Detecting multiple change-points in general causal time series using penalized quasi-likelihood
Jean-Marc Bardet, William Charky Kengne, Olivier Wintenberger

To cite this version:

HAL Id: hal-00507759
https://hal.archives-ouvertes.fr/hal-00507759
Submitted on 30 Jul 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 (num-
ber 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 consis-
tency 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 ob-
tained. 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

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 $(x_i)_{i \in \mathbb{N}} \in \mathbb{R}^N$, $M((x_i)_{i \in \mathbb{N}})$ is a $(m \times p)$
non-zero real matrix and $f((x_i)_{i \in \mathbb{N}}) \in \mathbb{R}^m$. Let $T \subset \mathbb{Z}$ and $(\xi_i)_{i \in \mathbb{Z}}$ be a sequence of centered inde-
pendent 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-
maximum likelihood estimator; Model selection by penalized likelihood.
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:

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

The existence and properties of these general affine processes were studied in Bardet and Wintenberger \cite{2} as a particular case of chains with infinite memory considered in Doukhan and Wintenberger \cite{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 \((X_1, \ldots, X_n)\) of \( X = (X_t)_{t \in \mathbb{Z}} \) is observed where

\[
X \in \mathcal{M}_{T^*_j}(\Theta, \mathcal{T}^*_j) \quad \text{for all } j = 1, \ldots, K^*,
\]

- \( K^* \in \mathbb{N}^* \), \( T^*_j = \{t^*_j-1 + 1, t^*_j-1 + 2, \ldots, t^*_j\} \) with \( 0 < t^*_1 < \ldots < t^*_K^* \), \( t^*_j \in \mathbb{N} \) 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 \( j = 1, \ldots, K^* \).

The aim in the problem is the estimation of the unknown parameters \( (K^*, (t^*_j)_{1 \leq j \leq K^*}, (\theta^*_{j})_{1 \leq j \leq K^*}) \).

In the literature it is generally supposed that \( X \) is a stationary process on each set \( T^*_j \) and is independent on each \( T^*_k \), \( k \neq j \) (for instance in \cite{18}, \cite{15}, \cite{6} and \cite{7}). Here the problem \cite{13} does not induce such assumption and thus the framework is closer to the applications, see Remark 1 in \cite{7}.

In the problem of change-point detection, numerous papers were devoted to the CUSUM procedure (see for instance Kokozska and Leipus \cite{13} in the specific case of ARCH(\infty) processes). In Lavielle and Ludena \cite{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. \cite{6} proposed a likelihood ratio as the estimator of break points for an AR(p) process. Lavielle and Moulines \cite{18} consider a general contrast using the mean square errors for estimating the parameters. In Davis et al. \cite{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 \cite{14},
Berkes et al. [1], Franck and Zakoian [11] or generalizations of GARCH processes (see Mikosch and Straumann [23], Robinson and Zaffaroni [22]). Bardet and Wintenberger [3] study the asymptotic normality of the parameter of models very dependent of their whole past (see Remark 3.3) for any compact set \( \Theta \subseteq \mathbb{R}^d \) and for any \( g : \Theta \rightarrow \mathbb{R}^d \); \( \| g \|_{\Theta} = \sup_{\theta \in \Theta}(\| g(\theta) \|) \). Hence the minimization of the penalized contrast leads to an estimator (see (3.2)) of the parameter \( K^* \). Classical heuristics such as the BIC one or the MDL one lead to choose an “optimal” penalty \( \beta_n \) which overpenalizes the number of breaks to avoid artificial breaks in cases of models very dependent of their whole past (see Remark 3.3).

The main results of the paper are the following: under Lipschitzian condition on \( f_\theta \) and \( M_\theta \), the estimator \( (\hat{K}_n, (\hat{t}_j/n)_{1 \leq j \leq \hat{K}_n-1}, (\hat{\theta}_j)_{1 \leq j \leq \hat{K}_n}) \) is consistent when the moment of order \( r \) on the innovations and \( X \) is larger than 2. If moreover Lipschitzian conditions are satisfied by the derivatives of \( f_\theta \) and \( M_\theta \) and if \( r \geq 4 \), then the convergence rate of \( (\hat{t}_j/n)_{1 \leq j \leq \hat{K}_n-1} \) is \( O_P(n^{-1}) \) and a Central Limit Theorem (CLT) for \( (\hat{\theta}_j)_{1 \leq j \leq \hat{K}_n} \) (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(\( \infty \)), ARCH(\( \infty \)) and TARCH(\( \infty \)) 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 \( \mathcal{M}_t(f_\theta, M_\theta) \). Let \( \theta \in \mathbb{R}^d \) and \( M_\theta \) and \( f_\theta \) be numerical functions such that for all \( (x_i)_{i \in \mathbb{N}} \in \mathbb{R}^\mathbb{N} \), \( M_\theta((x_i)_{i \in \mathbb{N}}) \neq 0 \) and \( f_\theta((x_i)_{i \in \mathbb{N}}) \in \mathbb{R} \). We use the following different norms:

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

Let \( \Psi_\theta = f_\theta, M_\theta \) and \( i = 0, 1, 2 \), then for any compact set \( \Theta \subseteq \mathbb{R}^d \), define
From \([2]\) we have:
\[
(2.3)
\]

causal models and ARCH-type models simultaneously: for ARCH-type models existence of Markov chains. The subset
\[
(2.2)
\]

Let
\[
(\theta \in \Theta)\]

and under Assumption 
\[
(2.1)
\]

following assumption holds on \(h_{\theta} = M_{\theta}^2\):

**Assumption \(A_i(h_{\theta}, \Theta)\):** Assume that \(\|\partial^i h_{\theta}(0)/\partial \theta^i\|_{\Theta} < \infty\) and there exists a sequence of non-negative real number \((\alpha_k^{(i)}(h_{\theta}, \Theta))_{i \geq 1}\) such as \(\sum_{k=1}^{\infty} \alpha_k^{(i)}(h_{\theta}, \Theta) < \infty\) satisfying
\[
\left\| \frac{\partial^i h_{\theta}(x)}{\partial \theta^i} - \frac{\partial^i h_{\theta}(y)}{\partial \theta^i} \right\|_{\Theta} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(h_{\theta}, \Theta) |x_k - y_k| \quad \text{for all } x, y \in \mathbb{R}^N.
\]

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

**Assumption \(A_i(h_{\theta}, \Theta)\):** Assume that \(\|\partial^i h_{\theta}(0)/\partial \theta^i\|_{\Theta} < \infty\) and there exists a sequence of non-negative real number \((\alpha_k^{(i)}(h_{\theta}, \Theta))_{i \geq 1}\) such as \(\sum_{k=1}^{\infty} \alpha_k^{(i)}(h_{\theta}, \Theta) < \infty\) satisfying
\[
\left\| \frac{\partial^i h_{\theta}(x)}{\partial \theta^i} - \frac{\partial^i h_{\theta}(y)}{\partial \theta^i} \right\|_{\Theta} \leq \sum_{k=1}^{\infty} \alpha_k^{(i)}(h_{\theta}, \Theta) |x_k^2 - y_k^2| \quad \text{for all } x, y \in \mathbb{R}^N.
\]

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:
\[
\beta^{(i)}(\theta) := \sum_{k \geq 1} \beta_k^{(i)}(\theta) \quad \text{where } \quad \beta_k^{(i)}(\theta) := \alpha_k^{(i)}(f_{\theta}, \{\theta\}) + (\mathbb{E}|\xi_0|^r)^{1/r} \alpha_k^{(i)}(M_{\theta}, \{\theta\}),
\]
and under Assumption \(A_i(h_{\theta}, \Theta)\)
\[
\tilde{\beta}^{(i)}(\theta) := \sum_{k \geq 1} \tilde{\beta}_k^{(i)}(\theta) \quad \text{where } \quad \tilde{\beta}_k^{(i)}(\theta) := (\mathbb{E}|\xi_0|^r)^{2/r} \alpha_k^{(i)}(h_{\theta}, \{\theta\}).
\]

The dependence with respect to \(r\) of \(\beta_k^{(i)}(\theta)\) and \(\tilde{\beta}_k^{(i)}(\theta)\) are omitted for notational convenience. Then define:
\[
\Theta(r) := \{\theta \in \Theta, \; A_0(f_{\theta}, \{\theta\}) \text{ and } A_0(M_{\theta}, \{\theta\}) \text{ hold with } \beta^{(0)}(\theta) < 1\}
\]
\[
\cup \{\theta \in \mathbb{R}^d, \; f_{\theta} = 0 \; \text{and } \; A_0(h_{\theta}, \{\theta\}) \text{ holds with } \tilde{\beta}^{(0)}(\theta) < 1\}.
\]

From \([3]\) 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 \((\xi_t)_{t \geq t}\) 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 \(A_i(\Psi, \Theta)\) are classical when studying the existence of solution of general models. For instance, Duflo \([3]\) 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(\(\Theta\)) \(\exists h > 0\) such that \(\inf_{\theta \in \Theta} \langle |h_\theta(x)| \rangle \geq h\) for all \(x \in \mathbb{R}^N\).

Assumption Id(\(\Theta\)): For all \(\theta, \theta' \in \Theta^2\),
\[
(f_\theta(X_0, X_{-1}, \cdots) = f_{\theta'}(X_0, X_{-1}, \cdots) \text{ and } h_\theta(X_0, X_{-1}, \cdots) = h_{\theta'}(X_0, X_{-1}, \cdots) \ a.s.) \Rightarrow \theta = \theta'.
\]

Assumption Var(\(\Theta\)): For all \(\theta \in \Theta\),
\[
\left(\frac{\partial f_\theta}{\partial \theta(i)}(X_0, X_{-1}, \cdots)\right)_{1 \leq i \leq d} \neq 0 \quad \text{or} \quad \left(\frac{\partial h_\theta}{\partial \theta(i)}(X_0, X_{-1}, \cdots)\right)_{1 \leq i \leq d} \neq 0 \quad \text{a.s.}
\]

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

2.2. Existence of the solution to the problem \((\mathbf{1.2})\). Consider the problem \((\mathbf{1.2})\) 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 \(\theta_1^*\), 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 \(K \geq 2\), \(\mathcal{F}_K\) = \(\{t = (t_1, \ldots, t_{K-1}) \mid 0 < t_1 < \ldots < t_{K-1} < n\}\). In particular, \(\mathbf{t}^* = (t_1^*, \ldots, t_{K^* - 1}) \in \mathcal{F}_{K^*}\) is the true vector of instants of change;
- For \(K \in \mathbb{N}^*\) and \(\mathbf{t} \in \mathcal{F}_K\), \(T_k = \{t \in \mathbb{Z}, t_{K-1} < t \leq t_k\}\) and \(n_k = \text{Card}(T_k)\) with \(1 \leq k \leq K\). In particular; \(T_j^* = \{t \in \mathbb{Z}, t_{j-1}^* < t \leq t_j^*\}\) and \(n_j^* = \text{Card}(T_j^*)\) for \(1 \leq j \leq K^*\). For all \(1 \leq k \leq K\) and \(1 \leq j \leq K^*\), let \(n_{kj} = \text{Card}(T_j^* \cap T_k)\);

The following proposition establishes the existence of the nonstationary solution of the problem \((\mathbf{1.2})\) and its moment properties.

Proposition 2.2 Consider the problem \((\mathbf{1.2})\). 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 \((\mathbf{1.2})\) 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 \((\mathbf{1.2})\) 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 \(X_1 = 0\) 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 \(\mathbf{t}^*\) and the parameters \(\theta^*\) is based on the minimum of a penalized contrast. It is clear that if \((\xi_t)_t\) is a Gaussian process and if \(X \in \mathcal{M}_T(f_\theta, M_\theta)\) then for \(s \in\)
\( T = \{ t, t+1, \ldots, t' \} \), the distribution of \( X_s \mid (X_{s-j})_{j \in \mathbb{N}^*} \) is \( N(f_{\theta}(X_{s-1}, \ldots), h_{\theta}(X_{s-1}, \ldots)) \). Therefore, with the notation \( f_{\theta} = f_{\theta}(X_{s-1}, X_{s-2}, \ldots) \), \( M_{\theta} = M_{\theta}(X_{s-1}, X_{s-2}, \ldots) \) and \( h_{\theta} = M_{\theta}^2 \), we deduce the conditional log-likelihood on \( T \) (up to an additional constant)

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

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 \((X_{s-j})_{j \in \mathbb{N}^*}\). We approximate it by:

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

with \( \hat{f}_{\theta} = f_{\theta}(X_{t-1}, \ldots, X_{1}, u) \), \( \hat{M}_{\theta} = M_{\theta}(X_{t-1}, \ldots, X_{1}, u) \) and \( \hat{h}_{\theta} = (\hat{M}_{\theta})^2 \) 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 [12] or in [13]. 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 \( \ell \in \mathcal{F}_K \), \( \theta \in \Theta(r)^K \), define the contrast function \( \hat{J}_n \) by the expression:

\[
\hat{J}_n(K, \ell, \theta) := -2 \sum_{k=1}^{K} \hat{J}_n(T_k, \theta_k) = -2 \sum_{k=1}^{K} \sum_{j=1}^{K^*} \hat{L}_n(T_k \cap T_j^*, \theta_k),
\]

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 \( K \in \{1, \ldots, K_{\max}\} \) and \((\ell, \theta) \in \mathcal{F}_K \times \Theta(r)^K \) (\( \Theta(r) \) is supposed to be a compact set) define the penalized contrast \( \tilde{J}_n \) by

\[
\tilde{J}_n(K, \ell, \theta) := \hat{J}_n(K, \ell, \theta) + \frac{n}{v_n} K = \hat{J}_n(K, \ell, \theta) + \beta_n K
\]

and the penalized contrast estimator \((\hat{K}_n, \hat{\ell}_n, \hat{\theta}_n)\) of \((K^*, \ell^*, \theta^*)\) as

\[
(\hat{K}_n, \hat{\ell}_n, \hat{\theta}_n) = \arg\min_{1 \leq K \leq K_{\max}} \arg\min_{(\ell, \theta) \in \mathcal{F}_K \times \Theta(r)^K} (\tilde{J}_n(K, \ell, \theta)) \quad \text{and} \quad \hat{\theta}_n = \hat{\theta}_n(\hat{K}_n, \hat{\ell}_n).
\]

For \( K \geq 1 \) and \( \ell \in \mathcal{F}_K \), denote \( \hat{\theta}_n(\ell) = (\hat{\theta}(T_1), \ldots, \hat{\theta}(T_K)) := \arg\min_{\theta \in \Theta(r)^K} (\tilde{J}_n(\ell, \theta)) = \arg\min_{\theta \in \Theta(r)^K} (\hat{J}_n(\ell, \theta)). \)

Then, \( \hat{\theta}_n(T_k) \) is the QMLE of \( \theta_k^* \) computed on \( T_k \) and \( \hat{\theta}_n(\ell^*) \) is the QMLE of \( \theta^* \).  

**Remark 3.2** If \( K^* \) is known, the estimator of \((\ell^*, \theta^*)\) may be obtained by minimizing \( \hat{J}_n \) instead of \( \tilde{J}_n \). However the knowledge of \( K^* \) does not improve the asymptotic results established in this paper.
3.2. Consistency of $\hat{H}_n$. For establishing the consistency, we add the couple of following classical assumptions in the problem of break detection:

**Hypothesis B**: \( \min_{j=1,\ldots,K^*-1} \|\theta_j^* - \theta_j^\prime\| > 0. \)

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

**Hypothesis C**: there exists \( \tau_1^*, \ldots, \tau_{K^* - 1}^* \) with \( 0 < \tau_1^* < \ldots < \tau_{K^* - 1}^* < 1 \) such that for \( j = 1, \ldots, K^* \), \( t_j^* = [n\tau_j^*] \) (where \( [x] \) is the floor of \( x \)). The vector \( \tau^* = (\tau_1^*, \ldots, \tau_{K^* - 1}^*) \) 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 \( t_k^* \), \( T_k^* \) 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 \( \alpha_i(h_\Theta, \Theta) = 0 \) and if \( A_i(h_\Theta, \Theta) \) holds then \( \alpha_i(M_\Theta, \Theta) = 0 \), define:

**Hypothesis H**: \( i = 0, 1, 2 \): For \( 0 \leq p \leq i \), the assumptions \( A_p(f_\Theta, \Theta) \), \( A_p(M_\Theta, \Theta) \) (or \( A_p(h_\Theta, \Theta) \)) hold and for all \( j = 1, \ldots, K^* \) there exists \( r \geq 1 \) such that \( \theta_j^\prime \in \Theta(r) \). Denoting \( c^* = \min_{j=1,\ldots,K^*} \left( - \log(\beta_0(\theta_j^*)) / 8 \right) \) and \( \min_{j=1,\ldots,K^*} \left( - \log(\beta_0(\theta_j^*)) / 8 \right) \)

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

\[
(3.4) \quad \sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} \sum_{\ell / k^c \leq r/4} \beta_0^{(\ell)}(\theta_j^*) \right) < \infty \quad \text{and} \quad \sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} \sum_{\ell / k^2 \leq r/4} \left( \alpha_0^{(\ell)}(f_\Theta, \Theta(r)) + \alpha_0^{(\ell)}(M_\Theta, \Theta(r)) + \alpha_0^{(\ell)}(h_\Theta, \Theta(r)) \right) \right) < \infty.
\]

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 \( M(f_\Theta, M_\Theta) \) 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 \( (v_n)_{n \in \mathbb{N}} \) are deduced from a result of Kounias [10]. The conditions on \( (v_n)_{n \in \mathbb{N}} \) are not too restrictive:

(1) geometric case: if \( \alpha_i^{(i)}(f_\Theta, \Theta(r)) + \alpha_i^{(i)}(M_\Theta, \Theta(r)) + \alpha_i^{(i)}(h_\Theta, \Theta(r)) = O(a^i) \) with \( 0 \leq a < 1 \), then any \( (v_n)_{n \in \mathbb{N}} \) such as \( v_n = o(n) \) can be chosen (for instance \( v_n = n(\log(n))^{-1} \)).

(2) Riemanian case: if \( \alpha_i^{(i)}(f_\Theta, \Theta(r)) + \alpha_i^{(i)}(M_\Theta, \Theta(r)) + \alpha_i^{(i)}(h_\Theta, \Theta(r)) = O(\ell^{-\gamma}) \) with \( \gamma > 1 \),
• if \( \gamma > 1 + (1 \sqrt{4r^{-1}}) \), then all sequence \((v_n)_n\) such as \( v_n = o(n) \) can be chosen (for instance \( v_n = n (\log n)^{-1} \)).
• if \( 1 \sqrt{4r^{-1}} < \gamma \leq 1 + (1 \sqrt{4r^{-1}}) \), then any \((v_n)_n\) such as \( v_n = O(n^{\gamma/(1+4r^{-1})} (\log n)^{-\delta}) \) with \( \delta > 1 \sqrt{4r^{-1}} \) 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:

\[
(\hat{K}_n, \hat{\theta}_n, \hat{\tau}_n) \overset{P}{\to} (K^*, \theta^*, \tau^*).
\]

**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 Riemannian 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 \( \hat{\theta}_n \) and \( \hat{\tau}_n \), 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 \( (\|\hat{L}_n - \tau^*\|_m)_{n > 1} \) is uniformly tight in probability, i.e.

\[
\lim_{\delta \to \infty} \lim_{n \to \infty} P(\|\hat{L}_n - \tau^*\|_m > \delta) = 0.
\]

This theorem induces that \( w_n^{-1} \|\hat{L}_n - \tau^*\|_m \overset{P}{\to} 0 \) for any sequence \((w_n)_n\) such as \( w_n \to \infty \) and therefore \( \|\hat{L}_n - \tau^*\|_m = o_P(w_n) \): 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 \( \theta^*_j \). By convention if \( \hat{K}_n < K^* \), set \( \hat{T}_j = \hat{T}_{\hat{K}_n} \) for \( j \in \{\hat{K}_n, \ldots, K^*\} \). 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 \( \theta^*_j \in \hat{\Theta}(r) \) for all \( j = 1, \ldots, K^* \), we have

\[
\sqrt{n}^j (\hat{\theta}_n(T_j) - \theta^*_j) \overset{D}{\to} \mathcal{N}_d (0, F(\theta^*_j)^{-1} G(\theta^*_j) F(\theta^*_j)^{-1}),
\]

where, using \( q_{0,j} \) defined in \((3.2)\), the matrix \( F \) and \( G \) are such as

\[
(F(\theta^*_j))_{k,l} = \mathbb{E}\left( \frac{\partial^2 q_{0,j}(\theta^*_j)}{\partial \theta_k \partial \theta_l} \right) \quad \text{and} \quad (G(\theta^*_j))_{k,l} = \mathbb{E}\left( \frac{\partial q_{0,j}(\theta^*_j)}{\partial \theta_k} \frac{\partial q_{0,j}(\theta^*_j)}{\partial \theta_l} \right).
\]
**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 H2 it excludes models with finite moments \( r \geq 4 \) satisfying: 
\[
\ell^{-\gamma} = O(\alpha_i^0(f_{\theta}, \Theta(r)) + \alpha_i^0(M_{\theta}, \Theta(r)) + \alpha_i^0(h_{\theta}, \Theta(r)))
\]
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 \( \hat{\theta}_n \) hold but we do not get any rate of convergence for \( \tilde{\theta}_n \).

4. Some examples.

4.1. AR(\( \infty \)) models. Consider AR(\( \infty \)) with \( K^* - 1 \) breaks defined by the equation:
\[
X_t = \sum_{k \geq 1} \phi_k(\theta^*_j)X_{t-k} + \xi_t, \quad t_{j-1}^* < t \leq t_j^*, \quad j = 1, \ldots, K^*.
\]
It corresponds to the problem (1.2) with models \( M_{T^*}(f_{\theta}, M_{\theta}) \) where \( f_{\theta}(x_1, \ldots) = \sum_{k \geq 1} \phi_k(\theta)x_k \) and \( M_{\theta} = 1 \). Assume that \( \Theta \) 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 \( E|\xi_0|^{r} < \infty \). Then Assumptions D(\( \Theta \)) and A0(\( f_{\theta}, \Theta \)) hold automatically with \( \beta = 1 \) and \( \alpha_k^0(f_{\theta}, \Theta(r)) = \| \phi_k(\theta) \|_{\Theta} \). Then,

- Assume that Id(\( \Theta \)) holds and that there exists \( r \geq 2 \) such that \( E|\xi_0|^{r} < \infty \). 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 \( (\hat{K}_n, \hat{\xi}_n, \hat{\theta}_n) \).

- Moreover, if \( r \geq 4, \gamma > 3/2 \) and \( \phi_k \) 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 \( \hat{\xi}_n \) and the CLT (3.7) satisfied by \( \hat{\theta}_n(\hat{T}_j) \) for all \( j \).

Note that this problem of change detection was considered by Davis et al. in [8] under moments of order greater than 4 is required. In Davis et al. [9], 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(\( \infty \)) models. Consider an ARCH(\( \infty \)) model with \( K^* - 1 \) breaks defined by:
\[
X_t = \left( \psi_0(\theta^*_j) + \sum_{k=1}^{\infty} \psi_k(\theta^*_j)x_{t-k}^2 \right)^{1/2} \xi_t, \quad t_{j-1}^* < t \leq t_j^*, \quad j = 1, \ldots, K^*.
\]
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 \( E(\xi_0^2) = 1 \). 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 \( \Theta \) 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 \( (\hat{K}_n, \hat{\xi}_n, \hat{\theta}_n) \) when \( \theta^*_j \in \Theta \) for all \( j \).
Moreover, if \( r \geq 4 \) and \( \psi_k \) 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 \( \theta^*_j \in \Theta(4) \) for all \( j \), then the penalization \( v_n = \beta_n = \sqrt{n} \) as in Remark 3.3 ensures the convergence (3.6) of \( \hat{L}_n \) and the CLT (3.7) satisfied by \( \hat{\gamma}_n(\hat{T}_j) \) 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:

\[
X_t = \sigma_t \xi_t, \quad \sigma_t^2 = a_{0,j} + \sum_{k=1}^q a_{k,j} X_{t-k}^2 + \sum_{k=1}^p b_{k,j} \sigma_{t-k}^2 \quad t_j - 1 < t \leq t_j, \quad j = 1, \ldots, K^*
\]

with \( E(\xi_0^2) = 1 \). 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 [14]) a nonnegative sequence \( (\psi_k(\theta))_k \) such that \( \sigma_t^2 = \psi_0(\theta) + \sum_{k \geq 1} \psi_0(\theta) X_{t-k}^2 \). Remark that this sequence is twice differentiable with respect to \( \theta \) and that its derivatives are exponentially decreasing. Moreover for any \( \theta \in \Theta \) it holds \( \sum_{k \geq 1} \psi_k(\theta) \leq \left( \sum_{k=1}^q a_k \right) / \left( 1 - \sum_{k=1}^p b_k \right) \) and one can consider:

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

Then if \( \sum_{k=1}^q a_k + \sum_{k=1}^p b_k < 1 \) 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 \( (E|\xi_0|^4)^{1/2} \sum_{k=1}^q a_k + \sum_{k=1}^p b_k < 1 \) 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 [2] can be chosen in this case.

### 4.3. Estimates breaks in TARCH(\( \infty \)) model.

Consider a TARCH(\( \infty \)) model with breaks defined by:

\[
X_t = \sigma_t \xi_t, \quad \sigma_t = b_0(\theta^*_j) + \sum_{k \geq 1} \left( b^+_k(\theta^*_j) \max(X_{t-k}, 0) - b^-_k(\theta^*_j) \min(X_{t-k}, 0) \right),
\]

for any \( t_{j-1} < t \leq t_j, \quad j = 1, \ldots, K^* \) and where \( \sum_{k \geq 1} \max(\| b^+_k(\theta) \|_{\Theta}, \| b^-_k(\theta) \|_{\Theta}) < \infty \). 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 \( \| \xi_0 \|_{r}, \sum_{k \geq 1} \max(\| b^+_k(\theta) \|_{\Theta}, \| b^-_k(\theta) \|_{\Theta}) < 1 \) for \( r \geq 2 \). If there exists \( \gamma > 1 \) 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 \( \left( \hat{K}_n, \hat{\gamma}_n, \hat{\theta}_n \right) \) 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 \( \beta_n = \sqrt{n} \) ensures the convergence (3.4) of \( \hat{L}_n \) and the CLT (3.7) satisfied by \( \hat{\gamma}_n(\hat{T}_j) \) for all \( j \) (with \( \theta^*_j \in \Theta(4) \)).
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 \[\text{2}\]). Therefore, X is defined by induction as follows:

\[
X_t := M_{\theta_j}^0 (X_{t-1}, X_{t-2}, \cdots) \xi_t + f_{\theta_j}^0(X_{t-1}, X_{t-2}, \cdots), \forall t \in T_j^*; \ j = 2, \cdots K^*.
\]

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 \[\text{3}\] we remark that

\[
\|X_t\|_r \leq \frac{\|Z_{t,1}\|_r}{1 - \beta(0) (\theta_1^*)}
\]

for \(t \leq t_1^*\), with \(Z_{t,j} := M_{\theta_j}^0 (0, 0, \cdots) \xi_t + f_{\theta_j}^0 (0, 0, \cdots)\) for all \(j = 1, \ldots, K^*\). Assume that there exists \(C_{r,t} > 0\) such that \(C_{r,t} = \sup_{t < \ell} \|X_\ell\|_r\) and let \(t \in T_j^*\), then

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

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

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

Then, we have:

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

We deduce that

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

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 \(\|X_t\|_r \leq \max_{1 \leq j \leq K^*} \{\|Z_{t,j}\|_r + C_{r,t+1} \beta(0) (\theta_j^*)\}\). Thus, by definition of \(C_{r,t+1} = \sup_{t \leq \ell} \|X_\ell\|_r\) we obtain

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

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

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

Finally we obtain the desired result with

\[
C = \max_{1 \leq j \leq K^*} \frac{\|M_{\theta^*}(0, 0, \cdots) \xi_0 + f_{\theta^*}(0, 0, \cdots)\|_{r}}{1 - \beta^{(0)}(\theta^*)} \wedge \max_{1 \leq j \leq K^*} \frac{\|M_{\theta^*}(0, 0, \cdots) \xi_0\|_{r}}{\sqrt{1 - \beta^{(0)}(\theta^*)}}.
\]

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

for all \( t \in \mathbb{Z}, \quad \mathbb{E}\left( \sup_{\theta \in \Theta(r)} |q_t(\theta)| \right) \leq C.
\]

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

\[
\|f_\theta\|_{\Theta(r)}^2 \leq 2 \left( \|f_\theta - f_\theta(0, \ldots)\|_{\Theta(r)}^2 + \|f_\theta(0, \ldots)\|_{\Theta(r)}^2 \right)
\]

\[
\leq 2 \left( \sum_{i \geq 1} \alpha_i^{(0)}(f_\theta, \Theta(r)) \right) \cdot \sum_{i \geq 1} \alpha_i^{(0)}(f_\theta, \Theta(r)) |X_{t-i}|^2 + \|f_\theta(0, \ldots)\|_{\Theta(r)}^2,
\]

therefore

\[
\mathbb{E}\|f_\theta\|_{\Theta(r)}^2 \leq 2 \left( \sum_{i \geq 1} \alpha_i^{(0)}(f_\theta, \Theta(r)) \right)^2 + \|f_\theta(0, \ldots)\|_{\Theta(r)}^2.
\]

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

\[
|\log(h_\theta)| = |\log(h) + \log(h_\theta^t)| \leq 1 + |\log(h)| + \frac{1}{h} h_\theta^t.
\]

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

\[
\mathbb{E}\left( \sup_{\theta \in \Theta(r)} |q_t(\theta)| \right) \leq 1 + |\log(h)| + \frac{1}{h} (\mathbb{E}\|h_\theta\|_{\Theta(r)} + 2\mathbb{E}|X_t|^2 + 2\mathbb{E}|f_\theta^t|^2_{\Theta(r)}) \leq C.
\]

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 [4] that the equation

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

has \( r \) order stationary solution \( (X_{t,j})_{t \in \mathbb{Z}} \) 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 that \( \theta^*_j \in \Theta(r) \) for \( j = 1, \ldots, K^* \) for \( r \geq 2 \). Then:
1. $X_t = X_{t,1}$ for all $t \leq t^*_j$;
2. There exists $C > 0$ such that for any $j \in \{2, \cdots, K^*\}$, for all $t \in T^*_j$,
\[
\|X_t - X_{t,j}\|_r \leq C \left( \inf_{1 \leq p \leq t - t^*_j} \left\{ \beta^{(0)}(\theta^*_j)^{(t-t^*_j)/p} + \sum_{i \geq p} \beta^{(0)}_i(\theta^*_j) \right\} \right)
\]
\[
\|X^2_t - X^2_{t,j}\|_{r/2} \leq C \left( \inf_{1 \leq p \leq t - t^*_j} \left\{ \tilde{\beta}^{(0)}(\theta^*_j)^{(t-t^*_j)/p} + \sum_{i \geq p} \beta^{(0)}_i(\theta^*_j) \right\} \right).
\]

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 $C_r > 0$ such that $\|X^2_t - X^2_{t,j}\|_{r/2} \leq \|X_t\|_r + \|X_{t,j}\|_r \leq C + \max_{1 \leq j \leq K^*} \|X_{0,j}\|_r \leq C_r$ for all $j = 1, \ldots, K^*$ and $t \in \mathbb{Z}$. For $1 \leq p \leq t - t^*_j - 1$ let $u_t := \sup_{t - t^*_j - 1 \leq j \leq t^*_j} \|X_t - X_{t,j}\|_r$. Then $\|X_t - X_{t,j}\|_r \leq u_{(t-t^*_j-1)/p}$ and for any $t \leq i \leq t^*_j$:
\[
\|X_t - X_{i,j}\|_r \leq \sum_{k \geq 1} \beta^{(0)}_k(\theta^*_j)\|X_{t-k} - X_{t-k,j}\|_r
\]
\[
\leq \sum_{k=1}^p \beta^{(0)}_k(\theta^*_j)\|X_{t-k} - X_{t-k,j}\|_r + C_r \sum_{k>p} \beta^{(0)}_k(\theta^*_j)
\]
\[
\leq \beta^{(0)}(\theta^*_j)u_{(t-t^*_j-1)/p} + C_r \sum_{k>p} \beta^{(0)}_k(\theta^*_j).
\]

Similarly, it is easy to show that for all $1 \leq \ell \leq [(t - t^*_j - 1)/p]$ we have
\[
u_t \leq \beta^{(0)}(\theta^*_j)u_{\ell-1} + C_r \sum_{k>p} \beta^{(0)}_k(\theta^*_j).
\]

Denote $a = \beta^{(0)}(\theta^*_j) < 1$, $b = C_r \sum_{k>p} \beta^{(0)}_k(\theta^*_j)$ such that $u_t \leq au_{t-1} + b$. Considering $w_0 = u_0$ and $w_t = au_{t-1} + b$, then $w_t = a^t w_0 + b(1 - a^{-1})(1 - a) \leq a^t w_0 + b/(1 - a)$. Since $u_0 \leq C_r$ by definition and $u_t \leq w_t$ for any $t$, we have:
\[
u_t \leq a^tu_0 + b \frac{1}{1-a} \leq (\beta^{(0)}(\theta^*_j))^\ell C_r + \frac{C_r}{1 - \beta^{(0)}(\theta^*_j)} \sum_{k>p} \beta^{(0)}_k(\theta^*_j)
\]
\[
\leq \frac{C_r}{1 - \beta^{(0)}(\theta^*_j)} \left( \beta^{(0)}(\theta^*_j)^\ell + \sum_{k>p} \beta^{(0)}_k(\theta^*_j) \right).
\]

Thus for all $1 \leq p \leq t - t^*_j - 1$
\[
\|X^2_t - X^2_{t,j}\|_{r/2} \leq C_r \|X_t - X_{t,j}\|_r \leq C_r u_{(t-t^*_j-1)/p} \leq C \left( \beta^{(0)}(\theta^*_j)^{(t-t^*_j-1)/p} + \sum_{i \geq p} \beta^{(0)}_i(\theta^*_j) \right)
\]

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
\[
\|X^2_t - X^2_{t,j}\|_{r/2} \leq \sum_{k \geq 1} \beta^{(0)}_k(\theta^*_j)\|X^2_{t-k} - X^2_{t-k,j}\|_{r/2}.
\]
For all $j = 1, \ldots, K^*$ and $t \in \mathbb{Z}$, by Proposition 2.2, $\|X_t^2 - X_{t,j}^2\|_{r/2} \leq C_r^2$ and therefore

$$\hat{u}_t \leq \tilde{\beta}^{(0)}(\theta^*_t)\hat{u}_{t-1} + C_r^2 \sum_{k>p} \tilde{\beta}^{(0)}_k(\theta^*_t)$$

for $\hat{u}_t = \sup_{t_p \leq t \leq t'_p} \|X_t^2 - X_{t,j}^2\|_{r/2}$ and Lemma 5.2 is proved.

5.4. The asymptotic behavior of the likelihood. For the process $(X_{i,j})_{t \in T_j', j=1,\ldots,K^*}$, for any $j \in \{1, \ldots, K^*\}$ and $s \in T_j^*$ denote:

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

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 $T \subseteq T_j^*$, 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 $r \geq 2$ holds then for all $j = 1, \ldots, K^*$:

$$\frac{v_{n,j}^\ast}{n_j^\ast} \left\| L_n(T_j^*, \theta) - L_{n,j}(T_j^*, \theta) \right\|_{\Theta(r)} \overset{a.s.}{\underset{n \to \infty}{\to}} 0.$$

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

$$\frac{v_{n,j}^\ast}{n_j^\ast} \left\| \frac{\partial^i L_n(T_j^*, \theta)}{\partial \theta^i} - \frac{\partial^i L_{n,j}(T_j^*, \theta)}{\partial \theta^i} \right\|_{\Theta(r)} \overset{a.s.}{\underset{n \to \infty}{\to}} 0.$$

Proof

1-) For any $\theta \in \Theta(r)$, $\left| \frac{1}{n_j^\ast} L_n(T_j^*, \theta) - \frac{1}{n_j^\ast} L_{n,j}(T_j^*, \theta) \right| \leq \frac{1}{n_j^\ast} \sum_{k=1}^{n_j^\ast} |q_{t_j^* - 1 + k}(\theta) - q_{t_j^* - 1 + k,j}(\theta)|$.

Then:

$$\frac{v_{n,j}^\ast}{n_j^\ast} \left\| \frac{1}{n_j^\ast} L_n(T_j^*, \theta) - \frac{1}{n_j^\ast} L_{n,j}(T_j^*, \theta) \right\|_{\Theta(r)} \leq \frac{v_{n,j}^\ast}{n_j^\ast} \sum_{k=1}^{n_j^\ast} \left\| q_{t_j^* - 1 + k}(\theta) - q_{t_j^* - 1 + k,j}(\theta) \right\|_{\Theta(r)}.$$

By Corollary 1 of Kounias [10], with $r \leq 4$ and no loss of generality, it is sufficient that

$$\sum_{k=1}^{\infty} \left( \frac{v_k}{k} \right)^{r/4} \mathbb{E}(\left\| q_{t_j^* - 1 + k}(\theta) - q_{t_j^* - 1 + k,j}(\theta) \right\|_{\Theta(r)}^{r/4}) < \infty.$$

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

$$(5.3) \quad |q_{s,j}(\theta) - q_{s,j}(\theta)| \leq \frac{1}{h_{\theta}^{s,j}} \left| X_s - f_{\theta}^{s,j} \right|^2 |h_{\theta}^{s,j} - h_{\theta}^{s,j}|$$

$$+ \frac{1}{h_{\theta}} \left( \left| X_s^2 - X_{s,j}^2 \right| + \left| f_{\theta}^{s,j} - f_{\theta}^{s,j} \right| \left| f_{\theta}^{s,j} + f_{\theta}^{s,j} + 2X_s \right| + 2\left| f_{\theta}^{s,j} \right| \left| X_s - X_{s,j} \right| + \left| h_{\theta}^{s,j} - h_{\theta}^{s,j} \right| \right).$$
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(\theta)} \leq C \left( 1 + |X_{s,j}| + |X_s|^2 + \|f_\theta^{s-j}\|_{\Theta(\theta)} + \|f_\theta\|^2\right)
\times \left( |X_s - X_{s,j}| + \|f_\theta^s - f_\theta^{s-j}\|_{\Theta(\theta)} + \|h_\theta^s - h_\theta^{s-j}\|_{\Theta(\theta)}\),
$$

and by Cauchy-Schwarz Inequality,

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

Using Proposition 2.2 and the argument of the proof of Lemma 5.1, we claim that $\mathbb{E}|X_s|^r \leq C$, $\mathbb{E}|f_\theta|^r_{\Theta(\theta)} \leq C$ and that $\mathbb{E}|f_\theta^{s-j}|^{r/2}_{\Theta(\theta)} \leq C$. Thus:

$$(5.4) \quad (\mathbb{E}|q_s(\theta) - q_{s,j}(\theta)|^{\gamma/4})^2 \leq C (\mathbb{E}|X_s - X_{s,j}|^{r/2} + \mathbb{E}|f_\theta^s - f_\theta^{s-j}|^{r/2} + \mathbb{E}|h_\theta^s - h_\theta^{s-j}|^{r/2}).$$

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

$$
|X_s - X_{s,j}|^{r/2} \leq |X_s - X_{s,j}|_r \leq C \inf_{1 \leq p \leq k} \left\{ \beta^{(0)}(\theta)^{k/p} + \sum_{i \geq p} \beta_i^{(0)}(\theta^*) \right\}
\leq C \inf_{1 \leq p \leq k/2} \left\{ \beta^{(0)}(\theta^*)^{k/(2p)} + \sum_{i \geq p} \beta_i^{(0)}(\theta^*) \right\}.
$$

(5.5) $\quad \mathbb{E}|X_s - X_{s,j}|^{r/2} \leq C \left( \inf_{1 \leq p \leq k/2} \left\{ \beta^{(0)}(\theta^*)^{k/(2p)} + \sum_{i \geq p} \beta_i^{(0)}(\theta^*) \right\} \right)^{r/2}.
$$

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

$$(5.6) \quad \|h_\theta^s - h_\theta^{s,j}\|_{\Theta(\theta)}^{r/2} \leq C \sum_{i \geq 1} \alpha_i^{(0)}(M_\theta, \Theta(\theta)) \|X_{s-i} - X_{s-i,j}\|_r.
$$

From (5.6) we obtain:

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

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 5.2:

$$
|X_{s-i} - X_{s-i,j}|_r \leq C \inf_{1 \leq p \leq k-i} \left\{ \beta^{(0)}(\theta^*)^{(k-i)/p} + \sum_{i \geq p} \beta_i^{(0)}(\theta^*) \right\}
\leq C \inf_{1 \leq p \leq k/2} \left\{ \beta^{(0)}(\theta^*)^{k/(2p)} + \sum_{i \geq p} \beta_i^{(0)}(\theta^*) \right\}.
$$

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

$$(5.7) \quad \mathbb{E}|h_\theta^s - h_\theta^{s,j}|^{\gamma/2}_{\Theta(\theta)} \leq C \left( \inf_{1 \leq p \leq k/2} \left\{ \beta^{(0)}(\theta^*)^{k/(2p)} + \sum_{i \geq p} \beta_i^{(0)}(\theta^*) \right\} + \sum_{i \geq k/2} \alpha_i^{(0)}(M_\theta, \Theta(\theta)) \right)^{r/2}.
$$
Similarly, we obtain:

\( (5.8) \quad \mathbb{E}\|f_\theta^* - f_\theta^{j, k}\|_{\Theta(r)}^{r/2} \leq C\left( \inf_{1 \leq p \leq k/2} \{ \beta_0(0)(\theta_j^* k/(2p)) + \sum_{i \geq p} \beta_i(0)(\theta_j^*) \} + \sum_{i \geq k/2} \alpha_i(0)(f_\theta, \Theta(r)) \right)^{r/2}. \)

Relations (5.4), (5.5), (5.7) and (5.8) give (the same inequality holds with \( h_\theta \) replaced by \( M_\theta \)):

\( (5.9) \quad \mathbb{E}\|q_s(\theta) - q_{s,j}(\theta)\|_{\Theta(r)}^{r/4} \leq C\left( \left( \inf_{1 \leq p \leq k/2} \{ \beta_0(0)(\theta_j^* k/(2p)) + \sum_{i \geq p} \beta_i(0)(\theta_j^*) \} \right)^{r/4} + \left( \sum_{i \geq k/2} \alpha_i(0)(f_\theta, \Theta(r)) \right)^{r/4} + \left( \sum_{i \geq k/2} \alpha_i(0)(M_\theta, \Theta(r)) \right)^{r/4} \right]. \)

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

\( \sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} \left( \beta_0(0)(\theta_j^*) \right)^{rk/8u_k} < \infty. \)

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

\[ \sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} \mathbb{E}\|q_{s,j+k}(\theta) - q_{s,j+k}(\theta)\|_{\Theta(r)}^{r/4} \leq \sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} \left( \beta_0(0)(\theta_j^*) \right)^{rk/8u_k} \]

\[ + \sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} \left( \sum_{i \geq u_k} \beta_i(0)(\theta_j^*) \right)^{r/4} + \sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} \left( \sum_{i \geq k/2} \alpha_i(0)(f_\theta, \Theta(r)) + \alpha_i(0)(M_\theta, \Theta(r)) \right)^{r/4}. \]

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 as previously remarking that (5.3) has the simplified form:

\[ |q_s(\theta) - q_{s,j}(\theta)| \leq \frac{1}{k^2} X^2_s h_\theta - h_s^j + \frac{1}{h} \|X_s^2 - X_{s,j}^2\| + \frac{1}{h} |h_s^j - h_s^{j,j}|. \]

Then

\[ \mathbb{E}\|q_s(\theta) - q_{s,j}(\theta)\|_{\Theta(r)}^{r/4} \leq C \mathbb{E}\left[ |X_s^2 - X_{s,j}^2| + \|h_s^j - h_s^{j,j}\|_{\Theta(r)}^{r/2} \right]. \]

As \( \|h_s^j - h_s^{j,j}\|_{\Theta(r)} \leq C \sum_{i \geq 1} \alpha_i(0)(h_\theta, \Theta(r)) \|X_{s-i}^2 - X_{s-i,j}^r \|_{\Theta(r)}^{r/2} \) we derive from Lemma 5.2:

\[ \mathbb{E}\|q_s(\theta) - q_{s,j}(\theta)\|_{\Theta(r)}^{r/4} \leq C\left( \inf_{1 \leq p \leq k/2} \{ \tilde{\beta}(0)(\theta_j^* k/(2p)) + \sum_{i \geq p} \tilde{\beta}_i(0)(\theta_j^*) \} \right)^{r/4} + \left( \sum_{i \geq k/2} \alpha_i(0)(h_\theta, \Theta(r)) \right)^{r/4}. \]

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:

\[ \frac{v_{n_j^*}}{n_j^*} \left\| \frac{\partial L_n(T_j^*, \theta)}{\partial \theta_i} - \frac{\partial L_m(T_j^*, \theta)}{\partial \theta_i} \right\|_{\Theta(r)} \leq \frac{v_{n_j^*}}{n_j^*} \sum_{k=1}^{n_j^*} \left\| \frac{\partial q_{s,j+k-1}(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j+k-1}(\theta)}{\partial \theta_i} \right\|_{\Theta(r)} . \]
By Corollary 1 of Kounias (1969), when \( r \leq 4 \) with no loss of generality, it suffices to show

\[
\sum_{k \geq 1} \left( \frac{v_k}{k} \right)^{r/4} E \left( \left\| \frac{\partial q_{t_{j-1}+k}^*}{\partial \theta_i} \right\|_{\Theta(r)}^{r/4} \right) < \infty.
\]

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

\[
\frac{\partial q_s(\theta)}{\partial \theta_i} = -2 \frac{(X_s - f_{\theta}^s)}{h_{\theta}^s} \frac{\partial f_{\theta}^s}{\partial \theta_i} - \frac{(X_s - f_{\theta}^s)^2}{(h_{\theta}^s)^2} \frac{\partial h_{\theta}^s}{\partial \theta_i} + \frac{1}{h_{\theta}^s} \partial h_{\theta}^s \frac{\partial f_{\theta}^s}{\partial \theta_i},
\]

\[
\frac{\partial q_{s,j}(\theta)}{\partial \theta_i} = -2 \frac{(X_{s,j} - f_{\theta}^{s,j})}{h_{\theta}^{s,j}} \frac{\partial f_{\theta}^{s,j}}{\partial \theta_i} - \frac{(X_{s,j} - f_{\theta}^{s,j})^2}{(h_{\theta}^{s,j})^2} \frac{\partial h_{\theta}^{s,j}}{\partial \theta_i} + \frac{1}{h_{\theta}^{s,j}} \partial h_{\theta}^{s,j} \frac{\partial f_{\theta}^{s,j}}{\partial \theta_i},
\]

Thus, using \(|a_1 b_1 | c_1 - a_2 b_2 c_2 | \leq |a_1 - a_2| b_2 | c_2 | + |b_1 - b_2| a_1 | c_2 | + |c_1 - c_2| a_1 | b_1 |\),

\[
\left\| \frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i} \right\|_{\Theta(r)} \leq 2 \left( \frac{1}{h_{\theta}^s} \left\| h_{\theta}^s - h_{\theta}^{s,j} \right\|_{\Theta(r)} \left\| X_s - f_{\theta}^s \right\|_{\Theta(r)} + \frac{1}{h_{\theta}^{s,j}} \left\| h_{\theta}^{s,j} - h_{\theta}^s \right\|_{\Theta(r)} \left\| X_{s,j} - f_{\theta}^{s,j} \right\|_{\Theta(r)} \right)
\]

So for all \( s \geq t_{j-1}^* \) it holds:

\[
\left( E \left\| \frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i} \right\|_{\Theta(r)}^{r/4} \right)^2 \leq C \left( 1 + \left\| X_s \right\|_{\Theta(r)}^2 + \left\| X_{s,j} \right\|_{\Theta(r)}^2 + \left\| f_{\theta}^s \right\|_{\Theta(r)}^2 + \left\| f_{\theta}^{s,j} \right\|_{\Theta(r)}^2 \right) \left( \left\| \frac{\partial f_{\theta}^s}{\partial \theta_i} \right\|_{\Theta(r)}^2 + \left\| \frac{\partial f_{\theta}^{s,j}}{\partial \theta_i} \right\|_{\Theta(r)}^2 \right)
\]

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 \( s \geq t_{j-1}^* \),

\[
E \left\| \frac{\partial q_s(\theta)}{\partial \theta_i} - \frac{\partial q_{s,j}(\theta)}{\partial \theta_i} \right\|_{\Theta(r)}^{r/4} \leq C \left[ \left( \inf_{1 \leq p \leq k/2} \left\{ \beta(0)(\theta_j)^{k/(2p)} + \sum_{i \geq 1} \beta(i)(\theta_j)^{k/(2p)} \right\} \right)^{r/4} + \left( \sum_{i \geq k/2} \alpha(i)(\theta_j) \right)^{r/4} \right]
\]
Corollary 5.1

We conclude the proof of Proposition 5.1 using

\(L\)

For all

\(v_n = 1\) for all \(n\) in the penalization of (3.2) as the penalization term does not matter at all.

Proof

Let us first give the following useful corollary of Lemma 5.3

\[2\]

\[1\]

\[\text{Corollary 5.1} \quad i-) \text{ under the assumptions of Lemma 5.3}\]

\[1\]

\[\text{Proof:}\]

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

\[
\sum_{k \geq 1} \left(\frac{v_k}{k}\right)^{r/4} \mathbb{E} \left( \left\| \frac{\partial q_{i_{k-1}+k(j)}}{\partial \theta_i} - \frac{\partial q_{i_{k-1}+k,j}}{\partial \theta_i} \right\|_{(\Theta(\cdot))}^{r/4} \right) < \infty.
\]

\[\text{5.5. Consistency when the breaks are known.}\]

When the breaks are known, we can choose \(v_n = 1\) for all \(n\) in the penalization of (3.2) as the penalization term does not matter at all.

Proposition 5.1

For all \(j = 1, \ldots, K^*\), under the assumptions of Lemma 5.3 \(i-)\) with \(v_n = 1\) for all \(n\), if the assumption \(\text{Id}(\Theta(\cdot))\) holds then

\[
\hat{\theta}_n(T^*_j) \xrightarrow{a.s.} \theta^*_j.
\]

Proof

Let us first give the following useful corollary of Lemma 5.3

\[\text{Corollary 5.1} \quad i-) \text{ under the assumptions of Lemma 5.3}\]

\[1\]

\[\text{Proof:}\]

\[\text{We conclude the proof of Proposition 5.1 using } \mathcal{L}_j(\theta) = -\frac{1}{2} \mathbb{E} \left( q_{0,j}(\theta) \right).\]

\[\text{We conclude the proof of Proposition 5.1 using } \mathcal{L}_j(\theta) = -\frac{1}{2} \mathbb{E} \left( q_{0,j}(\theta) \right) \text{ has a unique maximum in } \theta^*_j \text{ (see [4]). From the almost sure convergence of the quasi-likelihood in } i-) \text{ of Corollary 5.1, it comes:}\]

\[\hat{\theta}_n(T^*_j) = \text{Argmax}_{\theta \in \Theta(r)} \left( \frac{1}{n_j^s} \mathcal{L}_n(T^*_j, \theta) \right) \xrightarrow{a.s.} \theta^*_j.\]

\[\text{Proof of Corollary 5.1}\]

Note that the proof of Lemma 5.3 can be repeated by replacing \(L_n\) by the quasi-likelihood \(\mathcal{L}_n\). Thus, we obtain for \(i = 0, 1, 2, \ldots\),

\[1\]

\[\text{i-) Let } j \in 1, \ldots, K^*. \text{ From [4], we have:}\]

\[\left\| \frac{1}{n_j^s} \mathcal{L}_{n,j}(T^*_j, \theta) - \mathcal{L}_j(\theta) \right\|_{(\Theta(\cdot))} \xrightarrow{a.s.} 0.\]

\[\text{Using (5.10), the convergence to the limit likelihood follows.}\]

\[\text{ii-) From Lemma 4 and Theorem 1 of [4], } \left\| \frac{1}{n_j^s} \mathcal{L}_{n,j}(T^*_j, \theta) - \mathcal{L}_j(\theta) \right\|_{(\Theta(\cdot))} \xrightarrow{a.s.} 0 \text{ for } i = 1, 2 \text{ and we conclude from (5.10).}\]
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 \( \left( \hat{K}_n, \hat{\theta}_n \right) \xrightarrow{p_{n \to \infty}} (\theta^*, \theta^*) \). In part (2), \( K^* \) is unknown and we show \( \hat{K}_n \xrightarrow{p_{n \to \infty}} K^* \) which ends the proof of Theorem 3.1.

**Part (1).** Assume that \( K^* \) is known and denote for any \( t \in F_{K^*} \):

\[
\hat{I}_n(t) := \hat{I}_n(K^*, t, \hat{\theta}_n(t)) = -2 \sum_{k=1}^{K^*} \sum_{j=1}^{K^*} \hat{L}_n(T_k \cap T_j^*, \hat{\theta}_n(T_k))
\]

It comes that \( \hat{L}_n = \text{Argmin} \left( \hat{I}_n(t) \right) \). We show that \( \hat{L}_n \xrightarrow{p_{n \to \infty}} \tau^* \) as it implies \( \hat{\theta}_n(T_{n,j}) \xrightarrow{p_{n \to \infty}} \theta^* \) 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 \( 0 < \eta < 1 \)

\[
V_{\eta,u_n} = \{ t \in \mathbb{Z} / |t - t^*| > \eta n ; u_n \leq t \leq n - u_n \},
\]

\[
W_{\eta,u_n} = \{ t \in \mathbb{Z} / |t - t^*| > \eta n ; 0 < t < u_n \ or \ n - u_n < t \leq n \}.
\]

Asymptotically, we have \( \mathbb{P}(\|\hat{L}_n - \tau^*\|_m > \eta) \approx \mathbb{P}(|\hat{L}_n - t^*| > \eta n) \). But

\[
\mathbb{P}(\|\hat{L}_n - t^*\| > \eta n) \leq \mathbb{P}(\hat{L}_n \in V_{\eta,u_n}) + \mathbb{P}(\hat{L}_n \in W_{\eta,u_n})
\]

\[
\leq \mathbb{P}(\min_{t \in V_{\eta,u_n}} (\hat{L}_n(t) - \hat{L}_n(t^*)) \leq 0) + \mathbb{P}(\min_{t \in W_{\eta,u_n}} (\hat{L}_n(t) - \hat{L}_n(t^*)) \leq 0)
\]

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

Let \( t \in V_{\eta,u_n} \) satisfying \( t^* \leq t \) (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:

\[
(5.11) \quad \hat{L}_n(t) - \hat{L}_n(t^*) = 2(\hat{L}_n(T_1^*, \hat{\theta}_n(T_1^*)) - \hat{L}_n(T_1^*, \hat{\theta}_n(T_1))) + \hat{L}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_2))
\]

As \( \#T_1^* = t^* \), \( \#(T_1 \cap T_2^*) = t - t^* \), \( \#T_2 = n - t \geq u_n \), each term tends to 0 with \( n \). Using Proposition 3.3 and Corollary 3.1, we get the following convergence, uniformly on \( V_{\eta,u_n} \),

\[
\hat{\theta}_n(T_1^*) \xrightarrow{a.s.} \theta_1^*, \hat{\theta}_n(T_2^*) \xrightarrow{a.s.} \theta_2^*, \hat{\theta}_n(T_2) \xrightarrow{a.s.} \theta_2^* \quad \text{and} \quad \frac{\hat{L}_n(T_1^*, \theta)}{n} \xrightarrow{a.s.} 0,
\]

\[
\left| \frac{\hat{L}_n(T_1 \cap T_2^*, \theta)}{t - t^*} - \mathcal{L}_2(\theta) \right| \xrightarrow{a.s.} 0, \quad \left| \frac{\hat{L}_n(T_2, \theta)}{n - t} - \mathcal{L}_2(\theta) \right| \xrightarrow{a.s.} 0.
\]

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

\[
\left| \frac{\hat{L}_n(T_1^*, \theta)}{n} - \tau_1^* \mathcal{L}_1(\theta) \right| < \frac{\varepsilon}{6}; \quad \left| \frac{\hat{L}_n(T_1 \cap T_2^*, \theta)}{t - t^*} - \mathcal{L}_2(\theta) \right| < \frac{\varepsilon}{6}; \quad \left| \frac{\hat{L}_n(T_2, \theta)}{n - t} - \mathcal{L}_2(\theta) \right| < \frac{\varepsilon}{6} \quad \text{and} \quad \left| \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1^*))}{n} - \tau_1^* \mathcal{L}_1(\theta_1^*) \right| < \frac{\varepsilon}{6}.
\]
Then, from (5.11) and inequalities (5.12), (5.13) and (5.14) we obtain uniformly in $t$:

$$\left| \frac{\hat{L}_n(T_1 \cap T_2, \hat{\theta}_n(T_2))}{t - t^*} - \mathcal{L}_2(\theta_2^*) \right| < \frac{\varepsilon}{6}, \quad \frac{n - t}{n} \left| \frac{\hat{L}_n(T_2, \hat{\theta}_n(T_2)) - \hat{L}_n(T_2, \hat{\theta}_n(T_2))}{n - t} \right| < \frac{\varepsilon}{6}.$$ 

Thus, for $n > N_0$,

$$\tau_1^* \mathcal{L}_1(\theta_1^*) - \tau_1^* \mathcal{L}_1(\hat{\theta}_n(T_1)) = \tau_1^* \mathcal{L}_1(\theta_1^*) - \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1^*))}{n} + \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1))}{n} + \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1))}{n} - \tau_1^* \mathcal{L}_1(\theta_1^*)$$

$$\leq \frac{\varepsilon}{6} + \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1))}{n} - \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1))}{n} + \frac{\varepsilon}{6}.$$ 

Then,

(5.12) \[ \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1^*))}{n} - \frac{\hat{L}_n(T_1^*, \hat{\theta}_n(T_1))}{n} > \tau_1^* \left( \mathcal{L}_1(\theta_1^*) - \mathcal{L}_1(\hat{\theta}_n(T_1)) \right) - \frac{\varepsilon}{3}. \]

Similarly, for $n > N_0$:

(5.13) \[ \frac{\hat{L}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_2))}{n} - \frac{\hat{L}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_2))}{n} > \eta \left( \mathcal{L}_2(\theta_2^*) - \mathcal{L}_2(\hat{\theta}_n(T_1)) \right) - \frac{\varepsilon}{3}. \]

Finally, for $n > N_0$,

(5.14) \[ \frac{\hat{L}_n(T_2, \hat{\theta}_n(T_2)) - \hat{L}_n(T_2, \hat{\theta}_n(T_2))}{n} > \frac{\varepsilon}{6}, \]

and from (2.11) and inequalities (5.12), (5.13) and (5.14) we obtain uniformly in $t$:

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

Since $\theta_1^* \neq \theta_2^*$, let $V_1, V_2$ be two open neighborhoods and disjoint of $\theta_1^*$ and $\theta_2^*$ respectively,

$$\delta_i := \inf_{\theta \in V_i} \left( \mathcal{L}_i(\theta_i^*) - \mathcal{L}_i(\theta) \right) > 0 \quad \text{for} \quad i = 1, 2,$$

since the function $\theta \mapsto \mathcal{L}_j(\theta)$ has a strict maximum in $\theta_j^*$ (see [14]). With $\varepsilon = \min(\tau_1^* \delta_1, \eta \delta_2)$, we get

- if $\hat{\theta}_n(T_1) \in V_1$ i.e. $\hat{\theta}_n(T_1) \in V_2$, then $\frac{\hat{L}_n(t) - \hat{L}_n(t^*)}{n} > \eta \delta_2 - \frac{5}{6} \varepsilon \geq \frac{\varepsilon}{6};$
- if $\hat{\theta}_n(T_1) \notin V_1$ i.e. $\hat{\theta}_n(T_1) \in V_1^c$, then $\frac{\hat{L}_n(t) - \hat{L}_n(t^*)}{n} > \tau_1^* \delta_1 - \frac{5}{6} \varepsilon \geq \frac{\varepsilon}{6}.$

In any case we prove that $\frac{\hat{L}_n(t) - \hat{L}_n(t^*)}{n} \geq \frac{\varepsilon}{6} n$ for $n > N_0$ and all $t \in V_{\eta,u_n}$. It implies that

$$\mathbb{P} \left( \min_{t \in V_{\eta,u_n}} \left( \hat{L}_n(t) - \hat{L}_n(t^*) \right) \leq 0 \right) \to 0$$

and we show similarly

$$\mathbb{P} \left( \min_{t \in W_{\eta,u_n}} \left( \hat{L}_n(t) - \hat{L}_n(t^*) \right) \leq 0 \right) \to 0.$$

It follows directly that $\mathbb{P}(\|\hat{L}_n - \tau^*\|_m > \eta) \to 0$ 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 $K \geq 1$, $(\hat{\theta}_n, \hat{\theta}_n)$ obtained by the minimization of $\hat{J}_n(t, \theta)$ on $F_K \times \Theta(r)^K$ and $\hat{\theta}_n = \hat{\theta}_n/n$. Under assumptions of Theorem \[\ref{mainthm}], $\|\hat{\theta}_n - \theta^*\|_n \xrightarrow{n \to \infty} 0$ 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 \[\ref{mainthm}\] i), for any $K \geq 2$, there exists $C_K > 0$ such as:

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

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

$$P(\hat{\theta}_n = K) \leq P\left( \inf_{(t, \theta) \in F_K \times \Theta(r)^K} (\hat{J}_n(K, t, \theta)) \leq \hat{J}_n(K^*, t^*, \theta^*) \right)$$

(5.15)

$$\leq P\left( \inf_{(t, \theta) \in F_K \times \Theta(r)^K} (\hat{J}_n(K, t, \theta) - \hat{J}_n(K^*, t^*, \theta^*)) \leq \frac{n}{v_n} (K^* - K) \right).$$

i-) For $K < K^*$, we decompose $\hat{J}_n(K, t, \theta) - \hat{J}_n(K^*, t^*, \theta^*) = n(u_n(t, \theta) + e_n(t, \theta))$ where $u_n$ is defined in Lemma \[\ref{mainthm}\] and

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

It comes from the relation (5.15) that:

(5.16) $P(\hat{\theta}_n = K) \leq P\left( \inf_{(t, \theta) \in F_K \times \Theta(r)^K} (u_n(t, \theta) + e_n(t, \theta)) \leq \frac{\beta_n}{n} (K^* - K) \right).$

Corollary \[\ref{corollary}\] ensures that $e_n(t, \theta) \to 0$ a.s. and uniformly on $F_K \times \Theta(r)^K$. By Lemma \[\ref{mainthm}\], there exists $C_K > 0$ such that $u_n(t, \theta) \geq C_K \|t - t^*\|_n / n$ for all $(t, \theta) \in F_K \times \Theta(r)^K$. But, since $K < K^*$, for any $t \in F_K$, we have $\|t - t^*\|_n / n = \|t - t^*\|_\infty \geq \min_{1 \leq j \leq K^*} (\tau_j^* - \tau_{j-1}^*/2)$ that is positive by assumption. Then $u_n(t, \theta) > 0$ for all $(t, \theta) \in F_K \times \Theta(r)^K$ and since $1/v_n \xrightarrow{n \to \infty} 0$, we deduce from (5.16) that $P(\hat{\theta}_n = K) \xrightarrow{n \to \infty} 0$.

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

(5.17) $P(\hat{\theta}_n = K) \leq P\left( |\hat{J}_n(K, \hat{\theta}_n) - \hat{J}_n(K^*, \hat{\theta}^*_{\hat{\theta}_n})| \geq \frac{n}{v_n} (K - K^*) \right).$

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

$$\hat{J}_n(K, \hat{\theta}_n, \hat{\theta}_n) - \hat{J}_n(K^*, t^*, \theta^*) = 2 \left( \sum_{j=1}^{K^*} \hat{L}_n(T_j^*, \theta_j^*) - \sum_{k=1}^{K} \hat{L}_n(\hat{\theta}_{n,k}) \right).$$


We conclude in two steps:

$$2 \sum_{j=1}^{K^*} \left[ \hat{L}_n(T_j^*, \theta_j^*) - \sum_{k=k_j-1+1}^{k_j} \hat{L}_n(T_{n,k}, \hat{\theta}_{n,k}) \right]$$

and from (5.17) we deduce that:

$$\Pr(\hat{K}_n = K) \leq \frac{2v_n}{n} \sum_{j=1}^{K^*} \mathbb{E}\left| \hat{L}_n(T_j^*, \theta_j^*) - \sum_{k=k_j-1+1}^{k_j} \hat{L}_n(T_{n,k}, \hat{\theta}_{n,k}) \right|$$

$$\leq C \sum_{j=1}^{K^*} \frac{v_n^*}{n_j} \mathbb{E}\left| \hat{L}_n(T_j^*, \theta_j^*) - \sum_{k=k_j-1+1}^{k_j} \hat{L}_n(T_{n,k}, \hat{\theta}_{n,k}) \right|.$$ 

Since for any $j = 1, \ldots, K^* - 1$, it comes from Lemma 5.3 that

$$\frac{v_n^*}{n_j} \mathbb{E}\left| \hat{L}_n(T_j^*, \theta_j^*) - \sum_{k=k_j-1+1}^{k_j} \hat{L}_n(T_{n,k}, \hat{\theta}_{n,k}) \right| \rightarrow 0,$$

and therefore $\Pr(\hat{K}_n = K) \rightarrow 0$. $\blacksquare$

**Proof of Lemma 5.5** Let $K \geq 1$ and consider the real function $\nu$ define on $\Theta \times \Theta$ by:

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

The function $\nu$ has positive values and $\nu(\theta, \theta') = 0$ if and only if $\theta = \theta'$ since the function $\theta \mapsto \mathcal{L}_j(\theta)$ has a strict maximum in $\theta_j^*$ (see [14]). By Lemma 3.3 of [17], there exists $C_{\theta^*} > 0$ such that for any $(\ell, \ell') \in \mathcal{F}_K \times \Theta_K$

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

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

### 5.7. Proof of Theorem 5.3

Assume with no loss of generality that $K^* = 2$. Denote $(u_n)_n$ a sequence satisfying $u_n \rightarrow \infty$, $u_n/n \rightarrow 0$ and $\Pr(\hat{L}_n - \ell^* > u_n) \rightarrow 0$ (for example $u_n = n \sqrt{\max(|\hat{\tau}_n - \tau^*|, n^{-1})}$). For $\delta > 0$, as we have

$$\Pr(\hat{L}_n - \ell^* > \delta) \leq \Pr(\delta < \hat{L}_n - \ell^* \leq u_n) + \Pr(\hat{L}_n - \ell^* \mid u_n)$$

it suffices to show that $\lim_{\delta \rightarrow 0} \Pr(\delta < \hat{L}_n - \ell^* \leq u_n) = 0$.

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

$$\Pr(\delta < |\hat{L}_n - t^*| \leq u_n) \leq \Pr\left( \min_{t \in V_{\delta, u_n}} (\hat{L}_n(t) - \hat{L}_n(t^*)) \leq 0 \right).$$

Let $t \in V_{\delta, u_n}$ (for example $t \geq t^*$). With the notation of the proof of Theorem 5.1, we have

$$\hat{L}_n(T_1^*, \hat{\theta}_n(T_1^*)) \geq \hat{L}_n(T_1^*, \hat{\theta}_n(T_1))$$

and from (5.11) we obtain:

$$\hat{L}_n(t) - \hat{L}_n(t^*) \geq \frac{2}{t - t^*} \left( \hat{L}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_2^*)) - \hat{L}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_1)) + \hat{L}_n(T_2, \hat{\theta}_n(T_2^*)) - \hat{L}_n(T_2, \hat{\theta}_n(T_2)) \right).$$

We conclude in two steps:
We conclude by the Taylor expansion on \( \hat{\lambda}_n(1 \cap T_2^*, \hat{\theta}_n(T_2^*)) - \hat{\lambda}_n(1 \cap T_2^*, \hat{\theta}_n(T_1)) \) > 0 for \( n \) large enough.

Then 
\[
\frac{\hat{\lambda}_n(T_1, \theta)}{n} = \frac{t^* \hat{\lambda}_n(T_1^*, \theta)}{t^*} + \frac{t - t^* \hat{\lambda}_n(T_1 \cap T_2^*, \theta)}{t - t^*} \quad \text{and since} \quad \frac{t - t^*}{n} \leq \frac{a_n}{n \to \infty} \quad \text{and}
\]
\[
\hat{\theta}_n(T_1) = \text{Argmax}_{\theta \in \Theta(r)} \left( \frac{1}{n} \hat{\lambda}_n(T_1, \theta) \right) \quad \overset{a.s.}{\longrightarrow} \quad \theta_1^*.
\]

It comes that 
\[
\frac{1}{t - t^*} \left( \hat{\lambda}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_2^*)) - \hat{\lambda}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_1)) \right) \quad \overset{a.s.}{\longrightarrow} \quad 0.
\]

For large value of \( n \), we remark that \( \hat{\theta}_n(T_2) \in \Theta^o \) so that \( \partial \hat{\lambda}_n(T_2, \hat{\theta}_n(T_2))/\partial \theta = 0 \). The mean value theorem on \( \partial \hat{\lambda}_n/\partial \theta_i \) for any \( i = 1, \ldots, d \) gives the existence of \( \tilde{\theta}_{n,i} \in [\hat{\theta}_n(T_2), \hat{\theta}_n(T_2)] \) such that:
\[
(5.18) \quad 0 = \frac{\partial^2 \hat{\lambda}_n(T_2, \tilde{\theta}_{n,i})}{\partial \theta_i} (\hat{\theta}_n(T_2)) - \tilde{\theta}_{n,i} \quad \forall i = 1, \ldots, d,
\]

where for \( a, b \in \mathbb{R}^d \), \([a, b] = \{(1 - \lambda)a + \lambda b : \lambda \in [0, 1]\}\). Using the equalities \( \hat{\lambda}_n(T_2^*, \theta) = \hat{\lambda}_n(T_1 \cap T_2^*, \theta) + \hat{\lambda}_n(T_2, \theta) \) and \( \partial \hat{\lambda}_n(T_2, \hat{\theta}_n(T_2))/\partial \theta = 0 \), it comes from (5.18):
\[
\frac{\partial \hat{\lambda}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_2))}{\partial \theta_i} = \frac{\partial^2 \hat{\lambda}_n(T_2, \hat{\theta}_n(T_2))}{\partial \theta_i} (\hat{\theta}_n(T_2)) - \tilde{\theta}_{n,i} \quad \forall i = 1, \ldots, d,
\]

and it follows:
\[
(5.19) \quad \frac{1}{t - t^*} \frac{\partial \hat{\lambda}_n(T_1 \cap T_2^*, \hat{\theta}_n(T_2))}{\partial \theta_i} = \frac{n - t}{t - t^*} A_n \cdot (\hat{\theta}_n(T_2) - \hat{\theta}_n(T_2^*))
\]

with \( A_n := \left( \frac{1}{n - t} \frac{\partial^2 \hat{\lambda}_n(T_2, \hat{\theta}_n(T_2))}{\partial \theta_i} \right)_{1 \leq i \leq d} \). Corollary 5.1 ii-) gives that:
\[
(5.20) \quad \frac{n - t}{t - t^*} (\hat{\theta}_n(T_2) - \hat{\theta}_n(T_2^*)) \overset{a.s.}{\longrightarrow} 0.
\]

We conclude by the Taylor expansion on \( \hat{\lambda}_n \) that gives
\[
|\frac{1}{t - t^*} \hat{\lambda}_n(T_2, \hat{\theta}_n(T_2)) - \hat{\lambda}_n(T_2, \hat{\theta}_n(T_2^*))| \leq \frac{1}{2(t - t^*)} \| \hat{\theta}_n(T_2) - \hat{\theta}_n(T_2^*) \|^2 \quad \sup_{\theta \in \Theta(r)} \left| \frac{\partial^2 \hat{\lambda}_n(T_2, \theta)}{\partial \theta^2} \right| \to 0 \quad \text{a.s.} \]
5.8. Proof of Theorem 5.3. First, \((\hat{\theta}_n(T_j) - \theta_j^*) = (\hat{\theta}_n(T_j) - \hat{\theta}_n(T_j^*)) + (\hat{\theta}_n(T_j^*) - \theta_j^*)\) for any \(j \in \{1, \cdots, K^*\}\). By Theorem 5.2 it comes \(\hat{t}_j - t_j^* = o_P(\log(n))\). Using relation (5.20), we obtain: \(\hat{\theta}_n(T_j) - \hat{\theta}_n(T_j^*) = o_P(\log(n))\). Hence, \(\sqrt{n}\left(\hat{\theta}_n(T_j) - \hat{\theta}_n(T_j^*)\right) \overset{P}{\to} 0\) and it suffices to show that \(\sqrt{n}\left(\hat{\theta}_n(T_j^*) - \theta_j^*\right) \overset{d}{\to} N_d(0, F(\theta_j^*)^{-1}G(\theta_j^*)F(\theta_j^*)^{-1})\) to conclude.

For large value of \(n\), \(\hat{\theta}_n(T_j^*) \in \Theta(r)\). By the mean value theorem, there exists \((\hat{\theta}_{n,k})_{1 \leq k \leq d} \in [\hat{\theta}_n(T_j^*), \theta_j^*]\) such that

\[
\frac{\partial L_n(T_j^*, \hat{\theta}_n(T_j^*))}{\partial \theta_k} = \frac{\partial L_n(T_j^*, \theta_j^*)}{\partial \theta_k} + \frac{\partial^2 L_n(T_j^*, \hat{\theta}_{n,k})}{\partial \theta \partial \theta_k} (\hat{\theta}_n(T_j^*) - \theta_j^*).
\]

Let \(F_n = -2\left(\frac{1}{n_j} \frac{\partial^2 L_n(T_j^*, \hat{\theta}_{n,k})}{\partial \theta \partial \theta_k}\right)\) by Lemma 5.3 and Corollary 5.1, \(F_n \overset{a.s.}{\to} F(\theta_j^*)\) (where \(F(\theta_j^*)\) is defined by (5.3)). 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

\[
\sqrt{n_j}(\hat{\theta}_n(T_j^*) - \theta_j^*) = -2F_n^{-1}\left[\frac{1}{\sqrt{n_j}}\left(\frac{\partial L_n(T_j^*, \hat{\theta}_n(T_j^*))}{\partial \theta} - \frac{\partial L_n(T_j^*, \theta_j^*)}{\partial \theta}\right)\right].
\]

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

\[
\frac{1}{\sqrt{n_j}} \frac{\partial L_n(T_j^*, \theta_j^*)}{\partial \theta} \overset{d}{\to} N_d(0, G(\theta_j^*))
\]

where \(G(\theta_j^*)\) is given by (5.3). Thus, since \(\hat{\theta}_n(T_j^*, \hat{\theta}_n(T_j)) / \partial \theta = 0\), we have:

\[
\frac{1}{\sqrt{n_j}} \frac{\partial L_n(T_j^*, \hat{\theta}_n(T_j^*))}{\partial \theta} = \frac{1}{\sqrt{n_j}}\left(\frac{\partial L_n(T_j^*, \hat{\theta}_n(T_j^*))}{\partial \theta} - \frac{\partial L_n(T_j^*, \hat{\theta}_n(T_j))}{\partial \theta}\right) \overset{a.s.}{\to} 0.
\]

We conclude using Lemma 5.3 and the fact that \(1/\sqrt{n} = O(v_n/n)\).


