

Financial Risk

Lecture 4

Extremal Index

- So far we have assumed that we have i.i.d. data which is not a correct assumption
- Typically big losses come in (volatility) clusters
- It can be shown for a stationary sequence $\{r_t\}$ satisfying a condition of sufficiently fast decay of long range dependence of exceedance clusters and an i.i.d. series $\{\tilde{r}_t\}$ with the same marginal distribution as $\{r_t\}$ that, if we have convergence

$$P\left(\frac{M_n - \beta_n}{\alpha_n} \leq x\right) \rightarrow F_*(x) = (\tilde{F}_*(x))^\theta$$

then $F_*(x)$ is GEV and can be expressed as a power $0 < \theta \leq 1$ of the GEV towards which the maximum of the i.i.d. sequence converges

- We call θ the extremal index and it may be interpreted as the reciprocal of the mean cluster length

Extremal Index (GEV)

- Since

$$\tilde{F}_*(x) = \exp \left[- \left(1 + \xi \frac{x - \beta}{\alpha} \right)^{-1/\xi} \right]$$

we get (exercise)

$$F_*(x) = (\tilde{F}_*(x))^\theta = \exp \left[- \left(1 + \xi_* \frac{x - \beta_*}{\alpha_*} \right)^{-1/\xi_*} \right]$$

where $\xi_* = \xi$, $\alpha_* = \alpha\theta^\xi$ and $\beta_* = \beta - \alpha(1 - \theta^\xi)/\xi$

In practice

- We will use some declustering techniques to estimate the extremal index and use cluster maxima (byproduct) for PoT
- Block method; divide observations into blocks of fixed length. All values in a block that exceed the threshold u is a cluster
- Block-runs method; The first cluster starts at the first exceedance of u and contains all exceedances within a fixed length r thereafter, and so on
- Runs method; The first cluster starts with the first exceedance of u and stops as soon as there is a value below u , the next starts with the next exceedance of u and so forth

In practice

- We get an estimate of the extremal index as

$$\hat{\theta} = \frac{\#clusters}{N(u)}$$

- We may use the cluster maxima for PoT and GPD for the stationary series
- The block sizes and thresholds used for declustering are not necessarily related to the ones used to give the best GEV and GP fits.

GEV VaR for a stationary series

- So we may estimate scale, location and shape for the GEV as if data were i.i.d. (i.e. using block maxima) and then adjust parameter estimates using our estimate of the extremal index

$$VaR_p = \begin{cases} \beta_n - \frac{\alpha_n}{\xi_n} \{1 - [-n\theta \ln(p)]^{-\xi_n}\} \\ \beta_n - \alpha_n \ln[-n\theta \ln(p)] \end{cases}$$

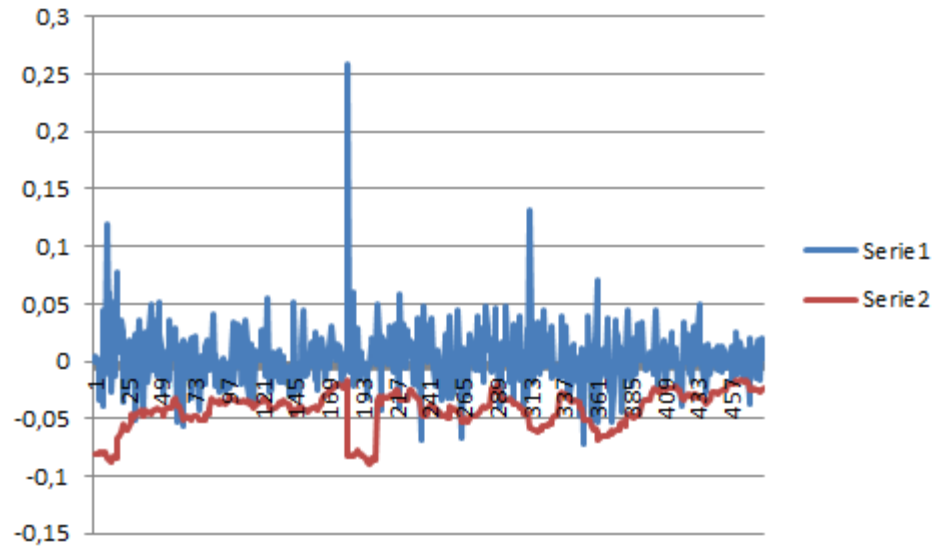
Decile 9

- Using block declustering with block size 10 and $u = 0.08$ yields $\hat{\theta} = 0.77$
- As a check that the declustering is reasonable one may check the fit of the GPD to the cluster maxima (minus threshold)
- Using $\hat{\theta} = 0.77$ in the above ($\xi_n = 0$) VaR formula gives the 95% and 99% VaR for the decile9 data; 0.0838 and 0.1270
- We note that these numbers higher than the ones given without using the extremal index.

FB-data in Excel

- As an example we use the block declustering method with 30-day blocks gives the extremal index 0.542 as there are 13 clusters and 24 exceedances of the threshold 1.5 (there are approximately 5% standardized negated returns above 1.5)
- Using 30-day block maxima and plugging in the estimated extremal index gave a 95% quantile for the standardized negated returns of 1.743
- The choice of 30-day block declustering was made after some tweaking...

FB-data in Excel



There are 18 violations or exceedances of VaR in 480 observations giving a percentage of approximately 4 which is a slight over-estimation of the VaR

PoT VaR for stationary series

- PoT for stationary series is done by replacing threshold exceedances (minus threshold) by cluster maxima (minus threshold)
- Also the number of exceedances $N(\eta)$ should be replaced by the number of clusters in the formulas for quantiles and VaR
- Letting N_c denote the number of clusters, we have (for the stationary series)

$$VaR_p = \begin{cases} \eta - \frac{\alpha + \xi(\eta - \beta)}{\xi} \left\{ 1 - \left[\frac{T}{N_c} (1 - p) \right]^{-\xi} \right\} \\ \eta - \alpha \ln \left(\frac{T}{N_c} (1 - p) \right) \end{cases}$$

PoT VaR for stationary series

- Of course different declustering methods may yield different results and some tweaking may be needed
- Again, it should be noted that thresholds used for declustering do not have to be equal to the ones used when fitting GPD and that block sizes used for declustering do not have to be equal to the block sizes used in the block maxima method

Taking one step further

- So far we know how to use extremes for stationary data
- We know that log return data is typically not stationary, i.e. distribution is not constant over time, and trying to estimate distributions from non stationary data is not a good idea since non stationarity implies that distribution parameters change over time...
- To remedy this we may first "devolatilize" data using some mean and volatility series $\{\hat{\mu}_t\}$ and $\{\hat{\sigma}_t\}$

$$z_t = \frac{-r_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

Taking one step further

- Hopefully the series $\{z_t\}$ appears stationary and we fit some (maybe EVT) distribution F to this series
- We will then get

$$VaR_{p,t} = \hat{\mu}_t + \hat{\sigma}_t F^{-1}(p)$$

where $F^{-1}(p)$ denotes the p -quantile of F , i.e. the VaR_p for the stationary series $\{z_t\}$

- This is what we probably would use in real life to get a VaR-series that tracks returns and volatility clusters, instead of a constant VaR which would typically make our bank lose lots of money during volatility clusters and setting aside unnecessary regulatory capital in tranquil periods

Imposing stationarity

- There are several ways of “creating” the normalizing mean and volatility series $\{\hat{\mu}_t\}$ and $\{\hat{\sigma}_t\}$
- One simple, but often effective, way is to use windows of observations (of returns), for which one computes means and standard deviations

Imposing stationarity

- So if we have a series of returns $\{r_t\}$ we get the mean and volatility series $\{\hat{\mu}_t\}$ and $\{\hat{\sigma}_t\}$ from

$$\hat{\mu}_t = \frac{1}{\tau} \sum_{i=t-\tau}^{t-1} r_i$$

and

$$\hat{\sigma}_t = \sqrt{\frac{1}{\tau - 1} \sum_{i=t-\tau}^{t-1} (r_i - \hat{\mu}_t)^2}$$

- For daily returns the window length, τ , is typically chosen to be 21 days, which amounts to an approximate trading month
- However adjustments of the window length may improve performance

Using EVT distributions

- Using the estimation schemes covered in earlier lectures we may use block-maxima, PoT and extremal index methods to estimate quantiles $F^{-1}(p)$ to be used for the VaR-series
- It will likely be the case that different methods perform differently on different data sets
- For a "nice" data set an i.i.d. normal assumption for the standardized negated returns may suffice but in many cases one will need GEV or GP distributions and the extremal index and declustering.

Are the parameters statistically significant?

- To check this we may use profile likelihood
- The idea is to use that twice the difference between the log likelihood function for all parameters and the log likelihood function for all parameters but one follows a chi-squared distribution with one degree of freedom
- The 95% quantile for the chi-squared with one degree of freedom is 3.84 which means that we find a 95% CI for θ say by finding the values of θ for which the difference of the log likelihood function with all ML estimates plugged in and the log likelihood where we all the ML estimates except the one for θ is $3.84/2=1.92$

ML inference: Profile likelihood confidence intervals (Plots from Coles)

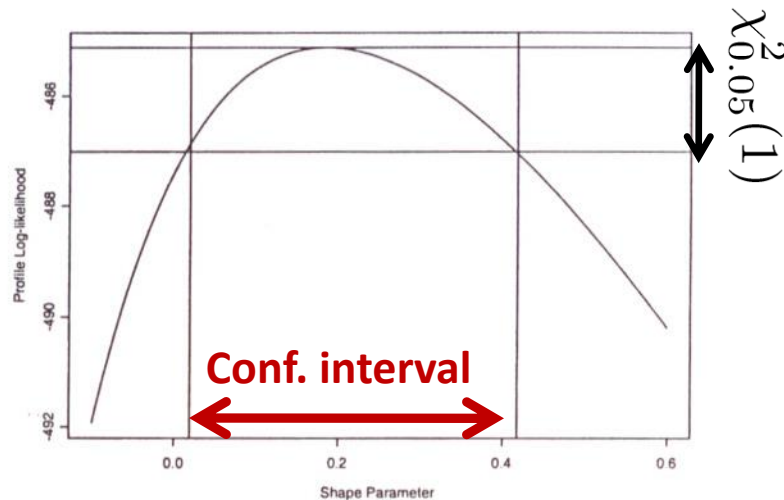


FIGURE 4.3. Profile likelihood for ξ in threshold excess model of daily rainfall data.

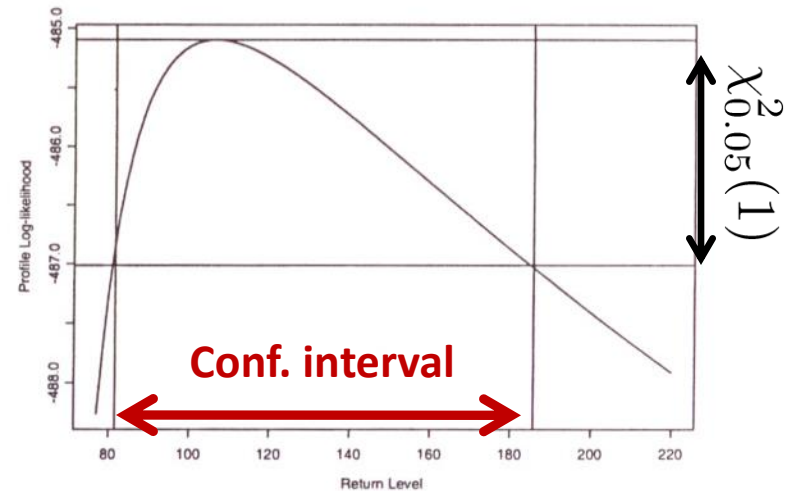


FIGURE 4.4. Profile likelihood for 100-year return level in threshold excess model of daily rainfall data.

Profile likelihood confidence intervals for the shape parameter and return level in the PoT (GPD) model.

A real life example

- Let's assume that we are working in a bank and want to find a reasonable model for 95% VaR of the Nikkei 225 index
- We will focus on fitting models to historical data in order to find a model to use for forecasts
- We are free to use pretty much any software we want but have to appreciate the fact that using Excel will probably not give us as much flexibility as using some more advanced package in which we can do ML-estimates
- Below we have used SAS for some plots and Matlab for the main analysis
- To fit GEV and GP we use the commands "gevfit" and "gpfit" which give ML-estimates and confidence intervals for parameters at hand
- QQ-plots are made using what has been covered on the slides together with the command "polyfit" to give reference lines

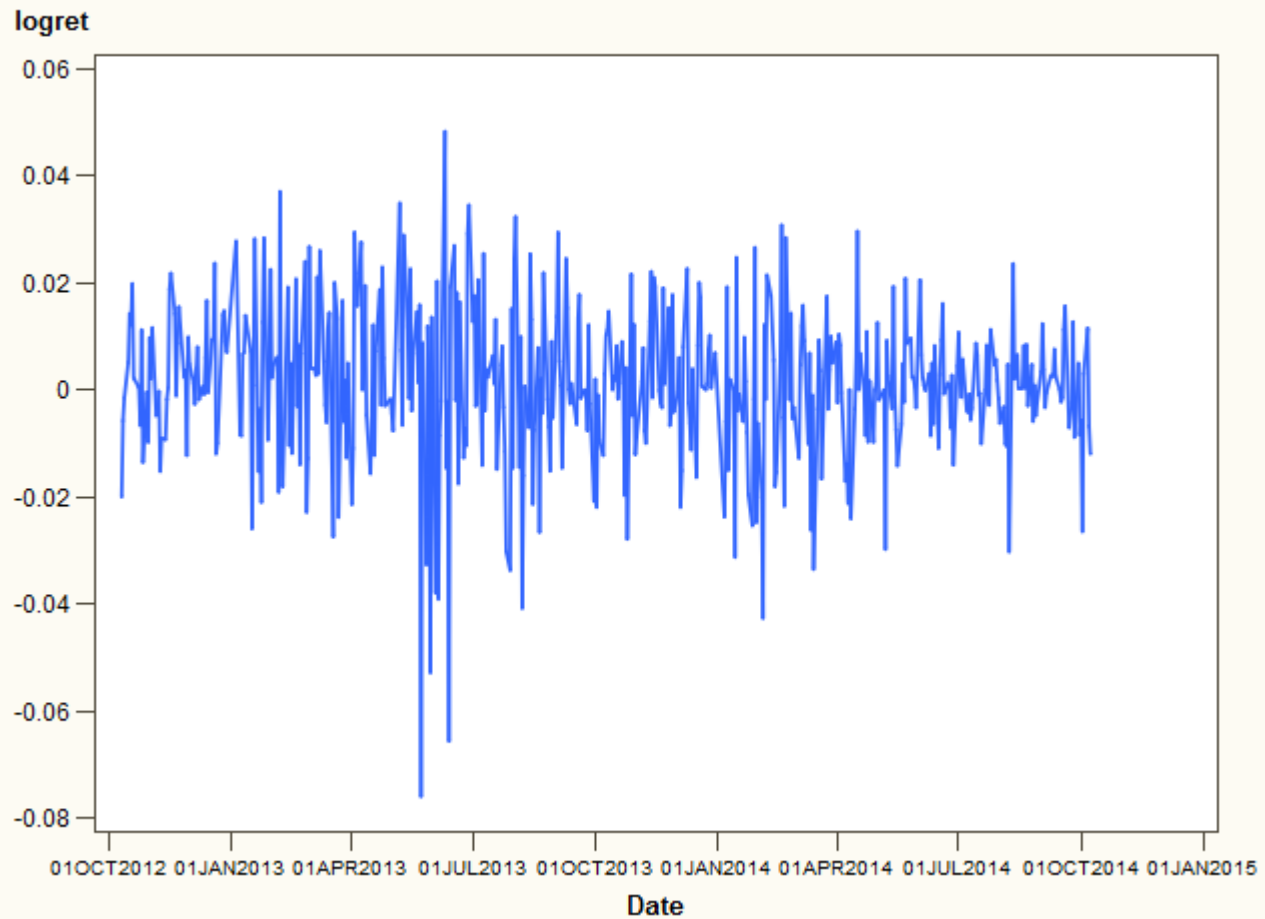
Nikkei 225

Nikkei 225 121009 to 141009

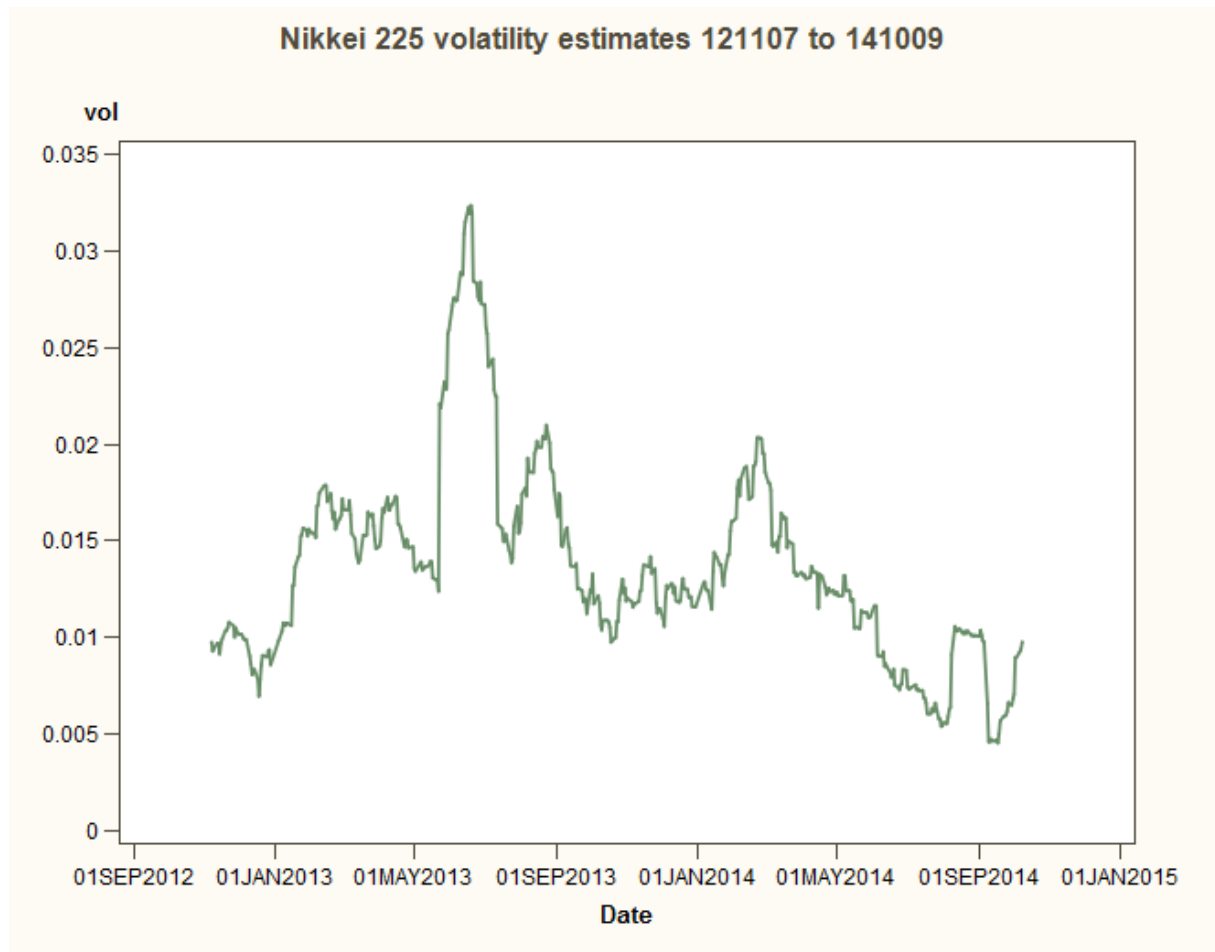


Nikkei 225

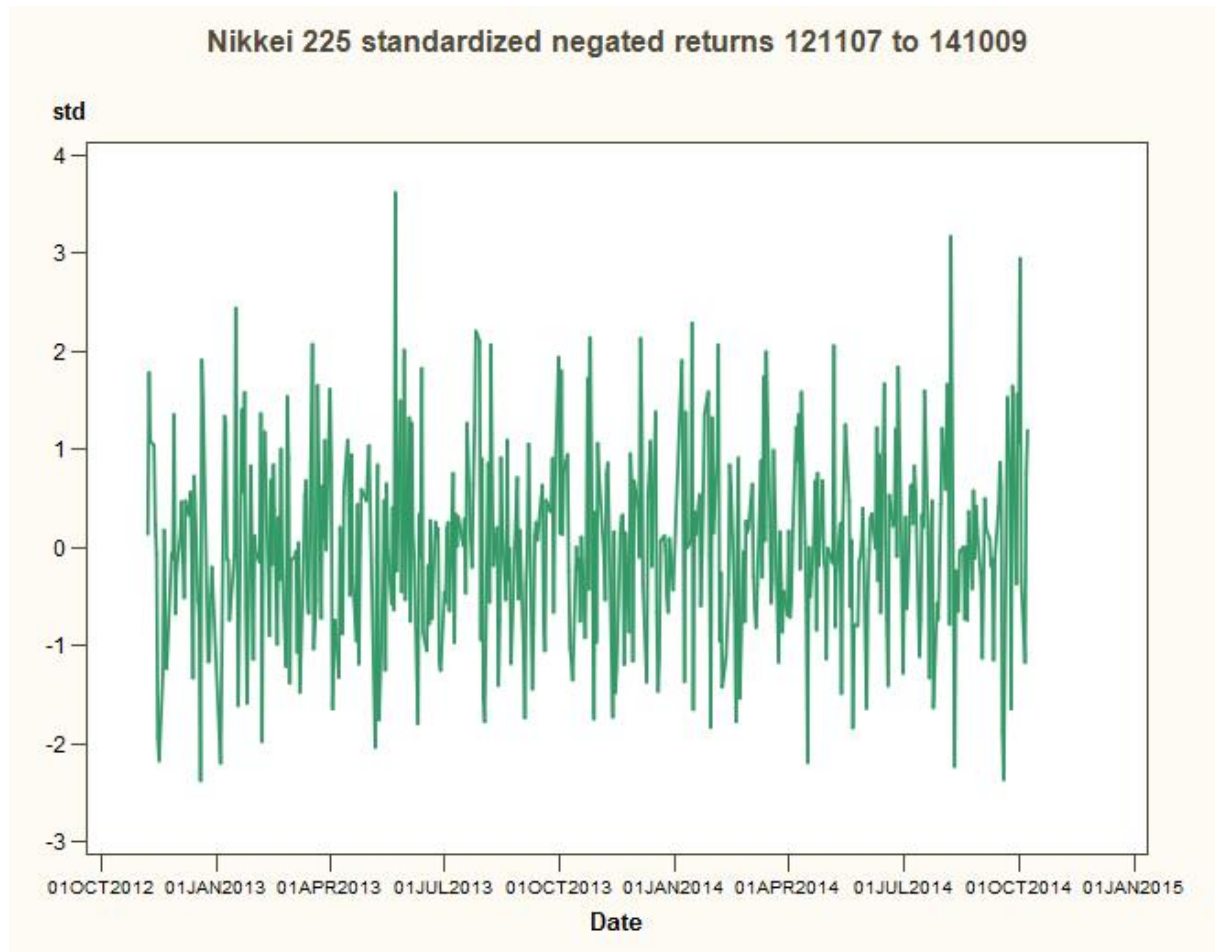
Nikkei 225 log-returns 121010 to 141009



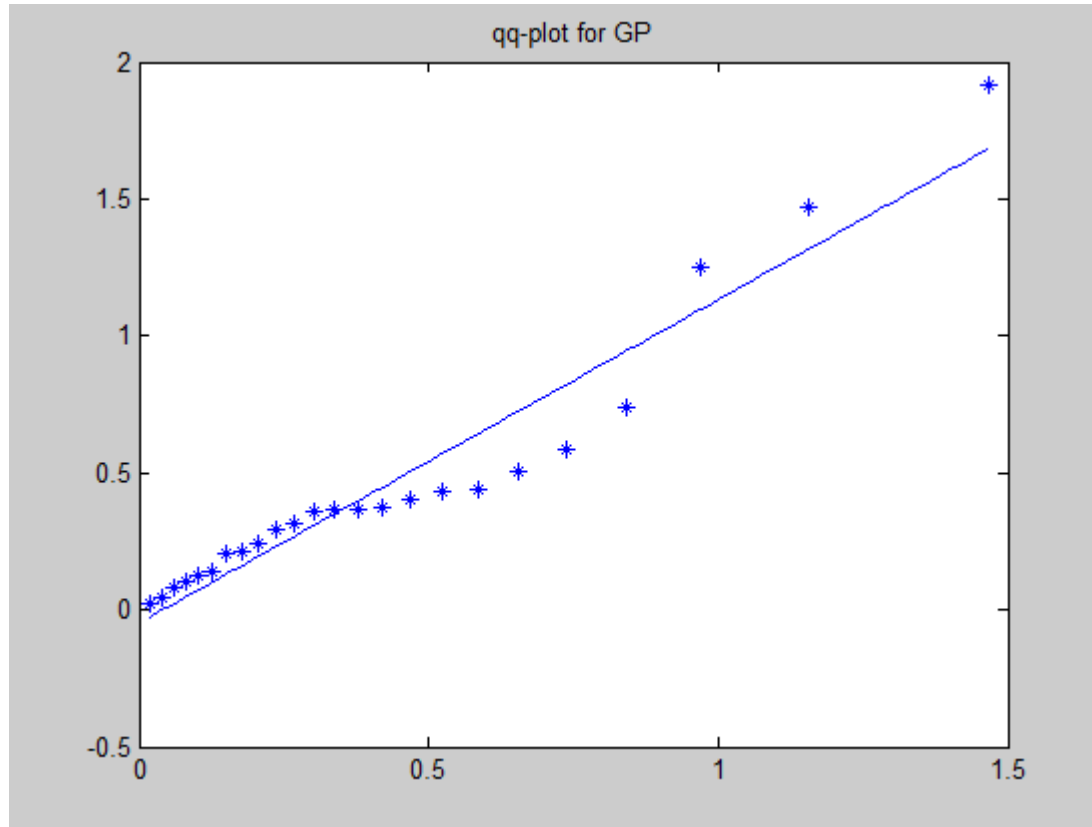
Nikkei 225



Nikkei 225

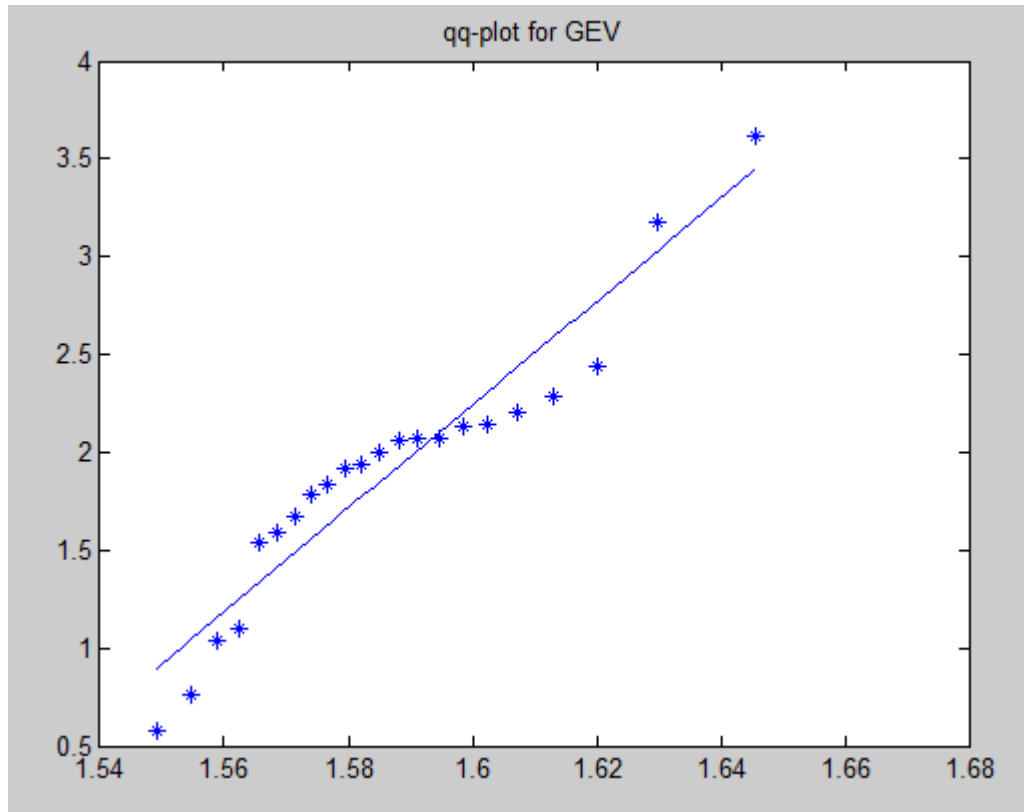


PoT for N225

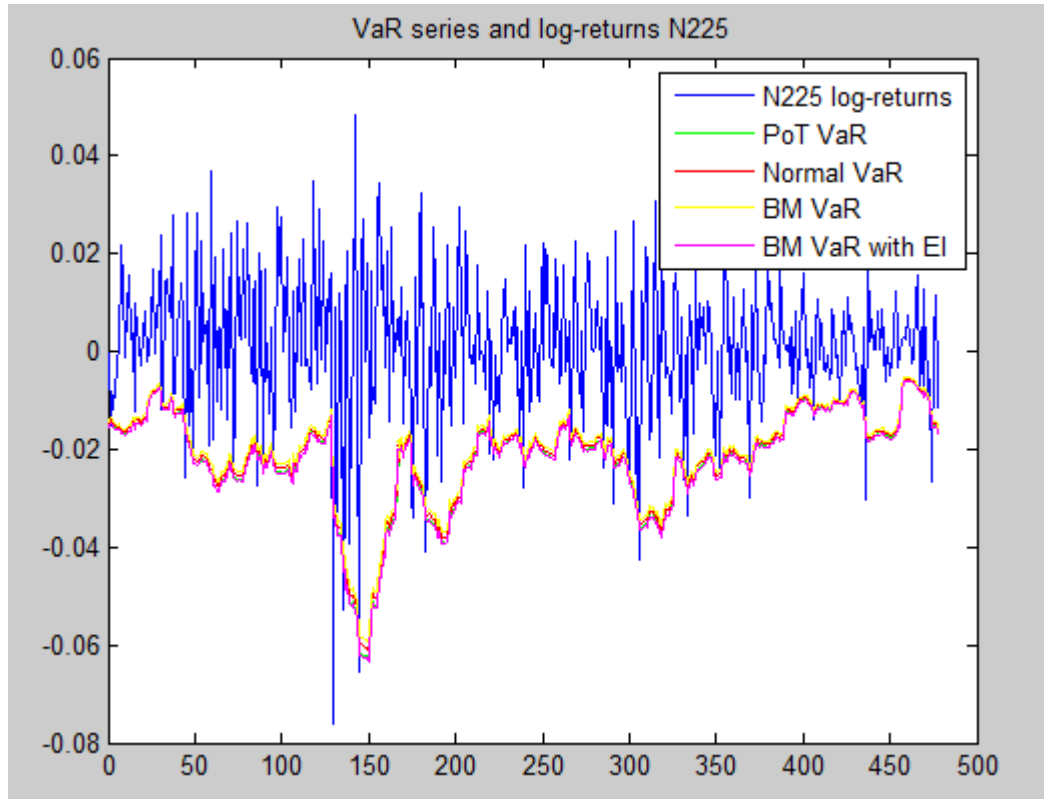


Here we used 95%-quantile as threshold and no declustering

BM for N225



95% VaR



Model	Violations
PoT	5.0%
Normal	5.9%
BM	6.5%
BM w EI	5.0%

The extremal index was estimated using the runs declustering

Conclusion

- There is no universal solution
- In calm periods normal VaR may work well
- In more volatile periods EVT methods probably yield better results
- Tweaking (which will be frustrating and time-consuming) is needed to find good models
- In practice/real life one would probably do daily or weekly re-estimate parameters for distributions and VaR used for forecasting
- There are more ways to evaluate models than the ones presented here

Summary and about tech proj 1

- What has been covered today will be included in technical project 1 and in particular one must know how to do what has been covered today, i.e. using the extremal index to capture dependencies in data when creating VaR series, starting from an arbitrary set of stock prices in order to get a grade higher than 4/VG
- Hence it is recommended that you download some data set, preferably two years of prices, from e.g. yahoo finance and starting playing around with it.
- Note that file downloaded from yahoo finance are csv-files that you may transfer to xlsx by using "text to columns" under "data" in Excel
- To get started check out the xlsx-file available on facebook and at the course web page