Submitted to the Annals of Statistics

DETECTING MULTIPLE CHANGE-POINTS IN GENERAL CAUSAL TIME SERIES USING PENALIZED QUASI-LIKELIHOOD

BY JEAN-MARC BARDET WILLIAM KENGNE AND OLIVIER WINTENBERGER

SAMM, Université Paris 1 Panthéon-Sorbonne, France
CEREMADE, Université Paris-Dauphine, FRANCE

Abstract This paper is devoted to the off-line multiple change-point detection in a semiparametric framework. The time series is supposed to belong to a large class of models including AR(∞), ARCH(∞), TARCH(∞),... models where the coefficients change at each instant of breaks. The different unknown parameters (number of changes, change dates and parameters of successive models) are estimated using a penalized contrast built on conditional quasi-likelihood. Under Lipshitzian conditions on the model, the consistency of the estimator is proved when the moment order r of the process satisfies $r \geq 2$. If $r \geq 4$, the same convergence rates for the estimators than in the case of independent random variables are obtained. The particular cases of AR(∞), ARCH(∞) and TARCH(∞) show that our method notably improves the existing results.

1. Introduction. The problem of the detection of change-points is a classical problem as well as in the statistic than in the signal processing community. If the first important result in this topic was obtained by Page [20] in 1955, real advances have been done in the seventies, notably with the results of Hinkley (see for instance Hinkley [13]) and the topic of change detection became a distinct and important field of the statistic since the eighties (see the book of Basseville and Nikiforov [3] for a large overview).

Two approaches are generally considered for solving a problem of change detection: an 'on-line' approach leading to sequential estimation and an 'off-line' approach which arises 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 a general class $\mathcal{M}_T(M, f)$ of causal (non-anticipative) time series. Let M and f be a measurable functions such that for all $(x_i)_{i\in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$, $M\big((x_i)_{i\in \mathbb{N}}\big)$ is a $(m \times p)$ non-zero real matrix and $f\big((x_i)_{i\in \mathbb{N}}\big) \in \mathbb{R}^m$. Let $T \subset \mathbb{Z}$ and $(\xi_t)_{t\in\mathbb{Z}}$ be a sequence of centered independent and identically distributed (iid) $\mathbb{R}^p$-random vectors called the innovations and satisfying

AMS 2000 subject classifications: Primary 62M10, 62F12

Keywords and phrases: Change detection; Causal processes; ARCH(∞) processes; AR(∞) processes; Quasi-

$\mathrm{var}(\xi_0) = I_p$ (the identity matrix of dimension p). Then, define

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

$$X_{t+1} = M\big((X_{t-i})_{i\in\mathbb{N}}\big)\xi_t + f\big((X_{t-i})_{i\in\mathbb{N}}\big) \quad \text{for all } t \in T. \tag{1.1}$$

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.

The problem of change-point detection is the following: assume that a trajectory $(X_1, \cdots, X_n)$ of $X = (X_t)_{t\in\mathbb{Z}}$ is observed where

$$X \in \mathcal{M}_{T_j^*}(M_{\theta_j^*}, f_{\theta_j^*}) \quad \text{for all } j = 1, \ldots, K^*, \qquad \text{with} \tag{1.2}$$

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

The aim in the problem is the estimation of the unknown parameters $\big(K^*, (t_j^*)_{1\leq j\leq K^*-1}, (\theta_j^*)_{1\leq j\leq K^*}\big)$. In the literature it is generally supposed that X is a stationary process on each set T_j^* and is independent on each T_j^* from the other T_k^*, $k \neq j$ (for instance in [18], [15], [6] and [7]). Here the problem (1.2) does not induce such assumption and thus the framework is closer to the applications, see Remark 1 in [7].

In the problem of change-point detection, numerous papers were devoted to the CUSUM procedure (see for instance Kokozska and Leipus [15] in the specific case of ARCH(∞) processes). In Lavielle and Ludena [17] a "Whittle" contrast is used for estimating the break 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 break points 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 model.

We consider here a semiparametric estimator based on a penalized contrast using the quasi-likelihood function. For usual stationary time series, the conditional quasi-likelihood is constructed as follow:

1. Compute the conditional likelihood (with respect to $\sigma\{X_0, X_{-1}, \ldots\}$) as if $(X_t)_{t\in\mathbb{Z}}$ is known and when the process of innovations is a Gaussian sequence;
2. Approximate this computation for a sample $(X_1, \ldots, X_n)$;
3. Apply this approximation even if the process of innovations is not a Gaussian sequence.

The quasi-maximum likelihood estimator (QMLE) obtained by maximizing the quasi-likelihood function has convincing asymptotic properties in the case of GARCH processes (see Jeantheau [14]

Berkes *et al.* [4], Franck and Zakoian [11]) or generalizations of GARCH processes (see Mikosch and Straumann [23], Robinson and Zaffaroni [22]). Bardet and Wintenberger [2] study the asymptotic normality of the QMLE of θ applied to $\mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$. Thus, when K^* is known, a natural estimator of the parameter $\big((t_j^*)_{1\leq j\leq K^*-1}, (\theta_j^*)_{1\leq j\leq K^*}\big)$ for a process satisfying (1.2) is the QMLE on every intervals $[t_j+1,\ldots,t_{j+1}]$ and every parameters θ_j for $1\leq j\leq K^*$. However we consider here that K^* is unknown and such method cannot be directly used. The solution chosen is to penalize the contrast by an additional term $\beta_n K$, where $(\beta_n)_{n\in I\!N}$ is an increasing sequence of real numbers (see the final expression of the penalized contrast in (3.2)). Such procedure of penalization was previously used for instance by Yao [24] 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 (3.3)) of the parameters $\big(K^*, (t_j^*)_{1\leq j\leq K^*-1}, (\theta_j^*)_{1\leq j\leq K^*}\big)$.
Classical heuristics such as the BIC one or the MDL one of [7] lead to choose $\beta_n \propto \log n$. In our study, such penalizations are excluded in some cases, when the models $\mathcal{M}_T(M,f)$ are very dependent of their whole past, see Remark 3.3 for more details. Finally, we will show that an "optimal" penalization is $\beta_n \propto \sqrt{n}$ which overpenalizes the number of breaks to avoid artificial breaks in cases of models very dependent of their whole past (see Remark 3.5).

The main results of the paper are the following: under Lipshitzian condition on f_θ and M_θ, 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 when the moment of order r on the innovations and X is larger than 2. If moreover Lipshitzian conditions are satisfied by the derivatives of f_θ and M_θ and if $r\geq 4$, then the convergence rate of $(\widehat{t}_j/n)_{1\leq j\leq \widehat{K}_n-1}$ is $O_P(n^{-1})$ 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 they are the same than in an independent setting.

Section 2 is devoted to the presentation of the model and the assumptions and the study of the existence of a nonstationary solution of the problem (1.2). 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. Section 5 contains the main proofs.

2. Assumptions and existence of a solution of the change process.

2.1. *Assumptions on the class of models $\mathcal{M}_{\mathbb{Z}}(f_\theta, M_\theta)$.* Let $\theta\in I\!R^d$ and M_θ and f_θ be numerical functions such that for all $(x_i)_{i\in I\!N}\in I\!R^{I\!N}$, $M_\theta\big((x_i)_{i\in I\!N}\big)\neq 0$ and $f_\theta\big((x_i)_{i\in I\!N}\big)\in I\!R$. We use the following different norms:

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

Let $\Psi_\theta=f_\theta$, M_θ and $i=0,\,1,\,2$, then for any compact set $\Theta\subseteq I\!R^d$, define

Assumption $\mathbf{A}_i(\Psi_\theta,\Theta)$: *Assume that* $\|\partial^i\Psi_\theta(0)/\partial\theta^i\|_\Theta<\infty$ *and there exists a sequence of non-negative real number* $(\alpha_i^{(k)}(\Psi_\theta,\Theta))_{i\geq 1}$ *such that* $\sum\limits_{k=1}^\infty \alpha_k^{(i)}(\Psi_\theta,\Theta)<\infty$ *satisfying*

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

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

Assumption $\mathbf{A}_i(h_\theta,\Theta)$: *Assume that* $\|\partial^i h_\theta(0)/\partial\theta^i\|_\Theta<\infty$ *and there exists a sequence of non-negative real number* $(\alpha_i^{(k)}(h_\theta,\Theta))_{i\geq 1}$ *such as* $\sum\limits_{k=1}^\infty \alpha_k^{(i)}(h_\theta,\Theta)<\infty$ *satisfying*

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

Now, for any $i=0,1,2$ and $\theta\in\Theta$, under Assumptions $A_i(f_\theta,\Theta)$ and $A_i(M_\theta,\Theta)$, denote:

$$\beta^{(i)}(\theta):=\sum_{k\geq 1}\beta_k^{(i)}(\theta)\quad\text{where}\quad \beta_k^{(i)}(\theta):=\alpha_k^{(i)}(f_\theta,\{\theta\})+(\mathbb{E}|\xi_0|^r)^{1/r}\alpha_k^{(i)}(M_\theta,\{\theta\}),\tag{2.1}$$

and under Assumption $A_i(h_\theta,\Theta)$

$$\tilde\beta^{(i)}(\theta):=\sum_{k\geq 1}\tilde\beta_k^{(i)}(\theta)\quad\text{where}\quad \tilde\beta_k^{(i)}(\theta):=(\mathbb{E}|\xi_0|^r)^{2/r}\alpha_k^{(i)}(h_\theta,\{\theta\}).\tag{2.2}$$

The dependence with respect to r of $\beta_k^{(i)}(\theta)$ and $\tilde\beta_k^{(i)}(\theta)$ are omitted for notational convenience. Then define:

$$\begin{aligned}\Theta(r):=\{\theta\in\Theta,\ &A_0(f_\theta,\{\theta\})\text{ and }A_0(M_\theta,\{\theta\})\text{ hold with }\beta^{(0)}(\theta)<1\}\\ &\cup\{\theta\in\mathbb{R}^d,\ f_\theta=0\text{ and }A_0(h_\theta,\{\theta\})\text{ holds with }\tilde\beta^{(0)}(\theta)<1\}.\end{aligned}\tag{2.3}$$

From [2] we have:

Proposition 2.1 *If* $\theta\in\Theta(r)$ *for some* $r\geq 1$, *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* $\|X_0\|_r<\infty$.

Remark 2.1 *The Lipschitz-type hypothesis* $A_i(\Psi_\theta,\Theta)$ *are classical when studying the existence of solution of general models. For instance, Duflo [9] used such a Lipschitz-type inequality to show the existence of Markov chains. The subset* $\Theta(r)$ *is defined as a reunion to consider accurately general causal models and ARCH-type models simultaneously: for ARCH-type models* $A_0(h_\theta,\{\theta\})$ *is less restrictive than* $A_0(M_\theta,\{\theta\})$. *However, remark that* $A_0(h_\theta,\{\theta\})$ *is still not optimal for ensuring the existence of a stationary solution for ARCH-type models.*

Let $\theta\in\Theta(r)$ 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 $x \in I\!\!R^{I\!\!N}$.

Assumption Id(Θ): For all $\theta, \theta' \in \Theta^2$,

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

Assumption Var(Θ): For all $\theta \in \Theta$,

$$\Big(\frac{\partial f_\theta}{\partial \theta^{(i)}}(X_0, X_{-1}, \cdots)\Big)_{1\leq i\leq d} \neq 0 \quad \text{or} \quad \Big(\frac{\partial h_\theta}{\partial \theta^{(i)}}(X_0, X_{-1}, \cdots)\Big)_{1\leq i\leq d} \neq 0 \quad a.s.$$

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 problem* (1.2). Consider the problem (1.2) and let $(X_1, \ldots, X_n)$ be an observed path of X. 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 $t^*_1, \cdots, t^*_{k^*-1}$ of breaks and parameters $\theta^*_1, \cdots, \theta^*_{K^*}$ are unknown. Consider first the following notation.

Notation.

- *For* $K \geq 2$, $\mathcal{F}_K = \big\{\underline{t} = (t_1, \ldots t_{K-1})\ ;\ 0 < t_1 < \ldots < t_{K-1} < n\big\}$. *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* $K \in I\!\!N^*$ *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 nonstationary solution of the problem (1.2) and its moments properties.

Proposition 2.2 *Consider the problem* (1.2)*. Assume there exists* $r \geq 1$ *such that* $\theta^*_j \in \Theta(r)$ *for all* $j = 1, \ldots, K^*$*. Then*

(i) there exists a process $X = (X_t)_{t\in\mathbb{Z}}$ *solution of the model* (1.2) *such as* $\|X_t\|_r < \infty$ *for* $t \in \mathbb{Z}$ *and* X *is a causal time series.*

(ii) there exists a constant $C > 0$ *such that for all* $t \in \mathbb{Z}$ *we have* $\|X_t\|_r \leq C$*.*

Remark 2.2 *The problem* (1.2) *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* $t = 0$*) or, for instance, if we set* $X_t = 0$ *for* $t \leq 0$*.*

3. Asymptotic results of the estimation procedure.

3.1. *The estimation procedure.* 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

$T=\{t,t+1,\dots,t'\}$, the distribution of $X_s \mid (X_{s-j})_{j\in\mathbb{N}^*}$ is $\mathcal{N}\big(f_\theta\big(X_{s-1},\dots\big),h_\theta\big(X_{s-1},\dots\big)\big)$. Therefore, with the notation $f^s_\theta=f_\theta\big(X_{s-1},X_{s-2}\dots\big)$, $M^s_\theta=M_\theta\big(X_{s-1},X_{s-2}\dots\big)$ and $h^s_\theta={M^s_\theta}^2$, we deduce the conditional log-likelihood on T (up to an additional constant)

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

By convention, we set $L_n(\emptyset,\theta_k):=0$. Since only $X_1,\dots,X_n$ 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:

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

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

Remark 3.1 *For convenience, in the sequel we chose $u=(u_n)_{n\in\mathbb{N}}$ with $u_n=0$ for all $n\in\mathbb{N}$ as in [11] or in [2]. Indeed, this choice has no effect on the asymptotic behavior of estimators.*

Now, even if the process $(\xi_t)_t$ is non-Gaussian and for any number of breaks $K-1\geq 1$ and any $\underline{t}\in\mathcal{F}_K$, $\underline{\theta}\in\Theta(r)^K$, define the contrast function $\widehat{J}_n$ by the expression:

$$\widehat{J}_n(K,\underline{t},\underline{\theta}):=-2\sum_{k=1}^{K}\widehat{L}_n(T_k,\theta_k)=-2\sum_{k=1}^{K}\sum_{j=1}^{K^*}\widehat{L}_n(T_k\cap T^*_j,\theta_k),\tag{3.1}$$

Finally, let $(v_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be sequences satisfying $v_n\geq 1$ and $\beta_n:=n/v_n\to\infty$ $(n\to\infty)$. Let $K_{\max}\in\mathbb{N}^*$ and for $K\in\{1,\cdots,K_{\max}\}$ and $(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta(r)^K$ ($\Theta(r)$ is supposed to be a compact set) define the penalized contrast $\widetilde{J}_n$ by

$$\widetilde{J}_n(K,\underline{t},\underline{\theta}):=\widehat{J}_n(K,\underline{t},\underline{\theta})+\frac{n}{v_n}K=\widehat{J}_n(K,\underline{t},\underline{\theta})+\beta_n K\tag{3.2}$$

and the penalized contrast estimator $(\widehat{K}_n,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n)$ of $(K^*,\underline{t}^*,\underline{\theta}^*)$ as

$$(\widehat{K}_n,\widehat{\underline{t}}_n,\widehat{\underline{\theta}}_n)=\underset{1\leq K\leq K_{\max}}{\text{Argmin}}\ \underset{(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta(r)^K}{\text{Argmin}}\big(\widetilde{J}_n(K,\underline{t},\underline{\theta})\big)\quad\text{and}\quad\widehat{\underline{\tau}}_n=\frac{\widehat{\underline{t}}_n}{n}.\tag{3.3}$$

For $K\geq 1$ and $\underline{t}\in\mathcal{F}_K$, denote $\widehat{\underline{\theta}}_n(\underline{t})=(\widehat{\theta}(T_1),\dots,\widehat{\theta}(T_K)):=\underset{\underline{\theta}\in\Theta(r)^K}{\text{Argmin}}\Big(\widetilde{J}_n(\underline{t},\underline{\theta})\Big)=\underset{\underline{\theta}\in\Theta(r)^K}{\text{Argmin}}\Big(\widehat{J}_n(\underline{t},\underline{\theta})\Big)$. Then, $\widehat{\theta}_n(T_k)$ is the QMLE of θ^*_k computed on T_k and $\widehat{\underline{\theta}}_n(\underline{t}^*)$ is the $QMLE$ of $\underline{\theta}^*$.

Remark 3.2 *If K^* is known, the estimator of $(\underline{\tau}^*,\underline{\theta}^*)$ may be obtained by minimizing $\widehat{J}_n$ instead of $\widetilde{J}_n$. However the knowledge of K^* does not improve the asymptotic results established in this paper.*

3.2. *Consistency of* $(\widehat{K}_n, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n)$. For establishing the consistency, we add the couple of following classical assumptions in the problem of break detection:

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

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

Hypothesis C: *there exists* $\tau^*_1, \dots, \tau^*_{K^*-1}$ *with* $0 < \tau^*_1 < \dots < \tau^*_{K^*-1} < 1$ *such that for* $j = 1, \cdots, K^*$, $t^*_j = [n\tau^*_j]$ *(where* $[x]$ *is the floor of* x*). The vector* $\underline{\tau}^* = \big(\tau^*_1, \dots, \tau^*_{K^*-1}\big)$ *is called the vector of breaks.*

Even if the length of T^*_j has asymptotically the same order than n, the dependences with respect to n of t^*_j, t_k, T^*_j and T_k are omitted for notational convenience.
Finally we make a technical non classical assumption. Using the convention: if $A_i(M_\theta, \Theta)$ holds then $\alpha^{(i)}_\ell(h_\theta, \Theta) = 0$ and if $A_i(h_\theta, \Theta)$ holds then $\alpha^{(i)}_\ell(M_\theta, \Theta) = 0$, define:

Hypothesis $\mathbf{H}_i$ ($i = 0, 1, 2$): *For* $0 \le p \le i$, *the assumptions* $A_p(f_\theta, \Theta)$, $A_p(M_\theta, \Theta)$ *(or* $A_p(h_\theta, \Theta)$*) hold and for all* $j = 1, \dots, K^*$ *there exists* $r \ge 1$ *such that* $\theta^*_j \in \Theta(r)$. *Denoting*

$$c^* = \min_{j=1,\cdots,K^*} \big(-\log(\beta^{(0)}(\theta^*_j))/8 \big) \wedge \min_{j=1,\cdots,K^*} \big(-\log(\tilde{\beta}^{(0)}(\theta^*_j))/8 \big)$$

the sequence $(v_n)_{n \in I\!N}$ *used in* (3.2) *satisfies for all* $j = 1, \cdots K^*$*:*

$$\sum_{k\ge1} \Big(\frac{v_k}{k}\Big)^{r/4\wedge1} \Big(\sum_{\ell \ge kc^*/\log(k)} \beta^{(0)}_\ell(\theta^*_j) \Big)^{r/4} \wedge \sum_{k\ge1} \Big(\frac{v_k}{k}\Big)^{r/4\wedge1} \Big(\sum_{\ell \ge kc^*/\log(k)} \tilde{\beta}^{(0)}_\ell(\theta^*_j) \Big)^{r/4} < \infty \quad \textit{and}$$
$$\sum_{k\ge1} \Big(\frac{v_k}{k}\Big)^{r/4\wedge1} \Big(\sum_{\ell \ge k/2} \big(\alpha^{(p)}_\ell(f_\theta, \Theta(r)) + \alpha^{(p)}_\ell(M_\theta, \Theta(r)) + \alpha^{(p)}_\ell(h_\theta, \Theta(r))\big) \Big)^{r/4} < \infty. \tag{3.4}$$

The assumption H_i is interesting as it links the decrease rate of the Lipschitz coefficients and the penalization term of (3.2). The classical BIC penalization and the one coming from the MDL approach (see [7]) correspond to a sequence $v_n \propto n/\log(n)$. 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 Remark below). However, if the decrease of the Lipschitz coefficients is slower, our method can exclude such a choice and an heavier term $\beta_n = n/v_n >> \log(n)$ in the penalization has to be chosen.

Remark 3.3 *Conditions* (3.4) *satisfied by* $(v_n)_n$ *are deduced from a result of Kounias [16]. The conditions on* $(v_n)_n$ *are not too restrictive:*

(1) *geometric case: if* $\alpha^{(i)}_\ell(f_\theta, \Theta(r)) + \alpha^{(i)}_\ell(M_\theta, \Theta(r)) + \alpha^{(i)}_\ell(h_\theta, \Theta(r)) = O(a^\ell)$ *with* $0 \le a < 1$, *then any* $(v_n)_n$ *such as* $v_n = o(n)$ *can be chosen (for instance* $v_n = n(\log n)^{-1}$*).*

(2) *Riemanian case: if* $\alpha^{(i)}_\ell(f_\theta, \Theta(r)) + \alpha^{(i)}_\ell(M_\theta, \Theta(r)) + \alpha^{(i)}_\ell(h_\theta, \Theta(r)) = O(\ell^{-\gamma})$ *with* $\gamma > 1$

- *if* $\gamma > 1 + (1 \vee 4r^{-1})$, *then all sequence* $(v_n)_n$ *such as* $v_n = o(n)$ *can be chosen (for instance* $v_n = n(\log n)^{-1}$*).*
- *if* $1 \vee 4r^{-1} < \gamma \leq 1 + (1 \vee 4r^{-1})$, *then any* $(v_n)_n$ *such as* $v_n = O(n^{\gamma-(1\vee 4r^{-1})}(\log n)^{-\delta})$ *with* $\delta > 1 \vee 4r^{-1}$ *can be chosen.*

We are now ready to prove the consistency of the penalized QMLE:

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

$$(\widehat{K}_n, \underline{\widehat{\tau}}_n, \underline{\widehat{\theta}}_n) \xrightarrow[n\to\infty]{\mathcal{P}} (K^*, \underline{\tau}^*, \underline{\theta}^*). \tag{3.5}$$

Remark 3.4 *If* K^* *is known, we can relax the assumptions for the consistency by taking* $v_n = 1$ *for all* n *as the penalization term in* (3.2) *does not matter. If* K^* *is unknown then a reasonable choice in any geometric or Riemanian cases is* $v_n \propto \log n$ *(therefore* $\beta_n \propto n(\log n)^{-1}$*), see Remark 3.3.*

3.3. *Rate of convergence of the estimators.* To state a rate of convergence of the estimators $\underline{\widehat{\tau}}_n$ and $\underline{\widehat{\theta}}_n$, we need to work under stronger moment and regularity assumptions.

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

$$\lim_{\delta\to\infty} \lim_{n\to\infty} \mathbb{P}(\|\underline{\widehat{t}}_n - \underline{t}^*\|_m > \delta) = 0. \tag{3.6}$$

This theorem induces that $w_n^{-1} \|\underline{\widehat{t}}_n - \underline{t}^*\|_m \xrightarrow{P} 0$ for any sequence $(w_n)_n$ such as $w_n \to \infty$ 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 for mixing processes in [18].

Let us turn now 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 the hypothesis* $D(\Theta(r))$, $Id(\Theta(r))$, B, C *and* H_2 *are satisfied with* $r \geq 4$ *and* $\sqrt{n} = O(v_n)$. *Then if* $\theta_j^* \in \overset{\circ}{\Theta}(r)$ *for all* $j = 1, \cdots, K^*$, *we have*

$$\sqrt{n_j^*}\,\big(\widehat{\theta}_n(\widehat{T}_j) - \theta_j^*\big) \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}_d\big(0, F(\theta_j^*)^{-1} G(\theta_j^*) F(\theta_j^*)^{-1}\big), \tag{3.7}$$

where, using $q_{0,j}$ *defined in* (5.2), *the matrix* F *and* G *are such as*

$$(F(\theta_j^*))_{k,l} = I\!E\Big(\frac{\partial^2 q_{0,j}(\theta_j^*)}{\partial\theta_k \partial\theta_l}\Big) \text{ and } (G(\theta_j^*))_{k,l} = I\!E\Big(\frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta_k}\frac{\partial q_{0,j}(\theta_j^*)}{\partial\theta_l}\Big). \tag{3.8}$$

Remark 3.5 *In Theorem 3.3, a condition on the rate of convergence of v_n is added. The optimal choice for the penalization term corresponds to $v_n \propto \sqrt{n}$ as it corresponds to the most general problem* (1.2)*, see Remark 3.3. However, by assumption H_2 it excludes models with finite moments $r \geq 4$ satisfying: $\ell^{-\gamma} = O(\alpha_\ell^{(i)}(f_\theta, \Theta(r)) + \alpha_\ell^{(i)}(M_\theta, \Theta(r)) + \alpha_\ell^{(i)}(h_\theta, \Theta(r)))$ with $1 < \gamma \leq 3/2$ for some $i = 0, 1, 2$. For these models the consistency and the rate of convergence of order n for $\widehat{\underline{\tau}}_n$ hold 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:

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

It correponds to the problem (1.2) with models $\mathcal{M}_{T_i^*}(f_\theta, M_\theta)$ where $f_\theta(x_1, \cdots) = \sum_{k \geq 1} \phi_k(\theta) x_k$ and $M_\theta \equiv 1$. Assume that Θ is a compact set such that $\sum_{k \geq 1} \|\phi_k(\theta)\|_\Theta < 1$. Thus $\Theta(r) = \Theta$ for any $r \geq 1$ satisfying $\mathbb{E}|\xi_0|^r < \infty$. Then Assumptions D($\Theta$) and $A_0(f_\theta, \Theta)$ hold automatically with $\underline{h} = 1$ and $\alpha_k^{(0)}(f_\theta, \Theta(r)) = \|\phi_k(\theta)\|_\Theta$. Then,

- Assume that Id(Θ) holds and that there exists $r \geq 2$ 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 $k \geq 1$, then the penalization $v_n = \log n$ (or $\beta_n = n/\log n$) ensures the consistency of $(\widehat{K}_n, \widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n)$.
- Moreover, if $r \geq 4$, $\gamma > 3/2$ and ϕ_k twice differentiable satisfying $\|\phi_k'(\theta)\|_\Theta = O(k^{-\gamma})$ and $\|\phi_k''(\theta)\|_\Theta = O(k^{-\gamma})$, then the penalization $v_n = \beta_n = \sqrt{n}$ ensures the convergence (3.6) of $\widehat{\underline{t}}_n$ and the CLT (3.7) 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] under moments of order greater than 4 is 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:

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

where for any $\theta \in \Theta$, $\psi_0(\theta) > 0$ and $(\psi_k(\theta))_{k \geq 1}$ is a sequence of positive real number 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 $\sum_{k \geq 1} \|\psi_k(\theta)\|_\Theta < 1$, 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 $k \geq 1$, then the penalization $v_n = \log n$ (or $\beta_n = n/\log n$) 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 $r \geq 4$ and ψ_k is twice differentiable satisfying $\|\psi'_k(\theta)\|_\Theta = O(k^{-\gamma})$ and $\|\psi''_k(\theta)\|_\Theta = O(k^{-\gamma})$ with $\gamma > 3/2$, if $\Theta(4)$ is a compact such that $\theta^*_j \in \overset{\circ}{\Theta}(4)$ for all j, then the penalization $v_n = \beta_n = \sqrt{n}$ as in Remark 3.3 ensures the convergence (3.6) of $\widehat{\underline{t}}_n$ and the CLT (3.7) satisfied by $\widehat{\underline{\theta}}_n(\widehat{T}_j)$ for all j.

This problem of break detection was already studied by Kokoszka and Leipus in [15] 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:*

$$X_t = \sigma_t\,\xi_t\ ,\ \sigma_t^2 = a^*_{0,j} + \sum_{k=1}^{q} a^*_{k,j} X^2_{t-k} + \sum_{k=1}^{p} b^*_{k,j}\sigma^2_{t-k} \quad t^*_{j-1} < t \leq t^*_j\ ,\ j = 1, \cdots, K^*$$

with $\mathbb{E}(\xi_0^2) = 1$. Assume that for any $\theta = (a_0, \ldots, a_q, b_1, \ldots, b_p) \in \Theta$ then $a_k \geq 0$, $b_k \geq 0$ 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 $\sigma_t^2 = \psi_0(\theta) + \sum_{k\geq 1} \psi_k(\theta) X^2_{t-k}$. Remark that this sequence is twice differentiable with respect to θ and that its derivatives are exponentially decreasing. Moreover for any $\theta \in \Theta$ it holds $\sum_{k\geq1}\psi_k(\theta) \leq \big(\sum_{k=1}^q a_k\big)/\big(1-\sum_{k=1}^p b_k\big)$ and one can consider:

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

Then if $\sum_{k=1}^q a^*_{k,j} + \sum_{k=1}^p b^*_{k,j} < 1$ for all j (case $r \geq 2$), our estimation procedure associated with a penalization term $\beta_n K$ for any $1 << \beta_n << n$ 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 penalization $1 << \beta_n = 0(\sqrt{n})$ allows the same rates of convergence than in the case where (X_t) are independent r.v. For example, a penalization $\beta_n \propto \log n$ as in [7] can be chosen in this case.

4.3. *Estimates breaks in TARCH(∞) model.* Consider a TARCH(∞) model with breaks defined by:

$$X_t = \sigma_t\,\xi_t\ ,\ \sigma_t = b_0(\theta^*_j) + \sum_{k\geq1}\Big(b^+_k(\theta^*_j)\max(X_{t-k},0) - b^-_k(\theta^*_j)\min(X_{t-k},0)\Big),$$

for any $t^*_{j-1} < t \leq t^*_j$, $j = 1, \cdots, K^*$ and where $\sum_{k\geq1}\max(\|b^+_k(\theta)\|_\Theta, \|b^-_k(\theta)\|_\Theta) < \infty$. Then $f_\theta = 0$ and $(A_0(M_\theta, \Theta))$ holds with $\alpha^{(0)}_k(M_\theta,\Theta) = \max(\|b^+_k(\theta)\|_\Theta, \|b^-_k(\theta)\|_\Theta)$.

- Assume $\|\xi_0\|_r \sum_{k\geq1}\max(\|b^+_k(\theta)\|_\Theta, \|b^-_k(\theta)\|_\Theta) < 1$ for $r \geq 2$. If there exists $\gamma > 1 \vee 4r^{-1}$ such as $\max(\|b^+_k(\theta)\|_\Theta, \|b^-_k(\theta)\|_\Theta) = O(k^{-\gamma})$ for all $k \geq 1$, then a penalization $v_n = \log n$ (or $\beta_n = n/\log n$) leads to the consistency of $(\widehat{K}_n, \widehat{\underline{\tau}}_n, \widehat{\underline{\theta}}_n)$ when $\theta^*_j \in \Theta(2)$ for all j.
- Moreover, if $r \geq 4$ 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 $v_n = \beta_n = \sqrt{n}$ ensures the convergence (3.6) of $\widehat{\underline{t}}_n$ and the CLT (3.7) satisfied by $\widehat{\underline{\theta}}_n(\widehat{T}_j)$ for all j (with $\theta^*_j \in \overset{\circ}{\Theta}(4)$).

5. Proofs of the main results. In the sequel C denotes a positive constant whom value may differ from one inequality to another.

5.1. *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:

$$X_t := M_{\theta_j^*}(X_{t-1}, X_{t-2}, \cdots)\xi_t + f_{\theta_j^*}(X_{t-1}, X_{t-2}, \cdots),\ \forall t \in T_j^*;\ j = 2, \cdots K^*. \tag{5.1}$$

Thus, X_t is independent of $(\xi_j)_{j>t}$ 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 $Z_{t,j} := M_{\theta_j^*}(0,0,\cdots)\xi_t + f_{\theta_j^*}(0,0,\cdots)$ for all $j = 1, \ldots, K^*$. Assume that there exists $C_{r,t} > 0$ such that $C_{r,t} = \sup_{i<t} \|X_i\|_r$ and let $t \in T_j^*$, then

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

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

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

Then, we have:

$$\|M_{\theta_j^*}(X_{t-1}, \cdots) - M_{\theta_j^*}(0,0,\cdots)\|_r \leq \sum_{i=1}^{\infty} \alpha_i^{(0)}(M_{\theta_j^*}, \theta_j^*)\|X_{t-i}\|_r \leq C_{r,t} \sum_{i=1}^{\infty} \alpha_i^{(0)}(M_{\theta_j^*}, \theta_j^*),$$

$$\|f_{\theta_j^*}(X_{t-1}, \cdots) - f_{\theta_j^*}(0,0,\cdots)\|_r \leq \sum_{i=1}^{\infty} \alpha_i^{(0)}(f_{\theta_j^*}, \theta_j^*)\|X_{t-i}\|_r \leq C_{r,t} \sum_{i=1}^{\infty} \alpha_i^{(0)}(f_{\theta_j^*}, \theta_j^*).$$

We deduce that

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

Thus, $\|X_t\|^r < \infty$, $C_{r,t+1} < \infty$ and $\|X_t\|_r \leq \|Z_{t,j}\|_r + C_{r,t+1}\beta^{(0)}(\theta_j^*)$ since $C_{r,t} \leq C_{r,t+1}$. Similarly for any $i < t$, we have $C_{r,i} \leq C_{r,t+1}$ and $\|X_i\|_r \leq \max_{1\leq j\leq K^*}\left\{\|Z_{t,j}\|_r + C_{r,t+1}\beta^{(0)}(\theta_j^*)\right\}$. Thus, by definition of $C_{r,t+1} = \sup_{i\leq t}\|X_t\|_r$ we obtain

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

and the Proposition 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 starting from the inequality

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

Finally we obtain the desired result with

$$C = \max_{1\leq j\leq K^*} \frac{\|M_{\theta_j^*}(0,0,\cdots)\ \xi_0 + f_{\theta_j^*}(0,0,\cdots)\|_r}{1-\beta^{(0)}(\theta_j^*)} \wedge \max_{1\leq j\leq K^*} \frac{\|M_{\theta_j^*}(0,0,\cdots)\ \xi_0\|_r}{\sqrt{1-\tilde\beta^{(0)}(\theta_j^*)}}. \qquad \blacksquare$$

5.2. *Some preliminary result.* The following technical lemma is useful in the sequel:

Lemma 5.1 *Suppose that $\theta_j^* \in \Theta(r)$ for $j = 1,\ldots,K^*$ with $r \geq 2$ and under the assumptions $A_0(f_\theta,\Theta)$, $A_0(M_\theta,\Theta)$ (or $A_0(h_\theta,\Theta)$) and $D(\Theta(r))$, then there exists $C > 0$ such that*

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

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

$$\begin{aligned}\big\|f_\theta^t\big\|^2_{\Theta(r)} &\leq 2\Big(\big\|f_\theta^t - f_\theta(0,\ldots)\big\|^2_{\Theta(r)} + \big\|f_\theta(0,\ldots)\big\|^2_{\Theta(r)}\Big)\\ &\leq 2\Big(\Big(\sum_{i\geq 1}\alpha_i^{(0)}(f_\theta,\Theta(r))\Big)\cdot\sum_{i\geq 1}\alpha_i^{(0)}(f_\theta,\Theta(r))|X_{t-i}|^2 + \big\|f_\theta(0,\ldots)\big\|^2_{\Theta(r)}\Big),\end{aligned}$$

therefore

$$\mathbb{E}\big\|f_\theta^t\big\|^2_{\Theta(r)} \leq 2\Big(C\Big(\sum_{i\geq 1}\alpha_i^{(0)}(f_\theta,\Theta(r))\Big)^2 + \big\|f_\theta(0,\ldots)\big\|^2_{\Theta(r)}\Big).$$

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

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

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

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

5.3. *Comparison with stationary solutions.* In the following, we assume that $\theta_j^* \in \Theta(r)$ for all $j = 1,\ldots,K^*$ with $r \geq 1$. It comes from [2] that the equation

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

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

Lemma 5.2 *Assume that the assumptions $A_0(f_\theta,\Theta)$, $A_0(M_\theta,\Theta)$ (or $A_0(h_\theta,\Theta)$) hold and*

1. $X_t = X_{t,1}$ *for all* $t \leq t_1^*$;
2. *There exists* $C > 0$ *such that for any* $j \in \{2, \cdots, K^*\}$, *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 $j \in \{2, \cdots, K^*\}$, 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 $\|X_t^2 - X_{t,j}^2\|_{r/2} \leq \|X_t\|_r + \|X_{t,j}\|_r \leq C + \max_{1\leq j\leq K^*}\|X_{0,j}\|_r \leq C_r$ for all $j = 1, \ldots, K^*$ 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 $u_\ell \leq au_{\ell-1} + b$. Considering $w_0 = u_0$ and $w_l = aw_{\ell-1} + b$, 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 $1 \leq p \leq t - t_{j-1}^*$

$$\|X_t^2 - X_{t,j}^2\|_{r/2} \leq C_r\|X_t - X_{t,j}\|_r \leq C_r u_{[(t-t_{j-1}^*)/p]} \leq C\big(\beta^{(0)}(\theta_j^*)^{(t-t_{j-1}^*)/p} + \sum_{i\geq p}\beta_i^{(0)}(\theta_j^*)\big)$$

and Lemma 5.2 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

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

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

$$\tilde{u}_\ell \le \tilde{\beta}^{(0)}(\theta_j^*)\tilde{u}_{\ell-1} + C_r^2\sum_{k>p}\tilde{\beta}_k^{(0)}(\theta_j^*)$$

for $\tilde{u}_\ell=\sup_{t_{j-1}^*+\ell p\le i\le t_j^*}\|X_i^2-X_{i,j}^2\|_{r/2}$ and Lemma 5.2 is proved. ■

5.4. *The asymptotic behavior of the likelihood.* For the process $\big(X_{t,j}\big)_{t\in T_j^*,\,j=1,\ldots,K^*}$, for any $j\in\{1,\cdots,K^*\}$ and $s\in T_j^*$ denote:

$$q_{s,j}(\theta) := \frac{\big(X_{s,j}-f_\theta^{s,j}\big)^2}{h_\theta^{s,j}} + \log\big(h_\theta^{s,j}\big) \tag{5.2}$$

with $f_\theta^{s,j}:=f_\theta(X_{s-1,j},X_{s-2,j},\ldots)$, $h_\theta^{s,j}:=(M_\theta^{s,j})^2$ where $M_\theta^{s,j}:=M_\theta(X_{s-1,j},X_{s-2,j},\ldots)$. 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 5.3 *Assume that the hypothesis $D(\Theta(r))$ holds.*

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

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

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

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

Proof 1-) For any $\theta\in\Theta(r)$, $\Big|\frac{1}{n_j^*}L_n\big(T_j^*,\theta\big)-\frac{1}{n_j^*}L_{n,j}\big(T_j^*,\theta\big)\Big| \le \frac{1}{n_j^*}\sum_{k=1}^{n_j^*}|q_{t_{j-1}^*+k}(\theta)-q_{t_{j-1}^*+k,j}(\theta)|$. Then:

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

By Corollary 1 of Kounias [16], with $r\le 4$ and no loss of generality, it is sufficient that

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

For any $\theta\in\Theta(r)$, we have:

$$\begin{aligned}|q_s(\theta)-q_{s,j}(\theta)| \le{}& \frac{1}{\underline{h}^2}|X_s-f_\theta^s|^2|h_\theta^s-h_\theta^{s,j}| \\ &+\frac{1}{\underline{h}}\big(|X_s^2-X_{s,j}^2|+|f_\theta^s-f_\theta^{s,j}||f_\theta^s+f_\theta^{s,j}+2X_s|+2|f_\theta^{s,j}||X_s-X_{s,j}|+|h_\theta^s-h_\theta^{s,j}|\big).\end{aligned} \tag{5.3}$$

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

$$\|q_s(\theta) - q_{s,j}(\theta)\|_{\Theta(r)} \le C\big(1 + |X_{s,j}| + |X_s|^2 + \|f_\theta^{s,j}\|_{\Theta(r)} + \|f_\theta^s\|^2_{\Theta(r)}\big) \times \big(|X_s - X_{s,j}| + \|f_\theta^s - f_\theta^{s,j}\|_{\Theta(r)} + \|h_\theta^s - h_\theta^{s,j}\|_{\Theta(r)}\big),$$

and by Cauchy-Schwartz Inequality,

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

Using Proposition (2.2) and the argument of the proof of Lemma (5.1) we claim that $\mathbb{E}|X_s|^r \le C$, $\mathbb{E}\|f_\theta^s\|^r_{\Theta(r)} \le C$ and that $\mathbb{E}\|f_\theta^{s,j}\|^r_{\Theta(r)} \le C$. Thus:

$$(\mathbb{E}\|q_s(\theta) - q_{s,j}(\theta)\|^{r/4}_{\Theta(r)})^2 \le C\big(\mathbb{E}|X_s - X_{s,j}|^{r/2} + \mathbb{E}\|f_\theta^s - f_\theta^{s,j}\|^{r/2}_{\Theta(r)} + \mathbb{E}\|h_\theta^s - h_\theta^{s,j}\|^{r/2}_{\Theta(r)}\big). \tag{5.4}$$

Since $r/2 \ge 1$, we will use the $L^{r/2}$ norm. By Lemma 5.2:

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

$$\Longrightarrow \quad \mathbb{E}|X_s - X_{s,j}|^{r/2} \le C\Big(\inf_{1\le p\le k}\big\{\beta^{(0)}(\theta_j^*)^{k/p} + \sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\big\}\Big)^{r/2}. \tag{5.5}$$

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

$$\|\|h_\theta^s - h_\theta^{s,j}\|_{\Theta(r)}\|_{r/2} \le C\sum_{i\ge1}\alpha_i^{(0)}(M_\theta, \Theta(r))\|X_{s-i} - X_{s-i,j}\|_r. \tag{5.6}$$

From (5.6) we obtain:

$$\|\|h_\theta^s - h_\theta^{s,j}\|_{\Theta(r)}\|_{r/2} \le C\Big(\sum_{i=1}^{k/2-1}\alpha_i^{(0)}(M_\theta, \Theta(r))\|X_{s-i} - X_{s-i,j}\|_r + \sum_{i\ge k/2}\alpha_i^{(0)}(M_\theta, \Theta(r))\|X_{s-i} - X_{s-i,j}\|_r\Big).$$

For all $s \ge t^*_{j-1}$ and $1 \le i \le k/2-1$, then $s - i > t^*_{j-1}$, $s - i > k/2$ and by Lemma 5.2:

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

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

$$\mathbb{E}\|h_\theta^s - h_\theta^{s,j}\|^{r/2}_{\Theta(r)} \le C\Big(\inf_{1\le p\le k/2}\big\{\beta^{(0)}(\theta_j^*)^{k/(2p)} + \sum_{i\ge p}\beta_i^{(0)}(\theta_j^*)\big\} + \sum_{i\ge k/2}\alpha_i^{(0)}(M_\theta, \Theta(r))\Big)^{r/2}. \tag{5.7}$$

Similarly, we obtain:

$$(5.8)\quad I\!\!E\|f^s_\theta - f^{s,j}_\theta\|^{r/2}_{\Theta(r)} \leq C\Big(\inf_{1\leq p\leq k/2}\big\{\beta^{(0)}(\theta^*_j)^{k/(2p)} + \sum_{i\geq p}\beta^{(0)}_i(\theta^*_j)\big\} + \sum_{i\geq k/2}\alpha^{(0)}_i(f_\theta,\Theta(r))\Big)^{r/2}.$$

Relations (5.4), (5.5), (5.7) et (5.8) give (the same inequality holds with h_θ replaced by M_θ):

$$(5.9)\quad I\!\!E\ \|q_s(\theta) - q_{s,j}(\theta)\|^{r/4}_{\Theta(r)} \leq C\Big[\Big(\inf_{1\leq p\leq k/2}\big\{\beta^{(0)}(\theta^*_j)^{k/(2p)} + \sum_{i\geq p}\beta^{(0)}_i(\theta^*_j)\big\}\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(0)}_i(f_\theta,\Theta(r))\Big)^{r/4} + \Big(\sum_{i\geq k/2}\alpha^{(0)}_i(M_\theta,\Theta(r))\Big)^{r/4}\Big].$$

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

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

Choosing $p = u_k$ in (5.9) we obtain:

$$\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4} I\!\!E\big(\|q_{t^*_{j-1}+k}(\theta) - q_{t^*_{j-1}+k,j}(\theta)\|^{r/4}_{\Theta(r)}\big) \leq \sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\big(\beta^{(r)}(\theta^*_j)\big)^{rk/8u_k}$$
$$+\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\Big(\sum_{i\geq u_k}\beta^{(0)}_i(\theta^*_j)\Big)^{r/4}+\sum_{k\geq 1}\big(\frac{v_k}{k}\big)^{r/4}\Big(\sum_{i\geq k/2}\big(\alpha^{(0)}_i(f_\theta,\Theta(r))+\alpha^{(0)}_i(M_\theta,\Theta(r))\big)\Big)^{r/4}.$$

This bound is finite by assumption and the result follows by using Corollary 1 of [16] .

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 (5.3) has the simplified form:

$$|q_s(\theta) - q_{s,j}(\theta)| \leq \frac{1}{\underline{h}^2}X^2_s|h^s_\theta - h^{s,j}_\theta| + \frac{1}{\underline{h}}|X^2_s - X^2_{s,j}| + \frac{1}{\underline{h}}|h^s_\theta - h^{s,j}_\theta|.$$

Then

$$(I\!\!E\|q_s(\theta) - q_{s,j}(\theta)\|^{r/4}_{\Theta(r)})^2 \leq C I\!\!E\big[\big(|X^2_s - X^2_{s,j}| + \|h^s_\theta - h^{s,j}_\theta\|_{\Theta(r)}\big)^{r/2}\big].$$

As $\|\|h^s_\theta - h^{s,j}_\theta\|_{\Theta(r)}\|_{r/2} \leq C\sum_{i\geq 1}\alpha^{(0)}_i(h_\theta,\Theta(r))\|X^2_{s-i} - X^2_{s-i,j}\|_{r/2}$ we derive from Lemma 5.2,

$$I\!\!E\ \|q_s(\theta)-q_{s,j}(\theta)\|^{r/4}_{\Theta(r)} \leq C\Big[\Big(\inf_{1\leq p\leq k/2}\big\{\tilde\beta^{(0)}(\theta^*_j)^{k/(2p)}+\sum_{i\geq p}\tilde\beta^{(0)}_i(\theta^*_j)\big\}\Big)^{r/4}+\Big(\sum_{i\geq k/2}\alpha^{(0)}_i(h_\theta,\Theta(r))\Big)^{r/4}\Big].$$

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

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

By Corollary 1 of Kounias (1969), when $r \leq 4$ with no loss of generality, it suffices to show

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

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

$$\frac{\partial q_s(\theta)}{\partial \theta_i} = -2\frac{(X_s - f^s_\theta)}{h^s_\theta}\frac{\partial f^s_\theta}{\partial \theta_i} - \frac{(X_s - f^s_\theta)^2}{(h^s_\theta)^2}\frac{\partial h^s_\theta}{\partial \theta_i} + \frac{1}{h^s_\theta}\frac{\partial h^s_\theta}{\partial \theta_i}$$

$$\frac{\partial q_{s,j}(\theta)}{\partial \theta_i} = -2\frac{(X_{s,j} - f^{s,j}_\theta)}{h^{s,j}_\theta}\frac{\partial f^{s,j}_\theta}{\partial \theta_i} - \frac{(X_{s,j} - f^{s,j}_\theta)^2}{(h^{s,j}_\theta)^2}\frac{\partial h^{s,j}_\theta}{\partial \theta_i} + \frac{1}{h^{s,j}_\theta}\frac{\partial h^{s,j}_\theta}{\partial \theta_i}.$$

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

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

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

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

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

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

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

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

$$+\Big(\sum_{i\geq k/2}\alpha_i^{(0)}(M_\theta,\Theta(r))\Big)^{r/4}+\Big(\sum_{i\geq k/2}\alpha_i^{(1)}(f_\theta,\Theta(r))\Big)^{r/4}+\Big(\sum_{i\geq k/2}\alpha_i^{(1)}(M_\theta,\Theta(r))\Big)^{r/4}\Big]$$

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

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

■

5.5. *Consistency when the breaks are known.* When the breaks are known, we can chose $v_n=1$ for all n in the penalization of (3.2) as the penalization term does not matter at all.

Proposition 5.1 *For all $j=1,\dots,K^*$, under the assumptions of Lemma 5.3 1-) with $v_n=1$ for all n, if the assumption Id($\Theta(r)$) 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 5.3

Corollary 5.1 *i-) under the assumptions of Lemma 5.3 1-) we have:*

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

ii-) Under assumptions of Lemma 5.3 2-) we have:

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

We conclude the proof of Proposition 5.1 using $\mathcal{L}_j(\theta)=-\frac{1}{2}\mathbb{E}\left(q_{0,j}(\theta)\right)$ has a unique maximum in θ^*_j (see [14]). From the almost sure convergence of the quasi-likelihood in i-) of Corollary 5.1, it comes:

$$\widehat{\theta}_n(T^*_j)=\underset{\theta\in\Theta(r)}{\text{Argmax}}\left(\frac{1}{n^*_j}\widehat{L}_n\big(T^*_j,\theta\big)\right)\xrightarrow[n\to\infty]{\text{a.s.}}\theta^*_j.$$

■

Proof of Corollary 5.1 Note that the proof of Lemma 5.3 can be repeated by replacing L_n by the quasi-likelihood $\widehat{L}_n$. Thus, we obtain for $i=0,1,2$,

$$\frac{v_{n^*_j}}{n^*_j}\left\|\frac{\partial^i\widehat{L}_n\big(T^*_j,\theta\big)}{\partial\theta^i}-\frac{\partial^i L_{n,j}\big(T^*_j,\theta\big)}{\partial\theta^i}\right\|_{\Theta(r)}\xrightarrow[n\to\infty]{}0.\tag{5.10}$$

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

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

Using (5.10), the convergence to the limit likelihood follows.

ii-) From Lemma 4 and Theorem 1 of [2], $\Big\|\frac{1}{n^*_j}\frac{\partial^i L_{n,j}(T^*_j,\theta)}{\partial\theta^i}-\frac{\partial^i\mathcal{L}_j(\theta)}{\partial\theta^i}\Big\|_{\Theta(r)}\xrightarrow[n\to\infty]{\text{a.s.}}0$ for $i=1,2$ and we conclude from (5.10).

■

5.6. *Proof of Theorem 3.1.* This proof is divided into two parts. In **part (1)** K^* is assumed to be known and we show $(\underline{\widehat{\tau}}_n, \underline{\widehat{\theta}}_n) \xrightarrow[n\to\infty]{\mathcal{P}} (\underline{\tau}^*, \underline{\theta}^*)$. 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 $\underline{t} \in \mathcal{F}_{K^*}$:

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

It comes that $\underline{\widehat{t}}_n = \underset{\underline{t}\in\mathcal{F}_{K^*}}{\operatorname{Argmin}} \left(\widehat{I}_n(\underline{t})\right)$. We show that $\underline{\widehat{\tau}}_n \xrightarrow[n\to\infty]{\mathrm{P}} \underline{\tau}^*$ as it implies $\widehat{\theta}_n(\widehat{T}_{n,j}) - \widehat{\theta}_n(T_j^*) \xrightarrow[n\to\infty]{\mathrm{P}} 0$ and from Proposition 5.1 $\widehat{\theta}_n(\widehat{T}_{n,j}) \xrightarrow[n\to\infty]{\mathrm{P}} \theta_j^*$ for all $j = 1, \cdots, K^*$. 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 $\mathbb{P}(\|\underline{\widehat{\tau}}_n - \underline{\tau}^*\|_m > \eta) \simeq \mathbb{P}(|\widehat{t}_n - t^*| > \eta n)$. But

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

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

$$\begin{aligned} \widehat{I}_n(t) - \widehat{I}_n(t^*) = 2\Big(&\widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1^*)) - \widehat{L}_n(T_1^*, \widehat{\theta}_n(T_1)) + \widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_2^*)) \\ &- \widehat{L}_n(T_1 \cap T_2^*, \widehat{\theta}_n(T_1)) + \widehat{L}_n(T_2, \widehat{\theta}_n(T_2^*)) - \widehat{L}_n(T_2, \widehat{\theta}_n(T_2))\Big). \end{aligned} \tag{5.11}$$

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 5.1 and Corollary 5.1, we get the following convergence, uniformly on V_{η,u_n},

$$\widehat{\theta}_n(T_1^*) \xrightarrow[n\to\infty]{a.s.} \theta_1^*,\ \ \widehat{\theta}_n(T_2^*) \xrightarrow[n\to\infty]{a.s.} \theta_2^*,\ \ \widehat{\theta}_n(T_2) \xrightarrow[n\to\infty]{a.s.} \theta_2^* \ \text{ and } \ \Big\|\frac{\widehat{L}_n\big(T_1^*, \theta\big)}{n} - \tau_1^* \mathcal{L}_1(\theta)\Big\|_{\Theta(r)} \xrightarrow[n\to\infty]{a.s.} 0,$$
$$\Big\|\frac{\widehat{L}_n\big(T_1 \cap T_2^*, \theta\big)}{t - t^*} - \mathcal{L}_2(\theta)\Big\|_{\Theta(r)} \xrightarrow[n\to\infty]{a.s.} 0,\ \ \Big\|\frac{\widehat{L}_n\big(T_2, \theta\big)}{n - t} - \mathcal{L}_2(\theta)\Big\|_{\Theta(r)} \xrightarrow[n\to\infty]{a.s.} 0.$$

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

$$\Big\|\frac{\widehat{L}_n\big(T_1^*, \theta\big)}{\cdots} - \tau_1^* \mathcal{L}_1(\theta)\Big\| \cdots < \frac{\varepsilon}{\cdots};\ \ \Big\|\frac{\widehat{L}_n\big(T_1 \cap T_2^*, \theta\big)}{\cdots} - \mathcal{L}_2(\theta)\Big\| \cdots < \frac{\varepsilon}{\cdots};\ \ \Big|\frac{\widehat{L}_n\big(T_1^*, \widehat{\theta}_n(T_1^*)\big)}{\cdots} - \tau_1^* \mathcal{L}_1(\theta^*)\Big| < \frac{\varepsilon}{\cdots}$$

$$\Big|\frac{\widehat{L}_n\big(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*)\big)}{t-t^*}-\mathcal{L}_2(\theta_2^*)\Big|<\frac{\varepsilon}{6};\qquad \frac{n-t}{n}\Big|\frac{\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))-\widehat{L}_n(T_2,\widehat{\theta}_n(T_2))}{n-t}\Big|<\frac{\varepsilon}{6}$$

Thus, for $n>N_0$,

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

Then,

$$\frac{\widehat{L}_n\big(T_1^*,\widehat{\theta}_n(T_1^*)\big)}{n}-\frac{\widehat{L}_n\big(T_1^*,\widehat{\theta}_n(T_1)\big)}{n} > \tau_1^*\Big(\mathcal{L}_1(\theta_1^*)-\mathcal{L}_1(\widehat{\theta}_n(T_1))\Big)-\frac{\varepsilon}{3}. \tag{5.12}$$

Similarly, for $n>N_0$:

$$\frac{\widehat{L}_n\big(T_1\cap T_2^*,\widehat{\theta}_n(T_2^*)\big)}{n}-\frac{\widehat{L}_n\big(T_1\cap T_2^*,\widehat{\theta}_n(T_1)\big)}{n}>\eta\Big(\mathcal{L}_2(\theta_2^*)-\mathcal{L}_2(\widehat{\theta}_n(T_1))\Big)-\frac{\varepsilon}{3}. \tag{5.13}$$

Finally, for $n>N_0$,

$$\frac{\widehat{L}_n(T_2,\widehat{\theta}_n(T_2^*))-\widehat{L}_n(T_2,\widehat{\theta}_n(T_2))}{n} > -\frac{\varepsilon}{6}, \tag{5.14}$$

and from (5.11) and inequalities (5.12), (5.13) and (5.14) we obtain uniformly in t:

$$\frac{\widehat{I}_n(t)-\widehat{I}_n(t^*)}{n}>\tau_1^*\Big(\mathcal{L}_1(\theta_1^*)-\mathcal{L}_1(\widehat{\theta}_n(T_1))\Big)+\eta\Big(\mathcal{L}_2(\theta_2^*)-\mathcal{L}_2(\widehat{\theta}_n(T_1))\Big)-\frac{5}{6}\varepsilon,\quad n>N_0.$$

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

In any case we prove that $\widehat{I}_n(t)-\widehat{I}_n(t^*)>\dfrac{\varepsilon}{6}n$ for $n>N_0$ and all $t\in V_{\eta,u_n}$. It implies that $\mathbb{P}\Big(\min\limits_{\underline{t}\in V_{\eta,u_n}}(\widehat{I}_n(t)-\widehat{I}_n(t^*))\leq 0\Big)\underset{n\to\infty}{\longrightarrow}0$ and we show similarly $\mathbb{P}\Big(\min\limits_{\underline{t}\in W_{\eta,u_n}}(\widehat{I}_n(t)-\widehat{I}_n(t^*))\leq 0\Big)\underset{n\to\infty}{\longrightarrow}0$. 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 $K\geq 2$, $x=(x_1,\cdots,x_{K-1})\in\mathbb{R}^{K-1}$, $y=(y_1,\cdots,y_{K^*-1})\in\mathbb{R}^{K^*-1}$, 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 5.4 *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 $\mathcal{F}_K \times \Theta(r)^K$ 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 5.5 which is proved below (see also [18]):

Lemma 5.5 *Under the assumptions of Lemma 5.3 i-), for any $K \geq 2$, there exists $C_K > 0$ such as:*

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

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

$$\begin{aligned} \mathrm{P}(\widehat{K}_n = K) &\leq \mathrm{P}\Big(\inf_{(\underline{t},\underline{\theta}) \in \mathcal{F}_K \times \Theta(r)^K} (\widetilde{J}_n(K, \underline{t}, \underline{\theta})) \leq \widetilde{J}_n(K^*, \underline{t}^*, \underline{\theta}^*) \Big) \\ &\leq \mathrm{P}\Big(\inf_{(\underline{t},\underline{\theta}) \in \mathcal{F}_K \times \Theta(r)^K} (\widehat{J}_n(K, \underline{t}, \underline{\theta}) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*)) \leq \frac{n}{v_n} (K^* - K). \Big). \end{aligned} \tag{5.15}$$

i-) For $K < K^*$, we decompose $\widehat{J}_n(K, \underline{t}, \underline{\theta}) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*) = n(u_n(\underline{t}, \underline{\theta}) + e_n(\underline{t}, \underline{\theta}))$ where u_n is defined in Lemma 5.5 and

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

It comes from the relation (5.15) that:

$$\mathrm{P}(\widehat{K}_n = K) \leq \mathrm{P}\Big(\inf_{(\underline{t},\underline{\theta}) \in \mathcal{F}_K \times \Theta_K} (u_n(\underline{t}, \underline{\theta}) + e_n(\underline{t}, \underline{\theta})) \leq \frac{\beta_n}{n} (K^* - K) \Big). \tag{5.16}$$

Corollary 5.1 ensures that $e_n(\underline{t}, \underline{\theta})) \to 0$ a.s. and uniformly on $\mathcal{F}_K \times \Theta(r)^K$. By Lemma 5.5, there exists $C_K > 0$ such that $u_n(\underline{t}, \underline{\theta}) \geq C_K \|\underline{t} - \underline{t}^*\|_\infty / n$ for all $(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta(r)^K$. 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 $u_n(\underline{t}, \underline{\theta}) > 0$ for all $(\underline{t}, \underline{\theta}) \in \mathcal{F}_K \times \Theta(r)^K$ and since $1/v_n \xrightarrow[n\to\infty]{} 0$, we deduce from (5.16) that $\mathrm{P}(\widehat{K}_n = K) \xrightarrow[n\to\infty]{} 0$.

ii-) Now let $K^* < K \leq K_{\max}$. from (5.16) and the Markov Inequality we have:

$$\begin{aligned} \mathrm{P}(\widehat{K}_n = K) &\leq \mathrm{P}\Big(\widehat{J}_n(K, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*) + \frac{n}{v_n}(K - K^*) \leq 0 \Big) \\ &\leq \mathrm{P}\Big(|\widehat{J}_n(K, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*)| \geq \frac{n}{v_n} \Big) \\ &\leq \frac{v_n}{n} \mathbb{E} |\widehat{J}_n(K, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*)|. \end{aligned} \tag{5.17}$$

Denote $\widehat{\underline{t}}_n = (\widehat{t}_{n,1}, \cdots, \widehat{t}_{n,K})$. By Lemma 5.4, there exists some subset $\{k_j \ , 1 \leq j \leq K^* - 1\}$ of $\{1, \cdots, K-1\}$ such that for any $j = 1, \cdots, K^* - 1$, $\widehat{t}_{n,k_j}/n \to \tau_j^*$. Denoting $k_0 = 0$ and $k_{K^*} = K$, we have:

$$\widehat{J}_n(K, \widehat{\underline{t}}_n, \widehat{\underline{\theta}}_n) - \widehat{J}_n(K^*, \underline{t}^*, \underline{\theta}^*) = 2\Big(\sum^{K^*} \widehat{L}_n(T_j^*, \theta_j^*) - \sum^{K} \widehat{L}_n(\widehat{T}_{n,k}, \widehat{\theta}_{n,k}) \Big)$$

$$= 2\sum_{j=1}^{K^*}\Big[\widehat L_n(T_j^*,\theta_j^*)-\sum_{k=k_{j-1}+1}^{k_j}\widehat L_n(\widehat T_{n,k},\widehat\theta_{n,k})\Big]$$

and from (5.17) we deduce that:

$$\begin{aligned}\mathrm{P}(\widehat K_n=K) &\le \frac{2v_n}{n}\sum_{j=1}^{K^*}\mathbb{E}\Big|\widehat L_n(T_j^*,\theta_j^*)-\sum_{k=k_{j-1}+1}^{k_j}\widehat L_n(\widehat T_{n,k},\widehat\theta_{n,k})\Big| \\ &\le C\sum_{j=1}^{K^*}\frac{v_{n_j^*}}{n_j^*}\mathbb{E}\Big|\widehat L_n(T_j^*,\theta_j^*)-\sum_{k=k_{j-1}+1}^{k_j}\widehat L_n(\widehat T_{n,k},\widehat\theta_{n,k})\Big|.\end{aligned}$$

Since for any $j=1,\cdots,K^*-1$, it comes from Lemma 5.3 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 5.5 Let $K\ge1$ and consider the real function υ define on $\Theta\times\Theta$ by:

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

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 [14]). By Lemma 3.3 of [17], there exists $C_{\theta^*}>0$ such that for any $(\underline{t},\underline{\theta})\in\mathcal{F}_K\times\Theta_K$

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

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

5.7. *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

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

it suffices to show that $\lim\limits_{\delta\to\infty}\lim\limits_{n\to\infty}\mathbb{P}(\delta<|\widehat t_n-t^*|\le u_n)=0$.
Denote $V_{\delta,u_n}=\{\ t\in\mathbb{Z}/\ \ \delta<|t-t^*|\le u_n\ \}$. Then,

$$\mathbb{P}(\delta<|\widehat t_n-t^*|\le u_n)\le\mathbb{P}\Big(\min_{t\in V_{\delta,u_n}}(\widehat I_n(t)-\widehat I_n(t^*))\le0\Big).$$

Let $t\in V_{\delta,u_n}$ (for example $t\ge t^*$). With the notation of the proof of Theorem 3.1, we have $\widehat L_n(T_1^*,\widehat\theta_n(T_1^*))\ge\widehat L_n(T_1^*,\widehat\theta_n(T_1))$ and from (5.11) we obtain:

$$\frac{\widehat I_n(t)-\widehat I_n(t^*)}{t-t^*}\ge\frac{2}{t-t^*}\Big(\widehat L_n(T_1\cap T_2^*,\widehat\theta_n(T_2^*))-\widehat L_n(T_1\cap T_2^*,\widehat\theta_n(T_1))+\widehat L_n(T_2,\widehat\theta_n(T_2^*))-\widehat L_n(T_2,\widehat\theta_n(T_2))\Big).$$

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

$$\widehat\theta_n(T_1)=\underset{\theta\in\Theta(r)}{\mathrm{Argmax}}\left(\frac1n\widehat L_n(T_1,\theta)\right)\quad\underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}}\theta_1^*.$$

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 $\widehat\theta_n(T_2)\in\overset{\circ}{\Theta}(r)$ 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:

$$0=\frac{\partial\widehat L_n(T_2,\widehat\theta_n(T_2^*))}{\partial\theta_i}+\frac{\partial^2\widehat L_n(T_2,\widetilde\theta_{n,i})}{\partial\theta\partial\theta_i}(\widehat\theta_n(T_2))-\widehat\theta_n(T_2^*))\tag{5.18}$$

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 (5.18):

$$\frac{\partial\widehat L_n(T_1\cap T_2^*,\widehat\theta_n(T_2^*))}{\partial\theta_i}=\frac{\partial^2\widehat L_n(T_2,\widetilde\theta_{n,i})}{\partial\theta\partial\theta_i}(\widehat\theta_n(T_2))-\widehat\theta_n(T_2^*)),\qquad\forall i=1,\dots,d,$$

and it follows:

$$\frac{1}{t-t^*}\frac{\partial\widehat L_n(T_1\cap T_2^*,\widehat\theta_n(T_2^*))}{\partial\theta}=\frac{n-t}{t-t^*}A_n\cdot(\widehat\theta_n(T_2)-\widehat\theta_n(T_2^*))\tag{5.19}$$

with $A_n:=\Big(\frac{1}{n-t}\frac{\partial^2\widehat L_n(T_2,\widetilde\theta_{n,i})}{\partial\theta\partial\theta_i}\Big)_{1\leq i\leq d}$. Corollary 5.1 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 $A_n\underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}}-\frac12\mathbb E\Big(\frac{\partial^2q_{0,2}(\theta_2^*)}{\partial\theta^2}\Big)$. Under assumption (Var), $\mathbb E\Big(\frac{\partial^2q_{0,2}(\theta_2^*)}{\partial\theta^2}\Big)$ is a non-singular matrix (see [2]). Then, we deduce from (5.19) that

$$\frac{n-t}{t-t^*}(\widehat\theta_n(T_2)-\widehat\theta_n(T_2^*))\underset{n,\delta\to\infty}{\overset{a.s.}{\longrightarrow}}0.\tag{5.20}$$

We conclude by the Taylor expansion on $\widehat L_n$ that gives

$$\frac{1}{t-t^*}|\widehat L_n(T_2,\widehat\theta_n(T_2))-\widehat L_n(T_2,\widehat\theta_n(T_2^*))|$$
$$\leq\frac{1}{2(t-t^*)}\|\widehat\theta_n(T_2))-\widehat\theta_n(T_2^*)\|^2\sup_{\theta\in\Theta(r)}\Big\|\frac{\partial^2\widehat L_n(T_2,\theta)}{\partial\theta^2}\Big\|\to0\quad\text{a.s.}\qquad\blacksquare$$

5.8. *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 $j\in\{1,\cdots,K^*\}$. By Theorem 3.2 it comes $\widehat t_j-t^*_j=o_P(\log(n))$. Using relation (5.20), we obtain: $\widehat\theta_n(\widehat T_j)-\widehat\theta_n(T^*_j)=o_P(\dfrac{\log(n)}{n})$. Hence, $\sqrt{n^*_j}\big(\widehat\theta_n(\widehat T_j)-\widehat\theta_n(T^*_j)\big)\xrightarrow[n\to\infty]{P}0$ 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, $\widehat\theta_n(T^*_j)\in\overset{\circ}{\Theta}(r)$. 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

$$\frac{\partial L_n(T^*_j,\widehat\theta_n(T^*_j))}{\partial\theta_k}=\frac{\partial L_n(T^*_j,\theta^*_j)}{\partial\theta_k}+\frac{\partial^2 L_n(T^*_j,\widetilde\theta_{n,k})}{\partial\theta\partial\theta_k}(\widehat\theta_n(T^*_j)-\theta^*_j). \tag{5.21}$$

Let $F_n=-2\Big(\dfrac{1}{n^*_j}\dfrac{\partial^2 L_n(T^*_j,\widetilde\theta_{n,k})}{\partial\theta\partial\theta_k}\Big)_{1\le k\le d}$. By Lemma 5.3 and Corollary 5.1, $F_n\xrightarrow[n\to\infty]{\text{a.s.}}F(\theta^*_j)$ (where $F(\theta^*_j)$ is defined by (3.8)). But, under (Var), $F(\theta^*_j)$ is a non singular matrix (see [2]). Thus, for n large enough, F_n is invertible and (5.21) 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 (3.8). 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 5.3 and the fact that $1/\sqrt{n}=O(v_n/n)$. ■

REFERENCES

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

[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.

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

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

[5] Billingsley. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, 1968.

[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.

[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.

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

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

[10] Feller, W. *An Introduction to Probability Theory and its Applications*, vol. 2. Wiley, 1966.

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

[12] Giraitis, L., Leipus, R., and Surgailis, D. Recent advances in ARCH modelling. In *Long-Memory in Economics* (2006), G. Teyssière and A. Kirman, Eds., Springer Verlag., pp. 3–38.

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

[14] Jeantheau, T. Strong consistency of estimators for multivariate arch models. *Econometric Theory 14*, 1 (1998), 70–86.

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

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

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

[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.

[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.

[21] Page, E. S. On problems in which a change in a parameter occurs at an unknown point. *Biometrika 44*, (1957), 248–252.

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

[23] Straumann, D., and Mikosch, T. Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist. 34*, 5 (2006), 2449–2495.

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

SAMM, Université Paris 1
90 Rue de Tolbiac
75634 Paris Cedex 13, FRANCE
E-mail: Jean-Marc.Bardet@univ-paris1.fr
URL: http://samm.univ-paris1.fr/-Jean-Marc-Bardet

E-mail: William.Kengne@malix.univ-paris1.fr
URL: http://samm.univ-paris1.fr/William-Kengne

CEREMADE, Université Paris-Dauphine
Place du Maréchal de Lattre de Tassigny
75016 Paris, FRANCE
E-mail: Olivier.Wintenberger@ceremade.fr
URL: http://wintenberger.fr