Long–Memory and Change–Points in Volatility Processes

Gilles Teyssière

stats@gillesteyssiere.net

September 2009

Introduction

- So far we have considered homogeneous long-memory processes, i.e., processes characterized by a single set of parameters,
- This assumption is unrealistic in finance, as financial time series display also (local) trends, changes in regime, etc;
- We then have to consider methods that
  1. Detect changes in regime for strongly dependent data,
  2. Provide an unbiased estimate of the memory parameter for non-homogeneous time series.

The change-point problem

- Consider the realized time series $\{Y_1, \ldots, Y_T\}$,
- Is this series characterized by a constant vector parameter θ
- Or is this vector changing over time?
- For financial time series, the hypothesis of a constant vector parameter is unlikely.
- Change-point detection of a GARCH process allows to estimate the parameters of this process on the largest interval of homogeneity, and then obtain an unbiased estimate of the volatility.
- This is of interest for
  1. Practitioners using GARCH models for risk management,
  2. Traders using GARCH models for correcting the bias in the Black-Scholes formula.

Change-point detection: the FTSE 100 index

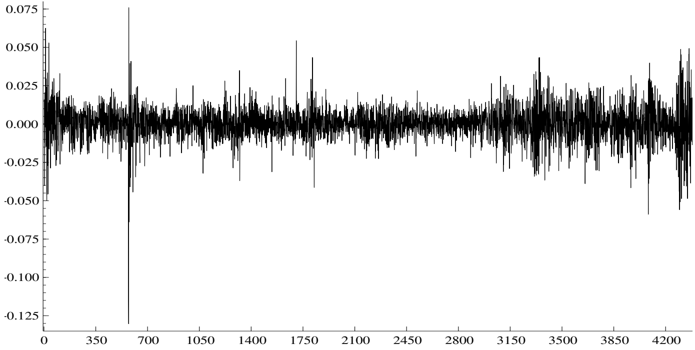


Figure: Returns on the FTSE 100 index $X_t = \log(P_t/P_{t-1})$ (1986–2002)

We observe intermittency of the volatility process: large variations are followed by variations of smaller magnitude.

Splitting the series in shorter intervals with homogeneous variance

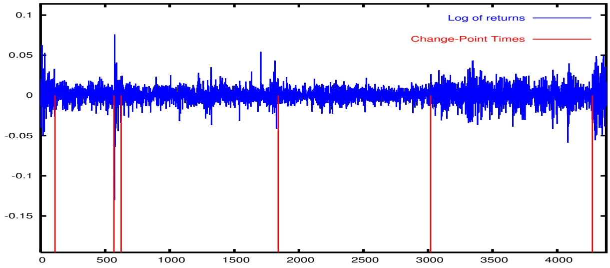


Figure: Returns on the FTSE 100 index $X_t = \log(P_t/P_{t-1})$ (1986–2002)

Note

The Gaussian adaptive method used for this splitting will be exposed later.

Empirical properties of strong dependence in volatility

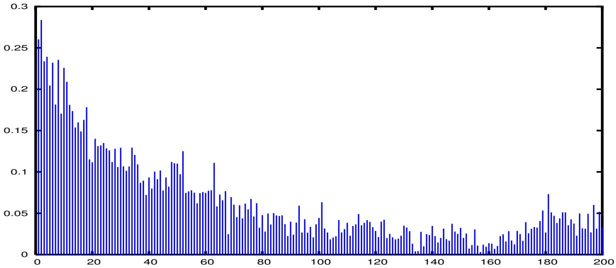


Figure: ACF of $|X_t|$ over the time interval (1986–2002)

Remark

ACF decays hyperbolically to zero, like a long-memory process

Characterization of the dependence structure with the ACF

- For a short-range dependent process, the ACF decays quickly to zero, the decay rate is said exponential.

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| < \infty$$

Example

AR(1) process $Y_t = a_1 Y_{t-1} + \varepsilon$*, which has a stationary solution if* $|a_1| < 1$*,*

$$\gamma(k) = \frac{\sigma^2}{1 - a_1^2} a_1^k, \quad \sigma^2 = Var(\varepsilon)$$

- For a second order stationary strongly dependent process, the ACF decays as follows:

$$\gamma(k) \sim k^{2d-1}, \quad d \in (0, 1/2),$$

Consider the ACF on sub-intervals

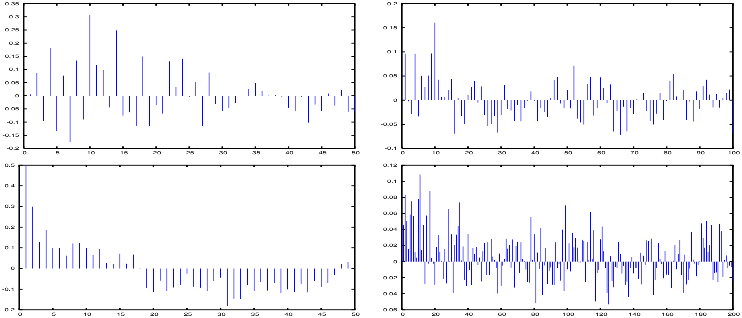


Figure: ACF of $|X_t|$ over the time intervals :
a: Top, left: $[1, 112]$; b: Top, right: $[113, 568]$
c: Bottom,left: $[569, 624]$; d: Bottom, right: $[625, 1840]$

Temporary conclusions

1. Financial time series, when considered as the realization of a single homogeneous process, display some characteristics similar to long-memory processes.
2. However, when studied over sub-intervals, these properties of strong dependence are less obvious.
3. In our particular case, the dependence property was inferred from the asymptotic behavior of the empirical ACF

$$\hat{\rho}_Y(k) = \frac{\hat{\gamma}_Y(k)}{\hat{\gamma}_Y(0)}, \quad \hat{\gamma}_Y(k) = T^{-1} \sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y}), \quad \hat{\gamma}_Y(0) = \mathrm{Var}\,(Y).$$

4. One could wonder whether these methods are appropriate,
5. First, ACF is not much informative if the process is "not very close" to being Gaussian (see Samorodnitsky, 2002) which is the case of financial time series.
6. Next, if the process is not second order stationary, the conclusions drawn from the asymptotic behavior of the ACF are wrong.
7. Finally, the volatility process could mix long-range dependence and change-points.

Change-point detection: parametric tests

- Tests based on the assumption that the process that generates the data is characterized by a finite number of parameters.
- We make the assumption that this process is an ARCH-type process
- These tests are based on the comparison of either the residual process of the ARCH (GARCH) process or likelihood of the process on a moving window splitting the observed sample in two.
- These tests have a drawback, as they consider only the case of a single-change point.

Tests based on the empirical process of squared residuals of ARCH sequences I

- Process ARCH(p) defined by:

$$X_t = \sigma_t \varepsilon_t, \quad \sigma_t^2 = b_0 + \sum_{j=1}^{p} b_j X_{t-j}^2, \quad \varepsilon_t \text{ iid}, \quad E\,\varepsilon_0 = 0, \quad E\,\varepsilon_0^2 = 1,$$

  $b_0 > 0$ and $b_i \geqslant 0$.

- Empirical process of squared residuals $\hat{\varepsilon}_t^2 = \dfrac{X_t^2}{\hat{\sigma}_t^2}$

$$\hat{K}_T(s,t) = \frac{1}{\sqrt{T}} \sum_{1 \leqslant i \leqslant [Ts]} \left[\mathbb{I}\left\{ \hat{\varepsilon}_i^2 \leqslant t \right\} - F(t) \right], \quad 0 < s \leqslant 1.$$

- F repartition function of ε_0^2

Tests based on the empirical process of squared residuals of ARCH sequences II

- Tests based on the process:

$$\hat{T}_T(s,t) = \sqrt{T} \cdot \frac{[Ts]}{T}\left(1 - \frac{[Ts]}{T}\right)\left(\hat{F}_{[Ts]}(t) - \hat{F}^*_{T-[Ts]}(t)\right),$$

with

$$\hat{F}_{[Ts]}(t) = \frac{1}{[Ts]} \sum_{1 \leqslant i \leqslant [Ts]} \mathbf{I}\left\{\hat{\varepsilon}_i^2 \leqslant t\right\}$$

$F^*_{T-[Ts]}(t)$ is analogously defined using indices greater than $[Ts]$.

- The test compares the empirical distribution function of $\hat{\varepsilon}_1^2, \ldots, \hat{\varepsilon}_{[Ts]}^2$ with the one of $\hat{\varepsilon}_{[Ts]+1}^2, \ldots, \hat{\varepsilon}_T^2$
- $\hat{T}_T(s,t)$ has the same limit as $\hat{K}_T(s,t) - \frac{[Ts]}{T}\hat{K}_T(1,t)$

Tests based on the empirical process of squared residuals of ARCH sequences III

Assumptions

1. *Distribution function F of ε_0^2 has a derivative $f(t) = F^{'}(t)$ continuous over* $(0, \infty)$,
2. $\lim_{t\to 0} tf(t) = 0$ *and* $\lim_{t\to\infty} tf(t) = 0$,
3. *The vector of parameters b is estimated by a unbiased estimator $\hat{b} = (\hat{b}_0, \ldots, \hat{b}_p)$ which admits the representation:*

$$\hat{b}_i - b_i = \frac{1}{n} \sum_{1\leqslant j\leqslant n} l_i(\varepsilon_j^2) f_i(\varepsilon_{j-1}, \varepsilon_{j-2}, \ldots) + o_P(T^{-1/2}), \quad 0 \leqslant i \leqslant p,$$

   (The PML estimator satisfies this hypothesis.)
4. *The functions l_i are regular in the following sense*

$$E\, l_i(\varepsilon_0^2) = 0, \quad E\,[l_i(\varepsilon_0^2)]^2 < \infty, \quad E[f_i(\varepsilon_0, \varepsilon_{-1}, \ldots)]^2 < \infty, \quad 0 \leqslant i \leqslant p,$$

5. $E\, \varepsilon_0^4 < \infty$,
6. $(E\, \varepsilon_0^4)^{1/2} \sum_{1\leqslant j\leqslant p} b_j < 1$.

Tests based on the empirical process of squared residuals of ARCH sequences IV

- Under the previous hypotheses, the asymptotic distribution of the empirical process of the squared residuals of an ARCH sequence is

$$\hat{K}_T(s,t) \xrightarrow{d} K(s,t) + stf(t)\xi,$$

  f is the density of ε_i^2, ξ is a Gaussian VA correlated with $K(s,t)$.

- Since

$$\hat{K}_T(s,t) \xrightarrow{d} K(s,t) + stf(t)\xi,$$

  then

$$\begin{aligned} \hat{T}_T(s,t) &= \hat{K}_T(s,t) - \frac{[Ts]}{T}\hat{K}_T(1,t) \\ &\sim (K(s,t) + stf(t)\xi) - s\,(K_T(1,t) + tf(t)\xi) \\ &= K(s,t) - sK(1,t), \end{aligned}$$

- For $1 \leqslant k \leqslant T$ define (change-point date k is unknown)

$$\hat{F}_k(t) = \frac{1}{k}\#\{i \leqslant k : \hat{\varepsilon}_i^2 \leqslant t\}, \quad \hat{F}_k^*(t) = \frac{1}{T-k}\#\{i > k : \hat{\varepsilon}_i^2 \leqslant t\},$$

$$\hat{T}_T(k,t) = \sqrt{T} \cdot \frac{k}{T}\left(1 - \frac{k}{T}\right)\left|\hat{F}_k(t) - \hat{F}_k^*(t)\right|;$$

$$\hat{M}_T = \sup_{0 \leqslant t \leqslant \infty} \max_{1 \leqslant k \leqslant T} |\hat{T}_T(k,t)| = \max_{1 \leqslant k \leqslant T} \max_{1 \leqslant j \leqslant T} |\ \hat{T}_T(k, \hat{\varepsilon}_j^2)\ | \ .$$

- The asymptotic distribution of $\hat{M}_T$ is the same as the generalized Kolmogorov-Smirnov statistic.

Remark

Estimator of the change-point date, (supposed unique):

$$\hat{k}_M = \max\left\{k : \max_{1\leqslant k\leqslant T}\max_{1\leqslant j\leqslant T} \mid \hat{T}_T(k,\hat{\varepsilon}_j^2) \mid\right\} \quad \text{if} \quad \max_{1\leqslant k\leqslant T}\max_{1\leqslant j\leqslant T} \mid \hat{T}_T(k,\hat{\varepsilon}_j^2) \mid \geqslant \hat{c}_T(\alpha),$$

$\hat{c}_T(\alpha)$: *asymptotic or bootstrap critical value at the level* α

- The Cramér – von Mises statistic is defined as:

$$\hat{B} := \int_0^1 \left\{ \frac{1}{T} \sum_{i=1}^{T} [\hat{\mathcal{K}}([Ts], \hat{\varepsilon}_i^2)]^2 \right\} ds$$

which has approximatively the following asymptotic distribution:

$$\hat{B} \sim B := \int_0^1 \int_0^1 \mathcal{K}^2(s, u) duds,$$

where $\mathcal{K}$ is the Kiefer process.

- Critical values are obtained from Blum, Kiefer and Rosenblatt (1961).

Tests based on the generalized likelihood ratio I

- We consider as null hypothesis a GARCH(1,1) model with constants coefficients:

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \quad \sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2,$$

- Under the alternative hypothesis, for $t > t_0$, the process is defined as:

$$X_t = \sigma_t \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2), \quad \sigma_t^2 = \omega^* + \beta^* \sigma_{t-1}^2 + \alpha^* \varepsilon_{t-1}^2,$$

where $\omega^* \neq \omega$, or $\beta^* \neq \beta$, or $\alpha^* \neq \alpha$.

Definition

The generalized likelihood ratio statistic is defined as:

$$\Lambda_{t_0} = \frac{\textit{maximum of the likelihood function under the null hypothesis}}{\textit{maximum of the likelihood function if change-point in } t_0}.$$

see Csörgő and Horváth (1997)

Tests based on the generalized likelihood ratio II

Notations

- $\hat{\omega}$, $\hat{\alpha}$, $\hat{\beta}$: estimated values from the whole sample $X_1, \ldots, X_T$,
- $\tilde{\omega}$, $\tilde{\alpha}$, $\tilde{\beta}$: estimated values from the sub-sample $X_1, \ldots, X_{t_0}$,
- $\bar{\omega}$, $\bar{\alpha}$, $\bar{\beta}$: estimated values from the sub-sample $X_{t_0+1}, \ldots, X_T$.
- Define $\hat{\sigma}_t^2 = \hat{\omega} + \hat{\beta}\hat{\sigma}_{t-1}^2 + \hat{\alpha}\hat{\varepsilon}_{t-1}^2$,
- Define $(\bar{\sigma}_t^2, \tilde{\sigma}_t^2)$ similarly as $\hat{\sigma}_t^2$.

Since the likelihood is Gaussian

$$-2\ln\Lambda_{t_0} = -\left[\sum_{t=1}^{t_0}(\ln\tilde{\sigma}_t^2 - \ln\hat{\sigma}_t^2) + \sum_{t=t_0+1}^{T}(\ln\bar{\sigma}_t^2 - \ln\hat{\sigma}_t^2)\right] + \sum_{t=1}^{n}\frac{\hat{\varepsilon}_t^2}{\hat{\sigma}_t^2} - \sum_{t=1}^{t_0}\frac{\hat{\varepsilon}_t^2}{\tilde{\sigma}_t^2} - \sum_{t=t_0+1}^{T}\frac{\hat{\varepsilon}_t^2}{\bar{\sigma}_t^2}$$

Tests based on the generalized likelihood ratio III

- Since the change-point date t_0 is unknown, we consider the statistic:

$$\Lambda_T^* = \max_{1 \leq k \leq T} -2 \ln \Lambda_k$$

- Even if observations are iid, the statistic Λ_T^* satisfies an Erdös-type limit theorem,with an exponential distribution as limit.
- The critical value, over which H_0 is rejected, is:

$$c_T(\alpha) = \frac{[D_d(\log T) - \log[-\log(1-\alpha)] + \log 2]^2}{2 \log \log T},$$

$$\text{with} \quad D_d(x) = 2 \log x + \frac{d}{2} \log \log x - \log \Gamma\left(\frac{d}{2}\right).$$

Remark

The convergence rate of this limit is very slow, and asymptotic critical values are far greater than the ones obtained in small sample by simulation, see Gombay and Horváth (1996).

Tests based on the generalized likelihood ratio IV

- The generalized likelihood ratio statistic is based on the process

$$\left\{ h(1-h)(-2\ln\Lambda_{[Th]}),\ 0<h<1 \right\}$$

- If observations are independent, with a density depending on b parameters, then this process can be approximated by

$$\left\{ \sum_{i=1}^{b} (W_i^0)^2(h),\ 0<h<1 \right\}$$

where $W_i^0(\cdot), i=1,\dots,b$ are independent Brownian bridges on $[0,1]$.

- Under the null hypothesis of constancy of parameters

$$\Delta_T^* := T^{-3} \sum_{k=1}^{T-1} k(T-k)(-2\ln\Lambda_k) \xrightarrow{d} \int_0^1 \sum_{i=1}^{b} (W_i^0)^2(h)dh.$$

- Critical values for Δ_T^* are obtained from Kiefer (1959).

Test by Inclan and Tiao (1994)

- Based on the process $\{D_T(h), h \in [0,1]\}$

$$D_T(h) := \frac{\sum_{j=1}^{[Th]} X_j^2}{\sum_{j=1}^{T} X_j^2} - \frac{[Th]}{T}, \quad h \in [0,1].$$

- Under the null hypothesis H_0 of constant variance, the process $\{D_T(h), h \in [0,1]\}$ converges to a Brownian bridge over $[0,1]$.
- A test of constancy of the unconditional variance is based on the following functional of the process $\{D_T(h)\}$, which under H_0 converges to the supremum of a Brownian bridge over $[0,1]$.

$$\sqrt{T/2} \sup_{0 \le h \le 1} |D_T(h)| \xrightarrow{d} \sup_{0 \le h \le 1} \left| W^0(h) \right| .$$

CUSUM test by Kokoszka and Leipus (1999) I

- We make the assumption that $\{X_t\}$ is an ARCH(∞) process defined by:

$$
\begin{aligned}
X_t &= \sigma_t \varepsilon_t, \quad \varepsilon_t \sim \text{iid}, \quad E\varepsilon_0 = 0, \quad \text{Var } \varepsilon_0 = 1, \\
\sigma_t^2 &= \omega + \sum_{j=1}^{\infty} \alpha_j X_{t-j}^2, \quad t = 1, \ldots, t_0, \\
\sigma_t^2 &= \omega^\star + \sum_{j=1}^{\infty} \alpha_j^\star X_{t-j}^2, \quad t = t_0 + 1, \ldots, T,
\end{aligned}
$$

  and we make the additional assumption that the unconditional variance changes at time t_0, supposed unknown:

$$
\Delta(n) = \frac{\omega}{1 - \sum_{j=1}^{\infty} \alpha_j} - \frac{\omega^\star}{1 - \sum_{j=1}^{\infty} \alpha_j^\star} \neq 0.
$$

- The null hypothesis is $H_0 : \omega = \omega^\star, \ \alpha_j = \alpha_j^\star$ for all j, while the alternative hypothesis is that there exists j such that $H_A : \omega \neq \omega^\star$ or $\alpha_j \neq \alpha_j^\star$.

CUSUM test by Kokoszka and Leipus (1999) II

- The CUSUM test is based on the process $\{U_T(h), h \in [0,1]\}$ defined by:

$$U_T(h) := \sqrt{T}\frac{[Th](T-[Th])}{T^2}\left(\frac{1}{[Th]}\sum_{j=1}^{[Th]} X_j^2 - \frac{1}{T-[Th]}\sum_{j=[Th]+1}^{T} X_j^2\right),$$

- The test statistic is based on the following functional of the process $\{U_T(h), h \in [0,1]\}$

$$\sup_{0\le h\le 1} |U_T(h)|/\sigma \xrightarrow{d} \sup_{0\le h\le 1} \left|W^0(h)\right|,$$

- The variance σ^2 is usually estimated by spectral methods, i.e., we use the non-parametric estimator $s_T^2(q)$.

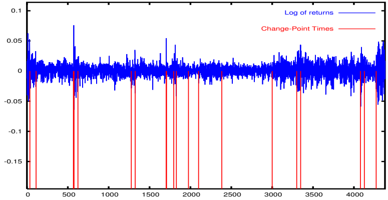
The binary segmentation algorithm

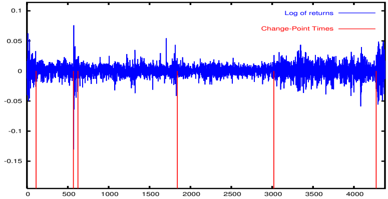
Algorithm proposed by Vostrikova (1981), which allows to extend single change-point methods to the multiple change-point case

Algorithm

1. Apply the single change-point procedure on the whole sample
2. If a change-point is detected, divide the sample in two at the estimated change-point date,
3. Apply the single change-point procedure on these two subsamples,
4. Apply iteratively the procedure until no further change-point is found on the new segments.

Drawbacks of the local method





- Top : Binary segmentation procedure
- Bottom : Adaptive method

Open question

Which segmentation (dimension of the model) is the right one?

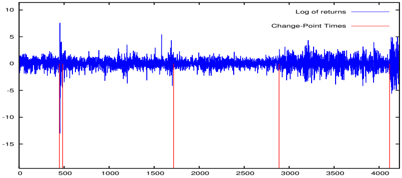
Global method

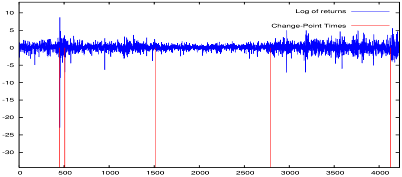
Global method has some interesting properties:

1. This method can detect several change-points at unknown dates,
2. Even if the data are strongly dependent,
3. This method is based on a Gaussian likelihood, and then can be adapted to the multivariate case,
4. This multivariate case is relevant if the series have common change-point times,
5. Finally, even if the data are non Gaussian, this methods works practically well, (interesting results on financial data)

Relevance of the global method

An empirical example: the FTSE 100 and S&P 500 over the time interval 1986–2002





- Adaptive detection of multiple change-point in the univariate case:
- Top: returns on the FTSE 100 index,
- Bottom: returns on the S&P 500 index,
- (The estimated change-point times are represented by vertical lines)

Remark

Change-point times for both series look common.

The global method

- We assume that the m–dimensional process $\{\mathbf{Y}_t = (Y_{1,t}, \ldots, Y_{m,t})'\}$ abruptly changes and is characterized by a (vector) parameter $\theta \in \Theta$, which is constant between two changes.
- Let K be an integer and let $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_{K-1}\}$ be an ordered sequence of integers that verify $0 < \tau_1 < \tau_2 < \ldots < \tau_{K-1} < T$.
- For all $1 \leqslant k \leqslant K$, let $U(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k}; \theta)$ be a contrast function used for estimating the true value of the parameter over the segment k.
- The minimum contrast estimator (MCE) of $\hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k})$, evaluated over the k^{th} segment of $\boldsymbol{\tau}$, is defined as the solution of the following minimization problem:

$$U\left(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k})\right) \leqslant U(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k}; \theta)\,,\ \forall \theta \in \Theta.$$

The global method: contrast function I

- For all $1 \leqslant k \leqslant K$, define G as follows:

$$G(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k}) = U\left(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k}; \hat{\theta}(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k})\right).$$

- We define the contrast function as $J(\boldsymbol{\tau}, \mathbf{Y})$:

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^{K} G(\mathbf{Y}_{\tau_{k-1}+1}, \ldots, \mathbf{Y}_{\tau_k}),$$

with $\tau_0 = 0$ and $\tau_K = T$.

The global method: contrast function II

- We consider the general case of change-point in the covariance matrix of the series $\{\mathbf{Y}_t\}$,
- More precisely, we assume that there exists an integer $K^\star$,, a sequence $\boldsymbol{\tau}^\star = \{\tau_1^\star, \tau_2^\star, \ldots, \tau_{K^\star}^\star\}$ with $\tau_0^\star = 0 < \tau_1^\star < \ldots < \tau_{K^\star-1}^\star < \tau_{K^\star}^\star = T$ and $K^\star$ $(m \times m)$ covariance matrices $\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_{K^\star}$ such that $\mathrm{Cov}\mathbf{Y}_t = \mathbb{E}(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))(\mathbf{Y}_t - \mathbb{E}(\mathbf{Y}_t))' = \boldsymbol{\Sigma}_k$ for $\tau_{k-1}^\star + 1 \leqslant t \leqslant \tau_k^\star$.
- We consider two particular configurations:

The global method: contrast function III

Model M1: There exists an m-dimensional vector μ such that $\mathbb{E}(\mathbf{Y}_t) = \mu$ for $t = 1, 2, \ldots, T$. Further, $\mathbf{\Sigma}_k \neq \mathbf{\Sigma}_{k+1}$ for $1 \leqslant k \leqslant K^\star - 1$.
In the simplest case of change-points in the covariance matrix, without change in the mean, which is of interest for multivariate volatility processes, we use the following contrast function based on the Gaussian log-likelihood function:

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^{K} n_k \log |\widehat{\mathbf{\Sigma}}_{\tau_k}|, \qquad (C_1)$$

where $n_k = \tau_k - \tau_{k-1}$ is the length of the segment k, $\widehat{\mathbf{\Sigma}}_{\tau_k}$ is the $(m \times m)$ covariance matrix evaluated on the segment k:

$$\widehat{\mathbf{\Sigma}}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}})(\mathbf{Y}_t - \bar{\mathbf{Y}})'.$$

$\bar{\mathbf{Y}} = T^{-1} \sum_{t=1}^{T} \mathbf{Y}_t$: sample mean of the m–dimensional vector $\mathbf{Y}_t$ evaluated on the whole sample.

The global method: contrast function IV

Model M2: There exists $K^\star$ m–dimensional vectors $\mu_1, \ldots \mu_{K^\star}$ such that $\mathbb{E}(\mathbf{Y}_t) = \mu_k$ for $\tau^\star_{k-1} + 1 \leqslant t \leqslant \tau^\star_k$. Furthermore, $(\mu_k, \mathbf{\Sigma}_k) \neq (\mu_{k+1}, \mathbf{\Sigma}_{k+1})$ for $1 \leqslant k \leqslant K^\star - 1$.
For the detection of change-points in the vector of mean and/or covariance matrix of a multivariate sequence of random variables, the contrast function reduces to;

$$J(\boldsymbol{\tau}, \mathbf{Y}) = \frac{1}{T} \sum_{k=1}^{K} n_k \log |\widehat{\mathbf{\Sigma}}_{\tau_k}| \qquad (C_2)$$

where the covariance matrix $\widehat{\mathbf{\Sigma}}_{\tau_k}$ of dimension $(m \times m)$ is evaluated over the segment k :

$$\widehat{\mathbf{\Sigma}}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})'$$

$\bar{\mathbf{Y}}_{\tau_k} = n_k^{-1} \sum_{t=\tau_{k-1}+1}^{\tau_k} \mathbf{Y}_t$: sample mean of the m–dimensional vector $\mathbf{Y}_t$ over that segment.

The global method: contrast function V

Asymptotic results for the minimum contrast estimator of $\boldsymbol{\tau}^\star$ are obtained in the following framework:

- **A1** For all $1 \leqslant i \leqslant m$ and all $1 \leqslant t \leq T$, define $\eta_{t,i} = Y_{t,i} - \mathbb{E}\,(Y_{t,i})$. There exists $C > 0$ and $1 \leqslant h < 2$ such that for all $u \geq 0$ and all $s \geq 1$,

$$\mathbb{E}\left(\sum_{t=u+1}^{u+s} \eta_{t,i}\right)^2 \leqslant C(\theta)s^h.$$

  (**A1** is verified for $h = 1$ for both weakly dependent processes and $1 < h < 2$ for strongly dependent processes.)

- **A2** There exists a sequence $0 < a_1 < a_2 < \ldots < a_{K^\star-1} < a_{K^\star} = 1$ such that for all $T \geqslant 1$ and for all $1 \leqslant k \leqslant K^\star - 1$, $\tau_k^\star = [Ta_k]$.

The global method: contrast function VI

When the real number of segments $K^\star$ is known, we have the following result on the convergence rate of the MCE of $\boldsymbol{\tau}^\star$:

Theorem

Suppose that **A1-A2** *are verified. In the case* **M1** *(resp.* **M2***), let $\hat{\boldsymbol{\tau}}_T$ be the dates that minimize the empirical contrast $J(\boldsymbol{\tau}, \mathbf{Y})$ defined by equation (C_1) (resp. (C_2)). Then, the sequence $\{T\|\hat{\boldsymbol{\tau}}_T - \boldsymbol{\tau}^\star\|_\infty\}$ is uniformly tight in probability:*

$$\lim_{T\to\infty} \lim_{\delta\to\infty} \mathrm{P}\left(\max_{1\leqslant k\leqslant K^\star-1} |\hat{\tau}_{T,k} - \tau_k^\star| > \delta \right) = 0. \qquad (C_3)$$

(Here, $J(\boldsymbol{\tau}, \mathbf{Y})$ is minimized over all possible sequences $\boldsymbol{\tau}$ with length $K^\star$)

The global method: contrast function VII

We use a Gaussian contrast function:

- Univariate case: $J(\boldsymbol{\tau}, \mathbf{Y}) = \dfrac{1}{T} \sum_{k=1}^{K} T_k \log \sigma_k^2$, $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_{K-1}\}$ is an ordered sequence of integers satisfying $0 < \tau_1 < \tau_2 < \ldots < \tau_{K-1} < T$, $n_k = \tau_k - \tau_{k-1}$ is the length of the segment k,
- Multivariate case: $J(\boldsymbol{\tau}, \mathbf{Y}) = \dfrac{1}{T} \sum_{k=1}^{K} T_k \log |\widehat{\boldsymbol{\Sigma}}_{\tau_k}|$

  The $(m \times m)$ empirical covariance matrix $\widehat{\boldsymbol{\Sigma}}_{\tau_k}$ is evaluated over the segment k

$$\widehat{\boldsymbol{\Sigma}}_{\tau_k} = \frac{1}{n_k} \sum_{t=\tau_{k-1}+1}^{\tau_k} (\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{\tau_k})'$$

  $\bar{\mathbf{Y}}_{\tau_k} = T_k^{-1} \sum_{t=\tau_{k-1}+1}^{\tau_k} \mathbf{Y}_t$: sample mean of the m–dimensional vector $\mathbf{Y}_t$ over that segment.

The global method: contrast function VIII

Change-point times are estimated by minimizing the penalized contrast function

$$J(\boldsymbol{\tau}, \mathbf{y}) + \beta \mathrm{pen}(\boldsymbol{\tau}) = J(\boldsymbol{\tau}, \mathbf{y}) + \beta_T K$$

where

1. $\beta_T K$: penalty term that controls the level of resolution of the segmentation $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_{K-1}\}$.
2. If β is a function of T that tends to 0 when T tends to ∞, the following theorem states that the number of estimated segments converges in probability to $K^\star$.

Penalty term I

Theorem

Let $\{\beta_T\}$ *be a sequence of positive real numbers such that*

$$\beta_T \underset{T\to\infty}{\longrightarrow} 0 \text{ and } T^{2-h}\beta_T \underset{T\to\infty}{\longrightarrow} \infty, \quad 1 \leqslant h < 2.$$

Then, under the hypotheses **A1-A2**, *the estimated number of segments* $K(\hat{\boldsymbol{\tau}}_T)$, *where* $\hat{\boldsymbol{\tau}}_T$ *is the penalized MCE estimator of* $\boldsymbol{\tau}^\star$, *i.e., obtained by minimizing* $J(\boldsymbol{\tau}, \mathbf{Y}) + \beta_T \mathrm{pen}(\boldsymbol{\tau})$, *converges in probability to* $K^\star$.

Remark

The contrast function $J(\boldsymbol{\tau}, \mathbf{Y})$ *if minimized over the set of all possible sequences* $\boldsymbol{\tau}$ *and for all possible* K, $1 \leq K \leq K_{max}$, K_{max} *is a finite upper bound of* $K^\star$

Penalty term II

Standard choice for β (that overestimate the number of change-points)

- $\beta_T = \log(T)/T$ (BIC)
- $\beta_T = 4\log(T)/T^{1-2d}$ for strongly dependent series.

- The parameter d is unknown,
- How to estimate d from the data?
- If the process is not stationary, standard estimation method seen in the previous lectures provide incorrect results that overestimate the parameter d and artificially increase β.
- We can use methods robust to nonstationarity, like wavelets methods based on the scaling properties of the wavelets coefficients, but these methods require a large sample size.
- We then resort to adaptive methods, i.e., methods such that the segmentation does not depend too much on β.
- These methods consist in finding a change in the curvature of (K, J_K) : K is chosen so that J_K ceases to significantly decrease.

Adaptive choice for the penalty parameter I

- Assume that the penalization $\mathrm{pen}(\boldsymbol{\tau})$ depends only on the dimension of the model, i.e., the number of segments K. Set

$$
\begin{aligned}
J_K &= J(\hat{\boldsymbol{\tau}}_K, \mathbf{Y}), \\
p_K &= \mathrm{pen}(\boldsymbol{\tau}), \quad \forall \boldsymbol{\tau} \in \mathcal{T}_K \\
\hat{p}_K &= \mathrm{pen}(\hat{\boldsymbol{\tau}}_K).
\end{aligned}
$$

- Thus, for any penalty parameter $\beta > 0$, the solution $\hat{\boldsymbol{\tau}}(\beta)$ minimize the penalized contrast:

$$
\begin{aligned}
\hat{\boldsymbol{\tau}}(\beta) &= \arg\min_{\boldsymbol{\tau}} (J(\boldsymbol{\tau}, \mathbf{Y}) + \beta \mathrm{pen}(\boldsymbol{\tau})) \\
&= \hat{\boldsymbol{\tau}}_{\hat{K}(\beta)}
\end{aligned}
$$

where

$$
\hat{K}(\beta) = \arg\min_{K \geqslant 1} \{J_K + \beta p_K\}.
$$

Adaptive choice for the penalty parameter II

- The solution $\hat{K}(\beta)$ is a piecewise constant function of β. If $\hat{K}(\beta) = K$,

$$J_K + \beta p_K \quad < \quad \min_{L \neq K}(J_L + \beta p_L).$$

- The, β verifies

$$\max_{L>K} \frac{J_K - J_L}{p_L - p_K} < \beta < \min_{L<K} \frac{J_L - J_K}{p_K - p_L}.$$

- Then, there exists a sequence $\{K_1 = 1 < K_2 < \ldots\}$, and a sequence $\{\beta_0 = \infty > \beta_1 > \ldots\}$, with

$$\beta_i = \frac{J_{K_i} - J_{K_{i+1}}}{p_{K_{i+1}} - p_{K_i}} \ , \quad i \geqslant 1,$$

  such that $\hat{K}(\beta) = K_i$, $\forall \beta \in [\beta_i, \beta_{i-1})$.
- The set $\{(p_{K_i}, J_{K_i}), i \geqslant 1\}$ is the convex envelope of the set $\{(p_K, J_K), K \geqslant 1\}$.

Adaptive choice for the penalty parameter III

- The estimated sequence $\hat{\boldsymbol{\tau}}(\beta)$ should not depend too much on the choice for the penalty parameter β (a small variation of β should not yield a very different solution $\hat{\boldsymbol{\tau}}$).
- This stability of the solution with respect to the choice for β will be insured if we consider the largest intervals $[\beta_i, \beta_{i-1}), i \geqslant 1$.

We use the following procedure:

1. for $K = 1, 2, \ldots, K_{MAX}$, compute $\hat{\boldsymbol{\tau}}_K$, $J_K = J(\hat{\boldsymbol{\tau}}_K, \mathbf{Y})$ and $\hat{p}_K = \mathrm{pen}(\hat{\boldsymbol{\tau}}_K)$,
2. compute the sequence $\{K_i\}$ and $\{\beta_i\}$, and the lengths $\{l_{K_i}\}$ of the intervals $[\beta_i, \beta_{i-1})$,
3. consider the highest value of K_i such that $l_{K_i} \gg l_{K_j}$, for $j > i$.

Adaptive choice for the penalty parameter: heuristic approach I

Graphical method for selecting the dimension K could be summarized as follows:

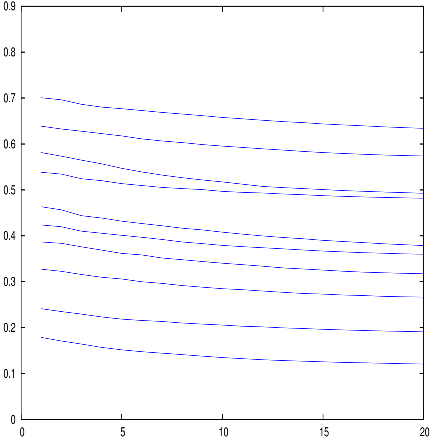
1. examine the decrease of the contrast J_K when K increases
2. select K for which J_K ceases to significantly decrease.

- This amounts to look for the maximal curvature of $(p_K\,,\,J_K)$,
- Second derivative of that curve is linked to the length of the interval $([\beta_i, \beta_{i-1}), i \geqslant 1)$,
- Looking for a change in the decrease amounts to look for a change in the slope of the curve.

Adaptive choice for the penalty parameter: heuristic approach II

- This requires a careful inspection of the curve, and is difficult to automatize.
- Consider an alternative approach more easy to automatize
- Principle of the method: model the decrease of the sequence of contrasts $\{J_K\}$ when there is no change point in the series $\{\mathbf{Y}_t\}$ and look for which value of K this model adjusts the $\{J_K\}$
- No analytical solution for finding the distribution of the $\{J_K\}$
- Some simulations show that this sequence decreases as $c_1 K + c_2 K \log(K)$.

Adaptive choice for the penalty parameter: heuristic approach III



- Ten sequences of contrast function $\{J_K\}$ computed from 10 sequences of Gaussian iid. random variable with coefficient of correlation $\rho = 0.5$
- The fit to the function $c_1 K + c_2 K \log(K)$ is almost perfect ($r^2 > 0.999$). (estimated coefficients $\hat{c}_1$ and $\hat{c}_2$ are different for each of these series)

Adaptive choice for the penalty parameter: heuristic approach IV

Algorithm

For $i = 1, 2, \ldots,$

1. Fit the model:

$$J_K = c_1 K + c_2 K \log(K) + e_K,$$

to the series $\{J_K, K \geqslant K_i\}$, assuming that $\{e_K\}$ is a sequence of Gaussian iid. and centered random variables.

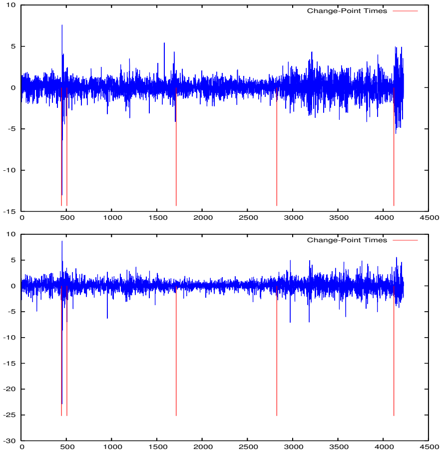
2. Evaluate the probability that J_{K_i-1} follows this model, i.e., the probability

$$\mathcal{P}_{K_i} = P\big(e_{K_i-1} \geqslant J_{K_i-1} - \hat{c}_1(K_i - 1) + \hat{c}_2(K_i - 1)\log(K_i - 1)\big),$$

under the estimated model.

3. The estimated number of segments is the highest value of K_i such that the P–value $\mathcal{P}_{K_i}$ is lower than a threshold α. (We set $\alpha = 10^{-7}$ and $K_{MAX} = 20$)

Application to the bivariate series of the FT100 and S&P 500 indices



- Adaptive detection of change-points in multivariate series
- Top: returns on FTSE 100
- Bottom: returns on S&P 500

Practical applications

- On multivariate GARCH processes, this method detects the change points with a great precision (with a greater precision than in the univariate case)
- This method provides better results than parametric tests extended to the multivariate case (e.g., the multivariate generalized likelihood ratio test)
- On real data, the automatic method detects the significant changes (stock market crashes)
- However, the user could tune the level of resolution by choosing the level of the P–value $\mathcal{P}_{K_i}$.

Wavelet analysis for long-memory processes I

Definition

Let $\{Y_t, t \in \mathbb{R}\}$ be a second–order stationary process. This process is a long–memory process if its spectrum $f_Y(\lambda)$ is such that in a close positive neighborhood of the zero frequency,

$$f_Y(\lambda) \sim c_f \lambda^{-\alpha}, \quad \lambda \to 0_+, \quad c_f \in (0, \infty),$$

or equivalently, if its autocorrelation function $\rho_Y(k)$ has the following hyperbolic rate of decay[a]

$$\rho_Y(k) \asymp k^{\alpha - 1},$$

with $\alpha \in (0, 1)$.

[a] $x_k \asymp y_k$ means that $\exists$ two constants C_1, C_2 such that $C_1 y_k \leqslant x_k \leqslant C_2 y_k$, $k \to \infty$.

Wavelet analysis for long-memory processes II

- Statistical tools based on global statistics are not robust to trends, breaks, etc,
- We then consider an alternative statistical method more robust to these nonlinearities,
- Wavelets based methods are robust to trend, change-points and nonlinearities
- So that they allow us to adjudicate between long-range dependence, non-stationarities, nonlinearities, change-points (Teyssiere and Abry, 2005).

Wavelet estimator of d

- The mother wavelet ψ_0 has $\mathcal{N}$ moments of order zero, with $\mathcal{N} \geq 1$, i.e.,

$$\int t^k \psi_0(t) dt \equiv 0, \quad k = 0, \ldots, \mathcal{N} - 1.$$

- $\psi_{j,k}$ is a family of waveforms $\left\{\psi_{j,k} = 2^{-j/2}\psi_0(2^{-j}t - k)\right\}$, i.e., a collection of dilations and translations of the mother wavelet ψ_0
- $j = 1, \ldots, J$ are the octaves, $k \in \mathbb{Z}$
- Discrete wavelet coefficients are defined by: $d_X(j,k) = \int_{\mathbb{R}} X(t)\psi_{j,k}(t)\, dt$
- Veitch an Abry (1999) estimator of $\alpha = 2d$, based on the DWT and the properties of independence of wavelet coefficients $d_X(j,k)$ of Gaussian fractional processes.

Daubechies wavelets

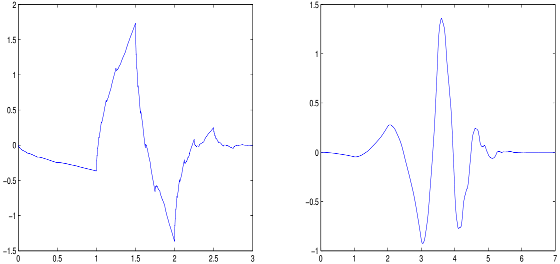


Figure: Daubechies 2 (left) and Daubechies 4 (right)

- We use here Daubechies wavelets,
- Daubechies wavelets have a support of minimum size for a given number of $\mathcal{N}$ vanishing moments.

Moments of order zero of a wavelet

- The moments of order zero $\mathcal{N}$ is a key variable for wavelet analysis.

Moments of order zero of a wavelet

- The moments of order zero $\mathcal{N}$ is a key variable for wavelet analysis.
- This number has to be chosen after (visual) inspection of the series

Moments of order zero of a wavelet

- The moments of order zero $\mathcal{N}$ is a key variable for wavelet analysis.
- This number has to be chosen after (visual) inspection of the series
- By definition of $\mathcal{N}$, the coefficients of any polynomial of order $P < \mathcal{N}$ are equal to zero, $d_P(j, k) \equiv 0$.

Moments of order zero of a wavelet

- The moments of order zero $\mathcal{N}$ is a key variable for wavelet analysis.
- This number has to be chosen after (visual) inspection of the series
- By definition of $\mathcal{N}$, the coefficients of any polynomial of order $P < \mathcal{N}$ are equal to zero, $d_P(j,k) \equiv 0$.
- For a strongly dependent process (centered) X_t, the surimposition of a non stationary polynomial does not affect the estimation of the long memory parameter as long as the degree of the polynomial is less than the number of non-null moments $\mathcal{N}$.

Moments of order zero of a wavelet

- The moments of order zero $\mathcal{N}$ is a key variable for wavelet analysis.
- This number has to be chosen after (visual) inspection of the series
- By definition of $\mathcal{N}$, the coefficients of any polynomial of order $P < \mathcal{N}$ are equal to zero, $d_P(j,k) \equiv 0$.
- For a strongly dependent process (centered) X_t, the surimposition of a non stationary polynomial does not affect the estimation of the long memory parameter as long as the degree of the polynomial is less than the number of non-null moments $\mathcal{N}$.
- We choose $\mathcal{N}$ for obtaining an estimate of the long-memory parameter which cannot be affected by trends and non-stationarities.

Example of strongly dependent time series with a trend

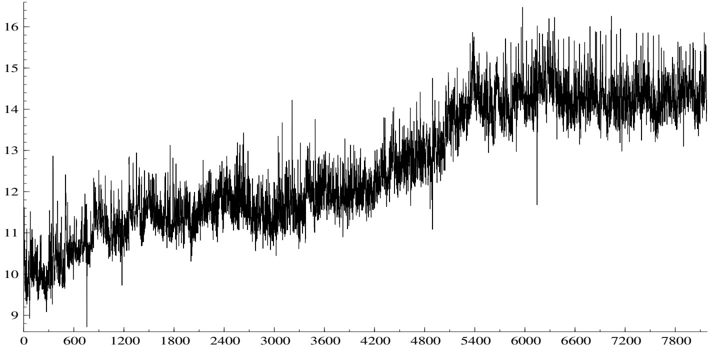


Figure: Logarithm of the volume of transactions on AT&T

Estimation of d

- The process $d_X(j,k)$ is stationary (if $\mathcal{N} \geqslant (\alpha-1)/2$)
- For the largest octaves, its variance satisfies the following power law:

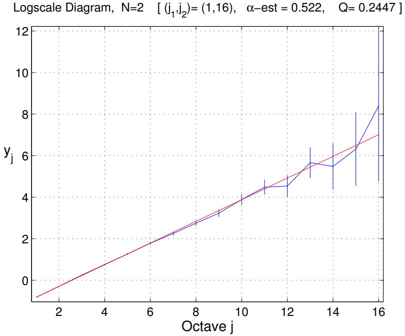
$$Ed_X(j,\cdot)^2 = 2^{j\alpha} c_f C, \quad \text{as } 2^j \to \infty, \qquad C = \int |\lambda|^{-\alpha} |\Psi_0(\lambda)|^2 d\lambda,$$

  ($\Psi_0(\lambda)$ is the Fourier transform of the mother wavelet ψ_0.)
- The scaling parameter α is estimated from the slope of the following linear regression:

$$\log_2\left(Ed_X(j,\cdot)^2\right) = j\alpha + \log_2(c_f C) \qquad \text{called “logscale diagram”}$$

- Estimation of this regression by weighted least squares

Example of a logscale diagram



- "Logscale diagram" of a fractional Gaussian noise, $d = 0.25$
- $\hat{\alpha} = 2\hat{d} = 0.522$
- We could select all octaves as the scaling law appears from the first octave.
- As we will see later, for nonlinear LRD processes the first octaves are affected by the presence of nonlinearities
- For nonlinear LRD processes, the scaling law clearly appears in the largest octaves.

Asymptotic distribution of the estimator

- Define: $S_X(j) = \frac{1}{n_j} \sum_{k=1}^{n_j} d_X(j,k)^2$

  n_j: number of wavelet coefficients $d_X(j,k)$ available at octave j, $n_j = O(2^{-j}T)$
- Wavelet estimator (for the range of octaves $[j_1, j_2]$)

$$\hat{\alpha}_W = \sum_{j=j_1}^{j_2} w_j(\log_2 S_X(j) - (\psi(n_j/2)/\log 2 - \log_2(n_j/2)))$$

$$w_j = \frac{1}{a_j}\frac{S_0 j - S_1}{S_0 S_2 - S_1^2}, \quad S_p = \sum_{j=j_1}^{j_2} j^p/a_j, \quad a_j = \zeta(2, n_j/2)$$

- This estimator has "approximatively" the following asymptotic distribution:

$$(\hat{\alpha} - \alpha) \sim N\left(0, \frac{1}{T \ln^2(2) 2^{1-j_1}}\right),$$

- j_1 is the lowest octave, the long range behavior is captured by the octaves greater than j_1.

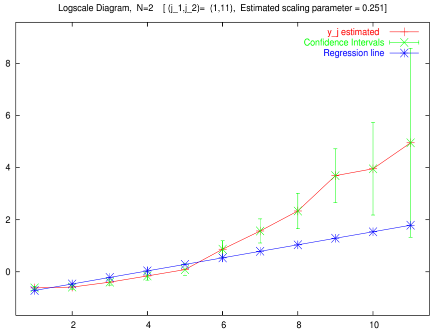
How to select the lowest octave j_1 ?

- If j_1 is too small : strong bias as the interval contains some octave that do not verify the scaling law, but only short term dependencies and non-linearities.
- If j_1 is too "large": bias is reduced but the variance becomes large
- Selection of j_1 in relation with the problem of optimal bandwidth selection for the local Whittle estimator in the frequency domain
- The octave associated with the optimal bandwidth m_{LW}^{opt} of the local Whittle estimator is then equal to m_{LW}^{opt}/T, and matches the octave 2^{-j_1}.
- Using m_{LW}^{opt} we define the optimal lower octave :

$$j_1^{opt} = \left[\frac{\log T - \log m_{LW}^{opt}}{\log 2} \right],$$

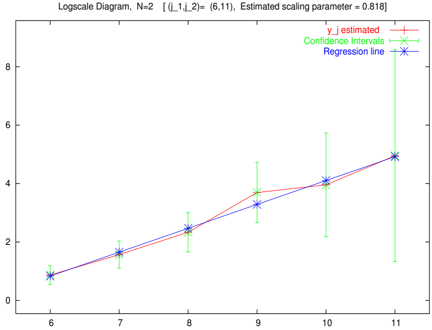
- This gives satisfactory results. (So far, there is no alternative method)

Example: selecting j_1 for nonlinear models I

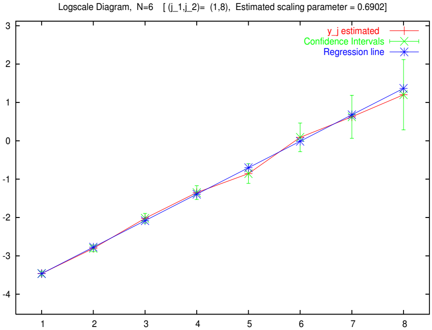


- "Logscale diagram" of a long-memory stochastic volatility model (LMSV) with $\alpha = 2d = 0.90$
- $\hat{\alpha} = 2\hat{d} = 0.25$,
- $\hat{\alpha} \ll \alpha$ as we wrongly select all octaves while the scaling law does not appear in the first octaves,
- The scaling law appears in the largest octaves,
- The first octaves are affected by the short-range nonlinearities of the LMSV.

Example: selecting j_1 for nonlinear models II

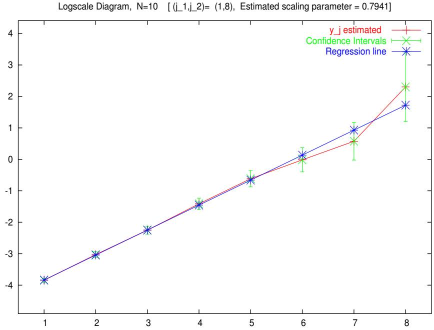


- "Logscale diagram" of a long-memory stochastic volatility model (LMSV) $\alpha = 2d = 0.90$
- $\hat{\alpha} = 2\hat{d} = 0.818$
- We set $j_1 = 6$ as the scaling law appears after that octave.
- Same issue is present for other nonlinear LRD processes; See Teyssière and Abry (2005) for further details.



- Logscale diagram for the log-volume of transactions on AT&T shares , $j_1 = 1$, $\mathcal{N} = 6$; $\hat{\alpha} = 0.6902$
- We select $\mathcal{N} = 6$ because of the trend (estimation results are stable for $\mathcal{N} \geqslant 6$)
- We select all octaves as the scaling law appears from the first octave.

Volume of transactions: how to select $\mathcal{N}$? II

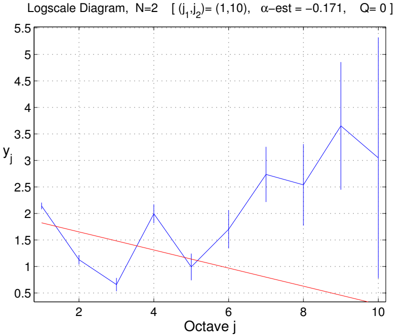


- Logscale diagram for the log-volume of transactions on IBM shares , $j_1 = 1$, $\mathcal{N} = 10$; $\hat{\alpha} = 2\hat{d} = 0.7941$
- We select $\mathcal{N} = 10$ because of the trend (estimation results are stable for $\mathcal{N} \geqslant 10$)
- We select all octaves as the scaling law appears from the first octave.

Are volume and volatility sharing the same long-memory properties ?

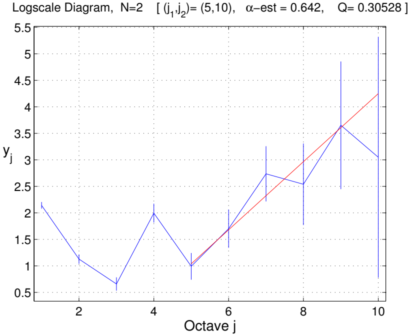
- Previous works on the common long-memory properties of prices and volume
- Both processes are supposed to inherit their long-memory properties from a news arrival process (the so-called mixture of distribution hypothesis, MDH)
- We apply the wavelet estimator to the same series used by Lobato and Velasco (2000).

Logscale diagrams for volatility series I



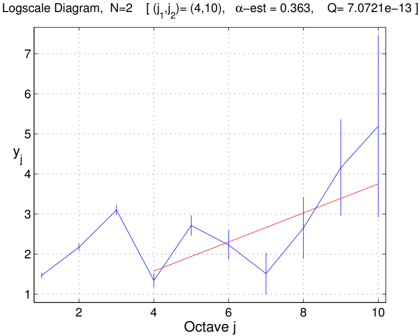
- Logscale diagram for the absolute returns on AT&T shares, $j_1 = 1$, $\mathcal{N} = 2$; $\hat{\alpha} = 2\hat{d} = -0.171$
- It is obvious that the scaling law does not appear from the first octave
- The volatility process is less “nice” than the volume process.

Logscale diagrams for volatility series II



- Logscale diagram for the absolute returns on AT&T shares, $j_1 = 5$, $\mathcal{N} = 2$; $\hat{\alpha} = 2\hat{d} = 0.642$
- Volume and volatility processes appear to have different scaling properties. (See Teyssière and Abry (2005) for further details).

Logscale diagrams for volatility series III



- Logscale diagram for the absolute returns on IBM shares, $j_1 = 4$, $\mathcal{N} = 2$; $\hat{\alpha} = 2\hat{d} = 0.363$
- Volume and volatility processes appear to have different scaling properties.
- Volatility series appear to have a lower degree of long-memory than was is usually claimed using other non robusts estimators. (See Teyssière and Abry (2005) for other examples).

References I

Abry, P., Veitch, D. (1998). Wavelet analysis of long-range dependent traffic. *IEEE Transactions on Information Theory.*, **44**, 2–15.

Abry, P., Flandrin, P., Taqqu, M.S., Veitch, D. (2000). Wavelets for the analysis, estimation, and synthesis of scaling data. In: Park, K. and Willinger, W. (Eds), *Self Similar Network Traffic and Performance Evaluation*, 39–88, Wiley.

Abry, P., Teyssière, G. (2009). Wavelet analysis of high-frequency financial data. *Preprint*

Blum, J. R., Kiefer, J., Rosenblatt, M. (1961). Distribution free tests of independence based on the sample distribution function. *The Annals of Mathematical Statistics*, **32**, 485–498.

Csörgö, M., Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley.

Gombay, E., Horváth, L. (1996). On the rate of approximations for maximum likelihood tests in change-point models. *Journal of Multivariate Analysis*, **56**, 120–152.

Horváth, L., Kokoszka, P.S., Teyssière, G. (2001). Empirical processes for the squared residuals of an ARCH sequence. *The Annals of Statistics*, **29**, 445–469.

Horváth, L., Kokoszka, P.S., Teyssière, G. (2004). Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. *Journal of Statistical Computation and Simulation*, **74**, 469–485.

Inclán, C., Tiao, G.C. (1994). Use of cumulative sums of squares for retrospective detection of changes in variance. *Journal of the American Statistical Association*, **89**, 913–923.

References II

- Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing: Detection Theory*. Prentice Hall.
- Kiefer, J. (1959). K-sample analogues of the Kolmogorov–Smirnov and Cramér-v. Mises tests. *Annals of Mathematical Statistics*, **30**, 420–447.
- Kokoszka, P. S., Leipus, R. (1999). Testing for parameter changes in ARCH models. *Lithuanian Mathematical Journal*, **39**, 231–247.
- Kokoszka, P.S., Teyssière, G. (2002). Change–point detection in GARCH models: Asymptotic and bootstrap tests. *Preprint*.
- Lavielle, M., Teyssière, G. (2006). Detection of multiple change–points in multivariate time Series. *Lithuanian Mathematical Journal*, **46**, 287–306.
- Lobato, I. N., Velasco, C. (2000). Long-memory in stock-market trading volume. *Journal of Business & Economic Statistics*, **18**, 410–427.
- Mallat, S. (2009). *A Wavelet Tour of Signal Processing. The Sparse Way*. Academic Press. Third edition.
- Samorodnitsky, G. (2002). *Lecture Notes on Long Range Dependence*, Copenhagen.
- Teyssière, G., Abry, P. (2005). Wavelet analysis of nonlinear long–range dependent processes. Applications to financial time Series. In *Long Memory in Economics*. Springer Verlag, Berlin, pp. 173–238.
- Vostrikova, L. Yu. (1981). Detecting disorder in multidimensional random processes. *Soviet Mathematics Doklady*, **24**, 55–59.