Sains Malaysiana 42(7)(2013): 981–987

Modeling Repairable System Failure with Repair History and Covariates

(Model Sistem Kegagalan Dibaiki dengan Sejarah Pembaikan dan Kovariat)

SAMIRA EHSANI*, JAYANTHI ARASAN & NOOR AKMA IBRAHIM

ABSTRACT

In this paper, we extended a repairable system model under general repair that is based on repair history, to incorporate covariates. We calculated the bias, standard error and RMSE *of the parameter estimates of this model at different sample sizes using simulated data. We applied the model to a real demonstration data and tested for existence of time trend, repair and covariate effects. Following that we also conducted a coverage probability study on the Wald confidence interval estimates. Finally we conducted hypothesis testing for the parameters of the model.The results indicated that the estimation procedure is working well for the proposed model but the Wald interval should be applied with much caution.*

Keywords: Covariate; general repair; repairable system

ABSTRAK

Dalam kertas ini, kami melanjutkan model sistem dibaiki di bawah pembaikan am yang berdasarkan sejarah pembaikan, dengan menggabungkan kovariat. Kami mengira ralat, sisihan piawai dan PMRKD *bagi penganggar parameter-parameter model ini pada sampel yang berbeza saiz dengan menggunakan data simulasi. Kami menguna pakai model ini kepada data demonstrasi sebenar dan telah menguji kewujudan kecenderungan masa, kesan pembaikan dan kovariat. Berikutan itu kami juga menjalankan kajian liputan kebarangkalian bagi anggaran selang keyakinan 'Wald'. Akhirnya kami menjalankan pengujian hipotesis bagi parameter-parameter model. Keputusan yang diperoleh menunjukkan bahawa prosedur penganggaran berjalan lancar bagi model yang dicadangkan tetapi selang 'Wald' harus digunakan dengan berhati-hati.*

Kata kunci: Kovariat; perbaikan umum; sistem diperbaiki

INTRODUCTION

A system is said to be repairable when it can be resorted back to functionality after it has failed to perform at least one to its intended functions. Repair action can bring the system to one of the following states: as bad as old (minimal repair), as good as new (perfect repair) or better than old but worse than new (general repair) (Høyland & Rausand 1994). Many stochastic models have been developed for repairable systems assuming different repair effects. The period when the system is unable to function is referred to as repair time and is assumed to be negligible. The proportional intensity (PI) model and the virtual age model are popularly used to account for general repair effect. Lawless and Thiagarajah (1996) introduced a proportional intensity model that incorporates both time trends and renewal type behavior. Guo et al. (2007) later proposed a new general repair model based on the expected cumulative number of failures to capture the repair history. The virtual age models by Kijima (1989) and Kijima and Sumita (1986) express the repair effect by a reduction of the system age unlike the PI models where the repair effect is expressed by a reduction of the system failure intensity (Guo et al. 2007). Other literatures on the repairable system models and recurrent events are Brown (1975), Cox and Lewis (1966), Crow (1974), Gasmi et al. (2003), Kaminskiy and Krivtsov (1998), Wang and Pham (1996) and Yañez et al. (2002).

Most repairable system models do not include covariates or other factors that affect repair times. Røstum (2000) showed how the used covariates such as length or diameter of pipe, age and presence of clay can be very useful in analyzing pipe failures in water networks. The aim of this paper was to extend a repairable system model based on repair history to incorporate the effect of covariates. The baseline intensity function can be described using several forms such as the log-linear, power law or linear. In this paper we used the log-linear intensity function due to its flexibility and wide application.

We briefly review the proportional intensity model for repairable systems. Then, we extend the model proposed by Guo et al. (2007) to incorporate several covariates. Following that, a simulation study is conducted to assess the accuracy and efficiency of the parameter estimates. Finally we present a numerical example with some applications of the proposed model.

PROPORTIONAL INTENSITY MODEL

The PI Model was first introduced by Cox (1972). PI model is used to model the intensity process of failures and repairs of a repairable system which incorporates explanatory variables. Vlok et al. (2004) introduced a PI model for both non-repairable and repairable systems utilizing historic failure data and corresponding diagnostic measurements. Lawless and Thiagarajah (1996) introduced a proportional intensity model where the failure intensity function conditional on the history up to time t, H_t is $\lambda(t, H_t) = e^{\theta' x(t)}$, where $x(t) = (x_1(t),\ldots,x_p(t))'$ is a vector of functions that may depend on both t and H_t and $\theta = (\theta_1,\ldots,\theta_p)'$ is the vector of unknown parameters. He then studied a PI model that incorporates both time trends and renewal type behavior with the following failure intensity,

$$\lambda(t) = \exp(\alpha + \beta t + \gamma(t - t_{N(t^-)}), \tag{1}$$

where α, β, γ are the parameters of the model and $t - t_{N(t^-)}$ is the time of the last failure before t. Guo et al. (2007) proposed another general repair model based on expected cumulative number of repairs or failures where,

$$\lambda(t) = \lambda_0(t) \exp[\theta' x(t)] = \lambda_0(t) \exp[\gamma m(t)], \tag{2}$$

and $\lambda_0(t)$ is the baseline failure intensity function and $m(t) = E[N(t)]$, where $N(t)$ is the cumulative number of failures up to time t.

MODEL DEVELOPMENT

Maintenance action often brings a system somewhere between new and old state. Therefore a general repair model is more realistic for describing maintenance effort. All covariates that could have influence on the rate of occurrence of failures should be included in the statistical model. This would help in understanding the significance of each covariate on the failure history. Most of the systems are influenced by different covariates that can be either time varying or fixed. Fixed covariates have values that are independent of time for each failure. Here, we assume that the system is a network consisting of smaller components. Also, we have a series of $i = 1, 2, \ldots, n$ events triggered by different components with covariate value x_i at the i^{th} failure. For the log-linear baseline intensity function, the effect of the covariates can be incorporated in the failure intensity function as given below,

$$\lambda(t) = e^{\beta' x + gt + \gamma m(t)}. \tag{3}$$

Here $x' = (x_0, x_1, \ldots, x_p)$ is the vector of covariate values, where $x_0 = 1$ and $\beta' = (\beta_0, \beta_1, \ldots, \beta_p)$, g and γ are the parameters of the model and $m(t)$ is the expected cumulative number of failures up to time t. Suppose we have a series of $i = 1, 2, \ldots, n$ events. If the cumulative number of failures up to time t_i is $(i-1)$, $h_{1i} = e^{\beta' x_i + g t_{i-1} + \gamma(i-1)}$, $h_{2i} = e^{\beta' x_i + g t_i + \gamma(i-1)}$, the conditional reliability function before the i^{th} failure is,

$$R(t_i \mid t_{i-1}) = \exp\left[\frac{1}{g}\left(h_{1i} - h_{2i}\right)\right]. \tag{4}$$

The conditional pdf for the i^{th} failure is then,

$$f(t_i \mid t_{i-1}) = h_{2i} \exp\left[\frac{1}{g}\left(h_{1i} - h_{2i}\right)\right], \tag{5}$$

and the corresponding log-likelihood function for observed data on n events is,

$$l(\beta, g) = \sum_{i=1}^{n}\left(\beta' x_i + g t_i + \gamma(i-1) + \frac{h_{1i} - h_{2i}}{g}\right). \tag{6}$$

The first and second derivations of the log-likelihood function with respect to parameters are,

$$\frac{\partial l}{\partial \beta_j} = \sum_{i=1}^{n}\left(x_i^j + \frac{h_{1i} - h_{2i}}{g}\right), \quad j = 0, 1, \ldots, p,$$

$$\frac{\partial l}{\partial g} = \sum_{i=1}^{n}\left(t_i + \frac{t_{i-1}h_{1i} - t_i h_{2i}}{g} - \frac{h_{1i} - h_{2i}}{g}^2\right),$$

$$\frac{\partial l}{\partial g} = \sum_{i=1}^{n}\left(t_i + \frac{t_{i-1}h_{1i} - t_i h_{2i}}{g} - \frac{h_{1i} - h_{2i}}{g}^2\right),$$

$$\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} = \sum_{i=1}^{n}\left(x_i^{j+k} + \frac{h_{1i} - h_{2i}}{g}\right), \; j = 0, 1, \ldots, p, \; k = 0, 1, \ldots, p,$$

$$\frac{\partial^2 l}{\partial \beta_j \partial g} = \sum_{i=1}^{n} x_i^j \left(\frac{t_{i-1}h_{1i} - t_i h_{2i}}{g} - \frac{h_{1i} - h_{2i}}{g}^2\right), \; j = 0, 1, \ldots, p,$$

$$\frac{\partial^2 l}{\partial \beta_j \partial \gamma} = \sum_{i=1}^{n} x_i^j (i-1)\left(\frac{h_{1i} - h_{2i}}{g}\right), \quad j = 0, 1, \ldots, p,$$

$$\frac{\partial^2 l}{\partial g^2} = \sum_{i=1}^{n}\left[\frac{t_{i-1}^2 h_{1i} - t_i^2 h_{2i}}{g} - 2\left(\frac{t_{i-1}h_{1i} - t_i h_{2i}}{g^2}\right) + 2\left(\frac{h_{1i} - h_{2i}}{g}^3\right)\right],$$

$$\frac{\partial^2 l}{\partial g \partial \gamma} = \sum_{i=1}^{n}\left[-\frac{(i-1)\left(-g t_{i-1}h_{1i} + g t_i h_{2i} + h_{1i} - h_{2i}\right)^2}{\beta_2}\right],$$

$$\frac{\partial^2 l}{\partial \gamma^2} = \sum_{i=1}^{n}\left[-\frac{(i-1)^2\left(-h_{1i} + h_{2i}\right)}{g}\right].$$

SIMULATION STUDY

Simulation studies using 1000 samples of n events where n=50,80,100,150 and 200 were conducted for the new model with one covariate. The covariate values were simulated from the standard normal distribution. The values of -3.2, 0.05, 0.045 and -0.05 were chosen as the initial values of parameters β_0, β_1, g and γ. These values were chosen specifically to give us failure times that are similar to those found in pipeline failures. Suppose we have $i = 1,2,\ldots, n$ failures. Random numbers, u_0, were generated from the uniform distribution on the interval (0, 1), to produce t_i as follows,

$$t_i = \frac{ln\left(e^{\beta_0+\beta_1 x_i+g t_{i-1}+\gamma i} - ln(u_i) g e^{\gamma}\right) - (\beta_0 + \beta_1 x_i + \gamma i)}{g}.$$

SIMULATION RESULTS

Table 1 shows the results of the bias, standard error and RMSE for the parameter estimates of the new model at different sample sizes. Both bias and std.error contribute to the average error size of an estimator, thus the RMSE = $\sqrt{se^2 + bias^2}$ is used to measure the average overall error of the parameter estimates.

TABLE 1. Bias of the estimates

n	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{g}$	$\hat{\gamma}$
50	-0.78158	0.00094	0.02056	-0.03629
80	-0.55653	0.00389	0.01219	-0.01757
100	-0.47831	0.00568	0.00968	-0.01284
150	-0.40591	-0.00063	0.00758	-0.00938
200	-0.35134	-0.00273	0.00620	-0.00736

TABLE 2. Standard error of the estimates

n	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{g}$	$\hat{\gamma}$
50	0.88721	0.14737	0.02307	0.04548
80	0.75070	0.18239	0.01551	0.02356
100	0.74675	0.13589	0.01510	0.02191
150	0.60570	0.11509	0.01127	0.01499
200	0.53769	0.09170	0.00902	0.01104

TABLE 3. RMSE of the estimates

n	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{g}$	$\hat{\gamma}$
50	1.18237	0.14737	0.03090	0.05818
80	0.93449	0.18243	0.01973	0.02939
100	0.88680	0.13601	0.01794	0.02540
150	0.72913	0.11509	0.01359	0.01768
200	0.64230	0.09174	0.01095	0.01327

From the results we can see that all values of bias, standard error and RMSE were relatively low for all the parameter estimates. When n increase, the standard error and RMSE values clearly decreased. This decreasing trend is also apparent for the values of bias of $\hat{\beta}_0, \hat{g}$ and $\hat{\gamma}$ but not that clear for $\hat{\beta}_1$. However, overall we can conclude that the estimation procedure is working well for the proposed model (Tables 2 & 3).

CONFIDENCE INTERVAL ESTIMATES

Let $\hat{\theta}$ be the maximum likelihood estimator for parameter θ and $l(\theta)$ the log-likelihood function of θ. $\hat{\theta}$ 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. 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 $\left(1-\frac{\alpha}{2}\right)$ quantile of the standard normal distribution the 100(1 – α)% confidence interval for θ_j is given by the following,

$$\hat{\theta}_J - z_{1-\frac{\alpha}{2}} \sqrt{I^{-1}\left(\hat{\theta}\right)_{jj}} < \theta_j < \hat{\theta}_J + z_{1-\frac{\alpha}{2}} \sqrt{I^{-1}\left(\hat{\theta}\right)_{jj}}. \quad (7)$$

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 – α). We conducted a coverage probability study with N = 1000 samples to assess the performance of the intervals. Following that, we calculated the estimated total error probabilities by adding the number of times in which 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) end point was more(less) than the true parameter value divided by the total number of samples, N. Figures 1 to 4 show the estimated coverage probability for parameters β_0, β_1, g and γ at $\alpha = 0.05$.

The Wald interval is known to work well only when sample sizes are rather large and is prone to be highly asymmetrical (Arasan & Lunn 2009). Thus, we calculated the estimated coverage probabilities for sample sizes n= 50, 80, 100, 150 and 200. The results indicated that the Wald interval works rather well for parameters β_0 and β_1, although there are quite a few asymmetrical intervals. However, for parameters g and γ the performance is rather poor with several anti conservative and asymmetrical intervals but starts to improve when sample size increases (Table 4).

The Wald does not produce any conservative intervals and when sample size increases, the number of anti-conservative intervals decreases. Thus, the Wald interval should be applied with caution especially for parameters g and γ (Table 5). Arasan and Lunn (2008) showed that other methods can work better than the Wald for moderate and low sample sizes. Therefore, other confidence interval estimation techniques, such as the parametric bootstrap can be investigated in future to compare their performance with the Wald interval.

NUMERICAL EXAMPLE

A pipeline failure demo data is fit to the proposed model with two covariates. This data set contains structural data from the pipes, such as diameter and length of pipe, which are the covariates of the model and maintenance data is the

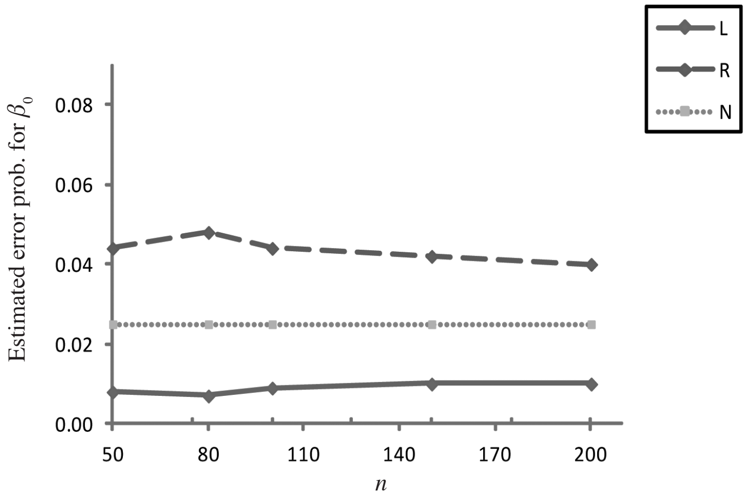

FIGURE 1. Estimated error probabilities for β_0 at $\alpha = 0.05$

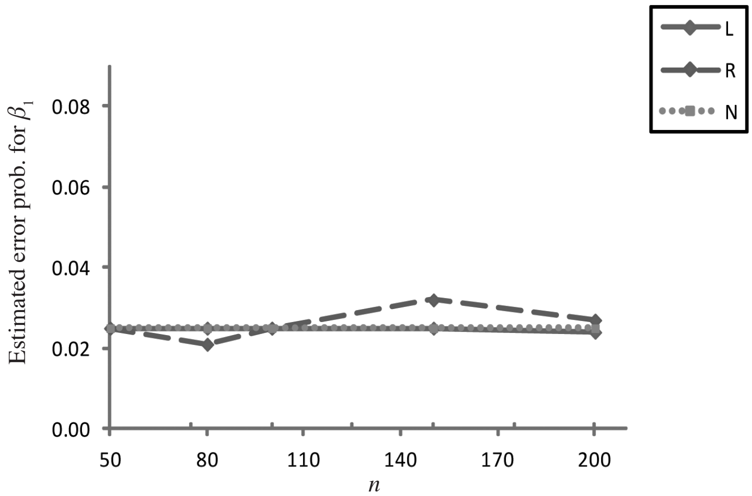

FIGURE 2. Estimated error probabilities for β_1 at $\alpha = 0.05$

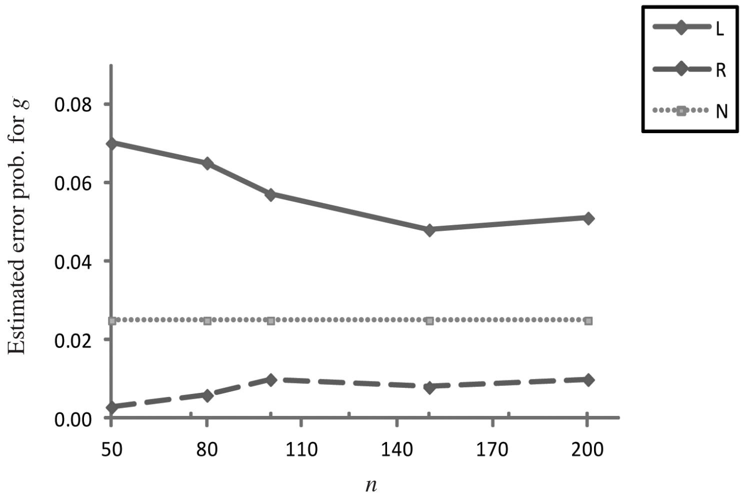

FIGURE 3. Estimated error probabilities for g at $\alpha = 0.05$

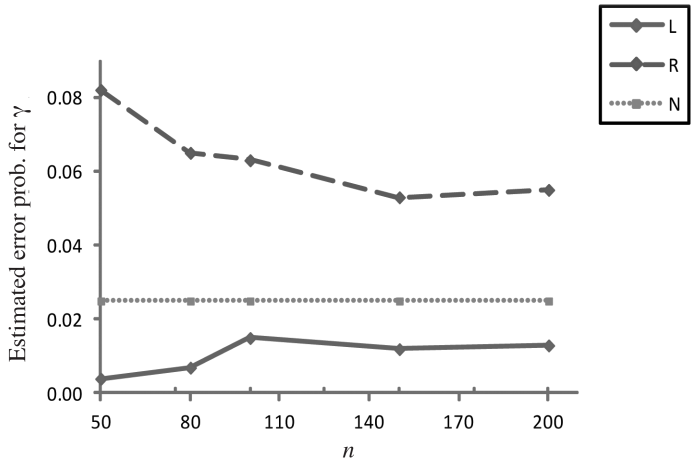

FIGURE 4. Estimated error probabilities for γ at $\alpha = 0.05$

TABLE 4. Summary of internal estimates at $\alpha = 0.05$

Parameter	Conservative	Anti Conservative	Asymmetrical
β_0	0	0	5
β_1	0	0	0
g	0	2	5
γ	0	4	5

TABLE 5. Summary of interval estimates at $\alpha = 0.05$ for different sample sizes

sample size	Conservative	Anti-conservative	Asymmetrical
50	0	2	3
80	0	2	3
100	0	1	3
150	0	0	3
200	0	1	3

time to failures of the pipe network. The failure intensity $\lambda(t) = e^{\beta_0+\beta_1 x_1+\beta_2 x_2+gt+\gamma(i-1)}$ is assumed where β_0, β_1, β_2, g and γ are the parameters and x_1 and x_2 are the covariates (diameter and length) and t is the failure time. Table 6 gives the value of the parameter estimates for this model.

TESTS FOR REPAIR AND COVARIATE EFFECTS USING LIKELIHOOD RATIO (LR) TEST

The likelihood ratio (LR) test is used to test the significance of adding an additional parameter to the model. 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 null hypothesis, H_0.

Let $\hat{\theta}_r$ be the maximum likelihood estimator of the restricted model under H_0 and $\hat{\theta}_f$ the maximum likelihood estimator of the full model. The maximized likelihood of the reduced model can never exceed the maximized likelihood of the full model 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. If l is the log-likelihood function, the likelihood ratio statistic for testing H_0 versus H_1 is the given by the following,

$$\Lambda = -2[l(\hat{\theta}_r) - l(\hat{\theta}_f)].$$

The LR statistics, Λ, follows the χ_k^2 distribution with k degree of freedom, where k is the number of parameters in the full model minus the number of parameters in the reduced model. The following test can be conducted to see if there is a significant repair effect,

H_0: No repair effect exists ($\gamma = 0$).

H_1: Repair effect exists ($\gamma \neq 0$).

TABLE 6. Estimates of model for pipe network failures

Parameter	Estimates	Std.errors	Wald intervals-90%	Wald intervals-95%
β_0	–5.52743	1.40145	(-7.83283,-3.22204)	(-8.27428,-2.78058)
β_1	–0.08334	0.18450	(-0.38685,0.22017)	(-0.44497,0.27829)
β_2	0.19457	0.24926	(-0.21547,0.60460)	(-0.29399,0.68312)
g	0.00133	0.00055	(0.00041,0.00224)	(0.00024,0.00241)
γ	–0.15379	0.06450	(-0.25989,-0.04769)	(-0.28020,-0.02737)

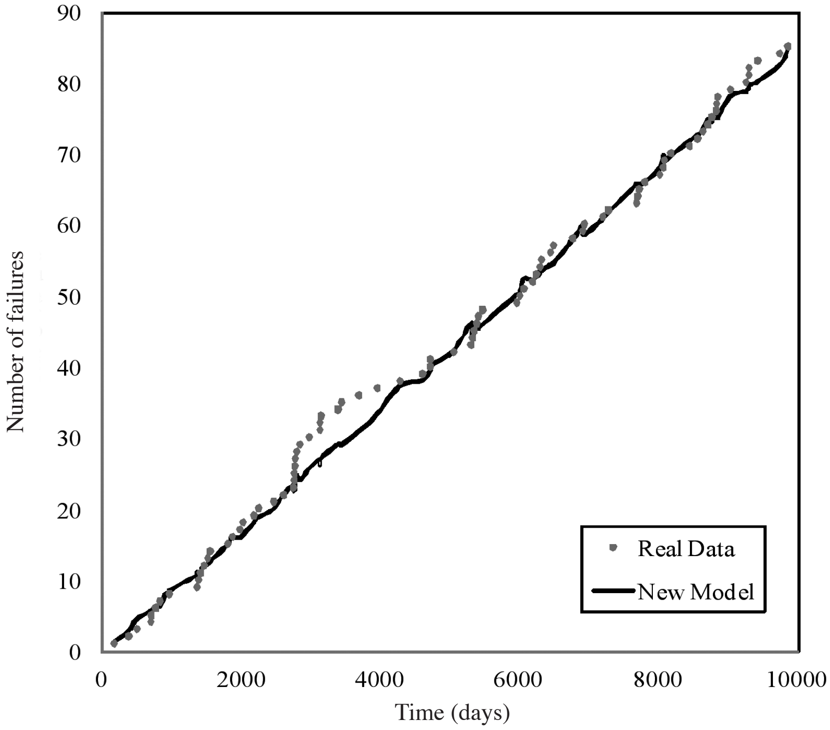

FIGURE 5. Estimates of the expected number of failures

Under the null hypothesis the test statistic $\Lambda = 6.172$ is greater than $\chi^2_{1,0.05} = 3.84$. Therefore, H_0 is rejected at $\alpha = 0.05$ and this means that the repair effect is significant.

Similarly, the following tests can be implemented to check the significance of the covariates, diameter

H_0: No covariate (diameter) effect exists ($\beta_1 = 0$).

H_1: Covariate (diameter) effect exists ($\beta_1 \neq 0$).

and length,

H_0: No covariate (length) effect exists ($\beta_2 = 0$).

H_1: Covariate (length) effect exists ($\beta_2 \neq 0$).

In the test involving the diameter(length) as covariate, $\Lambda = 4.202(4.587)$ and is greater than $\chi^2_{1,0.05} = 3.84$. This indicates that both diameter and length have significant effect on the pipefailures. These tests show the evidence of repair and covariate effects in the model. Therefore the full model can be utilized for the subsequent analysis. Figure 5 illustrates the estimates of the expected number of failure $m(t)$ for the data set with 2 covariates. It seems that the model fits the data quite well.

CONCLUSION

In this paper, we extended a general repair model based on repair (failure) history to incorporate fixed covariates. This model will be practical since it allows us to capture the effect of covariates in addition of capturing the time trend and repair effects thus enabling us to understand how they contribute to failures. The simulation study clearly showed that the model works well and is easy to use with no convergence or computational problems. The model also worked well when it was applied to real demo data from pipe failures where it indicated that the covariates, diameter and length have significant effect on pipe failures. This model will be very useful in real industrial applications because of its ability to capture time trend, repair and covariate effects simultaneously.

REFERENCES

Arasan, J. & Lunn, M. 2008. Alternative interval estimation for parameters of bivariate exponential model with time varying covariate. *Computational Statistics* 23: 605-622.

Arasan, J. & Lunn, M. 2009. Survival model of a parallel system with dependent failures and time varying covariates. *Journal of Statistical Planning and Inference* 139(3): 944-951.

Brown, C.C. 1975. On the use of indicator variables for studying the time-dependence of parameters in a response-time model. *Biometrics* 31: 863-872.

Cox, D.R. 1972. The statistical analysis of dependencies in point processes. In *Stochastic Point Processes,* edited by Lewis, P.A.W. New York: Wiley.

Cox, D.R. & Lewis, P.A.W. 1966. *The Statistical Analysis of Series of Events*. London: Methuen Press.

Crow, L.H. 1974. Reliability analysis for complex, repairable systems. In *Reliability and Biometry*, edited by Proschan, F. & Serfling, R.J. Philadelphia, PA: SIAM: 379-410.

Gasmi, S., Love, C.E. & Kahle, W. 2003. A general repair, proportional-hazards, framework to model complex repairable systems. *IEEE Transactions on Reliability* 52(1): 26-32.

Guo, H.R., Liao, H., Zhao, W. & Mettas, A. 2007. A new stochastic model for systems under general repairs. *IEEE Transactions on Reliability* 56(1): 40-49.

Høyland, A. & Rausand, M. 1994. *System Reliability Theory*. New York: Wiley Press.

Kaminskiy, M.P. & Krivtsov, V.V. 1998. A Monte Carlo approach to repairable system reliability analysis. In *Probabilistic Safety Assessment and Management*, edited by Mosleh, A. & Bari, R.A. New York: Springer.

Kijima, M. 1989. Some results for repairable systems with general repair. *Journal of Applied Probability* 26: 89-102.

Kijima, M. & Sumita, U. 1986. A useful generalization of renewal theory: Counting processes governed by non-negative Markovian increments. *Journal of Applied Probability* 23: 71-88.

Lawless, J.F. & Thiagarajah, K. 1996. A point-process model incorporating renewals and time trends, with application to repairable systems. *Technometrics* 38(2): 131-138.

Røstum, J. 2000. Statistical Modeling of Pipe Failures in Water Networks. PhD thesis, Norwegian University of Science and Technology (unpublished).

Vlok, P.J., Wnek, M. & Zygmunt, M. 2004. Utilising statistical residual life estimates of bearings to quantify the influence of preventive maintenance actions. *Mechanical Systems and Signal Processing* 18(4): 833-847.

Wang, H. & Pham, H. 1996. Optimal age-dependent preventive maintenance policies with imperfect maintenance. *International Journal of Reliability, Quality and Safety Engineering* 3(2): 119-135.

Yañez, M., Joglar, F. & Modarres, M. 2002. Generalized renewal process for analysis of repairable systems with limited failure experience. *Reliability Engineering & System Safety* 77(2): 167-180.

Samira Ehsani*
Department of Mathematics
Faculty of Science
University Putra Malaysia
43400, Serdang, Selangor D.E.
Malaysia

Jayanthi Arasan & Noor Akma Ibrahim
Laboratory of Computational Statistics and Operations Research
Institute for Mathematical Research
University Putra Malaysia
43400, Serdang, Selangor D.E.
Malaysia

*Corresponding author; email: ehsani_samira@yahoo.com

Received: 4 March 2010
Accepted: 2 February 2013