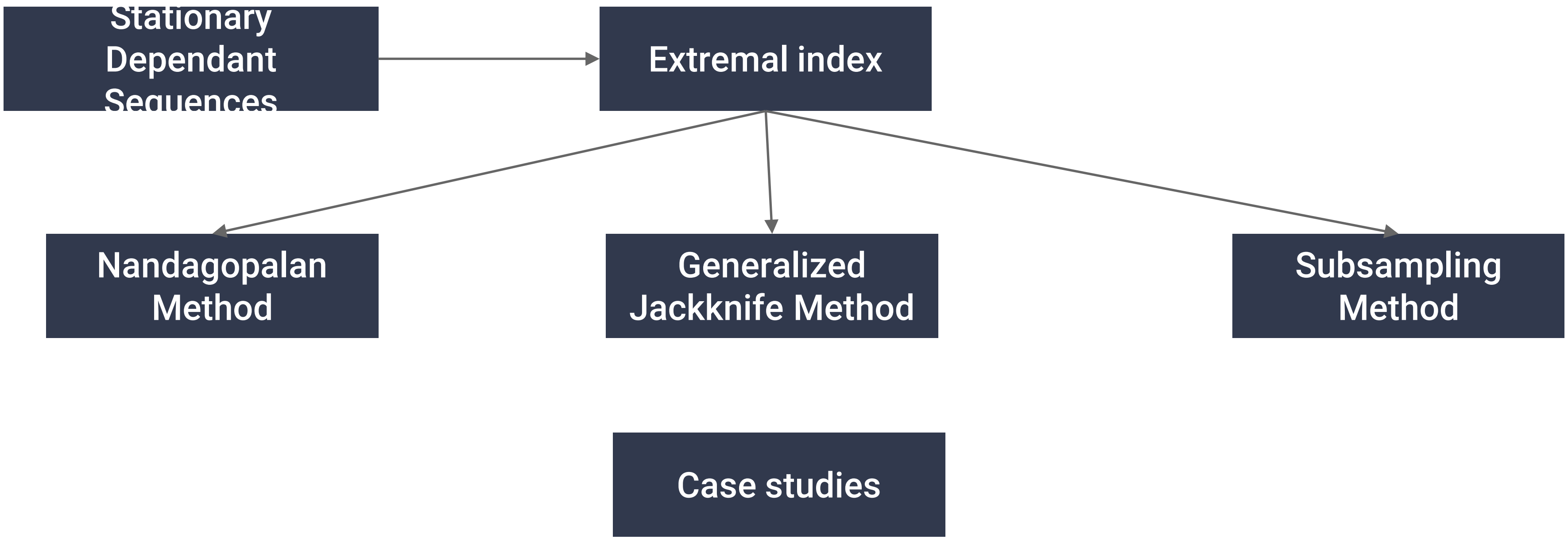
# Computational Statistics (MA51109)

## Term Paper

# Abstract


- Poisson Process: A type of random mathematical object that consists of points randomly located on a mathematical space.
- Compound Poisson Process: Process in which each arrival in an ordinary Poisson process comes with an associated real-valued random variable that represents the value of the arrival in a sense.
- There exists the extremal index $\theta$, $0 \leq \theta \leq 1$, which is directly related to the clustering of exceedances of high values for such stationary sequences.
- $\theta = 1$, for independent, identically distributed sequences and $\theta > 0$, for most of the cases.
- Estimation of extremal index through the use of the Generalized Jackknife methodology, possibly together with the use of subsampling techniques.

# Flow of the paper



# Scope of the Paper

- Assumption of the bias term of the extremal index estimator same as the bias term of the Nandagopalan's estimator $(\theta_N)$ for the stationary ARMAX sequences and i.i.d. sequences. There is an important tradeoff between bias and variance, linked to the choice of the threshold.
- Bias reduction and stability over a considerable set of thresholds is the **major objective** of this paper.

**Other objectives:**

- To Perform Generalized Jackknife methodology for the estimation of the extremal index for ARMAX process as well as for i.i.d random samples.
- To use subsampling techniques in the estimation process of the extremal index and look for the effects and changes.

# Definitions

**Defintion 1.1.** The $\mathbf{D}(u_n)$ condition holds for the stationary sequence $\mathbf{X}$ if for any integers $p, q$ and $i_1 < i_2 < \cdots < i_p < j_1 < j_2 < \cdots j_q < n$, for which $j_1 - i_p \geqslant l = l_n$, we have

$$|F_{i_1,i_2,\ldots,i_p,j_1,j_2,\ldots,j_q}(u_n,\ldots,u_n) - F_{i_1,i_2,\ldots,i_p}(u_n,\ldots,u_n)F_{j_1,j_2,\ldots,j_q}(u_n,\ldots,u_n) \leqslant \alpha_{n,l},$$

with $\alpha_{n,l} \underset{n\to\infty}{\longrightarrow} 0$, for some sequence $\{l_n = \mathrm{o}(n)\}$.

**Defintion 1.2.** The local dependence condition $\mathbf{D}''(u_n)$ holds for a stationary $\mathbf{D}(u_n)$-sequence $\mathbf{X}$ if there exists a sequence of integers $\{s_n\}$, with $s_n \to \infty$, $s_n \alpha_{n,l_n} \to 0$, $s_n l_n/n \to 0$, $s_n(1 - F(u_n)) \to 0$, as $n \to \infty$, such that

$$\lim_{n\to\infty} n \sum_{j=3}^{r_n} \mathbb{P}(X_1 > u_n, X_{j-1} \leqslant u_n < X_j) = 0,$$

with $r_n = [n/s_n]$.

# Tail index

For Y = $\{Y_n\}_{n\geq1}$, sequence of i.i.d. r.v.'s from an underlying parent d.f. F. Let $Y_{i:n}$ , 1 i n, be the set of associated ascending order statistics (o.s.). Tail index is defined as:

$$\mathbb{P}[Y_{n:n} := \max(Y_1, Y_2, \ldots, Y_n) \leqslant x] = F^n(x) \approx EV_\gamma\left(\frac{x - b_n}{a_n}\right),$$

which holds for large values of $n$, and appropriate sequences $a_n > 0$, $b_n \in \mathbb{R}$, with

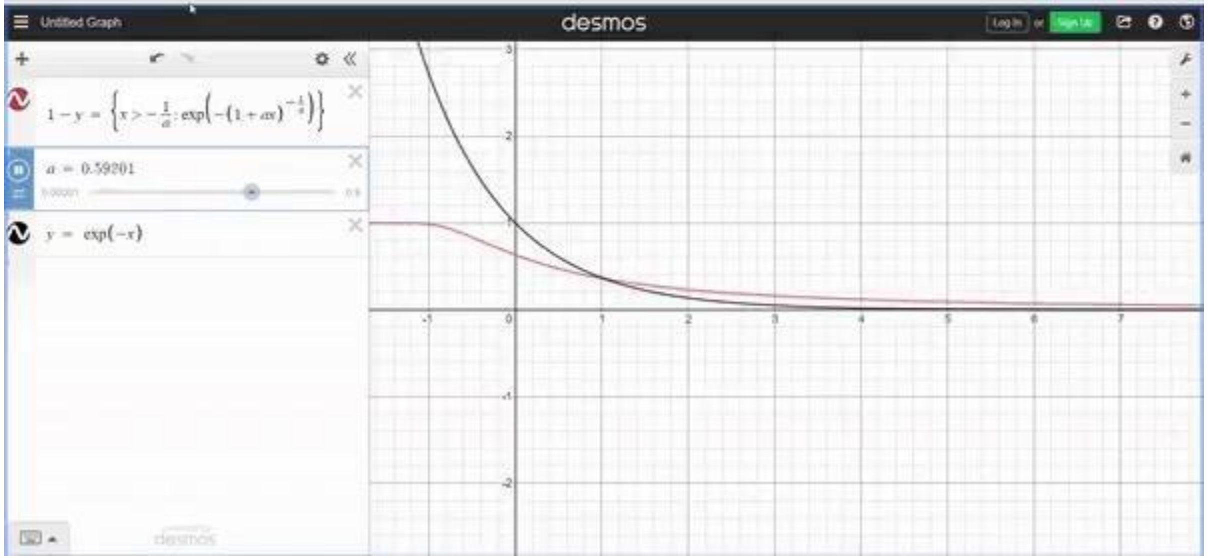
$$EV_\gamma(x) = \begin{cases} \exp\{-(1+\gamma x)^{-1/\gamma}\}, & 1+\gamma x > 0 \quad \text{if } \gamma \neq 0, \\ \exp(-\exp(-x)), & x \in \mathbb{R} \quad \text{if } \gamma = 0, \end{cases}$$

The tail index $\gamma$ is thus directly related to the weight of the right tail of the underlying model F and as $\gamma$ increases the right tail becomes heavier and heavier.

If the sequence of maximum values $Y_{n:n}$ , linearly normalized, converges towards a r.v. has an $EV_\gamma$ d.f. defined above.

# Background Search

- **Heavy Tailed Distributions:** The distribution which have their CCDF (Complementary Cumulative Density Function) tail above that exponential distribution, or in other terms they are the distributions whose tails are not exponentially bounded.
- CCDF of exponential distribution = 1 - CDF = $e^{-\lambda x}$




Figurative verification to heavy tailedness of Extreme Value function

# Extremal Index

Let X = $\{X_n\}_{n\geq 1}$ stationary sequence comes from an underlying d.f. F and the limiting d.f. of the maximum $X_{n:n} := \max(X_1, X_2, \ldots, X_n)$ may be directly related to the limiting d.f. Of the maximum, $Y_{n:n}$, of the associated i.i.d. sequence, through a new parameter, the so-called **extremal index.**

**Defintion 1.3.** The stationary sequence $\{X_n\}_{n\geqslant 1}$ is said to have an extremal index $\theta$ $(0 < \theta \leqslant 1)$ if, for all $\tau > 0$, we may find a sequence of levels $u_n = u_n(\tau)$ such that

$$\mathbb{P}(Y_{n:n} \leqslant u_n) = F^n(u_n) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\tau} \quad \text{and} \quad \mathbb{P}(X_{n:n} \leqslant u_n) \underset{n\to\infty}{\longrightarrow} \mathrm{e}^{-\theta\tau}. \tag{2}$$

From Definition 1.3 it follows that the *extremal index* $\theta$ may be informally defined by the approximations

$$P[\max(X_1, X_2, \ldots, X_n) \leqslant x] \approx F^{n\theta}(x) \approx EV_\gamma^\theta\left(\frac{x - b_n}{a_n}\right)$$
$$= EV_\gamma\left(\frac{x - b_n'}{a_n'}\right), \quad \begin{cases} a_n' = a_n\theta^\gamma, \\ b_n' = b_n + a_n\left(\frac{\theta^\gamma - 1}{\gamma}\right). \end{cases}$$

# Extremal Index

Under independence, this point process converges to a homogeneous Poisson process, as n → ∞, but when there is a slightly stronger local dependence, clusters of exceedances may occur and the limiting process may be a compound Poisson process.

The extremal index can then also be defined as the reciprocal of the "mean time of duration of extreme events".

$$\theta = \frac{1}{\text{limiting mean size of clusters}}$$

$$= \lim_{n\to\infty} P(X_2 \leqslant u_n | X_1 > u_n) \quad \text{(downcrossings)}, \tag{3}$$

$$= \lim_{n\to\infty} P(X_1 \leqslant u_n | X_2 > u_n) \quad \text{(upcrossings)}, \tag{4}$$

where $u_n$, a sequence of values such that

$$F(u_n) = 1 - \tau/n + o(1/n) \quad \text{as } n \to \infty, \tag{5}$$

a so-called *normalized level*, i.e., a level such that $n(1 - F(u_n)) \to \tau > 0$, or equivalently, a level such that $F^n(u_n) \to \exp(-\tau)$, as $n \to \infty$, i.e., (2) holds.

# Nandagopalan's Extremal Index

**Nandagopalan's estimator:** For given $(X_1, X_2, \ldots, X_n)$ samples, non-parametric estimator of θ, given by the ratio between the number of down-crossings/up-crossings and the number of exceedances over a high threshold is

$$\theta_n^N = \theta_n^N(u) := \frac{\sum_{j=1}^{n-1} I_{[X_j > u, X_{j+1} \leqslant u]}}{\sum_{j=1}^{n} I_{[X_j > u]}} = \frac{\sum_{j=1}^{n-1} I_{[X_j \leqslant u < X_{j+1}]}}{\sum_{j=1}^{n} I_{[X_j > u]}},$$

$u = u_n$ must be such that $n(1 - F(u_n)) = c_n\tau = \tau_n$, $\tau_n \to \infty$ and $\tau_n/n \to 0$.

If $\tau$ is fixed, the Nandagopalan's estimator may not be consistent. So we need to assume $\tau = \tau_n \to \infty$, as $n \to \infty$. We shall here consider a deterministic level $u \in [X_{n-k:n}, X_{n-k+1:n})$, the interval between the $(k + 1)$ and the $k^{th}$ top order statistics.

# Nandagopalan’s Extremal Index

The extremal index estimator is then a function of k, the number of top o.s. higher than the chosen threshold. The estimator can be modified as

$$\hat{\theta}_n^N(k) \equiv \theta_n^N(u) := \frac{1}{k}\sum_{j=1}^{n-1} I_{[X_j \leqslant u < X_{j+1}]}, \quad u \in [X_{n-k:n}, X_{n-k+1:n}).$$

**Or**

$$\hat{\theta}_n^N(k) = \frac{1}{k}\sum_{j=1}^{n-1} I_{[X_j \leqslant X_{n-k:n} < X_{j+1}]}.$$

Consistency is attained only if k is intermediate, i.e., $k = k_n \to \infty$, $k = o(n)$ as $n \to \infty$.

# Fréchet Process

$$F(x) \equiv \Phi_\gamma(x) = \exp(-x^{-1/\gamma}),\, x > 0,\, \gamma > 0.$$

- The cdf of the Frechet Parent distribution we have utilised to simulate from.
- We have utilised the inverse transform method to simulate from this distribution.
- The data simulated from this distribution are iid stationary.

```
def cdf_inv_fr(u, gamma):
  return ((pow(-np.log(u) , -gamma)))

def iid_fr(gamma , n , random_state = 124):
  x = np.zeros(n)
  r = np.random.RandomState(random_state)
  t = 0
  for i in range(n):
    r2 = np.random.RandomState(random_state + i)
    u = r2.uniform(0,1,1)[0]
    xi = cdf_inv_fr(u,gamma)
    x[t] = xi
    t = t+1
  return x
```

# ARMAX Process

An ARMAX process is based on an i.i.d. sequence of innovations $\{Z_i\}_{i>=1}$, with degree of freedom H, and is defined through the following relation:

$$\boxed{X_i = \beta \max(X_{i-1}, Z_i), \quad i \geqslant 1,\ 0 < \beta < 1.}$$

ARMAX sequence has a stationary distribution F, it depends on the d.f. H of the i.i.d. sequence $\{Z_i\}_{i>=1}$ through the relation:

$$F(x)/F(x/\beta) = H(x/\beta)$$

If we consider Fréchet innovations, such that $H(x) = \Phi_\gamma^{\beta^{-1/\gamma}-1}(x)$, we then get $F(x) = \Phi_\gamma(x)$, and

$$\theta = \lim_{x\to\infty} \frac{P(X_i > x, X_{i+1} \leqslant x)}{P(X_i > x)} = 1 - \lim_{x\to\infty} \frac{1 - F(x/\beta)}{1 - F(x)} = 1 - \beta^{1/\gamma}.$$

For the particular case $\gamma = 1$, considered later on for illustration, we thus get $\theta = 1 - \beta$.

# ARMAX Process

Let us consider the ARMAX model. As said before, the use of the stationarity equation in which enables us to write:

$$\boxed{F(x) = \exp(-x^{-1/\gamma}) \iff H(x) = \frac{F(\beta x)}{F(x)} = \exp(-x^{-1/\gamma}(\beta^{-1/\gamma} - 1)).}$$

Then,

$$F(u) = 1 - \tau/n \iff u = [-\ln(1 - \tau/n)]^{-\gamma}.$$

Consequently, $F(u/\beta) = (1 - \tau/n)^{\beta^{1/\gamma}}$, $H(u) = (1 - \tau/n)^{\beta^{-1/\gamma}-1}$ and

$$H(u/\beta) = (1 - \tau/n)^{1-\beta^{1/\gamma}} =: (1 - \tau/n)^{\theta}, \quad \theta := 1 - \beta^{1/\gamma}.$$

# Generalized Jackknife Method

The main objectives of the Jackknife methodology are the following:

1. Bias and variance estimation, together with the derivation of the sampling distribution of a certain statistic, only through manipulation of the observed data x.
2. The building of estimators with bias and mean squared error smaller than those of an initial set of estimators.

Jackknife Methodology is thus a **resampling methodology**, which usually gives a positive answer to the question whether the combination of information can improve the quality of estimators of a certain parameter or functional.

# Generalized Jackknife Method

According to the Generalized Jackknife methodology, we need to have access to three estimators of , $T_n^{(1)}$ , $T_n^{(2)}$, $T_n^{(3)}$ , with the same type of bias.

**Defintion 3.1.** Given three biased estimators of $\theta$, $T_n^{(1)}$, $T_n^{(2)}$ and $T_n^{(3)}$, such that

$$E[T_n^{(i)} - \theta] = d_1(\theta)\varphi_1^{(i)}(n) + d_2(\theta)\varphi_2^{(i)}(n), \quad i = 1, 2, 3,$$

Generalized Jackknife statistic (of order 2) is given by the equation on the right with ||A|| denoting, as usual, the determinant of the matrix A.

$$\longrightarrow \quad T_n^{GJ} := \frac{\begin{Vmatrix} T_n^{(1)} & T_n^{(2)} & T_n^{(3)} \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{Vmatrix}}{\begin{Vmatrix} 1 & 1 & 1 \\ \varphi_1^{(1)} & \varphi_1^{(2)} & \varphi_1^{(3)} \\ \varphi_2^{(1)} & \varphi_2^{(2)} & \varphi_2^{(3)} \end{Vmatrix}},$$

# Generalized Jackknife Method

The class of estimators are found by the equation below:

- The choice of $\delta$ equal to 0.25
- $\delta$ is regarded as a tuning parameter

$$\hat{\theta}_n^{GJ(\delta)}(k) := \frac{(\delta^2+1)\hat{\theta}_n^N([\delta k]+1) - \delta(\hat{\theta}_n^N([\delta^2 k]+1) + \hat{\theta}_n^N(k))}{(1-\delta)^2}$$

# Simulated Distributional Behavior of the Generalized Jackknife Extremal Index Estimator

Relative Efficiency ($REFF_0^{GJ}$) is defined defined as

$$REFF_0 \equiv REFF_0^{GJ} = \sqrt{\frac{MSE_s[\hat{\theta}_{n0}^N]}{MSE_s[\hat{\theta}_{n0}^{GJ}]}},$$

Bias reduction Indicator ($BRI_0^{GJ}$)

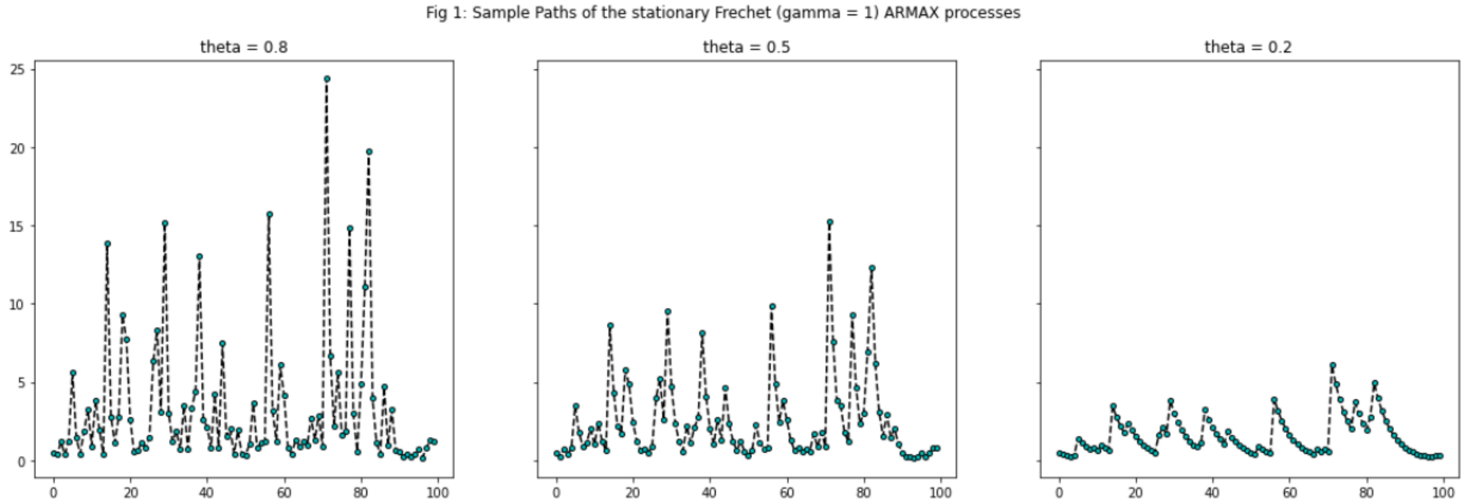
$$BRI_0 \equiv BRI_0^{GJ} = \left|\frac{BIAS_s[\hat{\theta}_{n0}^N]}{BIAS_s[\hat{\theta}_{n0}^{GJ}]}\right|$$

Indicator to illustrate the loss of sensitivity of the new estimators to the choice of the level k

$$STI^{GJ} = \frac{\sum_{k=1}^{n-1} I_{\{|E_s(\hat{\theta}_n^{GJ}(k))-\theta|\leqslant 0.01\}}}{\sum_{k=1}^{n-1} I_{\{|E_s(\hat{\theta}_n^{N}(k))-\theta|\leqslant 0.01\}}}$$

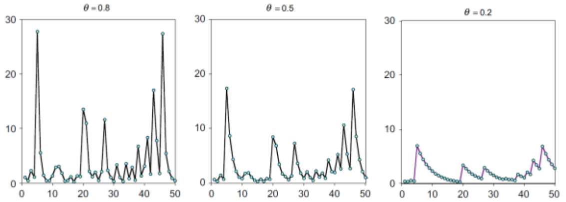
**Note: Higher the indicators, the better the GJ estimator.**

# Fig 1: The Armax Simulation

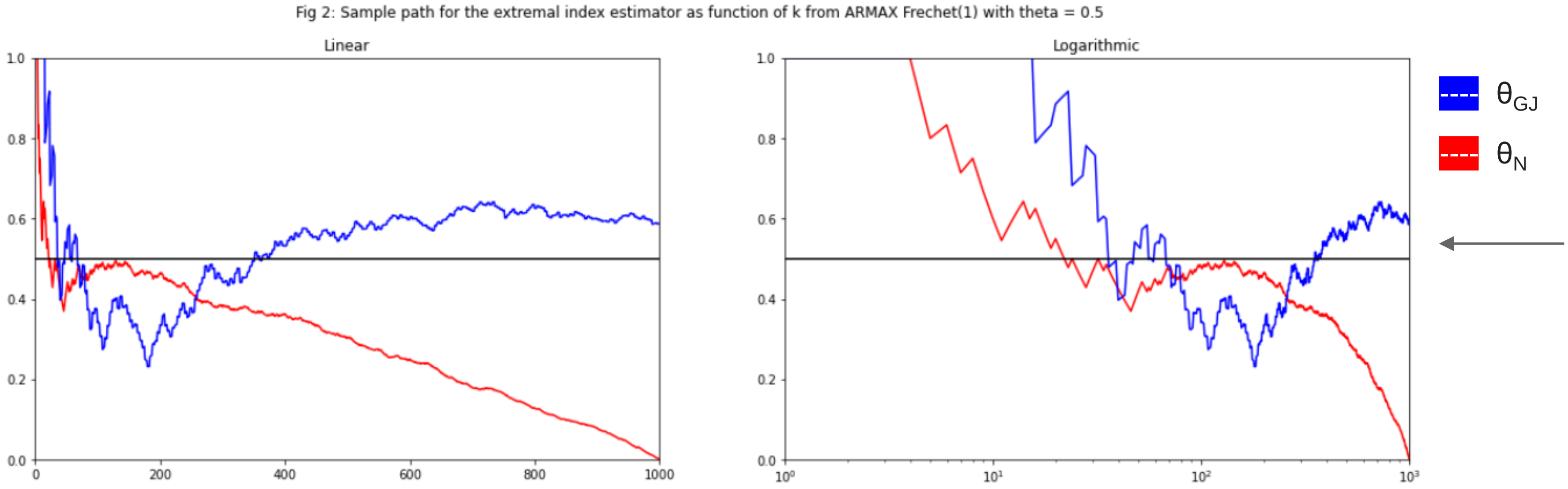



Our results

The results in the paper

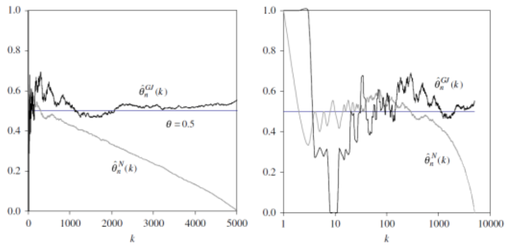


# Fig 2: Sample path of extremal index estimator

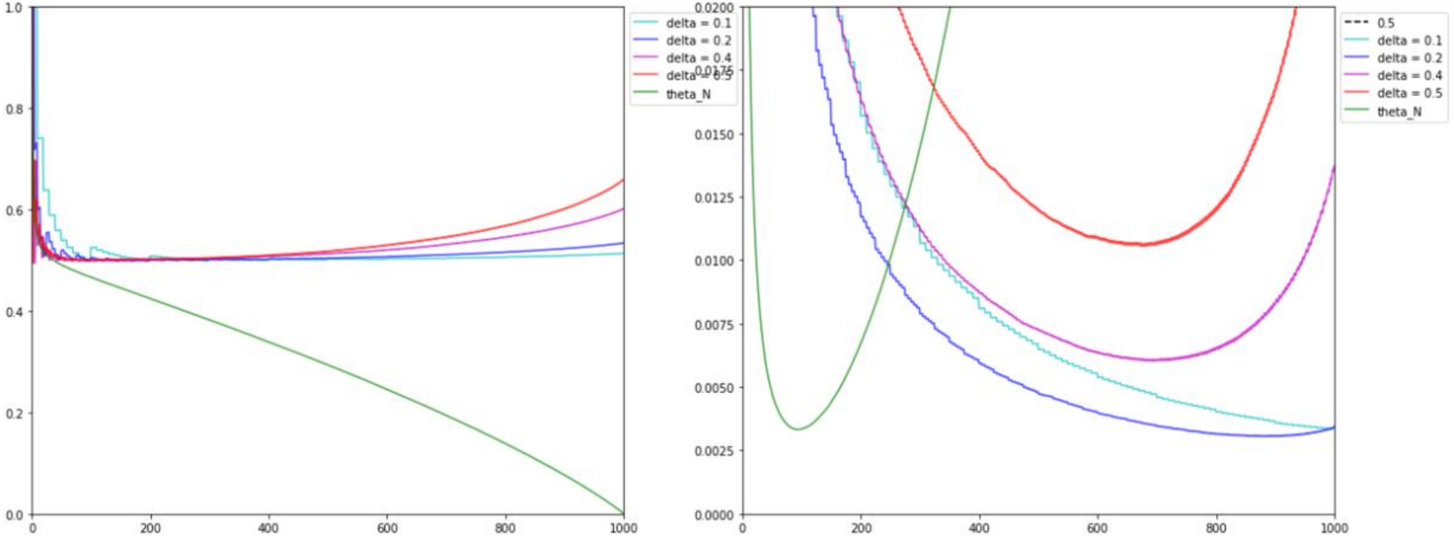



**Our results**

**The results in the paper**

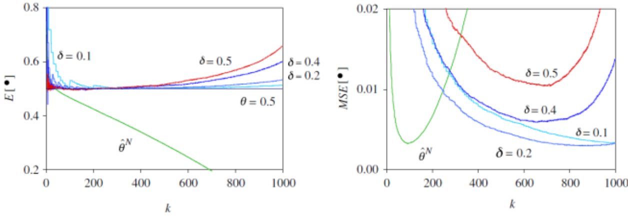


# Fig 3: Variation with delta, Fréchet(1) theta = 0.5



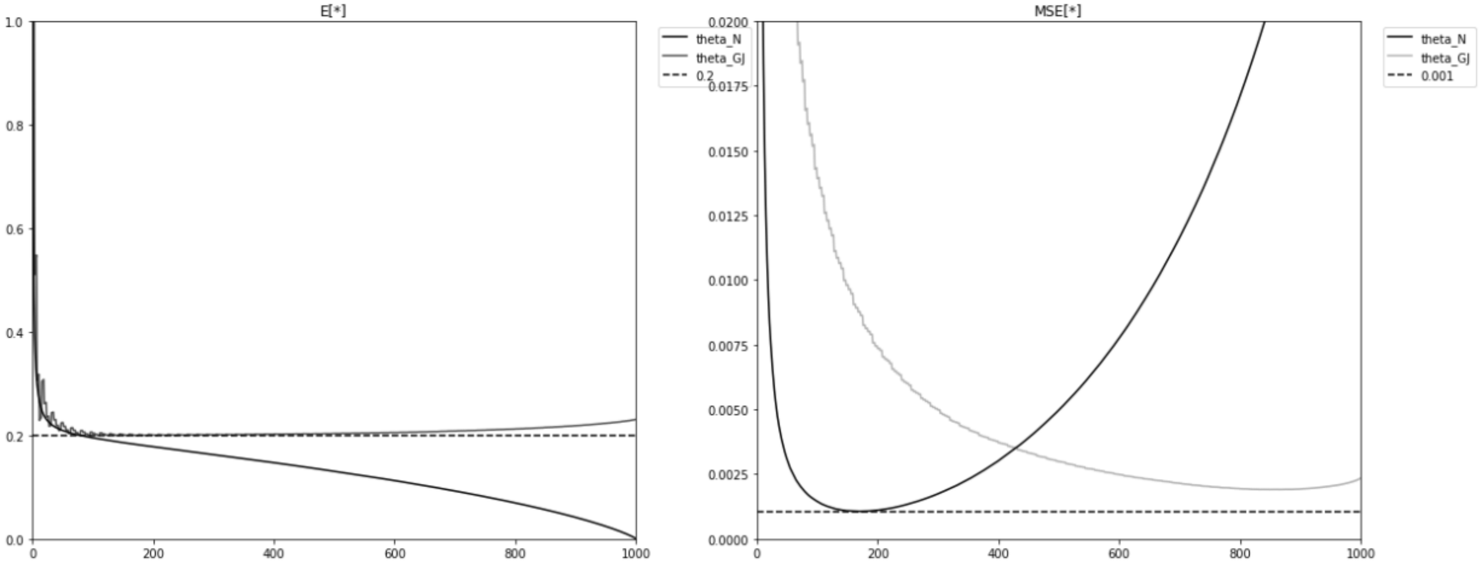

Our results

The results in the paper



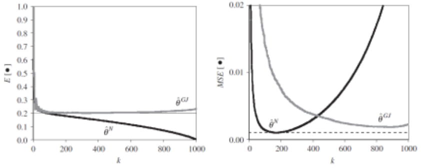
# Fig 4: Mean and MSE, Fréchet(1) theta = 0.2

Figure 4: E[*] and MSE[*] for a Fretchet sequence (gamma = 1) and theta = 0.2



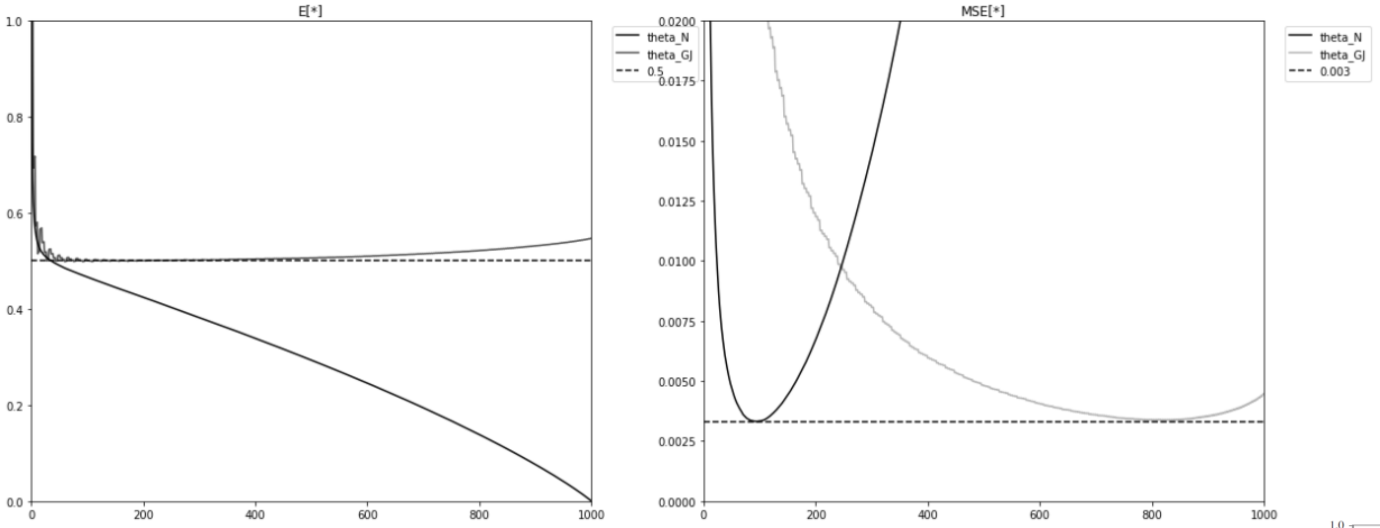

Our results

The results in the paper



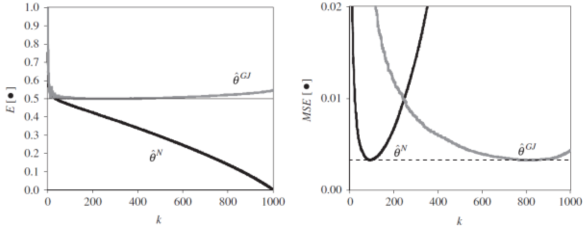
# Fig 5: Mean and MSE, Fréchet(1) theta = 0.5

Figure 5: E[*] and MSE[*] for a Fretchet sequence (gamma = 1) and theta = 0.5



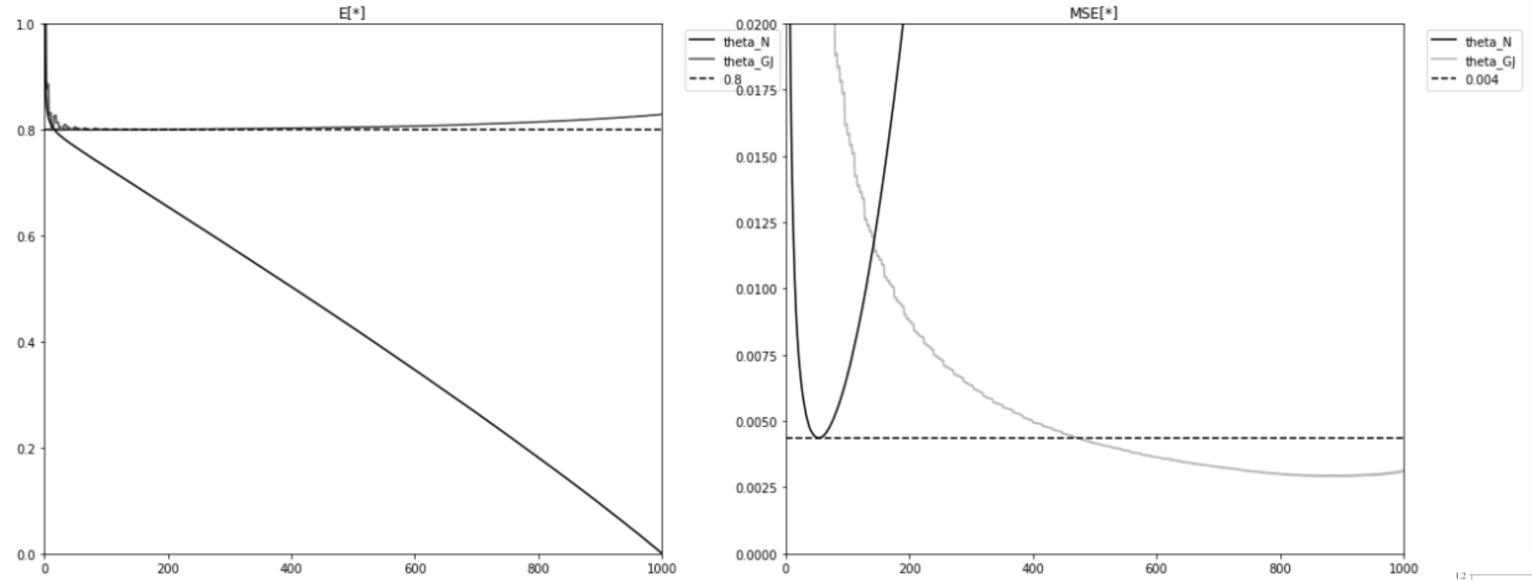

Our results

The results in the paper

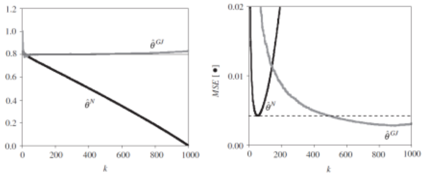


# Fig 6: Mean and MSE, Fréchet(1) theta = 0.8

Figure 6: E[*] and MSE[*] for a Fretchet sequence (gamma = 1) and theta = 0.8




Our results




The results in the paper

# Table 1: Simulated values (theta = 0.2)

| | N = 100 | N = 200 | N = 500 | N = 1000 | N = 2000 |
|---|---|---|---|---|---|
| **k/n** | 0.3600 | 0.2800 | 0.2180 | 0.1720 | 0.1365 |
| **k_gj/n** | 0.9700 | 0.9850 | 0.9300 | 0.8690 | 0.8085 |
| **E_n** | 0.1698 | 0.1761 | 0.1787 | 0.1825 | 0.1858 |
| **E_GJ** | 0.2227 | 0.2247 | 0.2205 | 0.2164 | 0.2127 |
| **MSE_n** | 0.0038 | 0.0026 | 0.0016 | 0.0011 | 0.0007 |
| **MSE_GJ** | 0.0136 | 0.0074 | 0.0034 | 0.0019 | 0.001 |
| **REFF** | 0.5255 | 0.5919 | 0.6786 | 0.7425 | 0.822 |
| **BRI** | 1.3324 | 0.9698 | 1.0419 | 1.0686 | 1.1132 |
| **STI** | 3.1429 | 7.4286 | 8.4474 | 8.6418 | 9.0440 |

# Table 1: Simulated values (theta = 0.5)

| | N = 100 | N = 200 | N = 500 | N = 1000 | N = 2000 |
|---|---|---|---|---|---|
| **k/n** | 0.2100 | 0.1600 | 0.1220 | 0.0950 | 0.0780 |
| **k_gj/n** | 0.9700 | 0.9650 | 0.8980 | 0.8170 | 0.7445 |
| **E_n** | 0.4419 | 0.4530 | 0.4609 | 0.4683 | 0.4737 |
| **E_GJ** | 0.5310 | 0.5348 | 0.5296 | 0.5224 | 0.5176 |
| **MSE_n** | 0.0130 | 0.0087 | 0.0051 | 0.0034 | 0.0021 |
| **MSE_GJ** | 0.0232 | 0.0129 | 0.0060 | 0.0034 | 0.0019 |
| **REFF** | 0.7476 | 0.8217 | 0.9207 | 0.9956 | 1.0599 |
| **BRI** | 1.8731 | 1.3531 | 1.3203 | 1.4162 | 1.4959 |
| **STI** | 20.00 | 20.00 | 20.9231 | 22.76 | 21.6481 |

# Table 1: Simulated values (theta = 0.8)

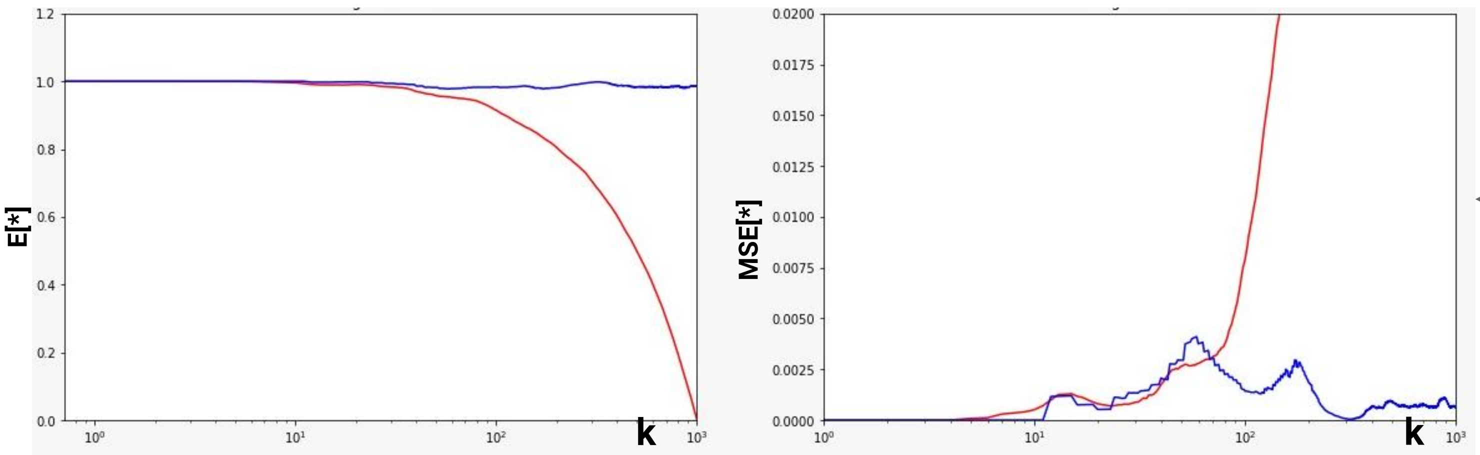
| | N = 100 | N = 200 | N = 500 | N = 1000 | N = 2000 |
|---|---|---|---|---|---|
| **k/n** | 0.1200 | 0.0950 | 0.0660 | 0.0530 | 0.0430 |
| **k_gj/n** | 0.9700 | 0.9850 | 0.9700 | 0.9050 | 0.8105 |
| **E_n** | 0.7344 | 0.7443 | 0.7592 | 0.7654 | 0.7718 |
| **E_GJ** | 0.8129 | 0.8203 | 0.8228 | 0.8199 | 0.8148 |
| **MSE_n** | 0.0195 | 0.0123 | 0.0068 | 0.0043 | 0.0027 |
| **MSE_GJ** | 0.0227 | 0.0119 | 0.0053 | 0.0029 | 0.0016 |
| **REFF** | 0.9256 | 1.0175 | 1.1379 | 1.2155 | 1.3037 |
| **BRI** | 5.0710 | 2.7433 | 1.7870 | 1.7364 | 1.9117 |
| **STI** | 66.0000 | 46.6667 | 50.0000 | 52.6923 | 49.1429 |

# Table 2: Asymptotic and simulated optimal sample fractions for Nandagopalan's estimator

| n | Theta = 0.2 | Theta = 0.2 | Theta = 0.5 | Theta = 0.5 | Theta = 0.8 | Theta = 0.8 |
|---|---|---|---|---|---|---|
| | k_ass/n | k/n | k_ass/n | k/n | k_ass/n | k/n |
| 100 | 0.3816 | 0.3600 | 0.2071 | 0.2100 | 0.1156 | 0.1200 |
| 200 | 0.3029 | 0.2800 | 0.1644 | 0.1600 | 0.0917 | 0.0950 |
| 500 | 0.2231 | 0.2180 | 0.1211 | 0.1220 | 0.0676 | 0.0660 |
| 1000 | 0.1771 | 0.1730 | 0.0961 | 0.0950 | 0.0536 | 0.0530 |
| 2000 | 0.1406 | 0.1365 | 0.0763 | 0.0780 | 0.0426 | 0.0430 |

$$k_{0n}^{N} \sim \left\{ \frac{2(1-\theta)n^2}{\theta(1+\theta)^2} \right\}^{1/3} =: k_{\mathrm{ass}}^{N},$$

# Fig 7: Simulation from Fréchet(1) i.i.d. framework

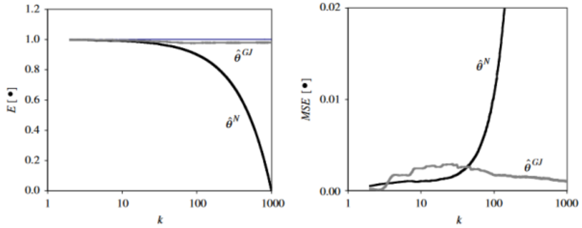



$\theta_N$

$\theta_{GJ}$

Our results

The results in the paper



# Autoregressive Process of Order 1 (AR~1)

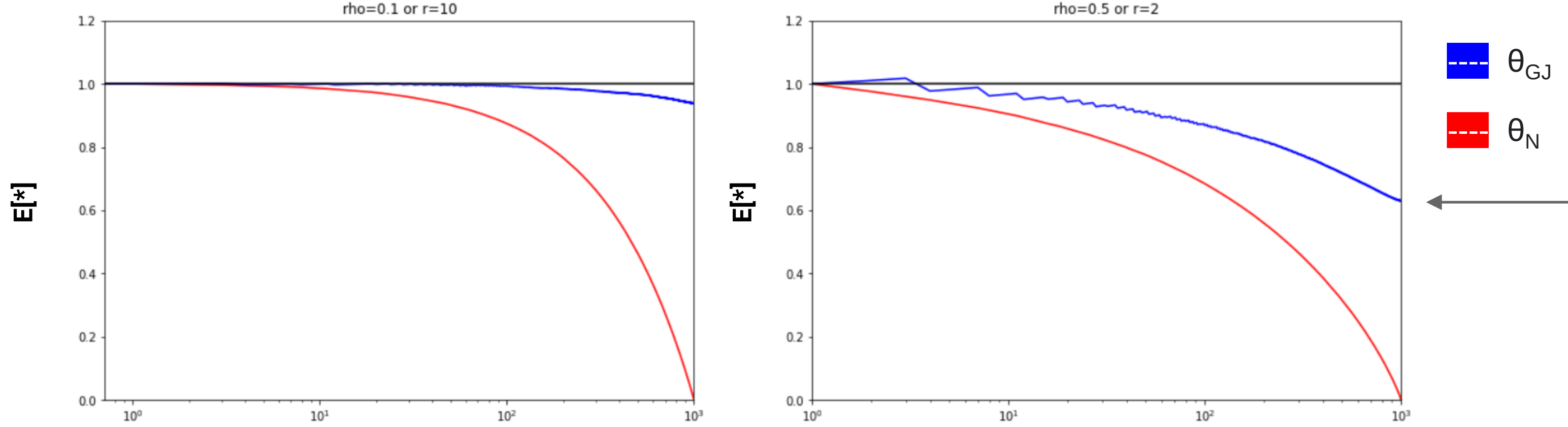
We have simulated for sample size n = 1000 , the class of exponential autoregressive processes of order 1 with with $\{\varepsilon_j\}_{j>=1}$ standard exponential random variables. For this type of sequences we have **$\theta$ = 1 and D''** satisfied.

$$\boxed{X_j = \rho X_{j-1} + \varepsilon_j, \quad j \geqslant 1, \; X_0 \frown \text{Exponential}(1),}$$

In next slide, we present the simulated mean values of the **extremal index estimators** on a logarithmic scale, for the **autoregressive processes** in the equation above, with $\theta$ **= 0.1 and 0.5.**

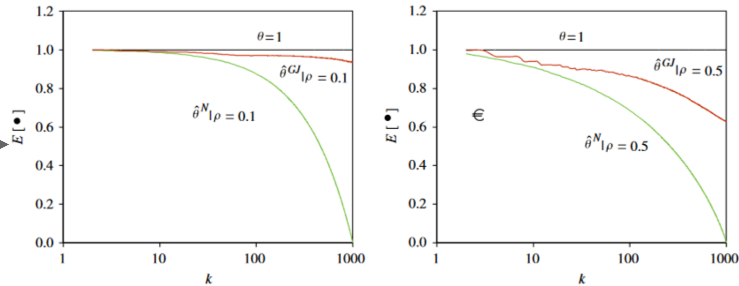
# Fig 8: AR(1) Process

Fig 8: Simulated mean values of the extremal index estimators under study on a logarithmic scale, for samples of sizen=1000 from the AR processes in (20), with rho=0.1 (left) and rho=0.5 (right).




Our results

The results in the paper



# Autoregressive Process ($AR_r$)

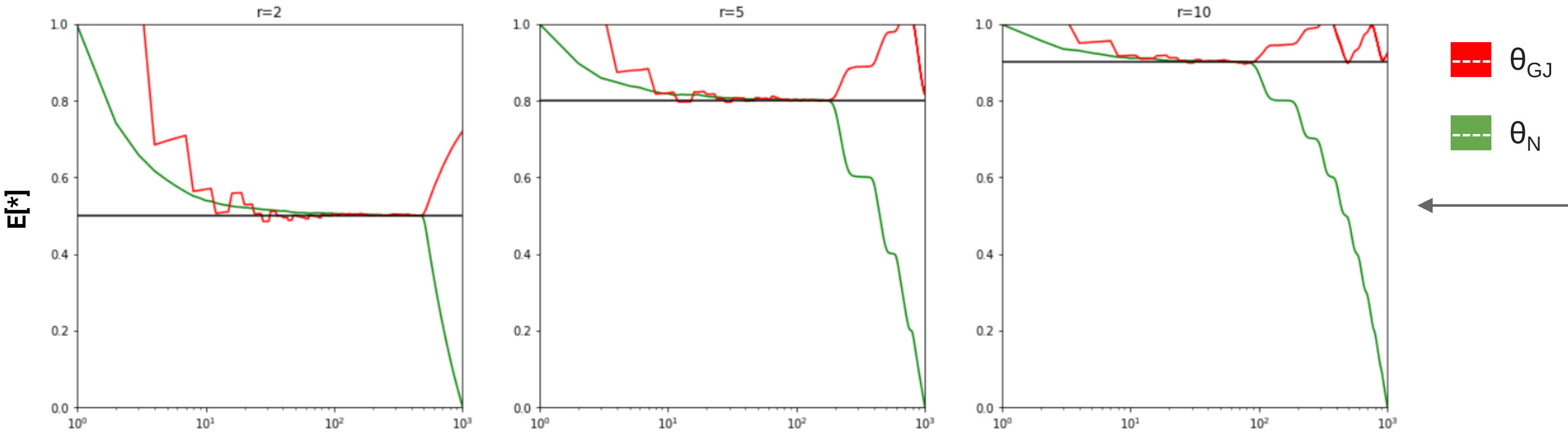
For a fixed $r>=1$, and a sequence of i.i.d. Random variables $\{\varepsilon_n\}_{n>=1}$ such that $P(\varepsilon_1 = k/r) = 1/r$, $k = 0, 1, \ldots, r-1$. We consider the process below:

$$\boxed{X_j := \frac{1}{r} X_{j-1} + \varepsilon_j, \quad j \geqslant 1 \quad \text{and} \quad X_0 \frown \text{Uniform}(0, 1).}$$

In next slide, we present the simulated mean values of the extremal index estimators on a logarithmic scale, for samples of size **n = 1000** from the **$AR_r(1)$** processes in the equation above with **r = 2, 5 and 10.**

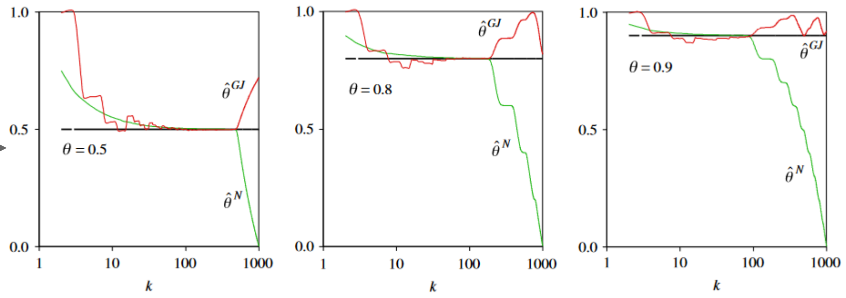
# Fig 9 $AR_r(1)$

Fig. 9. Simulated mean values of the extremal index estimators under study in a lnk-scale, for samples of size n=1000 from the ARr(1) processes in (21), withr=2 (left),r=5 (center) and r=10 (right)




Our results

The results in the paper



# Subsampling

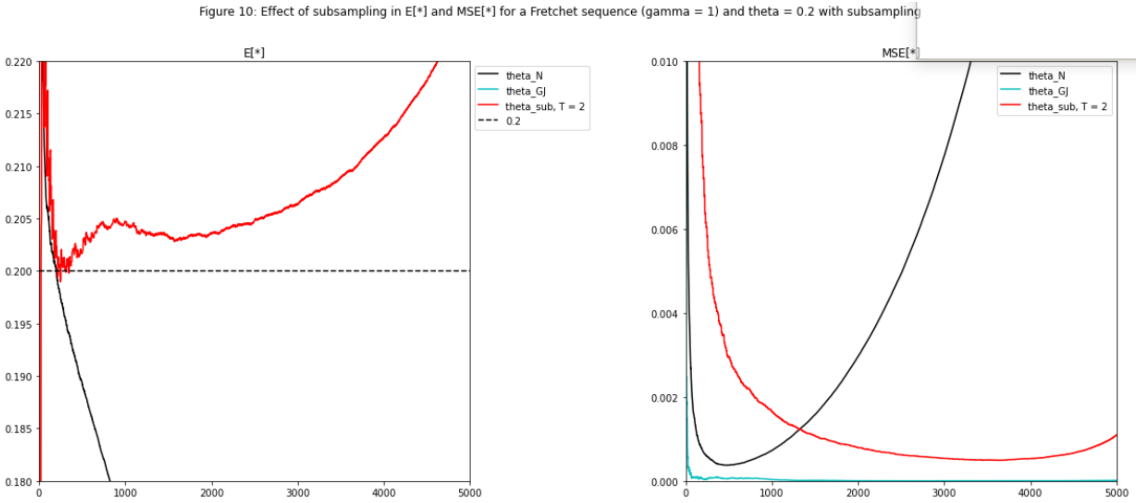
Subsampling may thus improve the performance of an estimator, through the consideration of averages, which often enable a decrease in variance, and consequently in mean squared error.

The following sub-sampling algorithm has been used in the paper:

- fix $T$ (possibly $T = 2$), and compute $r = [n/T]$;
- consider the $T$ subsamples of size $r$, $\mathbf{V}_i = (X_i, X_{T+i}, \ldots, X_{(r-1)T+i})$, for $i = 1, 2, \ldots, T$, and compute the estimates $\hat{\theta}^{GJ}_{\mathbf{V}_i}(j)$, $j = 1, 2, \ldots, r-1$, with $\hat{\theta}^{GJ}_{\mathbf{V}}$, the estimator in (16) with $\delta = \frac{1}{4}$, applied to the subsample under consideration;
- compute $\hat{\theta}^{GJ}_{\text{sub}|T}((j-1)T+1) = 1 - \frac{1}{T}\sum_{i=1}^{T}(1 - \hat{\theta}^{GJ}_{\mathbf{V}_i}(j))^{1/T}$, $j = 1, 2, \ldots, r-1$. Next, fill the gaps, considering that, for any value $k \in ((j-1)T+1,\ jT]$, $\hat{\theta}^{GJ}_{\text{sub}|T}(k) = \hat{\theta}^{GJ}_{\text{sub}|T}((j-1)T+1)$.

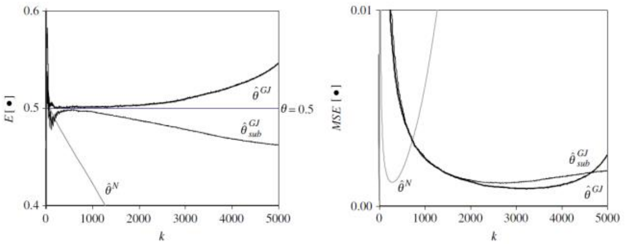
If $\theta$ <= 0.2, we are able to overpass the original estimator at optimal levels, when we consider the Generalized Jackknife statistic in with δ=1/4,  together with the use of subsampling techniques with T= 2, 3 and 4. A graph with representation along with the table has been computed for the same.

# Fig 10: Effect of subsampling, Fréchet(1) theta 0.5



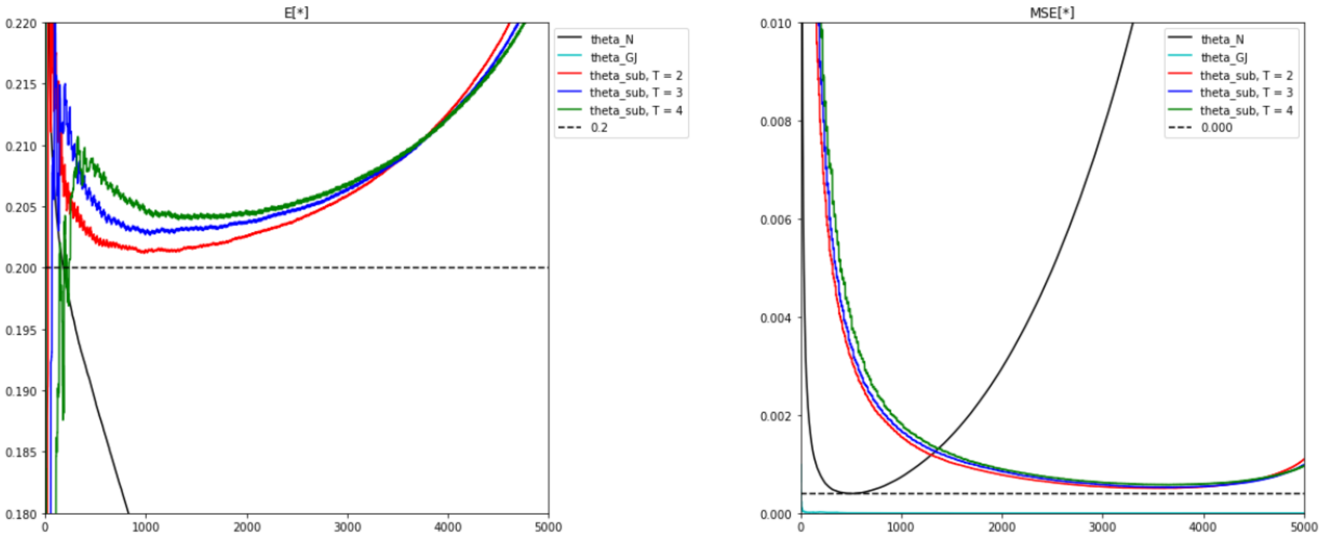

Our results

The results in the paper



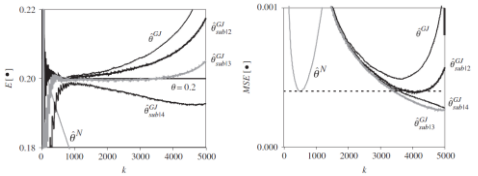
# Fig 11: Effect of subsampling, Fréchet(1) theta 0.2

Figure 11: E[*] and MSE[*] for a Fretchet sequence (gamma = 1) and theta = 0.2 with subsampling




Our results

The results in the paper



| | N = 1000 | N = 2000 |
|---|---|---|
| REFF, no subsample | 0.7428 | 0.8220 |
| REFF, T = 2 | 0.7348 | 0.8089 |
| REFF, T = 3 | 0.7031 | 0.7729 |
| REFF, T = 4 | 0.6437 | 0.7263 |
| BRI, no subsample | 1.1668 | 1.1132 |
| BRI, T = 2 | 1.0539 | 1.1589 |
| BRI, T = 3 | 1.0313 | 1.1105 |
| BRI, T = 4 | 0.9122 | 1.0140 |
| STI, no subsample | 8.6418 | 9.0440 |
| STI, T = 2 | 7.6667 | 8.6026 |
| STI, T = 3 | 6.8462 | 7.9108 |
| STI, T = 4 | 3.8462 | 7.8500 |

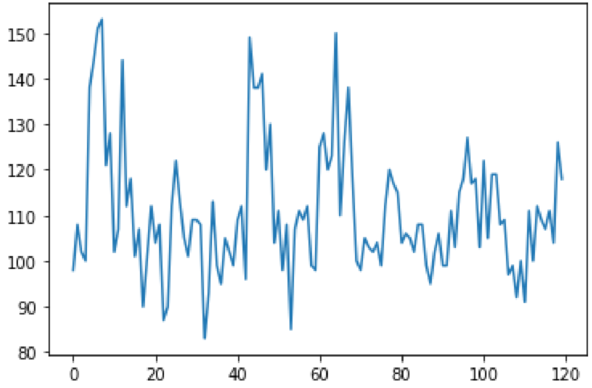
Table 3: REFF, BRI and STI for estimators, theta = 0.2

# Conclusion

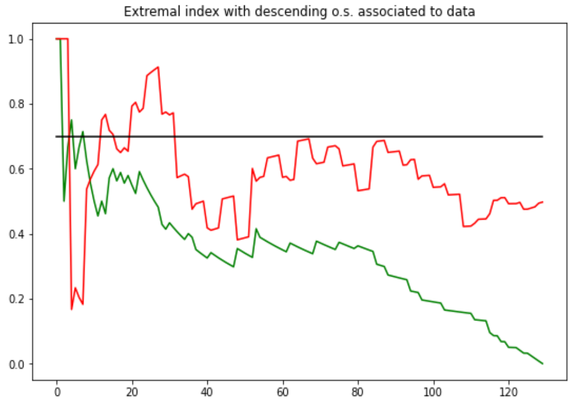
1. Two most attractive features of the proposed Generalized Jackknife estimator of the extremal index are:
   - Stable sample path (for a wide region of k-values), close to the target value
   - Wide “bathtub” patterns of their MSE(k) functions, which make less relevant the choice of the optimal sample fraction $k_0/n$
1. Regarding MSE, the initial GJ estimator does not overpass the original estimator, when both estimators are considered at their optimal levels ($\theta$ is small)
2. For the estimator with $\delta = 1/4$ or its subsampling variants with a moderate value of T, both the BRI and the STI indicators are greater than 1 for all simulated models and sample sizes
3. The above insensitivity of the mean value (and sample path) to changes in k is indeed a new feature of extremal index estimators

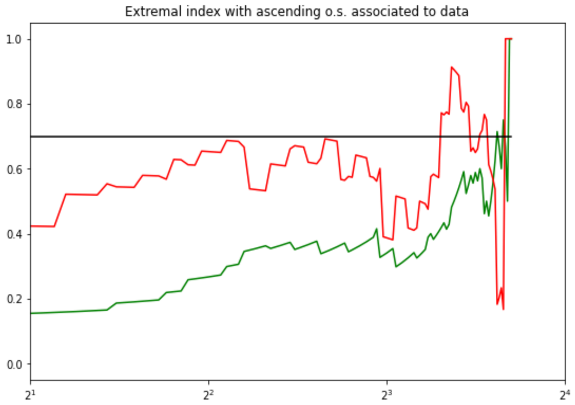
# Ozone Case Study

- Performance of the estimators has been examined for n = 120 weekly maxima of hourly averages of ozone concentrations measured in parts per million




Data used




Result: $\theta$ converges to 0.7 for GJ estimator

$\theta_n$

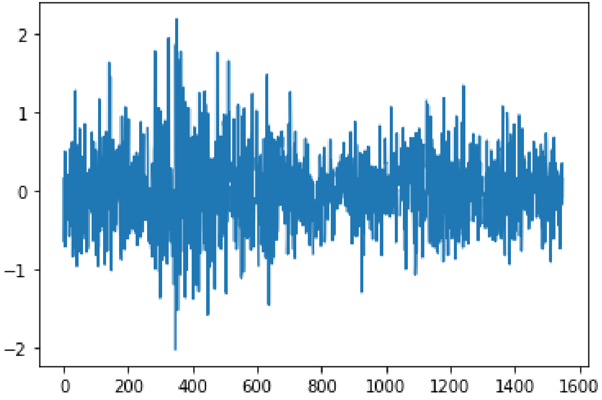
$\theta_{GJ}$

# Ozone Case Study: Conclusions

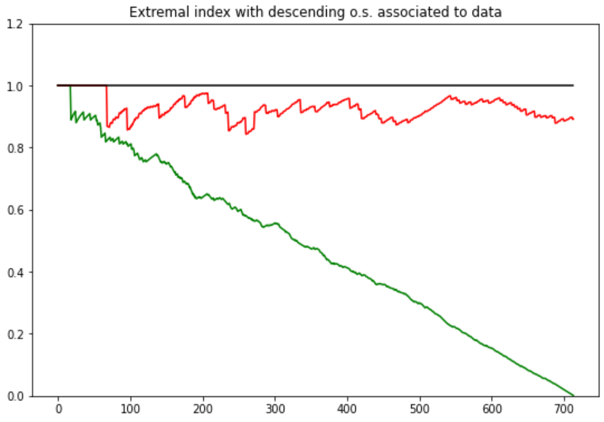
1. Higher stability of the GJ estimates, around a value close to 0.7, is clear
2. Stability appears only for reasonably large values of k, whereas for the classical estimates, the sample path exhibits a very small region of stability around the value  = 0.7 (for small values of k)
3. The estimates seem to work in a way similar to the one achieved for data under conditions D and D''
4. Provide heuristic support for the validity of the above-mentioned conditions

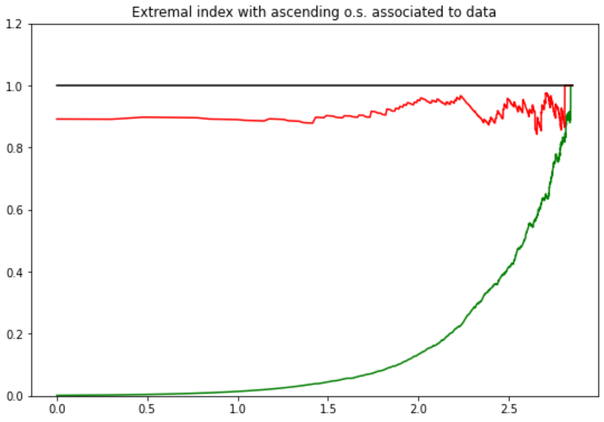
# Financial Log returns Case study

Performance of the above mentioned estimators in the analysis of the Euro–UK pound daily exchange rates from January 4, 1999, until December 14, 2004.

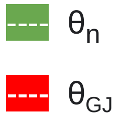



Data used




Result: $\theta$ converges to 1 for GJ estimator



# Financial Log Returns: Conclusions

1. Higher stability of the Generalized Jackknife estimates, around a value close to 1, is clear
2. The stability appears both for small and large values of k
3. The value of extremal index (=1) could be interpreted as efficiency of the financial market under study and would enable us to deal with the sample of log-returns as approximately i.i.d.
4. However, essentially because of their varying volatility, financial time series are very unlikely to satisfy condition D
5. Thus, a derived estimate =1 can be a consequence of assuming D instead of a genuine feature of the process, as we expect an overestimation of $\theta$ in case condition D'' does not hold

# References


- https://www.sciencedirect.com/science/article/pii/S0167947305000733
- https://rss.onlinelibrary.wiley.com/doi/epdf/10.1111/1467-9868.00223
- https://link.springer.com/article/10.1023/A:1016592028871
- https://link.springer.com/article/10.1023/A:1011470010228
- https://www.sciencedirect.com/science/article/pii/S0378375803001940
- http://emis.dsd.sztaki.hu/journals/PM/59f4/pm59f402.pdf
- https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9892.00320