

Testing for a change in the tail parameter of regularly varying time series with long memory

Davide Giraudo

joint work with Annika Betken and Rafał Kulik

Ruhr-Universität Bochum

Research supported by the Collaborative Research Center 823 "Statistical modelling of nonlinear dynamic processes"

Telecom ParisTech, February, 27th 2020
Seminar S^2A

Hypothesis

Given observations X_1, \dots, X_n , assume that $\mathbb{P}(X_j > x) = x^{-\alpha_j} L(x)$, $j = 1, \dots, n$, where $\alpha_j > 0$ and L is a slowly varying function ($L(cx)/L(x) \rightarrow 1$ for all positive c as $x \rightarrow +\infty$). The number α_j is called the tail index of X_j .

We consider the testing problem (H, A) :

$$H: \alpha_1 = \dots = \alpha_n$$

against

$$A: \alpha_1 = \dots = \alpha_k \neq \alpha_{k+1} = \dots = \alpha_n \\ \text{for some } k \in \{1, \dots, n-1\}.$$

Estimation of the tail index of a random variable

Let X be a random variable with distribution function F having regularly varying tail with index $-\alpha$, $\alpha > 0$, that is

$\bar{F}(x) := \mathbb{P}(X > x) = x^{-\alpha}L(x)$, where L is slowly varying at infinity. It can be shown that

$$\lim_{u \rightarrow \infty} \mathbb{E} \left[\log \left(\frac{X}{u} \right) \mid X > u \right] = \lim_{u \rightarrow \infty} \frac{\mathbb{E} \left[\log \left(\frac{X}{u} \right) \mathbf{1}_{\{X > u\}} \right]}{\mathbb{P}(X > u)} = \frac{1}{\alpha}.$$

Thus, the tail index of a stationary time series X_j , $j \in \mathbb{N}$, with marginal distribution F having regularly varying tail with index $-\alpha$, $\alpha > 0$, can be estimated using the estimator $\hat{\gamma}$ defined by

$$\hat{\gamma} = \frac{1}{\sum_{j=1}^n \mathbf{1}_{\{X_j > u_n\}}} \sum_{j=1}^n \log \left(\frac{X_j}{u_n} \right) \mathbf{1}_{\{X_j > u_n\}},$$

where u_n , $n \in \mathbb{N}$, is a sequence such that $u_n \rightarrow \infty$ and $n\bar{F}(u_n) \rightarrow \infty$.

Test statistic

Let

$$\hat{\gamma}_{n,k} = \frac{1}{\sum_{j=1}^k \mathbf{1}\{X_j > u_n\}} \sum_{j=1}^k \log\left(\frac{X_j}{u_n}\right) \mathbf{1}\{X_j > u_n\}$$

and

$$\Gamma_{k,n} = \frac{k}{n} \left| \frac{\hat{\gamma}_{n,k}}{\hat{\gamma}_{n,n}} - 1 \right|.$$

Since the location of the change of tail parameter k is unknown, we define the test-statistic

$$\Gamma_n := \max_{1 \leq k \leq n-1} \Gamma_{k,n}.$$

Empirical process

In order to find the limiting distribution of the test statistics, similar to Kulik and Soulier (2011) we consider the two-parameter tail empirical process (TEP)

$$e_n(s, t) = \{ \tilde{T}_n(s, t) - T(s, t) \}, \quad s \geq 1, t \in [0, 1],$$

where

$$\tilde{T}_n(s, t) = \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{[nt]} \mathbf{1}\{X_j > u_n s\}$$

and

$$T(s, t) = ts^{-\alpha}.$$

Then

$$\int_1^{+\infty} \frac{1}{s} \tilde{T}_n(s, t) ds = \sum_{j=1}^n \log\left(\frac{X_j}{u_n}\right) \mathbf{1}\{X_j > u_n\}.$$

Stochastic volatility model

We will consider the case of the stochastic volatility model:

$$X_j = \sigma(Y_j) \varepsilon_j,$$

where

- $(\varepsilon_j)_{j \in \mathbb{Z}}$ is an i.i.d. sequence of random variables with $\mathbb{E}[\varepsilon_1] = 0$;
- σ is a non-negative measurable function;
- $Y_j, j \geq 1$, is a stationary, long-range dependent Gaussian process, that is,

$$Y_j = \sum_{k=1}^{\infty} c_k \eta_{j-k}, \quad \sum_{k=1}^{\infty} c_k^2 = 1,$$

for i.i.d. Gaussian random variables $\eta_j, j \in \mathbb{Z}$, with $\mathbb{E}[\eta_1] = 0$, $\text{var } \eta_1 = 1$, $((\varepsilon_j, \eta_j))_{j \in \mathbb{Z}}$ is independent and

$$\gamma_Y(k) := \text{cov}(Y_j, Y_{j+k}) = \sum_{\ell \geq 0} c_\ell c_{\ell+k} = k^{-D} L_\gamma(k),$$

where $D \in (0, 1)$ and L_γ slowly varying at ∞ .

Covariance of the process $(\sigma(Y_j))_{j \geq 1}$

Let φ be the density of the standard normal distribution. Every $G \in L^2(\mathbb{R}, \varphi(x)dx)$ has an expansion in Hermite polynomials, i.e. for $G \in L^2(\mathbb{R}, \varphi(x)dx)$ and X standard normally distributed, we have

$$G(X) = \sum_{r=0}^{\infty} \frac{J_r(G)}{r!} H_r(X),$$

where the so-called *Hermite coefficient* $J_r(G)$ is given by

$$J_r(G) := \langle G, H_r \rangle_{L^2} = \mathbb{E}[G(X)H_r(X)],$$

and H_r is the r -th Hermite polynomial.

Let $m := \min \{k \geq 1 : J_k(G) \neq 0\}$ be the Hermite rank of G . Then

$$d_n^2 := \text{var} \left(\sum_{j=1}^n H_m(Y_j) \right) \sim c_m n^{2-mD} L^m(n), \quad c_m = \frac{2m!}{(1-Dm)(2-Dm)}.$$

Type of assumptions: slowly varying function

We assume that

$$\mathbb{P}(X_1 > x) = cx^{-\alpha} \exp\left(\int_1^x \frac{\eta^*(u)}{u} du\right),$$

where $\eta^*(u) = u^{-\rho} L_{\eta^*}(u)$, $\rho > 0$, L_{η^*} slowly varying;

Example

Assume that $\eta^*(x) = x^{-\alpha\beta}$ for some $\beta > 0$; then, for $x \rightarrow \infty$, $\mathbb{P}(X_1 > x) = C(x^{-\alpha} + \mathcal{O}(x^{-\alpha(\beta+1)}))$.

Such an assumption gives a good comparison between $\mathbb{P}(X_1 > x)$ and $x^{-\alpha}$:

$$\forall t \geq 1, \forall z > 0, \quad \frac{\bar{F}_Z(zt)}{\bar{F}_Z(t)} \leq z^{-\alpha} + C_\varepsilon z^{-\alpha-\rho} (z \vee z^{-1})^\varepsilon$$

$$\left| \frac{\bar{F}_Z(at) - \bar{F}_Z(bt)}{\bar{F}_Z(t)} - (a^{-\alpha} - b^{-\alpha}) \right| \leq C\eta^*(t) (\min\{a; 1\})^{-\alpha-\rho-\varepsilon} (b-a).$$

Type of assumptions: moments of $\sigma(Y_1)$

The inequalities

$$\forall t \geq 1, \forall z > 0, \quad \frac{\bar{F}_Z(zt)}{\bar{F}_Z(t)} \leq z^{-\alpha} + C_\varepsilon z^{-\alpha-\rho} (z \vee z^{-1})^\varepsilon$$

$$\left| \frac{\bar{F}_Z(at) - \bar{F}_Z(bt)}{\bar{F}_Z(t)} - (a^{-\alpha} - b^{-\alpha}) \right| \leq C\eta^*(t) (\min\{a; 1\})^{-\alpha-\rho-\varepsilon} (b-a).$$

with $t = 1/\sigma(Y_j)$ leads to the control of partial sums of type $\sum_{j=1}^n \sigma(Y_j)^p$. We require

- $\mathbb{E}[\sigma^{\alpha+(\alpha \vee \rho)+\delta}(Y_1)] < \infty$ for some $\delta > 0$;
- $\mathbb{E}\left[\sigma(Y_1)^{2(\alpha-\rho)} \left(\sigma(Y_1) \vee (\sigma(Y_1))^{-1}\right)^{\varepsilon_0}\right] < \infty$ for some $\varepsilon_0 > 0$.
- $\mathbb{E}\left[\sigma(Y_1)^{-2-\delta}\right]$ is finite for a positive δ .

Example

If $0 < c \leq \sigma(y) \leq C$ for all $y \in \mathbb{R}$, then all the moment assumptions are satisfied.

Type of assumptions: η^*

$$\mathbb{P}(X_1 > x) = cx^{-\alpha} \exp\left(\int_1^x \frac{\eta^*(u)}{u} du\right),$$

where $\eta^*(u) = u^{-\rho} L_{\eta^*}(u)$, $\rho > 0$, L_{η^*} slowly varying;

$$d_n^2 := \text{var}\left(\sum_{j=1}^n H_m(Y_j)\right) \sim c_m n^{2-mD} L^m(n), \quad c_m = \frac{2m!}{(1-Dm)(2-Dm)}.$$

We assume that $\eta^*(u_n) = o\left(\frac{d_n}{n} + \frac{1}{\sqrt{n\bar{F}(u_n)}}\right)$.

Limit of the empirical process

Theorem (Betken, G., Kulik (2020+))

Assume that the technical assumptions hold. Let

$$e_n(s, t) := \frac{1}{n\bar{F}(u_n)} \sum_{j=1}^{\lfloor nt \rfloor} \mathbf{1}\{X_j > u_n s\} - ts^{-\alpha}.$$

- If $\frac{n}{d_n} = o\left(\sqrt{n\bar{F}(u_n)}\right)$,

$$\frac{n}{d_n} e_n(s, t) \Rightarrow \frac{s^{-\alpha}}{\mathbb{E}[\sigma^\alpha(Y_1)]} \frac{J_q(\Psi)}{q!} Z_q(t),$$

where \Rightarrow denotes weak convergence in $D([1, \infty] \times [0, 1])$, $\Psi(y) = \sigma^\alpha(y)$, q is the Hermite rank of Ψ and Z_q a Hermite process of order q .

- If $\sqrt{n\bar{F}(u_n)} = o\left(\frac{n}{d_n}\right)$,

$$\sqrt{n\bar{F}(u_n)} e_n(s, t) \Rightarrow B_{s^{-\alpha}, t}$$

in $D([1, \infty] \times [0, 1])$, where B denotes a standard Brownian sheet.

Explanation of the two cases

We consider the following decomposition:

$$\begin{aligned}e_n(s, t) &= \frac{1}{n\bar{F}_Z(u_n)} \sum_{j=1}^{\lfloor nt \rfloor} (\mathbf{1}\{X_j > u_n s\} - \mathbb{E}[\mathbf{1}\{X_j > u_n s\} | \mathcal{F}_{j-1}]) \\ &\quad + \frac{1}{n\bar{F}_Z(u_n)} \sum_{j=1}^{\lfloor nt \rfloor} (\mathbb{E}[\mathbf{1}\{X_j > u_n s\} | \mathcal{F}_{j-1}] - \bar{F}(u_n s)) \\ &= M_n(s, t) + L_n(s, t),\end{aligned}$$

where

$$\mathcal{F}_j := \sigma(\varepsilon_k, \eta_k, k \in \mathbb{Z}, k \leq j).$$

We call M_n the *martingale part*, while we refer to L_n as the *long memory part*.

If $\frac{n}{d_n} = o\left(\sqrt{n\bar{F}(u_n)}\right)$, the martingale part is negligible.

Long memory part

Recall that

$$L_n(s, t) = \frac{1}{n\bar{F}_Z(u_n)} \sum_{j=1}^{\lfloor nt \rfloor} (\mathbb{E}[\mathbf{1}\{\sigma(Y_j)\varepsilon_j > u_n s\} \mid \mathcal{F}_{j-1}] - \bar{F}(u_n s)).$$

Since $\mathcal{F}_{j-1} = \sigma(\varepsilon_k, \eta_k, k \in \mathbb{Z}, k \leq j-1)$, we get

$$L_n(s, t) = \frac{1}{n\bar{F}_Z(u_n)} \sum_{j=1}^{\lfloor nt \rfloor} \left(\bar{F}_{\varepsilon_0} \left(\frac{u_n s}{\sigma(Y_j)} \right) - \bar{F}(u_n s) \right).$$

With the assumptions on \bar{F}_Z and \bar{F}_{ε_0} , we are reduced to treat

$$L'_n(s, t) = s^{-\alpha} \frac{1}{n} \sum_{j=1}^{\lfloor nt \rfloor} (\sigma(Y_j)^\alpha - \mathbb{E}[\sigma(Y_j)^\alpha]).$$

This can be done by writing the Hermite expansion of $y \mapsto \sigma(y)^\alpha$.

Martingale part (1)

Recall that

$$M_n(s, t) = \frac{1}{n\bar{F}_Z(u_n)} \sum_{j=1}^{\lfloor nt \rfloor} (\mathbf{1}\{X_j > u_n s\} - \mathbb{E}[\mathbf{1}\{X_j > u_n s\} | \mathcal{F}_{j-1}]).$$

- finite dimensional distribution: central limit theorem for martingale arrays;
- convergence in $D([1, R] \times [0, 1])$: we use a result of Davydov and Zitikis (2008). We have to control the moments of

$$\sqrt{n\bar{F}_Z(u_n)} |M_n(s, t) - M_n(s', t')| \text{ for } s, s', t \text{ and } t' \text{ such that } |s - s'| + |t - t'| \geq 1/(n\bar{F}_Z(u_n)).$$

Letting

$$\Delta_n(s, t) := \frac{1}{\sqrt{n\bar{F}_Z(u_n)}} \sum_{j=1}^{\lfloor nt \rfloor} \mathbb{E}[\mathbf{1}\{X_j > u_n(1 + Rs)\} | \mathcal{F}_{j-1}],$$

we have to show that the following goes to zero in probability

$$\max_{1 \leq i, j \leq 1/(n\bar{F}_Z(u_n))} \left| \Delta_n\left(\frac{i}{n\bar{F}_Z(u_n)}, \frac{j}{n\bar{F}_Z(u_n)}\right) - \Delta_n\left(\frac{i-1}{n\bar{F}_Z(u_n)}, \frac{j}{n\bar{F}_Z(u_n)}\right) \right|$$

Martingale part (2)

Recall that

$$M_n(s, t) = \frac{1}{n\bar{F}_Z(u_n)} \sum_{j=1}^{\lfloor nt \rfloor} (\mathbf{1}\{X_j > u_n s\} - \mathbb{E}[\mathbf{1}\{X_j > u_n s\} | \mathcal{F}_{j-1}])$$

and

$$\Delta_n(s, t) := \frac{1}{\sqrt{n\bar{F}_Z(u_n)}} \sum_{j=1}^{\lfloor nt \rfloor} \mathbb{E}[\mathbf{1}\{X_j > u_n(1 + Rs)\} | \mathcal{F}_{j-1}].$$

We also have to show that

$$\max_{1 \leq i, j \leq 1/(n\bar{F}_Z(u_n))} \left| \Delta_n\left(\frac{i}{n\bar{F}_Z(u_n)}, \frac{j}{n\bar{F}_Z(u_n)}\right) - \Delta_n\left(\frac{i}{n\bar{F}_Z(u_n)}, \frac{j-1}{n\bar{F}_Z(u_n)}\right) \right| \rightarrow 0$$

in probability. We use the assumption on \bar{F}_Z and integrability of powers of $\sigma(Y_1)$.

Convergence of the tail estimator

$$\hat{\gamma}_n(t) = \frac{\sum_{j=1}^{\lfloor nt \rfloor} \log \left(\frac{X_j}{u_n} \right) \mathbf{1}\{X_j > u_n\}}{\sum_{j=1}^{\lfloor nt \rfloor} \mathbf{1}\{X_j > u_n\}}.$$

Corollary (Betken, G., Kulik (2020+))

Under the technical assumptions,

- if $\frac{n}{d_n} = o\left(\sqrt{n\bar{F}(u_n)}\right)$, then

$$\frac{n}{d_n} t (\hat{\gamma}_n(t) - \gamma) \rightarrow 0 \text{ in probability in } D[0, 1];$$

- if $\sqrt{n\bar{F}(u_n)} = o\left(\frac{n}{d_n}\right)$, then

$$\sqrt{n\bar{F}(u_n)} t (\hat{\gamma}_n(t) - \alpha^{-1}) \Rightarrow \frac{1}{\alpha} B \text{ in distribution in } D[0, 1],$$

where B is a standard Brownian motion.

Convergence of the test statistic

Corollary (Betken, G., Kulik (2020+))

Under the technical assumptions,

- if $\frac{n}{d_n} = o\left(\sqrt{n\bar{F}(u_n)}\right)$, then

$$\frac{n}{d_n} \sup_{t \in [0,1]} t \left| \frac{\hat{\gamma}_n(t)}{\hat{\gamma}_n} - 1 \right| \rightarrow 0 \text{ in probability;}$$

- if $\sqrt{n\bar{F}(u_n)} = o\left(\frac{n}{d_n}\right)$, then

$$\sqrt{n\bar{F}(u_n)} \sup_{t \in [0,1]} t \left| \frac{\hat{\gamma}_n(t)}{\hat{\gamma}_n} - 1 \right| \xrightarrow{\mathcal{D}} \sup_{t \in [0,1]} |B(t) - tB(1)|$$

where B is a standard Brownian motion.

Simulations: setting (1)

$$X_j = \sigma(Y_j)\varepsilon_j, \quad j \geq 1,$$

where

- $\varepsilon_j, j \geq 1$, is an i.i.d. Pareto distributed sequence generated by the function **rgpd** (**fExtremes** package in **R**);
- $Y_j, j \geq 1$, is a fractional Gaussian noise sequence generated by the function **simFGNO** (**longmemo** package in **R**) with Hurst parameter H ;
- $\sigma(y) = \exp(y)$.

Under the alternative, we insert a change of height h at location $k = \lfloor n\tau \rfloor$ by simulating i.i.d. Pareto distributed observations $\varepsilon_j, j \geq 1$, where $\varepsilon_j, j = 1, \dots, k$, with tail index $\alpha_1 = \dots = \alpha_k = \alpha$ and $\varepsilon_j, j = k + 1, \dots, n$, with tail index $\alpha_{k+1} = \dots = \alpha_n = \alpha + h$.

Simulations: setting (2)

Under the alternative, we insert a change of height h at location $k = \lfloor n\tau \rfloor$ by simulating i.i.d. Pareto distributed observations ε_j , $j \geq 1$, where ε_j , $j = 1, \dots, k$, with tail index $\alpha_1 = \dots = \alpha_k = \alpha$ and ε_j , $j = k + 1, \dots, n$, with tail index $\alpha_{k+1} = \dots = \alpha_n = \alpha + h$.

We base test decisions on the statistic $\Gamma_n := \max_{1 \leq k \leq n-1} \Gamma_{k,n}$, where

$$\Gamma_{k,n} = \frac{k}{n} \left| \frac{\hat{\gamma}_{\text{Hill}}\left(\frac{k}{n}\right)}{\hat{\gamma}_{\text{Hill}}(1)} - 1 \right| \quad \text{with} \quad \hat{\gamma}_{\text{Hill}}(t) = \frac{1}{\lfloor k_n t \rfloor} \sum_{i=1}^{\lfloor k_n t \rfloor} \log \left(\frac{X_{\lfloor nt \rfloor : \lfloor nt \rfloor - i + 1}}{X_{\lfloor nt \rfloor : \lfloor nt \rfloor - \lfloor k_n t \rfloor}} \right)$$

i.e., $\hat{\gamma}_{\text{Hill}}(t)$ is the Hill estimator based on the observations $X_1, \dots, X_{\lfloor nt \rfloor}$ and where $X_{N:N} \geq X_{N:N-1} \geq \dots \geq X_{N:1}$ for all $N \geq 1$. We choose $k_n = \lfloor np \rfloor$, i.e., p defines the proportion of the data that the estimation of the tail index is based on.

Simulations (1)

		$\alpha = 2.5$					
	p	n	$h = 0$	$h = 0.5$	$h = 1$	$h = -0.5$	$h = -1$
$H = 0.6$	0.1	300	0.088	0.088	0.086	0.109	0.192
		500	0.078	0.069	0.065	0.105	0.249
		1000	0.071	0.063	0.059	0.106	0.391
	0.2	300	0.071	0.065	0.058	0.078	0.176
		500	0.049	0.059	0.059	0.076	0.227
		1000	0.044	0.050	0.055	0.086	0.387
$H = 0.7$	0.1	300	0.112	0.103	0.096	0.137	0.217
		500	0.093	0.086	0.087	0.123	0.262
		1000	0.084	0.069	0.070	0.118	0.385
	0.2	300	0.087	0.083	0.083	0.105	0.196
		500	0.075	0.080	0.071	0.099	0.256
		1000	0.068	0.063	0.067	0.109	0.408

Table: Rejection rates of the change-point test based on the statistic Γ_n , $k_m = \lfloor mp \rfloor$, for LMSV time series (Pareto distributed ε_j , $j \geq 1$) of length n with Hurst parameter H , tail index α and a shift in the mean of height h after a proportion $\tau = 0.5$. The calculations are based on 5,000 simulation runs.

Simulations (2)

		$\alpha = 2.5$					
	p	n	$h = 0$	$h = 0.5$	$h = 1$	$h = -0.5$	$h = -1$
$H = 0.6$	0.1	300	0.088	0.086	0.085	0.104	0.127
		500	0.078	0.071	0.071	0.083	0.129
		1000	0.071	0.058	0.060	0.076	0.151
	0.2	300	0.071	0.069	0.068	0.075	0.099
		500	0.049	0.052	0.059	0.063	0.120
		1000	0.044	0.050	0.052	0.056	0.160
$H = 0.7$	0.1	300	0.112	0.100	0.110	0.124	0.139
		500	0.093	0.091	0.092	0.100	0.145
		1000	0.084	0.074	0.075	0.092	0.176
	0.2	300	0.0868	0.081	0.073	0.087	0.113
		500	0.075	0.071	0.076	0.080	0.122
		1000	0.068	0.068	0.073	0.084	0.187

Table: Rejection rates of the change-point test based on the statistic Γ_n , $k_m = \lfloor mp \rfloor$, for LMSV time series (Pareto distributed ε_j , $j \geq 1$) of length n with Hurst parameter H , tail index α and a shift in the mean of height h after a proportion $\tau = 0.25$. The calculations are based on 5,000 simulation runs.

Conclusion

What we have done:

- Convergence of the empirical process.
- Convergence of the tail estimator and test statistic when the martingale part dominates.

Remaining questions:

- Find the good normalization for the tail estimator when the long memory part dominates.
- Treat the case where n/d_n and $\sqrt{n\bar{F}(u_n)}$ are equivalent.

References

- A. BETKEN, D. GIRAUDO, R. KULIK (2020+). Testing for change in the tail parameter for regularly varying time series with long memory via Hill statistics. In preparation
- A. BETKEN, R. KULIK (2019). Testing for change in stochastic volatility with long range dependence. *Journal of Time Series Analysis*, DOI: 10.1111/jtsa.12449.
- V. PIPIRAS, M. S. TAQQU (2017). Long-Range Dependence and Self-Similarity. Vol. 45. *Cambridge university press*.
- R. KULIK, P. SOULIER (2011). The tail empirical process for long memory stochastic volatility sequences. *Stochastic Processes and their Applications*, 121:109 - 134.
- A. C. HARVEY (2002). Long memory in stochastic volatility. In: *Forecasting Volatility in the Financial Markets*, 307-320.
- F. BREIDT, N. CRATO, AND P. DE LIMA (1998). The detection and estimation of long memory in stochastic volatility. *Journal of Econometrics*, 83:325 - 348.
- M. S. TAQQU (1979). Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 50:55 - 83.