Extreme value theory and time series: prediction after extreme events

Philippe Soulier (Université Paris-Ouest)

Essec, 25 novembre 2010

I. Statement of the problem

II. Regularly varying time series

III. Measures of extremal dependence

IV. Estimation of limit conditional laws

V. Long memory in volatility

I. Statement of the problem

Let $\{X_k\}$ be a stationary time series (possibly multivariate).

We consider the problem of the inference on future values $X_m, \ldots, X_{m+h'}$ given that the recent past $X_{-h}, \ldots, X_{-1}$ is "extreme".

Examples

Given that ozone level is extremely high today,
- how high will it be tomorrow?
- how high will be another pollutant tomorrow?
- how high will be some combination of several pollutant.

Same type of questions with financial time series.

Mathematical formulation

For a fixed event A, does there exist normalizing sequences $a(t)$ and $m(t)$ and a limiting distribution H_A such that

$$\lim_{t\to\infty} \mathbf{P}((X_m, \ldots, X_{m+h'}) \in m(t) + a(t)B \mid (X_{-h}, \ldots, X_0) \in tA) = H_A(B) \text{ ?}$$

- The dilatation tA expresses the extreme nature of the conditioning event.

- The normalizing sequences a and m and the limiting distribution H_A are a priori unknown and must be estimated.

Examples

•The simplest case case is the conditional distribution of X_m given X_0 is extreme: are there normalizing functions a, m and a non degenerate distribution function Ψ_m such that

$$\Psi_m(x) = \lim_{t\to\infty} \mathbf{P}(X_m \leq m(t) + a(t)x \mid X_0 > t) \ .$$

• The conditionned and the conditioning events can be multivariate. Taking $A = \{(u,v) \in [0,\infty)^2 \mid u > 1; v > 1\}$. Then we can look for a and m such that there exists

$$\lim_{t\to\infty} \mathbf{P}\left((X_m,\dots,X_{m+h'}) \in m(t) + a(t)B \mid X_1 > 1; X_0 > t\right) \ .$$

Take $A = \{(u,v) \in [0,\infty)^2 \mid u + v > 1\}$. Then we study

$$\lim_{t\to\infty} \mathbf{P}\left((X_m,\dots,X_{m+h'}) \in m(t) + a(t)B \mid X_{-1} + X_0 > t\right) \ .$$

These problems can be investigated for time series with light-tailed marginal distributions (e.g. Gaussian) or for regularly varying time series.

We consider only the latter case.

II. Regularly varying time series

A stationary (possibly vector valued) time series $\{X_t\}$ is said to be regularly varying if all the finite dimensional distributions are regularly varying. This means that there exists a normalizing function a and a sequence of nondegenerate Radon measures $\{\nu_k\}$ on $[-\infty, \infty]^k \setminus \{\mathbf{0}\}$ called exponent measures such that

$$t\mathbf{P}\left(\frac{(X_1, \ldots, X_k)}{a(t)} \in \cdot\right) \to_v \nu_k$$

vaguely on $[-\infty, \infty]^k \setminus \{\mathbf{0}\}$.

The measure ν_k is necessarily nhomogeneous with degree $-\alpha > 0$ and the function a is regularly varying at infinity with index $1/\alpha$.

W.l.o.g. we assume that $\nu_k(([-\infty, \epsilon]^k)^c) > 0$ for some $\epsilon > 0$.

Equivalent formulation

The distribution function F_k of $(X_1, \dots, X_k)$ is in the maximum domain of attraction of a multivariate Frechet law.

Let $\{\mathbf{X}_j^*\}$ be an i.i.d. sequence of random vectors with marginal distribution F_k. Then there exists a max-stable distribution G with Frechet marginals such that

$$a(n)^{-1} \max_{1 \le i \le n} \mathbf{X}_i^* \to_d G$$

with

$$G(\mathbf{x}) = \exp\{-\nu_k((-\infty, \mathbf{x}]^c)\} \; .$$

for $x \in [0, \infty]^k \setminus \{0\}$. The multivariate distribution G has Frechet marginals.

Asymptotic dependence vs Asymptotic independence

This definition induces two very different cases. The regularly varying k-dimensional random vector $(X_1, \ldots, X_k)$ is said to be asymptotically independent if its exponent measure ν_k satisfies

$$\nu_k(([-\infty, 0) \cup (0, \infty])^k) = 0 \ . \tag{1}$$

If (1) does not hold, the vector is said to be asymptotically dependent.

For $k = 2$, (1) is equivalent to the measure ν_k being concentrated on the axes.

In general, if the measure ν_k is concentrated on the axes, then the max-stable distribution G has independent marginals:

$$G(x_1, \ldots, x_k) = \prod_{i=1}^{k} \mathrm{e}^{-c_i x_i^{-\alpha}}$$

with $c_i = \nu_k(\mathbf{R}^{i-1} \times (1, \infty) \times \mathbf{R}^{k-i})$.

We will refer to this case as pairwise asymptotic independence.

Interpretation

A random vector $\mathbf{X}$ is asymptotically dependent if large values of all its component can occur simultaneously.

A random vector $\mathbf{X}$ is asymptotically independent if large values of all its component cannot occur simultaneously.

It is pairwise asymptotically independent if large values of any two components cannot occur simultaneously.

A time series which is not pairwise asymptotically independent exhibits clustering of extremes.

Examples

- An i.i.d. sequence with regularly varying marginal distribution is pairwise asymptotically independent.

- An MA(q) process with regularly varying innovation has asymptotically dependent k-dimensional marginal distributions if $k \leq q$, and asymptotically independent (but not pairwise independent) otherwise. High values of X_t are due to single high values of $\{Z_t\}$ and will come in q-tuples.

- Linear processes with heavy tailed innovation

$$X_t = \sum_{j \in \mathbf{Z}} c_j Z_{t-j}$$

where $\{Z_j\}$ is an i.i.d. sequence with regularly varying marginal distribution and the coefficients c_j satisfy some summability condition.

If an infinite number of coefficients c_j is non zero, then all the marginal distributions are asymptotically dependent.

GARCH Processes

A GARCH(p, q) process $\{X_t\}$ is defined by

$$X_t = \sigma_t Z_t$$

$$\sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2 + \sum_{j=1}^{q} b_j \sigma_{t-j}^2 \ ,$$

with $a_0 > 0$, $a_j, b_j \geq 0$ and $\sum_{j=1}^{p} a_j + \sum_{j=1}^{q} b_j < 1$.

Then, under some other technical assumptions on the distribution of Z, there exists κ such that the one-dimensional marginal distributions of X_t and σ_t are regularly varying with index κ and the finite dimensional distributions are also regularly varying if κ is not an integer.

The finite dimensional distributions are asymptotically dependent: clustering of exceedences.

Extension to solutions of Stochastic Recurrence Equations.

Basrak, Davis and Mikosch (2002)

Stochastic volatility processes

The stochastic volatility (SV) process has been introduced in econometric literature by Taylor (1986). The log-prices are modeled as

$$X_t = Z_t \sigma_t \ ,$$

where $\{Z_t\}$ is an i.i.d. sequence of zero mean random variables, independent of $\{\sigma_t\}$, which stands as a proxy for the volatility. It is usually modeled as $\sigma_t = \sigma(\xi_t)$, where $\{\xi_t\}$ is a stationary Gaussian process.
A frequent choice for the function σ is $\sigma(x) = \mathrm{e}^{x/2}$.

Contrary to GARCH-type process, this is a latent variable model; $\sigma(\xi_t)$ is unobservable and it is not necessarily the conditional variance. The advantage of this model is that it allows a more flexible modelization of the tails (by $\{Z_t\}$) and of the memory (by the Gaussian process $\{\xi_j\}$).

This model can generalized by assuming that $\{\xi_k\}$ is a linear, possibly non Gaussian process and/or by introducing some dependence between the sequences $\{Z_k\}$ and $\{\xi_k\}$. Specifically, it can be assumed that

$$\xi_k = \sum_{j=1}^{\infty} c_j \eta_{k-j}$$

where $\{\eta_k\}$ is a Gaussian white noise and $\{(\eta_k, Z_k)\}$ is an i.i.d. sequence of bivariate vectors.
This allows for some form of leverage.

Choosing $\sigma(x) = \mathrm{e}^{x/2}$ yields the EGARCH process.
Nelson (1991).

For brevity, we will not discuss this case, though most of our results hold.

- Extremal properties of the SV model

If Z_1 has a regularly varying right tail:

$$\mathbf{P}(Z_1 > x) = L(x)x^{-\alpha}$$

where L is slowly varying.

Then, by Breiman's Lemma if $\mathbf{E}[\sigma^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$,

$$\mathbf{P}(X_1 > x) \sim \mathbf{E}[\sigma^{\alpha}]\mathbf{P}(Z_1 > z) .$$

X_1 has the same tail index as Z_1.

Asymptotic independence

The SV process is pairwise asymptotically independent:

$$\forall j \geq 1\ , \quad \lim_{t \to \infty} t\mathbf{P}(X_0 > a(t)x\ ,\ X_j > a(t)y) = 0$$

where the function a is such that $t\mathbf{P}(Z > a(t)) \sim 1$.

Asymptotic independence

The SV process is pairwise asymptotically independent:

$$\forall j \geq 1 \, , \quad \lim_{t \to \infty} t\mathbf{P}(X_0 > a(t)x \, , \, X_j > a(t)y) = 0$$

where the function a is such that $t\mathbf{P}(Z > a(t)) \sim 1$.

Proof: Let $\mathcal{X}$ be the sigma-field generated by $\{\sigma_j\}$. Then X_0 and X_j are independent condionally on $\mathcal{X}$. By a bivariate version of Breiman's Lemma, for $j > 0$, if $\mathbf{E}[\sigma_0^{\alpha+\epsilon}\sigma_t^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$, we obtain

$$\begin{aligned} t\mathbf{P}(X_0 > a(t)x \, , \, X_j > a(t)y) \\ &= \mathbf{E}[t\mathbf{P}(X_0 > a(t)x \, , X_j > a(t)y \mid \mathcal{X})] \\ &= t^{-1}\mathbf{E}[t\mathbf{P}(\sigma_0 Z_0 > a(t)x \mid \mathcal{X})t\mathbf{P}(\sigma_j Z_j > a(t)y \mid \mathcal{X})] \\ &\sim t^{-1}\mathbf{E}[\sigma_0^{\alpha}\sigma_j^{\alpha}] \to 0 \, . \end{aligned}$$

Contrary to the GARCH process, there is no clustering of extremes for the SV model.

In particular, the maximum behaves as in the case of an i.i.d. sequence.

$$\mathbf{P}(\max_{j=1,\ldots,n} X_j \le a_n x) \to \mathrm{e}^{-x^{-\alpha}}$$

with $a_n = F_X^{\leftarrow}(1 - 1/n)$.

This holds regardless of the memory of $\{\xi_j\}$.

Breidt and Davis (1998), Davis and Mikoch (2001)

GARCH vs SV

	Memory	Extremes
GARCH	short	asymp. dep.
SV	possibly long	asymp. indep.

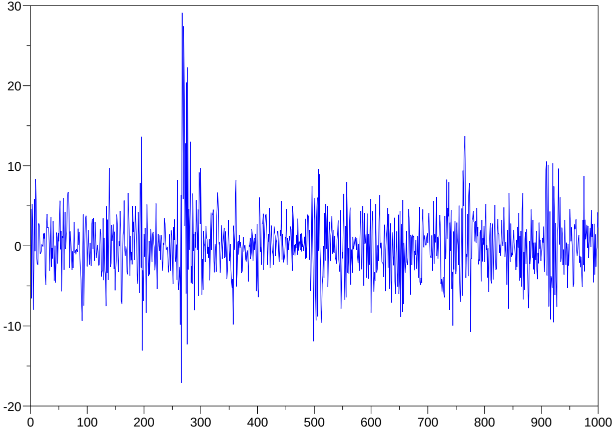


Figure: Sample path of GARCH(1,1) process with $a_1 + b_1 = .96$.

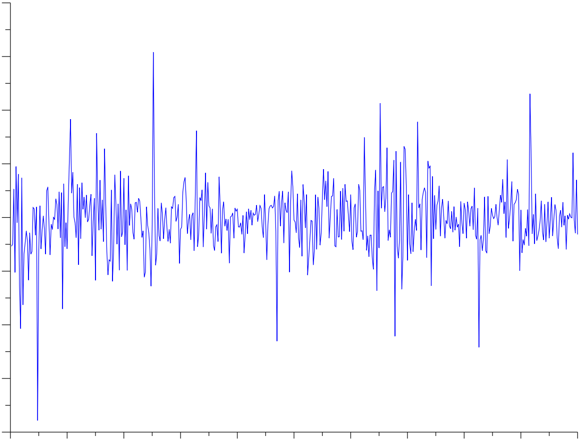


Figure: Sample path of LMSV process with $\alpha = 3$ and $H = .85$

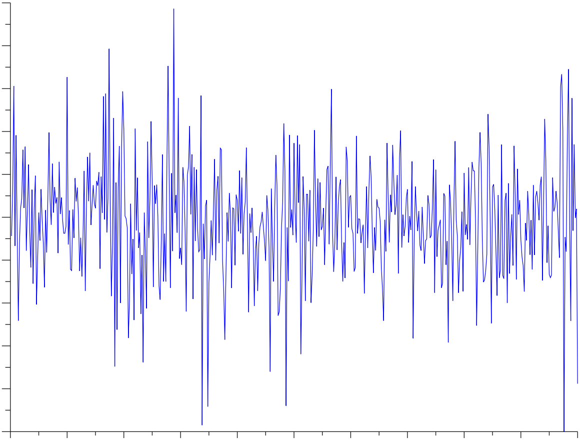


Figure: The real data

III. Measures of extremal dependence

For regularly varying time series, several quantities have been introduced to measure the extremal dependence.

- The tail process $\{Y_j\}$, defined as the weak limit (of finite marginals) of $\{X_j\}$ given X_0 large:

$$(Y_1, \ldots, Y_m) = w - \lim_{t \to \infty} x^{-1}(X_1, \ldots, X_m) \text{ given } X_0 > x \,.$$

  Basrak and Segers (2009).

- The extremogram. For $m > h$, $A \subset \mathbf{R}^{h'}$ and $B \subset \mathbf{R}^{h}$, define

$$\rho(A, B, m) = \lim_{t \to \infty} \mathbf{P}((X_{m+1}, \ldots, X_{m+h'}) \in tA \mid (X_1, \ldots, X_h) \in tB)$$

  Davis and Mikosch (2009)

The tail process and the extremogram are related to each other and to the sequence of exponent measures ν_k.

- Extreme dependence function

$$\chi_{k_1,\ldots,k_d}(x_1,\ldots,x_d) = \lim_{z\to\infty} \frac{\mathbf{P}(X_{k_1} > tx_1,\ldots,X_{k_d} > tx_d)}{\mathbf{P}(X_1 > t)}$$

Fasen et al. (2010)

- Extremal dependence measure

$$EDM(h) = \lim_{t\to\infty} \mathbf{E}\left[\frac{X_0 X_h}{\|(X_0,X_h)\|^2} \mid \|(X_0,X_h)\| > t\right]$$

where $\|\cdot\|$ denotes any norm on $\mathbf{R}^2$.
Larsen and Resnick (2009)

• For regularly varying time series with asymptotic dependence, these quantities are non trivial and can be estimated by standard extreme value techniques (under some form of mixing conditions).

• In the case of asymptotic independence, they become trivial. For instance, the extremal dependence function $\chi_{k_1,\dots,k_d}$ and the extremal dependence measure $EDM(k)$ are identically zero, the extremogram $\rho(A, B, m)$ is zero for most choices of sets A and B and lag m. The definition of these extremal quantities must be adapted to asymptotic independence or new quantities must be defined.

• In both cases, these quantities can be expressed in terms of the conditional limit laws.

Conditional limit laws in the case of asymptotic dependence

In the case of asymptotic dependence, the extremogram and other measures of dependence can be expressed in terms of conditional limit laws and of the exponent measures.

For $m > h$, $A \subset \mathbf{R}^h$ such that $\nu_h(A) > 0$ and $B \subset \mathbf{R}^{h'}$,

$$
\begin{aligned}
&\rho(A, B, m) \\
&= \lim_{t\to\infty} \mathbf{P}((X_{m+1}, \ldots, X_{m+h'}) \in tB \mid (X_1, \ldots, X_h) \in tA) \\
&= \lim_{t\to\infty} \frac{t\mathbf{P}((X_1, \ldots, X_h) \in a(t)A\ ;\ (X_{m+1}, \ldots, X_{m+h'}) \in a(t)B)}{t\mathbf{P}((X_1, \ldots, X_h) \in a(t)A)} \\
&= \frac{\nu_{m+h'}(A \otimes B)}{\nu_h(A)}\ .
\end{aligned}
$$

Conditional limit laws for the SV process

In this context, different normalization must be found and the limit laws cannot be expressed by means of the exponent measures.

Let $\mathcal{C}$ be a "reasonable" subcone of $[0,\infty]^q \setminus \{0\}$.

Then there exists a nondegenerate Radon measure $\boldsymbol{\nu}_{\mathcal{C}}$ on $\mathcal{C}$, homogeneous with degree $-\alpha\boldsymbol{\beta}_{\mathcal{C}}$ such that

$$\lim_{t\to\infty} \mathbf{P}\left((X_m,\ldots,X_{m+h}) \in B \mid (X_{-q},\ldots,X_{-1}) \in tA\right) = \frac{\mathbf{E}[\mathbf{1}_{\{(X_m,\ldots,X_{m+h})\in B\}}\boldsymbol{\nu}_{\mathcal{C}}(\boldsymbol{\sigma}_q^{-1}\cdot A)]}{\mathbf{E}[\boldsymbol{\nu}_{\mathcal{C}}(\boldsymbol{\sigma}_q^{-1}\cdot A)]}$$

where $\boldsymbol{\sigma}_q = (\sigma_{-q},\ldots,\sigma_{-1})$, and for all locally compact subset A of $\mathcal{C}$ such that $\mathbf{E}[\boldsymbol{\nu}_{\mathcal{C}}(\boldsymbol{\sigma}_q^{-1}\cdot A)] \neq 0$, where

$$\boldsymbol{\sigma}_q^{-1}\cdot A = \left\{\mathbf{x}\in\mathbf{R}^q \mid \prod_{i=1}^{q}\sigma_{-i}x_i \in A\right\} .$$

Examples

- For $m > 0$, $h \geq 0$, and any Borel subset B, if $\mathbf{E}[\sigma_0^{\alpha+\epsilon}] < \infty$ for some $\epsilon > 0$, then

$$\lim_{t\to\infty} \mathbf{P}((X_m, \ldots, X_{m+h}) \in B \mid X_0 > t) = \frac{\mathbf{E}[\mathbf{1}_{\{(X_m,\ldots,X_{m+h})\in B\}}\sigma_0^{\alpha}]}{\mathbf{E}[\sigma_0^{\alpha}]} .$$

Here, $\mathcal{C} = (0, \infty]$, $\beta_{\mathcal{C}} = 1$ and $A = (1, \infty)$.

- For $m > 1$, $h \geq 0$, and any Borel subset B, if $\mathbf{E}[\sigma_0^{\alpha+\epsilon}\sigma_1^{\alpha+\epsilon}] < \infty$,

$$\lim_{t\to\infty} \mathbf{P}((X_m, \ldots, X_{m+h}) \in B \mid X_1 > t, X_0 > t) = \frac{\mathbf{E}[\mathbf{1}_{\{(X_m,\ldots,X_{m+h})\in B\}}\sigma_0^{\alpha}\sigma_1^{\alpha}]}{\mathbf{E}[\sigma_0^{\alpha}\sigma_1^{\alpha}]} .$$

Here $\mathcal{C} = (0, \infty]^2$, $\beta_{\mathcal{C}} = 2$, $A = (1, \infty)^2$.

- For $m > 1$, $h \geq 0$, and any Borel subset B, if $\mathbf{E}[\sigma_0^{\alpha+\epsilon}] < \infty$, then

$$\lim_{t\to\infty} \mathbf{P}((X_m, \ldots, X_{m+h}) \in B \mid X_0 + X_1 > t) = \frac{\mathbf{E}[\mathbf{1}_{\{(X_m,\ldots,X_{m+h})\in B\}}\{\sigma_0^{\alpha} + \sigma_1^{\alpha}\}]}{\mathbf{E}[\sigma_0^{\alpha} + \sigma_1^{\alpha}]}.$$

Here, $\mathcal{C} = [0, \infty]^2 \setminus \{0\}$, $\beta_{\mathcal{C}} = 1$ and $A = \{x, y > 0 \mid x + y > 1\}$.

- In general, the limiting disribution of $(X_m, \ldots, X_{m+h})$ given $(X_1, \ldots, X_h) \in tA$ will have the following form:

$$\lim_{t\to\infty} \mathbf{P}(X_m \leq x_0, \ldots, X_{m+h'} \leq x_h \mid (X_1, \ldots, X_h) \in tA) = \frac{\mathbf{E}[\prod_{i=0}^{h'} F_Z(x_i/\sigma_{m+i}) g(\sigma_1, \ldots, \sigma_h)]}{\mathbf{E}[g(\sigma_1, \ldots, \sigma_h)]}$$

where g is a $\alpha\beta$-homogeneous function which depends on A.

The SV model is a very particular case of asymptotically independent regularly varying time series, where normalizing functions are not needed to obtain non degenerate limit laws. In the general case of asymptotic independence, it is necessary to look for functions a and m such that, e.g.

$$\mathbf{P}((X_m, \ldots, X_{m+h'}) \in m(t) + b(t) \cdot \mid X_0 > t)$$

has a non degenerate limit as $t \to \infty$.

In practice, these functions are unknown and must be estimated. This has not been done in a time series context.

In the case of asymptotic dependence, $m(t) = 0$ and $b(t) = t$.

IV. Estimation of limit conditional laws

We consider only asymptotically dependent regularly varying time series and SV process.

For clarity, we only consider the case of the limiting conditional distribution of X_h given X_0 is extreme.

- Asymptotic dependence

$$\Psi(x) = \lim_{t\to\infty} \mathbf{P}(t^{-1}X_h \le x \mid X_0 > t)$$

- SV process

$$\Psi(x) = \lim_{t\to\infty} \mathbf{P}(X_h \le x \mid X_0 > t) = \frac{\mathbf{E}[\sigma_0^\alpha F_Z(x/\sigma_h)]}{\mathbf{E}[\sigma_0^\alpha]}\ .$$

Davis and Mikosch (2009) have partially studied the estimation of the extremogram $\rho(A, B, m)$ for fixed sets A, B in the case of asymptotic dependence and for strongly mixing time series.

Let $X_{(n:1)} \leq \cdots \leq X_{(n:n)}$ be the order statistics of the observations $X_1, \ldots, X_n$. A natural estimator is given by

- Asymptotic dependence

$$\hat{\Psi}_n(x) = \frac{1}{k} \sum_{j=1}^{n} \mathbf{1}_{\{X_j > X_{(n:n-k)}\}} \mathbf{1}_{\{X_{j+h} \leq X_{(n:n-k)} x\}} \; .$$

- SV process

$$\hat{\Psi}_n(x) = \frac{1}{k} \sum_{j=1}^{n} \mathbf{1}_{\{X_j > X_{(n:n-k)}\}} \mathbf{1}_{\{X_{j+h} \leq x\}} \; .$$

The threshold k is user chosen.

To study $\hat{\Psi}_n$, we introduce the empirical process

$$K_n(s,x) = \frac{1}{k}\sum_{j=1}^{n} \mathbf{1}_{\{X_j > u_n s\}} \mathbf{1}_{\{X_{j+h} \le v_n x\}}$$

with u_n a deterministic threshold such that $k = n\mathbf{P}(X_1 > u_n)$ and $k \to \infty$ and $v_n = u_n$ (asymp. dep.) or $v_n \equiv 1$ (SV). Then

$$\hat{\Psi}_n(x) = K_n(u_n^{-1} X_{(n:n-k)}, u_n^{-1} X_{(n:n-k)} x) \text{ (asymp. dep.)} \quad ,$$
$$\hat{\Psi}_n(x) = K_n(u_n^{-1} X_{(n:n-k)}, x) \text{ (SV)} \quad .$$

and

$$e_n(s) = K_n(u_n^{-1} X_{(n:n-k)} s, \infty) = \frac{1}{k}\sum_{i=1}^{n} \mathbf{1}_{\{X_i > u_n s\}}$$

is the usual univariate tail empirical process.

Results in the asymptotically dependent case

In this case, the standard extreme value theory for dependent data applies. See e.g. Drees (2002,2003), Rootzen (2009).

Three kind of conditions are needed to obtain asymptotic results.

- Mixing conditions. (e.g. strong mixing)
- Some technical conditions specific to extreme value theory, which can be rather hard to check.
- Second order conditions to control the nonparametric bias. Davis and Mikosch (2009) consider "pre-asymptotics" to avoid this issue.

Then if k is chosen according to the (unverifiable) second order conditions, $\sqrt{k}\{K_n(s,x) - K(s,x)\}$ converges to a Gaussian process on $[1,\infty) \times \mathbf{R}$.

This entails the convergence of $\sqrt{k}\{\hat{\Psi}_n(x) - \Psi(x)\}$ to a Gaussian process on $\mathbf{R}$.

Results for the (long memory) stochastic volatility process

The SV process may not be mixing, in particular when the volatility is of the form

$$\sigma_t = \sigma(\xi_t)$$

and $\{\xi_t\}$ is a long memory process.

Nevertheless, under very weak conditions, if k is suitably chosen, i.e. $k \asymp n^{\gamma}$ where γ is related to the second order condition and to the memory of the process $\{\xi_j\}$, then we obtain the same results as in the i.i.d. case

$$\sqrt{k}(K_n(s,x) - s^{-\alpha}\Psi(x)) \Rightarrow \mathbb{W}(s^{-\alpha}, \Psi(x))$$

$$k^{1/2}(\hat{\Psi}_n - \Psi) \Rightarrow \mathbb{B}_1 \circ \Psi \ ,$$

$$k^{1/2}(e_n(s) - s^{-\alpha}) \Rightarrow \mathbb{B}_2(s^{-\alpha}) \ .$$

where $\mathbb{W}$ is a standard Brownain sheet and $\mathbb{B}_i$ are independent standard Brownian bridges.

Kulik and Soulier, forthcoming

Tail empirical process

In the case of the univariate tail empirical process

$$e_n(s) = \frac{1}{k} \sum_{j=1}^{n} \mathbf{1}_{\{X_j > X_{(n:n-k)} s\}} ,$$

long memory does not affect the choice of the threshold k at all: under the *same second order condition on the marginal distribution as in the case of i.i.d. obervations*,

$$\sqrt{k}\{e_n(s) - s^{-\alpha}\} \Rightarrow \mathbb{B}(s^{-\alpha}) .$$

This implies in particular that the Hill estimator of the extreme value index $\gamma = \alpha$ of the marginal distribution of the LMSV model is consistent and asymptotically normal, with the same asymptotic variance and the same rate of convergence as if the observations were i.i.d. Indeed:

$$\hat{\gamma}_n = \int_1^\infty \frac{e_n(s)}{s} \, \mathrm{d}s .$$

Kulik and Soulier (2011).

The theory is relatively well matched by the practice.

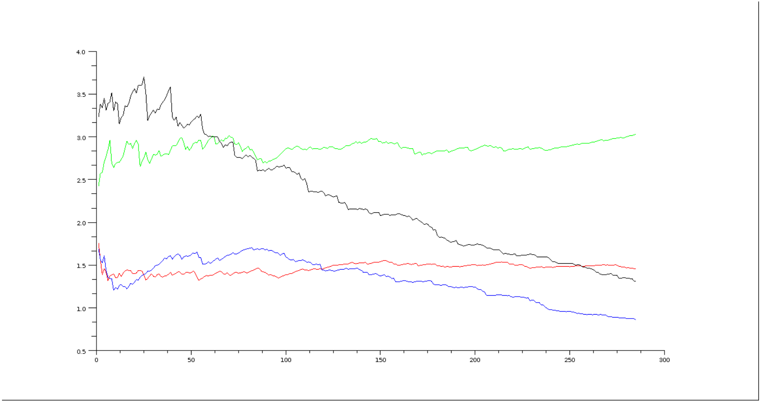


Figure: Hill estimator for i.i.d. (green and red) and LMSV process with $H = .85$ (black and blue) and $\alpha = 3$ or $\alpha = 1.5$.

Test of asymptotic independence

Asymptotic independence for a stationary time series $\{Y_k\}$ can be defined by $\lambda(m) = 0$ for all m, where λ is the extremal coefficient function of Fasen et al (2010):

$$\lambda(m) = \lim_{x \to \infty} \mathbf{P}(Y_m > x \mid Y_0 > x) .$$

For the (LM)SV process, this quantity can be estimated by

$$\hat{\lambda}(m) = \frac{1}{k} \sum_{j=1}^{n} \mathbf{1}_{\{Y_j > Y_{(n:n-k)}\}} \mathbf{1}_{\{Y_{j+m} > Y_{(n:n-k)}\}} .$$

where k is a user chosen threshold.

- For asymptotically dependent time series, under some mixing conditions, and some conditions on k, one can expect that $\lim_{n\to\infty} \hat{\lambda}(m) = \lambda(m) > 0$.
- For the stochastic volatility process, if $k \to \infty$ and $k = o(\sqrt{n})$, then

$$\lim_{n\to\infty} \sqrt{n}\,\hat{\lambda}(t) \to \mathbf{N}(0, \Sigma^2) ,$$

with

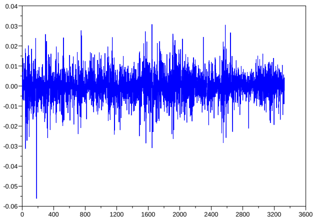
$$\Sigma^2 = \frac{\mathbf{E}[\sigma_0^\alpha \sigma_m^\alpha]}{\mathbf{E}[\sigma_0^\alpha]} .$$

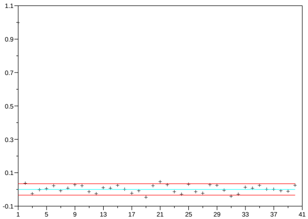
- If the process $\{X_j\}$ is asymptotically dependent, then $\sqrt{n}\,\hat{\lambda}(m) \to \infty$. This provides the basis for a test of stochastic volatility process against asymptotically dependent time series.

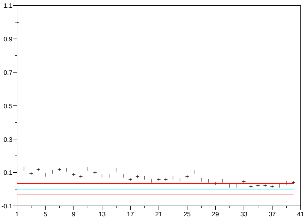
V. Long memory in volatility

Long memory in volatility is one of the “stylised fact” of financial modeling: log-returns of financial time series are white noise sequence but non linear functions (absolute value, powers) can be strongly correlated.

This property is usually evidenced by autocovariance plots.







log-returns (top) autocorrelation (left)
autocorrelation of absolute values (right)

Modeling long memory

In the present framework, short memory will be summability of the autocovariance function of the Gaussian process $\{\xi_j\}$, and long memory means non summability:

$$\sum_{j=1}^{\infty} |\mathrm{cov}(\xi_0, \xi_j)| = \infty \ .$$

We consider here both cases of short and long memory.

It is often assumed more specifically that the autocovariance function is regularly varying at infinity:

$$\mathrm{cov}(\xi_0, \xi_k) = k^{2H-2} L(k)$$

where L is a slowly varying function and $H \in (1/2, 1)$ is called the Hurst index of the process $\{\xi_k\}$.

The desired feature that non linear functions may be correlated is obtained. For instance, if $\sigma(x) = \mathrm{e}^{x/2}$,

$$\mathrm{cov}(\log X_0^2, \log X_k^2) = \mathrm{cov}(\xi_0, \xi_k) \ .$$

The correlation of log-squares is exactly the correlation of the process $\{\xi_k\}$. If $\{\xi_k\}$ has long memory, then so has $\{\log X_k^2\}$.

In the general case, the memory of non linear functions of $\{X_k\}$ can be the same as that of $\{\xi_k\}$ or weaker.

For other functions σ and other non linear functional G, define

$$\bar{G}(x) = \mathbf{E}[G(Z_0\sigma(x))] .$$

Then, if $X_k = Z_k\sigma(X_k)$,

$$\mathrm{cov}(G(X_0), G(X_k)) = \mathrm{cov}(\bar{G}(\xi_0), \bar{G}(\xi_k)) .$$

If τ is the Hermite rank of $\bar{G}$, then

$$|\mathrm{cov}(G(X_0), G(X_k))| = |\mathrm{cov}(\bar{G}(\xi_0), \bar{G}(\xi_k))| \leq \mathrm{var}(G(X_0))\rho_k^{\tau} .$$

Asymptotic theory for long memory processes

If the Gaussian stationary process $\{\xi_j\}$ has long memory Hurst index $H \in (1/2, 1)$, then

$$\mathrm{var}\left(\sum_{j=1}^{n} \xi_0\right) = L(n)n^{2H} \ ,$$

where L is slowly varying and $L(n) \sim H^{-1}(2H-1)^{-1}\ell(n)$ and

$$L^{-1/2}(n)n^{-H}\sum_{k=1}^{[nt]} \xi_k \Rightarrow B_H(t)$$

where B_H is the FBM with Hurst index H, i.e. the only H-self-similar stationary increments Gaussian process.

Subordinated processes

For a function $g : \mathbf{R}^q \to \mathbf{R}$ the behaviour of

$$\sum_{k=1}^{[nt]} g(\xi_{k+1}, \ldots, \xi_{k+q})$$

depends on the Hermite rank τ of g with respect to the distribution of $(\xi_1, \ldots, \xi_q)$.

• The Hermite rank of g is the smallest integer $d \geq 1$ for which there exists a polynomial P of degree d such that

$$\mathbf{E}[P(\xi_1, \ldots, \xi_q) g(\xi_1, \ldots, \xi_q)] \neq 0 \ .$$

- Examples of Hermite ranks.

$$g(x_1, \ldots, x_q) = \prod_{j=1}^{q} \mathrm{e}^{x_j} \ , \quad \tau = 1 \ ,$$

$$g(x_1, \ldots, x_q) = \sum_{j=1}^{q} \mathrm{e}^{x_j} \ , \quad \tau = 1 \ ,$$

If g is componentwise even then $\tau \geq 2$.

Weak convergence of functionals

- If $\tau(1-H) > 1/2$ (which implies $\tau \geq 2$), then

$$n^{-1/2}\sum_{k=1}^{[nt]} g(\xi_{k+1},\ldots,\xi_{k+q}) \Rightarrow cB(t)$$

  where B is the standard Brownian motion.
- If $\tau(1-H) < 1/2$, then

$$n^{-1+\tau(1-H)}\ell^{\tau/2}(n)\sum_{k=1}^{[nt]} g(\xi_{k+1},\ldots,\xi_{k+q}) \Rightarrow Z_\tau(t)$$

  where Z_τ is the FBM wiht Hurst index H if $\tau = 1$ or a non Gaussian, $1-\tau(1-H)$ self-similar process.
  Arcones (1994)

Back to the stochastic volatility model

Recall that we need to study the empirical process

$$K_n(s,x) = \frac{1}{k}\sum_{j=1}^{n} \mathbf{1}_{\{X_j > u_n s\}} \mathbf{1}_{\{X_{j+m} \le x\}} \ .$$

For small k, the behaviour of the empirical process is the same as in the case of i.i.d. observations; for large k, long memory comes into play.

To see this dichotomy, we decompose the empirical process into an "i.i.d. part" and a "LRD part".

Some notation

Let $\mathcal{X}$ be the sigma-field generated by the Gaussian process $\{\xi_j\}$ with covariance $\rho_n = \mathrm{cov}(\xi_0, \xi_n)$. Define

$$G_n(s, x, x_1, x_2) = \frac{\mathbf{E}[X_0 > u_n s\,,\ X_m \le x \mid \xi_0 = x_1; \xi_m = x_2]}{\mathbf{P}(X_0 > u_n)}$$
$$= \frac{\mathbf{P}(Z_0\sigma(x_1) > u_n s\,,\ Z_m\sigma(x_2) \le x)}{\mathbf{P}(X_0 > u_n)}\,.$$

Then for all $x \in (-\infty, \infty)$ and $s \ge 1$,

$$\lim_{n\to\infty} G_n(s, x, x_1, x_2) = \frac{\sigma^\alpha(x_1)\mathbf{P}(Z_m\sigma(x_2) \le x)s^{-\alpha}}{\mathbf{E}[\sigma^\alpha(\xi_0)]}$$
$$\lim_{n\to\infty} \mathbf{E}[G_n(s, x, \xi_0, \xi_n)] = \Psi(x)s^{-\alpha}$$

where Ψ is the limiting conditional distribution of X_m given X_0 is extreme:

$$\Psi(x) = \lim_{t\to\infty} \mathbf{P}(X_m \le x \mid X_0 > x)\,.$$

Decomposition of the empirical process

$$
\begin{aligned}
K_n(s,x) - \Psi(x)s^{-\alpha} \\
&= \frac{1}{k}\sum_{j=1}^{n}\Big\{\mathbf{1}_{\{X_j>u_n s\}}\mathbf{1}_{\{X_{j+h}\le x\}} - \mathbf{P}(X_j > u_n s\,,\; X_{j+h} \le x \mid \mathcal{X})\Big\} \\
&\quad + \frac{1}{n}\sum_{j=1}^{n}\{G_n(s,x,\xi_j,\xi_{j+m}) + \mathbf{E}[G_n(s,x,\xi_j,\xi_{j+m})]\} \\
&\quad + \frac{1}{n}\sum_{j=1}^{n}\mathbf{E}[G_n(s,x,\xi_j,\xi_{j+m})] - \Psi(x)s^{-\alpha} \\
&=: E_{1,n}(s,x) + E_{2,n}(s,x) + B_n(s,x)\ .
\end{aligned}
$$

The last term $B_n(s,x)$ is a bias term and must be dealt with using second order conditions.

The term $E_{1,n}$ is a sum of conditionally independent summands. A central limit theorem can be easily proved by usual techniques (Lindeberg-Lyapounov CLT).

$$k^{1/2}E_{1,n}(s,x) \Rightarrow W(s^{-\alpha}, \Psi(x))$$

where W is a standard Brownian sheet and the convergence is in the Skorohod space $\mathcal{D}((-\infty,\infty) \times [1,\infty))$.

The term $E_{2,n}$ is a sum of functions of bivariate Gaussian vectors. Central and non central limit theorems can be proved using the results of Arcones (1994).

The rate of convergence depends on the Hermite rank τ of the function σ^α.

- If $\tau(1-H) > 1/2$, then $E_{2,n} = O_P(n^{-1/2}) = o_P(k^{-1/2})$, uniformly on compact sets. The term $E_{1,n}$ dominates.
- If $\tau(1-H) < 1/2$, then $\rho_n^{-\tau/2} E_{2,n}(s,x)$ converges to a degenerate process of the form $J(x)s^{-\alpha}\Xi$, where J is a deterministic function and Ξ is a non Gaussian random variable if $\tau > 1$.

We thus have the following dichotomy.

- If $k\rho_n^{\tau} \to 0$ (k *small*), then $E_{2,n} \ll E_{1,n}$ and

$$k^{1/2}\{K_n(s,x) - \mathbf{E}[K_n(s,x)]\} \Rightarrow \mathbb{W}(s^{-\alpha}, \Psi(x)) ,$$

  as for an i.i.d. sequence;

- If $\tau(1-H) < 1/2$ and $k\rho_n^{\tau} \to \infty$ (k *large*), then $E_{1,n} \ll E_{2,n}$ and

$$\rho_n^{-\tau/2}\{K_n(s,x) - \mathbf{E}[K_n(s,x)]\} \to J(x)\Psi(x)\Xi .$$

By Vervaat's Lemma, in both cases, $X_{(n:n-k)}/u_n \to_P 1$. Thus it is possible to replace the unobservable threshold u_n by the order statistic $X_{(n:n-k)}$.

Given a second order condition and a suitable choice of k, we obtain

- If $k\rho_n^\tau \to 0$
$$k^{1/2}\{\hat{\Psi}_n - \Psi\} \Rightarrow \mathbb{B} \circ \Psi$$
where $\mathbb{B}$ is the standard Brownian bridge.
- If $\tau(1-H) < 1/2$ and $k\rho_n^\tau \to \infty$, then $\rho_n^{-\tau/2}\{\hat{\Psi}_n - \Psi\}$ converges to some non Gaussian distribution.

- For the univariate tail empirical process, the long memory part cancels because of the degeneracy of the limiting process. Given the same second order condition as if the observations were i.i.d., we have
$$k^{1/2}\{e_n(s) - s^{-\alpha}\} \Rightarrow \mathbb{B}(s^{-\alpha}) .$$

The choice of k is dictated by two considerations:

- get rid of the bias;
- avoid the LRD zone.

In practice, the LRD convergence is not very useful, so the former results suggest to make an even more conservative (small) choice of k than when LRD is not suspected.

We are currently working on methods for selecting k in the presence of LRD.