Projet: *Risk, rare events and extremes*

Alessandro La Farciola

January 4, 2023

Contents

Introduction

In this project we are going to analyze 66 years of data of averaged daily returns (units in $\times 100\%$) from January 1950 to December 2015. In particular, the dataset is composed of two series of financial daily returns related to industry portfolios such as *Beer, Liquor* and *Healthcare, Medical Equipment, Pharmaceutical Products* (through the report we refer to the first series as *Beer* and the second as *Health* for brevity). The following plots show the two series.

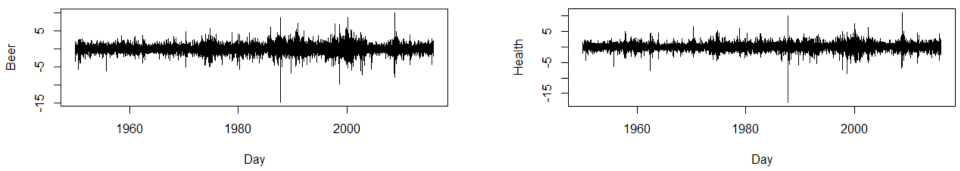


Figure 1.1: Beer and Health industries daily returns (units in $\times 100\%$).

The analysis is divided in two parts. In the first one, the goal of the project is 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. In particular, we try to model with stationary assumption, but our mainly focus is on fitting data in a non-stationary environment.

In the second part, we analyze both series in a bivariate environment using multivariate extreme value statistics techniques in order to model the dependence between the two series. The following figure represents the plot of the vector with the two series as components.

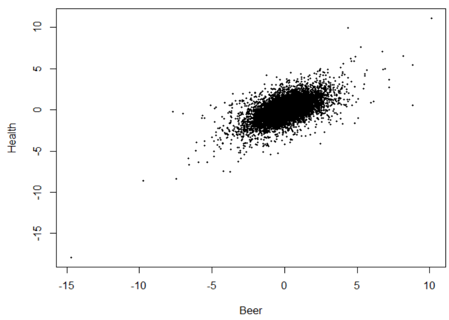
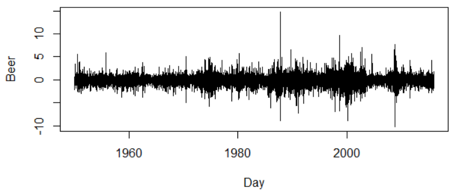


Figure 1.2: Bivariate plot of Beer and Health industries daily returns.

Univariate statistics

In this section, we focus on each series in order to estimate the one-step-ahead conditional quantiles of extreme negative returns. We use a Peaks-over-Thresholds approach and we compare the results using monthly and annual maxima. Since we are interested on the negative returns, the first thing to do is to consider data with opposite sign, both to model exceedances or maxima[1]. The following plots shows data after the previous transformation.



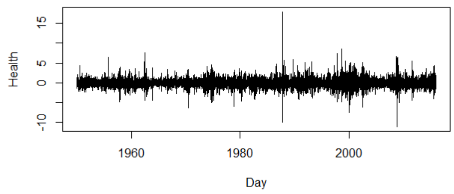


Figure 2.1: Negative Beer and Health industries daily returns (units in $\times 100\%$).

One first look at data highlights the typical high volatility of financial daily returns, with some big peaks, as for example at the end on 1980s and around 2007-2008. This events are well known as consequences of market crashes, or even financial crisis.

2.1 Beer

Now, we consider the *Beer* series.

2.1.1 PoT: stationary model

The first approach that we can have is to ignore the non-stationarity environment, supposing a stationary setting and fitting data with this assumption. In particular, we can model with a Peaks-over-Thresholds (PoT) approach, fitting data over a certain threshold with GPD (Generalized Pareto distribution). One first issue is the choice of the threshold. In order to fix a *good* threshold, we should consider that since the GPD is a limiting distribution, if the GPD is a valid model for the exceedances over a certain threshold u, then for all $v > u$, the GPD also must be an appropriate distribution for the exceedances over v. For this reason, we might expect a certain stability when fitting the GPD to the

[1] From now on, every time we refer to data we, actually, mean data with opposite sign.

exceedances over a sequence of increasing thresholds. We can investigate this stability in two ways:

- *Mean excess*: if $X \sim GPD(\sigma, \xi)$, then $\mathbb{E}(X - u \mid X > u) = (\sigma + \xi u)/(1 - \xi)$. Hence, we should expect a linear relationship between the threshold u and the empirical mean excess, with slope $\xi/(1 - \xi)$.

- *Parameter estimates*: if a distribution $GPD(\tilde{\sigma}, \xi)$ is valid for $X \mid X > u_0$, estimates of parameters should be almost constant with thresholds $u > u_0$.

The following plots shows the *mean residual life plot* and the *parameter stability plots*. The choice of the threshold is crucial, since we should pick a threshold as low as possible in order to have more data available for estimation (decreasing variance), but also need that the GPD remains a valid asymptotic model (ensuring small bias). From figures 2.2 the choice of threshold is not so clear, but they suggest that the quantile at level 0.99, equal to 2.81, can be a reasonable choice.

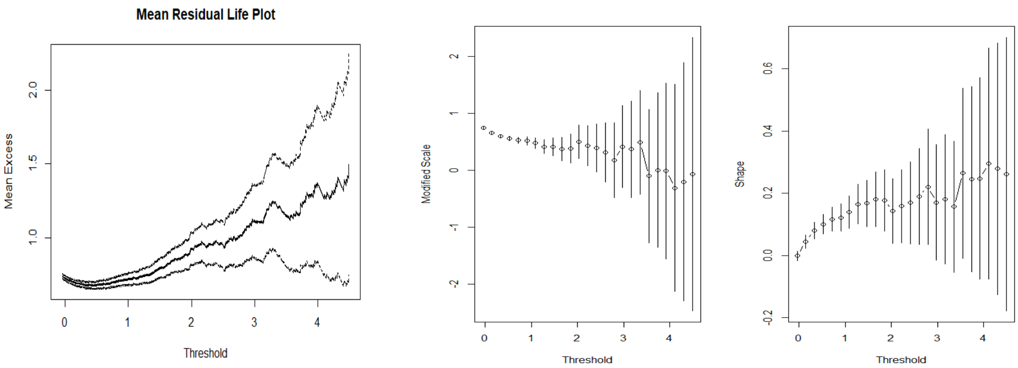


Figure 2.2: Mean residual life plot and parameter stability plot.

We are, now, ready to fit the model and we need to specify the number of observations per year. In our case, we have 16694 observations and 66 years and so we consider throughout this report 253 observations per year. Using *ismev* library in *R* we obtain the following estimates with standard errors in brackets:

$$\hat{\sigma} = 0.82(0.10) \quad \hat{\xi} = 0.21(0.09).$$

Diagnostic plots of the fitting are shown in Figure 2.3.

A first look at normal-based confidence intervals suggests that we can reject the case $\xi = 0$:

$$CI\ for\ \hat{\sigma} : 0.82 \pm 1.96 \times 0.10 = (0.63, 1.02)$$
$$CI\ for\ \hat{\xi} : 0.21 \pm 1.96 \times 0.09 = (0.02, 0.39).$$

This hypothesis is also confirmed by the 95% confidence intervals obtained by the profile likelihood of parameters as we can see in Figure 2.4.

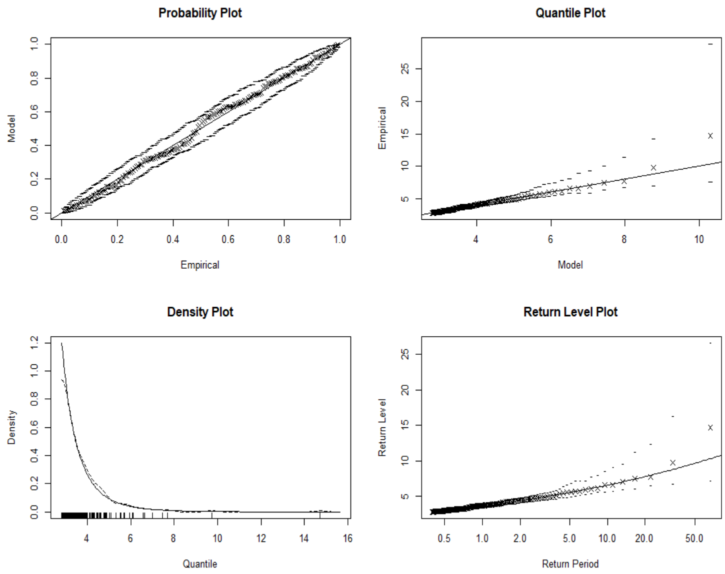


Figure 2.3: Diagnostic plots of stationary PoT model.

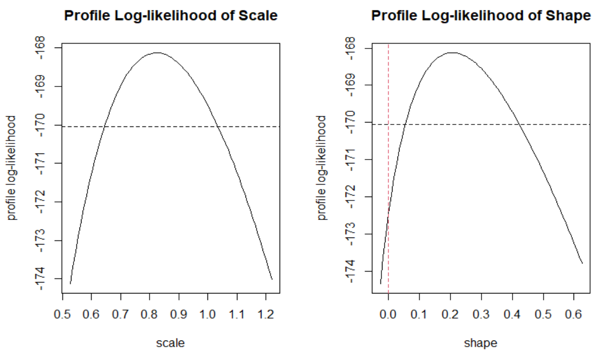


Figure 2.4: Profile likelihood for $\hat{\sigma}$ and $\hat{\xi}$ and their 95%-confidence intervals.

Nevertheless, since the estimation of the ξ parameter is very close to 0, it can be interesting to explore also the Gumbel model with the constraint $\xi = 0$. Diagnostic plots, shown in Figure 2.5, indicate a worse fit since several points in the tail falls outside of confidence intervals of the fitted return level. In addition, also the probability plot suggests a worse fit. Therefore, we prefer using the 2-parameter model, confirming the hypothesis that $\xi > 0$. Moreover, even if we could consider the Gumbel fit as plausible, this does not imply that other models are not. Hence, it is usually preferable to choose a 2-parameter model since accepting more uncertainty about the shape parameter can provide a more realistic quantification of the uncertainty of the model.

Now, it is possible to compute the m-year return level and it is given by

$$x_m = u + \frac{\sigma}{\xi}\{(m\zeta_u)^{\xi} - 1\}, \tag{2.1.1}$$

where ζ_u is the probability of exceeding the threshold u. In our case, we have that for 1 year $m = 253$, 10 years $m = 10 \cdot 253$ and for 100 years $m = 100 \cdot 253$. Hence, applying

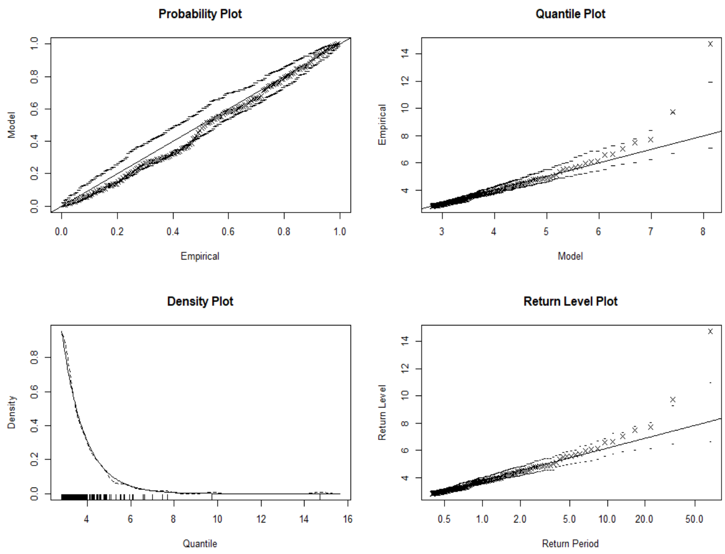


Figure 2.5: Diagnostic plots of Gumbel stationary PoT model.

formula (2.1.1), we obtain that

$$x_1 = 3.65, \qquad x_{10} = 6.58, \qquad x_{100} = 11.31.$$

A plot of the profile likelihood for return level can give information about the confidence interval, showing the typical asymmetry.

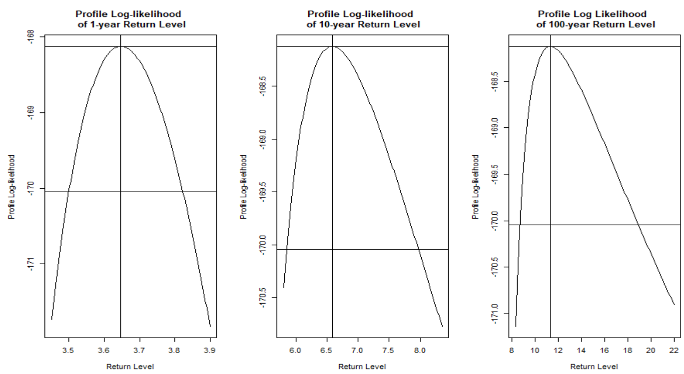


Figure 2.6: Profile likelihood for 1,10,100-years return levels and their 95%-confidence intervals.

From the previous plots it is possible to extrapolate the following CI for return levels:

$$x_1 : (3.50, 3.82) \qquad x_{10} : (5.8, 7.9) \qquad x_{100} : (8.65, 18.9).$$

In order to measure the validity of the previous return levels, one can count the number of observations above the estimated return levels and it turnes out that the 1-year return level is exceeded 72 times, the 10-years return level 6 times and the 100-year return level only once.

2.1.2 PoT: non-stationary models

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. The following Figure shows the exceedances from the previous stationary model and it is clearly possible to note this tendency.

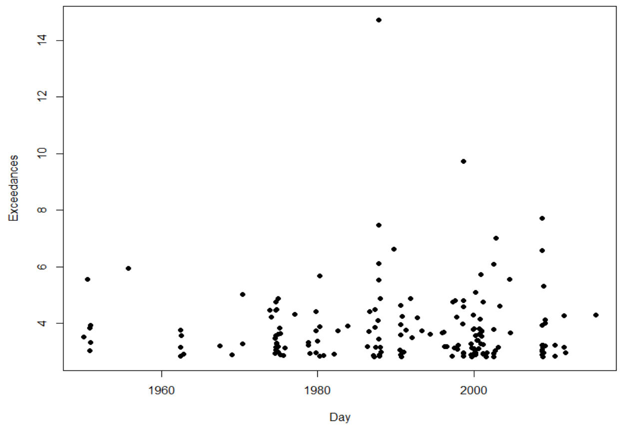


Figure 2.7: Plot of exceedances above threshold according to stationary model

One way to take into account this features can be modelling in a non-stationary environment and this is the main goal of this report. Another limit of the previous model was the small amount of data available for estimation due to the choice of a single threshold. In a non-stationary environment we let the threshold parameter to be varying in time and this can bring more data available for the GPD fitting. In particular, we set a certain time lag l in a way so that the rolling time window spans a certain period. Namely, for each time t, the threshold u_t depends on the previous say month, three months or a year. After choosing a threshold u_t, we should model exceedances using Generalized Pareto distribution.

In the previous setting, we let $Y_t = X_t - u_t$ for $t = l+1, \dots, n$, and let $T_1, \dots, T_{N_u}$ denote the exceedance times. Due to, as we already said, the strong dependence from the most recent past, it seems reasonable to assume a Markov-like structure for the threshold exceedances Y_{T_k}. This means that, calling $\mathcal{F}_t^X$ the σ-algebra generated by observations $X_1, \dots, X_t$,

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

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

$$\sigma_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 u_{t-1}, \qquad t = l+2, \dots, n. \tag{2.1.2}$$

Firstly, we can present the results of the previous model with a rolling window of 1 month. Since we have 253 observations per year, as we mentioned above, this means taking the lag $l = 21$ and setting the threshold u_t to be a high quantile of previous observations up to l. In order to choose the level of quantile for the threshold, we can use the same argument of before about stability plot (see Figure 2.8). In particular, we are plotting the estimated parameters of the model (2.1.2) with the normally-based 95%-confidence intervals with different choices of quantile level, from 0.9 to 0.99. The plots show a fair stability of parameters and this suggest that choosing 0.95 as quantile level can be a reasonable choice.

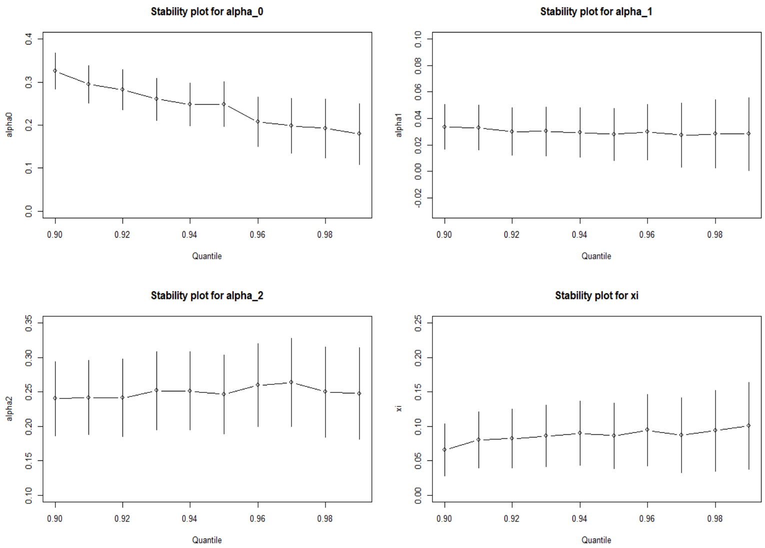


Figure 2.8: Stability plot for parameters in non-stationary model

When we model in a non-stationary environment, another covariate that we can take into account is time, namely we can investigate if there is a linear trend in time. In general, the financial sector is exposed to both rational and irrational human behaviour, causing high volatility in daily returns that usually doesn't depend on time. This is quite clear in series plots, see Figure 1.1 and Figure 2.1. Moreover, if we try to model the scale parameter adding time covariate in (2.1.2), i.e.

$$\sigma_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 u_{t-1} + \alpha_3(t-1),$$

we have that the estimation of the parameter α_3 is of order 10^{-6}, confirming this point.

Now, if we model with GPD distribution the exceedances above the threshold u_t according on what was mentioned before, we end up with the following estimations.

$$\begin{aligned} \hat{\alpha}_0 &= 0.24\ (0.03) \\ \hat{\alpha}_1 &= 0.03\ (0.01) \\ \hat{\alpha}_2 &= 0.25\ (0.03) \\ \hat{\xi} &= 0.09\ (0.02). \end{aligned}$$

Diagnostic plots for the previous model are shown in Figure 2.9.

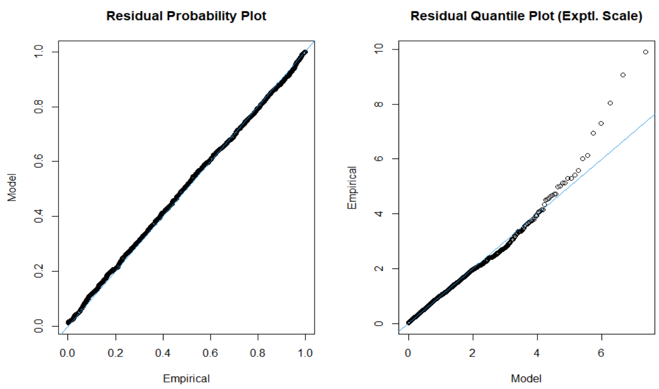


Figure 2.9: Diagnostic plots for non-stationary model with lag $l = 21$

Another way to handle this non-stationarity would be to reparameterize the scale parameter as

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

In this way we don't allow the scale parameter to take negative values. The diagnostic plots for this model are shown in Figure 2.10 and the estimated parameters are the following:

$$\begin{aligned} \hat{\beta}_0 &= -0.61 \ (0.04) \\ \hat{\beta}_1 &= 0.08 \ (0.02) \\ \hat{\beta}_2 &= 0.34 \ (0.04) \\ \hat{\xi} &= 0.08 \ (0.02). \end{aligned}$$

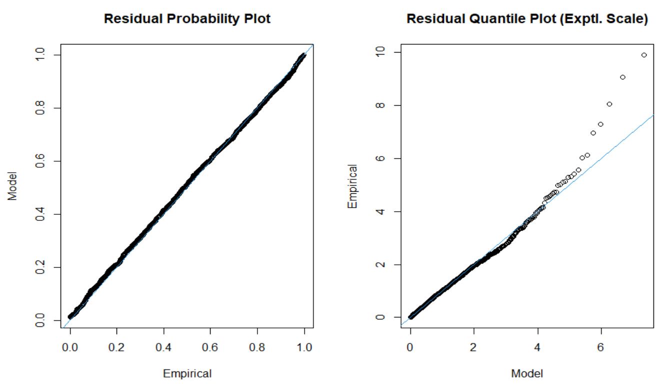


Figure 2.10: Diagnostic plots for non-stationary model with lag $l = 21$ and *siglink*= *exp*

First of all, we can notice the ξ parameter estimations are similar for both models. Moreover, one way to compare them can be considering the *Akaike Information criterion* $AIC = 2\{dim(\theta) - \hat{\ell}\}$. For the first one it is equal to 1116.11, while for the second one it is 1109.6 and this suggests to prefer the exponential one.

A more detailed analysis can be done considering different models with different time lag l. More specifically, we can repeat the previous fitting setting for thresholds u_t with a rolling window of 3 months or a year, namely $l = 21 \cdot 3$ and $l = 21 \cdot 12$.

Tables 2.1 and 2.2 summarize the results for each model.

	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

Table 2.1: Comparison between models

	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

Table 2.2: Comparison between models with *siglink* = *exp*.

As we can see from the previous tables, the best fit according to the *Akaike information crtiterion* is picking $l = 21 \cdot 3$, i.e. a rolling window of 3 months. Figure 2.11 shows the diagnostic plots of the fitting confirming the improvement with respect to Figure 2.10.

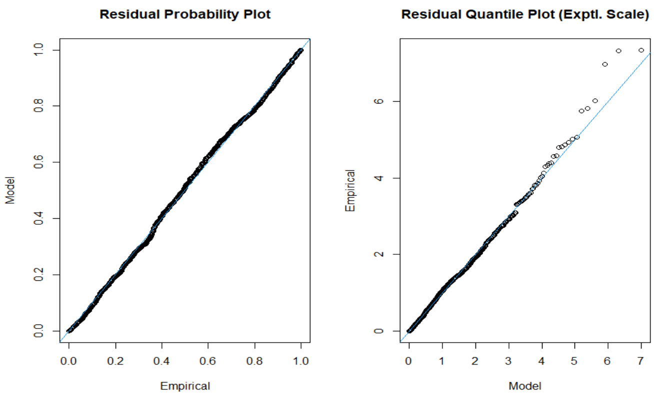


Figure 2.11: Diagnostic plots for non-stationary model with lag $l = 21 \cdot 3$.

Moreover, it is possible to note that the estimations of parameters according to previous models are not far one from each other, especially for the shape parameter, suggesting to consider them reasonable estimations.

Annual and monthly maxima

It is possible to model non-stationarity also using a block maxima approach, extrapolating, for example, the annual 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. Like we did previously, we can model parameters letting them depending on maxima of the previous block, according to

$$\mu_n = \beta_0 + \beta_1 M_{n-1}, \qquad \sigma_n = \alpha_0 + \alpha_1 M_{n-1}. \tag{2.1.3}$$

For the annual maxima model, the estimated parameters are

$$\begin{aligned}
\hat{\beta}_0 &= 2.30\ (0.04)\\
\hat{\beta}_1 &= 0.17\ (0.02)\\
\hat{\alpha}_0 &= -0.17\ (0.04)\\
\hat{\alpha}_1 &= 0.07\ (0.07)\\
\hat{\xi} &= 0.17\ (0.09).
\end{aligned}$$

Moreover, Figure 2.13 shows the diagnostic plots.

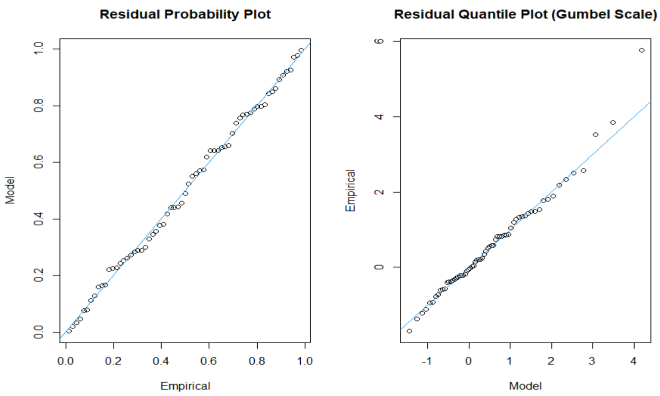


Figure 2.12: Diagnostic plots for non-stationary model annual maxima.

On the other hand, it is possible to model also with monthly maxima, still according to non-stationary model (2.1.3), and obtain the following estimations for parameters:

$$\begin{aligned}
\hat{\beta}_0 &= 0.80\ (0.04)\\
\hat{\beta}_1 &= 0.29\ (0.03)\\
\hat{\alpha}_0 &= 0.25\ (0.04)\\
\hat{\alpha}_1 &= 0.19\ (0.03)\\
\hat{\xi} &= 0.18\ (0.03).
\end{aligned}$$

Figure 2.13 represents diagnostic plots.

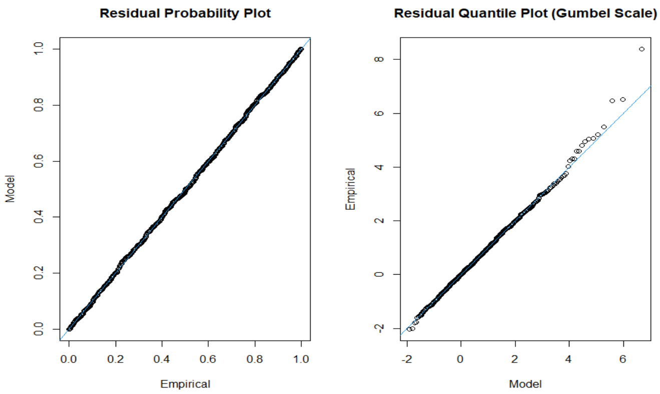


Figure 2.13: Diagnostic plots for non-stationary model monthly maxima.

The AIC information criterion suggest us to consider more reasonable the annual maxima model, even if the diagnostic plots don't seem to make one particularly preferable.

Further considerations about models will be discussed in Section 2.1.4 where we take into account the return levels estimation.

2.1.3 Extremal index

Each of the extreme value models applied so far has been obtained through the assumption that an underlying process consists of a sequence of independent random variables. However, in general, the types of data to which extreme value models are common applied, temporal independence can be an unrealistic assumption. This phenomenon is particularly evident in extreme financial returns and it often shows up in the tendency of extremes returns to occur in clusters. We, briefly, presented this problem at the beginning of Section 2.1.2 and we tried to face it via modelling in a non-stationary environment, assuming a Markov-like structure for the threshold exceedances. Now, we want to discuss a little bit further the asymptotic dependence of extreme negative returns, in particular estimating the extremal index. Indeed, one way of interpreting the extremal index θ is exactly 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, where limiting is in the sense of clusters of exceedances of increasingly high thresholds. With *exiplot* function in *evd* library of *R*, it is possible to estimate the extremal index, plotting it at a sequence of increasing thresholds using the *runs-method*. The result of this plot is showed in Figure 2.14.

As we can see from the previous plot, the estimated extremal index increases with high thresholds as we could expect, since it would mean that higher returns are asymptotically more independent that lower ones. However, it is possible to note that starting from a reasonable choice of threshold 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. Moreover, as 0.9 is close to 1 (for independence), we can say that the asymptotic dependence is rather weak. Furthermore, we can plot the estimated extremal index also for the exceedances after choosing the threshold in the non-stationary environment (see Figure 2.15), where we used the threshold model with lag $l = 21 \cdot 3$. In this case we focus more on the estimated θ as threshold is close to 0 and we can see how the model can

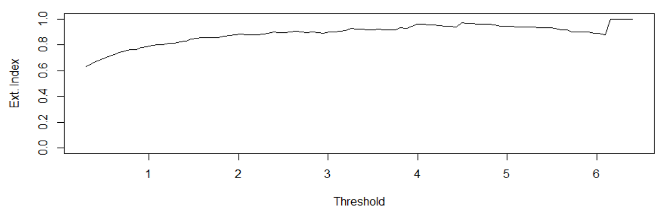


Figure 2.14: Estimated extremal index against threshold (using the runs-method with r = 1).

capture this cluster effect and the estimated parameter is closer to 1 starting from the beginning.

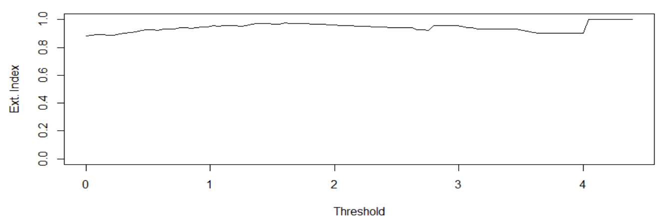


Figure 2.15: Estimated extremal index of exceedances against threshold (using the runs-method with r = 1).

Finally, it is important to remark that, unfortunately, one limit of *runs-method* for estimating θ, is that it is sensitive to the choice of runs r. Of course, a consequence of this problem can be the difficulty of estimating the extremal index and so the possibility of getting to slightly inaccurate conclusions.

2.1.4 Return level

An important issue of extreme theory is the estimation of the high return level. In this section we want to investigate this issue, testing the previous non-stationary models.

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

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

In the last equation, according to our GPD model, we have that for $x \geq u_k$

$$P(X_k \leq x \mid \mathcal{F}_{k-1}^X) = 1 - \frac{N_u}{n} \left(1 + \hat{\xi}\, \frac{x - u_k}{\hat{\sigma_k}}\right)^{-1/\hat{\xi}}. \tag{2.1.4}$$

Now, our goal is to compute the conditional quantiles corresponding to 1,10 and 100-years return levels for each time point $k \in \{l+2, \ldots, n\}$. Firstly, 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

$$\tilde{Y}_{T_k} := Y_{T_k} / \hat{\sigma}_{T_k} \mid \mathcal{F}_{k-1}^X \sim GPD(1, \hat{\xi}). \tag{2.1.5}$$

This scaled variable $\tilde{Y}_{T_k}$ can now be considered as a more stationary series. This is confirmed by the following Figure 2.16 where we are plotting the scaled Beer data according to the transformation (2.1.5).

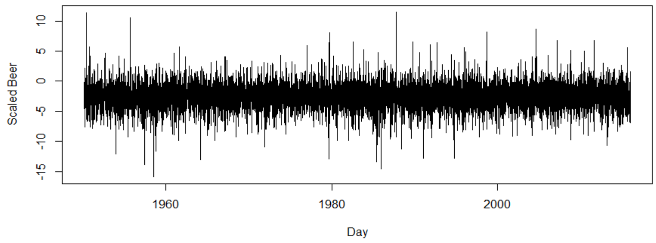


Figure 2.16: Scaled Beer series.

Due to this stationary assumption on $\tilde{Y}$, we can compute the returns level as we usually do in the stationary setting, i.e.

$$x_m = u + \frac{\hat{\sigma}}{\hat{\xi}} \left\{ (m n_y \zeta_u)^{\hat{\xi}} - 1 \right\}, \tag{2.1.6}$$

where $\zeta_u = N_u / n$ represents the probability to exceed the threshold and n_y is the number of observation per year. In our case, if we select a model, say $l = 21 \cdot 3$, we firstly scale data according to (2.1.5), then 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 $\hat{\xi} = 0.11$ (see Table 2.1). Now, we can compute m-observation return level according to equation (2.1.6), for $m = 1, 10, 100$. Figure 2.17 represents the profile likelihood of the scaled model return levels with their 95%-confidence intervals showing the usual asymmetry of such intervals.

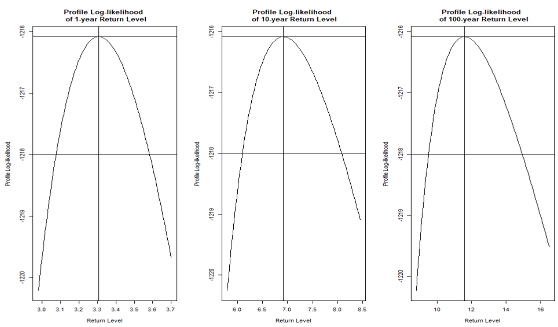


Figure 2.17: Profile likelihood for 1,10,100-years return levels and their 95%-confidence intervals.

From these return levels x_1, x_{10}, x_{100} and their CI $[x_1^{min}, x_1^{max}]$, $[x_{10}^{min}, x_{10}^{max}]$, $[x_{100}^{min}, x_{100}^{max}]$, it is possible to make the back transformation to variable X_{T_k} multiplying every value for $\hat{\sigma}_k$ and adding the threshold u_k, for each time k. 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.

Table 2.3 shows the computed return levels with theirs CI for the stationary series $\tilde{Y}_t$, where we applied the procedure described above for each model presented in section 2.1.2.

	1-year return level	10-year return level	100-year return level
$l = 21$	3.63 [3.40,3.90]	6.98 [6.30,7.90]	11.07 [9.50,13.40]
$l = 21$ *exp*	3.60 [3.38,3.85]	6.80 [6.21,7.75]	10.80 [9.30,13.05]
$l = 21 \cdot 3$	3.30 [3.04,3.64]	6.90 [6.10,8.10]	11.60 [9.50,14.90]
$l = 21 \cdot 3$ *exp*	3.36 [3.10,3.65]	7.11 [6.25,8.35]	12.10 [9.80,15.70]
$l = 21 \cdot 12$	3.30 [3.01,3.60]	7.50 [6.45,8.97]	13.50 [10.70,18.30]
$l = 21 \cdot 12$ *exp*	3.35 [3.08,3.65]	7.75 [6.65,9.35]	14.26 [11.20,19.40]

Table 2.3: Comparison between models estimating return level.

From the results presented in Table 2.3, after the back transformation, 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. In particular, in Table 2.4 the first column represents the number of observations above the *1-year return level - 10-years return level - 100-years return level.* While, the second and third columns represent the same results but for the extreme values of the confidence intervals, respectively the minimum and the maximum point.

	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

Table 2.4: Comparison between models validating return level.

It is possible to use Table 2.4 also for a comparison between the different models. Indeed, since we are validating the estimated return levels for each model and looking at data we should expect that they exceed the 1-year return level about 66 times, since we have 66 years of data available, 10-years return level about 6 or 7 times and 100-years return level once or never.

First of all, we can notice that all models are not terrible at estimating the m-years return level, even if it is possible to say that the $l = 21$ is the "worst" one, especially estimating the 100-years return level, probably underestimating the level of exceedances. On the other hand, $l = 21 \cdot 3$ seems to be the "best" (consistently with what was said in Section 2.1.2). Figure 2.18 shows an example on how the return level and their confidence intervals appear with respect to data[2] and the chosen threshold. The plot represents a window of 1 year, in particular 1998, where we can notice an observation in September exceeding the 10-year return level but not the upper confidence interval.

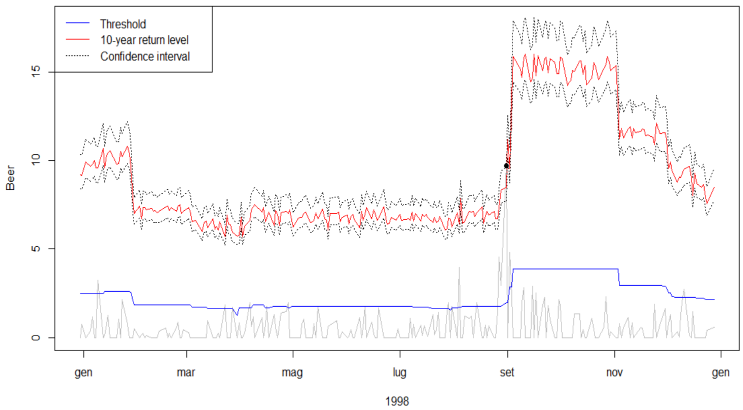


Figure 2.18: 10-years return level with confidence interval for year 1998

Block maxima approach

Now, we can try to estimate the 10-years and 100-years observation return levels with block maxima models discussed previously. Also in this case, as we did for PoT approach, we need to transform data in order to obtain a stationary series, compute return levels with uncertainty, then come back to the original series and validate them. Unlike before, we are modelling maxima M_n with distribution $GEV(\eta_n, \tau_n, \xi)$ and in order to scale the series we apply the transformation $\tilde{M}_n = (M_n - \eta_n)/\tau_n$. Now, if we fit a GEV model to the maxima $\tilde{M}$, we can see, as expected, that the estimation of location parameter is almost 0 (order of 10^{-6}), the scale parameter is approximately 1 and the shape parameter is 0.17 for the annual maxima and 0.18 for the monthly maxima (same estimations of before). Table 2.5 shows the computed return levels with their 95%-confidence intervals estimated with profile likelihood for the scaled maxima. It is possible to note a remarkable difference between the estimated return levels for the two models but this is due to the non trivial difference between estimated parameters between the models (see previous Section).

	10-year return level	100-year return level
Annual Maxima	2.72 [2.40,3.90]	6.83 [6.01,12.90]
Monthly Maxima	7.66 [6.60,9.10]	14.57 [11.70,19.10]

Table 2.5: Comparison between block maxima models estimating return level

On the other hand, in Table 2.6 it is possible to find the result of the 10-years and 100-years return level validation, after back transformation, as we did for the PoT models.

[2]In order to make a clearer plot, instead of plotting data we plotted the maximum between them and 0. So the grey lines represents only the negative returns.

We can notice that, even in this case, annual maxima model seems to perform a little bit better than the monthly maxima one.

	x_m validation	x_m^{min} validation	x_m^{max} validation
Annual Maxima	6 - 1	11 - 3	3 - 0
Monthly Maxima	9 - 1	11 - 3	4 - 1

Table 2.6: Comparison between models validating return level

2.2 Health

Following the same approach used for the *Beer* series, in this section we present the results also for the *Health* series. In particular, we focus on non-stationary models.

Regarding the Peaks-over-Threshold approach, we try the same models presented in the previous section and Tables 2.7 - 2.10 show the results obtained. In particular, we can find estimated parameters, AIC values, estimated 1,10,100-years return levels for scaled series and validation of such return levels after back transformation. We recall that, fix a certain lag value l, for models in Tables 2.7-2.8 we let the scale parameter to be

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

While, for models in Tables 2.9-2.10 we let scale parameters to be

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

	α_0	α_1	α_2	ξ	AIC
$l = 21$	0.33 (0.02)	0.04 (0.02)	0.18 (0.03)	0.05 (0.02)	1161.43
$l = 21 \cdot 3$	0.19 (0.04)	0.05 (0.01)	0.25 (0.04)	0.10 (0.03)	922.20
$l = 21 \cdot 12$	0.07 (0.07)	0.05 (0.01)	0.36 (0.06)	0.13 (0.03)	1043.08

Table 2.7: Comparison between models.

	1,10,100- years Return levels	x_m validation
$l = 21$	3.47 - 6.39 - 9.70	46 - 9 - 4
$l = 21 \cdot 3$	3.32 - 6.80 -11.20	55 - 8 - 3
$l = 21 \cdot 12$	3.20 - 7.03 - 12.23	53 - 9 - 2

Table 2.8: Comparison between models in estimating return levels.

	β_0	β_1	β_2	ξ	AIC
$l = 21$	-1.03 (0.05)	0.08 (0.03)	0.33 (0.04)	0.05 (0.02)	1155.05
$l = 21 \cdot 3$	-1.24 (0.08)	0.11 (0.03)	0.41 (0.06)	0.10 (0.03)	923.67
$l = 21 \cdot 12$	-1.36 (0.12)	0.11 (0.03)	0.53 (0.08)	0.13 (0.03)	1042.56

Table 2.9: Comparison between models with *siglink* = *exp*.

	1,10,100- years Return levels	x_m validation
$l = 21$	3.43 - 6.26 - 9.42	47 - 9 - 4
$l = 21 \cdot 3$	3.31 - 6.75 - 11.10	53 - 9 - 3
$l = 21 \cdot 12$	3.20 - 7.05 - 12.30	53 - 9 - 2

Table 2.10: Comparison between models with *siglink* = *exp* in estimating return levels.

From the *AIC* criterion, as for *Beer* series, the model with lag $l = 21 \cdot 3$, i.e. a rolling window that spans a period of 3 months, seems to be the "best". However, return levels validation suggest that the fit is not so good. This is also confirmed by the diagnostic plots (see Figure 2.19).

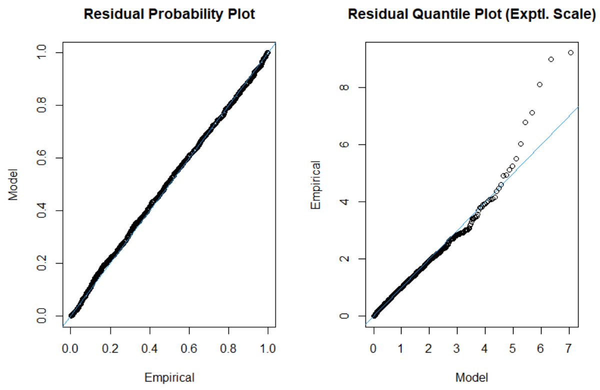


Figure 2.19: Diagnostic plots for non-stationary PoT model with lag $l = 21$, Health series.

Now, regarding the estimation of the extremal index, Figure 2.20, shows above the estimated extremal index against thresholds for the original series. In this case, as for *Beer* series, starting from a reasonable choice of threshold 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. Moreover, as 0.9 is close to 1, we can say that the asymptotic dependence is rather weak. While, the second plot shows the estimated extremal index for the exceedances after choosing the threshold in the non-stationary environment according to the pervious model. Also in this case we can see how the model can capture this cluster effect and the estimated parameter is closer to 1 starting from the beginning.

Finally, regarding block maxima approach, we can model annual maxima and monthly maxima with GEV distribution. In particular, as before, we let the location and the scale parameter varying in time according to the following expressions:

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

Table 2.11 shows the estimated parameters and the 10,100-years return level validation according to the procedure explained in the previous Section.

	β_0	β_1	α_1	α_2	ξ	Ret. lev.
Annual maxima	1.8 (0.30)	0.30 (0.09)	0.32 (0.30)	0.18 (0.10)	0.28 (0.11)	5 - 1
Monthly maxima	0.80 (0.05)	0.30 (0.03)	0.38 (0.04)	0.14 (0.02)	0.13 (0.02)	9 - 2

Table 2.11: Comparison between block maxima models, Health series.

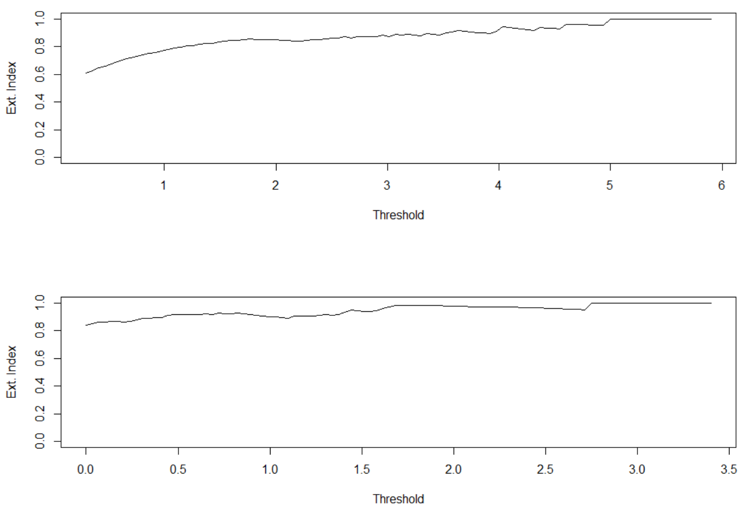


Figure 2.20: Estimated extremal against threshold (using the runs-method with r = 1), Health series.

From these results, it seems that annual maxima can be considered a more reasonable block maxima model and Figure 2.21 shows the diagnostic plots.

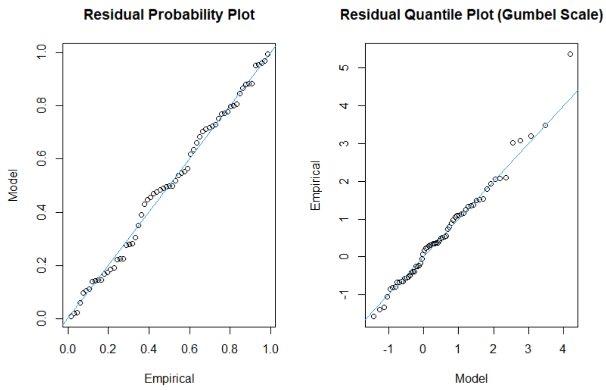


Figure 2.21: Diagnostic plots for non-stationary annual maxima model, Health series.

Multivariate statistics

In this second part of the report, we want to model the dependence between the two series *Beer* and *Health* using multivariate extreme value statistics.

A first quick look at data (see Figure 1.1 and 1.2) suggests that we can distinguish some dependence between the two time series. This is supported also by the correlation parameter that is 0.59 and by the fact that both data are expression of the same financial market and so we expect that high negative returns may come from the same events. Of course this remarks are useful to have a first impression on dependence, but it is not clear whether or not there is really "extremal dependence". Our goal is trying to quantify this extremal dependence in order to understand how deep is the link between the two portfolios in analysis.

3.1 PoT approach

In case of a POT approach, as we did for computing return levels, we need to consider the scaled series that are now stationary series. More specifically, we recall 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

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

Therefore, in order to model bivariate extremes in a non-stationary setting, we firstly consider the best non-stationary model for each marginal (as discussed in previous sections) and then we scale the series according to the previous definition with the respectively estimated parameters. After fitting stationary models to the scaled series with threshold $u = 0$, we then apply marginal transformations to unit Fréchet marginal distributions. In particular we fit *GPDs* above these thresholds, giving fitted distribution

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

for each $d = 1, 2$, based on the fitted parameters $(\hat{\sigma}_d, \hat{\xi}_d)$. Now, we can apply the following component-wise transformation for variables

$$z_j = -1/\log\{\hat{F}_d(x_j)\} \quad for\ j = 1, \ldots, n,$$

that have approximate unit Fréchet marginal distribution. After the previous transformation to standard Fréchet variables, the two series are plotted on logarithmic scale in Figure 3.1, where also the transformed thresholds are added. Since we have one threshold for each variable, when each component exceeds, the respective observation lies in the region at top-right of the plot.

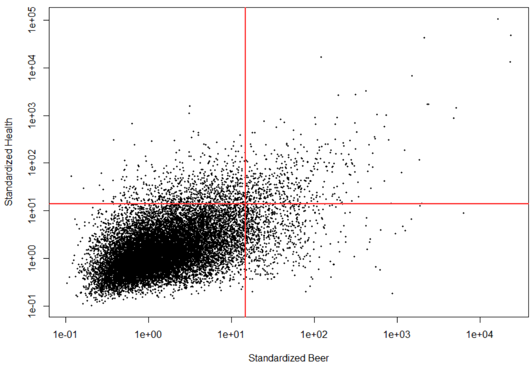


Figure 3.1: Bivariate plot of Beer and Health industries daily returns after transformation to unit Fréchet scale.

A first look at the previous plot suggests us the existence of dependency at high extremes between the two series. However, the occurrence of several observations in regions at bottom-right and top-left of the quadrant may indicate that the extremal dependence is not so strong.

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

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

Fitting with logistic dependence structure with threshold likelihood, we get an estimated dependence parameter equal to

$$\hat{\alpha} = 0.726\ (0.011),$$

implying that a normally-based confidence interval for $\hat{\alpha}$ is $(0.71, 0.75)$. We can notice that the standard error, in this case, is particularly small, giving a short confidence interval for the dependence estimation. Moreover, since 1 is outside the confidence band, we can conclude that there is extremal dependence, even if it may be not very strong. Now, comparing this estimation also with other models, we would like to give a more precise measure of this dependence. Figure 3.2 shows diagnostic plots for the previous fit, where the top-right plot represents Pickands dependence function $A(t)$. We recall that $A(t) = 1$ corresponds to asymptotic independence and $A(t) = \max\{t, 1-t\}$ corresponds to complete dependence.

The bottom-right plot represents the spectral density, consistently with the estimated dependence parameter, this model suggest a not very strong asymptotic dependence. Moreover, the estimated marginal parameters are

$$\hat{\sigma}_1 = 1.02(0.04), \quad \hat{\xi}_1 = 0.09(0.03), \quad \hat{\sigma}_2 = 1.01(0.04), \quad \hat{\xi}_2 = 0.10(0.02).$$

The estimated marginal parameters are close to what we expected since we scaled data with the non-stationary estimated parameter of Section 2.1.2 and, in addition, it is possible to note that the shape parameters are really close one to each other. This suggest us

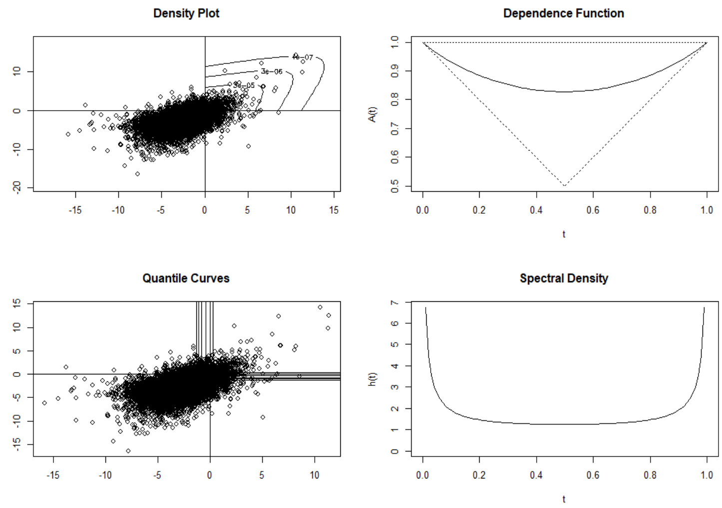


Figure 3.2: Diagnostic plots for logistic models.

to try to fit a simpler model with a common shape parameter. In this case the estimations are the following:

$$\hat{\sigma}_1 = 1.02(0.04), \quad \hat{\sigma}_2 = 1.01(0.04), \quad \hat{\xi}_2 = 0.10(0.02) \quad \hat{\alpha} = 0.726(0.011).$$

First of all, we can notice that the estimated parameter are really close to the previous ones, especially for the dependence parameter. Furthermore, since the two models are nested, a likelihood ratio test can give us information about the improvement of the fit. The difference between the number of the two parameters is 1 and so we compare the difference between the deviance of the two models and the 95%-quantile of the χ^2-distribution. The first value is 0.02, significantly small with respect to 3.84. This means that we cannot reject the null hypothesis - common shape parameter - confirming the equivalence of the two models. For this reasons, in this case, it is preferable to fit with a simpler model, i.e. common shape.

A second model that we can try to fit is the asymmetric logistic one, in order to see the difference with the previous one. Also in this case, we can compare the model with common shape parameter and the one with different ones. For the all parameters model the estimated dependence parameter is

$$\hat{\alpha} = 0.692 \ (0.032),$$

and the asymmetry parameter estimates are:

$$\hat{\phi}_1 = 0.868(0.083), \quad \hat{\phi}_2 = 0.937(0.103).$$

While, for the model with common shape parameter we have

$$\hat{\alpha} = 0.693 \ (0.032), \quad \hat{\phi}_1 = 0.87(0.08), \quad \hat{\phi}_2 = 0.938(0.103).$$

Firstly, we can note that the estimations are very close and, as before, a likelihood ratio test (0.08) can confirm that the common shape parameter is a preferable model. Now, since the symmetric and the asymmetric models are nested, also in the case a likelihood ratio test can help us choosing the best fit. Also in this case, the deviance statistics (2.39) is less than the 95%-quantile of the χ^2-distribution (5.99). This suggest that we can't reject the symmetric hypothesis in favor of asymmetry. Therefore, the asymmetric model cannot be considered significantly better than the logistic one and so we prefer the parsimonious (symmetric) logistic model.

Table 3.1 shows the *AIC* value for different models used to fit data. Based on the *Akaike infromation criterion* the logistic model, presented before, appear to be the "best" one for this dataset.

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

Table 3.1: Comparison between different models

3.2 Yearly Maxima approach

Now, we can compare the previous bivariate extreme analysis with a yearly maxima approach. In this case, we extrapolate yearly maxima for both series, fit GEV marginals and estimate the dependence parameters. As before, we can model with different joint distribution functions. Figure 3.3 shows diagnostic plots for logistic model, with whom we obtain an estimation for the dependence parameter equal to

$$\hat{\alpha} = 0.537 \ (0.063),$$

with normally-based CI for equal to $(0.41, 0.66)$. Unlike PoT approach, we can notice that the standard error for the dependence parameter is bigger and this cause a bigger uncertainty in the estimation. However, in this case, the asymptotic dependence seem to appear much stronger than before. This observation is also confirmed by the spectral density that presents a huge curvature at around 0.5.

Now, looking at the estimated Pickands dependence function, it seems to depart a little bit from the symmetric logistic function. This suggests to try modelling with an asymmetric distribution function. If we model with the asymmetric logistic, we obtain an estimation of dependence parameter equal to

$$\hat{\alpha} = 0.488 \ (0.087).$$

As before, a likelihood ratio test (deviance statistics equal to 0.4) suggest us that the symmetric model can't be rejected and so it is preferable to fit with the parsimonious (symmetric) model.

We can try to apply the same different models as we did above for PoT and we obtain the *AIC* values according to Table 3.1. Even in this case, from these results, we can select as "best" model the logistic one.

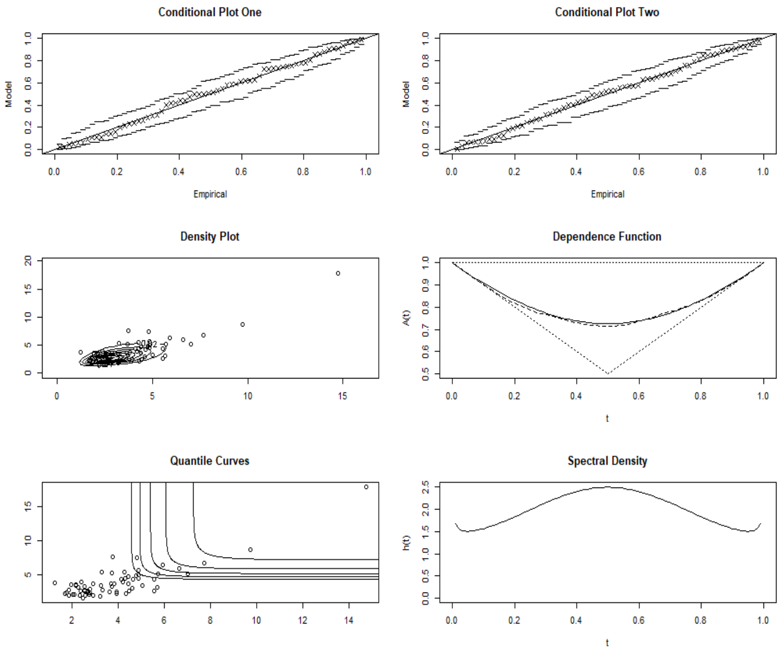


Figure 3.3: Diagnostic plots for logistic models (yearly maxima approach.

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 3.2: Comparison between different models

Finally, we can say that probably the reason why the annual maxima approach gives an estimation of extremal dependency much stronger than PoT approach is that considering only yearly maxima we are considering the very high extreme returns. Indeed, as we can from Figure 3.4, the annual maxima series looks much more correlated (correlation parameter is 0.84) than the whole series. Therefore, it is reasonable to expect stronger asymptotic dependence for very extreme values, since they can be results of the same big market crashes, or even global financial crises that involve different investment areas.

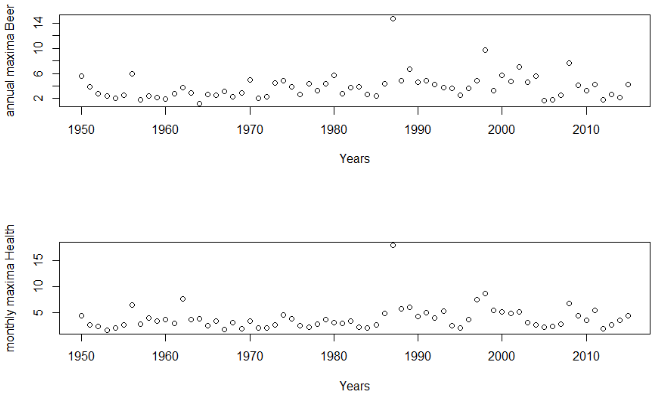


Figure 3.4: Annual maxima of Beer and Health series.

Conclusion

In this report we wanted to analyze 66 years of data of averaged financial daily returns (units in $\times 100\%$) and in particular of two specifics portfolios, one related to *Beer and Liquor* industry and the other related to *Healthcare and Pharmaceutical products*. The analysis was divided in two parts.

In the first one, 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. In particular, we firstly 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. For this reason, we tried to model letting the scale parameter for GPD (PoT appraoch) and the location and scale parameter for GEV (Block maxima approach) to be varying in time. More specifically, in the first case, we set a certain time lag l in a way so that the rolling time window spans a certain period, say 3 months, and then we set the threshold u_t to be a high quantile of previous observations up to l. Hence, we modeled some of the distribution parameters depending on previous data and previous threshold, assuming a Markov-like structure. Therefore, we presented different results, such as estimated parameters, AIC values, and we discussed how to estimate the 1,10,100-years return levels with their uncertainty for each model. In particular, we saw that non-stationary models were better fit for the *Beer* series than the *Health* one. Moreover, we noticed that in both case a rolling window of 3 months was a preferable model. In addition, we discussed also how we could interpret the cluster phenomenon through the estimation of the extremal index, for both series.

In the second part of the report, we wanted to use multivariate extreme value statistics in order to model the dependence between the two series. In particular, we used different extremal models in order to estimate the possible dependence between the extremal negative daily returns of both portfolios. As we did for the univariate case, also here we compared the results between PoT approach and Annual maxima approach. In conclusion, 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. Actually, 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. However, in both cases we discussed different models, trying to understand whether a symmetric model was sufficient for the fitting or it was necessary to fit with an asymmetric one.

At the end, we can say, as we already noticed, that the non-stationary model fitted in this report, especially for the *Health* series, presented some problems. While, regarding block maxima approach, we could say that, especially for the annual maxima, probably the extrapolated maxima looks more stationary series and so this could bring to some models that are more consistent. This uncertainty could affect also the bivariate analysis, probably overestimating the actual dependence parameter in the PoT approach. To solve this possible problems one can try to fit non-stationarity with other models, for example choosing different covariates, and compare the results with those we found in this report to have a more detailed overview of the estimations.

Bibliography

[1] Davison A., (2021), Lecture notes of the course *Risk, Rare events and extremes*.

[2] Coles S., (2001) *An Introduction to Statistical Modelling of Extreme Values*, London: Springer.

[3] Finkenstädt and Rootzén, (2004) *Extreme Values in Finance, Telecommunications and the Environment*, CRC.