5 Extreme value theory

For more theoretical details, see Embrechts et al. (1997), especially Chapters 2 and 3.

5.1 Maxima

Consider losses $(X_k)_{k\in\mathbb{N}}$ (e.g., negative log-returns).

5.1.1 Generalized extreme value distribution

Convergence of sums

Let $(X_k)_{k\in\mathbb{N}}$ be i.i.d. with $\mathbb{E}[X_1^2] < \infty$ (mean μ, variance σ^2) and

$$S_n = \sum_{k=1}^{n} X_k.$$

Note that $\bar{X}_n \underset{n\uparrow\infty}{\overset{\text{a.s.}}{\to}} \mu$ by the Strong Law of Large Numbers (SLLN), so $(\bar{X}_n - \mu)/\sigma \underset{n\uparrow\infty}{\overset{\text{a.s.}}{\to}} 0$. By the CLT,

$$\sqrt{n}\frac{\bar{X}_n-\mu}{\sigma}=\frac{S_n-n\mu}{\sqrt{n}\sigma}\underset{n\uparrow\infty}{\overset{\mathsf{d}}{\to}}\mathrm{N}(0,1),$$

that is,

$$\lim_{n\to\infty}\mathbb{P}((S_n-d_n)/c_n\le x)=\Phi(x),\quad x\in\mathbb{R},$$

i.e., $c_n=\sqrt{n}\sigma$, $d_n=n\mu$, $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^x e^{-z^2/2}dz$. More generally ($\sigma^2=\infty$), the limiting distributions for appropriately normalized sums are the class of α-stable distributions ($\alpha\in(0,2]$; $\alpha=2$: normal distribution).

Convergence of maxima

QRM is concerned with maximal losses (worst-case losses). Let $(X_i)_{i\in\mathbb{N}}\overset{\text{ind.}}{\sim}F$ (can be relaxed to a strictly stationary time series) and F continuous. Then the *block maxima* is given by

$$M_n=\max\{X_1,\dots,X_n\}.$$

Clearly, $\mathbb{P}(M_n\le x)=\mathbb{P}(X_1\le x,\dots,X_n\le x)=F^n(x)\underset{n\uparrow\infty}{\to}\mathbb{1}_{\{x\ge x_F\}}$, where $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} = F^{-}(1) \le \infty$ denotes the *right endpoint of F*; typically $x_F = \infty$ in finance/insurance. Thus, $M_n \underset{n\uparrow\infty}{\overset{\mathrm{d}}{\to}} x_F$.

One can show that, since x_F is a constant, $M_n \underset{n\uparrow\infty}{\overset{\mathrm{P}}{\to}} x_F$, and, since $(M_n) \nearrow$, $M_n \underset{n\uparrow\infty}{\overset{\text{a.s.}}{\to}} x_F$ (similar as in the SLLN). Is there a "CLT" for block maxima?

Idea CLT: What about linear transformations (the simplest possible)?

Definition 5.1 (Maximum domain of attraction)
Suppose we find *normalizing sequences* of real numbers $(c_n) > 0$ and (d_n) such that $(M_n - d_n)/c_n$ converges in distribution, i.e.,

$$\mathbb{P}((M_n - d_n)/c_n \le x) = \mathbb{P}(M_n \le c_n x + d_n) = F^n(c_n x + d_n) \underset{n\uparrow\infty}{\to} H(x),$$

for some *non-degenerate (n.d.)* df H (not a unit jump). Then F is in the *maximum domain of attraction* of H ($F \in \mathrm{MDA}(H)$).

H is determined up to location/scale, i.e., H specifies a unique *type* of distribution. In particular, we can always choose $(c_n), (d_n)$ such that the limit of $\frac{M_n - d_n}{c_n}$ appears in a location-scale transformed way.

Theorem 5.2 (Convergence to Types)

Suppose $(M_n)_n$ is a sequence of rvs such that $\frac{M_n - d_n}{c_n} \xrightarrow{\text{d}} Y$ for a rv Y and $d_n \in \mathbb{R}$, $c_n > 0$. Then

$$\frac{M_n - \delta_n}{\gamma_n} \xrightarrow{\text{d}} Z$$

for a rv Z and $\delta_n \in \mathbb{R}$, $\gamma_n > 0$ if and only if

$$(c_n/\gamma_n) \to c \in [0, \infty), \quad (d_n - \delta_n)/\gamma_n \to d \in \mathbb{R},$$

in which case $Z \overset{\text{d}}{=} cY + d$ (i.e., Y and Z *are of the same type*) and c, d are the unique such constants.

Proof. See Embrechts et al. (1997, p. 554). □

How does H look like? (Fisher and Tippett (1928), Gnedenko (1943))

Theorem 5.3 (Fisher–Tippett–Gnedenko)
If $F \in \mathrm{MDA}(H)$ for some n.d. H, then H must be of GEV type, i.e., $H = H_\xi$ for some $\xi \in \mathbb{R}$ (see later).

Proof. Non-trivial. For a sketch, see Embrechts et al. (1997, p. 122). □

- **Interpretation:** If location-scale transformed maxima converge in distribution to a n.d. limit, the limit distribution must be a GEV distribution.
- We can always choose normalizing sequences $(c_n) > 0$, (d_n) such that H_ξ appears in standard form.
- All commonly encountered continuous distributions are in the MDA of a GEV distribution.
- The following is often a useful result:
$$\lim_{n\uparrow\infty} F^n(c_n x+d_n) = H(x) \underset{-\log x\approx 1-x}{\overset{-\log(\cdot)}{\Longleftrightarrow}} \lim_{n\uparrow\infty} n\bar{F}(c_n x+d_n) = -\log H(x).$$

Definition 5.4 (Generalized extreme value (GEV) distribution)

The (standard) *generalized extreme value (GEV) distribution* is given by

$$H_\xi(x) = \begin{cases} \exp(-(1+\xi x)^{-1/\xi}), & \text{if } \xi \neq 0, \\ \exp(-e^{-x}), & \text{if } \xi = 0, \end{cases}$$

where $1 + \xi x > 0$ (MLE!). A three-parameter family is obtained by a location-scale transform $H_{\xi,\mu,\sigma}(x) = H_\xi((x-\mu)/\sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

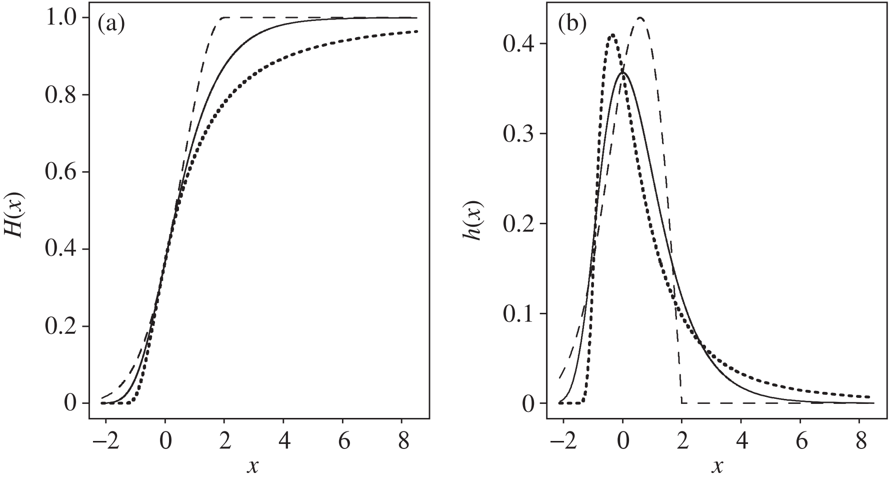
- The parameterization is continuous in ξ (simplifies statistical modeling).
- The larger ξ, the heavier tailed H_ξ (if $\xi > 0$, $\mathbb{E}[X^k] = \infty$ iff $k \geq \frac{1}{\xi}$).
- ξ is the *shape* (determines moments, tail). Special cases:
  1) $\xi < 0$: the Weibull df, short-tailed, $x_{H_\xi} < \infty$;
  2) $\xi = 0$: the Gumbel df, $x_{H_0} = \infty$, decays exponentially;
  3) $\xi > 0$: the Fréchet df, $x_{H_\xi} = \infty$, heavy-tailed ($\bar{H}_\xi(x) \approx (\xi x)^{-1/\xi}$), most important case for practice (typically, $\xi \in (1/5, 1/2)$).

- For $1+\xi x>0$, the density h_ξ of H_ξ is given by

$$h_\xi(x)=\begin{cases}(1+\xi x)^{-1/\xi-1}H_\xi(x), & \text{if } \xi\neq 0,\\ e^{-x}H_0(x), & \text{if } \xi=0.\end{cases}$$

- One can show that *tail equivalence* $\lim_{x\uparrow x_F=x_G}\frac{\bar{F}(x)}{\bar{G}(x)}=c\in(0,\infty)$ implies $F\in\mathrm{MDA}(H)\Leftrightarrow G\in\mathrm{MDA}(H)$ with the same normalizing sequences, i.e., tail equivalent distributions belong to $\mathrm{MDA}(H_\xi)$ for the same ξ.
- Minima: $-X_1,\dots,-X_n\overset{\text{ind.}}{\sim}\bar{F}(-x)=1-F(-x)$. If $\bar{F}(-x)\in\mathrm{MDA}(H_\xi)$, then, properly normalized, the limiting distribution of $\min\{X_1,\dots,X_n\}=-\max\{-X_1,\dots,-X_n\}$ is a type of $1-H_\xi(-x)$.

(a): H_ξ; (b): density h_ξ; for $\xi \in \{-0.5, 0, 0.5\}$ (dashed, solid, dotted)



Example 5.5 (Exponential distribution)

For $(X_i)_{i\in\mathbb{N}} \overset{\text{ind.}}{\sim} \mathrm{Exp}(\lambda)$, choosing $c_n = 1/\lambda$, $d_n = \log(n)/\lambda$, one obtains

$$\begin{aligned}F^n(c_n x + d_n) &= \big(1 - \exp(-\lambda((1/\lambda)x + \log(n)/\lambda))\big)^n \\ &= (1 - \exp(-x)/n)^n \underset{n\uparrow\infty}{\to} \exp(-e^{-x}) = H_0(x).\end{aligned}$$

Therefore, $F \in \mathrm{MDA}(H_0)$ (Gumbel).

Example 5.6 (Pareto distribution)

For $(X_i)_{i\in\mathbb{N}} \overset{\text{ind.}}{\sim} \mathrm{Par}(\theta,\kappa)$ with $F(x) = 1 - (\frac{\kappa}{\kappa+x})^\theta$, $x \ge 0$, $\theta,\kappa > 0$, choosing $c_n = \kappa n^{1/\theta}/\theta$, $d_n = \kappa(n^{1/\theta} - 1)$, one obtains

$$\begin{aligned}F^n(c_n x + d_n) &= \Bigg(1 - \bigg(\frac{\kappa}{\kappa + x\kappa n^{1/\theta}/\theta + \kappa(n^{1/\theta} - 1)}\bigg)^\theta\Bigg)^n \\ &= \Bigg(1 - \bigg(\frac{1}{1 + xn^{1/\theta}/\theta + n^{1/\theta} - 1}\bigg)^\theta\Bigg)^n \\ &= \Bigg(1 - \frac{(1/(x/\theta))^\theta}{n}\Bigg)^n = \Bigg(1 - \frac{(\theta/x)^\theta}{n}\Bigg)^n \underset{n\uparrow\infty}{\to} \exp(-(\theta/x)^\theta),\end{aligned}$$

which equals $H_{1/\theta,\theta,1}$ and hence $F \in \mathrm{MDA}(H_{1/\theta})$ (Fréchet).

We could have equally well chosen $c_n = \kappa(n^{1/\theta} - 1)$ and $d_n = 0$, since

$$
\begin{aligned}
F^n(c_n x + d_n) &= \Bigl(1 - \Bigl(\frac{\kappa}{\kappa + \kappa(n^{1/\theta} - 1)x}\Bigr)^{\theta}\Bigr)^n \\
&= \Bigl(1 - \Bigl(\frac{1}{1 - x + n^{1/\theta}x}\Bigr)^{\theta}\Bigr)^n = \Bigl(1 - \frac{\bigl(\frac{1}{(1-x)/n^{1/\theta}+x}\bigr)^{\theta}}{n}\Bigr)^n \\
&\underset{n\uparrow\infty}{\to} \exp(-(1/x)^{\theta}),
\end{aligned}
$$

which equals $H_{1/\theta,1,1/\theta}$.

Lurking in the background: $(a_n) \in \mathbb{C}$, $a_n \to a \in \mathbb{C} \Rightarrow (1 + a_n/n)^n \to e^a$.

5.1.2 Maximum domains of attraction

All commonly applied continuous F belong to $\mathrm{MDA}(H_\xi)$ for some $\xi \in \mathbb{R}$. μ, σ can be estimated, but how can we characterize/determine ξ? All $F \in \mathrm{MDA}(H_\xi)$ for $\xi > 0$ have an elegant characterization involving the following notions.

Definition 5.7 (Slowly/regularly varying functions)

1) A positive, Lebesgue-measurable function L on $(0, \infty)$ is *slowly varying at* ∞ if $\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1$, $t > 0$. The class of all such functions is denoted by SV; e.g., $c, \log \in \mathrm{SV}$.
2) A positive, Lebesgue-measurable function h on $(0, \infty)$ is *regularly varying at* ∞ *with index* $\alpha \in \mathbb{R}$ if $\lim_{x\to\infty} \frac{h(tx)}{h(x)} = t^\alpha$, $t > 0$. The class of all such functions is denoted by RV_α; $x^\alpha L(x) \in \mathrm{RV}_\alpha$.

The Fréchet case

Theorem 5.8 (Fréchet MDA, Gnedenko (1943))
For $\xi > 0$, $F \in \mathrm{MDA}(H_\xi)$ if and only if $\bar{F}(x) = x^{-1/\xi}L(x)$ for some $L \in \mathrm{SV}$. If $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$, the normalizing sequences can be chosen as $c_n = F^{-}(1 - 1/n)$ and $d_n = 0$, $n \in \mathbb{N}$.

Proof. Non-trivial. For a sketch, see Embrechts et al. (1997, p. 131). □

- **Interpretation:** If $\bar{F} \in \mathrm{RV}_{-1/\xi}$ (decay like a power function; Pareto like), then $F \in \mathrm{MDA}(H_\xi)$ for $\xi > 0$; $\alpha = 1/\xi$ is known as *tail index*.
- L can destroy the power, but not too much (in (statistical) practice it matters, though)
- If $X \sim F$, $X \geq 0$, $\bar{F} \in \mathrm{RV}_{-\alpha}$, $\alpha > 0$ (equivalently, $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$), then $\mathbb{E}[X^k] < \infty$ if $k < \alpha = 1/\xi$, $\mathbb{E}[X^k] = \infty$ if $k > \alpha = 1/\xi$; see Embrechts et al. (1997, p. 568).

- One can show the *von Mises condition*:
  If F has density f with $\lim_{x\uparrow\infty} \frac{xf(x)}{\bar{F}(x)} = \alpha > 0$, then $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$.
- **Examples in $\mathrm{MDA}(H_\xi)$, $\xi > 0$:** inverse gamma, Student t, log-gamma, F, Cauchy, α-stable with $0 < \alpha < 2$, Burr and Pareto

Example 5.9 (Pareto distribution)

For $F = \mathrm{Par}(\theta, \kappa)$, $\bar{F}(x) = (\kappa/(\kappa + x))^\theta = (1 + x/\kappa)^{-\theta} = x^{-\theta}L(x)$, $x \ge 0$, $\theta, \kappa > 0$, where $L(x) = (\kappa^{-1} + x^{-1})^{-\theta} \in \mathrm{SV}$. We see (again) that $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$.

The Gumbel case

The characterization of this class is more complicated. One can show the following result non-trivial; see Embrechts et al. (1997, p. 142).

Theorem 5.10 (Gumbel MDA)

$F \in \mathrm{MDA}(H_0)$ if and only if there exists $z < x_F \le \infty$ such that

$$\bar{F}(x) = c(x) \exp\Big(-\int_z^x \frac{g(t)}{a(t)}\,dt\Big), \quad x \in (z, x_F),$$

where c and g are measurable functions satisfying $c(x) \to c > 0$, $g(x) \to 1$ for $x \uparrow x_F$ and $a(x) > 0$ with density a' satisfying $\lim_{x\uparrow x_F} a'(x) = 0$.

If $F \in \mathrm{MDA}(H_0)$, the normalizing sequences can be chosen as $c_n = a(d_n)$ for $a(x) = \int_x^{x_F} \bar{F}(t)\,dt/\bar{F}(x)$, $x < x_F$, (the mean excess function), and $d_n = F^{-}(1 - 1/n)$, $n \in \mathbb{N}$.

- Essentially $\mathrm{MDA}(H_0)$ contains dfs whose tails decay roughly exponentially *(light-tailed)*, but the tails can be quite different (up to moderately heavy). All moments exist for distributions in the Gumbel class, but both $x_F < \infty$ and $x_F = \infty$ are possible.

- **Examples in $\mathrm{MDA}(H_0)$:** normal, log-normal, exponential, gamma (exponential, Erlang, χ^2), standard Weibull, Benktander type I and II, generalized hyperbolic (not: Student t).

The Weibull case

Theorem 5.11 (Weibull MDA)
For $\xi < 0$, $F \in \mathrm{MDA}(H_\xi)$ if and only if $x_F < \infty$ and $\bar{F}(x_F - 1/x) = x^{1/\xi} L(x)$ for some $L \in \mathrm{SV}$. If $F \in \mathrm{MDA}(H_\xi)$, $\xi < 0$, the normalizing sequences can be chosen as $c_n = x_F - F^{-}(1 - 1/n)$ and $d_n = x_F$, $n \in \mathbb{N}$.

Proof. Non-trivial. For a sketch, see Embrechts et al. (1997, p. 135). □

- One can show the *von Mises condition*:
  If F has density f which is positive on some finite interval (z, x_F) and if $\lim_{x \uparrow x_F} \frac{(x_F - x) f(x)}{\bar{F}(x)} = \alpha > 0$, then $F \in \mathrm{MDA}(H_\xi)$, $\xi < 0$.

- **Examples in $\mathrm{MDA}(H_\xi)$, $\xi < 0$:** beta (uniform). All $F \in \mathrm{MDA}(H_\xi)$, $\xi < 0$, share $x_F < \infty$.

5.1.3 Maxima of strictly stationary time series

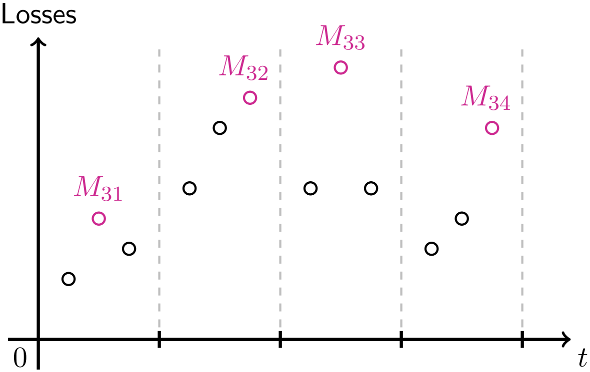
What about maxima of strictly stationary time series?

- Let $(X_k)_{k\in\mathbb{Z}}$ denote a strictly stationary time series with stationary distribution $X_k \sim F$, $k \in \mathbb{Z}$.
- Let $\tilde{X}_k \overset{\text{ind.}}{\sim} F$, $k \in \mathbb{Z}$, and $\tilde{M}_n = \max\{\tilde{X}_1, \dots, \tilde{X}_n\}$. For many processes one can show that there exists a real number $\theta \in (0, 1]$ such that $\lim_{n\uparrow\infty} \mathbb{P}((M_n - d_n)/c_n \leq x) = H^\theta(x)$ if and only if $\lim_{n\uparrow\infty} \mathbb{P}((\tilde{M}_n - d_n)/c_n \leq x) = H(x)$ (n.d.); θ is known as *extremal index*.
- If $F \in \mathrm{MDA}(H_\xi)$ for some $\xi \Rightarrow \tilde{M}_n$ converges in distribution to H_ξ $\Rightarrow M_n$ converges in distribution to H_ξ^θ. Since H_ξ^θ is of the same type as H_ξ, the limiting distribution of the block maxima of the dependent series is the same as in the i.i.d. case (only location/scale may change).

- For large n, $\mathbb{P}((M_n - d_n)/c_n \leq x) \approx H^{\theta}(x) \approx F^{n\theta}(c_n x + d_n)$, so the distribution of M_n from a time series with extremal index θ can be approximated by the distribution $\tilde{M}_{n\theta}$ of the maximum of $n\theta < n$ observations from the associated i.i.d. series. $\Rightarrow n\theta$ counts the number of roughly independent clusters in n observations (θ is often interpreted as “1/mean cluster size”).
- If $\theta = 1$, large sample maxima behave as in the i.i.d. case; if $\theta \in (0, 1)$, large sample maxima tend to cluster.
- **Examples** (see Embrechts et al. (1997, pp. 216, pp. 415, pp. 476))
  - Strict white noise (iid rvs): $\theta = 1$;
  - ARMA processes with (ε_t) strict white noise: $\theta = 1$ (Gaussian); $\theta \in (0, 1)$ (if df of ε_t is in $\mathrm{MDA}(H_\xi)$, $\xi > 0$);
  - (G)ARCH processes: $\theta \in (0, 1)$.

5.1.4 The block maxima method (BMM)

The basic idea in a picture based on losses $X_1, \dots, X_{12}$:



Consider the maximal loss from each block and fit $H_{\xi,\mu,\sigma}$ to them.

Fitting the GEV distribution

- Suppose $(x_i)_{i\in\mathbb{N}}$ are realizations of $(X_i)_{i\in\mathbb{N}} \overset{\text{ind.}}{\sim} F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$, or of a process with an extremal index such as GARCH. The Fisher–Tippett–Gnedenko Theorem implies that

$$\mathbb{P}(M_n \le x) = \mathbb{P}((M_n - d_n)/c_n \le (x - d_n)/c_n) \underset{n\,\text{large}}{\approx} H_{\xi,\mu=d_n,\sigma=c_n}(x).$$

- For fitting $\boldsymbol{\theta} = (\xi, \mu, \sigma)$, we assume our realizations can be divided into m blocks of size n denoted by $M_{n1}, \dots, M_{nm}$ (often naturally the case, e.g., in hydrology: daily water levels $\Rightarrow$ yearly maxima; in finance: daily log-returns $\Rightarrow$ monthly/quarterly/yearly maxima).
- Assume the block size n to be sufficiently large so that (regardless of whether the underlying data are dependent or not), the block maxima can be considered independent.

- The log-likelihood is

$$\ell(\boldsymbol{\theta}; M_{n1}, \dots, M_{nm}) = \sum_{i=1}^{m} \log\Bigl(\frac{1}{\sigma} h_\xi\Bigl(\frac{M_{ni}-\mu}{\sigma}\Bigr)\mathbb{1}_{\{1+\xi(M_{ni}-\mu)/\sigma>0\}}\Bigr).$$

Maximize w.r.t. $\boldsymbol{\theta} = (\xi, \mu, \sigma)$ to get $\hat{\boldsymbol{\theta}} = (\hat{\xi}, \hat{\mu}, \hat{\sigma})$.

Remark 5.12

1) Sufficiently many/large blocks require large amounts of data.
2) Bias and variance must be traded off (*bias-variance tradeoff*):
   - Block size $n\uparrow \Rightarrow$ GEV approximation more accurate $\Rightarrow$ bias $\downarrow$
   - Number of blocks $m\uparrow \Rightarrow$ more data for MLE $\Rightarrow$ variance $\downarrow$
3) There is no general best strategy known to find the optimal block size.
4) The support of the density depends on the parameters $\Rightarrow$ not differentiable; classical MLE regularity conditions for consistency and asymptotic efficiency do not applied. For $\xi > -1/2$ (fine for practice), Smith (1985) showed that the MLE is regular.

Example 5.13 (Block maxima analysis of S&P500)

Suppose it is Friday 1987-10-16. The S&P 500 index fell by 9.12% this week. On that Friday alone the index is down 5.16% on the previous day (largest one-day fall since 1962). We fit a GEV distribution to annual maxima of daily negative returns $X_t = S_t/S_{t-1} - 1$ since 1960.

Analysis 1: Based on annual maxima ($m = 28$; including the latest from the incomplete year 1987): $\hat{\boldsymbol{\theta}} = (0.29, 2.03, 0.72) \Rightarrow$ heavy-tailed Fréchet distribution (infinite third moment). The corresponding standard errors are $(0.21, 0.0016, 0.0014) \Rightarrow$ High uncertainty (m small) for estimating ξ.

Analysis 2: Based on biannual maxima ($m = 56$): $\hat{\boldsymbol{\theta}} = (0.33, 1.68, 0.55)$ with standard errors $(0.14, 0.0009, 0.0007) \Rightarrow$ Hints at even heavier tails.

Return levels and stress losses (exceedances)

The fitted GEV model can be used to estimate:

1) The size of an event with prescribed frequency (*return-level problem*)
2) The frequency of an event with prescribed size (*return-period problem*)

Definition 5.14 (Return level, return period)
Let $M_n \sim H$ (exact). The *k n-block return level* is $r_{n,k} = H^{-}(1 - 1/k)$.
The *return period* of the event $\{M_n > u\}$ is $k_{n,u} = 1/\bar{H}(u)$.

- $r_{n,k}$ is the level which is exceeded (on average) in one out of every k n-blocks, so $r_{n,k}$ solves $\mathbb{P}(M_n > r_{n,k}) = 1/k$ (e.g., 10-year return level $r_{260,10}$ = level exceeded in one out of every 10 years; 260d $\approx$ 1 year).
- $k_{n,u}$ is the number of n-blocks for which we expect to see a single n-block exceeding u, so $k_{n,u}$ solves $r_{n,k_{n,u}} = H^{-}(1 - 1/k_{n,u}) = u$.

- Parametric estimators are given by

$$\hat{r}_{n,k} = H^{-}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(1-1/k) = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}((-\log(1-1/k))^{-\hat{\xi}} - 1),$$
$$\hat{k}_{n,u} = 1/\bar{H}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(u).$$

  Confidence intervals for $r_{n,k}$, $k_{n,u}$ can be constructed via profile-likelihoods; see Davison (2003, pp. 126) and McNeil et al. (2005, p. 274).

Example 5.15 (Block maxima analysis of S&P500 (continued))

- The 10-year return level $r_{260,10}$ based on data up to and including Friday 1987-10-16 is estimated as $\hat{r}_{260,10} = 4.32\%$. The next trading day is Black Monday (1987-10-19), the event of an index drop of 20.47% is far beyond $\hat{r}_{260,10}$. One can show that 20.47% is in the 95% confidence interval of $r_{260,50}$ (estimated as $\hat{r}_{260,50} = 7.23\%$), but the 28 maxima are too few to get a reliable estimate of a once-in-50-years event.

- If we estimate the return period $k_{260,0.2047}$ of a loss of 20.47%, the point estimate is $\hat{k}_{260,0.2047} = 1629$ years. One can show that the 95%

confidence interval encompasses everything from 45 years to essentially never! $\Rightarrow$ Very high uncertainty involved in estimating $k_{260,0.2047}$.

- In summary, on 1987-10-16 we simply did not have enough data to say anything meaningful about an event of this magnitude. This illustrates the difficulties of attempting to quantify events beyond our empirical experience.

5.2 Threshold exceedances

The BMM is wasteful of data (only the maxima of large blocks are used). It has been largely superseded in practice by methods based on threshold exceedances (*peaks-over-threshold (POT)* approach), where all data above a designated high threshold u are used.

5.2.1 Generalized Pareto distribution

Definition 5.16 (Generalized Pareto distribution (GPD))

The *generalized Pareto distribution (GPD)* is given by

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi}, & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0, \end{cases}$$

where $\beta > 0$, and the support is $x \geq 0$ when $\xi \geq 0$ and $x \in [0, -\beta/\xi]$ when $\xi < 0$.

- The parameterization is continuous in ξ.
- The larger ξ, the heavier tailed $G_{\xi,\beta}$ (if $\xi > 0$, $\mathbb{E}[X^k] = \infty$ iff $k \geq \frac{1}{\xi}$; if $\xi < 1$, then $\mathbb{E}[X] = \beta/(1-\xi)$).
- ξ is known as *shape*; β as *scale*. Special cases:

  1) $\xi > 0$: $\mathrm{Par}(1/\xi, \beta/\xi)$

  2) $\xi = 0$: $\mathrm{Exp}(1/\beta)$

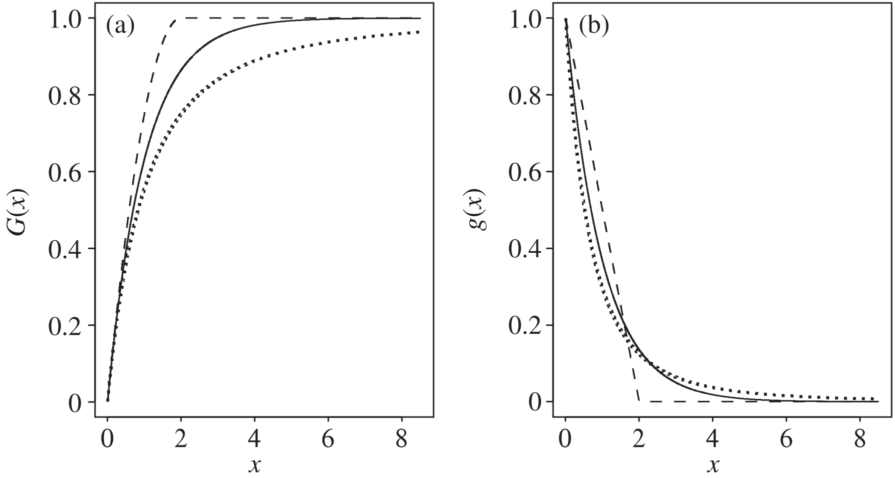
  3) $\xi < 0$: short-tailed Pareto type II distribution

- The density $g_{\xi,\beta}$ of $G_{\xi,\beta}$ is given by

$$g_{\xi,\beta}(x) = \begin{cases} \frac{1}{\beta}(1+\xi x/\beta)^{-1/\xi-1}, & \text{if } \xi \neq 0, \\ \frac{1}{\beta}\exp(-x/\beta), & \text{if } \xi = 0, \end{cases}$$

  where $x \geq 0$ when $\xi \geq 0$ and $x \in [0, -\beta/\xi)$ when $\xi < 0$ (MLE!).

- $G_{\xi,\beta} \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$ (follows from Theorems 5.8, 5.10 and 5.11) (a): $G_{\xi,1}$; (b): density $g_{\xi,1}$; for $\xi \in \{-0.5, 0, 0.5\}$ (dashed, solid, dotted)



Definition 5.17 (Excess distribution over u, mean excess function)

Let $X \sim F$. The *excess distribution over the threshold u* is defined by

$$F_u(x) = \mathbb{P}(X - u \leq x \,|\, X > u) = \frac{F(x+u) - F(u)}{1 - F(u)}, \quad x \in [0, x_F - u).$$

If $\mathbb{E}|X| < \infty$, the *mean excess function* is defined by

$$e(u) = \mathbb{E}[X - u \,|\, X > u] \quad \text{(mean w.r.t. } F_u\text{)}$$

Interpretation

F_u describes the distribution of the loss over u (excess), given that u is exceeded. $e(u)$ is the mean of F_u as a function in u.

- One can show the useful formula $e(u) = \frac{1}{\bar{F}(u)} \int_u^{x_F} \bar{F}(x)\, dx$.
- For continuous $X \sim F$ with $\mathbb{E}|X| < \infty$, the following formula holds:

$$\mathrm{ES}_\alpha(X) = e(\mathrm{VaR}_\alpha(X)) + \mathrm{VaR}_\alpha(X), \quad \alpha \in (0,1); \qquad (12)$$

- The results of the following example are easy to check.

Example 5.18 (F_u, $e(u)$ for $\mathrm{Exp}(\lambda)$, $G_{\xi,\beta}$)

1) If F is $\mathrm{Exp}(\lambda)$, then $F_u(x) = 1 - e^{-\lambda x}$, $x \ge 0$ (so again $\mathrm{Exp}(\lambda)$; lack-of-memory property). The mean excess function is $e(u) = 1/\lambda = \mathbb{E}X$.

2) If F is $G_{\xi,\beta}$, then $F_u(x) = G_{\xi,\beta+\xi u}(x)$, $x \ge 0$ (so again GPD, with the same shape, only the scale grows linearly in u) $\Rightarrow$ Important for (re)insurance (u denotes the threshold determined by an insurance contract; everything above needs to be covered by reinsurance). This will also allow us to conduct estimation of risk measures lower in the tail and then scale up (see later; one of the core applications of EVT).

- The mean excess function of $G_{\xi,\beta}$ is

$$e(u) = \frac{\beta + \xi u}{1 - \xi}, \quad \text{for all } u : \beta + \xi u > 0,$$

which is linear in u (this is a characterizing property of the GPD and used to determine u). Note that ξ determines the slope of $e(u)$.

Theorem 5.19 (Pickands–Balkema–de Haan (1974/75))
There exists a positive, measurable function $\beta(u)$, such that

$$\lim_{u \uparrow x_F} \sup_{0 \le x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0.$$

if and only if $F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$.

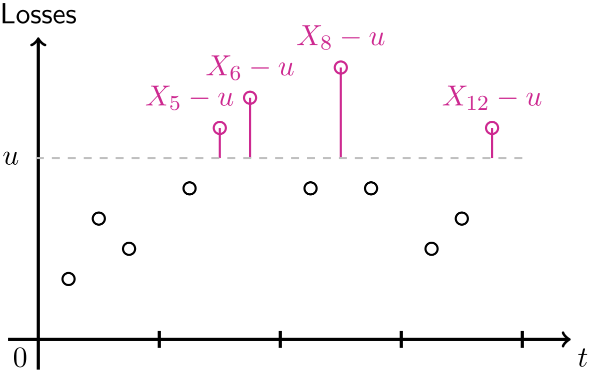
Proof. Non-trivial; see, e.g., Pickands (1975) and Balkema and de Haan (1974). □

Interpretation

- GPD = Canoncial df for modeling excess losses over high u.
- The result is also a characterization of $\mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$. All $F \in \mathrm{MDA}(H_\xi)$ form a set of df for which the excess distribution converges to the GPD $G_{\xi,\beta}$ with the same ξ as in H_ξ as the threshold u is raised.

5.2.2 Modeling excess losses

The basic idea in a picture based on losses $X_1, \dots, X_{12}$.



Consider all excesses over u and fit $G_{\xi,\beta}$ to them.

The method

- Given losses $X_1, \dots, X_n \sim F \in \mathrm{MDA}(H_\xi)$, $\xi \in \mathbb{R}$, let
  - $N_u = |\{i \in \{1, \dots, n\} : X_i > u\}|$ denote the *number of exceedances* over the (given; see later) threshold u;
  - $\tilde{X}_1 < \dots < \tilde{X}_{N_u}$ denote the ordered *exceedances*; and
  - $Y_k = \tilde{X}_k - u$, $k \in \{1, \dots, N_u\}$, the corresponding *excesses*.
- If $Y_1, \dots, Y_{N_u}$ are i.i.d. and (roughly) distributed as $G_{\xi,\beta}$, the log-likelihood is given by

$$\ell(\xi, \beta; Y_1, \dots, Y_{N_u}) = \sum_{k=1}^{N_u} \log g_{\xi,\beta}(Y_k)$$

$$= -N_u \log(\beta) - (1 + 1/\xi) \sum_{k=1}^{N_u} \log(1 + \xi Y_k/\beta)$$

$\Rightarrow$ Maximize w.r.t. $\beta > 0$ and $1 + \xi Y_k/\beta > 0$ for all $k \in \{1, \dots, N_u\}$.

Non-i.i.d. data

- If $X_1, \dots, X_n$ are serially dependent and show no tendency of clusters of extreme values (extremal index $\theta = 1$), asymptotic theory of point processes suggests a limiting model for high-level threshold exceedances, in which exceedances occur according to a Poisson process and the excess losses are i.i.d. generalized Pareto distributed.
- If extremal clustering is present ($\theta < 1$; e.g., (G)ARCH processes), the assumption of independent excess losses is less satisfactory. Easiest approach: neglect the problem, simply apply MLE which is then a quasi-MLE (QMLE) (likelihood misspecified); point estimates should still be reasonable, standard errors may be too small.
- See Section 5.3 for more details on threshold exceedances.

Excesses over higher thresholds

Once a model is fitted to F_u, we can infer a model for F_v, $v \geq u$.

Lemma 5.20

Assume, for some u, $F_u(x) = G_{\xi,\beta}(x)$ for $0 \leq x < x_F - u$. Then $F_v(x) = G_{\xi,\beta+\xi(v-u)}(x)$ for all $v \geq u$.

Proof. Recall that $F_u(x) = \mathbb{P}(X - u \leq x \,|\, X > u) = \frac{F(u+x)-F(u)}{\bar{F}(u)}$, so $\bar{F}_u(x) = \bar{F}(u+x)/\bar{F}(u)$. For $v \geq u$, we have

$$\bar{F}_v(x) = \frac{\bar{F}(v+x)}{\bar{F}(v)} = \frac{\bar{F}(u+(v+x-u))}{\bar{F}(u)} \frac{\bar{F}(u)}{\bar{F}(u+(v-u))}$$
$$= \frac{\bar{F}_u(v+x-u)}{\bar{F}_u(v-u)} = \frac{\bar{G}_{\xi,\beta}(x+v-u)}{\bar{G}_{\xi,\beta}(v-u)} \underset{\text{check}}{=} \bar{G}_{\xi,\beta+\xi(v-u)}(x) \qquad \square$$

$\Rightarrow$ The excess distribution over $v \geq u$ remains GPD with the same ξ (and β growing linearly in v); makes sense for a limiting distribution for $u \uparrow$.

If $\xi < 1$, the mean excess function is given by

$$e(v) = \frac{\xi}{1-\xi} v + \frac{\beta - \xi u}{1-\xi}, \quad v \in [u, \infty) \text{ if } \xi \in [0,1), \tag{13}$$

and $v \in [u, u - \beta/\xi]$ if $\xi < 0$. This forms the bases for a graphical method for choosing u.

Sample mean excess plot and choice of the threshold

Definition 5.21 (Sample mean excess function, mean excess plot)
Based on positive loss data $X_1, \dots, X_n$, the *sample mean excess function* is defined by

$$e_n(v) = \frac{\sum_{i=1}^n (X_i - v)\mathbb{1}_{\{X_i > v\}}}{\sum_{i=1}^n \mathbb{1}_{\{X_i > v\}}}, \quad X_{(n)} > v.$$

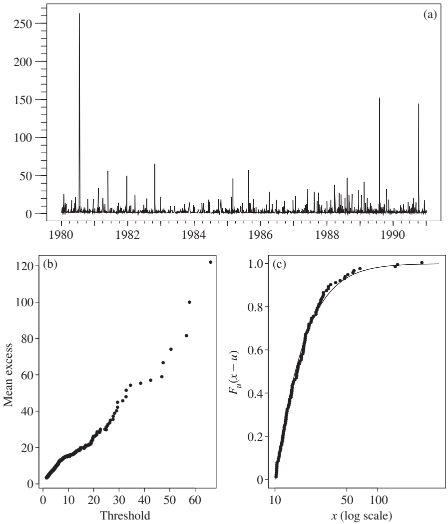
The *mean excess plot* is the plot of $\{(X_{(i)}, e_n(X_{(i)})) : 1 \le i \le n-1\}$, where $X_{(i)}$ denotes the ith order statistic.

- If the data supports the GPD model over u, the $e_n(v)$ should become increasingly "linear" for higher values of u. An upward/zero/downward trend indicates $\xi > 0/\xi = 0/\xi < 0$.
- The sample mean excess plot is rarely perfectly linear (particularly for large u where one averages over a small number of excesses).
- The choice of a good threshold u is as difficult as finding an adequate block size for the Block Maxima method. There are data-driven tools (e.g., sample mean excess plot), but there is no general method to determine an optimal threshold (without second-order assumptions on $L \in \mathrm{SV}$).
- Typically, select u as the smallest point where $e_n(v)$, $v \geq u$, becomes linear. Rule-of-thumb: One needs a couple of thousand data points and can often take u around the 0.9-quantile.
- One should always analyze the data for several u and check the sensitivity of the choice of u.

Example 5.22 (Danish fire loss data)

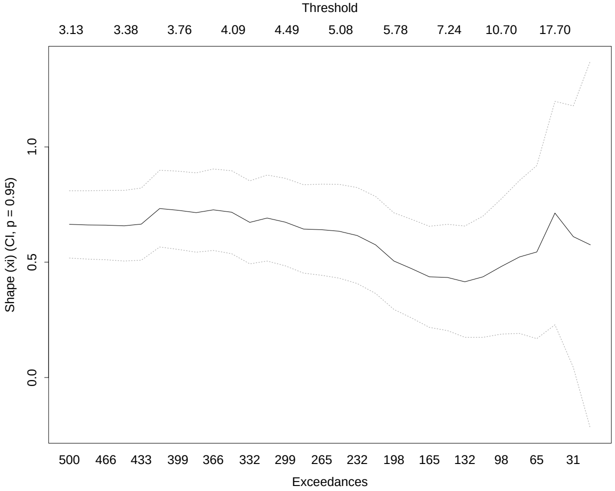
- 2156 fire insurance losses over 1M Danish kroner from 1980-01-03 to 1990-12-31; combined loss for a building and its contents, in some cases also a loss of business earnings. The losses are inflation adjusted to reflect values as of 1985.
- The mean excess function shows a "kink" below 10; "straightening out" above 10 $\Rightarrow$ Our choice is $u = 10$ (so 10M Danish kroner).
- MLE $(\hat{\xi}, \hat{\beta}) = (0.50, 7.0)$ (with standard errors $(0.14, 1.1)$) $\Rightarrow$ very heavy-tailed, infinite-variance model
- We can then estimate the expected loss given exceedance of 10M kroner or any higher threshold (via $e(v)$ in (13) based on $\hat{\xi}, \hat{\beta}$ and the chosen u), even beyond the data. $\Rightarrow$ EVT allows us to estimate "in the data" and then "scale up".

(a): Losses ($>$ 1M; in M); (b): $e_n(u)$ ($\uparrow$); (c) empirical $F_u(x-u)$, $G_{\hat{\xi},\hat{\beta}}$



$\Rightarrow$ Choose the threshold $u = 10$

Sensitivity of the estimated shape parameter $\hat{\xi}$ to changes in u:



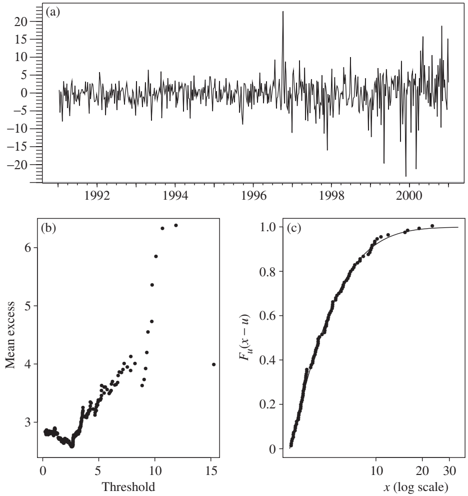
Example 5.23 (AT&T weekly loss data)

- Let (X_t) denote weekly log-returns and consider the percentage one-week loss as a fraction of S_t, given by

$$100L_{t+1}/S_t \underset{(1)}{=} 100(-S_t(\exp(X_{t+1})-1))/S_t = 100(1-\exp(X_{t+1})).$$

- We have 521 such losses (period 1991–2000).
- The estimated GPD parameters are $\hat{\xi} = 0.22$ and $\hat{\beta} = 2.1$ (MLEs) with standard errors 0.13 and 0.34, respectively. The fitted model is thus close to having an infinite fourth moment.
- Note that we ignored here that monthly AT&T data over 1993–2000 is actually not consistent with the i.i.d. assumption (absolute values of log-returns reject the hypothesis of serial uncorrelatedness via the Ljung–Box test).

(a): % losses (1991–2000); (b): $e_n(u)$; (c): empirical $F_u(x-u)$, $G_{\hat{\xi},\hat{\beta}}$.



$\Rightarrow$ Choose the threshold $u = 2.75\%$ (102 exceedances)

5.2.3 Modeling tails and measures of tail risk

- How can the fitted GPD model be used to estimate the tail of the loss distribution F and associated risk measures?
- Assume $F_u(x) = G_{\xi,\beta}(x)$ for $0 \le x < x_F - u$, $\xi \neq 0$ and some u.
- We obtain the following GPD-based formula for tail probabilities:

$$\begin{aligned}\bar{F}(x) &= \mathbb{P}(X > u)\mathbb{P}(X > x \,|\, X > u)\\ &= \bar{F}(u)\mathbb{P}(X - u > x - u \mid X > u) = \bar{F}(u)\bar{F}_u(x-u)\\ &= \bar{F}(u)\Big(1 + \xi\frac{x-u}{\beta}\Big)^{-1/\xi}, \quad x \ge u. \end{aligned} \tag{14}$$

- Assuming we know $\bar{F}(u)$, inverting this formula for $\alpha \ge F(u)$ leads to

$$\mathrm{VaR}_\alpha = F^{-}(\alpha) = u + \frac{\beta}{\xi}\Big(\Big(\frac{1-\alpha}{\bar{F}(u)}\Big)^{-\xi} - 1\Big), \tag{15}$$

$$\mathrm{ES}_\alpha = \frac{\mathrm{VaR}_\alpha}{1-\xi} + \frac{\beta - \xi u}{1-\xi}, \quad \xi < 1. \tag{16}$$

The formula for ES_α can also be obtained from $e(\cdot)$ via (12) and (13).

- $\bar{F}(x)$, VaR_α and ES_α are all of the form $g(\xi, \beta, \bar{F}(u))$. If we have sufficient samples above u, we obtain semi-parametric plug-in estimators via $g(\hat{\xi}, \hat{\beta}, N_u/n)$.
- We hope to gain over empirical estimators by using a kind of extrapolation based on the GPD for more extreme tail probabilities and risk measures.
- For example, based on (14), Smith (1987) proposed the semi-parametric *tail estimator*

$$\hat{\bar{F}}(x) = \frac{N_u}{n}\left(1 + \hat{\xi}\,\frac{x-u}{\hat{\beta}}\right)^{-1/\hat{\xi}}, \quad x \geq u;$$

also known as the *Smith estimator* (note that it is only valid for $x \geq u$).

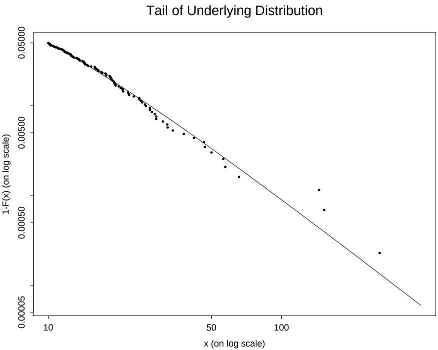
$\Rightarrow$ Bias-variance tradeoff: $u \uparrow \Rightarrow$ bias of parametrically estimating $\bar{F}_u(x-u) \downarrow$, but variance of non-parametrically estimating $\bar{F}(u) \uparrow$

- GPD-based $\widehat{\mathrm{VaR}}_\alpha$, $\widehat{\mathrm{ES}}_\alpha$ for $\alpha \geq 1 - N_u/n$ can be obtained similarly from (15), (16).

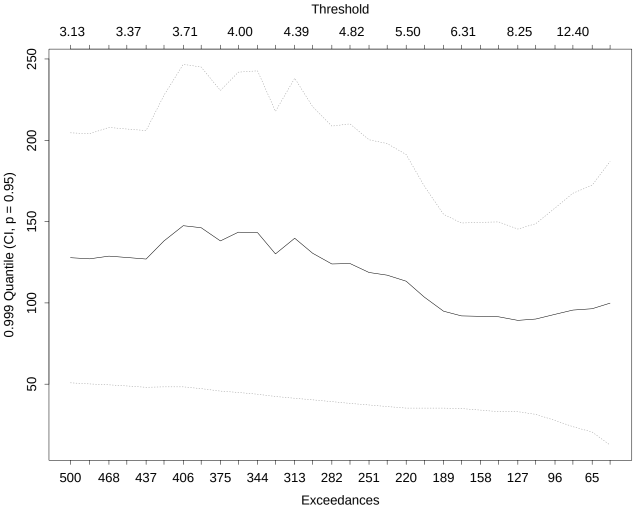
- Confidence intervals for $\bar{F}(x)$, $x \geq u$, VaR_α, ES_α can be obtained likelihood-based (neglecting the uncertainty in N_u/n): Reparametrize the GPD model in terms of $\phi = g(\xi, \beta, N_u/n)$ and construct a confidence interval for ϕ based on the likelihood ratio test.

Example 5.24 (Danish fire loss data (continued))

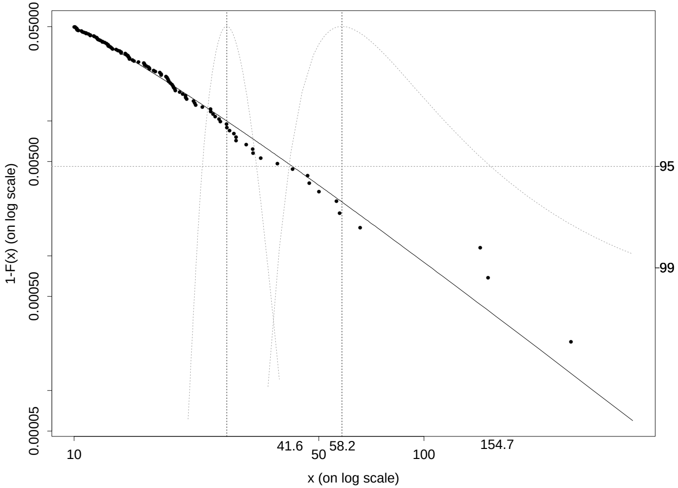
The semi-parametric Smith/tail estimator $\hat{\bar{F}}(x)$, $x \geq u$ is given by:



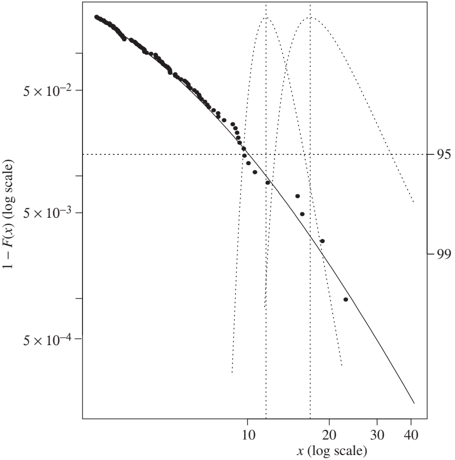
It is important to check the sensitivity of $\hat{\bar{F}}$ (or $\widehat{\mathrm{VaR}}_\alpha$, $\widehat{\mathrm{ES}}_\alpha$) w.r.t. u.



Here are $\hat{\bar{F}}(x)$, $x \geq u$, $\widehat{\mathrm{VaR}}_{0.99}$, $\widehat{\mathrm{ES}}_{0.99}$ including confidence intervals.



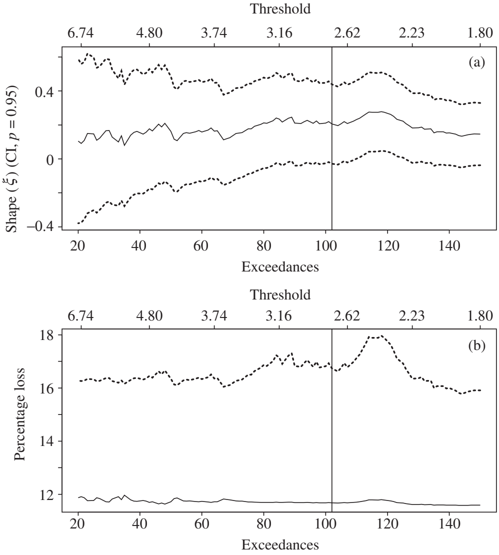
Example 5.25 (AT&T weekly loss data (continued))



- Fitted GPD model as in Example 5.23.
- Plot of $\hat{\bar{F}}(x)$.
- Vertical lines: $\widehat{\mathrm{VaR}}_{0.99}$, $\widehat{\mathrm{ES}}_{0.99}$
- log-log scale often good: $\bar{F}(x) = x^{-\alpha}L(x)$ and therefore

  $\log \bar{F}(x) = -\alpha \log(x) + \log L(x)$

  $\approx$ linear in $\log(x)$



- Sensitivity w.r.t. u
- **Top:** $\hat{\xi}$ for different u or N_u, including a 95% CI based on standard error
- **Bottom:** Corresponding $\widehat{\mathrm{VaR}}_{0.99}$ (solid line), $\widehat{\mathrm{ES}}_{0.99}$ (dotted line)

5.2.4 The Hill estimator

- Assume $F \in \mathrm{MDA}(H_\xi)$, $\xi > 0$, so that $\bar{F}(x) = x^{-\alpha}L(x)$, $\alpha > 0$.
- Let e^* be the mean excess function for $\log X$. Using partial integration ($\int H\,dG = [HG] - \int G\,dH$), we obtain

$$
\begin{aligned}
e^*(\log u) &= \mathbb{E}(\log X - \log u \,|\, \log X > \log u)\\
&= \frac{1}{\bar{F}(u)} \int_u^\infty (\log x - \log u)\, dF(x) = -\frac{1}{\bar{F}(u)} \int_u^\infty \log\Big(\frac{x}{u}\Big)\, d\bar{F}(x)\\
&= -\frac{1}{\bar{F}(u)} \Bigg(\underbrace{\Big[\log\Big(\frac{x}{u}\Big)\bar{F}(x)\Big]_u^\infty}_{=0} - \int_u^\infty \bar{F}(x)\frac{1}{x}\,dx \Bigg)\\
&= \frac{1}{\bar{F}(u)} \int_u^\infty \frac{\bar{F}(x)}{x}\,dx = \frac{1}{\bar{F}(u)} \int_u^\infty x^{-\alpha-1}L(x)\,dx.
\end{aligned}
$$

For u sufficiently large, $L(x) \approx L(u)$, $x \geq u$ (Karamata's Theorem), so

$$
e^*(\log u) \underset{u \text{ large}}{\approx} \frac{L(u)u^{-\alpha}/\alpha}{\bar{F}(u)} = \frac{1}{\alpha}.
$$

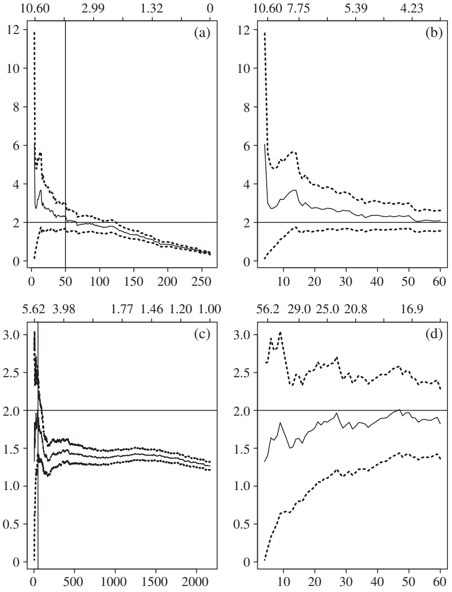
- For n large, k sufficiently small, use $u = X_{[k]}$, so

$$\frac{1}{\alpha} \approx e_n^*(\log X_{[k]}) = \frac{\sum_{i=1}^n (\log X_i - \log X_{[k]}) \mathbb{1}_{\{\log X_i > \log X_{[k]}\}}}{\sum_{i=1}^n \mathbb{1}_{\{\log X_i > \log X_{[k]}\}}}$$
$$= \frac{\sum_{i=1}^{k-1} (\log X_{[i]} - \log X_{[k]})}{k-1} = \frac{1}{k-1} \sum_{i=1}^{k-1} \log X_{[i]} - \log X_{[k]}$$

- The standard form of the *Hill estimator* of the tail index α is

$$\hat{\alpha}_{k,n}^{(\mathsf{H})} = \left(\frac{1}{k} \sum_{i=1}^{k} \log X_{[i]} - \log X_{[k]}\right)^{-1}, \quad 2 \le k \le n, \ k \text{ sufficiently small.}$$

- Choosing k: Find a small k where the *Hill plot* $\{(k, \hat{\alpha}_{k,n}^{(\mathsf{H})}) : 2 \le k \le n\}$ stabilizes (typically, $k = \lceil \beta n \rceil$, $\beta \in [0.01, 0.05]$).
- Interpreting Hill plots can be didfficult. If F does not have a regularly varying tail (or if it has serial dependence), Hill plots can be very misleading.



- Hill plots showing estimates of $\alpha = 1/\xi$ for (a), (b) the AT&T data and (c),(d) the Danish fire loss data (rhs = expanded version of the lhs).
- (a),(b): suggests estimates of $\alpha \in [1.5, 2]$ ($\xi \in [1/2, 2/3]$; close to the estimated $\hat{\xi} = 0.50$, see Example 5.22); (c),(d): suggests estimates of $\alpha \in [2, 4]$ ($\xi \in [1/4, 1/2]$; larger than the estimated $\hat{\xi} = 0.22$, see Example 5.23)

Hill-based tail and risk measure estimates

- Assume $\bar{F}(x) = cx^{-\alpha}$, $x \ge u > 0$ (replacing L by a constant). Estimate α by $\hat{\alpha}^{(\mathsf{H})}_{k,n}$ and u by $X_{[k]}$ (for k sufficiently small).
- Note that $c = u^{\alpha}\bar{F}(u)$ so $\hat{c} = X_{[k]}^{\hat{\alpha}^{(\mathsf{H})}_{k,n}} \hat{\bar{F}}_n(X_{[k]}) \approx X_{[k]}^{\hat{\alpha}^{(\mathsf{H})}_{k,n}} \frac{k}{n}$. We thus obtain the semi-parametric *Hill tail estimator*
$$\hat{\bar{F}}(x) = \frac{k}{n}\left(\frac{x}{X_{[k]}}\right)^{-\hat{\alpha}^{(\mathsf{H})}_{k,n}}, \quad x \ge X_{[k]}.$$
- From this result we obtain the semi-parametric *Hill VaR estimator*
$$\widehat{\mathrm{VaR}}_{\alpha}(X) = \left(\frac{n}{k}(1-\alpha)\right)^{-\frac{1}{\hat{\alpha}^{(\mathsf{H})}_{k,n}}} X_{[k]}, \quad \alpha \ge F(u) \approx 1 - \frac{k}{n},$$
and, for $\hat{\alpha}^{(\mathsf{H})}_{k,n} > 1$, $\alpha \ge F(u) \approx 1 - \frac{k}{n}$, the semi-param. *Hill ES estimator*
$$\widehat{\mathrm{ES}}_{\alpha}(X) = \frac{\left(\frac{n}{k}\right)^{-\frac{1}{\hat{\alpha}^{(\mathsf{H})}_{k,n}}} X_{[k]}}{1-\alpha} \int_{\alpha}^{1} (1-z)^{-\frac{1}{\hat{\alpha}^{(\mathsf{H})}_{k,n}}}\, dz = \frac{\hat{\alpha}^{(\mathsf{H})}_{k,n}}{\hat{\alpha}^{(\mathsf{H})}_{k,n} - 1} \widehat{\mathrm{VaR}}_{\alpha}(X).$$

Interlude: Scaling of the risk measures $\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha$

- Again assume $\bar{F}(x) = cx^{-\alpha}$, $x \ge u > 0$, and let $\hat{\alpha}$ denote a tail index estimator.
- As $\frac{\bar{F}(u)}{\bar{F}(x)} = (\frac{x}{u})^\alpha$, using $x := \mathrm{VaR}_\beta(X)$ and $u := \mathrm{VaR}_{\beta_u}(X)$ implies

$$\mathrm{VaR}_\beta(X) = \Big(\frac{1-\beta_u}{1-\beta}\Big)^{\frac{1}{\alpha}} \mathrm{VaR}_{\beta_u}(X). \tag{17}$$

  This allows one to estimate VaR_β at $\beta_u \le \beta$ (for $\beta_u \ge F(u)$):

$$\widehat{\mathrm{VaR}}_\beta(X) = \Big(\frac{1-\beta_u}{1-\beta}\Big)^{\frac{1}{\hat{\alpha}}} \widehat{\mathrm{VaR}}_{\beta_u}(X).$$

- For $\alpha > 1$, $\beta \ge \beta_u \ge F(u)$, a similar scaling for $\mathrm{ES}_\beta(X)$ is

$$\mathrm{ES}_\beta(X) \underset{(17)}{=} \frac{(1-\beta_u)^{\frac{1}{\alpha}} \mathrm{VaR}_{\beta_u}(X)}{1-\beta} \underbrace{\int_\beta^1 (1-\tilde{\beta})^{-\frac{1}{\alpha}}\, d\tilde{\beta}}_{=\frac{\alpha}{\alpha-1}(1-\beta)^{1-\frac{1}{\alpha}}} \underset{(17)}{=} \frac{\alpha}{\alpha-1} \mathrm{VaR}_\beta(X)$$

5.2.5 Simulation study of EVT quantile estimators

We compare estimators for ξ (Study 1) and $\mathrm{VaR}_{0.99}$ (Study 2) based on

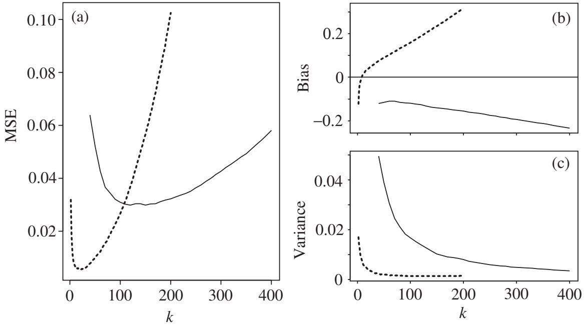
$$\begin{aligned}\mathrm{MSE}[\hat\theta] &= \mathbb{E}[(\hat\theta-\theta)^2] = \mathbb{E}[(\hat\theta-\mathbb{E}[\hat\theta]+\mathbb{E}[\hat\theta]-\theta)^2]\\ &= \mathbb{E}[(\hat\theta-\mathbb{E}[\hat\theta])^2] + \mathbb{E}[2(\hat\theta-\mathbb{E}[\hat\theta])(\mathbb{E}[\hat\theta]-\theta)] + \mathbb{E}[(\mathbb{E}[\hat\theta]-\theta)^2]\\ &= (\mathbb{E}[\hat\theta]-\theta)^2 + \mathrm{Var}[\hat\theta] = \mathrm{bias}[\hat\theta]^2 + \mathrm{Var}[\hat\theta]\end{aligned}$$

with a Monte Carlo study (Sample size $N = 1000$; drawn from a t_4 distribution with corresponding true $\xi = 1/4$); analytical evaluation of bias and variance is not possible.

Study 1: Estimating ξ

We estimate ξ with a fitted GPD (via MLE; $k \in \{30, 40, \dots, 400\}$) and with the Hill estimator ($\hat\xi = 1/\hat\alpha^{(\mathrm{H})}_{k,n}$; $k \in \{2, 3, \dots, 200\}$). Note that the t_4 distribution has a well-behaved regularly varying tail.

(a): $\widehat{\mathrm{MSE}}[\hat{\xi}]$; (b): $\widehat{\mathrm{bias}}[\hat{\xi}]$; (c): $\widehat{\mathrm{Var}}[\hat{\xi}]$ (solid: GPD; dotted: Hill)

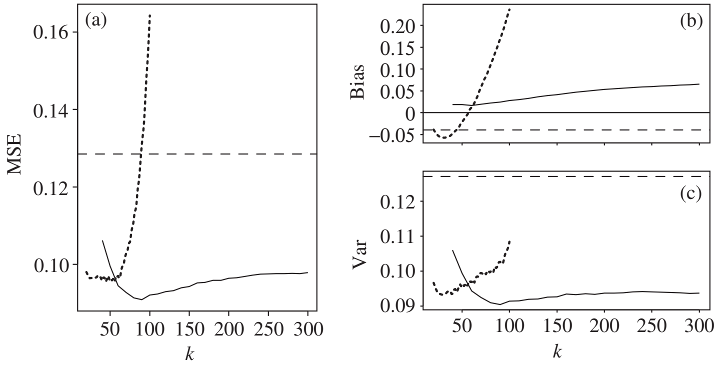


- The Hill estimator outperforms the GPD estimator (optimal k around 20–30) according to the variance for small k (number of order statistics)
- The biases are closer, with the Hill (GPD) estimator tending to overestimate (underestimate) ξ.
- For the GPD method, the optimal u is around 100–150 exceedances.

Study 2: Estimating $\mathrm{VaR}_{0.99}$

Estimate $\mathrm{VaR}_{0.99}$ based on a fitted GPD, with the Hill VaR estimator and with the empirical quantile estimator. Here the situation changes.

(a): $\widehat{\mathrm{MSE}}[\widehat{\mathrm{VaR}}_{0.99}]$; (b): $\widehat{\mathrm{bias}}[\widehat{\mathrm{VaR}}_{0.99}]$; (c): $\widehat{\mathrm{Var}}[\widehat{\mathrm{VaR}}_{0.99}]$ (solid: GPD; dotted: Hill; dashed: empirical quantile estimator)



- The empirical $\mathrm{VaR}_{0.99}$ estimator has a negative bias.
- The Hill $\mathrm{VaR}_{0.99}$ estimator has a negative bias for small k but a rapidly growing positive bias for larger k.
- The GPD $\mathrm{VaR}_{0.99}$ estimator has a positive bias which grows much more slowly.
- The GPD $\mathrm{VaR}_{0.99}$ estimator attains lowest MSE for a value of k around 100, but the MSE is very robust to the choice of k (because of the slow growth of the bias) $\Rightarrow$ Choice of u less critical!
- The Hill $\mathrm{VaR}_{0.99}$ estimator performs well for $20 \le k \le 75$ (we only use k values that lead to a quantile estimate beyond the effective threshold $X_{[k]}$) but then deteriorates rapidly.
- Both EVT methods outperform the empirical quantile estimator.

5.2.6 Conditional EVT for financial time series

- The GPD method is an unconditional approach for estimating $\bar{F}$ and associated risk measures. A conditional (time-dependent) risk-measurement approach may be more appropriate.
- We now consider a simple adaptation of the GPD method to obtain conditional risk-measure estimates in a GARCH context.
- Assume $X_{t-n+1}, \dots, X_t$ are negative log-returns generated by a strictly stationary time series process (X_t) of the form

$$X_t = \mu_t + \sigma_t Z_t,$$

  where μ_t and σ_t are $\mathcal{F}_{t-1}$-measurable and $Z_t \overset{\text{ind.}}{\sim} F_Z$; e.g., ARMA model with GARCH errors. Furthermore, let $Z \sim F_Z$.
- VaR_α^t and ES_α^t based on $F_{X_{t+1}|\mathcal{F}_t}$ are given by

$$\mathrm{VaR}_\alpha^t(X_{t+1}) = \mu_{t+1} + \sigma_{t+1}\,\mathrm{VaR}_\alpha(Z),$$
$$\mathrm{ES}_\alpha^t(X_{t+1}) = \mu_{t+1} + \sigma_{t+1}\,\mathrm{ES}_\alpha(Z).$$

- To obtain estimates $\widehat{\mathrm{VaR}}^t_\alpha(X_{t+1})$ and $\widehat{\mathrm{ES}}^t_\alpha(X_{t+1})$, proceed as follows:
  1) Fit an ARMA-GARCH model (via exponential smoothing or QMLE based on normal innovations (since we do not assume a particular innovation distribution)) $\Rightarrow$ Estimates of μ_{t+1} and σ_{t+1}.
  2) Fit a GPD to F_Z (treat the residuals from the GARCH fitting procedure as i.i.d. from F_Z) $\Rightarrow$ GPD-based estimates of $\mathrm{VaR}_\alpha(Z)$ (see (15)) and $\mathrm{ES}_\alpha(Z)$ (see (16)).

5.3 Point process models

So far: loss size distribution. Now: loss frequency distribution

5.3.1 Threshold exceedances for strict white noise

- Consider a strict white noise $(X_i)_{i\in\mathbb{N}}$ (i.i.d. from $F \in \mathrm{MDA}(H_\xi)$; can be extended to dependent processes with extremal index $\theta = 1$).
- Let $u_n(x) = c_n x + d_n$ (x fixed). We know $F^n(u_n(x)) \underset{n\uparrow\infty}{\to} H_\xi(x)$. Taking $-\log(\cdot)$ and using $-\log y \approx 1 - y$ for $y \to 1$, we obtain $n\bar{F}(u_n(x)) \approx -n\log F(u_n(x)) = -\log\big(F^n(u_n(x))\big) \underset{n\uparrow\infty}{\to} -\log H_\xi(x)$.
- $N_{u_n(x)}$ (exceedances among $X_1, \dots, X_n$) fulfills $N_{u_n(x)} \sim \mathrm{B}(n, \bar{F}(u_n(x)))$
- The Poisson Limit Theorem ($n \to \infty$, $p = \bar{F}(u_n(x)) \to 0$, $np = n\bar{F}(u_n(x)) \to \lambda = -\log H_\xi(x)$) implies $N_{u_n(x)} \underset{n\uparrow\infty}{\to} \mathrm{Poi}(-\log H_\xi(x))$.
- One can show: Not only is $N_{u_n(x)}$ asymptotically Poisson, but the exceedances occur according to a Poisson process.

On point processes

- Suppose $Y_1, \dots, Y_n$ take values in some *state space* $\mathcal{X}$ (e.g., $\mathbb{R}$, $\mathbb{R}^2$). Define for any $A \subseteq \mathcal{X}$, the counting rv

$$N(A) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_i \in A\}}.$$

  Under technical conditions, see Embrechts et al. (1997, pp. 220), $N(\cdot)$ defines a point process.

- $N(\cdot)$ is a *Poisson point process* on $\mathcal{X}$ with *intensity measure* Λ if:

  1) For $A \subseteq \mathcal{X}$ and $k \geq 0$,

$$\mathbb{P}(N(A) = k) = \begin{cases} e^{-\Lambda(A)} \frac{\Lambda(A)^k}{k!}, & \text{if } \Lambda(A) < \infty, \\ 0, & \text{if } \Lambda(A) = \infty. \end{cases}$$

  2) $N(A_1), \dots, N(A_m)$ are independent for any mutually disjoint subsets $A_1, \dots, A_m$ of $\mathcal{X}$.

- Note that $\mathbb{E}N(A) = \Lambda(A)$. Also, the *intensity (function)* is the function $\lambda(\boldsymbol{x})$ which satisfies $\Lambda(A) = \int_A \lambda(\boldsymbol{x})\, d\boldsymbol{x}$.

Asymptotic behavior of the point process of exceedances

- For $n \in \mathbb{N}$ and $i \in \{1, \dots, n\}$ let $Y_{i,n} = \frac{i}{n} \mathbb{1}_{\{X_i > u_n(x)\}}$. The point process of exceedances over u_n is the process $N_n(\cdot)$ with state space $\mathcal{X} = (0, 1]$ given by

$$N_n(A) = \sum_{i=1}^{n} \mathbb{1}_{\{Y_{i,n} \in A\}}, \quad A \subseteq \mathcal{X}.$$

- N_n is an element of the sequence of point processes (N_n). N_n counts the exceedances with time of occurrence in A and we are interested in the behaviour of N_n as $n \to \infty$.
- Embrechts et al. (1997, Theorem 5.3.2) show that $N_n(\cdot)$ converges in distribution on $\mathcal{X}$ to a Poisson process $N(\cdot)$ with intensity $\Lambda(\cdot)$ satisfying $\Lambda(A) = (t_2 - t_1)\lambda(x)$ for $A = (t_1, t_2) \subseteq \mathcal{X}$, $\lambda(x) = -\log H_\xi(x)$.

- In particular, $\mathbb{E}N_n(A) \underset{n\uparrow\infty}{\to} \mathbb{E}N(A) = \Lambda(A) = (t_2 - t_1)\lambda(x)$. λ does not depend on time and takes the constant value $\lambda = \lambda(x)$.
- We refer to the limiting process as a *homogeneous Poisson process* with *intensity* (or rate) λ.

Application of the result in practice

- Fix a large n and $u = c_n x + d_n$ for some x.
- Approximate N_u by a Poisson rv and the point process of exceedances of u by a homogeneous Poisson process with rate $\lambda = -\log H_\xi(x) = -\log H_\xi((u - d_n)/c_n) = -\log H_{\xi,\mu=d_n,\sigma=c_n}(u)$.

  $\Rightarrow$ Relationship between the GEV model and a Poisson model for the occurrence in time of exceedances of u.
- We see that exceedances of i.i.d. data over u are separated by i.i.d. exponential waiting times.

5.3.2 The POT model

- Putting the pieces together, we obtain an asymptotic model for threshold exceedances in regularly spaced i.i.d. data (or data with $\theta = 1$).
- This so-called *peaks-over-threshold (POT) model* makes the following assumptions:
  1) Exceedances times occur according to a homogeneous Poisson process.
  2) Excesses above u are i.i.d. and independent of exceedance times.
  3) The excess distribution is generalized Pareto.
- This model can also be viewed as a *marked Poisson point process* (exceedance times = points; GPD-distributed excesses = marks) or a (non-homogeneous) *two-dimensional Poisson* point process (point (t, x) = (time, magnitude of exceedance))

Two-dimensional Poisson formulation of POT model

- Assume that, on the state space $\mathcal{X} = (0,1] \times (u,\infty)$, the point process defined by $N(A) = \sum_{i=1}^n \mathbb{1}_{\{(i/n, X_i) \in A\}}$ is a Poisson process with intensity at (t,x) given by

$$\lambda(x) = \lambda(t,x) = \begin{cases} \frac{1}{\sigma}\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi-1}, & \text{if } (1+\xi(x-\mu)/\sigma) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

- For $A = (t_1,t_2) \times (x,\infty) \subseteq \mathcal{X}$, the intensity measure is

$$\Lambda(A) = \int_{t_1}^{t_2} \int_x^\infty \lambda(y)\,dy\,dt = -(t_2 - t_1)\log H_{\xi,\mu,\sigma}(x)$$

  Thus, for any $x \geq u$, the one-dimensional process of exceedances of x is a homogeneous Poisson process with intensity $\tau(x) = -\log H_{\xi,\mu,\sigma}(x)$.

- $\bar{F}_u(x)$ can be calculated as the ratio of the rates of exceeding $u+x$ and u via

$$\bar{F}_u(x) = \frac{\tau(u+x)}{\tau(u)} = \left(1 + \frac{\xi x}{\sigma + \xi(u-\mu)}\right)^{-1/\xi} = \bar{G}_{\xi,\sigma+\xi(u-\mu)}(x)$$

This is precisely the POT model.

- The model also implies the GEV model. Consider $\{M_n \le x\}$ for some $x \ge u$, i.e., the event that there are no points in $A = (0,1] \times (x,\infty)$. Thus, $\mathbb{P}(M_n \le x) = \mathbb{P}(N(A) = 0) = \exp(-\Lambda(A)) = H_{\xi,\mu,\sigma}(x)$, $x \ge u$, which is precisely the GEV model.

Statistical estimation of the POT model

- Given the exceedances $\tilde{X}_1 < \cdots < \tilde{X}_{N_u}$, $A = (0,1] \times (u,\infty)$ and $\Lambda(A) = \tau(u) =: \tau_u$, the likelihood $L(\xi,\sigma,\mu;\tilde{X}_1,\dots,\tilde{X}_{N_u})$ is

$$\underbrace{N_u!}_{\substack{\text{ordered}\\ \text{sample}}}\ \underbrace{e^{-\Lambda(A)}\frac{\Lambda(A)^{N_u}}{N_u!}}_{\text{prob. of } N_u \text{ samples}} \prod_{i=1}^{N_u} \underbrace{\frac{\lambda(\tilde{X}_i)}{\Lambda(A)}}_{\text{density of } \tilde{X}_i} = e^{-\Lambda(A)} \prod_{i=1}^{N_u} \lambda(\tilde{X}_i) = e^{-\tau_u} \prod_{i=1}^{N_u} \lambda(\tilde{X}_i).$$

- Reparametrizing λ by $\tau_u = -\log H_{\xi,\mu,\sigma}(u) = (1 + \xi\frac{u-\mu}{\sigma})^{-1/\xi}$ and

$\beta = \sigma + \xi(u-\mu)$, we obtain

$$\lambda(x) = \frac{1}{\sigma}\Big(1+\xi\frac{x-\mu}{\sigma}\Big)^{-\frac{1}{\xi}-1} = \frac{1}{\sigma}\bigg(\Big(1+\xi\frac{u-\mu}{\sigma}\Big)\Big(1+\frac{\xi\frac{x-u}{\sigma}}{1+\xi\frac{u-\mu}{\sigma}}\Big)\bigg)^{-\frac{1}{\xi}-1}$$

$$= \frac{\tau_u}{\sigma(1+\xi\frac{u-\mu}{\sigma})}\Big(1+\frac{\xi\frac{x-u}{\sigma}}{1+\xi\frac{u-\mu}{\sigma}}\Big)^{-\frac{1}{\xi}-1} = \frac{\tau_u}{\beta}\Big(1+\frac{\xi(x-u)}{\sigma+\xi(u-\mu)}\Big)^{-\frac{1}{\xi}-1}$$

$$= \frac{\tau_u}{\beta}\Big(1+\frac{\xi(x-u)}{\beta}\Big)^{-\frac{1}{\xi}-1} = \tau_u g_{\xi,\beta}(x-u),$$

where $\xi \in \mathbb{R}$ and $\tau_u, \beta > 0$. Therefore, $\ell(\xi,\sigma,\mu;\tilde{X}_1,\dots,\tilde{X}_{N_u})$ equals

$$= -\tau_u + \sum_{i=1}^{N_u}\log\lambda(\tilde{X}_i) = -\tau_u + N_u\log\tau_u + \sum_{i=1}^{N_u}\overbrace{(\log\lambda(\tilde{X}_i) - \log\tau_u)}^{=\log g_{\xi,\beta}(\tilde{X}_i-u)}$$

$$= \ell_{\mathrm{Poi}}(\tau_u; N_u) - N_u\log(T) + \ell_{\mathrm{GPD}}(\xi,\beta;\tilde{X}_1-u,\dots,\tilde{X}_{N_u}-u), \tag{18}$$

where ℓ_{Poi} is the log-likelihood for a one-dimensional homogeneous Poisson process with rate τ_u and ℓ_{GPD} is the log-likelihood for fitting a GPD to the excesses $\tilde{X}_i - u$, $i \in \{1,\dots,N_u\}$.

- We can thus separate inferences about (ξ, β) and τ_u. Estimate (ξ, β) in a GPD analysis and then τ_u by its MLE N_u. Use these estimates to infer estimates of $\mu = u - \beta(1 - \tau_u^{\xi})/\xi$ and $\sigma = \tau_u^{\xi}\beta$.

Advantages of the POT model formulation

- One advantage of the two-dimensional Poisson point process model is that ξ, μ and σ do not depend on u (unlike β in the GPD model).

  $\Rightarrow$ In practice, we would expect the estimated parameters of the Poisson model to be roughly stable over a range of high thresholds.

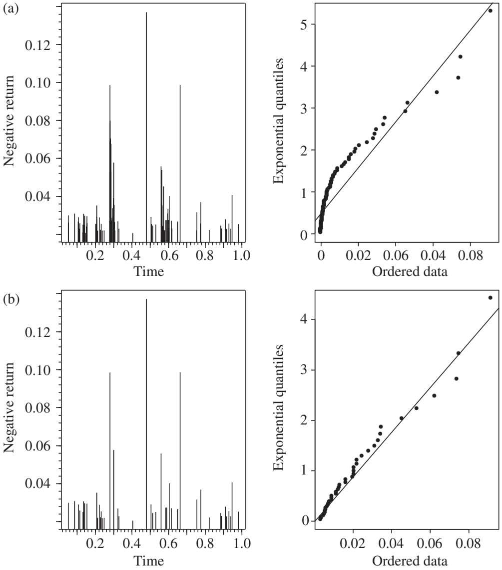
- The intensity λ is thus often used to introduce covariates to obtain Poisson processes which are non-homogeneous in time, e.g., by replacing μ and σ by parameters that vary over time as functions of covariates; see, e.g., Chavez-Demoulin et al. (2013).

Applicability of the POT model to return series data

- Returns do not really form genuine point events in time (in contrast to, e.g., water levels). They are discrete-time measurements that describe short-term changes (a day or a week). Nonetheless, assume that under a longer-term perspective, such data can be approximated by point events in time.
- Exceedances of u for daily financial return series do not necessarily occur according to a homogeneous Poisson process. They tend to cluster. Thus the standard POT model is not directly applicable.
- For stochastic processes with extremal index $\theta < 1$, e.g., GARCH processes, the extremal clusters themselves should occur according to a homogeneous Poisson process in time $\Rightarrow$ Individual exceedances occur according to a *Poisson cluster process*; see Leadbetter (1991). Thus a suitable model for the occurrence and magnitude of exceedances in a

financial return series might be some form of marked Poisson cluster process.

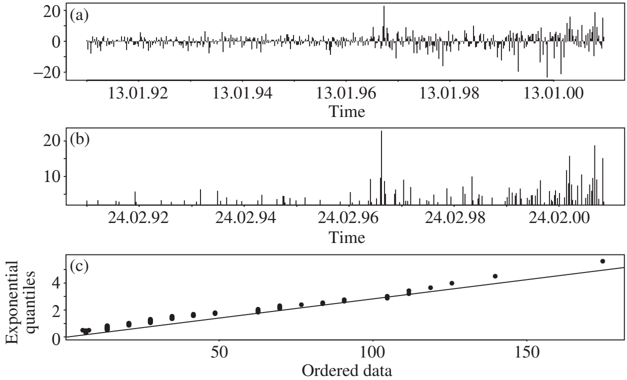
- *Declustering* may circumvent the problem. One identifies clusters (ad hoc; not easy) of exceedances and then applis the POT model to cluster maxima only.
- A possible declustering algorithm is the *runs method*. A run size r is fixed and two successive exceedances are said to belong to two different clusters if they are separated by a run of at least r values below u; see Embrechts et al. (1997, pp. 422).
- In the following figure the DAX daily negative returns have been declustered with $r = 10$ trading days; this reduces the 100 exceedances to 42 cluster maxima.



- (a): DAX daily negative returns and a Q-Q plot of their spacings
- (b): Declustered data (runs method with $r = 10$ trading days $\Rightarrow$ spacings are more consistent with a Poisson model)
- However, by neglecting the modeling of cluster formation, we cannot make more dynamic statements about the intensity of occurrence of exceedances.

Example 5.26 (POT analysis of AT&T weekly losses (continued))

Consider the 102 weekly percentage losses exceeding $u = 2.75\%$:



- Inter-exceedance times seem to follow an exponential distribution.
- But exceedances become more frequent over time ($\lightning$ to homogeneous Poisson process $\Rightarrow$ Possibly consider an inhomogeneous Poisson process).

- Using the log-likelihood (18), we fit a two-dimensional Poisson model to the 102 exceedances of $u = 2.75\%$. The parameter estimates are $\hat{\xi} = 0.22$, $\hat{\mu} = 19.9$ and $\hat{\sigma} = 5.95$.
- The implied GPD scale parameter is $\hat{\beta} = \hat{\sigma} + \hat{\xi}(u - \hat{\mu}) = 2.1 \Rightarrow$ The same $\hat{\xi}$ and $\hat{\beta}$ as in Example 5.23.
- The estimated exceedance rate over $u = 2.75$ is $\hat{\tau}(u) = -\log H_{\hat{\xi},\hat{\mu},\hat{\sigma}}(u) = 102$ ($=$ number of exceedances; as theory suggests).
- Higher thresholds, e.g., 15%: Since $\hat{\tau}(15) = 2.50$, losses exceeding 15% occur as a Poisson process with rate 2.5 losses per 10-year period ($\approx$ a four-year event). $\Rightarrow$ The Poisson model provides an alternative method of defining the return period of an event.
- Similarly, estimate return levels: If the 10-year return level is the level which is exceeded according to a Poisson process with rate one loss per 10 years, estimate the level by solving $\hat{\tau}(u) = 1$ w.r.t. u, so

$u = H^{-1}_{\hat{\xi},\hat{\mu},\hat{\sigma}}(\exp(-1)) = 19.9$ so the 10-year event is a weekly loss of roughly 20%.

- Confidence intervals for such quantities can be constructed via profile likelihoods.