MICHIGAN BUSINESS SCHOOL

Product Number wp03-20
Released July, 2003

From the Department of
OPERATIONS AND MANAGEMENT SCIENCE

WORKING PAPER SERIES

A Non-Homogeneous Poisson Process Predictive Model for Automobile Warranty Claims

by

Karl D. Majeske

LECTURER OF OPERATIONS AND MANAGEMENT SCIENCE

Automobile warranties and thus lifetimes are characterized in the two-dimensional space of time and mileage. This paper presents a non-homogenous Poisson Process (NHPP) predictive model for automobile warranty claims consisting of two components: a population size function and a failure or warranty claim rate. The population size function tracks the population in the time domain and accounts for mileage by removing vehicles from the population when they exceed the warranty mileage limitation. The model uses the intensity function of a NHPP—the instantaneous probability of failure—to model the occurrence of warranty claims. The approach was developed to support automobile manufacturers' process of using claims observed during the early portion (first seven months) of vehicle life to predict claims for the remainder of coverage, typically between three and five years. This paper uses manufacturer provided warranty data from a luxury car to demonstrate the NHPP model by predicting claims for three vehicle subsystems. Warranty predictions are then compared with the actual observed values.

A Non-Homogeneous Poisson Process Predictive Model for Automobile Warranty Claims

Karl D. Majeske
The University of Michigan Business School
701 Tappan Street
Ann Arbor, MI 48109-1234
Phone: 734-647-6978
E-mail: kdm@umich.edu

ABSTRACT

Automobile warranties and thus lifetimes are characterized in the two-dimensional space of time and mileage. This paper presents a non-homogenous Poisson Process (NHPP) predictive model for automobile warranty claims consisting of two components: a population size function and a failure or warranty claim rate. The population size function tracks the population in the time domain and accounts for mileage by removing vehicles from the population when they exceed the warranty mileage limitation. The model uses the intensity function of a NHPP – the instantaneous probability of failure - to model the occurrence of warranty claims. The approach was developed to support automobile manufacturers' process of using claims observed during the early portion (first seven months) of vehicle life to predict claims for the remainder of coverage, typically between three and five years. This paper uses manufacturer provided warranty data from a luxury car to demonstrate the NHPP model by predicting claims for three vehicle subsystems. Warranty predictions are then compared with the actual observed values.

Keywords: Automobile warranty data, Two-dimensional lifetime, Non-homogeneous poisson process, Warranty claim predictions.

1. Introduction

Manufacturers of consumer durable goods use many metrics to assess the quality and reliability of their products. One class of measures, referred to as field performance, represent how well a product has functioned during its life relative to customer expectations. Global automobile manufacturers track customer satisfaction with the *Initial Quality Study*[1], a third party measure of vehicle field performance. This survey primarily identifies items that annoy customers or vehicle features that fail to meet their expectations, and manufacturers use this information as an input into future designs. Warranty claims also measure the field performance of automobiles. Items repaired under warranty cause not only an inconvenience to the customer, but result in a direct deduction to the manufacturer's profit generated by the sale. These claims then represent an opportunity for cost reduction that manufacturers can glean through product and process design changes. To quantify the impact on customer satisfaction and cost of warranty, manufacturers use statistical techniques to model and predict the number and cost of warranty claims. Manufacturers use warranty predictions for a variety of purposes: to assess customer satisfaction, identify product and process design changes, and predict financial liability.

This research focused on predicting the number of automobile warranty claims, which automobile manufacturers use as a measure of quality during the various stages of the product development life cycle. After completing product design, automobile manufactures make pre-production prototype vehicles. These vehicles are used in crash testing (to satisfy safety standards) and reliability testing (to establish field performance). Manufacturers also make warranty predictions based on the performance of these pre-production prototype vehicles. They use these predictions to motivate the need for, and justify the expense of, product design changes prior to the commencement of mass production. These early warranty predictions are also an

[1] *The Initial Quality Study,* an across manufacturer comparative study of new vehicle quality, is published annually by J. D. Power and Associates, Agoura Hills, CA.

input into product pricing and the viability of the vehicle in the marketplace. Sarawgi and Kurtz (1995) propose a method for predicting warranty claims using bench test data. Majeske and Herrin (1998) show an example where field performance dramatically differs from predictions using bench test data.

Once a vehicle model reaches the marketplace, manufacturers predict warranty claims and costs based on data observed early in the product life (the first six or seven months of usage). Financial functions use these predictions to determine the amount of money to reserve to cover the future warranty liability. Engineering functions use these predictions for many purposes: identifying product design changes, identifying manufacturing or assembly process changes, cost justifying various changes, and evaluating the effectiveness of these changes once implemented. Wu and Meeker (2002) propose a method for early detection of reliability problems using warranty data. They argue the earlier problems are detected, the smaller the financial and goodwill costs that will result from the problems.

Current trends in automobile manufacturing and marketing have increased basic warranty coverage from 12 months / 12,000 miles to as much as five years / 50,000 miles on the complete vehicle and ten years / 100,000 miles on the power-train. This increase in coverage has highlighted the lack of predictive validity of some automobile warranty forecasting techniques. The motivation for this research was the need to develop a technique that provided the manufacturer more accurate warranty predictions, especially in the later stages of warranty coverage. This paper contains the result of that work, a non-homogeneous Poisson Process (NHPP) predictive model for automobile warranty claims. This technique forecasts the number of warranty claims that a population of vehicles will experience based on the claims observed early in the vehicle lifetime. In addition to providing manufacturers a prediction of the total number of claims, the technique also forecasts the timing of claims during the vehicle lifetime.

The remainder of this paper is organized as follows. Section 2 defines the warranty lifetime of an automobile in the two-dimensional space defined by time and mileage, and introduces the usage function. Section 3 provides an overview of warranty claim predictive models with a focus on automobile warranty. Section 4 develops the NHPP predictive model that consists of two components. The first component is a time based population size that accounts for mileage by censoring vehicles from the population when they exceed the warranty mileage limitation. The second component, the failure rate or the warranty claim rate, is captured by the intensity function of a NHPP. The model then predicts claims by integrating the product of the population size and the intensity function over the time domain of the warranty coverage. Section 5 presents an application of the approach to three subsystems of a luxury automobile using manufacturer supplied warranty data. Section 6 contains the conclusion.

2. Automobile Lifetime Under a Two-Dimensional Warranty

Manufacturers provide product warranties that can be described in a multitude of ways (Blischke and Murthy 1992). Basic coverage describes the warranty included in the purchase price of a product. This coverage represents a service contract that a customer must buy if they wish to purchase the product. Extended coverage represents coverage not included in the purchase price of the product (Kelly and Conant 1991). Moskowitz and Chun (1994) define three major types of warranties: pro rata, free replacement, and lump sum. The pro rata warranty defines a cost sharing function between the manufacturer and seller with the most common setup transferring the repair burden from the manufacturer to the customer linearly over the coverage (e.g., automobile batteries and tires). With free replacement coverage, the manufacturer agrees to replace or repair the product during the coverage at no cost to the customer (e.g., automobiles, computers). The lump sum warranty provides the customer with a fixed payment, determined at time of sale, if the product fails.

Manufacturers express warranty coverage using a quantitative definition of product lifetime. When one variable quantifies product life (e.g., hours of use for a boat motor) the product carries a one-dimensional warranty (Blischke and Murthy 1994). Defining product life in a single metric like "time" has a great deal of intuitive appeal and facilitates warranty data modeling. When two metrics define product life, the product carries a two-dimensional warranty (Blischke and Murthy 1996, Moskowitz and Chun 1994, and Singpurwalla and Wilson 1993). Manufacturers sell automobiles with a two-dimensional basic coverage characterized by time and mileage that offers free replacement. The two-dimensional automobile warranty provides customers a flexible coverage based on their personal usage pattern. The two-dimensional automobile warranty also protects the manufacturer from replacing components on high mileage newer vehicles and limits long-term manufacturer liability on low usage vehicles.

Figure 1 shows the two-dimensional lifetime space for automobiles characterized by time and mileage. The date of final assembly establishes the physical existence of an automobile but the warranty lifetime doesn't begin until a customer purchases the vehicle. Let the random variable L represent the sales lag, the time from final vehicle assembly to vehicle sale. Sales lag is an unknown value at time of assembly; however, all vehicles at risk for warranty claims have an observed value because they have been sold to a customer. Robinson and McDonald (1991) show that the larger the sales lag, the more likely a vehicle will observe warranty claims. The cutoff date is the calendar date of a warranty prediction. Let the random variable T represent time in service – the elapsed time from product sale to the cutoff date – which captures product exposure or the amount of time the vehicle was eligible for warranty claims. Due to the random nature of sales lag, a population of vehicles assembled together will have varying values for time in service at a given cutoff date. This phenomenon of varying values for time in service can complicate the analysis and prediction of automobile warranty claims.

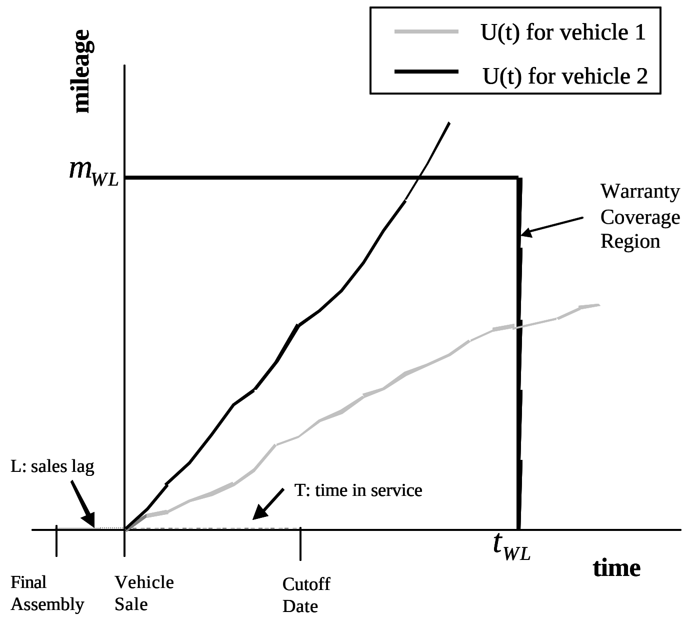

Figure 1: Two Dimensional Automobile Lifetime Space

Figure 1 shows the usage function $U(t)$ - the non-decreasing path through the lifetime space that relates time to mileage - for two hypothetical vehicles. Defining the time limitation and the mileage limitation of the two-dimensional coverage as t_{WL} and m_{WL}, respectively, results in a rectangular coverage region. In this example, vehicle 1 leaves the warranty coverage region by exceeding the time limit while vehicle 2 leaves coverage by exceeding the mileage limit. Because vehicles may leave coverage by exceeding the mileage limit prior to the time limit $U(t) > m_{WL}$ for $t < t_{WL}$, the duration of time that a vehicle will remain under warranty coverage $U^{-1}(m_{WL})$ is an unknown random variable at time of sale. The length of warranty exposure varies within a population depending on the vehicles' usage function $U(t)$, further complicating predictions. Vehicles repaired while inside the basic coverage region, $U(t)$ such that $(\mathrm{t} \leq \mathrm{t}_{\mathrm{WL}} \cap \mathrm{U}(t) \leq \mathrm{m}_{\mathrm{WL}})$, (may) result in warranty claims.

At a cutoff date, each vehicle is at an unknown point on their $U(t)$ function. Automobile manufacturers track the sale of each vehicle and thus know vehicle location in the time domain.

However, due to the uncertain usage patterns of customers, manufacturers do not know a vehicles location in the mileage domain. Because not all vehicles stay at risk for claims through t_{WL}, the remaining coverage – in both time and mileage – is unknown for each vehicle when making a prediction. When vehicles have experienced one or more warranty claims prior to the cutoff date, the manufacturer has an observed mileage corresponding to the time of the repair(s). For vehicles without warranty claims, the manufacturer has no information regarding the mileage domain of the lifetime path. Due to this limited information on mileage, many automobile manufacturers make warranty predictions purely in the time domain.

3. Warranty Claim Predictive Models

Murthy and Djamaludin (2002) provide a comprehensive review of literature related to the various aspects of new product warranty. This section presents some specific statistical models applied to warranty claims, with a focus on automobile warranty claims. This section is divided into three sub-sections based on the type of statistical model used: Poisson or discrete models, stochastic process models, and repair per thousand vehicle models.

3.1 Poisson Models

Poisson prediction models have intuitive appeal to many reliability engineers. These models assume X, the number of claims for a product, follows a Poisson distribution with probability mass function

$$p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \qquad x = 0,1,2,\ldots$$

For a one-dimensional warranty, reliability engineers allow the Poisson parameter lambda λ to be a function of time in service t

$$p(x) = e^{-\lambda t} \frac{(\lambda t)^x}{x!} \qquad x = 0,1,2,\ldots \qquad t > 0. \tag{1}$$

Equation (1) allows fitting a Poisson distribution to a population with varying values of time in service. Assuming l is independent of t, you can predict the expected claims through the warranty time limit t_{WL} for one vehicle as $l t_{WL}$, and for a population of N independent and identically distributed vehicles as $N l t_{WL}$.

Kalbfleisch, Lawless, and Robinson (1991) develop a Poisson model for predicting automobile warranty claims in the time domain similar to Equation (1) above. They note that their Poisson approach yields biased estimates due to vehicles leaving the warranty coverage region by exceeding the mileage limit. Moskowitz and Chun (1994) suggest a Poisson model for two-dimensional warranty coverage. They formulate the warranty problem as a Poisson regression by fitting the cumulative Poisson parameter m with various functions of time t and mileage m. One can easily estimate the parameters of the linear model

$$m = b_1 t + b_2 m$$

using ordinary least squares. Chen and Papova (2002) fit the bi-variate Poisson model to two-dimensional warranty data with maximum likelihood estimation. In application, the lack of data on the mileage domain limits the ability to use this bi-variate approach to model and predict automobile warranty data.

3.2 Stochastic Process Models

Stochastic processes model a product or system that generates a series of (time) ordered observations. Applications of stochastic processes to field performance data appear in the literature as repairable systems reliability. The Poisson Process (Kingman 1993) assumes times between failures are independent and identically distributed (i.i.d.) exponential random variables. A renewal process relaxes the exponential assumption for time between failures but still requires theses times to be i.i.d. The non-homogeneous Poisson Process (NHPP) does not require i.i.d. time between failures and provides a great deal of flexibility when fitting observational data. The NHPP is characterized by the intensity function

$$n(t) = \lim_{\Delta t \to 0} \frac{\Pr(\text{observe a failure in interval } t, t + \Delta t)}{\Delta t} \tag{2}$$

that represents the instantaneous probability of a failure. The expected number of failures (warranty claims) for the NHPP is the integral of the intensity function. For a one-dimensional warranty, the expected number of warranty claims through the warranty limit t_{WL} is

$$E[C(t_{WL})] = \int_0^{t_{WL}} n(t)dt . \tag{3}$$

Nelson (1988) suggested a graphical technique for evaluating either the total cost or number of repairs for a population of products. This technique plots the mean cumulative cost $mcc(t)$ or mean cumulative repair $mcr(t)$ versus time t. Nelson noted the derivatives of both $mcc(t)$ and $mcr(t)$ become $n(t)$, the intensity function of a NHPP. Wang, Suzuki, and Yamamoto (2002) fit a NHPP to warranty data for products with an unknown sales date or products that don't observe a value for sales lag.

The power law process (Rigdon and Basu 1989a), a NHPP fit by Roberts and Mann (1993) to repairable system data, has the intensity function

$$n(t) = (at)^b . \tag{4}$$

Crow (1974) showed that the power law process has the property that T_1, the time to first failure, follows a Weibull distribution (Blischke and Murthy 2000, Meeker and Escobar 1998) with the cumulative distribution function

$$F(t) = 1 - e^{-(at)^b} \qquad t \geq 0 . \tag{5}$$

Authors have referred to the power law process as the Weibull Process (Crow 1982); however, only the time to first failure follows a Weibull distribution. Huang (2001) presents a Bayesian approach for fitting a power law process to repairable system data. Muralidharan (2001) extends the power law process to allow the failure or repair affect the failure rate of the product.

3.3 Repairs per Thousand Vehicles

The most common automobile warranty-data reporting scheme uses cumulative warranty claims per thousand vehicles. While the notation differs between manufacturers, this approach appears to be the industry standard. Majeske, Riches, and Annadi (2003) suggest the following model for $R(t)$ - repairs (warranty claims) per thousand vehicles. Let N_i represent the number of vehicles in the population with at least *i* months in service (MIS). Notice that N_i can decrease as *i* increases due to varying values for sales lag and vehicles exceeding the mileage limit. Let f_i represent the observed number or frequency of claims in time period *i* – manufacturers generally track claims by month - for the N_i vehicles. Then calculate

$$R(t) = \sum_{i=0}^{t} \frac{f_i}{N_i} * 1000 .$$

One method used to predict $R(t)$ values utilizes transformed data. By regressing $\log[R(t)]$ on $\log(t)$, one can fit the simple regression model

$$\log(R(t)) = b_0 + b_1 \log(t) + e_t . \qquad (6)$$

To predict a future $R(t)$ value, denoted $\hat{R}(t)$, simply extend the fit line and transform back to the original units of $R(t)$ using

$$\hat{R}(t) = \exp(\hat{b}_0 + \hat{b}_1 \log(t)) . \qquad (7)$$

Wasserman (1992) developed a dynamic linear predictive model for $R(t)$ using data from multiple product years of the same model vehicle. He shows that incorporating data from previous model years can increase the accuracy of predictions. Singpurwalla and Wilson (1993) developed a bivariate failure model for automobile warranty data indexed by time t and mileage m. They derived the marginal failure distributions and presented a method for predicting $R(t)$ using a log-log model similar to Equation (6).

4. Automobile Warranty Predictive Model

This section presents a non-homogenous Poisson Process (NHPP) predictive model for automobile warranty claims. The approach presented here makes predictions in the time domain but accounts for mileage by removing vehicles from the population when they exceed the mileage limit. The technique models the warranty claims for an automobile with a NHPP that provides a great deal of flexibility when fitting observational data.

4.1 Service Time based Population Size

Automobile manufacturers form vehicle populations based on assembly date. This allows relating warranty claims and costs to time based events such as product design changes, component manufacturing process changes, assembly process changes and vendor issues. Given the variation in sales lag and the random nature of when vehicles leave warranty coverage, the population size varies as it moves through calendar time. Specifically, the population size would start at zero, and increase by one each time a vehicle is sold, and decrease by one each time a vehicle leaves warranty coverage. To place vehicles in a common metric for warranty predictions, automobile manufacturers characterize lifetime using time in service. As previously shown in Figure 1, vehicle sale begins the warranty coverage and defines the origin for time in service. Due to the random nature of sales lag, a population of vehicles assembled on the same day will have different time in service values when making a prediction.

Let $N(t)$ represent the population size or the number of vehicles at risk for warranty claims at time in service t. For a one-dimensional coverage in the time domain, the manufacturer would have complete information on $N(t)$. When making a prediction of total claims for a population of size N,

$$N(t) = \begin{cases} N & t \le t_{WL} \\ 0 & t > t_{WL} \end{cases}. \tag{8}$$

When performing warranty data analysis, many automobile manufacturers use a population definition similar to Equation (8) due to the limited information on vehicle mileage. Using Equation (8) fails to remove vehicles from the population when they exceed the mileage limit thus over estimating the population size at higher time in service values.

Define u as the rate a vehicle accumulates mileage, which is the derivative or slope of the usage function $U(t)$. Assuming a constant usage, a vehicle is no longer in the warranty coverage region if the mileage exceeds the mileage limit or

$$ut > m_{WL}. \tag{9}$$

By letting u_i represent the (constant) usage rate of the ith vehicle, the variable

$$d_i = \begin{cases} 1 & \text{if } u_i t \le m_{WL} \\ 0 & \text{else} \end{cases} \tag{10}$$

indicates if a vehicle remains in the population; that is, Equation (10) indicates vehicles still in the warranty coverage region at time t. To determine the population size at time t, sum the indicator variables of Equation (10) to obtain

$$N(t,m) = \begin{cases} \sum_{i=1}^{N} d_i & t \le t_{WL} \\ 0 & t > t_{WL} \end{cases}. \tag{11}$$

Equations (10) and (11) define a time in service based population size that removes vehicles when they exceed the mileage limit. $N(t,m)$ characterizes the population size in the time domain while taking into account the effects of mileage on warranty coverage.

Automobile manufacturers do not have complete information on the population size $N(t,m)$ as they do for $N(t)$, due to the lack of information on vehicle usage. To use $N(t,m)$, the manufacturer must estimate usage for each vehicle in the population. Vehicles with a warranty claim have an observed value on the lifetime path that can be used to estimate u. To estimate usage for vehicles without a claim, identify a parametric model for u - such as a lognormal or gamma distribution – and estimate the distribution parameters using the observed data. Then estimate usage values for vehicles without claims according to the probability distribution fit to the sample data. Once each vehicle has a usage value (either observed or estimated) you can determine $N(t,m)$ using Equations (10) and (11).

4.2 Time Based Failure Rate

The second component of the predictive technique models the warranty claims associated with a vehicle. A non-homogeneous poisson process (NHPP) provides a flexible model to capture the varied warranty claim patterns observed in automobile warranty data. A NHPP is characterized by the intensity function (instantaneous probability of a failure) of Equation (2). To use the predictive approach in this paper, one must identify an intensity function that captures the time-based pattern of warranty claims. Once a manufacturer has identified the functional form of the intensity function, they must estimate the parameters. Rigdon (1989b) estimated intensity function parameters with multiple observations from a single system and developed confidence intervals on parameters. Hossain and Dahiya (1993) develop a NHPP model for software reliability with a data set that represented one system with 34 failures. These two techniques provide methods for fitting a NHPP to a system with many failures. Automobile manufacturers make predictions early in the vehicle life. At six or seven months in service, a relatively small percentage of vehicles will have experienced multiple warranty claims. This limits the ability to use the above techniques to directly fit a NHPP to warranty data.

Thompson (1981) shows that for any NHPP, the intensity function and the hazard function of T_1, the time to first failure, have the same functional form

$$n(t) = h(t_1)\,. \tag{12}$$

This provides an alternate approach to characterizing the intensity function. First, estimate the distribution parameters for time to first warranty claim (failure). Then use these coefficients to characterize a NHPP intensity function. Majeske (2003) describes how to use maximum likelihood estimation to fit the right-censored data encountered in warranty data analysis. By including non-failed vehicles as censored observations, this incorporates vehicles without a warranty claim when fitting the stochastic process. In some applications, manufacturers use a parametric model for time to failure, which they could use as a starting point for modeling warranty claims. To identify a parametric model for time to first warranty claim, one could use the hazard plot method suggested by Majeske, Lynch-Caris, and Herrin (1997).

4.3 Predicting Warranty Claims with the Model

Let the function $C(t)$ represent the cumulative number of warranty claims observed by the population through t time in service. Using this definition, $C(t)$ is a strictly increasing step function that starts at 0. Let $\hat{C}(t)$ represent the predicted number of claims through time t as the expected value of the cumulative warranty claim function

$$\hat{C}(t) = E[C(t)].$$

Modeling a single vehicle with a NHPP, you can predict claims by integrating the intensity function as shown in Equation (3). To predict claims for a population assume homogeneity; that is, all vehicles in the population have the same intensity function or failure rate.

For a one-dimensional warranty coverage - use the population size function $N(t)$ of Equation (8) - the expected number of claims is the product of the population size and the integral of the intensity function

$$\hat{C}(t_{WL}) = E[C(t_{WL})] = N \int_0^{t_{WL}} n(t)dt \quad . \tag{13}$$

For the two-dimensional warranty coverage of automobiles, one can predict the expected number of warranty claims by integrating the product of $N(t,m)$ - the population size function of Equation (11) - and the intensity function

$$\hat{C}(t_{WL}) = E[C(t_{WL})] = \int_0^{t_{WL}} N(t,m)n(t)dt \ . \qquad (14)$$

In some cases, manufacturers may want to relax the homogeneity assumption. Allowing the model parameters to be a function of $\vec{X}$ (covariates such as vehicle option content) would allow the manufacturer to make inference on how these variables affect warranty claim rates. Including covariates makes predicting claims somewhat more difficult because vehicles are at risk for different lengths of time. Letting t_i represent the time the ith vehicle is at risk for warranty claims, the expected number of claims for the population of vehicles is

$$\hat{C}(t_{WL}) = E[C(t_{WL})] = \sum_{i=1}^{N} \int_0^{t_i} v(t, \vec{X}_i)dt \ . \qquad (15)$$

The complicating factor with this approach is that the t_i values depend on the vehicle mileage, a random variable with limited or no information.

This research suggests predicting warranty claims by vehicle subsystem (e.g., powertrain, wiring, fit and finish, etc.) and then aggregating subsystem predictions to obtain a prediction for the population of vehicles. To account for differences in option content (e.g., V6 versus V8 engine) the manufacturer can stratify the population for each subsystem and make a prediction with Equation (14) for each strata within each subsystem. Then, aggregate the predictions across strata (option content differences) and subsystem (portions of the vehicle) to arrive at a prediction for the population of vehicles.

5. Application of Predictive Technique

This section assesses the predictive validity of the NHPP using a set of 9,168 luxury cars. To emulate the manufacturer's prediction process, this section makes predictions using the data available to the manufacturer at a cutoff date approximately seven months after vehicle assembly, termed the current date. Using the model developed in the previous section, warranty claim predictions are made through a cutoff approximately 45 months after assembly, called the future cutoff. Because the vehicles have already observed the claims through the future cutoff, comparing predictions with actual claims allows assessing the predictive technique.

5.1 Estimating the Population

To estimate the population function $N(t, m)$ through the future cutoff date, a usage value u was calculated for each vehicle with a warranty claim prior to the current cutoff date. Using maximum likelihood estimation, a two parameter gamma distribution, with probability density function

$$f(r) = \frac{1}{\Gamma(a)} r^{a-1} e^{-r/g} \qquad r > 0, \tag{16}$$

was fit to the data yielding the parameter estimates $\hat{a} = 3.548$ and $\hat{g} = 16.629$. Figure 2 shows a histogram of the observed values and the gamma distribution. To generate usage values for vehicles without a claim, assume that the vehicles with an observed usage value (56% of the population) are representative of the entire population. Usage rates were then simulated for vehicles without an observed value from a gamma distribution with $a = 3.548$ and $g = 16.629$ to match the observed data. Table 1 contains the $N(t, m)$ values through the future cutoff date calculated using Equations (10) and (11). Table 1 also contains $N(t)$, the population size ignoring the effect of mileage. Figure 3 plots the data of Table 1 to provide a graphical comparison of the mileage censored population size $N(t, m)$ with the time only population size $N(t)$.

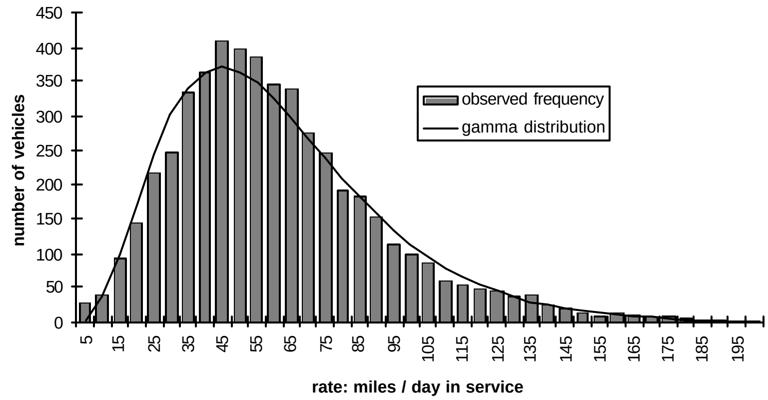

Figure 2: Histogram of observed usage rates and best fit Gamma distribution

Table 1: Population size by time in service (months)

Time in Service (Months)	*N(t,m)*	*N(t)*	Time in Service (Months)	*N(t,m)*	*N(t)*	Time in Service (Months)	*N(t,m)*	*N(t)*
0	9168	9168	16	8480	9168	32	4561	9168
1	9168	9168	17	8296	9168	33	4328	9168
2	9168	9168	18	8082	9168	34	4098	9168
3	9168	9168	19	7875	9168	35	3903	9168
4	9168	9168	20	7620	9168	36	3705	9168
5	9168	9168	21	7360	9168	37	3538	9168
6	9166	9168	22	7131	9168	38	3340	9168
7	9166	9168	23	6849	9168	39	3173	9168
8	9158	9168	24	6601	9168	40	2979	9052
9	9141	9168	25	6341	9168	41	2810	8895
10	9109	9168	26	6059	9168	42	2586	8691
11	9061	9168	27	5794	9168	43	2383	8450
12	8992	9168	28	5523	9168	44	2156	8165
13	8881	9168	29	5290	9168	45	1963	7813
14	8754	9168	30	5054	9168			
15	8620	9168	31	4806	9168			

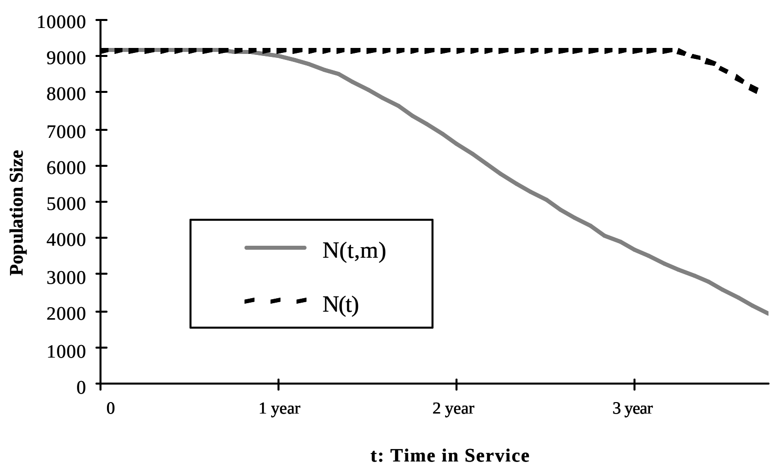

Figure 3: Estimated population size with and without adjusting for mileage

5.2 Estimating the NHPP Intensity Function

This section characterizes the NHPP intensity function for three subsystems of the luxury cars. First, a functional form is identified for time to first claim using empirical hazard plots. The parameters are then estimated for each subsystem. A figure containing the empirical hazard function $\hat{h}(t)$ - calculated using life tables with an interval width of seven days (Lawless 1982) - and the fit line graphically shows the aptness of the model.

Subsystem 1:

Majeske (2003) derives the Weibull-uniform mixture model with cumulative distribution function

$$F(t) = \begin{cases} p * \left(\dfrac{t}{A}\right) + q * \left(1 - e^{-(at)^b}\right) & 0 \le t < A \\ p + q * \left(1 - e^{-(at)^b}\right) & t \ge A \end{cases} \tag{17}$$

to fit warranty claims from subsystem 1 of these same luxury cars. The distribution was developed to model warranty claims as a mixture of manufacturing defects and usage related failures using the data available about two years after vehicle assembly. The mixture distribution of Equation (17) defines a NHPP with intensity function

$$n(t) = \begin{cases} \dfrac{\dfrac{p(1-q)}{T^*} + (1-p)[a\,b\,(at)^{b-1} e^{-(at)^b}}{1 - p\left(q + \dfrac{(1-q)t}{T^*}\right) - (1-p)\left(1 - e^{-(at)^b}\right)} & 0 \le t < T^* \\ a\,b(a\mathrm{t})^{b\text{-}1} & t \ge T^* \end{cases} \tag{18}$$

To apply the prediction method, the mixture model was fit to the data available seven months after vehicle assembly. Table 2 contains the mixture distribution parameter estimates obtained via maximum likelihood estimation and the standard errors estimated with the observed Fisher information matrix as outlined in Majeske (2003). Figure 4 shows the empirical hazard function

$\hat{h}(t)$ and the mixture hazard function using the parameter estimates from Table 2. The mixture model appears to provide a good fit to the data available at 7 months in service.

Table 2: Parameter Estimates for Time to First Warranty Claim.

Subsystem 1 Mixture Distribution			Subsystem 2 Weibull Distribution			Subsystem 3 Linear Hazard		
Parameter	Estimate	Std Error	Parameter	Estimate	Std Error	Parameter	Estimate	Std Error
a	0.00027	0.000021	a	0.00324	0.00006	b_0	0.0001558	0.0000379
b	1.006	0.0339	b	1.057	0.0154	b_1	0.0000174	0.0000022
p	0.047	0.0048						
q	0.232	0.0295						

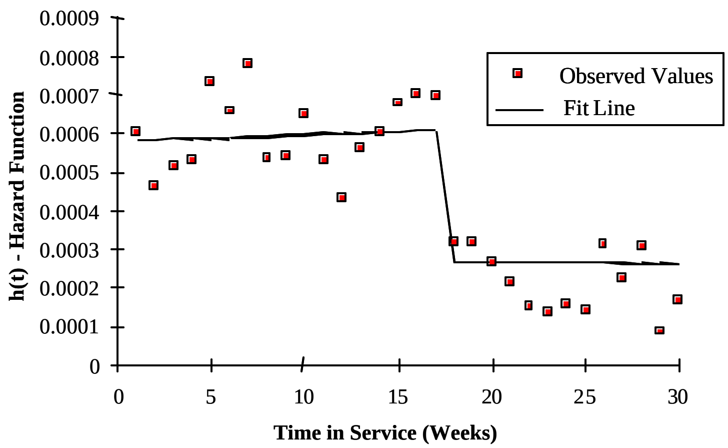

Figure 4: Subsystem 1 – Empirical and mixture model hazard functions

Subsystem 2:

Reliability engineers often use the Weibull distribution of Equation (5) to model field failure and bench test data for subsystem 2. Modeling time to first failure with a Weibull distribution results in a special case of the NHPP, the power law process, with intensity function of Equation (4). The Weibull distribution was fit to the time to first subsystem 2 claim data available at the current cutoff. The parameter estimates and associated standard errors appear in

Table 2. Figure 5 plots the Weibull hazard with the observed values for subsystem 2. Notice that the Weibull does not capture the apparent decreasing hazard in weeks 15 through 30.

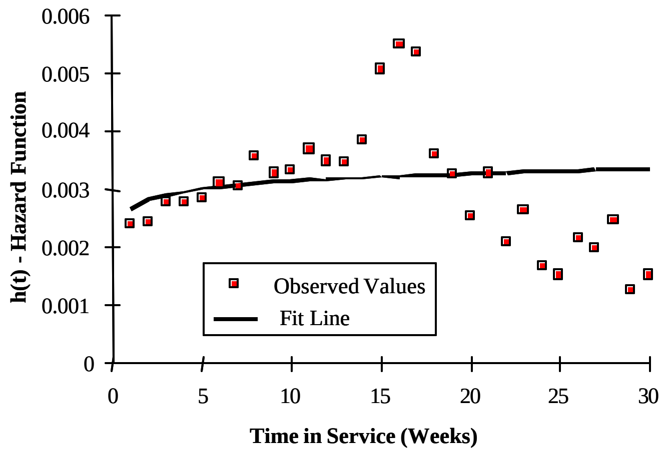

Figure 5: Subsystem 2 – Empirical and Weibull hazard functions

Subsystem 3:

The manufacturer did not use a parametric model for subsystem 3 claims; rather, they used the log-log $R(t)$ approach of Equations (6) and (7). After evaluating the empirical hazard plot shown in Figure 6, the linear hazard function

$$h(t) = b_0 + b_1 t \qquad (19)$$

was identified for the Subsystem 3 data. The linear hazard of Equation (19) defines a NHPP with intensity function

$$n(t) = b_0 + b_1 t\,. \qquad (20)$$

The parameters of the linear hazard model were estimated using ordinary least squares by fitting a line to the empirical hazard function. The parameter estimates and associated standard errors appear in Table 2. Figure 6 shows the empirical hazard along with the linear hazard, which appears to provide a good fit to the data.

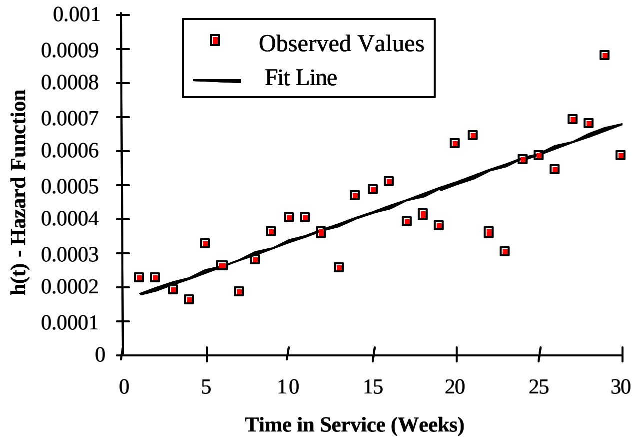

Figure 6: Subsystem 3 – Empirical and linear hazard functions

5.3 Predicting Warranty Claims

This section predicts warranty claims through 45 months in service using the population size $N(t,m)$ and NHPP intensify functions developed in sections 5.1 and 5.2 respectively. Because months in service don't correspond to calendar time, predictions were made for 30 day (month) intervals. To make predictions, Equation (14) was evaluated for each monthly interval. Table 3 contains the observed claims by month at the current cutoff, the predicted claims through the future cutoff, and the observed claims at the future cutoff for each of the three subsystems. Comparing the total claims through 45 months, the NHPP approach predicted within 7% of observed claims for all three subsystems.

Figures 7, 8 and 9 graphically depict the data in Table 3 by comparing the observed and predicted claims by month for subsystems 1, 2, and 3 respectively. Evaluating Figure 7, the Weibull portion of the mixture ($t > 4$ MIS) over predicts in the period 9 - 22 MIS and under predicts in the period 28 - 45 MIS. While the model has some correlation structure in the residuals, it does provide a good fit to the subsystem 1 warranty claims. From Figure 8, the power law process under predicts the subsystem 2 infant mortality failures and does not predict the downward trend at 10 MIS. However, the model very accurately predicts claims from 30 to

45 MIS and provides an accurate prediction of total claims. Figure 9 shows the linear intensity model lends a good fit to the subsystem 3 data, yet slightly over predicts claims after 30 MIS.

6. Conclusion

Automobile manufacturers rely on predictions for the number, timing, and cost of future warranty claims. Omitting the mileage data contained on automobile warranty claims, and making predictions purely in the time domain, ignores vehicles leaving coverage prior to the warranty time limit. Incorporating a two-dimensional aspect will result in a predictive model that better represents the process being modeled. This paper develops a NHPP predictive model that has a parametric component (time to first failure) and provides a great deal of flexibility in application. The NHPP approach also allows incorporating past experience when identifying the failure process and bases predictions on early field performance of the products. Using the NHPP will result in more accurate predictions of warranty claims and support decision-making when implementing engineering design and manufacturing process changes. More accurate predictions will also assist the manufacturer in allocating reserves to pay for repairs covered under warranty.

Table 3 Observed and Predicted Warranty Claims by Month in Service

	Subsystem 1			Subsystem 2			Subsystem 3		
Time in Service (Months)	Observed Claims at 6 MIS	Predicted Claims through 45 MIS	Observed Claims at 45 MIS	Observed Claims at 6 MIS	Predicted Claims through 45 MIS	Observed Claims at 45 MIS	Observed Claims at 6 MIS	Predicted Claims through 45 MIS	Observed Claims at 45 MIS
0	81	100	81	414	414	414	90	90	90
1	188	155	189	1010	786	1022	66	53	69
2	199	158	200	1039	844	1080	71	74	79
3	164	160	170	1241	869	1301	103	94	119
4	185	162	190	1576	886	1664	122	115	151
5	93	96	98	1059	899	1143	133	135	159
6	41	73	51	578	909	723	141	156	182
7	50	73	74	469	917	775	138	176	230
8		73	84		924	903		197	264
9		73	64		929	854		217	247
10		73	60		931	881		236	258
11		73	58		932	856		256	291
12		72	57		929	694		274	252
13		71	61		922	587		290	238
14		70	40		913	605		306	263
15		69	43		903	554		320	263
16		68	41		891	522		334	322
17		67	49		875	521		345	307
18		65	32		855	522		355	420
19		63	41		836	468		363	389
20		61	43		811	534		369	391
21		59	45		786	530		372	350
22		58	39		764	529		377	407
23		55	56		735	507		377	393
24		53	42		710	461		378	345
25		51	43		684	437		378	348
26		49	52		655	491		374	329
27		47	38		628	502		371	360
28		45	47		600	501		366	359
29		43	50		576	526		363	371
30		41	46		551	567		358	427
31		39	43		525	618		351	420
32		37	54		499	603		343	395
33		35	62		474	561		335	325
34		33	46		450	522		327	290
35		32	51		429	530		320	290
36		30	43		408	476		312	261
37		29	48		390	436		306	212
38		27	35		369	399		296	166
39		26	33		351	384		289	174
40		24	36		330	351		278	149
41		23	36		312	397		268	181
42		21	40		287	374		253	204
43		19	33		265	388		238	154
44		17	19		240	373		220	150
45		16	21		219	332		205	131
	1001	2786	2784	7386	30415	28418	864	12810	12175

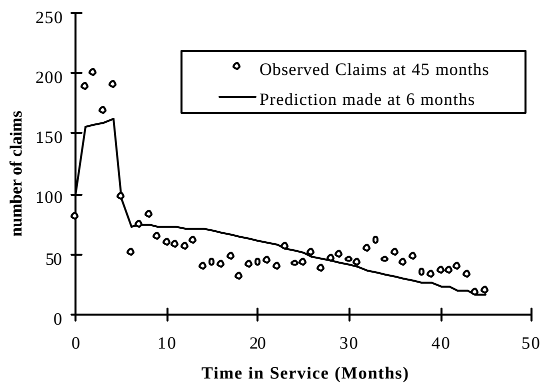

Figure 7: Subsystem 1 – Mixture Model Prediction with Observed Claims

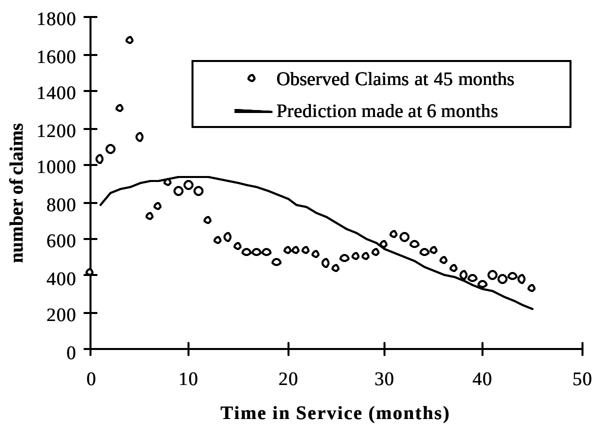

Figure 8: Subsystem 2 – Power Law Process Prediction with Observed Claims

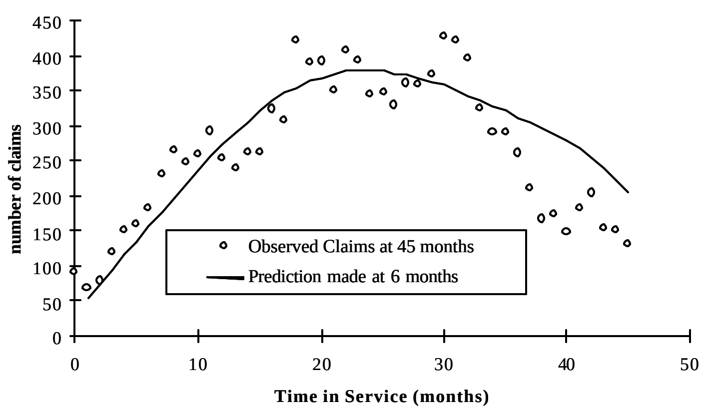

Figure 9: Subsystem 3 – Linear Intensity Function Model Prediction with Observed Claims

References

Blischke, W. R., D. N. P. Murthy. 1992. Product warranty management - I: a taxonomy for warranty policies. *European Journal of Operational Research*, **62** 127-148.

Blischke, W. R., D. N. P. Murthy. 1994. *Warranty Cost Anaysis*. Marcel Dekker, New York.

Blischke, W. R., D. N. P. Murthy. 1996. *Product Warranty Handbook*. Marcel Dekker, New York.

Blischke, W. R., D. N. P. Murthy. 2000. *Reliability Modeling, Prediction and Optimization*, John Wiley and Sons, New York.

Chen, T., E. Popova. 2002. Maintenance policies with two-dimensional warranty. *Reliability Engineering and System Safety* **77** 61-69.

Crow, L. H. 1974. Reliability Analysis for Complex Repairable Systems. *Reliability and Biometry*. SIAM.

Crow, L. H. 1982. Confidence interval procedures for the weibull process with application to reliability growth. *Technometrics* **24** 67-72.

Hossain, S. A., R. C. Dahiya. 1993.Estimating the parameters of a non-homogeneous poisson-process model for software reliability. *IEEE Transactions on Reliability* **42** 604-612.

Huang, Y. 2000. A decision model for deteriorating repairable systems. *IIE Transactions* **33** 479-485.

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

Kelley, C. A., J. S. Conant. 1991. Extended warranties: consumer and manufacturer perceptions. *The Journal of Consumer Affairs* **25** 68-83.

Kingman, J. F. C. 1993. *Poisson Processes*. Clarendon Press, Oxford University Press Inc., New York, New York.

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

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

Majeske, K. D., Herrin G. D. 1998. Determining warranty benefits for automobile design changes. *Proceedings of the Annual Reliability and Maintainability Symposium* 94-99.

Majeske, K. D., M. D. Riches, H. P. Annandi. 2003. Ford's reliability improvement process – a case study on automotive wheel bearings. *Case Studies in Reliability and Maintenance* 545-569. John Wiley and Sons, New York, New York.

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*, John Wiley and Sons, New York.

Moskowitz, H., Y. H. Chun. 1994. A Poisson regression model for two-attribute warranty policies. *Naval Research Logistics* **41** 355-376.

Muralidharan, K. 2001. On testing of parameter in modulated power law process. *Applied Staochastic Models in Business and Industry* **17** 331-343.

Murthy, D.N.P, I Djamaludin. 2002. New product warranty: a literature review. *International Journal of Production Economics* **79** 231-260.

Nelson, W. 1988. Graphical analysis of system repair data. *Journal of Quality Technology* **20** 24-35.

Rigdon, S. E., A. P. Basu. 1989a. The power law process: a model for the reliability of repairable systems. *Journal of Quality Technology* **21** 251-260.

Rigdon, S. E., A. P. Basu. 1989b. Mean squared errors of estimators of the intensity function of a non-homogeneous poisson process. *Statistics & Probability Letters* **8** 445-449.

Roberts, W. T. Jr., L Mann, Jr. 1993. Failure predictions in repairable multi-component systems. *International Journal of Production Economics* **29** 103-110.

Robinson, J. A., G. C. McDonald. 1991. Issues related to field reliability and warranty data. *Data Quality Control: Theory and Pragmatics*. 69-90. Marcel Dekker, New York.

Sarawgi, N., S. K. Kurtz. 1995. A simple method for predicting the cumulative failures of consumer products during the warranty period. *Proceedings Annual Reliability and Maintainability Symposium* 384-390.

Singpurwalla, N. D., S. Wilson. 1993. The warranty problem: its statistical and game theoretic aspects. *SIAM Review* **35** 17-42.

Thompson, W. A. Jr. 1981. On the foundations of reliability. *Technometrics* **23** 1 - 13.

Wang, L., K. Suzuki, W. Yamamoto. 2002. Age-based warranty data analysis without date-specific sales information. *Applied Stochastic Models in Business and Industry* **18** 323-337.

Wasserman, G. S. 1992. An application of dynamic linear models for predicting warranty claims. *Computers and Industrial Engineering* **22** 37-47.

Wu, H., W. Q. Meeker. 2002. Early detection of reliability problems using information from warranty databases. *Technometrics* **44**(2) 120-133.