МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2011
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2011
Proceedings of the Fortieth Jubilee Spring Conference
of the Union of Bulgarian Mathematicians
Borovetz, April 5–9, 2011

AUTO WARRANTY DATA: MEAN CUMULATIVE FUNCTION*

Stefanka Chukova, Her Guan Teo

In this study, we review and extend previous work on truncation that is typical for automotive warranty data. To deal with the problem of incomplete mileage information, we consider a linear approach within nonparametric framework. We evaluate the mean cumulative warranty cost (per vehicle) and its standard error as a function of age, of mileage, and of actual (calendar) time.

Introduction. Nowadays product warranty plays an increasingly important role in the world of business and its use is widespread. In this study, we emphasize on automotive warranty which allows for free repairs subject to both age and mileage limitations. In the USA, till recently, the most common limit was 36 months or 36 000 miles, whichever comes first. As sales records are retained, vehicle's age is known for all sold vehicles at all time. However, odometer readings are recorded and entered in the warranty database at the dealerships only at the time of placing a claim. Therefore, the study of automotive warranty involves analysis of two variables (age and mileage), and the information on one of them (mileage) is incomplete [4]. The main ideas of warranty analysis are given in [1] and [2]. In [7] authors also deal with the incomplete mileage information problem by using a simple linear mileage accumulation model. The bias due to reporting delay of claim in the analysis of warranty claims and costs is studied in [6]. The study in [3] allows for variation in the rate of mileage accumulation over a vehicle's lifetime using a piece-wise linear model.

1. The Robust Estimator. Next, we introduce Hu and Lawless [5] *robust estimator*, which forms the basis for our study. Let $n_i(t)$ be the total warranty cost (or number of claims) for vehicle i at time t, which is assumed to be discrete, that is $t = 1, 2, \ldots$. Let $N_i(t)$ be the accumulated warranty cost (or number of claims) up to and including time t for vehicle i. Note that "time" here can be either age, or mileage of the vehicle, not necessarily the calendar time. Suppose M vehicles have been under observation and their records are included in the warranty database. Let τ_i, $i = 1, 2, \ldots, M$, be the time that vehicle i has been under observation, that is from the vehicle's sale date until the time it is out of coverage or until the "cut-off" date of the dataset. We call τ_i the *observation time* of vehicle i. Its precise definition depends on whether "time" is age, mileage or actual time.

****2000 Mathematics Subject Classification:** 62G05, 60K10.
Key words: Mean Cumulative Function, incomplete information, warranty data.

Let $\hat{\Lambda}(t)$ be the estimator of $\Lambda(t) = E[N_i(t)]$, the mean cumulative warranty cost (or number of claims). In the discrete time case, the incremental rate function is $\lambda(t) = \Lambda(t) - \Lambda(t-1)$ with an initial condition $\Lambda(0) = 0$. Let $\delta_i(t)$ be the indicator of whether vehicle i is under observation at time t and, hence, eligible to generate a claim. For "time" is age case and "time" is mileage case, we have $\delta_i(t) = I(\tau_i \geq t)$. Then, the total warranty cost (or number of claims) at time t for all M vehicles is given by $n(t) = \sum_{i=1}^{M} \delta_i(t) n_i(t)$. Note that the products $\delta_i(t)n_i(t)$ is always known. Suppose if the observation process is independent of the event process, then the rate function can be estimated by $\hat{\lambda}(t) = \frac{n(t)}{MP(t)}$, where $P(t)$ is the probability that a vehicle is eligible to generate a claim at time t. This is the robust estimator proposed by [5], and $P(t)$ is assumed to be known. In practice, $P(t)$ is usually unknown and needs to be estimated.

Now, if $M(t) = MP(t)$, that is the number of vehicles that are eligible to generate a claim at time t, then, we get $\hat{\lambda}(t) = \frac{n(t)}{M(t)}$, and the associated mean cumulative function estimator is $\hat{\Lambda}(t) = \sum_{s=1}^{t} \hat{\lambda}(s)$, $t = 1, 2, \ldots, \tau_{\max}$, where $\tau_{\max} = \max(\tau_i)$ for $i = 1, 2, \ldots, M$. Under mild conditions and assuming known $M(t)$, [5] proved the asymptotic normality of $\hat{\Lambda}(t)$, with a standard error estimated by the square root of

$$\widehat{Var}[\hat{\Lambda}(t)] = \sum_{i=1}^{M} \left\{ \sum_{s=1}^{t} \left[\frac{\delta_i(s) n_i(s)}{M(s)} - \frac{\hat{\lambda}(s)}{M} \right] \right\}^2 . \tag{1}$$

Unless specified otherwise, the 95% confidence intervals in this article are evaluated using Eq. (1). We compute $\hat{M}(t)$ as an estimate of $M(t)$ from the warranty data and substitute $\hat{M}(t)$ to obtain $\hat{\lambda}(t)$, $\hat{\Lambda}(t)$, and the standard error of $\hat{\Lambda}(t)$. Note that the validity of Eq. (1), with $M(t)$ replaced by $\hat{M}(t)$, has not been proved yet. From now on, we use t to denote age, m mileage, and x the actual (calendar) time.

2. Mean Cumulative Function: A linear approach. "Time" is Age Case. Suppose we ignore the withdrawals from warranty coverage due to exceeding the mileage limit. Then, the estimate of the number of vehicles eligible to generate a claim at the target age t, $t \leq l_a$, is simply the number of vehicles age t or older, that is $\hat{M}(t) = \sum_{i=1}^{M} I(a_i \geq t)$, where a_i is the current age of vehicle i (on the "cut-off" date). This is the unadjusted estimator of $M(t)$. To get the true warranty claim rate, we need to adjust for withdrawals from warranty due to exceeding the mileage limit. Note that all adjustments made here and later will always be to $\hat{M}(t)$.

Let M_1 denote the number of vehicles with at least one claim and M_2 denote the number of vehicles without claims, such that $M_1 + M_2 = M$. Recall that the observation time is given by $\tau_i = \min(a_i, l_a, y_i)$, where y_i is the age at which the vehicles exceeds (or would exceed) the mileage limit. Since odometers are not monitored continuously, y_i is usually not known even for vehicles with claims. Thus, for a vehicle with at least one claim, we simply extrapolate y_i linearly using the age and mileage at the time of the most recent claim. Let $r_i = \beta_i/\alpha_i$, where α_i and β_i are the age and mileage of the vehicle at the latest claim respectively. Then, r_i is the estimated mileage accumulation rate (in miles per day) for vehicle i. Subsequently, at the target age t, vehicle i contributes to $\hat{M}(t)$ if it is old enough and if its mileage at age t is estimated to have been within the mileage limit l_m. Thus the contribution of vehicle i to $\hat{M}(t)$ is $I(a_i \geq t) I\left(r_i \leq \dfrac{l_m}{t}\right)$.

Figure 1 illustrates the above idea graphically using four hypothetical vehicles. The large square represents the warranty region, the little black circles represent the age and mileage for the latest claim of these vehicles, the little squares represent the extrapolated mileages for these vehicles at their current age (on the "cut-off" date) and the straight lines represent the trajectories of the vehicles. It can be seen that two of the vehicles are older than the target age t_q. But, one of them is estimated to leave the warranty coverage due to exceeding the mileage limit before age t_q, and, hence, it does not contribute to the adjusted (for mileage) value of $\hat{M}(t_q)$.

Now, we need to consider those vehicles that have not experienced a claim. By using the information on the vehicles with claims, we can construct an empirical distribution function for mileage accumulation rate as follows: $\hat{F}(r) = \dfrac{1}{M_1} \displaystyle\sum_{i=1}^{M_1} I(r_i \leq r)$, where M_1 is the number of vehicles with claims. Consequently, the probability that a typical vehicle remains in warranty coverage at age t is $\hat{F}\left(\dfrac{l_m}{t}\right)$, and, hence, the contribution to $\hat{M}(t)$ for a vehicle without claims is $I(a_i \geq t)\hat{F}\left(\dfrac{l_m}{t}\right)$.

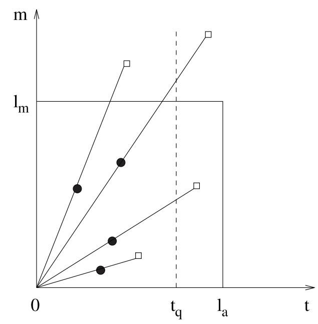

Fig. 1. Age case

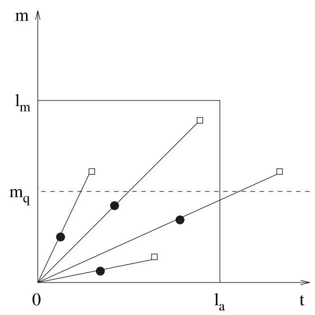

Fig. 2. Mileage case

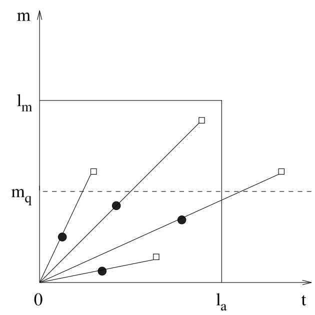

Fig. 3. Actual time

3. "Time" is Mileage Case. Let $\hat{M}(m)$ be the number of vehicles eligible to generate a claim at the target mileage m, $m \leq l_m$. For the unadjusted case, for a vehicle with at least one claim, its current mileage can be estimated by $a_i r_i$, where a_i is the current age (on the "cut-off" date) and r_i is the mileage accumulation rate based on the latest claim. Hence, the contribution to $\hat{M}(m)$ for the vehicle can be estimated by $I\left(r_i \geq \dfrac{m}{a_i}\right)$. Then, the contribution for a vehicle with no claims is given by $1 - \hat{F}\left(\dfrac{m}{a_i}\right)$. To adjust for withdrawals due to exceeding the age limit l_a, we simply replace a_i in the unadjusted case by $\min(a_i, l_a)$, the minimum of the vehicle's current age and the warranty age limit.

Figure 2 illustrates the idea for the "time" is mileage case. It can be seen that three of the vehicles are estimated to have exceeded the target mileage m_q at their current ages. But, one of them is estimated to have exceeded the age limit before mileage m_q, and hence it will not contribute to the adjusted (for age) value of $\hat{M}(m_q)$.

4. New Model: Actual Time Case. Next, we develop a new model for estimating the mean cumulative warranty cost per vehicle in the actual time case, $\Lambda(x)$. Again, all adjustments will be made to $\hat{M}(x)$. Let X denote the current time (the "cut-off" date). Suppose we ignore the withdrawals from warranty coverage due to mileage, then the estimate of the number of vehicles eligible to generate a claim at the target time x, $x \leq X$, is simply the number of vehicles sold before or at time x and still within the age limit, i.e.,

$$\hat{M}(x) = \sum_{i=1}^{M} I(s_i \leq x) I(z_i \leq l_a), \tag{2}$$

where s_i is the sale date and z_i is the age at the target time of vehicle i. This is the unadjusted estimator of $M(x)$ (or we can say that this is the adjusted for age estimator, since we have taken into account the age limit).

To get the true warranty claim rate, the adjustment for withdrawals due to mileage is: for a vehicle with at least one claim $I(s_i \leq x)I(z_i \leq l_a)I\left(r_i \leq \frac{l_m}{z_i}\right)$ and for a vehicle with no claim $I(s_i \leq x)I(z_i \leq l_a)\hat{F}\left(\frac{l_m}{z_i}\right)$, where $\hat{F}(r)$ is the empirical distribution function for mileage accumulation rate as given in Section 2.

Figure 3 illustrates the idea graphically using four hypothetical vehicles. It can be seen that only the vehicle sold at time s_3 is under warranty coverage at the target time x_q, and will contribute to $M(x_q)$. The other three vehicles are all out of warranty coverage at time x_q. Both of the vehicles sold at time s_1 leave warranty coverage due to age, and one of them is also estimated to have exceeded the mileage limit before time x_q. The vehicle sold at time s_2 is still within the age limit, but it is estimated to have exceeded the mileage limit before time x_q.

5. Example: The P-claims Dataset. Table 1 shows the summary of a warranty dataset up to four different "cuts" in time: 1 January 2001, 1 January 2002, 1 January

Table 1. Summary of the P-claims dataset

	01/01/2001	01/01/2002	01/01/12003	24/10/2003
Number of vehicles sold	16764	44761	44879	44890
Number of vehicles with claims	1669	12628	18882	21736
Number of claims	2554	25518	46820	59144
Total cost of claims	86122	751145	1464578	1953220
Number of vehicles with P-claims	48	508	974	1247
Total number of P-claims	50	579	1166	1510
Total cost of P-claims	14512	123292	222825	264539
Median vehicle age (days)	92	322	687	983
Median MAR (miles per day)	40	42	41	38
Median reporting delay (days)	11	8	8	8

2003, and the actual “cut-off” date 24 October 2003 (the latest process date of a claim). We assume that this actual “cut-off” date is the “current date”. By using this dataset, we analyse the warranty cost on one major system of the vehicle, which is not identified here but referred to as System P with P-claims. Note that the median mileage accumulation rate (MAR) is around 40 miles per day over the four time cuts. This is more than the rate of 33 miles per day, which corresponds approximately to reaching the 36 000-mile limit in three years. Thus, most vehicles leave the warranty coverage due to mileage. Also, we should emphasize that we use mileage information form all claims, not just those for System P. The table below provides a summary of the P-claims dataset.

Figure 4 shows the unadjusted and adjusted $\hat{\Lambda}(t)$ up to the age limit of $t = 1095$ days with its 95% confidence intervals (CI). Figure 6 shows the unadjusted and adjusted

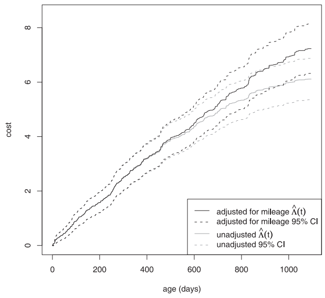

Fig. 4. Unadjusted/adjusted $\hat{\Lambda}(t)$

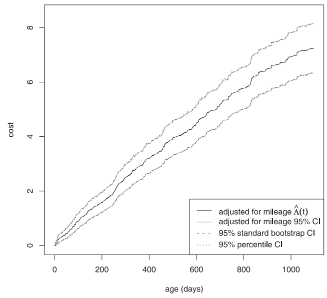

Fig. 5. Adjusted $\hat{\Lambda}(t)$, different 95% CI

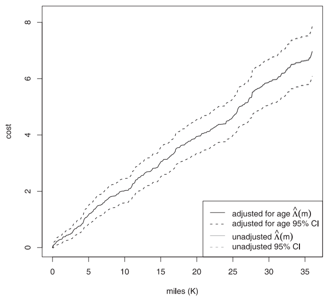

Fig. 6. Unadjusted/adjusted $\hat{\Lambda}(m)$

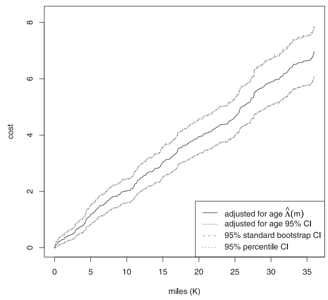

Fig. 7. Adjusted $\hat{\Lambda}(m)$, different 95% CI

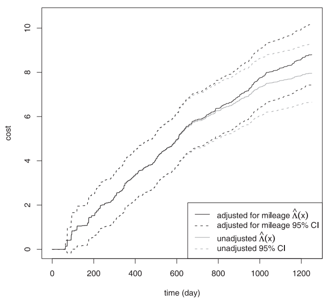

Fig. 8. Unadjusted/adjusted $\hat{\Lambda}(x)$

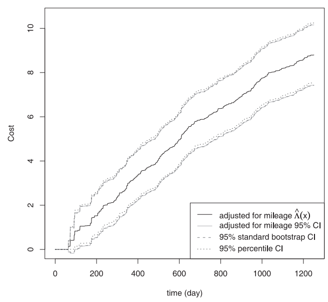

Fig. 9. Adjusted $\hat{\Lambda}(x)$, different 95% CI

$\hat{\Lambda}(m)$ up to the mileage limit of $m = 36\ 000$ miles with its 95% CI. Figure 8 shows the unadjusted and adjusted $\hat{\Lambda}(x)$ from 22 May 2000 ($x = 1$, the first sale date) until 24 October 2003 ($x = 1251$, the "cut-off" date) with its 95% CI. Figures 5, 7, and 9 show the 95% confidence intervals evaluated using Eq. (1), the 95% standard bootstrap confidence intervals, and the 95% percentile confidence intervals for the three estimates.

REFERENCES

[1] W. Blischke, P. Murthy. Warranty Cost Analysis, New York, Marcel Dekker, 1994.

[2] W. Blischke, P. Murthy. Product Warranty Handbook, New York, Marcel Dekker, 1996.

[3] D. Christozov, S. Chukova, J. Robinson. Automotive Warranty Data: Estimation of the Mean Cumulative Function using Stratification Approach, Technical Report, SMSCS, Victoria University of Wellington, 2008, 1–20.

[4] S. Chukova, J. Robinson. Estimating Mean Cumulative Functions from Truncated Automotive Warranty Data, In: Modern Statistical and Mathematical Methods in Reliability, Eds. A. Wilson, N. Limnios, S. Keller-McNulty, Y. Armijo, Singapore, World Scientific, 2005, 121–135.

[5] X. Hu, J. Lawless. Estimation of Rate and Mean Functions from Truncated Recurrent Event Data, *Journal of the American Statistical Association*, **91** (1996), 300–310.

[6] J. Lawless. Statistical Analysis of Product Warranty Data, *International Statistical Review*, **66** (1998), 41–60.

[7] J. Lawless, X. Hu, J. Cao. Methods for the Estimation of Failure Distributions and Rates from Automotive Warranty Data, *Lifetime Data Analysis*, **1** (1995), 227–240.

Stefanka Chukova
Her Guan Teo
School of Mathematics, Statistics and Operations Research
Victoria University of Wellington, New Zealand
Kelburn Parade Gate 6
Wellington 6040, New Zealand
e-mail: stefanka@gmail.com
teoherguan@gmail.com

АВТО ГАРАНЦИОННИ ДАННИ: СРЕДНА КУМУЛАТИВНА ФУНКЦИЯ

Стефанка Чукова, Хър Гуан Тео

В това изследване разглеждаме и разширяваме предишната ни работа по цензуриране, типично за авто гаранционни данни. За да разрешим проблема с непълната информация за километража, използваме линеен подход в непараметрични рамки. Оценяваме средните кумулативни гаранционни разходи (за превозно средство) и стандартната им грешка като функция на възрастта, на километража и на реалното (календарно) време.