Extremes for Time Series [1]

Olivier Wintenberger

Sorbonne University

http://wintenberger.fr/

Based on a long-term collaboration with T. Mikosch

[1]CUSO, February 2023

Contents

1. Heavy tails in real-life data

1.1. Finance.

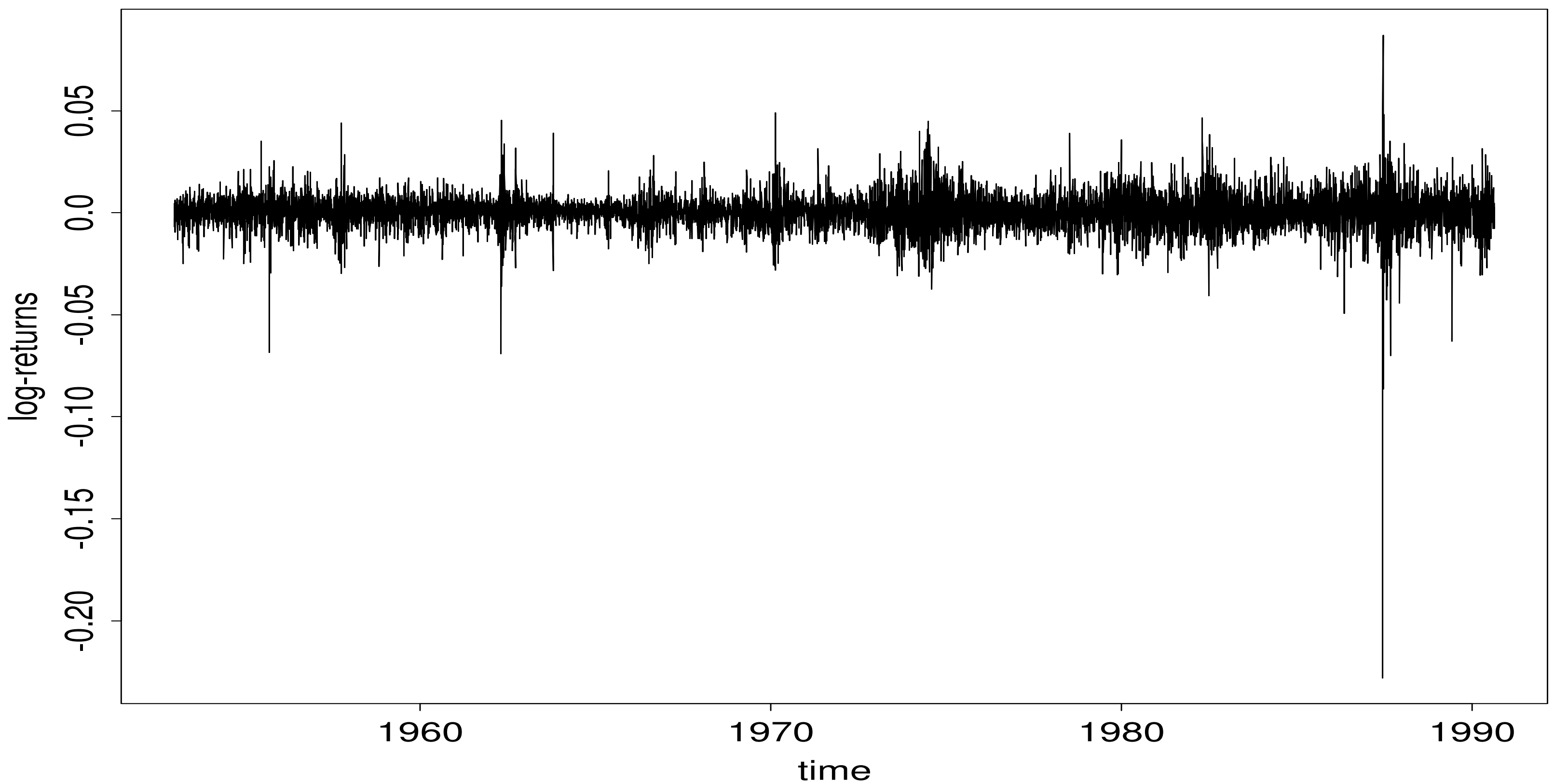
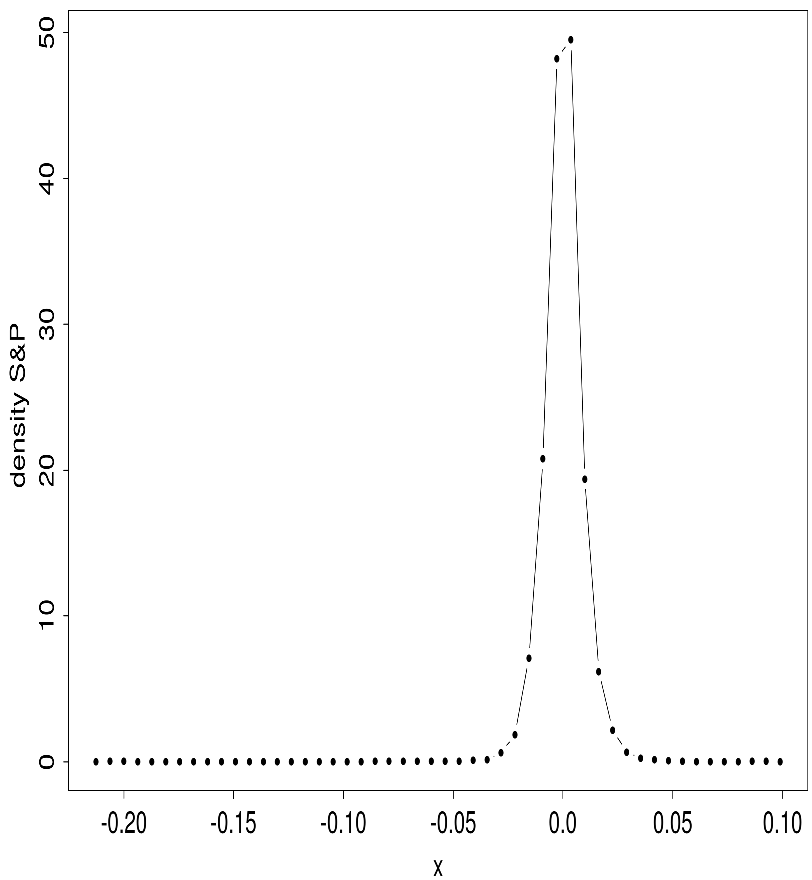


Figure 1. Plot of **9558** *S&P500* daily log-returns from January *2, 1953,* to December *31, 1990.* The year marks indicate the beginning of the calendar year.



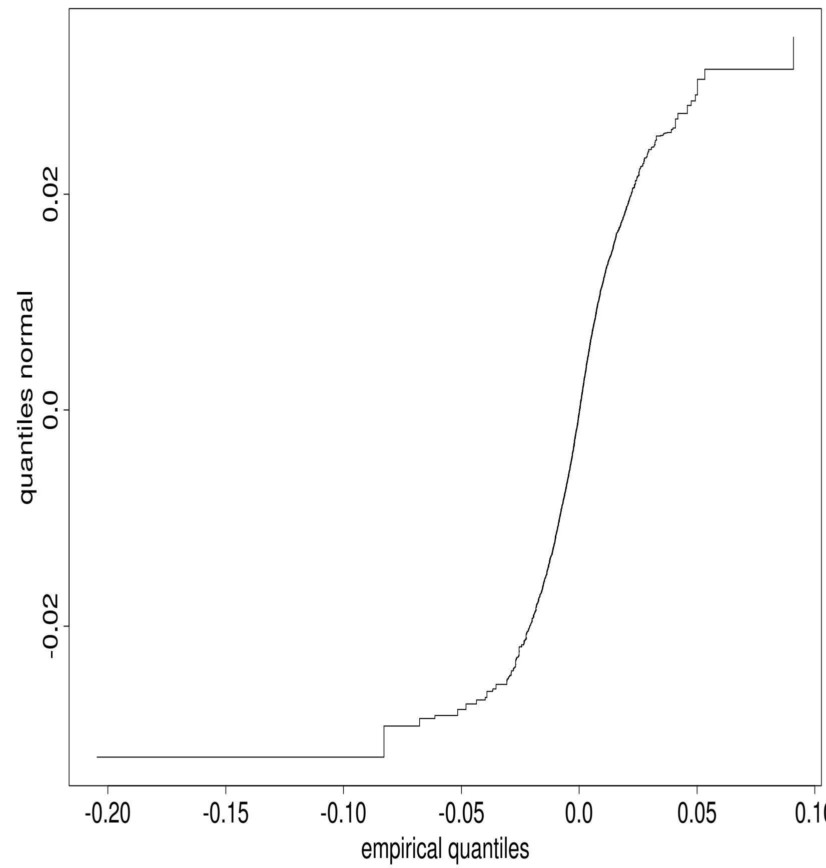


FIGURE 2. Left: Density plot of the *S&P500* data. The limits on the $\boldsymbol{x}$-axis indicate the range of the data. QQ-plot of the *S&P500* data against the normal distribution.

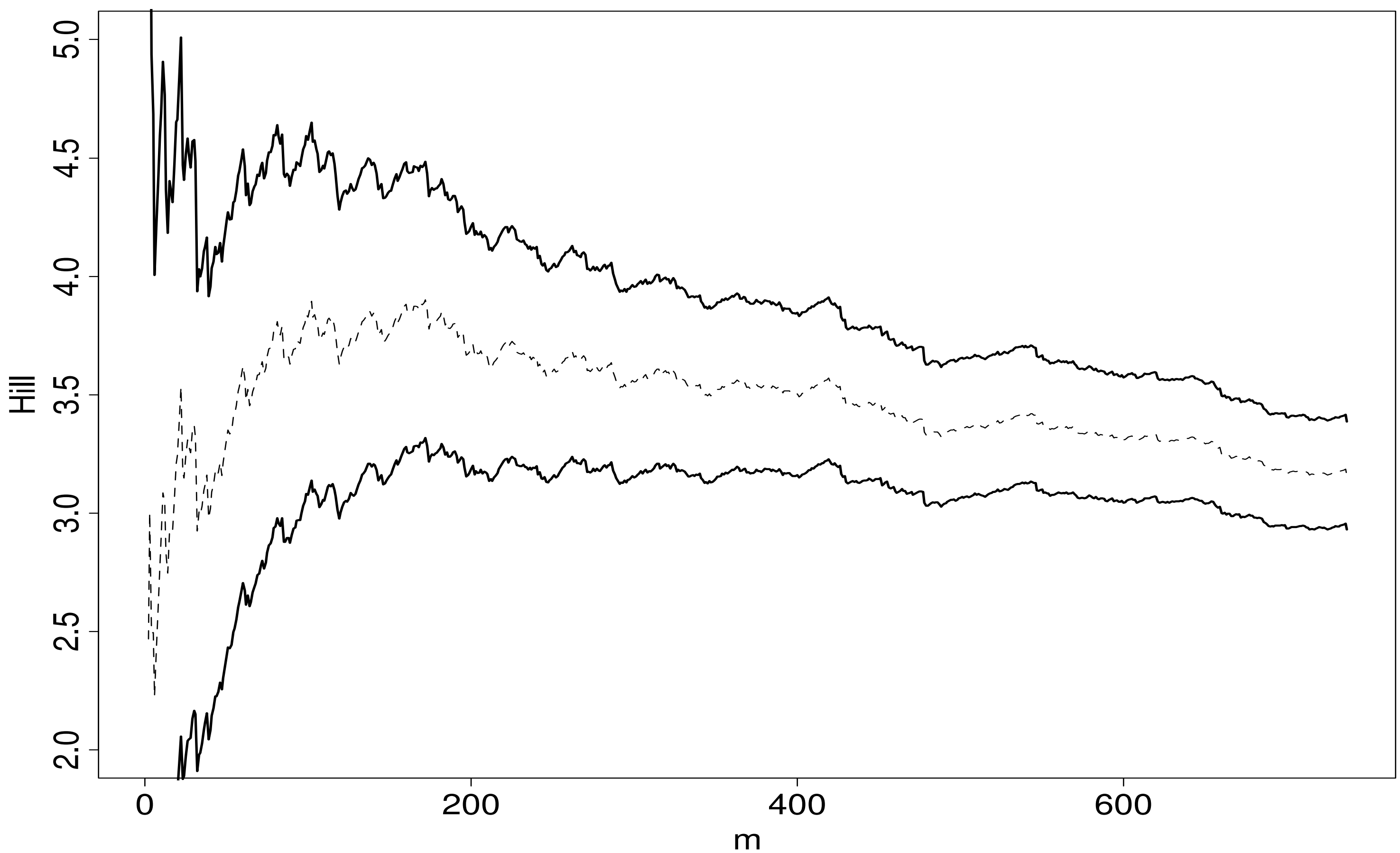


FIGURE 3. Hill plot (dotted line) for the *S&P500 data* with **95%** asymptotic confidence bounds. The Hill estimator approximates the tail index α in the model $\mathbb{P}(X > x) \sim c\,x^{-\alpha}$ as a function of the m upper order statistics in the return sample.

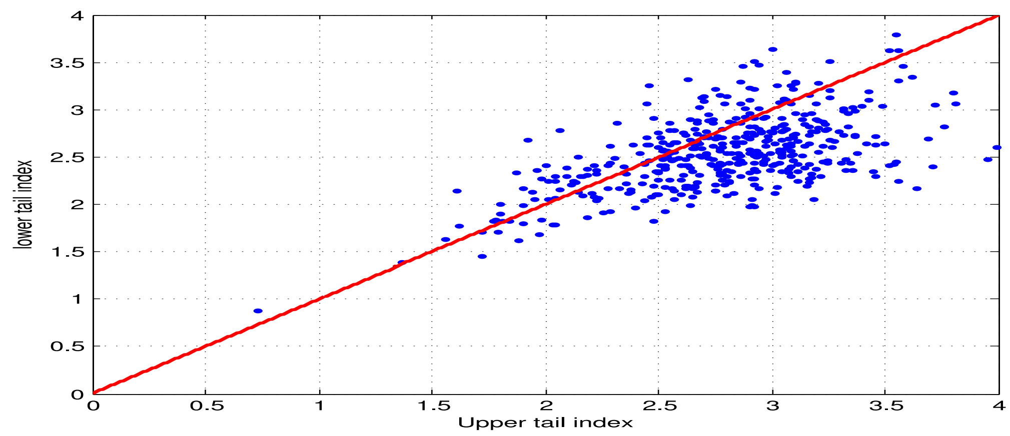


FIGURE 4. Hill estimates of the upper and lower tail indices (gains and losses) for the 500 time series of the *S&P500 index*.

1.2. **Insurance.**

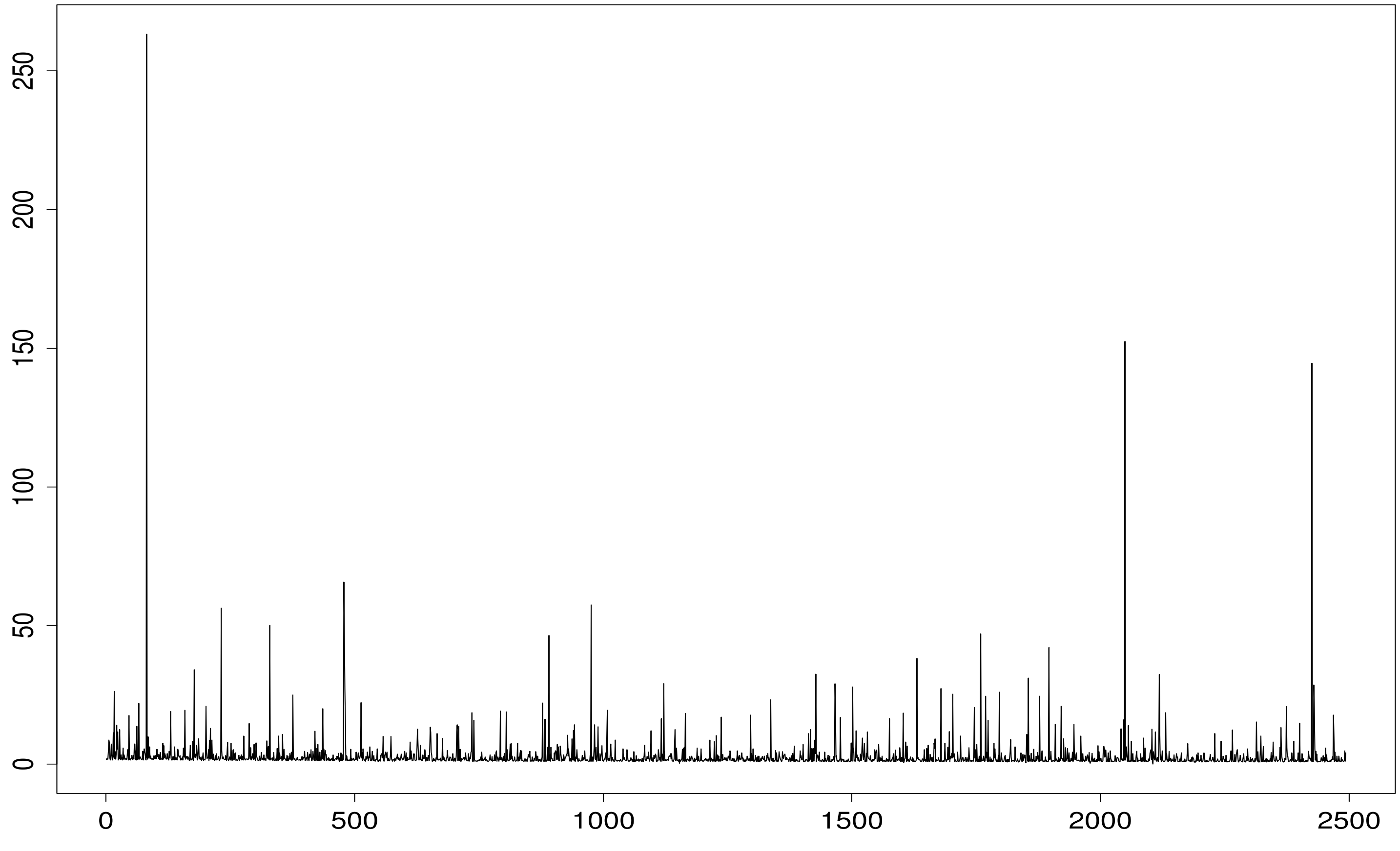


Figure 5. Danish fire insurance data losses.

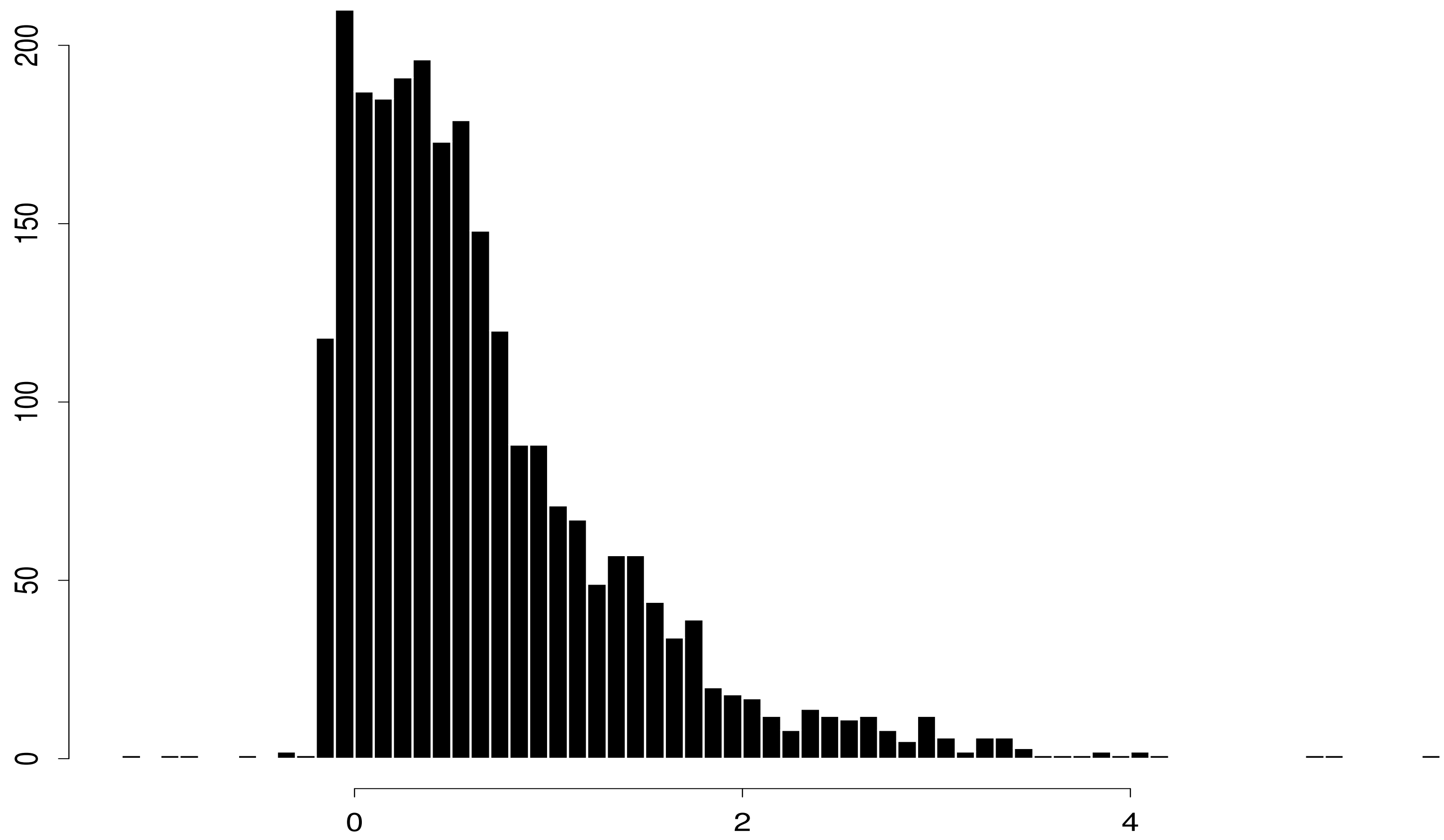


FIGURE 6. Histogram of the logarithmic Danish fire insurance losses.

2. Extremal dependence/independence in real-life data

2.1. Independence in insurance data.

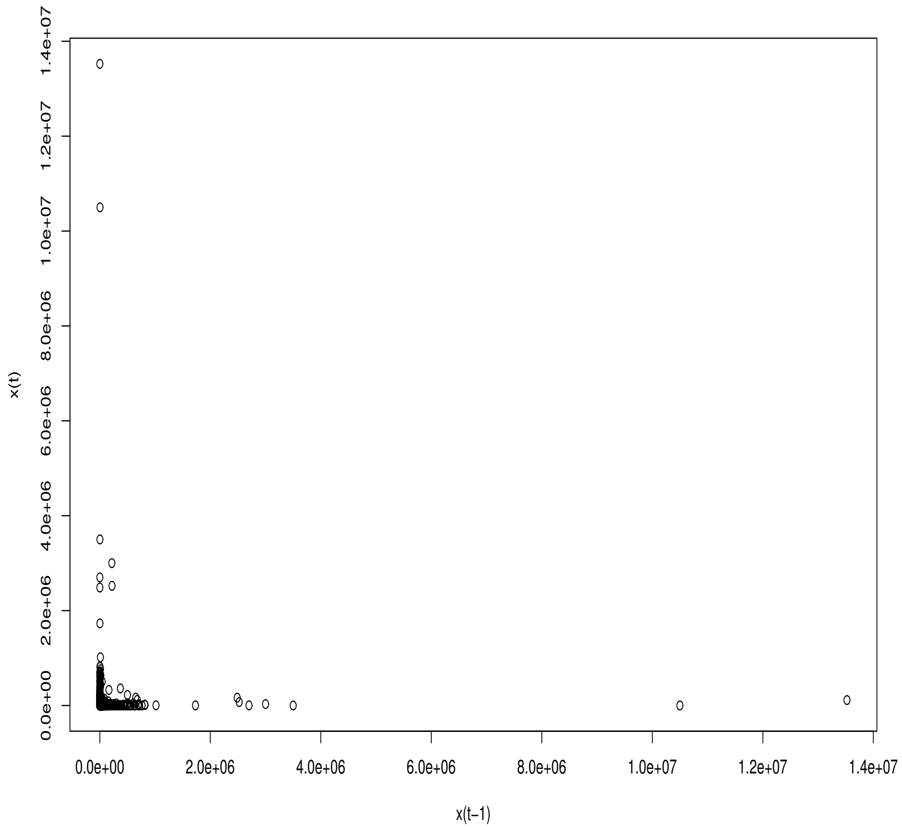


Figure 7. Scatterplot of one day-lag US fire insurance losses - independence.

2.2. Extremal dependence in financial data.

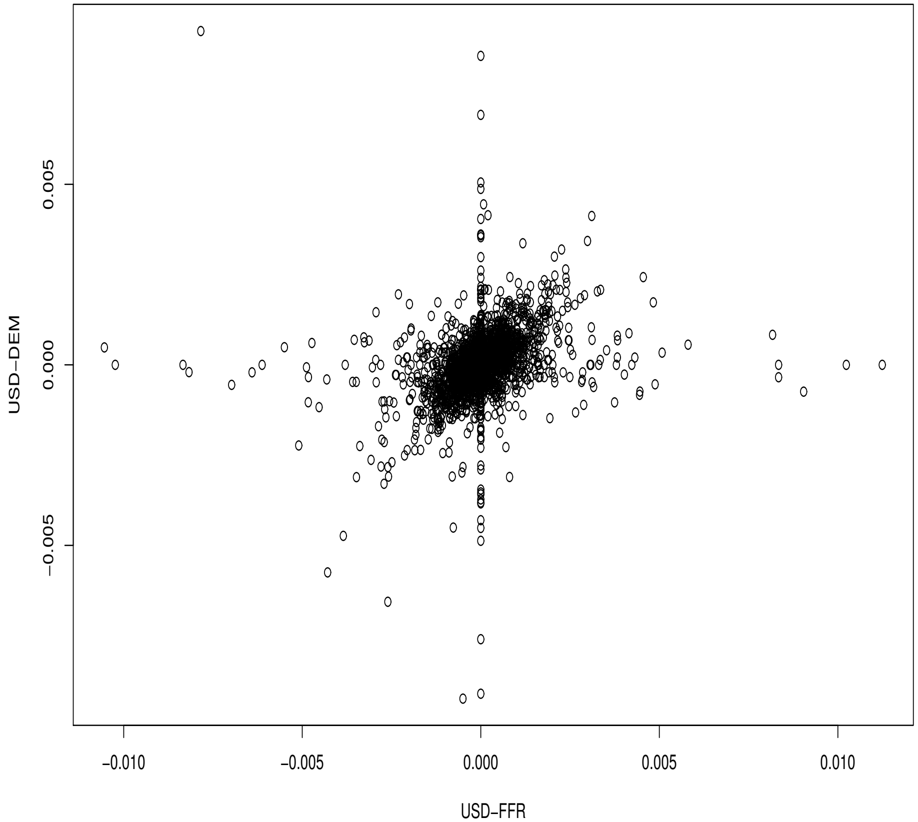


Figure 8. Scatterplot of 5 minute foreign exchange rate log-returns, USD-DEM against USD-FRF.

3. Extreme value theory for an iid univariate sequence Resnick (1987, 2007), Embrechts, Klüppelberg, Mikosch (1997), de Haan, Ferreira (2006)

3.1. Max-stable distributions (extreme value distributions).

- **A random variable X and its distribution F are max-stable if for every $n \geq 2$ there exist $c_n > 0$, $d_n \in \mathbb{R}$, such that for iid copies (X_i) of X,**

$$c_n^{-1}(M_n - d_n) = c_n^{-1}\big(\max_{i=1,\dots,n} X_i - d_n\big) \stackrel{d}{=} X \,.$$

- **Any max-stable distribution belongs to the location/scale family of one of the three standard max-stable distributions:**

$$\begin{array}{llll} \Phi_\alpha(x) = e^{-x^{-\alpha}}, & x > 0, & \alpha > 0 & \text{Fréchet} \\ \Psi_\alpha(x) = e^{-|x|^{\alpha}}, & x < 0, & \alpha > 0 & \text{Weibull} \\ \Lambda(x) = e^{-e^{-x}}. & x \in \mathbb{R}, & & \text{Gumbel} \end{array}$$

- **The max-stable distributions are the only possible non-degenerate weak limits for standardized maxima of an iid sequence (Fisher-Tippett Theorem 1928, Gnedenko (1943)).**

3.2. Maximum domains of attraction (MDA).

- **The distribution F of X is in the maximum domain of attraction of the max-stable distribution $G \in \{\Phi_\alpha, \Psi_\alpha, \Lambda\}$ ($F \in \mathrm{MDA}(G)$) if there exist constants $a_n > 0$, $b_n \in \mathbb{R}$ such that**

$$\lim_{n\to\infty} \mathbb{P}(a_n^{-1}(M_n - b_n) \le x) \to G(x)\,, \quad x \in \mathbb{R}\,.$$

- **Examples:**

MDA(Φ_α): Student with α degrees of freedom,

Cauchy ($\alpha = 1$),

infinite variance α-stable distributions,

Pareto $\overline{F}(x) = x^{-\alpha}$, $x > 1$,

log-gamma distribution.

MDA(Ψ_α): uniform, β-distribution.

MDA(Λ): log-normal distribution,

Weibull $\overline{F}(x) = \mathrm{e}^{-x^\tau}$, $x > 0$, $\tau > 0$,

gamma distribution,

normal distribution.

- **$F \in \mathrm{MDA}(\Phi_\alpha)$: Regular variation of the right tail**

$$\overline{F}(x) = 1 - F(x) = \mathbb{P}(X > x) = x^{-\alpha} L(x)\,, \quad x > 0\,,$$

for a slowly varying function L:

$L(xy)/L(x) \to 1$, $x \to \infty$, for every $y > 0$.

Then moments of order $\alpha - \delta$, $\delta > 0$, are finite, and $\alpha + \delta$, $\delta > 0$, are infinite.

4. Univariate regular variation

4.1. **Definition and properties.** **The random variable X and its distribution F are regularly varying with (tail) index $\alpha \geq 0$, $X \in \mathrm{RV}(\alpha)$, if**

$$\mathbb{P}(\pm X > x) \sim p_{\pm} \frac{L(x)}{x^{\alpha}}, \qquad x \to \infty,$$

where L is slowly varying and $p_+ + p_- = 1$.

- **Equivalently,**

$$\lim_{x\to\infty} \frac{\mathbb{P}(\pm X > c\,x)}{\mathbb{P}(|X| > x)} = p_\pm\, c^{-\alpha} \qquad \text{for every } c > 0.$$

or for $0 < a < b$

$$\lim_{x\to\infty} \frac{\mathbb{P}(x^{-1}X \in (a, b])}{\mathbb{P}(|X| > x)} = p_+(a^{-\alpha} - b^{-\alpha})$$

$$\lim_{x\to\infty} \frac{\mathbb{P}(x^{-1}X \in (-b, -a])}{\mathbb{P}(|X| > x)} = p_-(a^{-\alpha} - b^{-\alpha})\,.$$

• **Equivalently,**

$$\lim_{x\to\infty} n\,\mathbb{P}(\pm X > c\,a_n) = p_\pm\, c^{-\alpha} \qquad \text{for every } c > 0,$$

for some sequence $a_n \to \infty$ **satisfying** $n\,\mathbb{P}(|X| > a_n) \to 1$,

e.g. $a_n = F^{\leftarrow}_{|X|}(1 - 1/n)$,

and $F \in MDA(\Phi_\alpha)$:

$$\lim_{n\to\infty} \mathbb{P}(a_n^{-1} M_n \le x) \to \Phi_\alpha^{p_+}(x)\,, \quad x \in \mathbb{R}\,.$$

4.2. Operations on regularly varying random variables.

4.2.1. *Convolution.*

- **Assume $X \in \mathrm{RV}(\alpha)$ and**
  - **either Y is independent of X and $Y \in \mathrm{RV}(\alpha)$**
  - **or $\mathbb{P}(|Y| > x) = o(\mathbb{P}(|X| > x))$ as $x \to \infty$.**

  Then

Feller's convolution lemma

$$\mathbb{P}(X + Y > x) \sim \mathbb{P}(X > x) + \mathbb{P}(Y > x)\,, \qquad x \to \infty\,.$$

- **Example**

  X_i iid, $X \in \mathrm{RV}(\alpha)$ and $p_+p_- > 0$. Then for all $n \geq 2$,

$$\mathbb{P}\big(\pm(X_1 + \cdots + X_n) > x\big) \sim n\,\mathbb{P}(\pm X > x)\,, \qquad x \to \infty\,.$$

- **Example**

  If $\mathbb{P}(|Y| > x) = o(\mathbb{P}(|X| > x))$ then

$$\mathbb{P}(\pm(X + Y) > x) \sim \mathbb{P}(\pm X > x)\,, \qquad x \to \infty\,.$$

- Example
  - Assume (Z_t) iid with $Z \in \mathrm{RV}(\alpha)$ for some $\alpha > 0$, $\mathbb{E}[Z] = 0$ (if expectation exists) and real (ψ_j) such that $\sum_j \psi_j^{2\wedge(\alpha-\varepsilon)} < \infty$.
  - Then the strictly stationary linear process

$$X_t = \sum_{j=0}^{\infty} \psi_j \, Z_{t-j} \,, \qquad t \in \mathbb{Z}$$

is well defined and by Feller's lemma

$$\begin{aligned} \mathbb{P}(X > x) &\sim \sum_{j=0}^{\infty} \mathbb{P}(\psi_j Z_j > x) \\ &\sim \mathbb{P}(|Z| > x) \sum_{j=0}^{\infty} \left[p_+ \, (\psi_j)_+^{\alpha} + p_- \, (\psi_j)_-^{\alpha} \right]. \end{aligned}$$

– **For example, (X_t) is AR(1):**

$$X_t = \varphi X_{t-1} + Z_t\,, \qquad t \in \mathbb{Z}\,, \qquad |\varphi| < 1\,.$$

Then $X_t = \sum_{j=0}^{\infty} \varphi^j Z_{t-j}$ **and**

$$\mathbb{P}(X > x) \sim \mathbb{P}(|Z| > x) \sum_{j=0}^{\infty} \left[p_+ (\varphi^j)_+^\alpha + p_- (\varphi^j)_-^\alpha\right].$$

The single big jump heuristics for a linear process with $X = \sum_j \psi_j Z_j$ Mikosch, W. (2023+)

$$\lim_{x\to\infty} \mathbb{P}\Big(\bigcup_{j\in\mathbb{Z}} |\psi_j Z_j| > x \,\Big|\, |X| > x\Big) = 1.$$

$$\lim_{x\to\infty} \mathbb{P}\Big(\bigcup_{j\neq k\in\mathbb{Z}} \{|\psi_j Z_j| > x \,, |\psi_k Z_k| > x\} \,\Big|\, |X| > x\Big) = 0.$$

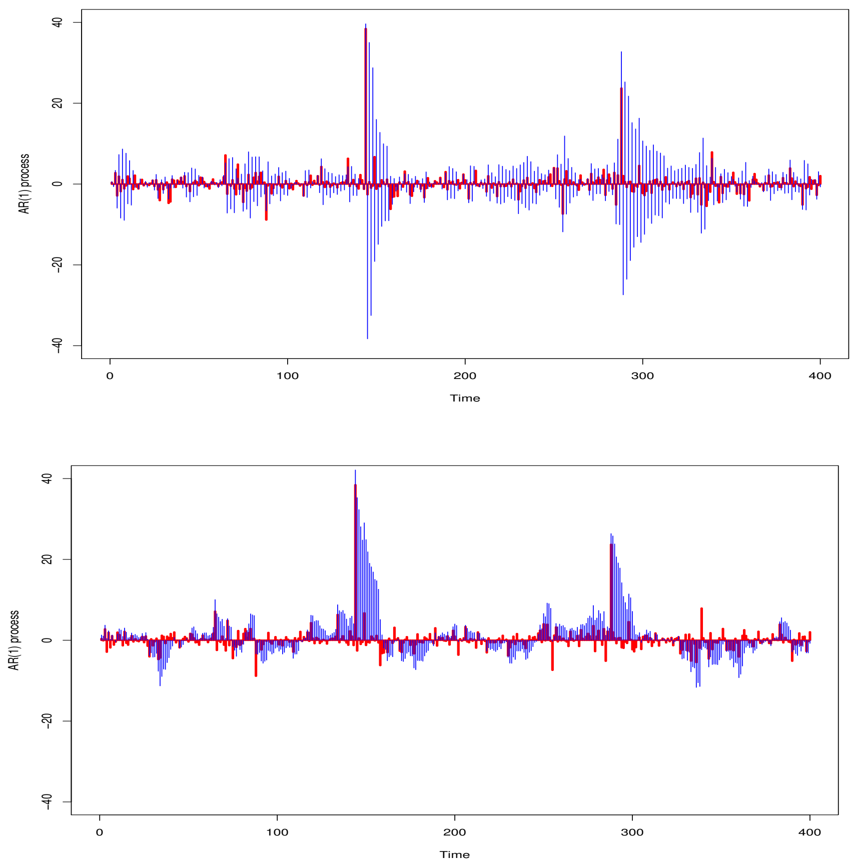


FIGURE 9. Visualization of a sample path of an AR(1) process $\boldsymbol{X_t = \varphi X_{t-1} + Z_t}$, $\boldsymbol{t = 1, \ldots, 400}$ (blue) with $\boldsymbol{\varphi = -0.9}$ **(top)** and $\boldsymbol{\varphi = 0.9}$ **(bottom)**. The sample path of the noise $\boldsymbol{(Z_t)_{t=1,\ldots,400}}$ (red) comes from a Student(2) distribution and is the same in both graphs.

4.2.2. *Multiplication.* **Assume $X_1, X_2 > 0$ independent, $X_1 \in \mathrm{RV}(\alpha)$ for some $\alpha > 0$**

- **If in addition**
  - **either $X_2 \in \mathrm{RV}(\alpha)$**
  - **or $\mathbb{P}(X_2 > x) = o(\mathbb{P}(X_1 > x))$**

  then $X_1 X_2 \in \mathrm{RV}(\alpha)$.
- **If in addition $\mathbb{E}[X_2^{\alpha+\varepsilon}] < \infty$ for some $\varepsilon > 0$ then**

Breiman's lemma

$$\mathbb{P}(X_1 X_2 > x) \sim \mathbb{E}[X_2^{\alpha}]\,\mathbb{P}(X_1 > x)\,, \qquad x \to \infty\,.$$

- Example
  - Consider a strictly stationary positive volatility process (σ_t) independent of the iid sequence (Z_t) with $Z \in \mathrm{RV}(\alpha)$.
  - The stochastic volatility process

$$X_t = \sigma_t Z_t, \qquad t \in \mathbb{Z},$$

is strictly stationary.

  - If $\mathbb{E}[\sigma^{\alpha+\varepsilon}] < \infty$ for some $\varepsilon > 0$ then by Breiman

$$\mathbb{P}(\pm X_t > x) \sim \mathbb{E}[\sigma^\alpha]\, \mathbb{P}(\pm Z > x), \qquad x \to \infty.$$

  - The latter result holds for a log-normal σ, e.g.

$$\log \sigma_t = \sum_{j=0}^{\infty} \psi_j \, \eta_{t-j}, \qquad t \in \mathbb{Z},$$

with $\sum_j \psi_j^2 < \infty$ and iid standard normal (η_t).

- Example
  - Consider (Z_t) iid standard normal.
  - Assume the affine stochastic recurrence equation for the squared volatility sequence (σ_t^2) has a strictly stationary solution for suitable positive $\alpha_0, \alpha_1, \beta_1$:[2]

$$\sigma_t^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta_1)\,\sigma_{t-1}^2\,, \qquad t \in \mathbb{Z}\,.$$

The strictly stationary process given by

$$X_t = \sigma_t\, Z_t\,, \qquad t \in \mathbb{Z}\,,$$

is a GARCH(1, 1) process Bollerslev (1986).[a]

[a] Generalized Auto Regressive Conditionally Heteroscedastic process of order (1,1).

[2] A positive solution exists iff $\mathbb{E}[\log(\alpha_1 Z^2 + \beta_1)] < 0$ and $\alpha_0 > 0$.

The equation $\mathbb{E}[(\alpha_1 Z^2 + \beta_1)^{\alpha/2}] = 1$ **has a unique positive solution** α **and for some** $c_0 > 0$**,**

The Kesten-Goldie Theorem[a] **Kesten (1973), Goldie (1991)**

$$\mathbb{P}(\sigma_t > x) \sim c_0 \, x^{-\alpha}, \qquad x \to \infty .$$

[a] See also Buraczewski, Damek, Mikosch (2016)

and by Breiman's lemma

$$\mathbb{P}(\pm X_t > x) = \mathbb{P}(\pm \sigma_t \, Z_t > x) \sim \mathbb{E}[(Z_\pm)^\alpha] \, \mathbb{P}(\sigma > x), \quad x \to \infty .$$

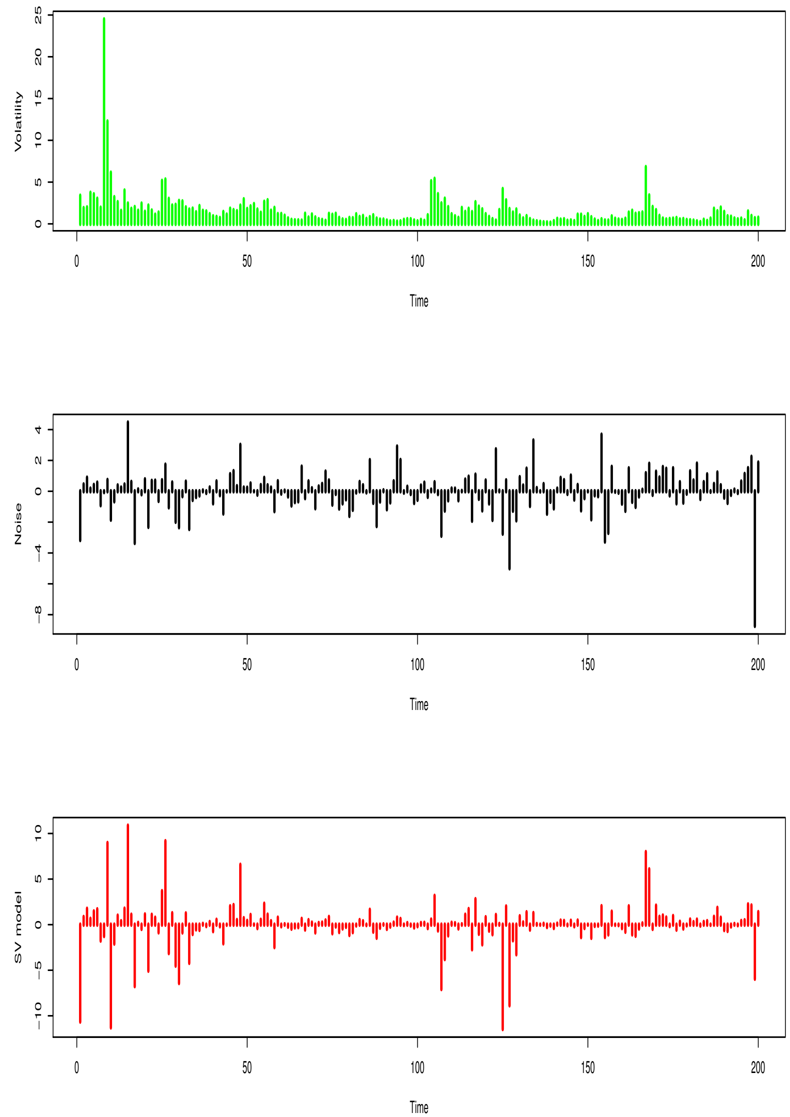


FIGURE 10. Visualization of a stochastic volatility process $(X_t)_{1\le t\le 200}$ **(bottom)** with Student noise (Z_t) with 2 degrees of freedom **(middle)** and the corresponding volatility process (σ_t) **(top)**. The processes (Z_t) and (σ_t) are independent. The log-volatility process is an AR(1) process given by the difference equation $\log \sigma_t = 0.9 \log \sigma_{t-1} + \eta_t$, $t \in \mathbb{Z}$, with iid centered exponential noise (η_t) with mean $1/4$. An application of Breiman's result shows that $\mathbb{P}(\sigma > x) \sim c\, x^{-4}$ as $x \to \infty$. We see that the extreme values of (Z_t) trigger the extremes of (X_t).

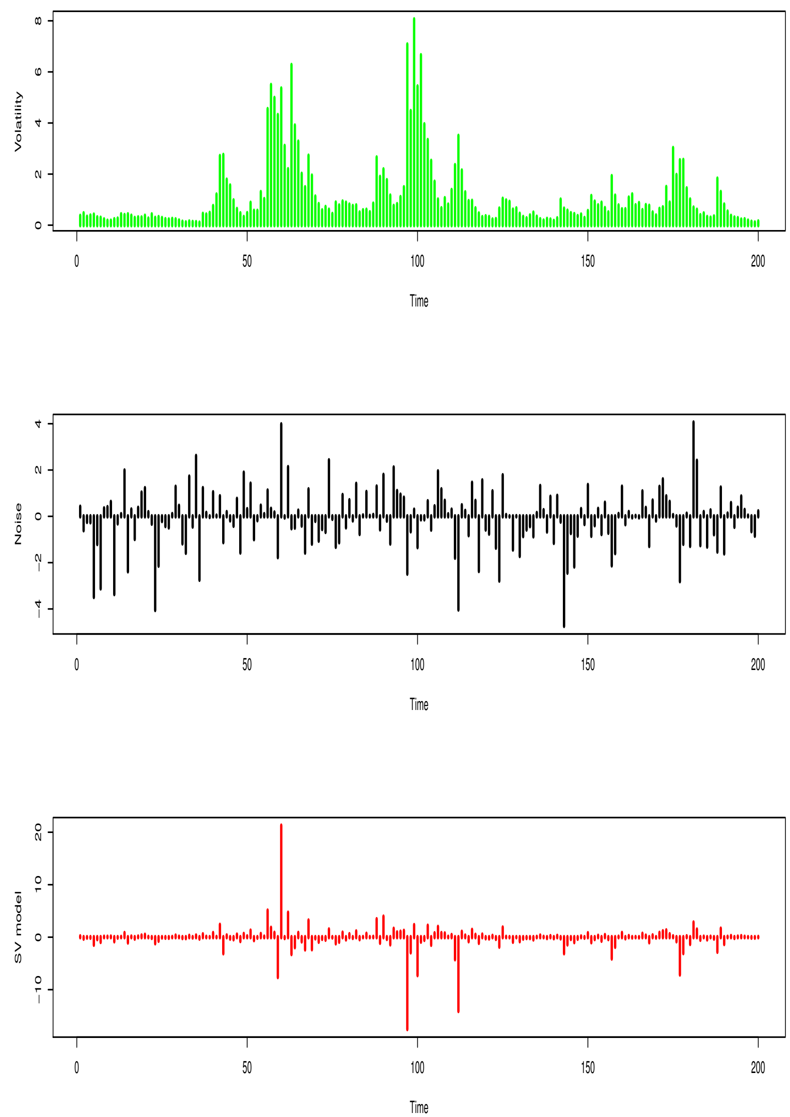


FIGURE 11. Visualization of a stochastic volatility process $(X_t)_{1\le t\le 200}$ **(bottom)** with Student noise (Z_t) with 8 degrees of freedom **(middle)** and the corresponding volatility process (σ_t) **(top)**. The processes (Z_t) and (σ_t) are independent. The log-volatility process is an AR(1) process given by the difference equation $\log\sigma_t = 0.9\,\log\sigma_{t-1}+\eta_t$, $t\in\mathbb{Z}$, with iid centered exponential noise (η_t) with mean $1/4$. An application of Breiman's result shows that $\mathbb{P}(\sigma > x)\sim c\,x^{-4}$ as $x\to\infty$. We see that the extreme values of (σ_t) follow the same patterns as the extremes of (X_t). Compare with Figure 10; in the latter case the innovation dominated the volatility.

5. Point processes and their weak convergence Resnick (1987,2007)

5.1. Preliminaries.

A point process is a random counting measure on some state space E:[a]

- **For random vectors $\xi_i \in E$,**

$$N(A) = \sum_{i=1}^{M} \varepsilon_{\xi_i}(A) = \#\{i \leq M : \xi_i \in A\}, \qquad A \subset E,$$

for M finite or infinite, ε_x Dirac measure at x.

- **$N(A)$ is finite for compact subsets A of E.**

[a] We typically assume $E \subset \mathbb{R}^d$.

5.2. Binomial and Poisson processes.

Binomial point process

$$N = \sum_{i=1}^{M} \varepsilon_{\xi_i}, \qquad (\xi_i) \text{ iid}, \quad M \text{ finite}$$

Then

$$N(A) = \sum_{i=1}^{M} 1(\xi_i \in A) \sim \mathrm{Bin}(M, \mathbb{P}(\xi \in A))$$

Poisson process or Poisson random measure with mean measure μ on $(E, \mathcal{E})$, PRM(μ),[a]

- $N(A) \sim \text{Poisson}(\mu(\text{A}))$ **for** $A \in \mathcal{E}$.
- **If** $A_1, \ldots, A_m \in \mathcal{E}$ **are disjoint then** $N(A_1), \ldots, N(A_m)$ **are independent.**

[a] μ is Radon on E.

- **Homogeneous Poisson process on $E \subset \mathbb{R}^d$:**

  $\mu = \lambda\,\mathrm{Leb}$ on E, λ is the **intensity** On $E = \mathbb{R}_+$, the HPP has representation

$$N = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i}\,, \quad \Gamma_i = E_1 + \cdots + E_i\,,\ (E_k)\ \text{iid Exp}(\lambda)$$

- Assume $\lambda = 1$. The process

$$N_\alpha = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-1/\alpha}}$$

  is $\mathrm{PRM}(\mu_\alpha)$ on $\mathbb{R}_+$ with $\mu_\alpha(x,\infty) = x^{-\alpha}$, $x > 0$, since

$$\begin{aligned} N_\alpha(x\,,\infty) &= \#\{i \ge 1 : \Gamma_i^{-1/\alpha} > x\} \\ &= \#\{i \ge 1 : \Gamma_i \le x^{-\alpha}\} = N(0, x^{-\alpha}]\,, \end{aligned}$$

  and $N(0, x^{-\alpha}] \sim \mathrm{Poisson}(x^{-\alpha})$.

  Note: $\mathbb{P}(\Gamma_1^{-1/\alpha} \le x) = \mathrm{e}^{-x^{-\alpha}} = \Phi_\alpha(x)$.

5.3. Operations acting on Poisson processes.

Transformation of the points of a PRM

If $N = \sum_{i=1}^{\infty} \varepsilon_{\xi_i}$ is $\mathrm{PRM}(\mu)$ on $E \subset \mathbb{R}^d$ and $f : E \to E' \subset \mathbb{R}^{d'}$ is measurable and such that $\mu_f(A) = \mu(f^{-1}(A))$ is finite for compact $A \subset E'$ then

$$N_f = \sum_{i=1}^{\infty} \varepsilon_{f(\xi_i)} \sim \mathrm{PRM}(\mu_f) \quad \text{on } E'.$$

Marking of a PRM

If $N = \sum_{i=1}^{\infty} \varepsilon_{\xi_i}$ is $\mathrm{PRM}(\mu)$ on $E \subset \mathbb{R}^d$ and (V_i) is an iid $\mathbb{S}$-valued sequence independent of (ξ_i) then

$$N_{\xi,V} = \sum_{i=1}^{\infty} \varepsilon_{\xi_i, V_i} \sim \mathrm{PRM}(\mu \times F_V) \text{ on } E \times \mathbb{S}.$$

5.4. Max-stable process with Fréchet marginals.

- **Mark PRM N_α: for iid $V_i > 0$,**

$$N = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-1/\alpha}, V_i} \sim \mathrm{PRM}(\mu_\alpha \times F_V)$$

- **For $y > 0$ define $A_y = \{(x, v) \in \mathbb{R}_+^2 : x\, v > y\}$.**
- **If $\mathbb{E}[V^\alpha] < \infty$ then**

$$\begin{aligned}(\mu_a \otimes F_V)(A_y) &= \int_{v=0}^{\infty} \Big(\int_{x=y/v}^{\infty} \alpha x^{-\alpha-1}\, dz \Big) F_V(dv) \\ &= \int_{v=0}^{\infty} (y/v)^{-\alpha}\, F_V(dv) = y^{-\alpha}\, \mathbb{E}[V^\alpha]\,.\end{aligned}$$

- **But**

$$
\begin{aligned}
N(A_y) &= \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-1/\alpha}, V_i}(A_y) \\
&\sim \text{Poisson}((\mu_\alpha \times F_V)(A_y)) \\
&= \text{Poisson}(y^{-\alpha}\, \mathbb{E}[V^\alpha])\,.
\end{aligned}
$$

- **Hence**

$$
\begin{aligned}
\mathbb{P}\Big(\sup_{i\ge 1} \Gamma_i^{-1/\alpha} V_i \le y\Big) &= \mathbb{P}(N(A_y) = 0) \\
&= \exp\big(-\mathbb{E}[N(A_y)]\big) \\
&= \exp\big(-y^{-\alpha}\, \mathbb{E}[V^\alpha]\big) = \Phi_\alpha^{\mathbb{E}[V^\alpha]}(y)\,.
\end{aligned}
$$

Max-stable process de Haan (1984)

Let (V_i) be iid positive stochastic processes on $T=\mathbb{Z}$ or $T\subset\mathbb{R}^s$ and $\mathbb{E}[V^\alpha(t)]<\infty$ for all $t\in T$. The process

$$\xi(t)=\sup_{i\ge 1}\Gamma_i^{-1/\alpha}\,V_i(t)\,,\qquad t\in T$$

is max-stable with Fréchet marginals:

$$\mathbb{P}(\xi(t)\le y)=\Phi_\alpha^{\mathbb{E}[V^\alpha(t)]}(y)\,,\qquad y>0\,.$$

- **Brown-Resnick process**:

$$V_i(t)=\exp(W_i(t)-0.5\,t)\,,\qquad t\ge 0\,,$$

where (W_i) are iid standard Brownian motions on $T=\mathbb{R}$. The BR process is a stationary process; see Kabluchko (2009), Kabluchko, Schlather, de Haan (2009), Stoev (2008), Stoev, Taqqu (2005).

5.5. Weak convergence of binomial processes to a Poisson process.

(N_n) point processes on E converge in distribution to a point process N on E ($N_n \stackrel{d}{\to} N$) if $\forall A_i \in \mathcal{E}$ with $N(\partial A_i) = 0$ a.s. and $m \geq 1$,

$$(N_n(A_1), \ldots, N_n(A_m)) \stackrel{d}{\to} (N(A_1), \ldots, N(A_m))$$

- For each n, let $(X_{ni})_{i=1,2,\ldots}$ be an iid sequence. Then $N_n = \sum_{i=1}^{n} \varepsilon_{X_{ni}}$ is binomial.

Binomial processes converge weakly to Poisson process

$$N_n \stackrel{d}{\to} N \sim \mathrm{PRM}(\mu)$$

if and only if for any μ-continuity set $A \subset E$,[a] Resnick (2007)

$$\mathbb{E}[N_n(A)] = n\,\mathbb{P}(X_{n1} \in A) =: \mu_n(A) \to \mu(A) = \mathbb{E}[N(A)]\,. \tag{5.1}$$

[a] See also p. 109.

5.6. Vague convergence of measures Resnick (1987, 2007).

A limit relation of the type (5.1) is called

vague convergence of (μ_n) to μ: $\mu_n \xrightarrow{v} \mu$.

- **Here it is assumed that μ_n, μ are finite on compact sets $A \subset E$: Radon measures.**
- **Vague convergence $\mu_n \xrightarrow{v} \mu$ can often be verified on particular subsets of E, for example on the μ-continuous rectangles $(\mathrm{a}, \mathrm{b}] \subset E$.**
- **Example: Weak convergence of maxima in MDA(Φ_α).**
  - **Assume (X_i) iid, $F \in \text{MDA}(\Phi_\alpha)$: $\overline{F}(x) = L(x)x^{-\alpha}$, $x > 0$.**
  - **Choose (a_n) such that $n\mathbb{P}(X > a_n) \to 1$.**

– Regular variation and the definition of (a_n) imply for any $(c,d] \subset \mathbb{R}_+$:

$$\begin{aligned}\mu_n(c,d] &:= n\,\mathbb{P}(a_n^{-1}X \in (c,d]) \\ &\sim \frac{\mathbb{P}(X > a_n\,c)}{\mathbb{P}(X > a_n)} - \frac{\mathbb{P}(X > a_n\,d)}{\mathbb{P}(X > a_n)} \\ &\to c^{-\alpha} - d^{-\alpha} = \mu_\alpha(c,d]\,.\end{aligned}$$

Note: Every interval $(c,d]$ is a μ_α-continuity set.

– The relations $\mu_n(c,d] \to \mu_\alpha(c,d]$ for $0 < c < d$ imply $\mu_n \stackrel{v}{\to} \mu_\alpha$ on $E = \mathbb{R}_+$.

– Hence, for $X_{ni} = a_n^{-1}X_i$,

$$N_n = \sum_{i=1}^n \varepsilon_{a_n^{-1}X_i} \stackrel{d}{\to} N_\alpha = \sum_{i=1}^\infty \varepsilon_{\Gamma_i^{-1/\alpha}} \sim \mathrm{PRM}(\mu_\alpha)\,.$$

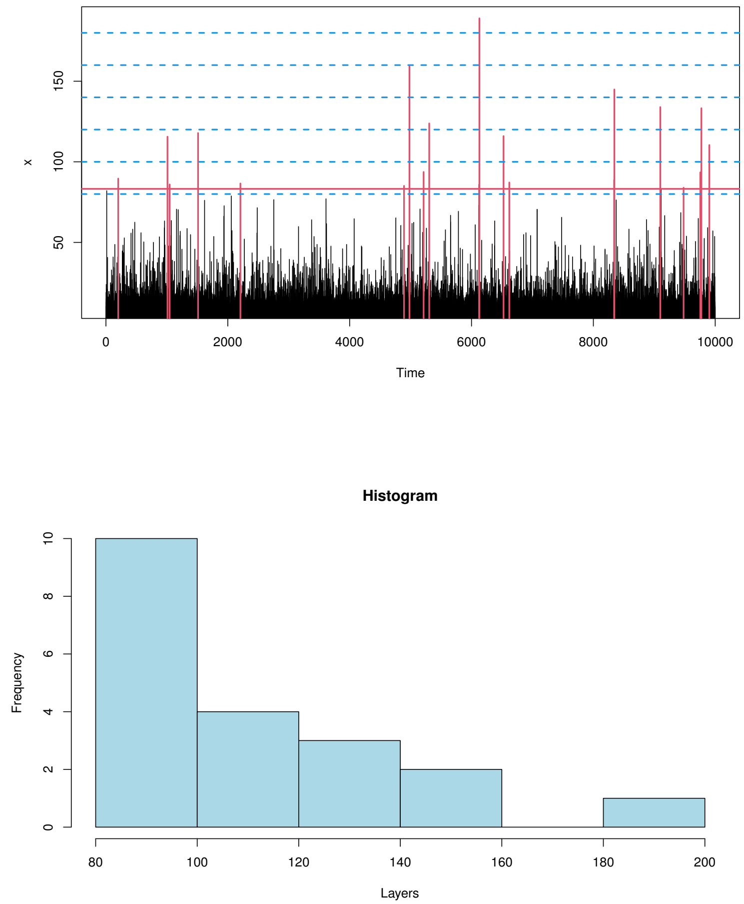


FIGURE 12. **Left:** The exceedances larger than the 20th order statistic of an iid sample with Pareto(3) distribution, i.e., $\mathbb{P}(X > x) = (x/10)^{-3}$, $x > 10$. This order statistic corresponds to the red line. The stippled blue lines define the layers corresponding to $x_1 = 80$, $x_2 = 100$, $\ldots x_7 = 200$. **Right:** Counts of the point process $\widetilde{N}_n$ in the distinct layers $(x_1, x_2]$, $\ldots$, $(x_6, x_7]$.

– **Joint convergence of order statistics**

Let $X_{(n)} \le \cdots \le X_{(1)}$ be the order statistics of $X_1, \ldots, X_n$.

Then for $x_1 > x_2 > \ldots > x_k$,

$$\mathbb{P}\big(a_n^{-1} X_{(1)} \le x_1 \,, \ldots \,, a_n^{-1} X_{(k)} \le x_k\big)$$

$$= \mathbb{P}\big(N_n(x_1, \infty) = 0 \,, N_n(x_2, \infty) \le 1 \,, \ldots , N_n(x_k, \infty) \le k - 1\big)$$

$$\to \mathbb{P}\big(N(x_1, \infty) = 0 \,, N(x_2, \infty) \le 1 \,, \ldots , N(x_k, \infty) \le k - 1\big)$$

$$= \mathbb{P}\big(\Gamma_1^{-1/\alpha} \le x_1, \ldots , \Gamma_k^{-1/\alpha} \le x_k\big)$$

– **This means**

$$a_n^{-1}\big(X_{(1)}, \ldots , X_{(k)}\big) \stackrel{d}{\to} \big(\Gamma_1^{-1/\alpha}, \ldots , \Gamma_k^{-1/\alpha}\big) \,.$$

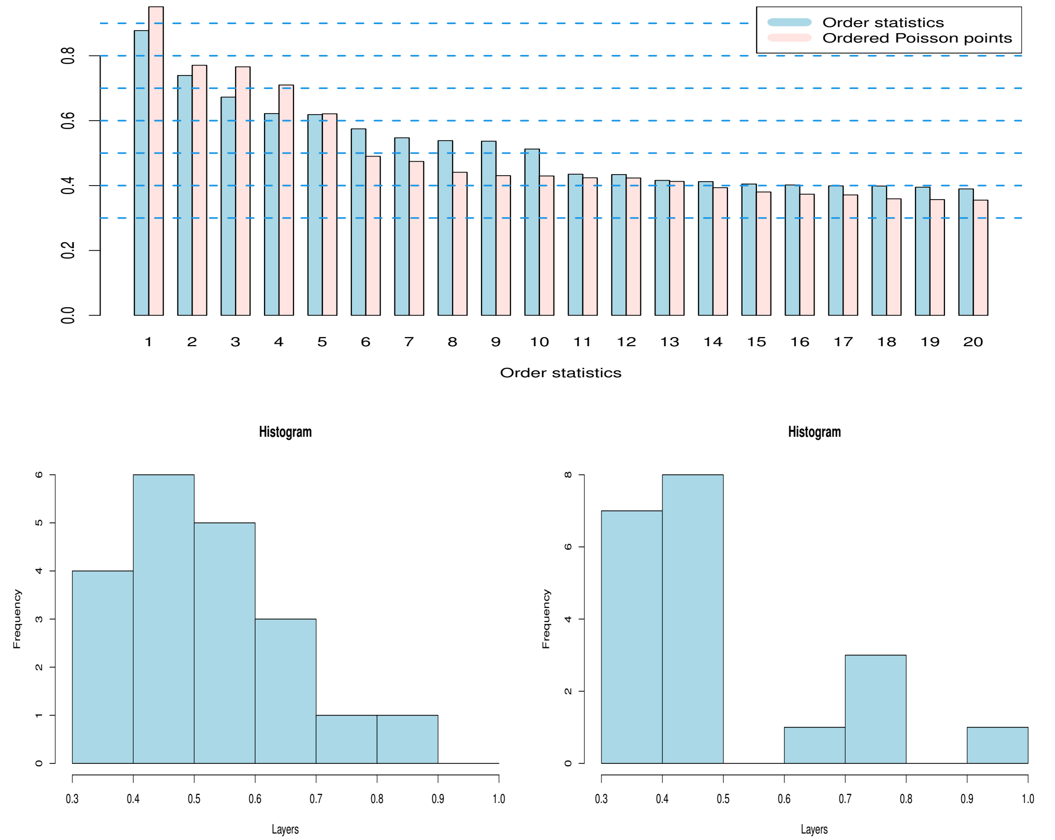


FIGURE 13. **Top:** The 20 largest order statistics of an iid rescaled sample X_i/a_n of size $n = 1000$ with Pareto(3) distribution, i.e., $\mathbb{P}(X > x) = (x/10)^{-3}$, $x > 10$, and a_n defined by $\mathbb{P}(X > a_n) = 1/1000$ (blue bars). The 20 largest points $\Gamma_i^{-1/3}$ of the limit point process (pink bars). The stippled blue lines define the layers corresponding to $x_1 = 0.3$, $x_2 = 0.4$, ..., $x_7 = 1$. **Bottom left:** Counts of the point process N_n in the distinct layers $(x_1, x_2]$, ..., $(x_6, x_7]$. **Bottom right:** Counts of the point process N_{Φ_α} in the distinct layers $(x_1, x_2]$, ..., $(x_6, x_7]$. Compare also with the unscaled point process $\widetilde{N}_n$ in Figure 12.

6. Convergence of component-wise maxima

- **Consider an iid sequence $(\mathbf{X}_i)$ of $\mathbb{R}_+^d$-valued random vectors.**
- **Assume for the moment that the components of $\mathbf{X}$ have identical distribution and $\mathbb{P}(X^{(1)} > x) = x^{-\alpha} L(x)$ for some $\alpha > 0$ and a slowly varying function L.**
- **Choose (a_n) such that $n\,\mathbb{P}(X_1 > a_n) \to 1$.**
- **For a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ define the componentwise maxima:**

$$\mathbf{M}_n = \Big(\max_{i=1,\ldots,n} X_i^{(j)} \Big)_{j=1,\ldots,d}\,.$$

- **In particular, for each $j = 1, \ldots, d$,**

$$\mathbb{P}\Big(a_n^{-1} \max_{i=1,\ldots,n} X_i^{(j)} \le x \Big) \to \Phi_\alpha(x)\,, \qquad x \in \mathbb{R}\,.$$

- **When do the components of $\mathbf{M}_n$ converge jointly?**

- For $\mathbf{x} = (x_1, \ldots, x_d) \geq \mathbf{0}$ the distribution function of $\mathbf{M}_n$ is

$$
\begin{aligned}
\mathbb{P}(a_n^{-1}\mathbf{M}_n \leq \mathbf{x}) &= \Big(\bigcap_{j=1}^{d} \Big\{ a_n^{-1} \max_{i=1,\ldots,n} X_i^{(j)} \leq x_j \Big\} \Big) \\
&= \mathbb{P}\Big(\bigcap_{i=1}^{n} \bigcap_{j=1}^{d} \{ a_n^{-1} X_i^{(j)} \leq x_j \} \Big) \\
&= \mathbb{P}\Big(\bigcap_{i=1}^{n} \{ a_n^{-1} \mathbf{X}_i \in [\mathbf{0}, \mathbf{x}] \} \Big) \\
&= \big[\mathbb{P}(a_n^{-1}\mathbf{X} \in [\mathbf{0}, \mathbf{x}]) \big]^n \\
&= \Big[1 - \frac{n\, \mathbb{P}(a_n^{-1}\mathbf{X} \in [\mathbf{0}, \mathbf{x}]^c)}{n} \Big]^n .
\end{aligned}
$$

- The right-hand side converges to a non-degenerate distribution function $H(\mathbf{x})$ for all but countably many $\mathbf{x}$ if and only if

$$
\mu_n([\mathbf{0}, \mathbf{x}]^c) = n\, \mathbb{P}(a_n^{-1}\mathbf{X} \in [\mathbf{0}, \mathbf{x}]^c) \to \mu([\mathbf{0}, \mathbf{x}]^c) \tag{6.1}
$$

- $\mu([0,\mathbf{x}]^c)$ **is non-zero for some** $\mathbf{x} \geq \mathbf{0}$ **and**

$$H(\mathbf{x}) = \exp\big(-\mu([0,\mathbf{x}]^c)\big)\,, \qquad \mathbf{x} \geq \mathbf{0}$$

has Fréchet Φ_α-marginals:

H **is a multivariate Fréchet distribution.**

$\mu([0,\mathbf{x}]^c)$, $\mathbf{x} \geq \mathbf{0}$, **can be extended to a Radon measure** μ **on** $\mathbb{R}^d_{+,0} = \mathbb{R}^d_+ \backslash \{\mathbf{0}\}$**:**

the exponent or tail measure of *H*.

(6.1) can be extended to any μ-continuity set $A \subset \mathbb{R}^d_{+,0}$ bounded away from zero:

$$\mu_n(A) = n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in A) \to \mu(A) \Longleftrightarrow \mu_n \stackrel{v}{\to} \mu\,.$$

- We observe for any $t > 0$ and μ-continuity set A,[3]

$$\mu_n(t\,A) \;\sim\; t^{-\alpha}\left[nt^{\alpha}\,\mathbb{P}\big(a_{[t^{\alpha}n]}^{-1}\mathbf{X} \in A\big)\right]$$

$$\to t^{-\alpha}\,\mu(A)\,.$$

On the other hand, $\mu_n(t\,A) \to \mu(tA)$.

The exponent measure μ is homogeneous:

for any Borel set $A \subset \mathbb{R}^d_{+,0}$:

$$\mu(t\,A) = t^{-\alpha}\,\mu(A)\,, \qquad t > 0\,.$$

---

[3] (a_n) is a regularly varying sequence: $a_n = n^{1/\alpha}L(n)$.

7. Multivariate regular variation

7.1. Definition and equivalences.

The $\mathbb{R}^d$-valued random vector $\mathbf{X}$ and its distribution are

regularly varying

if $|\mathbf{X}|$ is regularly varying and there exists a non-null Radon measure μ on $\mathbb{R}_0^d = \mathbb{R}^d \backslash \{0\}$ such that

$$\frac{\mathbb{P}(x^{-1}\mathbf{X} \in \cdot)}{\mathbb{P}(|\mathbf{X}| > x)} \stackrel{v}{\to} \mu(\cdot)\,, \qquad x \to \infty\,.$$

- **Equivalently, for any sequence (a_n) such that**

$n\,\mathbb{P}(|\mathbf{X}| > a_n) \to 1$,

$$n\,\mathbb{P}(a_n^{-1}\mathbf{X} \in \cdot) \stackrel{v}{\to} \mu(\cdot) \qquad (7.1)$$

- μ **is homogeneous:** since for any small $t > 0$

$$\frac{\mathbb{P}(|\mathbf{X}| > t\,x)}{\mathbb{P}(|\mathbf{X}| > x)} \to \mu(\{\mathbf{x} : |\mathbf{x}| > t\}) > 0\,, \qquad x \to \infty\,,$$

regular variation calculus Bingham, Goldie, Teugels (1987) **yields that the limit is proportionnal to** $t^{-\alpha}$ **for some** $\alpha \geq 0$.

- **Therefore as** $x \to \infty$

$$\frac{\mathbb{P}(x^{-1}\mathbf{X} \in t\,A)}{\mathbb{P}(|\mathbf{X}| > x)} \to \mu(t\,A)\,,$$

$$\frac{\mathbb{P}(x^{-1}\mathbf{X} \in t\,A)}{\mathbb{P}(|\mathbf{X}| > x)} = \frac{\mathbb{P}((tx)^{-1}\mathbf{X} \in A)}{\mathbb{P}(|\mathbf{X}| > t\,x)} \frac{\mathbb{P}(|\mathbf{X}| > t\,x)}{\mathbb{P}(|\mathbf{X}| > x)} \to \mu(A)\,t^{-\alpha}$$

Homogeneity of μ: $\mu(t\,\cdot) = t^{-\alpha}\,\mu(\cdot)\,,\ t > 0$, **for some $\alpha \geq 0$.**

We write: $\mathbf{X} \in \mathrm{RV}(\alpha, \mu)$.

- **Regular variation in spherical coordinates:** for any fixed norm $|\cdot|$ and $t > 0$,

$$\frac{\mathbb{P}\Big(|\mathbf{X}| > t\,x\,,\frac{\mathbf{X}}{|\mathbf{X}|} \in \cdot\Big)}{\mathbb{P}(|\mathbf{X}| > x)} \xrightarrow{w} \mu\Big(\{\mathbf{x} : |\mathbf{x}| > t\,,\frac{\mathbf{x}}{|\mathbf{x}|} \in \cdot\}\Big)$$

$$= t^{-\alpha}\,\mu\Big(\{\mathbf{x} : |\mathbf{x}| > 1\,,\frac{\mathbf{x}}{|\mathbf{x}|} \in \cdot\}\Big)$$

$$=: t^{-\alpha}\mathbb{P}(\boldsymbol{\Theta} \in \cdot)\,.$$

$\mathbb{P}(\boldsymbol{\Theta} \in \cdot)$ **is the spectral distribution/measure or angular measure of $\mathbf{X}$ on the unit sphere $\mathbb{S}^{d-1} = \mathbb{S}^{d-1}_{|\cdot|}$.**

Assume $\alpha > 0$. Then $\mathrm{X} \in \mathrm{RV}(\alpha, \mu)$ if and only if as $x \to \infty$

$$\mathbb{P}\Big(\Big(\frac{|\mathrm{X}|}{x}, \frac{\mathrm{X}}{|\mathrm{X}|}\Big) \in \cdot \,\Big|\, |\mathrm{X}| > x\Big) \xrightarrow{w} \mathbb{P}((Y, \Theta) \in \cdot)$$

for independent Y, Θ, Pareto(α)-distributed Y: $\mathbb{P}(Y > y) = y^{-\alpha}$.

From (7.1) on p. 49 and (5.1) on p. 39 we have:

Assume $\alpha > 0$ and $n\,\mathbb{P}(|\mathrm{X}| > a_n) \to 1$.

Then $\mathrm{X} \in \mathrm{RV}(\alpha, \mu)$ if and only if

$$N_n = \sum_{i=1}^{n} \varepsilon_{a_n^{-1}\mathrm{X}_i} \xrightarrow{d} N \sim \mathrm{PRM}(\mu)\,, \qquad n \to \infty\,.$$

7.2. Operations on regularly varying random vectors.

7.2.1. *Convolution.* Assume

- $\mathbf{X} \in \mathrm{RV}(\alpha, \mu_{\mathbf{X}})$, $\alpha > 0$,
- there is $c_0 \geq 0$ such that
$$\lim_{x\to\infty} \frac{\mathbb{P}(|\mathbf{Y}| > x)}{\mathbb{P}(|\mathbf{X}| > x)} = c_0\,,$$
- if also $c_0 > 0$, $\mathbf{Y} \in \mathrm{RV}(\alpha, \mu_{\mathbf{Y}})$ and $\mathbf{X}, \mathbf{Y}$ are independent.

Then

$$\frac{\mathbb{P}(\mathbf{X}+\mathbf{Y} \in \cdot)}{\mathbb{P}(|\mathbf{X}+\mathbf{Y}| > x)} \xrightarrow{v} \frac{1}{1+c_0}\mu_{\mathbf{X}}(\cdot) + \frac{c_0}{1+c_0}\mu_{\mathbf{Y}}(\cdot)\,,$$

and for $\mu_{\mathbf{X}+\mathbf{Y}}$-continuity sets A,

Multivariate Feller's lemma

$$\mathbb{P}(x^{-1}(\mathbf{X}+\mathbf{Y}) \in A) \sim \mathbb{P}(x^{-1}\mathbf{X} \in A) + \mathbb{P}(x^{-1}\mathbf{Y} \in A)\,, \qquad x \to \infty\,.$$

7.2.2. *Multiplication.* **Assume**

- $\mathrm{X} \in \mathrm{RV}(\alpha, \mu_{\mathrm{X}})$ **for some** $\alpha > 0$,
- **the** $d' \times d$ **random matrix** A **and** $\mathrm{X} \in \mathbb{R}^d$ **are independent,**
- $\mathbb{E}[\|\mathrm{A}\|^{\alpha+\epsilon}] < \infty$ **for some** $\epsilon > 0$.

Then

Multivariate Breiman's lemma Basrak, Davis, Mikosch (2002)

$$\frac{\mathbb{P}(x^{-1}\,\mathrm{A}\,\mathrm{X} \in \cdot)}{\mathbb{P}(|\mathrm{X}| > x)} \xrightarrow{v} \mathbb{E}[\mu_{\mathrm{X}}\{\mathrm{x} : \mathrm{A}\,\mathrm{x} \in \cdot\}] = \mathbb{E}[\mu_{\mathrm{X}}(\mathrm{A}^{-1}\cdot)]\,, \qquad x \to \infty\,.$$

- **If** $\mathbb{E}[\mu_{\mathrm{X}}(\mathrm{A}^{-1}\cdot)]$ **is non-null then** $\mathrm{AX} \in \mathrm{RV}(\alpha, \mu_{\mathrm{AX}})$ **where**

$$\mu_{\mathrm{AX}}(\cdot) = \frac{\mathbb{E}[\mu_{\mathrm{X}}(\mathrm{A}^{-1}\cdot)]}{\mathbb{E}[\mu_{\mathrm{X}}(\{\mathrm{x} : |\mathrm{A}\,\mathrm{x}| > 1\})]}\,.$$

- **Example: Regularly varying AR(1) process**
- **Assume** $X_t = \varphi X_{t-1} + Z_t$, $t \in \mathbb{Z}$, (Z_t) **iid**, $Z \in \mathrm{RV}(\alpha)$, **and** $|\varphi| < 1$.
- **Then for** $h \geq 0$,

$$
\begin{aligned}
\mathbf{X}_h = (X_0, \ldots, X_h) &= X_0 \,(1, \varphi, \ldots, \varphi^h) \\
&\quad + (0, Z_1, Z_2 + \varphi Z_1, \ldots, Z_h + \varphi Z_{h-1} + \varphi^{h-1} Z_1) \\
&= X_0 \begin{pmatrix} 1 \\ \varphi \\ \varphi^2 \\ \vdots \\ \varphi^h \end{pmatrix} + Z_1 \begin{pmatrix} 0 \\ 1 \\ \varphi \\ \vdots \\ \varphi^{h-1} \end{pmatrix} + \cdots + Z_h \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}
\end{aligned}
$$

- **Feller's lemma:**

$$\mathbb{P}(x^{-1}\mathbf{X}_h \in \cdot) \sim \mathbb{P}(x^{-1}X_0\,(1,\varphi,\ldots,\varphi^h) \in \cdot) + \sum_{i=1}^{h} \mathbb{P}(x^{-1}Z(0,\ldots,0,1,\varphi,\ldots\,\varphi^{h-i}) \in \cdot)$$

- **Breiman's lemma:** $\dfrac{\mathbb{P}(\pm X_0 > x)}{\mathbb{P}(|Z| > x)} \to \widetilde{p}_\pm, \; \dfrac{\mathbb{P}(\pm Z > x)}{\mathbb{P}(|Z| > x)} \to p_\pm$

$$\mu_Z(dx) = \big(p_+ 1(x > 0) + p_- 1(x < 0)\big)\alpha|x|^{-\alpha-1}\,dx\,,$$

$$\mu_X(dx) = \big(\widetilde{p}_+ 1(x > 0) + \widetilde{p}_- 1(x < 0)\big)\alpha|x|^{-\alpha-1}\,dx\,,$$

$$\frac{\mathbb{P}(x^{-1}\mathbf{X} \in \cdot)}{\mathbb{P}(|Z| > x)} \to \mu_X\big(\{y \in \mathbb{R} : y(1,\varphi,\ldots,\varphi^h) \in \cdot\}\big) + \sum_{i=1}^{h} \mu_Z\big(\{y \in \mathbb{R} : y\,(0,\ldots,0,1,\varphi,\ldots\,\varphi^{h-i}) \in \cdot\}\big)$$

- **The case $h = 1$.**

$$\mathbb{P}\Big(\boldsymbol{\Theta} = \frac{(\mathbf{1}, \boldsymbol{\varphi})}{|(\mathbf{1}, \boldsymbol{\varphi})|}\Big) = \frac{\widetilde{\boldsymbol{p}}_+ |(\mathbf{1}, \boldsymbol{\varphi})|^{\boldsymbol{\alpha}}}{\mathbf{1} + |(\mathbf{1}, \boldsymbol{\varphi})|^{\boldsymbol{\alpha}}}$$

$$\mathbb{P}\Big(\boldsymbol{\Theta} = \frac{(-\mathbf{1}, -\boldsymbol{\varphi})}{|(\mathbf{1}, \boldsymbol{\varphi})|}\Big) = \frac{\widetilde{\boldsymbol{p}}_- |(\mathbf{1}, \boldsymbol{\varphi})|^{\boldsymbol{\alpha}}}{\mathbf{1} + |(\mathbf{1}, \boldsymbol{\varphi})|^{\boldsymbol{\alpha}}}$$

$$\mathbb{P}(\boldsymbol{\Theta} = (\mathbf{0}, +\mathbf{1})) = \frac{\boldsymbol{p}_+}{\mathbf{1} + |(\mathbf{1}, \boldsymbol{\varphi})|^{\boldsymbol{\alpha}}}$$

$$\mathbb{P}(\boldsymbol{\Theta} = (\mathbf{0}, -\mathbf{1})) = \frac{\boldsymbol{p}_-}{\mathbf{1} + |(\mathbf{1}, \boldsymbol{\varphi})|^{\boldsymbol{\alpha}}}$$

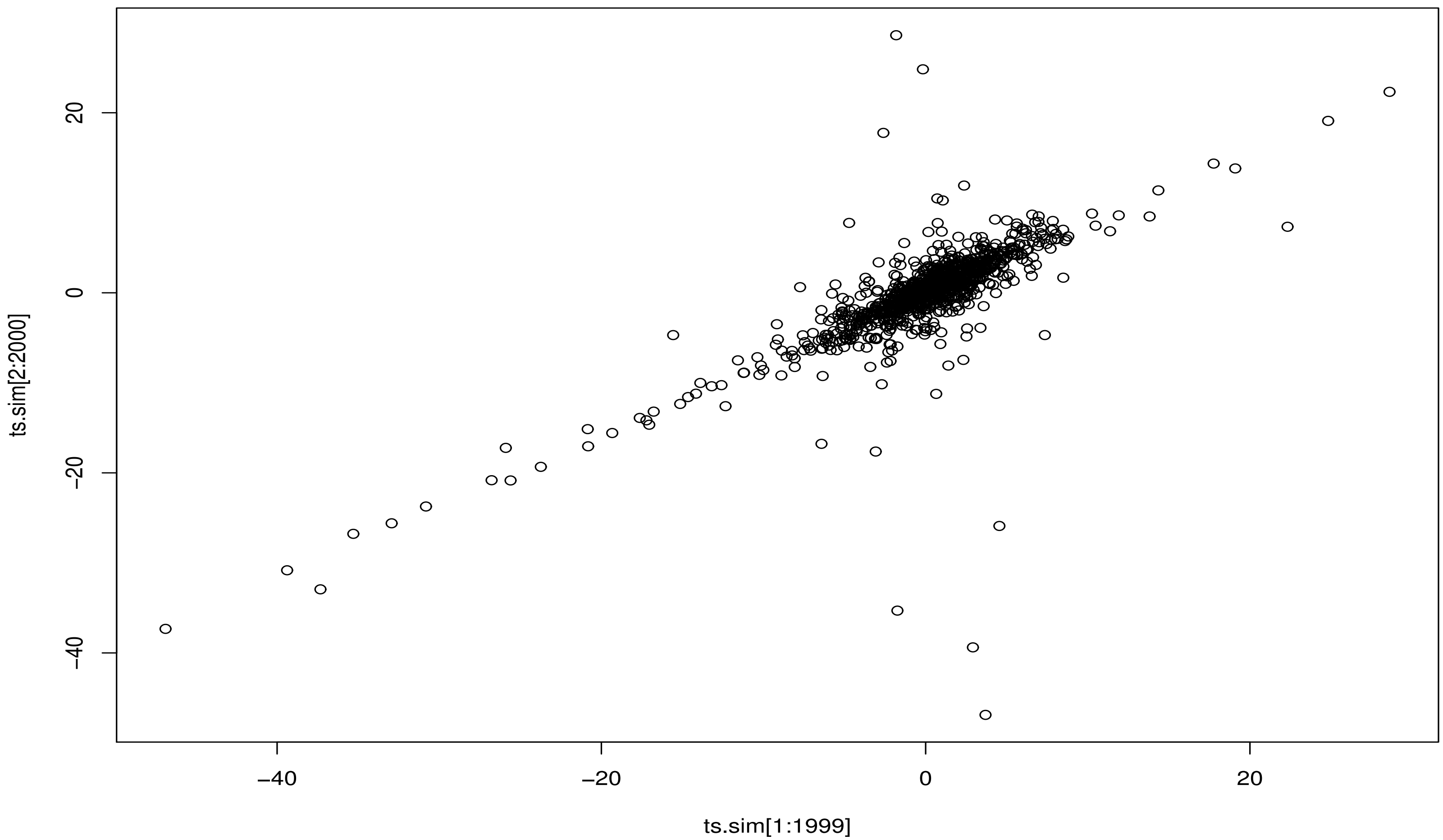


FIGURE 14. Scatterplot of AR(1) process $X_t = 0.8X_{t-1} + Z_t$ with iid student(2) (Z_t). Compare with Figure 2.2.

- Example: The stochastic volatility model
- Assume
  - $X_t = \sigma_t Z_t$, $t \in \mathbb{Z}$, is iid with $Z \in \mathrm{RV}(\alpha)$.
  - (σ_t) is positive stationary and $\mathbb{E}[\sigma^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$.
  - (σ_t) and (Z_t) are independent.
- We already know by Breiman: $\mathbb{P}(\pm X > x) \sim \mathbb{E}[\sigma^\alpha]\,\mathbb{P}(\pm Z > x)$, i.e. $X \in \mathrm{RV}(\alpha)$.
- We have by Markov's inequality

$$\begin{aligned}
\mathbb{P}(X_0 > \delta x\,, X_h > \delta x) &= \mathbb{P}(X_0 \wedge X_h > \delta\, x) \\
&\le \mathbb{P}\big((\sigma_0 \vee \sigma_h)\,(Z_0 \wedge Z_h) > \delta x\big) \\
&\le \mathbb{E}[(\sigma_0 \vee \sigma_h)^{\alpha+\epsilon}]\,\mathbb{E}[(Z_0 \wedge Z_h)^{\alpha+\epsilon}]\,(\delta\, x)^{-(\alpha+\epsilon)}
\end{aligned}$$

- **Therefore, choosing the Euclidean norm $|\cdot|_2$ and**

$$\mathbf{X}_h = (X_0, \ldots, X_{h+1}),$$

$$\frac{\mathbb{P}(\min(|X_0|, \ldots, |X_h|) > \delta\, x)}{\mathbb{P}(|\mathbf{X}_h|_2 > x)} \le \frac{\mathbb{P}(\min(|X_0|, \ldots, |X_h|) > \delta\, x)}{\mathbb{P}(|X_0| > x)} \to 0\,, \qquad x \to \infty\,.$$

If $\mathbf{X}_h \in \mathrm{RV}(\alpha, \mu_h)$ we would have $\mu_h((\delta, \infty)^{h+1}) = 0$ for all $\delta > 0$, and μ_h would be concentrated on the axes only.

- **For any Borel sets $A_0, \ldots, A_h$ such that $B = A_1 \times \cdots \times A_h$ is bounded away from zero,**

$$\mu_h(B) = \sum_{k=0}^{h} \prod_{i=0}^{k-1} \varepsilon_0(A_i)\, \mu_\alpha(A_k) \prod_{j=k+1}^{h} \varepsilon_0(A_j)\,,$$

where $\mu_\alpha(dx) = (p_+ 1(x > 0) + p_- 1(x < 0))\, \alpha |x|^{-\alpha-1}$.

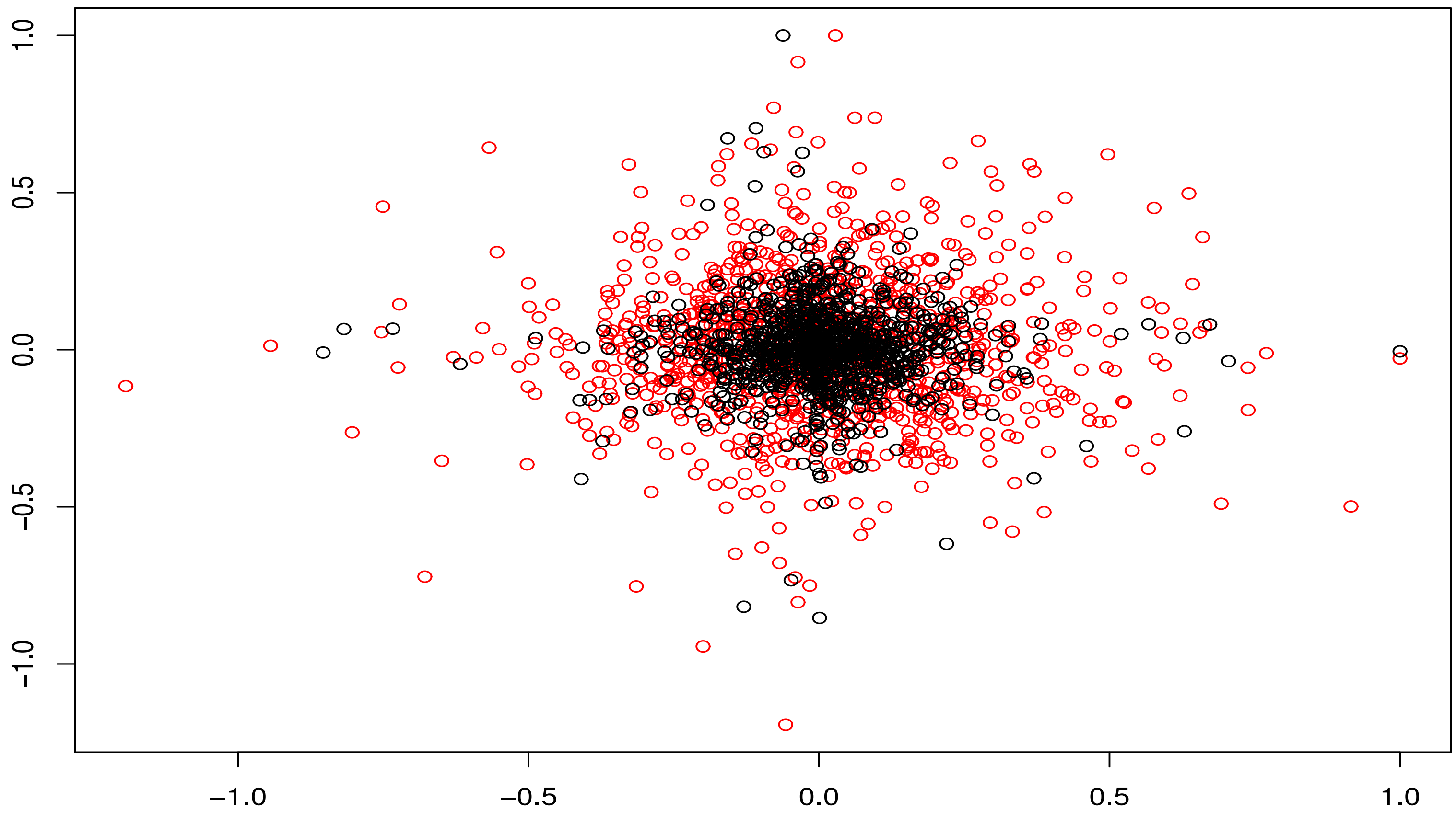


FIGURE 15. Scatterplot of stochastic volatility processes $\boldsymbol{X}_t = \boldsymbol{\sigma}_t \boldsymbol{Z}_t$ with innovations dominating (black) and volatility dominating (red).

If $\mathbf{X} \in \mathrm{RV}(\alpha, \mu_{\mathbf{X}})$ and $\mu_{\mathbf{X}}$ is supported on the axes then we say that components of $\mathbf{X}$ are asymptotically independent.

8. Regularly varying time series

8.1. Definition and preliminaries. .

An $\mathbb{R}^d$-valued stationary time series is regularly varying with index $\alpha > 0$ if for any $h \geq 0$, $\mathbf{Y}_h = (\mathbf{X}_0, \ldots, \mathbf{X}_h) \in \mathrm{RV}(\alpha, \mu_h)$:

$$\frac{\mathbb{P}(x^{-1}\mathbf{Y}_h \in \cdot)}{\mathbb{P}(|\mathbf{X}| > x)} \xrightarrow{v} \mu_h(\cdot)\,, \qquad x \to \infty\,.$$

- **We note that $\mathbb{P}(|\mathbf{Y}_h| > x)/\mathbb{P}(|\mathbf{X}| > x) \to c_h > 0$.**
- **The normalization with $\mathbb{P}(|\mathbf{X}| > x)$ instead of $\mathbb{P}(|\mathbf{Y}_h| > x)$ ensures that the family (μ_h) is consistent:**

$$\mu_{h+1}(\mathbb{R}^d \times \cdot) = \mu_{h+1}(\cdot \times \mathbb{R}^d) = \mu_h(\cdot)\,.$$

- **General regularly varying time series were studied first in** Davis, Hsing (1995)**.**
- **Before 1995, EVT for regularly varying linear processes was considered by** Rootzén (1978,1986); Davis, Resnick (1985,1986); **and for solutions to stochastic recurrence equation by** de Haan, Resnick, Rootzén, de Vries (1989).
- **In recent years there has been proposed two alternative definitions of stationary regularly varying time series, the one based on the tail process** Basrak, Segers (2009) **and the one on the tail measure** Kulik, Soulier (2020)**.**

8.2. **The tail process** Basrak, Segers (2009). **Let Y be Pareto(α) distributed: $\mathbb{P}(Y > y) = y^{-\alpha}$, $y > 1$.**

An $\mathbb{R}^d$-valued stationary sequence (X_t) is regularly varying with index $\alpha > 0$ if and only if one of the following conditions holds:

(1) There exist a Pareto(α) random variable Y and an $\mathbb{R}^d$-valued sequence $(\Theta_t)_{t\geq 0}$ such that Y, $(\Theta_t)_{t\geq 0}$ are independent and

$$\mathbb{P}\big(x^{-1}(X_0,\dots,X_h) \in \cdot \mid |\mathrm{X}_0| > x\big) \xrightarrow{w} \mathbb{P}\big(Y(\Theta_0,\dots,\Theta_h) \in \cdot\big).$$

(2) There exist a Pareto(α) random variable Y and an $\mathbb{R}^d$-valued sequence $(\Theta_t)_{t\leq 0}$ such that Y, $(\Theta_t)_{t\leq 0}$ are independent and

$$\mathbb{P}\big(x^{-1}(X_{-h},\dots,X_0) \in \cdot \mid |\mathrm{X}_0| > x\big) \xrightarrow{w} \mathbb{P}\big(Y(\Theta_{-h},\dots,\Theta_0) \in \cdot\big).$$

(3) There exist a Pareto(α) random variable Y and an $\mathbb{R}^d$-valued sequence $(\boldsymbol{\Theta}_t)_{t\in\mathbb{Z}}$ such that Y, $(\boldsymbol{\Theta}_t)_{t\in\mathbb{Z}}$ are independent and

$$\mathbb{P}\big(x^{-1}(\boldsymbol{X}_{-h},\ldots,\boldsymbol{X}_h)\in\cdot \mid |\mathbf{X}_0|>x\big)\xrightarrow{w}\mathbb{P}\big(Y(\boldsymbol{\Theta}_{-h},\ldots,\boldsymbol{\Theta}_h)\in\cdot\big)\,.$$

The processes $(\boldsymbol{\Theta}_t)_{t\in T}$ with $T=\{0,1,\ldots,\},\{\ldots,-1,0\},\mathbb{Z}$ are the respective

- **forward spectral tail process ,**
- **backward spectral tail process ,**
- **spectral tail process of $(\mathbf{X}_t)$.**

The processes $(Y_t)_{t\in T}=Y(\boldsymbol{\Theta}_t)_{t\in T}$ are the corresponding tail processes of $(\mathbf{X}_t)$.
$\boldsymbol{\Theta}_0\in\mathbb{S}^{d-1}$ has the spectral distribution of $\mathbf{X}_0$.

A proof of sufficiency for $h = 1$ Assume $(\mathbf{X}_t)$ is regularly varying with index $\alpha > 0$. Then

$$\begin{aligned}\mathbb{P}(x^{-1}(\mathbf{X}_0, \mathbf{X}_1) \in A \mid |\mathbf{X}_0| > x) &= \frac{\mathbb{P}(x^{-1}(\mathbf{X}_0, \mathbf{X}_1) \in A\,, |x^{-1}\mathbf{X}_0| > 1)}{\mathbb{P}(|\mathbf{X}_0| > x)} \\ &\to \mu_1\big(\{\mathbf{x} \in A : |x_0| > 1\big) \\ &=: \mathbb{P}((\mathbf{Y}_0, \mathbf{Y}_1) \in A)\,,\end{aligned}$$

and $|\mathbf{Y}_0|$ is Pareto(α) since for $t > 1$,

$$\frac{\mathbb{P}(x^{-1}|\mathbf{X}_0| \in (t, \infty))}{\mathbb{P}(|\mathbf{X}_0| > x)} \to t^{-\alpha} = \mathbb{P}(|\mathbf{Y}_0| > t)\,.$$

Write $Y = |\mathbf{Y}_0|$ and $(\mathbf{Y}_0, Y_1) = Y\,(\boldsymbol{\Theta}_0\,, \boldsymbol{\Theta}_1)$. Then

$$\begin{aligned}\mathbb{P}(Y > y\,, (\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1) \in B) &= \mu_1\Big((\mathbf{x}_0, \mathbf{x}_1) : |\mathbf{x}_0/y| > 1\,, \frac{(\mathbf{x}_0, \mathbf{x}_1)/y}{|\mathbf{x}_0/y|} \in B\Big) \\ &= y^{-\alpha}\,\mu_1\Big((\mathbf{x}_0, \mathbf{x}_1) : |\mathbf{x}_0| > 1\,, \frac{(\mathbf{x}_0, \mathbf{x}_1)}{|\mathbf{x}_0|} \in B\Big) \\ &= \mathbb{P}(Y > y)\,\mathbb{P}\big((\boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1) \in B\big)\end{aligned}$$

- Example: Regularly varying AR(1) process; see p. 55.
- Assume $X_t = \varphi X_{t-1} + Z_t$, $t \in \mathbb{Z}$, (Z_t) iid, $Z \in \mathrm{RV}(\alpha)$, and $|\varphi| < 1$.
- Then for $h \geq 0$,

$$
\begin{aligned}
\mathbf{X}_h = (X_0, \ldots, X_h) &= X_0\,(1, \varphi, \ldots, \varphi^h) \\
&\quad +(0, Z_1, Z_2 + \varphi Z_1, \ldots, Z_h + \varphi Z_{h-1} + \varphi^{h-1} Z_1) \\
&= X_0 \begin{pmatrix} 1 \\ \varphi \\ \varphi^2 \\ \vdots \\ \varphi^h \end{pmatrix} + Z_1 \begin{pmatrix} 0 \\ 1 \\ \varphi \\ \vdots \\ \varphi^{h-1} \end{pmatrix} + \cdots + Z_h \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \\
&= X_0\,(1, \varphi, \ldots, \varphi^h) + \mathrm{Q}(\mathrm{Z})\,.
\end{aligned}
$$

- **Note that for $\delta > 0$,**

$$\mathbb{P}(|x^{-1}\mathrm{Q}(Z)| > \delta \mid |X_0| > x) = \mathbb{P}(|x^{-1}\mathrm{Q}(Z)| > \delta) \to 0\,.$$

- **Hence for a continuity set A with respect to the limit,**

$$\mathbb{P}(x^{-1}\mathrm{X}_h \in A \mid |X_0| > x)$$

$$= \mathbb{P}(x^{-1}X_0\,(1, \varphi, \ldots, \varphi^h) + o_{\mathbb{P}}(1) \in A \mid |X_0| > x)$$

$$\to \mathbb{P}\big(Y\,\Theta_0\,(1, \varphi, \ldots, \varphi^h) \in A\big)\,,$$

where $\mathbb{P}(\Theta_0 = \pm 1) = \widetilde{p}_\pm$.

The forward spectral tail process of an AR(1) process:

$$(\Theta_0, \ldots, \Theta_h) = \Theta_0\,(1, \varphi, \ldots, \varphi^h)$$

- **Example: The extremogram** Davis, Mikosch (2009, 2012)
- **For an $\mathbb{R}^d$-valued stationary regularly varying sequence $(\mathbf{X}_t)$ and Borel sets A, B bounded away from zero the extremogram is the limit function**

$$\rho_{AB}(h) = \lim_{x \to \infty} \mathbb{P}(x^{-1}\mathbf{X}_h \in B \mid x^{-1}\mathbf{X}_0 \in A)\,, \qquad h \in \mathbb{Z}\,.$$

By regular variation of $(\mathbf{X}_t)$ it is well defined.

- **For** $A = B$, ρ_{AA} **is the autocorrelation function of some stationary process:**

$$
\begin{aligned}
&\operatorname{corr}\big(\mathbf{1}(x^{-1}\mathbf{X}_h \in A)\,,\mathbf{1}(x^{-1}\mathbf{X}_0 \in A)\big)\\
&= \frac{\operatorname{cov}\big(\mathbf{1}(x^{-1}\mathbf{X}_h \in A)\,,\mathbf{1}(x^{-1}\mathbf{X}_0 \in A)\big)}{\big[\mathbb{P}(x^{-1}\mathbf{X}_h \in A)\,\mathbb{P}(x^{-1}\mathbf{X}_0 \in A)\big]^{1/2}}\\
&= \frac{\mathbb{P}\big(x^{-1}\mathbf{X}_h \in A\,, x^{-1}\mathbf{X}_0 \in A\big)}{\mathbb{P}(x^{-1}\mathbf{X}_0 \in A)} - \frac{\big[\mathbb{P}(x^{-1}\mathbf{X}_0 \in A)\big]^2}{\mathbb{P}(x^{-1}\mathbf{X}_0 \in A)}\\
&= \mathbb{P}(x^{-1}\mathbf{X}_h \in A \mid x^{-1}\mathbf{X}_0 \in A) + o(1)\,.
\end{aligned}
$$

The limit function is non-negative definite, hence the autocorrelation function of a stationary process.

- One can use the notions of time series analysis to describe the extremal dependence structure in a stationary sequence, e.g. long/short range dependence or time series spectral distribution for extremal events.
- For $d = 1$ and $A = (1, \infty)$,

$$\begin{aligned}\rho_{AA}(h) &= \lim_{x\to\infty} \mathbb{P}(x^{-1}X_h > 1 \mid X_0 > x)\\ &= \mathbb{P}(Y\,\Theta_h > 1)\\ &= \int_1^\infty \mathbb{P}(y\,\Theta_h > 1)\,\alpha\,y^{-\alpha-1}\,dy\\ &= \mathbb{E}[(\Theta_h \wedge 1)_+^\alpha]\end{aligned}$$

- **For the AR(1) process and $h \geq 1$,**

$$
\begin{aligned}
\rho_{AA}(h) &= \mathbb{E}[(\Theta_0\,\varphi^h)_+^\alpha] \\
&= \begin{cases} \widetilde{p}_+\,\varphi^{\alpha\,h}\,, & \varphi \in (0,1)\,, \\ |\varphi|^{\alpha\,h}\big[\widetilde{p}_+\,1(h\ \text{even}) + \widetilde{p}_-\,1(h\ \text{odd})\big]\,, & \varphi \in (-1,0)\,. \end{cases}
\end{aligned}
$$

- **Recall:** $\mathrm{corr}(X_0, X_h) = \varphi^h$, $h \geq 1$.

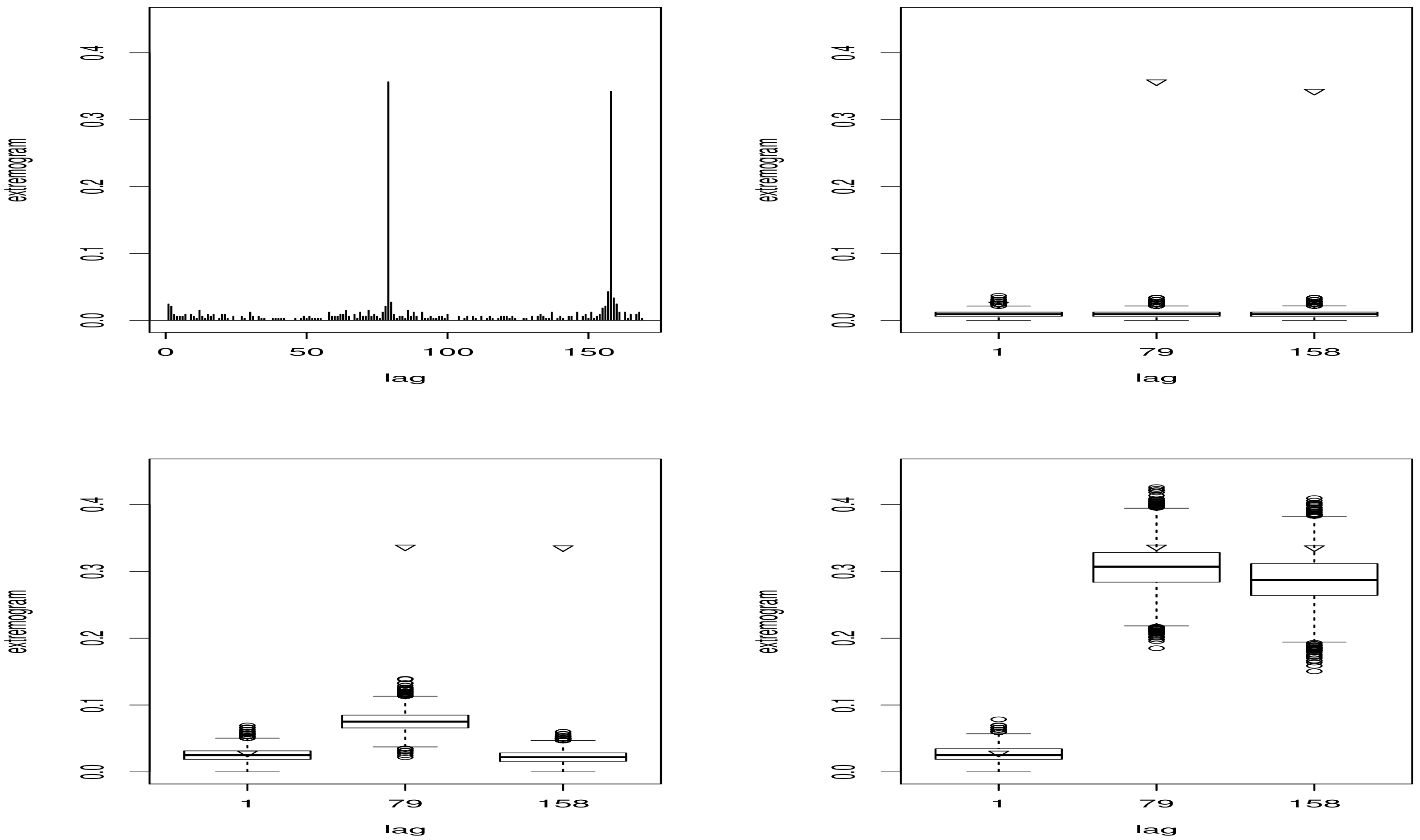


FIGURE 16. Top left: Sample extremogram for **5**-minute GE log-returns. Boxplots at lags 1, 79, 158 using permutations (top right) and stationary bootstrap with mean block size **50** and **200** (bottom).

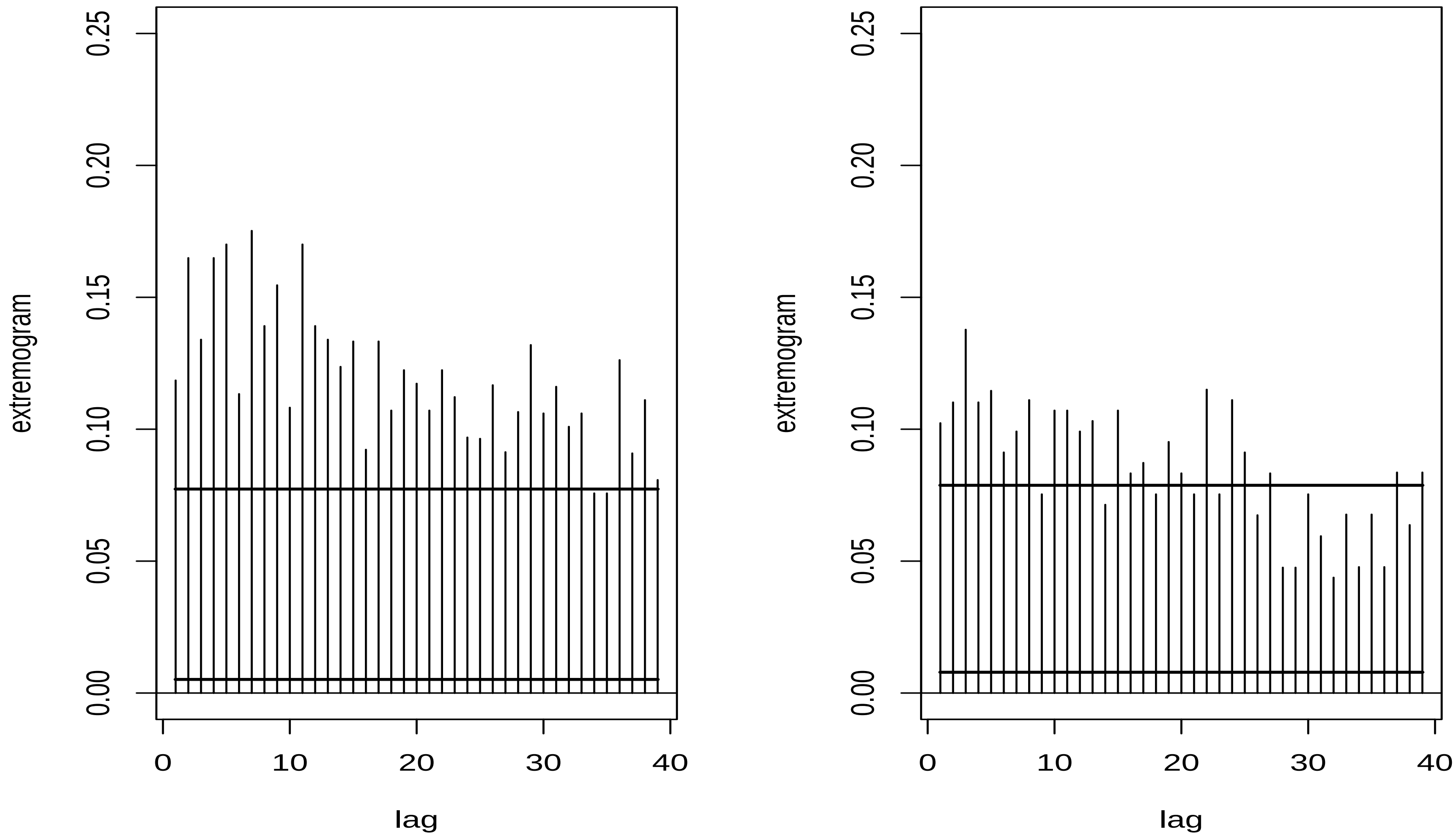


FIGURE 17. (Left) Ratio sample extremogram with $\boldsymbol{A} = \boldsymbol{B} = (\mathbf{1}, \boldsymbol{\infty})$ for **5** minute returns of USD-DEM foreign exchange rates; see also Figure 2.2 on p. 10. The extremogram alternates between large values at even lags and small ones at odd lags. This is an indication of AR behavior with negative leading coefficient. (Right) Ratio sample extremogram for the daily log-returns of the SP500 index.

8.3. **Examples.** **Take $A = B = (1, \infty)$.**

- **The extremogram of a GARCH$(1,1)$ process is not very explicit, but $\gamma_{AA}(h)$ decays exponentially fast to zero. This is in agreement with the geometric β-mixing property of GARCH.**

  Short serial extremal dependence

- **The stochastic volatility model with stationary Gaussian $(\log \sigma_t)$ and iid regularly varying (Z_t) with index $\alpha > 0$ has extremogram $\gamma_{AA}(h) = 0$ as in the iid case.**

  No serial extremal dependence

- **Recall: $\operatorname{corr}(X_0, X_h) = 0$, $h \geq 1$.**

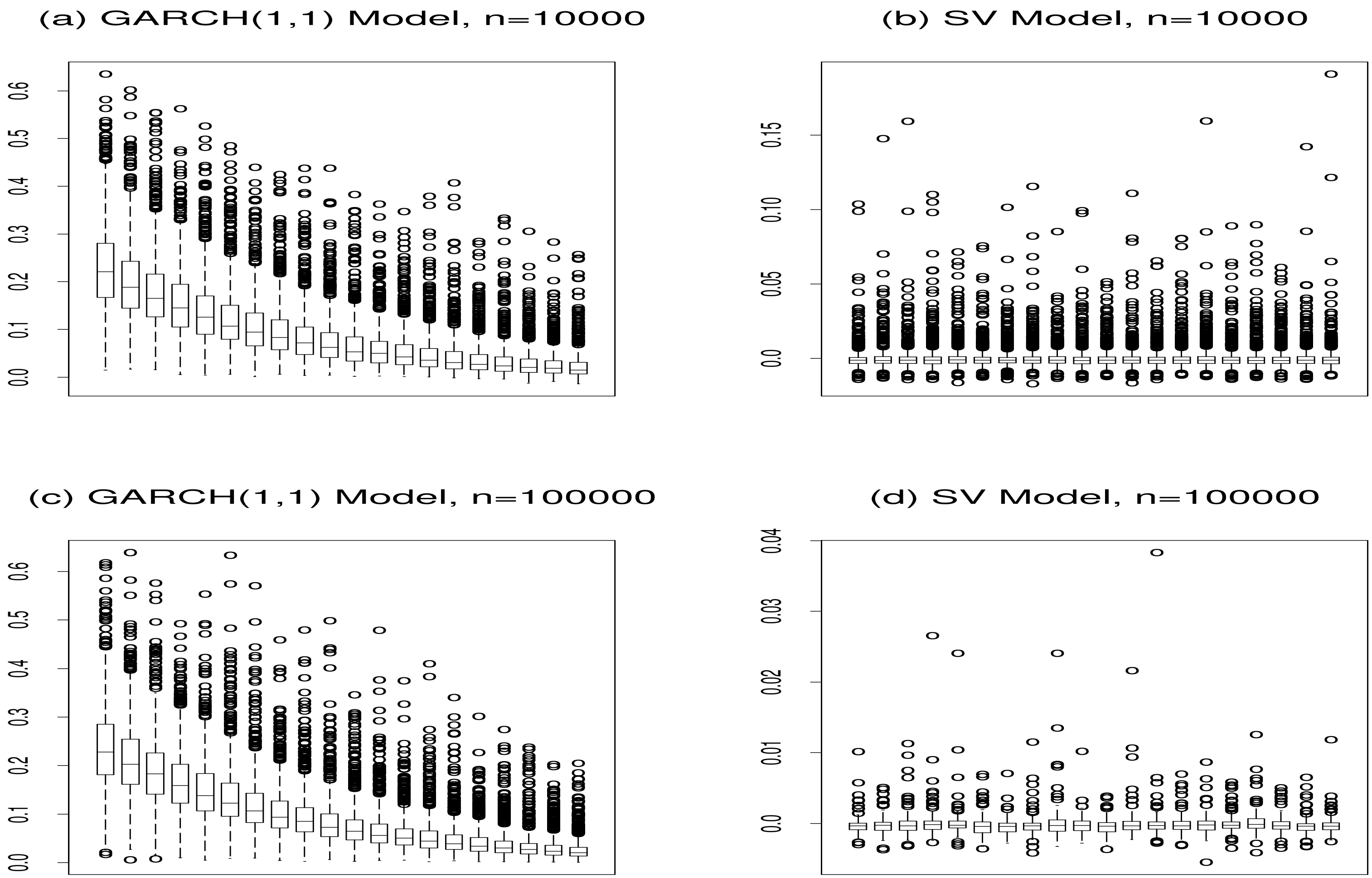


FIGURE 18. Ratio sample extremogram with $A = B = (1, \infty)$ for simulations of GARCH(1,1) and SV models. GARCH$(1,1)$ process $X_t = (0.0001 + 0.1X_{t-1}^2 + 0.9\sigma_{t-1}^2)^{0.5} Z_t$ for iid standard normal (Z_t). Stochastic volatility process $X_t = \sigma_t Z_t$ for iid student (Z_t) with 4 degrees of freedom, Gaussian ARMA(1,1) process $\log \sigma_t = 0.5 \log \sigma_{t-1} + 0.3\eta_{t-1} + \eta_t$.

Take-home messages

- One can use the extremogram to justify the selection of a particular time series model.
- For example, the autocorrelation functions of a GARCH$(1,1)$ process and a stochastic volatility model can be very similar, the extremogram of a stochastic volatility model vanishes while it is not the case for a GARCH$(1,1)$ process.
- One could define long/short range dependence in some meaningful way or the spectral distribution for extremal events in a strictly stationary sequence .

8.4. The sample extremogram.

- Let (X_t) be strongly mixing (possibly vector-valued) regularly varying.
- Assume $m = m_n \to \infty$ and $m_n/n \to 0$, and $a_m \to \infty$ satisfies $P(|X_0| > a_m) \sim m^{-1}$. Then

$$\widehat{P}_m(C) = \frac{m}{n} \sum_{t=1}^{n} I_{\{X_t/a_m \in C\}}$$

is a consistent estimator of

$$\mu_1(C) = \lim_{m\to\infty} m\, P(X_0/a_m \in C)\,.$$

- **In particular,**

$$E\widehat{P}_m(C) \rightarrow \mu_1(C)\,,$$

$$\mathrm{var}(\widehat{P}_m(C)) \sim \frac{m}{n}\sigma^2(C) = \frac{m}{n}\Big[\mu_1(C) + 2\sum_{h=1}^{\infty}\tau_h(C)\Big]$$

for

$$\tau_h(C) = \mu_{h+1}(C \times \mathbb{R}_0^{d(h-1)} \times C)\,.$$

- **For μ_1-continuity sets C bounded away from zero,**

$$\Big(\frac{n}{m}\Big)^{1/2}\,[\widehat{P}_m(C) - m\,P(a_m^{-1}\mathrm{X}_0 \in C)] \stackrel{d}{\rightarrow} N(0, \sigma^2(C))\,.$$

(pre-asymptotic central limit theorem).

- **An analogous result holds for finitely many sets $C_1, \ldots, C_h$.**

• The **ratio sample extremogram**

$$\widehat{\rho}_{AB}(h) = \frac{\frac{m}{n}\sum_{t=1}^{n-h} I_{\{a_m^{-1}\mathrm{X}_{t+h}\in B, a_m^{-1}\mathrm{X}_t\in A\}}}{\frac{m}{n}\sum_{t=1}^{n} I_{\{a_m^{-1}\mathrm{X}_t\in A\}}} = \frac{\sum_{t=1}^{n-h} I_{\{a_m^{-1}\mathrm{X}_{t+h}\in B, a_m^{-1}\mathrm{X}_t\in A\}}}{\sum_{t=1}^{n} I_{\{a_m^{-1}\mathrm{X}_t\in A\}}}, \quad h \geq 0\,,$$

estimates

$$\rho_{AB}(h) = \lim_{n\to\infty} P(a_n^{-1}X_h \in B \mid a_n^{-1}X_0 \in A) = \frac{\mu_{h+1}(A \times \overline{\mathbb{R}}_0^{d(h-1)} \times B)}{\mu_{h+1}(A \times \overline{\mathbb{R}}_0^{dh})}, \quad h \geq 0\,.$$

- **Pre-asymptotic limit theory for the ratio estimator follows from the previous central limit theory**

$$\Big(\frac{n}{m}\Big)^{1/2}\Big(\widehat{\rho}_{AB}(i)-\rho_{AB:m}(i)\Big)_{i=0,\ldots,h}\stackrel{d}{\rightarrow} N(0,\Sigma)\,,$$

where $\rho_{AB:m}(h)=P(a_m^{-1}X_h\in B\mid a_m^{-1}X_0\in A)$.

Problems

(1) The central limit theorem for the ratio sample estimator is pre-asymptotic. (For applications, the pre-asymptotic centering $\rho_{AB:m}(h) = P(a_m^{-1} X_h \in B \mid a_m^{-1} X_0 \in A)$ is more relevant than its limit $\rho_{AB}(h)$.)

(2) The asymptotic variance-covariance structure of the ratio sample estimator depends on expressions which are unknown.

Two methods to overcome (2):

random permutations and stationary bootstrap.

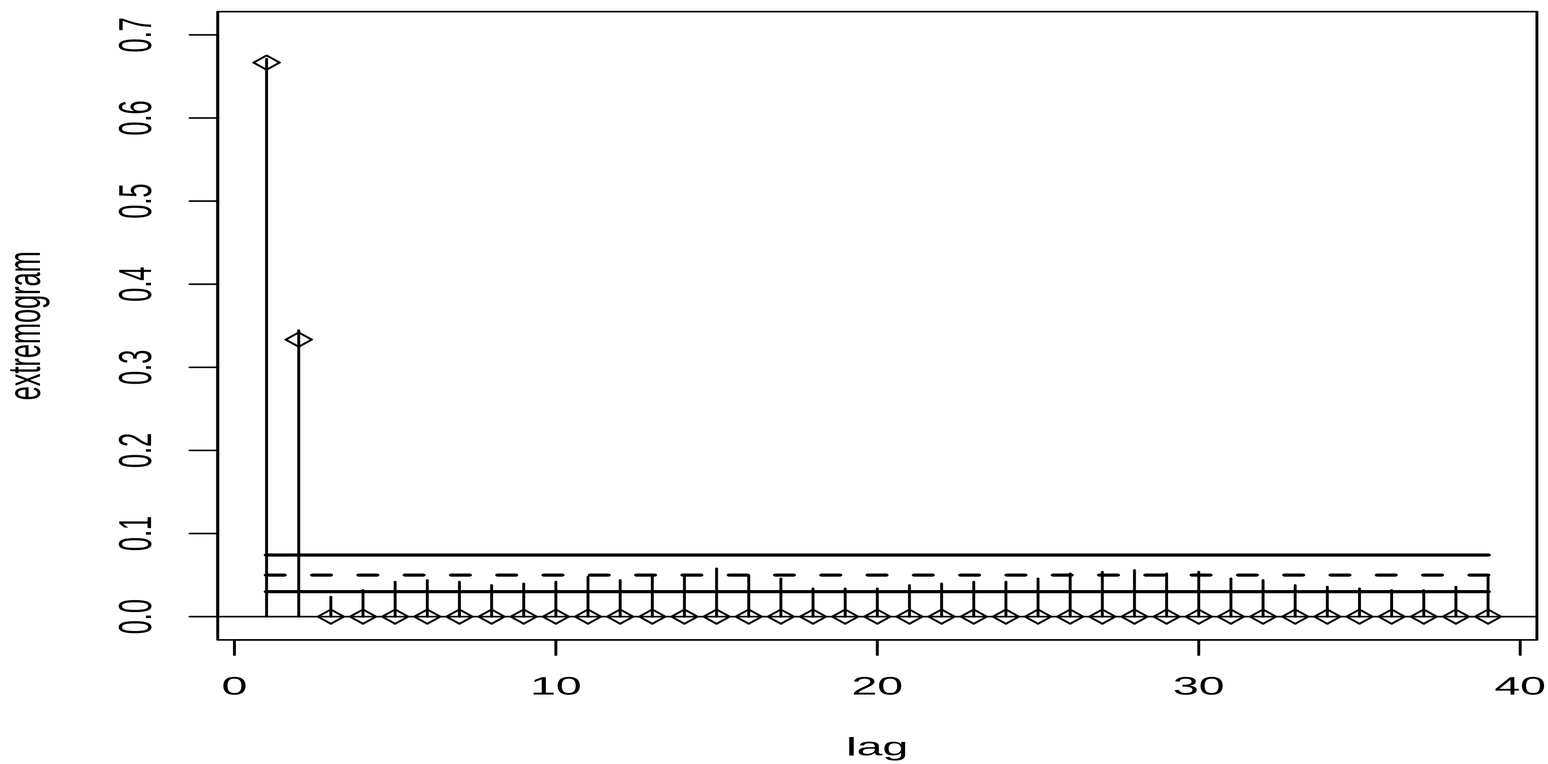


FIGURE 19. Sample extremogram for the max-moving average MMA(2) process. The diamonds superimposed on the figure represent the population extremogram values. Confidence bands are based on random permutations of the data.

Question for experts in empirical processes.

For practical reasons we would prefer

$$\widetilde{\rho}_{AB}(h) = \frac{\sum_{t=1}^{n-h} I_{\{|\mathrm{X}|_{(m)}^{-1}\mathrm{X}_{t+h}\in B,|\mathrm{X}|_{(m)}^{-1}\mathrm{X}_t\in A\}}}{\sum_{t=1}^{n} I_{\{|\mathrm{X}|_{(m)}^{-1}\mathrm{X}_t\in A\}}}, \quad h \geq 0,$$

where $|\mathrm{X}|_{(m)}$ is the m-th largest order statistic. Under mixing and anti-clustering (see below) condition one can easily show that

$$|\mathrm{X}|_{(m)}/a_{n/m} \stackrel{P}{\to} 1, m \to \infty, m/n \to 0.$$

For using uniform CLT on triangular arrays we would need to control the entropy of classes

$$\{xA : x \in (1-\varepsilon, 1+\varepsilon)\}$$

for every fixed set A that is a μ_1-continuity set.

9. The extremal index of a stationary process

9.1. Generalities.

- Assume (X_t) real-valued stationary, $X \sim F$, with right endpoint $x_F = \sup\{x \in \mathbb{R} : F(x) < 1\}$, and write

$$M_n = \max_{i=1,\dots,n} X_i \,, \qquad n \geq 1\,.$$

- Newell (1964), Loynes (1965), O'Brien (1974) observed for numerous examples and suitable sequences $u_n \uparrow x_F$ that

$$\mathbb{P}(M_n \leq u_n) \approx [\mathbb{P}(X \leq u_n)]^{n\,\theta_X}\,.$$

for numbers $\theta_X \in (0,1]$.

- Leadbetter (1983) made this fact precise.[4]

[4]See also the monograph Leadbetter, Lindgren, Rootzén (1983)

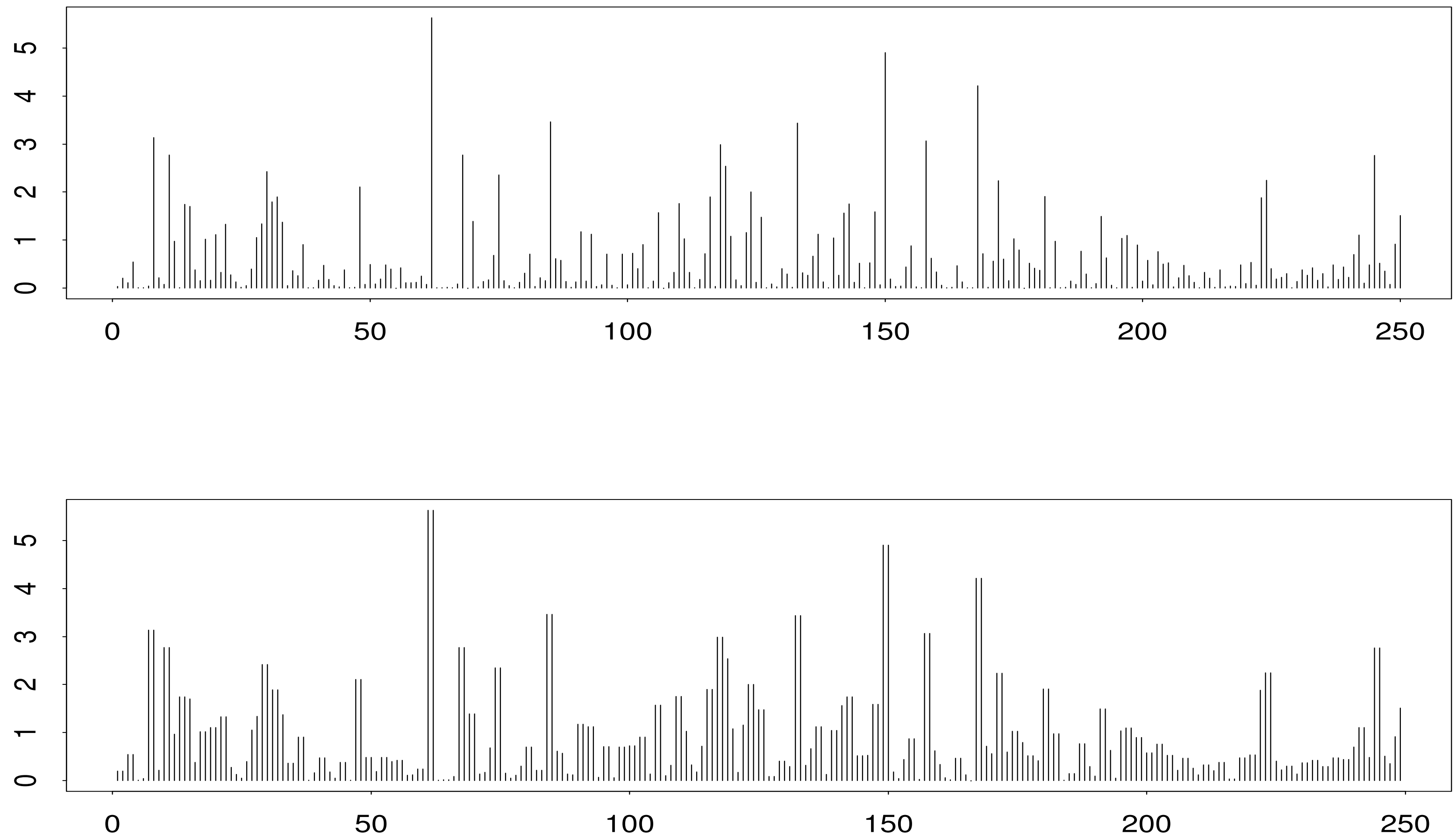


Figure 20. A sequence of iid random variables $\boldsymbol{Y_i}$ (Top) with distribution function $\sqrt{\boldsymbol{F}}$, where $\boldsymbol{F}$ is standard exponential. Bottom: the sequence of pairwise maxima $\mathbf{max}(\boldsymbol{Y_i}, \boldsymbol{Y_{i+1}})$ with distribution $\boldsymbol{F}$. By construction, extremes appear in clusters of size 2. The extremal index is $\boldsymbol{\theta_X} = \mathbf{1/2}$.

9.2. Leadbetter's condition D and definition of extremal index.

- **Idea 1:** Split the sample $X_1,\ldots,X_n$ into $k_n=[n/r_n]$ smaller blocks of length or size $r=r_n\to\infty$ while $k_n\to\infty$:

$$\underbrace{X_1,\ldots,X_{r_n}}_{\text{Block 1}},\underbrace{X_{r_n+1},\ldots,X_{2\,r_n}}_{\text{Block 2}},\ldots,\underbrace{X_{(k_n-1)\,r_n+1},\ldots,X_{k_n\,r_n}}_{\text{Block }k_n}.$$

- **Idea 2:** Approximate these k_n dependent blocks by k_n iid copies of the first block $(X_t)_{1\le t\le r_n}$: the blocks method S.N. Bernstein (1926), some kind of mixing.

For $\ell \in \mathbb{N}_+$ define

$$\alpha_{n,\ell} = \max_{A_1,A_2} \Big| \mathbb{P}\Big(\max_{t \in A_1 \cup A_2} X_t \le u_n \Big) - \mathbb{P}\Big(\max_{t \in A_1} X_t \le u_n \Big) \mathbb{P}\Big(\max_{t \in A_2} X_t \le u_n \Big) \Big|,$$

where the maximum is taken over all sets $A_1, A_2 \subset \{1, \dots, n\}$ such that A_1 and A_2 consist of integers $1 \le i_1 < \cdots < i_p$ and $j_1 < \cdots < j_q$, respectively, with the property $j_1 - i_p = \ell$.

Condition $D(u_n)$ holds if $\alpha_{n,\ell_n} \to 0$ for some integer sequence $\ell_n = o(n)$.

Asymptotic independence of block maxima under $D(u_n)$.

Assume the stationary sequence (X_n) satisfies $D(u_n)$ and $(n\,\overline{F}(u_n))$ is bounded. Then

$$\mathbb{P}(M_n \le u_n) = \big[\mathbb{P}(M_{r_n} \le u_n)\big]^{k_n} + o(1)\,, \qquad n \to \infty\,,$$

for any block sizes $r_n \to \infty$ and, in the notation used in formulating $D(u_n)$, such that $\ell_n / r_n \to 0$ and $k_n\, \alpha_{n,\ell_n} \to 0$.

Leadbetter's theorem

Assume

(1) for any $\tau > 0$ there exists $((u_n(\tau))$ such that $n\,\overline{F}(u_n(\tau)) \to \tau$

(2) $D(u_n(\tau))$ holds for any $\tau > 0$

(3) $\lim_{n\to\infty} \mathbb{P}\big(M_n \leq u_n(\tau)\big)$ exists.

Then there exists $\theta_X \in [0,1]$ such that this limit coincides with $\mathrm{e}^{-\theta_X\,\tau}$.

The extremal index of a stationary sequence

Assume (1), (2) and

$$\lim_{n\to\infty} \mathbb{P}\big(M_n \leq u_n(\tau)\big) = \mathrm{e}^{-\theta_X\,\tau} \tag{9.1}$$

for any $\tau > 0$ and some $\theta_X \in [0,1]$. Then θ_X is the extremal index of (X_t).

- **For an iid sequence (X_t), (9.1) with $\theta_X = 1 \iff n\,\overline{F}(u_n) \to \tau$.**
- $F \in \mathrm{MDA}(H)$ **holds** $\iff n\,\overline{F}(a_n\,x + d_n) \to \log H(x)$, $\forall x$, **for suitable** $a_n > 0, d_n \in \mathbb{R}$.
- **If** $\theta_X > 0$ **exists:** $\mathbb{P}(a_n^{-1}(M_n - d_n) \le x) \to H^{\theta_X}(x)$, $\forall x$, **and** H^{θ_X} **is of the same type as** H.

9.3. The extremal index as reciprocal of the expected cluster size above high thresholds.

- **What is an extremal cluster?**
- **Folklore: θ_X is the reciprocal of the expected cluster size above high thresholds**

- An adhoc statistical answer:

> **Blocks method:**
>
> $$\underbrace{X_1,\ldots,X_{r_n}}_{\text{Block }1},\underbrace{X_{r_n+1},\ldots,X_{2\,r_n}}_{\text{Block }2},\ldots,\underbrace{X_{(k_n-1)\,r_n+1},\ldots,X_{k_n\,r_n}}_{\text{Block }k_n}.$$
>
> An extremal cluster in the sample $X_1,\ldots,X_n$: a block if there is at least one exceedance of $u=u_n\uparrow x_F$ in this block.

- By stationarity of (X_t) the expected cluster size in one block is given by

$$\mathbb{E}\Big[\sum_{t=1}^{r_n}\mathbf{1}(X_t>u_n)\,\Big|\,M_r>u_n\Big]=\sum_{t=1}^{r_n}\frac{\mathbb{P}(X_t>u_n\,,M_r>u_n)}{\mathbb{P}(M_r>u_n)}$$

$$=\sum_{t=1}^{r_n}\frac{\mathbb{P}(X_t>u_n)}{\mathbb{P}(M_r>u_n)}$$

$$=\frac{r_n\,\mathbb{P}(X>u_n)}{\mathbb{P}(M_r>u_n)}=\theta_n^{-1}\,.$$

- Leadbetter (1983): **under mild regularity conditions on (X_t) and if $u_n \uparrow x_F$ the limit**

$$\theta = \lim_{n\to\infty} \theta_n = \lim_{n\to\infty} \frac{\mathbb{P}(M_r > u_n)}{r_n\,\mathbb{P}(X > u_n)} \in [0,1]$$

exists and $\theta = \theta_X$.

For this reason, the extremal index θ_X is often referred to as

the reciprocal of the expected extremal cluster size above high thresholds.

Assume

- $\forall \tau > 0$ **there exists** $u_n = u_n(\tau)$ **such that** $n\,\overline{F}(u_n) \to \tau$.
- **Anticlustering condition (AC):**[a]

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n} > u_n \mid X_0 > u_n\big) = 0\,.$$

- **Mixing condition (M):**[b]

$$\mathbb{P}(M_n \le u_n) = \big[\mathbb{P}(M_{r_n} \le u_n)\big]^{k_n} + o(1)\,.$$

Then

- **If (AC) holds then**

$$\lim_{k\to\infty} \limsup_{n\to\infty} \Big|\theta_n - \mathbb{P}(M_k \le u_n \mid X_0 > u_n)\Big| = 0\,,$$

  and $\liminf_{n\to\infty} \theta_n > 0$.
- **If also (M) holds and** $\theta = \lim_{n\to\infty} \theta_n$ **exists then** θ_X **exists and** $\theta = \theta_X$.

[a] $M_{s,t} = \max_{s\le i\le t} X_i$ for $s \le t$.

[b] Satisfied if conditions on p. 89 hold.

- (AC) is easily verified when (X_t) is m-dependent or Ψ-mixing. Then also (M) holds.

- Example: Regularly varying AR(1) process. Assume $X_t = \varphi X_{t-1} + Z_t$, $|\varphi| < 1$, (Z_t) iid and $Z \in \mathrm{RV}(\alpha)$, $n\,\mathbb{P}(|X| > a_n) \to 1$.

  Recall that $X_t = \varphi^t X_0 + Q_t(\mathrm{Z})$ and $\max_{t \le r_n} |Q_t(\mathrm{Z})| = O_{\mathbb{P}}(1)$ as $t \to \infty$. Then for a Pareto(α)-distributed Y,

$$
\begin{aligned}
&\mathbb{P}(M_{k,r_n} > a_n \mid X_0 > a_n) \\
&\quad \le c\,\mathbb{P}(a_n^{-1}|X_0|\,|\varphi|^k + o_{\mathbb{P}}(1) > 1 \mid |X_0| > a_n) \\
&\quad \to \mathbb{P}(Y\,|\varphi|^k > 1)\,, \qquad n \to \infty\,, \\
&\quad \to 0\,, \qquad k \to \infty\,.
\end{aligned}
$$

- (AC) is much weaker than Leadbetter's $D'(u_n)$ condition:

$$\lim_{\ell\to\infty}\limsup_{n\to\infty}\sum_{t=1}^{[n/\ell]}\mathbb{P}(X_t > u_n \mid X_0 > u_n) = 0\,.$$

Under $D(u_n), D'(u_n)$, for a stationary Gaussian sequence (X_t) with standard normal marginals, with $n\,\overline{\Phi}(d_n) \sim 1$, $u_n = x/d_n + d_n$, and $\mathrm{corr}(X_0, X_h) = o(1/\log h)$: Leadbetter, Lindgren, Rootzén (1983)

$$d_n(M_n - d_n) \xrightarrow{d} G \sim \Lambda \qquad \text{and } \theta_X = 1$$

No extremal clustering for all reasonable Gaussian sequences.

- **Example: The extremal index of a regularly varying sequence**

  Assume

  – (X_t) stationary and non-negative.

  – (AC) holds for $u_n = a_n$.

  Then $\theta = \lim_{n\to\infty} \theta_n > 0$ exists and has representation

$$\theta = \mathbb{P}\Big(Y \sup_{t\geq 1} \Theta_t \leq 1\Big) = \mathbb{E}\Big[\Big(1 - \sup_{t\geq 1} \Theta_t^\alpha\Big)_+\Big]$$
$$= \mathbb{E}\Big[\sup_{t\geq 0} \Theta_t^\alpha - \sup_{t\geq 1} \Theta_t^\alpha\Big].$$

  If also (M) holds for $u_n = x\,a_n$, $x > 0$, then θ_X exists and $\theta_X = \theta$.

- **Proof.** By regular variation, for fixed k,

$$\mathbb{P}(M_k \leq a_n \mid X_0 > a_n) \to \mathbb{P}\Big(Y \max_{t=1,\ldots,k} \Theta_t \leq 1\Big)$$

But if (AC) holds,

$$\lim_{k\to\infty} \limsup_{n\to\infty} \left| \theta_n - \mathbb{P}(M_k \le u_n \mid X_0 > u_n) \right| = 0\,,$$

and $\liminf_{n\to\infty} \theta_n > 0$. **Hence**

$$\begin{aligned}\theta &= \lim_{n\to\infty} \theta_n \\ &= \lim_{k\to\infty} \mathbb{P}\Big(Y \max_{t=1,\ldots,k} \Theta_t \le 1\Big) \\ &= \mathbb{P}\Big(Y \sup_{t\ge 1} \Theta_t \le 1\Big)\,,\end{aligned}$$

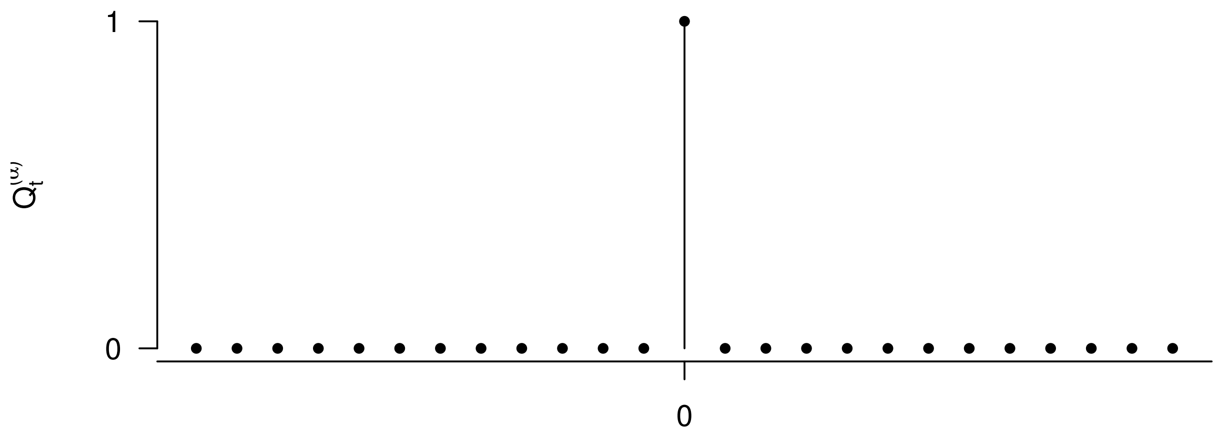
and $\theta > 0$.

If also (M) holds, $\theta = \theta_X$.

- **Example:** **Assume asymptotic independence. Then $\Theta_t = 0$ a.s. for $t \neq 0$. Hence**

$$\theta = \mathbb{P}\Big(Y \sup_{t \geq 1} \Theta_t \leq 1\Big) = \mathbf{1}(0 \leq 1) = 1\,,$$

and $\theta = \theta_X$



10. Time-change properties of the spectral tail process Basrak, Segers (2009)

Change of measure

For any process $(\mathbf{W}_t)$ we denote

$$\mathbb{P}_t^{\alpha}(\mathbf{W} \in \cdot) = \mathbb{E}\Big[\frac{|\boldsymbol{\Theta}_t|^{\alpha}}{\mathbb{E}\big[|\boldsymbol{\Theta}_t|^{\alpha}\big]}\mathbf{1}(\mathbf{W} \in \cdot)\Big]$$

$$= \mathbb{E}\Big[\frac{|\boldsymbol{\Theta}_t|^{\alpha}}{\mathbb{E}\big[|\boldsymbol{\Theta}_t|^{\alpha}\big]}\mathbf{1}(\mathbf{W} \in \cdot\,, \boldsymbol{\Theta}_t \neq 0)\Big]\,.$$

The spectral tail process (Θ_t) of an $\mathbb{R}^d$-valued regularly varying stationary sequence has the time-change properties:

1. For any $t \in \mathbb{Z}$,

$$\mathbb{E}[|\Theta_t|^\alpha] = \mathbb{P}(\Theta_{-t} \neq 0)\,.$$

2. For any $t \in \mathbb{Z}$ such that $\mathbb{P}(\Theta_{-t} \neq 0) > 0$, and any $h \geq 0$,

$$\mathbb{P}((\Theta_{-h}, \ldots, \Theta_h) \in \cdot \mid \Theta_{-t} \neq 0) \;=\; \mathbb{P}_t^\alpha\Big(\frac{(\Theta_{t-h}, \ldots, \Theta_{t+h})}{|\Theta_t|} \in \cdot\Big)\,.$$

Moreover, $\mathbb{E}[|\Theta_t|^\alpha] = 1$ if and only if for any $h \geq 0$,

$$\mathbb{P}((\Theta_{-h}, \ldots, \Theta_h) \in \cdot) = \mathbb{P}_t^\alpha\Big(\frac{(\Theta_{t-h}, \ldots, \Theta_{t+h})}{|\Theta_t|} \in \cdot\Big)\,.$$

- **Example: Stationary solution to stochastic recurrence equation.**
- $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))_{t\in\mathbb{Z}}$, $A, B > 0$, **and**
  - **the conditions of Kesten-Goldie hold, in particular** $\mathbb{E}[A^\alpha] = 1$ **for some** $\alpha > 0$,
- **Then** (X_t) **is regularly varying with index** α **and the forward spectral tail process is given for** $h \geq 0$,[5]

$$(\Theta_0, \ldots, \Theta_h) = (1, \Pi_1, \ldots, \Pi_h), \qquad \Pi_h = A_1 \cdots A_h\,.$$

[5] Proof is similar to an AR(1) process; see p. 69.

- **The backward spectral tail process for $h \geq 0$, $t = h$, by time-change:**

$$
\begin{aligned}
&\mathbb{P}((\boldsymbol{\Theta}_{-h}, \ldots, \boldsymbol{\Theta}_h) \in \cdot \mid \boldsymbol{\Theta}_{-h} \neq \mathbf{0}) \\
&\quad = \mathbb{P}((\boldsymbol{\Theta}_{-h}, \ldots, \boldsymbol{\Theta}_h) \in \cdot) \\
&\quad = \mathbb{P}_h^{\alpha}\Big(\frac{(\boldsymbol{\Theta}_0, \ldots, \boldsymbol{\Theta}_{2h})}{\boldsymbol{\Theta}_h} \in \cdot\Big) \\
&\quad = \mathbb{E}\Big[\frac{\boldsymbol{\Pi}_h^{\alpha}}{\mathbb{E}[\boldsymbol{\Pi}_h^{\alpha}]}\mathbf{1}\Big(\frac{(1, \boldsymbol{\Pi}_1, \ldots, \boldsymbol{\Pi}_{2h})}{\boldsymbol{\Pi}_h} \in \cdot\Big)\Big]. \\
&\quad = \mathbb{E}\Big[\boldsymbol{\Pi}_h^{\alpha}\mathbf{1}\Big(\Big(\frac{1}{\boldsymbol{\Pi}_h}, \frac{\boldsymbol{\Pi}_1}{\boldsymbol{\Pi}_h}, \ldots, 1, \frac{\boldsymbol{\Pi}_{h+1}}{\boldsymbol{\Pi}_h}, \ldots, \frac{\boldsymbol{\Pi}_{2h}}{\boldsymbol{\Pi}_h}\Big) \in \cdot\Big)\Big].
\end{aligned}
$$

since $\mathbb{E}[\boldsymbol{\Pi}_h^{\alpha}] = (\mathbb{E}[A^{\alpha}])^h = 1$ **and** $\mathbb{P}(\boldsymbol{\Theta}_{-h} \neq 0) = \mathbb{E}[\boldsymbol{\Theta}_h^{\alpha}] = 1$.

- **Write for $i \geq 1$,**

$$\boldsymbol{\Pi}_i = \mathrm{e}^{S_i}\,, \quad S_i = \sum_{t=1}^{i} \log A_t\,, \qquad S_{-i} = \sum_{t=-i}^{-1} \log A_t\,.$$

- **Then**

$$\begin{aligned}
&\mathbb{P}((\boldsymbol{\Theta}_{-h},\ldots,\boldsymbol{\Theta}_{-1}) \in C_1\,,(\boldsymbol{\Theta}_0,\ldots,\boldsymbol{\Theta}_h) \in C_2)\\
&= \mathbb{E}\Big[\mathrm{e}^{\alpha\, S_h}\mathbf{1}\Big((\mathrm{e}^{-S_h}\,,\ldots\,,\mathrm{e}^{S_{h-1}-S_h}) \in C_1)\\
&\qquad\qquad \mathbf{1}\Big((1,\mathrm{e}^{S_{h+1}-S_h},\ldots,\mathrm{e}^{S_{2h}-S_h}) \in C_2\Big)\Big]\\
&= \mathbb{E}\Big[\mathrm{e}^{\alpha\, S_{-h}}\,\mathbf{1}\Big((\mathrm{e}^{-S_{-h}}\,,\ldots\,,\mathrm{e}^{-S_{-1}}) \in C_1\Big)\Big]\\
&\qquad\qquad \mathbb{P}\Big((1,\mathrm{e}^{S_1},\ldots,\mathrm{e}^{S_h}) \in C_2\Big)\Big]
\end{aligned}$$

Forward and backward spectral processes are independent.

Recall: proved by de Haan, Resnick, Rootzén, de Vries (1989)

$$\theta_X = \mathbb{E}\Big[\Big(1 - \sup_{t\ge 1} \Theta_t^\alpha\Big)_+\Big] = \mathbb{E}\Big[\Big(1 - \sup_{t\ge 1} \Pi_t^\alpha\Big)_+\Big].$$

- θ_X can also be written as $(S_0 = 0)$

$$\theta_X = \mathbb{P}^\alpha\Big(-\min_{t\le -1} S_t < 0\Big)\mathbb{P}\Big(\min_{t\le -1} S_t \le 0\Big),$$

where $\mathbb{P}^\alpha(A_{-t} \in \cdot) = \mathbb{E}[A_t^\alpha 1(A_t \in \cdot)]$, $t \ge 1$.

- For $A_t = \exp(\sqrt{2}N_t - 1)$ we have $\alpha = 1$, and Chang, Peres (1997)

$$\theta_X = \mathbb{P}\Big(\min_{t\le -1} S_t \le 0\Big)^2 \approx \frac{1}{2}\exp\Big(\frac{\zeta(0.5)}{\sqrt{2\pi}}\Big) \approx 0.2792.$$

ℓ^α-properties of the spectral tail process Janßen (2019)

Denote $\|\mathbf{x}\|_\alpha$ the ℓ^α-norm of any sequence $\mathbf{x} = (\mathbf{x}_t)_{t\in\mathbb{Z}} \in (\mathbb{R}^d)^{\mathbb{Z}}$:

$$\|\mathbf{x}\|_\alpha^\alpha = \sum_{t\in\mathbb{Z}} |\mathbf{x}_t|^\alpha \,.$$

We have the equivalence between the assertions:

- $|\Theta_t| \to 0$ **as** $t \to \infty$ **a.s.,**
- $|\Theta_t| \to 0$ **as** $t \to -\infty$ **a.s.,**
- $\|\Theta\|_\alpha < \infty$ **a.s.**

11. Point process convergence for stationary regularly varying sequences Davis, Hsing (1995), Basrak, Segers (2009)

11.1. Point process convergence.

- **The Laplace functional of a point process N on $E \subset \mathbb{R}^d$:**[6]

$$\Psi_N(f) = \mathbb{E}\exp\Big(-\int_E f\,dN\Big)\,, \qquad f \in \mathbb{C}_K^+\,,$$

determines the distribution of N.

- **$N_n \xrightarrow{d} N$ on E if and only if $\Psi_{N_n}(f) \to \Psi_N(f)$ for a suitable class of functions f, e.g. $f \in \mathbb{C}_K^+$.**
- **Example: $N = \sum_{i=1}^{\infty} \varepsilon_{\xi_i}$. Then**

$$\Psi_N(f) = \mathbb{E}\Big[\exp\Big(-\sum_{i=1}^{\infty} f(\xi_i)\Big)\Big]\,.$$

[6] $\mathbb{C}_K^+$ consists of the continuous functions $f \geq 0$ on E with compact support.

Example: $(\mathbf{X}_{ni})$ **triangular array of row-wise iid random vectors and** $n\,\mathbb{P}(\boldsymbol{X}_{n1}\in\cdot)\stackrel{v}{\to}\mu(\cdot)$ **for some Radon measure** μ **on** E**.**

Then for $N_n=\sum_{i=1}^n \varepsilon_{\mathbf{X}_{ni}}$, Resnick (2007)

$$\begin{aligned}
\Psi_{N_n}(f) &= \mathbb{E}\Big[\exp\Big(-\sum_{i=1}^n f(\boldsymbol{X}_{ni})\Big)\Big] \\
&= \Big(\mathbb{E}[\exp(-f(\boldsymbol{X}_{n1}))\Big)^n \\
&= \Big(1-\frac{n\big(1-\mathbb{E}[\mathrm{e}^{-f(\boldsymbol{X}_{n1})}]\big)}{n}\Big)^n \\
&= \Big(1-\frac{\int_E\big(1-\mathrm{e}^{-f(\mathrm{x})}\big)\,\big[n\,\mathbb{P}(\mathbf{X}_{n1}\in d\mathrm{x})\big]}{n}\Big)^n \\
&\to \exp\Big(-\int_E\big(1-\mathrm{e}^{-f(x)}\big)\mu(d\mathrm{x})\Big) = \Psi_N(f)\,.
\end{aligned}$$

- Ψ_N **is the Laplace functional of a PRM(**μ**).**

- If (X_t) iid, $\mathrm{X} \in \mathrm{RV}(\alpha, \mu_{\mathrm{X}})$, $\mathrm{X}_{nt} = a_n^{-1}\mathrm{X}_t$, then $n\,\mathbb{P}(\mathrm{X}_{n1} \in \cdot) = n\,\mathbb{P}(a_n^{-1}\boldsymbol{X} \in \cdot) \xrightarrow{v} \mu_{\mathrm{X}}$ on $\boldsymbol{E} = \mathbb{R}_0^d$ and

$$N_n = \sum_{t=1}^{n} \varepsilon_{a_n^{-1}\mathrm{X}_t} \xrightarrow{d} N \sim \mathrm{PRM}(\mu_{\mathrm{X}})\,.$$

- Now consider an $\mathbb{R}^d$-valued regularly varying stationary sequence (X_t).
- Recall the blocks method: for $r_n \to \infty$, $k_n = [n/r_n]$,

$$\underbrace{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{r_n}}_{\text{Block } 1}, \underbrace{\boldsymbol{X}_{r_n+1}, \ldots, \boldsymbol{X}_{2\,r_n}}_{\text{Block } 2}, \ldots, \underbrace{\boldsymbol{X}_{(k_n-1)\,r_n+1}, \ldots, \boldsymbol{X}_{k_n\,r_n}}_{\text{Block } k_n}.$$

Mixing condition (MC) for point processes.

Assume $n\,\mathbb{P}(|\mathbf{X}| > a_n) \to 1$ **and for** $f \in \mathbb{C}_K^+$,

$$\Psi_{N_n}(f) = (\Psi_{\widehat{N}_{r_n}}(f))^{k_n} + o(1)\,, \qquad n \to \infty\,,$$

where $\widehat{N}_{r_n} = \sum_{t=1}^{r_n} \varepsilon_{a_n^{-1}\mathbf{X}_t}$.

- $(\Psi_{\widehat{N}_{r_n}}(f))^n$ **is the Laplace functional of**

$$\widetilde{N}_n = \sum_{i=1}^{k_n} \widehat{N}_{r_n,i}\,, \qquad (\widehat{N}_{r_n,i}),\ i = 1,\ldots,k_n \text{ iid copies of } \widehat{N}_{r_n}.$$

Anticlustering condition (AC):[a]

$$\lim_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}\big(M_{k,r_n}^{|\mathbf{X}|} > a_n \mid |\mathbf{X}_0| > a_n\big) = 0\,.$$

[a] $M_{s,t}^{|\mathbf{X}|} = \max_{s\le i\le t} |\mathbf{X}|_i$ for $s \le t$.

Remark the same anticlustering condition as for the existence of the extremal index for the threshold $u_n = a_n$.

Point process convergence

Davis, Hsing (1995), Basrak, Segers (2009) **Assume**

- $(\mathbf{X}_t)$ is an $\mathbb{R}^d$-valued regularly varying stationary process $(\mathbf{X}_t)$ with index $\alpha > 0$
- (AC) and (MC)
- $n\,\mathbb{P}(|\mathbf{X}| > a_n) \to 1$.

Then $N_n \xrightarrow{d} N$ where

$$\Psi_N(f) = \exp\Big(- \int_0^\infty \mathbb{E}\big[e^{-\sum_{t=1}^\infty f(y\,\Theta_t)} \big(1 - e^{-f(y\,\Theta_0)}\big)\big]\, d(-y^{-\alpha})\Big)\,.$$

The process N is a Poisson cluster process:

$$N = \sum_{i=1}^\infty \sum_{j=1}^\infty \varepsilon_{\Gamma_i^{-1/\alpha} \mathbf{Q}_{ij}}\,, \qquad n \to \infty\,,$$

where

- (Γ_i) are the points of a unit rate homogeneous Poisson process on $(0,\infty)$
- independent of the iid sequence $(\mathbf{Q}_{ij})_{j\in\mathbb{Z}}$ versions of the spectral cluster process

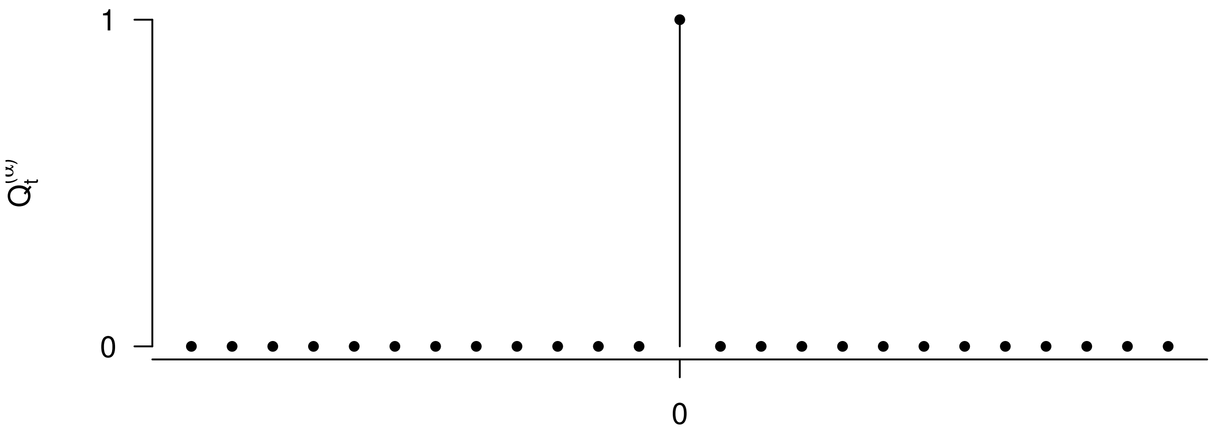
$$\mathbf{Q} = \frac{\Theta}{\|\Theta\|_\alpha}\,.$$

11.2. The spectral cluster process Buriticá, Meyer, Mikosch, W. (2021).

- If $\Theta_t = 0$ a.s. for $t \neq 0$ (asymptotic independence) then

$$N = \sum_{i=1}^{\infty} \varepsilon_{\Gamma_i^{-1/a} \boldsymbol{\Theta}_{i0}} \sim \mathrm{PRM}(\mu_{\mathrm{X}})\,.$$

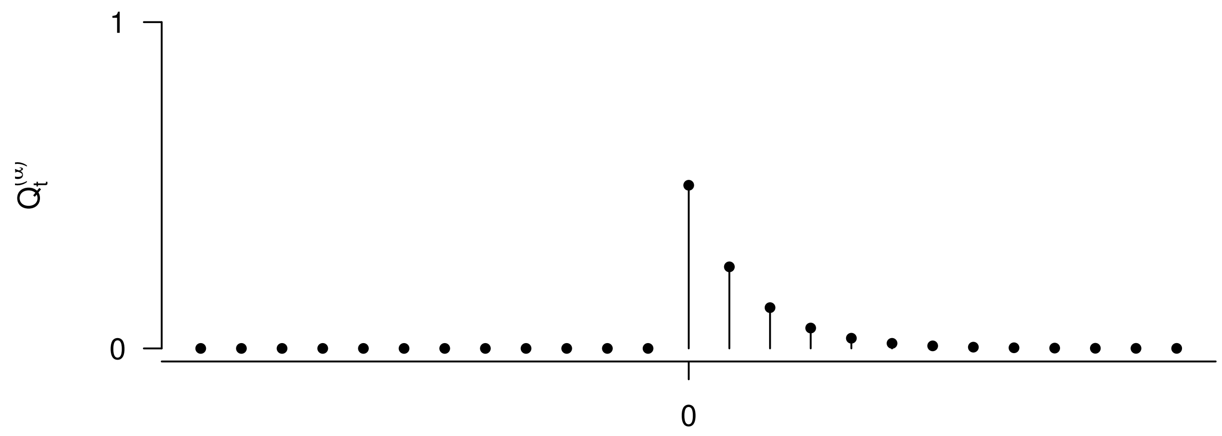
Then $\mathbf{Q} = \boldsymbol{\Theta} = (\ldots, 0, \boldsymbol{\Theta}_0, 0, \ldots)$



- (X_t) a stationary AR(1), $X_t = \varphi X_{t-1} + Z_t$ with $\varphi \in (0,1)$, and (Z_t) iid satisfying RV_α,

$$Q_t^{(\alpha)} = \Theta_t / \|\Theta\|_\alpha \quad = \quad \varphi^t \Theta_0^Z \, 1(J + t \geq 0)\,(1 - \varphi^\alpha)^{1/\alpha},$$

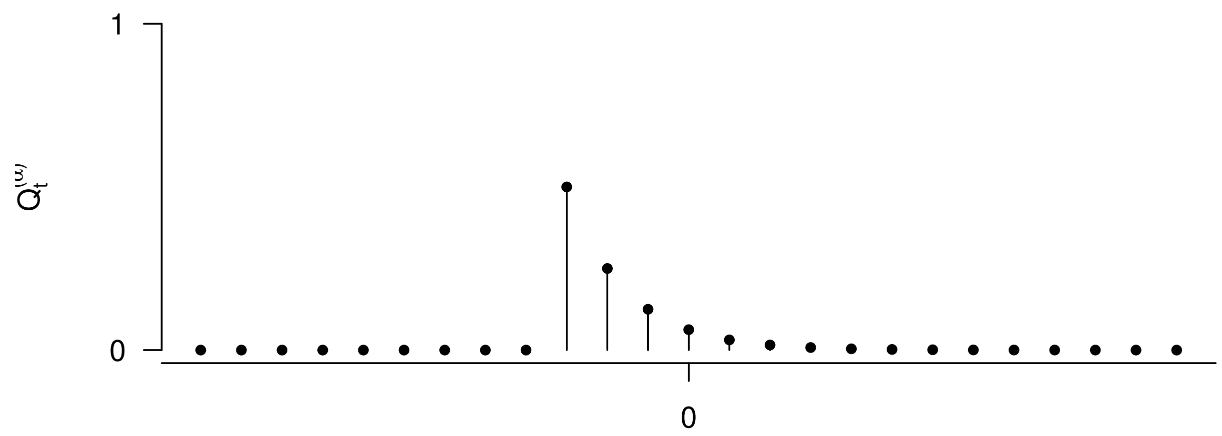
J independent of Θ_0^Z, $\mathbb{P}(J = j) = (1 - \varphi^\alpha)\varphi^{j\alpha}, j \geq 0.$



- (X_t) a stationary AR(1), $X_t = \varphi X_{t-1} + Z_t$ with $\varphi \in (0,1)$, and (Z_t) iid satisfying RV_α,

$$Q_t^{(\alpha)} = \Theta_t / \|\Theta\|_\alpha \quad = \quad \varphi^t \Theta_0^Z \, 1(J + t \geq 0) \, (1 - \varphi^\alpha)^{1/\alpha},$$

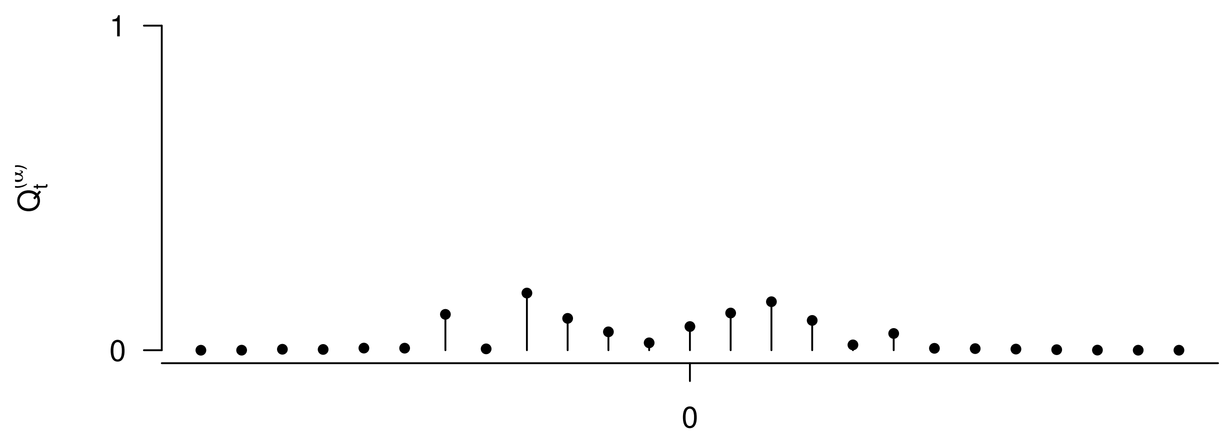
J independent of Θ_0^Z, $\mathbb{P}(J = j) = (1 - \varphi^\alpha)\varphi^{j\alpha}, j \geq 0.$



- (X_t) causal solution to SRE, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive iid and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \;=\; A_t \cdots A_1\,, \quad t \geq 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.

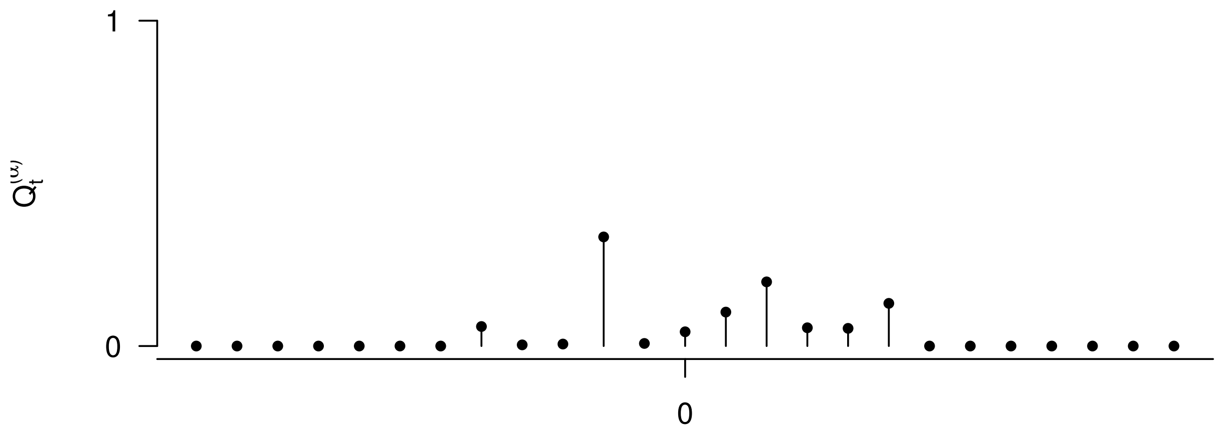


We take $A_t = e^{N_t - 1/2}$ such that (N_t) is iid gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leq 0$.

- (X_t) causal solution to SRE, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive iid and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \ = \ A_t \cdots A_1 \, , \quad t \geq 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.

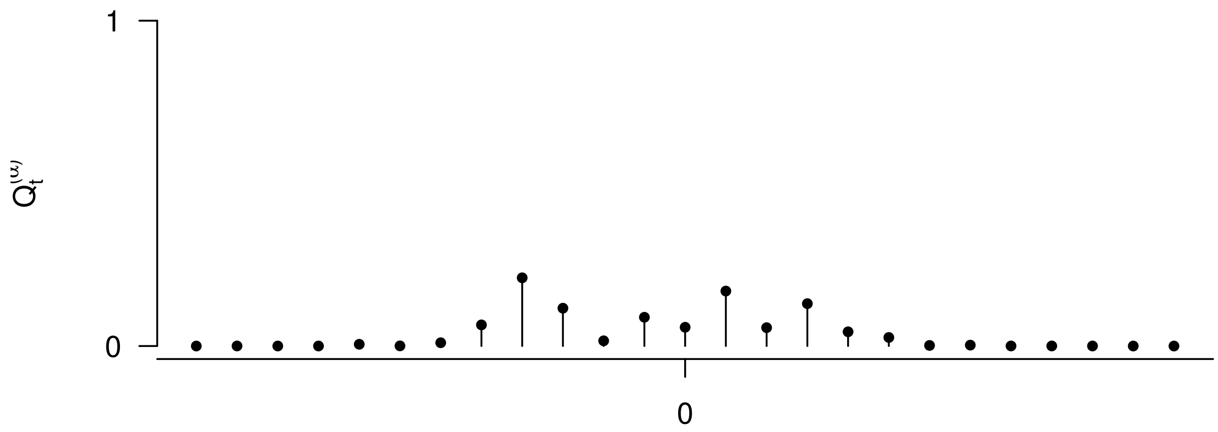


We take $A_t = e^{N_t - 1/2}$ such that (N_t) is iid gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leq 0$.

- (X_t) causal solution to SRE, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive iid and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \; = \; A_t \cdots A_1 \, , \quad t \geq 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.

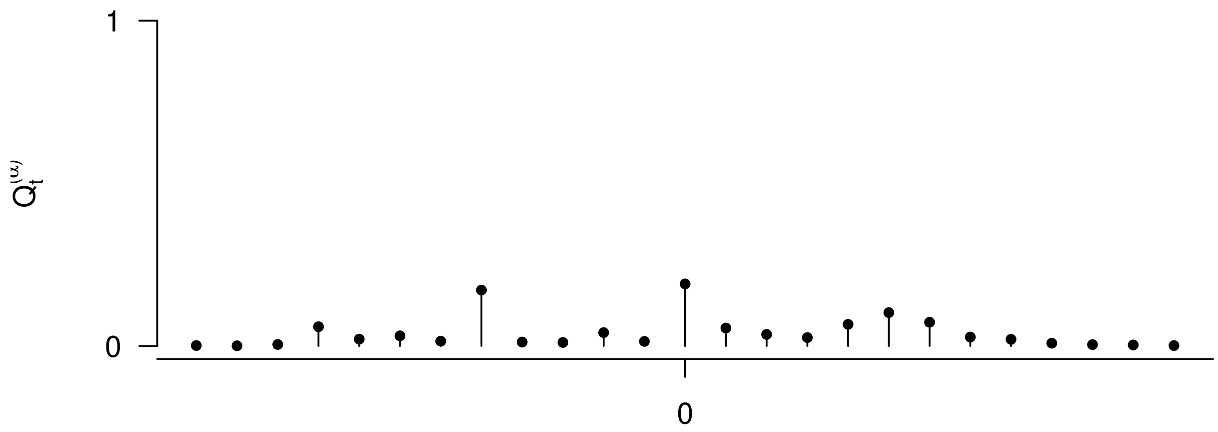


We take $A_t = e^{N_t - 1/2}$ such that (N_t) is iid gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leq 0$.

- (X_t) causal solution to SRE, $X_t = A_t X_{t-1} + B_t$, $((A_t, B_t))$ positive iid and $((A, B))$ satisfies Kesten-Goldie theory then

$$\Theta_t \; = \; A_t \cdots A_1 \, , \quad t \geq 0,$$

and $\Theta_t \to 0$ a.s. since $\mathbb{E}[\log(A_1)] < 0$ holds.



We take $A_t = e^{N_t - 1/2}$ such that (N_t) is iid gaussian noise, and we follow Example 6.1. in Janßen and Segers (2014) where $\Theta_{-t} = A_{-t} \cdots A_{-1}$, for $t \leq 0$.

11.3. Extremal index of $(|\mathrm{X}_t|)$.

Let $(\boldsymbol{X}_t)$ satisfy the conditions of Theorem. For $x > 0$, denoting $M_n^{|\mathrm{X}|} = \max(|\mathrm{X}_1|, \ldots, |\mathrm{X}_n|)$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}(M_n \le x\, a_n) \\
&= \lim_{n\to\infty} \mathbb{P}\big(N_n\big((x,\infty)\big) = 0\big) \\
&= \mathbb{P}\big(N\big((x,\infty)\big) = 0\big) \\
&= \mathbb{P}\Big(\max_{i\ge 0} \Gamma_i^{-1/\alpha} \max_{j\in\mathbb{Z}} |\mathrm{Q}_{ij}| \le x\Big) \\
&= \mathbb{P}\Big(0 \le \max_{i\ge 0} \Gamma_i^{-1/\alpha} \big(x^{-1} \max_{j\in\mathbb{Z}} |\mathrm{Q}_{ij}|\big) \le 1\Big) \\
&= \exp\big(-x^{-\alpha} \mathbb{E}\big[\max_{j\in\mathbb{Z}} |\mathrm{Q}_j|^{\alpha}\big]\big)\,,
\end{aligned}$$

because $\boldsymbol{A} = \{\mathrm{x} \in \mathbb{R}^d : |\mathrm{x}| > x\}$ is a continuity set.

$\boldsymbol{\theta}_{|\mathrm{X}|} = \mathbb{E}[\max_{t\in\mathbb{Z}} |\mathrm{Q}_t|^{\alpha}] \in (0,1]$ **is the extremal index of** $(|\mathrm{X}_t|)$.

12. Large deviations Hult, Lindskog, Mikosch, Samorodnitsky (2005)

12.1. Heavy-tailed large deviations in $\mathbb{R}$, A.V. Nagaev (1969), S.V. Nagaev (1979), Cline, Hsing (1998).

- Z_n iid positive, regularly varying with index $\alpha > 0$,

  $S_n = (Z_1 + \cdots + Z_n) - b_n$, $b_n = n\mathbb{E}[Z]$ for $\alpha > 1$, $= 0$ for $\alpha < 1$.

- If $\lambda_n = \sqrt{a\, n\, \log n}$ and $a > \alpha - 2$, $\alpha > 2$,

$$\sup_{x \ge \lambda_n} \left| \frac{P(\lambda_n^{-1} S_n > x)}{n\, P(Z_1 > x)} - 1 \right| \to 0\,.$$

- Equivalently, for $\mu(x, \infty) = x^{-\alpha}$,

$$\sup_{x \ge 1} \left| \frac{P(\lambda_n^{-1} S_n \in (x, \infty))}{n\, P(Z_1 > \lambda_n)} - \mu(x\,, \infty) \right| \to 0\,.$$

- Analogous results exist for $\alpha \ge 2$ and general regularly varying Z_n.

12.2. Extensions to stationary sequences.

- Davis, Hsing (1995) **for certain mixing sequences, $\alpha < 2$**
- Mikosch, Samorodnitsky (2000) **for linear processes.**
- Konstantinides, Mikosch (2004) **for solution to stochastic recurrence equations**

$$X_t = A_t X_{t-1} + B_t ,$$

where (A_t, B_t) are iid pairs, B_t regularly varying with index $\alpha > 0$, $\mathbb{E}|A_1|^{\alpha+\epsilon} < \infty$.

- Mikosch, W. (2013) **for general Markov chains, including the stochastic recurrence equation where $\mathbb{E}[A_1^\alpha] = 1$.**

Large deviation and cluster index

Bartciewicz, Jakubowski, Mikosch, W. (2011), Mikosch, W. (2014), Buriticá, Mikosch, W. (2023+)

Assume

- (X_t) is an $\mathbb{R}_+$-valued regularly varying stationary process (X_t) with index $\alpha > 0$,
- the anti-clustering condition (AC),
- the vanishing-small-values condition (CS_1): for all $\delta > 0$

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{\mathbb{P}(\sum_{t=1}^{n}(X_t 1(X_t \le \varepsilon\lambda_n) - \mathbb{E}[X_t 1(X_t \le \varepsilon\lambda_n)] > \delta\,\lambda_n)}{n\,\mathbb{P}(\mathrm{X} > \lambda_n)} = 0\,.$$

Then the large deviation principle holds

$$\frac{P(\lambda_n^{-1} S_n > \lambda_n)}{n\,P(X_1 > \lambda_n)} \to c(1) > 0\,.$$

If $c(1) < \infty$ then $c(1) = \mathbb{E}\Big[\Big(\sum_{t\in\mathbb{Z}}|\mathrm{Q}_t|\Big)^{\alpha}\Big]$ is called the cluster index.

Remark that

- **If $c(1)$ is finite then $c(1) = \mathbb{E}[\|Q\|_1^\alpha]$ whereas $\theta_X = \mathbb{E}[\|Q\|_\infty^\alpha]$. The cluster index shares with the extremal index a similar simple expression with respect to (Q_t).**
- **$c(1) = \infty$ is possible when $\theta_X > 0$.**
- **The temporal dependence is responsible for:**

  - a negative dependence in the tail of the maxima and partial sums ($\alpha \leq 1$) that are smaller than in the iid case because $\theta_X \leq 1$, $\alpha > 0$, and $c(1) \leq 1$, $\alpha < 1$,

  - a positive dependence in the tail of the partial sums ($\alpha \geq 1$) that are larger than in the iid case becasue $c(1) \geq 1$.

12.3. Heavy-tailed large deviations for stochastic processes.

- $\mathbf{X}^n \in \mathbb{R}^d$ satisfies a large deviation principle if there exist $\gamma_n, \lambda_n \to \infty$ and a non-null Radon measure μ such that

$$\gamma_n \, P(\lambda_n^{-1}\mathbf{X}^n \in \cdot) \stackrel{v}{\to} \mu(\cdot)\,.$$

- $\mathbf{X}^n \in \mathbb{D} = (\mathbb{D}([0,1],\mathbb{R}^d), J_1)$ satisfies a large deviation principle if there exist $\gamma_n, \lambda_n \to \infty$ and a non-null boundedly finite measure m such that

$$\gamma_n \, P(\lambda_n^{-1}\mathbf{X}^n \in \cdot) \stackrel{\hat{w}}{\to} m(\cdot)\,.$$

- **This convergence can be expressed in terms of the finite-dimensional distributions and tightness analogous to regular variation of stochastic processes.** Hult and Lindskog (2005).
- **The continuous mapping theorem holds:**

$$\gamma_n P(h(\lambda_n^{-1}\mathbf{X}^n) \in \cdot) \xrightarrow{\hat{w}} m \circ h^{-1}(\cdot)\,.$$

for a.e. continuous mappings $h : \mathbb{D}\backslash\{0\} \to E$ ensuring that $h^{-1}(B)$ is bounded in $\mathbb{D}\backslash\{0\}$ for bounded $B \subset E$.

- **The temporal extremal dependence for stochastic processes is an active research topic** Basrak, Planinic and Soulier (2018), Soulier (2022)**.**

Example: Regularly varying random walks in $\mathbb{R}^d$

- $\mathbf{Z}_n \in \mathbb{R}^d$ iid regularly varying with index $\alpha > 0$ and limiting measure μ in $\mathbb{R}^d \backslash \{0\}$

$$\mathbf{S}_0 = 0\,, \quad \mathbf{S}_n = \mathbf{Z}_1 + \cdots + \mathbf{Z}_n\,, \quad n \geq 1\,.$$

- The corresponding random walk process in $\mathbb{D}$ (Donsker process)

$$\mathbf{S}^n(t) = \mathbf{S}_{[nt]}\,, \quad 0 \leq t \leq 1\,.$$

Assume $\lambda_n^{-1}\mathbf{S}_n \stackrel{P}{\to} 0$ and in addition

$$\lambda_n/\sqrt{n^{1+\gamma}} \to \infty \text{ for some } \gamma > 0 \text{ if } \alpha = 2$$

$$\lambda_n/\sqrt{n\,\log n} \to \infty \text{ if } \alpha > 2$$

- **Then in $\mathbb{D}\backslash\{0\}$**

$$\frac{P(\lambda_n^{-1}\mathbf{S}^n \in \cdot)}{n\,P(|\mathbf{Z}_1| > \lambda_n)} \stackrel{\hat{w}}{\rightarrow} m(\cdot)\,.$$

- **The measure m satisfies**

$$m\left(\{\mathbf{x} \in \mathbb{D} : \mathbf{x} = \mathbf{y}\,\mathbf{1}_{[v,1]}\,, v \in [0,1]\,, \mathbf{y} \in \mathbb{R}^d\backslash\{0\}\}^{\mathrm{C}}\right) = 0\,.$$

- **This supports the idea of heavy-tailed large deviation heuristics: The random walk $\mathbf{S}^n$ reaches the rare set $\lambda_n A \subset \mathbb{R}^d\backslash\{0\}$ by one jump due to exactly one extraordinarily large step size $\mathbf{Z}_i$.**

12.4. Ruin probabilities in $\mathbb{R}$.

- Z_n iid positive regularly varying with index $\alpha > 1$.
- Then for $c > 0$, $S_n = Z_1 + \cdots + Z_n - nEZ_1$, as $u \to \infty$, Embrechts, Veraverbeke (1982)

$$\psi_u = P\left(\sup_{n\geq 1}(S_n - c\,n) > u\right) \sim \frac{1}{c}\,\frac{1}{\alpha - 1}\,u\,P(Z_1 > u)\,.$$

- In an insurance context, the random walk with negative drift $S_n - c\,n$ describes the cash balance between arriving claims and linearly growing premium income.

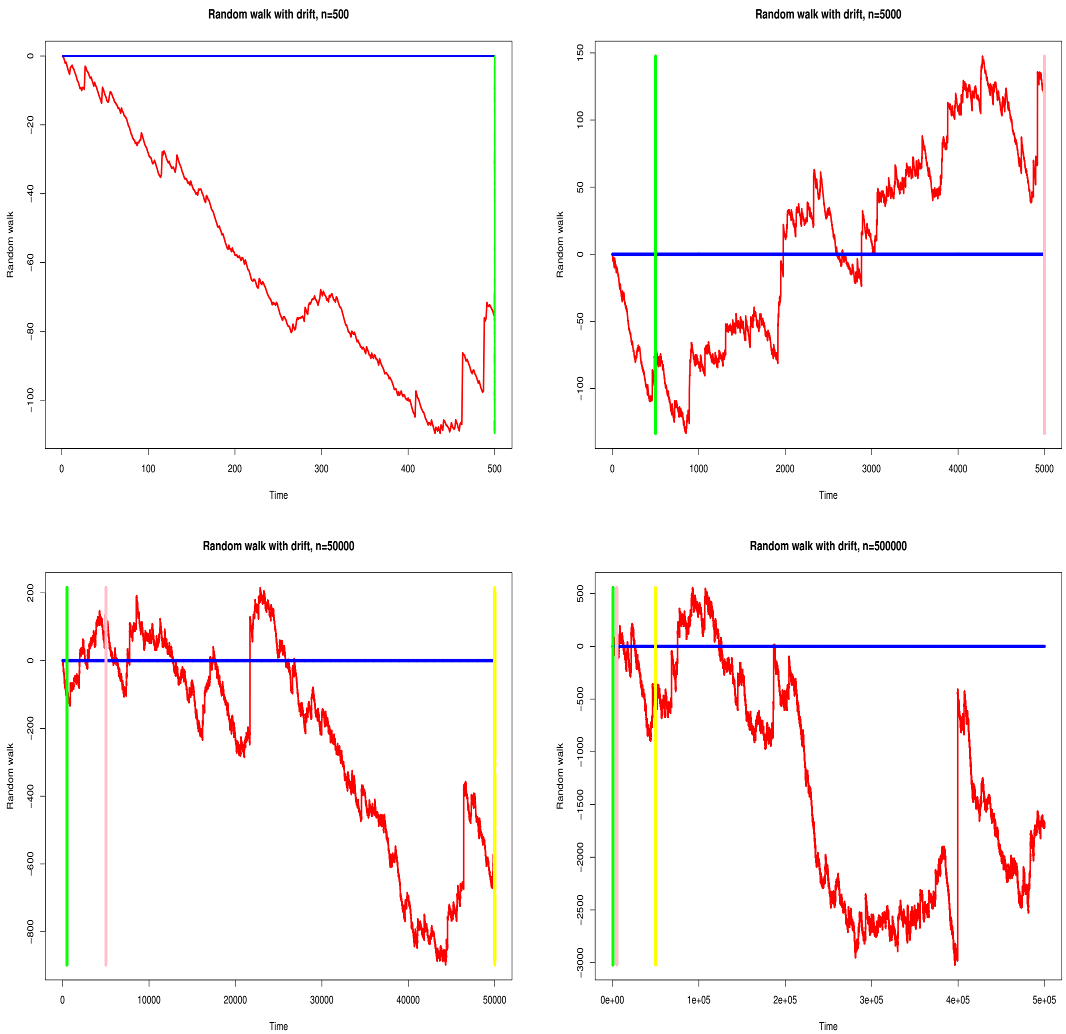


FIGURE 21. Random walk based on iid Pareto(2) step sizes (X_i) with expectation $\mathbb{E}[X] = 2$. The graphs show the random walk $(T_i)_{i=1,\dots,n}$ with negative drift based on the step sizes $Y_i = X_i - (\mathbb{E}[X] + \delta) = X_i - 2.005$ for different n. **Top:** $n = 5 \times 10^2$ (left), $n = 5 \times 10^3$ (right). **Bottom:** $n = 5 \times 10^4$ (left), $n = 5 \times 10^5$ (left).

Ruin probability and cluster index

Mikosch, W. (2014)

Assume

- (X_t) is an $\mathbb{R}_+$-valued regularly varying stationary process (X_t) with index $\alpha > 1$,
- the anti-clustering condition (AC),
- the vanishing-small-values condition (CS_1),
- $c(1) < \infty$.

Then for $c > 0$, $S_n = Z_1 + \cdots + Z_n - nEZ_1$, as $u \to \infty$,

$$\psi_u = P\left(\sup_{n\geq 1}(S_n - c\,n) > u\right) \sim \frac{c(1)}{c}\,\frac{1}{\alpha - 1}\,u\,P(Z_1 > u)\,.$$

In this setting we have $c(1) \geq 1$ and $c(1) = 1$ in the asymptotically independent case only.

13. The point process of exceedances

13.1. The point process with time stamps.

Consider the point process with time stamps

$$\overline{N}_n = \sum_{i=1}^{n} \varepsilon_{(a_n^{-1} X_i, i/n)}\,, \qquad n \geq 1\,,$$

Condition $\overline{\mathcal{A}}(a_n)$

For all sets $B_j \times A_j \subset \mathbb{R}^d \times (0,1]$, $j = 1,\ldots,k$, $k \geq 1$, such that $A_j = (s_j, t_j]$ for $0 \leq s_1 < t_1 \leq \cdots \leq s_k < t_k \leq 1$ and B_j is any finite union of rectangles of the type $(\mathrm{a}, \mathrm{b}]$ bounded away from zero, we have

$$\mathbb{E}\big[\mathrm{e}^{-\sum_{j=1}^{k} \overline{N}_n(B_j \times A_j)}\big] - \prod_{j=1}^{k} \mathbb{E}\big[\mathrm{e}^{-\overline{N}_n(B_j \times A_j)}\big] \to 0\,, \quad n \to \infty\,.$$

Convergence of the point process with time stamps

We assume the conditions of the point process convergence Theorem and $\overline{\mathcal{A}}(a_n)$. Then, on the state space $\mathbb{R}_{\mathbf{0}}^d \times (0,1]$, we have

$$\overline{N}_n \stackrel{d}{\to} \overline{N} = \sum_{i=1}^{\infty} \sum_{j=-\infty}^{\infty} \varepsilon_{(\Gamma_i^{-1/\alpha} \mathrm{Q}_{ij}, U_i)}$$

where (U_i) is an iid $U(0,1)$ sequence independent of $(\Gamma_i)_{i\geq 1}$, $(\mathrm{Q}_{ij})_{i\geq 1, j\in\mathbb{Z}}$.

Remark that the limiting mean measure is diffuse:

$$\mu \otimes \Lambda(\{(\mathrm{x}, t)\}) = 0.$$

13.2. The point process of exceedances.

Following Hsing (1993), for $x > 0$ consider the point process of exceedances with state space $(0, 1]$:

$$\eta_{n,x}(\cdot) := \overline{N}_n\big(\{\mathrm{y} : |\mathrm{y}| > x\} \times \cdot\big) \;=\; \sum_{i=1}^{n} \varepsilon_{i/n}(\cdot)\, \mathbf{1}(|\mathrm{X}_i| > x\, a_n)\,.$$

Under the previous assumptions, using a continuity argument we obtain

$$\begin{aligned}\eta_{n,x}(\cdot) \;\stackrel{d}{\to}\; \eta_x(\cdot) \;&:= \overline{N}\big(\{\mathrm{y} : |\mathrm{y}| > x\} \times \cdot\big)\\ &= \sum_{i=1}^{\infty} \sum_{j\in\mathbb{Z}} \mathbf{1}\big(\Gamma_i^{-1/\alpha} |\mathrm{Q}_{ji}| > x\big)\, \varepsilon_{U_i}(\cdot)\,.\end{aligned}$$

Then

$$\mathbb{E}\Big[\exp\Big(-\int f(u)\eta_x(u)\Big)\Big]$$
$$= \exp\Big(-\int_0^1\int_0^\infty \mathbb{E}\big[1-\mathrm{e}^{-f(u)\sum_{j=-\infty}^{\infty} 1(y|Q_j|>x)}\big]\,du\,d(-y^{-\alpha})\Big)$$
$$= \exp\Big(-x^{-\alpha}\int_0^1\int_0^\infty \mathbb{E}\big[1-\mathrm{e}^{-f(u)\sum_{j=-\infty}^{\infty} 1(y|Q_j|>1)}\big]\,du\,d(-y^{-\alpha})\Big).$$

We recognize the Laplace tranform of a compound Poisson process.

One can rewrite the limit as

$$\eta_x(t) = \sum_{i=1}^{N_x(t)} \xi_i\,, \qquad 0 < t \le 1\,,$$

where

- N_x **is a Poisson process on** $(0,1]$ **with intensity** $x^{-\alpha}$**,**
- **for an iid sequence** (Y_i) **of Pareto(**α**)-distributed random variables which is also independent of** (Q_i)**,**

$$\xi_i = \sum_{j\in\mathbb{Z}} 1(Y_i\,|\mathrm{Q}_{ij}| > 1)\,,$$

- N_x**,** (ξ_i) **are independent.**

13.3. Probability of exceedances.

We deduce the relations, using the order statistics $|\mathbf{Q}|_{(1)} \geq |\mathbf{Q}|_{(2)} \geq \cdots$ of the tail cluster process,

- $\mathbb{P}(\xi_1 > 0) = \mathbb{P}(Y \max_{t\in\mathbb{Z}} |\mathbf{Q}_t| > 1) = \mathbb{E}[|\mathbf{Q}|_{(1)}^{\alpha}] = \theta_{|\mathbf{X}|}$,
- $\mathbb{P}(\xi_1 = j) = \mathbb{E}[|\mathbf{Q}|_{(j)}^{\alpha} - |\mathbf{Q}|_{(j+1)}^{\alpha}] =: \quad \pi_j$.

The expression of the statistic π_j, $j \geq 1$, in terms of Q is simple.

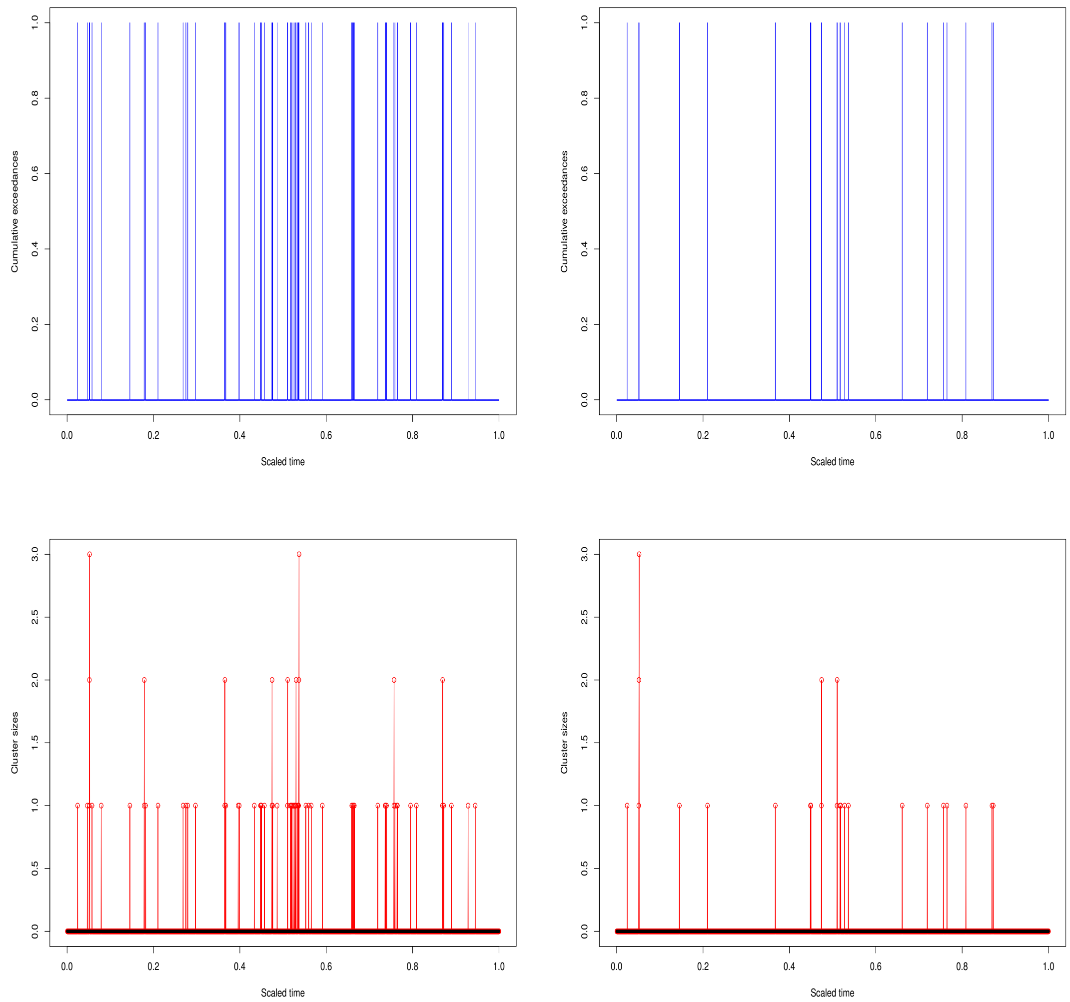


Figure 22. We consider the log-returns of the Bit Coin USD stock prices from 17 September 2014 until 8 January 2021. **Top:** Graphs $(i/n, \mathbf{1}(|X_i| > q))$, $i = 1, \ldots, n$, for the 97% **(left)** and 99% **(right)** empirical quantiles of the absolute values of the sample. The extremal clusters of high level exceedances are well visible. **Bottom:** The corresponding graphs for the cluster lengths of the exceedances of the 97% **(left)** and 99% **(right)** empirical quantiles of the absolute values of the sample.

In practice we only observe the ξ_i given that $\xi_i > 0$. We have

$$\begin{aligned}\mathbb{E}[\xi_1 \mid \xi_1 > 0] &= \sum_{j\in\mathbb{Z}} \mathbb{P}(Y_i\,|\mathrm{Q}_{1j}| > 1 \mid \xi_1 > 0) \\ &= \frac{\mathbb{E}[\sum_{j\in\mathbb{Z}} |\mathrm{Q}_{1j}|^\alpha]}{\mathbb{P}(\xi_1 > 0)} \\ &= \frac{1}{\theta_{|\mathrm{X}|}}\,.\end{aligned}$$

Similarly

$$\mathbb{P}(\xi_1 = j \mid \xi_1 > 0) = \frac{\pi_j}{\theta_{|\mathrm{X}|}}\,.$$

The statistic $\pi_j/\theta_{|\mathrm{X}|}$ can be understood as the probability of recording a cluster of length j.

14. Cluster inference

We observe a sample $(X_1, \dots, X_n)$ from a stationary regularly varying time series (X_t) of order $\alpha > 0$.

14.1. Large deviations of ℓ^α-blocks Buriticá, Mikosch, W. (2023+).

For inference purposes, let $r_n \to \infty$, $k_n := \lceil n/r_n \rceil \to \infty$.

$$X_{[1,n]} = (\underbrace{X_{[1,r_n]}}_{\mathcal{B}_{1,r_n}}, \underbrace{X_{[r_n+1,2r_n]}}_{\mathcal{B}_{2,r_n}}, \dots, \underbrace{X_{[n-r_n+1,n]}}_{\mathcal{B}_{k_n,r_n}}).$$

In the following we use a Peak Over Threshold method over blocks with large

$$\|\mathcal{B}_{j,r_n}\|_\alpha = \Big(\sum_{t=jr_n+1}^{(j+1)r_n} |X_t|^\alpha \Big)^{1/\alpha}.$$

Large deviations for ℓ^α-norms of blocks

Assume (X_t) and (x_n) satisfy **AC**, $\mathbf{CS}_\alpha$,

$$\lim_{\epsilon\to 0}\limsup_{n\to\infty}\frac{\mathbb{P}(\sum_{t=1}^n(|X_t|^\alpha \mathbf{1}(|X_t|^\alpha \le \varepsilon x_n^\alpha) > \delta\, x_n^\alpha)}{n\,\mathbb{P}(X>\lambda_n)} = 0\,,$$

and $\mathbb{P}(\|\mathcal{B}_{1,r_n}\|_\alpha > x_{r_n}) \to 0$.

- Then,

$$\mathbb{P}(\|\mathcal{B}_{1,r_n}\|_\alpha > x_{r_n})/(r_n\mathbb{P}(|X_0| > x_{r_n})) \to c(\alpha),$$

where $c(\alpha) = \mathbb{E}[\|Q\|_\alpha^\alpha] = 1$.

- **Moreover,**

$$\mathbb{P}(\|\mathcal{B}_{1,r_n}\|_\alpha > y\, x_{r_n}, \|\mathcal{B}_{1,r_n}\|_\alpha^{-1}\mathcal{B}_{1,r_n} \in \cdot \mid \|\mathcal{B}_{1,r_n}\|_\alpha > x_{r_n})$$
$$\to y^{-\alpha}\,\mathbb{P}(\mathrm{Q} \in \cdot)\,, \qquad n \to \infty\,,$$

and the convergence holds for a family of shift-invariant ℓ^α-continuity sets.

14.2. Bias - variance analysis.

Let $r_n \to \infty$, $k_n := \lceil n/r_n \rceil \to \infty$.

$$\mathbf{X}_{[1,n]} = (\underbrace{\mathbf{X}_{[1,r_n]}}_{\mathcal{B}_{1,r_n}}, \underbrace{\mathbf{X}_{[r_n+1,2r_n]}}_{\mathcal{B}_{2,r_n}}, \cdots, \underbrace{\mathbf{X}_{[n-r_n+1,n]}}_{\mathcal{B}_{k_n,r_n}}).$$

Aim:
Infer $\mathbb{E}[f(Y\mathbf{Q})]$ for suitable cluster functionals $f : \ell^\alpha \to \mathbb{R}$.

We propose to estimate the statistic $f^{\mathbf{Q}} = \mathbb{E}[f(Y\mathbf{Q})]$ by

$$\widehat{f^{\mathbf{Q}}} := \frac{1}{m}\sum_{t=1}^{k_n} f\big(\mathcal{B}_t/\|\mathcal{B}_t\|_{\alpha,(m+1)}\big)\mathbf{1}(\|\mathcal{B}_t\|_\alpha > \|\mathcal{B}_t\|_{\alpha,(m+1)}),$$

where $\|\mathcal{B}\|_{\alpha,(1)} \geq \cdots \geq \|\mathcal{B}\|_{\alpha,(k_n)}$ and $m = m_n \to \infty$.

Asymptotic normality Buriticá, W. (2023+), Kulik, Soulier (2020), Cissokho, Kulik (2021)

Assume AC, CS$_\alpha$, and further mixing and bias conditions. There exists $m = m_n \to \infty$, $k_n/m_n \to \infty$, such that for suitable $f : \ell^\alpha \to \mathbb{R}$,

$$\sqrt{m}(\widehat{f^{Q}} - f^{Q}) \xrightarrow{d} \mathcal{N}\big(0, \mathrm{var}(f(YQ))\big),$$

with $m := m_n = \lfloor k_n \mathbb{P}(\|\mathcal{B}_{1,r_n}\|_\alpha > x_{r_n}) \rfloor$.

(1) The asymptotic variance $\mathrm{var}(f(Y\mathbf{Q}))$ can be degenerate to 0 for simple spectral tail processes $(\mathbf{Q}_t)$,

(2) We promote the use of order statistics of α-norm blocks such that

$$\|\mathcal{B}\|_{\alpha,(m)}/x_{r_n} \xrightarrow{\mathbb{P}} 1.$$

where $m_n = \lceil k_n \mathbb{P}(\|\mathcal{B}_{1,r_n}\|_\alpha > x_{r_n}) \rceil$,

(3) The α-cluster approach allows to choose r_n achieving a good bias-variance trade-off using the estimator $\widehat{f^{\mathbf{Q}}}$ and the estimation of the asymptotic variance $\mathrm{var}(f(Y\mathbf{Q}))$.

Heuristic on the number of extreme blocks:

For iid sequence, using the single big jump principle

$$m_n = \left\lceil k_n \mathbb{P}(\|\mathcal{B}_{1,r_n}\|_\alpha > x_{r_n}) \right\rceil \sim n\mathbb{P}(|\mathbf{X}_0| > x_{r_n})\,.$$

By the large deviation principle (because $c(\alpha) = 1$!) we also have

$$m_n \sim k_n \mathbb{P}(\|\mathcal{B}_{1,r_n}\|_\alpha > x_{r_n}) \sim n\mathbb{P}(|\mathbf{X}_0| > x_{r_n}).$$

The tuning parameter r_n does not dependent on the underlying time dependencies within the ℓ^α-block!

14.3. Cluster-based inference Buriticá, W. (2023+).

Extremal index inference.

If $f : (\mathrm{x}_t) \mapsto \|(\mathrm{x}_t)\|_\infty^\alpha / \|(\mathrm{x}_t)\|_\alpha^\alpha$, **then,**

$$\widehat{f^{\mathrm{Q}}} = \quad \mathbb{E}[\|\mathrm{Q}\|_\infty^\alpha] \quad = \theta_{|\mathrm{X}|}.$$

New estimator of the extremal index based on extremal ℓ^α**-blocks:**

$$\widehat{\theta}_{|\mathrm{X}|,\alpha} \;=\; \tfrac{1}{m}\sum_{t=1}^{k_n} \tfrac{\|\mathcal{B}_t\|_\infty^\alpha}{\|\mathcal{B}_t\|_\alpha^\alpha} \mathbf{1}(\|\mathcal{B}_t\|_\alpha > \|\mathcal{B}\|_{\alpha,(m+1)}),$$

For linear processes, $\|\mathrm{Q}\|_\infty$ **is deterministic and the asymptotic variance is null**

$$\mathrm{var}(f(Y\mathrm{Q})) = \mathrm{var}(\|\mathrm{Q}\|_\infty^\alpha) = 0\,.$$

Recall the relation

$$(\mathbb{E}[\sum_{t\in\mathbb{Z}} \mathbf{1}(|Y\,\mathbf{Q}^{(\infty)}| > 1)])^{-1} \quad = \quad \theta_{|\mathrm{X}|}\,,$$

yielding the estimator, for large threshold u,

$$\widetilde{\theta}_{|\mathrm{X}|,\infty}(u) = \frac{\sum_{t=1}^{k_n} \mathbf{1}(\|\mathcal{B}_t\|_\infty > u)}{\sum_{t=1}^{n} \mathbf{1}(|\mathrm{X}_t| > u)}.$$

The so-called blocks estimator of Hsing (1993) is defined letting

$u = |\mathrm{X}|_{(m+1)}$.

We use a better variant replacing the threshold $u = |\mathrm{X}|_{(m+1)}$ with

$u = \|\mathcal{B}_t\|_{\infty,(m+1)}$:

$$\widehat{\theta}_{|\mathrm{X}|,\infty} \;=\; \Big(\frac{1}{m}\sum_{t=1}^{n}\mathbf{1}\big(|\mathrm{X}_t| > \|\mathcal{B}_t\|_{\infty,(m+1)}\big)\Big)^{-1}.$$

Cluster index inference.

If $f : (\mathbf{x}_t) \mapsto \|(\mathbf{x}_t)\|_1^\alpha / \|(\mathbf{x}_t)\|_\alpha^\alpha$, then,

$$\widehat{f^{\mathbf{Q}}} = \quad \mathbb{E}[\|\mathbf{Q}\|_1^\alpha] \quad = c(1).$$

New estimator of the extremal index based on extremal ℓ^α-blocks:

$$\widehat{c(1)}_\alpha = \frac{1}{m} \sum_{t=1}^{k_n} \frac{\|\mathbf{B}_t\|_1^\alpha}{\|\mathbf{B}_t\|_\alpha^\alpha} \mathbb{1}\big(\|\mathbf{B}_t\|_\alpha > \|\mathbf{B}\|_{\alpha,(m+1)}\big) \,.$$

Using a version of Cissokho and Kulik (2021) with a block dependent threshold we also consider the alternative

$$\widehat{c(1)}_\infty = \frac{\sum_{t=1}^{k_n} \mathbb{1}\big(\|\mathbf{B}_t\|_1 > \|\mathbf{B}\|_{\infty,(m+1)}\big)}{\sum_{t=1}^{n} \mathbb{1}\big(|\mathbf{X}_t| > \|\mathbf{B}\|_{\infty,(m+1)}\big)} \,.$$

14.4. Numerical experiments.

Simulation setup

- **We simulate 1 000 AR(1) trajectories $(X_t)_{t=1,\dots,n}$, $X_t = \varphi X_{t-1} + Z_t$, for $n = 3\,000$.**
- **We fix $m = m_n = \lceil n/r_n^2 \rceil$ and we use that $m_n \sim n/r_n \mathbb{P}(\|\mathcal{B}_1\|_\alpha > x_{r_n}) = o(n/r_n)$.**
- **The α-cluster based approach requires the estimation of α. We use the estimator from de Haan, Mercadier, Zhou (2016).**

Simulation results

Extremal index inference.

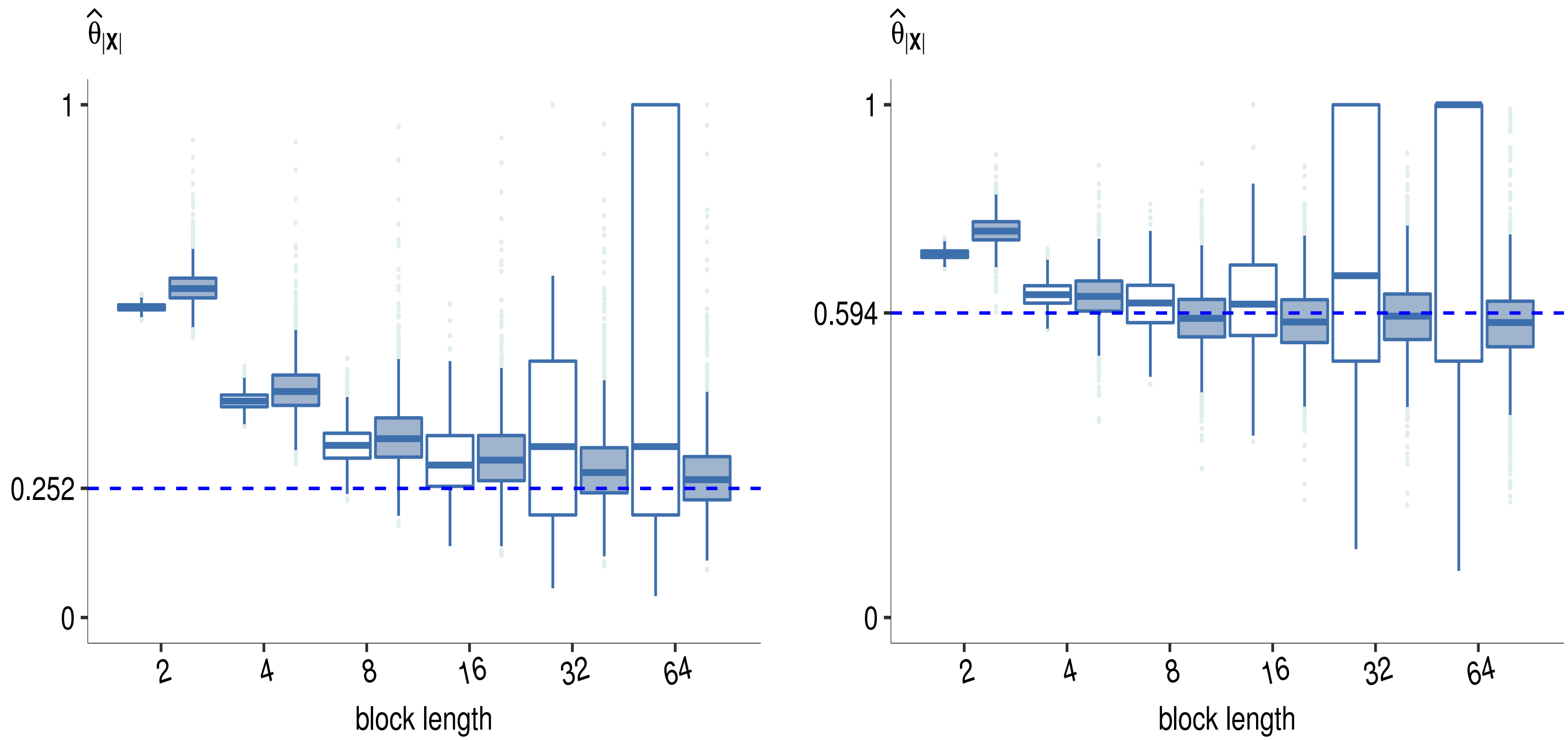


FIGURE 23. Boxplots of estimates $\widehat{\boldsymbol{\theta}}_{|\mathbf{X}|,\widehat{\alpha}}$ (blue) and $\widehat{\boldsymbol{\theta}}_{|\mathbf{X}|,\infty}$ (white), from observations $(\mathbf{X}_t)_{t=1,\dots,n}$ from a causal AR(1) model with student$(\boldsymbol{\alpha})$ noise, $\boldsymbol{\alpha} = \mathbf{1.3}$ and $\boldsymbol{\varphi} = \mathbf{0.8}$ (left), $\boldsymbol{\varphi} = \mathbf{0.5}$ (right), such that $\boldsymbol{n} = \mathbf{3\,000}$.

Cluster index inference.

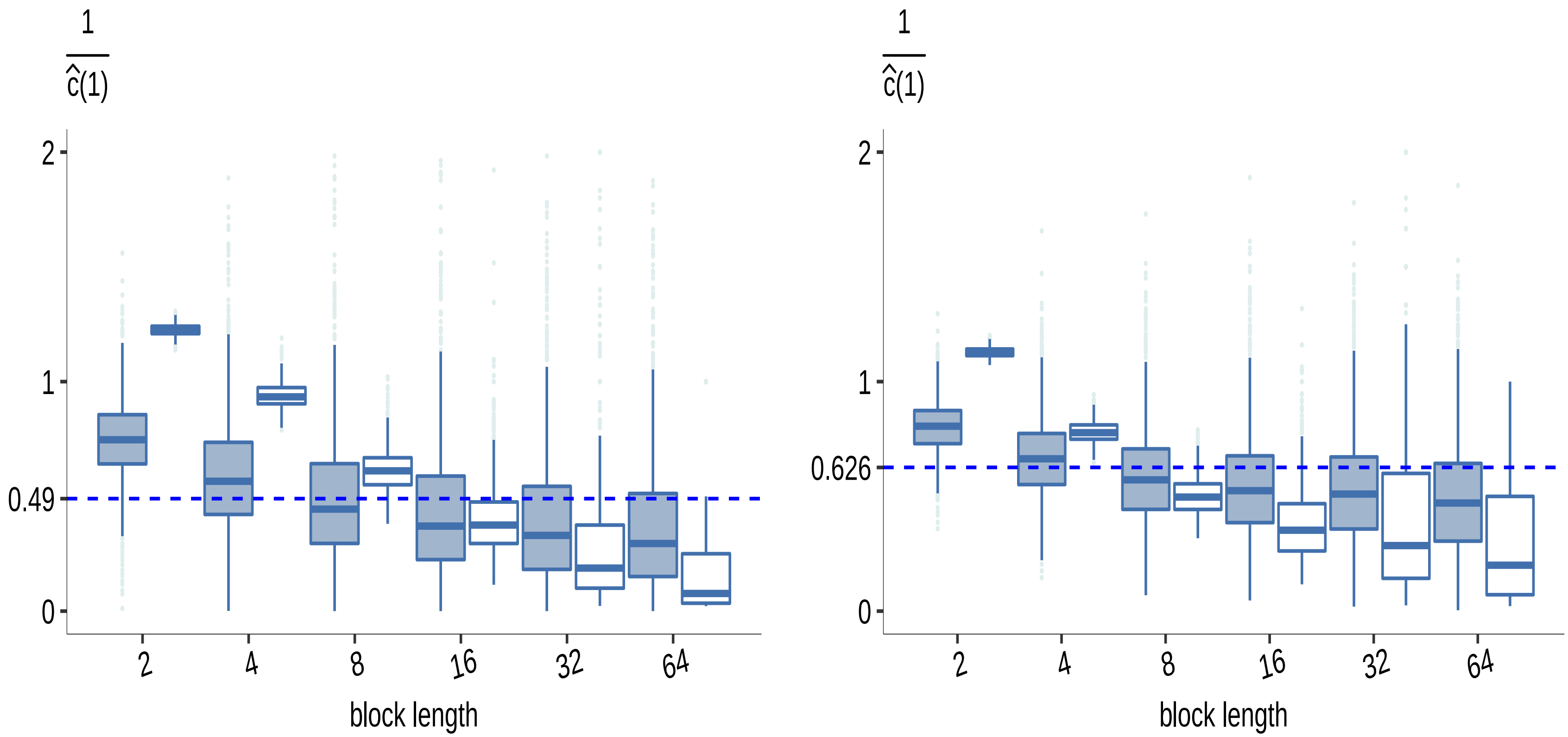


FIGURE 24. Boxplots of estimates $\mathbf{1/\widehat{c(1)}_{\widehat{\alpha}}}$ (blue) and $\mathbf{1/\widehat{c(1)}_{\infty}}$ (white), from observations $\mathbf{(X_t)_{t=1,\dots,n}}$ from a causal AR(1) model with student$\boldsymbol{(\alpha)}$ noise, $\boldsymbol{\alpha = 1.3}$ and $\boldsymbol{\varphi = 0.8}$ (left), $\boldsymbol{\varphi = 0.5}$ (right), such that $\boldsymbol{n = 3\,000}$.

15. Conclusions

- **Motivated by risk analysis it is mandatory to better understand extremal dependence in time.**
- **Recent advances in applied probability clarify the objects of interest.**
- **A book in collaboration with T. Mikosch is almost finished...**
- **Many statistical problems remain open!**

Thanks for your attention!