J SYST SCI SYST ENG
Vol. 28, No. 6, December 2019, pp. 715–730
DOI: https://doi.org/10.1007/s11518-019-5434-8
ISSN: 1004-3756 (paper), 1861-9576 (online)
CN 11-2983/N

Warranty Claims Forecasting for New Products Sold with a Two-Dimensional Warranty

Anshu Dai,[a] Zhaomin Zhang,[b] Pengwen Hou,[a] Jingyi Yue,[a] Shuguang He,[c] Zhen He[c]

[a]Co-Innovation Center for Computable Modeling in Management Science, Tianjin University of Finance and Economics, Tianjin 300222, China
anshudaitju@163.com, pengwen@tjufe.edu.cn, yuejingyi25@163.com

[b]College of Aeronautic Engineering, Civil Aviation University of China, Tianjin 300300, China
zhangzhaozhaomin@tju.edu.cn

[c]College of Management and Economics, Tianjin University, Tianjin 300072, China
shuguanghe@126.com (✉), zhhe@tju.edu.cn

Abstract. Warranty claims forecasting plays an increasingly important role not only for preparing financial plans but also for optimizing warranty policy and improving after-sale services. In the case of new products, an important feature is that the new generation of products often has a close connection with the previous generations of products it replaces. Thus, the warranty claims data of the previous generations of products can be used for extracting reliability information of new products. In this context, we propose a warranty claims forecasting model considering usage rate for new products sold with a two-dimensional warranty. The accelerate failure time model is introduced to investigate the effect of usage rate on product degradation. The non-homogeneous Poisson process is used to model failure counts of repairable products and the constrained maximum likelihood estimation method is used to estimate model parameters. The results of data experiments based on both simulation and real data collected from an automobile manufacturer in China show that the proposed model considering the varying usage rate outperforms the traditional models in forecasting the number of warranty claims.

Keywords: Non-homogeneous poisson process, new products, two-dimensional warranty, usage rate, warranty data analysis

1. Introduction

Great changes have taken place in information and communication technologies with deepening economy globalization. Beyond price reduction, high quality and excellence performance, attractive post-sale service has played a key role in encouraging market demand. In this context, most durable products are sold with a warranty. A warranty is a contractual obligation incurred by a manufacturer (vendor or seller) in connection with the sale of a product (Wu 2012). It requires the manufacturer to remedy any problems or failures that occur during a specified period of warranty coverage (Karim and Suzuki 2005). Warranty policy can be broadly categorized into one-dimension and two-dimensional policies. One-dimensional warranty is usually characterized by an interval that is defined in terms of single variable-for example, age, or usage. Unlike the one-dimensional warranty, a two-dimensional warranty is characterized by a region in two dimensions with the axes representing age and usage, respectively. Such a warranty policy has been successfully used for relatively high-priced products including excavators, automobiles and printers.

Nowadays, an increasing number of manufacturers, especially for complex products manufacturers, regard warranty as an effective tool to boost sales and build brand loyalty (Bian et al. 2015). While a satisfactory warranty term could attract more customers, such a warranty obligation generates additional cost to the manufacturer. According to a survey of "Warranty Week" (http://www.warrantyweek.com/) in 2014, the cost of warranty as a percentage of the sale price can typically vary between 2% and 5% for some big companies in the world. Thus, accurately forecasting warranty claims becomes increasingly important to decisions making on designing warranty policy and preparing financial plans, as well as product development.

There is a vast literature that deals with approaches developed to forecasting warranty claims. Stochastic process has received considerable attention with regard to warranty claims forecasting. In the case of one-dimensional warranty, Kalbfleisch et al. (1991) applied a Poisson process to forecast and analyze warranty claims, where the parameter of the Poisson process is a function of service time. Kaminskiy and Krivtsov (2000) discussed three warranty claim forecasting models with the G-renewal process, the ordinary renewal process and the non-homogeneous Poisson process (NHPP), respectively. Fredette and Lawless (2007) proposed a warranty claims forecasting model using flexible NHPP, where possible heterogeneity among the individuals was modeled using random effects. Akbarov and Wu (2012a) predicted warranty claims based on weighted maximum likelihood estimation (MLE), in which they gave more weight on claims reported in recent months and less weight on claims in earlier months. Akbarov and Wu (2012b) proposed a NHPP model for warranty claims forecasting considering dynamic over-dispersion, where the variance to mean ratio is changing over time. In the case of two-dimensional warranty, Moskowitz and Chun (1994) discussed a Poisson regression method for warranty claims modeling of products, which can be expressed by a regression function of the usage and age for a two-dimensional warranted product. Lawless et al. (1995) developed a method for modeling the occurrence of warranty claims based on both age and mileage, and utilized supplemental information about mileage accumulation to estimate survival rates. Majeske (2007) proposed a NHPP model to predict warranty claims in the time domain but accounted for mileage by removing vehicles from the user population when they reach the mileage limit. Based on warranty data, Baik and Murthy (2008) developed a method to estimate component reliability by using an accelerated failure time (AFT) model considering the effect of usage rate on the failure process of the product. Jack et al. (2009) designed a repair-replace strategy based on usage rate for products sold with a two-dimensional warranty. Lawless JF et al. (2009) proposed models that can be used to assess the dependence on age or usage in heterogeneous populations of products, and developed methods to estimate model parameters based on different types of field data. Yang et al. (2016) employed two generalized renewal process methods to model the imperfect repair, and applied them to forecast the warranty claims for two-dimensional warranted product. Dai et al. (2017) used a stochastic expectation-maximization algorithm to estimate the parameters of reliability model based on two-dimensional warranty data with censoring times, and they adopted the proposed method to predict warranty claims.

Additionally, the lifetime distribution approaches are also used in the field of warranty claims forecasting. Kleyner and Sandborn (2005) developed a warranty prediction model based on a piecewise application of *Weibull* and *Exponential* distributions. Rai (2009) presented a flexible warranty cost forecasting approach considering seasonality, business days per month and sales ramp-up in addition to age and usage.

In term of machine learning, Wu and Akbarov (2011) introduced a weighted support vector regression (SVR) model and a weighted SVR-based time series model to forecast warranty claims. Rai and Singh (2005b) adopted radial basis function network for warranty claims forecasting. With regard to time series models, Wasserman (1992) utilized the $log - logR(t)$ model, first-order autoregression time series models, and the Kalman filter models to predict the number of claims.

However, existing research has paid less attention to new products. New products usually have similar components and production methods with the earlier generation products. Taking the Volkswagen engine EAXXX (specific model is not disclosed) for example, the earlier generation engines have similar mechanical structure with the previous ones, the difference is the length of the crankshaft and piston connecting rod. The length of connecting rod of new engines is shorter than the one of earlier generation engines for increasing the amount of exhaust. Thus, it is feasible to use the warranty claims data of the earlier generation products to get some inference about the reliability of new products. In this context, Wu and Akbarov (2012) developed a two-phase modeling algorithm for warranty claims forecasting for new products. The approach presented in Wu and Akbarov (2012) made predictions in the time domain where the product degradation depends solely on the age of the product. Different from one-dimensional warranted products, products sold with a two-dimensional warranty usually deteriorate with the age and usage, that is to say, the reliability of two-dimensional warranted products are influenced by both the age and usage. Therefore, the existing warranty claims forecasting approach for new products does not apply to two-dimensional warranted new products.

The main contribution of this paper is that we develop a warranty claims forecasting model for new products with a two-dimensional warranty by taking both the age and the usage into consideration, which is an extension of Wu and Akbarov (2012)'s study. Since the complexity of the proposed forecasting model, the CMLE method is adopted to settle parameter estimation. Besides, we validate the proposed forecasting model using artificially generated data and real warranty data collected from an automobile manufacturer in China.

The outline of the paper is as follows. Section 2 introduces problem description. Section 3 defines the warranty claims forecasting model considering usage rate. Section 4 applies the proposed model to both simulation and real data collected from an automobile manufacturer in China, and compares the performance to the traditional models. Finally, Section 5 contains the concluding remarks of this work.

2. Problem Description

In this study, we focus on two-dimensional warranted new products. In the case of two-dimensional warranty, the warranty is characterized by a rectangular region $\Omega \equiv [0, W_0) \times [0, U_0)$ where W_0 is the age limit and U_0 is the usage limit. The warranty ends when the age of the product exceeds W_0 or usage reaches U_0, whichever comes first. This study assumes that all warranty claims are valid, and each failure has an impact on product performance and results in a warranty claim (Wang et al. 2015, Shahanaghi et al. 2013).

The form of warranty data is shown in Table 1. We assume that warranty data are aggregated on a monthly basis, denoted as $1, 2, \cdots, i$ and the related production amount of the products are denoted as $S_1, S_2, \cdots, S_i$. We also assume that there is no lag between the date of production and the date of sale. Therefore, the production time is set as the time origin in the analysis. The number of warranty claims of the products produced in i month and used for j MIS are referred to as $d_{i,j}$. The proposed model is to predict the value of $d_{i,j}$ $(j = 1, 2, \cdots, W_0)$ for a given production month i and stated as Table 1.

Table 1 The Form of Warranty Claims Data

Production month	Production amount	Month in service 1	2	3	$\cdots$	W_0
1	S_1	$d_{1,1}$	$d_{1,2}$	$d_{1,3}$	$\cdots$	d_{1,w_0}
2	S_2	$d_{2,1}$	$d_{2,2}$	$d_{2,3}$	$\cdots$	d_{2,w_0}
3	S_3	$d_{3,1}$	$d_{3,2}$	$d_{3,3}$	$\cdots$	d_{3,w_0}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	S_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	$\cdots$	d_{i,w_0}

3. Warranty Claims Forecasting Model

The NHPP-based forecasting model considering usage rate is presented to predict the expected number of claims within the warranty coverage and the CMLE method is used to estimate model parameters.

3.1 Usage Rate

Usage may be defined in a number of ways, which can be viewed as output produced. For example, the distance traveled is termed as usage for an automobile, while the number of copies made could capture usage of a photocopier. Under the circumstances, usage rate is defined as the output per unit time.

This study assumes the usage rate is constant for each customer over the warranty coverage, but it varies across the customer population (Iskandar et al. 2005, Baik and Murthy 2008, Yang et al. 2016). Under this assumption, the two-dimensional failure process could be effectively reduced to a one-dimensional one. The usage rate can be treated as a random variable following an appropriate distribution. The usage rate is defined as

$$r_i = \frac{u_i}{t_i}, i = 1, 2, 3, \cdots, n \tag{1}$$

where u_i represents usage accumulation of the ith product until t_i and t_i is the corresponding MIS. Denote R as a random variable of the usage rate with lower bound r_{min} and upper bound r_{max}, $G(r)$ and $g(r)$ are its cumulative distribution function and probability density function respectively.

When manufacturing data, failure time data and accumulated usage data are known, the usage rate distribution can be obtained. However, warranty data considered in this study are available in the form of aggregated claims, which poses a major challenge for collecting the customer usage rate information. In order to obtain more comprehensive information concerning usage rate of the products, accumulated usage data are collected from multiple sources including maintenance and customer survey. Usage information can be recorded when customers bring their products for regular maintenance during and after the warranty coverage. This information is for the un-failed product since it is collected during regular maintenance only. With the development of information technology, after-sales service department could contact customers easily and request for returning accumulated usage information. Although implementing customer survey is costly, it contains a great source of information for both failed and un-failed products. In addition, it is worth noting that the time-in-service information can be obtained from the sales records of manufacturers.

Once all the usage and time-in-service information are available, a distribution that fits best to the data can be selected. Rai and Sing (2005a), Shahanaghi et al. (2013), and Limon S et al. (2016) suggested that $Weibull$, $Normal$, $Lognormal$ and $Exponential$ distributions are possible distributions when fitting the usage rate on available historical data.

3.2 Estimate the Number of Products under Warranty

Because products with high usage rate will reach their usage limit U_0 first, the number of products under warranty will varying over time. We take the accumulated usage data of the automobile engines in 2010 for example, which is collected from a specific automobile manufacturer in China. The mileage accumulation for each inspection is plotted as shown in Figure 1. It indicates that those black points that locating in the upper left section of the rectangle will reach usage limit U_0 first because of high usage rates, while those grey points locating in the lower right section will exceed warranty until age limit W_0. Almost 20 percent of these products cease warranty due to reaching usage limit.

We assume that all N products are manufactured as a batch and launched into market at the exact same time. Let $N(t)$ represent the number of products under warranty at t MIS. As it is costly to obtain the usage rate information of all the products, manufacturers usually do not have complete information about $N(t)$. Concerning this issue, parametric models for estimating $N(t)$ are adopted (Rai and Singh 2006, Majeske 2007). Therefore, the usage rate distribution obtained from available collected data is then assumed to represent the usage information of all the products in warranty coverage. Once the distribution parameters are known, the probability of a product exceeding the usage limit at t MIS can be expressed as:

$$P(R \geq U_0/t) = \int_{U_0/t}^{\infty} g(x)dt,\ t \leq W_0 \tag{2}$$

Then, the expected number of products still in warranty at t MIS is,

$$E[N(t)] = N \times P(R \leq U_0/t) \tag{3}$$

3.3 Model Product Failures

Under the two-dimensional warranty policy, the degradation is a function of the age and usage. For a given value of usage rate r, the conditional hazard function (failure rate) for the time to the first failure is given by $h(t;r)$, which is non-decreasing function of both t and r. A counting process can model product failures over time. The products are supposed to be repairable, then the counting process is characterized by a conditional intensity function $\lambda_r(t)$. Because we assume that repair times are negligible and minimal repair is performed when a failure occurs under warranty, i. e., after the repair, the product would be restored to its condition immediately before the failure occurred, then $\lambda_r(t) = h(t;r)$ (Thompson WA 1981, Barlow and Hunter 1960).

A product is usually composed of several interconnected components. Product reliability is a function of the reliabilities of the various components of the product. During the design phase, decisions concerning component reliability are made to ensure that a product has the desired reliability at a nominal usage rate r_0. In the case of two-dimensional warranty, the nominal usage rate r_0 can be viewed as the ratio of the age limit to the usage limit, which means that the warranty expires when both the maximum age and usage limits are exceeded simultaneously. As the usage rate increases (decreases), the rate of degradation is likely to be faster (slower). When modeling the effect of the usage rate on the time to failure and the number of failures for a specified time interval, the AFT model is a popular choice (Jack et al. 2009, Wu 2014, Yang et al. 2016).

Denote T_0 as the time to the first failure under r_0, T_r is the time to the first failure under a given usage rate r. Using the AFT formulation, we have:

$$\frac{T_r}{T_0} = (\frac{r_0}{r})^{\gamma} \tag{4}$$

where γ is the parameter of the AFT model and $\gamma \geq 1$.

If the density function of T_0 is given by $f_{r_0}(t \mid \alpha_0, \beta_0, \gamma)$, where α_0 and β_0 are respectively the scale and shape parameters of T_0. According to Eq.(4), the density function for the failure time T_r is

$$f_r(t \mid \alpha_0, \beta_0, \gamma) \equiv (\frac{r}{r_0})^{\gamma} f_{r_0}((r/r_0)^{\gamma} t \mid \alpha_0, \beta_0, \gamma), t > 0 \tag{5}$$

Then the cumulative distribution function for T_r is $F_r(t \mid \alpha_0, \beta_0, \gamma)$, the scale parameter $\alpha_r = \alpha_0(r_0/r)^{\gamma}$, the shape parameter is unaffected because the products are assumed to have the same failure mode (Baik and Murthy 2008). The hazard function $h_r(t \mid \alpha_0, \beta_0, \gamma)$

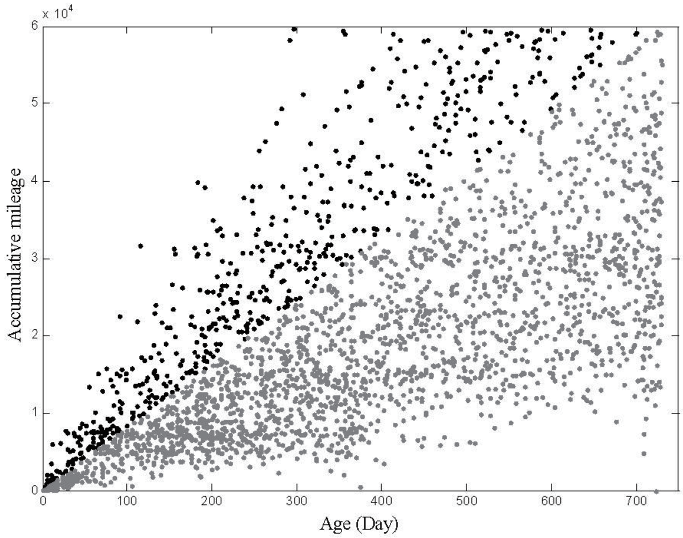

Figure 1 Mileage Accumulations of Automobiles in 2010

associated with $F_r(t \mid \alpha_0, \beta_0, \gamma)$ is given by:

$$h_r(t \mid \alpha_0, \beta_0, \gamma) = \frac{f_r(t \mid \alpha_0, \beta_0, \gamma)}{1 - F_r(t \mid \alpha_0, \beta_0, \gamma)}, t > 0 \tag{6}$$

Since failed products are assumed to be repaired minimally and the repair times are negligible, the breakdown process can be described by a NHPP (Akbarov and Wu 2012a, Iskandar et al. 2005, Baik et al. 2004). The intensity function has the same functional form as the hazard function for the time to the first failure (Thompson WA 1981, Barlow and Hunter 1960). Therefore, for a given usage rate r, the failure intensity function can be defined as:

$$\lambda_r(t \mid \alpha_0, \beta_0, \gamma) = h_r(t \mid \alpha_0, \beta_0, \gamma) \tag{7}$$

It is common to use the $Weibull$ distribution to model product failure in reliability analysis because of its flexibility and applicability to various failure processes. Modeling time to the first failure with a $Weibull$ distribution leads to the power-law process, a special issue of NHPP (Majeske 2007, Jack et al. 2009, Krivtsov 2000). In this study, T_0 is assumed to follow a $Weibull$ distribution. According to the AFT formulation, the hazard function of T_r is given as,

$$h_r(t \mid \alpha_0, \beta_0, \gamma) = \frac{\beta_0}{\alpha_0 (r_0/r)^\gamma} \left(\frac{t}{\alpha_0 (r_0/r)^\gamma}\right)^{\beta_0 - 1} \tag{8}$$

Therefore, the expected number of failures during a specified time interval is given by the conditional cumulative hazard function of the time to the first failure of the product.

3.4 Predict Warranty Claims

When making predictions for grouped products, the intensity function and the number of products still in warranty coverage are required to be integrated. Let d_k be the number of claims observed by the products under warranty in month k and $\hat{d}_k$ be the corresponding predicted value. Since we assume that all the products manufactured in the same batch have the same intensity function, the expected number of claims in k MIS is (Akbarov and Wu 2012ab, Majeske 2007, Wu 2012)

$$\hat{d}_k = E(d_k) = E[N(k)] \int_{k-1}^{k} \lambda_{\bar{r}_t}(t) dt \tag{9}$$

and the average rate $\bar{r}_t$ at time t of the products still in the warranty coverage is

$$\bar{r}_t = \int_{R_{\min}}^{U/t} \frac{x g(x)}{G(U/t) - G(r_{\min})} dx, t = 1, 2, \cdots, W_0 \quad (10)$$

where r_{min} is the lower bound of R and $G(\bullet)$ is known based on available historical data.

From aggregated data, we could not be able to obtain exact failure time of the products, but only knows the age lies within an interval. When analyzing aggregated claims data, the number of warranty claims is commonly assumed to follow a NHPP distribution (Ascher and Feingold 1984, Wu and Akbarov 2011, Akbarov and Wu 2012b, Wu S 2013, Kalbfleisch et al. 1991). Based on the observed claims of the new product in the T months, the likelihood function of the NHPP model is thus expressed as follows:

$$L_{NHPP}(\Theta) = \Pi_{k=1}^{T} \frac{(E[N(k)] \int_{k-1}^{k} \lambda_{\bar{r}_x}(x)dx)^{d_k} \exp(-E[N(k)] \int_{k-1}^{k} \lambda_{\bar{r}_x}(x)dx)}{d_k!} \quad (11)$$

where Θ is the column vector of parameters of lifetime distribution and d_k is the observed value at k MIS.

The log-likelihood function then becomes,

$$\log L(\Theta) = \sum_{k=1}^{T} d_k \log(E[N(k)] \int_{k-1}^{k} \lambda_{\bar{r}_x}(x)dx) - \sum_{k=1}^{T} E[N(k)] \int_{k-1}^{k} \lambda_{\bar{r}_x}(x)dx - \sum_{k=1}^{T} \log d_k \quad (12)$$

Using the expected number of products under warranty and the failure intensity function given in Eqs.(3) and (8), Eq.(12) can be rewritten as

$$\begin{aligned}\log L(\alpha_0, \beta_0, \gamma) &= \sum_{k=1}^{T} d_k \log(\frac{E[N(k)]}{\alpha_0^{\beta_0}} (\frac{\bar{r}_t}{r_0})^{\gamma\beta_0} (k^{\beta_0} - (k-1)^{\beta_0})) \\ &- \sum_{k=1}^{T} \frac{E[N(k)]}{\alpha_0^{\beta_0}} (\frac{\bar{r}_t}{r_0})^{\gamma\beta_0} (k^{\beta_0} - (k-1)^{\beta_0}) - \sum_{k=1}^{T} \log d_k\end{aligned} \quad (13)$$

The MLE of $(\alpha_0, \beta_0, \gamma)$ can be obtained from directly maximizing the likelihood function. Taking the derivatives of Eq.(13) with respect to α_0, β_0 and γ, a system of three equations of unknown variables can be obtained. Since the expressions are complex, it is quite difficult to find the solution of likelihood equations for $(\alpha_0, \beta_0, \gamma)$. We adopt the nonlinear constrained programming method to estimate the parameters of the model as has been done in Wang and Yang (2012). Mathematical model of finding minimum value of nonlinear constrained programing is

$$\begin{cases} \min(-\log L) \\ \text{s. t. } \alpha_0 > 0, 0 < \beta_0 < 10, \gamma \geq 1 \end{cases} \quad (14)$$

The objective function is a negative log-likelihood that is related to the model parameters and the constrained variables are model parameters. The estimated values of the model parameters can be obtained via solving this nonlinear constrained programming problem. In this study, α_0 can be set greater than zero. β_0 is set to be between the values of 0 and 10, because almost all expected solutions are in this range. According to the AFT model, the acceleration factor γ is set to be greater or equal to one.

In product manufacturing, the new generation of products usually has a close relationship with the previous generations of products it replaces because of similar components and manufacturing technology. In this context, Wu and Akbarov (2012) suggested that the warranty data of the earlier generations products can be used to extract reliability information of new products and developed the CMLE method for estimating parameters. Here, we adopt the CMLE method to settle parameter estimation for two-dimensional warranted products, as discussed below.

3.4.1 Phase I: Selecting Reference Products

The claims data of the earlier generations of products are used, referred to as candidate reference products that have been in the market. We assume that claim rates of reference prod-

Table 2 Parameters in the Simulation

Warranty coverage	$W_0 = 24(months), U_0 = 6(\times 10^4 km)$
Warranty policy	Free repair via minimal repair
Nominal usage rate	$r_0 = 0.25(10^4 km/month)$
Parameter of the AFT model	$\gamma = 2$
Number of the target product sold	$N = 10000$
Number of candidate reference products	$M = 29$

ucts are complete from 1 to k MIS, which can be obtained from warranty database. We then calculate the upper and lower of bounds of the claim rates of the candidate reference products. Denote M as the number of candidate reference products and $\lambda_{k,n}$ as the claim rate of product n $(n = 1, 2, \cdots, M)$ in k $(k = 1, 2, \cdots, T)$ MIS. Let λ_k be the claim rate of the new product in k MIS. The Pearson product-moment correlation coefficient is introduced to select reference products, which yields,

$$\rho_n = \frac{\sum_{k=1}^{T} (\lambda_{k,n} - \mu_n)(\lambda_k - \mu)}{\sqrt{\sum_{k=1}^{T} (\lambda_{k,n} - \mu_n)^2 \sum_{k=1}^{T} (\lambda_k - \mu)^2}} \tag{15}$$

where $\mu_n = \frac{1}{T} \sum_{k=1}^{T} \lambda_{k,n}$ and $\mu = \frac{1}{T} \sum_{k=1}^{T} \lambda_k$. We set C be the cut-off value. If $\rho_n > C$, product n is selected as one of the reference products. Then, there are M_0 products satisfying $\rho_n > C$ $(n = 1, 2, \cdots, M_0)$ and M_0 products have claim rates $\lambda_{k,1}, \lambda_{k,2}, \cdots, \lambda_{k,M_0}$. To compute $\mu_k^{'} = \frac{1}{M_0} \sum_{i=1}^{M_0} \lambda_{k,i}$ and $\sigma_k^{'} = \frac{1}{M_0 - 1} \sum_{i=1}^{M_0} \sqrt{(\lambda_{k,i} - \mu_k^{'})^2}$.

3.4.2 Phase II: Build a Mathematical Model Subject to Constrains

Find λ_k^l and λ_k^u so that $\lambda_k^l \leq \lambda_k \leq \lambda_k^u$, where the lower bound λ_k^l and the upper bound λ_k^u are chosen from $\mu_k^{'} - 3\sigma_k, \mu_k^{'} - 2\sigma_k, \cdots,$ $\mu_k^{'} + 2\sigma_k, \mu_k^{'} + 3\sigma_k$. The mathematical model is built under the assumption that the claim rates of the new product lie in $[\lambda_k^l, \lambda_k^u]$, as shown in Eq.(16).

$$\begin{aligned} &\max(L_{\text{NHPP}}) \\ &\text{s.t. } \hat{\lambda}_k^l \leq \int_{k-1}^{k} \hat{\lambda}_{\bar{r}_t}(t)dt \leq \hat{\lambda}_k^u \\ &\quad \alpha_0 > 0, 0 < \beta_0 < 10, \gamma \geq 1 \\ &\text{for } k = 1, 2, \cdots, T \end{aligned} \tag{16}$$

According to the mentioned approach suggested by Wang and Yang (2012), the solution of Eq.(16) can be obtained by solving a nonlinear constrained programming problem as following

$$\begin{aligned} &\min(-\log L) \\ &\text{s.t. } y_k^l \leq \int_{k-1}^{k} \lambda_{\bar{r}_t}(t)dt \leq y_k^u \\ &\quad \alpha_0 > 0, 0 < \beta_0 < 10, \gamma \geq 1 \\ &\text{for } k = 1, 2, \cdots, T \end{aligned} \tag{17}$$

When the model parameters are estimated, the expected number of warranty claims at k MIS can be derived using Eq.(9).

4. Data Experiments

In this section, data experiments are conducted that are based on both simulation and real data collected from an automobile manufacturer to validate the proposed forecasting model.

4.1 Simulation Study

We assume that the product under consideration is a new generation of automobiles sold with a two-dimensional warranty policy. The parameter values of the proposed model are given in Table 2.

In this section, we utilize the estimated parameters of usage rate distribution based on the accumulated usage data in Section 3.2.

The usage rate of the historical data follows a *Lognormal* distribution ($\mu_r = -1.86, \sigma_r = 0.58$) ranging from 0.003 to 1.032. Suppose that time to the first failure under the nominal usage rate r_0 of the new products I and II follow the *Weibull* distribution with $\alpha_0 = 100, \beta_0 = 1.5$ and $\alpha_0 = 300, \beta_0 = 1.2$, respectively. M products are selected as the candidate reference products for each new product. In accordance with Wu and Akbarov (2012), three values of C, 0.6, 0.7 and 0.8 are considered.

To enable a comparison of results among the proposed model using the CMLE, the NHPP model using the MLE and the $\log-\log R(t)$ model, NRMSE is calculated using

$$NRMSE = \sqrt{\frac{\sum_{k=1}^{T}(d_k - \hat{d}_k)^2}{\sum_{k=1}^{T} d_k^2}} \tag{18}$$

where $\hat{d}_k$ is the predicted value of d_k, T is set to be 24. The smaller NRMSE values means higher predicting performance.

Many automobile manufacturers usually adopt the $\log-\log R(t)$ model to predict warranty claims. They track cumulative claims by month in t *MIS* using,

$$R(t) = \sum_{i=0}^{t} \frac{d_i}{N(i)} 1000 \tag{19}$$

One way used to predict $\hat{R}(t)$ utilizes transformed data via regressing $\log R(t)$ on $\log(t)$. Starting at 1 *MIS*, to fit the simple regression model

$$\log R(t) = a + b\log(t) + \xi_t \tag{20}$$

where a and b are regression coefficients, ξ_t is stochastic error term.

To obtain the estimate of $R(t)$, we simply extend the fit line and transform back to the original units of $R(t)$ using

$$\hat{R}(t) = \exp(\hat{a} + \hat{b}\log(t)) \tag{21}$$

The number of warranty claims received in k *MIS* of each new product is sampled from a *Poisson* distribution with mean $y_k = N(k)\int_{k-1}^{k}\lambda_{\bar{r}_x}(x)dx$ for $k = 1, 2, \cdots, 24$, where $\lambda_{\bar{r}_k}(k)$ is the intensity function and defined as:

$$\lambda_{\bar{r}_k}(k) = \beta_0(\frac{\bar{r}_k}{r_0})^{\gamma\beta_0}(\frac{k^{\beta_0-1}}{\alpha_0^{\beta_0}}) \tag{22}$$

where α_0 is the scale parameter and β_0 is the shape parameter of $\lambda_{\bar{r}_k}(k)$ for the new products I, II. The values of $\bar{r}_k$ in k *MIS* could be obtained by Eq.(10) based on known usage rate distribution. Let us generate the warranty claims of 29 candidate reference products from the intensity function as given in Eq.(22) with different combinations of $(\alpha_0, \beta_0, \gamma)$. The scale parameter α_0, the shape parameter β_0 and the AFT parameter γ are randomly selected from $(100, 600)$, $(0, 4)$ and $(1, 5)$, respectively.

Firstly, we predict values for the total number of products still under warranty $\hat{N}(t)$ through a time period of 24 months using Eqs.(2) and (3). Figure 2 shows the effect of various usage rate on $\hat{N}(t)$. Since the usage rate varies across the customer population, the number of products still in warranty coverage is altering over time. As shown in Figure 2, $\hat{N}(t)$ with various usage rate estimates 20 percent of the products have left warranty coverage due to mileage exceeding the limit. Nevertheless, without considering various usage rates would lead to overestimating.

The predictions of warranty claims of new products I and II are made based on Eqs.(9), (13) and (17) accordingly. Please note that the solution of *MLE* can be obtained via solving a nonlinear constrained programming problem as shown in Eq.(14). Table 3 shows the prediction performance (i.e. the values of NRMSE) of each forecasting model incorporating varying usage rate based on the artificially generated dataset. Furthermore, Table 3 contains the NRMSE values derived from these forecasting

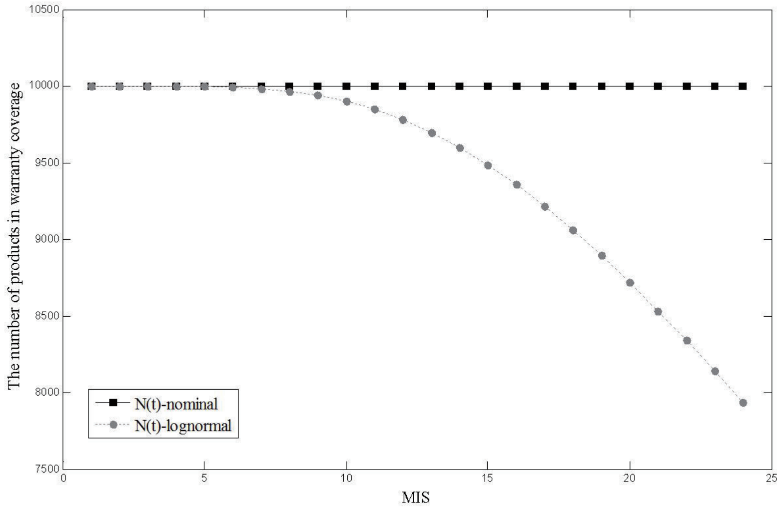

Figure 2 Estimated Number of Products in Warranty

Table 3 Simulation Results Based on the Forecasting Models with and without Considering the Usage Rate

	Product I		Product II	
	With usage rate considered	Without usage rate considered	With usage rate considered	Without usage rate considered
MLE	0.338	0.421	0.492	0.506
$\log - \log R(t)$	0.479	0.611	0.536	0.885
CMLE ($C = 0.6$)	0.260	0.410	0.389	0.493
CMLE ($C = 0.7$)	0.253	0.405	0.398	0.497
CMLE ($C = 0.8$)	0.258	0.409	0.402	0.502

models with a nominal usage rate. It is worth noting that the models with a nominal usage rate are equivalent to the models without usage rate considered (Yang et al. 2016). Those NRMSE values are the average values of running the simulation for 1000 times.

Table 3 shows that when considering various usage rate, the proposed model using CMLE has the lowest NRMSE. The results further show that these forecasting models considering varying usage rate are characterized by greater prediction accuracy than the models based on a nominal usage rate. In addition, the estimated values and mean squared error of estimates of parameter of these forecasting models considering varying usage rate are presented in Table 4. We can observe that average estimate of $(\alpha_0, \beta_0, \gamma)$ derived from the proposed model using CMLE are closer to the settings and has smaller standard errors. These facts indicate that the proposed model using CMLE which considers various usage rate has the best forecasting performance.

4.2 Case Study

We use the accumulated usage data of automobile engines as mentioned in Section 3.2. The manufacturer provides a two-dimensional warranty of 24-month or $6(\times 10^4 km)$ since the products are purchased. During the year of 2010, the usage and time-in service information of 3131 automobiles are recorded. The collected data set includes 1253 maintenance data and 1878 customer survey data. These data are used to estimate the usage rate distribution parameters representing the entire customer population. The average usage rate of the product with two or more failures can be

Table 4 Estimated Parameters, the Standard Errors Derived from the Forecasting Models Considering Usage Rate

Setting $\alpha_0 = 100, \beta_0 = 1.5, \gamma = 2$		Estimate value	Std. error	Setting $\alpha_0 = 300, \beta_0 = 1.2, \gamma = 2$		Estimate value	Std. error
$\log - \log R(t)$	a	−0.78	0.35	$\log - \log R(t)$	a	−0.53	0.28
	b	1.36	0.13		b	1.08	0.10
	α_0	160.8	61.9		α_0	175.9	95.7
MLE	β_0	1.38	0.18	*MLE*	β_0	1.40	0.17
	γ	1.76	0.46		γ	2.49	0.59
	α_0	96.9	31.6		α_0	238.2	84.1
CMLE ($C = 0.6$)	β_0	1.54	0.13	CMLE ($C = 0.6$)	β_0	1.27	0.10
	γ	2.13	0.41		γ	2.25	0.43
	α_0	102.6	35.5		α_0	242.9	86.4
CMLE ($C = 0.7$)	β_0	1.52	0.11	CMLE ($C = 0.7$)	β_0	1.29	0.10
	γ	2.05	0.33		γ	2.28	0.47
	α_0	90.0	28.2		α_0	254.8	87.0
CMLE ($C = 0.8$)	β_0	1.58	0.12	CMLE ($C = 0.8$)	β_0	1.27	0.12
	γ	2.13	0.39		γ	2.34	0.46

Table 5 NRMSE Values from the Forecasting Models with or without Considering Usage Rate for Case1 and Case2

	Case1: Product I		Case2: Product II	
	With usage rate considered	Without usage rate considered	With usage rate considered	Without usage rate considered
MLE	0.243	0.482	0.170	0.386
$\log - \log R(t)$	0.604	0.721	0.533	0.632
CMLE ($C = 0.6$)	0.134	0.480	0.105	0.382
CMLE ($C = 0.7$)	0.117	0.477	0.102	0.360
CMLE ($C = 0.8$)	0.125	0.438	0.103	0.367

obtained as:

$$\hat{r} = \frac{\sum_{i=1}^{m} (u_i/t_i)}{m} \tag{23}$$

where m is the number of claims experienced by the product.

The histograms (as shown in Figure 3) of the studied automobiles manufactured are presented to identify the possible distribution of the usage rate. To obtain best-fit distribution, the probability plots are shown in Figure 4. It can be visually observed from the plots that the usage rate follows a *Lognormal* distribution. Moreover, the Kolmogorov-Smirnov $(K - S)$ goodness of fit test is used on the observations for evaluating the fitness of the considered distributions. The results of $K - S$ test at a significance level of 95% show that the *lognormal* distribution ($p = 0.215$) fits best. The estimates of the location, and the shape parameters of the *lognormal* distribution are obtained as −1.86 and 0.578, respectively.

The products sold in Dec.2009 is considered to be new product I and products sold in Feb.2010 is new product II, referred to as Case 1 and Case 2, respectively. We utilize usage rate distribution parameters estimated earlier to calculate the average rate $\bar{r}_t$ at t *MIS* of the products under warranty. Nine products are selected as candidate reference products for each new product, respectively.

Table 5 shows the measures of the prediction accuracy for new products I and II using NRMSE. It can be seen from the table that when

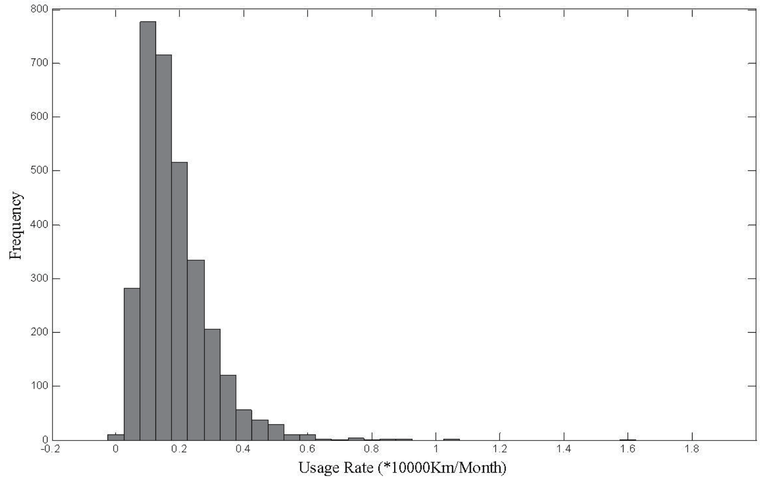

Figure 3 Histogram of Usage Rate

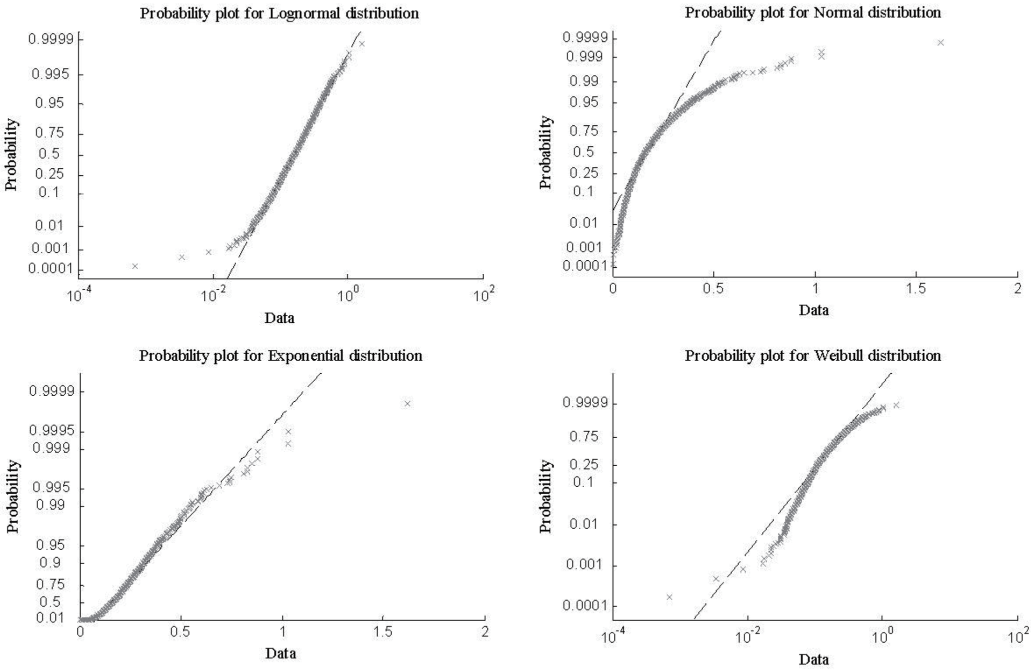

Figure 4 Four Probability Plots of Usage Rate

considering varying usage rate, the proposed model using the CMLE has the lowest NRMSE value. Moreover, both the proposed model using the CMLE and the NHPP model using the MLE perform better than the $\log - \log R(t)$ model for all cases.

As shown in Table 5, the NRMSE values derived from each forecasting model considering varying usage rate are all smaller than those models with a nominal usage rate. This supports our statement that to accurately forecast warranty claims, it is necessary to take varying usage rate into consideration especially for products having higher usage rate. We also provide the estimated parameters and standard errors of these three models considering

Table 6 Estimated Parameters, the Standard Errors in the Actual Data for Case1 and Case2

Case 1		Estimate value	Std. error	Case 2		Estimate value	Std. error
$\log - \log R(t)$	a	0.29	–	$\log - \log R(t)$	a	0.15	–
	b	1.15	–		b	1.11	–
MLE	α_0	232.2	41.7	*MLE*	α_0	236.9	62.4
	β_0	1.06	0.10		β_0	1.02	0.08
	γ	1.70	0.35		γ	2.66	0.38
CMLE ($C = 0.6$)	α_0	95.1	12.6	CMLE ($C = 0.6$)	α_0	141.9	27.1
	β_0	1.22	0.05		β_0	1.16	0.08
	γ	3.0	0.20		γ	3.0	0.17
CMLE ($C = 0.7$)	α_0	99.8	9.2	CMLE ($C = 0.7$)	α_0	131.5	21.3
	β_0	1.21	0.03		β_0	1.17	0.05
	γ	2.9	0.27		γ	3.0	0.22
CMLE ($C = 0.8$)	α_0	112.5	14.8	CMLE ($C = 0.8$)	α_0	138.5	20.8
	β_0	1.2	0.05		β_0	1.20	0.06
	γ	3.0	0.17		γ	3.0	0.30

varying usage rate for different cases in Table 6. Overall, the proposed model using CMLE is statistically more efficient than the other models.

To compare the results of three forecasting models incorporating various usage rate discussed more directly, the observed claims and the related prediction values are depicted in Figures 5 and 6, respectively.

Figure 5 shows the proposed model using the CMLE does exhibit a good fit to the actual warranty claims especially for $C = 0.7$. Besides, we can see that majority of claims occurs during the initial lifetime of the products, which seems to be caused by manufacturing or assembly defects. Moreover, Figure 5 indicates that the number of claims decreases after 10 MIS. This is due to fact that an increasing number of automobiles leaves warranty coverage. Please note that because of high claims rates in the early product lifetime, the $\log - \log R(t)$ model accurately predicts claims from 1 to 10 MIS, however, consistently overestimates claims after 11 MIS. This result agrees with the findings of Majeske (2007).

Figure 6 shows the observed and predicted claims by MIS for Case 2. Observe from Figure 6 that the proposed model provides a good fit for the actual warranty claims when $C = 0.7$. In addition, it also reveals that the $\log - \log R(t)$ model overestimates the number of warranty claims after 9 MIS because of the high claim rate at the beginning of the warranty period.

5. Conclusions

According to the connection between new products and the previous generations of products, we propose a warranty claims forecasting model considering usage rate for new products sold with a two-dimensional warranty. In the proposed model, we take the usage rate as a random variable and assume the usage rate keeps constant of a product and varies among different customers. Under the assumption of minimal repair on the failed products, the NHPP is used to model the number of failures of repairable products where the CMLE is adopted to estimate model parameters. The effects of the usage rate on both the number of products under warranty and product reliability are taken into consideration.

Data experiments based on both simulation and real warranty data collected from an automobile manufacturer in China are conducted

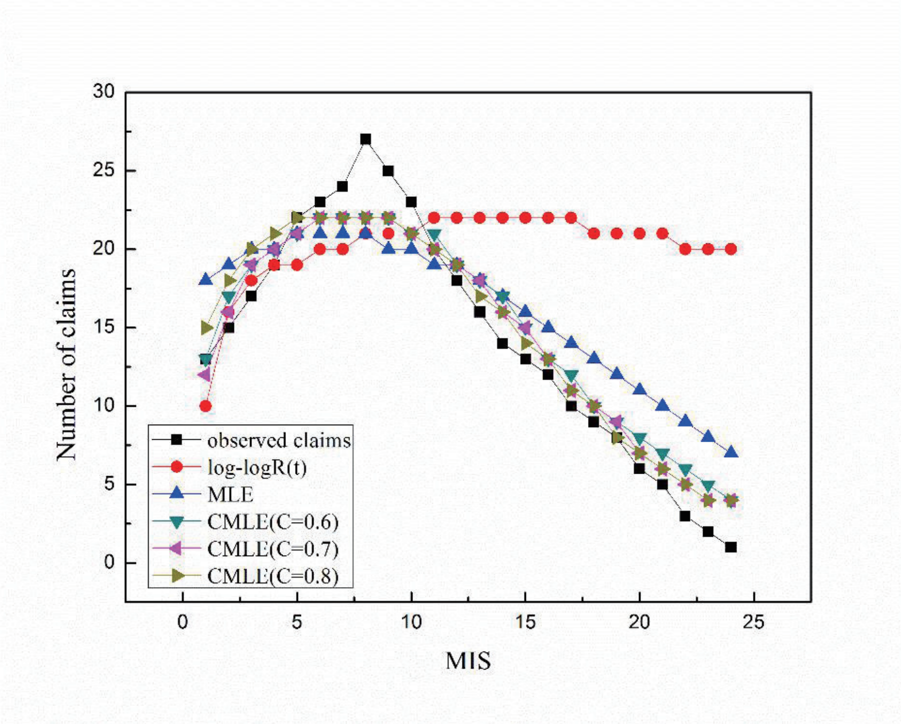

Figure 5 Prediction Validation for Three Models Considering Usage Rate Against Actual Data for Case 1

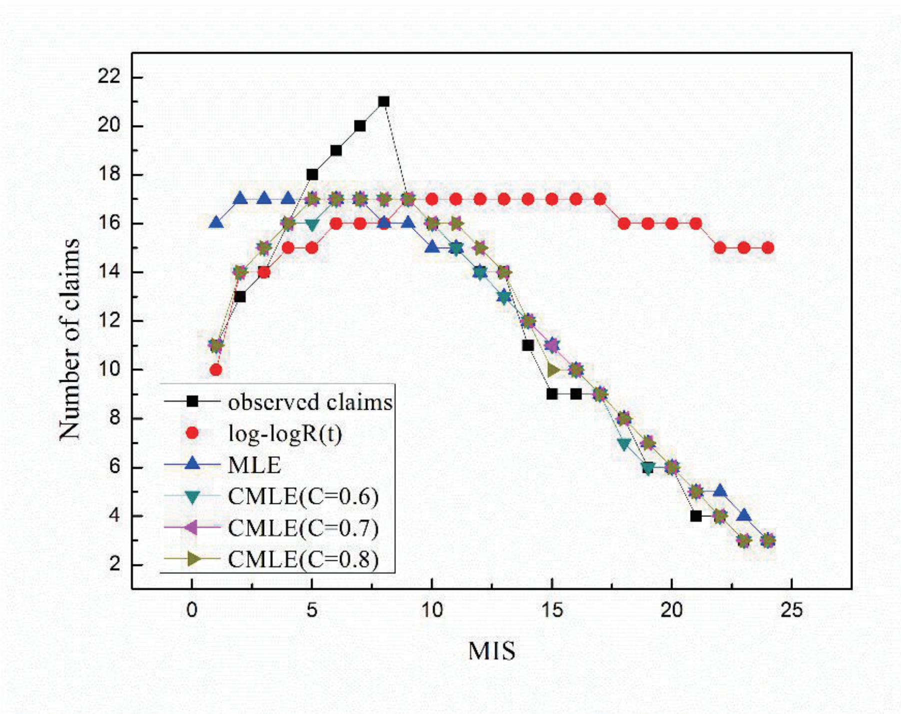

Figure 6 Prediction Validation for Three Models Considering Usage Rate Against Actual Data for Case 2

to test the performance of the proposed model. The results of data experiments show that the proposed model using CMLE which considers the varying usage rate has a higher predicting accuracy than the NHPP model using the MLE and the $\log - \log R(t)$ model. The results further show that the proposed forecasting models that consider varying usage rate outperform the models that do not consider the usage rate.

Some future topics are worth investigating in this area including consideration of different actions taken for failed items such as imperfect repair or replacement. Moreover, we will analyze censored and incomplete warranty data and develop more refined techniques for spare parts optimization using the information of lifetime distribution.

6. Acknowledgments

This work is supported by National Natural Science Foundation of China under Grant Nos. 71802145, 71902180, 71872123, and 71801064, 71532008. The author would like to thank the reviewers for their constructive comments on the earlier version of the manuscript.

References

Akbarov A, Wu S (2012a). Warranty claim forecasting based on weighted maximum likelihood estimation. *Quality and Reliability Engineering International* 28:663-669.

Akbarov A, Wu S (2012b). Forecasting warranty claims considering dynamic over-dispersion. *International Journal of Production Economics* 139(2):615-622.

Ascher H, Feingold H (1984). Repairable systems reliability: Modeling, inference, misconceptions and their causes. *Marcel Dekker: New York.*

Baik J, Murthy DNP, Jack N (2004). Two-dimensional failure modeling with minimal repair. *Naval Research Logistics* 51(3):345-362.

Barlow RE, Hunter L (1960). Optimum preventive maintenance policies. *Operations Research* 8(1):90-100.

Baik J, Murthy DNP (2008). Reliability assessment based on two-dimensional warranty data and an accelerated failure time model. *International Journal of Reliability and Safety* 2(3):190-208.

Bian Y, Yan S, Zhang W, Xu H (2015). Warranty strategy in a supply chain when two retailer's exrended warranties bundled with the products. *Journal of Systems Science and Systems Engineering* 24(3):364-388.

Dai A, He Z, Liu Z, Yang D, He S (2017). Field reliability modeling based on two-dimensional warranty data with censoring times. *Quality Engineering* 29(3):468-483.

Fredette M, Lawless JF (2007). Finite-horizon prediction of recurrent events, with application to forecasts of warranty claims. *Technometrics* 1706 (49):66-80.

Iskandar BP, Murthy DNP, Jack N (2005). A new repair-replace strategy for items sold with a two-dimensional warranty. *Computers and Operations Research* 32:669-682.

Jack N, Iskandar BP, Murthy DNP (2009). A repair-replace strategy based on usage rate for items sold with a two-dimensional warranty. *Reliability Engineering and System Safety* 94(2):611-617.

Kalbfleisch JD, Lawless JF, Robinson JA (1991). Methods for the analysis and prediction of warranty claims. *Technometrics* 33(3):273-285.

Kaminskiy MP, Krivtsov VV (2000). G-renewal process as a model for statistical warranty claim prediction. *Proceedings Annual Reliablity and Maintainability Symposium.*

Karim MR, Suzuki K (2005). Analysis of warranty claim data: A literature review. *International Journal of Quality and Reliability Management* 22(7): 667-686.

Kleyner A, Sandborn P (2005). A warranty forecasting model based on piecewise statistical distributions and stochastic simulation. *Reliability Engineering and System Safety* 88:207-214.

Krivtsov V (2000). A monte carlo approach to modeling and estimation of the generalized renewal process in repairable system reliability analysis. PhD Dissertation, University of Maryland.

Lawless J, Hu J, Cao J (1995). Methods for the estimation of failure distributions and rates from automobile warranty data. *Lifetime Data Analysis* 1(3):227-240.

Lawless JF, Crowder MJ, Lee KA (2009). Analysis of reliability and warranty claims in products with age and usage scales. *Technometrics* 51(1): 14-24.

Limon S, Yadav OP, Zuo MJ, Muscha J, Honeyman R (2016). Reliability estimation considering usage rate profile and warranty claims. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 230(3):297-308.

Majeske KD (2007). A non-homogeneous poisson process predictive model for automobile warranty claims. *Reliability Engineering and System Safety* 92 (2):243-251.

Moskowitz H, Chun YH (1994). A poisson regression model for two-attribute warranty policies. *Naval Research Logistics* 41(3):355-376.

Rai BK (2009). Warranty spend forecasting for subsystem failures influenced by calendar month seasonality. *IEEE Transactions on Reliability* 58(4): 649-657.

Rai B, Sing N (2005a). A modeling framework for assessing the impact of new time/mileage warranty limits on the number and cost of automotive warranty claims. *Reliability Engineering and System Safety* 88(2):157-169.

Rai B, Singh N (2005b). Forecasting automobile warranty performance in presence of "maturing data" phenomena using multilayer perceptron neural network. *Journal of Systems Science and Systems Engineering* 14(2):159-176.

Rai B, Singh N (2006). Customer-rush near warranty expiration limit, and nonparametric hazard rate estimation from known mileage accumulation rates. *IEEE Transactions on Reliability* 55(3):480-489.

Shahanaghi K, Noorossana R, Jalali-Naini SG, Heydari M (2013). Failure modeling and optimizing preventive maintenance strategy during two-dimensional extended warranty contracts. *Engineering Failure Analysis* 28:90-102.

Thompson WA (1981). On the foundations of reliability. *Technometrics* 23:1-13.

Wang Y, Liu Z, Liu Y (2015). Optimal preventive maintenance strategy for repairable items under two-dimensional warranty. *Reliability Engineering and System Safety* 142:326-333.

Wang ZM, Yang JG (2012). Numerical method for weibull generalized renewal process and its applications in reliability analysis of NC machine tools. *Computers and Industrial Engineering* 63(4):1128-1134.

Wasserman GS (1992). An application of dynamic linear models for predicting warranty claims. *Computers and Industrial Engineering* 22(1):37-47.

Wu S (2012). Warranty data analysis: A review. *Quality and Reliability Engineering International* 28(8):795-805.

Wu S (2013). A review on coarse warranty data and analysis. *Reliability Engineering System Safety* 114:1-11.

Wu S, Akbarov A (2011). Support vector regression for warranty claim forecasting. *European Journal of Operational Research* 213(1):196-204.

Wu S, Akbarov A (2012). Forecasting warranty claims for recently launched products. *Reliability Engineering and System Safety* 106:160-164.

Wu S (2014). Construction of asymmetric copulas and its application in two-dimensional reliability modelling. *European Journal of Operational Research* 238(2):476-485.

Yang D, He Z, He S (2016). Warranty claims forecasting based on a general imperfect repair model considering usage rate. *Reliability Engineering and System Safety* 145:147-154.

Anshu Dai is an assistant professor in the Co-innovation Center for Computable Modeling in Management Science at Tianjin University of Finance and Economics, China. She received her Ph.D. degree from Tianjin University, China. Her research interests include warranty data analysis and warranty policy design.

Zhaomin Zhang is an assistant professor in the College of Aeronautic Engineering at Civil Aviation University of China. He received his Ph.D. degree from Tianjin University, China. His current research interests include warranty policy design and maintenance service.

Pengwen Hou is an assistant professor in the Co-innovation Center for Computable Modeling in Management Science at Tianjin University of Finance and Economics, China. He received his Ph.D. degree from Tianjin University, China. His research interests include operations research and management mathematics.

Jingyi Yue is a Ph.D. candidate in the Co-innovation Center for Computable Modeling in Management Science at Tianjin University of Finance and Economics, China. Her current research include control theory and optimization algorithm.

Shuguang He is a professor in College of Management and Economics at Tianjin University, China. He received his Ph.D. degree from Tianjin University, China. His research interests include statistical process controls, information management, and warranty data analysis.

Zhen He is currently an academician of International Academy for Quality and a distinguished professor in College of Management and Economics at Tianjin University, China.He received the B.S., M.S., and Ph.D. degrees from Tianjin University, Tianjin, China, 1988,1991, and 2000, respectively. His current research interests include warranty management,statistical process controls, design of experiments, and response surface methodology.