Insurance: Mathematics and Economics 51 (2012) 43–52

Contents lists available at SciVerse ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

A new class of models for heavy tailed distributions in finance and insurance risk

Soohan Ahn[a], Joseph H.T. Kim[b,*], Vaidyanathan Ramaswami[c]

[a] Department of Statistics, University of Seoul, Seoul 130-743, South Korea

[b] Department of Statistics and Actuarial Science, University of Waterloo, Waterloo ON, Canada

[c] AT&T Labs-Research, 180 Park Avenue, Florham Park, NJ 07932, USA

ARTICLE INFO

Article history:
Received November 2011
Received in revised form
January 2012
Accepted 4 February 2012

Keywords:
Generalized Pareto distribution
Log phase-type distribution
Heavy tail
Data fitting
Extreme value theory

ABSTRACT

Many insurance loss data are known to be heavy-tailed. In this article we study the class of Log phase-type (LogPH) distributions as a parametric alternative in fitting heavy tailed data. Transformed from the popular phase-type distribution class, the LogPH introduced by Ramaswami exhibits several advantages over other parametric alternatives. We analytically derive its tail related quantities including the conditional tail moments and the mean excess function, and also discuss its tail thickness in the context of extreme value theory. Because of its denseness proved herein, we argue that the LogPH can offer a rich class of heavy-tailed loss distributions without separate modeling for the tail side, which is the case for the generalized Pareto distribution (GPD). As a numerical example we use the well-known Danish fire data to calibrate the LogPH model and compare the result with that of the GPD. We also present fitting results for a set of insurance guarantee loss data.

1. Introduction

Using the normal distribution to model risk is inadequate in many insurance applications because, as historical records do attest, insurance loss is typically characterized by asymmetry and thick tails. In particular, losses are often non-negative, right-skewed and leptokurtic. While there are numerous different parametric families to choose from in modeling insurance loss data, see e.g., Klugman et al. (2004), the list of possible parametric candidates is not exhaustive.

The task of accurately fitting the tail is an important task in actuarial risk modeling as the losses in the tail, though rare in frequency, are indeed the ones that have the most impact on the operations of an insurer and could lead to possible bankruptcy of the company. For example, the insured loss caused by hurricane Katrina in 2005 amounts to $72 billion dollars, and the loss from the 9/11 attacks are about $23 billions (Swiss Re, 2011). Modeling jumbo losses and their analyses therefore have been of great interest to actuaries in order to determine the appropriate level of premiums and reserves, and re-insurance.

Traditional parametric models actuaries have employed for heavy-tailed loss data include the Weibull, Loggamma, Burr, and the popular generalized Pareto distribution (GPD) in the context of Extreme Value Theory (EVT) (Embrechts et al., 1997). Various techniques of EVT have been applied successfully to many disciplines such as finance, actuarial science, and hydrology.

In this article we consider the Log phase-type (LogPH) distribution as an alternative parametric model choice for modeling insurance claim data. The LogPH class was introduced by Ramaswami and first reported formally in Ghosh et al. (2011); see Ramaswami (2011) for several other successful uses of the LogPH distribution in the context of queuing and reliability. The attractiveness of this class is due to its ability to model both the tail and the head of the data using a common single family of distributions and also due to certain denseness properties to be discussed below. Such a distribution is obtained by treating the underlying variable to be such that its (natural) logarithm follows a phase type (PH) distribution, i.e., is distributed as the absorption time in a suitably chosen finite state Markov chain with one absorbing state. Phase type distributions are discussed in detail in Chapter 2 of Latouche and Ramaswami (1999), and include the exponential distribution and its mixtures and convolutions as special cases. In addition, they are closed under many of the usual operations such as mixing, convolutions, formation of maxima and minima, etc., and dense in the class of all distributions on the nonnegative axis. A major disadvantage of the class, however, is the exponential decay of the tail, a fact remedied by the LogPH construction.

Some advantages of the LogPH over the GPD, at least from the practitioner's perspective, include the following: (1) we have one common parametric family of distributions generalizing the Pareto distribution and with certain known denseness properties; (2) the class has a power law tail, suitable for heavy-tail data, with

* Correspondence to: Department of Applied Statistics, Yonsei University, Seoul 120-749, South Korea.

E-mail addresses: jhtkim@yonsei.ac.kr, jhtkim@uwaterloo.ca (J.H.T. Kim).

0167-6687/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.insmatheco.2012.02.002

a power law index that is arbitrary; (3) the model could fit the entire range of data and not just the tail, eliminating the need to select the tail threshold from which the GPD starts; (4) we can use the full data as opposed to some subset of it; (5) there is no need to invoke the technical assumption underlying the EVT. While the general definition of the LogPH would possibly show a large number of parameters and their reduction is an area for further research, the number of parameters can be limited by restricting to certain subsets of distributions such as the generalized Coxian distributions for which the Markov chain of the PH distribution is acyclic. Also, many of the reported examples of fitting from diverse fields have shown that a fitting based on EM has resulted in only a few distinct parameters in the model; see Ramaswami (2011).

The article is organized as follows. Section 2 provides a brief summary of PH distributions. In Section 3 the LogPH distribution is introduced and its properties are studied, including a parameter estimation method and risk measure expression. Its tail behavior is also investigated. To illustrate its practical use, in Section 4, we fit the LogPH model to the Danish fire data and compare the calibrated model with the generalized Pareto alternative. We also use a set of insurance guarantee loss data and show how the LogPH model fits the data well.

2. PH-type distribution

The phase-type (PH) distribution was introduced by Neuts (1975, 1981) and is defined as the distribution of the time until absorption in a continuous time Markov chain with an absorbing state. That is, we consider a continuous time Markov chain with a state space $\{1, \ldots, p, p+1\}$, which has an initial probability vector $(\alpha, 0)$, $\alpha\mathbf{1} = 1$ and also an infinitesimal generator

$$Q = \begin{pmatrix} T & t \\ \mathbf{0} & 0 \end{pmatrix}, \quad t = -T\mathbf{1},$$

where $\mathbf{0}$ is a p dimensional row vector consisting of zeros, and $\mathbf{1}$ is a p dimensional column vector consisting of ones. Note that the state $p+1$ is an absorbing state, α is a p dimensional row vector and T is a $p \times p$ matrix. It is assumed that T is an infinitesimal generator such that the off-diagonal elements of T are non-negative and the diagonal elements of T are strictly negative. We also assume that the real part of each eigenvalue of T is strictly negative and define $-\eta_T < 0$ to be the eigenvalue of T closest to 0. These assumptions ensure that the first p states are transient and that absorption occurs almost surely (a.s.). If we let X be the time until absorption in the Markov chain, then the distribution of X is called the PH-distribution with (parameters) (α, T) and is denoted by $PH(\alpha, T)$.

The PH distributions have received much attention in the applied literature related to queues, dams, insurance risk, reliability, etc., due to their useful properties such as closure properties, computational tractability and denseness. Note that the denseness property means that the class of PH distributions is dense in the Prohorov metric of weak convergence in the set of all probability distributions on $[0, \infty)$. For more details, we refer to Latouche and Ramaswami (Latouche and Ramaswami, 1999). Moreover, recently, an efficient method based on the EM algorithm was reported by Asmussen et al. (1996) for fitting PH distributions to data and is gaining much popularity. For previous other approaches in fitting, see e.g. Panchenko and Thuümmler (2007) and Bobbio et al. (2003). Recently, Lee and Lin (2010) present a fitting algorithm using a class of mixtures of Erlang distributions which is a subclass of the class of the phase type distributions. They use a modified EM algorithm for parameter estimation and discuss the efficiency of their algorithm relative to the general algorithm of Asmussen et al. (1996).

If X is PH-type distributed with (α, T), then $f_X(x)$ and $F_X(x)$, which are the density function and distribution function of X respectively, have the following forms

$$F_X(x) = 1 - \alpha e^{Tx}\mathbf{1}, \quad x > 0,$$
$$f_X(x) = \alpha e^{Tx} t, \quad x > 0.$$

From this, one can show that the Laplace–Stieltjes transform of $PH(\alpha, T)$ is given by

$$\phi(s) = E[e^{-sX}] = \alpha(s\mathbf{I} - T)^{-1} t \quad \text{for } Re(s) \geq 0 \tag{1}$$

where $\mathbf{I}$ is the identity matrix, and the moments

$$E[X^k] = k!\,\alpha(-T^{-1})^k\mathbf{1} \quad \text{for } k \geq 1.$$

We can also see from the form of $1-F_X(x)$ that PH distributions have an exponentially decaying tail, that is, $1 - F_X(x) = O(e^{-\eta_T x})$ as $x \to \infty$. The fact that PH distributions have a light tail render their use in modeling heavy tailed distributions difficult. Even though the class of PH distributions can approximate a heavy-tailed distribution to any desired accuracy due to their denseness property, modeling heavy tails may require an excessive number of states in the Markov chain and thereby a large number of parameters, leading to impractical models.

3. LogPH-type distribution

The LogPH distribution, to be denoted by LogPH(α, T), was introduced by Ramaswami – see Ghosh et al. (2011); Ramaswami (2011) – and is defined as the distribution of the random variable $Y = \exp(X)$ where X has a PH distribution with (α, T). We note vice versa that, if Y is LogPH distributed then $\log(Y)$ is a PH distributed random variable.

From the definition and the properties of a PH distribution, we can easily show that a LogPH random variable Y with (α, T) has its distribution function and density function as

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\log Y \leq \log y) \\ &= 1 - \alpha e^{T \log y}\mathbf{1}, \quad \text{for } y \geq 1 \end{aligned} \tag{2}$$

and

$$f_Y(y) = \frac{1}{y}\alpha e^{T \log y} t, \quad y \geq 1, \ t = -T\mathbf{1}. \tag{3}$$

The requirement $Y \geq 1$ seen above is not really a major constraint. In many practical situations, this can be achieved by a rescaling of the underlying variable. An alternative is to use bilateral phase type distributions introduced by Ahn and Ramaswami (2005) and proceed in ways similar to our discussion by exploiting the fact that the positive and negative parts of a bilateral phase type random variable are each phase type. An example of an instance where this more general model may be required is one where we are modeling a real valued random variable with heavy tails both on the left and on the right. To keep things simple, for the remainder of this paper, we restrict ourselves to the LogPH case only.

3.1. Moments

We consider the moments of the LogPH distribution in this section. If we use the change of variable technique, then we have

$$\int_1^\infty y^k e^{T \log y} dy = \int_0^\infty e^{ky} e^{Ty} e^y dy = \int_0^\infty e^{[(k+1)\mathbf{I}+T]y} dy.$$

Here, the finiteness and nonnegativity of the integral depends on the eigenvalues of T. That is, if the integral is finite, then its value is equal to

$$-((k+1)\mathbf{I}+T)^{-1}, \tag{4}$$

which is possible only when all the eigenvalues of $(k+1)I+T$ are negative.

We recall that the real parts of all the eigenvalues of T have negative real parts, and that $-\eta_T < 0$ is the eigenvalue of T closest to 0. Thus, if we define k^* as

$$k^* = \arg\min_{k\in N}\{k < \eta_T\},$$

then the LogPH distribution has moments of up to order k^*. Also, for $k \le k^*$, we have

$$E[Y^k] = \int_1^\infty y^k \frac{1}{y}\alpha e^{T\log y} t dy = \alpha \int_1^\infty y^{k-1} e^{T\log y} dy\, t$$
$$= \alpha(k\mathbf{I}+T)^{-1}T\mathbf{1}. \tag{5}$$

3.2. Conditional tail moments

The Conditional Tail Expectation (CTE)[1] is a widely accepted risk measure in insurance and finance applications. For a continuous loss distribution function F_Y corresponding to a random variable Y and a suitably chosen confidence level $p \in (0, 1)$, the CTE at p is defined as

$$CTE_p(Y) = E(Y|Y > Q_p(Y)) \tag{6}$$

where $Q_p(Y)$ is the pth quantile of F_Y. For insurance companies, for example, the CTE quantifies the expected loss when an exceptional loss does occur and can be used to determine required risk reserves, re-insurance, etc.

For univariate and multivariate PH distributions the CTE is analytically available (Cai and Li, 2005). It is easy to show that, if X has a PH distribution with (α, T), then the excess of loss $X - d|X > d$ is again PH distributed with parameter (α_d, T) where

$$\alpha_d = \frac{\alpha\, e^{dT}}{\alpha\, e^{dT}\mathbf{1}} \tag{7}$$

for any constant $d > 0$. This yields that

$$E(X|X>d) = d + E(X-d|X>d) = d - \frac{\alpha\, e^{dT}T^{-1}\mathbf{1}}{\alpha\, e^{dT}\mathbf{1}}, \tag{8}$$

where $E(X - d|X > d)$ is commonly called the mean excess function. The CTE of X at p is therefore obtained by setting $d = Q_p(X)$, the p quantile of the loss X. Note that, Eq. (8) cannot be used directly to determine the CTE of the LogPH variable $Y = e^X$, since the expectation operator does not commute with transformations in general.

Instead, we derive the analytic form for the CTE at p of the LogPH random variable using the fact that the excess risk of the PH variable is also PH distributed. Recall first that the moment generating function (MGF) of the PH distribution, from its Laplace–Stieltjes transform (1), is given by

$$M_X(s) = E(e^{sX}) = -\alpha(s\mathbf{I}+T)^{-1}t. \tag{9}$$

Since the excess risk at threshold d is also PH distributed, from (7) and (9), we have

$$E[e^{s(X-d)}|X>d] = -\alpha_d(s\mathbf{I}+T)^{-1}t \tag{10}$$

[1] Also known as the Expected Shortfall, Conditional VaR, and Tail VaR in the finance literature.

subject to the existence of the inverse in the right side of Eq. (10); when the inverse fails to exist, so does the expected value on the left. The left side of the previous equation is now rewritten as

$$E[e^{s(X-d)} \mid X > d] = e^{-sd}E[e^{sX} \mid X > d]$$
$$= e^{-sd}E[e^{sX} \mid e^X > e^d]$$
$$= e^{-sd}E[Y^s \mid Y > e^d].$$

Hence, plugging this back to (10) gives, subject to its existence, that

$$E[Y^s|Y > e^d] = -e^{sd}\alpha_d(s\mathbf{I}+T)^{-1}t. \tag{11}$$

Let us set $d = Q_p(X)$. Also, assume that the eigenvalue $\eta_T > s$, so that the moments of Y up to order s exist. Using that the quantile is preserved under a monotonic transformation, that is, the pth quantile of $Y = e^X$ equals $e^d = Q_p(Y)$, Eq. (11) can be rewritten as

$$CTE_p[Y^s] = -(Q_p(Y))^s\alpha_{Q_p(X)}(s\mathbf{I}+T)^{-1}t. \tag{12}$$

This yields the conditional tail moments of an arbitrary order s of the LogPH distribution, and reduces to the CTE formula by setting $s = 1$.

The mean excess function (MEF) and its higher moment counterparts of the LogPH are readily available from the previous result. In particular, we find from (11) that, subject to its existence,

$$E[Y-e^d|Y>e^d] = -e^d[\alpha_d(\mathbf{I}+T)^{-1}t+1]. \tag{13}$$

By replacing $u = e^d$, we obtain

$$E[Y-u|Y>u] = -u[\alpha_{\log u}(\mathbf{I}+T)^{-1}t+1], \tag{14}$$

the MEF of the LogPH distribution. A further discussion on the MEF will follow shortly.

Note also that since the distribution function of X is continuous, increasing and also equal to 0 at 0, it is easy to verify that

$$\lim_{p\to 0} d = \lim_{p\to 0} Q_p(X) = 0 \quad \text{and} \quad \lim_{p\to 0}\alpha_d = \alpha,$$

leading to

$$\lim_{p\to 0} CTE_p(Y) = -\alpha(I+T)^{-1}t = E[Y]. \tag{15}$$

We comment that the general tail moments are relevant in some approaches to the portfolio selection problem in the financial economics literature. For example, the lower partial moment of order s with threshold d of a portfolio return defined by

$$E[(d-Y)^s \cdot I(Y<d)], \quad s > 0 \tag{16}$$

has been investigated as a measure of the investment risk instead of the traditional return variance; see, e.g., Bawa and Lindenberg (1977), Harlow and Rao (1989) and Grootveld and Hallerbach (1999).

3.3. Tail behavior

As mentioned earlier, the tail functions of the PH and LogPH distribution function have respectively exponential and power law decaying tails. We start with a brief overview of several tail-related quantities in the context of extreme value theory (EVT).

We use the notation $\bar{F}(y)$ to denote the tail (or survival) function of a given distribution function $F(y)$ which is defined as $\bar{F}(y) = 1 - F(y)$, $-\infty < y < \infty$. We say that a tail function $\bar{F}(y)$ is regularly varying with index $\rho > 0$, and we write $\bar{F} \in \mathcal{R}_{-\rho}$, if

$$\lim_{y\to\infty}\frac{\bar{F}(y\lambda)}{\bar{F}(y)} = \lambda^{-\rho}, \quad \lambda > 0. \tag{17}$$

When $\rho = 0$ the tail is called slowly varying, or $\bar{F} \in \mathcal{R}_0$. Using this we can represent a regularly varying distribution as

$\bar{F}(y) \sim L(y)y^{-\rho}$ where $L(\cdot) \in \mathcal{R}_0$. We note that $f(y) \sim g(y)$ means $\lim_{y\to\infty} f(y)/g(y) = 1$. Heavy-tailed distributions such as Pareto, Generalized Pareto, and Stable laws are examples of regularly varying distributions. The EVT states that when $\bar{F} \in \mathcal{R}_{-\rho}$ the distribution is in the maximum domain of attraction of the Fréchet distribution, a member of the generalized extreme value (GEV) distribution family. Further, the celebrated Balkema–de Haan–Pickands theorem asserts that the excess of loss $X - d|X > d$ from such a distribution with a large threshold $d > 0$ converges to the Generalized Pareto distribution (GPD) with positive Pareto index $\xi > 0$. That is,

$$\lim_{d\uparrow y_F} \sup_{0<y<y_F-d} |Pr(Y - d < y|Y > d) - G_{\xi,\beta}(y)| = 0, \tag{18}$$

where $y_F \leq \infty$ is the right endpoint of F and $G_{\xi,\beta}(y)$ is the distribution function of the GPD:

$$G_{\xi,\beta}(y) = 1 - \left(1 + \frac{\xi}{\beta}y\right)^{-1/\xi}, \quad y > 0,\ \xi > 0. \tag{19}$$

See, e.g., Sections 3.3 and 3.4 of Embrechts et al. (1997) for theoretical relationships between the GEV and the GPD with different ranges of the ξ parameter. Among other properties, the GPD is well-known for its linear mean excess function:

$$E(Y - d|Y > d) = \frac{\xi d + \beta}{1 - \xi}, \quad \xi < 1, \tag{20}$$

a preliminary tool to examine the existence of a heavy tail and to identify the tail threshold d.

This theoretical result provides a major foundation for EVT and has been widely utilized in many insurance and finance applications as a tool to fit the tail part of the data whenever heavy tail is indicated. In particular, the tail part is separately fitted using the GPD from a threshold u which is determined from the data at hand. The selection of the threshold, however, is a challenging task; by necessity it is determined empirically at a large value where data points are rare, and this affects the estimated parameters and their stability.

To examine the tail thickness of the LogPH distribution we first observe from Theorem 2.7.2 in Latouche and Ramaswami (1999), O'Cinneide (1991) and (2) that its tail function decays with rate $(\log y)^k y^{-\eta_T}$ for some k, $0 \leq k \leq m_T - 1$ where $-\eta_T < 0$ is the eigenvalue of T closest to 0 and m_T is its multiplicity. Hence it is clear that

$$\lim_{y\to\infty} \frac{\bar{F}(y\lambda)}{\bar{F}(y)} = \lambda^{-\eta_T}, \quad \lambda > 0 \tag{21}$$

indicating that the LogPH is a regularly varying distribution with index η_T, and thus that the LogPH distribution is in the maximum domain of attraction of the Fréchet distribution. Based on this tail behavior we find that the LogPH is a heavy-tailed distribution well aligned with extreme value theory.

In fact, the tail thickness of the LogPH is comparable to the GPD not just in the sense of the tail variation (17). A closer look at the MEF of the LogPH in (14) reveals that if the first moment exists, then

$$E[Y - u|Y > u] = -u[\alpha_{\log u}(\mathbf{I} + T)^{-1}t + 1] \sim c \cdot u, \quad u \uparrow \infty \tag{22}$$

for some positive c, by observing from (7) that $\alpha_{\log u}$ converges to a constant. Thus the MEF of the LogPH tends to a linear function as the threshold u becomes larger, another similarity to the GPD. In passing, we also note that the MEF of the PH distribution is almost constant from (8) as the threshold becomes larger, a resemblance to the Exponential distribution. The MEF graphs of the LogPH and PH shown in Fig. 1, using the parameters calibrated for the Danish fire data in Section 5, clearly shows this pattern. In particular, the left panel reveals that the LogPH MEF looks like a straight line from the very beginning to our eyes. Therefore, if the sample MEF plot shows a linear pattern asymptotically, the LogPH may be a good candidate as a parametric choice that does not require a separate model for the tail part as is the case for the GPD. Moreover, combined with its denseness in Theorem 1 to be given in the next section and flexibility in the number of parameters, the LogPH distributions form a rich class of distributions that are useful in insurance loss modeling.

4. Fitting LogPH distribution

In this paper, fitting LogPH distribution to data is carried out by fitting a PH distribution to the log transformed data. That is, for n observations $y_1, \ldots, y_n$ from a LogPH distribution, we take log transformation as

$$x_1 = \log(y_1), \ldots, x_n = \log(y_n)$$

and then fit a PH distribution to $x_1, \ldots, x_n$. We use the obtained estimators of the parameters (α, T) as the estimators for LogPH distribution. This is due to the fact that the log of the LogPH distributed random variable is a PH distributed random variable, which is mentioned in Section 3.

For fitting PH distribution, we use the algorithm developed by Asumussen et al.,[2] which is called the EMPHT algorithm. The EMPHT algorithm uses the EM (Expectation–Maximization) algorithm because of the incompleteness of the observations, and this is an iterative method for maximum-likelihood estimation. This algorithm is known to be the first successful algorithm for fitting the PH distribution to data, and we will briefly describe it in the following subsection.

Note that the class of PH distributions is known to be dense in the set of all distributions on $[0, \infty)$. That is, any distribution on $[0, \infty)$ can be approximated by a sequence of PH distributions. Note also that a PH distribution has an exponentially decaying tail which asymptotically goes to 0 as $e^{-\eta_T x}$ where η_T is the real eigenvalue of T closest to 0. Thus, if we apply the EMPHT algorithm directly to data, then this provides a way to get approximating distributions with exponentially decaying tail to the distribution of the data.

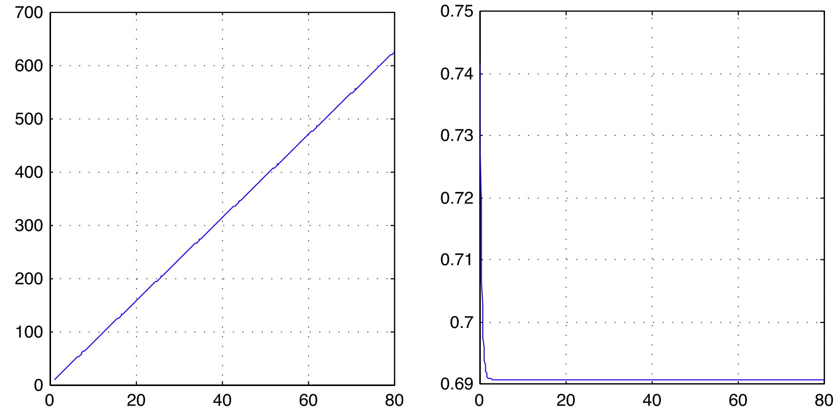
Meanwhile, if consider the distribution of the log transform of any random variable defined on $[1, \infty)$, then the distribution can be approximated also by a PH distributions by the denseness property of PH. Thus, the approach in this section provides a way for finding approximating distributions with a power law decaying tail to the distribution of data. Basically, this approach relies on the following theorem.

Theorem 1. *The class of LogPH distributions is dense in the set of all distribution functions defined on* $[1, \infty)$.

Proof. For a given distribution function G of a random variable $U \geq 1$, we can think of the random variable $V = \log(U)$ and its distribution function H. Since the class of phase type distribution is dense in the set all distribution functions defined on $[0, \infty)$, there is a sequence of phase type random variables $X_1, X_2, \ldots$ which converges in distribution to the random variable V.

Now, we take exponential transformation of X_i's such as $Y_i = \exp(X_i)$, $i \geq 1$ and denote its LogPH distribution function by F_i respectively. Then, by the continuous mapping theorem in Whitt (2002), the sequence of random variables Y_i's converges weakly to $\exp(V) = U$. In other words, the sequence of distribution functions

[2] EMpht, http://home.imf.au.dk/asmus/pspapers.html.



(a) MEF of LogPH distribution.

(b) MEF of PH distribution.

Fig. 1. Mean excess function of LogPH and PH.

F_i's converses to G in the Prohorov metric of weak convergence, which completes the proof. For more details on the continuous mapping theorem and the Prohorov metric of weak convergence, we recommend the readers to refer to the Ref. Whitt (2002). □

In practice, users may face losses that are positive but less than one, and there is also the possibility of a probability mass at the minimum value of data. In these cases, taking the logarithm and proceeding as described above is not directly feasible. Several remedies can be used in such cases as follows. Conditional modeling can be adopted when it is enough for the analysis to consider the conditional tail probability given that the data is greater than 1. One could also use a procedure based on shifting, in which we consider the shifted data $x_i - a_x + 1$ where a_x is the possible minimum value of data and take a proper inversion after fitting LogPH distribution to the transformed data. In addition, if there exists a positive probability mass at the minimum value a_x, we can think of estimating the mass by the relative number of data of the value a_x which is a consistent estimator, fitting LogPH distribution to the remaining data and then taking the mixture of the fitted distributions as the final fitted distribution to data. Finally, note that the phase type distribution construct does allow a mass at 0 for the PH distribution by choosing α such that $\alpha\mathbf{1} \leq 1$.

4.1. EMPHT algorithm — Maximum likelihood estimation

In this subsection, we briefly describe the EMPHT algorithm developed by Asmussen et al. (1996) for fitting a PH distribution to data. We consider n observations $x_1, \ldots, x_n$ and assume that these are generated by the PH distribution $PH(\alpha, T)$ with a $p+1$ dimensional state space as defined in Section 2. Assuming that we can observe the whole path of the underlying Markov Chain, the complete sample corresponding to a given observation x_i of the PH distribution can be represented as

$$(j_0^{(i)}, \ldots, j_{m_i-1}^{(i)}; s_0^{(i)}, \ldots, s_{m_i-1}^{(i)})$$

where m_i is the number of jumps until the Markov chain hits the absorbing state 0, $j_k^{(i)}$ is the k-th state visited by the Markov Chain, and $s_k^{(i)}$ is the corresponding sojourn time. Note that $j_{m_i}^{(i)} = 0$, $s_{m_i} = \infty$, and $x_i = s_0^{(i)} + \cdots + s_{m_i-1}^{(i)}$. Then, the density of the i-th complete sample can be represented by

$$[\alpha]_{j_0^{(i)}} \exp\left[[T]_{j_0^{(i)} j_0^{(i)}} s_0^{(i)}\right] [T]_{j_0^{(i)} j_1^{(i)}} \cdots \times \exp\left[[T]_{j_{m_i-1}^{(i)} j_{m_i-1}^{(i)}} s_{m_i-1}^{(i)}\right] [t]_{j_{m_i-1}^{(i)}}$$

where we use the notations $[B]_{i,j}$ and $[\mathbf{b}]_i$ to denote (i,j)-th element and i-th element of a given matrix B and a vector $\mathbf{b}$ respectively.

If we use this and denote by $\mathbf{x}$ the complete sample of all of n observations as

$$\mathbf{x} = (j_0^{(1)}, \ldots, j_{m_1-1}^{(1)}; s_0^{(1)}, \ldots, s_{m_1-1}^{(1)}; \cdots; j_0^{(n)}, \ldots, j_{m_n-1}^{(n)}; s_0^{(n)}, \ldots, s_{m_n-1}^{(n)})$$

then the density of the complete sample can be written in the form

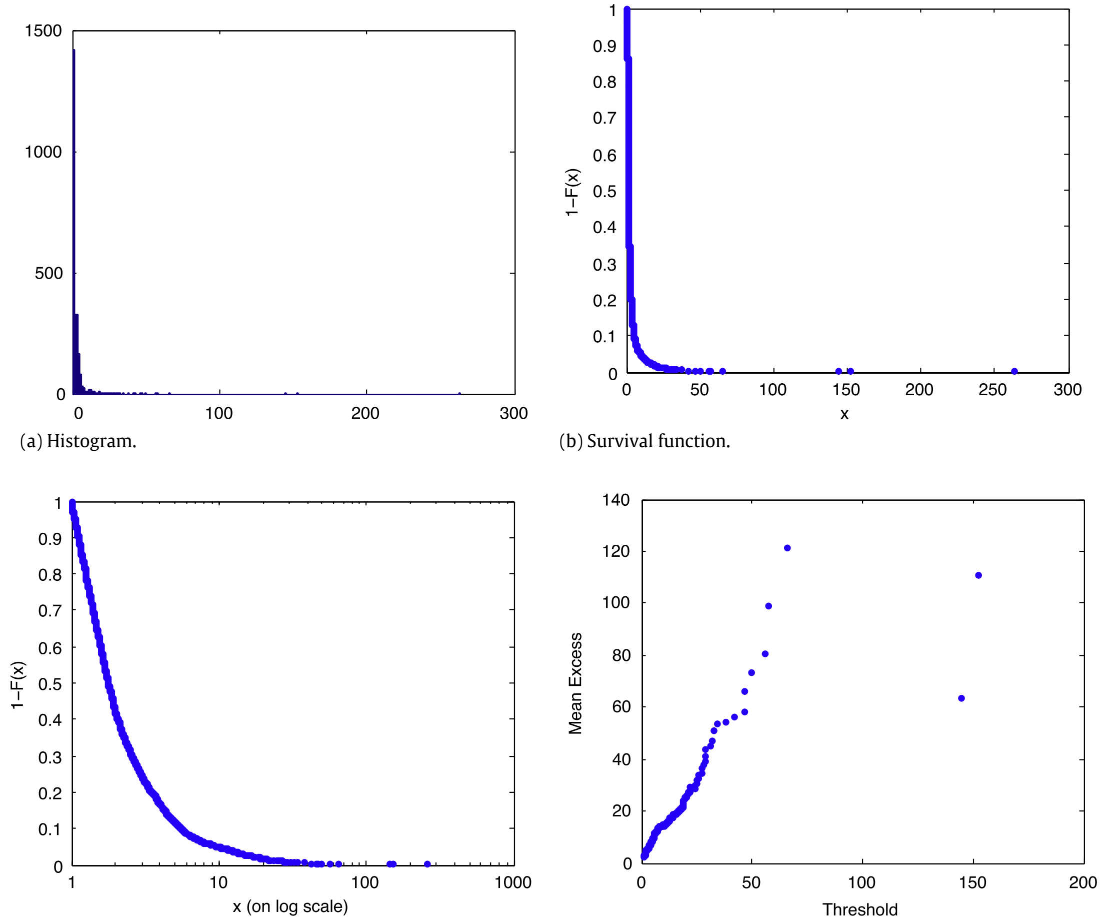
$$f(\mathbf{x}|\alpha, T) = \prod_{i=1}^{p} [\alpha]_i^{B_i} \prod_{i=1}^{p} \exp\left[[T]_{i,i} Z_i\right] \prod_{i=1}^{p} \prod_{\substack{j=1 \\ j \neq i}}^{p} [T]_{i,j}^{N_{i,j}} [t]_i^{N_{i,0}} \tag{23}$$

where B_i, $i = 1, \ldots, p$ denote the number of Markov Chains starting in state i, Z_i, $i = 1, \ldots, p$ the total time spent by Markov chains in state i, and $N_{i,j}$ the total number of jumps from state i to state j for $i \neq j$, $i = 1, \ldots, p$ and $j = 0, 1, \ldots, p$. We can observe that the density in (23) is a member of the multi-parameter exponential family with sufficient statistics B_i's, Z_i's, and $N_{i,j}$'s. This yields that the maximum likelihood estimators based on the complete sample $\mathbf{x}$ are given by

$$\begin{aligned} \hat{\alpha}_i &= \frac{B_i}{n}; \qquad [\hat{T}]_{i,j} = \frac{N_{i,j}}{Z_i}, \quad i \neq j; \\ [\hat{t}]_i &= \frac{N_{i,0}}{Z_i}; \qquad [\hat{T}]_{ii} = -[\hat{t}]_i - \sum_{\substack{j=1 \\ j\neq i}} [\hat{T}]_{i,j}. \end{aligned} \tag{24}$$

We note again that the observed data $x_1, \ldots, x_n$ are incomplete and the EM algorithm works for calculating the maximum likelihood estimators of α and T based on the incomplete data. The EM algorithm is an iterative procedure that maximizes in each step the conditional expectation of the log-likelihood given the observed sample $x_1, \ldots, x_n$. Hence there are in each iteration two steps, one of which is the E-step for calculating the conditional expectation of the log-likelihood given the observations which are absorption times, and the other is M-step for maximization. From the density function in (23), the log-likelihood of the complete sample is in the form of

$$\log f(\mathbf{x}|\alpha, T) = \sum_{i=1}^{p} B_i \log([\alpha]_i) + \sum_{i=1}^{p} [T]_{i,i} Z_i + \sum_{i=1}^{p} \sum_{\substack{j=1 \\ j \neq i}}^{p} N_{i,j} \log([T]_{i,j}) + \sum_{i=1}^{p} N_{i,0} \log([t]_i)$$



(a) Histogram.

(b) Survival function.

(c) Survival function on log scale.

(d) Sample mean excess function.

Fig. 2. Danish fire data summary. MEF excludes the largest observation.

and we observe that the log-likelihood function is linear in the sufficient statistics. Hence, in the E-step, calculating the conditional expectation of the log-likelihood given the observations reduces to calculating the conditional expectations of the sufficient statistics, the exact forms of which are given in Asmussen et al. (1996).

Now, we describe how the EM algorithm works here. Given any initial parameters (α_0, T_0), in E-step, the conditional expectations of the sufficient statistics given the observations are calculated and then are plugged in the likelihood as the sufficient statistics for complete sample. Then, in M-step, the maximum likelihood estimators are obtained using the equations in (24). These two steps are iterated until the convergence of the maximum-likelihood estimators.

As mentioned in Bladt (2005), though the EM algorithm always converges, it does not necessarily converge to the maximum likelihood estimator due to the possible presence of local maxima. In this case, the problem may be solved by a suitable variation of the initial values of (α_0, T_0). Also, since the representation of a given PH distribution is not unique, the estimated values of the parameters could be different depending on the initial values of the parameters though it represents the same distribution function. The minimal representation of the PH distributions still remains an open problem.

5. Numerical examples

To illustrate the use of the LogPH distribution, we consider two numerical examples in this section. The first one is the well-known Danish fire insurance data, and the second one is a simulated loss sample from a block of segregated fund of a Canadian life insurer.

5.1. Danish fire data

The Danish fire data[3] has been popular in modeling insurance claim size and appeared in the literature, including McNeil (1997), Resnick (2005) and Embrechts et al. (1997). The data, collected at Copenhagen Reinsurance, comprise 2167 fire losses between 1980 and 1990, both years inclusive. The numbers have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish Krone (Fig. 2). The minimum of the data is 1, so no adjustment is needed in fitting the LogPH distribution. In this section we compare the LogPH to the GPD with focus on the tail part. Since the GPD only fits the excess over a threshold we do not make a formal comparison about the fitting below the threshold. However the denseness of the PH distribution with more parameters would make the fitting of the center of data superior.

We start with fitting the GPD as a benchmark case. We consider two choices for the tail threshold of the Danish data as suggested in Embrechts et al. (1997); see Section 6.5 of the book for details of the GPD threshold selection and its application to the Danish fire data. The first threshold is at a loss of $u = 10$ with 109 exceedances, in which case the estimated parameters using the maximum likelihood estimation are $\hat{\xi} = 0.497$ and $\hat{\beta} = 6.975$. As the kth moment of the GPD is finite only when $k < 1/\xi$ for $\xi > 0$, it is indicated that, for $\hat{\xi} = 0.497$, only the first two

[3] Retrieved from Dr. A. McNeil's webpage: http://www.ma.hw.ac.uk/~mcneil/data.html.

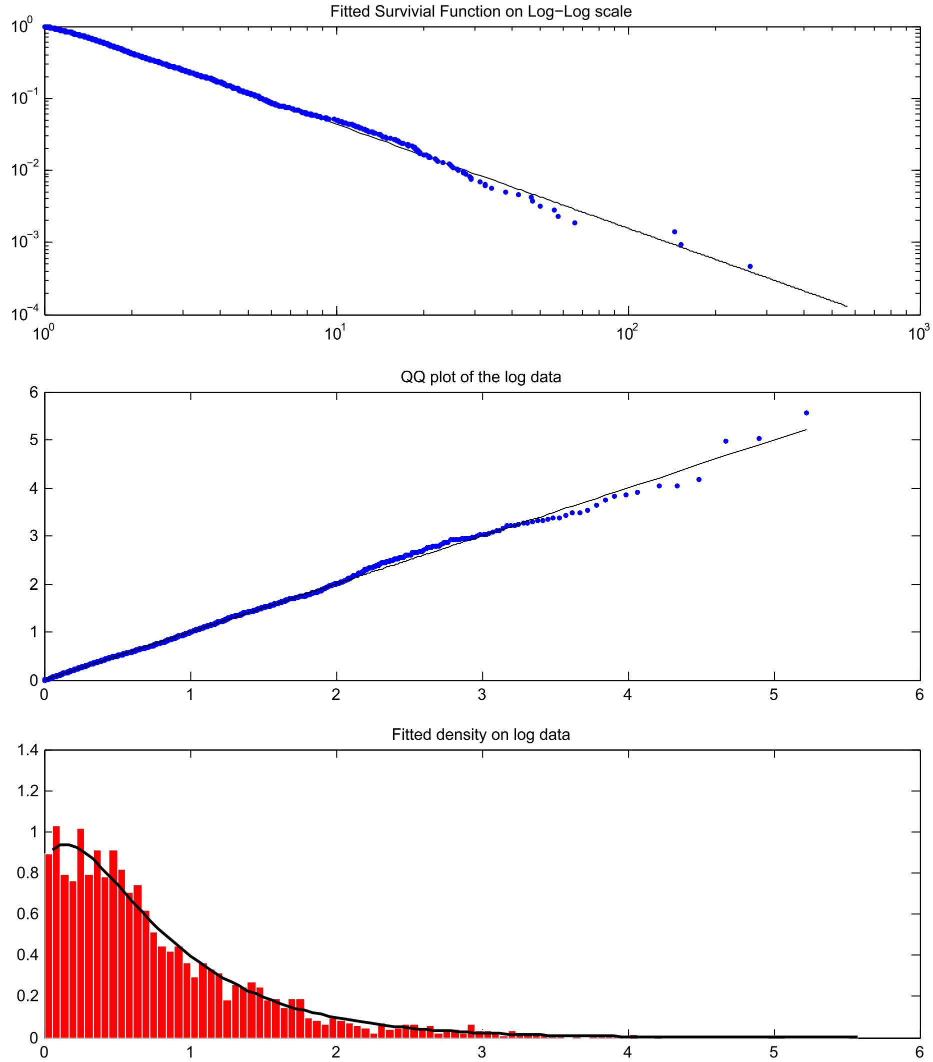


Fig. 3. Fitted LogPH model for Danish fire data.

moments exist. For the second choice, the threshold is set at $u = 18$ with 47 exceedances, the parameters are given by $\hat{\xi} = 0.735$ and $\hat{\beta} = 7.350$, indicating the existence of the first moment only.

Now we fit the whole data range – including the tail and center – with the LogPH using the method in the previous section. The data contains eleven identical minimum losses lying at 1, but does not seem to indicate a discrete mass at this point because ten of them are from year 1985 where much of the data are rounded at one or two decimal places.[4] We included these points in the parameter estimation without modification as this would not substantially alter the resulting parameter estimates. For the selection of the model we carry out the generalized likelihood ratio test based on the Wilks Theorem, which uses the asymptotic properties of the χ^2 statistic; see, e.g., Klugman et al. (2004) for various loss model selection procedures. The likelihood ratio test leads to the selection of a LogPH model with two phases. The estimated parameters are $\hat{\alpha} = (0.622, 0.378)$ and

$$\hat{T} = \begin{bmatrix} -4.000 & 3.564 \\ 0.267 & -1.813 \end{bmatrix}.$$

The eigenvalues of $\hat{T}$ are given by -4.37 and -1.44. Hence, from the moment condition in Section 3, $\hat{\eta}_T = 1.44$ and only the first moment seems to exist, confirming a considerably heavy tail. To be more precise, the moment exists up to about 1.4. This is in fact quite close to the moments existence range for the second choice of the GPD parameter estimation where $1/\hat{\xi} = 1/0.735 = 1.36$, indicating the existence of moments up to about 1.36. Thus, it is interesting to see that, at least from the moment perspective, the calibrated LogPH's tail thickness is in the middle of those two alternative GPD distributions, with it being very close to the second one.

Fig. 3 shows the fitted LogPH against the data. The tail part as well as the center of the data are well represented by the fitted LogPH distribution. In Fig. 4 we compare the tail fitting for the three models: the LogPH and the GPD with two different parameter choices as explained above. We see that the tail probability under the first choice for the GPD (dashed line in the picture) decreases fastest of the three models, leading to the existence of the first two moments. Both the second GPD parameter choice (dotted) and the LogPH case (solid) exhibit heavier tail, leading to the existence of the first moment only, affected more by the largest three losses in the picture. The tail of the dotted GPD is lighter than the LogPH for the early part of the tail, but becomes heavier in the deep tail area. As noted earlier, however, the tail thickness of both models in terms of the moments is comparable. At any rate, the advantage of the LogPH in fitting the overall data is evident from the picture.

In Table 1 we also present the conditional log-likelihood values for the three candidate models – two GPDs and one LogPH –

[4] No explanation about this discrepancy seems available in the literature.

Table 1
Conditional log-likelihood values for the Danish fire data. Number of exceedances is 109 when $u = 10$, and 47 when $u = 18$.

	$u = 10$		$u = 18$
GPD $\hat{\xi} = 0.497$, $\hat{\beta} = 6.975$	−374.89	GPD $\hat{\xi} = 0.735$, $\hat{\beta} = 7.350$	−175.30
LogPH	−375.98	LogPH	−177.75

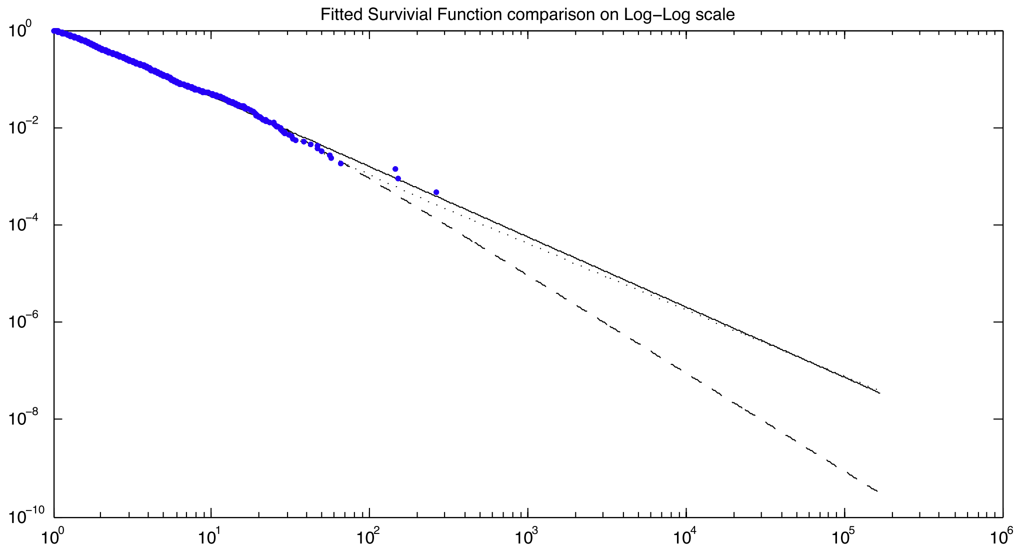


Fig. 4. Fitted model tail comparison for the Danish fire data: Solid line is the fitted LogPH model; Dashed line is the GPD with $\hat{\xi} = 0.497$, $\hat{\beta} = 6.975$, $u = 10$; Dotted line is the GPD with $\hat{\xi} = 0.735$, $\hat{\beta} = 7.350$, $u = 18$.

with two different threshold values. As GPD fitting ignores the losses below the given threshold, only the losses beyond the threshold are included in the likelihood calculation. Similarly, the LogPH likelihood contains only the losses beyond the given threshold with the likelihood contribution adjusted by the survival function at the threshold. The resulting values in the table show that the LogPH and the GPD have comparable fits. Although the GPD fit is slightly better than the LogPH, its advantage seems marginal, especially when we look at tail area around and beyond the sample maximum. One may carry out other likelihood-based model selection methods such as the Akaike Information Criterion (AIC),[5] but these methods however are not straightforward in our case because of a few reasons. First, as we consider the fit of the tail part only, the GPD has an advantage over the LogPH as its MLE maximizes the likelihood of the tail data directly, whereas the MLE of the LogPH maximizes the whole data's likelihood with the likelihood contributions of the data below the threshold being dominant. Second, the number of parameters one should use under likelihood-based criteria, such as the AIC where additional parameters are discouraged through a penalty, is uncertain when only the tail part is considered. This is because the LogPH fits the whole data range, rather than just the tail part, using more parameters.

Overall, for the Danish fire data, we find that the LogPH can be a promising alternative to model heavy-tailed insurance loss data. In particular, its Pareto-like tail is thick enough to be comparable to the GPD, but there is no need to fit the tail part and the rest separately. With flexibility in its parameter choice and denseness, the LogPH can easily model both the body and the heavy tail at the same time.

Table 2
Basic statistics based on the GMMB data (Unit: CDN dollar).

Mean	95.31
Variance	8699.20
Stand. dev.	93.27
Skewness	4.49
Kurtosis	33.98
Minimum	10.23
Maximum	1206.90

5.2. *Insurance guarantee loss data*

Segregated funds (a.k.a variable annuities) that offer insurance guarantees are often required to hold risk capitals by regulation based on simulated loss data from internal models. We consider a loss data set simulated from a block of an anonymous Canadian life insurer.[6] In particular, the block consists of 1055 policies with the level of the Guaranteed Minimum Maturity Benefit (GMMB) set at 100% of the initial fund value. The economic scenarios are based on the Wilkie stochastic model. After 2000 scenario runs we obtained a GMMB guarantee cost sample of the same size in Canadian dollars. A numerical and graphical summary of this data is given in Table 2, and Figs. 5 and 6. All of these indicate that the loss data is heavy-tailed and right-skewed.

To fit a LogPH the data has been first log-transformed and further shifted by the minimum, so that the smallest log data value lies at zero. Then PH distributions were fitted using different number of phases using the EMPHT algorithm in Section 4. From these we selected a LogPH with 9 phases based on the likelihood ratio test. We report the parameters in the Appendix. The resulting model is compared to the actual sample in Fig. 7. As in the first example, the model fits the data well in both the body and the tail. The closest eigenvalue of T is −4.4947, indicating that the first four moments exist.

5 The AIC selects the model that maximizes $l - k$, where l is the log-likelihood and k is the number of parameters of the model.

6 With thanks to David Gilliland and Wes Leong at GGY, Toronto.

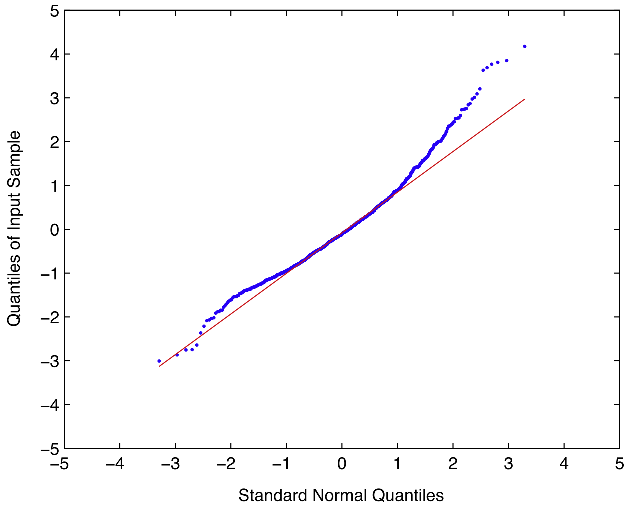


Fig. 5. Standardized Q–Q plot after log transform of the GMMB data.

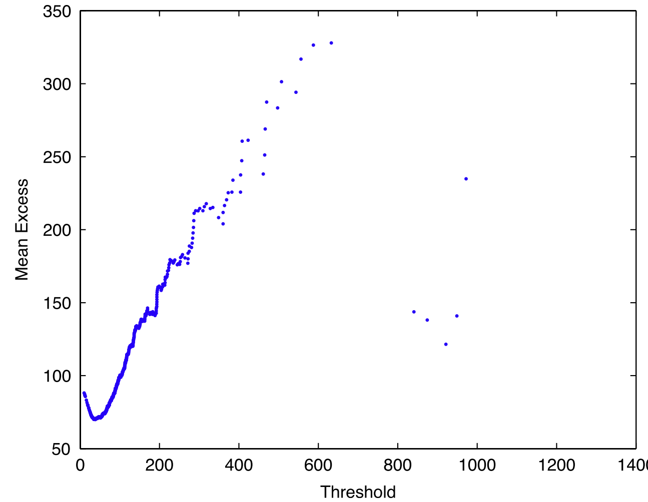


Fig. 6. Sample mean excess function of the GMMB data.

6. Concluding remarks

Fitting heavy-tailed loss data with parametric distributions is an important analytical task for insurers. Several possible methods are available in the literature to deal with tail estimation, including extreme value theory (EVT). In this article we propose the use of the LogPH class of distributions. This class exhibits a heavy tail and has some nice tail properties that are well aligned with the EVT. Since the LogPH can fit the whole data in a straightforward manner, it has an advantage over the EVT in that in the latter the tail threshold must be identified first from the data and that is a challenging task. In addition, combined with its denseness as proven in the article, the LogPH can offer a rich class of heavy-tailed loss distributions without separate modeling for the tail side.

Acknowledgments

The authors thank two anonymous referees for their constructive comments. Joseph Kim is grateful for the support of the Natural Sciences and Engineering Research Council of Canada. Soohan Ahn

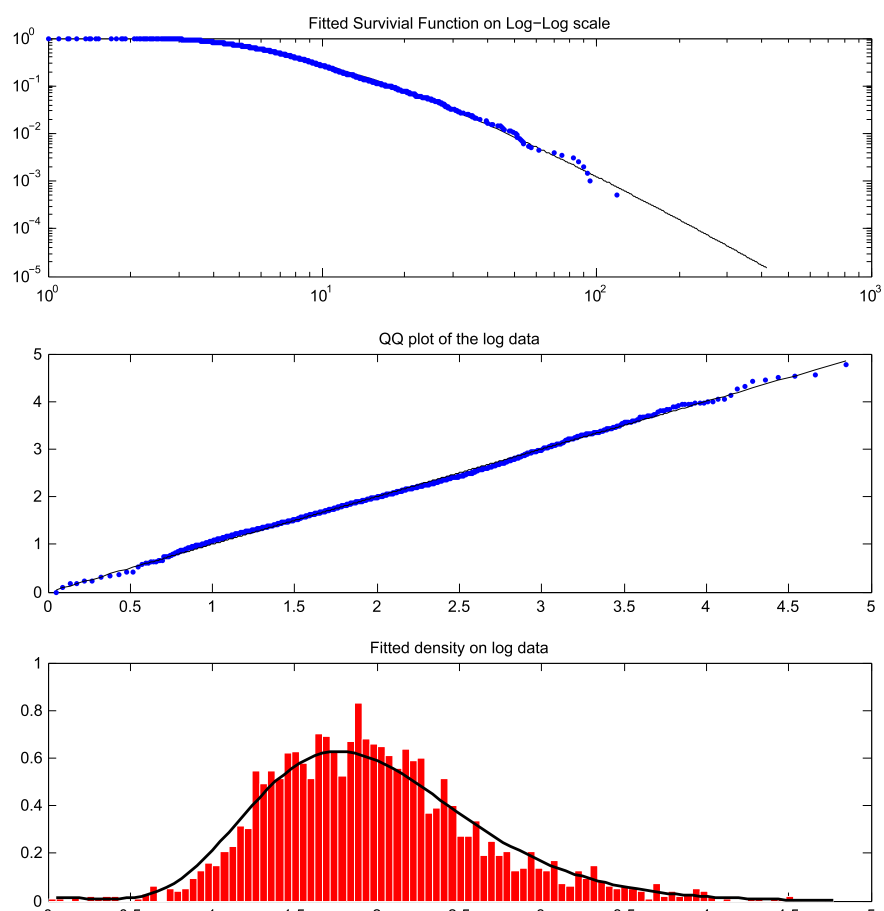


Fig. 7. Fitted LogPH model for the insurance guarantee loss.

$$\alpha = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0.9979 & 0 & 0 & 0.0021 \end{pmatrix}$$

$$T = \begin{pmatrix} -4.4947 & 0 & 4.4947 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -4.4947 & 0 & 0 & 4.4947 & 0 & 0 & 0 & 0 \\ 0 & 4.4947 & -4.4947 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4.4947 & 0 & 0 & 0 & 4.4947 & 0 \\ 0 & 0 & 0 & 4.4947 & -4.4947 & 0 & 0 & 0 & 0 \\ 4.5835 & 0 & 0 & 0 & 0 & -4.6034 & 0 & 0 & 0.0199 \\ 0 & 0 & 0 & 0 & 0 & 0 & -4.4947 & 0 & 4.4947 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4.4947 & -4.4947 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4.6034 \end{pmatrix}$$

Box I.

was supported by the 2011 sabbatical year research grant of the University of Seoul. Soohan Ahn also greatly appreciate the hospitality of the Department of Statistics, University of Toronto, during his sabbatical year.

Appendix

The parameters of the calibrated LogPH model for the insurance guarantee loss data are given in Box I.

References

Ahn, S., Ramaswami, V., 2005. Bilateral phase type distributions. Stochastic Models 21, 239–259.

Asmussen, S., Nerman, O., Olsson, M., 1996. Fitting phase-type distributions via the EM algorithm. Scandinavian Journal of Statistics 23 (4), 419–441.

Bawa, V.S., Lindenberg, E.B., 1977. Capital market equilibrium in a mean-lower partial moment framework. Journal of Financial Economics 5 (2), 189–200.

Bladt, M., 2005. A review on phase-type distributions and their use in risk theory. Astin Bulletin 35 (1), 145–161.

Bobbio, A., Horv́th, A., Scarpa, M., Telek, M., 2003. Acyclic discrete phase type distributions: properties and a parameter estimation algorithm. Performance Evaluation 54 (1), 1–32.

Cai, J., Li, H., 2005. Conditional tail expectations for multivariate phase-type distributions. Journal of Applied Probability 42 (3), 810–825.

Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. Modelling Extremal Events. Springer, New York.

Ghosh, A., Jana, R., Ramaswami, V., Rowland, J., Shankaranarayanan, N.K., 2011. Modeling and characterization of large-scale Wi–Fi traffic in public hot-spots. In: Proc. IEEE INFOCOM 2011 Shanghai, China, pp. 2921–2929.

Grootveld, H., Hallerbach, W., 1999. Variance vs downside risk: is there really that much difference? European Journal of Operational Research 114 (2), 304–319.

Harlow, W., Rao, R., 1989. Asset pricing in a generalized mean-lower partial moment framework: theory and evidence. Journal of Financial and Quantitative Analysis 24 (03), 285–311.

Klugman, S., Panjer, H., Willmot, G., 2004. Loss Models, second ed. John Wiley, New York.

Latouche, G., Ramaswami, V., 1999. Introduction to Matrix Analytic Methods in Stochastic Modeling. SIAM & ASA, Philadelphia.

Lee, S.C.K., Lin, Sheldon X., 2010. Modeling and evaluating insurance losses via mixtures of Erlang distributions. North American Actuarial Journal 14 (1), 107–130.

McNeil, A., 1997. Estimating the tails of loss severity distributions using extreme value theory. ASTIN Bulletin 27 (1), 117–137.

Neuts, M.F., 1975. Probability distributions of phase type. In: Liver Amicorum Prof. Emeritus H. Florin, University of Louvain, Belgium, pp. 173–206.

Neuts, M.F., 1981. Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach. The Johns Hopkins University Press, Baltimore, MD.

O'Cinneide, C.A., 1991. Phase type distributions and invariant polytopes. Advances in Applied Probability 23, 515–535.

Panchenko, A., Thuümmler, A., 2007. Efficient phase-type fitting with aggregated traffic traces. Performance Evaluation 64 (7–8), 629–645.

Ramaswami, V., 2011. Applied probability and statistics: some challenges and opportunities (Where angels fear to tread?). In: Nadarajan, R., Lekshmi, R.S., Sai Sundara Krishnan, G. (Eds.), Computational and Mathematical Modelling. Narosa Publishing House, New Delhi, pp. 459–474.

Resnick, S., 2005. Discussion of the danish data on large fire insurance losses. ASTIN Bulletin 27 (1), 139–152.

Swiss Re, 2011. Natural catastrophes and man-made disasters 2010. Sigma 1/2011, Zurich, Switzerland.

Whitt, W., 2002. Stochastic Process Limits—An Introduction to Stochastic-Process Limits and their Application to Queues. Springer.