Bernoulli **23**(2), 2017, 1408–1447
DOI: 10.3150/15-BEJ782

A robust approach for estimating change-points in the mean of an AR(1) process

S. CHAKAR[1,2,*], E. LEBARBIER[1,**], C. LÉVY-LEDUC[1,†] and S. ROBIN[1,‡]

[1]*AgroParisTech/INRA, UMR 518 MIA, F-75005 Paris, France.*
E-mail: [*]*souhil.chakar@imag.fr;* [**]*emilie.lebarbier@agroparistech.fr;*
[†]*celine.levy-leduc@agroparistech.fr;* [‡]*stephane.robin@agroparistech.fr*
[2]*INP Grenoble, Laboratoire Jean Kuntzmann, Grenoble cedex 09, France*

We consider the problem of multiple change-point estimation in the mean of an AR(1) process. Taking into account the dependence structure does not allow us to use the dynamic programming algorithm, which is the only algorithm giving the optimal solution in the independent case. We propose a robust estimator of the autocorrelation parameter, which is consistent and satisfies a central limit theorem in the Gaussian case. Then, we propose to follow the classical inference approach, by plugging this estimator in the criteria used for change-points estimation. We show that the asymptotic properties of these estimators are the same as those of the classical estimators in the independent framework. The same plug-in approach is then used to approximate the modified BIC and choose the number of segments. This method is implemented in the R package `AR1seg` and is available from the Comprehensive R Archive Network (CRAN). This package is used in the simulation section in which we show that for finite sample sizes taking into account the dependence structure improves the statistical performance of the change-point estimators and of the selection criterion.

Keywords: auto-regressive model; change-points; model selection; robust estimation of the AR(1) parameter; time series

1. Introduction

Change-point detection problems arise in many fields, such as genomics [8,9,31], medical imaging [20], earth sciences [15,36], econometrics [17,19] or climate [25,28]. In many of these problems, the observations cannot be assumed to be independent. Indeed the autocovariance structure of the time series display more complex patterns and might be taken into account in change-point estimation.

An abundant literature exists about the statistical theory of change-point detection. Only speaking about Gaussian processes, various frameworks have been considered ranging from the independent case with changes in the mean [6], to more complex structural changes [3], dependent processes [19] or processes with changes in all parameters [5].

[19,21] proved that, if the number of changes is known, the least-squares estimators of the change-point locations and of the parameters of each segment are consistent under very mild conditions on the autocovariance structure of the process with changes in the mean. A quasi-likelihood approach is also proved to provide consistent estimates for the model with changes in

all parameters [5]. Many model selection criteria have also been proposed to estimate the number of changes, mostly in the independent case (see, for example, [20,22,37,39]).

Change-point detection also raises algorithmic issues as the determination of the optimal set of change-point locations is a discrete optimization problem. Several approximate solutions to this problem exist [12,16,29,35]. However, the dynamic programming algorithm introduced by [2] and its refinements [27,32] are the only efficient and exact algorithms to recover this optimal segmentation. The computational complexity of this algorithms is quadratic relatively to the length of the series.

However, the dynamic programming algorithm only applies when (i) the loss function (e.g., the negative log-likelihood) is additive with respect to the segments and when (ii) no parameter to be estimated is common to several segments. These requirements are met by the least-square criterion (which corresponds to the negative log-likelihood in the Gaussian homoscedastic independent model with changes in the mean) or by the model and criterion considered by [5]. In other cases, iterative and stochastic procedures are needed [4,24].

In this paper, we consider the segmentation of an AR(1) process with homogeneous autocorrelation coefficient $\rho^\star$:

$$y_i = \mu_k^\star + \eta_i, \qquad t_{n,k}^\star + 1 \le i \le t_{n,k+1}^\star, 0 \le k \le m^\star, 1 \le i \le n, \tag{1}$$

where $(\eta_i)_{i\in\mathbb{Z}}$ is a zero-mean second-order stationary AR(1) process defined as the solution of

$$\eta_i = \rho^\star \eta_{i-1} + \epsilon_i, \tag{2}$$

where $|\rho^\star| < 1$ and $(\epsilon_i)_{i\in\mathbb{Z}}$ is a white noise with variance $\sigma^{\star 2}$. We further also assume that y_0 is a r.v. with mean $\mu_0^\star$ and variance $\sigma^{\star 2}/(1-\rho^{\star 2})$. Actually, most of the results we provide in this paper hold without the Gaussian assumption, even if the Gaussian likelihood is considered.

Note that this model is different from the ones considered by [12] and [5]. Indeed, [12] considered the segmentation issue of a non-stationary time series which consists of blocks of different autoregressive processes where all the parameters of the autoregressive processes change from one segment to the other. [5] proposed a methodology for estimating the change-points of a non-stationary time series built from a general class of models having piecewise constant parameters. In this framework, all the parameters may change jointly at each change-point. This differs from our model (1) where the parameters $\rho^\star$ and $\sigma^\star$ are not assumed to change from one segment to the other. The direct maximum-likelihood inference for such a process violates both requirements (i) and (ii). Indeed the log-likelihood is not additive with respect to the segments because of the dependence that exists between data from neighbor segments and the unknown coefficient $\rho^\star$ needs to be estimated jointly over all segments.

Our aim is to propose a methodology for estimating both the change-point locations $\mathbf{t}_{n^\star} = (t_{n,k}^\star)_{1\le k\le m^\star}$ and the means $\boldsymbol{\mu}^\star = (\mu_k^\star)_{0\le k\le m^\star}$, accounting for the existence of the autocorrelation $\rho^\star$.

In the sequel, we shall use the following conventions: $t_{n,0}^\star = 0$, $t_{n,m^\star+1}^\star = n$ and assume that there exists $\boldsymbol{\tau}^\star = (\tau_k^\star)_{0\le k\le m+1}$ such that, for $0 \le k \le m+1$, $t_{n,k}^\star = \lfloor n\tau_k^\star \rfloor$, $\lfloor x \rfloor$ denoting the integer part of x. Consequently, $\tau_0^\star = 0$ and $\tau_{m^\star+1}^\star = 1$.

If $\rho^\star$ was known, the series could be decorrelated and the dynamic programming algorithm could then be used for the segmentation of this decorrelated series $(y_i - \rho^\star y_{i-1})_{i\ge 1}$. Here, $\rho^\star$ is unknown, but is estimated, and this estimator is then used to decorrelate the series.

To this aim, we borrow techniques from robust estimation [26]. Briefly speaking, we consider the data observed at the change-point locations as outliers and propose an estimate of $\rho^\star$ that is robust to the presence of such outliers. We shall prove that the estimate we propose is consistent and satisfies a central limit theorem.

We shall prove that the resulting change-point estimators satisfy the same asymptotic properties as those proposed by [5,21]. Finally, we propose a model selection criterion inspired by the one proposed by [39] and prove some asymptotic properties of this criterion.

This method is implemented in the R package `AR1seg` and is available from the Comprehensive R Archive Network (CRAN).

This paper is organized as follows. In Section 2, we propose a robust estimator for $\rho^\star$ and establish its asymptotic properties in the Gaussian case. In Section 3, we prove that the change-point estimators defined in (11) are consistent in both the Gaussian and the non-Gaussian case. In Section 4, we provide a consistent model selection criterion in the non-Gaussian case and derive an approximation of a Gaussian criterion. In Section 5, we illustrate by a simulation study the performance of this approach for time series having a finite sample size.

2. Robust estimation of the parameter $\rho^\star$

The aim of this section is to provide an estimator of $\rho^\star$ which can deal with the presence of change-points in the data. In the absence of change-points ($m^\star = 0$ in (1)), a consistent estimator of $\rho^\star$ could be obtained by using the classical autocorrelation function estimator of $(y_i)_{0\le i\le n}$ computed at lag 1. Since change-points can be seen as outliers in the AR(1) process, we shall propose a robust approach for estimating $\rho^\star$. [26] propose a robust estimator of the autocorrelation function of a stationary time series based on the robust scale estimator proposed by [33]. More precisely, the approach of [26] would result in the following estimate of $\rho^\star$:

$$\widehat{\rho}_{\mathrm{MG}} = \frac{Q_n^2(y^+) - Q_n^2(y^-)}{Q_n^2(y^+) + Q_n^2(y^-)},$$

where $y^+ = (y_{i+1} + y_i)_{0\le i\le n-1}$, $y^- = (y_{i+1} - y_i)_{0\le i\le n-1}$ and Q_n is the scale estimator of [33] which is such that $Q_n(x)$ is proportional to the first quartile of

$$\{|x_i - x_j|; 0 \le i < j \le n\}.$$

The asymptotic properties of this estimator are studied by [23] for Gaussian stationary processes with either short-range or long-range dependence. However, as we shall see in the simulation section we can provide an estimator of $\rho^\star$ which is more robust to the presence of change-points than $\widehat{\rho}_{\mathrm{MG}}$. The asymptotic properties of this novel robust estimator are given in Proposition 2.1.

Proposition 2.1. *Let $y_0, \dots, y_n$ be $(n+1)$ jointly Gaussian observations satisfying (1) and let*

$$\widetilde{\rho}_n = \frac{(\mathrm{med}_{0\le i\le n-2}|y_{i+2} - y_i|)^2}{(\mathrm{med}_{0\le i\le n-1}|y_{i+1} - y_i|)^2} - 1, \qquad (3)$$

where medx_i *denotes the median. Then,* $\widetilde{\rho}_n$ *satisfies the following central limit theorem*

$$\sqrt{n}\big(\widetilde{\rho}_n - \rho^\star\big) \xrightarrow{d} \mathcal{N}\big(0, \widetilde{\sigma}^2\big), \qquad \text{as } n \to \infty, \tag{4}$$

where

$$\widetilde{\sigma}^2 = \mathbb{E}\big[\Psi(\eta_0, \eta_1, \eta_2)^2\big] + 2\sum_{k\geq 1} \mathbb{E}\big[\Psi(\eta_0, \eta_1, \eta_2)\Psi(\eta_k, \eta_{k+1}, \eta_{k+2})\big], \tag{5}$$

and the function Ψ *is defined by*

$$\begin{aligned}&\Psi : (x_0, x_1, x_2)\\ &\mapsto -\frac{(1+\rho^\star)}{\Phi^{-1}(3/4)\varphi(\Phi^{-1}(3/4))}\big[\mathbb{1}_{\{|x_2-x_0|\leq\sqrt{2\sigma^{\star 2}}\Phi^{-1}(3/4)\}} - \mathbb{1}_{\{|x_1-x_0|\leq\sqrt{2\sigma^{\star 2}/(1+\rho^\star)}\Phi^{-1}(3/4)\}}\big],\end{aligned} \tag{6}$$

where Φ *and* φ *denote the cumulative distribution function and the probability distribution function of a standard Gaussian r.v., respectively.*

The proof of Proposition 2.1 is given in Appendix A.1.

Remark 2.1. Observe that the classical correlation could not be used here. Indeed, let $\widehat{\rho}_n(1)$ be the standard sample lag 1 autocorrelation of the series $y_0, \dots, y_n$ satisfying (1). One can easily show that $(\widehat{\rho}_n(1))$ tends in probability to $\frac{\gamma(1)+A}{\gamma(0)+A}$, where γ is the autocovariance function of $(\eta_i)_{i\in\mathbb{Z}}$ and $A = \sum_{k=0}^{m^\star}(\tau_{k+1}^\star - \tau_k^\star)\mu_k^{\star 2} - (\sum_{k=0}^{m^\star}(\tau_{k+1}^\star - \tau_k^\star)\mu_k^\star)^2$, as n tends to infinity. Our estimator avoids this asymptotic bias since it is robust to the presence of changes in the mean.

Remark 2.2. Note that the asymptotic distribution given in (4) allows to define a test of (H_0): '$\rho^\star = 0$' as the asymptotic variance $\widetilde{\sigma}^2$ does not depend on any unknown parameter under H_0.

Remark 2.3. Since the estimator (3) involves differences of the process (y_i) at different instants, it can only be used in the case of stable distributions as defined by [14]. Among them, we can quote the Cauchy, Lévy and Gaussian distributions, where the Gaussian distribution is the only one to have a finite second order moment. We give some hints in Appendix A.2 to explain why, in the case of the Cauchy distribution, taking $\widetilde{\widetilde{\rho}}_n$ defined as follows leads to an accurate estimator of $\rho^\star$:

$$\widetilde{\widetilde{\rho}}_n = \begin{cases} -1 + \sqrt{1 + \widetilde{\rho}_n}, & \text{if } \widetilde{\rho}_n \geq 0,\\ -\sqrt{1 - \sqrt{1 + \widetilde{\rho}_n}}, & \text{if } \widetilde{\rho}_n < 0,\end{cases} \tag{7}$$

where $\widetilde{\rho}_n$ is defined by (3). Some simulations are also provided in Section 5.4 to illustrate the finite sample size properties of this estimator.

3. Change-points and expectations estimation

In this section, the number of change-points $m^\star$ is assumed to be known. In the sequel, for notational simplicity, $m^\star$ will be denoted by m. Our goal is to estimate both the change-points and the means in model (1). A first idea consists in using the following criterion which is based on a Gaussian quasi-likelihood conditioned on y_0 and on the reparametrization $\delta_k = (1-\rho)\mu_k$, and to minimize it with respect to ρ:

$$\sum_{k=0}^{m}\sum_{i=t_k+2}^{t_{k+1}} (y_i - \rho y_{i-1} - \delta_k)^2 + \sum_{k=1}^{m}\left\{\left(y_{t_k+1} - \frac{\delta_k}{1-\rho}\right) - \rho\left(y_{t_k} - \frac{\delta_{k-1}}{1-\rho}\right)\right\}^2 + (y_1 - \rho y_0 - \delta_0)^2.$$

Due to the term that involves both δ_{k-1} and δ_k, this criterion cannot be efficiently minimized. Therefore, we propose to use an alternative criterion defined as follows:

$$SS_m(y, \rho, \boldsymbol{\delta}, \mathbf{t}) = \sum_{k=0}^{m}\sum_{i=t_k+1}^{t_{k+1}} (y_i - \rho y_{i-1} - \delta_k)^2. \tag{8}$$

Note that $SS_m(z, \rho, (1-\rho)\boldsymbol{\mu}, \mathbf{t})$ corresponds to $-n/2$ times the Gaussian log-likelihood of the following model maximized with respect to σ:

$$z_i - \mu_k^\star = \rho^\star\big(z_{i-1} - \mu_k^\star\big) + \epsilon_i, \qquad t_{n,k}^\star + 1 \le i \le t_{n,k+1}^\star, 0 \le k \le m, 1 \le i \le n, \tag{9}$$

and where z_0 is a r.v. with mean $\mu_0^\star$ and variance $\sigma^{\star 2}/(1-\rho^{\star 2})$ and $(\epsilon_i)_{i\in\mathbb{Z}}$ is a white noise with variance $\sigma^{\star 2}$. In this model, which is a subset of a model belonging to the class considered in [5], the expectation changes are not abrupt anymore as in model (1).

The parameter ρ, involved in each term of (8), is still a problem in order to minimize SS_m with respect to $\rho, \boldsymbol{\delta}$ and $\mathbf{t}$. This minimization problem is a complex discrete and global optimization problem. Dynamic Programming [2] cannot be used in this case. Only iterative methods are suitable to this minimization problem, without any guarantee to converge to the global minimum.

However, if ρ is replaced by an estimator $\overline{\rho}_n$, $SS_m(y, \overline{\rho}_n, \boldsymbol{\delta}, \mathbf{t})$ can be minimized with respect to $\boldsymbol{\delta}$ and $\mathbf{t}$ by Dynamic Programming. Proposition 3.2 gives asymptotic results for the estimators resulting from this method.

Proposition 3.1. *Let $z = (z_0, \ldots, z_n)$ be a finite sequence of real-valued r.v.'s satisfying* (9) *and $(\overline{\rho}_n)$ a sequence of real-valued r.v.'s. Let $\widehat{\boldsymbol{\delta}}_n(z, \overline{\rho}_n)$ and $\widehat{\mathbf{t}}_n(z, \overline{\rho}_n)$ be defined by*

$$\big(\widehat{\boldsymbol{\delta}}_n(z, \overline{\rho}_n), \widehat{\mathbf{t}}_n(z, \overline{\rho}_n)\big) = \mathop{\arg\min}_{(\boldsymbol{\delta}, \mathbf{t}) \in \mathbb{R}^{m+1} \times \mathcal{A}_{n,m}} SS_m(z, \overline{\rho}_n, \boldsymbol{\delta}, \mathbf{t}), \tag{10}$$

$$\widehat{\boldsymbol{\tau}}_n(z, \overline{\rho}_n) = \frac{1}{n}\widehat{\mathbf{t}}_n(z, \overline{\rho}_n), \tag{11}$$

where

$$\mathcal{A}_{n,m} = \big\{(t_0, \ldots, t_{m+1}); t_0 = 0 < \cdots < t_{m+1} = n, \forall k = 1, \ldots, m+1, t_k - t_{k-1} \ge \Delta_n\big\} \tag{12}$$

and where (Δ_n) *is a real sequence such that* $n^{-1}\Delta_n \longrightarrow_{n\to\infty} 0$ *and* $n^{-\alpha}\Delta_n \longrightarrow_{n\to\infty} +\infty$ *with* $\alpha > 0$. *Assume that*

$$(\overline{\rho}_n - \rho^\star) = O_P(n^{-1/2}), \tag{13}$$

as n tends to infinity. Then,

$$\|\widehat{\boldsymbol{\tau}}_n(z, \overline{\rho}_n) - \boldsymbol{\tau}^\star\| = O_P(n^{-1}), \qquad \|\widehat{\boldsymbol{\delta}}_n(z, \overline{\rho}_n) - \boldsymbol{\delta}^\star\| = O_P(n^{-1/2}),$$

where $\|\cdot\|$ *is the Euclidean norm and*

$$\boldsymbol{\delta}^\star = (\delta_0^\star, \dots, \delta_{m+1}^\star) = (1-\rho^\star)(\mu_0^\star, \dots, \mu_{m+1}^\star).$$

Proposition 3.2. *The results of Proposition 3.1 still hold under the same assumptions when z is replaced with y satisfying (1).*

Remark 3.1. Observe that, similarly to most of the papers dedicated to multiple change-points estimation [3,21,38], we assume that the number of change-points m is fixed and does not depend on the number of observations.

The proofs of Propositions 3.1 and 3.2 are given in Appendix A.3 and A.4, respectively. Note that the estimators defined in these propositions have the same asymptotic properties as those of the estimators proposed by [21]. In the Gaussian framework, the estimator $\widetilde{\rho}_n$ defined in Section 2 satisfies the same properties as $\overline{\rho}_n$ and can thus be used in the criterion SS_m for providing consistent estimators of the change-points and of the means.

4. Selecting the number of change-points

We now consider the selection of the number of change-points. We first propose a penalized contrast criterion, which we prove to be consistent in the non-Gaussian case. The penalty has a general form, which needs to be specified for a practical use. Therefore, we also derive an adaptation of the modified BIC criterion proposed by [39] in the Gaussian context. This criterion does not rely on any tuning parameter and has been shown to be efficient in practical cases [30].

4.1. Consistent model selection criterion

We propose to select the number of change-points m as follows

$$\widehat{m} = \mathop{\arg\min}_{0\le m\le m_{\max}} \frac{1}{n} SS_m(z, \overline{\rho}_n) + \beta_n m, \tag{14}$$

where $m_{\max} \ge m^\star$, $(\beta_n)_{n\ge 1}$ is a sequence of positive real numbers, $\overline{\rho}_n$ satisfies the assumptions of Proposition 3.1 and

$$SS_m(z, \rho) = \min_{\boldsymbol{\delta}, \mathbf{t} \in \mathcal{A}_{n,m}} SS_m(z, \rho, \boldsymbol{\delta}, \mathbf{t}), \tag{15}$$

$\mathcal{A}_{n,m}$ being defined in (12). In practice, $m_{\max}$ is an upper bound provided by the user.

Proposition 4.1. *Under the assumptions of Proposition 3.1, and if*

$$\beta_n \underset{n\to\infty}{\longrightarrow} 0, \qquad n^{1/2}\beta_n \underset{n\to\infty}{\longrightarrow} +\infty, \qquad \Delta_n\beta_n \underset{n\to\infty}{\longrightarrow} +\infty,$$

where Δ_n is defined in Proposition 3.1, $\widehat{m}$ defined by (14) converges in probability to $m^\star$.

Proposition 4.2. *The result of Proposition 4.1 still holds under the same assumptions when z is replaced by y satisfying (1).*

The proofs of Propositions 4.1 and 4.2 are given in Appendix A.5 and A.6, respectively.

Remark 4.1. If $\beta_n = n^{-\beta}$, the assumptions of Propositions 4.1 and 4.2 are fulfilled if and only if $0 < \beta < \min(\alpha, 1/2)$, where α is defined in Proposition 3.1. α stands for the usual bound for the control of the minimal segment length [21]. The $1/2$ bound is the price to pay for the estimation of $\rho^\star$.

4.2. Modified BIC criterion

Reference [39] proposed a modified Bayesian information criterion (mBIC) to select the number m of change-points in the particular case of segmentation of an independent Gaussian process x. This criterion is defined in a Bayesian context in which a non informative prior is set for the number of segments m. mBIC is derived from an $O_P(1)$ approximation of the Bayes factor between models with m and 0 change-points, respectively. The mBIC selection procedure consists in choosing the number of change-points as:

$$\widehat{m} = \arg\max_m C_m(x, 0), \tag{16}$$

where the criterion $C_m(y, \rho)$ is defined for a process y as

$$C_m(y,\rho) = -\frac{n-m+1}{2}\log SS_m(y,\rho) + \log\Gamma\left(\frac{n-m+1}{2}\right) - \frac{1}{2}\sum_{k=0}^{m}\log n_k\big(\widehat{\mathbf{t}}(y,\rho)\big) - m\log n,$$

where Γ is the usual Gamma function. In the latter equation

$$n_k\big(\widehat{\mathbf{t}}(y,\rho)\big) = \widehat{t}_{k+1}(y,\rho) - \widehat{t}_k(y,\rho), \tag{17}$$

where $\widehat{\mathbf{t}}(y,\rho) = (\widehat{t}_1(y,\rho), \dots, \widehat{t}_m(y,\rho))$ is defined as $\widehat{\mathbf{t}}(y,\rho) = \arg\min_{\mathbf{t}\in\mathcal{A}_{n,m}} \min_\delta SS_m(y,\rho,\delta,\mathbf{t})$.

Note that, in model (9), the criterion could be directly applied to the decorrelated series $v^\star = (v_i^\star)_{1\le i\le n} = (y_i - \rho^\star y_{i-1})_{1\le i\le n}$ since

$$C_m\big(y,\rho^\star\big) = C_m\big(v^\star, 0\big).$$

We propose to use the same selection criterion, replacing $\rho^\star$ by some relevant estimator $\overline{\rho}_n$. The following two propositions show that this plug-in approach result in the same asymptotic properties under both model (9) and (1).

Proposition 4.3. *For any positive m, for a Gaussian process z satisfying* (9) *and under the assumptions of Proposition* 3.1, *we have*

$$C_m(z, \overline{\rho}_n) = C_m\big(z, \rho^\star\big) + O_P(1), \qquad \textit{as } n \to \infty.$$

Proposition 4.4. *For any positive m, for a Gaussian process y satisfying* (1) *and under the assumptions of Proposition* 3.2, *we have*

$$C_m(y, \overline{\rho}_n) = C_m\big(y, \rho^\star\big) + O_P(1), \qquad \textit{as } n \to \infty.$$

The proofs of Propositions 4.3 and 4.4 are given in Appendix A.7 and A.8, respectively.

In practice, we propose to take $\overline{\rho}_n = \widetilde{\rho}_n$ which satisfies the condition of Proposition 4.4 to estimate the number of segments by

$$\begin{aligned}\widehat{m} = \arg\max_m \Bigg[& -\left(\frac{n-m+1}{2}\right) \log SS_m(y, \widetilde{\rho}_n) + \log \Gamma\left(\frac{n-m+1}{2}\right) \\ & - \frac{1}{2} \sum_{k=0}^{m} \log n_k\big(\widehat{\mathbf{t}}(y, \widetilde{\rho}_n)\big) - m \log n \Bigg],\end{aligned} \tag{18}$$

where $SS_m(\cdot,\cdot)$ and $n_k(\cdot,\cdot)$ are defined in (15) and (17), respectively.

Remark 4.2. Since the definition of the original mBIC criterion is intrinsically related to normality, we did not study precisely the quality of our approximation without the normality assumption.

5. Numerical experiments

5.1. Practical implementation

Our decorrelation procedure introduces spurious change-points in the series, at distance 1 of the true change-points (see Figure 1, top). Since these artifacts may affect our procedure, we propose a post-processing to the estimated change-points $\widehat{\mathbf{t}_n}$, which consists in removing segments of length 1:

$$PP(\widehat{\mathbf{t}_n}) = \{\widehat{t}_{n,k} \in \widehat{\mathbf{t}_n}\} \setminus \{\widehat{t}_{n,i} \text{ such that } \widehat{t}_{n,i} = \widehat{t}_{n,i-1} + 1 \text{ and } \widehat{t}_{n,i+1} \neq \widehat{t}_{n,i} + 1\}.$$

This post-processing results in a smaller number of change-points. Figure 1 summarizes the whole processing.

In practice, it may also be useful to have some guidance on how to check that the assumptions underpinning our approach are satisfied for a given data set. A possible approach is to subtract

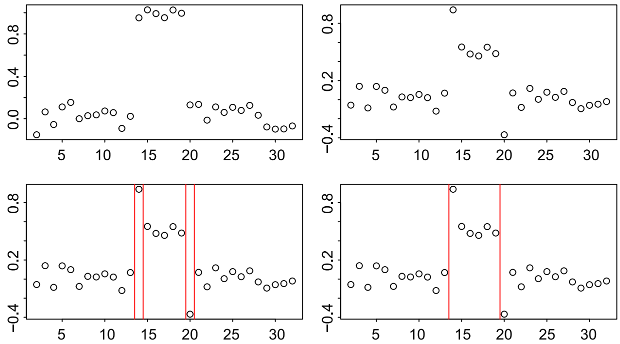

Figure 1. Top left: a series around two changes, with $\rho^\star = 0.5$. Top right: the decorrelated series in the same region. Bottom left: before post-processing, two pairs of adjacent change-points are found. Bottom right: post-processing removes the last change-point of each pair of adjacent ones.

the estimated piecewise constant function from the original series. If the model is the expected one, this new series should be a realization of a centered AR(1) process. Hence, the residuals built by decorrelation of this series should be mutually uncorrelated. One way to check this is to perform a Portmanteau test on this series of residuals.

5.2. Simulation design

To assess the performance of the proposed method, we used a simulation design inspired from the one conceived by [18]. We considered Gaussian series of length $n \in \{100, 200, 400, 800, 1600\}$ with autocorrelation at lag 1, denoted by $\rho^\star$, ranging from -0.9 to 0.9 (by steps of 0.1) and residual standard deviation $\sigma^\star$ between 0.1 and 0.6 (by steps of 0.1). All series were affected by $m^\star = 6$ change-points located at fractions $1/6 \pm 1/36, 3/6 \pm 2/36, 5/6 \pm 3/36$ of their length. Each combination was replicated $S = 100$ times. The mean within each segment alternates between 0 and 1, starting with $\mu_1 = 0$.

Estimation of $\rho^\star$

For each generated series, two different estimates $\overline{\rho}_n$ of $\rho^\star$ were computed: the original estimate $\overline{\rho}_n = \widehat{\rho}_{\mathrm{MG}}$ proposed by [26] and our revised version $\overline{\rho}_n = \widetilde{\rho}_n$. We carried the same study on series with no change-point (centered series).

Estimation of the segmentation parameters

For each generated series, we estimated the change-point locations $\widehat{\boldsymbol{\tau}}_n(y, \overline{\rho}_n)$ using Proposition 3.1 for each m from 1 to $m_{\max} = 75$ and with different choices of $\overline{\rho}_n$: $\widetilde{\rho}_n$ (our estimator), $\rho^\star$ (the true value) and zero (which does not take into account for the autocorrelation). For each choice of $\overline{\rho}_n$, we then selected the number of change-points $\widehat{m}$ using (18). Actually, the last choice $\overline{\rho}_n = 0$ corresponds to the classical least-squares framework. In addition, we shall also use the post-processing described in Section 5.1 for the cases where $\overline{\rho}_n = \widetilde{\rho}_n$ and $\rho^\star$.

To study the quality of the proposed model selection criterion, we computed the distribution of $\widehat{m}$ for each estimate $\overline{\rho}_n \in \{\widetilde{\rho}_n, \rho^\star, 0\}$ with post-processing or not for the first two estimates of $\rho^\star$.

In order to assess the performance of the estimation of the change-point locations, we computed the Hausdorff distance defined in the segmentation framework as follows, see [7,16]:

$$d\big(\boldsymbol{\tau}^\star, \widehat{\boldsymbol{\tau}}_n(y, \overline{\rho}_n)\big) = \max\big(d_1\big(\boldsymbol{\tau}^\star, \widehat{\boldsymbol{\tau}}_n(y, \overline{\rho}_n)\big), d_2\big(\boldsymbol{\tau}^\star, \widehat{\boldsymbol{\tau}}_n(y, \overline{\rho}_n)\big)\big), \tag{19}$$

where

$$d_1(\mathbf{a}, \mathbf{b}) = \sup_{b \in \mathbf{b}} \inf_{a \in \mathbf{a}} |a - b|, \quad \text{and} \tag{20}$$

$$d_2(\mathbf{a}, \mathbf{b}) = d_1(\mathbf{b}, \mathbf{a}). \tag{21}$$

d_1 close to zero means that an estimated change-point is likely to be close to a true change-point. A small value of d_2 means that a true change-point is likely to be close to each estimated change-point. A perfect segmentation results in both null d_1 and d_2. Over-segmentation results in a small d_1 and a large d_2. Under-segmentation results in a large d_1 and a small d_2, provided that the estimated change-points are correctly located.

5.3. Results

Estimation of $\rho^\star$

In Figure 2, we compare the performance of our robust estimator of $\rho^\star$: $\widetilde{\rho}_n$ with the ones of the estimator $\widehat{\rho}_{\mathrm{MG}}$ in the case where there are no change-points in the observations. More precisely, in this case, the observations y are generated under the model (1) with $\mu_k^\star = 0$, for all k. We observe that the estimator proposed by [26] performs better than our robust estimator. However, it is not the case anymore in the presence of change-points in the data as we can see in Figure 3. In the latter case, our robust estimator $\widetilde{\rho}_n$ outperforms the estimator $\widehat{\rho}_{\mathrm{MG}}$ for almost all values of $\rho^\star$.

Model selection

In Figures 4 and 5, we compare the estimated number of change-points $\widehat{m}$ in two different configurations of signal-to-noise ratio ($\sigma^\star = 0.1$ and $\sigma^\star = 0.5$) and with three different values of $\rho^\star$ ($\rho^\star = 0.3$, 0.6 and 0.8). In these figures, the notation LS, Robust and Oracle correspond to the cases where $\overline{\rho}_n = 0$, $\overline{\rho}_n = \widetilde{\rho}_n$ and $\overline{\rho}_n = \rho^\star$, respectively. Moreover, we use the notation-P when the post-processing described in Section 5.1 is used. In the situations where $\sigma^\star$ and $\rho^\star$ are small, all the methods provide an accurate estimation of the number of change-points. In the other cases, LS tends to strongly overestimate the number of change-points. Robust and Oracle tend to select twice the true number of change-points due to the artifactual presence of change-points in the decorrelated series as explained in Section 5.1. This is corrected by the post-processing and Robust-P provides the correct number of change-points in most of the considered configurations. Moreover, we also observe that the performance of Robust and Robust-P are similar to these of Oracle and Oracle-P: the robust decorrelation procedure we propose performs as well as if $\rho^\star$ was known for $n = 1600$. It has to be noted that the post-processing would not improve the performance on LS so we did not considered it.

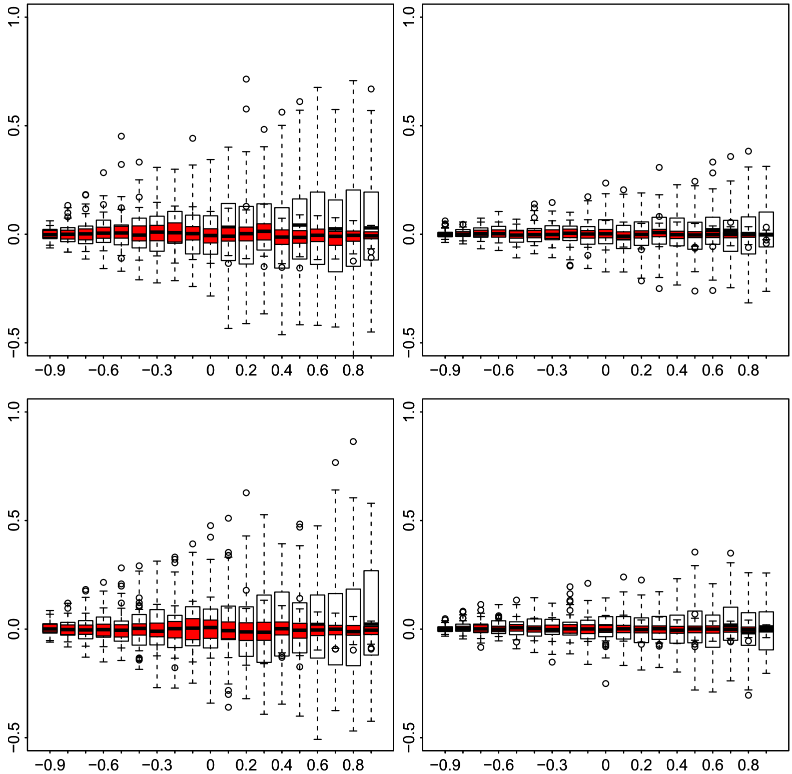

Figure 2. Boxplots of $\widehat{\rho}_{\mathrm{MG}} - \rho^\star$ in red and $\widetilde{\rho}_n - \rho^\star$ in black for different values of $\rho^\star$ in the case where there are no change-points in the data with $n = 400$ (plots on the left), $n = 1600$ (plots on the right), $\sigma^\star = 0.2$ (top) and $\sigma^\star = 0.6$ (bottom).

Change-point locations

Figures 6 and 8 display the boxplots of the two parts d_1 and d_2 of the Hausdorff distance defined in (20) and (21), respectively for different values of $\rho^\star$ when $\sigma^\star = 0.5$. d_2 is displayed in Figure 7 for $\sigma^\star = 0.1$; for this value of $\sigma^\star$, d_1 was found null for all methods and all values of $\rho^\star$.

When the noise is small ($\sigma^\star = 0.1$), the robust procedure we propose performs well for the whole range of correlation. On the contrary, the performance of LS are deprecated when the correlation increases, whereas these of LS* still provide accurate change-point locations. This shows that the least-square approach only fails because it turns to overestimate the number of change-points. This is all the more true for LS when the variance of the noise is large ($\sigma^\star = 0.5$). When the problem gets difficult (both $\sigma^\star$ and $\rho^\star$ large), our robust procedure tends to underestimate the number of change-points (which was expected) and the estimated change-points are close to true ones.

An other way to illustrate the performance of the estimation of the change-point locations is the histograms of these estimates. We provide these plots only for LS, Robust-P and Oracle-P,

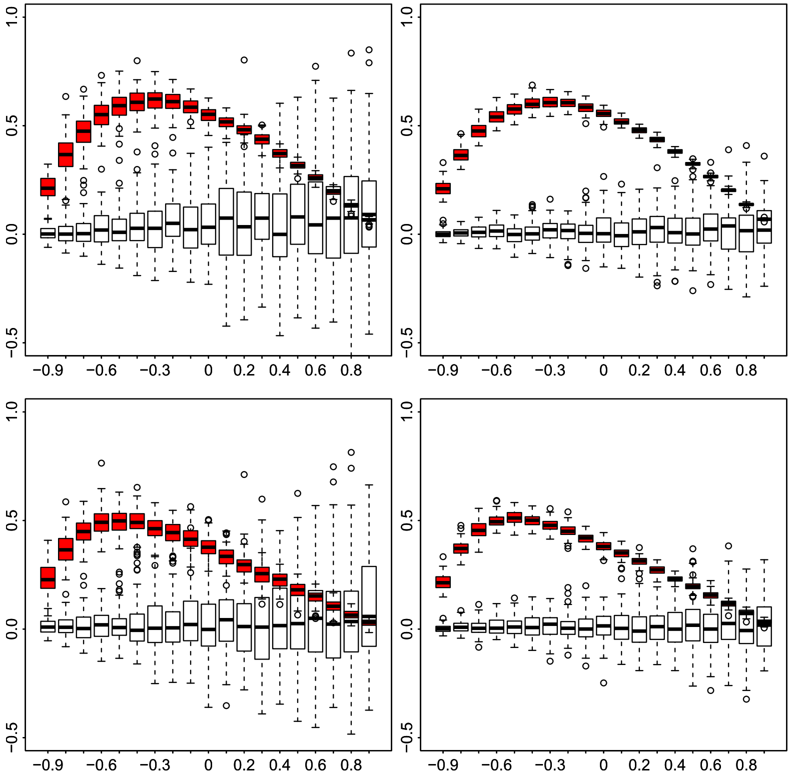

Figure 3. Boxplots of $\widehat{\rho}_{\text{MG}} - \rho^\star$ in red and $\widetilde{\rho}_n - \rho^\star$ in black for different values of $\rho^\star$ in the case where there are change-points in the data with $n = 400$ (plots on the left), $n = 1600$ (plots on the right), $\sigma^\star = 0.2$ (top) and $\sigma^\star = 0.6$ (bottom).

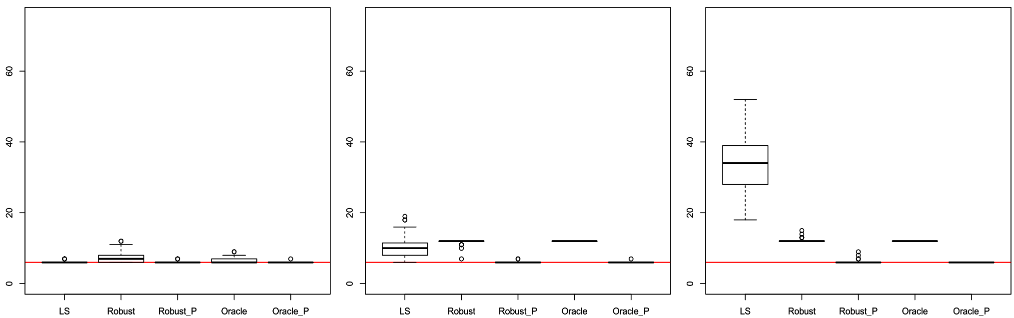

Figure 4. Boxplots for the estimated number of change-points for $n = 1600$ when $\overline{\rho}_n = 0$ (LS), $\overline{\rho}_n = \widetilde{\rho}_n$ (Robust and Robust-P with post-processing) and $\overline{\rho}_n = \rho^\star$ (Oracle and Oracle-P with post-processing) with $\sigma^\star = 0.1$ and $\rho^\star = 0.3$ (left), $\rho^\star = 0.6$ (middle) and $\rho^\star = 0.8$ (right). The true number of change-points is equal to 6 (red horizontal line).

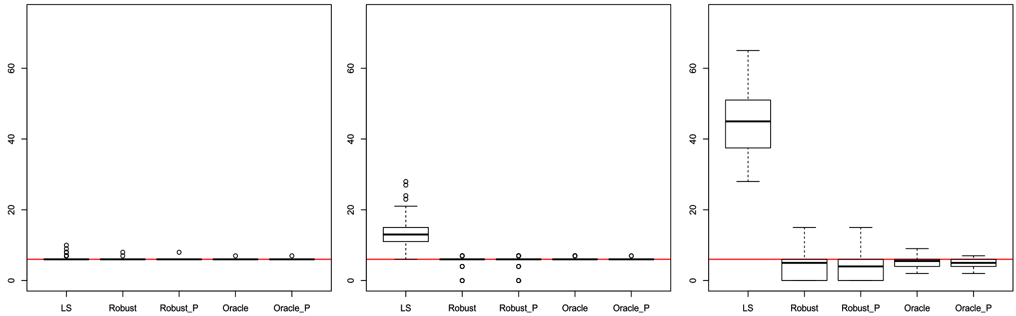

Figure 5. Boxplots for the estimated number of change-points for $n = 1600$ when $\overline{\rho}_n = 0$ (LS), $\overline{\rho}_n = \widetilde{\rho}_n$ (Robust and Robust-P with post-processing) and $\overline{\rho}_n = \rho^\star$ (Oracle and Oracle-P with post-processing) with $\sigma^\star = 0.5$ and $\rho^\star = 0.3$ (left), $\rho^\star = 0.3$ (middle) and $\rho^\star = 0.8$ (right). The true number of change-points is equal to 6 (red horizontal line).

because Post-processing does not change significantly LS estimates, and, furthermore, Robust (respectively, Oracle) method's histograms with or without Post-Processing are very similar, see Figures 9 and 10. These figures illustrate that in case of over-estimation of the number of changes by LS method, the additional change-points seem to be uniformly distributed.

5.4. Additional simulation studies

5.4.1. *Comparison with* [5]

The quasi-maximum likelihood method proposed by [5], when applied to an AR(1) process with changes in the mean $(y_0, \ldots, y_n)$, consists in the minimization with respect to $\boldsymbol{\rho} =$

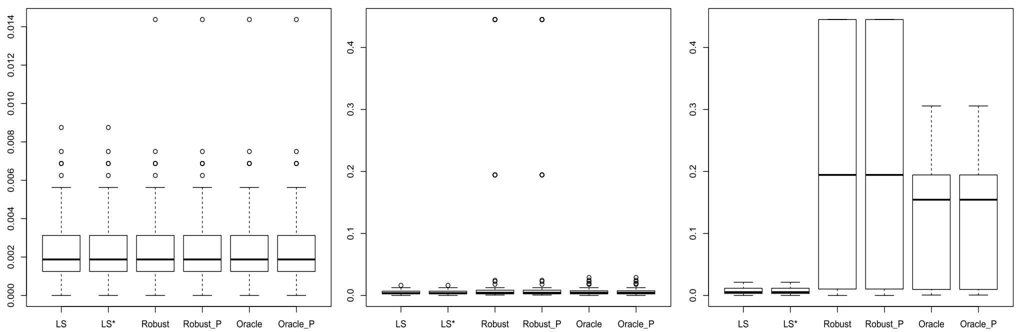

Figure 6. Boxplots for the first part of the Hausdorff distance (d_1) for $n = 1600$ when $\overline{\rho}_n = 0$ (LS and LS* when the true number of change-points is known), $\overline{\rho}_n = \widetilde{\rho}_n$ (Robust and Robust-P with post-processing) and $\overline{\rho}_n = \rho^\star$ (Oracle and Oracle-P with post-processing) with $\sigma^\star = 0.5$ and $\rho^\star = 0.3$ (left), $\rho^\star = 0.6$ (middle) and $\rho^\star = 0.8$ (right).

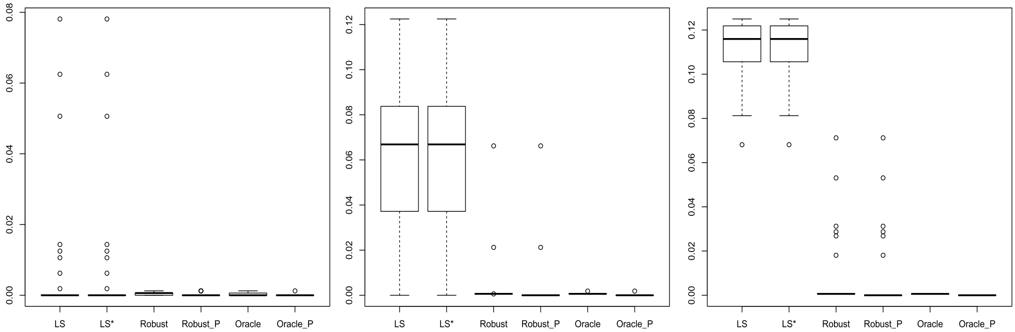

Figure 7. Boxplots for the second part of the Hausdorff distance (d_2) for $n = 1600$ when $\overline{\rho}_n = 0$ (LS and LS* when the true number of change-points is known), $\overline{\rho}_n = \widetilde{\rho}_n$ (Robust and Robust-P with post-processing) and $\overline{\rho}_n = \rho^\star$ (Oracle and Oracle-P with post-processing) with $\sigma^\star = 0.1$ and $\rho^\star = 0.3$ (left), $\rho^\star = 0.6$ (middle) and $\rho^\star = 0.8$ (right).

$(\rho_0, \dots, \rho_m)$, $\boldsymbol{\sigma} = (\sigma_0, \dots, \sigma_m)$, $\boldsymbol{\delta} = (\delta_0, \dots, \delta_m)$ and $\mathbf{t} = (t_0, \dots, t_m)$ of the following function:

$$(\boldsymbol{\rho}, \boldsymbol{\sigma}, \boldsymbol{\delta}, \mathbf{t}) \mapsto \sum_{k=0}^{m} \left\{ (t_{k+1} - t_k) \log\bigl(\sigma_k^2\bigr) + \frac{1}{\sigma_k^2} \sum_{i=t_k+1}^{t_{k+1}} (y_i - \rho_k y_{i-1} - \delta_k)^2 \right\}. \tag{22}$$

Indeed, in the class of models considered in [5], changes in all the parameters are possible at each change-point. Using this method to estimate the change-point locations for data satisfying model (1) or (9) boils down to ignore the stationarity of $(\eta_i)_{i\geq 0}$ as defined in (2). It can lead to a poor estimation of change-point locations, especially when there are many changes close to each other. To illustrate this fact, we compared our estimator of change-point locations to the estimates given by the minimization of (22). We generated 100 Gaussian series of length 400, under model (1), with $\rho^\star = 0.3$ and $\sigma^\star = 0.4$. The number of change-points, their locations and

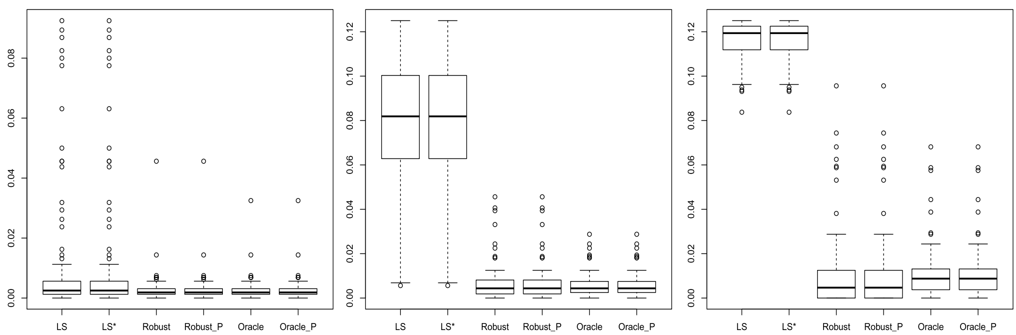

Figure 8. Boxplots for the second part of the Hausdorff distance (d_2) when $\overline{\rho}_n = 0$ (LS and LS* when the true number of change-points is known), $\overline{\rho}_n = \widetilde{\rho}_n$ (Robust and Robust-P with post-processing) and $\overline{\rho}_n = \rho^\star$ (Oracle and Oracle-P with post-processing) with $\sigma^\star = 0.5$ and $\rho^\star = 0.3$ (left), $\rho^\star = 0.6$ (middle) and $\rho^\star = 0.8$ (right).

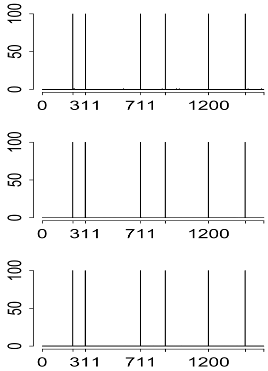

Figure 9. Frequencies of each possible change-point estimator, with $\sigma^\star = 0.1$ and $n = 1600$. Tick-marks on bottom-side axis represent the true change-point locations. $\overline{\rho}_n = 0$ (LS, top line), $\overline{\rho}_n = \widetilde{\rho}_n$ (Robust-P, middle line) and $\overline{\rho}_n = \rho^\star$ (Oracle-P, bottom line) with $\rho^\star = 0.3$ (left), $\rho^\star = 0.6$ (middle) and $\rho^\star = 0.8$ (right).

the means within segments are the same as in Section 5.2. The number of changes is assumed to be known and we did not post-process the estimates. Simulations show that using the method of [5] in this case can lead to a poor estimation of close change-points, while our method is less affected by the length of segments (see Figure 11). For example, the boundaries of the smallest segment are recovered in less than half of the simulations when minimizing (22).

5.4.2. *Robustness to model mis-specification*

In this section, we study the behaviour of our proposed robust procedure (Robust-P) when the signal is corrupted by an AR(2) Gaussian process, for example, in model 1, η_i is a zero-mean

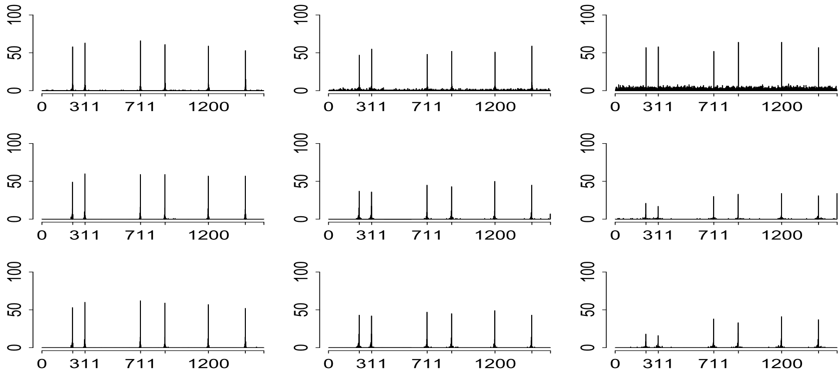

Figure 10. Frequencies of each possible change-point estimator, with $\sigma^\star = 0.5$ and $n = 1600$. Tick-marks on bottom-side axis represent the true change-point locations. $\overline{\rho}_n = 0$ (LS, top line), $\overline{\rho}_n = \widetilde{\rho}_n$ (Robust-P, middle line) and $\overline{\rho}_n = \rho^\star$ (Oracle-P, bottom line) with $\rho^\star = 0.3$ (left), $\rho^\star = 0.6$ (middle) and $\rho^\star = 0.8$ (right).

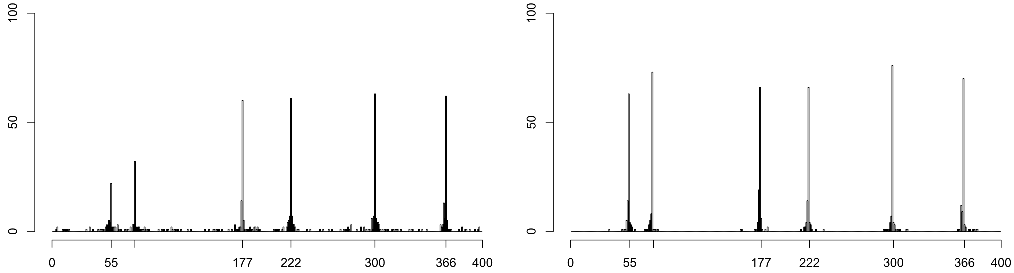

Figure 11. Frequencies of each possible change-point location estimate. Tick-marks on bottom-side axis represent the true change-point locations. Left: Estimation by the minimization of (22). Right: Our method.

stationary process such that

$$\eta_i = \phi_1^\star \eta_{i-1} + \phi_2^\star \eta_{i-2} + \epsilon_i,$$

where $|\phi_2| < 1$, $\phi_1 + \phi_2 < 1$ and $\phi_2 - \phi_1 < 1$. We considered series of fixed length $n = 1600$, a residual standard deviation $\sigma^\star = 0.1$, $\phi_1^\star = 0.3$ and $\phi_2^\star$ in $\{-0.9, -0.8, -0.7, \ldots, 0.5, 0.6\}$. We used the same segmentation design as in Section 5.1. Each combination was replicated 100 times. All the results are displayed in Figure 12.

The procedure performs well when $\phi_2^\star$ belongs to the interval $[-0.5, 0.2]$ as expected (similar to the case of AR(1)): the estimated segmentation is close to the true one. When $\phi_2^\star > 0.2$, it tends to over-estimate the number of change-points. The true change-points are detected (d_1 is close to zero, e.g., the decorrelation procedure with the obtained negative estimation of $\rho^\star$ leads to an increasing in the mean differences) but false change-points are added (large d_2). When $\phi_2^\star < -0.5$, under-segmentation is observed: the decorrelation procedure with a large estimated value of $\rho^\star$ leads to a difficult segmentation problem.

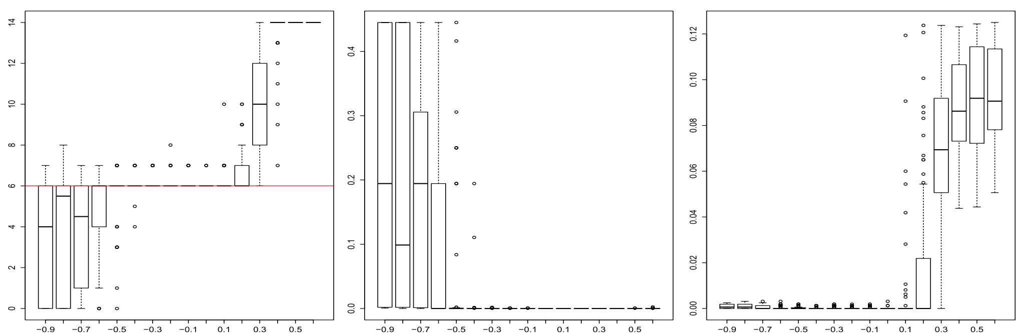

Figure 12. Left: Boxplots for the estimated number of change-points. Center and right: Boxplots for the first part of the Hausdorff distance (d_1) and for the second part of the Hausdorff distance (d_2) with $n = 1600$, $\sigma^\star = 0.1$ and $\phi_1^\star = 0.3$ with respect to different values of $\phi_2^\star$.

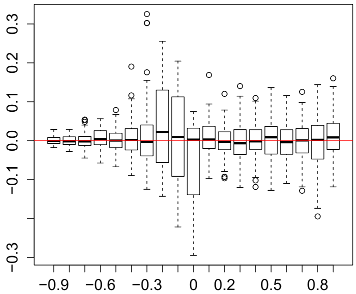

Figure 13. Boxplots of $\widetilde{\widetilde{\rho}}_n - \rho^\star$ for different values of $\rho^\star$ when $n = 1600$ and $\sigma^\star = 0.1$.

5.4.3. *Estimator of $\rho^\star$ in the case of the Cauchy distribution*

In Section 2, an analogous estimator of $\rho^\star$ in the case of Cauchy distributed observations is proposed. We follow the simulation design described in Section 5.2, where the Gaussian r.v.'s are replaced by Cauchy r.v.'s. More precisely, the expectation parameters are replaced by the location parameters of the Cauchy distribution and $\sigma^\star$ is replaced by the scale parameter of the Cauchy distribution. We can see from Figure 13 that $\widetilde{\widetilde{\rho}}_n$ is an accurate estimator of $\rho^\star$ except when $\rho^\star$ is close to zero. When this estimator of $\rho^\star$ is used in our change-point estimation method, it leads to poor estimations of the change-points since the Cauchy distribution does not have finite second order moment (simulations not shown).

6. Conclusion

In this paper, we propose a novel approach for estimating multiple change-points in the mean of an AR(1) process. Our approach is based on two main stages. The first one consists in building a robust estimator of the autocorrelation parameter which is used for whitening the original series. In the second stage, we apply the inference approach commonly used to estimate change-points in the mean of independent random variables. The Gaussianity assumption is only made in the Propositions 2.1 (proposed auto-correlation estimator), 4.3 and 4.4 (mBIC criterion). In the course of this study, we have shown that our approach, which is implemented in the R package AR1seg, is a very efficient technique both on a theoretical and practical point of view. More precisely, it has two main features which make it very attractive. First, the estimators that we propose have the same asymptotic properties as the classical estimators in the independent framework which means that the performances of our estimators are not affected by the dependence assumption. Second, from a practical point of view, AR1seg is computationally efficient and exhibits better performance on finite sample size data than existing approaches which do not take into account the dependence structure of the observations.

Appendix: Proofs

A.1. Proof of Proposition 2.1

Let F_1 and F_2 denote the cumulative distribution functions (c.d.f.) of $(|y_{i+1} - y_i|)$ for $i \neq t^\star_{n,1}, \dots, t^\star_{n,m^\star}$ and $(|y_{i+2} - y_i|)$ for $i \neq t^\star_{n,1} - 1, \dots, t^\star_{n,m^\star} - 1$, respectively. By (1), $(y_i -$

$\mathbb{E}(y_i))_{0\leq i\leq n}$ are $(n+1)$ observations of a AR(1) stationary Gaussian process thus for any $i \neq t^\star_{n,1}, \ldots, t^\star_{n,m^\star}$, $(y_{i+1} - y_i)$ and for any $i \neq t^\star_{n,1} - 1, \ldots, t^\star_{n,m^\star} - 1$, $(y_{i+2} - y_i)$ are zero-mean Gaussian r.v.'s with variances equal to $2\sigma^{\star 2}/(1+\rho^\star)$ and $2\sigma^{\star 2}$, respectively. Hence, for all t in $\mathbb{R}$,

$$F_1 : t \mapsto 2\Phi\left(t\sqrt{\frac{1+\rho^\star}{2\sigma^{\star 2}}}\right) - 1 \quad \text{and} \quad F_2 : t \mapsto 2\Phi\left(t\sqrt{\frac{1}{2\sigma^{\star 2}}}\right) - 1, \tag{23}$$

where Φ denotes the cumulative distribution function of a standard Gaussian r.v.

Let also denote by $F_{1,n}$ and $F_{2,n-1}$ the empirical cumulative distribution functions of $(|y_{i+1} - y_i|)_{0\leq i\leq n-1}$ and $(|y_{i+2} - y_i|)_{0\leq i\leq n-2}$, respectively. Observe that for all t in $\mathbb{R}$,

$$\begin{aligned}
&\sqrt{n}\big(F_{1,n}(t) - F_1(t)\big) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \big(\mathbb{1}_{\{|y_{i+1}-y_i|\leq t\}} - F_1(t)\big) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{i\in\{t^\star_{n,1},\ldots,t^\star_{n,m^\star}\}} \big(\mathbb{1}_{\{|y_{i+1}-y_i|\leq t\}} - F_1(t)\big) + \frac{1}{\sqrt{n}} \sum_{\substack{0\leq i\leq n-1 \\ i\notin\{t^\star_{n,1},\ldots,t^\star_{n,m^\star}\}}} \big(\mathbb{1}_{\{|y_{i+1}-y_i|\leq t\}} - F_1(t)\big) \\
&\quad = \frac{1}{\sqrt{n}} \sum_{0\leq i\leq n-1} \big(\mathbb{1}_{\{|z_i|\leq t\}} - F_1(t)\big) + R_n(t),
\end{aligned} \tag{24}$$

where $\sup_{t\in\mathbb{R}} |R_n(t)| = o_p(1)$, the $z_i = y_{i+1} - y_i$ except for $i = t^\star_{n,1}, \ldots, t^\star_{n,m^\star}$, where $z_i = \eta_{i+1} - \eta_i$, (η_i) being defined in (2).

Thus, by using the theorem of [11], we obtain that the first term in the right-hand side of (24) converges in distribution to a zero-mean Gaussian process G in the space of càdlàg functions equipped with the uniform norm. Since the second term in the right-hand side tends uniformly to zero in probability, we get that $\sqrt{n}(F_{1,n} - F_1)$ converges in distribution to a zero-mean Gaussian process in the space of càdlàg functions equipped with the uniform norm and that the same holds for $\sqrt{n-1}(F_{2,n-1} - F_2)$.

By Lemma 21.3 of [34], the quantile function $T : F \mapsto F^{-1}(1/2)$ is Hadamard differentiable at F tangentially to the set of càdlàg functions h that are continuous at $F^{-1}(1/2)$ with derivative $T'_F(h) = -h(F^{-1}(1/2))/F'(F^{-1}(1/2))$. By applying the functional delta method [34], Theorem 20.8, we get that $\sqrt{n}(T(F_{1,n}) - T(F_1))$ converges in distribution to $T'_{F_1}(G)$. Moreover, by the Continuous mapping theorem, it is the same for $T'_{F_1}\{\sqrt{n}(F_{1,n} - F_1)\}$. Thus,

$$\begin{aligned}
\sqrt{n}\big(F_{1,n}^{-1}(1/2) - F_1^{-1}(1/2)\big) &= T'_{F_1}\big\{\sqrt{n}(F_{1,n} - F_1)\big\} + o_p(1) \\
&= -\frac{1}{\sqrt{n}} \frac{\sum_{i=0}^{n-1} (\mathbb{1}_{\{|y_{i+1}-y_i|\leq F_1^{-1}(1/2)\}} - 1/2)}{F_1'(F_1^{-1}(1/2))} + o_p(1).
\end{aligned}$$

In the same way,

$$\sqrt{n-1}\big(F_{2,n-1}^{-1}(1/2)-F_2^{-1}(1/2)\big) = -\frac{1}{\sqrt{n-1}}\frac{\sum_{i=0}^{n-2}(\mathbb{1}_{\{|y_{i+2}-y_i|\le F_2^{-1}(1/2)\}}-1/2)}{F_2'(F_2^{-1}(1/2))}+o_p(1).$$

By applying the Delta method [34], Theorem 3.1, with the transformation $f(x)=x^2$, we get

$$\sqrt{n}\big(F_{1,n}^{-1}(1/2)^2-F_1^{-1}(1/2)^2\big) = -\frac{2F_1^{-1}(1/2)}{\sqrt{n}}\frac{\sum_{i=0}^{n-1}(\mathbb{1}_{\{|y_{i+1}-y_i|\le F_1^{-1}(1/2)\}}-1/2)}{F_1'(F_1^{-1}(1/2))}+o_p(1),$$

$$\sqrt{n-1}\big(F_{2,n-1}^{-1}(1/2)^2-F_2^{-1}(1/2)^2\big) = -\frac{2F_2^{-1}(1/2)}{\sqrt{n-1}}\frac{\sum_{i=0}^{n-2}(\mathbb{1}_{\{|y_{i+2}-y_i|\le F_2^{-1}(1/2)\}}-1/2)}{F_2'(F_2^{-1}(1/2))}+o_p(1).$$

Note that by (23), we obtain that

$$F_1^{-1}(1/2)=\sqrt{\frac{2\sigma^{\star 2}}{1+\rho^\star}}\Phi^{-1}(3/4) \quad \text{and} \quad F_2^{-1}(1/2)=\sqrt{2\sigma^{\star 2}}\Phi^{-1}(3/4). \tag{25}$$

Moreover,

$$\begin{aligned} F_1'\big(F_1^{-1}(1/2)\big)&=2\sqrt{\frac{1+\rho^\star}{2\sigma^{\star 2}}}\varphi\big(\Phi^{-1}(3/4)\big) \quad \text{and} \\ F_2'\big(F_2^{-1}(1/2)\big)&=2\sqrt{\frac{1}{2\sigma^{\star 2}}}\varphi\big(\Phi^{-1}(3/4)\big), \end{aligned} \tag{26}$$

where φ denotes the probability distribution function of a standard Gaussian r.v.

Observe that $\sqrt{n}(\widetilde{\rho}_n-\rho^\star)$ can be rewritten as follows:

$$\begin{aligned} \sqrt{n}\big(\widetilde{\rho}_n-\rho^\star\big)&=\sqrt{n}\frac{F_{2,n}^{-1}(1/2)^2-(1+\rho^\star)F_{1,n}^{-1}(1/2)^2}{F_{1,n}^{-1}(1/2)^2} \\ &=\sqrt{n}\frac{(F_{2,n-1}^{-1}(1/2)^2-F_2^{-1}(1/2)^2)-(1+\rho^\star)(F_{1,n}^{-1}(1/2)^2-F_1^{-1}(1/2)^2)}{F_{1,n}^{-1}(1/2)^2} \\ &\quad+\sqrt{n}\frac{F_2^{-1}(1/2)^2-(1+\rho^\star)F_1^{-1}(1/2)^2}{F_{1,n}^{-1}(1/2)^2}. \end{aligned} \tag{27}$$

By (25) the last term in the right-hand side of (27) is equal to zero. Thus,

$$
\begin{aligned}
&F_{1,n}^{-1}(1/2)^2\sqrt{n}\big(\widetilde{\rho}_n-\rho^\star\big)\\
&\quad=\frac{1}{\sqrt{n-1}}\sum_{i=0}^{n-2}\big\{a_2(\mathbb{1}_{\{|y_{i+2}-y_i|\le F_2^{-1}(1/2)\}}-1/2)-a_1\big(1+\rho^\star\big)(\mathbb{1}_{\{|y_{i+1}-y_i|\le F_1^{-1}(1/2)\}}-1/2)\big\}\\
&\qquad+o_p(1),
\end{aligned}
$$

where, by (26),

$$
\begin{aligned}
a_2&=-\frac{2F_2^{-1}(1/2)}{F_2'(F_2^{-1}(1/2))}=-2\sigma^{\star 2}\frac{\Phi^{-1}(3/4)}{\varphi(\Phi^{-1}(3/4))}\quad\text{and}\\
a_1&=-\frac{2F_1^{-1}(1/2)}{F_1'(F_1^{-1}(1/2))}=-\frac{2\sigma^{\star 2}}{1+\rho^\star}\frac{\Phi^{-1}(3/4)}{\varphi(\Phi^{-1}(3/4))}.
\end{aligned}
$$

By (25), $\sqrt{n}(\widetilde{\rho}_n-\rho^\star)$ can thus be rewritten as follows:

$$
F_{1,n}^{-1}(1/2)^2\sqrt{n}\big(\widetilde{\rho}_n-\rho^\star\big)=\frac{2\sigma^{\star 2}\Phi^{-1}(3/4)^2}{1+\rho^\star}\frac{1}{\sqrt{n-1}}\sum_{0\le i\le n-2}\Psi(\eta_i,\eta_{i+1},\eta_{i+2})+o_p(1),
$$

where Ψ is defined in (6) and (η_i) is defined in (2). Since Ψ is a function on $\mathbb{R}^3$ with Hermite rank greater than 1 and $(\eta_i)_{i\ge 0}$ is a stationary AR(1) Gaussian process, and since $F_{1,n}^{-1}(1/2)^2$ converges in probability to $F_1^{-1}(1/2)^2$, (4) follows by applying [1], Theorem 4, Slutsky's lemma and equation (25).

A.2. Hints for (7)

Note that if X has a Cauchy(x_0,γ) distribution then the characteristic function φ_X of X can be written as $\varphi_X(t)=\mathrm{e}^{\mathrm{i}x_0t-\gamma|t|}$. Moreover, the c.d.f. F_X of X is such that $F_X^{-1}(3/4)=x_0+\gamma$. Thus, $\eta_i=\sum_{k\ge 0}(\rho^\star)^k\epsilon_{i-k}$ has a Cauchy$(\frac{x_0}{1-\rho^\star},\frac{\gamma}{1-|\rho^\star|})$ distribution and $(\rho^\star-1)\eta_i$ has a Cauchy$(-x_0,\frac{\gamma|\rho^\star-1|}{1-|\rho^\star|})$ distribution. Since $\eta_{i+1}-\eta_i=(\rho^\star-1)\eta_i+\epsilon_i$ is a sum of two independent Cauchy r.v.'s, it is distributed as a Cauchy$(0,\gamma(1+|\frac{\rho^\star-1}{1-|\rho^\star|}|))$ distribution. In the same way, $\eta_{i+2}-\eta_i=(\rho^{\star 2}-1)\eta_i+\rho^\star\epsilon_i+\epsilon_{i+2}$ is a sum of three independent Cauchy r.v.'s and has thus a Cauchy$(0,2\gamma(1+|\rho^\star|))$. Let F_1 and F_2 denote the c.d.f. of $(\eta_{i+1}-\eta_i)$ and $(\eta_{i+2}-\eta_i)$, respectively. By using the properties of the c.d.f. of a Cauchy distribution, we get, on the one hand, that $F_2^{-1}(3/4)=2\gamma(1+|\rho^\star|)$ and, on the other hand, that

$$
F_1^{-1}(3/4)=\begin{cases}2\gamma, & \text{if } \rho^\star>0,\\[4pt] \dfrac{2\gamma}{1+\rho^\star}, & \text{if } \rho^\star<0.\end{cases}
$$

From this we get that

$$\left(\frac{F_2^{-1}(3/4)}{F_1^{-1}(3/4)}\right)^2 - 1 = \begin{cases} \rho^\star(2+\rho^\star), & \text{if } \rho^\star > 0, \\ \rho^{\star 2}(\rho^{\star 2}-2), & \text{if } \rho^\star < 0. \end{cases}$$

The definition of $\widetilde{\widetilde{\rho}}_n$ comes by inverting these last two functions.

A.3. Proof of Proposition 3.1

In the sequel, we need the following definitions, notations and remarks. Observe that (9) can be rewritten as follows:

$$z = \rho^* Bz + T(\mathbf{t}_n^\star)\boldsymbol{\delta}^\star + \epsilon, \tag{28}$$

where

$$z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}, \quad Bz = \begin{pmatrix} z_0 \\ \vdots \\ z_{n-1} \end{pmatrix}, \quad \boldsymbol{\delta}^\star = \begin{pmatrix} \delta_0^\star \\ \vdots \\ \delta_m^\star \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \tag{29}$$

where $\delta_k^\star = (1-\rho^\star)\mu_k^\star$, for $0 \le k \le m$, and $T(\mathbf{t})$ is an $n \times (m+1)$ matrix where the kth column is $(\underbrace{0,\ldots,0}_{t_{k-1}} \underbrace{1,\ldots,1}_{t_k - t_{k-1}} \underbrace{0,\ldots,0}_{n-t_k})^T$.

Let us define the exact and estimated decorrelated series by

$$w^\star = z - \rho^\star Bz, \tag{30}$$

$$\overline{w} = z - \overline{\rho}_n Bz. \tag{31}$$

For any vector subspace E of $\mathbb{R}^n$, let π_E denote the orthogonal projection of $\mathbb{R}^n$ on E. Let also $\|\cdot\|$ be the Euclidean norm on $\mathbb{R}^n$, $\langle\cdot,\cdot\rangle$ the canonical scalar product on $\mathbb{R}^n$ and $\|\cdot\|_\infty$ the sup norm. For x a vector of $\mathbb{R}^n$ and $\mathbf{t} \in \mathcal{A}_{n,m}$, let

$$J_{n,m}(x,\mathbf{t}) = \frac{1}{n}\left(\left\|\pi_{E_{\mathbf{t}_n^\star}}(x)\right\|^2 - \left\|\pi_{E_\mathbf{t}}(x)\right\|^2\right), \tag{32}$$

written $J_n(x,\mathbf{t})$ in the sequel for notational simplicity. In (32), $E_{\mathbf{t}_n^\star}$ and $E_\mathbf{t}$ correspond to the linear subspaces of $\mathbb{R}^n$ generated by the columns of $T(\mathbf{t}_n^\star)$ and $T(\mathbf{t})$, respectively. We shall use the same decomposition as the one introduced in [21]:

$$J_n(x,\mathbf{t}) = K_n(x,\mathbf{t}) + V_n(x,\mathbf{t}) + W_n(x,\mathbf{t}), \tag{33}$$

where

$$K_n(x,\mathbf{t}) = \frac{1}{n}\left\|(\pi_{E_{\mathbf{t}_n^\star}} - \pi_{E_\mathbf{t}})\mathbb{E}x\right\|^2, \tag{34}$$

$$V_n(x,\mathbf{t}) = \frac{1}{n}\left(\left\|\pi_{E_{\mathbf{t}_n^\star}}(x-\mathbb{E}x)\right\|^2 - \left\|\pi_{E_\mathbf{t}}(x-\mathbb{E}x)\right\|^2\right), \tag{35}$$

$$W_n(x,\mathbf{t}) = \frac{2}{n}\big(\langle \pi_{E_{\mathbf{t}_n^\star}}(x-\mathbb{E}x), \pi_{E_{\mathbf{t}_n^\star}}(\mathbb{E}x)\rangle - \langle \pi_{E_{\mathbf{t}}}(x-\mathbb{E}x), \pi_{E_{\mathbf{t}}}(\mathbb{E}x)\rangle\big). \tag{36}$$

We shall also use the following notations:

$$\underline{\lambda} = \min_{1\le k\le m}\big|\delta_k^\star - \delta_{k-1}^\star\big|, \tag{37}$$

$$\overline{\lambda} = \max_{1\le k\le m}\big|\delta_k^\star - \delta_{k-1}^\star\big|, \tag{38}$$

$$\Delta_{\boldsymbol{\tau}^\star} = \min_{1\le k\le m+1}\big(\tau_k^\star - \tau_{k-1}^\star\big), \tag{39}$$

$$\mathcal{C}_{\nu,\gamma,n,m} = \big\{\mathbf{t}\in\mathcal{A}_{n,m};\, \nu\underline{\lambda}^{-2} \le \|\mathbf{t}-\mathbf{t}_n^\star\| \le n\gamma\Delta_{\boldsymbol{\tau}^\star}\big\}, \tag{40}$$

$$\mathcal{C}'_{\nu,\gamma,n,m} = \mathcal{C}_{\nu,\gamma,n,m} \cap \big\{\mathbf{t}\in\mathcal{A}_{n,m};\, \forall k=1,\dots,m,\, t_k \ge t_{n,k}^\star\big\}, \tag{41}$$

$$\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I}) = \big\{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m};\ \forall k\in\mathcal{I},\, \nu\underline{\lambda}^{-2} \le t_k - t_{n,k}^\star \le n\gamma\Delta_{\boldsymbol{\tau}^\star} \text{ and } \forall k\notin\mathcal{I},\, t_k - t_{n,k}^\star < \nu\underline{\lambda}^{-2}\big\}, \tag{42}$$

for any $\nu > 0$, $0 < \gamma < 1/2$ and $\mathcal{I}\subset\{1,\dots,m\}$. We shall also need the following lemmas in order to prove Proposition 3.1 which are proved below.

Lemma A.1. *Let* $(z_0,\dots,z_n)$ *be defined by* (1) *or* (9), *then*

$$\|Bz\| = O_P\big(n^{1/2}\big), \tag{43}$$

$$\|z\| = O_P\big(n^{1/2}\big), \tag{44}$$

as n *tends to infinity, where* Bz *and* z *are defined in* (29).

Lemma A.2. *Let* $(z_0,\dots,z_n)$ *be defined by* (1) *or* (9) *then, for all* $\mathbf{t}\in\mathcal{A}_{n,m}$,

$$\big|J_n(\overline{w},\mathbf{t}) - J_n(w^\star,\mathbf{t})\big| \le 2\frac{|\rho^\star - \overline{\rho}|}{n}\|Bz\|\big(\big|\rho^\star+\overline{\rho}\big|\|Bz\| + 2\|z\|\big) = O_P\big(n^{-1/2}\big) = o_P(1),$$

as $n\to\infty$, *where* J_n *is defined in* (32), Bz *and* z *are defined in* (29), $w^\star$ *is defined in* (30) *and* $\overline{w}$ *is defined in* (31).

Lemma A.3. *Under the assumptions of Proposition* 3.1, $\|\overline{\boldsymbol{\tau}}_n - \boldsymbol{\tau}^\star\|_\infty$ *converges in probability to* 0, *as* n *tends to infinity.*

Lemma A.4. *Under the assumptions of Proposition* 3.1 *and for any* $\nu > 0$, $0 < \gamma < 1/2$ *and* $\mathcal{I}\subset\{1,\dots,m\}$,

$$P\Big(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})}\Big(\frac{1}{2}K_n\big(w^\star,\mathbf{t}\big) + V_n\big(w^\star,\mathbf{t}\big) + W_n\big(w^\star,\mathbf{t}\big)\Big) \le 0\Big) \longrightarrow 0, \qquad \text{as } n\to\infty,$$

where $\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})$ *is defined in* (42) *and* $w^\star$ *is defined in* (30).

Lemma A.5. *Under the assumptions of Proposition 3.1 and for any $\nu > 0$, $0 < \gamma < 1/2$ and $\mathcal{I} \subset \{1, \dots, m\}$,*

$$P\Big(\min_{\mathbf{t} \in \mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} J_n(\overline{w}, \mathbf{t}) \le 0\Big) \longrightarrow 0, \qquad \text{as } n \to \infty,$$

where $\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})$ is defined in (42) and $\overline{w}$ is defined in (31).

Lemma A.6. *Under the assumptions of Proposition 3.1,*

$$\big\|\widehat{\boldsymbol{\tau}}_n(z, \overline{\rho}_n) - \boldsymbol{\tau}^\star\big\|_\infty = O_P\big(n^{-1}\big).$$

Proof of Lemma A.1. Without loss of generality, assume $(z_0, \dots, z_n)$ is defined by (9). $\|z\|^2 - \|Bz\|^2 = z_n^2 - z_0^2 = O_P(1)$ thus we only need to prove (43). $\|Bz\|^2 = \sum_{i=0}^{n-1} z_i^2$ then Markov inequality implies that $\|Bz\|^2 = O_P(n)$. □

Proof of Lemma A.2. By (30), $\overline{w} = w^\star + (\rho^\star - \overline{\rho}_n)Bz$. Thus, by (32), we get

$$\begin{aligned} & J_n(\overline{w}, \mathbf{t}) - J_n\big(w^\star, \mathbf{t}\big) \\ &\quad = \frac{(\rho^\star - \overline{\rho}_n)^2}{n}\big\|\pi_{E_{\mathbf{t}^\star}}(Bz)\big\|^2 + \frac{2(\rho^\star - \overline{\rho}_n)}{n}\big\langle \pi_{E_{\mathbf{t}^\star}}\big(z - \rho^\star Bz\big), \pi_{E_{\mathbf{t}^\star}}(Bz)\big\rangle \\ &\qquad - \frac{(\rho^\star - \overline{\rho}_n)^2}{n}\big\|\pi_{E_{\mathbf{t}}}(Bz)\big\|^2 - \frac{2(\rho^\star - \overline{\rho}_n)}{n}\big\langle \pi_{E_{\mathbf{t}}}\big(z - \rho^\star Bz\big), \pi_{E_{\mathbf{t}}}(Bz)\big\rangle. \end{aligned} \tag{45}$$

Observe that the sum of the first two term in the right-hand side of (45) can be rewritten as follows:

$$\begin{aligned} & \frac{1}{n}\big(\rho^\star - \overline{\rho}_n\big)\big\langle \pi_{E_{\mathbf{t}^\star}}(Bz), \big(\rho^\star - \overline{\rho}_n\big)\pi_{E_{\mathbf{t}^\star}}(Bz) + 2\pi_{E_{\mathbf{t}^\star}}\big(z - \rho^\star Bz\big)\big\rangle \\ &\quad = \frac{1}{n}\big(\rho^\star - \overline{\rho}_n\big)\big\langle \pi_{E_{\mathbf{t}^\star}}(Bz), \pi_{E_{\mathbf{t}^\star}}\big(2z - \big(\rho^\star + \overline{\rho}_n\big)Bz\big)\big\rangle. \end{aligned}$$

Since the same can be done for the last two terms in the right-hand side of (45), the Cauchy–Schwarz inequality and the 1-Lipschitz property of projections give

$$\big|J_n(\overline{w}, \mathbf{t}) - J_n\big(w^\star, \mathbf{t}\big)\big| \le 2\frac{|\rho^\star - \overline{\rho}_n|}{n}\|Bz\|\big(\big|\rho^\star + \overline{\rho}_n\big|\|Bz\| + 2\|z\|\big).$$

The conclusion follows from (13) and Lemma A.1. □

Proof of Lemma A.3. [21], proof of Theorem 3, give the following bounds for any $\mathbf{t} \in \mathcal{A}_{n,m}$:

$$K_n\big(w^\star, \mathbf{t}\big) \ge \underline{\lambda}^2 \min\bigg(\frac{1}{n}\max_{1 \le k \le m}\big|t_k - t^\star_{n,k}\big|, \Delta_{\boldsymbol{\tau}^\star}\bigg), \tag{46}$$

$$V_n(w^\star, \mathbf{t}) \geq -\frac{2(m+1)}{n\Delta_n}\left(\max_{1\leq s\leq n}\left(\sum_{i=1}^{s}\epsilon_i\right)^2 + \max_{1\leq s\leq n}\left(\sum_{i=n-s}^{n}\epsilon_i\right)^2\right), \tag{47}$$

$$\left|W_n(w^\star, \mathbf{t})\right| \leq \frac{3(m+1)^2\overline{\lambda}}{n}\left(\max_{1\leq s\leq n}\left|\sum_{i=1}^{s}\epsilon_i\right| + \max_{1\leq s\leq n}\left|\sum_{i=n-s}^{n}\epsilon_i\right|\right), \tag{48}$$

where $\underline{\lambda}$, $\overline{\lambda}$ and $\Delta_{\boldsymbol{\tau}^\star}$ are defined in (37)–(39). For any $\nu > 0$, define, as in [21], proof of Theorem 3,

$$\mathcal{C}_{n,m,\nu} = \left\{\mathbf{t} \in \mathcal{A}_{n,m};\ \left\|\mathbf{t} - \mathbf{t}_n^\star\right\|_\infty \geq n\nu\right\}. \tag{49}$$

For $0 < \nu < \Delta_{\boldsymbol{\tau}^\star}$, we have:

$$\begin{aligned}
P\left(\left\|\widehat{\mathbf{t}}_n(z, \overline{\rho}_n) - \mathbf{t}_n^\star\right\|_\infty \geq n\nu\right) &\leq P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}} J_n(\overline{w}, \mathbf{t}) \leq 0\right) \\
&\leq P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}} \left(J_n(\overline{w}, \mathbf{t}) - J_n(w^\star, \mathbf{t})\right) \leq -\nu\underline{\lambda}^2\right) \\
&\quad + P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}} \left(V_n(w^\star, \mathbf{t}) + W_n(w^\star, \mathbf{t})\right) \leq -\nu\underline{\lambda}^2\right) \\
&\leq P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}} \left(J_n(\overline{w}, \mathbf{t}) - J_n(w^\star, \mathbf{t})\right) \leq -\nu\underline{\lambda}^2\right) \\
&\quad + P\left(\max_{1\leq s\leq n}\left(\sum_{i=1}^{s}\epsilon_i\right)^2 + \max_{1\leq s\leq n}\left(\sum_{i=n-s}^{n}\epsilon_i\right)^2 \geq c\underline{\lambda}^2 n\Delta_n\nu\right) \\
&\quad + P\left(\max_{1\leq s\leq n}\left|\sum_{i=1}^{s}\epsilon_i\right| + \max_{1\leq s\leq n}\left|\sum_{i=n-s}^{n}\epsilon_i\right| \geq c\underline{\lambda}^2 n\nu\overline{\lambda}^{-1}\right)
\end{aligned} \tag{50}$$

for some positive constant c. The last two terms of this sum go to 0 when n goes to infinity [21], proof of Theorem 3. To show that the first term shares the same property, it suffices to show that $J_n(\overline{w}, \mathbf{t}) - J_n(w^\star, \mathbf{t})$ is bounded uniformly in $\mathbf{t}$ by a sequence of r.v.'s which converges to 0 in probability. This result holds by Lemma A.2. □

Proof of Lemma A.4. Using equations (64)–(66) of [21], one can show the bound (73) of [21] on

$$P\left(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} \left(K_n(w^\star, \mathbf{t}) + V_n(w^\star, \mathbf{t}) + W_n(w^\star, \mathbf{t})\right) \leq 0\right).$$

Using the same arguments, we have the same bound on

$$P\left(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} \left(\frac{1}{2}K_n(w^\star, \mathbf{t}) + V_n(w^\star, \mathbf{t}) + W_n(w^\star, \mathbf{t})\right) \leq 0\right).$$

We conclude using equations (67)–(71) of [21]. □

Proof of Lemma A.5. By (33),

$$P\Big(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} J_n(\overline{w},\mathbf{t})\le 0\Big)$$
$$\le P\Big(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})}\Big(J_n(\overline{w},\mathbf{t})-J_n\big(w^\star,\mathbf{t}\big)+\frac{1}{2}K_n\big(w^\star,\mathbf{t}\big)\Big)\le 0\Big)$$
$$+P\Big(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})}\Big(\frac{1}{2}K_n\big(w^\star,\mathbf{t}\big)+V_n\big(w^\star,\mathbf{t}\big)+W_n\big(w^\star,\mathbf{t}\big)\Big)\le 0\Big).$$

By Lemma A.4, the conclusion thus follows if

$$P\Big(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})}\Big(J_n(\overline{w},\mathbf{t})-J_n\big(w^\star,\mathbf{t}\big)+\frac{1}{2}K_n\big(w^\star,\mathbf{t}\big)\Big)\le 0\Big)\longrightarrow 0,\qquad \text{as } n\to\infty.$$

Since $\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} K_n(w^\star,\mathbf{t})\ge(1-\gamma)\Delta_{\boldsymbol{\tau}^\star}\nu$ [21], equation (65),

$$P\Big(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})}\Big(J_n(\overline{w},\mathbf{t})-J_n\big(w^\star,\mathbf{t}\big)+\frac{1}{2}K_n\big(w^\star,\mathbf{t}\big)\Big)\le 0\Big)$$
$$\le P\Big(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})}\big(J_n(\overline{w},\mathbf{t})-J_n\big(w^\star,\mathbf{t}\big)\big)\le\frac{1}{2}(\gamma-1)\Delta_{\boldsymbol{\tau}^\star}\nu\Big),$$

and we conclude by Lemma A.2. □

Proof of Lemma A.6. For notational simplicity, $\widehat{\mathbf{t}}_n(z,\overline{\rho}_n)$ will be replaced by $\bar{\mathbf{t}}_n$ in this proof. Since for any $\nu>0$,

$$P\big(\big\|\bar{\mathbf{t}}_n-\mathbf{t}_n^\star\big\|_\infty<\nu\underline{\lambda}^{-2}\big)=P\big(\big\|\bar{\mathbf{t}}_n-\mathbf{t}_n^\star\big\|_\infty\le n\gamma\Delta_{\boldsymbol{\tau}^\star}\big)-P(\bar{\mathbf{t}}_n\in\mathcal{C}_{\nu,\gamma,n,m}),$$

it is enough, by Lemma A.3, to prove that

$$P(\bar{\mathbf{t}}_n\in\mathcal{C}_{\nu,\gamma,n,m})\longrightarrow 0,\qquad \text{as } n\to\infty,$$

for all $\nu>0$ and $0<\gamma<1/2$. Since

$$\mathcal{C}_{\nu,\gamma,n,m}=\bigcup_{\mathcal{I}\subset\{1,\dots,m\}}\mathcal{C}_{\nu,\gamma,n,m}\cap\big\{\mathbf{t}\in\mathcal{A}_{n,m};\forall k\in\mathcal{I},t_k\ge t_{n,k}^\star\big\},$$

we shall only study one set in the union without loss of generality and prove that

$$P\big(\bar{\mathbf{t}}_n\in\mathcal{C}'_{\nu,\gamma,n,m}\big)\longrightarrow 0,\qquad \text{as } n\to\infty,$$

where $\mathcal{C}'_{\nu,\gamma,n,m}$ is defined in (41). Since $\mathcal{C}'_{\nu,\gamma,n,m}=\bigcup_{\mathcal{I}\subset\{1,\dots,m\}}\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})$, we shall only study one set in the union without loss of generality and prove that

$$P\big(\bar{\mathbf{t}}_n\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})\big)\longrightarrow 0,\qquad \text{as } n\to\infty.$$

Since

$$P\left(\bar{\mathbf{t}}_n \in \mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})\right) \leq P\left(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} J_n(\overline{w},\mathbf{t}) \leq 0\right),$$

the proof is complete by Lemma A.5. □

Proof of Proposition 3.1. For notational simplicity, $\widehat{\boldsymbol{\delta}}_n(z,\overline{\rho}_n)$ will be replaced by $\overline{\boldsymbol{\delta}}_n$ in this proof. By Lemma A.6, the last result to show is

$$\left\|\overline{\boldsymbol{\delta}_n} - \boldsymbol{\delta}^\star\right\| = O_P\left(n^{-1/2}\right),$$

that is, for all k, $\overline{\delta}_{n,k} - \delta_k^\star = O_P(n^{-1/2})$. By (30) and (31),

$$\begin{aligned}
\overline{\delta}_{n,k} &= \frac{1}{\bar{t}_{n,k+1} - \bar{t}_{n,k}} \sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} \overline{w}_i \\
&= \frac{1}{n(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})} \left(\sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} w_i^\star + \left(\rho^\star - \overline{\rho}_n\right) \sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} z_{i-1}\right).
\end{aligned}$$

By the Cauchy–Schwarz inequality,

$$\left|\sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} z_{i-1}\right| \leq (\bar{t}_{n,k+1} - \bar{t}_{n,k})^{1/2}\left(z_{\bar{t}_{n,k}}^2 + \cdots + z_{\bar{t}_{n,k+1}-1}^2\right)^{1/2} \leq n^{1/2}\|Bz\| = O_P(n),$$

where the last equality comes from Lemma A.1. Hence by (13) and Lemma A.6,

$$\begin{aligned}
\overline{\delta}_{n,k} &= \frac{1}{n(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})} \sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} w_i^\star + O_P\left(n^{-1/2}\right) \\
&= \frac{1}{n(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})} \left(\sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} \mathbb{E}w_i^\star + \sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} \epsilon_i\right) + O_P\left(n^{-1/2}\right),
\end{aligned}$$

where the last equality comes from (28) and (30).

Let us now prove that

$$\frac{1}{n(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})} \sum_{i=\bar{t}_{n,k}+1}^{\bar{t}_{n,k+1}} \epsilon_i = O_P\left(n^{-1/2}\right). \tag{51}$$

By Lemma A.3, $n^{-1}(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})^{-1} = O_P(n^{-1})$. Moreover,

$$\sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}} \epsilon_i = \sum_{i=t^\star_{n,k}+1}^{t^\star_{n,k+1}} \epsilon_i \pm \sum_{i=\overline{t}_{n,k}+1}^{t^\star_{n,k}} \epsilon_i \pm \sum_{i=t^\star_{n,k+1}}^{\overline{t}_{n,k+1}+1} \epsilon_i. \tag{52}$$

By the Chebyshev inequality, the first term in the right-hand side of (52) is $O_P(n^{1/2})$. By using the Cauchy–Schwarz inequality, we get that the second term of (52) satisfies: $|\sum_{i=\overline{t}_{n,k}+1}^{t^\star_{n,k}} \epsilon_i| \leq |t^\star_{n,k} - \overline{t}_{n,k}|^{1/2}(\sum_{i=1}^{n} \epsilon_i^2)^{1/2} = O_P(1)O_P(n^{1/2}) = O_P(n^{1/2})$, by Lemma A.6. The same holds for the last term in the right-hand side of (52), which gives (51). Hence,

$$\begin{aligned}
\overline{\delta}_{n,k} - \delta^\star_k &= \frac{1}{n(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})} \sum_{i=\overline{t}_{n,k}+1}^{\overline{t}_{n,k+1}} \left(\mathbb{E}w^\star_i - \delta^\star_k\right) + O_P\left(n^{-1/2}\right) \\
&= \frac{1}{n(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})} \sum_{i\in\{\overline{t}_{n,k}+1,\ldots,\overline{t}_{n,k+1}\}\setminus\{t^\star_{n,k}+1,\ldots,t^\star_{n,k+1}\}} \left(\mathbb{E}w^\star_i - \delta^\star_k\right) + O_P\left(n^{-1/2}\right),
\end{aligned}$$

and then

$$\begin{aligned}
\left|\overline{\delta}_{n,k} - \delta^\star_k\right| \leq{}& \frac{1}{n(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})} \sharp\{\overline{t}_{n,k}+1,\ldots,\overline{t}_{n,k+1}\} \setminus \left\{t^\star_{n,k}+1,\ldots,t^\star_{n,k+1}\right\} \max_{l=0,\ldots,m} \left|\delta^\star_l - \delta^\star_k\right| \\
&+ O_P\left(n^{-1/2}\right).
\end{aligned}$$

We conclude by using Lemma A.6 to get $\sharp\{\overline{t}_{n,k}+1,\ldots,\overline{t}_{n,k+1}\} \setminus \{t^\star_{n,k}+1,\ldots,t^\star_{n,k+1}\} = O_P(1)$ and Lemma A.3 to get $(\overline{\tau}_{n,k+1} - \overline{\tau}_{n,k})^{-1} = O_P(1)$. □

A.4. Proof of Proposition 3.2

The connection between models (1) and (9) is made by the following lemmas.

Lemma A.7. *Let* $(y_0,\ldots,y_n)$ *be defined by* (1) *and let*

$$v^\star_i = y_i - \rho^\star y_{i-1}, \tag{53}$$

$$\Delta^\star_i = \begin{cases} -\rho^\star\left(\mu^\star_k - \mu^\star_{k-1}\right), & \text{if } i = t^\star_{n,k}+1, \\ 0, & \text{otherwise}, \end{cases} \tag{54}$$

where the $\mu^\star_k$*'s are defined in* (1), *then the process*

$$w^\star_i = v^\star_i + \Delta^\star_i \tag{55}$$

has the same distribution as $z_i - \rho^\star z_{i-1}$ *where* $(z_0, \dots, z_n)$ *is defined by* (9). *Such a process* $(z_0, \dots, z_n)$ *can be constructed recursively as*

$$\begin{cases} z_0 = y_0, \\ z_i = w_i^\star + \rho^\star z_{i-1}, \qquad \text{for } i > 0. \end{cases} \tag{56}$$

Lemma A.8. *Let* $(y_0, \dots, y_n)$ *be defined by* (1) *and let* z *be defined by* (53–56). *Then*

$$\overline{w}_i = \overline{v}_i + \overline{\Delta}_i, \tag{57}$$

where

$$\overline{v}_i = y_i - \overline{\rho}_n y_{i-1}, \tag{58}$$

$$\overline{w}_i = z_i - \overline{\rho}_n z_{i-1}, \tag{59}$$

$$\overline{\Delta}_i = \Delta_i^\star + \big(\rho^\star - \overline{\rho}_n\big)(z_{i-1} - y_{i-1}). \tag{60}$$

Lemma A.9. *Let* $\overline{\Delta} = (\overline{\Delta}_i)_{0 \le i \le n}$ *as defined in* (60). *Then* $\|\overline{\Delta}\| = O_P(1)$.

Proof of Lemma A.7. Let z being defined by (56). Using (55), we get, for all $0 \le k \le m$, $t^\star_{n,k} < i \le t^\star_{n,k+1}$,

$$\begin{aligned} \big(z_i - \mu_k^\star\big) - \rho^\star\big(z_{i-1} - \mu_k^\star\big) &= \big(y_i - \mu_k^\star\big) - \rho^\star\big(y_{i-1} - \mu_k^\star\big) + \Delta_i^\star \\ &= \begin{cases} \big(y_i - \mu_k^\star\big) - \rho^\star\big(y_{i-1} - \mu_{k-1}^\star\big), & \text{if } i = t^\star_{n,k} + 1, \\ \big(y_i - \mu_k^\star\big) - \rho^\star\big(y_{i-1} - \mu_k^\star\big), & \text{otherwise.} \end{cases} \end{aligned}$$

This expression equals $(y_i - \mathbb{E}(y_i)) - \rho^\star(y_{i-1} - \mathbb{E}(y_{i-1})) = \eta_i - \rho^\star \eta_{i-1} = \epsilon_i$ by (1) and (2). Then z satisfies (9). □

Proof of Lemma A.8. The proof of Lemma A.8 is straightforward. □

Proof of Lemma A.9. (60) can be written as

$$\overline{\Delta} = \Delta^\star + \big(\rho^\star - \overline{\rho}_n\big)(By - Bz),$$

where $\Delta^\star = (\Delta_i^\star)_{1 \le i \le n}$, $By = (y_{i-1})_{1 \le i \le n}$ and Bz is defined in (29). By the triangle inequality,

$$\|\overline{\Delta}\| \le \big\|\Delta^\star\big\| + \big|\rho^\star - \overline{\rho}_n\big|\big(\|By\| + \|Bz\|\big). \tag{61}$$

Since $\|\Delta^\star\|$ is constant it is bounded. The conclusion follows from (61), (13) and Lemma A.1. □

Proof of Proposition 3.2. Let y, z, $\overline{v}$, $\overline{w}$ and $\overline{\Delta}$ be defined in Lemma A.8.

Using (32) and Lemma A.8, we get

$$J_n(\overline{v}, \mathbf{t}) = J_n(\overline{w}, \mathbf{t}) + J_n(\overline{\Delta}, \mathbf{t}) - \frac{2}{n}\Big(\big\langle \pi_{E_{\mathbf{t}_n^\star}}(\overline{w}), \pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\big\rangle - \big\langle \pi_{E_{\mathbf{t}}}(\overline{w}), \pi_{E_{\mathbf{t}}}(\overline{\Delta})\big\rangle\Big).$$

By the Cauchy–Schwarz inequality and the 1-Lipschitz property of projections, we have

$$\left|J_n(\overline{\Delta},\mathbf{t})\right| \le \frac{2}{n}\|\overline{\Delta}\|^2,$$
$$\left|\langle \pi_{E_{\mathbf{t}_n^\star}}(\overline{w}), \pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle - \langle \pi_{E_{\mathbf{t}}}(\overline{w}), \pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right| \le 2\|\overline{\Delta}\|\|\overline{w}\|.$$

Note that $\overline{w} = z - \overline{\rho}_n Bz$ thus by the triangle inequality

$$\|\overline{w}\| \le \|z\| + |\overline{\rho}_n|\|Bz\|.$$

Since $|\overline{\rho}_n| = O_P(1)$, we deduce from Lemma A.1 that $\|\overline{w}\| = O_P(n^{1/2})$. Since, by Lemma A.9, $\|\overline{\Delta}\| = O_P(1)$, we obtain that

$$\sup_{\mathbf{t}}\left|J_n(\overline{\Delta},\mathbf{t}) - \frac{2}{n}\left(\langle \pi_{E_{\mathbf{t}_n^\star}}(\overline{w}), \pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle - \langle \pi_{E_{\mathbf{t}}}(\overline{w}), \pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right)\right| = O_P\left(n^{-1/2}\right). \tag{62}$$

For $0 < \nu < \Delta_{\boldsymbol{\tau}^\star}$, using (33) and (49), we get:

$$\begin{aligned}
P\left(\left\|\bar{\mathbf{t}}_n - \mathbf{t}^\star\right\|_\infty \ge \nu\right) &\le P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}} J_n(\overline{v},\mathbf{t}) \le 0\right)\\
&\le P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}}\left\{J_n(\overline{w},\mathbf{t}) + J_n(\overline{\Delta},\mathbf{t})\right.\right.\\
&\qquad \left.\left.- \frac{2}{n}\left(\langle \pi_{E_{\mathbf{t}_n^\star}}(\overline{w}), \pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle - \langle \pi_{E_{\mathbf{t}}}(\overline{w}), \pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right)\right\} \le 0\right)\\
&\le P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}}\left\{K_n(\overline{w},\mathbf{t}) + V_n(\overline{w},\mathbf{t}) + W_n(\overline{w},\mathbf{t}) + J_n(\overline{\Delta},\mathbf{t})\right.\right.\\
&\qquad \left.\left.- \frac{2}{n}\left(\langle \pi_{E_{\mathbf{t}_n^\star}}(\overline{w}), \pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle - \langle \pi_{E_{\mathbf{t}}}(\overline{w}), \pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right)\right\} \le 0\right)\\
&\le P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}}\left\{\frac{1}{2}K_n(\overline{w},\mathbf{t}) + V_n(\overline{w},\mathbf{t}) + W_n(\overline{w},\mathbf{t})\right\} \le 0\right)\\
&\qquad + P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}}\left\{\frac{1}{2}K_n(\overline{w},\mathbf{t}) + J_n(\overline{\Delta},\mathbf{t})\right.\right.\\
&\qquad \left.\left.- \frac{2}{n}\left(\langle \pi_{E_{\mathbf{t}_n^\star}}(\overline{w}), \pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle - \langle \pi_{E_{\mathbf{t}}}(\overline{w}), \pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right)\right\} \le 0\right).
\end{aligned}$$

Following the proof of Lemma A.3, one can prove that

$$P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}}\left\{\frac{1}{2}K_n(\overline{w},\mathbf{t}) + V_n(\overline{w},\mathbf{t}) + W_n(\overline{w},\mathbf{t})\right\} \le 0\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{63}$$

Using (46), we get that

$$P\left(\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}}\left\{\frac{1}{2}K_n(\overline{w},\mathbf{t})+J_n(\overline{\Delta},\mathbf{t})-\frac{2}{n}\left(\langle\pi_{E_{\mathbf{t}_n^\star}}(\overline{w}),\pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle-\langle\pi_{E_{\mathbf{t}}}(\overline{w}),\pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right)\right\}\leq 0\right)$$
$$\leq P\left(\frac{1}{2}\underline{\lambda}^2\nu+\min_{\mathbf{t}\in\mathcal{C}_{n,m,\nu}}\left\{J_n(\overline{\Delta},\mathbf{t})-\frac{2}{n}\left(\langle\pi_{E_{\mathbf{t}_n^\star}}(\overline{w}),\pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle-\langle\pi_{E_{\mathbf{t}}}(\overline{w}),\pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right)\right\}\leq 0\right) \tag{64}$$

which goes to zero when n goes to infinity by (62).

Then Lemma A.3 still holds if y is defined by (1). To show the rate of convergence, we use the same decomposition. As in the proof of Lemma A.6, $P(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} J_n(\overline{v},\mathbf{t})\leq 0)\longrightarrow_{n\to\infty} 0$ for all $\nu>0$ and $0<\gamma<1/2$ is a sufficient condition for proving that $P(\widehat{\mathbf{t}}_n(y,\overline{\rho}_n)\in\mathcal{C}_{\nu,\gamma,n,m})\longrightarrow_{n\to\infty} 0$, which allows us to conclude on the rate of convergence of the estimated change-points. Note that

$$P\left(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}(\mathcal{I})} J_n(\overline{v},\mathbf{t})\leq 0\right)\leq P\left(\min_{\mathbf{t}\in\mathcal{C}'_{\nu,\gamma,n,m}}\left\{\frac{1}{2}K_n(\overline{w},\mathbf{t})+V_n(\overline{w},\mathbf{t})+W_n(\overline{w},\mathbf{t})\right\}\leq 0\right)$$
$$+P\left(\frac{1}{2}\underline{\lambda}^2\nu+J_n(\overline{\Delta},\mathbf{t})\right.$$
$$\left.-\frac{2}{n}\left(\langle\pi_{E_{\mathbf{t}_n^\star}}(\overline{w}),\pi_{E_{\mathbf{t}_n^\star}}(\overline{\Delta})\rangle-\langle\pi_{E_{\mathbf{t}}}(\overline{w}),\pi_{E_{\mathbf{t}}}(\overline{\Delta})\rangle\right)\leq 0\right).$$

In the latter equation, the second term of the right-hand side goes to zero as n goes to infinity by (62).

The first term of right-hand side goes to zero when n goes to infinity by following the same line of reasoning as the one of Lemma A.5. This concludes the proof of Proposition 3.2. □

A.5. Proof of Proposition 4.1

We shall used in this section the notations introduced in Sections A.3 and 4.1. The result derives directly from Lemmas A.10 and A.11.

Lemma A.10. *Under the assumptions of Proposition* 4.1, $P(\widehat{m}=m)\longrightarrow_{n\to\infty} 0$ *if* $m<m^\star$.

Lemma A.11. *Under the assumptions of Proposition* 4.1, $P(\widehat{m}=m)\longrightarrow_{n\to\infty} 0$ *if* $m>m^\star$.

Proof of Lemma A.10. If $\widehat{m}=m<m^\star$, then

$$\frac{1}{n}SS_m(z,\overline{\rho}_n)+\beta_n m\leq\frac{1}{n}SS_{m^\star}(z,\overline{\rho}_n)+\beta_n m^\star,$$

where SS_m is defined in (15). In particular, there exists $\mathbf{t}\in\mathcal{A}_{n,m}$ such that

$$\frac{1}{n}\min_{\boldsymbol{\delta}} SS_m(z,\overline{\rho}_n,\boldsymbol{\delta},\mathbf{t})+\beta_n m\leq\frac{1}{n}\min_{\boldsymbol{\delta}} SS_m\left(z,\overline{\rho}_n,\boldsymbol{\delta},\mathbf{t}_n^\star\right)+\beta_n m^\star.$$

From (32), we get

$$J_n(\overline{w}, \mathbf{t}) \le \beta_n\big(m^\star - m\big).$$

Since (β_n) converges to zero, for any $\varepsilon > 0$, $\beta_n(m^\star - m) \le \varepsilon$ for a large enough n, and so

$$J_n(\overline{w}, \mathbf{t}) \le \varepsilon.$$

One can check that there exist $0 < \nu < \Delta_{\boldsymbol{\tau}^\star}$ such that, for a large enough n, there exists $\mathbf{t}' \in \mathcal{C}_{n,m^\star,\nu}$ such that $E_{\mathbf{t}} \subset E_{\mathbf{t}'}$ (that is the change-points of $\mathbf{t}$ are change-points of $\mathbf{t}'$) for all $\mathbf{t} \in \mathcal{A}_{n,m}$, where $\mathcal{C}_{n,m^\star,\nu}$ is defined in (49). From (32) and $E_{\mathbf{t}} \subset E_{\mathbf{t}'}$, we get $J_n(\overline{w}, \mathbf{t}') \le J_n(\overline{w}, \mathbf{t})$. Then, the following inequality holds for all $\varepsilon > 0$ and any large enough n:

$$P(\widehat{m} = m) \le P\big(\exists \mathbf{t}' \in \mathcal{C}_{n,m^\star,\nu}, J_n\big(\overline{w}, \mathbf{t}'\big) \le \varepsilon\big). \tag{65}$$

We then follow the steps of (50), $-\nu\underline{\lambda}^2$ being replaced by $\varepsilon - \nu\underline{\lambda}^2$. The convergence of $P(\exists \mathbf{t}' \in \mathcal{C}_{n,m^\star,\nu}, J_n(\overline{w}, \mathbf{t}') \le \varepsilon)$ to zero holds with $\varepsilon < \nu\underline{\lambda}^2$. We can conclude with (65). □

Proof of Lemma A.11. Following the proof of Lemma A.10, if $\widehat{m} = m > m^\star$, there exists $\mathbf{t} \in \mathcal{A}_{n,m}$ such that $J_n(\overline{w}, \mathbf{t}) \le \beta_n(m^\star - m)$ and then $J_n(\overline{w}, \mathbf{t}) + \beta_n \le 0$ since $m > m^\star$. Then

$$P(\widehat{m} = m) \le P\big(\exists \mathbf{t} \in \mathcal{A}_{n,m}, J_n(\overline{w}, \mathbf{t}) + \beta_n \le 0\big). \tag{66}$$

Adding the change-points of $\mathbf{t}_n^\star$ to those of such a $\mathbf{t}$, one can get $t' \in \mathcal{A}_{n,m'}$ with $m^\star < m \le m' \le m + m^\star$ such that $E_{\mathbf{t}} \cup E_{\mathbf{t}_n^\star} \subset E_{\mathbf{t}'}$, provided that $(m + m^\star)\lceil \Delta_n \rceil \le n$, where $\lceil \cdot \rceil$ is the ceiling function, this condition being fulfilled for any sufficiently large n under the assumptions of Proposition 4.1 since $n^{-1}\Delta_n$ converges to zero. Since $E_{\mathbf{t}} \subset E_{\mathbf{t}'}$, we derive $J_n(\overline{w}, \mathbf{t}') + \beta_n \le J_n(\overline{w}, \mathbf{t}) + \beta_n$ from (32). Then, from (66), we get

$$\forall m' > m^\star, \qquad P\big(\exists \mathbf{t}' \in \mathcal{A}_{n,m'}, E_{\mathbf{t}_n^\star} \subset E_{\mathbf{t}'}, J_n\big(\overline{w}, \mathbf{t}'\big) + \beta_n \le 0\big) \underset{n\to\infty}{\longrightarrow} 0 \tag{67}$$

is a sufficient condition to prove the lemma. Let us prove (67). Let $m' > m^\star$ and such a $\mathbf{t}'$. We compare $J_n(\overline{w}, \mathbf{t}')$ to $J_n(w^\star, \mathbf{t}')$. Since $\mathbb{E}w^\star \in E_{\mathbf{t}_n^\star} \subset E_{\mathbf{t}'}$, $K_n(w^\star, \mathbf{t}') = 0$ by (34). By (36) and $\mathbb{E}w^\star \in E_{\mathbf{t}_n^\star} \subset E_{\mathbf{t}'}$,

$$\begin{aligned}
W_n\big(w^\star, \mathbf{t}'\big) &= \frac{2}{n}\big(\big\langle \pi_{E_{\mathbf{t}_n^\star}}\big(w^\star - \mathbb{E}w^\star\big), \pi_{E_{\mathbf{t}_n^\star}}\big(\mathbb{E}w^\star\big)\big\rangle - \big\langle \pi_{E_{\mathbf{t}'}}\big(w^\star - \mathbb{E}w^\star\big), \pi_{E_{\mathbf{t}'}}\big(\mathbb{E}w^\star\big)\big\rangle\big) \\
&= \frac{2}{n}\big\langle \pi_{E_{\mathbf{t}_n^\star}}\big(w^\star - \mathbb{E}w^\star\big) - \pi_{E_{\mathbf{t}'}}\big(w^\star - \mathbb{E}w^\star\big), \pi_{E_{\mathbf{t}_n^\star}}\big(\mathbb{E}w^\star\big)\big\rangle \\
&= -\frac{2}{n}\big\langle \pi_{E^\perp_{\mathbf{t}_n^\star}}\pi_{E_{\mathbf{t}'}}\big(w^\star - \mathbb{E}w^\star\big), \pi_{E_{\mathbf{t}_n^\star}}\big(\mathbb{E}w^\star\big)\big\rangle \\
&= 0,
\end{aligned}$$

where $E^\perp$ is the (Euclidean) orthogonal complement of the vector subspace E. Then $J_n(w^\star, \mathbf{t}') = V_n(w^\star, \mathbf{t}')$ and

$$J_n\big(\overline{w}, \mathbf{t}'\big) = V_n\big(w^\star, \mathbf{t}'\big) + \big(J_n\big(\overline{w}, \mathbf{t}'\big) - J_n\big(w^\star, \mathbf{t}'\big)\big). \tag{68}$$

Using (47), $V_n(w^\star, \mathbf{t}) \geq -\frac{2(m'+1)}{n\Delta_n} M_n$, where

$$M_n = M_{n,1} + M_{n,2},$$
$$M_{n,1} = \max_{1\leq s\leq n}\left(\sum_{i=1}^{s}\epsilon_i\right)^2,$$
$$M_{n,2} = \max_{1\leq s\leq n}\left(\sum_{i=n-s}^{n}\epsilon_i\right)^2.$$

We define $D_n = \sup_{\mathbf{t}'\in\mathcal{A}_{n,m'}} |J_n(\overline{w}, \mathbf{t}') - J_n(w^\star, \mathbf{t}')|$. Then, using (68),

$$J_n\big(\overline{w}, \mathbf{t}'\big) \geq -\frac{2(m+1)}{n\Delta_n} M_n - D_n,$$

which implies

$$P\big(\exists \mathbf{t}' \in \mathcal{A}_{n,m'}, E_{\mathbf{t}_n^\star} \subset E_{\mathbf{t}'}, J_n\big(\overline{w}, \mathbf{t}'\big) + \beta_n \leq 0\big) \leq P\left(-\frac{2(m'+1)}{n\Delta_n} M_n - D_n + \beta_n \leq 0\right)$$
$$\leq P\left(\frac{2(m'+1)}{n\Delta_n} M_n \geq \frac{\beta_n}{2}\right) + P\left(D_n \geq \frac{\beta_n}{2}\right).$$

By Lemma A.2, $D_n = O_P(n^{-1/2})$ and then $P(D_n \geq \frac{\beta_n}{2})$ tends to zero as n tends to infinity since $n^{1/2}\beta_n \longrightarrow_{n\to\infty} +\infty$. Let us now prove that $P(\frac{2(m+1)}{n\Delta_n} M_n \geq \frac{\beta_n}{2})$ tends to zero as n tends to infinity, which concludes the proof. Note that

$$P\left(\frac{2(m'+1)}{n\Delta_n} M_n \geq \frac{\beta_n}{2}\right) \leq P\left(M_{n,1} \geq \frac{n\Delta_n\beta_n}{8(m'+1)}\right) + P\left(M_{n,2} \geq \frac{n\Delta_n\beta_n}{8(m'+1)}\right).$$

We prove the convergence for each term in the right-hand side of the above equation. We shall prove it for the first term in the right-hand side since the arguments for the other term are the same. From Kolmogorov's maximal inequality [13], Theorem 2.5.2, since $(\epsilon_i)_{i\geq 0}$ is a sequence of independent r.v.'s with zero-mean and finite variance $\sigma^{\star 2}$,

$$\forall \delta > 0, \qquad P\big(M_{n,1} \geq \delta^2\big) \leq \frac{n\sigma^{\star 2}}{\delta^2}. \tag{69}$$

Letting $\delta^2 = \frac{n\Delta_n\beta_n}{8(m'+1)}$ in (69), we get

$$P\left(M_{n,1} \geq \frac{n\Delta_n\beta_n}{8(m'+1)}\right) \leq \frac{8(m'+1)\sigma^{\star 2}}{\Delta_n\beta_n},$$

which goes to 0 as n tends to infinity because $\Delta_n\beta_n \longrightarrow_{n\to\infty} +\infty$. The proof of the convergence of $P(M_{n,2} \geq \frac{n\Delta_n\beta_n}{8(m'+1)})$ follows the same lines. □

A.6. Proof of Proposition 4.2

The proof directly derives from the following lemmas.

Lemma A.12. *Under the assumptions of Proposition 4.2,* $P(\widehat{m}=m)\longrightarrow_{n\to\infty} 0$ *if* $m<m^\star$.

Lemma A.13. *Under the assumptions of Proposition 4.2,* $P(\widehat{m}=m)\longrightarrow_{n\to\infty} 0$ *if* $m>m^\star$.

Proof of Lemma A.12. Following the proof of Lemma A.10 and replacing $\overline{w}$ by $\overline{v}$, we get, for any $\varepsilon>0$,

$$P(\widehat{m}=m)\le P\big(\exists \mathbf{t}'\in\mathcal{C}_{n,m^\star,\nu},\ J_n\big(\overline{v},\mathbf{t}'\big)\le\varepsilon\big) \tag{70}$$

$$\begin{aligned}&\le P\left(\exists \mathbf{t}'\in\mathcal{C}_{n,m^\star,\nu},\ \frac{1}{2}K_n\big(\overline{w},\mathbf{t}'\big)+V_n\big(\overline{w},\mathbf{t}'\big)+W_n\big(\overline{w},\mathbf{t}'\big)\le\frac{\varepsilon}{2}\right)\\&\quad+P\left(\exists \mathbf{t}'\in\mathcal{C}_{n,m^\star,\nu},\ \frac{1}{2}K_n\big(\overline{w},\mathbf{t}'\big)+J_n\big(\overline{v},\mathbf{t}'\big)-J_n\big(\overline{w},\mathbf{t}'\big)\le\frac{\varepsilon}{2}\right),\end{aligned} \tag{71}$$

since

$$J_n\big(\overline{v},\mathbf{t}'\big)=\tfrac{1}{2}K_n\big(\overline{w},\mathbf{t}'\big)+V_n\big(\overline{w},\mathbf{t}'\big)+W_n\big(\overline{w},\mathbf{t}'\big)+\tfrac{1}{2}K_n\big(\overline{w},\mathbf{t}'\big)+J_n\big(\overline{v},\mathbf{t}'\big)-J_n\big(\overline{w},\mathbf{t}'\big).$$

From (63) and (71), it suffices to prove that

$$P\left(\exists \mathbf{t}'\in\mathcal{C}_{n,m^\star,\nu},\ \frac{1}{2}K_n\big(\overline{w},\mathbf{t}'\big)+J_n\big(\overline{v},\mathbf{t}'\big)-J_n\big(\overline{w},\mathbf{t}'\big)\le\frac{\varepsilon}{2}\right)\underset{n\to\infty}{\longrightarrow}0$$

to conclude the proof. It follows from (62) and (64), $\frac{1}{2}\underline{\lambda}^2\nu$ being replaced by $\frac{1}{2}(\underline{\lambda}^2\nu-\varepsilon)$, which is positive if $\varepsilon<\underline{\lambda}^2\nu$. □

Proof of Lemma A.13. As in the proof of Lemma A.11, it suffices to show that

$$P\big(\exists\mathbf{t}\in\mathcal{A}_{n,m},\ J_n(\overline{v},\mathbf{t})+\beta_n\le 0\big)\underset{n\to\infty}{\longrightarrow}0.$$

Since

$$J_n(\overline{v},\mathbf{t})\ge J_n(\overline{w},\mathbf{t})-\sup_t\big|J_n(\overline{v},\mathbf{t})-J_n(\overline{w},\mathbf{t})\big|,$$

the result follows from

$$P\left(\exists\mathbf{t}\in\mathcal{A}_{n,m},\ J_n(\overline{w},\mathbf{t})+\frac{1}{2}\beta_n\le 0\right)\underset{n\to\infty}{\longrightarrow}0, \tag{72}$$

$$P\left(\sup_t\big|J_n(\overline{v},\mathbf{t})-J_n(\overline{w},\mathbf{t})\big|\ge\frac{1}{2}\beta_n\right)\underset{n\to\infty}{\longrightarrow}0. \tag{73}$$

Equation (72) follows from the Proof of Lemma A.11, replacing β_n by $\frac{1}{2}\beta_n$. Equation (73) follows from (62) and from $n^{1/2}\beta_n \longrightarrow_{n\to\infty} +\infty$. □

A.7. Proof of Proposition 4.3

We first give some lemmas which are useful for the proof of Proposition 4.3.

Lemma A.14. *Under the assumptions of Proposition 4.3 with SS_m given by (15), we have, for any positive m,*

$$SS_m(z, \overline{\rho}_n) = SS_m\big(z, \rho^\star\big) + O_P(1), \qquad \text{as } n \to \infty.$$

Lemma A.15. *Under the assumptions of Proposition 4.3 with SS_m given by (15), we have, for any positive m,*

$$SS_m\big(z, \rho^\star\big)^{-1} = O_P\big(n^{-1}\big), \qquad \text{as } n \to \infty.$$

Proof of Lemma A.14. The proof of this lemma follows exactly this of Lemma A.16. The difference is that, in (9), the term $\Delta^\star$ appearing in the decomposition (77) vanishes. □

Proof of Lemma A.15. We first define

$$SS_m(z, \rho, \mathbf{t}) = \arg\min_{\delta} SS_m(z, \rho, \delta, \mathbf{t}).$$

We have, for any positive M,

$$\begin{aligned}
P\left(\frac{n}{SS_m(z, \rho^\star)} > M\right) &\le P\left(\left\{\frac{SS_m(z, \rho^\star)}{SS_m(z, \rho^\star, \mathbf{t}^\star)} > 1\right\} \cap \left\{\frac{n}{SS_m(z, \rho^\star)} > M\right\}\right) \\
&\quad + P\left(\left\{\frac{SS_m(z, \rho^\star)}{SS_m(z, \rho^\star, \mathbf{t}^\star)} < 1\right\} \cap \left\{\frac{n}{SS_m(z, \rho^\star)} > M\right\}\right) \\
&\le P\left(\frac{n}{SS_m(z, \rho^\star, \mathbf{t}^\star)} > M\right) + P\left(\frac{SS_m(z, \rho^\star)}{SS_m(z, \rho^\star, \mathbf{t}^\star)} < 1\right).
\end{aligned}$$

Under the assumptions of Proposition 3.1, a by product of the proof of Theorem 3 in [21] is that

$$P\left(\frac{SS_m(z, \rho^\star)}{SS_m(z, \rho^\star, \mathbf{t}^\star)} < 1\right) = P\big(SS_m\big(z, \rho^\star\big) - SS_m\big(z, \rho^\star, \mathbf{t}^\star\big) < 0\big) \le \kappa n^{-\alpha},$$

where κ is a positive constant depending on $\boldsymbol{\delta}^\star$ and $\mathbf{t}^\star$, and α is a positive constant. Furthermore, as $\sigma^{\star-2} SS_m(z, \rho^\star, \mathbf{t}^\star)$ has a χ^2_{n-m-1} distribution, $n^{-1} SS_m(z, \rho^\star, \mathbf{t}^\star) = \sigma^{\star 2} + o_P(1)$ and thus $n^{-1} SS_m(z, \rho^\star, \mathbf{t}^\star) = O_P(1)$, which concludes the proof. □

Proof of Proposition 4.3. We have to prove that, for a given positive m, $C_m(z,\rho^\star) - C_m(z,\overline{\rho}_n) = O_P(1)$. Observe that, since $\widehat{\tau}_k(z,\rho) = \widehat{t}_k(z,\rho)/n$,

$$\begin{aligned}\sum_{k=0}^{m}\log n_k\big(\widehat{t}(z,\overline{\rho}_n)\big) - \sum_{k=0}^{m}\log n_k\big(\widehat{t}\big(z,\rho^\star\big)\big) \\ = \sum_{k=0}^{m}\log\big(\widehat{\tau}_{k+1}(z,\overline{\rho}_n) - \widehat{\tau}_k(z,\overline{\rho}_n)\big) - \sum_{k=0}^{m}\log\big(\widehat{\tau}_{k+1}\big(z,\rho^\star\big) - \widehat{\tau}_k\big(z,\rho^\star\big)\big).\end{aligned} \tag{74}$$

By Proposition 3.1, both quantities of the previous equation converge in probability to

$$\sum_{k=0}^{m}\log\big(\tau_{k+1}^\star - \tau_k^\star\big)$$

thus

$$\sum_{k=0}^{m}\log n_k\big(\widehat{t}(z,\overline{\rho}_n)\big) - \sum_{k=0}^{m}\log n_k\big(\widehat{t}\big(z,\rho^\star\big)\big) = O_P(1). \tag{75}$$

Further note that

$$\log SS_m(z,\overline{\rho}_n) - \log SS_m\big(z,\rho^\star\big) = \log\left(\frac{SS_m(z,\overline{\rho}_n)}{SS_m(z,\rho^\star)}\right) = R\left(\frac{SS_m(z,\overline{\rho}_n) - SS_m(z,\rho^\star)}{SS_m(z,\rho^\star)}\right),$$

where $R(x) = \log(1+x)$. Lemma A.14 states that $SS_m(z,\overline{\rho}_n) - SS_m(z,\rho^\star) = O_P(1)$ and Lemma A.15 that $[SS_m(z,\rho^\star)]^{-1} = O_P(n^{-1})$ so, by [34], Lemma 2.12, we get that

$$\log SS_m(z,\overline{\rho}_n) - \log SS_m\big(z,\rho^\star\big) = O_P\big(n^{-1}\big).$$

Hence

$$\frac{n-m+1}{2}\log SS_m(z,\overline{\rho}_n) - \frac{n-m+1}{2}\log SS_m\big(z,\rho^\star\big) = O_P(1),$$

which with (75) concludes the proof of Proposition 4.3. □

A.8. Proof of Proposition 4.4

We first give some lemmas which are useful for the proof of Proposition 4.4.

Lemma A.16. *Under the assumptions of Proposition 4.3 with SS_m given by (15), we have, for any positive m,*

$$SS_m(y,\overline{\rho}_n) = SS_m\big(y,\rho^\star\big) + O_P(1), \qquad \text{as } n\to\infty.$$

Lemma A.17. *If $(y_0,\ldots,y_n)$ is defined by (1) and $(z_0,\ldots,z_n)$ is defined as in Lemma A.7, then*

$$SS_m\big(y,\rho^\star\big) = SS_m\big(z,\rho^\star\big) + O_P(1), \qquad \text{as } n\to\infty.$$

Lemma A.18. *Let* (X_n) *and* (Y_n) *be two sequences of r.v.'s such that* $X_n - Y_n = O_P(1)$. *If* $Y_n^{-1} = O_P(n^{-1})$ *then* $X_n^{-1} = O_P(n^{-1})$.

Proof of Lemma A.16. Using the matrix notations from the proof of Lemma A.9, we have

$$SS_m(y, \rho^\star) = \min_{T,\delta} \| y - \rho^\star By - T\delta \|^2, \qquad SS_m(y, \overline{\rho}_n) = \min_{T,\delta} \| y - \overline{\rho}_n By - T\delta \|^2,$$

where all minimizations are achieved over all segmentations with m change points belonging to $\mathcal{A}_{n,m}$. Let us define $(\widehat{T}^\star, \widehat{\delta}^\star)$ and $(\overline{T}, \overline{\delta})$ by

$$(\widehat{T}^\star, \widehat{\delta}^\star) = \arg\min_{T,\delta} \| y - \rho^\star By - T\delta \|, \qquad (\overline{T}, \overline{\delta}) = \arg\min_{T,\delta} \| y - \overline{\rho}_n By - T\delta \|.$$

Note that $\widehat{T}^\star$ and $\overline{T}$ refer to $\widehat{t}(y, \rho^\star)$ and $\widehat{t}(y, \overline{\rho}_n)$, respectively. We have

$$\begin{aligned} \left| SS_m(y, \overline{\rho}_n) - SS_m(y, \rho^\star) \right| &= \left| \min_{T,\delta} \| y - \overline{\rho}_n By - T\delta \|^2 - \min_{T,\delta} \| y - \rho^\star By - T\delta \|^2 \right| \\ &\leq \max\Big(\big| \| y - \overline{\rho}_n By - \widehat{T}^\star \widehat{\delta}^\star \|^2 - \| y - \rho^\star By - \widehat{T}^\star \widehat{\delta}^\star \|^2 \big|, \\ &\quad \big| \| y - \overline{\rho}_n By - \overline{T\delta} \|^2 - \| y - \rho^\star By - \overline{T\delta} \|^2 \big| \Big). \end{aligned} \tag{76}$$

We now have to prove that this upper bound is $O_P(1)$. We first prove it for the second term of in the right-hand side of (76). To do so, observe that $\| y - \overline{\rho}_n By - \overline{T\delta} \|^2 = \| y - \rho^\star By - \overline{T\delta} + (\rho^\star - \overline{\rho}_n) By \|^2$. Thus,

$$\| y - \overline{\rho}_n By - \overline{T\delta} \|^2 - \| y - \rho^\star By - \overline{T\delta} \|^2 = (\overline{\rho}_n - \rho^\star)^2 \| By \|^2 + 2(\rho^\star - \overline{\rho}_n) \langle By, y - \rho^\star By - \overline{T\delta} \rangle.$$

Since, by (28) and Lemma A.7, $y - \rho^\star By - \overline{T\delta} = \epsilon - \Delta^\star + (T^\star \delta^\star - \overline{T\delta}) = \epsilon - \Delta^\star + T^\star(\delta^\star - \overline{\delta}) + (T^\star - \overline{T})\overline{\delta}$, where $\Delta^\star$ is the n-dimensional vector with entries $\Delta_i^\star$, we get

$$\begin{aligned} &\| y - \overline{\rho}_n By - \overline{T\delta} \|^2 - \| y - \rho^\star By - \overline{T\delta} \|^2 \\ &\quad = (\overline{\rho}_n - \rho^\star)^2 \| By \|^2 + 2(\rho^\star - \overline{\rho}_n) \big(\langle By, \epsilon \rangle + \langle By, T^\star(\delta^\star - \overline{\delta}) \rangle \\ &\qquad + \langle By, (T^\star - \overline{T})\overline{\delta} \rangle - \langle By, \Delta^\star \rangle \big). \end{aligned} \tag{77}$$

Let us now prove that each term in the right-hand side of (77) is $O_P(1)$.

(i) Let us study the first term of (77). Using Lemma A.1 and (13), we get that

$$(\overline{\rho}_n - \rho^\star)^2 \| By \|^2 = O_P(1). \tag{78}$$

(ii) Let us now study the second term of (77). Observe that $\langle By, \epsilon \rangle = \sum_{i=1}^n y_{i-1}\epsilon_i = \sum_{i=1}^n (y_{i-1} - \mathbb{E}(y_{i-1}))\epsilon_i + \sum_{i=1}^n \mathbb{E}(y_{i-1})\epsilon_i$. By using the central limit theorem for i.i.d. r.v.'s and since there is a finite number of change-points, the second term is $O_P(\sqrt{n})$. As for the first term, since $(y_{i-1} - \mathbb{E}(y_{i-1}))$ is a causal AR(1) process, then by using the beginning of the proof of Proposition 8.10.1 of [10], we get that $\sum_{i=1}^{n}(y_{i-1} - \mathbb{E}(y_{i-1}))\epsilon_i = O_P(\sqrt{n})$. Thus,

$$\langle By, \epsilon\rangle = O_P(\sqrt{n}). \tag{79}$$

Furthermore, we have $\|T^\star(\delta^\star - \overline{\delta})\|^2 = \sum_{k=0}^{m}(t_{k+1}^\star - t_k^\star)(\delta_k^\star - \overline{\delta}_k)^2$ where each term of the sum is $O_P(1)$, thanks to Proposition 3.2, and so is the sum. Now using Lemma A.1 and the Cauchy–Schwarz inequality, we get

$$\langle By, T^\star(\delta^\star - \overline{\delta})\rangle = O_P(\sqrt{n}). \tag{80}$$

The convergence rate of $\widehat{t}(y, \overline{\rho}_n)$ given in Proposition 3.2 ensures that, for any $\varepsilon > 0$ there exists a positive M such that each column of $(T^\star - \overline{T})$ has at most M non-zero coefficients with probability greater than $1 - \varepsilon$. By using Proposition 3.2, we obtain that with probability greater than $1 - \varepsilon$

$$\|(T^\star - \overline{T})\overline{\delta}\|^2 \le M\sum_k \overline{\delta}_k^2 = 2M\sum_k(\overline{\delta}_k - \delta_k^\star)^2 + 2M\sum_k \delta_k^{\star 2} \le MM', \tag{81}$$

where M' is a positive constant. By the Cauchy–Schwarz inequality, (81) and Lemma A.1, we get

$$\langle By, (T^\star - \overline{T})\overline{\delta}\rangle = O_P(\sqrt{n}). \tag{82}$$

As $\Delta^\star$ has only m non-zero entries, $\langle By, \Delta^\star\rangle$ is the sum of m Gaussian r.v.'s and is therefore $O_P(1)$.

Thus, combining (79), (80) and (82) with (13), we get

$$(\rho^\star - \overline{\rho}_n)(\langle By, \epsilon\rangle + \langle By, T^\star(\delta^\star - \overline{\delta})\rangle + \langle By, (T^\star - \overline{T})\overline{\delta}\rangle - \langle By, \Delta^\star\rangle) = O_P(1).$$

To complete the proof, we need to consider the first term of (76). As $\rho^\star$ satisfies the same assumptions as $\overline{\rho}_n$, using the same line of reasoning as for the second term holds so we get

$$\|y - \overline{\rho}_n By - \widehat{T}^\star\widehat{\delta}^\star\|^2 - \|y - \rho^\star By - \widehat{T}^\star\widehat{\delta}^\star\|^2 = O_P(1). \qquad \square$$

Proof of Lemma A.17. The proof follows the same line of reasoning as the proof of Lemma A.16.

Let us define $(\widehat{T}^y, \widehat{\delta}^y)$ and $(\widehat{T}^z, \widehat{\delta}^z)$ by

$$(\widehat{T}^y, \widehat{\delta}^y) = \arg\min_{T,\delta}\|y - \rho^\star By - T\delta\|^2, \qquad (\widehat{T}^z, \widehat{\delta}^z) = \arg\min_{T,\delta}\|z - \rho^\star Bz - T\delta\|^2.$$

We have

$$\begin{aligned}|SS_m(y, \rho^\star) - SS_m(z, \rho^\star)\| \le \max\Big(&\big|\|y - \rho^\star By - \widehat{T}^y\widehat{\delta}^y\|^2 - \|z - \rho^\star Bz - \widehat{T}^y\widehat{\delta}^y\|^2\big|,\\ &\big|\|y - \rho^\star By - \widehat{T}^z\widehat{\delta}^z\|^2 - \|z - \rho^\star Bz - \widehat{T}^z\widehat{\delta}^z\|^2\big|\Big).\end{aligned}$$

According to Lemma A.7, we have $y - \rho^\star By = z - \rho^\star Bz - \Delta^\star$ where $\Delta^\star = (\Delta_i^\star)$. As for the first term

$$\begin{aligned}&\|y - \rho^\star By - \widehat{T}^y\widehat{\delta}^y\|^2 - \|z - \rho^\star Bz - \widehat{T}^y\widehat{\delta}^y\|^2\\&= \|\Delta^\star\|^2 - 2\big(\langle\Delta^\star, \epsilon\rangle + \langle\Delta^\star, T^\star(\delta^\star - \widehat{\delta}^y)\rangle + \langle\Delta^\star, (T^\star - \widehat{T}^y)\widehat{\delta}^y\rangle\big),\end{aligned}$$

the first term of which is a constant and all other terms being $O_P(1)$, which can be proved following the same line as the proof of Lemma A.16. The control of $\|y - \rho^\star By - \widehat{T}^z\widehat{\delta}^z\|^2 - \|z - \rho^\star Bz - \widehat{T}^z\widehat{\delta}^z\|^2$ follows the same lines. □

Proof of Lemma A.18. Observe that

$$X_n^{-1} = \big(Y_n + (X_n - Y_n)\big)^{-1} = Y_n^{-1}\big(1 + Y_n^{-1}(X_n - Y_n)\big)^{-1}.$$

Since, by assumption, $Y_n^{-1}(X_n - Y_n) = O_P(n^{-1})$, the terms inside the parentheses converges in probability to one. Thus, $(1 + Y_n^{-1}(X_n - Y_n))^{-1}$ is in particular $O_P(1)$ which concludes the proof. □

Proof of Proposition 4.4. As for the proof of Proposition 4.3, denoting $\widehat{\tau}_k(y,\rho) = \widehat{t}_k(y,\rho)/n$, the decomposition (74) still holds, replacing z with y. Then, by Proposition 3.2, we have

$$\sum_{k=0}^{m}\log n_k\big(\widehat{t}(y,\overline{\rho}_n)\big) - \sum_{k=0}^{m}\log n_k\big(\widehat{t}(y,\rho^\star)\big) = O_P(1).$$

For a process y under model (1), we construct a process z under model (9) using Lemma A.7. The proof relies on the fact that y inherits some properties of z. Again, we note that

$$\log SS_m(y,\overline{\rho}_n) - \log SS_m\big(y,\rho^\star\big) = R\left(\frac{SS_m(y,\overline{\rho}_n) - SS_m(y,\rho^\star)}{SS_m(y,\rho^\star)}\right).$$

Lemma A.16 states that $SS_m(y,\overline{\rho}_n) - SS_m(y,\rho^\star) = O_P(1)$. To conclude the proof we need to further show that $[SS_m(y,\rho^\star)]^{-1} = O_P(n^{-1})$. We first show that $[SS_m(y,\rho^\star) - SS_m(z,\rho^\star)] = O_P(1)$ in Lemma A.17 and, because $[SS_m(z,\rho^\star)]^{-1} = O_P(n^{-1})$, we conclude using Lemma A.18. □

References

[1] Arcones, M.A. (1994). Limit theorems for nonlinear functionals of a stationary Gaussian sequence of vectors. *Ann. Probab.* **22** 2242–2274. MR1331224

[2] Auger, I.E. and Lawrence, C.E. (1989). Algorithms for the optimal identification of segment neighborhoods. *Bull. Math. Biol.* **51** 39–54. MR0978902

[3] Bai, J. and Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66** 47–78. MR1616121

[4] Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *J. Appl. Econometrics* **18** 1–22.

[5] Bardet, J.-M., Kengne, W.C. and Wintenberger, O. (2012). Multiple breaks detection in general causal time series using penalized quasi-likelihood. *Electron. J. Stat.* **6** 435–477.

[6] Basseville, M. and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Application. Prentice Hall Information and System Sciences Series*. Englewood Cliffs, NJ: Prentice Hall. MR1210954

[7] Boysen, L., Kempe, A., Liebscher, V., Munk, A. and Wittich, O. (2009). Consistencies and rates of convergence of jump-penalized least squares estimators. *Ann. Statist.* **37** 157–183. MR2488348

[8] Braun, J.V., Braun, R.K. and Müller, H.-G. (2000). Multiple changepoint fitting via quasilikelihood, with application to DNA sequence segmentation. *Biometrika* **87** 301–314. MR1782480

[9] Braun, J.V. and Müller, H.-G. (1998). Statistical methods for DNA sequence segmentation. *Statist. Sci.* **13** 142–162.

[10] Brockwell, P.J. and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd ed. *Springer Series in Statistics*. New York: Springer. MR1093459

[11] Csörgő, S. and Mielniczuk, J. (1996). The empirical process of a short-range dependent stationary sequence under Gaussian subordination. *Probab. Theory Related Fields* **104** 15–25. MR1367664

[12] Davis, R.A., Lee, T.C.M. and Rodriguez-Yam, G.A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041

[13] Durrett, R. (2010). *Probability: Theory and Examples*, 4th ed. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: Cambridge Univ. Press. MR2722836

[14] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, *Vol.* 2, 2nd ed. New York: Wiley.

[15] Gazeaux, J., Williams, S., King, M., Bos, M., Dach, R., Deo, M., Moore, A.W., Ostini, L., Petrie, E., Roggero, M., Teferle, F.N., Olivares, G. and Webb, F.H. (2013). Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment. *J. Geophys. Res.* (*Solid Earth*) **118** 2397–2407.

[16] Harchaoui, Z. and Lévy-Leduc, C. (2010). Multiple change-point estimation with a total variation penalty. *J. Amer. Statist. Assoc.* **105** 1480–1493. MR2796565

[17] Lai, T.L., Liu, H. and Xing, H. (2005). Autoregressive models with piecewise constant volatility and regression parameters. *Statist. Sinica* **15** 279–301. MR2190207

[18] Lai, W.R., Johnson, M.D., Kucherlapati, R. and Park, P.J. (2005). Comparative analysis of algorithms for identifying amplifications and deletions in array CGH data. *Bioinformatics* **21** 3763–3770.

[19] Lavielle, M. (1999). Detection of multiple changes in a sequence of dependent variables. *Stochastic Process. Appl.* **83** 79–102. MR1705601

[20] Lavielle, M. (2005). Using penalized contrasts for the change-point problem. *Signal Process.* **85** 1501–1510.

[21] Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *J. Time Ser. Anal.* **21** 33–59. MR1766173

[22] Lebarbier, É. (2005). Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing* **85** 717–736.

[23] Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M.S. and Reisen, V.A. (2011). Robust estimation of the scale and of the autocovariance function of Gaussian short- and long-range dependent processes. *J. Time Series Anal.* **32** 135–156. MR2807883

[24] Li, S. and Lund, R. (2012). Multiple changepoint detection via genetic algorithms. *J. Climate* **25** 674–686.

[25] Lu, Q., Lund, R. and Lee, T.C.M. (2010). An MDL approach to the climate segmentation problem. *Ann. Appl. Stat.* **4** 299–319. MR2758173

[26] Ma, Y. and Genton, M.G. (2000). Highly robust estimation of the autocovariance function. *J. Time Ser. Anal*. **21** 663–684. MR1801704
[27] Maidstone, R., Hocking, T., Rigaill, G. and Fearnhead, P. (2014). On optimal multiple changepoint algorithms for large data. Available at arXiv:1409.1842.
[28] Mestre, O. (2000). Méthodes statistiques pour l'homogénéisation de longues séries climatiques. Ph.D. thesis, Univ. Toulouse 3.
[29] Olshen, A.B., Venkatraman, E.S., Lucito, R. and Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5** 557–572.
[30] Picard, F., Lebarbier, E., Hoebeke, M., Rigaill, G., Thiam, B. and Robin, S. (2011). Joint segmentation, calling, and normalization of multiple CGH profiles. *Biostatistics* **12** 413–428.
[31] Picard, F., Robin, S., Lavielle, M., Vaisse, C. and Daudin, J.-J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinformatics* **6** 27.
[32] Rigaill, G., Lebarbier, E. and Robin, S. (2012). Exact posterior distributions and model selection criteria for multiple change-point detection problems. *Stat. Comput*. **22** 917–929. MR2913792
[33] Rousseeuw, P.J. and Croux, C. (1993). Alternatives to the median absolute deviation. *J. Amer. Statist. Assoc*. **88** 1273–1283. MR1245360
[34] Van der Vaart, A.W. (2000). *Asymptotic Statistics* **3**. Cambridge: Cambridge Univ. Press.
[35] Venkatraman, E.S. (1992). Consistency results in multiple change-point problems. Ph.D. thesis, Stanford Univ.
[36] Williams, S. (2003). Offsets in global positioning system time series. *J. Geophys. Res*. (*Solid Earth*) **108**.
[37] Yao, Y.-C. (1988). Estimating the number of change-points via Schwarz' criterion. *Statist. Probab. Lett*. **6** 181–189. MR0919373
[38] Yao, Y.-C. and Au, S.T. (1989). Least-squares estimation of a step function. *Sankhyā Ser. A* **51** 370–381. MR1175613
[39] Zhang, N.R. and Siegmund, D.O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63** 22–32, 309. MR2345571

Received February 2015 and revised October 2015