanding the model with fixed or time-dependent covariate. Therefore, in this work, we conduct a study to extend the bathtub hazard model by incorporating fixed covariates with the presence of right-censored data. In the earlier work, Ismail *et al*. (2022) examined the bathtub hazard model, assessing its parameter performance through a simulation study. The study also constructed Wald and bootstrap confidence intervals using real data on lung cancer with right-censored observations. This present study expands the previous work by exploring an additional common interval estimation method based on likelihood ratio (LR) and further compare the performance of this method with the Wald method via a coverage probability study. A real dataset on multiple myeloma with right-censored observation is employed to demonstrate the applicability of the bathtub hazard model.

2. Methodology

Bathtub hazard model with random variable T which denoting the lifetimes has the following distribution function:

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

with two positive parameters, where $\lambda > 0$ is a parameter that does not affect the shape of failure rate and γ is defined as the shape parameter.

The corresponding probability density function (pdf) and the survival function are respectively given by,

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

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

Accordingly, the failure rate function is given by,

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

The failure rate function becomes bathtub-like as $\gamma < 1$ and is increasing when $\gamma \geq 1$ (Chen 2000).

Let the parameter λ be a function of the covariates to incorporate the effects of covariates on survival time in failure rate of the bathtub hazard model. Thus, we can express the function as

$$\lambda_i = e^{-\beta_0-\beta_1 x_i} \tag{5}$$

Henceforth, the failure rate function for a data set with a fixed covariate x_i where $i = 1, 2, ..., n$ can be written as,

$$h(t_i) = \lambda_i\gamma t^{\gamma-1}e^{t_i^{\gamma}} = e^{-\beta_0-\beta_1 x_i}\gamma t_i^{\gamma-1}e^{t_i^{\gamma}} \tag{6}$$

Let $\theta = (\gamma, \beta_0, \beta_1)$ be the vector of the model parameters. Maximum likelihood method was used to estimate the three parameters. The likelihood function for the full sample if there are no censored observations can be written as

$$L(\theta) = \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{n}\left\{e^{\left(-\beta_0-\beta_1 x_i\right)}\gamma t_i^{\gamma-1}e^{t_i^{\gamma}}e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_i^{\gamma}}\right)}\right\} \tag{7}$$

In order to incorporate right-censored data to the likelihood function, a censoring indicator need to be defined. Hence, for the i^{th} observation, the censoring indicator denoted by S is given by,

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

Suppose t_i is the observed survival time for the i^{th} subject, the likelihood function for uncensored and right-censored is,

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \{f(t_i)\}^{S_i} \{S(t_i)\}^{1-S_i} \\ &= \prod_{i=1}^{n} \left\{ \left[e^{\left(-\beta_0-\beta_1 x_i\right)} \gamma t_i^{\gamma-1} e^{t_i^{\gamma}} e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_i^{\gamma}}\right)} \right] \right\}^{S_i} \times \left\{ \left[e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_i^{\gamma}}\right)} \right] \right\}^{1-S_i} \end{aligned} \tag{8}$$

and the associated log-likelihood function becomes

$$\begin{aligned} \ell(\theta) = \sum_{i=1}^{n} S_i & \left[\left(-\beta_0-\beta_1 x_i\right) + \ln(\gamma) + (\gamma-1)\ln(t_i) + t_i^{\gamma} + e^{\left(-\beta_0-\beta_1 x_i\right)}(1-e^{t_i^{\gamma}}) \right] \\ & + (1-S_i) \left[e^{\left(-\beta_0-\beta_1 x_i\right)}(1-e^{t_i^{\gamma}}) \right] \end{aligned} \tag{9}$$

The maximum likelihood estimators of all parameters can be obtained by setting the first derivative of the log-likelihood function equal to zero. However, explicit equations cannot be provided due to the complex and nonlinear forms of the equation. Therefore, a numerical known as Newton-Raphson method can be used to compute approximate estimates of parameters.

3. Confidence Interval Estimation

In this section, we discuss two techniques (Wald and likelihood ratio) that were used to construct confidence interval for each parameter in the two-parameter bathtub hazard model with fixed covariate in the presence of right-censored data.

3.1. *Asymptotic interval estimation (Wald method)*

Let $\hat{\theta}$ be the maximum likelihood estimator for the vector of parameters θ and $\ell(\theta)$ is the log-likelihood function of θ. Under mild regularity conditions, $\hat{\theta}$ is asymptotically normally distributed with mean θ and covariance matrix $I^{-}(\theta)$ where $I(\theta)$ is defined as the Fisher information matrix, evaluated at the true value of the parameter (Cox and Hinkley 1974). The matrix $I(\theta)$ can be estimated by the observed information matrix $I(\hat{\theta})$ whose

$(j,k)^{th}$ element can be found from the second partial derivatives of the log-likelihood function evaluated at $\hat{\theta}$.

If $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution, then the asymptotic $100(1-\alpha)\%$ confidence interval for θ_j

$$\hat{\theta}_j \pm z_{(1-\alpha/2)}\sqrt{I^{-1}(\hat{\theta}_{jj})} \tag{10}$$

3.2. *Likelihood ratio method*

For a parameter of interest θ, the likelihood ratio statistic in testing the null hypothesis $H_0:\theta=\theta_0$ versus the alternative, $H_1:\theta\neq\theta_0$ is given as follows:

$$\Psi = -2[l(\theta_0,\tilde{\eta}) - l(\hat{\theta},\hat{\eta})] \tag{11}$$

where l is the log-likelihood function, η is known as the vector of nuisance parameters, $(\hat{\theta},\hat{\eta})$ is the MLE of (θ,n), and $\tilde{\eta}$ is the restricted maximum likelihood estimator of η under H_0. For a large sample size, Ψ is approximately $\chi^2_{(1)}$ under H_0 and a $100(1-\alpha)\%$ confidence interval for θ can be determined by finding a set of two values of θ_0, for which H_0 is not rejected at the specified significance level, that is, the values that satisfy:

$$l(\theta_0,\tilde{\eta}) = \ell(\hat{\theta},\hat{\eta}) - \tfrac{1}{2}\chi^2(1;1-\alpha), \tag{12}$$

with the lower confidence limit, $\theta_L<\theta$ and the upper confidence limit $\theta_U>\theta$.

4. Simulation Study

A simulation study using $N=1000$ replications with five different sample sizes, was conducted to assess the performance of the parameters of the bathtub hazard model with covariate and right-censored data. Also, the simulation study was carried out at different level of censoring proportion (cp). The values of the parameters γ,β_0 and β_1 were particularly set at 0.4, 3.3. and 0.9, respectively. The R programming language was used to conduct the simulation study. In this study, we apply the following simulation algorithm:

(1) Generate covariate values x_i from a standard normal distribution.
(2) Generate a sequence of random numbers u_i from a standard uniform distribution on the unit interval (0,1) to obtain the event times, t_i for $i=1,2,...,n$.
(3) Generate censoring times, c_i from an exponential distribution with the value of μ would be modified to obtain the desired censoring proportion (cp).
(4) Generate survival time t_i by:

$$t_i = \left(\ln\left(1 - \frac{\ln(1-U_i)}{e^{(-\beta_0-\beta_1 x_i)}}\right)\right)^{1/\gamma} \tag{13}$$

Table 1-3 presents the bias, SE and RMSE values for all the parameter estimates compared at various sample sizes and different levels of censoring proportion. The bias values decrease when the sample size is increased. Meanwhile, as censoring proportion increases, the bias values are also increasing.

As depicted in Table 2, the value of SE decreases with increasing sample size. When the censoring proportion increases, the SE value also increases. Similar results can be observed for the RMSE values, see Table 3. This indicates better estimates can be obtained at higher sample size and smaller censoring proportion.

Table 1: Summary of bias values for the parameter estimates at different n and cp

Estimates	n	cp=0%	cp=10%	cp=20%	cp=30%	cp=40%	cp=50%
$\hat{\gamma}$	50	0.00704	0.00710	0.00744	0.00848	0.00959	0.01005
	100	0.00336	0.00373	0.00388	0.00458	0.00464	0.00545
	150	0.00165	0.00204	0.00276	0.00299	0.00300	0.003370
	200	0.00144	0.00160	0.00195	0.00216	0.00198	0.00242
	250	0.00134	0.00159	0.00171	0.00174	0.00189	0.00238
$\hat{\beta}_0$	50	0.08089	0.08215	0.08419	0.09462	0.10769	0.11272
	100	0.03534	0.04004	0.04162	0.04749	0.04838	0.05646
	150	0.01735	0.01994	0.02619	0.02809	0.02855	0.03248
	200	0.01327	0.01432	0.01743	0.01971	0.01961	0.02444
	250	0.01188	0.01507	0.01662	0.01736	0.01802	0.01945
$\hat{\beta}_1$	50	0.04374	0.04437	0.04638	0.05105	0.05729	0.06238
	100	0.01544	0.01666	0.01761	0.02136	0.02186	0.02187
	150	0.00926	0.01178	0.01457	0.01669	0.01941	0.02269
	200	0.00724	0.00651	0.00924	0.00997	0.00999	0.01283
	250	0.00434	0.00622	0.00747	0.00782	0.00695	0.00926

Table 2: Summary of standard error (SE) values for the parameter estimates at different n and cp

Estimates	n	cp=0%	cp=10%	cp=20%	cp=30%	cp=40%	cp=50%
$\hat{\gamma}$	50	0.02373	0.02431	0.02688	0.03011	0.03459	0.03862
	100	0.01661	0.01812	0.01880	0.02079	0.02141	0.02418
	150	0.01278	0.01402	0.01628	0.01799	0.01980	0.02254
	200	0.01123	0.01209	0.01360	0.01502	0.01618	0.01921
	250	0.00991	0.01060	0.01205	0.01375	0.01472	0.01685
$\hat{\beta}_0$	50	0.35728	0.36198	0.38342	0.40408	0.43248	0.45666
	100	0.23872	0.25247	0.25742	0.27133	0.27572	0.29219
	150	0.18881	0.19776	0.21201	0.22233	0.23138	0.24467
	200	0.16253	0.16940	0.18023	0.18878	0.19467	0.21154
	250	0.14692	0.15176	0.16222	0.17269	0.17709	0.18630
$\hat{\beta}_1$	50	0.20150	0.20247	0.21532	0.41501	0.25190	0.26988
	100	0.13423	0.14329	0.14612	0.15548	0.15948	0.17408
	150	0.09998	0.10632	0.11384	0.12149	0.13011	0.13907
	200	0.09391	0.09800	0.10615	0.11336	0.11696	0.12805
	250	0.07826	0.08226	0.08857	0.09599	0.09937	0.10945

Table 3: Summary of root mean square error (RMSE) values for the parameter estimates at different n and cp

Estimates	n	cp=0%	cp=10	cp=20	cp=30	cp=40	cp=50%
	50	0.02474	0.02532	0.02789	0.05105	0.03589	0.03990
	100	0.01694	0.01850	0.01919	0.02129	0.02190	0.02479
$\hat{\gamma}$	150	0.01289	0.01417	0.01651	0.01824	0.02003	0.02279
	200	0.01132	0.01212	0.01374	0.01518	0.01630	0.01936
	250	0.01000	0.01072	0.01217	0.01386	0.01484	0.01700
	50	0.36633	0.37118	0.39255	0.22987	0.44568	0.47037
	100	0.24132	0.25563	0.26076	0.27546	0.27993	0.29760
$\hat{\beta}_0$	150	0.18960	0.19877	0.21362	0.22410	0.23313	0.24681
	200	0.16307	0.17000	0.18107	0.18981	0.19565	0.21295
	250	0.14740	0.15251	0.16307	0.17356	0.17800	0.18730
	50	0.20620	0.20727	0.22026	0.23547	0.25833	0.27699
	100	0.13511	0.14426	0.14718	0.15694	0.16097	0.17545
$\hat{\beta}_1$	150	0.10041	0.10697	0.11477	0.12263	0.13155	0.14091
	200	0.09419	0.09821	0.10655	0.11380	0.11739	0.12869
	250	0.07839	0.08250	0.08888	0.09630	0.09961	0.10984

5. Coverage Probability Study

A coverage probability study was carried out to assess and compare the performance of Wald, and likelihood ratio (LR) confidence interval using $N = 1000$ replication of 7 different sample sizes $n = 30, 40, 50, 100, 150, 200,$ and 250. In addition, three levels of censoring proportion were used. The study was conducted at two nominal probabilities, $\alpha = 0.05$ and 0.1. A coverage probability error of a confidence interval is the probability that the interval range contains the true value of parameter. The estimated error probabilities on the left and right are obtained by adding the number of times for the left (right) endpoint that is more (less) that the true parameter value divided by the total number of samples. Thus, the total error probability is merely the total of left and right error probabilities. As given in Doganaksoy and Schmee (1993), the standard error of the estimated coverage error probability is $se(\hat{\alpha}) = [\alpha(1-\alpha)/N]^{1/2}$. Following that, the interval is called anti-conservative (AC) if total error is greater than $\alpha + 2.58se(\hat{\alpha})$, conservative (C) if the total error probability is smaller than $\alpha - 2.58se(\hat{\alpha})$. If the larger error probability (left or right) is more than 1.5 times the smaller one, then the interval is asymmetrical (AS). The overall performance of different confidence interval method is evaluated based on total numbers of anti-conservative, conservative and asymmetrical intervals. In coverage probability study, a good confidence interval is considered if it has the least number of anti-conservative, conservative and asymmetrical intervals. Also, the value of left and error probabilities are closer to 0.025 (0.05) and the value of total error probabilities is closer to nominal probability 0.05 (0.1). Table 4 and 5 display performance of interval estimation techniques at nominal error probability of 0.05 and 0.1, respectively.

Table 4: Performance of interval estimation techniques at different *n* and cp for $\alpha = 0.05$

		cp=0%			cp=5%			cp=10%		
n	Technique	AC	C	AS	AC	C	AS	AC	C	AS
30	Wald	1	0	2	1	0	2	1	0	1
	LR	3	0	1	3	0	1	3	0	1
40	Wald	0	0	3	0	0	3	0	0	3
	LR	2	0	1	2	0	1	2	0	1
50	Wald	0	0	2	1	0	3	0	0	3
	LR	2	0	1	2	0	1	2	0	1
100	Wald	0	0	1	0	0	1	1	0	2
	LR	1	0	1	1	0	1	1	0	1
150	Wald	0	0	2	0	0	2	0	0	1
	LR	1	0	1	1	0	1	1	0	1
200	Wald	0	0	2	0	0	1	0	0	1
	LR	1	0	1	1	0	1	0	0	1
250	Wald	0	0	3	0	0	2	0	0	3
	LR	1	0	1	1	0	1	1	0	1

Table 5: Performance of interval estimation techniques at different *n* and cp for $\alpha = 0.1$

		cp=0%			cp=5%			cp=10%		
n	Technique	AC	C	AS	AC	C	AS	AC	C	AS
30	Wald	1	0	2	1	0	2	1	0	3
	LR	3	0	1	3	0	1	3	0	1
40	Wald	0	0	3	0	0	3	0	0	3
	LR	3	0	1	3	0	1	3	0	1
50	Wald	0	0	1	0	0	2	0	0	3
	LR	2	0	1	2	0	1	2	0	1
100	Wald	0	0	1	0	0	1	0	0	1
	LR	2	0	1	2	0	1	2	0	1
150	Wald	0	0	1	0	0	1	0	0	1
	LR	1	0	1	1	0	1	1	0	1
200	Wald	0	0	1	0	0	1	0	0	1
	LR	2	0	1	1	0	1	1	0	1
250	Wald	0	0	1	0	0	1	0	0	2
	LR	0	0	1	2	0	1	2	0	1

More summarized results according to each level of censoring proportion and sample size are given Table 6 and 7.

Table 6: Summary of performance for each interval estimation techniques at different cp

		Wald			LR		
α	cp(%)	AC	C	AS	AC	C	AS
0.05	0	1	0	15	11	0	7
	5	2	0	14	11	0	7
	10	2	0	14	10	0	7
0.1	0	1	0	10	13	0	7
	5	1	0	11	14	0	7
	10	1	0	14	14	0	7

Referring to Table 6, the Wald method tends to generate more asymmetrical intervals across all levels of censoring proportion and both nominal probabilities, while also yielding few anti-conservative intervals. Conversely, the LR method produces many anti-conservative intervals at all censoring proportion levels and nominal probabilities.

From Table 7, the Wald method works better when $n > 50$ at both nominal levels, starting to produce none of anti-conservative as the sample size increases. The Wald method demonstrates better performance at a nominal probability of 0.01, as indicated by the decrease in asymmetrical intervals generated when $n > 50$. In the same manner, LR method produces fewer anti-conservative intervals when $n > 50$ at both nominal probabilities. The LR method persistently yields asymmetrical intervals, irrespective of the sample size, for both nominal probabilities.

Table 7: Summary of performance for each interval estimation techniques at different *n*

		Wald			LR		
α	n	AC	C	AS	AC	C	AS
	30	3	0	5	9	0	3
	40	1	0	9	6	0	3
	50	1	0	8	6	0	3
0.05	100	0	0	5	3	0	3
	150	0	0	5	3	0	3
	200	0	0	4	2	0	3
	250	0	0	8	2	0	6
	30	3	0	7	9	0	3
	40	0	0	9	9	0	3
	50	0	0	6	6	0	3
0.1	100	0	0	3	6	0	3
	150	0	0	3	3	0	3
	200	0	0	3	4	0	3
	250	0	0	4	4	0	3

Table 8: Summary of performance of Wald method for all parameters at $\alpha = 0.05$

		cp=0			cp=5%			cp=10%		
α	n	AC	C	AS	AC	C	AS	AC	C	AS
	30	*		*	*		*	*		*
	40			*			*	*		*
	50			*	*		*			*
γ	100			*			*			*
	150			*			*			*
	200			*			*			*
	250			*			*			*
	30									
	40			*			*			*
	50			*			*			*
β_0	100									
	150			*			*			
	200									
	250			*			*			*
	30			*			*			
	40			*			*			*
	50						*			*
β_1	100									*
	150									
	200			*						
	250			*						*

Depicted in Table 8 and 9 are the performance of the interval estimation techniques according to each parameter for nominal probability of 0.05. The asterisk (*) symbol is used to illustrate the occurrence of anticonservative, conservative or asymmetrical intervals at different level of censoring proportion and sample sizes for each parameter. The Wald method consistently produce asymmetrical intervals for parameter γ at all censoring proportion levels, see Table 8. Meanwhile, for both β_0 and β_1, there are fewer asymmetrical intervals generated. There are few anti-conservative intervals observed for parameter γ at each censoring proportion levels.

As can be seen in Table 9, LR method also consistently produces asymmetrical intervals for parameter γ at all censoring proportion levels and sample sizes. The occurrence of anti-conservative intervals can be seen for parameter γ at n=30 across all censoring proportion levels. In contrast, for parameter β_0 and β_1, no asymmetrical intervals are generated across all sample sizes and censoring proportion levels. The presence of anti-conservative intervals is observed for both parameters at specific censoring proportion levels and sample sizes.

Further aspects can be examined which is by looking at the graphical display of estimated error probabilities in Figure 1-3. For γ, Wald and LR show a better performance when the sample size is larger since the estimated left and right error probabilities are getting closer to $\alpha/2$. For both β_0 and β_1, Wald and LR method also perform better as sample size is getting larger. In the case of LR method, the two lines representing left and right error probabilities for both parameter β_0 and β_1 coincide on the line plot due to the identical error values produced by the method. For β_0 and β_1, the error probabilities associated with the Wald method appear to be slightly smaller.

Table 9: Summary of performance of LR method for all parameters at $\alpha = 0.05$

		cp=0			cp=5%			cp=10%		
α	n	AC	C	AS	AC	C	AS	AC	C	AS
γ	30	*		*	*		*	*		*
	40			*			*			*
	50			*			*			*
	100			*			*			*
	150			*			*			*
	200			*			*			*
	250			*			*			*
β_0	30	*			*			*		
	40	*			*			*		
	50	*			*			*		
	100									
	150	*			*			*		
	200									
	250	*			*			*		
β_1	30	*			*			*		
	40	*			*			*		
	50	*			*			*		
	100	*			*			*		
	150									
	200	*			*					
	250									

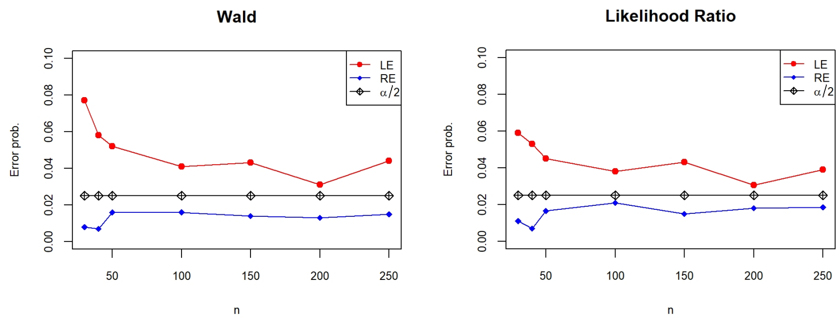

Figure 1: Estimated error probabilities for parameter γ at cp=10% and $\alpha = 0.05$

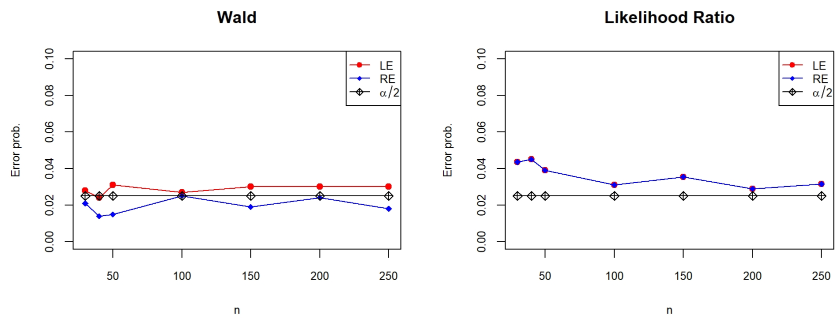

Figure 2 : Estimated error probabilities for parameter β_0 at cp=10% and $\alpha = 0.05$

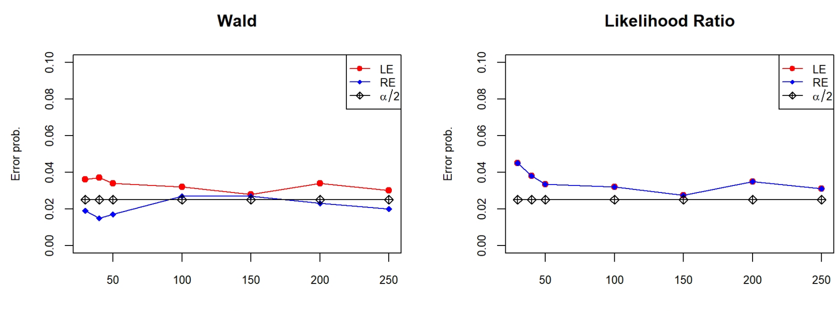

Figure 3: Estimated error probabilities for parameter β_1 at cp=10% and $\alpha = 0.05$

6. Real Data Analysis

In this section, a myeloma data was employed for the illustrative purpose. The data have been obtained from Collet (2003) which was analyzed by Krall *et al*. (1975). The data represent the survival time (in months) of 48 patients who aged between 50 and 80 years with multiple myeloma. Multiple myeloma is a type of blood cancer that can be characterized by the abnormal plasma cells, a type of white blood cell found in the bone marrow. In multiple myeloma, these abnormal plasma cells can crowd out the healthy blood cells in the bone marrow. While advances in treatment have improved outcomes for many patients, the disease can still lead to death. In the dataset, 11 (22.9%) patients were still alive at the end of the study contributing to right-censored survival times. Figure 4 illustrates the survival probabilities obtained using a non-parametric Kaplan-Meier (solid line) and based on the bathtub hazard model (blue dotted line).

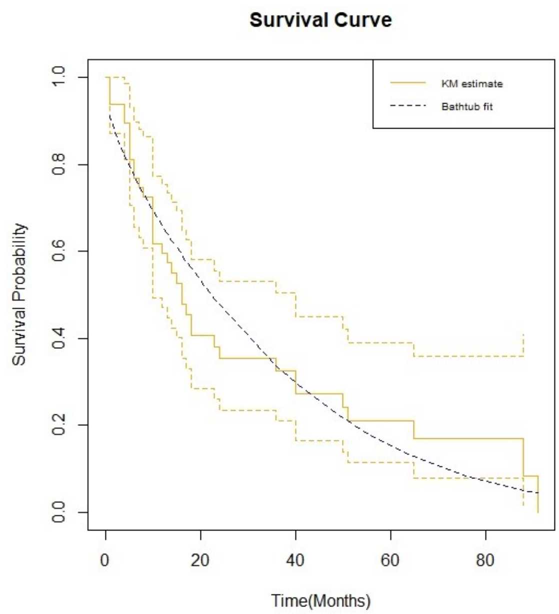

Figure 4: Plot of estimated survival probabilities based on the fitted distribution and Kaplan-Meier estimator

In Figure 4, the bathtub estimates based on the bathtub hazard model appear to be roughly close to Kaplan-Meier estimates suggesting the bathtub hazard distribution is appropriate for the respective myeloma data.

Next, we obtained the bathtub hazard model with covariate, to examine the impact of gender on survival time of patients with multiple myeloma. Based on the results presented in Table 10, gender does not exhibit significant impact on the survival time of patients with multiple myeloma (p-value=0.346). Figure 5 complements the findings by providing insight into the comparison of survival rates in multiple myeloma patients based on gender. As depicted in the figure, the slight difference in survival rates between male patients (represented by red solid line) and female patients (indicated by the green dotted line) supports the previous discussion on the non-significance of covariate gender, as evidenced by the results in Table 10. According to Pasvolsky *et al*. (2023), the incidence of multiple myeloma is slightly higher in men compared to women, with rates of 8.8 versus 5.7 new cases per 100,000 persons per year, respectively. However, the influence of gender on the outcomes of multiple myeloma patients remains uncertain and unknown (Derman *et al*. 2021).

Table 10: MLE of multiple myeloma data

Parameters	Estimates (Standard Error)	t value	p-value
γ	0.3137(0.0220)	14.241	$< 2\times 10^{-16}$
β_0	3.3688(0.5938)	5.673	0.0000
β_1	-0.3061(0.3375)	-0.907	0.346

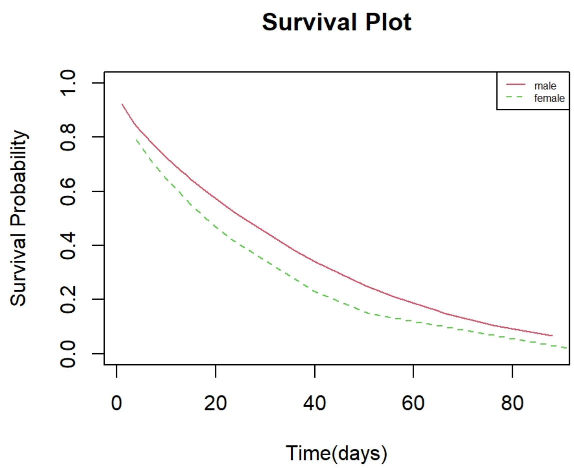

Figure 5: Survival plot (Gender)

Apart from the discussion above, confidence interval estimation method can also be employed to assess the significance of covariate (gender) on survival time for multiple myeloma patients. Table 11 shows the 95% confidence intervals using the Wald method. It is apparent that the confidence interval estimates for parameter β_1 contains 0, implying that gender do not significantly impact the survival time of multiple myeloma patients.

Table 11: 95% confidence intervals

Parameters	Wald
γ	(0.2704, 0.3569)
β_0	(2.2050, 4.5326)
β_1	(-0.9676,0.3554)

7. Conclusion

In this study, we extend a two-parameter bathtub hazard model by incorporating fixed covariate in the presence of right-censored data. The results from simulation study indicates that the value of bias, SE and RMSE increase with the increase in censoring proportion and decrease in sample size. Also, this study achieved one of its objectives of assessing the performance of two confidence interval estimation methods via a coverage probability study. Based on the preceding findings discussed in previous section, the Wald method starts to perform well when $n > 50$. This finding corroborates the work done by (Kiani *et al.* 2012; Loh *et al.* 2015). The Wald method generates quite a number of asymmetrical intervals at all censoring proportion levels. As discussed in Doganaksoy and Schmee (1993), Wald is known to be highly asymmetrical as compared to LR especially when dealing with censored data. On the other hand, LR method yields several anti-conservative intervals at all censoring proportions and sample sizes. Particularly, the occurrence of anti-conservative intervals produced by LR method decreases as sample size exceeds 50. In the comparison of Wald and LR methods, Wald method slightly outperforms LR method at greater sample size as no anti-conservative intervals were produced and lesser asymmetrical intervals specifically at nominal probability of 0.01. Arasan (2009) found that the LR method outperforms the Wald method when handling censored data. However, the present study utilized low censoring proportions of 0%, 5%, and 10%, potentially limiting the ability to assess the performance of LR method with higher censoring proportions. Future research is suggested to explore data with more complicated structures such as doubly censored data and higher censoring proportions. Application of the extended bathtub hazard model to the multiple myeloma data set has shown that the model fits the data set well. Also, we obtained bathtub hazard model with covariate in order to examine the significance of covariate (gender) on survival time of multiple myeloma patients.

References

Arasan, J. 2009. Interval Estimation for Parameters of a Bivariate Time Varying Covariate Model. *Pertanika J. Sci. & Technol*. **17**(2): 313–323.

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):1-21.

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.

Collet D. 2003. *Modelling Survival Data in Medical Research*. New York: Chapman & Hall/CRC.

Cox D.R. & Hinkley D.V. 1974. *Theoretical Statistics*. New York: Chapman & Hall/CRC.

Derman, B. A., Langerman, S. S., Maric, M., Jakubowiak, A., Zhang, W., & Chiu, B. C. 2021. Sex differences in outcomes in multiple myeloma. *British Journal of Haematology* **192**(3): e66–e69.

Dhillon, B. S. 1981. Life Distributions. IEEE Transactions on *Reliability* **R-30**(5): 457–460.

Dimitrakopoulou, T., Adamidis, K. & Loukas, S. 2007. A lifetime distribution with an upside-down bathtub-shaped hazard function. *IEEE Transactions on Reliability* **56**(2):308–311.

Doganaksoy, N. & Schmee, J. 1993. Comparisons of Approximate Confidence Intervals for Distributions Used in Life-Data Analysis. *Technometrics* **35**(2):175–184.

Deevey, E. S. (1947). Life Tables for Natural Populations of Animals. *The Quarterly Review of Biology* **22**(4): 283–314.

Efron, B. 1981a. Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods. *Biometrika* **68**(3): 589–599.

Efron, B. 1981b. Nonparametric Standard Errors and Confidence Intervals. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* **9**(2): 139–158.

Efron, B. 1985. Bootstrap Confidence Intervals for a Class of Parametric Problems. *Biometrika* 72(1):45–58.

Efron, B. & Tibshirani, R. J. 1994. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Ismail, I., Arasan, J., Mustafa, M. S., Aslam, M., & Safari, M. 2022. Bathtub Hazard Model With Covariate and Right. *Journal of Quality Measurement and Analysis* **18**(3): 1–15.

Jeng, S.-L. & Meeker, W. Q. 2000. Comparisons of Approximate Confidence Interval Procedures for Type I Censored Data. *Technometrics* **42**(2): 135–148.

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.

Krall, J. M. Uthoff, V. A. & Harley, J. B. 1975. A step-up procedure for selecting variables associated with survival. *Biometrics* **31**(1): 49–57.

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.

Manoharan, T., Arasan, J., Midi, H. & Adam, M. B. 2015. A Coverage Probability on the Parameters of the Log-Normal Distribution in the Presence of Left-Truncated and Right- Censored Survival Data. *Malaysian Journal of Mathematical Sciences* **9**(1): 127–144.

Maurya, S. K., Singh, S. K. & Singh, U. (2020). A New Right-Skewed Upside Down Bathtub Shaped Heavy-tailed Distribution and its Applications. *Journal of Modern Applied Statistical Methods* **19**(1): 1–23.
Mudholkar, G. S. & Srivastava, D. K. 1993. Exponentiated Weibull Family for Analyzing Bathtub Failure-Rate Data. IEEE Transactions on Reliability 42(2): 299–302.
Paranjape, S., Rajarshi, M. B., & Gore, A. P. 1985. On a Model for Hazard Rates. *Biometrical Journal* **27**(8): 913–917.
Paranjape, S. & Rajarshi, M. B. 1986. Modelling Non-Monotonic Survivorship Data With Bathtub Distributions. *Ecology* **67**(6): 1693–1695.
Pasvolsky, O., Saliba, R. M., Masood, A., Mohamedi, A. H., Tanner, M. R., Bashir, Q., Srour, S., Saini, N., Ramdial, J., Nieto, Y., Lee, H. C., Patel, K. K., Kebriaei, P., Thomas, S. K., Weber, D. M., Orlowski, R. Z., Shpall, E. J., Champlin, R. E., & Qazilbash, M. H. 2023. Impact of gender on outcomes of patients with multiple myeloma undergoing autologous Haematopoietic stem cell transplant. British Journal of *Haematology* **201**(4): e37–e41.
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.
Sarhan, A. M. & Mustafa, A. 2022. A new extension of the two-parameter bathtub hazard shaped distribution. *Scientific African* 17: 1-19.
Smith, R. M. & Bain, L. J. 1975. An exponential power life-testing distribution. *Communications in Statistics* 4(5): 469–481.
Wang, R. Sha, N. Gu, B. & Xu, X. 2014. Statistical Analysis of a Weibull Extension with Bathtub-Shaped Failure Rate Function. *Advances in Statistics* **2014**:1–15.
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.
Zhang, Z. & Gui, W. 2022. Statistical Analysis of the Lifetime Distribution with Bathtub-Shaped Hazard Function under Lagged-Effect Step-Stress Model. *Mathematics* **2022**(5): 1-23.

Department of Mathematics and Statistics
Faculty of Science
Universiti Putra Malaysia
43400 UPM Serdang
Selangor DE, MALAYSIA.
E-mail : idari512@uitm.edu.my [*], *jayanthi@upm.edu.my, mshafie@upm.edu.my, aslam.safari@upm.edu.my*

Mathematical Sciences Studies,
College of Computing, Informatics and Media
Universiti Teknologi MARA Cawangan Kelantan
Kampus Machang, Bukit Ilmu,
18500 Machang
Kelantan DN, MALAYSIA.
E-mail : idari512@uitm.edu.my [*]

Received: 17 July 2023
Accepted: 24 May 2024

[*]Corresponding author