Modeling Repairable System Failures with Repair Effect and Time Dependent Covariates

Jayanthi Arasan and Samira Ehsani

***Abstract*—In this paper we extend a repairable system model that incorporates both time trend and repair history to include a time dependent covariate. We calculated the bias, standard error and RMSE of the parameter estimates of this model at different sample sizes using simulated data. Following that, we studied the Wald method of constructing confidence interval estimates for the parameters of this model. Finally the model is fit to real data from the pipeline network failure.**

***Index Terms*—repairable, covariate, Wald.**

I. Introduction

A System is said to be repairable when it can be restored back to functionality by some repair process or maintenance action after a failure has occurred in the system. The "repair time" which is the period where the system is unable to function is assumed to be negligible. In most cases a repair action can only bring the system back to the state it was prior to the failure, also known as "as-bad-as-old". This type of repair action is also referred to as minimal, imperfect or general repair. Thus, modeling these types of repair has received a lot of attention recently. The general repair model for repairable systems by using the idea of the virtual age process of the system was developed by [1].

Lawless and Thiagarajah [2] introduced a proportional intensity model that incorporates both time trends and renewal type behavior. Guo et al. [3] later proposed a new general repair model based on the expected cumulative number of failures to capture the repair history. Other literatures on the repairable system models and recurrent events are [4], [5], [6], [7], [8], [9], [10] and [11].

Most repairable system models do not take into account other factors that affect repair times, more popularly known as covariates or concomitant variables. In some analysis involving repairable systems, covariates can be very useful in indicating the cause of failures. Røstum [12] showed how the use of covariates such as length or diameter of pipes, age and presence of clay can be very useful in analyzing pipe failures in water networks. It is rather common in any analysis to find covariates that do not remain at a fixed value over time. These types of covariates are known as time dependent covariates, for example, age, level of erosion, water pressure and velocity.

II. The Models

Lawless & Thiagarajah [13] introduced a proportional intensity model where the failure intensity function conditional on the history up to time t, H_t is $\lambda(t, H_t) = e^{\boldsymbol{\theta}'\boldsymbol{x}(t)}$, where $\boldsymbol{x}(t) = (x_1(t), ..., x_p(t))'$ is a vector of functions that may depend on both t and H_t and $\boldsymbol{\theta} = (\theta_1, ..., \theta_p)'$ is the vector of unknown parameters. This idea was later extended by [3] to obtain a parametric model based on the expected cumulative number of repairs (failures). In this research, we extend this model to include a time dependent covariate.

For the proportional intensity model the effect of a time dependent covariate can be added to the model in different ways depending on the nature of the covariate. Let z be the value of a time dependent covariate at the time the study started and $z(t)$ its value at time t. Then, the failure intensity function is,

$$\lambda(t) = \lambda_0(t) \exp(\gamma\mu(t) + \beta z(t)), \tag{1}$$

where,

$$\lambda_0(t) = e^{a+bt}. \tag{2}$$

and a, b, γ and β are the parameters of the model. $\mu(t)$ is the expected cumulative number of failures up to time t.

In this work we assume that the system is a network consisting of smaller components. Suppose we have a series of $i = 1, 2, \ldots, n$ events triggered by different compenents with covariate value z_i at start of study and $z(t_i)$ at the i^{th} failure. Let us consider a model with $z(t_i) = z_i + t_i$, where the cumulative number of failures up to time t_i is $(i-1)$. , If $h_i = e^{(a+bt_{i-1}+\gamma(i-1)+\beta(z_i+t_{i-1}))}$ and $w_i = e^{(a+bt_i+\gamma(i-1)+\beta(z_i+t_i))}$, the conditional pdf of the i^{th} failure is,

$$\begin{aligned} f(t_i|t_{i-1}) &= \exp\Big(a + bt_i + \gamma(i-1) + \beta(z_i + t_i) \\ &+ \frac{h_i + w_i}{b+\beta}\Big). \end{aligned} \tag{3}$$

Following that the corresponding log-likelihood function for observed data of n events is,

$$\begin{aligned} l(a, b, \beta, \gamma) &= \sum_{i=1}^{n} a + bt_i + \gamma(i-1) + \beta(z_i + t_i) \\ &+ \frac{h_i + w_i}{b+\beta} \end{aligned} \tag{4}$$

Then, the first and second derivatives of the log-likelihood function would be as follows,

$$\frac{\partial l}{\partial a} = \sum_{i=1}^{n} 1 + \frac{h_i - w_i}{b+\beta}$$

$$\frac{\partial l}{\partial b} = \sum_{i=1}^{n} t_i + \frac{t_{i-1}h_i - t_i w_i}{b+\beta} + \frac{-h_i + w_i}{(b+\beta)^2}$$

Manuscript received April 01, 2011; revised April 27, 2011.

J. Arasan is with the Department of Mathematics, Faculty of Science, University Putra Malaysia, Serdang, 43400, Selangor, Malaysia (e-mail: jayanthi@math.upm.edu.my).

S. Ehsani is with the Department of Mathematics, Faculty of Science, University Putra Malaysia, Serdang, 43400, Selangor, Malaysia e-mail: (samira@math.upm.edu.my).

$$\frac{\partial l}{\partial \beta} = \sum_{i=1}^{n} z_i + t_i + \frac{(z_i - t_{i-1})h_i + (z_i + t_i)w_i}{b+\beta} + \frac{-h_i + w_i}{(b+\beta)^2}$$

$$\frac{\partial l}{\partial \gamma} = \sum_{i=1}^{n} i - 1 + \frac{(i-1)h_i - (i-1)w_i}{b+\beta}$$

$$\frac{\partial^2 l}{\partial a^2} = \sum_{i=1}^{n} \frac{h_i - w_i}{b+\beta}$$

$$\frac{\partial^2 l}{\partial a \partial b} = \sum_{i=1}^{n} \frac{t_{i-1}h_i - t_i w_i}{b+\beta} + \frac{-h_i + w_i}{(b+\beta)^2}$$

$$\frac{\partial^2 l}{\partial a \partial \beta} = \sum_{i=1}^{n} \frac{(z_i + t_{i-1})h_i - (z_i + t_i)w_i}{b+\beta} + \frac{-h_i + w_i}{(b+\beta)^2}$$

$$\frac{\partial^2 l}{\partial a \partial \gamma} = \sum_{i=1}^{n} \frac{(i-1)h_i - (i-1)w_i}{b+\beta}$$

$$\frac{\partial^2 l}{\partial b^2} = \sum_{i=1}^{n} \frac{t_{i-1}^2 h_i + t_i^2 w_i}{b+\beta} - \frac{2(-t_{i-1}h_i + t_i w_i)}{(b+\beta)^2} - \frac{2(-h_i + w_i))}{(b+\beta)^3}$$

$$\frac{\partial l}{\partial b \partial \beta} = \sum_{i=1}^{n} \frac{t_{i-1}(z_i + t_{i-1})h_i - t_i(z_i + t_i)w_i}{b+\beta} + \frac{-t_{i-1}h_i + t_i w_i}{(b+\beta)^2} + \frac{-(z_i + t_{i-1})h_i + (z_i + t_i)w_i}{(b+\beta)^2} - \frac{2(-h_i + w_i)}{(b+\beta)^3}$$

$$\frac{\partial l}{\partial b \partial \gamma} = \sum_{i=1}^{n} \frac{t_{i-1}(i-1)h_i - t_i(i-1)w_i}{b+\beta} + \frac{-(i-1)h_i + (i-1)w_i}{(b+\beta)^2}$$

$$\frac{\partial l}{\partial \beta^2} = \sum_{i=1}^{n} \frac{(z_i + t_{i-1})^2 h_i - (z_i + t_i)^2 w_i}{b+\beta} + \frac{2(-(z_i + t_{i-1})h_i + (z_i + t_i)w_i)}{(b+\beta)^2} - \frac{2(-h_i + w_i)}{(b+\beta)^3}$$

$$\frac{\partial l}{\partial \beta \partial \gamma} = \sum_{i=1}^{n} \frac{(z_i + t_{i-1})(i-1)h_i - (z_i + t_i)(i-1)w_i}{b+\beta} + \frac{-(i-1)h_i + (i-1)w_i)}{(b+\beta)^2}$$

$$\frac{\partial^2 l}{\partial \gamma^2} = \sum_{i=1}^{n} \frac{(i-1)^2 h_i - (i-1)^2 w_i}{b+\beta}$$

A. Simulation Study

A simulation study using 1000 samples of sizes $n = 50, 80, 100, 150$ and 200 was conducted with one time dependent covariate. The covariate values were simulated from the standard normal distribution. The values of -2.6, 0.008, 0.0005 and -0.07 were chosen as the parameter values of a, b, β and γ. These particular values of the parameters were chosen to give us failure times that are similar to those found in pipeline failures where a common time dependent covariate is the age of the pipes.

TABLE I
BIAS, STANDARD ERROR AND RMSE OF THE PARAMETER ESTIMATES

	n	$\hat{a}$	$\hat{b}$	$\hat{\beta}$	$\hat{\gamma}$
Bias	50	-0.29991	0.00926	0.00869	-0.16431
	80	-0.27372	0.00599	0.00443	-0.09086
	100	-0.25182	0.00594	0.00194	-0.06732
	150	-0.19407	0.00767	-0.00255	-0.04303
	200	-0.15992	0.00394	-0.00016	-0.03150
s.e	50	0.32474	0.24454	0.24135	0.09727
	80	0.30150	0.18375	0.18346	0.03925
	100	0.30366	0.11857	0.11858	0.03176
	150	0.22181	0.12272	0.12246	0.02950
	200	0.22166	0.09309	0.09265	0.02849
RMSE	50	0.44204	0.24471	0.24150	0.19094
	80	0.40722	0.18384	0.18351	0.09897
	100	0.39449	0.11872	0.11859	0.07443
	150	0.29472	0.12296	0.12249	0.05217
	200	0.27333	0.09318	0.09265	0.04247

Random numbers, u_i, were generated from the uniform distribution on the interval $(0, 1)$, to produce t_i as follows,

$$t_i = \frac{1}{b+\beta}\Bigg(\ln\left[e^{a+\gamma(i-1)+\beta(t_{i-1}+z_i)+bt_{i-1}} - \ln(u_i)(b+\beta)\right] - (a + \beta z_i + \gamma(i-1))\Bigg).$$

Table I shows the bias, standard errors and RMSE of the parameter estimate at various sample sizes. From the results we can see that both bias and standard error values are relatively low for all the parameter estimates. When n increases, the decrease in the value of the bias, standard error and RMSE is clear. Thus, we can conclude that the estimation procedure is working well for the proposed model.

III. CONFIDENCE INTERVAL ESTIMATES

Based on the asymptotic normality of the MLE, the inverse of the observed information matrix, which can be obtained from the second partial derivatives of the log-likelihood function evaluated at $\hat{a}$,$\hat{b}$,$\hat{\beta}$ and $\hat{\gamma}$, provides us with the estimators for the variance and covariance,

$$\widehat{var}(\hat{a}, \hat{b}, \hat{\beta}, \hat{\gamma}) = [i(\hat{a}, \hat{b}, \hat{\beta}, \hat{\gamma})]^{-1}.$$

The confidence interval estimate based on the asymptotic normality of the maximum likelihood estimates is also known as the Wald interval. However, it is widely known that the Wald intervals can be highly asymmetrical, see [14]. Also, in many cases it can be totally unreliable especially for short series of events which is rather common in the modeling of repairable systems or any recurrent events in general. Thus, we would like to assess its performance before applying the Wald intervals to the parameters of this model.

Let $\hat{\theta}$ be the maximum likelihood estimator for parameter θ and $l(\theta)$ the log-likelihood function of θ. Under mild regularity conditions, $\hat{\theta}$ is asymptotically normally distributed with mean θ and covariance matrix $I^{-1}(\theta)$, where $I(\theta)$ is the Fisher information matrix evaluated at the true value of the parameter θ, [5]. The matrix $I(\theta)$ which is not available can be replaced by the observed information matrix $I(\hat{\theta})$

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

The estimate of var$(\hat{\theta}_j)$ is then given by the $(j,j)^{th}$ element of $I^{-1}(\hat{\theta})$. If $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$ quantile of the standard normal distribution the $100(1-\alpha)\%$ confidence interval for θ_j is given by the following,

$$\hat{\theta}_j - z_{1-\frac{\alpha}{2}}\sqrt{I^{-1}(\hat{\theta})_{jj}} < \theta_j < \hat{\theta}_j + z_{1-\frac{\alpha}{2}}\sqrt{I^{-1}(\hat{\theta})_{jj}}.$$

A. Coverage Probability Study

We conducted a coverage probability study using study using $N = 1000$ samples of sizes $n = 50, 80, 100, 150$ and 200 to compare the performance of the confidence interval estimates at $\alpha = 0.05$ and $\alpha = 0.10$ where α is the nominal error probability. Following that, we calculated the estimated total error probabilities by adding the number of times an interval did not contain the true parameter value divided by the total number of samples.

The estimated left(right) error probability was calculated by adding the number of times the left(right) endpoint was more(less) than the true parameter value divided by the total number of samples, N. If the total error probability is greater than $\alpha + 2.58\ s.e(\hat{\alpha})$, then the method is termed anticonservative and if it is lower than $\alpha - 2.58\ s.e(\hat{\alpha})$, the method is termed conservative, [15].

B. Results and Discussion

The coverage probability of a confidence interval is the probability that the interval contains the true parameter value and should preferably be equal or close to the nominal coverage probability, $(1-\alpha)$. Figure 1 illustrates the estimated left and right error probabilities for parameters a, b, β and γ when $\alpha = 0.05$. Tables 2-3 show the results obtained from the coverage probability study. The overall performances of the different methods were judged based on the total number of anticonservative, conservative and asymmetrical intervals.

TABLE II
SUMMARY OF THE INTERVAL ESTIMATES AT $\alpha = 0.05$.

n	Anti conservative	Conservative	Asymmetrical
50	4	0	2
80	4	0	2
100	2	0	2
150	3	0	2
200	2	0	3

From figure 1 we can observe that the Wald interval works quite well for the parameters b and β but rather poorly for parameters a and γ. We can also see that the number of anti conservative and asymmetrical intervals are rather high especially at lower sample sizes, see table 2. Thus, we should apply the Wald intervals with much caution especially for parameters a and γ.

IV. Application with Real Data

We analyzed the demo pipeline data obtained from Trondheim's Gemini VA database using the repairable system model with a time dependent covariate. This database contains 82 failures from a pipe network starting from the year 1975 to 2001. The time dependent covariate used in the model is the age of the pipes. All variables were measuerd in days.

Table 4 shows the values of the parameter estimates and standard errors when the data was fitted to the proposed model. The table also shows the 90% and 95% Wald confidence interval estimates for the model parameters. We know that if b is not significant then there is no evidence of time trend and if the parameter γ is not significant then there is no evidence of repair effect within the proposed model. Here, $b > 0$ shows that the failure rate is increasing over time. The value of $\gamma < 0$ implies that the repair action reduces the failure intensity. The parameter β shows the effect of the time dependent covariate. The value $\beta < 0$ indicates that older pipes have higher resistance against failure . The Wald

TABLE III
ESTIMATED ERROR PROBABILITIES AT $\alpha = 0.05$.

Para.	n	Left	Right	Total
a	50	0.074	0.113	0.187
	80	0.037	0.100	0.137
	100	0.030	0.101	0.131
	150	0.026	0.093	0.119
	200	0.019	0.073	0.092
b	50	0.039	0.033	0.072
	80	0.035	0.033	0.068
	100	0.037	0.025	0.062
	150	0.039	0.027	0.066
	200	0.034	0.022	0.056
β	50	0.041	0.033	0.074
	80	0.034	0.030	0.064
	100	0.027	0.033	0.060
	150	0.028	0.038	0.066
	200	0.026	0.034	0.060
γ	50	0.000	0.318	0.318
	80	0.002	0.248	0.250
	100	0.001	0.201	0.202
	150	0.001	0.158	0.159
	200	0.002	0.131	0.133

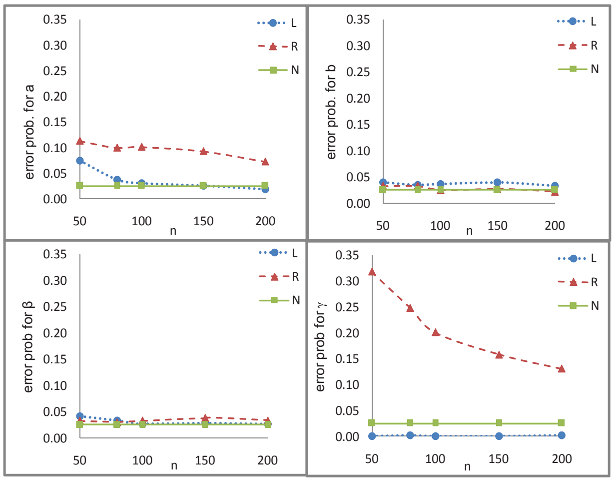

Fig. 1. Estimated error probabilities for a, b, β and γ at $\alpha = 0.05$.

TABLE IV
ESTIMATES OF PARAMETERS FOR PIPE NETWORK FAILURES WITH TIME DEPENDENT COVARIATE.

Para.	Est.	Std.err.	Wald intervals (90%) Wald intervals (95%)
a	-5.359030	0.38410	$(-5.99087, -4.72719)$ $(-6.11187, -4.60619)$
b	0.001500	0.00056	$(0.00058, 0.00242)$ $(0.00040, 0.00260)$
β	-0.000004	0.00001	$(-0.00001, 0.00001)$ $(-0.00002, 0.00002)$
γ	-0.173530	0.06564	$(-0.28151, -0.06555)$ $(-0.30218, -0.04487)$

confidence interval estimates suggests that parameters b and γ are significant whereas β is not significant at both 90% and 95% confidence intervals. However, as we can see from the results of the coverage probability study , the Wald intervals are not very reliable. Thus, we carry out the likelihood ratio test to check the significance of the parameters of this model.

This test is known to perform better than the Wald test since it has better statistical properties, [13]. The basic idea of a likelihood ratio test is to compare the maximized likelihood of two nested models, the full model and the reduced model. The reduced model is restricted by certain conditions given in H_0.
Let $\widehat{\theta}_r$ be the maximum likelihood estimator of the restricted model under H_0 and $\widehat{\theta}_f$ the maximum likelihood estimator of the full model. The maximized likelihood of the reduced model, $l(\widehat{\theta}_r)$ can never exceed the maximized likelihood of the full model, $l(\widehat{\theta}_f)$, because it is a subset of the full model. Thus, the ratio of the maximized likelihood of the reduced model to the full model is bounded between 0 and 1. A ratio close to 1 indicates that the reduced model is close to the full model whereas a ratio close to 0 indicates that the two models are quite different and that the reduced model is unacceptable. The likelihood ratio statistic for testing H_0 versus H_l is given by the following,

$$\Psi = -2[L(\widehat{\theta}_r) - L(\widehat{\theta}_f)].$$

For large sample size, Ψ is approximately $\chi^2_{(\nu)}$, where ν is the number of parameters in the full model minus the number of parameters in the reduced model.

We are interested mainly in testing whether the time trend, repair effect and the covariate effect are significant. Table 5 displays the likelihood ratio test results and whether H_0 is rejected at $\alpha = 0.05$. The results indicate that both time trend and repair effect are significant at $\alpha = 0.05$. However the effect of the time dependent covariate, is not significant at $\alpha = 0.05$. So, the full model is not necessary and we can use the reduced model by omitting the time dependent covariate.

Table 6 shows the values of the parameter estimates and standard error for the reduced model. Figure 2 shows the estimates of the expected number of failures for the reduced model. We can clearly see that expected number of failures appear to fit the real data very well.

V. CONCLUSION

In this paper, we proposed the use of a repairable system model with time dependent covariate. This model allows us

TABLE V
LIKELIHOOD RATIO TEST RESULTS.

Hypothesis	Ψ	Conclusion
1.H_0: $b = 0$ H_1: $b \neq 0$	7.640	*Reject* H_0 *at* $\alpha = 0.05$
2.H_0: $\gamma = 0$ H_1: $\gamma \neq 0$	7.346	*Reject* H_0 *at* $\alpha = 0.05$
3.H_0: $\beta = 0$ H_1: $\beta \neq 0$	0.114	*Fail to reject* H_0 *at* $\alpha = 0.05$

TABLE VI
PARAMETERS ESTIMATES FOR PIPE NETWORK FAILURES FOR THE FINAL MODEL.

Para.	Est.	Std.err.
a	-5.43490	0.31465
b	0.00150	0.00056
γ	-0.17373	0.06589

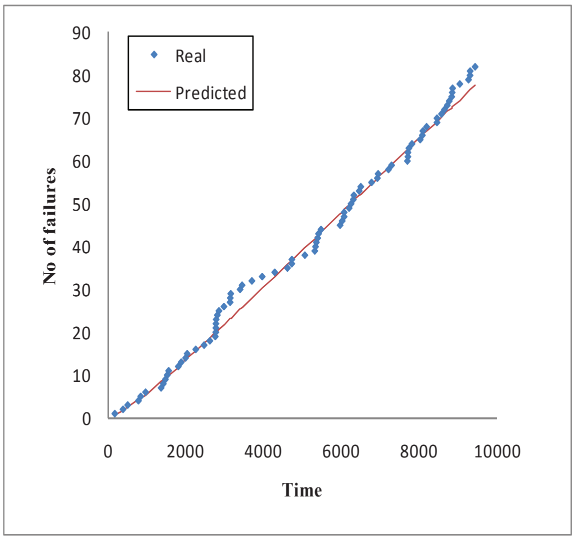

Fig. 2. Estimates of expected number of failures for reduced model.

to incorporate the effect of a time dependent covariate in the modeling of repairable system failures. More work should be done to investigate the use of other types of models that can incorporate time dependent covariates. We should also extend these models to include several types of covariates, for example fixed and time dependent covariates or covariates that may take values that follow a step function instead of continuously changing with time.

Discussion on the confidence interval procedure for this model was based on the asymptotic normality of the maximum likelihood method. This method depends heavily on asymptotic properties, thus, for the moderate and small sample sizes the standard error estimates may not be particularly good. Alternative methods based on the parametric bootstrap should be investigated since they have proved to be more useful in many instances especially for short series of events which is rather common in the modeling of repairable systems or any recurrent events in general.

REFERENCES

[1] M. Kijima, “Some results for repairable systems with general repair,” *Journal of Applied Probability*, vol. 26, pp. 89–102, 1989.

[2] J. F. Lawless and K. Thiagarajah, “A point-process model incorporating renewals and time trends, with application to repairable systems,” *Technometrics*, vol. 38, no. 2, pp. 131–138, 1996.

[3] H.R. Guo, H. Liao, W. Zhao, and A. Mettas, “A new stochastic model for systems under general repairs,” *IEEE Trans. on Reliability*, vol. 56, no. 1, pp. 40–49, 2007.

[4] C. Brown, “On the use of indicator variable for studying the time-dependence of parameters in a response-time model,” *Biometrics*, vol. 31, pp. 863–872, 1975.

[5] D. R. Cox and P. A. W. Lewis, *The Statistical Analysis of Series of Events*. London, U.K.: Methuen, 1966.

[6] L. H. Crow, “Reliability analysis for complex, repairable systems,” *in Reliability and Biometry* , F. Proschan and R.J. Serfling, Eds, Philadelphia, PA: SIAM, pp. 379–410, 1974.

[7] S. Gasmi, C.E. Love, and W. Kahle, “A general repair, proportional-hazards, framework to model complex repairable systems,” *IEEE Trans. on Reliability*, vol. 52, no. 1, pp. 26–32, 2003.

[8] M. Kaminskiy and V. Krivtsov, “A Monte Carlo approach to repairable system repairable analysis,” *in Probabilistic Safety Assessment and Management.* New York: Springer, pp. 1063–1068, 1998.

[9] M. Kijima and N. Sumita, “A useful generalization of renewal theory, counting process governed by non-negative markovian increments,” *Journal of Applied Probability*, vol. 23, pp. 71–88, 1986.

[10] H. Wang and N. Pham, “Optimal age-dependent preventive maintanance policies with imperfect maintanance,” *International Journal of Reliability, Quality and Safety Engineering*, vol. 3, pp. 119–135, 1996.

[11] M. Yanez, F. Joglar, and M. Modarres, “Generalized renewal process for analysis of repairable systems with limited failure experience,” *Reliability Engineering and System Safety*, vol. 77, pp. 167–180, 2002.

[12] J. Røstum, *Statistical Modelling of Pipe Failures in Water Networks*. PhD Dissertation, Norwegian University of Science and technology, 2000.

[13] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*. New York: John Wiley, 1982.

[14] J. Arasan and M. Lunn, “Survival model of a parallel system with dependent failures and time varying covariates,” *J. Statist. Plann. Inference*, vol. 139, no. 3, pp. 944–951, 2009.

[15] N. Doganaksoy and J. Schmee, “Comparison of approximate confidence intervals for distributions used in life-data analysis,” *Technometrics*, vol. 35, no. 2, pp. 175–184, 1993.