The Journal of Risk (63–80) Volume 14/Number 3, Spring 2012

Fitting the generalized Pareto distribution to commercial fire loss severity: evidence from Taiwan

Wo-Chiang Lee
Department of Banking and Finance, Tamkang University, 151 Yin-Chuan Road,
Tamsui District, New Taipei City, Taiwan 25137, Republic of China;
email: wclee@mail.tku.edu.tw

This paper focuses on modeling and estimating tail parameters of loss distributions from Taiwanese commercial fire loss severity. Using extreme value theory, we employ the generalized Pareto distribution (GPD) and compare it with standard parametric modeling based on lognormal, exponential, gamma and Weibull distributions. In an empirical study, we determine the thresholds of the GPD using mean excess plots and Hill plots. Kolmogorov–Smirnov and likelihood ratio goodness-of-fit tests are conducted, and value-at-risk and expected shortfall are calculated. We also construct confidence intervals for the estimates using the bootstrap method.

1 INTRODUCTION

For a non-life insurance company, just a few individual claims made upon a portfolio often make up the majority of the indemnities paid out by the company. Among the largest insurance claims, commercial fire insurance has the highest value. Hence, gaining an understanding of the tail distribution of fire loss severity is useful for the pricing and risk management of a non-life insurance company.

Historical data on loss severity in insurance is often modeled using lognormal, exponential, Weibull and gamma distributions. However, these distributions appear to overestimate or underestimate the tail probability. In terms of fitting the tail of a loss function, a pioneering and well-known work by Hogg and Klugman (1984) focused on fitting the size of loss distributions to the data. They used a truncated Pareto distribution to fit the loss function. However, Boyd (1988) argued that they seriously underestimated the tail region of the fitted loss distribution. Hogg and Klugman compared two methods of estimation, namely, maximum likelihood estimation (MLE) and method of moment. The issue of whether extreme value theory (EVT) or the generalized Pareto distribution (GPD) is better for measuring loss severity has also been discussed extensively in the literature. Several early studies argued that EVT can provide a number of sensible approaches to this problem. Bassi *et al* (1998), McNeil (1997), Resnick (1997), McNeil and Saladin (1997) and Embrechts *et al* (1997, 1999) suggested that it was preferable to use a GPD in order to estimate the tail measure of loss data. Beirlant *et al* (2004) pointed out that insurance loss data usually exhibits heavy tails. They tested the method on a variety of simulated heavy-tailed distributions to show what kinds of thresholds are required and what sample sizes are necessary to give accurate estimates of quantiles. Therefore, it is the key to many risk management problems related to insurance, reinsurance and finance, as shown by Embrechts *et al* (1999).

Furthermore, many early researchers experimented with operational loss data on insurance. Beirlant and Teugels (1992) modeled large claims in non-life insurance using an extreme value model. Zajdenweber (1996) used extreme values in business interruption insurance. Rootzen and Tajvidi (2000) used extreme value statistics to fit wind-storm losses. Moscadelli (2004) showed that the tails of loss distribution functions are, in the first approximation, of heavy-tailed Pareto type. Patrick *et al* (2004) examined the empirical regularities in operational loss data and found that loss data by event type is quite similar across institutions. Nešlehová *et al* (2006) used EVT and the overall quantitative risk management consequences of extremely heavy-tailed data. Chava *et al* (2008) focused on modeling and predicting the loss distribution for credit-risky assets such as bonds or loans. They also analyzed the dependence between the default probabilities and recovery rates and showed that they are negatively correlated. Dahen *et al* (2010) analyzed US bank data and showed that US banks could suffer, on average, more than four major losses a year. They also used the extreme distribution to fit the operational losses and estimated annual insurance premiums. Lee and Fang (2010) focused on modeling and estimating the tail parameters of Taiwan's commercial bank operation loss severity. They also measured the capital for operational risk.

In an early work on fire loss, Mandelbrot (1964) used the random walks concept and some tail distributions to model and discuss fire damage and related phenomena. To measure the loss severity of commercial fire insurance loss, we attempt to answer the following questions. Which techniques fit the loss data statistically and also result in meaningful capital estimates? Are there models that can be considered to be appropriate loss risk measures? How well does the method accommodate a wide variety of empirical loss data?

For the purposes of our empirical study, we measure commercial fire insurance loss using a data-driven loss distribution approach (LDA). By estimating commercial fire loss insurance risk on business-line and event-type levels, we are able to present the estimates in a more balanced fashion. The LDA framework has three essential components: a distribution of the annual number of losses, a distribution of the dollar amount of loss severity and an aggregate loss distribution that combines the two. Strictly speaking, we utilize EVT to analyze the tail behavior of commercial fire insurance loss. The results may help non-life insurance companies to manage their risk. For the purposes of comparison, we consider the following one- and two-parameter distributions to model the loss severity: lognormal, exponential, gamma and Weibull. These were chosen due to their simplicity and applicability to other areas of economics and finance. Distributions such as the exponential, Weibull and gamma are unlikely to fit heavy-tailed data, but provide a nice comparison to heavier-tailed distributions such as the GPD and generalized extreme value (GEV) distribution.

We succeeded fitting the GPD using exceedingly high thresholds of 5.969×10^5, 5.185×10^6 and 2.376×10^7. We show that the GPD can be fitted to commercial fire insurance loss severity. When the loss data exceeds high thresholds, the GPD is a useful method for estimating the tails of loss severity distributions. This means that the GPD is a theoretically well-supported technique for fitting a parametric distribution to the tail of an unknown underlying distribution.

The remainder of the paper is organized as follows. Section 2 introduces EVT and goodness of fit. Section 3 gives some empirical results and analysis. Section 4 gives a few concluding remarks and ideas for future work.

2 EXTREME VALUE THEORY

We now proceed to use EVT to estimate the tail of a loss severity distribution. Extreme event risk is present in all areas of risk management. Whether we are concerned with market, credit, operational or insurance risk, one of the greatest challenges for a risk manager is to implement risk management models that allow for rare but damaging events and permit the measurement of their consequences.

The oldest group of extreme value models is block maxima models. These are models for the largest observations collected from large samples of identically distributed observations. The asymptotic distribution of a series of maxima is modeled, and under certain conditions the distribution of the standardized maximum of the series is shown to converge to the Gumbel, Frechet or Weibull distribution. The GEV distribution is a standard form of these three distributions.

The GPD was developed as a distribution for modeling tails of a wide variety of distributions. Suppose that $F(x)$ is the cumulative distribution function for a random variable x and that threshold μ is a value of x on the right tail of the distribution. The probability that x lies between u and $u + y$, $y > 0$, is $F(u + y) - F(u)$. The probability of x being greater than u is $1 - F(u)$. Define $F_u(y)$ as the probability that x is between u and $u + y$, conditional on $x > u$. We have:

$$F_u(y) = \Pr\{x - u \leqslant y \mid x > u\} = \frac{F(u+y) - F(u)}{1 - F(u)} \tag{2.1}$$

Once the threshold is estimated, the conditional distribution F_u converges to the GPD. We can find a limit $F_u(y) \approx G_{\xi,\sigma(u)}(y)$ as $u \to \infty$ (Pickands (1975) and Balkema and de Haan (1974)):

$$G_{\xi,\sigma(u)}(y) = \begin{cases} 1 - \left(1 + \xi \dfrac{y}{\sigma}\right)^{-1/\xi} & \text{if } \xi \neq 0 \\ 1 - \mathrm{e}^{-y/\sigma} & \text{if } \xi = 0 \end{cases} \tag{2.2}$$

where ξ is the shape parameter and determines the heaviness of the tail of the distribution, and σ is a scale parameter. When $\xi = 0$, the random variable x has a standard exponential distribution. As the tails of the distribution become heavier (or longer tailed), the value of ξ increases. The parameters can be estimated using MLE (for a more detailed description of the model, see Neftci (2000)).

One of the most difficult problems in the practical application of EVT is choosing the appropriate threshold for where the tail begins. The most widely used methods for exploring the data are graphical methods, ie, quantile–quantile (Q–Q) plots, Hill plots and the distribution of mean excess. These methods involve creating several plots of the data and using heuristics to choose the appropriate threshold.

In EVT and its applications, the Q–Q plot is typically plotted against the exponential distribution to measure the fat-tailedness of a distribution (eg, an exponential distribution with a medium-sized tail). If the data is taken from an exponential distribution, the points on the graph would lie along a straight line. If the graph is concave, this indicates a fat-tailed distribution, whereas a convex shape is an indication of a short-tailed distribution. In addition, if the Q–Q plot deviates significantly from a straight line, then either the estimate of the shape parameter is inaccurate or the model selection is untenable.

Selecting an appropriate threshold is a critical problem with the peaks-over-threshold method. There are two graphical tools used to choose the threshold: the Hill plot and mean excess plot. The Hill plot displays an estimate of ξ for different exceedance levels and is the maximum likelihood estimator for a GPD. Hill (1975) proposed the following estimator for ξ. The Hill estimator is the maximum likelihood estimator for a GPD since the extreme distribution converges to a GPD over a high threshold u.

Let $x_1 > \cdots > x_n$ be the ordered statistics of independent and identically distributed random variables. We set $k < n$ and define the Hill estimator of the tail index

$1/\xi$ based on upper-order statistics as:

$$\left.\begin{aligned} H_{k,n} &= \frac{1}{k}\sum_{i=1}^{k}\ln\left(\frac{x_{i,n}}{x_{k+1,n}}\right) \\ \xi &\cong H_{k,n}^{-1} \quad \text{when } n \to \infty,\ k/n \to 0 \end{aligned}\right\} \tag{2.3}$$

The number of upper-order statistics used in the estimation is $k + 1$ and n is the sample size.[1] A Hill plot is constructed such that the estimated ξ is plotted as a function either of k upper-order statistics or of the threshold. More precisely, the Hill graph is defined by the set of points, and hopefully the graph is stable so that a value of ξ can be chosen. The Hill plot also helps us to choose the data threshold and the parameter value. The parameter should be chosen where the plot looks stable:

$$\{(k, H_{k,n}^{-1}),\ 1 \leqslant k \leqslant n\} \tag{2.4}$$

The mean excess plot introduced by Davidson and Smith (1990) graphs the conditional mean of the data above different thresholds. The sample mean excess function (MEF) is defined as:

$$e_{n_u}(u) = \frac{\sum_{i=1}^{n_u}(x_i - u)}{\sum_{i=1}^{n_u} I_{u(x_i > u)}} \tag{2.5}$$

where $I = 1$ if $\xi > u$, and 0 otherwise, and where n_u denotes the number of data points that exceed the threshold u. The MEF is the sum of the excesses over the threshold u divided by n_u. It is an estimate of the MEF that describes the expected overshoot of a threshold once an exceedance occurs. If the empirical MEF has a positive gradient above a certain threshold u, it is an indication that the data follows the GPD with a positive shape parameter ξ. On the other hand, exponentially distributed data would show a horizontal MEF, while short-tailed data would have a negatively sloped line.

Following Equation (2.2), the probability that $x > u + y$ conditional on $x > u$ is $1 - G_{\xi,\sigma(u)}(y)$, while the probability that $x > u$ is $1 - F(u)$, and the unconditional probability that $x > u + y$ is therefore:

$$F(x > u + y) = [1 - F(u)][1 - G_{\xi,\sigma,u}(y)] \tag{2.6}$$

If n is the total number of observations, an estimate of $1 - F(u)$ calculated from the empirical data is n_u/n. The unconditional probability that $x > u + y$ is therefore:

$$\frac{n_u}{n}[1 - G_{\xi,\sigma}(y)] = \frac{n_u}{n}\left(1 + \hat{\xi}\frac{y}{\sigma}\right)^{-1/\hat{\xi}} \tag{2.7}$$

[1] Beirlant *et al* (1996) proposed estimating the optimal k from the minimum value of the sequence of weighted mean square error expressions.

which means that our estimator of the tail for the cumulative probability distribution is:

$$F(x) = 1 - \frac{n_u}{n}\left(1 + \hat{\xi}\frac{x-u}{\sigma}\right)^{-1/\hat{\xi}} \tag{2.8}$$

To calculate value-at-risk (VaR) with a confidence level q it is necessary to solve the equation:

$$F(\text{VaR}) = q$$

From Equation (2.8), we have:

$$q = 1 - \frac{n_u}{n}\left(1 + \hat{\xi}\frac{\text{VaR}-u}{\sigma}\right)^{-1/\hat{\xi}} \tag{2.9}$$

The VaR is therefore:

$$\text{VaR} = u + \frac{\sigma}{\xi}\left(\left(\frac{n}{n_u}(1-q)\right)^{-\xi} - 1\right) \tag{2.10}$$

Expected shortfall (ES) is a concept used in finance and, more specifically, in the field of financial risk measurement to evaluate the market risk of a portfolio. It is an alternative to VaR. The expected shortfall at the $p\%$ level is the expected return on the portfolio in the worst $p\%$ of the cases. For example, $\text{ES}(0.05)$ is the expectation of the worst 5 out of 100 events. Expected shortfall is also called conditional value-at-risk and expected tail loss.

In our case, we define the excess shortfall as the expected loss size, given that VaR is exceeded:

$$\text{ES}_q = E(L \mid L > \text{VaR}_q) \tag{2.11}$$

where $q(= 1 - p)$ is the confidence level. Furthermore, we obtain the following ES estimator:

$$\text{ES}_q = \frac{\text{VaR}_q}{1-\xi} + \frac{\sigma - \xi u}{1-\xi} \tag{2.12}$$

One can attempt to fit any particular parametric distribution to data; however, only certain distributions will have a good fit. There are two ways of assessing this goodness of fit: either by using graphical methods or by using formal statistical goodness-of-fit tests. The former method (a Q–Q plot or a normalized probability–probability (P–P) plot, for example) helps an individual to determine whether a fit is very poor, but may not reveal whether a fit is good in the formal sense of statistical fit. Examples of the latter method are the Kolmogorov–Smirnov (KS) test or the likelihood ratio (LR) test. The Q–Q plot depicts the match or mismatch between the observed values in the data and the estimated value given by the hypothesized fitted distribution. The KS test is a nonparametric supremum test based on the empirical cumulative distribution

TABLE 1 Frequencies of commercial fire loss.

Range of loss amount (NT$)	Number of loss events	Percentage (%)	Sum of loss amount (NT$)	Percentage (%)
0–100 000	2618	62.90	74 154 281	0.63
100 001–200 000	387	9.29	54 611 060	0.46
200 001–500 000	401	9.64	127 755 196	1.08
500 001–1 000 000	198	4.75	143 612 390	1.21
1 000 001–5 000 000	335	8.05	779 265 293	6.57
5 000 001–10 000 000	75	1.81	543 222 505	4.58
⩾10 000 001	148	3.56	10 134 086 981	85.47
Total	4162	100	11 856 707 706	100

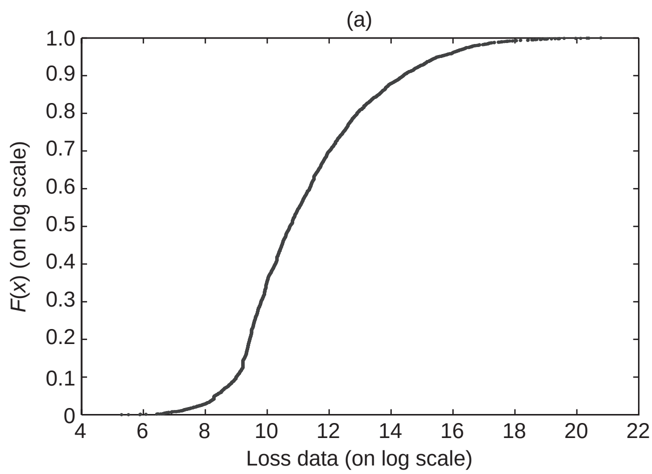
function. The LR test is based on exceedances over a threshold u or on the $k+1$ largest-order statistics. In the GPD model, we test H_0 ($\xi = 0$) against H_1 ($\xi \neq 0$), with unknown scale parameters $\sigma > 0$.

3 EMPIRICAL RESULTS AND ANALYSIS

There are 4612 observations in the data set. All commercial fire insurance loss data sets used in this study were obtained from a non-life insurance company in Taiwan. The data is made up of five years' worth of fire losses. Table 1 reports the frequency and percentage of loss events. The last two columns represent the sum and percentage of loss amounts. The data shows that most loss events have a value of less than NT$100 000 (New Taiwan dollars), whereas, for loss amounts, the figure is above NT$10 000 000, with a percentage of 85.47%.

The empirical distribution in part (a) of Figure 1 on the next page summarizes the cumulated distribution function on a log–log plot of the loss data set. We can ascertain the threshold of the tail distribution with a phenomenological analysis of the figure. For example, for values over 10 (on a log scale), the cumulated probability is near to 1. Part (b) of Figure 1 on the next page shows a scatter plot of loss data. The series indicates that there are several particularly large assessments of loss over NT$1 million. The figure also shows us that the skewness of a loss set lacks symmetry, and positive values for skewness in Table 2 on the next page indicate that data that is skewed to the right (skewness coefficient of 23.113). Right-skewedness means that the right tail is long relative to the left tail. In addition, kurtosis is a measure of whether the data is peaked or flat relative to a normal distribution. The loss data sets with high kurtosis tend to have a distinct peak near the mean, decline rather rapidly and have heavy tails.

FIGURE 1 (a) Empirical distribution of fire loss data and (b) scatter plot of fire loss amount.



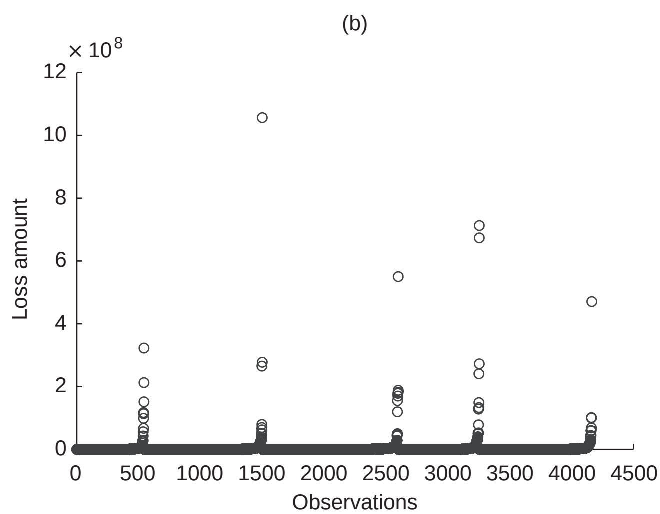
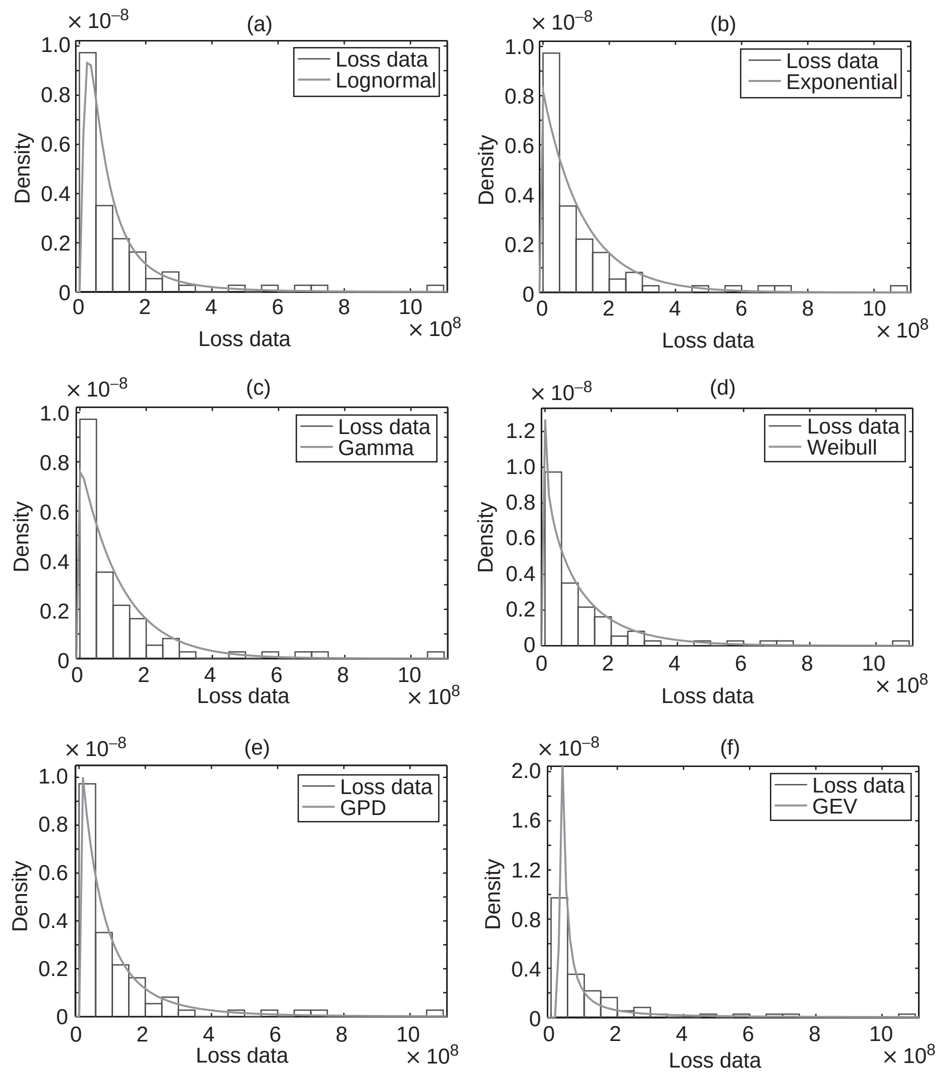


TABLE 2 Summary statistics.

Mean	Standard deviation	Kurtosis	Skewness	Minimum	Maximum	Number of observations
284 800.51	28 623 111	664.794	23.113	199	1.056×10^9	4162

Values in New Taiwan dollars.

FIGURE 2 Probability density function plots of loss amounts.



(a) Lognormal. (b) Exponential. (c) Gamma. (d) Weibull. (e) GPD. (f) GEV.

It is practically impossible to experiment with every possible parametric distribution that we know of. An alternative way of conducting such an exhaustive search could be to fit general class distributions to the loss data in the hope that the distributions will be flexible enough to conform to the underlying data in a reasonable way. For the

TABLE 3 Parametric estimations for fitted functions.

(a)

Distribution	Lognormal	Exponential	Gamma
Loglikelihood	–55 913.3	–66 019.3	–58 236
Parameter 1	$\mu = 11.2174$	$\mu = 2.8488 \times 10^6$	$\xi = 0.20164$
Parameter 2	$\sigma = 2.22117$	—	$\sigma = 1.41281 \times 10^7$

(b)

Distribution	Weibull	GPD	GEV
Loglikelihood	–56 766.8	–55 690	–55 607.6
Parameter 1	$\xi = 245\,204$	$\xi = 1.77364$	$\xi = 1.68294$
Parameter 2	$\lambda = 0.379161$	$\sigma = 40406.7$	$\sigma = 48\,620.7$
Parameter 3	$\mu = 0$	$\mu = 569\,600$	$\mu = 569\,506$

purposes of comparison, we have used lognormal, exponential, Weibull and gamma distributions as a benchmark.

We then fit the probability density function (PDF) plot of the above distributions. Figure 2 on the preceding page shows the poor fit of the exponential, gamma, Weibull and GEV distributions, and shows that other distributions fit the loss data much better, especially the GPD distribution.

Table 3 lists the parametric estimations for fitted functions. The goodness-of-fit loglikelihood value shows that the GEV model is highest, followed by the GPD model, lognormal, Weibull and gamma functions. The exponential function has the lowest value. However, the estimation of the GPD model depends on the choice of threshold. In the following section we discuss the parameter estimation of the GPD further.

We use the GPD model to evaluate the VaR of fire loss severity. The first step is to select the threshold. The MEF plots the sample mean excesses against thresholds. In Figure 3 on the facing page we can see that the mean excess of the fire loss data against threshold values shows an upward sloping MEF. The plot indicates a heavy tail in the sample distribution. At the upward sloping point, we find three segments (for example, in the first segment, the threshold value is almost equal to 5.969×10^5). The other two threshold values are 5.185×10^6 and 2.376×10^7.

The Hill plot in Figure 4 on the facing page displays an estimate of ξ for different exceedances; a threshold is selected from the plot where the shape parameter ξ is fairly stable. The number of upper-order statistics or thresholds can be restricted in order to investigate the stable part of the Hill plot. Figure 5 on page 74 plots the

FIGURE 3 The mean excess function of loss amount.

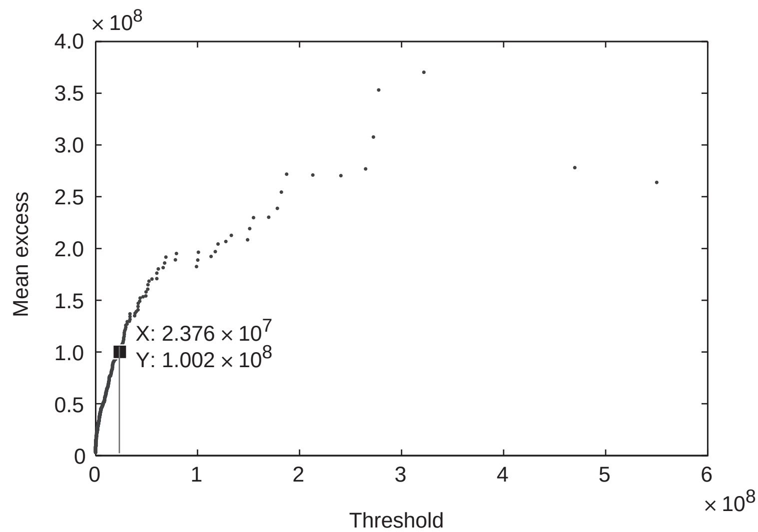


FIGURE 4 The Hill plot of the loss amount.

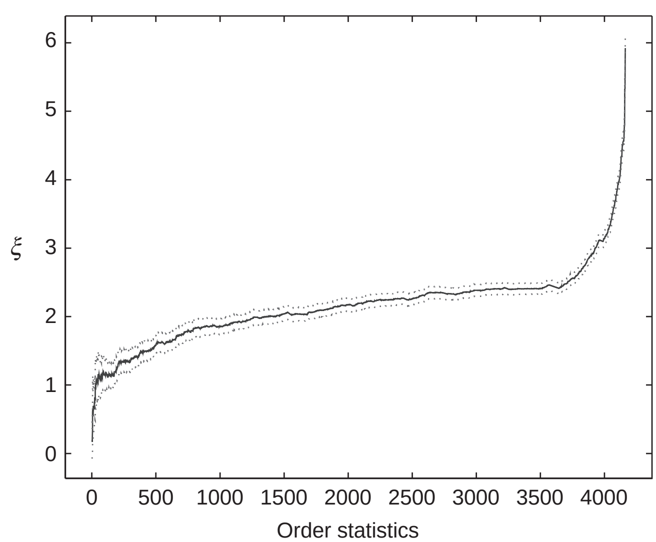
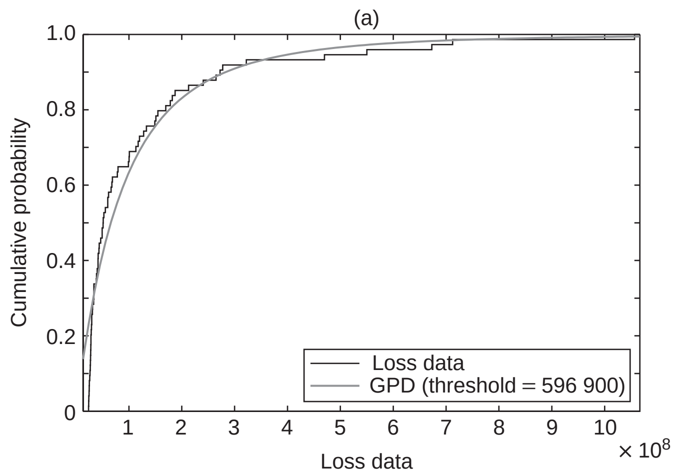
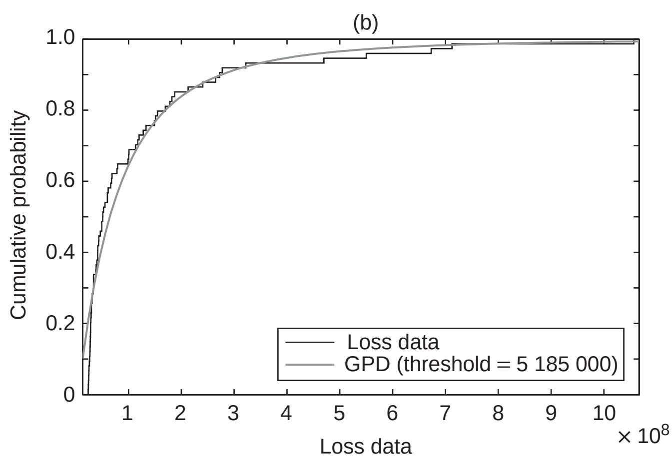


FIGURE 5 Cumulative density function of the estimated GPD model and the loss data over thresholds. [Figure continues on next page.]



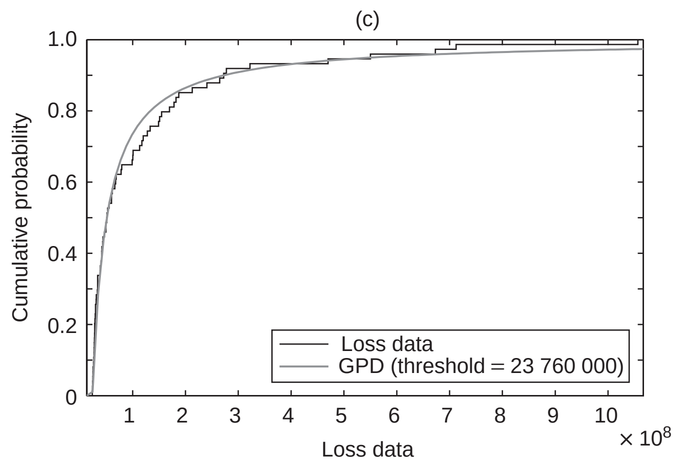


(a) Threshold = 5.969×10^5. (b) Threshold = 5.185×10^6.

cumulative density function of the estimated GPD model and the loss data over three thresholds. We find that the GPD model also fits reasonably well.

Table 4 on the facing page reports some estimate results for the GPD model. For example, when the threshold is set to 5.969×10^5, the number of exceedances is 706. We also calculate the VaR and ES at the 95%, 97.5% and 99% confidence levels using Equations (2.9) and (2.11). The results are also shown in Table 4 on the facing page.

FIGURE 5 Continued.



(c) Threshold = 2.376×10^7.

TABLE 4 VaR and ES of the GPD.

	N_u		
	706	**216**	**74**
Threshold	5.969×10^5	5.185×10^6	2.376×10^7
σ scaling parameter	1.5892×10^6 (1.3256×10^5)	9.9444×10^6 (1.3048×10^6)	2.7023×10^7 (7.2302×10^6)
ξ shape parameter	1.2947 (0.0890)	0.9581 (0.1298)	1.0160 (0.2684)
VaR (95%)	5.3383×10^6	5.5622×10^6	6.4654×10^6
VaR (97.5%)	1.4013×10^7	1.5703×10^7	1.5976×10^7
VaR (99%)	4.7326×10^7	4.5081×10^7	4.4890×10^7
ES (95%)	2.0885×10^7	2.5152×10^8	5.8426×10^8
ES (97.5%)	5.0320×10^7	4.9355×10^8	1.1787×10^9
ES (99%)	1.6336×10^8	1.1947×10^9	2.9858×10^9

Figures in parentheses are standard deviation. N_u denotes the number of exceedances. VaR (95%), VaR (97.5%) and VaR (99%) denotes the value-at-risk at the 95%, 97.5% and 99% confidence levels, respectively. ES (95%) denotes the expected shortfall at the 95% level, and so on.

Table 5 on the next page presents results for the goodness of fit for the GPD model. The fact that The KS test does not reject H_0 at the 5% significance level means that the loss data has a GPD distribution. The P-value of the LR test is smaller than all

TABLE 5 Goodness of fit for the GPD model.

	N of exceedances		
	706	**216**	**74**
Threshold	5.969×10^5	5.185×10^6	2.376×10^7
KS test	1	1	1
(P-value)	(1.0000)	(1.0000)	(1.0000)
LR test	5.3970×10^3	5.3455×10^2	1.2139×10^2
(P-value)	0.0000*	0.0000*	0.0000*

The null hypothesis for the Kolmogorov–Smirnov test is that the loss data has a GPD distribution. The alternative hypothesis is that the loss data does not have that distribution. The asterisk denotes significance at the 5% level.

TABLE 6 Bootstrap confidence intervals for GPD.

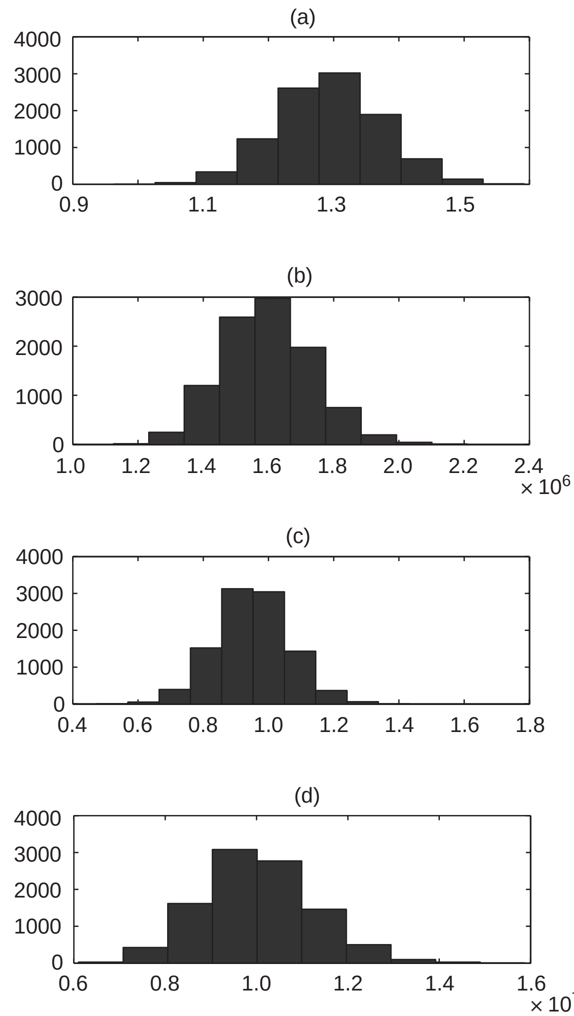
	Threshold		
	$\mathbf{5.969 \times 10^5}$	$\mathbf{5.185 \times 10^6}$	$\mathbf{2.376 \times 10^7}$
σ scaling parameter	$[1.3495, 1.8715] \times 10^6$ (1.5892×10^6)	$[0.7689, 1.2861] \times 10^7$ (0.9444×10^7)	$[1.5995, 4.5654] \times 10^7$ (2.7023×10^7)
ξ shape parameter	[1.1202, 1.4690] (1.2946)	[0.7037, 1.2124] (0.9581)	[0.4900, 1.5420] (1.0160)

Bootstrap confidence intervals at a significance level 5% for parameters. Figures in parentheses are the actual scaling parameter.

the significance levels. It also shows that the GPD is good for model fitting. If the parameters are unknown, but consistently estimated, the bootstrap distribution function is a reliable approximation of the true sampling distribution. We therefore take the bootstrap method into account to estimate the confidence interval of parameters.[2] Table 6 shows the confidence intervals of parameters ξ and σ for the GPD model at the 5% significance level. The results from Table 6 indicate that the bootstrap critical values are consistent estimates of the actual ones. Figure 6 on the facing page shows that the bootstrap estimates for ξ and σ appear acceptably close to normality. The mean values of parameters from bootstrap estimates are close to the actual ones. Hence, the thresholds that we have chosen are optimal and reasonable.

[2] We generate 10 000 duplicate data sets by resampling from y_i (exceedances over the threshold u) to fit the GPD.

FIGURE 6 Histogram of bootstrap for parameter ξ and σ at different thresholds (5.969×10^5, 5.185×10^6 and 2.376×10^7). [Figure continues on next page.]

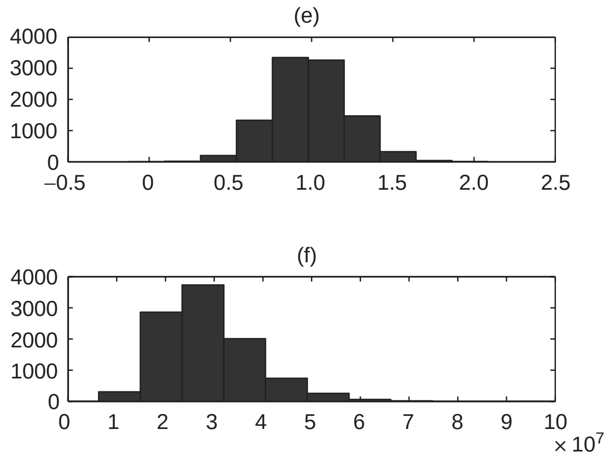


(a) Bootstrap of ξ for 5.969×10^5. (b) Bootstrap of σ for 5.969×10^5. (c) Bootstrap of ξ for 5.185×10^6. (d) Bootstrap of σ for 5.18×10^6.

4 CONCLUDING REMARKS

In many applications of loss data distributions, a key concern is fitting the loss data in the tail. As mentioned above, good estimates of the tails of fire loss severity distributions are essential for pricing and risk management of commercial fire insurance

FIGURE 6 Continued.



(e) Bootstrap of ξ for 2.376×10^7. (f) Bootstrap of σ for 2.376×10^7.

loss. In this paper we have described parametric curve-fitting methods for modeling extreme historical losses using an LDA. We first execute an exploratory loss data analysis using a Q–Q plot of lognormal, exponential, gamma, Weibull, GPD and GEV distributions. The Q–Q plot and loglikelihood function value revealed the exponential and Weibull distribution to be poorly fitted, while other distributions can be seen to fit the loss data much better. Furthermore, we determined the optimal thresholds and parameter value of GPD model using a Hill plot and a mean excess function plot. The Hill plot is gratifyingly stable and concentrated in a narrow range. The selection of thresholds suggested by the MEF plot also provided successful fittings of the GPD. In addition, we also took the bootstrap method into account in order to estimate the confidence interval of parameters. We had some success in fitting the GPD using high thresholds of 5.969×10^5, 5.185×10^6 and 2.376×10^7.

Last but not least, we showed that the GPD can be fitted to commercial fire insurance loss severity. When the loss data exceeds high thresholds, the GPD is a useful method for estimating the tails of loss severity distributions. It also means that the GPD is a theoretically well-supported technique for fitting a parametric distribution to the tail of an unknown underlying distribution.

Finally, we suggest some interesting directions for further research. First, it would be useful to model the tail loss distribution for other forms of insurance. Second, from a risk management viewpoint, constructing a useful management system for avoiding large fire claims would be an interesting line of further research.

REFERENCES

Balkema, A. A., and de Haan, L. (1974). Residual life time at great age. *Annals of Probability* **2**, 792–804.

Bassi, F., Embrechts, P., and Kafetzaki, M. (1998). Risk management and quantile estimation. In *A Practical Guide to Heavy Tails*, Adler, R. J., Feldman, F., and Taqqu, M. (eds), pp. 111–130. Birkhäuser.

Beder, T. S. (1995). VaR: seductive but dangerous. *Financial Analysts Journal* **51**(5), 12–13.

Beirlant, J., and Teugels, J. L. (1992). Modeling large claims in non-life insurance. *Insurance: Mathematics and Economics* **11**, 17–29.

Beirlant, J., Vynckier, P., and Teugels, J. (1996). Excess function and estimation of the extreme values index. *Bernoulli* **2**(4), 293–318.

Beirlant, J., Joossens, E., and Segers, J. (2004). Generalized Pareto fit to the society of actuaries' large claims database. *North American Actuarial Journal* **8**(2), 108–111.

Boyd, V. (1988). Fitting the truncated Pareto distribution to loss distributions. *Journal of the Staple Inn Actuarial Society* **31**, 151–158.

Chava, S., Stefanescu, C., and Turnbull, S. (2008). Modeling the loss distribution. Working Paper. URL: http://faculty.london.edu/cstefanescu/Chava_Stefanescu_Turnbull.pdf.

Cruz, M. G. (2002). *Modeling, Measuring and Hedging Operational Risk*. John Wiley & Sons.

Dahen, H., Dionne, G., and Zajdenweber, D. (2010). A practical application of extreme value theory to operational risk in banks. *The Journal of Operational Risk* **5**(2), 1–16.

Davidson, A. C., and Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B* **52**, 393–442.

Embrechts, P., Kluppelberg, C., and Mikosch, T. (1997). *Modeling Extreme Events for Insurance and Finance*. Springer.

Embrechts, P., Resnick, S. I., and Samorodnitsky, G. (1999). Extreme value theory as a risk management tool. *North American Actuarial Journal* **3**(2), 30–41.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **46**, 1163–1173.

Hogg, R., and Klugman, S. (1984). *Loss Distributions*. John Wiley & Sons.

Lee, W. C., and Fang, C. J. (2010). The measurement of capital for operational risk of Taiwanese commercial banks. *The Journal of Operational Risk* **5**(2), 79–102.

Mandelbrot, B. (1964). Random walks, fire damage and related phenomena. *Operations Research* **12**, 582–585.

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

McNeil, A. J., and Saladin, T. (1997). The peaks over thresholds method for estimating high quantiles of loss distributions. Preprint, Department Mathematik, ETH Zentrum, Zurich.

Moscadelli, M. (2004). The modelling of operational risk: experience with the analysis of the data collected by the Basel Committee. Working Paper no. 517, Bank of Italy.

Neftci, S. N. (2000). Value at risk calculations, extreme events and tail estimation. *Journal of Derivatives* **7**(3), 23–38.

Nešlehová, J., Embrechts, P., and Chavez-Demoulin, V. (2006). Infinite-mean models and the LDA for operational risk. *The Journal of Operational Risk* **1**(1), 3–25.

Patrick, D. F., Jordan, J. S., and Rosengren, E. S. (2004). Implications of alternative operational risk modeling techniques. Working Paper W11103, National Bureau of Economic Research.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119–131.

Resnick, S. I. (1997). Discussion of the Danish data on large fire insurance losses. *ASTIN Bulletin* **27**(1), 139–151.

Rootzen, H., and Tajvidi, N. (2000). Extreme value statistics and wind storm losses: a case study. In *Extremes and Integrated Risk Management*, Embrechts, P. (ed). Risk Books, London.

Zajdenweber, D. (1996). Extreme values in business interruption insurance. *Journal of Risk and Insurance* **63**(1), 95–110.