

# 5 Extreme value theory

## 5.1 Maxima

## 5.2 Threshold exceedances

## 5.1 Maxima

Consider a series of financial losses  $(X_k)_{k \in \mathbb{N}}$ .

### 5.1.1 Generalized extreme value distribution

#### Convergence of sums

Let  $(X_k)_{k \in \mathbb{N}}$  be iid with  $\mathbb{E}(X_1^2) < \infty$  (mean  $\mu$ , variance  $\sigma^2$ ) and  $S_n = \sum_{k=1}^n X_k$ . As  $n \rightarrow \infty$ ,  $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$  by the Strong Law of Large Numbers (SLLN), so  $(\bar{X}_n - \mu)/\sigma \xrightarrow{\text{a.s.}} 0$ . By the CLT,

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n \uparrow \infty]{d} N(0, 1) \text{ or } \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{S_n - d_n}{c_n} \leq x\right) = \Phi(x),$$

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

## Convergence of maxima

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

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

One can show that, for  $n \rightarrow \infty$ ,  $M_n \xrightarrow{\text{a.s.}} x_F$  (similar as in the SLLN) where  $x_F := \sup\{x \in \mathbb{R} : F(x) < 1\} = F^{\leftarrow}(1) \leq \infty$  denotes the *right endpoint of  $F$*  (similar to the SLLN).

**Question:** 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.

$$\mathbb{P}((M_n - d_n)/c_n \leq x) = \mathbb{P}(M_n \leq c_n x + d_n) = F^n(c_n x + d_n) \xrightarrow[n \uparrow \infty]{} H(x),$$

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

One can show that  $H$  is determined up to location/scale, i.e.  $H$  specifies a **unique type** of distribution. This is guaranteed by the **convergence to types theorem**; see the appendix.

**Question:** What does  $H$  look like?

## Definition 5.2 (Generalized extreme value (GEV) distribution)

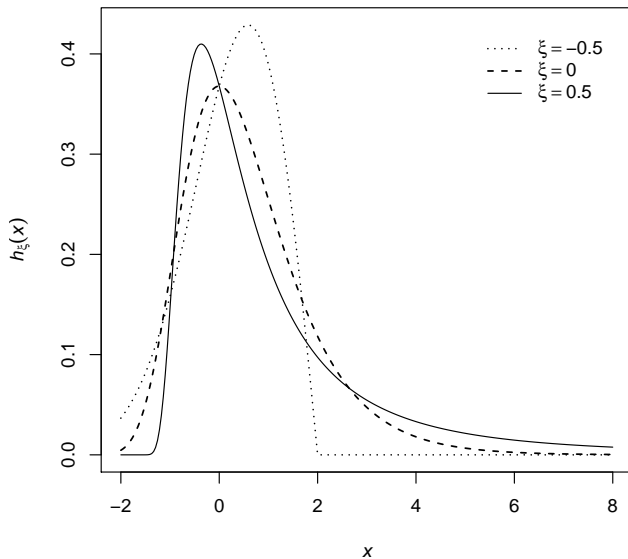
The (standard) *generalized extreme value (GEV) distribution* is given by

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

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  $\xi$  (simplifies statistical modelling).
- The larger  $\xi$ , the heavier tailed  $H_{\xi}$  (if  $\xi > 0$ ,  $\mathbb{E}(X^k) = \infty$  iff  $k \geq \frac{1}{\xi}$ ).
- $\xi$  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

Density  $h_\xi$  for  $\xi \in \{-0.5, 0, 0.5\}$  (dotted, dashed, solid)



### Theorem 5.3 (Fisher–Tippett–Gnedenko)

If  $F \in \text{MDA}(H)$  for some non-degenerate  $H$ , then  $H$  must be of GEV type, i.e.  $H = H_\xi$  for some  $\xi \in \mathbb{R}$ .

*Proof.* Non-trivial. For a sketch, see Embrechts et al. (1997, p. 122).  $\square$

- **Interpretation:** If location-scale transformed maxima converge in distribution to a non-degenerate limit, the limiting distribution must be a GEV distribution.
- We can always choose normalizing sequences  $(c_n) > 0$ ,  $(d_n)$  such that  $H_\xi$  appears in standard form.
- All commonly encountered continuous distributions are in the MDA of a GEV distribution.

### Example 5.4 (Exponential distribution)

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

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

### Example 5.5 (Pareto distribution)

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

$$\begin{aligned} &\left(1 - \left(\frac{\kappa}{\kappa + x\kappa n^{1/\theta}/\theta + \kappa(n^{1/\theta} - 1)}\right)^\theta\right)^n \\ &= \left(1 - \left(\frac{1}{1 + xn^{1/\theta}/\theta + n^{1/\theta} - 1}\right)^\theta\right)^n = \left(1 - \left(\frac{1}{n^{1/\theta}(1 + x/\theta)}\right)^\theta\right)^n \\ &= \left(1 - \frac{(1 + x/\theta)^{-\theta}}{n}\right)^n \xrightarrow{n \uparrow \infty} \exp(-(1 + x/\theta)^{-\theta}) = H_{1/\theta}(x) \text{ (Fréchet)} \end{aligned}$$

Therefore,  $F \in \text{MDA}(H_{1/\theta})$ .



## 5.1.2 Maximum domains of attraction

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

### Definition 5.6 (Slowly/regularly varying functions)

- 1) A positive, Lebesgue-measurable function  $L$  on  $(0, \infty)$  is *slowly varying at  $\infty$*  if  $\lim_{x \rightarrow \infty} \frac{L(tx)}{L(x)} = 1$ ,  $t > 0$ . The class of all such functions is denoted by  $\mathcal{R}_0$ ; e.g.  $c, \log \in \mathcal{R}_0$ .
- 2) A positive, Lebesgue-measurable function  $h$  on  $(0, \infty)$  is *regularly varying at  $\infty$  with index  $\alpha \in \mathbb{R}$*  if  $\lim_{x \rightarrow \infty} \frac{h(tx)}{h(x)} = t^\alpha$ ,  $t > 0$ . The class of all such functions is denoted by  $\mathcal{R}_\alpha$ ; e.g.  $x^\alpha L(x) \in \mathcal{R}_\alpha$ .

If  $\bar{F} \in \mathcal{R}_{-\alpha}$ ,  $\alpha > 0$ , the tail of  $F$  decays like a power function (Pareto like).

## The Fréchet case

### Theorem 5.7 (Fréchet MDA, Gnedenko (1943))

For  $\xi > 0$ ,  $F \in \text{MDA}(H_\xi)$  if and only if  $\bar{F}(x) = x^{-1/\xi}L(x)$  for some  $L \in \mathcal{R}_0$ . If  $F \in \text{MDA}(H_\xi)$ ,  $\xi > 0$ , the normalizing sequences can be chosen as  $c_n = F^{\leftarrow}(1 - 1/n)$  and  $d_n = 0$ ,  $n \in \mathbb{N}$ .

*Proof.* Non-trivial. For a sketch, see Embrechts et al. (1997, p. 131).  $\square$

- **Interpretation:** Distributions in  $\text{MDA}(H_\xi)$ ,  $\xi > 0$ , are those whose tails decay like power functions;  $\alpha = 1/\xi$  is known as *tail index*.
- If  $X \sim F \in \text{MDA}(H_\xi)$ ,  $\xi > 0$ ,  $X \geq 0$ , 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).
- **Examples in  $\text{MDA}(H_\xi)$ ,  $\xi > 0$ :** Inverse gamma, Student  $t$ , log-gamma,  $F$ , Cauchy,  $\alpha$ -stable with  $0 < \alpha < 2$ , Burr and Pareto

### Example 5.8 (Pareto distribution)

For  $F = \text{Par}(\theta, \kappa)$ ,  $\bar{F}(x) = (\kappa/(\kappa + x))^\theta = (1 + x/\kappa)^{-\theta} = x^{-\theta}L(x)$ ,  $x \geq 0$ ,  $\theta, \kappa > 0$ , where  $L(x) = (\kappa^{-1} + x^{-1})^{-\theta} \in \mathcal{R}_0$ . We (again) see that  $F \in \text{MDA}(H_\xi)$ ,  $\xi > 0$ .

### The Gumbel case

- The **characterization** of this class is **more complicated**; see the appendix and Embrechts et al. (1997, p. 142).
- Essentially  $\text{MDA}(H_0)$  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  $\text{MDA}(H_0)$ :** Normal, log-normal, exponential, gamma (exponential, Erlang,  $\chi^2$ ), standard Weibull, Benktander type I and II, generalized hyperbolic (except Student  $t$ ).

## The Weibull case

### Theorem 5.9 (Weibull MDA)

For  $\xi < 0$ ,  $F \in \text{MDA}(H_\xi)$  if and only if  $x_F < \infty$  and  $\bar{F}(x_F - 1/x) = x^{1/\xi} L(x)$  for some  $L \in \mathcal{R}_0$ ; the normalizing sequences can be chosen as  $c_n = x_F - F^{\leftarrow}(1 - 1/n)$  and  $d_n = x_F$ ,  $n \in \mathbb{N}$ .

*Proof.* **Non-trivial.** For a sketch, see Embrechts et al. (1997, p. 135).  $\square$

**Examples in  $\text{MDA}(H_\xi)$ ,  $\xi < 0$ :** **beta** (uniform). All  $F \in \text{MDA}(H_\xi)$ ,  $\xi < 0$ , 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  $X_k \sim F$ ,  $k \in \mathbb{Z}$ .

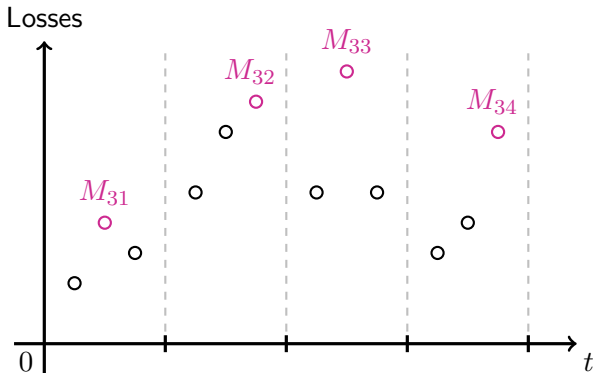
- Let  $\tilde{X}_k \stackrel{\text{ind.}}{\sim} F$ ,  $k \in \mathbb{Z}$ , and  $\tilde{M}_n = \max\{\tilde{X}_1, \dots, \tilde{X}_n\}$ . For many processes one can show that there exists a real number  $\theta \in (0, 1]$  such that  $\lim_{n \uparrow \infty} \mathbb{P}((M_n - d_n)/c_n \leq x) = H^\theta(x)$  if and only if  $\lim_{n \uparrow \infty} \mathbb{P}((\tilde{M}_n - d_n)/c_n \leq x) = H(x)$  (non-degenerate);  $\theta$  is known as the *extremal index*.
- If  $F \in \text{MDA}(H_\xi)$  for some  $\xi \Rightarrow M_n$  converges in distribution to  $H_\xi^\theta$ . Since  $H_\xi^\theta$  is of the same type as  $H_\xi$ , the limiting distribution of the block maxima of the dependent series is the same as in the iid 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  $\theta$  can be approximated by the distribution  $\tilde{M}_{n\theta}$  of the maximum of  $n\theta < n$  observations from the associated iid series.  $\Rightarrow n\theta$  counts the number of roughly independent clusters in  $n$  observations ( $\theta$  is often interpreted as “1/mean cluster size”).
- If  $\theta = 1$ , large sample maxima behave as in the iid 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  $(\varepsilon_t)$  strict white noise:  $\theta = 1$  (Gaussian);  $\theta \in (0, 1)$  (if df of  $\varepsilon_t$  is in  $\text{MDA}(H_\xi)$ ,  $\xi > 0$ );
  - ▶ GARCH 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  $(X_i)_{i \in \mathbb{N}} \stackrel{\text{ind.}}{\sim} F \in \text{MDA}(H_\xi)$ ,  $\xi \in \mathbb{R}$ , or of a process with an extremal index such as GARCH. The **Fisher–Tippett–Gnedenko Theorem** implies that

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

- For fitting  $\theta = (\xi, \mu, \sigma)$ , divide the realizations into  $m$  blocks of size  $n$  denoted by  $M_{n1}, \dots, M_{nm}$  (e.g. daily log-returns  $\Rightarrow$  monthly 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 density  $h_\xi$  of  $H_\xi$  is

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



The **log-likelihood** is thus

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

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

### Remark 5.10

- 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) **MLE regularity conditions** for consistency and asymptotic efficiency were shown by Smith (1985) for  $\xi > -1/2$  (fine for practice).

## Return levels and stress losses (exceedances)

Let  $M_n \sim H$  (exact or estimated).  $H$  can be used to estimate the...

- 1) ... size of an event with prescribed frequency (*return-level problem*)
  - The level  $r_{n,k}$  which is expected to be exceeded in one out of every  $k$  blocks of size  $n$  satisfies  $\mathbb{P}(M_n > r_{n,k}) = 1/k$  (e.g. 10-year return level  $r_{260,10}$  = level exceeded in one out of every 10y; 260d  $\approx$  1y).
  - $r_{n,k} = H^{\leftarrow}(1 - 1/k)$  is known as  *$k$   $n$ -block return level* with parametric estimator  $\hat{r}_{n,k} = H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}^{\leftarrow}(1 - 1/k) = \hat{\mu} + \frac{\hat{\sigma}}{\hat{\xi}}((-\log(1 - 1/k))^{-\hat{\xi}} - 1)$ .
- 2) ... frequency of an event with prescribed size (*return-period problem*)
  - The number  $k_{n,u}$  of  $n$ -blocks for which we expect to see a single  $n$ -block exceeding  $u$  satisfies  $r_{n,k_{n,u}} = u$ .
  - $k_{n,u} = 1/\bar{H}(u)$  is known as *return period* of the event  $\{M_n > u\}$  with parametric estimator  $\hat{k}_{n,u} = 1/\bar{H}_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}(u)$ .

### Example 5.11 (Block maxima analysis of S&P500)

Suppose it is Friday 1987-10-16; the Friday before Black Monday (1987-10-19). The S&P 500 index fell by 9.12% this week. On that Friday alone the index is down 5.16%. We fit a GEV distribution to (bi)annual maxima of daily negative log-returns  $X_t = \log(S_t/S_{t-1})$  since 1960-01-01.

**Analysis 1:** Annual maxima ( $m = 28$ ; including the latest from the incomplete year 1987):  $\hat{\theta} = (0.30, 0.02, 0.007) \Rightarrow$  heavy-tailed Fréchet distribution (infinite fourth moment). The corresponding standard errors are  $(0.21, 0.002, 0.001) \Rightarrow$  High uncertainty ( $m$  small) for estimating  $\xi$ .

**Analysis 2:** Biannual maxima ( $m = 56$ ):  $\hat{\theta} = (0.34, 0.02, 0.006)$  with standard errors  $(0.14, 0.0009, 0.0008) \Rightarrow$  Even heavier tails. In what follows we work with the annual maxima.

- What is the probability that next year's maximal risk-factor change exceeds all previous ones?  $1 - H_{\hat{\xi}, \hat{\mu}, \hat{\sigma}}(\text{"all previous maxima"})$

- Was a risk-factor change of the size as of Black Monday foreseeable?
    - ▶ Based on data up to and including Friday 1987-10-16, the 10-year return level  $r_{260,10}$  is estimated as  $\hat{r}_{260,10} = 4.42\%$ .
    - ▶ Index drop Black Monday:  $20.47\% \Rightarrow X_{t+1} = 22.9\% \gg \hat{r}_{260,10}$ .
    - ▶ One can show that  $22.9\%$  is in the 95% confidence interval of  $r_{260,50}$  (estimated as  $\hat{r}_{260,50} = 7.49\%$ ), but the 28 maxima are too few to get a reliable estimate of a once-in-50-years event.
  - Based on the available data, what is the (estimated) return period of a risk-factor change at least as large as on Black Monday?
    - ▶ The estimated return period  $k_{260,0.229}$  is  $\hat{k}_{260,0.229} = 1877$  years.
    - ▶ One can show that the 95% confidence interval encompasses everything from 45y to essentially never!  $\Rightarrow$  Very high uncertainty!
- $\Rightarrow$  On 1987-10-16 we did not have enough data to say anything meaningful about such an event. Quantifying such events is difficult.

## 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 5.12 (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  $\xi < 0$ .

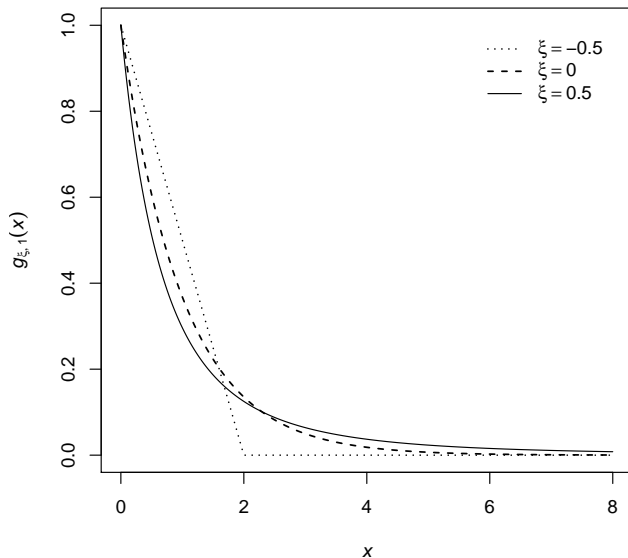
- The parameterization is continuous in  $\xi$ .
- The larger  $\xi$ , the heavier tailed  $G_{\xi,\beta}$  (if  $\xi > 0$ ,  $\mathbb{E}(X^k) = \infty$  iff  $k \geq \frac{1}{\xi}$ ; if  $\xi < 1$ , then  $\mathbb{E}X = \beta/(1 - \xi)$ ).
- $\xi$  is known as *shape*;  $\beta$  as *scale*. Special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

- $G_{\xi,\beta} \in \text{MDA}(H_\xi)$ ,  $\xi \in \mathbb{R}$ .

Density  $g_{\xi,1}$  for  $\xi \in \{-0.5, 0, 0.5\}$  (dotted, dashed, solid)



### Definition 5.13 (Excess distribution over $u$ , mean excess function)

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

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

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

$$e(u) = \mathbb{E}(X - u \mid X > u) \quad (\text{i.e. the mean w.r.t. } F_u)$$

### Interpretation

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

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

$$\text{ES}_\alpha(X) = e(\text{VaR}_\alpha(X)) + \text{VaR}_\alpha(X), \quad \alpha \in (0, 1) \quad (10)$$



**Example 5.14** ( $F_u$ ,  $e(u)$  for  $\text{Exp}(\lambda)$ ,  $G_{\xi,\beta}$ )

- 1) If  $F$  is  $\text{Exp}(\lambda)$ , then  $F_u(x) = 1 - e^{-\lambda x}$ ,  $x \geq 0$  (so again  $\text{Exp}(\lambda)$ ; lack-of-memory property). The mean excess function is  $e(u) = 1/\lambda = \mathbb{E}X$ .
- 2) If  $F$  is  $G_{\xi,\beta}$ , then  $F_u(x) = G_{\xi,\beta+\xi u}(x)$ ,  $x \geq 0$  (so again GPD, with the same shape, only the scale grows linearly in  $u$ ). 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  $\xi$  determines the slope of  $e(u)$ .

### Theorem 5.15 (Pickands–Balkema–de Haan (1974/75))

There exists a positive, measurable function  $\beta(u)$ , such that

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

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

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

### Interpretation

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

## 5.2.2 Modelling 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  $X_1, \dots, X_n \sim F \in \text{MDA}(H_\xi)$ ,  $\xi \in \mathbb{R}$ , 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 *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 iid and (roughly) distributed as  $G_{\xi, \beta}$ , the *log-likelihood* is given by

$$\begin{aligned}\ell(\xi, \beta; Y_1, \dots, Y_{N_u}) &= \sum_{k=1}^{N_u} \log g_{\xi, \beta}(Y_k) \\ &= -N_u \log(\beta) - (1 + 1/\xi) \sum_{k=1}^{N_u} \log(1 + \xi Y_k / \beta)\end{aligned}$$

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

## Excesses over higher thresholds

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

### Lemma 5.16

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

$$\begin{aligned}\bar{F}_v(x) &= \frac{\bar{F}(v+x)}{\bar{F}(v)} = \frac{\bar{F}(u+(v+x-u))}{\bar{F}(u)} \frac{\bar{F}(u)}{\bar{F}(u+(v-u))} \\ &= \frac{\bar{F}_u(v+x-u)}{\bar{F}_u(v-u)} = \frac{\bar{G}_{\xi,\beta}(x+v-u)}{\bar{G}_{\xi,\beta}(v-u)} \stackrel{\text{check}}{=} \bar{G}_{\xi,\beta+\xi(v-u)}(x) \quad \square\end{aligned}$$

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

If  $\xi < 1$  (so if it exists), the mean excess function is given by

$$e(v) = \frac{\xi}{1-\xi}v + \frac{\beta - \xi u}{1-\xi}, \quad v \in [u, \infty) \text{ if } \xi \in [0, 1), \quad (11)$$

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

## Sample mean excess plot and choice of the threshold

### Definition 5.17 (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

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

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

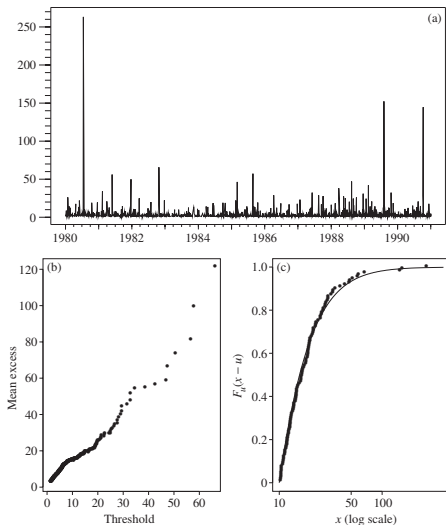
- If the data supports the GPD model over  $u$ ,  $e_n(v)$  should become increasingly “linear” for higher values of  $v \geq u$ . An upward/zero/downward trend indicates whether  $\xi > 0/\xi = 0/\xi < 0$ .
- Select  $u$  as the smallest point where  $e_n(v)$ ,  $v \geq u$ , becomes linear. Rule-of-thumb: One needs a couple of thousand data points and can often take  $u$  around the 0.9-quantile.
- 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  $L \in \mathcal{R}_0$ ).
- One should always analyze the data for several  $u$  and check the sensitivity of the choice of  $u$ .

### Example 5.18 (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 (11) 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”.

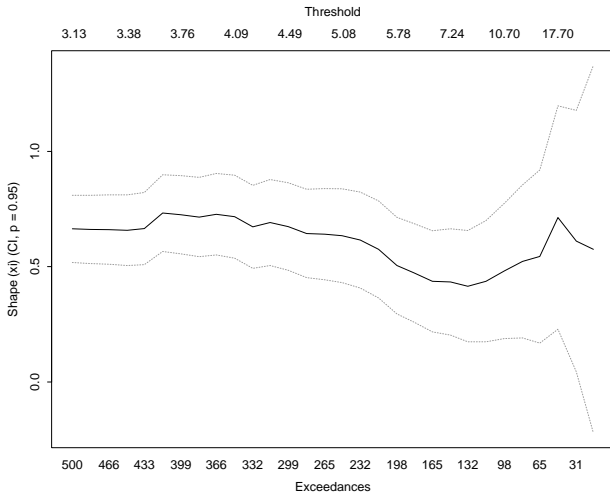


(a): Losses ( $> 1M$ ; in M); (b):  $e_n(u)$  ( $\uparrow$ ); (c) empirical  $F_u(x - u)$ ,  $G_{\hat{\xi}, \hat{\beta}}$



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

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



$\Rightarrow$  The higher  $u$ , the wider the confidence intervals (also support  $u = 10$ ).

### Example 5.19 (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 data over 1993–2000 is not consistent with the iid 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 Modelling 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  $0 \leq x < x_F - u$ ,  $\xi \neq 0$  and some  $u$ .
- We obtain the following GPD-based formula for tail probabilities:

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

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

$$\text{VaR}_\alpha = F^{\leftarrow}(\alpha) = u + \frac{\beta}{\xi} \left( \left( \frac{1 - \alpha}{\bar{F}(u)} \right)^{-\xi} - 1 \right), \quad (12)$$

$$\text{ES}_\alpha = \frac{\text{VaR}_\alpha}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi}, \quad \xi < 1. \quad (13)$$

The formula for  $\text{ES}_\alpha$  can also be obtained from  $e(\cdot)$  via (10) and (11).

- $\bar{F}(x)$ ,  $\text{VaR}_\alpha$  and  $\text{ES}_\alpha$  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.
- In this spirit, Smith (1987) proposed the *tail estimator*

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

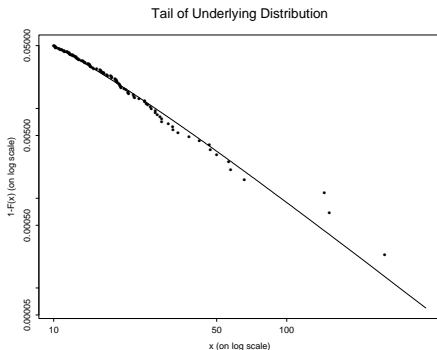
also known as the *Smith estimator* (note that it is only valid for  $x \geq u$ ). It faces a **bias-variance tradeoff**: If  $u$  is increased, the bias of parametrically estimating  $\bar{F}_u(x - u)$  decreases, but the variance of it and the nonparametrically estimated  $\bar{F}(u)$  increases.

- Similarly, GPD-based  $\widehat{\text{VaR}}_\alpha$ ,  $\widehat{\text{ES}}_\alpha$  for  $\alpha \geq 1 - N_u/n$  can be obtained from (12), (13).

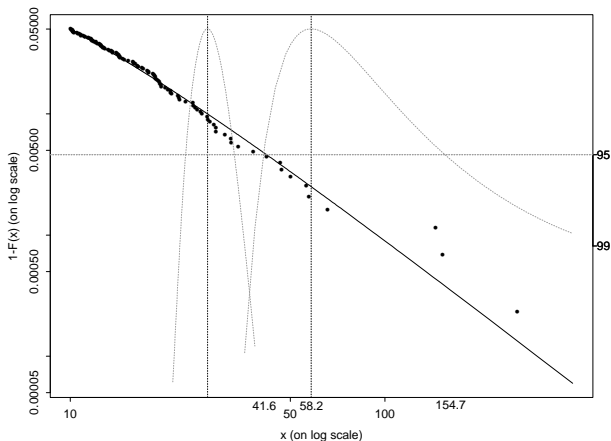
- Confidence intervals for  $\bar{F}(x)$ ,  $x \geq u$ ,  $\text{VaR}_\alpha$ ,  $\text{ES}_\alpha$  can be obtained likelihood-based (neglecting the uncertainty in  $N_u/n$ ): Reparametrize the GPD model in terms of  $\phi = g(\xi, \beta, N_u/n)$  and construct a confidence interval for  $\phi$  based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

### Example 5.20 (Danish fire loss data (continued))

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



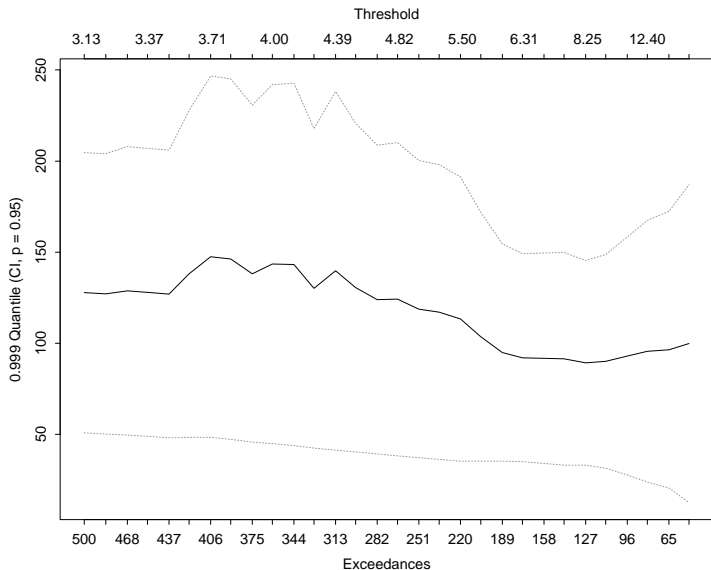
Here are  $\hat{\bar{F}}(x)$ ,  $x \geq u$ ,  $\widehat{\text{VaR}}_{0.99}$ ,  $\widehat{\text{ES}}_{0.99}$  including confidence intervals.



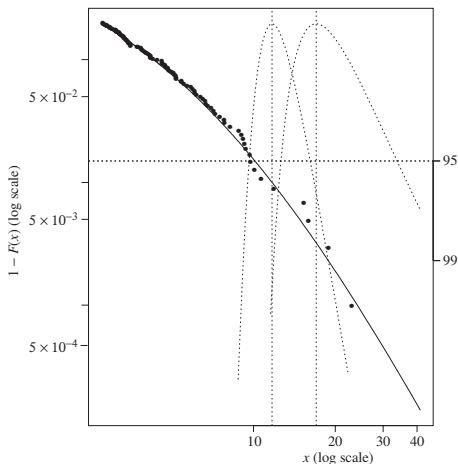
Log-log scale often helpful: If  $\bar{F}(x) = x^{-\alpha}L(x)$ ,  $\log \bar{F}(x) = -\alpha \log(x) + \log L(x)$  which is approximately linear in  $\log x$ .



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



## Example 5.21 (AT&T weekly loss data (continued))



- Fitted GPD model as in Example 5.19.
- Plot of  $\hat{F}(x)$ .
- Vertical lines:  $\widehat{\text{VaR}}_{0.99}$ ,  $\widehat{\text{ES}}_{0.99}$



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

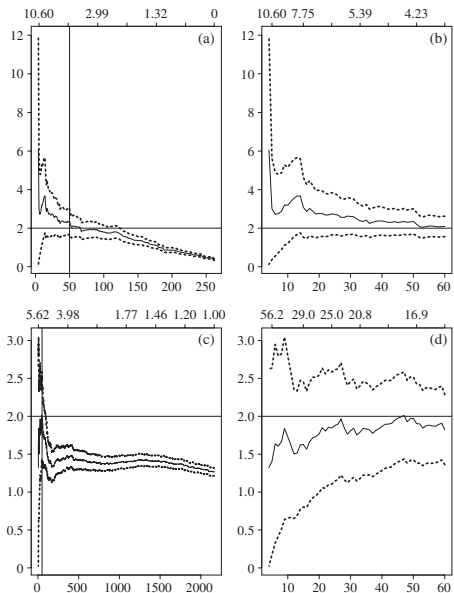
## 5.2.4 The Hill estimator

- Assume  $F \in \text{MDA}(H_\xi)$ ,  $\xi > 0$ , so that  $\bar{F}(x) = x^{-\alpha}L(x)$ ,  $\alpha > 0$ .
- The standard form of the *Hill estimator of the tail index  $\alpha$*  is

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

**Idea:** This can be derived by noting that the mean excess function  $e(\log u)$  of  $\log X$  at  $\log u$  is roughly  $1/\alpha$  for large  $u$  (by Karamata's Theorem), then using  $e_n(\log X_{k,n})$  as an estimator for  $e(\log u)$  and solving for  $\alpha$ ; see the appendix. Note:  $X_{1,n} \geq \dots \geq X_{n,n}$ .

- Choosing  $k$ : Find a small  $k$  where the *Hill plot*  $\{(k, \hat{\alpha}_{k,n}^{(H)}) : 2 \leq k \leq n\}$  stabilizes (typically,  $k = \lceil \beta n \rceil$ ,  $\beta \in [0.01, 0.05]$ ).
- Interpreting Hill plots can be difficult. 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 = zoomed-in version of the lhs).
- (a),(b) suggest estimates of  $\alpha \in [2, 4]$  ( $\xi \in [1/4, 1/2]$ ; clarger than the estimated  $\hat{\xi} = 0.22$ , see Example 5.19); (c),(d) suggest estimates of  $\alpha \in [1.5, 2]$  ( $\xi \in [1/2, 2/3]$  (infinite variance!); close to the estimated  $\hat{\xi} = 0.50$ , see Example 5.18)

## Hill-based tail and risk measure estimates

- Assume  $\bar{F}(x) = cx^{-\alpha}$ ,  $x \geq u > 0$  (replacing  $L$  by a constant). Estimate  $\alpha$  by  $\hat{\alpha}_{k,n}^{(H)}$  and  $u$  by  $X_{k,n}$  (for  $k$  sufficiently small).
- Note that  $c = u^\alpha \bar{F}(u)$  so  $\hat{c} = X_{k,n}^{\hat{\alpha}_{k,n}^{(H)}} \hat{\bar{F}}_n(X_{k,n}) \approx X_{k,n}^{\hat{\alpha}_{k,n}^{(H)}} \frac{k}{n}$ . We thus obtain the semi-parametric *Hill tail estimator*

$$\hat{\bar{F}}(x) = \frac{k}{n} \left( \frac{x}{X_{k,n}} \right)^{-\hat{\alpha}_{k,n}^{(H)}}, \quad x \geq X_{k,n}.$$

- From this result we obtain the semi-parametric *Hill VaR estimator*

$$\widehat{\text{VaR}}_\alpha(X) = \left( \frac{n}{k} (1 - \alpha) \right)^{-\frac{1}{\hat{\alpha}_{k,n}^{(H)}}} X_{k,n}, \quad \alpha \geq F(u) \approx 1 - \frac{k}{n},$$

and, for  $\hat{\alpha}_{k,n}^{(H)} > 1$ ,  $\alpha \geq F(u) \approx 1 - \frac{k}{n}$ , the semi-param. *Hill ES estimator*

$$\widehat{\text{ES}}_\alpha(X) = \frac{\left( \frac{n}{k} \right)^{-\frac{1}{\hat{\alpha}_{k,n}^{(H)}}} X_{k,n}}{1 - \alpha} \int_\alpha^1 (1 - z)^{-\frac{1}{\hat{\alpha}_{k,n}^{(H)}}} dz = \frac{\hat{\alpha}_{k,n}^{(H)}}{\hat{\alpha}_{k,n}^{(H)} - 1} \widehat{\text{VaR}}_\alpha(X).$$

## 5.2.5 Simulation study of EVT quantile estimators

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

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

with a Monte Carlo study (based on 1000 samples from a  $t_4$  distribution with corresponding true  $\xi = 1/4$ ) since analytical evaluation of bias and variance is not possible.

### Study 1: Estimating $\xi$

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

(a):  $\widehat{\text{MSE}}(\hat{\xi})$ ; (b):  $\widehat{\text{bias}}(\hat{\xi})$ ; (c):  $\widehat{\text{var}}(\hat{\xi})$  (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: the Hill (GPD) estimator tends to overestimate (underestimate)  $\xi$ .
- For the GPD method, the optimal  $u$  is around 100–150 exceedances.



## Study 2: Estimating $\text{VaR}_{0.99}$

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

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



- The empirical  $\text{VaR}_{0.99}$  estimator has a negative bias.
- The Hill  $\text{VaR}_{0.99}$  estimator has a negative bias for small  $k$  but a rapidly growing positive bias for larger  $k$ .
- The GPD  $\text{VaR}_{0.99}$  estimator has a positive bias which grows much more slowly.
- The GPD  $\text{VaR}_{0.99}$  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  $\text{VaR}_{0.99}$  estimator performs well for  $20 \leq k \leq 75$  (we only use  $k$  values that lead to a quantile estimate beyond the effective threshold  $X_{k,n}$ ) 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  $\mu_t$  and  $\sigma_t$  are  $\mathcal{F}_{t-1}$ -measurable and  $Z_t \stackrel{\text{ind.}}{\sim} F_Z$ ; e.g. ARMA model with GARCH errors. Furthermore, let  $Z \sim F_Z$ .

- $\text{VaR}_\alpha^t$  and  $\text{ES}_\alpha^t$  based on  $F_{X_{t+1}|\mathcal{F}_t}$  are given by

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

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

- To obtain estimates  $\widehat{\text{VaR}}_{\alpha}^t(X_{t+1})$  and  $\widehat{\text{ES}}_{\alpha}^t(X_{t+1})$ , 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  $\mu_{t+1}$  and  $\sigma_{t+1}$ .
  - 2) Fit a GPD to  $F_Z$  (treat the residuals from the GARCH fitting procedure as iid from  $F_Z$ )  $\Rightarrow$  GPD-based estimates of  $\text{VaR}_{\alpha}(Z)$  (see (12)) and  $\text{ES}_{\alpha}(Z)$  (see (13)).