International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 22(2), November, 2022, pp 13-24

An Approach for the Estimation of Reliability Function for Two-dimensional Warranty Claims Data

Tahsina Aziz[1] and Md. Rezaul Karim[1*]

[1]Department of Statistics, University of Rajshahi, Rajshahi-6205, Bangladesh
Email: taziz_stat@yahoo.com and mrezakarim@yahoo.com

[*]Correspondence should be addressed to Md. Rezaul Karim
(Email: mrezakarim@yahoo.com)

[Received March 14, 2021; Accepted May 26, 2021]

Abstract

Warranty is a critical element in the implementation of marketing strategy and assurance to customers used by manufacturers. It also assures to customers that the manufacturer will provide compensation, through repair, replacement, or refund, for purchased products that fail within the warranty period. There are situations where two-lifetime variables are considered together for offering the warranty period. For example, for automobiles, sometimes warranty coverage has both age and mileage or usage limits, whichever occurs first. The warranty policy characterized by a region in a two-dimensional plane with one axis representing product age and the other axis representing product usage is known as a "two-dimensional" or "two-attribute" warranty policy. This paper aims to analyse a set of two-dimensional warranty claims data of a component of an automobile. It derives the joint probability models for age and usage and hence estimates the reliability function of the component. It also estimates the fractiles of the joint probability distribution that focus on the practical use of the information regarding the two-dimensional warranty limits of the component.

Keywords: Automobile component, Reliability, Two-dimensional warranty, Weibull model.

AMS Classification: 62N05, 90B25.

1. Introduction

Manufacturers use warranty services at the time of sale to indicate the high level of quality and reliability of their products. The warranty declaration ensures that the customer undertakes all or part of the cost of some specified conditional failures in a specified warranty period/region. For a one-dimensional warranty, this period only involves the one lifetime variable of the product and for a two-dimensional warranty, the period involves the two-lifetime variables simultaneously of the product. Generally, the larger warranty region indicates an attractive compensation to the customers and helps the manufacturers to increase the volume of sales. However, without having sufficient reliability of the product, the margin profit for manufacturers will hit severely, increase the warranty cost and also decrease the goodwill of the manufacturers. Therefore, it is important to determine the optimum warranty period for the product from both manufacturers' and customers' perspectives.

Much of the literature on warranty analysis considers one-dimensional warranty which is indexed by a single lifetime variable, such as age or usage. The age is measured by calendar time such as day, month, year, and so on, and the usage is measured by real operating time in terms of mileage, the number of copies, etc. However, there are situations where two-lifetime variables are considered together for offering the warranty period. For example, for automobiles, sometimes warranty coverage has both age and mileage limits, whichever occurs first (such as a five-year/50,000-mile protection plan). The warranty policy characterized by a region in a two-dimensional plane with one axis representing product age and the other axis representing product usage is known as a "two-dimensional" or "two-attribute" warranty policy (Karim and Suzuki, 2005). Moskowitz and Chun (1994) assumed that the number of events under the two-attribute warranty policies is distributed as Poisson and suggested a Poisson regression model. Lawless et al. (1995) discussed methods to model the dependence of failures on age and mileage and to estimate survival distributions and rates from warranty claims data using supplemental information about mileage accumulation. Singpurwalla and Wilson (1998) proposed an approach for developing probabilistic models indexed by two variables, time and amount of use, and applied these variables in an additive hazard model. Kim and Rao (2000) considered the two-dimensional warranty policy and implemented the expected warranty cost analysis based on a bivariate exponential distribution. Pal and Murthy (2003) applied Gumbel's bivariate exponential distribution for estimating the warranty cost of motorcycles under the two-dimensional warranty policy. More on two-dimensional warranty analysis can be found in Blischke and Murthy (1994), Mitra and Patankar (2010), Blischke, Karim and Murthy (2011), Gupta and Chatterjee (2014), Wang and Xie (2017), Muhammed and Almetwally (2020) and Lin and Chen (2021).

Automotive manufacturing companies utilized the warranty database as a prime source of field reliability data, which can be collected economically and efficiently through repair service networks. They analyze these data to enhance the quality and field reliability of their products and to improve customer satisfaction. In this paper, an approach is discussed for modelling the reliability of an automotive component based on two-dimensional warranty claims data. It derives the joint probability model for age and usage and hence estimates the reliability function of the component.

The outline of the paper is as follows. Section 2 describes the warranty claims data set of an automobile component. The paper analyses this data set. Section 3 presents the preliminary analysis results of the data set. Sections 4 and 5 derive the joint probability density function and reliability function, respectively. Section 6 concludes the paper with a discussion and possible implementation issues for future research.

2. Description of Data

In this paper we consider a set of warranty claims data for an automobile component. The component was produced during one year, sold during 26-month period and warranty claims were recorded during four years observational period under the two-dimensional warranty with age=18 months (age limit) and usage=100000 km (usage limit). Therefore, the warranty region can be represented by a rectangular constructed by the points (0, 0) and (18, 100000) and the failed components are replaced free of charge if they failed within this region. There are 4240 failed observations and 45760 censored observations for the component. The main limitation of the database is that for all failed observations the age and usage are known but for any censored observation the age is known whereas the usage is unknown. The structure of the age-based aggregated warranty claims and censored data are shown in Table 1, where t, d_t and r_t denote the age in month, number of units failed at t and number of units that are right-censored at t, respectively. The usage of each failed component is also given in the warranty database.

Table 1: Aggregated warranty claims data of the automobile component

Age in month (t)	No. of failures at t (d_t)	No. of censored at t (r_t)
1	d_1	r_1
2	d_2	r_2
⋮	⋮	⋮
18	d_{18}	r_{18}
Total	n=4240	r=45760

3. Preliminary Analysis of Data

Two-lifetime variables, Age denoted by T, and Usage denoted by X, are considered in this paper for modelling the reliability of the component. The descriptive statistics of these two variables based on only failure data are shown in Table 2. In Table 2, the measurement units for Age and Usage are respectively, month and kilometer (km). The descriptive statistics given in Table 2 are conditional estimates in the sense that they are estimated based on items that failed during the warranty period and led to claims but the censored observations are ignored. This means the summary statistics for the variables Age and Usage given that the Age is less than or equal to 18 months and Usage is less than or equal to 100000 kilometers. Table 1 indicates that the conditional average and median lifetimes with respect to Age are 10.768 and 11.0 months and Usage are 27001 and 25128 kilometers. In the case of the Usage variable, the mean exceeds the median, indicating skewness to the right. On the other hand, Age variable shows negative skewness.

Table 2: Descriptive statistics of lifetime variables for n=4240 observations

Statistics	Variable	
	Age (T)	Usage (X)
Mean	10.768	27001
Standard deviation	4.588	15938
Minimum	1	26
First quartile	7	15665
Median	11	25128
Third quartile	15	36381
Maximum	18	96394
Skewness	-0.33	0.82
Kurtosis	-0.82	0.96

Figure 1 shows the marginal plot of Usage versus Age, which is a scatterplot with histograms in the margins of the *X*- and *Y*-axes. This figure can be used to assess the relationship between two variables and examine their distributions.

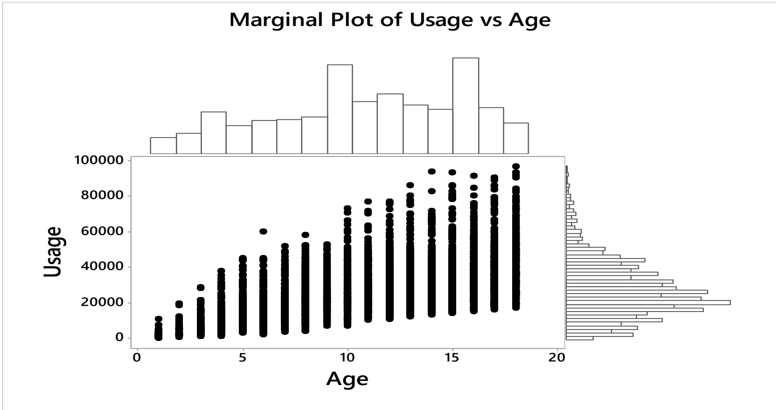

Figure 1: Marginal plot of Usage versus Age

The scatterplot in Figure 1 indicates a positive correlation between Age and Usage; that is, the Usage increases as Age increases. The numerical value of the Pearson correlation coefficient between Age and Usage is 0.603 with the p-value of 0.000 for the hypothesis test of the correlations being zero indicates sufficient evidence at α=0.01 that the correlation is not zero. The marginal distributions have clusters of points (about 19000 - 21000 for Usage and about 15 - 16 for Age).

4. Joint Probability Density Function

In this section, we derive the bivariate joint probability density function (pdf) and joint cumulative density function (cdf) of the lifetime variables Age (*T*) and Usage (*X*). To find the joint pdf, first, we find the suitable distribution of *T* and then derive the conditional distribution of *X* given Age *T*. It is observed that the Weibull distribution can be considered as the best fitted distribution for *T* among the four distributions, Weibull, Lognormal, Exponential and Normal, in the sense that it gives the smallest value of the adjusted Anderson-Darling (AD) statistic. The AD values for the competitive four distributions are for Weibull = 93655.506, Lognormal = 93655.555, Exponential = 93656.762, and for Normal = 93655.904. In this case, the variable *T* with its corresponding frequency and failure/censored indicator (δ_t=1 for failure and δ_t=0 for censored at age *t*) are considered.

If η and β denote respectively the scale and shape parameters of the Weibull distribution, then the pdf of *T* can be written as,

$$f(t;\eta,\beta) = \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right];\ t \geq 0,\ \eta,\beta > 0 \tag{1}$$

We apply the maximum likelihood estimation method to estimate the parameters of the model. The maximum likelihood estimates (MLE) of the parameters of Weibull distribution (1) are

$\hat{\eta} = 82.2409$ and $\hat{\beta} = 1.57953$. The MLE of the shape parameter of Weibull distribution greater than one indicates an increasing failure rate (IFR) with respect to the Age of the component.

Next, we find the conditional distribution of Usage (X) given Age (T), $f(x|T=t)$. It is noted that for a given Age in month, the number of censored observations are known but the censored usage are unknown in the warranty database. Therefore, to estimate the censored usage and hence the conditional distribution, $f(x|T=t)$, we propose the following step-by-step procedure.

Step 1: Find the suitable probability distributions for the usages (X) of failed components separately for each age in months, t, t=1, 2, …, 18. Here we assume the Weibull distributions with the scale and shape parameters, η_t^0 and β_t^0, for t=1, 2, …, 18.

Step 2: Find the maximum likelihood (ML) estimates of the parameters, η_t^0 and β_t^0, for t=1, 2, …, 18 of Weibull distribution. These estimates are based on only failure usages for each age.

Step 3: Derive the relationships between the ML estimates of both parameters and the age in month. We assume the linear relationships which are shown in Figure 2. Figure 2 indicates that the scale parameter is a linear function and the shape parameter can be expressed roughly as linear of the age in month. According to Figure 2, the relationship between the shape parameter β_t^0 and age (t) can be approximated as

$$\beta_t^0 = 1.39314 + 0.080058 \times t \tag{2}$$

Similarly, the relationship between the scale parameter η_t^0 and age (t) can be approximated as

$$\eta_t^0 = 3407.634 + 2496.967 \times t \tag{3}$$

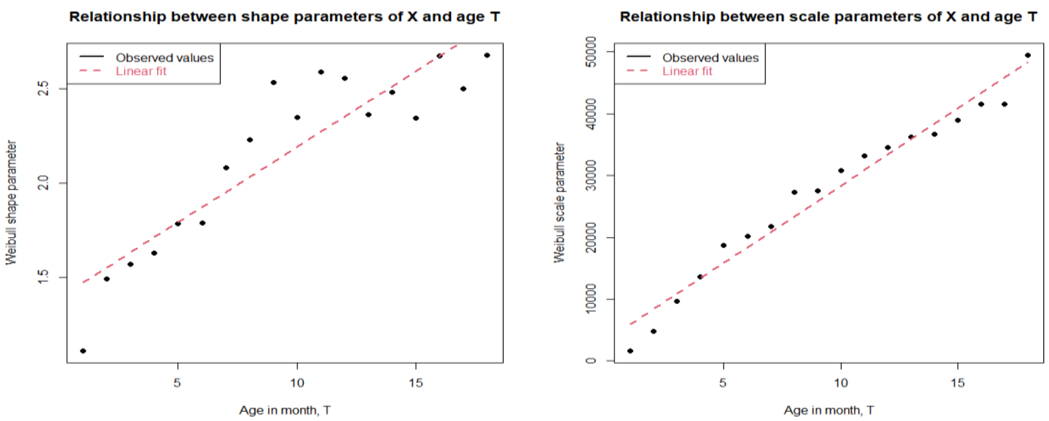

Figure 2: Linear relationships for MLEs of the parameters and age for failure data only

Step 4: Generate Weibull random variates X_t of size ($n_t + r_t$) with parameters β_t^0 and η_t^0, given in (2) and (3), for t=1, 2, …, 18, and arrange the values of each X_t in ascending order. Repeat this for a large number of times, e.g., k =10000 times, compute averages of k sets of ordered X_t and denote it by X_t^*, t=1, 2, …, 18.

Step 5: Divide X_t^* in two sets, the first set (A_{ft}) with the smallest n_t and the second set (A_{ct}) with the largest r_t observations, for t=1, 2, …, 18. Assume the second set (A_{ct}) as the estimated censored usages. Here it is assumed that for a given age, the distributions of the failure and censored usages are the same but the mean usage of censored units would be higher than that of the mean usage of failure units.

Step 6: Fit Weibull distributions and estimate parameters based on the n_t observed failure usage and r_t censored usage (set A_{ct}, estimated in Step 5) for each age t=1, 2, …, 18.

Step 7: Like Step 3, derive the relationships between the estimates of both parameters and the age in month. Here we again assume the linear relationships which are shown in Figure 3. According to Figure 3, the relationship between the shape parameter β_t and age (t) and the relationship between the scale parameter η_t and age (t) can be approximated[1], respectively as,

$$\beta_t = 1.07298 + 0.09297 \times t \tag{4}$$

and

$$\eta_t = \text{-}1783.000 + 4386.329 \times t \tag{5}$$

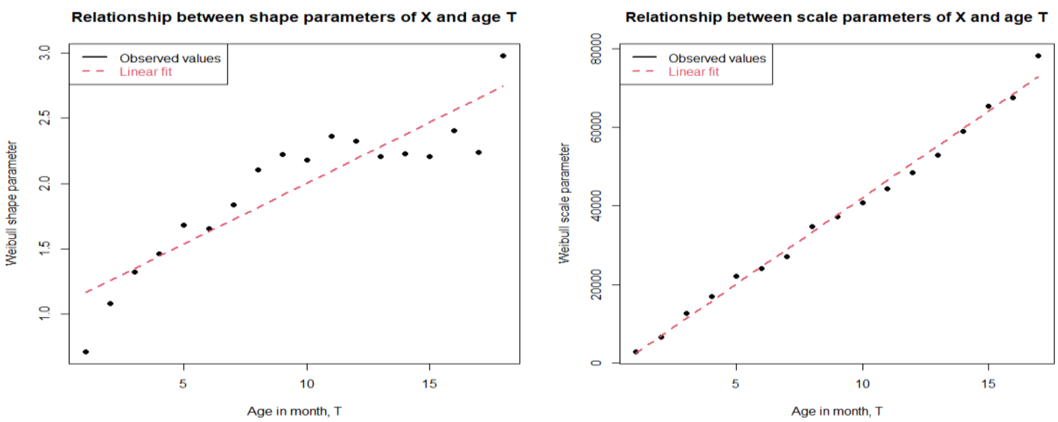

Figure 3: Linear relationships for MLEs of the parameters and age for failure and censored data

Step 8: Consider the conditional distribution of usage given age, $f(x|T=t)$, as Weibull with shape parameter β_t and scale parameter η_t. That is, the pdf of X given T is

$$f(x|T=t;\eta_t,\beta_t) = \frac{\beta_t}{\eta_t}\left(\frac{x}{\eta_t}\right)^{\beta_t-1} \exp\left[-\left(\frac{x}{\eta_t}\right)^{\beta_t}\right]; \quad t, x \geq 0,\ \eta_t, \beta_t > 0 \tag{6}$$

[1] To derive the relationship for the scale parameter, the estimate of the scale parameter for age 18 months has been excluded assuming it is an outliner as it gives a very large value compared with others because of having a huge number of censored observations for this age.

After implementing the above eight-steps, the joint probability density function for T and X can be obtained by using the marginal and conditional density functions as follows.

$$f(t,x) = f(x \mid T = t) \times f(t), \quad t, x \geq 0 \tag{7}$$

Inserting the density functions (1) and (6) in (7), we get

$$\begin{aligned} f(t,x) &= \frac{\beta_t}{\eta_t}\left(\frac{x}{\eta_t}\right)^{\beta_t - 1} \exp\left[-\left(\frac{x}{\eta_t}\right)^{\beta_t}\right] \times \frac{\beta}{\eta}\left(\frac{t}{\eta}\right)^{\beta - 1} \exp\left[-\left(\frac{t}{\eta}\right)^{\beta}\right] \\ &= \frac{\beta_t \beta}{\eta_t \eta}\left(\frac{x}{\eta_t}\right)^{\beta_t - 1}\left(\frac{t}{\eta}\right)^{\beta - 1} \exp\left[-\left\{\left(\frac{x}{\eta_t}\right)^{\beta_t} + \left(\frac{t}{\eta}\right)^{\beta}\right\}\right]; t, x \geq 0 \end{aligned} \tag{8}$$

where the two shape parameters $\beta, \beta_t > 0$ and two scale parameters $\eta, \eta_t > 0$. The maximum likelihood estimates of the parameters β and η are obtained using the failure and censored data likelihood for T derived based on the pdf (1) and β_t and η_t are obtained from (4) and (5).

A plot of the joint probability density function $f(t,x)$ given in (8) is shown in Figure 4.

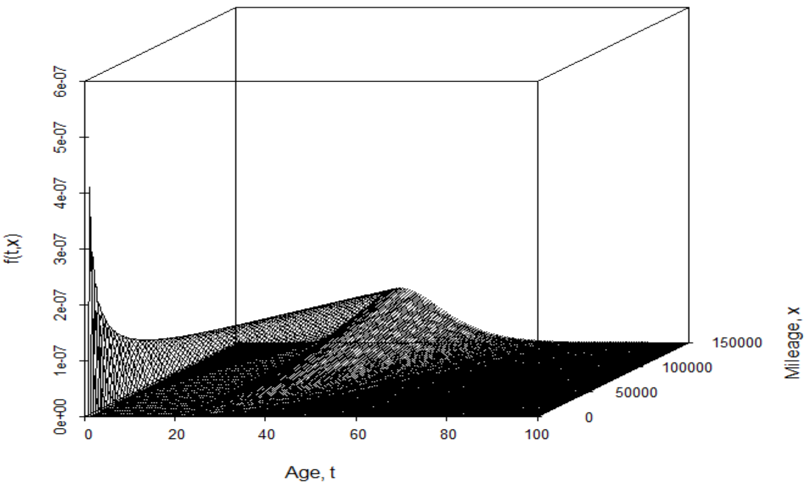

Figure 4: Plot of joint pdf of T and X, $f(t,x)$

According to (8), the probability of failure of a component in the interval ($t_{i-1} < T \leq t_i$, $x_{j-1} < X \leq x_j$) is

$$\begin{aligned} \Pr\left(t_{i-1} < T \leq t_i,\ x_{j-1} < X \leq x_j\right) &= \int_{t_{i-1}}^{t_i} \int_{x_{j-1}}^{x_j} f(t,x)\,dtdx \\ &= F(t_i, x_j) - F(t_i, x_{j-1}) - F(t_{i-1}, x_j) + F(t_{i-1}, x_{j-1}) \end{aligned} \tag{9}$$

where for this data set i=1, 2, …, I; j=1, 2, …, J; I and J denote the number of intervals for T and X, respectively, with t_0=0, x_0=0, and $F(t,x)$ denotes the joint cumulative density function of T and X defined by

$$F(t,x)=\Pr\left(T\le t,\ X\le x\right)=\int_0^t\int_0^x f(u,v)dudv \tag{10}$$

If n denotes the total number of failures, the expected number of failures for the component (n_{ij}) in the ijth interval ($t_{i-1}<T\le t_i$, $x_{j-1}<X\le x_j$) for the variables lifetime variables T and X can be estimated by using (9) as

$$n_{ij}=n\times\Pr\left(t_{i-1}<T\le t_i,\ x_{j-1}<X\le x_j\right),\text{for } i=1,2,...,18\ \&\ j=1,2,...,6 \tag{11}$$

5. Joint Reliability Function

The joint reliability function of T and X becomes

$$R(t,x)=\Pr\left(T>t,\ X>x\right)=\int_t^\infty\int_x^\infty f(u,v)dudv=1-F(t,x), t\ge 0, x\ge 0. \tag{12}$$

Note that the closed-form solutions of the cdf $F(t,x)$ and the reliability function $R(t,x)$ cannot be obtained. To estimate these functions numerically, we apply the "integral2( )" function given in the R program.

To compare the nonparametric and parametric estimates of reliability function numerically, we make 10 groups of Usage with 10,000 kilometers intervals, given in Table 3.

Table 3: Ten different groups of Usage

Group No.	Usage limit	No. of observations
1	0 – 10000	534
2	10000 - 20000	981
3	20000 – 30000	1130
4	30000 – 40000	758
5	40000 - 50000	523
6	50000 - 60000	156
7	60000 - 70000	81
8	70000 - 80000	52
9	80000 - 90000	17
10	90000 - 100000	8
	Total	4240

For testing the equality of the mean ages in month for ten uses groups, we use the analysis of variance (ANOVA) procedure and the output is given in Table 4.

Table 4: ANOVA Table for Age with ten groups of Usage

Source	DF	Adj SS	Adj MS	F-Value	p-Value
Usage group	9	40815	4534.97	396.08	0.000
Error	4230	48431	11.45		
Total	4239	89246			

In the ANOVA Table 4, the p-value (0.000) indicates that there is sufficient evidence that not all the means of the Ages are equal when alpha is set at 0.05. The means and confidence intervals (CI) of Ages for ten different Usage groups are given in Table 5.

Table 5: Means and confidence intervals (CI) of Ages for ten Usage groups

Usage group	*n*	Mean	Standard deviation	95% CI
1	534	3.633	2.299	(3.3459, 3.9200)
2	981	9.521	3.840	(9.3090, 9.7330)
3	1130	11.693	3.577	(11.496, 11.890)
4	758	12.649	3.490	(12.408, 12.890)
5	523	13.208	3.183	(12.918, 13.498)
6	156	14.487	3.024	(13.956, 15.018)
7	81	14.827	2.355	(14.096, 15.564)
8	52	15.423	2.396	(14.503, 16.343)
9	17	15.824	1.629	(14.215, 17.432)
10	8	16.750	1.581	(14.405, 19.095)

The confidence intervals for groups 6 to 10 indicate that they are overlapping. Therefore, next, we merge the groups 6 to 10 and perform the ANOVA again based on the new six groups, where the new group No. 6 means the merged of the old groups 6-10. In this case, the p-value (0.000) again indicates that there is sufficient evidence that not all the means of the Ages are equal when alpha is set at 0.05, and the means and confidence intervals of Ages for six different Usage groups are given in Table 6.

Table 6: Means and confidence intervals (CI) of Ages for six Usage groups

Usage group	*n*	Mean	Standard deviation	95% CI
1	534	3.633	2.299	(3.3458, 3.9201)
2	981	9.521	3.840	(9.3090, 9.7330)
3	1130	11.693	3.577	(11.496, 11.890)
4	758	12.649	3.490	(12.408, 12.890)
5	523	13.208	3.183	(12.918, 13.499)
6	314	14.860	2.708	(14.485, 15.234)

Table 6 shows that all the confidence intervals are distinct (non-overlapping). Hence the next analyses (to find the conditional distribution of Age given Usage), will be conducted based on these six Usage groups.

To evaluate the performance of the proposed method, we estimate the parametric reliability function (12) by applying the maximum likelihood estimation method and compare it with the nonparametric estimate of the reliability function which can be easily estimated from the observed data. Figure 5 compares these estimates.

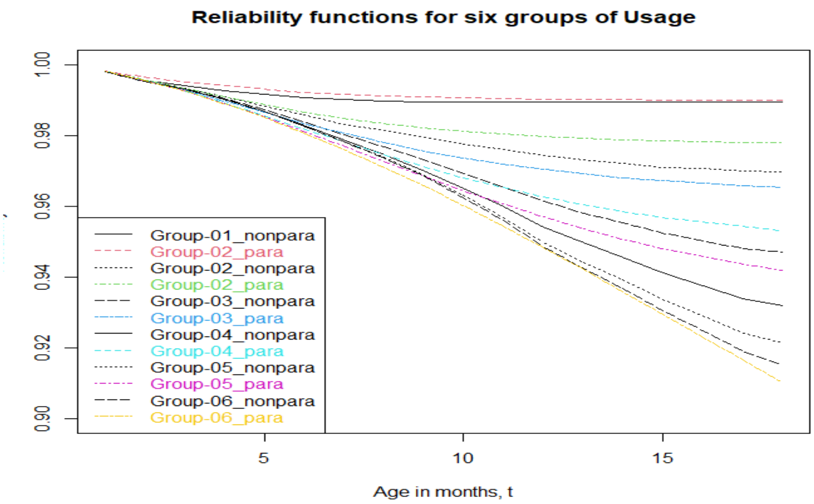

Figure 5: Plots of nonparametric and parametric reliability functions for six groups of usage

Figure 5 indicates that the nature and estimates of the parametric reliability functions (colour lines) are approximately similar to the nonparametric reliability functions (black lines) for the six groups of usage. This implies the applicability of the proposed method. Parametric estimates would be more close to the nonparametric estimates if much better approximations of the parameters, Eqs. (2) (5), can be used.

5.1 Fractiles of Distribution

The p-fractile of the continuous bivariate probability distribution, $F(t,x)$, is any pair of values of the random variables T and X, call (t_p, x_p), such that $F(t_p, x_p) = p$, where $0 \leq p \leq 1$. The related terms of fractile are percentile, decile, and quantile. The p-fractile of a sample indicates that at least a proportion p of the sample lies at or below the values (t_p, x_p) of the random variables T and X, and at least a proportion (1- p) lies at or above of the values (t_p, x_p). The estimating equation for (t_p, x_p) can be expressed as $\{t_p, x_p\} = F^{-1}(p, p)$, where $F^{-1}(.,.)$ denotes the inverse function of the cumulative distribution function $F(.,.)$.

Figure 6 illustrates the quantiles for p=0.05 (left-hand side) and p=0.10 (right-hand side). There are many possible pairs of values for $\{t_p, x_p\}$. For example, for p=0.05, some possible values are $\{t_p, x_p\}$={(15, 51000), (18, 44000), (24, 40000)}, and for p=0.10, some possible values are $\{t_p, x_p\}$={(24, 80000), (30, 72000), (36,70000)}. These estimates and Figure 6, as for example, indicate that 5% of the component will fail by 18 months and 44000 kms and 10% will fail by 30 months and 72000 kms. The estimates of fractiles can be used for deciding on the suitable two-dimensional warranty limits of the component.

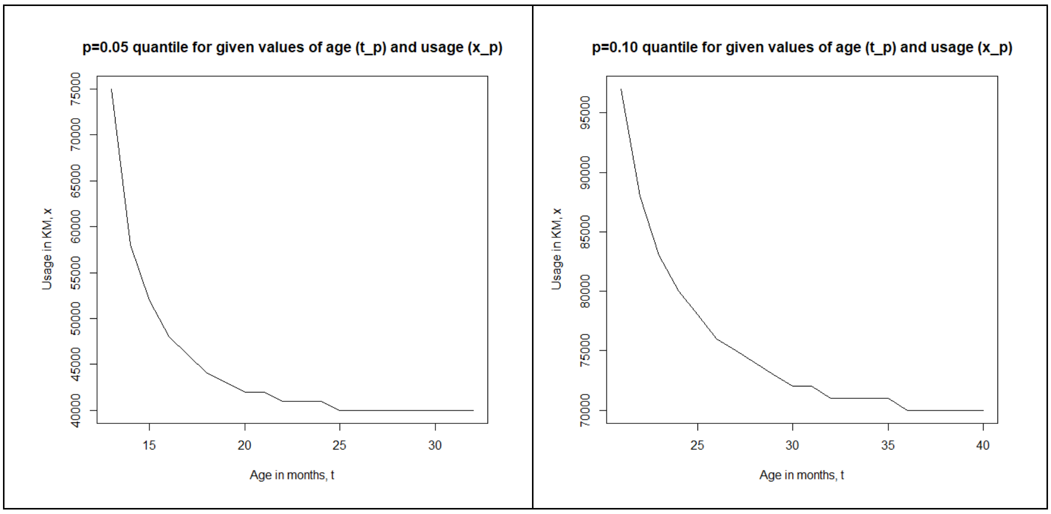

Figure 6: Plots of 0.05-fractile (left-hand side) and 0.10-fractile (right-hand side)

6. Conclusions

This paper analyzed a set of two-dimensional warranty claims data of a component of an automobile, where two lifetime variables, age and usage, are considered together for offering the warranty period. It proposed a method to derive the approximated joint distribution of age and usage and then this distribution is applied to estimate the reliability function of the component. The estimates of fractiles of the bivariate lifetime distribution would be useful to the manufacturer for selecting suitable two-dimensional warranty limits and deciding on the optimum maintenance policy for the component. A comparison of the parametric reliability function with that of the nonparametric estimate implies the applicability of the proposed method. The relationships between the ML estimates of both scale and shape parameters and the age in month are assumed linear because of easy computation and interpretation. However, better approximations would improve the applicability of the method.

An extension of the method concerning more lifetime distributions would be valuable. Also, simulation studies can be performed for investigating the performance of the approach.

Acknowledgment

The authors wish to thank the anonymous reviewer for careful reading the manuscript and constructive comments and suggestions.

References

[1] Blischke, W. R., Karim, M. R., Murthy, D. N. P. (2011). Warranty Data Collection and Analysis. Springer-Verlag London Limited.

[2] Blischke, W. R., Murthy, D. N. P. (1994). Warranty cost analysis. Marcel Dekker, New York.

[3] Gupta, S. K., De, S. and Chatterjee, A. (2014). Warranty forecasting from incomplete two-dimensional warranty data. Reliability Engineering & System Safety, 126, 1–13.

[4] Karim, M. R., and Suzuki, K. (2005). Analysis of warranty claim data: a literature review, International Journal of Quality & Reliability Management, Vol. 22, No. 7, pp. 667-686.

[5] Kim, H. G. and Rao, B. M. (2000). Expected warranty cost of a two-attribute free-replacement warranties based on a bi-variate exponential distribution. Comput Ind Eng 38:425–434.

[6] Lawless, J. F., Hu, J. and Cao, J. (1995). Methods for the estimation of failure distributions and rates from automobile warranty data. Lifetime Data Anal 1:227–240.

[7] Lin, K. and Chen, Y. (2021). Analysis of two-dimensional warranty data considering global and local dependence of heterogeneous marginals. Reliability Engineering & System Safety, 207, 107327. doi: 10.1016/j.ress.2020.107327.

[8] Mitra, A. and Patankar, J. G. (2010). An integrated multiobjective model for two-dimensional warranty policies, Lawrence, K.D. and Kleinman, G. (Ed.) Applications in Multicriteria Decision Making, Data Envelopment Analysis, and Finance (Applications of Management Science, Vol. 14), Emerald Group Publishing Limited, Bingley, pp. 35-57.

[9] Moskowitz, H. and Chun, Y. H. (1994). A Poisson regression model for two-attribute warranty policies. Nav Res Logist 41:355–376.

[10] Muhammed, H. Z. and Almetwally, E. M. (2020). Bayesian and Non-Bayesian Estimation for the Bivariate Inverse Weibull Distribution Under Progressive Type-II Censoring. Ann. Data Sci. doi: 10.1007/s40745-020-00316-72020.

[11] Pal, S. and Murthy, G. S. R. (2003). An application of Gumbel's bivariate exponential distribution in estimation of warranty cost of motorcycles. Int J Qual Reliab Manag 20: 488–502.

[12] Singpurwalla, N. D. and Wilson, S.P. (1998). Failure models indexed by two scales. Adv Appl Probab 30: 1058–1072.

[13] Wang, X. and Xie, W. (2017). Two-dimensional warranty: A literature review. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 232(3), 284–307.