Forecasting warranty claims considering dynamic over-dispersion

Artur Akbarov, Shaomin Wu *

School of Applied Sciences, Cranfield University, Bedfordshire MK43 0AL, UK

Abstract

Forecasting warranty claims is vitally important for manufacturers in preparing their fiscal plans as well as in managing their inventory. One of the widely used forecasting models is the non-homogeneous Poisson process (NHPP), which assumes that the mean and the variance of the numbers of warranty claims at any given time interval are equal. However, this is not always the case. Warranty claim data often exhibit a phenomenon known as *over-dispersion*, which implies that the variance to mean ratio is larger than one. Furthermore, this ratio might change over time and can have a trend or a clearly discernible functional form, which has not yet been considered in the existing literature on warranty claims forecasting.

This paper presents a warranty claim forecasting approach that tackles the problem of the dynamic over-dispersion exhibited in warranty claims data. It considers the application of both mixed NHPP and Cox process models to warranty claims and assumes that the intensity of the mixed NHPP follows a gamma distribution and the intensity of the Cox process follows a gamma process. Warranty claim data collected from an electronics product manufacturer are used validate the models, which show that these models outperform conventional NHPP models.

Key words: Warranty data, Poisson process, Cox process, mixed Poisson process, warranty forecasting, warranty prediction

1 Introduction

Product warranty has become a ubiquitous feature of product sales and serves many different purposes (see Murthy and Djamaludin (2002); Wu (2011), for

* Suggested citation: Akbarov, A., Wu, S. Forecasting warranty claims considering dynamic over-dispersion (2012) International Journal of Production Economics, 139 (2), pp. 615-622.
Corresponding author: s.m.wu@kent.ac.uk

example). Accurately forecasting the number of warranty claims can help manufacturers/warranty suppliers in preparing their fiscal plans and stocking their inventories. Starting with Kalbfleisch et al. (1991), research on forecasting warranty claims has received considerable attention (see Stephens and Crowder (2004); Majeske (2007); Fredette and Lawless (2007); Wu and Akbarov (2011), for example).

Warranty data can often be represented as a contingency table shown in Table 2. Count data in such a table is commonly modelled with Poisson processes (Bishop et al., 1975; Lawless and Kalbfleisch, 1992; Wang et al., 2002).

A stochastic process might exhibit a phenomenon called over-dispersion, which has the variance to mean ratio of the process at any time interval larger than 1. Modelling a stochastic process exhibiting the over-dispersion phenomenon has been addressed by previous studies such as Kalbfleisch et al. (1991), Fredette and Lawless (2007), and Kalbfleisch and Lawless (1996). However, the dynamic nature of the over-dispersion phenomenon in warranty claims data has received little attention. In the existing literature, when dealing with the over-dispersion, authors normally assume a stochastic process with a constant variance. However, this might not be true in reality. In this paper, we consider a measure of the over-dispersion, which changes with time. In particular, we consider a mixed Poisson process where the variance of the mixing distribution changes with time, and a Cox process model where the parameters of the mixing distribution for each time period are different.

Some previous studies use non-parametric modelling techniques to predict warranty claims. Wasserman and Sudjianto (1996) use the multi-layer perceptron (MLP) neural network to build warranty forecasting models. Rai and Singh (2005) use the radial basis function (RBF) neural network to forecast warranty claims. Hrycej et al. (2007) also use a MLP neural network to forecast warranty cost based on individual vehicle variables (i.e. age, monthly mileage rate, and road condition index) and the overall manufacturing quality fluctuation risk (i.e. different technical groups). Wu and Akbarov (2011) use support vector regression to build time series models and regression models to predict warranty claims and conclude that these models outperform MLP and RBF neural networks. It is known that, when the form of the failure rate is known, that is, the underlying failure generating process is known, the parametric methods can outperform the non-parametric methods. We will consider such a comparison in our future work.

In this study, we focus only on Poisson processes, which are often used as forecasting tools in applications such as forecasting demand for inventory control of spare parts (Kennedy et al., 2002; Lindsey and Pavur, 2009; Syntetos et al., 2010) and forecasting insurance claims (Fahrmeir and Echavarria, 2006). More specifically, we focus on the non-homogeneous Poisson process (NHPP) and its extensions. The NHPP models are widely used in reliability and warranty claim data analysis (see Kalbfleisch et al. (1991); Lawless (1998); Majeske (2007); Fredette and Lawless (2007); Yun et al. (2008), for example).

The remainder of this paper is structured as follows. Section 2 discusses the over-dispersion phenomenon in warranty claims data and presents a brief review of literature concerned with modelling warranty data considering this phenomenon. Section 3 presents models that can deal with over-dispersion, namely, mixed non-homogeneous Poisson process and Cox process models. Section 4 presents case studies based on warranty claims data collected from an electronics manufacturer. Section 5 discusses the strengths and weaknesses of the models presented in Section 3. Section 6 draws conclusions from this study.

Table 1
Notation

t	months since the date of manufacture.
$d_{i,t}$	number of warranty claims in month t from production batch of month i.
d_t	$= \sum_i d_{i,t}$, number of warranty claims in month t summed over all i.
s_i	number of products shipped in month i.
S	$= \sum_i^n s_i$, total number of products shipped out.
$h(x)$	intensity function of the Poisson process.
M_t	random variable, which is the number of warranty claims in month t.

2 Problem statement and prior work

The warranty data used for the case study in this paper are collected from a leading electronics manufacturer and consist of two parts, see Table 2. The first part is monthly records of warranty claims matched to the product's date of manufacture, and the second is the number of monthly shipments. Table 2 shows the format of the available data. In this study we assume that the number of monthly shipments adequately represents the number of manufactured products in corresponding months.

2.1 Problem statement

The non-homogenous Poisson process (NHPP) assumes that the mean and the variance of a stochastic process at a given time interval are equal. That is, the variance to mean ratio is 1. However, in some cases, count data such as

Table 2
Warranty data: s_i shipments in month i, $\mathbf{S}$ total shipments, $d_{i,t}$ warranty claims in month t for products produced in month i, and $\mathbf{d}_t$ total claims in month t.

Manufacture	Shipment	Months since the date of manufacture				
date	amount	1	2	...	n-1	n
1	s_1	$d_{1,1}$	$d_{1,2}$	...	$d_{1,n-1}$	$d_{1,n}$
2	s_2	$d_{2,1}$	$d_{2,2}$	...	$d_{2,n-1}$	$d_{2,n}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	s_n	$d_{n,1}$	$d_{n,2}$	...	$d_{n,n-1}$	$d_{n,n}$
Total	$\mathbf{S}$	$\mathbf{d_1}$	$\mathbf{d_2}$	...	$\mathbf{d_{n-1}}$	$\mathbf{d_n}$

insurance data and warranty data might exhibit over-dispersion (Kalbfleisch et al., 1991), where the variance to mean ratio is larger than 1.

Warranty data, as shown in Table 2, include monthly warranty claims, $d_{i,t}$, and monthly shipments, s_i. The over-dispersion phenomenon can be detected using Pearson residuals. If there is no over-dispersion, Pearson residuals distribute according to a normal distribution with variance 1 (Kalbfleisch et al., 1991; Bishop et al., 1975). The Pearson residuals are $r_{i,t} = (d_{i,t} - \hat{d}_{i,t})^2/\hat{d}_{i,t}$, where $\hat{d}_{i,t}$ can be estimated in two different ways. The first is to use warranty claim data only, that is $\hat{d}_{i,t} = (\sum_i d_{i,t} \sum_t d_{i,t})/\sum_i \sum_t d_{i,t}$, see Bishop et al. (1975). This method is referred to as non-parametric method. The second is to use warranty claim data along with the shipment data, that is $\hat{d}_{i,t} = s_i(\sum_i d_{i,t}/\sum_i s_i)$. This method is referred to as parametric method. The non-parametric method is used to estimate the mean of a Poisson distribution when the sample sizes are not known, that is, monthly manufactured amounts are not known. Since, monthly shipments only roughly represent the manufactured amounts in a month and we use both methods to estimate the Pearson residuals.

Figure 1 shows the variance of the Pearson residuals for one of the products (Product 1) using both the non-parametric and parametric methods. It is clear from the figure that the variances of the Pearson residuals are larger than 1. It can also be noted that the variances estimated with the parametric method are much larger than the variances estimated with the non-parametric method. This is due to the fact that shipment amounts do not accurately represent manufactured amounts, and thus introduce additional variation into the data. Similar results were obtained for other products.

Figure 2 shows the variance to mean ratio for Product 2. It is clear from the figure that the variance is larger than the mean at any given month t.

Figure 3 is the Q-Q plot for the Pearson residuals estimated with the non-parametric method for Product 1 at time $t = 5$. It is clear from the figure that

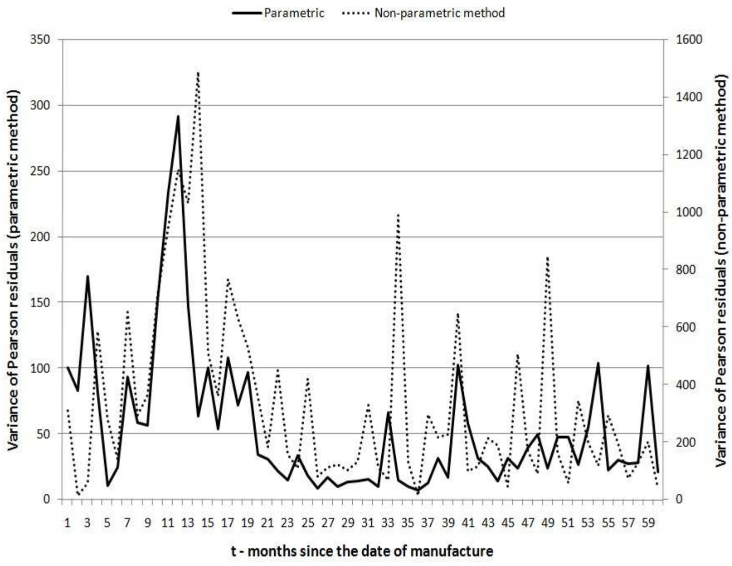

Fig. 1. Variances of Pearson residuals for Product 1.

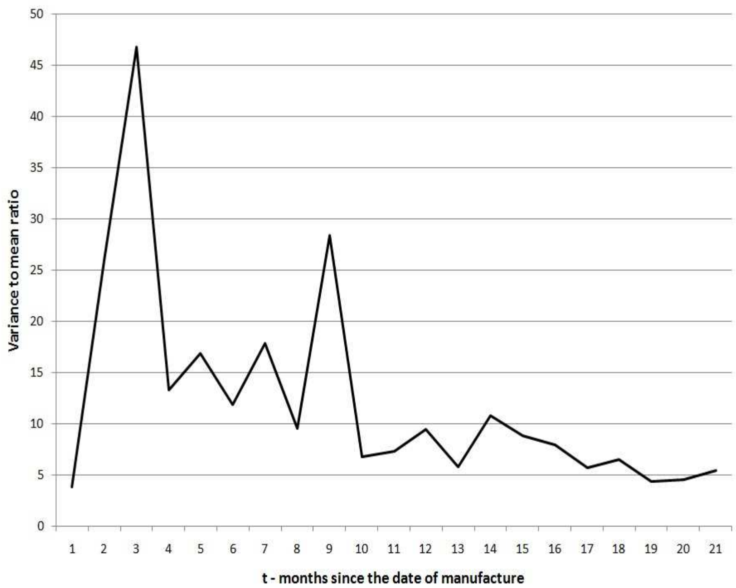

Fig. 2. Variance to mean ratio for Product 2.

the residuals are not normally distributed. Similar results were obtained for other products at different time periods.

2.2 Prior work

This section gives a brief literature review on the use of Poisson processes for modelling warranty data. It also includes the review of methods for tackling

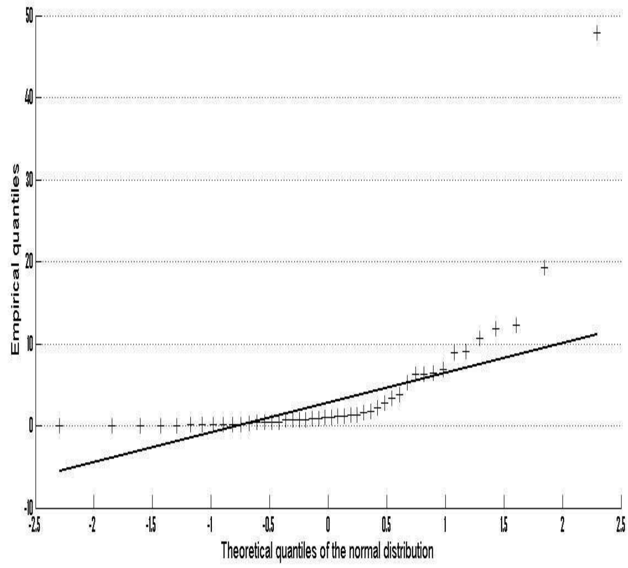

Fig. 3. Q-Q plot of Pearson residuals for Product 1 at time $t = 5$.

the issue of over-dispersion exhibited in warranty data.

Kalbfleisch et al. (1991) use the Poisson model to analyse automobile warranty data with reporting delays, where the reporting delay is the time from a product failure to the time when the failure claim is entered into a database. Hence, the observed data represents the time from the date of sale until the date of a failure plus the reporting delay. The authors recognise that the repair rates of individual cars will vary, however, the repair counts obtained by considering a large fleet of cars are expected to be close to Poisson counts when the repair rates are small. The authors also consider the presence of over-dispersion in the data and tackle the issue by introducing an unobservable random variable α_i, which is associated with each automobile unit i. The α_is are assumed to be independent and identically distributed with $\mathrm{E}(\alpha_i) = 1$ and $\mathrm{Var}(\alpha_i) = \sigma^2$. So that, the ith unit is assumed to generate claims according to a Poisson model with expected claims at a given age t_a, $\alpha_i \lambda(t_a)$. σ^2 is constant for all units. The extra variation in the data is thought to have risen due to the heterogeneity of automobile units.

Kalbfleisch and Lawless (1996) also consider a NHPP that allows for over-dispersion. The over-dispersion is handled by assuming that the number of claims at age t_a for cars sold at time t_s, $n(t_s, t_a)$ is a random variable with mean $\mu(t_s, t_a) = N(t_s)\lambda(t_a)$ and variance $\mathrm{Var}(n(t_s, t_a)) = \sigma^2 \mu(t_s, t_a)$. When $\sigma^2 = 1$, the model is a Poisson model, and when $\sigma^2 > 1$, the model allows for extra variation. σ^2 is estimated as the variance of the whole data set; thus, it is constant for all t_a. The authors also state that over-dispersion arises from several sources, including inherent variation in the robustness of units, variations in usage environment, and non-Poisson claim patterns for individual units. Lawless (1998) considers similar models as in Kalbfleisch et al. (1991) and Kalbfleisch and Lawless (1996) on a product unit level.

Fredette and Lawless (2007) propose a mixed NHPP model for dealing with over-dispersion caused by heterogeneity of products. The expected number of warranty claims for each product (or equivalently for a single process) is assumed to be given by $\lambda(t) = \alpha f(t)$, where $\lambda(t)$ is the intensity function of the NHPP, α represents the overall frequency of failures, and $f(t)$ describes the shape of the intensity function. For a collection of products, the expected number of failures at time t can be expressed as $\alpha_i f(t)$, where α_is are independently and identically distributed random variables with the same gamma distribution parametrised so that $\mathrm{E}(\alpha_i) = a/b$ and $\mathrm{Var}(\alpha_i) = a/b^2$.

Lawless et al. (2009) consider a mixed Poisson process model for repeated events based on age and usage scales. They accommodate heterogeneity with random effects Z_i for each i^{th} unit, where each Z_i is independently and identically distributed (iid) according to the same probability distribution $G(\cdot)$. They consider an intensity function conditioned on Z_i given by $\lambda(t|Z_i) = Z_i^{\beta}\lambda_0(tZ^{\beta};\gamma)$. They deal with over-dispersion, which is thought to be due to heterogeneity of users and the usage environment, by introducing a new iid random variable v_i with mean 1 and variance ϕ so that $\lambda(t|Z_i) = v_i Z_i^{\beta}\lambda_0(tZ^{\beta};\gamma)$, leading to a mixed Poisson process model.

Lawless and Crowder (2010) also deal with over-dispersion by introducing a random variable Z_i for each unit i, with mean 1 and variance ϕ, with the same gamma distribution for all i, $Ga(\phi, \phi^{-1})$.

Some authors use the Poisson model to estimate warranty claims but they do not consider the phenomenon of over-dispersion. For example, Karim et al. (2001) consider the application of NHPP to analyse automobile warranty data. They estimate the probability of failure of a unit at an age using the marginal counts data. Wang et al. (2002) estimate warranty claims based on claims given in terms of time to failure from the date of sale with no monthly sales information but with the number of total sales. Majeske (2007) proposes an NHPP model with a parametric component— time to first failure, for analysing automobile warranty data. He considers three subsystems for luxury cars with intensity rates of Weibull-Uniform, power law and linear hazard functions.

From the above literature, one can find that the over-dispersion in the warranty claims data can arise mainly due to the following two reasons. The first is the heterogeneity of products, or in other words, the differences in the intrinsic reliability of individual products. And the second is the heterogeneity of users as products used by different users can have different usage intensity and operating environments.

Although the phenomenon of the over-dispersion in warranty claims data has been addressed previously, its dynamic nature seems to have received little attention. In the present paper, we consider stochastic processes derived from the non-homogeneous Poisson process that can deal with dynamic over-dispersion.

3 Modelling warranty data

Consider a discrete-time Poisson process denoted as $\{N_t, t = 1, 2, ...\}$ with $N_0 = 0$. Let $M_t = N_t - N_{t-1}$ $(t = 1, 2, ...)$ represent the increments of the process at consecutive time periods of unit one, a month in this case. For a non-homogenous Poisson process, each M_t follows a Poisson distribution with mean μ_t, $M_t \sim \text{Poi}(\mu_t)$.

The over-dispersion exhibited in the data can be handled by assuming μ_t for each t to be a random variable. The resulting marginal expected value and the variance of M_t are given by:

$$\text{E}(M_t) = \text{E}(\mu_t) \quad \text{and} \quad \text{Var}(M_t) = \text{E}(\mu_t) + \text{Var}(\mu_t) \tag{1}$$

The variance of M_t is larger than its mean as long as $\text{Var}(\mu_t) > 0$. When $\text{Var}(\mu_t) = 0$ for all t, we have a conventional non-homogeneous Poisson process.

3.1 The non-homogeneous Poisson process

One of the most popular stochastic processes in reliability analysis is the non-homogeneous Poisson process (NHPP). It is often used to model the lifetime of products that are subject to minimal repair. A minimal repair assumes that the hazard rate of a failed item is restored to what it was just before the failure.

The increments of the NHPP are independent from each other. Let the intensity function of the NHPP be $\mu_t = S \int_{t-1}^{t} h(x)dx$, where S is the total number of products shipped out. The probability of observing n claims in any given month is given by:

$$P(M_t = n) = \frac{\mu_t^n e^{-\mu_t}}{n!}. \tag{2}$$

The mean of the NHPP is variable with time as opposed to a constant mean of the homogenous Poisson process. The expected value and the variance of M_t are equal, $\text{E}(M_t) = \text{Var}(M_t) = \mu_t$, that is, the variance to mean ratio is 1 for any given t.

3.2 *The mixed non-homogeneous Poisson process*

The mixed non-homogeneous Poisson processes (MNHPP) are often used to model the heterogeneity of the intrinsic reliability of the products and the heterogeneity of users.

The increments of the mixed Poisson process are not independent. Let the intensity function of the MNHPP be $\mu_t = \alpha S \int_{t-1}^{t} h(x)dx$, where $\alpha \sim \mathrm{Ga}(a, b)$ with $\mathrm{E}(\alpha) = a/b$ and $\mathrm{Var}(\alpha) = a/b^2$. The choice of the gamma distribution is justified by its flexibility and the resulting mathematical tractability. Then, the probability of observing n claims in any given month is given by (see Appendix A for derivation):

$$P(M_t = n) = \frac{\Gamma(a + \sum_{i=1}^{t-1} d_i + n)}{n!\Gamma(a + \sum_{i=1}^{t-1} d_i)} \times \frac{(b + S\int_0^{t-1} h(x)dx)^{a+\sum_{i=1}^{t-1} d_i} (S\int_{t-1}^{t} h(x)dx)^n}{(b + S\int_0^{t-1} h(x)dx + S\int_{t-1}^{t} h(x)dx)^{a+n+\sum_{i=1}^{t-1} d_i}} \tag{3}$$

The expected value of M_t is given by

$$\mathrm{E}(M_t) = \frac{a + \sum_{i=1}^{t-1} d_i}{b + S\int_0^{t-1} h(x)dx} S \int_{t-1}^{t} h(x)dx, \tag{4}$$

which can be derived based on $\mu_t = \alpha S \int_{t-1}^{t} h(x)dx$ as given above and the law of total expectation. And the variance is given by

$$\mathrm{Var}(M_t) = \frac{a + \sum_{i=1}^{t-1} d_i}{b + S\int_0^{t-1} h(x)dx} S \int_{t-1}^{t} h(x)dx + \frac{a + \sum_{i=1}^{t-1} d_i}{(b + S\int_0^{t-1} h(x)dx)^2} (S \int_{t-1}^{t} h(x)dx)^2, \tag{5}$$

which can be derived using Eq. (4) and the law of total variance.

We can therefore obtain the variance to mean ratio, which is given by

$$\frac{\mathrm{Var}(M_t)}{\mathrm{E}(M_t)} = 1 + \frac{S\int_{t-1}^{t} h(x)dx}{b + S\int_0^{t-1} h(x)dx}. \tag{6}$$

It is clear that the variance to mean ratio is larger than 1 for $b > 0$ and

$\int_{t-1}^{t} h(x)dx > 0$ for all t. Furthermore, this ratio is dynamic in time, and depends on $h(t)$.

3.3 The Cox process

The warranty claims of electronic products might be affected by external factors besides the intrinsic reliability of the products. For example, Wu (2011) finds that warranty claims are often related to the human behaviour such as product failures that are not reported as warranty claims (FBNR — failed but nor reported) and claims that might not be due to by product failure (RBNF — reported but not failed). Also, rapid technological developments in the electronics industry can lead to early obsolescence. Thus, in the later stages of the product life, product failures may not be reported. The over-dispersion can also be due to the heterogeneity of products and users. Since the usage intensity patterns can vary over time, it is reasonable to assume that the variance resulting from such heterogeneity be variable with time.

The increments of the Cox process are independent from each other. Let $\mu_t = \alpha_t S \int_{t-1}^{t} h(x)dx$, where $\alpha_t \sim \mathrm{Ga}(a_t, b_t)$ with $\mathrm{E}(\alpha) = a_t/b_t$ and $\mathrm{Var}(\alpha) = a_t/b_t^2$, then the probability of observing n claims in interval $(t-1, t]$ is given by:

$$P(M_t = n) = \frac{\Gamma(a_t + n)}{n!\Gamma(a_t)} \frac{b_t^{a_t} (S \int_{t-1}^{t} h(x)dx)^n}{(b_t + S \int_{t-1}^{t} h(x)dx)^{a_t+n}} \tag{7}$$

The expected value of M_t is given by

$$\mathrm{E}(M_t) = \frac{a_t}{b_t} S \int_{t-1}^{t} h(x)dx, \tag{8}$$

and the variance is given by

$$\mathrm{Var}(M_t) = \frac{a_t}{b_t} S \int_{t-1}^{t} h(x)dx + \frac{a_t}{b_t^2} (S \int_{t-1}^{t} h(x)dx)^2. \tag{9}$$

The variance to mean ratio is given by

$$\frac{\mathrm{Var}(M_t)}{\mathrm{E}(M_t)} = 1 + \frac{S \int_{t-1}^{t} h(x)dx}{b_t}. \tag{10}$$

It is clear that this ratio is dynamic and depends on both b_t and $h(x)$.

4 Prediction and prediction intervals

At a given time T, the aim of a warranty forecasting project is to forecast the number of warranty claims in the next K months. Let D_K denote the random variable which represents the number of warranty claims in the time interval $(T, T+K]$. Predictions can be obtained based on the expected values of D_K, $\mathrm{E}(D_K)$. The prediction intervals can be estimated by inverting the cdf (cumulative distribution function) of D_K, $F_{D_K}(x)$, at given set of cumulative probabilities. For example, 90% prediction intervals can span from the lower limit given by $L_{D_K} = F_{D_K}^{-1}(0.05)$ and the upper limit given by $U_{D_K} = F_{D_K}^{-1}(0.95)$.

The distribution of D_K is estimated based on the available data. Therefore, the uncertainty of D_K should also reflect the uncertainty of the parameter estimates. However, when the available data is large enough, we can assume that the uncertainty of the parameter estimates is negligible. In this paper, we assume that the uncertainty of the parameter estimates is negligible and that the estimated parameters are the "true" parameters.

In the case of NHPP, D_K is distributed according to a Poisson distribution with mean $S\int_T^{T+K} h(x)dx$. Therefore, the expected value of D_K is $\mathrm{E}(D_K) = S\int_T^{T+K} h(x)dx$. The 90% prediction interval can be determined by obtaining 5^{th} and 95^{th} percentiles of the Poisson distribution.

In the case of mixed NHPP, D_K is distributed according to a negative binomial distribution, which can be thought of as a gamma mixture of Poisson distributions. Its expected value is given by

$$\mathrm{E}(D_K) = \frac{a + \sum_{i=t}^{T} d_i}{b + S\int_0^T h(x)dx} S \int_T^{T+K} h(x)dx. \tag{11}$$

Prediction intervals can be estimated by obtaining the appropriate percentiles of the cdf, and its corresponding pdf (probability density function) is given by Eq. (17). This can be done by evaluating the cdf using Monte Carlo simulation. That is, $F_{D_K}(N_{T+K} - N_T = n) = \sum_{x=0}^{n} \int_0^\infty P(x|\mu) g_\mu(\mu) d\mu$ can be evaluated by generating random numbers from $g_\mu(\cdot)$, where $P(x|\mu)$ is the distribution function of the Poisson distribution, and $g_\mu(\mu)$ is the pdf of the gamma distribution.

In the case of the Cox process, the distribution of D_K is the convolution of several random variables. Its expected value is the sum of expected values of the compounding variables and given by:

$$\mathrm{E}(D_K) = S \sum_{t=T}^{T+K} \frac{a_t}{b_t} \int_{t-1}^{t} h(x)dx. \tag{12}$$

That is, $\mathrm{E}(D_K) = \sum_{i=T}^{T+K} \mathrm{E}(M_t)$. Since, $\mathrm{E}(M_t)$ is a gamma random variable for all t, $\mathrm{E}(D_K)$ is the sum of independent gamma random variables. The analytical form of the pdf of the sum of independent gamma random variables is given in (Sim, 1992). As the case with the mixed Poisson process, the quantiles of the $F_{D_K}(\cdot)$ can be found using Monte Carlo simulation.

5 Case studies

We consider five products from the same manufacturer. These products are electronics products with lifetime warranties. Upon a failure, the product is repaired (as a minimal repair), and then returned to the customer. The warranty claims have been aggregated on a monthly basis. We use data of the first 24 months for model-fitting. This choice is common, especially in the electronics industry, where companies are interested in being able to predict warranty claims after the first 2 years since the product launch. For comparing the prediction accuracy, we use data of the next 12 months.

We compare the forecasting performance of the following models:

- Non-homogeneous Poisson process (NHPP).
- Mixed non-homogeneous Poisson process (MPP) with $\mathrm{E}(\alpha) = 1$.
- Cox process (CP) with $\mathrm{E}(\alpha_t) = 1$ for all t, and $\mathrm{Var}(\alpha_t) = 1/c_t$, where $c_t = a_t = b_t$. For the models considered here we assume that $c_t = \gamma t$. The main reason for this is that we expect the variance of α_t to decrease with time, which leads to decreasing over-dispersion over time, as observed in the warranty claim dataset we have.

For products considered in this study, the claim rate increases in the first several months and then starts to drop off. As a result we have a unimodal curve for the claim rates. Such a curve can be modelled by several different functions including some probability density functions. However, after some tests on curve fitting, we have selected the model that has the hazard rate function of the inverse-Weibull distribution. This function is flexible and can readily be interpreted in the context of product failures.

For all of the above models, the $h(x)$ is chosen to be the hazard rate function

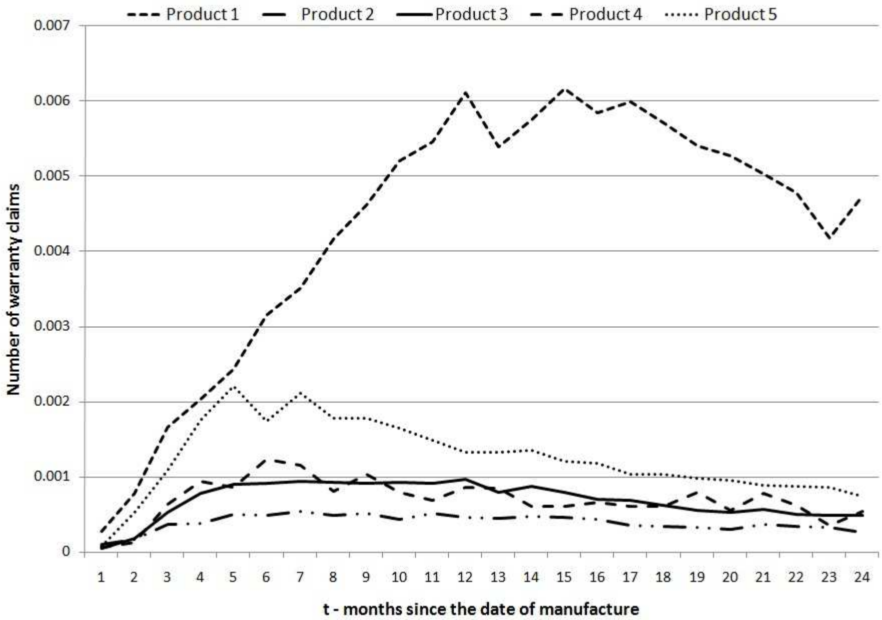

Fig. 4. Claim rates of the five electronics products

of the inverse Weibull distribution and given by:

$$h(t) = \gamma\beta^{\gamma}t^{-(\gamma+1)}e^{-(\frac{\beta}{t})^{\gamma}}(1 - e^{-(\frac{\beta}{t})^{\gamma}})^{-1}. \tag{13}$$

The claim rates for the products considered in this paper are shown in Figure 4. Thus, the NHPP model has two parameters, whereas the MPP and CP models have three parameters, respectively.

The log-likelihood functions for the above models are the logarithm of the corresponding likelihood functions given by $L = \prod_{t=1}^{T} P(M_t = d_t)$.

The model performance is assessed with both the log-likelihood value and the commonly used Akaike information criterion (AIC). The prediction accuracy is measured in terms of normalised rooted mean squared error (NRMSE).

$$NRMSE = \sqrt{\frac{(D_K^* - \mathrm{E}(D_K))^2}{D_K^{*2}}}, \tag{14}$$

where D_K^* is the observed number of claims in interval $(T, T+K]$.

5.1 *Results of model fitting*

Table 3 shows the estimated log-likelihood and the AIC for all of the five products modelled with NHPP, the mixed Poisson process (MPP) and the Cox process (CP). It can be seen that for the first three products the CP fits the data best as it has the smallest AIC. For the last two products, MPP has the smallest AIC. As it can be expected, NHPP does not fit the data so well as the models that take into account the dynamic over-dispersion.

Table 3
Log-likelihood and Akaike information criterion (AIC) estimated using non-homogeneous Poisson process (NHPP), mixed non-homogeneous Poisson process (MPP), and Cox process (CP). AIC* indicates the smallest AIC.

Product	NHPP		MPP		CP	
	lnL	AIC	lnL	AIC	lnL	AIC
1	143671.20	-287338.40	143721.04	-287436.07	143870.97	-287735.95*
2	10874.60	-21745.19	10891.93	-21777.85	10893.66	-21781.32*
3	64655.71	-129307.42	64928.56	-129851.12	64949.58	-129893.17*
4	2754.49	-5504.98	2773.45	-5540.91*	2772.57	-5539.14
5	27471.77	-54939.55	27671.23	-55336.46*	27665.66	-55325.31

5.2 *Results of prediction*

Table 4 shows the measures of the prediction accuracy using NRMSE for $K = 12$. It can be seen from the table that on average the CP model has the lowest NRMSE. Both the MPP and CP models perform better than the NHPP model for all products.

Table 4
Normalised rooted mean squared error (NRMSE).

Product	NHPP	MPP	CP
1	0.210	0.370	0.149
2	0.609	0.302	0.461
3	0.284	0.092	0.066
4	0.336	0.069	0.188
5	0.265	0.158	0.006
Average	**0.341**	**0.198**	**0.174**

6 Discussion

Warranty claims data can often be affected by factors that are not related to the intrinsic reliability of the products. This can be due to customer behaviours towards warranty claims, levels of expertise of technicians that deal with warranty claims, or market and environmental conditions. Also, product reliability itself can vary across production batches if small changes are incorporated into product design. All of these factors coming together can result in over-dispersion in warranty claims data. In this paper, we have considered the warranty claim forecasting problem for warranty data of electronics products that exhibit over-dispersion. These warranty claim data have shown that the over-dispersion is dynamic and changes over time. Some products can clearly exhibit a trend in the over-dispersion, which can be detected by estimating the variance to mean ratio over different time periods. We have presented models that tackle the dynamic over-dispersion. These models, in general, fit the data better than the conventional non-homogeneous Poisson process models and can result in better prediction results.

The Cox process models offer a certain degree of flexibility in modelling the dynamic over-dispersion as both the shape and the scale parameters (a_t and b_t) of the mixing distribution are time dependant. More research needs to be done to investigate different formulations of these parameters and α_t.

It is also possible to let the over-dispersion itself be a random variable, for example, by assuming the variance of α or α_t be a random variable. This, however, requires more computational effort as the probability distributions involved become intractable.

7 Conclusions and future work

We can draw the following conclusions from this study.

- Over-dispersion in warranty data can often have a dynamic nature with a possible trend.
- The over-dispersed data can be modelled with both mixed Poisson processes and Cox processes. The Cox processes offer more flexibility and allow to set a certain functional structure on the dynamic over-dispersion.
- The case study shows that models specifically tailored for dealing with over-dispersion fit the data better and have better prediction accuracy than the models based on NHPP.

In our future studies we will consider the forecasting performance of Poisson processes against non-parametric methods such as neural networks and support vector regression.

As a further avenue of investigation for formulating the warranty forecasting problem as univariate time series resulting from a count distribution that displays conditional heteroscedasticity (i.e. the dynamic over-dispersion pattern) in the residual pattern. There has been some work in this respect. For example, Cameron and Trivedi (1998) provide a treatment of INARMA(integer-valued autoregressive moving average) processes that can be extended to include a GARCH (autoregressive conditional heteroscedasticity) error structure. To account for the 'vintages' provided by the production batches they extend this framework to multivariate series. Zhu (2011) adapts the integer-valued time-scale model to account for over-dispersion and volatility. We would like to investigate these modelling techniques in our future work.

Acknowledgement

We are grateful to the anonymous referees for their valuable comments and suggestions that led to improvement of this paper. This research is supported by Engineering and Physical Sciences Research Council (EPSRC) of the United Kingdom (EPSRC Grant reference: EP/G039674/1).

References

Bishop, Y., Fienberg, S., Holland, P., 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Massachusets.

Cameron, C., and Trivedi, P.K., 1998. Regression Analysis of Count Data, Econometric Society Monograph No.30, Cambridge University Press.

Fahrmeir, L., Echavarria, O. L., 2006. Structured additive regression for overdispersed and zero-inflated count data. Applied Stochastic Models in Business and Industry 22, 351–369.

Fredette, M., Lawless, J. F., Feb. 2007. Finite-Horizon Prediction of Recurrent Events, With Application to Forecasts of Warranty Claims. Technometrics 49 (1), 66–80.

Hrycej, T., Ag, D., Grabert, M., 2007. Warranty Cost Forecast Based on Car Failure Data. In: Proceedings of International Joint Conference on Neural Networks. pp. 12–17.

Kalbfleisch, J. D., Lawless, J. F., 1996. Statistical analysis of warranty claims data. In: Blishke, W. R., Murthy, D. N. P. (Eds.), Product Warranty Handbook. M. Dekker, New York, pp. 231–259.

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

Karim, R., Yamamoto, W., Suzuki, K., 2001. Statistical analysis of marginal count failure data. Lifetime Data Analysis 7, 173–186.

Kennedy, W. J., Patterson, J. W., Fredendall, L. D., 2002. An overview of recent literature on spare parts inventories. International Journal of Production Economics 76, 201–215.

Lawless, J. F., Apr. 1998. Statistical Analysis of Product Warranty Data. International Statistical Review / Revue Internationale de Statistique 66 (1), 41.

Lawless, J. F., Crowder, M. J., Oct. 2010. Models and estimation for systems with recurrent events and usage processes. Lifetime data analysis 16 (4), 547–70.

Lawless, J. F., Crowder, M. J., Lee, K. A., 2009. Analysis of reliability and warranty claims in products with age and usage scales. Technometrics 51, 14–24.

Lawless, J. F., Kalbfleisch, J. D., 1992. Some issues in the collection and analysis of field reliability data. Kluwer, pp. 141–152.

Lindsey, M., Pavur, R., 2009. Prediction intervals for future demand of existing products with an observed demand of zero. International Journal of Production Economics 119, 75–89.

Majeske, K., Feb. 2007. A non-homogeneous Poisson process predictive model for automobile warranty claims. Reliability Engineering and System Safety 92 (2), 243–251.

Murthy, D. N., Djamaludin, I., 2002. New product warranty: A literature review. International Journal Of Production Economics 79 (3), 231–260.

Rai, B., Singh, N., Jun. 2005. Forecasting warranty performance in the presence of the 'maturing data' phenomenon. International Journal of Systems Science 36 (7), 381–394.

Sim, C., 1992. Point processes with correlated gamma interarrival times. Statistics and Probability Letters 15, 135–141.

Stephens, D., Crowder, M., Jan. 2004. Bayesian analysis of discrete time warranty data. Journal of the Royal Statistical Society: Series C (Applied Statistics) 53 (1), 195–217.

Syntetos, A. A., Babai, M. Z., Davies, J., Stephenson, D., 2010. Forecasting and stock control: A study in wholesaling context. International Journal of Production Economics 127, 103–111.

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

Wasserman, G. S., Sudjianto, A., 1996. A comparison of three strategies for forecasting warranty claims. IIE Transactions (Institute of Industrial Engineers) 28 (12), 967–977.

Wu, S., 2011. Warranty claim analysis considering human factors. Reliability Engineering and System Safety 96, 131–138.

Wu, S., Akbarov, A., Mar. 2011. Support vector regression for warranty claim forecasting. European Journal of Operational Research.
Yun, W. Y., Murthy, D. N. P., Jack, N., 2008. Warranty servicing with imperfect repair. International Journal of Production Economics 111, 159–169.
Zhu, F., 2011. A negative binomial integer-valued GARCH model (2011) Journal of Time Series Analysis, 32(1), 54-67.

Appendix A

Consider a Poisson process with $\mu(t) = \alpha H(t)$, where $\alpha \sim \text{Ga}(a, b)$ such that $\text{E}(\alpha) = a/b$ and $H(t)(= S \int_{t-1}^{t} h(x)dx)$ is the expected number of events in the interval $[0, t]$. The probability of observing n number of events in the interval $[0, t]$ is given by:

$$\begin{aligned} P(N(t) = n) &= \int_0^\infty \frac{(\alpha H(t))^n e^{-\alpha H(t)}}{n!} \frac{b^a}{\Gamma(a)} \alpha^{a-1} e^{-b\alpha} d\alpha \\ &= \frac{\Gamma(n+a)}{n!\Gamma(a)} \frac{b^a H(t)^n}{(b+H(t))^{a+n}} \end{aligned} \tag{15}$$

Using the Bayes' theorem we can derive the probability of observing n events in $(t, t+\Delta t]$ given $N(t) = N$. $g(\alpha|N(t) = N)$ is proportional to:

$$\begin{aligned} g(\alpha|N(t) = N) &\propto \frac{(\alpha H(t))^N e^{-\alpha H(t)}}{N!} \frac{b^a}{\Gamma(a)} \alpha^{a-1} e^{-b\alpha} \\ &= \frac{H(t)^N b^a}{N!\Gamma(a)} \alpha^{a+N-1} e^{-\alpha(b+H(t))} \\ &\propto \frac{(b+H(t))^{(a+N)}}{\Gamma(a+N)} \alpha^{(a+N-1)} e^{-(b+H(t))\alpha} \end{aligned} \tag{16}$$

which is a gamma distribution, $\text{Ga}(a+N, b+H(t))$, with $\text{E} = (a+N)/(b+H(t))$.

Thus, the probability of observing n claims in $(t, t+\Delta t]$ is given by

$$\begin{aligned} &P(N(t+\Delta t) - N(t) = n|N(t) = N) \\ &\quad = \int_0^\infty \frac{(\alpha(H(t+\Delta t) - H(t)))^n e^{-\alpha(H(t+\Delta t)-H(t))}}{n!} g(\alpha|N(t) = N) \end{aligned}$$

$$= \frac{\Gamma(a+N+n)}{n!\Gamma(a+N)} \frac{(b+H(t))^{a+N}(H(t+\Delta t)-H(t))^n}{(b+H(t+\Delta t))^{a+N+n}} \tag{17}$$

Since the available data are recorded on a monthly basis, we have $\Delta t = 1$. Let $M_t = N(t) - N(t-1)$, for $t = 1, 2...$, and $h(x)$ be the intensity function. The expected number of events in interval $[0, t]$ is given by $S \int_0^t h(x)dx$, thus:

$$\begin{aligned} P(\, M_t = n) = & \frac{\Gamma(a + \sum_{i=1}^{t-1} d_i + n)}{n!\Gamma(a + \sum_{i=1}^{t-1} d_i)} \times \\ & \times \frac{(b + S\int_0^{t-1} h(x)dx)^{a+\sum_{i=1}^{t-1} d_t} (S\int_{t-1}^{t} h(x)dx)^n}{(b + S\int_0^{t-1} h(x)dx + S\int_{t-1}^{t} h(x)dx)^{a+\sum_{i=1}^{t-1} d_i + n}} \end{aligned} \tag{18}$$