

The multiple change-points problem for the spectral distribution

MARC LAVIELLE¹ and CARENNE LUDEÑA²

¹*Equipe de Probabilités, Statistique et Modélisation, Université Paris Sud, Bât. 425, 91400 Orsay, France. E-mail: marc.lavielle@math.u-psud.fr*

²*Departamento de Matemáticas, Instituto Venezolano de Investigaciones Científicas, Apartado 21827, Caracas 1020-A, Venezuela. E-mail: cludena@cauchy.ivic.ve*

We consider the problem of detecting an unknown number of change-points in the spectrum of a second-order stationary random process. To reach this goal, some maximal inequalities for quadratic forms are first given under very weak assumptions. In a parametric framework, and when the number of changes is known, the change-point instants and the parameter vector are estimated using the Whittle pseudo-likelihood of the observations. A penalized minimum contrast estimate is proposed when the number of changes is unknown. The statistical properties of these estimates hold for strongly mixing and also long-range dependent processes. Estimation in a nonparametric framework is also considered, by using the spectral measure function. We conclude with an application to electroencephalogram analysis.

Keywords: detection of change-points; long range dependence; maximal inequality; nonparametric spectral estimation; penalized minimum contrast estimate; quadratic forms; Whittle likelihood

1. Introduction

Estimating change-points in the spectrum of a time series has been widely studied as a particular topic of the general change-point problem: see, for example, Brodsky and Darkhovsky (1993) or Basseville and Nikiforov (1993) for a thorough survey.

In Picard (1985) and later on in a series of articles by Giraitis and Leipus (1990; 1992), a posteriori detection of a change in the spectral distribution function is considered, based on a Kolmogorov–Smirnov type test statistic. Functional limit theorems under the null hypothesis of no changes are given under very general conditions. Giraitis and Leipus (1992) also consider the behaviour of the statistics under local alternatives. These results are based on a functional limit theorem for quadratic forms under L^2 type conditions. Brodsky and Darkhovsky (1993) consider additionally the problem of multiple changes and construct consistent estimators for the number of changes for strongly mixing sequences, under certain conditions on larger-order cumulants.

For strongly dependent processes the changes in the spectrum may affect the long-range dependence parameter. Depending on the value of this parameter, limiting processes can

change – see, for example, Fox and Taqqu (1997) and Terrin and Taqqu (1991) – and thus the distribution of the test statistic.

Following the papers of Lavielle and Moulines (1999) and Lavielle (1999), on detecting change-points in the marginal distribution of a sequence of (possibly dependent) observations, we study in this paper the detection of multiple change-points in a piecewise stationary centred time series which may exhibit long-range dependence. This kind of problem arises, for example, in network data where it is reasonable to assume that the dependence structure of the data changes in time (Taqqu *et al.*, 1995).

We assume that the distribution of the process depends on a parameter θ that changes abruptly at some unknown instants $(t_j^*, j \geq 1)$. To obtain asymptotical results, we consider here the case where the sequence (t_j^*) depends on the volume of the observations. If the number of segments K^* is known, change-points are found together with the sequence of parameters $(\theta_j^*, 1 \leq j \leq K^*)$ by minimizing a certain contrast function constructed on the basis of the Whittle pseudo-likelihood of the observations. This approximation of the likelihood implicitly asserts that we are assuming that observations in different time segments are asymptotically uncorrelated. Furthermore, we also require that the fourth-order cumulants decay fast enough.

Assume there exist $0 < \tau_1 < \dots < \tau_{K^*-1} < 1$, such that $t_j^* = [n\tau_j^*]$. Then the length of each segment tends to infinity at the same rate as the total number of observations ($t_{j+1}^* - t_j^* = O(n)$). It is shown, under very weak conditions, that the minimum contrast estimator of the normalized change-points sequence $(\hat{t}_{n,1}/n, \dots, \hat{t}_{n,K^*-1}/n)$ converges to the true normalized sequence $(\tau_1^*, \dots, \tau_{K^*-1}^*)$. The estimated vector of parameters also converges to the true vector of parameters $(\theta_j^*, 1 \leq j \leq K^*)$. Furthermore, it is shown that the errors of location $\hat{t}_{n,j} - t_j^* = O_p(1)$, for a wide class of (weakly or strongly) dependent processes.

When the number of changes is unknown, it is estimated by minimizing a penalized contrast function. The penalization term has the form $\beta_n K$, where K is the number of segments, that is, the number of parameters in the model, and where (β_n) is a positive sequence decreasing to 0. Penalized estimation has been considered, for example, when estimating the order of an ARMA process (see Akaike 1974; or Hannan 1980), or the order of a mixture of populations (see Dacunha-Castelle and Gassiat 1997). Then this problem of change-points detection can be seen as a problem of model selection via penalization, as in Lavielle (1999). We show that the estimated number of change-points converges to the true number of change-points if β_n goes to 0 at an appropriate rate that depends on the covariance structure of the process.

This kind of method also extends to semi-parametric type tests. Indeed, if we consider that the spectral distribution function changes abruptly in certain given frequency bands, we can build a contrast function by using the integrated periodogram. An application to electroencephalogram (EEG) analysis illustrates the ability of the proposed method to detect changes in the electrical activity of the brain.

In Section 2 we present some useful results dealing with quadratic forms, based on certain maximal inequalities for sums of (possibly dependent) random variables due to Móricz *et al.* (1982). Sections 3 and 4 are respectively devoted to the parametric and nonparametric contexts.

2. Some preliminary results for quadratic forms

In order to give asymptotic results concerning the estimated change-points, we must control uniformly in probability the quadratic forms which appear in the empirical contrasts. Here, uniformity is both with respect to the parameter space and the change-point instants. A first result, given in Lemma 2.2, deals with uniform inequalities in probability over all the possible change-point configurations. Uniform results with respect to the parameter space are then established in Lemma 2.3.

2.1. Some maximal inequalities for quadratic forms

Let $(X_t, t \in \mathbb{N})$ be a real second-order stationary process with zero mean and autocovariance function γ . Let $(b(u), u \in \mathbb{Z})$ be a bounded sequence of real numbers such that $b(u) = b(-u)$, and such that $\sum_{\mathbb{Z}} b(u)\gamma(u) < \infty$.

We consider here the quadratic form $Q_{1:m}$ defined by

$$Q_{1:n} = \sum_{t=1}^n \sum_{s=1}^n b(t-s)X_t X_s - n \sum_{u=-\infty}^{+\infty} b(u)\gamma(u). \quad (1)$$

Then, for any $0 \leq m \leq n$, we have

$$Q_{1:n} = Q_{1:m} + Q_{m+1:m} + 2R_{m,n},$$

where

$$R_{m,n} = \sum_{t=1}^m \sum_{s=m+1}^n b(t-s)X_t X_s.$$

In this section, we are interested in controlling $Q_{1:n}$ and $R_{m,n}$ in probability. To do this, we assume the following hypotheses are satisfied:

Hypothesis 1.

- (i) There exist a $D > 0$ and a $\beta \geq 1$ such that, for any $u \geq 0$, $|b(u)| \leq Du^{-\beta}$.
- (ii) There exist a $C > 0$ and an $\alpha > 0$ such that, for any $u \geq 0$, $|\gamma(u)| \leq Cu^{-\alpha}$.
- (iii) There exists a $\kappa > 0$ such that, for any $(t, s, t', s') \in \mathbb{Z}^4$,

$$|\text{cum}(X_t, X_s, X_{t'}, X_{s'})| \leq \kappa |t-s|^{-\alpha} |t'-s'|^{-\alpha}.$$

We have the following results:

Lemma 2.1. Under Hypothesis 1, for any $\varepsilon > 0$, let

$$h = \begin{cases} 4 - 2\alpha - 2\beta & \text{if } \alpha + \beta < \frac{3}{2}, \\ 1 + \varepsilon & \text{if } \alpha + \beta = \frac{3}{2}, \\ 1 & \text{if } \alpha + \beta > \frac{3}{2}; \end{cases} \quad h' = \begin{cases} 2 - \alpha & \text{if } \alpha < 1, \\ 1 + \varepsilon & \text{if } \alpha = 1, \\ 1 & \text{if } \alpha > 1. \end{cases}$$

Then, there exists a $C_1 > 0$ and a $D_1 > 0$ such that, for all $n \geq 1$, for all $0 < m \leq n$,

$$EQ_{1:n}^2 \leq C_1 n^h, \quad (2)$$

$$ER_{m,n}^2 \leq D_1 m^{h'}. \quad (3)$$

Remark. In the case $\alpha + \beta = \frac{3}{2}$ (for h) or $\alpha = 1$ (for h'), the bounds in (2) and (3) can be sharpened to obtain $C_1 n \log n$ and $D_1 m \log m$. However, this improved result will not change the rates given in Lemma 2.2 below, and therefore the rates of convergence of the estimated change-points given in Section 3.

Proof. Let \mathcal{C} denote any non-negative constant. Assume that $m \leq n/2$. Under Hypothesis 1(i) and 1(ii), we have:

$$\begin{aligned} EQ_{1:n} &= \sum_{t=1}^n \sum_{s=1}^n \gamma(t-s)b(t-s) \\ &\leq \mathcal{C} n^{(2-\alpha-\beta)\vee 1}, \\ ER_{m,n} &= \sum_{t=1}^m \sum_{s=m+1}^n \gamma(t-s)b(t-s) \\ &= \sum_{u=1}^m u\gamma(u)b(u) + m \sum_{u=m+1}^{n-m} \gamma(u)b(u) + \sum_{u=n-m+1}^n (n-u)\gamma(u)b(u) \\ &\leq \mathcal{C} m^{(2-\alpha-\beta)\vee 0}. \end{aligned}$$

On the other hand, under Hypothesis 1(iii) and following, for example, Terrin and Taqqu (1991),

$$\begin{aligned} \text{var } Q_{1:n} &= \sum_{t=1}^n \sum_{s=1}^n \sum_{t'=1}^n \sum_{s'=1}^n (\text{cum}(X_t, X_s, X_{t'}, X_{s'}) + 2\gamma(t-t')\gamma(s-s'))b(t-s)b(t'-s') \\ &\leq \mathcal{C} \sum_{t=1}^n \sum_{s=1}^n \sum_{t'=1}^n \sum_{s'=1}^n |t-t'|^{-\alpha} |s-s'|^{-\alpha} |b(t-s)b(t'-s')| \\ &\leq \mathcal{C} n^h. \end{aligned}$$

We show (3) in the same way. Indeed, we have

$$\begin{aligned} \text{var } R_{m,n} &\leq \mathcal{C} \sum_{t=1}^m \sum_{s=m+1}^n \sum_{t'=1}^m \sum_{s'=m+1}^n |t-t'|^{-\alpha} |s-s'|^{-\alpha} |b(t-s)b(t'-s')| \\ &\leq \mathcal{C} \sum_{t=1}^m \sum_{t'=1}^m |t-t'|^{-\alpha} \sum_{s=m+1}^n \sum_{s'=m+1}^n |s-s'|^{-\alpha} |b(t-s)b(t'-s')|. \end{aligned}$$

Note that

$$\sum_{t=1}^m \sum_{t'=1}^m |t - t'|^{-\alpha} \leq \mathcal{C} m^{h'}.$$

On the other hand,

$$\begin{aligned} \sum_{s=m+1}^n \sum_{s'=m+1}^n |s - s'|^{-\alpha} |b(t-s)b(t'-s')| &\leq \sum_{u=1}^n b(u)^2 + 2 \sum_{u=1}^n |b(u)| \sum_{v=u+1}^n |b(v)| |u - v|^{-\alpha} \\ &\leq \mathcal{C} \left(1 + 2 \sum_{u=1}^n u^{-\beta} u^{1-\beta-\alpha} \right) < \infty, \end{aligned}$$

because $1 - 2\beta - \alpha < -1$. This yields the required bound. \square

Using Lemma 2.1, we can now obtain some useful maximal inequalities:

Lemma 2.2. *Under Hypothesis 1, for any $\varepsilon > 0$, let $a = \max(4 - 2\alpha - 2\beta, 1 + \varepsilon)$ and $a' = \max(2 - \alpha, 1 + \varepsilon)$. Then, there exist $C_2 > 0$, $C_3 > 0$, $D_2 > 0$ and $D_3 > 0$ such that, for all $n > 0$, for all $m > 0$, and for all $\delta > 0$, the following inequalities are satisfied:*

$$\mathbb{P} \left(\max_{1 \leq k \leq n} |Q_{1:k}| > \delta \right) \leq C_2 \frac{n^a}{\delta^2}, \tag{4}$$

$$\mathbb{P} \left(\max_{k \geq m} \frac{|Q_{1:k}|}{k} > \delta \right) \leq C_3 \frac{m^{a-2}}{\delta^2}, \tag{5}$$

$$\mathbb{P} \left(\max_{1 \leq k < n} \max_{k+1 \leq l \leq n} |R_{k,l}| > \delta \right) \leq D_2 \frac{n^a}{\delta^2}, \tag{6}$$

$$\mathbb{P} \left(\max_{k \geq n} \frac{|R_{k,n}|}{k} > \delta \right) \leq D_3 \frac{m^{a'-2}}{\delta^2}. \tag{7}$$

Proof. First we show that, under the hypothesis of Lemma 2.2, there exists a constant $A > 0$ such that, for any $\delta > 0$ and for any positive and decreasing sequence $d_1 \geq d_2 \geq \dots \geq d_n > 0$,

$$\mathbb{P} \left(\max_{1 \leq k \leq n} d_k |Q_{1:k}| > \delta \right) \leq C_2 \frac{n^{a-1}}{\delta^2} \sum_{t=1}^n d_t^2.$$

This result is an extension of the following Hájek–Rényi type inequality for partial sums (see Levielle and Moulines 1999):

Theorem 2.3. *Let $(X_{t,t} \in \mathbb{N})$ be a sequence of zero-mean random variables. Assume that there exist $A_1 > 0$ and $1 < a < 2$ such that, for all $1 \leq i \leq j$, $\mathbb{E}(\sum_{t=i}^j X_t)^2 \leq A_1 |j - i + 1|^a$. Then, there exists a constant $A_2 \geq 1$ such that, for any $n \geq 1$, for any $\delta > 0$, and for any positive and decreasing sequence $d_1 \geq d_2 \geq \dots \geq d_n > 0$, we have the inequality*

$$P\left(\max_{1 \leq k \leq n} d_k \left| \sum_{t=1}^k X_t \right| > \delta\right) \leq A_2 \frac{n^{a-1}}{\delta^2} \sum_{t=1}^n d_t^2.$$

For any $1 \leq r \leq k \leq n$, we can write $Q_{1:k} = Q_{1:r} + Q_{r+1:k} + 2R_{r,k}$. Using Lemma 2.1,

$$\begin{aligned} ER_{r,k}^2 &\leq E(Q_{1:k}^2 + Q_{1:r}^2 + Q_{r+1:k}^2) \\ &\leq 2C_1 k^a. \end{aligned}$$

On the other hand, $R_{r,k}$ can be expressed as a sum: $R_{r,k} = \sum_{t=1}^k \eta_{r,t}$ where $\eta_{r,1} = \eta_{r,2} = \dots = \eta_{r,r} = 0$ and $\eta_{r,t} = \sum_{s=1}^r X_t X_s b(t-s)$ for $t > r$. Thus, the maximal inequality proposed in Theorem 2.3 holds for $R_{r,k}$: there exists $D > 0$ such that

$$P\left(\max_{1 \leq k \leq n} d_k |R_{r,k}| > \delta\right) \leq D \frac{n^{a-1}}{\delta^2} \sum_{t=1}^n d_t^2.$$

Following the proof advanced by Móricz *et al.* (1982), we have

$$\begin{aligned} P\left(\max_{1 \leq k \leq n} d_k |Q_{1:k}| > \delta\right) &\leq P\left(\max_{1 \leq k \leq r-1} d_k |Q_{1:k}| > \delta\right) + P(d_r |Q_{1:r}| > p_1 \delta) \\ &\quad + P\left(\max_{r+1 \leq k \leq n} d_k |Q_{r+1:k}| > p_2 \delta\right) + P\left(\max_{r+1 \leq k \leq n} d_k |R_{r,k}| > p_3 \delta\right), \end{aligned}$$

where p_1, p_2 and p_3 are positive numbers such that $p_1 + p_2 + p_3 = 1$. By induction, we have

$$\begin{aligned} P\left(\max_{1 \leq k \leq n} d_k |Q_{1:k}| > \delta\right) &\leq \frac{C_2}{\delta^2} \left((r-1)^{a-1} \sum_{t=1}^{r-1} d_t^2 + \frac{(n-r)^{a-1}}{p_2^2} \sum_{t=r+1}^n d_t^2 + \frac{C_1 r^a}{C_2 p_1^2} d_r^2 + \frac{D n^{a-1}}{C_2 p_3^2} \sum_{t=1}^n d_t^2 \right). \end{aligned}$$

Choose r such that $(r-1)^{a-1} \sum_{t=1}^{r-1} d_t^2 \leq n^{a-1} \sum_{t=1}^n d_t^2 / 2^a$ and $(n-r)^{a-1} \sum_{t=r+1}^n d_t^2 \leq n^{a-1} \sum_{t=1}^n d_t^2 / 2^a$. Then, since (d_t) is a decreasing sequence, $r^a d_r^2 \leq r^{a-1} \sum_{t=1}^r d_t^2$ and

$$P\left(\max_{1 \leq k \leq N} d_k |Q_{1:k}| > \delta\right) \leq \frac{C_2}{\delta^2} \left(\frac{1}{2^a} \left(1 + \frac{1}{p_2^2} \right) + \frac{C_1}{C_2 p_1^2} + \frac{D}{C_2 p_3^2} \right) n^{a-1} \sum_{t=1}^n d_t^2.$$

Now choose p_2 close enough to 1 and C_2 large enough, in order to ensure that

$$\frac{1}{2^a} \left(1 + \frac{1}{p_2^2} \right) + \frac{C_1}{C_2 p_1^2} + \frac{D}{C_2 p_3^2} \leq 1$$

(note that this is always possible, since $a > 1$).

We obtain (4) with $d_k = 1$ for any $k \geq 0$. Setting $d_k = 1/k$, we obtain

$$P\left(\max_{1 \leq k \leq n} \frac{|Q_{1:k}|}{k} > \delta\right) \leq C_2 \frac{n^{a-1}}{\delta^2} \sum_{t=1}^n \frac{1}{t^2}. \tag{8}$$

Then (5) is a direct consequence of (8), since

$$\begin{aligned} \mathbb{P}\left(\max_{k \geq m} \frac{|Q_{1:k}|}{k} > \delta\right) &\leq \sum_{p=0}^{\infty} \mathbb{P}\left(\max_{2^p m \leq k < 2^{p+1} m} \frac{|Q_{1:k}|}{k} > \delta\right) \\ &\leq \frac{C_2 m^{a-2}}{\delta^2(1-2^{a-2})}. \end{aligned}$$

Finally, (6) and (7) are shown by using the same kind of arguments as for (4) and (5). \square

2.2. Uniform convergence of parametric quadratic forms

Assume now that the function b depends on a parameter θ that belongs to a compact subset Θ of \mathbb{R}^d . For any $\theta \in \Theta$, assume that $(b(u, \theta), u \in \mathbb{N})$ is a bounded sequence of real numbers such that $b(u, \theta) = b(-u, \theta)$ and $\sum_{\mathbb{Z}} b(u, \theta)\gamma(u) < \infty$.

For any $\theta \in \mathbb{R}$, for any $n > 0$ and any $0 \leq m \leq n$, set

$$\begin{aligned} Q_{1:n}(\theta) &= \sum_{t=1}^n \sum_{s=1}^n b(t-s, \theta) X_t X_s - n \sum_{k=-\infty}^{+\infty} b(u, \theta)\gamma(u), \\ R_{m,n}(\theta) &= \sum_{t=1}^m \sum_{s=m+1}^n b(t-s, \theta) X_t X_s. \end{aligned}$$

We assume that the following hypothesis is satisfied:

Hypothesis 2.

- (i) For any $\theta \in \Theta$, Hypothesis 1 is satisfied.
- (ii) The function b is continuously differentiable with respect to θ (on the interior of Θ). Furthermore, there exists a $D' > 0$ such that, for any $u \geq 0$, for any $\theta \in \Theta$, $|\partial_{\theta} b(u, \theta)| \leq D' u^{-1}$.

Lemma 2.4. Under Hypothesis 2, for all $\delta > 0$,

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{\theta \in \Theta} \max_{1 \leq k \leq n} \frac{|Q_{1:k}(\theta)|}{n} > \delta\right) = 0, \tag{9}$$

$$\lim_{m \rightarrow \infty} \mathbb{P}\left(\sup_{\theta \in \Theta} \max_{k \geq m} \frac{|Q_{1:k}(\theta)|}{k} > \delta\right) = 0, \tag{10}$$

$$\lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{\theta \in \Theta} \max_{1 \leq k < n} \max_{k+1 \leq l \leq n} \frac{|R_{k,l}(\theta)|}{n} > \delta\right) = 0, \tag{11}$$

$$\lim_{m \rightarrow \infty} \mathbb{P}\left(\sup_{\theta \in \Theta} \max_{k \geq m} \frac{|R_{k,n}(\theta)|}{k} > \delta\right) = 0, \tag{12}$$

Proof. Equation (9) follows directly from the following two remarks:

(i) By using Lemma 2.2, for any θ in Θ , for any $\delta > 0$, we have that, under Hypothesis 2,

$$\lim_{n \rightarrow \infty} P\left(\max_{1 \leq k < n} \frac{|Q_{1:k}(\theta)|}{n} > \delta\right) = 0.$$

(ii) For any $\varepsilon > 0$, let $\omega(k, \varepsilon) = \sup_{\|\theta - \theta'\| \leq \varepsilon} n^{-1} |Q_{1:k}(\theta) - Q_{1:k}(\theta')|$, where $\|\cdot\|$ is the Euclidian norm. Then, there exist two sequences $\xi(v)$ and $\varepsilon(v)$, $\xi(v) \rightarrow 0$ and $\varepsilon(v) \rightarrow 0$ when $v \rightarrow \infty$, such that $\overline{\lim} P(\max_{1 \leq k < n} \omega(k, \varepsilon(v)) > \xi(v)) = 0$. Indeed, let $\tilde{Q}_{1:n}$ be the quadratic form defined in (1) with $b(u) = u^{-1}$. Under Hypothesis 2 there exists $C > 0$ such that

$$\omega(k, \varepsilon) \leq C \left\{ \frac{|\tilde{Q}_{1:k}|}{n} + 2\varepsilon \sum_{u=-\infty}^{+\infty} \frac{|\gamma(u)|}{u} \right\}.$$

By Lemma 2.1 we have $n^{-1} \max_{1 \leq k \leq n} |\tilde{Q}_{1:k}|$ converges to 0. Now use Hypothesis 1 to conclude that $\sum |\gamma(u)|/u$ converges.

We show (10), (11) and (12) in exactly the same way. □

3. Parametric estimation in the presence of change-points

3.1. Description of the model

Assume that the spectrum of the zero-mean process $X = (X_t, t \in \mathbb{Z})$ depends on a parameter θ which changes abruptly at some unknown instants $t_1^*, t_2^*, \dots, t_{K^*-1}^*$ with $0 < t_1^* < \dots < t_{K^*-1}^* < n$. These changes affect the covariance of process X , and we furthermore assume the process is asymptotically uncorrelated before and after a change. More precisely, we assume the following hypothesis is satisfied:

Hypothesis 3.

(i) There exists a vector $(\theta_1^*, \dots, \theta_{K^*}^*)$, such that θ_j^* takes its values in the interior of a compact subset Θ of \mathbb{R}^d for any $1 \leq j \leq K^*$, and such that

$$EX_s X_t = \gamma(t - s, \theta_j^*) \quad \text{if } t_{j-1}^* + 1 \leq s \leq t \leq t_j^*.$$

(ii) There exists a $G > 0$ and a $\gamma > 0$ such that

$$EX_s X_t \leq G|t - s|^{-\gamma} \quad \text{if there exists } 1 \leq j \leq K^* - 1 \text{ such that } s \leq t_j^* < t.$$

(iii) There exists a $\kappa > 0$ such that for any $(t, s, t', s') \in \mathbb{Z}^4$,

$$|\text{cum}(X_t, X_s, X_{t'}, X_{s'})| \leq \kappa |E(X_t X_s) E(X_{t'} X_{s'})|.$$

As an example, consider the AR(1) process X_t which is the solution to the following equation:

$$X_t = \mathfrak{G}^*(t)X_{t-1} + \varepsilon_t$$

where (ε_t) is a white noise, and where \mathfrak{G}^* is a time-varying piecewise constant function that takes its values in $(-1, +1)$. This example is a particular case of the more general problem of estimating the function \mathfrak{G}^* which has been considered by Dahlhaus (1997), assuming that \mathfrak{G}^* is smooth enough to guarantee that the process is *locally stationary*.

We define the configuration of normalized change-points by $\boldsymbol{\tau}^* = (\tau_1^*, \dots, \tau_{K^*-1}^*)$, where $t_j^* = [n\tau_j^*]$, and $0 < \tau_1 < \dots < \tau_{K^*-1} < 1$. The problem consists in estimating the vector of parameters $\boldsymbol{\theta}^*$ and estimating the configuration of normalized change-points $\boldsymbol{\tau}^*$ from n observations X_1, X_2, \dots, X_n . The aim of this section is to study the behaviour of an estimator of $(\boldsymbol{\tau}^*, \boldsymbol{\theta}^*)$, as $n \rightarrow \infty$.

We use index j for the true configuration of change-points $\boldsymbol{t}^* = (t_j^*, 1 \leq j \leq K^* - 1)$ and for the true sequence of parameters $\boldsymbol{\theta}^* = (\theta_j^*, 1 \leq j \leq K^*)$. We use index k for any other configuration. We denote by T_j^* the set of indices that belong to segment j in the configuration \boldsymbol{t}^* , and by T_k the set of indices that belong to segment k in the configuration \boldsymbol{t} :

$$T_j^* = \{t \in \mathbb{Z}, t_{j-1}^* + 1 \leq t \leq t_j^*\},$$

$$T_k = \{t \in \mathbb{Z}, t_{k-1} + 1 \leq t \leq t_k\}.$$

The lengths of T_j^* and T_k are respectively n_j^* and n_k . Denote by $T_{kj} = T_j^* \cap T_k$ the vector of indices that belong to segment j in the configuration \boldsymbol{t}^* and to segment k in the configuration \boldsymbol{t} . The length of T_{kj} is n_{kj} .

The dependence with respect to n of t_j^*, t_k, T_j^*, T_k and T_{kj} is omitted for notational convenience. For the same reason, the dependence of T_k and T_{kj} with respect to the configuration of change-points \boldsymbol{t} is also omitted.

Let $f(\lambda, \theta_j^*)$ be the spectral density of X in segment j of the true configuration $\boldsymbol{t}^* = (t_1^*, \dots, t_{K^*-1}^*)$. For any $\eta > 0$, let $\mathcal{L}(\eta)$ be the set of Lipschitz functions of order η over $(-\pi, \pi]$. That is, $g \in \mathcal{L}(\eta)$ if there exists a K such that $\sup_{\lambda \in (-\pi, \pi]} |g(\lambda + \mu) - g(\lambda)| \leq K|\mu|^\eta$.

We shall assume that the following hypothesis on f and γ is satisfied:

Hypothesis 4.

- (i) The parametrization is proper: if $\theta \neq \theta'$ then $f(\lambda, \theta) \neq f(\lambda, \theta')$ over a set of positive Lebesgue measure.
- (ii) Θ is compact and, for any $1 \leq j \leq K^*$, $\theta_j^* \in \mathring{\Theta}$, the interior of $\Theta \subseteq \mathbb{R}^d$. Furthermore, $\theta_j^* \neq \theta_{j+1}^*$ for any $1 \leq j \leq K^* - 1$.
- (iii) $1/f(\lambda, \theta)$ is three times continuously differentiable with respect to θ .
- (iv) For any $\theta \in \Theta$, there exists $\eta(\theta)$ such that, for all $1 \leq i, l \leq d$, and all $\varepsilon > 0$,

$$\frac{1}{f(\cdot, \theta)} \in \mathcal{L}(\eta(\theta)), \quad \frac{\partial}{\partial \theta_i} \left(\frac{1}{f(\cdot, \theta)} \right) \in \mathcal{L}(\eta(\theta) - \varepsilon), \quad \frac{\partial^2}{\partial \theta_i \partial \theta_l} \left(\frac{1}{f(\cdot, \theta)} \right) \in \mathcal{L}(\eta(\theta) - \varepsilon),$$

- (v) There exists a function $\alpha: \Theta \rightarrow (0, +\infty)$ such that, for any $u > 0$ and any $1 \leq j \leq K^*$,

$$|\gamma(u, \theta_j^*)| = O(u^{-\alpha(\theta_j^*)}).$$

A couple of comments are in order. Concerning Hypothesis 4(ii), it might sometimes be useful for practical purposes to assume that there exists a minimum jump size, independent of the true configuration θ^* . That is, there exists a Δ such that

$$\inf_{1 \leq j \leq K^*-1} \|\theta_{j+1}^* - \theta_j^*\| \geq \Delta.$$

However, as the number of change-points is fixed, there always exists a constant C^* , which bounds this minimum jump size.

Turning to Hypothesis 4(iv), call $b(k, \theta)$ the k th Fourier coefficient of $1/f(\cdot, \theta)$. Following this notation, call $\partial_{\theta_i} b(k, \theta)$ and $\partial_{\theta_i}^2 b(k, \theta)$ the k th Fourier coefficients of $\partial_{\theta_i} 1/f(\cdot, \theta)$ and $\partial_{\theta_i}^2 1/f(\cdot, \theta)$, respectively. Then Hypothesis 4(iv) yields that for all $\varepsilon > 0$, $b(u, \theta) = O(u^{-1-\eta(\theta)})$, $\partial_{\theta_i} b(u, \theta) = O(u^{-1-\eta(\theta)+\varepsilon})$ and $\partial_{\theta_i}^2 b(u, \theta) = O(u^{-1-\eta(\theta)+\varepsilon})$.

Hypothesis 4 is satisfied for a wide class of second-order stationary processes. In particular, it is satisfied if X is an ARMA process. It is also satisfied if X is a piecewise strongly dependent process according to the following definition:

Definition 1. We say that $Y = (Y_t, t \geq 0)$ is a strongly dependent process if the following hold:

- (i) Y is a centred linear process: there exists an independently and identically distributed sequence of centred random variables ζ with fourth-order cumulant $\kappa_4 < \infty$ and a sequence (a_t) of $\ell^2(\mathbb{Z})$, such that formally $Y_t = \sum_{s \in \mathbb{Z}} a_s \zeta_{t-s}$.
- (ii) The spectral density $f(\lambda, \theta^*)$ of Y , $\theta^* \in \Theta$, belongs to a parametric family such that $f(\lambda, \theta)$ and $1/f(\lambda, \theta)$ can be written as

$$f(\lambda, \theta) = C_1(\lambda, \theta)L_1(\lambda, \theta)|\lambda|^{\alpha(\theta)-1}$$

$$\frac{1}{f(\lambda, \theta)} = C_2(\lambda, \theta)L_2(\lambda, \theta)|\lambda|^{1-\alpha(\theta)}$$

where $\alpha: \Theta \rightarrow (0, 1)$ is three times continuously differentiable, C_1 and C_2 are strictly positive and three times continuously differentiable with respect to (λ, θ) , and L_1 and L_2 are strictly positive and slowly varying functions which are three times continuously differentiable with respect to (λ, θ) , for $\lambda \neq 0$.

In particular, we have that (see Fox and Taqqu 1987) the covariance function of Y , under Definition 1, has the form

$$\gamma(s, \theta^*) = EY_t Y_{t+s} = O(C(\theta^*)|s|^{-\alpha(\theta^*)}L_1(s, \theta^*)).$$

Furthermore, if $a_s = O(|s| - (1 + \alpha(\theta^*))/2)$, then Hypothesis 3(iii) is satisfied.

3.2. Estimation when the number of change-points is known

3.2.1. The estimator

For a given number of segments K , let \mathcal{T}_K be the set of configurations of change-points and Θ_K the space of the parameters,

$$\begin{aligned} \mathcal{T}_K &= \{ \mathbf{t} = (t_0, t_1, \dots, t_K), t_0 = 0 < t_1 < t_2 < \dots < t_K = n \}, \\ \Theta_K &= \{ \boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_K), \theta_k \in \Theta \}, \end{aligned}$$

where Θ was introduced in Section 3.1.

Let $t \in \mathcal{T}_K$ be any configuration of the $K - 1$ change-points, and let $\boldsymbol{\theta} \in \Theta_K$ be any vector of K parameters. Then, for any $1 \leq k \leq K$, let

$$I_n(T_k, \lambda) = \frac{1}{n_k} \left| \sum_{t \in T_k} X_t e^{-ei\lambda} \right|^2 \tag{13}$$

be the periodogram computed over the window T_k , and

$$W_n(T_k, \theta_k) = \int_{(-\pi, \pi]} \left(\log f(\lambda, \theta_k) + \frac{I_n(T_k, \lambda)}{f(\lambda, \theta_k)} \right) d\lambda \tag{14}$$

be minus the Whittle log-likelihood of $(X_t, t \in T_k)$ evaluated at θ_k .

For a given configuration \mathbf{t} of \mathcal{T}_K , let $\hat{\theta}_n(T_k)$ be the value of $\theta \in \Theta$ that minimizes $W_n(T_k, \theta)$, that is, $\hat{\theta}_n(T_k)$ is the Whittle estimate of θ_k^* computed over the k th segment of \mathbf{t} . Define $w_\vartheta(\theta)$ by

$$w_\vartheta(\theta) = \int_{(-\pi, \pi]} \left(\log f(\lambda, \theta) + \frac{f(\lambda, \vartheta)}{f(\lambda, \theta)} \right) d\lambda. \tag{15}$$

Under Hypothesis 4(i), we have that θ_j^* is the value of $\theta \in \Theta$ that minimizes $w_{\theta_j^*}(\theta)$. For any $\varepsilon > 0$ and $1 \leq j \leq K$, set

$$a(\theta_j^*) = \begin{cases} 2(1 - \alpha(\theta_j^*) - \eta(\theta_j^*)) & \text{if } \alpha(\theta_j^*) + \eta(\theta_j^*) < \frac{1}{2}, \\ 1 + \varepsilon & \text{if } \alpha(\theta_j^*) + \eta(\theta_j^*) = \frac{1}{2}, \\ 1 & \text{if } \alpha(\theta_j^*) + \eta(\theta_j^*) > \frac{1}{2}. \end{cases}$$

Note that when the change-point sequence is known, the Whittle estimate of θ_j^* converges at rate $n_j^{a(\theta_j^*)/2-1}$, for any $1 \leq j \leq K^*$:

Lemma 3.1. *Under Hypothesis 4, there exists a $C > 0$ such that, for any $1 \leq j \leq K^*$,*

$$\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P} \left(n_j^{1-a(\theta_j^*)/2} \|\hat{\theta}_n(T_j^*) - \theta_j^*\| > \delta \right) = 0.$$

Remark. We shall see in Theorem 3.4 that we obtain the same rates when the change-point sequence is unknown.

Proof. The proof uses standard minimum contrast estimator arguments. For any $1 \leq j \leq K^*$, for any $\theta \in \Theta$, set

$$S_n(T_j^*, \theta) = \sum_{t \in T_j^*} \sum_{s \in T_j^*} X_t X_s b(t-s, \theta) - n_j^* \sum_{u=-\infty}^{+\infty} \gamma(u, \theta_j^*) b(u, \theta). \tag{16}$$

Here $b(u, \theta)$ is the u th Fourier coefficient of $1/f(\cdot, \theta)$. The estimator $\hat{\theta}_n(T_j^*)$ minimizes $W_n(T_j^*, \theta)$. On the other hand, using equations (13)–(16), $W_n(T_j^*, \theta)$ can be decomposed as a deterministic contrast function plus a random fluctuation:

$$W_n(T_j^*, \theta) = w_{\theta_j^*}(\theta) + \frac{S_n(T_j^*, \theta)}{n_j^*}.$$

It is enough to see that $w_{\theta_j^*}(\theta)$ reaches its minimum for $\theta = \theta_j^*$ and that the random fluctuation $S_n(T_j^*, \theta)/n_j^*$ tends to 0 in probability, uniformly in θ , at the specified rate. \square

In many interesting cases (under mixing conditions, or under Definition 1), $a(\theta_j^*) = 1$ and under some mild additional technical conditions a central limit theorem is available: there exists a $d \times d$ limiting covariance matrix $\Gamma(\theta_j^*)$ such that

$$\sqrt{n}(\hat{\theta}_n(T_j^*) - \theta_j^*) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma(\theta_j^*)).$$

See Giraitis and Surgailis (1990) or Ludeña and Lavielle (1999) for an extension to random fields.

Define $J_n(\mathbf{t}, \boldsymbol{\theta})$ as

$$J_n(\mathbf{t}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{k=1}^K n_k W_n(T_k, \theta_k). \tag{17}$$

When \mathbf{t}^* is unknown the estimate $(\hat{\mathbf{t}}_n, \hat{\boldsymbol{\theta}}_n)$ is the value of $(\mathbf{t}, \boldsymbol{\theta})$ that minimizes $J_n(\mathbf{t}, \boldsymbol{\theta})$ in $\mathcal{F}_K \times \Theta_K$. Thus, $\hat{\theta}_{n,k} \stackrel{\text{def}}{=} \hat{\theta}_n(\hat{T}_{n,k})$ is the Whittle estimate of θ_k^* computed in the segment k of the estimated configuration of change-points $\hat{\mathbf{t}}_n$, with $\hat{T}_{n,k} = \{t \in \mathbb{Z}, \hat{t}_{n,k-1} + 1 \leq t \leq \hat{t}_{n,k}\}$.

3.2.2. Consistency of the estimator

We first establish the consistency of the estimator when the number of segments K^* is known, but when the location of the change-points is unknown.

Theorem 3.2. *Let $\hat{\mathbf{t}}_n$ be the estimate of the change-points sequence and $\hat{\boldsymbol{\theta}}_n$ be the estimate of the parameters in the different segments, obtained as the solution of the following minimization problem:*

$$J_n(\hat{\mathbf{t}}_n, \hat{\boldsymbol{\theta}}_n) \leq J_n(\mathbf{t}, \boldsymbol{\theta}), \forall (\mathbf{t}, \boldsymbol{\theta}) \in \mathcal{F}_{K^*} \times \Theta_{K^*},$$

where J_n was defined in (17). Let $\hat{\boldsymbol{\tau}}_n = \hat{\mathbf{t}}_n/n$. Then, under Hypotheses 3 and 4, $(\hat{\boldsymbol{\tau}}_n, \hat{\boldsymbol{\theta}}_n)$ converges in probability to $(\boldsymbol{\tau}^*, \boldsymbol{\theta}^*)$.

Proof. For any $1 \leq k, k' \leq K$, for any $1 \leq j, j' \leq K^*$, for any $\theta_k \in \Theta$, set

$$S_n(T_{kj}, \theta_k) = \sum_{t \in T_{kj}} \sum_{s \in T_{kj}} X_t X_s b(t-s, \theta_k) - n_{kj} \sum_{u=-\infty}^{+\infty} \gamma(u, \theta_j^*) b(u, \theta_k), \quad (18)$$

$$R_n(T_{kj}, T_{k'j'}, \theta_k) = \sum_{t \in T_{kj}} \sum_{s \in T_{k'j'}} X_t X_s b(t-s, \theta_k). \quad (19)$$

Here $b(u, \theta_k)$ is the u th Fourier coefficient of $1/f(\cdot, \theta_k)$. Then, using equations (13), (14), (15), (18) and (19), we have

$$W_n(T_k, \theta_k) = \sum_{j=1}^{K^*} \frac{n_{kj}}{n_k} w_{\theta_j^*}(\theta_k) + \frac{S_n(T_k, \theta_k)}{n_k},$$

$$W_n(T_j^*, \theta_j^*) = w_{\theta_j^*}(\theta_j^*) + \frac{S_n(T_j^*, \theta_j^*)}{n_j^*},$$

where

$$S_n(T_k, \theta_k) = \sum_{j=1}^{K^*} S_n(T_{kj}, \theta_k) + \sum_{j=1}^{K^*} \sum_{j' \neq j} R_n(T_{kj}, T_{k'j'}, \theta_k), \quad (20)$$

$$S_n(T_j^*, \theta_j^*) = \sum_{k=1}^K S_n(T_{kj}, \theta_j^*) + \sum_{k=1}^K \sum_{k' \neq k} R_n(T_{kj}, T_{k'j'}, \theta_j^*), \quad (21)$$

For any $(\theta, \theta') \in \Theta \times \Theta$, set

$$v(\theta', \theta) = w_\theta(\theta') - w_\theta(\theta). \quad (22)$$

Observe that $v(\theta', \theta) \geq 0$ and $v(\theta', \theta) = 0$ if and only if $\theta' = \theta$, under Hypothesis 4(i).

For technical reasons, instead of $J_n(\mathbf{t}, \boldsymbol{\theta})$ we will use the contrast function $U_n(\mathbf{t}, \boldsymbol{\theta})$ defined by

$$U_n(\mathbf{t}, \boldsymbol{\theta}) = J_n(\mathbf{t}, \boldsymbol{\theta}) - J_n(\mathbf{t}^*, \boldsymbol{\theta}^*).$$

Since $J_n(\mathbf{t}^*, \boldsymbol{\theta}^*)$ is a constant, $(\hat{\mathbf{t}}_n, \hat{\boldsymbol{\theta}}_n)$ also minimizes $U_n(\mathbf{t}, \boldsymbol{\theta})$.

Using (22), we can write

$$U_n(\mathbf{t}, \boldsymbol{\theta}) = u(\mathbf{t}, \boldsymbol{\theta}) + e_n(\mathbf{t}, \boldsymbol{\theta}), \quad (23)$$

with

$$\begin{aligned}
 u(\mathbf{t}, \boldsymbol{\theta}) &= \sum_{k=1}^K \sum_{j=1}^{K^*} \frac{n_{kj}}{n} v(\theta_k, \theta_j^*) \\
 e_n(\mathbf{t}, \boldsymbol{\theta}) &= \frac{1}{n} \left[\sum_{k=1}^K S_n(T_k, \theta_k) - \sum_{j=1}^{K^*} S_n(T_j^*, \theta_j^*) \right]. \tag{24}
 \end{aligned}$$

Equation (23) can be interpreted as a deterministic term $u(\mathbf{t}, \boldsymbol{\theta})$ plus an error term $e_n(\mathbf{t}, \boldsymbol{\theta})$ which we will refer to as the fluctuation. In most change-point problems that consider only the marginal distribution of the observations (see Lavielle 1999), the fluctuation terms $e_n(\mathbf{t}, \boldsymbol{\theta})$ can generally be written as a sum of (possibly dependent) variables. In our case, as we are interested in changes in the spectrum, we must deal with quadratic forms. However, from (20) and (21) we have that the fluctuation term can be written as a sum of ‘good’ quadratic fluctuation terms $S_n(T_{kj}, \theta_k)$ and $S_n(T_{kj}, \theta_j^*)$ plus certain remainder cross terms $R_n(T_{kj}, T_{k'j}, \theta_k)$ and $R_n(T_{kj}, T_{k'j}, \theta_j^*)$ which can be controlled in probability.

Consistency of $(\hat{\mathbf{t}}_n, \hat{\boldsymbol{\theta}}_n)$ is then a direct consequence of the following two facts. First, the following lemma which was shown in Lavielle (1999):

Lemma 3.3. *Let v be a function from $\Theta \times \Theta$ to \mathbb{R} such that $v(\theta', \theta) \geq 0$ and $v(\theta', \theta) = 0$ if and only if $\theta = \theta'$. Then, with the above notation, there exists a constant $C_{\theta^*} > 0$ such that, for any $(\mathbf{t}, \boldsymbol{\theta}) \in \mathcal{T}_K \times \Theta_K$,*

$$u(\mathbf{t}, \boldsymbol{\theta}) = \sum_{k=1}^K \sum_{j=1}^{K^*} \frac{n_{kj}}{n} v(\theta_k, \theta_j^*) \geq \frac{C_{\theta^*}}{n} \|\mathbf{t} - \mathbf{t}^*\|_{\infty},$$

where $\|\mathbf{t} - \mathbf{t}^*\| = \max_{1 \leq j \leq K^* - 1} \min_{1 \leq k \leq K - 1} |t_k - t_j^*|$.

Second, if Hypotheses 3 and 4 are satisfied, the fluctuation term $e_n(\mathbf{t}, \boldsymbol{\theta})$ converges to 0, uniformly in $(\mathbf{t}, \boldsymbol{\theta})$. Indeed, using (20), (21) and (24), we remark that, by Lemma 2.4, $e_n(\mathbf{t}, \boldsymbol{\theta})$ can be decomposed as a finite sum of terms that converge to 0, uniformly in $(\mathbf{t}, \boldsymbol{\theta})$.

For any $\delta > 0$, let us define $\mathcal{T}_{K,\delta} = \{\mathbf{t} \in \mathcal{T}_K; \|\mathbf{t} - \mathbf{t}^*\|_{\infty} > n\delta\}$. Thus, using Lemma 3.3, for any $\delta > 0$, we have,

$$\begin{aligned}
 \mathbb{P}(\|\hat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}^*\|_{\infty} > \delta) &\leq \mathbb{P}\left(\inf_{\boldsymbol{\theta} \in \Theta_K} \min_{\mathbf{t} \in \mathcal{T}_{K,\delta}} U_n(\mathbf{t}, \boldsymbol{\theta}) < 0\right) \\
 &\leq \mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \Theta_K} \max_{\mathbf{t} \in \mathcal{T}_{K,\delta}} |e_n(\mathbf{t}, \boldsymbol{\theta})| > \inf_{\boldsymbol{\theta} \in \Theta_K} \min_{\mathbf{t} \in \mathcal{T}_{K,\delta}} u(\mathbf{t}, \boldsymbol{\theta})\right) \\
 &\leq \mathbb{P}\left(\sup_{\boldsymbol{\theta} \in \Theta_K} \max_{\mathbf{t} \in \mathcal{T}_{K,\delta}} |e_n(\mathbf{t}, \boldsymbol{\theta})| > C_{\theta^*} \delta\right). \tag{25}
 \end{aligned}$$

Thus, the right-hand side of (25) goes to 0 as $n \rightarrow \infty$, and $\hat{\boldsymbol{\tau}}_n$ converges to $\boldsymbol{\tau}^*$.

Since the estimated change-point sequence converges to the true change-point sequence,

$\hat{\theta}_n(\hat{T}_{n,j}) - \hat{\theta}_n(T_j^*)$ converges to 0 in probability. Using the fact that $\hat{\theta}_n(T_j^*)$ converges to θ_j^* , we conclude that $\hat{\theta}_{n,j} = \hat{\theta}_n(\hat{T}_{n,j})$ converges to θ_j^* in probability. \square

3.2.3. The rate of convergence

Once the consistency has been established, we can give bounds for the rate of convergence.

Theorem 3.4. Assume that Hypotheses 3 and 4 are satisfied. For any $1 \leq j \leq K$, let $h_j^* = h(\theta_j^*) = \max(2 - 2\eta(\theta_j^*) - 2\alpha(\theta_j^*), 1)$. Then, the sequences $\{\hat{t}_{n,j} - t_j^*\}$ and $\{\sqrt{n^{2-h_j^*}} \|\hat{\theta}_{n,j} - \theta_j^*\|\}$ are uniformly tight in probability:

- (i) $\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} P(|\hat{t}_{n,j} - t_j^*| \geq \delta) = 0;$
- (ii) $\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} P(\sqrt{n^{2-h_j^*}} \|\hat{\theta}_{n,j} - \theta_j^*\| \geq \delta) = 0.$

Remark. Theorem 3.4 states that the rate of convergence of the estimated change-point sequence $\hat{\tau}_n$ is n^{-1} under very general conditions. Furthermore, this rate does not depend on the dependence structure of the process. On the other hand, the fact that the change-point instants must be estimated does not affect the rate of convergence of $\hat{\theta}_n$: the rate is still the rate of the Whittle estimate, computed with the true change-points.

Proof. By Theorem 3.2, for big enough n , the estimated change-point must belong to a contiguous time segment. Thus, without any loss of generality, we consider the case of a unique change-point t^* , and we show that $P(\hat{t}_n - t^* \geq \delta)$ goes to 0 when δ and n go to ∞ (the case $P(t_n^* - \hat{t} \geq \delta)$ is obviously identical).

Since there is only one change-point, we have $T_1^* = \{1, 2, \dots, t^*\}$ and $T_2^* = \{t^* + 1, \dots, n\}$. For any $1 \leq t \leq n$, set $T_1 = \{1, 2, \dots, t\}$ and $T_2 = \{t + 1, \dots, n\}$.

Assume that $1 \leq t^* \leq t \leq n$. We then have $T_{11} = T_1^*$, $T_{12} = \{t^* + 1, \dots, t\}$ and $T_{22} = T_2$. Also $n_{11} = t^*$, $n_{12} = t - t^*$ and $n_{22} = n - t$.

For any $1 \leq t \leq n$, let $\hat{\theta}_n(T_1)$ and $\hat{\theta}_n(T_2)$ be the values of θ_1 and θ_2 that minimize $W_n(T_1, \theta_1)$ and $W_n(T_2, \theta_2)$. Thus, \hat{t}_n is the value of t that minimizes

$$J_n(t) = J_n(t, \hat{\theta}_n(T_1), \hat{\theta}_n(T_2)) = \frac{n_1}{n} W_n(T_1, \hat{\theta}_n(T_1)) + \frac{n_2}{n} W_n(T_2, \hat{\theta}_n(T_2)),$$

with $n_1 = t$ and $n_2 = n - t$. For any $t^* \leq t \leq n$, we have

$$\begin{aligned} J_n(t) &= \frac{n_{11}}{n} W_n(T_{11}, \hat{\theta}_n(T_1)) + \frac{n_{12}}{n} W_n(T_{12}, \hat{\theta}_n(T_1)) \\ &\quad + \frac{R_n(T_{11}, T_{12}, \hat{\theta}_n(T_1))}{n} + \frac{n_{22}}{n} W_n(T_{22}, \hat{\theta}_n(T_2)) \end{aligned} \tag{26}$$

$$\begin{aligned}
 J_n(t^*) &= \frac{n_{11}}{n} W_n(T_{11}, \hat{\theta}_n(T_1^*)) + \frac{n_{12}}{n} W_n(T_{12}, \hat{\theta}_n(T_2^*)) \\
 &\quad + \frac{R_n(T_{11}, T_{12}, \hat{\theta}_n(T_2^*))}{n} + \frac{n_{22}}{n} W_n(T_{22}, \hat{\theta}_n(T_2^*)). \tag{27}
 \end{aligned}$$

Since $\hat{\tau}_n$ converges to τ^* in probability, there exists a positive sequence (m_n) such that $m_n \rightarrow +\infty$ and $m_n/n \rightarrow 0$, and such that $P(\hat{t}_n - t^* > m_n) \rightarrow 0$, when $n \rightarrow +\infty$. Write $\mathcal{F}_{n,\delta} = \{t \in \mathbb{Z}, t^* + \delta < t < t^* + m_n\}$. Part (i) will follow if we show that

$$\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} P\left(\min_{t \in \mathcal{F}_{n,\delta}} J_n(t) - J_n(t^*) < 0\right) = 0. \tag{28}$$

We will establish (28) by studying each one of the terms obtained from the decompositions of $J_n(t^*)$ and $J_n(t)$ proposed in (26) and (27):

- (a) $W_n(T_{11}, \hat{\theta}_n(T_1^*)) \leq W_n(T_{11}, \hat{\theta}_n(T_1))$. Indeed, $\hat{\theta}_n(T_1^*)$ minimizes $W_n(T_1^*, \theta_1)$, where $T_1^* = T_{11}$.
- (b) $W_n(T_{12}, \hat{\theta}_n(T_1)) - W_n(T_{12}, \hat{\theta}_n(T_2^*))$ converges to $v(\theta_1^*, \theta_2^*)$, uniformly on $\mathcal{F}_{n,\delta}$. Indeed, $\hat{\theta}_n(T_i)$ converges uniformly on $\mathcal{F}_{n,\delta}$ to θ_i^* , $i = 1, 2$. Thus, by Lemma 2.4, $W_n(T_{12}, \hat{\theta}_n(T_i))$ converges uniformly to $w_{\theta_2^*}(\theta_i^*)$.
- (c) $R_n(T_{11}, T_{12}, \theta_n(T_1))/n_{12}$ and $R_n(T_{22}, T_{12}, \theta_n(T_2^*))/n_{12}$ converge to 0, uniformly on $\mathcal{F}_{n,\delta}$. This is a direct application of Lemma 2.4.
- (d) $(n_{22}/n_{12})(W_n(T_{22}, \hat{\theta}_n(T_2)) - W_n(T_{22}, \hat{\theta}_n(T_2^*)))$ converges to 0, uniformly over $\mathcal{F}_{n,\delta}$. To see this, write

$$|W_n(T_{22}, \hat{\theta}_n(T_2)) - W_n(T_{22}, \hat{\theta}_n(T_2^*))| \leq \frac{1}{2} \|\hat{\theta}_n(T_2) - \hat{\theta}_n(T_2^*)\|^2 \sup_{\theta \in \Theta} |\partial_{\theta^2}^2 W_n(T_{22}, \theta)|. \tag{29}$$

Since $\partial_{\theta} W_n(T_2, \hat{\theta}_n(T_2)) = \partial_{\theta} W_n(T_2^*, \hat{\theta}_n(T_2^*)) = 0$, there exists $\tilde{\theta} \in \Theta$ such that

$$\begin{aligned}
 &n_{22}(\hat{\theta}_n(T_2^*) - \hat{\theta}_n(T_2)) \frac{\partial^2}{\partial \theta^2} W_n(T_2, \tilde{\theta}) \\
 &\quad + n_{12} \left(\frac{\partial}{\partial \theta} W_n(T_{12}, \hat{\theta}_n(T_2^*)) + \frac{1}{n_{12}} \frac{\partial}{\partial \theta} R_n(T_{12}, T_{22}, \hat{\theta}_n(T_2^*)) \right) = 0,
 \end{aligned}$$

and $(n_{22}/n_{12})(\hat{\theta}_n(T_2^*) - \hat{\theta}_n(T_2)) = o_P(1)$, uniformly on $\mathcal{F}_{n,\delta}$, again by Lemma 2.4.

Set $C^* = v(\theta_1^*, \theta_2^*) \sup_{\theta \in \Theta} |\partial_{\theta^2}^2 W_n(T_2^*, \theta)|$. Then, the proof of (28) is achieved by observing that

$$\begin{aligned}
 P\left(\min_{t \in \mathcal{F}_{n,\delta}} J_n(t) - J_n(t^*) < 0\right) &\leq P\left(\frac{n_{22}}{n_{12}} \|\hat{\theta}_n(T_2) - \hat{\theta}_n(T_2^*)\|^2 > C^*/3\right) \\
 &\quad + P(R_n(T_{11}, T_{12}, \hat{\theta}_n(T_1))/n_{12} > C^*/3) + P(R_n(T_{11}, T_{12}, \hat{\theta}_n(T_1))/n_{12} > C^*/3).
 \end{aligned}$$

Part (ii) follows directly. Indeed since the error $n_{k,k+1} = \hat{t}_k - t_k^* = O_P(1)$, we deduce that $\hat{\theta}_n(\hat{T}_k) - \hat{\theta}_n(T_k^*) = O_P(n^{-1})$. Thus, both $\hat{\theta}_n(\hat{T}_k)$ and $\hat{\theta}_n(T_k^*)$ converge to θ_j^* at the rate of $n^{1-h_j^*/2}$ by Lemma 3.1. □

3.3. Estimation of the number of change-points

The results presented in this subsection are straightforward extensions of similar results already obtained in different contexts by Lavielle (1999) and Lavielle and Moulines (1999). We denote by K^* the true unknown number of segments, which is assumed to be bounded by a finite constant. First, we can remark that Theorem 3.2 can be generalized to the case $K \geq K^*$.

Lemma 3.5. *Let $(\hat{\mathbf{t}}_n, \hat{\boldsymbol{\theta}}_n)$ be the estimate of $(\mathbf{t}, \boldsymbol{\theta})$, obtained by minimizing $J_n(\mathbf{t}, \boldsymbol{\theta})$ defined in (17), over $\mathcal{T}_K \times \Theta_K$. Let $\hat{\boldsymbol{\tau}}_n = \hat{\mathbf{t}}_n/n$. Then, under Hypotheses 3 and 4, $\max_{1 \leq j \leq K^* - 1} \min_{1 \leq k \leq K - 1} |\hat{\tau}_{n,k} - \tau_j^*|$ converges in probability to 0 when $n \rightarrow \infty$, if $K \geq K^*$.*

Lemma 3.5 means that, even if the number of segments is fixed to a value K greater than the true value K^* , a subfamily $(\hat{\tau}_{n,k_j}, 1 \leq j \leq K^* - 1)$ of $\hat{\boldsymbol{\tau}}_n$ still converges to the true configuration $\boldsymbol{\tau}^*$ under the hypothesis of Theorem 3.2.

We will estimate the number of segments K^* together with $(\mathbf{t}^*, \boldsymbol{\theta}^*)$ by minimizing the penalized contrast function

$$\tilde{J}_n(K, \mathbf{t}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{k=1}^K n_k W_n(T_k, \theta_k) + \beta_n K. \quad (30)$$

A convenient choice for the sequence (β_n) will ensure the convergence of the estimate.

Theorem 3.6. *Assume that Hypotheses 3 and 4 hold. Let $a^* = \max_{1 \leq j \leq K^*} h_j^*$ and let (β_n) be a positive sequence such that*

$$\beta_n \xrightarrow{n \rightarrow \infty} 0; \quad n^{a^* - 2} \beta_n^2 \xrightarrow{n \rightarrow \infty} +\infty.$$

Then, the minimum penalized contrast estimator $(\hat{K}_n, \hat{\mathbf{t}}_n, \hat{\boldsymbol{\theta}}_n)$ obtained by minimizing the function $\tilde{J}_n(K, \boldsymbol{\tau}, \boldsymbol{\theta})$, defined in (30), over $\{1, 2, \dots, \bar{K}\} \times \tau_K \times \Theta_K$ converges in probability to $(K^, \boldsymbol{\tau}^*, \boldsymbol{\theta}^*)$ if $\bar{K} \geq K^*$.*

Proof. We will show that $P(\hat{K}_n \neq K^*)$ goes to 0 when $n \rightarrow \infty$.

For any $K < K^*$, we have

$$\begin{aligned} P(\hat{K}_n = K) &\leq P\left(\inf_{\boldsymbol{\theta} \in \Theta_K} \min_{\mathbf{t} \in \tau_K} \tilde{J}_n(K, \mathbf{t}, \boldsymbol{\theta}) < \tilde{J}_n(K^*, \mathbf{t}^*, \boldsymbol{\theta}^*)\right) \\ &\leq P\left(\inf_{\boldsymbol{\theta} \in \Theta_K} \min_{\mathbf{t} \in \mathcal{T}_K} (u(\mathbf{t}, \boldsymbol{\theta}) + e_n(\mathbf{t}, \boldsymbol{\theta})) < \beta_n(K^* - K)\right). \end{aligned}$$

Let $\Delta_{\boldsymbol{\tau}}^* = \min_{1 \leq j \leq K^*} (\tau_j^* - \tau_{j-1}^*)$. From Lemma 3.3, $u(\mathbf{t}, \boldsymbol{\theta}) \geq C_{\theta^*} \Delta_{\boldsymbol{\tau}}^* / 2 > 0$. On the other hand, $e_n(\mathbf{t}, \boldsymbol{\theta})$ converges to 0 uniformly on $\mathcal{T}_K \times \Theta_K$. We use the condition $\beta_n \rightarrow 0$ to conclude that $\lim_{n \rightarrow \infty} P(\hat{K}_n < K^*) = 0$.

We now have to show that $\lim_{n \rightarrow \infty} P(\hat{K}_n > K^*) = 0$. From Lemma 3.5, we know that a subfamily of $\hat{\boldsymbol{\tau}}_n$ converges to $\boldsymbol{\tau}^*$. Then we have to show that the penalty term allows us to suppress any spurious change-point in any segment of the true configuration. Thus, without

loss of generality, we shall consider a particular configuration t^* with only one segment and show that the probability of detecting a change-point goes to 0 as $n \rightarrow \infty$. Let θ^* be the true value of θ and let $T^* = \{1, 2, \dots, n\}$. For any $K \geq 2$ and for any $(t, \theta) \in \Theta_K \times \mathcal{T}_K$, let

$$U_n(t, \theta) = \sum_{k=1}^K \frac{n_k}{n} W_n(T_k, \theta_k) - W_n(T^*, \theta^*).$$

Then

$$P(\hat{K}_n = K) \leq P\left(\inf_{\theta \in \Theta_K} \min_{t \in \mathcal{T}_K} U_n(t, \theta) + (K - 1)\beta_n < 0\right).$$

We can easily verify that

$$nU_n(t, \theta) = \sum_{k=1}^K (n_k v(\theta_k, \theta^*) + S_n(T_k, \theta_k) - S_n(T_k, \theta^*)) - \sum_{k=1}^K \sum_{k' \neq k} R_n(T_k, T_{k'}, \theta^*).$$

For any $1 \leq k \leq K$,

$$|S_n(T_k, \theta_k) - S_n(T_k, \theta^*)| \leq \|\theta_k - \theta^*\| \left| \frac{\partial}{\partial \theta} S_n(T_k, \theta^*) \right| + \frac{\|\theta_k - \theta^*\|^2}{2} \left| \sup_{\theta \in \Theta} \frac{\partial^2}{\partial \theta^2} S_n(T_k, \theta) \right|.$$

Using Lemma 2.2, we have that, for any $1 \leq k, k' \leq K$,

$$P\left(\max_{t \in \mathcal{T}_K} \left| \frac{\partial}{\partial \theta} S_n(T_k, \theta^*) \right| > n\beta_n\right) \leq C_2 \frac{n^{2-h^*}}{\beta_n^2},$$

$$P\left(\max_{t \in \mathcal{T}_K} |R_n(T_k, T_{k'}, \theta^*)| > n\beta_n\right) \leq 2C_2 \frac{n^{2-h^*}}{\beta_n^2}.$$

Since $n^{a^*-2}\beta_n^2 \rightarrow +\infty$, these two probabilities go to 0 when $n \rightarrow \infty$. Thus, it remains to check that, for any $1 \leq k \leq K$,

$$\lim_{n \rightarrow \infty} P\left(\inf_{\theta_k \in \Theta_K} \min_{t \in \mathcal{T}_K} \left\{ n_k v(\theta_k, \theta^*) + \|\theta_k - \theta^*\|^2 \sup_{\theta \in \Theta} \frac{\partial^2}{\partial \theta^2} S_n(T_k, \theta) \right\} + n\beta_n < 0\right) = 0.$$

Since Θ is compact and from the definition of v , there exist $C > 0$ and $\Delta_\Theta < \infty$ such that, for any $(\theta, \theta') \in \Theta \times \Theta$, $\Delta_\Theta \geq v(\theta, \theta') \geq C\|\theta - \theta'\|^2$. This yields

$$P\left(\inf_{\theta_k \in \Theta_K} \min_{t \in \mathcal{T}_K} \left\{ n_k v(\theta_k, \theta^*) + \|\theta_k - \theta^*\|^2 \sup_{\theta \in \Theta} \frac{\partial^2}{\partial \theta^2} S_n(T_k, \theta) \right\} + n\beta_n < 0\right) \leq P_{n,1} + P_{n,2},$$

where

$$\begin{aligned}
 P_{n,1} &= \mathbb{P} \left(\sup_{\theta \in \Theta_K} \max_{n^{h^*/2} \leq n_k \leq n} \left| \frac{\partial^2 S_n(T_k, \theta)}{\partial \theta^2} \right| > C \right) \\
 P_{n,2} &= \mathbb{P} \left(\sup_{\theta \in \Theta_K} \max_{n_k \leq n^{h^*/2}} \Delta_{\Theta} \left| \frac{\partial^2 S_n(T_k, \theta)}{\partial \theta^2} \right| > n\beta_n \right) \\
 &= \mathbb{P} \left(\sup_{\theta \in \Theta_K} \max_{n_k \leq n^{h^*/2}} \Delta_{\Theta} \left| \frac{\partial^2 S_n(T_k, \theta)}{\partial \theta^2} \right| > \sqrt{n^{2-h^*} \beta_n^2} \right).
 \end{aligned}$$

We use Lemma 2.2 and the fact that $n^{a^*-2}\beta_n^2 \rightarrow +\infty$ to conclude that both $P_{n,1}$ and $P_{n,2}$ go to 0 when $n \rightarrow \infty$. □

3.4. A numerical experiment

We assume that X is a Gaussian ARIMA(0, d , 0) process. That is, X is solution of the equation

$$(1 - B)^d X_n = \varepsilon_n,$$

where B is the shift operator defined by $B(X_n) = X_{n-1}$, and where ε is a Gaussian white noise. Furthermore, we assume that an unknown number of changes affect the parameter d . We simulated a series $X = (X_t, 1 \leq t \leq 2000)$ with the following values of d : $d = 0.25$ for $t \in [1, 500]$; $d = 0.4$ for $t \in [501, 1000]$; $d = 0.15$ for $t \in [1001, 2000]$. The variance of the white noise remains constant: $\sigma_{\varepsilon}^2 = 1$. A realization of this process is displayed Figure 1(a).

The estimated configuration of change-points \hat{t}_n is obtained by minimizing the function $\tilde{J}_n(K, \tau, \theta)$ defined in (30).

A histogram of \hat{t}_n obtained with 100 realizations of X and $\beta_n = 8$ is displayed in Figure 1(b). We point out that two changes were well detected 83 times ($\hat{K}_n = 3$), one false alarm occurred 13 times ($\hat{K}_n = 4$), 2 false alarms occurred twice ($\hat{K}_n = 5$) and the first change was not detected twice ($\hat{K}_n = 2$).

4. Semi-parametric estimation in presence of change-points

4.1. The model and the estimate

Using the same methodology, we are also able to address a semi-parametric change-point problem. Specifically make the following assumption:

Hypothesis 5.

(i) There exist K^* covariance functions $\gamma_1^*, \gamma_2^*, \dots, \gamma_{K^*}^*$ such that

$$EX_s X_t = \gamma_j^*(t - s) \quad \text{if } t_{j-1}^* + 1 \leq s \leq t \leq t_j^*.$$

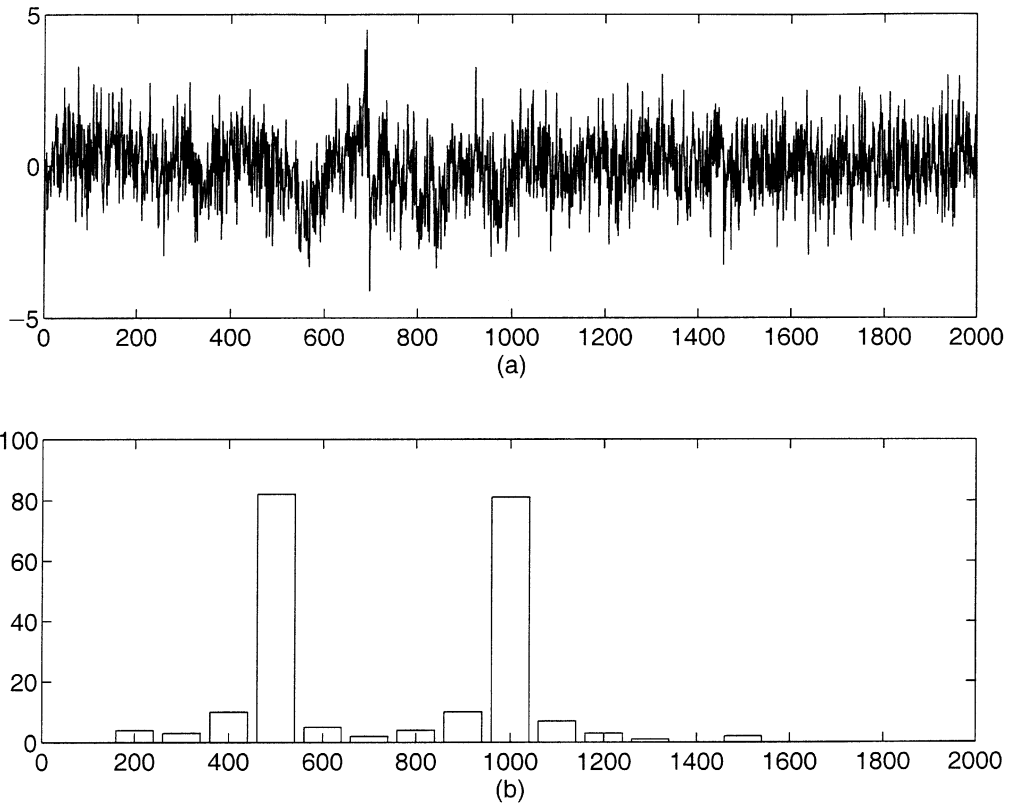


Figure 1. An example with several change-points. (a) The simulated series X : changes are present at $t_1^* = 500$ and $t_2^* = 1000$. (b) Histogram of the estimated change-point instants \hat{t}_n .

(ii) There exist a $G > 0$ and a $\gamma > 0$ such that

$$EX_s X_t \leq G(t - s)^{-\gamma} \quad \text{if there exists } 1 \leq j \leq K^* - 1 \text{ such that } s \leq t_j^* < t.$$

(iii) There exists a $\kappa > 0$ such that, for any $(t, s, t', s') \in \mathbb{Z}^4$,

$$|\text{cum}(X_t, X_s, X_{t'}, X_{s'})| \leq \kappa |E(X_t X_s) E(X_{t'} X_{s'})|.$$

Denote by f_j^* the spectral density of X in segment j of the true configuration t^* . We assume that the changes affect the spectrum of the process over certain previously specified frequency bands $[\lambda_i, \lambda_{i+1}]$, $i = 1, \dots, N$.

For the sake of simplicity, we assume here that the changes affect the spectrum on a unique frequency band $[0, \lambda]$. The method we develop can be extended directly to the case of several frequency bands, as will be done below with an application to real data.

Let

$$F_j^*(\lambda) = \int_0^\lambda f_j^*(\mu) d\mu.$$

We assume the following hypothesis is satisfied by $(F_j^*(\lambda))$ and (γ_j^*) :

Hypothesis 6.

- (i) For any $1 \leq j \leq K^* - 1$, $F_j^*(\lambda) \neq F_{j+1}^*(\lambda)$.
- (ii) There exist $\alpha_1 > 0, \dots, \alpha_{K^*} > 0$ and a constant $C > 0$ such that, for any $u > 0$ and any $1 \leq j \leq K^*$,

$$|\gamma_j^*(u)| \leq Cu^{-\alpha_j}.$$

For any configuration of change-points $\mathbf{t} \in \mathcal{T}_K$, and for any $1 \leq k \leq K$, set

$$\hat{F}_n(T_k, \lambda) = \int_0^\lambda I_n(T_k, \mu) d\mu,$$

where $I_n(T_k, \cdot)$ was defined in (13) as the periodogram computed over segment k of \mathbf{t} . We know that $\hat{F}_n(T_j^*, \lambda)$ converges almost surely to $F_j^*(\lambda)$ when $n \rightarrow \infty$, for any $1 \leq j \leq K^*$. Then it is quite natural to define a contrast function based on this estimate, for detecting jumps in $(F_j^*(\lambda))$. Indeed, we shall estimate \mathbf{t}^* by minimizing the contrast function J_n defined by

$$J_n(\mathbf{t}) = -\frac{1}{n} \sum_{k=1}^K n_k \hat{F}_n^2(T_k, \lambda), \tag{31}$$

when the number of change-points is not estimated. On the other hand, K^* and \mathbf{t}^* can be estimated simultaneously by minimizing the penalized contrast function

$$\tilde{J}_n(K, \mathbf{t}) = -\frac{1}{n} \sum_{k=1}^K n_k \hat{F}_n^2(T_k, \lambda) + \beta_n K, \tag{32}$$

where conditions over the sequence β_n will be given below. Since λ is a known fixed value, it can be treated as a constant, and the dependence of J_n and \tilde{J}_n with respect to λ is omitted.

In a parametric context, a lower bound for the length of the estimated segments was not required. In the nonparametric setting, however, consistent estimators will be found if minimization of the contrast functions proposed in (31) and (32) are carried out over the set $\mathcal{T}_{K, \Delta_n}$ defined by

$$\mathcal{T}_{K, \Delta_n} = \{\mathbf{t} \in \mathcal{T}_K, t_k - t_{k-1} \geq \Delta_n\},$$

where Δ_n must be chosen according to the rate of convergence of $(\hat{F}_n(T_k, \lambda))$.

The following theorem summarizes the properties of these two estimates:

Theorem 4.1. Assume that Hypotheses 5 and 6 hold. Let $h_j^* = \max(2 - 2\alpha_j^*, 1)$ for any $1 \leq j \leq K^*$, and let $a^* = \max_{1 \leq j \leq K^*} h_j^*$. Also let (Δ_n) be a positive sequence such that $n^{1-h^*} \Delta_n \rightarrow \infty$ as $n \rightarrow \infty$.

- (i) Let $\hat{\mathbf{t}}_n$ be the estimate of the change-point sequence, obtained as the solution of the following minimization problem:

$$J_n(\hat{\mathbf{t}}_n) \leq J_n(\mathbf{t}), \quad \forall \mathbf{t} \in \mathcal{T}_{K^*, \Delta_n},$$

where J_n was defined in (31). Then, $|\hat{t}_{n,j} - t_j^*| = O_P(1)$ for any $1 \leq j \leq K^* - 1$.
 (ii) Let (β_n) be positive sequence such that

$$\beta_n \xrightarrow[n \rightarrow \infty]{} 0; \quad n^{a^*-2} \beta_n^2 \xrightarrow[n \rightarrow \infty]{} +\infty.$$

Then the minimum penalized contrast estimator $(\hat{K}_n, \hat{\mathbf{t}}_n/n)$ obtained by minimizing the function $(\tilde{J}_n(K, \mathbf{t}))$, defined in (32), over $\{1, 2, \dots, \bar{K}\} \times \mathcal{T}_{K, \Delta_n}$ converges in probability to $(K^*, \boldsymbol{\tau}^*)$ if $\bar{K} \geq K^*$.

Proof. We shall show (i) first. For any $\mathbf{t} = (t_k, 1 \leq k \leq K - 1)$, set

$$S_n(T_k) = n_k \hat{F}_n(T_k, \lambda) - \sum_{j=1}^{K^*} n_{kj} F_j^*(\lambda).$$

Then, the estimate $\hat{\mathbf{t}}_n$ minimizes

$$\begin{aligned} U_n(\mathbf{t}) &= J_n(\mathbf{t}) - J_n(\mathbf{t}^*) \\ &= u(\mathbf{t}) + e_n(\mathbf{t}) + \eta_n(\mathbf{t}), \end{aligned}$$

where

$$u(\mathbf{t}) = \sum_{k=1}^K \sum_{j=1}^{K^*} \frac{n_{kj}}{n_k} (F_k^*(\lambda) - F_j^*(\lambda))^2,$$

$$e_n(\mathbf{t}) = \frac{2}{n} \sum_{j=1}^{K^*} F_j^*(\lambda) S_n(T_j^*) - \frac{2}{n} \sum_{k=1}^K \left(\sum_{j=1}^{K^*} \frac{n_{kj}}{n_k} F_j^*(\lambda) \right) S_n(T_k), \tag{33}$$

$$\eta_n(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^{K^*} \frac{S_n^2(T_j^*)}{n_j^*} - \frac{1}{n} \sum_{k=1}^K \frac{S_n^2(T_k)}{n_k}. \tag{34}$$

The arguments used for the consistency of $\hat{\mathbf{t}}_n$ in a parametric framework still apply. First, from Lemma 3.3, there exists a constant $C_{\theta^*} > 0$ such that, for any configuration $\mathbf{t} \in \mathcal{T}_K$, $u(\mathbf{t}, \boldsymbol{\theta}) \geq C_{\theta^*} n^{-1} \|\mathbf{t} - \mathbf{t}^*\|_\infty$. Second, if Hypotheses 5 and 6 are satisfied, the fluctuations $e_n(\mathbf{t})$ and $\eta_n(\mathbf{t})$ converge to 0, uniformly over $\mathcal{T}_{K^*, \Delta_n}$. Indeed let $(b_\lambda(u), u \in \mathbb{Z})$ be the sequence defined by

$$b_\lambda(0) = \lambda, \quad b_\lambda(u) = \frac{\sin \lambda u}{u}, \quad u \neq 0.$$

Then, the decomposition of S_n proposed in (20) and (21) still holds:

$$S_n(T_k) = \sum_{j=1}^{K^*} S_n(T_{kj}) + \sum_{j=1}^{K^*} \sum_{j' \neq j} R_n(T_{kj}, T_{kj'}),$$

where

$$S_n(T_{kj}) = \sum_{t \in T_{kj}} \sum_{s \in T_{kj}} X_t X_s b_\lambda(t-s) - n_{kj} F_j^*(\lambda),$$

$$R_n(T_{kj}, T_{kj'}) = \sum_{t \in T_{kj}} \sum_{s \in T_{kj'}} X_t X_s b_\lambda(t-s).$$

Since $|b_\lambda(u)| \leq |u^{-1}|$ for any $u > 0$, Lemma 2.2 holds with $\beta = 1$: there exists a constant C_S such that, for any $1 \leq k \leq K^*$, for any $n > 0$ and for any $\delta > 0$,

$$P\left(\max_{1 \leq t_{k-1} \leq n} \max_{t_{k-1}+1 \leq t_k \leq n} \frac{S_n(T_k)}{n} \geq \delta\right) \leq C_S \frac{n^{h^*-2}}{\delta^2}.$$

We deduce from (33) that $e_n(\mathbf{t})$ converges to 0 when $n \rightarrow \infty$, uniformly over \mathcal{T}_{K^*} . On the other hand, again by Lemma 2.2, we have that, for any $1 \leq k \leq K^*$, for any $n > 0$ and for any $\delta > 0$,

$$P\left(\max_{1 \leq t_{k-1} \leq n - \Delta_n} \max_{t_{k-1} + \Delta_n \leq t_k \leq n} \frac{S_n^2(T_k)}{n(t_k - t_{k-1})} \geq \delta\right) \leq C_S \frac{n^{h^*-1}}{\delta \Delta_n}.$$

Since $n^{1-h^*} \Delta_n \rightarrow \infty$, we deduce from (34) that $\eta_n(\mathbf{t})$ converges to 0 when $n \rightarrow \infty$, uniformly over $\mathcal{T}_{K^*, \Delta_n}$.

Once the consistency of $\hat{\mathbf{t}}_n$ has been established, we can assume that $\eta_n(\hat{\mathbf{t}}_n)$ is negligible with respect to $e_n(\hat{\mathbf{t}}_n)$ when $n \rightarrow \infty$. Then, a slight adaptation of the proof of Theorem 3.4 achieves the proof of (i).

The proof of Theorem 3.6 can also be easily adapted in order to prove (ii). □

Remark. In many applications $h^* = 1$, and then it is enough to require that $\Delta_n \rightarrow \infty$ as $n \rightarrow \infty$ to ensure the consistency of $\hat{\mathbf{t}}_n$.

4.2. Application to EEG analysis

It is well known that the EEG recordings are non-stationary signals (see Biscay *et al.* 1995). Epileptogenic transients are among the phenomena that make the EEG non-stationary. The spectral characteristics of the observed series change abruptly at some random instants, and the detection of such changes is a crucial step for the interpretation of the recorded brain electrical activity.

The following frequency bands of brain activity are conventionally observed: delta (1.5–3.5 Hz), theta (3.5–7.5 Hz), alpha (7.5–12.5 Hz), and beta (12.5–19.5 Hz).

In a more general framework, assume that the spectral measure of the observed process changes over some previously chosen frequency bands $([\lambda_l, \lambda_{l+1}], 0 \leq l \leq L)$. Write $\lambda = (\lambda_l, 0 \leq l \leq L)$, and define the vector

$$F_j^*(\lambda) = \left(\int_{\lambda_l}^{\lambda_{l+1}} f_j^*(\mu) d\mu \right),$$

for any $1 \leq j \leq K^*$. We assume that Hypothesis 6 is satisfied: $F_j^*(\lambda) \neq F_{j+1}^*(\lambda)$ for any $1 \leq j \leq K^* - 1$. For any configuration of change-points $\mathbf{t} \in \mathcal{T}_K$, and for any $1 \leq k \leq K$, we also define

$$\hat{F}_n(T_k, \lambda) = \left(\int_{\lambda_l}^{\lambda_{l+1}} I_n(T_k, \mu) d\mu \right).$$

Since $\hat{F}_n(T_k, \lambda)$ is a vector, $\hat{F}_n^2(T_k, \lambda)$ is replaced by $\|\hat{F}_n(T_k, \lambda)\|^2$ in (31) and (32).

We present in Figure 2 the segmentation of two EEG recordings. We display the observed series together with the estimated change-point times (marked by vertical lines), and also the estimated spectral measure $\hat{F}_n(T_k, \lambda)$.

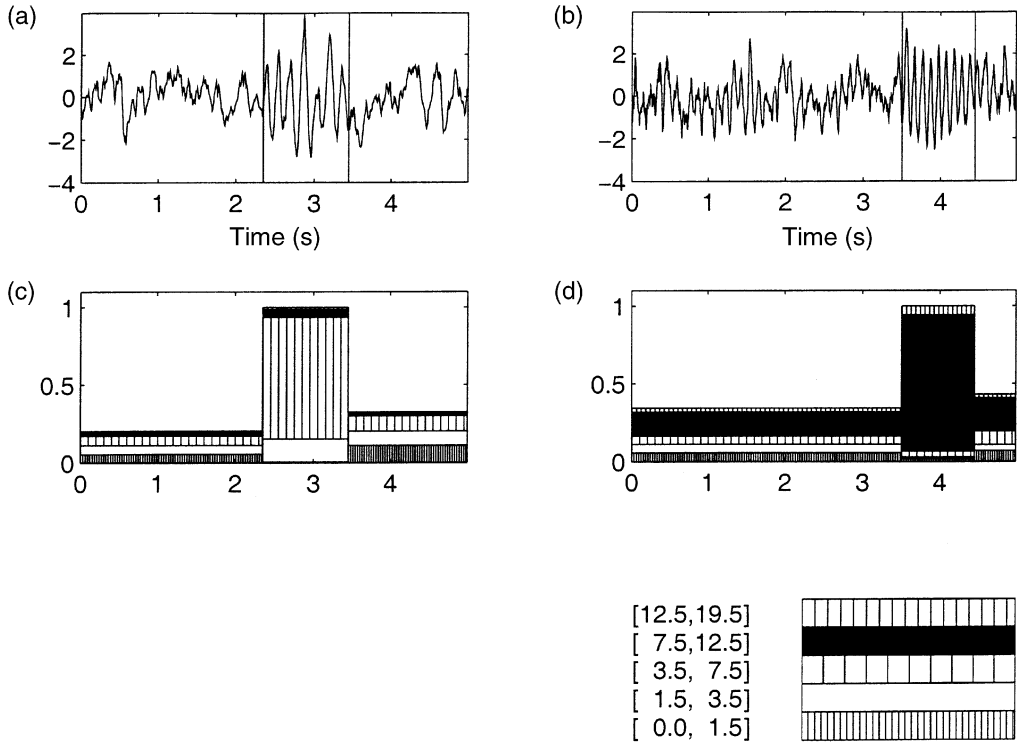


Figure 2. Segmentation of EEG recordings. (a) Segmentation of theta activity (3.5–7.5 Hz). (b) Segmentation of alpha activity (7.5–12.5 Hz). (c), (d) The estimated spectral distributions computed in the estimated segments.

Acknowledgements

We thank Rolando Biscay and Peter Valdés from the Cuban Neuroscience Center, Havana, for the many helpful comments concerning the application of the method to EEG data and for supplying real EEG data.

References

- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Automat. Control*, **19**, 716–723.
- Basseville, M. and Nikiforov, N. (1993) *The Detection of Abrupt Changes – Theory and Applications*. Englewood Cliffs, NJ: Prentice Hall.
- Biscay, R., Lavielle, M., González, A., Clark, I. and Valdés, P. (1995) Maximum a posteriori estimation of change points in the EEG. *Int. J. of Bio-Medical Computing*, **38**, 189–196.
- Brodsky, B. and Darkhovsky, B. (1993) *Nonparametric Methods in Change-Point Problems*. Dordrecht: Kluwer Academic Publishers.
- Dacunha-Castelle, D. and Gassiat, E. (1997) The estimation of the order of a mixture model. *Bernoulli*, **3**, 279–299.
- Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Ann. Statist.*, **25**, 1–37.
- Fox, R. and Taqqu, M. (1987) Central limit theorems for quadratic forms in random variables having long range dependence. *Probab. Theory Related Fields*, **74**, 213–240.
- Giraitis, L. and Leipus, R. (1990) Functional CLT for nonparametric estimates of the spectrum and change-point problem for the spectral function. *Lithuanian Math. J.*, **30**, 674–679.
- Giraitis, L. and Leipus, R. (1992) Testing and estimating in the change-point problem of the spectral function. *Lithuanian Math. J.*, **32**, 20–38.
- Giraitis, L. and Surgailis, D. (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields*, **86**, 87–104.
- Hannan, J. (1980) The estimation of the order of an ARMA process. *Ann. Statist.*, **8**, 1071–1081.
- Lavielle, M. (1999) Detection of multiple change in a sequence of dependent variables. *Stochastic Process. Appl.*, **83**, 79–102.
- Lavielle, M. and Moulines, E. (1999) Least squares estimation of an unknown number of shifts in a time series. *J. Time Series Anal.* To appear.
- Ludeña, C. and Lavielle, M. (1999) Some comments on the Whittle estimator for strongly dependent Gaussian fields. *Scand. J. Statist.*, **26**, 433–450.
- Móricz, F., Serfling, R. and Stout, W. (1982) Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. *Ann. Probab.*, **10**, 1032–1040.
- Picard, D. (1985) Testing and estimating change points in time series. *J. Appl. Probab.*, **17**, 841–867.
- Taqqu, M., Beran, J., Sherman, R. and Willenger, W. (1995) Long range dependence in variable-bit-rate video traffic. *IEEE Trans. Commun.*, **43**, 1566–1579.
- Terrin, N. and Taqqu, M. (1991) A non central limit theorem for quadratic forms of Gaussian stationary sequences. *J. Theoret. Probab.*, **3**, 449–475.

Received May 1998 and revised May 1999