Automotive Warranty Data: Estimation of the Mean Cumulative Function using Stratification Approach

Dimitar Christozov
American University in Bulgaria, BG
dchistozov@aubg.bg

Stefanka Chukova
Victoria University of Wellington, NZ
stefanka@mcs.vuw.ac.nz

Jeff Robinson
General Motors R&D Center, USA
jeffrey.a.robinson@gm.com

March 17, 2008

Abstract

This study extends previous work that accounted for mileage accumulation in automotive warranty data analysis. Coverage is typically limited by age as well as mileage. Age is known for all sold vehicles all the time, but mileage is only observed for a vehicle with a claim and only at the time of the claim. Here we consider either age or mileage as the usage measure. We evaluate the mean cumulative number of claims or cost of claims and its standard error as functions of the usage measure. Within a nonparametric framework, we account the rate of mileage accumulation and allow for variation in the rate of mileage accumulation over a vehicle's lifetime. We illustrate the ideas with real data on four cases based on whether the usage measure is age or miles and whether the results are adjusted for withdrawals from warranty coverage.

Key words: automotive warranty data analysis, mean cumulative function, nonparametric framework

1 Introduction

Warranty data is of considerable interest to corporations for several reasons. Warranty claims are a liability incurred at the time of sale and represent a cost of doing business, so forecasting those costs is of interest. For engineers a secondary but important use of warranty data is to assess the reliability of products in the field. A third characteristic of warranty coverage is that it is a product attribute valued by customers and affecting their buying decisions. For example increasing warranty coverage may attract more buyers but also increase costs. See Robinson and McDonald [10] and Blischke and Murthy [1] for more discussion on these general points.

Automotive warranties in many other countries generally offer a free repair subject to age and mileage limits. Age is measured from the time of sale to the customer. At this writing the most common limits in the U.S.A. for "bumper-to-bumper" coverage are thirty-six months or thirty-six thousand miles, whichever comes first. Some manufacturers offer longer warranties across the board. Others offer longer warranties only on their luxury models or for selected components such as powertrain. An important, but seldom discussed distinction is that in some cases the warranty coverage is transferable to a subsequent buyer, but in other cases it is not.

Vehicle ages are known at all times because sales records are retained. It is also becoming technically feasible to track mileage accumulation on all vehicles in the field, but this is currently not a common practice for cost and privacy reasons. But for vehicles that generate a warranty claim, the mileage at the time of a warranty repair are recorded at the dealership and included in the warranty database. Thus from a modeling standpoint we have two usage measures (age and mileage), but one of them (mileage) is incompletely observed. As is commonly done warranty claims are modelled here as recurrent events from a repairable system. Also, we take a nonparametric approach because sample sizes are large. We note however that warranty fo recasting, which requires extrapolation beyond the oldest age/mileage in the field, requires either a parametric model or the incorporation of past-model data on older vehicles. We deal explicitly with the problem of incomplete mileage information and also with the problem that repairs made beyond the age or mileage limits will not be part of the warranty database.

This modeling approach and data structure has been discussed extensively before. The model and estimation procedure are based on the "robust estimator" discussed by Hu and Lawless [3]. The consideration of the number of units at risk due to say mileage limitations and based on incomplete mileage information extends Nelson's [8] standard estimator. Lawless, Hu and Cao [6] also dealt with the incomplete data problem and specify a simple linear mileage accumulation model, which is generalized in this study. The general survey paper by Lawless [5] includes a discussion of the bias caused by reporting delay, which was analyzed earlier in Lawless and Nadeau [7] and Kalbfleisch, Lawless and Robinson [4]. Robinson [9] discussed finite population corrections to the variance estimators.

Chukova and Robinson [11] adopted the robust estimator and the simple linear mileage accumulation model to estimate the number of units at risk at any given time from the incomplete mileage data. This information was used to provide explicit expressions for the estimated mean cumulative number of claims per vehicle with or without adjustments for reporting delay. Here we relax the linearity assumption for mileage accumulation, proposing instead a piece-wise linear model with nodes occuring at the observed mileages corresponding to warranty repairs. In Chukova and Robinson only the last warranty claim was used to estimate a vehicle's mileage accumulation rate. For the piece-wise model, we use all claims in the database to characterise mileage accumulation. We define vehicle strata based on mileage accumual tion rates as well as the variability of mileage accumulation rates over time. We compare estimated mean cumulative functions based on this stratification approach with the more basic ones given by Chukova and Robinson.

Section 2 defines the basic notation and summarizes the Hu and Lawless model as adopted by Chukova and Robinson. Then Section 3 presents the more general model based on the stratification approach. The compuations for the stratification model are described in Section 4. Section 5 includes an example based on the same dataset analyzed by Chukova and Robinson comparing the two methods. A summary discussion is found in Section 6.

2 The Model

Using notations and terminology from Hu and Lawless [4] and Chukova and Robinson [2], let $n_i(t)$ be the number of claims at time t for vehicle i. (In the formulas that follow $n_i(t)$ could also denote the cost of the claims at time t for vehicle i). It is convenient and not restrictive to think of time as discrete, i.e. $t = 1, 2, ...$ Let $N_i(t)$ be the accumulated number of claims (or cost) up through and including time t for vehicle i. "Time" is either age or mileage of the vehicle, not the calendar time. Let M be the number of vehicles under observation, whose records make up the warranty database. Let τ_i be the"time" that unit i has been under observation. Precise definition of the τ's will depend on whether time is age or mileage.

Hu and Lawless [4] and Chukova and Robinson [2] obtained estimators $\hat{\Lambda}(t)$ of the population mean cumulative function $\Lambda(t) = EN_i(t)$. The incremental rate function can be written as $\lambda(t) = \Lambda(t) - \Lambda(t-1)$ with the initial condition $\Lambda(0) = 0$. Let $\delta_i(t) = I(\tau_i \geq t)$ be the indicator of whether car i is under observation at time t. Then

$$n_{.}(t) = \sum_{i=1}^{M} \delta_i(t) n_i(t)$$

is the total number of claims observed at time t for all M vehicles. Note that $\delta_i(t)$ may be unknown for some of the cases, but the product $\delta_i(t)n_i(t)$ is always known. That is for the numerator in the rate calculations we do not distinguish between a vehicle not under observation at time t versus one that is under observation but produces no claims.

Let $M(t) = M$ denote the number of vehicles eligible to generate a claim at time t and denote its generic estimator by $\hat{M}(t)$. Then the rate function estimate is given by

$$\hat{\lambda}(t) = \frac{n_{.}(t)}{\hat{M}(t)} \tag{1}$$

with associated mean cumulative function estimator

$$\hat{\Lambda}(t) = \sum_{s=1}^{t} \hat{\lambda}(s), \quad t = 1, 2, \ldots, \max_{1 \leq i \leq M}(\tau_i). \tag{2}$$

Assuming known $M(t)$ and some other mild conditions, Hu and Lawless [3] show asymptotic normality of $\hat{\Lambda}(t)$ with a standard error given by the square root of

$$\hat{V}ar[\hat{\Lambda}(t)] = \sum_{i=1}^{M} \left(\sum_{s=1}^{t} \left[\frac{\delta_i(s) n_i(s)}{M(s)} - \frac{\hat{\lambda}(s)}{M} \right] \right)^2. \tag{3}$$

Chukova and Robinson estimated $M(t)$ from the warranty data itself and substituted this estimate into (3) to obtain approximated standard errors. We follow the same course here but propose an alternative estimator of $M(t)$, which as mentioned earlier, makes use of essentially all available mileage accumulation information. The stratification approach used to derive this alternative estimation is discussed in the next section.

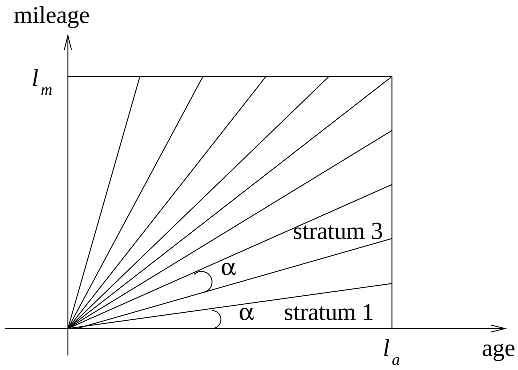


Figure 1: Warranty region with strata

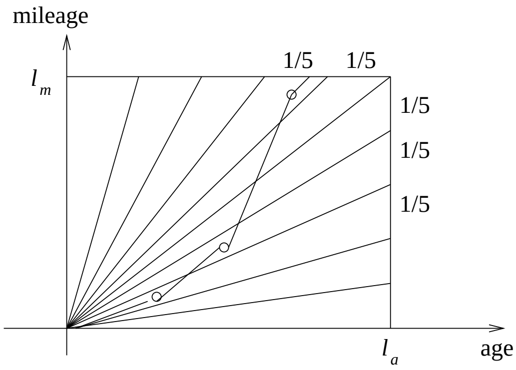


Figure 2: Trajectory of a 5_{10}-stable car

3 Modeling mileage accumulation: A Stratification Approach

Next, we extend the ideas of Chukova and Robinson [2] and Hu and Lawless [4], by relaxing the linearity assumption regarding the automobile mileage accumulation. We propose an empirical piece-wise linear model to deal with this issue. We model the vehicle driving pattern (behaviour) by focusing on two factors:

- the mileage accumulation rate,
- the variability of the mileage accumulation rate.

We split the group of M cars into two subgroups, so that the first one consists of M_1 cars with a claim record and the second one consists of M_2 cars with no claim record, where $M = M_1 + M_2$. Next, we propose a model for the mileage accumulation based on the observed driving patterns of the cars with claims and impose this model on the cars with no claims.

3.1 Grouping the cars with claims according to their driving patterns

First we identify several groups of cars depending on the pattern of their mileage accumulation rate, which is observed through their claim records. We measure the variability of the mileage accumulation rate by introducing a partition of the warranty region into strata, as shown in Figure 1. The variability of the mileage accumulation rate for a car is measured by assigning the car to a particular group. These groups are associated with the number of strata the trajectory of the car goes through during its warranty "life", where the trajectory of the car is assumed to be piece-wise linear between its consecutive claims and remains within the stratum of its last claim, as shown in Figure 2.

Definition 1. We say that a car is stable with respect to a certain range if the trajectory of this car remains within this range throughout its warranty "life".

Let P_k be a regular strata partition of the warranty region with a stratum angle equal to $\alpha = \frac{\pi}{2k}$. For example, in Figure 1 the partition of the warranty region is P_{10}. This partition is regular because the corresponding strata are of equal size, i.e., the size of the stratum angle $\alpha = \frac{\pi}{20}$ is the same for all strata. We say that a car is 1_k-stable with respect to partition P_k, where k is a positive integer number, if all claims of this car belong to a single stratum of this partition. We denote the size of this group by M_1^1. In addition, we denote by $O_{1,s}^1$, $s = 1, 2, \ldots, k$ the number of 1_k-stable cars, for which claims fall into the s^{th} stratum of P_k, where

$$O_{1,1}^1 + O_{1,2}^1 + \ldots + O_{1,k}^1 = M_1^1.$$

These counts $O_{1,s}^1$ will be used to estimate a strata distribution, which reflects the proportion of vehicles within each of the strata.

Next, we consider the group of cars, say of size $M_1^{k_i}$, k_i are positive integer numbers, $1 < k_1 < k_2 < \ldots < k$, such that their claims are spread within exactly k_i strata of the partition P_k. These are the k_{i_k}-stable cars, i.e., they are stable with respect to an aggregated stratum consisting of k_i of P_k's strata. In order to estimate the strata distribution we need to take into account the contribution of the k_i-stable vehicles to the number of vehicles in the strata. This contribution is assumed to be uniform over the k_i strata included in the aggregated stratum with respect to which this vehicle is k_i-stable. In other words, each k_i-stable vehicle contributes a fraction of $\frac{1}{k_i}$ to each of these k_i strata. The piece-wise linear trajectory of such a car goes through k_i strata, starting with the stratum that contains the earliest claim. Figure 2 depicts the trajectory of an 5_{10}-stable car, which goes through strata 3 to 7. The contribution of this car to the count of cars in each of these strata is equal to $\frac{1}{5}$, i.e., the contribution of this car is uniformly distributed over the strata, which consist of its trajectory. Therefore, the corresponding counts $O_{1,s}^{k_i}$, $s = 1, 2, \ldots, n_i$ are also well defined.

We continue in the same way and identify groups of cars that are stable with respect to a prespecified size of an aggregated stratum. As a result of this procedure we partition the set of all M_1 cars with claims into non-overlapping groups, called driving patterns groups (DPG). Within each of these DPG, the variability of the automobile mileage accumulation follows a similar pattern. Assume that k_d is the size of the largest aggregated stratum needed to describe the stable driving patterns of the cars with claims. The value of k_d reflects on how accurate the model accounts for the variability of the driving patterns of the vehicles.

Then, following the above notations, we will have several groups of cars, such that, the first one consists of M_1^1 1_k-stable cars with $O_{1,s}^1$, $s = 1, 2, \ldots, k$ number of cars in the s^{th} stratum, the second one consists of $M_1^{k_1}$ k_{1_k}-stable cars with $O_{1,s}^{k_1}$, $s = 1, 2, \ldots, k$ number of cars in the s^{th} stratum and so on and the last one consists of $M_1^{k_d}$ k_{d_k}-stable cars with $O_{1,s}^{k_d}$, $s = 1, 2, \ldots, k$ number of cars in the s^{th} stratum.

After these groups are identified, there will be a set of cars, say of size $M_1^{k_{d+1}}$, with unstable driving pattern. We say that these cars are unstable because their driving trajectory during their warranty 'life" goes through more than k_d strata of the initial partition P_k. For this group of cars we apply the model for the mileage accumulation in Chukova and Robinson [2], i.e., we assume that their trajectories are linear, determined by their last claim and identify the counts $O_{1,s}^{k_{d+1}}$, $s = 1, 2, \ldots, n_i$ of unstable cars in the s^{th} stratum. Thus, the whole set of cars with claims is partitioned into non-overlapping groups, such that

$$M_1^1 + M_1^{k_1} + \ldots M_1^{k_d} + M_1^{k_{d+1}} = M_1.$$

Moreover, the contribution of these groups to each of the strata is also known. Therefore the number of cars with claims $O_{1,s}$ within the s^{th} stratum is

$$O_{1,s} \;=\; O_{1,s}^1 + O_{1,s}^{k_1} + \ldots O_{1,s}^{k_d} + O_{1,s}^{k_{d+1}} \text{ for } s = 1, ..., k. \tag{4}$$

Note: The notion of unstable cars is not important. We can build up the model by looking at all possible DPG for a given regular partition P_k. For a regular partition with k strata, we need to look at 1_k-stable cars, 2_k-stable cars, ..., k_k stable cars, i.e., the total number of DPG needed to place each car in one of the DPG is equal to k. Thus, using this approach the number of unstable vehicles will be equal to zero.

3.2 Estimating the strata distribution

Now we are ready to estimate the strata distribution. Using the strata counts $O^{1}_{1,s}$, $O^{k_1}_{1,s}, \ldots,$ $O^{k_d}_{1,s}, O^{k_{d+1}}_{1,s}$ we can estimate the strata distribution $p = (p_1, p_2, \ldots, p_k)$, where

$p_s = Pr$(a vehicle with claim belongs to s^{th} stratum). Thus, we obtain

$$p_s \;=\; \frac{O_{1,s}}{M_1}, \text{ for } s = 1, 2, \ldots, k, \tag{5}$$

where $O_{1,s}$ for $s = 1, 2, \ldots, k$ are given by (4). The strata distribution reflects the mileage accumulation rate for the cars with claims. While estimating this distribution, the variability of the driving patterns is taken into account. In addition, we assume that the strata distribution is time independent, i.e., it remains the same over different age intervals.

3.3 The mileage accumulation model for cars with no claims

Often in the warranty database, there is a large group of cars of size M_2, with no claim records. We assume that the driving behaviour of these cars is probabilistically identical to the driving behaviour of the cars with claims. In other words, we assume that the strata distribution is a reasonable representation of the driving patterns for the cars with no claims, and therefore, the strata distribution describes the driving patterns of all cars in the database.

4 Computing the mean cumulative number of claims or cost of claims

Next, we propose a procedure for computing the mean cumulative number of claims, and related to it, mean cumulative cost of claims under two different meanings of the parameter "time", namely, the "time" is age of the vehicles or the "time" is the mileage of the vehicles.

4.1 "Time" is Age Case

Define a regular partition $0 = a_0 < a_1 < \ldots < a_{n-1} < a_n = l_a$ of size n, i.e.,

$$a_i - a_{i-1} = h^{(a)}, \;\; i = 1, 2, \ldots, n, \quad \text{where } h^a = \frac{l_a}{n}, \;\; a_0 = 0,$$

and l_a is the warranty age-wise limit, as given in Figure 3. If necessary, we can extend this partition beyond the warranty age limit l_a. The "time" discretization is defined by the step $h^{(a)}$, i.e., the "time" t assumes discrete values $t_i = i\, h^{(a)}$, such as $0 = t_0 = a_0, t_1 = a_1, \ldots, t_{n-1} = a_{n-1}, t_n = a_n = l_a, t_{n+1} = l_a + h^{(a)}, \ldots$. Moreover, since the age of the vehicles is known, no matter whether it is a vehicle with a claim record or with no claim record, the number of vehicles N_{a_i} with age within an age-bin $\triangle^{(a)}_i = [t_{i-1}, t_i)$, $i = 1, 2, \ldots, n$, is known.

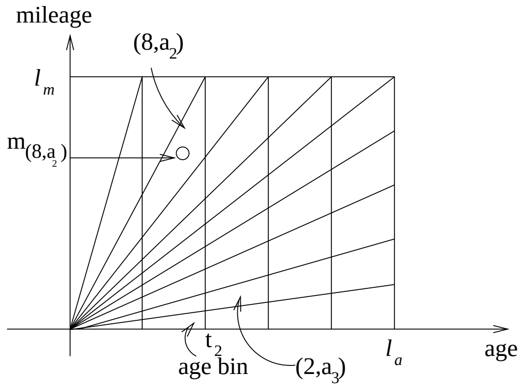


Figure 3: Age bins

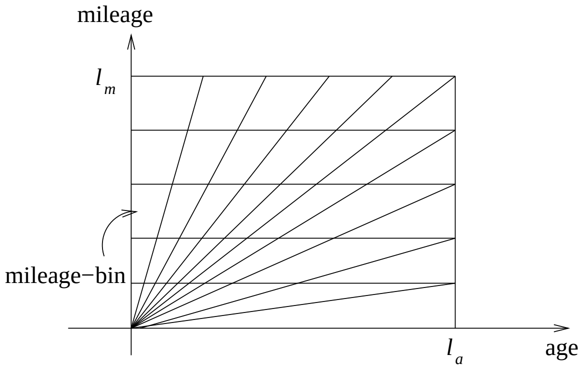


Figure 4: Mileage bins

If we ignore withdrawals from coverage due to mileage, then the number of vehicles eligible to generate a claim at age t is simply the number of vehicles age t or older, i.e.,

$$\hat{M}(t_l) = \sum_{i=1}^{M} \delta_i(t_l). \tag{6}$$

This estimator correctly characterizes how warranty claim rates and costs actually accrue as a function of age, and it may be useful for prediction. However, in addition to the inherent claim rate as a function of age, it is influenced by the accumulation of mileage because this affects how long vehicles remain in warranty coverage. To illustrate the point, suppose hypothetically that warranty claims occur due to age and not miles, and suppose that drivers begin to drive more. The inherent reliability of the population of vehicles would not change, but the warranty claim rate would go down because more vehicles would drop out of coverage sooner due to the mileage limit.

To get at this "true" warranty claim rate, we will adjust for the fact that some vehicles leave coverage by exceeding the mileage limit. The numerators in (1) are available because they are the number of claims or cost for the vehicles at age t and are available in the database. Here and later the adjustment will always be to $\hat{M}(t)$ in (1). To capture this rate the "time" under observation is now defined as $\tau_i = \min(age_i, y_i)$, where y_i is the age in days at which the i^{th} vehicle exceeds (or would exceed) the mileage limit of $l_m = 36,000$ miles and age_i is its current age.

Since odometers are not monitored continuously, y_i is not known even for vehicles that have had a claim. For a target age t, a vehicle 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. Therefore, based on our stratification model for the mileage accumulation, the estimator for the denominator of (1) is given by

$$\hat{M}(t_l) = \left(M - \sum_{i=1}^{l} N_{a_i} \right) \left(\sum_{j=1}^{k-l} p_j \right). \tag{7}$$

For example, looking at Figure 3, assume that we want to estimate $\hat{M}(t_2)$. We need to estimate the number of vehicles age t_2 or older, which at time t_2 are still within the warranty region, i.e., at time t_2 they have not exceeded the mileage limit l_m. The number of vehicles age t_2 or older is equal to $M - \sum_{i=1}^{2} N_{a_i}$. At time t_2, vehicles with driving patterns associated with the 10^{th} and 9^{th} strata would have left the warranty coverage, whereas vehicles with driving patterns associated with strata $1-8$ will be still within the warranty. Therefore, the proportion of vehicles within the warranty region at time t_2 is $\sum_{i=1}^{8} p_i$. Therefore the required estimate is

$$\hat{M}(t_2) = \left(M - \sum_{i=1}^{2} N_{a_i} \right) \left(\sum_{j=1}^{8} p_j \right). \tag{8}$$

4.2 "Time" is Miles Case

If warranty claim rates are more closely related to mileage than age, then we may wish to analyze by miles. This could occur, for example, in some engineering applications where we would expect warranty incidents to occur more as a result of usage than of age. Throughout the remainder of the paper we will use the argument "m" for mileage. The fact that the exact mileage is unknown except at the time of a claim complicates the calculation of $\hat{M}(m)$, but the mileage accumulation model used in the previous subsection can produce reasonable results. Again, the numerators in (1) are available because they are the number of claims or cost of claims for the vehicles at mileage m and are available in the database.

For the unadjusted for age case the "time" under observation for vehicle i is $\tau_i = m_i$, the current mileage. As before it will be convenient to think of mileage as a discrete variable. We discretize the mileage by creating a regular partition $0 = m_0 < m_1 < \ldots < m_{n-1} < m_n = l_m$ of size n, i.e.,

$$m_i - m_{i-1} = h^{(m)}, \;\; i = 1, 2, \ldots, n, \quad \text{where } h^{(m)} = \frac{l_m}{n}, \;\; m_0 = 0,$$

and l_m is the warranty mileage-wise limit, as given in Figure 4. If necessary, we can extend this partition beyond the warranty mileage limit l_m. The "time" discretization is defined by step $h^{(m)}$, i.e., the "time" t assumes discrete values $m_i = i\,h^{(m)}$ for $i = 1, 2, \ldots, n, \ldots$, such as $0 = m_0, m_1, \ldots, m_{n-1}, m_n = l_m, m_{n+1} = l_m + h^{(m)}, \ldots$. In addition, consider an age-strata grid, say (s, a_i), determined by the stratum s, $s = 1, 2, \ldots, k$ and the age-bin $\triangle_i^{(a)}$ for $i = 1, 2, \ldots, n, \ldots$, as shown in Figure 3. We use this age-strata grid to estimate the number of cars within appropriately specified mileage-bins $\triangle_i^{(m)} = [m_{i-1}, m_i)$. As in the case of time is "age", in order to provide an estimator for $\hat{M}(m)$, we need to know the number of vehicles N_{m_i} with mileage within a mileage-bin $\triangle_i^{(m)} = [m_{i-1}, m_i)$, $i = 1, 2, \ldots, n$. The current mileage is not known exactly, even for vehicles with claims, but it can be estimated by using the strata distribution and the age-strata grid (s, a).

We estimate the N_{m_i}'s by using ideas similar to the ideas of analysing group data. For each cell (s, a_i) from the age-strata grid we identify a typical mile-representative, say $m_{(s,a_i)}$. For example, $m_{(8,a_2)}$ represents the mileage for the $(8, a_2)$'s cell, as shown in Figure 3. Therefore, we estimate that

$$N_{m_{(s,a_i)}} = p_s N_{a_i} \tag{9}$$

is the number of cars with current mileage equal to $m_{(s,a_i)}$. Hence, for each cell of the age-strata grid we identify three numbers: the strata number s, the average mileage $m_{(s,a_i)}$ and corresponding number $N_{m_{(s,a_i)}}$. Next, we estimate the number of vehicles N_{m_i} with mileage within a mileage-bin $\triangle_i^{(m)}$ for $i = 1, 2, \ldots, n, \ldots$, by simply adding the numbers of cars with

Year	1998	1999	2000	2001
Number of sold vehicles	40 048	44 755	34 807	44 890
Number of cars with claims	25 011	25 504	21 473	21 103
Number of claims	62 123	60 127	43 814	45 430
Total cost of the claims	6 978 476	6 670 940	2 217 578	2 091 608

Table 1: Summary of 1998 - 2001 warranty data

typical mile-representatives that fall within the i^{th} mileage-bin $\triangle_i^{(m)}$, i.e.,

$$N_{m_i} = \sum_{m_{(s,a_i)} \in \triangle_i^{(m)}} N_{m_{(s,a_i)}} \tag{10}$$

For the unadjusted case, we estimate N_{m_i} for $i = 1, 2, \ldots, n, \ldots$, by extending the age-strata grid beyond the warranty age limit l_a to cover all claims including those that are outside of the warranty coverage, i.e. the estimator for $\hat{M}(m)$ is given by

$$\hat{M}(m) = M - \sum_{i=1}^{k^*} N^*_{m_i}, \tag{11}$$

where the counts $N^*_{m_i}$ within $\triangle_i^{(m)}$ are estimated using the extended age-strata grid. The value of k^* is equal to the number of age-bins in the extended age-strata grid.

To adjust $\hat{M}(m)$ for vehicles leaving coverage due to age, we want to assure that the target mileage m is reached before the vehicle leaves the warranty due to age, i.e., we consider only claims that are within the warranty coverage. Thus, the age adjustment of the estimator is achieved by

$$\hat{M}(m) = M - \sum_{i=1}^{k} N_{m_i}, \tag{12}$$

where k is the number of age-bins within the warranty coverage and N_{m_i} are estimated using only these k age-bins.

5 Example

We illustrate the ideas on a set of actual warranty data for a particular vehicle make over four year models. Table 1 provides a summary of these data.

\ DPG MAR \	S	M_1	M_3	M_6	$M_{>6}$	Total
< 450	41.14	131.16	99.08	111.69	229.20	122.45
450 - 1080	41.65	115.32	136.57	160.93	152.19	121.33
1080 - 2600	51.49	149.11	138.00	166.57	165.36	134.10
> 2600	29.75	72.37	89.06	69.56	199.85	92.07
Total	45.65	131.44	132.36	155.88	165.91	126.25

Table 2: Average cost per vehicle - 2001 DPG

\ DPG MAR \	S	M_1	M_3	M_6	$M_{>6}$	Total
< 450	25.86	90.22	127.91	124.77	249.24	123.60
450 - 1080	29.96	112.51	143.74	173.07	170.39	125.93
1080 - 2600	34.34	128.91	163.80	173.95	215.18	143.23
> 2600	30.70	114.23	152.30	114.92	367.07	155.85
Total	30.21	111.47	146.94	146.68	250.47	137.15

Table 3: Average cost per vehicle - 2000 DPG

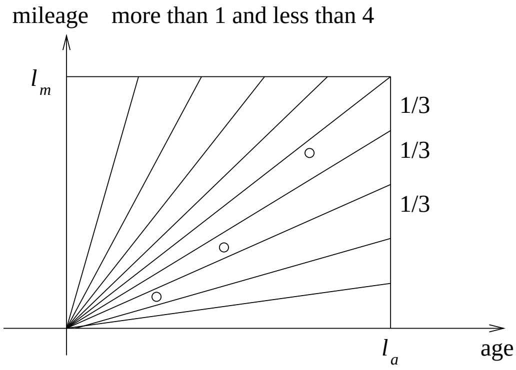


Figure 5: Strata contribution for M_3

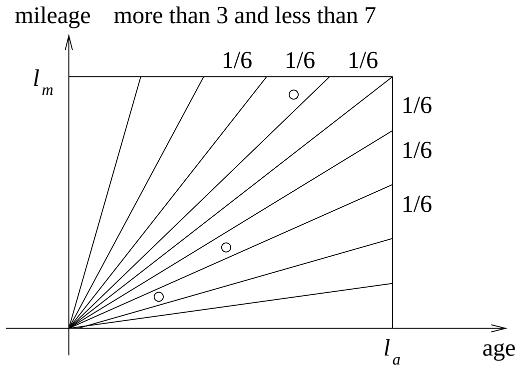


Figure 6: Strata contribution for M_6

5.1 The strata distribution

The warranty region is divided into age-bins of size of one month and mileage-bins of size of 1000 miles. The strata partition is chosen to be of size $k = 72$, i.e, each stratum is reasonable narrow with an angle equal to

$$\frac{\pi}{2k} = \frac{\pi}{144} = 0.0218166 \text{ radians.}$$

The set of all cars with clams is partitioned into several non-overlapping driving patterns groups (DPG) as follows:

- cars with a single claim - denoted as group S
- cars with multiple claims
  - cars with all claims within one stratum - group M_1
  - cars with all claims within three strata - group M_3
  - cars with all claims within six strata - group M_6
  - cars with claims spread over more than six strata - group $M_{>6}$

Next, we aim to estimate the strata distribution for the above set of DPG. We determine the contribution of a car to a stratum in the following way:

- For cars with a single claim - group S
  - each car belongs to a single stratum determined by its claim
- For cars with multiple claims
  - cars with all claims within one stratum - DPG M_1
    * each car belongs to a single stratum determined by its claim
  - cars with all claims within three strata - DPG M_3
    * each car is uniformly distributed over three strata, starting from the stratum consisting of its earliest claim
  - cars with all claims within six strata - DPG M_6

    - each car is uniformly distributed over six strata, starting from the stratum consisting of its earliest claim
  - cars with claims spread over more than six strata - group $M_{>6}$
    - unstable cars - each car belongs to a single stratum determined by its last claim

Figure 5 illustrates the contribution of a car from DPG M_3 to the strata distribution, whereas Figure 6 shows this contribution for a car from DPG M_6.

Figure 7 depicts the strata distributions for the three consecutive 1999 - 2001 year models, and it is easy to see that these distributions are very similar.

Next, we explore the relationship between the variability of the driving pattern and the warranty cost. We extracted the information needed to study this relationship from the database and its summary for 2000 and 2001 is given in Table 2 and Table 3. In these tables the mileage accumulation rate (MAR) is in miles per month. Similar summaries for 1998 and 1999, not shown here, were also available. The pictorial representation of these results is given in Figure 9 and Figure 11. These figures differ by the number of DPG considered in the model, which reflects on how variability of the driving patterns is accounted for. Figure 9 shows the warranty cost per car for different driving patterns, where the driving patterns are represented by S, M_1, M_3, M_6 and $M_{>6}$ groups as defined above. On the other hand, Figure 11 shows the warranty cost per car for different DPG, where the driving patterns are represented by $M_1, M_2, M_3, \ldots, M_{11}$ and $M_{>11}$ groups. These figures suggest that the variability of the driving pattern affects the warranty cost. The upward trend of the plots suggests that a higher variability leads the higher warranty cost. This is a very interesting observation that needs to be studied further. At the same time, it suggests that in modelling the mileage accumulation its variability should be taken into account and that modelling only the "average" driving behaviour does not provide enough information to study the relationship between the driving behaviour and related warranty costs.

Figure 10 depicts the warranty cost as a function of mileage accumulation rate (MAR) for years 2000 and 2001 with partition of MAR ($< 450, 450-1080, 1080-2600, > 2600$). We have looked at similar plots for 1998 - 2001 with much finer partition of MAR, shown in Figure 12, but, in this study, no consistent relationship between warranty cost and MAR has been detected.

5.2 The "P-claims" Dataset

We illustrate the methods for calculating the mean cumulative number of claims or cost of claims and its standard error as functions of the usage measure on a set of warranty data with 44,890 records taken from model year 2001 vehicles sold mainly in calendar years 2000 and 2001. We examine the warranty claims on one major system of the vehicle, which is not identified. It will be referred to as "System P". Table 4 summarizes the dataset. In order to illustrate our points we created versions of the dataset as it would have existed at four different "cuts" in time: Jan. 1, 2001; Jan. 1, 2002; Jan. 1, 2003; and the actual "cut" date for our original dataset, Oct. 24, 2003. These are displayed in Table 4 along with the descriptive statistics of the datasets up through the respective dates. For proprietary reasons a few vehicles were randomly selected, and their records were deleted from the original dataset. Also, the costs have been re-scaled. These precautions do not affect the authentic nature of the data.

The median mileage accumulation rate is around 40 miles per day, and declines slightly over calendar time. This is more than the rate of 33 miles per day, which corresponds roughly to exhausting the 36,000-mile limit in exactly three years. So most cars leave coverage due to

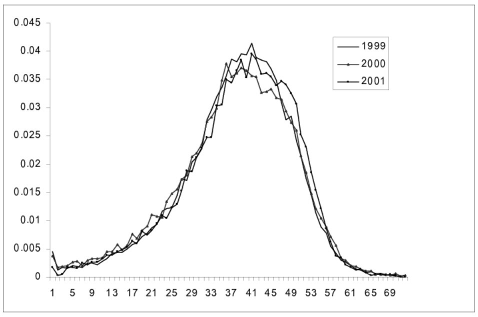


Figure 7: 1999 - 2001 strata distributions

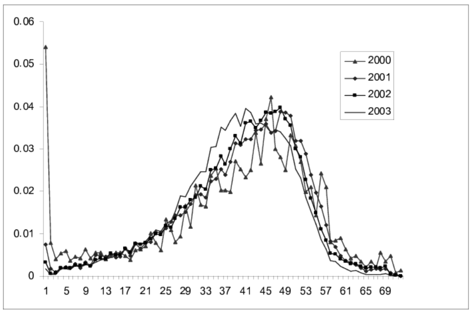
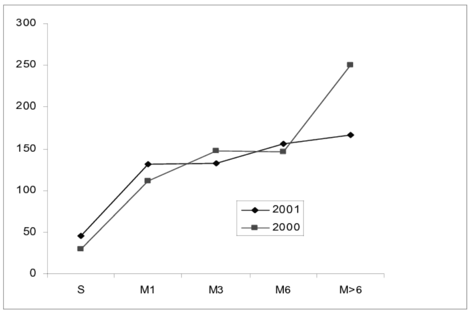
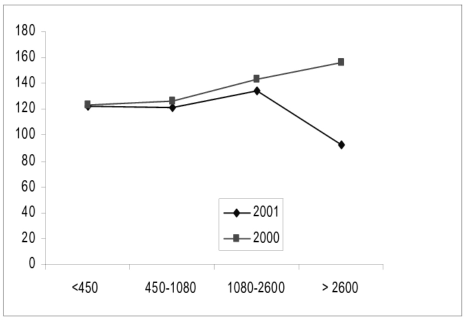


Figure 8: 2001 time cuts strata distributions



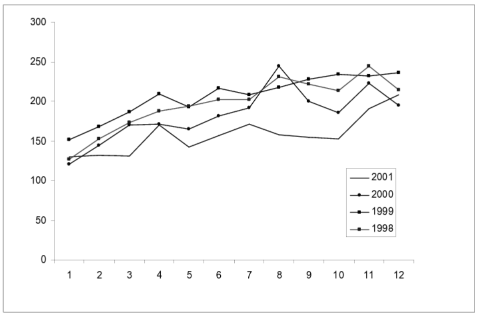
7

Figure 9: Cost per DPG - 4 DPG



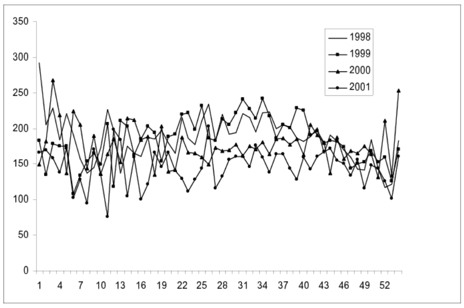
9

Figure 10: Cost per MAR



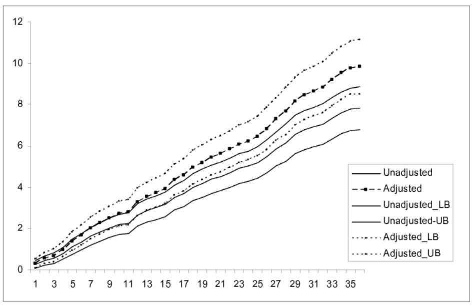
8

Figure 11: Cost per DPG - 12 DPG



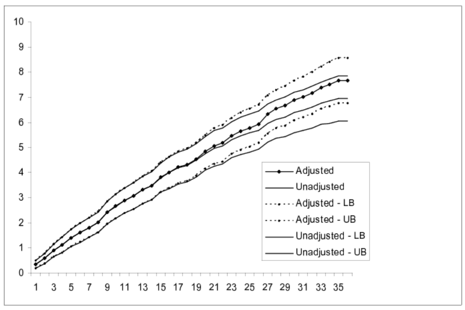
10

Figure 12: Cost per MAR



6

Figure 13: Unadjusted/adjusted $\Lambda(t)$ 95% CL



3

Figure 14: Unadjusted/adjusted $\Lambda(m)$ 95% CL

Descriptive\ Time Statistics \Cuts	Jan. 1, 2001	Jan. 1, 2002	Jan. 1, 2003	Oct. 24, 2003
Number of P - claims	53	597	1206	1639
Cost	14 552	129 527	236 459	291 279
Number of sold vehicles	16 73	44 761	44 879	44 890
Median vehicle age (days)	91	333	697	992

Table 4: Summary of 2001 Warranty File: Time cuts and p-claims

mileage. Also, we should emphasize that in the estimation of the strata distribution we use mileage information from any claim, not just those from System P.

5.2.1 Examples for the "Time" is Age Case

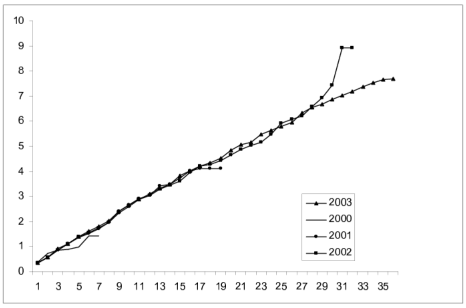
In this section and the next, we illustrate the calculations for cost per car. The results are similar for the number of claims per car. Figure 13 illustrates the effect of the adjustment for withdrawals from coverage due to mileage, shown for the most recent time cut. The adjusted curve is significantly higher than the unadjusted one. We have provided 95% confidence limits on the adjusted and unadjusted $\Lambda(t)$'s, roughly indicating that the differences due to the mileage adjustment is statistically significant at 95% significance level. This result is consistent with the findings in Chukova and Robinson [2] based on the same data.

Figure 15 shows the adjusted curves for all four time cuts to illustrate how the results would unfold over calendar time and Figure 16 provides similar results for the unadjusted curves.

5.2.2 Examples for the "Time" is Miles Case

Figure 14 shows the cumulative cost per car by mileage for the unadjusted case (with 95% confidence limits) along with the adjustment for withdrawal from coverage due to age. In "time" is mileage case, the adjustment is not statistically significant at 95% significance level. Although, the two curves differ much more than under Chukova and Robinson [2] model.

We have observed that the effect of the adjustment for mileage in "time" is age case and the adjustment for age in "time" is mileage case on the curves is different. The reason is that in our dataset relatively few vehicles, only about 33%, leave coverage due to age. Also, the age limit adjustment does not begin until the oldest car reaches the age limit of warranty coverage, which is three years from the first sale for our dataset. This is in contrast to the mileage adjustment for the "time" is age case, where the adjustment begins to have an effect when the first car might possibly reach the mileage limit. In the Chukova and Robinson [2] model the estimation of the percentage of vehicles leaving the coverage due to age is 38%.



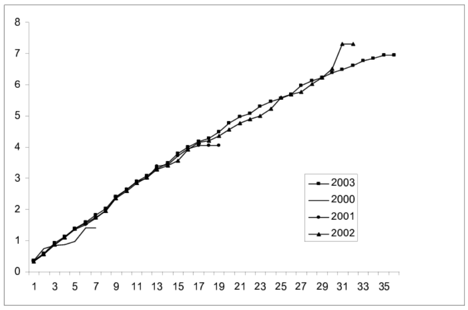


Figure 15: Adjusted $\Lambda(t)$ by timecuts   Figure 16: Unadjusted $\Lambda(t)$ by timecuts

6 Discussion

We have proposed an extension to the Chukova and Robinson [2] and the Hu and Lawless [3] "robust" estimator for the mean cumulative function and its associated standard error. Without requiring any supplemental source of data for mileage accumulation, we deal with the problem of incomplete information for mileages as it typically occurs in automotive warranty data. We also utilize all available information and deal with the issue of changing mileage accumulation rates throughout a vehicle's life. As in previous studies, the adjustments that we propose are with respect to the number of cars at risk.

Also, in Chukova and Robinson's [2] model the linearity assumption imposed on the mileage accumulation rate did not allow the variability of vehicle driving patterns to be taken into account. The stratification approach in modelling the mileage accumulation allows us to account for this variability and led to an interesting observation, namely that a higher variability in the driving pattern may lead to a higher warranty cost. Of course, this observation is based on a single dataset and requires more study.

Other future work could include a model to allow for dynamic cost estimation and prediction of the mean cumulative number of claims or cost of claims and its standard error as functions of the usage measure. Also, with the availability of appropriate data, a full bivarite treatment of this problem would be possible.

References

[1] W. R. Blischke and D. N. P. Murthy, *Product Warranty Handbook* (Marcel Dekker, New York, 1996).

[2] S. Chukova and J. Robinson, in *Modern Statistical and Mathematical Methods in reliability*, Ed. A. Wilson, N. Limnios, S. Keller-McNulty, Y.Armijo, World Scientific, Singapore, 2005, p. 121- 135

[3] X. Hu and J. Lawless, *Biometrika*, **83**, 747 (1996).

[4] X. Hu and J. Lawless, *J. Amer. Statist. Assoc.*, **91**, 300 (1996).

[5] J. Kalbfleisch, J. Lawless and J. Robinson, *Technometrics*, **33**, 273 (1991).

[6] J. Lawless, *Int. Stat. Rev.* **66**, 41 (1998).

[7] J. Lawless, X. Hu and J. Cao, *Lifetime Data Analysis* **1**, 227 (1995).

[8] J. Lawless and J. Nadeau, *Techometrics* **37**, 158 (1995).

[9] W. Nelson, *Reccurent Events Data Analysis for Products Repairs, Disease Reccurences, and Other Applications* (ASA-SIAM, Philadelphia, 2003).

[10] J. Robinson, in *Recent Advances in Life-Testing and Reliability* , Ed. N. Balakrishnan (CRC Press, Boca Raton, London, Tokyo, 1995), p. 195.

[11] J. Robinson and G. McDonald, in *Data Quality Control: Theory and Pragmatics*, Ed. G. E. Liepins and V. R. R. Uppuluri (Marcel Dekker, New York, 1991), p. 69.