

Contrast estimation of time-varying in finite 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 in finite memory processes. 2021. hal-02586009v2

HAL Id: hal-02586009

https://hal.science/hal-02586009v2

Preprint submitted on 16 Oct 2020 (v2), last revised 15 Jun 2021 (v3)

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.

Submitted to Bernoulli

Contrast estimation of general locally stationary processes using coupling

JEAN-MARC BARDET [1], PAUL DOUKHAN [2] and OLIVIER WINTENBERGER [3]

[1] *University Paris 1 Panthéon-Sorbonne, SAMM, France, E-mail:* bardet@univ-paris1.fr

[2] *University Cergy-Pontoise, AGM, France, E-mail:* paul.doukhan@u-cergy.fr

[3] *Sorbonne University, LPSM, France, E-mail:* olivier.wintenberger@upmc.fr

This paper aims at providing statistical guarantees for a kernel based estimation of time varying parameters driving the dynamic of very generals classes of local stationary processes. We consider coupling arguments in order to extend the results of Dahlhaus *et al.* [13] to the local stationary version of the infinite memory processes in Doukhan and Wintenberger [21]. The estimators are computed as localized M-estimators of any contrast satisfying appropriate regularity conditions. We prove the uniform consistency and 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 local-stationary processes as ARMA, AR(∞), GARCH, ARCH(∞), ARMA-GARCH, LARCH(∞), ..., and integer valued processes are also considered. Numerical experiments demonstrate the efficiency of the estimators on both simulated and real data sets.

MSC 2010 subject classifications: Primary 62G05, 62G20; secondary 62M05.

1. Introduction

Following the seminal paper of Dahlhaus [7] local-stationarity is considered as a natural set of conditions for introducing non-stationarity in times series. The chapter of Dahlhaus [11] yields an exhaustive survey about new results obtained between 1992 and 2012 on this topics. Dahlhaus and its co-authors have developed a consistent framework studying definitions and properties of local stationary time-varying models (See [14], [12] and [27] for instance) as well as associated statistical issues such as identification and estimation (See [8], [9], [14], [12], [10] and [27]). Note that, except in Dahlhaus and Subba-Rao [14], 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 the important paper of Dahlhaus et al. [13], allowing to get rid off the linearity condition on the models. Time non-homogeneous Markovian observations are the considered non-linear time varying models. Under some contraction condition, the observations forget their own past exponentially fast. Anyway many important models like GARCH-type models are not

taken in 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, do not fall into the setting considered by Dahlhaus et al. [13]. 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. [13] to such an infinite memory setting. We first extend the concept of local stationarity beyond the Markovian case. More precisely, the stationary models introduced by Doukhan and Wintenberger in [21] are extended to time varying infinite memory causal processes defined as $(X_t^{(n)})_{1\le t\le n}$ a recursive solution to the equation

$$X_t^{(n)} = F_{\boldsymbol{\theta}_t^{(n)}}\big((X_{t-k}^{(n)})_{k\in\mathbb{N}^*};\xi_t\big), \qquad 1\le t\le n, \qquad n\in\mathbb{N}^*=\mathbb{N}\setminus\{0\}$$

where $(\boldsymbol{\theta}_t^{(n)})_{0\le t\le n,\, n\in\mathbb{N}^*}$ is a deterministic family with $\boldsymbol{\theta}_t^{(n)}\in\Theta\subset\mathbb{R}^d$ for any $0\le t\le n$ and $n\in\mathbb{N}^*$, $F_{\boldsymbol{\theta}}$ is a known real-valued function and the innovations ξ_t constitute an independent and identically distributed (i.i.d.) sequence. For ease of writing we will consider $X_t^{(n)}=0$ for $t\le 0$, but the arbitrary choice of any deterministic initial values does not change the asymptotic behavior. In order to make it tractable, a secondary aim of the paper is to keep the conditions as simple as possible. For instance, in our setting local-stationarity consists simply in the existence of a function $\boldsymbol{\theta}^*$ such as for $u\in(0,1)$, $\boldsymbol{\theta}_t^{(n)}\simeq\boldsymbol{\theta}^*(u)$ when $t/n\simeq u$ (see the more precise condition (7) in the so-called Assumption (LS(ρ))). Under this assumption, we define a kernel based estimator $\widehat{\boldsymbol{\theta}}(u)$ of $\boldsymbol{\theta}^*(u)$ obtained by the minimization of a localized sum of contrast Φ (see its definition in (11)). 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 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. Processes with infinite memory may provide sharp conditions of convergence of estimation procedures in such models. Indeed, as any contrast is a function of the observations only, the contrast itself has infinite memory (see Bardet and Wintenberger [5] for a detailed discussion in the stationary setting). We first consider least squares and least absolute values contrasts time-varying LARCH(∞) processes and we notably obtained an efficient asymptotic estimation for these infinite memory chains. Considering quasi log-likelihood contrast also offers numerous estimation results for time-varying finite or infinite memory processes: we obtain the uniform consistency and the asymptotic normality for time varying AR(∞) and ARCH(∞). For finite memory time varying ARMA(p,q) or GARCH(p,q), our results recover previous ones of Dahlhaus and co-authors [8], [14] and [13], valid only for ARMA and ARCH models, and extend them to the important GARCH model class. Dahlhaus et

al. [13] used functional dependence conditions, new coupling arguments developed from Dedecker and Prieur [16] are at the basis of the work, since no strong mixing condition may be expected and longer memories than exponential decaying ones have to be tackled. One advantage of our approach is that it goes around the derivative process, a drawback is that the rates of convergence in the asymptotic normality are sub-optimal. Finally we also apply our strategy to prove estimation convergence for time varying ARMA-GARCH processes and time varying integer valued Poisson-GLM type 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 non-parametric 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 typical 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 nonparametric 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 V is a vector space then $V^{\infty} = \{(x_n)_{n\in\mathbb{N}} \in V^{\mathbb{N}}; \exists N \in \mathbb{N},\ x_k = 0,\ \text{for all } k > N\}$;
- 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: $\|Z\|_p = \left[\mathbb{E}(\|Z\|^p)\right]^{1/p}$;
- 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 Θ denotes a subset of $\mathbb{R}^d$, and $\overset{\circ}{\Theta}$ is the interior of Θ. If V is a Banach space then $\mathcal{C}(\Theta, V)$ denotes the Banach space of V-valued continuous functions on Θ 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 $\mathbb{E}\left[\|f\|_\Theta^p\right] < \infty$.
- For $\boldsymbol{\theta} \in \Theta$ and $\Psi_{\boldsymbol{\theta}} :\ \mathbb{R}^\infty \to V$ a Borel function with values in a finite dimensional vector space V, $\partial^k_{\boldsymbol{\theta}} \Psi_\theta(x)$ denotes respectively for $k = 0, 1, 2$, in case they exist, $\Psi_{\boldsymbol{\theta}}(x)$, $\partial \Psi_{\boldsymbol{\theta}}(x)/\partial \boldsymbol{\theta}$ and $\partial^2 \Psi_{\boldsymbol{\theta}}(x)/\partial \boldsymbol{\theta}^2$ 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 $\boldsymbol{\theta} \in \mathbb{R}^d$ and let a function $F_{\boldsymbol{\theta}}$ be defined as follows

$$F_{\boldsymbol{\theta}} : \ (x, y) \in \mathbb{R}^{\infty} \times \mathbb{R} \mapsto F_{\boldsymbol{\theta}}(x, y) \in \mathbb{R}. \tag{1}$$

Doukhan and Wintenberger proved in [21] the existence and uniqueness of the stationary solution of the recurrence equation

$$X_t = F_{\boldsymbol{\theta}}\big((X_{t-k})_{k\in\mathbb{N}^*}, \xi_t\big), \qquad \text{for all } t \in \mathbb{Z}, \tag{2}$$

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 $F_{\boldsymbol{\theta}}$. As a consequence, we define the following family of assumptions $\mathbf{(A}_k\mathbf{(\Theta))}$ for $k = 0, 1, 2$ and some compact subset Θ of $\mathbb{R}^d$:

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

- $$C_k(\Theta) = \Big\| \sup_{\boldsymbol{\theta}\in\Theta} \big\|\partial^k_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(0, \xi_0)\big\| \Big\|_p < \infty \tag{3}$$
- $$\Big\| \sup_{\boldsymbol{\theta}\in\Theta} \big\|\partial^k_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(x, \xi_0) - \partial^k_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(y, \xi_0)\big\| \Big\|_p \leq \sum_{j=1}^{\infty} b_j^{(k)}(\Theta)\, \|x_j - y_j\|_p, \tag{4}$$

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

Thus, from [21], under the uniform contraction conditions $\mathbf{(A}_0\mathbf{(\Theta))}$ with $B_0(\Theta) < 1$, there exists a unique stationary solution of (2) in $\mathbb{L}^p$ (defined almost surely).

2.3. Local stationary 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 $\mathbf{(A}_0\mathbf{(\Theta))}$ with $B_0(\Theta) < 1$ ensure the existence of a non-stationary $\mathbb{L}^p$-process. More precisely, we define the triangular array $(X_t^{(n)})_{1\leq t\leq n,\, n\in\mathbb{N}^*}$ such as:

$$X_t^{(n)} = F_{\boldsymbol{\theta}_t^{(n)}}\big((X_{t-k}^{(n)})_{k\in\mathbb{N}^*}; \xi_t\big), \qquad 1 \leq t \leq n,\, n \in \mathbb{N}^*\,, \tag{5}$$

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

$$X_t^{(n)} = 0, \quad \text{for } t \le 0. \tag{6}$$

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 (6), the nonstationary triangular array $(X_t^{(n)})_{0\le t\le n,\, n\in\mathbb{N}^*}$, solution of (5), remains in $\mathbb{L}^p$ and it satisfies*

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

2.4. The stationary version

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

Assumption (LS(ρ)): *There exist $K_{\boldsymbol{\theta}} > 0$ and a continuous function $\boldsymbol{\theta}^* :\ u \in [0,1] \mapsto \boldsymbol{\theta}^*(u) \in \mathbb{R}^d$, such as*

$$\left\|\boldsymbol{\theta}_t^{(n)} - \boldsymbol{\theta}^*(u)\right\| \le K_{\boldsymbol{\theta}} \left|u - \frac{t}{n}\right|^{\rho}, \quad \textit{for any } \ n \in \mathbb{N}^* \textit{ and } 1 \le t \le n. \tag{7}$$

This condition describes an Hölder type behavior for the approximation of $\boldsymbol{\theta}_t^{(n)}$ by the function $\boldsymbol{\theta}^*$. When $t/n \underset{n\to+\infty}{\longrightarrow} u$ then $\boldsymbol{\theta}_t^{(n)} \underset{n\to+\infty}{\longrightarrow} \boldsymbol{\theta}^*(u)$ and the behavior of $(X_t^{(n)})$ is similar to its so-called stationary version.

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

$$\widetilde{X}_t(u) = F_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{t-k}(u))_{k\ge1}, \xi_t\big), \qquad t \in \mathbb{Z}\,, \tag{8}$$

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

Note that from [21] the existence of the stationary version is implied by 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 is such that $\mathbf{(A_0(\Theta))}$ holds with $B_0(\Theta) < 1$, then there exists a.s. a unique stationary stationary version $(\widetilde{X}_t(u))_{t\in\mathbb{Z}}$ satisfying (8) and

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

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 [5]. Let (X_t) be a stationary solution of the infinite memory model (2) with parameter $\boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^d$ such that $\mathbf{(A_0(\Theta))}$ holds with $B_0(\Theta) < 1$. We estimate $\boldsymbol{\theta}^*$ using an M-estimator based on an observed path $(X_1, \dots, X_n)$.

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

Space $\mathrm{Lip}_p(\Theta)$: *A Borel function* $h : \mathbb{R}^\infty \times \Theta \to \mathbb{R}$ *belongs to* $Lip_p(\Theta)$ *if there exist 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* $g : [0,\infty)^2 \to [0,\infty)$ *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* $\sup_{s\ge1}\{\|U_s\|_p \vee \|V_s\|_p\} < \infty$*, one obtains:*

$$\begin{cases} \mathbb{E}\big[\sup_{\boldsymbol{\theta}\in\Theta} \big|h(0,\boldsymbol{\theta})\big|\big] < \infty; \\ \mathbb{E}\big[\sup\limits_{\boldsymbol{\theta}\in\Theta} \big|h(U,\boldsymbol{\theta}) - h(V,\boldsymbol{\theta})\big|\big] \le g\big(\sup\limits_{s\ge1}\big\{\|U_s\|_p \vee \|V_s\|_p\big\}\big) \sum\limits_{s=1}^\infty \alpha_s(h,\Theta)\,\|U_s - V_s\|_p. \end{cases} \tag{9}$$

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

Assumption $\mathbf{(Co(\Phi,\Theta))}$: *The function* $\Phi \in Lip_p(\Theta)$ *for* $p \ge 1$ *is such that for* $(X_t)_{t\in\mathbb{Z}}$ *satisfying the infinite memory model (2) with parameter* $\boldsymbol{\theta}^* \in \overset{\circ}{\Theta}$ *and with* $\mathcal{F}_0 = \sigma\big((X_{-k})_{k\in\mathbb{N}}\big)$*,*

$$\boldsymbol{\theta}^* \text{ is the unique minimum of the function } \boldsymbol{\theta} \in \Theta \mapsto \mathbb{E}\big[\Phi\big((X_{1-k})_{k\in\mathbb{N}}, \boldsymbol{\theta}\big)|\mathcal{F}_0\big] \text{ in } \overset{\circ}{\Theta}. \tag{10}$$

As a consequence, this condition is depending on the function $\boldsymbol{\theta} \in \Theta \mapsto F_{\boldsymbol{\theta}}(\cdot)$ driving the infinite memory model (2). Therefore the contrast function Φ will be chosen with respect to this function $F_{\boldsymbol{\theta}}$. This is thus natural to define the M-estimator of $\boldsymbol{\theta}$ by

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

Note that the factor $1/n$ aims at establishing a Law of Large Numbers. Indeed, if the

almost sure convergence holds, *i.e.*

$$\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{n}\sum_{t=1}^{n}\Phi((X_{t-i})_{i\in\mathbb{N}},\boldsymbol{\theta})-\mathbb{E}\big[\Phi((X_{-i})_{i\in\mathbb{N}},\boldsymbol{\theta})\big]\Big| \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} 0,$$

then usual arguments imply $\widehat{\boldsymbol{\theta}}_n \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} \boldsymbol{\theta}^*$.

3.2. The local stationary case

We extend the notion of contrast function Φ to the non-stationary process $\left(X_t^{(n)}\right)_{t\in\mathbb{Z}}$ 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 (5) under condition* $\mathbf{(A_0(\Theta))}$ *with $B_0(\Theta)<1$ and let $\Phi\in Lip_p(\Theta)$ with $p\geq 1$. Then for any $\boldsymbol{\theta}\in\Theta$, the sequence of contrasts $\big(\Phi\big((X_{t-k}^{(n)})_{k\in\mathbb{N}},\boldsymbol{\theta}\big)\big)_{t\in\mathbb{Z}}$ exists in $\mathbb{L}^1$. Moreover, under Assumption* $\mathbf{(LS(\rho))}$*, and with $(\widetilde{X}_t(u))_{t\in\mathbb{Z}}$ the stationary version defined in (8), $\big(\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)\big)_{t\in\mathbb{Z}}$ is a stationary ergodic process.*

Under the local stationary assumption $\mathbf{(LS(\rho))}$, we can expect estimating $\boldsymbol{\theta}^*(u)$ with $0<u<1$ defined in (7) 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 $K:\mathbb{R}\to\mathbb{R}$ be such as:*

- *K has a compact support, i.e. there exists $c>0$ such as $K(x)=0$ for $|x|\geq c$;*
- *$K:\mathbb{R}\to\mathbb{R}$ is piecewise differentiable with $\int_{\mathbb{R}}K(x)dx=1$, $C_K=\sup_{x\in\mathbb{R}}|K(x)|<\infty$.*

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

$$\widehat{\boldsymbol{\theta}}(u)=\underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\,\frac{1}{nh_n}\sum_{j=1}^{n}\Phi\big((X_{j-i}^{(n)})_{i\in\mathbb{N}},\boldsymbol{\theta}\big)\,K\Big(\frac{\frac{j}{n}-u}{h_n}\Big),\qquad u\in(0,1). \tag{11}$$

Under weak conditions, this estimator is uniformly consistent:

Theorem 3.1. *Let $(X_t^{(n)})_{t\in\mathbb{N}}$ be the solution of the non-stationary infinite memory model (5) which satisfies Assumption* $\mathbf{(A_0(\Theta))}$ *with $B_0(\Theta)<1$ and $\sum_{t=2}^{\infty}t\log(t)b_t(\Theta)<\infty$ and Assumption* $\mathbf{(A_1(\Theta))}$*, with also Assumption* $\mathbf{(LS(\rho))}$*. If for $p\geq 1$, $\Phi\in Lip_p(\Theta)$*

with $\sum_{s\geq 0} s\,\alpha_s(\Phi,\Theta) < \infty$ *satisfying Assumption* $(\mathbf{Co}(\Phi,\Theta))$ *then, for any* $u \in (0,1)$, $\widehat{\boldsymbol{\theta}}(u)$ *consistently estimates* $\boldsymbol{\theta}^*(u)$*:*

$$\widehat{\boldsymbol{\theta}}(u) \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} \boldsymbol{\theta}^*(u), \quad \text{if} \quad h_n \underset{n\to+\infty}{\longrightarrow} 0 \quad \text{and} \quad nh_n \underset{n\to+\infty}{\longrightarrow} \infty.$$

Moreover, if $p > 1$ *and* $n^{1-1/p}h_n \underset{n\to+\infty}{\longrightarrow} \infty$ *then for any* $\varepsilon > 0$ *then* $\widehat{\boldsymbol{\theta}}$ *uniformly consistently estimates* $\boldsymbol{\theta}^*$*:*

$$\sup_{u\in[\varepsilon,1-\varepsilon]} \left\|\widehat{\boldsymbol{\theta}}(u) - \boldsymbol{\theta}^*(u)\right\| \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0\,. \tag{12}$$

Remark 3.1. *We notice that the uniform consistency of the kernel estimator was already obtained for Markov processes by Dahlhaus et al. in [13] under a different set of assumptions. Our extra condition on the bandwidth* $n^{1-1/p}h_n \to \infty$, $n \to \infty$, *corresponds to the extra condition in* (ii) *of Theorem 5.2 of [13] with* $M = 0$*. We notice also that our extra condition implicitly requires that* $p > 1$*.*

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

$$B_0(\{\boldsymbol{\theta}^*(u)\}) < 1\,, \qquad u \in (0,1)\,. \tag{13}$$

Then a continuity argument under $\mathbf{(A_0(\Theta))}$ *yeilds 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 (11) *with* $\{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*(u)\| \leq \varepsilon\}$ *in place of* Θ *consistently estimates* $\theta^*(u)$ *by an application of Theorem 3.1. That this M-estimator coincides with* $\widehat{\boldsymbol{\theta}}(u)$ *for* n *sufficiently large follows by a compacity argument of Pfanzagl [28]; The compact set* Θ *is covered by finitely many compact sets of diameter* ε *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* (11) *belongs to the unique compact set containing* $\boldsymbol{\theta}^*(u)$ *for* n *large enough. See [5] for details. Note that we were not able to extend the uniform consistency of* $\widehat{\boldsymbol{\theta}}$ *under pointwise contraction, which is consistent with [13] that worked under uniform contraction as well. Uniform contraction seems neceassary for proving uniform moments 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 Φ and the integrability of its derivatives. We have:

Theorem 3.2. *Let the assumptions of Theorem 3.1 hold with* Θ *a compact set. Assume also 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* $\partial_{\boldsymbol{\theta}}\Phi \in Lip_p(\Theta)$ *with* $\sum_{s=1}^{\infty} s\,\alpha_s(\partial_{\boldsymbol{\theta}}\Phi,\Theta) < \infty$*, and for any* $u \in (0,1)$*, with* $\rho \in (0,1]$ *defined in Assumption* $\mathbf{(LS(}\rho\mathbf{))}$*,*

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

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

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

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

$$nh_n \underset{n\to+\infty}{\longrightarrow} \infty \quad \textit{and} \quad nh_n^{1+2\rho} \underset{n\to+\infty}{\longrightarrow} 0, \tag{14}$$

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

$$\sqrt{nh_n}\,\big(\widehat{\boldsymbol{\theta}}(u)-\boldsymbol{\theta}^*(u)\big) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}_d\Big(0\,,\,\Gamma^{-1}(\boldsymbol{\theta}^*(u))\Sigma\big(\boldsymbol{\theta}^*(u)\big)\Gamma^{-1}(\boldsymbol{\theta}^*(u))\Big). \tag{15}$$

Remark 3.3. *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* $\rho\in(0,1]$*. 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 [31] and Dahlhaus* et al. *[13] derive expressions for an equivalent of the bias; in this case i.e.* $nh_n^{1+2\rho} \underset{n\to+\infty}{\longrightarrow} \ell\neq 0$*, one may use the classical minimax bandwidth and, then, a non-centred Gaussian limit theorem occurs.*

Remark 3.4. *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 and not the uniform consistency of* $\widehat{\boldsymbol{\theta}}$*. Thus the asymptotic normality (15) can be extending under the pointwise contraction (13).*

Remark 3.5. *Considering* $(u_1,\ldots,u_m)$ *instead of* u*, a multidimensional central limit theorem could also be obtained extending (15). 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 locally stationary infinite memory models with contrast functions $\Phi\in \mathrm{Lip}_p(\Theta)$ 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 tvAR(1)) processes defined by

$$X_t^{(n)} = \theta_t^{(n)} X_{t-1}^{(n)} + \xi_t, \quad \text{for } 1 \le t \le n,\, n \in \mathbb{N}^*, \tag{16}$$

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

4.1.1. Least Square contrast

When Φ_{LS} is the Least Square (LS) contrast defined as

$$\Phi_{LS}(x, \theta) = (x_1 - \theta\, x_2)^2\,, \tag{17}$$

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

$$\mathbb{E}\big[|\sup_{\theta\in\Theta}\big|\Phi_{LS}(U,\theta)-\Phi_{LS}(V,\theta)\big|\big] \le (1+r)\max_{1\le s\le 2}\big\{\big\|U_i\|_2\vee\big\|V_i\|_2\big\}\big(\|U_1-V_1\|_2+r\|U_2-V_2\|_2\big),$$

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

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

After elementary algebra, we obtain:

$$\mathbb{E}\Big[\sup_{\theta\in[-r,r]} \big|\partial_\theta \Phi_{LS}(U,\theta) - \partial_\theta \Phi_{LS}(U,\theta)\big|\Big] \le 4\left(\|U_1 - V_1\|_2 + \|U_2 - V_2\|_2\right)\left(\|U_1\|_2 + \|V_1\|_2 + \|U_2\|_2 + \|V_2\|_2\right)$$

from Hölder's inequality. Analogously,

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

ensuring that $\partial_\theta \Phi_{LS}$ and $\partial^2_{\theta^2} \Phi_{LS}$ are included in $\mathrm{Lip}_2\big([-r, r]\big)$. If $\mathbb{E}[\xi_0^2] < \infty$, both the matrix $\Sigma\big(\theta^*(u)\big) = 4\left(\int_{\mathbb{R}} K^2(x)dx\right)\sigma_\xi^4(1-\theta^*(u)^2)^{-1}$ 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* **(LS(ρ))**, *the localized least square estimator is asymptotically normal when the sequence $(h_n)_n$ satisfies (14) and we obtain for any $u \in (0, 1)$*

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

Here, we recover for $0 < \rho \le 1$ the results on tvAR(1) models obtained by Bardet and Doukhan in [3], which are also valid for functions $u \in (0, 1) \to \widehat{\theta}(u)$ with Hölderian derivatives.

4.1.2. Least Absolute Value contrast

In the framework of tvAR(1) processes (16) 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 $0 < r < 1$,

$$\Phi_{LAV}(x, \theta) = \big|x_1 - \theta\, x_2\big|. \tag{18}$$

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

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

implying $\alpha_1(\Phi_{LAV}, \Theta) = 1$ and $\alpha_2(\Phi_{LAV}, \Theta) = r$ and $\alpha_j(\Phi_{LAV}, \Theta) = 0$ for $j \ge 3$. Since Φ_{LAV} is not a differentiable function, we will restrict our purpose to the uniform consistency of $\widehat{\theta}(u)$. 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* **(LS(ρ))**, *then the localized LAV estimators $\widehat{\theta}_n$ satisfies* (12).

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 [5] as

$$X_t = M_{\boldsymbol{\theta}}\big((X_{t-i})_{i\ge1}\big)\, \xi_t + f_{\boldsymbol{\theta}}\big((X_{t-i})_{i\ge1}\big), \qquad \text{for any } t \in \mathbb{Z}, \boldsymbol{\theta} \in \Theta \tag{19}$$

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

$$\sup_{\boldsymbol{\theta}\in\Theta} \big|K_{\boldsymbol{\theta}}(x) - K_{\boldsymbol{\theta}}(y)\big| \le \sum_{i=1}^{\infty} \beta_i(K, \Theta)\, \big|x_i - y_i\big|, \tag{20}$$

for any x, $y \in \mathbb{R}^{\infty}$. Then, (X_t) satisfies the infinite memory model (2) 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

$$b_j^{(0)}(\Theta) \le \beta_j(f, \Theta) + \|\xi_0\|_p\, \beta_j(M, \Theta) \quad \text{for any } j \in \mathbb{N}^*. \tag{21}$$

Therefore, (X_t) is a stationary and $\mathbb{L}^p$ solution of the causal affine model (19) when

$$\sum_{j=1}^{\infty} \beta_j(f, \Theta) + \|\xi_0\|_p\, \beta_j(M, \Theta) < 1. \tag{22}$$

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

$$\Phi_G(x,\boldsymbol{\theta}) = \log\left(M^2_{\boldsymbol{\theta}}\big((x_i)_{i\geq 2}\big)\right) + \frac{\big(x_1 - f_{\boldsymbol{\theta}}\big((x_i)_{i\geq 2}\big)\big)^2}{M^2_{\boldsymbol{\theta}}\big((x_i)_{i\geq 2}\big)}. \tag{23}$$

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 [5] (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.* [4] (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

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

hence, there exists a function g such that

$$\begin{aligned}\mathbb{E}\big[\sup_{\theta\in\Theta}\big|\Phi_G(U,\boldsymbol{\theta}) - \Phi_G(V,\boldsymbol{\theta})\big|\big] \\ \leq g\big(\sup_{i\geq 1}\big\{\|U_i\|_3 \vee \|V_i\|_3\big\}\big)\Big(\|U_1-V_1\|_3 + \sum_{i=2}^{\infty} b^{(0)}_k(\Theta)\,\|U_i-V_i\|_3\Big),\end{aligned} \tag{24}$$

using Hölder inequality and with $b^{(0)}_k(\Theta) = \beta_k(f_{\boldsymbol{\theta}},\Theta) + \|\xi_0\|_p\,\beta_k(M_{\boldsymbol{\theta}},\Theta)$ the Lipschitz coefficients of the function $F_{\boldsymbol{\theta}}$ given in $\mathbf{(A_0(\Theta))}$. Therefore, according to (9) with $\alpha_k(\Phi_G,\Theta) = b^{(0)}_k(\Theta)$ for $k\geq 2$ and $\alpha_1(\Phi_G,\Theta)=1$, we check that $\Phi_G \in \mathrm{Lip}_3(\Theta)$ since $\mathbf{(A_0(\Theta))}$ holds and $B_0(\Theta) = \sum_k b^{(0)}_k(\Theta) < \infty$.
Now we consider a time varying causal affine processes, that is the local stationary extension of causal affine processes defined in (19), *i.e.*

$$X^{(n)}_t = M_{\boldsymbol{\theta}^{(n)}_t}\big((X^{(n)}_{t-i})_{1\leq i}\big)\,\xi_t + f_{\boldsymbol{\theta}^{(n)}_t}\big((X^{(n)}_{t-i})_{1\leq i}\big), \qquad \text{for any } t\in\mathbb{Z}, \tag{25}$$

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

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

Proposition 4.1. *Let $(X^{(n)}_t)$ satisfy (25) where $f_{\boldsymbol{\theta}}$, $M_{\boldsymbol{\theta}}$, $\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}$, $\partial_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}}$, $\partial^2_{\boldsymbol{\theta}^2} f_{\boldsymbol{\theta}}$ and $\partial^2_{\boldsymbol{\theta}^2} M_{\boldsymbol{\theta}}$ satisfy Lipschitz inequalities (20) and under Assumption $\mathbf{(LS(\rho))}$. Assume also:*

1. $\|\xi_0\|_4 < \infty$ *where the probability distribution of* ξ_0 *is absolutely continuous with respect to the Lebesgue measure and* Θ *is a bounded set included in* $\big\{\boldsymbol{\theta} \in \mathbb{R}^d,\ \sum_{j=1}^{\infty} \big(\beta_j(f_{\boldsymbol{\theta}}, \{\boldsymbol{\theta}\}) + \|\xi_0\|_4\, \beta_j(M_{\boldsymbol{\theta}}, \{\boldsymbol{\theta}\})\big) < 1\big\}$;
2. *there exists* $\underline{M} > 0$ *such as* $M_{\boldsymbol{\theta}} \geq \underline{M}$ *for any* $\boldsymbol{\theta} \in \Theta$;
3. *For all* $\boldsymbol{\theta}$, $\boldsymbol{\theta'} \in \Theta$,

$$\big(f_{\boldsymbol{\theta}} = f_{\boldsymbol{\theta'}} \quad \text{and} \quad M_{\boldsymbol{\theta}} = M_{\boldsymbol{\theta'}}\big) \implies \boldsymbol{\theta} = \boldsymbol{\theta'}; \tag{26}$$

4. *We have*

$$\Big(\sum_{j=1}^{d} \mu_j \frac{\partial}{\partial \boldsymbol{\theta}_j} f_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}}\big) = 0 \quad a.s. \implies \mu_j = 0,\ j = 1, \ldots, d\Big),$$

$$\text{or} \quad \Big(\sum_{j=1}^{d} \mu_j \frac{\partial}{\partial \boldsymbol{\theta}_j} M_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}}\big) = 0 \quad a.s. \implies \mu_j = 0,\ j = 1, \ldots, d\Big). \tag{27}$$

Consider $\Phi = \Phi_G$ *as* (-2) *times the Gaussian conditional log-density* (23). *Then, if*

$$\sum_{j=1}^{\infty} j\left(\log j\left(\beta_j(f, \Theta) + \beta_j(M, \Theta)\right) + j\left(\beta_j(\partial_{\boldsymbol{\theta}} f, \Theta)\right)\right.$$
$$\left. + \sum_{j=1}^{\infty}\big(\beta_j(\partial_{\boldsymbol{\theta}} M, \Theta)\big) + \beta_j(\partial^2_{\boldsymbol{\theta}^2} f, \Theta) + \beta_j(\partial^2_{\boldsymbol{\theta}^2} M, \Theta)\big) < \infty, \tag{28}$$

with $\beta_j(\cdot, \Theta)$ *defined in* (20), *the localized QMLE* $\widehat{\boldsymbol{\theta}}(u)$ *satisfies the central limit theorem* (15).

Note that the QML contrast Φ 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 [5] 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, tvAR(∞), tvARCH(∞) and tvARMA-GARCH models.

4.2.1. Time varying AR(∞) and time varying ARMA(p, q) processes

In such the case of time varying AR(∞) (or tvAR(∞)) or time varying (invertible) ARMA(p, q) processes, we have $M_{\boldsymbol{\theta}} = \sigma(\boldsymbol{\theta}) > \underline{\sigma} > 0$ and $f_{\boldsymbol{\theta}}((x_i)_{i\geq 1}) = \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})\, x_j$,

where $(a_j(\boldsymbol{\theta}))_{j\geq 1}$ is a sequence of real numbers, implying

$$X_t^{(n)} = \sigma(\boldsymbol{\theta}_t^{(n)})\,\xi_t + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}_t^{(n)})\,X_{t-j}^{(n)}, \quad \text{for } 1\leq t\leq n,\ n\in\mathbb{N}^*, \tag{29}$$

with $X_t^{(n)}=0$ for any $t\leq 0$. Thus, the Lipschitz coefficients satisfy $\beta_j(f,\Theta)=\sup_{\boldsymbol{\theta}\in\Theta}|a_j(\boldsymbol{\theta})|$ and $\beta_j(M,\Theta)=0$. Then we obtain the asymptotic normality of $\widehat{\boldsymbol{\theta}}(u)$ from primitive conditions on functions a_j and σ by an application of Proposition 4.1:

Corollary 4.3. *Let $(X_t^{(n)})$ be a tvAR(∞) process defined in (29). If $\|\xi_0\|_4<\infty$, let Θ be a bounded subset of $\mathbb{R}^d$ included in $\big\{\boldsymbol{\theta}\in\mathbb{R}^d,\ \sum_{j=1}^{\infty}a_j(\boldsymbol{\theta})<1\big\}$. If for each $j\in\mathbb{N}^*$ the functions $\boldsymbol{\theta}\in\Theta\mapsto a_j(\boldsymbol{\theta})\in\mathbb{R}$ and $\boldsymbol{\theta}\in\Theta\mapsto\sigma(\boldsymbol{\theta})\in[\underline{\sigma},\infty)$ are $\mathcal{C}^2(\Theta)$ functions such as $\big(a_j(\boldsymbol{\theta})=a_j(\boldsymbol{\theta}'),\ \forall j\in\mathbb{N}^*$ and $\sigma(\boldsymbol{\theta})=\sigma(\boldsymbol{\theta}')\big)$ imply $(\boldsymbol{\theta}=\boldsymbol{\theta}')$, if $\boldsymbol{\theta}_t^{(n)}$ satisfies the assumption of local stationarity* **(LS(ρ))**, *and if*

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

then the central limit (15) holds for any $u\in(0,1)$ under condition (14).

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 $\kappa\geq\underline{\kappa}>2$, $\mu\leq(\sum_{j=1}^{\infty}j^{-\underline{\kappa}})^{-1}$ and $\sigma\geq\underline{\sigma}>0$. Then the previous corollary implies the asymptotic normality (15) of $\big(\widehat{\mu}(u),\widehat{\kappa}(u),\widehat{\sigma}(u)\big)$ under condition (14) when $\big(\mu_t^{(n)},\kappa_t^{(n)},\sigma_t^{(n)}\big)$ satisfies Assumption **(LS(ρ))**.

An important subclass of tvAR(∞) models is the one of an invertible tvARMA(p,q) models defined as

$$X_t^{(n)}+\phi_{1,t}^{(n)}\,X_{t-1}^{(n)}+\cdots+\phi_{p,t}^{(n)}\,X_{t-p}^{(n)}=\sigma_t^{(n)}\xi_t+\psi_{1,t}^{(n)}\,\xi_{t-1}+\cdots+\psi_{q,t}^{(n)}\,\xi_{t-q} \tag{30}$$

(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 [7]. We consider the set of parameters $\boldsymbol{\theta}_t^{(n)}=\big(\phi_{1,t}^{(n)},\ldots,\phi_{p,t}^{(n)},\psi_{1,t}^{(n)},\ldots,\psi_{q,t}^{(n)},\sigma_t^{(n)}\big)'$ and the subset $\Theta_{ARMA}^{(p,q)}$ of $\mathbb{R}^d$ with $d=p+q+1$ defined by:

$$\Theta_{ARMA}^{(p,q)}=\Big\{(\phi_1,\ldots,\phi_p,\psi_1,\ldots,\psi_q,\sigma)\in\mathbb{R}^{p+q+1},\\ 1+\phi_1\,z+\cdots+\phi_p\,z^p\neq 0 \text{ and } 1+\psi_1\,z+\cdots+\psi_q\,z^q\neq 0 \text{ for all } |z|\leq 1\Big\}.$$

Then if $\boldsymbol{\theta}_t^{(n)}\in\Theta$ for any $1\leq t\leq n$, $n\in\mathbb{N}^*$ and Θ a compact subset of $\Theta_{ARMA}^{(p,q)}$, then $\sup_{n,t}\|X_t^{(n)}\|_p<\infty$ for any $p\geq 1$ when $\|\xi_0\|_p<\infty$ since $X_t^{(n)}$ can be written

as a tvAR(∞) process (29) with finite sum of absolute values of coefficients. Moreover, from classical analytic arguments it is well known that the corresponding Lipschitz coefficients $\beta_j(f,\Theta)$, $\beta_j(\partial_{\boldsymbol{\theta}} f,\Theta)$ and $\beta_j(\partial^2_{\boldsymbol{\theta}^2} f,\Theta)$ decrease exponentially fast so that the condition (28) is automatically satisfied.

As a consequence of Proposition 4.1 we obtain:

Corollary 4.4. *If $(X_t^{(n)})$ is a tvARMA(p,q) process defined in (30), Θ is a bounded subset of $\Theta^{(p,q)}_{ARMA}$, if $\|\xi_0\|_4 < \infty$ and $\boldsymbol{\theta}_t^{(n)}$ satisfies the assumption of local stationarity* **(LS(ρ))**, *which is implied by the local stationarity* **(LS(ρ))** *of 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 (15) holds for any $u\in(0,1)$ under condition (14).*

This result is a QMLE version of the results obtained by Dahlhaus in [9] for Gaussian tvARMA processes (using Whittle likelihood approximation) and by Azrak and Mélard in [1]; we use quasi likelihood contrasts similarly as those authors.

4.2.2. Time varying ARCH(∞) and time varying GARCH(p,q) processes

Time varying ARCH(∞) (tvARCH(∞)) or time varying GARCH(p,q) (tvGARCH(p,q)) processes correspond to $f_{\boldsymbol{\theta}}((x_i)_{i\geq 1})=0$ and $M_{\boldsymbol{\theta}}((x_i)_{i\geq 1}) = \left(a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})\, x_j^2\right)^{1/2}$, where $(a_j(\boldsymbol{\theta}))_{j\geq 0}$ is a sequence of non negative real numbers and $a_0(\cdot)\geq \underline{a}>0$ implying

$$X_t^{(n)} = \xi_t\left(a_0(\boldsymbol{\theta}_t^{(n)}) + \sum_{i\geq 1} a_i(\boldsymbol{\theta}_t^{(n)})\big(X_{t-i}^{(n)}\big)^2\right)^{1/2} \quad \text{for } 1\leq t\leq n,\ n\in\mathbb{N}^*, \tag{31}$$

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 **(LS(ρ))**. For more details about stationary ARCH(∞) or GARCH(p,q) processes, or for the transition from GARCH(p,q) to ARCH(∞), see [22]. We are going to specify again the conditions of Proposition 4.1 in such a case. Firstly, we consider Lipschitz properties on $\left((X_t^{(n)})^2\right)_t$ rather than $(X_t^{(n)})_t$ as in [5] in order to recover natural constraint on the parameters set Θ (similar as the moments condition of [24])

$$\Theta \text{ is a compact subset of } \left\{\boldsymbol{\theta}\in\mathbb{R}^d,\ \|\xi_0\|_4^2\sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})<1\right\}. \tag{32}$$

Secondly, the Lipschitz coefficients of Φ, $\partial_{\boldsymbol{\theta}}\Phi$ and $\partial^2_{\boldsymbol{\theta}^2}\Phi$ can be expressed in terms of $|U_i^2-V_i^2|$ following the same computations than in (24) 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 $\left(M_{\boldsymbol{\theta}}((U_i)_{i\geq 1})\times M_{\boldsymbol{\theta}}((V_i)_{i\geq 1})\right)^{-1}$ appears in the function g, we deduce that Φ, $\partial_{\boldsymbol{\theta}}\Phi$ and $\partial^2_{\boldsymbol{\theta}^2}\Phi$ are respectively included in $\mathrm{Lip}_3(\Theta)$, $\mathrm{Lip}_4(\Theta)$ and $\mathrm{Lip}_4(\Theta)$ with coefficients $\alpha_s(\cdot,\Theta)$ defined in (9) satisfying for $s\geq 2$,

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

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

Corollary 4.5. *Let $(X_t^{(n)})$ be a tvARCH(∞) process defined in (31). We assume that $\|\xi_0\|_4 < \infty$ and we consider Θ a compact subset of $\left\{\boldsymbol{\theta} \in \mathbb{R}^d,\ \|\xi_0\|_4^2 \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta}) < 1\right\}$. If for $j \in \mathbb{N}^*$ the functions $\boldsymbol{\theta} \in \Theta \mapsto a_j(\boldsymbol{\theta}) \in [0,\infty)$ and $\boldsymbol{\theta} \in \Theta \mapsto a_0(\boldsymbol{\theta}) \in [\underline{a},\infty)$ are $\mathcal{C}^2(\Theta)$ functions such that $\left(a_j(\boldsymbol{\theta}) = a_j(\boldsymbol{\theta}'),\ \forall j \in \mathbb{N}\right)$ implies $(\boldsymbol{\theta} = \boldsymbol{\theta}')$, if $\boldsymbol{\theta}_t^{(n)}$ satisfies the assumption of local stationarity* **(LS(ρ))**, *and if*

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

then the localized QMLE $\widehat{\boldsymbol{\theta}}(u)$ is asymptotically normal and (15) holds for any $u \in (0,1)$ and any (h_n) satisfying (14).

To the best of our knowledge, this result is new. In [14] the existence of tvARCH(∞) processes has been studied and the asymptotic normality has been obtained for tvARCH(p) processes.

We specialized the previous result to the cases where $(X_t^{(n)})$ is a tvGARCH(p,q) process. We assume that $\|\xi_0\|_4 < \infty$ and we consider the model

$$\left\{\begin{array}{rcl} X_t^{(n)} & = & \sigma_t^{(n)}\,\xi_t \\ \left(\sigma_t^{(n)}\right)^2 & = & c_{0,t}^{(n)} + c_{1,t}^{(n)}\left(X_{t-1}^{(n)}\right)^2 + \cdots + c_{p,t}^{(n)}\left(X_{t-p}^{(n)}\right)^2 + d_{1,t}^{(n)}\left(\sigma_{t-1}^{(n)}\right)^2 + \cdots + d_{q,t}^{(n)}\left(\sigma_{t-q}^{(n)}\right)^2 \end{array}\right. \tag{33}$$

where $(c_{i,t}^{(n)})_{0\le i\le p}$ and $(d_{j,t}^{(n)})_{1\le j\le q}$ are non negative real number for any $1 \le t \le n$, $n \in \mathbb{N}^*$, with $X_t^{(n)} = 0$ for any $t \le 0$. Consider $\boldsymbol{\theta}_t^{(n)} = \left(c_{0,t}^{(n)}, c_{1,t}^{(n)}, \ldots, c_{p,t}^{(n)}, d_{1,t}^{(n)}, \ldots, d_{q,t}^{(n)}\right)'$ with $c_{0,t}^{(n)} \ge \underline{c} > 0$ and the subset $\Theta_{GARCH}^{(p,q)}$ of $\mathbb{R}^{p+q+1}$ defined by:

$$\Theta_{GARCH}^{(p,q)} = \left\{(c_0, c_1, \ldots, c_p, d_1, \ldots, d_q) \in \mathbb{R}^{p+q+1},\ \sum_{j=1}^{q} d_j + \|\xi_0\|_4^2 \sum_{i=1}^{p} c_i < 1\right\}.$$

If $\boldsymbol{\theta}_t^{(n)} \in \Theta$ for any $1 \le t \le n$, $n \in \mathbb{N}^*$ with Θ a bounded set included in $\Theta_{GARCH}^{(p,q)}$ then we have $\sup_{t,n} \|X_t^{(n)}\|_4 < \infty$. Moreover, $(X_t^{(n)})_t$ can be written as a tvARCH(∞) (see [22]). Moreover the coefficients $a_i(\boldsymbol{\theta}_t^{(n)})$ in the tvARCH(∞) decrease exponentially fast. As $\alpha_s(\Phi,\Theta)$, $\alpha_s(\partial_{\boldsymbol{\theta}}\Phi,\Theta)$ and $\alpha_s(\partial^2_{\boldsymbol{\theta}^2}\Phi,\Theta)$ can be expressed from $a_s(\cdot)$ and their derivatives, which are also exponentially decreasing, this implies the following corollary:

Corollary 4.6. *Let $(X_t^{(n)})$ be a tvGARCH(p,q) process defined in (33) and Θ be a bounded set included in $\Theta_{GARCH}^{(p,q)}$ where $\|\xi_0\|_4 < \infty$. If $\boldsymbol{\theta}_t^{(n)}$ satisfies the assumption*

of local stationarity **(LS(ρ))**, *which is implied by the local stationarity* **(LS(ρ))** *on all functions* $c_{0,t}^{(n)},\dots,c_{p,t}^{(n)},d_{1,t}^{(n)},\dots,d_{q,t}^{(n)}$, *then the central limit* (15) *holds for any* $u\in(0,1)$ *under condition* (14).

This result can be compared for instance with those of [14] for tvARCH(p), which are obtained under the same procedure but under the condition $\|\xi_0\|_{4(1+\delta)}<\infty$, or those of [30] for tvGARCH(p,q), which are obtained from a local polynomial estimation and under the condition $\|\xi_0\|_8<\infty$. Note that [33] also obtained asymptotic normality under very sharp conditions in a special case of tvARCH(p) process.

4.2.3. Time varying ARMA(p,q)-GARCH(p',q') processes

This model was introduced in the stationary framework by [17] and developed in [26]. The model consists on a tvARMA(p,q) where the pure white noise (ξ_t) is replaced by a weak white noise $(\varepsilon_t^{(n)})$ that is a tvGARCH(p',q') process, *i.e.* $(X_t^{(n)})$ is defined by

$$\left\{\begin{array}{rcl} X_t^{(n)} & = & -\phi_{1,t}^{(n)}\,X_{t-1}^{(n)}-\cdots-\phi_{p,t}^{(n)}\,X_{t-p}^{(n)}+\varepsilon_t^{(n)}+\psi_{1,t}^{(n)}\,\varepsilon_{t-1}^{(n)}+\cdots+\psi_{q,t}^{(n)}\,\varepsilon_{t-q}^{(n)}\\ \varepsilon_t^{(n)} & = & \sigma_t^{(n)}\,\xi_t\\ \left(\sigma_t^{(n)}\right)^2 & = & c_{0,t}^{(n)}+c_{1,t}^{(n)}\left(\varepsilon_{t-1}^{(n)}\right)^2+\cdots+c_{p',t}^{(n)}\left(\varepsilon_{t-p'}^{(n)}\right)^2+d_{1,t}^{(n)}\left(\sigma_{t-1}^{(n)}\right)^2+\cdots+d_{q',t}^{(n)}\left(\sigma_{t-q'}^{(n)}\right)^2 \end{array}\right. \tag{34}$$

for any $1\le t\le n$, $n\in\mathbb{N}^*$, with $X_t^{(n)}=0$ for any $t\le 0$. As previously, $c_{0,t}^{(n)}\ge\underline{c}>0$ and $(c_{j,t}^{(n)})$ is a family of non-negative real numbers. Consider

$$\boldsymbol{\theta}_t^{(n)}=\left(\phi_{1,t}^{(n)},\dots,\phi_{p,t}^{(n)},\psi_{1,t}^{(n)},\dots,\psi_{q,t}^{(n)},c_{0,t}^{(n)},c_{1,t}^{(n)},\dots,c_{p',t}^{(n)},d_{1,t}^{(n)},\dots,d_{q',t}^{(n)}\right)'.$$

If $\|\xi_0\|_4<\infty$, define the subset $\Theta_{ARMAGARCH}^{(p,q,p',q')}$ of $\mathbb{R}^d$ with $d=p+q+p'+q'+1$ by

$$\Theta_{ARMAGARCH}^{(p,q,p',q')}=\Big\{\boldsymbol{\theta}\in\mathbb{R}^d,\ \sum_{j=1}^{q'}d_j+\|\xi_0|_4^2\sum_{j=1}^{p'}c_j<1,\\ \text{and}\quad\left(1+\sum_{j=1}^{p}\phi_j z^j\right)\left(1+\sum_{j=1}^{q}\psi_j z^j\right)\neq 0\ \text{for all}\ |z|\le 1\Big\}.$$

Then, if Θ is a bounded set included in $\Theta_{ARMAGARCH}^{(p,q,p',q')}$ then the GARCH(p',q') process $(\varepsilon_t^{(n)})_t$ satisfies $\sup_{t,n}\|\varepsilon_t^{(n)}\|_4<\infty$ (see previously) when $\boldsymbol{\theta}_t^{(n)}\in\Theta$ for any $1\le t\le n$, $n\in\mathbb{N}^*$. Moreover, $X_t^{(n)}$ can be written as a linear filter of $(\varepsilon_t^{(n)})_t$ when the ARMA coefficients satisfies the condition required in $\Theta_{ARMAGARCH}^{(p,q,p',q')}$, and these coefficients decrease exponentially fast. Therefore, when $\boldsymbol{\theta}_t^{(n)}\in\Theta$ for any $1\le t\le n$, $n\in\mathbb{N}^*$ with Θ a bounded set included in $\Theta_{ARMAGARCH}^{(p,q,p',q')}$, then $\sup_{t,n}\|X_t^{(n)}\|_4<\infty$.
Moreover, following Lemma 2.1. of [2], we know that a stationary ARMA(p,q)-GARCH(p',q') process is a stationary affine causal process with functions $f_{\boldsymbol{\theta}}$ and $M_{\boldsymbol{\theta}}$ satisfying the Lipschitz condition (20) with Lipschitz coefficients decreasing exponentially fast, as well as

their derivatives. This is also the same case for a time varying ARMA(p,q)-GARCH(p',q') process. Therefore, we obtain the following result:

Corollary 4.7. *Let $(X_t^{(n)})$ be a time varying ARMA(p,q)-GARCH(p',q') process defined in (34) with $\|\xi_0\|_4 < \infty$ and Θ be a bounded set included in $\Theta_{ARMAGARCH}^{(p,q,p',q')}$. Moreover, if $\boldsymbol{\theta}_t^{(n)}$ satisfies the assumption of local stationarity* **(LS(ρ))**, *which is implied by the same local stationarity property satisfied by $\phi_{1,t}^{(n)},\dots,\phi_{p,t}^{(n)}$, $\psi_{1,t}^{(n)},\dots,\psi_{q,t}^{(n)}$, $c_{0,t}^{(n)},c_{1,t}^{(n)},\dots,c_{p',t}^{(n)}$, $d_{1,t}^{(n)},\dots,d_{q',t}^{(n)}$, then the localized QMLE is asymptotically normal as (15) holds for any $u\in(0,1)$ and (h_n) satisfying (14).*

4.3. Time varying LARCH(∞) processes and LS-contrast

Here we consider a LARCH(∞) process introduced by Robinson in [29] and studied intensively by Giraitis *et al.* in [25]. The model is defined as

$$X_t = \xi_t\left(a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})\, X_{t-j}\right) \quad \text{for any } t\in\mathbb{Z}, \tag{35}$$

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 $a_0(\boldsymbol{\theta})\geq 0$ 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 $\boldsymbol{\theta}\in\Theta$,

$$\|\xi_0\|_r \sum_{j=1}^{\infty} |a_j(\boldsymbol{\theta})| < 1. \tag{36}$$

Even if a LARCH(∞) 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 close to 0 and this does not allow asymptotic results for such contrasts (see more details in Francq and Zakoïan [23]). Even if 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

$$\Phi_{LARCH}(x,\boldsymbol{\theta}) = \left(x_1^2 - \left(a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})\, x_{j+1}\right)^2\right)^2. \tag{37}$$

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

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

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

$$\Big(a_0(\boldsymbol{\theta}) + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})\, X_{1-j} = a_0(\boldsymbol{\theta}') + \sum_{j=1}^{\infty} a_j(\boldsymbol{\theta})\, X_{1-j} \ a.s.\Big) \implies \boldsymbol{\theta} = \boldsymbol{\theta}', \tag{38}$$

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

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

Hence

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

and therefore $\Phi_{LARCH} \in \mathrm{Lip}_4(\Theta)$ 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 (36).

We consider now the time varying LARCH(∞) process defined by:

$$X_t^{(n)} = \xi_t \Big(a_0(\boldsymbol{\theta_t^{(n)}}) + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta_t^{(n)}})\, X_{t-i}^{(n)}\Big), \qquad \text{for any } t \in \mathbb{Z}, \tag{39}$$

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 Θ is a bounded subset of the set

$$\Theta_{LARCH} = \Big\{\boldsymbol{\theta} \in \mathbb{R}^d,\ \|\xi_0\|_8 \sum_{i=1}^{\infty} |a_i(\boldsymbol{\theta})| < \infty\Big\}. \tag{40}$$

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

Proposition 4.2. *Assume that* $\boldsymbol{\theta} \in \mathbb{R}^d \mapsto a_j(\boldsymbol{\theta}) \in \mathbb{R}$ *are* $\mathcal{C}^2$ *functions for any* $j \in \mathbb{R}$, $a_0(\cdot) \geq 0$ *and*

1. $\|\xi_0\|_8 < \infty$ *where the probability distribution of* ξ_0 *is absolutely continuous with respect to the Lebesgue measure and* Θ *is a bounded set included in* Θ_{LARCH}*;*
2. *For all* $\boldsymbol{\theta}$, $\boldsymbol{\theta'} \in \Theta$,

$$(a_i(\boldsymbol{\theta}) = a_i(\boldsymbol{\theta'}),\ \textit{for all } i \in \mathbb{N}) \implies (\boldsymbol{\theta} = \boldsymbol{\theta'}); \tag{41}$$

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

$$(\partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta}) = \partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta'}),\ \textit{for all } i \in \mathbb{N}) \implies (\boldsymbol{\theta} = \boldsymbol{\theta'}). \tag{42}$$

Let $(X_t^{(n)})$ *be a tvLARCH process defined following* (39) *where* $\boldsymbol{\theta}_t^{(n)}$ *satisfies Assumption* **(LS(**ρ**))**. *Consider* $\Phi = \Phi_{LARCH}$ *as in* (37). *If*

$$\sum_{j=1}^{\infty} j \log j \sup_{\boldsymbol{\theta} \in \Theta} |a_j(\boldsymbol{\theta}| + j \sup_{\boldsymbol{\theta} \in \Theta} \|\partial_{\boldsymbol{\theta}} a_j(\boldsymbol{\theta}\| + \sup_{\boldsymbol{\theta} \in \Theta} \|\partial^2_{\boldsymbol{\theta}^2} a_j(\boldsymbol{\theta}\| < \infty, \tag{43}$$

then $\widehat{\boldsymbol{\theta}}(u)$ *is asymptotically normal as* (15) *for any* $u \in (0,1)$ *and* (h_n) *satisfying* (14).

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 time varying GLARCH(p,q) process, natural extension of stationary GLARCH(p,q) processes (see for instance [25]) is also interesting and straightforward:

Corollary 4.8. *If* $\|\xi_0\|_8 < \infty$ *and the probability distribution of* ξ_0 *is absolutely continuous with respect to the Lebesgue measure and if* $(X_t^{(n)})$ *is a tvGLARCH*(p,q) *process defined by*

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

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

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

which satisfies the assumption of local stationarity **(LS(**ρ**))**. *Then the central limit* (15) *holds for any* $u \in (0,1)$ *under condition* (14).

This result is due to the exponential decay of the sequences $(\alpha_s(\Phi_{LARCH}, \Theta))_s$ and $(\alpha_s(\partial_{\boldsymbol{\theta}} \Phi_{LARCH}, \Theta))_s$ in such as case (see [25]).

4.4. Time varying integer valued processes and Poisson QMLE

Finally, we consider the integer valued process (X_t) defined so that the conditional distribution of X_t is a Poisson distribution with parameter $\lambda_{\boldsymbol{\theta}}\big((X_{t-i})_{i\geq 1}\big)$, *i.e.*

$$X_t \mid ((X_{t-i})_{i\geq 1}) \overset{\mathcal{L}}{\sim} \mathcal{P}(\lambda_{\boldsymbol{\theta}}\big((X_{t-i})_{i\geq 1}\big)), \quad \text{for any } t\in\mathbb{Z}, \tag{44}$$

where for any $\boldsymbol{\theta}\in\Theta, U\in\mathbb{R}^{\infty}\mapsto\lambda_{\boldsymbol{\theta}}(U)\in[\underline{\lambda},\infty)$, $\underline{\lambda}>0$, is once again a Lipschitz function on Θ satisfying (20) with Lipschitz coefficients $(\beta_i(\lambda_{\boldsymbol{\theta}},\Theta))_{i\geq 1}$ such that $\sum_{i=1}^{\infty}\beta_i(\lambda_{\boldsymbol{\theta}},\Theta)<1$ (see for instance Doukhan and Kengne [19]).
Then, we consider as a contrast the opposite of the log-likelihood of the process, *i.e.*

$$\Phi_P(x,\theta) = -x_1\,\log\big(\lambda_{\boldsymbol{\theta}}\big((x_i)_{i\geq 2}\big)\big) + \lambda_{\boldsymbol{\theta}}\big((x_i)_{i\geq 2}\big). \tag{45}$$

Then, after classical computations we obtain for $p=2$:

$$\begin{aligned}
&\mathbb{E}\big[\sup_{\boldsymbol{\theta}\in\Theta}\big|\Phi_P(U,\boldsymbol{\theta})-\Phi_P(V,\boldsymbol{\theta})\big|\big]\\
&\quad\leq C\Big(\sup_{i\geq 1}\big\{\|U_i\|_2\vee\|V_i\|_2\big\}\,\|U_1-V_1\|_2+(1+\|V_1\|_2^2)^{1/2}\sum_{i=2}^{\infty}\beta_i(\lambda_{\boldsymbol{\theta}},\Theta)\,\|U_i-V_i\|_2\Big)\\
&\qquad\leq g\big(\sup_{i\geq 1}\big\{\|U_i\|_2\vee\|V_i\|_2\big\}\big)\sum_{i=1}^{\infty}\alpha_i(\Phi,\Theta)\,\|U_i-V_i\|_2,
\end{aligned}$$

with $\alpha_1(\Phi_P,\Theta)=1$ and $\alpha_i(\Phi_P,\Theta)=\beta_{i-1}(\lambda_{\boldsymbol{\theta}},\Theta)$ for $i\geq 2$, inducing $\Phi_P\in\mathrm{Lip}_2(\Theta)$.

We extend the structural recursive equation (44) and Φ_P to the time varying framework as follows. This example shows that our results apply to integer valued local stationary processes, which is an original and interesting extension. Hence, we consider

$$X_t^{(n)} \mid (X_{t-i}^{(n)})_{i\geq 1} \overset{\mathcal{L}}{\sim} \mathcal{P}\Big(\lambda_{\boldsymbol{\theta}_t^{(n)}}\big((X_{t-i}^{(n)})_{i\geq 1}\big)\Big), \quad \text{for any } 1\leq t\leq n, \text{ and all } n\in\mathbb{N}^*, \tag{46}$$

where $X_t^{(n)}=0$ for $t\leq 0$. If $\lambda_{\boldsymbol{\theta}}(\cdot)$ satisfies the uniform Lipschitz property (20) with Lipschitz coefficients $(\beta_i(\lambda,\Theta))_{i\geq 1}$, we define

$$\Theta_{\mathcal{P}} = \Big\{\boldsymbol{\theta}\in\mathbb{R}^d,\ \sum_{i=1}^{\infty}\beta_i(\lambda,\{\boldsymbol{\theta}\})<1\Big\}. \tag{47}$$

Then, using Theorem 2.1 of Doukhan *et al.* [18] and Lemma 2.1, we deduce that if $\boldsymbol{\theta}\in\Theta_{\mathcal{P}}$ for any $1\leq t\leq n$ and $n\in\mathbb{N}^*$, then $\sup_{t,n}\|X_t^{(n)}\|_p<\infty$, for any $p\geq 1$. Using the Poisson QMLE defined by (45), we obtain the following asymptotic result:

Proposition 4.3. *Let $(X_t^{(n)})$ satisfy (25) where $\lambda_{\boldsymbol{\theta}}$, $\partial_{\boldsymbol{\theta}}\lambda_{\boldsymbol{\theta}}$ and $\partial^2_{\boldsymbol{\theta}^2}\lambda_{\boldsymbol{\theta}}$ satisfy Lipschitz inequalities (20) and under Assumption* **(LS(ρ))**. *Assume also:*

1. Θ *is a bounded set included in* $\Theta_{\mathcal{P}}$*;*
2. *there exists* $\underline{\lambda} > 0$ *such as* $\lambda_{\boldsymbol{\theta}} \geq \underline{\lambda}$ *for any* $\boldsymbol{\theta} \in \Theta$*;*
3. *For all* $\boldsymbol{\theta}$, $\boldsymbol{\theta'} \in \Theta$, $\lambda_{\boldsymbol{\theta}} = \lambda_{\boldsymbol{\theta'}}$ *implies* $\boldsymbol{\theta} = \boldsymbol{\theta'}$*;*
4. *For any* $i = 1, \ldots, d$, $\frac{\partial}{\partial \boldsymbol{\theta}_i} \lambda_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}}\big) \neq 0$ *a.s.*

Consider $\Phi = \Phi_P$ *as it was defined in* (45)*. Then, if*

$$\sum_{j=1}^{\infty} (j\ \log j)\, \beta_j(\lambda, \Theta) + j\, \beta_j(\partial_{\boldsymbol{\theta}} \lambda, \Theta) + \beta_j(\partial^2_{\boldsymbol{\theta}^2} \lambda, \Theta) < \infty, \tag{48}$$

for any $u \in (0,1)$ *and under condition* (14)*.*

$$\sqrt{n\, h_n}\big(\widehat{\boldsymbol{\theta}}(u) - \boldsymbol{\theta}^*(u)\big) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\Big(0, E\Big[\frac{\partial_{\boldsymbol{\theta}} \lambda_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}}\big)^t \partial_{\boldsymbol{\theta}} \lambda_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}}\big)}{\lambda^2_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}}\big)}\Big]\Big). \tag{49}$$

This central limit theorem can notably be applied for:

- Time varying integer-valued GARCH(p,q) processes (tvINGARCH(p,q) processes), where

$$\lambda_{\boldsymbol{\theta}_t^{(n)}}\big((X^{(n)}_{t-j})_{j\geq 1}\big) = a^{(n)}_{0,t} + \sum_{i=1}^{p} a^{(n)}_{i,t}\, X^{(n)}_{t-i} + \sum_{i=1}^{q} b^{(n)}_{i,t}\, \lambda_{\boldsymbol{\theta}_t^{(n)}}\big((X^{(n)}_{t-i-j})_{j\geq 1}\big), \tag{50}$$

  for any $1 \leq t \leq n$ and $n \in \mathbb{N}^*$, with $X^{(n)}_t = 0$ for all $t \leq 0$. Here

$$\boldsymbol{\theta}^{(n)}_t = \big(a^{(n)}_{0,t}, \ldots, a^{(n)}_{p,t}, b^{(n)}_{1,t}, \ldots, b^{(n)}_{q,t}\big).$$

  Here, following [18], it is possible to consider a sharper set of parameters than the one given by (47); hence let Θ be a bounded subset of

$$\Theta_{INGARCH} = \Big\{\boldsymbol{\theta} = (a_0, a_1, \ldots, a_p, b_1, \ldots, b_q) \in \mathbb{R}^{p+q+1},\ \sum_{i=1}^{p} a_i + \sum_{i=1}^{q} b_i < 1\Big\}. \tag{51}$$

  In such a case, the Lipschitz coefficients $\beta_j(\lambda, \Theta)$, $\beta_j(\partial_{\boldsymbol{\theta}} \lambda, \Theta)$ and $\beta_j(\partial^2_{\boldsymbol{\theta}^2} \lambda, \Theta)$ exist and decrease exponentially fast (see [18]), and all the conditions are satisfied for obtaining (49).
- Integer valued threshold GARCH(p,q) processes where, with ℓ a positive fixed integer,

$$\begin{aligned}
&\lambda_{\boldsymbol{\theta}_t^{(n)}}\big((X^{(n)}_{t-j})_{j\geq 1}\big) \\
&= a^{(n)}_{0,t} + \sum_{i=1}^{p} a^{(n)}_{i,t}\, \lambda_{\boldsymbol{\theta}_t^{(n)}}\big((X^{(n)}_{t-i-j})_{j\geq 1}\big) + \sum_{i=1}^{q} \big(b^{(n)}_{i,t}\ \max(X^{(n)}_{t-i} - \ell, 0) - c^{(n)}_{i,t}\ \min(X^{(n)}_{t-i}, \ell)\big),
\end{aligned}$$

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

$$\boldsymbol{\theta}_t^{(n)} = \big(a_{0,t}^{(n)}, \ldots, a_{p,t}^{(n)}, b_{1,t}^{(n)}, \ldots, b_{q,t}^{(n)}, c_{1,t}^{(n)}, \ldots, c_{q,t}^{(n)}\big).$$

As previously, the Lipschitz coefficients $\beta_j(\lambda, \Theta)$, $\beta_j(\partial_{\boldsymbol{\theta}}\lambda, \Theta)$ and $\beta_j(\partial^2_{\boldsymbol{\theta}^2}\lambda, \Theta)$ exist and decrease exponentially fast (see [18]), and all the conditions are satisfied for obtaining (49).

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 $n\,h_n^3 \underset{n\to+\infty}{\longrightarrow} 0$, 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 tvGARCH(1, 1). Here, with the notation of equation (33), assume:

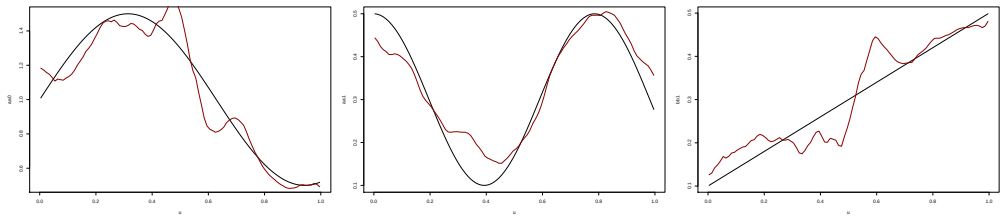
$$c_{0,t}^{(n)} = 1 + 0.5\,\sin\big(5\,\frac{t}{n}\big), \quad c_{1,t}^{(n)} = 0.1 + 0.4\cos^2\big(4\,\frac{t}{n}\big) \quad \text{and} \quad d_{1,t}^{(n)} = 0.1 + 0.4\,\frac{t}{n},$$

for any $1 \leq t \leq n$ and $n \in \mathbb{N}^*$. Clearly, $c_0^*(u) = 1 + 0.5\,\sin(5\,u)$, $c_1^*(u) = 0.1 + 0.4\,\cos^2(4\,u)$ and $d_1^*(u) = 0.1 + 0.4\,u$. Moreover, we assume that (ξ_t) is a sequence of i.i.d.r.v. following $\mathcal{N}(0, 1)$ distribution.
We independently replicated 1000 trajectories of such process, for $n = 2000,\ 5000$ and 10000 and computed the Gaussian QMLE estimators $\widehat{c}_0(u)$, $\widehat{c}_1(u)$ and $\widehat{d}_1(u)$ for $u = k/50$ with $k = 1, \ldots, 49$. 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 $\widehat{\theta}^U(u)$ and $\widehat{\theta}^E(u)$.

Table 1 contains the results of these Monte Carlo experiments where we computed the square root of the mean integrated squared error (RSMISE). 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 $\widehat{c}_0^E(u)$, $\widehat{c}_1^E(u)$ and $\widehat{d}_1^E(u)$ when $n = 5000$.

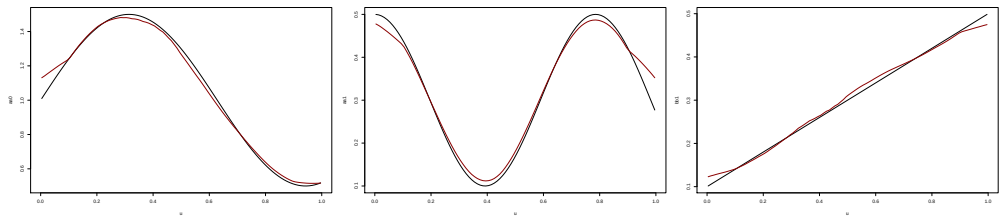
	c_0		c_1		d_1	
n	$\widehat{c}_0^U$	$\widehat{c}_0^E$	$\widehat{c}_1^U$	$\widehat{c}_1^E$	$\widehat{d}_1^U$	$\widehat{d}_1^E$
1000	0.493	0.455	0.126	0.122	0.230	0.208
3000	0.363	0.323	0.081	0.077	0.167	0.146
10000	0.259	0.224	0.052	0.048	0.118	0.101

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



(a) Trajectory of $\widehat{c}_0^E$. (b) Trajectory of $\widehat{c}_1^E$. (c) Trajectory of $\widehat{d}_1^E$.

Figure 1: Paths of functions c_0, c_1, d_1 (in black), and a path of $\widehat{c}_0^E$, $\widehat{c}_1^E$ and $\widehat{d}_1^E$ (in red) for $n = 10000$



(a) Mean of $\widehat{c}_0^E$. (b) Mean of $\widehat{c}_1^E$. (c) Mean of $\widehat{d}_1^E$.

Figure 2: Paths of functions c_0, c_1, d_1 (in black), and the mean trajectories over 1000 replications of $\widehat{c}_0^E$, $\widehat{c}_1^E$ and $\widehat{d}_1^E$ (in red) for $n = 5000$

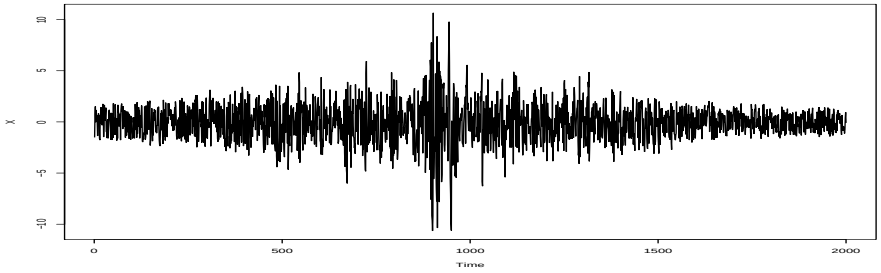


Figure 3: On trajectory of a tvARCH(∞)-process for $n = 2000$

n	$\widehat{c}_0^U$	$\widehat{c}_0^E$	$\widehat{c}_1^U$	$\widehat{c}_1^E$	$\widehat{p}^U$	$\widehat{p}^E$
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 $\widehat{c}_0^U$, $\widehat{c}_0^E$, $\widehat{c}_1^U$, $\widehat{c}_1^E$, $\widehat{p}^U$ and $\widehat{p}^E$ for the tvARCH(∞) processes for $n = 1000$, 2000, 5000 and 10000 computed from 1000 independent replications.

2. A tvARCH(∞) example. With the notation of equation (31), assume:

$$\boldsymbol{\theta} = \big(c_0, c_1, p\big), \quad a_0(\boldsymbol{\theta}) = c_0 \quad \text{and} \quad a_j(\boldsymbol{\theta}) = c_1\, j^{-p} \quad \text{for } j \in \mathbb{N}^*$$

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

for any $1 \le t \le n$ and $n \in \mathbb{N}^*$. Therefore $c_0^*(u) = 1 + 0.5\, \sin(5\,u)$, $c_1^*(u) = 0.1{+}0.5\, \cos^2(4\,u)$ and $p^*(u) = 2.1{+}u$. Moreover, we assume that (ξ_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 $\widehat{c}_0^E(u)$, $\widehat{c}_1^E(u)$ and $\widehat{p}^E(u)$ for $u = k/50$ with $k = 1, \ldots, 49$.

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

3. The example of one integer valued process.
A tvINGARCH$(1, 0)$-process (tvINARCH(1)) as defined in (50). Here we choose

$$a_{0,t}^{(n)} = 1 + 0.5\, \sin\big(5\,\frac{t}{n}\big) \quad \text{and} \quad a_{1,t}^{(n)} = 0.3 + 0.5\,\frac{t}{n}$$

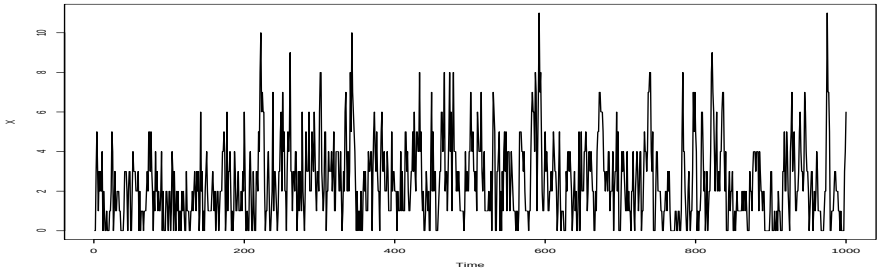


Figure 4: A trajectory of a tvINGARCH$(1,0)$-process for $n = 1000$

n	$\widehat{a}_0^U$	$\widehat{a}_0^E$	$\widehat{a}_1^U$	$\widehat{a}_1^E$
1000	0.144	0.135	0.051	0.058
2000	0.111	0.103	0.045	0.041
5000	0.079	0.073	0.032	0.030
10000	0.061	0.056	0.025	0.022

Table 3. Square root of the MISE of $\widehat{a}_0^U$, $\widehat{a}_0^E$, $\widehat{a}_1^U$ and $\widehat{a}_1^E$ for tvINGARCH$(1,0)$ processes for $n = 1000$, 2000, 5000 and 10000 computed from 1000 independent replications.

for any $1 \leq t \leq n$ and $n \in \mathbb{N}^*$. Note that we only consider here the Poisson distribution case. Figure 4 exhibits a trajectory of such a process for $n = 1000$. Using the same procedure than in the previous examples, Table 3 contains the RSMISE of the estimators computed with Uniform and Epanechnikov kernels, $\widehat{a}_0^U$, $\widehat{a}_0^E$, $\widehat{a}_1^U$ and $\widehat{a}_1^E$.

Conclusion: We can globally conclude from these Monte Carlo experiments:

- The consistency of the estimators and their convergence rate are established;
- The Epanechnikov's kernel is preferable to the uniform one.

5.2. Application to financial data

We apply our local non-parametric 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 5 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 Zakoïan [22] for more details). As a consequence, we used a tvGARCH$(1,1)$ process (see (33)) to take into account the changes in economic and financial conjectures over 20 years on such a model (think in particular of the September 2008 crisis and of the spring 2020 COVID crisis).

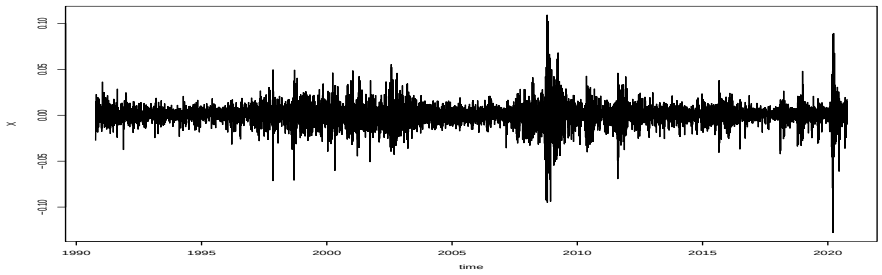


Figure 5: The log-returns of daily closing values of S&P500 index between October 1990 and October 2020

Figure 6 exhibits the evolution of the three estimators computed with Uniform and Epanechnikov kernels, *i.e.* $\widehat{c}_0^U$, $\widehat{c}_1^U$, $\widehat{d}_1^U$, $\widehat{c}_0^E$, $\widehat{c}_1^E$, $\widehat{d}_1^E$ from Gaussian QMLE.

We draw the evolution of $c_1 + d_1$ in order to get a visual indicator of the variability of the S&P500 index. The larger $c_1 + d_1$ the worst the moment properties of the tvGARCH$(1, 1)$. The variability may be seen as an indicator of instability of the financial markets and thus of the crisis. Indeed the maximum of the $c_1 + d_1$ 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 c_1 and d_1 separately. Then we observe that 2003 corresponds to a higher value for the coefficient d_1 and 2008 to a higher value for the coefficient c_1. We note that d_1 is the coefficient of persistence in the volatility whereas c_1 transfers external shocks in the volatility. Finally, also surprisingly, the COVID crisis does not seem to change the last five years evolution of c_1 and d_1.

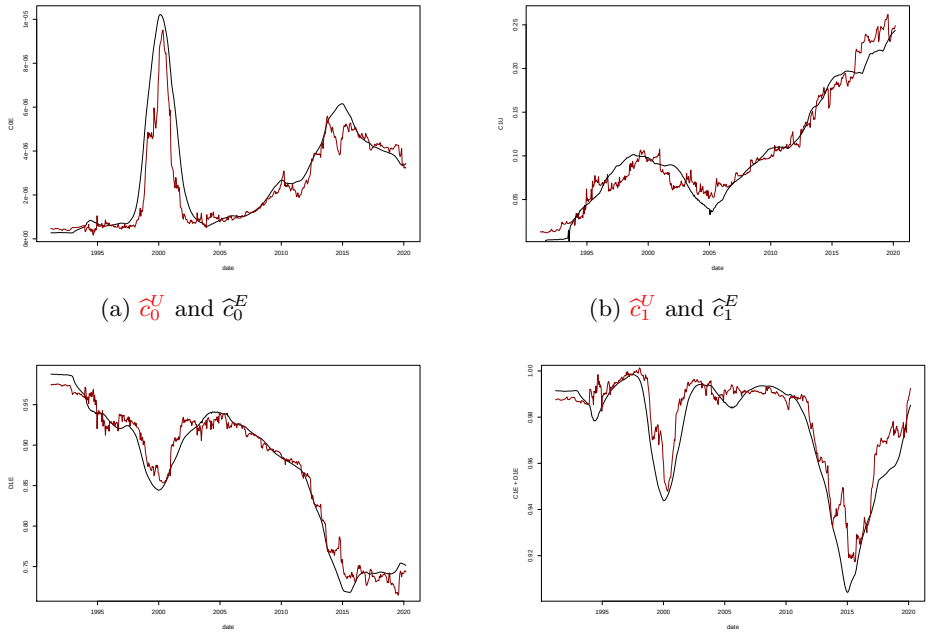
6. Moments and coupling properties of non stationary infinite memory processes

6.1. Proof of the moments properties in Lemma 2.1

We start this section with the proof of Lemma 2.1 which follows similar arguments than in [21] and that we give here for completeness.

Proof of Lemma 2.1. Under the assumption $(\mathbf{A}_0(\Theta))$, for any $n \in \mathbb{N}^*$ and $0 \le t \le n$ we have

$$\|X_t^{(n)} - F_{\boldsymbol{\theta}_t^{(n)}}(0; \xi_t)\|_p \le \sum_{s=1}^{\infty} b_s^{(0)} \|X_{t-s}^{(n)}\|_p \,.$$



(a) $\widehat{c}_0^U$ and $\widehat{c}_0^E$ (b) $\widehat{c}_1^U$ and $\widehat{c}_1^E$

(c) $\widehat{d}_1^U$ and $\widehat{d}_1^E$ (d) $\widehat{c}_1^U + \widehat{d}_1^U$ and $\widehat{c}_1^E + \widehat{d}_1^E$

Figure 6: Estimators of c_0, c_1, d_1 and $c_1 + d_1$ 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

$$\|X_t^{(n)}\|_p \le \sum_{s=1}^{\infty} b_s^{(0)} \|X_{t-s}^{(n)}\|_p + \sup_{\boldsymbol{\theta}\in\Theta} \|F_{\boldsymbol{\theta}}(0,\xi_0)\|_p.$$

As a consequence we get

$$\|X_t^{(n)}\|_p \le \sum_{s=1}^{\infty} b_s^{(0)} \max_{j\le t-1} \|X_j^{(n)}\|_p + \sup_{\boldsymbol{\theta}} \|F_{\boldsymbol{\theta}}(0,\xi_0)\|_p \le B_0(\Theta) \max_{j<t} \|X_j^{(n)}\|_p + C_0(\Theta).$$

With $M_t = \max_{j\le t} \|X_j^{(n)}\|_p$, a recursion entails that $M_t \le B_0(\Theta)\, M_{t-1} + C_0(\Theta)$ where $0 \le B_0(\Theta) < 1$, which implies with $M_0 = 0$ that for any $0 \le t \le n$,

$$M_t \le C_0(\Theta) \sum_{k=0}^{t} b_k^{(0)} \le \frac{C_0(\Theta)}{1 - B_0(\Theta)} < \infty$$

and this achieves the proof. We refer to [21] for more details. □

6.2. Weak dependence properties of the stationary version

Any stationary infinite memory process (2) admits fruitful coupling properties. In this section we quantify them thanks to the τ-weak dependence properties introduced in [16]. The reader is deferred to the lecture notes [15] for complements and details on coupling, based on the Wasserstein distance between probabilities. Conditional coupling for stationary time series is defined as follows.

Definition 6.1 ([16]). *Let $(\Omega, \mathcal{C}, \mathbb{P})$ be a probability space, $\mathcal{M}$ a σ-subalgebra of $\mathcal{C}$ and Z a random variable with values in E. Denote $\Lambda_1(E)$ the space of 1-Lipschitz functions from E to $\mathbb{R}$. Assume that $\|Z\|_p < \infty$ and define the coupling coefficient $\tau^{(p)}$ as*

$$\tau^{(p)}(\mathcal{M}, Z) = \Big\| \sup_{f\in\Lambda_1(E)} \Big\{ \Big| \int f(x) \mathbb{P}_{Z|\mathcal{M}}(dx) - \int f(x) \mathbb{P}_Z(dx) \Big| \Big\} \Big\|_p.$$

The dependence between the past of the sequence $(Z_t)_{t\in\mathbb{Z}}$ and its future k-tuples may be assessed by using the coupling coefficient $\tau^{(p)}$: Consider the norm $\|x - y\| = \|x_1 - y_1\| + \cdots + \|x_k - y_k\|$ on E^k, set $\mathcal{M}_p = \sigma(Z_t, t \le p)$ and define

$$\tau_Z^{(p)}(r) \quad = \quad \sup_{k>0} \Big\{ \max_{1\le l\le k} \frac{1}{l} \sup \Big\{ \tau^{(p)}(\mathcal{M}_p, (Z_{j_1}, \ldots, Z_{j_l})) \text{ with } p + r \le j_1 < \cdots < j_l \Big\} \Big\}.$$

Finally, the time series $(Z_t)_{t\in\mathbb{Z}}$ is said to be $\tau_Z^{(p)}$*-weakly dependent* in case the sequence of coefficients $\tau_Z^{(p)}(r)$ tend to 0 as r tends to infinity.

The τ-dependence coefficients of the stationary process $(\widetilde{X}_t(u))_{t\in\mathbb{Z}}$ are bounded above by using the following coupling scheme. Hence, if $(\xi_t^\circ)_{t\in\mathbb{Z}}$ is an independent replication of $(\xi_t)_{t\in\mathbb{Z}}$, define $(\widetilde{X}_t^\circ(u))_{t\in\mathbb{Z}}$ such as:

$$\widetilde{X}_t^\circ(u) = \begin{cases} F_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{t-k}^\circ(u))_{k\geq 1}, \xi_t^\circ\big), & \text{for } t \leq 0; \\ F_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{t-k}^\circ(u))_{k\geq 1}, \xi_t\big), & \text{for } t > 0. \end{cases} \tag{52}$$

Then for $s \geq 0$, we have the upper-bound

$$\tau^{(p)}_{\widetilde{X}(u)}(s) \leq \big\|\widetilde{X}_s(u) - \widetilde{X}_s^\circ(u)\big\|_p \,. \tag{53}$$

In the following, we mimic the proof of Theorem 3.1 of [21] in order to get an $\mathbb{L}^p$-estimate uniform over $u \in [0,1]$ of the approximation of $\widetilde{X}_s(u)$ by $\widetilde{X}_s^\circ(u)$. We start by estimating the moments $\|\sup_{u\in[0,1]} |\widetilde{X}_s(u)|\|_p$ in the following Lemma.

Lemma 6.1. *Let $\Theta \subset \mathbb{R}^d$ be such that $(\mathbf{A}_0(\Theta))$ holds with $B_0(\Theta) < 1$ and assume that (LS(ρ)) also holds. Then the stationary version $(\widetilde{X}_t(u))_{t\in\mathbb{Z}}$ solution of (8) satisfies*

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

Proof of Lemma 6.1. We adapt the fixed point approach of [20] to our setting. We refer to [20] for details. We consider $\mathbb{L}^p(\mathcal{C}([0,1],\mathbb{R}))$ the Banach space of random continuous functions $H : [0,1] \to \mathbb{R}$ that admits finite p 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 $(\xi_t)_{t\in\mathbb{Z}}$ such that $\mathbb{E}[\sup_{u\in[0,1]} |H_u((\xi_t)_{t\in\mathbb{Z}})|^p] < \infty$. We denote L the lag operator on sequences $(x_t)_{t\in\mathbb{Z}}$ of $\mathbb{R}^{\mathbb{Z}}$ such that $L((x_t)_{t\in\mathbb{Z}}) = (x_{t-1})_{t\in\mathbb{Z}}$. We denote Φ the function from $\mathbb{L}^p(\mathcal{C}([0,1],\mathbb{R}))$ such that

$$\Phi(H)(u) = F_{\boldsymbol{\theta}^*(u)}((H_u \circ L^j)_{j\geq 0}, \pi_0) \,, \qquad u \in [0,1] \,,$$

where π_0 is the projection $\pi_0\big((x_t)_{t\in\mathbb{Z}}\big) = x_0$. That $u \mapsto \Phi(H)(u)$ is continuous follows from the continuity of $\boldsymbol{\theta} \mapsto F_{\boldsymbol{\theta}}$ and $u \mapsto \boldsymbol{\theta}^*(u)$ under $(\mathbf{A}_0(\Theta))$ and (LS(ρ)). That $\sup_{u\in[0,1]} |\Phi(H)|(u)$ admits finite moments of order p follows from similar arguments than in Lemma 1 of [20] under $(\mathbf{A}_0(\Theta))$ that holds uniformly in $u \in [0,1]$. One can apply the Picard fixed point theorem to Φ which is a contraction under $B_0(\Theta) < 1$. We derive the existence of $\widetilde{X}_t(u)$ in the Banach space $\mathbb{L}^p(\mathcal{C}([0,1],\mathbb{R}))$ and the desired estimate on its norm. □

Notice that the same uniform estimate also holds on the coupling version $\widetilde{X}_s^\circ(u)$ so that one can consider the approximation, for any $s \in \mathbb{N}^*$ and any $r \in \mathbb{N}^*$. Now let us set the

uniform coupling τ-coefficients as $\tau_{\widetilde{X}}^{(p)}(s) \equiv \big\| \sup_{u\in[0,1]} |\widetilde{X}_s(u) - \widetilde{X}_s^\circ(u)| \big\|_p$, then,

$$\begin{aligned}
\tau_{\widetilde{X}}^{(p)}(s) &\le \Big\| \sup_{u\in[0,1]} \big| F_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{s-k}(u))_{k\ge1}, \xi_t\big) - F_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}^\circ_{s-k}(u))_{k\ge1}, \xi_t^\circ\big) \big| \Big\|_p \\
&\le \Big\| \sup_{u\in[0,1]} \big| F_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}_{s-k}(u))_{k\ge1}, \xi_s\big) - F_{\boldsymbol{\theta}^*(u)}\big((\widetilde{X}^\circ_{s-k}(u))_{k\ge1}, \xi_s^\circ\big) \big| \Big\|_p \\
&\le \sum_{k=1}^{\infty} b_t^{(0)}(\Theta) \Big\| \sup_{u\in[0,1]} \big| \widetilde{X}_{s-k}(u) - \widetilde{X}^\circ_{s-k}(u) \big| \Big\|_p \\
&\le B_0(\Theta) \max_{s-r\le t\le s-1} \Big\| \sup_{u\in[0,1]} \big| \widetilde{X}_t(u) - \widetilde{X}_t^\circ(u) \big| \Big\|_p \qquad (54) \\
&\quad + 2 \sum_{k=r+1}^{\infty} b_t^{(0)}(\Theta) \Big\| \sup_{u\in[0,1]} |\widetilde{X}_0(u)| \Big\|_p .
\end{aligned}$$

By a recursive argument we easily derive that $\max_{t\ge0} \big\| \sup_{u\in[0,1]} \big| \widetilde{X}_t(u) - \widetilde{X}_t^\circ(u) \big| \big\|_p < \infty$. Then we extend the bound so that

$$\begin{aligned}
&\max_{t\ge s} \Big\| \sup_{u\in[0,1]} |\widetilde{X}_t(u) - \widetilde{X}_t^\circ(u)| \Big\|_p \\
&\quad \le B_0(\Theta) \max_{t\ge s-r} \Big\| \sup_{u\in[0,1]} |\widetilde{X}_t(u) - \widetilde{X}_t^\circ(u)| \Big\|_p + 2 \sum_{k=r+1}^{\infty} b_t^{(0)}(\Theta) \Big\| \sup_{u\in[0,1]} |\widetilde{X}_0(u)| \Big\|_p . \qquad (55)
\end{aligned}$$

Given any $r \in \mathbb{N}^*$, a recursive argument on the sequence $\big(\max_{t\ge kr} \big\| \sup_{u\in[0,1]} |\widetilde{X}_t(u) - \widetilde{X}_t^\circ(u)| \big\|_p \big)_{k\ge0}$ based on (55) together with the coupling properties of the τ-dependent coefficient (53) yields the following Lemma that is analog to Theorem 3.1 of [21]. We refer to [21] for more details.

Lemma 6.2. *Let $\Theta \subset \mathbb{R}^d$ such that $(\mathbf{A}_0(\Theta))$ holds with $B_0(\Theta) < 1$ and assume that (LS(ρ)) also hold. Then the stationary version $(\widetilde{X}_t(u))_{t\in\mathbb{Z}}$ is uniformly approximated by its coupling version $(\widetilde{X}_t^\circ(u))$ so that we have*

$$\tau_{\widetilde{X}}^{(p)}(s) \le C\,\lambda_s \qquad \textit{where} \quad \lambda_s = \inf_{1\le r\le s} \Big(B_0(\Theta)^{s/r} + \sum_{t=r+1}^{\infty} b_t^{(0)}(\Theta) \Big) \quad \textit{for } s \ge 1. \qquad (56)$$

Remark that the bound on the τ-coefficients does not depend on $u \in (0,1)$. Notice also that:

- if $\quad b_t^{(0)}(\Theta) = O\big(t^{-\kappa}\big)$ with $\kappa > 1$, then $\tau_{\widetilde{X}}^{(p)}(s) \le \lambda_s = O\big(s^{1-\kappa}\log s\big)$;
- if $\quad b_t^{(0)}(\Theta) = O\big(r^t\big)$ with $0 < r < 1$, then $\tau_{\widetilde{X}}^{(p)}(s) \le \lambda_s = O\big(e^{\sqrt{s\,\log(r)\,\log(B_0(\Theta))}}\big)$.

(57)

The SLLN in [21] is implied by the summability of the τ-dependence coefficients. We will use the following Lemma on the τ-dependence coefficients $\tau_{\widetilde{X}}^{(p)}(s)$, $s \ge 1$.

Lemma 6.3. *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

$$\lambda_s \leq s^{-C\log(1/B_0(\Theta))} + \sum_{t=\lfloor s/C\log(s)\rfloor}^{\infty} b_t^{(0)}(\Theta).$$

For $C > 0$ sufficiently large and since $B_0(\Theta) < 1$, we get $\sum_{s=1}^{\infty} s^{-C\log(1/B_0(\Theta))} < \infty$. Moreover, for $s > e$ quote that if $t = s/C\log(s)$ we have $s > Ct\log(t)$ and inverting of sums yields:

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

and the desired result follows. □

6.3. Coupling of local-stationary processes

The τ-dependence properties of the stationary process comes from the coupling scheme where we uniformly approximate the stationary version $(\widetilde{X}_t(u))$ with a copy $(\widetilde{X}_t^*(u))$ that is independent of the past $(\widetilde{X}_t(u))_{t\leq 0}$. The weak dependence notion will be used in order to get the uniform SLLN over functional of $(\widetilde{X}_t(u))$. The goal of this section is to extend such coupling approach to non-stationary processes $(X_t^{(n)})$ in the sense that one approximates $(X_t^{(n)})$ in $\mathbb{L}^p$ with a certain coupled version. A useful remark is that we do not use the stationarity of the coupling process $(\widetilde{X}_t^*(u))$ for obtaining the recursion (55). Thus a similar coupling scheme can be extended to the local stationary process $(X_t^{(n)})$ but locally only. In order to localize, we define $u \in [\varepsilon, 1-\varepsilon]$, $\varepsilon > 0$ and n large enough the quantities

$$i_n(u) := [n(u - ch_n)] \geq 1 \quad \text{and} \quad j_n(u) := [n(u + ch_n)] \leq n, \tag{58}$$

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

Definition 6.2 (Coupling process)**.** *In the time-window* $\{i_n(u), i_n(u)+1, \ldots, j_n(u)\}$ *define the process* $(X_t^*(u))_{i_n\leq t\leq j_n}$ *by*

$$X_t^*(u) = \begin{cases} X_t^{(n)}, & t < i_n(u), \\ F_{\boldsymbol{\theta}^*(u)}\big((X_{t-k}^*(u))_{k\geq 1}, \xi_t\big), & i_n(u) \leq t \leq j_n(u). \end{cases} \tag{59}$$

Notice that for the ease of notation and as n is fixed sufficiently large in this section, we suppress the dependence in n on the coupling process. We first have to prove the existence of the process $(X_t^*(u))$.

Lemma 6.4. *Let $\Theta \subset \mathbb{R}^d$ be such that $(\mathbf{A}_0(\Theta))$ holds with $B_0(\Theta) < 1$ and assume that (LS(ρ)) also holds. Then, for any $u \in (0,1)$, there exists a.s. a unique coupling process $(X_t^*(u))_{t\in\mathbb{Z}}$ satisfying (59) and there exists a positive constant $C^* > 0$ such that*

$$\Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^*_{i_n(u)+s}(u)\big| \Big\|_p \le C^*\, n^{1/p} \qquad \textit{for all } \ 0 \le s \le 2c\,nh_n.$$

Proof of Lemma 6.4. We use a chaining argument adapted to our framework. We denote (u_k) where $u_k = (k+1/2+ch_n)/n$ for $k \in U_n(\varepsilon) = \big\{[\varepsilon n - ch_n - 1/2], \dots, [(1-\varepsilon)n - ch_n - 1/2]\big\}$ a grid of points of the segment $[\varepsilon, 1-\varepsilon]$ and therefore $\big|U_n(\varepsilon)\big| \simeq (1-2\varepsilon)n \le n$. Moreover, for each $u \in [\varepsilon, 1-\varepsilon]$ there exists u_k such that $|u-u_k| \le 1/(2n)$ and therefore $i_n(u) = i_n(u_k)$ with $i_n(u)$ defined in (58). Then a chaining argument shows that

$$\sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^*_{i_n(u)+s}(u)\big| \le \max_{k\in U_n(\varepsilon)} \big|X^*_{i_n(u_k)+s}(u_k)\big| + \sup_{u,v:\, i_n(u)=i_n(v)} \big|X^*_{i_n(u)+s}(u) - X^*_{i_n(v)+s}(v)\big|. \tag{60}$$

Using the inequality $\max(|x|,|y|) \le |x|+|y|$ and similar argument than in the proof of Lemma 2.1, we get that

$$\begin{aligned}
\Big\| \max_{k\in U_n(\varepsilon)} \big|X^*_{i_n(u_k)+s}(u_k)\big| \Big\|_p &\le \Big\| \sum_{k\in U_n(\varepsilon)} \big|X^*_{i_n(u_k)+s}(u_k)\big| \Big\|_p \\
&\le \Big(\sum_{k\in U_n(\varepsilon)} \big\|X^*_{i_n(u_k)+s}(u_k)\big\|_p^p \Big)^{1/p} \\
&\le n^{1/p}\, \frac{C_0(\Theta)}{1-B_0(\Theta)},
\end{aligned} \tag{61}$$

since $\|X^*_{i_n(u)+s}(u)\|_p \le C_0(\Theta)/(1-B_0(\Theta))$ for any $u \in [\varepsilon, 1-\varepsilon]$ and $0 \le s \le 2c\,nh_n$. Set $\delta_n = \Big\| \sup_{u,v:\, i_n(u)=i_n(v)} \big|X^*_{i_n(u)+s}(u) - X^*_{i_n(v)+s}(v)\big| \Big\|_p$. We derive from an application of

the chaining argument,

$$\begin{aligned}
\delta_n \;\le\;& \Big\| \sup_{u,v:\, i_n(u)=i_n(v)} \big| F_{\theta^*(i_n(u)+s)}\big((X^*_{i_n(u)+s-k}(u))_{k\ge1}, \xi_{i_n(u)+s}\big) \\
&\qquad\qquad - F_{\theta^*(i_n(u)+s)}\big((X^*_{i_n(v)+s-k}(v))_{k\ge1}, \xi_{i_n(v)+s}\big)\big| \Big\|_p \\
+\;& \Big\| \sup_{u,v:\, i_n(u)=i_n(v)} \big| F_{\theta^*(i_n(u)+s)}\big((X^*_{i_n(v)+s-k}(v))_{k\ge1}, \xi_{i_n(v)+s}\big) \\
&\qquad\qquad - F_{\theta^*(i_n(v)+s)}\big((X^*_{i_n(v)+s-k}(v))_{k\ge1}, \xi_{i_n(v)+s}\big)\big| \Big\|_p \\
\le\;& \sum_{k=1}^{\infty} b_k^{(0)}(\Theta) \Big\| \sup_{u,v:\, i_n(u)=i_n(v)} \big| X^*_{i_n(u)+s-k}(u) - X^*_{i_n(v)+s-k}(v)\big| \Big\|_p \\
+\;& \sup_{u,v:\, i_n(u)=i_n(v)} \big\|\boldsymbol{\theta}^*(u) - \boldsymbol{\theta}^*(v)\big\| \\
&\qquad \times \Big(\sum_{k=1}^{\infty} b_k^{(1)}(\Theta) \Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big| X^*_{i_n(u)+s-k}(u)\big| \Big\|_p + \Big\| \sup_{\boldsymbol{\theta}\in\Theta} \big\|\partial^1_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(0;\xi_0)\big\| \Big\|_p \Big) \\
\le\;& B_0(\Theta) \Big\| \sup_{u,v:i_n(u)=i_n(v)} \big| X^*_{i_n(u)+s-k}(u) - X^*_{i_n(v)+s-k}(v)\big| \Big\|_p \\
+\;& K_{\boldsymbol{\theta}}\, n^{-\rho} \Big\| \sup_{\boldsymbol{\theta}\in\Theta} \big\|\partial^1_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(0;\xi_0)\big\| \Big\|_p \\
+\;& K_{\boldsymbol{\theta}} n^{-\rho} B_1(\Theta) \Big(\frac{C_0(\Theta) n^{1/p}}{1-B_0(\Theta)} + \Big\| \sup_{u,v:i_n(u)=i_n(v)} \big| X^*_{i_n(u)+s-k}(u) - X^*_{i_n(v)+s-k}(v)\big| \Big\|_p \Big),
\end{aligned}$$

from (7), (60) and (61). Collecting all those bounds for n sufficiently large in order that $n^{-\rho}$ is sufficiently small, we get

$$\begin{aligned}
&\Big\| \sup_{u,v:\, i_n(u)=i_n(v)} \big| X^*_{i_n(u)+s}(u) - X^*_{i_n(v)+s}(v)\big| \Big\|_p \\
&\qquad\qquad \le \frac{\dfrac{K_{\boldsymbol{\theta}} B_1(\Theta) C_0(\Theta)}{1-B_0(\Theta)}\, n^{1/p-\rho} + K_{\boldsymbol{\theta}}\, n^{-\rho} \big\| \sup_{\boldsymbol{\theta}\in\Theta} |\partial^1_{\boldsymbol{\theta}} F_{\boldsymbol{\theta}}(0;\xi_0)| \big\|_p}{1 - B_0(\Theta) - K_{\boldsymbol{\theta}}\, n^{-\rho} B_1(\Theta)}\,.
\end{aligned}$$

Finally, applying again the chaining argument we obtain

$$\| \max_{k\in U_n(\varepsilon)} \big| X^*_{i_n(u_k)+s}(u_k)\big| \|_p \le n^{1/p}\, \frac{C_0(\Theta)}{1-B_0(\Theta)} + O(n^{1/p-\rho})\,.$$

□

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

Lemma 6.5. *Under Assumptions* $(\mathbf{A_0(\Theta)})$ *with* $B_0(\Theta) < 1$, $(\mathbf{A_1(\Theta)})$ *and* $(\mathbf{LS}(\rho))$, *with* $(X_t^*(u))$ *the coupling process defined in* (59), *there exists a positive constant* $C' > 0$ *such that*

$$\Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^{(n)}_{i_n(u)+s} - X^*_{i_n(u)+s}(u)\big| \Big\|_p \le C'\, h_n^{\rho} n^{1/p} \qquad \text{for all } 0 \le s \le 2c\,nh_n. \tag{62}$$

Proof of Lemma 6.5. Define $\Delta_s^* = 0$ for any $s \le 0$ and for $0 \le s \le 2c\,nh_n$ set the quantity $\Delta_s^* = \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^{(n)}_{i_n(u)+s} - X^*_{i_n(u)+s}(u)\big|$. For $0 \le s \le 2c\,nh_n$ we decompose

$$\begin{aligned}
\Delta_s^* \;=\; & \sup_{u\in[\varepsilon,1-\varepsilon]} \Big|F_{\boldsymbol{\theta}^{(n)}_{i_n(u)+s}}\big((X^{(n)}_{i_n(u)+s-k})_{k\ge1}, \xi_{i_n(u)+s}\big) \\
& \qquad\qquad - F_{\boldsymbol{\theta}^*(u)}\big((X^*_{i_n(u)+s-k}(u))_{k\ge1}, \xi_{i_n(u)+s}\big)\Big| \\
\le\; & \sup_{u\in[\varepsilon,1-\varepsilon]} \Big|F_{\boldsymbol{\theta}^{(n)}_{i_n(u)+s}}\big((X^{(n)}_{i_n(u)+s-k})_{k\ge1}, \xi_{i_n(u)+s}\big) \\
& \qquad\qquad - F_{\boldsymbol{\theta}^{(n)}_{i_n(u)+s}}\big((X^*_{i_n(u)+s-k}(u))_{k\ge1}, \xi_{i_n(u)+s}\big)\Big| \\
+\; & \sup_{u\in[\varepsilon,1-\varepsilon]} \Big|F_{\boldsymbol{\theta}^{(n)}_{i_n(u)+s}}\big((X^*_{i_n(u)+s-k}(u))_{k\ge1}, \xi_{i_n(u)+s}\big) \\
& \qquad\qquad - F_{\boldsymbol{\theta}^*(u)}\big((X^*_{i_n(u)+s-k}(u))_{k\ge1}, \xi_{i_n(u)+s}\big)\Big|.
\end{aligned}$$

Then we derive

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

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

Now define $M_t^* = \max_{s\leq t} \big\|\Delta_s^*\big\|_p$. We derive from an application of Lemma 6.4 that for any $0 \leq t \leq 2c\,nh_n$ it holds

$$
\big\|\Delta_t^*\big\|_p \leq B_0(\Theta)\, M^*_{t-1} + \sup_{u\in[\varepsilon,1-\varepsilon]} \big\|\boldsymbol{\theta}^{(n)}_{i_n(u)+t} - \boldsymbol{\theta}^*(u)\big\| \times \Big(C^*\, n^{1/p} + C_1(\Theta) \Big).
$$

We have $\sup_{u\in[\varepsilon,1-\varepsilon]} \big\|\boldsymbol{\theta}^{(n)}_{i_n(u)+t} - \boldsymbol{\theta}^*(u)\big\| \leq c\, K_\theta\, h_n^\rho$ from condition (7) of Assumption $(\mathbf{LS}(\rho))$. As a consequence, for any $0 \leq t \leq 2c\,nh_n$, we have

$$
M_t^* \leq B_0(\Theta)\, M^*_{t-1} + c\, K_\theta \Big(C^* n^{1/p} + C_1(\Theta) \Big)\, h_n^\rho.
$$

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

$$
M_t^* \leq \frac{c\, K_\theta}{1 - B_0(\Theta)} \Big(C^* n^{1/p} + C_1(\Theta) \Big)\, h_n^\rho.
$$

This completes the proof of the Lemma 6.5. □

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

Lemma 6.6. *Under Assumptions* $(\mathbf{A_0}(\boldsymbol{\Theta}))$ *with* $B_0(\Theta) < 1$, $(\mathbf{A_1}(\boldsymbol{\Theta}))$ *and* $(\mathbf{LS}(\rho))$ *there exists a positive constant* $C'' > 0$ *such that*

$$\Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^{(n)}_{i_n(u)+s} - \widetilde{X}_{i_n(u)+s}(u)\big| \Big\|_p \le C''\, n^{1/p}\big(h_n^\rho + \lambda_s\big), \quad \textit{for all } 0 \le s \le 2c\,nh_n. \tag{63}$$

Proof of Lemma 6.6. The approximation is derived since the coupling process is a non stationary coupling version of $\widetilde{X}_t^\circ(u)$ defined in (52). Indeed, using $(X_t^*(u))$ the coupling process defined in (59) and repeating the same arguments than above, we obtain a similar recursive relation than (55) on the sequence

$$\Big(\max_{t\ge kr} \Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|\widetilde{X}_{i_n(u)+t}(u) - X^*_{i_n(u)+t}(u)\big| \Big\|_p\Big)_{k\ge 0}$$

given any $r \in \mathbb{N}^*$, using Lemma 6.4 and the estimates

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

We obtain

$$\Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^*_{i_n(u)+s}(u) - \widetilde{X}_{i_n(u)+s}(u)\big| \Big\|_p \le C\, n^{1/p}\, \lambda_s\,, \qquad \text{for} \quad 0 \le s \le 2c\,nh_n,$$

with λ_s defined in (56). Combining this result with (63) we bound the $\mathbb{L}^p$ norm of the approximation of $X_s^{(n)}$ with the stationary version $\widetilde{X}_s(u)$, namely for for all $0 \le s \le 2c\,nh_n$,

$$\begin{aligned}&\Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^{(n)}_{i_n(u)+s} - \widetilde{X}_{i_n(u)+s}(u)\big| \Big\|_p \\ &\le \Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^{(n)}_{i_n(u)+s} - X^*_{i_n(u)+s}(u)\big| \Big\|_p + \Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \big|X^*_{i_n(u)+s}(u) - \widetilde{X}_{i_n+s}(u)\big| \Big\|_p \\ &\le C\, n^{1/p}\big(h_n^\rho + \lambda_s\big).\end{aligned}$$

□

7. Proofs for the Section 3

7.1. Some useful lemmas

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

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

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

$$\big\|\phi_{t,m_2} - \phi_{t,m_1}\big\|_1 \;\le\; g\big(\sup_{0\le s\le m_2}\{\|X_{t-s}^{(n)}\|_p\}\big)\sum_{k=m_1+1}^{m_2}\alpha_k(h,\Theta)\,\|X_{t-k}^{(n)}\|_p \tag{64}$$

$$\le\; C\sum_{k=m_1+1}^{m_2}\alpha_k(h,\Theta) \tag{65}$$

from Lemma 2.1, since if $s<0$ then $X_s=0$, thus the corresponding supremum bound extends over each $s\le n$.
As $\sum_{k=1}^{\infty}\alpha_k(h,\Theta)<\infty$ we deduce that for any $\varepsilon>0$, $\sum_{k=m_1+1}^{m_2}\alpha_k(h,\Theta)\le\varepsilon$ 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 $\Phi\big((X_{t-k}^{(n)})_{k\ge0},\boldsymbol{\theta}\big)$.

When $\boldsymbol{\theta}_t^{(n)}=\boldsymbol{\theta}^*(u)$ for any t,n, we consider $\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)$ that also exists in $\mathbb{L}^1$. Moreover, as $(\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}}$ is a stationary ergodic process, this is also the case for $\big(\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)\big)_{t\in\mathbb{Z}}$ (see Corollary 2.1.3. in [32]). □

Lemma 7.1. *Let $i_n(u))$ and $j_n(u)$ defined in (58).*

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 $0\le u\le1$. If $Z(u)$ is an ergodic process continuous with respect to u and satisfying $\mathbb{E}[\sup_{0\le u\le1}\|Z_0(u)\|]<\infty$ then we have*

$$\sup_{0<u<1}\left\|\frac{1}{n\,h_n}\sum_{t=i_n(u)}^{j_n(u)}Z_t(u)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\right\| \;\xrightarrow[n\to+\infty]{a.s.}\; 0. \tag{66}$$

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

$$\frac{1}{\sqrt{n\,h_n}}\sum_{t=i_n(u)}^{j_n(u)}Z_t\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big) \;\xrightarrow[n\to+\infty]{\mathcal{L}}\; \mathcal{N}\Big(0\,,\,\Big(\int_{\mathbb{R}}K^2(x)\,dx\Big)\,\Sigma\Big). \tag{67}$$

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 $n(u-c\,h_n)\ge1$ and $nu+c\,nh_n\le n$. Then, for $j\in\{1,\dots,\ell\}$, we denote $I_j^{(\ell)}=\big[-c+2c\,\frac{j-1}{\ell}\,,\,-c+2c\,\frac{j}{\ell}\big]$, $T_j^{(\ell)}=\Big\{t\in\mathbb{N},\ \frac{\frac{t}{n}-u}{h_n}\in I_j\Big\}$ and

$$S_n^{(\ell)}(u)=\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}Z_t(u)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\quad\text{and}\quad S_{n,j}^{(\ell)}(u)=\frac{1}{nh_n}\sum_{t\in T_j^{(\ell)}}Z_t(u)\,. \tag{68}$$

Below, we will omit the reference to 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 $\mathrm{Card}(T_j^{(\ell)}) \simeq 2cnh_n/\ell \underset{n\to+\infty}{\longrightarrow} \infty$ we apply the uniform ergodic theorem and since $\mathbb{E}[Z_0(u)] = 0$ for any $0 < u < 1$ we obtain

$$\sup_{0\le u\le 1} \|S_{n,j}^{(\ell)}(u)\| \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} 0. \tag{69}$$

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

$$S_n^{(\ell)}(u) = \sum_{j=1}^{\ell} K\Big(\frac{\frac{t_j^{(\ell)}}{n}-u}{h_n}\Big) S_{n,j}^{(\ell)}(u) + \sum_{j=1}^{\ell} \frac{1}{nh_n} \sum_{t\in T_j^{(\ell)}} Z_t(u)\Big[K\Big(\frac{\frac{t}{n}-u}{h_n}\Big) - K\Big(\frac{\frac{t_j^{(\ell)}}{n}-u}{h_n}\Big)\Big]. \tag{70}$$

First, since K is a bounded function and from (69), then for any $\ell \in \mathbb{N}^*$ we obtain

$$\sup_{0\le u\le 1} \Big\| \sum_{j=l}^{\ell} K\Big(\frac{\frac{t_j^{(\ell)}}{n}-u}{h_n}\Big) S_{n,j}^{(\ell)}(u)\Big\| \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} 0. \tag{71}$$

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

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

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

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

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

$$\frac{1}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \sup_{0<u<1} \|Z_t(u)\| \underset{n\to+\infty}{\overset{a.s.}{\longrightarrow}} \mathbb{E}\Big[\sup_{0<u<1}\|Z_0(u)\|\Big].$$

Thus, for any $\varepsilon > 0$, there exists a.s. (ℓ_0, n_0) such as for any $\ell \ge \ell_0$ and $n \ge n_0$,

$$\sup_{0<u<1} \Big\| \sum_{j=1}^{\ell} \frac{1}{nh_n} \sum_{t\in T_j^{(\ell)}} Z_t(u) \Big[K\Big(\frac{\frac{t}{n}-u}{h_n}\Big) - K\Big(\frac{\frac{t_j^{(\ell)}}{n}-u}{h_n}\Big)\Big]\Big\| \le \varepsilon \qquad a.s. \tag{72}$$

From (70), (71) and (72), we deduce (66).

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).

$$S_n^{(\ell)} = \sum_{j=1}^{\ell} a_j \, S_{n,j}^{(\ell)}, \tag{73}$$

with $S_{n,j}^{(\ell)}$ defined in (68). Using $\mathrm{Card}(T_j^{(\ell)}) \sim 2cnh_n/\ell \underset{n\to+\infty}{\longrightarrow} \infty$, for any $j \in \{1, \ldots, \ell\}$,

$$\Big(\frac{\ell}{2cnh_n}\Big)^{1/2} \sum_{t \in T_j^{(\ell)}} Z_t = \sqrt{\frac{nh_n\ell}{2c}}\, S_{n,j}^{(\ell)} \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\Big(0\,,\, \sigma^2\Big), \tag{74}$$

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 $j, j' \in \{1, \ldots, \ell\}$ such as $j \neq j'$,

$$\begin{aligned} nh_n \,\mathrm{Cov}\left(S_{n,j}^{(\ell)}\,,\, S_{n,j'}^{(\ell)}\right) &= \frac{1}{nh_n} \sum_{t\in T_j^{(\ell)}} \sum_{t'\in T_{j'}^{(\ell)}} \mathbb{E}[Z_t Z_{t'}] = \frac{1}{nh_n} \sum_{t\in T_j^{(\ell)}} \sum_{t'\in T_{j'}^{(\ell)}} \mathbb{E}[Z_0 Z_{t'-t}] \\ \Longrightarrow \quad \Big|nh_n \,\mathrm{Cov}\left(S_{n,j}^{(\ell)}\,,\, S_{n,j'}^{(\ell)}\right)\Big| &\le C \Big|\sum_{k > T_j^{(\ell)}} \mathbb{E}[Z_0 Z_k]\Big| \underset{n\to+\infty}{\longrightarrow} 0, \end{aligned}$$

as soon as $\ell = o(nh_n)$ since $\sum_{t\in\mathbb{Z}} \big|\mathbb{E}[Z_0 Z_t]\big| < \infty$. Hence a central limit theorem holds for any linear combinations of $S_{n,j}^{(\ell)}$. E.g. for $S_n^{(\ell)}$ defined in (73),

$$\begin{aligned} &\sqrt{\frac{nh_n\ell}{2c}}\, S_n^{(\ell)} \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\Big(0\,,\, \sum_{j=1}^{\ell} a_j^2 \sum_{t\in\mathbb{Z}} \mathbb{E}[Z_0 Z_t]\Big) \\ &\qquad\qquad \Longrightarrow \quad \sqrt{nh_n}\, S_n^{(\ell)} \underset{n\to+\infty}{\overset{\mathcal{L}}{\longrightarrow}} \mathcal{N}\Big(0\,,\, \sigma^2 \int_{\mathbb{R}} K_\ell^2(x)\, dx\Big), \end{aligned} \tag{75}$$

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

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

$$\int_{\mathbb{R}} K_\ell^2(x)\, dx \underset{\ell\to+\infty}{\longrightarrow} \int_{\mathbb{R}} K^2(x)\, dx.$$

The result will follow from Theorem 3.2 in [6]. For this we check that for any $\varepsilon > 0$

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \mathbb{P}\Big(\Big|S_n^{(\ell)} - S_n\Big| \ge \varepsilon/\sqrt{nh_n}\Big) = 0, \qquad \text{with} \qquad S_n = \frac{1}{nh_n} \sum_{t=1}^{n} Z_t\, K\Big(\frac{\frac{t}{n} - u}{h_n}\Big).$$

From Markov inequality

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

with

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

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

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

Thus

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

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

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

Now the extension to $d>1$ is standard: consider $r=(r_1,\ldots,r_d)^T\in\mathbb{R}^d$ and a linear combination $\mathcal{Z}_t = r_1 Z_t^{(1)} + \cdots + r_d Z_t^{(d)}$ where $Z_t = \big(Z_t^{(1)},\ldots,Z_t^{(d)}\big)^T$ 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. □

7.2. Proofs of the main results

We will prove (12) in Theorem 3.1 only since the consistency is achieved directly by simple arguments. We need the following Lemma that is a strong law of large number on the contrast as if the stationary versions were observed:

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

$$\sup_{\varepsilon\le u\le 1-\varepsilon}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}\Phi(\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta})K\big(\tfrac{\frac{t}{n}-u}{h_n}\big)-\mathbb{E}\big[\Phi\big((\widetilde{X}_{-k}(u))_{k\ge 0},\boldsymbol{\theta}\big)\big]\Big|$$
$$\xrightarrow[n\to+\infty]{a.s.} 0. \quad (76)$$

Proof. The expression I in (76) tends to 0 a.s., if is it the case for I_1 and I_2 such that,

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

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

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 $Z_t(\boldsymbol{\theta},u)=\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)-\mathbb{E}\big[\Phi\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)\big]$; this is a centered ergodic stationary process on the Banach space of the continuous function over $\Theta\times[0,1]$ equipped with the uniform norm. Using $\mathbb{E}\big[\sup_{(\boldsymbol{\theta},u)\in\Theta\times[0,1]}|Z_0(\boldsymbol{\theta},u)|\big]<\infty$ since $\Phi\in \mathrm{Lip}_p(\Theta)$, with Theorem 2.2.1. in [32] we apply the part 1. of Lemma 7.1 to get

$$I_1 \xrightarrow[n\to+\infty]{a.s.} 0. \quad (77)$$

2. For the term I_2, notice that

$$\begin{aligned} I_2 &\le \sup_{\varepsilon\le u\le 1-\varepsilon}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\Big(1-\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\Big)\mathbb{E}\big[\Phi\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)\big]\Big| \\ &\le C\sup_{\varepsilon\le u\le 1-\varepsilon}\Big|1-\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\Big|\le\frac{C}{nh_n}, \end{aligned} \quad (78)$$

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 (77) and (78). □

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 $(\widetilde{X}_t(u))_t$ denoting the stationary process defined in (8), we obtain*

$$\sup_{u\in[\varepsilon,1-\varepsilon]}\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{nh_n}\sum_{k=1}^{n}\Phi\big((X_{k-t}^{(n)})_{t\ge 0},\boldsymbol{\theta}\big)K\Big(\frac{\frac{k}{n}-u}{h_n}\Big)-\mathbb{E}\big[\Phi\big((\widetilde{X}_{-k}(u))_{k\ge 0},\boldsymbol{\theta}\big)\big]\Big| \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \quad (79)$$

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

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

From Assumption $(\mathbf{A_0}(\boldsymbol{\Theta}))$ and $B_0(\Theta)<1$,

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

for $j \leq 0$ using similar arguments as in the proof Lemma 6.6. Moreover with $(\mathbf{A_1}(\boldsymbol{\Theta}))$ and $(\mathbf{LS}(\rho))$, we apply 6.6 in order to get

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

for $j \geq 1$. Therefore,

$$\Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{nh_n} \sum_{t=0}^{2c\,nh_n} \big(\Phi\big((X_{t-k}^{(n)})_{k\geq 0},\boldsymbol{\theta}\big) - \Phi\big((\widetilde{X}_{t-k}(u))_{k\geq 0},\boldsymbol{\theta}\big)\big) K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\Big\|_1$$
$$\leq \frac{C}{nh_n} \sum_{t=0}^{2c\,nh_n} C_K\, C^* \Big(\sum_{s=1}^{t} \alpha_s(\Phi,\Theta)\, C\, n^{1/p}\big(h_n^\rho + \lambda_{t+1-s}\big) + \sum_{s=t+1}^{t} \alpha_s(\Phi,\Theta)\, C\, n^{1/p}\Big). \tag{80}$$

Then we deduce

$$\Big\| \sup_{u\in[\varepsilon,1-\varepsilon]} \sup_{\boldsymbol{\theta}\in\Theta} \frac{1}{nh_n} \sum_{t=0}^{2c\,nh_n} \big(\Phi\big((X_{t-k}^{(n)})_{k\geq 0},\boldsymbol{\theta}\big) - \Phi\big((\widetilde{X}_{t-k}(u))_{k\geq 0},\boldsymbol{\theta}\big)\big)\, K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\Big\|_1$$
$$\leq \frac{C\, n^{1/p}}{nh_n} \sum_{t=i_n(u)}^{j_n(u)} \Big(\sum_{s=1}^{t-i_n} \alpha_s(\Phi,\Theta)\,\big(\lambda_{t-s-i_n} + h_n^\rho\big) + \sum_{s=t-i_n+1}^{\infty} \alpha_s(\Phi,\Theta)\Big)$$
$$\leq \frac{C\, n^{1/p}}{nh_n} \Big(\sum_{k=1}^{j_n-i_n} \lambda_k \sum_{i=1}^{k} \alpha_i(\Phi,\Theta) + h_n^\rho \sum_{k=1}^{j_n-i_n} \sum_{i=1}^{k} \alpha_i(\Phi,\Theta) + \sum_{i=1}^{\infty} i\alpha_i(\Phi,\Theta)\Big)$$
$$\leq \frac{Cn^{1/p}}{nh_n} \Big(\sum_{k=1}^{\infty} \lambda_k \sum_{i=1}^{\infty} \alpha_i(\Phi,\Theta) + \big(h_n^\rho + 1\big) \sum_{i=1}^{\infty} i\alpha_i(\Phi,\Theta)\Big)$$
$$\leq C\, \frac{n^{1/p}}{nh_n}. \tag{81}$$

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 $\sum_{t=1}^{\infty} t\log(t) b_t(\Theta) < \infty$ and follows from Lemma 6.3. Finally, using (80), (81) and the almost sure convergence (76) we obtain the weak consistency result (79). □

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

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

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

$$\sup_{u\in[\varepsilon,1-\varepsilon]} \sup_{\boldsymbol{\theta}\in\Theta} \Big| \frac{1}{nh_n} \sum_{k=1}^{n} \Phi\big((X_{k-t}^{(n)})_{t\geq 0},\boldsymbol{\theta}\big)\, K\Big(\frac{\frac{k}{n}-u}{h_n}\Big) - \mathbb{E}\big[\Phi\big((\widetilde{X}_{-k}(u))_{k\geq 0},\boldsymbol{\theta}\big)\big]\Big| \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0\,,$$

see the discussion in the Appendix of [13]. 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

$$\sup_{u\in[\varepsilon,1-\varepsilon]} \sup_{\boldsymbol{\theta}\in\Theta} \Big| \frac{1}{nh_n} \sum_{k=1}^{n} \Phi\big((\widetilde{X}_{k-t}(u))_{t\geq 0},\boldsymbol{\theta}\big)\, K\Big(\frac{\frac{k}{n}-u}{h_n}\Big) - \mathbb{E}\big[\Phi\big((\widetilde{X}_{-k}(u))_{k\geq 0},\boldsymbol{\theta}\big)\big]\Big| \underset{n\to+\infty}{\overset{\mathbb{P}}{\longrightarrow}} 0\,.$$

From usual arguments, see for instance [6], this uniform version of the weak law of large number obtained in Lemma 7.2 will follow from the equicontinuity of the family

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

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

$$\|\widetilde{X}_t(u) - \widetilde{X}_t(u')\|_p \leq \frac{\big\|\boldsymbol{\theta}^*(u') - \boldsymbol{\theta}^*(u)\big\|}{1 - B_0(\Theta)} \Big(\frac{B_1(\Theta)\, C_0(\Theta)}{1 - B_0(\Theta)} + C_1(\Theta) \Big)\,, \tag{82}$$

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

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

We upper-bound

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\partial^1_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(\widetilde{X}_{t-1}(u),\widetilde{X}_{t-2}(u),\dots,\xi_t)\right\|_p \leq \sum_{k=1}^{\infty} b_k^{(1)}(\Theta)\left\|\widetilde{X}_{t-k}(u)\right\|_p + \sup_{\boldsymbol{\theta}\in\Theta}\left\|\partial^1_{\boldsymbol{\theta}}F_{\boldsymbol{\theta}}(0,0,\dots,\xi_t)\right\|_p .$$

By a similar argument than in the proof of Lemma 6.5, we deduce that (82) holds. □

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 $\partial_{\boldsymbol{\theta}}H_t\big(\boldsymbol{\theta}^*(u)\big)=\Big(\frac{\partial}{\partial\theta_i}H_t\big(\boldsymbol{\theta}^*(u)\big)\Big)_{1\le i\le d}$ for $i_n\le t\le j_n$ satisfies a multidimensional central limit theorem, where we denote $H_t\big(\boldsymbol{\theta}\big)=\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)$. We notice first that $\partial_{\boldsymbol{\theta}}\mathbb{E}\big[\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}^*(u)\big)\mid\mathcal{F}_0\big]=0$ as $\boldsymbol{\theta}^*(u)$ is the unique minimizer of $\mathbb{E}\big[\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}^*(u)\big)\mid\mathcal{F}_0\big]$ over the open set $\overset{o}{\Theta}$.
The function $\boldsymbol{\theta}\in\Theta\mapsto\mathbb{E}\big[\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}^*(u)\big)\mid\mathcal{F}_0\big]$ is differentiable under the condition $\big\|\partial_{\boldsymbol{\theta}}\Phi\big\|\in\mathrm{Lip}_p(\Theta)$. Thus $\partial_{\boldsymbol{\theta}}H_t\big(\boldsymbol{\theta}^*(u)\big)$ constitutes a differences of martingale sequence. We also have $\mathbb{E}\big[\big\|\partial_{\boldsymbol{\theta}}\Phi\big\|^2\big]<\infty$ and we can apply the CLT for differences of martingale sequences (See for instance [6]). As a consequence we can apply the point 2. of Lemma 7.1 and obtain the multidimensional central limit theorem

$$\frac{1}{\sqrt{nh_n}}\sum_{t=i_n(u)}^{j_n(u)}\partial_{\boldsymbol{\theta}}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}^*(u)\big)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\xrightarrow[n\to+\infty]{\mathcal{L}}\mathcal{N}\Big(0\,,\,\Sigma\big(\boldsymbol{\theta}^*(u)\big)\Big)\quad(83)$$

$$\text{with}\quad\Sigma\big(\boldsymbol{\theta}^*(u)\big)=\int_{\mathbb{R}}K^2(x)dx \times\sum_{t\in\mathbb{Z}}\Big(\mathrm{Cov}\,\big[\frac{\partial}{\partial\theta_i}\Phi\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}^*(u)\big)\,,\,\frac{\partial}{\partial\theta_j}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}^*(u)\big)\big]\Big)_{1\le i,j\le d}.$$

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

$$\begin{aligned}\frac{1}{\sqrt{nh_n}}&\sum_{t=i_n(u)}^{j_n(u)}\partial_{\boldsymbol{\theta}}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\widehat{\boldsymbol{\theta}}(u)\big)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\\&=\frac{1}{\sqrt{nh_n}}\sum_{t=i_n(u)}^{j_n(u)}\partial_{\boldsymbol{\theta}}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}^*(u)\big)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\\&+\sqrt{nh_n}\cdot\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}\partial^2_{\boldsymbol{\theta}^2}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\overline{\boldsymbol{\theta}}(u)\big)K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\big(\widehat{\boldsymbol{\theta}}(u)-\boldsymbol{\theta}^*(u)\big),\quad(84)\end{aligned}$$

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

$$\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}\big(\partial^2_{\boldsymbol{\theta}^2}\Phi((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\overline{\boldsymbol{\theta}}(u))-\mathbb{E}\big[\partial^2_{\boldsymbol{\theta}^2}\Phi((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\overline{\boldsymbol{\theta}}(u))\big]\big)K\big(\tfrac{\frac{t}{n}-u}{h_n}\big) \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{85}$$

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

$$\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}\partial^2_{\boldsymbol{\theta}^2}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\overline{\boldsymbol{\theta}}(u)\big)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big) \xrightarrow[n\to+\infty]{\mathbb{P}} \Gamma(\boldsymbol{\theta}^*(u)).$$

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

$$\frac{1}{nh_n}\sum_{t=i_n(u)}^{j_n(u)}\partial_{\boldsymbol{\theta}}\Phi\big((X^{(n)}_{t-k}(u))_{k\in\mathbb{N}},\widehat{\boldsymbol{\theta}}(u)\big)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big) = 0. \tag{86}$$

Using the assumptions on the Lipschitz coefficients of $\partial_{\boldsymbol{\theta}}\Phi$, the same inequalities as (80) and (81) in the proof of Proposition 7.1 lead for a convenient constant $C>0$ to:

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

As a consequence we deduce that:

$$\begin{aligned}
&\Big\|\frac{1}{\sqrt{nh_n}}\sum_{t=i_n(u)}^{j_n(u)}\partial_{\boldsymbol{\theta}}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\widehat{\boldsymbol{\theta}}(u)\big)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big)\Big\|_1 \leq C\left(\frac{1}{\sqrt{nh_n}}+h_n^{\rho}\sqrt{nh_n}\right)\\
&\qquad\Longrightarrow \frac{1}{\sqrt{nh_n}}\sum_{t=i_n(u)}^{j_n(u)}\partial_{\boldsymbol{\theta}}\Phi\big((\widetilde{X}_{t-k}(u))_{k\in\mathbb{N}},\widehat{\boldsymbol{\theta}}(u)\big)\,K\Big(\frac{\frac{t}{n}-u}{h_n}\Big) \xrightarrow[n\to+\infty]{\mathbb{P}} 0,
\end{aligned} \tag{87}$$

by using (14). Finally, from (84), using (85), (87), Slutsky Lemma and (83), we deduce:

$$\sqrt{nh_n}\;\Gamma(\boldsymbol{\theta}^*(u))\big(\widehat{\boldsymbol{\theta}}(u)-\boldsymbol{\theta}^*(u)\big) \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}\Big(0\,,\,\Sigma\big(\boldsymbol{\theta}^*(u)\big)\Big),$$

and this leads to Theorem 3.2. □

Proof of Proposition 4.2. We already proved in Section 4.3 that $\Phi_{LARCH} \in \mathrm{Lip}_4(\Theta)$ as well as Assumption $\mathbf{Co}(\Phi_{LARCH}, \Theta)$ when condition (41) 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 $\partial_{\boldsymbol{\theta}} \Phi_{LARCH} \in \mathrm{Lip}_4(\Theta)$. Indeed we use the estimates

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

Therefore, using Hölder and Minkowski Inequalities, we obtain

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

This inequality implies that $\partial_{\boldsymbol{\theta}} \Phi_{LARCH} \in \mathrm{Lip}_4(\Theta)$ under the assumptions of Proposition 4.2.

We also have to establish that under conditions of Proposition 4.2,

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

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

$$\begin{aligned}
\partial_{\boldsymbol{\theta}} \Phi_{LARCH}\big((\widetilde{X}_k(u))_{k\le 0}, \boldsymbol{\theta}^*(u)\big) \;=\; & -4(\xi_0^2 - 1)\Big(a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} a_i(\boldsymbol{\theta}^*(u))\, \widetilde{X}_{-i}(u)\Big)^3 \\
& \times \Big(\partial_{\boldsymbol{\theta}} a_0(\boldsymbol{\theta}^*(u)) + \sum_{i=1}^{\infty} \partial_{\boldsymbol{\theta}} a_i(\boldsymbol{\theta}^*(u))\, \widetilde{X}_{-i}(u)\Big)^2 \qquad (88)
\end{aligned}$$

Therefore, using Hölder and Minkowski Inequalities and independence of ξ_0 and $(\widetilde{X}_k(u))_{k\le -1}$

together with $\mathcal{E}_u \equiv \mathbb{E}\Big[\big\|\partial_{\boldsymbol{\theta}}\Phi_{LARCH}\big((\widetilde{X}_k(u))_{k\leq 0},\boldsymbol{\theta}^*(u)\big)\big\|^2\Big]$ we derive that

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

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

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

As a consequence, similar arguments as previously entail

$$\mathbb{E}\Big[\big\|\partial^2_{\boldsymbol{\theta}^2}\Phi_{LARCH}\big((\widetilde{X}_k(u))_{k\leq 0},\boldsymbol{\theta}^*(u)\big)\big\|\Big]<\infty$$

from Hausdorff and Minkowski inequalities. Finally we checked the conditions of Theorem 3.2 since the asymptotic covariance matrix $\Sigma\big(\boldsymbol{\theta}^*(u)\big)$ and $\Gamma(\boldsymbol{\theta}^*(u))$ are positive definite matrix from (42) using (88). □

Proof of Proposition 4.1. We proved in Section 4 that $\alpha_k(\Phi_G,\Theta)=b_k^{(0)}(\Theta)$ and $\Phi_G\in\mathrm{Lip}_3(\Theta)$ when $f_{\boldsymbol{\theta}}$ and $M_{\boldsymbol{\theta}}$ satisfy Lipschitz inequalities (20). But we also have:

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

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

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

Thus

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

from Jensen inequality and since we assume that $f_{\boldsymbol{\theta}}$, $M_{\boldsymbol{\theta}}$, $\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}}$ and $\partial_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}}$ satisfy Lipschitz inequalities (20). As a consequence we derive

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

therefore $\partial_{\boldsymbol{\theta}}\Phi_G \in \mathrm{Lip}_4(\Theta)$. From these computations and with the inequality (21) we also deduce that condition (28) implies $B_0(\Theta) < 1$, and $\sum_{t=2}^{\infty} t\log(t) b_t(\Theta) < \infty$ follows from $\sum_{s\geq 0} s\,\alpha_s(\Phi,\Theta) < \infty$, required in Theorems 3.1 and 3.2. Similar calculations also entail

$$\mathbb{E}\Big[\big\|\partial^2_{\boldsymbol{\theta}^2}\Phi\big((\widetilde{X}_{-k}(u))_{k\in\mathbb{N}},\boldsymbol{\theta}\big)\big\|^2\Big] < \infty, \quad \text{for any} \quad \boldsymbol{\theta}\in\Theta,$$

since $p = 4$ and $\partial^2_{\boldsymbol{\theta}^2} f_{\boldsymbol{\theta}}$ and $\partial^2_{\boldsymbol{\theta}^2} M_{\boldsymbol{\theta}}$ satisfy Lipschitz inequalities (20). We also have

$$\mathbb{E}\big[\big\|\partial_{\boldsymbol{\theta}}\Phi\big((\widetilde{X}_k(u))_{k\le 0},\boldsymbol{\theta}^*(u)\big)\big\|^2\big] \le \frac{12}{(1\vee \underline{M})^2}\Big(\big\|\partial_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}^*(u)}((\widetilde{X}_k(u))_{k\le -1})\big\|_2^2 + \big\|\partial_{\boldsymbol{\theta}} f_{\boldsymbol{\theta}^*(u)}((\widetilde{X}_k(u))_{k\le -1})\big\|_2^2 \big\|\xi_0\big\|_2^2 + \big\|\partial_{\boldsymbol{\theta}} M_{\boldsymbol{\theta}^*(u)}((\widetilde{X}_k(u))_{k\le -1})\big\|_2^2 \big\|\xi_0\big\|_4^4\Big),$$

since $\widetilde{X}_0(u) - f_{\boldsymbol{\theta}^*(u)}((\widetilde{X}_k(u))_{k\le -1}) = M_{\boldsymbol{\theta}^*(u)}((\widetilde{X}_k(u))_{k\le -1})\,\xi_0$, with $M_{\boldsymbol{\theta}^*(u)}((\widetilde{X}_k(u))_{k\le -1})$ and ξ_0 which are independent. Therefore, we obtain

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

since $p = 4$. Finally (27) ensures that asymptotic covariance matrix $\Sigma\big(\boldsymbol{\theta}^*(u)\big)$ and $\Gamma(\boldsymbol{\theta}^*(u))$ are positive definite matrix (see [5]) and (26) implies the existence and the uniqueness of $\boldsymbol{\theta}^*(u)$ as the minimum of $\boldsymbol{\theta} \in \Theta \mapsto \mathbb{E}\big[\Phi\big((\widetilde{X}_k(u))_{k\le 1}),\boldsymbol{\theta}\big) \mid \mathcal{F}_0\big]$ defined in (10) (see also [5]). This ends the checking of the conditions of Theorem 3.2. □

Aknowledgments 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] Azrak, R. and Melard, G. (2006). Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients. *Statistical Inference for Stochastic Processes* **9** 279-330.

[2] Bardet, J. M., Boularouk, Y. and Djaballah, K. (2017). Asymptotic behavior of the Laplacian quasi-maximum likelihood estimator of affine causal processes. *Electronic Journal of Statistics* **11** 452-479.

[3] Bardet, J. M. and Doukhan, P. (2018). Non-parametric estimation of time varying AR(1)–processes with local stationarity and periodicity. *Electronic Journal of Statistics* **12** 2323-2354.

[4] Bardet, J. M., Kengne, W. and Wintenberger, O. (2012). Multiple breaks detection in general causal time series using penalized quasi-likelihood. *Elec. Journal Statist.* 435–477.

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

[6] Billingsley, P. (1999). *Convergence of probability measures*. Second edition, Wiley and sons.

[7] Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. *Stochastic Processes and Applications* **62** 139-168.

[8] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Annals of Statistics* **25** 1-37.
[9] Dahlhaus, R. (2000). A Likelihood Approximation for Locally Stationnary Processes. *Annals of Statistics* **28** 1762-1794.
[10] Dahlhaus, R. (2009). Empirical spectral processes for locally stationary time series. *Bernoulli* **15** 1-39.
[11] Dahlhaus, R. (2012). Locally Stationary Processes. In *Time Series Analysis: Methods and Applications*, (T. Subba Rao, S. Subba Rao and C. R. Rao, eds.). *Handbook of statistics* **30** Elsevier.
[12] Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *The Annals of Statistics* **34** 2790–2824.
[13] Dahlhaus, R., Richter, S. and Wu, W. B. (2019). Towards a general theory for nonlinear locally stationary processes. *Bernoulli* **25** 1013–1044.
[14] Dahlhaus, R. and Subba Rao, S. (2006). Statistical Inference for Time Varying ARCH processes. *The Annals of Statistics* **34** 1075-1114.
[15] Dedecker, J., Doukhan, P., Lang, G., León, J. R., Louhichi, S. and Prieur, C. (2007). *Weak dependence: With Examples and Applications*. Lecture Notes in Statistics **190**, Springer-Verlag, New York.
[16] Dedecker, J. and Prieur, C. (2004). Coupling for $\tau-$Dependent Sequences and Applications. *Journal of Theoretical Probability* **17** 861-885.
[17] Ding, Z., Granger, C. W. J. and Engle, R. (1993). A long memory property of stock market returns and a new model. *J. Empirical Finance* **1** 83-106.
[18] Doukhan, P., Fokianos, K. and Tjøstheim, D. (2012). On weak dependence conditions for Poisson autoregressions. *Statistics and Probability Letters* **82** 942-948.
[19] Doukhan, P. and Kengne, W. (2015). Inference and testing for structural change in general Poisson autoregressive models. *Electronic Journal of Statistics* **9** 1267-1314.
[20] Doukhan, P. and Truquet, L. (2007). A fixed point approach to model random fields. *ALEA Lat. Am. J. Probab. Math. Stat.* **3** 111–132.
[21] Doukhan, P. and Wintenberger, O. (2008). Weakly dependent chains with infinite memory. *Stochastic Processes and Applications* **118** 1997-2013.
[22] Francq, C. and Zakoïan, J. M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.
[23] Francq, C. and Zakoïan, J.-M. (2010). Inconsistency of the MLE and inference based on weighted LS for LARCH models. *Journal of Econometrics* **159** 151-165.
[24] Giraitis, L., Kokoszka, P. and Leipus, R. (2000). Stationary ARCH models: dependence structure and Central Limit Theorem. *Econometric Theory* **1** 3–22.
[25] Giraitis, L., Leipus, R., Robinson, P. and Surgailis, D. (2003). LARCH, leverage and long memory. *Journal of Financial Econometrics* **2** 177–210.
[26] Ling, S. and McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric Theory* **19** 280-310.
[27] Paraschakisa, K. and Dahlhaus, R. (2012). Frequency and phase estimation in time series with quasi periodic components. *Journal of Time Series Analysis* **33**

13-31.

[28] Pfanzagl, J. (1969). On the mesurability and consistency of minimum contrast estimates. *Metrika* **14** 249–334.

[29] Robinson, P. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* **47** 67–84.

[30] Rohan, N. (2013). A time varying GARCH() model and related statistical inference. *Statistics & Probability Letters* **83** 1983-1990.

[31] Rosenblatt, M. (1991). *Stochastic curve estimation* **3**. NSF-CBMS Regional Conference Series in Probability and Statistics.

[32] Straumann, D. (2005). *Estimation in conditionally heteroscedastic time series models. Lecture Notes in Statistics* **181**. Springer-Verlag, Berlin.

[33] Truquet, L. (2017). Parameter stability and semiparametric inference in time-varying ARCH models. *Journal of Royal Statistical Society Series B* **79** 1391–1414.