International Journal of Management and Applied Science, ISSN: 2394-7926 Volume-3, Issue-3, Mar.-2017
http://iraj.in

MULTIVARIATE EXTREME VALUE THEORY AND APPLICATION TO ENVIRONMENT

[1]SAMIA AYARI, [2]MOHAMED BOUTAHAR

Business and Economic Statistics Modeling Laboratory, High Institute of Management of Tunis, University of Tunis, Tunisia,
Institute of Mathematics of Marseille, Faculty of Sciences, University of Aix-Marseille, France.
E-mail: samia.ihec@hotmail.fr, mohamed.boutahar@univ-amu.fr

Abstract— In this paper, we are interested in modeling the dependence structure for a multivariate extreme value distribution. A Monte Carlo simulation is conducted to compare the performance of different nonparametric estimators of the dependence function. Experimental results indicate that the Zhang estimator has better performance than the other dependence function estimators. We finally carry out an empirical application to the air pollution data in Tunisia, in order to illustrate the usefulness of the dependence function estimators to describe the link between extreme ozone concentrations observed in some polluted stations.

Index Terms— multivariate extreme value distribution, dependence function, air pollution, ozone modeling

I. INTRODUCTION

Air pollution constitutes an important environmental problem due to the harmful effects caused by exposing to air pollutants existing in urban areas. Several epidemiological studies have indicated a strong link between increased air pollution and increased mortality and morbidity from all causes. Regarding the air quality problem, it's necessary to provide information about the concentration of these pollutants in order to alert the public of the onset, severity and duration of poor air quality conditions. The causes of air pollution episodes are complex and depend on various factors including solar radiation, chemical processes, emissions, meteorological parameters, etc... that's why forecasters faced big difficulties in air quality time series modeling.

Going through the literature, a lot of research efforts go towards the understanding of air quality phenomenon and the ability to forecast it.

Several deterministic and statistical models have been developed to evaluate and predict the pollutant dispersion and especially peak pollution levels such as regression models: [8], neural networks: [10], Kalman filter: [18], extreme value models: [13], etc...

The extreme value theory was made about 50 years ago; it is a discipline that seeks to model the stochastic behavior of extreme values and to estimate the probability of the rare events more extreme than those observed in the past. This theory has attracted the interest of many specialists in different fields such as: finance: [3], environment: [1], engineering: [9] and economy: [6], etc...There are two approaches in the extreme value theory: the univariate extreme value theory and the multivariate extreme value theory. Up to now, the multivariate extreme value theory hasn't be applied in air pollution field and especially in modeling ozone dependence data.

We aim in this study to model the dependence structure of multivariate extreme value distribution. Different nonparametric estimators of the dependence function of multivariate extreme value distribution are developed in the R software. An application to the air pollution data in Tunisia will be made in order the model the dependence structure for ozone data.

II. MULTIVARIATE EXTREME VALUE THEORY

Classical univariate extreme value models require strong and unrealistic assumptions such as the independence of observations. It was noticed that in most applications of rare events, more than one random variable should be considered. In this context [17] points that most real extreme value problems are inherently multivariate by nature. Within this framework, researches in extreme value theory are more oriented towards multivariate extreme value analysis, Stuart and [16], [4]. In this context, the copula theory has played an important role in modeling multivariate extreme value distributions, [16], [5]. Copulas can decompose a distribution into two components: the marginal distributions, on one hand, and a mathematical function that links them by modeling their dependence, on the other hand.

Assume that X_j=($X_{1,j}$,..., $X_{n,j}$), let F_j , $1 \le j \le p$, be the common extreme value distribution of $X_{i,j}$, $1 \le i \le n$. The joint distribution of X=(X_1 ,…, X_p) satisfies F(x_1,…,x_p)= C(F_1 (x_1),…,F_p (x_p)) where C is a multivariate extreme value copula. According to [11], the copula C depends on (p-1) dimensional dependent function A(s_1 ,..., s_{p-1}) as following

$$C(F_1(x_1)\ldots,F_p(x_p)) = C(u_1,\ldots,u_p) = \exp\left(\sum_{k=1}^{p} \log(u_k) \quad A\left(\frac{\log u_1}{\sum_{k=1}^{p}\log(u_k)},\ldots,\frac{\log u_{p-1}}{\sum_{k=1}^{p}\log(u_k)}\right)\right)$$

Note that if $A(s_1,\ldots,s_{p-1}) = 1$ then the random vectors X_j, $1 \le j \le p$, are independents. We will compare the following four estimators of the dependent function $A(s_1,\ldots,s_{p-1})$:

1. The Pickands estimator, [11]:

$$A^P(s_1,\ldots,s_{p-1}) = \frac{n}{\sum_{i=1}^{n} \min_{1\le j\le p} \frac{Y_{ij}}{s_j}}$$

2. The Hall-Tajvidi estimator,[7]:

$$A^{HT}(s_1,\ldots,s_{p-1}) = \frac{n}{\sum_{i=1}^{n} \min_{1\le j\le p}\left(\frac{Y_{ij}/\bar{Y}_j}{s_j}\right)}$$

3. The Deheuvels estimator, [2]:

$$A^D(s_1,\ldots,s_{p-1}) = \frac{n}{\sum_{i=1}^{n} \min_{1\le j\le p}(Y_{ij}/s_j) - n\sum_{j=1}^{p} s_j\bar{Y}_{ij} + n}$$

4. The Zhang estimator,[19]:

$$A^Z(s_1,\ldots,s_{p-1}) = \frac{\prod_{j=1}^{p}\left(\sum_{i=1}^{n} Y_{ij}\right)^{s_j}}{\sum_{i=1}^{n} \min_{1\le j\le p}(Y_{ij}/s_j)}$$

where

$$Y_{ij} = -\log(F_j(X_{ij})), \bar{Y}_j = \frac{1}{n}\sum_{i=1}^{n} Y_{ij}, 1 \le j \le p, \sum_{j=1}^{p} s_j = 1$$

III. SIMULATION STUDY

The first step of our study is to simulate trivariate extreme value distributions whose dependence function are logistic-type.

[14] has developed algorithms for simulating symmetric versions of multivariate extreme value distribution. The algorithms are included in the R “evd” package.

We are limited to the logistic version of multivariate extreme value distributions since it represents the most flexible existing parametric forms for general dimensions, [14].

Explicitly, the logistic dependence function of trivariate extreme value distributions is presented as follows:

$$A(s_1,s_2) = \{\theta^r s_1^r + \varphi^r s_2^r\}^{\frac{1}{r}} + \{\theta^r s_2^r + \varphi^r s_3^r\}^{\frac{1}{r}} + \{\theta^r s_3^r + \varphi^r s_1^r\}^{\frac{1}{r}} + \psi\{s_1^r + s^r{}_2 + s_3^r\}^{\frac{1}{r}} + 1 - \theta - \phi - \psi,$$

where $s_3 = 1 - s_1 - s_2$.

Here we consider the symmetric logistic dependence function with r=3, $\theta = \phi = 0$ and $\psi = 1$.

The simulation is repeated for different sample sizes : n=100, n=200 and n=500.

For each case, we generate 10000 data sets, the four estimators cited above are compared.

To assess the performance of the estimators, the mean integrated square errors MISE is used in our study.

Table1. Mean integrated square errors MISE for different sample sizes.

Estimator	n=100	n=200	n=500
Pickands	0.01468744	0.01463039	0.01459532
Deheuvels	0.04719728	0.04718369	0.0472438
Hall-Tajvidi	0.0146899	0.01462972	0.0145951
Zhang	0.01468065	0.01462637	0.01459444

As claimed in [7], the Hall-Tajvidi estimator has better performance than the Deheuvels estimator in terms of MISE for different sample sizes.

Table1 shows that the Deheuvels estimator has worse performance than the Pickands and the Zhang estimator for different sample sizes. The Pickands, the Hall-Tajvidi and the Zhang estimators have similar performance in terms of MISE. The best one is the Zhang estimator.

IV. AN APPLICATION TO THE AIR POLLUTION DATA IN TUNISIA

Tunisia has experienced in recent decades a significant urban growth in vehicles and road traffic even for industrial activities. This has caused majority air pollution by oxides of sulfur, ozone, nitrogen dioxides and suspended particulates. The preservation of the atmosphere is one of the priorities of the environmental policy in Tunisia. One of the taken measure to protect the air quality is monitoring the air. This measure has been taken by the specialists of environment in Tunisia, in order to provide information about the concentration of pollutants and to alert the public of the onset, severity and duration of poor air quality conditions. In light of this situation, the National Agency for Protection of the Environment has started a project which consists on the development of a device for monitoring air quality. The national network for monitoring air quality, which was established in 1996, consists of 15 fixed stations and a mobile laboratory. These stations are equipped with analyzers and instruments for measuring pollutants such as sulfur dioxide, oxides of nitrogen, the solid particles, nitrogen monoxide and ozone.

A. Descriptive statistics of ozone data in Tunisia

Ozone is a natural component of the troposphere, produced by the photochemical reaction of nitrogen oxides and volatile organic compounds. Numerous scientific studies have linked ozone exposure to a variety of problems, including: increased respiratory symptoms, such as irritation of the airways, coughing, or difficulty breathing, aggravated asthma, development of chronic bronchitis, irregular heartbeat, nonfatal heart attacks, premature deaths in people with heart or lung disease. It can also have harmful effects on animals, plants, air quality...etc

From a regulatory perspective, Tunisia has set its own limit value 235 $\mu g.m^{-3}$ and guideline value (150-200 $\mu g.m^{-3}$) for hourly ozone concentrations.

Descriptive statistics show that ozone recorded several exceedances for both guideline and limit value of the Tunisian standard NT 106.04. Nahli Park, Ghazela and Mourouj Park are the most polluted stations which recorded high concentration levels of ozone. So, we are interested in modeling the dependence structure of ozone concentrations observed in these stations.

B. Modeling the dependence structure of ozone data in Tunisia

Here we explore the dependence structure of ozone data observed in three stations in Tunisia: Nahli Park, Mourouj Park and Ghazela stations from 2005 to 2012. At first, the monthly maximum ozone concentrations were recorded from 2005 to 2012. Second, we take maximum-likelihood fitting for the generalized extreme value distributions for monthly maximum ozone concentrations observed in each station.

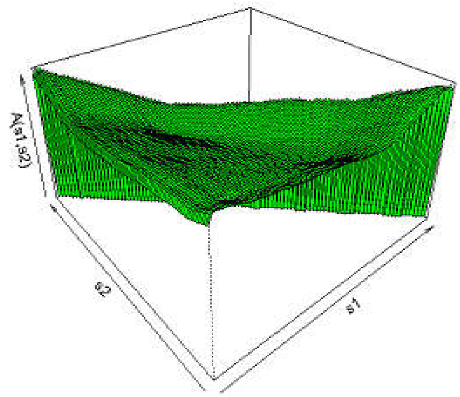

Figure1. The Zhang estimator of the dependent function A.

The figure 1 indicates that there is a small dependence, (since $A^Z(s_1, s_2)$ is nearly equal to 1) between extreme ozone concentrations observed in Nahli Park, Mourouj Park and Ghazela station. Similar results can also be found by the Pearson's and the Spearman's coefficients for bivariate maximum ozone concentrations.

REFERENCES

[1] S. Coles and E. Casson, “Extreme value modelling of hurricane wind speeds,” Structural Safety, vol. 20, pp. 283-296, 1998.

[2] P. Deheuvels, “On the limiting behavior of the Pickands estimator for bivariate extreme-value distributions,” Statistics and Probability Letters, vol. 12, pp. 429-439, 1991.

[3] P. Embrechts, C. kluppelberg and T. Mikosch, “Modeling Extremal Events for Insurance and Finance,” Springer, Berlin, 1997.

[4] H. Ferreira, “Dependence between two multivariate extremes, “Statistics and Probability Letters, vol. 81, pp. 586 -59, 2011.

[5] E.W Frees and E.A. Valdez, “Understanding relationships using copulas,”North American Actuarial Journal., vol. 2, pp. 1-25, 1998.

[6] R. Gençay and F. Selçuk, “Extreme value theory and value-at-risk: relative performance in emerging markets ,” Forecasting Economic and Financial Time Series Using Nonlinear Methods, vol. 20, pp. 287-303, 2004.

[7] P. Hall and N. Tajvidi. “Distribution and dependence-function estimation for bivariate extreme value distributions,” Bernouilli., vol. 5, pp. 835 –84, 2000.

[8] Hubbard and W. Cobourm, “Development of a regression model to forecast ground-level ozone concentration in Louisville, KY,“ Atmospheric Environment, vol. 32, pp. 14-15, 1996.

[9] M. Onorato, A.R. Osborne and M. Serio, “Extreme wave events in directional random oceanic sea states,” Physics of Fluids, vol. 14, pp. 25-28, 2002.

[10] P. Perez, A. Trier and J. Reyes,”Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile,” Atmospheric Environment, vol. 34, pp. 1189-1196, 2000.

[11] J. Pickands, “Multivariate Extreme Value Distributions,” Bulletin of the International Statistical Institute, vol. 49, pp. 859-878, 1981.

[12] J. Pickands, “Statistical inference using extreme order statistics,” The Annals of Statistics, vol. 3, pp. 119-131. 1975.

[13] T. Rao, G. Sistla, V. Pagnoiti, W.B. Petersen, J.S. Irwin and D.B. Turner, “Resampling and extreme Value statistics in air quality model performance evaluation,” Atmospheric Environment, vol. 19, pp. 1503—1518, 1985.

[14] S. A. Stephenson, “Simulating multivariate extreme Value distributions of logistic type,” Extremes, vol. 6, pp. 49—59, 2003.

[15] C.G. Stuart and J.A Tawn, “Modeling extreme multivariate events,“ Journal of the Royal Statistical Society, Series B (Methodological), vol. 53, pp. 377-392, 1991

[16] J.A. Tawn, “Modelling multivariate extreme value distribution,” Biometrica, vol. 77, pp. 245-253, 1990.

[17] J.A. Tawn, “Applications of multivariate extremes in extreme value theory and its applications,” J. Galambos, Kluwer, 1994.

[18] J.T. Van del Wal and L.H.J.M. Janssen,”Analysis of spatial and temporal variations of PM 10 concentrations in the Netherlands using Kalman filtering,” Atmospheric Environment, vol. 34, pp. 3675—368, 2000.

[19] D. Zhang, M.T. Wells and L. Peng, “Nonparametric estimation of the dependence function for a multivariate extreme value distribution,“ Journal of Multivariate Analysis , vol. 99, pp. 577-588, 2008.