ERASMUS UNIVSITTY ROTTERDAM

Erasmus School of Economics

Bachelor thesis Econometrie

High dimensional Risk Management with Multivariate Generalized Pareto

Author:
Onno Lijmbach *(482567)*

Supervisor:
Jochem Oorschot

Second assessor:
Rutger-Jan Lange

Abstract

Using Extreme Value theory to describe risk metrics such as the Value at Risk and Expected Shortfall is not novel. One way is to make use of Multivariate Generalized Pareto distribution to describe these risk metrics, however one pitfall is that it can only handle moderate dimensions. We propose two methods which are able to handle larger dimensions. The first method uses clustering to reduce the amount of dimensions. The second method creates sparsity be implementing a tree structure on the assets in the portfolio. We compare these 2 methods with the univariate generalized Pareto distribution, where we use the data from 16 companies in the AEX. We see that the 2 models can accurately describe the Value at Risk.

5-7-2020

Contents

1 Introduction

Assessing the risks of portfolios and indices is an pinnacle task for investors. They need to know how much risk they take by investing, therefore they use certain risk metrics like the Value at Risk or the Expected Shortfall. These metrics tell us how much money they lose given a specific probability. Extreme value theory can address this problem by researching how the most extreme events behave.

A lot of research has been done by studying the extreme returns of financial instruments in an univariate manner, however portfolios and indices consist of multiple correlated assets. In the univariate setting this correlation is not directly taken into account. If we are able to also model the correlation between these assets, we can estimate the risks in a more efficient way.

A way to estimate the Value at Risk and Expected Shortfall using extreme value theory is by means of fitting the Generalized Pareto distribution. This can both be done in the univariate case as well as in a multivariate manner. In this research we focus on the multivariate case, which is introduced by Rootzén, Tajvidi, et al. (2006) and applied to finance by Kiriliouk et al. (2019), however these methods are only computational feasible for moderate dimensions.

Therefore we apply two methods to make the Generalized Pareto distribution feasible for higher dimensions. The first method uses dimension reduction by means of K-means clustering to convert the high number of assets to a lower number of clusters, after that we apply the methods from Kiriliouk et al. (2019). The second method uses the graph structure introduced by Engelke and Hitz (2018), which creates sparsity in the model. It uses the notion of conditional independence, which tells us that two assets can be independent if we know the return on another asset.

We use the data of 16 companies in the AEX at 31 March 2020, of which we make an adjusted-AEX index. Of this index we want to estimate the Value at Risk and Expected Shortfall. We do not incorporate all the assets inside the AEX, due to the absence of data in the most volatile periods. We include all the assets which had their initial public offering before 1 January 2008. We select this date such that we can include the financial crisis of 2008.

We compare these two methods with the usual univariate Generalized Pareto fit and we find that all three methods correctly describe the Value at Risk, where the Tree model has the lowest p-values. We see that the Clustering and the Tree model in the case of 99 % coverage have significant excesses bigger than the VaR.

We find that both the Clustering method, as well as the method which creates a tree structure can predict the Value at Risk in a more efficient way than if it is done univariatly. We contribute to the current literature by applying two methods which can deal with the high dimensions that occur in extreme value theory for the calculation of risk metrics, whereas before this could only be achieved for moderate dimensions.

In the next section we give an extensive literature review, regarding the methods that have been proposed in high dimensional extreme value theory. Thereafter we analyse the data of the adjusted-AEX. Then we give the information regarding the methods of both Kiriliouk et al. (2019) and Engelke and Hitz (2018). Next we discuss the results of both methods and we conclude with a summary of our research with suggestions for further research.

2 Literature Review

There are 2 main techniques to determine the distribution of an extreme event. The first method is called block maxima, where we take the maximum of every time period and then fit a distribution over all these maxima (Gumbel 1958). The second method is the peaks-over-threshold method, where we fit a distribution

to all data which exceeds a high threshold (Smith 1987). This distribution converges for high thresholds to a Generalized Pareto Distribution (Balkema and De Haan (1974), Pickands III et al. (1975), Smith (1984) and Davison and Smith (1990)).

The peaks over threshold model has seen use in a wide variety of branches. It has been used to model extreme floods (Katz, Parlange, and Naveau 2002), model waves and water levels (Hawkes et al. 2002) and in finance to model the risk of assets (McNeil, Frey, and Embrechts 2015). But also sectors that would not come immediately to mind, such as wind engineering (Ragan and Manuel 2008), strength of materials (Anderson, De Maré, and Rootzén 2005) and traffic safety (Gordon et al. 2013).

However the impact of a flooding is determined by the flooding of many dikes, not only one. In this example the damage is caused by multiple components, which are not independent. In financial risk management a similar argument comes to mind; a portfolio consists of different correlated financial instruments, therefore there is a need to model the returns of these assets together.

One option is to use the multivariate Generalized Pareto Distribution, which is introduced by Rootzén, Tajvidi, et al. (2006). The main difference between the univariate Generalized Pareto distribution and the multivariate one is that the dependence of the extremes of different variables is explicitly modelled. Further work on the multivariate Generalized Pareto distribution is done by Falk and Guillou (2008), Ferreira, De Haan, et al. (2014), Rootzén, Segers, and Wadsworth (2018a) and Rootzén, Segers, and Wadsworth (2018b). Although the growing probabilistic literature of the multivariate Generalized Pareto distribution, there has not been much use as a statistical model. Furthermore when the model was used (Davison and Huser (2015), Fondeville and Davison (2018)), it was only within a single family of a GP distribution. Kiriliouk et al. (2019) applies the method to more than a single family of GP distributions.

However a downside of this method is that it only works in moderate dimensions due to the absence of sparsity, which is a major drawback in modelling portfolios or indices. Therefore we apply the methods as described in Engelke and Hitz (2018), which require a shape parameter bigger than 0. Gilli et al. (2006) showed that the returns of assets usually have a shape parameter bigger than 0, therefore this method is viable. The method relies on the notion of conditional independence (Dawid 1979). It tells us that two variables are not independent by themself, however when these two variables are conditioned on other variables they are. If this is put together with a graphical structure, it creates a simple probabilistic structure and sparsity (Lauritzen (1996) and Wainwright, Jordan, et al. (2008)).

The Hammersley-Clifford theorem (Hammersley and Clifford 1971) tells us that under the assumption of conditional independence on decomposable graphs, there is a equivalence with a factorization of the high dimensional density. In Engelke and Hitz (2018) they prove an extremal variant of the Hammersley-Clifford property. This results in an increase in sparsity, which tackles the problem that occurs in Kiriliouk et al. (2019).

As said in the previous paragraph, to apply the Hammersley-Clifford theorem we need to have a decomposable graph. In many cases, like in finance, the underlying graph is not known and therefore needs to be estimated. One option is to apply the algorithm of Kruskal (1956), which creates a minimum spanning tree of all the variables. This algorithms chooses the tree between all possible assets, where the sum of all the weighs is minimized. As weights a bivariate likelihood score can be chosen, in this way it creates a tree with a high likelihood.

A distribution that is widely used in extreme value theory is the Hüsler-Reiss distribution (Hüsler and Reiss 1989), it takes the place of the normal distribution in the extreme world. It is parameterised with the variogram Γ. Engelke and Hitz (2018) shows that the underlying graph can be represented with zero patterns

in the variogram Γ.

Creating a graph structure is not novel in finance, Bessler and Yang (2003) used it to research the dependence structure of several financial markets. Further work on this field has been done by Abdelwahab, Amor, and Abdelwahed (2008) and Talih and Hengartner (2005). However to the best of our knowledge, nobody has applied it to extremes in finance.

The idea of creating parsimonious models to describe extreme values is not novel. Spatial max-stable random fields have been widely studied by Schlather (2002), Kabluchko, Schlather, De Haan, et al. (2009) and Opitz (2013). However these methods require a underlying knowledge of the underlying structure and therefore can not be applied to a financial setting. Closely related to the work of Engelke and Hitz (2018) is the work of Coles and Tawn (1991) and Smith, Tawn, and Coles (1997). They propose a Markov chain, where all the bivariate marginals are extreme value distributions.

The other option of modelling extremes, is the method of block maxima. However there are multiple reasons why the peaks over threshold model is more suitable in our case. There has not been much research done on a multivariate variant of the block maxima method, Bücher and Segers (2014) uses a method where they take a block maxima of each component and then fit a extreme dependence structure on it. Furthermore, Madsen, Rasmussen, and Rosbjerg (1997) showed that if the shape parameter is bigger than 0, POT is more efficient than block maxima.

3 Data

We consider the negative returns as the main variable of our research:

$$R_{i,t} = 1 - \frac{S_{i,t}}{S_{i,t-1}} \tag{1}$$

Where $R_{i,t}$ is the negative return of stock i at week t and $S_{i,t}$ is the value of stock i at time t.

The companies we consider are noted in the Amsterdam Exchange Index (AEX) in 2020. Due to different dates of the initial public offering, some stocks are longer prevalent in the market. However the more data we have the better the parameter estimates are, especially for the periods with extreme negative returns. Therefore we only include stocks which had their initial public offering before the 1'st of January 2008. We choose this day because it includes volatile moments such as the financial crisis of 2008 and the corona crisis in 2020. In Table 1 we can see all the companies which where present in the AEX at 31 of March in 2020, with their respective share in the AEX in the index. The initial public offering date (IPO DATE) is also included. 16 Companies (*) are included in the research. We use the data returns of these companies from the timespan of 25 July of 2005 untill 4 May of 2020

Company	Share (%)	Adjusted Share	IPO DATE
ABN AMRO	0.55	0.00	20-11-2015
ADYEN	3.84	0.00	13-6-2018
AEGON*	0.76	0.93	27-3-1995
Ahold Delhaize	4.57	0.00	20-10-2008
Akzo Nobel*	2.33	2.85	27-3-1995
Arcelor Mittel*	1.12	1.37	25-7-2005
ASM*	0.87	1.07	20-7-1998
ASML*	16.37	20.05	4-1-1999
ASR	0.6	0.00	10-6-2014
DSM*	3.66	4.48	3-1-2000
Galapagos	1.48	0.00	6-5-2005
Heineken*	3.42	4.19	27-3-1995
IMCD	0.66	0.00	27-6-2014
ING*	3.63	4.45	27-3-1995
Just Eat Takeaway	1.72	0.00	30-9-2016
KPN*	1.51	1.85	27-3-1995
NN	1.4	0.00	2-7-2014
Phillips*	6.45	7.90	8-5-1995
Prosus	5	0.00	11-9-2019
Randstad*	0.69	0.84	8-5-1995
Relx*	7.85	9.61	27-3-1995
Royal dutch Shell *	14.6	17.88	18-7-2005
Unibail-Rodamco*	1.39	1.70	31-12-1990
Unilever*	12.12	14.84	20-7-1998
Wolters Kluwe*	3.41	4.18	27-3-1995

Table 1: Companies of the AEX

With these 16 companies we cover 80.18 % of the AEX. In Table 1 the adjusted weight is also stated, where the weight is 0 if this company is not included in the adjusted-AEX.

To justify the use of the Pareto models from Engelke and Hitz (2018), we need to substantiate that the shape parameter is bigger than 0. This implies that the distribution of returns has no end point. In Figure 1 we see the histogram of all the returns with a fitted normal distribution over it. The data has a kurtosis of 8.87, while a normal distribution has a kurosis of 3. This implies that in the return distribution there is more mass in the tails than there would be with a Gaussian. Hereby we conclude that the assumption of a shape parameter being bigger than 0 is likely.

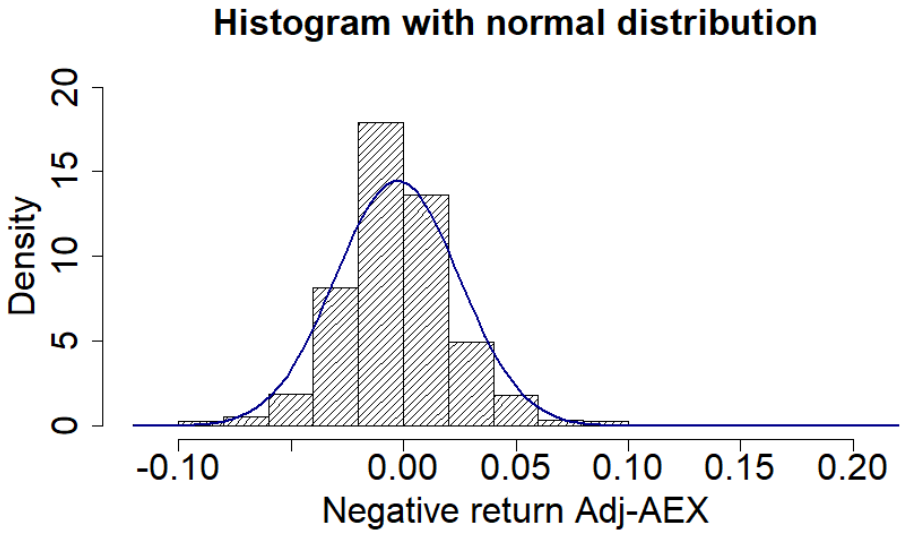


Figure 1: Histogram

We can see the returns of the adjusted-AEX in Figure 2. We can see that there are 2 periods where there are extreme returns: the financial crisis (2008) and the corona crisis (2020). The stock price of the adjusted-AEX can be found in Figure 3, where we also see the effect of these crises.

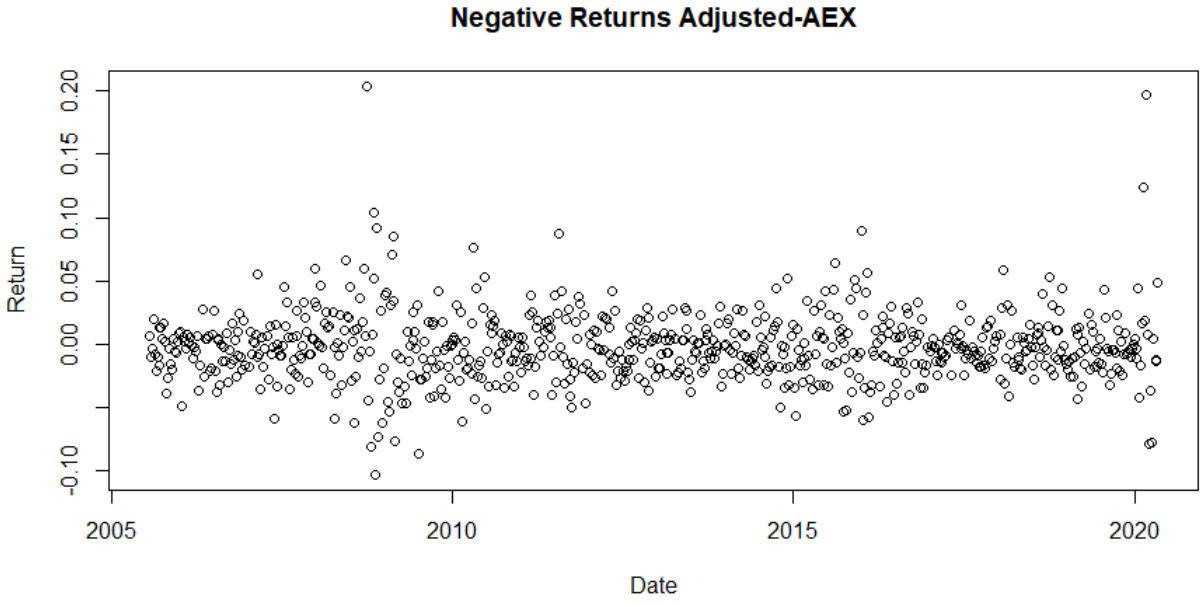


Figure 2: Negative Returns Adjusted-AEX

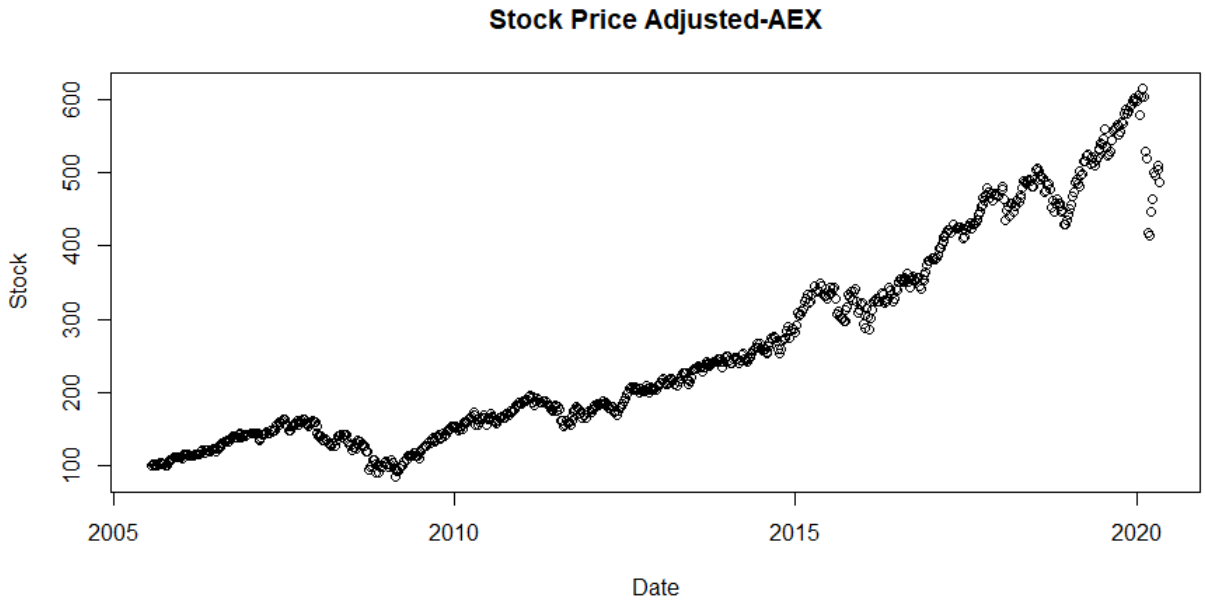


Figure 3: Stock price Adjusted-AEX

We train all the methods with the data from 25 July of 2005 untill 31 December 2017 (T=649) and we use the data from 1 January 2018 untill 4 May 2020 (T=123) to test the methods out of sample.

4 Methodology

The methodology is divided into four different parts. We start with a general discription of univariate Generalized Pareto, which is used as a benchmark. The first method is the clustering method where we first cluster the assets into 3 groups, thereafter we apply the methodology of Kiriliouk et al. (2019). The second method imposes a tree structure on the available assets, thereby creating sparsity (Engelke and Hitz 2018). Finally we discuss how the parameter estimates lead to the Value at Risk and Expected Shortfall.

4.1 Pareto distribution

It is shown that the distribution of $Z - u > x|Z > u$ follows a Generalized Pareto distribution for a large threshold u. This tells us that:

$$P(Z - u > x|Z > u) = (1 + \frac{\gamma x}{\sigma})^{-\frac{1}{\gamma}} \tag{2}$$

Where γ is the shape parameter and σ the scale parameter of the Generalized Pareto distribution. This is the generalized version of the Pareto distribution, for the methodology regarding Engelke and Hitz (2018) they assume that the shape parameter γ is strictly positive.

We fit the returns of the adjusted-AEX with an univariate Pareto distribution as a benchmark for the other two methods.

4.2 Cluster method

We follow the methodology regarding Kiriliouk et al. (2019). We see weeks which have one asset with the quantile of a return higher than q as an extreme week. Ideally we want to directly implement this methodology, however their methods are only computationally capable of handling a moderate number of assets at the same time. Therefore we cluster the 16 assets in three different groups.

4.2.1 Clustering

Ideally we want to directly implement this methodology, however the methods of Kiriliouk et al. (2019) are only capable of handling moderate dimensions. Therefore we group the 16 assets in three different clusters. The clusters are made using the K-means algorithm (MacQueen et al. 1967).

K-means is an unsupervised learning technique that puts together observations, which have approximately the same variables. It minimizes the following criterion:

$$\arg\min_{S} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2 \tag{3}$$

Where k are the number of clusters, x are the observations, S=$\{S_1, S_2, ... , S_k\}$ is the set of clusters with S_i the set of observations in cluster i and μ_i is the average of cluster i.

We want to cluster the assets which have a high dependency in the tail, therefore we construct the following variable:

$$X_{i,t} = \begin{cases} Q(X_{i,t}), \text{ if } \quad Q(X_{i,t}) > 0.8 \\ 0, \text{ if } \quad Q(X_{i,t}) \leq 0.8 \end{cases}$$

where $Q(X_{i,t})$ is the quantile of the negative return of asset i at time t. By censoring from 0.8 we emphasize the correlations in the tale of the return distribution.

In total we have the returns of 649 different weeks (dimensions) for 16 different assets (observations), where we group the assets in terms of the variable X_i as defined above.

After we performed K-means, we have three groups of assets. However we need to weigh the assets so that they represent the weights as in the adjusted-AEX. We weigh the companies by their AEX share, such that each cluster has a share and each company has it's own share within each cluster. Thereafter we calculate for each cluster what the return would have been.

4.2.2 Threshold selection

In order to apply the Multivariate Generalized Pareto distribution we need to choose a threshold u. The threshold u needs to be chosen such that $Y - u|Y \not\leq u$ approximates a GP distribution. We exploit the stability property of multivariate GP distributions and use the χ measure of asymptotic dependence:

$$\chi(q) := \frac{P(F(R_1) > q, ..., F(R_d) > q)}{1 - q} \tag{4}$$

To guide us in choosing the appropriate threshold q, we make use of the following property: GP distributions are threshold stable, this means that we can change the thresholds of the distribution, it still remains a GP distribution with new parameters. If we choose a new threshold w, the distribution of $X - W|X \not\leq w$ is a GP distribution with parameters $\sigma + \gamma w$ and γ. A special role is played by the level $w = \sigma(t^\gamma - 1)/\gamma$.

These levels have a special property: For any set $A \subset \{x \in R^d : x \nleq 0\}$ it holds that, for $t \geq 1$ that.

$$P(X \in w_t + t^\gamma A) = \frac{P(x \in A)}{t} \tag{5}$$

This implies that when we reach a high threshold q the χ measure should be equal. We estimate the χ function empirically by:

$$\widehat{\chi(q)} = \frac{\sum_{i=1}^{n} (\widehat{F(R_1)} > q, ..., \widehat{F(R_d)} > q)}{n(1-q)} \tag{6}$$

where $\widehat{F(R_i)}$ represents the empirical cumulative distribution of the returns R_i. We use the empirical distribution $\widehat{F(R_i)}$ as an estimator for the real distribution $F(R_i)$.

4.2.3 Density

We use two different constructors to make the Generalized Pareto densities. These constructors require a transformation of the data, such that the parameters of the Generalized Pareto distribution of the transformed data are $\sigma = 1$ and $\gamma = 0$. The transformation is as follows:

$$X = \sigma \frac{e^{\gamma X_0} - 1}{\gamma} \tag{7}$$

where X_0 is the standardized data. This describes the relation between the real data X and the standardized data X_0, however we want to create X_0 without explicitly knowing the shape and scale parameters. Hence we construct X_0 in the following manner:

$$X_0 = E + T + max(T) \tag{8}$$

where E is a unit exponential, T are the observed exceedances, and max(T) equals the maximum exceedances of all the asssets. Now that we have the standardized density, we use 2 constructions method to create the density for the Generalized Multivariate Pareto distribution (Rootzén, Segers, and Wadsworth 2018b). The density of the first construction method T is as follow:

$$h_T(x, 1, 0) = \frac{\mathbb{1}(max(x) > 0)}{e^{max(x)}} \int_0^\infty f_T(x + log(t)) t^{-1} dt \tag{9}$$

Where h_t is the density of X_0 and f_T can be any density, thus providing a flexible aspect to the model. The density of the second construction method U is as follows:

$$h_U(x, 1, 0) = \frac{\mathbb{1}(max(x) > 0)}{E(e^{max(U)})} \int_0^\infty f_u(x + log(t)) dt \tag{10}$$

These 2 methods required the standardized data, if we revert the data back to the normal scale we obtain the following density, which can be optimised :

$$h(x; \sigma; \gamma) = h(\frac{1}{\gamma} log(1 + \gamma x/\sigma); 1; 0) \prod_{i=1}^{d} \frac{1}{\sigma_j + \gamma_j x_j} \tag{11}$$

Note that $h(*)$ can be different densities, therefore we consider the following densities: Gumbel, Reverse Exponential and Multivariate Normal. Next to the dependence parameters, all the distributions also

have their own parameters which need to be estimated. The data is then directly applied to the threshold exceedances $(R - u | R \nleq u)$.

4.2.4 Censoring

We only use the density in equation 11 if one of the returns are above their respective thresholds u. We apply censoring to avoid the weight given to non extreme data, furthermore the bias of the dependence parameters can be larger without censoring (Huser, Davison, and Genton 2016). Therefore we use the following likelihood contribution:

$$h^C(y - u; \sigma; \gamma) = \int_{j \in C \times (-\infty, u_j)} h(y - u; \sigma; \gamma) d(y_c); \tag{12}$$

Where C is the set of indices that falls below the threshold u. This results that for the censored indices that the integral will be taken, whereas for the other components the regular density will suffice. This will then result in the following likelihood, which needs to be optimised by means of maximum likelihood:

$$L(\theta, \sigma, \gamma) = \prod_{i=1}^{n} h^C(y_{D/C} - u_{D/C}; 0_C; \theta, \sigma, \gamma) \tag{13}$$

Where D is the set of all indices and C is the set which is censored. 0_C indicates that the set which is censored is censored to 0.

4.2.5 Model selection

To select the model we want to use, we need to consider a few steps to come to our final model. There can be a lot simplifications done in both the distribution as well as in the dependence parameters. Furthermore there are 3 distributions which we can fit the model to and 2 different constructors. Therefore we do the following steps to come to the final model:

- Transform the returns R for every asset i to its respective quantile function U_i and transform it to Pareto scale, e.g $R_i^{par} = \frac{1}{1-U_i}$. Now every asset has the same scale.
- Select a threshold v on the Pareto scale as described in section 4.2.2.
- Fit every combination of distribution and constructor on the Pareto transformed data with the maximum number of parameters to to the standardized Pareto data $R^{par} - v | max(R^{par}) > v$. We use the condition, because we are only interested in the weeks with one or more high quantiles.
- To select the combination of distribution and constructor, which we want to investigate further, we select the model in standardized Pareto scale with the lowest AIC.
- Use likelihood ratio tests to see if we can simplify the parameters of the distribution. This can result in equal shape or location parameters, which increases sparsity.
- Select a quantile c on the normal scale for every asset. Simultaneously fit the GP margins to $\mathbf{R} - \mathbf{c} | \mathbf{R} \nleq \mathbf{c}$ by maximizing equation 13. This means that the return of one asset should be higher than the respective threshold c to be included in the model.
- Use likelihood ratio tests to see if the marginal parameters can be simplified.

This does not guarantee that the optimal model will be chosen, however if the steps above are followed it results in a model that is one of the best options, while increasing sparsity wherever possible.

4.3 Tree model

The method from Kiriliouk et al. (2019) has one major drawback, the computation time for higher dimensions is not feasible. Therefore we apply a method from Engelke and Hitz (2018), where we parsimoniously estimate the relationship between extremal returns. This relationship can be described by a tree structure. It relies on the notion of conditional (extremal) independence. It states that for a graph G = (V, E) with nodes V and edges E, we say that Y is an extremal graphical model if it satisfies that pairwise Markov property:

$$Y_i \perp_e Y_j | Y_{\neg(i,j)}, (i,j) \notin E \tag{14}$$

Here the symbol $\perp_e$ states the conditional independence. Equation 14 tells us that if we know all the other variables, two variables are independent if they do not share a edge. This is especially helpfull in the financial setting, where most stocks are correlated with each other; through this property we can make a parsimonious model, because if there is not an edge between two vertices then the 2 variables are independent, conditioned on all the other variables.

To model this extreme dependence structure, we use the family of Hüsler-Reiss distribution. This distribution is arguably the standard normal distribution of the extreme world. It is parameterised with the variogram Γ. The exponent measure of this distribution is defined as, with k as the reference index:

$$\lambda(\mathbf{y}) = y_k^{-2} \prod_{i \neq k} y_i^{-1} \phi_{d-1}(\overline{y}; \Sigma^k) \tag{15}$$

Where $\phi_{d-1}(*, \Sigma)$ is defined as the multivariate normal distribution. $\overline{y}$ is defined as $log(y_i/y_k + \Gamma_{i,k}/2)$. The covariance is defined as:

$$\Sigma^k = \frac{1}{2}(\Gamma_{i,k} + \Gamma_{j,k} - \Gamma_{i,j})_{i,j \neq k} \tag{16}$$

The exponent measure connects to decomposable graphs in the following manner (Engelke and Hitz 2018): there is a 3 way equivalence if we say that the total density Y satisfies the pairwise markov property for every combination of vertices if and only if the density satisfies the global markov property if and only if the density Y of the graph G can be written as:

$$f_Y(y) = \frac{1}{\Lambda(1)} \frac{\prod_{c \in \mathcal{C}} \lambda_C(Y_C)}{\prod_{d \in \mathcal{D}} \lambda_D(Y_D)} \tag{17}$$

Where $\mathcal{C}$ is the clique set of the graph and $\mathcal{D}$ is the separator set of the graph. The clique set are all the groups of vertices which are all connected to each other. An interpretation is that within a certain clique the vertices are not conditionally independent, because they do not satisfy the pairwise markov property. The separator set is the set of vertices which seperates the cliques in the graph and are the reason conditional independence could occur. Furthermore:

$$\Lambda(1) = 2\phi(\frac{\sqrt{\Gamma_{12}}}{2}) \tag{18}$$

If we use a tree as our graph structure, the cliques are the edges in the graph and the seperator set are

vertices. This implies that the density of a tree equals:

$$f_Y(y) = \frac{1}{\Lambda(1)} \prod_{(i,j)\in E} \frac{\lambda_{i,j}(y_i, y_j)}{y_i^{-2} y_j^{-2}} \prod_{i\in V} y_i^{-2} \tag{19}$$

Here we can see how this method creates sparsity: only the inputs of the variogram were there is an edge in the graph are needed to determine the density,
For now it is assumed that all the variables were on Pareto scale, e.g $\frac{1}{1-u}$, this implies that the shape parameter equals 0 and the scale parameter equals 1. However to asses the risk metrics we need to know these mariginal parameters. Therefore we estimate a different density where we also estimate the scale and shape parameter on the MGPD scale , e.g $\mathbf{X} - \mathbf{v}|\mathbf{X} \nleq \mathbf{v}$. To get the parameters of the joint distribution we can estimate the dependence structure and the marginal parameters of the exceedences simultaneously. The conditional density of X-t given that there is one component that satisfies $\frac{X}{t} > 1$ equals:

$$f_{MGPD}(x, \theta, \sigma, \xi) = f_Y((1 + \xi \frac{x}{\sigma_u})^{\frac{1}{\xi}}, \theta) \prod_{j\in V} \frac{1}{\sigma_j} (1 + \xi \frac{x}{\sigma_u})^{\frac{1}{\xi}-1} \tag{20}$$

Where θ are the dependence parameters and ξ_j and σ_j are the marginal parameters.

4.3.1 Censoring

We use censoring for the same reasons as explained in section 4.2.4 . The density therefore becomes:

$$f^C_{MGPD}(x, \theta, \sigma, \xi) = \int_{j\in C\times(-\infty,0)} f_{MGPD}((1 + \xi \frac{x}{\sigma_u})^{\frac{1}{\xi}}_{D/C}, 0_C, \theta; \sigma; \xi) \tag{21}$$

Where D is the set of all indices and C is the set of indices which are smaller than 0. 0_C indicates that all the assets which are censored get the integral value up to 0 from $-\infty$. Y is on the MGPD scale. This results in the following likelihood:

$$L(x, \theta, \sigma, \xi) = \prod_{i=1}^{n} f^C_{MGPD}(x_{D_i/C_i}, 0_{C_i}, \theta, \sigma, \xi) \tag{22}$$

which needs to be maximised by means of maximum likelihood.

4.4 Tree

Until now the structure of the graph was known, however this is usually not the case. We construct a minimum spanning tree (Kruskal 1956). The algorithm constructs edges such that all vertices are included and the sum of the weights is minimized:

$$\mathcal{T} = \underset{\mathcal{T}=(V,E)}{\arg\min} \sum_{(i,j)\in E} w_{ij} \tag{23}$$

The weights determine the tree structure and need to be chosen carefully. It is common to choose the conditional independence structure that maximizes the likelihood Kirshner, Smyth, and Robertson (2012). We use the following weights on Pareto scale:

$$w_{i,j} = -L(\hat{\Gamma_{i,j}}, y^{(1)}, ..., y^{(n)}) - 2 \sum_{y_i>1} log(y_i) - 2 \sum_{y_j>1} log(y_j) \tag{24}$$

4.5 Model selection

We first select a tree by the Kruskal algorithm (Kruskal 1956) on standardized Pareto data, this allows us to first determine the tree structure before we estimate the marginal parameters. We censor the tree on a quantile of 0.9, the values which fall below this quantile are censored and also contribute to the likelihood in a censored fashion. After we obtain the tree structure of these assets, we estimate the marginal parameters as well as the dependence parameters by means of maximizing equation 22. Here we also censor all the data below the quantiles of 0.9 as will be discussed in the results.

4.6 Risk Metrics

Ultimately we want to calculate the risk of a given portfolio. To calculate the Value at Risk and Expected Shortfall we use the property of sum stability under shape property. It states that when X follows a multivariate Generalized Pareto distribution with scale parameter σ and shape parameter $\gamma = \gamma 1$, then the weighted sum of different components of X also follows a Generalized Pareto distribution:

$$\sum_{j=1}^{d} a_j X_j | \sum_{j=1}^{d} a_j X_j > 0 \sim GP(\sum_{j=1}^{d} a_j \sigma_j, \gamma) \tag{25}$$

This helps us to calculate the risk metrics of a portfolio consisting of assets with certain weights. We consider 2 different risk metrics: Value at Risk and Expected Shortfall. The Value at Risk is calculated as follows, if $\sum_{j=1}^{d} a_j (Y_{t,j} - u_j) > 0$ is $GP(\sum_j a_j \sigma_j, \gamma)$, then:

$$VaR(p) = \sum_{j=1}^{d} a_j u_j + \frac{\sum_{j=1}^{d} a_j \sigma_j}{\gamma} ((\frac{\phi}{p})^{\gamma} - 1) \tag{26}$$

where u_j is the individual threshold for asset j and phi is the probability that an outlier occurs. and the expected shortfall is calculated in the following manner:

$$ES(p) = VaR(p) + \frac{\sum_{j=1}^{d} a_j \sigma_j + \gamma(VaR(p) - \sum_{j=1}^{d} a_j u_j)}{1 - \gamma} \tag{27}$$

We want to see if these methods are usefull in estimating the Value at Risk and Expected Shortfall. Therefore we test the results of these methods on the test set, which includes the negative weekly returns from 1 January 2018 untill 5 May 2020.

For the Value at Risk we use the unconditional coverage test of Christoffersen (1998). It is a likelihood ratio test with the null hypothesis that the exceedence probability of the q% VaR is exactly 1-q%. If the null hypothesis is rejected the VaR estimate is not sufficient.

For the Expected Shortfall we use the test of McNeil and Frey (2000). It tests whether the expected value of the exceedences on the VaR are significantly bigger than zero. It uses bootstrapping to calculate the standard errors. It is used to test the Expected Shortfall.

5 Results

We begin with discussing the results of the clustering method, thereafter we discuss the results of the tree model. Finally we look how accurately these models predict the Value at Risk and Expected Shortfall.

5.1 Results Clustering method

We begin with describing the clusters that are made, thereafter we discuss the model selection and we conclude with the Value at Risk and Expected Shortfall.

5.1.1 Results K-means

The 3 clusters that are constructed by means of K-means clustering are as follows:

	Cluster 1	Cluster 2	Cluster 3
	KPN	AKZO	Aegon
	Unibail	ASML	DSM
		Unilever	ING
		Heineken	Phillips
		Wolters	Shell
		Relx	ARM
			Randstad
			ASM
Weight adj-AEX	0.036	0.567	0.397

In the first cluster we see the telecom company KPN and the real estate company Unibail. The second cluster we see the publishers Wolters-Kluwe and Relx, but also companies which are present in the consumer markets, like Heineken and Unilever, also two tech companies are included with Akzo Nobel and ASML. The third cluster consists of financial intermediaries, such as insurer Aegon and the bank ING, companies which are dependent on the oil price with Shell and Arcellor-Mittel. Furthermore DSM, Phillips, Randstad and ASM are also included.

5.1.2 Threshold

In Figure 4 we see the empirical χ plot of the returns of the 3 clusters with 95 % confidence intervals.

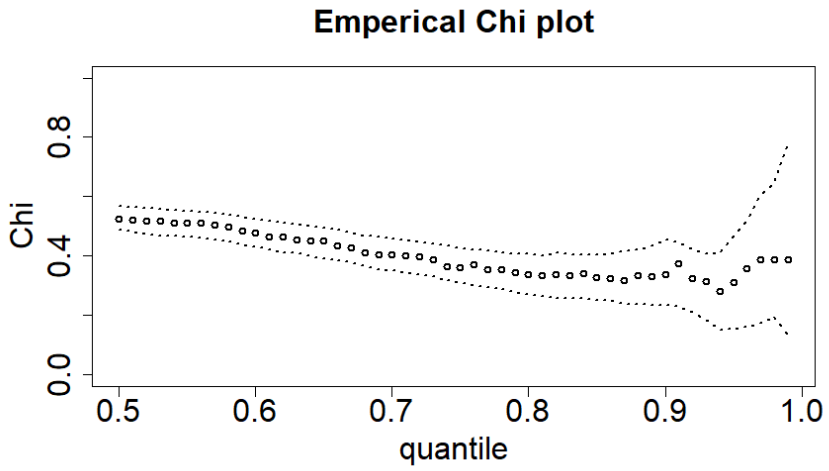


Figure 4: Emperical Chi plot

We look for the quantile where the χ plot straightens, we see that this happens at the 90% quantile. Therefore in the remainder of the results, we censor the data below the quantile of 0.9.

5.1.3 Densities

To choose the density for the model we first fit the model with the most parameters and look at the Akaike Information Criterion (AIC) to see which model has the best fit. The Akaike information criterion is the negative likelihood plus two times the number of parameters in the model.

	NLL	AIC
Gumbel T	594.07	604.07
MVGaus T	605.08	615.08
Rev Exp T	598.44	608.44
Rev EXP U	600.87	610.87
Gumbel U	599.45	609.45

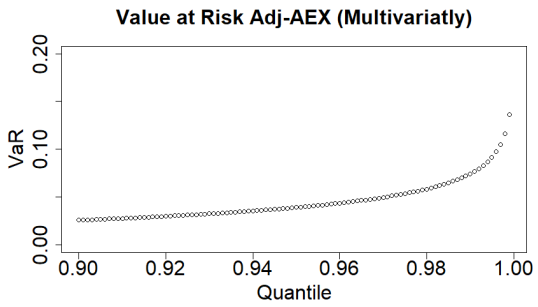
Table 2: Negative Log likelihoods with AIC of densities

We see that the Gumbel Distribution with the T constructor has the lowest AIC, therefore we continue with this combination of density and constructor to further evaluate the model.
To determine with what parameterazation of the model we want continue, we estimate (1) the Gumbel distribution with the same location parameter, but unique scale parameters.(2) The Gumbel distribution with unique location parameters with the same scale.(3) The Gumbel distribution with just one free parameter. After these we conclude by means of a likelihood ratio test that there is no significant difference between the model with 1 parameter and the full model, hence we continue with the model with only 1 parameter.
We continue with the estimation on the multivariate Generalized Pareto scale (e.g the difference of Returns minus thresholds, given that one cluster surpasses the threshold.), where we estimate the parameter of the Gumbel distribution and the marginal parameters of the Pareto distribution simultaneously. The negative log likelihood of the model with individual scale and shape parameters equals -341.08, the model with an equal shape parameter has a negative log likelihood of -338.85. If we apply the likelihood ratio test we do not reject the null hypothesis of equal shape parameters therefore, we continue with the model with equal shape parameters.

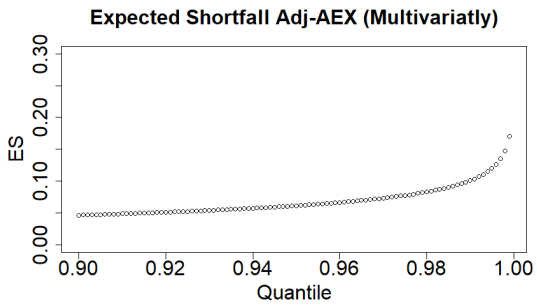
5.2 Risk Metrics

We determine the Value at Risk and the Expected Shortfall by means of equation 26 and 27. To do this we need the marginal parameters γ and the σ_i for each asset i.

The Value at Risk can be seen in Figure 5a and the Expected Shortfall can be found in Figure 5b.



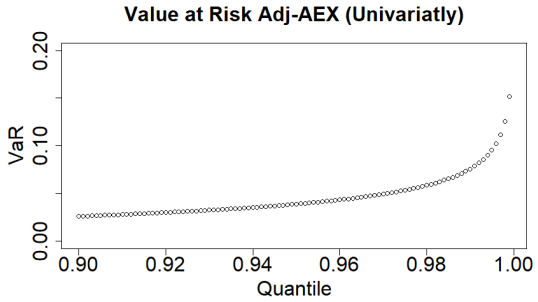
(a) Value at Risk Clustering model



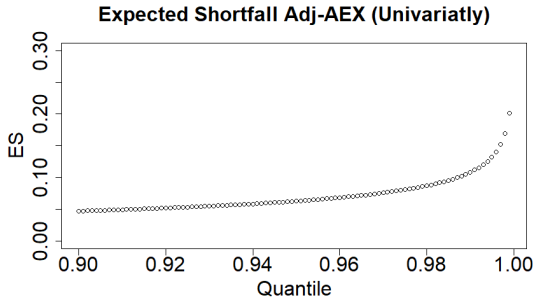
(b) Expected Shortfall Clustering model

Figure 5: Results Cluster model

The Value at Risk of the Univatiate Generalized Pareto can be seen in Figure 6a and the Expected Shortfall can be found in Figure 6b.



(a) Value at Risk univariatly fitted.



(b) Expected Shortfall univariatly fitted.

Figure 6: Results Univariate model

5.3 Tree model

In Figure 7 we see how the companies from the adjusted AEX are connected.

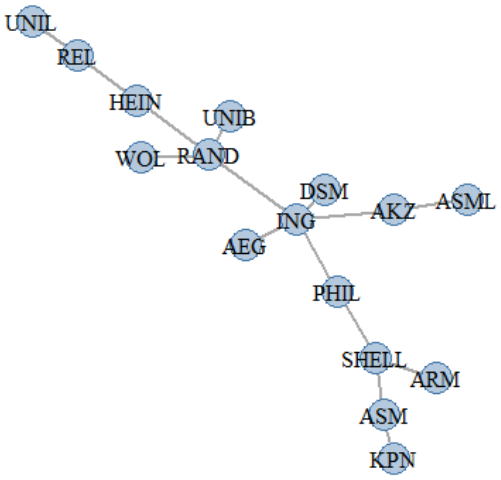


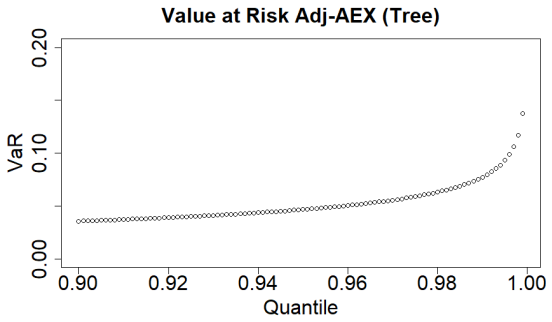
Figure 7: Companies Extremal Tree

In the center of the graph we see the Bank ING, this is in line with our expectations because the banking sector connects all the different industries. It is connected with the insurer Aegon, which tells us the close relation between financial intermediaries. On the right ING is connected to 2 big chemical companies, DSM and Akzo Nobel, which is on it self connected to the chip producer ASML. On the top left corner, ING is connected to the employment agency Randstad. Randstad is connected to both the real estate agent Unibail and the publisher Wolters Kluwe. It is also connected to the brewer Heineken, which is connected to information company Relx, where that is again connected to Unilever. On the bottom right corner we see that ING is connected with electronics fabricant Phillips. Phillips is connected to Oil company Shell, which is connected with steel producer Arcellor Mittel and Chemical company ASM. Asm is connected to KPN.

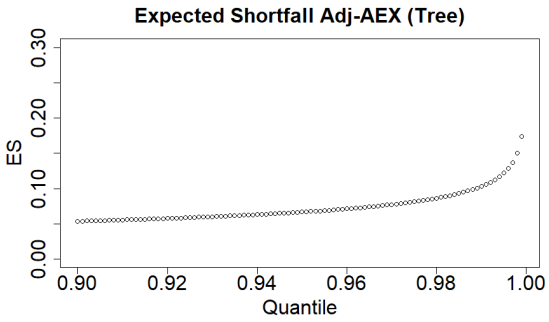
5.3.1 Risk Metrics

We want to calculate the Value at Risk and Expected Shortfall by means of equation 26 and 27. Therefore we need the marginal parameters, we estimate these jointly by maximizing equation 22, where we make use of censoring and also estimate the variogram parameters. To make use of the sum stability property of the Pareto distributions we need to test whether or not all the assets have the same γ. Therefore we estimate equation 22 both with equal gamma and individual gamma and we use a LR test to determine if the γ parameters are significantly different. The Log likelihood of the model with individual γ parameters equal 135676, whereas it is 98574 for the model with equal γ parameters. With these log likelihoods we reject the null hypothesis of equal γ parameters.

However, to still do inference on the Value at Risk and Expected Shortfal,l we continue with the model where all the γ parameters are equal. The Value at Risk of the adjusted-AEX of the Tree model is calculated using equation 26 and can be found in Figure 8a and the Expected Shortfall calculated with equation 27 can



(a) Value at Risk of Adjusted-AEX (Tree)



(b) Expected Shortfall of Adjusted-AEX (Tree)

Figure 8: Results Tree model

be seen in Figure 8b.

5.4 Backtesting

The results of the backtest can be seen in Table 3, 4, 5 for the 90 %, 95%, 99% respectively. In the first column the expected amount of exceedances is given, in the second column the realized exceedances, in the third column the p-value of the unconditional coverage test (Christoffersen 1998), in the fourth column the p-value of the test, which determines if the exceedances of the VaR are significantly bigger than zero (McNeil and Frey 2000).

	Expected	Realized	P-value (VaR)	P-value (ES)
Tree	12	9	0.3	0.139
Clustering	12	14	0.251	0.197
Univariate	12	14	0.251	0.213

Table 3: 90% Test

	Expected	Realized	P-value (VaR)	P-value (ES)
Tree	6	5	0.623	0.145
Clustering	6	9	0.268	0.252
Univariate	6	9	0.268	0.378

Table 4: 95% Test

	Expected	Realized	P-value (VaR)	P-value (ES)
TREE	1	2	0.522	0.0478
Clustering	1	2	0.522	0.042
Univariate	1	2	0.522	0.0614

Table 5: 99% Test

For the test of correct unconditional coverage of the Value at Risk, we do not reject the null hypothesis of correct coverage for any of the three models on all three levels. We do see that the p-value is greater or equal for the Tree model than for the Clustering or Univariate model. For the Expected shortfall we do not

reject the null hypothesis that the excesses have a mean that is 0 for the 90% and 95% coverage, however for the Tree and Clustering model in the 99% case this does occur.

6 Conclusion

The literature on multivariate extreme value theory is not very broad, especially on the application of finance. Therefore we apply two methods which are computationally efficient. Hereby we extend the literature of multivariate Pareto distributions, which are computationally efficient for higher dimensions, which is needed for portfolio management where there are a lot of assets.

The first method is based on two steps. We first cluster the number of assets in the portfolio into a smaller amount of clusters and then we apply the methodology of Kiriliouk et al. (2019). A pitfall of this method was that it could only deal with moderate dimensions, however by clustering we work around this problem by reducing the amount of dimensions and creating a smaller amount of sub-portfolios. The second method does not need clustering to come up with a computationally efficient answer. It relies on the methodology of Engelke and Hitz (2018), where we create sparsity by creating a tree structure in the assets of the portfolio.

We use the data from 16 companies of the AEX, where we create a portfolio scaled to the weights these companies had in the AEX at 31 March of 2020. We train the 2 models with data from 25 July 2005 untill 31 December 2017 and we test the models for accurate Value at Risk and Expected Shortfall metrics on the returns from 1 January 2018 untill 5 May 2020. We also compare these methods to the univariate Generalized Pareto distribution.

We see that the 2 methods can accurately describe the Value at Risk for both 90%, 95% and 99% coverage. We see that the model where we created a tree has lower p values than the univariate and clustering model. Furthermore we see that the predicted Expected Shortfall is significant for the 99% model.

For further extensions we suggest to make more complicated graphs, as is noted the graph must be undirected and acyclical, but by adding edges to the minimum spanning tree the tree can be extended. Moreover we still had a relatively small asset space (N=16), it would be interesting to see how these methods cope with even larger number of assets (say N=100). Furthermore, a method at which the tree structure and the marginal parameters are estimated simultaneously would be helpfull.

References

[1] Allali Abdelwahab, Oueslati Amor, and Trabelsi Abdelwahed. "The analysis of the interdependence structure in international financial markets by graphical models". In: *International research journal of finance and economics* 15.2 (2008), pp. 1–306.

[2] CW Anderson, Jacques De Maré, and Holger Rootzén. "Methods for estimating the sizes of large inclusions in clean steels". In: *Acta Materialia* 53.8 (2005), pp. 2295–2304.

[3] August A Balkema and Laurens De Haan. "Residual life time at great age". In: *The Annals of probability* (1974), pp. 792–804.

[4] David A Bessler and Jian Yang. "The structure of interdependence in international stock markets". In: *Journal of international money and finance* 22.2 (2003), pp. 261–287.

[5] Axel Bücher and Johan Segers. "Extreme value copula estimation based on block maxima of a multivariate stationary time series". In: *Extremes* 17.3 (2014), pp. 495–528.

[6] Peter F Christoffersen. "Evaluating interval forecasts". In: *International economic review* (1998), pp. 841–862.

[7] Stuart G Coles and Jonathan A Tawn. "Modelling extreme multivariate events". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 53.2 (1991), pp. 377–392.

[8] Anthony C Davison and Raphaël Huser. "Statistics of extremes". In: *Annual Review of Statistics and its Application* 2 (2015), pp. 203–235.

[9] Anthony C Davison and Richard L Smith. "Models for exceedances over high thresholds". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 52.3 (1990), pp. 393–425.

[10] A Philip Dawid. "Conditional independence in statistical theory". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 41.1 (1979), pp. 1–15.

[11] Sebastian Engelke and Adrien S Hitz. "Graphical models for extremes". In: *arXiv preprint arXiv:1812.01734* (2018).

[12] Michael Falk and Armelle Guillou. "Peaks-over-threshold stability of multivariate generalized Pareto distributions". In: *Journal of Multivariate Analysis* 99.4 (2008), pp. 715–734.

[13] Ana Ferreira, Laurens De Haan, et al. "The generalized Pareto process; with a view towards application and simulation". In: *Bernoulli* 20.4 (2014), pp. 1717–1737.

[14] Raphaël de Fondeville and Anthony C Davison. "High-dimensional peaks-over-threshold inference". In: *Biometrika* 105.3 (2018), pp. 575–592.

[15] Manfred Gilli et al. "An application of extreme value theory for measuring financial risk". In: *Computational Economics* 27.2-3 (2006), pp. 207–228.

[16] TJ Gordon, LP Kostyniuk, PE Green, MA Barnes, DF Blower, SE Bogard, AD Blankespoor, DJ LeBlanc, BR Cannon, and SB McLaughlin. *A multivariate analysis of crash and naturalistic driving data in relation to highway factors*. Tech. rep. Transportation Research Board, Washington, DC, 2013.

[17] Emil Julius Gumbel. *Statistics of extremes*. Courier Corporation, 1958.

[18] John M Hammersley and Peter Clifford. "Markov fields on finite graphs and lattices". In: *Unpublished manuscript* 46 (1971).

[19] Peter J Hawkes, Ben P Gouldby, Jonathan A Tawn, and Michael W Owen. "The joint probability of waves and water levels in coastal engineering design". In: *Journal of hydraulic research* 40.3 (2002), pp. 241–251.

[20] Raphaël Huser, Anthony C Davison, and Marc G Genton. "Likelihood estimators for multivariate extremes". In: *Extremes* 19.1 (2016), pp. 79–103.

[21] Jürg Hüsler and Rolf-Dieter Reiss. "Maxima of normal random vectors: between independence and complete dependence". In: *Statistics & Probability Letters* 7.4 (1989), pp. 283–286.

[22] Zakhar Kabluchko, Martin Schlather, Laurens De Haan, et al. "Stationary max-stable fields associated to negative definite functions". In: *The Annals of Probability* 37.5 (2009), pp. 2042–2065.

[23] Richard W Katz, Marc B Parlange, and Philippe Naveau. "Statistics of extremes in hydrology". In: *Advances in water resources* 25.8-12 (2002), pp. 1287–1304.

[24] Anna Kiriliouk, Holger Rootzén, Johan Segers, and Jennifer L Wadsworth. "Peaks over thresholds modeling with multivariate generalized Pareto distributions". In: *Technometrics* 61.1 (2019), pp. 123–135.

[25] Sergey Kirshner, Padhraic Smyth, and Andrew Robertson. "Conditional Chow-Liu tree structures for modeling discrete-valued vector time series". In: *arXiv preprint arXiv:1207.4142* (2012).

[26] Joseph B Kruskal. "On the shortest spanning subtree of a graph and the traveling salesman problem". In: *Proceedings of the American Mathematical society* 7.1 (1956), pp. 48–50.

[27] Steffen L Lauritzen. *Graphical models*. Vol. 17. Clarendon Press, 1996.

[28] James MacQueen et al. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA. 1967, pp. 281–297.

[29] Henrik Madsen, Peter F Rasmussen, and Dan Rosbjerg. "Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling". In: *Water resources research* 33.4 (1997), pp. 747–757.

[30] Alexander J McNeil and Rüdiger Frey. "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". In: *Journal of empirical finance* 7.3-4 (2000), pp. 271–300.

[31] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[32] Thomas Opitz. "Extremal t processes: Elliptical domain of attraction and a spectral representation". In: *Journal of Multivariate Analysis* 122 (2013), pp. 409–413.

[33] James Pickands III et al. "Statistical inference using extreme order statistics". In: *the Annals of Statistics* 3.1 (1975), pp. 119–131.

[34] Patrick Ragan and Lance Manuel. "Statistical extrapolation methods for estimating wind turbine extreme loads". In: *Journal of Solar Energy Engineering* 130.3 (2008).

[35] Holger Rootzén, Johan Segers, and Jennifer L Wadsworth. "Multivariate generalized Pareto distributions: Parametrizations, representations, and properties". In: *Journal of Multivariate Analysis* 165 (2018), pp. 117–131.

[36] Holger Rootzén, Johan Segers, and Jennifer L Wadsworth. "Multivariate peaks over thresholds models". In: *Extremes* 21.1 (2018), pp. 115–145.

[37] Holger Rootzén, Nader Tajvidi, et al. "Multivariate generalized Pareto distributions". In: *Bernoulli* 12.5 (2006), pp. 917–930.

[38] Martin Schlather. "Models for stationary max-stable random fields". In: *Extremes* 5.1 (2002), pp. 33–44.

[39] Richard L Smith. "Estimating tails of probability distributions". In: *The annals of Statistics* (1987), pp. 1174–1207.

[40] Richard L Smith. "Threshold methods for sample extremes". In: *Statistical extremes and applications*. Springer, 1984, pp. 621–638.

[41] Richard L Smith, Jonathan A Tawn, and Stuart G Coles. "Markov chain models for threshold exceedances". In: *Biometrika* 84.2 (1997), pp. 249–268.

[42] Makram Talih and Nicolas Hengartner. "Structural learning with time-varying components: tracking the cross-section of financial time series". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.3 (2005), pp. 321–341.

[43] Martin J Wainwright, Michael I Jordan, et al. "Graphical models, exponential families, and variational inference". In: *Foundations and Trends® in Machine Learning* 1.1–2 (2008), pp. 1–305.