Financial Times Series

Lecture 9

Extreme Values

- For a time series $\{r_t\}$ of log returns we may be interested in the behavior of

$$M_n = max\{r_1, \dots, r_n\}$$

or

$$m_n = min\{r_1, \dots, r_n\}$$

- Of course these are of special interest in risk management when trying to hedge big losses

Extreme Values

- If we assume that the log returns are serially independent and have distribution function

$$F(x) = P(r_t \leq x)$$

it holds that

$$P(M_n \leq x) = P(r_1 \leq x, \dots, r_n \leq x) = \prod_{i=1}^{n} P(r_i \leq x)$$

$$= [F(x)]^n$$

Degeneration

- But what happens if we let the number of observations increase, i.e. let $n \to \infty$?

- Then $[F(x)]^n \to 0$ or $[F(x)]^n \to 1$ depending on if $x < u$ or $x \geq u$ where u is the upper end point of r_t (typically $u = \infty$ for log returns)

- So we need something more to get a non-trivial limit…

Appropriate sequences

- We need sequences $\{\alpha_n\}, \{\beta_n\}$ such that

$$M_{n*} = \frac{M_n - \beta_n}{\alpha_n}$$

converges to a non-trivial limit

- We sometimes refer to $\{\alpha_n\}$ and $\{\beta_n\}$ and the scaling and location sequences, respectively

Limiting distributions

- It turns out that if the limit exists its distribution function will be (Generalized extreme value distribution)

$$F_*(x) = exp\left[-(1+\xi x)^{-1/\xi}\right]$$

for $x < -1/\xi$ if $\xi < 0$ and for $x > -1/\xi$ if $\xi > 0$

- The special case $\xi = 0$ gives

$$F_*(x) = exp[-exp(-x)]$$

for $-\infty < x < \infty$

Three types

- Type I, $\xi = 0$, the Gumbel distribution

$$F_*(x) = exp[-exp(-x)], -\infty < x < \infty$$

- Type II, $\xi > 0$, the Fréchet distribution

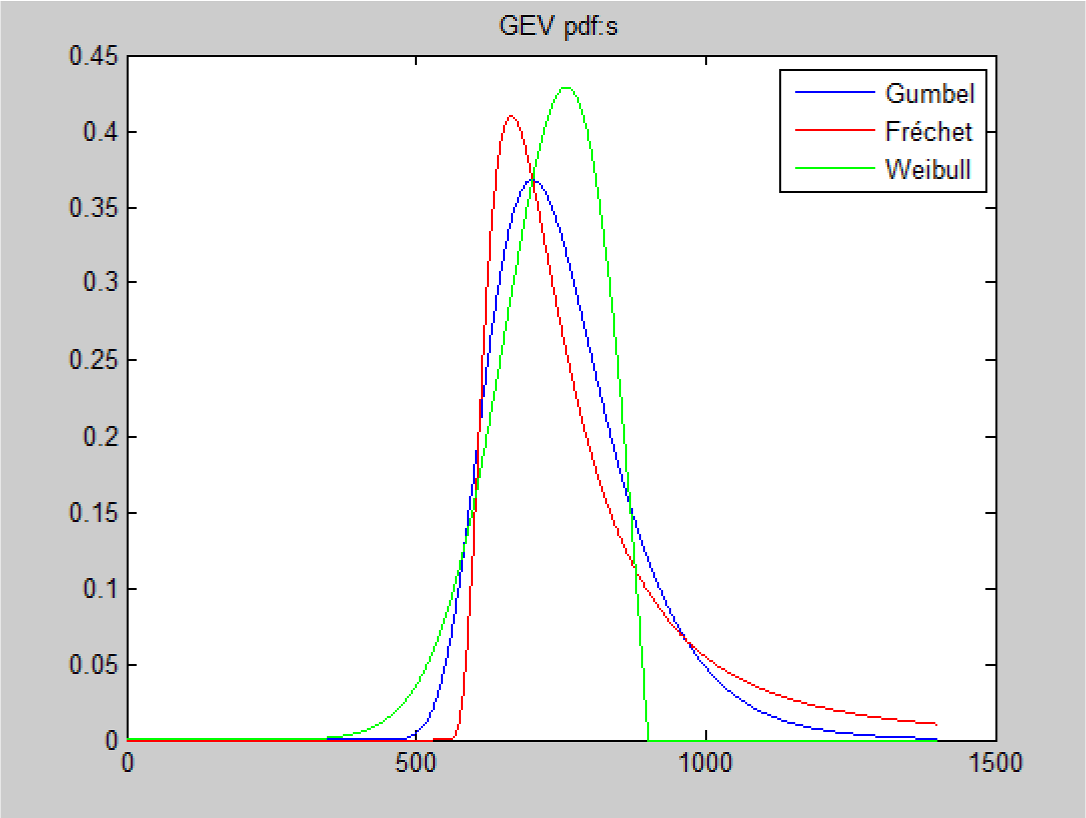
$$F_*(x) = \begin{cases} exp\left[-(1+\xi x)^{-1/\xi}\right], x > -1/\xi \\ 0 \qquad\qquad\qquad\qquad\quad , otherwise \end{cases}$$

- Type III, $\xi < 0$, the Weibull distribution

$$F_*(x) = \begin{cases} exp\left[-(1+\xi x)^{-1/\xi}\right], x < -1/\xi \\ 1 \qquad\qquad\qquad\qquad\quad , otherwise \end{cases}$$

Three types

- The pdf:s for $\xi = 0, \xi = 0.5$ and $\xi = -0.5$



In practice

- In real world problems we typically do not have to worry about what

$$F(x) = P(r_t \leq x)$$

looks like in order to fit extreme value distributions

- What we do have to worry about however is that returns typically not are independent or stationary

- It turns out that we may use the same distributions for maxima of dependent sequences if we can adjust for the dependence, at least if the dependence is not too strong

In practice

- Also, we do not really have to worry about what the sequences $\{\alpha_n\}, \{\beta_n\}$ look like

- It is so since if M_{n*} has limiting distribution $F_*(\cdot)$ we have for $y = \alpha_n x + \beta_n$ that

$$P(M_n \leq y) \approx F_*\left(\frac{y - \beta_n}{\alpha_n}\right)$$

- This means that the values of α_n and β_n may be involved in parameter estimations but since parameters are unknown anyway this is not a problem

Estimation

- For a given sample there is just one maximum or minimum
- Of course we cannot estimate parameters using just one observation
- One way of circumventing this problem is to divide the sample into non-overlapping blocks and then use the maximum from each block to estimate parameters

$$\left\{r_1, \dots, r_n \middle| r_{n+1}, \dots, r_{2n} \middle| \cdots \middle| r_{(k-1)n+1}, \dots, r_{kn}\right\}$$

Estimation

- The method described above is referred to as the ”Block maxima method”.

- For sufficiently large blocks the block maxima should follow the GEV distribution

- The block maxima may be considered a sample from the GEV

Estimation

- If we denote the block maxima $\{M_{1,n}, \dots, M_{k,n}\}$, the pdf needed for the ML estimation is (exercise)

$$f(M_{i,n}) = \frac{1}{\alpha_n}\left[1 + \frac{\xi_n(M_{i,n} - \beta_n)}{\alpha_n}\right]^{-\left(\frac{1}{\xi_n}+1\right)} exp\left[-\left(1 + \frac{\xi_n(M_{i,n} - \beta_n)}{\alpha_n}\right)^{-1/\xi_n}\right]$$

if $\xi_n \neq 0$ where it has to hold that $1 + \frac{\xi_n(M_{i,n}-\beta_n)}{\alpha_n} > 0$ and

$$f(M_{i,n}) = \frac{1}{\alpha_n} exp\left[-\frac{M_{i,n} - \beta_n}{\alpha_n} - exp\left(-\frac{M_{i,n} - \beta_n}{\alpha_n}\right)\right]$$

if $\xi_n = 0$

Estimation

- The likelihood function is then

$$L(\alpha_n, \beta_n, \xi_n | M_{1,n}, \dots, M_{k,n}) = \prod_{i=1}^{k} f(M_{i,n})$$

- The estimates will be unbiased and asymptotically normal

- Estimations may also be made using regression or non-parametric techniques, see Tsay 3rd ed p.347-348

Example

- Fitting the Gumbel distribution to the Decile1,2,9,10 data using 21-day blocks gives

Data	**Scale** α_n	**Location** β_n
Decile1	0.0807	0.1291
Decile2	0.0633	0.0983
Decile9	0.0265	0.0787
Decile10	0.0286	0.0728

Checking model fit

- One can define residuals as

$$e_i = \left(1 + \frac{\xi_n\left(M_{i,n} - \beta_n\right)}{\alpha_n}\right)^{-1/\xi_n}$$

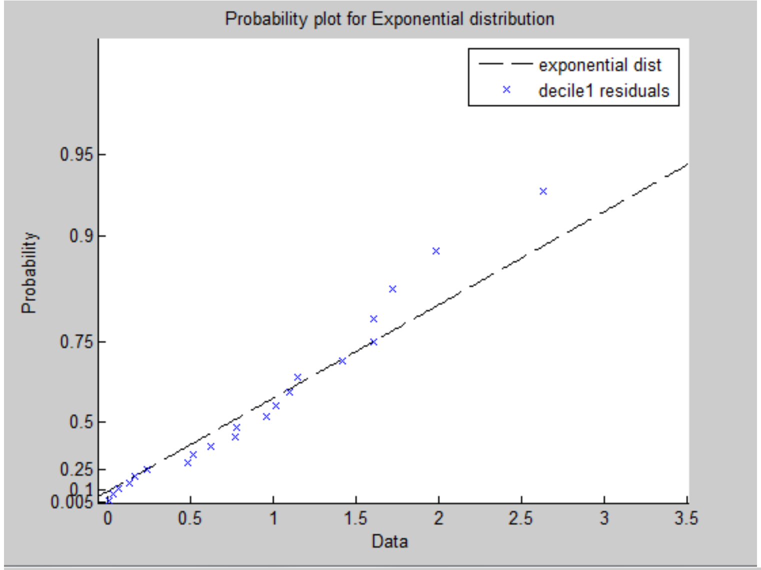
if $\xi_n \neq 0$ and

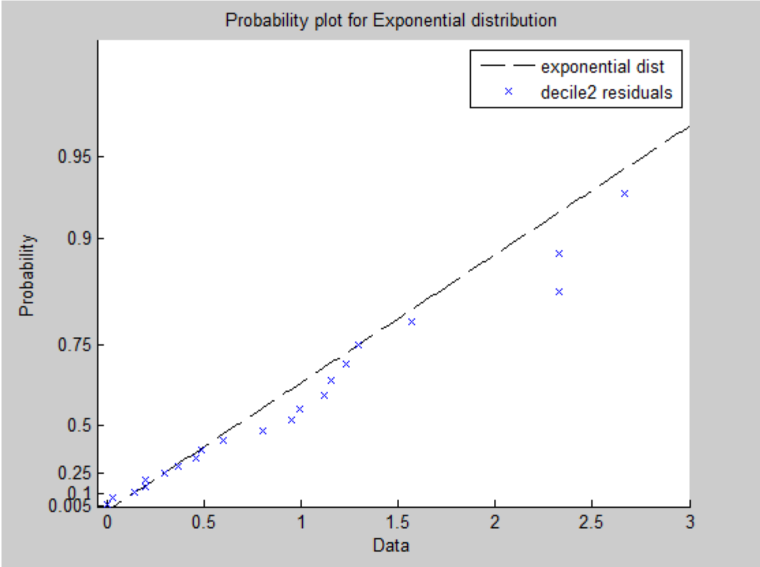
$$e_i = exp\left(-\frac{M_{i,n} - \beta_n}{\alpha_n}\right)$$

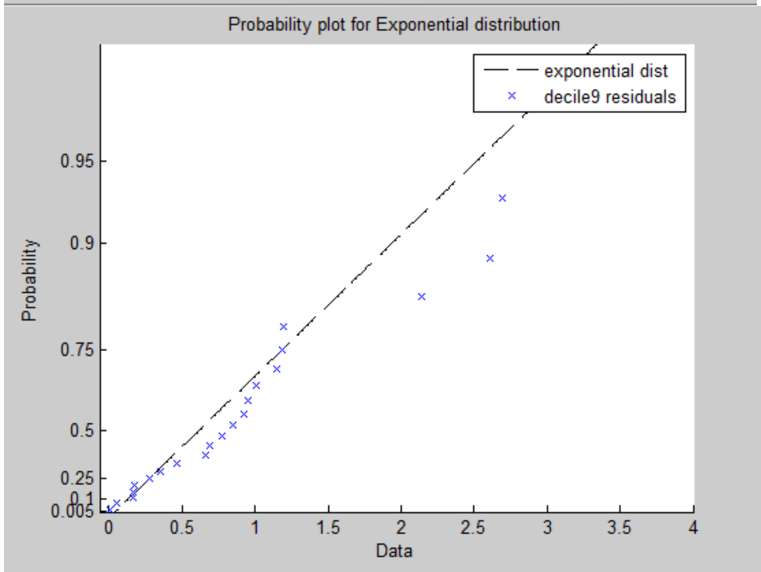
if $\xi_n = 0$

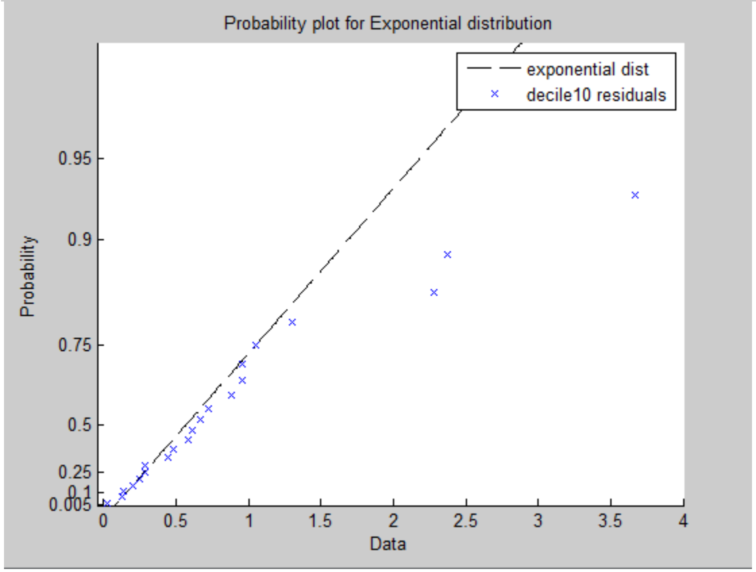
- Residuals should follow an exponential distribution if the model is correctly specified

PP-plots


Probability plot for Exponential distribution
exponential dist
decile1 residuals
Probability
0.95
0.9
0.75
0.5
0.25
0.1
0.005
0
0.5
1
1.5
2
2.5
3
3.5
Data


Probability plot for Exponential distribution
exponential dist
decile2 residuals
Probability
0.95
0.9
0.75
0.5
0.25
0.1
0.005
0
0.5
1
1.5
2
2.5
3
Data


Probability plot for Exponential distribution
exponential dist
decile9 residuals
Probability
0.95
0.9
0.75
0.5
0.25
0.1
0.005
0
0.5
1
1.5
2
2.5
3
3.5
4
Data


Probability plot for Exponential distribution
exponential dist
decile10 residuals
Probability
0.95
0.9
0.75
0.5
0.25
0.1
0.005
0
0.5
1
1.5
2
2.5
3
3.5
4
Data

Using Block-Maxima for VaR

- In VaR we are interested in quantiles

- Using GEV distributions and assuming that we have negated returns so that a high return is a big loss we let p^* be the (small) probability of a great loss and write

$$1 - p^* = \begin{cases} exp\left\{-\left[1 + \frac{\xi_n({M^*}_n - \beta_n)}{\alpha_n}\right]^{-1/\xi_n}\right\} \\ exp\left\{-exp\left[-\frac{{M^*}_n - \beta_n}{\alpha_n}\right]\right\} \end{cases}$$

Using Block-Maxima for VaR

- Solving for the quantile ${M^*}_n$ we get

$$
{M^*}_n = \begin{cases} \beta_n - \dfrac{\alpha_n}{\xi_n}\left\{1 - [-\ln(1-p^*)]^{-\xi_n}\right\} \\ \beta_n - \alpha_n \ln[-\ln(1-p^*)] \end{cases}
$$

- Now this is the quantile for the number (n) of observations in each the block so we have to transform it to use it for one-day VaR

Using Block-Maxima for VaR

- Under the assumption of independent returns we may use that

$$1 - p^* = P\left(M_{i,n} \leq {M^*}_n\right) = [P(r_t \leq {M^*}_n)]^n$$

- So if we want $P(r_t \leq {M^*}_n) = 1 - p$ we get

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

Using Block-Maxima for VaR

- Using the decile data and the Gumbel model for block maxima of negated returns we get

Data	95% VaR	99% VaR
Decile1	0.0737	0.1538
Decile2	0.0711	0.1485
Decile9	0.0592	0.1091
Decile10	0.0530	0.1006

Return Level

- We may be interested in what levels losses are expected to exceed within in a certain time frame

- We refer to this as the return level $L_{n,k}$ where k denotes the number of periods of length n and

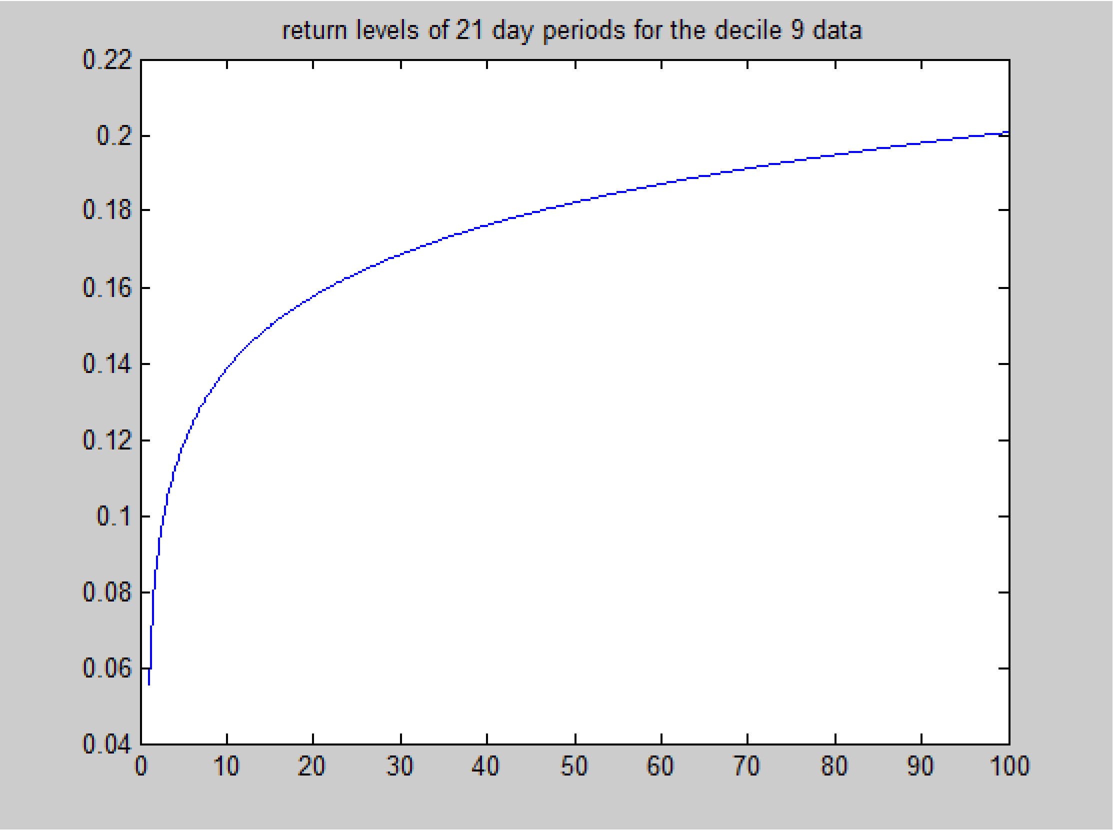
$$P\left(M_{i,n} > L_{n,k}\right) = \frac{1}{k}$$

Return Level

- If n is large enough so that the distribution of maxima is GEV we have that

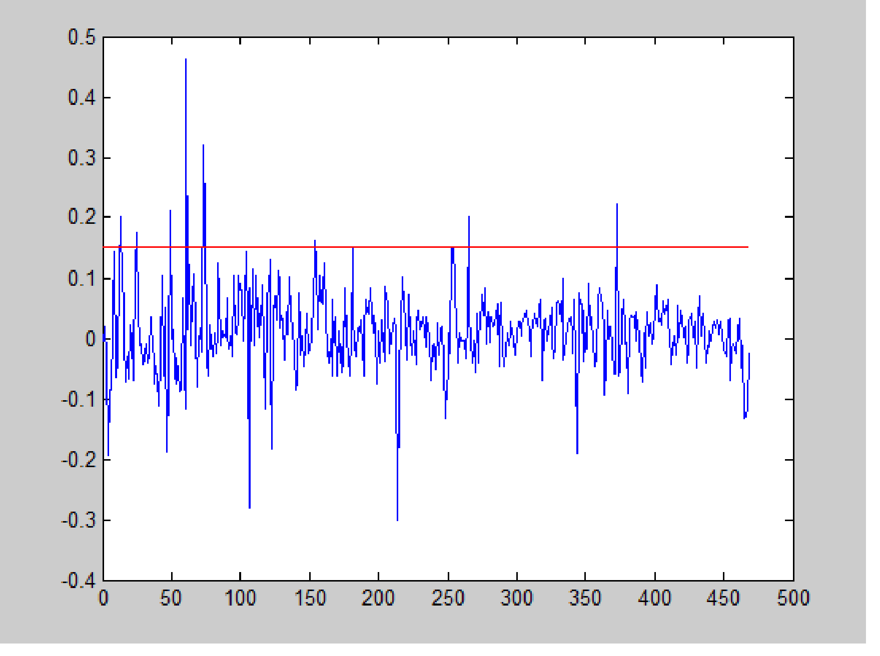
$$L_{n,k} = \begin{cases} \beta_n - \frac{\alpha_n}{\xi_n}\left\{1 - \left[-\ln\left(1 - \frac{1}{k}\right)\right]^{-\xi_n}\right\} \\ \beta_n - \alpha_n \ln\left[-\ln\left(1 - \frac{1}{k}\right)\right] \end{cases}$$

Return Level



Peaks Over Threshold (POT)

- Instead of using block maxima to model behaviour we may use observations above certain level or threshold



Peaks Over Threshold (POT)

- Letting t_i denote a time where an exceedance of the threshold η occurs we will use the data

$$\{r_{t_1} - \eta, \ldots, r_{t_n} - \eta\}$$

- In POT we do not have to choose a block size but a threshold η and different choices will give different estimates but it turns out that VaR done with POT may not be as sensitive to the threshold choice as VaR done with block maxima is to the choice of block length
- Starting out with a value of η so that 5% of the sample is left is a rule of thumb

Which distribution?

- Since we will just use returns above η the distribution we will use is conditional on the event $r_t > \eta$

$$P(r_t \leq x + \eta \,|r_t > \eta\,) = \frac{P(r_t \leq x + \eta) - P(r_t \leq \eta)}{1 - P(r_t \leq \eta)}$$

- Using $P(r_t \leq x) = exp\left[-\left(1 + \xi \frac{x-\beta}{\alpha}\right)^{-1/\xi}\right]$ and the approximation $e^x \approx 1 - x$ gives (exercise)

$$P(r_t \leq x + \eta \,|r_t > \eta\,) \approx 1 - \left[1 + \frac{\xi x}{\alpha + \xi(\eta - \beta)}\right]^{-1/\xi}$$

where $x > 0$ and $\alpha + \xi(\eta - \beta) > 0$

GDP

- We refer to the above distribution as the Generalized Pareto Distribution

- The limiting case $\xi = 0$ gives

$$P(r_t \leq x + \eta \,|r_t > \eta\,) \approx 1 - exp[-x/\alpha]$$

- We sometimes write (the scale parameter)

$$\psi(\eta) = \alpha + \xi(\eta - \beta)$$

Important properties

- If a certain threshold η_0 yields a shape parameter ξ and scale parameter $\psi(\eta_0)$, a higher threshold $\eta > \eta_0$ will yield the shape parameter

$$\psi(\eta) = \psi(\eta_0) + \xi(\eta - \eta_0)$$

- For the case $\xi = 0$ the GPD is just an exponential distribution so peaks over threshold should in this case behave according to an exponential distribution

Mean Excess Function

- Another tool for choosing threshold is using the mean excess over the threshold η_0, assuming $\xi < 1$ (exercise)

$$E(r_t - \eta_0 | r_t > \eta_0) = \frac{\psi(\eta_0)}{1 - \xi}$$

- Then for any $\eta > \eta_0$ we have the mean excess function

$$e(\eta) = E(r_t - \eta | r_t > \eta) = \frac{\psi(\eta_0) + \xi(\eta - \eta_0)}{1 - \xi}$$

- So for a fixed ξ the mean excess function is linear in $\eta - \eta_0$

Empirical Mean Excess Function

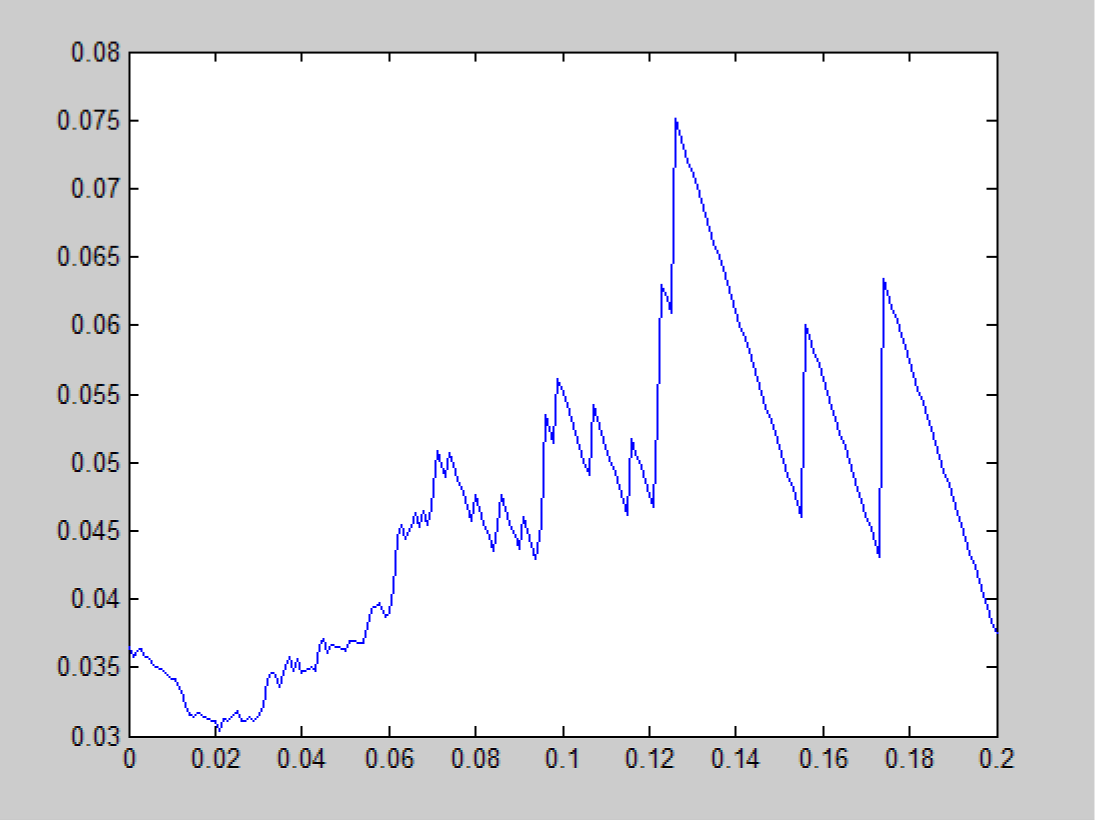
- In practice we will use the empirical mean excess function

$$e_T(\eta) = \frac{1}{N(\eta)} \sum_{i=1}^{N(\eta)} \left(r_{t_i} - \eta\right)$$

where $N(\eta)$ is the number of exceedances of η

- We plot $e_T(\eta)$ against η and choose the threshold η_0 as the value for which $\eta > \eta_0$ gives a linear appearance of $e_T(\eta)$

For negated decile 9 data



Decile 9

- We are working with negated returns since we are interested in risks of big losses

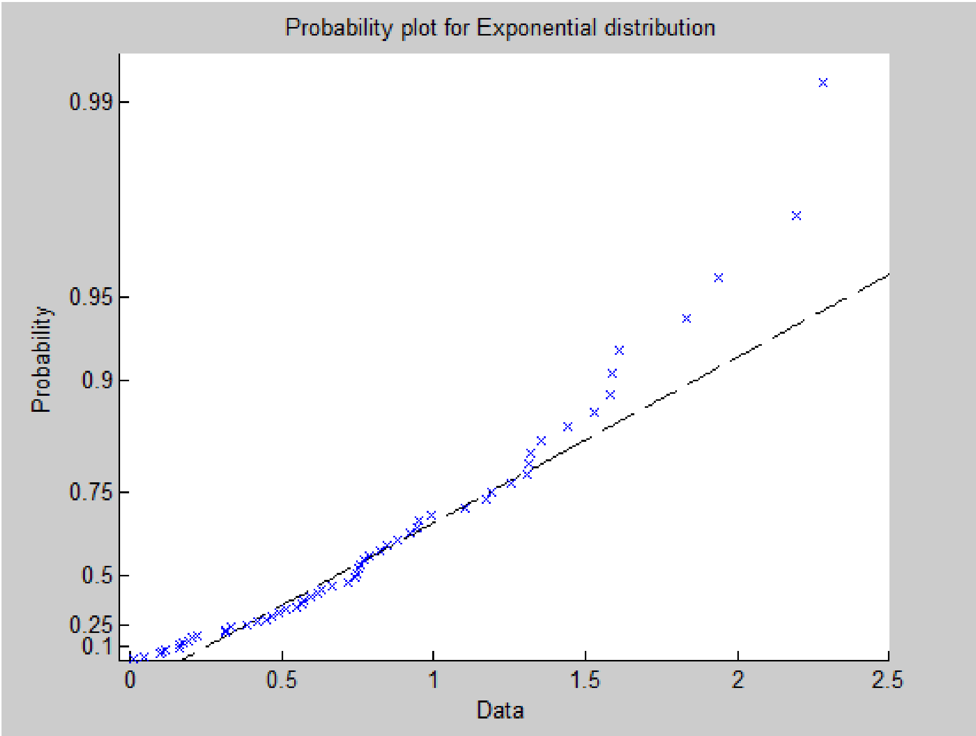
- Based on the the mean excess plot we choose $\eta_0 = 0.04$ and use ML to fit the GPD to the series $\{r_{t_i} - 0.04\}$ for those r_{t_i} that exceed the threshold

- The parameter estimates are

$$\hat{\alpha} = -0.0275, \hat{\beta} = 0.0029, \hat{\xi} = 1.2049$$

Model checking

- We define residuals as $\frac{1}{\hat{\xi}}\ln\left(1+\hat{\xi}\,\frac{r_{t_i}-\eta_0}{\hat{\alpha}+\hat{\xi}(\eta_0-\hat{\beta})}\right)$ which should follow an exponential distribution



Model checking

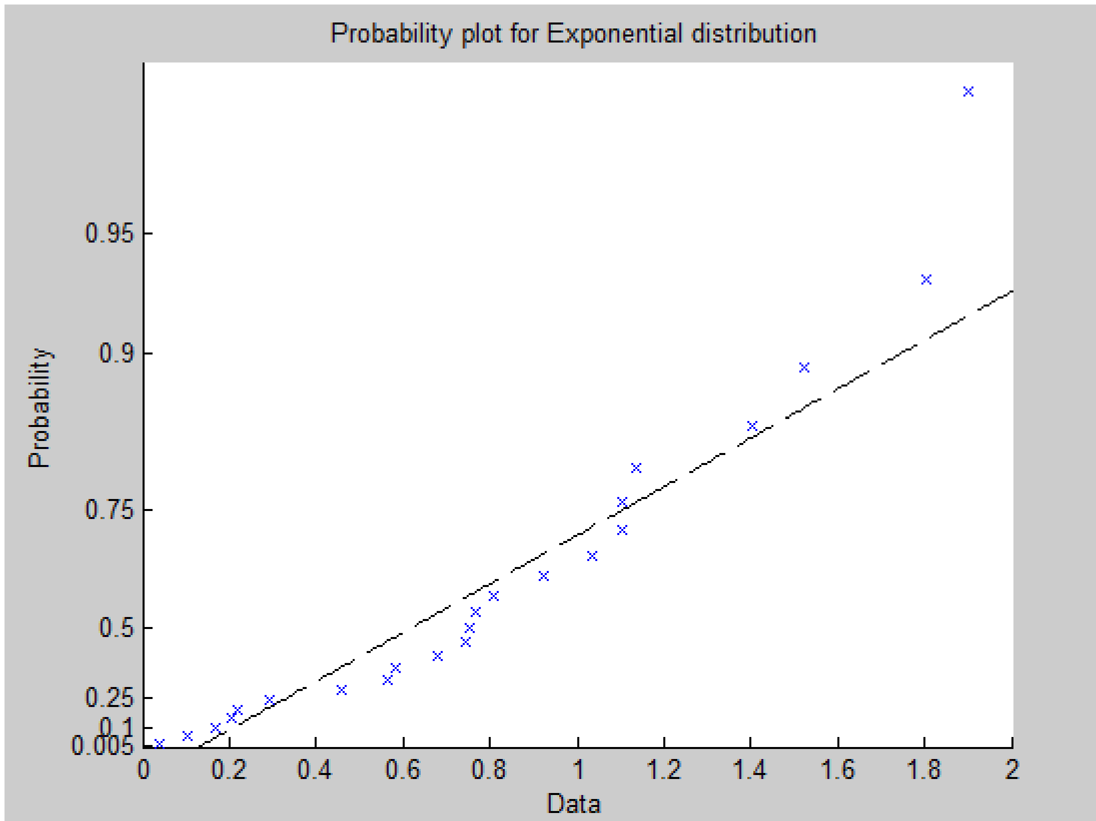
- The above probability plot was made using the estimated parameters using the threshold 0.04 suggested by the mean excess plot

- However the residual plot suggests trying a larger threshold (to be able to capture the big losses) and we will use the 5% rule of thumb which yields $\eta_0 = 0.0644$

- Re-estimation gives

$$\hat{\alpha} = -0.0492, \hat{\beta} = 0.0030, \hat{\xi} = 1.2072$$

Model checking



VaR using POT

- So we know how to estimate $P(r_t \leq x + \eta \,|r_t > \eta\,)$ since, letting $y = x + \eta$,

$$P(r_t \leq y|r_t > \eta\,) = \frac{F(y) - F(\eta)}{1 - F(\eta)} \approx 1 - \left[1 + \frac{\xi x}{\alpha + \xi(\eta - \beta)}\right]^{-1/\xi}$$

- This gives (why?)

$$F(y) \approx 1 - \frac{N(\eta)}{T}\left[1 + \frac{\xi(y - \eta)}{\alpha + \xi(\eta - \beta)}\right]^{-1/\xi}$$

VaR using POT

- So for a small upper tail probability p, letting $q = 1 - p$ we have

$$VaR_q = \eta - \frac{\alpha + \xi(\eta - \beta)}{\xi}\left\{1 - \left[\frac{T}{N(\eta)}(1 - q)\right]^{-\xi}\right\}$$

- Also the associated expected shortfall is given by

$$ES_q = \frac{VaR_q + \alpha - \xi\beta}{1 - \xi}$$

Decile 9

- Using the parameter estimates above and that we have 23 exceedances in 468 observations we get we get the 95% and 99% VaR for the decile 9 data; 0.0640 and 0.1849

- Compare to 0.0592 and 0.1091 from Block-Maxima method

Extremal Index

- So far we have assumed that we have IID data which is not a correct assumption
- Typically exceedances 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 IID series $\{\tilde{r}_t\}$ with the same marginal distribution as $\{r_t\}$ that, if we have convergence

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

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 IID sequence converges

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

Extremal Index

- Since

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

we get (exercise)

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

Where $\xi_* = \xi$, $\alpha_* = \alpha\theta^{\xi}$ and $\beta_* = \beta - \alpha\left(1 - \theta^{\xi}\right)/\xi$

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 firs exceedance of u and stops as soon as there is a value below u, the next starts with the next exceedance of u an 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

GEV VaR for a stationary series

- So we may estimate scale, location and shape for the GEV as if data were IID and then adjust parameter estimates using our estimate of the extremal index

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

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.

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 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 ”devolatize” (negated) data using one of our volatility models

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

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

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

where $F^{-1}(q)$ denotes the q-quantile of F

- This is what we probably would use (and have used for GARCH and its relatives) in real life...