A two-sample test for comparison of long memory parameters

F. Lavancier[1], A. Philippe[1], D. Surgailis[2]

[1]Laboratoire Jean Leray, Université de Nantes

[2]Institute of Mathematics and Informatics, Vilnius

JSTAR 2008

1 Introduction

Introduction

Let $X(t)$, $t \in \mathbb{Z}$, be a second order, stationary time series and let $\gamma(h)$ be its covariance function, i.e.

$$\gamma(h) = \mathrm{cov}(X(t), X(t+h)).$$

We say that X exhibit long memory when its covariance function is not summable :

$$\sum_{h \in \mathbb{Z}} |\gamma(h)| = \infty.$$

<u>Exemple :</u> FARIMA(p,d,q) processes, defined for $p \in \mathbb{N}$, $q \in \mathbb{N}$, $d \in (-1/2, 1/2)$, verify $\gamma(h) \sim h^{2d-1}$ when $h \to \infty$ and long memory occurs when $0 < d < 1/2$.

Let

- X_1 a FARIMA(p_1, d_1, q_1) with $0 \leq d_1 < 1/2$,
- X_2 a FARIMA(p_2, d_2, q_2) with $0 \leq d_2 < 1/2$.

We want to test the null hypothesis

$$\mathrm{H}_0 : d_1 = d_2$$

Let

- X_1 a FARIMA(p_1, d_1, q_1) with $0 \leq d_1 < 1/2$,
- X_2 a FARIMA(p_2, d_2, q_2) with $0 \leq d_2 < 1/2$.

We want to test the null hypothesis

$$\mathrm{H}_0 : d_1 = d_2$$

Our framework :

- X_1 and X_2 may not be independent.
- We do not restrict to FARIMA models (see the assumptions later)

2 Test Statistic

$$\mathrm{H}_0 : d_1 = d_2$$

First idea :

- estimate d_1 and d_2 by $\hat{d}_1$ and $\hat{d}_2$
  (different estimators are available : log-periodogram, Whittle, GPH, FEXP, etc.)
- evaluate $(\hat{d}_1 - \hat{d}_2)$ to conclude

$$\mathrm{H}_0 : d_1 = d_2$$

First idea :

- estimate d_1 and d_2 by $\hat{d}_1$ and $\hat{d}_2$ (different estimators are available : log-periodogram, Whittle, GPH, FEXP, etc.)
- evaluate $(\hat{d}_1 - \hat{d}_2)$ to conclude

Drawbacks :

- the joint probability law of $\hat{d}_1$ and $\hat{d}_2$ in the dependent case is not known.
- the behavior of $(\hat{d}_1 - \hat{d}_2)$ is strongly sensitive to the short-memory part of the induced processes X_1 and X_2 (e.g. the ARMA part of a FARIMA), leading to a bad size of the test.

Our approach : If X exhibits long memory,

$$S_n(\tau) = \sum_{t=1}^{[n\tau]} (X(t) - EX(t))$$

does not have a standard asymptotic behavior.

Our approach : If X exhibits long memory,

$$S_n(\tau) = \sum_{t=1}^{[n\tau]} (X(t) - EX(t))$$

does not have a standard asymptotic behavior.
For testing

$$H_0 : \ d = 0 \text{ (short memory) vs } H_1 : \ d \neq 0 \text{ (long memory)}$$

several procedures rely on the variations of S_n.

- R/S (Lo, 1991) : based on the range of S_n,
- KPSS (Kwiatkowski *et al.*, 1992) : based on $E(S_n^2)$,
- V/S (Giraitis *et al.*, 2003) : based on $Var(S_n)$.

Our approach : If X exhibits long memory,

$$S_n(\tau) = \sum_{t=1}^{[n\tau]} (X(t) - EX(t))$$

does not have a standard asymptotic behavior.
For testing

$$H_0 : \; d = 0 \text{ (short memory) vs } H_1 : \; d \neq 0 \text{ (long memory)}$$

several procedures rely on the variations of S_n.

- R/S (Lo, 1991) : based on the range of S_n,
- KPSS (Kwiatkowski *et al.*, 1992) : based on $E(S_n^2)$,
- V/S (Giraitis *et al.*, 2003) : based on $Var(S_n)$.

In the same spirit, for testing $\mathrm{H}_0 : d_1 = d_2$ our statistic is

$$T_{n,q} \quad = \quad \frac{V_1/S_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{V_1/S_{1,q}},$$

where $V_1/S_{1,q}$ is the standard V/S statistic for X_1,
$V_2/S_{2,q}$ is the standard V/S statistic for X_2.

More precisely

$$T_{n,q} = \frac{V_1/S_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{V_1/S_{1,q}}.$$

For i=1,2, $\overline{X}_i$ denotes the sample mean of X_i
$\hat{\gamma}_i(h)$ the empirical covariance function of X_i.

$$V_i = n^{-2}\sum_{k=1}^{n}\left(\sum_{t=1}^{k}(X_i(t)-\overline{X_i})\right)^2 - n^{-3}\left(\sum_{k=1}^{n}\sum_{t=1}^{k}(X_i(t)-\overline{X_i})\right)^2$$

(V_i is the empirical variance of the partial sums of X_i)

$$S_{i,q} = \sum_{h=-q}^{q}\left(1-\frac{|h|}{q+1}\right)\hat{\gamma}_i(h) = \frac{1}{q+1}\sum_{h,\ell=1}^{q+1}\hat{\gamma}_i(h-\ell).$$

($S_{i,q}$ estimates the variance of the limiting law of the partial sums of X_i)

The dependent case

Consider the cross-covariance estimator

$$S_{12,q} = \sum_{h=-q}^{q} \left(1 - \frac{|h|}{q+1}\right) \hat{\gamma}_{12}(h) = \frac{1}{q+1} \sum_{h,\ell=1}^{q+1} \hat{\gamma}_{12}(h-\ell)$$

where, $\hat{\gamma}_{12}(h) = n^{-1} \sum_{t=1}^{n-h} (X_1(t) - \overline{X}_1)(X_2(t+h) - \overline{X}_2)$, $h > 0$.

The dependent case

Consider the cross-covariance estimator

$$S_{12,q} = \sum_{h=-q}^{q} \left(1 - \frac{|h|}{q+1}\right) \hat{\gamma}_{12}(h) = \frac{1}{q+1} \sum_{h,\ell=1}^{q+1} \hat{\gamma}_{12}(h-\ell)$$

where, $\hat{\gamma}_{12}(h) = n^{-1} \sum_{t=1}^{n-h} (X_1(t) - \overline{X}_1)(X_2(t+h) - \overline{X}_2)$, $h > 0$.
When X_1 and X_2 are dependent, we introduce

$$\tilde{X}_1(t) = X_1(t) - (S_{12,q}/S_{2,q}) X_2(t), \qquad t = 1, \ldots, n.$$

so that the partial sums of $\tilde{X}_1$ and X_2 are uncorrelated.

The dependent case

Consider the cross-covariance estimator

$$S_{12,q} = \sum_{h=-q}^{q} \left(1 - \frac{|h|}{q+1}\right) \hat{\gamma}_{12}(h) = \frac{1}{q+1} \sum_{h,\ell=1}^{q+1} \hat{\gamma}_{12}(h-\ell)$$

where, $\hat{\gamma}_{12}(h) = n^{-1} \sum_{t=1}^{n-h} (X_1(t) - \overline{X}_1)(X_2(t+h) - \overline{X}_2)$, $h > 0$.
When X_1 and X_2 are dependent, we introduce

$$\tilde{X}_1(t) = X_1(t) - (S_{12,q}/S_{2,q}) X_2(t), \qquad t = 1, \ldots, n.$$

so that the partial sums of $\tilde{X}_1$ and X_2 are uncorrelated.
Then we consider

$$\tilde{T}_n = \frac{\tilde{V}_1/\tilde{S}_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{\tilde{V}_1/\tilde{S}_{1,q}},$$

where $\tilde{V}_1$ and $\tilde{S}_{1,q}$ are the same as before but with respect to $\tilde{X}_1$.

3 Consistency of the test

Assumptions

<u>Assumption $A(d_1, d_2)$</u> There exist $d_i \in [0, 1/2), i = 1, 2$ such that for any $i, j = 1, 2$ the following limits exist

$$1) \qquad c_{ij} = \lim_{n\to\infty} \frac{1}{n^{1+d_i+d_j}} \sum_{t,s=1}^{n} \gamma_{ij}(t-s).$$

Moreover, when $q \to \infty, n \to \infty, n/q \to \infty$,

$$2) \qquad \frac{\sum_{k,l=1}^{q} \hat{\gamma}_{ij}(k-l)}{\sum_{k,l=1}^{q} \gamma_{ij}(k-l)} \to_p 1$$

This assumption claims that

1) the second moment of the partial sums of X_i converge with the proper normalization,

2) the natural estimation of this second moment is consistent.

Assumptions

<u>ASSUMPTION B(d_1, d_2)</u> The partial sums of X_1 and X_2

$$\left(n^{-d_1-(1/2)} \sum_{t=1}^{[n\tau]} (X_1(t) - EX_1(t)), n^{-d_2-(1/2)} \sum_{t=1}^{[n\tau]} (X_2(t) - EX_2(t))\right)$$

converge (jointly) in finite dimensional distribution to

$$\left(\sqrt{c_{11}} B_{1,d_1}(\tau), \sqrt{c_{22}} B_{2,d_2}(\tau)\right),$$

where $(B_{1,d_1}(\tau), B_{2,d_2}(\tau))$ is a bivariate fractional Brownian motion with parameters d_1, d_2 and the correlation coefficient $\rho = c_{12}/\sqrt{c_{11}c_{22}}$.

<u>Some questions :</u>

- What is a bivariate fractional Brownian motion ? see later.
- Is this assumption restrictive (especially the joint convergence) ?
  $\rightarrow$ We will show that it holds for linear processes.

Recall that $\mathrm{H}_0 : d_1 = d_2$ and the test statistic is

$$\tilde{T}_n = \frac{\tilde{V}_1/\tilde{S}_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{\tilde{V}_1/\tilde{S}_{1,q}},$$

Recall that $\mathrm{H}_0 : d_1 = d_2$ and the test statistic is

$$\tilde{T}_n = \frac{\tilde{V}_1/\tilde{S}_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{\tilde{V}_1/\tilde{S}_{1,q}},$$

Proposition (Consistency of the test)

(i) Let Assumptions $A(d_1, d_2)$ and $B(d_1, d_2)$ be satisfied with some $d_1 = d_2 = d \in [0, 1/2)$. Then, as $n, q, n/q \to \infty$,

$$\tilde{T}_n \to_{\mathrm{law}} T = \frac{U_1}{U_2} + \frac{U_2}{U_1},$$

where

$$U_i = \int_0^1 (B^0_{i,d}(\tau))^2 \mathrm{d}\tau - \left(\int_0^1 B^0_{i,d}(\tau)\mathrm{d}\tau\right)^2 \qquad (i = 1, 2)$$

and where $B^0_{1,d}(\tau)$, $B^0_{2,d}(\tau)$ are mutually independent fractional bridges with the same parameter d.

Recall that $H_0 : d_1 = d_2$ and the test statistic is

$$\tilde{T}_n = \frac{\tilde{V}_1/\tilde{S}_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{\tilde{V}_1/\tilde{S}_{1,q}},$$

Proposition (Consistency of the test)

(i) Let Assumptions $A(d_1, d_2)$ and $B(d_1, d_2)$ be satisfied with some $d_1 = d_2 = d \in [0, 1/2)$. Then, as $n, q, n/q \to \infty$,

$$\tilde{T}_n \to_{\text{law}} T = \frac{U_1}{U_2} + \frac{U_2}{U_1},$$

where

$$U_i = \int_0^1 (B^0_{i,d}(\tau))^2 \mathrm{d}\tau - \left(\int_0^1 B^0_{i,d}(\tau)\mathrm{d}\tau\right)^2 \qquad (i = 1, 2)$$

and where $B^0_{1,d}(\tau)$, $B^0_{2,d}(\tau)$ are mutually independent fractional bridges with the same parameter d.

(ii) Let Assumptions $A(d_1, d_2)$ and $B(d_1, d_2)$ be satisfied with $d_1 \neq d_2$ $(d_1, d_2 \in [0, 1/2))$. Then, as $n, q, n/q \to \infty$,

$$|\tilde{T}_n| \to_p \infty.$$

4 The bivariate fractional Brownian motion

Definition

A bi-fBm $\left(B_{1,d_1}(s), B_{2,d_2}(s)\right), s \in \mathbb{R}$ with parameters $d_i \in (-1/2, 1/2)$, $i = 1, 2$, is a Gaussian process with (for $s_1, s_2 > 0$ and $d_1 + d_2 \neq 0$)

$$EB_{1,d_1}(s) = EB_{2,d_2}(s) = 0$$

$$EB_{1,d_1}(s_1)B_{1,d_1}(s_2) = (1/2)(|s_1|^{2d_1+1} + |s_2|^{2d_1+1} - |s_1 - s_2|^{2d_1+1}),$$

$$EB_{2,d_2}(s_1)B_{2,d_2}(s_2) = (1/2)(|s_1|^{2d_2+1} + |s_2|^{2d_2+1} - |s_1 - s_2|^{2d_2+1}),$$

Definition

A bi-fBm $\big(B_{1,d_1}(s), B_{2,d_2}(s)\big), s \in \mathbb{R}$ with parameters $d_i \in (-1/2, 1/2)$, $i = 1, 2$, is a Gaussian process with (for $s_1, s_2 > 0$ and $d_1 + d_2 \neq 0$)

$$EB_{1,d_1}(s) = EB_{2,d_2}(s) = 0$$

$$EB_{1,d_1}(s_1)B_{1,d_1}(s_2) = (1/2)(|s_1|^{2d_1+1} + |s_2|^{2d_1+1} - |s_1 - s_2|^{2d_1+1}),$$

$$EB_{2,d_2}(s_1)B_{2,d_2}(s_2) = (1/2)(|s_1|^{2d_2+1} + |s_2|^{2d_2+1} - |s_1 - s_2|^{2d_2+1}),$$

$$EB_{1,d_1}(s_1)B_{2,d_2}(s_2) = \begin{cases} c_1|s_1|^{d_1+d_2+1} + c_2|s_2|^{d_1+d_2+1} - c_1|s_1 - s_2|^{d_1+d_2+1}, & \text{if } s_1 \geq s_2, \\ c_1|s_1|^{d_1+d_2+1} + c_2|s_2|^{d_1+d_2+1} - c_2|s_1 - s_2|^{d_1+d_2+1}, & \text{if } s_1 \leq s_2, \end{cases}$$

where c_1, c_2 are some constants yielding the positive definiteness.

Definition

A bi-fBm $(B_{1,d_1}(s), B_{2,d_2}(s)), s \in \mathbb{R}$ with parameters $d_i \in (-1/2, 1/2)$, $i = 1, 2$, is a Gaussian process with (for $s_1, s_2 > 0$ and $d_1 + d_2 \neq 0$)

$$EB_{1,d_1}(s) = EB_{2,d_2}(s) = 0$$

$$EB_{1,d_1}(s_1)B_{1,d_1}(s_2) = (1/2)(|s_1|^{2d_1+1} + |s_2|^{2d_1+1} - |s_1 - s_2|^{2d_1+1}),$$

$$EB_{2,d_2}(s_1)B_{2,d_2}(s_2) = (1/2)(|s_1|^{2d_2+1} + |s_2|^{2d_2+1} - |s_1 - s_2|^{2d_2+1}),$$

$$EB_{1,d_1}(s_1)B_{2,d_2}(s_2) = \begin{cases} c_1|s_1|^{d_1+d_2+1} + c_2|s_2|^{d_1+d_2+1} - c_1|s_1 - s_2|^{d_1+d_2+1}, & \text{if } s_1 \geq s_2, \\ c_1|s_1|^{d_1+d_2+1} + c_2|s_2|^{d_1+d_2+1} - c_2|s_1 - s_2|^{d_1+d_2+1}, & \text{if } s_1 \leq s_2, \end{cases}$$

where c_1, c_2 are some constants yielding the positive definiteness.

- the definition extends to all $(s_1, s_2) \in \mathbb{R}^2$ and to $d_1 + d_2 = 0$,
- the case $d_1 + d_2 = 0$ involves logarithm functions,
- the extension from a bi-fBm to a p-fBm is straightforward.
- What we proved about the domain of definition of c_1, c_2 :

  $|c_1 + c_2| \leq 1$ is necessary but clearly too large ;

  It suffices that (c_1, c_2) belongs to an elliptic domain centered at 0.

Characterization

Theorem

Let $X(t) = (X_1(t), X_2(t))_{t\geq 0}$ *be a centered, 2nd order process, null at* 0*. Assume that*

- X *has stationary increments : For any* $t, h_1, h_2 \geq 0$

$$(X_1(h_1+t) - X_1(t), X_2(h_2+t) - X_2(t)) =_{\text{fdd}} (X_1(h_1), X_2(h_2))$$

- X *is scale invariant : For any* $t, \lambda > 0$,

$$(X_1(\lambda t), X_2(\lambda t)) =_{\text{fdd}} \left(\lambda^{d_1+1/2} X_1(t), \lambda^{d_2+1/2} X_2(t)\right),$$

Moreover, assume that $t \mapsto EX_1(t)X_2(1)$ *and* $t \mapsto EX_2(t)X_1(1)$ *are continuously differentiable on* $(0,1) \cup (1,\infty)$.

Then X *has the same covariance structure as the bi-fBm defined above.*

Therefore, the bi-fBm is the unique Gaussian process satisfying a stationary increments and scale invariant property.

5 The case of bivariate linear models

Definition of a bivariate linear process

We consider bivariate linear models $(X_1(t), X_2(t)),\ t \in \mathbb{Z}$ as given by

$$X_1(t) = \sum_{k=0}^{\infty} a_{11}(k)\xi_1(t-k) + \sum_{k=0}^{\infty} a_{12}(k)\xi_2(t-k),$$
$$X_2(t) = \sum_{k=0}^{\infty} a_{21}(k)\xi_1(t-k) + \sum_{k=0}^{\infty} a_{22}(k)\xi_2(t-k),$$

where $a_{ij}(k)$ are real coefficients with $\sum_{k=0}^{\infty} a_{ij}^2(k) < \infty$ and $(\xi_1(t), \xi_2(t))$, $t \in \mathbb{Z}$ is a bivariate (weak) white noise :

- For $i = 1, 2 \quad E\xi_i(t) = 0$ and $E(\xi_i(t)^2) = 1$,
- $E(\xi_1(t), \xi_2(t)) = \rho \in (-1, 1)$ and $E(\xi_1(t), \xi_2(s)) = 0$ if $s \neq t$.

We assume that $(\xi_1(t), \xi_2(t))$, $t \in \mathbb{Z}$ is a sequence of i.i.d. random vectors and that there exists $d_{ij} \in [0, 1/2)$ such that :

$$\sum_{k=0}^{\infty} |a_{ij}(k)| < \infty, \qquad \text{if} \quad d_{ij} = 0,$$

$$a_{ij}(k) = (\alpha_{ij} + o(1))\, |k|^{d_{ij}-1} \quad (k \to \infty) \qquad \text{if} \quad d_{ij} \in (0, 1/2),$$

where $\alpha_{ij} \neq 0$ are some numbers, $i, j = 1, 2$.

Proposition

If there exists $p > 1$ such that $E|\xi_i(t)|^{2p} < \infty$ $(i = 1, 2)$, then $(X_1(t), X_2(t))$ satisfies Assumptions $A(d_1, d_2)$ and $B(d_1, d_2)$, with

$$d_i = \max\{d_{i1}, d_{i2}\} \in [0, 1/2) \qquad (i = 1, 2).$$

$\Longrightarrow$ For such bivariate linear processes, our test is consistent.

Some examples

Example

Let $c_{ij} \in \mathbb{R}$ $(i = 1, 2)$ be some constants, and let

$$X_i(t) = (1 - L)^{-d_i}(c_{i1}\xi_1(t) + c_{i2}\xi_2(t)) \qquad (i = 1, 2)$$

be FARIMA$(0, d_i, 0)$ processes with $d_1, d_2 \in (0, 1/2)$ may be different.

Example

$$\begin{aligned}(1 - L)^{d'_{11}} X_1(t) + \beta(1 - L)^{d'_{12}} X_2(t) &= \xi_1(t),\\ (1 - L)^{d'_{22}} X_2(t) &= \xi_2(t),\end{aligned}$$

where $d'_{ij} \in [0, 1/2), \beta \in \mathbb{R}$ are parameters, $d'_{22} + d'_{12} - d'_{12} < 1/2$.
A stationary solution of the above equation is given by

$$\begin{aligned}X_2(t) &= (1 - L)^{-d'_{22}}\xi_2(t),\\ X_1(t) &= (1 - L)^{-d'_{11}}\xi_1(t) - \beta(1 - L)^{d'_{12}-d'_{11}-d'_{22}}\xi_2(t).\end{aligned}$$

6 Practical implementation of the test

Recall that we want to test $\mathrm{H}_0 : d_1 = d_2$ with the test statistic

$$\tilde{T}_n \quad = \quad \frac{\tilde{V}_1/\tilde{S}_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{\tilde{V}_1/\tilde{S}_{1,q}}.$$

Under H_0, $\tilde{T}_n \quad \to_{\text{law}} \; U_d$ which depends on $d = d_1 = d_2$.

Recall that we want to test $\mathrm{H}_0 : d_1 = d_2$ with the test statistic

$$\tilde{T}_n = \frac{\tilde{V}_1/\tilde{S}_{1,q}}{V_2/S_{2,q}} + \frac{V_2/S_{2,q}}{\tilde{V}_1/\tilde{S}_{1,q}}.$$

Under H_0, $\tilde{T}_n \rightarrow_{\text{law}} U_d$ which depends on $d = d_1 = d_2$.

For a practical implementation, given a sample and a signifiance level $\alpha \in (0, 1)$, we must :

- first choose the parameter q
- compute $\tilde{T}_n$
- estimate d by a consistent estimator $\hat{d}$
- test whether $\tilde{T}_n > c_\alpha(\hat{d})$ (the critical region),

where $c_\alpha(d)$ is the upper quantile of order α of U_d.

Choice of q

The choice of q is crucial.
It appears mainly in the asymptotic behaviour of S in V/S.
From the theory, we must have $q, n/q \to \infty$ when $n \to \infty$.
In Giraitis et al (2006), $q = [n^{1/3}]$ is suggested.

Choice of q

The choice of q is crucial.
It appears mainly in the asymptotic behaviour of S in V/S.
From the theory, we must have $q, n/q \to \infty$ when $n \to \infty$.
In Giraitis et al (2006), $q = [n^{1/3}]$ is suggested.

But simulations show that

- n being fixed, d has a strong effect on the optimal choice of q,
- the short memory part is important (e.g. the ARMA part of a FARIMA).

Choice of q

The choice of q is crucial.
It appears mainly in the asymptotic behaviour of S in V/S.
From the theory, we must have $q, n/q \to \infty$ when $n \to \infty$.
In Giraitis et al (2006), $q = [n^{1/3}]$ is suggested.

But simulations show that

- n being fixed, d has a strong effect on the optimal choice of q,
- the short memory part is important (e.g. the ARMA part of a FARIMA).

We optimize q under H_0 to guarantee a correct size of the test.
We focus on the ratio $S_{1,q}/S_{2,q}$ that appears in $\tilde{T}_n$.
Starting from a result of Disasto et al (2008), we obtain the linear expansion of

$$E\left(\frac{S_{1,q}}{S_{2,q}} * \frac{c_{22}}{c_{11}} - 1\right)^2.$$

We choose q which minimizes the first term in this expansion.

Choice of q

This scheme leads to the choice

$$q = \begin{cases} 0.3\sqrt{|\hat{I}_1 - \hat{I}_2|}\, n^{1/(3+4\hat{d})}, & if\ \hat{d} < 1/4, \\ 0.3\sqrt{|\hat{I}_1 - \hat{I}_2|}\, n^{1/2-\hat{d}}, & if\ \hat{d} > 1/4. \end{cases}$$

where $\hat{d} = (\hat{d}_1 + \hat{d}_2)/2$ is the adaptive FEXP estimator (see Louditsky et al, 2001) and

$$\hat{I}_i = \int_0^\pi x^{-2\hat{d}} \sin^{-2}(x/2) \frac{\hat{g}_i(x)}{\hat{g}_i(0)} dx,$$

where $\hat{g}_i$ estimates the short memory part of the spectral density of X_i.

Choice of q

This scheme leads to the choice

$$q = \begin{cases} 0.3\sqrt{|\hat{I}_1 - \hat{I}_2|}\, n^{1/(3+4\hat{d})}, & if\ \hat{d} < 1/4, \\ 0.3\sqrt{|\hat{I}_1 - \hat{I}_2|}\, n^{1/2-\hat{d}}, & if\ \hat{d} > 1/4. \end{cases}$$

where $\hat{d} = (\hat{d}_1 + \hat{d}_2)/2$ is the adaptive FEXP estimator (see Louditsky et al, 2001) and

$$\hat{I}_i = \int_0^\pi x^{-2\hat{d}} \sin^{-2}(x/2) \frac{\hat{g}_i(x)}{\hat{g}_i(0)} dx,$$

where $\hat{g}_i$ estimates the short memory part of the spectral density of X_i.

For $\hat{g}_i$, we choose the spectral density of the best AR process approaching this short memory part. We proceed in a two steps procedure :

- we first estimate d by the adaptative FEXP estimator
- then we fit an AR process to $(1-L)^{\hat{d}} X_i$ by BIC criterion.

Let us sum-up the testing procedure.

Let two series $X_1(t)$, $X_2(t)$ and a signifiance level $\alpha \in (0,1)$

- First estimate d_1 and d_2 with the adaptative FEXP estimator
- Estimate g_1 and g_2 which approximate the short-memory part in the spectral density of X_1 and X_2
- This leads to the choice of q
- Compute $\tilde{T}_n$ and compare to $c_\alpha((\hat{d}_1 + \hat{d}_2)/2)$

7 Some simulations

Simulations

We compute the test with independent X_1 and X_2 where

$$X_1 \sim FAR(1, d_1, 0)$$
$$X_2 \sim FAR(1, d_2, 0)$$

i.e. $(1 - a_i L)(1 - L)^{d_i} X_i(n) = \epsilon_i(n)$, where ϵ_i is a white noise.

Several values of a_i and d_i are tested :

$$a_i \in \{-0.4, 0, 0.4\} \text{ and } d_i \in \{0, 0.1, 0.2, 0.3, 0.4\}.$$

The probability of rejection is evaluated on 1000 replications of the test where the signifiance level is fixed at 5%.
The sample size of X_1 and X_2 is 4096.

For fixed a_1, a_2 , each cell contains the **probability of rejection of** H_0 for different parameters (d_1, d_2) with $d_i \in \{0, 0.1, 0.2, 0.3, 0.4\}$ and $d_1 \leq d_2$

	$a_1 = -0.4$					$a_1 = 0$					$a_1 = 0.4$				
	.057														
	.192	.050													
$a_2 = -0.4$	.483	.148	.056												
	.774	.387	.113	.057											
	.911	.678	.356	.095	.029										
	.047					.051									
	.118	.061				.233	.041								
$a_2 = 0$	.354	.092	.046			.589	.204	.048							
	.620	.290	.083	.041		.857	.488	.144	.043						
	.811	.568	.261	.078	.033	.958	.766	.422	.112	.029					
	.057					.052					.035				
	.101	.035				.108	.042				.201	.057			
$a_2 = 0.4$	.293	.083	.046			.355	.109	.048			.573	.192	.042		
	.575	.246	.073	.043		.697	.342	.108	.052		.840	.536	.165	.040	
	.792	.475	.231	.061	.033	.882	.641	.302	.092	.033	.951	.778	.478	.143	.030

Mean-Value of q chosen for the above simulations

	a=-0.4					a=0					a=0.4				
a=-0.4	4.3														
	3.7	3.3													
	3.2	2.8	2.7												
	2.8	2.6	2.2	1.6											
	2.7	2.2	1.7	1.0	0.5										
a=0	10.9					3.2									
	9.1	7.9				2.7	2.1								
	7.9	6.9	6.2			2.3	2.0	1.7							
	6.9	6.3	5.2	3.9		1.9	1.8	1.4	1.0						
	6.2	5.3	3.9	2.5	1.5	1.8	1.5	1.0	0.5	0.3					
a=0.4	15.8					11.2					5.4				
	13.1	11.2				9.0	7.5				4.4	3.7			
	11.1	9.5	8.6			7.5	6.3	5.3			3.6	3.0	2.7		
	9.6	8.5	7.0	5.1		6.2	5.3	4.3	2.9		3.1	2.6	2.0	1.4	
	8.4	7.1	5.0	3.3	2.0	5.3	4.4	2.9	1.8	1.0	2.6	2.0	1.4	0.8	0.4

Simulations on dependent samples

We evaluate the test with

$$X_1(n) = (1-p)Y_1(n) + pY_2(n)$$
$$X_2(n) = (1-p)Y_2(n) + pY_1(n)$$

where Y_i are independent $\mathrm{F}(d_i)$ with $d_i \in \{0, 0.1, 0.2, 0.3, 0.4\}$ and $p \in [0, 1/2)$.

p=0.05	.055				
	.242	.062			
	.639	.207	.057		
	.866	.571	.159	.046	
	.966	.837	.464	.104	.045
p=0.25	.053				
	.247	.061			
	.727	.214	.049		
	.945	.629	.185	.052	
	.993	.894	.493	.138	.043
p=0.45	.055				
	.836	.046			
	.983	.629	.059		
	.996	.957	.330	.031	
	1.000	.993	.850	.138	.044

- DISTASO, W., GIRAITIS, L. AND ABADIR, K. (2008) Two estimators of long-run variance : beyond short memory. Preprint.
- GIRAITIS, L., KOKOSZKA, P., LEIPUS, R. AND TEYSSIÈRE, G. (2003) Rescaled variance and related tests for long memory in volatility and levels. *J. Econometrics* **112**, 265–294.
- GIRAITIS, L., LEIPUS, R. AND PHILIPPE, A. (2006) A test for stationarity versus trends and unit roots for a wide class of dependent errors. *Econometric Th.* **22**, 989–1029.
- LOUDITSKY, A., MOULINES, E. AND SOULIER, PH. (2001) Adaptive estimation of the fractional differencing coefficient *Bernoulli* **7**,(5), 699-731.
- KWIATKOWSKI, D., PHILLIPS, P.C.B., SCHMIDT, P. AND SHIN, Y. (1992) Testing the null hypothesis of stationarity against the alternative of a unit root : how sure are we that economic time series have a unit root ? *J. Econometrics* **54**, 159–178.
- LO, A.W. (1991) Long-term memory in stock market prices. *Econometrica* **59**, 1279–1313.