Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 06, pp. 17342-17346, June 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

ISSN: 0975-833X

RESEARCH ARTICLE

DETECTION OF A SERIOUS SINISTER IN A RATE BOX: AN APPLICATION OF THE EXTREME THEORY VALUES IN CAR INSURANCE

*Diawara Daouda

Zhongnan University of Economics and Law Wuhan 182# Nanhu Avenue, East Lake High-tech Development Zone, Wuhan 430073, China

ARTICLE INFO

Article History:
Received 26th March, 2015
Received in revised form
12th April, 2015
Accepted 29th May, 2015
Published online 30th June, 2015

Key words:
Serious sinister, Pareto Distribution, Auto Insurance, POT method,
Convex combination of thresholds.

ABSTRACT

Generally boxes made from characteristics of the insured and the vehicle, are assumed to be homogeneous in terms of claims. The presence of large claims in a class is not only disrupting this homogeneity assumption of the boxes, but also the stability of risk indicators such as the pure premium. The extreme value theory provides a rigorous probabilistic mathematical basis on which we can build statistical models to predict the intensity and frequency of extreme events. Our method provides a comparison between the variance of a single threshold and that of a convex combination. Acceptable threshold is the threshold which will have a smaller variance.

Citation: Diawara Daouda, 2015. "Detection of a serious sinister in a rate box: An application of the extreme theory values in car insurance", *International Journal of Current Research,* 7, (6), 17342-17346.

INTRODUCTION

In motor insurance, classes formed from characteristics of the insured and the vehicle are assumed to be homogeneous in terms of claims. The presence of extreme events disturb the homogeneity of the portfolio and their detection is essential to avoid errors in pricing and interpretation that may have the impact of tariff changes useful or useless opposite.

Permanent questions arise are:

- What art is a serious disaster in a given class?
- What is the threshold at which a claim will be considered serious in a rate case?

It is very important to know the severity of a disaster in a rate case to ensure stability of damage indicators and thus a match between the reference premium and loss experience. In our article we will firstly make an overview of the extreme value theory. Using this theory to draw the line between ordinary claims and serious claims. The approach of exceeding the threshold noted POT (Peaks Over Threshold) of the extreme value theory allows to find a reasonable threshold. Once the threshold is chosen, the choice is often a compromise, we will make an estimate of the tail function. One note that this estimate is accurate the threshold should be well chosen. Finally, we propose a contribution from the convex combination of two thresholds method that minimizes the variance, two thresholds obtained from the extreme value theory. Our method is a comparison between the variance of a convex combination of two thresholds and the variance of a threshold. So the reasonable threshold will be one that has the smallest variance.

***Corresponding author: Diawara Daouda**
Zhongnan University of Economics and Law Wuhan 182# Nanhu Avenue, East Lake High-tech Development Zone,Wuhan 430073, China

Law of extreme statistics

In this part we are interested in the limit distributions of order statistics when $n \to +\infty$.

Therefore, let us consider random variables $X = (X_1, X_2, ..., X_n)$ with distribution function $F(x) = P(X \le x)$ and the density function f.

Let $M_n = \max(X_1, X_2, ..., X_n)$ and $m_n = \min(X_1, X_2, ..., X_n)$. The value $d_n = M_n - m_n$ is called extreme deviation. The law that follows M_n and m_n are:

$$F_{M_n}(x)=P(M_n\le x)=\left[F_X(x)\right]^n$$

$$F_{m_n}(x)=1-\left(1-F_X(x)\right)^n$$

We can therefore conclude that M_n is a random variable distribution function F^n.

$$\begin{cases} F_{M_n}(x)=F^n(x) \\ f_{M_n}(x)=n\,F^{n-1}(x)\,f(x) \end{cases}$$

So if the distribution function is known then we can find a simple way of maximum (See [3],[6],[3]).

Problem: The distribution function F of X is generally unknown therefore impossible to determine the distribution of the maximum M_n from this result.

So you have to be interested in the asymptotic distribution of the maximum.

So you have to be interested in the asymptotic distribution of the maximum by making n tend to infinity. We have:

$$\lim_{x\to\infty}F_{M_n}(x)=\lim_{x\to\infty}\left(F_X(x)\right)^n=\begin{cases}0 & if\quad F(x)<1\\ 1 & if\quad F(x)=1\end{cases}$$

At infinity extreme law $F_{M_n}\in\{0,1\}$, therefore it is called degenerate. This causes a lack of information on the extreme (See [3],[3]).

Asymptotic distribution of the maximum

The distribution function of maximum obtained at infinity leads to degenerate law, we look for a non-degenerate law for maximum .This non-degenerate limit law is provided by the following theorem (Theorem of Gnedenko, 1943).

Theorem 1

Whether $X_1,X_2,...,X_n$ is a sequence of random variables i.i.d with distribution function F and $M_n=\max(X_1,X_2,...,X_n)$. If a sequence $(a_n)_{n\ge1}>0$ of positive terms exists, a real sequence $(b_n)_{n\ge1}$ and a non-degenerate distribution function G, so that:

$$\lim_{x\to\infty}P\left(\frac{M_n-b_n}{a_n}\right)=\lim_{x\to\infty}F^n(a_nx+b_n)=G(x)$$

$$P\left(\frac{M_n-b_n}{a_n}\le x\right)=P(M_n\le a_nx+b_n)$$
$$=F^n(a_nx+b_n)\xrightarrow{d}G(x)$$

Then the only possible forms of G are the Gumbel, Fréchet or Weibull distributions, also called type I, II and III distributions respectively (See [7]).

The variable $\alpha_n=\dfrac{M_n-b_n}{a_n}$ is called normalized maximum.

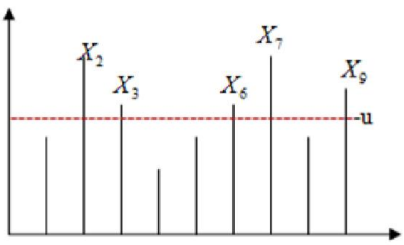
Jenkinson (1955) which gives a unique general form of limit laws: This family can be described by a single term to three parameters:

$$G_{\mu,\sigma,\xi}(x)=\exp\left(-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right),x\in\mathbb{R}$$

Here $x\in\mathbb{R}$, $\mu\in\mathbb{R}$ and $\sigma>0$. The case $\xi>0,\xi<0$ ($\xi<0$) corresponds to the Fréchet (Weibull)-type distribution function. $\xi=0$ Corresponds to the Gumbel distribution(see [6] and [7]).

Conditional excess distribution

The second part of the extreme value theory called POT (Peaks Over Threshold) is to choose an appropriate threshold and use observations that exceed this threshold, called excess.



POT Method

Fig. 1. Threshold

From figure 1 we see that u is the threshold and the numbers X_2,X_3,X_6,X_7 and X_9 are in excess of the threshold. The random variables X_2,X_3,X_6,X_7 and X_9 are extreme values. The Insurance companies have a tradition of using a threshold to differentiate the two types of claims (the very frequent regular claims no high costs and serious claims whose probability of occurrence is very low but resulting in very high costs for mutual insurance).

Suppose that $X_1,X_2,...,X_n$ are n independent realizations of a random variable X with a distribution function $F(x)$. Let $x_F=\sup\{x:F(x)<1\}$ be the finite or infinite right endpoint of the distribution F.

The distribution function of the excesses over certain (high) threshold u (F_u) is given by:

$$F_u(y)=P(X-u\le y/X>u)=\begin{cases}\dfrac{F(y+u)-F(u)}{1-F(u)} & \text{if } y\ge0\\ 0 & \text{if } \ y<0\end{cases}$$

Figure 2 shows the conditional function of the excess F_u which depends of the threshold and the distribution function F .

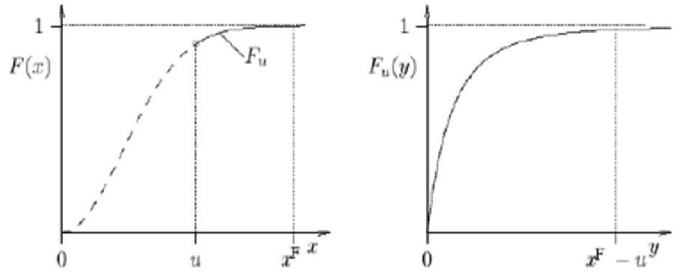


Fig. 2. Distribution function and function Conditional distribution F_u

The purpose of the POT method is defined by what probability distribution can be approached .Balkema and de Haan (1974), Pickands (1975) proposed a theorem states that the conditional distribution of excess (See [4],[6],[7]).

Theorem 2

The Pickands-Balkema-de Haan theorem (Balkema & de Haan 1974; Pickands 1975) states that if the distribution function $F \in DA\left(G_{\xi}\right)$ then $\exists$ a positive measurable function $\sigma\left(u\right)$ such that:

$$\lim_{u \to x_F} \sup_{y \in [0, x_F - u]} \left| F_u(y) - G_{\xi,\sigma(u)}(y) \right| = 0$$

and vice versa, where $G_{\xi,\sigma(u)}\left(y\right)$ denote the Generalized Pareto distribution. The Generalized Pareto Distribution (GPD) is given:

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

With $y \geq 0$ for $\xi > 0$ *and* $0 \leq y \leq -\dfrac{\sigma(u)}{\xi}$ for $\xi < 0$

In other words well chosen for a threshold, the excesses of law can be approximated by a GPD whose extreme index is the same as that of the GEV law. The main difficulties of this model are the choice of the threshold and the estimation method parameters (see [6] and [7]).

Threshold Selection

The extreme value theory provides different methods for estimating a threshold above observation will be considered extreme. If one chooses a low threshold, certain non-outliers will be declared as extreme and involve an under the pure premium. If the threshold should be big but not too high to have enough data beyond this threshold (enough data to a good estimate of the model). A threshold selecting tools is the graph of the sample mean excess function $e_n\left(u\right)$ (ME-plot).

Definition: The ME-plot is defined as follows:

$$\left\{\left(u; e_n\left(u\right)\right); X_1 < u < X_n\right\}$$

X_1 and X_n are respectively the minimum and the maximum of the sample is given by the formula:

$$e_n(u) = \frac{\sum_{i=1}^{n} (X_i - u)^+}{\sum_{i=1}^{n} \left(I_{(X_i > u)}\right)} = \frac{1}{N_u} \sum_{i=1}^{n} (X_i - u)^+$$

In other words the sum of the excess over the threshold u divided by the number N_u of data that exceeds .The sample mean excess function $e_n\left(u\right)$ is an empirical estimate of the mean excess function: $e\left(u\right) = E\left(X - u \mid X > u\right)$. The mean excess function of GPD is:

$$e(u) = \frac{\sigma_{(u)} + \xi u}{1 - \xi}$$

This is a linear function at u .Then the choice of threshold corresponds to the beginning of the linearity observed on the graph defined by $\left(u, e\left(u\right)\right)$.The threshold u is determined from the time when the graph of the function has an affine part unchanged (see [1], [2] and [7]).

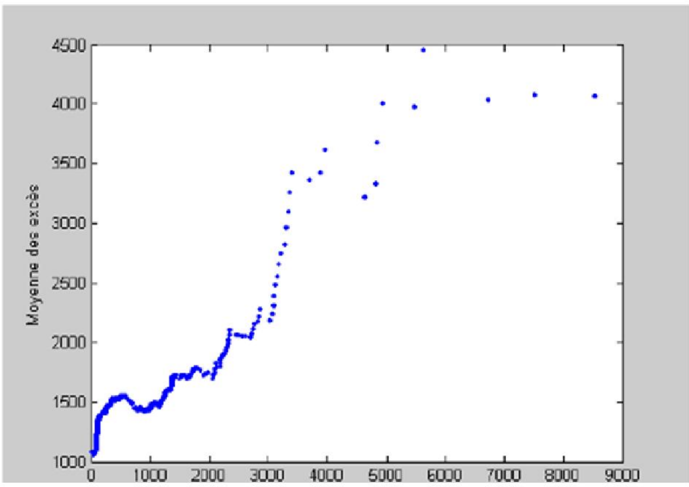


Fig. 3. Mean Excess Function of GPD

From fig.3 we have an representation of the empirical mean excess function of a GPD distribution with a positive parameter ξ. It is found that this feature is stable for a threshold of the order of 5500. So we can choose as a threshold beyond which any observation (Loss) is considered extreme. Here the mean excess function is stable of a threshold of the order of 5 500. So we can choose as a threshold $u = 5500$ beyond which any observation (Loss) is considered extreme ([2]).

Numerical applications

The database provides a sample of 2020 observations for 4 wheel vehicle for personal use during the year 2013.Les data come from a Malian insurance company and concern the amounts of claims caused by the insured of a risk class. This file contains only the amounts of claims during the insurance year. Risk boxes are constructed from the vehicle characteristics and other variables. For confidentiality reasons, the company would not give us the other variables.

Statistical analysis of the data

Table 1. Statistics of data

N	Valid	2020
	Missing	2
Mean		10.076699471
Median		10.074040950
Minimum		3.5112916
Maximum		16.3227467
Percentiles	25	8.722730170
	50	10.074040950
	75	11.438886175

The table 1 shows that the amounts of claims are increasing on average with the number of disaster.

Detection of extreme value

We will initially use the boxplot which allows a simple reading to detect the presence of extreme values.

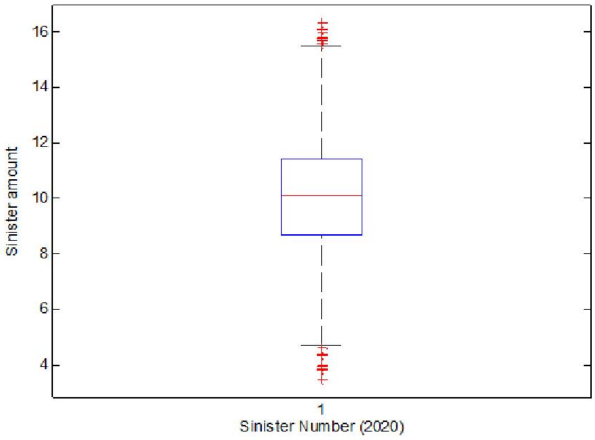


Fig. 4. Boxplot of data

This figure shows the upper limit and lower limit of the simple boxplot. If our data do not contain extreme values then all data will be between the upper limit and lower limit, this is not the case of fig.4, therefore we can clearly see the presence of extreme values (see [1]). After having detected the presence of extreme values, the average excess function allows to predict the approximate threshold (beginning of linearity).

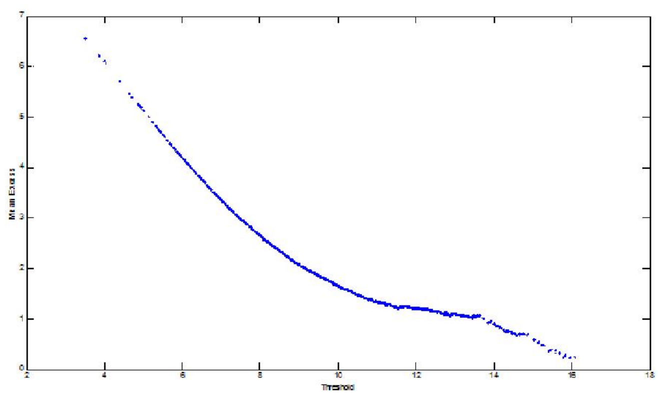


Fig. 5. Mean Excess Function of Data

The mean excess function of all data plotted below (Fig. 5) gives a threshold beyond which a claim can be considered serious. This choice must satisfy criteria: keeping a large enough data up for a good estimate of the model. From this figure we can retain as a threshold value 12. There is a stability curve from this value (See [2] and [3]). Plotting the Mean Residual Life helps support the choice of threshold (fig.6)

Fig. 6. Mean Residual Life Plot

Looking at the plot (fig.6), 12 is a reasonable selection for the threshold The choice of this threshold gives the following statistical table:

Table 2. Statistic of data after the choice of threshold

```
Threshold Call: 12
  Number Above: 323
Proportion Above: 0.1599
```

The statistical data table after choosing a threshold gives 323 observations above (approximately 16% the total of all observations) the threshold. Given the total number of sample observations we can say that the observations beyond the threshold are important to a good approximation of the model. So GPD can be adjusted by the maximum likelihood method from 323 values exceeding the threshold (See [5]).

Adjusting the GPD

GPD is adjusted by the maximum likelihood method from the 323 values exceeding the threshold $u = 12$, among the *2020* sample values (Using R).

Table 3. Results of the estimate by the likelihood method of arameters

```
Varying Threshold: FALSE

  Threshold Call: 12
    Number Above: 323
Proportion Above: 0.1599

Estimates
  scale     shape
 1.5805   -0.2762

Standard Error Type: observed

Standard Errors
  scale      shape
0.12529    0.05852
```

This table shows that-linearity is characterized by a negative slope so the data belong to MDA (Weibull) (see [3]).

New threshold selection approach

As reported since the beginning of this article the main difficulty of the problem lies in the choice of threshold and parameter estimation of the Pareto distribution. For our part we assume that the extreme value theory is known and we use mathematical reasoning that says if: $a \in [\mu, \eta]$ then μ and η are approximate values a and $r = \frac{\mu + \eta}{2}$ is a more reasonable approximation of a. The real r is simply a convex combination of μ and η. Let u_1 and u_2 two random thresholds obtained using the extreme theory values, α a positive real ($0 < \alpha < 1$) and $u = \alpha u_1 + (1-\alpha)u_2$ a convex combination and .So u is a reasonable random threshold than u_1 and u_2. Our method provides a comparison between the variance of a single threshold and that of a convex combination. Acceptable threshold is the threshold which will have a smaller variance.

You can have two cases:

$$Var(u_1) = Var(u_2) = \sigma^2 \text{ and } Var(u_1) \neq Var(u_2)$$

First case: $Var(u_1) = Var(u_2) = \sigma^2$.Let

$$f(\alpha) = \frac{Var(\alpha u_1 + (1-\alpha)u_2)}{\sigma^2} = \frac{\alpha^2 Var(u_1) + (1-\alpha)^2 Var(u_2) + 2\alpha(1-\alpha)Cov(u_1, u_2)}{\sigma^2}$$

$$= \alpha^2 + (1-\alpha)^2 + 2\alpha(1-\alpha)\frac{Cov(u_1, u_2)}{\sigma^2}$$

We know that: $Cov(u_1, u_2) < \sigma^2 \Rightarrow f(\alpha) < 1$

And $P(\alpha) = \alpha^2 + (1-\alpha)^2$, $P'(\alpha) = 2\alpha - 2(1-\alpha) = 0 \Rightarrow \alpha = \frac{1}{2}$

In conclusion we can say that the function $f(\alpha) < 1$ and is minimal for $\alpha = \frac{1}{2}$. So the convex combination minimizes the variance.

Second case: $Var(u_1) \neq Var(u_2)$.We has two possibilities:

$Var(u_1) < Var(u_2)$ or $Var(u_1) > Var(u_2)$

Let: $g(\alpha) = Var(u) = \alpha^2 Var(u_1) + (1-\alpha)^2 Var(u_2) + 2\alpha(1-\alpha)Cov(u_1, u_2)$

This case has been much study in the literature (see [10],[8])

Conclusions and Discussions

The extreme value theory based on the generalized Pareto distribution is a very valuable tool for modeling rare events but does not totally solve the problems in one go. It is interesting from the perspective of prediction. But we must recognize that its use in practice is not an easy thing, especially in an unstable economy in transition. It is the same for our technique, based on a reduction of the variance of the convex combination of two random thresholds, the comparison of the variance of a single threshold and that of a convex combination, even if it seems to be a good empirical compromise in quality between the average method based on excess and that using the generalized Pareto distribution.

REFERENCES

Alexander J., McNeil, Saladin T. 1997. The peaks over thersholds for estimating high quantiles of loss distribution *International ASTIN Colloquium*, pp 70-94.

Aurelie Muller Comportement asymptotique de la distribution des pluies extrêmes en France

Christophe Pallier Exemples d'Analyses de Variance avec R 25 août 2002

De Haan, L. and Ferreira, A. Extreme value theory: An introduction

Efron, B. 1979. Bootstrap methods: another look at the Jackknife The Annals of Statistics, 7, 1-26

Embrechts, P., Kluppelberg, C. and Mikosch, T. Modeling extremal events: for insurance and finance

Noureddine Benlagha Michel Grun-Réhomme et Olga Vasechko Université Paris 2, ERMES-UMR7181-CNRS, 12 place du Panthéon, 75005 Paris, France

Noureddine Benlagha and Michel Grun-Réhomme Université Paris 2, ERMES-UMR7181-CNRS, 92 rue d'Assas 75006 Paris, France

The POT Package Octobre 30, 2007

Tukey J.W. 1977, Exploratory Data Analysis, Ed. Addison-Wesley

****