Electronic Journal of Statistics
Vol. 6 (2012) 435–477
ISSN: 1935-7524
DOI: 10.1214/12-EJS680

Multiple breaks detection in general causal time series using penalized quasi-likelihood

Jean-Marc Bardet and William Kengne

SAMM, Université Paris 1 Panthéon-Sorbonne, France
e-mail: Jean-Marc.Bardet@univ-paris1.fr; William-Charky.Kengne@malix.univ-paris1.fr

Olivier Wintenberger

CEREMADE, Université Paris-Dauphine, France
e-mail: wintenberger@ceremade.dauphine.fr

Abstract: This paper is devoted to the off-line multiple breaks detection for a general class of models. The observations are supposed to fit a parametric causal process (such as classical models AR(∞), ARCH(∞) or TARCH(∞)) with distinct parameters on multiple periods. The number and dates of breaks, and the different parameters on each period are estimated using a quasi-likelihood contrast penalized by the number of distinct periods. For a convenient choice of the regularization parameter in the penalty term, the consistency of the estimator is proved when the moment order r of the process satisfies $r \geq 2$. If $r \geq 4$, the length of each approximative segment tends to infinity at the same rate as the length of the true segment and the parameters estimators on each segment are asymptotically normal. Compared to the existing literature, we added the fact that a dependence is possible over distinct periods. To be robust to this dependence, the chosen regularization parameter in the penalty term is larger than the ones from BIC approach. We detail our results which notably improve the existing ones for the AR(∞), ARCH(∞) and TARCH(∞) models. For the practical applications (when n is not too large) we use a data-driven procedure based on the slope estimation to choose the penalty term. The procedure is implemented using the dynamic programming algorithm. It is an $O(n^2)$ complexity algorithm that we apply on AR(1), AR(2), GARCH$(1,1)$ and TARCH(1) processes and on the FTSE index data.

AMS 2000 subject classifications: Primary 62M10, 62F12.
Keywords and phrases: Change detection, causal processes, ARCH(∞) processes, AR(∞) processes, quasi-maximum likelihood estimator, model selection by penalized likelihood.

Received May 2011.

Contents

1. Introduction

The breaks detection is a classical problem as well as in the statistic than in the signal processing community. The first important result in this topic was obtained by Page [20] in 1955 and real advances have been done during the seventies, notably with the results of Hinkley (see for instance [12]) and the break detection became a distinct and important area of research in statistic (see the book of Basseville and Nikiforov [3] for a large overview).

Two approaches are generally considered for solving a problem of breaks detection: an 'on-line' approach leading to sequential estimation and an 'off-line' approach when the series of observations is complete. Concerning this last approach, numerous results were obtained for independent random variables in a parametric frame (see for instance Bai and Perron [1]). The case of the off-line detection of multiple change-points in a parametric or semiparametric frame for dependent variables or time series also provided an important literature. The present paper is a new contribution to this problem.

In this paper, we consider the following change-point problem: for $j = 1, 2, \ldots, K^*$,

$$X_t = g_{\theta_j^*}(\xi_t, X_{t-1}, X_{t-2}, \ldots) \quad \text{for all } t \in \{t_{j-1}^* + 1, t_{j-1}^* + 2, \ldots, t_j^*\} \tag{1}$$

where g_θ is a parametric function satisfying assumptions detailed in Section 2, $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of centered independent and identically distributed (iid) $\mathbb{R}^p$-random vectors called the innovations, $K^*-1\in\mathbb{N}$ is the unknown number of the breaks, $t_0^*=0<t_1^*<\cdots<t_{K^*-1}^*<n=t_{K^*}^*$ with $(t_j^*)_{1\le j\le K^*-1}\in\mathbb{N}$ are the K^*-1 unknown dates of the breaks, $\theta_j^*\in\Theta\subset\mathbb{R}^d$ for $j=1,\dots,K^*$ are the unknown parameters of the model. Note that the assumptions on g_θ are weaker enough for X to be for instance AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH or bilinear processes on each period.

The aim of our statistical procedure is the estimation of the unknown parameters $(K^*,(t_j^*)_{1\le j\le K^*-1},(\theta_j^*)_{1\le j\le K^*})$ in the problem (1). In the literature, it is generally supposed that X is a stationary process on each set $\{t_{j-1}^*+1\dots,t_j^*\}$ and is independent on each $\{t_{i-1}^*+1,\dots,t_i^*\}$ of the other $\{t_{k-1}^*+1,\dots,t_k^*\}$, $k\neq i$ (for instance in [18], [14] and [7]). Here the problem (1) does not induce such assumptions and thus the framework is closer to the applications, see Remark 1 in [7].

In the problem of change-points detection, numerous papers were devoted to the CUSUM procedure (see for instance Kokozska and Leipus [14] in the specific case of ARCH(∞) processes). In Lavielle and Ludena [17] a "Whittle" contrast is used for estimating the breaks dates in the spectral density of piecewise long-memory processes (in a semi-parametric framework). Davis *et al.* [6] proposed a likelihood ratio as the estimator of breaks for an AR(p) process. Lavielle and Moulines [18] consider a general contrast using the mean square errors for estimating the parameters. In Davis *et al.* [7], the criteria called Minimum Description Length (MDL) is applied to a large class of nonlinear time series.

We consider here a semiparametric estimator based on a penalized contrast (so-called $penQLIK$ in the sequel) using the quasi-likelihood function. For usual stationary time series, the conditional quasi-likelihood (so-called $QLIK$ in the sequel) is constructed as follow:

1. Assume the process $(\xi_t)_{t\in\mathbb{Z}}$ is a Gaussian sequence and compute the conditional likelihood (with respect to $\sigma\{X_0,X_{-1},\dots\}$) based on the unobservable infinite realization of $(X_t)_{t\in\mathbb{Z}}$;
2. Approximate this computation for a sample $(X_1,\dots,X_n)$;
3. Apply this approximation even if the process of the innovations is not a Gaussian sequence.

The quasi-maximum likelihood estimator (QMLE) obtained by maximizing the $QLIK$ has convincing asymptotic properties in the case of GARCH processes (see Jeantheau [13], Berkes *et al.* [5], Franck and Zakoian [10]) or generalizations of GARCH processes (see Mikosch and Straumann [22], Robinson and Zaffaroni [21]). Bardet and Wintenberger [2] study the asymptotic normality of the QMLE of θ applied to the class of models considered here. Thus, when K^* is known, a natural estimator of the parameter $(\underline{t}^*,\underline{\theta}^*)=((t_j^*)_{1\le j\le K^*-1},(\theta_j^*)_{1\le j\le K^*})$ for a process satisfying (1) is the QMLE on every intervals $[t_j+1,\dots,t_{j+1}]$ and every parameters θ_j for $1\le j\le K^*$. However we consider here that K^* is unknown and such method cannot be directly used. The chosen solution is to penalize the contrast by an additional term $\kappa_n K$, where the regularization parameters κ_n

form an increasing sequence of real numbers (see the final expression of the penalized contrast in (4)). Such procedure of penalization was previously used for instance by Yao [23] to estimate the number of change-points with the Schwarz criterion and by Lavielle and Moulines [18]. Hence the minimization of the penalized contrast leads to an estimator (see (5)) of the parameters $(K^*, \underline{t}^*, \underline{\theta}^*)$.

Classical heuristics such as the BIC one lead to choose $\kappa_n \propto \log n$. In our study, such penalties terms are excluded in some cases, when the models in (1) are very dependent on their whole past, see Section 3 (and simulation results) for more details. Roughly speaking, an explanation of this can be provided by the simple relation:

$$\begin{aligned} penQLIK(K, \underline{t}, \underline{\theta}) &= QLIK(K, \overline{t}, \underline{\theta}) + \kappa_n K \\ &= \Big(QLIK(K, \underline{t}, \underline{\theta}) - \widetilde{QLIK}(K, \underline{t}, \underline{\theta}) \Big) + \widetilde{QLIK}(K, \underline{t}, \underline{\theta}) + \kappa_n K \end{aligned}$$

where $\widetilde{QLIK}$ is the conditional quasi-likelihood of a process following (1) except that it is composed by stationary time series on each period which are independent of the stationary processes defined on the other periods. Using moment bounds we will prove in Section 6 that $\big| QLIK(K, \underline{t}, \underline{\theta}) - \widetilde{QLIK}(K, \underline{t}, \underline{\theta}) \big| = O_P(u_n)$ with $u_n \to \infty$ and $u_n/n \to 0$, where $(u_n)_{n \in \mathbb{N}}$ depends on the Lipshitzian behavior of g_θ. Since $\widetilde{QLIK}(K, \underline{t}, \underline{\theta}) \sim C\, n \quad a.s.$ when $n \to \infty$ from results obtained in [2], it is clear that the penalty term can play a role only if $\kappa_n >> u_n$. Finally, we will show that under weak conditions on the model, the regularization parameter $\kappa_n \propto \sqrt{n}$ over-penalizes the number of breaks for avoiding artificial breaks in cases of models very dependent on their whole past (see Section 3 for details). Such a choice of κ_n is robust to the (possibly strong) dependence.

The main results of the paper are the following: under Lipshitzian condition on g_θ and when the moments of order $r \geq 2$ of the innovations and X are finite, the estimator $\big(\widehat{K}_n, (\widehat{t}_j/n)_{1 \leq j \leq \widehat{K}_n - 1}, (\widehat{\theta}_j)_{1 \leq j \leq \widehat{K}_n}\big)$ is consistent. If moreover Lipshitzian conditions are also satisfied by the derivatives of g_θ and if $r \geq 4$, then the convergence rate of $(\widehat{t}_j/n)_{1 \leq j \leq \widehat{K}_n - 1}$ is $O_P(w_n)$ for any sequence $(w_n)_n$ such that $w_n >> n^{-1}$ and a Central Limit Theorem (CLT) for $(\widehat{\theta}_j)_{1 \leq j \leq \widehat{K}_n}$ with a $\sqrt{n}$-convergence rate is established. These results are "optimal" in the sense that the convergence rate is the same than in an independent setting.

After detailing the particular cases of AR(∞), ARCH(∞) and TARCH(∞) satisfying the break-point problem (1), the estimator is applied to generated trajectories of such time series. Two difficulties appeared. Firstly, the computation time was very long and exponentially increased with K. We solved this problem by using a dynamic programming algorithm which is a $O(n^2)$ complexity algorithm. We also considered only small length trajectories ($n \leq 2000$). Secondly, we obtained the consistency of the estimator of K^* as a theoretical result in all considered model regardless of their dependence properties when $\kappa_n \propto \sqrt{n}$ when $n \to \infty$. We will see that for particular models such that ARMA(p, q) or GARCH(p, q) a BIC-type penalty with $\kappa_n \propto \log n$ is also possible, but $\kappa_n \propto \sqrt{n}$ ensures the convergence for a larger class of models (including AR(∞), ARCH(∞) or TARCH(∞) processes).

However, for n not too large (for instance $n = 1000$) the choice of $\kappa_n = \sqrt{n}$ very often led to $\widehat{K}_n \neq K^*$. Hence we chose to implement a data-driven procedure for estimating κ_n (denoted $\hat{\kappa}_n$ in the sequel) using a slope estimation method (see [4]), such procedure being nowadays often used in the model selection frame. In such a way, the results of simulations are clearly satisfying (see Section 5). The estimation procedure is also applied to financial data and this provides estimating dates of breaks corresponding with key dates of financial crisis.

The following Section 2 is devoted to the assumptions and the study of the existence of a nonstationary solution of the change point problem (1). The definition of the estimator and its asymptotic properties are studied in Section 3. The particular examples of AR(∞), ARCH(∞) and TARCH(∞) processes are detailed in Section 4, while the concrete estimation procedure and numerical applications are presented in Section 5. Finally, Section 6 contains the main proofs.

2. Assumptions and existence of a non-stationary solution

2.1. Notation and assumptions

Let $\theta \in \mathbb{R}^d$ and M_θ and f_θ be real-valued measurable functions such that for all $(x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $M_\theta\big((x_i)_{i\in\mathbb{N}}\big) \neq 0$. In this paper, we consider a general class $\mathcal{M}_T(M_\theta, f_\theta)$ of causal (non-anticipative) time series. Let $T \subset \mathbb{Z}$ and $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of centered independent and identically distributed (iid) random variables called the innovations and satisfying $\mathrm{var}(\xi_0) = 1$. Define

Class $\mathcal{M}_T(M_\theta, f_\theta)$: *The process $X = (X_t)_{t\in\mathbb{Z}}$ belongs to $\mathcal{M}_T(M_\theta, f_\theta)$ if it satisfies the relation:*

$$X_{t+1} = M_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big)\xi_t + f_\theta\big((X_{t-i})_{i\in\mathbb{N}}\big) \quad \text{for all } t \in T. \tag{2}$$

The existence and properties of these general affine processes were studied in Bardet and Wintenberger [2] as a particular case of chains with infinite memory considered in Doukhan and Wintenberger [8]. Numerous classical real valued time series are included in $\mathcal{M}_{\mathbb{Z}}(M, f)$: for instance AR($\infty$), ARCH($\infty$), TARCH($\infty$), ARMA-GARCH or bilinear processes.

For obtaining conditions of existence of a process included in $\mathcal{M}_T(M_\theta, f_\theta)$ first define the following different norms:

1. $\|\cdot\|$ applied to a vector denotes the Euclidean norm of the vector;
2. for any compact set $\mathcal{K} \subseteq \mathbb{R}^d$ and for any $g : \mathcal{K} \longrightarrow \mathbb{R}^{d'}$; $\|g\|_{\mathcal{K}} = \sup_{\theta\in\mathcal{K}}(\|g(\theta)\|)$;
3. for all $x = (x_1, \ldots, x_K) \in \mathbb{R}^K$, $\|x\|_m = \max_{i=1,\ldots,K} |x_i|$;
4. if Y is a random vvector with finite r-order moments, we set $\|Y\|_r = (\mathbb{E}\|Y\|^r)^{1/r}$.

Let $\Psi_\theta = M_\theta$, f_θ and $i = 0,\, 1,\, 2$, then for any compact set $\mathcal{K} \subseteq \mathbb{R}^d$, define

Assumption A$_i(\Psi_\theta, \mathcal{K})$**:** *Assume that* $\|\partial^i \Psi_\theta(0)/\partial\theta^i\|_{\mathcal{K}} < \infty$ *and there exists a sequence of non-negative real numbers* $(\alpha_k^{(i)}(\Psi_\theta, \mathcal{K}))_{k\ge 1}$ *such that* $\sum_{k=1}^\infty \alpha_k^{(i)} \times (\Psi_\theta, \mathcal{K}) < \infty$ *satisfying*

$$\Big\| \frac{\partial^i \Psi_\theta(x)}{\partial\theta^i} - \frac{\partial^i \Psi_\theta(y)}{\partial\theta^i} \Big\|_{\mathcal{K}} \le \sum_{k=1}^{\infty} \alpha_k^{(i)}(\Psi_\theta, \mathcal{K}) |x_k - y_k| \quad \text{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

In the sequel we refer to the particular case called "ARCH-type process", if $f_\theta = 0$ and the following assumption holds on $h_\theta = M_\theta^2$:

Assumption A$_i(h_\theta, \mathcal{K})$**:** *Assume that* $f_\theta = 0$, $\|\partial^i h_\theta(0)/\partial\theta^i\|_{\mathcal{K}} < \infty$ *and there exists a sequence of non-negative real numbers* $(\alpha_k^{(i)}(h_\theta, \mathcal{K}))_{k\ge 1}$ *satisfying* $\sum_{k=1}^\infty \alpha_k^{(i)}(h_\theta, \mathcal{K}) < \infty$ *and*

$$\Big\| \frac{\partial^i h_\theta(x)}{\partial\theta^i} - \frac{\partial^i h_\theta(y)}{\partial\theta^i} \Big\|_{\mathcal{K}} \le \sum_{k=1}^{\infty} \alpha_k^{(i)}(h_\theta, \mathcal{K}) |x_k^2 - y_k^2| \quad \text{for all } x, y \in \mathbb{R}^{\mathbb{N}}.$$

Now, for any $i = 0, 1, 2$ and any compact $\mathcal{K} \subset \mathbb{R}^d$, under Assumptions $A_i(f_\theta, \mathcal{K})$ and $A_i(M_\theta, \mathcal{K})$, denote:

$$\beta^{(i)}(\mathcal{K}) := \sum_{k\ge 1} \beta_k^{(i)}(\mathcal{K}) \quad \text{where} \quad \beta_k^{(i)}(\mathcal{K}) := \alpha_k^{(i)}(f_\theta, \mathcal{K}) + (\mathbb{E}|\xi_0|^r)^{1/r} \alpha_k^{(i)}(M_\theta, \mathcal{K}),$$

and under Assumption $A_i(h_\theta, \mathcal{K})$

$$\tilde\beta^{(i)}(\mathcal{K}) := \sum_{k\ge 1} \tilde\beta_k^{(i)}(\mathcal{K}) \quad \text{where} \quad \tilde\beta_k^{(i)}(\mathcal{K}) := (\mathbb{E}|\xi_0|^r)^{2/r} \alpha_k^{(i)}(h_\theta, \mathcal{K}).$$

The dependence with respect to r of the coefficients $\beta^{(i)}$ and $\tilde\beta^{(i)}$ are omitted for notational convenience. From now on let us fix Θ a compact subset of $\mathbb{R}^d$ satisfying some contraction properties:

Assumption A: *Assume there exists* $r \ge 1$ *such that for all* $\theta \in \Theta$ *either* $A_0(f_\theta, \{\theta\})$ *and* $A_0(M_\theta, \{\theta\})$ *hold with* $\beta^{(0)}(\{\theta\}) < 1$ *either* $f_\theta = 0$ *and* $A_0(h_\theta, \{\theta\})$ *holds with* $\tilde\beta^{(0)}(\{\theta\}) < 1$.

From [2] we have:

Proposition 2.1. *If* $\theta \in \Theta$ *satisfies* A *for some* $r \ge 1$*, there exists a unique causal (non anticipative, i.e.* X_t *is independent of* $(\xi_i)_{i>t}$ *for* $t \in \mathbb{Z}$*) solution* $X = (X_t)_{t\in\mathbb{Z}} \in \mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$ *which is stationary, ergodic and satisfies* $\|X_0\|_r < \infty$.

The assumption A is classical when studying the existence of stationary solution of general models. For instance, Duflo [9] used such a Lipschitz-type inequality to show the existence of Markov chains. The elements of the compact set Θ satisfies one Lipschitz-type condition specified either for general causal models either for ARCH-type models. This distinction is adequate as for ARCH-type models $A_0(h_\theta, \{\theta\})$ is less restrictive than $A_0(M_\theta, \{\theta\})$. Remark that assumption $\tilde{\beta}^{(0)}(\theta) < 1$ is optimal for the stationarity of order $r \geq 1$ but not for the strict stationarity of the solution of an ARCH-type model.

Let $\theta \in \Theta$ and $X = (X_t)_{t \in \mathbb{Z}}$ a stationary solution included in $\mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$. For studying QMLE properties, it is convenient to assume the following assumptions:

Assumption D(Θ)**:** $\exists \underline{h} > 0$ such that $\inf_{\theta \in \Theta}(|h_\theta(x)|) \geq \underline{h}$ for all $x \in \mathbb{R}^{\mathbb{N}}$.

Assumption Id(Θ)**:** For all $\theta, \theta' \in \Theta$,

$$\Big(f_\theta(X_0, X_{-1}, \ldots) = f_{\theta'}(X_0, X_{-1}, \ldots) \text{ and } \\ h_\theta(X_0, X_{-1}, \ldots) = h_{\theta'}(X_0, X_{-1}, \ldots) \text{ a.s.} \Big) \ \Rightarrow \ \theta = \theta'.$$

Assumption Var(Θ)**:** For all $\theta \in \Theta$, one of the families $\left(\frac{\partial f_\theta}{\partial \theta^i}(X_0, X_{-1}, \ldots)\right)_{1 \leq i \leq d}$ or $\left(\frac{\partial h_\theta}{\partial \theta^i}(X_0, X_{-1}, \ldots)\right)_{1 \leq i \leq d}$ is a.e. linearly independent.

Assumption D(Θ) will be required to define the QMLE, Id(Θ) to show the consistence of the QMLE and Var(Θ) to show the asymptotic normality.

2.2. Existence of the solution to the change-point problem

Using the class $\mathcal{M}_T(M_\theta, f_\theta)$, the problem (1) of change-point detection can be formulated as follows: assume that a trajectory $(X_1, \ldots, X_n)$ of $X = (X_t)_{t \in \mathbb{Z}}$ is observed where

$$X \in \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*}) \quad \text{for all } j = 1, \ldots, K^*, \qquad \text{with} \tag{3}$$

- $K^* \in \mathbb{N}^*$, $T_j^* = \{t_{j-1}^* + 1, t_{j-1}^* + 2, \ldots, t_j^*\}$ with $0 < t_1^* < \cdots < t_{K^*-1}^* < n$, $t_j^* \in \mathbb{N}$ and by convention $t_0^* = -\infty$ and $t_{K^*}^* = \infty$;
- $\theta_j^* = (\theta_{j,1}^*, \ldots, \theta_{j,\,d}^*) \in \Theta \subset \mathbb{R}^d$ for $j = 1, \ldots, K^*$.

Consider the problem (3). Then the past of X before the time $t = 0$ depends on θ_1^* and the future after $t = n$ depends on $\theta_{K^*}^*$. The number $K^* - 1$ of breaks, the instants $t_1^*, \ldots, t_{k^*-1}^*$ of breaks and parameters $\theta_1^*, \ldots, \theta_{K^*}^*$ are unknown. Consider first the following notation.

Notation.

- *For* $K \geq 2$, $\mathcal{F}_K = \big\{ \underline{t} = (t_1, \ldots t_{K-1}) \ ; \ 0 < t_1 < \cdots < t_{K-1} < n \big\}$. *In particular,* $\underline{t}^* = \big(t_1^*, \ldots, t_{K^*-1}^*\big) \in \mathcal{F}_{K^*}$ *is the true vector of instants of change;*
- *For* $K \in \mathbb{N}^*$ *and* $\underline{t} \in \mathcal{F}_K$, $T_k = \big\{ t \in \mathbb{Z}, t_{k-1} < t \leq t_k \big\}$ *and* $n_k = Card(T_k)$ *with* $1 \leq k \leq K$. *In particular;* $T_j^* = \big\{ t \in \mathbb{Z}, \ t_{j-1}^* < t \leq t_j^* \big\}$ *and* $n_j^* = Card(T_j^*)$ *for* $1 \leq j \leq K^*$. *For all* $1 \leq k \leq K$ *and* $1 \leq j \leq K^*$, *let* $n_{kj} = Card(T_j^* \cap T_k)$*;*

The following proposition establishes the existence of the non stationary solution of the problem (3) and its moments properties.

Proposition 2.2. *Consider the problem* (3) *with* $\theta_j^* \in \Theta$ *for all* $j = 1, \ldots, K^*$, Θ *satisfying* A *for some* $r \geq 1$. *Then*

(i) there exists a solution $X = (X_t)_{t\in\mathbb{Z}}$ *of the model* (3) *and* X *is a causal time series.*
(ii) there exists a constant $C > 0$ *such that for all* $t \in \mathbb{Z}$ *we have* $\|X_t\|_r \leq C$.

The problem (3) distinguishes the case $t \in T_1^* = \{1, \ldots, t_1^*\}$ to the other ones since it is easy to see that $(X_t)_{t\in T_1^*}$ is a stationary process while $(X_t)_{t>t_1^*}$ is not. However, all the results of this paper hold if $(X_t)_{t\in T_1^*}$ is defined as the other $(X_t)_{t\in T_j^*}$, $j \geq 2$ (by defining a break in $t = 0$ setting $X_t = 0$ for $t \leq 0$ for instance).

3. The estimation procedure and the asymptotic behavior of the estimator

3.1. The penalized QLIK contrast

The estimation procedure of the number of breaks $K^* - 1$, the instants of breaks $\underline{t}^*$ and the parameters $\underline{\theta}^*$ is based on the minimum of a penalized QLIK contrast.

By definition, if $X \in \mathcal{M}_T(f_\theta, M_\theta)$ then the conditional (to the past values of X) mean and the variance are given by, respectively, $f_\theta(X_{s-1}, \ldots)$ and $h_\theta(X_{s-1}, \ldots)$. Therefore, with the notation $f_\theta^s = f_\theta\big(X_{s-1}, X_{s-2}\ldots\big)$, $M_\theta^s = M_\theta\big(X_{s-1}, X_{s-2}\ldots\big)$ and $h_\theta^s = {M_\theta^s}^2$, we deduce the quasi-likelihood of X on a period T:

$$L_n(T, \theta) := -\frac{1}{2} \sum_{s\in T} q_s(\theta) \ \text{ with } \ q_s(\theta) := \frac{(X_s - f_\theta^s)^2}{h_\theta^s} + \log\left(h_\theta^s\right).$$

By convention, we set $L_n(\emptyset, \theta_k) := 0$. Since only $X_1, \ldots, X_n$ are observed, $L_n(T, \theta)$ cannot be computed because it depends on the past values $(X_{-j})_{j\in\mathbb{N}}$. We approximate it by the QLIK criteria on a period T:

$$\widehat{L}_n(T, \theta) := -\frac{1}{2} \sum_{s\in T} \widehat{q}_s(\theta) \quad \text{where} \quad \widehat{q}_s(\theta) := \frac{\big(X_s - \widehat{f}_\theta^s\big)^2}{\widehat{h}_\theta^s} + \log\big(\widehat{h}_\theta^s\big)$$

with $\widehat{f}_\theta^t = f_\theta\big(X_{t-1}, \ldots, X_1, u\big)$, $\widehat{M}_\theta^t = M_\theta\big(X_{t-1}, \ldots, X_1, u\big)$ and $\widehat{h}_\theta^t = (\widehat{M}_\theta^t)^2$ for any deterministic sequence $u = (u_n)$ with finitely many non-zero values.

Remark 3.1. For convenience, in the sequel we choose $u = (u_n)_{n\in\mathbb{N}}$ with $u_n = 0$ for all $n \in \mathbb{N}$ as in [10] or in [2]. Indeed, this choice has no effect on the asymptotic behavior of estimators.

Now, for any number of periods $K \geq 1$, any instants of breaks $\underline{t} \in \mathcal{F}_K$ and any parameters on each periods $\underline{\theta} \in \Theta^K$, the global QLIK contrast $\widehat{J}_n$ is given by the expression:

$$(QLIK) \quad \widehat{J}_n(K, \underline{t}, \underline{\theta}) := -2 \sum_{k=1}^{K} \widehat{L}_n(T_k, \theta_k).$$

Since K^* has to be estimated, define the QLIK contrast penalized by the number of periods, called $penQLIK$ contrast, by

$$(penQLIK) \quad \widetilde{J}_n(K, \underline{t}, \underline{\theta}) := \widehat{J}_n(K, \underline{t}, \underline{\theta}) + \kappa_n \, K \tag{4}$$

where $\kappa_n \leq n$ is called the regularization parameter and will be fixed later. Suppose that an upper bound $K_{\max} > 0$ of the number of periods is known. Our estimator is defined as one of the minimizers of the penalized contrast:

$$(\widehat{K}_n, \underline{\widehat{t}}_n, \underline{\widehat{\theta}}_n) \in \underset{1 \leq K \leq K_{\max}}{\operatorname{Argmin}} \ \underset{(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta^K}{\operatorname{Argmin}} \big(\widetilde{J}_n(K, \underline{t}, \underline{\theta}) \big) \quad \text{and} \quad \underline{\widehat{\tau}}_n = \frac{\underline{\widehat{t}}_n}{n}. \tag{5}$$

As very often in model selection problems, the whole estimation procedure deeply depends on the choice of the regularization parameters (κ_n). To be robust in possible dependence over distinct periods, the regularization parameters (κ_n) have to be carefully chosen:

Assumption $\mathbf{H}_i$ $(i = 0, 1, 2)$**:** *For $0 \leq p \leq i$, the assumptions $A_p(f_\theta, \Theta)$, $A_p(M_\theta, \Theta)$ (or respectively $A_p(h_\theta, \Theta)$) hold and $\theta^*_j \in \Theta$ for all $j = 1, \ldots, K^*$, Θ satisfying A for some $r \geq 1$. Denoting $c^* > 0$ a real number satisfying*

$$c^* = \min_{j=1,\ldots,K^*} \big(-\log(\beta^{(0)}(\theta^*_j))/8 \big) \quad \textit{or resp.} \quad \min_{j=1,\ldots,K^*} \big(-\log(\tilde{\beta}^{(0)}(\theta^*_j))/8 \big)$$

the regularization parameters (κ_n) used in (4) satisfy $\kappa_n \wedge n \, \kappa_n^{-1} \to \infty$ with $n \to \infty$ and for all $j = 1, \ldots K^*$:

$$\sum_{k \geq 2} \kappa_k^{-(r/4 \wedge 1)} \Big(\sum_{\ell \geq k c^* / \log(k)} \beta^{(p)}_\ell(\Theta) \Big)^{(r/4 \wedge 1)} < \infty \tag{6}$$

$$\text{or resp.} \qquad \sum_{k \geq 2} \kappa_k^{-(r/4 \wedge 1)} \Big(\sum_{\ell \geq k c^* / \log(k)} \tilde{\beta}^{(p)}_\ell(\Theta) \Big)^{(r/4 \wedge 1)} < \infty. \tag{7}$$

The assumption H_i is interesting as it links the decrease rate of the Lipschitz coefficients and the penalty term of (4). The classical BIC corresponds to regularization parameters of the order of $\log(n)$. This choice is possible if the Lipschitz coefficients decrease exponentially fast, which hold for all models $\mathrm{M}(f_\theta, M_\theta)$ with finite order (see below). However, if the decrease of the Lipschitz coefficients is polynomial only regularization parameters satisfying $\kappa_n >> \log(n)$ satisfy H_i. Moreover, whatever the decay of the Lipschitzian coefficients, the estimation is more robust (with respect to the dependence over distinct segments) for the largest regularization parameter. More precisely, consider the following two paradigmatic examples for which (κ_n) satisfies conditions (7) (also used in [16]):

(1) geometric case: if $\alpha_\ell^{(i)}(f_\theta,\Theta)+\alpha_\ell^{(i)}(M_\theta,\Theta)+\alpha_\ell^{(i)}(h_\theta,\Theta)=O(a^\ell)$ with $0\le a<1$, then any choice of regularization parameters (κ_n) such that $\kappa_n\to\infty$ and $\kappa_n=o(n)$, satisfy (7) (for instance κ_n of order $\log n$ as in the BIC approach).

(2) Riemanian case: if $\alpha_\ell^{(i)}(f_\theta,\Theta)+\alpha_\ell^{(i)}(M_\theta,\Theta)+\alpha_\ell^{(i)}(h_\theta,\Theta)=O(\ell^{-\gamma})$ with $\gamma>1$,
- if $\gamma>1+(1\vee 4r^{-1})$, then any choice of (κ_n) such that $\kappa_n\to\infty$ and $\kappa_n=o(n)$ satisfy (7).
- if $(1\vee 4r^{-1})<\gamma\le 1+(1\vee 4r^{-1})$, then any choice of (κ_n) such that $O(\kappa_n)=n^{1-\gamma+(1\vee 4r^{-1})}(\log n)^\delta$ with $\delta>\gamma-1+(1\vee 4r^{-1})$ and $\kappa_n=o(n)$ can be chosen. However any of these choices satisfy $\kappa_n>>\log n$.

Remark 3.2. The sequence (δ_n) with $\delta_n:=nc^*/\log n$ appearing in (7) is the size of "small" blocks that are excluded from the original observations to deal with the possible dependence between period. It is the theoretically size below which we do not distinguish the breaks due to the dependence. This size depends on the real model and is unknown.

3.2. Consistency of the estimators

For establishing the consistency, we add the couple of following classical assumptions in the problem of break detection:

Assumption B: $\min_{j=1,\dots,K^*-1}\|\theta^*_{j+1}-\theta^*_j\|>0$.

Furthermore, the distance between instants of breaks cannot be too small:

Assumption C: *there exists a vector* $\underline{\tau}^*=(\tau_1^*,\dots,\tau^*_{K^*-1})$ *with* $0<\tau_1^*<\dots<\tau^*_{K^*-1}<1$ *called the vector of breaks such that for* $j=1,\dots,K^*$, $t_j^*=[n\tau_j^*]$ *(where* $[x]$ *is the floor of* x*). The is called the vector of breaks.*

Even if the length of T_j^* has asymptotically the same order than n, the dependence with respect to n of t_j^*, t_k, T_j^* and T_k are omitted for notational convenience.

Remark 3.3. The assumption C implies that the length of each segment tends to infinity at the same rate as n. We will introduce a size $u_n<<n$ which represents the lower bound on the accuracy of the approximation of the lengths of the segments. This minimum size is needed for the numerical computation of the criteria. For the ARMA and GARCH model, $u_n=\mathcal{O}((\log n)^\delta)$ can be chosen for $1\le\delta\le 2$.

We are now ready to prove the consistency of the penalized QLIK contrast:

Theorem 3.1. *Assume that $D(\Theta)$, $Id(\Theta)$, B, C and H_0 are satisfied with $r \geq 2$. If $K_{\max} \geq K^*$ then:*

$$(\widehat{K}_n, \underline{\widehat{\tau}}_n, \underline{\widehat{\theta}}_n) \xrightarrow[n\to\infty]{\mathcal{P}} (K^*, \underline{\tau}^*, \underline{\theta}^*).$$

Note that if K^* is known, we can relax the assumptions for the consistency by taking $\kappa_n = 1$ for all n as the penalty term in (4) does not matter. If K^* is unknown and $r = 2$, then a robust choice to any geometric or Riemanian dependence is $\kappa_n \propto n/\log n$. However, such large regularization parameters always over-penalized in practice.

3.3. Rates of convergence of the estimators

To state the rates of convergence of the estimators $\underline{\widehat{\tau}}_n$ and $\underline{\widehat{\theta}}_n$, we need to work under stronger moment and regularity assumptions. By convention, if the vectors $\underline{\widehat{t}}_n$ and $\underline{t}^*$ do not have the same length, complete the shorter of the 2 vectors with n before computing the norm $\|\underline{\widehat{t}}_n - \underline{t}^*\|_m$.

Theorem 3.2. *Assume that $D(\Theta)$, $Id(\Theta)$, B, C and H_2 are satisfied with $r \geq 4$. If $K_{\max} \geq K^*$ then the sequence $(\|\underline{\widehat{t}}_n - \underline{t}^*\|_m)_{n>1}$ is uniformly tight in probability,* i.e.

$$\lim_{\delta\to\infty} \lim_{n\to\infty} \mathbb{P}(\|\underline{\widehat{t}}_n - \underline{t}^*\|_m > \delta) = 0. \tag{8}$$

This theorem induces that $w_n^{-1} \|\underline{\widehat{t}}_n - \underline{t}^*\|_m \xrightarrow{P} 0$ for any sequence $(w_n)_n$ such that $w_n \to \infty$ and therefore $\|\underline{\widehat{t}}_n - \underline{t}^*\|_m = o_P(w_n)$: the convergence rate is arbitrary close to $O_P(1)$. This is the same convergence rate as in the case where $(X_t)_t$ is a sequence of independent r.v. (see for instance [1]). Such convergence rate was already reached in the frame of piecewise linear regression with innovations satisfying a mixing property in [18].

Let us turn now to the convergence rate of the estimator of parameters θ_j^*. By convention if $\widehat{K}_n < K^*$, set $\widehat{T}_j = \widehat{T}_{\widehat{K}_n}$ for $j \in \{\widehat{K}_n, \dots, K^*\}$. Then,

Theorem 3.3. *Assume that $D(\Theta)$, $Id(\Theta)$, B, C and H_2 are satisfied with $r \geq 4$ and $\kappa_n = O(\sqrt{n})$. Then if $\theta_j^* \in \overset{\circ}{\Theta}$ for all $j = 1, \dots, K^*$, we have*

$$\sqrt{n_j^*}\,\big(\widehat{\theta}_n(\widehat{T}_j) - \theta_j^*\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d\big(0, F(\theta_j^*)^{-1} G(\theta_j^*) F(\theta_j^*)^{-1}\big), \tag{9}$$

where, using $q_{0,j}$ defined in (12), the matrix F and G are such that

$$(F(\theta_j^*))_{k,l} = \mathbb{E}\Big(\frac{\partial^2 q_{0,j}(\theta_j^*)}{\partial\theta_k \partial\theta_l}\Big) \;\; \textit{and} \;\; (G(\theta_j^*))_{k,l} = \mathbb{E}\Big(\frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta_k} \frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta_l}\Big). \tag{10}$$

In Theorem 3.3, a condition on the rate of convergence of κ_n is added. The most robust choice for the regularization parameter corresponds to $\kappa_n \propto \sqrt{n}$ as it corresponds to the most general problem (3) (see above). However, by assumption H_2 it excludes models with finite moments $r \geq 4$ satisfying: $\ell^{-\gamma} =$

$O(\alpha_\ell^{(i)}(f_\theta,\Theta)+\alpha_\ell^{(i)}(M_\theta,\Theta))$ or $\ell^{-\gamma}=\alpha_\ell^{(i)}(h_\theta,\Theta))$ with $1<\gamma\le 3/2$ for some $i=0,1,2$. For these models the consistency for $\widehat{\underline{\tau}}_n$ holds but we do not get any rate of convergence for $\widehat{\underline{\theta}}_n$.

4. Some examples

4.1. AR(∞) models

Consider $AR(\infty)$ with K^*-1 breaks defined by the equation:

$$X_t=\sum_{k\ge 1}\phi_k(\theta_j^*)X_{t-k}+\xi_t,\quad t_{j-1}^*<t\le t_j^*,\ j=1,\dots,K^*.$$

This is the problem (3) with models $\mathcal{M}_{T_i^*}(f_\theta,M_\theta)$ where $f_\theta(x_1,\dots)=\sum_{k\ge 1}\times\phi_k(\theta)x_k$ and $M_\theta\equiv 1$. Assume that Θ is a compact set such that $\sum_{k\ge 1}\|\phi_k(\theta)\|_\Theta<1$. Thus $\Theta=\Theta$ for any $r\ge 1$ satisfying $\mathbb{E}|\xi_0|^r<\infty$. Then Assumptions $\mathrm{D}(\Theta)$ and $A_0(f_\theta,\Theta)$ hold automatically with $\underline{h}=1$ and $\alpha_k^{(0)}(f_\theta,\Theta)=\|\phi_k(\theta)\|_\Theta$. Then,

- Assume that $\mathrm{Id}(\Theta)$ holds and that there exists $r\ge 2$ such that $\mathbb{E}|\xi_0|^r<\infty$. If there exists $\gamma>1\vee 4r^{-1}$ such that $\|\phi_k(\theta)\|_\Theta=O(k^{-\gamma})$ for all $k\ge 1$, the choice $\kappa_n=n/\log n$ ensures the strong consistency of $(\widehat{K}_n,\widehat{\underline{\tau}}_n,\widehat{\underline{\theta}}_n)$.
- Moreover, if $\mathbb{E}|\xi_0|^4<\infty$, $\gamma>3/2$ and ϕ_k twice differentiable satisfying $\|\phi_k'(\theta)\|_\Theta=O(k^{-\gamma})$ and $\|\phi_k''(\theta)\|_\Theta=O(k^{-\gamma})$, the choice $\kappa_n=\sqrt{n}$ ensures the convergence (8) of $\widehat{\underline{t}}_n$ and the CLT (9) satisfied by $\widehat{\underline{\theta}}_n(\widehat{T}_j)$ for all j.

Note that this problem of change detection was considered by Davis *et al.* in [6] but moments $r>4$ are required. In Davis *et al.* [7], the same problem for another break model for AR processes is studied. However, in both these papers, the process is supposed to be independent from one block to another and stationary on each block.

4.2. ARCH(∞) models

Consider an $ARCH(\infty)$ model with K^*-1 breaks defined by:

$$X_t=\Big(\psi_0(\theta_j^*)+\sum_{k=1}^{\infty}\psi_k(\theta_j^*)X_{t-k}^2\Big)^{1/2}\xi_t,\quad t_{j-1}^*<t\le t_j^*,\ j=1,\dots,K^*,$$

where $(\psi_k(\theta))_{k\ge 0}$ is a sequence of positive real numbers and $\mathbb{E}(\xi_0^2)=1$. Note that $h_\theta((x_k)_{k\in\mathbb{N}})=\psi_0(\theta)+\sum_{k=1}^{\infty}\psi_k(\theta)x_k^2$ and $f_\theta=0$. Assume that Θ is a compact set such that $\sum_{k\ge 1}\|\psi_k(\theta)\|_\Theta<1$, then $\Theta(2)=\Theta$. Assume that $\inf_{\theta\in\Theta}\psi_0(\theta)>0$ which ensures that $D(\Theta)$ and $Id(\Theta)$ hold.

- If there exists $\gamma>2$ such that $\|\psi_k(\theta)\|_\Theta=O(k^{-\gamma})$ for all $k\ge 1$, then the choice of $\kappa_n=n/\log n$ leads to the consistency of $(\widehat{K}_n,\widehat{\underline{\tau}}_n,\widehat{\underline{\theta}}_n)$ when $\theta_j^*\in\Theta$ for all j.

- Moreover, if $\mathbb{E}|\xi_0|^4 < \infty$, $\Theta(4)$ is a compact set such s $\theta_j^* \in \overset{\circ}{\Theta}(4)$ for all j, and if ψ_k is a twice differentiable function satisfying $\|\psi_k'(\theta)\|_\Theta = O(k^{-\gamma})$ and $\|\psi_k''(\theta)\|_\Theta = O(k^{-\gamma})$ with $\gamma > 3/2$, then the choice of $\kappa_n = \sqrt{n}$ ensures the convergence (8) and the CLT (9) satisfied by $\underline{\widehat{\theta}}_n(\widehat{T}_j)$ for all j.

This problem of break detection was already studied by Kokoszka and Leipus in [14] but they obtained the consistency of their procedure under stronger assumptions.

Example 1. Let us detail the GARCH(p,q) model with K^*-1 breaks defined by:

$$X_t = \sigma_t\,\xi_t,\ \sigma_t^2 = a_{0,j}^* + \sum_{k=1}^q a_{k,j}^* X_{t-k}^2 + \sum_{k=1}^p b_{k,j}^* \sigma_{t-k}^2 \quad t_{j-1}^* < t \le t_j^*,\ j = 1, \dots, K^*$$

with $\mathbb{E}(\xi_0^2) = 1$. Assume that for any $\theta = (a_0, \dots, a_q, b_1, \dots, b_p) \in \Theta$ then $a_k \ge 0$, $b_k \ge 0$ and $\sum_{k=1}^p b_k < 1$. Then, there exists (see Nelson and Cao [19]) a nonnegative sequence $(\psi_k(\theta))_k$ such that $\sigma_t^2 = \psi_0(\theta) + \sum_{k\ge 1} \psi_k(\theta) X_{t-k}^2$. Remark that this sequence is twice differentiable with respect to θ and that its derivatives are exponentially decreasing. Moreover for any $\theta \in \Theta$, $\sum_{k\ge 1}\psi_k(\theta) \le \big(\sum_{k=1}^q a_k\big)/\big(1 - \sum_{k=1}^p b_k\big)$ and

$$\Theta = \left\{\theta \in \Theta\ ,\ (\mathbb{E}|\xi_0|^r)^{2/r} \sum_{k=1}^q a_k + \sum_{k=1}^p b_k < 1\right\}.$$

Then if $\sum_{k=1}^q a_{k,j}^* + \sum_{k=1}^p b_{k,j}^* < 1$ for all j (case $r \ge 2$), our estimation procedure associated with a regularization parameter $\kappa_n K$ for any $1 << \kappa_n << n$ is consistent. Moreover, if $(\mathbb{E}|\xi_0|^4)^{1/2} \sum_{k=1}^q a_{k,j}^* + \sum_{k=1}^p b_{k,j}^* < 1$ for all j, then our procedure with a regularization parameter $1 << \kappa_n = O(\sqrt{n})$ allows the same rates of convergence than in the case where (X_t) are independent random variables. For example, a BIC-type regularization parameter $\kappa_n \propto \log n$ as in [7] can be chosen in this case.

4.3. Estimates breaks in TARCH(∞) model

Consider a TARCH(∞) model with breaks defined by:

$$X_t = \sigma_t\,\xi_t,\ \sigma_t = b_0(\theta_j^*) + \sum_{k\ge 1}\Big(b_k^+(\theta_j^*)\max(X_{t-k}, 0) - b_k^-(\theta_j^*)\min(X_{t-k}, 0)\Big),$$

with $t_{j-1}^* < t \le t_j^*$, $j = 1, \dots, K^*$, where $(b_k^+)_{k\ge 0}$, $(b_k^-)_{k\ge 0}$ are sequences of positive real numbers satisfying $\|b_0(\theta)\|_\Theta > 0$ and $\sum_{k\ge 1}\max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < \infty$. Then $f_\theta = 0$ and $(A_0(M_\theta, \Theta))$ holds with $\alpha_k^{(0)}(M_\theta, \Theta) = \max(\|b_k^+(\theta)\|_\Theta,$ $\|b_k^-(\theta)\|_\Theta)$

- Assume $\|\xi_0\|_r^{1/r}\sum_{k\ge 1}\max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) < 1$ for $r \ge 2$. If there exists $\gamma > 1 \vee 4r^{-1}$ such that $\max(\|b_k^+(\theta)\|_\Theta, \|b_k^-(\theta)\|_\Theta) = O(k^{-\gamma})$ for all

$k \geq 1$, then $\kappa_n = n/\log n$ leads to the consistency of $(\widehat{K}_n, \underline{\widehat{\tau}}_n, \underline{\widehat{\theta}}_n)$ when $\theta_j^* \in \Theta(2)$ for all j.
- Moreover, if $\mathbb{E}|\xi_0|^4 < \infty$ and b_k^+, b_k^- are twice differentiable satisfying $\|\partial b_k^+(\theta)/\partial\theta\|_\Theta = O(k^{-\gamma})$ and $\|\partial^2 b_k^+(\theta)/\partial\theta^2\|_\Theta = O(k^{-\gamma})$ with $\gamma > 3/2$ (the same for b_k^-), then $\kappa_n = \sqrt{n}$ ensures the convergence (8) and the CLT (9) satisfied by $\underline{\widehat{\theta}}_n(\widehat{T}_j)$ for all j (with $\theta_j^* \in \overset{\circ}{\Theta}(4)$).

To our knowledge, these results are the first one concerning the change detection for TARCH(∞).

5. Some simulations results

The procedure is implemented on the R software (developed by the CRAN project). Since we proceed with not so large samples ($n \leq 2000$), the consistency of $\widehat{K}_n$ is often not obtained for the most robust theoretical choice of $\kappa_n = \sqrt{n}$. As a consequence, for numerical applications, we chose a data-driven procedure for computing the regularization parameter κ_n. Thus, κ_n is calibrated using the slope estimation procedure of Baudry et al. [4]. Once obtained the regularization parameter κ_n, the dynamic programming algorithm (see [15]) is used to minimize the criteria. Remark that we could also use the genetic algorithm and the approximated likelihood of [7] to speed up the procedure.

5.1. The slope estimation procedure

The heuristic of the procedure is that the criteria (here QLIK) is a linear transformation of the penalties (here the number of periods K) for the most complex models (with K close to $K_{\max}$). This slope should be close to $-\kappa_n/2$. This procedure has already been used in [4] for breaks detection in an i.i.d. context. We adapt it to the case of dependence (details are omitted for the common part with the iid case and we refer the interested reader to [4]).

By construction, the procedure is very sensitive to the choice of $K_{\max}$ as only complex models are used to estimate the slope. As discussed in the Remark 3.3, we only consider periods of length larger than u_n and we can a priori fix K_{max} smaller than $[n/u_n]$. Therefore, the slope estimation procedure consider only the linear part of $-QLIK$ with $K \leq K_{max}$. The concrete procedure (see examples below) is:

1. For each $1 \leq K \leq K_{max}$, draw $(K, -\min_{\underline{t},\underline{\theta}} QLIK(K))_{1\leq K\leq K_{max}}$. Then compute the slope of the linear part: this slope is $\hat{\kappa}_n/2$.
2. Using $\kappa_n = \hat{\kappa}_n$, draw $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))_{1\leq K\leq K_{max}}$. This curve has a global minimum at $\widehat{K}_n$.

5.2. Implementation details

We assume that the regularization parameter is known (for instance $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$ or $\kappa_n = \sqrt{n}$). In this section, we give more details on how to compute $\widehat{K}_n$ and the optimal configuration of the breaks by using the dynamic programming algorithm. The basic idea of this algorithm is that: for a given $1 \leq K \leq K_{max}$, if $(t_1, \ldots, t_{K-1}, t)$ is an optimal configuration of $X_1, \ldots, X_t$ into K segments, then $(t_1, \ldots, t_{K-1})$ is an optimal configuration of $X_1, \ldots, X_{t_{K-1}}$ into $K-1$ segments.

For $1 \leq i \leq l \leq n$, denote $T_{i,l} = \{i, i+1, \ldots, l\}$ and let ML be the upper triangular matrix of dimension $n \times n$ with $ML_{i,l} = \widehat{L}(T_{i,l}, \widehat{\theta}_n(T_{i,l}))$ for $i \leq l$. The estimated number of segment $\widehat{K}_n$ and the corresponding optimal configuration can be obtained as follow:

1. Let C be an upper triangular matrix of dimension $K_{max} \times n$. For $1 \leq K \leq K_{max}$ and $K \leq t \leq n$, $C_{K,t}$ will be the minimum penalized criteria of $X_1, \ldots, X_t$ into K segments. Therefore, for $t = 1, \ldots, n$ $C_{1,t} = -2ML_{1,t} + \kappa_n$ and the relation $C_{K+1,t} = \min_{K \leq l \leq t-1}(C_{K,l} - 2ML_{l+1,t} + \kappa_n)$ is satisfied. Hence, $\widehat{K}_n = \operatorname{Argmin}_{1 \leq K \leq K_{max}}(C_{K,n})$.
2. Let Z be an upper triangular matrix of dimension $(K_{max}-1) \times n$. For $1 \leq K \leq (K_{max}-1)$ and $K+1 \leq t \leq n$, $Z_{K,t}$ will be the Kth potential break-point of $X_1, \ldots, X_t$. Therefore, the relation $Z_{K,t} = \operatorname{Argmin}_{K \leq l \leq t-1}(C_{K,l} - 2ML_{l+1,t} + \kappa_n)$ is satisfied for $K = 1, \ldots, K_{max} - 1$ and the break-point are obtained as follow: set $\widehat{t}_{\widehat{K}_n} = n$, $\widehat{t}_1 = 1$ and for $K = \widehat{K}_n - 1, \ldots, 2$, $\widehat{t}_K = Z_{K, \widehat{t}_{K+1}}$.

Note that the above procedure requires $\mathcal{O}(n^2)$ operations, instead of $\mathcal{O}(n^{K_{max}})$ if the standard procedure is used.

Remark 5.1. The minimum description length (MDL in the sequel) criterion (see [7]) is defined in our setting by:

$$MDL(K, \underline{t}, \underline{\theta}) := \log_+(K-1) + K \log n + \frac{d}{2} \sum_{k=1}^{K} \log n_k - \sum_{k=1}^{K} \widehat{L}_n(T_k, \theta_k)$$

where $\log_+(x) = 0$ if $x \leq 1$ and $\log_+(x) = \log x$ if $x > 1$. We can also write:

$$MDL(K, \underline{t}, \underline{\theta}) = \widetilde{MDL}(K, \underline{t}, \underline{\theta}) + K \log n$$

where

$$\widetilde{MDL}(K, \underline{t}, \underline{\theta}) = -\sum_{k=1}^{K} \Big(\widehat{L}_n(T_k, \theta_k) - \frac{d}{2} \log n_k - \log_+(k-1) + \log_+(k-2) \Big).$$

Hence, the MDL criterion can be seen as a penalized criterion and the dynamic programming algorithm described above can be used to find the optimal configuration.

5.3. Results of simulations

AR(1) models: we consider the problem (3) for a AR(1):

$$X_t = \theta_j^* X_{t-1} + \xi_t \qquad \forall t \in T_j, \quad \forall j \in \{1, \ldots, K^*\}.$$

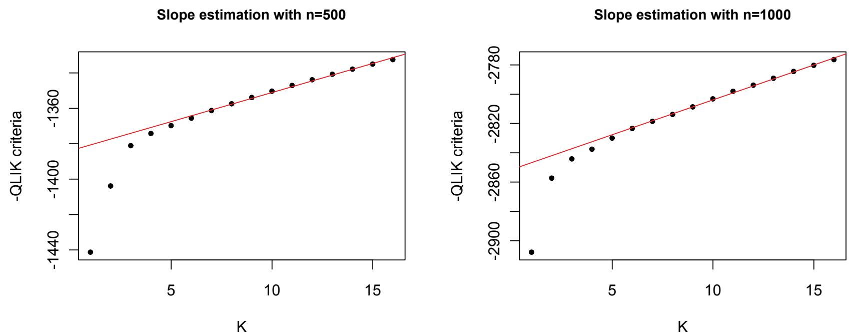

Fig 1. *The curve of* $-\min_{\underline{t},\underline{\theta}} QLIK$ *for* $1 \leq K \leq K_{max}$ *for AR(1) process in scenario* A_4*. The solid line represents the linear part of this curve with slope* $\hat{\kappa}_n/2 = 3.47$ *when* $n = 500$ *and* $\hat{\kappa}_n/2 = 4.90$ *when* $n = 1000$

For $n = 500$ and $n = 1000$, we generate a sample $(X_1, \ldots, X_n)$ in the following situations:

- **scenario** $\mathbf{A}_0$: $\theta^*(1) = 0.5$ is constant $(K^* = 1)$;
- **scenario** $\mathbf{A}_1$: $\theta^*(1) = 0.5$ changes to $\theta^*(2) = 0.2$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario** $\mathbf{A}_2$: $\theta^*(1) = 0.7$ changes to $\theta^*(2) = 0.9$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario** $\mathbf{A}_3$: $\theta^*(1) = 0.5$ changes to $\theta^*(2) = 0.3$ at $t_1^* = 0.3n$ which changes to $\theta^*(3) = 0.7$ at $t_2^* = 0.7n$ $(K^* = 3)$;
- **scenario** $\mathbf{A}_4$: $\theta^*(1) = 0.7$ changes to $\theta^*(2) = 0.9$ at $t_1^* = 0.3n$ which changes to $\theta^*(3) = 0.6$ at $t_2^* = 0.7n$ $(K^* = 3)$.

The regularization parameter is chosen by using the slope estimation presented above (Subsection 5.1). Figure 1 represents the slope of the linear part of the $-QLIK$ criteria (minimized in $(\underline{t}, \underline{\theta})$) in scenario A_4 for $n = 500$ and $n = 1000$. Thus, by referring to the Figure 1 we obtain $\hat{\kappa}_n \approx 7.0$ for $n = 500$ and $\hat{\kappa}_n \approx 9.8$ for $n = 1000$.

We are going to minimize the $penQLIK$ in $(K, \underline{t}, \underline{\theta})$, with $1 \leq K \leq K_{max}$ and $\kappa_n = \hat{\kappa}_n$. Figure 2 represents the points $(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$ for $1 \leq K \leq K_{max} = 10$.

One can easily read on the Figure 2, the estimated values $\widehat{K}_n = 4$ for $n = 500$ and $\widehat{K}_n = 3$ for $n = 1000$ (the estimated number of break is $\widehat{K}_n - 1$). Moreover, the estimated instants of break are $\widehat{\underline{t}}_n = (146, 228, 357)$ $(\underline{t}^* = (150, 350))$ for $n = 500$ and $\widehat{\underline{t}}_n = (282, 687)$ $(\underline{t}^* = (300, 700))$ for $n = 1000$. Figure 3 shows the estimated break points for the trajectories ($n = 500$ and $n = 1000$) for AR(1) processes following scenario $\mathbf{A}_4$ with two changes.

Now, 100 independent replications of a AR(1) process are generated following the scenarios $\mathbf{A}_0$-$\mathbf{A}_4$. For each replication, the estimated number of segments is computed using QLIK criteria with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and using MDL procedure and Table 1 indicates the proportions of number of replications (frequencies) where the true number of breaks is achieved following the scenarios $\mathbf{A}_0$-$\mathbf{A}_4$.

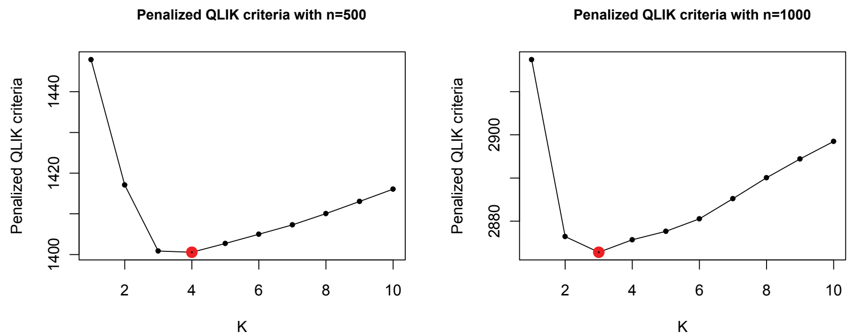

FIG 2. *The graph*$(K, \min_{\underline{t},\underline{\theta}} penQLIK(K))$ *for* $1 \leq K \leq K_{max} = 10$ *for AR(1) in scenario* A_4.

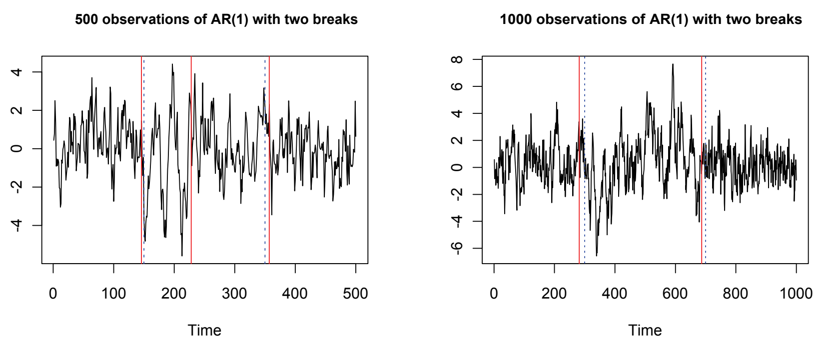

FIG 3. *The estimated of breakpoints for a trajectory of* $AR(1)$ *process in scenario* A_4. *The solid lines represent the estimated break instants and the dotted lines represent the true ones.*

For the replications of scenario $\mathbf{A}_4$, where the true number of break is fitted ($\widehat{K}_n = 3$), the average of the estimated parameters are computed and shown in Table 2.

AR(2) models: we consider the problem (3) for a AR(2):

$$X_t = \phi_1^*(j)\, X_{t-1} + \phi_2^*(j)\, X_{t-2} + \xi_t \qquad \forall t \in T_j, \quad \forall j \in \{1, \ldots, K^*\}.$$

Denote $\theta^*(j) = (\phi_1^*(j), \phi_2^*(j))$. For $n = 500$ and $n = 1000$, we generate a sample $(X_1, \ldots, X_n)$ in the following situations:

- **scenario** $\mathbf{B}_0$: $\theta^*(1) = (0.4, 0.3)$ is constant ($K^* = 1$);
- **scenario** $\mathbf{B}_1$: $\theta^*(1) = (0.4, 0.3)$ changes to $\theta^*(2) = (0.1, 0.3)$ at $t^* = 0.5n$ ($K^* = 2$);
- **scenario** $\mathbf{B}_2$: $\theta^*(1) = (0.4, 0.3)$ changes to $\theta^*(2) = (0.2, 0.5)$ at $t^* = 0.5n$ ($K^* = 2$);
- **scenario** $\mathbf{B}_3$: $\theta^*(1) = (0.4, 0.3)$ changes to $\theta^*(2) = (0.6, 0.1)$ at $t^* = 0.5n$ ($K^* = 2$).

TABLE 1
Frequencies of the number of breaks estimated after 100 replications for AR(1) process following scenarios $\boldsymbol{A}_0$-$\boldsymbol{A}_4$

Model			$\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$
scenario A_0 ($K^* = 1$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.74	0.00	0.26
		$\kappa_n = \log n$	0.50	0.00	0.50
		$\kappa_n = \sqrt{n}$	0.94	0.00	0.06
		MDL	0.95	0.00	0.05
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.81	0.00	0.20
		$\kappa_n = \log n$	0.43	0.00	0.57
		$\kappa_n = \sqrt{n}$	1.00	0.00	0.00
		MDL	0.97	0.00	0.03
scenario A_1 ($K^* = 2$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.52	0.06	0.42
		$\kappa_n = \log n$	0.40	0.04	0.56
		$\kappa_n = \sqrt{n}$	0.23	0.77	0.00
		MDL	0.44	0.56	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.78	0.00	0.22
		$\kappa_n = \log n$	0.40	0.00	0.60
		$\kappa_n = \sqrt{n}$	0.38	0.62	0.00
		MDL	0.87	0.13	0.00
scenario A_2 ($K^* = 2$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.48	0.00	0.52
		$\kappa_n = \log n$	0.17	0.00	0.83
		$\kappa_n = \sqrt{n}$	0.29	0.71	0.00
		MDL	0.56	0.44	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.76	0.00	0.24
		$\kappa_n = \log n$	0.06	0.00	0.94
		$\kappa_n = \sqrt{n}$	0.57	0.43	0.00
		MDL	0.89	0.07	0.04
scenario A_3 ($K^* = 3$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.45	0.32	0.23
		$\kappa_n = \log n$	0.37	0.26	0.37
		$\kappa_n = \sqrt{n}$	0.00	1.00	0.00
		MDL	0.01	0.99	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.61	0.13	0.26
		$\kappa_n = \log n$	0.39	0.00	0.61
		$\kappa_n = \sqrt{n}$	0.00	1.00	0.00
		MDL	0.20	0.80	0.00
scenario A_4 ($K^* = 3$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.53	0.12	0.35
		$\kappa_n = \log n$	0.28	0.06	0.66
		$\kappa_n = \sqrt{n}$	0.04	0.96	0.00
		MDL	0.09	0.91	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.75	0.00	0.25
		$\kappa_n = \log n$	0.12	0.00	0.88
		$\kappa_n = \sqrt{n}$	0.06	0.94	0.00
		MDL	0.54	0.46	0.00

100 independent replications of a AR(2) process are generated following the scenarios $\mathbf{B}_0$-$\mathbf{B}_3$. It is evaluated the performance of the procedure using QLIK criteria with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and the one of MDL procedure. Table 3 indicates the proportions of number of replications (frequencies) where the true number of breaks is achieved following the scenarios $\mathbf{B}_0$-$\mathbf{B}_3$.

GARCH(1,1) models: we consider examples of problem (3) when X is a $GARCH(1,1)$ process on each period:

$$X_t = \sigma_t Z_t, \qquad \sigma_t^2 = a_0^*(j) + a_1^*(j) X_t^2 + b_1^*(j) \sigma_t^2 \qquad \forall t \in T_j^*, \;\; \forall j \in \{1, \ldots, K^*\}. \tag{11}$$

TABLE 2
The estimated parameters for the replications of $AR(1)$ processes following $\mathbf{A}_4$ satisfying $\widehat{K}_n = 3 = K^*$ (two changes estimated)

scenario A_4		Mean ± s.d. $\widehat{\tau}_1$	Mean ± s.d. $\widehat{\tau}_2$	Mean $\lVert\widehat{\underline{\tau}}_n - \underline{\tau}^*\rVert$	Mean $\widehat{\theta}(j)$, $j=1,2,3$
$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.310 ± 0.078	0.719 ± 0.049	0.070	0.668 ; 0.894 ; 0.571
	$\kappa_n = \log n$	0.308 ± 0.064	0.718 ± 0.049	0.065	0.666 ; 0.898 ; 0.567
	$\kappa_n = \sqrt{n}$	0.323 ± 0.037	0.678 ± 0.016	0.045	0.600 ; 0.935 ; 0.561
	MDL	0.316 ± 0.044	0.680 ± 0.013	0.042	0.637 ; 0.926 ; 0.577
$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.297 ± 0.078	0.691 ± 0.025	0.063	0.694 ; 0.894 ; 0.613
	$\kappa_n = \log n$	0.317 ± 0.038	0.710 ± 0.033	0.045	0.708 ; 0.874 ; 0.598
	$\kappa_n = \sqrt{n}$	0.341 ± 0.078	0.714 ± 0.023	0.046	0.640 ; 0.905 ; 0.528
	MDL	0.340 ± 0.085	0.702 ± 0.029	0.062	0.676 ; 0.911 ; 0.586

TABLE 3
Frequencies of the number of breaks estimated after 100 replications for $AR(2)$ process following scenarios $\boldsymbol{B}_0$-$\boldsymbol{B}_3$

Model			$\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$
scenario B_0 ($K^* = 1$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.61	0.00	0.39
		$\kappa_n = \log n$	0.08	0.00	0.92
		$\kappa_n = \sqrt{n}$	0.94	0.00	0.06
		MDL	0.92	0.00	0.08
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.79	0.00	0.21
		$\kappa_n = \log n$	0.06	0.00	0.94
		$\kappa_n = \sqrt{n}$	0.98	0.00	0.02
		MDL	0.97	0.00	0.03
scenario B_1 ($K^* = 2$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.63	0.04	0.33
		$\kappa_n = \log n$	0.15	0.00	0.85
		$\kappa_n = \sqrt{n}$	0.37	0.63	0.00
		MDL	0.38	0.62	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.83	0.00	0.17
		$\kappa_n = \log n$	0.03	0.00	0.97
		$\kappa_n = \sqrt{n}$	0.60	0.40	0.00
		MDL	0.85	0.15	0.00
scenario B_2 ($K^* = 2$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.57	0.15	0.28
		$\kappa_n = \log n$	0.09	0.01	0.90
		$\kappa_n = \sqrt{n}$	0.10	0.90	0.00
		MDL	0.11	0.89	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.78	0.08	0.14
		$\kappa_n = \log n$	0.05	0.00	0.95
		$\kappa_n = \sqrt{n}$	0.17	0.83	0.00
		MDL	0.24	0.76	0.00
scenario B_3 ($K^* = 3$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.41	0.25	0.34
		$\kappa_n = \log n$	0.08	0.03	0.89
		$\kappa_n = \sqrt{n}$	0.07	0.93	0.00
		MDL	0.08	0.92	0.00
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.75	0.08	0.17
		$\kappa_n = \log n$	0.03	0.00	0.97
		$\kappa_n = \sqrt{n}$	0.19	0.81	0.00
		MDL	0.22	0.78	0.00

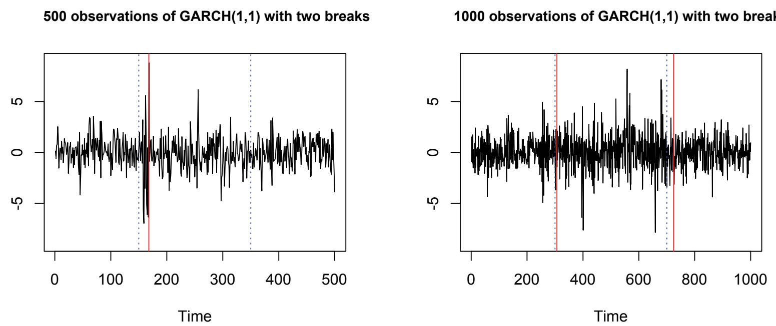

Fig 4. *A GARCH$(1,1)$ process with 2 breaks ($K^*=3$) following the scenario* $\mathbf{G}_4$. *The solid lines represent the estimated break instants and the dotted lines represent the true ones.*

Thus $\theta^* = (a_0^*, a_1^*, b_1^*)$. For $n = 500$ and $n = 1000$, we generate $(X_1, \ldots, X_n)$ in the following situation:

- **scenario** $\mathbf{G}_0$: $\theta^*(1) = (0.5, 0.2, 0.2)$ is constant $(K^* = 1)$;
- **scenario** $\mathbf{G}_1$: $\theta^*(1) = (0.5, 0.2, 0.2)$ changes to $\theta^*(2) = (0.5, 0.2, 0.6)$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario** $\mathbf{G}_2$: $\theta^*(1) = (0.5, 0.6, 0.2)$ changes to $\theta^*(2) = (1, 0.6, 0.2)$ at $t^* = 0.5n$ $(K^* = 2)$;
- **scenario** $\mathbf{G}_3$: $\theta^*(1) = (0.5, 0.2, 0.2)$ changes to $\theta^*(2) = (0.5, 0.2, 0.0)$ at $t_1^* = 0.3n$ which changes to $\theta^*(3) = (0.1, 0.2, 0.0)$ at $t_2^* = 0.7n$ $(K^* = 3)$;
- **scenario** $\mathbf{G}_4$: $\theta^*(1) = (0.5, 0.6, 0.2)$ changes to $\theta^*(2) = (1, 0.6, 0.2)$ (at $t_1^* = 0.3n$) which changes to $\theta^*(3) = (1, 0.2, 0.2)$ at $t_2^* = 0.7n$ $(K^* = 3)$.

Figure 4 shows an example of scenario $\mathbf{G}_4$ where one break is fitted with $\hat{\kappa}_n \approx 12.7$ for $n = 500$ and two breaks with $\hat{\kappa}_n \approx 18.3$ for $n = 1000$; we obtain, $\widehat{\underline{t}}_n = 168$ (while $\underline{t}^* = (150, 350)$) for $n = 500$ and $\widehat{\underline{t}}_n = (307, 725)$ (while $\underline{t}^* = (300, 700)$) for $n = 1000$.

Now, 100 independent replications of GARCH$(1,1)$ processes are generated following the scenarios $\mathbf{G}_0$-$\mathbf{G}_4$. For each replication, the estimated number of segment is computed using QLIK criteria with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and using MDL procedure and Table 4 indicates the proportions of replications (frequencies) when the true number of breaks is achieved following the scenarios $\mathbf{G}_0$-$\mathbf{G}_4$.

For the replications of the scenario $\mathbf{G}_2$, when the true number of break is fitted ($\widehat{K}_n = 2 = K^*$), the average of the estimated parameters are computed and shown in Table 5.

Finally, recall that in [7], the process is stationary on each segment and assumed to be independent from a segment to another. Davis *et al.* (2008) used the genetic algorithm to approximate the optimal values of the MDL criteria. We consider three of their scenarios with $n = 1000$ for GARCH$(1,1)$ processes:

- **scenario** $\mathbf{A}$: $\theta^*(1) = (0.4, 0.1, 0.5)$ is constant $(K^* = 1)$;

TABLE 4
Frequencies of the number of breaks estimated after 100 replications for GARCH(1,1) processes following the scenarios $\boldsymbol{G}_0$-$\boldsymbol{G}_4$

Model			$\widehat{K}_n = K^*$	$\widehat{K}_n < K^*$	$\widehat{K}_n > K^*$
scenario G_0 ($K^* = 1$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.44	0.00	0.56
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.58	0.00	0.42
		MDL	0.51	0.00	0.49
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.60	0.00	0.40
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.75	0.00	0.25
		MDL	0.63	0.00	0.37
scenario G_1 ($K^* = 2$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.42	0.12	0.46
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.52	0.05	0.00
		MDL	0.55	0.35	0.10
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.65	0.00	0.35
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.74	0.10	0.00
		MDL	0.67	0.09	0.24
scenario G_2 ($K^* = 2$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.52	0.20	0.28
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.39	0.44	0.17
		MDL	0.44	0.40	0.16
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.56	0.10	0.34
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.42	0.48	0.10
		MDL	0.57	0.31	0.12
scenario G_3 ($K^* = 3$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.41	0.28	0.31
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.40	0.60	0.00
		MDL	0.53	0.39	0.08
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.70	0.26	0.04
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.43	0.57	0.00
		MDL	0.59	0.37	0.04
scenario G_4 ($K^* = 3$)	$n = 500$	$\kappa_n = \hat{\kappa}_n$	0.30	0.55	0.15
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.08	0.90	0.02
		MDL	0.16	0.77	0.07
	$n = 1000$	$\kappa_n = \hat{\kappa}_n$	0.53	0.29	0.18
		$\kappa_n = \log n$	0.00	0.00	1.00
		$\kappa_n = \sqrt{n}$	0.10	0.90	0.00
		MDL	0.27	0.66	0.07

- **scenario C**: $\theta^*(1) = (0.4, 0.1, 0.5)$ changes to $\theta^*(2) = (0.4, 0.1, 0.6)$ at $t^* = 0.5n$ ($K^* = 2$);
- **scenario J**: $\theta^*(1) = (0.1, 0.1, 0.8)$ changes to $\theta^*(2) = (0.5, 0.1, 0.8)$ at $t^* = 0.5n$ ($K^* = 2$).

Table 6 shows the results obtained with our penQLIK method with $\kappa_n = \hat{\kappa}_n$, $\kappa_n = \log n$, $\kappa_n = \sqrt{n}$ and the results of the MDL procedure (obtained after 500 replications) taken in Table I of [7].

TABLE 5
The estimated parameters for the replications of $GARCH(1,1)$ processes following the scenario G_2 and satisfying $\widehat{K}_n = 2 = K^*$ (one break fitted)

Model	n		Mean ± s.d. $\widehat{\tau}$	Mean $\lvert\widehat{\tau}-\tau^*\rvert$	Mean $\widehat{\theta}(j)$, $j=1,2$
scenario G_2	$n=500$	$\kappa_n=\hat{\kappa}_n$	0.428 ± 0.245	0.176	(0.489, 0.513, 0.202)
					(1.027, 0.560, 0.220)
		$\kappa_n=\log n$	NA	NA	NA
		$\kappa_n=\sqrt{n}$	0.297 ± 0.260	0.253	(0.374, 0.376, 0.234)
					(1.026, 0.606, 0.182)
		MDL	0.372 ± 0.275	0.222	(0.439, 0.437, 0.203)
					(0.980, 0.579, 0.224)
	$n=1000$	$\kappa_n=\hat{\kappa}_n$	0.466 ± 0.130	0.058	(0.455, 0.528, 0.201)
					(1.038, 0.595, 0.186)
		$\kappa_n=\log n$	NA	NA	NA
		$\kappa_n=\sqrt{n}$	0.302 ± 0.266	0.215	(0.276, 0.395, 0.354)
					(1.033, 0.582, 0.182)
		MDL	0.469 ± 0.133	0.059	(0.455, 0.528, 0.201)
					(1.052, 0.597, 0.181)

TABLE 6
*Frequencies of the number of breaks estimated after 100 replications for $GARCH(1,1)$ processes with $n = 1000$ following the scenarios **A**, **C** and **J** of Davis et al. (2008) [7]. The results of MDL procedure were taken in Table I of [7]*

Model		$\widehat{K}=1$ (no break)	$\widehat{K}=2$ (one break)	$\widehat{K}\geq 3$ (more than 2 breaks)
scenario A	$\kappa_n=\hat{\kappa}_n$	0.560	0.390	0.050
($K^*=1$)	$\kappa_n=\log n$	0.000	0.000	1.000
	$\kappa_n=\sqrt{n}$	0.600	0.390	0.010
	MDL	0.958	0.042	0.000
scenario C	$\kappa_n=\hat{\kappa}_n$	0.290	0.550	0.160
($K^*=2$)	$\kappa_n=\log n$	0.000	0.000	1.000
	$\kappa_n=\sqrt{n}$	0.770	0.330	0.000
	MDL	0.804	0.192	0.004
scenario J	$\kappa_n=\hat{\kappa}_n$	0.050	0.630	0.320
($K^*=2$)	$\kappa_n=\log n$	0.000	0.000	1.000
	$\kappa_n=\sqrt{n}$	0.280	0.620	0.100
	MDL	0.008	0.952	0.040

Conclusion of simulations for AR(1), AR(2) and GARCH(1,1) processes: The results of QLIK criteria with $\hat{\kappa}_n$ and $\sqrt{n}$ penalty show that the probability $\mathbb{P}(\widehat{K}_n = K^*)$ increases as n increases in all scenarios as it can be deduced from the theory. This is not the case for $\log n$ penalty (see for instance the scenario $\mathbf{A}_2$). Comparing the results of scenarios $\mathbf{A}_1$ and $\mathbf{A}_2$ (or scenarios $\mathbf{A}_3$ and $\mathbf{A}_4$), the BIC penalty ($\kappa_n = \log n$) under-penalizes the number of breaks when the process is sufficiently dependent on its own past. More dependent the process, larger the probability to fit the true number of breaks with $\sqrt{n}$ or $\hat{\kappa}_n$ penalty (except in the case $\mathbf{G}_2$ for $\sqrt{n}$ penalty). However in the case of two breaks, the $\sqrt{n}$ penalty over-penalizes the number of breaks contrarily with $\hat{\kappa}_n$ penalty which provides the best results as well for AR(1) as for GARCH(1,1) processes.

For the scenarios $\mathbf{A}_1$-$\mathbf{A}_4$, the change in the parameter induces a change in the variance of the stationary solution of the model. In these cases, the Table 1 shows that the MDL procedure provides satisfactory results when there is one break in the model. But this procedure is not really efficient in the case of two breaks (see scenarios $\mathbf{A}_3$ and $\mathbf{A}_4$). In Table 3, we also consider two scenarios ($\mathbf{B}_2$ and $\mathbf{B}_3$) of AR(2) process where there is a change in the parameters but the variances of the stationary solutions are very close. As can be seen, the penalty $\widehat{\kappa}_n$ still works well whereas the MDL procedure provides poor results. Moreover, the Table 6 shows that, the MDL procedure provides sometimes excellent results (scenarios **A** and **J**), but also very weak result (scenario **C**).

Finally, one can see that if the $\hat{\kappa}_n$ penalty does not always provide the best results, its results in all scenarios remain satisfactory, in the sense that in all considered scenarios, the estimated probability to fit the true number of break is greater than 0.50 for $n = 1000$. The use of our method with $\hat{\kappa}_n$ is clearly the best possible trade-off for one break models. In the case of two breaks, the $\hat{\kappa}_n$ penalty provides best results. Contrary to the MDL procedure, the QLIK criteria with $\widehat{\kappa}_n$ penalty works well in the AR models even when the changes in the parameters does not induce a change in the variance of the stationary solution. For all these reasons, we recommend to use our procedure with the penalty term $\kappa_n = \widehat{\kappa}_n$.

TARCH(1) models: we consider an example of problem (3) where X is a $TARCH(1)$ with one change: $\forall t \in T_j, \quad j = 1, 2 = K^*$,

$$X_t = \sigma_t \, \xi_t, \;\; \sigma_t = b_0^*(j) + b_1^{+*}(j) \max(X_{t-1}, 0) - b_1^{-*}(j) \min(X_{t-1}, 0).$$

The vector of parameter is $\theta^* = (b_0^*, b_1^{+*}, b_1^{-*})$. Here we assume that the number of breaks is known, *i.e.* $K = K^* = 2$ but the break instant t^* and parameters θ^* are unknown. For $n = 1000$ and $n = 2000$, we generate 100 independent replications of $(X_1, \ldots, X_n)$ with parameters $\theta^*(1) = (0.01, 0.05, 0.04)$ for $t \leq t^* = 0.4n$ and $\theta^*(2) = (0.01, 0.05, 0.1)$ for $t > t^*$. Table 5 provides the sample mean and the standard deviation of $\widehat{\tau}_n$, the sample mean of the error $|\widehat{\tau}_n - \tau^*|$ and the sample means of $\widehat{\theta}_n(1)$ and $\widehat{\theta}_n(2)$.

We can see that the results obtained for AR(1) and GARCH$(1,1)$ models are much better than those obtained for TARCH(1) process even when K^* is known and $K^* = 2$ instead of $K^* = 3$. This is explained by the fact that this model provides an asymmetric function of the past observations. Thus, some asymmetric effects can be confused with breaks.

TABLE 7
The estimated parameters for a TARCH(1) process with one break from 100 independent replications. The parameter $\theta^*(1) = (0.01, 0.05, 0.04)$ changes to $\theta^*(2) = (0.01, 0.05, 0.1)$ at $t^* = 0.4n$

	Mean $\widehat{\tau}_n \pm$ s.d.	Mean $\lvert\widehat{\tau}_n - \tau^*\rvert$	Mean $\widehat{\theta}(1)$	Mean $\widehat{\theta}(2)$
$n = 1000$	0.436 ± 0.126	0.093	(0.056, 0.071, 0.044)	(0.067, 0.057, 0.103)
$n = 2000$	0.419 ± 0.063	0.044	(0.059, 0.052, 0.051)	(0.066, 0.061, 0.098)

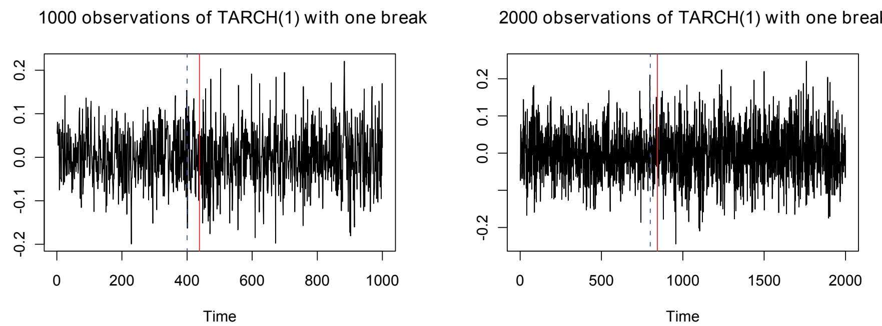

FIG 5. *Example of a the trajectory of TARCH(1) with one change (the red line represents the estimated break instant and the dotted line represents the true one).*

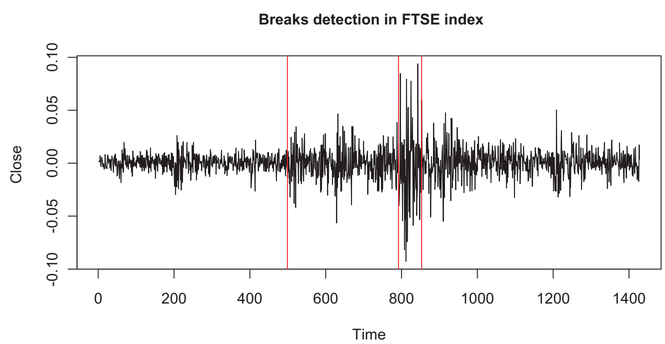

FIG 6. *The log-ratios of the closing daily price of the FTSE index. The vertical red lines represent the estimated instant of breaks.*

However, Table 5 shows that the change is correctly detected and the decay rate of the error $|\widehat{\tau}_n - \tau^*|$ is confirmed. Figure 5 presents an example of such TARCH(1) process with one break.

5.4. Application to financial data: FTSE index analysis

Now we apply our detection of changes methodology to the series of the log-returns of the closing values of the FTSE index: the share index of the 100 most highly capitalized UK companies listed on the London Stock Exchange, with the aim of investigating whether and how any detected breakpoints correspond to the milestones of the recent financial crisis. This is a trajectory composed with $n = 1428$ observations ranging from 27 July 2005 to 18 March 2011, *i.e.* roughly 6 trading years and uploaded from Yahoo finance (see Figure 6). We studied the log-ratio of the closing daily prices. Remark that we completely treat the period studied in [11].

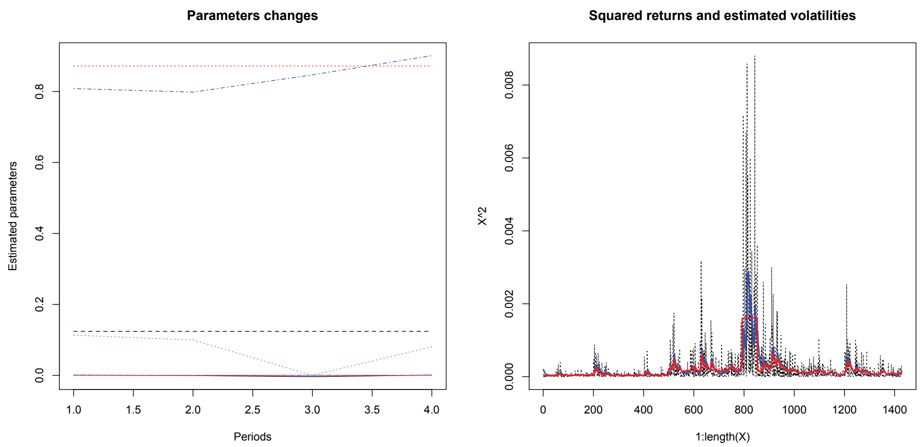

Fig 7. *The right plot represents the values of the parameters. The black line represents the values of μ, the red line of a_0, the green line of a_1 and the blue line of b_1. The left plot corresponds to the squared log-returns and the fitted volatilities, in blue with estimations over the whole sequence, in red with breaks.*

The *penQLIK* contrast is applied for a GARCH$(1,1)$ model (see (11) for a formal definition). The slope estimation procedure applied with $u_n = [n/(4 * \log(n))] = 49$ and $K_{max} = 25$ returns the values $\hat{\kappa} \approx 15$ and $\hat{K} = 4$, i.e. three breaks $\hat{t}_1 = 499$, $\hat{t}_2 = 792$ and $\hat{t}_3 = 853$. These values are close to the three breaks obtained in [11]:

$\hat{t}_1 = 499$, corresponding to 16 July 2007. From Wikipedia: "During the week of July 16, 2007, Bear Stearns disclosed that the two subprime hedge funds had lost nearly all of their value amid a rapid decline in the market for subprime mortgages."

$\hat{t}_2 = 792$, corresponding to 11 September 2008. From Wikipedia "On September 15, 2008, Lehman Brothers filed for Chapter 11 bankruptcy protection following the massive exodus of most of its clients, drastic losses in its stock, and devaluation of its assets by credit rating agencies".

$\hat{t}_3 = 853$, corresponding to the 5 December 2008. From Wikipedia "In the final quarter of 2008, the financial crisis saw the G-20 group of major economies assume a new significance as a focus of economic and financial crisis management."

Remark that our first two breaks are closer to the events identified in [11] than their own breaks. Analyzing the estimated values of coefficients (see Figure 7), breaks are due to changes of the coefficients a_1 and b_1 in the GARCH$(1,1)$ model (11). There is no break for the mean μ and the a_0 coefficients, valued close to 0. Next, we compare the fitted volatilities of the parameters estimations over the whole sequence and within the 4 periods. The third period, corresponding to a change of the value of the parameter a_1 ($a_1(3)$ is not significantly different from 0), leads to an estimated volatility satisfying the recurrence equation $\sigma_t \approx a_0(3)+b_1(3)\sigma_{t-1}$. In this period of high volatility, the estimated volatilities have different behaviors whether we take the break into account or not.

6. Proofs of the main results

In the sequel C denotes a positive constant whom value may differ from one inequality to another and (v_n) is a sequence such that $v_n = n/\kappa_n$ for all $n \geq 1$.

6.1. Proof of Proposition 2.2

(i) It is clear that $\{X_t,\, t \leq t_1^*\}$ exists and is causal, stationary with finite moments of order r (see [2]). Therefore, X is defined by induction as follows:

$$X_t := M_{\theta_j^*}(X_{t-1}, X_{t-2}, \ldots)\xi_t + f_{\theta_j^*}(X_{t-1}, X_{t-2}, \ldots),\ \forall t \in T_j^*;\ j = 2, \ldots K^*.$$

Thus, X_t is independent on $(\xi_j)_{j>t}$ and it suffices to prove (ii) which immediately leads existence of moments.

(ii) Let us first consider the general case when $A_0(f_\theta, \{\theta\})$ and $A_0(M_\theta, \{\theta\})$ hold with $\beta^{(0)}(\theta) < 1$. As in [8] we remark that

$$\|X_t\|_r \leq \frac{\|Z_{t_j^*,1}\|_r}{1-\beta^{(0)}(\theta_1^*)}$$

for $t \leq t_1^*$, with $Z_{t,j} := M_{\theta_j^*}(0,0,\ldots)\xi_t + f_{\theta_j^*}(0,0,\ldots)$ for all $j = 1, \ldots, K^*$. Assume that there exists $C_{r,t} < \infty$ such that $C_{r,t} = \sup_{i<t} \|X_i\|_r$. We will prove that $C_{r,t+1} < \infty$ and by induction that $C_{r,t} < \infty$ for any $t \leq 0$. Let $t \in T_j^*$, then

$$|X_t - Z_{t,j}| \leq |M_{\theta_j^*}(X_{t-1}, \ldots) - M_{\theta_j^*}(0,0,\ldots)||\xi_t| + |f_{\theta_j^*}(X_{t-1}, \ldots) - f_{\theta_j^*}(0,0,\ldots)|.$$

We obtain for all t, by independence of $(\xi_j)_{j>t}$ and X_t:

$$\begin{aligned}\|X_t - Z_t\|_r \leq \|M_{\theta_j^*}(X_{t-1}, \ldots) - M_{\theta_j^*}(0,0,\ldots)\|_r \|\xi_t\|_r \\ + \|f_{\theta_j^*}(X_{t-1}, \ldots) - f_{\theta_j^*}(0,0,\ldots)\|_r.\end{aligned}$$

Then, we have:

$$\begin{aligned}\|M_{\theta_j^*}(X_{t-1}, \ldots) - M_{\theta_j^*}(0,0,\ldots)\|_r &\leq \sum_{i=1}^{\infty} \alpha_i^{(0)}(M_{\theta_j^*}, \theta_j^*)\|X_{t-i}\|_r \\ &\leq C_{r,t} \sum_{i=1}^{\infty} \alpha_i^{(0)}(M_{\theta_j^*}, \theta_j^*), \\ \|f_{\theta_j^*}(X_{t-1}, \ldots) - f_{\theta_j^*}(0,0,\ldots)\|_r &\leq \sum_{i=1}^{\infty} \alpha_i^{(0)}(f_{\theta_j^*}, \theta_j^*)\|X_{t-i}\|_r \\ &\leq C_{r,t} \sum_{i=1}^{\infty} \alpha_i^{(0)}(f_{\theta_j^*}, \theta_j^*).\end{aligned}$$

We deduce that

$$\|X_t\|_r \le \|Z_{t,j}\|_r + C_{r,t}\left(\sum_{i=1}^{\infty}\alpha_i^{(0)}(f_{\theta_j^*},\{\theta_j^*\}) + (\mathbb{E}\|\xi_0\|^r)^{1/r}\sum_{i=1}^{\infty}\alpha_i^{(0)}(M_{\theta_j^*},\{\theta_j^*\})\right).$$

Thus, $\|X_t\|^r < \infty$ and $C_{r,t+1} < \infty$. As (ξ_t) is iid, remark that $\|Z_{t,j}\|_r = \|Z_{0,j}\|_r =: \|Z_j\|_r$. For any $i \le t$, we obtain similarly that $\|X_i\|_r \le \|Z_j\|_r + C_{r,i}\beta^{(0)}(\theta_j^*) \le \|Z_j\|_r + C_{r,t+1}\beta^{(0)}(\theta_j^*)$ since $C_{r,i} \le C_{r,t+1} < \infty$. Thus, by definition of $C_{r,t+1} = \sup_{i\le t}\|X_t\|_r$ we obtain

$$C_{r,t+1} \le \max_{1\le j\le K^*}\left\{\|Z_{t,j}\|_r + C_{r,t+1}\beta^{(0)}(\theta_j^*)\right\}.$$

The Proposition is established with $C = \max_{1\le j\le K^*}\|Z_{t,j}\|_r/(1-\beta^{(0)}(\theta_j^*))$.

In the ARCH-type case when $f_\theta = 0$ and $A_0(h_\theta,\{\theta\})$ holds with $\tilde\beta^{(0)}(\theta) < 1$, we follow the same reasoning than previously starting from the inequality

$$\|X_t^2 - (M_{\theta_j^*}(0,0,\ldots)\xi_t)^2\|_{r/2} \le \|h_{\theta_j^*}(X_{t-1},\ldots) - h_{\theta_j^*}(0,0,\ldots)\|_{r/2}\|\xi_t^2\|_{r/2}.$$

We obtain the desired result with $C = \max_{1\le j\le K^*}\|M_{\theta_j^*}(0,0,\ldots)\ \xi_0\|_r/\sqrt{1-\tilde\beta^{(0)}(\theta_j^*)}$ in this case. □

6.2. Some preliminary result

The following technical lemma is useful in the sequel:

Lemma 6.1. *Suppose that $\theta_j^* \in \Theta$ for $j = 1,\ldots,K^*$, Θ satisfying A with $r \ge 2$ and under the assumptions $A_0(f_\theta,\Theta)$, $A_0(M_\theta,\Theta)$ (or $A_0(h_\theta,\Theta)$) and $D(\Theta)$, then there exists $C > 0$ such that*

$$\text{for all } t \in \mathbb{Z},\quad \mathbb{E}\Big(\sup_{\theta\in\Theta}\big|q_t(\theta)\big|\Big) \le C.$$

Proof. Using the inequality $(a+b)^2 \le 2(a^2+b^2)$, we have for all $t \in \mathbb{Z}$:

$$\begin{aligned}\big\|f_\theta^t\big\|_\Theta^2 &\le 2\Big(\big\|f_\theta^t - f_\theta(0,\ldots)\big\|_\Theta^2 + \big\|f_\theta(0,\ldots)\big\|_\Theta^2\Big)\\ &\le 2\Big(\Big(\sum_{i\ge1}\alpha_i^{(0)}(f_\theta,\Theta)\Big)\cdot\sum_{i\ge1}\alpha_i^{(0)}(f_\theta,\Theta)|X_{t-i}|^2 + \big\|f_\theta(0,\ldots)\big\|_\Theta^2\Big),\end{aligned}$$

therefore

$$\mathbb{E}\big\|f_\theta^t\big\|_\Theta^2 \le 2\Big(C\Big(\sum_{i\ge1}\alpha_i^{(0)}(f_\theta,\Theta)\Big)^2 + \big\|f_\theta(0,\ldots)\big\|_\Theta^2\Big).$$

Thus $\mathbb{E}\big\|f_\theta^t\|_\Theta^2 \le C$ for all $t \in \mathbb{Z}$ and similarly $\mathbb{E}(\|h_\theta^t\|_\Theta) = \mathbb{E}(\|M_\theta^t\|_\Theta^2) \le C$. Yet, under assumption $(D(\Theta))$, we have: $|q_t(\theta)| \le \frac{1}{\underline{h}}|X_t - f_\theta^t|^2 + |\log(h_\theta^t)|$ and using inequality $\log x \le x-1$ for all $x > 0$, it follows:

$$|\log(h_\theta^t)| = \Big|\log(\underline{h}) + \log(\frac{h_\theta^t}{\underline{h}})\Big| \le 1 + |\log(\underline{h})| + \frac{1}{\underline{h}}h_\theta^t.$$

Finally, we have for all $t \in \mathbb{Z}$:

$$\mathbb{E}\big(\sup_{\theta\in\Theta}|q_t(\theta)|\big) \quad \leq \quad 1+|\log \underline{h}|+\frac{1}{\underline{h}}\big(\mathbb{E}\|h_\theta^t\|_\Theta+2\mathbb{E}|X_t|^2+2\mathbb{E}\|f_\theta^t\|_\Theta^2\big)\leq C.$$

□

6.3. Comparison with stationary solutions

In the following, we assume that $\theta_j^*\in\Theta$ for all $j=1,\dots,K^*$, Θ satisfying A with $r\geq 1$. It comes from [2] that the equation

$$X_{t,j}=M_{\theta_j^*}\big((X_{t-k,j})_{k\in\mathbb{N}^*}\big)\cdot\xi_t+f_{\theta_j^*}\big((X_{t-k,j})_{k\in\mathbb{N}^*}\big)\quad\text{for all } t\in\mathbb{Z}$$

has r order stationary solution $\big(X_{t,j}\big)_{t\in\mathbb{Z}}$ for any $j=1,\dots,K^*$. Then

Lemma 6.2. *Assume that the assumptions $A_0(f_\theta,\Theta)$, $A_0(M_\theta,\Theta)$ (or $A_0(h_\theta,\Theta)$) hold and that $\theta_j^*\in\Theta$ for $j=1,\dots,K^*$, Θ satisfying A for $r\geq 2$. Then:*

1. *$X_t=X_{t,1}$ for all $t\leq t_1^*$;*
2. *There exists $C>0$ such that for any $j\in\{2,\dots,K^*\}$, for all $t\in T_j^*$,*

$$\begin{aligned}\|X_t-X_{t,j}\|_r &\leq C\Big(\inf_{1\leq p\leq t-t_{j-1}^*}\Big\{\beta^{(0)}(\theta_j^*)^{(t-t_{j-1}^*)/p}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\Big\}\Big)\\ \|X_t^2-X_{t,j}^2\|_{r/2} &\leq C\Big(\inf_{1\leq p\leq t-t_{j-1}^*}\Big\{\tilde\beta^{(0)}(\theta_j^*)^{(t-t_{j-1}^*)/p}+\sum_{i\geq p}\tilde\beta_i^{(0)}(\theta_j^*)\Big\}\Big).\end{aligned}$$

Proof. 1. It is obvious from the definition of X.
2. Let $j\in\{2,\dots,K^*\}$, we proceed by induction on $t\in T_j^*$.

First consider the general case where $A_0(f_\theta,\{\theta\})$ and $A_0(M_\theta,\{\theta\})$ hold with $\beta^{(0)}(\theta)<1$. By Proposition 2.2, there exists $C_r\geq 0$ such that $\|X_t-X_{t,j}\|_r\leq\|X_t\|_r+\|X_{t,j}\|_r\leq C+\max_{1\leq j\leq K^*}\|X_{0,j}\|_r\leq C_r$ for all $j=1,\dots,K^*$ and $t\in\mathbb{Z}$. For $1\leq p\leq t-t_{j-1}^*$ let $u_\ell:=\sup_{t_{j-1}^*+\ell p\leq i\leq t_j^*}\|X_i-X_{i,j}\|_r$. Then $\|X_t-X_{t,j}\|_r\leq u_{[(t-t_{j-1}^*)/p]}$ and for any $t\leq i\leq t_j^*$:

$$\begin{aligned}\|X_i-X_{i,j}\|_r &\leq\sum_{k\geq 1}\beta_k^{(0)}(\theta_j^*)\|X_{i-k}-X_{i-k,j}\|_r\\ &\leq\sum_{k=1}^{p}\beta_k^{(0)}(\theta_j^*)\|X_{i-k}-X_{i-k,j}\|_r+C_r\sum_{k>p}\beta_k^{(0)}(\theta_j^*)\\ &\leq\beta^{(0)}(\theta_j^*)u_{[(t-t_{j-1}^*)/p]-1}+C_r\sum_{k>p}\beta_k^{(0)}(\theta_j^*).\end{aligned}$$

Similarly, it is easy to show that for all $1\leq\ell\leq[(t-t_{j-1}^*)/p]$ we have

$$u_\ell\leq\beta^{(0)}(\theta_j^*)u_{\ell-1}+C_r\sum_{k>p}\beta_k^{(0)}(\theta_j^*).$$

Denote $a = \beta^{(0)}(\theta_j^*) < 1$, $b = C_r \sum_{k>p} \beta_k^{(0)}(\theta_j^*)$ such that $u_\ell \leq a u_{\ell-1} + b$. Considering $w_0 = u_0$ and $w_l = a w_{\ell-1} + b$, then $w_\ell = a^\ell w_0 + b(1-a^{\ell-1})/(1-a) \leq a^\ell w_0 + b/(1-a)$. Since $u_0 \leq C_r$ by definition and $u_\ell \leq w_\ell$ for any ℓ, we have:

$$\begin{aligned} u_\ell &\leq a^\ell u_0 + \frac{b}{1-a} \leq \big(\beta^{(0)}(\theta_j^*)\big)^\ell C_r + \frac{C_r}{1-\beta^{(0)}(\theta_j^*)} \sum_{k>p} \beta_k^{(0)}(\theta_j^*) \\ &\leq \frac{C_r}{1-\beta^{(0)}(\theta_j^*)} \Big(\big(\beta^{(0)}(\theta_j^*)\big)^\ell + \sum_{k>p} \beta_k^{(0)}(\theta_j^*) \Big). \end{aligned}$$

Thus for all $1 \leq p \leq t - t_{j-1}^*$

$$\begin{aligned} \|X_t - X_{t,j}\|_r &\leq \beta^{(0)}(\theta_j^*) u_{[(t-t_{j-1}^*)/p]-1} + C_r \sum_{k>p} \beta_k^{(0)}(\theta_j^*) \\ &\leq C\big(\beta^{(0)}(\theta_j^*)^{(t-t_{j-1}^*)/p} + \sum_{i\geq p} \beta_i^{(0)}(\theta_j^*)\big) \end{aligned}$$

and Lemma 6.2 is proved.

In the ARCH-type case when $f_\theta = 0$ and $A_0(h_\theta, \{\theta\})$ holds with $\tilde\beta^{(0)}(\theta) < 1$, we follow the same reasoning than previously starting from the inequality

$$\|X_i^2 - X_{i,j}^2\|_{r/2} \leq \sum_{k\geq 1} \tilde\beta_k^{(0)}(\theta_j^*) \|X_{i-k}^2 - X_{i-k,j}^2\|_{r/2}.$$

For all $j = 1, \ldots, K^*$ and $t \in \mathbb{Z}$, by Proposition 2.2, $\|X_i^2 - X_{i,j}^2\|_{r/2} \leq C_r^2$ and therefore

$$\tilde u_\ell \leq \tilde\beta^{(0)}(\theta_j^*) \tilde u_{\ell-1} + C_r^2 \sum_{k>p} \tilde\beta_k^{(0)}(\theta_j^*)$$

for $\tilde u_\ell = \sup_{t_{j-1}^*+\ell p \leq i \leq t_j^*} \|X_i^2 - X_{i,j}^2\|_{r/2}$ and Lemma 6.2 is proved. ☐

6.4. The asymptotic behavior of the likelihood

For the process $\big(X_{t,j}\big)_{t\in T_j^*,\, j=1,\ldots,K^*}$, for any $j \in \{1, \ldots, K^*\}$ and $s \in T_j^*$ denote:

$$q_{s,j}(\theta) := \frac{\big(X_{s,j} - f_\theta^{s,j}\big)^2}{h_\theta^{s,j}} + \log\big(h_\theta^{s,j}\big) \tag{12}$$

with $f_\theta^{s,j} := f_\theta(X_{s-1,j}, X_{s-2,j}, \ldots)$, $M_\theta^{s,j} := M_\theta(X_{s-1,j}, X_{s-2,j}, \ldots)$ and $h_\theta^{s,j} := (M_\theta^{s,j})^2$. For any $T \subset T_j^*$, denote

$$L_{n,j}(T, \theta) := -\frac{1}{2} \sum_{s\in T} q_{s,j}(\theta)$$

the likelihood of the j^{th} stationary model computed on T.

Lemma 6.3. *Assume that $D(\Theta)$ holds.*

1. *If the assumption H_0 with $r \geq 2$ holds then for all $j = 1, \dots, K^*$:*

$$\frac{v_{n_j^*}}{n_j^*} \left\| L_n\big(T_j^*, \theta\big) - L_{n,j}\big(T_j^*, \theta\big) \right\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$$

2. *For $i = 1, 2$, if the assumption H_i with $r \geq 4$ holds then for all $j = 1, \dots, K^*$:*

$$\frac{v_{n_j^*}}{n_j^*} \Big\| \frac{\partial^i L_n\big(T_j^*, \theta\big)}{\partial \theta^i} - \frac{\partial^i L_{n,j}\big(T_j^*, \theta\big)}{\partial \theta^i} \Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$$

Proof. 1-) For any $\theta \in \Theta$,

$$\Big| \frac{1}{n_j^*} L_n\big(T_j^*, \theta\big) - \frac{1}{n_j^*} L_{n,j}\big(T_j^*, \theta\big) \Big| \leq \frac{1}{n_j^*} \sum_{k=1}^{n_j^*} |q_{t_{j-1}^*+k}(\theta) - q_{t_{j-1}^*+k,j}(\theta)|.$$

Then:

$$v_{n_j^*} \Big\| \frac{1}{n_j^*} L_n\big(T_j^*, \theta\big) - \frac{1}{n_j^*} L_{n,j}\big(T_j^*, \theta\big) \Big\|_\Theta \leq \frac{v_{n_j^*}}{n_j^*} \sum_{k=1}^{n_j^*} \|q_{t_{j-1}^*+k}(\theta) - q_{t_{j-1}^*+k,j}(\theta)\|_\Theta.$$

By Corollary 1 of Kounias and Weng [16], with $r \leq 4$ and no loss of generality, the proof of Lemma 6.3 1-) is achieved if

$$\sum_{k\geq 1} \big(\frac{v_k}{k}\big)^{r/4} \mathbb{E}\big(\|q_{t_{j-1}^*+k}(\theta) - q_{t_{j-1}^*+k,j}(\theta)\|_\Theta^{r/4}\big) < \infty. \tag{13}$$

Let us prove (13). For any $\theta \in \Theta$, we have:

$$\begin{aligned} &|q_s(\theta) - q_{s,j}(\theta)| \leq \frac{1}{\underline{h}^2} |X_s - f_\theta^s|^2 |h_\theta^s - h_\theta^{s,j}| \\ &+ \frac{1}{\underline{h}} \big(|X_s^2 - X_{s,j}^2| + |f_\theta^s - f_\theta^{s,j}||f_\theta^s + f_\theta^{s,j} + 2X_s| + 2|f_\theta^{s,j}||X_s - X_{s,j}| + |h_\theta^s - h_\theta^{s,j}|\big). \end{aligned} \tag{14}$$

First consider the general case with $A_0(f_\theta, \{\theta\})$ and $A_0(M_\theta, \{\theta\})$ hold and $\beta^{(0)}(\theta) < 1$:

$$\begin{aligned} \|q_s(\theta) - q_{s,j}(\theta)\|_\Theta \leq{}& C\big(1 + |X_{s,j}| + |X_s|^2 + \|f_\theta^{s,j}\|_\Theta + \|f_\theta^s\|_\Theta^2\big) \\ &\times \big(|X_s - X_{s,j}| + \|f_\theta^s - f_\theta^{s,j}\|_\Theta + \|h_\theta^s - h_\theta^{s,j}\|_\Theta\big), \end{aligned}$$

and by Cauchy-Schwartz Inequality,

$$\begin{aligned} (\mathbb{E}\|q_s(\theta) - q_{s,j}(\theta)\|_\Theta^{r/4})^2 \leq{}& C\mathbb{E}\big[\big(1 + |X_{s,j}| + |X_s|^2 + \|f_\theta^{s,j}\|_\Theta + \|f_\theta^s\|_\Theta^2\big)^{r/2}\big] \\ &\times \mathbb{E}\big[\big(|X_s - X_{s,j}| + \|f_\theta^s - f_\theta^{s,j}\|_\Theta + \|h_\theta^s - h_\theta^{s,j}\|_\Theta\big)^{r/2}\big]. \end{aligned}$$

Using Proposition (2.2) and the argument of the proof of Lemma (6.1) we claim that $\mathbb{E}|X_s|^r \leq C$, $\mathbb{E}\|f_\theta^s\|_\Theta^r \leq C$ and that $\mathbb{E}\|f_\theta^{s,j}\|_\Theta^r \leq C$. Thus:

$$(\mathbb{E}\|q_s(\theta)-q_{s,j}(\theta)\|_\Theta^{r/4})^2 \leq C\big(\mathbb{E}|X_s-X_{s,j}|^{r/2}+\mathbb{E}\|f_\theta^s-f_\theta^{s,j}\|_\Theta^{r/2}+\mathbb{E}\|h_\theta^s-h_\theta^{s,j}\|_\Theta^{r/2}\big). \tag{15}$$

Since $r/2 \geq 1$, we will use the $L^{r/2}$ norm. By Lemma 6.2:

$$\begin{aligned}\|X_s-X_{s,j}\|_{r/2} \leq \|X_s-X_{s,j}\|_r &\leq C \inf_{1\leq p\leq k}\big\{\beta^{(0)}(\theta_j^*)^{k/p}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big\}\\ &\leq C \inf_{1\leq p\leq k/2}\big\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big\}.\end{aligned}$$

$$\Longrightarrow \quad \mathbb{E}|X_s-X_{s,j}|^{r/2} \leq C\Big(\inf_{1\leq p\leq k}\big\{\beta^{(0)}(\theta_j^*)^{k/p}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big\}\Big)^{r/2}. \tag{16}$$

Moreover, as $(A_0(M_\theta,\Theta))$ holds, we have:

$$\|\|h_\theta^s-h_\theta^{s,j}\|_\Theta\|_{r/2} \leq C\sum_{i\geq 1}\alpha_i^{(0)}(M_\theta,\Theta)\|X_{s-i}-X_{s-i,j}\|_r. \tag{17}$$

From (17) we obtain:

$$\begin{aligned}\big\|\|h_\theta^s-h_\theta^{s,j}\|_\Theta\big\|_{r/2} \leq C\Big(&\sum_{i=1}^{k/2-1}\alpha_i^{(0)}(M_\theta,\Theta)\|X_{s-i}-X_{s-i,j}\|_r\\ &+\sum_{i\geq k/2}\alpha_i^{(0)}(M_\theta,\Theta)\|X_{s-i}-X_{s-i,j}\|_r\Big).\end{aligned}$$

For all $s \geq t_{j-1}^*$ and $1 \leq i \leq k/2-1$, then $s-i > t_{j-1}^*$, $s-i > k/2$ and by Lemma 6.2:

$$\begin{aligned}\|X_{s-i}-X_{s-i,j}\|_r &\leq C \inf_{1\leq p\leq k-i}\big\{\beta^{(0)}(\theta_j^*)^{(k-i)/p}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big\}\\ &\leq C \inf_{1\leq p\leq k/2}\big\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big\}\end{aligned}$$

Thus, we can find $C>0$ not depending on s such that:

$$\begin{aligned}&\mathbb{E}\|h_\theta^s-h_\theta^{s,j}\|_\Theta^{r/2}\\ &\quad\leq C\Big(\inf_{1\leq p\leq k/2}\big\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big\}+\sum_{i\geq k/2}\alpha_i^{(0)}(M_\theta,\Theta)\Big)^{r/2}.\end{aligned} \tag{18}$$

Similarly, we obtain:

$$\begin{aligned}&\mathbb{E}\|f_\theta^s-f_\theta^{s,j}\|_\Theta^{r/2}\\ &\quad\leq C\Big(\inf_{1\leq p\leq k/2}\big\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big\}+\sum_{i\geq k/2}\alpha_i^{(0)}(f_\theta,\Theta)\Big)^{r/2}.\end{aligned} \tag{19}$$

Relations (15), (16), (18) et (19) give (the same inequality holds with h_θ replaced by M_θ):

$$\mathbb{E}\,\|q_s(\theta)-q_{s,j}(\theta)\|_\Theta^{r/4}\le C\Big[\Big(\inf_{1\le p\le k/2}\big\{\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\big\}\Big)^{r/4}+\Big(\sum_{i\ge k/2}\alpha_i^{(0)}(f_\theta,\Theta)\Big)^{r/4}+\Big(\sum_{i\ge k/2}\alpha_i^{(0)}(M_\theta,\Theta)\Big)^{r/4}\Big]. \quad (20)$$

By definition $u_k=kc^*/\log(k)$ ($\le k/2$ for large value of k) satisfies the relation

$$\sum_{k\ge 1}\big(\frac{v_k}{k}\big)^{r/4}\big(\beta^{(0)}(\theta_j^*)\big)^{rk/8u_k}<\infty.$$

Choosing $p=u_k$ in (20) we obtain:

$$\begin{aligned}\sum_{k\ge 1}\big(\frac{v_k}{k}\big)^{r/4}&\mathbb{E}\big(\|q_{t_{j-1}^*+k}(\theta)-q_{t_{j-1}^*+k,j}(\theta)\|_\Theta^{r/4}\big)\\ &\le\sum_{k\ge 1}\big(\frac{v_k}{k}\big)^{r/4}\big(\beta^{(0)}(\theta_j^*)\big)^{rk/8u_k}+\sum_{k\ge 1}\big(\frac{v_k}{k}\big)^{r/4}\Big(\sum_{i\ge u_k}\beta_i^{(0)}(\theta_j^*)\Big)^{r/4}\\ &\qquad+\sum_{k\ge 1}\big(\frac{v_k}{k}\big)^{r/4}\Big(\sum_{i\ge k/2}\big(\alpha_i^{(0)}(f_\theta,\Theta)+\alpha_i^{(0)}(M_\theta,\Theta)\big)\Big)^{r/4}.\end{aligned}$$

This bound is finite by assumption and therefore (13) is established.

In the ARCH-type case when $f_\theta=0$ and $A_0(h_\theta,\{\theta\})$ holds with $\tilde\beta^{(0)}(\theta)<1$, we follow the same reasoning than previously remarking that (14) has the simplified form:

$$|q_s(\theta)-q_{s,j}(\theta)|\le\frac{1}{\underline{h}^2}X_s^2|h_\theta^s-h_\theta^{s,j}|+\frac{1}{\underline{h}}|X_s^2-X_{s,j}^2|+\frac{1}{\underline{h}}|h_\theta^s-h_\theta^{s,j}|.$$

Then

$$(\mathbb{E}\|q_s(\theta)-q_{s,j}(\theta)\|_\Theta^{r/4})^2\le C\mathbb{E}\big[\big(|X_s^2-X_{s,j}^2|+\|h_\theta^s-h_\theta^{s,j}\|_\Theta\big)^{r/2}\big].$$

As $\|\|h_\theta^s-h_\theta^{s,j}\|_\Theta\|_{r/2}\le C\sum_{i\ge 1}\alpha_i^{(0)}(h_\theta,\Theta)\|X_{s-i}^2-X_{s-i,j}^2\|_{r/2}$ we derive from Lemma 6.2,

$$\mathbb{E}\,\|q_s(\theta)-q_{s,j}(\theta)\|_\Theta^{r/4}\le C\,\Big[\Big(\inf_{1\le p\le k/2}\big\{\tilde\beta^{(0)}(\theta_j^*)^{k/(2p)}+\sum_{i\ge p}\tilde\beta_i^{(0)}(\theta_j^*)\big\}\Big)^{r/4}+\Big(\sum_{i\ge k/2}\alpha_i^{(0)}(h_\theta,\Theta)\Big)^{r/4}\Big].$$

We easily conclude to the result by choosing $p=u_k$ as above.

2-) We detail the proof for one order derivation in the general case where $A_0(f_\theta,\{\theta\})$ and $A_0(M_\theta,\{\theta\})$ hold with $\beta^{(0)}(\theta)<1$. The proofs of the other cases follow the same reasoning.

Let $j\in\{1,\dots,K^*\}$ and $i=1,\dots,d$, we have:

$$\frac{v_{n_j^*}}{n_j^*}\Big\|\frac{\partial L_n(T_j^*,\theta)}{\partial\theta_i}-\frac{\partial L_{n,j}(T_j^*,\theta)}{\partial\theta_i}\Big\|_\Theta\leq\frac{v_{n_j^*}}{n_j^*}\sum_{k=1}^{n_j^*}\Big\|\frac{\partial q_{t_{j-1}^*+k}(\theta)}{\partial\theta_i}-\frac{\partial q_{t_{j-1}^*+k,j}(\theta)}{\partial\theta_i}\Big\|_\Theta.$$

As previously, using Corollary 1 of [16], when $r\leq 4$ with no loss of generality, Lemma 6.3 2-) will be established if

$$\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\mathbb{E}\Big(\Big\|\frac{\partial q_{t_{j-1}^*+k}(\theta)}{\partial\theta_i}-\frac{\partial q_{t_{j-1}^*+k,j}(\theta)}{\partial\theta_i}\Big\|_\Theta^{r/4}\Big)<\infty. \tag{21}$$

For any $s\geq t_{j-1}^*$ denote $k=s-t_{j-1}^*$. For any $\theta\in\Theta$, we have:

$$\begin{aligned}\frac{\partial q_s(\theta)}{\partial\theta_i}&=-2\frac{(X_s-f_\theta^s)}{h_\theta^s}\frac{\partial f_\theta^s}{\partial\theta_i}-\frac{(X_s-f_\theta^s)^2}{(h_\theta^s)^2}\frac{\partial h_\theta^s}{\partial\theta_i}+\frac{1}{h_\theta^s}\frac{\partial h_\theta^s}{\partial\theta_i}\\ \frac{\partial q_{s,j}(\theta)}{\partial\theta_i}&=-2\frac{(X_{s,j}-f_\theta^{s,j})}{h_\theta^{s,j}}\frac{\partial f_\theta^{s,j}}{\partial\theta_i}-\frac{(X_{s,j}-f_\theta^{s,j})^2}{(h_\theta^{s,j})^2}\frac{\partial h_\theta^{s,j}}{\partial\theta_i}+\frac{1}{h_\theta^{s,j}}\frac{\partial h_\theta^{s,j}}{\partial\theta_i}.\end{aligned}$$

Thus, using $|a_1b_1c_1-a_2b_2c_2|\leq|a_1-a_2||b_2||c_2|+|b_1-b_2||a_1||c_2|+|c_1-c_2||a_1||b_1|$,

$$\begin{aligned}&\Big\|\frac{\partial q_s(\theta)}{\partial\theta_i}-\frac{\partial q_{s,j}(\theta)}{\partial\theta_i}\Big\|_\Theta\leq 2\Big(\frac{1}{\underline{h}^2}\|h_\theta^s-h_\theta^{s,j}\|_\Theta\|X_{s,j}-f_\theta^{s,j}\|_\Theta\Big\|\frac{\partial f_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta\\ &\quad+\frac{1}{\underline{h}}(|X_s-X_{s,j}|+\|f_\theta^s-f_\theta^{s,j}\|_\Theta)\Big\|\frac{\partial f_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta+\frac{1}{\underline{h}}\Big\|\frac{\partial f_\theta^s}{\partial\theta_i}-\frac{\partial f_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta\|X_s-f_\theta^s\|_\Theta\Big)\\ &\quad+\frac{2}{\underline{h}^3}\|h_\theta^s-h_\theta^{s,j}\|_\Theta\|X_{s,j}-f_\theta^{s,j}\|_\Theta^2\Big\|\frac{\partial h_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta\\ &\quad+\frac{1}{\underline{h}}(|X_s-X_{s,j}|+\|f_\theta^s-f_\theta^{s,j}\|_\Theta)(|X_s+X_{s,j}|+\|f_\theta^s+f_\theta^{s,j}\|_\Theta)\Big\|\frac{\partial f_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta\\ &\quad+\frac{1}{\underline{h}^2}\Big\|\frac{\partial h_\theta^s}{\partial\theta_i}-\frac{\partial h_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta\|X_s-f_\theta^s\|_\Theta^2+\frac{1}{\underline{h}^2}\|h_\theta^s-h_\theta^{s,j}\|_\Theta\Big\|\frac{\partial h_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta\\ &\quad+\frac{1}{\underline{h}}\Big\|\frac{\partial h_\theta^s}{\partial\theta_i}-\frac{\partial h_\theta^{s,j}}{\partial\theta_i}\Big\|_\Theta.\end{aligned}$$

So for all $s \geq t^*_{j-1}$ it holds:

$$\Big\|\frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i}\Big\|_\Theta \leq C\Big(1 + |X_s|^2 + |X_{s,j}|^2 + \|f^s_\theta\|^2_\Theta + \|f^{s,j}_\theta\|^2_\Theta + \Big\|\frac{\partial f^s_\theta}{\partial \theta_i}\Big\|^2_\Theta + \Big\|\frac{\partial f^{s,j}_\theta}{\partial \theta_i}\Big\|^2_\Theta + \Big\|\frac{\partial h^s_\theta}{\partial \theta_i}\Big\|^2_\Theta + \Big\|\frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|^2_\Theta\Big) \times \Big(|X_s - X_{s,j}| + \|f^s_\theta - f^{s,j}_\theta\|_\Theta + \|h^s_\theta - h^{s,j}_\theta\|_\Theta + \Big\|\frac{\partial f^s_\theta}{\partial \theta_i} - \frac{\partial f^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta + \Big\|\frac{\partial h^s_\theta}{\partial \theta_i} - \frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|_\Theta\Big).$$

Since the processes admits finite moments of order r, by Cauchy-Schwartz Inequality:

$$\Big(\mathbb{E}\Big\|\frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i}\Big\|^{r/4}_\Theta\Big)^2 \leq C\Big(\mathbb{E}|X_s - X_{s,j}|^{r/2} + \mathbb{E}(\|f^s_\theta - f^{s,j}_\theta\|^{r/2}_\Theta) + \mathbb{E}(\|h^s_\theta - h^{s,j}_\theta\|^{r/2}_\Theta) + \mathbb{E}\Big\|\frac{\partial f^s_\theta}{\partial \theta_i} - \frac{\partial f^{s,j}_\theta}{\partial \theta_i}\Big\|^{r/2}_\Theta + \mathbb{E}\Big\|\frac{\partial h^s_\theta}{\partial \theta_i} - \frac{\partial h^{s,j}_\theta}{\partial \theta_i}\Big\|^{r/2}_\Theta\Big).$$

As $(A_0(M_\theta, \Theta))$ and $(A_1(M_\theta, \Theta))$ hold necessarily in this case, with the arguments of the proof of 1-), for all $s \geq t^*_{j-1}$,

$$\mathbb{E}\Big\|\frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i}\Big\|^{r/4}_\Theta \leq C\Big[\Big(\inf_{1\leq p\leq k/2}\big\{\beta^{(0)}(\theta^*_j)^{k/(2p)} + \sum_{i\geq p}\beta^{(0)}_i(\theta^*_j)\big\}\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(0)}_i(f_\theta, \Theta)\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(0)}_i(M_\theta, \Theta)\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(1)}_i(f_\theta, \Theta)\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(1)}_i(M_\theta, \Theta)\Big)^{r/4}\Big]$$

Choosing $p = u_k = kc^*/\log(k)$, we show (as in proof of 1-)) that:

$$\sum_{k\geq 1}\Big(\frac{v_k}{k}\Big)^{r/4}\mathbb{E}\Big(\Big\|\frac{\partial q_{t^*_{j-1}+k}(\theta)}{\partial \theta_i} - \frac{\partial q_{t^*_{j-1}+k,j}(\theta)}{\partial \theta_i}\Big\|^{r/4}_\Theta\Big) < \infty.$$

Therefore (21) is proved and Lemma 6.3 2-) also. □

6.5. Consistency when the breaks are known

When the breaks are known, we can choose $v_n = 1$ for all n; in (4), the penalty term does not matter at all.

Proposition 6.1. *For all $j = 1, \dots, K^*$, under the assumptions of Lemma 6.3 1-) with $v_n = 1$ for all n, if the assumption $Id(\Theta)$ holds then*

$$\widehat{\theta}_n(T^*_j) \xrightarrow[n\to\infty]{a.s.} \theta^*_j.$$

Proof. Let us first give the following useful corollary of Lemma 6.3

Corollary 6.1. *i-) under the assumptions of Lemma 6.3 1-) we have:*

$$\Big\| \frac{1}{n_j^*} \widehat{L}_n\big(T_j^*, \theta\big) - \mathcal{L}_j(\theta) \Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0 \quad \textit{with} \quad \mathcal{L}_j(\theta) = -\frac{1}{2} \mathbb{E}\left(q_{0,j}(\theta)\right).$$

ii-) Under assumptions of Lemma 6.3 2-) we have:

$$\Big\| \frac{1}{n_j^*} \frac{\partial^i \widehat{L}_n(T_j^*, \theta)}{\partial \theta^i} - \frac{\partial^i \mathcal{L}_j(\theta)}{\partial \theta^i} \Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0 \quad \textit{with} \quad \frac{\partial^i \mathcal{L}_j(\theta)}{\partial \theta^i} = -\frac{1}{2} \mathbb{E}\left(\frac{\partial^i q_{0,j}(\theta)}{\partial \theta^i} \right).$$

We conclude the proof of Proposition 6.1 using $\mathcal{L}_j(\theta) = -\frac{1}{2}\mathbb{E}\left(q_{0,j}(\theta)\right)$ has a unique maximum in θ_j^* (see [13]). From the almost sure convergence of the quasi-likelihood in i-) of Corollary 6.1, it comes:

$$\widehat{\theta}_n(T_j^*) = \operatorname*{Argmax}_{\theta \in \Theta} \left(\frac{1}{n_j^*} \widehat{L}_n\big(T_j^*, \theta\big) \right) \xrightarrow[n\to\infty]{\text{a.s.}} \theta_j^*.$$

□

Proof of Corollary 6.1. Note that the proof of Lemma 6.3 can be repeated by replacing L_n by the quasi-likelihood $\widehat{L}_n$. Thus, we obtain for $i = 0, 1, 2$,

$$\frac{v_{n_j^*}}{n_j^*} \left\| \frac{\partial^i \widehat{L}_n\big(T_j^*, \theta\big)}{\partial \theta^i} - \frac{\partial^i L_{n,j}\big(T_j^*, \theta\big)}{\partial \theta^i} \right\|_\Theta \xrightarrow[n\to\infty]{} 0. \tag{22}$$

i-) Let $j \in 1, \dots, K^*$. From [2], we have:

$$\Big\| \frac{1}{n_j^*} L_{n,j}\big(T_j^*, \theta\big) - \mathcal{L}_j(\theta) \Big\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

Using (22), the convergence to the limit likelihood follows.

ii-) From Lemma 4 and Theorem 1 of [2], $\big\| \frac{1}{n_j^*} \frac{\partial^i L_{n,j}(T_j^*, \theta)}{\partial \theta^i} - \frac{\partial^i \mathcal{L}_j(\theta)}{\partial \theta^i} \big\|_\Theta \xrightarrow[n\to\infty]{\text{a.s.}} 0$ for $i = 1,\ 2$ and we conclude from (22).

□

6.6. Proof of Theorem 3.1

This proof is divided into two parts. In **part (1)** K^* is assumed to be known and we show $(\widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n) \xrightarrow[n\to\infty]{\mathcal{P}} (\underline{\tau}^*, \underline{\theta}^*)$. In **part (2)**, K^* is unknown and we show $\widehat{K}_n \xrightarrow[n\to\infty]{\mathcal{P}} K^*$ which ends the proof of Theorem 3.1.

Part (1). Assume that K^* is known and denote for any $\underline{t} \in \mathcal{F}_{K^*}$:

$$\widehat{I}_n(\underline{t}) := \widehat{J}_n(K^*, \underline{t}, \widehat{\underline{\theta}}_n(\underline{t})) = -2 \sum_{k=1}^{K^*} \sum_{j=1}^{K^*} \widehat{L}_n\left(T_k \cap T_j^*, \widehat{\theta}_n(T_k)\right)$$

It comes that $\widehat{\underline{t}}_n = \mathrm{Argmin}_{\underline{t}\in\mathcal{F}_{K^*}}(\widehat{I}_n(\underline{t}))$. We show that $\widehat{\underline{\tau}}_n \xrightarrow[n\to\infty]{\mathcal{P}} \underline{\tau}^*$ as it implies $\widehat{\theta}_n(\widehat{T}_{n,j}) - \widehat{\theta}_n(T_j^*) \xrightarrow[n\to\infty]{\mathcal{P}} 0$ and from Proposition 6.1 $\widehat{\theta}_n(\widehat{T}_{n,j}) \xrightarrow[n\to\infty]{\mathcal{P}} \theta_j^*$ for all $j = 1,\dots,K^*$. Without loss of generality, assume that $K^* = 2$ and let (u_n) be a sequence of positive integers satisfying $u_n \to \infty$, $u_n/n \to 0$ and for some $0 < \eta < 1$

$$\begin{aligned} V_{\eta,u_n} &= \{\ t\in\mathbb{Z}/\ \ |t-t^*| > \eta n\ ;\ u_n \le t \le n-u_n\ \}, \\ W_{\eta,u_n} &= \{\ t\in\mathbb{Z}/\ \ |t-t^*| > \eta n\ ;\ 0 < t < u_n \ \text{ or }\ n-u_n < t \le n\ \}. \end{aligned}$$

Asymptotically, we have $\mathbb{P}(\|\widehat{\underline{\tau}}_n - \underline{\tau}^*\|_m > \eta) \simeq \mathbb{P}(|\widehat{t}_n - t^*| > \eta n)$. But

$$\begin{aligned} \mathbb{P}(|\widehat{t}_n - t^*| > \eta n) &\le \mathbb{P}\Big(\widehat{t}_n \in V_{\eta,u_n}\Big) + \mathbb{P}\Big(\widehat{t}_n \in W_{\eta,u_n}\Big) \\ &\le \mathbb{P}\Big(\min_{t\in V_{\eta,u_n}}(\widehat{I}_n(t) - \widehat{I}_n(t^*)) \le 0\Big) \\ &\qquad\qquad + \mathbb{P}\Big(\min_{t\in W_{\eta,u_n}}(\widehat{I}_n(t) - \widehat{I}_n(t^*)) \le 0\Big) \end{aligned}$$

we show with similar arguments that these two probabilities tend to 0. We only detail below the proof of $\mathbb{P}(\min_{t\in V_{\eta,u_n}}(\widehat{I}_n(t) - \widehat{I}_n(t^*)) \le 0) \to 0$ for shortness.

Let $t \in V_{\eta,u_n}$ satisfying $t^* \le t$ (with no loss of generality), then $T_1 \cap T_1^* = T_1^*$, $T_2 \cap T_1^* = \emptyset$ and $T_2 \cap T_2^* = T_2$. We decompose:

$$\begin{aligned} \widehat{I}_n(t) - \widehat{I}_n(t^*) = 2\Big(&\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*)) - \widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1)) + \widehat{L}_n(T_1\cap T_2^*, \widehat{\theta}_n(T_2^*)) \\ &- \widehat{L}_n(T_1\cap T_2^*, \widehat{\theta}_n(T_1)) + \widehat{L}_n(T_2, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2, \widehat{\theta}_n(T_2))\Big). \end{aligned} \tag{23}$$

As $\# T_1^* = t^*$, $\#(T_1 \cap T_2^*) = t - t^*$, $\# T_2 = n - t \ge u_n$, each term tends to ∞ with n. Using Proposition 6.1 and Corollary 6.1, we get the following convergence, uniformly on V_{η,u_n},

$$\widehat{\theta}_n(T_1^*) \xrightarrow[n\to\infty]{a.s.} \theta_1^*,\quad \widehat{\theta}_n(T_2^*) \xrightarrow[n\to\infty]{a.s.} \theta_2^*,\quad \widehat{\theta}_n(T_2) \xrightarrow[n\to\infty]{a.s.} \theta_2^*$$

$$\text{and}\quad \Big\|\frac{\widehat{L}_n(T_1^*,\theta)}{n} - \tau_1^*\mathcal{L}_1(\theta)\Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0,\quad \Big\|\frac{\widehat{L}_n(T_1\cap T_2^*,\theta)}{t-t^*} - \mathcal{L}_2(\theta)\Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0,$$
$$\text{and}\quad \Big\|\frac{\widehat{L}_n(T_2,\theta)}{n-t} - \mathcal{L}_2(\theta)\Big\|_\Theta \xrightarrow[n\to\infty]{a.s.} 0.$$

For any $\varepsilon > 0$, there exists an integer N_0 such that for any $n > N_0$,

$$\begin{aligned} &\Big\|\frac{\widehat{L}_n(T_1^*,\theta)}{n} - \tau_1^*\mathcal{L}_1(\theta)\Big\|_\Theta < \frac{\varepsilon}{6},\quad \Big\|\frac{\widehat{L}_n(T_1\cap T_2^*,\theta)}{t-t^*} - \mathcal{L}_2(\theta)\Big\|_\Theta < \frac{\varepsilon}{6}, \\ &\Big|\frac{\widehat{L}_n(T_1^*,\widehat{\theta}_n(T_1^*))}{n} - \tau_1^*\mathcal{L}_1(\theta_1^*)\Big| < \frac{\varepsilon}{6},\quad \Big|\frac{\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{t-t^*} - \mathcal{L}_2(\theta_2^*)\Big| < \frac{\varepsilon}{6} \\ &\qquad \text{and}\qquad \frac{n-t}{n}\Big|\frac{\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2,\widehat{\theta}_n(T_2))}{n-t}\Big| < \frac{\varepsilon}{6}. \end{aligned}$$

Thus, for $n > N_0$,

$$\begin{aligned}\tau_1^*\mathcal{L}_1(\theta_1^*) - \tau_1^*\mathcal{L}_1(\widehat{\theta}_n(T_1)) &= \tau_1^*\mathcal{L}_1(\theta_1^*) - \frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1^*)\big)}{n} + \frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1^*)\big)}{n} \\ &\quad - \frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1)\big)}{n} + \frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1)\big)}{n} - \tau_1^*\mathcal{L}_1(\widehat{\theta}_n(T_1)) \\ &\le \frac{\varepsilon}{6} + \frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1^*)\big)}{n} - \frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1)\big)}{n} + \frac{\varepsilon}{6}.\end{aligned}$$

Then,

$$\frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1^*)\big)}{n} - \frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1)\big)}{n} > \tau_1^*\Big(\mathcal{L}_1(\theta_1^*) - \mathcal{L}_1(\widehat{\theta}_n(T_1))\Big) - \frac{\varepsilon}{3}. \quad (24)$$

Similarly, for $n > N_0$:

$$\frac{\widehat{L}_n\big(T_1 \cap T_2^*, \widehat{\theta}_n(T_2^*)\big)}{n} - \frac{\widehat{L}_n\big(T_1 \cap T_2^*, \widehat{\theta}_n(T_1)\big)}{n} > \eta\Big(\mathcal{L}_2(\theta_2^*) - \mathcal{L}_2(\widehat{\theta}_n(T_1))\Big) - \frac{\varepsilon}{3}. \quad (25)$$

Finally, for $n > N_0$,

$$\frac{\widehat{L}_n(T_2, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2, \widehat{\theta}_n(T_2))}{n} \quad > \quad -\frac{\varepsilon}{6}, \quad (26)$$

and from (23) and inequalities (24), (25) and (26) we obtain uniformly in t and for $n > N_0$:

$$\frac{\widehat{I}_n(t) - \widehat{I}_n(t^*)}{n} > \tau_1^*\Big(\mathcal{L}_1(\theta_1^*) - \mathcal{L}_1(\widehat{\theta}_n(T_1))\Big) + \eta\Big(\mathcal{L}_2(\theta_2^*) - \mathcal{L}_2(\widehat{\theta}_n(T_1))\Big) - \frac{5}{6}\varepsilon.$$

Since $\theta_1^* \neq \theta_2^*$, let $\mathcal{V}_1$, $\mathcal{V}_2$ be two open neighborhoods and disjoint of θ_1^* and θ_2^* respectively,

$$\delta_i := \operatorname*{Inf}_{\theta \in \mathcal{V}_i^c}\Big(\mathcal{L}_i(\theta_i^*) - \mathcal{L}_i(\theta)\Big) > 0 \quad \text{for} \quad i = 1, 2,$$

since the function $\theta \mapsto \mathcal{L}_j(\theta)$ has a strict maximum in θ_j^* (see [13]). With $\varepsilon = \min(\tau_1^*\delta_1,\ \eta\delta_2)$, we get

- if $\widehat{\theta}_n(T_1) \in \mathcal{V}_1$ i.e. $\widehat{\theta}_n(T_1) \in \mathcal{V}_2^c$, then $\frac{\widehat{I}_n(t) - \widehat{I}_n(t^*)}{n} > \eta\delta_2 - \frac{5}{6}\varepsilon \ge \frac{\varepsilon}{6}$;
- If $\widehat{\theta}_n(T_1) \notin \mathcal{V}_1$ i.e. $\widehat{\theta}_n(T_1) \in \mathcal{V}_1^c$, then $\frac{\widehat{I}_n(t) - \widehat{I}_n(t^*)}{n} > \tau_1^*\delta_1 - \frac{5}{6}\varepsilon \ge \frac{\varepsilon}{6}$.

In any case we prove that $\widehat{I}_n(t) - \widehat{I}_n(t^*) > \frac{\varepsilon}{6}n$ for $n > N_0$ and all $t \in V_{\eta,u_n}$. It implies that $\mathbb{P}(\min_{\underline{t} \in V_{\eta,u_n}}(\widehat{I}_n(t) - \widehat{I}_n(t^*)) \le 0) \underset{n\to\infty}{\longrightarrow} 0$ and we show similarly $\mathbb{P}(\min_{\underline{t} \in W_{\eta,u_n}}(\widehat{I}_n(t) - \widehat{I}_n(t^*)) \le 0) \underset{n\to\infty}{\longrightarrow} 0$. It follows directly that $\mathbb{P}(\|\widehat{\underline{\tau}}_n - \underline{\tau}^*\|_m > \eta) \underset{n\to\infty}{\longrightarrow} 0$ for all $\eta > 0$.

Part (2). Now K^* is unknown. For $K \geq 2$, $x = (x_1, \dots, x_{K-1}) \in \mathbb{R}^{K-1}$, $y = (y_1, \dots, y_{K^*-1}) \in \mathbb{R}^{K^*-1}$, denote

$$\|x - y\|_\infty = \max_{1 \leq j \leq K^*-1} \min_{1 \leq k \leq K-1} |x_k - y_j|.$$

The following Lemma follows directly from **Part (1)** and the definition of $\|\cdot\|_\infty$:

Lemma 6.4. *Let $K \geq 1$, $(\widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n)$ obtained by the minimization of $\widehat{J}_n(\underline{t}, \underline{\theta})$ on $\mathcal{F}_K \times \Theta^K$ and $\widehat{\underline{\tau}}_n = \widehat{\underline{t}}_n/n$. Under assumptions of Theorem 3.1, $\|\widehat{\underline{\tau}}_n - \underline{\tau}^*\|_\infty \xrightarrow[n\to+\infty]{P} 0$ if $K \geq K^*$.*

Now we use the following Lemma 6.5 which is proved below (see also [18]):

Lemma 6.5. *Under the assumptions of Lemma 6.3 i-), for any $K \geq 2$, there exists $C_K > 0$ such that:*

$$\forall (\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta^K, \quad e_n(\underline{t}, \underline{\theta}) = 2 \sum_{j=1}^{K^*} \sum_{k=1}^{K} \frac{n_{kj}}{n} (\mathcal{L}_j(\theta_j^*) - \mathcal{L}_j(\theta_k)) \geq \frac{C_K}{n} \|\underline{t} - \underline{t}^*\|_\infty.$$

Continue with the proof of **Part (2)** shared in two parts, *i.e.* we show that $\mathrm{P}(\widehat{K}_n = K) \xrightarrow[n\to+\infty]{} 0$ for $K < K^*$ and $K^* < K \leq K_{\max}$ separately. In any case, we have

$$\begin{aligned}\mathrm{P}(\widehat{K}_n = K) &\leq \mathrm{P}\Big(\inf_{(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta^K} (\widetilde{J}_n(K, \underline{t}, \underline{\theta})) \leq \widetilde{J}_n(K^*, \underline{t}^*, \underline{\theta}^*) \Big) \\ &\leq \mathrm{P}\Big(\inf_{(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta^K} (\widehat{J}_n(K, \underline{t}, \underline{\theta}) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*)) \leq \frac{n}{v_n}(K^* - K).\Big). \end{aligned} \tag{27}$$

i-) For $K < K^*$, we decompose $\widehat{J}_n(K, \underline{t}, \underline{\theta}) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*) = n(d_n(\underline{t}, \underline{\theta}) + e_n(\underline{t}, \underline{\theta}))$ where e_n is defined in Lemma 6.5 and

$$\begin{aligned} d_n(\underline{t}, \underline{\theta}) = 2 \Bigg[&\sum_{j=1}^{K^*} \frac{n_j^*}{n} \Big(\frac{\widehat{L}_n(T_j^*, \theta_j^*)}{n_j^*} - \mathcal{L}_j(\theta_j^*) \Big) \\ &+ \sum_{k=1}^{K} \sum_{j=1}^{K^*} \frac{n_{kj}}{n} \Big(\mathcal{L}_j(\theta_k) - \frac{\widehat{L}_n(T_j^* \cap T_k, \theta_k)}{n_{kj}} \Big) \Bigg]. \end{aligned}$$

It comes from the relation (27) that:

$$\mathrm{P}(\widehat{K}_n = K) \leq \mathrm{P}\Big(\inf_{(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta_K} (d_n(\underline{t}, \underline{\theta}) + e_n(\underline{t}, \underline{\theta})) \leq \frac{1}{v_n}(K^* - K) \Big). \tag{28}$$

Corollary 6.1 ensures that $d_n(\underline{t}, \underline{\theta})) \to 0$ a.s. and uniformly on $\mathcal{F}_K \times \Theta^K$. By Lemma 6.5, there exists $C_K > 0$ such that $e_n(\underline{t}, \underline{\theta}) \geq C_K \|\underline{t} - \underline{t}^*\|_\infty / n$ for all $(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta^K$. But, since $K < K^*$, for any $\underline{t} \in \mathcal{F}_K$, we have

$\|\underline{t}-\underline{t}^*\|_\infty/n = \|\underline{\tau}-\underline{\tau}^*\|_\infty \geq \min_{1\leq j\leq K^*}(\tau_j^*-\tau_{j-1}^*)/2$ that is positive by assumption. Then $e_n(\underline{t},\underline{\theta})>0$ for all $(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta^K$ and since $1/v_n \underset{n\to\infty}{\longrightarrow} 0$, we deduce from (28) that $\mathrm{P}(\widehat{K}_n=K)\underset{n\to\infty}{\longrightarrow} 0$.

ii-) Now let $K^*<K\leq K_{\max}$. From (28) and the Markov Inequality we have:

$$\begin{aligned}\mathrm{P}(\widehat{K}_n=K) &\leq \mathrm{P}\Big(\widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n)-\widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*)+\frac{n}{v_n}(K-K^*)\leq 0\Big)\\ &\leq \mathrm{P}\Big(|\widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n)-\widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*)|\geq \frac{n}{v_n}\Big)\\ &\leq \frac{v_n}{n}\mathbb{E}|\widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n)-\widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*)|. \qquad (29)\end{aligned}$$

Denote $\widehat{\underline{t}}_n=(\widehat{t}_{n,1},\ldots,\widehat{t}_{n,K})$. By Lemma 6.4, there exists some subset $\{k_j\ ,1\leq j\leq K^*-1\}$ of $\{1,\ldots,K-1\}$ such that for any $j=1,\ldots,K^*-1$, $\widehat{t}_{n,k_j}/n\to\tau_j^*$. Denoting $k_0=0$ and $k_{K^*}=K$, we have:

$$\begin{aligned}\widehat{J}_n(K,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n)-\widehat{J}_n(K^*,\underline{t}^*,\underline{\theta}^*) &= 2\Big(\sum_{j=1}^{K^*}\widehat{L}_n(T_j^*,\theta_j^*)-\sum_{k=1}^{K}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big)\\ &= 2\sum_{j=1}^{K^*}\Big[\widehat{L}_n(T_j^*,\theta_j^*)-\sum_{k=k_{j-1}+1}^{k_j}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big]\end{aligned}$$

and from (29) we deduce that:

$$\begin{aligned}\mathrm{P}(\widehat{K}_n=K) &\leq \frac{2v_n}{n}\sum_{j=1}^{K^*}\mathbb{E}\Big|\widehat{L}_n(T_j^*,\theta_j^*)-\sum_{k=k_{j-1}+1}^{k_j}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big|\\ &\leq C\sum_{j=1}^{K^*}\frac{v_{n_j^*}}{n_j^*}\mathbb{E}\Big|\widehat{L}_n(T_j^*,\theta_j^*)-\sum_{k=k_{j-1}+1}^{k_j}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big|.\end{aligned}$$

Since for any $j=1,\ldots,K^*-1$, it comes easily from the proof of Lemma 6.3 that

$$\frac{v_{n_j^*}}{n_j^*}\mathbb{E}\Big|\widehat{L}_n(T_j^*,\theta_j^*)-\sum_{k=k_{j-1}+1}^{k_j}\widehat{L}_n(\widehat{T}_{n,k},\widehat{\theta}_{n,k})\Big|\underset{n\to\infty}{\longrightarrow}0,$$

and therefore $\mathrm{P}(\widehat{K}_n=K)\underset{n\to\infty}{\longrightarrow}0$. □

Proof of Lemma 6.5. Let $K\geq 1$ and consider the real function υ define on $\Theta\times\Theta$ by:

$$\upsilon(\theta,\theta')=\begin{cases}\min\limits_{1\leq j\leq K^*}[\max(\mathcal{L}_j(\theta_j^*)-\mathcal{L}_j(\theta),\ \mathcal{L}_j(\theta_j^*)-\mathcal{L}_j(\theta'))] & \text{if } \theta\neq\theta'\\ 0 \ \text{ if } \ \theta=\theta'.\end{cases}$$

The function υ has positive values and $\upsilon(\theta,\theta')=0$ if and only if $\theta=\theta'$ since the function $\theta\mapsto\mathcal{L}_j(\theta)$ has a strict maximum in θ^*_j (see [13]). By Lemma 3.3 of [17], there exists $C_{\theta^*}>0$ such that for any $(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta_K$

$$\sum_{j=1}^{K^*}\sum_{k=1}^{K}\frac{n_{kj}}{n}\upsilon(\theta_k,\theta^*_j)\geq\frac{C_{\theta^*}}{n}\|\underline{t}-\underline{t}^*\|_\infty.$$

Moreover, for any $j=1,\ldots,K^*$ and $\theta\in\Theta$, $\mathcal{L}_j(\theta^*_j)-\mathcal{L}_j(\theta)\geq\upsilon(\theta,\theta^*_j)$ and denoting $C_K=2C_{\theta^*}$ the result follows immediately. $\square$

6.7. Proof of Theorem 3.2

Assume with no loss of generality that $K^*=2$. Denote $(u_n)_n$ a sequence satisfying $u_n\underset{n\to\infty}{\longrightarrow}\infty$, $u_n/n\underset{n\to\infty}{\longrightarrow}0$ and $\mathbb{P}(|\widehat{\underline{t}}_n-\underline{t}^*|>u_n)\underset{n\to\infty}{\longrightarrow}0$ (for example $u_n=n\sqrt{\max(\mathbb{E}|\widehat{\tau}_n-\tau^*|,n^{-1})}$). For $\delta>0$, as we have

$$\mathbb{P}(|\widehat{\underline{t}}_n-\underline{t}^*|>\delta)\leq\mathbb{P}(\delta<|\widehat{\underline{t}}_n-\underline{t}^*|\leq u_n)+\mathbb{P}(|\widehat{\underline{t}}_n-\underline{t}^*|_m>u_n)$$

it suffices to show that $\lim\limits_{\delta\to\infty}\ \lim\limits_{n\to\infty}\mathbb{P}(\delta<|\widehat{t}_n-t^*|\leq u_n)=0$.

Denote $V_{\delta,u_n}=\{\ t\in\mathbb{Z}/\ \ \delta<|t-t^*|\leq u_n\ \}$. Then,

$$\mathbb{P}(\delta<|\widehat{t}_n-t^*|\leq u_n)\leq\mathbb{P}\Big(\min_{t\in V_{\delta,u_n}}(\widehat{I}_n(t)-\widehat{I}_n(t^*))\leq 0\Big).$$

Let $t\in V_{\delta,u_n}$ (for example $t\geq t^*$). With the notation of the proof of Theorem 3.1, we have $\widehat{L}_n(T_1^*,\widehat{\theta}_n(T_1^*))\geq\widehat{L}_n(T_1^*,\widehat{\theta}_n(T_1))$ and from (23) we obtain:

$$\frac{\widehat{I}_n(t)-\widehat{I}_n(t^*)}{t-t^*}\geq\frac{2}{t-t^*}\Big(\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))-\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_1))\\+\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))-\widehat{L}_n(T_2,\widehat{\theta}_n(T_2))\Big).$$

We conclude in two steps:

i-) We show that $\frac{1}{t-t^*}\big(\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))-\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_1))\big)>0$ for n large enough. Then $\frac{\widehat{L}_n(T_1,\theta)}{n}=\frac{t^*}{n}\frac{\widehat{L}_n(T_1^*,\theta)}{t^*}+\frac{t-t^*}{n}\frac{\widehat{L}_n(T_1\cap T_2^*,\theta)}{t-t^*}$ and since $\frac{t-t^*}{n}\leq\frac{u_n}{n}\underset{n\to\infty}{\longrightarrow}0$ and

$$\widehat{\theta}_n(T_1)=\underset{\theta\in\Theta}{\mathrm{Argmax}}\left(\frac{1}{n}\widehat{L}_n(T_1,\theta)\right)\ \underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}}\ \theta_1^*.$$

It comes that $\frac{1}{t-t^*}\big(\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))-\widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_1))\big)$ converges a.s. and uniformly on V_{δ,u_n} to $\mathcal{L}_2(\theta_2^*)-\mathcal{L}_2(\theta_1^*)>0$.

ii-) We show that $\frac{1}{t-t^*}\big(\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2,\widehat{\theta}_n(T_2))\big) \underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}} 0$. For large value of n, we remark that $\widehat{\theta}_n(T_2) \in \overset{\circ}{\Theta}$ so that $\partial \widehat{L}_n(T_2,\widehat{\theta}_n(T_2))/\partial\theta = 0$. The mean value theorem on $\partial \widehat{L}_n/\partial\theta_i$ for any $i = 1,\dots,d$ gives the existence of $\widetilde{\theta}_{n,i} \in [\widehat{\theta}_n(T_2),\widehat{\theta}_n(T_2^*)]$ such that:

$$0 \;=\; \frac{\partial \widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))}{\partial\theta_i} + \frac{\partial^2 \widehat{L}_n(T_2,\widetilde{\theta}_{n,i})}{\partial\theta\partial\theta_i}(\widehat{\theta}_n(T_2)) - \widehat{\theta}_n(T_2^*)) \quad (30)$$

where for $a,b \in \mathbb{R}^d$, $[a,b] = \{(1-\lambda)a+\lambda b;\ \lambda\in[0,1]\}$. Using the equalities $\widehat{L}_n(T_2^*,\theta) = \widehat{L}_n(T_1\cap T_2^*,\theta) + \widehat{L}_n(T_2,\theta)$ and $\partial\widehat{L}_n(T_2^*,\widehat{\theta}_n(T_2^*))/\partial\theta = 0$, it comes from (30):

$$\frac{\partial \widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{\partial\theta_i} = \frac{\partial^2 \widehat{L}_n(T_2,\widetilde{\theta}_{n,i})}{\partial\theta\partial\theta_i}(\widehat{\theta}_n(T_2))-\widehat{\theta}_n(T_2^*)), \quad \forall i = 1,\dots,d,$$

and it follows:

$$\frac{1}{t-t^*}\frac{\partial \widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{\partial\theta} = \frac{n-t}{t-t^*}A_n\cdot(\widehat{\theta}_n(T_2) - \widehat{\theta}_n(T_2^*)) \quad (31)$$

with $A_n := \big(\frac{1}{n-t}\frac{\partial^2\widehat{L}_n(T_2,\widetilde{\theta}_{n,i})}{\partial\theta\partial\theta_i}\big)_{1\le i\le d}$. Corollary 6.1 ii-) gives that:

$$\frac{1}{t-t^*}\frac{\partial \widehat{L}_n(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*))}{\partial\theta} \underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}} \frac{\partial \mathcal{L}_2(\theta_2^*)}{\partial\theta} = 0$$

and $A_n \underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}} -\frac{1}{2}\mathbb{E}\big(\frac{\partial^2 q_{0,2}(\theta_2^*)}{\partial\theta^2}\big)$. Under assumption (Var), $\mathbb{E}\big(\frac{\partial^2 q_{0,2}(\theta_2^*)}{\partial\theta^2}\big)$ is a nonsingular matrix (see [2]). Then, we deduce from (31) that

$$\frac{n-t}{t-t^*}(\widehat{\theta}_n(T_2) - \widehat{\theta}_n(T_2^*)) \underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}} 0. \quad (32)$$

We conclude by the Taylor expansion on $\widehat{L}_n$ that gives

$$\begin{aligned}&\frac{1}{t-t^*}|\widehat{L}_n(T_2,\widehat{\theta}_n(T_2)) - \widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))|\\ &\le \frac{1}{2(t-t^*)}\|\widehat{\theta}_n(T_2)) - \widehat{\theta}_n(T_2^*)\|^2 \sup_{\theta\in\Theta}\Big\|\frac{\partial^2\widehat{L}_n(T_2,\theta)}{\partial\theta^2}\Big\| \to 0 \quad \text{a.s.}\end{aligned}$$ ☐

6.8. Proof of Theorem 3.3

First, $\big(\widehat{\theta}_n(\widehat{T}_j)-\theta_j^*\big) = \big(\widehat{\theta}_n(\widehat{T}_j)-\widehat{\theta}_n(T_j^*)\big)+\big(\widehat{\theta}_n(T_j^*)-\theta_j^*\big)$ for any $j\in\{1,\dots,K^*\}$. By Theorem 3.2 it comes $\widehat{t}_j - t_j^* = o_P(\log(n))$. Using relation (32), we obtain: $\widehat{\theta}_n(\widehat{T}_j) - \widehat{\theta}_n(T_j^*) = o_P(\frac{\log(n)}{n})$. Hence, $\sqrt{n_j^*}\big(\widehat{\theta}_n(\widehat{T}_j) - \widehat{\theta}_n(T_j^*)\big) \underset{n\to\infty}{\overset{P}{\longrightarrow}} 0$ and it suffices to show that $\sqrt{n_j^*}\big(\widehat{\theta}_n(T_j^*)-\theta_j^*\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d\big(0,F(\theta_j^*)^{-1}G(\theta_j^*)F(\theta_j^*)^{-1}\big)$ to conclude.

For large value of n, $\widehat{\theta}_n(T_j^*)\in\overset{\circ}{\Theta}$. By the mean value theorem, there exists $(\widetilde{\theta}_{n,k})_{1\le k\le d}\in[\widehat{\theta}_n(T_j^*),\theta_j^*]$ such that

$$\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta_k}=\frac{\partial L_n(T_j^*,\theta_j^*)}{\partial\theta_k}+\frac{\partial^2 L_n(T_j^*,\widetilde{\theta}_{n,k})}{\partial\theta\partial\theta_k}(\widehat{\theta}_n(T_j^*)-\theta_j^*). \tag{33}$$

Let $F_n=-2\big(\frac{1}{n_j^*}\frac{\partial^2 L_n(T_j^*,\widetilde{\theta}_{n,k})}{\partial\theta\partial\theta_k}\big)_{1\le k\le d}$. By Lemma 6.3 and Corollary 6.1, $F_n \xrightarrow[n\to\infty]{\text{a.s.}} F(\theta_j^*)$ (where $F(\theta_j^*)$ is defined by (10)). But, under (Var), $F(\theta_j^*)$ is a non singular matrix (see [2]). Thus, for n large enough, F_n is invertible and (33) gives

$$\sqrt{n_j^*}\big(\widehat{\theta}_n(T_j^*)-\theta_j^*\big)=-2F_n^{-1}\Big[\frac{1}{\sqrt{n_j^*}}\Big(\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}-\frac{\partial L_n(T_j^*,\theta_j^*)}{\partial\theta}\Big)\Big].$$

As in proof of Lemma 3 of [2], it is now easy to show that:

$$\frac{1}{\sqrt{n_j^*}}\frac{\partial L_n(T_j^*,\theta_j^*)}{\partial\theta}\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}_d\big(0,G(\theta_j^*)\big)$$

where $G(\theta_j^*)$ is given by (10). Thus, since $\partial\widehat{L}_n(T_j^*,\widehat{\theta}_n(T_j^*))/\partial\theta=0$, we have:

$$\frac{1}{\sqrt{n_j^*}}\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}=\frac{1}{\sqrt{n_j^*}}\Big(\frac{\partial L_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}-\frac{\partial\widehat{L}_n(T_j^*,\widehat{\theta}_n(T_j^*))}{\partial\theta}\Big)\xrightarrow[n\to\infty]{\text{a.s.}}0.$$

We conclude using Lemma 6.3 and the fact that $1/\sqrt{n}=O(v_n/n)$. □

Acknowledgements

The authors are grateful to the referees and the associate editor for many relevant suggestions and comments which helped to improve the contents of the paper.

References

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

[2] Bardet, J.-M. and Wintenberger, O. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist. 37*, (2009), 2730–2759. MR2541445

[3] Basseville, M. and Nikiforov, I. Detection of Abrupt Changes: Theory and Applications. Prentice Hall, Englewood Cliffs, NJ, 1993. MR1210954

[4] Baudry, J.-P., Maugis, C. and Michel, B. Slope Heuristics: overview and implementation. *RR-INRIA*, (2010), n°7223.

[5] Berkes, I., Horváth, L., and Kokoszka, P. GARCH processes: structure and estimation. *Bernoulli 9* (2003), 201–227. MR1997027

[6] Davis, R. A., Huang, D. and Yao, Y.-C. Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist. 23*, (1995), 282–304. MR1331669

[7] Davis, R. A., Lee, T. C. M. and Rodriguez-Yam, G. A. Break detection for a class of nonlinear time series models. *Journal of Time Series Analysis 29*, (2008), 834–867. MR2450899

[8] Doukhan, P. and Wintenberger, O. Weakly dependent chains with infinite memory. *Stochastic Process. Appl. 118*, (2008) 1997-2013. MR2462284

[9] Duflo, M. Méthodes récursives aléatoires. *Masson, Paris*, 1990; English Edition, Springer, 1996. MR1082344

[10] Francq, C., and Zakoïan, J.-M. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli 10* (2004), 605–637. MR2076065

[11] Fryzlewicz, P., and Subba Rao, S. BaSTA: consistent multiscale multiple change-point detection for piecewise-stationary ARCH processes. *Preprint* (2010). `http://stats.lse.ac.uk/fryzlewicz/basta/basta.pdf`

[12] Hinkley, D. V. Inference about the change-point in a sequence of random variables. *Biometrika 57*, (1970), 1–17. MR0273727

[13] Jeantheau, T. Strong consistency of estimators for multivariate arch models. *Econometric Theory 14* (1998), 70–86. MR1613694

[14] Kokoszka, P. and Leipus, R. Change-point estimation in ARCH models. *Bernoulli 6*, (2000), 513–539. MR1762558

[15] Kay, S. M. Fundamentals of Statistical Signal Processing, vol. 2. *Prentice-Hall, Englewood Cliffs, NJ*, (1998)

[16] Kounias, E. G. and Weng, T.-S. An inequality and almost sure convergence. *Annals of Mathematical Statistics 40*, (1969), 1091–1093. MR0245058

[17] Lavielle, M. and Ludena, C. The multiple change-points problem for the spectral distribution. *Bernoulli 6*, (2000) 845–869. MR1791905

[18] Lavielle, M. and Moulines, E. Least squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis 21*, (2000) 33–59. MR1766173

[19] Nelson, D. B. and Cao, C. Q. Inequality Constraints in the Univariate GARCH Model. *Journal of Business & Economic Statistics 10*, (1992), 229–235.

[20] Page, E. S. A test for a change in a parameter occurring at an unknown point. *Biometrika 42*, (1955), 523–526. MR0072412

[21] Robinson, P., and Zaffaroni, P. Pseudo-maximum likelihood estimation of ARCH(∞) models. *Ann. Statist. 34* (2006), 1049–1074. MR2278351

[22] Straumann, D., and Mikosch, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist. 34* (2006), 2449–2495. MR2291507

[23] Yao, Y.C. Estimating the number of change-points via Schwarz criterion. *Statistics & Probability Letters 6*, (1988), 181–189. MR0919373