Applying Generalized Pareto Distribution to the Risk Management of Commerce Fire Insurance

Wo-Chiang Lee
Associate Professor, Department of Banking and Finance
Tamkang University
151,Yin-Chuan Road, Tamsui Taipei County, Taiwan 25137, R.O.C.
TEL:+886-2-26215656 Ext. 3327
FAX:+886-2-26214755
E-mail:wclee@mail.tku.edu.tw

March 30, 2009
Preliminary draft
Comments and suggestions are welcome

Abstract

This paper focus on modeling and estimating tail parameters of commercial fire loss severity. Using extreme value theory, we centralized on the generalized Pareto distribution (GPD) and compare with standard parametric modeling based on Lognormal, Exponential, Gamma and Weibull distributions. In empirical study, we determine the thresholds of GPD through mean excess plot and Hill plot. Kolmogorv-Smirnov and LR goodness-of-fit test are conducted to assess how good the fit is. VaR and expected shortfall are also calculated. We also take into account bootstrap method to estimate the confidence interval of parameters. Empirical results show that the GPD method is a theoretically well supported technique for fitting a parametric distribution to the tail of an unknown underlying distribution. It can capture the tail behavior of commercial fire insurance loss very well.

Keywords: Generalized Pareto distribution, mean excess function, Hill plot
JEL classification : C12, C13, C19
1. Introduction

For a non-life insurance company, a few claims hitting a portfolio usually represent the most part of the indemnities paid by the company. Among the large claim insurance, commerce fire insurance is the most amount. Hence, to understand the tail distribution of fire loss severity can help for pricing and risk management to non-life insurance company.

Historical data on loss severity in insurance are often modeled with Lognormal, Weibull and Gamma distribution. However, these distributions appear overestimating or underestimating in tail probability. In the tail of loss function fit, a pioneer and well-known reference on fitting size of loss distributions to data is Hogg and Klugman (1984). They used truncated Pareto distribution to fit the loss function. Nonetheless, Boyd (1988) pointed out that they underestimated serious the area in the tail of a fitted loss distribution. They compared two estimate methods, namely, Maximum Likelihood Method (MLE) and Method of Moment. Nonetheless, an important question is that which is better of extreme value theory (hence EVT) and generalized Pareto distribution (hence GPD) for loss severity?

Several early studies have argued that EVT motivates 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) also suggested to use GPD distribution in the estimation of tail measure of loss data. Cebrian et al. (2003) pointed out that insurance loss data usually present heavy tails. They test 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).

What is more, many early researchers have experimented with operational loss data or insurance. Beirlant and Teugels (1992) modeled large claims in non-life insurance with extreme value model. Zajdenweber (1996) used extreme values in business interruption insurance. Rootzen and Tajvidi (2000) used extreme value statistics to fit the wind storm losses. Moscadelli (2004) showed the tails of the loss distribution functions are in first approximation heavy-tailed Pareto-type. Patrick et al. (2004) examined the empirical regularities in operational loss data and found that loss data by event types are quite similar across institutions. Neštěhová et al. (2006) used extreme value theory and the overall quantitative risk management consequences of extremely heavy-tailed data. Chava et al. (2008) focuses on modeling and predicting the loss distribution for credit risky assets such as bonds or loans. They also analyze the dependence between the default probabilities and recovery rates and showed that they are negatively correlated.

To measure the loss severity of commercial fire insurance loss, we want to reply the following questions. Which techniques fit the loss data statistically and also result in meaningful capital estimates? Are there models that can be considered appropriate loss risk measures? How well is the method able to reasonably accommodate a wide variety of empirical loss data? Which goodness-of-fit criteria are most appropriately used for loss data?

For the purpose of empirical study, we want to measure the commercial fire insurance loss towards a data-driven loss distribution approach (hence LDA). By estimating commercial fire loss insurance risk at the business line and event type levels as well, 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.
Concretely speaking, we utilize extreme value theory to analyze the tail behavior of commercial fire insurance loss. The result would help non-life insurance company to manage its risk. For the sake of comparison, we considered the following one and two parameter distributions to model the loss severity, namely Lognormal, Exponential, Gamma, and Weibull distribution. They were chosen because of their simplicity and applicability to other areas of economics and finance. These distributions such as the Exponential, Weibull, and Gamma are unlikely to t heavy-tailed data, but provide a nice comparison to the heavier-tailed distributions such as the GPD and generalized extreme value (hence GEV).

The remainder of the paper is as follows. Section 2 introduces the extreme value theory and goodness of fit. Section 3 is empirical result and analysis. Followed by a few concluding remarks and ideas on future work.

2. Extreme Value Theory

We go further to use extreme value theory to estimate the tail of 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 to the risk manager is to implement risk management models which allow for rare but damaging events, and permit the measurement of their consequences.

The oldest group of extreme value models is the 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 distributions. A standard form of these three distributions is called the GEV distribution.

The generalized Pareto distribution was developed as a distribution that can model 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 in 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 for x greater than u is 1-F(u). Define F_u(y) as the probability that x between u and u+y conditional on x>u. We have

\[ F_u(y) = \Pr[x \leq y | x > u] = \frac{F(u+y)-F(u)}{1-F(u)} \] (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; Balkema and de Haan, 1974).

\[ G_{\xi,\sigma(u)}(y) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma} \right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}} & \text{if } \xi = 0 \end{cases} \] (2)

where the two parameters, \( \xi \) is shape parameter and determines the heaviness of the tail of
the distribution, $\sigma$ 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 long-tailed), the value of $\xi$ increases. The parameters can be estimated using maximum-likelihood methods.\(^1\)

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 to explore the data is graphical methods, i.e. 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 the extreme value theory and applications, the Q-Q plot is typically plotted against the exponential distribution to measure the fat-tailness of a distribution.\(^2\) If the data is from an exponential distribution, the points on the graph would lie along a straight line. If there is a concave presence, this would indicate a fat-tailed distribution, whereas a convex departure is an indication of short-tailed distribution. In addition, if there is a stronger deviation of the Q-Q plot 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 POT method. There are two graphical tools used to choose the threshold, the Hill plot and mean excess plot. The Hill plot, which displays an estimate of $\xi$ for different exceedance levels and is the maximum likelihood estimator for a GPD. Hill (1975) proposed the following estimator for $\xi$. The Hill estimator is the maximum likelihood estimator for a GPD and since the extreme distribution converges to a GPD over a high threshold $u$.

Let $x_1 \geq \ldots \geq x_n$ be the ordered statistics of random variables iid. We pick $k < n$ and define the Hill estimator of the tail index $1/\xi$ based on upper order statistics to be

\[
H_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \ln \left( \frac{x_{i,n}}{x_{i+k,n}} \right)
\]

\[
\xi \equiv H_{k,n}^{-1} \quad \text{when} \quad n \to \infty, \quad k/n \to 0
\]

The number of upper order statistics used in the estimation is $k+1$ and $n$ is the sample size.\(^3\) A Hill-plot is constructed such that estimated $\xi$ is plotted as a function of either $k$ upper order statistics or of the threshold. More precisely, the Hill graph is defined by the set of points and hope the graph looks stable so you can pick out a value of $\xi$. On the other side, the Hill plot helps us to choose the data threshold and the parameter value as well. The parameter should be chosen where the plot looks stable.

\[
\{(k, H_{k,n}^{-1}) : k \leq n\}
\]

The mean excess plot, introduced by Davidson and Smith (1990), which graphs the

---

\(^1\) For more detailed description of the model, see Neftci (2000).

\(^2\) i.e, a distribution like exponential with a medium-sized tail.

\(^3\) Beirlant et al. (1996) proposed to estimate the optimal $k$ from the minimum value of the sequence of weighted mean square error expressions.
conditional mean of the data above different thresholds, the sample mean excess function (MEF) which is defined by

$$e_{n_u}(u) = \frac{\sum_{i=1}^{n_u} (x_i - u)}{\sum_{i=1}^{n_u} I_{u(x_i > u)}}$$  \hspace{1cm} (5)

where $I = 1$ if $x_i > u$ and 0, otherwise. $n_u$ : the number of data points which 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 mean excess function which 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 $\xi$. 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), the probability that $x > u + y$ conditional that $x > u$ is $1 - G_{\xi,\sigma,u}(y)$. Whereas the probability that $x > u$ is $1 - F(u)$ and the unconditional probability that $x > u + y$ is therefore

$$F(x > u + y) = [1 - F(u)][1 - G_{\xi,\sigma,u}(y)]$$ \hspace{1cm} (6)

If $n$ is 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

$$\frac{n_u}{n}[1 - G_{\xi,\sigma}(y)] = \frac{n_u}{n}(1 + \frac{\xi}{\sigma} \frac{y}{\sigma})^{-\frac{1}{\xi}}$$ \hspace{1cm} (7)

Which means that our estimator of the tail for cumulative probability distribution is?

$$F(x) = 1 - \frac{n_u}{n}(1 + \frac{\xi}{\sigma} \frac{x - u}{\sigma})^{-\frac{1}{\xi}}$$ \hspace{1cm} (8)

To calculate VaR with a confidence level $q$, it is necessary to solve the equation

$$F(VaR) = q$$

From equation (8), we have

$$q = 1 - \frac{n_u}{n}(1 + \frac{\xi}{\sigma} \frac{VaR - u}{\sigma})^{-\frac{1}{\xi}}$$ \hspace{1cm} (9)

The VaR is therefore

$$VaR = u + \frac{\sigma}{\xi}\left(\left(\frac{n}{n_u}(1 - q)^{-\frac{1}{\xi}}\right)^{-\frac{1}{\xi}} - 1\right)$$ \hspace{1cm} (10)
Expected shortfall (hence 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 p% level is the expected return on the portfolio in the worst p% of the cases. For example, $ES_{0.05}$ is the expectation in the worst 5 out of 100 events. Expected shortfall is also called Conditional Value at Risk (CVaR) and Expected Tail Loss (ETL).

In our case, we define the excess shortfall as the excepted loss size, given that VaR is excess as equation (11)

$$ES_q = E(L | L > VaR_q)$$  \hspace{1cm} (11)

where q (=1-p) is the confidence level. Furthermore, we could get the following expected shortfall estimator

$$ES_q = \frac{VaR_q}{1-\xi} + \frac{\sigma - \xi \mu}{1-\xi}$$  \hspace{1cm} (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: using graphical methods and using formal statistical goodness-of-fit tests. The former such as a Q-Q plot or a normalized probability-probability (P-P) plot can help an individual determine whether a fit is very poor, but may not reveal whether a fit is good in the formal sense of statistical fit. The latter like the Kolmogorov-Smirnov (K-S) test and the Likelihood ratio (LR) test. Further to say, 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 K-S test is a non-parametric supremum test based on the empirical cumulative distribution function (CDF). Likelihood ratio test is based on exceedances over a threshold \(u\) or on the \(k+1\) largest order statistics. Within the GPD model, one tests $H_0: \xi = 0$ 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 datasets used in this study are obtained from a non-life insurance company in Taiwan. The data contain five year fire loss amount. Table 1 reports the frequency of loss event and its percentage; the last two columns represent the sum of loss amount and its percentage. The data shows that most of loss event centralized less than NT$100000, whereas for loss amount is centralized above NT$1000000 with a percentage of 85.47%. 4

The empirical distribution in Figure 1(a) is to graphically summarize the distribution of loss data set. We find that it indications of the strong skewed right and long tailed for the data. It also shows an asymmetric shape. Figure 1(b) shows the scatter plot of loss data. The series indicate that there several particularly large assessments of loss over NT$1 million. The

4 NTS is New Taiwan dollars.
figure also tell us, the skewness of loss set lacks of symmetry and positive values for the skewness in table 2 indicate data that are skewed right. (skewness coefficient of 23.113). Skewed right means that the right tail is long relative to the left tail. In addition, kurtosis is a measure of whether the data are 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.

It is practically impossible to experiment with every possible parametric distribution that we know of. An alternative way to conduct such an exhaustive search could be to fit general class distributions to the loss data with the hope that these distributions are flexible enough to confirm to the underlying data in a reasonable way. For the purposes of comparison, we have used Lognormal, Exponential, Weibull and Gamma distribution as a benchmark.

As mentioned above, Q-Q plot can help an individual determine how good or poor a fit is, but may not reveal whether a fit is good in the formal sense of statistical fit. In the following Figure 2(a) to 2(f), represents the Exponential and Weibull distribution poor fitted. Whereas for other distributions can fit the loss data much better.

Table 3 lists parametric estimations for fitted functions, the goodness of fit, log-likelihood value shows that GEV model is highest, and then is GPD model, Lognormal, Weibull and Gamma function. The lowest value is Exponential function. However, the estimation of GPD model depends on the choice of threshold. In the following section, we will go further to discuss the parameters estimation of GPD.

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. We see the Figure 3 and find that the mean exceed of the fire loss data against threshold values show an upward sloping mean excess function as in this plot indicates a heavy tail in the sample distribution. At the upward sloping point, we find three segments, for example, in the first segment, threshold value almost equal to 5.969e+5. We further to find the other two threshold values 5.185e+6 and 2.376e+7, respectively.

The Hill plot in Figure 4 displays an estimate of $\xi$ for different exceedances; a threshold is selected from the plot where the shape parameter $\xi$ is fairly stable. The number of upper order statistics or thresholds can be restricted to investigate the stable part of the Hill-plot. Whereas Figure 5 plots the cumulative density function of the estimated GPD model and the loss data over three thresholds. We also find that the GPD model fitted reasonable well.
Table 4 reports some estimate results of GPD model. For example, when the threshold is set to 5.969e+5, the number of exceedances is 706. We also calculate the VaR and ES at 95%, 97.5%, and 99% confidence level by equation (9) and equation (11), and the results are also shown in Table 4.

Table 5 represents the goodness-of-fit for GPD model. The K-S test does not reject Ho at 5% significance level means that the loss data has a GPD distribution. The P-value of LR test are smaller than all the significantly levels. It also shows that 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. Therefore, we further take into account bootstrap method to estimate the confidence interval of parameters.\textsuperscript{5} Table 6 shows the confidence intervals of parameter $\xi$ and $\sigma$ for GPD model at the significance level 5%. The results from Table 6 indicate that the bootstrap critical values are consistent estimates of the actual ones. The figure also shows that the bootstrap estimates for $\xi$ and $\sigma$ appear acceptably close to normality. These mean of parameters from bootstrap estimate are close to the actual ones. In a word, these thresholds we choose are optimal and reasonable.

4. Concluding Remarks

In many applications, fitting the loss data in the tail is the main concern. As mentioned above, good estimates for the tails of fire loss severity distributions are essential for pricing and risk management of commercial fire insurance loss. In this paper, we describe parametric curve-fitting methods for modeling extreme historical losses by way of LDA. We first execute exploratory loss data analysis using Q-Q plot of Lognormal, Exponential, Gamma, Weibull, GPD and GEV distributions. The Q-Q plot and log-likelihood function value reveal that the Exponential and Weibull distribution are poor fitted. Whereas for other distributions can fit the loss data much better. Furthermore, we determine the optimal thresholds and parameter value of GPD model with Hill plot and the mean excess function plot. The Hill plot is gratifyingly stable and is in a tight neighborhood. The sample mean excess function for the loss data suggests. In addition, we also take into account bootstrap method to estimate the confidence interval of parameters. We may have success fitting the GPD using exceed high thresholds of 5.969e+5, 5.185e+6 and 2.376e+7, respectively.

Last but not the last, we show that GPD can fitting to commercial fire insurance loss severity. Which exceed high thresholds is a useful method for estimating the tails of loss severity distributions. It also means that GPD is a theoretically well supported technique for fitting a parametric distribution to the tail of an unknown underlying distribution.

With regard to this topic, we would like to suggest some interesting directions for further research. First, we will try to model the tail loss distribution of other insurance. Second, from

\textsuperscript{5} We generate 10000 replicate datasets by resampling from $y_i$ (exceedances over the threshold $u$) to fit GPD.
risk management viewpoint, how to construct a useful management system for avoiding large claim is in the further work.

References


### Table 1 Frequencies of commerce fire loss

<table>
<thead>
<tr>
<th>Range of loss amount (NT$)</th>
<th>Number of loss event.</th>
<th>Percentage (%)</th>
<th>Sum of loss amount</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-100000</td>
<td>2618</td>
<td>0.6290</td>
<td>74,154,281</td>
<td>0.6254</td>
</tr>
<tr>
<td>100001-200000</td>
<td>387</td>
<td>0.0929</td>
<td>54,611,060</td>
<td>0.4605</td>
</tr>
<tr>
<td>200001-500000</td>
<td>401</td>
<td>0.0963</td>
<td>127,755,196</td>
<td>1.0774</td>
</tr>
<tr>
<td>500001-1000000</td>
<td>198</td>
<td>0.0475</td>
<td>143,612,390</td>
<td>1.2112</td>
</tr>
<tr>
<td>1000001-5000000</td>
<td>335</td>
<td>0.0804</td>
<td>779,265,293</td>
<td>6.5723</td>
</tr>
<tr>
<td>5000001-10000000</td>
<td>75</td>
<td>0.0180</td>
<td>543,222,505</td>
<td>4.5185</td>
</tr>
<tr>
<td>&gt; 100000000</td>
<td>148</td>
<td>0.0355</td>
<td>10,134,086,981</td>
<td>85.4713</td>
</tr>
<tr>
<td>total</td>
<td>4162</td>
<td>100</td>
<td>11,856,707,706</td>
<td>100</td>
</tr>
</tbody>
</table>

Note: Value in New Taiwan dollars (NT$)

### Table 2 Summary statistics

<table>
<thead>
<tr>
<th>Items</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Kurtosis</th>
<th>Skewness</th>
<th>Minimum</th>
<th>Maximum</th>
<th>N of Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amount</td>
<td>284800.51</td>
<td>28623111</td>
<td>664.794</td>
<td>23.113</td>
<td>199</td>
<td>1.056*10^9</td>
<td>4162</td>
</tr>
</tbody>
</table>

Note: Value in New Taiwan dollars; Std. Dev. is standard deviation.

### Table 3 Parametric estimations for fitted functions

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Lognormal</th>
<th>Exponential</th>
<th>Gamma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Likelihood</td>
<td>-55913.3</td>
<td>-66019.3</td>
<td>-58236</td>
</tr>
<tr>
<td>Mean</td>
<td>876934</td>
<td>2.8488e+6</td>
<td>2.8488e+6</td>
</tr>
<tr>
<td>Variance</td>
<td>1.0603e+014</td>
<td>8.11566e+12</td>
<td>4.02482e+013</td>
</tr>
<tr>
<td>Parameter-1</td>
<td>μ=11.2174</td>
<td>μ=2.8488e+6</td>
<td>ξ=0.20164</td>
</tr>
<tr>
<td>Parameter-2</td>
<td>σ=2.22117</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Weibull</th>
<th>GPD</th>
<th>GEV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-Likelihood</td>
<td>-56766.8</td>
<td>-55690</td>
<td>-55607.6</td>
</tr>
<tr>
<td>Mean</td>
<td>951188</td>
<td>Inf</td>
<td>inf</td>
</tr>
<tr>
<td>Variance</td>
<td>1.07038e+013</td>
<td>Inf</td>
<td>inf</td>
</tr>
<tr>
<td>Parameter-1</td>
<td>ξ=245204</td>
<td>ξ=1.77364</td>
<td>ξ=1.68294</td>
</tr>
<tr>
<td>Parameter-2</td>
<td>λ=0.379161</td>
<td>σ=40406.7</td>
<td>σ=48620.7</td>
</tr>
<tr>
<td>Parameter-3</td>
<td>μ=0</td>
<td>μ=199</td>
<td>μ=28228.2</td>
</tr>
</tbody>
</table>
### Table 4: The VaR and ES of GPD

<table>
<thead>
<tr>
<th>N&lt;sub&gt;n&lt;/sub&gt;</th>
<th>706</th>
<th>216</th>
<th>74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>5.969e+5</td>
<td>5.185 e+6</td>
<td>2.376e+7</td>
</tr>
<tr>
<td>σ: scaling parameter</td>
<td>1.5892e+006</td>
<td>9.9444e+006</td>
<td>2.7023e+007</td>
</tr>
<tr>
<td>ξ: shape parameter</td>
<td>(1.3256e+005)&lt;sup&gt;&lt;i&gt;b&lt;/i&gt;&lt;/sup&gt;</td>
<td>(1.3048e+006)</td>
<td>(7.2302e+006)</td>
</tr>
<tr>
<td>VaR (95%)</td>
<td>5.3383e+006</td>
<td>5.5622e+006</td>
<td>6.4654e+006</td>
</tr>
<tr>
<td>VaR (97.5%)</td>
<td>1.4013e+007</td>
<td>1.5703e+007</td>
<td>1.5976e+007</td>
</tr>
<tr>
<td>VaR (99%)</td>
<td>4.7326e+007</td>
<td>4.5081e+007</td>
<td>4.4890e+007</td>
</tr>
<tr>
<td>ES (95%)&lt;sup&gt;d&lt;/sup&gt;</td>
<td>2.0885e+007</td>
<td>2.5152e+008</td>
<td>5.8426e+008</td>
</tr>
<tr>
<td>ES (97.5%)</td>
<td>5.0330e+007</td>
<td>4.9355e+008</td>
<td>1.1787e+009</td>
</tr>
<tr>
<td>ES (99%)</td>
<td>1.6336e+008</td>
<td>1.1947e+009</td>
<td>2.9858e+009</td>
</tr>
</tbody>
</table>

**Note:**
- a: N<sub>n</sub>: the number of exceedances.
- b: Figures in parentheses are standard deviation.
- c: VaR (95%), VaR (97.5%) and VaR (99%) represent the Value at Risk at the 95%, 97.5% and 99% confidence level, respectively.
- d: ES (95%) is the expected shortfall at 95% level, and so on.

### Table 5: Goodness-of-fit for GPD model

<table>
<thead>
<tr>
<th>N of exceedances</th>
<th>706</th>
<th>216</th>
<th>74</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>5.969e+5</td>
<td>5.185 e+6</td>
<td>2.376e+7</td>
</tr>
<tr>
<td>K-S Test&lt;sup&gt;a&lt;/sup&gt; (P-value)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1 (1.0000)</td>
<td>1 (1.0000)</td>
<td>1 (1.0000)</td>
</tr>
<tr>
<td>LR Test&lt;sup&gt;b&lt;/sup&gt; (P-value)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.0000*</td>
<td>0.0000*</td>
<td>0.0000*</td>
</tr>
</tbody>
</table>

**Note:**
- a. The null hypothesis for the Kolmogorov-Smirnov test is that loss data has a GPD distribution. The alternative hypothesis that loss data does not have that distribution.
- b. denote: significance at the 5% level, does not reject Ho.
- c. * denote: significance at the 5% level.

### Table 6: Bootstrap confidence intervals for GPD

<table>
<thead>
<tr>
<th>Threshold</th>
<th>5.969e+5</th>
<th>5.185 e+6</th>
<th>2.376e+7</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ: scaling parameter</td>
<td>[1.3495, 1.8715]e+6</td>
<td>[0.7689, 1.2861]e+7</td>
<td>[1.5995, 4.5654]e+7</td>
</tr>
<tr>
<td>ξ: shape parameter</td>
<td>(1.5892e+006)&lt;sup&gt;b&lt;/sup&gt;</td>
<td>(0.9444e+007)</td>
<td>(2.7023e+007)</td>
</tr>
<tr>
<td>parameter</td>
<td>[1.1202, 1.4690]</td>
<td>[0.7037, 1.2124]</td>
<td>[0.4900, 1.5420]</td>
</tr>
<tr>
<td>ξ: shape parameter</td>
<td>(1.2946)&lt;sup&gt;c&lt;/sup&gt;</td>
<td>(0.9581)</td>
<td>(1.0160)</td>
</tr>
</tbody>
</table>

**Note:**
- a. Bootstrap confidence intervals at a significance level 5% for parameters.
- b. Figures in parentheses are actual scaling parameter.
- c. Figures in parentheses are actual true scaling parameter.
Figure 1(a) The Empirical distribution fire loss data  
Figure 1(b) The scatter plot of fire loss amount

Figure 2(a) The Logistic Q-Q plot of loss amount  
Figure 2(b) The Exponential Q-Q plot of loss amount

Figure 2(c) The Gamma Q-Q Plot of Loss Amount  
Figure 2(d) The Weibull Q-Q Plot of Loss Amount

Figure 2(e) The GPD Q-Q plot of loss amount  
Figure 2(f) The GEV Q-Q plot of loss amount
Figure 3 The mean excess function of loss amount

Figure 4 The Hill plot of loss amount

Figure 5 Cumulative density function of the estimated GPD model and the loss data over thresholds (5.969e+5, 5.185e+6, 2.376e+7)
Figure 6 Histogram of bootstrap for parameter ξ and σ at different thresholds (5.969e+5, 5.185e+6, 2.376e+7)
應用一般化柏拉圖分配於商業火災保險之風險管理

李沃牆
淡江大學財金系副教授
251 台北縣新北市新莊區 151 號
Associate Professor, Department of Banking and Finance
TEL:+886-2-26215656 Ext. 3327
FAX:+886-2-26214755
E-mail:wclee@mail.tku.edu.tw

March 30, 2009
Preliminary draft
Comments and suggestions are welcome

摘要

本研究主要重點在於商業火災險損失的分配建構及尾端參數的估計。透過極端值理論，本文應用一般化柏拉圖分配(簡稱 GPD)並與 Lognormal, Exponential, Gamma 及 Weibull 等參數化分配模型進行比較。在實證研究中，應用平均餘額函數及 Hill plot 來挑選門檻值，並透過 Kolmogorv-Smirnov 及 LR 適合度檢定來決定最適門檻值，同時也計算損失的風險值及預期損失；最後，也應用拔靴複製法來估計參數的信賴區間。實證結果顯示，GPD 模型不僅能夠配適未知分配的尾端行為，亦能夠合理配適商業火災險損失分配。

關鍵字：一般化柏拉圖分配，平均餘額函數，Hill plot
JEL 分類: C12, C13, C19