Risk Analysis Method by the Extreme Data of Dependent Exogenous Variables

Submitted: 26th July 2021; accepted: 10th March 2022

Ihor Tereshchenko, Anton Tereshchenko, Nataliya Bilous, Svetlana Shtangey, Zygmunt L. Warsza

DOI: 10.14313/JAMRIS/3-2021/18

Abstract:
Many practical tasks of data multivariate statistical analysis from the standpoint of a risk-oriented process approach (in accordance with ISO 9001: 2015, 31000: 2018) requires the definition of the risk values for the dependent exogenous variables of some processes. This paper proposes the method, which consist of original stages sequence for calculating value-at-risk (VaR) or conditional-value-at-risk (CVaR) of dependent exogenous variables, presented of the extreme data frame of critical manufacture process parameters or other parameters, for example, extreme data of environmental monitoring and etc. Risk analysis method by the extreme data of dependent exogenous variables, presented of the data matrix, uses the result of solving the formalized problem of defines the tails parameters of the joint distributions of exogenous variables as components of a bivariate random variable. It can be argued that the tails parameters of the joint distributions of dependent exogenous variables make the validated corrections of the VaR and CVaR estimates for such variables. This method expands the practical application of extreme value theory for the value at risk analysis of any dependent variables as process parameters.

Keywords: *exogenous variables; risk-oriented process approach; extreme value theory; tailed distribution*

Abbreviations:

VaR:	Value at Risk
CVaR:	Conditional Value at Risk
GPD:	Generalized Pareto Distribution
GEVD:	Generalized Extreme Value Distribution
POT:	Peaks-Over-Threshold
EVS:	Extreme Value Statistics
ES:	Expected Shortfall
EVT:	Extreme Value Theory
QMS:	Quality Management System
EI:	Extremal Index
EVI:	Extremal Value Index
Q-Q:	Quantile-Quantile
CI:	Confidence Interval

Nomenclature:

$\{X_i\}_{i\geq 1}$:	The sequence of independent and identically distributed (i.i.d.) vector of random exogenous variables
$X_i = (x_{i1},\dots,x_{iD})$:	Vector of random exogenous variables
$i = 1,\dots,n$:	The sequence time numbers of multivariate elements (exogenous variables)
$D\,(1 \leq d \leq D)$:	Dimension the vector of random exogenous variables
$F()$:	The marginal distribution function of exogenous variable
x_{id}:	The exogenous variable values
$M_n = (M_{n1},\dots,M_{nD})$:	The data-frame of componentwise (exogenous variables) maximums
$M_{nd} = \max\{x_{i1},\dots,x_{iD}\}$	Vector of exogenous variables maximums
$M = \{M_{nd}\}$:	The $n\times D$-dimensional array of maximums (extremes) of *n* statistical *D*-dimensional exogenous variable observations
a_n, b_n:	The constants for a non-degenerate distribution function *G*
$G(\,)$:	The generalised extreme value distribution function
ξ_d, μ_d, σ_d:	Extremal index, location parameter, scale parameter of marginal exogenous variables distributions respectively
χ:	The probability of one variable being extreme given that the other is extreme
u:	Threshold
Z:	The standardized residuals from the fitted model of Heffernan & Tawn

Introduction

By variables further mean exogenous variables whose cause is external to the model and whose role is to explain other variables or outcomes in the model [1].
The extreme values of process's variables are the values which equal to, close to or exceed the limit values acceptable to certain requirements (for example, the

requirements of branch standards for product's quality, environmental standards, etc.). Therefore mentally and functionally investigations are based on the general concept of risk-oriented management by ISO 9001: 2015/ISO 31000: 2018 and the methodology of the process approach [2-4].

A software solution reviews for the extreme value statistics [5-7] and related risk analysis [8-11] indicates to a positive trend of the researchers' interest to these areas of statistics for the information technologies and a high relevance level of such investigations.

The evolution of the risk assessment methodology and the market success of the RiskMetrics Group [12, 13] can serve as a good example of the dynamic growth of progress and the usefulness of applied risk analysis.

Dynamic trading strategies can be considered as branch trends of risk analysis with a good analogy of application [14]. These strategies targets on a predefined level of risk as measured by volatility, Value at Risk (VaR) or Conditional Value at Risk (CVaR). Investigations have shown that targeting increases the risk-adjusted performance and heightens utility gains for mean-variance investors. As such, the target factorial direction of researches is accented.

Many solutions useful for practical application in this area of statistics are concentrated in modern software products of the R programming language.

So, it is necessary to note the "fExtremes" software package [15]. This software package contains some functions were implemented from Alec Stephenson's R-package "evir" [16] imported from Alexander McNeil's S library "EVIS" [17], extreme values in S programming language, some are from Alec Stephenson's R-package "ismev" [18] based on Stuart Coles code from his book "Introduction to Statistical Modeling of Extreme Values" [19] and some were written by Diethelm Wuertz. The topics of "fExtremes" package includes: data preprocessing, explorative data analysis, peak over threshold modeling, block maxima modeling, estimation of VaR/CVaR and the computation of the extreme index.

It is necessary to note the content of the R-package "ReIns" (February 10, 2020) [20, 21] for risk analysis. The important position of the book [20] is that in the risk analysis, a global fit that appropriately captures the body and the tail of the variables distributions is essential. The whole range modeling of the variables using a standard distribution is usually very hard and often impossible due to the specific characteristics of the body and the tail of the distributions of variables. A possible solution is to combines two distributions in a splicing model [21]: a light-tailed distribution for the body which covers light and moderate variables, and a heavy-tailed distribution for the tail to capture large variables.

Note, that the proposed solution is an adequate alternative to the approximation of tail values by the Pareto distribution.

The R-package "texmex" [22] also has the actual content of the approaches that have been implemented. This software package for the statistical extreme value modeling of threshold excesses, maxima and multivariate extremes uses univariate models for threshold excesses and maxima are the Generalized Pareto Distribution (GPD) and Generalized Extreme Value Distribution (GEVD). Also, for serially dependent sequences, the intervals declustering algorithm of Ferro & Segers [23, 24] is provided, with diagnostic support to aid selection of threshold, declustering horizon and the computation of the Extremal Index (EI). Multivariate modeling is performed via the conditional approach of Heffernan & Tawn [25, 26], with graphical tools for threshold selection and to diagnose estimation convergence.

The R-package "tsxtreme" also allows characterizing the extremes values dependence structure of time series via the Peaks-Over-Threshold (POT) methods [27]. It uses the Heffernan & Tawn conditional approach [26] which is flexible in terms of extremal and asymptotic dependence structures, and Bayesian methods improves efficiency and allow for deriving measures of uncertainty. For example, the EI, related to the size of clusters in time, can be estimated and samples from its posterior distribution obtained.

Thus, the use of the conditional Heffernan & Tawn's approach is justified for solving the relevant problem of the risk analysis by the extreme data of dependent exogenous variables as the main goal of this article.

It is important that the Heffernan & Tawn's method, evaluated and extrapolated the distribution of a two-dimensional random variable is developed a semi-parametric approach that overcomes the limitations to arguments when components of the two-dimensional variable become large at the same rate.

For the software packages that are considered it is important to note, that the POT methodology for Extreme Value Statistics (EVS) is applied, which uses more data and allows evaluating the behavior of extreme values above some high threshold [28, 29]. POT has an advantage because it allows to inference directly on the distribution of variables extremes [28].

So, at statistical estimates the extremes of the multivariate data the problem of the dependence (in particular, correlation) of the components (covariates of variables) of multivariate observations are inevitably arises. For this currently relevant problem, the review [30] considers the most interesting examples of strongly correlated variables for which there are very few exact of the EVS.

In the paper [31] a test for detecting sequential correlations in multivariate time series was proposed. This test uses Spearman's rank correlation properties and the theory of extreme values.

Risk assessment can be performed using the model for maxima that can be obtained by combining the GEVD for the univariate marginal distributions with extreme-value copulas to describe their dependence structure, as justified by the theory of multivariate extreme values [32]. Here it is advisable to note the use of the copula mathematical tool [33–36] for analysis and assessments of nonlinear dependencies of variables. At present, paired copula models are mainly used, for which a mathematical tools and software methods have been developed.

The paper [37] describes a multivariate statistical dependency model for hydrological observations, which used to estimate flood losses (i.e. risks) in a large and heterogeneous region.

So, the methodology of the extreme values theory and the subject of risk analysis are currently relevant and are considered in the overwhelming majority of cases for the financial, in particular the exchange, branches [7, 11]. Therefore, one of the article goals is to expand the application of the extreme values methodology and risk analysis to data monitoring of dependent variables for the industrial, ecological and other branches.

Note that the closest to the topic of our article are the sources [30–37], in which the relationship between the observed extreme values and the influence of exogenous factors, including on risk values, is investigated. Review of the literature allows to conclusions:

– The relevance of risk analysis and the need to expand it for monitoring dependent process parameters/exogenous variables not only in the financial branches.

– The POT advantage for risk methods analyses because it uses more data, allows evaluating the behavior of extreme values above some high threshold and inferences to be made directly on the distribution of exogenous extremes as a risk values. Another advantage of the POT approach is that common risk measures like VaR and Expected Shortfall (ES/CVaR) may be computed [28].

– The copula mathematical tools provides convenient instruments for analyzing pairwise dependences of exogenous variables and calculating the parameters of asymptotic dependence.

– The conditional approach of Heffernan & Tawn is flexible in terms of extremal and asymptotic dependence structures together with Ferro & Seeger's algorithm which helps to choose the threshold and the EI of variables extremes distribution.

It is also important that the statement of the problem and the method for solving it are performed from the perspective of an object-oriented analysis, which accords to modern tendencies of the process approach to support the product Quality Management System (QMS).

Risk Analysis Problem Statement

So, will be used the interpretation for multivariate sequences proposed in thesis [38]. Let $\{X_i\}_{i\geq 1}$ be the sequence of independent and identically distributed (i.i.d.) vector of random exogenous variables $X_i = (x_{i1},\ldots,x_{iD})$ with dimension $D(1 \leq d \leq D)$ and marginal distribution function *F*, where $i = 1,\ldots,n$ is the sequence time numbers of multivariate elements (exogenous variables). Many definitions are possible for the maximum of *n* consecutive elements of a multivariate sequence.

For estimating the dependence of the exogenous variables values x_{id}, will use the conditional approach proposed by Heffernan & Tawn [25, 26], when maximum/extreme of *n* consecutive elements is defined as the data-frame of componentwise (i.e. for each exogenous variables) maximums:

$$M_n = (M_{n1},\ldots,M_{nD})'$$

where $M_{nd} = \max\{x_{i1},\ldots,x_{iD}\}$ for each $1 \leq d \leq D$.

Thus, will form $M = \{M_{nd}\}$ as an $n\times D$-dimensional array of maximums (extremes) of *n* statistical *D*-dimensional observations are known for process's variables. Such data-set covers the examples of situations in which dependence at extreme levels is a consequence of proximity in space, time or dependence on a common covariate/variable [25].

Non-degenerate limits are obtained for the distribution function

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) = F^n(a_n + b_n), \quad (1)$$

where $x = (x_1,\ldots,x_D)$, $a_n = (a_{n1},\ldots,a_{nD}) > 0$ and $b_n = (b_{n1},\ldots,b_{nD})$.

All operations are performed componentwise, so $a_n > 0$ is translated as $a_{n1} > 0,\ldots,a_{nD} > 0$. The Haan and Resnick theorem [39] characterizes all of the possible limit distributions. If there exist sequences of constants $a_n > 0$ and b_n such that

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \ as\, n \to \infty, \quad (2)$$

for a non-degenerate distribution function *G*, then each of the *D* one-dimensional component distributions of *G* is a GEVD function

$$G_d(x_d) = \exp\left[-\left\{1+\xi_d\left(\frac{x_d-\mu_d}{\sigma_d}\right)\right\}^{-1/\xi_d}\right], \quad (3)$$

$$for\ 1 \leq d \leq D$$

with standard Frechet components, where ξ_d is a EI/EVI or shape parameter (shape), μ_d is a location parameter (location) and σ_d is a scale parameter (scale) of marginal exogenous variables distributions.

Note, that for high threshold values that are given by high quantiles (0.95 and higher), the distribution of probabilities for random maximums of exogenous variables values is close to the generalized Gaussian or Pareto distributions [28, 40].

The problem statement as follows: for an array of observed maximum values of variables – *M*, taking into account their pairwise dependence, develop and investigation the method for calculating VaR/CVaR as high quantiles of the respectively joint distribution functions of maximum values of variables.

Assuming, for the moment [25, 26], that the marginal distributions of variables are identical and one natural measure of their pairwise dependence is a parameters of paired conditional distributions of dependent variables.

Proposed Method

The method proposes the sequence of statistical computational procedures of processing the observations of multivariate exogenous variables extreme values.

Consider the content of the implementation stages for proposed method.

Stage 1: definition the dependent variables of the *M* extreme data array using the copula mathematical tools.

Generalized univariate limiting distribution of extreme values for describe the dependence structure of the variables can be combined with copulas of this extreme values [35].

In this case, copula (*C*) is the joint distribution function of variables random vectors *X* and *Y* after transformation to variables *U* and *V*, with uniform [0, 1] margins, via $(U,V)=\{F_X(x),F_Y(y)\}$ [25].

Here *X* and *Y* are M_{nd} for different components *d*. Then the pairs $(u_i,v_i), i=1,\ldots,n$ are independent realizations with approximate distribution *C*.

Copula *C* is used to compute function values that give an empirical measure of the type and strength of the tail dependence exhibited by the data.

In particular, was calculated the values of the functions ChiBar ($\overline{\chi}(u)$) and Chi ($\chi(u)$) [22] described by Coles, Heffernan & Tawn [25]: χ is the probability of one variable being extreme given that the other is extreme

$$\chi=\lim_{u\to 1}\Pr(V>u\,|\,U>u)' \tag{4}$$

$$\overline{\chi}=\lim_{u\to 1}\left(\frac{2\log\Pr(U>u)}{\log\Pr(U>u,V>u)}-1\right), \tag{5}$$

where $-1\le\overline{\chi}(u)\le 1$ fo all $0\le u\le 1$.

The both measures ChiBar and Chi are needed in order to obtain a summary that is informative for variables which may be either asymptotically independent or asymptotically dependent.

Stage 2: fitting of the conditional distributions functions of the most dependent variables extreme values to the observed data via the Heffernan & Tawn's method in accordance with results of the stage 1.

Note that the simulated Heffernan & Town's model data above the prediction threshold can be obtained for both point and bootstrap estimates of the dependency model parameters. Q-Q diagnostics allows comparing the simulated distributions and estimating their compliance with the Pareto distribution.

Thus the simulated values of the dependent variables are created, given that the conditioning variable is above its high quantile. This collection of values is interpreted as a tail distribution.

Stage 3: definition the parameters of the conditional distributions functions of the dependent variables extreme values: the scale parameter is sigma (σ) and the shape parameter is *xi* (ξ) (EVI/EI).

At this stage, the parameters are determined for the conditional distribution functions of pairwise dependent variables. The probability that one variable will be extreme (exceeds a certain threshold) is determined given that the other variable is also extreme or exceeds a certain threshold. Thus, a complete description of the dependent variables, the values of which exceed the established thresholds, is available for analysis.

Note that for further risk analysis the EI estimating is a great importance. In paper, EI values were estimated using the Ferro & Seeger's method for dependent extremes [23, 26]. According to these methods, EI value defines for variable, above which EI is stable for higher values of this variable [40]. It is recommended to choose such values of variables as a "high enough threshold" *u* [41].

Stage 4: the VaR and CVaR calculations as quantile-dependent values using the parameters of the conditional distributions functions of the dependent variables extreme values based on Heffernan & Town's model.

Thus, the VaR and CVaR were determined based on the parameters of the conditional Pareto distribution that is the shape parameter ξ and the scale parameter σ, for some high enough threshold *u* [28]. For the simulated tail Pareto distribution of the dependent variables values the VaR and CVaR are calculated by the formulas [28]:

$$VaR=u+\frac{\sigma}{\xi}\cdot\left(\left(\frac{n}{r}(1-q)\right)^{-\xi}-1\right) \tag{6}$$

$$CVaR=\frac{\sigma+\xi(VaR-u)}{1-\xi}, \tag{7}$$

where *r* is a quantity of exceedances over the threshold *u*, *q* is a quantile probability.

At this stage, note the advantage of POT analysis, which is the ability to draw conclusions directly about the all distribution of variables extreme values over a given threshold [28].

Stage 5: the VaR/CVaR analyzes for the dependent and independent exogenous variables.

Comment: the VaR/CVaR without taking into account the dependence of exogenous variables are calculated using the parameters of the marginal distributions of extreme variables values.

Thus, the above sequence of the method's stages allows solving the problem posed in section 2. An important goal is also to test the operation of the method created for proprietary data on well-known data.

Experimental Results

The paper uses five-dimensional air quality monitoring data comprising the measurements series of ground level ozone (O_3), nitrogen dioxide (NO_2), nitrogen oxide (NO), sulphur dioxide (SO_2) and particulate matter (PM10) in Leeds (UK) city center, during the years 1994–1998 inclusively [22, 26].

In particular, was used the daily maxima for dependent NO and NO_2 variables during winter periods 1994–1998 inclusively. Note, the gases are recorded in parts per billion and the particulate matter in micrograms per cubic meter.

Researches represent a synergy of parametric and nonparametric statistical procedures as well as uses descriptive and inferential statistic tools for data.

At the first stage, for *M* array an estimate of the variables dependence was obtained by combining the generalized extreme value distribution for the univariate marginal distributions with extreme value copulas to describe their dependence structure, as justified by the theory of multivariate extreme values [35]. So, Fig. 1 visualizes the character of dependence of five exogenous variables. For example, variables NO and NO_2 (NO/NO_2) demonstrate a clear dependence.

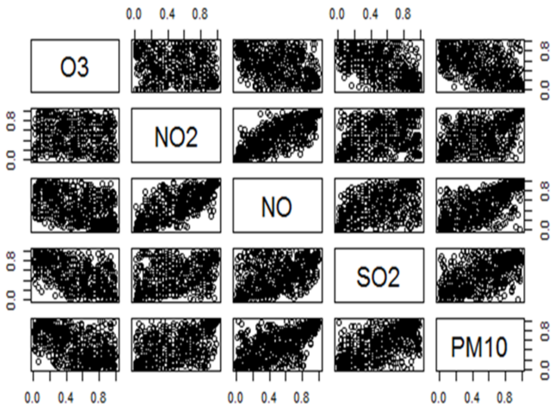


Fig. 1. Exploratory copula-diagram for proposed data

Next, the graphs of the ChiBar ( ) and Chi ( ) functions are plotted, which give an empirical measure of the type and strength of the tail dependence demonstrated by the data NO/NO_2, as shown in Fig. 2.

Parameter u is the upper limit of the general marginal distribution support and is the probability of one variable being extreme given that the other is extreme. In the case $\chi = 0$ the variables are said to be asymptotically independent [25].

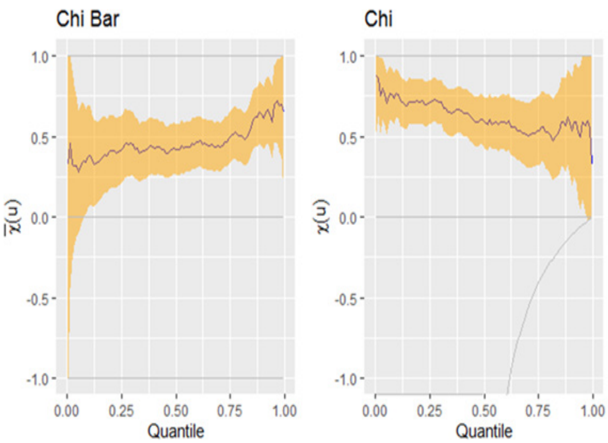


Fig. 2. Graphs the ChiBar and Chi values dependencies on 100 quantiles of the random value u distribution at the confidence interval (brown background) background

The significance level for the confidence interval is 0.05. A limiting value of ChiBar equal to 1 indicates asymptotic dependence of NO/NO_2, in which case the limiting value of Chi gives a measure of the strength of dependence in this class, as shown in Fig. 2. In the case that a limiting value of ChiBar of less than 1 is it indicates asymptotic independence in which case Chi is irrelevant and the limiting value of ChiBar gives a measure of the strength of dependence [25]. In the case of the confidence interval for ChiBar excluding the value 1 for all of the largest quantiles, the plot of the Chi function is shown in grey on Fig.2.

It is important that the ChiBar and Chi values are integral estimates of the relationship of two exogenous variables that are permutation invariant to the choice of the conditioning variable (i.e. the estimates for NO/NO2 are equivalent to those for NO_2/NO) [22, 26].

At the second stage the fitting of the conditional distributions functions of the most dependent variables extreme values to the observed data via the Heffernan & Tawn's method [26] in accordance with the results of stage 1 is produced.

Three diagnostic plots are produced for dependent variables NO/NO_2, as shown in Fig. 3 [22].

Fig. 3a shows scatterplots of the residuals Z from the fitted model of Heffernan & Tawn are plotted against the quantile of the conditioning variable (NO_2), with a lowess curve showing the local mean of these points.

Fig. 3b shows the absolute value of $|Z - mean\ (Z)|$ is also plotted again with the lowess curve showing the local mean of these points. Any trend in the location or scatter of these variables with the conditioning variable indicates a violation of the model assumption that the residuals Z are independent of the conditioning variable [28]. This can be indicative of the dependence threshold used being too low.

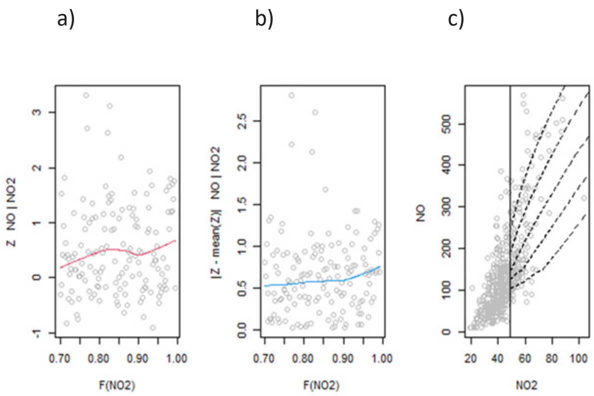


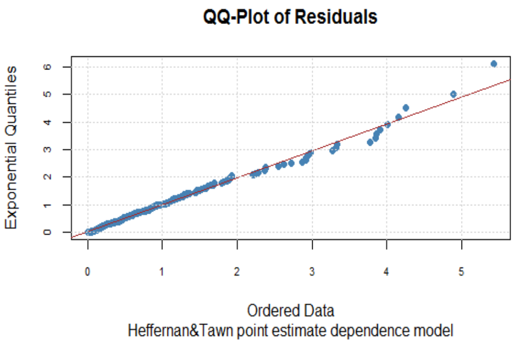
Fig. 3. Diagnostic plots for dependent variables NO/NO2

Fig. 3c shows the original data (on the original scale) and the fitted quantiles (specified by quantiles) of the conditional distribution of dependent variable given the conditioning variable. A model that fits well will have

good agreement between the distribution of the raw data (shown by the scatter plot) and the fitted quantiles. Simulated data for a collection of 5 generalized Pareto models (dashed lines) is generated above the conditioning variable threshold (vertical line).

Note that the simulated data above the prediction threshold can be obtained for both point and bootstrap estimates the parameters of the dependent variables conditional distribution model by Heffernan & Town. Despite of the Fig. 3 informativeness, this visualization is not convincing enough to assess the accordance of the simulated data to the Pareto distribution. Therefore, Q-Q diagnosis is used for simulated values are based on the point and bootstrap estimates of the dependence model parameters. Q-Q diagnostics allows you to compare the simulated distributions and assess their accordance with the Pareto distribution, as shown in Fig. 4.

a)



b)

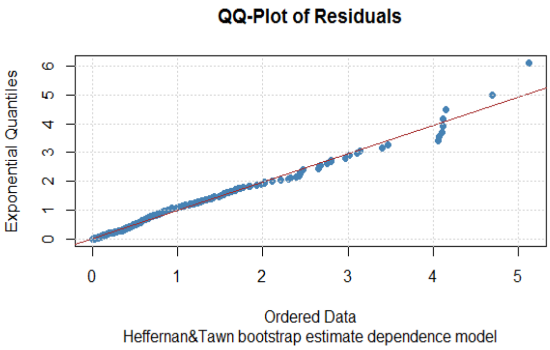


Fig. 4. Graphical Q-Q-verification the accordance of the reference distribution to the simulated distributions by Heffernan & Tawn's dependence model for point (a) and bootstrap (b) estimates of the dependence model parameters

The Q-Q-plot using the Pareto distribution as a reference distribution is graphical technique to infer the tail behavior of observed variables [28]. If the excesses over thresholds are from a thin-tailed distribution, then the GPD is exponential with $\xi = 0$ and the Q-Q-plot should be linear. Departures from linearity in the Q-Q-plot then indicate either fat-tailed behavior ($\xi > 0$) or bounded tails ($\xi < 0$). The simulated data show nearly linear convergence to the reference values for excesses over thresholds with bounded-tailed ($\xi = -0.12$).

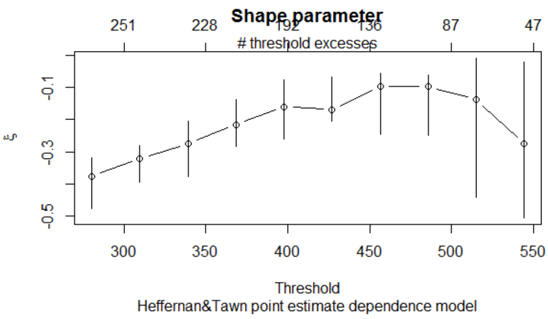
Note that the simulated values of variables for the procedure of fitting the generalized Pareto models to the original data according to the Heffernan & Towne's method based on the point estimation of the dependency model parameters demonstrate close identity with the simulated values, which additionally contain simulated repeating datasets in accordance to bootstrap-estimation of dependency model parameters.

At the third stage estimates the EI of a dependent series of observations above a given threshold. An appropriate choice of threshold is one above which the estimated EI is stable over further higher thresholds. This can be analyzed using Fig. 5, where the EI is shown for each threshold.

Graphs were plotted for the Heffernan & Tawn's model for point and bootstrap estimates of the Pareto conditional distribution parameters, as shown in Fig.5 a and b respectively. The values are equal to $\xi = -0.142$ for point model and $\xi = -0.164$ for bootstrap model. Uncertainty in the estimation of the EI and GPD parameters where assessed by using a bootstrap scheme which accounts for uncertainty in the EI estimation, and the appropriate uncertainty in the declustering of the series [22].

At the fourth stage VaR/CVaR are calculated for point and bootstrap estimates of the Heffernan & Tawn's dependence model parameters. VaR/CVaR were also calculated for independent exogenous variables. The VaR/CVaR values are given in Table 1, where prob. – the probabilities (quantiles levels) of the conditional and marginal distributions quantiles for the VaR/CVaR calculation.

a)



b)

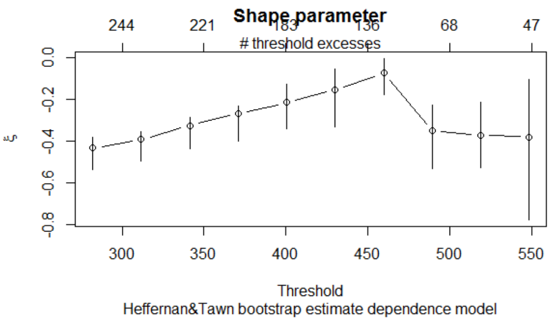


Fig. 5. The EI dependence graphs on the threshold exogenous variables values of the Heffernan & Tawn's model for the point (a) and bootstrap (b) estimates of the Pareto conditional distribution parameters

Tab. 1. VaR/CVaR for point and bootstrap estimates of the Heffernan & Tawn dependence model parameters and for independent exogenous variable

№	prob.	point model		bootstrap model		independent exogene NO	
		VaR	CVaR	VaR	CVaR	VaR	CVaR
1	0.9900	691.63	768.06	677.94	750.37	494.17	531.11
2	0.9950	746.83	819.84	730.07	799.55	526.04	553.13
3	0.9990	865.67	931.32	843.06	906.13	571.03	584.22
4	0.9995	913.09	975.80	888.45	948.94	582.41	592.08
5	0.9999	1015.16	1071.55	986.83	1041.74	598.47	603.18

Tab. 2. Values for 99%-th quantile level of VaR/CVaR estimates and 95% CIs

VAR			CVaR		
Lower CI	Estimate	Upper CI	Lower CI	Estimate	Upper CI
647.49	677.94	725.12	703.36	750.37	842.59

From Table 1 it can be seen that calculated VaR/CVaR based on the point estimate of the Heffernan & Tawn's dependence model parameters are fairly close to the calculated values based on the bootstrap estimate of the Heffernan & Tawn's dependence model parameters. The calculated VaR/CVaR for independent variable distribution is smaller than the risks values based on the estimates of the Heffernan & Tawn's dependence model parameters for all quantiles levels.

At the fifth stage makes the analysis of VaR/CVaR values for dependent and independent exogenous variables. This stage uses the advantage of POT analysis which is the ability to draw conclusions about the entire available distribution of extreme variables values over a given threshold. Estimated VaR/CVaR values and confidence intervals (CIs) are shown in Fig. 6.

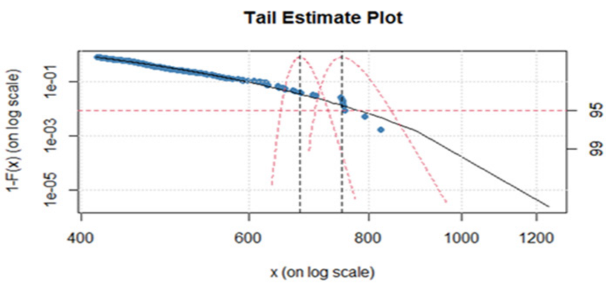


Fig. 6. Tail plot for exogenous variables conditional distribution over the threshold for the bootstrap estimate of the Heffernan & Tawn's dependence model parameters and CIs (red dotted line) with the estimates of VaR/CVaR high quantiles (vertical dotted line)

For further analysis a graphical representation for exogenous variables conditional distribution over the threshold (solid black curve) on Fig. 6 is added by the CIs for estimates of VaR/CVaR high quantiles for the variant of bootstrap estimate of the Heffernan & Tawn's dependence model parameters. The CIs values and estimates for VaR/CVaR in this case are characterized by the data in Table 2.

The sensitivity of the VaR estimates to changes in the threshold u to be investigated using the functions of R package "fExtremes" is shown in Fig. 7.

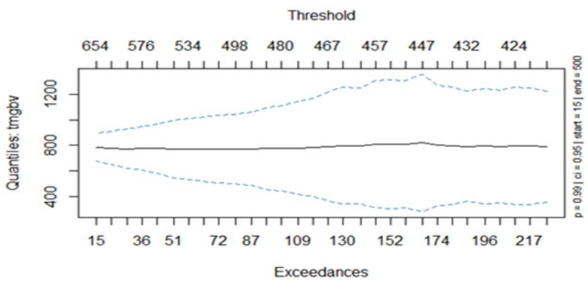


Fig. 7. Estimate plot for the varies of a high quantile in the tail of the simulated dataset of variables based on the Heffernan & Tawn's conditional model in depend on threshold or number of extremes (95%-th CI shows with dashed lines)

So, for analysis together with Fig. 6 is convenient to use Fig. 7 showing how the estimate of a high quantile for the tail of a conditional data based on the Heffernan & Tawn's dependence model varies with threshold or number of extremes. The VaR estimates for 99%-th quantile level are stable for thresholds more than 410.

It can be argued that EVT is a useful supplementary risk measure because it provides more appropriate distributions to fit extreme events [40]. This makes it possible to reduce the uncertainty of VaR/CVaR estimates, which can be seen in Fig. 7 when the VaR estimates for 99% quantile level remains stable over a wide range of thresholds.

5. Conclusion

In risk analysis, a global fit that appropriately captures the body and the tail of the distribution of exogenous variables is essential. At statistical estimates the extremes data of the exogenous variables the problem of the dependence of these variables as the monitoring result is inevitably arises.

The accounting of the joint exogenous variables distribution is a relevant clarification of the risk value as it provides more adequate estimates of the variables distribution fitted to the extreme events. Therefore the problem of risk-analysis by the extreme data of dependent exogenous variables as the main goal of the article is solved using the conditional Heffernan & Tawn's approach.

The result of this solution is the sequence and content of the actions (stages) that make up the proposed method. Studies have shown that the main trends in VaR/CVaR change for proprietary data of dependent exogenous variables are generally confirmed for well-known data.

When solving the problem, the following were used: the POT approach for analysis of extreme values, the flexibility of the Heffernan & Tawn's method to the structure of asymptotic dependencies, the advantages of the Ferro & Segers's algorithm for choosing a threshold and an EI for the conditional exogenous variables distribution.

Thus, the method expands the application of the risk-analysis methodology for monitoring data of dependent exogenous variables not only for the financial area, but also for industrial, environmental and other areas.

It is also important that the proposed method opens up the prospect of research on the influence of VaR/CVaR estimates on the accuracy of interval estimates of observation parameters.

Risk-adjusted dependent extreme value variables research targeting is optimizes the branches performance in terms of functional load on process elements and benefits stakeholders.

ACKNOWLEDGEMENTS

This paper is submitted in a competition framework of scientific and technological developments "Science for the safety of man and society" in accordance with the Resolution of the scientific council of the National Research Foundation of Ukraine (protocol №. 7 of May 11, 2020). These studies have been extended for the topic "Study of psychophysiological functions and intelligence of students in the conditions of life in a state of epidemic limitation".

AUTHORS

Ihor Tereshchenko – Infocommunication Engineering Department, Kharkiv National University of Radio Electronics, Kharkiv, 61166, Ukraine,
e-mail: ihor.tereshchenko@nure.ua.

Anton Tereshchenko – Management of Information and Cyber Security Department, State University of Telecommunications, Kyiv, 03680, Ukraine,
e-mail: arisen001@gmail.com.

Nataliya Bilous* – Software Engineering Department, Kharkiv National University of Radio Electronics, Kharkiv, 61166, Ukraine,
e-mail: nataliya.bilous@nure.ua.

Svitlana Shtangey – Infocommunication Engineering Department, Kharkiv National University of Radio Electronics, Kharkiv, 61166, Ukraine,
e-mail: svitlana.shtanhei@nure.ua.

Zygmunt L. Warsza* – Łukasiewicz Research Network – Industrial Research Institute for Automation and Measurements PIAP,
e-mail: zlw1936@gmail.com.

∗ Corresponding authors

REFERENCES

[1] C. Lleras, "Path Analysis". In: *Encyclopedia of Social Measurement*, Elsevier, 2005, 25–30, 10.1016/B0-12-369398-5/00483-7.

[2] "ISO 9000, Introduction and Support Package: Guidance on the Concept and Use of the Process Approach for management systems". ISO/TC 176/SC 2/N 544R3, https://www.iso.org/files/live/sites/isoorg/files/archive/pdf/en/04_concept_and_use_of_the_process_approach_for_management_systems.pdf. Accessed on: 2022-04-13.

[3] "ISO 9001:2015 Quality management systems". https://www.iso.org/standard/62085.html. Accessed on: 2022-04-13.

[4] "The new ISO 31000 keeps risk management simple". https://www.iso.org/cms/render/live/en/sites/isoorg/contents/news/2018/02/Ref2263.html. Accessed on: 2022-04-13.

[5] A. Stephenson and E. Gilleland, "Software for the analysis of extreme events: The current state and future directions", *Extremes*, vol. 8, 2005, 87–109, 10.1007/s10687-006-7962-0.

[6] E. Gilleland, M. Ribatet and A. G. Stephenson, "A software review for extreme value analysis", *Extremes*, vol. 16, no. 1, 2013, 103–119, 10.1007/s10687-012-0155-0.

[7] M. I. Gomes and A. Guillou, "Extreme Value Theory and Statistics of Univariate Extremes: A Review", *International Statistical Review*, vol. 83, no. 2, 2015, 263–292, 10.1111/insr.12058.

[8] P. Abad, S. Benito and C. López, "A comprehensive review of Value at Risk methodologies", *The Spanish Review of Financial Economics*, vol. 12, no. 1, 2014, 15–32, 10.1016/j.srfe.2013.06.001.

[9] D. K. Dey and J. Yan, *Extreme Value Modeling and Risk Analysis: Methods and Applications*, Chapman and Hall/CRC, 2016, 10.1201/b19721.

[10] N. G. Zrazhevska and A. G. Zrazhevsky, "Classification of methods for risk measures VaR and

CVaR calculation and estimation", *System research and information technologies*, no. 3, 2016, 126–141, 10.20535/SRIT.2308-8893.2016.3.11.

[11] B. G. Peterson et al., "PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis", https://CRAN.R-project.org/package=PerformanceAnalytics. Accessed on: 2022-04-13.

[12] J. Mina and J. Y. Xiao, "Return to RiskMetrics: The Evolution of a Standard". https://www.msci.com/documents/10199/dbb975aa-5dc2-4441-aa2d-ae34ab5f0945. Accessed on: 2022-04-13.

[13] B. Cordeiro and A. Kotoky, "MSCI to buy RiskMetrics for $1.55 billion | Reuters". https://www.reuters.com/article/us-riskmetrics-msci-idUSTRE62041J20100301. Accessed on: 2022-04-13.

[14] L. Rickenberg ,"Tail Risk Targeting: Target VaR and CVaR Strategies". https://papers.ssrn.com/abstract=3444999, DOI: 10.2139/ssrn.3444999. Accessed on: 2022-04-13.

[15] D. Wuertz, T. Setz and Y. Chalabi, "fExtremes: Rmetrics - Modelling Extreme Events in Finance". https://CRAN.R-project.org/package=fExtremes. Accessed on: 2022-04-13.

[16] B. Pfaff et al., "evir: Extreme Values in R". https://CRAN.R-project.org/package=evir. Accessed on: 2022-04-13.

[17] A. J. McNeil and R. Frey, "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach", *Journal of Empirical Finance*, vol. 7, no. 3-4, 2000, 271–300, 10.1016/S0927-5398(00)00012-8.

[18] "ismev: An Introduction to Statistical Modeling of Extreme Values". https://CRAN.R-project.org/package=ismev. Accessed on: 2022-04-13.

[19] S. Coles, *An Introduction to Statistical Modeling of Extreme Values*, Springer London, 2001, 10.1007/978-1-4471-3675-0.

[20] T. Reynkens et al., "ReIns: Functions from "Reinsurance: Actuarial and Statistical Aspects"." https://CRAN.R-project.org/package=ReIns. Accessed on: 2022-04-13.

[21] T. Reynkens, R. Verbelen, J. Beirlant and K. Antonio, "Modelling censored losses using splicing: A global fit strategy with mixed Erlang and extreme value distributions", *Insurance Mathematics and Economics*, vol. 77, 2017, 65–77, 10.1016/j.insmatheco.2017.08.005.

[22] H. Southworth et al., "texmex: Statistical Modelling of Extreme Values". https://CRAN.R-project.org/package=texmex. Accessed on: 2022-04-13.

[23] C. A. T. Ferro and J. Segers, "Inference for clusters of extreme values", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 65, no. 2, 2003, 545–556, 10.1111/1467-9868.00401.

[24] M. Ferreira, "Analysis of estimation methods for the extremal index", *Electronic Journal of Applied Statistical Analysis*, vol. 11, no. 1, 2018, 296–306. 10.1285/i20705948v11n1p296.

[25] S. Coles, J. Heffernan and J. Tawn, "Dependence Measures for Extreme Value Analyses", *Extremes*, vol. 2, no. 4, 1999, 339–365, 10.1023/A:1009963131610.

[26] J. E. Heffernan and J. A. Tawn, "A conditional approach for multivariate extreme values (with discussion)", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 66, no. 3, 2004, 497–546, 10.1111/j.1467--9868.2004.02050.x.

[27] T. Lugrin, "tsxtreme: Bayesian Modelling of Extremal Dependence in Time Series". https://CRAN.R-project.org/package=tsxtreme. Accessed on: 2022-04-13.

[28] E. Zivot and J. Wang, *Modeling Financial Time Series with S-PLUS®*, Springer New York, 2006, 10.1007/978-0-387-32348-0.

[29] N. Bezak, M. Brilly and M. Šraj, "Comparison between the peaks-over-threshold method and the annual maximum method for flood frequency analysis", *Hydrological Sciences Journal*, vol. 59, no. 5, 2014, 959–977, 10.1080/02626667.2013.831174.

[30] S. N. Majumdar, A. Pal and G. Schehr, "Extreme value statistics of correlated random variables: A pedagogical review", *Physics Reports*, vol. 840, 2020, 1–32, 10.1016/j.physrep.2019.10.005.

[31] R. S. Tsay, "Testing serial correlations in high-dimensional time series via extreme value theory", *Journal of Econometrics*, vol. 216, no. 1, 2020, 106–117, 10.1016/j.jeconom.2020.01.008.

[32] E. C. Brechmann and U. Schepsmeier, "Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine", *Journal of Statistical Software*, vol. 52, no. 3, 2013, 10.18637/jss.v052.i03.

[33] L. Deng, C. Ma and W. Yang, "Portfolio Optimization via Pair Copula-GARCH-EVT-CVaR Model", *Systems Engineering Procedia*, vol. 2, 2011, 171–181, 10.1016/j.sepro.2011.10.020.

[34] G. Meissner, *Correlation Risk Modeling and Management: an applied guide including the Basel III correlation framework – with Interactive Cor-*

relation Models in Excel/VBA, John Wiley & Sons, 2014, 10.1002/9781118809204.

[35] J. Carreau and G. Toulemonde, "Extra-parametrized extreme value copula: Extension to a spatial framework", *Spatial Statistics*, vol. 40, 2020, 10.1016/j.spasta.2020.100410.

[36] R. B. Nelsen, *An Introduction to Copulas*, Springer New York, 2006, 10.1007/0-387-28678-0.

[37] N. Quinn, P. D. Bates, J. Neal, A. Smith, O. Wing, C. Sampson, J. Smith and J. Heffernan, "The Spatial Dependence of Flood Hazard and Risk in the United States", *Water Resources Research*, vol. 55, no. 3, 2019, 1890–1911, 10.1029/2018WR024205.

[38] C. A. T. Ferro, "Statistical Methods for Clusters of Extreme Values", PhD Thesis, Lancaster University, September 2003, http://empslocal.ex.ac.uk/people/staff/ferro/Publications/Thesis/thesis.pdf. Accessed on: 2022-04-13.

[39] L. de Haan and S. I. Resnick, "Limit theory for multivariate sample extremes", *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 40, no. 4, 1977, 317–337, 10.1007/BF00533086.

[40] Y. Bensalah, "Steps in Applying Extreme Value Theory to Finance: A Review". https://www.banqueducanada.ca/wp-content/uploads/2010/01/wp00-20.pdf. Accessed on: 2022-04-13.

[41] I. V. Tereshchenko, A. I. Tereshchenko and S. V. Shtangey, "Risks Estimation Method by Clustered Extreme Data of Process Covariates", *Radio Electronics, Computer Science, Control*, no. 2, 2020, 51–64, 10.15588/1607-3274-2020-2-6.