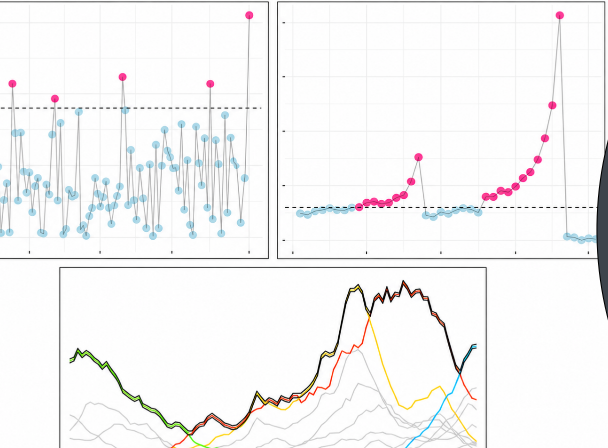
University of Stuttgart
Germany



Ioan Scheffel

Dependence in Extreme Value Theory: Tail Clustering and Maxima over time-changed Hidden Events

Mai 12 2026

Research Seminar unibz

Extreme value theory: modelling rare events

Statistical problem

- data contain only few extremes;

Extreme value theory: modelling rare events

Statistical problem

- data contain only few extremes;
- the target lies close or beyond the observed range;

Extreme value theory: modelling rare events

Statistical problem

- data contain only few extremes;
- the target lies close or beyond the observed range;
- naive empirical estimation fails.

Extreme value theory: modelling rare events

Statistical problem

- data contain only few extremes;
- the target lies close or beyond the observed range;
- naive empirical estimation fails.

Extreme value theory

- tail models from asymptotic theory;
- fit tail models to large observations;
- extrapolate .

Extreme value theory: modelling rare events

Statistical problem

- data contain only few extremes;
- the target lies close or beyond the observed range;
- naive empirical estimation fails.

Extreme value theory

- tail models from asymptotic theory;
- fit tail models to large observations;
- extrapolate .

observed (extreme) data $\longrightarrow$

Extreme value theory: modelling rare events

Statistical problem

- data contain only few extremes;
- the target lies close or beyond the observed range;
- naive empirical estimation fails.

Extreme value theory

- tail models from asymptotic theory;
- fit tail models to large observations;
- extrapolate .

observed (extreme) data $\longrightarrow$ fit tail model

Extreme value theory: modelling rare events

Statistical problem

- data contain only few extremes;
- the target lies close or beyond the observed range;
- naive empirical estimation fails.

Extreme value theory

- tail models from asymptotic theory;
- fit tail models to large observations;
- extrapolate .

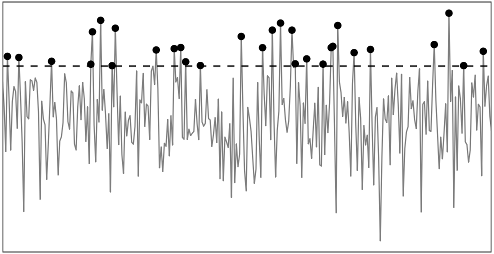
observed (extreme) data $\longrightarrow$ fit tail model $\longrightarrow$ rare-event extrapolation

Two extrapolation approaches

Peaks–over–Threshold (PoT)

$$X > u$$

First part of the talk



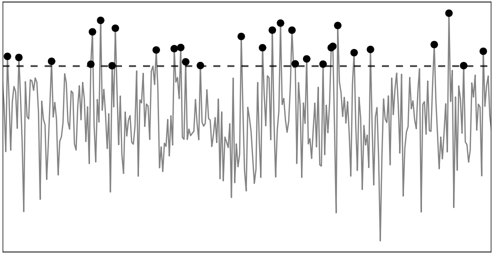
model fit from exceedances

Two extrapolation approaches

Peaks–over–Threshold (PoT)

$$X > u$$

First part of the talk

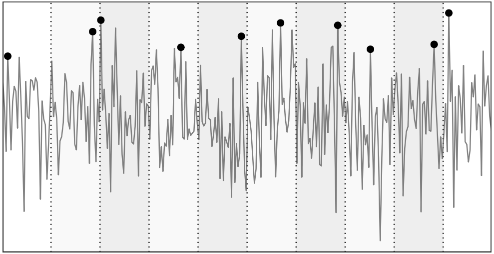


model fit from exceedances

Block maxima

$$M_n = \max_{1 \le i \le n} X_i$$

Second part of the talk



model fit from block maxima

Outline

Peaks-over-Threshold—Introduction

1

Regular variation

Assume that the tail is regularly varying (heavy tails):

$$\overline{F}(x) \;=\; \mathbb{P}(X > x) \;\approx\; x^{-1/\gamma} \qquad \text{with extreme value index} \qquad \gamma > 0\,.$$

Regular variation

Assume that the tail is regularly varying (heavy tails):

$$\overline{F}(x) \;=\; \mathbb{P}(X > x) \;\approx\; x^{-1/\gamma} \qquad \text{with extreme value index} \qquad \gamma > 0\,.$$

Examples:

- $\overline{F}(x) = x^{-1/\gamma} \cdot \mathbb{1}\{x > 1\}$ (Pareto distribution with shape $1/\gamma$)

Regular variation

Assume that the tail is regularly varying (heavy tails):

$$\overline{F}(x) \;=\; \mathbb{P}(X > x) \;\approx\; x^{-1/\gamma} \qquad \text{with extreme value index} \qquad \gamma > 0\,.$$

Examples:

- $\overline{F}(x) = x^{-1/\gamma} \cdot \mathbb{1}\{x > 1\}$ (Pareto distribution with shape $1/\gamma$)
- $\overline{F}(x) = (1 - \exp(-x^{-1/\gamma})) \cdot \mathbb{1}\{x > 0\}$ (Fréchet distribution with shape $1/\gamma$)

Regular variation

Assume that the tail is regularly varying (heavy tails):

$$\overline{F}(x) \;=\; \mathbb{P}(X > x) \;\approx\; x^{-1/\gamma} \qquad \text{with extreme value index} \qquad \gamma > 0\,.$$

Examples:

- $\overline{F}(x) = x^{-1/\gamma} \cdot \mathbb{1}\{x > 1\}$ (Pareto distribution with shape $1/\gamma$)
- $\overline{F}(x) = (1 - \exp(-x^{-1/\gamma})) \cdot \mathbb{1}\{x > 0\}$ (Fréchet distribution with shape $1/\gamma$)
- t-distribution with $\nu = 1/\gamma$ degrees of freedom

Regular variation

Assume that the tail is regularly varying (heavy tails):

$$\overline{F}(x) \;=\; \mathbb{P}(X > x) \;\approx\; x^{-1/\gamma} \qquad \text{with extreme value index} \qquad \gamma > 0\,.$$

Examples:

- $\overline{F}(x) = x^{-1/\gamma} \cdot \mathbb{1}\{x > 1\}$ (Pareto distribution with shape $1/\gamma$)
- $\overline{F}(x) = (1 - \exp(-x^{-1/\gamma})) \cdot \mathbb{1}\{x > 0\}$ (Fréchet distribution with shape $1/\gamma$)
- t-distribution with $\nu = 1/\gamma$ degrees of freedom
- normal distribution is not regularly varying (light tails)

Regular variation

Assume that the tail is regularly varying (heavy tails):

$$\overline{F}(x) \;=\; \mathbb{P}(X > x) \;\approx\; x^{-1/\gamma} \qquad \text{with extreme value index} \qquad \gamma > 0\,.$$

Examples:

- $\overline{F}(x) = x^{-1/\gamma} \cdot \mathbb{1}\{x > 1\}$ (Pareto distribution with shape $1/\gamma$)
- $\overline{F}(x) = (1 - \exp(-x^{-1/\gamma})) \cdot \mathbb{1}\{x > 0\}$ (Fréchet distribution with shape $1/\gamma$)
- t-distribution with $\nu = 1/\gamma$ degrees of freedom
- normal distribution is not regularly varying (light tails)
  Intuition: normal distribution is the limit of t-distribution for $\nu \to \infty$, that is, $\gamma \to 0$.

Regular variation

Assume that the tail is regularly varying (heavy tails):

$$\overline{F}(x) \; = \; \mathbb{P}(X > x) \; \approx \; x^{-1/\gamma} \qquad \text{with extreme value index} \qquad \gamma > 0\,.$$

Examples:

- $\overline{F}(x) = x^{-1/\gamma} \cdot \mathbb{1}\{x > 1\}$ (Pareto distribution with shape $1/\gamma$)
- $\overline{F}(x) = (1 - \exp(-x^{-1/\gamma})) \cdot \mathbb{1}\{x > 0\}$ (Fréchet distribution with shape $1/\gamma$)
- t-distribution with $\nu = 1/\gamma$ degrees of freedom
- normal distribution is not regularly varying (light tails)
  Intuition: normal distribution is the limit of t-distribution for $\nu \to \infty$, that is, $\gamma \to 0$.

Question: How to estimate extreme value index γ?

Idea: For Pareto distribution with shape $1/\gamma$ it holds

$$\int_u^\infty \frac{\overline{F}(x)}{x}\,\mathrm{d}x \;=\; \int_u^\infty x^{-1-1/\gamma}\,\mathrm{d}x \;=\; u^{-1/\gamma}\gamma \;=\; \overline{F}(u)\cdot\gamma\,.$$

Idea: For Pareto distribution with shape $1/\gamma$ it holds

$$\int_u^\infty \frac{\overline{F}(x)}{x}\,\mathrm{d}x \;=\; \int_u^\infty x^{-1-1/\gamma}\,\mathrm{d}x \;=\; u^{-1/\gamma}\gamma \;=\; \overline{F}(u)\cdot\gamma\,.$$

Generally: If F has a density f, it follows from partial integration

$$\int_u^\infty \frac{\overline{F}(x)}{x}\,\mathrm{d}x \;=\; \int_u^\infty (\log(x)-\log(u))\,f(x)\,\mathrm{d}x \;=\; \mathbb{E}[\log(X/u)\mathbb{1}\{X>u\}]$$

Idea: For Pareto distribution with shape $1/\gamma$ it holds

$$\int_u^\infty \frac{\overline{F}(x)}{x}\,\mathrm{d}x \;=\; \int_u^\infty x^{-1-1/\gamma}\,\mathrm{d}x \;=\; u^{-1/\gamma}\gamma \;=\; \overline{F}(u)\cdot\gamma\,.$$

Generally: If F has a density f, it follows from partial integration

$$\int_u^\infty \frac{\overline{F}(x)}{x}\,\mathrm{d}x \;=\; \int_u^\infty (\log(x)-\log(u))\,f(x)\,\mathrm{d}x \;=\; \mathbb{E}[\log(X/u)\mathbb{1}\{X>u\}]$$

Conclusion: The estimator

$$\frac{1}{n\mathbb{P}[X>u_n]}\sum_{i=1}^n \log(X_i/u_n)\mathbb{1}\{X_i>u_n\}$$

is unbiased for γ in the case of the Pareto distribution. It is called **Hill estimator.**

Idea: For Pareto distribution with shape $1/\gamma$ it holds

$$\int_u^\infty \frac{\overline{F}(x)}{x}\,\mathrm{d}x \;=\; \int_u^\infty x^{-1-1/\gamma}\,\mathrm{d}x \;=\; u^{-1/\gamma}\gamma \;=\; \overline{F}(u)\cdot\gamma\,.$$

Generally: If F has a density f, it follows from partial integration

$$\int_u^\infty \frac{\overline{F}(x)}{x}\,\mathrm{d}x \;=\; \int_u^\infty (\log(x)-\log(u))\,f(x)\,\mathrm{d}x \;=\; \mathbb{E}[\log(X/u)\mathbb{1}\{X>u\}]$$

Conclusion: The estimator

$$\frac{1}{n\mathbb{P}[X>u_n]}\sum_{i=1}^n \log(X_i/u_n)\mathbb{1}\{X_i>u_n\}$$

is unbiased for γ in the case of the Pareto distribution. It is called **Hill estimator.**

$\rightsquigarrow$ in general only asymptotically unbiased (Karamata's theorem)

Hill estimator

For threshold sequence (u_n) with $u_n \to \infty$ and $n\mathbb{P}[X_0 > u_n] \to \infty$ as $n \to \infty$

$$\frac{1}{n\mathbb{P}[X > u_n]} \sum_{i=1}^{n} \log\left(X_i/u_n\right) \mathbb{1}\{X_i > u_n\} \qquad \text{(deterministic thresholds)}$$

Hill estimator

For threshold sequence (u_n) with $u_n \to \infty$ and $n\mathbb{P}[X_0 > u_n] \to \infty$ as $n \to \infty$

$$\frac{1}{n\mathbb{P}[X > u_n]} \sum_{i=1}^{n} \log\left(X_i/u_n\right) \mathbb{1}\{X_i > u_n\} \qquad \text{(deterministic thresholds)}$$

In practice: From sample order statistics

$$X_{1:n} \leq \cdots \leq X_{n:n}\,,$$

choose k-th upper order statistic $X_{n-k:n}$ as threshold.

Hill estimator

For threshold sequence (u_n) with $u_n \to \infty$ and $n\mathbb{P}[X_0 > u_n] \to \infty$ as $n \to \infty$

$$\frac{1}{n\mathbb{P}[X > u_n]} \sum_{i=1}^{n} \log\left(X_i/u_n\right) \mathbb{1}\{X_i > u_n\} \qquad \text{(deterministic thresholds)}$$

In practice: From sample order statistics

$$X_{1:n} \leq \cdots \leq X_{n:n} \,,$$

choose k-th upper order statistic $X_{n-k:n}$ as threshold.

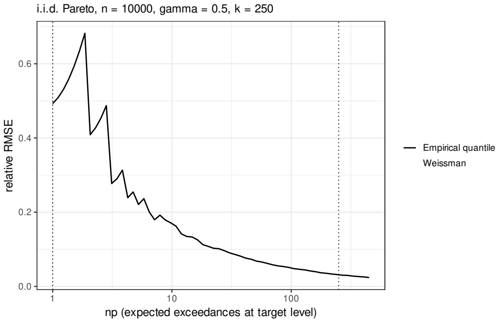
$$\rightsquigarrow \qquad \widehat{\gamma}_{k,n} \;=\; \frac{1}{k} \sum_{i=1}^{k} \log\left(X_{n-i+1:n}/X_{n-k:n}\right) \qquad \text{(random thresholds)}$$

Extrapolation vs. empirical quantile: the Weissman estimator

Setting:

- estimate large quantile q_{1-p}

  sample quantile $X_{n-\lfloor np \rfloor:n}$



Extrapolation vs. empirical quantile: the Weissman estimator

Setting:

- estimate large quantile q_{1-p}

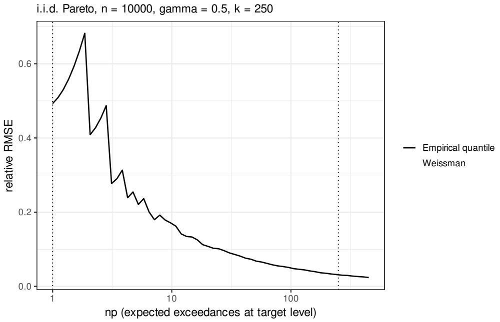
sample quantile $X_{n-\lfloor np \rfloor:n}$

Weissman estimator

$$X_{n-k:n}\left(\frac{k}{np}\right)^{\widehat{\gamma}_{k,n}}.$$

Idea:

Extrapolate from intermediate sample quantile $X_{n-k:n}$ by multiplying estimated tail decay $(k/(np))^{\widehat{\gamma}_{k,n}}$



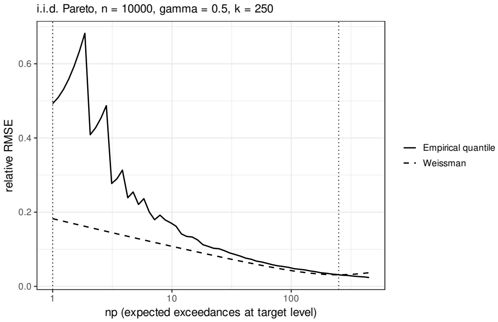
Extrapolation vs. empirical quantile: the Weissman estimator

Setting:

- estimate large quantile q_{1-p}
- compare sample quantile $X_{n-\lfloor np \rfloor:n}$ with *Weissman estimator*

$$X_{n-k:n}\left(\frac{k}{np}\right)^{\widehat{\gamma}_{k,n}}.$$

Idea:
Extrapolate from intermediate sample quantile $X_{n-k:n}$ by multiplying estimated tail decay $(k/(np))^{\widehat{\gamma}_{k,n}}$



$\rightsquigarrow$ Weissman estimator $X_{n-k:n}\left(\frac{k}{np}\right)^{\widehat{\gamma}_{k,n}}$ superior to sample quantile $X_{n-\lfloor np \rfloor:n}$ for estimating high quantiles

CLT for the Hill estimator under Tail Clustering

2

Central Limit Theorem for Hill estimator with i.i.d. Data

CLT for Hill Estimator with i.i.d. data

If X_0 has extreme value index $\gamma > 0$ and satisfies some second-order condition, and $n\mathbb{P}(X_0 > u_n) \to \infty$,

$$\sqrt{n\mathbb{P}(X_0 > u_n)}\left(\frac{\sum_{t=1}^{n}\log(X_t/u_n)\mathbb{1}\{X_t > u_n\}}{n\mathbb{P}(X_0 > u_n)} - \gamma\right) \to_d \mathcal{N}(0,\gamma^2) \qquad \text{as } n \to \infty\,.$$

Central Limit Theorem for Hill estimator with i.i.d. Data

CLT for Hill Estimator with i.i.d. data

If X_0 has extreme value index $\gamma > 0$ and satisfies some second-order condition, and $n\mathbb{P}(X_0 > u_n) \to \infty$,

$$\sqrt{n\mathbb{P}(X_0 > u_n)} \left(\frac{\sum_{t=1}^{n} \log(X_t/u_n) \mathbb{1}\{X_t > u_n\}}{n\mathbb{P}(X_0 > u_n)} - \gamma \right) \to_d \mathcal{N}(0, \gamma^2) \qquad \text{as } n \to \infty\,.$$

Conclusion:

effective sample size is smaller and speed of convergence is slower for PoT estimators (compared to classical estimators)

Long Memory Linear Time Series

Setting: (Causal) linear time series

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}\,, \qquad t \in \mathbb{N}_0$$

with

- symmetric i.i.d. innovations (ε_k)

Long Memory Linear Time Series

Setting: (Causal) linear time series

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}\,, \qquad t \in \mathbb{N}_0$$

with

- symmetric i.i.d. innovations (ε_k) with potentially heavy tails

$$\alpha := \sup\{m \in (1,2] \mid \text{the } m\text{-th moment of } \varepsilon \text{ exists}\}$$

$\rightsquigarrow$ we allow for distribution with infinite variance

Long Memory Linear Time Series

Setting: (Causal) linear time series

$$X_t = \sum_{k=0}^{\infty} a_k \varepsilon_{t-k}, \qquad t \in \mathbb{N}_0$$

with

- symmetric i.i.d. innovations (ε_k) with potentially heavy tails

$$\alpha := \sup\{m \in (1,2] \mid \text{the } m\text{-th moment of } \varepsilon \text{ exists}\}$$

$\rightsquigarrow$ we allow for distribution with infinite variance

- slowly decaying coefficients (a_k) with $a_k \sim k^{-1+d}$ with long memory parameter

$$0 < d < 1 - 1/\alpha$$

$\rightsquigarrow$ similar to fractional differencing parameter in fractional ARIMA

Asymptotics for Long Memory Linear Time Series

CLT for Hill Estimator (S., Oesting & Stupfler 2026)

Let (X_t) be a long memory linear time series with extreme value index $\gamma > 0$.

Under technical conditions on the distribution of the innovations (ε_k) and the growth rate of (u_n) it holds

$$n^{1-d-1/\alpha}\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}\left(\frac{\sum_{t=1}^n \log(X_t/u_n)\mathbb{1}\{X_t > u_n\}}{n\mathbb{P}(X_0 > u_n)} - \underbrace{\mathbb{E}(\log(X_0/u_n) \mid X_0 > u_n)}_{\to\gamma}\right)$$

$$\to_d \frac{1}{1/\gamma + 1} \cdot Z_\alpha$$

as $n \to \infty$, where

$$Z_\alpha \sim \begin{cases} S\alpha S\,, & \alpha \in (1,2)\,, \\ \mathcal{N}(0,1), & \alpha = 2 \text{ and variance exists.} \end{cases}$$

The Rate of Convergence

The rate of convergence

$$n^{1-d-1/\alpha}\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

can be decomposed into two factors:

- $n^{1-d-1/\alpha}$ corresponds to the rate in "classical" CLTs with long memory

  $\rightsquigarrow$ long memory slows down rate of convergence ($n^{1-d-1/\alpha} < n^{1/2}$)

- the ratio of tail function and marginal density

$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

$\rightsquigarrow$ describes the additional "PoT effect"

The Rate of Convergence

The rate of convergence

$$n^{1-d-1/\alpha}\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

can be decomposed into two factors:

- $n^{1-d-1/\alpha}$ corresponds to the rate in “classical” CLTs with long memory

  $\rightsquigarrow$ long memory slows down rate of convergence ($n^{1-d-1/\alpha} < n^{1/2}$)

- the ratio of tail function and marginal density

$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

$\rightsquigarrow$ describes the additional “PoT effect”

Reminder: In the i.i.d. setting, PoT effect slows down speed of convergence.

The Rate of Convergence

The rate of convergence

$$n^{1-d-1/\alpha}\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

can be decomposed into two factors:

- $n^{1-d-1/\alpha}$ corresponds to the rate in "classical" CLTs with long memory

  $\rightsquigarrow$ long memory slows down rate of convergence ($n^{1-d-1/\alpha} < n^{1/2}$)

- the ratio of tail function and marginal density

$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

$\rightsquigarrow$ describes the additional "PoT effect"

The Rate of Convergence

The rate of convergence

$$n^{1-d-1/\alpha}\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

can be decomposed into two factors:

- $n^{1-d-1/\alpha}$ corresponds to the rate in "classical" CLTs with long memory

  $\rightsquigarrow$ long memory slows down rate of convergence ($n^{1-d-1/\alpha} < n^{1/2}$)
- the ratio of tail function and marginal density

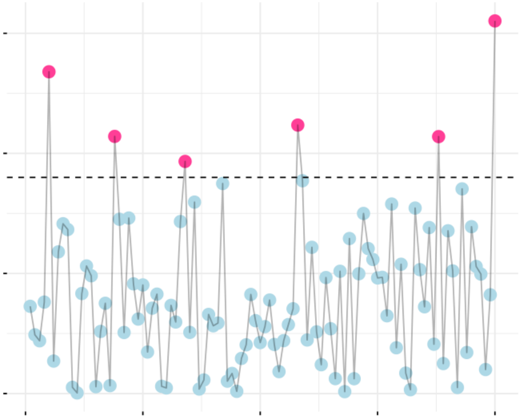
$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)}$$

$\rightsquigarrow$ describes the additional "PoT effect"

Reminder: In the i.i.d. setting, PoT effect slows down speed of convergence.

Ratio of Tail Function and Density

Case I: $X_0 \sim \mathcal{N}(0,1)$

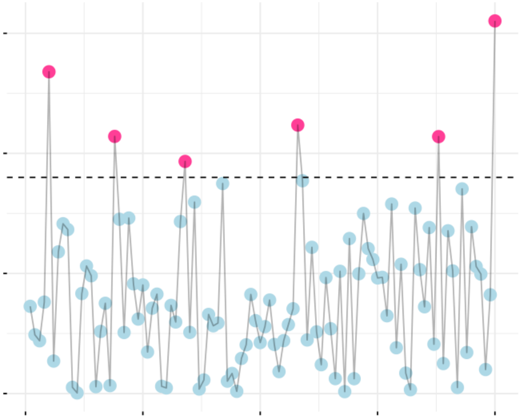


$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)} \sim \frac{1}{u_n}$$

as $n \to \infty$.

Ratio of Tail Function and Density

Case I: $X_0 \sim \mathcal{N}(0,1)$



$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)} \sim \frac{1}{u_n} \to 0$$

as $n \to \infty$.

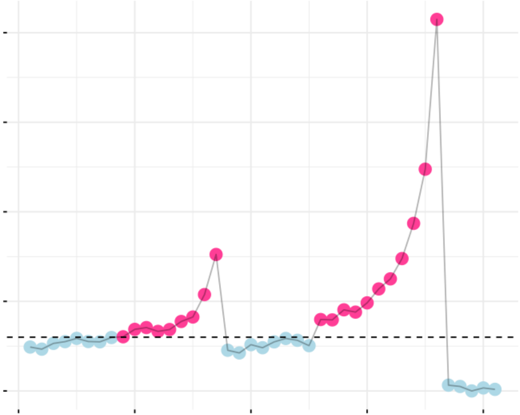
$\rightsquigarrow$ speed of convergence slows down due to “PoT effect” (similarly to the i.i.d. case)

Ratio of Tail Function and Density

Case II: X_0 with extreme value index $\gamma > 0$

$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)} \sim \gamma u_n$$

as $n \to \infty$.



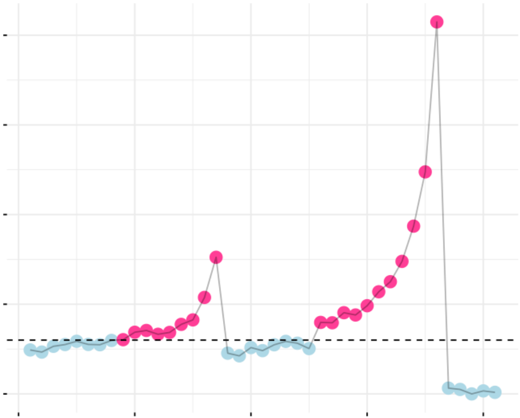
Ratio of Tail Function and Density

Case II: X_0 with extreme value index $\gamma > 0$

$$\frac{\mathbb{P}(X_0 > u_n)}{f_{X_0}(u_n)} \sim \gamma u_n \to \infty$$

as $n \to \infty$.

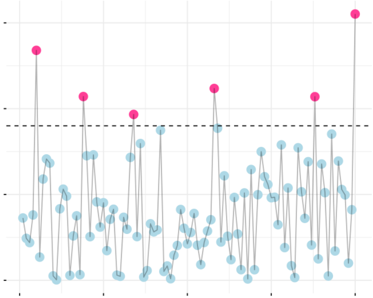
$\rightsquigarrow$ speed of convergence gets faster due to "PoT effect" (in contrast to the i.i.d. case)

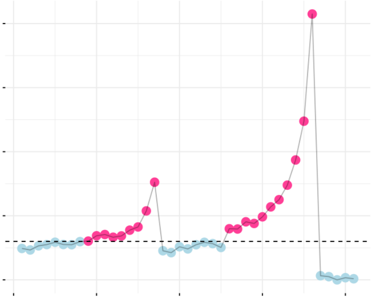


Summary

In Long Memory Linear Time Series . . .

- . . . there are both long memory and PoT effects affecting convergence rates of PoT estimators
- . . . convergence rate of PoT estimators depends on heavy/light tails.
- . . . with heavy tails there is even a speed-up when looking at extremes.





Block Maxima—Introduction

3

Max-stable limits: asymptotic models for block maxima

Setting:
$X_1, \ldots, X_n$ i.i.d. block maxima.

Max-stable limits: asymptotic models for block maxima

Setting:
$X_1, \ldots, X_n$ i.i.d. block maxima.
Assume that there exist constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, as $n \to \infty$,

$$\frac{\max_{i=1,\ldots,n} X_i \; - \; b_n}{a_n} \qquad \text{converges to a non-degenerate random variable} \qquad Z\,.$$

Max-stable limits: asymptotic models for block maxima

Setting:
$X_1, \ldots, X_n$ i.i.d. block maxima.
Assume that there exist constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, as $n \to \infty$,

$$\frac{\max_{i=1,\ldots,n} X_i \; - \; b_n}{a_n} \qquad \text{converges to a non-degenerate random variable} \qquad Z\,.$$

(Fisher–Tippett 1928, Gnedenko 1943)

In the above setting, Z has a generalized extreme value distribution (GEV)

Max-stable limits: asymptotic models for block maxima

Setting:
$X_1, \ldots, X_n$ i.i.d. block maxima.
Assume that there exist constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, as $n \to \infty$,

$$\frac{\max_{i=1,\ldots,n} X_i \; - \; b_n}{a_n} \qquad \text{converges to a non-degenerate random variable} \qquad Z\,.$$

(Fisher–Tippett 1928, Gnedenko 1943)

In the above setting, Z has a generalized extreme value distribution (GEV), that is, the three parametric family

$$\mathbb{P}[Z \le z] \;=\; \begin{cases} \exp\left(-\left(1+\xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi}\right), & \xi \neq 0\,, \\ \exp\left(-\exp\left(-\frac{x-\mu}{\sigma}\right)\right), & \xi = 0\,, \end{cases}$$

with location $\mu \in \mathbb{R}$, scale $\sigma > 0$, and shape $\xi \in \mathbb{R}$, and with support on $\{1+\xi(x-\mu)/\sigma > 0\}$.

Examples:

- Gumbel distribution $G(x) = \exp(-\exp(x))$

Examples:

- Gumbel distribution $G(x) = \exp(-\exp(x))$
- Fréchet distribution $G(x) = \exp(-x^{-\alpha})$ with shape $\alpha > 0$

Examples:

- Gumbel distribution $G(x) = \exp(-\exp(x))$
- Fréchet distribution $G(x) = \exp(-x^{-\alpha})$ with shape $\alpha > 0$

We say that a distribution F is in the **max-domain of attraction** (MDA) of a GEV G

Examples:

- Gumbel distribution $G(x) = \exp(-\exp(x))$
- Fréchet distribution $G(x) = \exp(-x^{-\alpha})$ with shape $\alpha > 0$

We say that a distribution F is in the **max-domain of attraction** (MDA) of a GEV G, if there exist constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, as $n \to \infty$,

$$\left(F(x \cdot a_n + b_n)\right)^n \;\to\; G(x)\,.$$

Examples:

- Gumbel distribution $G(x) = \exp(-\exp(x))$
- Fréchet distribution $G(x) = \exp(-x^{-\alpha})$ with shape $\alpha > 0$

We say that a distribution F is in the **max-domain of attraction** (MDA) of a GEV G, if there exist constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, as $n \to \infty$,

$$(F(x \cdot a_n + b_n))^n \ \to \ G(x)\,.$$

Alternative Definition: For i.i.d. $X_1, X_2, \ldots$ having distribution F

$$\mathbb{P}\left[\frac{\max_{i=1,\ldots,n} X_i - b_n}{a_n} \ \le \ x\right] \ = \ \mathbb{P}\left[\max_{i=1,\ldots,n} X_i \ \le \ x \cdot a_n + b_n\right]$$
$$= \ \mathbb{P}\left[\forall_{i=1,\ldots,n} X_i \ \le \ x \cdot a_n + b_n\right] \overset{\text{i.i.d}}{=} (F(x \cdot a_n + b_n))^n \overset{n\to\infty}{\to} G(x) \ = \ \mathbb{P}[Z \le x]$$

Examples:

- Gumbel distribution $G(x) = \exp(-\exp(x))$
- Fréchet distribution $G(x) = \exp(-x^{-\alpha})$ with shape $\alpha > 0$

We say that a distribution F is in the **max-domain of attraction** (MDA) of a GEV G, if there exist constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, as $n \to \infty$,

$$(F(x \cdot a_n + b_n))^n \ \to \ G(x)\,.$$

Alternative Definition: For i.i.d. $X_1, X_2, \ldots$ having distribution F

$$\mathbb{P}\left[\frac{\max_{i=1,\ldots,n} X_i - b_n}{a_n} \ \leq \ x\right] = \mathbb{P}\left[\max_{i=1,\ldots,n} X_i \ \leq \ x \cdot a_n + b_n\right]$$
$$= \mathbb{P}\left[\forall_{i=1,\ldots,n} X_i \ \leq \ x \cdot a_n + b_n\right] \overset{\text{i.i.d}}{=} (F(x \cdot a_n + b_n))^n \overset{n\to\infty}{\to} G(x) \ = \ \mathbb{P}[Z \leq x]$$

The *univariate* theory of Fisher–Tippett–Gnedenko describes **marginal extremes.**

Examples:

- Gumbel distribution $G(x) = \exp(-\exp(x))$
- Fréchet distribution $G(x) = \exp(-x^{-\alpha})$ with shape $\alpha > 0$

We say that a distribution F is in the **max-domain of attraction** (MDA) of a GEV G, if there exist constants $a_n > 0$, $b_n \in \mathbb{R}$, such that, as $n \to \infty$,

$$(F(x \cdot a_n + b_n))^n \ \to\ G(x)\,.$$

Alternative Definition: For i.i.d. $X_1, X_2, \ldots$ having distribution F

$$\mathbb{P}\left[\frac{\max_{i=1,\ldots,n} X_i - b_n}{a_n} \ \leq\ x\right] \ =\ \mathbb{P}\left[\max_{i=1,\ldots,n} X_i \ \leq\ x \cdot a_n + b_n\right]$$

$$=\ \mathbb{P}\left[\forall_{i=1,\ldots,n} X_i \ \leq\ x \cdot a_n + b_n\right] \overset{\text{i.i.d}}{=} (F(x \cdot a_n + b_n))^n \overset{n\to\infty}{\to} G(x) \ =\ \mathbb{P}[Z \leq x]$$

The *univariate* theory of Fisher–Tippett–Gnedenko describes **marginal extremes.**

To describe **extremal dependence**, we must pass to **joint limits across several sites.**

From one site to several sites

Setting: Let S be a (spatio-)temporal domain, and let

$$X_i = \{X_i(s) : s \in S\}, \qquad i = 1, 2, \ldots,$$

be i.i.d. random processes, for example, collecting block maxima across all sites $s \in S$.

From one site to several sites

Setting: Let S be a (spatio-)temporal domain, and let

$$X_i = \{X_i(s) : s \in S\}, \qquad i = 1, 2, \dots,$$

be i.i.d. random processes, for example, collecting block maxima across all sites $s \in S$.
For each site $s \in S$, define

$$M_n(s) = \max_{1 \le i \le n} X_i(s).$$

We have seen that, at each fixed site $s \in S$, if the rescaled maximum

$$\frac{M_n(s) - b_n(s)}{a_n(s)} \qquad \text{converges to a non-degenerate limit} \qquad Z(s)\,,$$

then $Z(s)$ has GEV distribution.

From one site to several sites

Setting: Let S be a (spatio-)temporal domain, and let

$$X_i = \{X_i(s) : s \in S\}, \qquad i = 1, 2, \dots,$$

be i.i.d. random processes, for example, collecting block maxima across all sites $s \in S$. For each site $s \in S$, define

$$M_n(s) = \max_{1 \le i \le n} X_i(s).$$

We have seen that, at each fixed site $s \in S$, if the rescaled maximum

$$\frac{M_n(s) - b_n(s)}{a_n(s)} \qquad \text{converges to a non-degenerate limit} \qquad Z(s)\,,$$

then $Z(s)$ has GEV distribution.

Question: What about the dependence structure of $(Z(s))_{s \in S}$?

Spectral Representation of Max-Stable Processes (de Haan 1984)

Let $Z = \{Z(s) : s \in S\}$ be a max-stable process with standard Gumbel margins. Then Z can be represented as

$$Z(s) = \bigvee_{i=1}^{\infty} \left\{ U^{(i)} + W^{(i)}(s) \right\}, \qquad s \in S,$$

where $\{U_i\}_{i \geq 1}$ are the points of a Poisson point process on $\mathbb{R}$ with intensity

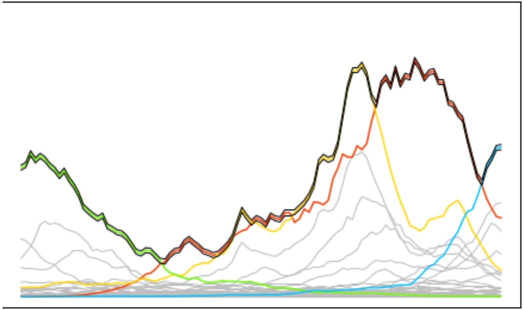
$$e^{-u}\, du,$$

and $W^{(1)}, W^{(2)}, \ldots$ are independent copies of a process W satisfying

$$\mathbb{E}\left[e^{W(s)} \right] = 1, \qquad s \in S.$$

Interpretation: Maxima over hidden events

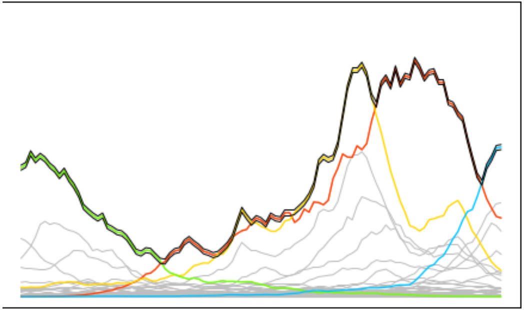
$$Z(s) = \bigvee_{i=1}^{\infty} \left\{ \underbrace{U_i}_{\text{event magnitude}} + \underbrace{W_i(s)}_{\text{event shape}} \right\}.$$



Interpretation: Maxima over hidden events

$$Z(s) = \bigvee_{i=1}^{\infty} \left\{ \underbrace{U_i}_{\text{event magnitude}} + \underbrace{W_i(s)}_{\text{event shape}} \right\}.$$

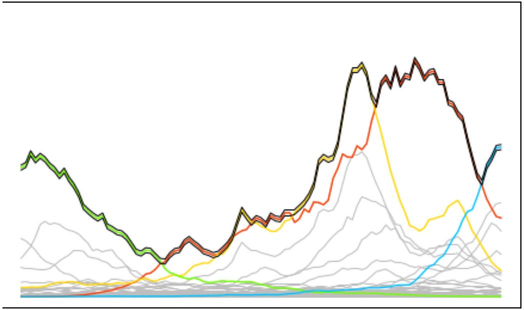
- Event **evolves independently** of the event magnitude $U^{(i)}$.



Interpretation: Maxima over hidden events

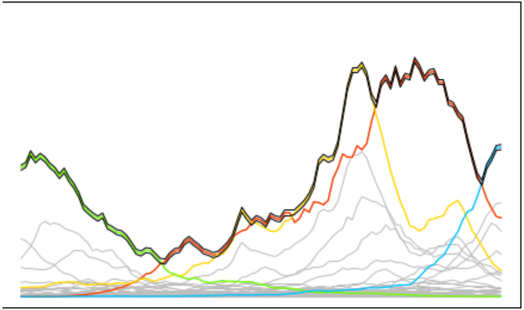
$$Z(s) = \bigvee_{i=1}^{\infty} \left\{ \underbrace{U_i}_{\text{event magnitude}} + \underbrace{W_i(s)}_{\text{event shape}} \right\}.$$

- Event **evolves independently** of the event magnitude $U^{(i)}$.
- Stationary law of W $\Rightarrow$ stationarity of Z



Interpretation: Maxima over hidden events

$$Z(s) = \bigvee_{i=1}^{\infty} \left\{ \underbrace{U_i}_{\text{event magnitude}} + \underbrace{W_i(s)}_{\text{event shape}} \right\}.$$

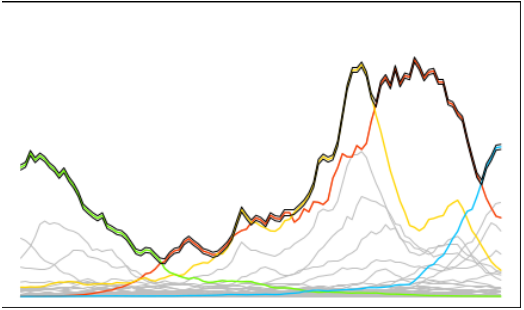


- Event **evolves independently** of the event magnitude $U^{(i)}$.
- Stationary law of W
  $\Rightarrow$ stationarity of Z
- **Non-stationary** law of W
  $\Rightarrow$ **still stationarity** of Z
  under additional conditions
  see Kabluchko, Schlather, and de Haan (2009)

Interpretation: Maxima over hidden events

$$Z(s) = \bigvee_{i=1}^{\infty} \left\{ \underbrace{U_i}_{\text{event magnitude}} + \underbrace{W_i(s)}_{\text{event shape}} \right\}.$$

- Event **evolves independently** of the event magnitude $U^{(i)}$.
- Stationary law of W
  $\Rightarrow$ stationarity of Z
- **Non-stationary** law of W
  $\Rightarrow$ **still stationarity** of Z
  under additional conditions
  see Kabluchko, Schlather, and de Haan (2009)



Question: How to model evolution depending on event magnitude?

Maxima over time-changed hidden events

4

Idea:
Keep the hidden-event interpretation, but allow for dependent event evolution:

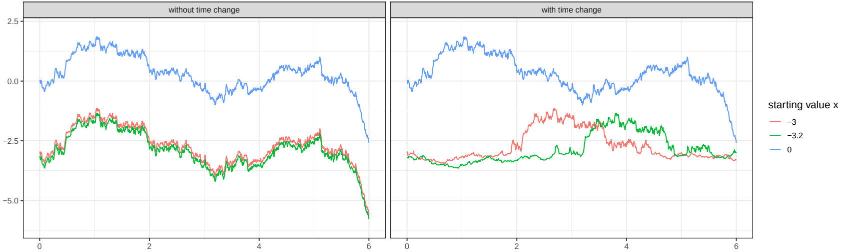
Idea:

Keep the hidden-event interpretation, but allow for dependent event evolution:

$$\underbrace{A_x(t)}_{\text{inverse random clock}} = \int_0^t \underbrace{\alpha(x + W(s))}_{\text{mass function}} \, \mathrm{d}s \,, \qquad \underbrace{T_x(t)}_{\text{random clock}} = A_x^{-1}(t)$$

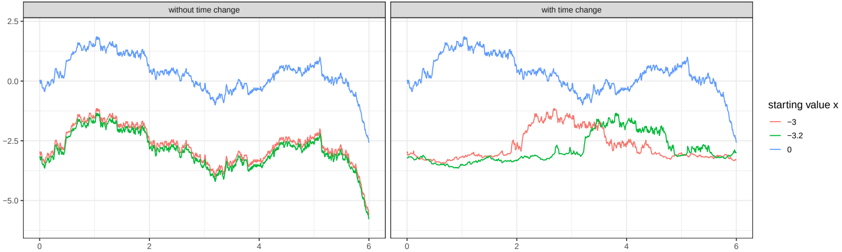
Idea:
Keep the hidden-event interpretation, but allow for dependent event evolution:

$$\underbrace{A_x(t)}_{\text{inverse random clock}} = \int_0^t \underbrace{\alpha(x+W(s))}_{\text{mass function}} \, \mathrm{d}s \,, \qquad \underbrace{T_x(t)}_{\text{random clock}} = A_x^{-1}(t)$$



Idea:
Keep the hidden-event interpretation, but allow for dependent event evolution:

$$\underbrace{A_x(t)}_{\text{inverse random clock}} = \int_0^t \underbrace{\alpha(x + W(s))}_{\text{mass function}} \, \mathrm{d}s \,, \qquad \underbrace{T_x(t)}_{\text{random clock}} = A_x^{-1}(t)$$



Consider adapted process $X(s) = \bigvee_{i=1}^{\infty} \{ \underbrace{U^{(i)}}_{\text{event magnitude}} + \underbrace{W^{(i)}(T_{U^{(i)}}(t))}_{\text{reactive event evolution}} \}.$

Construction of stationary processes in the MDA (S. 2026+)

Let $S = [0, \infty)$, and let W be a Lévy process satisfying $\mathbb{E}[\exp(W(1)] = 1$. Let α be an admissible mass function, and write

$$A_x(t) \;=\; \int_0^t \alpha(x + W(s))\,\mathrm{d}s\,, \qquad T_x(t) \;=\; A_x^{-1}(t)\,.$$

Consider the process

$$X(s) = \bigvee_{i=1}^{\infty} \left\{ U^{(i)} + W^{(i)}(T_{U^{(i)}}(t)) \right\}, \qquad s \in S,$$

where $\{U_i\}_{i \geq 1}$ are the points of a Poisson point process on $\mathbb{R}$ with intensity

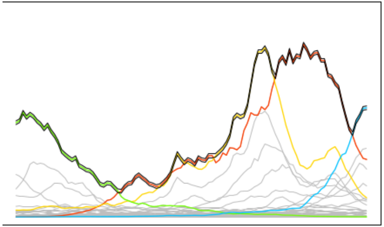
$$\alpha(u)e^{-u}\,du,$$

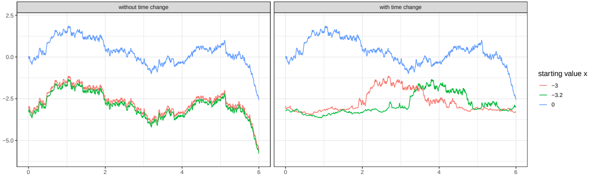
and $W^{(1)}, W^{(2)}, \ldots$ are i.i.d. copies of W.
Then $X = \{X(s)\}_{s \in S}$ is well-defined, stationary and $X \in \mathrm{MDA}(Z)$

Summary

Maxima over time-changed hidden events...

- ... yield flexible stationary models and preserving extremal behaviour trough MDA
- ... are built to preserve stationarity





References

L. D. Haan.
A Spectral Representation for Max-stable Processes.
The Annals of Probability, 12(4):1194–1204.

Z. Kabluchko, M. Schlather, and L. de Haan.
Stationary max-stable fields associated to negative definite functions.
The Annals of Probability, 37(5):2042–2065, 2009.

I. Scheffel.
Maxima of stationary systems of randomly time-changed lévy particles.
arXiv preprint, 2026.

I. Scheffel, M. Oesting, and G. Stupfler.
Central limit theory for peaks-over-threshold partial sums of long memory linear time series.
Stochastic Processes and their Applications, 197:104929, 2026.