



**HAL**  
open science

# Contrast estimation of time-varying infinite memory processes

Jean-Marc Bardet, Paul Doukhan, Olivier Wintenberger

► **To cite this version:**

Jean-Marc Bardet, Paul Doukhan, Olivier Wintenberger. Contrast estimation of time-varying infinite memory processes. Stochastic Processes and their Applications, 2022. hal-02586009v3

**HAL Id: hal-02586009**

**<https://hal.science/hal-02586009v3>**

Submitted on 15 Jun 2021

**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.

# Contrast estimation of time-varying infinite memory processes

Jean-Marc Bardet <sup>a</sup>, Paul Doukhan <sup>b</sup>, Olivier Wintenberger <sup>c</sup>

<sup>a</sup> *University Paris 1 Panthéon-Sorbonne, SAMM, France, e1*

<sup>b</sup> *CY University, AGM, Cergy-Pontoise, France, e2*

<sup>c</sup> *Sorbonne University, LPSM, France, e3*

---

**Abstract** This paper aims at providing statistical guarantees for a kernel-based estimation of time-varying parameters driving the dynamic of infinite memory processes introduced by Doukhan and Wintenberger [1]. We then extend the results of Dahlhaus *et al.* [2] on local stationary Markov processes to other important models such as the GARCH model. The estimators are computed as localized M-estimators of any contrast satisfying appropriate regularity conditions. We prove the uniform consistency and the pointwise asymptotic normality of such kernel-based estimators. We apply our results to usual contrasts such as least-square, least absolute value, or quasi-maximum likelihood contrasts. Various time-varying models such as  $AR(\infty)$ ,  $ARCH(\infty)$  and  $LARCH(\infty)$  are considered. We discuss their approximation of locally stationary ARMA and GARCH models under contraction conditions. Numerical experiments demonstrate the efficiency of the estimators on both simulated and real data sets.

*Keywords:* [class=MSC2010] [Primary] 62G05 62G20 [secondary] 62M05

---

## 1. Introduction

Following the seminal paper of Dahlhaus [3] local-stationarity is considered as a natural set of conditions for introducing non-stationarity in times series. The chapter [4] of Dahlhaus is an exhaustive survey for new results between 1992 and 2012 on this topic. Dahlhaus and its co-authors have developed a consistent framework studying definitions and properties of local stationary models (See [5], [6] and [7] for instance) as well as related statistical issues such as identification and estimation (see [8], [9], [5], [6], [10] and [7]). Note that, except in Dahlhaus and Subba-Rao [5], the estimators introduced in the previous papers are based on a spectral approximation of the Gaussian likelihood, *i.e.* Whittle type estimators. Moreover, models considered in this early literature are linear filters of independent inputs. More recently a general approach based on derivative processes has been developed in an important paper by Dahlhaus *et al.* [2]. It allows to get rid off the linearity condition on the models. Time

non-homogeneous Markovian observations are the considered non-linear time-varying models. Under an adequate contraction condition on the kernel, the observations forget their past exponentially fast and the efficiency of kernel-based estimators is proven. Anyway, many important models like GARCH-type models are not taken into consideration in the above considered literature. The reason is that the Markov representation, adjoining the volatility process, is not observed and, like Hidden Markov Models, does not fall into the setting considered by Dahlhaus et al. [2]. Indeed such models may be considered as infinite memory processes, a general class that also models longer memories than the exponential decaying ones.

The current paper aims at extending the work of Dahlhaus et al. [2] to such an infinite memory setting. We extend the concept of local stationarity beyond the Markovian case. More precisely, the stationary models introduced by Doukhan and Wintenberger in [1] are extended to time-varying infinite memory causal processes defined as  $(X_t^{(n)})_{1 \leq t \leq n}$  a recursive solution to the equation

$$X_t^{(n)} = F_{\theta_t^{(n)}}((X_{t-k}^{(n)})_{k \in \mathbb{N}^*}; \xi_t), \quad 1 \leq t \leq n, \quad n \in \mathbb{N}^* = \mathbb{N} \setminus \{0\}. \quad (1)$$

where  $(\theta_t^{(n)})_{0 \leq t \leq n, n \in \mathbb{N}^*}$  is a deterministic family with  $\theta_t^{(n)} \in \Theta \subset \mathbb{R}^d$  for any  $0 \leq t \leq n$  and  $n \in \mathbb{N}^*$ ,  $F_{\theta}$  is a known real-valued function and the innovations  $\xi_t$  constitute an independent and identically distributed (i.i.d.) sequence. For ease of writing, we will consider  $X_t^{(n)} = 0$  for  $t \leq 0$ , but the arbitrary choice of any deterministic initial values does not change the asymptotic behavior. To make it tractable, a secondary aim of the paper is to keep the conditions as simple as possible. For instance, in our setting time-variation consists in the existence of a Hölder-continuous function  $\theta^*$  over  $u \in (0, 1)$  such that  $\theta_t^{(n)} = \theta^*(t/n)$ , see Assumption  $(\mathcal{H}(\rho))$ . Under this assumption, we define a kernel-based estimator  $\hat{\theta}(u)$  of  $\theta^*(u)$  obtained by the minimization of a localized sum of contrast  $\Phi$  (see its definition in (12)). We then establish the uniform consistency and the asymptotic normality of this estimator, which is minimax rate optimal, under sharp and general conditions.

The generality of the setting and the relative simplicity of the conditions allow us to recover existing results on several classes of examples and extend them to infinite memory processes. Clearly, any Markov process with a contractive kernel is an infinite memory model (with memory one if the observations are Markovian, with exponential decaying memory if there is a hidden state). The infinite memory representation for GARCH processes is quite appealing since it holds on the observations whereas the Markovian representation holds by adjoining the volatility process. As any contrast is a function of the observations only, the contrast itself has infinite memory but is not Markovian (see Bardet and Wintenberger [11] for a detailed discussion in the stationary setting). Considering quasi log-likelihood contrast offers consistent and efficient estimation for time-varying infinite memory processes: we obtain the uniform consistency and the asymptotic normality for time-varying AR( $\infty$ ) and ARCH( $\infty$ ). For finite memory time-varying ARMA( $p, q$ ) or ARCH( $p$ ), our results recover previous ones

of Dahlhaus and co-authors [8], [5] and [2], and extend them to the important GARCH model class. We also consider least squares and least absolute values contrasts for time-varying LARCH( $\infty$ ) processes and we notably obtained an efficient asymptotic estimation for these infinite memory processes.

Dahlhaus et al. [2] used functional dependence conditions. For time-varying infinite memory processes satisfying (1), we introduce the tangent process, a process that respects the infinite memory in (1) and "stationarizes" its future. The approach is appealing for its simplicity and generality. It is doable since in Eqn. (1) the time variation of the auto-regressive coefficient  $\theta_t^{(n)}$  is independent of the lag  $k$ . The model (1) is slightly different than the usual markovian local stationary setting where some auto-regressive coefficients  $\theta_t^{(n)}$  in the infinite memory representation may depend on the lag  $k$ . However, our approach coincides with the usual one for local stationary finite memory processes under additional contraction conditions. We apply our strategy to prove estimation convergence for local stationary ARMA and GARCH processes.

Numerical studies are also proposed. Firstly, Monte-Carlo experiments show the accuracy of the estimator in several cases of time-varying processes. However, these simulations also exhibit that such a kernel-based estimate requires sufficient large sample sizes (at least one thousand in many cases). Secondly, an application to financial data (the S&P500 data from October 1990 to October 2020) demonstrates the evolution of the parameters in case a GARCH(1,1)-model is used.

The forthcoming Section 2 is devoted to the definition and existence of new non-stationary models. In Section 3, the definition of the kernel-based estimator as well as its uniform consistency and asymptotic normality are stated, while Section 4 reviews several important cases. Numerical experiments are proposed in Section 5 and proofs are postponed in Sections 6 and 7.

## 2. Preliminaries

### 2.1. Notation

Some standard notation is used:

- The symbol  $0$  denotes any null vector in any vector space;
- If  $S$  is an arbitrary space  $S^\infty = \{(x_n)_{n \in \mathbb{N}} \in S^\mathbb{N}; \exists N \in \mathbb{N}, x_k = 0, \text{ for all } k > N\}$ ;
- The symbol  $\|\cdot\|$  denotes the usual Euclidean norm of a vector or the associated norm of a matrix;
- For  $p \geq 1$  and  $Z$  a random vector in  $\mathbb{R}^m$ , denote:  $\|Z\|_p = [\mathbb{E}(\|Z\|^p)]^{1/p}$ .
- For the measurable vector- or matrix-valued function  $g$  defined on some set  $U$ ,  $\|g\|_U = \sup_{u \in U} \|g(u)\|$ .

- From now on  $\Theta$  denotes a subset of  $\mathbb{R}^d$ , and  $\overset{\circ}{\Theta}$  is the interior of  $\Theta$ . If  $V$  is a Banach space then  $\mathcal{C}(\Theta, V)$  denotes the Banach space of  $V$ -valued continuous functions on  $\Theta$  equipped with the uniform norm  $\|\cdot\|_{\Theta}$  and  $\mathbb{L}^p(\mathcal{C}(\Theta, V))$  ( $p \geq 1$ ) denotes the Banach space of random a.e. continuous functions  $f$  such that  $\mathbb{E}[\|f\|_{\Theta}^p] < \infty$ .
- For  $\theta \in \Theta$  and  $\Psi_{\theta} : \mathbb{R}^{\infty} \rightarrow V$  a Borel function with values in a finite dimensional vector space  $V$ ,  $\partial_{\theta}^k \Psi_{\theta}(x)$  denotes respectively for  $k = 0, 1, 2$ , in case they exist,  $\Psi_{\theta}(x)$ ,  $\partial \Psi_{\theta}(x)/\partial \theta$  and  $\partial^2 \Psi_{\theta}(x)/\partial \theta^2$  for  $x \in \mathbb{R}^{\infty}$ .

## 2.2. Stationary infinite memory processes

In all the sequel, we will consider a given real number  $p \geq 1$ .

Set  $\theta \in \mathbb{R}^d$  and let a function  $F_{\theta}$  be defined as follows

$$F_{\theta} : (x, y) \in \mathbb{R}^{\infty} \times \mathbb{R} \mapsto F_{\theta}(x, y) \in \mathbb{R}. \quad (2)$$

Doukhan and Wintenberger proved in [1] the existence and the uniqueness of the stationary solution of the recurrence equation

$$X_t = F_{\theta}((X_{t-k})_{k \in \mathbb{N}^*}, \xi_t), \quad \text{for all } t \in \mathbb{Z}, \quad (3)$$

where  $(X_t)$  is a process with values in  $\mathbb{R}$  and where  $(\xi_t)_{t \in \mathbb{Z}}$  is a sequence of i.i.d. random variables (r.v.). This framework provides a parametric representation of models such as nonlinear autoregressive or conditionally heteroskedastic time series for instance.

The existence of a stationary solution in  $\mathbb{L}^p$  of the above equation relies on a contraction argument on the function  $F_{\theta}$ . As a consequence, we define the following family of assumptions  $(\mathbf{A}_k(\Theta))$  for  $k = 0, 1, 2$  and some compact subset  $\Theta$  of  $\mathbb{R}^d$ :

$(\mathbf{A}_k(\Theta))$  For  $\theta \in \Theta$ , we assume that the functions  $\partial_{\theta}^k F_{\theta}$  exist on  $\mathbb{R}^{\infty} \times \mathbb{R}$  for  $k = 0, 1, 2$ . Moreover, for any  $\theta \in \Theta$ , there exists a sequence  $(b_j^{(k)}(\theta))_j$  of nonnegative numbers such that for all  $x, y \in \mathbb{R}^{\infty}$

$$\bullet \quad C_k(\Theta) = \sup_{\theta \in \Theta} \|\partial_{\theta}^k F_{\theta}(0, \xi_0)\|_p < \infty \quad (4)$$

$$\bullet \quad \left\| \partial_{\theta}^k F_{\theta}(x, \xi_0) - \partial_{\theta}^k F_{\theta}(y, \xi_0) \right\|_p \leq \sum_{j=1}^{\infty} b_j^{(k)}(\theta) \|x_j - y_j\|_p, \quad (5)$$

$$\text{with } B_k(\Theta) = \left\| \sum_{j=1}^{\infty} b_j^{(k)} \right\|_{\Theta} < \infty.$$

Thus, from [1], under the uniform contraction conditions  $(\mathbf{A}_0(\Theta))$  with  $B_0(\Theta) < 1$ , there exists a unique stationary solution of (3) in  $\mathbb{L}^p$  (defined almost surely).

### 2.3. Time-varying infinite memory process

If we replace now  $\boldsymbol{\theta}$  by the time-varying  $\boldsymbol{\theta}_t^{(n)}$  such that  $\boldsymbol{\theta}_t^{(n)} \in \Theta$ , then the uniform contraction conditions  $(\mathbf{A}_0(\Theta))$  with  $B_0(\Theta) < 1$  ensure the existence of a non-stationary  $\mathbb{L}^p$ -process. More precisely, from a recursion we define the triangular array  $(X_t^{(n)})_{1 \leq t \leq n, n \in \mathbb{N}^*}$ :

$$X_t^{(n)} = F_{\boldsymbol{\theta}_t^{(n)}}((X_{t-k}^{(n)})_{k \in \mathbb{N}^*}; \xi_t), \quad 1 \leq t \leq n, n \in \mathbb{N}^*, \quad (6)$$

where  $(\boldsymbol{\theta}_t^{(n)})_{0 \leq t \leq n, n \in \mathbb{N}^*}$  is a family of real numbers  $\boldsymbol{\theta}_t^{(n)} \in \Theta \subset \mathbb{R}^d$  for any  $0 \leq t \leq n$  and  $n \in \mathbb{N}^*$ . To make this recursion possible we also need initial conditions:

$$X_t^{(n)} = 0, \quad \text{for } t \leq 0. \quad (7)$$

The solution of the above equations is no longer stationary. However, we can establish the following result (its proof as well as all the other ones are postponed in the Sections 6 and 7):

**Lemma 2.1.** *Let  $\Theta \subset \mathbb{R}^d$  such that  $(\mathbf{A}_0(\Theta))$  holds with  $B_0(\Theta) < 1$ . Then, under the assumption (7), the nonstationary triangular array  $(X_t^{(n)})_{0 \leq t \leq n, n \in \mathbb{N}^*}$ , solution of (6), remains in  $\mathbb{L}^p$  and it satisfies*

$$\sup_{n \in \mathbb{N}^*, 0 \leq s \leq n} \|X_s^{(n)}\|_p \leq \frac{C_0(\Theta)}{1 - B_0(\Theta)}.$$

### 2.4. The stationary approximation

We introduce a function  $u \mapsto \boldsymbol{\theta}^*(u)$  on  $[0, 1]$ ; this is a continuous time approximation of the triangular array of parameters  $(\boldsymbol{\theta}_t^{(n)})_{0 \leq t \leq n, n \in \mathbb{N}^*}$ . We consider  $\rho \in (0, 1]$  and will assume the following assumption:

**Assumption  $(\mathcal{H}(\rho))$ :** *There exist  $K_\theta > 0$  and a Hölder continuous function  $\boldsymbol{\theta}^* : [0, 1] \rightarrow \mathbb{R}^d$ , such as*

$$\|\boldsymbol{\theta}_t^{(n)} - \boldsymbol{\theta}^*(u)\| \leq K_\theta \left| u - \frac{t}{n} \right|^\rho, \quad \text{for any } n \in \mathbb{N}^* \text{ and } 1 \leq t \leq n. \quad (8)$$

This condition imposes an Hölder type behavior for  $\boldsymbol{\theta}_t^{(n)} = \boldsymbol{\theta}^*(t/n)$ ,  $0 \leq t/n \leq 1$ .

**Definition 2.1.** *If it exists, we define  $(\tilde{X}_t(u))_{t \in \mathbb{Z}}$  as any solution of the recursion*

$$\tilde{X}_t(u) = F_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{t-k}(u))_{k \geq 1}, \xi_t), \quad t \in \mathbb{Z}, \quad (9)$$

and we call it the stationary version of  $(X_t^{(n)})$  at  $u \in [0, 1]$ .

Note that from [1] the existence of the stationary version is inferred from the contraction assumptions. Namely, if the function  $\boldsymbol{\theta}^*$  satisfies  $\boldsymbol{\theta}^*(u) \in \Theta \subset \mathbb{R}^d$  for each  $u \in [0, 1]$  and if it is such that  $(\mathbf{A}_0(\Theta))$  holds with  $B_0(\Theta) < 1$ , then there exists a.s. a unique stationary stationary version  $(\tilde{X}_t(u))_{t \in \mathbb{Z}}$  satisfying (9) and

$$\sup_{t \in \mathbb{Z}} \|\tilde{X}_t(u)\|_p \leq \frac{C_0(\Theta)}{1 - B_0(\Theta)}, \quad u \in [0, 1].$$

### 3. M-estimation for infinite memory processes

#### 3.1. The stationary case

We recall the framework of contrast estimation for infinite memory chain as in Bardet and Wintenberger [11]. Let  $(X_t)$  be a stationary solution of the infinite memory model (3) with parameter  $\theta^* \in \Theta \subset \mathbb{R}^d$  such that  $(\mathbf{A}_0(\Theta))$  holds with  $B_0(\Theta) < 1$ . We estimate  $\theta^*$  using an M-estimator based on an observed path  $(X_1, \dots, X_n)$ .

We define a contrast function  $\Phi(x, \theta)$  that satisfies a set of regularity assumptions combined in the definition of the space  $\text{Lip}_p(\Theta)$  for  $\Theta \subset \mathbb{R}^d$  (always with  $1 \leq p$ ):

**Space  $\text{Lip}_p(\Theta)$ :** A Borel function  $h : \mathbb{R}^\infty \times \Theta \rightarrow \mathbb{R}$  belongs to  $\text{Lip}_p(\Theta)$  if there exists a sequence of non-negative numbers  $(\alpha_i(h, \Theta))_{i \in \mathbb{N}}$  where  $\sum_{s=1}^{\infty} \alpha_s(h, \Theta) < \infty$  and a function  $g : [0, \infty)^2 \rightarrow [0, \infty)$  such as for any sequences  $U = (U_i)_{i \in \mathbb{N}^*} \in (\mathbb{L}^p)^\infty$  and  $V = (V_i)_{i \in \mathbb{N}^*} \in (\mathbb{L}^p)^\infty$  satisfying  $\sup_{s \geq 1} \{\|U_s\|_p \vee \|V_s\|_p\} < \infty$ , one obtains:

$$\begin{cases} \mathbb{E}[\sup_{\theta \in \Theta} |h(0, \theta)|] < \infty; \\ \mathbb{E}[\sup_{\theta \in \Theta} |h(U, \theta) - h(V, \theta)|] \leq g(\sup_{s \geq 1} \{\|U_s\|_p \vee \|V_s\|_p\}) \sum_{s=1}^{\infty} \alpha_s(h, \Theta) \|U_s - V_s\|_p. \end{cases} \quad (10)$$

Note that if  $h \in \text{Lip}_{p'}(\Theta)$  then  $h \in \text{Lip}_p(\Theta)$  when  $p \leq p'$  thanks to Jensen's inequality. It is possible (see below the general case for non-stationary models) to prove that if  $\Phi \in \text{Lip}_p(\Theta)$  and if the stationary solution  $(X_t)$  admits finite  $p$  moments, then  $\Phi((X_{-t})_{t \in \mathbb{N}}, \theta)$  exists in  $\mathbb{L}^1$  for any  $\theta \in \Theta$ . The existence of first order moments is crucial for ensuring that  $\Phi$  is a proper score function which is implied by the following condition:

**Assumption  $(\mathbf{Co}(\Phi, \Theta))$ :** The function  $\Phi \in \text{Lip}_p(\Theta)$  for  $p \geq 1$  is such that for  $(X_t)_{t \in \mathbb{Z}}$  satisfying the infinite memory model (3) with parameter  $\theta^* \in \overset{\circ}{\Theta}$  and with  $\mathcal{F}_0 = \sigma((X_{-k})_{k \in \mathbb{N}})$ ,

$\theta^*$  is the unique minimum of the function

$$\theta \in \Theta \mapsto \mathbb{E}[\Phi((X_{1-k})_{k \in \mathbb{N}}, \theta) | \mathcal{F}_0] \text{ in } \overset{\circ}{\Theta}. \quad (11)$$

As a consequence, this condition is depending on the function  $\theta \in \Theta \mapsto F_\theta(\cdot)$  driving the infinite memory model (3). Therefore the contrast function  $\Phi$  will be chosen with respect to this function  $F_\theta$ . This is thus natural to define the M-estimator of  $\theta$  by

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{Argmin}} \frac{1}{n} \sum_{t=1}^n \Phi((X_{t-i})_{i \in \mathbb{N}}, \theta).$$

Note that the factor  $1/n$  aims at establishing a Law of Large Numbers. Indeed, if the almost sure convergence holds, *i.e.*

$$\sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \Phi((X_{t-i})_{i \in \mathbb{N}}, \boldsymbol{\theta}) - \mathbb{E}[\Phi((X_{-i})_{i \in \mathbb{N}}, \boldsymbol{\theta})] \right| \xrightarrow[n \rightarrow +\infty]{a.s.} 0,$$

then usual arguments imply  $\widehat{\boldsymbol{\theta}}_n \xrightarrow[n \rightarrow +\infty]{a.s.} \boldsymbol{\theta}^*$ .

### 3.2. The time-varying case

We extend the notion of contrast function  $\Phi$  to the non-stationary process  $(X_t^{(n)})_{t \in \mathbb{Z}}$  for time-varying parameters  $\boldsymbol{\theta}_t^{(n)}$ . The first step below is to prove the integrability.

**Lemma 3.1.** *Let  $(X_t^{(n)})_{t \in \mathbb{Z}}$  satisfy the non-stationary infinite memory model (6) under condition  $(\mathbf{A}_0(\Theta))$  with  $B_0(\Theta) < 1$  and let  $\Phi \in \text{Lip}_p(\Theta)$  with  $p \geq 1$ . Then for any  $\boldsymbol{\theta} \in \Theta$ , the sequence of contrasts  $(\Phi((X_{t-k}^{(n)})_{k \in \mathbb{N}}, \boldsymbol{\theta}))_{t \in \mathbb{Z}}$  exists in  $\mathbb{L}^1$ . Moreover, under Assumption  $(\mathcal{H}(\rho))$ , and with  $(\tilde{X}_t(u))_{t \in \mathbb{Z}}$  the stationary version defined in (9),  $(\Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta}))_{t \in \mathbb{Z}}$  is a stationary ergodic process.*

Under the Hölder assumption  $(\mathcal{H}(\rho))$ , we can expect estimating  $\boldsymbol{\theta}^*(u)$  with  $0 < u < 1$  defined in (8) thanks to a M-estimator based on the observations  $X_t^{(n)}$  for  $t/n \simeq u$ . The previous M-estimator has to be localized around  $t$  such as  $t \simeq nu$  using a convolution kernel  $K$  with a compact support (for simplicity):

**Definition 3.1.** *Let a kernel function  $K : \mathbb{R} \rightarrow \mathbb{R}$  be such as:*

- $K$  has a compact support, *i.e.* there exists  $c > 0$  such as  $K(x) = 0$  for  $|x| \geq c$ ;
- $K : \mathbb{R} \rightarrow \mathbb{R}$  is piecewise differentiable with

$$\int_{\mathbb{R}} K(x) dx = 1, \quad C_K = \sup_{x \in \mathbb{R}} |K(x)| < \infty.$$

Then, with a bandwidth sequence  $(h_n)_{n \in \mathbb{N}}$  of positive numbers, we define the kernel-based estimator of  $\boldsymbol{\theta}^*(u)$  as

$$\widehat{\boldsymbol{\theta}}(u) = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmin}} \frac{1}{nh_n} \sum_{j=1}^n \Phi((X_{j-i}^{(n)})_{i \in \mathbb{N}}, \boldsymbol{\theta}) K\left(\frac{j}{h_n} - u\right), \quad u \in (0, 1). \quad (12)$$

Under weak conditions, this estimator is consistent. For the uniform consistency, we require a stronger assumption:

$(\tilde{\mathbf{A}}_k(\Theta))$  For  $\boldsymbol{\theta} \in \Theta$ , we assume that the functions  $\partial_{\boldsymbol{\theta}}^k F_{\boldsymbol{\theta}}$  exist on  $\mathbb{R}^\infty \times \mathbb{R}$



for  $k = 0, 1$ . Moreover, for any  $\theta \in \Theta$ , there exists a sequence  $(b_j^{(k)}(\theta))_j$  of nonnegative numbers such that for all  $x, y \in \mathbb{R}^\infty$

$$\bullet \quad \tilde{C}_k(\Theta) = \left\| \left\| \partial_{\tilde{\theta}}^k F_{\theta}(0, \xi_0) \right\|_{\Theta} \right\|_p < \infty \quad (13)$$

$$\bullet \quad \left\| \left\| \partial_{\tilde{\theta}}^k F_{\theta}(x, \xi_0) - \partial_{\tilde{\theta}}^k F_{\theta}(y, \xi_0) \right\|_{\Theta} \right\|_p \leq \sum_{j=1}^{\infty} b_j^{(k)}(\theta) \|x_j - y_j\|_p, \quad (14)$$

with  $\tilde{B}_k(\Theta) = \sum_{j=1}^{\infty} b_j^{(k)}(\theta) < \infty$ .

Note that the assumptions  $(\tilde{\mathbf{A}}_k(\Theta))$  are stronger versions of the assumptions on the uniform condition  $(\mathbf{A}_k(\Theta))$ . Indeed we can consider without loss of much generality that  $B_k(\Theta) = \left\| \sum_{j=1}^{\infty} b_j^{(k)} \right\|_{\Theta}$  whereas  $\tilde{B}_k(\Theta) = \sum_{j=1}^{\infty} \|b_j^{(k)}\|_{\Theta}$ . In many examples of applications Assumptions  $(\mathbf{A}_0(\Theta))$  and  $(\tilde{\mathbf{A}}_0(\Theta))$  are both likely to hold together with the uniform contraction condition  $B_0(\Theta) < 1$ . However the strong uniform contraction condition  $\tilde{B}_0(\Theta) < 1$  is rather restrictive but is required only to derive uniform consistency.

**Theorem 3.1.** *Let  $(X_t^{(n)})_{t \in \mathbb{N}}$  be the solution of the non-stationary infinite memory model (6) which satisfies Assumption  $(\mathbf{A}_0(\Theta))$  with  $B_0(\Theta) < 1$  and Assumption  $(\mathbf{A}_1(\Theta))$  for  $p \geq 1$ , with also Assumption  $(\mathcal{H}(\rho))$ . For the same  $p \geq 1$  assume  $\Phi \in \text{Lip}_p(\Theta)$  with  $\sum_{s \geq 0} s \alpha_s(\Phi, \Theta) < \infty$  satisfying together the Assumption  $(\mathbf{Co}(\Phi, \Theta))$  and  $\left\| \sum_{t=1}^{\infty} t \log(t) b_t^{(0)} \right\|_{\Theta} < \infty$  then, for any  $u \in (0, 1)$ ,  $\hat{\theta}(u)$  consistently estimates  $\theta^*(u)$ :*

$$\hat{\theta}(u) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta^*(u), \quad \text{if } h_n \xrightarrow[n \rightarrow +\infty]{} 0 \quad \text{and} \quad nh_n \xrightarrow[n \rightarrow +\infty]{} \infty.$$

Moreover, if  $p > 1$ , Assumption  $(\tilde{\mathbf{A}}_0(\Theta))$  with  $\tilde{B}_0(\Theta) < 1$ , Assumption  $(\tilde{\mathbf{A}}_1(\Theta))$ ,  $\sum_{t=1}^{\infty} t \log(t) b_t^{(0)}(\Theta) < \infty$  and  $n^{1-1/p} h_n \xrightarrow[n \rightarrow +\infty]{} \infty$  hold then  $\hat{\theta}$  uniformly consistently estimates  $\theta^*$

$$\sup_{u \in [\varepsilon, 1-\varepsilon]} \left\| \hat{\theta}(u) - \theta^*(u) \right\| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \quad \varepsilon > 0. \quad (15)$$

**Remark 3.1.** *The uniform consistency of the kernel estimator was already obtained for Markov processes by Dahlhaus et al. in [2] under a different set of assumptions. Our extra condition on the bandwidth  $n^{1-1/p} h_n \rightarrow \infty$ ,  $n \rightarrow \infty$ , corresponds to the extra condition in (ii) of Theorem 5.2 of [2] with  $M = 0$ .*

**Remark 3.2.** *Consistency of  $\hat{\theta}(u)$  for any  $u \in (0, 1)$  can be achieved relaxing the uniform contraction condition  $B_0(\Theta) < 1$  of Theorem 3.1. Instead assume  $B_0(\Theta) < \infty$  and the pointwise contraction*

$$B_0(\{\theta^*(u)\}) < 1, \quad u \in (0, 1). \quad (16)$$

Then, dominated convergence argument and  $(\mathbf{A}_0(\Theta))$  yield, for any  $u \in (0, 1)$ , the existence of  $\varepsilon > 0$  such that

$$B_0(\{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*(u)\| \leq \varepsilon\}) < 1.$$

The  $M$ -estimator defined in (12) with  $\{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*(u)\| \leq \varepsilon\}$  in place of  $\Theta$  consistently estimates  $\boldsymbol{\theta}^*(u)$  by an application of Theorem 3.1. That this  $M$ -estimator coincides with  $\widehat{\boldsymbol{\theta}}(u)$  for  $n$  sufficiently large follows by a compactness argument of Pfanzagl [12]; The compact set  $\Theta$  is covered by finitely many compact sets of diameter  $\varepsilon$  among which only one contains  $\boldsymbol{\theta}^*(u)$ . Applying the SLLN of Lemma 7.1 1., the uniqueness in Condition  $(\mathbf{Co}(\Phi, \Theta))$  implies that the minimizer of (12) belongs to the unique compact set containing  $\boldsymbol{\theta}^*(u)$  for  $n$  large enough. See [11] for details.

**Remark 3.3.** We were not able to extend the uniform consistency of  $\widehat{\boldsymbol{\theta}}$  under  $B_0(\Theta) < 1$ , which is consistent with [2] that also worked under strong uniform contraction of the Markov kernel. The strong uniform contraction condition  $\widehat{B}_0(\Theta) < 1$  is restrictive but it seems necessary to prove the finiteness of uniform moments  $\left\| \sup_{u \in [0, 1]} |\widetilde{X}_t(u)| \right\|_p < \infty$  as in Lemma 6.1.

For establishing the asymptotic normality of  $\widehat{\boldsymbol{\theta}}(u)$ , we analogously need extra assumptions on the differentiability of the contrast  $\Phi$  and the integrability of its derivatives. We have:

**Theorem 3.2.** Let  $(X_t^{(n)})_{t \in \mathbb{N}}$  be the solution of the non-stationary infinite memory model (6) which satisfies Assumption  $(\mathbf{A}_0(\Theta))$  with  $B_0(\Theta) < 1$  and Assumption  $(\mathbf{A}_1(\Theta))$ ,  $p \geq 1$ , with also Assumption  $(\mathcal{H}(\rho))$ . For the same  $p \geq 1$  assume  $\Phi \in \text{Lip}_p(\Theta)$  with  $\sum_{s \geq 0} s \alpha_s(\Phi, \Theta) < \infty$  satisfying Assumption  $(\mathbf{Co}(\Phi, \Theta))$  and  $\|\sum_{t=1}^{\infty} t \log(t) b_t^{(0)}\|_{\Theta} < \infty$ . Assume that for any  $x \in \mathbb{R}^{\infty}$ ,  $\boldsymbol{\theta} \in \Theta \mapsto \Phi(x, \boldsymbol{\theta})$  is a  $\mathcal{C}^2(\Theta)$  function such as  $\partial_{\boldsymbol{\theta}} \Phi \in \text{Lip}_p(\Theta)$  with  $\sum_{s=1}^{\infty} s \alpha_s(\partial_{\boldsymbol{\theta}} \Phi, \Theta) < \infty$ . For any  $u \in (0, 1)$ , with  $\rho \in (0, 1]$  defined in Assumption  $(\mathcal{H}(\rho))$ , assume that

- $\mathbb{E}[\|\partial_{\boldsymbol{\theta}} \Phi((\widetilde{X}_k(u))_{k \leq 0}, \boldsymbol{\theta}^*(u))\|^2] < \infty$  and  $\Sigma(\boldsymbol{\theta}^*(u))$  is a definite positive matrix with

$$\begin{aligned} \Sigma(\boldsymbol{\theta}^*(u)) = \\ \int_{\mathbb{R}} K^2(x) dx \mathbb{E} \left[ \partial_{\boldsymbol{\theta}} \Phi((\widetilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta}^*(u)) (\partial_{\boldsymbol{\theta}} \Phi((\widetilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta}^*(u)))^{\top} \right]; \end{aligned}$$

- $\Gamma(\boldsymbol{\theta}^*(u)) = \mathbb{E}[\partial_{\boldsymbol{\theta}^2}^2 \Phi((\widetilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta}^*(u))]$  is a positive definite matrix.

If  $(h_n)_n$  is a sequence of positive numbers such that

$$nh_n \xrightarrow{n \rightarrow +\infty} \infty \quad \text{and} \quad nh_n^{1+2\rho} \xrightarrow{n \rightarrow +\infty} 0, \quad (17)$$

then, for any  $u \in (0, 1)$ ,

$$\sqrt{nh_n} (\widehat{\boldsymbol{\theta}}(u) - \boldsymbol{\theta}^*(u)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_d \left( 0, \Gamma^{-1}(\boldsymbol{\theta}^*(u)) \Sigma(\boldsymbol{\theta}^*(u)) \Gamma^{-1}(\boldsymbol{\theta}^*(u)) \right). \quad (18)$$

**Remark 3.4.** A first consequence of this result is that the convergence rate of  $\widehat{\boldsymbol{\theta}}(u)$  is  $o(n^{-\rho/(2\rho+1)})$ , which is just below the classical minimax convergence rate in a non parametric framework for any regularity  $\rho \in (0, 1]$ . Then the optimal choice of the bandwidth satisfies  $h_n = o(n^{-1/(2\rho+1)})$  and the estimator is also uniformly consistent whenever  $p > (2\rho + 1)/(2\rho)$ . Under additional conditions, Rosenblatt [13] and Dahlhaus et al. [2] derive expressions for an equivalent of the bias; in this case i.e.  $nh_n^{1+2\rho} \xrightarrow{n \rightarrow +\infty} \ell \neq 0$ , one may use the classical minimax bandwidth and, then, a non-centred Gaussian limit theorem occurs.

**Remark 3.5.** A precise inspection of the proof of Theorem 3.2 shows that only the consistency of  $\widehat{\boldsymbol{\theta}}(u)$ ,  $u \in (0, 1)$ , is used. Thus the asymptotic normality (18) can be extended under the pointwise contraction (16).

**Remark 3.6.** Considering  $(u_1, \dots, u_m)$  instead of  $u$ , a multidimensional central limit theorem could also be obtained extending (18). Such a result could be interesting for testing the goodness-of-fit ( $H_0 : \boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ ) or the stationarity ( $H_0 : \boldsymbol{\theta}^* = C_0 \in \mathbb{R}^d$ ) of the process. This will be the subject of a forthcoming paper.

#### 4. Examples

Here we develop several examples of time-varying infinite memory models with contrast functions  $\Phi \in \text{Lip}_p(\Theta)$  for which the Assumption  $(\mathbf{Co}(\Phi, \Theta))$  is satisfied. We also check the conditions of Theorems 3.1 and 3.2 in order to assert the uniform consistency and the asymptotic normality of the localized M-estimator.

##### 4.1. Time-varying AR(1) processes

In the case of time-varying AR(1) (denoted further as tv-AR(1)) processes defined by

$$X_t^{(n)} = \theta_t^{(n)} X_{t-1}^{(n)} + \xi_t, \quad \text{for } 1 \leq t \leq n, n \in \mathbb{N}^*, \quad (19)$$

with  $X_t^{(n)} = 0$  for any  $t \leq 0$ ,  $\theta_t^{(n)} = \theta^*(t/n)$  and  $\Theta = [-r, r]$  with  $0 < r < 1$ .

*Least Square contrast.*

When  $\Phi_{LS}$  is the Least Square (LS) contrast defined as

$$\Phi_{LS}(x, \theta) = (x_1 - \theta x_2)^2, \quad (20)$$

we obtain the usual Yule-Walker (or Least Square) estimator of  $\theta^*(u)$  if the stationary version  $(\tilde{X}_t(u))$  were observed. Clearly Assumption  $(\mathbf{Co}(\Phi, \Theta))$  holds and using Hölder Inequality, we obtain that  $\Phi_{LS} \in \text{Lip}_p(\Theta)$  with  $p = 2$ ,

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\theta \in \Theta} |\Phi_{LS}(U, \theta) - \Phi_{LS}(V, \theta)| \right] \\ & \leq (1+r) \max_{1 \leq s \leq 2} \{ \|U_s\|_2 \vee \|V_s\|_2 \} (\|U_1 - V_1\|_2 + r \|U_2 - V_2\|_2), \end{aligned}$$

and therefore  $\alpha_1(\Phi_{LS}, \Theta) = 1$ ,  $\alpha_2(\Phi_{LS}, \Theta) = r$  and  $\alpha_j(\Phi_{LS}, \Theta) = 0$  for  $j \geq 3$ . From basic calculation we also have

$$\partial_\theta \Phi_{LS}(x, \theta) = 2x_2(\theta x_2 - x_1) \quad \text{and} \quad \partial_{\theta^2}^2 \Phi_{LS}(x, \theta) = 2x_2^2.$$

After elementary algebra, we obtain:

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\theta \in [-r, r]} |\partial_\theta \Phi_{LS}(U, \theta) - \partial_\theta \Phi_{LS}(V, \theta)| \right] \\ & \leq 4 (\|U_1 - V_1\|_2 + \|U_2 - V_2\|_2) (\|U_1\|_2 + \|V_1\|_2 + \|U_2\|_2 + \|V_2\|_2) \end{aligned}$$

from Hölder's inequality. Analogously,

$$\mathbb{E} \left[ \sup_{\theta \in [-r, r]} |\partial_{\theta^2}^2 \Phi_{LS}(U, \theta) - \partial_{\theta^2}^2 \Phi_{LS}(V, \theta)| \right] \leq 2 (\|U_2 - V_2\|_2) (\|U_2\|_2 + \|V_2\|_2),$$

ensuring that  $\partial_\theta \Phi_{LS}$  and  $\partial_{\theta^2}^2 \Phi_{LS}$  are included in  $\text{Lip}_2([-r, r])$ . If  $\mathbb{E}[\xi_0^2] < \infty$ , both the matrix  $\Sigma(\theta^*(u)) = 4 \left( \int_{\mathbb{R}} K^2(x) dx \right) \sigma_\xi^4 (1 - \theta^*(u)^2)^{-1}$  and  $\Gamma(\theta^*(u)) = 2\sigma_\xi^2 (1 - \theta^*(u)^2)^{-1}$  are positive definite. Then, by an application of Theorem 3.2 we obtain:

**Corollary 4.1.** *If  $\mathbb{E}[\xi_0^2] < \infty$  and if  $\theta_t^{(n)} \in \Theta = [-r, r]$  satisfies Assumption  $(\mathcal{H}(\rho))$ , the localized least square estimator is asymptotically normal when the sequence  $(h_n)_n$  satisfies (17) and we obtain for any  $u \in (0, 1)$*

$$\sqrt{n h_n} (\hat{\theta}(u) - \theta^*(u)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N} \left( 0, (1 - \theta^*(u)^2) \int_{\mathbb{R}} K^2(x) dx \right).$$

Here, we recover for  $0 < \rho \leq 1$  the results on tv-AR(1) models obtained by Bardet and Doukhan in [14], which are also valid for functions  $u \in (0, 1) \rightarrow \hat{\theta}(u)$  with Hölderian derivatives.

*Least Absolute Value contrast.*

In the framework of tv-AR(1) processes (19) a classical alternative of the LS contrast, known for its robustness, is the Least Absolute Values (LAV) contrast defined as follows on  $\theta \in \Theta = [-r, r]$  with  $0 < r < 1$ ,

$$\Phi_{LAV}(x, \theta) = |x_1 - \theta x_2|. \quad (21)$$

If the stationary version  $(\tilde{X}_t(u))$  were observed, we obtain the usual estimator of  $\theta$  and Assumption  $(Co(\Phi_{LAV}, \Theta))$  holds. In such a case,  $\Phi_{LAV} \in \text{Lip}_p(\Theta)$  for any  $1 \leq p$ , and we obtain

$$\mathbb{E} \left[ \sup_{\theta \in \Theta} |\Phi_{LAV}(U, \theta) - \Phi_{LAV}(V, \theta)| \right] \leq \|U_1 - V_1\|_p + r \|U_2 - V_2\|_p.$$

implying  $\alpha_1(\Phi_{LAV}, \Theta) = 1$  and  $\alpha_2(\Phi_{LAV}, \Theta) = r$  and  $\alpha_j(\Phi_{LAV}, \Theta) = 0$  for  $j \geq 3$ . Since  $\Phi_{LAV}$  is not a differentiable function, we will restrict our purpose to the uniform consistency of  $\hat{\theta}(u)$ . We obtain the following result from Theorem 3.1:

**Corollary 4.2.** *If  $\|\xi_0\|_1 < \infty$  and if  $(\theta_t^{(n)}) \in \Theta = [-r, r]$  satisfies Assumption  $(\mathcal{H}(\rho))$ , then the localized LAV estimators  $\hat{\theta}_n$  satisfies (15).*

#### 4.2. Causal affine processes and Gaussian QMLE

We consider the general class of causal affine processes  $(X_t)$  defined by Bardet and Wintenberger in [11] as

$$X_t = M_{\boldsymbol{\theta}}((X_{t-i})_{i \geq 1}) \xi_t + f_{\boldsymbol{\theta}}((X_{t-i})_{i \geq 1}), \quad \text{for any } t \in \mathbb{Z}, \boldsymbol{\theta} \in \Theta \quad (22)$$

with  $\Theta$  a compact subset of  $\mathbb{R}^d$ . We assume the existence of Lipschitz coefficient sequences  $(\beta_i(f, \Theta))_{i \in \mathbb{N}}$  and  $(\beta_i(M, \Theta))_{i \in \mathbb{N}}$  such as for  $K_{\boldsymbol{\theta}} = f_{\boldsymbol{\theta}}$  or  $M_{\boldsymbol{\theta}}$ ,

$$\sup_{\boldsymbol{\theta} \in \Theta} |K_{\boldsymbol{\theta}}(x) - K_{\boldsymbol{\theta}}(y)| \leq \sum_{i=1}^{\infty} \beta_i(K, \Theta) |x_i - y_i|, \quad (23)$$

for any  $x, y \in \mathbb{R}^{\infty}$ . Then,  $(X_t)$  satisfies the infinite memory model (3) with  $F_{\boldsymbol{\theta}}(x, \xi_0) = f_{\boldsymbol{\theta}}(x) + \xi_0 M_{\boldsymbol{\theta}}(x)$  and  $(\mathbf{A}_0(\Theta))$  holds when  $\sum_j \beta_j(f, \Theta) < \infty$  and  $\sum_j \beta_j(M, \Theta) < \infty$  since we have

$$b_j^{(0)}(\Theta) \leq \beta_j(f, \Theta) + \|\xi_0\|_p \beta_j(M, \Theta) \quad \text{for any } j \in \mathbb{N}^*. \quad (24)$$

Therefore,  $(X_t)$  is a stationary and  $\mathbb{L}^p$  solution of the causal affine model (22) when

$$\sum_{j=1}^{\infty} \beta_j(f, \{\boldsymbol{\theta}\}) + \|\xi_0\|_p \beta_j(M, \{\boldsymbol{\theta}\}) < 1. \quad (25)$$

In such a case this is interesting to consider  $\Phi$  as  $(-2)$  times the Gaussian conditional log-density, inducing

$$\Phi_G(x, \boldsymbol{\theta}) = \log(M_{\boldsymbol{\theta}}^2((x_i)_{i \geq 2})) + \frac{(x_1 - f_{\boldsymbol{\theta}}((x_i)_{i \geq 2}))^2}{M_{\boldsymbol{\theta}}^2((x_i)_{i \geq 2})}. \quad (26)$$

The M-estimator resulting from this contrast is the Gaussian Quasi-Maximum Likelihood estimator (QMLE), notably used for estimating the parameters of GARCH processes, but also for ARMA, APARCH, ARMA-GARCH processes... As this was already done in [11] (proof of Theorem 1), under identifiability conditions on  $f_{\boldsymbol{\theta}}$  and  $M_{\boldsymbol{\theta}}$  (see below), Assumption  $(Co(\Phi_G, \Theta))$  holds. Moreover, using Bardet *et al.* [15] (proof of Lemma 6.3), we obtain with  $p = 3$  assuming the existence of  $\underline{M} > 0$  such as  $M_{\boldsymbol{\theta}} \geq \underline{M}$  and with  $C$  a constant real number

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in \Theta} |\Phi_G(U, \boldsymbol{\theta}) - \Phi_G(V, \boldsymbol{\theta})| \leq \\ & C(1 + |U_1|^2 + |V_1|^2 + f_{\boldsymbol{\theta}}^2((U_i)_{i \geq 2}) + f_{\boldsymbol{\theta}}^2((V_i)_{i \geq 2})) \\ & \times \left( |U_1 - V_1| + |f_{\boldsymbol{\theta}}((U_i)_{i \geq 2}) - f_{\boldsymbol{\theta}}((V_i)_{i \geq 2})| + |M_{\boldsymbol{\theta}}((U_i)_{i \geq 2}) - M_{\boldsymbol{\theta}}((V_i)_{i \geq 2})| \right) \end{aligned}$$

hence, there exists a function  $g$  such that

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} |\Phi_G(U, \boldsymbol{\theta}) - \Phi_G(V, \boldsymbol{\theta})| \right] \\ & \leq g \left( \sup_{i \geq 1} \{ \|U_i\|_3 \vee \|V_i\|_3 \} \right) \left( \|U_1 - V_1\|_3 + \sum_{i=2}^{\infty} b_k^{(0)}(\Theta) \|U_i - V_i\|_3 \right), \quad (27) \end{aligned}$$

using Hölder inequality and with  $b_k^{(0)}(\Theta) = \beta_k(f_{\boldsymbol{\theta}}, \Theta) + \|\xi_0\|_p \beta_k(M_{\boldsymbol{\theta}}, \Theta)$  the Lipschitz coefficients of the function  $F_{\boldsymbol{\theta}}$  given in  $(\tilde{\mathbf{A}}_0(\Theta))$ . Therefore, according to (10) with  $\alpha_k(\Phi_G, \Theta) = b_k^{(0)}(\Theta)$  for  $k \geq 2$  and  $\alpha_1(\Phi_G, \Theta) = 1$ , we check that  $\Phi_G \in \text{Lip}_3(\Theta)$  since  $(\mathbf{A}_0(\Theta))$  holds and  $\tilde{B}_0(\Theta) = \sum_k b_k^{(0)}(\Theta) < \infty$ . Now we consider a time-varying causal affine processes, that is the local stationary extension of causal affine processes defined in (22), *i.e.*

$$X_t^{(n)} = M_{\boldsymbol{\theta}_t^{(n)}}((X_{t-i}^{(n)})_{1 \leq i}) \xi_t + f_{\boldsymbol{\theta}_t^{(n)}}((X_{t-i}^{(n)})_{1 \leq i}), \quad \text{for any } t \in \mathbb{Z}, \quad (28)$$

with  $\boldsymbol{\theta}_t^{(n)} \in \Theta$  a compact set of  $\mathbb{R}^d$  and  $X_t^{(n)} = \mathbf{0}$  for  $t \leq 0$ .

In the sequel, we will provide general sufficient conditions for the asymptotic normality of  $\hat{\boldsymbol{\theta}}(u)$  in terms of the functions  $f_{\boldsymbol{\theta}}$  and  $M_{\boldsymbol{\theta}}$  and of their derivatives.

**Proposition 4.1.** *Let  $(X_t^{(n)})$  satisfy (28) where  $f_{\boldsymbol{\theta}}$ ,  $M_{\boldsymbol{\theta}}$ ,  $\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}$ ,  $\partial_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}}$ ,  $\partial_{\boldsymbol{\theta}^2}^2 f_{\boldsymbol{\theta}}$  and  $\partial_{\boldsymbol{\theta}^2}^2 M_{\boldsymbol{\theta}}$  satisfy Lipschitz inequalities (23) and under Assumption  $(\mathcal{H}(\rho))$ . Assume also:*

1.  $\|\xi_0\|_4 < \infty$  where the probability distribution of  $\xi_0$  is absolutely continuous with respect to the Lebesgue measure and  $\Theta$  is a bounded set included in

$$\left\{ \boldsymbol{\theta} \in \mathbb{R}^d, \sum_{j=1}^{\infty} (\beta_j(f_{\boldsymbol{\theta}}, \{\boldsymbol{\theta}\}) + \|\xi_0\|_4 \beta_j(M_{\boldsymbol{\theta}}, \{\boldsymbol{\theta}\})) < 1 \right\};$$

2. There exists  $\underline{M} > 0$  such as  $M_{\boldsymbol{\theta}} \geq \underline{M}$  for any  $\boldsymbol{\theta} \in \Theta$ ;
3. For all  $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ ,

$$(f_{\boldsymbol{\theta}} = f_{\boldsymbol{\theta}'} \quad \text{and} \quad M_{\boldsymbol{\theta}} = M_{\boldsymbol{\theta}'}) \implies \boldsymbol{\theta} = \boldsymbol{\theta}'; \quad (29)$$

4. We have

$$\begin{aligned} & \left( \sum_{j=1}^d \mu_j \frac{\partial}{\partial \boldsymbol{\theta}_j} f_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}) = 0 \quad \text{a.s.} \implies \mu_j = 0, j = 1, \dots, d \right), \\ \text{or} & \left( \sum_{j=1}^d \mu_j \frac{\partial}{\partial \boldsymbol{\theta}_j} M_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}) = 0 \quad \text{a.s.} \implies \mu_j = 0, j = 1, \dots, d \right). \end{aligned} \quad (30)$$

Consider  $\Phi = \Phi_G$  as  $(-2)$  times the Gaussian conditional log-density (26). Then, if

$$\begin{aligned} & \sum_{j=1}^{\infty} j (\log j (\beta_j(f, \Theta) + \beta_j(M, \Theta)) + j (\beta_j(\partial_{\boldsymbol{\theta}} f, \Theta))) \\ & + \sum_{j=1}^{\infty} (\beta_j(\partial_{\boldsymbol{\theta}} M, \Theta) + \beta_j(\partial_{\boldsymbol{\theta}^2}^2 f, \Theta) + \beta_j(\partial_{\boldsymbol{\theta}^2}^2 M, \Theta)) < \infty, \end{aligned} \quad (31)$$

with  $\beta_j(\cdot, \Theta)$  defined in (23), the localized QMLE  $\hat{\boldsymbol{\theta}}(u)$  satisfies the central limit theorem (18).

Note that the QML contrast  $\Phi$  depends on  $f_{\boldsymbol{\theta}}$  and  $M_{\boldsymbol{\theta}}$ . This explains why the asymptotic normality can be obtained from conditions on  $f_{\boldsymbol{\theta}}$  and  $M_{\boldsymbol{\theta}}$  and their derivatives. Note also that the conditions required in Proposition 4.1 are essentially the same than those requested in Theorem 2 of [11] in the stationary framework. The asymptotic normality of the (localized) QMLE holds under natural conditions, the main difference here is the convergence rates which is  $\sqrt{n}$  in the stationary case but  $\sqrt{nh_n}$  in the non-stationary one, which follows from localisation. The minimax rate is  $o(n^{1/3})$  is obtained for  $\rho = 1$  for local stationary causal affine models; it is smaller than the usual parametric rate  $O(\sqrt{n})$  achieved by the QMLE in the stationary case.

In the sequel we will state with details the assumptions for three important specific models, tv-AR( $\infty$ ), tv-ARCH( $\infty$ ) and tv-ARMA-GARCH models, which will imply for instance asymptotic normality of the contrast estimator in cases of locally stationary ARMA( $p, q$ ) or GARCH( $p, q$ ) processes.

*Time-varying AR( $\infty$ ) and locally-stationary ARMA( $p, q$ ) processes.*

In the case of time-varying AR( $\infty$ ) (or tv-AR( $\infty$ )), we directly have  $M_{\boldsymbol{\theta}} = \sigma(\boldsymbol{\theta}) > \underline{\sigma} > 0$  and  $f_{\boldsymbol{\theta}}((x_i)_{i \geq 1}) = \sum_{j=1}^{\infty} d_j(\boldsymbol{\theta}) x_j$ , where  $(d_j(\boldsymbol{\theta}))_{j \geq 1}$  is a sequence of real numbers, implying

$$X_t^{(n)} = \sigma(\boldsymbol{\theta}_t^{(n)}) \xi_t + \sum_{j=1}^{\infty} d_j(\boldsymbol{\theta}_t^{(n)}) X_{t-j}^{(n)}, \quad \text{for } 1 \leq t \leq n, n \in \mathbb{N}^*, \quad (32)$$

with  $X_t^{(n)} = 0$  for any  $t \leq 0$ .

Then we obtain the asymptotic normality of  $\widehat{\boldsymbol{\theta}}(u)$  from crude conditions on functions  $d_j$  and  $\sigma$  by an application of Proposition 4.1 since  $\beta_j(f, \boldsymbol{\theta}) = d_j(\boldsymbol{\theta})$  and  $\beta_j(M, \boldsymbol{\theta}) = 0$ :

**Corollary 4.3.** *Let  $(X_t^{(n)})$  be a tv-AR( $\infty$ ) process defined in (32). If  $\|\xi_0\|_4 < \infty$ , let  $\Theta$  be a bounded subset of  $\mathbb{R}^d$  included in  $\{\boldsymbol{\theta} \in \mathbb{R}^d, \sum_{j=1}^{\infty} d_j(\boldsymbol{\theta}) < 1\}$ . If for each  $j \in \mathbb{N}^*$  the functions  $\boldsymbol{\theta} \in \Theta \mapsto d_j(\boldsymbol{\theta}) \in \mathbb{R}$  and  $\boldsymbol{\theta} \in \Theta \mapsto \sigma(\boldsymbol{\theta}) \in [\underline{\sigma}, \infty)$  are  $\mathcal{C}^2(\Theta)$  functions such as  $(d_j(\boldsymbol{\theta}) = d_j(\boldsymbol{\theta}'), \forall j \in \mathbb{N}^*$  and  $\sigma(\boldsymbol{\theta}) = \sigma(\boldsymbol{\theta}'))$  imply  $(\boldsymbol{\theta} = \boldsymbol{\theta}')$ , if  $\boldsymbol{\theta}_t^{(n)}$  satisfies the assumption of local stationarity  $(\mathcal{H}(\rho))$ , and if*

$$\sum_{j=1}^{\infty} \{j \log j \|d_j\|_{\Theta} + j \|\partial_{\boldsymbol{\theta}} d_j\|_{\Theta} + \|\partial_{\boldsymbol{\theta}^2}^2 d_j\|_{\Theta}\} < \infty,$$

*then the central limit (18) holds for any  $u \in (0, 1)$  under condition (17).*

This result is new, essentially because it deals with two difficulties: an infinite memory and also with a non exponential decrease memory. As an illustrative example consider  $\boldsymbol{\theta} = (\mu, \kappa, \sigma)'$  and  $a_j(\boldsymbol{\theta}) = \mu j^{-\kappa}$  for any  $j \geq 1$ , with  $\kappa \geq \underline{\kappa} > 2$ ,  $\mu \leq (\sum_{j=1}^{\infty} j^{-\kappa})^{-1}$  and  $\sigma \geq \underline{\sigma} > 0$ . Then the previous corollary implies the asymptotic normality (18) of  $(\widehat{\mu}(u), \widehat{\kappa}(u), \widehat{\sigma}(u))$  under condition (17)

when  $(\mu_t^{(n)}, \kappa_t^{(n)}, \sigma_t^{(n)})$  satisfies Assumption  $(\mathcal{H}(\rho))$ .

An important subclass of tv-AR( $\infty$ ) models is the one of an invertible locally-stationary ARMA( $p, q$ ) models defined as

$$X_t^{(n)} + \phi_{1,t}^{(n)} X_{t-1}^{(n)} + \cdots + \phi_{p,t}^{(n)} X_{t-p}^{(n)} = \sigma_t^{(n)} \xi_t + \psi_{1,t}^{(n)} \xi_{t-1} + \cdots + \psi_{q,t}^{(n)} \xi_{t-q} \quad (33)$$

(for  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$ ), with  $X_t^{(n)} = 0$  for any  $t \leq 0$ , as it was introduced in [3]. More generally, using  $B$  the usual backward operator such as for  $x = (x_i)_{i \in \mathbb{Z}}$ ,  $Bx = y$  with  $y = (y_i)_{i \in \mathbb{Z}}$  and  $y_i = x_{i-1}$  for any  $i \in \mathbb{Z}$ , the following Lemma details the link between usual representation of ARMA( $\infty, \infty$ ) process and its affine causal one:

**Lemma 4.1.** *For  $\theta \in \Theta \subset \mathbb{R}^d$ , consider  $(a_i(\theta))_{i \in \mathbb{N}^*}$  and  $(b_j(\theta))_{j \in \mathbb{N}^*}$  two sequences of  $\Theta \rightarrow \mathbb{R}$  functions and both the following power series*

$$P_\theta(x) = 1 + \sum_{i=1}^{\infty} a_i(\theta) x^i \quad \text{and} \quad Q_\theta(x) = 1 + \sum_{j=1}^{\infty} b_j(\theta) x^j.$$

*Assume that  $\sum_{i=1}^{\infty} |a_i(\theta)| < 1$  and  $\sum_{j=1}^{\infty} |b_j(\theta)| < 1$ . Then if  $\xi = (\xi_t)_{t \in \mathbb{Z}}$  is a white noise,  $\sigma(\theta) > 0$ , we can define the stationary linear process  $X = (X_t)_{t \in \mathbb{Z}}$  such as*

$$\begin{aligned} P_\theta(B)X &= Q_\theta(B)(\sigma(\theta)\xi) \\ \iff X_t + \sum_{i=1}^{\infty} a_i(\theta) X_{t-i} &= \sigma(\theta)\xi_t + \sum_{j=1}^{\infty} b_j(\theta)\sigma(\theta)\xi_{t-j}, \end{aligned} \quad (34)$$

*for any  $t \in \mathbb{Z}$ . Then there exists a power series  $Q_\theta^{-1}(x) = 1 + \sum_{j=1}^{\infty} c_j(\theta)x^j$  with  $\sum_{j=1}^{\infty} |c_j(\theta)| < \infty$  such as  $X$  admits an affine causal representation*

$$\begin{aligned} X &= \sigma(\theta)\xi + (I - Q_\theta^{-1}(B)P_\theta(B))X \\ \iff X_t &= \sigma(\theta)\xi_t + \sum_{j=1}^{\infty} d_j(\theta)X_{t-j}, \forall t \in \mathbb{Z}, \end{aligned} \quad (35)$$

*where  $I$  is the identity operator and with  $d_j(\theta) = -\sum_{i=0}^j a_i(\theta)c_{j-i}(\theta)$ , where  $a_0(\theta) = c_0(\theta) = 1$  implying*

$$\sum_{j=1}^{\infty} |d_j(\theta)| \leq \frac{\sum_{j=1}^{\infty} |a_j(\theta)| + \sum_{j=1}^{\infty} |b_j(\theta)|}{1 - \sum_{j=1}^{\infty} |b_j(\theta)|}. \quad (36)$$

Thus, the Lipschitz coefficients satisfy  $\beta_j(f, \{\theta\}) = d_j(\theta)$  and  $\beta_j(M, \{\theta\}) = 0$ . Note also that for a usual stationary ARMA( $p, q$ ) process, the condition  $\sum_{i=1}^p |a_i| < 1$  implies that the roots of  $P$  are outside the unit disc, and the condition  $\sum_{i=1}^q |b_i| < 1$  implies the invertibility of the process.



However, in the case of locally-stationary ARMA( $p, q$ ) or of locally-stationary ARMA( $\infty, \infty$ ) processes (say LS-ARMA( $p, q$ ) or LS-ARMA( $\infty, \infty$ )), a straightforward extension of the stationary affine causal representation to a local stationary affine causal one does not exactly provide the expression of a tv-AR( $\infty$ ) process. For instance, the affine causal representation of the invertible MA(1) process  $X_t = \varepsilon_t + b_1 \varepsilon_{t-1}$  where  $\text{Var}(\varepsilon_0) = \sigma^2$  and  $|b_1| < 1$  is:

$$X_t = \sigma \xi_t - \sum_{k=1}^{\infty} (-b_1)^k X_{t-k}, \quad t \in \mathbb{Z}, \quad \text{so that its tv-representation is}$$

$$\implies X_t^{(n)} = \sigma^{(n)}(t) \xi_t - \sum_{k=1}^{\infty} (-b_1^{(n)}(t))^k X_{t-k}^{(n)}, \quad \text{if } 1 \leq t \leq n, n \in \mathbb{N}^*, \quad (37)$$

with  $X_t^{(n)} = 0$  for any  $t \leq 0$ .

But if we write the "usual" local-stationary (LS) definition  $\bar{X}_t^{(n)} = \sigma^{(n)}(t) \xi_t + b_1^{(n)}(t) \sigma^{(n)}(t-1) \xi_{t-1}$  for  $1 \leq t \leq n, n \in \mathbb{N}^*$ , we obtain the following LS affine causal representation

$$\bar{X}_t^{(n)} = \sigma^{(n)}(t) \xi_t - \sum_{k=1}^{\infty} (-1)^k \left( \prod_{j=1}^k b_1^{(n)}(t-j) \right) \bar{X}_{t-k} \quad (38)$$

as soon as  $\sup_{0 \leq t \leq n, n \in \mathbb{N}} |b_1^{(n)}(t)| < 1$ . Then (38) is slightly different to (37). Fortunately under Assumption  $(\mathcal{H}(\rho))$ , the difference  $\|X_t^{(n)} - \bar{X}_t^{(n)}\|_p$  can be bounded as a negligible term. This is done in the following property for general LS-ARMA( $\infty, \infty$ ) processes:

**Proposition 4.2.** *Let  $(\boldsymbol{\theta}_t^{(n)})$  be a family of  $\mathbb{R}^d$  vectors satisfying Assumption  $(\mathcal{H}(\rho))$  with  $0 < \rho \leq 1$ . Consider also  $(a_i(\boldsymbol{\theta}))_{i \in \mathbb{N}^*}$  and  $(b_i(\boldsymbol{\theta}))_{i \in \mathbb{N}^*}$  two sequences of  $\Theta \rightarrow \mathbb{R}$  differentiable functions, as well as the positive function  $\sigma(\boldsymbol{\theta})$ , such as*

$$\bar{a} = \left\| \sum_{j=1}^{\infty} |a_j| \right\|_{\Theta} < 1, \quad \bar{b} = \left\| \sum_{j=1}^{\infty} |b_j| \right\|_{\Theta} < 1,$$

$$\left\| \sum_{j=1}^{\infty} \|\partial_{\boldsymbol{\theta}} a_j\| \right\|_{\Theta} < \infty, \quad \text{and} \quad \left\| \sum_{j=1}^{\infty} \|\partial_{\boldsymbol{\theta}} b_j\| \right\|_{\Theta} < \infty. \quad (39)$$

Define the usual LS-ARMA( $\infty, \infty$ ) process by

$$\bar{X}_t^{(n)} + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta}_t^{(n)}) \bar{X}_{t-i}^{(n)} = \sigma(\boldsymbol{\theta}_t^{(n)}) \xi_t + \sum_{j=1}^{\infty} b_j(\boldsymbol{\theta}_t^{(n)}) \sigma(\boldsymbol{\theta}_t^{(n)}) \xi_{t-j} \quad (40)$$

for  $1 \leq t \leq n, n \in \mathbb{N}^*$  and  $\bar{X}_t^{(n)} = 0$  for  $t \leq 0$ . Consider also the tv-ARMA( $\infty, \infty$ ) process induced by the affine causal representations (35) with

$$\bar{d} = \left\| \sum_{j=1}^{\infty} |d_j| \right\|_{\Theta} < 1,$$

$$X_t^{(n)} = \sigma(\theta_t^{(n)}) \xi_t + \sum_{j=1}^{\infty} d_j(\theta_t^{(n)}) X_{t-j}^{(n)}, \quad (41)$$

for  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  and  $X_t^{(n)} = 0$  for  $t \leq 0$ . Then there exists  $C > 0$  such as for any  $n \in \mathbb{N}^*$ ,

$$\sup_{1 \leq t \leq n} \|\bar{X}_t^{(n)} - X_t^{(n)}\|_p \leq C n^{-\rho}. \quad (42)$$

Now consider the particular case of tv-ARMA( $p, q$ ) defined in (33) as a particular case of (40). For obtaining the tv-AR( $\infty$ ) representation (35) of this process, define the triangular array of parameters  $\theta_t^{(n)} = (\phi_{1,t}^{(n)}, \dots, \phi_{p,t}^{(n)}, \psi_{1,t}^{(n)}, \dots, \psi_{q,t}^{(n)}, \sigma_t^{(n)})'$  and the subset  $\Theta_{ARMA}^{(p,q)}$  of  $\mathbb{R}^d$  with  $d = p + q + 1$  defined by:

$$\Theta_{ARMA}^{(p,q)} = \left\{ (\phi_1, \dots, \phi_p, \psi_1, \dots, \psi_q, \sigma) \in (-1, 1)^{p+q} \times (0, \infty), \right. \\ \left. \sum_{i=1}^p |\phi_i| + 2 \sum_{j=1}^q |\psi_j| < 1 \right\}.$$

Note that the definition of  $\Theta_{ARMA}^{(p,q)}$  is an immediate application of (36) and is clearly not optimal.

Then if  $\theta_t^{(n)} \in \Theta$  for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  and  $\Theta$  a compact subset of  $\Theta_{ARMA}^{(p,q)}$ , we know that  $\sup_{n,t} \|\bar{X}_t^{(n)}\|_p < \infty$  for any  $p \geq 1$  when  $\|\xi_0\|_p < \infty$  since  $(\bar{X}_t^{(n)})$  satisfies (32) with a sum of absolute values of Lipschitz coefficients smaller than 1. Moreover, from classical analytic arguments it is well known that the corresponding Lipschitz coefficients  $\beta_j(f, \Theta)$ ,  $\beta_j(\partial_{\theta} f, \Theta)$  and  $\beta_j(\partial_{\theta}^2 f, \Theta)$  decrease exponentially fast so that the condition (31) is automatically satisfied.

Finally, as a consequence of Proposition 4.1 applied on  $(\bar{X}_t^{(n)})$ , using the bound (42) and  $\sqrt{n} h_n n^{-\rho} \xrightarrow{n \rightarrow +\infty} 0$ , because  $n h_n^{1+2\rho} \xrightarrow{n \rightarrow +\infty} 0$ , we obtain:

**Corollary 4.4.** *If  $(\bar{X}_t^{(n)})$  is a LS-ARMA( $p, q$ ) process defined in (33),  $\Theta$  is a bounded subset of  $\Theta_{ARMA}^{(p,q)}$ , if  $\|\xi_0\|_4 < \infty$  and  $\theta_t^{(n)}$  satisfies the assumption  $(\mathcal{H}(\rho))$ , which is implied by the Holder assumption  $(\mathcal{H}(\rho))$  on all functions  $\phi_{1,t}^{(n)}, \dots, \phi_{p,t}^{(n)}, \psi_{1,t}^{(n)}, \dots, \psi_{q,t}^{(n)}, \sigma_t^{(n)}$ , then the central limit (18) holds for any  $u \in (0, 1)$  under condition (17).*

This result is a QMLE version of the results obtained in full generality by Dahlhaus in [9] for Gaussian LS-ARMA processes (using Whittle likelihood approximation) and by Azrak and Melard in [16].

More generally, we also have:

**Corollary 4.5.** *If  $(\bar{X}_t^{(n)})$  is a LS-ARMA( $\infty, \infty$ ) process defined in (33) and satisfying the assumptions of Proposition 4.2. Then the central limit (18) holds for any  $u \in (0, 1)$  under condition (17).*

This result, which is a new estimation result satisfied by LS-ARMA( $\infty, \infty$ ) and therefore LS-AR( $\infty$ ) and LS-MA( $\infty$ ) processes, exhibits how our techniques and results are derived for general infinite memory processes.

*Time-varying ARCH( $\infty$ ) and locally-stationary GARCH( $p, q$ ) processes.*

Time-varying ARCH( $\infty$ ) (tv-ARCH( $\infty$ )) processes correspond to  $f_{\boldsymbol{\theta}}((x_i)_{i \geq 1}) = 0$  and  $M_{\boldsymbol{\theta}}((x_i)_{i \geq 1}) = (d_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} d_j(\boldsymbol{\theta}) x_j^2)^{1/2}$ , where  $(d_j(\boldsymbol{\theta}))_{j \geq 0}$  is a sequence of non negative real numbers and  $d_0(\cdot) \geq \underline{d} > 0$  implying

$$X_t^{(n)} = \xi_t \left( d_0(\boldsymbol{\theta}_t^{(n)}) + \sum_{i \geq 1} d_i(\boldsymbol{\theta}_t^{(n)}) (X_{t-i}^{(n)})^2 \right)^{1/2} \quad \text{for } 1 \leq t \leq n, n \in \mathbb{N}^*, \quad (43)$$

with  $X_t^{(n)} = 0$  for any  $t \leq 0$ ,  $\boldsymbol{\theta}_t^{(n)} \in \mathbb{R}^d$  for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  satisfying Assumption  $(\mathcal{H}(\rho))$ . For more details about stationary ARCH( $\infty$ ) or GARCH( $p, q$ ) processes, or for the transition from GARCH( $p, q$ ) to ARCH( $\infty$ ), see [17]. We are going to specify again the conditions of Proposition 4.1 in such a case. Firstly, we consider Lipschitz properties on  $((X_t^{(n)})^2)_t$  rather than  $(X_t^{(n)})_t$  as in [11] in order to recover natural constraint on the parameters set  $\Theta$  (similar as the moments condition of [18]) we consider Lipschitz properties of  $((X_t^{(n)})^2)_t$  inducing  $b_j^{(0)}(\boldsymbol{\theta}) = d_j(\boldsymbol{\theta})$  and  $\xi_t^2 = \xi_t^2$  and therefore for  $p = 2$  in this framework

$$\Theta \text{ is a compact subset of } \Theta_{ARCH(\infty)} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^d, \|\xi_0\|_4^2 \sum_{j=1}^{\infty} d_j(\boldsymbol{\theta}) < 1 \right\}. \quad (44)$$

Secondly, the Lipschitz coefficients of  $\Phi$ ,  $\partial_{\boldsymbol{\theta}} \Phi$  and  $\partial_{\boldsymbol{\theta}^2}^2 \Phi$  can be expressed in terms of  $|U_i^2 - V_i^2|$  following the same computations than in (27) and in the proof of Proposition 4.1. But since  $|U_i^2 - V_i^2| = |U_i - V_i| |U_i + V_i|$  and each time  $(M_{\boldsymbol{\theta}}((U_i)_{i \geq 1}) \times M_{\boldsymbol{\theta}}((V_i)_{i \geq 1}))^{-1}$  appears in the function  $g$ , we deduce that  $\Phi$ ,  $\partial_{\boldsymbol{\theta}} \Phi$  and  $\partial_{\boldsymbol{\theta}^2}^2 \Phi$  are respectively included in  $\text{Lip}_3(\Theta)$ ,  $\text{Lip}_4(\Theta)$  and  $\text{Lip}_4(\Theta)$  with coefficients  $\alpha_s(\cdot, \Theta)$  defined in (10) satisfying for  $s \geq 2$ ,

$$\begin{cases} \alpha_s(\Phi, \Theta) & = \|d_s\|_{\Theta}, \\ \alpha_s(\partial_{\boldsymbol{\theta}} \Phi, \Theta) & = \left\| d_s + \|\partial_{\boldsymbol{\theta}} d_s\| \right\|_{\Theta}, \\ \alpha_s(\partial_{\boldsymbol{\theta}^2}^2 \Phi, \Theta) & = \left\| d_s + \|\partial_{\boldsymbol{\theta}} d_s\| + \|\partial_{\boldsymbol{\theta}^2}^2 d_s\| \right\|_{\Theta}. \end{cases} .$$

From an application of Proposition 4.1 we obtain the asymptotic normality of  $\hat{\boldsymbol{\theta}}(u)$  from primitive conditions on functions  $d_j$ :

**Corollary 4.6.** *Let  $(X_t^{(n)})$  be a tv-ARCH( $\infty$ ) process defined in (43). We assume that  $\|\xi_0\|_4 < \infty$  and we consider  $\Theta$  a compact subset of  $\Theta_{ARCH(\infty)}$ . If*

for  $j \in \mathbb{N}^*$  the functions  $\boldsymbol{\theta} \in \Theta \mapsto d_j(\boldsymbol{\theta}) \in [0, \infty)$  and  $\boldsymbol{\theta} \in \Theta \mapsto d_0(\boldsymbol{\theta}) \in [\underline{d}, \infty)$  are  $\mathcal{C}^2(\Theta)$  functions such that  $(d_j(\boldsymbol{\theta}) = d_j(\boldsymbol{\theta}'), \forall j \in \mathbb{N})$  implies  $(\boldsymbol{\theta} = \boldsymbol{\theta}')$ , if  $\boldsymbol{\theta}_t^{(n)}$  satisfies the assumption of local stationarity ( $\mathcal{H}(\rho)$ ), and if

$$\sum_{j=1}^{\infty} j \log j \|d_j\|_{\Theta} + j \|\partial_{\boldsymbol{\theta}} d_j\|_{\Theta} + \|\partial_{\boldsymbol{\theta}^2}^2 d_j\|_{\Theta} < \infty,$$

then the localized QMLE  $\widehat{\boldsymbol{\theta}}(u)$  is asymptotically normal and (18) holds for any  $u \in (0, 1)$  and for any sequence  $(h_n)$  satisfying (17).

To the best of our knowledge, this result is new. In [5] the existence of tv-ARCH( $\infty$ ) processes has been studied and the asymptotic normality has been obtained in the case of the tv-ARCH( $p$ ) processes.

We used the previous result to the cases where  $(\overline{X}_t^{(n)})$  is a LS-GARCH( $\infty, \infty$ ) process and more particularly a LS-GARCH( $p, q$ ) process. We assume that  $\|\xi_0\|_4 < \infty$  and we consider a LS-GARCH( $\infty, \infty$ ) process defined by

$$\begin{cases} \overline{X}_t^{(n)} &= \overline{\sigma}_t^{(n)} \xi_t \\ (\overline{\sigma}_t^{(n)})^2 &= a_0(\boldsymbol{\theta}_t^{(n)}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}_t^{(n)}) (\overline{X}_{t-j}^{(n)})^2 + \sum_{j=1}^{\infty} b_j(\boldsymbol{\theta}_t^{(n)}) (\overline{\sigma}_{t-j}^{(n)})^2 \end{cases} \quad (45)$$

where  $(a_j(\cdot))_{j \geq 0}$  and  $(b_j(\cdot))_{j \geq 1}$  are two sequences of non negative functions and with  $\overline{X}_t^{(n)} = 0$  for any  $t \leq 0$ . As for LS-ARMA processes, consider  $P_{\boldsymbol{\theta}}(B) = \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) B^{j-1}$  and  $Q_{\boldsymbol{\theta}}(B) = \sum_{j=1}^{\infty} b_j(\boldsymbol{\theta}) B^{j-1}$  and the corresponding tv-GARCH( $\infty, \infty$ ) process induced by the affine causal representation of  $\overline{X}_t^{(n)}$  and defined by

$$X_t^{(n)} = \xi_t \sqrt{a_0(\boldsymbol{\theta}_t^{(n)}) (1 - Q_{\boldsymbol{\theta}_t^{(n)}}(1))^{-1} + P_{\boldsymbol{\theta}_t^{(n)}}(B) \sum_{k=0}^{\infty} Q_{\boldsymbol{\theta}_t^{(n)}}^k(B) (X_{t-k-1}^{(n)})^2}, \quad (46)$$

for  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  and  $X_t^{(n)} = 0$  for  $t \leq 0$ .

As a particular case, we also define the LS-GARCH( $p, q$ ) (see for instance [19])

$$\begin{cases} \overline{X}_t^{(n)} &= \overline{\sigma}_t^{(n)} \xi_t \\ (\overline{\sigma}_t^{(n)})^2 &= \omega_t^{(n)} + \sum_{j=1}^p \alpha_{j,t}^{(n)} (\overline{X}_{t-j}^{(n)})^2 + \sum_{j=1}^q \beta_{j,t}^{(n)} (\overline{\sigma}_{t-j}^{(n)})^2 \end{cases} \quad (47)$$

where  $\boldsymbol{\theta}_t^{(n)} = (\omega_t^{(n)}, \alpha_{1,t}^{(n)}, \dots, \alpha_{p,t}^{(n)}, \beta_{1,t}^{(n)}, \dots, \beta_{q,t}^{(n)})'$  is a vector of non-positive real numbers for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  and  $\omega_t^{(n)} \geq \underline{\omega} > 0$ . As for ARMA processes, the local stationarity version of the ARCH( $\infty$ ) representation of the stationary GARCH( $\infty, \infty$ ) with parameter  $\boldsymbol{\theta}_t^{(n)}$  is not exactly the same as the tv-ARCH( $\infty$ ) representation deduced from (47). As a consequence, we first prove the following proposition:

**Proposition 4.3.** *Let  $(\boldsymbol{\theta}_t^{(n)})$  be a family of  $\mathbb{R}^d$  vectors satisfying Assumption  $(\mathcal{H}(\rho))$  with  $0 < \rho \leq 1$ . Consider also  $(a_i(\boldsymbol{\theta}))_{i \in \mathbb{N}^*}$  and  $(b_i(\boldsymbol{\theta}))_{i \in \mathbb{N}^*}$  two sequences of  $\Theta \rightarrow \mathbb{R}$  differentiable functions, as well as the positive function  $\sigma(\boldsymbol{\theta})$ , such as*

$$\begin{aligned} \bar{a} = \left\| \sum_{j=1}^{\infty} |a_j| \right\|_{\Theta} < 1, \quad \bar{b} = \left\| \sum_{j=1}^{\infty} |b_j| \right\|_{\Theta} < 1, \\ \left\| \sum_{j=1}^{\infty} \|\partial_{\boldsymbol{\theta}} a_j\| \right\|_{\Theta} < \infty \quad \text{and} \quad \left\| \sum_{j=1}^{\infty} \|\partial_{\boldsymbol{\theta}} b_j\| \right\|_{\Theta} < \infty. \end{aligned}$$

Let  $(\bar{X}_t^{(n)})$  be defined following (45) where  $\boldsymbol{\theta}_t^{(n)}$  is a family of  $\mathbb{R}^d$  vectors satisfying Assumption  $(\mathcal{H}(\rho))$  with  $0 < \rho \leq 1$  together with  $\left\| \sum_{j=1}^{\infty} a_j \right\|_{\Theta} < 1$  and  $\left\| \sum_{j=1}^{\infty} b_j \right\|_{\Theta} < 1$ . Consider also  $(X_t^{(n)})_{t,n}$  the tv-GARCH( $\infty, \infty$ ) process induced by the affine causal representation (46). Then if Assumption (44) holds for any  $p \geq 2$  there exists  $C > 0$  such as for any  $n \in \mathbb{N}^*$ ,

$$\sup_{1 \leq t \leq n} \|\bar{X}_t^{(n)} - X_t^{(n)}\|_4 \leq C n^{-\rho} \quad \text{and} \quad \sup_{1 \leq t \leq n} \|(\bar{X}_t^{(n)})^2 - (X_t^{(n)})^2\|_2 \leq C n^{-\rho}. \quad (48)$$

Now since  $(X_t^{(n)})$  follows (45),  $\|\xi_0\|_4^2 \sum_{j=1}^{\infty} d_j(\boldsymbol{\theta}) < 1$  is implied by

$$\|\xi_0\|_4^2 \frac{\sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})}{1 - \sum_{j=1}^{\infty} b_j(\boldsymbol{\theta})} < 1.$$

Now in the special case of LS-GARCH( $p, q$ ) process, define the subset  $\Theta_{GARCH}^{(p,q)}$  of  $\mathbb{R}^{p+q+1}$  defined by:

$$\Theta_{GARCH}^{(p,q)} = \left\{ (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q) \in (0, \infty) \times [0, \infty)^{p+q}, \right. \\ \left. \sum_{j=1}^q \beta_j + \|\xi_0\|_4^2 \sum_{i=1}^p \alpha_i < 1 \right\}.$$

If  $\boldsymbol{\theta}_t^{(n)} \in \Theta$  for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  with  $\Theta$  a bounded set included in  $\Theta_{GARCH}^{(p,q)}$  then we have  $\sup_{t,n} \|\bar{X}_t^{(n)}\|_4 < \infty$  and the coefficients  $a_i(\boldsymbol{\theta}_t^{(n)})$  in its tv-ARCH( $\infty$ ) representation decrease exponentially fast. As  $\alpha_s(\Phi, \Theta)$ ,  $\alpha_s(\partial_{\boldsymbol{\theta}} \Phi, \Theta)$  and  $\alpha_s(\partial_{\boldsymbol{\theta}^2}^2 \Phi, \Theta)$  can be expressed from  $a_s(\cdot)$  and their derivatives, which are also exponentially decreasing. Finally, using the bound (48), and with again  $\sqrt{n} h_n n^{-\rho} \xrightarrow{n \rightarrow +\infty} 0$  implying  $n h_n^{1+2\rho} \xrightarrow{n \rightarrow +\infty} 0$ , we can replace the asymptotic behavior of  $(\bar{X}_t^{(n)})$  by the one of the corresponding tv-GARCH( $p, q$ ):

**Corollary 4.7.** Let  $(\bar{X}_t^{(n)})$  be a LS-GARCH( $p, q$ ) process defined in (47) and  $\Theta$  be a bounded set included in  $\Theta_{GARCH}^{(p, q)}$  where  $\|\xi_0\|_4 < \infty$ . If  $\theta_t^{(n)} = (\omega_t^{(n)}, \alpha_{1,t}^{(n)}, \dots, \alpha_{p,t}^{(n)}, \beta_{1,t}^{(n)}, \dots, \beta_{q,t}^{(n)})'$  satisfies the assumption of local stationarity  $(\mathcal{H}(\rho))$ , which is implied by the local stationarity  $(\mathcal{H}(\rho))$  on all functions  $\omega_t^{(n)}, \alpha_{1,t}^{(n)}, \dots, \alpha_{p,t}^{(n)}, \beta_{1,t}^{(n)}, \dots, \beta_{q,t}^{(n)}$ , then the central limit result (18) holds for any  $u \in (0, 1)$  under condition (17).

This result can be compared for instance with those of [5] for tv-ARCH( $p$ ), which are obtained under the same procedure but under the condition  $\|\xi_0\|_{4(1+\delta)} < \infty$ , or those of [19] for LS-GARCH( $p, q$ ), which are obtained from a local polynomial estimation and under the condition  $\|\xi_0\|_8 < \infty$ . Note that [20] also obtained asymptotic normality under very sharp conditions in a special case of tv-ARCH( $p$ ) process.

More generally, we also have:

**Corollary 4.8.** If  $(\bar{X}_t^{(n)})$  is a LS-GARCH( $\infty, \infty$ ) process defined in (45) and satisfying the assumptions of Proposition 4.3. Then the central limit (18) holds for any  $u \in (0, 1)$  under condition (17).

#### 4.3. Time-varying LARCH( $\infty$ ) processes and LS-contrast

Here we consider a LARCH( $\infty$ ) process introduced by Robinson in [21] and studied intensively by Giraitis *et al.* in [22]. The model is defined as

$$X_t = \xi_t \left( a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) X_{t-j} \right) \quad \text{for any } t \in \mathbb{Z}, \quad (49)$$

where  $\boldsymbol{\theta} \in \mathbb{R}^d$  and assume  $\|\xi_0\|_2 = 1$ . Assume also that  $j \in \mathbb{N}$ ,  $\boldsymbol{\theta} \in \mathbb{R}^d \mapsto a_j(\boldsymbol{\theta}) \in \mathbb{R}$  are continuous functions and without lose of generality assume  $a_0(\boldsymbol{\theta}) > 0$  for any  $\boldsymbol{\theta} \in \mathbb{R}^d$ . Moreover, for ensuring the stationarity of  $(X_t)$  and the existence of  $\|X_t\|_r$  with  $r \geq 1$ , assume that for any  $\boldsymbol{\theta} \in \Theta$ ,

$$\|\xi_0\|_r \sum_{j=1}^{\infty} |a_j(\boldsymbol{\theta})| < 1, \quad (50)$$

since  $b_j^{(0)}(\boldsymbol{\theta}) = |a_j(\boldsymbol{\theta})|$ . Even if a LARCH( $\infty$ ) process is an affine causal process, the Gaussian QML contrast can not be used for estimating  $\boldsymbol{\theta}$ . Indeed, the conditional variance of  $X_t$  can not be bounded around 0 and this does not allow asymptotic results for such contrasts (see more details in Francq and Zakoïan [23]). Even if the weighted least square estimators can also be defined (see [23]), we consider here the following ordinary LS contrast of square values: for  $x \in \mathbb{R}^\infty$ , define

$$\Phi_{LARCH}(x, \boldsymbol{\theta}) = \left( x_1^2 - \left( a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) x_{j+1} \right)^2 \right)^2. \quad (51)$$

If the stationary version  $(\tilde{X}_t(u))$  were observed, for any  $\boldsymbol{\theta} \in \Theta$  the score associated to the LS-contrast is

$$\begin{aligned} & \mathbb{E}[\Phi_{LARCH}((\tilde{X}_{1-k}(u))_{k \geq 0}, \boldsymbol{\theta}) \mid \mathcal{F}_0] \\ &= \mathbb{E}[|\xi_1|^4 - 1] \left( a_0(\boldsymbol{\theta}^*) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}^*(u)) \tilde{X}_{1-j}(u) \right)^4 \\ &+ \left( (a_0(\boldsymbol{\theta}^*(u)) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}^*(u)) \tilde{X}_{1-j}(u))^2 - (a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) \tilde{X}_{1-j}(u))^2 \right)^2. \end{aligned}$$

We notice that the first term at the right side of the last equality does not depend on  $\boldsymbol{\theta}$ . Then since  $a_0(\cdot)$  is supposed to be non negative, if we assume

$$\left( a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) X_{1-j} = a_0(\boldsymbol{\theta}') + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}') X_{1-j} \text{ a.s.} \right) \implies \boldsymbol{\theta} = \boldsymbol{\theta}', \quad (52)$$

then  $\mathbb{E}[\Phi_{LARCH}((X_{1-k})_{k \geq 0}, \boldsymbol{\theta}) \mid \mathcal{F}_0]$  has a unique minimum that is  $\boldsymbol{\theta}^*$  and Assumption  $(Co(\Phi_{LARCH}, \Theta))$  holds. Moreover, after computations and use of Hölder Inequalities, if  $r = 4$ ,

$$\begin{aligned} & \sup_{\boldsymbol{\theta} \in \Theta} |\Phi_{LARCH}(U, \boldsymbol{\theta}) - \Phi_{LARCH}(V, \boldsymbol{\theta})| \\ & \leq \left( U_1^2 + V_1^2 + (a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) U_{j+1})^2 + (a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) V_{j+1})^2 \right) \\ & \quad \times \left( |U_1 + V_1| |U_1 - V_1| + \left| 2a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) (U_{j+1} + V_{j+1}) \right| \sum_{j=1}^{\infty} |a_j(\boldsymbol{\theta})| |U_{j+1} - V_{j+1}| \right) \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} |\Phi_L(U, \boldsymbol{\theta}) - \Phi_L(V, \boldsymbol{\theta})| \right] & \leq g \left( \sup_{i \geq 1} \{ \|U_i\|_4 \vee \|V_i\|_4 \} \right) \\ & \quad \times \left( \|U_1 - V_1\|_4 + \sum_{j=2}^{\infty} \sup_{\boldsymbol{\theta} \in \Theta} |a_{j-1}(\boldsymbol{\theta})| \|U_j - V_j\|_4 \right), \end{aligned}$$

and therefore  $\Phi_{LARCH} \in \text{Lip}_4(\Theta)$  with  $\alpha_1(\Phi_{LARCH}, \Theta) = 1$ , and  $\alpha_k(\Phi_{LARCH}, \Theta) = \sup_{\boldsymbol{\theta} \in \Theta} |a_{k-1}(\boldsymbol{\theta})|$ , for  $k \geq 2$  and  $\sum_k \alpha_k(\Phi_{LARCH}, \Theta) < \infty$ , from (50).

We consider now the time-varying LARCH( $\infty$ ) process defined by:

$$X_t^{(n)} = \xi_t \left( a_0(\boldsymbol{\theta}_t^{(n)}) + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta}_t^{(n)}) X_{t-i}^{(n)} \right), \quad \text{for any } t \in \mathbb{Z}, \quad (53)$$

with  $\boldsymbol{\theta}_t^{(n)} \in \Theta$  a compact set of  $\mathbb{R}^d$  and  $X_t^{(n)} = 0$  for  $t \leq 0$ . We also assume that  $a_0(\cdot)$  is a non negative function. An application of Theorem 3.1 implies

that the localized LS estimator is uniformly consistent when  $r = 4$ . To assert the asymptotic normality, we assume  $\|\xi_0\|_8 < \infty$  and  $\Theta$  is a bounded subset of the set

$$\Theta_{LARCH} = \left\{ \boldsymbol{\theta} \in \mathbb{R}^d, \|\xi_0\|_8 \sum_{i=1}^{\infty} |a_i(\boldsymbol{\theta})| < \infty \right\}. \quad (54)$$

Then, using classical computations and Hausdorff Inequalities (see the proof), we obtain the asymptotic behavior of the estimator:

**Proposition 4.4.** *Assume that  $\boldsymbol{\theta} \in \mathbb{R}^d \mapsto a_j(\boldsymbol{\theta}) \in \mathbb{R}$  are  $\mathcal{C}^2$  functions for any  $j \in \mathbb{N}$ ,  $a_0(\cdot) \geq 0$  and*

1.  $\|\xi_0\|_8 < \infty$  where the probability distribution of  $\xi_0$  is absolutely continuous with respect to the Lebesgue measure and  $\Theta$  is a bounded set included in  $\Theta_{LARCH}$ .
2. For all  $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ ,

$$(a_i(\boldsymbol{\theta}) = a_i(\boldsymbol{\theta}'), \text{ for all } i \in \mathbb{N}) \implies (\boldsymbol{\theta} = \boldsymbol{\theta}'); \quad (55)$$

3. For all  $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$ ,

$$(\partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta}) = \partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta}'), \text{ for all } i \in \mathbb{N}) \implies (\boldsymbol{\theta} = \boldsymbol{\theta}'). \quad (56)$$

Let  $(X_t^{(n)})$  be a tv-LARCH process defined following (53) where  $\boldsymbol{\theta}_t^{(n)}$  satisfies Assumption  $(\mathcal{H}(\rho))$ . Consider  $\Phi = \Phi_{LARCH}$  as in (51). If

$$\sum_{j=1}^{\infty} j \log j \sup_{\boldsymbol{\theta} \in \Theta} |a_j(\boldsymbol{\theta})| + j \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}} a_j(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}^2}^2 a_j(\boldsymbol{\theta})\| < \infty, \quad (57)$$

then  $\widehat{\boldsymbol{\theta}}(u)$  is asymptotically normal as in (18) for any  $u \in (0, 1)$  and  $(h_n)$  satisfying (17).

To our knowledge, this result is new, even in its stationary  $\boldsymbol{\theta}_t^{(n)} = \boldsymbol{\theta}^* \in \mathbb{R}^d$  version. The particular case of locally-stationary GLARCH( $p, q$ ) (LS-GLARCH( $p, q$ )) processes, natural extension of stationary GLARCH( $p, q$ ) processes (see for instance [22]) is also interesting and straightforward:

**Corollary 4.9.** *If  $\|\xi_0\|_8 < \infty$  and the probability distribution of  $\xi_0$  is absolutely continuous with respect to the Lebesgue measure and if  $(X_t^{(n)})$  is a LS-GLARCH( $p, q$ ) process defined by*

$$X_t^{(n)} = \xi_t \sigma_t^{(n)}, \quad \text{with} \quad \sigma_t^{(n)} = c_{0,t}^{(n)} + \sum_{i=1}^p c_{i,t}^{(n)} X_{t-i}^{(n)} + \sum_{j=1}^q d_{j,t}^{(n)} \sigma_{t-j}^{(n)}, \quad \text{for any } t \in \mathbb{N}^*,$$

with  $X_t^{(n)} = 0$  for  $t \leq 0$  and where  $\boldsymbol{\theta}_t^{(n)} = (c_{0,t}^{(n)}, \dots, c_{i,t}^{(n)}, d_{j,t}^{(n)}, \dots, d_{j,t}^{(n)}) \in \Theta$ , with  $\Theta$  a compact set in

$$\left\{ (c_0, c_1, \dots, c_p, d_1, \dots, d_q) \in (0, \infty) \times \mathbb{R}^{p+q}, \sum_{i=1}^q |d_i| + \|\xi_0\|_8 \sum_{i=1}^p |c_i| < 1 \right\},$$



which satisfies the assumption of local stationarity ( $\mathcal{H}(\rho)$ ). Then the central limit (18) holds for any  $u \in (0, 1)$  under condition (17).

As for the case of LS-GARCH( $p, q$ ) processes, this result is obtained by considering the local stationary LARCH( $\infty$ ) process  $(\bar{X}_t^{(n)})_t$  of the LS-GLARCH( $p, q$ ) process  $(X_t^{(n)})_t$  and exactly as in Proposition 4.3 by establishing

$$\sup_{1 \leq t \leq n} \|\bar{X}_t^{(n)} - X_t^{(n)}\|_8 \leq C n^{-\rho}.$$

Then the asymptotic behavior of the estimator  $\hat{\theta}(u)$  can be applied to  $(\bar{X}_t^{(n)})_t$  instead of  $(X_t^{(n)})_t$  and is also obtained from the exponential decay of the sequences  $(\alpha_s(\Phi_{LARCH}, \Theta))_s$  and  $(\alpha_s(\partial_{\theta}\Phi_{LARCH}, \Theta))_s$  in such as case (see [22]).

## 5. Numerical experiments

In the sequel we are going to apply our kernel-based estimator in several different cases of local stationary processes.

The window bandwidth  $h_n$  is a tuning parameter that requires to be chosen. In order to neglect the bias we chose  $h_n = n^{-\lambda}$  with  $\lambda = 0.35$ , inducing  $n h_n^3 \xrightarrow{n \rightarrow +\infty} 0$ , which is the uniform consistency and the asymptotic normality condition required for  $Lip_p$ -contrast and  $\mathcal{C}^{\rho}$  functions when  $p > 3/2$  and  $\rho = 1$ .

### 5.1. Monte Carlo simulations

Here we will consider three cases:

1. An example of tv-GARCH(1, 1). Here, with the notation of equation (47), assume:

$$\omega_t^{(n)} = 1 + 0.5 \sin\left(5 \frac{t}{n}\right), \quad \alpha_{1,t}^{(n)} = 0.1 + 0.4 \cos^2\left(4 \frac{t}{n}\right) \quad \text{and} \quad \beta_{1,t}^{(n)} = 0.1 + 0.4 \frac{t}{n},$$

for any  $1 \leq t \leq n$  and  $n \in \mathbb{N}^*$ . Clearly,  $\omega^*(u) = 1 + 0.5 \sin(5u)$ ,  $\alpha_1^*(u) = 0.1 + 0.4 \cos^2(4u)$  and  $\beta_1^*(u) = 0.1 + 0.4u$ . Moreover, we assume that  $(\xi_t)$  is a sequence of i.i.d.  $\mathcal{N}(0, 1)$  random variables.

We independently replicated 1000 trajectories of this process, for  $n = 2000, 5000$  and  $10000$  and we computed the Gaussian QMLE estimators  $\hat{\omega}(u)$ ,  $\hat{\alpha}_1(u)$  and  $\hat{\beta}_1(u)$  for  $u = k/50$  with  $k = 1, \dots, 49$ . Finally, we used the two well known kernels, the uniform kernel  $U(x) = \frac{1}{2} \mathbb{1}_{x \in [-1, 1]}$  and the Epanechnikov one  $E(x) = \frac{3}{4} (1 - x^2) \mathbb{1}_{x \in [-1, 1]}$  and denote respectively  $\hat{\theta}^U(u)$  and  $\hat{\theta}^E(u)$ .

Table 1 contains the results of these Monte-Carlo experiments where we computed the square root of the mean integrated squared error (RSMISE).

| $n$   | $\omega$         |                  | $\alpha_1$         |                    | $\beta_1$         |                   |
|-------|------------------|------------------|--------------------|--------------------|-------------------|-------------------|
|       | $\hat{\omega}^U$ | $\hat{\omega}^E$ | $\hat{\alpha}_1^U$ | $\hat{\alpha}_1^E$ | $\hat{\beta}_1^U$ | $\hat{\beta}_1^E$ |
| 1000  | 0.493            | 0.455            | 0.126              | 0.122              | 0.230             | 0.208             |
| 3000  | 0.363            | 0.323            | 0.081              | 0.077              | 0.167             | 0.146             |
| 10000 | 0.259            | 0.224            | 0.052              | 0.048              | 0.118             | 0.101             |

Table 1: Square root of the MISE for tv-GARCH(1,1) processes for  $n = 1000, 3000$  and  $10000$  computed from 1000 independent replications.

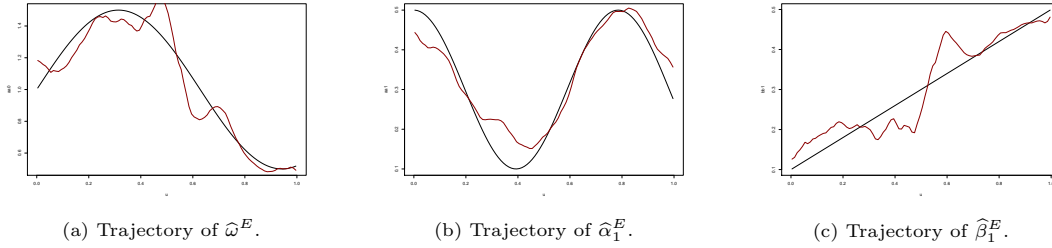


Figure 1: Paths of functions  $\omega, \alpha_1, \beta_1$  (in black), and a path of  $\hat{\omega}^E, \hat{\alpha}_1^E$  and  $\hat{\beta}_1^E$  (in red) for  $n = 10000$

In Figure 1 exhibits an example of particular trajectories of these estimators for  $n = 10000$ , while Figure 2 we also present the average trajectories of  $\hat{\omega}^E(u), \hat{\alpha}_1^E(u)$  and  $\hat{\beta}_1^E(u)$  when  $n = 5000$ .

2. A tv-ARCH( $\infty$ ) example. With the notation of equation (43), assume:

$$\boldsymbol{\theta} = (c_0, c_1, p), \quad d_0(\boldsymbol{\theta}) = c_0 \quad \text{and} \quad d_j(\boldsymbol{\theta}) = c_1 j^{-p} \quad \text{for } j \in \mathbb{N}^*$$

with  $c_{0,t}^{(n)} = 1 + 0.5 \sin\left(5 \frac{t}{n}\right)$ ,  $c_{1,t}^{(n)} = 0.1 + 0.5 \cos^2\left(4 \frac{j}{n}\right)$  and  $p_t^{(n)} = 2.1 + \frac{j}{n}$ .

for any  $1 \leq t \leq n$  and  $n \in \mathbb{N}^*$ . Therefore  $c_0^*(u) = 1 + 0.5 \sin(5u)$ ,  $c_1^*(u) =$

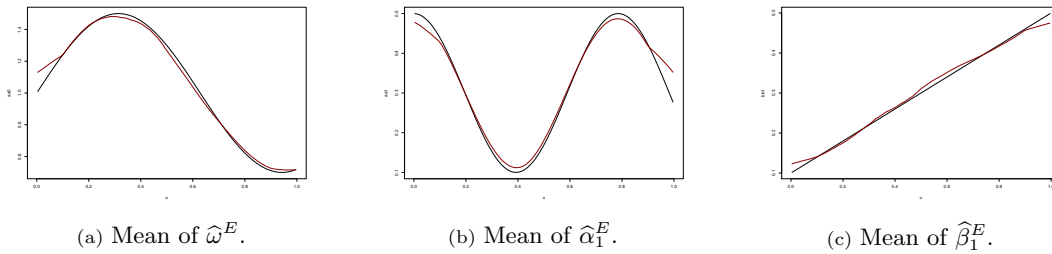
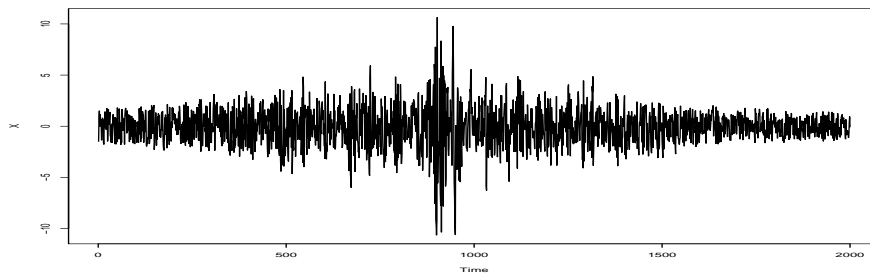


Figure 2: Paths of functions  $\omega, \alpha_1, \beta_1$  (in black), and the mean trajectories over 1000 replications of  $\hat{\omega}^E, \hat{\alpha}_1^E$  and  $\hat{\beta}_1^E$  (in red) for  $n = 5000$

Figure 3: On trajectory of a tv-ARCH( $\infty$ )-process for  $n = 2000$ 

| $n$   | $\hat{c}_0^U$ | $\hat{c}_0^E$ | $\hat{c}_1^U$ | $\hat{c}_1^E$ | $\hat{p}^U$ | $\hat{p}^E$ |
|-------|---------------|---------------|---------------|---------------|-------------|-------------|
| 1000  | 0.271         | 0.294         | 0.089         | 0.089         | 0.789       | 0.770       |
| 2000  | 0.227         | 0.261         | 0.066         | 0.069         | 0.712       | 0.696       |
| 5000  | 0.164         | 0.187         | 0.047         | 0.049         | 0.624       | 0.591       |
| 10000 | 0.128         | 0.144         | 0.037         | 0.039         | 0.555       | 0.517       |

Table 2: Square root of the MISE of  $\hat{c}_0^U$ ,  $\hat{c}_0^E$ ,  $\hat{c}_1^U$ ,  $\hat{c}_1^E$ ,  $\hat{p}^U$  and  $\hat{p}^E$  for the tv-ARCH( $\infty$ ) processes for  $n = 1000, 2000, 5000$  and  $10000$  computed from 1000 independent replications.

$0.1 + 0.5 \cos^2(4u)$  and  $p^*(u) = 2.1 + u$ . Moreover, we assume that  $(\xi_t)$  is a sequence of i.i.d.r.v. following  $\mathcal{U}([- \sqrt{3}, \sqrt{3}])$  (uniform) distribution.

As previously, we replicated 1000 trajectories of such process (see for instance one trajectory in Figure 3), for  $n = 2000, 5000$  and  $10000$  and computed the Gaussian QMLE estimators with Epanechnikov kernel  $\hat{c}_0^E(u)$ ,  $\hat{c}_1^E(u)$  and  $\hat{p}^E(u)$  for  $u = k/50$  with  $k = 1, \dots, 49$ .

Table 2 contains the results of these Monte-Carlo experiments where we computed the square root of mean integrated squared error (RSMISE).

## 5.2. Application to financial data

We apply our local kernel-based estimator to a trajectory of financial data. More precisely, we consider the log-returns of the daily closing values of S&P500 index between July 1999 and July 2019 (therefore  $n = 5031$ , see also Figure 4 for the graph of this trajectory). Many studies have shown that the GARCH(1,1) process is a relevant model for this type of data (We refer to the monograph of Francq and Zakoian [17] for more details). As a consequence, we used a tv-GARCH(1,1) process (see (47)) to take into account the changes in economic and financial conjectures over 20 years on such a model (think in particular

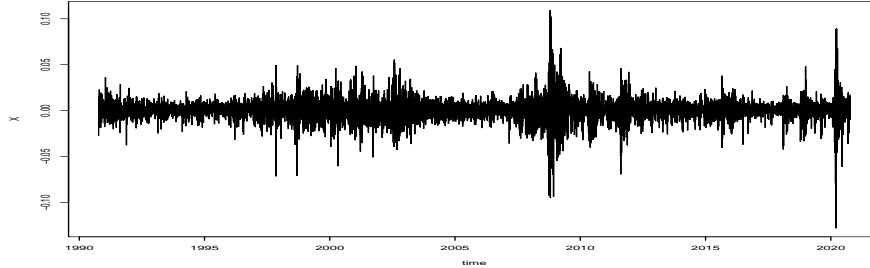


Figure 4: log-returns of daily closing values of S&P500 index between October 1990 and October 2020

of the September 2008 crisis and of the spring 2020 COVID crisis). Figure 5 exhibits the evolution of the three estimators computed with Uniform and Epanechnikov kernels, *i.e.*  $\hat{\omega}^U$ ,  $\hat{\alpha}_1^U$ ,  $\hat{\beta}_1^U$ ,  $\hat{\omega}^E$ ,  $\hat{\alpha}_1^E$ ,  $\hat{\beta}_1^E$  from Gaussian QMLE.

We draw the evolution of  $\alpha_1 + \beta_1$  in order to get a visual indicator of the variability of the S&P500 index. The larger  $\alpha_1 + \beta_1$  the worst the moment properties of the tv-GARCH(1, 1). The variability may be seen as an indicator of instability of the financial markets and thus of the crisis. Indeed the maximum of the  $\alpha_1 + \beta_1$  is achieved at the chore of the September 2008. More surprisingly, there is also a peak of variability as early as 2003. There the financial markets were renewing at their climate between 1998 and 2008 crisis. In order to distinguish between the two peaks of variability, one should observe that the curves of the coefficients  $\alpha_1$  and  $\beta_1$  separately. Then we observe that 2003 corresponds to a higher value for the coefficient  $\beta_1$  and 2008 to a higher value for the coefficient  $\alpha_1$ . We note that  $\beta_1$  is the coefficient of persistence in the volatility whereas  $\alpha_1$  transfers external shocks in the volatility. Finally, also surprisingly, the COVID crisis does not seem to change the last five years evolution of  $\alpha_1$  and  $\beta_1$ .

## 6. Moments properties of non stationary infinite memory processes

### 6.1. Proof of the properties of moments

*Proof of Lemma 2.1.* The proof of Lemma 2.1 follows from arguments extending those in [1] where more details may be found. From the iterative definition of  $X_t^{(n)}$  and under the assumption  $(\mathbf{A}_0(\Theta))$ , we obtain, for any  $n \in \mathbb{N}^*$  and  $0 \leq t \leq n$  :

$$\|X_t^{(n)} - F_{\theta_t^{(n)}}(0; \xi_t)\|_p \leq \sum_{s=1}^{\infty} b_s^{(0)}(\theta_t^{(n)}) \|X_{t-s}^{(n)}\|_p.$$

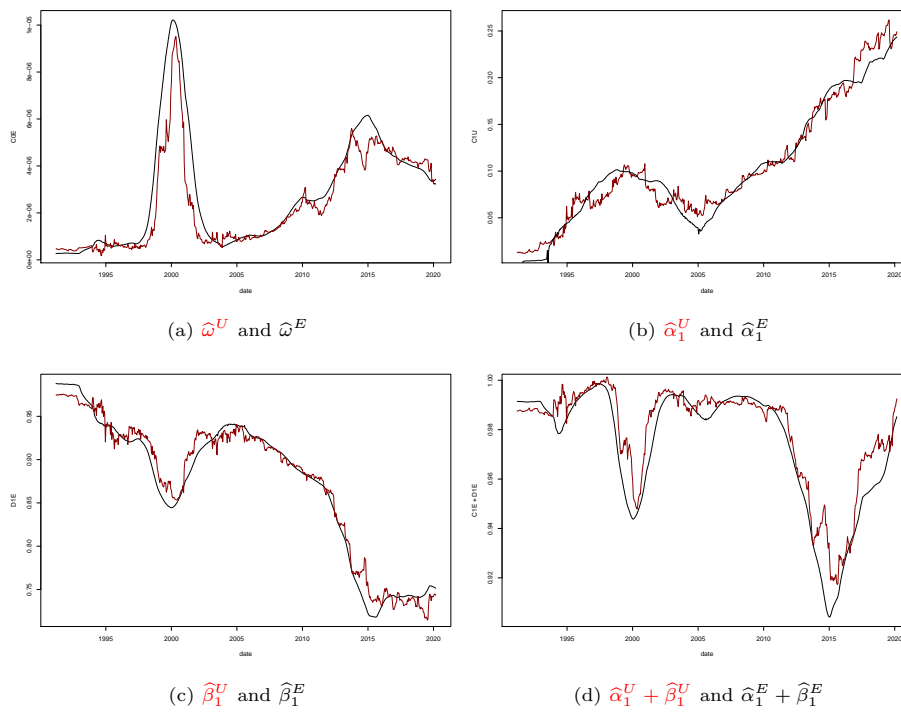


Figure 5: Estimators of  $\omega$ ,  $\alpha_1$ ,  $\beta_1$  and  $\alpha_1 + \beta_1$  for the log-returns of daily closing values of S&P500 index from October 1990 and October 2020

Thus from the triangle inequality we obtain

$$\begin{aligned} \|X_t^{(n)}\|_p &\leq \sum_{s=1}^{\infty} b_s^{(0)}(\boldsymbol{\theta}_t^{(n)}) \|X_{t-s}^{(n)}\|_p + \sup_{\boldsymbol{\theta} \in \Theta} \|F_{\boldsymbol{\theta}}(0, \xi_0)\|_p. \\ \|X_t^{(n)}\|_p &\leq B_0(\{\boldsymbol{\theta}_t^{(n)}\}) \max_{j < t} \|X_j^{(n)}\|_p + C_0(\Theta). \end{aligned} \quad (58)$$

A recursion entails with  $M_t = \max_{j \leq t} \|X_j^{(n)}\|_p$  that  $M_t \leq B_0(\Theta) M_{t-1} + C_0(\Theta)$  where  $0 \leq B_0(\Theta) < 1$ . Since  $M_0 = 0$ , for any  $0 \leq t \leq n$ ,

$$M_t \leq \frac{C_0(\Theta)}{1 - B_0(\Theta)} < \infty$$

and this achieves the proof.  $\square$

The next Lemma is a stronger moment inequality, useful only to prove the uniform consistency. It requires a uniform contraction condition stronger than  $B_0(\Theta) < 1$ , namely  $\tilde{B}_0(\Theta) < 1$ .

**Lemma 6.1.** *Let  $\Theta \subset \mathbb{R}^d$  be such that  $(\tilde{\mathbf{A}}_0(\Theta))$  holds with  $\tilde{B}_0(\Theta) < 1$  and assume that  $(\mathcal{H}(\rho))$  also holds. Then the stationary version  $(\tilde{X}_t(u))_{t \in \mathbb{Z}}$  solution of (9) satisfies*

$$\left\| \sup_{u \in [0,1]} |\tilde{X}_t(u)| \right\|_p \leq \frac{\tilde{C}_0(\Theta)}{1 - \tilde{B}_0(\Theta)}, \quad t \in \mathbb{Z}.$$

*Proof of Lemma 6.1.* We adapt here the fixed point approach in [24]. Consider the Banach space of random continuous functions  $H : [0, 1] \rightarrow \mathbb{R}$ ,  $\mathbb{L}^p(\mathcal{C}([0, 1], \mathbb{R}))$  with finite  $p$ th-moments equipped with the norm  $H \mapsto \|\sup_{u \in [0,1]} |H_u|\|_p$ . The underlying probability space is the one of the probability distribution of the iid sequence  $(\xi_t)_{t \in \mathbb{Z}}$ , i.e.  $H_u$  is a measurable function of  $(\xi_t)_{t \in \mathbb{N}}$  such that  $\mathbb{E}[\sup_{u \in [0,1]} |H_u((\xi_t)_{t \in \mathbb{N}})|^p] < \infty$ . We denote  $L$  the forward-lag operator on sequences  $(x_t)_{t \in \mathbb{N}}$  of  $\mathbb{R}^{\mathbb{N}}$  such that  $L((x_t)_{t \in \mathbb{N}}) = (x_{t+1})_{t \in \mathbb{N}}$  (quote that [24] use bidirectional sequences for non causal models and the current version is somehow simpler). Define:

$$\begin{aligned} \Phi : \mathbb{L}^p(\mathcal{C}([0, 1], \mathbb{R})) &\rightarrow \mathbb{L}^p(\mathcal{C}([0, 1], \mathbb{R})), \\ \Phi(H)(u) &= F_{\boldsymbol{\theta}^*(u)}((H_u \circ L^j)_{j \geq 0}, \pi_0), \quad \forall u \in [0, 1], \end{aligned}$$

with  $\pi_0$  the projection  $\pi_0((x_t)_{t \in \mathbb{N}}) = x_0$ . The continuity of  $u \mapsto \Phi(H)(u)$  follows from that of  $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}$  and  $u \mapsto \boldsymbol{\theta}^*(u)$  under  $(\tilde{\mathbf{A}}_0(\Theta))$  and  $(\mathcal{H}(\rho))$ . Finiteness of  $p$ th-moments for  $\sup_{u \in [0,1]} |\Phi(H)(u)|$  follows from similar arguments than in Lemma 1 of [24] since  $(\tilde{\mathbf{A}}_0(\Theta))$  holds uniformly wrt  $u \in [0, 1]$ . The Picard fixed point theorem applies to  $\Phi$  which is a contraction since  $\tilde{B}_0(\Theta) < 1$ . The existence of  $\tilde{X}_t(u)$  in the Banach space  $\mathbb{L}^p(\mathcal{C}([0, 1], \mathbb{R}))$  and the desired estimate on its norm follow.  $\square$

We note in passing that in any cases the stationary version  $(\tilde{X}_t(u))_{t \in \mathbb{Z}}$  is an ergodic process by its construction as a measurable function of the past of the innovations  $\tilde{X}_t(u) = H_u((\xi_{t-j})_{j \in \mathbb{N}})$ , see [25] for more details.

### 6.2. The tangent process

The tangent process is a stationarized version of the time-varying infinite memory process  $(X_t^{(n)})$  that respects the history of the original process. In order to localize, we define  $u \in [\varepsilon, 1 - \varepsilon]$ ,  $\varepsilon > 0$  and  $n$  large enough the quantities

$$1 \leq i_n(u) := [n(u - ch_n)] < j_n(u) := [n(u + ch_n)] \leq n, \quad (59)$$

where we recall that the compact support of the kernel  $K$  is included in  $[-c, c]$ .

**Definition 6.1.** *In the time-window  $\{i_n(u), i_n(u) + 1, \dots, j_n(u)\}$  we define*

$$X_t^*(u) = \begin{cases} X_t^{(n)}, & t < i_n(u), \\ F_{\theta^*(u)}((X_{t-k}^*(u))_{k \geq 1}, \xi_t), & i_n(u) \leq t \leq j_n(u). \end{cases} \quad (60)$$

As  $n$  is fixed and is sufficiently large in this section, for the ease of notation we omit the dependence on  $n$  concerning the quantities below, and especially for this process  $(X_t^*(u))_{t \in \mathbb{Z}}$ . We first prove its existence and uniform moment properties under the strong contraction condition  $\tilde{B}_0(\Theta) < 1$ . Note that its existence and non uniform moment properties extend easily under the weak contraction condition  $B_0(\Theta) < 1$  and  $(\mathbf{A}_0(\Theta))$ . We omit the proof of the latter case since it is similar but simpler than the following one.

**Lemma 6.2.** *Let  $\Theta \subset \mathbb{R}^d$  be such that  $(\tilde{\mathbf{A}}_0(\Theta))$  holds with  $\tilde{B}_0(\Theta) < 1$  and assume that  $(\mathcal{H}(\rho))$  also holds. Then, for any  $u \in (0, 1)$ , there exists a.s. a unique process  $(X_t^*(u))_{t \in \mathbb{Z}}$  satisfying (60) and there exists a positive constant  $C^* > 0$  such that*

$$\left\| \sup_{u \in [\varepsilon, 1 - \varepsilon]} |X_{i_n(u) + s}^*(u)| \right\|_p \leq C^* n^{1/p}, \quad \text{for all } 0 \leq s \leq 2cnh_n.$$

*Proof of Lemma 6.2.* We use a simple chaining argument. Denote  $u_k = (k + 1/2 + ch_n)/n$ , for  $k \in U_n(\varepsilon) = \{[\varepsilon n - ch_n - 1/2], \dots, [(1 - \varepsilon)n - ch_n - 1/2]\}$  a grid of points of the segment  $[\varepsilon, 1 - \varepsilon]$  and therefore  $\text{Card } U_n(\varepsilon) \simeq (1 - 2\varepsilon)n \leq n$ . Moreover, for each  $u \in [\varepsilon, 1 - \varepsilon]$  there exists  $u_k$  such that  $|u - u_k| \leq 1/(2n)$  and therefore  $i_n(u) = i_n(u_k)$  with  $i_n(u)$  defined in (59). Hence

$$\begin{aligned} \sup_{u \in [\varepsilon, 1 - \varepsilon]} |X_{i_n(u) + s}^*(u)| &\leq \max_{k \in U_n(\varepsilon)} |X_{i_n(u_k) + s}^*(u_k)| \\ &\quad + \sup_{u, v: i_n(u) = i_n(v)} |X_{i_n(u) + s}^*(u) - X_{i_n(v) + s}^*(v)|. \end{aligned} \quad (61)$$

Using  $\max(|x|, |y|) \leq |x| + |y|$  and an argument as in the proof of Lemma 2.1, we get that

$$\begin{aligned} \left\| \max_{k \in U_n(\varepsilon)} |X_{i_n(u_k)+s}^*(u_k)| \right\|_p &\leq \left\| \sum_{k \in U_n(\varepsilon)} |X_{i_n(u_k)+s}^*(u_k)| \right\|_p \\ &\leq \left( \sum_{k \in U_n(\varepsilon)} \|X_{i_n(u_k)+s}^*(u_k)\|_p^p \right)^{1/p} \leq n^{1/p} \frac{\tilde{C}_0(\Theta)}{1 - \tilde{B}_0(\Theta)}, \quad (62) \end{aligned}$$

since  $\|X_{i_n(u)+s}^*(u)\|_p \leq \tilde{C}_0(\Theta)/(1 - \tilde{B}_0(\Theta))$  for any  $u \in [\varepsilon, 1 - \varepsilon]$  and  $0 \leq s \leq 2cnh_n$ . Set  $\delta_n = \left\| \sup_{u,v: i_n(u)=i_n(v)} |X_{i_n(u)+s}^*(u) - X_{i_n(v)+s}^*(v)| \right\|_p$ . From an application of the chaining argument, we derive

$$\begin{aligned} \delta_n &\leq \left\| \sup_{u,v: i_n(u)=i_n(v)} |F_{\theta^*(i_n(u)+s)}((X_{i_n(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s}) \right. \\ &\quad \left. - F_{\theta^*(i_n(v)+s)}((X_{i_n(v)+s-k}^*(v))_{k \geq 1}, \xi_{i_n(v)+s}) \right\|_p \\ &+ \left\| \sup_{u,v: i_n(u)=i_n(v)} |F_{\theta^*(i_n(u)+s)}((X_{i_n(v)+s-k}^*(v))_{k \geq 1}, \xi_{i_n(v)+s}) \right. \\ &\quad \left. - F_{\theta^*(i_n(v)+s)}((X_{i_n(v)+s-k}^*(v))_{k \geq 1}, \xi_{i_n(v)+s}) \right\|_p \\ &\leq \sum_{k=1}^{\infty} b_k^{(0)}(\Theta) \left\| \sup_{u,v: i_n(u)=i_n(v)} |X_{i_n(u)+s-k}^*(u) - X_{i_n(v)+s-k}^*(v)| \right\|_p \\ &+ \sup_{u,v: i_n(u)=i_n(v)} \|\theta^*(u) - \theta^*(v)\| \left\| \sum_{k=1}^{\infty} b_k^{(1)}(\Theta) \right\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s-k}^*(u)| \left\|_p + \left\| \sup_{\theta \in \Theta} \|\partial_{\theta}^1 F_{\theta}(0; \xi_0)\| \right\|_p \\ &\leq \tilde{B}_0(\Theta) \left\| \sup_{u,v: i_n(u)=i_n(v)} |X_{i_n(u)+s-k}^*(u) - X_{i_n(v)+s-k}^*(v)| \right\|_p + \frac{K_{\theta}}{n^{\rho}} \left\| \sup_{\theta \in \Theta} \|\partial_{\theta}^1 F_{\theta}(0; \xi_0)\| \right\|_p \\ &+ K_{\theta} n^{-\rho} \tilde{B}_1(\Theta) \left( \frac{\tilde{C}_0(\Theta) n^{1/p}}{1 - \tilde{B}_0(\Theta)} + \left\| \sup_{u,v: i_n(u)=i_n(v)} |X_{i_n(u)+s-k}^*(u) - X_{i_n(v)+s-k}^*(v)| \right\|_p \right), \end{aligned}$$

from (8), (61) and (62). Collecting all those bounds for  $n$  sufficiently large in order that  $n^{-\rho}$  is sufficiently small, we get

$$\begin{aligned} &\left\| \sup_{u,v: i_n(u)=i_n(v)} |X_{i_n(u)+s}^*(u) - X_{i_n(v)+s}^*(v)| \right\|_p \\ &\leq \frac{K_{\theta} \tilde{B}_1(\Theta) \tilde{C}_0(\Theta)}{1 - \tilde{B}_0(\Theta)} n^{1/p-\rho} + K_{\theta} n^{-\rho} \left\| \sup_{\theta \in \Theta} \|\partial_{\theta}^1 F_{\theta}(0; \xi_0)\| \right\|_p \\ &\leq \frac{K_{\theta} \tilde{B}_1(\Theta) \tilde{C}_0(\Theta)}{1 - \tilde{B}_0(\Theta) - K_{\theta} n^{-\rho} \tilde{B}_1(\Theta)}. \end{aligned}$$

Finally, applying again the chaining argument we obtain

$$\left\| \max_{k \in U_n(\varepsilon)} |X_{i_n(u_k)+s}^*(u_k)| \right\|_p \leq n^{1/p} \frac{\tilde{C}_0(\Theta)}{1 - \tilde{B}_0(\Theta)} + O(n^{1/p-\rho}).$$



This ends the proof.  $\square$

We point out that  $(X_t^*(u))$  is not a copy of  $(X_t^{(n)})$  as it does not follow the same distribution. However this is a satisfactory non-stationary approximation of  $(X_t^{(n)})$  because we obtain the following moment bound:

**Lemma 6.3.** *Under Assumptions  $(\tilde{\mathbf{A}}_0(\Theta))$  with  $\tilde{B}_0(\Theta) < 1$ ,  $(\tilde{\mathbf{A}}_1(\Theta))$  and  $(\mathcal{H}(\rho))$ , with  $(X_t^*(u))$  the tangent process defined in (60), there exists a positive constant  $C' > 0$  such that*

$$\left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s}^{(n)} - X_{i_n(u)+s}^*(u)| \right\|_p \leq C' h_n^\rho n^{1/p}, \quad \text{for all } 0 \leq s \leq 2c n h_n. \quad (63)$$

Here again a similar non-uniform moment bound also holds under  $B_0(\Theta) < 1$ ,  $(\mathbf{A}_0(\Theta))$  and  $(\mathbf{A}_1(\Theta))$  following the same arguments than in the proof below.

*Proof of Lemma 6.3.* Define  $\Delta_s^* = 0$  for any  $s \leq 0$  and for  $0 \leq s \leq 2c n h_n$  set the quantity  $\Delta_s^* = \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s}^{(n)} - X_{i_n(u)+s}^*(u)|$ . For  $0 \leq s \leq 2c n h_n$  we decompose

$$\begin{aligned} \Delta_s^* &= \sup_{u \in [\varepsilon, 1-\varepsilon]} \left| F_{\boldsymbol{\theta}_{i_n(u)+s}^{(n)}} \left( (X_{i_n(u)+s-k}^{(n)})_{k \geq 1}, \xi_{i_n(u)+s} \right) \right. \\ &\quad \left. - F_{\boldsymbol{\theta}^*(u)} \left( (X_{i_n(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s} \right) \right| \\ &\leq \sup_{u \in [\varepsilon, 1-\varepsilon]} \left| F_{\boldsymbol{\theta}_{i_n(u)+s}^{(n)}} \left( (X_{i_n(u)+s-k}^{(n)})_{k \geq 1}, \xi_{i_n(u)+s} \right) \right. \\ &\quad \left. - F_{\boldsymbol{\theta}_{i_n(u)+s}^{(n)}} \left( (X_{i_n(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s} \right) \right| \\ &+ \sup_{u \in [\varepsilon, 1-\varepsilon]} \left| F_{\boldsymbol{\theta}_{i_n(u)+s}^{(n)}} \left( (X_{i_n(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s} \right) \right. \\ &\quad \left. - F_{\boldsymbol{\theta}^*(u)} \left( (X_{i_n(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s} \right) \right|. \end{aligned}$$

Then we derive

$$\begin{aligned}
\|\Delta_s^*\|_p &\leq \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} \left| F_{\boldsymbol{\theta}_{i_n(u)+s}}^{(n)} \left( (X_{t-k}^{(n)})_{k \geq 1}, \xi_t \right) \right. \right. \\
&\quad \left. \left. - F_{\boldsymbol{\theta}_{i_n(u)+s}}^{(n)} \left( (X_{i_n^*(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s} \right) \right| \right\|_p \\
&+ \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} \left| F_{\boldsymbol{\theta}_{i_n(u)+s}}^{(n)} \left( (X_{i_n^*(u)+s-k}^*(u))_{k \geq 1}, \xi_t \right) \right. \right. \\
&\quad \left. \left. - F_{\boldsymbol{\theta}^*(u)} \left( (X_{i_n^*(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s} \right) \right| \right\|_p \\
&\leq \left\| \sum_{k=1}^{\infty} b_k^{(0)}(\Theta) \sup_{u \in [\varepsilon, 1-\varepsilon]} \left| X_{i_n(u)+s-k}^{(n)} - X_{i_n^*(u)+s-k}^*(u) \right| \right\|_p \\
&+ \sup_{u \in [\varepsilon, 1-\varepsilon]} \left\| \boldsymbol{\theta}_{i_n(u)+s}^{(n)} - \boldsymbol{\theta}^*(u) \right\| \left\| \sup_{\boldsymbol{\theta} \in \Theta} \sup_{u \in [\varepsilon, 1-\varepsilon]} \left\| \partial_{\boldsymbol{\theta}}^1 F_{\boldsymbol{\theta}} \left( (X_{i_n^*(u)+s-k}^*(u))_{k \geq 1}, \xi_{i_n(u)+s} \right) \right\| \right\|_p \\
&\leq \left\| \sum_{k=1}^{\infty} b_k^{(0)}(\Theta) \Delta_{s-k}^* \right\|_p + \sup_{u \in [\varepsilon, 1-\varepsilon]} \left\| \boldsymbol{\theta}_{i_n(u)+s}^{(n)} - \boldsymbol{\theta}^*(u) \right\| \\
&\quad \times \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} \sum_{k=1}^{\infty} b_k^{(1)}(\Theta) \left| X_{i_n^*(u)+s-k}^*(u) \right| + \sup_{\boldsymbol{\theta} \in \Theta} \left\| \partial_{\boldsymbol{\theta}}^1 F_{\boldsymbol{\theta}}(0; \xi_{i_n(u)+s}) \right\| \right\|_p \\
&\leq \sum_{k=1}^{\infty} b_k^{(0)}(\Theta) \|\Delta_{s-k}^*\|_p + \sup_{u \in [\varepsilon, 1-\varepsilon]} \left\| \boldsymbol{\theta}_{i_n(u)+s}^{(n)} - \boldsymbol{\theta}^*(u) \right\| \\
&\quad \times \left( \sum_{k=1}^{\infty} b_k^{(1)}(\Theta) \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} \left| X_{i_n^*(u)+s-k}^*(u) \right| \right\|_p + \left\| \sup_{\boldsymbol{\theta} \in \Theta} \left\| \partial_{\boldsymbol{\theta}}^1 F_{\boldsymbol{\theta}}(0; \xi_0) \right\| \right\|_p \right)
\end{aligned}$$

by using the Assumptions  $(\tilde{\mathbf{A}}_0(\Theta))$  and  $(\tilde{\mathbf{A}}_1(\Theta))$ .

Now define  $M_t^* = \max_{s \leq t} \|\Delta_s^*\|_p$ . We derive from an application of Lemma 6.2 that for any  $0 \leq t \leq 2c n h_n$  it holds

$$\|\Delta_t^*\|_p \leq \tilde{B}_0(\Theta) M_{t-1}^* + \sup_{u \in [\varepsilon, 1-\varepsilon]} \left\| \boldsymbol{\theta}_{i_n(u)+t}^{(n)} - \boldsymbol{\theta}^*(u) \right\| \times \left( C^* n^{1/p} + \tilde{C}_1(\Theta) \right).$$

Remark that  $\sup_{u \in [\varepsilon, 1-\varepsilon]} \left\| \boldsymbol{\theta}_{i_n(u)+t}^{(n)} - \boldsymbol{\theta}^*(u) \right\| \leq c K_{\theta} h_n^{\rho}$  from condition (8) of Assumption  $(\mathcal{H}(\rho))$ . As a consequence, for any  $0 \leq t \leq 2c n h_n$ , we have

$$M_t^* \leq \tilde{B}_0(\Theta) M_{t-1}^* + c K_{\theta} \left( C^* n^{1/p} + \tilde{C}_1(\Theta) \right) h_n^{\rho}.$$

By definition,  $M_0^* = 0$ . Therefore we deduce for any  $0 \leq t \leq 2c n h_n$ ,

$$M_t^* \leq \frac{c K_{\theta}}{1 - \tilde{B}_0(\Theta)} \left( C^* n^{1/p} + \tilde{C}_1(\Theta) \right) h_n^{\rho}.$$

This completes the proof of the Lemma 6.3.  $\square$

Finally, the tangent process  $(X_t^*(u))_{t \in \mathbb{Z}}$  is also used for estimating the approximation of  $X_t^{(n)}$  with the stationary version  $\tilde{X}_t(u)$  for  $t/n \simeq u$ .

**Lemma 6.4.** *Under Assumptions  $(\tilde{\mathbf{A}}_0(\Theta))$  with  $\tilde{B}_0(\Theta) < 1$ ,  $(\tilde{\mathbf{A}}_1(\Theta))$  and  $(\mathcal{H}(\rho))$  there exists a positive constant  $C'' > 0$  such that*

$$\left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s}^{(n)} - \tilde{X}_{i_n(u)+s}(u)| \right\|_p \leq C'' n^{1/p} (h_n^\rho + \lambda_s), \quad \text{for all } 0 \leq s \leq 2cnh_n. \quad (64)$$

*Proof of Lemma 6.4.* Using the tangent process  $(X_t^*(u))$  defined in (60) and repeating the same arguments than above, we obtain a recursive relation on the sequence

$$\left( \max_{t \geq kr} \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |\tilde{X}_{i_n(u)+t}(u) - X_{i_n(u)+t}^*(u)| \right\|_p \right)_{k \geq 0}$$

given any  $r \in \mathbb{N}^*$ , using Lemma 6.2 and the estimates

$$\begin{aligned} \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |\tilde{X}_{i_n(u)+s}(u)| \right\|_p &\leq \left\| \max_{1 \leq t \leq n} \sup_{u \in [\varepsilon, 1-\varepsilon]} |\tilde{X}_t(u)| \right\|_p \\ &\leq \left( \sum_{1 \leq t \leq n} \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |\tilde{X}_t(u)| \right\|_p^p \right)^{\frac{1}{p}} \leq C n^{\frac{1}{p}}. \end{aligned}$$

We obtain

$$\left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s}^*(u) - \tilde{X}_{i_n(u)+s}(u)| \right\|_p \leq C n^{1/p} \lambda_s, \quad \text{for } 0 \leq s \leq 2cnh_n,$$

with  $\lambda_s = \inf_{1 \leq r \leq s} \left( B_0(\Theta)^{s/r} + \sum_{t=r+1}^{\infty} b_t^{(0)}(\Theta) \right)$  for  $s \geq 1$  defined as in Theorem 3.1 in [1]. Combining this result with (64) we bound the  $\mathbb{L}^p$  norm of the approximation of  $X_s^{(n)}$  with the stationary version  $\tilde{X}_s(u)$ , namely for all  $0 \leq s \leq 2cnh_n$ ,

$$\begin{aligned} \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s}^{(n)} - \tilde{X}_{i_n(u)+s}(u)| \right\|_p &\leq \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s}^{(n)} - X_{i_n(u)+s}^*(u)| \right\|_p \\ &\quad + \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+s}^*(u) - \tilde{X}_{i_n(u)+s}(u)| \right\|_p \leq C n^{1/p} (h_n^\rho + \lambda_s). \end{aligned}$$

□

## 7. Proofs for Sections 3 and 4

The section is organised as follows. § 7.1 proves the asymptotic behaviour of weighted partial sums of independent interest, and p Theorem 3.1 in the case of a step function kernel. § 7.2 includes the proof of the main Theorem 3.1 for general kernel functions. Then we prove Theorem 3.2. To this aim a specific additional approximation of the contrast, Proposition 7.1, is a last main step for those results. Then a last subsection § 7.2 proves our Proposition 4.4, useful for one essential example.

7.1. Some useful lemmas

*Proof of Lemma 3.1.* For  $\theta \in \Theta$ ,  $t \in \mathbb{Z}$  and  $m \in \mathbb{N}$ , define

$$\phi_{t,m} = \Phi(X_t^{(n)}, X_{t-1}^{(n)}, \dots, X_{t-m}^{(n)}, 0, 0, \dots; \theta).$$

As  $\Phi \in \text{Lip}_p(\Theta)$  the sequence  $(\phi_{t,m})_{m \in \mathbb{N}}$  is a Cauchy sequence in  $\mathbb{L}^q$  since for any  $m_2 > m_1$

$$\|\phi_{t,m_2} - \phi_{t,m_1}\|_1 \leq g \left( \sup_{0 \leq s \leq m_2} \|X_{t-s}^{(n)}\|_p \right) \sum_{k=m_1+1}^{m_2} \alpha_k(h, \Theta) \|X_{t-k}^{(n)}\|_p \quad (65)$$

$$\leq C \sum_{k=m_1+1}^{m_2} \alpha_k(h, \Theta) \quad (66)$$

from Lemma 2.1, since if  $s < 0$  then  $X_s = 0$ , thus the corresponding supremum bound extends over each  $s \leq n$ .

As  $\sum_{k=1}^{\infty} \alpha_k(h, \Theta) < \infty$  we deduce that for any  $\varepsilon > 0$ ,  $\sum_{k=m_1+1}^{m_2} \alpha_k(h, \Theta) \leq \varepsilon$  for  $m_1$  and  $m_2$  large enough. Using the completeness of  $\mathbb{L}^1$  we deduce the consistency of the sequence  $(\phi_{t,m})_{m \in \mathbb{N}}$  and the existence in  $\mathbb{L}^1$  of its limit  $\Phi((X_{t-k}^{(n)})_{k \geq 0}, \theta)$ .

When  $\theta_t^{(n)} = \theta^*(u)$  for any  $t, n$ , we consider  $\Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta)$  that also exists in  $\mathbb{L}^1$ . Moreover, as  $(\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}$  is a stationary ergodic process, this is also the case for  $(\Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta))_{t \in \mathbb{Z}}$  (see Corollary 2.1.3. in [25]).  $\square$

**Lemma 7.1.** *Let  $i_n(u)$  and  $j_n(u)$  defined in (59).*

1. *Let  $Z(u) = (Z_t(u))_{t \in \mathbb{N}}$  be a centered stationary process on a Banach space  $(\mathbb{B}, \|\cdot\|)$  for any  $0 \leq u \leq 1$ . If  $Z(u)$  is an ergodic process continuous with respect to  $u$  and satisfying  $\mathbb{E}[\sup_{0 \leq u \leq 1} \|Z_0(u)\|] < \infty$  then we have*

$$\sup_{0 < u < 1} \left\| \frac{1}{n h_n} \sum_{t=i_n(u)}^{j_n(u)} Z_t(u) K\left(\frac{\frac{t}{n} - u}{h_n}\right) \right\| \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \quad (67)$$

2. *Let  $Z = (Z_t)_{t \in \mathbb{N}}$  be a centered stationary process on  $\mathbb{R}^d$  such as  $\mathbb{E}[\|Z_0\|^2] < \infty$  and  $0 < u < 1$ . If  $\frac{1}{\sqrt{n}} \sum_{t=1}^n Z_t \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}_d(0, \Sigma)$  with  $\Sigma$  a positive definite symmetric matrix, then we have*

$$\frac{1}{\sqrt{n h_n}} \sum_{t=i_n(u)}^{j_n(u)} Z_t K\left(\frac{\frac{t}{n} - u}{h_n}\right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \left(\int_{\mathbb{R}} K^2(x) dx\right) \Sigma\right). \quad (68)$$

*Proof of Lemma 7.1.* Let  $\ell \in \mathbb{N}^*$ ,  $[-c, c]$  be the compact support of  $K$  and let  $n$  be large enough such as  $n(u - c h_n) \geq 1$  and  $nu + c n h_n \leq n$ . Then, for  $j \in$

$\{1, \dots, \ell\}$ , we denote  $I_j^{(\ell)} = [-c + 2c \frac{j-1}{\ell}, -c + 2c \frac{j}{\ell}]$ ,  $T_j^{(\ell)} = \left\{ t \in \mathbb{N}, \frac{t-u}{h_n} \in I_j \right\}$  and

$$S_n^{(\ell)}(u) = \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} Z_t(u) K\left(\frac{t-u}{h_n}\right) \quad \text{and} \quad S_{n,j}^{(\ell)}(u) = \frac{1}{nh_n} \sum_{t \in T_j^{(\ell)}} Z_t(u). \quad (69)$$

Below, we will omit to recall the dependence with respect to  $u$  when no confusion will be possible.

1. We notice that  $(Z_t(u))$  is a centered ergodic process on the Banach space  $\mathbb{L}^1(\mathcal{C}([0, 1], B))$ .

Thus for any fixed  $\ell \in \mathbb{N}^*$  such that  $\text{Card}(T_j^{(\ell)}) \simeq 2cnh_n/\ell \xrightarrow[n \rightarrow +\infty]{} \infty$  we apply the uniform ergodic theorem and since  $\mathbb{E}[Z_0(u)] = 0$  for any  $0 < u < 1$  we obtain

$$\sup_{0 \leq u \leq 1} \|S_{n,j}^{(\ell)}(u)\| \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \quad (70)$$

Denote  $t_j^{(\ell)} = -c + c \frac{2j-1}{\ell}$ , the midpoint of  $T_j^{(\ell)}$ , then we have

$$S_n^{(\ell)}(u) = \sum_{j=1}^{\ell} K\left(\frac{t_j^{(\ell)}}{h_n} - u\right) S_{n,j}^{(\ell)}(u) + \sum_{j=1}^{\ell} \frac{1}{nh_n} \sum_{t \in T_j^{(\ell)}} Z_t(u) \left[ K\left(\frac{t-u}{h_n}\right) - K\left(\frac{t_j^{(\ell)}}{h_n} - u\right) \right]. \quad (71)$$

Firstly, since  $K$  is a bounded function and from (70), then for any  $\ell \in \mathbb{N}^*$  we obtain

$$\sup_{0 \leq u \leq 1} \left\| \sum_{j=1}^{\ell} K\left(\frac{t_j^{(\ell)}}{h_n} - u\right) S_{n,j}^{(\ell)}(u) \right\| \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \quad (72)$$

Secondly, since  $K$  is a  $\mathcal{C}^1$  function on  $[-c, c]$  it holds

$$\sup_{0 \leq u \leq 1} \max_{1 \leq j \leq \ell} \sup_{t \in T_j^{(\ell)}} \left| K\left(\frac{t-u}{h_n}\right) - K\left(\frac{t_j^{(\ell)}}{h_n} - u\right) \right| \leq \frac{c}{\ell} \|K'\|_{\infty}$$

for any  $\ell \in \mathbb{N}^*$  and any  $n \in \mathbb{N}$ . Then we obtain

$$\begin{aligned} \sup_{0 \leq u \leq 1} \left\| \sum_{j=1}^{\ell} \frac{1}{nh_n} \sum_{t \in T_j^{(\ell)}} Z_t(u) \left[ K\left(\frac{t-u}{h_n}\right) - K\left(\frac{t_j^{(\ell)}}{h_n} - u\right) \right] \right\| \\ \leq \|K'\|_{\infty} \frac{c}{\ell} \cdot \frac{1}{nh_n} \sum_{t=1}^n \sup_{0 < u < 1} \|Z_t(u)\|. \end{aligned}$$

The ergodicity of  $(\sup_{0 < u < 1} \|Z_t(u)\|)_t$ , its stationarity, and  $\mathbb{E}[\sup_{0 < u < 1} \|Z_0(u)\|] < \infty$  together yield

$$\frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \sup_{0 < u < 1} \|Z_t(u)\| \xrightarrow[n \rightarrow +\infty]{a.s.} \mathbb{E} \left[ \sup_{0 < u < 1} \|Z_0(u)\| \right].$$

Thus, for any  $\varepsilon > 0$ , there exists a.s.  $(\ell_0, n_0)$  such as for any  $\ell \geq \ell_0$  and  $n \geq n_0$ ,

$$\sup_{0 < u < 1} \left\| \sum_{j=1}^{\ell} \frac{1}{nh_n} \sum_{t \in T_j^{(\ell)}} Z_t(u) \left[ K\left(\frac{t}{n} - u\right) - K\left(\frac{t_j^{(\ell)}}{n} - u\right) \right] \right\| \leq \varepsilon \quad a.s. \quad (73)$$

From (71), (72) and (73), we deduce (67).

2. Consider first  $K = K_\ell$  the piecewise constant function  $K_\ell(x) = \sum_{j=1}^{\ell} a_j \mathbb{1}_{x \in I_j^{(\ell)}}$ , and also assume first that  $d = 1$  with  $\Sigma = \sigma^2 > 0$  (unidimensional case).

$$S_n^{(\ell)} = \sum_{j=1}^{\ell} a_j S_{n,j}^{(\ell)}, \quad (74)$$

with  $S_{n,j}^{(\ell)}$  defined in (69). Using  $\text{Card}(T_j^{(\ell)}) \sim 2cnh_n/\ell \xrightarrow{n \rightarrow +\infty} \infty$ , for any  $j \in \{1, \dots, \ell\}$ ,

$$\left(\frac{\ell}{2cnh_n}\right)^{1/2} \sum_{t \in T_j^{(\ell)}} Z_t = \sqrt{\frac{nh_n\ell}{2c}} S_{n,j}^{(\ell)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2), \quad (75)$$

where  $\sigma^2 = \sum_{t \in \mathbb{Z}} \mathbb{E}[Z_0 Z_t]$  is such as  $0 < \sum_{t \in \mathbb{Z}} \mathbb{E}[Z_0 Z_t] < \infty$ . Moreover, using the stationarity of  $Z$ , we have for any  $j, j' \in \{1, \dots, \ell\}$  such as  $j \neq j'$ ,

$$\begin{aligned} nh_n \text{Cov}(S_{n,j}^{(\ell)}, S_{n,j'}^{(\ell)}) &= \frac{1}{nh_n} \sum_{t \in T_j^{(\ell)}} \sum_{t' \in T_{j'}^{(\ell)}} \mathbb{E}[Z_t Z_{t'}] = \frac{1}{nh_n} \sum_{t \in T_j^{(\ell)}} \sum_{t' \in T_{j'}^{(\ell)}} \mathbb{E}[Z_0 Z_{t'-t}] \\ &\implies \left| nh_n \text{Cov}(S_{n,j}^{(\ell)}, S_{n,j'}^{(\ell)}) \right| \leq C \left| \sum_{k > T_j^{(\ell)}} \mathbb{E}[Z_0 Z_k] \right| \xrightarrow[n \rightarrow +\infty]{} 0, \end{aligned}$$

as soon as  $\ell = o(nh_n)$  since  $\sum_{t \in \mathbb{Z}} |\mathbb{E}[Z_0 Z_t]| < \infty$ . Hence a central limit theorem holds for any linear combinations of  $S_{n,j}^{(\ell)}$ . E.g. for  $S_n^{(\ell)}$  defined in (74),

$$\begin{aligned} \sqrt{\frac{nh_n\ell}{2c}} S_n^{(\ell)} &\xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \sum_{j=1}^{\ell} a_j^2 \sum_{t \in \mathbb{Z}} \mathbb{E}[Z_0 Z_t]\right) \\ &\implies \sqrt{nh_n} S_n^{(\ell)} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \sigma^2 \int_{\mathbb{R}} K_\ell^2(x) dx\right), \quad (76) \end{aligned}$$

since for piecewise kernel we have  $\int_{\mathbb{R}} K_\ell^2(x) dx = \frac{2c}{\ell} \sum_{j=1}^{\ell} a_j^2$ .

Consider now a general piecewise differentiable kernel  $K$  and denote  $K_\ell$  such as  $K_\ell(x) = \sum_{j=1}^{\ell} a_j \mathbb{1}_{x \in I_j}$  for  $x \in \mathbb{R}$  with  $a_j = K(-c + c \frac{2j-1}{\ell})$  such that

$$\int_{\mathbb{R}} K_\ell^2(x) dx \xrightarrow{\ell \rightarrow +\infty} \int_{\mathbb{R}} K^2(x) dx.$$

The result will follow from Theorem 3.2 in [26]. For this we check that for any  $\varepsilon > 0$

$$\lim_{\ell \rightarrow \infty} \limsup_{n \rightarrow \infty} \mathbb{P}\left(\left|S_n^{(\ell)} - S_n\right| \geq \varepsilon/\sqrt{nh_n}\right) = 0, \quad \text{with } S_n = \frac{1}{nh_n} \sum_{t=1}^n Z_t K\left(\frac{\frac{t}{n} - u}{h_n}\right).$$

From Markov inequality

$$\mathbb{P}\left(\left|S_n^{(\ell)} - S_n\right| \geq \varepsilon/\sqrt{nh_n}\right) \leq nh_n \frac{\mathbb{E}[\Delta_n^2]}{\varepsilon^2}$$

with

$$\begin{aligned} \Delta_n &= \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} Z_t K\left(\frac{\frac{t}{n} - u}{h_n}\right) - \sum_{j=1}^{\ell} a_j S_{n,j}^{(\ell)} \\ &= \frac{1}{nh_n} \sum_{j=1}^{\ell} \sum_{t \in T_j} \underbrace{\left(K\left(\frac{\frac{t}{n} - u}{h_n}\right) - K\left(-c + 2c \frac{j-1}{\ell}\right)\right)}_{a_{t,j}(\ell)} Z_t. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{E}[\Delta_n^2] &= \frac{1}{(nh_n)^2} \sum_{j=1}^{\ell} \sum_{t \in T_j} \sum_{j'=1}^{\ell} \sum_{t' \in T_{j'}} a_{t,j}(\ell) a_{t',j'}(\ell) \mathbb{E}[Z_t Z_{t'}] \\ &\leq \frac{1}{(nh_n)^2} \sum_{j=1}^{\ell} \sum_{t \in T_j} \sum_{j'=1}^{\ell} \sum_{t' \in T_{j'}} |a_{t,j}(\ell)| |a_{t',j'}(\ell)| \mathbb{E}[Z_t Z_{t'}]. \end{aligned}$$

Since  $K$  is Lipschitz continuous, there exists a constant  $C > 0$  such that

$$|a_{t,j}(\ell)| \leq 2 \frac{cC}{\ell}, \quad \text{for } i_n \leq t \leq j_n \quad 1 \leq j \leq \ell.$$

Thus

$$\mathbb{E}[\Delta_n^2] \leq \frac{C^2}{(nh_n)^2} \left(\frac{2c}{\ell}\right)^2 \sum_{1 \leq t, t' \leq n} |\mathbb{E}[Z_t Z_{t'}]| \leq 8 \frac{C^2}{nh_n} \left(\frac{c}{\ell}\right)^2 \sum_{t \geq 0} |\mathbb{E}[Z_t Z_{t'}]|.$$

Then we obtain, again from Markov inequality that, for any  $\varepsilon > 0$ ,

$$\mathbb{P}\left(\left|S_n^{(\ell)} - S_n\right| \geq \varepsilon/\sqrt{nh_n}\right) \leq 2C^2 \left(\frac{c}{\ell}\right)^2 \sum_{t \geq 0} |\mathbb{E}[Z_t Z_{t'}]| \xrightarrow{\ell \rightarrow +\infty} 0.$$

Now the extension to  $d > 1$  is standard: consider  $r = (r_1, \dots, r_d)^T \in \mathbb{R}^d$  and a linear combination  $\mathcal{Z}_t = r_1 Z_t^{(1)} + \dots + r_d Z_t^{(d)}$  where  $Z_t = (Z_t^{(1)}, \dots, Z_t^{(d)})^T$  and apply the result obtained for  $d = 1$ . Then the asymptotic covariance matrix  $r^T \Sigma r > 0$  appears to be positive definite and this implies the multidimensional central limit theorem.  $\square$

## 7.2. Proofs of the main results

The following Lemma proves a strong law of large number on the contrast as if the stationary versions were observed: which implies the second point (uniform consistency) in Theorem 3.1. The first point (pointwise consistency) in Theorem 3.1 follows by similar arguments under the weaker assumptions as in Theorem 3.1. Thus the proof of the first point in Theorem 3.1 is omitted. The end of the section will aim at proving the CLT (15) in Theorem 3.1.

**Lemma 7.2.** *Under the assumptions of Theorem 3.1 we have*

$$\sup_{\varepsilon \leq u \leq 1-\varepsilon} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \Phi(\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta} \right) K\left(\frac{t-u}{h_n}\right) - \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \geq 0}, \boldsymbol{\theta})] \right| \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \quad (77)$$

*Proof.* The expression  $I$  in (77) tends to 0 a.s., if is the case for  $I_1$  and  $I_2$  such that,

$$I_1 = \sup_{\varepsilon \leq u \leq 1-\varepsilon} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \left( \Phi(\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta} \right) - \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta})] \right) K\left(\frac{t-u}{h_n}\right) \right|$$

$$I_2 = \sup_{\varepsilon \leq u \leq 1-\varepsilon} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} K\left(\frac{t-u}{h_n}\right) \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta})] - \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta})] \right|.$$

1. We use the part 1. of Lemma 7.1 to control  $I_1$ . For this we define  $Z(\boldsymbol{\theta}, u) = (Z_t(\boldsymbol{\theta}, u))_{t \in \mathbb{Z}}$  with  $Z_t(\boldsymbol{\theta}, u) = \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta}) - \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta})]$ ; this is a centred ergodic stationary process on the Banach space of the continuous function over  $\Theta \times [0, 1]$  equipped with the uniform norm.

Using  $\mathbb{E}[\sup_{(\boldsymbol{\theta}, u) \in \Theta \times [0, 1]} |Z_0(\boldsymbol{\theta}, u)|] < \infty$  since  $\Phi \in \text{Lip}_p(\Theta)$ , with Theorem 2.2.1. in [25] we apply the part 1. of Lemma 7.1 to get

$$I_1 \xrightarrow[n \rightarrow +\infty]{a.s.} 0. \quad (78)$$

2. For the term  $I_2$ , notice that

$$I_2 \leq \sup_{\varepsilon \leq u \leq 1-\varepsilon} \sup_{\boldsymbol{\theta} \in \Theta} \left| \left( 1 - \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} K\left(\frac{t-u}{h_n}\right) \right) \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta})] \right|$$

$$\leq C \sup_{\varepsilon \leq u \leq 1-\varepsilon} \left| 1 - \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} K\left(\frac{t-u}{h_n}\right) \right| \leq \frac{C}{nh_n}, \quad (79)$$

from the usual comparison of a Riemann sum and its integral: indeed  $K$  is Lipschitz because it is a piecewise differentiable function with a compact support. As a consequence, the proof is complete from (78) and (79).  $\square$



We also need the uniform approximation of the contrast with its stationary version stated in the next Proposition.

**Proposition 7.1.** *Under the assumptions of Theorem 3.1 with  $(\tilde{X}_t(u))_t$  denoting the stationary process defined in (9), we obtain*

$$\sup_{u \in [\varepsilon, 1-\varepsilon]} \sup_{\theta \in \Theta} \left| \frac{1}{nh_n} \sum_{k=1}^n \Phi((X_{k-t}^{(n)})_{t \geq 0}, \theta) K\left(\frac{\frac{k}{n} - u}{h_n}\right) - \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \geq 0}, \theta)] \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (80)$$

*Proof of Proposition 7.1.* Since  $\Phi \in \text{Lip}_p(\Theta)$  with  $p \geq 1$ , we have

$$\begin{aligned} & \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} \sup_{\theta \in \Theta} \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} (\Phi((X_{t-k}^{(n)})_{k \geq 0}, \theta) - \Phi((\tilde{X}_{t-k}(u))_{k \geq 0}, \theta)) K\left(\frac{\frac{t}{n} - u}{h_n}\right) \right\|_1 \\ & \leq \frac{C}{nh_n} \sum_{t=0}^{2c nh_n} \left| K\left(\frac{\frac{i_n(u)+t}{n} - u}{h_n}\right) \right| g\left(\sup_{s \leq j_n(u)} \{\|X_s^{(n)}\|_p \vee \|\tilde{X}_s(u)\|_p\}\right) \\ & \quad \times \sum_{s=1}^{\infty} \alpha_s(\Phi, \Theta) \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+t+1-s}^{(n)} - \tilde{X}_{i_n(u)+t+1-s}(u)| \right\|_p. \end{aligned}$$

From Assumption  $(\tilde{\mathbf{A}}_0(\Theta))$  and  $\tilde{B}_0(\Theta) < 1$ ,

$$\left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+j}^{(n)} - \tilde{X}_{i_n(u)+j}(u)| \right\|_p \leq C n^{1/p},$$

for  $j \leq 0$  using similar arguments as in the proof Lemma 6.4. Moreover with  $(\tilde{\mathbf{A}}_1(\Theta))$  and  $(\mathcal{H}(\rho))$ , we apply 6.4 in order to get

$$\left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} |X_{i_n(u)+j}^{(n)} - \tilde{X}_{i_n(u)+j}(u)| \right\|_p \leq n^{1/p} (h_n^\rho + \lambda_j).$$

for  $j \geq 1$ . Therefore,

$$\begin{aligned} & \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} \sup_{\theta \in \Theta} \frac{1}{nh_n} \sum_{t=0}^{2c nh_n} (\Phi((X_{t-k}^{(n)})_{k \geq 0}, \theta) - \Phi((\tilde{X}_{t-k}(u))_{k \geq 0}, \theta)) K\left(\frac{\frac{t}{n} - u}{h_n}\right) \right\|_1 \\ & \leq \frac{C}{nh_n} \sum_{t=0}^{2c nh_n} C_K C^* \left( \sum_{s=1}^t \alpha_s(\Phi, \Theta) C n^{1/p} (h_n^\rho + \lambda_{t+1-s}) + \sum_{s=t+1}^t \alpha_s(\Phi, \Theta) C n^{1/p} \right). \end{aligned} \quad (81)$$

Then we deduce

$$\begin{aligned}
& \left\| \sup_{u \in [\varepsilon, 1-\varepsilon]} \sup_{\boldsymbol{\theta} \in \Theta} \frac{1}{nh_n} \sum_{t=0}^{2c nh_n} (\Phi((X_{t-k}^{(n)}))_{k \geq 0}, \boldsymbol{\theta}) - \Phi((\tilde{X}_{t-k}(u))_{k \geq 0}, \boldsymbol{\theta}) K\left(\frac{t-u}{h_n}\right) \right\|_1 \\
& \leq \frac{C n^{1/p}}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \left( \sum_{s=1}^{t-i_n} \alpha_s(\Phi, \Theta) (\lambda_{t-s-i_n} + h_n^\rho) + \sum_{s=t-i_n+1}^{\infty} \alpha_s(\Phi, \Theta) \right) \\
& \leq \frac{C n^{1/p}}{nh_n} \left( \sum_{k=1}^{j_n-i_n} \lambda_k \sum_{i=1}^k \alpha_i(\Phi, \Theta) + h_n^\rho \sum_{k=1}^{j_n-i_n} \sum_{i=1}^k \alpha_i(\Phi, \Theta) + \sum_{i=1}^{\infty} i \alpha_i(\Phi, \Theta) \right) \\
& \leq \frac{C n^{1/p}}{nh_n} \left( \sum_{k=1}^{\infty} \lambda_k \sum_{i=1}^{\infty} \alpha_i(\Phi, \Theta) + (h_n^\rho + 1) \sum_{i=1}^{\infty} i \alpha_i(\Phi, \Theta) \right) \\
& \leq C \frac{n^{1/p}}{nh_n}. \tag{82}
\end{aligned}$$

Here we made use of the assumption  $\sum_{s=1}^{\infty} s \alpha_s(\Phi, \Theta) < \infty$  and of the fact that  $\sum_{k=1}^{\infty} \lambda_k < \infty$ . This last bound holds if  $\sum_{t=1}^{\infty} t \log(t) b_t^{(0)}(\Theta) < \infty$  and follows from Lemma 7.3. Finally, using (81), (82) and the almost sure convergence (77) we obtain the weak consistency result (80).  $\square$

**Lemma 7.3.** *If  $\sum_{t=2}^{\infty} t \log(t) b_t^{(0)}(\Theta) < \infty$  then  $\sum_{s=1}^{\infty} \lambda_s < \infty$ .*

*Proof.* Choosing  $r = \lfloor s/C \log(s) \rfloor$  for  $s \geq 2$  and  $C > 0$  we have

$$\lambda_s \leq s^{-C \log(1/\tilde{B}_0(\Theta))} + \sum_{t=\lfloor s/C \log(s) \rfloor}^{\infty} b_t^{(0)}(\Theta).$$

For  $C > 0$  large enough and since  $\tilde{B}_0(\Theta) < 1$ , we get  $\sum_{s=1}^{\infty} s^{-C \log(1/\tilde{B}_0(\Theta))} < \infty$ . Moreover, for  $s > e$  then  $t = s/C \log(s)$  implies  $s > Ct \log(t)$ . Thus:

$$\sum_{s=3}^{\infty} \sum_{t=\lfloor s/C \log(s) \rfloor}^{\infty} b_t^{(0)}(\Theta) \leq C \sum_{t=1}^{\infty} t \log(t) b_t^{(0)}(\Theta) < \infty,$$

and the desired result follows.  $\square$

*Proof of Theorem 3.1 (uniform consistency).* From the Assumption  $(\mathbf{Co}(\Phi, \Theta))$ , we have

$$\boldsymbol{\theta}^*(u) = \underset{\boldsymbol{\theta} \in \Theta}{\text{Argmin}} \mathbb{E}[\Phi((\tilde{X}_{-t}(u))_{t \geq 0}, \boldsymbol{\theta})].$$

The uniform weak law of large numbers implies the uniform convergence, and we need:

$$\sup_{u \in [\varepsilon, 1-\varepsilon]} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{nh_n} \sum_{k=1}^n \Phi((X_{k-t}^{(n)})_{t \geq 0}, \boldsymbol{\theta}) K\left(\frac{k-u}{h_n}\right) - \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \geq 0}, \boldsymbol{\theta})] \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0,$$

see the discussion in the Appendix of [2]. From an application of the approximation in Proposition 7.1, this uniform weak law of large number follows from the uniform weak law of large number on the stochastic version of the contrast, namely

$$\sup_{u \in [\varepsilon, 1-\varepsilon]} \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{nh_n} \sum_{k=1}^n \Phi((\tilde{X}_{k-t}(u))_{t \geq 0}, \boldsymbol{\theta}) K\left(\frac{k-u}{h_n}\right) - \mathbb{E}[\Phi((\tilde{X}_{-k}(u))_{k \geq 0}, \boldsymbol{\theta})] \right| \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0.$$

From usual arguments, see for instance [26], this uniform version of the weak law of large number obtained in Lemma 7.2 will follow from the equicontinuity of the family

$$\left( \sup_{\boldsymbol{\theta} \in \Theta} \left| \frac{1}{nh_n} \sum_{k=1}^n \Phi((\tilde{X}_{k-t}(u))_{t \geq 0}, \boldsymbol{\theta}) K\left(\frac{k-u}{h_n}\right) - \mathbb{E}[\Phi((\tilde{X}_{-t}(u))_{t \geq 0}, \boldsymbol{\theta})] \right| \right)_{u \in [\varepsilon, 1-\varepsilon]}.$$

This holds from Markov inequality as  $\Phi \in \text{Lip}_p(\Theta)$  and from the relation

$$\|\tilde{X}_t(u) - \tilde{X}_t(u')\|_p \leq \frac{\|\boldsymbol{\theta}^*(u') - \boldsymbol{\theta}^*(u)\|}{1 - \tilde{B}_0(\Theta)} \left( \frac{\tilde{B}_1(\Theta) \tilde{C}_0(\Theta)}{1 - \tilde{B}_0(\Theta)} + \tilde{C}_1(\Theta) \right), \quad (83)$$

as  $\boldsymbol{\theta}^*$  is equicontinuous. Indeed, under  $(\tilde{\mathbf{A}}_k(\boldsymbol{\Theta}))$ ,  $k = 1, 2$ , we have

$$\begin{aligned} \|\tilde{X}_t(u) - \tilde{X}_t(u')\|_p &\leq \|F_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{t-k}(u))_{k \geq 1}, \xi_t) - F_{\boldsymbol{\theta}^*(u')}((\tilde{X}_{t-k}(u'))_{k \geq 1}, \xi_t)\|_p \\ &\leq \|F_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{t-k}(u))_{k \geq 1}, \xi_t) - F_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{t-k}(u'))_{k \geq 1}, \xi_t)\|_p \\ &\quad + \|F_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{t-k}(u'))_{k \geq 1}, \xi_t) - F_{\boldsymbol{\theta}^*(u)}((\tilde{X}_{t-k}(u'))_{k \geq 1}, \xi_t)\|_p \\ &\leq \sum_{k=1}^{\infty} b_k^{(0)}(\Theta) \|\tilde{X}_{t-k}(u') - \tilde{X}_{t-k}(u)\|_p \\ &\quad + \|\boldsymbol{\theta}^*(u') - \boldsymbol{\theta}^*(u)\| \times \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}}^1 F_{\boldsymbol{\theta}}(\tilde{X}_{t-1}(u), \tilde{X}_{t-2}(u), \dots, \xi_t)\|_p. \end{aligned}$$

We upper-bound

$$\begin{aligned} \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}}^1 F_{\boldsymbol{\theta}}(\tilde{X}_{t-1}(u), \tilde{X}_{t-2}(u), \dots, \xi_t)\|_p \\ \leq \sum_{k=1}^{\infty} b_k^{(1)}(\Theta) \|\tilde{X}_{t-k}(u)\|_p + \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}}^1 F_{\boldsymbol{\theta}}(0, 0, \dots, \xi_t)\|_p. \end{aligned}$$

By a similar argument than in the proof of Lemma 6.3, we deduce that (83) holds.  $\square$

Now we are in position to prove Theorem 3.2.

*Proof of Theorem 3.2.* We follow the usual proof of asymptotic normality of a M-estimator. This will follow from the 3 forthcoming steps:

- I/ We establish that the family  $\partial_{\theta} H_t(\theta^*(u)) = \left( \frac{\partial}{\partial \theta_i} H_t(\theta^*(u)) \right)_{1 \leq i \leq d}$  for  $i_n \leq t \leq j_n$  satisfies a multidimensional central limit theorem, where we denote  $H_t(\theta) = \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta)$ .

We notice first that  $\partial_{\theta} \mathbb{E}[\Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta^*(u)) \mid \mathcal{F}_0] = 0$  as  $\theta^*(u)$  is the unique minimizer of  $\mathbb{E}[\Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta^*(u)) \mid \mathcal{F}_0]$  over the open set  $\overset{\circ}{\Theta}$ .

The function  $\theta \in \Theta \mapsto \mathbb{E}[\Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta^*(u)) \mid \mathcal{F}_0]$  is differentiable under the condition  $\|\partial_{\theta} \Phi\| \in \text{Lip}_p(\Theta)$ . Thus  $\partial_{\theta} H_t(\theta^*(u))$  constitutes a differences of martingale sequence. We also have  $\mathbb{E}[\|\partial_{\theta} \Phi\|^2] < \infty$  and we can apply the CLT for differences of martingale sequences (See for instance [26]). As a consequence we can apply the point 2. of Lemma 7.1 and obtain the multidimensional central limit theorem

$$\frac{1}{\sqrt{nh_n}} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\theta} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta^*(u)) K\left(\frac{t-u}{h_n}\right) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \Sigma(\theta^*(u))\right) \quad (84)$$

$$\text{with } \Sigma(\theta^*(u)) = \int_{\mathbb{R}} K^2(x) dx$$

$$\times \sum_{t \in \mathbb{Z}} \left( \text{Cov} \left[ \frac{\partial}{\partial \theta_i} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta^*(u)), \frac{\partial}{\partial \theta_j} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta^*(u)) \right] \right)_{1 \leq i, j \leq d}.$$

- II/ We use a Taylor-Lagrange expansion for establishing

$$\begin{aligned} & \frac{1}{\sqrt{nh_n}} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\theta} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \hat{\theta}(u)) K\left(\frac{t-u}{h_n}\right) \\ &= \frac{1}{\sqrt{nh_n}} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\theta} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta^*(u)) K\left(\frac{t-u}{h_n}\right) \\ &+ \sqrt{nh_n} \cdot \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\theta^2}^2 \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \bar{\theta}(u)) K\left(\frac{t-u}{h_n}\right) (\hat{\theta}(u) - \theta^*(u)), \end{aligned} \quad (85)$$

where  $\bar{\theta}(u)$  belongs to the segment with extremities  $\theta^*(u)$  and  $\hat{\theta}(u)$ . From Theorem 3.1, we have  $\bar{\theta}(u) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \theta^*(u)$ . Moreover, since  $\mathbb{E}[\|\partial_{\theta^2}^2 \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta)\|] < \infty$  for any  $\theta \in \Theta$  and  $\theta \in \Theta \mapsto \partial_{\theta^2}^2 \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \theta)$  is uniformly continuous because  $\Theta$  is a bounded set included in  $\mathbb{R}^d$ , we can apply Lemma 7.1 and then:

$$\frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} (\partial_{\theta^2}^2 \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \bar{\theta}(u)) \mathbb{E}[\partial_{\theta^2}^2 \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \bar{\theta}(u))]) K\left(\frac{t-u}{h_n}\right) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0. \quad (86)$$

Thus we get, with  $\Gamma(\boldsymbol{\theta}^*(u)) = \mathbb{E}[\partial_{\boldsymbol{\theta}^2}^2 \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta}^*(u))]$ ,

$$\frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\boldsymbol{\theta}^2}^2 \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \bar{\boldsymbol{\theta}}(u)) K\left(\frac{\frac{t}{n} - u}{h_n}\right) \xrightarrow[n \rightarrow +\infty]{\mathbb{P}} \Gamma(\boldsymbol{\theta}^*(u)).$$

Moreover, since  $\hat{\boldsymbol{\theta}}(u)$  minimizes the contrast function we have

$$\frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\boldsymbol{\theta}} \Phi((X_{t-k}^{(n)}(u))_{k \in \mathbb{N}}, \hat{\boldsymbol{\theta}}(u)) K\left(\frac{\frac{t}{n} - u}{h_n}\right) = 0. \quad (87)$$

Using the assumptions on the Lipschitz coefficients of  $\partial_{\boldsymbol{\theta}} \Phi$ , the same inequalities as (81) and (82) in the proof of Proposition 7.1 lead for a convenient constant  $C > 0$  to:

$$\begin{aligned} \left\| \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} (\partial_{\boldsymbol{\theta}} \Phi((X_{t-k}^{(n)}(u))_{k \in \mathbb{N}}, \hat{\boldsymbol{\theta}}(u)) - \partial_{\boldsymbol{\theta}} \Phi((X_{t-k}^*(u))_{k \geq 0}, \hat{\boldsymbol{\theta}}(u))) K\left(\frac{\frac{t}{n} - u}{h_n}\right) \right\|_1 &\leq Ch_n^\rho, \\ \left\| \frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} (\partial_{\boldsymbol{\theta}} \Phi((X_{t-k}^*(u))_{k \geq 0}, \hat{\boldsymbol{\theta}}(u)) - \partial_{\boldsymbol{\theta}} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \hat{\boldsymbol{\theta}}(u))) K\left(\frac{\frac{t}{n} - u}{h_n}\right) \right\|_1 &\leq \frac{C}{nh_n}. \end{aligned}$$

As a consequence we deduce that:

$$\begin{aligned} \left\| \frac{1}{\sqrt{nh_n}} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\boldsymbol{\theta}} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \hat{\boldsymbol{\theta}}(u)) K\left(\frac{\frac{t}{n} - u}{h_n}\right) \right\|_1 &\leq C \left( \frac{1}{\sqrt{nh_n}} + h_n^\rho \sqrt{nh_n} \right) \\ \implies \frac{1}{\sqrt{nh_n}} \sum_{t=i_n(u)}^{j_n(u)} \partial_{\boldsymbol{\theta}} \Phi((\tilde{X}_{t-k}(u))_{k \in \mathbb{N}}, \hat{\boldsymbol{\theta}}(u)) K\left(\frac{\frac{t}{n} - u}{h_n}\right) &\xrightarrow[n \rightarrow +\infty]{\mathbb{P}} 0, \quad (88) \end{aligned}$$

by using (17). Finally, from (85), using (86), (88), Slutsky Lemma and (84), we deduce:

$$\sqrt{nh_n} \Gamma(\boldsymbol{\theta}^*(u)) (\hat{\boldsymbol{\theta}}(u) - \boldsymbol{\theta}^*(u)) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \Sigma(\boldsymbol{\theta}^*(u))\right),$$

and this leads to Theorem 3.2.  $\square$

*Proof of Proposition 4.1.* We proved in Section 4 that  $a_k(\Phi_G, \Theta) = b_k^{(0)}(\Theta)$  and  $\Phi_G \in \text{Lip}_3(\Theta)$  when  $f_{\boldsymbol{\theta}}$  and  $M_{\boldsymbol{\theta}}$  satisfy Lipschitz inequalities (23). But we also have:

$$\begin{aligned} \partial_{\boldsymbol{\theta}} \Phi_G(x, \boldsymbol{\theta}) &= \frac{\partial_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}} \cdot ((x_k)_{k \geq 2})}{M_{\boldsymbol{\theta}}((x_k)_{k \geq 2})} + 2 \partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}((x_k)_{k \geq 2}) \cdot \frac{f_{\boldsymbol{\theta}}((x_k)_{k \geq 2}) - x_1}{M_{\boldsymbol{\theta}}^2((x_k)_{k \geq 2})} \\ &\quad - 2 \partial_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}}((x_k)_{k \geq 2}) \cdot \frac{(f_{\boldsymbol{\theta}}((x_k)_{k \geq 2}) - x_1)^2}{M_{\boldsymbol{\theta}}^3((x_k)_{k \geq 2})}. \end{aligned}$$

After computations and using  $M_\theta \geq \underline{M}$  as well as Hölder Inequalities, we obtain

$$\begin{aligned}
\mathbb{E} \left[ \sup_{\theta \in \Theta} \left\| \partial_\theta \Phi_G(U, \theta) - \partial_\theta \Phi_G(V, \theta) \right\| \right] &\leq \frac{1}{\underline{M}} \left\| \partial_\theta M_\theta((U_k)_{k \geq 2}) - \partial_\theta M_\theta((V_k)_{k \geq 2}) \right\|_1 \\
&\quad + \frac{\left\| \partial_\theta M_\theta((V_k)_{k \geq 2}) \right\|_2}{\underline{M}} \left\| M_\theta((U_k)_{k \geq 2}) - M_\theta((V_k)_{k \geq 2}) \right\|_2 \\
&\quad + 2 \left( \frac{\left\| f_\theta((U_k)_{k \geq 2}) - U_1 \right\|_2}{\underline{M}^2} \left\| \partial_\theta f_\theta((U_k)_{k \geq 2}) - \partial_\theta f_\theta((V_k)_{k \geq 2}) \right\|_2 \right. \\
+ 2 \cdot \frac{\left\| f_\theta((V_k)_{k \geq 2}) - V_1 \right\|_3 \left\| \partial_\theta f_\theta((V_k)_{k \geq 2}) \right\|_3}{\underline{M}^3} &\left. \left\| M_\theta((U_k)_{k \geq 2}) - M_\theta((V_k)_{k \geq 2}) \right\|_3 \right. \\
&\quad + \frac{\left\| \partial_\theta f_\theta((V_k)_{k \geq 2}) \right\|_2}{\underline{M}^2} \left( \left\| f_\theta((U_k)_{k \geq 2}) - f_\theta((V_k)_{k \geq 2}) \right\|_2 + \|U_1 - V_1\|_2 \right) \\
&\quad + 2 \left( \frac{\left\| \partial_\theta M_\theta((V_k)_{k \geq 2}) \right\|_3 \left\| f_\theta((U_k)_{k \geq 2}) + f_\theta((V_k)_{k \geq 2}) - U_1 - V_1 \right\|_3}{\underline{M}^3} \right. \\
&\quad \quad \left. \times \left( \left\| f_\theta((U_k)_{k \geq 2}) - f_\theta((V_k)_{k \geq 2}) \right\|_3 + \|U_1 - V_1\|_3 \right) \right. \\
+ 3 \cdot \frac{\left\| \partial_\theta M_\theta((V_k)_{k \geq 2}) \right\|_4 \left\| f_\theta((V_k)_{k \geq 2}) - V_1 \right\|_4^2}{\underline{M}^4} &\left. \left( \left\| M_\theta((U_k)_{k \geq 2}) - M_\theta((V_k)_{k \geq 2}) \right\|_4 \right) \right. \\
&\quad \left. + \frac{\left\| f_\theta((U_k)_{k \geq 2}) - U_1 \right\|_3^2}{\underline{M}^3} \left\| \partial_\theta M_\theta((U_k)_{k \geq 2}) - \partial_\theta M_\theta((V_k)_{k \geq 2}) \right\|_3 \right).
\end{aligned}$$

Thus

$$\begin{aligned}
\mathbb{E} \left[ \sup_{\theta \in \Theta} \left\| \partial_\theta \Phi_G(U, \theta) - \partial_\theta \Phi_G(V, \theta) \right\| \right] &\leq g \left( \sup_{i \geq 1} \{ \|U_i\|_4 \vee \|V_i\|_4 \} \right) \\
&\times \left( \left\| \partial_\theta M_\theta((U_k)_{k \geq 2}) - \partial_\theta M_\theta((V_k)_{k \geq 2}) \right\|_4 + \left\| \partial_\theta f_\theta((U_k)_{k \geq 2}) - \partial_\theta f_\theta((V_k)_{k \geq 2}) \right\|_4 \right. \\
&\left. + \left\| M_\theta((U_k)_{k \geq 2}) - M_\theta((V_k)_{k \geq 2}) \right\|_4 + \left\| f_\theta((U_k)_{k \geq 2}) - f_\theta((V_k)_{k \geq 2}) \right\|_4 + \|U_1 - V_1\|_4 \right)
\end{aligned}$$

from Jensen inequality and since we assume that  $f_\theta$ ,  $M_\theta$ ,  $\partial_\theta f_\theta$  and  $\partial_\theta M_\theta$  satisfy Lipschitz inequalities (23). As a consequence we derive

$$\begin{aligned}
\mathbb{E} \left[ \sup_{\theta \in \Theta} \left\| \partial_\theta \Phi_G(U, \theta) - \partial_\theta \Phi_G(V, \theta) \right\| \right] &\leq g \left( \sup_{i \geq 1} \{ \|U_i\|_4 \vee \|V_i\|_4 \} \right) \\
&\times \left( \|U_1 - V_1\|_4 + \sum_{i=2}^{\infty} (\beta_i(f, \Theta) + \beta_i(M, \Theta) + \beta_i(\partial_\theta f, \Theta) + \beta_i(\partial_\theta M, \Theta)) \|U_i - V_i\|_4 \right),
\end{aligned}$$

therefore  $\partial_\theta \Phi_G \in \text{Lip}_4(\Theta)$ . From these computations and with the inequality (24) we also deduce that condition (31) implies  $B_0(\Theta) < 1$ , and  $\sum_{t=2}^{\infty} t \log(t) b_t(\Theta) < \infty$  follows from  $\sum_{s \geq 0} s a_s(\Phi, \Theta) < \infty$ , required in Theorems 3.1 and 3.2. Similar calculations also entail

$$\mathbb{E} \left[ \left\| \partial_{\theta^2}^2 \Phi((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}, \theta) \right\|^2 \right] < \infty, \quad \text{for any } \theta \in \Theta,$$

since  $p = 4$  and  $\partial_{\theta^2}^2 f_{\theta}$  and  $\partial_{\theta^2}^2 M_{\theta}$  satisfy Lipschitz inequalities (23). We also have

$$\begin{aligned} \mathbb{E}[\|\partial_{\theta}\Phi((\tilde{X}_k(u))_{k \leq 0}, \theta^*(u))\|^2] &\leq \frac{12}{(1 \vee M)^2} \left( \|\partial_{\theta} M_{\theta^*(u)}((\tilde{X}_k(u))_{k \leq -1})\|_2^2 \right. \\ &\quad \left. + \|\partial_{\theta} f_{\theta^*(u)}((\tilde{X}_k(u))_{k \leq -1})\|_2^2 \|\xi_0\|_2^2 + \|\partial_{\theta} M_{\theta^*(u)}((\tilde{X}_k(u))_{k \leq -1})\|_2^2 \|\xi_0\|_4^4 \right), \end{aligned}$$

since  $\tilde{X}_0(u) - f_{\theta^*(u)}((\tilde{X}_k(u))_{k \leq -1}) = M_{\theta^*(u)}((\tilde{X}_k(u))_{k \leq -1}) \xi_0$ , with  $M_{\theta^*(u)}((\tilde{X}_k(u))_{k \leq -1})$  and  $\xi_0$  which are independent. Therefore, we obtain

$$\mathbb{E}[\|\partial_{\theta}\Phi((\tilde{X}_k(u))_{k \leq 0}, \theta^*(u))\|^2] < \infty$$

since  $p = 4$ . Finally (30) ensures that asymptotic covariance matrix  $\Sigma(\theta^*(u))$  and  $\Gamma(\theta^*(u))$  are positive definite matrix (see [11]) and (29) implies the existence and the uniqueness of  $\theta^*(u)$  as the minimum of  $\theta \in \Theta \mapsto \mathbb{E}[\Phi((\tilde{X}_k(u))_{k \leq 1}, \theta) \mid \mathcal{F}_0]$  defined in (11) (see also [11]). This ends the checking of the conditions of Theorem 3.2.  $\square$

*Proof of Proposition 4.2.* Denote  $P_{\theta}(B) = \sum_{j=1}^{\infty} a(\theta) B^{j-1}$  and  $R_{\theta}(x) = -\sum_{j=0}^{\infty} b_{j+1}(\theta) x^j$ . Using an iteration of (40), we obtain:

$$\bar{X}_t^{(n)} = \sigma(\theta_t^{(n)}) \xi_t + (I - P_{\theta_t^{(n)}}(B)) \bar{X}_t^{(n)} - \sum_{j=1}^{\infty} P_{\theta_{t-j}^{(n)}}(B) \left( \prod_{k=0}^{j-1} R_{\theta_{t-k}^{(n)}}(B) \right) \bar{X}_{t-j}^{(n)}, \quad (89)$$

for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  and  $\bar{X}_t^{(n)} = 0$  for  $t \leq 0$ . Following the same idea, we also have:

$$\begin{aligned} X_t^{(n)} &= \sigma(\theta_t^{(n)}) \xi_t + \sum_{i=1}^{\infty} d_i(\theta_t^{(n)}) X_{t-i} \\ &= \sigma(\theta_t^{(n)}) \xi_t + (I - P_{\theta_t^{(n)}}(B)) X_t^{(n)} - \sum_{j=1}^{\infty} P_{\theta_{t-j}^{(n)}}(B) \left( R_{\theta_{t-j}^{(n)}}(B) \right)^j X_{t-j}^{(n)}, \quad (90) \end{aligned}$$

for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$  and  $X_t^{(n)} = 0$  for  $t \leq 0$ . Therefore,

$$\begin{aligned} &\|\bar{X}_t^{(n)} - X_t^{(n)}\|_p \\ &\leq \left\| \sum_{j=1}^{\infty} P_{\theta_{t-j}^{(n)}}(B) \left( \prod_{k=0}^{j-1} R_{\theta_{t-k}^{(n)}}(B) \right) - P_{\theta_t^{(n)}}(B) (R_{\theta_t^{(n)}}(B))^j \right\| \bar{X}_{t-j}^{(n)} \Big\|_p \\ &+ \left\| (I - P_{\theta_t^{(n)}}(B)) (X_t^{(n)} - \bar{X}_t^{(n)}) - \sum_{j=1}^{\infty} P_{\theta_{t-j}^{(n)}}(B) (R_{\theta_{t-j}^{(n)}}(B))^j (X_{t-j}^{(n)} - \bar{X}_{t-j}^{(n)}) \right\|_p. \quad (91) \end{aligned}$$

Now let  $U_t(x) = \sum_{k=0}^{\infty} u_k(t)x^k$ ,  $V_t(x) = \sum_{k=0}^{\infty} v_k(t)x^k$  with  $\sup_{t \in \mathbb{N}} \sum_{k=0}^{\infty} |u_k(t)| < 1$  and  $\sup_{t \in \mathbb{N}} \sum_{k=0}^{\infty} |v_k(t)| < 1$ . Then, for any  $s, t \in \mathbb{N}$ ,

$$U_s(B)V_t(B) = \sum_{k=0}^{\infty} c_k(s, t) B^k,$$

$$\text{where } c_k(s, t) = \sum_{j=0}^k u_j(s) v_{k-j}(t), \text{ and } \sup_{s, t \in \mathbb{N}} \sum_{k=0}^{\infty} |c_k(s, t)| < 1.$$

Moreover, if  $\begin{cases} |u_j(s) - u_j(s')| \leq |\tilde{u}_j| \left(\frac{|s-s'|}{n}\right)^\rho \\ |v_j(s) - v_j(s')| \leq |\tilde{v}_j| \left(\frac{|s-s'|}{n}\right)^\rho \end{cases}$  for any  $j \in \mathbb{N}$ ,  $1 \leq s, s' \leq n$

with  $\sum_i |\tilde{u}_i| + |\tilde{v}_i| < \infty$  then if  $X = (X_\ell)$  is a sequence with  $\sup_{\ell \in \mathbb{Z}} \|X_\ell\|_p < \infty$ , for any  $\ell \leq n$ ,

$$\begin{aligned} \|U_s(B)V_t(B)X_n - U_{s'}(B)V_{t'}(B)X_\ell\|_p &\leq \left\| \sum_{k=0}^{\infty} (c_k(s, t) - c_k(s', t')) X_{\ell-k} \right\|_p \\ &\leq \sum_{k=0}^{\infty} \left( \left| \frac{s-s'}{n} \right|^\rho w_k^{(u)}(t) + \left| \frac{t-t'}{n} \right|^\rho w_k^{(v)}(s) \right) \|X_{\ell-k}\|_p, \end{aligned}$$

with  $w_k^{(u)}(t) = \sum_{j=0}^k |\tilde{u}_j| |v_{k-j}(t)|$  and  $w_k^{(v)}(s) = \sum_{j=0}^k |\tilde{v}_j| |u_{k-j}(s)|$ .

Now, if we denote  $\tilde{a}_k = K_\theta \sup_{0 \leq t, n, n \in \mathbb{N}^*} |a'_k(\theta_t^{(n)})|$  and  $\tilde{b}_k = K_\theta \sup_{0 \leq t, n, n \in \mathbb{N}^*} |b'_k(\theta_t^{(n)})|$ ,

implying  $\sum_k \tilde{a}_k + \tilde{b}_k < \infty$  from condition (39). We obtain for  $j \geq 2$ ,

$$P_{\theta_{t-j}^{(n)}}(B) \left( \prod_{k=0}^{j-1} R_{\theta_{t-k}^{(n)}}(B) \right) - P_{\theta_t^{(n)}}(B) \left( R_{\theta_t^{(n)}}(B) \right)^j = \sum_{\ell=0}^{\infty} c_\ell(t, j) B^\ell$$

$$\begin{aligned} \text{with } |c_\ell(t, j)| &\leq \sum_{i_0 + \dots + i_j = \ell} |b_{i_1}(\theta_t^{(n)})| \left( \left(\frac{j}{n}\right)^\rho \tilde{a}_{i_0} \prod_{k=2}^j b_{i_k}(\theta_t^{(n)}) \right. \\ &\quad \left. + \sum_{k=2}^j \left(\frac{k-2}{n}\right)^\rho a_{i_0}(\theta_{t-j}^{(n)}) \tilde{b}_{i_k} \prod_{m=2}^k |b_{i_m}(\theta_{t-m}^{(n)})| \prod_{m=k+1}^j |b_{i_m}(\theta_t^{(n)})| \right) \\ &\leq \left(\frac{j}{n}\right)^\rho \sum_{i_0 + \dots + i_j = \ell} |b_{i_1}(\theta_t^{(n)})| \left( \tilde{a}_{i_0} \prod_{k=2}^j |b_{i_k}(\theta_t^{(n)})| \right. \\ &\quad \left. + \sum_{k=2}^j |a_{i_0}(\theta_{t-j}^{(n)})| \tilde{b}_{i_k} \prod_{m=2}^k |b_{i_m}(\theta_{t-m}^{(n)})| \prod_{m=k+1}^j |b_{i_m}(\theta_t^{(n)})| \right). \end{aligned}$$



Then,

$$\begin{aligned}
\sum_{\ell=0}^{\infty} |c_{\ell}(t, j)| &\leq \left(\frac{j}{n}\right)^{\rho} \left[ \left( \sum_{k=0}^{\infty} \tilde{a}_k \right) \left( \sum_{k=0}^{\infty} |b_k(\boldsymbol{\theta}_t^{(n)})| \right)^j \right. \\
&\quad \left. + \sum_{k=0}^{\infty} |a_k(\boldsymbol{\theta}_{t-j}^{(n)})| \left( \sum_{k=0}^{\infty} \tilde{b}_k \right) \sum_{j'=2}^j \left( \sum_{k=0}^{\infty} |b_k(\boldsymbol{\theta}_t^{(n)})| \right)^{j-j'} \prod_{m=2}^{j'} \left( \sum_{k=0}^{\infty} |b_{j_m}(\boldsymbol{\theta}_{t-m}^{(n)})| \right) \right] \\
&\leq \left(\frac{j}{n}\right)^{\rho} \max \left\{ \sum_{k=0}^{\infty} \tilde{a}_k, \sum_{k=0}^{\infty} \tilde{b}_k \right\} j \bar{a} \bar{b}^j,
\end{aligned}$$

since  $\bar{a} = \sup_{\boldsymbol{\theta} \in \Theta} \sum_{k=0}^{\infty} |a_k(\boldsymbol{\theta})| < 1$  and  $\bar{b} = \sup_{\boldsymbol{\theta} \in \Theta} \sum_{k=0}^{\infty} |b_k(\boldsymbol{\theta})| < 1$  by assumption. Therefore,

$$\sum_{j=1}^{\infty} \sum_{\ell=0}^{\infty} |c_{\ell}(t, j)| \leq \frac{C}{n^{\rho}} \sum_{j=1}^{\infty} j^{1+\rho} \bar{b}^j \leq \frac{C'}{n^{\rho}},$$

since  $0 < \bar{b} < 1$ . Therefore with Lemma 2.1,

$$\begin{aligned}
\left\| \sum_{j=1}^{\infty} \left( P_{\boldsymbol{\theta}_{t-j}^{(n)}}(B) \left( \prod_{k=0}^{j-1} R_{\boldsymbol{\theta}_{t-k}^{(n)}}(B) \right) - P_{\boldsymbol{\theta}_t^{(n)}}(B) \left( R_{\boldsymbol{\theta}_t^{(n)}}(B) \right)^j \right) \bar{X}_{t-j}^{(n)} \right\|_p \\
\leq \frac{C'}{n^{\rho}} \frac{C_0(\Theta)}{1 - B_0(\Theta)} \leq \frac{C''}{n^{\rho}}. \quad (92)
\end{aligned}$$

Finally, with  $M_t^{(n)} = \sup_{s \leq t} \|X_{t-s}^{(n)} - \bar{X}_{t-s}^{(n)}\|_p$ ,

$$\begin{aligned}
\left\| (I - P_{\boldsymbol{\theta}_t^{(n)}}(B)) (X_t^{(n)} - \bar{X}_t^{(n)}) - \sum_{j=1}^{\infty} P_{\boldsymbol{\theta}_t^{(n)}}(B) \left( R_{\boldsymbol{\theta}_t^{(n)}}(B) \right)^j (X_{t-j}^{(n)} - \bar{X}_{t-j}^{(n)}) \right\|_p \\
= \left\| \sum_{j=1}^{\infty} d_j(\boldsymbol{\theta}_t^{(n)}) (X_{t-j}^{(n)} - \bar{X}_{t-j}^{(n)}) \right\|_p \leq \bar{d} \times M_{t-1}^{(n)}, \quad (93)
\end{aligned}$$

with  $0 \leq \bar{d} < 1$ . As a consequence, using (91), (92) and (93), for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$ ,

$$\|\bar{X}_t^{(n)} - X_t^{(n)}\|_p \leq \bar{d} \times M_{t-1}^{(n)} + \frac{C''}{n^{\rho}} \implies M_t^{(n)} \leq \bar{d} \times M_{t-1}^{(n)} + \frac{C''}{n^{\rho}}.$$

Since  $M_{t-1}^{(n)} = 0$ , this implies for any  $1 \leq t \leq n$ ,  $n \in \mathbb{N}^*$ ,

$$M_t^{(n)} \leq \frac{C''}{n^{\rho}} \sum_{k=0}^{t-1} \bar{d}^k \leq \frac{C''}{1 - \bar{d}} \times \frac{1}{n^{\rho}}.$$

And this achieves the proof.  $\square$

*Proof of Proposition 4.3.* We will sharply follow the same proof that the one of Proposition 4.2. For any  $n \in \mathbb{N}^*$  and  $1 \leq t \leq n$ , using the causal property of the processes, we have

$$\|\bar{X}_t^{(n)} - X_t^{(n)}\|_p = \|\xi_0\|_p \|\bar{\sigma}_t^{(n)} - \sigma_t^{(n)}\|_p \leq \frac{\|\xi_0\|_p}{2 \inf_{1 \leq t \leq n} a_0(\boldsymbol{\theta}_t^{(n)})} \left\| (\bar{\sigma}_t^{(n)})^2 - (\sigma_t^{(n)})^2 \right\|_p. \quad (94)$$

Using an iteration of (47), we also obtain for any  $n \in \mathbb{N}^*$  and  $1 \leq t \leq n$ ,

$$\begin{aligned} & (\bar{\sigma}_t^{(n)})^2 \\ &= \left( a_0(\boldsymbol{\theta}_t^{(n)}) + \sum_{k=0}^{\infty} a_0(\boldsymbol{\theta}_{t-k-1}^{(n)}) \prod_{j=0}^k Q_{t-j}(1) \right) + \sum_{k=0}^{\infty} P_{t-k}(B) \left( \prod_{j=0}^{k-1} Q_{t-j}(B) \right) (\bar{X}_{t-k-1}^{(n)})^2. \end{aligned} \quad (95)$$

As a consequence, we have

$$\begin{aligned} & \left\| (\bar{\sigma}_t^{(n)})^2 - (\sigma_t^{(n)})^2 \right\|_p \leq \left| a_0(\boldsymbol{\theta}_t^{(n)}) \left[ \sum_{k=0}^{\infty} \left( Q_t(1)^k - \prod_{j=0}^k Q_{t-j}(1) \right) \right] \right| \\ &+ \left\| \sum_{k=0}^{\infty} \left( P_t(B) Q_t^k(B) \right) (X_{t-k-1}^{(n)})^2 - P_{t-k}(B) \left( \prod_{j=0}^{k-1} Q_{t-j}(B) \right) (\bar{X}_{t-k-1}^{(n)})^2 \right\|_p. \end{aligned}$$

Now we can use the same inequalities and bounds as in the proof of Proposition 4.2 for obtaining (48).  $\square$

*Proof of Proposition 4.4.* We already proved in Section 4.3 that  $\Phi_{LARCH} \in \text{Lip}_4(\Theta)$  as well as Assumption  $\mathbf{Co}(\Phi_{LARCH}, \Theta)$  when condition (55) holds. We assumed that  $\boldsymbol{\theta} \in \Theta \mapsto a_i(\boldsymbol{\theta})$  are  $\mathcal{C}^2(\Theta)$  functions for any  $i \in \mathbb{N}$ . Thus in order to check the conditions of Theorem 3.2, we first have to prove that  $\partial_{\boldsymbol{\theta}} \Phi_{LARCH} \in \text{Lip}_4(\Theta)$ . Indeed we use the estimates

$$\begin{aligned} & \left\| \partial_{\boldsymbol{\theta}} \Phi_{LARCH}(U, \boldsymbol{\theta}) - \partial_{\boldsymbol{\theta}} \Phi_{LARCH}(V, \boldsymbol{\theta}) \right\| \\ & \leq 8 \left( |U_1| + |V_1| + 2a_0(\boldsymbol{\theta}) + \sum_{i=1}^{\infty} |a_i(\boldsymbol{\theta})| (|U_{i+1}| + |V_{i+1}|) \right)^2 \\ & \quad \times \left( \|\partial_{\boldsymbol{\theta}} a_0(\boldsymbol{\theta})\| + \sum_{i=1}^{\infty} \|\partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta})\| |U_{i+1}| \right) \left( |U_1 - V_1| + \sum_{i=1}^{\infty} |a_i(\boldsymbol{\theta})| |U_{i+1} - V_{i+1}| \right) \\ & \quad + 4 \left( |U_1| + |V_1| + 2a_0(\boldsymbol{\theta}) + \sum_{i=1}^{\infty} |a_i(\boldsymbol{\theta})| (|U_{i+1}| + |V_{i+1}|) \right)^3 \sum_{i=1}^{\infty} \|\partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta})\| |U_{i+1} - V_{i+1}|. \end{aligned}$$

Therefore, using Hölder and Minkowski Inequalities, we obtain

$$\begin{aligned} & \mathbb{E} \left[ \sup_{\boldsymbol{\theta} \in \Theta} \left\| \partial_{\boldsymbol{\theta}} \Phi_{LARCH}(U, \boldsymbol{\theta}) - \partial_{\boldsymbol{\theta}} \Phi_{LARCH}(V, \boldsymbol{\theta}) \right\| \right] \\ & \leq g \left( \sup_{i \geq 1} \{ \|U_i\|_4 \|V_i\|_4 \} \right) \left( |U_1 - V_1| + \sum_{i=1}^{\infty} \left( \sup_{\boldsymbol{\theta} \in \Theta} |a_i(\boldsymbol{\theta})| + \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta})\| \right) \|U_{i+1} - V_{i+1}\|_4 \right). \end{aligned}$$

This inequality implies that  $\partial_{\boldsymbol{\theta}}\Phi_{LARCH} \in \text{Lip}_4(\Theta)$  under the assumptions of Proposition 4.4.

We also have to establish that under conditions of Proposition 4.4,

$$\mathbb{E}\left[\left\|\partial_{\boldsymbol{\theta}}\Phi_{LARCH}\left((\tilde{X}_k(u))_{k \leq 0}, \boldsymbol{\theta}^*(u)\right)\right\|^2\right] < \infty$$

and  $\mathbb{E}\left[\left\|\partial_{\boldsymbol{\theta}^2}^2\Phi_{LARCH}\left((\tilde{X}_{-k}(u))_{k \in \mathbb{N}}, \boldsymbol{\theta}^*(u)\right)\right\|\right] < \infty$ . Indeed we have

$$\begin{aligned} \partial_{\boldsymbol{\theta}}\Phi_{LARCH}\left((\tilde{X}_k(u))_{k \leq 0}, \boldsymbol{\theta}^*(u)\right) &= -4(\xi_0^2 - 1)\left(a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right)^3 \\ &\quad \times \left(\partial_{\boldsymbol{\theta}}a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} \partial_{\boldsymbol{\theta}}a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right). \end{aligned} \quad (96)$$

Therefore, using Hölder and Minkowski Inequalities and independence of  $\xi_0$  and  $(\tilde{X}_k(u))_{k \leq -1}$  together with  $\mathcal{E}_u \equiv \mathbb{E}\left[\left\|\partial_{\boldsymbol{\theta}}\Phi_{LARCH}\left((\tilde{X}_k(u))_{k \leq 0}, \boldsymbol{\theta}^*(u)\right)\right\|^2\right]$  we derive that

$$\begin{aligned} \mathcal{E}_u &\leq 16 \mathbb{E}\left[(\xi_0^2 - 1)^2\right] \left(\mathbb{E}\left[\left(a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right)^8\right]\right)^{3/4} \\ &\quad \times \left(\mathbb{E}\left[\left\|\partial_{\boldsymbol{\theta}}a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} \partial_{\boldsymbol{\theta}}a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right\|^8\right]\right)^{1/4} \\ &\leq C \left(\sup_{\boldsymbol{\theta} \in \Theta} |a_0(\boldsymbol{\theta})| + \|\tilde{X}_0(u)\|_8 \sum_{i=1}^{\infty} \sup_{\boldsymbol{\theta} \in \Theta} |a_i(\boldsymbol{\theta})|\right)^6 \\ &\quad \times \left(\sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}}a_0(\boldsymbol{\theta})\| + \|\tilde{X}_0(u)\|_8 \sum_{i=1}^{\infty} \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}}a_i(\boldsymbol{\theta})\|\right)^2. \end{aligned}$$

Thus we obtain  $\mathbb{E}\left[\left\|\partial_{\boldsymbol{\theta}}\Phi_{LARCH}\left((\tilde{X}_k(u))_{k \leq 0}, \boldsymbol{\theta}^*(u)\right)\right\|^2\right] < \infty$  with  $r = 8$  under suitable conditions on  $(a_j)_j$ . The expression for the second derivatives is also derived

$$\begin{aligned} \partial_{\boldsymbol{\theta}^2}^2\Phi_{LARCH}\left((\tilde{X}_k(u))_{k \leq 0}, \boldsymbol{\theta}^*(u)\right) &= -4\left(a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right)^2 \\ &\times \left\{(\xi_0^2 - 3)\left(\partial_{\boldsymbol{\theta}}a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} \partial_{\boldsymbol{\theta}}a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right)\left(\partial_{\boldsymbol{\theta}}a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} \partial_{\boldsymbol{\theta}}a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right)\right\}' \\ &+ \left(a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right)\left(\partial_{\boldsymbol{\theta}^2}^2a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} \partial_{\boldsymbol{\theta}^2}^2a_i(\boldsymbol{\theta}^*(u)) \tilde{X}_{-i}(u)\right)\}. \end{aligned}$$

As a consequence, similar arguments as previously entail

$$\mathbb{E}\left[\left\|\partial_{\boldsymbol{\theta}^2}^2\Phi_{LARCH}\left((\tilde{X}_k(u))_{k \leq 0}, \boldsymbol{\theta}^*(u)\right)\right\|\right] < \infty$$

from Hausdorff and Minkowski inequalities. Finally we checked the conditions of Theorem 3.2 since the asymptotic covariance matrix  $\Sigma(\boldsymbol{\theta}^*(u))$  and  $\Gamma(\boldsymbol{\theta}^*(u))$  are positive definite matrix from (56) using (96).  $\square$

*Acknowledgments.* The authors wish to thank Rainer Dahlhaus, Thomas Mikosch and Lionel Truquet for many inspiring discussions.

The work of the second author was funded by CY Initiative of Excellence (grant "Investissements d'Avenir" ANR- 16-IDEX-0008), Project "EcoDep" PSI-AAP2020-0000000013.

## References

- [1] P. Doukhan, O. Wintenberger, Weakly dependent chains with infinite memory, *Stochastic Processes and Applications* 118 (2008) 1997–2013.
- [2] R. Dahlhaus, S. Richter, W. Wu, Towards a general theory for nonlinear locally stationary processes, *Bernoulli* 25 (2019) 1013–1044.
- [3] R. Dahlhaus, On the Kullback-Leibler information divergence of locally stationary processes, *Stochastic Processes and Applications* 62 (1996) 139–168.
- [4] R. Dahlhaus, Locally stationary processes, in: T. Subba Rao, S. Subba Rao, C. Rao (Eds.), *Time Series Analysis: Methods and Applications*, Vol. 30 of *Handbook of statistics*, Elsevier, 2012.
- [5] R. Dahlhaus, S. Subba Rao, Statistical Inference for Time Varying ARCH processes, *The Annals of Statistics* 34 (2006) 1075–1114.
- [6] R. Dahlhaus, W. Polonik, Nonparametric quasi-maximum likelihood estimation for gaussian locally stationary processes, *The Annals of Statistics* 34 (2006) 2790–2824.
- [7] K. Paraschakisa, R. Dahlhaus, Frequency and phase estimation in time series with quasi periodic components., *Journal of Time Series Analysis* 33 (2012) 13–31.
- [8] R. Dahlhaus, Fitting time series models to nonstationary processes, *Annals of Statistics* 25 (1997) 1–37.
- [9] R. Dahlhaus, A likelihood approximation for locally stationary processes, *Annals of Statistics* 28 (2000) 1762–1794.
- [10] R. Dahlhaus, Empirical spectral processes for locally stationary time series, *Bernoulli* 15 (2009) 1–39.
- [11] J.-M. Bardet, O. Wintenberger, Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes, *Ann. Statist.* 37 (2009) 2730–2759.

- [12] J. Pfanzagl, On the mesurability and consistency of minimum contrast estimates, *Metrika* 14 (1969) 249–334.
- [13] M. Rosenblatt, *Stochastic curve estimation*, Vol. 3, NSF-CBMS Regional Conference Series in Probability and Statistics, 1991.
- [14] J.-M. Bardet, P. Doukhan, Non-parametric estimation of time varying AR(1)-processes with local stationarity and periodicity, *Electronic Journal of Statistics* 12 (2) (2018) 2323–2354.
- [15] J.-M. Bardet, W. Kengne, O. Wintenberger, Multiple breaks detection in general causal time series using penalized quasi-likelihood, *Elec. Journal Statist.* (2012) 435–477.
- [16] R. Azrak, G. Mélard, Asymptotic properties of quasi-maximum likelihood estimators for arma models with time-dependent coefficients, *Statistical Inference for Stochastic Processes* 9 (2006) 279–330.
- [17] C. Francq, J.-M. Zakoïan, *GARCH Models: Structure, Statistical Inference and Financial Applications*, Wiley, 2010.
- [18] L. Giraitis, P. Kokoszka, R. Leipus, Stationary arch models: dependence structure and central limit theorem, *Econometric Theory* 1 (16) (2000) 3–22.
- [19] N. Rohan, A time varying GARCH() model and related statistical inference, *Statistics & Probability Letters* 83 (2013) 1983–1990.
- [20] L. Truquet, Parameter stability and semiparametric inference in time-varying arch models, *Journal of Royal Statistical Society Series B* 79 (2017) 1391–1414.
- [21] P. Robinson, Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression, *Journal of Econometrics* 47 (1) (1991) 67–84.
- [22] L. Giraitis, R. Leipus, P. Robinson, D. Surgailis, LARCH, leverage and long memory, *Journal of Financial Econometrics* 2 (2) (2003) 177–210.
- [23] C. Francq, J.-M. Zakoïan, Inconsistency of the mle and inference based on weighted ls for larch models, *Journal of Econometrics* 159 (2010) 151–165.
- [24] P. Doukhan, L. Truquet, A fixed point approach to model random fields, *ALEA Lat. Am. J. Probab. Math. Stat.* 3 (2007) 111–132.
- [25] D. Straumann, *Estimation in conditionally heteroscedastic time series models*, Vol. 181 of *Lecture Notes in Statistics*, Springer-Verlag, Berlin, 2005.
- [26] P. Billingsley, *Convergence of probability measures*, Second edition, Wiley and sons, 1999.