WARRANTY CLAIMS MODELLING

V. KULKARNI AND SIDNEY I. RESNICK

ABSTRACT. A company wishes to estimate or predict its financial exposure in a reporting period of length T (typically one quarter) due to warranty claims. We propose a fairly general random measure model which allows computation of the Laplace transform of the total claim made against the company in the reporting interval due to warranty claims. When specialized to a Poisson process of both sales and warranty claims, statistical estimation of relevant quantities is possible. The methodology is illustrated by analyzing automobile sales and warranty claims data from a large car manufacturer for a single car model and model year.

1. INTRODUCTION

A retail company needs to budget for warranty claims as part of their risk management policies. Excessive warranty claims indicate problems in manufacturing or procurement so distributional properties of the claims are needed. Financial requirements for predicting and then reporting quarterly results make it desirable for a company to be able to predict total quarterly warranty claims.

We analyze this problem using a random measure and clustered point process approach. A sale of an item generates a random measure $D(\cdot)$ in which $D((0,t])$ represents the total warranty claim experienced by the company for the sold item in the first t time units subsequent to the sale. Times of sales are treated as a point process with only simple points. Specialization of both the sale times and times of warranty claims relative to sales to non-homogeneous Poisson processes allows estimation of model parameters and, hence, estimation of the Laplace transform of the distribution of total claim exposure to the company in, say, a quarter. To guard against model risk arising from the Poisson assumptions, other tractable statistical models should be developed. See Ja, et al (2002) for earlier work on warranty costs arising out of non-stationary sales processes.

Similar models arise in other disciplines. For instance, in Internet modeling the times of user initiated connections are followed by a cluster of machine initiated connections giving rise to cluster point processes. Fitting realistic models to data in the face of a variety of statistical structures for the clusters is an important issue. In particular, cluster characteristics may or may not be dependent on cluster size and gaps between points in a cluster may or may not be independent.

Clearly there is a connection between warranty claims modeling developed here and insurance claims modeling. Instead of giving references to this vast literature, we refer the readers to two excellent books on this topic: Asmussen (2000), Rolski et al (1999). Insurance models typically do not account for non-stationary policy-issuing processes, which is what makes our models different.

Our paper is organized as follows: Section 2 outlines the general sales and claims model. Section 3 offers a theoretical formula for the total warranty claim against the company in a time period of $[0, T]$. This formula is specialized to the case where both times of sales and times of warranty claims follow non-homogeneous Poisson processes. This specialization yields a sufficiently explicit formula for the Laplace transform of the distribution of total warranty claims in a reporting interval; the Laplace transform can be numerically inverted to obtain, for instance, quantiles of the distribution of total claims in a reporting interval. Subsequent sections analyze two data sets of *sales dates* and *times and costs* of warranty claims

from a major car manufacturer for a single car model and model year. We show how model parameters may be estimated and use these estimates to compute the quantiles of the total warranty costs in each quarter for a this manufacturer. The last section contains the summary and conclusions.

The security protocols of the manufacturing company that supplied us with data required that the sales and warranty claim data be subjected to masking, scaling and some random deletions. Thus we view the methodology developed in this paper as a demonstration of feasibility. More definitive conclusions would require working closely with a manufacturer or retailer and obtaining more complete data records.

2. The sales and claims model

This section outlines the model.

2.1. **Sales.** Suppose there is a non-homogeneous Poisson process of sales. The Poisson counting process

$$\sum_j \epsilon_{S_j}(\cdot) \tag{2.1}$$

lives on $\mathbb{R}$ and has sales times $\{S_j\}$. The standard notation $\sum_j \epsilon_{S_j}$ for the counting function of the points $\{S_j\}$ is defined by

$$\epsilon_S(I) = \begin{cases} 1, & \text{if } S \in I, \\ 0, & \text{if } S \notin I, \end{cases}$$

for an interval I. The mean number of sales in an interval I is

$$\mu(I) = E\Big(\sum_j \epsilon_{S_j}(I)\Big),$$

where μ is a measure on $\mathbb{R}$ which is finite on finite intervals. The Poisson assumption means that the Laplace functional is of the form (Kallenberg (1983), Neveu (1977), Resnick (2006, 1987))

$$E(\exp\{-\sum_j f(S_j)\}) = \exp\Big\{-\int_{\mathbb{R}}(1 - e^{-f})d\mu\Big\} \tag{2.2}$$

for any non-negative function f defined on $\mathbb{R}$.

2.2. **Warranty claims.** For the jth sold item, we assume a general process of warranty claims. For the jth item sold, suppose there exist times $\{S_j + T_i^{(j)}, i \geq 1\}$ at which warranty claims are made. The time points $\{T_i^{(j)}\}$ are non-negative and non-decreasing in i. Assume there are finitely many points $\{S_j + T_i^{(j)}, i \geq 1, -\infty < j < \infty\}$ in any finite interval. There is an iid sequence of claim sizes $\{D_i^{(j)}, i \geq 1\}$ where we assume $D_i^{(j)}$ is the ith warranty claim for the jth sold item. For the jth item, there is a random measure $D_j(\cdot)$ on $[0, \infty)$ giving the total claims:

$$D_j(\cdot) = \sum_i D_i^{(j)} \epsilon_{T_i^{(j)}}(\cdot), \tag{2.3}$$

so that $D_j(I)$ is the total warranty claim amount during time interval I for the jth item sold. Later we will assume a standard warranty is of length W and that each of the iid random measures $D_j(\cdot)$ concentrates on $[0, W]$. The random measure giving total exposure to warranty claims in a time period I is

$$C(I) = \sum_j \sum_i D_i^{(j)} \epsilon_{T_i^{(j)}+S_j}(I) \tag{2.4}$$

so that the total exposure in a reporting interval $[0, T]$ (eg, 1 quarter or 1 year) is

$$C([0,T]) = \sum_j \sum_i D_i^{(j)} \epsilon_{T_i^{(j)}+S_j}([0,T]) = \sum_j D_j\big([(-S_j)^+, T - S_j]\big),$$

the aggregate across items sold of claims incurred in $[0, T]$.

3. The distribution of total claims.

3.1. **General warranty claims process.** Suppose $f : [0, \infty) \mapsto [0, \infty)$ is a test function; later it will be the indicator of an interval. The random measures $\{D_j\}$ given in (2.3) are iid. Define the Laplace functional of the random measure $D_1(\cdot)$ as

$$(3.1) \qquad \psi_D(f) = Ee^{-D_1(f)}.$$

As f varies, this determines the distribution of the random measure $D_1(\cdot)$. The Laplace functional of the total warranty claims random measure is

$$
\begin{aligned}
Ee^{-C(f)} &= Ee^{-\int_0^\infty f(s)C(ds)} = Ee^{-\sum_j \int_{s \geq S_j} f(s)D_j(ds - S_j)} \\
&= Ee^{-\sum_j \int_0^\infty f(y + S_j)D_j(dy)}.
\end{aligned}
$$

Now we condition on $\{S_j\}$. Let the conditional expectation with respect to the σ-field generated by $\{S_j\}$ be $E^{\{S_j\}}$. We get the above equal to

$$
= E\Big(E^{\{S_j\}}\big(e^{-\sum_j \int_0^\infty f(y + S_j)D_j(dy)}\big)\Big) = E\Big(\prod_j E^{\{S_j\}}\big(e^{-\int_0^\infty f(y + S_j)D_j(dy)}\big)\Big).
$$

Write $f_s(\cdot) = f(\cdot + s)$ and using (3.1), we recognize the forgoing as

$$
= E\Big(\prod_j \psi_D(f_{S_j})\Big) = Ee^{-\sum_j -\log \psi_D(f_{S_j})}.
$$

This being a function of Poisson points allows us to use the form in (2.2) and to write the above equal to

$$
\begin{aligned}
&= \exp\{-\int_{\mathbb{R}} \Big(1 - e^{-(-\log \psi_D(f_s))}\Big)\mu(ds)\} = \exp\{-\int_{\mathbb{R}} \Big(1 - \psi_D(f_s)\Big)\mu(ds)\} \\
&= \exp\{-\int_{\mathbb{R}} \Big(1 - \psi_D(f(\cdot + s))\Big)\mu(ds)\}.
\end{aligned}
$$

We summarize the above discussion in the following proposition.

Proposition 1. *Suppose $\{S_j\}$ are Poisson points with mean measure μ given by (2.1) and Laplace functional (2.2). Suppose $\{D_j(\cdot)\}$ are iid random measures representing cumulative warranty claim amounts as given by (2.3) with distribution determined by the Laplace functional (3.1). Then the total warranty claim random measure $C(\cdot)$ defined in (2.4) has Laplace functional*

$$(3.2) \qquad \psi_C(f) := \exp\{-\int_{\mathbb{R}} \Big(1 - \psi_D(f(\cdot + s))\Big)\mu(ds)\},$$

for $f \geq 0$.

3.2. **Specialization.** Setting $f = \zeta 1_{[0T]}$ for $\zeta > 0$ transforms

$$
\psi_C(f) = E(e^{-\zeta C[0,T]}), \quad \zeta > 0,
$$

into the Laplace transform of the random variable $C[0, T]$. We now see how explicit formula (3.2) can be made in certain special cases.

We assume the warranty claim sizes are iid random variables $\{D_i^{(j)}, i \geq 1, j \geq 1\}$ and suppose

$$(3.3) \qquad P[D_1^{(1)} \leq x], \quad x \geq 0$$

is the distribution with Laplace transform

$$
\hat{D}(\zeta) = E(e^{-\zeta D_1^{(1)}}), \quad \zeta > 0.
$$

Then the Laplace functional of the random measure $D_1(\cdot)$ given in (3.1) can be expressed as

$$
\psi_D(f) = E(e^{-D_1(f)}) = E(e^{-\sum_i D_i^{(1)} f(T_i^{(1)})})
$$

and conditioning on $\{T_i^{(1)}\}$ this is

$$=E\left(E^{\{T_i^{(1)}\}}\left(e^{-\sum_i D_i^{(1)} f(T_i^{(1)})}\right)\right)$$

$$=E\prod_i \hat{D}(f(T_i^{(1)})) = E\left(e^{-\sum_i -\log \hat{D}(f(T_i^{(1)}))}\right)$$

(3.4) $\psi_D(f) = \psi_T(-\log \hat{D} \circ f),$

where

$$\psi_T(g) = E(e^{-\sum_i g(T_i^{(1)})}), \quad g \geq 0,$$

is the Laplace functional of the point process with points $\{T_i^{(1)}\}$. Thus the Laplace functional of the random measure $D_1(\cdot)$ can be represented in terms of the Laplace functional of $\{T_i^{(1)}\}$ and the Laplace transform of the distribution $D(x)$.

Next we suppose $\{T_i^{(1)}\}$ are the points of a non-homogeneous Poisson process on $[0,\infty)$ and slightly more generally we suppose these are the points of a Poisson random measure with mean measure ν, a measure giving finite mass to bounded sets. We assume the warranty duration is a fixed number $W > 0$, longer than the length of the financial reporting interval T and so we assume ν concentrates on $[0,W]$ which amounts to considering the point process of claims to be $\sum_i \epsilon_{T_i^{(1)}}(\cdot \cap [0,W])$. (Variants where W is random with small variance about the mean could also be considered but we leave this for elsewhere.) Then for $g \geq 0$, the Poisson assumption means the Laplace functional of $\sum_i \epsilon_{T_i^{(1)}}(\cdot \cap [0,W])$ is

(3.5) $\psi_T(g) = e^{-\int_0^W (1-e^{-g})d\nu}.$

See Kallenberg (1983), Neveu (1977), Resnick (2006, 1987). With this Poisson assumption we have from (3.4),

$$\psi_D(f) = \psi_T(-\log \hat{D} \circ f) = \exp\{-\int_0^W \left(1 - e^{-(-\log \hat{D}\circ f(s))}\right)\nu(ds)\}$$

(3.6) $= e^{-\int_0^W \left(1 - \hat{D}\circ f(s)\right)\nu(ds)}.$

So assuming $\{T_i^{(1)}\}$ are Poisson points as well as $\{S_j\}$ being Poisson, and combining (3.2), (3.4) and (3.6), we get

$$\psi_C(f) = \exp\{-\int_{\mathbb{R}} 1 - \psi_D(f(s + \cdot))\mu(ds)\}$$

(3.7) $= \exp\{-\int_{\mathbb{R}} \left(1 - e^{-\int_0^W \left((1-\hat{D}(f(s+u)))\nu(du)\right)}\right)\mu(ds)\}$

To get the Laplace transform of $C[0,T]$, we set $f = \zeta 1_{[0,T]}$ for $\zeta > 0$ and get

$$Ee^{-\zeta C[0,T]} = \exp\{-\int_{\mathbb{R}} \left(1 - e^{-\int_0^W \left(1-\hat{D}(\zeta 1_{[0,T]}(s+u))\right)\nu(du)}\right)\mu(ds)\}$$

$$= \exp\{-\int_{\mathbb{R}} \left(1 - e^{-\left(1-\hat{D}(\zeta)\right)\int_0^W 1_{[0,T]}(s+u)\nu(du)}\right)\mu(ds)\}.$$

To analyze this further, suppose $W > T$, that is, the warranty period is larger than the reporting period. Decompose the integral over $\mathbb{R}$ as

$$\int_{\mathbb{R}} = \int_{s=0}^T + \int_{s<0} = A + B.$$

We find the term A is a standard convolution

(3.8) $A = \int_{s=0}^T \left(1 - e^{-(1-\hat{D}(\zeta))\nu([0,T-s])}\right)\mu(ds).$

For B, we get that $s < 0$ and $|s| < u \leq T + |s|$ and also $u < W$ and therefore,

$$B = \int_{s<0} \left(1 - e^{-(1-\hat{D}(\zeta)) \int_0^W 1_{[0,T]}(s+u)\nu(du)}\right) \mu(ds)$$

$$= \int_{s<0} \left(1 - e^{-(1-\hat{D}(\zeta))\nu(|s|,(T+|s|)\wedge W]}\right) \mu(ds)$$

$$(3.9) \qquad = \int_{\substack{s<0 \\ 0<|s|<W-T}} \left(1 - e^{-(1-\hat{D}(\zeta))\nu(|s|,(T+|s|)]}\right) \mu(ds) + \int_{\substack{s<0 \\ W-T<|s|<W}} \left(1 - e^{-(1-\hat{D}(\zeta))\nu(|s|,W]}\right) \mu(ds).$$

To summarize, we find that when both sales and claims follow non-homogeneous Poisson processes, we get an explicit formula for the Laplace transform of total warranty claims in a reporting period:

$$(3.10) \qquad\qquad Ee^{-\zeta C[0,T]} = e^{-(A+B)},$$

where A is given by (3.8) and B is given by (3.9).

If we specialize further and let $\sum_i \epsilon_{T_i^{(1)}}$ be homogeneous Poisson on $\mathbb{R}$ with $\nu(dt) = \lambda dt$ and $\sum_i \epsilon_{S_i}$ be homogeneous Poisson with $\mu(ds) = \theta ds$ so that the rates are $\lambda > 0$ and $\theta > 0$, we get

$$A = \int_{s=0}^T \left(1 - e^{-(1-\hat{D}(\zeta))\lambda(T-s)}\right)\theta ds$$

$$= \theta T - \frac{\theta(1 - e^{-(1-\hat{D}(\zeta))\lambda T})}{(1-\hat{D}(\zeta))\lambda}.$$

For B we get after transforming $s \mapsto -s$,

$$B = \int_{s=0}^{W-T} \left(1 - e^{-(1-\hat{D}(\zeta))\lambda T}\right)\theta ds + \int_{s=W-T}^W \left(1 - e^{-(1-\hat{D}(\zeta))\lambda(W-s)}\right)\theta ds$$

$$= \theta(W-T)(1 - e^{-(1-\hat{D}(\zeta))\lambda T}) + \int_0^T \left(1 - e^{-(1-\hat{D}(\zeta))\lambda s}\right)\theta ds$$

$$= \theta(W-T)(1 - e^{-(1-\hat{D}(\zeta))\lambda T}) + \theta T - \frac{\theta(1 - e^{-(1-\hat{D}(\zeta))T\lambda})}{\lambda(1-\hat{D}(\zeta))}.$$

4. Analysis of the claim size data

In an attempt to analyze the distribution of claim sizes $\{D_i^{(1)}\}$ and also to understand the process of sales $\{S_j\}$, we obtained and analyzed a year's worth of claims and sales data for a specific car model from a major car manufacturer. The claims data for model year 2000 comprises a spreadsheet called *cost2000* of length 73,167. The analysis for the fourth column *cost*, representing the cost to the company of each warranty claim, allows us to construct an estimate of the claim cost distribution $P[D_1^{(1)} \leq x]$ defined in (3.3). The usual methods from the statistics of extremes and quantitative risk management construct a distribution model using the empirical distribution for the center of the distribution and a generalized Pareto distribution for the part of the distribution above a threshold. See Coles (2001), Embrechts et al. (1997), McNeil et al. (2005).

4.1. **Initial analysis of warranty claim costs.** The time series plot for cost sizes shows a spikey structure without apparent trends. The autocorrelation plot taken out to 30 lags shows little dependence though there may be dependence at very small lags of 1 or 2 which may result from multiple claims for the same vehicle on the same day.

Initial analysis of the cost data summarized in Table 4.1 shows that the data has a wide range. Twenty four values in the data set are equal to 0. The units are unknown resulting from masking and scaling of the data by the manufacturer. Presumably there is a rational choice of upper bound based on the total value of a car. Despite the upper bound, it is still reasonable to model the cost data with a heavy tailed distribution based on the frequency of extreme values and the fact that there is a mix of vehicles with differing total
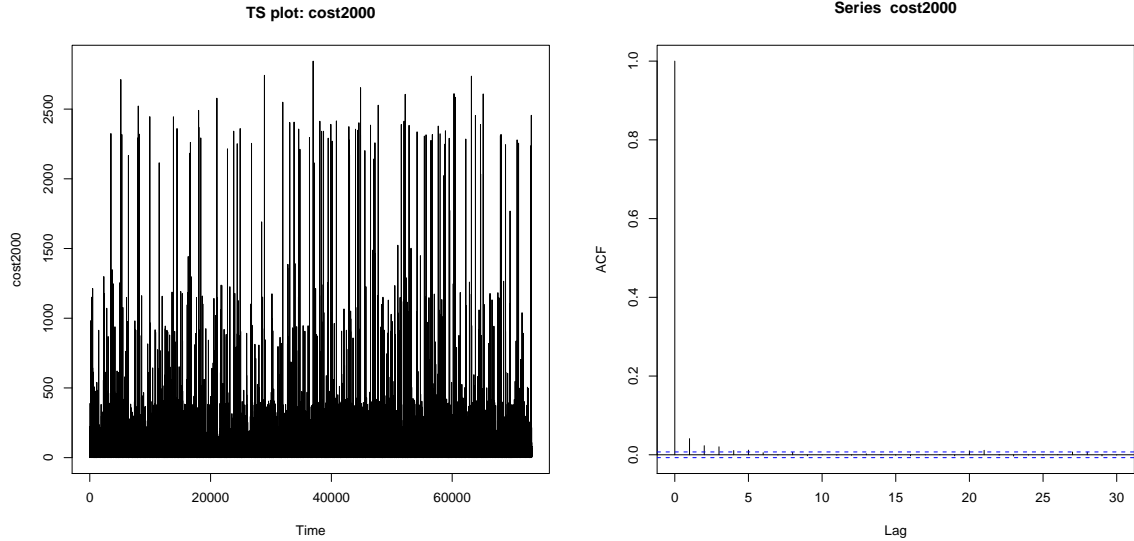
FIGURE 1. Time series plot of claims data (left) and autocorrelation plot (right) showing little dependence.

replacement value. An alternative which could be explored would be to fit a truncated Pareto distribution to threshold exceedances. This would produce similar statistical conclusions.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	6.573	11.650	35.610	31.830	2844.000

TABLE 1. Summary statistics for the data.

4.2. **Heavy tail analysis.** As a first step we made a QQ plot of the log transformed data against exponential quantiles. The upper 10,000 largest values follow a line whose slope indicates a value of $\alpha = 1.51$ for the model

$$P[\text{Cost} > x] \approx x^{-\alpha},$$

for large x. However, examination of the plot indicates that the 55 largest upper order statistics follow a line with slope which estimates $\alpha = 16.376$. Considering that 55 is a small percentage of 73167 or 10,000, it is reasonable to continue with the heavy tail analysis. Alternatively, one could consider a mixture model of two Pareto distributions but this may result in overfitting.

As an alternative exploratory procedure we made a Hill plot (Csörgő et al. (1985), Hill (1975), Mason and Turova (1994), Mason (1982), Resnick (2006, 1997)) of the cost data which plots the Hill estimator $H_{k,n}^{-1}$ of α as a function of k, the number of upper order statistics used in the estimation; the parameter n is the sample size. The plots in Figure 3 are quite stable. The altHill plot (Resnick and Stărică (1997), Resnick (2006)) replaces k by $[n^\theta]$ for $0.4 \leq \theta \leq 1$.

Next we fit a generalized Pareto distribution to the tail (see Coles (2001), Embrechts et al. (1997), McNeil et al. (2005)) following the standard peaks over threshold philosophy described as follows. For a threshold u and $x > u$ write

$$P[\text{Cost} > x] = P[\text{Cost} > x|\text{Cost} > u]P[\text{Cost} > u]$$
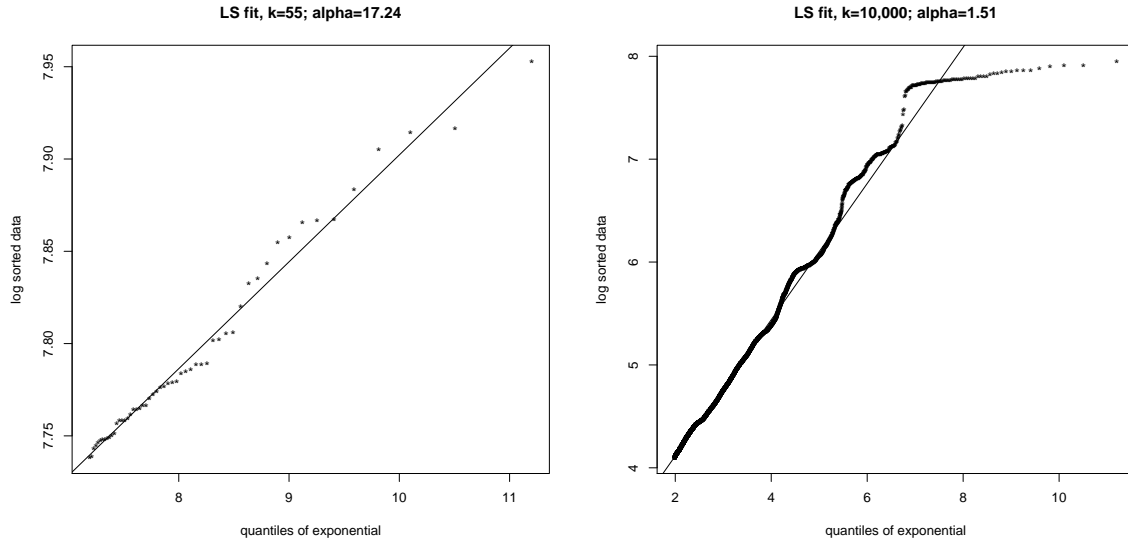$$= \bar{F}(u)\bar{F}^{[u]}(x),$$

FIGURE 2. QQ plots for the log transformed data. The left plot considers only the 55 upper order statistics while the right plot uses 10,000 upper order statistics.
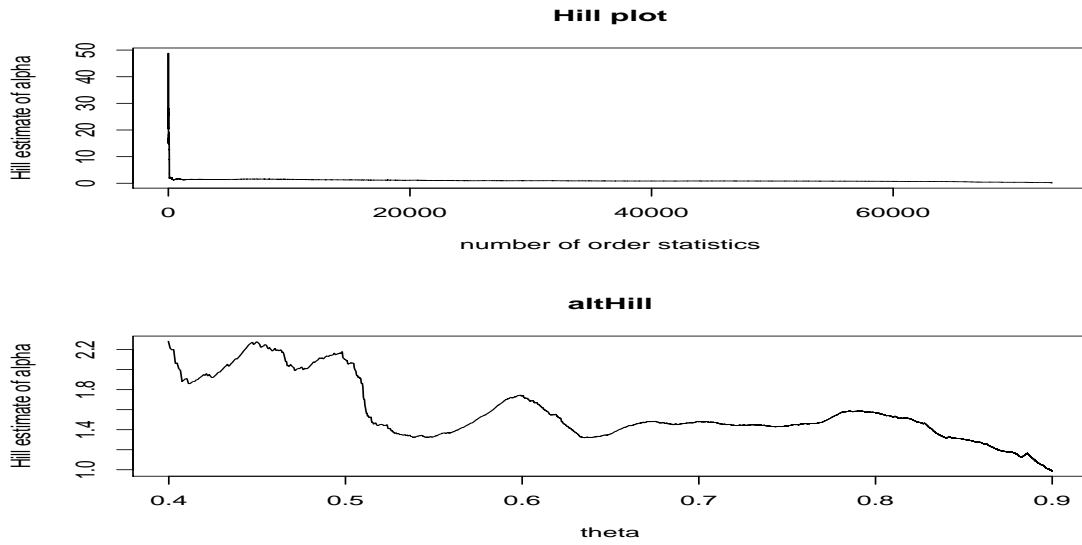

FIGURE 3. Hill plots for cost2000. The top plot plots estimates of α based on k, the number of upper order statistics used in the estimation while the bottom plot is in alt scale and replaces k by $[n^\theta]; 0.4 \le \theta \le 1$ and n is the sample size.

where $F(x)$ is the claim size distribution and $F^{[u]}(x)$ is the conditional distribution of claim cost given the claim is bigger than u; this is also the exceedance distribution. The exceedance distribution is approximated by the generalized Pareto distribution, effectively allowing extrapolation beyond the range of the data, yielding

$$(4.1) \qquad \approx \bar{F}(u)(1 + \frac{1}{\alpha\beta}(x - u))^{-\alpha},$$

for a scale parameter $\beta > 0$ and the Pareto shape parameter $\alpha > 0$. This means our estimate of the claim size distribution is

$$\hat{P}[\text{Cost } > x] = \begin{cases} \widehat{\overline{F}(x)}, & \text{if } x \leq u, \\ \widehat{\overline{F}(u)}(1 + \frac{1}{\hat{\alpha}\hat{\beta}}(x - u))^{-\hat{\alpha}}, & \text{if } x > u. \end{cases}$$

For many purposes, an adequate estimate, $\widehat{\overline{F}(x)}$, for $x \leq u$ is provided by the empirical distribution (# observations $> x$)/(sample size); however, relying on the empirical distribution may present storage problems when doing numerical Laplace transform inversion.

For the claim size data, we used the 10,000 upper order statistics, which corresponds to a threshold of $u = 60.262$, with 86% of the observations less than this threshold. We obtained by the method of maximum likelihood using the R-module *EVIR*, that $\hat{\alpha} = 1.54$ (as opposed to the QQ estimate of 1.51) and $\hat{\beta} = 41.4537$. The fitted excess distribution overlaid with the empirical distribution of excesses of the threshold is shown in Figure 4.
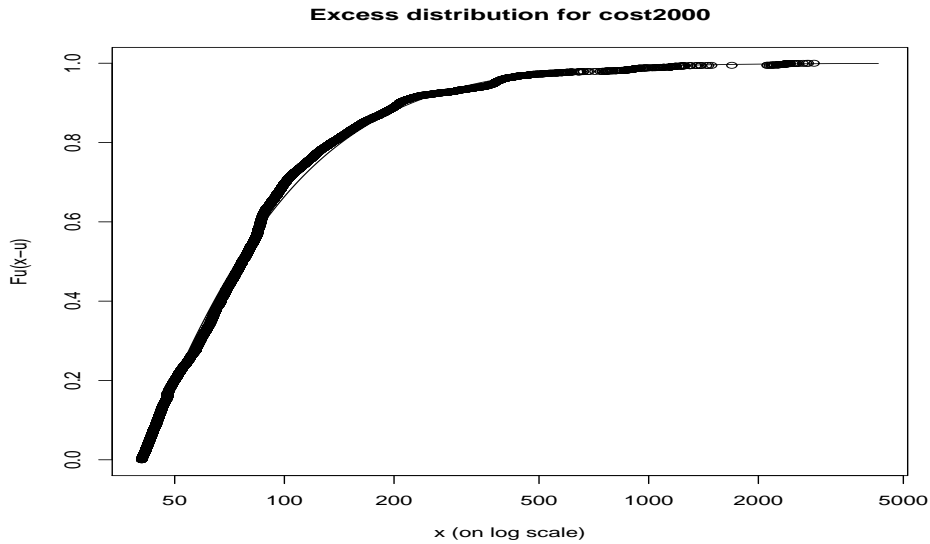
Excess distribution for cost2000

FIGURE 4. GPD fitted distribution of excesses compared with empirical distribution of excesses relative to the threshold 60.262.

We computed some some standard risk measures including quantiles of the fitted distribution and the *expected shortfall* (sfall) for each quantile; that is the expected cost given that the cost is greater than the quantile level. For example, from Table 2 we see that

$$P[\text{Cost } > 343.952] = 0.01$$

and the expected *excess* over 343.952 is 980.482-343.952=636.530.

For $\bar{F}(x), x < u$, in some circumstances, it is preferable to fit a parametric family rather than use the empirical distribution. We call the subset of the claims data corresponding to claim sizes less than the threshold $u = 62.262$ *claims2000small*. We fit a Gamma density

$$f_{a,s}(x) = \frac{1}{\Gamma(a)} s^{-a} x^{a-1} e^{-x/s}, \quad x > 0$$

with shape paramter a and scale parameter s. Using maximum likelihood via the R-function *fitdistr* we obtained the estimates

$$\widehat{\text{shape}} = \hat{a} = 1.25000, \quad \widehat{\text{scale}} = \hat{s} = 11.846.$$

p	quantile	sfall
0.800	46.276	138.001
0.900	74.619	218.218
0.950	118.993	343.804
0.990	343.952	980.482
0.999	1537.819	4359.364

TABLE 2. Quantiles of the fitted excess distribution and expected shortfalls.

This density assigns probability 0.9909 to values in $[0, u]$. Thus

$$P[\text{Cost} \leq v | \text{Cost} \leq u] = \frac{\hat{F}(v)}{\hat{F}(u)} = \int_0^v f_{\hat{a},\hat{s}}(y)dy, \quad 0 \leq v \leq u,$$

and therefore, on $[0, u]$ we estimate the density of $F(v)$ as

$$\hat{F}(u) \int_0^v f_{\hat{a},\hat{s}}(y)dy,$$

where recall $\hat{F}(u) = 0.862$, the percentage of observations below the threshold u.

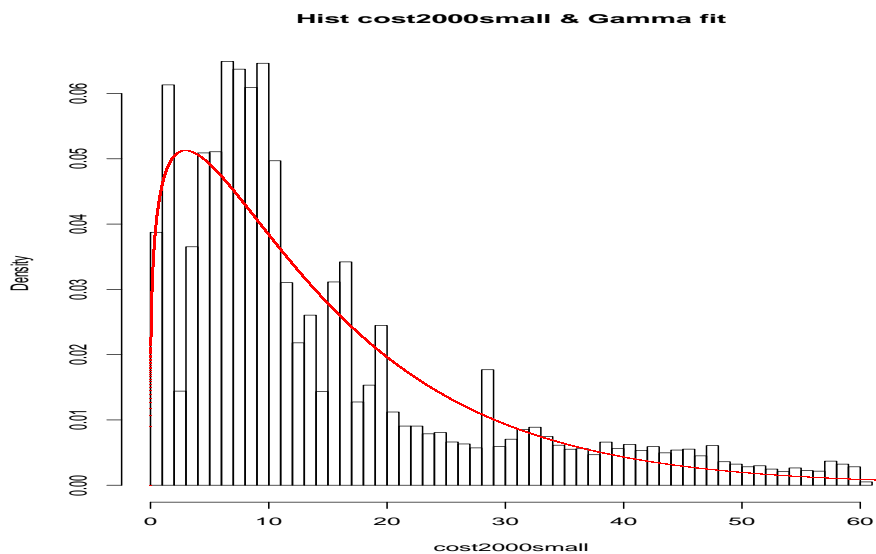The histogram of the prethreshold data *claims2000small* and the fitted Gamma density are displayed in Figure 5.

FIGURE 5. Histogram of the claim sizes data below threshold with the Gamma fitted density superimposed.

5. ANALYSIS OF SALES DATA

Having some understanding of the nature of warranty claims, we now turn to studying the sales process.

5.1. **Initial analysis.** The sales data for model year 2000 of a single car model consists of 34,807 records stretching over 1116 days or 3.05 years with an earliest date of 08/20/1999 and a latest date of 09/9/2002. Sales counts per day are given in Figure 6.
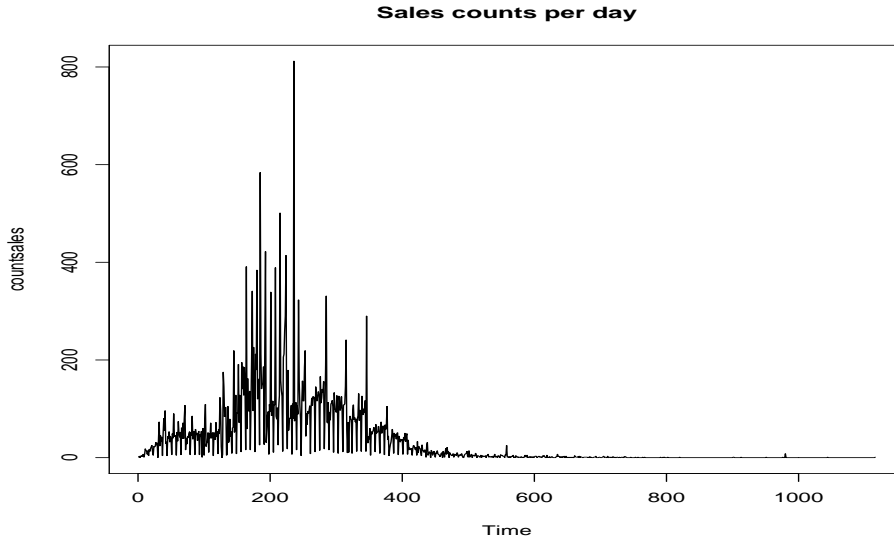
Sales counts per day

FIGURE 6. Sales per day for the 2000 model year stretching over 1116 days.

5.2. **The Bass sales rate model.** Bass (1969) proposed the following parametric model for sales of consumer goods after their market introduction. Let $\mu([0, t])$ be the cumulative sales of a particular model in t time units since the model introduction to the market. The sales rate model is then described by three parameters a, b and M as follows:

$$\frac{d\mu([0, t])}{dt} = a(\mu([0, t]) + b)(M - \mu([0, t])),$$

with initial condition $\mu(\{0\}) = 0$. The solution is:

$$(5.1) \qquad \mu([0, t]) = M\frac{b(1 - \exp\{-a(M + b)t\}}{b + M\exp\{-a(M + b)t\}}.$$

This model is designed so that the sales rate initially increases and then decreases and the total sales $\mu([0, t])$ approaches M as t approaches infinity. We can estimate a, b, M from sales data. For our purposes, a more convenient parameterization results from setting

$$a = \frac{A}{M}, \quad ab = B, \quad A + B = C,$$

and M, as before, is the total sales of the item. This gives

$$(5.2) \qquad \mu([0, t]) = M\frac{1 - \exp\{-Ct\}}{1 + \left(\frac{C}{B} - 1\right)\exp\{-Ct\}}.$$

Examples of different Bass sales rate curves are given in Figure 7.

Our intent is to use $\mu([0, t])$ as the mean measure of the Poisson process of sales. Comparing Figures 6 and 7, we see that there is excess variability in Figure 6 that is not captured by the smooth Bass sales rate plots so smoothing the data might be considered to get a better fit. Without smoothing the data, we perform maximum likelihood estimation of (B, C) using Poisson counts per day as the data. We use the R-function *nlm* and find estimated values of

$$(\hat{B}, \hat{C}) = (0.00041, 0.0163).$$

The plot giving observed daily counts vs fitted expected counts is in Figure 8. The fitted model does not capture all the variability, nor does it fit the big observed counts particularly well.
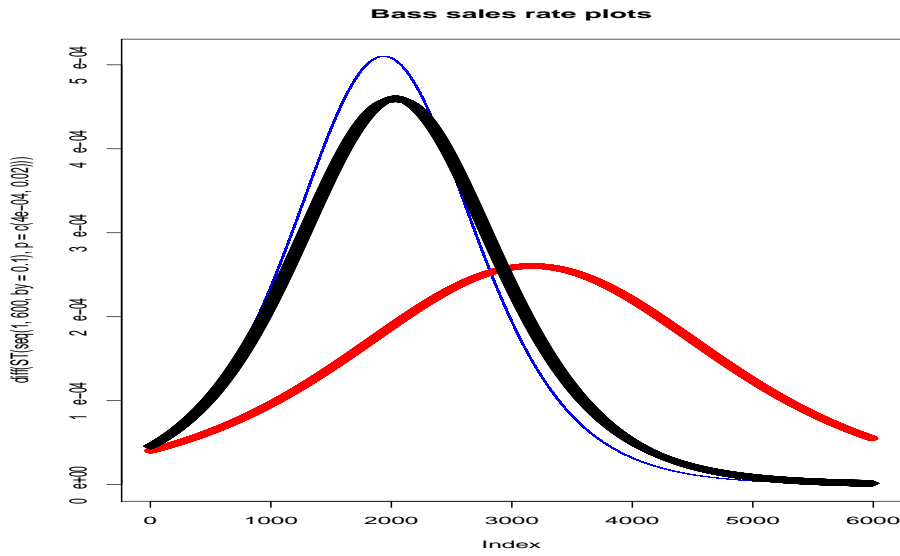
Bass sales rate plots

FIGURE 7. Different sales rate curves. The thin curve achieving the biggest maximum value correponds to $(B, C) = (.0004, .02)$; the middle, heavy curve is for $(B, C) = (.0004, .01)$; the lowest curve is for $(B, C) = (0.0004, 0.008)$.
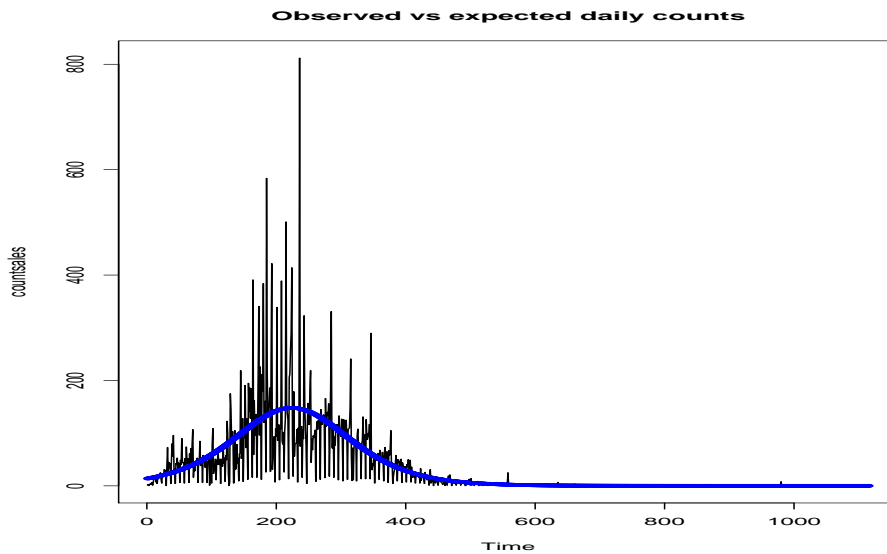
Observed vs expected daily counts

FIGURE 8. The spikey graph is the observed daily sales counts while the smooth curve gives the fitted Bass model.

A better visual fit to the Bass model comes from smoothing the data by aggregation into 12 day bins and considering sales per 12 day periods as the data. This yields very similar estimates of the parameters

$$(5.3) \qquad\qquad (\hat{B}, \hat{C}) = (0.00039, 0.0165)$$

and the observed 12 day sales counts vs expected plot looks better as shown in Figure 9. This increases our confidence that the parameter estimates of (B, C) are very reasonable.
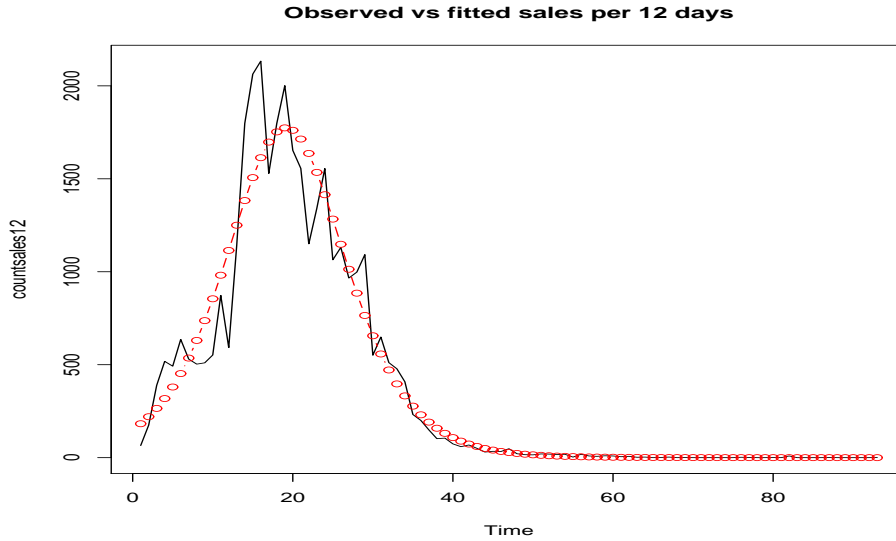
FIGURE 9. The continuous graph is the observed 12 day sales counts while the circles connected by lines gives the fitted Bass model.

6. ANALYSIS OF WARRANTY CLAIM DATA

This section analyzes warranty claim dates. We took warranty claim times of each vehicle and subtracted the time of sale to create times relative to zero for each car. The unit of measurement is one day. Each of these records was treated as a realization of a non-homogeneous Poisson process. Interestingly, this resulted in times which were negative (warranty claims presumably made by the dealer prior to sale) and times which exceeded the 3 year window. Owing to masking of the data by the manufacturer, not all sales dates could be matched with claims and subsets of sales and claims were selected corresponding to the same vehicle identification number.

6.1. **Initial analysis.** The amalgamation of all the warranty claim times relative to the vehicle purchase time produced a data set of length 65,351. The summary statistics for this data set follow in Table 6.1. Note the minimum is negative and the maximum exceeds 3 years or 1095 days.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-858.0	127.0	405.0	440.9	726.0	1459.0

TABLE 3. Summary statistics for the claim time data.

6.2. **Warranty claim counts.** The warranty claim times yield counts per day. The histogram of count frequencies for all the data is in left hand graph of Figure 10 which exhibits daily counts in the range (-900,1500). The right graph is a simplified histogram of count frequencies in which negative values are lumped to 0 and values exceeding 1095 days are lumped with 1096.

6.3. **Model fit.** The linear appearance of the warranty claims counts in the region $(0, 1095)$ exhibited in the left histogram of Figure 10 suggests estimating the measure ν appearing in (3.5) by

$$\nu(\{0\}) = w_1, \quad \nu(\{1096\}) = w_2, \quad \nu((i-1, i]) = mi + b, \ 1 \le i \le 1095,$$

where $\nu(\{1096\})$ results from all counts in the region (1095,1500]. Figure 11 gives the claim counts for days 1 through 1095 with a least squares fitted line yielding estimates of the intercept and slope $(75.69, -0.042)$.
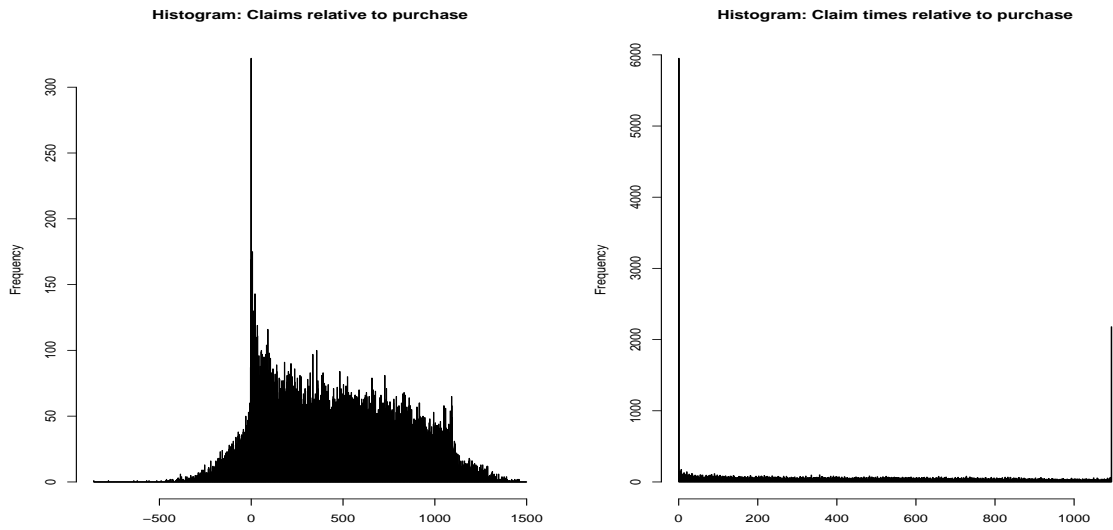
FIGURE 10. Histogram (left) of daily count frequencies in the range (-900,1500) and (right) histogram of daily counts in which negative values are counted as 0 and values exceeding 3 years are counted as 1095.

Keeping in mind that the data has been amalgamated over the 34,807 sales of cars in the data record provided, we divide these numbers by 34807 and get estimates of $\nu(\cdot)$ to be

$$(6.1) \quad \hat{\nu}(\{0\}) = 0.1663, \quad \hat{\nu}(\{1096\}) = 0.0612, \quad \hat{\nu}((i-1,i]) = (-1.218e^{-06})i + 0.002174, \ 1 \le i \le 1095.$$
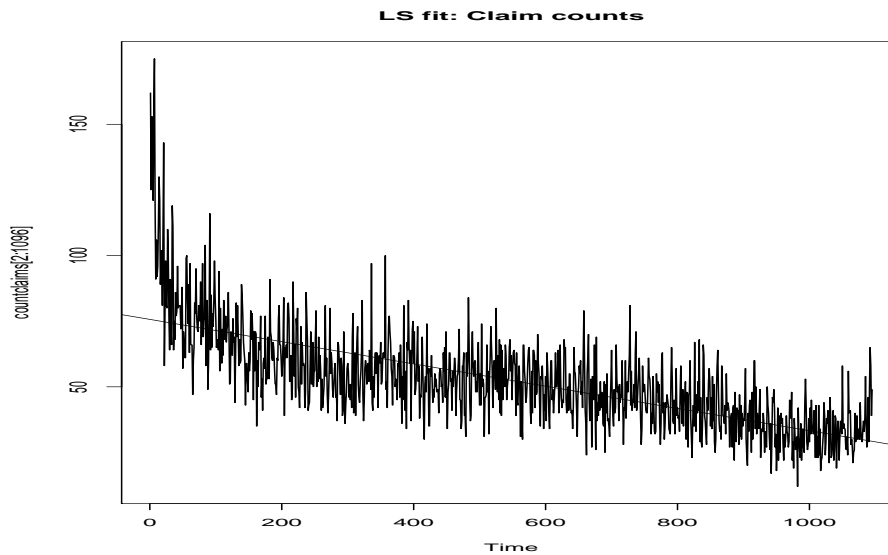

FIGURE 11. Claim counts for days 1 through 1095 and the least squares line.

7. TOTAL WARRANTY COSTS.

We assume that the sales start at time 0 and continue according to a non-homogeneous Poisson Process with rate function given in Equation (5.2), with parameters given in Equation (5.3). The warranty period is taken to be $W = 3$ years, and we compute the distribution of the total warranty costs for 12 quarters, the i-th quarter being the interval $[(i - 1) * T, i * T)$, with $T = 1/4$ year. The warranty claims from a single sale are assumed to arise according to a non-homogeneous Poisson process with mean measure given by the estimate in Equation (6.1), appropriately converted to years. Finally the distribution of individual claim size is assumed to be a combination of Gamma and generalized Pareto as described in Section 4.2. Using these values in the analysis as presented in Section 3.2 we compute the Laplace transform of the total warranty costs over different quarters. Then we use the numerical Laplace transform inversion program invlap (Hollenback (1998)) written in Matlab to compute the distribution of the total cost.

The table below gives the results about the mean, median and various quantiles of the costs for the 12 quarters. It shows that the warranty costs increase in the beginning and then decrease, as expected. There is a slight increase in the final quarter. This is due to the increase in the warranty claims at the end of warranty period.

Quarter	Mean	Median	75%	90%	95%	99%
1	23081	23036	24237	25355	26040	27358
2	63126	63081	65056	66868	67970	70069
3	83200	83155	85418	87491	88749	91139
4	46902	46856	48560	50130	51086	52912
5	14857	14811	15778	16685	17244	18325
6	4069	4024	4537	5036	5350	5940
7	1384	1338	1642	1953	2155	2568
8	743	697	922	1162	1322	1655
9	570	523	720	937	1082	1387
10	501	454	638	845	985	1277
11	454	408	583	783	919	1203
12	1236	1320	1621	1928	2128	2537

TABLE 4. Total claim quantiles for the 12 quarters

8. SUMMARY AND CONCLUSIONS

In this paper we have developed stochastic and statistical models to compute the warranty costs for a specified period, a quarter year in our case, arising out of a non-stationary sales process of the items under warranty. We have used the Bass model to estimate the sales rate as a function of time and then used this as the rate function of a non-homogeneous Poisson process representation of the sales process. We then developed a non-homogeneous Poisson process to model the claims process arising out of a single item during the warranty period. We also modeled the claim sizes as i.i.d. random variables with a combination of Gamma and Generalized Pareto distributions to account for the heavy tails manifest in the data. Finally we combined all these submodels to compute the quantiles of the warranty costs for the 12 quarters.

We believe that this methodology will help manufacturers with their financial planning in that they can more accurately account for and predict warranty liabilities in a proper fashion.

REFERENCES

S. Asmussen. *Ruin Probabilities*. World Scientific, Singapore, 2000.

F.M. Bass. A new product growth for model consumer durables. *Management Science*, 15:215–227, 1969.

S.G. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics. London: Springer. xiv, 210 p. , 2001.

S. Csörgő, P. Deheuvels, and D. Mason. Kernel estimates for the tail index of a distribution. *Ann. Statist.*, 13: 1050–1077, 1985.

P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extreme Events for Insurance and Finance.* Springer-Verlag, Berlin, 1997.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174, 1975.

Hollenbeck, K. J. INVLAP.M: A matlab function for numerical inversion of Laplace transforms by the de Hoog algorithm, http://www.isva.dtu.dk/staff/karl/invlap.htm, 1998.

S.-S. Ja, V. G. Kulkarni, A. Mitra, and J. Patankar. Warranty Reserves for Non-stationary Sales Processes. *NRLQ*, 49, No. 5, 499-513, 2002.

O. Kallenberg. *Random Measures.* Akademie-Verlag, Berlin, third edition, 1983. ISBN 0-12-394960-2.

D. Mason and T. Turova. Weak convergence of the Hill estimator process. In J. Galambos, J. Lechner, and E. Simiu, editors, *Extreme Value Theory and Applications*, pages 419–432. Kluwer Academic Publishers, Dordrecht, Holland, 1994.

D. Mason. Laws of large numbers for sums of extreme values. *Ann. Probab.*, 10:754–764, 1982.

A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management.* Princeton Series in Finance. Princeton University Press, Princeton, NJ, 2005. ISBN 0-691-12255-5. Concepts, techniques and tools.

J. Neveu. Processus ponctuels. In *École d'Été de Probabilités de Saint-Flour, VI—1976*, pages 249–445. Lecture Notes in Math., Vol. 598, Berlin, 1977. Springer-Verlag.

S.I. Resnick and C. Stărică. Smoothing the Hill estimator. *Adv. Applied Probab.*, 29:271–293, 1997.

S.I. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling.* Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2006. ISBN: 0-387-24272-4.

S.I. Resnick. *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag, New York, 1987.

S.I. Resnick. Heavy tail modeling and teletraffic data. *Ann. Statist.*, 25:1805–1869, 1997.

T. Rolski, H. Schmidli, V. Schmidt, and J. Teugels. *Stochastic Processes for Insurance and Finance* Wiley, New York, 1999.

V.G. KULKARNI, 209 SMITH BUILDING, DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH, UNIVERSITY OF NORTH CAROLINA, CHAPEL HILL, NC 27599 USA
E-mail address: vkulkarn@email.unc.edu

SIDNEY RESNICK, SCHOOL OF OPERATIONS RESEARCH AND INDUSTRIAL ENGINEERING, CORNELL UNIVERSITY, ITHACA, NY 14853 USA
E-mail address: sir1@cornell.edu