
Project

The financial sector is exposed to both rational and irrational human behaviour, with the latter often being the cause of market crashes, or even financial crises. Two such instances were the dot-com bubble, from the end of the 1990s until the beginning of the new millennium, and the 2007–2008 financial crisis. Such events can leave human society vulnerable to many adverse consequences. The goal of this project is to estimate the one-step-ahead conditional quantiles of extreme negative returns at 1-year, 10-year and 100-year return levels in a non-stationary environment, based on around 66 years of data of averaged daily returns (units in $\times 100\%$) of 30 industry portfolios, from January 1950 to December 2015. The data analysis project will involve using various techniques from the course to estimate some of these risks.

You will work in pairs, so find a partner and tell us who you are working with; you will then have to choose the two series you are going to work with from `Project2021.txt`, which you can save by using for instance `data2021<-read.table(file="Project2021.txt", sep='')`. A second file, `Siccodes30.txt`, contains information on the code-names that are used, and also regarding the industry sectors that compose every portfolio.

The data series will be assigned on a first-come-first-served basis. For example, if a group is first to select the Finance (Fin) and (Oil) series, then neither of them can be picked by a second group.

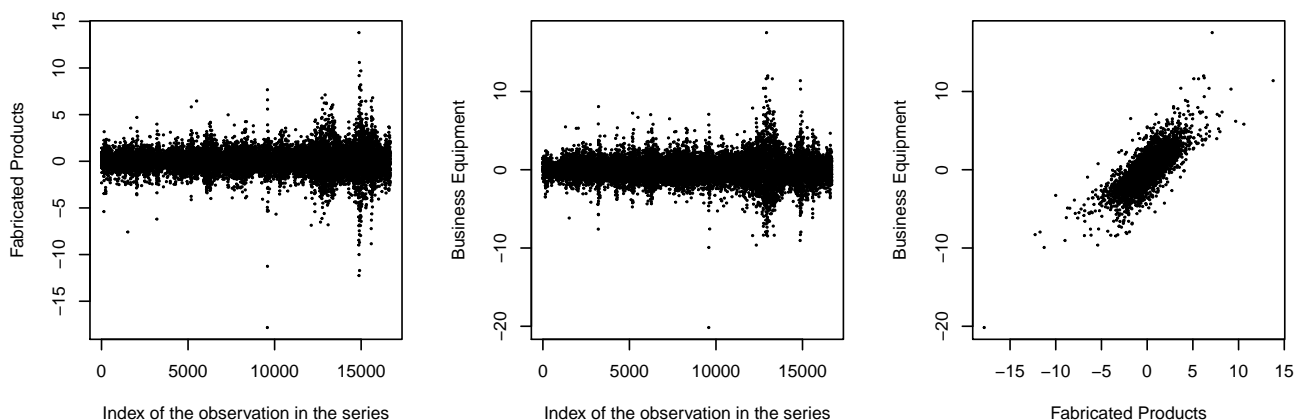


Figure 1: Time series of daily returns (%) from two selected portfolios.

Final deadline

By 17:00 on January 21, 2022 *at the latest*, send the PDF file containing the finished report `YourNameReport.pdf` and the txt file `YourNameCode.txt` containing the R code to `Mario.Krali@epfl.ch`. The code should be ready to run, and will be used to check that your results are honest, and that there has been no sharing of code.

You may discuss your work with other students, but the report and the code should be all your own work. Any signs of plagiarism or code-sharing will be heavily penalized.

Data

The file `Project2021.txt` contains time series of averaged daily returns of 30 industry portfolios compiled as part of the Kenneth French Data Library. Each portfolio is an aggregate measure composed of sectors within a specific industry; see the file `Siccodes30.txt`. Each group will be asked to analyse two series. After saving the data, using for instance `data2021<-read.table(file="Project2021.txt", sep='')`,

you can obtain a summary by typing `summary(data2021)`. There are 30 columns in total. Every column corresponds to a series of portfolio returns. You can also extract the date from the row names by using `date2021<-as.Date(rownames(data2021), "%Y%m%d")`.

Goal

The goal of your project is to estimate the one-step-ahead conditional quantiles of negative returns corresponding to 1-, 10-, and 100-year return levels in a non-stationary environment. With 66 years of daily data we suggest using a Peaks-over-Thresholds approach, though it would be wise to compare this with the results from using monthly or annual maxima. Non-stationarity could be modeled using regression techniques.

Non-stationarity

You could model the exceedances by using a Peaks-over-Thresholds (POT) approach. However, you must be careful when selecting exceedances in a non-stationary setting. In order to account for non-stationarity, you could set the threshold parameter to be varying in time, namely you could set u_t to be a high empirical quantile of previous observations up to a certain time lag, say l . One example would be to set l in a way so that the rolling time window spans a certain period, for instance, ranging from a month to three months, or six months, or more. For a high threshold u_t , it should then be feasible to model exceedances using a generalized Pareto distribution.

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. As shown in Figure 1, the financial market is characterised by periods of both stable returns, as well as high volatility which tends to occur in clusters. It seems reasonable to assume that the threshold exceedances Y_{T_k} possess Markov-like structure. This would allow one to capture the changing volatility in returns, for instance via the scale parameter. More specifically, letting \mathcal{F}_t^X denote the σ -algebra generated by X_1, \dots, X_t , i.e., the σ -algebra containing all past information, this implies that

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

You may try to model σ_{T_k} by using as covariates $X_{T_k-1}, Y_{T_k-1}, u_{T_k-1}$, by combining them, or by using other covariates from the past. One example would be to model the scale as

$$\sigma_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 Y_{t-1} + \alpha_3 u_{t-1}, \quad t = l + 2, \dots, n. \quad (2)$$

Non-stationary models can be fitted using the function `gpd.fit` (in the library `ismev`) using appropriate covariates (see the options `ydat`, `sigl`, `siglink`, etc.). You could then use AIC or other information criteria, aided by residual plots, to help choose a parsimonious model that fits the data adequately.

In a block-maxima setting you may also model non-stationarity using a GEV distribution with regression forms for its parameters, namely, location, scaling, and shape similar to eq. (2). As potential covariates you might try using a variable from the past similar to u_k from the POT framework, maxima of previous periods, or other covariates that may quantify high volatility.

Non-stationary models could be fitted using the function `gev.fit` (in the library `ismev`) with appropriate covariates (see the options `ydat`, `mul`, `mulink`, `muinit`, etc.). Similar to the POT case, the use of AIC or other information criteria, aided by residual plots, may help choose a suitable model.

Extremal index

Extreme financial returns tend to occur in clusters. Once you have fitted a model for the exceedances, it may be wise to estimate the extremal-index, to check whether they are asymptotically dependent. You may use the `exiplot`. Does it look stable? What can you conclude about the asymptotic dependence of extreme events?

Quantiles

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

$$Q_k(p) = \inf\{x \in \mathbb{R} : P(X_k \leq x \mid \mathcal{F}_{k-1}^X) \geq p\}, \quad (3)$$

where we recall that for $x \geq u_k$,

$$P(X_k \leq x \mid \mathcal{F}_{k-1}^X) = \left(1 - \frac{N_u}{n} (1 + \hat{\xi} \frac{x - u_k}{\hat{\sigma}_k})^{-1/\hat{\xi}}\right) \mathbb{1}\{x \geq u_k\}. \quad (4)$$

Once you compute the conditional quantiles corresponding to 1, 10, and 100-years return levels, for all time points $k \in \{l+2, \dots, n\}$, you may validate them by counting the number of observations above the one-step-ahead estimated quantiles. Can you think of ways to account for uncertainty in the estimated quantiles at each time point k ?

Now, for the moment, assume that maxima are independent across the years, and estimate the 10, and 100-years return levels using the fitted block-maxima model. Given the level of uncertainty over the time-periods, it might be informative to have confidence intervals for such quantiles. Compare the estimated return levels from the block-maxima approach with the conditional quantiles obtained above using eq.(4).

Modelling pairwise dependence in the extremes

In a second part of the project you should use multivariate extreme value statistics to model the dependence between the two series.

First, fit the same type of model to the comparison portfolio to find the best non-stationary model. Compare between the models you find for the two series. Could we assume the same shape parameter ξ for the two series?

Second, model the dependence between the extremes of the two series. Use the extremal models discussed in the lecture to estimate the possible dependence between the extremal negative daily returns of both portfolios.

In case of a POT approach, the threshold likelihood would be more appropriate. Since the `fbvpot` function cannot model non-stationarity with covariates, it may be useful to note the following: if $Y_{T_k} = X_{T_k} - u_{T_k} \mid \mathcal{F}_{T_k-1}^X \sim \text{GPD}(\hat{\sigma}_{T_k}, \hat{\xi})$, then $\tilde{Y}_{T_k} = Y_{T_k}/\hat{\sigma}_{T_k} \mid \mathcal{F}_{T_k-1}^X \sim \text{GPD}(1, \hat{\xi})$. The transformed \tilde{Y}_{T_k} are now a stationary series.

You may also compare the extremal dependence between yearly maxima and see if there are differences between the two approaches.

You can also fit a Gaussian model (Gaussian copula) to compare with the fit of the extremal model. Discuss the fit of these models.

Scope for initiative

Here are some ideas for things you could add to the project above:

- comparison with analysis of monthly/annual maxima. Are their results similar to those for the GPD analysis? What special problems arise?
- discussion of the uncertainty of your estimated quantiles. What components of the uncertainty are there, and how can you accommodate those that are missing?

If you want to try something else, check with us that it would be reasonable, and have a go!