A two-sample test for comparison of long memory parameters

F. Lavancier\textsuperscript{1}, A. Philippe\textsuperscript{1}, D. Surgailis\textsuperscript{2}

\textsuperscript{1}Laboratoire Jean Leray, Université de Nantes

\textsuperscript{2}Institute of Mathematics and Informatics, Vilnius

10th International Vilnius Conference on Probability and Mathematical Statistics
1 Introduction
The statistical problem

- \((X_1(t))_{t \in \mathbb{Z}}\) is a stationary process with long memory parameter \(d_1 \geq 0\)
- \((X_2(t))_{t \in \mathbb{Z}}\) is a stationary process with long memory parameter \(d_2 \geq 0\)
- \(X_1\) and \(X_2\) may be dependent.

We want to test the null hypothesis

\[ H_0 : d_1 = d_2 \]
2 The Testing procedure and its consistency
First idea : Testing from the estimation of $(d_1, d_2)$

$H_0 : d_1 = d_2$

Procedure

- 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
First idea : Testing from the estimation of \((d_1, d_2)\)

\[ H_0 : d_1 = d_2 \]

Procedure

- 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

Drawback

1. The joint probability law of \(\hat{d}_1\) and \(\hat{d}_2\) in the dependent case is not known.

2. The behavior of \(|\hat{d}_1 - \hat{d}_2|\) is strongly sensitive to the short-memory part of processes \(X_1\) and \(X_2\) (e.g. the ARMA part of a FARIMA), leading to a bad size of the test.
Our testing procedure

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 testing procedure

If $X$ exhibits long memory,

$$S_n(\tau) = \sum_{t=1}^{\lceil nt \rceil} (X(t) - EX(t))$$

does not have a standard asymptotic behavior.

Example for $F(d)$

$$\frac{1}{n^{d+1/2}} S_n(\tau) \xrightarrow{D[0,1]} W_d(\tau)$$
Our testing procedure

If $X$ exhibits long memory,

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

does not have a standard asymptotic behavior.

Example for $F(d)$

$$\frac{1}{n^{d+1/2}} S_n(\tau) \xrightarrow{D[0,1]} W_d(\tau)$$

In particular, typically,

$$Var(S_n(\tau)) = O(n^{2d+1} L(n)),$$

where $L$ is a slowly varying function.

⇒ The idea is to base the test on the variations of $S_n$. 
Univariate time series

For univariate time series, the variations of $S_n$ are used to test:

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

The most standard statistics

- 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 $Var(S_n)$. 
Univariate time series

For univariate time series, the variations of $S_n$ are used to test:

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

The most standard statistics

- 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 $\text{Var}(S_n)$.

In the same spirit, for testing $H_0 : d_1 = d_2$ our statistic is

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

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

\[ T_{n,q} = \frac{V_1}{S_1,q} + \frac{V_2}{V_2/S_2,q} + \frac{V_2}{V_1/S_1,q}. \]

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
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)

Basically, when \( n, q, n/q \to \infty \),

\[ \left( \frac{n}{q} \right)^{-2d_i} V_i/S_{i,q} \longrightarrow U(d_i) \]
Assumptions

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

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

Moreover, when \( q \to \infty, n \to \infty, n/q \to \infty, \)

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

**Remark.**

This assumption claims that

1) the second moment of the partial sums of \((X_1, X_2)\) converge with the proper normalization,

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

**ASSUMPTION B(d₁, d₂)** The partial sums of X₁ and X₂

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

converge (jointly) in finite dimensional distribution to

\[
(\sqrt{c_{11}}B_{1,d_1}(\tau), \sqrt{c_{22}}B_{2,d_2}(\tau)),
\]

where \((B_{1,d_1}(\tau), B_{2,d_2}(\tau))\) is a nonanticipative bivariate fractional Brownian motion with parameters \(d_1, d_2\) and the correlation coefficient

\[
\rho = \text{corr}(B_1(1), B_2(1)) = \frac{c_{12}}{\sqrt{c_{11}c_{22}}}
\]

**Remark.**

This is fulfilled for bivariate (super)linear processes under mild assumptions.
Asymptotic behavior of $T_n$

**Proposition**

Let Assumptions $A(d_1, d_2)$ and $B(d_1, d_2)$ be satisfied with some $d_1, d_2 \in [0, 1/2)$ and $\rho \in [-1, 1]$, and let $n, q, n/q \to \infty$.

(i) If $d_1 = d_2 = d$ then

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

where

$$U_i = \int_0^1 (B^0_i(\tau))^2 d\tau - \left(\int_0^1 B^0_i(\tau) d\tau\right)^2, \quad i = 1, 2,$$

where $(B^0_i(\tau) = B_i(\tau) - \tau B_i(1), \tau \in [0, 1])$, $i = 1, 2$ are fractional Brownian bridges obtained from bivariate fBm $((B_1(\tau), B_2(\tau)), \tau \in \mathbb{R})$ with the same memory parameters $d_1 = d_2 = d$ and correlation coefficient $\rho$.

(ii) If $d_1 \neq d_2$ then

$$T_n \to_p \infty.$$
The dependent case

When \( X_1 \) and \( X_2 \) are dependent, we introduce

\[
\tilde{X}_1(t) = X_1(t) - \frac{S_{12,q}}{S_{2,q}} X_2(t), \quad t = 1, \ldots, n.
\]

where

\[
S_{12,q} = \frac{1}{q+1} \sum_{h,\ell=1}^{q+1} \hat{\gamma}_{12}(h-\ell)
\]

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

\[\implies\] The partial sums of \( \tilde{X}_1 \) and \( X_2 \) are uncorrelated.

Then we consider

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

where \( \tilde{V}_1 \) and \( \tilde{S}_{1,q} \) are the same as before but with respect to \( \tilde{X}_1 \).
Proposition (Consistency of the test)

Let Assumptions $A(d_1, d_2)$ and $B(d_1, d_2)$ be satisfied with some $d_1, d_2 \in [0, 1/2)$ and $\rho \in [-1, 1]$, and let $n, q, n/q \to \infty$.

(i) If $d_1 = d_2 = d \in [0, 1/2)$, then

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

where

$$U_i = \int_0^1 (B_{i,d}^0(\tau))^2 d\tau - \left( \int_0^1 B_{i,d}^0(\tau) d\tau \right)^2 \quad (i = 1, 2)$$

and where $B_{1,d}^0(\tau)$, $B_{2,d}^0(\tau)$ are mutually independent fractional bridges with the same parameter $d$. 
Proposition (Consistency of the test)

Let Assumptions $A(d_1, d_2)$ and $B(d_1, d_2)$ be satisfied with some $d_1, d_2 \in [0, 1/2)$ and $\rho \in [-1, 1]$, and let $n, q, n/q \to \infty$.

(i) If $d_1 = d_2 = d \in [0, 1/2)$, then

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

where

$$U_i = \int_0^1 (B_{i,d}^0(\tau))^2 d\tau - \left( \int_0^1 B_{i,d}^0(\tau) d\tau \right)^2 \quad (i = 1, 2)$$

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

(ii) If $d_1 > d_2$, then

$$\tilde{T}_n \to_{p} \infty.$$
Proposition (Consistency of the test)

Let Assumptions \( A(d_1, d_2) \) and \( B(d_1, d_2) \) be satisfied with some \( d_1, d_2 \in [0, 1/2) \) and \( \rho \in [-1, 1] \), and let \( n, q, n/q \to \infty \).

(i) If \( d_1 = d_2 = d \in [0, 1/2) \), then

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

where

\[
U_i = \int_0^1 (B_{i,d}^0(\tau))^2 d\tau - \left( \int_0^1 B_{i,d}^0(\tau)d\tau \right)^2 \quad (i = 1, 2)
\]

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

(ii) If \( d_1 > d_2 \), then

\[
\tilde{T}_n \to_p \infty.
\]

Remark.

When \( d_1 < d_2 \), \( \tilde{T}_n \to_p \rho^2 + \rho^{-2} \).
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Test Statistic</th>
<th>In practice</th>
<th>Simulations</th>
</tr>
</thead>
</table>

3. Practical implementation of the test
We want to test $d_1 = d_2$ vs $d_1 > d_2$ 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 $H_0$, $\tilde{T}_n \rightarrow_{law} U_d$ which depends on $d = d_1 = d_2$. 
We want to test $d_1 = d_2$ vs $d_1 > d_2$ 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 significance level $\alpha \in (0, 1)$, we must:

- first choose the parameter $q$
- compute $\tilde{T}_n$
- estimate $d$ by a consistent estimator $\hat{d} = (\hat{d}_1 + \hat{d}_2)/2$
- test whether $\tilde{T}_n > c_\alpha(\hat{d})$ (the critical region),

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

The choice of $q$ is crucial.
From the theory, we must have $q, n/q \to \infty$ when $n \to \infty$. 
Choice of $q$

The choice of $q$ is crucial.
From the theory, we must have $q, n/q \to \infty$ when $n \to \infty$.

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. From the theory, we must have $q, n/q \to \infty$ when $n \to \infty$.

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 $q$ to guarantee a correct level of the test. We focus on the ratio $S_{1,q}/S_{2,q}$ that appears in $\tilde{T}_n$. We obtain the linear expansion of

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

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

This scheme leads to the choice

$$
\hat{q} = 0.3 \left| \hat{I} \right|^{1/2} \begin{cases} 
  n^{1/(3+4\hat{d})} & \text{if } \hat{d} \leq 1/4, \\
  n^{1/2-\hat{d}} & \text{if } \hat{d} \geq 1/4,
\end{cases}
$$

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

$$
\hat{I} = \int_{0}^{\pi} \left( \frac{\hat{g}_1(x)}{\hat{g}_1(0)} - \frac{\hat{g}_2(x)}{\hat{g}_2(0)} \right) \frac{dx}{x^{2\hat{d}} \sin^2(x/2)},
$$

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 adaptive FEXP estimator then we fit an AR process to $(1-L)^{\hat{d}}X_i$ by BIC criterion.
Choice of \( q \)

This scheme leads to the choice

\[
\hat{q} = 0.3 \, |\hat{I}|^{1/2} \begin{cases} 
\frac{n^{1/(3+4\hat{d})}}{\hat{n}^{1/2-\hat{d}}} & \text{if } \hat{d} \leq 1/4, \\
\frac{n^{1/(3+4\hat{d})}}{\hat{n}^{1/2-\hat{d}}} & \text{if } \hat{d} \geq 1/4,
\end{cases}
\]

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

\[
\hat{I} = \int_0^{\pi} \left( \frac{\hat{g}_1(x)}{\hat{g}_1(0)} - \frac{\hat{g}_2(x)}{\hat{g}_2(0)} \right) \frac{dx}{x^{2\hat{d}} \sin^2(x/2)},
\]

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.
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 $\epsilon_i$ 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 significance 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 $d_1 \leq d_2$

<table>
<thead>
<tr>
<th></th>
<th>$a_i = -0.4$</th>
<th>$a_i = 0$</th>
<th>$a_i = 0.4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_2 = -0.4$</td>
<td>.057</td>
<td>.192 .050</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.483 .148 .056</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.774 .387 .113 .057</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.911 .678 .356 .095 .029</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_2 = 0$</td>
<td>.047</td>
<td>.118 .061</td>
<td>.589 .204 .048</td>
</tr>
<tr>
<td></td>
<td>.354 .092 .046</td>
<td>.233 .041</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.620 .290 .083 .041</td>
<td>.857 .488 .144 .043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.811 .568 .261 .078 .033</td>
<td>.958 .766 .422 .112 .029</td>
<td></td>
</tr>
<tr>
<td>$a_2 = 0.4$</td>
<td>.057</td>
<td>.101 .035</td>
<td>.355 .109 .048</td>
</tr>
<tr>
<td></td>
<td>.101 .035</td>
<td>.108 .042</td>
<td>.573 .192 .042</td>
</tr>
<tr>
<td></td>
<td>.293 .083 .046</td>
<td>.355 .108 .052</td>
<td>.840 .536 .165 .040</td>
</tr>
<tr>
<td></td>
<td>.575 .246 .073 .043</td>
<td>.697 .342 .108 .052</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.792 .475 .231 .061 .033</td>
<td>.882 .641 .302 .092 .033</td>
<td>.951 .778 .478 .143 .030</td>
</tr>
</tbody>
</table>
Mean-Value of $q$ chosen for the above simulations

<table>
<thead>
<tr>
<th></th>
<th>a=-0.4</th>
<th>a=0</th>
<th>a=0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>a=-0.4</td>
<td>4.3</td>
<td>10.9</td>
<td>15.8</td>
</tr>
<tr>
<td></td>
<td>3.7 3.3</td>
<td>9.1 7.9</td>
<td>13.1 11.2</td>
</tr>
<tr>
<td></td>
<td>3.2 2.8 2.7</td>
<td>7.9 6.9 6.2</td>
<td>11.1 9.5 8.6</td>
</tr>
<tr>
<td></td>
<td>2.8 2.6 2.2 1.6</td>
<td>6.9 6.3 5.2 3.9</td>
<td>9.6 8.5 7.0 5.1</td>
</tr>
<tr>
<td></td>
<td>2.7 2.2 1.7 1.0 0.5</td>
<td>6.2 5.3 3.9 2.5 1.5</td>
<td>8.4 7.1 5.0 3.3 2.0</td>
</tr>
<tr>
<td>a=0</td>
<td>3.2</td>
<td>2.7 2.1</td>
<td>4.4 3.7</td>
</tr>
<tr>
<td></td>
<td>2.3 2.0 1.7</td>
<td>3.6 3.0 2.7</td>
<td>3.1 2.6 2.0 1.4</td>
</tr>
<tr>
<td>a=0.4</td>
<td>1.9 1.8 1.4 1.0</td>
<td>3.1 2.6 2.0 1.4</td>
<td>2.6 2.0 1.4 0.8 0.4</td>
</tr>
</tbody>
</table>
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 \( F(d_i) \) with \( d_i \in \{0, 0.1, 0.2, 0.3, 0.4\} \) and \( p \in [0, 1/2) \).


