Risk, rare events and extremes: final project

Alessandro La Farciola

EPFL

Structure of semester project

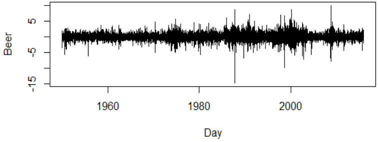
1. Learning main results about extreme theory
   - Lectures of the course *Risk, rare events and extremes*
   - Slides
   - Exercises and solutions
   - Weekly meetings

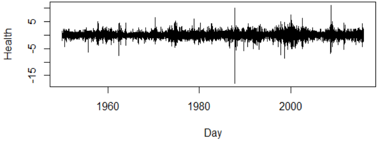
Structure of semester project

1. Learning main results about extreme theory
   - Lectures of the course *Risk, rare events and extremes*
   - Slides
   - Exercises and solutions
   - Weekly meetings
2. Project

Project

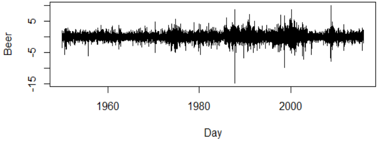
- Analyze 66 years of data of averaged daily returns (unit in $\times 100\%$) from January 1950 to December 2015.





Project

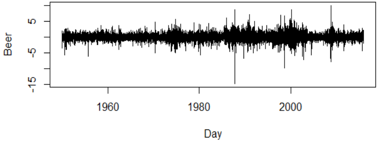
- Analyze 66 years of data of averaged daily returns (unit in $\times 100\%$) from January 1950 to December 2015.

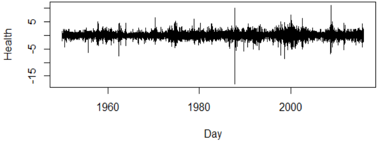


- First part: estimate one-step-ahead conditional quantiles of extreme negative returns at 1-year, 10-years, 100-years return levels through a univariate analysis.

Project

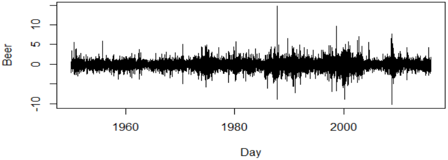
- Analyze 66 years of data of averaged daily returns (unit in $\times 100\%$) from January 1950 to December 2015.





- First part: estimate one-step-ahead conditional quantiles of extreme negative returns at 1-year, 10-years, 100-years return levels through a univariate analysis.
- Second part: analysis in a bivariate environment using multivariate extreme value statistics techniques in order to model dependence between the two series.

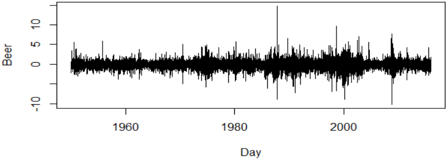
Univariate analysis: *Beer* series



- First approach: stationary model. In particular, PoT approach where we fit data over a certain threshold with GPD distribution

$$H(x) = \begin{cases} 1 - (1 + \xi x/\sigma)_+^{-1/\xi} & \textit{if } \xi \neq 0 \\ 1 - \exp(-x/\sigma) & \textit{if } \xi = 0 \end{cases} \quad \forall x > 0.$$

Univariate analysis: *Beer* series

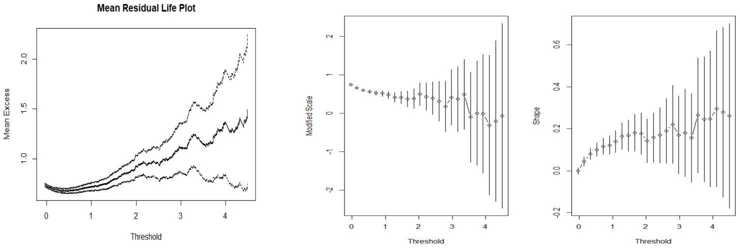


- First approach: stationary model. In particular, PoT approach where we fit data over a certain threshold with GPD distribution

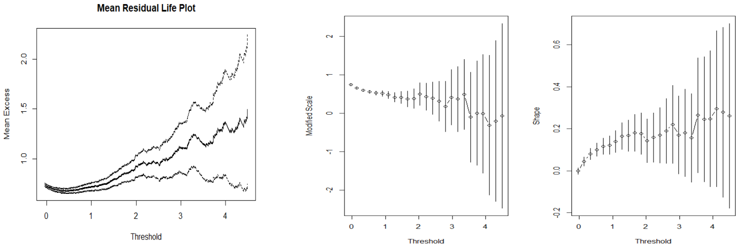
$$H(x) = \begin{cases} 1 - (1 + \xi x/\sigma)_+^{-1/\xi} & \textit{if } \xi \neq 0 \\ 1 - \exp(-x/\sigma) & \textit{if } \xi = 0 \end{cases} \qquad \forall x > 0.$$

- We need to choose a valid threshold u. We can investigate stability in terms of *Mean excess* and *Parameter estimates*.

Stationary PoT approach

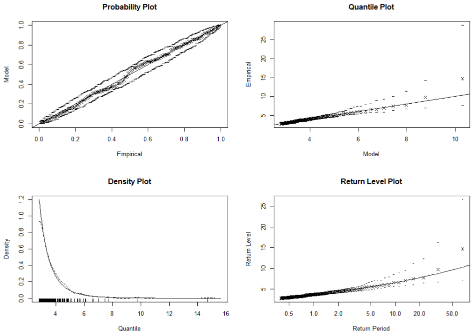


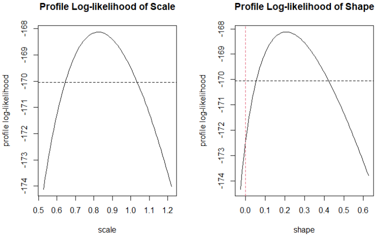
Stationary PoT approach



- The choice of threshold $u = 2.81$ seems reasonable.
- We have 16694 observations in 66 years and so we model with 253 observations per year.
- Estimated parameters are $\hat{\sigma} = 0.82(0.10)$ and $\hat{\xi} = 0.21(0.09)$, that gives normally based 95%-confidence intervals equal to $(0.63, 1.02)$ and $(0.02, 0.39)$.

Stationary PoT approach. diagnostic plots



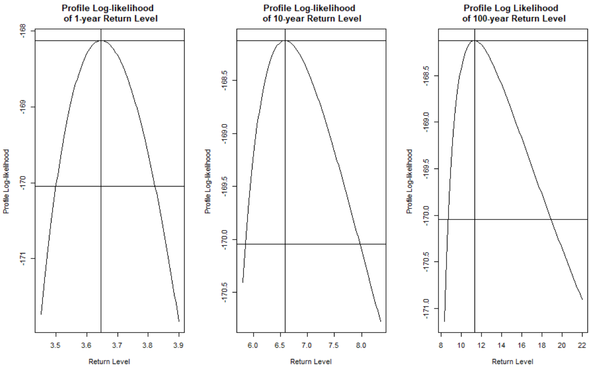


It is possible to explore also the Gumbel model with the constraint $\xi = 0$, but, as we could expect, diagnostic plots confirm that 2-parameter model is more reasonable.

We can compute the m-years return level that is given by

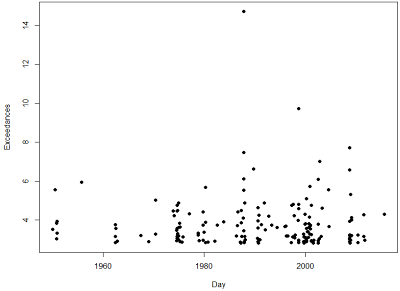
$$x_m = u + \frac{\hat{\sigma}}{\hat{\xi}}\{(m\hat{\zeta}_u)^{\hat{\xi}} - 1\},$$

where $\hat{\zeta}_u$ is the probability of exceeding the threshold u and $m = \tilde{m} \cdot 253$ with $\tilde{m} = 1, 10, 100$.



Non-stationary PoT approach

- One of the main features of financial returns is the tendency to occur in clusters. This is due to the strong dependence of daily returns from the most recent past.



- One way to take into account this features can be modelling in a non-stationary environment.

Non-stationary PoT approach

- Set a certain time lag l in a way so that the rolling time window spans a certain period. Namely, for each time t, we let the threshold u_t to be a high quantile of previous observations up to l, for example a month, three months or a year. After choosing a threshold u_t, we should model exceedances using Generalized Pareto distribution.
- Let $Y_t = X_t - u_t$ for $t = l+1, \ldots, n$, and let $T_1, \ldots, T_{N_u}$ denote the exceedance times. It seems reasonable to assume a Markov-like structure for the threshold exceedances Y_{T_k}. Calling $\mathcal{F}_t^X$ the σ-algebra generated by observations $X_1, \ldots, X_t$,

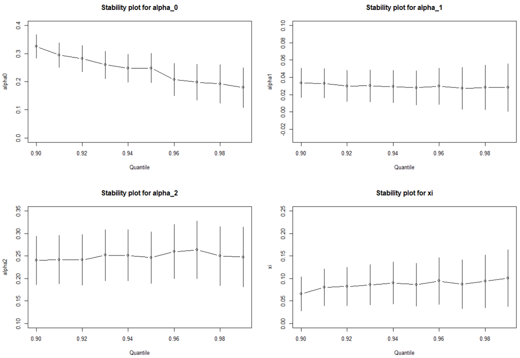
$$Y_{T_k} \mid \mathcal{F}_{T_k - 1}^X \sim GPD(\sigma_{T_k}, \xi).$$

- More specifically, we can model the scale parameter depending on some covariates, for example on previous data and previous threshold, according to the following expression:

$$\sigma_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 u_{t-1}, \qquad t = l+2, \ldots, n.$$

Non-stationary PoT approach

- For each choice of lag l, in order to choose a valid quantile level, we can use stability plot.



- This suggests that choosing 0.95 as quantile level can be a reasonable choice.

Non-stationary PoT approach

- According to model presented so far, we can use different choices of lag $l = 21, 21 \cdot 3, 21 \cdot 12$.

	α_0	α_1	α_2	ξ	AIC
$l = 21$	0.24 (0.03)	0.03 (0.01)	0.25 (0.03)	0.09 (0.02)	1116.11
$l = 21 \cdot 3$	-0.01 (0.04)	0.04 (0.01)	0.43 (0.05)	0.11 (0.03)	862.83
$l = 21 \cdot 12$	-0.02 (0.06)	0.02 (0.01)	0.40 (0.06)	0.16 (0.04)	905.65

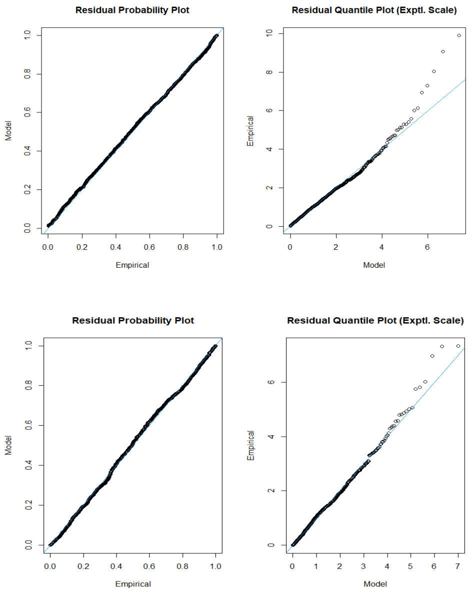
- We can also consider exponential models for the scale parameter ($siglink = exp$), according to

$$\sigma_t = \exp\left(\beta_0 + \beta_1 X_{t-1} + \beta_2 u_{t-1}\right), \qquad t = l + 2, \dots, n,$$

	β_0	β_1	β_2	ξ	AIC
$l = 21$	-1.20 (0.06)	0.07 (0.02)	0.47 (0.05)	0.08 (0.02)	1109.6
$l = 21 \cdot 3$	-1.60 (0.09)	0.07 (0.03)	0.69 (0.07)	0.12 (0.03)	874.17
$l = 21 \cdot 12$	-1.53 (0.12)	0.04 (0.03)	0.60 (0.08)	0.17 (0.04)	914.29

- We can compare the previous model with *Akaike Information criterion* $AIC = 2\{dim(\theta) - \hat{\ell}\}$.

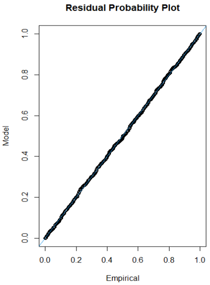
Non-stationary PoT approach: diagnostic plots

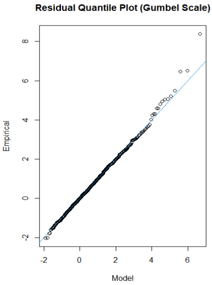


Non-stationary Block Maxima approach

- It is possible to model non-stationarity also using a block maxima approach, extrapolating, the annual maxima (or monthly maxima) M_n from data and fitting them with GEV distribution.
- In particular, we can fit GEV with regression forms for its parameters, namely letting the location and the scale parameter varying in time, according to

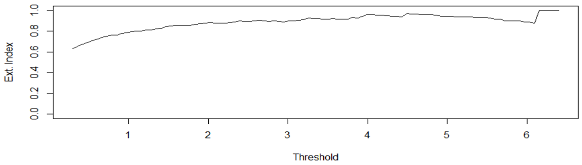
$$\mu_n = \beta_0 + \beta_1 M_{n-1}, \qquad \sigma_n = \alpha_0 + \alpha_1 M_{n-1}.$$





Extremal index

- One way of interpreting the extremal index θ is in terms of the propensity of the process to cluster at extreme levels. Namely, it turns out that θ^{-1} is equal to the limiting mean cluster size.
- It is possible to estimate the extremal index for example with the *exiplot* function in *evd* library of *R*. In particular, it plots the extremal index estimation at a sequence of increasing thresholds using the *runs-method*.



- The extremal index is fairly stable at about 0.9, suggesting that clusters of extreme events are approximately of size $1/0.9 = 1.11$ on average.

Return levels: PoT approach

- In the non-stationary setting, under the Markov-structure assumption, we have that, conditional on the past, the one-step-ahead quantiles are defined as

$$Q_k(p) = \inf\{x \in \mathbb{R} : P(X_k \leq x \mid \mathcal{F}_{k-1}^X) \geq p\}.$$

And so, we have that

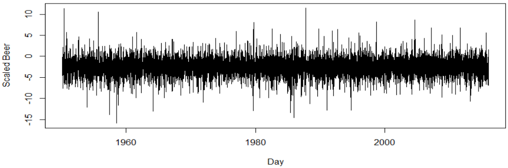
$$P(X_k \leq x \mid \mathcal{F}_{k-1}^X) = 1 - \frac{N_u}{n}\left(1 + \hat{\xi}\,\frac{x - u_k}{\hat{\sigma_k}}\right)^{-1/\hat{\xi}}.$$

- In order to compute the conditional quantiles corresponding to 1,10 and 100-years return levels, we notice that if $Y_{T_k} = X_{T_k} - u_{T_k} | \mathcal{F}_{k-1}^X \sim GPD(\hat{\sigma}_{T_k}, \hat{\xi})$, then

$$\tilde{Y}_{T_k} := Y_{T_k}/\hat{\sigma}_{T_k} \mid \mathcal{F}_{k-1}^X \sim GPD(1, \hat{\xi}).$$

This scaled variable $\tilde{Y}_{T_k}$ can now be considered as a more stationary series.

Return levels: PoT approach



- We fit a stationary GPD model on $\tilde{Y}$ with threshold $u = 0$, and as we could expect we obtain as estimated parameter $\hat{\sigma} = 1$ and the same shape parameter obtained previously.
- We compute m-observation return level according to

$$x_m = u + \frac{\hat{\sigma}}{\hat{\xi}} \left\{ (m n_y \hat{\zeta}_u)^{\hat{\xi}} - 1 \right\}$$

  and we obtain their 95%-confidence intervals via profile likelihood
- We make back transformation to variable X_k multiplying every value for $\hat{\sigma}_k$ and adding the threshold u_k, for each time k.

Return levels: PoT approach

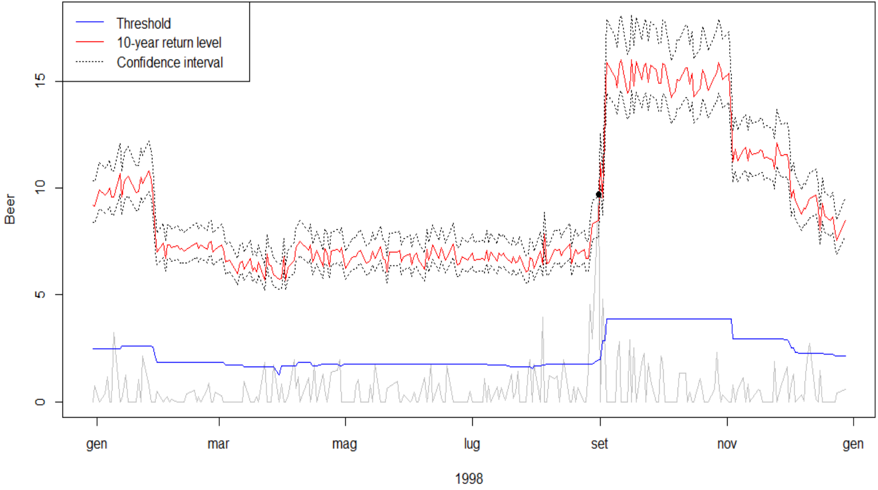
- In this way we are able to compute the conditional quantiles corresponding to 1,10 and 100-years return levels, taking into account also the uncertainty in the estimation.
- For each time k we obtained return levels x_1, x_{10}, x_{100} and their CI $[x_1^{min}, x_1^{max}]$, $[x_{10}^{min}, x_{10}^{max}]$, $[x_{100}^{min}, x_{100}^{max}]$.

Return levels: PoT approach

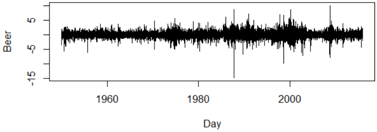
- In this way we are able to compute the conditional quantiles corresponding to 1,10 and 100-years return levels, taking into account also the uncertainty in the estimation.
- For each time k we obtained return levels x_1, x_{10}, x_{100} and their CI $[x_1^{min}, x_1^{max}]$, $[x_{10}^{min}, x_{10}^{max}]$, $[x_{100}^{min}, x_{100}^{max}]$.
- We can validate the computed m-return levels counting the number of observations above the one-step-ahead estimated quantiles in order to investigate the accuracy of these estimations.

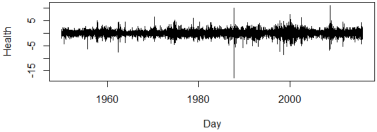
	x_m validation	x_m^{min} validation	x_m^{max} validation
$l = 21$	58 - 8 - 3	64 - 13 - 5	49 - 6 - 2
$l = 21$ *exp*	57 - 9 - 3	63 - 14 - 3	48 - 6 - 2
$l = 21 \cdot 3$	61 - 6 - 0	77 - 12 - 3	56 - 5 - 0
$l = 21 \cdot 3$ *exp*	65 - 7 - 0	73 - 12 - 2	52 - 4 - 0
$l = 21 \cdot 12$	66 - 5 - 1	81 - 9 - 2	54 - 5 - 0
$l = 21 \cdot 12$ *exp*	65 - 5 - 0	85 - 8 - 2	54 - 5 - 0

Return levels: PoT approach



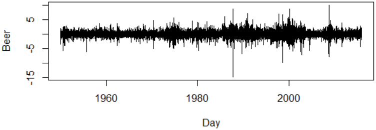
Multivariate analysis: PoT appraoch

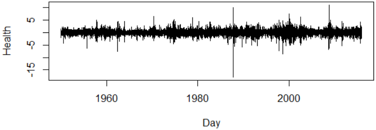




- Consider the best non-stationary model for each marginal and then we consider the standardized series (as discussed previously).

Multivariate analysis: PoT appraoch





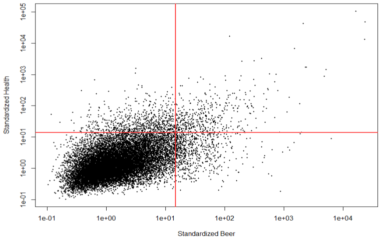
- Consider the best non-stationary model for each marginal and then we consider the standardized series (as discussed previously).
- Apply marginal transformations to unit Fréchet marginal distributions: fit *GPDs* above these thresholds according to

$$\hat{F}_d(x) = \begin{cases} \#\{j : x_{j,d} \leq x\}/n & x \leq 0 \\ 1 - \hat{p}_u \left\{1 + \hat{\xi}_d(x - u_d)/\hat{\sigma_d}\right\}_+^{-1/\hat{\xi}_d} & x > 0, \end{cases}$$

and apply the component-wise transformation for variables

$$z_j = -1/\log\{\hat{F}_d(x_j)\}.$$

Multivariate analysis: PoT approach



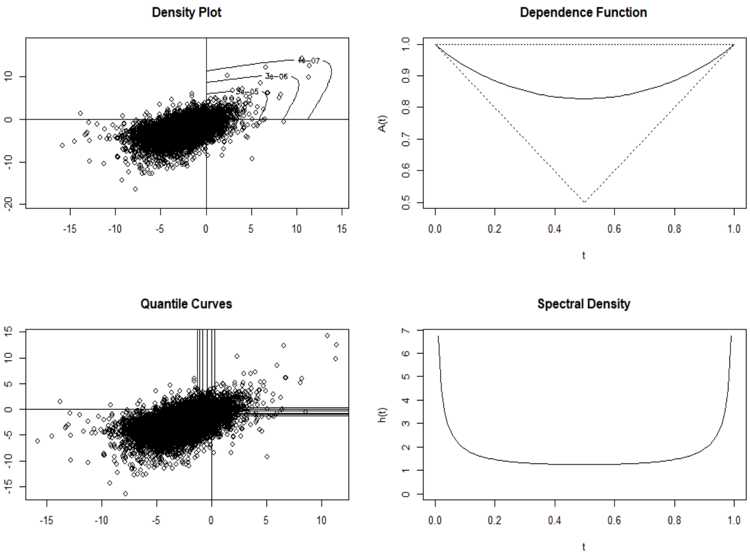
- The most simple and natural model that we can try to fit is the logistic. In this case, the bivariate extreme value distribution is given by

$$G(z_1, z_2) = \exp\left[-\left\{z_1^{-1/\alpha} + z_2^{-1/\alpha}\right\}^{\alpha}\right].$$

- We get an estimated dependence parameter equal to

$$\hat{\alpha} = 0.726\ (0.011) \qquad \Rightarrow \qquad CI:\ (0.71, 0.75).$$

Multivariate analysis: PoT approach



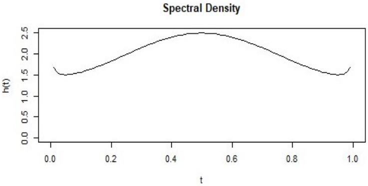
Multivariate analysis: PoT approach

- Try to fit a simpler model with a common shape parameter. The estimated parameters are very similar and a likelihood ratio test suggest that we cannot reject the null hypothesis, i.e. common shape parameter. The latter is preferable for the fitting since it is a simpler model.
- A second model that we can try to fit is the asymmetric logistic. In this case the estimated dependence parameter is $\hat{\alpha} = 0.692$ (0.032). The likelihood ratio test suggest that the asymmetric model cannot be considered significantly better than the logistic one and so we prefer the parsimonious (symmetric) logistic model.

Models	AIC
Logistic	20044.55
Negative Logistic	20055.55
Bilogistic	20045.97
Coles-Tawn	20049.53
Negative Bilogistic	20056.44
Husler-Reiss	20071.06

Multivariate analysis: Yearly maxima approach

- We extrapolate yearly maxima for both series, fit GEV marginals and estimate the dependence parameters.
- Fitting with logistic model we obtain an estimation of the dependence parameter equal to $\hat{\alpha} = 0.537$ (0.063).



Models	AIC
Logistic	447.45
Negative Logistic	451.85
Bilogistic	447.73
Coles-Tawn	449.20
Negative Bilogistic	449.28
Husler-Reiss	449.31

Table: Comparison between different models

Conclusions

- In the first part, we wanted to fit an univariate model for extreme negative returns for both series and estimate the one-step-ahead conditional quantiles of extreme negative returns at 1-year, 10-years, 100-years return levels.
- We tried to model with stationary assumptions and then we noticed that probably a non-stationary setting would have been more appropriate especially for capturing the usual tendency of extreme financial returns to occur in clusters.

Conclusions

- In the first part, we wanted to fit an univariate model for extreme negative returns for both series and estimate the one-step-ahead conditional quantiles of extreme negative returns at 1-year, 10-years, 100-years return levels.
- We tried to model with stationary assumptions and then we noticed that probably a non-stationary setting would have been more appropriate especially for capturing the usual tendency of extreme financial returns to occur in clusters.
- In the second part, we wanted to use multivariate extreme value statistics in order to model the dependence between the two series.
- According to PoT approach, we could say that the initial observation of asymptotic dependence between data was confirmed, even if that extremal dependence was rather weak.

Conclusions

- We realized that the estimated extremal dependence was much more stronger in the block maxima method, proabably due to the fact that considering only yearly maxima implies considering the very high extreme returns, which are more dependent instead of the exceedances above the threshold fixed for the PoT models.
- At the end, we can say, that the non-stationary model fitted, especially for the *Health* series, presented some problems. This uncertainty could affect also the bivariate analysis, probably overestimating the actual dependence parameter in the PoT approach. To go further one can try to fit non-stationarity with some other models, for example choosing different covariates.

THANK YOU FOR YOUR ATTENTION