应用概率统计 第二十八卷
第四期 2012年8月

Chinese Journal of Applied Probability
and Statistics Vol.28 No.4 Aug. 2012

A Statistical Model of Chinese Earthquake Loss Distribution *

Cai Quan Lin Zhengyan
(*Department of Mathematics, Zhejiang University, Hangzhou, 310027*)

Abstract

The estimation of loss distribution is always a big issue for insurance companies. Several parametric or nonparametric methods are introduced to fit loss distributions. In this paper, we propose a method by combining both parametric and nonparametric methods to solve this problem. We first determine the threshold between large and small losses by observing the graph of mean excess function, then use the generalized Pareto distribution, the parametric method, to fit excess data, and use kernel density estimation, the nonparametric method, to fit the distribution below threshold. Finally, we use a data set about Chinese annual earthquake loss to compare this method with other existing methods.

Keywords: Loss distribution, heavy tail, generalized Pareto, mean excess function, kernel density estimation.

AMS Subject Classification: 62N02.

§1. Introduction

In non-life insurance, estimation of loss distribution is a fundamental part of the business. In most situations, losses are small, and extreme losses are rarely observed, but the number and the size of extreme losses can have a substantial influence on the profit of the company. Therefore, it is of great importance to fit the loss distribution precisely, especially the tail part, since the high risk of tail part is transferred from insurance companies to reinsurance companies.

The most frequently discussed nonparametric method is kernel density estimation (KDE), with variety of local KDE, variable KDE and transformed KDE. Since the distributions of losses are often highly skewed and heavy tailed, several kinds of transformed KDE were discussed. The shifted power transformation was introduced by Wand et al. (1991).

*The project supported by National Natural Scince Foundation of China (11171303) and the Doctoral Program of Higher Education of China (20090101110020).

Received December 9, 2011. Revised February 13, 2012.

They showed that the classical kernel density estimator was improved substantially by applying a transformation and suggested the shifted power transformation family. Bolancé et al. (2003) improved the shifted power transformation for highly skewed data by proposing an alternative parameter selection algorithm. The Mobius-like transformation was introduced by Clements et al. (2003). Unlike the shifted power transformation, the Mobius-like transformation transforms $(0,\infty)$ into $(-1,1)$ and this method is designed to avoid boundary problems. Kernel density estimation with Champernowne transformation was discussed by Buch-Larsen et al. (2005). They showed advantage by transforming data through Champernowne function and divided data into three parts: big loss, middle loss and small loss. Among all these transformed KDE, Wand's shifted power transformation has the simplest form and has a good fitting outcome.

Another section of method is to apply extreme value theory to handle this issue. Beirlant and Teugels (1992), Embrechts and Kluppelberg (1993) have argued that extreme value theory (EVT) motivates a number of sensible approaches to this problem. In particular, the peak over threshold (POT) model has been advocated by Rootzén and Tajvidi (1997), McNeil and Saladin (1997) to choose the threshold. In the POT model, the excess losses over high thresholds are modeled with the generalized Pareto distribution (GPD). This distribution arises naturally in a key limit theorem in EVT.

In this paper, we combine this two different skills to solve the tail fitting problem, first use extreme value theory, the POT model to separate the data set into two parts, big loss and small loss, both of them are easy to fit. Then apply the generalized Pareto distribution to fit big loss data and kernel density estimation, transformed if necessary, to fit small loss data. Section 2 describes the methods used in fitting loss. In Section 3, an empirical study is presented and Section 4 is an error measurement of the method. Finally, Section 5 outlines the main conclusions.

§2. Methods

2.1 Classical Kernel Density Estimation

Suppose we have a sequence of independent and identically distributed observations $x_1,\ldots,x_n$, from an unknown density f. The kernel density estimator is

$$\widehat{f}_h(x)=\frac{1}{nh}\sum_{i=1}^{n}K\Big(\frac{x-x_i}{h}\Big)=:\frac{1}{n}\sum_{i=1}^{n}K_h(x-x_i), \tag{2.1}$$

where $K(\cdot)$ is the kernel function, $h=h_n$ is the bandwidth.

A number of kernel functions are commonly used: the probability density function (pdf) of uniform, triangular, biweight, triweight, Epanechnikov, Gaussian and others. In this paper we use Gaussian kernel:

$$K(t)=\frac{1}{\sqrt{2\pi}}\exp\Big(\frac{-t^2}{2}\Big),$$

which is easy to apply and performs well. The bandwidth is a free parameter, which exhibits a strong influence on the resulting estimate. The most common criterion used to select this parameter is the expected L_2 risk function, also known as the mean integrated squared error (MISE). Common rules to select bandwidth are rule of thumb (particularly for Gaussian kernel), unbiased cross-validation, biased cross-validation, direct plug-in and solve-the-equation rules. We try all these methods to select a best bandwidth for our estimation.

2.2 Transformation and Kernel Density Estimation

Classical kernel density estimation does not perform well when the true density is asymmetric. The lack of information in the right tail of the domain makes it difficult to obtain a reliable nonparametric estimate of the density in that area. Different papers have proposed different transformed kernel estimation methods of density function, based on parametric families (see [2], [3], [4] and [5]).

Let $g(\cdot)$ be an increasing and monotonic transformation function that has a first derivative $g'(\cdot)$. The shifted power transformation family used by Wand et al. (1991) is

$$g_\lambda(x)=\begin{cases}(x+\lambda_1)^{\lambda_2}, & \lambda_2\neq 0;\\ \ln(x+\lambda_1), & \lambda_2=0,\end{cases}\tag{2.2}$$

where $\lambda_1>-\min(x_1,\ldots,x_n)$ and $\lambda_2\leq 1$ for right-skewed distribution. This approach has a simple mathematical formulation and works particularly well for asymmetric distributions. In order to estimate the optimal parameters of the shifted power transformation function, the algorithm described in [3] can be used.

Let us assume a sample of n independent and identically distributed observations $x_1,\ldots,x_n$ is available. We also assume that a transformation function $g(\cdot)$ has been selected so that the data can be transformed by $y_i=g(x_i), i=1,\ldots,n$. We denote the transformed sample by $y_1,\ldots,y_n$.

Having the transformed data, we then estimate the density of the transformed data set using the classical kernel density estimator

$$\widehat{f}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} K_h(y - y_i) = \frac{1}{nh}\sum_{i=1}^{n} K\Big(\frac{y - y_i}{h}\Big). \tag{2.3}$$

So the estimator of the original density is obtained by back-transformation

$$\widehat{f}_X(x,\lambda) = g'_\lambda(x)\widehat{f}_Y(y,\lambda) = g'_\lambda(x) n^{-1}\sum_{i=1}^{n} K_h(g_\lambda(x) - g_\lambda(X_i)). \tag{2.4}$$

2.3 Selecting the Transformation Parameters and the Bandwidth

The key of shifted power transformation is to select the parameters $\lambda = (\lambda_1, \lambda_2)$. We restrict the set of transformation parameter $\lambda = (\lambda_1, \lambda_2)$ to those values that give approximately zero skewness for the transformed data $y_1, \ldots, y_n$.

We define skewness as

$$\gamma_y = \Big\{n^{-1}\sum_{i=1}^{n}(y_i - \overline{y})^3\Big\}\Big/\Big\{n^{-1}\sum_{i=1}^{n}(y_i - \overline{y})^2\Big\}^{3/2},$$

where $\overline{y}$ is the sample mean.

To select the parameter vector λ, we aim at minimizing the mean integrated square error (MISE) of $\widehat{f}_Y(y)$

$$\mathrm{MISE}_Y(\widehat{f}_Y) = \mathsf{E}\Big[\int_{-\infty}^{+\infty}(\widehat{f}_Y(y) - f_Y(y))^2 \mathrm{d}y\Big], \tag{2.5}$$

which, when h is asymptotically optimal (see [3]), can be approximated by

$$\frac{5}{4}[k_2\alpha(K)^2]^{2/5}\beta(f''_Y)^{1/5} n^{-4/5}, \tag{2.6}$$

where $k_2 = \displaystyle\int t^2K(t)\mathrm{d}t$, $\alpha(K) = \displaystyle\int K(t)^2\mathrm{d}t$ and $\beta(f''_Y) = \displaystyle\int_{-\infty}^{+\infty}[f''_Y(y)]^2\mathrm{d}y$.

Minimizing (2.6) with respect to the transformation parameters is equivalent to minimizing $\beta(f''_Y)$.

In [10] the following estimator for $\beta(f''_Y)$ is suggested as

$$\widehat{\beta}(f''_Y) = n^{-1}(n-1)^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} c^{-5}K * K\{c^{-1}(y_i - y_j)\}, \tag{2.7}$$

where

$$K * K(t) = \int_{-\infty}^{+\infty} K(t-s)K(s)\mathrm{d}s$$

is the kernel convolution and c is the bandwidth used in the estimation of $\beta(f''_Y)$, which can be estimated by minimizing the mean square error (MSE) of $\beta(f''_Y)$. When it is assumed that f_Y is a normal density as in the "rule of thumb" approach, c can be estimated by $\widehat{c} = \widehat{\sigma}_y[21/(40\sqrt{2}n^2)]^{1/13}$, where

$$\widehat{\sigma}_y = \sqrt{n^{-1}\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

Finally, we need to make the selection of the bandwidth that is going to be used for the transformation. Here we simply use the "rule of thumb" for a standard normal density. Since our transformation aims at a transformed density with zero skewness, this approach seems very plausible. Following [11], the estimator of the bandwidth b is $\widehat{b} = 1.059\widehat{\sigma}_x n^{-1/5}$.

2.4 Generalized Pareto Distribution

Here we also suppose we have a sequence of independent and identically distributed observations $x_1, \ldots, x_n$ from an unknown density f. We are interested in excess losses over a high threshold u. Let x_0 be the finite or infinite right endpoint of the distribution f. That is to say, $x_0 = \sup\{x \in R : F(x) < 1\} < \infty$. We define the distribution function of the excesses over the threshold u by

$$F_u(x) = \mathsf{P}\{X - u \le x | X > u\} = \frac{F(x+u) - F(u)}{1 - F(u)}, \tag{2.8}$$

for $0 \le x < x_0 - u$. $F_u(x)$ is thus the probability that a loss exceeds the threshold u by no more than an amount x, given that the threshold is exceeded.

The assuming distribution that modeling the excesses data is the Generalized Pareto distribution (GPD) which is usually expressed as a two parameter distribution with distribution function

$$G_{\xi,\sigma}(x) = \begin{cases} 1 - (1 + \xi x/\sigma)^{-1/\xi}, & \xi \ne 0, \\ 1 - \exp(-x/\sigma), & \xi = 0. \end{cases} \tag{2.9}$$

One tool for choosing suitable thresholds is the sample mean excess plot

$$\{(u, e_n(u)), X_{(1)} < u < X_{(n)}\}, \tag{2.10}$$

where $X_{(1)}$ and $X_{(n)}$ are the first and n-th order statistics of the data sample and $e_n(u)$ is the sample mean excess function defined by

$$e_n(u) = \mathsf{E}[X - u | X > u] = \frac{\sum\limits_{i=1}^{n}(X_i - u)^+}{\sum\limits_{i=1}^{n} 1_{\{X_i > u\}}}, \tag{2.11}$$

i.e. the sum of the excesses over the threshold u divided by the number of data points which exceed the threshold u.

If the empirical plot seems to take shape of a reasonably straight line with positive gradient above a certain value of u, then this is an indication that the excesses over this threshold follow a generalized Pareto distribution with positive shape parameter. This is clear since for the GPD

$$e(u) = (\sigma + \xi u)/(1 - \xi), \tag{2.12}$$

where $\sigma + \xi u > 0$.

For points in the tail of the distribution $(x > u)$ we note that

$$F(x) = \mathsf{P}\{X \le x\} = (1 - \mathsf{P}\{X \le u\})F_u(x - u) + \mathsf{P}\{X \le u\}, \tag{2.13}$$

thus we can estimate $F_u(x-u)$ by $g_{\xi,\sigma}(x)$, which is the probability density function of $G_{\xi,\sigma}(x)$ for u large. We can also estimate $\mathsf{P}\{X \le u\}$ from the data by $F_n(u)$, the empirical distribution function evaluated at u.

Thus for $x > u$ we can use the tail estimate

$$\widehat{F}(x) = (1 - F_n(u))G_{\xi,\sigma}(x) + F_n(u) = 1 - \frac{n_u}{n}\Big(1 + \xi\frac{x-u}{\sigma}\Big)^{-1/\xi}, \tag{2.14}$$

to approximate the distribution function $F(x)$.

Parameters ξ, σ can be estimated by MLE, and the maximum likelihood function is

$$l(\xi, \sigma; x) = -n\ln\sigma - \Big(\frac{1}{\xi} + 1\Big)\sum_{i=1}^{n}\ln\Big(1 + \frac{\xi}{\sigma}x_i\Big). \tag{2.15}$$

§3. Empirical Study

In this section, we present an empirical study. The data set contains earthquake loss in 2009 RMB (ten thousand yuan) each year from 1961 to 2009 in China, except 2008, since the loss in that year is much huger than any other year, and such terrible earthquake is not likely to happen in recent decades, in order to avoid overestimation and reduce error, we regard it as an outlier. Now we use our method to fit this data set.

We first list the descriptive statistics for the annual earthquake loss in Table 1. It shows that both kurtosis and standard deviation are big considering that there are only 48 observations. So we view the data set as highly skewed data with heavy tail.

Table 1 Descriptive statistics for annual earthquake loss in China (in 2009 RMB)

Stat	Mean	Std. Dev.	Skewness	Kurtosis	Minimum	Median	Maximum
Loss	129690.78	319499.07	5.359965	30.93493	4	40661.5	2158500

3.1 Classical Kernel Density Estimation

We apply classical kernel density estimation (KDE) to fit the given data set. Here we use Gaussian kernel and in choosing the bandwidth, the "rule of thumb" is implemented. The KDE graph is shown in Figure 1. We note that the density does not have a smooth shape, as it has some bumps around the sample observations. Besides, the selected bandwidth is as wide as 4.351×10^4, so it is not a quite accurate estimation. Based on this fact, we further explore the shifted power transformation to apply to our data set.

3.2 Transformation on Kernel Density Estimation

Before exploring shifted power transformation, we need to obtain parameters $\lambda = (\lambda_1, \lambda_2)$ first. The method which we propose to obtain the transformation parameters needs to search within a set of transformation parameters that generate transformed variables with zero skewness, to look for the pair of parameters minimizing expression (2.7). We find that the value of parameters $\lambda = (\lambda_1, \lambda_2) = (-3.989921, 0.25)$.

In Figure 2 we show the kernel density estimate of shifted power transformation of the data set. We can see the smoothed shape of the pdf estimated in all of the domain, including the right tail, where data is scarce.

3.3 Combination of Two Methods

Finally the recommend fitting method is given below. We combine Generalized Pareto distribution (GPD) and kernel density estimation (KDE) together to fit the data set. POT method is applied to choose the threshold between big loss and small loss. The sample mean excess plot is shown in Figure 3. Apart from the extreme right hand points which are averages of only a small number of large excesses, the whole plot is approximately linear, suggesting that the GPD might be fitted from a reasonably low threshold. We have chosen a threshold at 40000 (marked by a vertical line), since above this threshold, the data becomes scarce and almost linear. This threshold is exceeded by 24 losses and represents a threshold at the 50th percentile of the data set.

Then we use GPD to fit the excess data over threshold. First by maximizing the log likelihood function, we get the parameter for Pareto distribution. Then the GPD graph can be plotted in Figure 4. So we can use an smooth curve to estimate the tail part of the data set.

As to the part below threshold, we use bounded kernel density estimation (KDE) or transformation kernel estimation (TKE) to estimate the loss function. Here we simply plot bounded KDE with Gaussian kernel and rule-of-thumb method selected bandwidth. The kernel density estimation of loss below threshold is plotted in Figure 5. Other KDE methods, such as TKE with shifted power transformation, are illustrated above and has a similar way to apply. Here we do not list them again.

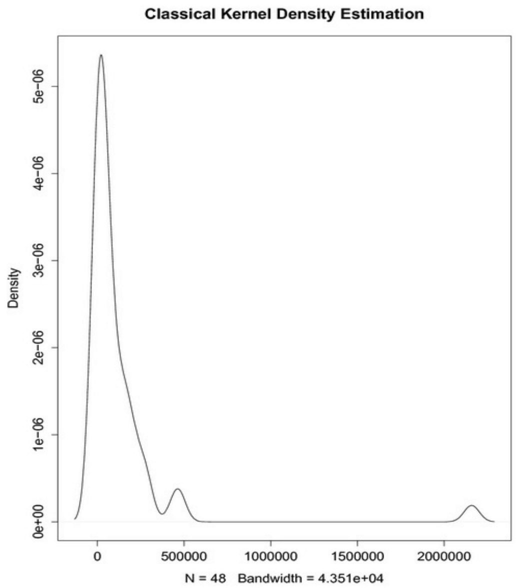

Figure 1　Classical kernel density to fit loss data

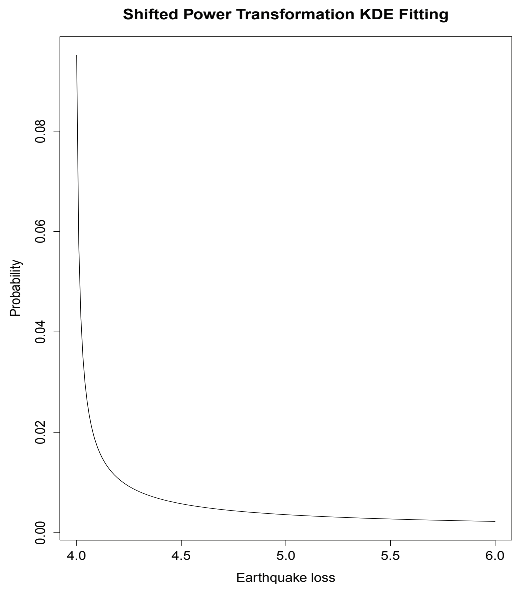

Figure 2　Transformed kernel density to fit loss data

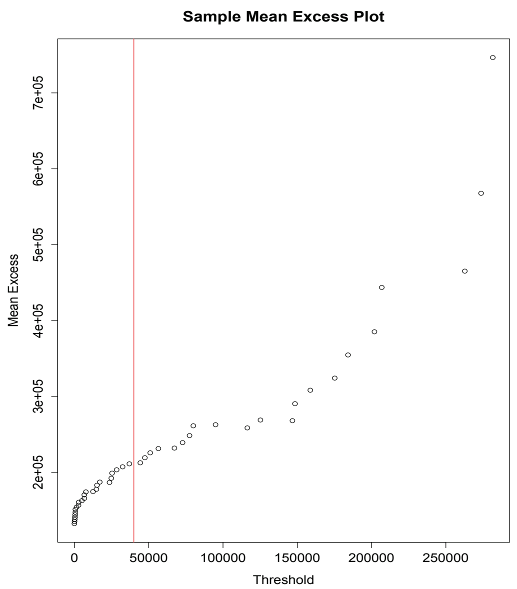

Figure 3　Sample mean excess plot of annual earthquake loss data

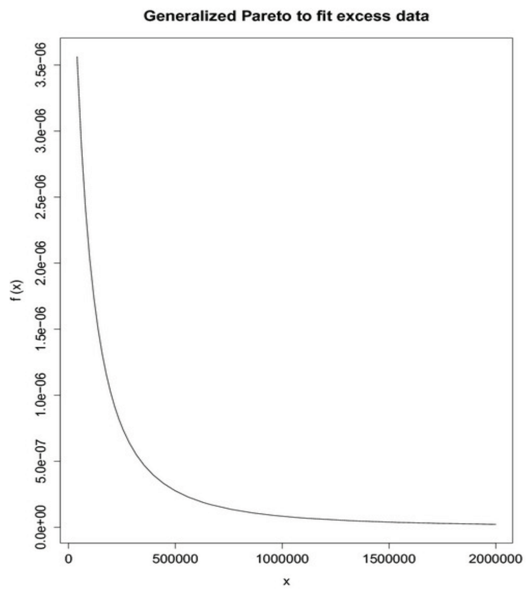

Figure 4　Generalized Pareto to fit excess data

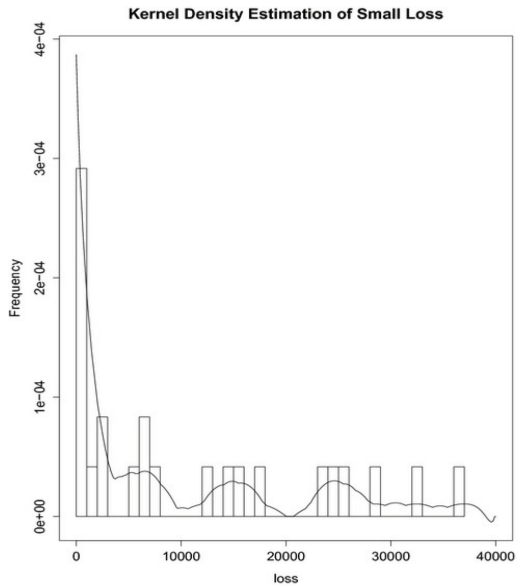

Figure 5 Kernel density estimation of loss below threshold

§4. Measuring the Error

In order to show the advantage of our method over other methods, we calculate the fitting errors of different methods. The first error measurement method are log-likelihood function and weighted log-likelihood function, which are easy to apply when estimating error of parametric model. The second error measurement method are integrated square error and weighted integrated square error, which are easy to apply when estimating error of nonparametric model.

4.1 Log-likelihood Function and Weighted Log-likelihood Function

Let us assume that we have a sample of n independent and identically distributed observations $x_1, \ldots, x_n$. $\widehat{f}_X$ is an estimate of the density for every point in the domain. Then, we can estimate the log-likelihood function as

$$\ln \widehat{L}(\widehat{f}_X; x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln \widehat{f}_X(x_i). \tag{4.1}$$

Similarly，if a transformation method was used as $\widehat{f}_X(x, \widehat{\lambda}; \widehat{h})$, the estimated log-likelihood function is

$$\ln \widehat{L}(\widehat{f}_X(\cdot); T_{\widehat{\lambda}}(\cdot); x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln \widehat{f}_X(x_i, \widehat{\lambda}; \widehat{h}). \tag{4.2}$$

4.2 Integrated Square Error and Weighted Integrated Square Error

However, log-likelihood function and weighted log-likelihood function are not good criteria to evaluate the performance of non-parametric estimation. So we measure the performance of the kernel density estimators by the error measures ISE, WISE_1 and WISE_2. $\widehat{f}_X(x)$ is defined as the estimate of the density for every point x in the domain and $f_X(x)$ is the true density. Then ISE, WISE_1 and WISE_2 are defined as follows

$$\mathrm{ISE} = \int_0^{\infty} (\widehat{f}_X(x) - f_X(x))^2 \mathrm{d}x, \tag{4.3}$$

$$\mathrm{WISE}_1 = \int_0^{\infty} (\widehat{f}_X(x) - f_X(x))^2 x \mathrm{d}x, \tag{4.4}$$

$$\mathrm{WISE}_2 = \int_0^{\infty} (\widehat{f}_X(x) - f_X(x))^2 x^2 \mathrm{d}x. \tag{4.5}$$

ISE weighs errors of the estimator in the whole domain equally, while WISE_1 and WISE_2 emphasizes the tail of the distribution, which is very relevant in practice when dealing with heavy-tailed data set, just like the data set we used in this paper.

In [10] it is proved that minimizing ISE is equivalent to minimizing the cross-validation (CV) function

$$\mathrm{CV} = \int_{-\infty}^{+\infty} [\widehat{f}_X(x)]^2 \mathrm{d}x - \frac{2}{n} \sum_{i=1}^{n} \widehat{f}_i(x_i), \tag{4.6}$$

where $\widehat{f}_i$ is the "leave-one-out" estimation, the kernel estimation of f_X without observation x_i.

Similarly, we can estimate WISE_1 and WISE_2 approximately with

$$\mathrm{WCV}_1 = \int_{-\infty}^{+\infty} [\widehat{f}_X(x)]^2 \mathrm{d}x - \frac{2}{n} \sum_{i=1}^{n} \widehat{f}_i(x_i) x_i, \tag{4.7}$$

$$\mathrm{WCV}_2 = \int_{-\infty}^{+\infty} [\widehat{f}_X(x)]^2 \mathrm{d}x - \frac{2}{n} \sum_{i=1}^{n} \widehat{f}_i(x_i) x_i^2. \tag{4.8}$$

4.3 Comparison of Methods

Here we compare two parametric methods by fitting with normal distribution and lognormal distribution; three nonparametric methods, which are classical KDE, TKE with log transformation and TKE with shifted power transformation; and the method suggested in this paper: choosing threshold first, use GPD to fit excess data afterwards and KDE or TKE to fit data below threshold.

We show the results of log-likelihood and weighted log-likelihood for all these densities for the earthquake loss in Table 2. A higher value indicates a better fit. We can easily observe from the first two rows that normal distribution and lognormal distribution does not fit this data set well, which means the parametric method is not applicable in this heavy-tailed fitting problem. From the third row to the fifth row, we can see an obvious improvement in error measurement. This is an evidence that the nonparametric method, KDE and TKE perform better than the parametric method in row one and row two. Row six to row eight show our method which combines both GPD to fit excess data and KDE or TKE to fit data below the threshold of 40000. Method I, II and III indicate classical KDE, TKE with log transformation and TKE with shifted power transformation. This method shows advantage over other methods, since the largest number of $\ln \widehat{L}$ is reached in Combination method III, $\ln_{w(1)} \widehat{L}$ and $\ln_{w(2)} \widehat{L}$ is reached in Combination method I.

Table 2 Comparison of both parametric and nonparametric methods

Method	$\ln \widehat{L}$	$\ln_{w(1)} \widehat{L}$	$\ln_{w(2)} \widehat{L}$
Normal	-675.9802	-999.6313	-1475.7821
Lognormal	-592.7322	-749.8355	-840.9582
Classical KDE	-633.8634	-758.7188	-782.9776
TKE with log transformation	-585.7816	-750.5531	-863.9291
TKE with shifted power transformation	-582.8955	-742.8866	-851.5888
Combination method I	-580.7832	-751.6342	-721.4960
Combination method II	-572.5750	-861.1948	-733.1351
Combination method III	-570.9885	-756.6023	-727.4210

From above analysis, we can see our method has advantage towards other methods, which is a more accurate method with less error than others. As to whether apply KDE or TKE, we propose the use of CV, WCV_1 and WCV_2 to compare the fit of classical KDE and TKE, which is a better method to measure nonparametric fits. The results for earthquake loss and the earthquake loss below threshold are found in Tables 3 and Table 4 respectively. The lower value indicates the better fit. In Table 3, the minimum value of CV for earthquake loss is found for TKE with shifted power transformation. While the minimum values of WCV_1 and WCV_2 for earthquake loss X are found in TKE with log transformation and classical KDE. In Table 4, the minimum values of CV and WCV_1 for earthquake loss below threshold are found in Combination method III. While the minimum values of WCV_2 are found in Combination method I. These two results indicate that there is no method in KDE or TKE outweighs others in all three error measurements. We may choose TKE with shifted power transformation if considering error in whole domain equally, and may choose Classical KDE if putting more weights on tail. Since we have already spread the large loss to reinsurance companies and only consider loss below threshold, we suggest to weigh all error in whole domain equally. In this case, TKE with shifted power transformation has the best fit.

Table 3 Comparison of different KDE and TKE

Method	CV	WCV_1	WCV_2
Classical KDE	-3.56×10^{-6}	-0.2227	-35359.53
TKE with log transformation	-3.43×10^{-6}	-0.2282	-24165.06
TKE with shifted power transformation	-3.24×10^{-3}	-0.0032	-25627.42

Table 4 Comparison of combination method from earthquake loss below 40000

Method	$\ln\widehat{L}$	$\ln_{w(1)}\widehat{L}$	$\ln_{w(2)}\widehat{L}$
Combination method I	-4.045547×10^{-5}	-0.3005022	-5602.093
Combination method II	-0.0001100273	-0.2251045	-3912.225
Combination method III	-0.00400114	-0.30979	-4799.462

§5. Conclusion

In this paper, we have introduced an alternative method for estimating loss distributions. The method, which first use extreme value theory to choose the threshold, then fit excess data with generalized Pareto distribution, and applies kernel density estimation with transformation to fit loss below threshold, perform well in earthquake loss fitting problem above. Since insurance companies reinsure large loss to reinsurance companies, it is a practical choice to select threshold and determine the scale over and below threshold.

References

[1] Wand, M.P. and Jones, M.C., *Kernel Smoothing*, Chapman & Hall, 1995.

[2] Wand, M.P., Marron, J.S. and Ruppert, D., Transformations in density estimation, *Journal of the American Statistical Association*, **86**(1991), 343–353.

[3] Bolancé, C., Guillén, M. and Nielsen, J.P., Kernel density estimation of actuarial loss functions, *Insurance: Mathematics and Economics*, **32**(2003), 19–36.

[4] Clements, A., Hurn, S. and Lindsay, K., Mobius-like mappings and their use in kernel density estimation, *Journal of the American Statistical Association*, **98**(2003), 993–1000.

[5] Buch-Larsen, T., Nielsen, J.P., Guillén, M. and Bolancé, C., Kernel density estimation for heavy-tailed distributions using the Champernowne transformation, *Statistics*, **39(6)**(2005), 503–518.

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

[7] Embrechts, P. and Kluppelberg, C., Some aspects of insurance mathematics, *Theory of Probability and Its Applications*, **38**(1993), 262–295.

[8] Rootzén, H. and Tajvidi, N., Extreme value statistics and wind storm losses: a case study, *Scandinavian Actuarial Journal*, **1**(1997), 70–94.

[9] McNeil, A.J. and Saladin, T., The peaks over thresholds method for estimating high quantiles of loss distributions, *Proceedings of XXVIIth International Astin Colloquium*, 1997, 23–43.

[10] Hall, P. and Marron, J.S., Estimation of integrated squared density derivatives, *Statistics & Probability Letters*, **6**(1987), 100–115.

[11] Silverman, B.W., *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, 1986.

基于中国地震的损失模型

蔡 铨 林正炎

(浙江大学数学系, 杭州, 310027)

对损失分布的估计一直是保险公司的重要问题. 有多种参数方法以及非参数方法拟合损失分布. 本文作者提出了结合参数和非参数的方法来解决损失分布拟合问题. 首先通过超额均值图确定大小损失之间的阈限, 再利用广义Pareto分布拟合阈值以上损失, 转换后的核密度估计拟合阈值以下损失. 最后, 通过实证分析将该方法和其他方法进行了误差分析比较, 取得了理想的结果.

关键词: 损失分布, 重尾, 广义Pareto, 超额均值图, 核密度估计.

学科分类号: O212.7.