Extremes of Stationary Time Series – Day 1.

Dr. Ioannis Papastathopoulos (with illustrations by Lambert de Monte)

February 27, 2023

Introduction

- In basic data analysis we often assume that observations are independent or even independent and identically distributed,

$$X_1, \ldots, X_n \overset{\text{ind}}{\sim} F_1, \ldots, F_n, \qquad X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2),$$

- Time series is the study of observations that arise in some order (usually time) and so are dependent: in other words

one d*** thing after another!**

- There are many more ways to be dependent than to be independent, and almost all data are collected in time order, so time series arise in a vast range of disciplines: climatology; economics; finance; marketing; epidemiology; biomedicine; genomics; environmental science; computer science; electrical engineering; physics; ...

Definition.
(a) A **stochastic process** $\{X_t\}_{t\in\mathcal{T}}$ with index set $\mathcal{T}$ is a family of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$.
(b) A **realisation** of $\{X_t\}$ is the outcome $\{x_t\}_{t\in\mathcal{T}} = \{X_t(\omega)\}_{t\in\mathcal{T}}$ for some $\omega \in \Omega$.
About the index set

- In time series usually $\mathcal{T} = \mathbb{R}, \mathbb{R}_+$ or $\mathbb{Z}$;
- owing to digitisation, $\mathcal{T}$ cannot in practice contain a sub-interval of $\mathbb{R}$, but the time step Δt can be very small in some applications;
- (almost-)continuous time series can be thinned by subsampling at the points of a grid, or, in some cases, by integration over intervals of width h (e.g., rainfall data);
- for general discussion, we take $\mathcal{T} = \mathbb{Z}$, so that X_t is recorded at times $0, \pm 1, \pm 2, \ldots$, and write a realisation of the process observed for finite period as $x_1, \ldots, x_n$

- We will explore the impact of dependence between values in the series on the extreme values when observed at long-range and short-range.

- Generally, one considers how dependence affects:
  - $M_n = \max(X_1, \ldots, X_n)$;
  - exceedances of a high threshold.

- We will explore the impact of dependence between values in the series on the extreme values when observed at long-range and short-range.

- Generally, one considers how dependence affects:
  - $M_n = \max(X_1, \ldots, X_n)$;
  - exceedances of a high threshold.

- We will explore the impact of dependence between values in the series on the extreme values when observed at long-range and short-range.

- Generally, one considers how dependence affects:
  - $M_n = \max(X_1, \ldots, X_n)$;
  - exceedances of a high threshold.

Stationarity and stationary sequences

- A process $\{X_t\}$ is said to be a stationary process if the joint distributions of

$$(X_{t_1}, \dots, X_{t_k}) \quad \text{and} \quad (X_{t_1+\tau}, \dots, X_{t_k+\tau})$$

are the same for any k, $t_1, \dots, t_k$, and τ.

- Throughout we will assume that the univariate marginal distribution function is F, *i.e.*,

$$F(x) = \mathbb{P}(X_t \leq x)$$

for all t.

Long-range asymptotic independence condition

We suppose that there exist normalising sequences $a_n > 0$ and b_n such that

$$\frac{M_n - b_n}{a_n}$$

has a non-degenerate limit distribution.

We want to characterise the limit behaviour of M_n.

Recall: $M_n = \max(X_1, \dots, X_n)$.

In the absence of any conditions that limit the amount of long-range dependence that can be present in the values of the series, any limit distribution can be obtained.

For example let $X_t = X_1$ for all t, then for all n,

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x) = F(x),$$

so the class of limit distributions covers all distributions.

- In the absence of any conditions that limit the amount of long-range dependence that can be present in the values of the series, any limit distribution can be obtained.

- For example let $X_t = X_1$ for all t, then for all n,

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x) = F(x),$$

so the class of limit distributions covers all distributions.

Asymptotic independence of Maxima (AIM)

- Let $M_{i,j} = \max(X_i, \ldots, X_j)$ and $u_n = a_n x + b_n$ for a_n, b_n defined above and x any real number.

- **AIM**(u_n) **condition**: there exists a sequence q_n of positive integers with $q_n = o(n)$ such that for all i and j

$$|\mathbb{P}(M_{1,i} \leq u_n, M_{i+q_n,i+q_n+j} \leq u_n) - \mathbb{P}(M_{1,i} \leq u_n)\mathbb{P}(M_{1,j} \leq u_n)| \to 0,$$

as $n \to \infty$.

- The condition ensures that separated groups of extreme points become increasingly close to being independent as their separation and level both increase at appropriate rates.

See Leadbetter et al. [1983] and O'Brien [1987] for other similar conditions.

Unified Extremal Types Theorem

Theorem Suppose that there exist normalising sequences a_n and b_n such that

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \to H(x), \quad n \to \infty,$$

where H is a non-degenerate distribution, and that the $\mathrm{AIM}(a_n x + b_n)$ condition holds, then H is of the same type[1] as

$$\exp[-(1+\xi x)_+^{-1/\xi}],$$

i.e. a GEV limit distribution.

[1]Recall: We say that distribution functions F_1 and F_2 are of the same type if there exist constants $a > 0$ and b such that $F_2(ax+b) = F_1(x)$ for all x.

Outline strategy of proof

1. Partition time into a series of alternating long and short blocks, k of each.

2. Let the numbers in long and short blocks be r_n and q_n respectively with $r_n \to \infty$, $q_n \to \infty$ and $q_n/r_n \to 0$.

3. In a neighbouring pair of long and short blocks, the maximum over the two blocks is likely to be in the long block.

4. All maxima tend to fall into the long blocks.

5. Long block maxima are approximately independent as they are separated in time by the short blocks, so satisfy $\mathrm{AIM}(u_n)$ condition.

6. Long block maxima satisfy the max-stability property, and hence are GEV.

1. Partition time into a series of alternating long and short blocks, k of each.

2. Let the numbers in long and short blocks be r_n and q_n respectively with $r_n \to \infty$, $q_n \to \infty$ and $q_n/r_n \to 0$.

3. In a neighbouring pair of long and short blocks, the maximum over the two blocks is likely to be in the long block.

4. All maxima tend to fall into the long blocks.

5. Long block maxima are approximately independent as they are separated in time by the short blocks, so satisfy AIM(u_n) condition.

6. Long block maxima satisfy the max-stability property, and hence are GEV.

1. Partition time into a series of alternating long and short blocks, k of each.
2. Let the numbers in long and short blocks be r_n and q_n respectively with $r_n \to \infty$, $q_n \to \infty$ and $q_n/r_n \to 0$.
3. In a neighbouring pair of long and short blocks, the maximum over the two blocks is likely to be in the long block.
4. All maxima tend to fall into the long blocks.
5. Long block maxima are approximately independent as they are separated in time by the short blocks, so satisfy AIM(u_n) condition.
6. Long block maxima satisfy the max-stability property, and hence are GEV.

1. Partition time into a series of alternating long and short blocks, k of each.
2. Let the numbers in long and short blocks be r_n and q_n respectively with $r_n \to \infty$, $q_n \to \infty$ and $q_n/r_n \to 0$.
3. In a neighbouring pair of long and short blocks, the maximum over the two blocks is likely to be in the long block.
4. All maxima tend to fall into the long blocks.
5. Long block maxima are approximately independent as they are separated in time by the short blocks, so satisfy AIM(u_n) condition.
6. Long block maxima satisfy the max-stability property, and hence are GEV.

Outline strategy of proof

1. Partition time into a series of alternating long and short blocks, k of each.
2. Let the numbers in long and short blocks be r_n and q_n respectively with $r_n \to \infty$, $q_n \to \infty$ and $q_n/r_n \to 0$.
3. In a neighbouring pair of long and short blocks, the maximum over the two blocks is likely to be in the long block.
4. All maxima tend to fall into the long blocks.
5. Long block maxima are approximately independent as they are separated in time by the short blocks, so satisfy AIM(u_n) condition.
6. Long block maxima satisfy the max-stability property, and hence are GEV.

1. Partition time into a series of alternating long and short blocks, k of each.
2. Let the numbers in long and short blocks be r_n and q_n respectively with $r_n \to \infty$, $q_n \to \infty$ and $q_n/r_n \to 0$.
3. In a neighbouring pair of long and short blocks, the maximum over the two blocks is likely to be in the long block.
4. All maxima tend to fall into the long blocks.
5. Long block maxima are approximately independent as they are separated in time by the short blocks, so satisfy AIM(u_n) condition.
6. Long block maxima satisfy the max-stability property, and hence are GEV.

- At a practical level this is all we need, the maximum of a stationary sequence that has some independence at long-range follows a GEV distribution.

- The result does not show how the dependence changes the behaviour of M_n.

- **We separate marginal and dependence features** for the remainder and subsequently fix the marginal distribution F of the $\{X_t\}$ to be Fréchet.

Extremes of Fréchet marginal variables

Let $\{X_t\}$ be a stationary process which has Fréchet marginal distributions, *i.e.*

$$F(x) = \exp(-1/x) \text{ for } x > 0.$$

Let $\widehat{M}_n$ denote the maximum of n IID variables with marginal distribution F, then

$$\mathbb{P}\left(\frac{\widehat{M}_n}{n} \leq x\right) = \{F(nx)\}^n = \exp(-1/x) = G(x),$$

i.e., $\widehat{M}_n/n$ has a Fréchet limit distribution for all n and we denote the limit distribution of the maximum of the IID variables by G.

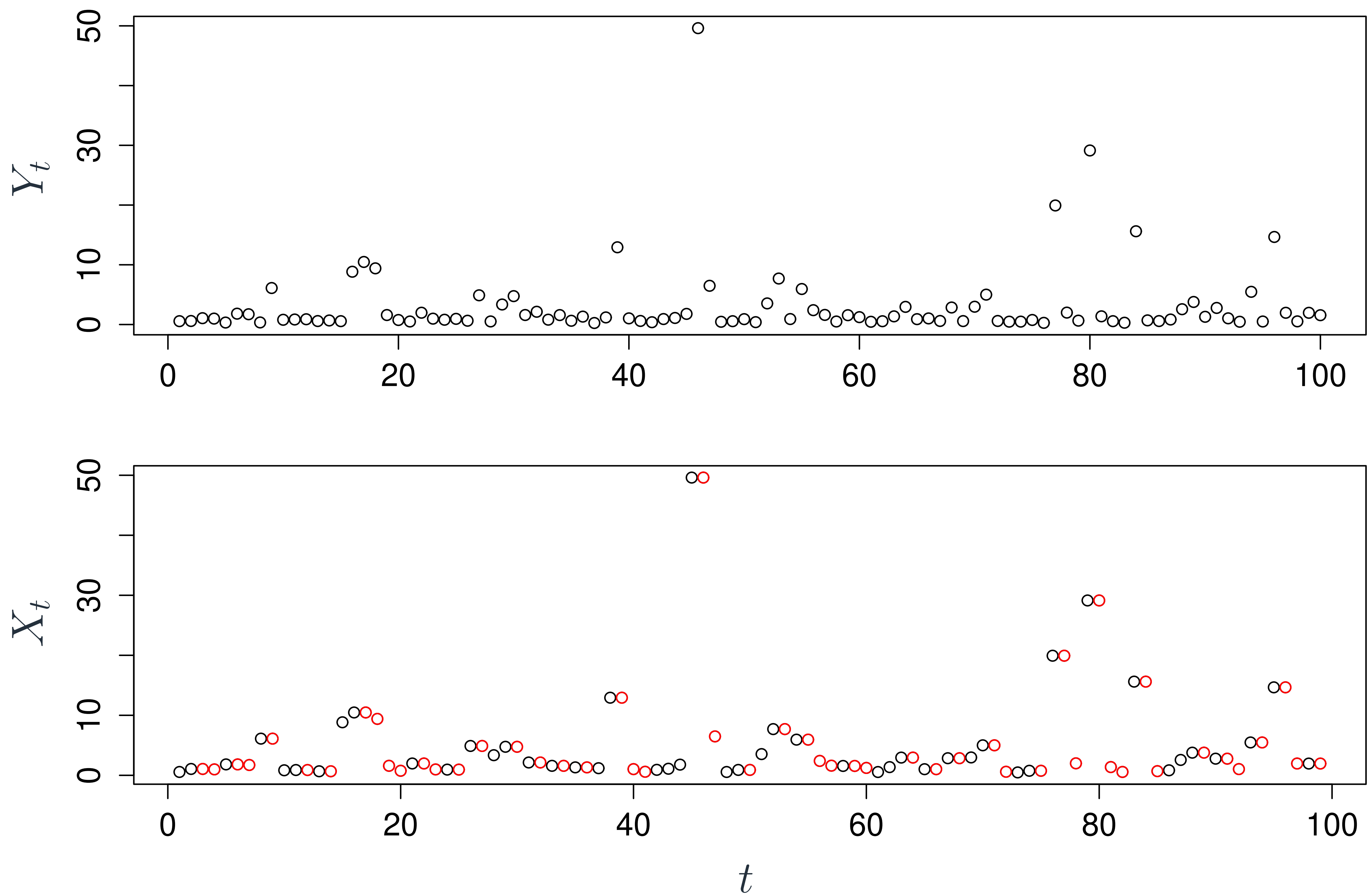
Suppose that $M_n = \max(X_1, \ldots, X_n)$, that $\{X_t\}$ satisfies the AIM(nx) condition and that

$$\mathbb{P}\left(\frac{M_n}{n} \leq x\right) \to H(x),$$

then we can assess the effect of dependence on M_n by looking at the difference between H and G.

Motivating example

Let $\{Y_t\}$ be IID with $F_Y(y) = \exp[-1/(2y)]$, *i.e.*, Fréchet type marginals. Define, for all t, $X_t = \max(Y_t, Y_{t-1})$.



Marginal properties

The marginal distribution of the $\{X_t\}$ process is:

$$
\begin{aligned}
\mathbb{P}(X_t \leq x) &= \mathbb{P}(Y_t \leq x, Y_{t-1} \leq x) \\
&= \{F_Y(x)\}^2 \\
&= \exp[-2/(2x)] \\
&= \exp(-1/x),
\end{aligned}
$$

i.e., Fréchet distributed.

Moving maxima dependence properties

- For $\tau > 1$, X_t is independent of $X_{t+\tau}$ as the associated Y variables are all different for each of the X's and hence the X's are independent.
- For $\tau = 1$, X_t is dependent with $X_{t+\tau}$ as both X_t and X_{t+1} are functions of Y_t.
- It is hence clear that long-range independence holds.
- Extreme values occur as groups or clusters of independent values, with two values per cluster for all the largest values.

$$\mathbb{P}\left(\frac{M_n}{n} \leq x\right) = \mathbb{P}(X_1 \leq nx, \ldots, X_n \leq nx) = \mathbb{P}(Y_0 \leq nx, Y_1 \leq nx, \ldots, Y_n \leq nx)$$

$$= [F_Y(nx)]^{n+1} = [\exp\{-1/(2nx)\}]^{n+1}$$

$$= \exp\{-(n+1)/(2nx)\} \to \exp\{-1/(2x)\} \qquad \text{as } n \to \infty$$

$$= \{G(x)\}^{1/2} = H(x).$$

Extremal index

- The previous example shows that short-range dependence in the extreme values of the process affects the limiting distribution of M_n.

- First we present a measure of short-range extremal dependence, termed the extremal index.

- Define the extremal index θ, for variables with Fréchet marginals, by

$$\theta = \lim_{n\to\infty} \mathbb{P}(M_{2,p_n} \leq n | X_1 > n)$$

$$= \lim_{n\to\infty} \{1 - \mathbb{P}(M_{2,p_n} > n | X_1 > n)\}$$

where $p_n = o(n)$.

- $0 \leq \theta \leq 1$.
- Larger values of θ correspond to weaker short-range extremal dependence.
- For an IID process $\theta = 1$ as

$$\begin{aligned} \theta &= \lim_{n\to\infty} \{F(n)\}^{p_n} \\ &= \lim_{n\to\infty} \exp(-p_n/n) \\ &= 1, \end{aligned}$$

  since $p_n/n \to 0$.
- If a process is independent for all lags $\tau \geq m$, for some finite m then $1/m \leq \theta \leq 1$.

Extremal index for moving maxima example

- From the realisation of the process it is clear that $\theta = 1/2$.
- A derivation is

$$\begin{aligned}
\mathbb{P}(M_{2,p_n} \le n | X_1 > n) &= \mathbb{P}(Y_1 \le n, \ldots, Y_{p_n} \le n | \max(Y_0, Y_1) > n) \\
&\approx \mathbb{P}(Y_1 \le n, \ldots, Y_{p_n} \le n | Y_0 > n)\frac{1}{2} + \mathbb{P}(Y_1 \le n, \ldots, Y_{p_n} \le n | Y_1 > n)\frac{1}{2} \\
&\to 1 \times \frac{1}{2} + 0 \times \frac{1}{2}, \quad n \to \infty \\
&= \frac{1}{2} = \theta.
\end{aligned}$$

Combining the above results and examples, the following result should appear to be quite natural.

For variables with Fréchet marginal distributions, provided that

$$\mathbb{P}\left(\frac{M_n}{n} \leq x\right) \to H(x), \quad n \to \infty$$

to a non-degenerate limit H and that

- AIM(nx) condition holds;
- θ exists,

then

$$H(x) = \{G(x)\}^{\theta}.$$

The implications of this result are that, except for a really weak long-range dependence condition:

1. for stationary processes the effect of dependence is through θ only;

2. θ can be absorbed into the normalising constants, so dependence does not change the limiting type;

3. the limit suggests
$$H(x) = \lim_{n\to\infty} [\{F(nx)\}^n]^\theta = \lim_{n\to\infty} F(nx)^{n\theta},$$
so $n\theta$ can be thought of as an effective number of independent variables.

Extreme value distributions

- Consider random variables $X_1, \ldots, X_n$ with arbitrary distribution function F and which satisfy the AIM(u_n) condition.

- As $n \to \infty$,

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq y\right) \to \exp\left[-\theta\left\{1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right\}_+^{-1/\xi}\right],$$

where if the variables are IID then $\theta = 1$.

We assume that

- $\{X_t\}$ is a stationary process with arbitrary distribution function F;
- the required norming constants for an IID process with marginal F are a_n and b_n with limit distribution G, a GEV$(0, 1, \xi)$ distribution;
- a long-range asymptotic independence condition (similar to the AIM$(a_n x + b_n)$ condition) holds.

Consider the point processes

$$P_n = \left\{ \left(\frac{i}{n+1}, \frac{X_i - b_n}{a_n} \right) : i = 1, \ldots, n \right\}$$

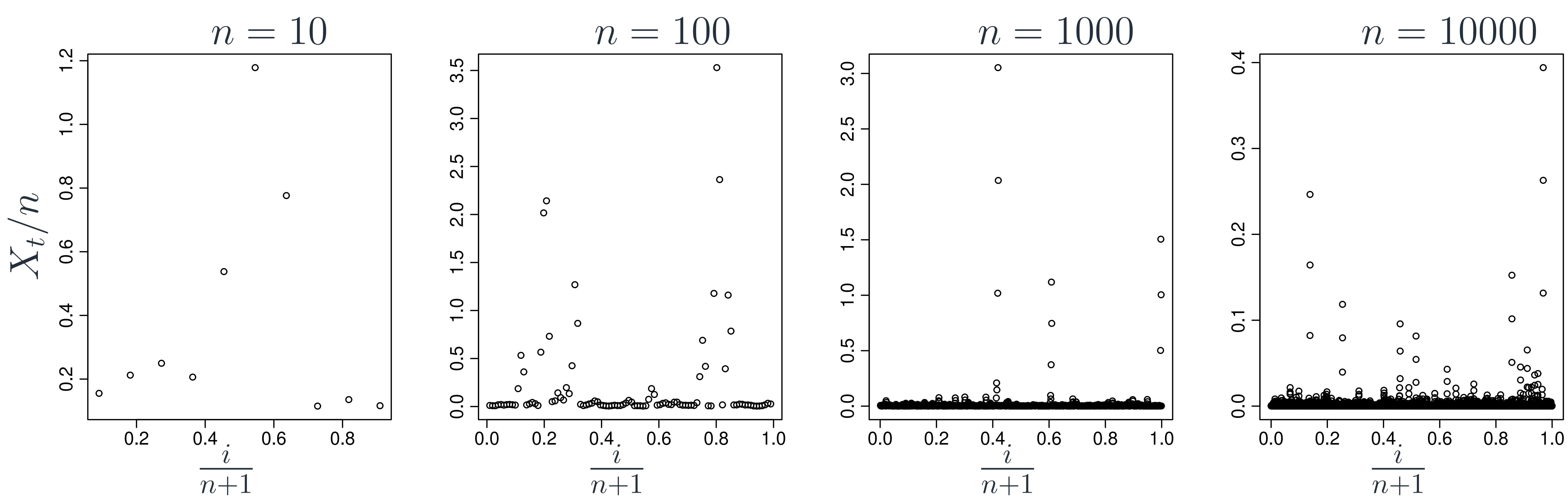
on $[0, 1] \times \mathbb{R}$.

Extremal index

Point process P_n for $n = 10, 100, 1000, 10000$ respectively with

$$X_i \sim \text{Moving Max}(\alpha_0, \alpha_1, \alpha_2) = \max(\alpha_0 Y_i, \alpha_1 Y_{i-1}, \alpha_2 Y_{i-2}),$$

with $\alpha_0 = 1/3, \alpha_1 = 1/2, \alpha_2 = 1/6.$



- Under the above conditions on P_n, on the set $[0,1] \times (u, \infty)$, where $u > b_l = \lim_{n\to\infty}(x_F - b_n)/a_n$, then [Hsing et al., 1988, Rootzén, 1988]

$$P_n \xrightarrow{d} P \qquad \text{as } n \to \infty,$$

where P is a clustered non-homogeneous Poisson process with cluster distribution π.

- The cluster distribution π is concentrated on $\{1, 2, \dots\}$ and has probability mass function

$$\pi(j) = \lim_{n\to\infty} \pi_n(j) = \lim_{n\to\infty} \mathbb{P}\left[\sum_{i=1}^{p_n} I\left(\frac{X_i - b_n}{a_n} > x\right) = j \,\middle|\, \sum_{i=1}^{p_n} I\left(\frac{X_i - b_n}{a_n} > x\right) > 0\right],$$

where $p_n = o(n)$ and $x > b_l$.

- The Laplace functional of the limiting Poisson process P is

$$-\log \mathbb{E}e^{-Pf} = \theta(1+\xi y)^{-1/\xi} \int_0^1 [1 - \phi(f(t))]\, dt, \qquad f \geq 0, \text{Borel}$$

where $\phi(s) = \sum_{j=1}^{\infty} e^{-sj}\pi(j)$ is the Laplace transform of the cluster distribution π.

- The extremal index satisfies $\theta^{-1} = \sum_{j=1}^{\infty} j\pi(j)$.
- The set of cluster maxima (the largest value in each vertical string of points in the point process) form a non-homogeneous Poisson process with intensity

$$\lambda(t, y) = \theta(1 + \xi y)_+^{-1-1/\xi}.$$

- For data over a high threshold u we absorb the location b_n and scale a_n normalisation into the parameters to give the intensity of the non-homogeneous Poisson Process for the cluster maxima as

$$\lambda(t, y) = \theta\sigma^{-1}\left[1 + \xi\left(\frac{y-\mu}{\sigma}\right)\right]_+^{-1-1/\xi}.$$

Fixing a threshold level $u > b_l$ we have that:

1. clusters are those set of points exceeding u which occur at the same normalised time. In the original series these are large points which occur within a short time period of one another.

2. the expected number of exceedances of u per cluster (for which the cluster maxima exceeds u) is θ^{-1}, irrespective of the level u;

3. relative to independent series, when $\theta < 1$ there are fewer clusters (by a factor θ) and more exceedances per cluster (by a factor θ^{-1});

4. values in one cluster are independent of values in another cluster;

5. values within a cluster are dependent.

Implications of the limit process: exceedances

Fixing a threshold level $u > b_l$ we have that:

1. the rate of cluster maxima exceeding the threshold u is
$$\theta[1 + \xi(u - \mu)/\sigma]_+^{-1/\xi};$$
2. the rate of arbitrary values exceeding the threshold u is
$$[1 + \xi(u - \mu)/\sigma]_+^{-1/\xi};$$
3. cluster maxima excesses over u are independent, and follow a GPD(σ_u, ξ), with shape parameter ξ; GPD(σ_u, ξ) model for cluster maxima excesses over u, where
$$\sigma_u = \sigma + \xi(u - \mu);$$
4. arbitrary (in time) excesses over u also follow a GPD(σ_u, ξ), *i.e.* with the same parameters as the cluster maxima.

Statistical inference for extremes of stationary processes

We consider the additional practical difficulties encountered due to dependence in analysing the extreme values of a stationary process.

In a practical data analysis, we must:

- select a high threshold;
- identify clusters of extreme values which are independent from one another;
- estimate the extremal index.

When clusters are defined we

- extract the cluster maxima and fit the GPD distribution to the cluster maxima over the threshold;
- estimate characteristics of the cluster;
- assess sensitivity of cluster characteristics to threshold selection.

There are a number of ways of identifying independent clusters of extreme values.
Here we will focus on:

- the runs method, proposed by Smith and Weissman [1994];
- the blocks and intervals methods, see Ferro and Segers [2003];
- the identification of conditions that can be tested for independence between clusters, see Ledford and Tawn [2003].

For a selected high threshold u we conclude that if consecutive exceedances of the threshold u are separated by a set of r consecutive observations below the threshold u then the exceedances belong to separate clusters.

Similarly, exceedances separated by less than r consecutive non-exceedances are deemed to be in the same cluster.

The choice of u and r is critical, r needs to be the time lag when the process is "independent in the extremes".

These motivate the runs estimator of the extremal index θ:

- Use the empirical form the extremal index,

$$\widehat{\theta}(u, r) = \widehat{\mathbb{P}}(M_{2,r+1} \leq u \mid X_1 > u),$$

  using the empirical probability.

- $\widehat{\theta}(u, r)$ is identical to $1/(\text{mean cluster size})$ with clusters defined by the runs method.

- $\widehat{\theta}(u, r)$ can be obtained using the function `exi(data,u,r)` from the `evd` R package.

- Another estimator can be constructed by exploting the fact that the extremal index is related to the times between threshold exceedances.
- Let $\{X_t\}_{t\geq 1}$ be a process for which we consider the exceedances of a threshold u. Let F denote de marginal distribution of the X_i.
- Assume W.L.O.G. that $X_1 > u$, and let $T(u)$ denote a random waiting time until the next exceedance of u.
- Then, $T(u) = \min\{t \geq 1 : X_{t+1} > u\}$ and

$$\mathbb{P}(T(u) > n) = \mathbb{P}(\max\{X_2, \ldots, X_{n+1}\} \leq u \mid X_1 > u).$$

- If the X_i are IID, then

$$\mathbb{P}(T(u) > n) = F(u)^n,$$

and it can be shown that $\{1 - F(u)\}T(u)$ is asymptotically unit exponential.

- When the X_i are not independent, *i.e.*, $\theta \in (0,1)$, it can be shown that

$$\{1 - F(u)\}T(u) \xrightarrow{d} T_\theta,$$

where T_θ denotes a random variable distributed according to the mixture

$$(1-\theta)\varepsilon_0 + \theta\mu_\theta$$

where ε_0 is the degenerate probability distribution at 0 and μ_θ is the exponential distribution with mean θ^{-1}.

- Note: θ is both the proportion of non-zero interexceedance times and the reciprocal of the mean of the non-zero interexceedance times.

- Suppose that we observe n_u exceedances of u at times $1 \leq S_1 < \ldots < S_{n_u} \leq n$.
- Compute the interexceedance times $T_i = S_{i+1} - S_i$ for $i = 1, \ldots, n_{u-1}$.
- Set

$$\widehat{\theta}_n^{\delta}(u) = \frac{2\left(\sum_{i=1}^{n_u-1} T_i\right)^2}{(n_u-1)\sum_{i=1}^{n_u-1} T_i^2} \quad \text{and} \quad \widehat{\theta}_n^{\star}(u) = \frac{2\left(\sum_{i=1}^{n_u-1} T_i - 1\right)^2}{(n_u-1)\sum_{i=1}^{n_u-1} (T_i-1)(T_i-2)}.$$

- Then, the Ferro & Segers estimator $\widehat{\theta}_{FS}$ is given by

$$\widehat{\theta}_{FS} = \begin{cases} \min\{1, \widehat{\theta}_n^{\delta}(u)\} & \text{if } \max\{T_i : 1 \leq i \leq n_{u-1}\} \leq 2 \\ \min\{1, \widehat{\theta}_n^{\star}(u)\} & \text{if } \max\{T_i : 1 \leq i \leq n_{u-1}\} > 2 \end{cases}$$

- $\widehat{\theta}_{FS}$ can be obtained using the function `exi(data,u,r=0)` from the `evd` R package.

- A convenient way to estimate uncertainty for an estimate $\widehat{\theta}$ is to perform a bootstrap procedure.

- In the case of stationary sequences, a block bootstrap procedure is chosen to account for short range dependence structures.
  - Consider n realisations $\boldsymbol{x}_{1:n} = \{x_t\}_{t=1}^{n}$ from a time-series $\{X_t \,:\, t = 1, 2, \ldots\}$.
  - Create a collection of moving blocks of size b, *e.g.*, $C = \{\boldsymbol{x}_{1\,:\,b}, \ldots, \boldsymbol{x}_{i\,:\,(i+b-1)}, \ldots, \boldsymbol{x}_{n\,:\,b}\}$.
  - Create a block bootstrap sample $\boldsymbol{X}_{1:n}$ by sampling n/b blocks from C with replacement.

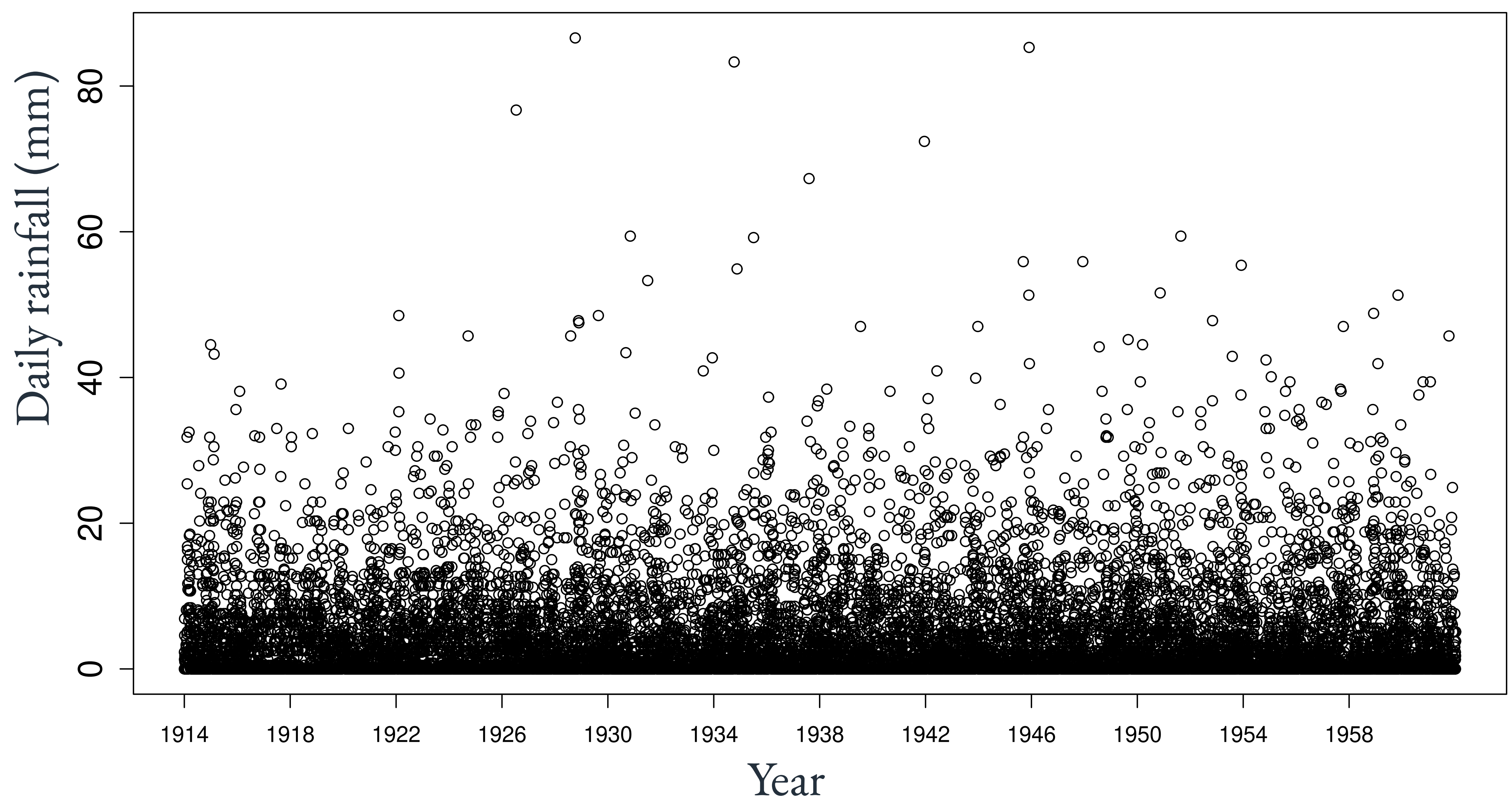
- For each of m block bootstrap samples, estimate the extremal index, resulting in a collection $\widehat{\boldsymbol{\theta}} = \{\widehat{\theta}_1, \ldots, \widehat{\theta}_m\}$.

- Construct a quantile-based confidence interval from $\widehat{\boldsymbol{\theta}}$.

Examples

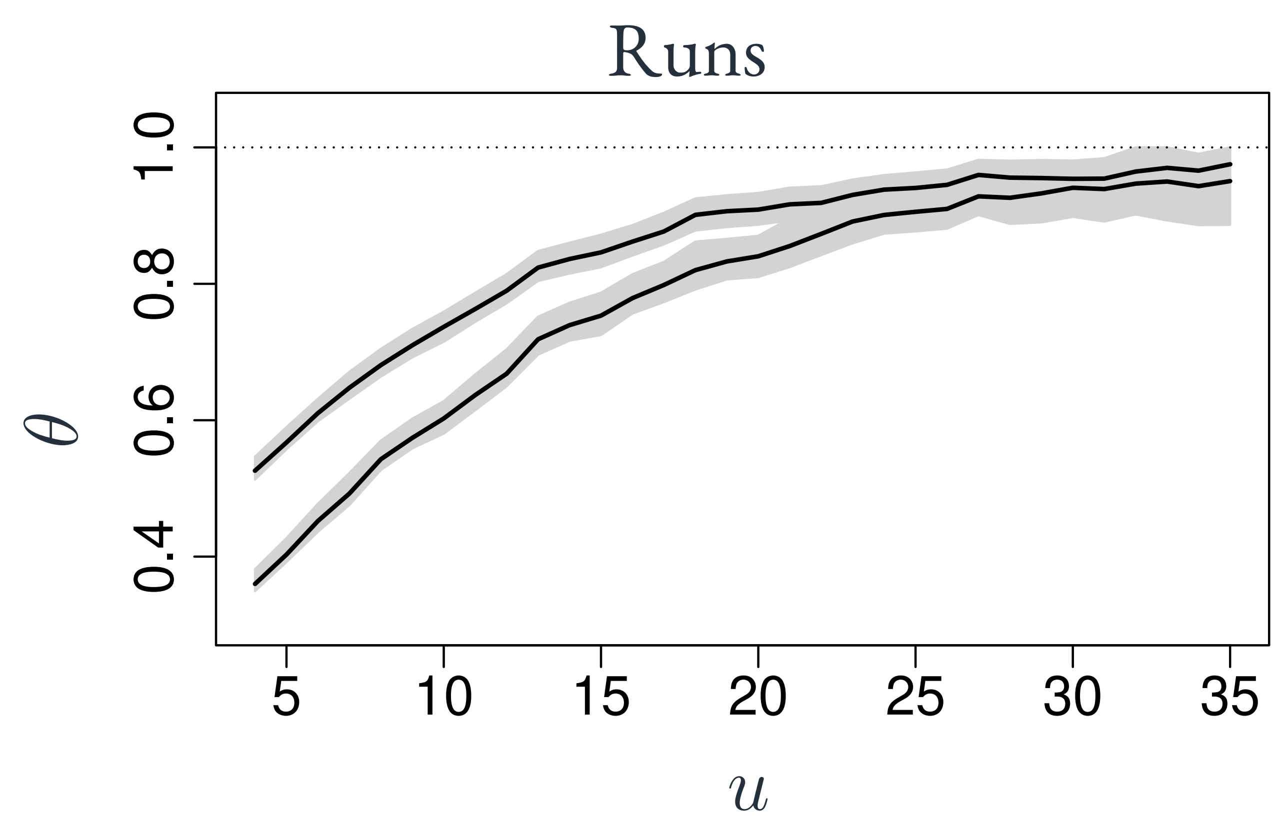
Example: daily rainfall data

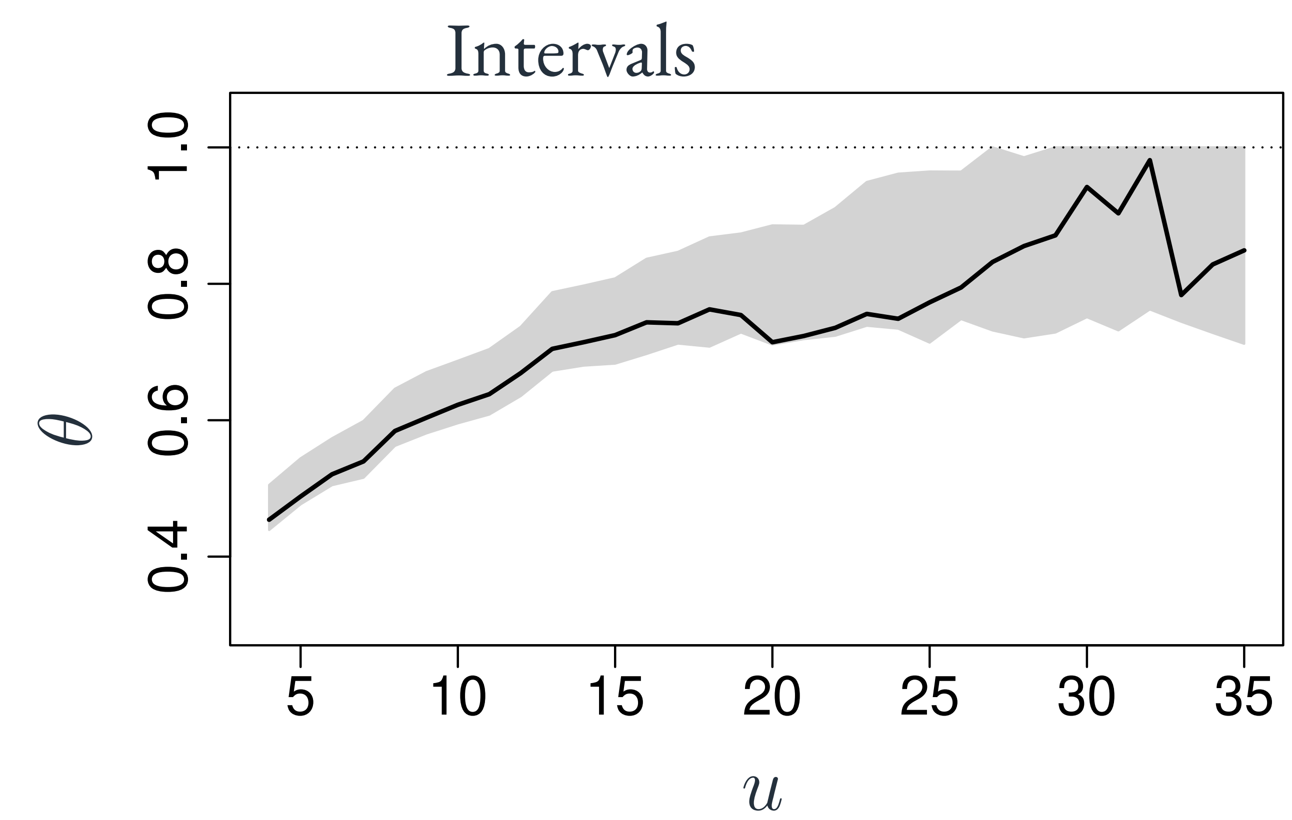
We illustrate the use of these cluster identification and extremal index estimation using a time series of daily rainfall accumulations at a location in south-west England, recorded during 1914-1962.



The series is assumed to be stationary over this observation period.

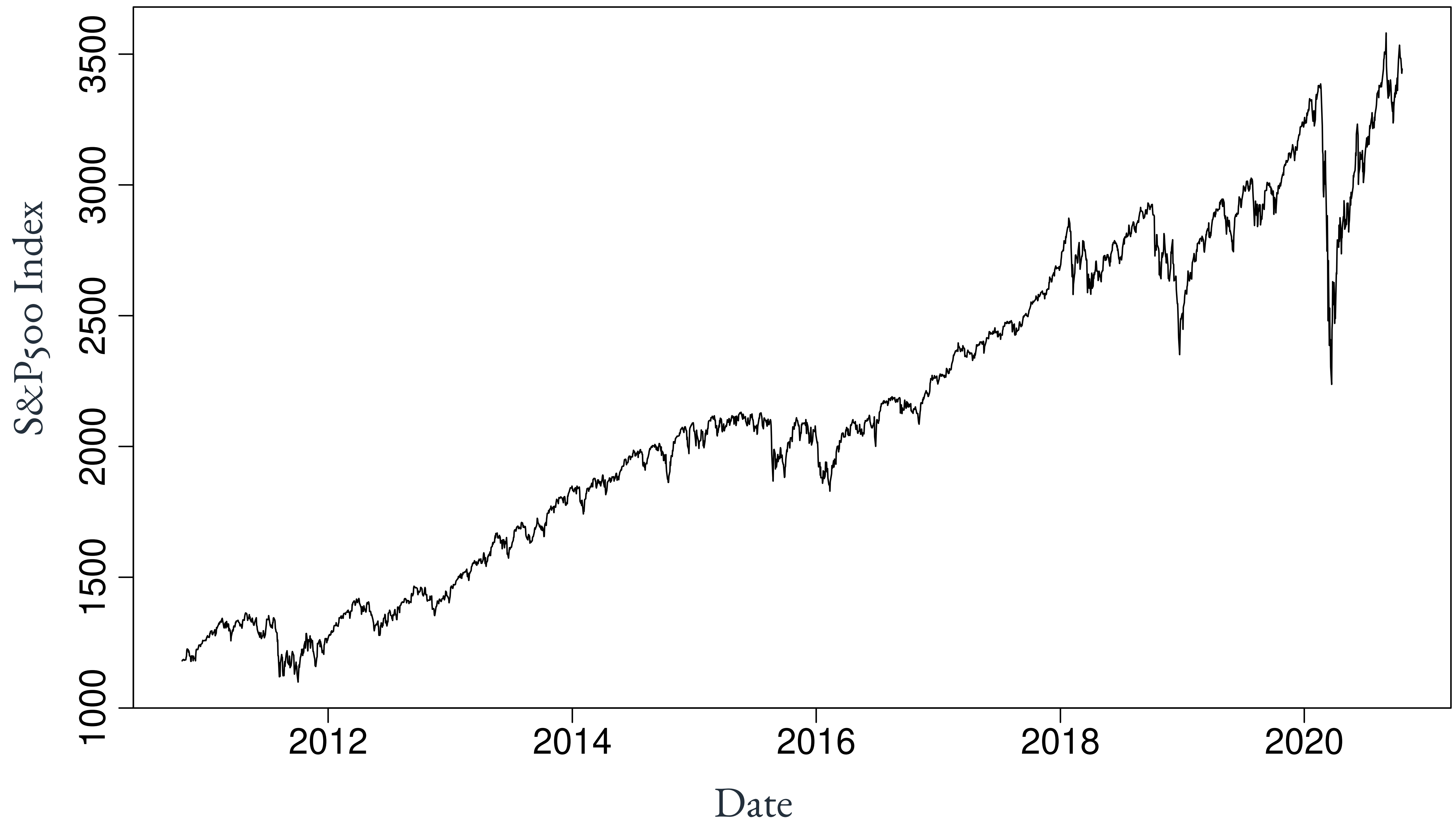
Extremal index for the rain data



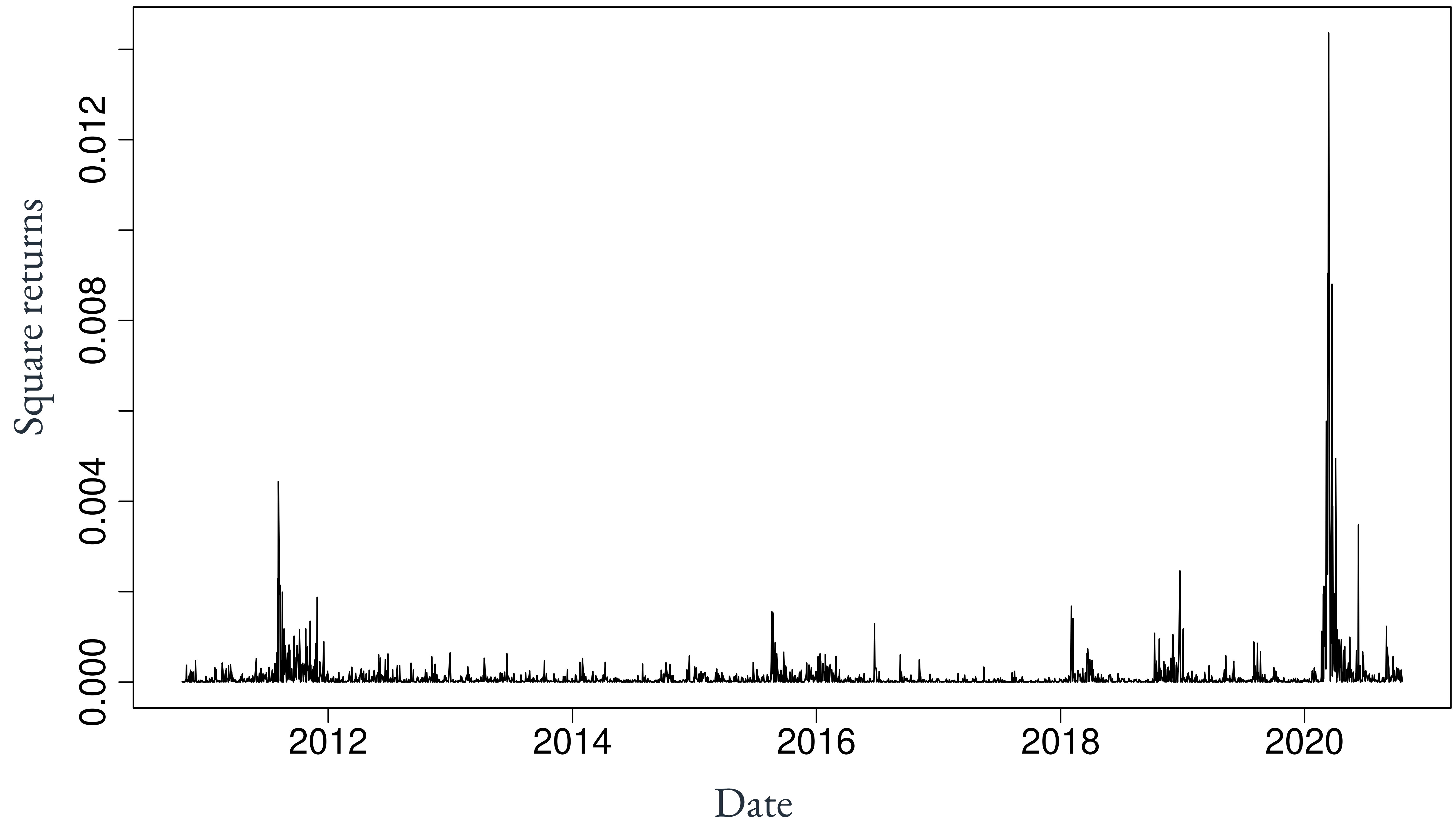


- $\widehat{\theta}(u, r)$ are threshold dependent;
- increasing r does not materially change our estimate $\widehat{\theta}(u, r)$;
- $r = 1$ is appropriate for cluster identification;
- the value of $\widehat{\theta}(u, r) \approx 1$ indicates weak short-range dependence with a limiting cluster size of 1 for higher thresholds;
- rainfall episodes tend to last a number of days, but this analysis tells us that extreme daily rainfall events tend to be isolated.

We now turn to estimating clustering for the financial time series data.

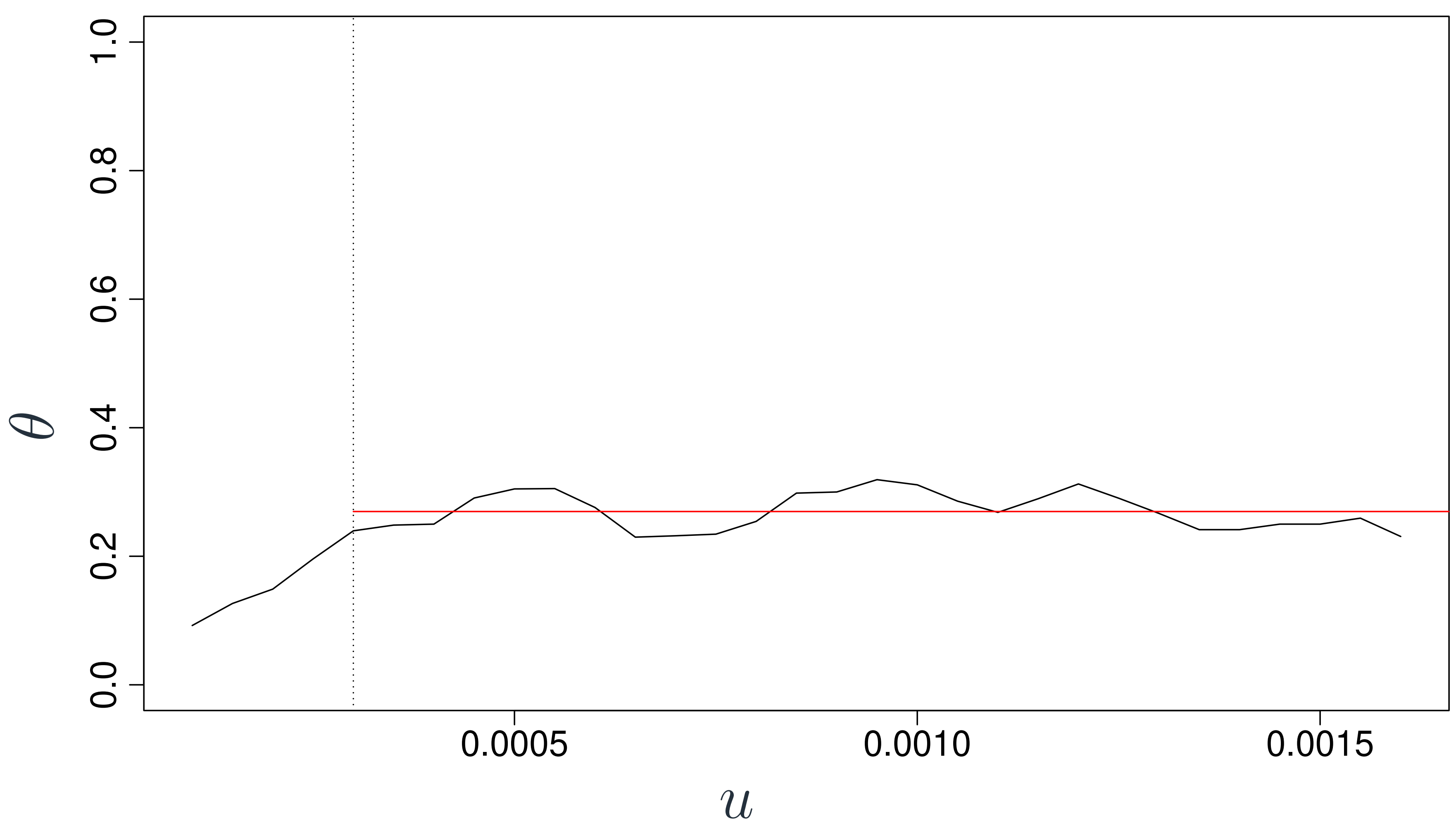


We focus on the squared returns from the S&P500 share index.



Recall: The return R_t for day t is $R_t = (X_{t-1} - X_t)/X_{t-1}$.

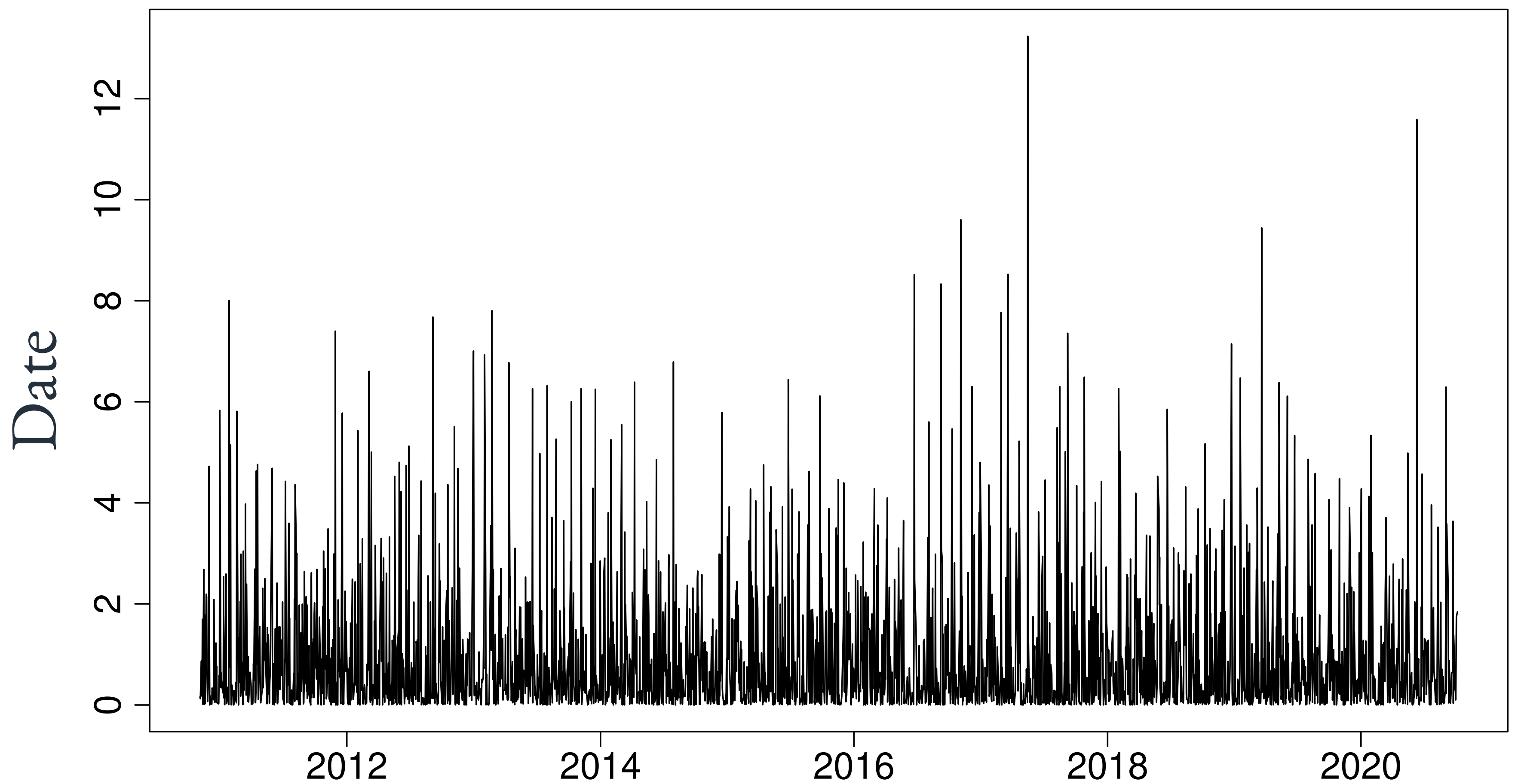
We compute the extremal index of the squared returns based on a runs estimator with $r = 10$.



- $\widehat{\theta}(u, r)$ is stable for $u \geq 0.0003$ (dotted line) corresponding to the 0.92 quantile of the squared returns.
- The value of $\widehat{\theta}(u \geq 0.0003, r) \approx 0.270$ (red line) suggests that the mean cluster size is around 3.70.
- $r = 10$ seems fine for cluster identification.

Squared standardised returns

We now investigate the source of serial dependence being the changing volatility of the process.We work with the squared standardised series:



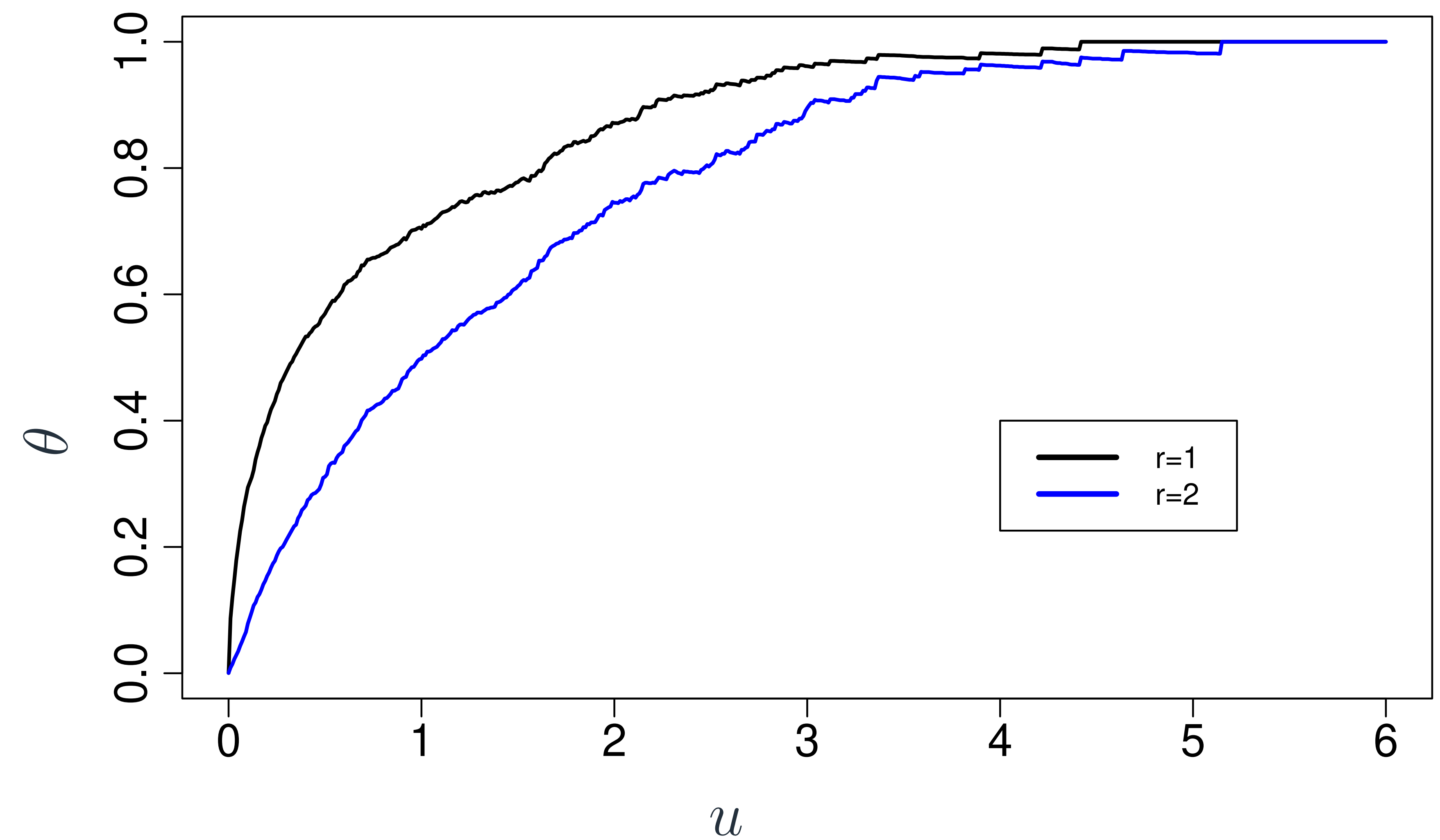
Squared standardised returns

We standardise the returns series by

- subtracting the local mean;
- dividing by the local standard deviation.

Local means and standard deviations were calculated using the $h = 21$ observations centred on the value to be standardised.

Extremal index for the squared standardised returns

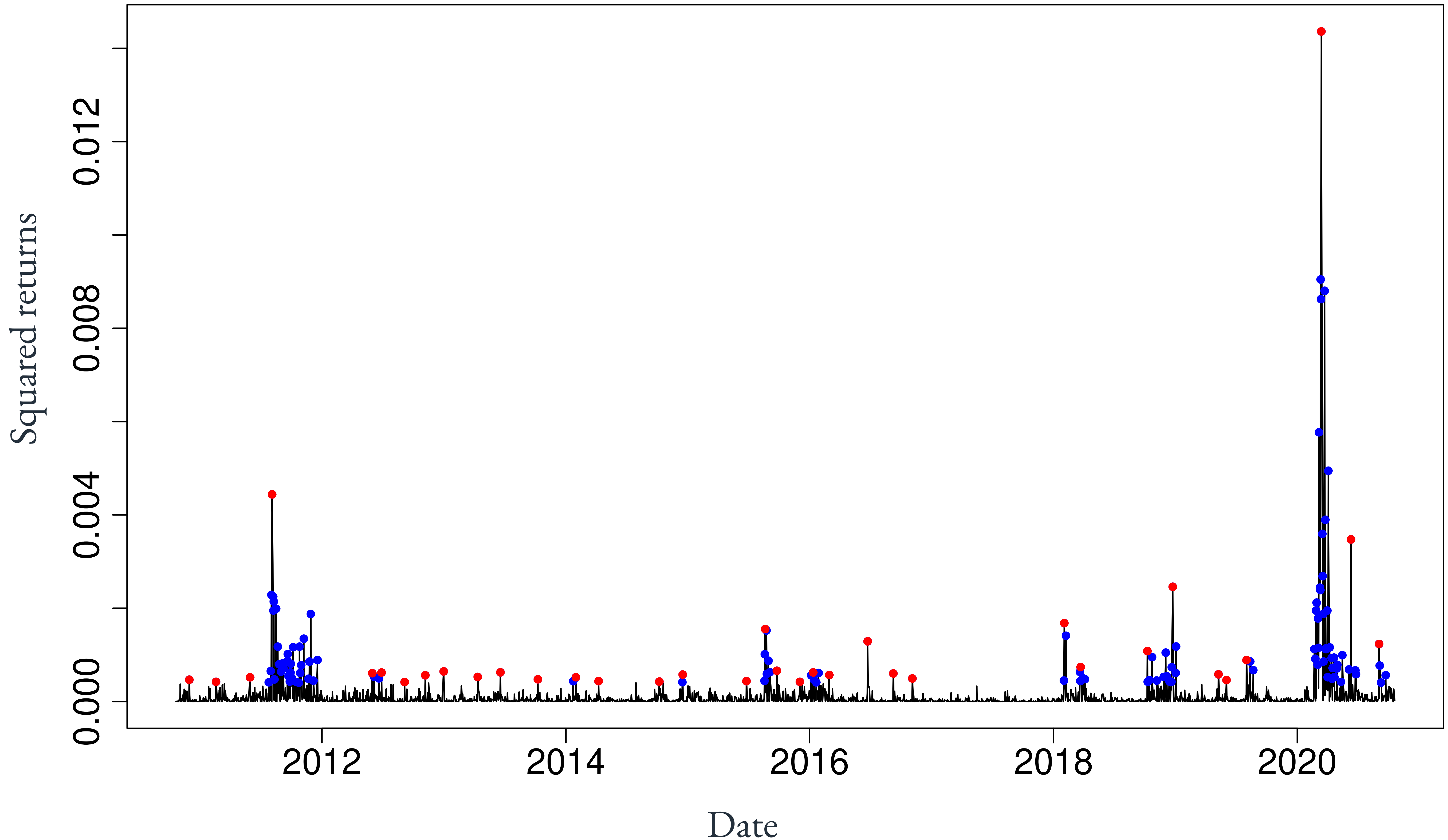


- Estimates of $\widehat{\theta}(u, r)$ are threshold dependent.
- Increasing r does not materially change our conclusions, $r = 1$ is fine for cluster identification.
- The limiting value of $\theta = 1$ indicates weak short-range independence with a limiting cluster size of 1.
- By undertaking the more sophisticated analysis and standardising the series we appear to have accounted for all the short-range dependence in the series.

Modelling cluster maxima and all exceedances

For a specific threshold u, we can

- model the cluster maxima:
  - identify the clusters for a given threshold u and run length r using the `clusters()` function from the `evd` package in R.
  - fit the $\mathrm{GPD}(\sigma_u, \xi)$ model for cluster maxima excesses over u;
  - standard errors can be obtained by usual methods as cluster maxima are independent.
- model all exceedances:
  - fit the $\mathrm{GPD}(\sigma_u, \xi)$ model for the excess of all exceedances over u;
  - the data are dependent so block bootstrap methods are needed for standard error evaluation.


Squared returns
0.000
0.004
0.008
0.012
2012
2014
2016
2018
2020
Date

Point process intensity estimation

To estimate μ and σ in the point process intensity, we need information in addition to that given by our estimate of σ_u and ξ.

This information is given by the expected number of cluster maxima exceeding u.

Equating the point process and sample values gives

$$\theta[1 + \xi(u - \mu)/\sigma]_+^{-1/\xi} = n_{\mathrm{cl},u}.$$

It follows that

$$\widehat{\sigma} = \widehat{\sigma}_u (n_{\mathrm{exc},u})^{\widehat{\xi}}$$

and

$$\widehat{\mu} = u + (\widehat{\sigma} - \widehat{\sigma}_u)/\widehat{\xi}.$$

θ is estimated by $\widehat{\theta}_u = n_{\mathrm{cl},u}/n_{\mathrm{exc},u}$, *i.e.*, the number of clusters above u divided by the number of exceedances of u.

Example: Squared S&P500 returns

For the squared S&P500 returns:

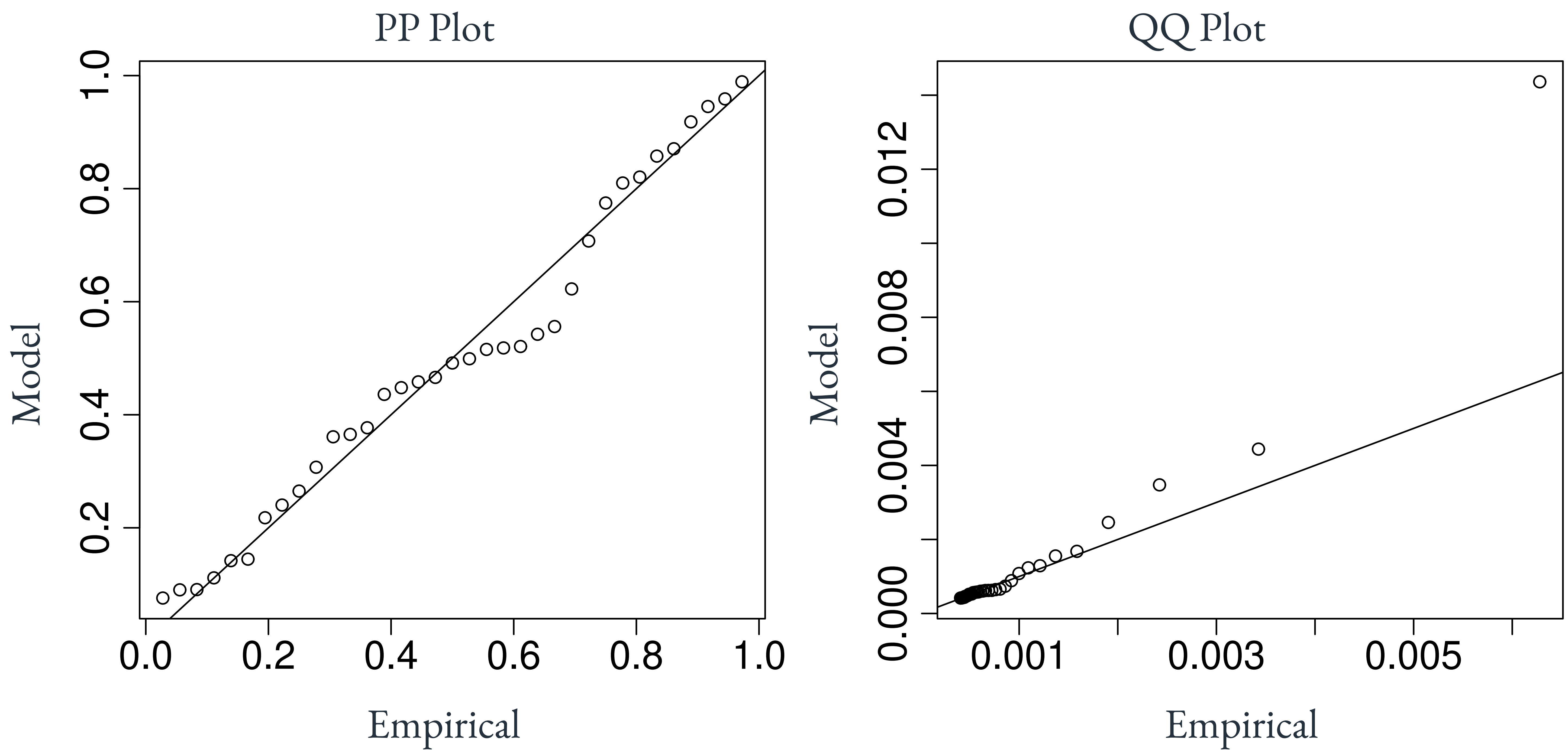
- using a declustering parameter of $r = 10$ and a threshold of $u = 0.0004$ corresponding to the 0.95 quantile of the squared returns gives $n_{\mathrm{cl},u} = 35$ independent clusters;
- there are $n_{\mathrm{exc},u} = 140$ exceedances of u;
- we estimate the extremal index as $\widehat{\theta} = 0.25$;
- fitting a GPD to cluster maxima gives $\widehat{\sigma}_u = 0.000215(0.00002)$, $\widehat{\xi} = 0.905(0.271)$;
- combining the above estimates gives

$$\widehat{\sigma} = \widehat{\sigma}_u (n_{\mathrm{exc},u})^{\widehat{\xi}} = 0.0189$$

and

$$\widehat{\mu} = u + (\widehat{\sigma} - \widehat{\sigma}_u)/\widehat{\xi} = 0.0211.$$

Diagnostics of GPD fit



References

C. A. T. Ferro and J. Segers. Inference for clusters of extreme values. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):545–556, 2003. doi: https://doi.org/10.1111/1467-9868.00401. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00401`.

T. Hsing, J. Hüsler, and M. Leadbetter. On the exceedance point process for a stationary sequence. *Probab. Theory Reltd Flds*, 78:97–112, 1988.

M. R. Leadbetter, G. Lindgren, and H. Rootzen. *Extremes and Related Properties of Random Sequences and Processes*. Springer Series in Statistics. Springer Verlag, 1983. ISBN 0387907319. `https://doi.org/10.1007/978-1-4612-5449-2`.

A. W. Ledford and J. A. Tawn. Diagnostics for dependence within time series extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65(2):521–543, 2003. doi: https://doi.org/10.1111/1467-9868.00400. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/1467-9868.00400`.

G. L. O'Brien. Extreme Values for Stationary and Markov Sequences. *The Annals of Probability*, 15(1):281 – 291, 1987. doi: 10.1214/aop/1176992270. URL `https://doi.org/10.1214/aop/1176992270`.

H. Rootzén. Maxima and exceedances of stationary Markov chains. *Adv. in Appl. Probab.*, 20:371–390, 1988.

R. L. Smith and I. Weissman. Estimating the extremal index. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3):515–528, 1994. doi: https://doi.org/10.1111/j.2517-6161.1994.tb01997.x. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1994.tb01997.x`.