Department of Decision Sciences, Bocconi University

An Introduction to Extreme Value Theory: from Basic Results to Tail Risk Inference in Time Series

COSMOSTATS 2023

Simone Padoan simone.padoan@unibocconi.it

13 September 2023

- Let $Y \sim F$ and
$$Q(1-p) := F^{\leftarrow}(1-p)$$
be the **$(1-p)$-quantile**, where $F^{\leftarrow}(x) = \inf\{y : F(y) \geq x\}$.

Remark

- In applied fields (finance, risk management etc), the quantile is called **Value at Risk (VaR)** and it is seen as a **risk measure** to quantify the intensity of **tail risk events**, setting p as small value.
- Let $Y_1, \ldots, Y_n$ be i.i.d. rvs with **unknown** distribution F. The estimation of Q is not obvious when $\boldsymbol{p}$ is **very small**, e.g. $p <= 1/n$.
- This is especially true when working with $X_1, \ldots, X_n$ **dependent** rvs.

- **What to do?**
We can assume to know F ... or in alternative rely on the **EVT**.

Review

Basic results of the univariate Extreme Value Theory (EVT)

- Let $Y_1, \ldots, Y_m$ be i.i.d. rvs with distribution F.
- Assume $F \in \mathcal{D}(G_\gamma)$, i.e. F belongs to the **DoA** of the `Generalised Extreme Value (GEV)` distribution G_γ.
- If there are norming functions $a(\cdot) > 0$ and $b(\cdot)$, such that

$$\lim_{n \to \infty} \mathbb{P}\left(\frac{\max(Y_1, \ldots, Y_m) - b(m)}{a(m)} \leq z \right) = G_\gamma(z), \qquad (1)$$

then

$$G_\gamma(z) = \begin{cases} \exp\left(-(1+\gamma z)_+^{-1/\gamma} \right), & \gamma \neq 0, \\ \exp\left(-\exp(-z)\right), & \gamma = 0, \end{cases}$$

- G_γ is: **short**- ($\gamma < 0$), **light**- ($\gamma = 0$) or **heavy-tailed** ($\gamma > 0$).

(1) Set $y = a(m)z + b(m)$ in (1). From

$$F(y) \approx G_\gamma^{1/m}\left(\frac{y - b(m)}{a(m)}\right), \quad n \to \infty,$$

and $1 - p = F(y)$, then for small enough p,

$$\begin{aligned} Q(1-p) &\approx b(m) + a(m)\frac{(-m\log(1-p))^{-\gamma} - 1}{\gamma} \\ &\approx b(m) + a(m)\frac{(mp)^{-\gamma} - 1}{\gamma} \end{aligned}$$

(2) **BM** approach suggests:

- Divide the sample of $Y_1, \ldots, Y_n$ rvs into k blocks of m rvs. Assume $k = k(n)$, $m = m(n)$, and $k \to \infty$ and $m \to \infty$ as $n \to \infty$ and $m = o(n)$ and $k = o(n)$.
- Compute k maxima to estimate $b(n/k)$, $a(n/k)$ and γ. Suitable estimators are: MLE, GPWM, etc. (**Jenkinson, 1969; Hosking et al., 1985**).

(3) Let $p := p(n)$ such that $p \to 0$ and $np \to c \geq 0$ as $n \to \infty$. Let $\tau_n' = 1 - p$ be an `extreme level`. Then, estimator for an `extreme quantile` is

$$\widehat{q}_{\tau_n'} = \widehat{b}_n(n/k) + \widehat{a}_n(n/k) \frac{\left(\frac{np}{k}\right)^{-\widehat{\gamma}_n} - 1}{\widehat{\gamma}_n}.$$

- $F \in \mathcal{D}(G_\gamma)$ can be equivalently formulated as follows.
- Let $y^* = \sup\{y : F(y) < 1\}$. For $u < y^*$, there is a scaling function $s(\cdot) > 0$ such that

$$\lim_{u \uparrow y^*} \mathbb{P}\left(\frac{Y-u}{s(u)} \leq z \middle| Y > u \right) = H_\gamma(z) \tag{2}$$

then

$$H_\gamma(z) = \begin{cases} 1 - (1+\gamma z)_+^{-1/\gamma}, & \text{if } \gamma \neq 0, \\ 1 - \exp(-z), & \text{if } \gamma = 0, \end{cases}$$

is the `Generalised Pareto (GP)` distribution (**de Haan and Ferreira, 2006, Ch.1**)

(1) Set $y = u + s(u)z$ in (2). From

$$F(y) \approx 1 - (1 - F(u))\left(1 - H_\gamma\left(\frac{y-u}{s(u)}\right)\right), \quad u \to y^*,$$

and $1 - p = F(y)$, then for small enough p,

$$Q(1-p) \approx u + s(u)\frac{\left(\frac{p}{1-F(u)}\right)^{-\gamma} - 1}{\gamma}.$$

(2) **Peaks-over-Threshold (PoT)** approach suggests:

- Set u so we can work with k **excess** variables from $Y_1, \ldots, Y_n$.
- A possibility is $u = Q(\tau_n)$, where $\tau_n = 1 - k/n$ is an `intermediate level`, with can be estimated by $X_{n-k,n}$.
- Approximate $1 - F(u) \approx 1 - F_n(X_{n-k,n}) = k/n$.
- Set $s(u) = a(1/(1 - F(u)))$, then $s(Q(\tau_n)) = a(n/k)$.
- Use the k **excesses** to estimate γ and $a(n/k)$. Suitable estimators are: method of moments, MLE, etc. (de Haan and Ferreira, 2006, Ch. 2,3)

(3) Then, estimator for an `extreme quantile` is

$$\widehat{q}_{\tau_n'} = X_{n-k,n} + \widehat{a}_n(n/k) \frac{\left(\frac{np}{k}\right)^{-\widehat{\gamma}_n} - 1}{\widehat{\gamma}_n}.$$

A first step forward

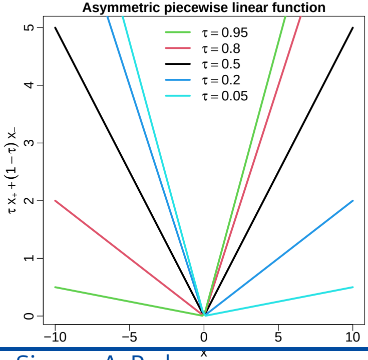
An alternative risk measure

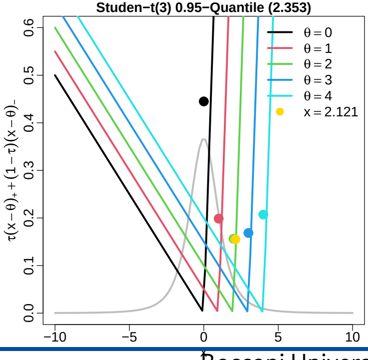
- The τth quantile of Y is the minimizers of an **asymmetric piecewise linear loss** function (Koenker and Bassett, 1978), i.e.

$$q_\tau \in \arg\min_{\theta \in \mathbb{R}} \tau \mathbb{E}\{(X-\theta)_+\} + (1-\tau)\mathbb{E}\{(X-\theta)_-\},$$

for $\tau \in (0,1)$, with the **median** obtained with $\tau = 1/2$.

- In other words



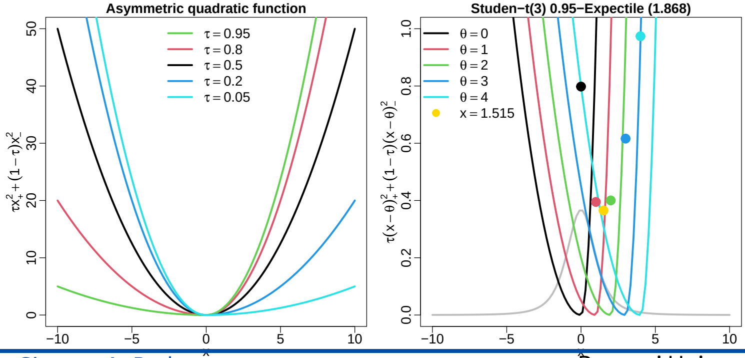


- The τth expectile of Y is the minimizer of an **asymmetric quadratic loss** function (Newey and Powell, 1987), i.e.

$$\xi_\tau = \arg\min_{\theta\in\mathbb{R}} \tau\mathbb{E}\{(X-\theta)_+^2\} + (1-\tau)\mathbb{E}\{(X-\theta)_-^2\},$$

for $\tau \in (0,1)$, with the **mean** obtained with $\tau = 1/2$.

- In other words



- From the first order condition

$$\tau\mathbb{E}\{(X-\theta)_+\} = (1-\tau)\mathbb{E}\{(X-\theta)_-\}$$

we obtain

$$\tau = \frac{\mathbb{E}(|X-\xi_\tau|\mathbb{1}(X \leq \xi_\tau))}{\mathbb{E}(|X-\xi_\tau|)}.$$

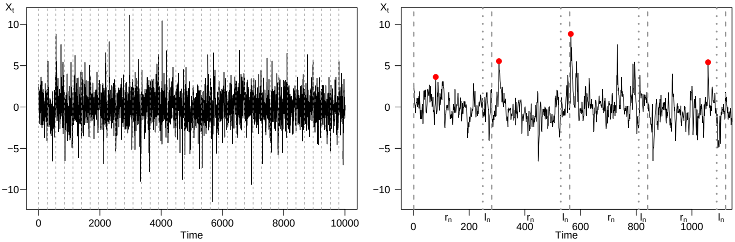
Pro and cons

risk measure property	quantile	expectile
law-invariant	✔	✔
elicitable	✔	✔
coherent	✗	✔
derivation	**tail probability**	**tail expectation**
constraints	$\gamma \in \mathbb{R}$	$\gamma < 1$

A second step forward

We now consider the case when data are temporarly dependent.

- Assume that the sequence $X_1, \ldots, X_n$ is:
  - stationary with marginal distribution $F \in \mathcal{D}(G_\gamma)$.
  - weakly time dependent, i.e. it satisfies a suitable mixing condition (a form of memorylessness property), e.g. **Leadbetter et al., (1983)**
- The classical result apply through the so-called big blocks separated by small blocks framework.



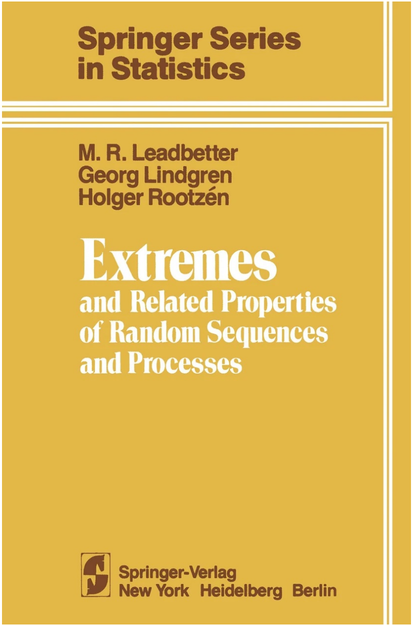
- In this case $F \in \mathcal{D}(G_\gamma)$ means that

$$\mathbb{P}\left(\frac{\max(X_1,\ldots,X_{r_n}) - b(r_n)}{a(r_n)} \leq x\right) \to G_\gamma^\theta(x), \quad n \to \infty,$$

where $\theta \in (0,1]$ is the **extremal index**, i.e. the reciprocal of the asymptotic mean cluster size of exceedances.

- An interpretation is, as $n \to \infty$, we have

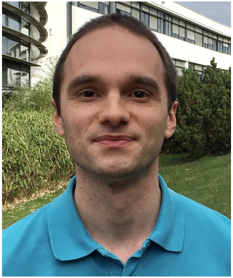
$$\mathbb{P}\left(\max_{1 \leq t \leq m} X_t \leq a(m)x + b(m)\right) \approx \mathbb{P}\left(\max_{1 \leq t \leq \lfloor \theta m \rfloor} Y_t \leq a(m)x + b(m)\right).$$


Springer Series
in Statistics
M. R. Leadbetter
Georg Lindgren
Holger Rootzén
Extremes
and Related Properties
of Random Sequences
and Processes
Springer-Verlag
New York Heidelberg Berlin

Statistical problem

We aim to derive statistical procedure for the extrapolation of **extreme expectile** in a time series framework.

Next part concerns some new results from the paper **Davison, P., Stupfler, (2022)** obtained for the case $\gamma > 0$. Thanks to



- The first step is to focus on the **intermedite level** τ_n:
  - The **Least Asymmetrically Weighted Squares (LAWS) estimator** is

$$\widetilde{\xi}_{\tau_n} = \arg\min_{\theta \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^{n} (\eta_{\tau_n}(X_t - \theta) - \eta_{\tau_n}(X_t)).$$

where $\eta_\tau(x) = |\tau - \mathbb{1}(x \leq 0)|x^2$ is the **expectile check function**.

  - The **expectile-quantile** tail equivalence result

$$\xi_\tau / q_\tau \approx (\gamma^{-1} - 1)^{-\gamma}, \quad \tau \to 1, \tag{3}$$

suggests to use the **quantile-based (QB) estimator** i.e.

$$\widehat{\xi}_{\tau_n} = (\widehat{\gamma}_n^{-1} - 1)^{-\widehat{\gamma}_n} \widehat{q}_{\tau_n}.$$

- From the **quantiles proportionality** result

$$\frac{q_{\tau'}}{q_\tau} \approx \left(\frac{1-\tau'}{1-\tau}\right)^{-\gamma}, \quad \tau' \gg \tau \to 1.$$

together with the **expectile-quantile** tail equivalence result we obtain that an **extrapolating** estimator of an **extreme expectile** is

$$\overline{\xi}^{\star}_{\tau_n'} = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\widehat{\gamma}_n} \overline{\xi}_{\tau_n},$$

where $\overline{\xi}_{\tau_n}$ can be either the **LAWS** or the **QB estimator** of ξ_{τ_n}.

Conditions

1. Let $(X_t)_{t\geq 1}$ be a β-mixing time series (e.g. Doukhan, P., 1994) with a continuous one-dimensional marginal heavy-tailed distribution.
2. Assume $0 < \gamma < 1$ and $\mathbb{E}|(\min(0, X))| < \infty$.
3. There are sequences $l_n := l(n)$, $r_n := r(n)$ such that as $n \to \infty$

$$l_n \to \infty, \quad r_n \to \infty, \quad \frac{l_n}{r_n} \to 0, \quad \frac{r_n}{n} \to 0, \quad \frac{n\beta(l_n)}{r_n} \to 0;$$

4. There exists the limit (tail copula function)

$$R_t(x, y) = \lim_{s \to 0} s^{-1} \mathbb{P}(F(X_1) > 1 - sx, F(X_{t+1}) > 1 - sy);$$

for all $t \geq 1$ and $(x, y) \in [0, \infty]^2 \setminus \{\infty, \infty\}$.

Conditions (continued...)

5. There are $\rho(r) \geq 0$, satisfying $\sum_t \rho(t) < \infty$, $D \geq 0$ such that

$$s^{-1}\mathbb{P}(F(X_1) > 1 - su, F(X_{t+1}) > 1 - sv) \leq \rho(t)\sqrt{uv} + sDuv,$$

as $s \longrightarrow 0$, for all $t \geq 1$ and $u, v \in [0, 1]$.

6. There is a measurable function A such that $A(s) \longrightarrow 0$ as $s \longrightarrow \infty$ and for $\rho \leq 0$ and all $y > 0$,

$$\lim_{s\to\infty} \frac{1}{A(1/\overline{F}(s))}\left[\frac{\overline{F}(sy)}{\overline{F}(s)} - y^{-1/\gamma}\right] = y^{-1/\gamma}\frac{y^{\rho/\gamma} - 1}{\gamma\rho}.$$

When $\rho = 0$, the right-hand reads as $y^{-1/\gamma}\log(y)/\gamma^2$.

7. $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \longrightarrow \lambda \in \mathbb{R}$ as $n \longrightarrow \infty$.

Theorem 1 (Davison, P., Stupfler, 2022)

Assume that $n(1-\tau_n) \to \infty$, $n(1-\tau_n') \to c \in [0,\infty)$, $\sqrt{n(1-\tau_n)}/\log((1-\tau_n)/(1-\tau_n')) \to \infty$ and $r_n(1-\tau_n) \to 0$ as $n \to \infty$. Then, under Conditions (1)–(7) one has

$$\frac{\sqrt{n(1-\tau_n)}}{\log((1-\tau_n)/(1-\tau_n'))} \log \frac{\overline{\xi}^{\star}_{\tau_n'}}{\xi_{\tau_n'}} \xrightarrow{d} \mathcal{N}\left(b,\, W(\gamma, R)\right),$$

where

$$b = \frac{\lambda}{1-\rho},$$

$$W(\gamma, R) = \gamma^2 \left(1 + 2\sum_{t\geq 1} R_t(1,1)\right).$$

Lemma 1 (Davison, P., Stupfler, 2022)

Under the conditions of Theorem 2 we have

$$\frac{1}{r_n(1-\tau_n)} \operatorname{Var}\left(\sum_{t=1}^{r_n} \mathbb{1}\{F(X_t) > \tau_n\}\right) \xrightarrow{n\to\infty} 1 + 2\sum_{t=1}^{\infty} R_t(1,1).$$

- Using "big blocks separated by small blocks" arguments we compute

$$Y_j = \sum_{t=1+j\ell_n}^{r_n+j\ell_n} \mathbb{1}(\widehat{F}_n(X_t) > \tau_n)$$

  for $j = 0, 1, \ldots, m_n - 1$, $m_n = \lfloor n/\ell_n \rfloor$ and $\ell_n = r_n + l_n$.
- Let S_n^2 be the empirical variance of $Y_0, \ldots, Y_{m_n-1}$. Then,

$$\widehat{W}_n(\gamma, R) = (r_n(1-\tau_n))^{-1} \widehat{\gamma}_n^2 \cdot S_n^2.$$

- For the bias term, we assume $A(s) = \gamma\beta s^{\rho}$. Given the estimates $\widehat{\beta}_n$ and $\widehat{\rho}_n$ and noting that $\lambda \approx \sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})$, then

$$\widehat{b}_n = \frac{\sqrt{n(1-\tau_n)}\widehat{\gamma}_n\widehat{\beta}_n(1-\tau_n)^{-\widehat{\rho}_n}}{1-\widehat{\rho}_n}$$

- Concluding, for large n,

$$\left[\overline{\xi}^{\star}_{\tau_n'}\left(\frac{1-\tau_n}{1-\tau_n'}\right)^{-\widehat{b}_n - z_{\alpha/2}\sqrt{\frac{\widehat{W}_n(\gamma,R)}{[n(1-\tau_n)]}}}, \overline{\xi}^{\star}_{\tau_n'}\left(\frac{1-\tau_n}{1-\tau_n'}\right)^{-\widehat{b}_n + z_{1-\alpha/2}\sqrt{\frac{\widehat{W}_n(\gamma,R)}{[n(1-\tau_n)]}}}\right],$$

is an **asymptotic** $(1-\alpha)100\%$ confidence interval estimator for the expectile at the extreme level.

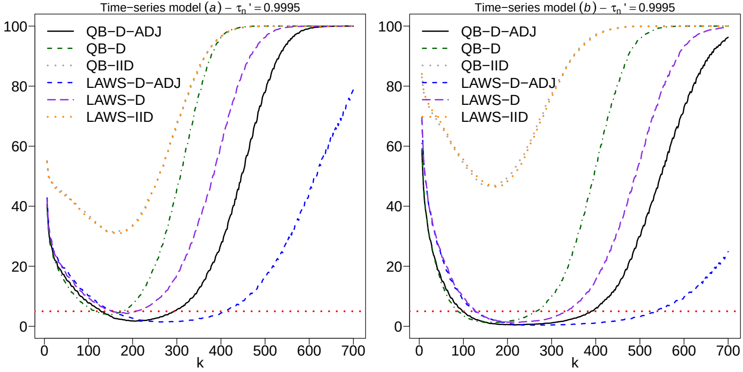
(a) The AR(1) model $Y_{t+1} = 0.8\,Y_t + \varepsilon_{t+1}$, where the innovations ε_t are i.i.d. with Student-t distribution and $\nu = 3$ df.

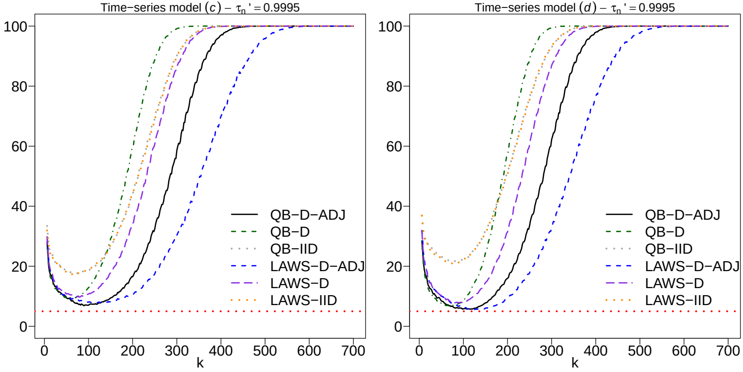
(b) The ARMA(1,1) model $Y_{t+1} = 0.95\,Y_t + \varepsilon_{t+1} + 0.9\,\varepsilon_t$, where the innovations ε_t are i.i.d. with symmetric Pareto distribution and shape parameter $\zeta = 3$.

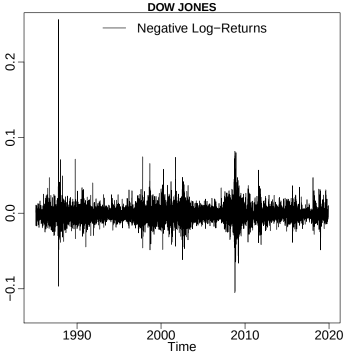
(c) The ARCH(1) model $Y_{t+1} = \sigma_{t+1}\varepsilon_{t+1}$, where $\sigma_{t+1}^2 = 0.4 + 0.6\,Y_t^2$, and ε_t are i.i.d. standard Gaussian innovations.

(d) The GARCH(1,1) model $Y_{t+1} = \sigma_{t+1}\varepsilon_{t+1}$, where $\sigma_{t+1}^2 = 0.1 + 0.4\,Y_t^2 + 0.4\,\sigma_t^2$, and ε_t are i.i.d. standard Gaussian innovations.

Non-coverage probability of CI

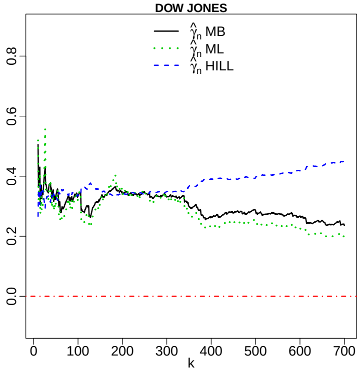



Time−series model (c) − τn ' = 0.9995
Time−series model (d) − τn ' = 0.9995
0
20
40
60
80
100
0
100
200
300
400
500
600
700
k
QB−D−ADJ
QB−D
QB−IID
LAWS−D−ADJ
LAWS−D
LAWS−IID

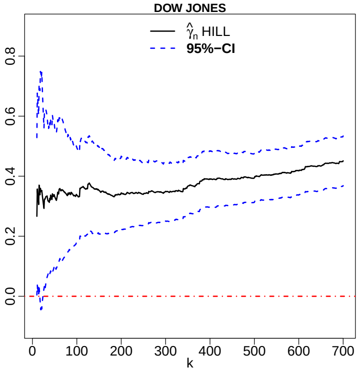


Let $(S_t, t \geq 1)$ be a price series. The **Negative Log-Returns** is

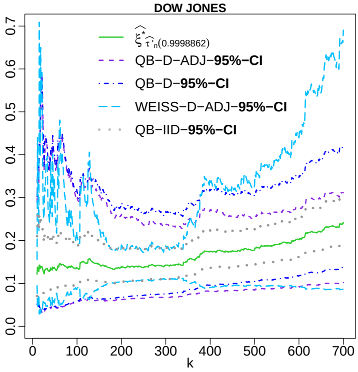
$$X_t = -\log(S_{t+1}/S_t), \quad t \geq 1$$

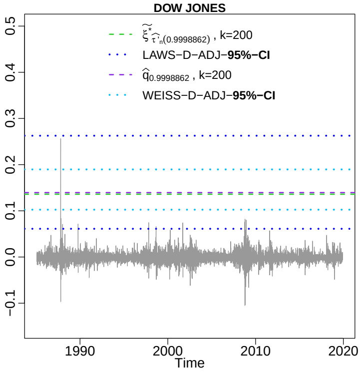


DOW JONES
$\hat{\gamma}_n$ MB
$\hat{\gamma}_n$ ML
$\hat{\gamma}_n$ HILL
0.0
0.2
0.4
0.6
0.8
0
100
200
300
400
500
600
700
k



DOW JONES
$\hat{\gamma}_n$ HILL
95%−CI
0.0
0.2
0.4
0.6
0.8
0
100
200
300
400
500
600
700
k


DOW JONES
QB−D−ADJ−95%−CI
QB−D−95%−CI
WEISS−D−ADJ−95%−CI
QB−IID−95%−CI
k
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
100
200
300
400
500
600
700


DOW JONES
, k=200
LAWS−D−ADJ−95%−CI
, k=200
WEISS−D−ADJ−95%−CI
Time
1990
2000
2010
2020
−0.1
0.0
0.1
0.2
0.3
0.4
0.5

1. In Davison, P., Stupfler, (2022) that are also estimation results for:
   - **dynamic** (conditional on the past) extreme expectiles based risk measures;
   - expectile-based **Marginal Expected Shortfall**, i.e.

$$\mathbb{E}(X|Y > \xi_\tau), \quad \tau \in (0, 1),$$

   where X is a loss return and Y is an aggregated loss return.

2. In Daouia, P., Stupfler, (2023) that are similar results for the case $\gamma < 0$;

3. Along with these papers there is an `R` package called `ExtremeRisks` that does the computation. Please, see

https://cran.r-project.org/web/packages/ExtremeRisks/index.html

...Thank you for your attention!

References I

- Daouiaa, A., Padoan, S.A. and Stupfler, G. (2023). Extreme expectile estimation for short-tailed data. https://arxiv.org/abs/2210.02056.
- Davison, C.A, Padoan, S. A. and Stupfler, G. (2022). Tail risk inference via expectiles in heavy-tailed time series. JBES, 41(3), 876–889.
- Doukhan, P. (1994), Mixing: Properties and Examples, Springer.
- de Haan, L. and A. Ferreira (2006). Extreme Value Theory: An Introduction. Springer.
- Leadbetter, M.R., Lindgren, G. and Rootzen, H. (1983). Extreme and Related Properties of random Sequences and Processes. Springer, Berlin.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. Econometrica, 46(1), 33–50.
- Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. Econometrica, 55(4), 819–847.