

# Data-driven semi-parametric detection of multiple changes in long-range dependent processes

Jean-Marc Bardet and Abdellatif Guenaizi

*SAMM, Université Panthéon-Sorbonne (Paris 1), 90 rue de Tolbiac, 75013 Paris, France*  
e-mail: [bardet@univ-paris1.fr](mailto:bardet@univ-paris1.fr); [guenaizi-abd@live.fr](mailto:guenaizi-abd@live.fr)

**Abstract:** This paper is devoted to the offline multiple changes detection for long-range dependent processes. The observations are supposed to satisfy a semi-parametric long-range dependent assumption with distinct memory parameters on each stage. A penalized local Whittle contrast is considered for estimating all the parameters, notably the number of changes. Consistency as well as convergence rates are obtained. Monte-Carlo experiments exhibit the accuracy of the estimators. They also show that the estimation of the number of breaks is improved by using a data-driven slope heuristic procedure of choice of the penalization parameter.

**MSC2020 subject classifications:** Primary 62G05, 62G20; secondary 62M05.

Received October 2019.

## 1. Introduction

There exists now a very large literature devoted to long-range dependent processes. Their most commonly used definition requires a second order stationary process  $X = (X_n)_{n \in \mathbb{Z}}$  with spectral density  $f$  such that:

$$f(\lambda) = |\lambda|^{-2d} L(|\lambda|) \quad \text{for any } \lambda \in [-\pi, \pi], \quad (1.1)$$

where  $0 < d < 1/2$  and  $L$  is a positive slowly varying function, *i.e.* satisfying  $\lim_{h \rightarrow 0} \frac{L(c|h|)}{L(|h|)} = 1$  for any  $c > 0$ . Typically  $L$  is a function with a positive limit or a logarithm.

From an observed trajectory  $(X_1, \dots, X_n)$  of a long-range dependent process, the estimation of the parameter  $d$  is an interesting statistical question. The case of a parametric estimator for which the explicit expression of the spectral density  $f$  is known was successively solved in many cases using maximum likelihood estimators (see for instance Dahlhaus, 1989) or Whittle estimators (see for instance Fox and Taqqu, 1987, Giraitis and Surgailis, 1990, or Giraitis and Taqqu, 1999).

However, with numerical applications in view, knowing the explicit form of the spectral density is not a realistic framework. A semi-parametric estimation of  $d$  where only the behaviour (1.1) is assumed should be preferred. Thus, numerous semi-parametric estimators of  $d$  were defined and studied, the main ones being

the log-periodogram (see [Geweke and Porter-Hudak, 1983](#), or [Robinson, 1995a](#)), the wavelet based (see [Bardet et al., 2000](#)) and the local Whittle estimators (see [Robinson, 1995b](#)).

This last one is a version of the Whittle estimator for which only asymptotically small frequencies are considered. It provides certainly the best trade-off between computation time and accuracy of the estimation (see for instance [Bardet et al., 2003b](#)). Its asymptotic normality was extended for numerous kinds of long-memory processes (see [Dalla et al., 2006](#)) and also non-stationary processes (see [Abadir et al., 2007](#)). However, there is still no satisfactory adaptive method of choice of the bandwidth parameter even if several interesting attempts have been developed (see for instance [Henry and Robinson, 1996](#), or [Henry, 2007](#)). Hence, the usual choice valid for FARIMA or Fractional Gaussian noise is commonly chosen.

In this paper, we consider the classical framework of offline multiple change detection. Let  $(X_1, \dots, X_n)$  be an observed trajectory of a process  $X$  that is partitioned into  $K^* + 1$  sub-trajectories on which it is a linear long-range dependent process whose long-memory parameters are distinct from one stage to another (see a more formal definition in (2.5)). Thus, there is dependence between all the sub-trajectories since all the different linear processes are constructed from the same white noise. The aim of this paper is to present a method for estimating from  $(X_1, \dots, X_n)$  the number  $K^*$  of abrupt changes, the  $K^*$  change-times  $(t_1^*, \dots, t_{K^*}^*)$  and the  $K^* + 1$  different long-memory parameters  $(d_1^*, \dots, d_{K^*+1}^*)$ , which are unknown.

The framework of offline multiple changes we have chosen has to be distinguished from that of the online one, for which a monitoring procedure is adopted and test of detection of change is successively applied (such as CUSUM procedure). The book of [Basseville and Nikiforov \(1993\)](#) is a good reference for an introduction on both online and offline methods. There exist several methods for building a sequential detector of long memory, in several frameworks, see for instance [Beran and Terrin \(1996\)](#), [Giraitis et al. \(2001\)](#), [Horvath \(2001\)](#), [Ray and Tsay \(2002\)](#), [Kokoszka and Leipus \(2003\)](#), [Yamaguchi \(2011\)](#), [Lavancier et al. \(2013\)](#) or [Bibinger et al. \(2017\)](#). In [Wang and Wang \(2006\)](#), a semi-parametric statistics based on a distance between two spectral estimators of the memory parameter allows to test if a change in this parameter occurs and also provides an estimator of the change date.

For our offline framework, following the previous purposes, we have chosen to set to build a penalized contrast based on a sum of successive local Whittle contrasts and to minimize it. The principle of this method, minimizing a penalized contrast, provides very convincing results in many frameworks: in case of mean changes with least squares contrast (see [Bai, 1998](#)), in case of linear models changes with least squares contrast (see [Bai and Perron, 1998](#), generalized by [Lavielle, 1999](#), and [Lavielle and Moulines, 2000](#)) or least absolute deviations (see [Bai, 1998](#)), in case of spectral densities changes with usual Whittle contrasts (see [Lavielle and Ludena, 2000](#)), in case of time series changes with quasi-maximum likelihood (see [Bardet et al., 2012](#)), . . . . Clearly, the remarkable paper of [Lavielle and Ludena \(2000\)](#) was the model of this article except

that we used a semi-parametric version of their Whittle contrast with the local Whittle contrast, and this engenders additional difficulties.

Restricting our paper to long-range dependent linear processes, we obtained several asymptotic results. First the consistency of the estimator has been established under assumptions on the second order term of the expansion of the spectral density close to 0. A convergence rate of the change times estimators is also provided, but we are not able to reach the usual  $O_{\mathbb{P}}(1)$  converge rate, which is obtained for instance in the parametric case (see [Lavielle and Ludena, 2000](#)).

Monte-Carlo experiments illustrate the consistency of the estimators. When the number of changes is known, the theoretical results concerning the consistencies of the estimator are satisfying and  $n = 5000$  provides very convincing results while they are still mediocre for  $n = 2000$  and bad for  $n = 500$ . This is not surprising since we considered a semi-parametric statistical framework. When the number of changes is unknown, although we choose an asymptotically consistent choice of penalization sequence, the consistency is not satisfying even for large sample such that  $n = 5000$ . The accuracy of the number of changes estimator is extremely dependent on the precise choice of the penalization sequence, even if this choice should not be important asymptotically. Then we have chosen to use a data-driven procedure for computing “optimal” penalty, the so-called “slope heuristic” procedure defined in [Arlot and Massart \(2009\)](#). It provides more accurate results than with a fixed penalization sequence and it leads to convincing results when  $n = 5000$ .

The following Section 2 is devoted to define the framework and the estimator. Its asymptotic properties are studied in Section 3. The concrete estimation procedure and numerical applications are presented in Section 4. Finally, Section 5 contains the main proofs.

## 2. Definitions and assumptions

### 2.1. The multiple changes framework

We consider in the sequel the case of multiple change long-range dependent linear processes. First we define a class  $L(d, \beta, c)$  of real sequences, where  $d \in (0, 1/2)$ ,  $\beta \in (0, 2]$  and  $c > 0$ :

**Class  $L(d, \beta, c)$ :** A sequence  $(a_i)_{i \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  belongs to the class  $L(d, \beta, c)$  if

- $|a_n| = cn^{d-1} + O(n^{d-1-\beta})$  when  $n \rightarrow \infty$ ;
- $\frac{\partial}{\partial \lambda} \alpha(\lambda) = O(|\lambda^{-1} \alpha(\lambda)|)$  when  $\lambda \rightarrow 0^+$  with  $\alpha(\lambda) = \sum_{j=0}^{\infty} a_j e^{ij\lambda}$ .

Note that the class  $L(d, \beta, c)$  is included in  $\ell^2(\mathbb{R})$ , the Hilbert space of square-summable sequences.

Now, for  $(a_i)_{i \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  a sequence of the class  $L(d, \beta, c)$ , it is possible to define a second order linear long-range dependent process. For this, consider  $(\varepsilon_t)_{t \in \mathbb{Z}}$

a sequence of independent and identically distributed random variables (iidrv) with zero mean satisfying:

$$\mathbb{E}(\varepsilon_0^4) = \mu_4 < \infty \quad \text{and} \quad \mathbb{E}(\varepsilon_0^2) = 1. \tag{2.1}$$

Then we can define  $Y = (Y_k)_{k \in \mathbb{Z}}$  such that

$$Y_k = \sum_{j=0}^{\infty} a_j \varepsilon_{k-j} \quad \text{for } k \in \mathbb{Z},$$

which is a consistent sum in  $\mathbb{L}^2(\Omega, \mathcal{A}, \mathbb{P})$  since its partial sums are Cauchy sequences:  $\mathbb{E}(|\sum_{j=n}^m a_j \varepsilon_{k-j}|^2) = \sum_{j=n}^m a_j^2 \xrightarrow{n \rightarrow \infty} 0$  because  $|a_n| = c n^{d-1} + O(n^{d-1-\beta})$  and  $0 < d < 1/2$ . Then  $Y$  is a zero mean stationary process, with autocovariance  $r(k) = \mathbb{E}(Y_0 Y_k)$  satisfying

$$r(n) = c^2 B(1 - 2d, d) n^{2d-1} + O(n^{2d-1-\beta}) \quad \text{when } n \rightarrow \infty, \tag{2.2}$$

with  $B(u, v)$  the usual Beta function (see for instance [Inoue, 1997](#)). It is also possible to define the spectral density  $f$  of  $Y$  in  $[-\pi, 0) \cup (0, \pi]$  and it satisfies for  $d \in (0, 1/2)$

$$f(\lambda) = \frac{c^2}{\pi} B(1 - 2d, d) \Gamma(2d) \sin\left(\frac{\pi}{2} - \pi d\right) |\lambda|^{-2d} + O(|\lambda|^{-2d+\beta})$$

when  $\lambda \rightarrow 0$ , (2.3)

using the Tauberian Theorem in [Zygmund \(1968\)](#) and with  $\Gamma(u)$  the usual Gamma function. By the way, we can also write that there exists  $c' > 0$  such that

$$f(\lambda) = c' |\lambda|^{-2d} + O(|\lambda|^{-2d+\beta}) \quad \text{when } \lambda \rightarrow 0, \tag{2.4}$$

that is the classical assumption required for instance in [Robinson \(1995b\)](#).

Using these definitions, we are going to give the following assumption satisfied by the trajectory  $(X_1, \dots, X_n)$  of the process  $X$  from we study the changes:

**Assumption A\*:** Let  $(\varepsilon_t)_{t \in \mathbb{Z}}$  be a sequence of satisfying (2.1). Denote also:

- $K^*$  a given integer number in  $\mathbb{N}$ ,  $\tau_0^* = 0 < \tau_1^* < \dots < \tau_{K^*}^* < 1 = \tau_{K^*+1}^*$ ;
- for any  $1 \leq i \leq K^* + 1$ ,  $d_i^* \in (0, 1/2)$ ,  $c_i^* > 0$  and  $\beta_i^* \in (0, 2]$ ;
- for any  $1 \leq i \leq K^* + 1$ ,  $(a_t^{(i)})_{t \in \mathbb{N}}$  is a sequence belonging to the class  $L(d_i^*, \beta_i^*, c_i^*)$ .

Define the process  $X = (X_t)_{1 \leq t \leq n}$  with  $n$  large enough ( $n > K^* + 1$ ) such that

1. for  $i = 1, \dots, K^* + 1$ ,

$$X_t = \sum_{j=0}^{\infty} a_j^{(i)} \varepsilon_{t-j} \quad \text{when } [n\tau_{i-1}^*] + 1 \leq t \leq [n\tau_i^*]. \tag{2.5}$$

2. For  $i = 1, \dots, K^*$ ,  $d_{i+1}^* - d_i^* \neq 0$  and denote

$$\Delta_d = \max_{1 \leq i \leq K^*} |d_{i+1}^* - d_i^*| > 0. \quad (2.6)$$

Remark that it should be more rigorous to write  $X_t^{(n)}$  instead of  $X_t$  and therefore to consider triangular arrays  $(X_t^{(n)})_{1 \leq t \leq n, n \in \mathbb{N}^*}$  instead of time series  $(X_t)_{t \in \mathbb{N}^*}$ . But in order not to overload already very technical formulas, we will rather keep the notation  $X_t$ .

Under Assumption  $A^*$ , there are  $K^*$  changes in the dynamic of  $(X_t)_{1 \leq t \leq n}$ . The first condition (2.5) is relative to the behavior  $(X_t)$  in each stage: it is a stationary linear long-range process with a spectral density satisfying (2.4) (where  $d = d_i^*$ ). Moreover there also exists a dependence for  $(X_t)$  from one stage to another one (see for instance the proof of Lemma 5.2 where the covariance between two subtrajectories of  $X$  is computed in (5.12)), which makes the model much more realistic than if the independence of successive regimes had been assumed. The second condition (2.6) is the key condition insuring that the framework is the one of multiple long-range dependence change.

## 2.2. Definition of the estimator

First we will add other notation:

For  $X$  satisfying Assumption  $A^*$ , denote:

- for any  $1 \leq i \leq K^* + 1$ ,  $t_i^* = [n\tau_i^*]$ ,  $T_i^* = \{t_{i-1}^* + 1, t_{i-1}^* + 2, \dots, t_i^*\}$  and  $n_i^* = t_i^* - t_{i-1}^*$ .

More generally, for  $K \in \{0, \dots, n-1\}$ ,  $t_0 = 1 < t_1 < \dots < t_K < t_{K+1} = n$  and for any  $i, j \in \{1, \dots, K+1\}$ , denote

- $T_i = \{t_{i-1} + 1, t_{i-1} + 2, \dots, t_i\}$  and  $n_i = t_i - t_{i-1}$ .
- $T_{ij} = \{t_{i-1} + 1, t_{i-1} + 2, \dots, t_i\} \cap \{t_{j-1}^* + 1, t_{j-1}^* + 2, \dots, t_j^*\}$  and  $n_{ij} = \#\{T_{ij}\}$ .
- $\mathbf{d} = (d_1, \dots, d_{K+1})$  and  $\mathbf{d}^* = (d_1^*, \dots, d_{K^*+1}^*)$ ,
- $\mathbf{t} = (t_1, \dots, t_K)$ ,  $\mathbf{t}^* = (t_1^*, \dots, t_{K^*}^*)$  and  $\boldsymbol{\tau}^* = (\tau_1^*, \dots, \tau_{K^*}^*)$ .

From Assumption  $A^*$ , denote by  $I_T$  the periodogram of  $X$  on the set  $T$  where  $T \subset \{1, \dots, n\}$ , and denote  $|T| = \#\{T\}$ :

$$I_T(\lambda) = \frac{1}{2\pi |T|} \left| \sum_{k \in T} X_k e^{-ik\lambda} \right|^2. \quad (2.7)$$

Using the seminal papers of Künsch (1987), Robinson (1995b) and Robinson and Henry (2003), we define a local Whittle estimator of  $d$ . For this, define for  $T \subset \{1, \dots, n\}$ ,  $d \in \mathbb{R}$  and  $m \in \{1, \dots, n\}$ ,

$$W_n(T, d, m) = \log(S_n(T, d, m)) - \frac{2d}{m} \sum_{k=1}^m \log(k/m) \quad (2.8)$$

$$\text{with } S_n(T, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} I_T(\lambda_j^{(n)}) \quad \text{and} \quad \lambda_k^{(n)} = 2\pi \frac{k}{n}. \quad (2.9)$$

The local Whittle objective function  $d \rightarrow W_n(T, d, m)$  can be minimized for estimating  $d$  on the set  $T$  providing the local Whittle estimator on  $T$  defined by:

$$\widehat{d}^{(T)} = \arg \min_{d \in [0, 0.5)} W_n(T, d, m).$$

**Remark 1.** Note that we use Fourier frequencies  $\lambda_k^{(n)} = 2\pi \frac{k}{n}$  in the definition of  $W_n(T, d, m)$ , while its common definition (see for instance Robinson, 1995b) consider the Fourier frequencies  $\lambda_k = 2\pi \frac{k}{|T|}$ . The explanation of this choice stems from the fact that in the definition of the following contrast  $L_n(K, \mathbf{t}, \mathbf{d}, m)$  on the whole trajectory  $(X_1, \dots, X_n)$  we will sum the local contrasts  $W_n(T_k, d_k, m)$ . This choice is required for allowing some simplifications in the proofs. But, as we assume that  $|T_i^*| = n_i^* \sim (\tau_i - \tau_{i-1}^*)n$ , we asymptotically use almost the usual frequencies.

Under Assumption  $A^*$ , we expect to estimate the distinct  $d_i^*$  on the different stages  $\{t_i^* + 1, \dots, t_{i+1}^*\}$  by using several local Whittle contrasts. In addition we will obtain a  $M$ -estimator for estimating  $d_i^*$  but also  $t_i^*$  and even  $K^*$ . Hence, for  $m \in \{1, \dots, n\}$ , we consider now a penalized local Whittle contrast defined by:

$$J_n(K, \mathbf{t}, \mathbf{d}, m) = \frac{1}{n} \sum_{k=1}^{K+1} n_k W_n(T_k, d_k, m) + K z_n, \quad (2.10)$$

where  $K \in \mathbb{N}$  is a number of changes,  $\mathbf{d} \in [0, 0.5)^{K+1}$ ,  $\mathbf{t} \in \mathcal{T}_K(0)$  and  $(z_n)$  is a sequence of positive real numbers that will be specified in the sequel.

This contrast is therefore a sum of local Whittle objective functions on the  $K + 1$  different stages  $T_k$ ,  $k = 1, \dots, K + 1$ , and a penalty term that is a linear function of the number of changes (and therefore of the number of estimated parameters). Then, with  $K_{\max} \in \mathbb{N}^*$  a chosen integer number, we define for  $a \geq 0$ :

$$\mathcal{T}_K(a) = \left\{ (t_1, \dots, t_K) \in \{2, \dots, n-1\}^K; \right. \\ \left. t_{i+1} > t_i \text{ and } |t_i - t_i^*| \geq a \text{ for all } i = 1, \dots, K \right\}. \quad (2.11)$$

Note that  $\mathcal{T}_K(0) = \{(t_1, \dots, t_K), 2 \leq t_1 < t_2 < \dots < t_K \leq n-1\}$ , which is a convenient notation, and for  $a > 0$ ,  $\mathcal{T}_K(a)$  requires the knowledge of the unknown  $(t_i^*)_{1 \leq i \leq K}$  but will be only used in the proofs. Then, define

$$(\widehat{K}_n, \widehat{\mathbf{t}}, \widehat{\mathbf{d}}) = \arg \min_{K \in \{0, \dots, K_{\max}\}, \mathbf{d} \in [0, 0.5)^{K+1}, \mathbf{t} \in \mathcal{T}_K(0)} J_n(K, \mathbf{t}, \mathbf{d}, m), \quad (2.12)$$

with  $\widehat{\mathbf{d}} = (\widehat{d}_1, \dots, \widehat{d}_{\widehat{K}_n+1})$  and  $\widehat{\mathbf{t}} = (\widehat{t}_1, \dots, \widehat{t}_{\widehat{K}_n})$ .

### 3. Asymptotic behaviors of the estimators

#### 3.1. Case of a known number of changes

We study first the case of a known number  $K^*$  of changes. In such a framework, let us define two particular cases of the minimization of the function  $J_n$ . First denote  $\tilde{\mathbf{t}} = (\tilde{t}_1, \dots, \tilde{t}_{K^*})$  and  $\tilde{\mathbf{d}} = (\tilde{d}_1, \dots, \tilde{d}_{K^*+1})$  obtained when the number of changes is known and  $\widehat{\mathbf{d}}^* = (\widehat{d}_i^*)_{1 \leq i \leq K^*+1}$  obtained when the number of changes and the change dates are known. They are defined by:

$$(\tilde{\mathbf{t}}, \tilde{\mathbf{d}}) = \arg \min_{\mathbf{d} \in [0,0.5]^{K^*+1}, \mathbf{t} \in \mathcal{T}_{K^*}(0)} J_n(K^*, \mathbf{t}, \mathbf{d}, m) \quad \text{and} \quad \widehat{\mathbf{d}}^* = \arg \min_{\mathbf{d} \in [0,0.5]^{K^*+1}} J_n(K^*, \mathbf{t}^*, \mathbf{d}, m). \quad (3.1)$$

Then, we can prove:

**Theorem 3.1.** For  $X$  satisfying Assumption  $A^*$ , with  $\tilde{\boldsymbol{\tau}} = (\tilde{\tau}_1, \dots, \tilde{\tau}_{K^*})$  where  $\tilde{\tau}_i = \frac{\tilde{t}_i}{n}$  for  $i = 1, \dots, K^*$ ,

$$(\tilde{\boldsymbol{\tau}}, \tilde{\mathbf{d}}) \xrightarrow[n, m, n/m \rightarrow \infty]{\mathcal{P}} (\boldsymbol{\tau}^*, \mathbf{d}^*).$$

This first theorem, whose proof as well as all other proofs can be found in Section 5, can be improved for specifying the rate of convergence of the estimators:

**Theorem 3.2.** For  $X$  satisfying Assumption  $A^*$ , if  $m = o(n^{2\underline{\beta}^*/(1+2\underline{\beta}^*)})$  where  $\underline{\beta}^* = \min_{1 \leq i \leq K^*+1} \beta_i^*$ , then for any  $\delta > 0$ ,

$$\lim_{\delta \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P} \left( \frac{\sqrt{m}}{n} \|\tilde{\mathbf{t}} - \mathbf{t}^*\| \geq \delta \right) = 0. \quad (3.2)$$

This result provides a bound of the “best” convergence rate of  $\tilde{\mathbf{t}}$  that is minimized by  $n^{(1+\underline{\beta}^*)/(1+2\underline{\beta}^*)}$ , *i.e.* the “best” convergence rate for  $\tilde{\boldsymbol{\tau}}$  is minimized by  $n^{-\underline{\beta}^*/(1+2\underline{\beta}^*)}$ .

**Remark 2.** This rate of convergence could be compared to the result obtained in the parametric framework of [Lavielle and Ludena \(2000\)](#) where the respective convergence rates (in probability) of  $\tilde{\mathbf{t}}$  and  $\tilde{\boldsymbol{\tau}}$  are 1 and  $n^{-1}$ . This is the price to pay for going from the parametric to the semi-parametric framework: the definition of the local Whittle estimator does not allow some simplifications as in the proof of Theorem 3.4 of [Lavielle and Ludena \(2000, p. 860\)](#). Indeed the random term of their classical used definition of Whittle contrast is  $\int_{\pi}^{\pi} I_T(\lambda)/f(\lambda)d\lambda$  while our random term is  $\log \left( \frac{1}{m} \sum_{j=1}^m (j/m)^{2d} I_T(\lambda_j^{(n)}) \right)$ : the logarithm term does not make possible their simplifications. However in the case of one change  $K^* = 1$  it is possible to get a sharper result.

**Corollary 1.** Under the assumptions of Theorem 3.2 but in the case of one change  $K^* = 1$ , we obtain: for any  $\delta > 0$ ,

$$\lim_{\delta \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P} \left( \frac{1}{n \gamma_n^-} (t_1^* - \tilde{t}_1) \geq \delta \right) + \mathbb{P} \left( \frac{1}{n \gamma_n^+} (\tilde{t}_1 - t_1^*) \geq \delta \right) = 0, \quad (3.3)$$

where

$$\begin{aligned} \bullet \text{ for } d_1^* > d_2^*, & \begin{cases} \gamma_n^+ = \frac{1}{\sqrt{m} \log(n/m)} \\ \gamma_n^- = \frac{1}{\sqrt{m}} \min \left\{ \left(\frac{n}{m}\right)^{2d_1^* - 2d_2^*}, \left(\frac{n}{m}\right)^{\frac{2d_1^* - 2d_2^*}{1+2d_2^*}} m^{-\frac{d_2^*}{1+2d_2^*}} \right\} \end{cases} ; \\ \bullet \text{ for } d_2^* > d_1^*, & \begin{cases} \gamma_n^- = \frac{1}{\sqrt{m} \log(n/m)} \\ \gamma_n^+ = \frac{1}{\sqrt{m}} \min \left\{ \left(\frac{n}{m}\right)^{2d_2^* - 2d_1^*}, \left(\frac{n}{m}\right)^{\frac{2d_2^* - 2d_1^*}{1+2d_1^*}} m^{-\frac{d_1^*}{1+2d_1^*}} \right\} \end{cases} . \end{aligned}$$

**Remark 3.** In the case of a unique change of mean of short memory stationary linear processes, Bai (1994) and Bai (1995) provided an explicit limit theorem in distribution of  $\tilde{t}$  (or  $\tilde{\tau}$ ) from a least squares and a least absolute deviation estimators. But this is obtained in case where  $\lambda_n = \mu_2^* - \mu_1^*$  decreases to 0 with  $n$  ( $\mu_1^*$  and  $\mu_2^*$  respectively denote the mean of the process before and after the change date  $t^*$ ). The limit theorem is then:

$$\lambda_n^2 (\hat{t}_n - t^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} C \arg \max_{u \in \mathbb{R}} \left\{ W(u) - \frac{1}{2} |u| \right\}$$

where  $W$  is a two-sided Brownian motion and  $C$  is a constant depending on the distribution of the process. It can be noted that in the parametric case of estimation of change in parameters of the spectral density, Lavielle and Ludena (2000) did not provide such asymptotic distribution of the change date. In Corollary 1 we only provide the convergence rate of  $\tilde{t}_1 - t_1^*$  and not an explicit limit theorem. The corollary proof, however, suggests that one might even hope for a limit theorem of  $\tilde{t}_1$ . Indeed, if a Donsker-type theorem could be obtained on the  $(Z_n^s)_n$  (see its definition in the proof) we could obtain

$$\begin{aligned} \bullet \text{ if } \tilde{t}_1 - t_1^* \geq 0 & \quad n \gamma_n^+ (\tilde{t}_1 - t_1^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \arg \min_{s \geq 0} (c^+ s + Z^s) \\ \bullet \text{ if } \tilde{t}_1 - t_1^* \leq 0 & \quad n \gamma_n^- (t_1^* - \tilde{t}_1) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \arg \min_{s \in \mathbb{R}} (c^- s + Z^s) \end{aligned}$$

where  $c^+$  and  $c^-$  are real numbers depending on  $d_1^*$ ,  $d_2^*$ ,  $\tau_1^*$ ,  $c_{0,1}^*$  and  $c_{0,2}^*$ .

**Remark 4.** An important thing to notice in the corollary result is that the convergence rate of  $t_1$  to  $t_1^*$  is not symmetrical according to the values taken by  $d_1^*$  and  $d_2^*$ , which was for example not at all the case in the results of Bai (1994) or Lavielle and Moulines (2000). This difference will induce a bias in the estimation of the change date a bias that depends on  $d_1^*$ ,  $d_2^*$  and other parameters of the process and this will be confirmed by Monte-Carlo simulations (see Section 4).

**Remark 5.** For  $\underline{\beta} = 2$  (typically the case of a FARIMA or a FGN processes), for  $\hat{t}_n - t^* \geq 0$ , by choosing  $m = n^\kappa$  with any  $\kappa < 2/3$ , we obtain a convergence rate:

- if  $d_1^* > d_2^*$ ,  $n \gamma_n^+ = n^{1-\kappa/2} / \log(n) = n^{\kappa'} / \log(n)$  with  $\kappa' > 2/3$ ;
- if  $d_2^* > d_1^*$ ,  $n \gamma_n^- = n^{\kappa'} n^{(2d_1^* - 2d_2^*)(1-\kappa)} = n^{0.5+\kappa''(0.5+2d_1^* - 2d_2^*)}$  with  $\kappa'' > 1/3$ : the converging rate depends on  $d_1^* - d_2^*$ . For instance, taking the extreme cases, when  $d_1^* \simeq 0$  and  $d_2^* \simeq 0.5$ , the convergence rate is  $n^\delta$  with  $\delta > 1/3$ , while when  $d_1^* \simeq d_2^*$ , it is  $n^\delta$  with  $\delta > 2/3$ .

Another consequence of Theorem 3.2 is that there is asymptotically a small loss on the convergence rates of the long memory parameter local Whittle estimators  $\tilde{d}_i$  when the change dates are estimated instead of being known. More formally, using the results of Robinson (1995b) improved by Dalla et al. (2006), we know that under conditions of Theorem 3.2,  $\hat{d}_i^*$  satisfies

$$\sqrt{m}(\hat{d}_i^* - d_i^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \frac{1}{4}). \quad (3.4)$$

Unfortunately, the rate of convergence obtained for  $\tilde{t}_i$  in Theorem 3.2 does not allow to keep this limit theorem when  $\hat{d}_i^*$  is replaced by  $\tilde{d}_i$ . We rather obtain:

**Theorem 3.3.** Under the assumptions of Theorem 3.2, for  $i = 1, \dots, K^* + 1$ ,

$$\lim_{M \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P}(\sqrt{m} |\tilde{d}_i - d_i^*| \leq M) = 1. \quad (3.5)$$

Hence, the rate of convergence of each  $\tilde{d}_i$  we established when multiple changes occur is a little bit less sharp than the one without change. Note that using the result of Corollary 1, this rate could be improved in case of one change.

### 3.2. Case of an unknown number of changes

Here we consider the case where  $K^*$  is unknown. For estimating  $K^*$ , the penalty term of penalized local Whittle contrast  $J_n$  defined in (2.10) is now essential. Indeed, we obtain:

**Theorem 3.4.** Under the assumptions of Theorem 3.2, if  $K_{\max} \geq K^*$  and if the sequence  $(z_n)_n$  defined in (2.10) is such that  $\max(z_n, \frac{1}{z_n \sqrt{m}}) \xrightarrow[n \rightarrow \infty]{} 0$ , using  $(\hat{K}, \hat{\mathbf{t}}, \hat{\mathbf{d}})$  defined in (2.12), then

$$(\hat{K}, \hat{\mathbf{t}}, \hat{\mathbf{d}}) \xrightarrow[n, m, n/m \rightarrow \infty]{\mathcal{P}} (K^*, \boldsymbol{\tau}^*, \mathbf{d}^*).$$

The conditions we obtained on  $m$  and  $z_n$  imply that  $n^{-\underline{\beta}^*/(1+2\underline{\beta}^*)} = o(z_n)$ , depending on  $\underline{\beta}^*$  that is generally unknown. However, the choice  $z_n = n^{-1/2}$  is a possible choice solving this problem. The provided proof does not allow to establish the consistency of a typical BIC criterion, which should be  $z_n = 2 \log n/n$  (and the forthcoming numerical results obtained using this BIC penalty are not surprisingly a disaster).

**Corollary 2.** Under the conditions of Theorem 3.4, the bounds (3.3) and (3.5) hold, *i.e.*, for  $i = 1, \dots, K^* + 1$ ,

$$\lim_{\delta \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\hat{\mathbf{t}} - \mathbf{t}^*\| \geq \delta\right) = 0$$

and  $\lim_{M \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P}(\sqrt{m} |\hat{d}_i - d_i^*| \leq M) = 1.$

Then the convergence rates of the estimators obtained in the case where the number of changes is unknown is the same as if the number of changes is known.

## 4. Numerical experiments

In the sequel we first describe the concrete procedure for applying the new multiple changes estimator, then we present the numerical results of Monte-Carlo experiments.

### 4.1. Concrete procedure of estimation

Several details require to be specified to concretely apply the multiple changes estimator. Indeed, we have done:

1. The choice of meta-parameters:
  - as we mainly studied the cases of FARIMA processes for which  $\beta = 2$ , we have chosen to set  $m = \lfloor n^{0.6} \rfloor$  since  $0.6 < 2\beta/(2\beta + 1) = 0.8$ ;
  - the number  $K_{\max} \geq K^*$  is crucial for the heuristic plot procedure (see below) and was chosen such that  $K_{\max} = 2(\lfloor \log(n) \rfloor - 1)$ , implying  $K_{\max} = 10, 12$  and  $14$  respectively for  $n = 500, 2000$  and  $5000$ . Those calibrations correspond to the results of numerous Monte Carlo experiments that we have chosen not to show here.
2. As the choice of the sequence  $(z_n)$  of the penalty term is not exactly specified but just has to satisfy  $\max(z_n, \frac{1}{z_n \sqrt{m}}) \xrightarrow[n \rightarrow \infty]{} 0$ . After other numerical simulations (not reported here), we have chosen to set  $z_n = 2/\sqrt{n}$  that offers the best results among our choices.
3. The dynamic programming procedure is implemented for allowing a significant decrease of the time consuming. Such procedure is very common in the offline multiple change context and has been described with details in [Kay \(1998\)](#).
4. For improving the procedure of selection of the changes number  $K^*$  for not too large samples, we implemented a data-driven procedure so-called “the slope heuristic procedure”. This procedure was introduced by [Arlot and Massart \(2009\)](#) in the framework of least squares estimation with fixed design, but that can be extended in many statistical fields (see [Baudry et al., 2012](#)). Applications in the multiple changes detection problem was already successfully done in [Baudry et al. \(2012\)](#) in an i.i.d. context and also for dependent time series in [Bardet et al. \(2012\)](#). In a general framework, it consists in computing  $-2 \log(\widehat{LIK}(K))$  where  $\widehat{LIK}(K)$  is the maximized likelihood for any  $K \in \{0, 1, \dots, K_{\max}\}$ . Here  $-2 \log(\widehat{LIK}(K))$  is replaced by  $\frac{1}{n} \sum_{k=1}^{K+1} n_k W_n(\tilde{T}_k, \tilde{d}_k, m)$ . Then for  $K > K^*$ , the decreasing of this contrast with respect to  $K$  is almost linear with a slope  $s$  (see [Figure 1](#) where the linearity can be observed when  $K > K^* = 4$ ), which can be estimated for instance by a least-squares estimator  $\hat{s}$ . Then  $\hat{K}_H$  is obtained by minimizing the penalized contrast

$J_n$  using  $\widehat{z}_n = 2\widehat{s}$ , i.e.

$$\widehat{K}_H = \arg \min_{0 \leq K \leq K_{\max}} \left\{ \frac{1}{n} \sum_{k=1}^{K+1} (\tilde{t}_{k+1} - \tilde{t}_k) W_n(\{\tilde{t}_k + 1, \dots, \tilde{t}_{k+1}\}, \tilde{\mathbf{d}}_k, m) + 2\widehat{s}K \right\}.$$

By construction, the procedure is sensitive to the choice of  $K_{\max}$  since a least squares regression is carried out for the “largest” values of  $K$  and we preferred to choose the largest reasonable value of  $K_{\max}$ .

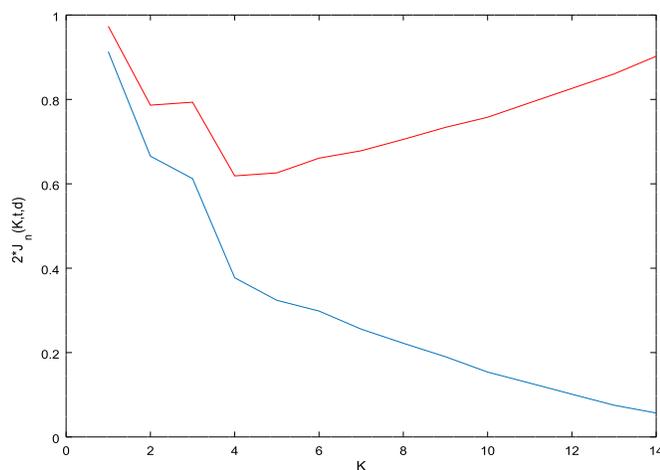


FIG 1. For  $n = 5000$ ,  $K^* = 4$  and a FARIMA(0,  $d$ , 0) process, the graph of  $2 \times J_n(K, \widehat{\mathbf{t}}, \widehat{\mathbf{d}}, m)$  (in blue), and the one of  $2 \times J_n(K, \widehat{\mathbf{t}}, \widehat{\mathbf{d}}, m) + 2 \times \widehat{s} \times K$  (in red).

A software was written with Octave software (also executable with Matlab software) and is available on

<http://samm.univ-paris1.fr/IMG/zip/detectchange.zip>.

#### 4.2. Monte-Carlo experiments in case of known number of changes

In the sequel we first exhibit the consistency of the multiple breaks estimator when the number of changes is known. Monte-Carlo experiments are realized in the following framework:

1. Three kinds of processes are considered: a FARIMA(0,  $d$ , 0) process, a FARIMA(1,  $d$ , 1) process with a AR coefficient  $\psi = -0.7$  and a MA coefficient  $\theta = 0.3$  (this refers to the familiar representation  $(1 - \psi B)X = (1 - B)^{-d}(1 + \theta B)\varepsilon$  where  $B$  is the backward operator) and a linear stationary process called  $X^{(d,1)}$  belonging to Class  $L(d, 1, 1)$ , since we chose a sequence  $(a_k)_{k \in \mathbb{N}}$  satisfying

$$a_k = (k + 1)^{d-1} + (k + 1)^{d-2} \quad \text{for all } k \in \mathbb{N}.$$

TABLE 1

RMSE of the estimators from 500 independent replications of processes, when the number  $K^*$  of changes is known. For  $K^* = 0$  the first line is obtained with  $d_1^* = 0.4$ , the second one with  $d_1^* = 0.1$

	$n$	FARIMA(0, $d$ , 0)			FARIMA(1, $d$ , 1)			$X^{(d,1)}$		
		500	2000	5000	500	2000	5000	500	2000	5000
$K^*=0$	$\tilde{d}_1$	0.070	0.047	0.034	0.098	0.090	0.066	0.077	0.048	0.035
	$\tilde{d}_1$	0.075	0.046	0.033	0.224	0.119	0.073	0.199	0.165	0.146
$K^*=1$	$\tilde{\tau}_1$	0.202	0.025	0.011	0.193	0.038	0.012	0.216	0.143	0.091
	$\tilde{d}_1$	0.178	0.055	0.043	0.099	0.096	0.082	0.189	0.130	0.092
	$\tilde{d}_2$	0.181	0.063	0.043	0.317	0.188	0.128	0.258	0.162	0.130
$K^*=3$	$\tilde{\tau}_1$	0.264	0.177	0.020	0.257	0.162	0.016	0.197	0.175	0.095
	$\tilde{\tau}_2$	0.231	0.144	0.035	0.231	0.134	0.011	0.223	0.208	0.141
	$\tilde{\tau}_3$	0.252	0.099	0.017	0.225	0.145	0.013	0.236	0.160	0.120
	$\tilde{d}_1$	0.182	0.075	0.047	0.117	0.095	0.087	0.283	0.200	0.103
	$\tilde{d}_2$	0.327	0.114	0.066	0.357	0.282	0.167	0.347	0.276	0.167
	$\tilde{d}_3$	0.414	0.206	0.055	0.165	0.097	0.088	0.470	0.257	0.105
	$\tilde{d}_4$	0.215	0.099	0.061	0.365	0.293	0.196	0.308	0.206	0.149

Note that both the FARIMA processes belongs to Class  $L(d, 2, c_0)$ .

2. For  $n = 500, 2000$  and  $5000$ , two cases are considered:

- Zero change,  $K^* = 0$  and  $d_1^* = 0.4$ , then  $d_1^* = 0.1$ , for obtaining a benchmark of the accuracy of local Whittle estimator of the long-range dependence parameter;
- One change,  $K^* = 1$  and  $(d_1^*, d_2^*) = (0.4, 0.1)$  and  $\tau_1^* = 0.5$ ;
- Three changes,  $K^* = 3$  and  $(d_1^*, d_2^*, d_3^*, d_4^*) = (0.4, 0.1, 0.4, 0.1)$  and  $(\tau_1^*, \tau_2^*, \tau_3^*) = (0.25, 0.5, 0.75)$ .

3. Each case is independently replicated 500 times and the RMSE, Root-Mean-Square Error, is computed for each estimator of the parameter. For a known parameter  $\theta^*$  and an estimator  $\hat{\theta}$  that is computed 500 times, we have:

$$RMSE(\hat{\theta}) = \left( \frac{1}{500} \sum_{i=1}^{500} (\hat{\theta}_i - \theta^*)^2 \right)^{1/2}.$$

The results of Monte-Carlo experiments are detailed in Table 1.

**Conclusion:** from Table 1, we conclude:

1. Even using the local Whittle estimator which is probably the most accurate in this framework, it is easy to verify that if the behaviour of the spectral density in 0 is not smooth, then even with a trajectory of size 5000, we keep a quadratic risk greater than 0.1 (see the case  $K^* = 0$  for a FARIMA(1,  $d$ , 1) or for the  $X^{(d,1)}$  process). We do not have to forget that the parameter  $d$  is relative to the long-memory behaviour of the process, in a semi-parametric setting.
2. If the number of changes is known, the estimators of  $\tau_i$  and  $d_i$  are consistent but their rates of convergence are significantly impacted by the

TABLE 2  
 Bias, standard deviation and RMSE of  $\tilde{t}_1$  for several values of  $d_1^*$  and  $d_2^*$  from 500 independent replications of processes FARIMA(0,  $d$ , 0)

n		1000	2000	4000	8000	16000
$d_1^* = 0.31$ and $d_2^* = 0.01$	Bias	0.0	-11.2	-12.5	-15.9	-17.1
	Standard Deviation	78.2	61.4	53.4	52.7	51.0
	RMSE	78.2	62.4	54.9	54.9	53.8
$d_1^* = 0.40$ and $d_2^* = 0.10$	Bias	4.4	-9.1	-9.7	-11.4	-13.0
	Standard Deviation	66.3	58.7	55.8	49.4	48.6
	RMSE	66.4	59.4	56.6	50.7	50.0
$d_1^* = 0.49$ and $d_2^* = 0.19$	Bias	5.7	-0.6	-1.3	-6.0	-9.3
	Standard Deviation	72.8	61.4	53.9	47.1	49.4
	RMSE	73.0	61.4	54.0	47.5	50.3
$d_1^* = 0.01$ and $d_2^* = 0.31$	Bias	14.4	17.2	18.1	22.3	27.0
	Standard Deviation	72.6	71.3	65.4	57.2	61.6
	RMSE	74.0	73.3	67.8	61.4	67.3
$d_1^* = 0.10$ and $d_2^* = 0.40$	Bias	14.3	14.8	23.2	29.4	27.3
	Standard Deviation	70.4	56.0	58.1	59.2	61.0
	RMSE	66.4	71.9	57.9	66.1	66.8
$d_1^* = 0.19$ and $d_2^* = 0.49$	Bias	11.4	17.7	20.8	22.2	33.4
	Standard Deviation	83.2	61.4	70.3	66.1	71.0
	RMSE	84.0	63.9	73.3	74.0	78.5

number of changes: as we could imagine, the largest  $K^*$  the largest the RMSE of the estimators. But finally, the case  $n = 5000$  provides extremely convincing results in FARIMA framework concerning the estimation of  $\tau_i$ , while the convergence rates for the process  $X^{(d,1)}$  are slow (since the asymptotic behavior of the spectral density around 0 is clearly rougher than in FARIMA framework).

#### 4.3. Convergence rate of $\tilde{t}_1$ in case of one change

In Theorem 3.2 an asymptotic bound is specified concerning the convergence rates of  $(\tilde{\tau}_i)$  and therefore of  $(\tilde{t}_i)$ : roughly speaking we have  $|\tilde{t}_i - t_i^*| = O_{\mathbb{P}}\left(\frac{n}{\sqrt{m}}\right)$ . In the case of one change ( $K^* = 1$ ) and with  $\beta = 2$ , this rate has been improved in Corollary 1 and is asymmetric. For instance, if  $d_1^* > d_2^*$ , with  $m = \lfloor n^{0.6} \rfloor$

- if  $\tilde{t}_i - t_i^* \geq 0$ , we have  $\tilde{t}_1 - t_1^* = O_{\mathbb{P}}\left(\frac{n^{0.7}}{\log n}\right)$ ;
- if  $\tilde{t}_i - t_i^* \leq 0$ , we rather obtain  $t_1^* - \tilde{t}_1 = O_{\mathbb{P}}\left(n^{0.7+0.8(d_1^*-d_2^*)}\right)$ .

We deduce that the convergence rate is asymptotically faster for  $\tilde{t}_i - t_i^* \leq 0$  than for  $\tilde{t}_i - t_i^* \geq 0$  and that there is a non symmetric asymptotic distribution of  $\tilde{t}_i - t_i^*$ . We wanted to study numerically the convergence rate in several configurations: with various values of  $n$  from 1000 to 16000, various values of  $d_1^*$  and  $d_2^*$ , keeping  $t_1^* = n/2$  and FARIMA(0,  $d$ , 0) processes. The results are reported in Table 2.

**Conclusion:** from Table 2, we conclude:

1. There exists an estimation-bias of  $\tilde{t}_1$  that increases with  $n$ . The sign of this bias is depending on the increasing (*i.e.* positive bias) or the decreases-

ing (i.e. negative bias) of the long-memory parameter after the change date, as it was established in Corollary 1 and Remarks 3, 4 and 5. This can be explained by the following fact: the local Whittle estimator of the long-memory parameter on a trajectory composed of long-memory processes with various parameters always provides an estimate of the largest parameter. Hence, following the sign of  $d_2^* - d_1^*$ , the transition zone may or may not be included in the break time estimation.

2. The standard deviation of  $\tilde{t}_1$  seems to weakly increase when  $n$  increases.
3. Even if we chose to keep a constant value  $|d_2^* - d_1^*| = 0.3$ , we can see that in terms of bias as well as standard deviation, the asymptotic behavior of  $\tilde{t}_1$  also depends on the values of  $d_1^*$  and  $d_2^*$  (but also on  $c_{0,1}^*$  and  $c_{0,1}^*$ ) and not only on the value of  $d_2^* - d_1^*$ .
4. However it should be noted that the results of the simulations cannot be very conclusive: with  $\beta = 2$ ,  $m = n^{0.6}$  and  $|d_1^* - d_2^*|$ , one obtains that  $\gamma_n^+/\gamma_n^- = n^{0.24}/\log(n)$  which increases from 0.76 to 1.05 when  $n$  increases from 1000 to 16000.

Finally, these numerical experiments seem to show a fairly fast convergence rate of  $\tilde{\tau}_1$  but also the difficulty to obtain a precise or estimated asymptotic distribution of  $\tilde{t}_1$ .

#### 4.4. Monte-Carlo experiments in case of unknown number of changes

In this subsection, we consider the result of the model selection using the penalized contrast for estimating the number of changes  $K^*$ . We reply exactly the same framework that in the previous subsection and notify the frequencies of the event ' $\hat{K} = K^*$ ', for:

- $\hat{K} = \hat{K}_n$  obtained directly by minimizing  $J_n$  with  $z_n = 2/\sqrt{n}$ ;
- $\hat{K} = \hat{K}_{BIC}$  obtained directly by minimizing  $J_n$  with  $z_n = 2 \log n/n$ , following the usual BIC procedure;
- $\hat{K} = \hat{K}_H$  obtained from the “slope heuristic” procedure described previously.

We obtained the results detailed in Table 3:

TABLE 3  
Frequencies of recognition of the true number of changes with several criteria from 500 independent replications of processes

	$n$	FARIMA(0, d, 0)			FARIMA(1, d, 1)			$X^{(d,1)}$		
		500	2000	5000	500	2000	5000	500	2000	5000
$K^*=1$	$\hat{K}_n$	0.11	0.21	0.51	0.21	0.45	0.67	0.05	0.05	0.01
	$\hat{K}_{BIC}$	0	0	0	0	0	0	0	0	0
	$\hat{K}_H$	0.35	0.91	0.92	0.49	0.77	0.81	0.25	0.47	0.57
$K^*=3$	$\hat{K}_n$	0.13	0.12	0.32	0.12	0.21	0.52	0.16	0.07	0.02
	$\hat{K}_{BIC}$	0	0	0	0	0	0	0	0	0
	$\hat{K}_H$	0.02	0.16	0.85	0.03	0.21	0.80	0.07	0.16	0.32

In the case  $n = 2000$  and  $K^* = 2$ , we also plotted the histogram (see Figure 2) of the values taken by  $\widehat{K}_H$  in the FARIMA(0,  $d$ , 0) case (after 500 independent replications). Even if  $K_{\max} = 12$  we see that the values of  $\widehat{K}_H$  are essentially concentrated from 0 to 3. It is however interesting to note that with the chosen choice of evolution of the long memory parameter ( $d_1^* = 0.4$ ,  $d_2^* = 0.1$  and  $d_3^* = 0.4$ ), the estimation of the break number is much more often 0 than 1. This can be explained by the fact that the local Whittle estimator of  $d$  on the area with  $d_2^* = 0.1$  and  $d_3^* = 0.4$  is closer to 0.4 than to 0.1.

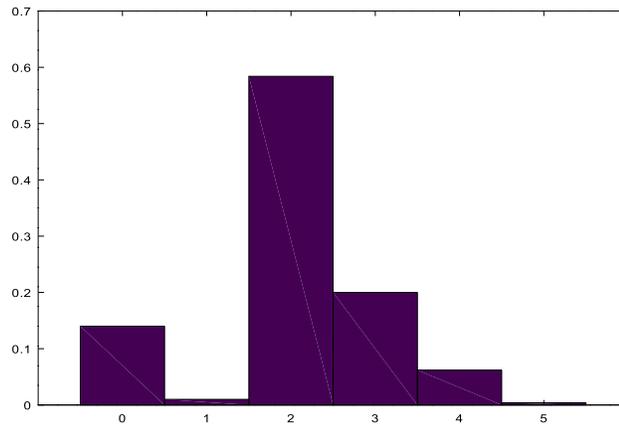


FIG 2. For  $n = 2000$ ,  $K^* = 2$ , histogram of the values taken by  $\widehat{K}_H$  in the FARIMA(0,  $d$ , 0) case.

**Conclusion:** from Table 3, we may conclude that:

1. The estimators of number of changes  $\widehat{K}_n$  and  $\widehat{K}_H$  have a satisfying behavior, meaning that they seem to converge to  $K^*$  when the sample length increases in the FARIMA framework. Once again, the convergence rates are significantly faster for small  $K^*$  values than for large ones. The results obtained with the “slope heuristic” procedure estimator  $\widehat{K}_H$  are almost the most accurate and provides very convincing results for  $n = 5000$ . Note also that the usual BIC penalty is not at all consistent, which can be explained by the use of local Whittle contrast that is not an approximation of the Gaussian likelihood as the usual Whittle contrast is. In case of process  $X^{(d,1)}$ , only  $\widehat{K}_H$  seems to be consistent while  $\widehat{K}_n$  is not able to detect the number of changes: this is due to the fact that the bandwidth parameter  $m$  can not be chosen as  $[n^{0.6}]$  for obtaining consistent estimators of long-memory parameters.
2. Finally we could underline that our detector based on a local Whittle contrast added to a “slope heuristic” data-driven penalization provides convincing results when  $n = 5000$  and not too bad when  $n = 2000$  (the case  $n = 500$  gives not significant estimation).

### 5. Proofs

Following the expansion (2.3), we denote in the sequel for  $i = 1, \dots, K^* + 1$ ,

$$c_{0,i}^* = \frac{c_i^{*2}}{\pi} B(1 - 2d_i^*, d_i^*) \Gamma(2d_i^*) \sin\left(\frac{\pi}{2} - \pi d_i^*\right). \tag{5.1}$$

We first provide the statements and the proofs of two useful lemmas:

**Lemma 5.1.** Under the assumptions of Theorem 3.1 and with  $S_n(T, d, m)$  defined in (2.9), for any  $i \in \{1, \dots, K^* + 1\}$  and  $T \subset T_i^*$ ,

$$\begin{aligned} \sup_{d \in (0, 1/2)} \min\left(T, \frac{n}{m}\right)^{-2d_i^*} \left| S_n(T, d, m) - \min\left(T, \frac{n}{m}\right)^{2d_i^*} \frac{c_{0,i}^* (2\pi)^{-2d_i^*}}{1 + 2d - 2d_i^*} \right| \\ = O_{\mathbb{P}}\left(\min\left(1, \frac{n}{m|T|}\right)^{1/2} + \left(\frac{m}{n}\right)^{\beta_i^*} + m^{-2d_i^*}\right). \end{aligned} \tag{5.2}$$

*Proof.* In the sequel, we will use intensively the notation and numerous proofs of Dalla et al. (2006). However, the results obtained in this paper have to be established again since, we consider  $\lambda_j^{(n)} = 2\pi \frac{j}{n}$  while they considered  $\lambda_j = 2\pi \frac{j}{|T|}$ .

We first define  $\eta_j^* = \frac{I_T(\lambda_j^{(n)})}{c_{0,i}^* (\lambda_j^{(n)})^{-2d_i^*}}$  and prove:

$$\mathbb{E} \left| \frac{1}{m} \sum_{j=1}^m (\eta_j^* - 1) \right| \leq C \left( \left(\frac{m}{n}\right)^{\beta_i^*} + \left(\frac{n}{|T|m}\right)^{1/2} \right) \tag{5.3}$$

where  $C > 0$  is a constant. For this we will go back to the proof of Proposition 5 in Dalla et al. (2006). Indeed, with the same notation, we have:

$$\mathbb{E} \left| \frac{1}{m} \sum_{j=1}^m (\eta_j^* - \mathbb{E}[\eta_j^*]) \right| \leq \frac{1}{m} (p_{|T|,1}(m) + p_{|T|,2}(m) + R_{|T|}(m))$$

where  $p_{|T|,1}(m) = 2\pi \sum_{j=1}^m I_{\varepsilon}(\lambda_j^{(n)})$ ,  $p_{|T|,2}(m) = \sum_{j=1}^m (\eta_j - 2\pi I_{\varepsilon}(\lambda_j^{(n)}))$ ,  $R_{|T|}(m) = \sum_{j=1}^m (\eta_j^* - \eta_j)$  with  $\eta_j = \frac{I_T(\lambda_j^{(n)})}{f(\lambda_j^{(n)})}$  and  $I_{\varepsilon}(\lambda_j^{(n)}) = \frac{1}{2\pi|T|} \left| \sum_{t=1}^{|T|} \varepsilon_t e^{it\lambda_j^{(n)}} \right|^2$ .

As in Proposition 5 of Dalla et al. (2006), we can write:

$$\begin{aligned} \mathbb{E}|R_{|T|}(m)| &\leq \sum_{j=1}^m \mathbb{E}|\eta_j^* - \eta_j| \\ &\leq C m \left(\frac{m}{n}\right)^{\beta_i^*}, \end{aligned}$$

and therefore

$$\mathbb{E}|R_{|T|}(m) - \mathbb{E}R_{|T|}(m)| \leq C m \left(\frac{m}{n}\right)^{\beta_i^*}. \quad (5.4)$$

Now, following also in Proposition 5 of Dalla et al. (2006), from Robinson (1995b, Relation (3.17)), adapted with our problem, i.e.  $j \leftrightarrow jT/n$  we have:

$$\begin{aligned} \mathbb{E}|\eta_j - 2\pi I_\varepsilon(\lambda_j^{(n)})| &\leq C |\log(1 + j|T|/n)|^{1/2} (j|T|/n)^{-1/2} \\ \implies \mathbb{E}|p_{|T|,2}(m)| &\leq C |\log(1 + j|T|/n)|^{1/2} (m|T|/n)^{-1/2}. \end{aligned} \quad (5.5)$$

Finally, we have to go back to the proof of (4.9) in Theorem 2 of Robinson (1995b) for bounding  $p_{|T|,1}(m)$ . Indeed, in this proof and using its notation we have

$$\begin{aligned} \mathbb{E}|p_{|T|,1}(m) - \mathbb{E}(p_{|T|,1}(m))| &= \mathbb{E}\left|\sum_{j=1}^m 2\pi I_\varepsilon(\lambda_j^{(n)}) - 1\right| \\ &\leq \sqrt{2} \left(\text{Var}\left(\frac{m}{|T|} \sum_{t \in T} (\varepsilon_t^2 - 1)\right) + \text{Var}\left(\sum_{s < t} d_{t-s} \varepsilon_t \varepsilon_s\right)\right)^{1/2}. \end{aligned}$$

But  $d_s = \frac{2}{|T|} \sum_{j=1}^m \cos(2\pi s j/n)$  and therefore we easily have  $|d_s| \leq 2m/|T|$ .

Using the usual expression of a sum of cosine functions, we also have  $|d_s| \leq \frac{2}{|T|} \left|\frac{\sin(\pi s m/n)}{\sin(\pi s/n)}\right| \leq \frac{2n}{\pi s |T|}$ . Therefore, using the variance expansion, we deduce that:

$$\text{Var}\left(\frac{m}{|T|} \sum_{t \in T} (\varepsilon_t^2 - 1)\right) \leq C \frac{m^2}{|T|},$$

while the variance of  $\sum_{s < t} d_{t-s} \varepsilon_t \varepsilon_s$  is

$$\begin{aligned} O\left(|T| \sum_{s=1}^{|T|} d_s^2\right) &= O\left(|T| \sum_{s=1}^{\lfloor n/m \rfloor} \left(\frac{2m}{|T|}\right)^2 + |T| \sum_{s \geq \lfloor n/m \rfloor} \left(\frac{2n}{\pi s |T|}\right)^2\right) \\ &= O\left(\frac{nm}{|T|} + \frac{nm}{|T|}\right). \end{aligned}$$

As a consequence we deduce:

$$\mathbb{E}|p_{|T|,1}(m) - \mathbb{E}(p_{|T|,1}(m))| \leq C \left(\frac{m}{|T|^{1/2}} + \left(\frac{nm}{|T|}\right)^{1/2}\right) \leq C \left(\frac{nm}{|T|}\right)^{1/2}. \quad (5.6)$$

Finally, using (5.4), (5.5) and (5.6), we deduce:

$$\mathbb{E}\left|\frac{1}{m} \sum_{j=1}^m (\eta_j^* - \mathbb{E}[\eta_j^*])\right| \leq \frac{C}{m} \left(m \left(\frac{m}{n}\right)^{\beta_i^*} + \frac{\log^{1/2}\left(\frac{m|T|}{n}\right)}{\left(\frac{m|T|}{n}\right)^{1/2}} + \left(\frac{nm}{|T|}\right)^{1/2}\right)$$

and this implies that (5.3) is established.

Now a straightforward application of Markov Inequality and Lemma 2 in Dalla et al. (2006) implies that for any  $d \in (0, 1/2)$ ,

$$\begin{aligned} & \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \eta_j^* - \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \right| = O_{\mathbb{P}}\left(\left(\frac{m}{n}\right)^{\beta_i^*} + \left(\frac{n}{m|T|}\right)^{1/2}\right) \\ \implies & \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \eta_j^* - \frac{1}{2d-2d_i^*+1} \right| \\ & = O_{\mathbb{P}}\left(\left(\frac{m}{n}\right)^{\beta_i^*} + \left(\frac{n}{m|T|}\right)^{1/2} + m^{2d-2d_i^*-1}\right). \end{aligned}$$

Since we have

$$S_n(T, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} I_T(\lambda_j^{(n)}) = \frac{c_{0,i}^*}{(2\pi)^{2d_i^*}} \left(\frac{n}{m}\right)^{2d_i^*} \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \eta_j^*,$$

we deduce that for any  $N \geq 1$ ,

$$\begin{aligned} & \sup_{|T| \geq N} \left| S_n(T, d, m) - \left(\frac{n}{m}\right)^{2d_i^*} \frac{c_{0,i}^* (2\pi)^{-2d_i^*}}{1+2d-2d_i^*} \right| \\ & = \left(\frac{n}{m}\right)^{2d_i^*} O_{\mathbb{P}}\left(\left(\frac{m}{n}\right)^{\beta_i^*} + \left(\frac{n}{mN}\right)^{1/2} + m^{2d-2d_i^*-1}\right). \quad (5.7) \end{aligned}$$

For small  $N$ , for instance such that  $N = o(n/m)$ , the random right side term is not bounded. However, for any  $T \subset T_i^*$ , we have  $\mathbb{E}(I_T(\lambda_j^{(n)})) \leq \sigma_i^2(1 + 2C \sum_{k=1}^{|T|} k^{2d_i^*-1}) \leq C|T|^{2d_i^*}$ . Thus, there exists  $C_i > 0$  such that for any  $\delta > 0$ ,

$$\mathbb{P}\left(\sup_{d \in (0, 1/2)} \left| S_n(T, d, m) - \frac{c_{0,i}^* (2\pi)^{-2d_i^*}}{1+2d-2d_i^*} |T|^{2d_i^*} \right| \geq \delta\right) \leq \frac{C_i}{\delta} |T|^{2d_i^*}. \quad (5.8)$$

Thus we deduce (5.2) and this achieves the proof of Lemma 5.1.  $\square$

In the sequel, we define:

$$\begin{aligned} R_n(T, T', d, m) &= \frac{1}{2\pi} \sum_{t \in T} \sum_{t' \in T'} X_t X_{t'} b_n(t' - t, d, m) \\ & \text{with } b_n(k, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} e^{-2\pi i \frac{jk}{n}}. \quad (5.9) \end{aligned}$$

Note that  $S_n(T, d, m)$ , which is defined in (2.9) can also be written as:

$$S_n(T, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} I_T(\lambda_j^{(n)}) = \frac{1}{2\pi|T|} \sum_{s \in T} \sum_{t \in T} X_s X_t b_n(t - s, d, m). \quad (5.10)$$

The following lemma establish an asymptotic bound for  $R_n$  when  $T$  and  $T'$  are included in distinct stages of the process:

**Lemma 5.2.** Under the assumptions of Theorem 3.1, there exists  $C > 0$  such that for any  $j, j' \in \{1, \dots, K^* + 1\}$  where  $j \neq j'$ , any  $T \subset T_j^*$  and  $T' \subset T_{j'}^*$ , and any  $N \in \mathbb{N}^*$ ,

$$\begin{aligned} \sup_{d \in (0, 1/2)} \max_{\min(|T|, |T'|) \geq N} \left( \min(|T|, |T'|, \frac{n}{m}) \right)^{-d_j^* - d_{j'}^*} \frac{1}{\min(|T|, |T'|)} |R_n(T, T', d, m)| \\ = O_{\mathbb{P}} \left( \left( \min \left( 1, \frac{n}{mN} \right) \right)^{1 - d_j^* - d_{j'}^*} \right). \end{aligned} \quad (5.11)$$

*Proof.* First, we can bound the covariance  $\text{Cov}(X_t, X_{t'})$  with  $t \in T \subset T_j^*$  and  $t' \in T' \subset T_{j'}^*$ , where  $j \neq j'$ . Indeed, assuming  $t < t'$ , define

$$\begin{aligned} \Gamma_{T, T'}(|t' - t|) &= \text{Cov}(X_t, X_{t'}) \\ &= \mathbb{E} \left( \sum_{k=0}^{\infty} a_k^{(j)} \varepsilon_{t-j} \sum_{k'=0}^{\infty} a_{k'}^{(j')} \varepsilon_{t'-j'} \right) \\ &= \sum_{k=0}^{\infty} a_k^{(j)} a_{t'-t+k}^{(j')}, \end{aligned}$$

since  $(\varepsilon_i)$  is supposed to be white noise with unit variance. Therefore, since  $a_k^{(j)} = c_j^* k^{d_j^* - 1} + O(k^{d_j^* - 1 - \beta_j^*})$  and  $a_k^{(j')} = c_{j'}^* k^{d_{j'}^* - 1} + O(k^{d_{j'}^* - 1 - \beta_{j'}^*})$ , there exists  $C$  such that

$$|a_k^{(j)} a_{t'-t+k}^{(j')}| \leq C k^{d_j^* - 1} (t' - t + k)^{d_{j'}^* - 1} \quad \text{for any } k \in \mathbb{N}^*.$$

As a consequence, there exist  $C' > 0$  and  $C'' > 0$  such that for  $t' > t$ ,

$$\begin{aligned} |\Gamma_{T, T'}(|t' - t|)| &\leq C' \sum_{k=1}^{\infty} k^{d_j^* - 1} (t' - t + k)^{d_{j'}^* - 1} \\ &= \frac{C'}{(t' - t)^{1 - d_j^* - d_{j'}^*}} \times \frac{1}{t' - t} \sum_{k=1}^{\infty} \left( \frac{k}{t' - t} \right)^{d_j^* - 1} \left( 1 + \frac{k}{t' - t} \right)^{d_{j'}^* - 1} \\ &\leq \left( C'' \int_0^{\infty} \frac{1}{x^{1 - d_j^*}} \frac{1}{(1 + x)^{1 - d_{j'}^*}} dx \right) \frac{1}{(t' - t)^{1 - d_j^* - d_{j'}^*}}. \end{aligned} \quad (5.12)$$

Now, using (5.9) and (5.12), we have:

$$\mathbb{E}(R_n(T, T', d, m)) = \frac{1}{2\pi} \sum_{t \in T} \sum_{t' \in T'} \Gamma_{T, T'}(t' - t) b_n(t' - t, d, m).$$

The right side term of the previous equality is only depending on  $(t' - t)$ . Therefore, using the notations  $\delta = -1 + \min\{|t - t'|, (t, t') \in T \times T'\} \geq 0$ ,  $\mu = \min\{|T|, |T'|\}$  and  $\nu = \max\{|T|, |T'|\}$ , it is possible to detail this term in the following way:

$$\mathbb{E}(R_n(T, T', d, m)) = \frac{1}{2\pi} \left( \sum_{k=1}^{\mu} k \Gamma_{T, T'}(\delta + k) b_n(\delta + k, d, m) \right)$$

$$\begin{aligned}
 & + \mu \sum_{k=\mu+1}^{\nu} \Gamma_{T,T'}(\delta+k) b_n(\delta+k, d, m) \\
 & + \sum_{k=\nu+1}^{\nu+\mu} (\nu+\mu-k) \Gamma_{T,T'}(\delta+k) b_n(\delta+k, d, m).
 \end{aligned}$$

But from usual calculations, for any  $d \in [-1/2, 1/2]$ , there exists  $C(d) > 0$  such that we have

$$|b_n(u, d, m)| \leq C(d) \min \left\{ 1, \left(\frac{n}{m}\right)^{1+2d} |u|^{-1-2d} \right\} \quad \text{for } u \in \mathbb{Z}. \tag{5.13}$$

As a consequence, if  $\mu + \nu \leq n/m$ , we obtain:

$$\begin{aligned}
 |\mathbb{E}(R_n(T, T', d, m))| & \leq C \left( \sum_{k=1}^{\mu} k^{d_j^*+d_{j'}^*} + 2\mu \sum_{k=\mu+1}^{\nu+\mu} k^{-1+d_j^*+d_{j'}^*} \right) \\
 & \leq C \mu \nu^{d_j^*+d_{j'}^*}.
 \end{aligned} \tag{5.14}$$

And when  $\mu \geq n/m$ , we can write:

$$\begin{aligned}
 & \left| \sum_{k=1}^{\mu} k \Gamma_{T,T'}(\delta+k) b_n(\delta+k, d, m) \right| \\
 & \leq C \sum_{k=1}^{[n/m]} k^{d_j^*+d_{j'}^*} + C \left(\frac{n}{m}\right)^{1+2d} \sum_{k=[n/m]}^{\mu} k \frac{|\delta+k|^{-1-2d}}{|\delta+k|^{1-d_j^*-d_{j'}^*}} \\
 & \leq C \left(\frac{n}{m}\right)^{1+d_j^*+d_{j'}^*} + C \left(\frac{n}{m}\right)^{1+2d} \sum_{k=[n/m]}^{\mu} k^{-1+d_j^*+d_{j'}^*-2d} \\
 & \leq C \left(\frac{n}{m}\right)^{d_j^*+d_{j'}^*} \mu \left(\frac{n}{m\mu}\right)^{1+\min(0, 2d-d_j^*-d_{j'}^*)} (\log(\mu))^{1_{2d=d_j^*+d_{j'}^*}}.
 \end{aligned}$$

Finally, by performing the same type of calculations several times, we obtain:

$$\begin{aligned}
 |\mathbb{E}(R_n(T, T', d, m))| & \leq C \mu \left( \min \left(\frac{n}{m}, \nu\right) \right)^{d_j^*+d_{j'}^*} \\
 & \quad \times \left( \min \left(1, \frac{n}{m\mu}\right) \right)^{1+\min(0, 2d-d_j^*-d_{j'}^*)} (\log(\mu))^{1_{2d=d_j^*+d_{j'}^*}}.
 \end{aligned}$$

Then we deduce

$$\sup_{d \in (0, 1/2)} |\mathbb{E}(R_n(T, T', d, m))| \leq C \mu \left( \min \left(\frac{n}{m}, \nu\right) \right)^{d_j^*+d_{j'}^*} \left( \min \left(1, \frac{n}{m\mu}\right) \right)^{1-d_j^*-d_{j'}^*}. \tag{5.15}$$

Now we are going to bound  $\text{Var}(R_n(T, T', d, m))$ . We have:

$$\text{Var}(R_n(T, T', d, m))$$

$$= \frac{1}{4\pi^2} \sum_{t \in T} \sum_{t' \in T'} \sum_{s \in T} \sum_{s' \in T'} \text{Cov}(X_t X_{t'}, X_s X_{s'}) b_n(t' - t, d, m) b_n(s' - s, d, m).$$

Without loss of generality, set  $t \leq s < t' \leq s'$ . We have:

$$\begin{aligned} & \text{Cov}(X_t X_{t'}, X_s X_{s'}) \\ &= \sum_{k=0}^{\infty} \sum_{\ell=0}^{\infty} \sum_{k'=0}^{\infty} \sum_{\ell'=0}^{\infty} a_k^{(j)} a_{\ell}^{(j)} a_{k'}^{(j')} a_{\ell'}^{(j')} \text{Cov}(\varepsilon_{t-k} \varepsilon_{t'-k'}, \varepsilon_{s-\ell} \varepsilon_{s'-\ell'}). \end{aligned}$$

Only two cases implies  $\text{Cov}(\varepsilon_{t-k} \varepsilon_{t'-k'}, \varepsilon_{s-\ell} \varepsilon_{s'-\ell'}) \neq 0$  since  $(\varepsilon_i)$  is a white noise. For the first one, it is equal to  $\mu_4 - \sigma^4$  and is obtained when  $t - k = t' - k' = s - \ell = s' - \ell'$ . For the second one, it is equal to  $\sigma^4$  and is obtained when  $(t - k = s - \ell) \neq (t' - k' = s' - \ell')$  or  $(t - k = s' - \ell') \neq (t' - k' = s - \ell)$ . As a consequence,

$$\begin{aligned} \text{Cov}(X_t X_{t'}, X_s X_{s'}) &= (\mu_4 - \sigma^4) \sum_{k=0}^{\infty} a_k^{(j)} a_{s-t+k}^{(j)} a_{t'-t+k}^{(j')} a_{s'-t+k}^{(j')} \\ &+ \sigma^4 \sum_{k=0}^{\infty} \sum_{k'=0, k' \neq k}^{\infty} a_k^{(j)} a_{s-t+k}^{(j)} a_{k'}^{(j')} a_{s'-t'+k'}^{(j')} \\ &+ \sigma^4 \sum_{k=0}^{\infty} \sum_{\ell=0, \ell \neq k}^{\infty} a_k^{(j)} a_{s'-t+k}^{(j')} a_{\ell}^{(j)} a_{t'-s+\ell}^{(j')} \end{aligned}$$

and this implies

$$\begin{aligned} |\text{Cov}(X_t X_{t'}, X_s X_{s'})| &\leq C \sum_{k=1}^{\infty} (k(s-t+k))^{d_j^*-1} ((t'-t+k)(s'-t+k))^{d_{j'}^*-1} \\ &+ C \left( \sum_{k=1}^{\infty} (k(s-t+k))^{d_j^*-1} \right) \left( \sum_{k'=1}^{\infty} (k'(s'-t'+k'))^{d_{j'}^*-1} \right) \\ &+ C \left( \sum_{k=1}^{\infty} k^{d_j^*-1} (s'-t+k)^{d_{j'}^*-1} \right) \left( \sum_{\ell=1}^{\infty} \ell^{d_j^*-1} (t'-s+\ell)^{d_{j'}^*-1} \right). \end{aligned}$$

Using the Cauchy-Schwarz Inequality, we have

$$\begin{aligned} & \sum_{k=1}^{\infty} (k(s-t+k))^{d_j^*-1} ((t'-t+k)(s'-t+k))^{d_{j'}^*-1} \\ & \leq \left( \sum_{k=1}^{\infty} (k(s-t+k))^{2d_j^*-2} \right)^{1/2} \left( \sum_{k=1}^{\infty} ((t'-t+k)(s'-t+k))^{2d_{j'}^*-2} \right)^{1/2} \end{aligned}$$

Now we apply the same trick as in (5.12) and obtain since  $s' > t'$ ,

$$\sum_{k=1}^{\infty} (k(s-t+k))^{d_j^*-1} ((t'-t+k)(s'-t+k))^{d_{j'}^*-1}$$

$$\leq C (s - t + 1)^{2d_j^* - 3/2} (t' - t + 1)^{2d_{j'}^* - 3/2},$$

and more generally,

$$\begin{aligned} |\text{Cov}(X_t X_{t'}, X_s X_{s'})| &\leq C \left( (s - t + 1)^{2d_j^* - 3/2} (t' - t + 1)^{2d_{j'}^* - 3/2} \right. \\ &\quad \left. + (s - t + 1)^{2d_j^* - 1} (s' - t' + 1)^{2d_{j'}^* - 1} + ((s' - t)(t' - s))^{d_j^* + d_{j'}^* - 1} \right). \end{aligned} \quad (5.16)$$

Therefore,

$$\begin{aligned} \text{Var}(R_n(T, T', d, m)) &\leq C \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} |\text{Cov}(X_t X_{t'}, X_s X_{s'}) b_n(t - t', d, m) b_n(s - s', d, m)| \\ &\leq C (J_1 + J_2 + J_3), \end{aligned} \quad (5.17)$$

with

$$\left\{ \begin{aligned} J_1 &= \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} (|t - s| + 1)^{2d_j^* - 3/2} (|t' - t| + 1)^{2d_{j'}^* - 3/2} \\ &\quad \times |b_n(t - t', d, m) b_n(s - s', d, m)| \\ J_2 &= \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} (1 + |t - s|)^{2d_j^* - 1} (1 + |t' - s'|)^{2d_{j'}^* - 1} \\ &\quad \times |b_n(t - t', d, m) b_n(s - s', d, m)| \\ J_3 &= \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} ((s' - t)(t' - s))^{d_j^* + d_{j'}^* - 1} \\ &\quad \times |b_n(t - t', d, m) b_n(s - s', d, m)| \end{aligned} \right. .$$

As a consequence, we can easily see that  $J_1$  is negligible with respect to  $J_2$  since  $2d - 3/2 < 2d - 1$ . Concerning  $J_2$  we use the same arguments than in [Lavielle and Ludena \(2000\)](#). Then,

$$\begin{aligned} J_2 &\leq C \sum_{t \in T} \sum_{s \in T} (1 + |t - s|)^{2d_j^* - 1} \\ &\quad \times \sum_{t' \in T'} \sum_{s' \in T'} (1 + |t' - s'|)^{2d_{j'}^* - 1} |b_n(t - t', d, m) b_n(s - s', d, m)| \\ &\leq C |T|^{2d_j^* + 1} \left( \sum_{u=0}^{|T|+|T'|} |b_n(u, d, m)|^2 \right. \\ &\quad \left. + 2 \sum_{u=0}^{|T|+|T'|} |b_n(u, d, m)| \sum_{v=u+1}^{|T|+|T'|} |b_n(v, d, m)| |v - u|^{2d_{j'}^* - 1} \right). \end{aligned}$$

Using (5.13),

$$\begin{aligned} \sum_{u=0}^{|T|+|T'|} |b_n(u, d, m)|^2 &\leq C \left( \sum_{u=0}^{\min(|T|+|T'|, n/m)} 1 + \sum_{u=\min(|T|+|T'|, n/m)}^{|T|+|T'|} (n/m)^{1+2d} |u|^{-1-2d} \right) \\ &\leq C(d) \min(|T| + |T'|, \frac{n}{m}). \end{aligned} \quad (5.18)$$

Moreover,

$$\begin{aligned}
& \sum_{u=0}^{|T|+|T'|} |b_n(u, d, m)| \sum_{v=u+1}^{|T|+|T'|} |b_n(v, d, m)| |u-v|^{2d_{j'}^*-1} \\
& \leq C \left\{ \sum_{u=0}^{\min(|T|+|T'|, n/m)} \left( \sum_{v=u+1}^{\min(|T|+|T'|, n/m)} (v-u)^{2d_{j'}^*-1} \right. \right. \\
& \quad \left. \left. + \left(\frac{n}{m}\right)^{1+2d} \sum_{v=\min(|T|+|T'|, n/m)}^{|T|+|T'|} v^{-1-2d} (v-u)^{2d_{j'}^*-1} \right) \right. \\
& \quad \left. + \left(\frac{n}{m}\right)^{2+4d} \sum_{u=\min(|T|+|T'|, n/m)}^{|T|+|T'|} \sum_{v=u+1}^{|T|+|T'|} (uv)^{-1-2d} (v-u)^{2d_{j'}^*-1} \right\} \\
& \leq C(d) \left( \min(|T| + |T'|, \frac{n}{m}) \right)^{1+2d_{j'}^*} \tag{5.19}
\end{aligned}$$

after classical computations. From (5.18) and (5.19), we obtain:

$$J_2 \leq C(d) |T|^{2d_j^*+1} \left( \min(|T| + |T'|, \frac{n}{m}) \right)^{1+2d_{j'}^*}. \tag{5.20}$$

Using the same decomposition of  $J_2$  but beginning with  $s', t' \in T'$  instead of  $s, t \in T$ , we can also replace  $T$  by  $T'$  in the previous bound. As a consequence, we obtain:

$$J_2 \leq C(d) \min \left\{ \mu^{2d_j^*+1} \left( \min(\nu, \frac{n}{m}) \right)^{1+2d_{j'}^*}, \mu^{2d_{j'}^*+1} \left( \min(\nu, \frac{n}{m}) \right)^{1+2d_j^*} \right\}. \tag{5.21}$$

Finally using symmetry reasons we also have  $J_3 = (\mathbb{E}(R_n(T, T', d, m)))^2$  and therefore:

$$J_3 \leq C \mu^2 \left( \min\left(\frac{n}{m}, \nu\right) \right)^{2d_j^*+2d_{j'}^*} \left( \min\left(1, \frac{n}{m\mu}\right) \right)^{2-2d_j^*-2d_{j'}^*}. \tag{5.22}$$

As a consequence, using (5.31), (5.32), (5.34) and (5.21), (5.22), we obtain that there exists  $C > 0$  such that:

$$\begin{aligned}
& \sup_{d \in (0, 1/2)} \text{Var}(R_n(T, T', d, m)) \\
& \leq C \mu^2 \left( \min\left(\frac{n}{m}, \nu\right) \right)^{2d_j^*+2d_{j'}^*} \left( \min\left(1, \frac{n}{m\mu}\right) \right)^{2-2d_j^*-2d_{j'}^*}. \tag{5.23}
\end{aligned}$$

Thus, with  $\mathbb{E}(R_n^2(T, T', d, m)) = \text{Var}(R_n(T, T', d, m)) + \mathbb{E}^2(R_n(T, T', d, m))$ , we have for any  $N \leq n$ ,

$$\sup_{d \in (0, 1/2)} \max_{\min(|T|, |T'|) \geq N} \frac{\left( \min(|T|, |T'|, \frac{n}{m}) \right)^{-d_j^*-d_{j'}^*}}{\min(|T|, |T'|)^2} \mathbb{E}(R_n^2(T, T', d, m))$$

$$\leq C \left( \min \left( 1, \frac{n}{mN} \right) \right)^{2-2d_j^*-2d_j^*} \quad (5.24)$$

with  $C > 0$ , and this achieves the proof of (5.11) using Lemma 2.2 and 2.4 in Lavielle and Ludena (2000).  $\square$

Now the proof of the consistency of  $\hat{\tau}$  can be established:

*Proof of Theorem 3.1. Mutatis mutandis*, we follow here a similar proof than in Lavielle and Ludena (2000). Denote

$$U_n(\mathbf{t}, \mathbf{d}, m) = J_n(K^*, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m), \quad (5.25)$$

where  $J_n$  is defined in (2.10). Then, using (5.10), we can write that for any  $\mathbf{d}$  and  $\mathbf{t}$ ,

$$U_n(\mathbf{t}, \mathbf{d}, m) = \frac{1}{n} \left[ \sum_{k=1}^{K^*+1} \left( n_k \log (S_n(T_k, d_k, m)) - n_k^* \log (S_n(T_k^*, d_k^*, m)) \right) \right] - \frac{\ell(m)}{n} \sum_{k=1}^{K^*+1} 2(n_k d_k - n_k^* d_k^*).$$

Now using a decomposition of each  $S_n$  on the ‘true’ periods, we can write:

$$S_n(T_k, d_k, m) = \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n_k} S_n(T_{kj}, d_k, m) + \frac{2}{n_k} \sum_{j=1}^{K^*+1} \sum_{j' \neq j}^{K^*} R_n(T_{kj}, T_{kj'}, d_k, m),$$

with  $R_n$  defined in (5.9). As a consequence,

$$\begin{aligned} U_n(\mathbf{d}, \mathbf{t}, m) &= \frac{1}{n} \sum_{k=1}^{K^*+1} \left[ n_k \log \left( \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n_k} \left( S_n(T_{kj}, d_k, m) + \sum_{j' \neq j}^{K^*+1} \frac{2}{n_{kj}} R_n(T_{kj}, T_{kj'}, d_k, m) \right) \right) \right. \\ &\quad \left. - n_k^* \log (S_n(T_k^*, d_k^*, m)) \right] + \frac{\ell(m)}{n} \sum_{k=1}^{K^*+1} 2(n_k d_k - n_k^* d_k^*) \\ &\geq \frac{1}{n} \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} \left[ n_{kj} \log \left( S_n(T_{kj}, d_k, m) + \sum_{j' \neq j}^{K^*+1} \frac{2}{n_{kj}} R_n(T_{kj}, T_{kj'}, d_k, m) \right) \right. \\ &\quad \left. - n_k^* \log (S_n(T_k^*, d_k^*, m)) \right] + \frac{\ell(m)}{n} \sum_{k=1}^{K^*+1} 2(n_k d_k - n_k^* d_k^*) \\ &\geq \frac{1}{n} \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} n_{kj} \left[ \log (S_n(T_{kj}, d_k, m)) \right. \end{aligned}$$

$$\begin{aligned}
& + \sum_{j'=1, j \neq j}^{K^*+1} \frac{2}{n_{kj}} R_n(T_{kj}, T_{kj'}, d_k, m) + 2d_k \ell(m) \Big] \\
& - \frac{1}{n} \sum_{k=1}^{K^*+1} n_k^* \left( \log(S_n(T_k^*, d_k^*, m)) + 2d_k^* \ell(m) \right) \quad (5.26)
\end{aligned}$$

using the concavity of  $x \mapsto \log(x)$  and with  $n_k = \sum_{j=1}^{K^*+1} n_{kj}$ . Now we are going to use Lemma 5.1 and 5.2. Therefore:

$$\begin{aligned}
\left(\frac{n}{m}\right)^{-2d_j^*} \left( S_n(T_{kj}, d_k, m) + \sum_{j'=1, j \neq j}^{K^*+1} \frac{2}{n_{kj}} R_n(T_{kj}, T_{kj'}, d_k, m) \right) \\
= \frac{c_{0,j}^* (2\pi)^{-2d_j^*}}{1 + 2d_k - 2d_j^*} + \varepsilon_{kj},
\end{aligned}$$

with  $\varepsilon_{kj} = O_{\mathbb{P}}(1)$  when  $n_{kj} = O(n/m)$  and  $\varepsilon_{kj} = o_{\mathbb{P}}(1)$  for  $n = o(n_{kj}m)$ . As a consequence, from (5.26), Lemma 5.1 and 5.2, we deduce that there exists a random variable  $D(m, n)$  such that  $D(m, n) \xrightarrow[n, m, n/m \rightarrow \infty]{\mathcal{P}} 0$  satisfying for any  $\mathbf{t}$  and  $\mathbf{d}$ ,

$$U_n(\mathbf{t}, \mathbf{d}, m) \geq \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (s(d_j^*, d_k) - s(d_j^*, d_j^*)) - |D(m, n)|$$

where for  $d \in (0, 1/2)$ ,

$$s(d_j^*, d) = 2d_j^* \log(n/m) + \log(c_{0,j}^* (2\pi)^{-2d_j^*}) - \log(1 + 2d - 2d_j^*) - 2d. \quad (5.27)$$

Now, simple computations also imply

$$U_n(\mathbf{t}, \mathbf{d}, m) \geq \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (u(d_j^*, d_k) - u(d_j^*, d_j^*)) - |D(m, n)| \quad (5.28)$$

with  $u(d_j^*, d) = -\log(1 + 2d - 2d_j^*) + 2d$ . Remark that  $u(d_j^*, d_k) - u(d_j^*, d_j^*) > 0$  for any  $d_j^* \neq d_k$  and of course  $u(d_j^*, d_j^*) - u(d_j^*, d_j^*) = 0$ . Now we could use Lemma 2.3 of Lavielle (1999, p. 88), adapted in Lemma 3.3 of Lavielle and Ludena (2000, p. 858) and we obtain that there exists  $C^* > 0$  depending only on  $\mathbf{d}^*$  such that

$$\sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (u(d_j^*, d_k) - u(d_j^*, d_j^*)) \geq \frac{C}{n} \|\mathbf{t} - \mathbf{t}^*\|_{\infty}, \quad (5.29)$$

and  $\|\mathbf{t} - \mathbf{t}^*\|_{\infty} = \max_{1 \leq k \leq K^*} \{|t_k - t_k^*|\}$ .

Therefore, it is also possible to write that for any  $\delta > 0$ ,

$$\mathbb{P}(\|\hat{\tau} - \tau^*\|_{\infty} > \delta) \leq \mathbb{P}\left(\inf_{\mathbf{d} \in (0, 1/2)^{K^*+1}} \min_{\mathbf{t} \in \mathcal{T}_{K^*}(n\delta)} U_n(\mathbf{t}, \mathbf{d}, m) < 0\right)$$

$$\begin{aligned} &\leq \mathbb{P}\left(\inf_{\mathbf{d} \in (0, 1/2)^{K^*+1}} \min_{\mathbf{t} \in \mathcal{T}_{K^*}(n\delta)} \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (u(d_j^*, d_k) - u(d_j^*, d_j^*)) \right. \\ &\qquad \qquad \qquad \left. - |D(m, n)| < 0\right) \\ &\leq \mathbb{P}\left(C\delta - |D(m, n)| < 0\right) \xrightarrow{n, m, n/m \rightarrow \infty} 0, \end{aligned}$$

since  $D(m, n) \xrightarrow[n, m, n/m \rightarrow \infty]{\mathcal{P}} 0$  and for  $\mathbf{t} \in \mathcal{T}_{K^*}(n\delta)$  we have  $\|\mathbf{t} - \mathbf{t}^*\|_\infty \geq \delta n$ . This achieves the proof.  $\square$

*Proof of Theorem 3.2.* Assume with no loss of generality that  $K^* = 1$ . From Theorem 3.1, there exists a sequence  $(u_n)_n$  of real numbers satisfying

$$u_n \sqrt{m}/n \xrightarrow{n \rightarrow \infty} \infty, \quad u_n/n \xrightarrow{n \rightarrow \infty} 0 \quad \text{and} \quad \mathbb{P}(|\tilde{t}_1 - t_1^*| > u_n) \xrightarrow{n \rightarrow \infty} 0.$$

For  $\delta > 0$ , as we have

$$\mathbb{P}\left(|\tilde{t}_1 - t_1^*| > \delta \frac{n}{\sqrt{m}}\right) \leq \mathbb{P}\left(\delta \frac{n}{\sqrt{m}} < |\tilde{t}_1 - t_1^*| \leq u_n\right) + \mathbb{P}(|\tilde{t}_1 - t_1^*| > u_n)$$

As a consequence, it is sufficient to show that  $\mathbb{P}\left(\delta \frac{n}{\sqrt{m}} < |\tilde{t}_1 - t_1^*| \leq u_n\right) \xrightarrow{n \rightarrow \infty} 0$ .

Denote  $V_{\delta, n, m} = \{t \in \mathbb{Z}/ \delta n/\sqrt{m} < |t_1 - t_1^*| \leq u_n\}$ . Then,

$$\begin{aligned} &\mathbb{P}\left(\delta \frac{n}{\sqrt{m}} < |\tilde{t}_1 - t_1^*| \leq u_n\right) \leq \\ &\mathbb{P}\left(\min_{t_1 \in V_{\delta, n, m}} (J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m)) \leq 0\right), \quad (5.30) \end{aligned}$$

where  $\hat{d}_i^*$  are defined in (3.1).

Let  $t_1 \in V_{\delta, n, m}$  and without losing generality, let's choose  $t_1 > t_1^*$ .

Then  $n_1 = t_1, n_2 = n - t_1, n_{11} = t_1^*, n_{12} = t_1 - t_1^*, n_{21} = 0$  and  $n_{22} = n - t_1, T_1^* = \{1, \dots, t_1^*\}, T_2^* = \{t_1^* + 1, \dots, n\}, T_1 = \{1, \dots, t_1\}, T_{11} = T_1^* = \{1, \dots, t_1^*\}, T_{12} = \{t_1^* + 1, \dots, t_1\}$  and  $T_2 = \{t_1 + 1, \dots, n\} = T_{22}$ .

On the one hand, using results of Lemma 5.1 and 5.2, since  $t_1/n \xrightarrow{n \rightarrow \infty} \tau_1$  and  $(t_1 - t_1^*)/n \xrightarrow{n \rightarrow \infty} 0$ , we can write

$$\frac{\frac{1}{t_1 - t_1^*} R_n(T_{11}, T_{12}, \tilde{d}_1, m)}{S_n(T_1^*, \tilde{d}_1, m)} = \left(\frac{n}{m}\right)^{1-2d_1^*} O_{\mathbb{P}}\left(\frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}}\right).$$

Therefore, using again the concavity of the logarithm function, we have:

$$\begin{aligned} &J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) \\ &= \frac{1}{n} \left\{ t_1 \log\left(\frac{t_1^*}{t_1} S_n(T_1^*, \tilde{d}_1, m)\right) + \frac{t_1 - t_1^*}{t_1} S_n(T_{12}, \tilde{d}_1, m) + \frac{2}{t_1} R_n(T_{11}, T_{12}, \tilde{d}_1, m) \right\} \end{aligned}$$

$$\begin{aligned}
 & + (n - t_1) \log (S_n(T_{22}, \tilde{d}_2, m)) + 2 \ell(m)(t_1 \tilde{d}_1 + (n - t_1) \tilde{d}_2) \Big\} \\
 \geq & \frac{1}{n} \left\{ t_1^* \log (S_n(T_1^*, \tilde{d}_1, m)) + (t_1 - t_1^*) \log (S_n(T_{12}, \tilde{d}_1, m)) \right. \\
 & \left. + (n - t_1) \log (S_n(T_{22}, \tilde{d}_2, m)) + 2 \ell(m)(t_1 \tilde{d}_1 + (n - t_1) \tilde{d}_2) \right. \\
 & \left. + \left(\frac{n}{m}\right)^{1-2d_1^*} O_{\mathbb{P}}\left(\frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}}\right) \right\} \\
 \geq & \frac{1}{n} \left\{ t_1^* W_n(T_1^*, \tilde{d}_1, m) + (t_1 - t_1^*) W_n(T_{12}, \tilde{d}_1, m) + (n - t_1) W_n(T_{22}, \tilde{d}_2, m) \right. \\
 & \left. + \left(\frac{n}{m}\right)^{1-2d_1^*} O_{\mathbb{P}}\left(\frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}}\right) \right\}.
 \end{aligned}$$

On the other hand, we also have:

$$J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) = \frac{1}{n} \left\{ t_1^* W_n(T_1^*, \hat{d}_1^*, m) + (n - t_1^*) W_n(T_2^*, \hat{d}_2^*, m) \right\}.$$

First we remark that from the definition of  $\hat{d}_1^*$ ,

$$W_n(T_1^*, \hat{d}_1^*, m) \leq W_n(T_1^*, \tilde{d}_1, m).$$

Therefore,

$$\begin{aligned}
 & \frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} \\
 & \geq \frac{1}{t_1 - t_1^*} \left\{ (t_1 - t_1^*) W_n(T_{12}, \tilde{d}_1, m) + (n - t_1) W_n(T_{22}, \tilde{d}_2, m) \right. \\
 & \left. - (n - t_1^*) W_n(T_2^*, \hat{d}_2^*, m) + \left(\frac{n}{m}\right)^{1-2d_1^*} O_{\mathbb{P}}\left(\frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}}\right) \right\}. \tag{5.31}
 \end{aligned}$$

Since  $t_1 \in V_{\delta, n, m}$ , we have

$$\frac{|T_{21}^*|}{n} \xrightarrow[\min(m, \frac{n}{m}) \rightarrow \infty]{\mathcal{P}} (1 - \tau_1^*) \quad \text{and} \quad \frac{|T_{22}|}{n} \xrightarrow[\min(m, \frac{n}{m}) \rightarrow \infty]{\mathcal{P}} (1 - \tau_1^*).$$

Then, Lemma 5.1 and more precisely inequality (5.7) can be applied. Therefore, conditionally to  $\tilde{d}_1, \tilde{d}_2$  and  $\hat{d}_2^*$ , we obtain:

$$\begin{aligned}
 W_n(T_{12}, \tilde{d}_1, m) &= 2d_2^* \log (n/m) + \log (c_{0,2}^* (2\pi)^{-2d_2^*}) - \log (1 + 2\tilde{d}_1 - 2d_2^*) - 2\tilde{d}_1 \\
 & \quad + O_{\mathbb{P}} \left[ \left(\frac{m}{n}\right)^{\beta_2^*} + \left(\frac{n}{m(t_1 - t_1^*)}\right)^{1/2} + m^{2\tilde{d}_1 - 2d_2^* - 1} \right] \\
 W_n(T_{22}, \tilde{d}_2, m) &= 2d_2^* \log (n/m) + \log (c_{0,2}^* (2\pi)^{-2d_2^*}) - \log (1 + 2\tilde{d}_2 - 2d_2^*) - 2\tilde{d}_2 \\
 & \quad + O_{\mathbb{P}} \left[ \left(\frac{m}{n}\right)^{\beta_2^*} + m^{-1/2} + m^{2\tilde{d}_2 - 2d_2^* - 1} \right] \\
 W_n(T_2^*, \hat{d}_2^*, m) &= 2d_2^* \log (n/m) + \log (c_{0,2}^* (2\pi)^{-2d_2^*}) - \log (1 + 2\hat{d}_2^* - 2d_2^*) - 2\hat{d}_2^* \\
 & \quad + O_{\mathbb{P}} \left[ \left(\frac{m}{n}\right)^{\beta_2^*} + m^{-1/2} + m^{2\hat{d}_2^* - 2d_2^* - 1} \right],
 \end{aligned}$$

since  $\ell(m) = \frac{1}{m} \sum_{j=1}^m \log(j/m) = -1 + O(m^{-1})$  that is negligible with respect to  $O_{\mathbb{P}}(m^{-1/2})$ . Therefore, (5.31) becomes:

$$\begin{aligned} & \frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} \\ & \geq \frac{1}{t_1 - t_1^*} \left\{ -(t_1 - t_1^*) (\log(1 + 2\tilde{d}_1 - 2d_2^*) - 2\tilde{d}_1) - (n - t_1) (\log(1 + 2\tilde{d}_2 - 2d_2^*) + 2\tilde{d}_2) \right. \\ & \quad + \frac{n}{m} O_{\mathbb{P}} \left( \frac{m^{\beta_2^* + 1}}{n^{\beta_2^*}} + m^{1/2} + m^{2\tilde{d}_2 - 2d_2^*} + m^{2\hat{d}_2^* - 2d_2^*} \right. \\ & \quad \left. \left. + \frac{n}{t_1 - t_1^*} m^{2\hat{d}_1^* - 2d_2^*} + \frac{n^{-2d_1^*}}{m^{-2d_1^*} (t_1 - t_1^*)^{1-d_1^* - d_2^*}} \right) \right. \\ & \quad \left. + (n - t_1^*) (\log(1 + 2\hat{d}_2^* - 2d_2^*) + 2\hat{d}_2^*) \right\}. \quad (5.32) \end{aligned}$$

But  $t_1$  is supposed to belong to  $V_{\delta, n, m}$  and therefore  $t_1 \geq t_1^* + \delta n / \sqrt{m}$ . Moreover, from Dalla et al. (2006, p. 221), when  $m$  is such that  $m = o(n^{2\beta_2^* / (1+2\beta_2^*)})$ , then:

$$\tilde{d}_2 = d_2^* + O_{\mathbb{P}}(m^{-1/2}) \quad \text{and} \quad \hat{d}_2^* = d_2^* + O_{\mathbb{P}}(m^{-1/2}). \quad (5.33)$$

Then, from (5.32), we obtain after computations,

$$\begin{aligned} & \frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} \\ & \geq 2(d_1^* - d_2^*) - \log(1 + 2(d_1^* - d_2^*)) + \frac{n}{m(t_1 - t_1^*)} O_{\mathbb{P}} \left( \frac{m^{1+\beta_2^*}}{n^{\beta_2^*}} + \sqrt{m} \right) \\ & \geq 2(d_1^* - d_2^*) - \log(1 + 2(d_1^* - d_2^*)) + O_{\mathbb{P}} \left( \frac{1}{\delta} + \frac{\sqrt{m}}{\delta} \left( \frac{m}{n} \right)^{\beta_2^*} \right). \end{aligned}$$

As  $m = o(n^{2\beta_2^* / (1+2\beta_2^*)})$  then  $\sqrt{m} \left( \frac{m}{n} \right)^{\beta_2^*} = o(1)$ . As a consequence, we finally obtain:

$$\begin{aligned} & \frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} \\ & \geq 2(d_1^* - d_2^*) - \log(1 + 2(d_1^* - d_2^*)) + O_{\mathbb{P}} \left( \frac{1}{\delta} \right). \quad (5.34) \end{aligned}$$

As  $\log(1 + x) < x$  for any  $x \in (-1, 0) \cup (0, 1)$ , and since  $d_1^* - d_2^* \neq 0$ , we obtain that

$$\lim_{\delta \rightarrow \infty} \mathbb{P} \left( \frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} < 0 \right) = 0$$

and therefore from (5.30) we deduce (3.3) and therefore the proof of Theorem 3.2 is achieved.  $\square$

*Proof of Corollary 1.* We will follow here the same reasoning than in Bai (1994) or Lavielle and Moulines (2000) in the case  $K^* = 1$  (one change point). Let

$s \in [-M, M]$  where  $M > 0$  and a sequence  $(\gamma_n)_n$  of positive real numbers such as  $\gamma_n \xrightarrow[n \rightarrow \infty]{} 0$ . Then, consider

$$\Delta_n(s) = J_n(1, t_1^* + [s n \gamma_n], (\tilde{d}_1, \tilde{d}_2), m) - J_n(1, t_1^*, (d_1^*, d_2^*), m). \quad (5.35)$$

We will prove that for appropriate sequences  $(\gamma_n)_n$ , there exist two sequences  $(a_n)_{n \in \mathbb{N}}$  and  $(b_n)_{n \in \mathbb{N}}$  not depending on  $s$  such that

$$a_n \Delta_n(s) + b_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \Delta(s) \quad \text{for any } s \in [-M, M].$$

We have

$$\begin{aligned} & J_n(1, t_1^* + [s n \gamma_n], (\tilde{d}_1, \tilde{d}_2), m) \\ &= \frac{1}{n} \left\{ ([n\tau_1^*] + [s n \gamma_n]) \left( 2\ell(m)\tilde{d}_1 + \log \left( \frac{[n\tau_1^*]}{[n\tau_1^*] + [s n \gamma_n]} S_n(T_1^*, \tilde{d}_1, m) \right. \right. \right. \\ & \quad \left. \left. \left. + \frac{[s n \gamma_n]}{[n\tau_1^*] + [s n \gamma_n]} S_n(T_{12}, \tilde{d}_1, m) + \frac{2}{[n\tau_1^*] + [s n \gamma_n]} R_n(T_{11}, T_{12}, \tilde{d}_1, m) \right) \right) \right. \\ & \quad \left. + (n - [n\tau_1^*] - [s n \gamma_n]) \left( 2\ell(m)\tilde{d}_2 + \log(S_n(T_{22}, \tilde{d}_2, m)) \right) \right\}. \end{aligned}$$

We will use the conditions  $m \xrightarrow[n \rightarrow \infty]{} \infty$ ,  $\gamma_n \xrightarrow[n \rightarrow \infty]{} 0$  and  $m = o(n^{2\beta_i^*/(1+2\beta_i^*)})$  implying  $\sqrt{m} \left(\frac{m}{n}\right)^{\beta_i^*} = o(1)$  for  $i = 1, 2$  and asymptotic results of the proofs of Lemma 5.1 and 5.2.

**I.** for  $0 \leq s \leq M$ :

$$\begin{aligned} & J_n(1, t_1^* + [s n \gamma_n], (\tilde{d}_1, \tilde{d}_2), m) \\ &= (\tau_1^* + s \gamma_n(1 + o(1))) \left( 2\tilde{d}_1(1 + o(m^{-1})) \right. \\ & \quad \left. + \log \left( \left( 1 - \frac{s}{\tau_1^*} \gamma_n(1 + o(1)) \right) \left( \frac{n}{m} \right)^{2d_1^*} \frac{c_{0,1}^*(2\pi)^{-2d_1^*}}{1 + 2\tilde{d}_1 - 2d_1^*} (1 + m^{-1/2} Z_{1,n}^s) \right. \right. \\ & \quad \left. \left. + \frac{s}{\tau_1^*} \gamma_n(1 + o(1)) \left( \min \left\{ n \gamma_n, \frac{n}{m} \right\} \right)^{2d_2^*} \frac{c_{0,2}^*(2\pi)^{-2d_2^*}}{1 + 2\tilde{d}_1 - 2d_2^*} (1 + m^{-1/2} Z_{2,n}^s) \right. \right. \\ & \quad \left. \left. + O_{\mathbb{P}} \left( \gamma_n \left( \min \left\{ n \gamma_n, \frac{n}{m} \right\} \right)^{d_1^* + d_2^*} \right) \right) + (1 - \tau_1^* - s \gamma_n(1 + o(1))) \right. \\ & \quad \left. \times \left( 2\tilde{d}_2(1 + o(m^{-1})) + \log \left( \left( \frac{n}{m} \right)^{2d_2^*} \frac{c_{0,2}^*(2\pi)^{-2d_2^*}}{1 + 2\tilde{d}_2 - 2d_2^*} (1 + m^{-1/2} Z_{3,n}^s) \right) \right) \right) \end{aligned}$$

where, for  $i = 1, 2, 3$ ,  $(Z_{i,n}^s)_n$  is a sequence of random variables depending on  $s$  satisfying  $Z_{i,n}^s \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Z_i^s$  with  $Z_i^s$  a random variable.

• if  $d_1^* > d_2^*$  we have

$$J_n(1, t_1^* + [s n \gamma_n], (\tilde{d}_1, \tilde{d}_2), m)$$

$$\begin{aligned}
 &= (\tau_1^* + s \gamma_n) \left( 2 \tilde{d}_1 + 2 d_1^* \log \left( \frac{n}{m} \right) + \log (c_{0,1}^* (2\pi)^{-2d_1^*}) - \log (1 + 2 \tilde{d}_1 - 2 d_1^*) \right) \\
 &\quad + \frac{Z_{1,n}^s}{\sqrt{m}} - s \frac{\gamma_n}{\tau_1^*} (1 + o_{\mathbb{P}}(1)) + (1 - \tau_1^* - s \gamma_n) \left( 2 \tilde{d}_2 + 2 d_2^* \log \left( \frac{n}{m} \right) \right. \\
 &\quad \left. + \log (c_{0,2}^* (2\pi)^{-2d_2^*}) - \log (1 + 2 \tilde{d}_2 - 2 d_2^*) + \frac{Z_{3,n}^s}{\sqrt{m}} \right) + o(m^{-1}).
 \end{aligned}$$

We also have:

$$\begin{aligned}
 J_n(1, t_1^*, (d_1^*, d_2^*), m) &= \tau_1^* \left( 2 d_1^* + 2 d_1^* \log \left( \frac{n}{m} \right) + \log (c_{0,1}^* (2\pi)^{-2d_1^*}) + \frac{Z'_{1,n}}{\sqrt{m}} \right) \\
 &\quad + (1 - \tau_1^*) \left( 2 d_2^* + 2 d_2^* \log \left( \frac{n}{m} \right) + \log (c_{0,2}^* (2\pi)^{-2d_2^*}) + \frac{Z'_{3,n}}{\sqrt{m}} \right) + O\left(\frac{\log(n)}{m}\right).
 \end{aligned} \tag{5.36}$$

As a consequence, with (5.35),

$$\begin{aligned}
 \Delta_n(s) &= \tau_1^* \left( 2 \tilde{d}_1 - 2 d_1^* - \log (1 + 2 \tilde{d}_1 - 2 d_1^*) + \frac{Z_{1,n}^s - Z'_{1,n}}{\sqrt{m}} \right) \\
 &\quad + (1 - \tau_1^*) \left( 2 \tilde{d}_2 - 2 d_2^* - \log (1 + 2 \tilde{d}_2 - 2 d_2^*) + \frac{Z_{3,n}^s - Z'_{3,n}}{\sqrt{m}} \right) \\
 &\quad + s \gamma_n \log \left( \frac{n}{m} \right) (2 d_1^* - 2 d_2^*) + O_{\mathbb{P}}\left(\gamma_n + \frac{\log(n)}{m}\right).
 \end{aligned} \tag{5.37}$$

Moreover, in such a framework, we have

$$\tilde{d}_1 = \arg \min_{d \in [0, 0.5]} \left\{ W_n(\{1, \dots, [n \tau_1^*] + [s n \gamma_n]\}, d, m) \right\},$$

and therefore, for  $s \geq 0$  and  $d_1^* > d_2^*$ , using the previous approximations,

$$\begin{aligned}
 \tilde{d}_1 &= \arg \min_{d \in (0, 0.5)} \left\{ \tau_1^* \left( 2 d - \log (1 + 2 d - 2 d_1^*) + \frac{Z_{1,n}^s}{\sqrt{m}} (1 + o_{\mathbb{P}}(1)) \right) \right. \\
 &\quad \left. + s \gamma_n \log \left( \frac{n}{m} \right) (2 d_1^* - 2 d_2^*) (1 + o_{\mathbb{P}}(1)) \right\},
 \end{aligned}$$

where the consistency is uniform in probability for  $d \in (0, 0.5)$ . Therefore, as the minimum of  $2 d - \log (1 + 2 d - 2 d_1^*)$  is obtained for  $d = d_1^*$  and since  $d \in (0, 1/2) \mapsto W_n(T, d, m)$  is a  $\mathcal{C}^1((0, 1/2))$  function, we deduce that

$$\tilde{d}_1 = \hat{d}_1^* + O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}} + s \gamma_n \log \left( \frac{n}{m} \right)\right). \tag{5.38}$$

We also have  $\tilde{d}_2 = \hat{d}_2^* + O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)$  from [Abadir et al. \(2007\)](#). As a consequence, with  $x - \log(1 + x) \simeq \frac{1}{2} x^2$  for  $x \rightarrow 0$ , (5.37) leads to

$$\Delta_n(s) = \left( s \gamma_n \log \left( \frac{n}{m} \right) (2 d_1^* - 2 d_2^*) + \frac{1}{\sqrt{m}} Z_n^s \right) (1 + o_{\mathbb{P}}(1)), \tag{5.39}$$

where  $(Z_n^s)$  is a sequence of random variables such as  $Z_n^s \xrightarrow[n \rightarrow \infty]{\mathcal{D}} Z^s$ . Hence, if

$$\frac{1}{\sqrt{m} \log(n/m)} = o(\gamma_n), \text{ then}$$

$$\arg \min_{s \in [0, M]} \Delta_n(s) = \arg \min_{s \in [0, M]} \left\{ s \gamma_n \log \left( \frac{n}{m} \right) (2 d_1^* - 2 d_2^*) \right\} = 0$$

with a minimum obtained for  $s = 0$ : therefore,  $\frac{1}{\sqrt{m} \log(n/m)}$  can not be negligible with respect to  $\gamma_n$ . If  $\gamma_n = \frac{1}{\sqrt{m} \log(n/m)}$ , then

$$\sqrt{m} \Delta_n(s) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} s (2 d_1^* - 2 d_2^*) + Z^s. \tag{5.40}$$

To prove that

$$\arg \min_{s \in [0, M]} \left\{ \sqrt{m} \Delta_n(s) \right\} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \arg \min_{s \in [0, M]} \left\{ s (2 d_1^* - 2 d_2^*) + Z^s \right\}, \tag{5.41}$$

we would need a Donsker-type theorem verified by  $Z_n^s$ . However we can still conclude that we necessarily have  $\gamma_n = \frac{1}{\sqrt{m} \log(n/m)}$  and therefore the convergence rate of  $\tilde{t}_1 - t_1^*$  is  $\frac{n}{\sqrt{m} \log(n/m)}$  when  $\tilde{t}_1 - t_1^* > 0$ .

• if  $d_2^* > d_1^*$ ,  $\left(\frac{n}{m}\right)^{2d_1^*}$  can be negligible with respect to  $\gamma_n \left(\min \left\{ n \gamma_n, \frac{n}{m} \right\}\right)^{2d_2^*}$  following the decreasing rate of  $(\gamma_n)_n$ . If  $\left(\frac{n}{m}\right)^{2d_1^*} = o\left(\gamma_n \left(\min \left\{ n \gamma_n, \frac{n}{m} \right\}\right)^{2d_2^*}\right)$  then using the previous decomposition:

$$\begin{aligned} \tilde{d}_1 = \arg \min_{d \in (0, 0.5)} \left\{ \tau_1^* \left( 2d - \log(1 + 2d - 2d_2^*) + \frac{Z_{2,n}^s}{\sqrt{m}} \right) \right. \\ \left. + O_{\mathbb{P}} \left( \frac{\left(\frac{n}{m}\right)^{2d_1^*}}{\gamma_n \left(\min \left\{ n \gamma_n, \frac{n}{m} \right\}\right)^{2d_2^*}} + \frac{1}{m} \right) \right\}, \end{aligned}$$

and therefore  $\tilde{d}_1 \xrightarrow[n \rightarrow \infty]{\mathcal{P}} d_2^*$ , implying

$$\begin{aligned} \Delta_n(s) = \tau_1^* \left( (2 d_2^* - 2 d_1^*) + \log \left( \gamma_n \left( \min \left\{ n \gamma_n, \frac{n}{m} \right\} \right)^{2d_2^*} \right) \right. \\ \left. + \log \left( \frac{s}{\tau_1^*} \right) + \log \left( c_{0,2}^* (2\pi)^{-2d_2^*} \right) \right) (1 + o_{\mathbb{P}}(1)). \end{aligned}$$

As a consequence the minimum in  $s$  of such  $\Delta_n(s)$  is obtained for  $s = 0$ , this indicates that  $(\gamma_n)_n$  was chosen too large.

Now consider that  $\begin{cases} \frac{1}{\sqrt{m}} \left(\frac{n}{m}\right)^{2d_1^*} &= o\left(\gamma_n \left(\min\left\{n\gamma_n, \frac{n}{m}\right\}\right)^{2d_2^*}\right) \\ o\left(\left(\frac{n}{m}\right)^{2d_1^*}\right) &= \gamma_n \left(\min\left\{n\gamma_n, \frac{n}{m}\right\}\right)^{2d_2^*} \end{cases}$ . In such a case, we obtain  $\tilde{d}_1 = d_1^* + O_{\mathbb{P}}\left(\gamma_n \left(\min\left\{n\gamma_n, \frac{n}{m}\right\}\right)^{2d_2^*} \left(\frac{n}{m}\right)^{-2d_1^*}\right)$  and therefore

$$\begin{aligned} \Delta_n(s) &= s \frac{c_{0,2}^*(2\pi)^{-2d_2^*}}{c_{0,1}^*(2\pi)^{-2d_1^*}(1+2d_1^*-2d_2^*)} \gamma_n \left(\min\left\{n\gamma_n, \frac{n}{m}\right\}\right)^{2d_2^*} \left(\frac{n}{m}\right)^{-2d_1^*} \\ &\quad + O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}} + \left(\gamma_n \left(\min\left\{n\gamma_n, \frac{n}{m}\right\}\right)^{2d_2^*} \left(\frac{n}{m}\right)^{-2d_1^*}\right)^2 + \gamma_n \log\left(\frac{n}{m}\right)\right). \end{aligned}$$

As a consequence, the minimum is once again obtain for  $s = 0$  implying we have to chose a “smaller” sequence  $(\gamma_n)_n$ .

**Case 1:** Now, if  $\underline{\beta}^* > 2d_2^* - 2d_1^*$ , we can chose  $m$  such as  $\sqrt{m} \left(\frac{n}{m}\right)^{2d_1^*-2d_2^*} \geq 1$  and  $\gamma_n = \frac{1}{\sqrt{m}} \left(\frac{n}{m}\right)^{2d_1^*-2d_2^*}$  implying  $\gamma_n \geq 1/m$ . Then for any  $s \in [0, M]$ ,

$$\sqrt{m} \Delta_n(s) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} s \frac{c_{0,2}^*(2\pi)^{2d_1^*-2d_2^*}}{c_{0,1}^*(1+2d_1^*-2d_2^*)} + Z^s, \tag{5.42}$$

with  $Z^s = \tau_1^* (Z_1^s - Z_1') + (1 - \tau_1^*) (Z_3^s - Z_3')$ . Once again we cannot use a Donsker-type theorem for obtaining the asymptotic behavior of  $\arg \min_{s \in [0, M]} \Delta_n(s)$  but we can deduce that  $\frac{n}{\sqrt{m}} \left(\frac{n}{m}\right)^{2d_1^*-2d_2^*}$  is the convergence rate of  $\tilde{t}_1 - t_1^*$  when  $\tilde{t}_1 - t_1^* \geq 0$ .

**Case 2:** If  $\underline{\beta}^* \leq 2d_2^* - 2d_1^*$ , we can chose  $m$  such as  $\sqrt{m} \left(\frac{n}{m}\right)^{2d_1^*-2d_2^*} < 1$ , and  $\gamma_n = \frac{1}{\sqrt{m}} \left(\frac{n}{m}\right)^{\frac{2d_1^*-2d_2^*}{1+2d_2^*}} m^{-\frac{d_2^*}{1+2d_2^*}}$  implying  $\gamma_n < 1/m$ . Then we also have (5.42), the same conclusion than in Case 1, except that the convergence rate of  $\tilde{t}_1 - t_1^*$  when  $\tilde{t}_1 - t_1^* \geq 0$  is  $\frac{n}{\sqrt{m}} \left(\frac{n}{m}\right)^{\frac{2d_1^*-2d_2^*}{1+2d_2^*}} m^{-\frac{d_2^*}{1+2d_2^*}}$ .

**II.** for  $-M \leq s \leq 0$ , the results can be deduced by symmetry and replacing  $d_1^*, d_2^*$  and  $\tau_1^*$  respectively by  $d_2^*, d_1^*$  and  $1 - \tau_1^*$ .  $\square$

**Remark 6.** If  $d_1^* > d_2^*$ , consider the asymptotic behavior of  $\mathbb{P}(\tilde{t}_1 > t_1^*)$ . We can write:

$$\begin{aligned} &\mathbb{P}(\tilde{t}_1 > t_1^*) \\ &= \mathbb{P}\left(\min_{u>0} J_n(1, [t_1^* + u], (\tilde{d}_1, \tilde{d}_2), m) < \min_{u<0} J_n(1, [t_1^* + u], (\tilde{d}_1, \tilde{d}_2), m)\right) \\ &= \mathbb{P}\left(\sqrt{m} \left(\min_{u>0} J_n(1, [t_1^* + u], (\tilde{d}_1, \tilde{d}_2), m) - J_n(1, t_1^*, (d_1^*, d_2^*), m)\right) \right. \\ &\quad \left. < \sqrt{m} \left(\min_{u<0} J_n(1, [t_1^* + u], (\tilde{d}_1, \tilde{d}_2), m) - J_n(1, t_1^*, (d_1^*, d_2^*), m)\right)\right). \end{aligned}$$

Using (5.40) and (5.42), we deduce that there exists  $0 < p_- < p_+ \leq 1$  depending on  $d_1^*, d_2^*, \tau_1^*, c_{0,1}^*$  and  $c_{0,2}^*$  such that for  $n$  large enough  $p_- \leq \mathbb{P}(\tilde{t}_1 > t_1^*) \leq p_+$ . However, if limit theorem such as (5.41) could be proven, we can deduce the precise limit to this probability when  $n \rightarrow \infty$ .

*Proof of Theorem 3.3.* Using Theorem 3.2, we can establish that  $\tilde{d}_i = \hat{d}_i^* + O_{\mathbb{P}}(m^{-1/2})$ . Indeed, once again without lose of generality, we can consider the case of one change. Using the notation and proof of Theorem 3.2, if we assume  $\tilde{t}_1 > t_1^*$ , knowing  $\tilde{t}_1 - t_1^* \leq C \frac{n}{\sqrt{m}}$ , then  $T_2 \subset T_2^*$  and therefore we can again write (5.33) and then  $|\tilde{d}_2 - \hat{d}_2^*| = O_{\mathbb{P}}(m^{-1/2})$ .

Concerning  $\tilde{d}_1$  and with the knowledge that  $\tilde{t}_1$  is such that  $0 \leq \tilde{t}_1 - t_1^* \leq C \frac{n}{\sqrt{m}}$ , we can write that

$$\tilde{d}_1 = \arg \min_{d \in [0, 0.5]} W_n(\{1, \dots, \tilde{t}_1\}, d, m).$$

But using computations of Theorem 3.2, we have

$$\begin{aligned} &W_n(\{1, \dots, \tilde{t}_1\}, d, m) \\ &= \log \left( \frac{t_1^*}{\tilde{t}_1} S_n(T_1^*, d, m) + \frac{\tilde{t}_1 - t_1^*}{\tilde{t}_1} S_n(\{t_1^* + 1, \dots, \tilde{t}_1\}, d, m) \right. \\ &\quad \left. + \frac{2}{\tilde{t}_1} R_n(\{1, \dots, t_1^*\}, \{t_1^* + 1, \dots, \tilde{t}_1\}, d, m) \right) + 2d \ell(m) \\ &= \log \left( S_n(T_1^*, d, m) \right) + D_{m,n,d} \left( \frac{\tilde{t}_1 - t_1^*}{t_1^*} \right) + 2d \ell(m) + \log(t_1^*/\tilde{t}_1) \\ &= W_n(T_1^*, d, m) + D_{m,n,d} \left( \frac{\tilde{t}_1 - t_1^*}{n} \right), \end{aligned}$$

where  $\sup_{d \in (0, 1/2)} |D_{m,n,d}| = O_{\mathbb{P}}(1)$  using Lemmas 5.1 and 5.2 and because we have  $t_1^* = \lfloor n\tau_1^* \rfloor$ .

Now, since  $\hat{d}_1^* = \arg \min_{d \in [0, 0.5]} W_n(T_1^*, d, m)$  and  $d \in (0, 1/2) \mapsto W_n(T, d, m)$  is a  $\mathcal{C}^1((0, 1/2))$  function, we deduce that  $\tilde{d}_1 = \hat{d}_1^* + \frac{1}{n} O_{\mathbb{P}}(|\tilde{t}_1 - t_1^*|) = \hat{d}_1^* + O_{\mathbb{P}}(m^{-1/2})$ . This achieves the proof of Theorem 3.3.  $\square$

*Proof of Theorem 3.4.* The proof is established if for any  $K \in \{0, \dots, K^* - 1, K^* + 1, \dots, K_{\max}\}$  the following consistency holds:

$$\mathbb{P} \left( J_n(K, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) < 0 \right) \xrightarrow[n \rightarrow \infty]{} 0, \tag{5.43}$$

for any  $\mathbf{t}$  and  $\mathbf{d}$ , with  $J_n$  defined as in (2.10).

Indeed, as  $J_n(K^*, \hat{\mathbf{t}}^*, \hat{\mathbf{d}}^*, m) \leq J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m)$  by definition, (5.43) is also satisfied by replacing  $J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m)$  by  $J_n(K^*, \hat{\mathbf{t}}^*, \hat{\mathbf{d}}^*, m)$ . We decompose the proof in two parts,  $K < K^*$  and  $K > K^*$ .

**Assume  $K < K^*$ .** Then, for any  $\mathbf{t}$  and  $\mathbf{d}$ , and using (5.26),

$$J_n(K, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m)$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{k=1}^{K+1} n_k \log \left( \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n_k} S_n(T_{kj}, d_k, m) \right. \\
&\quad \left. + \frac{2}{n_k} \sum_{j=1}^{K^*+1} \sum_{j'=1, j \neq j'}^{K^*+1} R_n(T_{kj}, T_{kj'}, d_k, m) \right) - \frac{1}{n} \sum_{j=1}^{K+1} n_j^* \log (S_n(T_j^*, d_j^*, m)) \\
&\quad + 2 \frac{\ell(m)}{n} \left( \sum_{k=1}^{K+1} n_k d_k - \sum_{j=1}^{K^*+1} n_j^* d_j^* \right) + (K - K^*) z_n \\
&\geq \sum_{k=1}^{K+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (s(d_j^*, d_k) - s(d_j^*, d_j^*)) - |D(m, n)| + (K - K^*) z_n \\
&\geq \sum_{k=1}^{K+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (u(d_j^*, d_k) - u(d_j^*, d_j^*)) - |D(m, n)| + (K - K^*) z_n
\end{aligned}$$

using (5.27) with  $D(m, n) \xrightarrow[n, m, n/m \rightarrow \infty]{\mathcal{P}} 0$  and since  $\sum_{j=1}^{K^*+1} n_{kj} = n_k$ ,  $\sum_{k=1}^{K+1} n_{kj} = n_j^*$  and  $u(d_j^*, d) = -\log(1 + 2d - 2d_j^*) + 2d \geq 2d_j^*$ .

Now, we use again Lemma 2.3 of Lavielle (1999, p. 88). This Lemma was obtained when  $K = K^*$  and we obtain that there exists  $C_d > 0$  such that

$$\sup_{\mathbf{d} \in \mathbf{t} \in} \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (u(d_j^*, d_k) - u(d_j^*, d_j^*)) \geq C_d \frac{1}{n} \|\mathbf{t} - \mathbf{t}^*\|_\infty$$

where  $\|\mathbf{t} - \mathbf{t}^*\|_\infty = \max_{1 \leq j \leq K^*} |t_j - t_j^*|$ . However, this result is still valide when  $K^*$  is replaced by  $K < K^*$  in the first sum, since it is sufficient to add  $K^* - K$  fictive times and consider  $t_{K+1} = t_{K+2} = \dots = t_{K^*} = t_K$  (and therefore  $n_{kj} = 0$  for  $k = K + 2, \dots, K^* + 1$ ). Therefore we obtain:

$$J_n(K, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) \geq \frac{1}{3} \min_{1 \leq i \leq K^*} |\tau_{i+1}^* - \tau_i^*| - |D(m, n)| + (K - K^*) z_n \quad (5.44)$$

since  $K < K^*$  and therefore

$$\|\mathbf{t} - \mathbf{t}^*\|_\infty \geq \frac{1}{2} \min_{1 \leq i \leq K^*} |t_{i+1}^* - t_i^*| \geq \frac{n}{3} \min_{1 \leq i \leq K^*} |\tau_{i+1}^* - \tau_i^*|$$

when  $n$  is large enough. Then, if  $z_n \xrightarrow[n \rightarrow \infty]{} 0$  then (5.43) is satisfied and this implies  $\mathbb{P}(\widehat{K} < K^*) \xrightarrow[n \rightarrow \infty]{} 0$ .

**Assume**  $K^* < K \leq K_{\max}$ . With  $\widehat{\mathbf{t}} = (\widehat{t}_1, \dots, \widehat{t}_K)$ , there exists some subset  $\{k_j, 1 \leq j \leq K^*\}$  of  $\{1, \dots, K\}$  such that for any  $j = 1, \dots, K^*$ ,  $|\frac{\widehat{t}_{k_j}}{n} - \tau_j^*| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)$ . To see this, consider the  $(\widehat{t}_{k_j})$  as the closest times among  $(\widehat{t}_1, \dots, \widehat{t}_K)$

to the  $(t_1^*, \dots, t_{K^*}^*)$ . The other  $K - K^*$  change dates  $\widehat{t}_i$  could be consider exactly as additional “false” changes (since the parameters  $d$  do not change at these times) and therefore the  $\widehat{t}_{k_j}$  minimize  $J_n(K, \mathbf{t}, \mathbf{d}, m)$  conditionally to those  $\widehat{t}_i$  with  $i \notin \{k_1, \dots, k_{K^*}\}$  as if the number of changes is known and is  $K^*$ . And therefore Theorem 3.2 holds for those  $\widehat{t}_{k_j}$ .

Then using the previous expansions detailed in the previous proofs, we obtain

$$\begin{aligned} & J_n(K, \widehat{\mathbf{t}}, \widehat{\mathbf{d}}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) \\ &= \frac{1}{n} \sum_{j=1}^{K^*+1} \left( \sum_{k=k_j+1}^{k_{j+1}} \widehat{n}_k \log(S_n(\widehat{T}_k, \widehat{d}_k, m)) - n_j^* \log(S_n(T_j^*, d_j^*, m)) \right) \\ & \quad + 2 \frac{\ell(m)}{n} \left( \sum_{i=1}^{K+1} \widehat{n}_i \widehat{d}_i - \sum_{j=1}^{K^*+1} n_j^* d_j^* \right) + (K - K^*) z_n \\ &\geq \frac{1}{n} \sum_{j=1}^{K^*+1} \left( \sum_{k=k_j+1}^{k_{j+1}} \widehat{n}_k s(d_j^*, \widehat{d}_k) - n_j^* s(d_j^*, d_j^*) \right) - |D(m, n)| \\ & \quad + 2 \frac{\ell(m)}{n} \left( \sum_{i=1}^{K+1} \widehat{n}_i \widehat{d}_i - \sum_{j=1}^{K^*+1} n_j^* d_j^* \right) + (K - K^*) z_n \\ &\geq \frac{1}{n} \sum_{j=1}^{K^*+1} \left( \sum_{k=k_j+1}^{k_{j+1}} \widehat{n}_k s(d_j^*, \widehat{d}_k) - n_j^* s(d_j^*, d_j^*) \right) - |D(m, n)| \\ & \quad + 2 \frac{\ell(m)}{n} \left( \sum_{i=1}^{K+1} \widehat{n}_i \widehat{d}_i - \sum_{j=1}^{K^*+1} n_j^* d_j^* \right) + (K - K^*) z_n \end{aligned}$$

with  $s$  defined in (5.27). Now, since  $\widehat{T}_k \subset \{\widehat{t}_{k_j+1}, \dots, \widehat{t}_{k_{j+1}}\}$ , we have from Theorem 3.4,  $\widehat{d}_k = d_j^* + O\left(\frac{1}{\sqrt{m}}\right)$ . As a consequence, for  $k = k_j + 1, \dots, k_{j+1}$  then  $s(d_j^*, \widehat{d}_k) = s(d_j^*, d_j^*) + O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)$ . Then,

$$\begin{aligned} & J_n(K, \widehat{\mathbf{t}}, \widehat{\mathbf{d}}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) \\ &\geq \frac{1}{n} \sum_{j=1}^{K^*+1} \left( s(d_j^*, d_j^*) + 2 d_j^* \frac{\ell(m)}{n} \right) \left( \sum_{k=k_j+1}^{k_{j+1}} \widehat{n}_k - n_j^* \right) \\ & \quad - |D(m, n)| - |E(m, n)| + (K - K^*) z_n \\ &\geq -|D(m, n)| - |E'(m, n)| + (K - K^*) z_n, \end{aligned}$$

with  $D(m, n) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)$  under condition  $m = o(n^{2\beta^*/(1+2\beta^*)})$  from the proof of Theorem 3.2,  $E(m, n) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)$  and therefore  $E'(m, n) = O_{\mathbb{P}}\left(\frac{1}{\sqrt{m}}\right)$  since  $\left| \sum_{k=k_j+1}^{k_{j+1}} \widehat{n}_k - n_j^* \right| = O_{\mathbb{P}}\left(\frac{n}{\sqrt{m}}\right)$ .

As a consequence if  $(z_n)$  is such that  $z_n \sqrt{m} \xrightarrow[n \rightarrow \infty]{} \infty$  then for any  $K > K^*$ ,

$$\mathbb{P}\left(J_n(K, \hat{\mathbf{t}}, \hat{\mathbf{d}}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) < 0\right) \xrightarrow[n \rightarrow \infty]{} 0.$$

This achieves the proof.  $\square$

*Proof of Corollary 2.* We can write

$$\begin{aligned} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\hat{\mathbf{t}} - \mathbf{t}^*\| \geq \delta\right) &= \sum_{K=0}^{K_{\max}} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\hat{\mathbf{t}} - \mathbf{t}^*\| \geq \delta \cap \hat{K} = K\right) \\ &= \sum_{K=0}^{K_{\max}} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\tilde{\mathbf{t}} - \mathbf{t}^*\| \geq \delta \mid \hat{K} = K\right) \mathbb{P}(\hat{K} = K). \end{aligned}$$

For  $K = K^*$  we have  $\lim_{\delta \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\tilde{\mathbf{t}} - \mathbf{t}^*\| \geq \delta\right) = 0$  and for  $K \neq K^*$ ,

$\lim_{m, n \rightarrow \infty} \mathbb{P}(\hat{K} = K) = 0$ . Then  $\lim_{\delta \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\hat{\mathbf{t}} - \mathbf{t}^*\| \geq \delta\right) = 0$  because  $K_{\max}$  is a fixed constant.

Similarly, with  $M > 0$  and for  $i = 1, \dots, K^* + 1$ ,

$$\begin{aligned} \mathbb{P}(\sqrt{m} |\hat{d}_i - d_i^*| \leq M) &= \sum_{K=0}^{K_{\max}} \mathbb{P}(\sqrt{m} |\hat{d}_i - d_i^*| \leq M \cap \hat{K} = K) \\ &= \sum_{K=0}^{K_{\max}} \mathbb{P}(\sqrt{m} |\tilde{d}_i - d_i^*| \leq M \mid \hat{K} = K) \mathbb{P}(\hat{K} = K). \end{aligned}$$

But  $\lim_{m, n \rightarrow \infty} \mathbb{P}(\hat{K} = K) = 1$  and  $\lim_{M \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P}(\sqrt{m} |\tilde{d}_i - d_i^*| \leq M) = 1$  for  $K = K^*$ , while for  $K \neq K^*$ ,  $\lim_{m, n \rightarrow \infty} \mathbb{P}(\hat{K} = K) = 0$ . Since  $K_{\max}$  is a fixed constant, this implies  $\lim_{M \rightarrow \infty} \lim_{m, n \rightarrow \infty} \mathbb{P}(\sqrt{m} |\hat{d}_i - d_i^*| \leq M) = 1$ .  $\square$

## Acknowledgements

The authors thank the Associate Editor and the Referees for their fruitful corrections, comments and suggestions, which notably improved the quality of the paper.

## References

- Abadir, K.M., Distaso, W. and Giraitis, L. (2007) Non-stationarity-extended local Whittle estimation. *J. Econometrics*, **141**, 1353–1384. [MR2413505](#)
- Arlot, S. and Massart, P. (2009) Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research*, **10**, 245–279.

- Bai J. (1994) Least squares estimation of a shift in linear processes. *J. of Time Series Anal.*, **15**, 453–472. [MR1292161](#)
- Bai J. (1995) Least absolute deviation estimation of a shift. *Econometric Theory*, **11**, 403–436. [MR1349930](#)
- Bai J. and Perron P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78. [MR1616121](#)
- Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A. and Taqqu, M.S. (2003a) Generators of long-range dependent processes: a survey. In: *Theory and Applications of Long-Range Dependence*, Birkhauser, Boston, MA, 579–623. [MR1957509](#)
- Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., Stoev, S. and Taqqu, M.S. (2003b) Semi-parametric estimation of the long-range dependence parameter: a survey. In: *Theory and Applications of Long-Range Dependence*, Birkhauser, Boston, MA, 557–577. [MR1957508](#)
- Bardet, J.-M., Kengne, W. and Wintenberger, O. (2012) Detecting multiple change-points in general causal time series using penalized quasi-likelihood. *Electronic Journal of Statistics*, **6**, 435–477. [MR2988415](#)
- Bardet, J.-M., Lang, G., Moulines, E. and Soulier, P. (2000) Wavelet estimator of long range-dependent processes. *Statist. Inference Stochast. Processes*, **3**, 85–99. [MR1819288](#)
- Basseville, M. and Nikiforov, I. (1993) *Detection of Abrupt Changes: Theory and Applications*, Prentice Hall, Englewood Cliffs, NJ, 1993. [MR1210954](#)
- Baudry, J.-P., Maugis, C. and Michel, B. (2012) Slope Heuristics: overview and implementation. *Statistics and Computing*, **22**, 455–470. [MR2865029](#)
- Beran, J. (1994) *Statistics for Long-Memory Processes*, Chapman and Hall, New York. [MR1304490](#)
- Beran, J. and Terrin, N. (1996) Testing for a change of the long-memory parameter. *Biometrika*, **83**, 627–638. [MR1423879](#)
- Bibinger, M., Jirak, M. and Vetter, M. (2017) Nonparametric change-point analysis of volatility. *Annals of Statistics*, **45**, 1542–1578. [MR3670188](#)
- Dalla, V., Giraitis, L. and Hidalgo, J. (2006) Consistent estimation of the memory parameter for nonlinear time series. *Journal of Time Series Analysis*, **27**, 211–251. [MR2235845](#)
- Doukhan, P., Oppenheim, G. and Taqqu M.S. (Editors) (2003) *Theory and Applications of Long-Range Dependence*, Birkhäuser. [MR1956041](#)
- Geweke, J. and Porter-Hudak, S. (1983) The estimation and application of long-memory time-series models. *J. Time Ser. Anal.*, **4**, 221–238. [MR0738585](#)
- Giraitis, L., Kokoszka, P. and Leipus, R. (2001) Testing for long memory in the presence of a general trend. *Journal of Applied Probability*, **38**, 1033–1054. [MR1876557](#)
- Giraitis, L., Koul, H. and Surgailis, D. (2012) *Large Sample Inference Memory Processes*, Imperial College Press. [MR2977317](#)
- Henry, M. and Robinson, P.M. (1996) Bandwidth choice in Gaussian semiparametric estimation of long-range dependence. In: *Athens Conference on Applied Probability and Time Series Analysis, Vol. II*, 220–232, Springer, New York. [MR1466748](#)

- Henry, M. (2007) Robust automatic bandwidth for long-memory. In: *Long Memory in Economics*, 157–172, Springer.
- Horvath, L. (2001) Change-point detection in long-memory processes. *Journal of Multivariate Analysis*, **78**, 218–234. [MR1859756](#)
- Inoue, A. (1997) Regularly varying correlation functions and KMO-Langevin equations. *Hokkaido Math. J.*, **26**, 457–482. [MR1463098](#)
- Inoue, A. (2000) Asymptotics for the partial autocorrelation function of a stationary process. *J. Anal. Math.*, **81**, 65–109. [MR1785278](#)
- Kay, S.M. (1998) *Fundamentals of Statistical Signal Processing*, 2, Prentice-Hall, Englewood Cliffs. [MR0130068](#)
- Kokoszka, P. and Leipus, R. (2003) Detection and estimation of changes in regime. In: P. Doukhan, G. Oppenheim, and M. S. Taqqu, editors, *Theory and Applications of Long-Range Dependence*, 325–337. [MR1957497](#)
- Künsch, H. (1987) Statistical aspects of self-similar processes. In: *Proceedings of the 1st World Congress of the Bernoulli Society*, 67–74, VNU Sci. Press, Utrecht. [MR1092336](#)
- Lavancier, F., Leipus, R., Philippe A. and Surgailis, D. (2013) Detection of non-constant long memory parameter. *Econometric Theory*, **29**, 1009–1056. [MR3148823](#)
- Lavielle, M. (1999) Detection of multiple change in a sequence of dependent variables. *Stochastic Process. Appl.*, **83**, 79–102. [MR1705601](#)
- Lavielle, M. and Ludena, C. (2000) The multiple change-points problem for the spectral distribution. *Bernoulli*, **6**, 845–869. [MR1791905](#)
- Lavielle, M. and Moulines, E. (2000) Least squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, **21**, 33–59. [MR1766173](#)
- Ray, B.K. and Tsay, R.S. (2002) Bayesian methods for change-point detection in long-range dependent processes. *Journal of Time Series Analysis*, **23**, 687–705. [MR1944606](#)
- Robinson, P.M. (1995a) Log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, **23**, 1048–1072. [MR1345214](#)
- Robinson, P.M. (1995b) Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics*, **23**, 1630–1661. [MR1370301](#)
- Robinson, P.M. and Henry, M. (2003) Higher-order kernel semiparametric M-estimation of long memory. *Journal of Econometrics*, **114**, 1–27. [MR1962371](#)
- Wang, L. and Wang, J. (2006) Testing and estimating for change in long memory parameter. *Journal of Statistical Computation and Simulation*, **76**, 317–329. [MR2201233](#)
- Yamaguchi, K. (2011) Estimating a change point in the long memory parameter. *Journal of Time Series Analysis*, **32**, 304–314. [MR2808257](#)
- Zygmund, A. (1968) *Trigonometric Series*, Cambridge University Press. [MR0236587](#)