NONPARAMETRIC MULTIVARIATE BREAKPOINT DETECTION FOR THE MEANS, VARIANCES, AND COVARIANCES OF A DISCRETE TIME STOCHASTIC PROCESS

VINCENT GUIGUES

Abstract. We introduce a nonparametric breakpoint detection method for the means and covariances of a multivariate discrete time stochastic process. Breakpoints are defined as left or right endpoints of maximal intervals of local time homogeneity for the means and covariances. The breakpoint detection method is an adaptive algorithm that estimates the last maximal interval of homogeneity. Applied recursively, it allows us to find an arbitrary number of breakpoints. We then study a second breakpoint detection algorithm that makes use of a sliding window. The quality of both methods is analyzed. For the adaptive algorithm, we provide the quality of the estimation of the one step ahead means and covariance matrix as well as upper bounds on the type I and type II errors when applying the procedure to a change-point model. Regarding the second method, the probability of correctly detecting the breakpoint of a change-point model is bounded from below. Numerical simulations assess the performance of both methods using simulated data.

AMS subject classifications: 62M10, 62G05.

1. Introduction

Detecting breakpoints in multidimensional time series allows us to identify structural changes in discrete time stochastic processes. Solving this challenging question is useful for a wide range of applications in bioinformatics (Fridlyand, Snijders, Pinkel, Albertson, and Jain 2004; Tibshirani and Wang 2007; Zhang and Siegmund 2007), finance (Mikosch and Starica 2000; Fan, Jiang, Zhang, and Zhou 2003; Mercurio and Spokoiny 2004), image processing (Désobry, Davy, and Doncarli 2005), or production management (Guigues 2009); see also (Basseville and Nikiforov 1993). More specifically, considering breakpoints yields a wider and more flexible class of models and is of interest for forecasting; for instance to generate short-term scenarios for stochastic optimization problems (Heitsch and Römisch 2009).

Breakpoint detection consists in looking for homogeneous segments where some or all the model parameters are constant or slowly varying in each segment. The definition of a breakpoint may vary from a study to another. In our case, a breakpoint is an endpoint of an *interval of local time homogeneity* (ILTH). The definition of an ILTH for a discrete time stochastic process dates back to Mercurio and Spokoiny (2004) in the one-dimensional case. In Guigues (2008), we extended the definition of ILTH to the *multivariate* case. This definition can be found in Section 2. Roughly speaking, an interval is said to be of local time homogeneity if the first two moments (mean and covariance matrix) are slowly varying on this interval.

Most of the works on breakpoint detection are based on parametric models and often use a strong a priori information. Such is the case of Hidden Markov Models using the Bayesian

Key words and phrases. Breakpoint detection and Time series and Multivariate data and Covariance matrix estimation and Adaptive algorithm.

Information Criterion (BIC) or Akaike Information Criterion (AIC) (Fridlyand et al. 2004) and of varying coefficient models (Fan and Zhang 1999; Cai, Fan, and Li 2000; Mercurio and Spokoiny 2004).

Nonparametric methods to detect sharp changes have also been developed. Popular examples from this class include kernel-based (Désobry, Davy, and Doncarli 2005) and wavelet-based (Wang 1995; Antoniadis and Gijbels 2002) algorithms. Wavelet methods deal in general with one-dimensional signals and detect changes looking at large values of wavelet coefficients at certain scales. These wavelet-based methods differ in the choice of the threshold above which wavelet coefficients are declared "large". Another popular and natural nonparametric breakpoint detection method useful both in the one-dimensional and the multidimensional case is based on the use of sliding windows (Müller 1992; Harchaoui, Bach, and Moulines 2008; Lévy-Leduc and Roueff 2009). In this context, typically, at each time, a test is performed to compare estimators on the left and right part of the corresponding window. However, most of the works mentioned above consider a finite and known number of breakpoints, which may appear as a strong limitation.

On the contrary, in this paper, we propose an *adaptive* algorithm to detect breakpoints in a nonparametric context for multidimensional data without assuming a specific number of breakpoints. This algorithm builds on our previous work Guigues (2008) where we extended the definition of ILTH to the multivariate case and presented an adaptive algorithm to estimate the ILTH in a nonparametric framework. Applied recursively, our algorithm allows us to estimate breakpoints defined as endpoints of maximal ILTH. The work Guigues (2008) is itself an extension of Mercurio and Spokoiny (2004). This latter paper explains how to obtain an adaptive estimation of an ILTH in a one-dimensional and parametric context with an application for forecasting the volatility of financial time series. Pointwise adaptive estimation methods first appeared in Lepski (1990) and were used in (Lepski and Spokoiny 1997; Spokoiny 1998). In this setting, the contributions of this paper are threefold.

First, we detail a modified version of the adaptive algorithm introduced in Guigues (2008) and show that the corresponding quality of the estimations of the one step ahead means and covariance matrix is theoretically controlled (Theorem 3.2).

Second, the procedure is applied to detect breakpoints in the means, variances, and covariances of a multivariate discrete time stochastic process in a nonparametric framework; a situation which has not received a lot of attention in the literature so far. In this context, for a change-point model, in Theorems 3.3 and 3.7, we bound from above the type I error (made when the algorithm sees an ILTH as an interval which is not of time homogeneity), and in Theorems 3.4 and 3.8, we bound from above the type II error (made when an interval which is not an ILTH is seen as ILTH).

Finally, we study a second breakpoint detection method which uses left and right sliding windows. Using the same probabilistic (nonparametric) framework as before, we bound from below the probability of correctly detecting the breakpoint for change-point models (Theorems 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, and 4.7).

The paper is organized as follows. The ILTH and breakpoints are defined in Section 2. The adaptive algorithm is presented and studied in Section 3. The second breakpoint detection method is studied in Section 4. Finally, in Section 5, we assess the efficiency of the proposed methods using simulated change-point models.

The proofs of the theorems as well as some new technical large deviation results (interesting *per se*) are given in the Appendix. We use the same notation for a random variable and for a particular realization; the context allowing us to know which concept is being referred to.

We start by setting down some notation:

- The integer part of $x \in \mathbb{R}$ will be denoted by $\mathrm{E}[x]$.
- For a random vector X, we denote its expectation by $\mathbb{E}[X]$.
- By $\|x\|_\infty$, we denote the infinity norm of the vector $x \in \mathbb{R}^n$, i.e., $\|x\|_\infty = \max(|x_i|, i = 1, \ldots, n)$.
- The cardinality of a set I is denoted by $|I|$.

2. ILTH and breakpoints

Let $r_t, t = 1, \ldots, N$, (with $N > 1$), be N independent observations of a time series generated by the model

$$r_t = \rho_t + \zeta_t, \quad \text{with } \mathbb{E}[r_t] = \rho_t \text{ and } \mathbb{E}[\zeta_t \zeta_t^\top] = Q_t, \quad t = 1, \ldots, N, \tag{1}$$

where ζ_t are independent random vectors in $\mathbb{R}^n$ with zero mean and $n \geq 2$. In what follows, depending on the context, we will make use of one of the following two assumptions:

(A1) For $t = 1, \ldots, N$, $\mathbb{E}[\|\zeta_t\|_\infty^p] \leq \sigma^p$ and $\|\rho_t\|_\infty \leq \sigma'$ for some finite $p > 4$, $\sigma > 0$, and $\sigma' > 0$.

(A2) For $t = 1, \ldots, N$, $\mathbb{E}[\|r_t\|_\infty^4] \leq \sigma^4$ for some finite $\sigma > 0$.

We first need to introduce the key notion of interval[1] of local time homogeneity (ILTH). The definition of an ILTH can be found in Guigues (2008) for the multivariate case and in Mercurio and Spokoiny (2004) for the one-dimensional case. An ILTH is an interval where the parameters ρ_t (the means) and Q_t (the covariance matrices) slowly vary. To define more precisely an ILTH, for any nonempty interval I, let $\hat{\rho}_I$ and $\hat{Q}_I$ be the following estimators of the mean and of the covariance matrix using the data of interval I:

$$\hat{\rho}_I = \frac{1}{|I|} \sum_{t \in I} r_t \quad \text{and} \quad \hat{Q}_I = \frac{1}{|I|} \sum_{t \in I} (r_t - \hat{\rho}_I)(r_t - \hat{\rho}_I)^\top. \tag{2}$$

We also set for any nonempty interval I

$$\Delta_I^\rho = \sqrt{\frac{1}{|I|} \sum_{t \in I} \|\rho_t - \rho_{N+1}\|_\infty^2} \quad \text{and} \quad \Delta_I^Q = \sqrt{\frac{1}{|I|} \sum_{t \in I} \|Q_t - Q_{N+1}\|_\infty^2}.$$

If the parameters ρ_t and Q_t slowly vary on a set I embedded in $\{1, \ldots, N+1\}$ with right endpoint $N+1$, we expect every ρ_t (resp. Q_t) for $t \in I$ to be close to ρ_{N+1} (resp. Q_{N+1}) and hence Δ_I^ρ (resp. Δ_I^Q) to be small. Similarly, for any subinterval J of an interval I where ρ_t and Q_t slowly vary, we expect Δ_J^ρ and Δ_J^Q to be small. To take into account the variability of the estimators $\hat{\rho}_I$ and $\hat{Q}_I$ of ρ_{N+1} and Q_{N+1} obtained using a set I embedded in $\{1, \ldots, N+1\}$ with right endpoint $N+1$, we introduce

$$V_I^\rho = \mathbb{E}[\|\hat{\rho}_I - \mathbb{E}[\hat{\rho}_I]\|_\infty] \quad \text{and} \quad V_I^Q = \mathbb{E}[\|\hat{Q}_I - \mathbb{E}[\hat{Q}_I]\|_\infty],$$

(notice that V_I^ρ simply reads $V_I^\rho = \mathbb{E}[\|\frac{1}{|I|}\sum_{t \in I} \zeta_t\|_\infty]$). For any nonempty interval I, let $\mathcal{I}(I)$ be a finite set of testing subintervals for I. An interval I embedded in $\{1, \ldots, N+1\}$ with right

[1]Since we deal with a discrete time stochastic process, the intervals I considered are discrete sets of consecutive time steps belonging to $\{1, \ldots, N+1\}$.

endpoint $N+1$ is said to be of local time homogeneity if $\Delta_J^\rho \leq DV_J^\rho$ and $\Delta_J^Q \leq DV_J^Q$ for $J = I$ and $J \in \mathcal{I}(I)$ where D is a fixed (small) constant. Different choices are possible for $\mathcal{I}(I)$. A possibility is to take all intervals of length proportional to m_0 ($m_0 > 0$ being fixed) strictly embedded in I with either the same left endpoint or the same right endpoint as I (see the numerical simulations of Section 5 for another possibility).

Next, we define a family $\mathcal{I}$ of candidate intervals of local time homogeneity as follows. Though we assume that $N+1$ belongs to the ILTH, since we do not have observations for time step $N+1$, time step N will be the right endpoint of all intervals from $\mathcal{I}$: the intervals in $\mathcal{I}$ have right endpoint N and length greater than or equal to m_0, i.e., they are of the form $\{N - m_0 - k, \ldots, N\}$ for some $k \in \mathbb{N}$.

In this context, the maximal ILTH denoted by $\mathbb{I}$ is defined as follows:

$$\mathbb{I} = \operatorname{argmax}\{|I| \mid I \in \mathcal{I},\ \Delta_J^\rho \leq DV_J^\rho,\ \Delta_J^Q \leq DV_J^Q, \text{ for } J = I \text{ and } J \in \mathcal{I}(I)\}. \tag{3}$$

We then define a breakpoint as a right or a left endpoint of a maximal ILTH. For $\mathbb{I}$ to be well defined, we assume that there is at least an ILTH (the smallest interval in $\mathcal{I}$ is an ILTH). By definition, the maximal ILTH is the largest interval $\mathbb{I}$, among a family of candidate intervals, such that the means, variances, and covariances slowly vary on interval $\mathbb{I}$. The estimation of this maximal ILTH, via an adaptive algorithm, is addressed in the next section. Under the hypothesis of LTH, since it is assumed that $N+1$ belongs to $\mathbb{I}$, this adaptive algorithm not only allows us to determine an estimation of the maximal ILTH but also estimations (on the basis of the available past data $r_t, t = 1, \ldots, N$) of the mean ρ_{N+1} and of the covariance matrix Q_{N+1} for the time step following the instant of the last observation.

Notice that instead of using one interval of homogeneity for both the mean and the covariance matrix, we could use two separate intervals of homogeneity $\mathbb{I}_\rho$ for the mean and $\mathbb{I}_Q$ for the covariance matrix. Intervals $\mathbb{I}_\rho$ and $\mathbb{I}_Q$ would be defined as follows:

$$\mathbb{I}_\rho = \operatorname{argmax}\{|I| \mid I \in \mathcal{I},\ \Delta_J^\rho \leq DV_J^\rho, \text{ for } J = I \text{ and } J \in \mathcal{I}(I)\}$$

and

$$\mathbb{I}_Q = \operatorname{argmax}\{|I| \mid I \subseteq \mathbb{I}_\rho,\ I \in \mathcal{I},\ \Delta_J^Q \leq DV_J^Q, \text{ for } J = I \text{ and } J \in \mathcal{I}(I)\}.$$

3. Adaptive algorithm

3.1. **Algorithm implementation and accuracy.** The adaptive algorithm determines an estimation $\hat{I}$ of the largest interval $\mathbb{I}$ among a family $\mathcal{I}$ of ordered candidate intervals $\mathcal{I}_1 \subset \mathcal{I}_2 \subset \ldots \subset \mathcal{I}_T$ such that the means, variances, and covariances slowly vary on $\mathbb{I}$. With this notation, $\mathcal{I}_1$ (the smallest candidate interval) is assumed to be an ILTH and $\mathbb{I} = \mathcal{I}_t$ where $\mathcal{I}_t$ is such that $\mathcal{I}_1, \ldots, \mathcal{I}_t$ are ILTH but $\mathcal{I}_{t+1}$ is not an ILTH. It remains to explain how to decide if on a given candidate interval $I \in \mathcal{I}$, we can consider that ρ_t and Q_t are slowly varying or not. This decision will be based on Theorem 3.1 below. We first introduce for $\lambda > 0$ the following notation:

$$K_1(\lambda) = 2 + \left(\frac{10\,p(\gamma+\lambda)}{3e(p-2-2\lambda)}\right)^{\frac{p}{2}} \text{ and } K_2(\lambda) = 1 + \left(\frac{10\,E[\frac{p}{2}](2\gamma+\lambda)}{3\,e(E[\frac{p}{2}]-2-2\lambda)}\right)^{\frac{E[\frac{p}{2}]}{2}}, \tag{4}$$

where γ is chosen such that $n \leq N^\gamma$ and p is the integer from Assumption **(A1)**.

Theorem 3.1. *Let $r_t, t = 1, \dots, N$, be N independent observations of a model generated by* (1). *Let Assumption **(A1)** hold, let $I \in \mathcal{I}$ be an ILTH and let $J \in \mathcal{I}(I)$. Let $\hat{\rho}_I$ and $\hat{Q}_I$ be the estimators defined in* (2). *We set $\sigma'' = \sigma^2 + (2\sigma' + \sigma)^2$ and $\sigma''' = 2(2\sigma' + \sigma)^2$. Then for every $\lambda > 0$ such that $10(\ln n + \lambda \ln m_0) \le 3m_0$, $p > 2(1+\lambda)$ and $\mu > 0$ such that $10(\ln n(n+1) + \mu \ln m_0) \le 3m_0$, $E[\frac{p}{2}] > 2(1+\mu)$ (m_0 being the length of the smallest interval of the set $\mathcal{I}(I)$), we have:*

$$\mathbb{P}\left(\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \ge \gamma_\rho(|I|, |J|, \lambda)\right) \le K_1(\lambda)\left(\frac{1}{|I|^\lambda} + \frac{1}{|J|^\lambda}\right), \tag{5}$$

$$\mathbb{P}\left(\|\hat{Q}_I - \hat{Q}_J\|_\infty \ge \gamma_Q(|I|, |J|, \mu)\right) \le (K_1(\mu) + K_2(\mu))\left(\frac{1}{|I|^\mu} + \frac{1}{|J|^\mu}\right), \tag{6}$$

where

$$\gamma_\rho(|I|, |J|, \lambda) = 2\sqrt{\frac{10}{3}}\sigma\left(\sqrt{\frac{\ln n + \lambda \ln |I|}{|I|}} + \sqrt{\frac{\ln n + \lambda \ln |J|}{|J|}}\right) + 2\sqrt{\frac{2}{\ln 2}} D\sigma \left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}}\right),$$

and where $\gamma_Q(|I|, |J|, \mu)$ is given by

$$k_1(\sigma, \sigma'')\left(\sqrt{\frac{\ln n(n+1) + \mu \ln |I|}{|I|}} + \sqrt{\frac{\ln n(n+1) + \mu \ln |J|}{|J|}}\right) + k_2(\sigma, \sigma''')D\left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}}\right),$$

with $k_1(\sigma, \sigma'') = 2\sqrt{\frac{10}{3}}(2\sigma^2 + \sigma'')$ and $k_2(\sigma, \sigma''') = \frac{16\sigma^2}{\ln 2} + \frac{4\sigma'''}{\sqrt{\ln 2}}$.

Now let $I \in \mathcal{I}$ be a candidate interval and let (λ, μ) be two positive parameters satisfying the conditions given in Theorem 3.1. Using this theorem, we will consider that the mean is slowly varying on interval I, if for every interval $J \in \mathcal{I}(I)$:

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \le \gamma_\rho(|I|, |J|, \lambda). \tag{7}$$

Similarly, we will consider that the covariance matrix is slowly varying on interval I if for every interval $J \in \mathcal{I}(I)$:

$$\|\hat{Q}_I - \hat{Q}_J\|_\infty \le \gamma_Q(|I|, |J|, \mu). \tag{8}$$

After running the adaptive algorithm, we end up with an estimation of a maximal interval I of homogeneity and we have detected a breakpoint. This breakpoint is a breakpoint in the mean (resp. in the covariance matrix) if inequality (7) (resp. (8)) is not satisfied for at least one subinterval $J \in \mathcal{I}(I)$.

We see that the adaptive algorithm has two positive parameters λ and μ which respectively control the calibration of the (one step ahead) mean ρ_{N+1} and of the (one step ahead) covariance matrix Q_{N+1}.

Once an estimation $\hat{I}$ of $\mathbb{I}$ is determined, the estimations of these one step ahead mean ρ_{N+1} and one step ahead covariance matrix Q_{N+1} are given by $\hat{\rho}_{\hat{I}}$ and $\hat{Q}_{\hat{I}}$.

In Theorem 3.2 which follows, we give the accuracy of the adaptive estimations $\hat{\rho}_{\hat{I}_\rho}$ of the mean ρ_{N+1} and $\hat{Q}_{\hat{I}_Q}$ of the covariance matrix Q_{N+1} when two intervals of homogeneity $\mathbb{I}_\rho$ and $\mathbb{I}_Q$ are used. A similar result can be given when one ILTH is used. In the sequel, we set $\mathcal{I}_+(I) := \mathcal{I}(I) \cup I$.

Theorem 3.2. *Let Assumption **(A1)** hold and let $\hat{\rho}_{\hat{I}_\rho}$ and $\hat{Q}_{\hat{I}_Q}$ be the empirical adaptive estimations of the mean ρ_{N+1} and of the covariance matrix Q_{N+1}. Let (λ, μ) be the parameters of the adaptive algorithm such that $\lambda > 0$, $p > 2(1+\lambda)$, $10(\ln n + \lambda \ln m_0) \le 3m_0$, $\mu > 0$, $E[\frac{p}{2}] > 2(1+\mu)$,* *and* $10(\ln n(n+1)+\mu \ln m_0) \leq 3m_0$. *Then*

$$\mathbb{P}\left(\|\hat{\rho}_{\hat{I}_\rho}-\rho_{N+1}\|_\infty > (6\sqrt{\tfrac{10}{3}}+6\sqrt{\tfrac{2}{\ln 2}}D)\sigma\sqrt{\tfrac{\ln n+\lambda \ln |\mathbb{I}_\rho|}{|\mathbb{I}_\rho|}}\right) \leq \sum_{I\in\mathcal{I}\ |I\subseteq \mathbb{I}_\rho}\ \sum_{J\in\mathcal{I}_+(I)} \frac{K_1(\lambda)}{|J|^\lambda}, \tag{9}$$

$$\mathbb{P}\left(\|\hat{Q}_{\hat{I}_Q}-Q_{N+1}\|_\infty > f(\sigma,\sigma')\sqrt{\tfrac{\ln n(n+1)+\mu \ln |\mathbb{I}_Q|}{|\mathbb{I}_Q|}}\right) \leq \sum_{I\in\mathcal{I}|I\subseteq \mathbb{I}_Q}\ \sum_{J\in\mathcal{I}_+(I)} \frac{K_1(\mu)+K_2(\mu)}{|J|^\mu}, \tag{10}$$

with $f(\sigma,\sigma')=6\sqrt{\frac{10}{3}}(2\sigma^2+\sigma'')+3D(\frac{16\sigma^2}{\ln 2}+\frac{4\sigma'''}{\sqrt{\ln 2}})$, $\sigma''=\sigma^2+(2\sigma'+\sigma)^2$, *and* $\sigma'''=2(2\sigma'+\sigma)^2$.

For fixed λ and μ, the confidence areas for the mean ρ_{N+1} and for the covariance matrix Q_{N+1} provided by the above theorem (of the form $\|\hat{\rho}-\rho_{N+1}\|_\infty \leq k_1\sigma\sqrt{\frac{\ln n+\lambda\ln|\mathbb{I}_\rho|}{|\mathbb{I}_\rho|}}$ and $\|\hat{Q}-Q_{N+1}\|_\infty \leq k_2(\sigma+\sigma')^2\sqrt{\frac{2\ln n+\mu\ln|\mathbb{I}_Q|}{|\mathbb{I}_Q|}}$ for some constants k_1 and k_2) are all the smaller as D and σ are small. Also, the volume of these confidence areas is slowly increasing when the number of components n increases and rapidly decreases when the lengths of $\mathbb{I}_\rho$ and $\mathbb{I}_Q$ increase. Finally, notice that the conditions $10(\ln n+\lambda\ln m_0)\leq 3m_0$ and $10(\ln n(n+1)+\mu\ln m_0)\leq 3m_0$ in the above theorem can be suppressed but this would lead to more complicated left-hand sides (see the Appendix). However, these conditions are not too restrictive. For instance, for $n=40$, if we take $m_0=n$ then we can take for μ values as large as 8.84. If $\mu=1$ and $m_0=n$ then it suffices for r_t to have more than 6 components ($n\geq 6$) to get $10(\ln n(n+1)+\mu\ln m_0)\leq 3m_0$.

3.2. **Type I and type II errors for a change-point model.** An interesting particular case where the hypothesis of LTH holds is when ρ_t and Q_t are piecewise constant functions. For a given time $N+1$, on the basis of N past observations $r_t, t=1,\ldots,N$, of a change-point time series satisfying Assumption **(A1)**, we would like to determine the last breakpoint in the mean or the covariance matrix. Using the adaptive algorithm, we can determine an estimation of the largest interval with right endpoint N without breakpoints in the means, variances, and covariances. It boils down to a multiple testing problem. In order to obtain a small type I error (made when a homogeneity interval is rejected), each test should be conducted with a small type I error. The following theorem provides an upper bound for the type I error if we use the adaptive algorithm with a change-point model.

Theorem 3.3. *Let* (r_t) *be a discrete time stochastic process generated by* (1). *Let Assumption* ***(A1)*** *hold and let* $I\in\mathcal{I}$ *be an ILTH with right endpoint* N *such that for every* $t,t'\in I$, $\rho_t=\rho_{t'}$ *and* $Q_t=Q_{t'}$. *Then if* $(\lambda,\mu)>0$ *are the parameters of the adaptive algorithm such that* $10(\ln n+\lambda\ln m_0)\leq 3m_0$, $p>2(1+\lambda)$, $10(\ln n(n+1)+\mu\ln m_0)\leq 3m_0$, *and* $E[\frac{p}{2}]>2(1+\mu)$ *(*m_0 *being the length of the smallest interval of the set* $\mathcal{I}(I)$*), we have*

$$\mathbb{P}(I \text{ is rejected}) \leq 2\,\frac{\mathrm{Card}(\mathcal{I}(I))}{m_0^{\min(\lambda,\mu)}}\left(K_1(\lambda)+K_1(\mu)+K_2(\mu)\right). \tag{11}$$

We would also like the algorithm to detect breakpoints in the means, variances, and covariances, if these changes are important enough. The following theorem provides an upper bound for the type II error when applying the adaptive algorithm to a change-point model with a sufficiently important breakpoint in the mean.

Theorem 3.4. *Let* (r_t) *be a discrete time stochastic process generated by* (1). *Let Assumption* ***(A1)*** *hold and let* $I=\{T_{bp}-m,\ldots,N\}$ *be an interval with a breakpoint at* T_{bp}*: for* $t=T_{bp}-$

$m, \ldots, T_{bp}-1$ *($m \geq m_0$ is a multiple of m_0), $\rho_t = m_1$ and for $t = T_{bp}, \ldots, N$, $\rho_t = m_2$. Let $m' = N+1-T_{bp} \geq m_0$ and let $\lambda > 0$ such that $10(\ln n + \lambda \ln m_0) \leq 3m_0$. If the changes in the means are sufficiently important, i.e., if $\|m_2 - m_1\|_\infty$ is greater than*

$$\frac{m+m'}{m'}\left(2f_1(m+m',\lambda)+2f_1(m,\lambda)+2\sqrt{\frac{2}{\ln 2}}D\sigma(\sqrt{\frac{\ln n}{m+m'}}+\sqrt{\frac{\ln n}{m}})\right), \tag{12}$$

where for any nonempty interval I

$$f_1(|I|,\lambda) = 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda \ln |I|}{|I|}}, \tag{13}$$

then the probability for interval I to be accepted as an interval of homogeneity is bounded from above as follows:

$$\mathbb{P}(I \text{ accepted}) \leq K_1(\lambda)\left(\frac{1}{m^\lambda}+\frac{1}{(m+m')^\lambda}\right).$$

A similar result can be given if we are interested in detecting breakpoints in the variances and covariances. We assumed m multiple of m_0 and $10(\ln n + \lambda \ln m_0) \leq 3m_0$ to simplify the presentation of the result but the type II error can be controlled without this last condition (see the Appendix).

The adaptive algorithm can also be implemented as in Guigues (2008) replacing $\hat{\rho}_I$ and $\hat{Q}_I$ by other (close) estimators. In this case, making Assumption **(A2)**, the type I and type II errors can be more easily controlled (see Theorems 3.7 and 3.8 which follow). These questions are addressed in the next section.

3.3. **A modified version of the adaptive algorithm.** Let us fix a positive parameter λ, and for any nonempty interval I, let $K_0(|I|)$ and $[\cdot]_K$, for $K > 0$, be the constant and the truncation operator defined by

$$K_0(|I|) = \sigma\Big(\tfrac{|I|}{\ln n(n+1)+\lambda \ln |I|}\Big)^{\frac{1}{4}}; \quad [x]_K = \begin{cases} K & \text{if } x > K, \\ -K & \text{if } x < -K, \\ x & \text{otherwise.} \end{cases} \tag{14}$$

Instead of using estimations $\hat{\rho}_I$ and $\hat{Q}_I$ given by (2), the adaptive algorithm can be implemented using (as in Guigues (2008))

$$\hat{\rho}_I = \frac{1}{|I|}\sum_{t\in I}\alpha_t^I \text{ and } \hat{Q}_I = \frac{1}{|I|}\sum_{t\in I}(\alpha_t^I - \hat{\rho}_I)(\alpha_t^I - \hat{\rho}_I)^\top, \tag{15}$$

where for $i = 1, \ldots, n$, and $t \in I$, $\alpha_t^I(i) = [r_t(i)]_{K_0(|I|)}$. In this context, λ becomes the only parameter of the adaptive algorithm. The estimations (15) are close to the estimations (2) when the length of I is large enough. It is explained in the Appendix (in the proof of Theorems 3.7 and 3.8 below) how this version of the adaptive algorithm is implemented. On the one hand, the upper bounds we get with this algorithm, in particular for the type I and type II errors when applied to a change-point model, are simpler than those obtained with the algorithm presented in the present paper. On the other hand, instead of using the estimators (15), this latter algorithm makes use of the more natural estimators (2) and has two parameters (λ, μ) instead of just one parameter λ in Guigues (2008). This additional degree of freedom allows us to improve the algorithm performance (see the numerical simulations in Section 5).

In the rest of this section, we focus on the adaptive algorithm from Guigues (2008) and bound the type I and type II errors obtained when applying it to a change-point model.

3.3.1. *Quality of the estimation.* Denoting by svec(Q) the symmetric vectorization of matrix Q, we start recalling (from Guigues (2008)) the quality of the estimation $\hat{\theta}_{\mathbb{I}} = (\hat{\rho}_{\mathbb{I}}^\top, \text{svec}(\hat{Q}_{\mathbb{I}})^\top)^\top$ that would be used for the parameter $\theta = (\rho_{N+1}^\top, \text{svec}(Q_{N+1})^\top)^\top$ if the ideal interval $\mathbb{I}$ of local time homogeneity was known.

Theorem 3.5. *Guigues (2008) Let Assumption **(A2)** hold. If $\lambda > 0$ is such that $\ln n(n+1) + \lambda \ln |\mathbb{I}| \leq |\mathbb{I}|$, then there is a constant $k(D)$ depending affinely on D such that*

$$\mathbb{P}\left(\|\hat{\theta}_{\mathbb{I}} - \theta\|_\infty \geq k(D) \max(\sigma, \sigma^2) \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} \right) \leq \frac{3}{|\mathbb{I}|^\lambda}. \tag{16}$$

The following theorem then gives the accuracy of the adaptive estimations.

Theorem 3.6. *Guigues (2008) Let Assumption **(A2)** hold. Let $\hat{I}$ be the interval selected by the adaptive algorithm and λ be the parameter involved in the definition of K_0. We suppose that $\ln n(n+1) + \lambda \ln m_0 \leq m_0$ where m_0 is the length of the smallest testing subinterval. Then there is a constant $k(D)$ depending affinely on D such that if $\hat{\theta}_{\hat{I}} = (\hat{\rho}_{\hat{I}}^\top, \text{svec}(\hat{Q}_{\hat{I}})^\top)^\top$ we get*

$$\mathbb{P}\left(\|\hat{\theta}_{\hat{I}} - \theta\|_\infty \geq k(D) \max(\sigma, \sigma^2) \sqrt{\frac{\ln n(n+1) + \lambda \ln |\mathbb{I}|}{|\mathbb{I}|}} \right) \leq \sum_{I \in \mathcal{I} \, | I \subseteq \mathbb{I}} \; \sum_{J \in \mathcal{I}_+(I)} \frac{3}{|J|^\lambda}. \tag{17}$$

Observe that the right-hand side of the above inequality depends on the number and lengths of the testing subintervals. It goes to 0 when λ goes to infinity. However, when λ increases we naturally obtain confidence areas of greater volume. For implementation purposes, we address the calibration of λ (of (λ, μ) with the version of the adaptive algorithm introduced in the previous section) in Section 5, which is dedicated to numerical simulations. Also observe that, as with the version of the adaptive algorithm from Section 3.1, the volume of these confidence areas naturally decreases when D and σ decrease, decreases when $|\mathbb{I}|$ increases, and slowly increases with the number n of components of the process. Finally, comparing Theorems 3.5 and 3.6, we see that the quality of the adaptive estimators is close to the quality of estimators $\hat{\rho}_{\mathbb{I}}$ and $\hat{Q}_{\mathbb{I}}$ that would be used if the ideal interval $\mathbb{I}$ for parameter estimation was known in advance. The adaptive algorithm can be viewed as an oracle that provides estimations of ρ_{N+1} and Q_{N+1} with controlled accuracy, on the basis of past observations $r_t, t = 1, \ldots, N$.

3.3.2. *Type I and type II errors for a change-point model.* The type I and type II errors are now controlled as follows:

Theorem 3.7. *Let (r_t) be a discrete time stochastic process generated by (1). Let Assumption **(A2)** hold and let I be an ILTH with right endpoint N such that for all $t, t' \in I$, $\rho_t = \rho_{t'}$ and $Q_t = Q_{t'}$. If $\lambda > 0$ is the parameter of the adaptive algorithm from Guigues (2008), then*

$$\mathbb{P}(I \text{ is rejected}) \leq \frac{6\,\text{Card}(\mathcal{I}(I))}{m_0^\lambda}. \tag{18}$$

The type I error is all the smaller as λ is large. More precisely, the upper bound given by Theorem 3.7 provides a way to choose λ, i.e.,

$$\lambda \geq \frac{\ln(\frac{6 \text{ Card}(\mathcal{I}(I))}{\alpha})}{\ln m_0},$$

such that the type I error is at most $\alpha > 0$.

The following theorem states that if on a given interval I, there is an important breakpoint in the mean and the covariance matrix at $T_{bp} \in I$, then for large enough values of λ and if we have a sufficient number of observations after the breakpoint, there is little chance that the adaptive algorithm from Guigues (2008) accepts this interval as an ILTH (that is to say that the algorithm makes a type II error).

Theorem 3.8. *Let* (r_t) *be a discrete time stochastic process generated by* (1)*. Let Assumption* ***(A2)*** *hold, let* $m \in \mathbb{N}^*$ *be a multiple of* m_0*, and let* $I = \{T_{bp} - m, \ldots, N\}$ *be an interval with a breakpoint in the means, variances, and covariances at* T_{bp}*: on* $J = \{T_{bp} - m, \ldots, T_{bp} - 1\}$*,* $\rho_t = m_1$*,* $Q_t = \mathcal{Q}_1$*, and on* $\{T_{bp}, \ldots, N\}$*,* $\rho_t = m_2$*,* $Q_t = \mathcal{Q}_2$*. Let* $\lambda > 0$ *be the parameter of the adaptive algorithm. We assume that* $\ln n(n+1) + \lambda \ln m_0 \leq m_0$*. If the breakpoint is sufficiently important, i.e., if* $\|m_2 - m_1\|_\infty$ *is greater than*

$$\frac{m+m'}{m'}\sigma\left((\frac{14}{3} + 2\sqrt{2})\left(f(m+m',\lambda) + f(m,\lambda)\right) + 4\sqrt{\frac{2}{\ln 2}}D(\sqrt{\frac{\ln n}{m+m'}} + \sqrt{\frac{\ln n}{m}})\right) \tag{19}$$

and

$$\|\mathcal{Q}_2 - \mathcal{Q}_1\|_\infty \geq \frac{m+m'}{m'}(2k'_Q + k_Q D)\sigma^2\left(f(m+m',\lambda) + f(m,\lambda)\right), \tag{20}$$

where $m' = N + 1 - T_{bp} \geq m_0$*,*

$$f(|I|,\lambda) = \sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}} \tag{21}$$

and where k_Q *and* k'_Q *are constants given in the Appendix, then*

$$\mathbb{P}(I \text{ accepted}) \leq 3\left(\frac{1}{m^\lambda} + \frac{1}{(m+m')^\lambda}\right).$$

The type II error naturally decreases when m and m' increase. When λ increases, the magnitude of the jump in the mean and covariance matrix increases and the type II error decreases.

4. Breakpoint detection using a sliding window

We study a nonparametric breakpoint detection method Müller (1992) for time series (r_t) generated by (1) and satisfying Assumption **(A1)** or Assumption **(A2)**. The method makes use of a *sliding window*. We bound from below the probability of correctly detecting the breakpoint when considering a change-point model with a breakpoint in the means or the covariance matrix.

More precisely, we study two particular change-point models. In the first of these models the covariance matrix $Q_t = \mathcal{Q}$ is constant and the mean changes at T_{bp}: for $t = 1, \ldots, T_{bp} - 1$, $\rho_t = m_1$ and for $t = T_{bp}, \ldots, N$, $\rho_t = m_2$. The second model corresponds to the symmetric case where the mean ρ_t is constant and the covariance matrix Q_t changes at T_{bp}: for $t = 1, \ldots, T_{bp} - 1$, $Q_t = \mathcal{Q}_1$ and for $t = T_{bp}, \ldots, N$, $Q_t = \mathcal{Q}_2$.

4.1. **Algorithm description.** Let $h \in \mathbb{N}^*$ be a positive parameter called the bandwidth corresponding to the length of the left and right sliding windows. For $h \leq t \leq N-h+1$, if there is a breakpoint in the covariance matrix at t, the mean being constant, then the estimation of the covariance matrix on a window of length h whose right endpoint is t (thus using the observations $r_k, k=t-h+1,\ldots,t$) and the estimation of the covariance matrix on a window of length h whose left endpoint is t (thus using the observations $r_k, k=t,\ldots,t+h-1$) should be quite different (if the breakpoint is sufficiently important). Besides, if for $k=t-h+1,\ldots,t+h-1$, $Q_k=\mathcal{Q}$ is constant then the left and right estimations at t should be close. This simple observation allows us to consider a simple nonparametric breakpoint detection method that looks for the instant where the difference between the left and right estimations is the most important. To define formally the breakpoint we need some more notation. For every $t=h,\ldots,N-h+1$, we define the left and right estimations $Q_\ell^t(h)$ and $Q_r^t(h)$ of the covariance matrix at t with

$$Q_\ell^t(h)=\frac{1}{h}\sum_{k=t-h+1}^{t}(r_k-\rho_\ell^t(h))(r_k-\rho_\ell^t(h))^\top,\;\; Q_r^t(h)=\frac{1}{h}\sum_{k=t}^{t+h-1}(r_k-\rho_r^t(h))(r_k-\rho_r^t(h))^\top, \tag{22}$$

where $\rho_\ell^t(h)$ and $\rho_r^t(h)$ are the left and right estimations of the mean:

$$\rho_\ell^t(h)=\frac{1}{h}\sum_{k=t-h+1}^{t}r_k,\;\; \rho_r^t(h)=\frac{1}{h}\sum_{k=t}^{t+h-1}r_k. \tag{23}$$

In this context, if the mean is constant, the estimation $T_{bp}(h)$ of the breakpoint in the covariance matrix when we use a window of length h is the time when the distance between the left and right estimations of the covariance matrix is maximized:

$$T_{bp}(h)=\underset{h\leq t\leq N-h+1}{\text{Argmax}}\;\|Q_\ell^t(h)-Q_r^t(h)\|_\infty. \tag{24}$$

Similarly, we use the following estimator to detect the most important breakpoint in the mean in the case when the variances and covariances are constant:

$$T_{bp}(h)=\underset{h\leq t\leq N-h+1}{\text{Argmax}}\;\|\rho_\ell^t(h)-\rho_r^t(h)\|_\infty. \tag{25}$$

4.2. **Quality of breakpoint detection.** We recall that $K_1(\lambda)$ and $K_2(\lambda)$ are defined by (4) in the previous section.

4.2.1. *Case where the covariance matrix is constant.* If the change in the mean is sufficiently important, we can bound from below as follows the probability for the estimator of T_{bp} to be in an interval of length $2h-2$ containing T_{bp}:

Theorem 4.1. *Let (r_t) be a discrete time stochastic process generated by (1). We assume that for $t=1,\ldots,T_{bp}-1$, $\rho_t=m_1, Q_t=\mathcal{Q}$ and for $t=T_{bp},\ldots,N$, $\rho_t=m_2, Q_t=\mathcal{Q}$. Let Assumption **(A1)** hold, let $\lambda>0$ and $h\in\mathbb{N}^*\backslash\{1\}$ be such that $10(\ln n+\lambda\ln h)\leq 3h$ and $p>2(1+\lambda)$. If $\|m_2-m_1\|_\infty>\frac{4h}{h-1}f_1(h,\lambda)$ where f_1 is defined by (13) then*

$$\mathbb{P}\left(T_{bp}(h)\in\{T_{bp}-h+1,\ldots,T_{bp}+h-2\}\right)\geq 1-\frac{2(N-4h+5)K_1(\lambda)}{h^\lambda}. \tag{26}$$

Using a more restrictive condition on the magnitude of the change in the means, we can also bound from below the probability that $T_{bp}(h)$ equals $T_{bp}-1$ or T_{bp}.

Theorem 4.2. *Let (r_t) be a discrete time stochastic process satisfying Assumption **(A1)**. We assume that for $t=1,\dots,T_{bp}-1$, $\rho_t=m_1, Q_t=\mathcal{Q}$ and for $t=T_{bp},\dots,N$, $\rho_t=m_2, Q_t=\mathcal{Q}$. Let $\lambda>0$ and $h\in\mathbb{N}^*\backslash\{1\}$ such that $10(\ln n+\lambda\ln h)\leq 3h$ and $p>2(1+\lambda)$. If $\|m_2-m_1\|_\infty > 4h\, f_1(h,\lambda)$ with f_1 defined by (13) then*

$$\mathbb{P}\left(\{T_{bp}(h)=T_{bp}-1\}\,\cup\,\{T_{bp}(h)=T_{bp}\}\right)\geq 1-\frac{2(N-2h+2)K_1(\lambda)}{h^{\lambda}}. \tag{27}$$

In what follows, for $h\leq t\leq N-h+1$, we denote by $I_\ell^t=\{t-h+1,\dots,t\}$ the interval of length h with right endpoint t and by $I_r^t=\{t,\dots,t+h-1\}$ the interval of length h with left endpoint t. Assume now that for the algorithm of Section 4.1, we use the left (resp. right) estimations of the covariance matrix and of the mean obtained replacing r_k by $\alpha_k^{I_\ell^t}$ (resp. $\alpha_k^{I_r^t}$) in the expressions of $Q_\ell^t(h)$ and $\rho_\ell^t(h)$ (resp. $Q_r^t(h)$ and $\rho_r^t(h)$) given by (22) and (23). In this case, the probability of detecting the breakpoint can be made arbitrarily large by choosing sufficiently large values of λ, i.e., if the magnitude of the breakpoint is sufficiently large. In this context, Theorems 4.3 and 4.4 which follow are analogous to Theorems 4.1 and 4.2:

Theorem 4.3. *Let (r_t) be a discrete time stochastic process satisfying Assumption **(A2)**. Suppose we use the left (resp. right) estimations of the covariance matrix and of the mean obtained replacing r_k by $\alpha_k^{I_\ell^t}$ (resp. $\alpha_k^{I_r^t}$) in the expressions of $Q_\ell^t(h)$ and $\rho_\ell^t(h)$ (resp. $Q_r^t(h)$ and $\rho_r^t(h)$) given by (22) and (23). We assume that for $t=1,\dots,T_{bp}-1$, $\rho_t=m_1, Q_t=\mathcal{Q}$ and for $t=T_{bp},\dots,N$, $\rho_t=m_2, Q_t=\mathcal{Q}$. Let $\lambda>0$ and $h\in\mathbb{N}^*\backslash\{1\}$. If $\|m_2-m_1\|_\infty>\frac{4h}{h-1}\,f_1'(h,\lambda)$ with*

$$f_1'(h,\lambda)=\sigma\left(\frac{7}{3}\left(\frac{\ln n(n+1)+\lambda\ln h}{h}\right)^{\frac{3}{4}}+\sqrt{2}\sqrt{\frac{\ln n(n+1)+\lambda\ln h}{h}}\right), \tag{28}$$

then

$$\mathbb{P}\left(T_{bp}(h)\in\{T_{bp}-h+1,\dots,T_{bp}+h-2\}\right)\geq 1-\frac{2(N-4h+5)}{h^{\lambda}}. \tag{29}$$

If $\ln n(n+1)+\lambda\ln h\leq h$, the above inequality holds replacing in (28) $f_1'(h,\lambda)$ by $f_1'(h,\lambda)=(\frac{7}{3}+\sqrt{2})\sigma\sqrt{\frac{\ln n(n+1)+\lambda\ln h}{h}}$.

Notice that f_1' increases with n (the number of components), λ and σ, and decreases with h, as expected. Also, when h tends to infinity, then $f_1'(h,\lambda)$ tends to 0. If h is not too large and λ sufficiently large, then this theorem shows that the estimator $T_{bp}(h)$ is of good quality. But when λ increases, so does f_1'. This means that the estimator will of course be all the more accurate as the change in the mean is important. Regarding the condition $\ln n(n+1)+\lambda\ln h\leq h$, if the number of components $n=30$, and if we choose $\lambda=0.5$, then the above condition implies $h\geq 16$. This condition is not too restrictive since the numerical simulations of the next section tend to show that "good" values of h are rather large.

Theorem 4.4. *Let Assumption **(A2)** hold and suppose we use the left (resp. right) estimations of the covariance matrix and of the mean obtained replacing r_k by $\alpha_k^{I_\ell^t}$ (resp. $\alpha_k^{I_r^t}$) in the expressions of $Q_\ell^t(h)$ and $\rho_\ell^t(h)$ (resp. $Q_r^t(h)$ and $\rho_r^t(h)$) given by (22) and (23). We assume that for $t=1,\dots,T_{bp}-1$, $\rho_t=m_1, Q_t=\mathcal{Q}$ and for $t=T_{bp},\dots,N$, $\rho_t=m_2, Q_t=\mathcal{Q}$. Let $\lambda>0$ and*

$h \in \mathbb{N}^*\backslash\{1\}$. *If* $\|m_2 - m_1\|_\infty > 4h\, f_1'(h,\lambda)$, *with* f_1' *given by* (28) *then*

$$\mathbb{P}\left(\{T_{bp}(h) = T_{bp} - 1\} \,\cup\, \{T_{bp}(h) = T_{bp}\}\right) \geq 1 - \frac{2(N-2h+2)}{h^\lambda}. \tag{30}$$

4.2.2. *Case where the mean is constant.* When the mean is constant, the following theorem provides a lower bound for the probability that the estimator of T_{bp} belongs to an interval of length $2h-2$ containing T_{bp}:

Theorem 4.5. *Let* (r_t) *be a discrete time stochastic process satisfying Assumption **(A1)**. Let* $\lambda > 0$ *and* $h \in \mathbb{N}^*\backslash\{1\}$ *such that* $10(\ln n(n+1) + \lambda \ln h) \leq 3h$ *and* $E[\frac{p}{2}] > 2(1+\lambda)$. *We assume that for* $t = 1, \ldots, T_{bp}-1$, $\rho_t = m, Q_t = \mathcal{Q}_1$ *and for* $t = T_{bp}, \ldots, N$, $\rho_t = m, Q_t = \mathcal{Q}_2$. *If* $\|\mathcal{Q}_2 - \mathcal{Q}_1\|_\infty > \frac{4h}{h-1}\, f_2(h,\lambda)$ *where*

$$f_2(h,\lambda) = 2\sqrt{\frac{10}{3}}(2\sigma^2 + \sigma'')\sqrt{\frac{\ln n(n+1) + \lambda \ln h}{h}}, \tag{31}$$

and $\sigma'' = \sigma^2 + (2\sigma' + \sigma)^2$, *then*

$$\mathbb{P}\left(T_{bp}(h) \in \{T_{bp} - h + 1, \ldots, T_{bp} + h - 2\}\right) \geq 1 - \frac{2(N-4h+5)(K_1(\lambda) + K_2(\lambda))}{h^\lambda}. \tag{32}$$

Using a more restrictive condition on the magnitude of the change in the covariance matrix, we can bound from below the probability that the estimator of T_{bp} equals T_{bp} or $T_{bp}-1$:

Theorem 4.6. *Let* (r_t) *be a discrete time stochastic process satisfying Assumption **(A1)**. Let* $\lambda > 0$ *and* $h \in \mathbb{N}^*\backslash\{1\}$ *such that* $10(\ln n(n+1) + \lambda \ln h) \leq 3h$ *and* $E[\frac{p}{2}] > 2(1+\lambda)$. *We assume that for* $t = 1, \ldots, T_{bp}-1$, $\rho_t = m, Q_t = \mathcal{Q}_1$ *and for* $t = T_{bp}, \ldots, N$, $\rho_t = m, Q_t = \mathcal{Q}_2$. *If* $\|\mathcal{Q}_2 - \mathcal{Q}_1\|_\infty > 4\,h\, f_2(h,\lambda)$ *where* f_2 *is defined by* (31) *then*

$$\mathbb{P}\left(\{T_{bp}(h) = T_{bp} - 1\} \,\cup\, \{T_{bp}(h) = T_{bp}\}\right) \geq 1 - \frac{2(N-2h+2)(K_1(\lambda) + K_2(\lambda))}{h^\lambda}. \tag{33}$$

Finally, the quality of covariance matrix breakpoint detection using $\alpha_k^{I_\ell^t}$ or $\alpha_k^{I_r^t}$ instead of r_k is given in the following theorem:

Theorem 4.7. *Let Assumption **(A2)** hold and suppose that we use the left (resp. right) estimations of the covariance matrix and of the mean obtained replacing* r_k *by* $\alpha_k^{I_\ell^t}$ *(resp.* $\alpha_k^{I_r^t}$*) in the expressions of* $Q_\ell^t(h)$ *and* $\rho_\ell^t(h)$ *(resp.* $Q_r^t(h)$ *and* $\rho_r^t(h)$*) given by* (22) *and* (23). *Let* $\lambda > 0$ *and* $h \in \mathbb{N}^*\backslash\{1\}$ *such that* $\ln n(n+1) + \lambda \ln h \leq h$. *If*

$$\|\mathcal{Q}_2 - \mathcal{Q}_1\|_\infty > \frac{4k_Q'}{h-1}\,\sigma^2\,\sqrt{h\,(\ln n(n+1) + \lambda \ln h)}$$

(where k_Q' *from Theorem 3.8 is defined in the Appendix), then the probability that* $T_{bp}(h)$ *belongs to* $\{T_{bp} - h + 1, \ldots, T_{bp} + h - 2\}$ *is greater than* $1 - \frac{4(N-4h+5)}{h^\lambda}$ *and if*

$$\|\mathcal{Q}_2 - \mathcal{Q}_1\|_\infty > 4k_Q'\,\sigma^2\,\sqrt{h\,(\ln n(n+1) + \lambda \ln h)},$$

then the probability that the estimator $T_{bp}(h)$ *equals* $T_{bp}-1$ *or* T_{bp} *is greater than* $1 - \frac{4(N-2h+2)}{h^\lambda}$.

5. Numerical simulations

In this section, we test the performance of the breakpoint detection methods studied in this paper using Gaussian change-point models with one or two breakpoints in the mean and/or the covariance matrix. Random vectors have $n = 30$ components and the tests are grouped in three sections:

- In Section 5.1, we test various implementations of the adaptive algorithm introduced in this paper.
- In Section 5.2, simulations are performed with the algorithm from Section 4.1.
- In the last Section 5.3, the adaptive algorithm of the present paper is compared with the adaptive algorithm from Guigues (2008) and with the algorithm from Section 4.1.

5.1. **Adaptive method.** We consider a Gaussian change-point model where the first T_1 observations are drawn from the Gaussian $\mathcal{N}(\beta_1\rho, \beta_2 Q)$ density and the last T_2 observations from the Gaussian $\mathcal{N}(\beta_1'\rho, \beta_2' Q)$ density for some positive parameters T_1, T_2 and different values for parameters $(\beta_1, \beta_2, \beta_1', \beta_2')$. We take for ρ and Q, the empirical mean and covariance matrix of the 3 month returns of the assets of the Dow Jones 30, computed on January 2, 2002, using 3 years of historical data, and assuming stationarity of the returns. The maximal mean return is 1.1 and the maximal variance 0.1. Such choice is motivated by the fact that the adaptive algorithm in the multivariate case Guigues (2008) was first applied to estimate the one step ahead mean return ρ and the one step ahead covariance matrix Q between the returns for a portfolio management model.

Intervals in $\mathcal{I}$ with length proportional to m_0. We first consider a change in the mean: $\beta_1' = 1.5$, $\beta_1 = \beta_2 = \beta_2' = 1$, $T_1 = 120$, $T_2 = 240$ and we choose for the set $\mathcal{I}$ at time step t intervals of the form $\{t - km_0, \ldots, t-1\}$ where $m_0 = 15$ and $k \in \mathbb{N}^*$. Our goal is to simulate such model and for each trajectory of the process to apply the adaptive algorithm as time goes by to determine estimations of the ILTH and of the one step ahead mean and covariance matrix.

Before reporting the results, it remains to explain how the parameters of the adaptive algorithm are chosen. Recall that the adaptive algorithm from Section 3.1 depends on two positive parameters λ and μ. Using Theorem 3.1, we accept I as an ILTH if for every $J \in \mathcal{I}(I)$, $\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \le k_1(\sigma + \sigma')(\sqrt{\frac{\ln n + \lambda \ln |I|}{|I|}} + \sqrt{\frac{\ln n + \lambda \ln |J|}{|J|}})$ and $\|\hat{Q}_I - \hat{Q}_J\|_\infty \le k_2(\sigma + \sigma')^2(\sqrt{\frac{\ln n(n+1) + \mu \ln |I|}{|I|}} + \sqrt{\frac{\ln n(n+1) + \mu \ln |J|}{|J|}})$ where (λ, μ, k_1, k_2) are positive constants. Parameter $\sigma + \sigma'$ will be estimated by the largest available realization $r_t(i)$ among all time steps t and components i. Theorem 3.1 provides constants k_1 and k_2 (independent on λ and μ) such that lower bounds are known for the probabilities that the corresponding above inequalities are satisfied. Of course, such inequalities will be satisfied with at least the same probability for larger values of k_1 and k_2. However, the choice of constants k_1 and k_2 provided by Theorem 3.1 may be conservative. The larger these constants, the more we will consider that an interval where the mean and covariance matrix are slowly varying is an ILTH but the least we will detect breakpoints. In the particular case of a constant one-dimensional process, for all I, J, $\hat{\rho}_I = \hat{\rho}_J$ and $\hat{Q}_I = \hat{Q}_J$, which means the above inequalities hold in this situation with probability one with $k_1 = k_2 = 0$. To prevent too large values for k_1 and k_2 while detecting the absence of breakpoints with a high probability, we should consider k_1 and k_2 as parameters of our algorithm (thus introducing two additional degrees of freedom).

We propose 4 error criteria to calibrate the parameters (λ, μ, k_1, k_2) of the adaptive algorithm. We take a discrete set of candidate values for parameters (λ, μ, k_1, k_2): $\{0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1\}$ for λ and μ and $\{0.01, 0.012, 0.014, 0.016, 0.018, 0.02, 0.03, 0.04, 0.05, 0.06,$

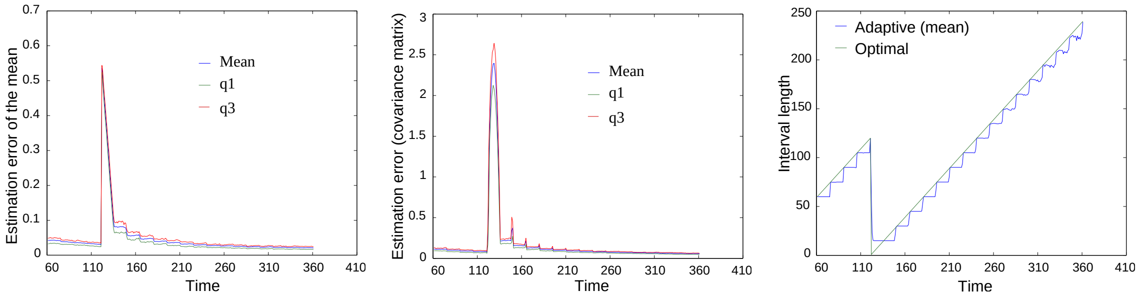

FIGURE 1. From left to right, estimation error of the mean, estimation error of the covariance matrix, and comparison between the optimal and estimated maximal intervals of homogeneity for a Gaussian process having a breakpoint in the mean.

$0.07, 0.08, 0.09, 0.1\}$ for k_1 and k_2. Considering this grid of values for (λ, μ, k_1, k_2), for each combination of the parameters, we make 100 simulations of the change-point model. For each combination, each simulation, and each time $t = t_0, \ldots, T_1 + T_2$, with $t_0 = 60$, on the basis of the past observations for time steps $1, \ldots, t-1$, we compute the adaptive one step ahead mean $\hat{\rho}_t$ and the adaptive one step ahead covariance matrix $\hat{Q}_t$ (as explained in Section 3). Setting $T = T_1 + T_2 - t_0 + 1$, we then define the 4 following error criteria:

$$MSE = \mathbb{E}[\frac{1}{T}\sum_{t=t_0}^{T_1+T_2} \|\hat{\rho}_t - \rho_t\|_\infty^2 + \frac{1}{T}\sum_{t=t_0}^{T_1+T_2} \|\hat{Q}_t - Q_t\|_\infty^2], \quad MSFE = \mathbb{E}[\frac{1}{T}\sum_{t=t_0}^{T_1+T_2} \|\hat{\rho}_t - r_t\|_\infty^2],$$

$$MAE = \mathbb{E}[\frac{1}{T}\sum_{t=t_0}^{T_1+T_2} \|\hat{\rho}_t - \rho_t\|_\infty + \frac{1}{T}\sum_{t=t_0}^{T_1+T_2} \|\hat{Q}_t - Q_t\|_\infty], \quad MAFE = \mathbb{E}[\frac{1}{T}\sum_{t=t_0}^{T_1+T_2} \|\hat{\rho}_t - r_t\|_\infty].$$

The chosen values of (λ, μ, k_1, k_2) are those providing the best (smallest) empirical MSE, MSFE, MAE, or MAFE.

Using the parameters minimizing the (estimated) MAE, we then make 100 new simulations of the change-point model. For each simulation and each time $t = t_0, \ldots, T_1 + T_2$, with $t_0 = 60$, on the basis of the past observations for time steps $1, \ldots, t-1$, we compute the adaptive one step ahead mean $\hat{\rho}_t$ and the adaptive one step ahead covariance matrix $\hat{Q}_t$. For this experiment, we report in Figure 1 the following results:

(i) an estimation of the mean and of the first and third quartiles (denoted here and in what follows by q_1 and q_3) of $\|\hat{\rho}_t - \rho_t\|_\infty$ for each instant t;

(ii) an estimation of the mean and of the first and third quartiles of $\|\hat{Q}_t - Q_t\|_\infty$ for each instant t;

(iii) the mean length of the estimated maximal interval of local time homogeneity together with the maximal (optimal) interval of local time homogeneity.

The estimation error of the mean and of the covariance matrix rapidly decreases after the breakpoint to re-find the error level before the breakpoint. The change in the mean is rapidly detected. Indeed, the mean value of the first instant t for which the homogeneity interval is not $\{1, \ldots, t\}$ is $t = 122$, i.e., two time steps after the breakpoint. Also, in mean, 15 time steps are necessary to obtain an ILTH whose left endpoint is $t = 121$. By definition of $\mathcal{I}$, the length of the homogeneity interval is at least $m_0 = 15$. With this implementation of the adaptive algorithm, $t = 135$ is the first time step for which our algorithm can provide the (optimal) maximal interval of local time homogeneity.

Intervals in $\mathcal{I}$ with length greater than or equal to m_0. In order to decrease the detection time, we now (and in the sequel) take for the set $\mathcal{I}$ at time step t intervals of form $\{t - m_0 - k + 1, \ldots, t - 1\}$ with $k \in \mathbb{N}^*$ (we recall that $m_0 = 15$).

We study 3 change-point models. For the first of these models, we choose $T_1 = 120$, $T_2 = 240$, $t_0 = 60$, $\beta_1' = 1.3$, and $\beta_1 = \beta_2 = \beta_2' = 1$ (change in the mean). The second model considers a change in the covariance matrix and is described by the following values of the parameters: $T_1 = 300$, $T_2 = 250$, $t_0 = 201$, $\beta_1 = \beta_1' = 1$, and $\beta_2 = 0.5$, $\beta_2' = 2$. We finally consider a breakpoint in both the mean and the covariance matrix taking $T_1 = 120$, $T_2 = 240$, $t_0 = 60$, $\beta_1 = 1$, $\beta_1' = 1.3$, $\beta_2 = 0.5$, and $\beta_2 = 2$. The parameters chosen to implement the adaptive algorithm are those minimizing the MAE in a preliminary simulation phase as explained above. For these models, the items (i), (ii), and (iii) defined above, i.e., the estimation error of the mean (mean and quartiles), the estimation error of the covariance matrix (mean and quartiles), as well as the the length of the estimated maximal interval of homogeneity (mean and quartiles) are represented in Figure 2.

We observe that breakpoints in the mean are more quickly detected. However, in all cases, when the last breakpoint occurred at a sufficiently remote time instant, the ILTH is well estimated, i.e., the empirical mean of the breakpoint estimator is close to the breakpoint instant and the empirical quartiles of this estimator are close to the empirical mean.

Test with two breakpoints. The algorithm can select different maximal intervals of homogeneity for the mean and for the covariances if the last important breakpoint in the mean occurred before the last important breakpoint in the covariance matrix. We illustrate this using a change-point Gaussian model. The first $T_1 = 140$ observations are drawn from a Gaussian $\mathcal{N}(\rho, 0.2Q)$ density (recall that ρ and Q are defined at the beginning of Section 5.1), the next $T_2 = 200$ observations from a Gaussian $\mathcal{N}(1.3\rho, 0.2Q)$ density, and the last $T_3 = 300$ observations from a Gaussian $\mathcal{N}(1.3\rho, 2Q)$ density. Choosing the same grid of values as before for the parameters of the adaptive algorithm, we simulate for each combination of these values 100 realizations of this process. The adaptive intervals are estimated for each instant $t \geq 80$. The values of the parameters providing the best MAE are $(\lambda, \mu, k_1, k_2) = (0.09, 0.08, 0.08, 0.014)$ and gave the quartiles of the adaptive interval length represented in Figure 3. Once again, we see that our algorithm more easily detects breakpoints in the mean than in the covariance matrix, the mean detection time being reasonable.

We refer to Guigues (2005) for additional numerical simulations that illustrate the calibration of the parameters when noises ζ_t are Gaussian with, in particular, the variances of the components in a given range. In these simulations, the following was naturally observed: First, the probability to reject an ILTH increases with the length of this interval. Second, this probability is a decreasing function of the grid step m_0. Indeed, for small values of the grid step, the homogeneity tests are more numerous. Finally, the mean detection time decreases with the magnitude of the change and for fixed m_0, k_1, k_2, and λ (resp. μ), the probability to reject an ILTH is a decreasing function of μ (resp. λ).

5.2. **Breakpoint detection using a sliding window.** We test the algorithm from Section 4.1 with the following change-point model. The observations r_t, $t = 1, \ldots, T_{bp} - 1 = 250$ are

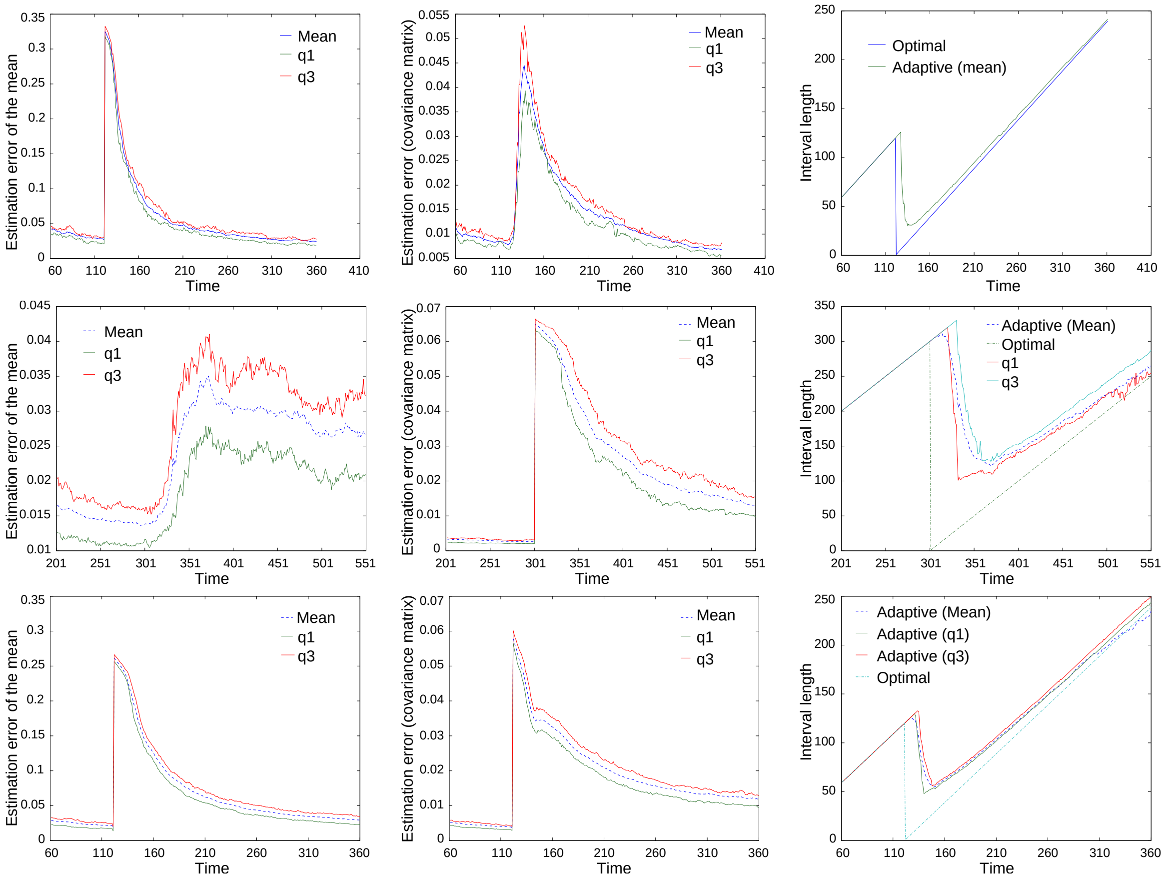

FIGURE 2. From top to bottom, breakpoint in the mean, in the covariance matrix, and in both. From left to right, estimation error of the mean, estimation error of the covariance matrix, and comparison between the lengths of the optimal and estimated maximal intervals of homogeneity.

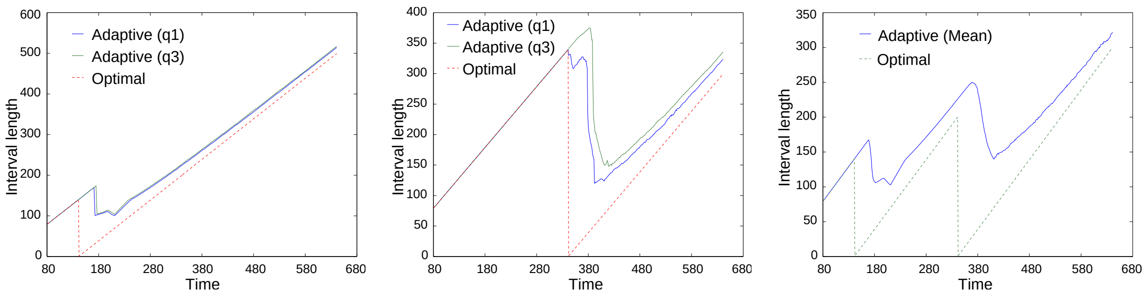

FIGURE 3. From left to right, adaptive interval mean length for respectively the mean, the covariance matrix and for both, compared with the optimal interval length.

drawn from a Gaussian $\mathcal{N}(m_i, Q_j)$ density and the remaining observations r_t, $t = 251, \ldots, 750$ from a Gaussian $\mathcal{N}(m_{i'}, Q_{j'})$ density for different combinations of the pairs (m_i, Q_j) where m_i and Q_j are means and covariance matrices defined as follows. If $\delta = \frac{0.05}{n}$ with $n = 30$ (the number of components of the random vectors), the means m_j are given for $i = 1, \ldots, n$, by: $m_1(i) = 1.15 + i\delta$, $m_2(i) = 1.2 + i\delta$, $m_3(i) = 1.4 + i\delta$, and $m_4(i) = 1.7 + i\delta$. The covariance matrices Q_j are diagonal matrices such that for $i = 1, \ldots, n$: $Q_1(i,i) = \frac{in(n+1)}{18}\delta^2$, $Q_2(i,i) = \frac{5in(n+1)}{36}\delta^2$, $Q_3(i,i) = \frac{2in(n+1)}{9}\delta^2$, and $Q_4(i,i) = \frac{in(n+1)}{2}\delta^2$. We obtain the matrices Q_2, Q_3, and Q_4 multiplying the standard deviations given in matrix Q_1 by $1.5, 2$, and 3 respectively. This data are close to those used in the numerical simulations of Ben-Tal and Nemirovski (1999).

We consider six change-point models and we measure the quality of breakpoint detection when the magnitude of the change increases. For the first three (resp. the last three) models, the covariance matrix (resp. the mean) is constant, set to Q_1 (resp. set to m_1), and the mean is set to m_1 (resp. the covariance matrix is set to Q_1) before T_{bp} while it is respectively m_2, m_3, and m_4 (Q_2, Q_3, and Q_4) after $T_{bp}-1$ for the first, second, and third models (resp. for the last three models). These models will be respectively denoted by (m_1, m_2, Q_1), (m_1, m_3, Q_1), (m_1, m_4, Q_1), (m_1, Q_1, Q_2), (m_1, Q_1, Q_3), and (m_1, Q_1, Q_4).

For each bandwidth $h = 2, 3, \ldots, 200$, we generate 400 simulations of each of these change-point models and we apply the algorithm from Section 4.1 to obtain estimations of the breakpoint for these simulations. We are interested in the evolution of the mean and standard deviation of the estimator of the breakpoint when h increases. The probability of correctly detecting the breakpoint, i.e., that $T_{bp}(h) = T_{bp}$ or that $T_{bp}(h) = T_{bp} - 1$ is also considered. The results are represented in Figures 4 and 5.

Breakpoints in the mean are very well detected. More precisely, for model (m_1, m_2, Q_1), for values of the bandwidth sufficiently large, say above 40, the empirical mean of the estimator of the breakpoint is close to T_{bp-1} (small bias) and the standard deviation of this estimator is small. When the magnitude of the breakpoint in the mean increases, i.e., for models (m_1, m_3, Q_1) and (m_1, m_4, Q_1), the instant of the breakpoint is correctly detected for nearly all values of h and all realizations of the process.

The breakpoints in the variances and covariances have to be important to obtain a good estimator of the breakpoint. However, for large values of the bandwidth h, say above 100, the corresponding estimator $T_{bp}(h)$ of the instant of the breakpoint has both a small bias and a small variance.

5.3. **Comparative study.** We consider the 3 change point models in the mean of the previous section and generate 400 trajectories of these processes. For each realization, the adaptive algorithm from Guigues (2008), the adaptive algorithm from Section 3.1, and the algorithm from Section 4.1 are used to estimate the instant of the breakpoint ($T_{bp} - 1 = 250$ or $T_{bp} = 251$). We are interested in the empirical mean and standard deviation of the estimator of the breakpoint. For the algorithm from Section 4.1, we choose a value of the bandwidth h yielding the smallest MAE given the data r_t, $t = 1, \ldots, 750$. For the other two algorithms, the parameters providing the smallest MAE given the data r_t, $t = 1, \ldots, 750$, are chosen. For this experiment, the empirical mean and standard deviation of the estimator of the breakpoint

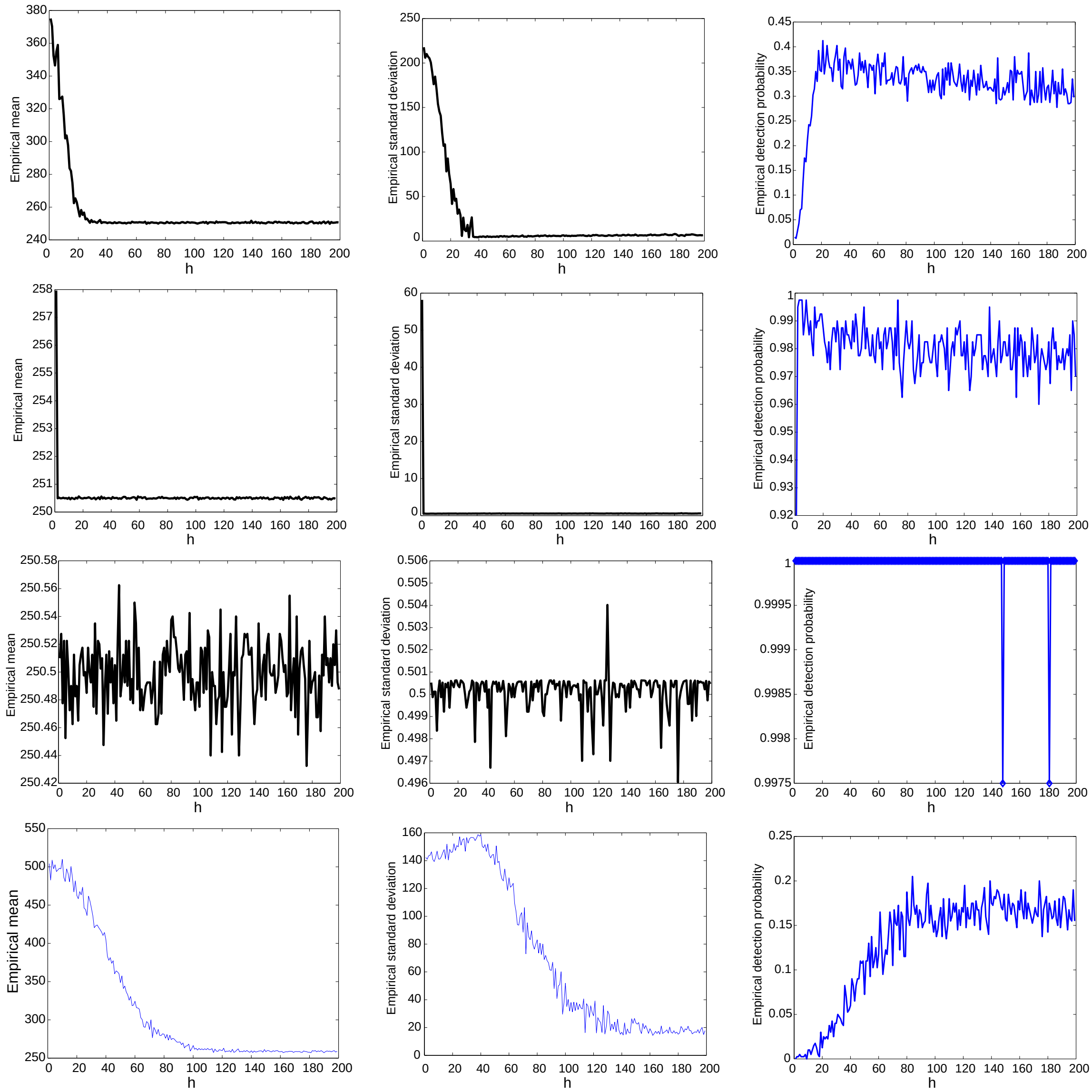

FIGURE 4. From left to right, the empirical mean and standard deviation of the estimator of the breakpoint as well as the empirical probability of correctly detecting the breakpoint are represented for values of h going from 2 to 200 using 400 realizations of Gaussian change-point models. From top to bottom, models (m_1, m_2, Q_1), (m_1, m_3, Q_1), (m_1, m_4, Q_1), and (m_1, Q_1, Q_2).

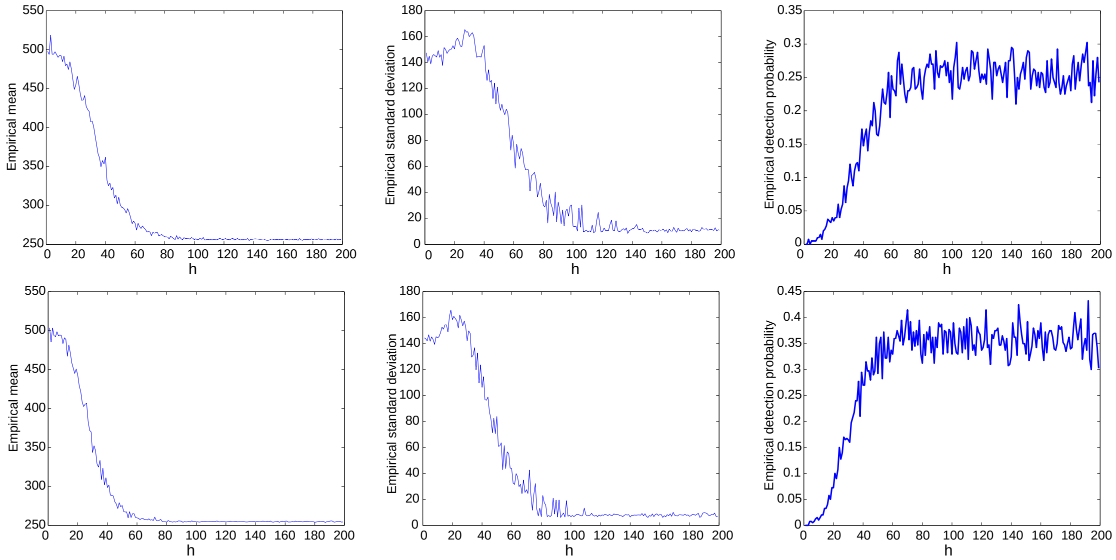

FIGURE 5. From left to right, the empirical mean and standard deviation of the estimator of the breakpoint as well as the empirical probability of correctly detecting the breakpoint are represented for values of h going from 2 to 200 using 400 realizations of Gaussian change-point models. From top to bottom, models (m_1, Q_1, Q_3) and (m_1, Q_1, Q_4).

are reported in Table 1. In this table, we also report the values of the parameters of the algorithm that were selected. We recall that the algorithm from Guigues (2008) has three parameters λ, k_1, and k_2 and that with this algorithm an interval I is accepted as an ILTH if for every $J \in \mathcal{I}(I)$, we have $\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \leq k_1 \sigma(\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}} + \sqrt{\frac{\ln n(n+1) + \lambda \ln |J|}{|J|}})$ and $\|\hat{Q}_I - \hat{Q}_J\|_\infty \leq k_2 \sigma^2(\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}} + \sqrt{\frac{\ln n(n+1) + \lambda \ln |J|}{|J|}})$. Parameter σ is estimated by the largest available realization $r_t(i)$ among all time steps t and components i. The values tested for λ, μ are $\{0.01, 0.04, 0.08, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1\}$ while the candidate values for k_1 and k_2 are $\{0.05, 0.055, 0.06, 0.065, 0.07, 0.075, 0.08, 0.085, 0.09, 0.095, 0.1\}$. Two values are tested for m_0: $m_0 = 15$ and $m_0 = 30$. Moreover, we consider two implementations of the adaptive algorithms corresponding to two different choices of $\mathcal{I}(I)$. For the first implementation (referred to as *implementation choice 1* in Table 1), we take for $\mathcal{I}(I)$ all the intervals of length proportional to m_0 strictly embedded in I with either the same left endpoint or the same right endpoint as I. For the second implementation (referred to as *implementation choice 2* in Table 1), we take two intervals in $\mathcal{I}(I)$ of length m_0: the first one has the same right endpoint as I and the second one has the same left endpoint as I.

We observe that the bias and standard deviation (s.d) of the estimator of the breakpoint are small for the algorithm from Section 4.1. This is in accordance with the results of the previous section and confirms the efficiency of this method when we know there is only one breakpoint in the time series. The adaptive algorithms perform less well on this example but still provide a small bias and a reasonably small s.d., especially when the magnitude of the change is important. We see that the adaptive algorithm from Section 3.1 which has one additional degree of freedom provides a better estimator of the breakpoint than Guigues (2008) for models (m_1, m_3, Q_1) and

Method	Model	m_1, m_2, Q_1	m_1, m_3, Q_1	m_1, m_4, Q_1
Sliding window	h	37	3	150
	Mean	250.17	250.47	250.43
	s.d	4.27	0.50	0.49
Adaptive	k_1	0.08	0.055	0.05
	k_2	0.08	0.04	0.05
	λ	0.04	0.5	0.8
	μ	0.04	0.6	0.1
	m_0	15	30	15
	Imp. choice	1	2	1
	Mean	240.62	244.60	249.10
	s.d	17.84	0.94	0.28
Guigues (2008)	k_1	0.06	0.05	0.05
	k_2	0.06	0.05	0.05
	λ	0.04	0.5	0.8
	m_0	15	15	15
	Imp. choice	1	1	1
	Mean	236.83	246.60	248.81
	s.d	7.69	1.58	0.44

TABLE 1. Empirical mean and standard deviation (s.d.) of the estimator of the breakpoint for three nonparametric multivariate breakpoint detection methods.

(m_1, m_4, Q_1). If the method based on a sliding window was the best in this example, one should recall that the hypotheses required to use it are more restrictive. Moreover, the adaptive methods allow us to determine successive breakpoints in the time series, which the sliding window-based algorithm is not able to do.

Finally, we also observed that the adaptive algorithm is very sensitive to the choice of its parameters. The quality of breakpoint detection could be improved testing for a larger number of candidate values for these parameters. This would however increase the computational time of the simulation phase aiming at calibrating these parameters.

6. Conclusion

We studied two nonparametric breakpoint detection methods in the means, variances, and covariances of a multivariate discrete time stochastic process in a nonparametric setting. We provided the theoretical and practical efficiency of these methods for a change-point time series.

The method using a sliding window allows us to detect the most important breakpoint in the mean, variances, and covariances on a given period. For this method to be efficient, it should be used when there is only one breakpoint and at least a few data come from the two distributions. The adaptive algorithm needs no assumption on the number of breakpoints. It aims to determine the breakpoints delimiting intervals of homogeneity. The performances of both algorithms depend on the choice of their parameters (the pair (λ, μ) for the adaptive algorithm and the bandwidth h for the other). They proved to be very efficient in detecting breakpoints in the mean. Breakpoints in the variances and covariances are less well detected though the simulations exhibited a small bias and a relatively small standard deviation for the estimators of breakpoints in the covariance matrix.

Acknowledgments

The author is grateful to Anatoli Juditski for helpful discussions. We would also like to thank the two reviewers and the Associate Editor for beneficial comments and suggestions.

References

[1] A. Antoniadis and I. Gijbels. Detecting abrupt changes by wavelet methods. *Journal of Nonparametric Statistics*, 14:7–29, 2002.

[2] M. Basseville and V. Nikiforov. *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, 1993.

[3] A. Ben-Tal and A. Nemirovski. Robust solutions to uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.

[4] Z. Cai, J. Fan, and R. Li. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95:888–902, 2000.

[5] F. Désobry, M. Davy, and C. Doncarli. An online kernel change detection algorithm. *IEEE Transactions on Signal Processing*, 53:2961–2974, 2005.

[6] J. Fan, J. Jiang, C. Zhang, and Z. Zhou. Time-dependent diffusion models for term structure dynamics. Statistical applications in financial econometrics. *Statistica Sinica*, 13:965–992, 2003.

[7] J. Fan and W. Zhang. Statistical estimation in varying coefficient models. *Annals of Statistics*, 27:1491–1518, 1999.

[8] J. Fridlyand, A. Snijders, D. Albertson D. Pinkel, and Jain A. Hidden Markov models approach to the analysis of array CGH data. *Journal of Multivariate Analysis*, 90:132–153, 2004.

[9] V. Guigues. Inférence statistique pour l'optimisation stochastique. Applications en finance et en gestion de production. *PhD thesis, Université Joseph Fourier*, 2005. `http://tel.archives-ouvertes.fr/tel-00098287/en/`.

[10] V. Guigues. Mean and covariance matrix adaptive estimation for a weakly stationary process. Application in stochastic optimization. *Statistics & Decision*, 26:109–143, 2008.

[11] V. Guigues. Robust production management. *Optimization and Engineering*, 10:505–532, 2009.

[12] Z. Harchaoui, F. Bach, and E. Moulines. Kernel change-point analysis. *In Advances in Neural Information Processing Systems*, 2008.

[13] H. Heitsch and W. Römisch. Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406, 2009.

[14] O. Lepski. On a problem of adaptive estimation in Gaussian white noise. *Theory of Probability and its Applications*, 35:454–466, 1990.

[15] O. Lepski and V. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, 25:2512–2546, 1997.

[16] C. Lévy-Leduc and F. Roueff. Detection and localization of change-points in high-dimensional network traffic data. *Annals of Applied Statistics*, 3:637–662, 2009.

[17] D. Mercurio and V. Spokoiny. Statistical inference for time-inhomogeneous volatility models. *Annals of Statistics*, 32:577–602, 2004.

[18] T. Mikosch and C. Starica. Change of structure in financial time series, long range dependence and the GARCH model. *Technical Report, Aarhus School of Business, Aarhus Univ.*, 58, 2000.

[19] H.G. Müller. Change-point in nonparametric regression analysis. *Annals of Statistics*, 20:737–761, 1992.

[20] V. Spokoiny. Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Annals of Statistics*, 26:1356–1378, 1998.

[21] R. Tibshirani and P. Wang. Spatial smoothing and hot spot detection for CGH data using the fused lasso. *Biostatistics*, 8:1–12, 2007.

[22] Y. Wang. Jump and sharp cusp detection by wavelets. *Biometrika*, 82:385–397, 1995.

[23] N.R. Zhang and D.O. Siegmund. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics*, 63:22–32, 2007.

Appendix: proof of the theorems

To show Theorem 3.1, we first need to show the following lemma which gives for the stationary case the (non-asymptotic) accuracy of the following estimators $\hat{\rho}_N$, $\hat{Q}_N^b$, and $\hat{Q}_N$:

$$\hat{\rho}_N = \frac{1}{N}\sum_{k=1}^{N} r_k, \quad \hat{Q}_N^b = \frac{1}{N}\sum_{k=1}^{N}(r_k - \rho_k)(r_k - \rho_k)^\top, \quad \hat{Q}_N = \frac{1}{N}\sum_{k=1}^{N}(r_k - \hat{\rho}_N)(r_k - \hat{\rho}_N)^\top. \tag{34}$$

$\hat{Q}_N^b$ is introduced to provide the accuracy of $\hat{Q}_N$.

Lemma 6.1. *Let $r_t, t = 1, \ldots, N$, be N independent observations generated from model* (1) *such that for every t, $\rho_t = \rho$, $Q_t = Q$. Let $\lambda > 0$ and let Assumption **(A1)** hold with $E[\frac{p}{2}] > 2 + 2\lambda$. We define*

$$K_1(\lambda) = 2 + \left(\frac{10\,p(\gamma+\lambda)}{3e(p-2-2\lambda)}\right)^{\frac{p}{2}}, \quad K_2(\lambda) = 1 + \left(\frac{10\,E[\frac{p}{2}](2\gamma+\lambda)}{3\,e(E[\frac{p}{2}]-2-2\lambda)}\right)^{\frac{E[\frac{p}{2}]}{2}}$$

where γ is chosen such that $n \le N^\gamma$. If N is sufficiently large, i.e., $3N \ge 10(\ln n(n+1) + \lambda \ln N)$, then the estimators $\hat{\rho}_N$, $\hat{Q}_N^b$, and $\hat{Q}_N$ defined in (34) *satisfy*

$$\mathbb{P}\left(\|\hat{\rho}_N - \rho\|_\infty \ge 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda\ln N}{N}}\right) \le \frac{K_1(\lambda)}{N^\lambda}, \tag{35}$$

$$\mathbb{P}\left(\|\hat{Q}_N^b - Q\|_\infty \ge 4\sqrt{\frac{10}{3}}\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda\ln N}{N}}\right) \le \frac{K_2(\lambda)}{N^\lambda}, \tag{36}$$

$$\mathbb{P}\left(\|\hat{Q}_N - Q\|_\infty \ge 8\sqrt{\frac{10}{3}}\sigma^2\sqrt{\frac{\ln n(n+1) + \lambda\ln N}{N}}\right) \le \frac{K_1(\lambda)+K_2(\lambda)}{N^\lambda}. \tag{37}$$

In what follows, to alleviate notation, the left-hand sides of inequalities of form (35), (36), and (37) will be written $\mathbb{P}(\|\hat{\rho}_N - \rho\|_\infty \ge \eta_1)$, $\mathbb{P}(\|\hat{Q}_N^b - Q\|_\infty \ge \eta_2')$, and $\mathbb{P}(\|\hat{Q}_N - Q\|_\infty \ge \eta_2)$.

Proof. Since $3N \ge 10(\ln n + \lambda \ln N)$ and $p \ge 2$,

$$\eta_1 = 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda\ln N}{N}} \ge \tilde{\eta}_1 + \frac{\sigma^p}{A^{p-1}},$$

with

$$\tilde{\eta}_1 = \sigma\sqrt{\frac{10}{3}}\sqrt{\frac{\ln n + \lambda\ln N}{N}} \quad \text{and} \quad A = \sigma\sqrt{\frac{3}{10}}\sqrt{\frac{N}{\ln n + \lambda\ln N}}.$$

We then define $\bar{\zeta}_t = \zeta_t 1(\|\zeta_t\|_\infty \le A)$. We have

$$\begin{aligned}\mathbb{P}\left(\|\frac{1}{N}\sum_{t=1}^{N}\zeta_t\|_\infty \ge \eta_1\right) &\le \mathbb{P}\left(\|\frac{1}{N}\sum_{t=1}^{N}\bar{\zeta}_t\|_\infty \ge \eta_1\right) + \mathbb{P}\left(\max_{1\le t\le N}\|\zeta_t\|_\infty > A\right)\\ &= p_1 + p_2.\end{aligned} \tag{38}$$

We then have for p_2

$$p_2 \le \sum_{t=1}^{N}\mathbb{P}(\|\zeta_t\|_\infty > A) \le \sum_{t=1}^{N}\frac{\mathbb{E}[\|\zeta_t\|_\infty^p]}{A^p} \le \frac{N\sigma^p}{A^p} \le \left(\frac{10(\gamma+\lambda)}{3}\right)^{\frac{p}{2}}\frac{(\ln N)^{\frac{p}{2}}}{N^{\frac{p}{2}-1}}.$$

Since $N \geq 2$, $\lambda > 0$, and $p > 2(1+\lambda)$, we obtain

$$p_2 \leq \left(\frac{10\, p(\gamma+\lambda)}{3e(p-2-2\lambda)} \right)^{\frac{p}{2}} \frac{1}{N^{\lambda}}. \tag{39}$$

Let us now estimate p_1. First notice that

$$\|\mathbb{E}[\bar{\zeta}_t]\|_\infty = \|\mathbb{E}[\zeta_t 1(\|\zeta_t\|_\infty > A)]\|_\infty \leq \frac{\sigma^p}{A^{p-1}},$$

and

$$\begin{array}{rcl} p_1 & \leq & \mathbb{P}\left(\|\frac{1}{N}\sum_{t=1}^{N} \bar{\zeta}_t\|_\infty \geq \tilde{\eta}_1 + \frac{\sigma^p}{A^{p-1}} \right) \\ & \leq & \mathbb{P}\left(\|\sum_{t=1}^{N} \bar{\zeta}_t - \mathbb{E}[\bar{\zeta}_t]\|_\infty \geq N\tilde{\eta}_1 \right). \end{array} \tag{40}$$

Next note that the random variables $\tilde{\xi}_t(i) = \bar{\zeta}_t(i) - \mathbb{E}[\bar{\zeta}_t(i)]$ satisfy

$$\mathbb{E}[\tilde{\xi}_t(i)] = 0, \;\; \mathbb{E}[\tilde{\xi}_t(i)^2] \leq \sigma^2 \;\text{ and }\; |\tilde{\xi}_t(i)| \leq 2\,A.$$

Since for every $i = 1, \ldots, n$, the variables $(\tilde{\xi}_t(i))_{1 \leq t \leq N}$ are independent, using Bernstein inequality we obtain

$$\mathbb{P}\left(\|\sum_{t=1}^{N} \tilde{\xi}_t\|_\infty \geq N\tilde{\eta}_1 \right) \leq 2\,n\, \exp\left(-\frac{1}{2}\frac{N\tilde{\eta}_1^2}{\sigma^2 + \frac{2}{3}A\tilde{\eta}_1} \right) = \frac{2}{N^{\lambda}}. \tag{41}$$

It then suffices to plug (41) into (40) and (40) and (39) into (38) to achieve the proof of (35). We now prove (36). We have

$$\|\hat{Q}_N^b - Q\|_\infty = \|\text{svec}(\hat{Q}_N^b) - \text{svec}(Q)\|_\infty = \|\frac{1}{N}\sum_{t=1}^{N} \xi_t\|_\infty$$

where $\xi_t \in \mathbb{R}^{\frac{n(n+1)}{2}}$ is defined by

$$\xi_t(j + \frac{k(2n-k+1)}{2}) = \zeta_t(k+1)\, \zeta_t(j+k) - Q(k+1, j+k), \;\; 0 \leq k \leq n-1, \;\; 1 \leq j \leq n-k. \tag{42}$$

For every $i = 1, \ldots, \frac{n(n+1)}{2}$, the variables $(\xi_t(i))_{1 \leq t \leq N}$ are independent and for $k = 1, \ldots, E[\frac{p}{2}]$

$$\mathbb{E}[\|\xi_t\|_\infty^k] \leq \mathbb{E}[\left(\|\zeta_t\|_\infty^2 + \sigma^2\right)^k] \leq \sum_{i=0}^{k} \;\; C_k^i\, \sigma^{2i}\sigma^{2(k-i)} = \left(2\sigma^2\right)^k.$$

It then suffices to follow the proof of (35) to prove (36). Finally, to show (37), we note that

$$\mathbb{P}(\|\hat{Q}_N - Q\|_\infty \geq \eta_2) \leq \mathbb{P}(\|\hat{\rho}_N - \rho\|_\infty \geq \eta_1) + \mathbb{P}(\|\hat{Q}_N^b - Q\|_\infty \geq \eta_2'). \tag{43}$$

□

Remark 6.2. *The condition $3N \geq 10(\ln n(n+1) + \lambda \ln N)$ in the above lemma can be suppressed but it would lead to more complicated right-hand sides.*

To proceed further, we need to recall the following lemma from Guigues (2008):

Lemma 6.3. *Guigues (2008) Let $X_t, t = 1, \ldots, N$, be N independent observations of a zero mean random vector in $\mathbb{R}^n$ with $n \geq 2$. If in addition we have $\mathbb{E}[\|X_t\|_\infty^2] \leq \sigma^2$ for every t, then*

$$\mathbb{E}[\|\frac{1}{N}\sum_{t=1}^{N} X_t\|_\infty^2] \leq \frac{8\sigma^2}{\ln 2}\frac{\ln n}{N}. \tag{44}$$

Proof of Theorems 3.1 and 3.2. We first prove Theorem 3.1. Using Lemmas 6.1 and 6.3, we have, for every nonempty interval I, and for every $\lambda > 0$ such that $10(\ln n + \lambda \ln |I|) \leq 3|I|$ and $p > 2(1+\lambda)$:

$$\mathbb{P}\left(\|\hat{\rho}_I - \bar{\rho}_I\|_\infty \geq f_1(|I|, \lambda) := 2\sqrt{\frac{10}{3}}\sigma\sqrt{\frac{\ln n + \lambda \ln |I|}{|I|}}\right) \leq \frac{K_1(\lambda)}{|I|^\lambda}, \tag{45}$$

$$\mathbb{E}[\|\hat{\rho}_I - \mathbb{E}[\hat{\rho}_I]\|_\infty^2] \leq \frac{8\sigma^2}{\ln 2}\left(\frac{\ln n}{|I|}\right), \tag{46}$$

where $\bar{\rho}_I = \frac{1}{|I|}\sum_{k \in I} \rho_k$. Moreover, we have

$$\|\hat{Q}_I - \bar{Q}_I\|_\infty \leq \|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2 + \|\tilde{Q}_I^b - \bar{Q}_I\|_\infty \tag{47}$$

with $\tilde{Q}_I^b = \frac{1}{|I|}\sum_{t \in I}(r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^\top$ and $\bar{Q}_I = \frac{1}{|I|}\sum_{k \in I} Q_k$. We can write $\|\tilde{Q}_I^b - \bar{Q}_I\|_\infty$ as $\|\frac{1}{|I|}\sum_{t \in I}\xi_t\|_\infty$ with $\xi_t = (r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^\top - Q_t$. Observing that for $k = 1, \ldots, E[\frac{p}{2}]$, $\mathbb{E}[\|\xi_t\|_\infty^k] \leq \sigma''^k$ with $\sigma'' = \sigma^2 + (\sigma + 2\sigma')^2$, we can, setting $A = \sigma''\sqrt{\frac{3|I|}{10(\ln n(n+1) + \lambda \ln |I|)}}$, follow the proof of (35) to show that if $10(\ln n(n+1) + \lambda \ln |I|) \leq 3|I|$ and $\mathrm{E}[\frac{p}{2}] > 2(1+\lambda)$

$$\mathbb{P}\left(\|\tilde{Q}_I^b - \bar{Q}_I\|_\infty \geq 2\sqrt{\frac{10}{3}}\sigma''\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}\right) \leq \frac{K_2(\lambda)}{|I|^\lambda} \tag{48}$$

with $K_2(\lambda) = 1 + \left(\frac{10\, E[\frac{p}{2}](2\gamma+\lambda)}{3\, e(E[\frac{p}{2}] - 2 - 2\lambda)}\right)^{\frac{E[\frac{p}{2}]}{2}}$. Using (45), (47) and (48), it follows that if $10(\ln n(n+1) + \lambda \ln |I|) \leq 3|I|$ and $\mathrm{E}[\frac{p}{2}] > 2(1+\lambda)$ we have

$$\mathbb{P}\left(\|\hat{Q}_I - \bar{Q}_I\|_\infty \geq f_2(|I|, \lambda) := 2\sqrt{\frac{10}{3}}(2\sigma^2 + \sigma'')\sqrt{\frac{\ln n(n+1) + \lambda \ln |I|}{|I|}}\right) \leq \frac{K_1(\lambda) + K_2(\lambda)}{|I|^\lambda}. \tag{49}$$

Finally

$$\begin{array}{rcl}\mathbb{E}[\|\hat{Q}_I - \mathbb{E}[\hat{Q}_I]\|_\infty] & \leq & 2\mathbb{E}[\|\hat{Q}_I - \tilde{Q}_I^b\|_\infty] + \mathbb{E}[\|\tilde{Q}_I^b - \mathbb{E}[\tilde{Q}_I^b]\|_\infty], \\ & \leq & 2\mathbb{E}[\|\hat{\rho}_I - \bar{\rho}_I\|_\infty^2] + \mathbb{E}[\|\frac{1}{|I|}\sum_{t \in I}\tilde{\xi}_t\|_\infty],\end{array} \tag{50}$$

setting $\tilde{\xi}_t = (r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^\top - \mathbb{E}[(r_t - \bar{\rho}_I)(r_t - \bar{\rho}_I)^\top]$. Since $\mathbb{E}[\|\tilde{\xi}_t\|_\infty^2] \leq \sigma'''^2$ with $\sigma''' = 2(\sigma + 2\sigma')^2$, following the proof of Lemma 6.3 (in Guigues (2008)), we show that

$$\mathbb{E}[\|\tilde{Q}_I^b - \mathbb{E}[\tilde{Q}_I^b]\|_\infty] \leq \frac{4\sigma'''}{\sqrt{\ln 2}}\sqrt{\frac{\ln n}{|I|}}. \tag{51}$$

Using (46), (50), and (51) we get

$$\mathbb{E}[\|\hat{Q}_I - \mathbb{E}[\hat{Q}_I]\|_\infty] \leq (\frac{16\sigma^2}{\ln 2} + \frac{4\sigma'''}{\sqrt{\ln 2}})\sqrt{\frac{\ln n}{|I|}}. \tag{52}$$

We now show (5). First note that

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \le \|\hat{\rho}_I - \bar{\rho}_I\|_\infty + \|\bar{\rho}_I - \rho_{N+1}\|_\infty + \|\bar{\rho}_J - \rho_{N+1}\|_\infty + \|\hat{\rho}_J - \bar{\rho}_J\|_\infty. \tag{53}$$

Also $\|\bar{\rho}_I - \rho_{N+1}\|_\infty \le \Delta_I^\rho$, $\|\bar{\rho}_J - \rho_{N+1}\|_\infty \le \Delta_J^\rho$ and since I is an ILTH and $J \in \mathcal{I}(I)$, using (46), we have

$$\|\bar{\rho}_I - \rho_{N+1}\|_\infty \le 2\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|I|}}, \quad \|\bar{\rho}_J - \rho_{N+1}\|_\infty \le 2\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|J|}}. \tag{54}$$

We end the proof of (5) plugging (54) into (53) and using (45). Inequality (6) can be shown in the same way using (49) and (52). This achieves the proof of Theorem 3.1.

Theorem 3.2 is then shown following the proof of Theorem 3.6 in Guigues (2008). Let us show for instance (9)((10) can be shown in the same way). We show that the event

$$\|\hat{\rho}_{\hat{I}_\rho} - \rho_{N+1}\|_\infty > 3f_1(|\mathbb{I}_\rho|, \lambda) + 4\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}_\rho|}} + \Delta^\rho_{\mathbb{I}_\rho}$$

implies the event

$$\bigcup_{I \in \mathcal{I} | I \subseteq \mathbb{I}_\rho} \bigcup_{J \in \mathcal{I}_+(I)} \left\{ \|\hat{\rho}_J - \bar{\rho}_J\|_\infty > f_1(|J|, \lambda) \right\},$$

which will prove (9). Let us suppose that for all I in $\mathcal{I}$ such that $I \subseteq \mathbb{I}_\rho$ and for $J \in \mathcal{I}_+(I)$, $\|\hat{\rho}_J - \bar{\rho}_J\|_\infty \le f_1(|J|, \lambda)$. We intend to prove that $\|\hat{\rho}_{\hat{I}_\rho} - \rho_{N+1}\|_\infty \le 3f_1(|\mathbb{I}_\rho|, \lambda) + 4\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}_\rho|}} + \Delta^\rho_{\mathbb{I}_\rho}$. First, note that $\mathbb{I}_\rho$ is not rejected. Indeed, for all $I \in \mathcal{I}$ such that $I \subseteq \mathbb{I}_\rho$ and for all $J \in \mathcal{I}(I)$:

$$\begin{aligned}
\|\hat{\rho}_I - \hat{\rho}_J\|_\infty &\le \|\hat{\rho}_I - \bar{\rho}_I\|_\infty + \|\bar{\rho}_I - \rho_{N+1}\|_\infty + \|\rho_{N+1} - \bar{\rho}_J\|_\infty + \|\bar{\rho}_J - \hat{\rho}_J\|_\infty \\
&\le f_1(|I|, \lambda) + f_1(|J|, \lambda) + \Delta_I^\rho + \Delta_J^\rho.
\end{aligned}$$

Now due to the definition of $\mathbb{I}_\rho$, $\Delta_I^\rho \le DV_I^\rho$, $\Delta_J^\rho \le DV_J^\rho$ and using (46) gives

$$\|\hat{\rho}_I - \hat{\rho}_J\|_\infty \le f_1(|I|, \lambda) + f_1(|J|, \lambda) + 2\sqrt{\frac{2}{\ln 2}} D\sigma \left(\sqrt{\frac{\ln n}{|I|}} + \sqrt{\frac{\ln n}{|J|}} \right).$$

Remember that I is accepted as an ILTH if (7) holds. It follows that for all $I \in \mathcal{I}$ such that $I \subseteq \mathbb{I}_\rho$, I is accepted and $\mathbb{I}_\rho$ is accepted so $\mathbb{I}_\rho \subseteq \hat{I}_\rho$. This implies

$$\begin{aligned}
\|\hat{\rho}_{\mathbb{I}_\rho} - \hat{\rho}_{\hat{I}_\rho}\|_\infty &\le f_1(|\mathbb{I}_\rho|, \lambda) + f_1(|\hat{I}_\rho|, \lambda) + 2\sqrt{\frac{2}{\ln 2}} D\sigma \left(\sqrt{\frac{\ln n}{|\mathbb{I}_\rho|}} + \sqrt{\frac{\ln n}{|\hat{I}_\rho|}} \right) \\
&\le 2f_1(|\mathbb{I}_\rho|, \lambda) + 4\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}_\rho|}},
\end{aligned}$$

since $f_1(|I|, \lambda)$ is a decreasing function of $|I|$. Next

$$\begin{aligned}
\|\hat{\rho}_{\hat{I}_\rho} - \rho_{N+1}\|_\infty &\le \|\hat{\rho}_{\hat{I}_\rho} - \hat{\rho}_{\mathbb{I}_\rho}\|_\infty + \|\hat{\rho}_{\mathbb{I}_\rho} - \bar{\rho}_{\mathbb{I}_\rho}\|_\infty + \|\bar{\rho}_{\mathbb{I}_\rho} - \rho_{N+1}\|_\infty \\
&\le 3f_1(|\mathbb{I}_\rho|, \lambda) + 4\sqrt{\frac{2}{\ln 2}} D\sigma \sqrt{\frac{\ln n}{|\mathbb{I}_\rho|}} + \Delta^\rho_{\mathbb{I}_\rho}.
\end{aligned}$$

□

Proof of Theorem 3.3. Using the notation introduced in Section 3, we want to bound from above

$$p = \mathbb{P}\left(\exists\, J \in \mathcal{I}(I) \mid \{\|\hat{\rho}_J - \hat{\rho}_I\|_\infty > \gamma_\rho(|I|, |J|, \lambda)\} \bigcup \left\{ \|\hat{Q}_J - \hat{Q}_I\|_\infty > \gamma_Q(|I|, |J|, \mu) \right\} \right).$$

We have

$$\begin{aligned} p &\leq \sum_{J\in\mathcal{I}(I)} \mathbb{P}(\|\hat{\rho}_J-\hat{\rho}_I\|_\infty > \gamma_\rho(|I|,|J|,\lambda)) + \mathbb{P}(\|\hat{Q}_J-\hat{Q}_I\|_\infty > \gamma_Q(|I|,|J|,\mu)),\\ &\leq \sum_{J\in\mathcal{I}(I)} K_1(\lambda)\left(\frac{1}{|I|^\lambda}+\frac{1}{|J|^\lambda}\right) + (K_1(\mu)+K_2(\mu))\left(\frac{1}{|I|^\mu}+\frac{1}{|J|^\mu}\right), \text{(from Theorem 3.1)},\\ &\leq \frac{2\ \mathrm{Card}(\mathcal{I}(I))}{m_0^{\min(\lambda,\mu)}}\left(K_1(\lambda)+K_1(\mu)+K_2(\mu)\right). \end{aligned}$$

□

Proof of Theorem 3.4. If interval I is accepted as an ILTH, we can consider the means, variances and covariances are close on $J=\{T_{bp}-m,\dots,T_{bp}-1\}$ and I. Thus, the probability p that I is accepted is bounded from above by $\mathbb{P}\left(\|\hat{\rho}_J-\hat{\rho}_I\|_\infty \leq \gamma_\rho(|I|,|J|,\lambda)\right)$. Besides,

$$\|\hat{\rho}_J-\hat{\rho}_I\|_\infty \geq \|\bar{\rho}_I-\bar{\rho}_J\|_\infty - \|\hat{\rho}_J-\bar{\rho}_J+\bar{\rho}_I-\hat{\rho}_I\|_\infty. \tag{55}$$

Since $\|\bar{\rho}_I-\bar{\rho}_J\|_\infty = \frac{m'}{m+m'}\|\rho_2-\rho_1\|_\infty$, using (55) and (12) yields

$$\begin{aligned} p &\leq \mathbb{P}(\|\hat{\rho}_J-\bar{\rho}_J+\bar{\rho}_I-\hat{\rho}_I\|_\infty \geq f_1(|J|,\lambda)+f_1(|I|,\lambda)),\\ &\leq \mathbb{P}(\|\hat{\rho}_J-\bar{\rho}_J\|_\infty \geq f_1(|J|,\lambda)) + \mathbb{P}(\|\hat{\rho}_I-\bar{\rho}_I\|_\infty \geq f_1(|I|,\lambda))\\ &\leq K_1(\lambda)\left(\frac{1}{m^\lambda}+\frac{1}{(m+m')^\lambda}\right) \quad \text{(using (45))}. \end{aligned}$$

□

Proof of Theorems 3.7 and 3.8. In Guigues (2008), expressions of $\gamma_\rho(|I|,|J|,\lambda)$ and $\gamma_Q(|I|,|J|,\lambda)$ are given such that if we use the modified version of the adaptive algorithm introduced in Section 3.3, an interval $I\in\mathcal{I}$ is accepted as an ILTH if for every $J\in\mathcal{I}(I)$: $\|\hat{\rho}_I-\hat{\rho}_J\|_\infty \leq \gamma_\rho(|I|,|J|,\lambda)$ and $\|\hat{Q}_I-\hat{Q}_J\|_\infty \leq \gamma_Q(|I|,|J|,\lambda)$, where $\lambda>0$ is now the only parameter of the adaptive algorithm. It is also shown in Guigues (2008) that if I is an ILTH and $J\in\mathcal{I}(I)$

$$\begin{aligned} &\mathbb{P}\left(\|\hat{\rho}_I-\hat{\rho}_J\|_\infty > \gamma_\rho(|I|,|J|,\lambda)\right) \leq \tfrac{1}{|I|^\lambda}+\tfrac{1}{|J|^\lambda},\\ &\mathbb{P}\left(\|\hat{Q}_I-\hat{Q}_J\|_\infty > \gamma_Q(|I|,|J|,\lambda)\right) \leq 2(\tfrac{1}{|I|^\lambda}+\tfrac{1}{|J|^\lambda}). \end{aligned}$$

Theorem 3.7 is then shown following the proof of Theorem 3.3. Further, if $\ln n(n+1)+\lambda\ln m_0 \leq m_0$, then we can implement the adaptive algorithm choosing

$$\begin{aligned} \gamma_\rho(|I|,|J|,\lambda) &= 4\sqrt{\frac{2}{\ln 2}}D\sigma\left(\sqrt{\frac{\ln n}{|I|}}+\sqrt{\frac{\ln n}{|J|}}\right) + (\frac{7}{3}+\sqrt{2})\sigma\left(f(|I|,\lambda)+f(|J|,\lambda)\right),\\ \gamma_Q(|I|,|J|,\lambda) &= (k_Q D + k_Q')\sigma^2(f(|I|,\lambda)+f(|J|,\lambda)), \end{aligned}$$

where $f(|I|,\lambda)=\sqrt{\frac{\ln n(n+1)+\lambda\ln|I|}{|I|}}$, $k_Q = 2+\frac{64}{\ln 2}+\frac{16}{\sqrt{\ln 2}}(2+\sqrt{2})$, and $k_Q' = \frac{160}{9}+\frac{26}{3}\sqrt{2}+8\sqrt{\frac{2}{\ln 2}}D$ (see Guigues (2008)). In this case (if $\ln n(n+1)+\lambda\ln m_0 \leq m_0$) then for every nonempty interval I, we have

$$\mathbb{P}\left(\|\hat{\rho}_I-\bar{\rho}_I\|_\infty > f_1'(|I|,\lambda)\right) \leq \frac{1}{|I|^\lambda}, \tag{56}$$

$$\mathbb{P}\left(\|\hat{Q}_I-\bar{Q}_I\|_\infty > f_2'(|I|,\lambda)\right) \leq \frac{2}{|I|^\lambda}, \tag{57}$$

with $f_1'(|I|,\lambda) = (\frac{7}{3}+\sqrt{2})\sigma f(|I|,\lambda)$ and $f_2'(|I|,\lambda) = k_Q'\sigma^2 f(|I|,\lambda)$. Now let us show Theorem 3.8. The probability p that I is accepted is bounded from above by

$$\mathbb{P}\left(\|\hat\rho_J - \hat\rho_I\|_\infty \le \gamma_\rho(|I|,|J|,\lambda) \bigcap \|\hat Q_J - \hat Q_I\|_\infty \le \gamma_Q(|I|,|J|,\lambda)\right)$$

which is bounded above by the sum of the probabilities. Besides,

$$\begin{aligned} \|\hat\rho_J - \hat\rho_I\|_\infty &\ge \|\bar\rho_I - \bar\rho_J\|_\infty - \|\hat\rho_J - \bar\rho_J + \bar\rho_I - \hat\rho_I\|_\infty, \\ \|\hat Q_J - \hat Q_I\|_\infty &\ge \|\bar Q_I - \bar Q_J\|_\infty - \|\hat Q_J - \bar Q_J + \bar Q_I - \hat Q_I\|_\infty. \end{aligned} \tag{58}$$

Next, we have $\|\bar\rho_I - \bar\rho_J\|_\infty = \frac{m'}{m+m'}\|\rho_2-\rho_1\|_\infty$ and $\|\bar Q_I - \bar Q_J\|_\infty = \frac{m'}{m+m'}\|Q_2 - Q_1\|_\infty$. Finally, using (58) and (12) yields

$$\begin{aligned} p &\le \mathbb{P}(\|\hat\rho_J - \bar\rho_J + \bar\rho_I - \hat\rho_I\|_\infty \ge f_1'(|J|,\lambda) + f_1'(|I|,\lambda)) \\ &\quad +\mathbb{P}(\|\hat Q_J - \bar Q_J + \bar Q_I - \hat Q_I\|_\infty \ge f_2'(|J|,\lambda) + f_2'(|I|,\lambda)), \\ &\le \mathbb{P}(\|\hat\rho_J - \bar\rho_J\|_\infty \ge f_1'(|J|,\lambda)) + \mathbb{P}(\|\hat\rho_I - \bar\rho_I\|_\infty \ge f_1'(|I|,\lambda)) \\ &\quad + \mathbb{P}(\|\hat Q_J - \bar Q_J\|_\infty \ge f_2'(|J|,\lambda)) + \mathbb{P}(\|\hat Q_I - \bar Q_I\|_\infty \ge f_2'(|I|,\lambda)), \\ &\le 3\left(\frac{1}{m^\lambda} + \frac{1}{(m+m')^\lambda}\right). \end{aligned}$$

□

Proof of Theorem 4.1. Let $\bar\rho_\ell^t(h) = \frac{1}{h}\sum_{k=t-h+1}^{t}\rho_k$ and $\bar\rho_r^t(h) = \frac{1}{h}\sum_{k=t}^{t+h-1}\rho_k$. We consider four cases.

- If $h \le t \le T_{bp} - h$, $\bar\rho_\ell^t(h) = \bar\rho_r^t(h) = m_1$ and

$$\begin{aligned} \mathbb{P}(\|\rho_\ell^t(h) - \rho_r^t(h)\|_\infty > 2f_1(h,\lambda)) &\le \mathbb{P}(\|\rho_\ell^t(h) - \bar\rho_\ell^t(h)\|_\infty + \|\rho_r^t(h) - \bar\rho_r^t(h)\|_\infty > 2f_1(h,\lambda)) \\ &\le \mathbb{P}(\|\rho_\ell^t(h) - \bar\rho_\ell^t(h)\|_\infty > f_1(h,\lambda)) \\ &\quad +\mathbb{P}(\|\rho_r^t(h) - \bar\rho_r^t(h)\|_\infty > f_1(h,\lambda)) \\ &\le \frac{2K_1(\lambda)}{h^\lambda}, \quad \text{using (45).} \end{aligned}$$

  Thus, there is a set $S_t \subset \Omega$ of probability at least $1 - \frac{2K_1(\lambda)}{h^\lambda}$ such that for $\omega \in S_t$, $\|\rho_\ell^t(h) - \rho_r^t(h)\|_\infty \le 2f_1(h,\lambda)$.
- Similarly, for $T_{bp} + h - 1 \le t \le N - h + 1$ there is a set $S_t \subset \Omega$ of probability at least $1 - \frac{2K_1(\lambda)}{h^\lambda}$ such that for $\omega \in S_t$, $\|\rho_\ell^t(h) - \rho_r^t(h)\|_\infty \le 2f_1(h,\lambda)$.
- For $T_{bp} \le t \le T_{bp} + h - 2$, we have

$$\bar\rho_r^t(h) = m_2 \text{ and } \bar\rho_\ell^t(h) = \frac{(t - T_{bp} + 1)m_2 + (h - t + T_{bp} - 1)m_1}{h}.$$

  Note that

$$\|\rho_\ell^t(h) - \rho_r^t(h)\|_\infty \ge -\|\rho_\ell^t(h) - \bar\rho_\ell^t(h) - \rho_r^t(h) + \bar\rho_r^t(h)\|_\infty + \left(\frac{h - t + T_{bp} - 1}{h}\right)\|m_2 - m_1\|_\infty,$$

  and that $\mathbb{P}(\|\rho_\ell^t(h) - \bar\rho_\ell^t(h) - \rho_r^t(h) + \bar\rho_r^t(h)\|_\infty > 2f_1(h,\lambda)) \le \frac{2K_1(\lambda)}{h^\lambda}$. Thus, there is a set $S_t \subset \Omega$ of probability at least $1 - \frac{2K_1(\lambda)}{h^\lambda}$ such that for $\omega \in S_t$,

$$\|\rho_\ell^t(h) - \rho_r^t(h)\|_\infty \ge -2f_1(h,\lambda) + \left(\frac{h - t + T_{bp} - 1}{h}\right)\|m_2 - m_1\|_\infty.$$

- Similarly, for $T_{bp}-h+1 \leq t \leq T_{bp}-1$, we can find a set $S_t \subset \Omega$ of probability at least $1-\frac{2K_1(\lambda)}{h^\lambda}$ such that for $\omega \in S_t$,

$$\|\rho_\ell^t(h)-\rho_r^t(h)\|_\infty \geq -2f_1(h,\lambda)+\left(\frac{h+t-T_{bp}}{h}\right)\|m_2-m_1\|_\infty.$$

Thus, if $\omega \in S_{T_{bp}-1}$ and $\|m_2-m_1\|_\infty > \frac{4h}{h-1}f_1(h,\lambda)$ then $\|\rho_\ell^{T_{bp}-1}(h)-\rho_r^{T_{bp}-1}(h)\|_\infty > 2f_1(h,\lambda)$. Finally, if $\omega \in \left(\bigcap_{t=h}^{T_{bp}-h} S_t\right)\bigcap\left(\bigcap_{t=T_{bp}+h-1}^{N-h+1} S_t\right)\bigcap S_{T_{bp}-1}$ then $T_{bp}(h) \in \{T_{bp}-h+1,\ldots,T_{bp}+h-2\}$. We thus get

$$\begin{array}{lll}\mathbb{P}(T_{bp}(h) \in \{T_{bp}-h+1, T_{bp}+h-2\}) & \geq & \mathbb{P}\left(\omega \in \left(\bigcap_{t=h}^{T_{bp}-h} S_t\right)\bigcap\left(\bigcap_{t=T_{bp}+h-1}^{N-h+1} S_t\right)\bigcap S_{T_{bp}-1}\right)\\ & \geq & 1-\dfrac{2K_1(\lambda)(N-4h+5)}{h^\lambda}.\end{array}$$

□

Proof of Theorem 4.2. Following the proof of Theorem 4.1, we can find sets S_t such that $\mathbb{P}(\omega \in S_t) \geq 1-\frac{2K_1(\lambda)}{h^\lambda}$ and

- If $h \leq t \leq T_{bp}-h$ or $T_{bp}+h-1 \leq t \leq N-h+1$ and $\omega \in S_t$, then $\|\rho_\ell^t(h)-\rho_r^t(h)\|_\infty \leq 2f_1(h,\lambda)$.
- If $T_{bp}+1 \leq t \leq T_{bp}+h-2$ or $T_{bp}-h+1 \leq t \leq T_{bp}-2$ and $\omega \in S_t$, then $\|\rho_\ell^t(h)-\rho_r^t(h)\|_\infty \leq 2f_1(h,\lambda)+\frac{h-2}{h}\|m_2-m_1\|_\infty$.
- If ($t=T_{bp}-1$ or $t=T_{bp}$) and $\omega \in S_t$,

$$\|\rho_\ell^t(h)-\rho_r^t(h)\|_\infty \geq -2f_1(h,\lambda)+\frac{h-1}{h}\|m_2-m_1\|_\infty > 2f_1(h,\lambda)+\frac{h-2}{h}\|m_2-m_1\|_\infty.$$

Thus, if $\omega \in \bigcap_{t=h}^{N-h+1} S_t$ and $\|m_2-m_1\|_\infty > 4hf_1(h,\lambda)$ then $T_{bp}(h) \in \{T_{bp}-1\} \cup \{T_{bp}\}$. □

Proof of Theorems 4.3 and 4.4. If Assumption **(A2)** holds, it is shown in Guigues (2008) that for any nonempty interval I then $\mathbb{P}(\|\hat{\rho}_I-\bar{\rho}_I\|_\infty > f_1'(|I|,\lambda)) \leq \frac{1}{|I|^\lambda}$, where $f_1'(|I|,\lambda)$ is given in (28). It then suffices to follow the proof of Theorem 4.1 to show Theorem 4.3 and to follow the proof of Theorem 4.2 to show Theorem 4.4. □

Proof of Theorems 4.5, 4.6, and 4.7. Theorems 4.5 and 4.6 are shown using (49) and following respectively the proofs of Theorems 4.1 and 4.2. Finally, we show Theorem 4.7 using (57) and following the proofs of Theorems 4.1 and 4.2. □

Vincent Guigues: UFRJ, Escola Politécnica, Departamento de Engenharia Industrial, Ilha do Fundão, CT, Bloco F, Rio de Janeiro, Brazil, Fundação Getulio Vargas, Escola de Matemática Aplicada, Praia de Botafogo, Rio de Janeiro, Brazil, Department of Statistics, LJK, BP 53, 38 041 Grenoble Cedex 9, France, vincent.guigues@imag.fr