International Journal of Statistics and Probability; Vol. 6, No. 1; January 2017
ISSN 1927-7032 E-ISSN 1927-7040
Published by Canadian Center of Science and Education

Extreme Value Theory: a New Characterization of the Distribution Function for the Mixed Method

Kané Ladji[1], Diawara Daouda[2], & Diallo Moumouni[3]

[1,3] Faculty of Economics and Management (F. S. E. G) Bamako-Mali

[2] Zhongnan University of Economics and Law Wuhan China

Correspondence: Kané Ladji, Assistant Master Professor, Faculty of Economics and Management, Bamako-Mali. E-mail: fsegmath@gmail.com

Received: November 11, 2016 Accepted: December 12, 2016 Online Published: December 21, 2016

doi:10.5539/ijsp.v6n1p71 URL: https://doi.org/10.5539/ijsp.v6n1p71

Abstract

Consider the sample $X_1, X_2, \dots, X_N$ of N independent and identically distributed (iid) random variables with common cumulative distribution function (cdf) F, and let F_u be their conditional excess distribution function. We define the ordered sample by $X_1 \leq X_2 \leq \cdots \leq X_N$. Pickands (1975), Balkema and de Haan (1974) posed that for a large class of underlying distribution functions F, and large u, F_u is well approximated by the Generalized Pareto Distribution. The mixed method is a method for determining thresholds. This method consists in minimizing the variance of a convex combination of other thresholds.

The objective of the mixed method is to determine by which probability distribution one can approach this conditional distribution. In this article, we propose a theorem which specifies the conditional distribution of excesses when the deterministic threshold tends to the end point.

Keywords: distribution function, Generalized Pareto Distribution (GPD), Mixed Method (MM).

2000 Mathematics Subject Classifications: 60G52, 60G70, 62G20, 62G32, 60E07, 62E20

1. Introduction

Pareto distribution is traditionally used by reinsurer's excess of loss mainly because of its good mathematical properties, particularly from the simplicity of the formulas resulting from its application. The new mixed method (MM) was proposed in [1, 2, 3, 4] to determine a threshold $U = \sum_{k=1}^{p} \alpha_k U_k + \alpha_3 U_3$ with $1 \leq p \leq 2$, 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%). This method allows a compromi between the GPD method and FME method, between a minimum strategy GPD and maximum strategy FME (Mean Excess Function). It is more correlated with the GPD method and relatively smooth.

2. Method

This article focuses on two major paragraphs. The first paragraph (see paragraph 3.1) is based on determining a threshold $U = \sum_{k=1}^{p} \alpha_k U_k + \alpha_3 U_3$ with $1 \leq p \leq 2$ by the mixed method (MM) and last paragraph (see paragraph 3.2) is to determine a distribution function of the laws of the mixed method. Let U_3: the threshold beyond which a unit is declared as extreme, obtained by the GPD function and U: the threshold beyond which a unit is declared as extreme, obtained by the mixed method (MM).Let $X_1, X_2, \dots, X_N$ N random variables (iid) common distribution function F. We are looking from the distribution F of X to define a conditional distribution F_{U_3} compared to U_3 threshold for random variables exceeding this threshold. It defines the excess over the threshold U_3 as the set of random variables y_j defined by: $y_j = X_j - U_3$ for $j \in E(U\_3) = \{j \in \{1,2, \dots, N\} / Xj > U3\}$. The function of distribution of the excess over the threshold U_3 is defined by:

$$F_{U_3}(y) = P(X - U_3 \leq y / X > U_3) = \begin{cases} \dfrac{F(U_3 + y) - F(U_3)}{1 - F(U_3)} & if \quad y \geq 0, \\ 0 & if \quad y < 0, \end{cases}$$

Thus, for large threshold U_3, the law of excess is approximated by a generalized Pareto law:

$$F_{U_3}(y) \approx F^{GPD}_{\xi,\sigma(U_3)}(y).$$

In this article, we will show that:

$$F_U(y) \approx F^{MM}_{\xi,\sigma(\mathrm{U})}(y),$$

where $F^{MM}_{\xi,\sigma(\mathrm{U})}(y)$ is the distribution function of the law of the mixed method and U is the threshold beyond which a unit is declared as extreme, obtained by the mixed method (MM).

Theorem Pickands (1975), Balkema and de Haan (1974) assures us that the law of the excess may be approaching a generalized Pareto law. In this article, we will use the theorem Pickands (1975), Balkema and de Haan (1974) to show that the law of the excess can be approached by a law of the mixed method.

3. Results

3.1. Determination of Threshold U By the Mixed Method (MM)

The new mixed method (MM) was proposed in [1, 2, 3, 4] to determine a threshold $U = \sum_{k=1}^{p} \alpha_k U_k + \alpha_3 U_3$ with $1 \leq p \leq 2$, 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%).

Let 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.

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

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

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

Consider the sample $X_1, X_2, \ldots, X_N$ of N independent and identically distributed (iid) random variables. We define the ordered sample by $X_1 \leq X_2 \leq \cdots \leq X_N$. 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 $X_1, X_2, \ldots, X_N$ and X_{U_j} ,which have been sorted in ascending order: $X_1 \leq X_2 \leq \cdots \leq X_k \leq U_j \leq \cdots \leq X_N$. We consider a statistical series 2 variables X and Y, taking the amount $X_1, X_2, \ldots, X_N$ and $Y_1, Y_2, \ldots, Y_N$. Which have been sorted in ascending order: $X_1 \leq X_2 \leq \cdots \leq X_N$ and $Y_1 \leq Y_2 \leq \cdots \leq Y_N$. We write:

- The means of X and Y are : $\bar{X} = \dfrac{1}{N} \sum_{i=1}^{N} X_i$ et $\bar{Y} = \dfrac{1}{N} \sum_{i=1}^{N} Y_i$

- 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: $Cov(X,Y)=\frac{1}{N}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})$

Example 1. Threshold Calculation

The data base provides a sample of 2020 observations for 4 wheel vehicle for personal use during the year 2013. The 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. 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. U_3 the threshold beyond which a unit is declared as extreme, obtained by the GPD function and U the threshold beyond which a unit is declared as extreme, obtained by the method MM.

Let N be the number of claims and $X_1, X_2, \dots, X_N$ the realizations of X, which is the random variable representing the amounts of loss. As usual we assume mutual independence of random variables.

Table 1. Determination of threshold U by the mixed method (MM)

Record values	Mean excess function	GPD function	MM method
U_1	U_2	U_3	$U=\sum_{k=1}^{p}\alpha_k U_k+\alpha_3 U_3$ with $1\le p\le 2$
$U_1=11{,}5$		$U_3=12{,}5$	$U=\alpha U_1+(1-\alpha)U_3=11{,}88$ with $\alpha=0{,}69$
	$U_2=12$	$U_3=12{,}5$	$U=\alpha U_2+(1-\alpha)U_3=12{,}19$ with $\alpha=0{,}63$
$U_1=11{,}5$	$U_2=12$	$U_3=12{,}5$	$U=\alpha U_1+(1+\alpha)U_2-2\alpha U_3=12{,}10$ with $\alpha=-0{,}06$

3.2 Law (distribution) of The Mixed Method

In this section, we will give the main result of this paper is to write a new law of the mixed method (MM). Let U_3*:* the threshold beyond which a unit is declared as extreme, obtained by the GPD function and U: the threshold beyond which a unit is declared as extreme, obtained by the mixed method (MM). Let $X_1, X_2, \dots, X_N$ N random variables (iid) common distribution function F. We are looking from the distribution F of X to define a conditional distribution F_{U_3} compared to U_3 threshold for random variables exceeding this threshold. It defines the excess over the threshold U_3 as the set of random variables y_j defined by: $y_j=X_j-U_3$ for $j\in E_{U_3}=\{j\in\{1,2,\dots,N\}/X_j>U_3\}$. It defines the excess over the threshold U_3 as the set of random variables y_j defined by:

$$F_{U_3}(y) = P(X - U_3 \le y / X > U_3) = \begin{cases} \dfrac{F(U_3 + y) - F(U_3)}{1 - F(U_3)} & if \quad y \ge 0, \\ 0 & if \quad y < 0, \end{cases}$$

The objective of the mixed method is to determine by which probability distribution one can approach this conditional distribution. In this article, we propose the following theorem (**Theorem 2**) which specifies the conditional distribution of excesses when the deterministic threshold tends to the end point X_F.

Theorem 1 (Pickands (1975), Balkema and de Haan (1974)): Let F_{U_3} be the conditional distribution of the excess over a threshold U_3, combined with unknown distribution function F. This function F belongs to the domain of attraction of G_ξ if and only if there exist a positive function σ such

$$\lim_{U_3 \to X_F} \operatorname{Sup}_{0 < y < X_{F - U_3}} \left| F_{U_3}(y) - F^{GPD}_{\xi,\sigma(U_3)}(y) \right| = 0,$$

Where $F^{GPD}_{\xi,\sigma(U_3)}$ is the distribution function of GPD, define by:

$$F^{GPD}_{\xi,\sigma(U_3)}\,(\mathrm{y}) = \begin{cases} 1 - \left(1 + \dfrac{y\xi}{\sigma}\right)^{\frac{-1}{\xi}} & if\ \xi \ne 0 \\ 1 - e^{\frac{-y}{\sigma}} & if\ \xi = 0 \end{cases}$$

for $y \in [0, (X_F - U_3)]$ if $\xi \ge 0$ and $y \in \left[0, Min\left(\frac{-\xi}{\rho}, X_F - U_3\right)\right]$ if $\xi < 0$ with $X_F = sup\{X \in \mathbb{R}, F(X) < 1\}$

Theorem 2: Let F_U be the conditional distribution of the excess over a threshold $U = \sum_{k=1}^{p} \alpha_k\, U_k + \alpha_3 U_3$ with $1 \le p \le 2$, combined with unknown distribution function F. This function F belongs to the domain of attraction of G_ξ if and only if there exists a positive function σ such

$$\lim_{U \to X_F} \operatorname{Sup}_{0 < y < X_{F - U}} \left| F_U(y) - F^{MM}_{\xi,\sigma(\mathrm{U})}(y) \right| = 0,$$

Where $F^{MM}_{\xi,\sigma(\mathrm{U})}$ is the distribution function of mixed method (MM), define by:

$$F^{MM}_{\xi,\sigma(\mathrm{U})}\,(\mathrm{y}) = \begin{cases} 1 - \left(1 + \dfrac{y\xi}{\sigma}\right)^{\frac{-1}{\xi}} & if\ \xi \ne 0 \\ 1 - e^{\frac{-y}{\sigma}} & if\ \xi = 0 \end{cases}$$

for $y \in [0, (X_F - U)]$ if $\xi \ge 0$ and $y \in \left[0, Min\left(\frac{-\xi}{\rho}, X_F - U\right)\right]$ if $\xi < 0$ with $X_F = sup\{X \in \mathbb{R}, F(X) < 1\}$.

Proof: The conditional distribution F_U of the excesses above the threshold U with is defined by:

$F_U(y) = P(X - U \le y / X > U) = \frac{F(U+y) - F(U)}{1 - F(U)}$ for $0 \le y \le X_F - U$.

This is equivalent to:

$F_U(x) = P(X < x / X > U) = \frac{F(x) - F(U)}{1 - F(U)}$ for $x \ge U$.

The proof of $F_U(y) \approx F^{MM}_{\xi,\sigma(\mathrm{U})}(y)$ results directly from the evidence of $F_{U_3}(y) \approx F^{GPD}_{\xi,\sigma(U_3)}(y)$ (theorem 1) and $U = \sum_{k=1}^{p} \alpha_k\, U_k + \alpha_3 U_3$ with $1 \le p \le 2$.

Example 2: Winting $\mathrm{F}^{\mathrm{MM}}_{\xi,\sigma(\mathrm{U})}$ (y).

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, *U_3*: the threshold beyond which a unit

is declared as extreme, obtained by the GPD function and U: the threshold beyond which a unit is declared as extreme, obtained by the MM function. Let N be the number of claims and $X_1, X_2, \dots, X_N$ the realizations of X, which is the random variable representing the amounts of sinister.

Table 2. Knowing the parameters (σ, ξ) and the thresholds, we can write the distribution functions of GPD and MM.

Parameters GPD $\xi,\ \sigma$	Threshold GPD U_3	Threshold MM U	distribution function of GPD $F^{GPD}_{\xi,\sigma(U_3)}(y) = 1 - \left(1 + \frac{y\xi}{\sigma}\right)^{\frac{-1}{\xi}}$ With $y = X - U_3$	distribution function of MM $F^{MM}_{\xi,\sigma(U)}(y) = 1 - \left(1 + \frac{y\xi}{\sigma}\right)^{\frac{-1}{\xi}}$ With $y = X - U$
$\xi = -0{,}3293$ $\sigma = 1{,}5576$	$U_3 = 12{,}5$	$U = 11{,}88$	$1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$ $y = X - 12{,}5$	$1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$ $y = X - 11{,}88$
$\xi = -0{,}3293$ $\sigma = 1{,}5576$	$U_3 = 12{,}5$	$U = 12{,}19$	$1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$ $y = X - 12{,}5$	$1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$ $y = X - 12{,}19$
$\xi = -0{,}3293$ $\sigma = 1{,}5576$	$U_3 = 12{,}5$	$U = 12{,}10$	$1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$ $y = X - 12{,}5$	$1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$ $y = X - 12{,}10$

Example 3: Threshold Calculation By the Graphical Method.

Knowledge of parameters (σ, ξ) allows to determine graphically the threshold U_3 by the GPD method and U by MM method (mixed method). To do this, we will write a program on the MAPLE software to determine these thresholds.

Distribution Function $F^{MM}(\xi, \sigma, \mathrm{u})$:

$$Distribution := \mathbf{proc}(\zeta, \sigma, u)\ \mathbf{local}\ i, j; plot\left(1 - \left(1 + \frac{\zeta.(U-u)}{\sigma}\right)^{\frac{-1}{\zeta}}, U = 0..20, thickness = 2, color = black\right) \mathbf{end};$$

$$Density := \mathbf{proc}(\zeta, \sigma, u)\ \mathbf{local}\ i, j; plot\left(\frac{1}{\sigma} \cdot \left(1 + \frac{\zeta.(x-u)}{\sigma}\right)^{\frac{-1-\zeta}{\zeta}}, = 0..20, thickness = 2, color = black\right) \mathbf{end};$$

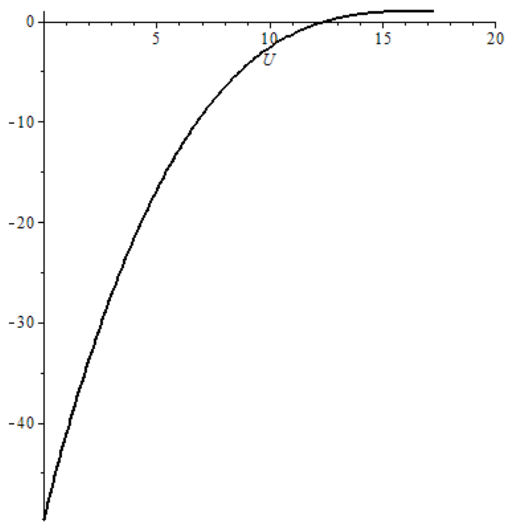
Table 3. Knowing the parameters (σ, ξ) and the thresholds, we can graphically read the thresholds

<table>
<tr>
<td>

GPD method:

$\xi = -0{,}3293,\ \sigma = 1{,}5576$ with $U_3 = 12{,}5$

$F^{GPD}_{\xi,\sigma(U_3)}(y) = 1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$



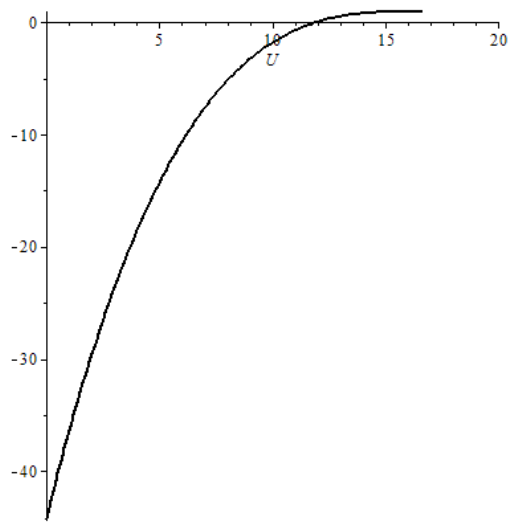
The threshold can be read graphically $U_3 = 12{,}5$

</td>
<td>

MM method:

$\xi = -0{,}3293,\ \sigma = 1{,}5576$ with $U = 11{,}88$

$F^{MM}_{\xi,\sigma(U)}(y) = 1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$



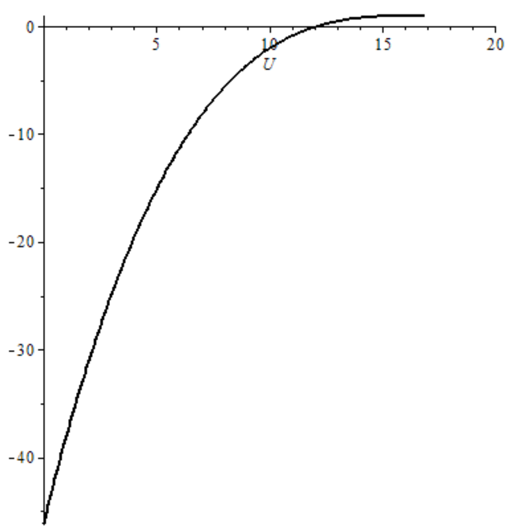
The threshold can be read graphically $U = 11{,}88$

</td>
</tr>
<tr>
<td>

MM method:

$\xi = -0{,}3293,\ \sigma = 1{,}5576$ with $U = 12{,}10$

$F^{MM}_{\xi,\sigma(U)}(y) = 1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$



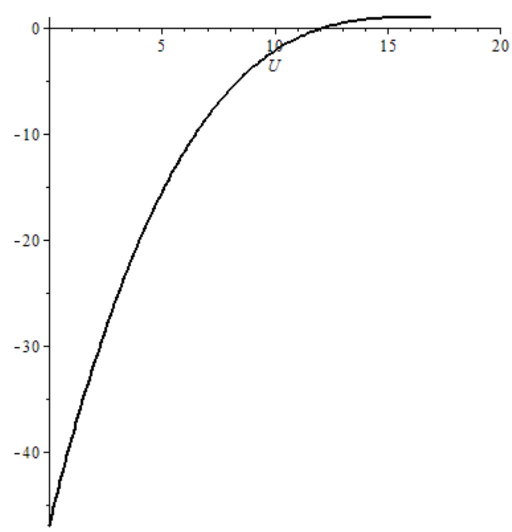
The threshold can be read graphically $U = 12{,}10$

</td>
<td>

MM method:

$\xi = -0{,}3293,\ \sigma = 1{,}5576$ with $U = 12{,}19$

$F^{MM}_{\xi,\sigma(U)}(y) = 1 - \left(1 - \frac{0{,}3293y}{1{,}5576}\right)^{\frac{1}{0{,}3293}}$



The threshold can be read graphically $U = 12{,}19$

</td>
</tr>
</table>

4. Conclusions

In the literature, various methods have been proposed to estimate the parameters (σ, ξ) of the GDP law: the method of maximum likelihood, method of moments, the Bayesian method, the estimator Pickands (1975) and the Hill estimator (1975). Note that these last two estimators can only be used for the index of the tail of the distribution ξ. Knowledge of parameters (σ, ξ) by the first two methods allows to determine graphically the threshold U_3 by the GPD method and U by MM method (mixed method). Therefore, we must carefully determine the parameters by different methods.

References

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

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

Balkema, A., & Dehaan, 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.

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

Embrechts, P., Kluppelberg, C., & Mikosch, T. (1997). *Modelingextremal 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, 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.

Xu, Y. M., Daouda, D., & Ladji, K. (2016). An application of extreme value theory in automobile insurance: A new approach to the convex combination of two variables thresholds minimizing the variance of this combination. *Advances and Applications in Statistics, 48*(2), 91–108.

Xu, Y. M., Ladji, K., & Daouda, D. (2016). Mixed method of extreme value theory, with application to the calculation of the portion of each claim payable by the reinsurance of excess of sinister. *International Journal of Statistics and Probability, 5*(2), 1927–7032.