LEAST-SQUARES ESTIMATION OF AN UNKNOWN NUMBER OF SHIFTS IN A TIME SERIES

By Marc Lavielle and Eric Moulines

Université Paris V and Université Paris-Sud Marc.Lavielle@math.v-psud.fr
and
Ecole Nationale Supérieure des Télécommunications, moulines@sig.enst.fr

First version received December 1996

Abstract. In this contribution, general results on the off-line least-squares estimate of changes in the mean of a random process are presented. First, a generalisation of the Hájek-Rényi inequality, dealing with the fluctuations of the normalized partial sums, is given. This preliminary result is then used to derive the consistency and the rate of convergence of the change-points estimate, in the situation where the number of changes is known. Strong consistency is obtained under some mixing conditions. The limiting distribution is also computed under an invariance principle. The case where the number of changes is unknown is then addressed. All these results apply to a large class of dependent processes, including strongly mixing and also long-range dependent processes.

Keywords. Detection of change points; Hájek-Rényi inequality; strongly mixing processes; strongly dependent processes; fractional Brownian motion; penalized least-squares estimate

1. INTRODUCTION

The problem of detecting and locating change-points in the mean of a random process, and estimating the magnitude of the jumps has been around for more than forty years. Most of the early efforts have been devoted to the detection/location of a single change-point in the mean of independent identically distributed random variables (see, among many other, Hinkley (1970), Sen Srivastava (1975), Hawkins (1977), Bhattacharya (1987)). Many recent contributions addressed the possible extensions of these methods and results to the detection/location of single/several change-points in the mean of a random (perhaps non-stationary) process. One of the pioneering contributions in that field is Picard (1985), who considered the detection of a single change-point in the mean a Gaussian AR process (whose order is known). The proposed method is based on maximum likelihood and is thus rather computationally demanding (moreover these results are not necessarily robust to deviations in the assumed model). A complete bibliography on change detection can be found in the books of Basseville and Nikiforov (1993) and Brodsky and Darkhovsky (1993).

Recently, Bai (1994) has proposed and studied a least-square estimate of the location of a single change-point in the mean of a linear process of general type under rather weak regularity assumption. This work has been later extended to multiple break points and weak dependent disturbance process (mixingale) by Bai and Perron (1996). From a practical point of view, least-squares estimates possess a main advantage over maximum likelihood methods: it does not require to specify the distribution of the error process ε. Furthermore, it is straightfoward to implement and is computationally efficient, even when the number of change-points is large. In this contribution, the results obtained by Bai (1994) are extended in two directions:

- While Bai (1994) considered the detection of a unique change-point, we consider the general case where multiple change-points can be present. As seen in the sequel, this extension is not trivial, and is worth being developed. When the number of change-points is unknown, the change-points problem becomes a problem of model selection and a penalized least-squares method is proposed (similar to that proposed in Yao (1988)).
- The results obtained by Bai (1994) and Bai and Perron (1996) hold only for weakly dependent processes. The main reason of this restriction is the use of a Hájek-Rényi type maximal inequality, extended by Bai (1994) to linear processes and by Bai and Perron (1996) to mixingales. We show in section II that it is possible to obtain this kind of inequality, under very mild assumption, including, for example, weakly and strongly (perhaps non stationary) dependent processes.

Exploiting this inequality, we show in Section III the consistency of the least-squares estimate when the number of change-points is known. Using a precise inequality obtained by Rio (1995) for stationary strongly mixing processes, we show, under some suitable conditions, the strong consistency of this estimator.

The rate of convergence of the change-points location estimator is then studied. It is shown, under very mild conditions (which includes, as a special case, long-range dependent processes), that the rate is n, where n is the number of observations (as in the case of independent and identically distributed random variables). The limiting distribution of the change-point location is then studied (when the magnitude of the jumps goes to zero at some specified rate).

Section IV is devoted to the number of change-points problem. This problem has been first addressed by Yao (1988), who proved the consistency of the Schwarz criterion when the disturbance is i.i.d. Gaussian with zero mean and unknown variance. In this contribution, the number of change-points is estimated by using a penalized least-squares approach. It is shown that an appropriately modified version of the Schwarz criterion yields a consistent estimator of the number of change-points under very weak conditions on the structure of the disturbance.

A small scale Monte-Carlo experiment is presented in Section V to support our claims. Some of the proofs are given in the appendix.

2. SOME RESULTS FOR THE FLUCTUATIONS OF PARTIAL SUMS

Let $\{\varepsilon_t\}_{t\geqslant 0}$ be a sequence of random variables. We define the partial sums by

$$S_{i:j} = \sum_{t=i}^{j} \varepsilon_t, \qquad 1 \leqslant i < j \leqslant n. \tag{1}$$

2.1. *A generalisation of the Hájek-Rényi inequality*

Let $\{b_k\}$ be a positive and decreasing sequence of real numbers. Hájek and Rényi (1955) have shown that, provided that $\{\varepsilon_t\}_{t\geqslant 0}$ is sequence of independent and identically distributed variables with zero-mean and finite variance $E\varepsilon_t^2 = \sigma^2 < \infty$,

$$\mathrm{P}\left(\max_{m\leqslant k\leqslant n} b_k|S_{1:k}| \geqslant \delta\right) \geqslant C_0 \frac{\sigma^2}{\delta^2}\left(mb_m^2 + \sum_{i=m+1}^{n} b_i^2\right), \tag{2}$$

with $C_0 = 1$. This result was extended to martingale increments by Birnbaum and Marshall (1961), and later to linear processes by Bai (1994); recall that $\{\varepsilon_t\}_{t\geqslant 0}$ is a linear process if

$$\varepsilon_t = \sum_{j=0}^{\infty} f_j \psi_{t-j} \tag{3}$$

where $\{\psi_t\}_{j\in\mathbb{Z}}$ is a sequence of independent variables with zero-mean and finite variance, such that $\sum_{j=0}^{\infty} j|f_j| < \infty$. In this context, the constant C_0 depends on the impulse response coefficients $\{f_j\}$ of the linear filter. It should be stressed that the result obtained by Bai deeply relies on the linear structure (3) of the process, and therefore, does not hold for non-linear processes. Moreover, the condition $\sum j|f_j| < \infty$ is a 'weak-mixing condition' (typically, the normalized sum $S_{1:k}/\sqrt{k}$ asymptotically converges to a Gaussian random variable under appropriate moment conditions); in particular, this condition does not hold for long-range dependent linear processes (in such case, $\sum f_j^2 < \infty$, but $\sum_j |f_j| = \infty$). We shall establish here an inequality of this kind, for a sequence (not necessarily stationary) $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ that satisfies, for $1 \leqslant \phi < 2$, the following condition:

- **H1**(ϕ) There exists $C(\varepsilon) < \infty$ such that, for all i, j,

$$\mathrm{E}(S_{i:j})^2 \leqslant C(\varepsilon)|j - i + 1|^{\phi}$$

Condition **H1**(ϕ) is fulfilled for a wide family of zero-mean processes $\varepsilon = \{\varepsilon\}_{t\in\mathbb{Z}}$. If ε is a second order stationary process, **H1**(ϕ) is fulfilled with $\phi = 1$ whenever the autocovariance function $\gamma(s) = \mathrm{E}\varepsilon_{t+s}\varepsilon_t$ satisfies $\sum_{s\geqslant 0}|\gamma(s)| < \infty$. This property is satisfied for linear processes of the form (3) such that $\sum_{j=0}^{\infty}|a_j| < \infty$, a class which includes, as a particular example, any ARMA process. Condition $\mathbf{H1}(\phi)$ is also satisfied for strongly mixing processes, under some conditions on the sequence of mixing coefficients $(\alpha(n))$ and on the moments of ε_t, see Doukhan (1994), or the quantile function of ε_t, see Rio (1995). For example, if ε_t is bounded with probability 1, $\mathbf{H1}(\phi)$ is satisfied with $\phi = 1$ if $\alpha(s) \leqslant M/s\log(s)$ (remark that, if $|\varepsilon_t| \leqslant C < \infty$ with probability 1, the autocovariance function is bounded by $|E\varepsilon_{t+s}\varepsilon_t| \leqslant 4C^2\alpha(s)$ and we have $\sum_{s\geqslant 0}|\gamma(s)| < \infty$). Finally, assumption $\mathbf{H1}(\phi)$ is also verified when ε is a zero-mean long-range dependent process, *i.e.*

$$\sup_{t\in\mathbb{Z}} \mathrm{E}(\varepsilon_{t+s}\varepsilon_t) \leqslant C'(\varepsilon)|s|^{2d-1}, \quad 0 < d < 1/2, \tag{4}$$

where d is the long-range dependence parameter (see, for example, Beran (1992)). Under these conditions, it is easy to see that there exists a constant $C(\varepsilon) < \infty$ such that, for $j > i$, $\mathrm{E}S_{i:j}^2 = C(\varepsilon)(j-i)^{1+2d}$ and $\mathbf{H1}(\phi)$ is satisfied with $\phi = 1 + 2d$.

THEOREM 1. *Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables that satisfies condition $\mathbf{H1}(\phi)$ for some $1 < \phi < 2$. Then, there exists a constant $A(\phi) \geqslant 1$ (that does not depend on ε) such that, for any $n \geqslant 1$, for any $1 \leqslant m \leqslant n$, for any $\delta > 0$, and for any positive and decreasing sequence $b_1 \geqslant b_2 \geqslant \ldots \geqslant b_n > 0$, we have the following inequalities:*

$$\mathrm{P}\Big(\max_{1\leqslant k\leqslant n} b_k|S_{i:k}| > \delta\Big) \leqslant \frac{A(\phi)C(\varepsilon)}{\delta^2} n^{\phi-1}\sum_{t=1}^{n} b_t^2, \tag{5}$$

$$\mathrm{P}\Big(\max_{m\leqslant k\leqslant n} b_k|S_{i:k}| > \delta\Big) \leqslant 4\frac{C(\varepsilon)m^{\phi}b_m^2}{\delta^2} + 4\frac{C(\varepsilon)A(\phi)}{\delta^2}(n-m)^{\phi-1}\sum_{t=m+1}^{n} b_t^2. \tag{6}$$

As a direct corollary of Theorem 1, we have

COROLLARY 2.1. *Assume that $\mathbf{H1}(\phi)$ holds for some $1 < \phi < 2$. Then, there exists a constant $C(\phi, \varepsilon) < \infty$, that depends on ε only through the constant $C(\varepsilon)$ such that, for any $m > 0$, any $\delta > 0$, and any $\beta > \phi/2$, we have:*

$$\mathrm{P}\Big(\max_{k\geqslant m} k^{-\beta}|S_{i:k}| > \delta\Big) \leqslant C(\phi, \varepsilon)m^{\phi-2\beta}. \tag{7}$$

REMARK. Let ε be a process satisfying $\mathbf{H1}(\phi)$ for some $1 < \phi < 2$ with some constant $C(\varepsilon)$, *i.e.* $\mathrm{E}(S_{i:j})^2 \leqslant C(\varepsilon)|j-i+1|^{\phi}$. Any time-shifted/time-reversed version of this process also verifies $\mathbf{H1}(\phi)$ with the *same* constants ϕ and $C(\varepsilon)$. Though the constants in the preceding theorems depend on ε only through $C(\varepsilon)$, the results in the preceding section are uniform with respect to the time-origin. In particular, we have:

$$\sup_{i\in\mathbb{Z}} \mathrm{P}\Big(\max_{k+i\geqslant m+i} k^{-\beta}|S_{i:k+i}| \geqslant \delta\Big) \leqslant C(\phi, \varepsilon)m^{\phi-2\beta}. \tag{8}$$

We shall use these results in the sequel with $b_k = 1$ for any k, and $b_k = 1/k$. We repeatedly use the following lemma, which is a direct application of Theorem 1 and its corollary:

LEMMA 2.2. *Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables that satisfies condition* **H1**(ϕ) *for some $1 < \phi < 2$. Then, there exist two constants $A(\phi, \varepsilon)$ and $B(\phi, \varepsilon)$ (that depend upon ε only through $C(\varepsilon)$) such that, for any $n > 0$ and any $\delta > 0$, we have:*

$$\sup_{i\in\mathbb{Z}} \mathrm{P}\left(\max_{i+1\leqslant k\leqslant n+i} |S_{i:k}| \geqslant \delta\right) \leqslant A(\phi, \varepsilon)\frac{n^{\phi}}{\delta^2} \tag{9}$$

$$\sup_{i\in\mathbb{Z}} \mathrm{P}\left(\max_{k\geqslant m+i-1} \frac{|S_{i:k}|}{k} \geqslant \delta\right) \leqslant B(\phi, \varepsilon)\frac{m^{\phi-2}}{\delta^2} \tag{10}$$

2.2. *A maximal inequality for strongly mixing stationary processes*

We consider now that ε is a strongly mixing stationary process (or α-*mixing*), see Doukhan (1994), or Rio (1995) for the definition of the sequence of strongly mixing coefficient $\{\alpha(n)\}_{n>0}$. Recall that ε is α-*mixing* if $\alpha(n) \to 0$ when $n \to \infty$.

We define also the quantile function Q: $\mathrm{P}(\varepsilon_1 > Q(u)) = u$, for $0 < u < 1$.

We make the following hypothesis:

- **H2** The process ε is α-*mixing*, and then exist $\rho > 0$ and $\gamma > 1$ such that:

a) There exists a constant C_Q and $u_0 > 0$ such that $Q(u) < C_Q u^{-1/\rho}$ for any $0 < u < u_0$.

b) There exists a constant C_α and $n_0 > 0$ such that $\alpha_n < C_\alpha n^{-\gamma}$ for any $n > n_0$.

c) $\rho > 4\gamma/(\gamma - 1)$.

Condition **H2** means that we control the tails of distribution of ε_t together with the mixing coefficients of ε. Furthermore, **H2-c** means that, the less mixing ε is (*i.e.*, the more dependent the sequence ε is), then, the more concentrated the marginal distribution of ε must be. (*i.e.* the lightest the tail of distribution of ε must be)

A sharp inequality obtained by Rio (1995) leads to the following result:

THEOREM 2. *Assume that* **H2** *is satisfied. Then, for any $\theta > 0$ such that*

$$\theta > \frac{\rho(3+\gamma) + 4\gamma}{2\rho(1+\gamma)}, \tag{11}$$

for any sequence (u_n) such that, for n large enough, $u_n > n^{\theta}$, and for any $\delta > 0$, we have:

$$\sum_{n=1}^{+\infty} \mathrm{P}\Big(\max_{1 \leqslant k \leqslant n} |S_{1:k}| \geqslant \delta u_n\Big) < +\infty. \tag{12}$$

Note that, under **H2-c**, the term in the right hand side of (11) is strictly bounded by 1. Hence, the sum in (12) is convergent for $u_n = n$.

3. ESTIMATION OF THE CHANGE-POINTS LOCATION

3.1. *Model assumptions and notations*

It is assumed the following model

$$Y_t = \mu_k^* + \varepsilon_t \qquad t_{k-1}^* + 1 \leqslant t \leqslant t_k^*,\ 1 \leqslant k \leqslant r \tag{13}$$

where we use the convention $t_0^* = 0$ and $t_{r+1}^* = n$. The indices of the break points and the mean values $\mu_1^*, \cdots, \mu_r^*$ are explicitly treated as unknown. It is assumed that $\min|\mu_{j+1}^* - \mu_j^*| > 0$. The purpose is to estimate the unknown means together with the break points when n observations $(Y_1, \cdots, Y_n)$ are available. In general, the number of breaks r can be treated as an unknown variable with true value r^*. However, for now, we treat it as known and will discuss methods to estimate it in later sections. It is assumed that there exists $\tau_1^*, \cdots, \tau_r^*$ such that, for $1 \leqslant k \leqslant r$, $t_k^* = [n\tau_k^*]$ ($[x]$ is the integer part of x). Following Bai and Perron (1996), (τ_k^*) are referred to as the *break fractions* and we let $\tau_0^* = 0$ and $\tau_{\tau+1}^* = 1$.

The method of estimation considered is that based on the least-square criterion. Let

$$\mathscr{A}_{n,r} = \{(t_0, t_1, \ldots, t_{r+1}),\ t_0 = 0 < t_1 < t_2 < \cdots < t_r < t_{r+1} = n\} \tag{14}$$

be the set of allowable r-partitions. In the sequel, the following set of allowable r-partitions is also considered.

$$\mathscr{A}_{n,r}^{\Delta_n} = \{(t_0, t_1, \ldots, t_{r+1});\ t_k - t_{k-1} \geqslant n\Delta_n\}. \tag{15}$$

where Δ_n is a sequence of non-increasing non-negative numbers such that $\Delta_n \to 0$ as $n \to \infty$ at some prescribed rate (in some cases, one may simply set $\Delta_n = 0$ for all n, and $\mathscr{A}_{n,r}^{\Delta_n} = \mathscr{A}_{n,r}$).

For each $\mathbf{t} \in \mathscr{A}_{n,r}$, the least-square estimates of the means are first obtained by minimizing the sum of square residuals, substituting them in the objective function and denoting the resulting sum as $Q_n(\mathbf{t})$

$$Q_n(\mathbf{t}) = \min_{(\mu_1,\cdots,\mu_{r+1})\varepsilon\mathbb{R}^{r+1}} \sum_{k=1}^{r+1} \sum_{t=t_{k-1}+1}^{t_k} (Y_t - \mu_k)^2$$

$$= \sum_{k=1}^{r+1} \sum_{t=t_{k-1}+1}^{t_k} (Y_t - \overline{Y}(t_{k-1}, t_k))^2. \tag{16}$$

where, for any sequence $\{u_t\}_{t\in\mathbb{Z}}$, we denote $\overline{u}(i, j)$ $(j > i)$ the average $\overline{u}(i, j) = (j - i)^{-1}\sum_{t=i+1}^{j} u_t$.

For any r-partitions $\mathbf{t}$, $\mathbf{t}' \in \mathscr{A}_{n,r}$, we define $\|\mathbf{t} - \mathbf{t}'\|_\infty = \max_{1\leqslant k\leqslant r}|t_k - t'_k|$.

THEOREM 3. *Assume* $\mathbf{H1}(\phi)$ *holds for some* $\phi < 2$. *Let* $\{\Delta_n\}_{n\geqslant 0}$ *be a positive non-increasing sequence such that* $\lim_{n\to\infty}\Delta_n = 0$ *and* $\lim_{n\to\infty} n^{2-\phi}\Delta_n = \infty$. *Let* $\hat{\mathbf{t}}_{\mathbf{n}}^{\Delta_n}$ *be the value of* $\mathbf{t}$ *that minimizes* $Q_n(\mathbf{t})$ over $\mathscr{A}_{n,r}^{\Delta_n}$. Then, $\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Delta_n} = \hat{\mathbf{t}}_{\mathbf{n}}^{\Delta_n}/n$ *converges in probability to* $\boldsymbol{\tau}^*$.

More precisely, denote $\Delta_\tau^* \triangleq \min_{1\leqslant k\leqslant r}|\tau_{k+1}^* - \tau_k^*|$. *There exists a constant* $K_1 < \infty$ *such that, for all* $(\mu_1^*, \cdots, \mu_{r+1}^*)$, *all* $0 < \delta \leqslant \Delta_\tau^*$ *and all n sufficiently large, it holds*

$$P(\|\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Delta_n} - \boldsymbol{\tau}^*\|_\infty \geqslant \delta) \leqslant K_1 n^{\phi-2}\underline{\lambda}^{-2}\delta^{-1}(\Delta_n^{-1} + \delta^{-1}(\overline{\lambda}/\underline{\lambda})^2) \tag{17}$$

where $\underline{\lambda} \triangleq \min_{1\leqslant k\leqslant r}|\mu_{k+1}^* - \mu_k^*|$ *and* $\overline{\lambda} \triangleq \max_{1\leqslant k\leqslant r}|\mu_{k+1} - \mu_k|$.

REMARK 1: In the above theorem, a minimum length $n\Delta_n$ between two successive change-points is imposed: instead of minimizing over all possible r-partitions, only the partitions such that $t_k - t_{k-1} \geqslant n\Delta_n$ are considered. Note that, $\Delta_n \to 0$ is chosen in such a way that consistent estimates of change fractions are obtained even when the lower bound for the change fractions is not known a priori. As seen later in this section, it is possible to remove this assumption by imposing either stronger conditions on the disturbance process or by constraining the estimates of the mean to lie within a compact set.

REMARK 2: Note that, since the constant K_1 *does not depend* on $(\lambda_1, \cdots, \lambda_r)$, the result (17) can be used in situations where these quantities depend on the sample size n (goes to zero at a certain rate with n). This property will be exploited later on, justifying the exact amount of effort needed to derive a uniform bound.

PROOF OF THEOREM 3: The proof is adapted from Bai (1994) and Bai and Perron (1996). We must verify that the contrast function associated to this contrast process has a unique minimum at the true value of the parameters, and that the contrast process converges uniformly to the contrast function. Define for any r-partition $\mathbf{t} \in \mathscr{A}_{n,r}$, the following quantities

$$J_n(\mathbf{t}) = n^{-1}(Q_n(\mathbf{t}) - Q_n(\mathbf{t}^*)), \tag{18}$$

$$K_n(\mathbf{t}) = n^{-1} \sum_{k+1}^{r+1} \sum_{t=t_{k-1}+1}^{t_k} (\mathrm{E} Y_t - \mathrm{E}\overline{Y}(t_{k-1}, t_k))^2, \tag{19}$$

$$V_n(\mathbf{t}) = n^{-1} \sum_{k+1}^{r+1} \left\{ \frac{(\sum_{t=t^*_{k-1}+1}^{t^*_k} \varepsilon_t)^2}{t^*_k - t^*_{k-1}} - \frac{(\sum_{t=t_{k-1}+1}^{t_k} \varepsilon_t)^2}{(t_k - t_{k-1})} \right\} \tag{20}$$

$$W_n(\mathbf{t}) = 2n^{-1} \sum_{k+1}^{r+1} \left\{ \left(\sum_{t=t^*_{k-1}+1}^{t^*_k} \varepsilon_t \right) \mu^*_k - \left(\sum_{t=t_{k-1}+1}^{t_k} \varepsilon_t \right) \mathrm{E}\overline{Y}(t_{k-1}, t_k) \right\}. \tag{21}$$

Using these notations, $J_n(\mathbf{t})$ may be decomposed as

$$J_n(\mathbf{t}) = K_n(\mathbf{t}) + V_n(\mathbf{t}) + W_n(\mathbf{t}). \tag{22}$$

We can show that $K_n(\mathbf{t})$ is lower bounded by

$$K_n(\mathbf{t}) \geqslant \min(n^{-1}\|\mathbf{t} - \mathbf{t}^*\|_\infty, \Delta^*_\tau)\underline{\lambda}^2. \tag{23}$$

Similarly, we need to obtain lower bound for $V_n(\mathbf{t})$ and $W_n(\mathbf{t})$. For $V_n(\mathbf{t})$, we have

$$V_n(\mathbf{t}) \geqslant -2n^{-2}\Delta_n^{-1}(r+1)\left(\max_{1 \leqslant s \leqslant n} \left(\sum_{t=1}^{s} \varepsilon_t \right)^2 + \max_{1 \leqslant s \leqslant n} \left(\sum_{t=n-s}^{n} \varepsilon_t \right)^2 \right) \tag{24}$$

Finally, note that

$$\begin{aligned} W_n(\mathbf{t}) = 2n^{-1} \sum_{k=1}^{r+1} & \left(\sum_{t=t^*_{k-1}+1}^{t^*_k} \varepsilon_t - \sum_{t=t_{k-1}+1}^{t_k} \varepsilon_t \right) (\mu^*_k - \mu^*) \\ & + \sum_{k=1}^{r+1} \left(\sum_{t=t_{k-1}+1}^{t_k} \varepsilon_t \right) (\mu^*_k - \mathrm{E}\overline{Y}(t_{k-1}, t_k)). \end{aligned}$$

where $\mu^* \stackrel{\triangle}{=} (\underline{r}+1)^{-1}\sum_{k=1}^{r+1}\mu^*_k$. Since for $1 \leqslant \underline{j},\ k \leqslant r+1$, it holds that $|\mu^*_j - \mu^*_k| \leqslant r\overline{\lambda}$, we have $|\mu^*_k - \mathrm{E}\overline{Y}(t_{k-1}, t_k)| \leqslant r\overline{\lambda}$ and $|\mu^*_k - \mu^*| \leqslant r\overline{\lambda}$, which implies

$$|W_n(\mathbf{t}) \geqslant 3n^{-1}(r+1)^2\overline{\lambda}\left(\max_{1 \leqslant s \leqslant n} \left| \sum_{t=1}^{s} \varepsilon_t \right| + \max_{1 \leqslant s \leqslant n} \left| \sum_{t=n-s}^{n} \varepsilon_t \right| \right) \tag{25}$$

For any $\delta > 0$, define

$$\mathscr{C}^{\Delta_n}_{n,\delta} \stackrel{\triangle}{=} \{\mathbf{t} \in \mathscr{A}^{\Delta_n}_{n,r},\ \|\mathbf{t} - \mathbf{t}^*\|_\infty \geqslant n\delta\}$$

Since $\Delta_n \to 0$, $\mathbf{t}^* \in \mathscr{A}_{n,r}^{\Delta_n}$ for sufficiently large n. Thus,

$$\mathrm{P}(\|\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Delta_n} - \boldsymbol{\tau}^*\|_\infty \geqslant \delta) \leqslant \mathrm{P}\left(\min_{t \in \mathscr{C}_{\mathbf{n},\delta}^{\Delta_n}} J_n(\mathbf{t}) \leqslant 0\right),$$

$$\leqslant \mathrm{P}\left(\max_{1 \leqslant s \leqslant n}\left(\sum_{t=1}^{s} \varepsilon_t\right)^2 + \max_{1 \leqslant s \leqslant n}\left(\sum_{t=n-s}^{n} \varepsilon_t\right)^2 \geqslant c\underline{\lambda}^2 n^2 \Delta_n \delta\right)$$

$$+ \mathrm{P}\left(\max_{1 \leqslant s \leqslant n}\left|\sum_{t=1}^{s} \varepsilon_t\right| + \max_{1 \leqslant s \leqslant n}\left|\sum_{t=n-s}^{n} \varepsilon_t\right| \geqslant c\underline{\lambda}^2 n \delta \overline{\lambda}^{-1}\right).$$

for some constant $c > 0$. The proof is concluded by applying Lemma 2.2 ■

3.2. *Alternative conditions*

As mentioned above, it is possible to remove the constraint on the minimum segment size Δ_n by imposing some additional conditions. As seen in the proof of Theorem 3, Δ_n is used to obtain a uniform bound of $V_n(\mathbf{t})$ (see (24)). One can obtain such bound under additional assumptions on the disturbance process. An example of such condition is given below; let $\{\beta_n\}$ be a non increasing sequence of numbers, such that $\beta_n \to 0$ and $n\beta_n \to \infty$. Consider the following assumption

- **H3**(β)

$$\lim_{n \to \infty} \mathrm{P}\left(\max_{0 \leqslant t_1 < t_2 \leqslant n} (t_2 - t_1)^{-1}\left|\sum_{t=t_1+1}^{t_2} \varepsilon_t\right|^2 \geqslant n\beta_n\right) = 0$$

Extending the result obtained in Lemma 1 in Yao (1988), **H3** is for example satisfied when ε is a zero-mean Gaussian process, provided that the covariance bound (4) holds. In such case, β_n may be set to: $\beta_n = 4C'(\varepsilon)\log(n)/n^{1-2d}$. We have

THEOREM 4. *Assume that* **H1**(ϕ) *and* **H3**(β) *hold for some* $\phi < 2$ *and some sequence* $\{\beta_n\}$ *such that* $\beta_n \to 0$ *and* $n\beta_n \to \infty$. *Let* $\hat{\mathbf{t}}_{\mathbf{n}}$ *be the value of* $\mathbf{t}$ *that minimizes* $Q_n(\mathbf{t})$ *over* $\mathscr{A}_{n,r}$. *Then,* $\hat{\boldsymbol{\tau}}_{\mathbf{n}} = \hat{\mathbf{t}}_{\mathbf{n}}/n$ *converges in probability to* $\boldsymbol{\tau}^*$.

The proof of Theorem 4 is a direct adaptation of Theorem 3. Uniform bounds (w.r.t. $(\lambda_1^*, \cdots, \lambda_r^*)$) similar to (17) can also be derived. To apply this theorem in a general setting, the following lemma proves more explicit conditions (in terms of the moments and of the dependence structure of the process) upon which **H3** is verified:

LEMMA 3.1. *Assume that there exist constants* $s > 0$, $1 \leqslant h_s < 2$ *and* $C_s > 0$ *such that* $s > 2/(2 - h_s)$ *and for all* $0 \leqslant t_1 < t_2 < +\infty$,

$$\mathrm{E}\left(\sum_{t=t_1+1}^{t_2} \varepsilon_t\right)^{2s} \leqslant C_s(t_2 - t_1)^{h_s s}. \tag{26}$$

Then, **H3** *holds with* $\beta_n = n^{(2+(h_s-2)s)/s} \log n$.

Proof of Lemma 3.1. The result follows directly from the relation

$$\mathrm{P}\left(\max_{1\leqslant t_1<t_2\leqslant n} (t_2 - t_1)^{-1}\left|\sum_{t=t_1+1}^{t_2} \varepsilon_t\right|^2 \geqslant n\beta_n\right)$$
$$\leqslant \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \mathrm{P}\left(\left|\sum_{t=t_1+1}^{t_2} \varepsilon_t\right|^2 \geqslant n\beta_n(t_2 - t_1)\right)$$

and the Markov inequality. ■

Remark: When $h_s = 1$, the relation (26) is a Rosenthal's type inequality. It holds for martingale increments with uniformly bounded moments of order $2s$. These inequalities also hold for weakly-dependent processes under appropriate conditions on the moments and on the rate of convergence of the mixing coefficients (sse, for example, Doukhan and Louichi (1997) for recent references). This inequality is also satisfied for (perhaps non-stationary) long-range dependent Gaussian processes, provided that (4) holds and for long-range dependent linear processes (3), under appropriate conditions on the f_j and on the moments of ψ_t.

Another solution consists in constraining the values of the estimated mean $\hat{\mu}_k$, $1 \leqslant k \leqslant r$ to lie within some compact subset Θ_r of $\mathbb{R}^{r+1}$. More specifically, for any $(\mathbf{t}, \boldsymbol{\mu}) \in \mathscr{A}_{n,r} \times \Theta_r$, denote

$$U_n(\mathbf{t}, \boldsymbol{\mu}) \stackrel{\triangle}{=} n^{-1} \sum_{k=1}^{r+1} \sum_{t=t_{k-1}+1}^{t_k} (Y_t - \mu_k)^2, \tag{27}$$

$$Q_n^{\Theta}(\mathbf{t}) \stackrel{\triangle}{=} \min_{\boldsymbol{\mu}\in\Theta_r} U_n(\mathbf{t}, \boldsymbol{\mu}) \tag{28}$$

$$\hat{\mathbf{t}}_{\mathbf{n}}^{\Theta} \stackrel{\triangle}{=} \operatorname*{argmin}_{\mathbf{t}\in\mathscr{A}_{n,r}} Q_n^{\Theta}(\mathbf{t}). \tag{29}$$

This criterion may be seen as a robustified least-square fitting procedure, which in a certain sense, trim the extreme values of the series by constraining the estimated means to lie inside a feasible set. In such case, it is not necessary to constrain the minimum length between the successive break points. Note however that there is a price to pay, because the change fraction estimator is consistent only if the subset Θ_r is chosen in such a way that $(\mu_1^*, \cdots, \mu_{r+1}^*) \in \Theta_r$.

THEOREM 5. *Assume* $\mathbf{H1}(\phi)$ *holds for some* $\phi < 2$. *Then, for any compact subset* Θ_r *of* $\mathbb{R}^{r+1}$, *such that* $(\mu_1^*, \ldots, \mu_{r+1}^*) \in \Theta_r$, $\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Theta} = \hat{\mathbf{t}}_{\mathbf{n}}^{\Theta}/n$ *converges in probability to* $\boldsymbol{\tau}^*$.

PROOF OF THEOREM 5. *Let* $\boldsymbol{\mu}^* \stackrel{\triangle}{=} (\mu_1^*, \cdots, \mu_{r+1}^*)$. By definition, $(\hat{\mathbf{t}}_{\mathbf{n}}^{\Theta}, \hat{\boldsymbol{\mu}}_{\mathbf{n}})$ minimizes $U_n(\mathbf{t}, \boldsymbol{\mu}) - U_n(\mathbf{t}^*, \boldsymbol{\mu}^*)$, where

$$U_n(\mathbf{t}, \boldsymbol{\mu}) - U_n(\mathbf{t}^*, \boldsymbol{\mu}^*) = \sum_{k=1}^{r+1}\sum_{j=1}^{r+1} \frac{n_{kj}}{n}(\mu_j^* - \mu_k)^2 - 2\sum_{k=1}^{r+1}\sum_{j=1}^{r+1} \frac{S_{kj}}{n}(\mu_k - \mu_j^*). \quad (30)$$

where,

$$n_{kj} = \#\{\{t_{k-1}+1, \cdots, t_k\} \cap (t_{j-1}^*, \cdots, t_j^*\}\}, \quad (31)$$

$$S_{kj} = \sum_{t \in \{t_{k-1}+1, \cdots, t_k\} \cap \{t_{j-1}^*, \cdots, t_j^*\}} \varepsilon_t \quad (32)$$

where, by convention, the sum over an empty set of indexes is zero. Note that the dependence of n_{kj} and S_{kj} on the r-partitions $\mathbf{t}$ and $\mathbf{t}^*$ is implicit. We can show that there exists a constant $C > 0$ such that, for all $(\mathbf{t}, \boldsymbol{\mu}) \in \mathscr{A}_{n,r} \times \Theta_r$ we have

$$\sum_{k=1}^{r+1}\sum_{j=1}^{r+1} \frac{n_{kj}}{n}(\mu_j^* - \mu_k)^2 \geqslant C \max(n^{-1}\|\mathbf{t} - \mathbf{t}^*\|_\infty, \|\boldsymbol{\mu} - \boldsymbol{\mu}^*\|^2). \quad (33)$$

On the other hand, $n^{-1}\sum_{k=1}^{r+1}\sum_{j=1}^{r+1}|S_{kj}|$ converges to 0, uniformly in $\mathbf{t} \in \mathscr{A}_{n,r}$. Thus, since Θ_t is compact, $n^{-1}\sum_{k=1}^{r+1}\sum_{j=1}^{r+1}|(\mu_k - \mu_j^*)S_{kj}|$ converges to 0, uniformly in $(\mathbf{t}, \boldsymbol{\mu})$ and $(\hat{\boldsymbol{\tau}}_{\mathbf{n}}, \hat{\boldsymbol{\mu}}_{\mathbf{n}})$ converges to $(\boldsymbol{\tau}^*, \boldsymbol{\mu}^*)$ if $\boldsymbol{\mu}^* \in \Theta_r$. Once again, it is also possible to obtain a uniform bound (w.r.t. $\boldsymbol{\mu}^* \in \Theta_r$) similar to (17). ■

Finally, strong consistency of the estimate is obtained under mixing conditions:

THEOREM 6. *Assume that* $\mathbf{H2}$ *is satisfied. For any* $\Delta > 0$, *let* $\hat{\mathbf{t}}_{\mathbf{n}}^{\Delta}$ *be the value of* $\mathbf{t}$ *that minimizes* $Q_n(\mathbf{t})$ *over* $\mathscr{A}_{n,r}^{\Delta}$. *Then, if* $\Delta \leqslant \Delta_\tau^*$, $\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Delta} = \hat{\mathbf{t}}_{\mathbf{n}}^{\Delta}/n$ *converges almost surely to* $\boldsymbol{\tau}^*$.

PROOF OF THEOREM 6. Since $\Delta \leqslant \Delta_\tau^*$, $\mathbf{t}^* \in \mathscr{A}_{n,r}^{\Delta}$. Thus, following the proof of Theorem 3, we have that, for any $\delta > 0$,

$$\mathrm{P}(\|\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Delta} - \boldsymbol{\tau}^*\|_\infty \geqslant \delta) \leqslant \mathrm{P}\left(\max_{1\leqslant s\leqslant n} \left(\sum_{t=1}^{s} \varepsilon_t \right)^2 + \max_{1\leqslant s\leqslant n} \left(\sum_{t=n-s}^{n} \varepsilon_t \right)^2 \geqslant c\underline{\lambda}^2 n^2 \Delta\delta \right)$$
$$+ \mathrm{P}\left(\max_{1\leqslant s\leqslant n} \left| \sum_{t=1}^{s} \varepsilon_t \right| + \max_{1\leqslant s\leqslant n} \left| \sum_{t=n-s}^{n} \varepsilon_t \right| \geqslant c\underline{\lambda}^2 n\delta\overline{\lambda}^{-1} \right)$$

for some constant $c > 0$. We conclude with Theorem 2 for the strong consistency of $\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Delta}$ under **H2**, by setting $u_n = n$, that is, $\theta = 1$. ∎

3.3. *Rate of convergence*

It is possible to derive the rate of convergence of the change-points estimate. It has been shown by Bai (1994), for a single change-point and weak-dependent disturbance that the rate of convergence is n (*i.e.* is linear with the sample size). This result has later been extended to multiple change-points (and more general linear regression models) by Bai and Perron (1996), under weak-dependence conditions for the additive disturbance. It is shown in the sequel that the rate of convergence of the change fraction is n under general assumptions on the disturbance process, which include, as particular examples, long-range dependence processes.

In this section, $\hat{\mathbf{t}}_{\mathbf{n}} \stackrel{\triangle}{=} \hat{\mathbf{t}}_{\mathbf{n}}(\mathscr{A}_{n,r}^{\Delta_n})$ is the estimate of $\mathbf{t}^*$ obtained by minimizing the contrast function $Q_n(\mathbf{t})$ over $\mathscr{A}_{n,r}^{\Delta_n}$. The following theorem also holds under the alternative conditions mentioned in the previous section.

THEOREM 7. *Assume that* **H1**(ϕ) *holds for some* $\phi < 2$. *Then, for all* $1 \leqslant j \leqslant r$, $\hat{t}_{n,j} - t_j^* = \mathrm{O_P}(1)$.

More precisely, denote $\Delta_\tau^* = \min_{1\leqslant j\leqslant r+1}|\tau_j^* - \tau_{j-1}^*|$, *and let* $0 < \gamma < 1/2$. *Then there exists a constant* $K_\gamma < \infty$ *such that for all* $(\mu_1^*, \cdots, \mu_{r+1}^*)$ *and all* $\delta > 0$, *it holds that, for large enough* n,

$$\mathrm{P}(\delta\underline{\lambda}^{2/\phi-2} \leqslant \|\hat{\mathbf{t}}_{\mathbf{n}} - \mathbf{t}^*\|_\infty \leqslant n\gamma\Delta_\tau^*) \leqslant K_\gamma(n^{\phi-2}\underline{\lambda}^{-2} + \delta^{\phi-2}(1 + (\overline{\lambda}/\underline{\lambda})^2) + n^{-1}\delta^{\phi-1}\underline{\lambda}^{2/\phi-2}) \quad (34)$$

where $\underline{\lambda} = \min_{1\leqslant k\leqslant r}|\mu_{k+1}^* - \mu_k^*|$ *and* $\overline{\lambda} = \max_{1\leqslant k\leqslant r}|\mu_{k+1}^* - \mu_k^*|$.

REMARK: Perhaps surprisingly, the rate of convergence of the estimator of the break fraction is not related to the rate of decay of the autocovariance function of the disturbance process ε. Once again, the need for a uniform upper bound (34) is justified by the need for a limit theory with steps λ_j going to zero with the sample size n.

PROOF OF THEOREM 7: Define

$$\mathscr{C}_{\delta,\gamma,n} = \{\mathbf{t} \in \mathscr{A}_{n,r},\ \delta\underline{\lambda}^{2/\phi-2} \leqslant \|\mathbf{t}-\mathbf{t}^*\|_\infty \leqslant n\gamma\Delta_\tau^*\}. \tag{35}$$

The proof consists in determining an upper bound for $P(\hat{\mathbf{t}}_\mathbf{n} \in \mathscr{C}_{\delta,\gamma,n})$. To that purpose, first decompose $\mathscr{C}_{\delta,\gamma,n}$ according as

$$\mathscr{C}_{\delta,\gamma,n} = \bigcup_{\mathscr{T}} \mathscr{C}_{\delta,\gamma,n} \cap \{\mathbf{t} \in \mathscr{A}_{n,r},\ t_k \geqslant t_k^*,\ \forall k \in \mathscr{T}\}$$

where the union is over all subsets $\mathscr{T}$ of the index set $\{1, \cdots, r\}$. We may compute an upper bound for each individual set $\mathscr{C}_{\delta,\gamma,n} \cap \{\mathbf{t} \in \mathscr{A}_{n,r},\ t_k \geqslant t_k^*,\ \forall k \in \mathscr{T}\}$. Of course, this upper bound does not depend on $\mathscr{T}$, and we consider, for notational simplicity, only the case where $\mathscr{T} = \{1, \cdots, r\}$. Denote $\mathscr{C}'_{\delta,\gamma,n} = \mathscr{C}_{\delta,\gamma,n} \cap \{\mathbf{t} \in \mathscr{A}_{n,r},\ t_k \geqslant t_k^*,\ \forall k \in \{1, \cdots, r\}\}$. We show in the Appendix that, for any $\delta > 0$,

$$P(\hat{\mathbf{t}}_\mathbf{n} \in \mathscr{C}'_{\delta,\gamma,n}) \leqslant K_\gamma(n^{\phi-2}\underline{\lambda}^{-2} + \delta^{\phi-2}(1 + (\overline{\lambda}/\underline{\lambda})^2) + n^{-1}\delta^{\phi-1}\underline{\lambda}^{2/\phi-2}). \tag{36}$$

3.4. *The limiting distribution*

In this section, we derive the asymptotic distribution $\hat{\mathbf{t}}_\mathbf{n}$ when the sample size goes to infinity. As shown by Picard (1985) and later by Bai (1994), Bai and Perron (1996), this limiting distribution can be used to construct confidence interval for change fractions. The limiting distribution also carried information on the way the estimator of the change fraction is linked with the other parameters in the model. Following Bai (1994), it is assumed in the sequel that the jump λ_j, $1 \leqslant j \leqslant r$, depends on the sample size n and diminishes as n increases. The limiting distribution for fixed jump size λ_j can be obtained in certain cases, but it depends in a very intricate way on the distribution of the disturbance and is thus of little practical use.

To stress the dependence of the jump in n, we use in this section the notation $\lambda_{n,j}$ instead of λ_j. We note $\overline{\lambda}_n = \max_{1\leqslant j\leqslant r}\lambda_{n,j}$ and $\underline{\lambda}_n = \min_{1\leqslant j\leqslant r}\lambda_{n,j}$. For some $\phi < 2$, we consider the following condition:

- **H4**(ϕ) For any $1 \leqslant j \leqslant r$ and for some $0 < \nu < 1$, it holds that

$$\overline{\lambda}_n \to 0 \qquad \text{and} \qquad n^{1-\nu}\underline{\lambda}_n^\gamma \underset{n\to\infty}{\longrightarrow} +\infty, \tag{37}$$

$$\underline{\lambda}_n^{-1}\lambda_{n,j} \underset{n\to\infty}{\longrightarrow} a_j \qquad 1 < a_j < \infty \tag{38}$$

where $\gamma \triangleq 2/(2-\phi)$.

The last requirement implies that the size of the jumps $\lambda_{n,j}$, $1 \leqslant j \leqslant r$ goes to zero with the sample size n at the same rate. For 'short-memory' processes ($\phi = 1$), one may set $\nu = 0$ (which is not allowed in the assumption above). In fact, it is in that case to weaken slightly this assumption. Typical conditions are, in such case $n^{-1}L(n)\underline{\lambda}_n^{-2} \to \infty$, where $L(n)$ is a slowly varying function as $n \to \infty$ (see Bai (1994)). Since the main emphasis here is on processes with long-range dependence, we do not pursue that road. In addition, the minimal size Δ_n of the interval between successive breakpoints (see Theorem 3) is set to $\Delta_n = n^{\nu(\phi-2)}$. Under these assumptions, one may show by applying Theorems 3 and 7 that the sequence $\{\underline{\lambda}_n^{\gamma}\hat{\mathbf{t}}_n\}$ is tight in the sense that

$$\lim_{\delta\to\infty} \lim_{n\to 0} \mathrm{P}(\underline{\lambda}_n^{\gamma}\|\hat{\mathbf{t}}_\mathbf{n} - \mathbf{t}^*\|_\infty \geqslant \delta) = 0. \tag{39}$$

Denote $\{B_\phi(s)\}_{s\geqslant 0}$ the fractional Brownian motion (fBm) with self-similarity exponent $\phi/2$ (for $\phi = 1$, $B_\phi(s)$ is the standard Brownian motion). Recall that the fBM is (the unique up to a scale factor) continuous Gaussian process with stationary increments satisfying $B_\phi(0) = 0$, $E(B_\phi(0)) = 0$ and $E(B_\phi(t)^2) = t^\phi$. Its covariance kernel is

$$\Gamma_\phi(t, s) = 1/2(s^\phi + t^\phi - |s-t|^\phi).$$

For the derivations that follow, we also need to introduce the two-sided fBM. This is, similarly, the unique (up to a scale factor) continuous Gaussian process with stationary increments satisfying $B_\phi(0) = 0$, $E\tilde{B}_\phi(t) = 0$, and $E\tilde{B}_\phi(t) = |t|^{2\phi}$. Its covariance kernel is given by

$$\mathrm{cov}(\tilde{B}_\phi(t), \tilde{B}_\phi(s)) = 1/2(|t|^\phi + |s|^\phi - |t-s|^\phi).$$

The results below deeply rely upon an *invariance principle, i.e.* a functional form of the Central Limit Theorem and a multi-dimensional Central Limit Theorem. Define for $n \in \mathbb{N}$ and $m \in \mathbb{N}$ the following sequences of polygonal interpolation functions $\{S_n(m, s)\}_{s\in\mathbb{R}}$, where $S_n(m, 0) = 0$ and

$$S_n(m, s) = \begin{cases} \sum_{t=m+1}^{m+[ns]}\varepsilon_t + \varepsilon_{m+[ns]+1}(ns - [ns]) & s > 0 \\ \sum_{t=m+1+[ns]}^{m}\varepsilon_t + \varepsilon_{m+[ns]+1}(ns - [ns]) & s < 0 \end{cases} \tag{40}$$

We assume that

- **H5-a**(ϕ) *(invariance principle)* $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a strict sense stationary process. In addition, there exists a constant $\sigma > 0$ such that, for all $m \in \mathbb{N}$,

$$n^{-\phi/2}S_n(m, s) \underset{n\to\infty}{\Longrightarrow} \sigma\tilde{B}_\phi(s), \qquad s \in [-1, 1],$$

where $\tilde{B}^{(k)}$ is a two-sided fractional Brownian motion. Furthermore, for any sequence of positive integers $\{m_n\}_{n\in\mathbb{N}}$ such that $m_n/n \to \infty$, $\{S_n(1, s)\}_{s\in[-1,+1]}$ and $\{S_n(m_n, s)\}_{s\in[-1,+1]}$ are asymptotically independent.

- **H5-b**(ϕ) (*multi-dimensional CLT*) For any positive integer r and any sequence of non-negative integers $\{m_{n,i}\}_{n\in\mathbb{N}}$, $1 \leqslant i \leqslant r$, such that for all $n \in \mathbb{N}$, $1 < m_{n,1} < \cdots < m_{n,r} < n$, and $\lim_{n\to\infty} m_{n,1}/n = m_i$ with $1 < m_1 < \cdots < m_r < 1$, it holds that

$$n^{-\phi/2}\left(\sum_{t=1}^{m_{1,n}}, \sum_{t=m_{1,n}+1}^{m_{2,n}} \varepsilon_t, \cdots, \sum_{t=m_{n,r}+1}^{n} \varepsilon_t\right) \xrightarrow{d} \mathcal{N}(0, \Gamma)$$

In the above expression, $\Longrightarrow$ denotes the weak-convergence in the space of continuous function on $[-1, +1]$ equipped with the uniform metric, and $\mathcal{N}(0, \Gamma)$ is the $(r+1)$-dimensional multivariate Gaussian distribution with covariance matrix Γ.

Assumption **H5**(ϕ) with $\phi = 1$ is verified for a wide-class of 'short-memory' processes, e.g. for *linear processes* (Eq. (3)) under the assumption that $\sum_{j=0}^{\infty} j|f_j| < \infty$ and $\{\psi_t\}_{t\in\mathbb{Z}}$ is either a sequence of i.i.d. random variables with zero-mean $E(\psi_t) = 0$ and finite variance $E(\psi_t^2) = \sigma^2$, or $\{\psi_t\}$ are martingale increments such that $E(\psi_t^2) = \sigma^2$, $\sup_{t\geqslant 0} E(|\psi_t|^{2+\delta}) < \infty$ and $n^{-1}\sum_{t=1}^{n} E(\psi_t^2|\mathscr{F}_{t-1}) \to \infty$, where $\mathscr{F}_t = \sigma(\psi_s, 1 \leqslant s \leqslant t)$. **H5**($\phi$) also holds with $\phi = 1$ under a wide range of mixing conditions, including mixingales (McLeish (1975), Bai and Perron (1996)), strong-dependent processes (Doukhan, (1994)), under appropriate conditions on the rate of decrease of the mixing coefficients and conditions on the moments (or on the tail of the distribution of ε_t). In all of these situations, the matrix Γ is diagonal.

For our discussion it is more interesting to ask whether these assumptions hold for strongly dependent processes. Invariance principles **H5**(ϕ) have been derived for interpolated sums of non-linear functions of Gaussian variables that exhibit a long-range dependence in Taqqu (1975; 1977). In that case, however, matrix Γ is no longer diagonal. Invariance principles have been also obtained by Taqqu (1977) and Ho and Hsing (1997) for non-Gaussian linear processes (Eq. (3)).

We have the following result:

THEOREM 8. *Assume that* **H1**(ϕ), **H4**(ϕ) *and* **H5**(ϕ) *hold for some* $\phi < 2$. *Then, for any* $1 \leqslant j \leqslant r$,

$$\underline{\lambda}_n^{2/2-\phi}(\hat{t}_{n,j} - t_j^*) \xrightarrow{d} \sigma^{2/2-\phi} \operatorname*{argmin}_{\nu\in\mathbb{R}}(|\nu| + 2\tilde{B}_\phi^{(j)}(v)) \qquad (41)$$

where $\tilde{B}_\phi^{(j)}$ *is a two-sided fractional Brownian motion. Furthermore,* $\hat{t}_{n,j}$ *and* $\hat{t}_{n,k}$ *are asymptotically independent if* $j \neq k$.

Denote $\overline{\boldsymbol{\mu}}_n$ *the vector of sample means:* $\overline{\boldsymbol{\mu}}_n \triangleq (\overline{Y}(1, \hat{t}_1), \overline{Y}(\hat{t}_1 + 1, \hat{t}_2), \cdots, \overline{Y}(\hat{t}_r + 1, n))$ *and* $\hat{\boldsymbol{\mu}}_n^* \triangleq (\overline{Y}(1, t_1^*), \overline{Y}(t_1^* + 1, t_2^*), \cdots, \overline{Y}(t_r^* + 1, n))$. *Then,* $n^{1-\phi/2}(\hat{\boldsymbol{\mu}}_n - \hat{\boldsymbol{\mu}}_n^{\Delta}) = o_P(1)$. *Assume in addition that* **H5-b**(ϕ) *holds. Then,* $n^{1-\phi/2}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}^*)$ *is asymptotically normal.*

Not surprisingly the limiting distribution of $\hat{\mathbf{t}}_\mathbf{n}$ depends upon the memory parameter ϕ through the normalizing constant $\gamma = 2/(2-\phi)$. This result is fairly intuitive: for a given value of the memory coefficient ϕ, the spread of the estimated change-point instant $\hat{t}_{n,j}$ increases as the magnitude of the jump λ_j

decreases; on the other hand, for a given jump magnitude λ_j, the spread of $\hat{t}_{n,j}$ increases as ϕ increases.

PROOF OF THEOREM 8: The proof is adapted from Bai (1994), Theorem 1. Denote $\boldsymbol{s} = (s_1 \cdots s_r) \in [-M, +M]$ and define

$$\tilde{K}_n(\boldsymbol{s}) \stackrel{\triangle}{=} K_n([t_1^* + s_1\lambda_{n,1}^{-\gamma}], \cdots, [t_r^* + s_r\lambda_{n,r}^{-\gamma}])$$

and define similarly $\tilde{V}_n(\boldsymbol{s})$, $\tilde{W}_n(\boldsymbol{s})$ and $\tilde{J}_n(\boldsymbol{s})$. The following result plays a key role in the derivations that follow.

LEMMA 3.2. *Assume that* $\mathbf{H1}(\phi)$ *holds for* $1 < \phi < 2$. *Then,*

$$\sup_{\boldsymbol{s}\in[-M,M]^r} \left(n\underline{\lambda}_n^{\gamma-2} \tilde{K}_n(\boldsymbol{s}) - \sum_{k=1}^{r} s_k a_k^{2-\gamma} \right) = \mathrm{o}(1), \tag{42}$$

$$\sup_{\boldsymbol{s}\in[-M,M]^r} \left(n\underline{\lambda}_n^{\gamma-2} \tilde{V}_n(\boldsymbol{s}) \right) = \mathrm{o}(1), \tag{43}$$

$$\sup_{\boldsymbol{s}\in[-M,M]^r} \left(n\underline{\lambda}_n^{\gamma-2} \tilde{W}_n(\boldsymbol{s}) - 2\sum_{k=1}^{r} a_k^{2-\gamma} \lambda_{n,k}^{\gamma\phi/2} S_{\lambda_{n,k}^{-\gamma}}(t_k^*, s_k) \right) = \mathrm{o}(1) \tag{44}$$

where $S_{\lambda_{n,k}^{-\gamma}}(t_k^*, s_k)$ *is defined in* (40).

Under assumption $(\mathbf{H5\text{-}a})(\phi)$, the process $\sum_{k=1}^r a_k^{2-\gamma} \lambda_{n,k}^{\gamma\phi/2} S_{\lambda_{n,k}^{-\gamma}}(t_k^*, s_k)$ converges (in the space of continuous function on $[-M, M]^r$ equipped with the supremum norm) to $\sigma \sum_{k=1}^r a_k^{2-\gamma} \tilde{B}_\phi^{(k)}(s_k)$, where $(\tilde{B}_\phi^{(k)}, 1 \leqslant k \leqslant r)$ are r independent copies of the fractional Brownian motion with self-similarity parameter $\phi/2$. Then, Lemma 3.2 implies that the process $\tilde{J}_n(s)$ (with polygonal interpolation) converges to $\sum_{k=1}^r a_k^{2-\gamma}(|s_k| + 2\sigma \tilde{B}_\phi^{(k)}(s_k))$. Let $v \stackrel{\triangle}{=} \sigma^{2/(2-\phi)} s$. Then, $\mathrm{argmin}_s(|s| + 2\sigma \tilde{B}\phi(s)) = \sigma^{2/(2-\phi)} \mathrm{argmin}_v(|v| + 2\tilde{B}_\phi(v))$. The result follows from the continuous mapping with argmin functionals, see Kim and Pollard (1990) (using the same arguments as in Bai (1994), Theorem 1).

Given the rate of convergence of the break dates, it is an easy matter to derive the asymptotic distributions of the sample mean, along the same lines as in Bai (1994). ■.

4. ESTIMATION OF THE NUMBER OF CHANGE-POINTS

In the previous section, the number of change-points is assumed to be known. In many applications however, the number of break fractions is not specified in advance, and inference about this parameter also is important. Estimation of the number of break points has been addressed by Yao (1988), which suggests the use of the Bayesian Information Criterion (also known as the Schwarz criterion).

Yao (1988) has shown that the estimate of the number of break points is consistent when the disturbance is a *Gaussian* white noise. This work has later been expanded by Liu, Wu and Zidek (1997). In both cases, the basic idea consists in adding a penalty term to the least-square criterion, in order to avoid over-segmentation.

The following result is a direct extension of Theorem 3:

LEMMA 4.1. *For any $r \geqslant 0$ and for any r-partition $\mathbf{t} \in \mathscr{A}_{n,r}$, let $\|\mathbf{t} - \mathbf{t}^*\|_\infty = \max_{1 \leqslant j \leqslant r^*} \min_{0 \leqslant k \leqslant r+1} |t_k - t_j^*|$. If $r \geqslant r^*$ and under the assumptions of Theorem 3, $\|\hat{\boldsymbol{\tau}}_\mathbf{n} - \boldsymbol{\tau}^*\|_\infty \to 0$ when $n \to \infty$. Moreover, the uniform bound* (17) *still holds.*

The proof is a straightforward adaptation of Theorem 3 and is omitted. Lemma 4.1 means that, even if the number of changes has been over-estimated, then, a sub-family $(\hat{\tau}_{k_1}, \hat{\tau}_{k_2}, \cdots, \hat{\tau}_{k_{r^*}})$ of $\hat{\boldsymbol{\tau}}_\mathbf{n}$ still converges to the true set of change fractions $\boldsymbol{\tau}^*$. Note also that the uniform bounds Eqs. (24) and (25) hold whatever the number r of estimated break fractions is. They even hold uniformly (w.r.t. to the number of break fractions), if this number is upper bounded.

We propose to estimate the configuration of change-points $\mathbf{t}^*$ and the number of changes r^* by minimizing a penalized least-square procedure. For any $R \in \mathbb{N}$, for any $\Delta \geqslant 0$ and any $\Theta_r \subset \mathbb{R}^{r+1}$, we consider the following estimates of $(\mathbf{t}^*, r^*)$:

$$(\hat{\mathbf{t}}_\mathbf{n}, \hat{r}_n) = \operatorname*{argmin}_{0 \leqslant r \leqslant R} \operatorname*{argmin}_{\mathbf{t} \in \mathscr{A}_{n,r}} \{Q_n(\mathbf{t}) + \beta_n r\}, \tag{45}$$

$$(\hat{\mathbf{t}}^\Delta_\mathbf{n}, \hat{r}^\Delta_n) = \operatorname*{argmin}_{0 \leqslant r \leqslant R} \operatorname*{argmin}_{\mathbf{t} \in \mathscr{A}^\Delta_{n,r}} \{Q_n(\mathbf{t}) + \beta_n r\}, \tag{46}$$

$$(\hat{\mathbf{t}}^\Theta_\mathbf{n}, \hat{r}^\Theta_n) = \operatorname*{argmin}_{0 \leqslant r \leqslant R} \operatorname*{argmin}_{\mathbf{t} \in \mathscr{A}_{n,r}} \{Q^\Theta_n(\mathbf{t}) + \beta_n r\}, \tag{47}$$

where the contrast functions Q_n and Q_n^Θ were defined in (16) and (28), and where $\{\beta_n\}$ is a decreasing sequence of positive real numbers. We denote $\hat{\boldsymbol{\tau}}_\mathbf{n}$, $\hat{\boldsymbol{\tau}}^\Delta_\mathbf{n}$ and $\hat{\boldsymbol{\tau}}^\Theta_\mathbf{n}$ the associated estimators of the break-fractions. The choice of the penalty is discussed in the next theorem:

THEOREM 9.

i) Assume that $\mathbf{H1}(\phi)$ *is satisfied for some $\phi < 2$. Then, for any sequence $\{\beta_n\}$ such that $\beta_n \to 0$ and $n\beta_n \to \infty$, and such that* $\mathbf{H3}(\beta)$ *holds, $(\hat{\boldsymbol{\tau}}_\mathbf{n}, \hat{r}_n)$ converges in probability to $(\boldsymbol{\tau}^*, r^*)$ if $r^* \leqslant R$.*

ii) Assume that $\mathbf{H1}(\phi)$ *is satisfied for some $\phi < 2$. Then, for any non increasing sequences $\{\Delta_n\}$ and $\{\beta_n)$ such that $\Delta_n \to 0$, $\beta_n \to 0$ and $n^{2-\phi}\Delta_n\beta_n \to +\infty$, $(\hat{\boldsymbol{\tau}}^{\Delta_n}_\mathbf{n}, \hat{r}^{\Delta_n}_n)$ converges in probability to $(\boldsymbol{\tau}^*, r^*)$ if $r^* \leqslant R$.*

iii) Assume that $\mathbf{H1}(\phi)$ *is satisfied for some $\phi < 2$. Assume also that $n^{2-\phi}\beta_n \to +\infty$. Then, for any compact subset Θ of $\mathbb{R}^{r+1}$, $(\hat{\boldsymbol{\tau}}^\Theta_\mathbf{n}, \hat{r}^\Delta_n)$ converges* *in probability to* $(\boldsymbol{\tau}^*, r^*)$ *if* $r^* \leqslant R$ *provided* $\boldsymbol{\mu}^* = (\mu_1^*, \ldots, \mu_{r^*+1}^*) \in \Theta$.

iv) Assume that **H2** *is satisfied. Assume also that* $\beta_n \geqslant n^{-\psi}$ *where*

$$\psi < \frac{\rho(\gamma - 1) - 4\gamma}{\rho(1+\gamma)}. \tag{48}$$

Then, for any $\Delta > 0$, $(\hat{\boldsymbol{\tau}}_{\mathbf{n}}^{\Delta}, \hat{r}_n^{\Delta})$ *converges almost-surely to* $(\boldsymbol{\tau}^*, r^*)$ *if* $r^* \leqslant R$ *provided* $\Delta \leqslant \Delta_\tau^*$.

Proof: We show (*i*) first. To establish the convergence in probability of $(\hat{\boldsymbol{\tau}}_{\mathbf{n}}, \hat{r}_n)$, it suffices to show that $\mathrm{P}(\hat{r}_n \neq r^*)$ goes to 0. By definition of $(\hat{\mathbf{t}}_{\mathbf{n}}, \hat{r}_n)$, we have

$$Q_n(\hat{\mathbf{t}}_{\mathbf{n}}) + \beta_n \hat{r}_n \leqslant Q_n(\mathbf{t}^*) + \beta_n r^*. \tag{49}$$

Using (18) and (22), this latter relation implies

$$K_n(\hat{\mathbf{t}}_{\mathbf{n}}) + V_n(\hat{\mathbf{t}}_{\mathbf{n}}) + W_n(\hat{\mathbf{t}}_{\mathbf{n}}) + \beta_n(\hat{r}_n - r^*) \leqslant 0. \tag{50}$$

Then,

$$\begin{aligned} \mathrm{P}(\hat{r}_n = r) &\leqslant \mathrm{P}(K_n(\hat{\mathbf{t}}_{\mathbf{n}}) + V_n(\hat{\mathbf{t}}_{\mathbf{n}}) + W_n(\hat{\mathbf{t}}_{\mathbf{n}}) + \beta_n(r - r^*) \leqslant 0), \\ &\leqslant \mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{A}_{n,r}} (K_n(\mathbf{t}) + V_n(\mathbf{t}) + W_n(\mathbf{t})) + \beta_n(r - r^*) \leqslant 0\right). \end{aligned} \tag{51}$$

Note that under $\mathbf{H3}(\beta)$ and (20), it holds that, for any $\delta > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{\mathbf{t}\in\mathscr{A}_{n,r}} |V_n(\mathbf{t})| \geqslant \delta\beta_n\right) = 0. \tag{52}$$

Assume first that $r < r^*$. Note that, for $r < r^*$, we have $\|\hat{\mathbf{t}}_{\mathbf{n}} - \mathbf{t}^*\|_\infty \geqslant \Delta_r^*/r^*$, which implies that

$$K_n(\hat{\mathbf{t}}_{\mathbf{n}}) \geqslant \frac{\Delta_\tau^*}{r^*}\underline{\lambda}^2. \tag{53}$$

Then, for any $r < r^*$,

$$\begin{aligned} \mathrm{P}(\hat{r}_n = r) &\leqslant \mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{A}_{n,r}} (V_n(\mathbf{t}) + W_n(\mathbf{t})) + \frac{\Delta_\tau^*}{2r^*}\underline{\lambda}^2 - \beta_n(r^* - r) \leqslant 0\right) \\ &\leqslant \mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{A}_{n,r}} |V_n(\mathbf{t})| \geqslant \Delta_\tau^*\underline{\lambda}^2/(2r^*)\right) + \mathrm{P}\left(\max_{\mathbf{t}\in\mathscr{A}_r} |W_n(\mathbf{t})| \geqslant \Delta_\tau^*\right). \end{aligned} \tag{54}$$

and $\mathrm{P}(\hat{r}_n = r) \to 0$ as $n \to \infty$, by application of (24), (25) and Lemma 2.2.

On the other hand, for any $r^* \leqslant r \leqslant R$,

$$\mathrm{P}(\hat{r}_n = r) \leqslant \mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{A}_{n,r}}\left(V_n(\mathbf{t}) + \frac{\beta_n}{2}\right) \leqslant 0\right) + \mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{A}_{n,r}}(W_n(\mathbf{t}) + K_n(\mathbf{t})) \geqslant \frac{\beta_n}{2}\right). \tag{55}$$

From (52), the first term in the right hand term of (55) goes to 0 when $n \to \infty$. The second term requires additional attention. For any $\mathbf{t} \in \mathscr{A}_{n,r}$, we have

$$K_n(\mathbf{t}) = \frac{1}{n}\sum_{k=1}^{r+1}\sum_{i=1}^{r^*+1}\sum_{j=1}^{r^*+1}\frac{n_{ki}n_{kj}}{n_k}\lambda_{ij}^2, \tag{56}$$

$$W_n(\mathbf{t}) = \frac{1}{2n}\sum_{k=1}^{r+1}\sum_{i=1}^{r^*+1}\sum_{j=1}^{r^*+1}\frac{n_{ki}}{n_k}\lambda_{ij}S_{kj}. \tag{57}$$

where $\lambda_{ij} \triangleq |\mu_i^* - \mu_j^*|$, and n_{kj} and S_{kj} are defined in Eqs. (31) and (32), respectively (the dependence of this quantity upon the r-partition $\mathbf{t}$ is implicit). Then, for any $1 \leqslant k \leqslant r$ and any $1 \leqslant i, j \leqslant r^*$, let

$$\mathscr{C}_{kj} = \{\mathbf{t} \in \mathscr{A}_{n,r},\ n_{kj} \leqslant n\beta_n\},$$

$$\overline{\mathscr{C}_{kj}} = \{\mathbf{t} \in \mathscr{A}_{n,r},\ n_{kj} \leqslant n\beta_n\}.$$

From Lemma 2.2, there exist two constants $A_1 > 0$ and $A_2 > 0$ such that, for any $c > 0$, for any $1 \leqslant k \leqslant r$ and any $1 \leqslant i, j \leqslant r^*$, such that $\lambda_{ij} > 0$,

$$\begin{aligned}\mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{C}_{kj}}\frac{n_{ki}}{nn_k}(\mu_i^* - \mu_j^*)S_{kj} + c(K_n(\mathbf{t}) + \beta_n) \leqslant 0\right) &\leqslant \mathrm{P}\left(\max_{n_{kj}\geqslant n\beta_n}\frac{|S_{kj}|}{n_{kj}} \geqslant c\lambda_{kj}\right)\\ &\leqslant \frac{A_1}{c^2}(n\beta_n)^{\phi-2}\end{aligned} \tag{58}$$

$$\begin{aligned}\mathrm{P}\left(\min_{\mathbf{t}\in\overline{\mathscr{C}_{kj}}}\frac{n_{ki}}{nn_k}(\mu_i^* - \mu_j^*)S_{kj} + c(K_n(\mathbf{t}) + \beta_n) \leqslant 0\right) &\leqslant \mathrm{P}\left(\max_{n_{kj}\geqslant n\beta_n}|S_{kj}| \geqslant \frac{c}{\lambda_{ij}}n\beta_n\right)\\ &\leqslant \frac{A_2}{c^2}(n\beta_n)^{\phi-2}.\end{aligned} \tag{59}$$

From (58) and (59), and using the fact that $n\beta_n \to +\infty$, we conclude that

$$\lim_{n\to\infty}\mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{A}_{n,r}}(W_n(\mathbf{t}) + K_n(\mathbf{t})) \geqslant \frac{\beta_n}{2}\right) = 0. \tag{60}$$

Finally, for any $0 \leqslant r \leqslant R$, $\mathrm{P}(\hat{r}_n = r) \to 0$ when $n \to \infty$ if $r \neq r^*$. *(ii)* and *(iii)* can be shown along the same lines

We finally show the strong consistency of $(\hat{\boldsymbol{\tau}}_\mathbf{n}, \hat{r}_\mathbf{n})$ using the fact that, for any $r \neq r^*$, $\sum_{n>0}\mathrm{P}(\hat{r}_\mathbf{n} = r) < \infty$ under **H2**, if (β_n) satisfies (48). Indeed, $n\beta_n > n^{1-\psi}$ with

$$1-\psi > \frac{\rho(3+\gamma)+4\gamma}{2\rho(1+\gamma)}.$$

We can apply Theorem 2 to $P(\hat{r}_n = r)$, with $r > r^*$, by setting $u_n = n\beta_n$. On the other hand, we can apply Theorem 2 to $P(\hat{r}_n = r)$, with $r < r^*$ by setting $u_n = n$ since $\beta_n \to 0$. ■

5. NUMERICAL EXAMPLES

In this section, we present a limited Monte-Carlo experiment. The disturbance ε is a fractional Gaussian noise, *i.e.* a covariance stationary process with zero-mean, and spectral density function given by

$$f(\lambda) = \frac{\sigma^2}{2\pi}(\sin\lambda)^{-2d}$$

d is the long-range dependence parameter. We use the Hoskings's method to simulate this time series (1992). In the simulation, we set $\sigma^2 = 1$ and $d = 0.3$. There are two change-points at time $\tau_1^* = 0.25$ and $\tau_2^* = 0.5$. The function μ is defined as follows:

$$\mu(t) = \begin{cases} 2 & 0 \leqslant t < 0.25 \\ 0 & 0.25 \leqslant t < 0.5 \\ 1 & 0.5 \leqslant t \leqslant 1 \end{cases}$$

We simulate 50 realizations of the sequence $Y_1, \ldots Y_n$, with different values of n. A typical realization ($n = 1000$ samples) is displayed in Figure 1.

Since the number of change-points is assumed to be unknown, the penalized least-squares estimator has been computed for each one of these realizations. If the statistical structure of the process were exactly known, an upper bound of the regularization factor could be computed:

$$\beta_n = 4\log(n)/n^{1-2d}.$$

Not surprisingly, the penalization is typically higher than in the i.i.d. case (see Yao (1988)). The histograms of the estimated change-points are displayed in Figure 2 while Table 1 gives the estimated numbers of change-points.

The coefficient of penalization β_n has been chosen by minimizing (45) in order to obtain approximatively the same number of over- and under-estimation of the change-points. It is clearly seen in this example that the estimated number of changes converges to the true value $r^* = 2$ when n increases. We remark also that the distribution of the estimated change-points concentrates around the true change-points $\tau_1^* = 0.25$ and $\tau_2^* = 0.5$ when $n \to \infty$.

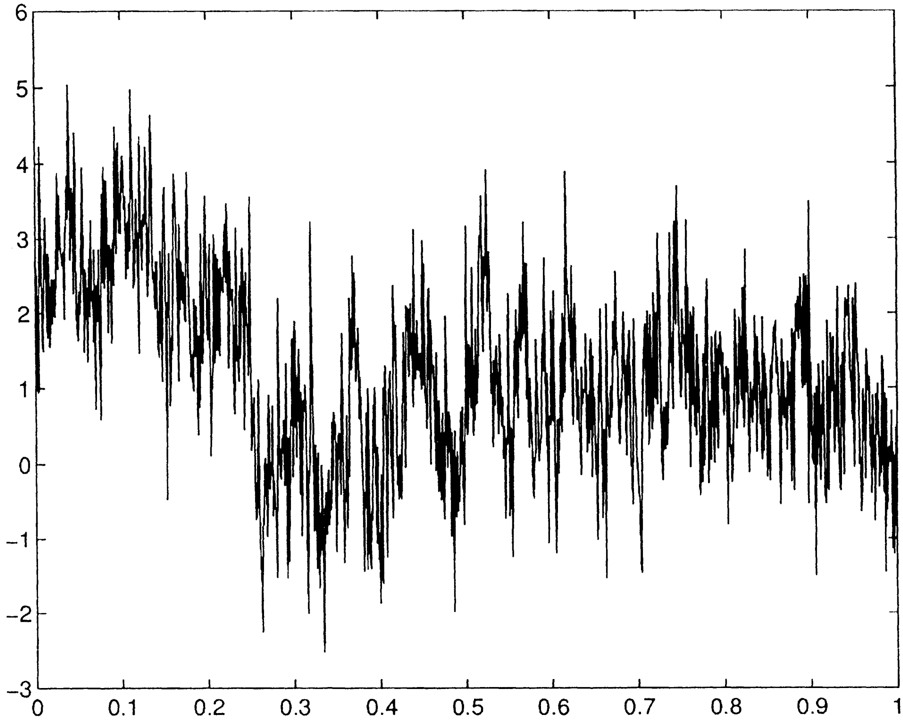

FIGURE 1. A realization of the process Y with $n = 1000$ and with change-points at $\tau_1^* = 0.25$ and $\tau_2^* = 0.5$.

6. ACKNOWLEDGEMENTS

The authors are deeply indebted to the two anonymous reviewers for their very constructive suggestions and for drawing our attention to relevant works, in particular the reference Bai and Perron (1996) which was a fruitful source of inspiration.

7. PROOFS

PROOF OF THEOREM 1: The proof of proposition (5) directly follows the proof proposed by Móricz, Serfling and Stout, (1982). Thus, for any $m \leqslant n$, we have:

$$\mathrm{P}\Big(\max_{m \leqslant k \leqslant n} b_k |S_{1:k}| > \delta\Big) \leqslant \mathrm{P} b_m |S_{1:m}| > \delta/2) + \mathrm{P}\Big(\max_{m+1 \leqslant k \leqslant n} b_k |S_{m+1:k}| > \delta/2\Big),$$

$$\leqslant 4 \frac{C(\varepsilon) m^{\phi} b_m^2}{\delta^2} + 4A(\phi) \frac{C(\varepsilon)}{\delta^2} (n-m)^{\phi - 1} \sum_{t=m+1}^{n} b_t^2. \quad \blacksquare \quad (61)$$

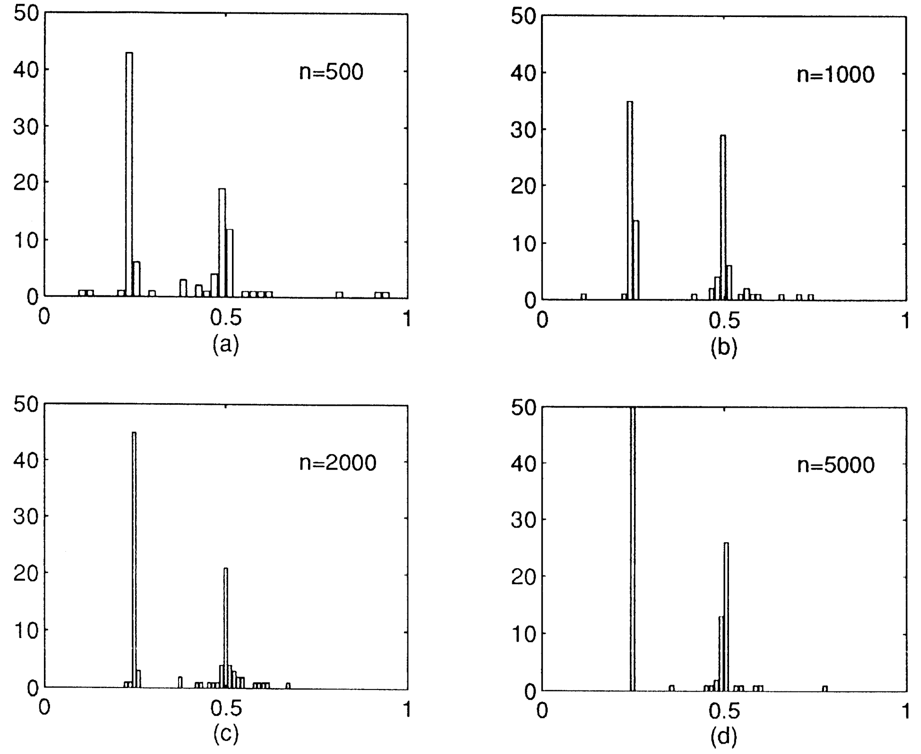

FIGURE 2. The empirical distributions of the estimated change-points obtained with different values of n: (a) $n = 500$, (b) $n = 1000$, (c) $n = 2000$, (d) $n = 5000$.

TABLE 1

ESTIMATION OF THE NUMBER r OF CHANGES, WITH DIFFERENT VALUES OF n. FOR EXAMPLE, 0.86 MEANS THAT, WITH $n = 1000$, WE OBTAINED 43 TIMES AN ESTIMATE $\hat{r} = 2$ AMONG THE 50 SIMULATIONS

n	β_n	$\hat{r} = 1$	$\hat{r} = 2$	$\hat{r} = 3$	$\hat{r} = 4$
500	20	0·12	0.74	0.14	0.00
1000	40	0.06	0.86	0.06	0.02
2000	100	0.04	0.96	0.00	0.00
5000	200	0.02	0.98	0.00	0.00

PROOF OF COROLLARY 2.1: Using Theorem 1, we have, for any $p \geqslant 0$, $m > 0$ and $\phi > 1$:

$$\mathrm{P}\Big(\max_{2^{\mathrm{p}} m \leqslant k < 2^{\mathrm{p}+1} m} b_k |S_{1:k}| > \delta\Big) \leqslant 4 \frac{C(\varepsilon)}{\delta^2} (2^{\mathrm{p}} m)^{\phi} b_{2^{\mathrm{p}} m}^2 + 4 \frac{A(\phi) C(\varepsilon)}{\delta^2} (2^{\mathrm{p}} m)^{\phi - 1} \sum_{t = 2^{\mathrm{p}} m + 1}^{2^{\mathrm{p}+1} m - 1} b_t^2 .$$

For $b_k = k^{-\beta}$ and $\beta > \phi/2$, we have:

$$\sum_{p=0}^{\infty}(2^{\mathrm{p}} m)^{\phi} b_{2^{\mathrm{p}} m}^2 = m^{\phi-2\beta} \sum_{p=0}^{\infty} 2^{\mathrm{p}(\phi-2\beta)} \leqslant \frac{m^{\phi-2\beta}}{1-2^{\phi-2\beta}}$$

On the other hand, there exists a constant $D < \infty$ such that:

$$\sum_{p=0}^{\infty}(2^{\mathrm{p}} m)^{\phi-1} \sum_{t=2^{\mathrm{p}} m+1}^{2^{\mathrm{p}+1} m-1} b_t^2 \leqslant D m^{(\phi-2\beta)}.$$

We conclude using the fact that:

$$\mathrm{P}\Big(\max_{k \geqslant m} b_k |S_{1:k}| > \delta\Big) \leqslant \sum_{p+0}^{\infty} \mathrm{P}\Big(\max_{2^{\mathrm{p}} m \leqslant k < 2^{\mathrm{p}+1} m} b_k |S_{1:k}| > \delta\Big). \qquad \blacksquare$$

Proof of Equation (36): The set $\mathscr{C}'_{\delta,\gamma,n}$ can be decomposed as $\mathscr{C}'_{\delta,\gamma,n} = \bigcup_{\mathscr{T}} \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})$, where the union is over all the subsets $\mathscr{T}$ of $\{1, \cdots, r\}$, and

$$\mathscr{C}'_{\delta,\gamma,n}(\mathscr{T}) = \{\mathbf{t} \in \mathscr{A}_{n,r},\ \delta \underline{\lambda}^{2/\phi-2} \leqslant t_k - t_k^* \leqslant n\gamma\Delta_\tau^*,$$

$$\forall k \in \mathscr{T},\ 0 \leqslant t_k - t_k^* < \delta\underline{\lambda}^{2/\phi-2},\ \forall k \notin \mathscr{T}\}$$

For any $\mathbf{t} \in \mathscr{C}'_{\delta,\gamma,n}$, denote $n_{kk} = t_k^* - t_{k-1}$, $n_{k,k+1} = t_k - t_k^*$, $n_k = t_k - t_{k-1}$ and $n_k^* = t_k^* - t_{k-1}^*$; the dependence of these quantities on $\mathbf{t}$ and $\mathbf{t}^*$ is implicit. Note that $n_k = n_{k,k} + n_{k,k+1}$ and $n_k^* = n_{k,k} + n_{k-1,k}$, and $n_{k,k}/n_k \geqslant (1-\gamma)\Delta_\tau^*$. For all $\mathbf{t} \in \mathscr{C}'_{\delta,\gamma,n}$, one may show that

$$K_n(\mathbf{t}) = \frac{1}{n}\sum_{k=1}^{r} \frac{n_{kk} n_{k,k+1}}{n_k} \lambda_k^2, \tag{62}$$

$$V_n(\mathbf{t}) = \frac{1}{n}\sum_{k=1}^{r} n_{k,k+1}\left(\frac{n_{kk}}{n_k}\left(\frac{S_{kk}}{n_{kk}} - \frac{S_{k,k+1}}{n_{k,k+1}}\right)^2 - \frac{n_{k+1,k+1}}{n_{k+1}^*}\left(\frac{S_{k+1,k+1}}{n_{k+1,k+1}} - \frac{S_{k,k+1}}{n_{k,k+1}}\right)^2\right) \tag{63}$$

$$W_n(\mathbf{t}) = \frac{2}{n}\sum_{k=1}^{r} \lambda_k \left(n_{k,k+1}\frac{S_{kk}}{n_k} + \frac{n_{kk}}{n_k} S_{kk+1}\right). \tag{64}$$

where, for $1 \leqslant i \leqslant j \leqslant r+1$, $S_{ij} = \sum_{t=t_{i-1}+1}^{t_j} \varepsilon_t$ and $\lambda_k = \mu_{k+1}^* - \mu_k^*$. In addition, we have, for all $\mathbf{t} \in \mathscr{C}'_{\delta,\gamma,n}$,

$$\min_{\mathbf{t} \in \mathscr{C}'_{\delta,\gamma,n}} K_n(\mathbf{t}) \geqslant (1-\gamma)\Delta_\tau^* \delta \underline{\lambda}^{2-2\phi/2-\phi}, \tag{65}$$

$$V_n(\mathbf{t}) \geqslant -\frac{1}{n}\sum_{k=1}^{r}\left(n_{k,k+1}\frac{S_{k+1,k+1}^2}{n_{k+1,k+1}^2}+\frac{S_{k,k+1}^2}{n_k}+2|S_{k,k+1}|\left(\frac{|S_{k+1,k+1}|}{n_{k+1,k+1}}+\frac{|S_{k,k}|}{n_{k,k}}\right)\right) \tag{66}$$

Using these expressions for $K_n(\mathbf{t})$, $V_n(\mathbf{t})$ and $W_n(\mathbf{t})$, and the above bounds, we obtain sharper bounds for $J_n(\mathbf{t})$ on $C'_{\delta,\gamma,n}$. First, by using Lemma 2.2 and its corollary, these exist finite constants C_1, C_2, C_3 such that, for all $1 \leqslant k \leqslant r$ and for all $c > 0$,

$$\mathrm{P}\left(\max_{\mathbf{t}\in\mathscr{C}'_{\delta,\gamma,n}}\frac{|S_{kk}|}{n_{kk}} \geqslant c\right) \leqslant \sup_{t\in\mathbb{Z}}\mathrm{P}\left(\sup_{s\geqslant n(1-\gamma)\Delta_\tau^*}\frac{|\sum_{i=t-s+1}^{t}\varepsilon_i|}{s} \geqslant c\right) \leqslant C_1\frac{n^{\phi-2}}{c^2}, \tag{67}$$

$$\mathrm{P}\left(\max_{\mathbf{t}\in\mathscr{C}'_{\delta,\gamma,n}}\frac{S_{kk}^2}{n_{kk}} \geqslant c\right) \leqslant \sup_{t\in\mathbb{Z}}\mathrm{P}\left(\sup_{s\leqslant n}\left|\sum_{i=t-s+1}^{t}\varepsilon_i\right| \geqslant \sqrt{n(1-\gamma)c\Delta_\tau^*}\right) \leqslant C_2\frac{n^{\phi-1}}{c} \tag{68}$$

$$\mathrm{P}\left(\max_{\mathbf{t}\in\mathscr{C}'_{\delta,\gamma,n}}\frac{S_{k,k+1}^2}{n_{k,k+1}} \geqslant c\right) \leqslant \sup_{t\in\mathbb{Z}}\mathrm{P}\left(\max_{1\leqslant s\leqslant n\gamma\Delta_\tau^*}\frac{(\sum_{i=t+1}^{t+s}\varepsilon_i)^2}{s} \geqslant c\right) \leqslant C_3\frac{n^{\phi-1}}{c}. \tag{69}$$

Let $\mathscr{T}$ be an arbitrary subset of $\{1, \cdots, r\}$. There exist finite constants C_4, C_5 (that do not depend upon $\underline{\lambda}$, $\overline{\lambda}$, nor on the subset $\mathscr{T}$) such that, for all $n \geqslant 1$

$$\begin{aligned}\mathrm{P}\left(\max_{k\in\mathscr{T}}\max_{\mathbf{t}\in\mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})}\frac{|S_{k,k+1}|}{n_{k,k+1}} \geqslant c\right) &\leqslant \sup_{t\in\mathbb{Z}}\mathrm{P}\left(\max_{s\geqslant\delta\underline{\lambda}^{2/\phi-2}}\frac{|\sum_{i=t+1}^{t+s}\varepsilon_i|}{s} \geqslant c\right)\\ &\leqslant C_4\frac{\delta^{\phi-2}}{c^2}\underline{\lambda}^2,\end{aligned} \tag{70}$$

$$\begin{aligned}\mathrm{P}\left(\max_{k\notin\mathscr{T}}\max_{\mathbf{t}\in\mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})}|S_{k,k+1}| \geqslant c\right) &\leqslant \sup_{t\in\mathbb{Z}}\mathrm{P}\left(\max_{k\notin\mathscr{T}}\max_{0\leqslant s\leqslant\delta\underline{\lambda}^{2/\phi-2}}\left|\sum_{i=t+1}^{t+s}\varepsilon_i\right| \geqslant c\right)\\ &\leqslant C_5\frac{\delta^{\phi}}{c^2}\underline{\lambda}^{2\phi/\phi-2}.\end{aligned} \tag{71}$$

Using (64), (65) and (66), there exists $c > 0$ small enough so that it holds

$$\mathrm{P}(\hat{\mathbf{t}}_\mathbf{n} \in \mathscr{C}'_{\delta,\gamma,n}) \leqslant \mathrm{P}\left(\min_{\mathbf{t}\in\mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})}(K_n(\mathbf{t}) + V_n(\mathbf{t}) + W_n(\mathbf{t})) \leqslant 0\right)$$

$$\leqslant \sum_{k=1}^{r} \mathrm{P}\left(c(1-\gamma)\Delta_\tau^* \underline{\lambda}^2 \leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} S_{k+1,k+1}^2 / n_{k+1,k+1}^2 \right)$$

$$+ \sum_{k \in \mathscr{T}} \mathrm{P}\left(c(1-\gamma) n \Delta_\tau^* \underline{\lambda}^2 \leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} S_{k,k+1}^2 / n_{k,k+1} \right)$$

$$+ \sum_{k \notin \mathscr{T}} \mathrm{P}\left(c(1-\gamma)\Delta_\tau^* \delta \underline{\lambda}^{2-2\phi/2-\phi} \leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} S_{k,k+1}^2 / n_k \right)$$

$$+ \sum_{k \in \mathscr{T}} \mathrm{P}\big(c(1-\gamma)\Delta_\tau^* \underline{\lambda}^2$$

$$\leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} |S_{k,k+1}| / n_{k,k+1} (|S_{k+1,k+1}| / n_{k+1,k+1} + |S_{k,k}| / n_{k,k}) \Big)$$

$$+ \sum_{k \notin \mathscr{T}} \mathrm{P}\Big(c(1-\gamma)\Delta_\tau^* \delta \underline{\lambda}^{2-2\phi/2-\phi}$$

$$\leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} |S_{k,k+1}| (|S_{k+1,k+1}| / n_{k+1,k+1} + |S_{k,k}| / n_{k,k}) \Big)$$

$$+ \sum_{k=1}^{r} \mathrm{P}\left(c\underline{\lambda} \leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} |S_{k,k}| / n_{k,k} \right)$$

$$+ \sum_{k \in \mathscr{T}} \mathrm{P}\left(c\underline{\lambda} \leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} |S_{k,k+1}| / n_{k,k+1} \right)$$

$$+ \sum_{k \notin \mathscr{T}} \mathrm{P}\left(c(1-\gamma)\Delta_\tau^* \delta \underline{\lambda}^{2\phi-2/\phi-2} \overline{\lambda}^{-1} \leqslant \max_{t \in \mathscr{C}'_{\delta,\gamma,n}(\mathscr{T})} |S_{k,k+1}| \right). \tag{73}$$

The proof is concluded by bounding each term in the previous sum using relations (67)–(71). ■

Proof of Lemma 3.2 We prove these properties for the positive orthant: $0 \leqslant s_i \leqslant M,\ 1 \leqslant i \leqslant r$ because of symmetry. Let

$$\mathscr{K}_n(M) \stackrel{\triangle}{=} \{\mathbf{t} \in \mathscr{A}_{n,r}^{\Delta_n},\ t_k = [t_k^* + s_k \lambda_{n,k}^{-1}],\ 0 \leqslant s_k \leqslant M,\ k \in \{1, \cdots, r\}\}.$$

Eq. (42) follows directly from (62). To prove Eq. (43), first note that (63) implies that

$$n\tilde{V}_n(s) \leqslant 88 \sum_{k+1}^{r} n_{k,k+1} \left(\frac{S_{k,k}^2}{n_{k,k}^2} + \frac{S_{k+1,k+1}^2}{n_{k+1,k+1}^2} \right) + 2|S_{k,k+1}| \left(\frac{|S_{k,k}|}{n_{k,k}} + \frac{|S_{k+1,k+1}|}{n_{k+1,k+1}} \right)$$

$$+ S_{k,k+1}^2 \left(\frac{1}{n_k} + \frac{1}{n_{k+1}} \right) \tag{74}$$

where, as before, $n_{ij} = \#(\mathscr{T}_{ij})$, with

$$\mathscr{T}_{ij} = \{[t_{i-1}^* + s_{i-1}\lambda_{n,i}^{-\gamma}] + 1, [t_i^* + s_i\lambda_{n,i}^{-\gamma}]\} \cap \{[t_{j-1} + s_{j-1}\lambda_{n,j}^{-\gamma}] + 1, [t_j + s_j\lambda_{n,j}^{-\gamma}]\}$$

and $S_{ij} = \sum_{t \in \mathscr{T}_{ij}} \varepsilon_t$. Now, for any $\delta > 0$, we have

$$(n\underline{\lambda}_n^{\gamma-2}|\tilde{V}_n(\mathbf{s})| \geqslant \delta) \leqslant \sum_{k+1}^{r} \mathrm{P}\left(\max_{\mathbf{t} \in \mathscr{K}_n(M)} \frac{S_{k,k}^2}{n_{k,k}^2} \geqslant c\delta\underline{\lambda}_n^2 \right) = (A)$$

$$+ \sum_{k+1}^{r} \mathrm{P}\left(\max_{\mathbf{t} \in \mathscr{K}_n(M)} |S_{k,k+1}| \left(\frac{|S_{k,k}|}{n_{k,k}} + \frac{|S_{k+1,k+1}|}{n_{k+1,k+1}} \right) \geqslant c\delta\underline{\lambda}_n^{2-\gamma} \right) = (B)$$

$$+ \sum_{k+1}^{r} \mathrm{P}\left(\max_{\mathbf{t} \in \mathscr{K}_n(M)} S_{k,k+1}^2 \geqslant c\delta\underline{\lambda}^{2-\gamma} \right) = (C). \tag{75}$$

where $c > 0$ is a sufficiently small constant. Using (67), we have $(A) = O(n^{\phi-2}\underline{\lambda}^{-2}) = o(1)$. Similarly we have by applying (68)–(69)

$$(B) \leqslant \sum_{k=1}^{r} \left(\max_{\mathbf{t} \in \mathscr{K}_n(M)} |S_{k,k+1}| \geqslant \underline{\lambda}_n^{1-\gamma} n^{\varepsilon/\gamma} \right) + \mathrm{P}\left(\max_{\mathbf{t} \in \mathscr{K}_n(M)} \frac{|S_{k,k}}{n_{k,k}} \geqslant c\delta\underline{\lambda}_n n^{-\varepsilon/\gamma} \right),$$

$$= O(n^{-2\varepsilon/\gamma}) + O(n^{(1-\varepsilon)(\phi-2)}\underline{\lambda}^{-2}) = o(1).$$

Finally, using again (67)–(71), we have $(C) = O(n^{-1}\underline{\lambda}_n^{-\gamma}) = o(1)$, concluding the proof of (43). The proof of (44) is along the same lines and is omitted. ■

REFERENCES

BAI, J., (1994), Least Square estimation of a shift in linear processes. *J. Time Series Analysis*, 15, 5, 453–472.

BAI, J. and PERRON, P., (1996), Estimating and testing linear models with multiple structural changes. *To appear in Econometrica*.

BASSEVILLE, M. and NIKIFOROV, N., (1993), *The Detection of abrupt changes – Theory and applications*. Prentice-Hall: Information and System sciences series.

BATTACHARYA, P., (1987), Maximum likelihood estimation of a change-point in the distribution of independent random variables: general multiparameter case. *J. of Multivariate Anal.*, 32, 183–208.

BERAN, J., (1992), Statistical methods for data with long-range dependence. *Statistical Science*, 7, 404–427.

BIRNAUM, Z. and MARSHALL, A., (1961), Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. *Ann. Math. Stat.*, 32, 682–703.

BRODKSY, B. and DARKHOVSKY, B., (1993), *Nonparametric methods in change-point problems*. Kluwer Academic Publishers, the Netherlands.

DOUKHAN, P., (1994), *Mixing, properties and examples*. Lecture notes in statistics, 85, Springer Verlag.
DOUKHAN, P. and LOUICHI, S., (1997), Weak dependence and moment inequalities. *Preprint University Paris-Sud*.
HÁJEK, J. and RÉNYI, A., (1955), Generalizations of an inequality of Kolmogorov. *Acta. Math. Acad. Sci.*, 6, 281–283.
HO, H. C. and HSING, T., (1997) Limit theorem for functional of moving averages, *The Annals of Probability* 25(4), 1636–1669.
HAWKINS, D., (1977), Testing a sequence of observations for a shift in location. *J. Am. Statist. Assoc.*, 72, 180–186.
HINKLEY, D., (1970), Inference about the change point in a sequence of random variables. *Biometrika*, 57, 1–17.
KIM, J. and POLLARD, D., (1990), Cube root asymptotics. *The Annals of Stat.*, 18, 191–219.
LIU, J., WU, S. and ZIDEK, J., (1997), On segmented multivariate regression. *Statistica Sinica*, 7, 497–525.
MCLEISH, D., (1975), A maximal inequality and dependent strong laws. *The Annals of Probability*, 5, 829–839.
MRICZ, F., SERFLING, R. and STOUT, W., (1982), Moment and probability bounds with quasi-superadditive structure for the maximum partial sum. *The Annals of Prob.*, 10, 4, 1032–1040.
PICARD, D., (1985), Testing and estimating change points in time series. *J. Applied Prob.*, 17, 841–867.
RIO, E., (1995), The functional law of iterated logarithm for strongly mixing processes. *The Annals of Prob.*, 23, 3, 1188–1203.
SEN, A. and SRIVASTAVA, M., (1975), On tests for detecting change in the mean. *The Annals of Stat.*, 3, 96–103.
TAQQU, M., (1975), Weak convergence to the fractional Brownian motion and to the Roseblatt process. *Z. Wahrsch. verw. Geb.*, 31, 287–302.
TAQQU, M., (1977), Law of the iterated logarithm for sum of non-linear functions of Gaussian variables than exhibit long range dependence. *Z. Wahrsch, verw. Geb.*, 40, 203–238.
YAO, Y., (1988), Estimating the number of change-points via Schwarz criterion. *Stat. & Probab. Lett.*, 6, 181–189.