



**HAL**  
open science

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

Jean-Marc Bardet, Abdellatif Guenaizi

► **To cite this version:**

Jean-Marc Bardet, Abdellatif Guenaizi. Data-driven semi-parametric detection of multiple changes in long-range dependent processes. *Electronic Journal of Statistics*, In press. hal-01676967v2

**HAL Id: hal-01676967**

**<https://hal.science/hal-01676967v2>**

Submitted on 29 Dec 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Jean-Marc Bardet and Abdellatif Guenaizi

bardet@univ-paris1.fr, Abdellatif.Guenaizi@malix.univ-paris1.fr

SAMM, Université Panthéon-Sorbonne (Paris I), 90 rue de Tolbiac, 75013 Paris, FRANCE

---

## Abstract

This paper is devoted to the offline multiple changes detection for long-range dependence processes. The observations are supposed to satisfy a semi-parametric long-range dependence 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. The 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.

*Keywords:* 62M10, 62M15, 62F12.

---

## 1. Introduction

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

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

where  $L$  is a positive slow varying function, satisfying for any  $c > 0$ ,  $\lim_{h \rightarrow 0} \frac{L(c|h|)}{L(|h|)} = 1$ , 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, 1987, 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.*, 2006). However there is still not satisfactory adaptive method of choice of the bandwidth parameter even if several interesting attempts have been developed (see for instance Henry and Robinson, 1998, 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. It consists on the observed trajectory  $(X_1, \dots, X_n)$  of a process  $X$  whose trajectory is partitioned into  $K^* + 1$  subtrajectories on which it is a linear long memory process whose long memory parameters are distinct from one area to another (see a more precise definition in (2.4)). Thus there is dependence between two subtrajectories 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 chose, 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-range memory, see for instance Giraitis *et al.* (2001), Kokoszka and Leipus (2003) or Lavancier *et al.* (2013).

For our offline framework, following the previous purposes, we chose to build a penalized contrast based on a sum 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-memory 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  $\mathcal{O}_P(1)$  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 chose an asymptotically consistent choice of penalization sequence, the consistency is not satisfying even for large sample such as  $n = 5000$ . 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 chose 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 very 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

- $|a_n| = c n^{d-1} + O(n^{d-1-\beta})$  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. Indeed, with  $(\varepsilon_t)_{t \in \mathbb{Z}}$  a sequence of independent and identically distributed random variables (iidrv) with zero mean and unit variance, we can define  $Y = (Y_k)_{k \in \mathbb{Z}}$  such as

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

Note that  $Y$  is a zero mean stationary process, with autocovariance  $r(k) = E(Y_0 Y_k)$  satisfying

$$r(n) = c^2 B(1 - 2d, d) n^{2d-1} + O(n^{2d-2-\beta}) \quad \text{when } n \rightarrow \infty, \quad (2.1)$$

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

$$f(\lambda) = \frac{c^2}{\pi} B(1 - 2d, d) \Gamma(2d) \sin\left(\frac{\pi}{2} - \pi d\right) |\lambda|^{-2d} + O(|\lambda|^{-2d+\beta}) \quad \text{when } \lambda \rightarrow 0, \quad (2.2)$$

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 as

$$f(\lambda) = c' |\lambda|^{-2d} + O(|\lambda|^{-2d+\beta}) \quad \text{when } \lambda \rightarrow 0, \quad (2.3)$$

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 iidrv with zero mean and unit variance. Denote also:

- $K^* \in \{0, \dots, n-1\}$ ,  $\tau_0^* = 0 < \tau_1^* < \dots < \tau_{K^*}^* < 1 = \tau_{K^*+1}^*$ ;
- $(d_i^*)_{1 \leq i \leq K^*+1} \in [0, 1/2)^{K^*+1}$ ,  $(c_i^*)_{1 \leq i \leq K^*+1} \in (0, \infty)^{K^*+1}$  and  $(\beta_i^*)_{1 \leq i \leq K^*+1} \in (0, 2)^{K^*+1}$
- $K^* + 1$  sequences  $(a_t^{(i)})$  such as  $(a_t^{(i)})_{t \in \mathbb{N}}$  belongs to the class  $L(d_i^*, \beta_i^*, c_i^*)$  for all  $i = 1, \dots, K^* + 1$ .

Define the process  $X = (X_t)_{1 \leq t \leq n}$  such as

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

$$X_t = \sum_{j=0}^{\infty} a_j^{(i)} \varepsilon_{t-j} \quad \text{when } [n\tau_{i-1}^*] + 1 \leq t \leq [n\tau_i^*]. \quad (2.4)$$

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

$$\Delta_d = \max_{1 \leq i \leq K^*} |d_{i+1}^* - d_i^*| > 0. \quad (2.5)$$

The first condition (2.4) is relative to the behavior  $(X_t)$  in each stage: it is a stationary linear long-range process with a spectral density satisfying (2.3) (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 (5.13)), which makes the model much more realistic than if the independence of successive regimes had been assumed. The second condition (2.5) 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:

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

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

- denote  $T_i = \{t_{i-1} + 1, t_{i-1} + 2, \dots, t_i\}$  and  $n_i = t_i - t_{i-1}$  for  $i = 1, \dots, K + 1$ .
- denote  $T_{ij} = \{t_{i-1} + 1, t_{i-1} + 2, \dots, t_i\} \cap \{t_{j-1}^* + 1, t_{j-1}^* + 2, \dots, t_j^*\}$  and  $n_{ij} = \#\{T_{ij}\}$  for  $i = 1, \dots, K + 1$  and  $j = 1, \dots, K^* + 1$ .

For  $K \in \{0, \dots, n\}$ , we will also use the following multidimensional notation:

- $\mathbf{d} = (d_1, \dots, d_{K+1})$  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  $|T| = \#\{T\}$ :

$$I_T(\lambda) = \frac{1}{2\pi |T|} \left| \sum_{k \in T} X_k e^{-ik\lambda} \right|^2. \quad (2.6)$$

Using the seminal papers of Kunsh (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  $m \in \{1, \dots, n\}$ ,

$$W_n(T, d, m) = \log(S_n(T, d, m)) - \frac{2d}{m} \sum_{k=1}^m \log(k/m) \quad (2.7)$$

$$\text{with } S_n(T, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} I_T(\lambda_j^{(n)}) \quad \text{and} \quad \lambda_k^{(n)} = 2\pi \frac{k}{n}. \quad (2.8)$$

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  $\hat{d} = \arg \min_{d \in [0, 0.5]} W_n(T, d, m)$  on  $T$ .

**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:

$$J_n(K, \mathbf{t}, \mathbf{d}, m) = \frac{1}{n} \sum_{k=1}^{K+1} n_k W_n(T_k, d_k, m) + K z_n, \quad (2.9)$$

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:

$$(\widehat{K}_n, \widehat{\mathbf{t}}, \widehat{\mathbf{d}}) = \underset{K \in \{0, \dots, K_{\max}\}, \mathbf{d} \in [0, 0.5]^{K+1}, \mathbf{t} \in \mathcal{T}_K(0)}{\arg \min} J_n(K, \mathbf{t}, \mathbf{d}, m), \quad (2.10)$$

with  $\widehat{\mathbf{d}} = (\widehat{d}_1, \dots, \widehat{d}_{\widehat{K}_n+1})$  and  $\widehat{\mathbf{t}} = (\widehat{t}_1, \dots, \widehat{t}_{\widehat{K}_n})$ , and where for  $a \geq 0$ ,

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

### 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  $\widetilde{\mathbf{t}} = (\widetilde{t}_1, \dots, \widetilde{t}_{K^*})$  and  $\widetilde{\mathbf{d}} = (\widetilde{d}_1, \dots, \widetilde{d}_{K^*+1})$  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:

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

Then, we can prove:

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

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

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  $\delta > 0$ ,*

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

This result provides a bound of the “best” convergence rate of  $\tilde{\mathbf{t}}$  which 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. But also the price to pay to the definition of local Whittle estimator which 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.*

Another consequence of this result is that there is asymptotically a small lose 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

$$\sqrt{m}(\hat{d}_i^* - d_i^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{1}{4}\right). \quad (3.3)$$

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 any  $i = 1, \dots, K^* + 1$ ,*

$$\sqrt{m} |\tilde{d}_i - d_i^*| = O_P(1). \quad (3.4)$$

### 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  $J_n$  is now essential. Indeed, we obtain:

**Theorem 3.4.** *Under the assumptions and notations of Theorem 3.1, if  $K_{\max} \geq K^*$ , with  $m = o(n^{2\underline{\beta}^*/(1+2\underline{\beta}^*)})$  where  $\underline{\beta}^* = \min_{1 \leq i \leq K^*+1} \beta_i^* z_n$  and  $\max\left(z_n, \frac{1}{z_n \sqrt{m}}\right) \xrightarrow[n \rightarrow \infty]{} 0$ , using  $(\hat{K}, \hat{\mathbf{t}}, \hat{\mathbf{d}})$  defined in (2.10), then*

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

Note that 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 1.** *Under the conditions of Theorem 3.4, the bounds (3.2) and (3.4) hold, i.e.,*

$$\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\hat{\mathbf{t}} - \mathbf{t}^*\| \geq \delta\right) = 0 \quad \text{and} \quad \sqrt{m} \|\hat{\mathbf{d}} - \mathbf{d}^*\| = O_P(1)$$

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: 1/ as we mainly studied the cases of FARIMA processes for which  $\beta = 2$ , we chose  $m = n^{0.65}$ ; 2/ the number  $K_{\max} \geq K^*$  is crucial for the heuristic plot procedure (see below) and was chosen such as  $K_{\max} = 2([\log(n)] - 1)$ , implying  $K_{\max} = 10, 12$  and  $14$  respectively for  $n = 500, 2000$  and  $5000$ .
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 many numerical simulations, we chose  $z_n = 2\sqrt{n}$  that offers 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 heuristic slop 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  $\hat{z}_n = 2\hat{s}$ , *i.e.*

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

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

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>.

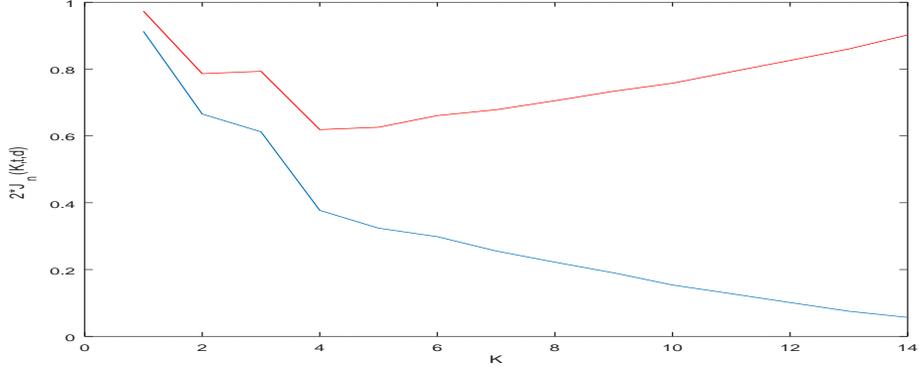


Figure 1: For  $n = 5000$ ,  $K^* = 4$  and a FARIMA(0,  $d$ , 0) process, the graph of  $2 \times J_n(K, \hat{\mathbf{t}}, \hat{\mathbf{d}}, m)$  (in blue), and the one of  $2 \times J_n(K, \hat{\mathbf{t}}, \hat{\mathbf{d}}, m) + 2 \times \hat{s} \times K$  (in red).

#### 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

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

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  $(\tau_1^*, \tau_2^*, \tau_3^*) = (0.25, 0.5, 0.75)$ .

Table 1: RMSE of the estimators from 500 independent replications of processes, when the number  $K^*$  of changes is known.

	$n$	FARIMA(0, $d$ , 0)			FARIMA(1, $d$ , 1)			$X^{(d,1)}$		
		500	2000	5000	500	2000	5000	500	2000	5000
$K^* = 0$	$\tilde{d}_1$ ( $d_1 = 0.4$ )	0.070	0.047	0.034	0.098	0.090	0.066	0.077	0.048	0.035
	$\tilde{d}_1$ ( $d_1 = 0.1$ )	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

- Each case is independently replicated 500 times and the RMSE, Root-Mean-Square Error, is computed for each estimator of the parameter.

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

#### 4.3. 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 ‘‘Heuristic Slope’’ procedure described previously.

We obtained the results detailed in Table 2:

Table 2: 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
$K^* = 1$	$\widehat{K}_n$	0.11	0.21	0.51	0.21	0.45	0.67	0.05	0.05	0.01
	$\widehat{K}_{BIC}$	0	0	0	0	0	0	0	0	0
	$\widehat{K}_H$	0.35	0.91	0.92	0.49	0.77	0.81	0.25	0.47	0.57
$K^* = 3$	$\widehat{K}_n$	0.13	0.12	0.32	0.12	0.21	0.52	0.16	0.07	0.02
	$\widehat{K}_{BIC}$	0	0	0	0	0	0	0	0	0
	$\widehat{K}_H$	0.02	0.16	0.85	0.03	0.21	0.80	0.07	0.16	0.32

#### 4.4. Conclusions of Monte-Carlo experiments

From Tables 1 and 2, we may conclude that:

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 slightly impacted by the 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).
3. 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 consistencies are slightly better for small  $K^*$  than for large  $K^*$ . 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  $n^{0.65}$  for obtaining consistent estimators of long memory parameters.

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

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

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 (2.8), for any  $i \in \{1, \dots, K^* + 1\}$  and  $T \subset T_i^*$ ,*

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

*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:

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

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:

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

where  $p_{|T|,1}(m) = 2\pi \sum_{j=1}^m I_\varepsilon(\lambda_j^{(n)})$ ,  $p_{|T|,2}(m) = \sum_{j=1}^m (\eta_j - 2\pi I_\varepsilon(\lambda_j^{(n)}))$  and  $R_{|T|}(m) = \sum_{j=1}^m (\eta_j^* - \eta_j)$  with

$$\eta_j = \frac{I_T(\lambda_j^{(n)})}{f(\lambda_j^{(n)})} \text{ and } I_\varepsilon(\lambda_j^{(n)}) = \frac{1}{2\pi|T|} \left| \sum_{t=1}^{|T|} \varepsilon_t e^{it\lambda_j^{(n)}} \right|^2.$$

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:

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

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

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

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:

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

and therefore (5.3) is established.

Now a straightforward application of Markov Inequality and Lemma 2 in Dalla *et al.* (2006) implies that

for any  $d \in [0, 1/2)$ ,

$$\begin{aligned} & \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \eta_j^* - \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \right| = O_P\left(\left(\frac{m}{n}\right)^{\beta_i^*} + \left(\frac{n}{m|T|}\right)^{1/2}\right) \\ \implies & \left| \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \eta_j^* - \frac{1}{2d-2d_i^*+1} \right| = O_P\left(\left(\frac{m}{n}\right)^{\beta_i^*} + \left(\frac{n}{m|T|}\right)^{1/2} + m^{2d-2d_i^*-1}\right). \end{aligned} \quad (5.7)$$

Since  $S_n(T, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} I_T(\lambda_j^{(n)}) = \frac{(2\pi)^{2d_i^*}}{c_{0,i}^*} \left(\frac{n}{m}\right)^{2d_i^*} \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d-2d_i^*} \eta_j^*$ , we deduce that for any  $N \geq 1$ ,

$$\sup_{|T| \geq N} \left| S_n(T, d, m) - \left(\frac{n}{m}\right)^{2d_i^*} \frac{c_{0,i}^* (2\pi)^{-2d_i^*}}{1+2d-2d_i^*} \right| = \left(\frac{n}{m}\right)^{2d_i^*} O_P\left(\left(\frac{m}{n}\right)^{\beta_i^*} + \left(\frac{n}{mN}\right)^{1/2} + m^{2d-2d_i^*-1}\right). \quad (5.8)$$

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

$$P\left(\sup_{d \in [0, 1/2)} \left| S_n(T, d, m) - \frac{c_{0,i}^* (2\pi)^{-2d_i^*}}{1+2d-2d_i^*} |T|^{2d_i^*} \right| \geq \delta\right) \leq \frac{C_i}{\delta} |T|^{2d_i^*}. \quad (5.9)$$

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

In the sequel, we define:

$$R_n(T, T', d, m) = \frac{1}{2\pi} \sum_{t \in T} \sum_{t' \in T'} X_t X_{t'} b_n(t' - t, d, m) \quad \text{with} \quad b_n(k, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} e^{-2\pi i \frac{jk}{n}}. \quad (5.10)$$

Note that  $S_n(T, d, m)$ , which is defined in (2.8) can also be written as:

$$S_n(T, d, m) = \frac{1}{m} \sum_{j=1}^m \left(\frac{j}{m}\right)^{2d} I_T(\lambda_j^{(n)}) = \frac{1}{2\pi |T|} \sum_{s \in T} \sum_{t \in T} X_s X_t b_n(t - s, d, m). \quad (5.11)$$

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  $N \in \mathbb{N}^*$ ,*

$$\sup_{d \in [0, 1/2)} \max_{\min(|T|, |T'|) \geq N} \left( \min(|T|, |T'|, \frac{n}{m}) \right)^{-d_j^* - d_{j'}^*} \frac{1}{\min(|T|, |T'|)} |R_n(T, T', d, m)| = O_P\left(\left(\min\left(1, \frac{n}{mN}\right)\right)^{1-d_j^* - d_{j'}^*}\right). \quad (5.12)$$

*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  $t < t'$ ,

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

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 as

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

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

$$\begin{aligned} |\Gamma_{T, T'}(|t' - t|)| &\leq C' \sum_{k=1}^{\infty} k^{d_j^*-1} (t' - t + k)^{d_{j'}^*-1} \\ &\leq \frac{C'}{(t' - t)^{1-d_j^*-d_{j'}^*}} \times \frac{1}{t' - t} \sum_{k=1}^{\infty} \left( \frac{k}{t' - t} \right)^{d_j^*-1} \left( 1 + \frac{k}{t' - t} \right)^{d_{j'}^*-1} \\ &\leq \left( C'' \int_0^{\infty} \frac{1}{x^{1-d_j^*}} \frac{1}{(1+x)^{1-d_{j'}^*}} dx \right) \frac{1}{(t' - t)^{1-d_j^*-d_{j'}^*}}. \end{aligned} \quad (5.13)$$

Now, using (5.10) and (5.13), we have:

$$\begin{aligned} \mathbb{E}(R_n(T, T', d, m)) &= \frac{1}{2\pi} \sum_{t \in T} \sum_{t' \in T'} \text{Cov}(X_t, X_{t'}) b_n(t' - t, d, m) \\ \implies |\mathbb{E}(R_n(T, T', d, m))| &= \frac{1}{2\pi} \sum_{t \in T} \sum_{t' \in T'} \Gamma_{T, T'}(t' - t) b_n(t' - t, d, m). \end{aligned}$$

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:

$$\begin{aligned} \mathbb{E}(R_n(T, T', d, m)) &= \frac{1}{2\pi} \left( \sum_{k=1}^{\mu} k \Gamma_{T, T'}(\delta + k) b_n(\delta + k, d, m) + \mu \sum_{k=\mu+1}^{\nu} \Gamma_{T, T'}(\delta + k) b_n(\delta + k, d, m) \right. \\ &\quad \left. + \sum_{k=\nu+1}^{\nu+\mu} (\nu + \mu - k) \Gamma_{T, T'}(\delta + k) b_n(\delta + k, d, m) \right). \end{aligned}$$

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

$$|b_n(u, d, m)| \leq C(d) \min \left\{ 1, \left( \frac{n}{m} \right)^{1+2d} |u|^{-1-2d} \right\} \quad \text{for } u \in \mathbb{Z}. \quad (5.14)$$

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

$$\begin{aligned} |\mathbb{E}(R_n(T, T', d, m))| &\leq C \left( \sum_{k=1}^{\mu} k^{d_j^* + d_{j'}^*} + 2\mu \sum_{k=\mu+1}^{\nu+\mu} k^{-1+d_j^* + d_{j'}^*} \right) \\ &\leq C \mu \nu^{d_j^* + d_{j'}^*}. \end{aligned} \quad (5.15)$$

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

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

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

$$\begin{aligned} |\mathbb{E}(R_n(T, T', d, m))| &\leq C \mu \left( \min \left( \frac{n}{m}, \nu \right) \right)^{d_j^* + d_{j'}^*} \left( \min \left( 1, \frac{n}{m\mu} \right) \right)^{1+\min(0, 2d-d_j^* - d_{j'}^*)} (\log(\mu))^{1_{2d=d_j^* + d_{j'}^*}} \\ \implies \sup_{d \in [0, 1/2)} |\mathbb{E}(R_n(T, T', d, m))| &\leq C \mu \left( \min \left( \frac{n}{m}, \nu \right) \right)^{d_j^* + d_{j'}^*} \left( \min \left( 1, \frac{n}{m\mu} \right) \right)^{1-d_j^* - d_{j'}^*}. \end{aligned} \quad (5.16)$$

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:

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

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,

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

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

Using the Cauchy-Schwarz Inequality, we have

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

Now we apply the same trick as in (5.13) 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,

$$\begin{aligned} |\text{Cov}(X_t X_{t'}, X_s X_{s'})| &\leq C \left( (s-t+1)^{2d_j^*-3/2} (t'-t+1)^{2d_{j'}^*-3/2} \right. \\ &\quad \left. + (s-t+1)^{2d_j^*-1} (s'-t'+1)^{2d_{j'}^*-1} + ((s'-t)(t'-s))^{d_j^*+d_{j'}^*-1} \right). \end{aligned} \quad (5.17)$$

$$\begin{aligned} \text{Var}(R_n(T, T', d, m)) &\leq C \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} |\text{Cov}(X_t X_{t'}, X_s X_{s'}) b_n(t-t', d, m) b_n(s-s', d, m)| \\ &\leq C (J_1 + J_2 + J_3), \end{aligned} \quad (5.18)$$

$$\text{with } \begin{cases} J_1 = \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} (|t-s|+1)^{2d_j^*-3/2} (|t'-t|+1)^{2d_{j'}^*-3/2} |b_n(t-t', d, m) b_n(s-s', d, m)| \\ J_2 = \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} (1+|t-s|)^{2d_j^*-1} (1+|t'-s'|)^{2d_{j'}^*-1} |b_n(t-t', d, m) b_n(s-s', d, m)| \\ J_3 = \sum_{t \in T} \sum_{s \in T} \sum_{t' \in T'} \sum_{s' \in T'} ((s'-t)(t'-s))^{d_j^*+d_{j'}^*-1} |b_n(t-t', d, m) b_n(s-s', d, m)| \end{cases} .$$

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,

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

As a consequence, using (5.14),

$$\sum_{u=0}^{|T|+|T'|} |b_n(u, d, m)|^2 \leq C \left( \sum_{u=0}^{\min(|T|+|T'|, n/m)} 1 + \sum_{u=\min(|T|+|T'|, n/m)}^{|T|+|T'|} (n/m)^{1+2d} |u|^{-1-2d} \right) \leq C(d) \min(|T| + |T'|, \frac{n}{m}). \quad (5.19)$$

Moreover,

$$\begin{aligned} &\sum_{u=0}^{|T|+|T'|} |b_n(u, d, m)| \sum_{v=u+1}^{|T|+|T'|} |b_n(v, d, m)| |u - v|^{2d_{j'}^* - 1} \\ &\leq C \left\{ \sum_{u=0}^{\min(|T|+|T'|, n/m)} \left( \sum_{v=u+1}^{\min(|T|+|T'|, n/m)} (v - u)^{2d_{j'}^* - 1} + \left(\frac{n}{m}\right)^{1+2d} \sum_{v=\min(|T|+|T'|, n/m)}^{|T|+|T'|} v^{-1-2d} (v - u)^{2d_{j'}^* - 1} \right) \right. \\ &\quad \left. + \left(\frac{n}{m}\right)^{2+4d} \sum_{u=\min(|T|+|T'|, n/m)}^{|T|+|T'|} \sum_{v=u+1}^{|T|+|T'|} (uv)^{-1-2d} (v - u)^{2d_{j'}^* - 1} \right\} \\ &\leq C(d) \left( \min(|T| + |T'|, \frac{n}{m}) \right)^{1+2d_{j'}^*} \end{aligned} \quad (5.20)$$

after classical computations. From (5.19) and (5.20), we obtain:

$$J_2 \leq C(d) |T|^{2d_j^* + 1} \left( \min(|T| + |T'|, \frac{n}{m}) \right)^{1+2d_{j'}^*}. \quad (5.21)$$

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:

$$J_2 \leq C(d) \min\left\{ \mu^{2d_j^* + 1} \left( \min\left(\nu, \frac{n}{m}\right) \right)^{1+2d_{j'}^*}, \mu^{2d_{j'}^* + 1} \left( \min\left(\nu, \frac{n}{m}\right) \right)^{1+2d_j^*} \right\}. \quad (5.22)$$

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

$$J_3 \leq C \mu^2 \left( \min\left(\frac{n}{m}, \nu\right) \right)^{2d_j^* + 2d_{j'}^*} \left( \min\left(1, \frac{n}{m\mu}\right) \right)^{2-2d_j^* - 2d_{j'}^*}. \quad (5.23)$$

As a consequence, using (5.32), (5.33), (5.35) and (5.22), (5.23), we obtain that there exists  $C > 0$  such as:

$$\sup_{d \in [0, 1/2]} \text{Var}(R_n(T, T', d, m)) \leq C \mu^2 \left( \min\left(\frac{n}{m}, \nu\right) \right)^{2d_j^* + 2d_{j'}^*} \left( \min\left(1, \frac{n}{m\mu}\right) \right)^{2-2d_j^* - 2d_{j'}^*}. \quad (5.24)$$

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

$$\sup_{d \in [0, 1/2]} \max_{\min(|T|, |T'|) \geq N} \left( \min(|T|, |T'|, \frac{n}{m}) \right)^{-d_j^* - d_{j'}^*} \frac{1}{\min(|T|, |T'|)^2} E(R_n^2(T, T', d, m)) \leq C \left( \min(1, \frac{n}{mN}) \right)^{2 - 2d_j^* - 2d_{j'}^*} \quad (5.25)$$

with  $C > 0$  that achieves the proof of (5.12) 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

$$U_n(\mathbf{t}, \mathbf{d}, m) = J_n(K^*, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m), \quad (5.26)$$

where  $J_n$  is defined in (2.9). Then, using (5.11), 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:

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

with  $R_n$  defined in (5.10). As a consequence,

$$\begin{aligned} U_n(\mathbf{d}, \mathbf{t}, m) &= \frac{1}{n} \sum_{k=1}^{K^*+1} \left[ n_k \log \left( \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n_k} \left( S_n(T_{kj}, d_k, m) + \sum_{j'=1, j \neq j'}^{K^*+1} \frac{2}{n_{kj}} R_n(T_{kj}, T_{kj'}, d_k, m) \right) \right) \right. \\ &\quad \left. - n_k^* \log(S_n(T_k^*, d_k^*, m)) \right] + \frac{\ell(m)}{n} \sum_{k=1}^{K^*+1} 2(n_k d_k - n_k^* d_k^*) \\ &\geq \frac{1}{n} \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} n_{kj} \log \left( S_n(T_{kj}, d_k, m) + \sum_{j'=1, j \neq j'}^{K^*+1} \frac{2}{n_{kj}} R_n(T_{kj}, T_{kj'}, d_k, m) \right) - n_k^* \log(S_n(T_k^*, d_k^*, m)) \\ &\quad + \frac{\ell(m)}{n} \sum_{k=1}^{K^*+1} 2(n_k d_k - n_k^* d_k^*) \\ &\geq \frac{1}{n} \sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} n_{kj} \left[ \log \left( S_n(T_{kj}, d_k, m) + \sum_{j'=1, j \neq j'}^{K^*+1} \frac{2}{n_{kj}} R_n(T_{kj}, T_{kj'}, d_k, m) \right) + 2 d_k \ell(m) \right] \\ &\quad - \frac{1}{n} \sum_{k=1}^{K^*+1} n_k^* \left( \log(S_n(T_k^*, d_k^*, m)) + 2 d_k^* \ell(m) \right) \end{aligned} \quad (5.27)$$

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:

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

with  $\varepsilon_{kj} = O_P(1)$  when  $n_{kj} = O(n/m)$  and  $\varepsilon_{kj} = o_P(1)$  for  $n = o(n_{kj}m)$ . As a consequence, from (5.27), Lemma 5.1 and 5.2, we deduce that there exists a random variable  $D(m, n)$  such as  $D(m, n) \xrightarrow[n, m, n/m \rightarrow \infty]{\mathcal{P}} 0$  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  $d \in [0, 1/2)$ ,

$$s(d_j^*, d) = 2d_j^* \log(n/m) + \log(c_{0,j}^* (2\pi)^{-2d_j^*}) - \log(1 + 2d - 2d_j^*) - 2d. \quad (5.28)$$

Now, simple computations also imply

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

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 as

$$\sum_{k=1}^{K^*+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (u(d_j^*, d_k) - u(d_j^*, d_j^*)) \geq \frac{C}{n} \|\mathbf{t} - \mathbf{t}^*\|_\infty, \quad (5.30)$$

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  $\delta > 0$ ,

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

since for  $\mathbf{t} \in \mathcal{T}_{K^*}(n\delta)$  we have  $\|\mathbf{t} - \mathbf{t}^*\|_\infty \geq \delta n$  and for any  $k \in \{1, \dots, K^*\}$ . 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  $(u_n)_n$  a sequence of real numbers satisfying  $u_n \sqrt{m}/n \xrightarrow{n \rightarrow \infty} \infty$ ,  $u_n/n \xrightarrow{n \rightarrow \infty} 0$  and  $\mathbb{P}(|\tilde{t}_1 - t_1^*| > u_n) \xrightarrow{n \rightarrow \infty} 0$ . 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  $\mathbb{P}(\delta n/\sqrt{m} < |\tilde{t}_1 - t_1^*| \leq u_n) \xrightarrow{n \rightarrow \infty} 0$ .

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

$$\mathbb{P}\left(\delta \frac{n}{\sqrt{m}} < |\tilde{t}_1 - t_1^*| \leq u_n\right) \leq \mathbb{P}\left(\min_{t_1 \in V_{\delta, n, m}} (J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m)) \leq 0\right), \quad (5.31)$$

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

Let  $t_1 \in V_{\delta, n, m}$  and with no loss of generality chose  $t_1 > t_1^*$ . Then  $n_1 = t_1$ ,  $n_2 = n - t_1$ ,  $n_{11} = t_1^*$ ,  $n_{12} = t_1 - t_1^*$ ,  $n_{21} = 0$  and  $n_{22} = n - t_1$ . Then  $T_1^* = \{1, \dots, t_1^*\}$ ,  $T_2^* = \{t_1^* + 1, \dots, n\}$ ,  $T_1 = \{1, \dots, t_1\}$ ,  $T_{11} = T_1^* = \{1, \dots, t_1^*\}$ ,  $T_{12} = \{t_1^* + 1, \dots, t_1\}$ ,  $T_2 = \{t_1 + 1, \dots, n\} = T_{22}$ .

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  $\frac{\frac{1}{t_1 - t_1^*} R_n(T_{11}, T_{12}, \tilde{d}_1, m)}{S_n(T_1^*, \tilde{d}_1, m)} = \left(\frac{n}{m}\right)^{1-2d_1^*} O_P\left(\frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}}\right)$ . Therefore, using again the concavity of the logarithm function, we have:

$$\begin{aligned} J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) &= \frac{1}{n} \left\{ t_1 \log \left( \frac{t_1^*}{t_1} S_n(T_1^*, \tilde{d}_1, m) + \frac{t_1 - t_1^*}{t_1} S_n(T_{12}, \tilde{d}_1, m) + \frac{2}{t_1} R_n(T_{11}, T_{12}, \tilde{d}_1, m) \right) \right. \\ &\quad \left. + (n - t_1) \log (S_n(T_{22}, \tilde{d}_2, m)) + 2 \ell(m) (t_1 \tilde{d}_1 + (n - t_1) \tilde{d}_2) \right\} \\ &\geq \frac{1}{n} \left\{ t_1^* \log (S_n(T_1^*, \tilde{d}_1, m)) + (t_1 - t_1^*) \log (S_n(T_{12}, \tilde{d}_1, m)) + (n - t_1) \log (S_n(T_{22}, \tilde{d}_2, m)) \right. \\ &\quad \left. + 2 \ell(m) (t_1 \tilde{d}_1 + (n - t_1) \tilde{d}_2) + \left(\frac{n}{m}\right)^{1-2d_1^*} O_P\left(\frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}}\right) \right\} \\ &\geq \frac{1}{n} \left\{ t_1^* W_n(T_1^*, \tilde{d}_1, m) + (t_1 - t_1^*) W_n(T_{12}, \tilde{d}_1, m) + (n - t_1) W_n(T_{22}, \tilde{d}_2, m) \right. \\ &\quad \left. + \left(\frac{n}{m}\right)^{1-2d_1^*} O_P\left(\frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}}\right) \right\}. \end{aligned}$$

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,

$$\begin{aligned} & \frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} \\ & \geq \frac{1}{t_1 - t_1^*} \left\{ (t_1 - t_1^*) W_n(T_{12}, \tilde{d}_1, m) + (n - t_1) W_n(T_{22}, \tilde{d}_2, m) \right. \\ & \quad \left. - (n - t_1^*) W_n(T_2^*, \hat{d}_2^*, m) + \left( \frac{n}{m} \right)^{1-2\hat{d}_1^*} O_P \left( \frac{1}{(t_1 - t_1^*)^{1-d_1^*-d_2^*}} \right) \right\}. \end{aligned} \quad (5.32)$$

Since  $t_1 \in V_{\delta, n, m}$ , implying  $|T_2^*|/n \xrightarrow[m, n/m \rightarrow \infty]{\mathcal{P}} (1 - \tau_1^*)$  and  $|T_{22}|/n \xrightarrow[m, n/m \rightarrow \infty]{\mathcal{P}} (1 - \tau_1^*)$ , Lemma 5.1 and more precisely inequality (5.8) can be applied. Then, conditionally to  $\tilde{d}_1, \tilde{d}_2$  and  $\hat{d}_2^*$ , we obtain:

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

since  $\ell(m) = \frac{1}{m} \sum_{j=1}^m \log(j/m) = -1 + O(m^{-1})$  which is negligible with respect to  $O_P(m^{-1/2})$ . Therefore, (5.32) becomes:

$$\begin{aligned} & \frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} \\ & \geq \frac{1}{t_1 - t_1^*} \left\{ - (t_1 - t_1^*) (\log(1 + 2\tilde{d}_1 + 2d_2^*) - 2\tilde{d}_1) - (n - t_1) (\log(1 + 2\tilde{d}_2 - 2d_2^*) + 2\tilde{d}_2) \right. \\ & \quad \left. + \frac{n}{m} O_P \left( \frac{m^{\beta_2^*+1}}{n^{\beta_2^*}} + m^{1/2} + m^{2\tilde{d}_2 - 2d_2^*} + m^{2\hat{d}_2^* - 2d_2^*} + \frac{n}{t_1 - t_1^*} m^{2\hat{d}_1^* - 2d_2^*} + \frac{n^{-2\hat{d}_1^*}}{m^{-2\hat{d}_1^*} (t_1 - t_1^*)^{1-d_1^*-d_2^*}} \right) \right. \\ & \quad \left. + (n - t_1^*) (\log(1 + 2\hat{d}_2^* - 2d_2^*) + 2\hat{d}_2^*) \right\}. \end{aligned} \quad (5.33)$$

$t_1$  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 as  $m = o(n^{2\beta_2^*/(1+2\beta_2^*)})$ , then:

$$\tilde{d}_2 = d_2^* + O_P(m^{-1/2}) \quad \text{and} \quad \hat{d}_2^* = d_2^* + O_P(m^{-1/2}). \quad (5.34)$$

Then, from (5.33), we obtain after computations,

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

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

$$\frac{n}{t_1 - t_1^*} \left\{ J_n(K^*, t_1, (\tilde{d}_1, \tilde{d}_2), m) - J_n(K^*, t_1^*, (\hat{d}_1^*, \hat{d}_2^*), m) \right\} \geq 2(d_1^* - d_2^*) - \log(1 + 2(d_1^* - d_2^*)) + O_P\left(\frac{1}{\delta}\right). \quad (5.35)$$

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

*Proof of Theorem 3.3.* Using Theorem 3.2, we can establish that  $\tilde{d}_i = \hat{d}_i^* + O_P(m^{-1/2})$ . Indeed, once again without loss 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 (5.34) and then  $|\tilde{d}_2 - \hat{d}_2^*| = O_P(m^{-1/2})$ .

Concerning  $\tilde{d}_1$  and with the knowledge that  $\tilde{t}_1$  is such as  $0 \leq \tilde{t}_1 - t_1^* \leq C \frac{n}{\sqrt{m}}$ , 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  $\sup_{d \in [0, 1/2]} |D_{m,n,d}| = O_P(1)$  using Lemmas 5.1 and 5.2 and because we have  $t_1^* = \lceil n\tau_1^* \rceil$ . 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  $\tilde{d}_1 = \hat{d}_1^* + \frac{1}{n} O_P(|\tilde{t}_1 - t_1^*|) = \hat{d}_1^* + O_P(m^{-1/2})$ . This achieves the proof of Theorem 3.3.  $\square$

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

$$\mathbb{P} \left( J_n(K, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) < 0 \right) \xrightarrow[n \rightarrow \infty]{} 0, \quad (5.36)$$

for any  $\mathbf{t}$  and  $\mathbf{d}$ , with  $J_n$  defined as in (2.9). Indeed, as  $J_n(K^*, \hat{\mathbf{t}}^*, \hat{\mathbf{d}}^*, m) \leq J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m)$  by definition, (5.36) 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 (5.27),

$$\begin{aligned}
& J_n(K, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) \\
&= \frac{1}{n} \sum_{k=1}^{K+1} n_k \log \left( \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n_k} S_n(T_{kj}, d_k, m) + \frac{2}{n_k} \sum_{j=1}^{K^*+1} \sum_{j'=1, j' \neq j}^{K^*+1} R_n(T_{kj}, T_{kj'}, d_k, m) \right) \\
&\quad - \frac{1}{n} \sum_{j=1}^{K+1} n_j^* \log (S_n(T_j^*, d_j^*, m)) + 2 \frac{\ell(m)}{n} \left( \sum_{k=1}^{K+1} n_k d_k - \sum_{j=1}^{K^*+1} n_j^* d_j^* \right) + (K - K^*) z_n \\
&\geq \sum_{k=1}^{K+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (s(d_j^*, d_k) - s(d_j^*, d_j^*)) - |D(m, n)| + (K - K^*) z_n \\
&\geq \sum_{k=1}^{K+1} \sum_{j=1}^{K^*+1} \frac{n_{kj}}{n} (u(d_j^*, d_k) - u(d_j^*, d_j^*)) - |D(m, n)| + (K - K^*) z_n
\end{aligned}$$

since  $\sum_{j=1}^{K^*+1} n_{kj} = n_k$  and  $\sum_{k=1}^{K+1} n_{kj} = n_j^*$  and using (5.28) with  $D(m, n) \xrightarrow[n, m, n/m \rightarrow \infty]{\mathcal{P}} 0$  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 exist  $C_d > 0$  such as

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

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:

$$J_n(K, \mathbf{t}, \mathbf{d}, m) - J_n(K^*, \mathbf{t}^*, \mathbf{d}^*, m) \geq \frac{1}{3} \min_{1 \leq i \leq K^*} |\tau_{i+1}^* - \tau_i^*| - |D(m, n)| + (K - K^*) z_n \quad (5.37)$$

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. Therefore, if  $z_n \xrightarrow[n \rightarrow \infty]{} 0$  then (5.36) is satisfied and therefore  $\mathbb{P}(\widehat{K} < K^*) \xrightarrow[n \rightarrow \infty]{} 0$ .

**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  $j = 1, \dots, K^*$ ,  $|\frac{\widehat{t}_{k_j}}{n} - \tau_j^*| = O_P\left(\frac{1}{\sqrt{m}}\right)$ . 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 (5.28). 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  $s(d_j^*, \widehat{d}_k) = s(d_j^*, d_j^*) + O_P\left(\frac{1}{\sqrt{m}}\right)$ . Then,

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

with  $D(m, n) = O_P\left(\frac{1}{\sqrt{m}}\right)$  under condition  $m = o(n^{2\beta_-^*/(1+2\beta_-^*)})$  from the proof of Theorem 3.2,  $E(m, n) = O_P\left(\frac{1}{\sqrt{m}}\right)$  and therefore  $E'(m, n) = O_P\left(\frac{1}{\sqrt{m}}\right)$  since  $\left| \sum_{k=k_j+1}^{k_{j+1}} \widehat{n}_k - n_j^* \right| = O_P\left(\frac{n}{\sqrt{m}}\right)$ . As a consequence if  $(z_n)$  is such that  $z_n \sqrt{m} \xrightarrow[n \rightarrow \infty]{} \infty$  then for any  $K > K^*$ ,

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

This achieves the proof. □

*Proof of Corollary 1.* The results are easily obtained by considering conditional probability with respect to the event  $\widehat{K} = K^*$ . □

## Acknowledgement

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

## 6. References

- [1] Abadir, K.M., Distaso, W. and Giraitis, L. (2007) Non-stationarity-extended local Whittle estimation. *J. Econometrics*, 141, 1353-1384.
- [2] Arlot, S. and Massart, P. (2009) Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research*, 10, 245-279.
- [3] Bai J. (1998) Least squares estimation of a shift in linear processes. *J. of Time Series Anal.*, 5, 453-472.
- [4] Bai J. and Perron P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, 66, 47-78.
- [5] Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A. and Taqqu, M.S. (2003a) Generators of long-range dependent processes: a survey. *Theory and applications of long-range dependence*, Birkhauser, Boston, MA, 579-623.
- [6] 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 *Theory and applications of long-range dependence*, Birkhauser, Boston, MA, 557-577.
- [7] 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.
- [8] 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.
- [9] Basseville, M. and Nikiforov, I. (1993). *Detection of Abrupt Changes: Theory and Applications*. Prentice Hall, Englewood Cliffs, NJ, 1993.
- [10] Baudry, J.-P., Maugis, C. and Michel, B. (2012) Slope Heuristics: overview and implementation. *Statistics and Computing*, 22, 455-470.
- [11] Beran, J. (1994) *Statistics for Long-Memory Processes*. Chapman and Hall, New York.
- [12] 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.

- [13] Doukhan, P., Oppenheim, G. and Taqqu M.S. (Editors) (2003) *Theory and applications of long-range dependence*, Birkhäuser.
- [14] Geweke, J. and Porter-Hudak, S. (1983), The estimation and application of long-memory time-series models, *J. Time Ser. Anal.*, 4, 221-238.
- [15] 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.
- [16] Giraitis, L., Koul, H. and Surgailis, D. (2012) *Large Sample Inference Memory Processes*, Imperial College Press.
- [17] 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.
- [18] Henry, M. (2007) Robust automatic bandwidth for long-memory. In: Long Memory in Economics, 157-172, Springer.
- [19] Inoue, A. (1997). Regularly varying correlation functions and KMO-Langevin equations. *Hokkaido Math. J.*, 26, 457-482.
- [20] Inoue, A. (2000). Asymptotics for the partial autocorrelation function of a stationary process. *J. Anal. Math.*, 81, 65-109.
- [21] Kay, S.M. (1998). *Fundamentals of Statistical Signal Processing*, 2, Prentice-Hall, Englewood Cliffs.
- [22] 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.
- [23] Künsch, H. (1987). Statistical aspects of self-similar processes. Proceedings of the 1st World Congress of the Bernoulli Society, 67–74, VNU Sci. Press, Utrecht.
- [24] Lavancier, F., Leipus, R., Philippe A. and Surgailis, D. (2013). Detection of non-constant long memory parameter. *Econometric Theory*, 29, 1009-1056.
- [25] Lavielle, M. (1999) Detection of multiple change in a sequence of dependent variables. *Stochastic Process. Appl.*, 83, 79-102.
- [26] Lavielle, M. and Ludena, C. (2000). The multiple change-points problem for the spectral distribution. *Bernoulli*, 6, 845-869.
- [27] 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.

- [28] Robinson, P.M. (1995a). Log-periodogram regression of time series with long-range dependence. *Annals of Statistics*, 23, 1048-1072.
- [29] Robinson, P.M. (1995b). Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics*, 23, 1630-1661.
- [30] Robinson, P.M. and Henry, M. (2003). Higher-order kernel semiparametric M-estimation of long memory. *Journal of Econometrics*, 114, 1-27.
- [31] Zygmund, A. (1968). *Trigonometric series*. Cambridge University Press.