Customer-rush Near Warranty Expiration Limit and Nonparametric Hazard Rate Estimation from Known Mileage Accumulation Rates

Bharatendra Rai and Nanua Singh, Industrial Engineering, Wayne State University

Time or mileage data obtained from warranty claims are generally more accurate for hard failures than for the soft failures. For soft failures automobile users sometimes delay reporting of warranty claims till the warranty coverage is about to expire. This results in an unusually high number of warranty claims near the end of warranty coverage. Since such a phenomena of customer-rush near the warranty expiration limit occurs due to user behavior rather than the vehicle design, it creates a bias in the warranty dataset. Design improvement activities that use field reliability studies based on such data can potentially obtain a distorted picture of the reality and lead to unwarranted and costly design changes. Research in the area of field reliability studies using warranty data provides several methods for warranty claims resulting from hard failures and assumes reported time or mileage as actual time or mileage at failure.

In this article we provide a methodology to obtain nonparametric hazard rate estimates addressing the phenomena of customer-rush near the warranty expiration limit that occurs due to soft failures. The methodology involves situations where estimates of mileage accumulation rates in the vehicle population are available. The claims influenced by soft failures are treated as left-censored and are identified using information in technician comments about the repair carried out and, if required, a more involved engineering analysis of field returned parts. Maximum likelihood estimates for the hazard function and their confidence limits are then obtained using Turnbull's iterative procedure. An application example illustrates use of the proposed methodology.

Keyword: Customer-rush, mileage accumulation rate, soft failures, hazard function, warranty data.

ACRONYMS

EGR Exhaust Gas Recirculation
MIS Months In Service
MLE Maximum Likelihood Estimator
PDF Probability Density Function
LN Lognormal

NOTATION

L_M Limit for warranty mileage
L_T Limit for warranty time
T Months in service ($t = 1, 2, \ldots, L_T$)
N Total number of vehicles in the field

V_t Number of vehicles in the field up to MIS = t
$N(t)$ Number of vehicles in the field without any claim at the beginning of MIS = t
a Mileage value below which warranty data is artificially truncated
L_M -b Mileage value above which warranty data is artificially truncated
n_t Number of first claims in the mileage interval (a, L_M-b) miles at MIS=t with $0<a<b<L_M$
n_t^* Number of first warranty claims at t
c_t Number of left-censored first claims at MIS = t
r_t Number of vehicles that have not yet completed t MIS
W Random variable denoting miles/month mileage accumulation rate in vehicle population
Y_t Random variable denoting miles driven by a vehicle at MIS = t
$\mathcal{L}$ Loglikelihood function
$R(t)$ Reliability function at MIS = t
$h(t)$ Hazard function
$H(t)$ Cumulative hazard function

1. Introduction

In reliability studies using automobile warranty data, the time or mileage at the time of warranty claim is assumed to be the time or mileage at failure (Suzuki 1985a, 1985b). Similarly an absence of warranty claim is treated as 'no failure' situation. Kalbfleisch, Lawless and Robinson (1991), Lawless, Hu and Cao (1995), Lu (1998), Kalbfleisch and Lawless (1992), Stephens and Crowder (2004), and Rai and Singh (2003, 2005) also provide warranty data analysis methods with examples based on such assumption.

However, the process leading to a warranty claim involves several factors. Two key factors being the type of failure and the chance of such failures getting detected by a user of the vehicle. Fig. 1. depicts four situations based on these two factors.

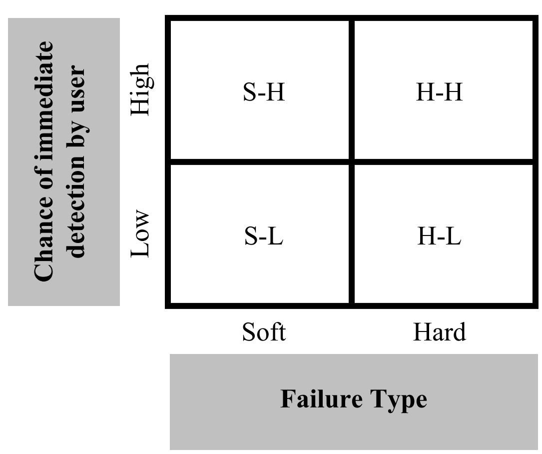


Fig. 1. Failure type and chance of detection by user

As shown in Fig. 1., the vehicle failures experienced by users can be mainly classified into two types viz., soft failures and hard failures. Soft failures are those that result into degraded performance, but the vehicle can still be operated. Some examples of soft failure are minor oil leaks, engine slow to start, unusual engine noise, etc. At the same time the pattern of reported failures in the warranty data may also be influenced by the chance of immediate detection of failure by the user. S-L represents the soft failures with a low chance of detection by the users that may or may not get reported and result into a warranty claim. Such soft failures may gradually aggravate over usage till the chance of detection becomes high as represented by S-H. Once a vehicle user detects a S-L or S-H type of failure, the user may decide to report it immediately or at a convenient date. Thus, for such soft failures time to claim minus time to failure is likely to be greater than zero.

Similarly hard failures are those that make a sub-system of a vehicle inoperative or unusable until repaired. For example, 'engine doesn't start' or 'engine stops running' are hard failures at vehicle level. H-L represents hard failures with a low chance of immediate detection by users. For example, the failure of one of the rear brake lights may be detected at the time of visit to the dealer for some other repair. Thus time to claim minus time to failure in such cases too could be greater than zero. H-H represents hard failures that have high chance of detection by the user, as a vehicle is unusable until repaired. Also due to largely immediate reporting at a repair center, time to claim minus time to failure is approximately zero, leading to a more accurate failure time or mileage data in warranty claims. However, it is to be noted that even hard failures with a high chance of detection by user may not be reported for claim immediately. For example, immediate reporting of 'engine stops running' failure may depend on whether the vehicle could be restarted or not after experiencing the failure.

S-L and S-H category of failures may lead to a situation where many such failures are not reported until the warranty coverage is about to expire. This sometimes leads to the occurrence of unusually high number of claims near the end of warranty coverage period. Such a phenomena is also termed as customer-rush near the warranty expiration limit. Iskandar and Blischke (2003) while carrying out reliability analysis using warranty data from motorcycle claims, point out delay in reporting of non-critical failures. They suggest that such delay in reporting may be modeled by fitting a mixture of distributions to the usage data. They, however, do not pursue such analysis in the paper. Rai and Singh (2004) and Rai (2004) provide a methodology to arrive at estimates of nonparametric hazard rates from warranty data that is biased due to the presence of customer-rush near the warranty expiration limit. The methodology used addresses situations when the estimates of mileage accumulation rates in the vehicle population, are not available.

This paper provides a methodology to arrive at nonparametric hazard rates based on the time to first failure for situations when the estimates of mileage accumulation rates in the vehicle population are available and the warranty data is influenced by the customer-rush phenomena. The time to first warranty claim is used in arriving at the hazard rate estimates to keep design or manufacturing quality separate from service quality (Majeske et al. 1997, Hu et al. 1998).

The remainder of the paper is organized as follows. In Section 2 we examine censoring, truncation, and certain other issues that characterize automobile warranty data and the role of hazard function. In Section 3 we discuss modeling of mileage accumulation rates in the vehicle population. We then propose a nonparametric methodology for obtaining maximum likelihood estimates for hazard rate along with their confidence limits in Section 4. We illustrate the use of the proposed methodology in Section 5 using an application example.

2. Censoring and Truncation In Automobile Warranty Data

2.1 Characterization of Warranty Data

When a warranty claim is made, vehicle or failure related information/data on over fifty different kinds are recorded in the warranty dataset. Some examples include production date, sale date, repair date, repair cost, repair time, mileage, etc. Two variables that are important in field reliability studies are months in service (MIS) of the vehicle from the date of sale, and mileage at the time of warranty claim. Fig. 2. shows a plot of MIS versus mileage data based on 924 first claims for a component level failure mode of a vehicle. The plot also gives histogram of frequency of vehicles in 1K (‘K’ represents mileage in 000s) mileage band increments and histogram of frequency of vehicles that failed at each MIS.

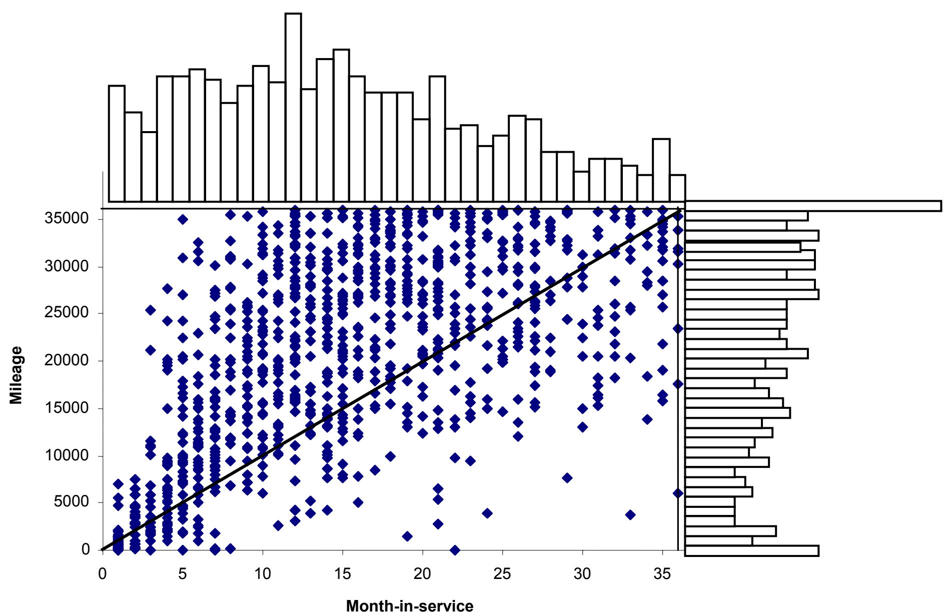


Fig. 2. MIS versus mileage plot of first warranty claims

It is observed from Fig. 2. that all the data points lie within the warranty period of 36 MIS and 36K miles. It is also observed that after about MIS = 10, many vehicles are likely to exceed the warranty mileage limit. When the non-failed vehicles within the warranty limits eventually fail outside the mileage limit, neither the number nor the mileage at first failure is known for the failures. Such datasets are said to be right truncated (Meeker and Escobar 1998, Nelson 1990, Cohen 1991).

The right side of the Fig. 2. shows ‘spikes’ in first warranty claims both at the beginning and toward the end of warranty mileage limit. The spike observed in the last mileage band of the histogram i.e., 35K-36K miles could be attributed to the customer-rush near the warranty limit (Rai and Singh 2004, Rai 2004). Such a phenomena is expected to occur especially for soft failures where vehicle users delay failure reporting till the warranty coverage is about to expire. Although actual failure for first claims affected by the customer-rush phenomena are expected to have occurred prior to the reported time, the actual MIS or mileage at the time of failure is not known. Such data are said to be left-censored.

The spike at the beginning can be attributed to the existence of manufacturing/assembly defects in addition to the usage related failures (Rai and Singh 2004, Majeske 2003). If the first claims below say *a* miles are screened out from further analysis to avoid mix-up between manufacturing/assembly defects and usage related failures, the dataset becomes left-truncated at each MIS value. Datasets that are both left and right truncated are said to be doubly truncated datasets.

Warranty data are known to be messy and unclean due to reasons including data entry errors, incorrect binning of claims, and inaccurate reporting of failures (Iskandar and Blischke 2003, Pal and Murthy 2003, Sander et al. 2003, Suzuki et al. 2001). Thus it is always helpful to screen the warranty data before undertaking any detailed statistical analysis.

2.2 Hazard Function

In field reliability studies hazard function denoted by *h(t)* plays an important role. Hazard function provide estimates of the distribution parameters, the proportion of units failing by a given age, percentiles of the distribution, the behavior of the failure rate as a function of their age, and conditional failure probabilities for units of any age (Blischke and Murthy 2000, Nelson 2000). Davis (2003) further notes that increasing, decreasing, or constant failure rate pattern obtained from hazard plot provides important clues for improving reliability and robustness of products.

For a continuous nonnegative random variable T representing lifetimes of individual items, the hazard function is defined as

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \le T \le t + \Delta t \mid T \ge t)}{\Delta t} \tag{1}$$

where Δt denotes the width of a very small time interval. The hazard function specifies the instantaneous rate of failure at time t, given that the item survives to time t. In other words, $h(t)\,\Delta t$ is the probability of failure in [t, t+Δt), given the item survives to time t.

Let N be the total number of vehicles in the field. Let n_t^* denote the number of first warranty claims at MIS = t and $N(t)$ denote the number of vehicles in the field without any claims at the beginning of MIS = t. Also let V_t denote the total number of vehicles in the field up to MIS = t. Assuming an unlimited warranty mileage limit for the vehicle population under study and a warranty time limit of L_T, an estimate of hazard function may be obtained as,

$$h_1(t) = \frac{n_t^*}{N(t)}, \ t = 1, 2, \ldots, L_T \text{ and } L_M = \infty \tag{2}$$

where $N(t) = V_t - \sum_{j=1}^{t-1} n_j^* = N(t-1) - n_{t-1}^*$ for $t = 2, 3, \ldots, L_T$ and $N(1) = V_1 = N$

In presence of a finite warranty mileage limit L_M, (2) requires a modification. The modification may be done either to the denominator or numerator respectively depending on whether or not the mileage accumulation rates in the vehicle population is known. When warranty mileage limit L_M is finite, $N(t)$ in (2) represents the number of vehicles 'at risk' of first failure and not the vehicles 'at risk' of first claim. A failure may occur within or outside the warranty mileage limit of L_M, whereas the claim would always be within L_M. When the mileage accumulation rate in the vehicle population is known, the denominator can be modified to represent the number of vehicles at risk of first claim within the warranty mileage limit L_M. When mileage accumulation rate in the vehicle population is not known, the numerator in (2) can be modified by estimating total number of first failures at MIS = t (Rai and Singh 2004). In subsequent sections, we discuss modeling of mileage accumulation rates and nonparametric estimation of hazard function in presence of known rates mileage accumulation in the vehicle population.

3. Modeling of Mileage Accumulation Rate

The information about the mileage accumulation rate in the vehicle population of interest helps to make adjustments in the denominator of (2) to arrive at the hazard rate. Let w_i (i = 1, 2, …, n ; $n \le N$) in miles per month denote the mileage accumulation rate of the ith vehicle. Let $g(w)$ and $G(w)$ be the probability density function (pdf) and distribution function (df) respectively for a random variable W. Mileage accumulation rate w_i for the ith vehicle is obtained as,

$$w_i = \frac{u_i}{t_i}, \text{ i} = 1, 2, \ldots, \text{n} \quad (3)$$

where u_i is the mileage of ith vehicle and t_i is the corresponding MIS. For example, a vehicle that has traveled 12K miles at the time of MIS = 10 will have a mileage accumulation rate of 12K/10 = 1200 miles per month. Although the vehicle is not expected to accumulate mileage at the rate of 1200 miles per month at every point in its life cycle, it is reasonable to assume a linear trend between mileage at $MIS = t$ versus t.

The parameters of $g(w)$ can be estimated by taking a random sample from the population of N vehicles. A random sample would ensure that both failed and non-failed vehicles form part of the sample. A sample from only failed or only non-failed vehicles may be biased if the mileage accumulation rate is correlated with the failure mode under study. Follow-up survey or recall data is generally used for estimating the parameters of $g(w)$. Assuming mileage accumulation rate for a given type of vehicle to be independent of the model year, one may use data from previous model year vehicle, if available.

Consider a situation where lognormal distribution provides a good fit for the mileage accumulation data. For the random variable $W \sim LN(\mu,\sigma^2)$ or $log(W) \sim N(\mu,\sigma^2)$, μ is called location parameter and σ is called scale parameter. Let $Y_t = W \times t$ $(t = 1, 2, \ldots, L_T)$ be a random variable representing total miles accumulated by a vehicle at MIS = t, assuming a linear mileage accumulation rate. The units for W and Y_t are miles per month and miles respectively. As t is constant and σ^2 is invariant to the change in location, we have

$$\log(Y_t) = [\log(W) + \log(t)] \sim \text{N}[\mu + \log(t), \sigma^2] \quad (4)$$

Thus for a lognormal distribution, estimates of proportion of vehicles within certain mileage limits at any MIS value can be obtained using $g(y_t)$. For situation where distribution other than the lognormal provides a good fit to mileage accumulation data, estimates of proportion of vehicles within certain mileage limit can be similarly obtained.

4. Nonparametric Estimation of the Hazard Function With Known Rates of Mileage Accumulation

4.1 Risk Set Adjustment for Hazard Rate Estimation

Using Y_t (t = 1, 2, 3, …, L_T), the denominator in (2) can be adjusted for vehicles that have crossed the warranty mileage limit of L_M miles. We can write (2) as,

$$h_2(t) = \frac{n_t^*}{N(t) - N(t) \times P[Y_t > L_M]} = \frac{n_t^*}{N(t) \times P[Y_t \le L_M]}, \; t = 1, 2, \ldots, L_T \quad (5)$$

The denominator in (5) represents the number of vehicles at risk of resulting in first claim within the warranty mileage limit of L_M miles at MIS = t. the mileage limit L_M is a fixed constant. It is assumed that time to first failure of the vehicle is the same as the MIS at its first warranty claim if the associated mileage is ≤ L_M. For hard failures it is plausible to assume that occurrence of failure at MIS = t is independent of the warranty mileage limit. However when the warranty data is influenced by the phenomena of customer-rush near warranty expiration limit, (5) requires further modification. Thus claims occurring in the last say *b* miles of the warranty mileage limit L_M, may be screened out to avoid bias in the estimation of hazard function. Similarly, to avoid mixing of manufacturing/ assembly related claims with usage related claims, the claims below say *a* miles at each MIS may be screened out. Let n_t denote the number of first claims between *a* miles and $(L_M - b)$ miles at MIS = *t*. We can now write (5) as,

$$h_3(t) = \frac{n_t}{N(t) \times P[a \le Y_t \le (L_M - b)]}, \; t = 1, 2, \ldots, L_T \quad (6)$$

The denominator in (6) represents number of vehicles at risk of resulting in first claim in the interval $(a, L_M - b)$ miles at each MIS.

In (6) *a* and *b* are considered fixed constants. In this paper we use a cut-off line in mileage for the values of *a* and *b* which is common to each MIS value. The choice of the values of *a* and *b* helps to avoid dependence of the failure time on warranty mileage limit. Thus occurrence of a failure at time MIS = t in the interval $(a, L_M - b)$ miles is assumed independent of the limits of the observation.

4.2 Incorporating Censoring Information in the Hazard Function Estimation

On certain occasions data on left-censored first claims due to delay in reporting of soft failures during $(L_M - b, L_M)$ are also available. Incorporating such left-censored first claims avoids underestimation of the hazard function. Two sources of such information that help in determination of left-censored first claims are

- technician comments available from the warranty claim records, and
- engineering analysis of the fields returned parts.

Reading details about the part requiring repair or replacement in the technician comments available from warranty claim records, can help to obtain insight about condition of the part. When information given in the technician comments are inconclusive regarding the failed part or subsystem, engineering analysis of the field returned parts are often used. Consider an example of an exhaust gas recirculation (EGR) valve that helps in emissions reduction and regulates exhaust gases by opening or closing operation in a diesel engine. Gradual accumulation of carbon deposits in the EGR valve can lead to its partial opening or closing that result in emissions and/or driveability problems. Heavy carbon deposits in an extreme case can leave the valve inoperative, which is a hard failure at component-level and soft failure at vehicle-level. Mention of inoperative EGR valve in the technician comments is an indication of advanced stage of carbon deposits. The reported failure time in such cases can be treated as left-censored. In absence of sufficient information in the warranty claim records, amount of carbon deposits in the field returned EGR valve using engineering analysis can be used for such a decision. It is not feasible to provide a standard guideline for deciding whether or not a reported failure is left-censored which is applicable to all failure modes. However depending on the nature of failure mode and using knowledge from subject-matter experts, sound operational definition can be developed.

A left-censored first warranty claim occurring at MIS = t requires an adjustment to the total number of first claims for each $MIS \leq t$. To enable such an adjustment for estimating hazard function, we make use of an iterative procedure developed by Turnbull (1974). The procedure helps to obtain maximum likelihood estimates of the reliability function. The method is useful when the data can be grouped naturally, as in the case of automobile warranty data where claims are grouped by MIS values.

1) Nonparametric Maximum Likelihood Estimation of Hazard Function in Presence of Left Censored First Claims and Right Censored Vehicles: Let $R(t) = P[T>t]$ be the reliability function at t = 1, 2, …, L_T. As defined earlier, n_t represents the number of first claims in the mileage interval (a, L_M-b) miles at MIS = t. Each of these n_t reported failures contribute a term $[R(t-1)-R(t)]$ to the likelihood function. Let r_t denote the number of vehicles that have not yet completed t months-in-service and are right-censored at MIS = t. It is assumed that the right-censoring of r_t vehicles occur at the beginning of MIS = t. Each of these r_t right-censored vehicles contribute a term $R(t)$ to the likelihood function. Similarly, let c_t be the number of left-censored first claims at MIS = t which contribute a term $[1-R(t)]$ to the likelihood function. It is assumed that the left-censored claims occur at the end of MIS = t. $N = \sum_{j=1}^{L_T}\left(n_j + r_j + c_j\right)$ is the total number of vehicles in the field. To obtain nonparametric maximum likelihood estimates of $R(t)$, the loglikelihood function $\mathcal{L}$ can be written as

$$\mathcal{L} = \log\left(\prod_{t=1}^{L_T}\left[R(t-1) - R(t)\right]^{n_t}[R(t)]^{r_t}[1 - R(t)]^{c_t}\right)$$

$$= \sum_{t=1}^{L_T}\left\{n_t \log[R(t-1) - R(t)] + r_t \log[R(t)] + c_t \log[1 - R(t)]\right\} \tag{7}$$

Using the results obtained by Turnbull (1974) we can obtain the maximum likelihood estimates $\hat{R}(t)$ for the reliability functions in (7) as,

$$R(t) = q_t \times R(t-1),\ t = 2, 3, \ldots, L_T \tag{8}$$

With, $R(0) = 1$; $R(1) = q_1$, and

$$q_t = \frac{\left(N'(t) - n_t^{'}\right)}{N'(t)}; \tag{9}$$

$$N'(t) = P\left[a \leq W_t \leq (L_M - b)\right] \times \sum_{i=t}^{L_T}(r_i + n_i^{'}) \tag{10}$$

$$n_t^{'} = n_t + \sum_{i=t}^{L_T} c_i \alpha_{ij} \quad (11)$$

$$\alpha_{it} = \frac{R(t-1) - R(t)}{1 - R(i)}, t \le i \quad (12)$$

The iteration procedure starts with obtaining $R(i^0)$ as initial estimates assuming $c_i = 0, \forall i = 1,2,..., L_T$. Then using (9)-(12), values for $R(i^l), l \ge 0$ is iteratively obtained until $\max_{1<i<L_T} \left| R(i^l) - R(i^{l-1}) \right| < \delta$, where δ is a very small number. From the estimates $\hat{R}(t)$, estimates of cumulative hazard function $\hat{H}(t)$ is obtained using the following relationship,

$$\hat{H}(t) = - \log[\hat{R}(t)], t = 1, 2, 3, \ldots, L_T \quad (13)$$

2) Confidence Limits for the Reliability Function: Variance of R(t) denoted by $Var[\hat{R}(t)]$ is used to develop confidence limits for $\hat{R}(t)$ (t = 1, 2, 3, …, L_T). The expressions to arrive at $Var[\hat{R}(t)]$ using second partial derivatives were developed by Turnbull (1974) and are as given below

$$Var[\hat{R}(t)] = \frac{n_t}{[R(t-1) - R(t)]^2} + \frac{n_{t+1}}{[R(t) - R(t+1)]^2} + \frac{r_t}{[R(t)]^2} + \frac{c_t}{[1 - R(t)]^2}, \text{ t} = 1, 2, ..., L_T\text{-}1 \quad (14)$$

$$Var[\hat{R}(L_T)] = \frac{n_{L_T}}{[R(L_T - 1) - R(L_T)]^2} + \frac{r_{L_T}}{[R(L_T)]^2} + \frac{c_{L_T}}{[1 - R(L_T)]^2}, \text{ t} = L_T \quad (15)$$

The approximate 95% confidence limits for $\hat{R}(t)$ are thus obtained at each t as

$$\hat{R}(t) \pm 1.96\sqrt{Var[\hat{R}(t)]}, \text{ t} = 1, 2, ..., L_T \quad (16)$$

5. An Application Example

To illustrate the application of the methodology proposed in previous sections, we consider field reliability analysis of a component level failure mode using warranty data. The vehicle and failure related details are not disclosed in the paper to protect the proprietary nature of information. The warranty coverage for the vehicle under study is two-dimensional with a time limit of 36 months-in-service and mileage limit of 36 000 miles, whichever occurs first. The number of vehicles in the field up to MIS = t denoted by V_t are rounded to the nearest thousands and are given in Table I below.

TABLE I

NUMBER OF VEHICLES IN THE FIELD

MIS (t)	V_t	MIS (t)	V_t	MIS (t)	V_t	MIS (t)	V_t
1	200,000	11	200,000	21	200,000	31	80,000
2	200,000	12	200,000	22	200,000	32	60,000
3	200,000	13	200,000	23	200,000	33	45,000
4	200,000	14	200,000	24	200,000	34	30,000
5	200,000	15	200,000	25	200,000	35	20,000
6	200,000	16	200,000	26	180,000	36	10,000
7	200,000	17	200,000	27	160,000		
8	200,000	18	200,000	28	140,000		
9	200,000	19	200,000	29	120,000		
10	200,000	20	200,000	30	100,000		

Table I shows that overall there are 200 000 vehicles in the field. Out of the total number of vehicles in the field, 10 000 vehicles have reached 36 MIS, whereas, all 200 000 vehicles have reached 25 MIS.

5.1 Mileage Accumulation Rate in the Vehicle Population

Recall data from a previous model year vehicle were used to obtain the estimates of mileage accumulation rate, as no such data was available from the current model year vehicle. Fig. 3. shows four-way probability plots of the mileage accumulation rates.

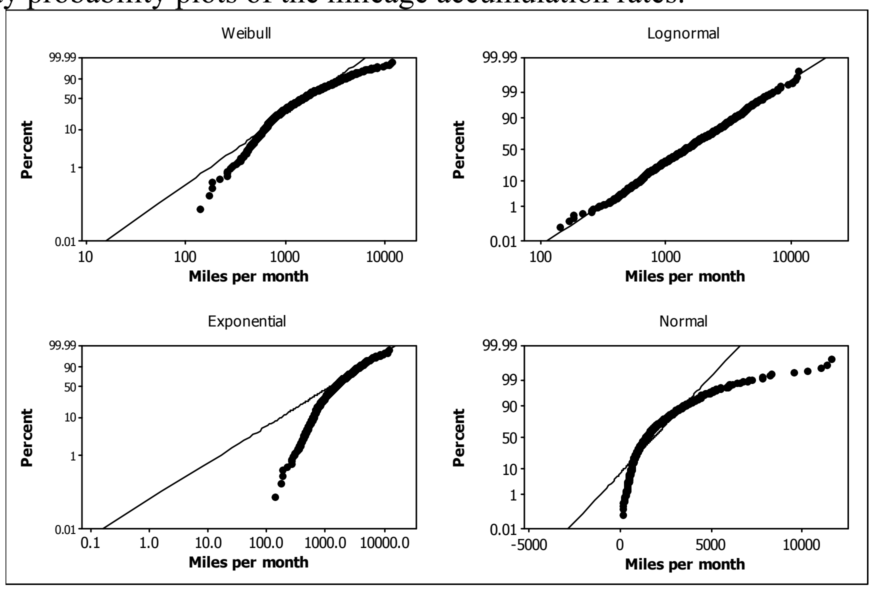
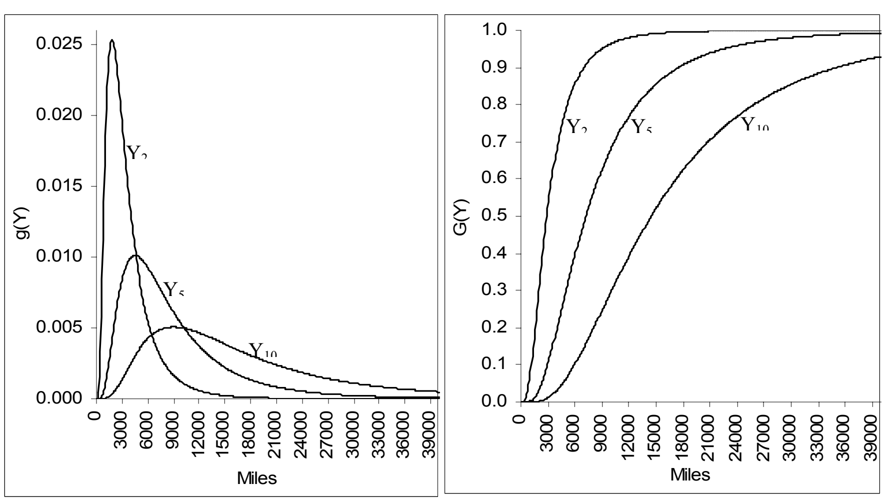


Fig. 3. Four-way probability plots of mileage accumulation rates



(a) (b)

Fig. 4. Mileage accumulation rates at MIS = 2, 5, and 10 for $\log(Y_t) \sim N[7.279+\log(t), 0.688^2]$
(a) Probability density function $g(y_t)$, (b) Cumulative density function $G(y_t)$

It can be visually observed from Fig. 3. that lognormal distribution provides a good fit for the mileage accumulation data. Based on lognormal distribution, an Anderson-Darling (AD) statistic value of 0.52 is obtained with significance level p of 0.19. AD test is a goodness-of-fit test based on empirical distribution function. It measures the 'distance' or 'discrepancy' between the empirical and the specified (or assumed) distributions (Kececioglu 1993). A value of significance level p close to one indicates closeness of the dataset to the chosen distribution and vice-versa.

The estimates of location and shape parameter of the lognormal distribution are obtained as 7.279 and 0.688 respectively. These estimates also indicate approximate mean mileage accumulation of 1840 miles per month and standard deviation of 1437 miles per month. Fig. 4. gives pdf $g(y_t)$ and cdf $G(y_t)$ at MIS = 2, 5, and 10.

5.2 Estimation of Cumulative Hazard

In this study the cut-off line in mileage with a = 1K and b = 1.5K is used. Out of a total of 924 first warranty claims, 38 claims are less than 1K miles and are screened out from further analysis. Of the remaining, there are 813 first claims in the interval (1K, 34.5K) and 73 first claims in the interval (34.5K, 36K). Using the information from technician comments and engineering analysis it is determined that 58 first claims out of 73 represent left-censored failures and are retained for further analysis. The remaining 15 out of 73 are screened out. Note that although in this example we consider first claims that are left-censored due to customer-rush phenomena, the methodology is applicable to left-censored claims at any time or mileage.

The estimates for cumulative hazard functions and their 95% confidence limits are obtained using (13)-(16). The cumulative hazard estimates are shown in Fig. 5. along with 95% confidence limits.

Fig. 5. Cumulative hazard rate

From Fig. 5. it is observed that the cumulative hazard plot shows a straight line up to about MIS = 30, indicating constant failure rate trend. There seems to be a slight upward trend after MIS = 30, indicating a beginning of increasing failure rate trend for the component level failure mode under study.

5.3 Comments on results

The effect of left artificial truncation point *a* and right artificial truncation point L_M-*b*, and inclusion of left censored claims can be observed from cumulative hazard plots depicted in Fig. 6(a). to 6(d).

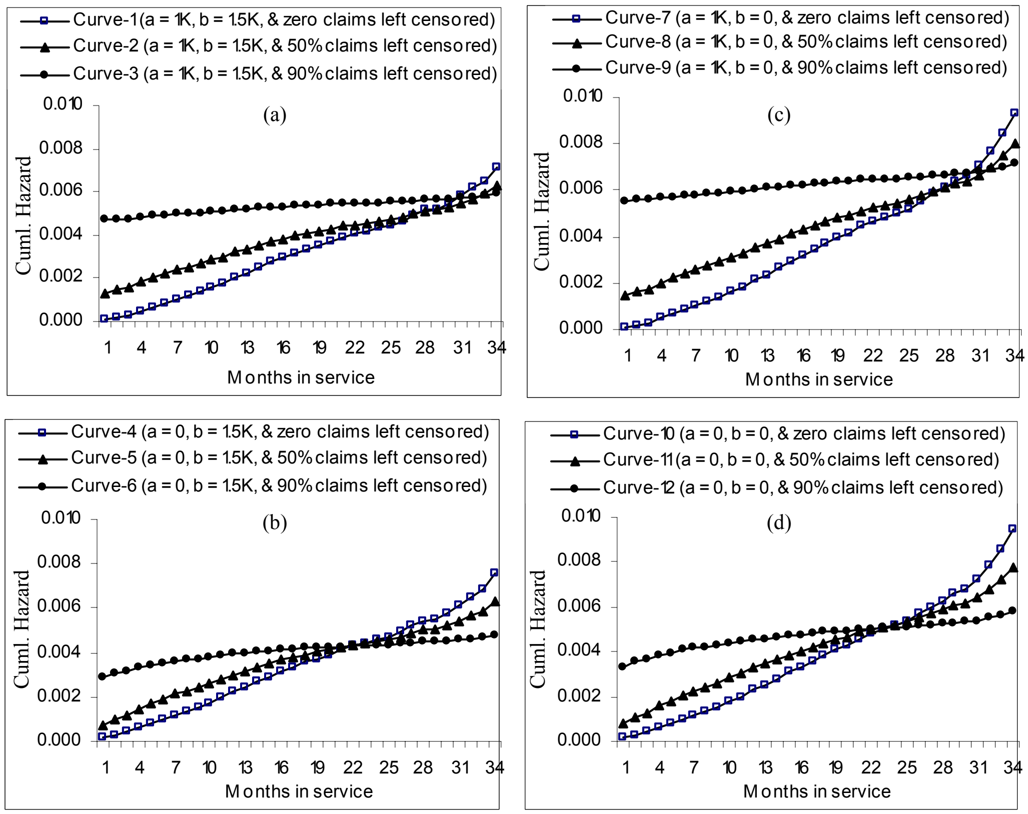


Fig. 6. Comparison of cumulative hazard rate estimates at different artificial truncation points and with different percentages of left-censored claims

Curve-1 in Fig. 6(a). shows cumulative hazard rates when the first claims are considered belonging to H-H category. It assumes that all the failures reported within mileage interval 1000 miles and 34 500 miles are result of usage related hard failures with high chance of detection by the users. Curve-2 considers 50% of the claims in the interval to belong to one of the remaining three categories viz., S-L, S-H, and H-L. It further assumes that all the failures within warranty period are reported and result in a warranty claim. The effect of considering certain percentage of claims as left-censored is to adjust the effective number of reported failures at different MIS values. The left-censored claims lead to an upward correction in the number of first claims observed at MIS values prior to the reported MIS and a similar downward correction at MIS values higher than the MIS at which the claim was originally reported. Similar pattern can be observed from Fig. 6(b). to 6(d). although with different magnitudes at different MIS values. Thus it is seen that treating the reported failures as hard failures or soft failures has considerable impact on the cumulative hazard rate pattern irrespective of whether the left or right artificial truncation points are considered or not.

Fig. 7. shows the curves 1, 4, 7, and 10 from Fig. 6. and compares the cumulative hazard plots for situations when left and/or right truncation points are considered and the failure mode belongs to H-H category.

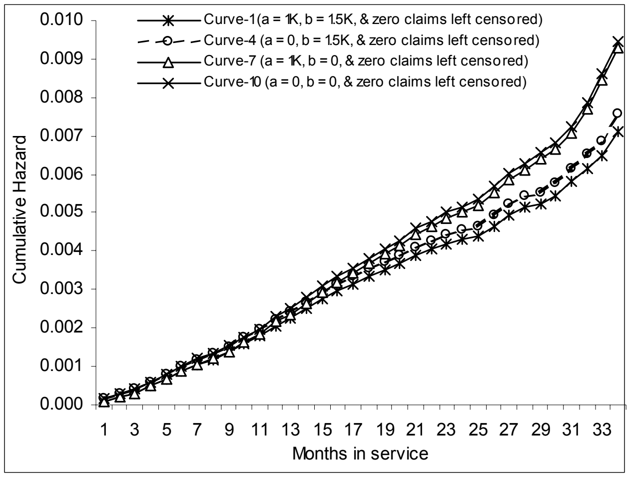


Fig. 7. Comparison of cumulative hazard rate estimates with different artificial truncation points and no left-censoring

Curve-1 in Fig. 7. addressees both spikes at the beginning and towards the end of the warranty coverage period while arriving at cumulative hazard rate. For curve-4, only the spike in claims due to customer-rush near the warranty expiration limit is addressed. Existence of manufacturing/ assembly defects in addition to usage related failures lead to slightly higher cumulative hazard rates in curve-4. Curve-7 ignores the phenomena of customer-rush and as a result shows higher cumulative hazard rates beyond about MIS = 20 than the curves 1 and 4. Below MIS = 15 curves 1, 4, 7, and 10 show relatively small difference in the cumulative hazard rates as not a significant portion of the vehicle population is expected to exceed 36K miles by then. In Fig. 7. the impact of ignoring spike in warranty claims towards the end of warranty coverage is more pronounced than ignoring the spike at the beginning. This is due to a larger accumulation of claims as a result of customer-rush phenomena. However when the warranty data shows a higher spike towards the beginning of the warranty period, the impact of changing left artificial truncation point denoted by *a* would be more pronounced.

6.0 Conclusions

Research in the area of field reliability studies using warranty data has traditionally assumed reported time or mileage at failure in warranty claims to be actual time or mileage at failure. Although this is largely true for hard failures, soft failures may lead to a delayed reporting of failures. In certain situations such delays can be to an extent that failures get reported when the warranty coverage is about to expire. Such situations lead to phenomena of customer-rush near the warranty expiration limit resulting in unusually high number of claims towards the end of warranty coverage period. Ignoring such phenomena may lead quality and reliability engineers to incorrectly obtain an increasing failure rate pattern from the cumulative hazard plots during field reliability studies. This in turn may further lead to making costly and uncalled for component or sub-system design changes. It is thus critical to distinguish between warranty claims belonging to soft failure or hard failure category and use suitable methods for field reliability studies. This paper presents a methodology that helps to obtain nonparametric estimates of cumulative hazard rates when the warranty claims may potentially be a result of soft failures. The methodology requires estimates of mileage accumulation rates in the vehicle population to be available. The methodology proposed also helps to obtain approximate confidence limits for the maximum likelihood estimates of cumulative hazard function.

References

Blischke, W.R., D. N. P. Murthy. 2000. *Reliability modeling, prediction, and optimization*. New York: John Wiley & Sons.

Cohen, C.A. 1991. *Truncated and censored samples: theory and applications*. New York: Marcel Dekker.

Davis, T.P. 2003. Reliability improvement in automotive engineering. *Global Vehicle Reliability – Prediction and Optimization Techniques*, J. E. Strutt and P. L. Hall, Eds., Professional Engineering Publishing, Chapter 4.

Hu, X.J., J. F. Lawless, K. Suzuki. 1998. Nonparametric estimation of a lifetime distribution when censoring times are missing. *Technometrics*, vol. 40, pp. 3-13.

Iskandar, B.P., W. R. Blischke. 2003. Reliability and warranty analysis of a motorcycle based on claims data. *Case Studies in Reliability and Maintainenance*, W. R. Blischke and D. N. P. Murthy, Eds. New Jersey: Wiley-Interscience, pp. 623-656.

Kalbfleisch, J.D., J. F. Lawless, J. A. Robinson. 1991. Methods for the analysis and prediction of warranty claims. *Technometrics,* vol. 33, pp. 273-285.

Kalbfleisch, J.D, J. F. Lawless. 1992. Some useful statistical methods for truncated data. *Journal of Quality Technology,* vol. 24, pp. 145-152.

Kececioglu, D. 1993. Reliability and life testing handbook, Vol. 1, PTR Prentice Hall, New Jersey.

Lawless, J.F., J. Hu, J. Cao. 1995. Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis*, vol. 1, pp. 227-240.

Lu, M.W. 1998. Automotive reliability prediction based on early field failure warranty data. *Quality and Reliability Engineering International*, vol. 14, pp. 103-108.

Majeske, K.D., T. Lynch-Caris, G. Herrin. 1997. Evaluating product and process design changes with warranty data. *International Journal of Production Economics*, vol. 50, pp. 79-89.

Majeske, K.D. 2003. A mixture model for automobile warranty data. *Reliability Engineering and System Safety*, 81, 71-77.

Meeker, W.Q., L. A. Escobar. 1998. *Statistical methods for reliability data*. New York: John Wiley & Sons.

Nelson, W. 1990. Hazard plotting of left truncated life data. *Journal of Quality Technology*, vol. 22, pp. 230-238.

Nelson, W. 2000. Theory and applications of hazard plotting for censored failure data. *Technometrics*, vol. 42, pp. 12-25.

Pal, S., G. S. R. Murthy. 2003. An application of Gumbel's bivariate exponential distribution in estimation of warranty cost of motor cycle. *The International Journal of Quality and Reliability Management*, vol. 20, no. 4, pp. 488-502.

Rai, B.K., N. Singh. 2003. Hazard rate estimation from incomplete and unclean warranty data. *Reliability Engineering & System Safety*, 81, 79-92.
Rai, B.K., N. Singh. 2005. A modeling framework for assessing the impact of new time / mileage warranty limits on the number and cost of automotive warranty claims. *Reliability Engineering & System Safety*, vol. 88, pp. 157-169.
Rai, B.K., N. Singh. 2004. Modeling and analysis of automobile warranty data in presence of bias due to customer-rush near warranty expiration limit. *Reliability Engineering & System Safety*, vol. 86, pp. 83-94.
Rai, B.K. 2004. Modeling, analysis, and prediction from automobile warranty datasets. Ph.D. dissertation, Dept. Industrial Engineering, Wayne State University, Detroit, MI.
Sander, P.C., L. M. Toscano, S. Luitjens, V. T. Petkova, A. Huijben, A. C. Brombacher. 2003. Warranty data analysis for assessing product reliability. *Case Studies in Reliability and Maintainenance*, W. R. Blischke and D. N. P. Murthy, Eds. New Jersey: Wiley-Interscience, pp. 591-621.
Stephens, D., M. Crowder. 2004. Bayesian analysis of discrete time warranty data. *Applied Statistics*, vol. 53, pp. 195-217.
Suzuki, K. 1985a. Nonparametric estimation of lifetime distributions from a record of failures and follow-ups. *Journal of American Statistical Association*, 80, 68-72.
Suzuki, K. 1985b. Estimation of lifetime parameters from incomplete field data. *Technometrics*, 27, 263-71.
Suzuki, K., Md. R. Karim, L. Wang. 2001. Statistical analysis of reliability warranty data. *Handbook of Statistics*, N. Balakrishnan and C. R. Rao, Eds. Amsterdam: Elsevier Science, Chapter 21.
Turnbull, B.W. 1974. Nonparametric estimation of a survivorship function with doubly censored data. *Journal of American Statistical Association*, 69, 169-73.