International Journal of Statistics and Probability; Vol. 5, No. 2; 2016
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Mixed Method of Extreme Value Theory, with Application to the Calculation of the Portion of Each Claim Payable by the Reinsurer of Excess of Sinister

Yingmei Xu[1], Kane Ladji[2] & Diawara Daouda[3]

[1,3]Zhongnan University of Economics and Law, Wuhan China

[2]Faculty of Economics and Management FSEG Bamako, Mali

Correspondence: Diawara Daouda, Ph.D Candidate Zhongnan University of Economics and Law, Wuhan China.

Tel: 86-132-6472-1138. E-mail : btddiawara@yahoo.fr

Received: January 26, 2016 Accepted: February 10, 2016 Online Published: February 25, 2016

doi:10.5539/ijsp.v5n2p100 URL: http://dx.doi.org/10.5539/ijsp.v5n2p100

Abstract

In the literature many determinists approaches (numerical and graphical methods), probability (the probability law, extreme value theory, Bayesian methods) exist for the detection of grave sinister. In this paper, we will give a new characterization of the mixed method of extreme value theory. These results are applied to the simulated data of a Malian insurance company.

Keywords :insurance, claims management, minimizing the variance, convex combination of two and three thresholds

2000 Mathematics Subject Classification: 60G52, 60G70, 62G20, 62G32

1. Introduction

Insurance is a transaction whereby a person undertakes to indemnify another person, the insured in case of occurrence of a specified risk, against the prior payment of a premium or a contribution. Insurance occupies a very important place in the modern economy: its mechanism helps increase the level of protection of all individuals, and practice has been made mandatory in many areas. The insurance mechanism is based on risk compensation if all policyholders are subject to risk, the likelihood that it will achieve for all insured is low. The victims are compensated thanks to contributions from the community of contributors. The insurer must be able to predict the loads of sinister it will have to bear because of the risks it covers when establishing its insurance policies. These evaluations are of great importance for the insurance company to avoid ruin and insurance solvency of its portfolio. So predict the occurrence of such expenses sinister is very important to take precautions. Usually these assessments are conducted by the insurer and the reinsurer. Reinsurance, Insurance whereby an insurer is guaranteed by another company own risk. The extreme value theory allows, in fact, establish the scenarios of the calculation of the portion of each sinister to load of the excess of loss reinsurer, allowing the insurance company to consider these sinister surpluses and keep its solvency.

Reinsurance excess of loss covers the portion of each individual claim excess a given priority, limited to a capacity granted by the reinsurer. We place ourselves in the collective risk model. Let N be the number of sinister and $X_1, X_2, ..., X_N$ the realizations of X, which is the random variable representing the amounts of loss. As usual we assume mutual independence of random variables. Let then D priority referred to above. Let then D priority referred to above, or deductible, cover and let C (capacity) offered by the reinsurer. Then, the portion of each sinister X_i $i = 1, ..., N$ is dependent reinsurer

$$R_i = \min(C, \max(0, X_i - D)) \qquad (1)$$

2. Mixed Method of Extreme Value Theory

We assume that the extreme value theory is known. This new method was proposed in (Noureddine and al., 2009) to determine a threshold, at which a unit is declared atypical minimizing the variance of a convex combination of thresholds obtained by the mean excess function and generalized Pareto distribution (extreme quantile were estimated with a probability of 99, 9% being an extreme value for the distribution of amounts of sinister with a confidence level of 95%). U_1 be the threshold beyond which a unit is declared as extreme, obtained by the mean excess function and U_2 the threshold beyond which a unit is declared as extreme, obtained by the GPD function.

Let:

$$U = \alpha U_1 + (1-\alpha) U_2 \,,\, 0 < \alpha < 1 \qquad (2)$$

α minimizes the variance of U (see Noureddine, and al., 2009).

We get $\alpha = \frac{V_2 - \mathrm{cov}(V_1, V_2)}{V_1 + V_2 - 2\,\mathrm{cov}(V_1, V_2)}$, where V_i is the variance of U_i. The variances and covariance are estimated using a bootstrap technique. This result generalizes to case of p random variables U_i.

2.1 New Method of Selection of a Threshold

In this section, we propose a new method to determine a threshold, at which a unit is declared atypical minimizing the variance of a convex combination of two thresholds and three thresholds. U_1 be the threshold beyond which a unit is declared as extreme, obtained by the record values, U_2 be the threshold beyond which a unit is declared as extreme, obtained by the mean excess function and U_3 the threshold beyond which a unit is declared as extreme, obtained by the GPD function with $U_1 < U_2 < U_3$. Let $U = \alpha U_p + (1-\alpha) U_q$ with $0 < \alpha < 1$, minimizes the variance U ,p, q=1,2,3 and p<q.

We get:

$$\alpha = \frac{V(X_{U_q}) - Cov(X_{U_p}, X_{U_q})}{V(X_{U_p}) + V(X_{U_q}) - 2Cov(X_{U_p}, X_{U_q})} \qquad (3)$$

For $U = \alpha U_1 + (1+\alpha) U_2 - 2\alpha U_3$ with $\alpha \in \mathbb{R}$ (α a real). We get:

$$\alpha = \frac{-V(X_{U_2}) - Cov(X_{U_1}, X_{U_2}) + 2Cov(X_{U_2}, X_{U_3})}{V(X_{U_1}) + V(X_{U_2}) + 4V(X_{U_3}) + 2Cov(X_{U_1}, X_{U_2}) - 4Cov(X_{U_1}, X_{U_3}) - 4Cov(X_{U_2}, X_{U_3})} \qquad (4)$$

Let N be the number of claims and $X_1, X_2, ..., X_N$ the realizations of X, which is the random variable representing the amounts of loss. As usual we assume mutual independence of random variables. Let X_{U_j} ,j=1,2,3 thresholds obtained by different methods. We consider a statistical series to a variable X_{U_j}, taking the amount $X_1, X_2, ..., X_N$ and X_{U_j} which have been sorted in ascending order: $X_1 \le X_2 \le ... \le X_k \le X_{U_j} \le ... \le X_N$. We define a statistical series 2 variables.

We consider a statistical series 2 variables X and Y, taking the amount $X_1, X_2, ..., X_N$ and $Y_1, Y_2, ..., Y_N$. which have been sorted in ascending order: $X_1 \le X_2 \le ... \le X_N$ and $Y_1 \le Y_2 \le ... \le Y_N$. We write:

- The means of X and Y are : $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ et $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$.
- The variances of X and Y are: $V(X) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2$ et $V(Y) = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2$.
- The covariance of X and Y is: $Cov(X,Y) = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$.

Let C the coverage (capacity) offered by the reinsurer. Then, the portion of each $X_j > U$ dependent the reinsurer is: $R_j = \min(C, X_j - U)$.

3. Numerical Application

The data base 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.

3.1 Exploratory Data Analysis

Table 1. Summary of the position of the initial sample.

N	Valid	2020
	Missing	2
Mean		10.077
Median		10.074
Minimum		3.511
Maximum		16.322
Percentiles	25	8.722
	50	10.074
	75	11.439

The scatter plot, the boxplot and histogram used to learn more.

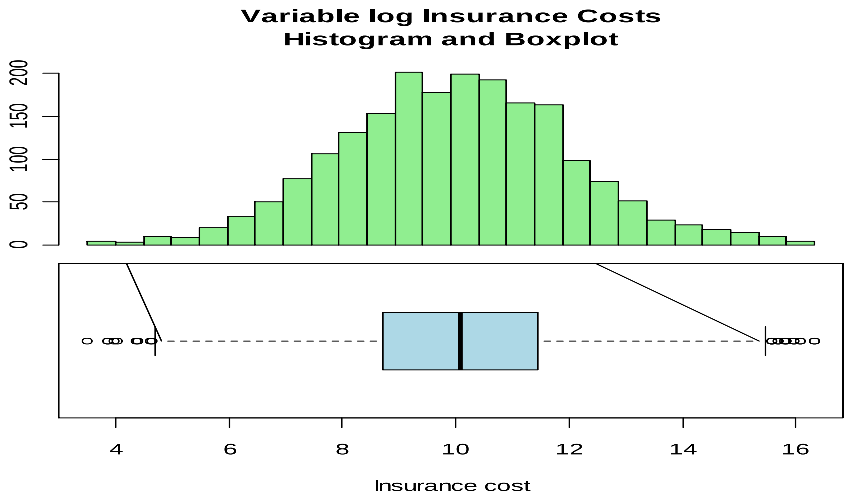

Figure 1. The scatter plot and Boxplot of Data

Each point is the graphical representation of a pair $(x, y), x \in]0, 2020[$ and $y \in]3,17[$, y means the amount. 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 figure 1, therefore we can clearly see the presence of extreme values**.**

The histogram represents the distribution of data in the intervals of length 0.5.We can see that the class which registers more data is the interval [9, 9.5 [.We also observe that from 12 from one class to the next (in ascending amount) the corresponding numbers are decrease considerably (about half of the workforce above).For more information about relatively extreme amounts, we construct the density and distribution curves of the sample.

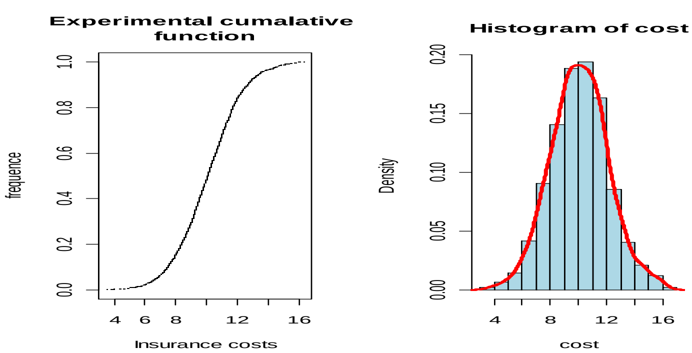

Figure 2. Empirical distribution function and density

The figures show the curves of empirical density function (right) and the curve of the empirical distribution function (left) .The function of the distribution curve gives the proportion of the interval [0; c] (where c is a value higher than 0), relative to the total sample. Note that for an amount of about 12, the frequency is in the order of 90th percentile. In other words, 90% of all the amounts are below 12F.

3.2 Extreme Statistics

The study of descriptive statistics enabled us to build the frequency curves of the different amounts and the distribution curve. Now we seek to model the tail of the distribution function. For this it is essential to determine an extreme threshold and once this threshold determined, it must be determined by the following the parameters of the theoretical distribution of excess.

3.3 Threshold Determination by the Threshold Estimation Methods

The extreme value theory, according to the approach offers different methods to estimate a threshold beyond which a case will be regarded as extreme value. One can distinguish the record values, the mean excess function and approximation GPD (generalized Pareto Distribution). This threshold should be large enough to use past results, but not too much in order to have a sufficient number of observations for quality estimates.

3.2.1 The Record Values

We assume that the extreme value theory is known (see Noureddine, et al., 2009 and Embrechts et al., 1997, p.307). With the method of record values, the threshold corresponds to a distribution value: Let the threshold $U_1 = 11{,}6$ according to data from the insurance company.

3.2.2 The Mean Excess Function

The function **mrlplot** available in the package **evir** gives a representation this graph from a sample with the software R.

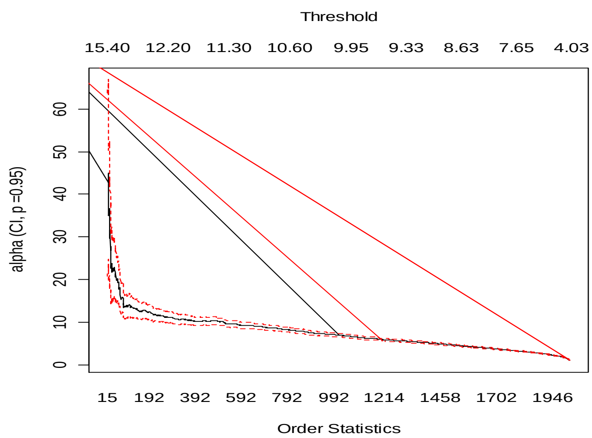

Figure 3. Hill-plot

This graph of Hill-plot, allows us to have the parameter estimates based on the statistical highest (number of excess), and we choose the most stable index. The case of the figure above is not informative enough because we do not observe stability. We can still use a graphic representation of the mean excess function.

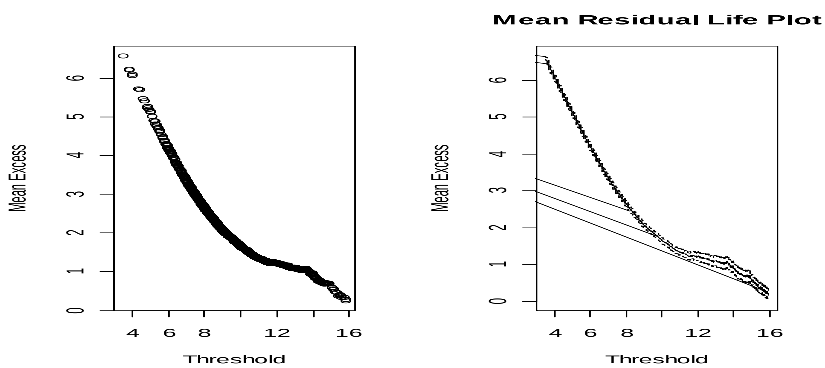

Figure 4. The mean excess function

Looking at the plot, a reasonable selection for the threshold would be 12. To complete our decision of the choice of threshold we used another tool which is that of the representation of the various parameters of a model (GPD) as a function of critical values observed on the graph that is proceeding to the limit of the linearity in the figure the mean excess function.

Using the function **tcplot** (Threshold Choice Plot), in our example we represented between 7 and 14 levels to better observe.

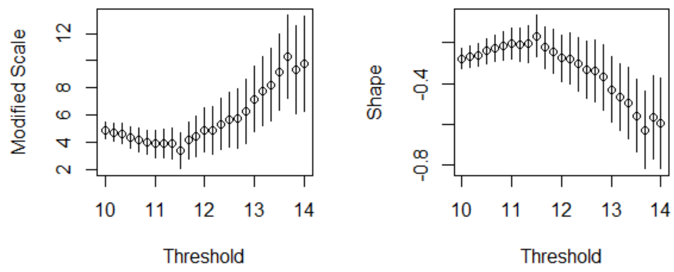

Figure 5. Choice of threshold

The combined linear stability of these representations enables us to take an equal threshold 12 for modeling. Let now a representation of the model with the threshold $u = 12$.

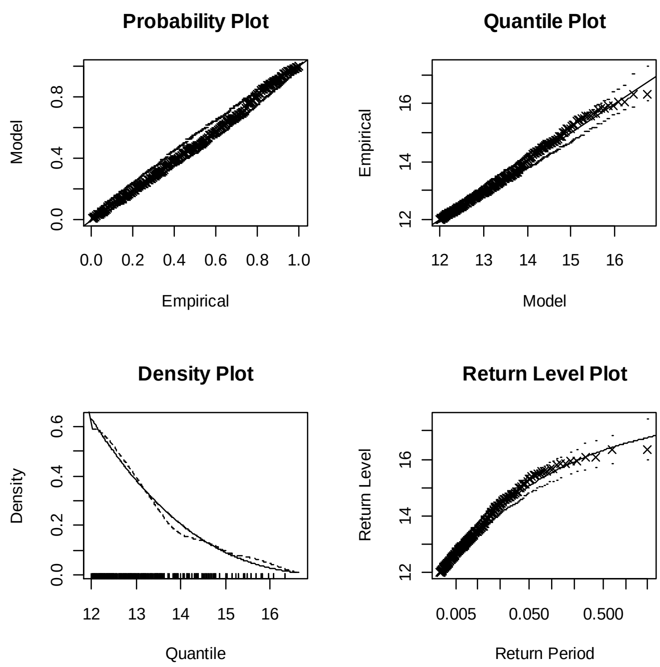

Figure 6. Representation of the model for a threshold

3.2.3 GPD Function

We observe that the graph quantile plot is linear, it can be concluded that the excess sample is adequate and approaches a model (GPD).

Once the threshold is selected, POT uses the fitgpd command to fit a GPD with the selected threshold observe that the graph quantile plot is linear, it can be concluded that

The excess sample is adequate and approaches a model (GPD).

An appropriate threshold is essential for the reliability of the excess sum model in this example:

Table 2. Results of the estimate by the likelihood method of parameters

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).The proportion of data above the threshold is described by the following table:

Table 2. Proportion of excess at the threshold 12.5

Number of Data	Threshold Call	Number Above	Proportion Above
2020	12.5	225	0.1114

The proportion above the threshold represents approximately 11.14% of the sample. The main objective of the next step is to model the tail of the distribution.

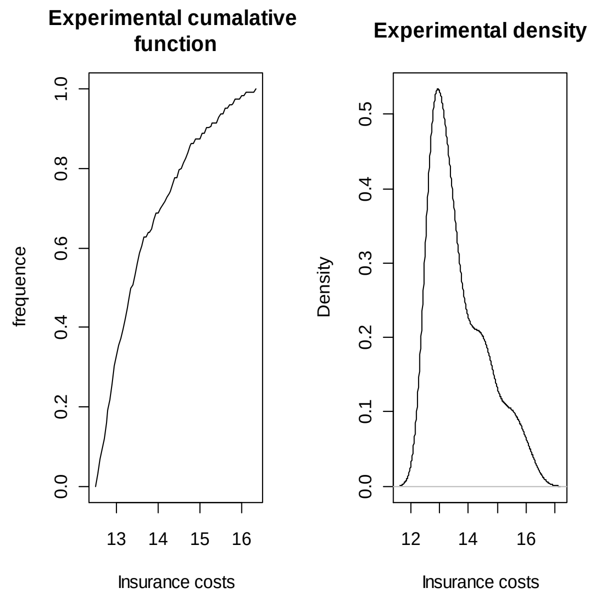

Figure 7. Tail of the empirical distribution function

Determination of the parameters of the extreme distribution

In this part, it is determine the parameters of the GPD model which modelize the excess. For this we have the fitgpdfunction, available on the R software package in POT. This function depends on several variables among which the estimation method. Here one uses three estimation methods: the method of maximum likelihood (MLE), the weighted moments method (MWP) and the method of moments (MOM) and must decide between the best methods. The best estimate is that which approach the sample excess. The results are summarized in the table below:

Summary of the GPD parameters estimation

Table 3. GPD parameters estimation

Estimator	Maximum Likelihood		Balanced moment without bias		Moment	
Parameters	Scale	Shape	Scale	Shape	Scale	Shape
	1.5576	-0.3293	1.3768	-1.1908	1.4154	-0.2241
Standard error	0.14818	0.07133	0.13987	0.08288	0.13085	0.06811
Convergence	successful		NA		NA	

The R software allows both parameter estimation and give an assessment of convergence at infinity. The software considers the asymptotic convergence of the maximum likelihood method. By cons for methods weighted moments and moments it has no information on convergence to infinity. Build the empirical density curve and the different curves probable densities, according to the estimation methods.

Curve relating to the maximum likelihood estimator (red)

Curve relating to the weighted moments unbiased estimator (blue)

Curve relating to the moments estimator (green).

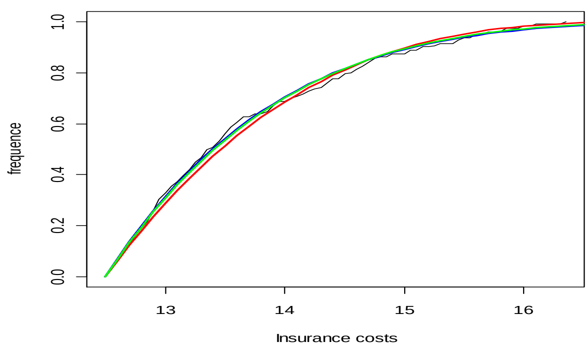

Figure 8. Empirical and theoretical distribution curves

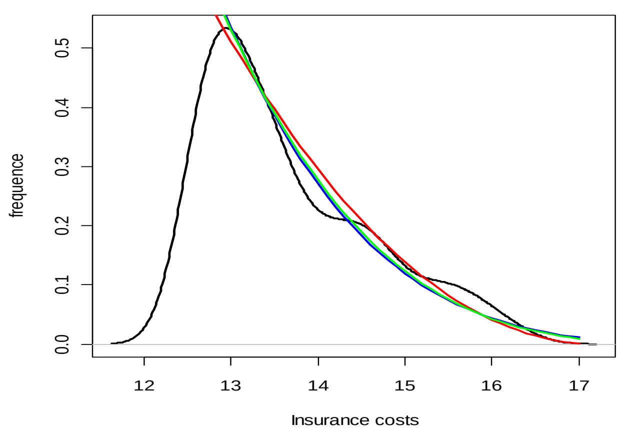

Figure 9. Empirical and theoretical densities Curves

On the two figures it can be observed the coincidence of the curves moments and that of the maximum likelihood. In addition, it can be seen that the curve weighted moments is further from the empirical curve relative to the other two.

Hence for our sample, the best method is the estimation method of maximum likelihood. Indeed, it not only gives curves (density and distribution) better approaching our excesses but also converges asymptotically. This good compromise is further supported by the results of the qq-plot function on R. She trace for a sample of size n, the pair:

$$\left\{\left(F^{-1}\left(\frac{n-k+1}{n+1}\right), R_{k,n}\right), k=1,2...,n\right\} \tag{5}$$

Where F is the distribution function of theoretical law and $R_{k,n}$ the k order statistics on the n data. A perfect fit between empirical distribution and the law tested is characterized by a perfectly linear representation of qq-plot.

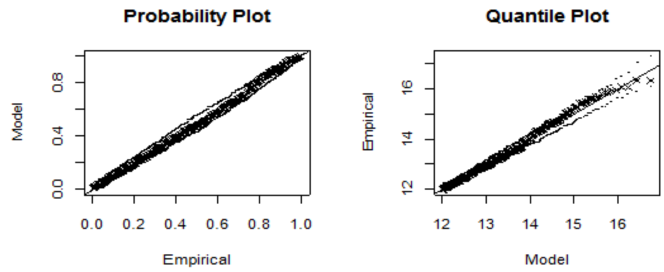

Figure 10. Comparison of distributions functions

Note that the curve of the theoretical model relative to the empirical model (respectively empirical to the theoretical model) follows the first bisecting line. Which indicates the similarity between the two models, hence the adequacy of parameters estimated by the maximum likelihood method. Summarize our findings in the following the table:

Table 4. Summary of the results

Threshold Call	Number Above	Proportion Above	Scale	shape
12.5	225	0.1114	1.5805	-0.2762

4. Calculation of the Portion of Each Claim Payable by the Reinsurer of Excess of Loss

Reinsurance excess of loss covers the portion of each individual claim excess a given priority, limited to a capacity granted by the reinsurer. We place ourselves in the collective risk model. Let N be the number of sinister and $X_1, X_2, ..., X_N$ the realizations of X, which is the random variable representing the amounts of loss. As usual we assume mutual independence of random variables.

Lemma: Let C the coverage (capacity) offered by the reinsurer. Then, the portion of each $X_j > U$ dependent the reinsurer is: $R_j = \min(C, X_j - U)$.

U_1=11,6 be the threshold beyond which a unit is declared as extreme, obtained by the record values, U_2=12 be the threshold beyond which a unit is declared as extreme, obtained by the mean excess function and U_3=12,5 the threshold beyond which a unit is declared as extreme, obtained by the GPD function.

Let $U = \alpha U_p + (1-\alpha)U_q$ with 0< α <1, minimizes the variance U, p, q=1,2,3 et p<q. We get

$$\alpha = \frac{V(X_{U_q}) - Cov(X_{U_p}, X_{U_q})}{V(X_{U_p}) + V(X_{U_q}) - 2Cov(X_{U_p}, X_{U_q})}$$ (see Noureddine and al., 2009)

For $U = \alpha U_1 + (1+\alpha)U_2 - 2\alpha U_3$ with $\alpha \in \mathbb{R}$ (α a real). We get :

$$\alpha = \frac{-V(X_{U_2}) - Cov(X_{U_1}, X_{U_2}) + 2Cov(X_{U_2}, X_{U_3})}{V(X_{U_1}) + V(X_{U_2}) + 4V(X_{U_3}) + 2Cov(X_{U_1}, X_{U_2}) - 4Cov(X_{U_1}, X_{U_3}) - 4Cov(X_{U_2}, X_{U_3})} \tag{6}$$

The thresholds obtained by the new method are the *U*.

$U=\alpha U_1+(1-\alpha)U_2$	$\alpha=0{,}58$	$U=11{,}77$	$R_j=\min(C,X_j-U)$
$U=\alpha U_1+(1-\alpha)U_3$	$\alpha=0{,}69$	$U=11{,}88$	$R_j=\min(C,X_j-U)\cdot$
$U=\alpha U_2+(1-\alpha)U_3$	$\alpha=0{,}63$	$U=12{,}19$	$R_j=\min(C,X_j-U)\cdot$
$U=\alpha U_1+(1+\alpha)U_2-2\alpha U_3$	$\alpha=-0{,}06$	$U=12{,}10$	$R_j\ \min(C,X_j-U)\cdot$

$R_j=\min(C,X_j-U)$

5. Conclusions and Recommendations

The method of convex combination minimizing the variance of two and three thresholds variables may be a very valuable tool for modeling the calculation of the portion of each sinister payable by the reinsurer of excess of sinister. It is applied after the selection of a number of different thresholds by the methods seen in the literature. Many critics have been formulated in the literature with respect to different threshold selection methods. Our technique based on a reduction of the variance of the convex combination of two and three random variables thresholds, even if it seems like a good empirical compromise quality, it must approach the experts in the field for future validation.

References

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

Allowen, A. (2008). Variations autour des Boxplots.Departement Systématique et Evolution.

Balkema, A., & de Haan, L. (1974). Residual life time at great age. *The Annals of probability, 2*(5), 792-804. http://dx.doi.org/10.1214/aop/1176996548

Charles, S. (2007). Assurances et probabilités. http://math.univ-lille1.fr /~suquet

De Haan, L., & Ferreira, A. (1984). *Extreme value theory.* Springer-Verlag.

Embrechts, P., Kluppelberg, C., & Mikosch,T. (1997). Modeming extremal events for insurance and finance. Springer, Berlin. http://dx.doi.org/10.1007/978-3-642-33483-2

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math., 44*, 423-453. http://dx.doi.org/10.2307/1968974

Mathieu, R. (2006). A User's Guide to the POT Package (Version 1.4)

Noureddine, B., & Michel, G., & Olga, V. (2009). Les sinistres graves en assurance automobile: Une nouvelle approche par la théorie des valeurs extrêmes, Revue MODULAD, Numéro 39.

Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist., 3*, 119-131. http://dx.doi.org/10.1214/aos/1176343003

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