Modelling component reliability using warranty data

Raymond Summit[1]

Abstract

Accelerated testing is often used to gauge the reliability of a product before it is released to the market. Warranty data provide valuable additional information about the reliability of a product and its constituent components via field data. In this study, warranty data provided by an Australian automobile manufacturer is used. This article considers the need to clean the data before commencing any analysis. It discusses the distribution of warranty claims and the use of warranty data to model the reliability of elemental components of an automobile. Various model options are considered before suitable modelling is adopted. The reality of modelling the reliability of a vehicle containing a large number of components is considered and a suitable method is adopted. Verification of the reliability modelling is made by comparing an estimated warranty cost with the actual cost.

Contents

1 Introduction

Warranties represent a significant cost to the manufacturer. In the North American automotive industry, spending has reached close to tens of billions of dollars and is increasing [16]. In order to properly manage a product's warranty liability, it is vital to model the reliability of a product and the cost of servicing the warranty on that product. This is particularly important in the Australian automotive industry, where warranty terms have recently been increasing. Foreign manufacturers have offered more generous warranty

terms than their local counterparts in an effort to establish themselves in the local market. This has forced local manufacturers to assure customers of their products' build quality by increasing their own warranty terms.

This study uses an Australian car manufacturer's warranty data to create component reliability models, which are then used to predict the overall warranty cost. It is important to commence any modelling by identifying the assumptions made. A brief outline of survival analysis techniques is provided in this article before they are applied to the manufacturer's database. The need to clean the data is established herein. We discuss how to analyse the database containing a large number of parts having few claims.

Often, a manufacturer's reliability model of a product is inaccurate. For example, Majeske and Herrin [13] found that a manufacturer's exponential reliability models predicted only half the actual failure rate of a car radio and underestimated the number of warranty claims on a brake system by a factor of four. They concluded that more accurate modelling can be achieved using the Weibull distribution and that predictive models obtained from bench testing need to be verified against field data. Various model options are considered in this study and then verified against the actual warranty cost.

2 Modelling considerations

2.1 Assumptions

In fitting a model to the warranty data, the following standard assumptions are made [2]:

- Each failure of the component during the warranty period leads to a claim and all claims are valid;

- There are no significant delays in making a claim;

- Failure times of the component are independent and identically distributed;

- Repair times are small compared to the mean time to failure of the component;

- All repairs are as good as new;

- The distance travelled by a vehicle before sale is negligible;

- The number of vehicles that are taken out of service during the study is negligible.

2.2 Dimensionality of the warranty

The manufacturer's warranty is for three years or $100,000\,\text{km}$, whichever comes first. This is an example of a two dimensional warranty, since it is limited by both time and usage restraints [1]. The manufacturer has accurate information about a vehicle's age, as this is derived from the manufacturing database. However, the distance field in the claims database contains inaccuracies. Evidently, some dealerships enter false figures into this mandatory field in order to complete the form. Moreover, if a claim has not been made on a vehicle, the distance it has travelled by the end of the experiment would need to be estimated. Censored data so formed would then be inaccurate [4, 8]. This makes time the preferred restraint upon which to model warranty claims. In addition, since most manufacturers have a monthly reporting period, time is the more convenient variable upon which to base warranty predictions [12].

Figure 1 plots distance against age of vehicles that are in the manufacturer's claims database. Although the data are incomplete (only claims data are represented), the figure does show that most users reach the time limitation before the distance limitation. This is supported by data from other studies [10, 11], where usage is calculated to be approximately $22,500\,\text{km/yr}$. This is a much lower rate than the $33,333\,\text{km/yr}$ allowance of the warranty in our

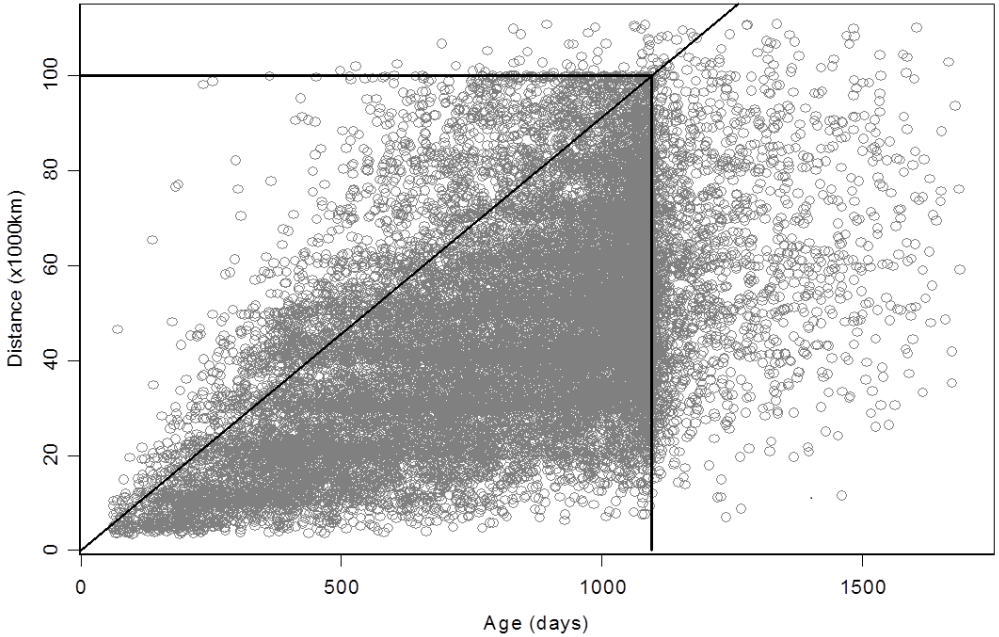Figure 1: Vehicle usage.

study. These observations again support the use of time as the dependent variable.

3 Survival analysis

3.1 Kaplan–Meier estimator

The discussion in this section is based on standard texts on lifetime data analysis [9, 14]. Survival analysis is characterised by the existence of observed failure times and censored times. The latter occurs when an item is removed

from further observation. In the case of warranty data, censoring occurs when a vehicle reaches the end of its warranty.

Various non-parametric survival estimators have been presented in the literature. The Kaplan–Meier (product-limit) estimator is widely accepted. The survival of an item to time t_i is the product of survival probabilities up to that time and is $\hat{S}(t) = \prod_{t_i \leqslant t} [k(t_i) - d_i]/k(t_i)$, where $k(t_i)$ and $d(t_i)$ are the number at risk and the number of deaths (or failures) in a time interval, respectively.

3.2 Parametric fit

The likelihood function $L(t_j; \theta) = \left(\prod_{j \in F} f(t_j; \theta) \right) \left(\prod_{j \in C} S(t_j; \theta) \right)$, where $f(t)$ is the failure density distribution with parameter set θ, $S(t)$ is the survival function, F is the set of observations reaching failure and C is the set of censored items. Maximising this function yields the maximum likelihood estimator (MLE) of the parameters of $f(t; \theta)$ [6].

For the exponential distribution, $f(t) = \lambda \exp(-\lambda t)$, with failure rate λ, the MLE of λ is found algebraically to be $\hat{\lambda} = r / \sum_{j=1}^{n} t_j$, where r is the number of items that reach failure. The MLE of the variance of λ is

$$\text{Var}\left(\hat{\lambda}\right) = \frac{\hat{\lambda}^2}{\sum_j \left[1 - \exp\left(-t_j \hat{\lambda}\right)\right]} \,.$$

The Weibull distribution may be written as

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

where β is the shape parameter and η is the scale parameter. The MLE of the Weibull parameters can only be obtained numerically. Standard references develop the equations for both exponential and Weibull distributions [14].

3.3 Renewal function formulation

A failed component is replaced by a new one, provided a vehicle is still within its warranty period. This process is called a renewal process and has been well documented [5, 18]. For a known failure distribution $F(t)$, the number of renewals is $M(t) = F(t) + \int_0^t M(t-x) \, dF(x)$.

This equation is solved algebraically for the exponential distribution, yielding the familiar result $M(t) = \lambda t$. A numerical solution is needed for the Weibull distribution. The one used in this study was provided by Xie [19].

4 The manufacturer's data

Having discussed modelling aspects and survival analysis techniques, we now turn to using the manufacturer's database to model the reliability of the components of the vehicle and the subsequent warranty cost.

4.1 Data source

The data used in this study has been made available by an Australian car manufacturer and consists of two databases. The first identifies $30,138$ vehicles produced in one year; the second contains the $62,456$ warranty claims made over the three year/$100,000$ km warranty and a little beyond. The data cover model variations (engine, body type, etc.) of a passenger vehicle. The claims database identifies the faulty part that led to a claim. Components provided by different suppliers have different part numbers and are analysed separately in this study. Although $3,040$ unique part numbers are identified, no one vehicle contains that many parts. The age of a component was calculated and used in the survival analysis of that component, whilst the age of a vehicle, a different quantity, was also monitored.

Table 1: Number of parts with given number of claims.

No. of claims	No. of parts	No. of claims	No. of parts
1–100	2964	1	1159
101–200	43	2	454
201–300	12	3	262
301–400	5	4	193
401–500	3	5	141
501–600	3	6	89
601–700	1	7	63
701–800	2	8	62
801–900	2	9	44
900–1000	0	10	31
> 1000	6	11–100	466

Exploratory data analysis of the manufacturer's claims database led us to discover errors, such as missing data, impossible date sequences, inconsistencies between a vehicle's age and distance travelled, and the existence of extreme outliers. The need to clean the data is well documented [15, 16]. As claims were entered manually by the dealerships, records are prone to data-entry errors. However, less than 2% of the records have been found to be in error. It is important to clean and retain as much of the data as possible because the removal of records would introduce a bias by reducing the number of claims. By cross-checking values in one field against that of other fields, it has been possible to clean and retain all the data.

4.2   Claims distribution

Table 1 shows the number of parts with the given claims frequency, with 1 to 100 claims being further broken down. Inspection of the table reveals that the vast majority of parts in the manufacturer's database proved to be very reliable during the warranty period.

Of the $62,456$ claims made during the warranty, $1,159$ parts were replaced once only. For these parts, the choice of parametric model is somewhat arbitrary. Similarly, for parts with few failures and numerous censored data, many models could fit the data. In such cases, it is practical to fit a simple exponential model.

4.3 Parametric models

Various models have been fitted to the observed data and compared via their maximum likelihood values. Table 2 reveals that the maximum likelihood values for the various parametric models are similar, with the few exceptions having values of a different order of magnitude. The reason for this occurrence has not been pursued in detail, but in most cases, it appears to be due to only one or two extraneous points. From the log likelihood values, no particular model stands out as providing a superior fit to the others.

The Weibull distribution is a widely accepted failure model in reliability applications [3]. It has flexibility and includes other distributions for particular values of its shape parameter. For example, when $\beta = 2$ the Weibull simplifies to the Rayleigh distribution. Table 2 shows that the Weibull model provides good log-likelihood values as well as consistently good fits, making it a suitable model to use with our data.

When few failures are observed, most models provide a reasonably good fit. Thus, the exponential distribution is as good a model as any other, and because of its mathematical tractability, is suitable under such circumstances [7]. Because our database contains a large number of parts, an onerous amount of computation time is required to fit parametric models. After careful examination of the claims distribution, it is clear that an exponential model is appropriate for parts with less than 20 claims.

Table 2: Log likelihood values.

Part	No. of claims	Extreme	Normal	Logistic	Expo	Weibull	Log normal	Log logistic	Rayleigh
A	5	$-8.4\mathrm{E}1$	$-8.3\mathrm{E}1$	$-8.4\mathrm{E}1$	$-8.3\mathrm{E}1$	$-8.2\mathrm{E}1$	$-8.2\mathrm{E}1$	$-8.2\mathrm{E}1$	$-8.2\mathrm{E}1$
B	10	$-1.6\mathrm{E}2$	$-1.6\mathrm{E}2$	$-1.6\mathrm{E}2$	$-1.6\mathrm{E}2$	$-1.6\mathrm{E}2$	$-6.8\mathrm{E}38$	$-1.6\mathrm{E}2$	$-1.6\mathrm{E}2$
C	15	$-2.3\mathrm{E}2$	$-2.3\mathrm{E}2$	$-2.3\mathrm{E}2$	$-2.3\mathrm{E}2$	$-2.3\mathrm{E}2$	$-2.7\mathrm{E}3$	$-2.3\mathrm{E}2$	$-2.3\mathrm{E}2$
D	20	$-3.0\mathrm{E}2$	$-2.9\mathrm{E}2$	$-3.0\mathrm{E}2$	$-3.0\mathrm{E}2$	$-2.9\mathrm{E}2$	$-2.9\mathrm{E}2$	$-2.9\mathrm{E}2$	$-2.9\mathrm{E}2$
E	30	$-4.2\mathrm{E}2$	$-4.2\mathrm{E}2$	$-4.2\mathrm{E}2$	$-4.4\mathrm{E}2$	$-4.2\mathrm{E}2$	$-4.2\mathrm{E}2$	$-4.2\mathrm{E}2$	$-4.3\mathrm{E}2$
F	50	$-7.1\mathrm{E}2$	$-7.1\mathrm{E}2$	$-7.1\mathrm{E}2$	$-7.2\mathrm{E}2$	$-7.1\mathrm{E}2$	$-7.1\mathrm{E}2$	$-7.1\mathrm{E}2$	$-7.1\mathrm{E}2$
G	75	$-1.0\mathrm{E}3$	$-1.0\mathrm{E}3$	$-1.0\mathrm{E}3$	$-1.0\mathrm{E}3$	$-1.0\mathrm{E}3$	$-3.4\mathrm{E}38$	$-1.0\mathrm{E}3$	$-1.0\mathrm{E}3$
H	102	$-1.4\mathrm{E}3$	$-1.4\mathrm{E}3$	$-1.4\mathrm{E}3$	$-1.3\mathrm{E}3$	$-1.3\mathrm{E}3$	$-1.3\mathrm{E}3$	$-1.3\mathrm{E}3$	$-1.4\mathrm{E}3$
I	152	$-2.0\mathrm{E}3$	$-2.0\mathrm{E}3$	$-2.0\mathrm{E}3$	$-2.0\mathrm{E}3$	$-2.0\mathrm{E}3$	$-2.0\mathrm{E}3$	$-2.0\mathrm{E}3$	$-2.0\mathrm{E}3$
J	200	$-2.7\mathrm{E}3$	$-2.7\mathrm{E}3$	$-2.7\mathrm{E}3$	$-2.6\mathrm{E}3$	$-2.2\mathrm{E}3$	$-2.2\mathrm{E}3$	$-2.2\mathrm{E}3$	$-3.3\mathrm{E}3$
K	302	$-3.7\mathrm{E}3$	$-3.7\mathrm{E}3$	$-3.7\mathrm{E}3$	$-3.8\mathrm{E}3$	$-3.7\mathrm{E}3$	$-3.7\mathrm{E}3$	$-3.7\mathrm{E}3$	$-3.7\mathrm{E}3$
L	499	$-5.8\mathrm{E}3$	$-5.8\mathrm{E}3$	$-5.8\mathrm{E}3$	$-6.0\mathrm{E}3$	$-6.0\mathrm{E}3$	$-6.0\mathrm{E}3$	$-6.0\mathrm{E}3$	$-6.0\mathrm{E}3$
M	705	$-8.1\mathrm{E}3$	$-8.1\mathrm{E}3$	$-8.1\mathrm{E}3$	$-8.2\mathrm{E}3$	$-8.1\mathrm{E}3$	$-8.1\mathrm{E}3$	$-8.1\mathrm{E}3$	$-8.1\mathrm{E}3$
N	1662	$-3.4\mathrm{E}38$	$-1.8\mathrm{E}4$	$-1.8\mathrm{E}4$	$-1.8\mathrm{E}4$	$-1.7\mathrm{E}4$	$-1.8\mathrm{E}4$	$-1.7\mathrm{E}4$	$-1.8\mathrm{E}4$
O	2035	$-2.1\mathrm{E}4$	$-2.1\mathrm{E}4$	$-2.1\mathrm{E}4$	$-2.1\mathrm{E}4$	$-2.1\mathrm{E}4$	$-2.1\mathrm{E}4$	$-2.1\mathrm{E}4$	$-2.1\mathrm{E}4$
P	2964	$-3.1\mathrm{E}4$	$-3.1\mathrm{E}4$	$-3.1\mathrm{E}4$	$-3.0\mathrm{E}4$	$-3.0\mathrm{E}4$	$-3.0\mathrm{E}4$	$-3.0\mathrm{E}4$	$-3.2\mathrm{E}4$
Q	8733	$-8.4\mathrm{E}4$	$-8.3\mathrm{E}4$	$-8.4\mathrm{E}4$	$-8.0\mathrm{E}4$	$-8.0\mathrm{E}4$	$-8.0\mathrm{E}4$	$-8.0\mathrm{E}4$	$-8.4\mathrm{E}4$

Table 3: Total warranty cost, variance and confidence limits.

	Warranty cost	Standard deviation	95% CI
Exponential	$272.84	$9.83	[$253.58, $292.11]
Weibull	$272.62	$5185.17	[$0, $10435]

5   Model verification

In order to verify the reliability models, the cost of the repair of each component, which is a stochastic quantity, is estimated. Repair cost depends on many factors, such as previous history of a problem, the mechanic's experience, and freight cost. We use the mean cost for a point estimate, but truncate the data if extreme outliers are present. The variance of the cost is used when calculating an interval estimate of warranty cost.

To verify the reliability modelling from the last section, the warranty cost per vehicle is calculated using the expected number of repairs and the expected cost of component repair. Two failure models are used for this comparison. The first models all component failures exponentially; the second uses the exponential distribution for components having less than 20 claims and the Weibull distribution for the rest. In each case, a 95% confidence interval (CI) is also calculated. The results are shown in Table 3. The two methods produced similar point estimates that closely match the actual cost per vehicle of $275.48. For the Weibull model, the coefficient of variation is approximately 20, which indicates that the standard deviation is too large to be of practical use. This is also evident from the large CI. We developed more useful CIs for the Weibull parameters using simulations [17].

It appears that the additional processing required in the Weibull model is not justified because the estimated warranty cost works out to be very similar to that of the exponential model. However, a large number of components have β estimates that lie outside the interval [0.95, 1.05]. Hence, it is preferable to use the Weibull distribution to obtain more accurate point estimates of the reliability of components with β values outside this range.

6 Summary

This article has shown how a manufacturers warranty database can be used to model the reliability of components, from which an estimate of the cost of a warranty can be made. The need to start the analysis with data cleaning has been demonstrated. The large number of components in the database produces copious computations. Thus a means of simplifying the computation has been established. Examination of the frequency distribution of components with claims within given ranges has proved to be insightful and has led to the use of two distributions to model the reliability of components. Using log-likelihood values, the fit of various models have been compared and it has been concluded that the Weibull model is suitable for components with a reasonable frequency of claims, whilst the exponential is suitable for components with fewer claims. This modelling has led to a fairly accurate estimate of the warranty cost when compared with the actual warranty cost.

References

[1] Blischke, W. R., Karim, M. R. and Murthy, D. N. P. *Warranty Data Collection and Analysis*. Springer, London, 2011. C440

[2] Blischke, W. R. and Murthy, D. N. P. *Warranty Cost Analysis*. Marcel Dekker, Inc, New York, 1994. C439

[3] Blischke, W. R. and Murthy, D. N. P. *Reliability: Modelling, Prediction, and Optimisation*. John Wiley and Sons, New York, 2000. C445

[4] Campean, I. F. and Brunson, D. Reliability modelling using warranty data. *International Conference on Statistics and Analytical Methods in Automotive Engineering*. IMechE Conference Transactions 2002, 225–236. C440

[5] Cox, D. R. *Renewal Theory*. Chapman and Hall, New York, 1982. C443

[6] Cox, D. R. and Oakes, D. *Analysis of Survival Data*. Chapman and Hall, London, 1984. C442

[7] Hill, V. L. and Blischke, W. R. An assessment of alternative models in warranty analysis. *Journal of Information and Optimization Sciences*, **8**, 1987, 33–55. C445

[8] Iskandar, B. P. and Blischke, W. R. Reliability and Warranty Analysis of a Motorcycle Based on Claims Data. In *Case Studies in Reliability and Maintenance* (eds W. R. Blischke and D. N. P. Murthy.) John Wiley and Sons, Hoboken, NJ, USA, 2003, 623–656. doi:10.1002/0471393002.ch27 C440

[9] Lawless, J. F. *Statistical Models and Methods for Lifetime Data*. John Wiley and Sons, New York, 1982. C441

[10] Lawless, J. F., Hu, J. and Cao, J. Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis*, **1**, 1995, 227–240. C440

[11] Lu, M. Automotive reliability prediction based on early field failure warranty data. *Quality And Reliability Engineering International*, **14**, 1998, 103–108. doi:10.1002/(SICI)1099-1638(199803/04)14:2¡103::AID-QRE147¿3.0.CO;2-5
C440

[12] Majeske, K. D. Evaluating product and process design changes with warranty data. *International Journal of Production Economics*, **50**, 1997, 79–89. doi:10.1016/S0925-5273(97)00034-0 C440

[13] Majeske, K. D. and Herrin, G. D. Determining the warranty benefits for automobile design changes. *Proceedings Annual Reliability and Maintainability Symposium*, 1998, 94–99.
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=653636&tag=1
C439

[14] Nelson, W. *Applied Life Data Analysis*. John Wiley and Sons, New York, 1982. C441, C442

[15] Rai, B. and Singh, N. Hazard rate estimation from incomplete and unclean warranty data. *Reliability Engineering and System Safety*, **81**, 2003, 79–92. doi:10.1016/S0951-8320(03)00083-8 C444

[16] Singh, N. and Rai, B. K. *Reliability Analysis and Prediction with Warranty Data: Issues, Strategies, and Methods*. CRC Press, 2009. doi:10.1201/9781439803264 C438, C444

[17] Summit, R. Estimating the reliability model parameters through a simulation of warranty claims: How much data is needed? In M. Nelson, M. Coupland, H. Sidhu, T. Hamilton, and A. J. Roberts, editors, *Proceedings of the 10th Biennial Engineering Mathematics and Applications Conference*, EMAC-2011, volume 53 of *ANZIAM Journal*, 2012, C451–C464. http://journal.austms.org.au/ojs/index.php/ANZIAMJ/article/view/5114 C447

[18] Tijms, H. C. *Stochastic Models: An Algorithmic Approach*. John Wiley and Sons, New York, 1994. C443

[19] Xie, M. On the Solution of Renewal-Type Integral Equations. *Communications in Statistics*, **18**, 1989, 281–293. doi:10.1080/03610918908812760 C443

Author address

1. **Raymond Summit**, WA School of Mines, Curtin University, Kalgoorlie, Australia.
   mailto:r.summit@curtin.edu.au