Extreme Value Analysis

Sidney Resnick
Cornell University
School of Operations Research and Industrial Engineering
sid@orie.cornell.edu
http://www.orie.cornell.edu/~sid

Books

1. J. Beirlant, J. Teugels, and P. Vynckier. *Practical Analysis of Extreme Values*. Leuven University Press, Leuven, 1996.

2. P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extreme Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997.

3. R.-D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values*. Birkhäuser Verlag, Basel, 2001. From insurance, finance, hydrology and other fields, With 1 Windows CD-ROM.

4. S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer-Verlag, New York, 1987.

Software

1. *Xtremes*: Package accompanying Reiss & Thomas book. Menu driven, easy to use. Some programming capability. Many nice data sets. Project under long term development.

2. *McNeil Splus module EVIS*: Module created for Splus by Alexander McNeil available free at http://www.math.ethz.ch/mcneil/. Versions for windows or unix. Professionally done. Requires some version of Splus or Splus2000. Easy to extend within the Splus environment.

Extreme Value Analysis

Build models where principle features of interest extremes, not central values.

Problem: How to make inferences well beyond the range of the data?

Examples.

1. T-year flood. Project Neptune in Netherlands had goal to reassess height of the Dutch dikes. The 10,000 year floot is the height u_{10000} such that the expected time between exceedances of u_{10000} is 10,000 years. Must estimate

$$F^{\leftarrow}(1 - \frac{1}{10,000}) = F^{\leftarrow}(\frac{9999}{10,000}),$$

where F is the distribution of the maximal height per year. If we had, say, 100 years worth of data, u_{10000} would be well outside range of data and could not be estimated non-parametrically using say the empirical cdf.

Examples (cont).

2. VAR-value-at-risk. Let

$$S_t = \text{ price of asset at } t$$

and define log-returns as

$$R_t = \log S_t - \log S_{t-1}.$$

Stylized facts:

Mg hypothesis

Volatility

Heavy tails

$\{R_t\}$ has minimal correlation

but $\{|R_t|\}$ and $\{R_t^2\}$ have LRD.

The loss variable is the "loss" expressed in positive units:

$$L_t = -(S_t - S_0) = \begin{cases} |S_t - S_0|, & \text{if } S_t - S_0 < 0, \\ -|S_t - S_0|, & \text{if } S_t - S_0 > 0. \end{cases}$$

So if L_t is negative, there is a profit. The value-at-risk parameter VaR(T, q) for the period T is the qth quantile of the loss distribution defined by

$$P[L_T \leq \text{VaR}(T, q)] = q.$$

How to compute: Define

$$F_T(x) = P[-\sum_{t=1}^{T} R_i \leq x],$$

and

$$\text{VaR}(T, q) = V_0\Big(1 - e^{-F_T^{\leftarrow}(q)}\Big).$$

For $q = 0.999$ say, there is only prob .0001 of losses exceeding VaR(T, q) over a time span length T. Must estimate very large quantile.

3. Expected shortfall. If the loss exceeds VaR, by how much?

Compute

$$E\Big(L_T | L_T > \text{VaR}(T,q)\Big),$$

or

$$E\Big(L_T - \text{VaR}(T,q) | L_T > \text{VaR}(T,q)\Big).$$

If L_T has distribution F, then

$$E\Big(L_T | L_T > \text{VaR}(T,q)\Big) = \int_{\text{VaR}(T,q)}^{\infty} x \frac{F(dx)}{\bar{F}(\text{VaR}(T,q))}.$$

4. CaR: Capital-at-risk. This is the maximal amount which may be invested so that a potential loss exceeds a given limit with given small probability. Given l and q, CaR(T,q,l) is the initial capital S_0 satisfying

$$q = P[L_T \le l] = P[\text{CaR}(T,q,l)(1-e^{\sum_{i=1}^{T} R_i}) \le l]$$

and it turns out

$$\text{CaR}(T,q,l) = \frac{l}{1 - e^{-F_T^{\leftarrow}(q)}} \approx \frac{l}{F_T^{\leftarrow}(q)}.$$

Basic Theory

Two inference methods for extremes:

1. Exceedances. Exceedance sequence = *partial duration series* or PDS. Analysis using exceedances is called the *peaks over threshold* or POT method.

2. Maxima. *Method of yearly maxima*. Sequence of maxima called the *annual maxima series* or AMS.

Suppose $X_1, \ldots, X_n$ are iid with common distribution $F(x)$ and define

$$M_n := \bigvee_{i=1}^{n} X_i = \max\{X_1, \ldots, X_n\}.$$

The distribution of M_n:

$$\begin{aligned} P[M_n \leq x] =& P[X_1 \leq x, \ldots, X_n \leq x] \\ =& P[X_1 \leq x] \cdots P[X_n \leq x] \\ =& F^n(x). \end{aligned}$$

Maxima

Say $F \in \mathcal{D}(G)$ if there exist scaling constants $a_n > 0$ and centering constants $b_n \in \mathbb{R}$ such that as $n \to \infty$

$$P[\frac{M_n - b_n}{a_n} \le x] = F^n(a_n x + b_n) \to G(x) \tag{DofA}$$

for all x such that $0 < G(x) < 1$, where we also must assume that G is a proper distribution whose probability mass is not concentrated at one point.

Remarks. (1) Importance? Sppse $X_1, \dots, X_n$ are iid with common (unknown or partly known) distribution F. We need the distribution of M_n. Write

$$F^n(a_n x + b_n) \approx G(x),$$

or changing variables $y = a_n x + b_n$

$$F^n(y) \approx G\Big(\frac{y - b_n}{a_n}\Big).$$

So if F^n unknown or hard to compute, instead deal with a location/scale family of distributions.

Remarks. (2) DofA holds for most F's (for some G): $\{$normal, log-normal, weibull, gamma, exponential, Gumbel$\}$, $\{$log-gamma, pareto, stable, Frechet$\}$, uniform.

Remarks. (3) Caution: the rate of convergence in DofA can vary enormously.

Remarks. (4) Caution: The iid assumption may not be sensible?

The method of yearly maxima: Suppose for the i-th "year" we have observations (claims, water levels, financial exposures)

$$X_j^{(i)},\ j = 1, \dots, m,$$

producing "yearly" maxima

$$Y_i = \bigvee_{j=1}^{m} X_j^{(i)},\ i = 1, \dots, n.$$

Perhaps $X_j^{(i)}$, $i = 1, \dots, n; j = 1, \dots, m$ is not observed or not retained. Must make inferences based on observed maxima over n years $y_i, 1 \le i \le n$.

If the X's $\sim F$, then the maxima

$$Y_1, Y_2, \dots, Y_n,$$

are a random sample size n from $F^m(x)$ and an approximate random sample from $G((y - b_m)/a_m)$, a location and scale family.

But: What is G?

Class of Extreme Value Distributions

If $F \in \mathcal{D}(G)$, then G is one of the types of the following classes of distributions called the *extreme value distributions*:

(i) Gumbel or EV0 class with an exponential tail:

$$G_0(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R}.$$

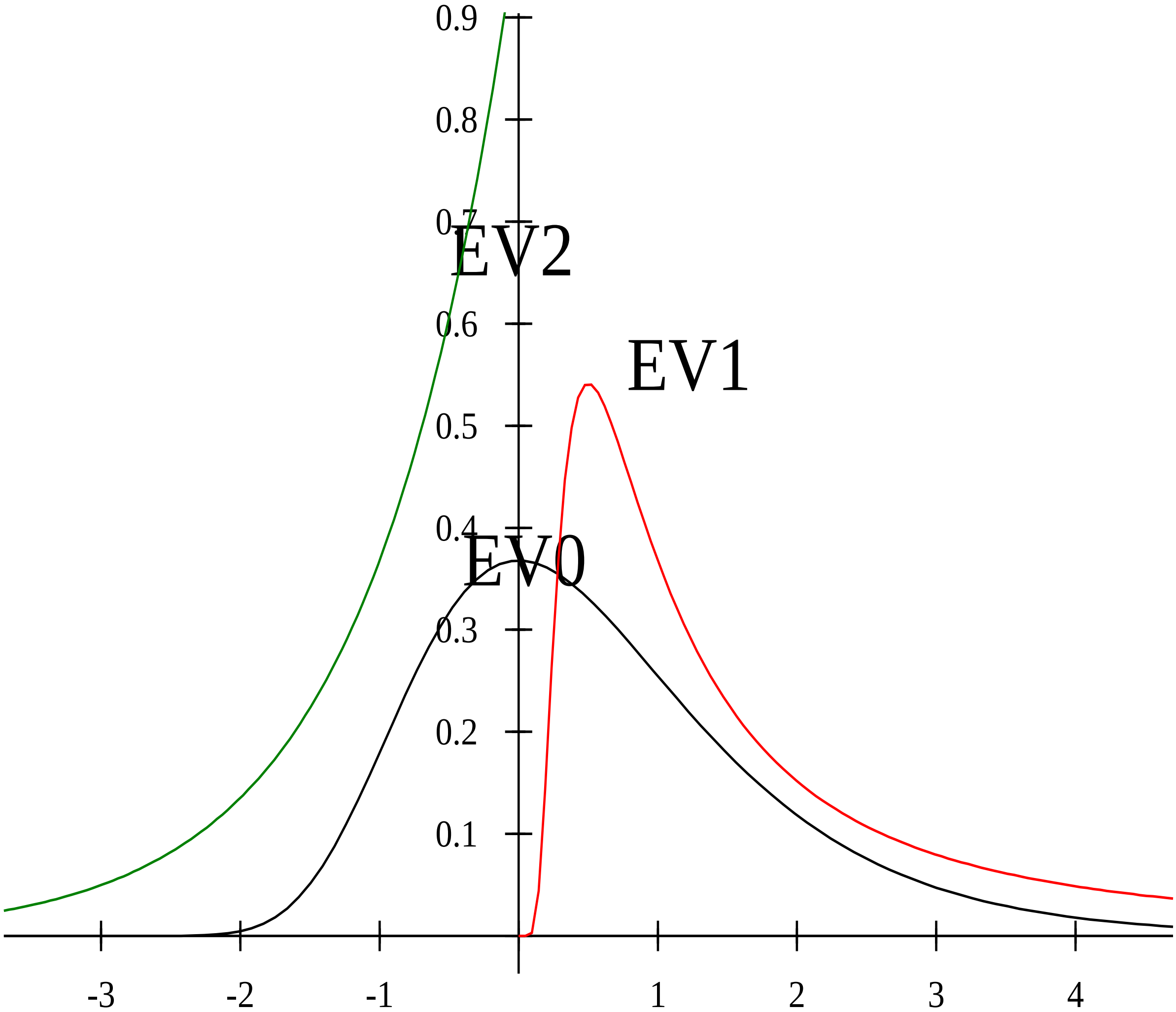
(ii) Frechet or EV1 class with a heavy tail and which is bounded below:

$$G_{1,\alpha}(x) = \begin{cases} e^{-x^{-\alpha}}, & \text{if } x > 0, \alpha > 0, \\ 0, & \text{if } x \leq 0. \end{cases}$$

(iii) Weibull or EV2 class which is bounded above:

$$G_{2,\alpha}(x) = \begin{cases} e^{-|x|^{-\alpha}}, & \text{if } x < 0, \alpha < 0, \\ 1, & \text{if } x \geq 0. \end{cases}$$

.



Extreme value densities: EV0, EV1, EV2.

Von Mises Parametrization

Various names: von Mises parameterization, γ-parameterization and Jenkinson parameterization.

Set $\gamma = 1/\alpha$, which is sometimes called the *extreme value (shape) parameter*. Without worrying about location and scale for the time being, define for $\gamma \in \mathbb{R}$,

$$G_\gamma(x) = e^{-(1+\gamma x)^{-1/\gamma}}, \quad 1 + \gamma x > 0.$$

When $\gamma = 0$, write

$$\lim_{\gamma \to 0} (1 + \gamma x)^{-1/\gamma} = e^{-x},$$

and so

$$G_0(x) = \exp\{-e^{-x}\}, \quad x \in \mathbb{R},$$

the Gumbel distribution. Set

$$G_{\gamma,\mu,\sigma}(x) = G_\gamma\left(\frac{x-\mu}{\sigma}\right),$$

and we get a 3-parameter family dependent on shape, location and scale.

Note

1. $\gamma > 0 \Rightarrow$ heavy tail, support $=(-\frac{1}{\gamma}, \infty)$

$$\bar{G}_\gamma(x) \sim x^{-1/\gamma}, \quad x \to \infty.$$

2. $\gamma = 0 \Rightarrow$ exponential tail, support $=(-\infty, \infty)$

$$\bar{G}_\gamma(x) \sim e^{-x}, \quad x \to \infty.$$

3. $\gamma < 0 \Rightarrow$ bounded above, support $=(-\infty, \frac{1}{|\gamma|})$.

Moral: The assumption $F \in \mathcal{D}(G)$ is mild and robust. Almost all text book F's satisfy this assumption. G is one of the extreme value distributions. If you need to fit a distribution to an annual maxima series, try fitting the 3-parameter family $G_{\gamma,\mu,\sigma}$ by MLE.

Exceedances

Pick level u and observations bigger than u are the *exceedances*. Names:

u =level, threshold,
priority level, retention level.

If $\{X_n\}$ iid $\sim F$, *exceedance times* $\{\tau_j, j \geq 1\}$ defined by

$$\begin{aligned}
\tau_1 &= \inf\{j \geq 1 : X_j > u\} \\
\tau_2 &= \inf\{j > \tau_1 : X_j > u\} \\
&\vdots \\
\tau_r &= \inf\{j > \tau_{r-1} : X_j > u\}.
\end{aligned}$$

$$\begin{aligned}
\{X_{\tau_j}, j \geq 1\} &= \textit{exceedances} \\
\{X_{\tau_j} - u, j \geq 1\} &= \textit{excesses.}
\end{aligned}$$

In <u>reinsurance</u>, excesses correspond to XL-treaty excesses of loss.

Distribution theory: $\{X_{\tau_j}, j \geq 1\}$ iid and

$$P[X_{\tau_j} > x] = \bar{F}^{[u]}(x) := P[X_1 > x | X_1 > u]$$
$$= \begin{cases} \frac{\bar{F}(x)}{\bar{F}(u)}, & \text{for } x > u, \\ 1, & \text{for } x < u. \end{cases}$$
$$=^d X_1 | X_1 > u.$$

Theorem. If $F \in \mathcal{D}(G)$, $Q = -\log G$, then as $n \to \infty$

$$P[X_{\tau_j^{(n)}} \leq a_n x + b_n] \to W(x)$$
$$= \begin{cases} 0, & \text{if } Q(x) > 1, \\ 1 - Q(x), & \text{if } Q(x) \leq 1. \end{cases}$$

Since G =EV distribution, we get the following possibilities for types of limiting exceedance distributions corresponding to $G_0, G_{1,\alpha}, G_{2,\alpha}$:

1. Exponential distribution (GP0):

$$W_0(x) = 1 - e^{-x}, \quad x \geq 0.$$

2. Pareto distribution (GP1):

$$W_{1,\alpha}(x) = 1 - x^{-\alpha}, \quad \alpha > 0, x \geq 1.$$

3. Beta distribution (GP2)

$$W_{2,\alpha}(x) = 1 - |x|^{-\alpha}, \quad \alpha < 0, -1 \leq x \leq 0.$$

Von Mises parameterization of GP distributions

Write for $\gamma \in \mathbb{R}$ the shape parameter family

$$Q_\gamma(x) = (1 + \gamma x)^{-1/\gamma}, \quad 1 + \gamma x > 0,$$

with the understanding that

$$Q_0(x) = e^{-x}.$$

Define

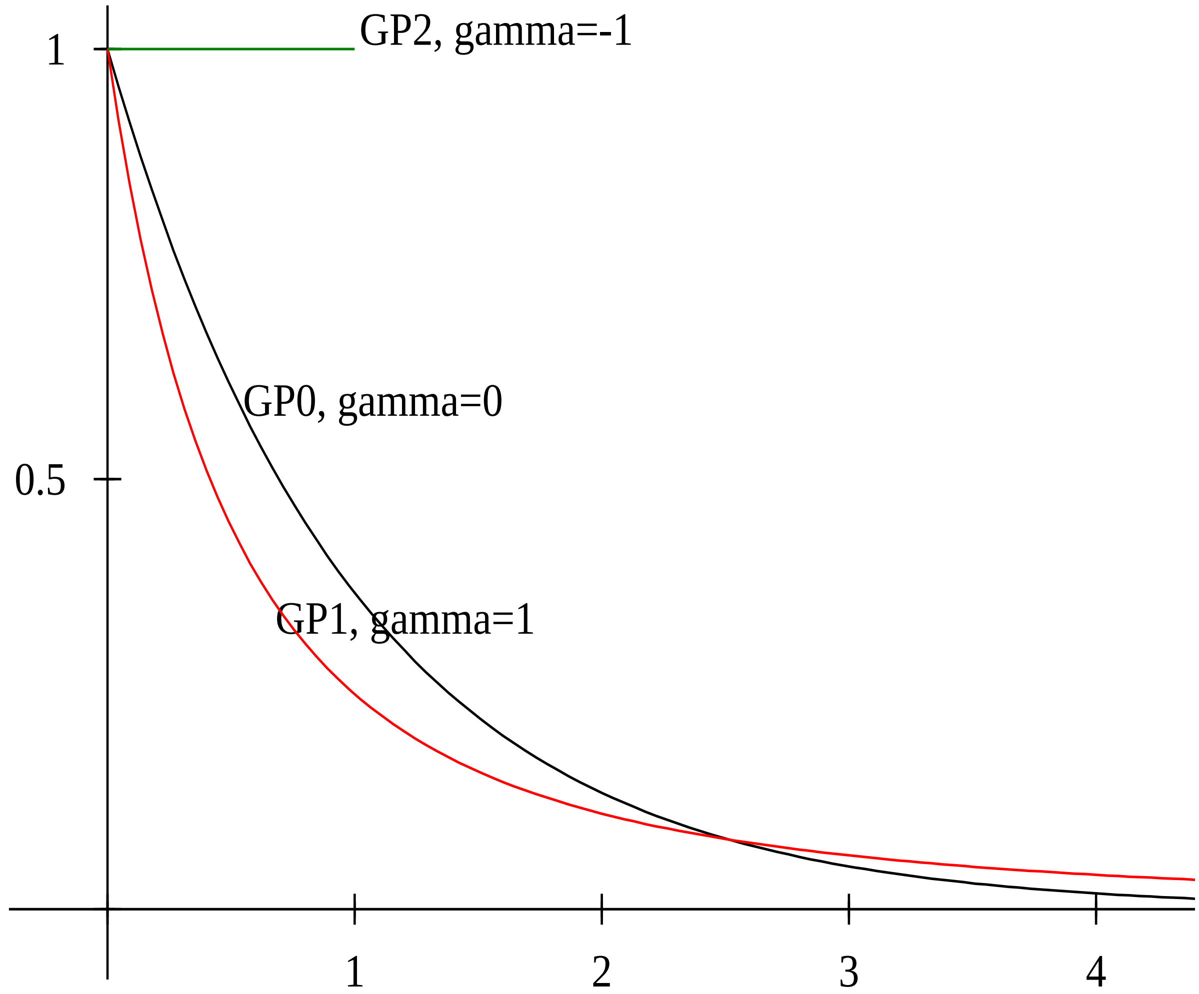
$$W_\gamma(x) = \begin{cases} 1 - Q_\gamma(x), & \text{if } 0 \le Q_\gamma(x) \le 1, \\ 0, & \text{if } Q_\gamma(x) > 1. \end{cases}$$
$$\begin{aligned} &= 1 - e^{-x}, \quad x > 0,\ \gamma = 0. \\ &= 1 - (1+\gamma x)^{-1/\gamma}, \quad x > 0,\ \gamma > 0. \\ &= 1 - (1+\gamma x)^{-1/\gamma}, \quad 0 < x < \frac{1}{|\gamma|},\ \gamma < 0. \end{aligned}$$

Three parameter GP family depending on location=μ, scale=σ, shape=γ:

$$W_{\gamma,\mu,\sigma}(x) = W_\gamma(\frac{x-\mu}{\sigma}).$$

Note $\gamma = 1/\alpha > 0$ corresponds to <u>heavy tails</u>. Cases:

$\alpha > 2 \Rightarrow$ finite variance
$1 < \alpha < 2 \Rightarrow$ infinite variance, finite mean,
$\alpha < 1 \Rightarrow$ infinite variance, infinite mean.



Generalized Pareto densities: GP0, $\gamma = 0$; GP1,$\gamma = 1$ (Pareto); GP2, $\gamma = -1$ (uniform).

Empirical CDF & Quantiles

The qth order quantile of a distribution $F(x)$ is

$$F^{\leftarrow}(q) = \inf\{s : F(s) \geq q\}, \quad 0 < q < 1.$$

Let $X_1, \ldots, X_n$ iid $\sim F(x)$, F unknown or partly known. Define for $x \in \mathbb{R}$ empirical cdf

$$\begin{aligned} \hat{F}_n(x) =& \frac{1}{n} \sum_{j=1}^{n} 1_{(-\infty,x]}(X_j) \\ =& \% \text{ of observations } \leq x. \end{aligned}$$

Then $\hat{F}_n(x) \to F(x)$ uniformly in x as $n \to \infty$. So

$$\hat{F}_n(x) \approx F(x)$$

and expect

$$\hat{F}_n^{\leftarrow}(x) \approx F^{\leftarrow}(x).$$

Define order statistics

$$X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}.$$

For $0 < q < 1$

$$\widehat{F}_n^{\leftarrow}(q) = X_{\lceil nq \rceil :n}$$

is an estimator of $F^{\leftarrow}(q)$. When does this make sense?

Example. Goal: estimate the <u>100</u> year flood

$$F^{\leftarrow}(1 - \frac{1}{100}) = F^{\leftarrow}(\frac{99}{100})$$

with 100 data points? According the the previous prescription, we use $X_{99:100}$ the 2nd largest order statistic.

Reminder: $\widehat{F}_n^{\leftarrow}(q) = X_{\lceil nq \rceil : n}$.

Goal: estimate <u>1000</u> year flood with 100 data points:

$$F^{\leftarrow}(1 - \frac{1}{1000}) = F^{\leftarrow}(\frac{999}{1000}) = F^{\leftarrow}(.999)$$

with $n = 100$? Then $q = .999$, $nq = 99.9$ and $\lceil 99.9 \rceil = 100$;
estimate is $X_{100:100}$, the largest observation.

Goal: estimate <u>10,000</u> year flood based on 100 observations:

$$F^{\leftarrow}(1 - \frac{1}{10,000}) = F^{\leftarrow}(\frac{9999}{10,000}) = F^{\leftarrow}(.9999)$$

and $q = .9999$, $nq = 99.99$, $\lceil nq \rceil = 100$
estimate still sample max.

This is not very clever.

Conclusion: Estimating extreme quantiles beyond the range of the data is not sensible using this (non-parametric) method based on order statistics. Extrapolate beyond data range using EVT.

Diagnostics & Estimation

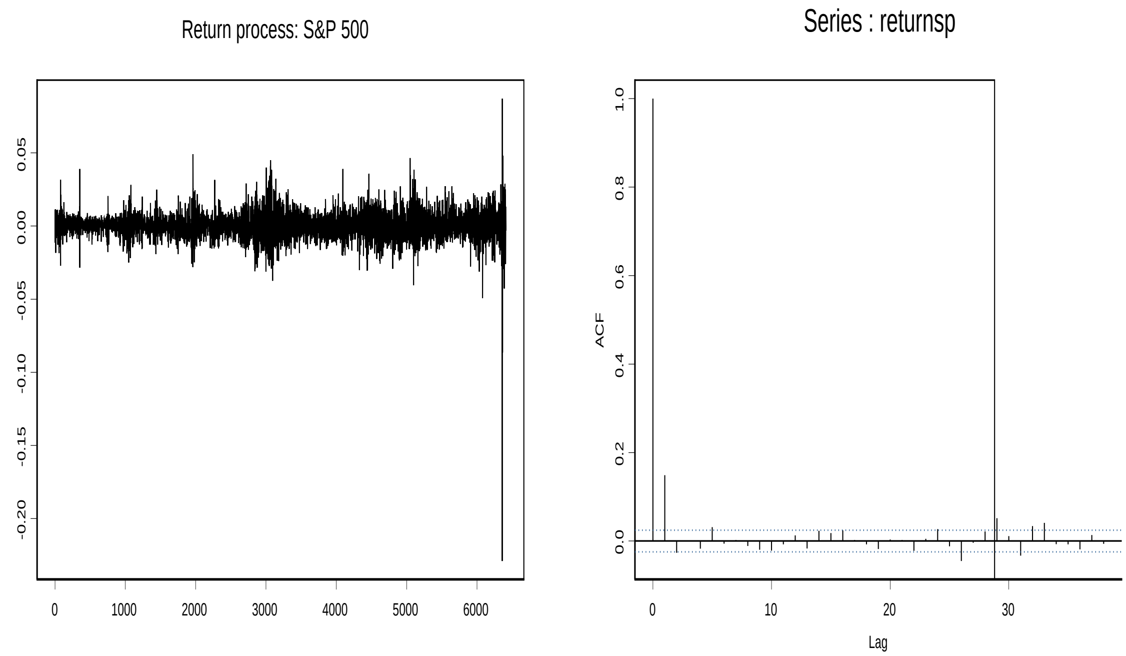
Some useful techniques:

0. Quick check of iid assumption with TS plot and ACF plot.
1. MLE estimation in 3-parameter model.
2. QQ plot as diagnostic or confirmatory technique. Is the data heavy tailed? For a correct model, empirical quantiles (of $\widehat{F}_n(x)$) plotted vs model quantiles should yield an approximate straight line.
3. Variant of QQ plot: mean excess plot; requires finite mean.
4. Hill plot and variants for heavy tailed analysis.
5. Quantile estimation using the fitted model tail.

Case Study: S&P 500

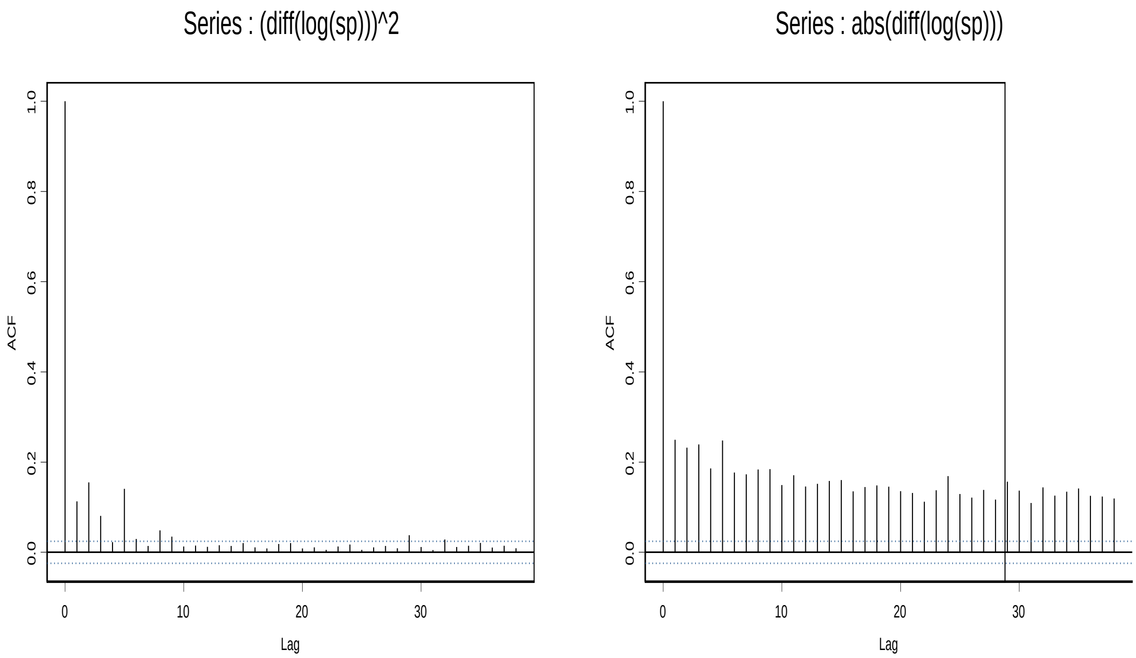
Standard & Poors 500 stock market index: daily data from July 1962 to December 1987; no corrections for weekends or other market closures. Log-Returns were computed by

returns = diff(log(S&P)).



Time series plot of S&P 500 return data (left) and the autocorrelation function (right).

Although the log-returns do not exhibit much correlation, this is not true for $(\text{log-returns})^2$ and $|(\text{log-returns})|$:

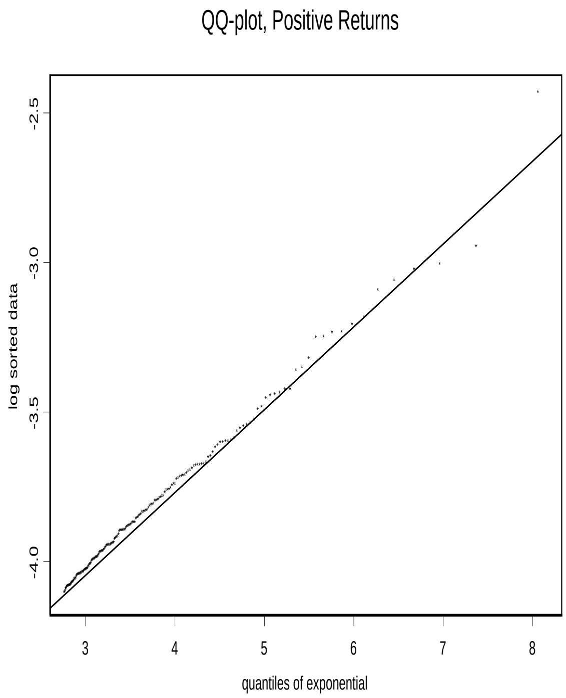


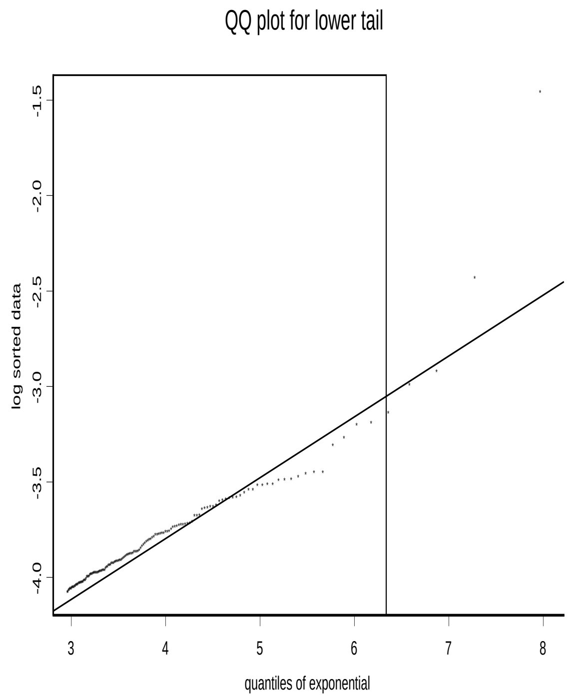
The autocorrelation function of the squared returns (left) and the autocorrelation function of the absolute values of the returns.

Are the tails of the log-return process heavy? QQ-plotting.

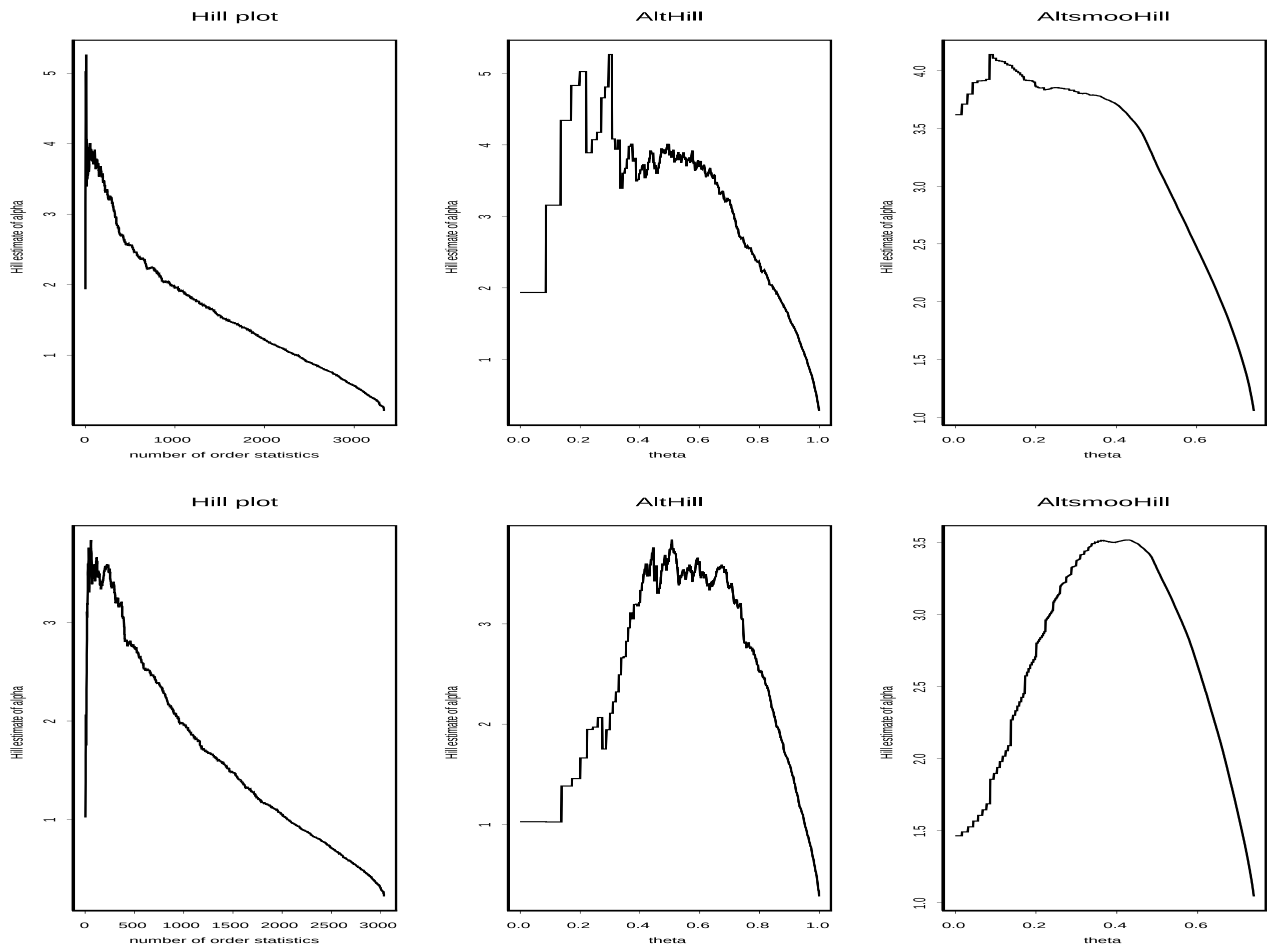
Left: positive returns, $k = 200$, slope estimate of $\widehat{\alpha} = 3.61$.

Right: abs(returns[returns¡0]), $k = 150$, $\widehat{\alpha} = 3.138$.





Hill, altHill and smooHill plots of the two tails.
Upper row: right tail, alt on log scale, smoothing: $r = 8$.
Lower row: left tail, alt on log scale, smoothing: $r = 8$.



Summary of estimates: Estimates of α from various methods. Gains if one can do estimation in a restricted family ($\alpha > 0$). Note the sensitivity of the estimates to the choice of k.

Est'r	k	$\widehat{\alpha}$	CI	MSE
QQ	200	3.61		
	100	3.63		
Hill(GP1)	200	3.45	$[2.9, 4.0]$	0.07
	100	3.81	$[3.2, 4.7]$	0.16
Mom(GP)	200	5.277	$[3.0, 35.9]$	142
	100	4.26	$[-11.2, 23.9]$	1914
MLE(GP)	200	5.85	$[-29.3, 44.9]$	1149
	100	4.32	$[-23.2, 45.6]$	1137

Right tail.

Estimator	k	$\widehat{\alpha}$	CI
QQ	150	3.138	
	100	2.98	
Hill(GP1)	150	3.387	$[2.90, 4.04]$
	100	3.48	$[2.85, 4.26]$
Moment(GP)	150	2.774	$[2.096, 7.02]$
	100	2.415	$[1.83, 7.27]$
MLE(GP)	150	3.25	$[2.002, 8.457]$
	100	2.75	$[1.462, 13.914]$

Left tail.

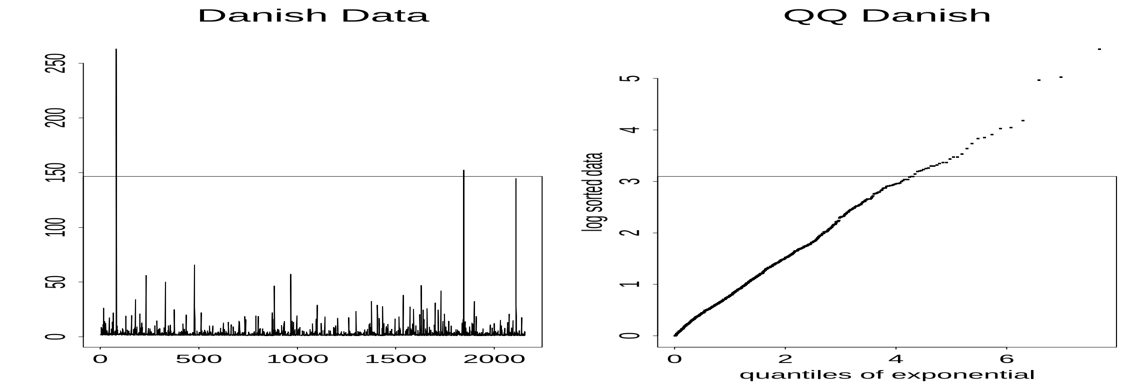
VaR Calculation–left tail quantile.

Estimator	k	$\widehat{\alpha}$	$q = .99$	q=.999
Hill(GP1)	150	3.387	.0214	.0423
	100	3.48	.0214	.0415
Moment(GP)	150	2.774	.0217	.046
	100	2.415	.0214	.0457
MLE(GP)	150	3.25	.0214	.0425
	100	2.75	.0212	.0425

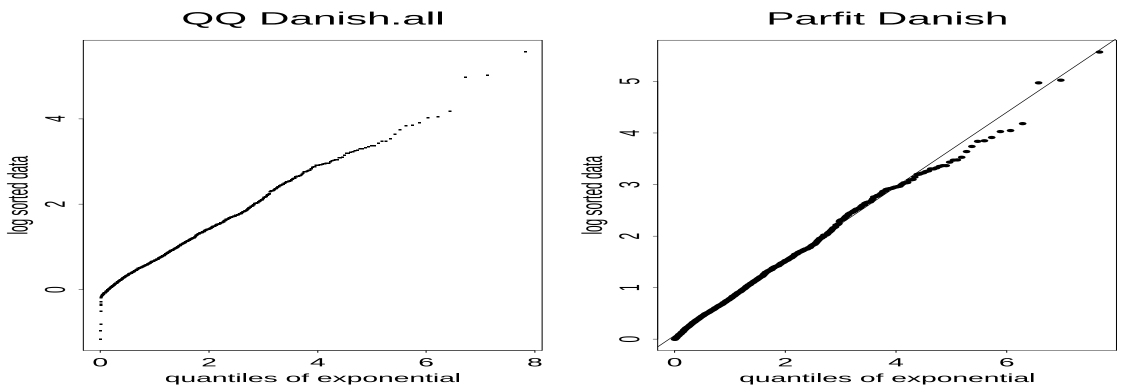
Comparing quantiles of the return distribution's left tail giving the approximation to VaR.

Case Study: Danish Fire Data

danish.all=total loss per event for claims 1980-1990 in 1985 krone; 2492 losses.
danish=exceedances over 1 million krone; 2156 losses.

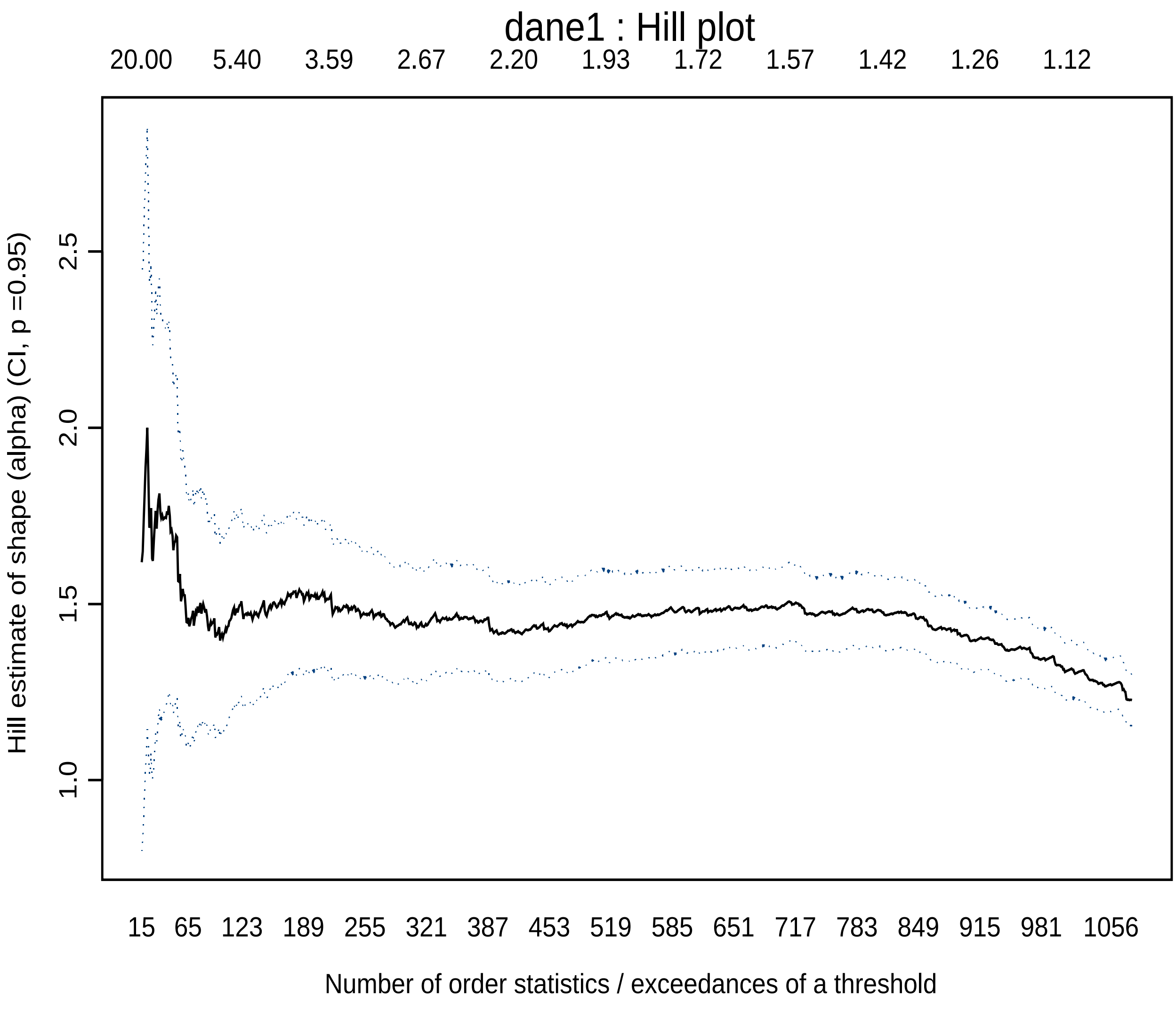


Tsplot and QQ-plot of Danish data.



QQ-plot Danish.all; parameter estimate gives $\hat{\alpha} = 1.38$ (infinite variance).

Hill plots. The QQ & Hill plots so stable, underlying distribution close to Pareto. Confirms estimate $\widehat{\alpha} \approx 1.4$.



Independence? Sample acf: Exploratory, informal method for testing for independence based on the *sample autocorrelation function* $\hat{\rho}_(h)$ where

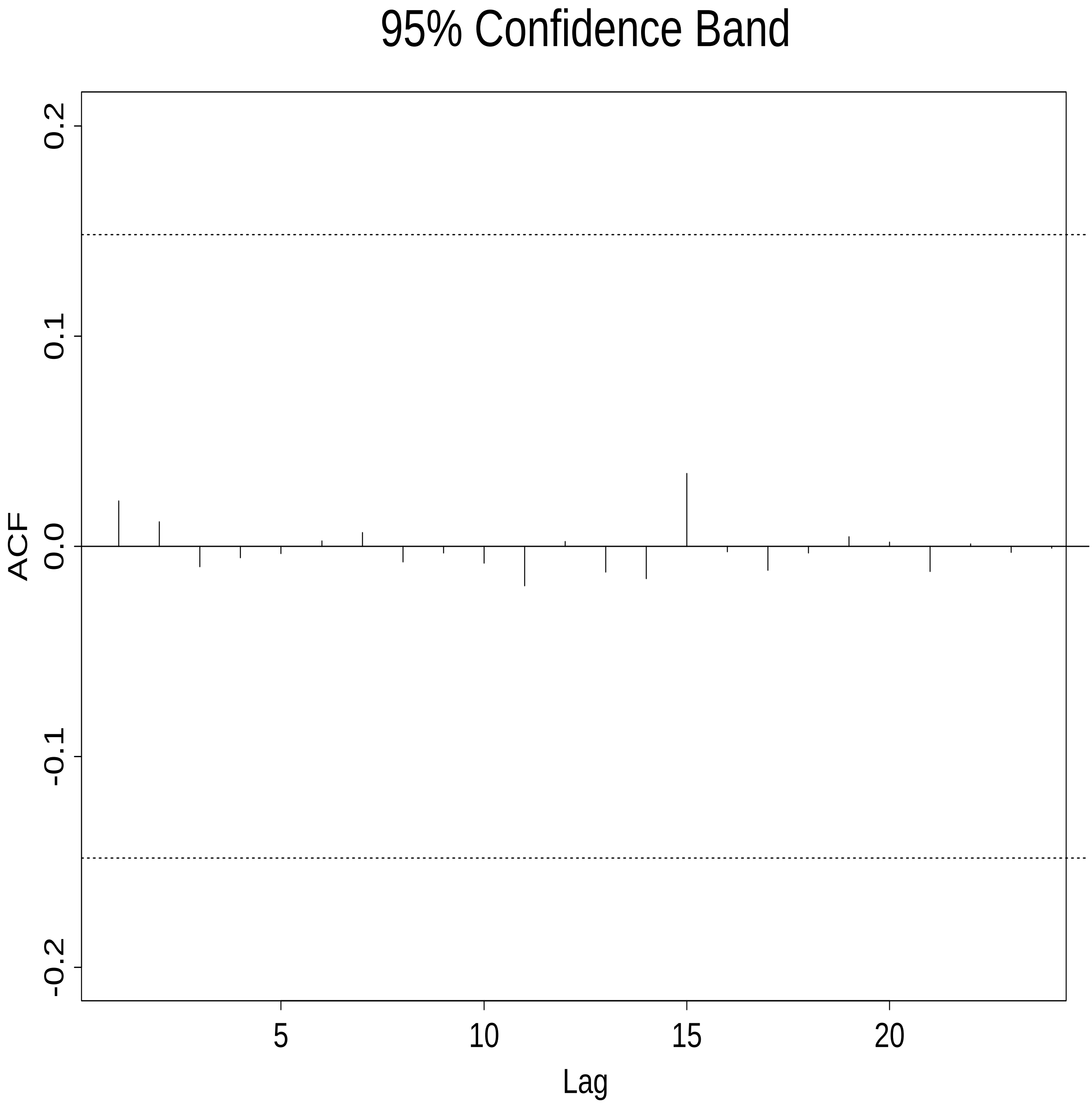
$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2}.$$

Note variance is infinite so mathematical correlations do not exist. However, when $\{X_n\}$ iid, heavy tailed (Davis and Resnick (1985a)),

$$\lim_{n\to\infty} \hat{\rho}(h) = \begin{cases} 1, & \text{if } h = 0, \\ 0, & \text{if } h \neq 0. \end{cases}$$

and for $h > 0$, $\hat{\rho}(h)$, suitably scaled, has a limit distribution corresponding to the ratio of 2 stable random variables.

Obtain quantiles of the limit distribution; form CI for $\hat{\rho}(h)$. Get magic window acf plot.

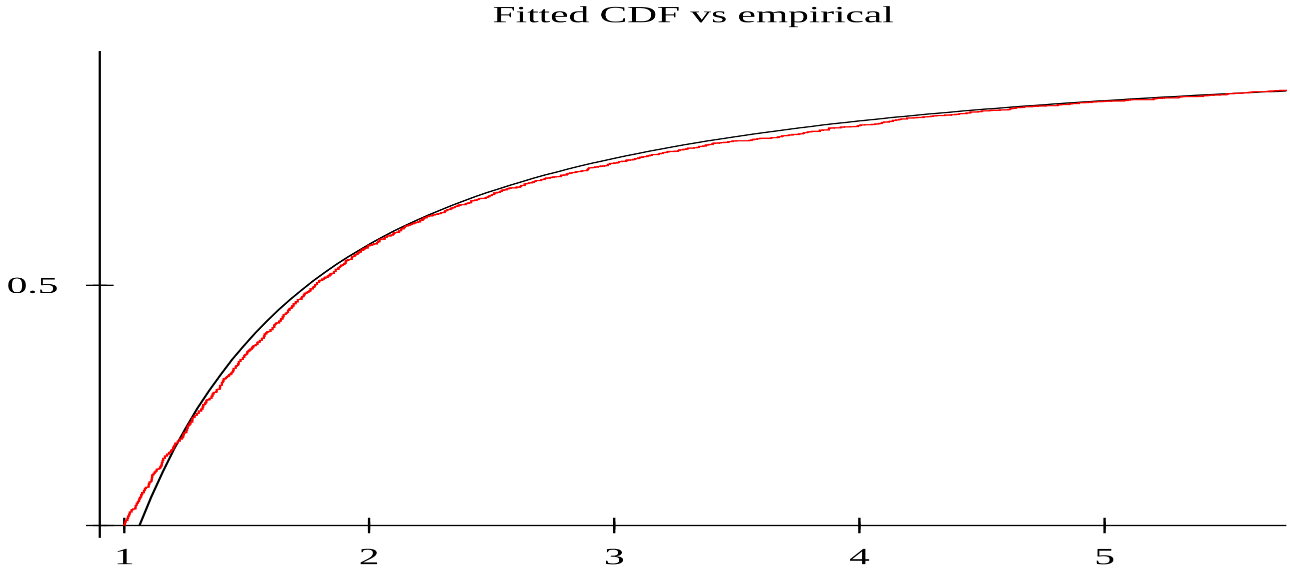


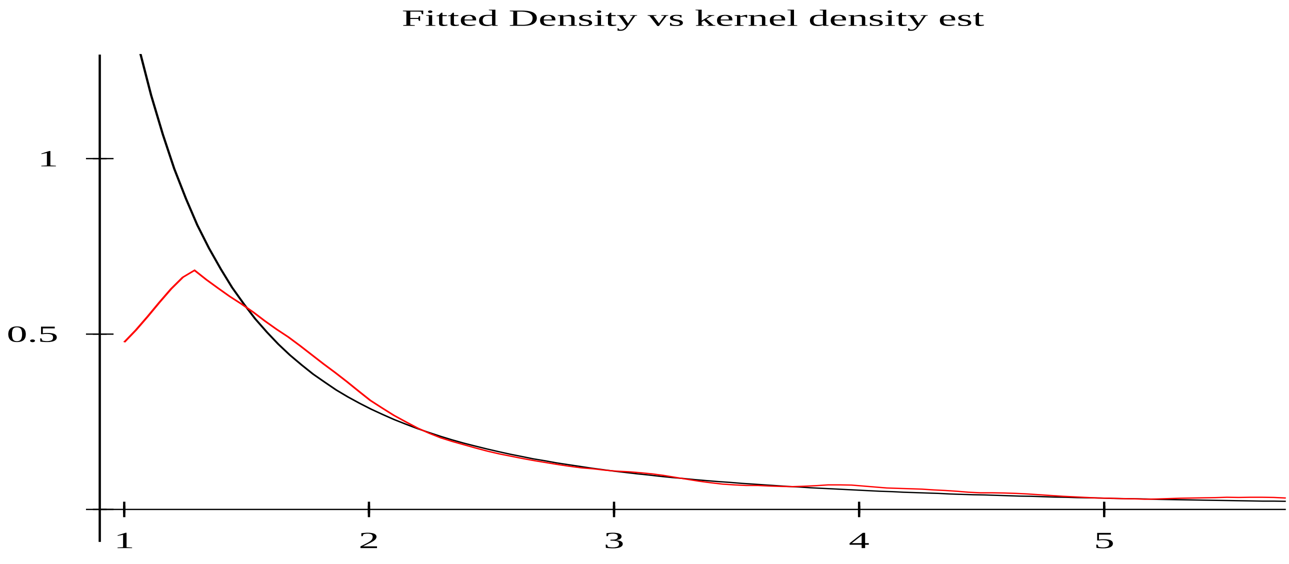
95% confidence band for the acf of the Danish loss data.

QQ-plot for exceedances so straight that hope gpd fits well:

black=fitted gpd; $\hat{\alpha} = 1.39$, $\hat{\sigma} = 1.06$,

red=empirical





Finale

From here can

$\rightarrow$ calculate mean excess of loss
$\rightarrow$ calculate quantiles
$\rightarrow$ evaluate sensitivity to choice of threshold

Conclusion: EVT offers

- a useful tool for extreme tail and quantile estimation beyond the range of the data.

- a technique with sound theoretical basis

- good fits

- potential reduction in ad hoc techniques

BUT!!

- model uncertainty

- parameter uncertainty

- sensitivity to choice of threshold or choice of number of upper order statistics

- when estimating beyond the range of the data, some religious conviction is helpful;

- dependencies ($\Rightarrow$ clustering) should be taken into account in more subtle analyses

Thanks.