

## CHANGE-POINT DETECTION, SEGMENTATION, AND RELATED TOPICS

JEAN-MARC BARDET<sup>1</sup>, VINCENT BRAULT<sup>2</sup>, SERGUEÏ DACHIAN<sup>3</sup>, FARIDA ENIKEEVA<sup>4</sup> AND BRUNO SAUSSEREAU<sup>5</sup>

**Abstract.** Recent contributions to change-point detection, segmentation and inference for non-regular models are presented. Various problems are considered including the multiple change-point estimation with adaptive penalty for time series with different dependency structures, estimation of the singularity point in cusp-type models, inference for thresholded autoregressive models, and cross-segmentation of matrices.

**Résumé.** Des contributions récentes dans les problèmes de détection de rupture, de ségmentation et de l'inférence pour des modèles non-réguliers sont présentées. Les problèmes considérés incluent l'estimation de plusieurs points de rupture avec une pénalité adaptative pour des séries temporelles avec différentes structures de dépendance, l'estimation d'un point de singularité pour des modèles de type cusp, l'inférence pour des modèles auto-régressifs à seuil et la segmentation croisée des matrices.

### INTRODUCTION

This article presents some recent results on the statistical inference for models with abrupt changes in the parameters and for related non-regular problems. The results were presented at the session “Change-point detection and segmentation” of the Journées MAS 2016, organized by Farida Enikeeva. Each section of the article is based on one of four talks given by Jean-Marc Bardet, Vincent Brault, Sergueï Dachian and Bruno Saussereau. Farida Enikeeva has written the introduction and coordinated the present article.

The models with irregularities such as the abrupt changes in a signal or noise parameters have received a lot of attention for decades. These problems arise naturally in many applications such as quality control, network traffic data analysis, seismography, analysis of DNA sequences, audio signal processing and many others [8]. A number of methods had been developed for the problems of detecting a single change in a parameter (often the signal mean) of an observed process in the iid noise case and under dependency conditions (see [17], [25]). Later on those methods were generalized to the problem of estimating multiple change points with known or bounded number of changes. The early techniques include, for example, the binary segmentation [81], the least squares estimation [62], and the dynamic programming approach [51] that allows to estimate the unknown change-points in quadratic time with respect to the sample length. Several techniques were proposed for the

---

<sup>1</sup> Equipe SAMM, Université Paris I, Bureau C20-11, 90 rue Tolbiac 75013 Paris, France

<sup>2</sup> Univ. Grenoble Alpes, CNRS, Grenoble INP\*, LJK, 38000 Grenoble, France

\* Institute of Engineering Univ. Grenoble Alpes

<sup>3</sup> Université de Lille, CNRS, UMR 8524 — Laboratoire Paul Painlevé, F-59000 Lille, France

<sup>4</sup> Laboratoire de Mathématiques et Applications, UMR 7348, CNRS, Université de Poitiers, Site du Futuroscope, Téléport 2, 11 Boulevard Marie et Pierre Curie, Bâtiment H3, TSA 61125, 86073 Poitiers Cedex 9, France

<sup>5</sup> Laboratoire de mathématiques de Besançon, UMR 6623, CNRS Université Bourgogne Franche-Comté, 16 route de Gray, 25030 Besançon, France

estimation in case of unknown number of change-points such as the wild binary segmentation [39], the SMUCE technique [38], the model selection approach [63], estimating with total variation penalty [44]. Another method of estimation of multiple change-points based on the slope heuristic procedure is presented in this article.

An important generalization of the classical change point detection is the problem of segmentation of multidimensional observations such as matrix or image segmentation (see for example [53] for some theoretical results). Other generalizations include thresholded models with unknown threshold parameter [78] and non-regular models with cusp-type singularity at an unknown location (see [27] and the numerous references therein). Let us note that unlike classical change-point models (where the change is discontinuous), in models with a cusp-type singularity, the change is continuous (though still quite abrupt). So, sometimes these models are classified as “smooth” change-point models.

The first section contains an overview of different results on the multiple change-point estimation using an adaptive penalty in several semi-parametric frameworks. This section was written by J.-M. Bardet and presents the results obtained in the joint works with C. Dion [4], A. Guenza [5], W. Kengne and O. Wintenberger [6], and with C. Faure, J. Lacaille and M. Olteanu [35].

The second section written by V. Brault is on the problem of estimating the boundaries of homogeneous blocks of an observed matrix by cross-segmentation. This contribution is based on the joint work with J.-C. Quinton and A. Samson and is already partially presented at two conferences (see [16] and [14]). In this section two models and the associated segmentation procedures are presented. The obtained results and the differences between two methods are discussed.

The author of the third section is S. Dachian. This part presents a survey of recent results on change-point location estimation for different observation models in presence of a cusp-type singularity in the driving function (density, signal, intensity function, drift, *etc.*) at the change-point. The survey follows a recent paper [27] by S. Dachian, N. Kordzakhia, Yu. Kutoyants and A. Novikov. For each model, the asymptotic behavior of the Bayesian and the maximum likelihood estimators is studied.

The last section is written by B. Sausseureau. It presents the results on inference for Threshold Autoregressive (TAR) models obtained jointly with A. M. Elmi [33]. The authors study asymptotic properties of least square estimators for weak TAR models and show that they are strongly consistent under some mixing assumptions.

## 1. OFFLINE DETECTION OF MULTIPLE CHANGES USING AN ADAPTIVE PENALTY

Consider the following example of a real dataset, *i.e.* log-ratio of the closing values of the FTSE (The Financial Times Stock Exchange 100 Index) index from 27 July 2005 to 18 March 2011 (Figure 1). In such an example, once the trajectory is observed, we might be interested in the detection of eventual changes in the dynamics of this time series. Using the semi-parametric methods that will be presented below, we obtain the division of the observed time series into 4 distinct zones, with the break instants corresponding to the dates of beginning or end of financial crises (typically September 2008).

In the sequel we present a general method for solving such a problem of offline detection of changes in three different semi-parametric frameworks (we denote  $(\xi_t)_{t \in \mathbb{Z}}$  a sequence of centered i.i.d.r.v.) once  $(X_1, \dots, X_n)$  is observed:

- (1) Signal + Noise model: Here, we assume that for any  $\theta \in \Theta \subset \mathbb{R}^d$ , there exists a function  $s_\theta : \mathbb{Z} \rightarrow \mathbb{R}$ , which is known, such as:

$$X_t = s_\theta(t) + \xi_t \quad \forall t \in \mathbb{Z};$$

- (2) Causal affine time series: Here, we assume that for any  $\theta \in \Theta \subset \mathbb{R}^d$ , there exists two functions  $f_\theta : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$  and  $M_\theta : \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$ , which are known, and such as

$$X_t = M_\theta(X_{t-1}, X_{t-2}, \dots) \xi_t + f_\theta(X_{t-1}, X_{t-2}, \dots) \quad \forall t \in \mathbb{Z};$$

- (3) Long memory process: Here we assume that  $(X_t)_{t \in \mathbb{Z}}$  is a long memory second order stationary process with a spectral density satisfying the following expansion: there exist  $d \in (0, 0.5)$ ,  $c_0$ ,  $c_1$  and  $\beta$  three

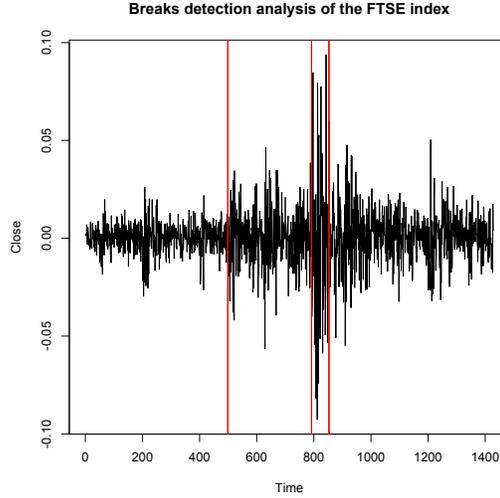


FIGURE 1. Log-ratio of the closing values of the FTSE index from July 27, 2005 to March 18, 2011 ( $n = 1428$ ).

positive real numbers such as

$$f_X(\lambda) = |\lambda|^{-2d}(c_0 + c_1|\lambda|^\beta + o(|\lambda|^\beta)) \quad \text{as } \lambda \rightarrow 0.$$

We assume that  $(X_1, \dots, X_n)$  has an unknown number  $K^* - 1$  of semi-parametric changes. More precisely, in each of the three frameworks,  $(X_t)$  is a time series depending on the changing parameter  $\theta$  with the values  $\theta_k^*$  for  $t \in \{t_{k-1}^* + 1, t_{k-1}^* + 2, \dots, t_k^*\}$ , where  $k = 1, \dots, K^*$ , with  $t_0^* = 0$  and  $t_{K^*}^* = n$  by convention. The number of changes  $K^* - 1$ , the change-points  $(t_k^*)_{1 \leq k \leq K^* - 1}$  and the changing parameters  $(\theta_k^*)_{1 \leq k \leq K^*}$  are unknown.

Our aim is to estimate  $K$ ,  $(t_k)_{1 \leq k \leq K-1}$ , and  $(\theta_k)_{1 \leq k \leq K}$ . In the following we will first consider the situation of known number of changes  $K^*$ . In Section 1.1 we give a general overview of the estimation methods based on the minimization of an appropriately contrast function for a known  $K^*$ . Next, in Section 1.2 we will consider several penalization criteria applied in the case of unknown number of changes. The results on the consistency and asymptotic of the obtained estimators of  $K$ ,  $(t_k)_{1 \leq k \leq K-1}$ , and  $(\theta_k)_{1 \leq k \leq K}$  will be given for each of three frameworks (1)–(3) stated above. Finally, in Section 1.3 the slope heuristic procedure of an adaptive penalization is presented.

### 1.1. The case of known number of changes

Let  $(Z_1, \dots, Z_m)$  be an observed sample of random variables  $(Z_t)_{t \in \mathbb{Z}}$  (in the sequel,  $(Z_1, \dots, Z_m)$  will notably correspond to certain portions of the trajectory  $(X_1, \dots, X_n)$ ), with a probability distribution depending on a true and unknown parameter  $\theta^* \in \Theta \subset \mathbb{R}^d$ . For any  $\theta \in \Theta \subset \mathbb{R}^d$ , define a contrast  $\Phi_\theta(Z_1, \dots, Z_m)$  such that

$$\theta^* = \lim_{m \rightarrow \infty} \underset{\theta \in \Theta}{\text{Argmin}} \mathbb{E} [\Phi_\theta(Z_1, \dots, Z_m)].$$

Typically, the following contrast functions are considered:

- $\Phi_\theta$  is a Least Squares (LS) criterion:  $\Phi_\theta(Z_1, \dots, Z_m) = \sum_{j=1}^m (Z_j - {}^t U_j \theta)^2$  or Least Absolute Value (LAV) criterion:  $\Phi_\theta(Z_1, \dots, Z_m) = \sum_{j=1}^m |Z_j - {}^t U_j \theta|$ , with  $(U_j)_{1 \leq j \leq m}$  a family of observed vectors (for instance observed exogenous variables);

- $\Phi_\theta(Z_1, \dots, Z_m) = -2 \log(L_\theta(Z_1, \dots, Z_m))$  where  $L_\theta$  is the likelihood;
- $\Phi_\theta(Z_1, \dots, Z_m) = -2 \log(\widehat{L}_\theta(Z_1, \dots, Z_m))$  where  $\widehat{L}_\theta$  is the quasi-likelihood.

We consider the general multiple change framework defined below.

**Assumption.** We observe a sample  $(X_1, \dots, X_n)$  such that  $X_t$  is a time series depending on the parameter  $\theta_k^*$  for  $t \in \{t_{k-1}^*+1, t_{k-1}^*+2, \dots, t_k^*\}$ ,  $k = 1, \dots, K^*$ , with the change-points  $0 = t_0^* < t_1^* < \dots < t_{K^*-1}^* < t_{K^*}^* = n$  and unknown parameters  $(\theta_k^*)_{1 \leq k \leq K^*} \in \Theta^{K^*}$  such that  $\theta_k^* \neq \theta_{k+1}^*$  for  $1 \leq k \leq K^* - 1$ .

In order to estimate all these parameters, for  $K \in \mathbb{N}^*$ ,  $1 \leq t_0 < t_1 < \dots < t_K = n$  and  $(\theta_i)_{1 \leq i \leq K} \in \Theta^K$ , define  $\widehat{I}_n(K, (t_i)_{1 \leq i \leq K-1}, (\theta_i)_{1 \leq i \leq K})$  as the sum of contrasts within each interval,

$$\widehat{I}_n(K, (t_i)_{1 \leq i \leq K-1}, (\theta_i)_{1 \leq i \leq K}) = \sum_{i=1}^K \Phi_{\theta_i}(X_{t_{i-1}+1}, \dots, X_{t_i})$$

The estimators of parameters minimize this sum of contrasts.

**Definition 1.1.** Let  $K^*$  be known. Define

$$((\widehat{t}_i)_{1 \leq i \leq K^*-1}, (\widehat{\theta}_i)_{1 \leq i \leq K^*}) = \underset{(t_i)_{1 \leq i \leq K^*-1}, (\theta)_{1 \leq i \leq K^*}}{\text{Argmin}} \widehat{I}_n(K^*, (t_i)_{1 \leq i \leq K^*-1}, (\theta_i)_{1 \leq i \leq K^*})$$

Note that a consequence of this definition is the following: if we denote by

$$\widetilde{\theta}_{t_{i-1}, t_i} = \underset{\theta \in \Theta}{\text{Argmin}} \Phi_\theta(X_{t_{i-1}+1}, \dots, X_{t_i}),$$

the estimate of  $\theta$  withing the interval  $(t_{i-1}, t_i]$ , then the estimates of change-points  $\widehat{t}_i$  are obtained as the solution of the following minimization problem:

$$(\widehat{t}_i)_{1 \leq i \leq K^*-1} = \underset{1 < t_1 < \dots < t_{K^*-1} < n}{\text{Argmin}} \widehat{I}_n(K^*, (t_i)_{1 \leq i \leq K^*}, (\widetilde{\theta}_{t_{i-1}, t_i})_{1 \leq i \leq K^*}).$$

Thus the initial minimization problem is reduced to the computation of  $(\widehat{t}_i)_{1 \leq i \leq K^*-1}$ , which normally requires to study  $O(n^{K^*-1})$  different values of  $(\widehat{t}_i)_{1 \leq i \leq K^*-1}$ . By use of the dynamic programming (see [35]), this problem is reduced to  $O(n^2)$  for any  $K^*$ .

## 1.2. The case of unknown number of changes

When  $K^*$  is unknown, we can define a penalized sum of contrasts, *i.e.* for  $0 \leq K \leq K_{\max}$ , where  $K_{\max}$  is a fixed real integer number supposed to be larger or equal to  $K^*$ , we set

$$\widehat{J}_n(K, (t_i)_{1 \leq i \leq K-1}, (\theta_i)_{1 \leq i \leq K}) = \widehat{I}_n(K, (t_i)_{1 \leq i \leq K-1}, (\theta_i)_{1 \leq i \leq K}) + \kappa_n \text{pen}(K)$$

where  $(\kappa_n)$  is a sequence of positive numbers and  $K \in \mathbb{N} \mapsto \text{pen}(K)$  is an increasing function.

**Definition 1.2.** Let  $K^*$  be unknown. Define

$$(\widehat{K}, (\widehat{t}_i)_{1 \leq i \leq \widehat{K}-1}, (\widehat{\theta}_i)_{1 \leq i \leq \widehat{K}}) = \underset{K, (t_i)_{1 \leq i \leq K-1}, (\theta)_{1 \leq i \leq K}}{\text{Argmin}} \widehat{J}_n(K, (t_i)_{1 \leq i \leq K-1}, (\theta_i)_{1 \leq i \leq K}).$$

The estimators of this type arise in three following well known situations.

- The case of multiple least square regression (see for instance [3]) with

$$\Phi_{\theta_i}(X_{t_{i-1}+1}, \dots, X_{t_i}) = \sum_{j=t_{i-1}+1}^{t_i} (X_j - {}^t U_j \theta_i)^2, \text{ pen}(K) = (d+1)K - 1 \text{ and } \kappa_n = 2\sigma^2,$$

always with  $(U_k)_{1 \leq k \leq n}$  a family of observed vectors in  $\mathbb{R}^d$ . This is the classical Mallows  $C_p$  criterion;

- The general case of maximum likelihood estimation with

$$\Phi_{\theta_i}(X_{t_{i-1}+1}, \dots, X_{t_i}) = -2 \log(L_{\theta_i}(X_{t_{i-1}+1}, \dots, X_{t_i})), \text{ pen}(K) = (d+1)K - 1 \text{ and } \kappa_n = 2.$$

This is the classical AIC criterion;

- The general case of maximum likelihood estimation with

$$\Phi_{\theta_i}(X_{t_{i-1}+1}, \dots, X_{t_i}) = -2 \log(L_{\theta_i}(X_{t_{i-1}+1}, \dots, X_{t_i})), \text{ pen}(K) = (d+1)K - 1, \text{ and } \kappa_n = \log(n).$$

This is the classical BIC criterion.

Let us make additional assumptions:

- $t_k^* = [n \tau_k^*]$  with  $\tau_0^* = 0 < \tau_1^* < \dots < \tau_{K^*}^* = 1$ ;
- the explicit form or expansion of functions  $s_\theta$  in the "signal + noise" model,  $f_\theta$  and  $M_\theta$  in the case of causal affine time series and the expansion of the spectral density  $f_X(\lambda)$  in case of long memory processes are known.
- $K^*, (\tau_k^*)_{1 \leq k \leq K^*-1}, (\theta_k^*)_{1 \leq k \leq K^*}$  and the distribution of  $\xi$  are unknown.

In the sequel we will use the following notation:

- $\underline{t}^* = (t_1^*, \dots, t_{K^*-1}^*), \widehat{\underline{t}} = (\widehat{t}_1, \dots, \widehat{t}_{\widehat{K}-1}), \underline{\tau}^* = (\tau_1^*, \dots, \tau_{K^*-1}^*)$  and  $\widehat{\underline{\tau}} = (\widehat{\tau}_1, \dots, \widehat{\tau}_{\widehat{K}-1})$ ;
- $\underline{\theta}^* = (\theta_1^*, \dots, \theta_{K^*}^*)$  and  $\widehat{\underline{\theta}} = (\widehat{\theta}_1, \dots, \widehat{\theta}_{\widehat{K}})$ .

Now, we consider the three different frameworks and state some asymptotic results.

### 1.2.1. Framework 1: signal + noise model

We observe  $(X_1, \dots, X_n)$  where the underlying process  $(X_t)_{t \in \mathbb{Z}}$  satisfies the relationship

$$X_t = s_{\theta_i^*}(t) + \xi_t \quad \text{for all } t \in \{t_{i-1}^* + 1, \dots, t_i^*\},$$

for  $i = 1, \dots, K^*$  and with an explicit function  $t \rightarrow s_\theta(t)$  depending on  $\theta \in \Theta \subset \mathbb{R}^d$ .

**Example 1.3.** In a typical situation we consider  $s_\theta(t) = \theta \in \mathbb{R}$  and therefore  $d = 1$ . Then, for different contrast functions  $\Phi_\theta$  we have the following estimators of  $\theta$ :

- for the contrast  $\Phi_\theta^{(2)}(X_1, \dots, X_n) = n \log \left( \sum_{t=1}^n (X_t - s_\theta(t))^2 \right)$  which is equivalent to the LS criterion, we have  $\widehat{\theta}_n = \frac{1}{n} \sum_{t=1}^n X_t$ ;
- for the contrast  $\Phi_\theta^{(1)}(X_1, \dots, X_n) = n \log \left( \sum_{t=1}^n |X_t - s_\theta(t)| \right)$  which is equivalent to the LAV criterion, we have  $\widehat{\theta}_n = \text{median}(X_1, \dots, X_n)$ .

Both for  $\Phi_\theta^{(1)}$  (see [2]) and  $\Phi_\theta^{(2)}$  (see [3] or [62]), the following theorem establishes the consistency of the corresponding estimators.

**Theorem 1.4.** *If  $\kappa_n \xrightarrow{n \rightarrow +\infty} \infty$  and  $\kappa_n/n \xrightarrow{n \rightarrow +\infty} 0$  with  $\text{pen}(K) = K$  and  $K_{\max} \geq K^*$ , then:*

$$(\widehat{K}_n, \widehat{\underline{t}}, \widehat{\theta}) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} (K^*, \underline{t}^*, \theta^*)$$

More precisely,  $\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\|\widehat{\underline{t}} - \underline{t}^*\|_d > \delta) = 0$ .

**Remark 1.5.** For the contrast  $\Phi_\theta^{(2)}$  the results are also valid under mixing conditions on  $(\xi_t)$ .

### 1.2.2. Framework 2: general causal affine models

We observe a sample  $(X_1, \dots, X_n)$  where

$$X_t = M_{\theta_i^*}(X_{t-1}, X_{t-2}, \dots) \xi_t + f_{\theta_i^*}(X_{t-1}, X_{t-2}, \dots) \quad \text{for all } t \in \{t_{i-1}^* + 1, \dots, t_i^*\},$$

for  $i = 1, \dots, K^*$  and explicit functions  $x \in \mathbb{R}^{\mathbb{N}} \rightarrow f_\theta(x)$  and  $x \in \mathbb{R}^{\mathbb{N}} \rightarrow M_\theta(x)$  for  $\theta \in \Theta$ .

**Example 1.6.** The following usual time series can be represented as particular cases of causal affine models:

- For AR( $\infty$ ) processes (and therefore ARMA processes), we have  $X_t = \xi_t + \sum_{i=1}^{\infty} a_i X_{t-i}$  and therefore  $f_\theta(X_{t-1}, X_{t-2}, \dots) = \sum_{i=1}^{\infty} a_i(\theta) X_{t-i}$  and  $M_\theta(X_{t-1}, X_{t-2}, \dots) = \sigma^2$ ;
- For ARCH( $\infty$ ) processes (and therefore GARCH processes):  $X_t = \xi_t (a_0 + \sum_{i=1}^{\infty} a_i X_{t-i}^2)^{1/2}$  and therefore  $f_\theta(X_{t-1}, X_{t-2}, \dots) = 0$  and  $M_\theta(X_{t-1}, X_{t-2}, \dots) = (a_0 + \sum_{i=1}^{\infty} a_i(\theta) X_{t-i}^2)^{1/2}$ ;
- the processes ARMA( $p, q$ )-GARCH( $p', q'$ ), APARCH( $p, \delta, q$ ), TARARCH( $p, q$ ) are other examples (see [7]).

Denote now for any  $\theta_i \in \Theta$ ,

$$\begin{cases} f_{\theta_i}^t = f_{\theta_i}(X_{t-1}, X_{t-2}, \dots) \\ M_{\theta_i}^t = M_{\theta_i}(X_{t-1}, X_{t-2}, \dots) \end{cases} \quad \text{for } t \in \{t_{i-1}^* + 1, \dots, t_i^*\}.$$

In case of general causal affine processes we can use a Gaussian Quasi-Maximum Likelihood (QML) estimation. We will use the short abbreviation QMLE is we talk about Quasi-Maximum Likelihood estimator.

(1) If  $(\xi_t)_t$  are i.i.d.r.v.  $\mathcal{N}(0, 1)$ , the conditional (w.r.t.  $(X_t)_{t \leq t_{i-1}}$ ) log-likelihood of  $(X_{t_{i-1}+1}, \dots, X_{t_i})$  is

$$L_i^G(\theta_i) = \sum_{t=t_{i-1}+1}^{t_i} q_t^G(\theta_i) \quad \text{with} \quad q_t^G(\theta_i) = -\frac{1}{2} \frac{(X_t - f_{\theta_i}^t)^2}{(M_{\theta_i}^t)^2} + \log(M_{\theta_i}^t).$$

However  $X_0, X_{-1}, \dots$  are unobserved! Thus define for  $\theta \in \Theta$

$$\begin{cases} \widehat{f}_\theta^t = f_\theta(X_{t-1}, \dots, X_1, 0, \dots) \\ \widehat{M}_\theta^t = M_\theta(X_{t-1}, \dots, X_1, 0, \dots) \end{cases} \quad \text{for } t \in \{1, \dots, n\}.$$

Then the Gaussian QML on  $\{t_{i-1} + 1, \dots, t_i\}$  is given by

$$\widehat{L}_i^G(\theta_i) = \sum_{t=t_{i-1}+1}^{t_i} \widehat{q}_t^G(\theta_i), \quad \text{where} \quad \widehat{q}_t^G(\theta_i) = -\frac{1}{2} \frac{(X_t - \widehat{f}_{\theta_i}^t)^2}{(\widehat{M}_{\theta_i}^t)^2} + \log(\widehat{M}_{\theta_i}^t)$$

and the Gaussian QMLE is  $\widehat{\theta}_i^G = \text{Argmax}_{\theta_i \in \Theta} \widehat{L}_i^G(\theta_i)$ .

(2) If  $(\xi_t)_t$  are i.i.d.r.v. from the Laplace distribution, the conditional log-likelihood of  $(X_1, \dots, X_n)$  is

$$L_i^L(\theta_i) = \sum_{t=t_{i-1}+1}^{t_i} q_t^L(\theta_i) \quad \text{with} \quad q_t^L(\theta_i) = -\frac{|X_t - f_{\theta_i}^t|}{M_{\theta_i}^t} + \log(M_{\theta_i}^t)$$

Then we can define the Laplacian QML by

$$\widehat{L}_i^L(\theta_i) = \sum_{t=t_{i-1}+1}^{t_i} \widehat{q}_t^L(\theta_i) \quad \text{with} \quad \widehat{q}_t^L(\theta_i) = -\frac{|X_t - \widehat{f}_{\theta_i}^t|}{\widehat{M}_{\theta_i}^t} + \log(\widehat{M}_{\theta_i}^t)$$

and the Laplacian QMLE as  $\widehat{\theta}_i^L = \text{Argmax}_{\theta_i \in \Theta} \widehat{L}_i^L(\theta_i)$ .

**Definition 1.7.** Define the contrasts  $\Phi_{\theta_i}^G$  and  $\Phi_{\theta_i}^L$  as follows,

$$\Phi_{\theta_i}^G(X_{t_{i-1}+1}, \dots, X_{t_i}) = -2\widehat{L}_i^G(\theta_i) \quad \text{and} \quad \Phi_{\theta_i}^L(X_{t_{i-1}+1}, \dots, X_{t_i}) = -\widehat{L}_i^L(\theta_i).$$

Finally, we will assume that there exists two sequences of Lipschitzian coefficients  $(\alpha_i(f))_{i \geq 1}$  and  $(\alpha_i(M))_{i \geq 1}$  respectively for  $f_\theta$  and  $g_\theta$ , i.e satisfying for any  $x, y \in \mathbb{R}^N$ ,

$$\sup_{\theta \in \Theta} |f_\theta(x) - f_\theta(y)| \leq \sum_{i=1}^{\infty} \alpha_i(f) |x_i - y_i| \quad \text{and} \quad \sup_{\theta \in \Theta} |M_\theta(x) - M_\theta(y)| \leq \sum_{i=1}^{\infty} \alpha_i(M) |x_i - y_i|.$$

**Theorem 1.8.** For  $\Phi_{\theta_i}^G$  we prove in [6] that

- if  $\text{pen}(K) = K$  and  $K_{\max} \geq K^*$ , if  $r \geq 2$ , under identifiability conditions on  $f_\theta$  and  $M_\theta$ , if there exists  $\underline{M} > 0$  such as for any  $x \in \mathbb{R}^N$  and any  $\theta \in \Theta$ ,  $M_\theta(x) \geq \underline{M}$  and if  $\kappa_n \wedge n \kappa_n^{-1} \rightarrow \infty$  and

$$\sum_{k \geq 2} \kappa_k^{-(r/4 \wedge 1)} \left( \sum_{\ell \geq k/\log(k)} \alpha_\ell(f) + \alpha_\ell(M) \right)^{(r/4 \wedge 1)} < \infty$$

then  $(\widehat{K}_n, \widehat{\tau}, \widehat{\theta}) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} (K^*, \tau^*, \theta^*)$ .

- Moreover, if  $r \geq 4$  and if  $\kappa_n = \sqrt{n}$ , under conditions on the first and second derivatives of  $f_\theta$  and  $M_\theta$  (see more details in [6]), then

$$\lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}(\|\widehat{t} - t^*\|_m > \delta) = 0 \quad \text{and} \quad \sqrt{n} \sqrt{\tau_j^* - \tau_{j-1}^*} (\widehat{\theta}_j - \theta_j^*) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}_d(0, G(\theta_j^*))$$

**Remark:** Note that the convergence rate is the same as if  $(X_1, \dots, X_n)$  was a family of independent random variables.

### 1.2.3. Framework 3: Long memory processes

In this case, the spectral density of  $(X_{t_{i-1}+1}, \dots, X_{t_i})$  is

$$f_X(\lambda) = |\lambda|^{-2d_i^*} (c_{0,i}^* + c_{1,i}^* |\lambda|^{\beta_i^*} + o(|\lambda|^{\beta_i^*})) \quad \text{as } \lambda \rightarrow 0,$$

where  $d_i^* \in (0, 1/2)$ ,  $c_{0,i}^*$ ,  $c_{1,i}^*$  and  $\beta_i^*$  are positive real numbers. Robinson [75] introduced the local Whittle contrast for estimating  $d_i^*$ . Hence, for  $d_i \in (0, 1/2)$ , and  $m \in \mathbb{N}^*$ , define

$$W_i(d_i, m) = \log \left( \frac{1}{m} \sum_{k=1}^m \left( \frac{k}{m} \right)^{2d_i} I_i(\lambda_k) \right) - \frac{2d_i}{m} \sum_{k=1}^m \log(k/m), \quad \text{where } \lambda_k = 2\pi \frac{k}{t_i - t_{i-1}}$$

and

$$I_i(\lambda) = \frac{1}{2\pi(t_i - t_{i-1})} \left| \sum_{k=t_{i-1}+1}^{t_i} X_k e^{-ik\lambda} \right|^2,$$

is the periodogram. Using this contrast we can define the local Whittle estimator of the long memory parameter  $d_i^*$

$$\hat{d}_i = \underset{d_i \in (0, 1/2)}{\text{Argmin}} \{W_i(d_i, m)\}.$$

**Definition 1.9.** Define the contrast  $\Phi_{\theta}^{LM}$  by

$$\Phi_{\theta_i}^{LM}(X_{t_{i-1}+1}, \dots, X_{t_i}) = (t_i - t_{i-1}) W_i(d_i, m).$$

In [5], we prove the following theorem.

**Theorem 1.10.** *If  $m = o(n^{2\beta^*/(1+2\beta^*)})$ , where  $\beta^* = \min_{1 \leq i \leq K^*+1} \beta_i^*$  and if  $\max\left(\frac{\kappa_n}{n}, \frac{n \log n}{\kappa_n \sqrt{m}}\right) \xrightarrow{n \rightarrow +\infty} 0$ , then*

$$(\hat{K}, \hat{\tau}, \hat{\underline{d}}) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} (K^*, \tau^*, \underline{d}^*).$$

We also obtain the following rates of convergence:

$$\sqrt{m} \|\hat{\underline{d}} - \underline{d}^*\| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0 \quad \text{and} \quad \lim_{\delta \rightarrow \infty} \lim_{n \rightarrow \infty} \mathbb{P}\left(\frac{\sqrt{m}}{n} \|\hat{\underline{t}} - \underline{t}^*\| \geq \delta\right) = 0.$$

### 1.3. An adaptive penalization: the slope heuristic procedure

The heuristic slope procedure has been introduced in [1]. It was applied in the three previous framework (see respectively [4] or [35], [6] and [5]) for deducing a data-driven penalization. In the sequel we detail the case of long-memory change detection.

Instead of a fixed sequence  $(\kappa_n)$ , we can use a data-driven penalty rate  $\hat{\kappa}_n = 2 \times |\hat{s}|$ , with  $\hat{s}$  the slope of the least square regression of  $\hat{I}_n(K, \hat{\underline{t}}, \hat{\underline{d}})$  onto  $K$  for  $K$  large enough. Let

$$\hat{K}_H = \underset{0 \leq K \leq K_{\max}}{\text{Argmin}} \left\{ \hat{I}_n(K, \hat{\underline{t}}, \hat{\underline{d}}) + 2 \times |\hat{s}| \times K \right\}$$

be an estimator of the number of changes using the data-driven procedure. Figure 1.3 shows a graphical illustration of this procedure in the long memory processes framework: Using the Monte-Carlo experiments we have shown that this data-driven procedure leads to more accurate results than the procedures based on penalization with a sequence  $(\kappa_n)$  chosen a priori (see in [4] and [35] for the result for Framework 1, [6] for Framework 2, and [5] for the case of Framework 3).

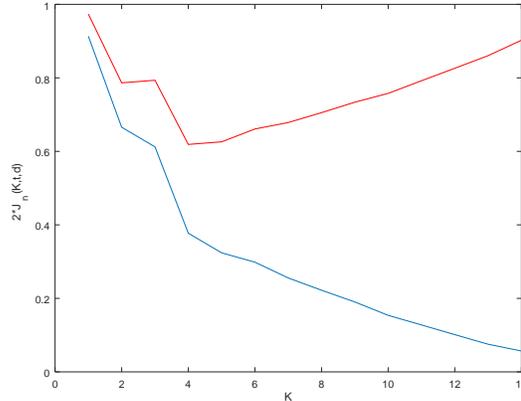


FIGURE 2. The graphs of  $\hat{I}_n(K, \hat{t}, \hat{d})$  (in black) and of  $\hat{I}_n(K, \hat{t}, \hat{d}) + 2 \times |\hat{s}| \times K$  (in black) for  $n = 5000$  observations with  $K^* = 4$  change-points of a FARIMA(0,  $d$ , 0) process.

## 2. CROSS-SEGMENTATION OF COLUMNS AND ROWS OF A MATRIX: COMPARISON OF TWO PROCEDURES WITH AN APPLICATION TO THE STUDY OF AUTONOMOUS VEHICLE

### 2.1. Introduction

In the research about the autonomous vehicles, the cost of a high performance GPS is a problem and a solution proposed is to study the environment of the vehicles to guide the latter in the case of regular trips. For this, we need to synthesize large videos footage of the environment (see [54]) by separating the distinct places (e.g. straight line, intersection...) (see [10]).

This problem is similar to the analysis of the biological Hi-C data (see [30]) and some algorithms exist. Notably, Brault et al. [12] have developed an algorithm of quick segmentation for a block-wise constant matrix, Brault et al. [15] have studied the segmentation method based on rank statistics.

In Figure 3, we display an example of a trip and the similarity matrix of the associated video images: the redder a cell  $(t_1, t_2)$  is, the stronger is the resemblance between the two associated images at times  $t_1$  and  $t_2$ . We can observe some color blocks in the matrix and our goal is to segment the rows and columns to bring out homogeneous blocks (either with uniform coloring or with the same types of colors).

To estimate the change-points in the matrix, two procedures have been developed for the Hi-C data called respectively **BlockSeg** (see [12]) and **MuChPoint** (see [15]). In this presentation, we compare the two methods with respect to their complexities and theoretical guarantees (see also [16] and [14]).

### 2.2. Statistical framework

In this part, we develop each modelization.

#### 2.2.1. BlockSeg method

In their article [12], the authors consider estimating  $\mathbf{n}_1^* = (n_{1,1}^*, \dots, n_{1,L_1^*}^*)$  and  $\mathbf{n}_2^* = (n_{2,1}^*, \dots, n_{2,L_2^*}^*)$  from the random matrix  $\mathbf{Y} = (Y_{i,j})_{1 \leq i,j \leq n}$  defined by

$$\mathbf{Y} = \mathbf{C} + \mathbf{E} \text{ where } C_{i,j} = \mu_{k,\ell}^* \quad \text{if } n_{1,k-1}^* \leq i < n_{1,k}^* \text{ and } n_{2,\ell-1}^* \leq j < n_{2,\ell}^*, \quad (1)$$

with the convention  $n_{1,0}^* = n_{2,0}^* = 1$  and  $n_{1,L_1^*+1}^* = n_{2,L_2^*+1}^* = n + 1$ . Then  $\mathbf{C} = (C_{i,j})_{1 \leq i,j \leq n}$  is a blockwise constant matrix and the entries  $E_{i,j}$  of the matrix  $\mathbf{E} = (E_{i,j})_{1 \leq i,j \leq n}$  are iid zero-mean random variables. With such a definition the  $Y_{i,j}$  are assumed to be independent random variables with a blockwise constant mean.

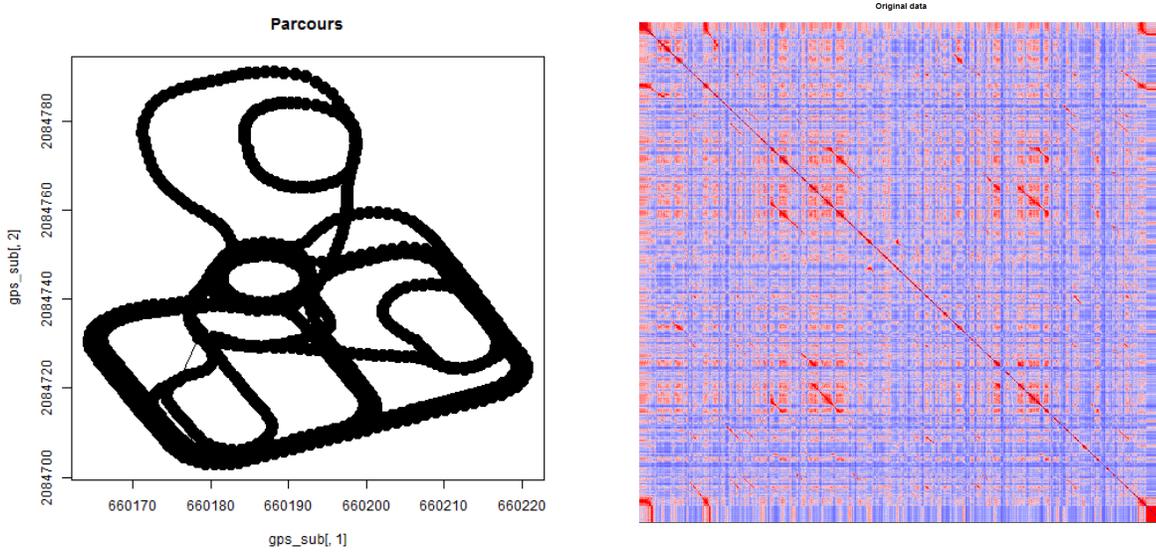


FIGURE 3. On the left are the GPS coordinates of the route made by a vehicle whose film was studied. On the right is the matrix of similarities of the associated video images.

Let  $\mathbf{T}$  be a  $n \times n$  lower triangular matrix with non-zero elements equal to one and  $\mathbf{B}$  a sparse matrix containing null entries except for the  $\mathbf{B}_{i,j}$  such that  $(i, j) \in \{n_{1,0}^*, \dots, n_{1,L_1}^*\} \times \{n_{2,0}^*, \dots, n_{2,L_2}^*\}$ . Then, (1) can be rewritten as follows:

$$\mathbf{Y} = \mathbf{T}\mathbf{B}\mathbf{T}^\top + \mathbf{E}, \quad (2)$$

where  $\mathbf{T}^\top$  denotes the transpose of the matrix  $\mathbf{T}$ . Let  $\text{Vec}(\mathbf{X})$  denote the vectorization of the matrix  $\mathbf{X}$  formed by stacking the columns of  $\mathbf{X}$  into a single column vector, then  $\text{Vec}(\mathbf{Y}) = \text{Vec}(\mathbf{T}\mathbf{B}\mathbf{T}^\top) + \text{Vec}(\mathbf{E})$ . Thus (2) can be rewritten as  $\mathcal{Y} = \mathcal{X}\mathcal{B} + \mathcal{E}$  with  $\mathcal{Y} = \text{Vec}(\mathbf{Y})$ ,  $\mathcal{X} = \mathbf{T} \otimes \mathbf{T}$ , where  $\otimes$  denotes the Kronecker product,  $\mathcal{B} = \text{Vec}(\mathbf{B})$  and  $\mathcal{E} = \text{Vec}(\mathbf{E})$ . By virtue of these transformations, Model (1) has thus been rephrased as a sparse high dimensional linear model where  $\mathcal{Y}$  and  $\mathcal{E}$  are  $n^2 \times 1$  column vectors,  $\mathcal{X}$  is a  $n^2 \times n^2$  matrix and  $\mathcal{B}$  is  $n^2 \times 1$  sparse column vectors. Multiple change-point estimation problem (1) can thus be addressed as a variable selection problem:

$$\hat{\mathcal{B}}(\lambda_n) = \underset{\mathcal{B} \in \mathbb{R}^{n^2}}{\text{Argmin}} \{ \|\mathcal{Y} - \mathcal{X}\mathcal{B}\|_2^2 + \lambda_n \|\mathcal{B}\|_1 \}, \quad (3)$$

where  $\|u\|_2^2$  and  $\|u\|_1$  are defined for a vector  $u$  in  $\mathbb{R}^{n^2}$  by  $\|u\|_2^2 = \sum_{i=1}^{n^2} u_i^2$  and  $\|u\|_1 = \sum_{i=1}^{n^2} |u_i|$ . Criterion (3) is related to the popular Least Absolute Shrinkage and Selection Operator (LASSO) in least-square regression. Due to the sparsity enforcing property of the  $\ell_1$ -norm, the estimator  $\hat{\mathcal{B}}$  of  $\mathcal{B}$  is expected to be sparse and to have non-zero elements matching with those of  $\mathcal{B}$ . Hence, retrieving the positions of the non zero elements of  $\hat{\mathcal{B}}$  thus provides estimators of  $(n_{1,k}^*)_{1 \leq k \leq L_1}$  and of  $(n_{2,k}^*)_{1 \leq k \leq L_2}$ . More precisely, let us define by  $\hat{\mathcal{A}}(\lambda_n)$  the set of active variables:

$$\hat{\mathcal{A}}(\lambda_n) = \{ j \in \{1, \dots, n^2\} : \hat{\mathcal{B}}_j(\lambda_n) \neq 0 \}.$$

For each  $j$  in  $\hat{\mathcal{A}}(\lambda_n)$ , consider the Euclidean division of  $(j-1)$  by  $n$ , namely  $(j-1) = nq_j + r_j$  then

$$\hat{\mathbf{n}}_1 = (\hat{n}_{1,k})_{1 \leq k \leq |\hat{\mathcal{A}}_1(\lambda_n)|} \in \{r_j + 1 : j \in \hat{\mathcal{A}}(\lambda_n)\}, \quad \hat{\mathbf{n}}_2 = (\hat{n}_{2,\ell})_{1 \leq \ell \leq |\hat{\mathcal{A}}_2(\lambda_n)|} \in \{q_j + 1 : j \in \hat{\mathcal{A}}(\lambda_n)\} \\ \text{where } \hat{n}_{1,1} < \hat{n}_{1,2} < \dots < \hat{n}_{1,|\hat{\mathcal{A}}_1(\lambda_n)|}, \quad \hat{n}_{2,1} < \hat{n}_{2,2} < \dots < \hat{n}_{2,|\hat{\mathcal{A}}_2(\lambda_n)|}. \quad (4)$$

In (4),  $|\widehat{\mathcal{A}}_1(\lambda_n)|$  and  $|\widehat{\mathcal{A}}_2(\lambda_n)|$  correspond to the number of distinct elements in  $\{r_j : j \in \widehat{\mathcal{A}}(\lambda_n)\}$  and  $\{q_j : j \in \widehat{\mathcal{A}}(\lambda_n)\}$ , respectively.

### 2.2.2. MuChPoint method

In their article [15], the authors assume that  $\mathbf{Y} = (Y_{i,j})_{1 \leq i, j \leq n}$  is a symmetric matrix such that the  $Y_{i,j}$ 's are independent random variables for  $i \geq j$ . Let  $n_0 = 1 < n_1 < n_2 < \dots < n_{L^*} < n+1 = n_{L^*+1}$  be a given  $L^*$ -uplet in  $\{2, \dots, n\}$ , the authors assume that for each  $\ell \in \{0, \dots, L^* - 1\}$ , there exists at least one row  $i \in \{1, \dots, n\}$  and two distinct laws  $\mathcal{L}_{i,\ell}$  and  $\mathcal{L}_{i,\ell+1}$  such that  $(Y_{i,j})_{j \in \{n_\ell, n_{\ell+1}, \dots, n_{\ell+1}-1\}}$  are independent and identically distributed following  $\mathcal{L}_{i,\ell}$ ,  $(Y_{i,j})_{j \in \{n_{\ell+1}, n_{\ell+1}+1, \dots, n_{\ell+2}-1\}}$  are independent and identically distributed following  $\mathcal{L}_{i,\ell+1}$  and the variables  $(Y_{i,j})_{j \in \{n_\ell, n_{\ell+1}, \dots, n_{\ell+2}-1\}}$  are independent.

To estimate the change-points, they propose to use a test statistic inspired by the one designed by Lung-Yut-Fong et al. [67] which extends the well-known Wilcoxon-Mann-Whitney rank-based test to deal with multivariate data and obtain the following statistic:

$$S_n(n_1, \dots, n_{L^*}) = \frac{4}{n^2} \sum_{\ell=0}^{L^*} (n_{\ell+1} - n_\ell) \sum_{i=1}^n \left( \overline{R}_\ell^{(i)} - \frac{n+1}{2} \right)^2$$

$$\text{with } \overline{R}_\ell^{(i)} = \frac{1}{n_{\ell+1} - n_\ell} \sum_{j=n_\ell}^{n_{\ell+1}-1} \sum_{k=1}^n \mathbb{I}_{\{Y_{i,k} \leq Y_{i,j}\}},$$

where  $\sum_{k=1}^n \mathbb{I}_{\{Y_{i,k} \leq Y_{i,j}\}}$  is the rank of  $Y_{i,j}$  in the row  $i$  and  $\overline{R}_\ell^{(i)}$  is its mean rank in the group  $\ell$ .

To estimate the change-points, the authors propose to maximize the statistic:

$$\widehat{n} = (\widehat{n}_1, \dots, \widehat{n}_{L^*}) \in \underset{1 < n_1 < n_2 < \dots < n_{L^*} < n+1}{\operatorname{argmax}} S_n(n_1, \dots, n_{L^*})$$

using dynamic programming strategy [51]. By symmetry, we obtain at the end  $(L^* + 1)^2$  blocks with the same distribution of elements for each block.

## 2.3. Comparison

In this part, we compare the consistency and the complexity of the two procedures.

### 2.3.1. Consistency

From a theoretical point of view, the two procedures are consistent but not under the same assumptions. We compare the assumptions according to the characteristics of each:

- **Assumption on the laws:**

(B1)  $(E_{i,j})$  are iid zero mean random variables such that  $\exists \beta > 0, \forall \nu \in \mathbb{R}$ ,  
 $\mathbb{E}[\exp(\nu E_{1,1})] \leq \exp(\beta \nu^2)$ .

(M1)  $Y_{i,j}$ 's are independent random variables when  $i \geq j$  and the cumulative distributions functions of the  $Y_{i,j}$ 's are continuous.

- **Assumption on the separation of the blocks:**

(B2) If we denote  $J_{\min}^* = \min_{1 \leq k \leq L_1^*} \|\mu_{k+1}^* - \mu_k^*\|_\infty \wedge \min_{1 \leq \ell \leq L_2^*} \|\mu_{\ell+1}^* - \mu_{\ell}^*\|_\infty$  where  $\|\cdot\|_\infty$  is the maximum norm, there exists  $(\delta_n)_{n \in \mathbb{N}} \in (\mathbb{R}^+)^{\mathbb{N}}$  non increasing such that  
 $\delta_n \xrightarrow[n]{n} 0$  and  $n / \log n \delta_n J_{\min}^{*2} \xrightarrow[n]{n} +\infty$ .

(M2) For each  $\ell \in \{0, \dots, L^*\}$ , there exists  $i \in \{1, \dots, n\}$ , such that if  $X$  follows the law  $\mathcal{L}_{i,\ell}$  and  $Y$  is distributed according to  $\mathcal{L}_{i,\ell+1}$ , then  $\mathbb{E}[F_X(Y)] \neq \mathbb{E}[F_Y(X)]$ , where  $F_X$  is the distribution function of  $X$

- **Assumption on the repartition of the observations:**

- (B3)  $\min_{0 \leq k \leq L_1^*} |n_{1,k+1}^* - n_{1,k}^*| \wedge \min_{0 \leq k \leq L_2^*} |n_{2,k+1}^* - n_{2,k}^*| \geq n\delta_n$ .  
(M3) There exists  $0 < \tau_1^* < \dots < \tau_{L^*}^* < 1$ , such that for each  $\ell \in \{1, \dots, L^*\}$ ,  $n_\ell^*/n \xrightarrow{n} \tau_\ell^*$ .

• **Assumption for the LASSO:**

- (B4) There exists  $(\lambda_n)_{n \in \mathbb{N}}$  such that  $|\widehat{\mathcal{A}}_1(\lambda_n)| = L_1^*$ , that  $|\widehat{\mathcal{A}}_2(\lambda_n)| = L_2^*$  and

$$\frac{\lambda_n}{n\delta_n J_{\min}^*} \xrightarrow{n} 0.$$

Under these assumptions, we obtain the following theorem:

**Theorem 2.1.** *Under the assumptions (B1), (B2), (B3) and (B4), the estimator of the BlockSeg procedure is consistent.*

*Under the assumptions (M1), (M2) and (M3), the estimator of the MuChPoint procedure is consistent.*

The proofs are available respectively in [12] (Proposition 1 of Section 2.2) and in [15] (Theorem 4 of Section 3.1).

### 2.3.2. Complexity

Since the  $\mathbf{B}$  is sparse (due to the number of blocks that is fixed), the complexity of the BlockSeg procedure is linear with the number of cases of the matrix while the complexity of the MuChPoint procedure is slightly worse.

**Proposition 2.2.** *The complexities of the BlockSeg procedure is  $\mathcal{O}(n^2)$  and of the MuChPoint procedure is  $\mathcal{O}(n^3)$ .*

The procedures are implemented in the software  respectively in the packages [11] and [13].

## 2.4. Perspectives

In theory, the BlockSeg procedure is faster than the MuChPoint procedure but its consistency conditions are more restrictive. In practice, the model selection in the first procedure requires estimating several times the change-point while in the second a model selection criterion can be used. Moreover, it is possible to improve the estimation of the BlockSeg procedure using the symmetry of the matrix. It is also important to automate the selection of the number of breaks. These two perspectives are the object of a future work.

## 3. ON PARAMETER ESTIMATION IN NON-REGULAR SITUATIONS OF CUSP TYPE

In this section we present a survey of recent (and less recent) results on *change-point location estimation* for different observation models (i.i.d. observations, signal in white Gaussian noise, inhomogeneous Poisson processes, ergodic diffusion processes, *etc.*), in presence of a *cuspid-type singularity* (the function is continuous, but has an infinite derivative) in the driving function (density, signal, intensity function, drift, *etc.*) at the change-point. The survey follows a recent paper [27] by Dachian, Kordzakhia, Kutoyants and Novikov, though here we consider a slightly more general situation of asymmetric (having different multiplicative constants at the left and at the right) cusp. For each model, we study the asymptotic behavior of the *maximum likelihood estimator (MLE)*, as well as that of the *Bayesian estimators (BEs)*.

Let us recall that it is well-known that in regular statistical models (irrespectively of the nature of the observations), both the MLE and the BEs are asymptotically normal and asymptotically efficient. This comes from the fact that regular statistical models are locally asymptotically normal (LAN).

The situation is quite different for non-regular statistical models. For classical change-point models (the function is discontinuous at the change-point), the limiting distributions of the MLE and of the BEs are no longer Gaussian, and we can cite at least three different kinds of asymptotic behavior.

- For i.i.d. observations with discontinuous density (see Chernoff and Rubin [22], as well as Ibragimov and Khasminskii [46, 47, 50]) and for Poisson processes with discontinuous intensity function (see Kutoyants [57, 59]), the limiting distributions of the MLE and of the BEs are given by some random variables which are (different) functionals of a two-sided Poisson process, and the BEs are asymptotically efficient.
- For a discontinuous signal in white Gaussian noise (see Ibragimov and Khasminskii [49, 50]), for change-point type dynamical systems with small noise (see Kutoyants [57, 58]), for ergodic diffusion processes with discontinuous drift (see Kutoyants [57, 60]), for change-point type delay differential equations with noise (see K uchler and Kutoyants [56]), for a discontinuous signal in time inhomogeneous diffusion (see H opfner and Kutoyants [45]), as well as for many other change-point models, the limiting distributions of the MLE and of the BEs are given by some random variables which are (different) functionals of a two-sided Brownian motion (Wiener process), and the BEs are asymptotically efficient.
- For the two-phase regression model (see Koul and Qian [55] for the linear case, as well as Ciuperca [24] for the nonlinear case) and for the threshold auto-regressive (TAR) model (see K.S. Chan [18], as well as N.H. Chan and Kutoyants [20, 21]), the limiting distributions of the MLE and of the BEs are given by some random variables which are (different) functionals of a two-sided compound Poisson process, and the BEs are asymptotically efficient.

Below, we consider the problem of change-point location estimation in presence of a cusp-type singularity at the change-point for several models of observation. Interestingly, unlike the classical (discontinuous) change-point models, and somewhat like regular statistical models, the asymptotic behavior of the MLE, as well as that of the BEs, is the same through all the models: the limiting distributions of the MLE and of the BEs are given by some random variables which are (different) functionals of a (two-sided) fractional Brownian motion (fBm), and the BEs are asymptotically efficient.

In all the considered models, the asymptotic behavior of the estimators is studied using the *likelihood ratio analysis* method introduced by Ibragimov and Khasminskii in [50]. The method consist in first showing that the *normalized likelihood ratio process* (with a suitable normalization) converge to some *limiting likelihood ratio process*, and then deducing the properties of the estimators (and, namely, the limiting distributions of the MLE and of the BEs given by some functionals of the limiting likelihood ratio process, as well as the asymptotic efficiency of the BEs). It turns out that in all the considered models, the limiting likelihood ratio process is the same and is an exponential of a fBm with a power drift (see below for more details).

Though historically the first model in which a cusp-type singularity at the change-point was studied was the model of i.i.d. observations (which will be considered in Section 3.2), we will start our survey by presenting in Section 3.1 the signal in white Gaussian noise model, in which the fBm based limiting likelihood ratio process appears the most naturally. Further, in Sections 3.3 and 3.4, we will consider inhomogeneous Poisson processes and ergodic diffusion processes, respectively. Finally in Section 3.5, we present some numerical simulations concerning the limiting distributions of the MLE and of the BEs.

In conclusion, let us note that some other models of observations were equally studied in presence of a cusp-type singularity at the change-point by different authors. For nonlinear regression models we refer to Prakasa Rao [73], D oring [31], as well as to D oring and Jensen [32]; for delay differential equations with noise we refer to Gushchin and K uchler [40]; and for dynamical systems with small noise we refer to Kutoyants [61]. These authors still obtain the same (fBm based) asymptotic behavior of the MLE and of the BEs. So, the asymptotic behavior seems to be universal in presence of a cusp-type singularity at the change-point, and it would be interesting to study more observation models to confirm (or infirm) this conjecture.

### 3.1. Signal in white Gaussian noise

Suppose we observe a *deterministic signal in white Gaussian noise*, that is, a realization of a stochastic process  $X^{(\varepsilon)} = (X_t^{(\varepsilon)}, t \in [0, T])$  starting from  $X_0^{(\varepsilon)} = 0$  and satisfying the equation

$$dX_t^{(\varepsilon)} = S_\theta(t) dt + \varepsilon dW_t, \quad t \in [0, T], \quad (5)$$

where  $T > 0$  is the observation time,  $\varepsilon > 0$  is the noise level,  $\theta \in \Theta = (\alpha, \beta) \subset (0, T)$  is the unknown parameter, and  $S_\theta$  is the deterministic signal supposed to have the form

$$S_\theta(t) = d(t - \theta)|t - \theta|^p + h(t - \theta) \quad (6)$$

with some continuously differentiable function  $h$ . Here and throughout the survey, the process  $(W_t, t \in \mathbb{R})$  is a standard two-sided Brownian motion (Wiener process),  $0 < p < 1/2$  (note that for  $p > 1/2$ , the Fisher information is finite and the statistical model becomes regular) and

$$d(x) = \begin{cases} a, & \text{if } x < 0, \\ b, & \text{if } x > 0, \end{cases} \quad (7)$$

with  $a^2 + b^2 > 0$ .

We are interested in the estimation of the parameter  $\theta$  in *small noise asymptotics*, that is, as  $\varepsilon \rightarrow 0$ . Let us note that the *long observation time asymptotics* can be reduced to this case in the following way. Suppose we observe  $Y^{(n\tau)} = (Y_t^{(n\tau)}, t \in [0, n\tau])$  starting from  $Y_0^{(n\tau)} = 0$  and satisfying the equation (5) with  $T = n\tau$ ,  $\varepsilon = 1$  and some  $\tau$ -periodic signal  $S_\theta$  (the period  $\tau > 0$  is supposed to be known). Then, putting

$$X_t = \frac{1}{n} \sum_{j=1}^n (Y_{(j-1)\tau+t} - Y_{(j-1)\tau}), \quad 0 \leq t \leq \tau,$$

we obtain a stochastic process satisfying (5) with  $T = \tau$  and  $\varepsilon = n^{-1/2}$  (and a different noise Wiener process).

The *likelihood* of our model is given by (see, for example, Liptser and Shiryaev [66])

$$L(\theta, X^{(\varepsilon)}) = \exp \left\{ \frac{1}{\varepsilon^2} \int_0^T S_\theta(t) dX_t^{(\varepsilon)} - \frac{1}{2\varepsilon^2} \int_0^T S_\theta^2(t) dt \right\}, \quad \theta \in \Theta.$$

Using this likelihood, we can introduce the maximum likelihood estimator (MLE)  $\widehat{\theta}_\varepsilon$  and the Bayes estimator (BE)  $\widetilde{\theta}_\varepsilon$  for a given strictly positive and continuous prior density  $p$  on  $\Theta$  (and for quadratic loss function) by usual relations

$$L(\widehat{\theta}_\varepsilon, X^{(\varepsilon)}) = \max_{\theta \in \Theta} L(\theta, X^{(\varepsilon)}) \quad \text{and} \quad \widetilde{\theta}_\varepsilon = \frac{\int_{\Theta} \theta p(\theta) L(\theta, X^{(\varepsilon)}) d\theta}{\int_{\Theta} p(\theta) L(\theta, X^{(\varepsilon)}) d\theta}. \quad (8)$$

In order to describe the properties of the MLE and of the BEs, we need to introduce some more notations which will be used throughout the survey.

For any  $H \in (0, 1)$ , we denote  $(W_H(u), u \in \mathbb{R})$  the fractional Brownian motion (fBm) of Hurst parameter  $H$ , that is, a centered Gaussian process with continuous trajectories and having covariance function of the form

$$\mathbf{E}[W_H(u_1) W_H(u_2)] = \frac{1}{2} [ |u_1|^{2H} + |u_2|^{2H} - |u_1 - u_2|^{2H} ]. \quad (9)$$

Further, we introduce the processes

$$Z_H(u) = \exp \left\{ W_H(u) - \frac{1}{2} |u|^{2H} \right\}, \quad u \in \mathbb{R},$$

and, for any  $\gamma > 0$ , the process

$$Z_{\gamma, H}(u) = Z_H(\gamma u), \quad u \in \mathbb{R}.$$

We also introduce the random variables  $\widehat{\xi}_H$  and  $\widetilde{\xi}_H$  by the relations

$$Z_H(\widehat{\xi}_H) = \max_{u \in \mathbb{R}} Z_H(u) \quad \text{and} \quad \widetilde{\xi}_H = \frac{\int_{\mathbb{R}} u Z_H(u) du}{\int_{\mathbb{R}} Z_H(u) du}.$$

Note that the random variable  $\widehat{\xi}_H$  is well defined, since with probability one the process  $Z_H$  attains its maximum in a unique point (see, for example, Ermakov [34]). Note also that similar random variables associated to the process  $Z_{\gamma,H}$  are nothing but  $\widehat{\xi}_H/\gamma$  and  $\widetilde{\xi}_H/\gamma$ .

Finally for  $0 < p < 1/2$  and  $a, b$  from (7), we introduce the constant

$$I_p(a, b) = \int_{-\infty}^{+\infty} \left( d(s-1) |s-1|^p - d(s) |s|^p \right)^2 ds.$$

Note that this constant have the following explicit expressions (see Ibragimov and Khasminskii [50]):

$$\begin{aligned} I_p(a, b) &= \frac{\Gamma(1+p) \Gamma(\frac{1}{2}-p)}{2^{2p} \sqrt{\pi} (2p+1)} (a^2 + b^2 - 2ab \cos(\pi p)) \\ &= B(p+1, p+1) \left( \frac{a^2 + b^2}{\cos(\pi p)} - 2ab \right). \end{aligned}$$

Now we can state the main result of this section.

**Theorem 3.1.** Denote  $\gamma = (I_p(a, b))^{1/(2p+1)}$ .

(1) We have the following asymptotic lower bound on the mean squared error of an arbitrary estimator  $\bar{\theta}_\varepsilon$ :

$$\lim_{\delta \searrow 0} \liminf_{\varepsilon \rightarrow 0} \sup_{|\theta - \theta_0| \leq \delta} \varepsilon^{-\frac{4}{2p+1}} \mathbf{E}_\theta (\bar{\theta}_\varepsilon - \theta)^2 \geq \mathbf{E}(\widetilde{\xi}_{p+\frac{1}{2}}/\gamma)^2, \quad \theta_0 \in \Theta.$$

(2) The MLE  $\widehat{\theta}_\varepsilon$  and the BEs  $\widetilde{\theta}_\varepsilon$  are consistent, have the following limiting distributions:

$$\varepsilon^{-\frac{2}{2p+1}} (\widehat{\theta}_\varepsilon - \theta) \xrightarrow{\mathcal{L}} \widehat{\xi}_{p+\frac{1}{2}}/\gamma \quad \text{and} \quad \varepsilon^{-\frac{2}{2p+1}} (\widetilde{\theta}_\varepsilon - \theta) \xrightarrow{\mathcal{L}} \widetilde{\xi}_{p+\frac{1}{2}}/\gamma,$$

the convergence of moments holds for the above convergences in law, and the BEs are asymptotically efficient, that is,

$$\lim_{\delta \searrow 0} \liminf_{\varepsilon \rightarrow 0} \sup_{|\theta - \theta_0| \leq \delta} \varepsilon^{-\frac{4}{2p+1}} \mathbf{E}_\theta (\widetilde{\theta}_\varepsilon - \theta)^2 = \mathbf{E}(\widetilde{\xi}_{p+\frac{1}{2}}/\gamma)^2, \quad \theta_0 \in \Theta.$$

The proof of this theorem can be found in Chernoyarov, Dachian and Kutoyants [23] and consist in checking the condition of Theorems 1.9.1, 1.10.1 and 1.10.2 of Ibragimov and Khasminskii [50], which provide all the desired results. The main ingredient hidden behind this theorems is the weak convergence (in a suitable functional space) of the *normalized likelihood ratio process*  $(Z^{(\varepsilon)}(u), u \in \mathbb{R})$  to the *limiting likelihood ratio process*  $Z_{\gamma,p+\frac{1}{2}}$ . The process  $Z^{(\varepsilon)}$  is defined by

$$Z^{(\varepsilon)}(u) = \frac{L(\theta + u\varphi_\varepsilon, X^{(\varepsilon)})}{L(\theta, X^{(\varepsilon)})} = \frac{1}{\varepsilon} \int_0^T [S_{\theta+u\varphi_\varepsilon}(t) - S_\theta(t)] dW_t - \frac{1}{2\varepsilon^2} \int_0^T [S_{\theta+u\varphi_\varepsilon}(t) - S_\theta(t)]^2 dt$$

for  $u \in (\varphi_\varepsilon^{-1}(\alpha - \theta), \varphi_\varepsilon^{-1}(\beta - \theta)) \uparrow \mathbb{R}$ , where  $\varphi_\varepsilon = \varepsilon^{\frac{2}{2p+1}}$ , and can be continuously extended to the whole  $\mathbb{R}$  so that it decreases to 0 at  $\pm\infty$ .

Let us give an heuristic argument explaining the convergence of the process  $Z^{(\varepsilon)}$  to the process  $Z_{\gamma,p+\frac{1}{2}}$ . It is based on the following representations of the fBm (see Kordzakhia, Kutoyants, Novikov and Hin [52]).

**Lemma 3.2.** The process

$$Y(u) = I_p^{-1/2}(a, b) \int_{-\infty}^{+\infty} [d(v-u)|v-u|^p - d(v)|v|^p] dW_v, \quad u \in \mathbb{R},$$

is a fBm of Hurst parameter  $p + \frac{1}{2}$ .

*Proof.* Obviously,  $(Y(u), u \in \mathbb{R})$  is a centered Gaussian process with continuous trajectories. It remains to check that its covariance function is of the form (9). Indeed, denoting  $g_{v,u} = d(v-u)|v-u|^p - d(v)|v|^p$  for the sake of shortness, we can write

$$\begin{aligned} \mathbf{E}[Y(u_1)Y(u_2)] &= I_p^{-1}(a, b) \int_{-\infty}^{+\infty} g_{v,u_1} g_{v,u_2} dv \\ &= \frac{1}{2I_p(a, b)} \int_{-\infty}^{+\infty} g_{v,u_1}^2 dv + \frac{1}{2I_p(a, b)} \int_{-\infty}^{+\infty} g_{v,u_2}^2 dv - \frac{1}{2I_p(a, b)} \int_{-\infty}^{+\infty} [g_{v,u_2} - g_{v,u_1}]^2 dv \\ &= \frac{1}{2} (|u_1|^{2p+1} + |u_2|^{2p+1} - |u_2 - u_1|^{2p+1}). \end{aligned}$$

Here we used the elementary equality  $2ab = a^2 + b^2 - (a-b)^2$  and the change of variable  $v = su_1$  (resp.  $v = su_2$  and  $v = u_1 + s(u_2 - u_1)$ ) in the first (resp. second and third) integral.  $\square$

For the process  $Z^{(\varepsilon)}$ , denoting  $s_{t,u} = d(t - \theta - u\varphi_\varepsilon)$  for the sake of shortness and noting that for  $h$  defined in (6)

$$\frac{1}{\varepsilon^2} \int_0^T [h(t - \theta - u\varphi_\varepsilon) - h(t - \theta)]^2 dt \leq C \frac{\varphi_\varepsilon^2}{\varepsilon^2} \leq C \varepsilon^{\frac{4}{2p+1}-2} \rightarrow 0,$$

we obtain

$$\ln Z^{(\varepsilon)}(u) = \frac{1}{\varepsilon} \int_0^T [s_{t,u}|t - \theta - u\varphi_\varepsilon|^p - s_{t,0}|t - \theta|^p] dW_t - \frac{1}{2\varepsilon^2} \int_0^T [s_{t,u}|t - \theta - u\varphi_\varepsilon|^p - s_{t,0}|t - \theta|^p]^2 dt + o(1).$$

Further,

$$\begin{aligned} \frac{1}{2\varepsilon^2} \int_0^T [s_{t,u}|t - \theta - u\varphi_\varepsilon|^p - s_{t,0}|t - \theta|^p]^2 dt &= \frac{u|u|^{2p} \varphi_\varepsilon^{2p+1}}{2\varepsilon^2} \int_{-\frac{\theta}{u\varphi_\varepsilon}}^{\frac{T-\theta}{u\varphi_\varepsilon}} [d(s-1)|s-1|^p - d(s)|s|^p]^2 ds \\ &\rightarrow \frac{|u|^{2p+1}}{2} I_p(a, b) = \frac{1}{2} |\gamma u|^{2p+1}, \end{aligned}$$

where we used the change of variable  $t = \theta + u\varphi_\varepsilon s$ .

Similarly, using the change of variable  $t = \theta + \varphi_\varepsilon v$ , we obtain

$$\begin{aligned} \frac{1}{\varepsilon} \int_0^T [s_{t,u}|t - \theta - u\varphi_\varepsilon|^p - s_{t,0}|t - \theta|^p] dW_t &= \frac{\varphi_\varepsilon^{p+\frac{1}{2}}}{\varepsilon} \int_{-\frac{\theta}{\varphi_\varepsilon}}^{\frac{T-\theta}{\varphi_\varepsilon}} [d(v-u)|v-u|^p - d(v)|v|^p] d\overline{W}_v \\ &\xrightarrow{\mathcal{L}} I_p^{1/2}(a, b) \overline{W}_{p+\frac{1}{2}}(u) = W_{p+\frac{1}{2}}(\gamma u), \end{aligned}$$

where  $\overline{W}$  is a Wiener process, while  $\overline{W}_{p+\frac{1}{2}}$  and  $W_{p+\frac{1}{2}}$  are fBms of Hurst parameter  $p + \frac{1}{2}$ .

Combining all the above, we finally get

$$Z^{(\varepsilon)}(u) \xrightarrow{\mathcal{L}} \exp \left\{ W_{p+\frac{1}{2}}(\gamma u) - \frac{1}{2} |\gamma u|^{2p+1} \right\} = Z_{\gamma, p+\frac{1}{2}}(u),$$

which is the desired convergence.

### 3.2. I.i.d. observations

Suppose we observe an *i.i.d. sample*  $X^{(n)} = (X_1, \dots, X_n)$  with marginal density  $f_\theta(x) = f(x - \theta)$ , where  $\theta \in (\alpha, \beta) \subset \mathbb{R}$  is the unknown parameter, and the function  $f$  is given by

$$f(t) = h(t) \exp\{d(t)|t|^p\}, \quad t \in \mathbb{R},$$

with some continuously differentiable function  $h$  satisfying  $h(0) \neq 0$ . The problem of estimation of the parameter  $\theta$  was considered by Prakasa Rao in [72], by Ermakov in [34], as well as by Ibragimov and Khasminskii in [46, 48, 50].

The MLE  $\widehat{\theta}_n$  and the BE  $\widetilde{\theta}_n$  for a given strictly positive and continuous prior density  $p$  on  $\Theta$  are defined by usual relations similar to (8) with the likelihood

$$L(\theta, X^{(n)}) = \prod_{i=1}^n f(X_i - \theta), \quad \theta \in \Theta,$$

and their asymptotic behavior as  $n \rightarrow +\infty$  is described by the following theorem.

**Theorem 3.3.** Denote  $\gamma = (e^{h(0)} I_p(a, b))^{1/(2p+1)}$ .

(1) We have the following asymptotic lower bound on the mean squared error of an arbitrary estimator  $\bar{\theta}_n$ :

$$\lim_{\delta \searrow 0} \liminf_{n \rightarrow +\infty} \sup_{|\theta - \theta_0| \leq \delta} n^{\frac{2}{2p+1}} \mathbf{E}_\theta (\bar{\theta}_n - \theta)^2 \geq \mathbf{E}(\widetilde{\xi}_{p+\frac{1}{2}}/\gamma)^2, \quad \theta_0 \in \Theta.$$

(2) The MLE  $\widehat{\theta}_n$  and the BEs  $\widetilde{\theta}_n$  are consistent, have the following limiting distributions:

$$n^{\frac{1}{2p+1}} (\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \widehat{\xi}_{p+\frac{1}{2}}/\gamma \quad \text{and} \quad n^{\frac{1}{2p+1}} (\widetilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \widetilde{\xi}_{p+\frac{1}{2}}/\gamma,$$

the convergence of moments holds, and the BEs are asymptotically efficient.

For the proof of this theorem we refer to Ibragimov and Khasminskii [50, Chapter 6].

### 3.3. Inhomogeneous Poisson processes

Suppose we observe  $n$  independent realizations on the interval  $[0, T]$  of an *inhomogeneous Poisson process* having a strictly positive intensity function  $\lambda_\theta$  of the form

$$\lambda_\theta(t) = d(t - \theta)|t - \theta|^p + \Psi(\theta, t), \quad t \in [0, T],$$

where  $T > 0$  is fixed,  $\theta \in \Theta = (\alpha, \beta) \subset (0, T)$  is the unknown parameter, and the function  $\Psi$  is continuous and, uniformly in  $t$ , Hölder continuous of order  $\mu > p + \frac{1}{2}$  with respect to  $\theta$ . We denote the observations  $X^{(n)} = (X^1, \dots, X^n)$ , where  $X^i = \{X_t^i, t \in [0, T]\}$  is the  $i$ -th realization,  $i = 1, \dots, n$ .

The likelihood of this model is given by (see, for example, Liptser and Shiryaev [66])

$$L(\theta, X^{(n)}) = \exp \left\{ \sum_{i=1}^n \int_0^T \ln \lambda_\theta(t) dX_t^i - n \int_0^T [\lambda_\theta(t) - 1] dt \right\}, \quad \theta \in \Theta.$$

Using this likelihood, the MLE  $\widehat{\theta}_n$  and the BE  $\widetilde{\theta}_n$  for a given strictly positive and continuous prior density  $p$  on  $\Theta$  are defined by usual relations similar to (8), and their asymptotic behavior as  $n \rightarrow +\infty$  is described by the following theorem.

**Theorem 3.4.** Denote  $\gamma_\theta = \left( \frac{I_p(a,b)}{\Psi(\theta,\theta)} \right)^{1/(2p+1)}$ .

(1) We have the following asymptotic lower bound on the mean squared error of an arbitrary estimator  $\bar{\theta}_n$ :

$$\lim_{\delta \searrow 0} \liminf_{n \rightarrow +\infty} \sup_{|\theta - \theta_0| \leq \delta} n^{\frac{2}{2p+1}} \mathbf{E}_\theta (\bar{\theta}_n - \theta)^2 \geq \mathbf{E}(\tilde{\xi}_{p+\frac{1}{2}}/\gamma_{\theta_0})^2, \quad \theta_0 \in \Theta.$$

(2) The MLE  $\hat{\theta}_n$  and the BEs  $\tilde{\theta}_n$  are consistent, have the following limiting distributions:

$$n^{\frac{1}{2p+1}} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \hat{\xi}_{p+\frac{1}{2}}/\gamma_\theta \quad \text{and} \quad n^{\frac{1}{2p+1}} (\tilde{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \tilde{\xi}_{p+\frac{1}{2}}/\gamma_\theta,$$

the convergence of moments holds, and the BEs are asymptotically efficient.

For the proof of this theorem we refer to Dachian [26].

### 3.4. Ergodic diffusion processes

Suppose we observe a realization  $X^T = (X_t, t \in [0, T])$  of the *ergodic diffusion process*

$$dX_t = S_\theta(X_t) dt + \sigma(X_t) dW_t, \quad X_0, \quad t \in [0, T],$$

where  $T > 0$  is the observation time,  $\theta \in \Theta = (\alpha, \beta) \subset \mathbb{R}$  is the unknown parameter, and the drift  $S_\theta$  is supposed to be of the form  $S_\theta(x) = d(x - \theta)|x - \theta|^p + h(x - \theta)$  with some Hölder continuous of order  $\mu > p + \frac{1}{2}$  function  $h$ . Moreover, we suppose that the conditions  $\mathcal{ES}$ ,  $\mathcal{EM}$  and  $\mathcal{A}_0(\Theta)$  from [60] are fulfilled. These conditions guarantee, in particular, the existence and uniqueness of the solution of the above equation, as well as the existence of the invariant density

$$f_\theta(x) = \frac{1}{G(\theta) \sigma^2(x)} \exp \left\{ 2 \int_\theta^x \frac{S_\theta(v)}{\sigma^2(v)} dv \right\}, \quad x \in \mathbb{R},$$

where  $G(\theta) > 0$  is the normalizing constant.

The likelihood of this model is given by (see, for example, Liptser and Shiryaev [66])

$$L(\theta, X^T) = \exp \left\{ \int_0^T \frac{S_\theta(X_t)}{\sigma^2(X_t)} dX_t - \frac{1}{2} \int_0^T \frac{S_\theta^2(X_t)}{\sigma^2(X_t)} dt \right\}, \quad \theta \in \Theta.$$

Using this likelihood, the MLE  $\hat{\theta}_T$  and the BE  $\tilde{\theta}_T$  for a given strictly positive and continuous prior density  $p$  on  $\Theta$  are defined by usual relations similar to (8), and their asymptotic behavior as  $T \rightarrow +\infty$  is described by the following theorem.

**Theorem 3.5.** Denote  $\gamma_\theta = \left( \frac{I_p(a,b)}{G(\theta) \sigma^4(\theta)} \right)^{1/(2p+1)}$ .

(1) We have the following asymptotic lower bound on the mean squared error of an arbitrary estimator  $\bar{\theta}_n$ :

$$\lim_{\delta \searrow 0} \liminf_{T \rightarrow +\infty} \sup_{|\theta - \theta_0| \leq \delta} T^{\frac{2}{2p+1}} \mathbf{E}_\theta (\bar{\theta}_T - \theta)^2 \geq \mathbf{E}(\tilde{\xi}_{p+\frac{1}{2}}/\gamma_{\theta_0})^2, \quad \theta_0 \in \Theta.$$

(2) The MLE  $\hat{\theta}_T$  and the BEs  $\tilde{\theta}_T$  are consistent, have the following limiting distributions:

$$T^{\frac{1}{2p+1}} (\hat{\theta}_T - \theta) \xrightarrow{\mathcal{L}} \hat{\xi}_{p+\frac{1}{2}}/\gamma_\theta \quad \text{and} \quad T^{\frac{1}{2p+1}} (\tilde{\theta}_T - \theta) \xrightarrow{\mathcal{L}} \tilde{\xi}_{p+\frac{1}{2}}/\gamma_\theta,$$

the convergence of moments holds, and the BEs are asymptotically efficient.

For the proof of this theorem we refer to Dachian and Kutoyants [28].

### 3.5. Numerical simulations

It is interesting to compare the limiting mean squared errors of the MLE and of the BEs for different values of  $H = p + \frac{1}{2}$ . In Novikov, Kordzakhia and Ling [69], it was shown via numerical simulations that  $\mathbf{E} \widehat{\xi}_H^2$  can be essentially larger than  $\mathbf{E} \widetilde{\xi}_H^2$ . The results are presented in Figure 4 for  $H \in (0.4, 1]$ .

In Figure 5, we present the densities of the random variables  $\widehat{\xi}_H$  and  $\widetilde{\xi}_H$  obtained by numerical simulations in Kordzakhia, Kutoyants, Novikov and Hin [52]. Note that on Panel B:  $H = 0.5$ , the solid curve is plotted using the analytic expression of the density of the MLE. This is the only case where the density is known explicitly.

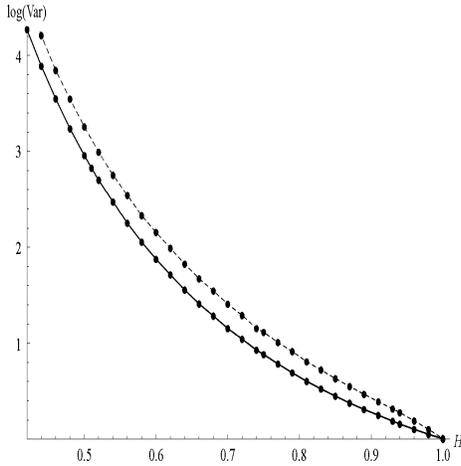


FIGURE 4. Plots of  $\ln \mathbf{E} \widehat{\xi}_H^2$  (solid curve) and  $\ln \mathbf{E} \widetilde{\xi}_H^2$  (dashed curve).

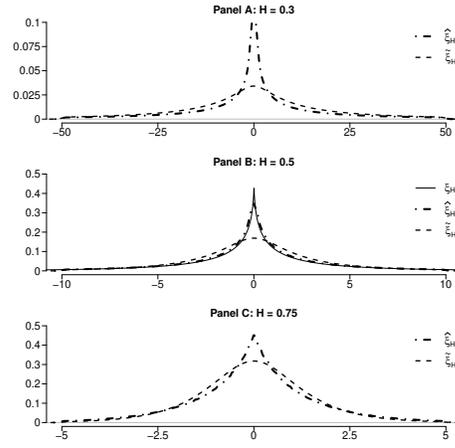


FIGURE 5. Densities of  $\widehat{\xi}_H$  and  $\widetilde{\xi}_H$  for  $H = 0.3$ ,  $H = 0.5$  and  $H = 0.75$ .

## 4. STATISTICAL INFERENCE FOR THRESHOLD D AUTOREGRESSIVE MODELS WITH INDEPENDENT INNOVATIONS

### 4.1. Introduction

The Threshold Autoregressive (TAR) models are introduced by [78] and were studied by many authors such that [18, 70, 71] and references therein. This model captures the dynamic behavior of time series by switching the regimes. The TAR model plays an important role in nonlinear time series and have been widely used to nonlinear phenomena in various fields, for example economics, environment, hydrology, physics, population dynamics, biological sciences, and among others. The TAR process is able to capture asymmetric limit cycles, as the main motivation for these models was to describe limit cycles of cyclical time-series [80]. For an update overview on TAR models, we can see [79]. The popularity of TAR models is due to the fact that they produce a simplified way of presenting a complex stochastic system in terms of decomposing it into a set smaller sub-system.

The main goal of TAR model related problems is to study the asymptotic properties of the estimated parameters and the estimated threshold. In [18], the author showed that under some regularity conditions, the least squares estimators of a stationary ergodic TAR models is strongly consistent. Qian, in [74], establishes the results similar to [18] for the maximum likelihood estimators for the same model under some regularity conditions on the errors density, not necessarily Gaussian. Moreover, [64] provided a numerical method to tabulate the limiting distribution of the estimated threshold in practice. In [41–43], the authors developed a statistical theory for threshold estimation in the context of regression. Under the assumption that the threshold effect is vanishingly small, he obtained the distribution and a parameter free limit of the estimated threshold.

Up to now, the existing papers have been treating the case where the models rely on strong assumptions on the noise processes, such as independence or martingale difference. A natural sequel is therefore to investigate the case where the strong hypothesis on the non-linear innovation do not hold. In other words, we will work in the framework of a noise sequence that is no more an independent sequence of random variables but just a sequence of uncorrelated random variables. This implies that the TAR process is no longer Markovian and is no more geometrically mixing. Therefore we adopt the framework of [36] in which the authors studied the Autoregressive and Moving-Average (ARMA) models, under a mixing property (and a stronger moment condition) on the observed process. They called these models weak ARMA models in opposition to strong ARMA models when the noise is an independent and identically distributed (iid for short) sequence. In link with weak ARMA models, we name these models weak TAR models. After the pioneering work of [36], many articles have been devoted to the study of weak models when one works with non independent innovation process. Nevertheless, to our knowledge, no study has addressed the question of weak TAR models.

In this work, we study the least square estimation (LSE) of weak TAR models and the asymptotic properties of the estimators. Under reasonable mixing assumptions for the time series process, we will prove that the LSE is strongly consistent and study the asymptotic laws. Although the consistency result is not really affected by our context, the asymptotic distribution needs further attention. For the parameters arising in the autoregressive formulation, we will be able to adapt the techniques of [36] using the mixing assumptions and as usual in this case, an extra moment assumption. Indeed, we shall require that the process has moments of order strictly greater than 4 whereas in the classical case, the fourth order moments are sufficient to investigate the asymptotic normality.

The asymptotic behaviour of the distribution of the threshold parameter is certainly the main novelty of our work. When the noise is strong (that means it is an iid sequence), the time series process is an ergodic Markov chain which is geometrically mixing. This is a stronger statement than the ones we do in our context because we will assume only  $\alpha$ -mixing property of the process. The results presented in [18, 64] heavily depend on the geometrically mixing property of the Markov chain. So we have to adapt their methodology in our case and this is feasible thanks to a weak convergence result of sums over triangular arrays to the compound Poisson limit under mixing assumptions. This technique is new in the time series context and we hope that it shall be used in other problems. All the proof of the announced results can be found in [33].

## 4.2. Model, assumptions and main results

A time series  $\{X_t\}_{t \in \mathbb{Z}}$  is said to be a weak TAR model if it satisfies

$$X_t = \begin{cases} \alpha_0 X_{t-1} + \varepsilon_t, & \text{for } X_{t-1} \leq r_0 \\ \beta_0 X_{t-1} + \varepsilon_t, & \text{for } X_{t-1} > r_0 \end{cases} \quad (10)$$

where the noise  $\varepsilon = (\varepsilon_t)_{t \in \mathbb{Z}}$  is a weak noise satisfying the following assumption.

**(H1):** The sequence  $(\varepsilon_t)_{t \in \mathbb{Z}}$  is strictly stationary, square integrable, and satisfies

- for any  $t$ ,  $\mathbb{E}(\varepsilon_t) = 0$
- for any  $s, t$ ,  $\mathbb{E}(\varepsilon_t \varepsilon_s) = \delta_{s,t} \sigma^2$  where  $\delta_{t,s} = 1$  if  $s = t$  and 0 otherwise
- for any  $t$ ,  $\mathbb{E}\varepsilon_t^4 < \infty$

In model (10),  $r_0$  is called the true threshold parameter and it is supposed to be unknown. Without loss of generality, we assume that there exist two finite constants  $\underline{r}, \bar{r}$  such that  $r_0 \in [\underline{r}, \bar{r}] := I$ . When  $r_0 = -\infty$  or  $r_0 = +\infty$ , the model reduces to a weak autoregressive (AR) model which is not of our interest.

The true parameter is denoted by  $\theta_0 = (\alpha_0, \beta_0, r_0)' \in \mathbb{R}^2 \times [\underline{r}, \bar{r}] = \mathbb{R}^2 \times I$  and a generic parameter is  $\theta = (\alpha, \beta, r) \in \mathbb{R}^2 \times I$ . We assume that the parameter space  $\Theta$  is a compact subspace of  $\mathbb{R}^2 \times I$ . We assume that  $\sigma$  is known (and equals 1). If  $\sigma$  is not known, it can be estimated by classical methods as soon as we know how to estimate the parameter  $\theta$ .

We will also use the restricted parameter  $\lambda = (\alpha, \beta)' \in \Lambda$ . One has  $\theta' = (\alpha, \beta, r) = (\lambda', r)$  and one may write that  $\Theta = \Lambda \times I$  where  $\Lambda$  is also assumed to be compact.

For any  $\theta = (\alpha, \beta, r) \in \Theta$ , we denote

$$\epsilon_t(\theta) = X_t - (\alpha + (\beta - \alpha)\mathbf{1}_{\{X_{t-1} > r\}})X_{t-1} = X_t - A_{t-1}(\theta)X_{t-1}, \tag{11}$$

where obviouslyly

$$A_{t-1}(\theta) = \alpha + (\beta - \alpha)\mathbf{1}_{\{X_{t-1} > r\}} = \beta + (\alpha - \beta)\mathbf{1}_{\{X_{t-1} \leq r\}}. \tag{12}$$

Let  $\{X_1, \dots, X_n\}$  be a sample from model (10) with the true parameter  $\theta_0$ . Given the initial value  $\mathcal{X}_0 = \{X_t; t \leq 0\}$  (that we may assume to be equal to 0 or, equivalently, that  $\epsilon_t(\theta) = 0$  for  $t \leq 0$ ), we consider the following sum of squares errors:

$$L_n(\theta) = \frac{1}{n} \sum_{t=1}^n \epsilon_t^2(\theta).$$

The minimizer  $\hat{\theta}_n$  of  $L_n(\theta)$  is called the least squares estimator of  $\theta_0$ , that is,

$$\hat{\theta}_n = \inf_{\theta \in \Theta} L_n(\theta)$$

Since the function  $L_n(\theta)$  is discontinuous in  $r$ , a way to obtain  $\hat{\theta}_n$  is as follows:

- for a fixed  $r$ , one minimizes  $L_n(\theta) = L_n(\lambda, r)$  and gets its minimizer  $\hat{\lambda}_n(r) = (\hat{\alpha}_n(r), \hat{\beta}_n(r))'$  and the minimum  $L_n^*(r) = L_n(\hat{\alpha}_n(r), \hat{\beta}_n(r), r)$ ,
- since  $L_n^*(r)$  takes only a finite number of possible values, its minimizer is easy to find, which will immediately lead to the estimator  $\hat{\theta}_n$ .

Throughout this work, we assume that  $|\alpha| + |\beta| < 1$ . This condition is sufficient to ensure the invertibility of the model (see Theorem A1 in [65]). Even if we do not use the invertibility in our proof, since the main assumptions will be made on the process  $X$ , this condition is necessary to prove that the initial values  $\mathcal{X}_0$  will not affect the asymptotic properties of the estimator  $\hat{\theta}_n$  (see [64] for further details).

Before stating the result on convergence of  $\hat{\theta}_n$  towards  $\theta$ , we will need further assumptions on the process  $(X_t)_{t \in \mathbb{Z}}$ .

- (H2):** The process  $(X_t)_{t \in \mathbb{Z}}$  is ergodic, strictly stationary and has fourth order moments. Moreover, for any  $t$ , the probability distribution function of  $X_t$  is absolutely continuous. Its density is denoted by  $\pi$  and is bounded away from 0 and  $\infty$  over each bounded set. The function  $\pi$  is also assumed to be a Lipschitz function.

Supposing that  $|\alpha| + |\beta| < 1$  implies the ergodicity in **(H2)** as it was noticed in [18]. See also the work of Chan and Tong [19] for some general sufficient conditions for stationarity and ergodicity.

We will also need the following hypothesis.

- (H3):** The threshold  $r_0$  in  $\mathbb{R}$  is the discontinuity point of autoregressive function, that is  $(\beta_0 - \alpha_0) \neq 0$ .

The above hypothesis is natural because if  $\alpha_0 = \beta_0$ , then our model becomes a simple auto-regressive AR(1) model. Under this assumption, we can state the following consistency result.

**Theorem 4.1.** *Let  $(X_t)_{t \in \mathbb{Z}}$  be the TAR process satisfying (10). We assume that **(H1)**, **(H2)** and **(H3)** hold. Then  $\hat{\theta}_n \rightarrow \theta_0$  a.s., as  $n \rightarrow +\infty$ .*

The proof of this result is classical. The fact that the noise is a weak noise does not affect the arguments of [18] and [74].

Next, we study the limiting distribution of  $\hat{\theta}_n$ .

We introduce the notation related to the mixing property that will be assumed. First, recall that for two random variables  $X$  and  $Y$ , the mixing coefficient  $\alpha(X, Y)$  is defined by

$$\alpha(X, Y) = \sup_{A \in \sigma(X), B \in \sigma(Y)} |P(A \cap B) - P(A)P(B)|,$$

where  $\sigma(X)$  is the sigma-field generated by  $X$ . We will make use of the Davydov inequality (see [29] or [37]) that states that for  $p, q$  and  $r$  three positive numbers such that  $1/p + 1/q + 1/r = 1$ , there exists a constant  $K$  such that we have

$$|\text{Cov}(X, Y)| \leq K \|X\|_{\mathbb{L}^p} \|Y\|_{\mathbb{L}^q} |\alpha(X, Y)|^{\frac{1}{r}}. \quad (13)$$

Now, let  $\mathcal{F}_{-\infty}^t$  and  $\mathcal{F}_{t+k}^{\infty}$  be the  $\sigma$ -fields generated by  $\{X_u : u \leq t\}$  and  $\{X_u : u \geq t+k\}$ , respectively. The strong mixing property coefficients  $(\alpha_X(k))_{k \in \mathbb{N}^*}$  of the stationary process  $(X_t)$  are defined by

$$\alpha_X(k) = \sup_{A \in \mathcal{F}_{-\infty}^t, B \in \mathcal{F}_{t+k}^{\infty}} |P(A \cap B) - P(A)P(B)|. \quad (14)$$

We formulate the following hypothesis.

**(H4):**  $(X_t)_{t \in \mathbb{Z}}$  satisfies the strong mixing condition: there exists  $\nu > 0$  such that

$$\sum_{k=0}^{\infty} \{\alpha_X(k)\}^{\frac{\nu}{2+\nu}} < \infty. \quad (15)$$

The above strong mixing condition will be used hereafter by means of the Davydov inequality (13). Therefore, the following moment condition will be also needed.

**(H5):**  $(X_t)_{t \in \mathbb{Z}}$  satisfies  $\mathbb{E}|X_t|^{4+2\nu} < \infty$  with the real  $\nu$  from Assumption **(H4)**.

When the noise is iid, the time series process satisfies a geometric mixing property which is stronger than our assumptions **(H4)** and **(H5)** and plays an important role in the work of [18, 64] and in all related works using their techniques.

In order to state the asymptotic normality result we need the following notation. We recall that we have denoted  $\lambda = (\alpha, \beta)$  the restricted parameter being such that  $\theta = (\lambda, r) \in \Lambda \times I = \Theta$ .

**Theorem 4.2.** *If Assumptions **(H1)** to **(H5)** hold, then  $\hat{\lambda}_n(\hat{r}_n) = \begin{pmatrix} \hat{\alpha}_n(\hat{r}_n) \\ \hat{\beta}_n(\hat{r}_n) \end{pmatrix}$  satisfies*

$$\sqrt{n}(\hat{\lambda}_n(\hat{r}_n) - \lambda_0) = \sqrt{n}(\hat{\lambda}_n(r_0) - \lambda_0) + o_{\mathbb{P}}(1)$$

and  $\sqrt{n}(\hat{\lambda}_n(r_0) - \lambda_0)$  has a normal limiting distribution with mean 0 and covariance matrix  $J^{-1}IJ^{-1}$  with

$$J = 2 \begin{pmatrix} \mathbb{E}(X_1^2 \mathbf{1}_{\{X_1 \leq r_0\}}) & 0 \\ 0 & \mathbb{E}(X_1^2 \mathbf{1}_{\{X_1 > r_0\}}) \end{pmatrix} \quad \text{and} \quad I = \lim_{n \rightarrow \infty} \text{Var} \left( \sqrt{n} \frac{\partial L_n(\lambda_0, r_0)}{\partial \lambda} \right).$$

### 4.3. Asymptotic behaviour of $\hat{r}_n$

Now, we study the limiting distribution of  $n(\hat{r}_n - r_0)$ . The arguments follow the ideas of [18]. Nevertheless, many precisions about this method are given in [64] and we further need to give more specific details due to our context.

In order to explain our the strategy, we consider the following profile sum of squares errors function defined for  $s \in \mathbb{R}$  by

$$\tilde{\phi}_n(s) = nL_n(\hat{\lambda}_n(r_0 + s/n), r_0 + s/n) - nL_n(\hat{\lambda}_n(r_0), r_0). \quad (16)$$

Suppose that the sequence of processes  $((\tilde{\phi}_n(s))_{s \in \mathbb{R}})_{n \geq 1}$  converges in the Skorohod space  $\mathbb{D}(\mathbb{R})$  of c.a.d.l.a.g. functions on  $\mathbb{R}$  (details will be given hereafter) to a process  $(\phi(s))_{s \in \mathbb{R}}$ . Then one uses a continuity result on the Skorohod space established in [76, 77]. This result asserts that the argmin will also converge to the argmin of the process  $(\phi(s))_{s \in \mathbb{R}}$  (if it exists).

To prove this convergence, we show that the sequence  $(\tilde{\phi}_n(\cdot))_{n \geq 1}$  can be approximated in  $\mathbb{D}(\mathbb{R})$  by the sequence of processes  $(\phi_n(\cdot))_{n \geq 1}$  defined by

$$\phi_n(s) = nL_n(\alpha_0, \beta_0, r_0 + s/n) - nL_n(\alpha_0, \beta_0, r_0). \quad (17)$$

Using (11) and easy calculus, one may write  $\phi_n$  as

$$\phi_n(s) = \sum_{t=1}^n \zeta_{1,t}(s) \mathbf{1}_{\{s < 0\}} + \sum_{t=1}^n \zeta_{2,t}(s) \mathbf{1}_{\{s \geq 0\}} \quad (18)$$

$$\zeta_{1,t}(s) = \left( 2X_t X_{t-1} (\alpha_0 - \beta_0) + X_{t-1}^2 (\beta_0^2 - \alpha_0^2) \right) \mathbf{1}_{\{r_0 + s/n < X_{t-1} \leq r_0\}} \quad (19)$$

$$\zeta_{2,t}(s) = \left( 2X_t X_{t-1} (\beta_0 - \alpha_0) + X_{t-1}^2 (\alpha_0^2 - \beta_0^2) \right) \mathbf{1}_{\{r_0 < X_{t-1} \leq r_0 + s/n\}}. \quad (20)$$

We also denote

$$\Gamma_{1,t} = 2X_t X_{t-1} (\alpha_0 - \beta_0) + X_{t-1}^2 (\beta_0^2 - \alpha_0^2) \quad (21)$$

$$\Gamma_{2,t} = 2X_t X_{t-1} (\beta_0 - \alpha_0) + X_{t-1}^2 (\alpha_0^2 - \beta_0^2). \quad (22)$$

We will prove that the  $\phi_n$  converges to a two sided compound Poisson process  $\phi$  in the Skorohod space. In order to define the limiting process, one introduces  $F_1(\cdot|r_0)$  the conditional distribution of  $\Gamma_{1,t}$  given  $X_{t-1} = r_0^-$  and  $F_2(\cdot|r_0)$  the conditional distribution of  $\Gamma_{2,t}$  given  $X_{t-1} = r_0^+$ . This measure exists and is the limiting conditional distribution of  $\Gamma_{2,t}$  given  $\{r_0 < X_{t-1} \leq r_0 + \delta\}$  as  $\delta \downarrow 0$ . Analogously,  $F_1(\cdot|r_0)$  exists as the limiting conditional distribution of  $\Gamma_{1,t}$  given  $\{r_0 - \delta < X_{t-1} \leq r_0\}$  as  $\delta \downarrow 0$ . The existence of this limit follows from the results of Neveu (see [68], page 124). By stationarity,  $F_2(\cdot|r_0)$  is also the conditional distribution of  $\Gamma_{2,2} = 2X_2 X_1 (\beta_0 - \alpha_0) + X_1^2 (\alpha_0^2 - \beta_0^2)$  given  $X_1 = r_0^+$ .

We define a two-sided compound Poisson process (CPP)  $(\phi(s))_{s \in \mathbb{R}}$  as follows:

$$\phi(s) = \phi_1(-s) \mathbf{1}_{\{s < 0\}} + \phi_2(s) \mathbf{1}_{\{s \geq 0\}},$$

where  $\{\phi_1(s), s \geq 0\}$  and  $\{\phi_2(s), s \geq 0\}$  are two independent Poisson processes with  $\phi_1(0) = \phi_2(0) = 0$  a.s., with the same jump rate  $\pi(r_0) > 0$ , where  $\pi(x)$  is the density of  $X_1$ .

As soon as we have proved that  $\phi_n$  converges to the two sided compound Poisson process  $\phi$  in the Skorohod space, we use Theorem 3.1 of [77]. Then there exists a unique random interval  $[M_-, M_+]$  on which the process  $\phi$  attains its global minimum a.s. and then  $n(\hat{r}_n - r_0)$  converges to  $M_-$ .

Now we can state our convergence result but we need an additional mixing assumption of the process  $X$ .

**(H6):** There exists a real  $a$  with  $\nu/(2 + \nu) < a < 1$ , a constant  $C$  and a real  $0 < \beta < 1$  such that for any  $u, r \in I$  we have

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \sum_{j; |j-k| \leq n^a, j \neq k} \mathbb{E} \left( \mathbf{1}_{\{r < X_{k-1} \leq r+1/n\}} \mathbf{1}_{\{r < X_{j-1} \leq r+1/n\}} \right) = 0 \quad \text{and} \quad (23)$$

$$\sum_{k=1}^n \sum_{j; |j-k| \leq n^a, j \neq k} \mathbb{E} \left( \mathbf{1}_{\{r_0+r/n < X_{j-1} \leq r_0+u/n\}} \mathbf{1}_{\{r_0+r/n < X_{k-1} \leq r_0+u/n\}} \right) \leq C(u-r)^\beta. \quad (24)$$

Let us make a few comments about this assumption. By stationarity, (23) is equivalent to

$$\lim_{n \rightarrow \infty} n \sum_{h=1}^{n^a} \mathbb{E} \left( \mathbf{1}_{\{r < X_1 \leq r+1/n\}} \mathbf{1}_{\{r < X_{h+1} \leq r+1/n\}} \right) = 0 .$$

This is a local mixing assumption and is clearly satisfied in the independent case and if **(H2)** is satisfied. In the context of [18, 64], it is deduced from a conditional argument and the Markovian context which implies that the process is geometrically mixing. Hence **(H6)** is a technical assumption but we emphasize the fact that this condition is written in the same spirit as Assumption (II) in [9] so it is quite natural in our non Markovian context.

Now we can state our other main result as follows:

**Theorem 4.3.** *We suppose that Assumptions **(H1)** to **(H6)** hold and that the density  $\pi$  is Lipschitz. Then  $n(\hat{r}_n - r_0) \rightarrow M_-$  and  $n(\hat{r}_n - r_0)$  is asymptotically independent of  $\sqrt{n}(\hat{\alpha}_n(r_0) - \alpha_0, \hat{\beta}_n(r_0) - \beta_0)'$  which is always asymptotically normally distributed (regardless of whether  $r_0$  is known or not).*

## REFERENCES

- [1] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *J. Mach. Learn. Res.*, 10:245–279, June 2009.
- [2] J. Bai. Estimation of multiple-regime regressions with least absolute deviation. *Journal of Statistical Planning and Inference*, 74:103–134, 10 1998.
- [3] J. Bai and P. Perron. Estimating and testing linear models with multiple structural changes. *Econometrica*, 66:47–78, 02 1998.
- [4] J.-M. Bardet and C. Dion. Robust semi-parametric multiple change-point detection. *Signal Processing*, 156, 11 2018.
- [5] J.-M. Bardet and A. Guenaizi. Data-driven semi-parametric detection of multiple changes in long-range dependent processes. *arXiv e-prints*, page arXiv:1801.02515, Jan 2018.
- [6] J.-M. Bardet, W. Kengne, and O. Wintenberger. Multiple breaks detection in general causal time series using penalized quasi-likelihood. *Electron. J. Statist.*, 6:435–477, 2012.
- [7] J.-M. Bardet and O. Wintenberger. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.*, 37:2730–2759, 2009.
- [8] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
- [9] S. M. Berman. A compound Poisson limit for stationary sums, and sojourns of Gaussian processes. *Ann. Probab.*, 8(3):511–538, 1980.
- [10] M. Birem, J.-C. Quinton, F. Berry, and Y. Mezouar. Sail-map: Loop-closure detection using saliency-based features. In *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4543–4548. IEEE, 2014.
- [11] V. Brault and J. Chiquet. blockseg: Two dimensional change-points detection. <https://cran.r-project.org/web/packages/blockseg/index.html>, 2018.
- [12] V. Brault, J. Chiquet, and C. Lévy-Leduc. Efficient block boundaries estimation in block-wise constant matrices: An application to hic data. *Electron. J. Statist.*, 11(1):1570–1599, 2017.
- [13] V. Brault, G. Cougoulat, S. Ouadah, and L. Sansonnet. Muchpoint: Multiple change point. <https://CRAN.R-project.org/package=MuchPoint>, 2018.
- [14] V. Brault, A. Leclercq-Samson, and J.-C. Quinton. Segmentation des lignes et des colonnes d’une matrice : application aux séquences de navigation visuelle. In *49èmes Journées de Statistique*, Avignon, France, May 2017. Société Française de Statistique.
- [15] V. Brault, S. Ouadah, L. Sansonnet, and C. Lévy-Leduc. Nonparametric multiple change-point estimation for analyzing large hi-c data matrices. *Journal of Multivariate Analysis*, 165:143 – 165, 2018.
- [16] V. Brault, A. Samson, and J.-C. Quinton. Modélisation statistique pour détecter des séquences vidéos similaires : application aux véhicules autonomes. In *Congrès de la Société Mathématique de France - SMF 2018*, Lille, France, June 2018. Société Mathématique de France.
- [17] B. E. Brodsky and B. S. Darkhovsky. *Nonparametric methods in change-point problems*. Kluwer Acad. Publ., the Netherlands, 1993.
- [18] K. S. Chan. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Ann. Statist.*, 21(1):520–533, 1993.
- [19] K. S. Chan and H. Tong. On the use of the deterministic Lyapunov function for the ergodicity of stochastic difference equations. *Adv. in Appl. Probab.*, 17(3):666–678, 1985.

- [20] N. H. Chan and Yu. A. Kutoyants. Recent developments of threshold estimation for nonlinear time series. *J. Japan Statist. Soc.*, 40(2):277–303, 2010.
- [21] N. H. Chan and Yu. A. Kutoyants. On parameter estimation of threshold autoregressive models. *Stat. Inference Stoch. Process.*, 15(1):81–104, 2012.
- [22] H. Chernoff and H. Rubin. The estimation of the location of a discontinuity in density. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I*, pages 19–37. University of California Press, Berkeley and Los Angeles, 1956.
- [23] O. V. Chernoyarov, S. Dachian, and Yu. A. Kutoyants. On parameter estimation for cusp-type signals. *Ann. Inst. Statist. Math.*, 70(1):39–62, 2018.
- [24] G. Ciuperca. Maximum likelihood estimator in a two-phase nonlinear random regression model. *Statist. Decisions*, 22(4):335–349, 2004.
- [25] M. Csörgő and L. Horváth. *Limit theorems in change-point analysis*. Wiley Series in Probability and Statistics, 1997.
- [26] S. Dachian. Estimation of cusp location by Poisson observations. *Stat. Inference Stoch. Process.*, 6(1):1–14, 2003.
- [27] S. Dachian, N. Kordzakhia, Yu. Kutoyants, and A. Novikov. Estimation of cusp location of stochastic processes: a survey. *Stat. Inference Stoch. Process.*, 21(2):345–362, 2018.
- [28] S. Dachian and Yu. A. Kutoyants. On cusp estimation of ergodic diffusion process. *J. Statist. Plann. Inference*, 117(1):153–166, 2003.
- [29] J. A. Davydov. The convergence of distributions which are generated by stationary random processes. *Teor. Veroyatnost. i Primenen.*, 13:730–737, 1968.
- [30] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.
- [31] M. Döring. Asymmetric cusp estimation in regression models. *Statistics*, 49(6):1279–1297, 2015.
- [32] M. Döring and U. Jensen. Smooth change point estimation in regression models with random design. *Ann. Inst. Statist. Math.*, 67(3):595–619, 2015.
- [33] M. Elmi and B. Sausseureau. Inference on threshold autoregressive models with dependent errors. Submitted, 2018.
- [34] M. S. Ermakov. On the asymptotic behavior of statistical estimates for samples having a density with singularities. *Theory Probab. Appl.*, 21(3):649–651, 1977.
- [35] C. Faure, J.-M. Bardet, J. Lacaille, and M. Olteanu. Comparison of three algorithms for parametric change-point detection. In *Proceeding ESANN 2016*, 04 2016.
- [36] C. Francq and J.-M. Zakoïan. Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference*, 68(1):145–165, 1998.
- [37] C. Francq and J.-M. Zakoïan. *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley, 2010.
- [38] K. Frick, A. Munk, and H. Sieling. Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 76(3):495–580, 2014. With discussions and a rejoinder by the authors.
- [39] P. Fryzlewicz. Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42(6):2243–2281, 12 2014.
- [40] A. A. Gushchin and U. Küchler. On estimation of delay location. *Stat. Inference Stoch. Process.*, 14(3):273–305, 2011.
- [41] B. E. Hansen. Inference in TAR models. *Stud. Nonlinear Dyn. Econom.*, 2(1):1–14, 1997.
- [42] B. E. Hansen. Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603, 2000.
- [43] B. E. Hansen. Threshold autoregression in economics. *Stat. Interface*, 4(2):123–127, 2011.
- [44] Z. Harchaoui and C. Lévy-Leduc. Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association*, 105(492):1480–1493, 2010.
- [45] R. Höpfner and Yu. A. Kutoyants. On LAN for parametrized continuous periodic signals in a time inhomogeneous diffusion. *Statist. Decisions*, 27(4):309–326, 2009.
- [46] I. A. Ibragimov and R. Z. Khasminskii. The asymptotic behavior of generalized Bayesian estimates. *Dokl. Akad. Nauk SSSR*, 194(2):257–260, 1970.
- [47] I. A. Ibragimov and R. Z. Khasminskii. Asymptotic behavior of statistical estimates for samples with a discontinuous density. *Math. USSR-Sb.*, 16(4):573–606, 1972.
- [48] I. A. Ibragimov and R. Z. Khasminskii. The asymptotical behavior of location parameter statistical estimators for samples with continuous density with singularities. *Zap. Nauchn. Sem. LOMI*, 41:67–93, 166–167, 1974. Problems of the theory of probability distributions, II.
- [49] I. A. Ibragimov and R. Z. Khasminskii. Parameter estimation for a discontinuous signal in white Gaussian noise. *Probl. Inf. Transm.*, 11(3):203–212, 1975.
- [50] I. A. Ibragimov and R. Z. Khasminskii. *Statistical estimation. Asymptotic theory*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York-Berlin, 1981.
- [51] S. Kay. *Fundamentals of statistical signal processing: detection theory*. Prentice-Hall, Inc., 1993.
- [52] N. E. Kordzakhia, Yu. A. Kutoyants, A. A. Novikov, and L.-Y. Hin. On limit distributions of estimators in irregular statistical models and a new representation of fractional Brownian motion. *Statist. Probab. Lett.*, 139:141–151, 2018.
- [53] A. P. Korostelev and A. B. Tsybakov. *Minimax Theory of Image Reconstruction*. Springer New York, 1993.

- [54] H. Korrapati, J. Courbon, S. Alizon, and F. Marmoiton. "the institut pascal data sets": un jeu de données en extérieur, multicapteurs et datées avec réalité terrain, données d'étalonnage et outils logiciels. In *Orasis, Congrès des jeunes chercheurs en vision par ordinateur*, 2013.
- [55] H. L. Koul and L. Qian. Asymptotics of maximum likelihood estimator in a two-phase linear regression model. *J. Statist. Plann. Inference*, 108(1-2):99–119, 2002. C. R. Rao 80th birthday felicitation volume, Part II.
- [56] U. Küchler and Yu. A. Kutoyants. Delay estimation for some stationary diffusion-type processes. *Scand. J. Statist.*, 27(3):405–414, 2000.
- [57] Yu. A. Kutoyants. *Parameter estimation for stochastic processes*, volume 6 of *Research and Exposition in Mathematics*. Heldermann Verlag, Berlin, 1984. Translated from the Russian and edited by B. L. S. Prakasa Rao.
- [58] Yu. A. Kutoyants. *Identification of dynamical systems with small noise*, volume 300 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.
- [59] Yu. A. Kutoyants. *Statistical inference for spatial Poisson processes*, volume 134 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.
- [60] Yu. A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Series in Statistics. Springer-Verlag London, Ltd., London, 2004.
- [61] Yu. A. Kutoyants. On cusp location estimation for perturbed dynamical system. *Scand. J. Stat.*, 2019. to appear.
- [62] M. Lavielle and E. Moulines. Least-squares estimation of an unknown number of shifts in time series. *Journal of Time Series Analysis*, 21:33 – 59, 01 2000.
- [63] E. Lebarbier. Detecting multiple change-points in the mean of gaussian process by model selection. *Signal Process.*, 85(4):717–736, Apr. 2005.
- [64] D. Li, S. Ling, and W. K. Li. Asymptotic theory on the least squares estimation of threshold moving-average models. *Econometric Theory*, 29(3):482–516, 2013.
- [65] S. Ling and H. Tong. Testing for a linear MA model against threshold MA models. *Ann. Statist.*, 33(6):2529–2552, 2005.
- [66] R. S. Liptser and A. N. Shiryaev. *Statistics of random processes. II. Applications*, volume 6 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 2nd expanded edition, 2001.
- [67] A. Lung-Yut-Fong, C. Lévy-Leduc, and O. Cappé. Homogeneity and change-point detection tests for multivariate data using rank statistics. *Journal de la Société Française de Statistique*, 156(4):133–162, 2015.
- [68] J. Neveu. *Mathematical foundations of the calculus of probability*. Translated by Amiel Feinstein. Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1965.
- [69] A. Novikov, N. Kordzakhia, and T. Ling. On moments of Pitman estimators: the case of fractional Brownian motion. *Theory Probab. Appl.*, 58(4):601–614, 2014.
- [70] J. D. Petrucci. On the consistency of least squares estimators for a threshold AR(1) model. *J. Time Ser. Anal.*, 7(4):269–278, 1986.
- [71] J. D. Petrucci and S. W. Woolford. A threshold AR(1) model. *J. Appl. Probab.*, 21(2):270–286, 1984.
- [72] B. L. S. Prakasa Rao. Estimation of the location of the cusp of a continuous density. *Ann. Math. Statist.*, 39(1):76–87, 1968.
- [73] B. L. S. Prakasa Rao. Estimation of cusp in nonregular nonlinear regression models. *J. Multivariate Anal.*, 88(2):243–251, 2004.
- [74] L. Qian. On maximum likelihood estimators for a threshold autoregression. *J. Statist. Plann. Inference*, 75(1):21–46, 1998.
- [75] P. M. Robinson. Gaussian semiparametric estimation of long range dependence. *Ann. Statist.*, 23(5):1630–1661, 10 1995.
- [76] E. Seijo and B. Sen. Change-point in stochastic design regression and the bootstrap. *Ann. Statist.*, 39(3):1580–1607, 2011.
- [77] E. Seijo and B. Sen. A continuous mapping theorem for the smallest argmax functional. *Electron. J. Stat.*, 5:421–439, 2011.
- [78] H. Tong. *Threshold models in nonlinear time series analysis*, volume 21 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1983.
- [79] H. Tong. *Nonlinear time series*, volume 6 of *Oxford Statistical Science Series*. The Clarendon Press, Oxford University Press, New York, 1990. A dynamical system approach, With an appendix by K. S. Chan, Oxford Science Publications.
- [80] R. S. Tsay. Testing and modeling threshold autoregressive processes. *J. Amer. Statist. Assoc.*, 84(405):231–240, 1989.
- [81] L. Vostrikova. Detecting 'disorder' in multidimensional random processes. *Soviet Math. Dokl.*, 24:55–59, 1981.