A new estimator for LARCH processes

Jean-Marc Bardet

A new estimator for LARCH processes

Jean-Marc Bardet, bardet@univ-paris1.fr,
University Paris 1 Panthéon-Sorbonne, SAMM, France

March 20, 2023

Abstract

The aim of this paper is to provide a new estimator of parameters for LARCH(∞) processes, and thus also for LARCH(p) or GLARCH(p, q) processes. This estimator results from minimising a contrast leading to a least squares estimator for the absolute values of the process. Strong consistency and asymptotic normality are shown, and convergence occurs at the rate $\sqrt{n}$ as well in short or long memory cases. Numerical experiments confirm the theoretical results and show that this new estimator significantly outperforms the smoothed quasi-maximum likelihood estimators or weighted least squares estimators commonly used for such processes.

Keywords: LARCH process; Semiparametric estimation; Long memory process
MSC2010: Primary 62F12, 62M10; Secondary 91B84

1 Introduction

Linear AutoRegressive Conditional Heteroskedastic (LARCH) processes were introduced by Robinson (1991) to model the long-range dependence of volatility and leverage. They are studied for their stationarity and dependence properties in Robinson and Zaffaroni (1997), Giraitis *et al.* (2000), Berkes and Horváth (2003), and Giraitis *et al.* (2004). A LARCH(∞) process $(X_t)_{t\in\mathbb{Z}}$ is defined by:

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

where $(\xi_t)_{t\in\mathbb{Z}}$ is a white noise. In what follows, we will consider a parametric class of such LARCH(∞) processes, *i.e.* $a_j = a_j(\theta)$ for any $j \in \mathbb{N}$ with $\theta \in \mathbb{R}^\ell$, where the functions $a_j(\cdot)$ are assumed to be known but the true parameter θ^* is unknown. This paper is devoted to studying the asymptotic properties of a new estimator of θ^* (rather than a component of θ^*, such as the location parameter, as was done in Beran, 2006) for LARCH(∞) processes when a trajectory $(X_1, \ldots, X_n)$ is observed (see more details in section 2).

LARCH(∞) processes, which conditionally represent heteroskedastic weak white noise, offer new perspectives for modeling financial data. This model has the advantage over GARCH formulations that volatility can be arbitrarily close to 0, which is nevertheless the case for certain financial series (see, for example, the example on CAC40 index returns discussed in section 4). The square of a LARCH(∞) process can also exhibit the long memory property (see below), which is impossible for a stationary ARCH(∞) process (see more theoretical details in Giraitis *et al.*, 2009, and the

same illustrative example in section 4).

For numerous affine causal processes, such as ARMA, GARCH, ARMA-GARCH, AR(∞), or ARCH(∞) processes, the Gaussian quasi-maximum likelihood (QML) provides a very accurate estimator (see further details in Bardet and Wintenberger, 2009). Even though a LARCH(∞) process or its special cases LARCH(p) or GLARCH(p, q) (see their definitions in (5) and (7)) are also causal affine time series, such a contrast cannot be used to estimate the parameter θ^*. Indeed, the conditional variance of X_t cannot be bounded near 0, and this does not allow asymptotic results for such contrasts (see more details on this point in Beran and Schützner, 2009, Truquet, 2014, and in particular in Francq and Zakoïan, 2010). Beran and Schützner (2009) and Truquet (2014) propose an interesting alternative estimator based on a family of smooth approximations of the QML estimate, and they establish the consistency and asymptotic normality of the estimator of θ^* in cases of short or long memory. Francq and Zakoïan (2010) preferred to construct weighted least squares estimators, for which they also show consistency and asymptotic normality. Note that they also extend their results to AR(p)-LARCH(q) processes as well as to Truquet.

We propose a new estimator obtained by minimizing a least-squares contrast of the absolute values of (X_t) (see its precise definition in (12), section 3). Under assumptions that are not too restrictive, especially considering short- and long-term memory, strong consistency and asymptotic normality are established for this estimator. Moreover, only a fourth-order white noise moment is required for asymptotic normality (order 4 in Truquet, 2014, and 5 in Beran and Schützner, 2009, for the smoothed QML estimator, no order condition in Francq and Zakoïan, 2010, for the weighted LS estimator when an appropriate weight family is chosen). A convergence rate $\sqrt{n}$ for this new estimator is proved for LARCH(∞) processes with short memory as well as for the smoothed QML estimator (Truquet, 2014) and for the weighted LS estimator (Francq and Zakoïan, 2010), but with different asymptotic covariance matrices. This rate of convergence $\sqrt{n}$ is also established for the LARCH(∞) process with long memory, while for the smoothed QMLE in Beran and Schützner (2009) only a rate n^β with $0 < \beta < 1/2$ is obtained. Note that such a rate of convergence is also obtained with QMLE for generalized quadratic ARCH processes in Grublyte *et al.* (2017).
Monte Carlo experiments confirm the asymptotic behavior of the estimator even for trajectories with not very large lengths. The performances of this new estimator are then compared with those obtained with the regularized QMLE (for which the choice of the regularization parameter is a real problem) and with those obtained with the weighted least squares estimator of Francq and Zakoïan (2010). The results of these comparisons undoubtedly show the much faster convergence of this new estimator, especially compared to the smoothed QMLE.

The following section 2 is devoted to the definition and stationarity conditions of the considered LARCH(∞) processes. The main results concerning the definition and the asymptotic behavior of the new estimator are given in section 3. Numerical experiments are proposed in section 4 and proofs are established in section 5.

2 LARCH(∞) processes

For $\ell \in \mathbb{N}^*$ and $u = (u_i)_{1\leq i\leq \ell} \in \mathbb{R}^\ell$, denote $\|u\| = \sqrt{u_1^2 + \cdots + u_\ell^2}$ the usual Euclidian norm. More generally, for $k \in \mathbb{N}^*$, if $u = (u_{i_1,\ldots,i_k})_{1\leq i_1,\ldots,i_k\leq \ell} \in \mathbb{R}^{\ell^k}$, denote $\|u\| = \sqrt{\sum_{1\leq i_1,\ldots,i_k\leq \ell} u_{i_1,\ldots,i_k}^2}$. Denote also $\|Z\|_p = \big(\mathbb{E}\big[\|Z\|^p\big]\big)^{1/p}$ for $p \geq 1$ where Z is a random vector. For any $m \in \mathbb{N}^*$, any $\ell \in \mathbb{N}^*$ and $\psi : \theta \in \Theta \subset \mathbb{R}^\ell \mapsto \psi(\theta) \in \mathbb{R}$ such as $\psi \in \mathcal{C}^m(\Theta)$, the space of m-times continuously differentiable functions on Θ, denote for $1 \leq k \leq m$:

$$\partial_\theta \psi(\theta) := \Big(\frac{\partial}{\partial \theta_i}\psi(\theta)\Big)_{1\leq i\leq \ell}$$

$$\text{and} \quad \partial^k_{\theta^k}\psi(\theta) := \Big(\partial^k_{\theta_{i_1}\cdots\theta_{i_k}}\psi(\theta)\Big)_{1\leq i_1,\ldots,i_k\leq \ell} = \Big(\frac{\partial^k}{\partial\theta_{i_1}\cdots\partial\theta_{i_k}}\psi(\theta)\Big)_{1\leq i_1,\ldots,i_k\leq \ell}.$$

Here we study a LARCH(∞) process introduced in Robinson (1991) and also studied in Giraitis *et al.* (2000), Giraitis *et al.* (2004), Beran and Schützner (2009), Francq and Zakoïan (2010) or Truquet (2014). For $r \geq 2$, we will consider the following assumption:

Assumption $A(r)$:

- $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of symmetric centered independent random variables with continuous distribution such as $\|\xi_0\|_1 = 1$ and $\|\xi_0\|_r < \infty$;
- For any $j \in \mathbb{N}$, $\theta \in \mathbb{R}^\ell \mapsto a_j(\theta) \in \mathbb{R}$ are known continuous functions and without loss of generality we will assume $a_0(\theta) > 0$ for any $\theta \in \mathbb{R}^\ell$.

We will define a LARCH(∞) process $(X_t)_{t\in\mathbb{Z}}$ using Assumption $A(r)$ with $r \geq 2$. Before this, define:

$$\Theta(2) = \Big\{\theta \in \mathbb{R}^\ell,\ \|\xi_0\|_2^2 \sum_{j=1}^\infty a_j^2(\theta) < 1\Big\}. \tag{1}$$

Then, under Assumption $A(r)$ with $r \geq 2$, for $\theta^* \in \Theta$ where Θ is a compact subset of $\Theta(2)$, we define a LARCH(∞) process $(X_t)_{t\in\mathbb{Z}}$ by:

$$X_t = \xi_t\Big(a_0(\theta^*) + \sum_{j=1}^\infty a_j(\theta^*)\, X_{t-j}\Big) \quad \text{for any } t \in \mathbb{Z}. \tag{2}$$

Under these conditions, Giraitis *et al.* (2004) have proved the stationarity of (X_t) and the existence of $\|X_0\|_2$. From now on we will assume that θ^* is unknown

Remark 2.1. *The case $a_0(\theta) = 0$ implies $X_t = 0$ a.s. for any $t \in \mathbb{Z}$, see Giraitis et al. (2000). This explains why, we assume $a_0(\theta) > 0$ for any $\theta \in \mathbb{R}^\ell$ in Assumption $A(r)$.*

Remark 2.2. *Note that in Assumption $A(r)$ we assume $\|\xi_0\|_1 = 1$ and not $\|\xi_0\|_2 = 1$ as is usually done. This is explained by the expression of the estimator we will consider. However, the difference between these two normalization options is only a new parametrization, because with $\xi'_t = \xi_t/\|\xi_0\|_2$ for any $t \in \mathbb{Z}$ and using the linearity of its expression, (2) could also be written as*

$$X_t = \xi'_t\Big(a'_0(\theta^*) + \sum_{j=1}^\infty a'_j(\theta^*)\, X_{t-j}\Big) \quad \textit{where } \|\xi'_0\|_2 = 1 \textit{ and } a'_j(\theta^*) = \|\xi_0\|_2\, a_j(\theta^*) \textit{ for any } j \in \mathbb{N}. \tag{3}$$

In the case of Gaussian white noise, for example, we have $\|\xi_0\|_2^2 = \sigma_\xi^2 = \pi/2$ *when* $\|\xi_0\|_1 = 1$.

In the sequel, under Assumption $A(r)$ with $r \geq 4$, we will also consider $\|X_0\|_4$ and for this we define

$$\Theta(4) = \Big\{\theta \in \mathbb{R}^\ell,\ \|\xi_0\|_4^4 \sum_{j=1}^{\infty} a_j^4(\theta) + 6\,\|\xi_0\|_2^2 \sum_{j=1}^{\infty} a_j^2(\theta) < 1\Big\}. \tag{4}$$

Remark 2.3. *Unless we consider a particular distribution for the noise, for example and typically the Gaussian* $\mathcal{N}\big(0, (\sqrt{\pi/2})^2\big)$*, the moments* $\|\xi_0\|_2 \leq \|\xi_0\|_4$ *are unknown and can take any value greater than or equal to* 1 *according to this distribution. So the sets* $\Theta(2)$ *and* $\Theta(4)$ *are indeed unknown if we consider only the hypothesis of a symmetric noise* (ξ_t) *with a moment of order* 4 *and satisfying* $\mathbb{E}\big[|\xi_0|\big] = 1$*. Note, however, that it is exactly the same when estimating the parameters of a GARCH*(p,q) *or ARCH*(∞) *process by quasi-maximum likelihood, e.g. when we define the stationarity set from the Lyapunov exponents depending on the noise distribution, or when we consider a condition of* $\|\xi_0\|_4$ *for asymptotic normality. Numerically (see section 4), one will use domains of minimisation in* θ *much larger than* $\Theta(2)$ *or* $\Theta(4)$*, for example* $0 \leq a_1,\, a_2 \leq 1$ *for a LARCH*(2) *process, but the condition of belonging to* $\Theta(2)$ *will be able to be checked a posteriori from the value taken by the estimator. The fact that this condition is not verified when applying a long-memory LARCH*(∞) *model to financial data led us to choose another more complex long-memory LARCH*(∞) *model.*

Three interesting special cases of LARCH(∞) processes can be mentioned:

1. A first special case of LARCH(∞) processes are the LARCH(p) processes defined by:

$$X_t = \xi_t\,\sigma_t \quad \text{with} \quad \sigma_t = a_0 + \sum_{i=1}^{p} a_i\,X_{t-i}, \qquad \text{for any } t \in \mathbb{Z} \tag{5}$$

   Therefore, a LARCH(p) process is a LARCH(∞) process defined in (2) with $a_k(\theta) = a_k$ for $0 \leq k \leq p$ and $\theta = {}^t\big(a_0, a_1, \ldots, a_p\big) \in (0,\infty) \times \mathbb{R}^p$. For such a process, the sets $\Theta(2)$ defined in (1) and $\Theta(4)$ defined in (4) become respectively:

$$\begin{aligned} \Theta_p(2) &= \Big\{\theta = {}^t\big(a_0, a_1, \ldots, a_p\big) \in (0,\infty) \times \mathbb{R}^p,\ \|\xi_0\|_2^2 \sum_{j=1}^{p} a_j^2 < 1\Big\} \\ \text{and}\ \ \Theta_p(4) &= \Big\{\theta = {}^t\big(a_0, a_1, \ldots, a_p\big) \in (0,\infty) \times \mathbb{R}^p,\ \|\xi_0\|_4^4 \sum_{j=1}^{p} a_j^4 + 6\,\|\xi_0\|_2^2 \sum_{j=1}^{p} a_j^2 < 1\Big\}. \end{aligned} \tag{6}$$

2. A natural extension of the LARCH(p) processes under consideration are GLARCH(p,q) processes, which follow the same procedure as the well-known transition from ARCH processes to GARCH processes. A GLARCH(p,q) process is defined by

$$X_t = \xi_t\,\sigma_t \quad \text{with} \quad \sigma_t = c_0 + \sum_{i=1}^{p} c_i\,X_{t-i} + \sum_{j=1}^{q} d_j\,\sigma_{t-j}, \qquad \text{for any } t \in \mathbb{Z}. \tag{7}$$

   To study such a process, one defines the polynomials $P(x) = 1 - \sum_{j=1}^{q} d_j\,x^j$ and $Q(x) = c_0 + \sum_{i=1}^{p} c_i\,x^i$. Then the previous iteration equation (7) is equivalent to $P(B)\,\sigma = Q(B)\,X$ where B is the usual backward operator. In the following we assume that P and Q are

coprime polynomials for $\theta = \theta^*$.
We define $\theta = \big(c_0, c_1, \ldots, d_1, \ldots, d_q\big)' \in (0,\infty) \times \mathbb{R}^{p+q}$ and the coefficients $a_k(\theta)$ decrease exponentially decrease to 0 when $k \to \infty$ (as it is usually known for ARMA(p,q) processes since the roots of P lie outside the unit circle).
For GLARCH(p,q) process, the assumption for obtaining a stationary 2nd-order solution of (7) is $\theta \in \Theta_{p,q}(2)$, with

$$\Theta_{p,q}(2) = \Big\{\theta \in (0,\infty) \times (-1,1)^{p+q},\ \sum_{i=1}^{q} d_i^2 + \|\xi_0\|_2 \sum_{j=1}^{p} c_j^2 < 1\Big\}. \tag{8}$$

The calculation of $\Theta(4)$ for such GLARCH(p,q) processes is not quite straightforward. In Giraitis *et al.* (2004) $\Theta(4)$ is simplified for GLARCH$(1,1)$ and it is established that:

$$\Theta_{1,1}(4) = \Big\{\theta = {}^t(c_0, c_1, d_1) \in (0,\infty) \times \mathbb{R}^2,\quad \|\xi_0\|_4^4 \,\frac{c_1^4}{1-d_1^4} + 6\,\|\xi_0\|_2^2\,\frac{c_1^2}{1-d_1^2} < 1\Big\}.$$

3. Another case we will study is that of LARCH(∞) with long memory, *i.e.* such that there are $d(\theta) \in (0,1/2)$ and $L_\theta(\cdot)$ a slowly varying function such that:

$$a_j(\theta) = L_\theta(j)\, j^{d(\theta)-1} \quad \text{for } j \in \mathbb{N}^*. \tag{9}$$

This case was considered in particular in Robinson and Zaffaroni (1997) and Beran and Schützner (2009). In this paper, a parametric estimation procedure was studied for the case where $a_j(\theta) = c\, j^{d-1}$ for $j \in \mathbb{N}^*$ and $\theta = {}^t(a_0, c, d)$ (see further details below). Note that in such a case $\sum_{j=1}^{\infty} |a_j(\theta)| = \infty$ but $\sum_{j=1}^{\infty} a_j^2(\theta) < \infty$.

Remark 2.4. *For this third type of example we focused on the long memory property and for that the parameter $d(\theta)$ (or simply the parameter d in Beran and Schützner's example) is necessarily in $(0,1/2)$. However, nothing prevents to extend the definition of the process to $d(\theta) \in (-\infty, 1/2)$ or to a compact set included in $(-\infty,1/2)$ when the estimator is applied. We would then lose the exclusive long memory character but we would gain in generality.*

3 A new estimator of LARCH parameters

3.1 Definition and consistency of the estimator

We consider here a special case of M-estimators for estimating θ^* from an observed trajectory $(X_1, \ldots, X_n)$ of a stationary solution of (2). For this purpose, let the following contrast function Φ for $x \in \mathbb{R}^{\mathbb{N}}$ and $\theta \in \mathbb{R}^\ell$ be defined by

$$\Phi(x,\theta) = \Big(|x_1| - \big|a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta)\, x_{j+1}\big|\Big)^2. \tag{10}$$

Remark 3.1. *The choice of this contrast function Φ follows from the fact that under a classical identifiability assumption (see Assumption Id(Θ) below), θ^* is the unique minimum in Θ of $\mathbb{E}\big[\Phi\big((X_{-k})_{k\geq 0}, \theta\big)\big]$ under the normalization condition $\|\xi_0\|_1 = 1$ (see the proof of Proposition 3.1, Part 3.).*

Other contrast functions satisfy this property, such as $\Phi_4(x,\theta) = \Big(x_1^2 - \big(a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta) x_{j+1}\big)^2\Big)^2$ *under the usual normalization condition* $\|\xi_0\|_2 = 1$. *Such M-estimators defined from* Φ_4 *require moments of order 8 to preserve their asymptotic normality, while for* Φ *a moment of order 4 is sufficient. Note that weighted quadratic contrast* Φ_{FZ} *derived from* Φ_4 *(see its definition in* (25)*) was defined in Francq and Zakoïan (2010) and allows, with appropriate weights, to obtain the asymptotic normality without a moment condition (except* $r \geq 1$*). However, Monte Carlo experiments show that the convergence rate of the estimator defined by* Φ *is faster than that of the estimator defined by* Φ_4 *or* Φ_{FZ}.
Note, however, that contrast functions such as

$$\Phi_2(x,\theta) = \Big(x_1 - \big(a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta) x_{j+1}\big)\Big)^2 \quad or \quad \Phi_1(x,\theta) = \Big|x_1 - \big(a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta) x_{j+1}\big)\Big|$$

are not such that $\mathbb{E}\big[\Phi\big((X_{-k})_{k\geq 0},\theta\big)\big]$ *has a unique minimum in* θ^* *on the set* Θ. *For example, quick calculations show that:*

$$\mathbb{E}\big[\Phi_2\big((X_{-k})_{k\geq 0},\theta\big)\big] = \big(\|\xi_0\|_2^2 - 1\big)\,\mathbb{E}\Big[\big(a_0(\theta^*) + \sum_{j=1}^{\infty} a_j(\theta^*) X_{-j}\big)^2\Big] + \mathbb{E}\Big[\big(a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta) X_{-j}\big)^2 - \big(a_0(\theta^*) + \sum_{j=1}^{\infty} a_j(\theta^*) X_{-j}\big)^2\Big],$$

which in general does not have a minimum at θ^*.

Remark 3.2. *Unlike, in particular, the papers by Francq and Zakoïan (2010) or Truquet (2014), the case of an AR(*p'*)-LARCH(*p*) process or, more generally, an AR(*p'*)-LARCH(*∞*) process is not treated here. Based on the* Φ *contrast chosen, it would also have been possible to extend the study to AR(*p'*)-LARCH(*∞*) processes by considering the* Φ' *contrast such that:*

$$\Phi'(x,\theta') = \Big(\Big|x_1 - \sum_{k=1}^{p'} b_k\, x_{1+k}\Big| - \Big|a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta)\,\Big|x_{j+1} - \sum_{k=1}^{p'} b_k\, x_{j+1+k}\Big|\Big)^2,$$

where $\theta' = {}^t\big(b_1,\ldots,b_{p'},{}^t\theta\big)$. *The convergence and asymptotic normality proofs for LARCH(*∞*)-processes proposed in the rest of the manuscript could then be extended to AR(*p'*)-LARCH(*∞*) processes, but this would make them even more technical and difficult to follow.*

Now we define the process $(\widetilde{X})_{t\in\mathbb{Z}}$ by:

$$\widetilde{X}_t = \begin{cases} X_t & \text{for } t \geq 1 \\ 0 & \text{for } t \leq 0 \end{cases}. \tag{11}$$

From now on, let us consider Θ as a compact subset of $\Theta(2)$, implying that $(X_t)_{t\in\mathbb{Z}}$ is a stationary ergodic process satisfying $\|X_0\|_2 < \infty$. Then define the following estimator:

$$\widehat{\theta}_n = \underset{\theta\in\Theta}{\text{Argmin}}\ \frac{1}{n}\sum_{t=1}^{n} \Phi\big((\widetilde{X}_{t-k})_{k\geq 0},\theta\big). \tag{12}$$

In the sequel, we add an assumption about the derivatives of functions $\theta \in \Theta \mapsto a_k(\theta)$ for $k \in \mathbb{N}$ (we use the convention $\partial^0_{\theta^0} a_k(\cdot) = a_k(\cdot)$). So for $i \in \mathbb{N}$ define:

Assumption $C_i(\Theta)$**:** For any $k \in \mathbb{N}$, the functions $a_k \in \mathcal{C}^i(\Theta)$ and there exist $C_a > 0$ and $\overline{d} < 1/2$ satisfying

$$\sup_{\theta\in\Theta}\Big\{\sum_{j=0}^{i}\big\|\partial^j_{\theta^j}a_k(\theta)\big\|\Big\} \leq C_a\, k^{\overline{d}-1} \quad \text{for any } k \in \mathbb{N}^*. \tag{13}$$

As already mentioned in Beran and Schützner (2009), to prove the consistency and asymptotic normality of $\widehat{\theta}_n$, it is necessary to take the derivatives in θ of $\theta \in \Theta \mapsto M_\theta(t)$ for $t \in \mathbb{Z}$, where

$$M_\theta(t) = a_0(\theta) + \sum_{k=1}^{\infty} a_k(\theta)\, X_{t-k} \quad \text{for } t \in \mathbb{Z} \text{ and } \theta \in \Theta. \tag{14}$$

Note that the convergence of this infinite sum in the $\mathbb{L}^2(\Omega)$-norm is done using a Volterra decomposition, as in Giraitis *et al.* (2000) or Giraitis *et al.* (2004). However, the existence of $M_\theta(t)$derivatives could be problematic, since the sequence $(a_k(\theta))$ is not summable in the case of a long memory. We therefore consider the assumption:

Assumption (S): For every $t \in \mathbb{Z}$, $(M_\theta(t))_{\theta\in\Theta}$ is a separable stochastic process on Θ.

Note that this assumption is not really restrictive, since a stochastic process can always be replaced by a separable version (see Remark 1 in Beran and Schützner, 2009). Moreover, as proved in Proposition 2 of Beran and Schützner (2009), Assumption (S) combined with Assumption $C_1(\Theta)$ (resp. $C_2(\Theta)$), where Θ is a compact set of the $\Theta(2) \subset \mathbb{R}^\ell$, implies that $\theta \in \Theta \mapsto M_\theta(t)$ is almost surely is differentiable once (resp. twice).
We add a classical identification condition:

Assumption Id(Θ): If θ, $\widetilde{\theta} \in \Theta$,

$$\big(a_i(\theta) = a_i(\widetilde{\theta}) \text{ for all } i \in \mathbb{N}\big) \implies \big(\theta = \widetilde{\theta}\big). \tag{15}$$

Then we obtain the following conditions for the consistency of the two estimators:

Proposition 3.1. Under Assumption $A(2)$, if $\theta^* \in \Theta$ is an unknown parameter, where Θ is a compact subset of $\Theta(2) \subset \mathbb{R}^\ell$ defined in (1), consider $(X_t)_{t\in\mathbb{Z}}$ as a stationary solution of (2) and $(X_1, \ldots, X_n)$ an observed trajectory of (X_t). Assume also Assumption $C_\ell(\Theta)$, (S) and Id(Θ). Then $\widehat{\theta}_n \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} \theta^*$ where $\widehat{\theta}_n$ is defined in (12).

Proposition 3.1 can be specified for the special cases considered earlier, starting with the LARCH(p) processes.

Corollary 3.1. Under Assumption $A(2)$, with $p \geq 1$, let (X_t) be a LARCH(p) process, solution of

$$X_t = \xi_t\, \sigma_t \quad \text{with} \quad \sigma_t = a_0^* + \sum_{i=1}^{p} a_i^*\, X_{t-i}, \qquad \text{for any } t \in \mathbb{Z}, \tag{16}$$

where $\theta^* = {}^t\big(a_0^*, a_1^*, \ldots, a_p^*\big) \in \Theta$, a compact subset of $\Theta_p(2)$ defined in (6). Let $(X_1, \ldots, X_n)$ be an observed trajectory of (X_t). Then $\widehat{\theta}_n = {}^t\big(\widehat{a}_0^{(n)}, \widehat{a}_1^{(n)}, \ldots, \widehat{a}_p^{(n)}\big) \underset{n\to\infty}{\overset{a.s.}{\longrightarrow}} \theta^*$ where $\widehat{\theta}_n$ is defined in (12)

Note that since the functions a_k considered in the Corollary 3.1 are constant functions, the Assumptions Id(Θ), $C_\ell(\Theta)$ and (S) are satisfied. The same is true if we consider the case of GLARCH(p,q) processes:

Corollary 3.2. Under Assumption $A(2)$, with $p \geq 1,\ q \geq 1$, let (X_t) be a GLARCH(p,q) process, solution of

$$X_t = \xi_t\, \sigma_t \quad \text{with} \quad \sigma_t = c_0^* + \sum_{i=1}^{p} c_i^*\, X_{t-i} + \sum_{j=1}^{q} d_j^*\, \sigma_{t-j}, \qquad \text{for any } t \in \mathbb{Z}, \tag{17}$$

where $\theta^* = {}^t\big(c_0^*, \ldots, c_p^*, d_1^*, \ldots, d_q^*\big) \in \Theta$, a compact subset of $\Theta_{p,q}(2)$ defined in (8). Let $(X_1, \ldots, X_n)$ be an observed trajectory of (X_t). Then $\widehat{\theta}_n = {}^t\big(\widehat{c}_0^{(n)}, \ldots, \widehat{c}_p^{(n)}, \widehat{d}_1^{(n)}, \ldots, \widehat{d}_q^{(n)}\big) \xrightarrow[n\to\infty]{a.s.} \theta^*$ where $\widehat{\theta}_n$ is defined in (12).

The consistency of $\widehat{\theta}_n$ for long memory LARCH(∞) can also be derived from Proposition 3.1:

Corollary 3.3. Assume $A(2)$ where the sequence $(a_k(\theta))_{k\in\mathbb{N}}$ satisfies (9), let $\theta^* \in \Theta$ where Θ is a compact subset of $\Theta(2)$ defined in (1). Let $(X_1, \ldots, X_n)$ be an observed trajectory of (X_t), which is a stationary long memory LARCH(∞) solution of (2). Then, under Assumptions $C_\ell(\Theta)$, Id(Θ) and (S), $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta^*$ where $\widehat{\theta}_n$ is defined in (12).

Corollary 3.4 (**Example of long memory LARCH(∞) studied in Beran and Schützner, 2009)**. Suppose that $(\xi_t)_{t\in\mathbb{Z}}$ is a sequence of symmetric centered independent random variables, such as $\|\xi_0\|_1 = 1$ and $\|\xi_0\|_2 < \infty$. Consider $\theta = {}^t(a_0, c, d)$ and the sequence $(a_j(\theta))_{j\in\mathbb{N}}$ with $a_0(\theta) = a_0$ and $a_j(\theta) = c\, j^{d-1}$ for $j \geq 1$. Define

$$\Theta = \Big\{ {}^t(a_0, c, d) \in [a_m, a_M] \times [-c_M, c_M] \times [0, d_M],\ \ c_M^2\, \|\xi_0\|_2^2 \sum_{k=1}^{\infty} k^{2\,d_M - 2} < 1 \Big\}, \tag{18}$$

where $0 < a_m \leq a_M < \infty$, $0 \leq d_M < 1/2$ and $0 \leq c_M < \infty$. Let $(X_1, \ldots, X_n)$ be an observed trajectory of $(X_t)_{t\in\mathbb{Z}}$ which is a stationary long-memory LARCH(∞) solution of (2) with parameter $\theta^* \in \Theta$. Then under Assumption (S), $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta^*$ where $\widehat{\theta}_n$ is defined in (12).

Remark 3.3. *The condition $\|\xi_0\|_2 < \infty$ required in Proposition 3.1 and Corollaries 3.1, 3.2, 3.3, and 3.4 can be compared with the conditions required in other works dealing with parametric estimation of LARCH processes. In Theorem 4.2. of Francq and Zakoïan (2010), θ^* is estimated using a weighted LS estimator (see its definition in* (25)*) and the consistency of this LS estimator is established under the condition $\|\xi_0\|_4 < \infty$, except for appropriate weights for which $\|\xi_0\|_1 < \infty$ is sufficient. In Theorem 4 of Beran and Schützner (2009), θ^* is estimated using a smoothed QML estimator (see its definition in* (24)*) and the condition $\|\xi_0\|_3 < \infty$ is required for $\mathbb{L}^1$ consistency of this QML estimator. Finally, in Truquet (2014), the strong consistency of a smoothed QML estimator for LARCH(p) processes is obtained under the condition $\|\xi_0\|_s < \infty$ with $s > 0$.*

3.2 Asymptotic normality of the estimator

Under Assumption $A(4)$, if $\theta^* \in \overset{\circ}{\Theta}$, the interior of Θ where Θ is a compact subset of $\Theta(4)$, and under Assumption (S) and $C_2(\Theta)$, define, if they exist, the following matrices:

$$\Gamma_1^* := \mathbb{E}\Big[\partial_\theta M_{\theta^*}(0)\,{}^t\partial_\theta M_{\theta^*}(0)\Big] = \partial_\theta a_0(\theta^*)\,{}^t\partial_\theta a_0(\theta^*) + \sigma_X^2 \sum_{k=1}^{\infty} \partial_\theta a_k(\theta^*)\,{}^t\partial_\theta a_k(\theta^*)$$

$$\text{with } \sigma_X^2 := \mathbb{E}\big[X_0^2\big] = \frac{a_0^2(\theta^*)\,\sigma_\xi^2}{1-\sigma_\xi^2\sum_{k=1}^{\infty} a_k^2(\theta^*)} \tag{19}$$

$$\text{and} \quad \Gamma_2^* := \mathbb{E}\Big[\big(M_{\theta^*}(0)\big)^2\,\partial_\theta M_{\theta^*}(0) \times {}^t\partial_\theta M_{\theta^*}(0)\Big]. \tag{20}$$

Then the asymptotic normality of $\widehat{\theta}_n$ can be established:

Theorem 3.1. Under Assumption $A(4)$, if $\theta^* \in \overset{\circ}{\Theta}$, is an unknown parameter, where Θ is a compact subset of $\Theta(4) \subset \mathbb{R}^\ell$, which is defined in (4), consider $(X_t)_{t\in\mathbb{Z}}$ as a stationary LARCH(∞) solution of (2) and $(X_1,\dots,X_n)$ an observed trajectory of (X_t). Assume that Assumption (S) is satisfied as well as Assumption $C_{\ell+2}(\Theta)$ and $\mathrm{Id}(\Theta)$. Then, if the matrices Γ_1^* and Γ_2^* defined in (19) and (20) are positive definite,

$$\sqrt{n}\,\big(\widehat{\theta}_n - \theta^*\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\Big(0\,,\,(\sigma_\xi^2-1)\,\big(\Gamma_1^*\big)^{-1}\Gamma_2^*\,\big(\Gamma_1^*\big)^{-1}\Big) \tag{21}$$

Remark 3.4. *The expression of Γ_2^* is not easy to simplify even in the simplest cases. This is not really a problem since, as usual, one can use the Slutsky Lemma, after defining the following estimators of σ_ξ^2, Γ_1^* and Γ_2^* by*

$$\widehat{\sigma}_\xi^2 := \frac{1}{n}\sum_{t=1}^{n} \frac{X_t^2}{\Big(a_0(\widehat{\theta}_n) + \sum_{k=1}^{t-1} a_k(\widehat{\theta}_n)\,X_{t-k}\Big)^2};$$

$$\widehat{\Gamma}_1 := \frac{1}{n}\sum_{t=1}^{n} \Big(\partial_\theta a_0(\widehat{\theta}_n) + \sum_{k=1}^{t-1}\partial_\theta a_k(\widehat{\theta}_n)\,X_{t-k}\Big)\,{}^t\Big(\partial_\theta a_0(\widehat{\theta}_n) + \sum_{k=1}^{t-1}\partial_\theta a_k(\widehat{\theta}_n)\,X_{t-k}\Big);$$

$$\widehat{\Gamma}_2 := \frac{1}{n}\sum_{t=1}^{n} \Big(a_0(\widehat{\theta}_n) + \sum_{k=1}^{t-1} a_k(\widehat{\theta}_n)\,X_{t-k}\Big)^2\Big(\partial_\theta a_0(\widehat{\theta}_n) + \sum_{k=1}^{t-1}\partial_\theta a_k(\widehat{\theta}_n)\,X_{t-k}\Big) \times$$
$$\times\,{}^t\Big(\partial_\theta a_0(\widehat{\theta}_n) + \sum_{k=1}^{t-1}\partial_\theta a_k(\widehat{\theta}_n)\,X_{t-k}\Big).$$

Note that the consistency of $\widehat{\sigma}_\xi^2$ was proved in Francq and Zakoïan (2010). The matrix $\widehat{\Gamma}_1$ and $\widehat{\Gamma}_2$ are also consistent estimators of Γ_1^* and Γ_2^* (see the proof in section 5), and therefore

$$\sqrt{n}\,\big(\widehat{\sigma}_\xi^2-1\big)^{-1/2}\big(\widehat{\Gamma}_1\big)^{1/2}\big(\widehat{\Gamma}_2\big)^{-1/2}\,\big(\widehat{\Gamma}_1\big)^{1/2}\big(\widehat{\theta}_n-\theta^*\big) \xrightarrow[n\to\infty]{\mathcal{L}} \mathcal{N}\big(0\,,\,I_d\big). \tag{22}$$

Such a central limit theorem (22)*, satisfied by $\widehat{\theta}_n$ allows the computation of asymptotic confidence intervals or test thresholds on θ.*

The special case of GLARCH(p,q) processes can be considered under simplified assumptions:

Corollary 3.5. Assume the conditions of Corollary 3.2, and also assume $\|\xi_0\|_4 < \infty$ Let Θ be a compact set included in $\Theta_{p,q}(4)$ and $\theta^* = {}^t\big(c_0^*, \ldots, c_p^*, d_1^*, \ldots, d_q^*\big) \in \overset{\circ}{\Theta}$. Then, with $\widehat{\theta}_n = {}^t\big(\widehat{c}_0^{(n)}, \ldots, \widehat{c}_p^{(n)}, \widehat{d}_1^{(n)}, \ldots, \widehat{d}_q^{(n)}\big)$, the central limit theorems (21) and (22) hold.

In the case of LARCH(p) processes, which are particular cases of GLARCH(p, q) processes, we can go further into the details of the conditions for asymptotic normality:

Corollary 3.6. Assume the conditions of Corollary 3.1, and suppose also $\|\xi_0\|_4 < \infty$, Θ defined by

$$\Theta = \Big\{ {}^t(a_0, a_1, \ldots, a_p) \in [\underline{a}, \overline{a}] \times \mathbb{R}^p \ \text{ with } \ \|\xi_0\|_4^4 \sum_{j=1}^{p} a_j^4 + 6\,\|\xi_0\|_2^2 \sum_{j=1}^{p} a_j^2 \leq r \Big\}, \tag{23}$$

where $0 < \underline{a} < \overline{a}$ and $0 < r < 1$, and $\theta^* = {}^t(a_0^*, a_1^*, \ldots, a_p^*) \in \overset{\circ}{\Theta}$. Then, with $\widehat{\theta}_n = {}^t\big(\widehat{a}_0^{(n)}, \widehat{a}_1^{(n)}, \ldots, \widehat{a}_p^{(n)}\big)$, the central limit theorems (21) and (22) hold.

As an example of computation of the asymptotic covariance, if we consider the case of a LARCH(1) process, we obtain:

$$\Gamma_1^* = \begin{pmatrix} 1 & 0 \\ 0 & \sigma_X^2 \end{pmatrix} \quad \text{and} \quad \Gamma_2^* = \begin{pmatrix} a_0^2 + \sigma_X^2 & 2\,a_0 a_1 \sigma_X^2 \\ 2\,a_0 a_1 \sigma_X^2 & a_0^2 \sigma_X^2 + \mathbb{E}[X_0^4] \end{pmatrix}.$$

This implies

$$\big(\Gamma_1^*\big)^{-1} \Gamma_2^* \big(\Gamma_1^*\big)^{-1} = \begin{pmatrix} a_0^{*2} + \sigma_X^2 & 2\,a_0^* a_1^* \\ 2\,a_0^* a_1^* & \frac{a_0^{*2}}{\sigma_X^2} + \frac{\mathbb{E}[X_0^4]}{\sigma_X^4} \end{pmatrix},$$

where $\left\{ \begin{array}{rcl} \sigma_X^2 & = & a_0^{*2}\,\sigma_\xi^2 \big(1 - \sigma_\xi^2 a_1^{*2}\big)^{-1} \\ \mathbb{E}[X_0^4] & = & a_0^{*4}\,\mathbb{E}[\xi_0^4]\,\big(1 + 5\,\sigma_\xi^2 a_1^{*2}\big)\big(1 - \sigma_\xi^2 a_1^{*2}\big)^{-1}\big(1 - \mathbb{E}[\xi_0^4] a_1^{*4}\big)^{-1} \end{array} \right.$.

The asymptotic normality of the estimator $\widehat{\theta}_n$ can also be obtained in the case of long-memory LARCH(∞):

Corollary 3.7. Assume the conditions of Corollary 3.3, and further assume that $\|\xi_0\|_4 < \infty$, Θ is a compact set included in $\Theta(4) \subset \mathbb{R}^\ell$ and $\theta^* \in \overset{\circ}{\Theta}$. Then, under Assumption $C_{\max(\ell,2)}(\Theta)$. and if Γ_1^* and Γ_2^* are positive definite matrices, the central limit theorems (21) and (22) hold.

Corollary 3.8 (**Example of long memory LARCH(∞) studied in Beran and Schützner, 2009)**. Under the assumptions of Corollary 3.4 and if $\overline{c}^4 \|\xi_0\|_4^4 \sum_{i=1}^{\infty} j^{4\overline{d}-4} + 6\overline{c}^2 \|\xi_0\|_2^2 \sum_{i=1}^{\infty} j^{2\overline{d}-2} < 1$, then the central limit theorems (21) and (22) hold.

Remark 3.5. *To our knowledge, the only result for estimating the memory parameter d in the case of LARCH(∞) processes with long memory was obtained in Beran and Schützner (2009) using the smoothed QML estimator. However, the expression of this estimator uses only a small part of the sample $(X_1, \ldots, X_n)$, namely the last n^β observations, where $\beta < 1 - 2d$, while d is unknown to account for the long memory of the process. This leads to a convergence rate of $n^{\beta/2}$ for this truncated QML estimator, which is far less interesting than the convergence rate of $\sqrt{n}$ obtained with $\widehat{\theta}_n$. Monte Carlo experiments (see section 4) will also show that $\widehat{\theta}_n$ performs numerically better than these other estimators in terms of convergence rate, especially when compared to estimators based on QML.*

4 Numerical experiments

4.1 Monte Carlo experiments

In this section, we report the results of Monte Carlo experiments conducted with different LARCH processes. More specifically, we considered:

- Three different LARCH processes:
  1. A LARCH(2) process, with parameters $a_0 = 5$, $a_1 = -0.2$ and $a_2 = 0.4$;
  2. A GLARCH(1, 1) process, with parameters $c_0 = 2$, $c_1 = 0.3$ and $d_1 = -0.6$;
  3. A long memory LARCH(∞) process, with $\theta = {}^t(a_0, c, d)$ and $a_0(\theta) = a_0$ and $a_k(\theta) = c\,k^{d-1}$. We choose $a_0 = 1$, $c = 0.2$ and $d = 0.1,\ 0.2,\ 0.3$ and 0.4, using the same example studied in Beran and Schützner (2009) for its numerical illustrations. To define numerically the trajectory of such processes, whose theoretical definition involves an infinite sum, a truncation of this infinite sum has been used by taking a sum of one million terms in the computation of $M_\theta(t)$, which also requires the generation of one million additional realizations of the white noise.
- Several trajectory lengths: $n = 200,\ 500,\ 1000,\ 2000$ and 5000 for LARCH(2) and GLARCH(1, 1) processes, and $n = 1000,\ 2500,\ 5000$ and 10000 for the LARCH(∞) process (as in Beran and Schützner, 2009);
- Two distributions for ξ_0 such as $\mathbb{E}[|\xi_0|] = 1$: a Gaussian $\mathcal{N}(0, \pi/2)$ distribution denoted $\mathcal{N}$ and a normalized Student $t(6)$ distribution with 6 freedom degrees.
- Choice of Θ for minimisation (12) and calculation of $\widehat{\theta}_n$: as already mentioned in the Remark 2.3, the set $\Theta(2)$ depends on the distribution of the noise, and an "extended" condition will be used in the minimisation algorithm. Hence, for the LARCH(2) process this means $0 \leq a_1,\, a_2 \leq 1$, for the GLARCH(1, 1) process it will be $0 \leq c_1,\, d_1 \leq 1$ and for the long memory LARCH(∞) process, $c < 1$ and $0 \leq d \leq 0.5$. After computing the estimator, and using the empirical variance of the residuals $\widehat{\xi}_t$ as an estimator of $\|\xi_0\|_2^2$, the condition $\widehat{\theta}_n \in \Theta(2)$ can be checked from the equation $\widehat{\sigma}^2_{\widehat{\xi}_t} \sum_{j=1}^{\infty} a_j^2(\widehat{\theta}_n) < 1$, which means for example that $\widehat{\sigma}^2_{\widehat{\xi}_t}\, \widehat{c}^{\,2}\, \zeta(2 - 2\,\widehat{d}) < 1$ for the long memory LARCH(∞) process, where ζ is the Riemann zeta function.

For each choice of process, length n and noise distribution, 1000 replications of independent trajectories of the LARCH process are generated, except for the long memory LARCH(∞) where only 300 replications are used due to the computational time of the estimator. Note that in this case, an algorithm based on the Fast Fourier Transform (FFT) has been developed in Nielsen and Noël (2021, section 2.2), which could have been used to speed up the computations significantly, or if we had applied the estimators to series larger than 10^4.

Two other estimators are to be compared with $\widehat{\theta}_n$:

1. Following Beran and Schützner (2009) and Truquet (2014), the first is the smooth approximation of the QMLE, which for $h > 0$ is given by

$$\widehat{\theta}_{QML}(h) := \underset{\theta\in\Theta}{\text{Argmin}}\ \frac{1}{n}\sum_{t=1}^{n}\frac{h+X_t^2}{h+(M_\theta(t))^2}+\log\big(h+(M_\theta(t))^2\big). \tag{24}$$

   The a priori choice of h or data-driven $\widehat{h}$ is not a straightforward task, although Truquet has provided some guidance in Truquet (2014). Therefore, we will present the results obtained for 2 different values of h that give the best performances. In the case of the considered LARCH(∞) process with long memory, Beran and Schützner (2009) proposed a modified version of $\widehat{\theta}_{QML}(h)$ and we will use their results.

2. Following Francq and Zakoïan (2010), the second is the weighted least squares estimator defined by:

$$\widehat{\theta}_{FZ} := \underset{\theta\in\Theta}{\text{Argmin}}\ \frac{1}{n}\sum_{t=1}^{n}\tau_t\left(X_t^2-(M_\theta(t))^2\right)^2, \tag{25}$$

   where the weights (τ_t) are obtained for LARCH(p) or GLARCH(p,q) using an empirical rule proposed in Ling (2007): $\tau_t = \Big(\max\big(1\,,\,\frac{1}{C}\sum_{i=1}^{p}|X_{t-i}|\,\mathbb{1}_{|X_{t-i}|>C}\big)\Big)^{-4}$, where C is computed as the 90% quantile of the absolute values $\big(|X_1|,\ldots,|X_n|\big)$. In case of long memory LARCH(∞), we replace p by $t-1$ in the definition of τ_t.

Remark 4.1. *The estimator $\widehat{\theta}_n$ can also be defined as a nonlinear least squares estimator for the regression $(|X_t|)_{1\le t\le n}$ on $\Big(\Big|a_0(\theta)+\sum_{j=1}^{\infty}a_j(\theta)\,X_{t-j}\Big|\Big)_{1\le t\le n}$. As it was already done by Francq and Zakoïan (2010) (see the estimator defined in (25)), a weighted version of $\widehat{\theta}_n$ could also be considered:*

$$\widehat{\theta}_n^w = \underset{\theta\in\Theta}{Arg\min}\ \frac{1}{n}\sum_{t=1}^{n}w_t\left(|X_t|-\Big|a_0(\theta)+\sum_{j=1}^{\infty}a_j(\theta)\,\widetilde{X}_{t-j}\Big|\right)^2,$$

where $w_t = W\big((X_{t-k})_{1\le k\le t-1}\big) > 0$, defining a sequence of selected weights. Inspired by their work, it seems that using weights of the form $w_t = \Big(\max\big(1\,,\,\frac{1}{C}\sum_{i=1}^{p}|X_{t-i}|\,\mathbb{1}_{|X_{t-i}|>C}\big)^{-2}$ would allow, in particular, remove the condition $r=4$ from the assumptions of asymptotic normality of $\widehat{\theta}_n$ by requiring only $r=1$ for $\widehat{\theta}_n^w$. To improve the convergence rate of the $\widehat{\theta}_n$, taking into account Remark 4.2 of Francq and Zakoïan (2010), one might think that $w_t = \big(M_{\theta^*}(t)\big)^{-2}$ would be the ideal sequence of weights, but infeasible in practice.

Remark 4.2. *The estimator $\widehat{\theta}_n$ is obtained with parameters defined under the normalization condition $\|\xi_0\|_1 = 1$, while the estimators $\widehat{\theta}_{QML}(h)$ or $\widehat{\theta}_{FZ}$ are basically defined under the normalization condition $\|\xi_0\|_2 = 1$. After remark 2.2, a suitable renormalization of the parameters of the LARCH process allows the transition from one condition to the other. So, for reference, consider the LARCH(p) equation with parameter $\theta = {}^t(a_0,\ldots,a_p)$ under condition $\|\xi_0\|_1 = 1$, $\widehat{\theta}_n$ is an estimator of θ, but $\widehat{\theta}_{QML}(h)$ or $\widehat{\theta}_{FZ}$ are estimators of $\theta/\|\xi_0\|_2$. Therefore, a comparison of the accuracies of the estimators is possible by considering $\widehat{\theta}_n$, $\|\xi_0\|_2\,\widehat{\theta}_{QML}(h)$ and $\|\xi_0\|_2\,\widehat{\theta}_{FZ}$. If the law of ξ_0 is unknown, the comparison with the estimator $\widehat{\sigma}_\xi$ defined in (22) is still possible.*

		$\widehat{\theta}_n$			$\widehat{\theta}_{FZ}$			$\widehat{\theta}_{QML}(2)$			$\widehat{\theta}_{QML}(1)$		
ξ_0 law	n	a_0	a_1	a_2	a_0	a_1	a_2	a_0	a_1	a_2	a_0	a_1	a_2
$\mathcal{N}$	200	0.326	0.047	0.064	0.423	0.092	0.100	1.015	0.136	0.123	1.832	0.248	0.168
	500	0.210	0.029	0.043	0.268	0.059	0.065	0.446	0.070	0.086	1.119	0.145	0.121
	1000	0.145	0.021	0.030	0.188	0.044	0.047	0.382	0.060	0.082	0.582	0.075	0.084
	2000	0.101	0.014	0.021	0.130	0.030	0.033	0.265	0.047	0.059	0.453	0.053	0.080
	5000	0.065	0.009	0.013	0.083	0.019	0.021	0.205	0.031	0.048	0.326	0.037	0.061
$t(6)$	200	0.433	0.061	0.091	0.624	0.130	0.133	2.111	0.268	0.205	1.916	0.258	0.186
	500	0.272	0.040	0.061	0.390	0.091	0.094	1.699	0.214	0.186	1.569	0.209	0.167
	1000	0.224	0.029	0.051	0.275	0.067	0.071	1.361	0.170	0.191	1.270	0.166	0.155
	2000	0.124	0.021	0.031	0.196	0.048	0.051	1.334	0.149	0.175	1.535	0.147	0.151
	5000	0.077	0.014	0.021	0.127	0.031	0.033	1.230	0.136	0.211	1.533	0.147	0.188

Table 1: Square roots of the MSE computed for each estimator of parameters $a_0 = 5$, $a_1 = -0.2$ and $a_2 = 0.4$ of a LARCH(2) process computed from 1000 independent replications.

The results are presented in tables 1, 2 and 3.

Conclusions of the Monte Carlo experiments:

- Looking at the decay of the square root of the MSE of the $\widehat{\theta}_n$ components towards 0 as n increases from 200 to 5000 (or from 1000 to 10000), we see that this decay is approximately $1/\sqrt{n}$, which corresponds to the theoretical convergence rate established in Theorem 3.1. This is true for a LARCH(2) process as well as for a GLARCH(1, 1) or a LARCH(∞) process with long memory, regardless of whether we consider a white noise with a Gaussian or Student distribution $t(6)$.

- In general, the square root of the MSE of $\widehat{\theta}_n$ converges to 0 twice as fast as that of $\widehat{\theta}_{FZ}$ when the white noise follows a Gaussian or Student $t(6)$ distribution, for the three processes considered. We note that $\widehat{\theta}_{FZ}$ gives convincing results in the case of a LARCH(∞) process with long memory, while this was not shown in Francq and Zakoïan (2010), although they are far outperformed by those of $\widehat{\theta}_n$. However, the numerical results for the convergence of the MSE of $\widehat{\theta}_{FZ}$ towards 0 become much worse when the white noise distribution of the LARCH processes is a Student $t(6)$ distribution.

- Finally, $\widehat{\theta}_{QML}$ gives satisfactory performances comparable to those of $\widehat{\theta}_n$ only in one case, namely for the LARCH(1, 1) process with Gaussian white noise, and this after choosing an optimal regularisation parameter. It should be noted, however, that the choice of this parameter can easily be made in the context of Monte Carlo experiments, but this would otherwise require a data-driven procedure that does not currently exist. For the estimation

		$\widehat{\theta}_n$			$\widehat{\theta}_{FZ}$			$\widehat{\theta}_{QML}(1)$			$\widehat{\theta}_{QML}(0.5)$		
ξ_0 law	n	c_0	c_1	b_1	c_0	c_1	b_1	c_0	c_1	b_1	c_0	c_1	b_1
$\mathcal{N}$	200	0.172	0.044	0.096	0.238	0.094	0.135	0.179	0.045	0.099	0.190	0.052	0.105
	500	0.108	0.028	0.057	0.158	0.068	0.081	0.102	0.031	0.055	0.114	0.040	0.061
	1000	0.071	0.019	0.039	0.113	0.050	0.055	0.066	0.018	0.034	0.081	0.029	0.048
	2000	0.052	0.013	0.028	0.087	0.040	0.039	0.045	0.012	0.023	0.050	0.019	0.024
	5000	0.033	0.008	0.017	0.065	0.030	0.025	0.028	0.007	0.014	0.027	0.006	0.012
$t(6)$	200	0.233	0.061	0.145	0.367	0.113	0.232	0.335	0.085	0.191	0.349	0.094	0.181
	500	0.142	0.042	0.081	0.196	0.077	0.116	0.255	0.072	0.156	0.275	0.095	0.156
	1000	0.091	0.029	0.051	0.147	0.063	0.082	0.207	0.064	0.125	0.261	0.096	0.144
	2000	0.064	0.020	0.033	0.096	0.043	0.052	0.153	0.059	0.097	0.214	0.092	0.119
	5000	0.039	0.013	0.022	0.071	0.029	0.037	0.116	0.036	0.095	0.184	0.073	0.119

Table 2: Square roots of the MSE computed for each estimator of parameters $c_0 = 2$, $c_1 = 0.3$ and $d_1 = -0.6$ of a GLARCH$(1, 1)$ process computed from 1000 independent replications.

		$\widehat{\theta}_n$			$\widehat{\theta}_{FZ}$			$\widehat{d}_{QML}$
d	n	a_0	c	d	a_0	c	d	d
$d = 0.1$	1000	0.035	0.024	0.089	0.092	0.054	0.160	0.357
	2500	0.020	0.015	0.048	0.089	0.059	0.119	0.292
	5000	0.016	0.010	0.036	0.040	0.031	0.084	0.217
	10000	0.013	0.010	0.021	0.033	0.029	0.053	0.198
$d = 0.2$	1000	0.041	0.023	0.060	0.103	0.059	0.147	1.449
	2500	0.028	0.017	0.043	0.052	0.033	0.088	0.733
	5000	0.016	0.010	0.024	0.033	0.027	0.050	0.559
	10000	0.014	0.008	0.017	0.032	0.024	0.045	0.257
$d = 0.3$	1000	0.060	0.024	0.053	0.119	0.054	0.110	-
	2500	0.045	0.017	0.037	0.091	0.035	0.087	-
	5000	0.026	0.011	0.022	0.057	0.029	0.050	-
	10000	0.016	0.008	0.017	0.039	0.024	0.042	-
$d = 0.4$	1000	0.100	0.025	0.047	0.189	0.051	0.081	-
	2500	0.071	0.017	0.030	0.158	0.041	0.074	-
	5000	0.041	0.011	0.018	0.122	0.032	0.057	-
	10000	0.034	0.009	0.015	0.067	0.024	0.036	-

Table 3: Square roots of the MSE computed for estimators $\widehat{\theta}_n$ and $\widehat{\theta}_{FZ}$ of parameters $a_0 = 1$, $c = 0.2$ and $d = 0.1$, 0.2, 0.3 and 0.4 of the LARCH(∞) process computed from 300 independent replications, and for $\widehat{d}_{QML}(h)$ already computed in Beran and Schützner (2009) for $d = 0.1$ and 0.2.

of the long memory parameter d in the case of a LARCH(∞) process, the QML estimator proposed in Beran and Schützner (2009) has truly disastrous performances compared to those obtained with $\widehat{\theta}_{FZ}$ and in particular those of $\widehat{\theta}_n$.

4.2 Application on the returns of a financial index

Let us look at the financial returns of an index called the CAC40, which is the French equivalent of the FTSE100, with 40 companies instead of 100. These are the closing values between 15 November 2002 and 15 November 2022, i.e. 20 years and $n = 5122$ data. The correlogram plot shows a weak white-noise type behaviour, which is also consistent with a conditionally heteroscedastic process. Firstly, we can see that these returns have values extremely close to 0 (70 are in absolute terms less

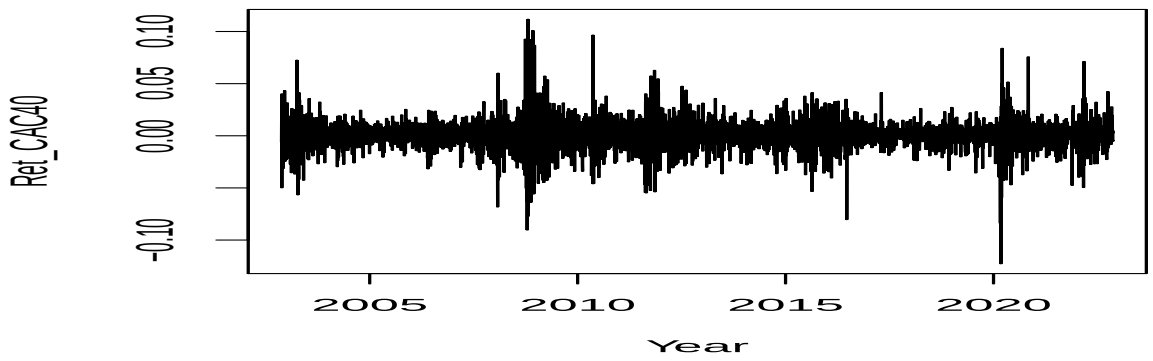

Figure 1: Financial returns of CAC40 index, between November 15, 2002 and November 15, 2022

than 1% of the standard deviation) and even twice exactly 0, which would be very unlikely with a GARCH process often used to model this data, for which the conditional variance is always greater than a positive constant.

Moreover, as already pointed out in Giraitis *i.e.* (2004) with the S&P500 returns data since 1928, we can observe a particular behaviour of the leverage estimate: it is negative for almost all lags and follows a power-law type distribution (see Figure 2). A non-linear least squares approximation gives the value of this power $\simeq -0.55$. If we note h_t the leverage, this would mean that $h_t \sim C\, t^{d-1}$ with $d \simeq 0.45$ and $C < 0$. Now, if the CAC40 returns follow the example of LARCH(∞) processes with long memory (X_t) studied in Beran and Schützner (2009), *i.e.* $a_j(\theta) = c\, j^{d-1}$ for $j \geq 1$, then we have:

$$\mathrm{Cov}\,(X_0, X_t) = 0, \quad h_t = \mathrm{Cov}\,(X_0, X_t^2) \simeq C\, t^{d-1} \quad \text{and} \quad \mathrm{Cov}\,(X_0^2, X_t^2) \simeq c'\, t^{2d-1},$$

with $0 < d < 1/2$, $C < 0$ and $c' > 0$ (see again Giraitis *i.e.*, 2004, or Robinson and Zaffaroni, 1997). We check that this last asymptotic behaviour is well verified by plotting the correlogram of the squares of the process, which is done in figure 3. Again we observe a power law type behaviour, and a non-linear least squares estimation of this power gives the result $\simeq -0.09$. However, if we use the

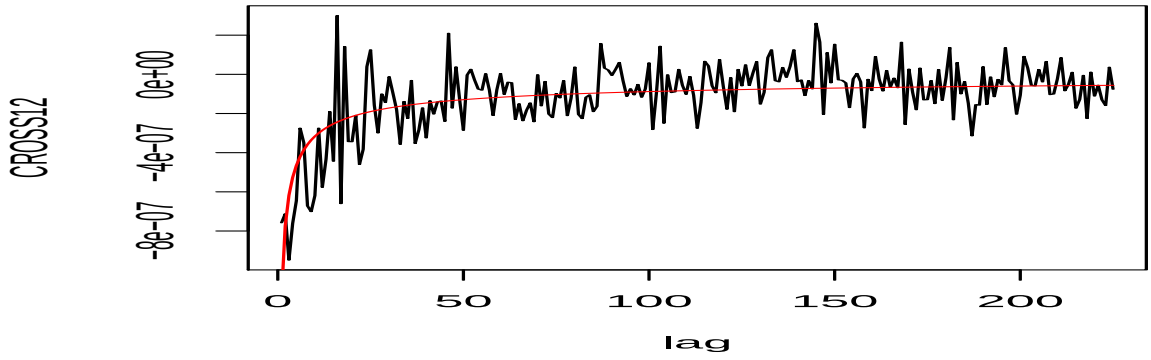

Figure 2: Leverage estimate of the financial returns of CAC40 index, between November 15, 2002 and November 15, 2022

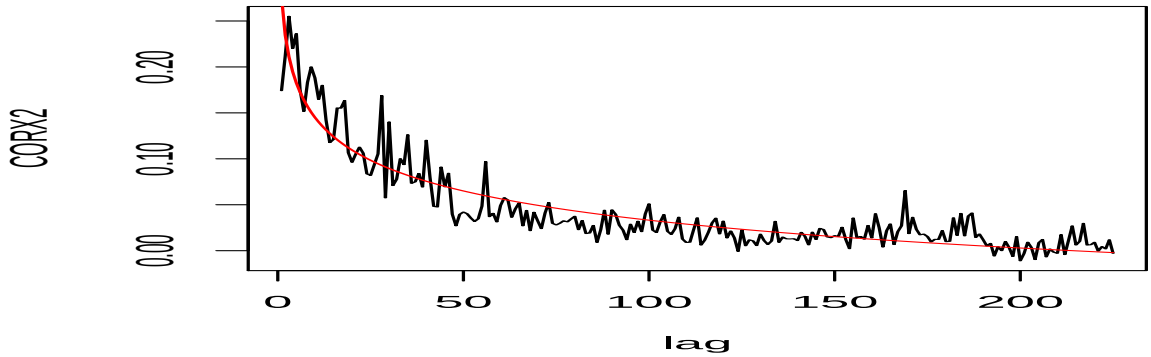

Figure 3: Correlogram of the squared financial returns of CAC40 index, between November 15, 2002 and November 15, 2022

value of $d = 0.445$ obtained numerically from h_t, we find that $2d - 1 \simeq -0.11$, a value very close to -0.09. There seems to be a good fit of a long memory LARCH(∞) model with these return data.

Fitting the returns of CAC40 index with two models of long memory LARCH(∞)

a. We first fitted these data with the long memory LARCH(∞) process studied in Beran and Schützner (2009), *i.e.*, such that

$$a_0(\theta) = a_0 \quad \text{and} \quad a_j(\theta) = c\, j^{d-1} \quad \text{for } j \geq 1.$$

For this model, we considered the estimator $\widehat{\theta}_n = {}^t\big(\widehat{a}_0, \widehat{c}, \widehat{d}\big)$ used in the results of Table 3 and we obtain

$$\widehat{a}_0 \simeq 0.010 \quad , \ \widehat{c} \simeq -0.159 \quad \text{and} \quad \widehat{d} \simeq 0.488.$$

This is quite consistent with the values of d previously obtained by nonlinear least squares. Note that we also obtained $\widehat{\theta}_{FZ} = {}^t\big(0.007, -0.143, 0.497\big)$, a value quite close to that of $\widehat{\theta}_n$, confirming

the long memory property of this time series. Note also that using (22) an estimate of the covariance matrix of these estimators can be computed. And we obtained the following 95% confidence intervals from $\widehat{\theta}_n$:

$$a_0 \in \big[0.00999\,,\,0.0110\big], \quad c \in \big[-0.192\,,\,-0.125\big] \quad \text{and} \quad d \in \big[0.433\,,\,0.544\big].$$

However, we notice that $\sum_{j=1}^{\infty} a_j^2(\widehat{\theta}_n) \simeq (0.159)^2\,\zeta(1.024) \simeq 1.068 > 1$ and also $\sum_{j=1}^{\infty} a_j^2(\widehat{\theta}_{FZ}) \simeq (-0.143)^2\,\zeta(1.006) > 1$. Thus, whatever the distribution of the noise, the second order stationarity condition is not verified. The long memory model LARCH(∞) process such as $a_j(\theta) = c\,j^{d-1}$ for $j \geq 1$ does not seem appropriate. This is also confirmed by the correlograms of the residuals $\widehat{\xi}_t = X_t/(\widehat{a}_0 + \widehat{c}\,\sum_{j=1}^{t-1} j^{\widehat{d}-1} X_{t-j})$ and their absolute values $|\widehat{\xi}_t|$ plotted in Figure 4. Indeed, and it is particularly clear with the correlogram of $|\widehat{\xi}_t|$, $(\widehat{\xi}_t)$ could not appear almost as white noise (note that Francq and Zakoïan, 2010, also proposed a goodness-of-fit test based on $(\widehat{\xi}_t^2)$, but only for AR(q)-LARCH(p) processes). Modelling these data with this model is therefore problematic.

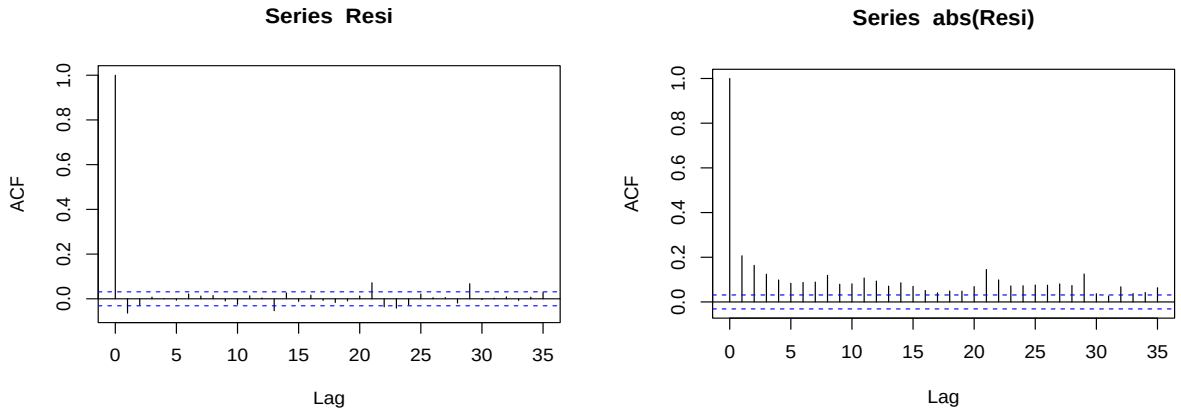

Figure 4: Correlograms of residuals and absolute values of residuals for the returns of CAC40 index, between November 15, 2002 and November 15, 2022, using $a_j(\theta) = c\,j^{d-1}$ for $j \geq 1$

b. We now consider another model of a long memory LARCH(∞) process. We will thus suppose that

$$a_0(\theta) = a_0 \quad \text{and} \quad a_j(\theta) = c\,j^{d-1}\big(1 + \frac{c'}{j}\big) \quad \text{for } j \geq 1,$$

with $\theta = (a_0, c, d, c') \in (0,\infty) \times \mathbb{R} \times (0, 1/2) \times \mathbb{R}$. It is clear that for $c' = 0$ we recover the previous model. For this model, we computed on data the estimator $\widehat{\theta}_n = {}^t\big(\widehat{a}_0, \widehat{c}, \widehat{d}, \widehat{c}'\big)$ and obtained

$$\widehat{a}_0 \simeq 0.010 \quad , \ \widehat{c} \simeq -0.302, \ \widehat{d} \simeq 0.346 \quad \text{and} \quad \widehat{c}' \simeq -0.678.$$

In contrast to the previous model, now $\sum_{j=1}^{\infty} a_j^2(\widehat{\theta}_n) \simeq (-0.302)^2\,\big(\zeta(1.308) + (0.678)^2\,\zeta(3.308) - 1.356\,\zeta(2.308)\big) \simeq 0.222 < 1$, and the second order stationary condition is generally satisfied (except when $\|\xi_0\|_2 \geq 4.5$). And using (22), an estimate of the covariance matrix of these estimators can be computed, and the following 95% confidence intervals of the parameters are obtained:

$$a_0 \in \big[0.0099\,,\,0.0110\big], \quad c \in \big[-0.383\,,\,-0.221\big], \quad d \in \big[0.275\,,\,0.416\big] \quad \text{and} \quad c' \in \big[-0.885\,,\,-0.472\big].$$

Finally, to visualise the goodness-of-fit of this second long memory LARCH(∞) model, the correlograms of the residuals and the absolute values of the residuals are plotted in Figure 5.

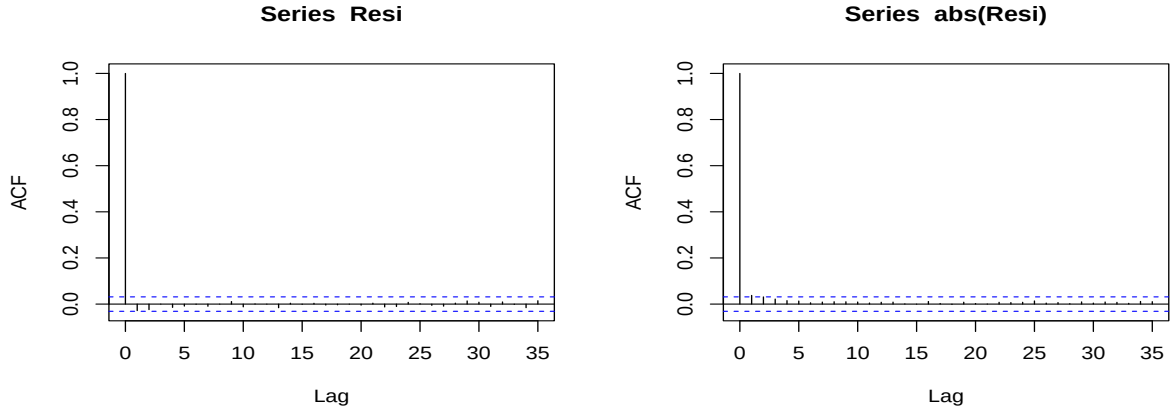

Figure 5: Correlograms of residuals and absolute values of residuals for the returns of CAC40 index, between November 15, 2002 and November 15, 2022, using $a_j(\theta) = c\, j^{d-1}\big(1 + \frac{c'}{j}\big)$ for $j \geq 1$

Conclusion: For these financial data, as observed by other authors, modelling by a long memory LARCH(∞) process seems relevant. However, we point out that the model studied by Beran and Schützner (2009), *i.e.* $a_j(\theta) = c\, j^{d-1}$ for $j \geq 1$, is not appropriate for these data, while a model with a slightly more complex behaviour with an additional parameter, *i.e.* $a_j(\theta) = c\, j^{d-1}\big(1 + \frac{c'}{j}\big)$ for $j \geq 1$, provides a quite satisfactory fit.

5 Proofs

In the sequel, with $\sum_{j=1}^{m} \cdot = 0$ for $m \leq 0$ by convention, for $t \in \mathbb{Z}$, we will many times consider

$$M_\theta(t) := a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta)\, X_{t-j} \quad \text{and} \quad \widetilde{M}_\theta(t) := a_0(\theta) + \sum_{j=1}^{t-1} a_j(\theta)\, X_{t-j}. \tag{26}$$

Lemma 5.1. Under the assumptions of Proposition 3.1 and with $\overline{d} < 1/2$ defined in (13), we obtain:

1. There exists $\widetilde{C} > 0$, such as for any $t \geq 1$ and $\theta \in \Theta$:

$$\mathbb{E}\Big[\big(M_\theta(t)\big)^2\Big] < \infty, \quad \mathbb{E}\Big[\big(\widetilde{M}_\theta(t)\big)^2\Big] < \infty \quad \text{and} \quad \mathbb{E}\Big[\big(M_\theta(t) - \widetilde{M}_\theta(t)\big)^2\Big] \leq \widetilde{C}\, t^{2\overline{d}-1}. \tag{27}$$

2. There exists $\widetilde{C} > 0$, such as for any $t \geq 1$ and $\theta \in \Theta$:

$$\mathbb{E}\Big[\big\|\partial_\theta M_\theta(t)\big\|^2\Big] < \infty, \quad \mathbb{E}\Big[\big\|\partial_\theta \widetilde{M}_\theta(t)\big\|^2\Big] < \infty \quad \text{and} \quad \mathbb{E}\Big[\big\|\partial_\theta M_\theta(t) - \partial_\theta \widetilde{M}_\theta(t)\big\|^2\Big] \leq \widetilde{C}\, t^{2\overline{d}-1}. \tag{28}$$

3. There exists $\overline{C} > 0$, such as for any $t \geq 1$,

$$\mathbb{E}\Big[\sup_{\theta\in\Theta}\big(M_\theta(t) - \widetilde{M}_\theta(t)\big)^2\Big] \leq \overline{C}\, t^{2\overline{d}-1}. \tag{29}$$

Proof of Lemma 5.1. 1. Since $(X_t)_{t\in\mathbb{Z}}$ is a weak white noise, it can be written directly that

$$\mathbb{E}\Big[\big(M_\theta(t)\big)^2\Big] = \mathbb{E}[X_0^2]\sum_{k=1}^{\infty} a_k^2(\theta) \le \mathbb{E}[X_0^2]\sum_{k=1}^{\infty}\sup_{\theta\in\Theta} a_k^2(\theta).$$

From Assumption $C_\ell(\Theta)$, $\sup_{\theta\in\Theta}\big|a_k(\theta)\big| \le C_a\, k^{\overline{d}-1}$ implying

$$\sum_{k=1}^{\infty}\sup_{\theta\in\Theta} a_k^2(\theta) \le C_a^2\sum_{k=1}^{\infty} k^{2\overline{d}-2} < \infty,$$

since $2\overline{d}-2<-1$. And therefore $\mathbb{E}\Big[\big(M_\theta(t)\big)^2\Big] < \infty$.
Using the same reasoning,

$$\mathbb{E}\Big[\big(\widetilde{M}_\theta(t)\big)^2\Big] = \mathbb{E}[X_0^2]\sum_{k=1}^{t-1} a_k^2(\theta) \le \mathbb{E}\Big[\big(M_\theta(t)\big)^2\Big] < \infty.$$

Finally, $\mathbb{E}\Big[\big(M_\theta(t)-\widetilde{M}_\theta(t)\big)^2\Big] = \mathbb{E}[X_0^2]\sum_{k=t}^{\infty} a_k^2(\theta) \le C_a^2\,\mathbb{E}[X_0^2]\sum_{k=t}^{\infty} k^{2\overline{d}-2}$ always from Assumption $C_\ell(\Theta)$. But for f a positive decreasing function, $\sum_{k=n_0}^{n_1} f(k) \le f(n_0) + \int_{n_0}^{n_1} f(t)dt$ for any integer numbers n_0 and n_1 such as $0 \le n_0 < n_1 \le \infty$. Therefore, for any $t \ge 1$,

$$\sum_{k=t}^{\infty} k^{2\overline{d}-2} \le t^{2\overline{d}-2} + \int_t^{\infty} x^{2\overline{d}-2}dx \le t^{2\overline{d}-2} + \frac{1}{1-2\overline{d}}\,t^{2\overline{d}-1} \le \Big(\frac{2-2\overline{d}}{1-2\overline{d}}\Big)\,t^{2\overline{d}-1}.$$

Therefore, for any $\theta\in\Theta$ and $t\ge 1$,

$$\mathbb{E}\Big[\big(M_\theta(t) - \widetilde{M}_\theta(t)\big)^2\Big] \le C_a^2\,\mathbb{E}[X_0^2]\Big(\frac{2-2\overline{d}}{1-2\overline{d}}\Big)\,t^{2\overline{d}-1} \le \widetilde{C}\,t^{2\overline{d}-1}.$$

2. Using Assumptions (S) and $C_\ell(\Theta)$, $\theta\in\Theta \mapsto \partial_\theta M_\theta(t)$ exists a.s., and for any $\theta\in\Theta$ and $t\ge 1$,

$$\partial_\theta M_\theta(t) = \partial_\theta a_0(\theta) + \sum_{k=1}^{\infty}\partial_\theta a_k(\theta)X_{t-k} \quad \text{and} \quad \partial_\theta\widetilde{M}_\theta(t) = \partial_\theta a_0(\theta) + \sum_{k=1}^{t-1}\partial_\theta a_k(\theta)X_{t-k}\ a.s. \tag{30}$$

Then $\mathbb{E}\Big[\big(\partial_\theta M_\theta(t)\big)^2\Big] = \sum_{j=1}^{\ell}\mathbb{E}\Big[\big(\partial_{\theta_j} M_\theta(t)\big)^2\Big]$. By replacing $a_k(\theta)$ by $\partial_{\theta_j}a_k(\theta)$ that also satisfies (13) for $j=1,\dots,\ell$, we can use the same reasoning as in 1. for obtaining $\mathbb{E}\Big[\big(\partial_{\theta_j} M_\theta(t)\big)^2\Big] < \infty$. And this can also be done for $\partial_\theta\widetilde{M}_\theta(t)$ and $\partial_\theta M_\theta(t) - \partial_\theta\widetilde{M}_\theta(t)$ and we obtain (28).

3. We use here ideas already present in Lemmas 1 and 2 of Beran and Schützner (2009). The result will be easily generalized, but to facilitate the writing of the proof, we restrict ourselves to the case where $\ell=2$ and $\theta=(\theta_1,\theta_2)\in\Theta\subset[\underline{\theta}_1,\overline{\theta}_1]\times[\underline{\theta}_2,\overline{\theta}_2]$. Denote $Z_\theta = M_\theta(t) - \widetilde{M}_\theta(t)$. Using Theorem 3.B of Parzen (1999), we can write that for any $t\ge 1$ and $(\theta_1,\theta_2)\in[\underline{\theta}_1,\overline{\theta}_1]\times[\underline{\theta}_2,\overline{\theta}_2]$,

$$Z^2_{(\theta_1,\theta_2)} \le \frac{1}{2}\Big(Z^2_{(\underline{\theta}_1,\theta_2)} + Z^2_{(\overline{\theta}_1,\theta_2)} + \int_{\underline{\theta}_1}^{\overline{\theta}_1} Z^2_{(u_1,\theta_2)} + \big(\partial_{\theta_1} Z_{(u_1,\theta_2)}\big)^2\,du_1\Big).$$

By applying again this theorem, and after computations, we finally obtain for any $(\theta_1,\theta_2)\in[\underline{\theta}_1,\overline{\theta}_1]\times[\underline{\theta}_2,\overline{\theta}_2]$,

$$\begin{aligned}4\,Z^2_{(\theta_1,\theta_2)} \le\ & Z^2_{(\underline{\theta}_1,\underline{\theta}_2)} + Z^2_{(\underline{\theta}_1,\overline{\theta}_2)} + Z^2_{(\overline{\theta}_1,\underline{\theta}_2)} + Z^2_{(\overline{\theta}_1,\overline{\theta}_2)} \\ &+ \int_{\underline{\theta}_2}^{\overline{\theta}_2}\Big(Z^2_{(\underline{\theta}_1,u_2)} + Z^2_{(\overline{\theta}_1,u_2)} + \big(\partial_{\theta_2}Z_{(\underline{\theta}_1,u_2)}\big)^2 + \big(\partial_{\theta_2}Z_{(\overline{\theta}_1,u_2)}\big)^2\Big)\,du_2 \\ &+ \int_{\underline{\theta}_1}^{\overline{\theta}_1}\Big(Z^2_{(u_1,\underline{\theta}_2)} + Z^2_{(u_1,\overline{\theta}_2)} + \big(\partial_{\theta_1}Z_{(u_1,\underline{\theta}_2)}\big)^2 + \big(\partial_{\theta_1}Z_{(u_1,\overline{\theta}_2)}\big)^2\Big)\,du_1 \\ &+ \int_{\underline{\theta}_1}^{\overline{\theta}_1}\int_{\underline{\theta}_2}^{\overline{\theta}_2}\Big(Z^2_{(u_1,u_2)} + \big(\partial_{\theta_1}Z_{(u_1,u_2)}\big)^2 + \big(\partial_{\theta_2}Z_{(u_1,u_2)}\big)^2 + \big(\partial^2_{\theta_1\theta_2}Z_{(u_1,u_2)}\big)^2\Big)\,du_1du_2.\end{aligned}$$

After taking expectations, we finally obtain that there exist positive real numbers $C_0(\Theta)$, $C_1(\Theta)$, $C_2(\Theta)$ and $C_{12}(\Theta)$ depending only on Θ such as

$$\begin{aligned}\mathbb{E}\big[\sup_{(\theta_1,\theta_2)\in\Theta} Z^2_{(\theta_1,\theta_2)}\big] \le\ & C_0(\Theta)\sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[Z^2_{(\theta_1,\theta_2)}\big] + C_1(\Theta)\sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[\big(\partial_{\theta_1}Z_{(\theta_1,\theta_2)}\big)^2\big] \\ &+ C_2(\Theta)\sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[\big(\partial_{\theta_2}Z_{(\theta_1,\theta_2)}\big)^2\big] + C_{12}(\Theta)\sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[\big(\partial^2_{\theta_1\theta_2}Z_{(\theta_1,\theta_2)}\big)^2\big]. \quad (31)\end{aligned}$$

Now, using the previous point 1., we obtain that for any $t\ge 1$:

$$\sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[Z^2_{(\theta_1,\theta_2)}\big] = \sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[\big(M^t_{(\theta_1,\theta_2)} - \widetilde{M}^t_{(\theta_1,\theta_2)}\big)^2\big] \le \widetilde{C}\,t^{2\overline{d}-1}.$$

From point 2., the same bound can be established as well for $\sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[\big(\partial_{\theta_1}Z_{(\theta_1,\theta_2)}\big)^2\big]$ and for $\sup_{(\theta_1,\theta_2)\in\Theta}\mathbb{E}\big[\big(\partial_{\theta_1}Z_{(\theta_1,\theta_2)}\big)^2\big]$. From a straightforward extension to $\partial^2_{\theta_1\theta_2}Z_{(\theta_1,\theta_2)}$, (31) implies that there exists $\overline{C}>0$:

$$\mathbb{E}\big[\sup_{(\theta_1,\theta_2)\in\Theta}\big(M^t_{(\theta_1,\theta_2)} - \widetilde{M}^t_{(\theta_1,\theta_2)}\big)^2\big] \le \overline{C}\,t^{2\overline{d}-1}.$$

□

Proof of Proposition 3.1. For $\theta\in\Theta$, denote:

$$I_n(\theta) := \frac{1}{n}\sum_{t=1}^n \Phi\big((X_{t-k})_{k\ge0},\theta\big) \ \text{ and } \ \widetilde{I}_n(\theta) := \frac{1}{n}\sum_{t=1}^n \Phi\big((\widetilde{X}_{t-k})_{k\ge0},\theta\big), \quad (32)$$

where Φ is defined in (10). The proof will be stepped in 3 points:

1. We prove here that $\sup_{\theta\in\Theta}\big|I_n(\theta)-I(\theta)\big| \xrightarrow[n\to\infty]{a.s.} 0$, with

$$I(\theta) := \mathbb{E}\big[\Phi\big((X_{-k})_{k\ge0},\theta\big)\big] \ \text{ for } \theta\in\Theta. \quad (33)$$

Indeed, from Doukhan and Wintenberger (2008), there exists a function $H:\mathbb{R}^{\mathbb{N}}\to\mathbb{R}$ such as for any $t\in\mathbb{Z}$, $X_t = H((\xi_{t-j})_{j\ge0})$ and therefore $(X_t)_{t\in\mathbb{Z}}$ is a second order ergodic stationary sequence since $r=2$. Then, for $\theta\in\Theta$, there exists $H_\Phi:\mathbb{R}^{\mathbb{N}}\to[0,\infty)$ such that

$$\Phi\big((X_{-k})_{k\ge0},\theta\big) = \Phi\big(\big((H(\xi_{k-j}))_{j\ge0}\big)_{k\ge0},\theta\big) = H_\Phi\big((\xi_{-j})_{j\ge0}\big),$$

with also $\mathbb{E}\big[\big|\Phi\big((X_{-k})_{k\geq 0},\theta\big)\big|\big]<\infty$. Then using Theorem 36.4 in Billingsley (1995), $\big(\Phi\big((X_{-k})_{k\geq 0},\theta\big)\big)_{t\in\mathbb{Z}}$ is an ergodic stationary sequence for any $\theta\in\Theta$. As a consequence, for any $\theta\in\Theta$,

$$I_n(\theta)\ \xrightarrow[n\to\infty]{a.s.}\ I(\theta).$$

Moreover, since Θ is a compact set, using Theorem 2.2.1. in Straumann (2005), we deduce that $\big(\Phi\big((X_{t-k})_{k\geq 0},\theta\big)\big)_{t\in\mathbb{Z}}$ also follows a uniform ergodic theorem and we obtain $\sup_{\theta\in\Theta}\big|I_n(\theta)-I(\theta)\big|\ \xrightarrow[n\to\infty]{a.s.}\ 0$.

2. We also have $\sup_{\theta\in\Theta}\big|I_n(\theta)-\widetilde{I}_n(\theta)\big|\ \xrightarrow[n\to\infty]{a.s.}\ 0$. We have

$$\big|I_n(\theta)-\widetilde{I}_n(\theta)\big| \ \leq\ \frac{1}{n}\sum_{t=1}^{n}\big|\Phi\big((X_{t-k})_{k\geq 0},\theta\big)-\Phi\big((\widetilde{X}_{t-k})_{k\geq 0},\theta\big)\big|, \tag{34}$$

and for any $\theta\in\Theta$,

$$\begin{aligned}\big|\Phi\big((X_{t-k})_{k\geq 0},\theta\big)-\Phi\big((\widetilde{X}_{t-k})_{k\geq 0},\theta\big)\big| &= \Big|\big(|X_t|-\big|M_\theta(t)\big|\big)^2-\big(|X_t|-\big|\widetilde{M}_\theta(t)\big|\big)^2\Big|\\ &\leq \big|M_\theta(t)-\widetilde{M}_\theta(t)\big|\Big(2\,|X_t|+\big|M_\theta(t)\big|+\big|\widetilde{M}_\theta(t)\big|\Big),\end{aligned}$$

with $M_\theta(t)$ and $\widetilde{M}_\theta(t)$ defined in (26). Therefore, using Cauchy-Schwarz and Minkowski inequalities,

$$\begin{aligned}&\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|\Phi\big((X_{t-k})_{k\geq 0},\theta\big)-\Phi\big((\widetilde{X}_{t-k})_{k\geq 0},\theta\big)\big|\Big]\\ &\quad\leq \mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\big|M_\theta(t)-\widetilde{M}_\theta(t)\big|\Big\}\Big(2\,|X_t|+\sup_{\theta\in\Theta}\Big\{\big|M_\theta(t)\big|+\big|\widetilde{M}_\theta(t)\big|\Big\}\Big)\Big]\\ &\quad\leq \Big(\mathbb{E}\Big[\sup_{\theta\in\Theta}\Big\{\big(M_\theta(t)-\widetilde{M}_\theta(t)\big)^2\Big\}\Big]\Big)^{1/2}\Big(2\,\|X_t\|_2+\Big(\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|M_\theta(t)\big|^2\Big]\Big)^{1/2}+\Big(\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|\widetilde{M}_\theta(t)\big|^2\Big]\Big)^{1/2}\Big)\end{aligned} \tag{35}$$

Now, since Θ is a compact set included in $\Theta(2)$, we have $\|X_t\|_2<\infty$. Moreover, using Lemma 5.1, point 3., for $t=1$, we have $\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|M_\theta(1)\big|^2\Big]=\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|M_\theta(t)\big|^2\Big]\leq\overline{C}$.
With $\big|\widetilde{M}_\theta(t)\big|^2\leq 2\,\big(M_\theta(t)-\widetilde{M}_\theta(t)\big)^2+2\,\big(M_\theta(t)\big)^2$, this also implies that $\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|\widetilde{M}_\theta(t)\big|^2\Big]\leq 4\,\overline{C}$. Therefore, there exists $C>0$ such as for any $t\geq 1$,

$$\Big(2\,\|X_t\|_2+\Big(\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|M_\theta(t)\big|^2\Big]\Big)^{1/2}+\Big(\mathbb{E}\Big[\sup_{\theta\in\Theta}\big|\widetilde{M}_\theta(t)\big|^2\Big]\Big)^{1/2}\Big)\leq C. \tag{36}$$

Therefore, from (35), (36) and Lemma 5.1, point 3., we deduce that there exists $C>0$ such as:

$$\mathbb{E}\big[\sup_{\theta\in\Theta}\big|\Phi\big((X_{t-k})_{k\geq 0},\theta\big)-\Phi\big((\widetilde{X}_{t-k})_{k\geq 0},\theta\big)\big|\big]\leq C\,t^{2\overline{d}-1}\quad\text{for any } t\geq 1.$$

Then, using this bound and since $\overline{d}<1/2$, there exists $C>0$ such as:

$$\sum_{t=1}^{n}\frac{1}{t}\,\mathbb{E}\big[\sup_{\theta\in\Theta}\big|\Phi\big((X_{t-k})_{k\geq 0},\theta\big)-\Phi\big((\widetilde{X}_{t-k})_{k\geq 0},\theta\big)\big|\big]\leq C\sum_{t=1}^{n}t^{2\overline{d}-2}\leq C\sum_{t=1}^{\infty}t^{2\overline{d}-2}<\infty. \tag{37}$$

In Corollary 1 of Kounias and Weng (1969), it is established that $\displaystyle\sum_{t=1}^{\infty}\frac{\mathbb{E}\big[|Z_t|\big]}{b_t}<\infty$ implies $\displaystyle\frac{1}{b_n}\sum_{t=1}^{n}Z_t\ \xrightarrow[n\to\infty]{a.s.}\ 0$ for a $\mathbb{L}^1$ sequence of r.v. $(Z_t)_t$ and $b_n\underset{n\to\infty}{\longrightarrow}\infty$. Therefore, with $b_t=t$ for

$t \in \mathbb{N}^*$, (37) leads to

$$\frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta} \left|\Phi\big((X_{t-k})_{k\geq 0},\theta\big) - \Phi\big((\widetilde{X}_{t-k})_{k\geq 0},\theta\big)\right| \xrightarrow[n\to\infty]{a.s.} 0,$$

and therefore to $\sup_{\theta\in\Theta} \left|I_n(\theta) - \widetilde{I}_n(\theta)\right| \xrightarrow[n\to\infty]{a.s.} 0$ from (34).

3. The two previous points show us that $\sup_{\theta\in\Theta} \left|\widetilde{I}_n(\theta) - I(\theta)\right| \xrightarrow[n\to\infty]{a.s.} 0$ with I defined in (33). The proof is achieved if we establish that θ^* is the unique minimum of $\theta \in \Theta \mapsto I(\theta)$. This is induced by the following computations:

$$\begin{aligned}
I(\theta) &= \mathbb{E}\big[\Phi\big((X_{-k})_{k\geq 0},\theta\big)\big] \\
&= \mathbb{E}\Big[\Big(|\xi_0|\,\big|a_0(\theta^*) + \sum_{j=1}^{\infty} a_j(\theta^*)\,X_{-j}\big| - \big|a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta)\,X_{-j}\big|\Big)^2\Big] \\
&= \mathbb{E}\big[\xi_0^2 - 1\big]\,\mathbb{E}\Big[\Big(\big|a_0(\theta^*) + \sum_{j=1}^{\infty} a_j(\theta^*)\,X_{-j}\big|\Big)^2\Big] \\
&\qquad + \mathbb{E}\Big[\Big(\big|a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta)\,X_{-j}\big| - \big|a_0(\theta^*) + \sum_{j=1}^{\infty} a_j(\theta^*)\,X_{-j}\big|\Big)^2\Big],
\end{aligned}$$

using the assumption $\|\xi_0\|_1 = 1$ and because (X_t) is a causal time series implying that ξ_0 independent to $\sigma\big\{(X_{-k})_{k\geq 1}\big\}$. The first term of the previous relationship does not depend on θ. The second one vanishes when $\theta = \theta^*$. It is also non negative and it vanishes if

$$\big|a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta)\,X_{-j}\big| = \big|a_0(\theta^*) + \sum_{j=1}^{\infty} a_j(\theta^*)\,X_{-j}\big| \quad a.s.$$

As we assumed that $a_0(\cdot)$ is a positive function, using also Assumption $\mathrm{Id}(\Theta)$, we deduce that $\theta = \theta^*$ is the only solution of the previous equality. As a consequence, θ^* is the unique minimizer of $I(\cdot)$ and since $\sup_{\theta\in\Theta} \left|\widetilde{I}_n(\theta) - I(\theta)\right| \xrightarrow[n\to\infty]{a.s.} 0$ and $\widehat{\theta}_n = \widehat{\theta}_n = \operatorname*{Argmin}_{\theta\in\Theta} \widetilde{I}_n(\theta)$, we deduce that $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta^*$. □

Proof of Corollary 3.1. In such a case, $\theta = {}^t(a_0,\ldots,a_p)$ and therefore $a_i(\theta) = a_i$ for any $0 \leq i \leq p$ and $a_i(\theta) = 0$ for $i \geq p+1$. Then, $\theta \in \Theta \mapsto a_k(\theta)$ are $\mathcal{C}^\infty$ functions on Θ for any $k \in \mathbb{N}$. Moreover, since $\sup_{\theta\in\Theta} |a_k(\theta)| = 0$ for $k \geq p+1$ as well as $\sup_{\theta\in\Theta} \left\|\partial^j_{\theta^j} a_k(\theta)\right\| = 0$ for any $1 \leq j \leq p+1$, therefore (13) is satisfied and Assumption $C_{p+1}(\Theta)$ holds. Since $\sum_{k=1}^{\infty} \sup_{\theta\in\Theta} |a_k(\theta)| < \infty$, Assumption (S) is also satisfied. Finally, Assumption $\mathrm{Id}(\Theta)$ is obviously satisfied: $a_i(\theta) = a_i(\widetilde{\theta})$ implies $a_i = \widetilde{a}_i$, which implies $\theta = {}^t(a_0,\ldots,a_p) = \widetilde{\theta} = {}^t(\widetilde{a}_0,\ldots,\widetilde{a}_p)$ for any $i \in \mathbb{N}$. Then the strong consistency of $\widehat{\theta}_n$ is established from Proposition 3.1. □

Proof of Corollary 3.2. Denote $\theta = \big(\theta_P, \theta_Q\big) \in \Theta$ where $\theta_P = {}^t\big(c_0, c_1, \ldots, c_p\big) \in (0,\infty)\times\mathbb{R}^p$, $\theta_Q = {}^t\big(d_1,\ldots,d_q\big) \in \mathbb{R}^q$. Then $\sigma = P^{-1}_{\theta_P}(B)\,Q_{\theta_Q}(B)X$. It is clear that $\theta_P \to P_{\theta_P}$ is an injective function, and it is the same for $\theta_P \to P^{-1}_{\theta_P}$ and $\theta_Q \to Q_{\theta_Q}$. Finally it is also the same for $\theta = {}^t\big(\theta_P, \theta_Q\big) \to P^{-1}_{\theta_P} \times Q_{\theta_Q}$, because P_{θ_P} and Q_{θ_Q} are not zero polynomial and because we assume that the $p+1$ components of θ_P are free of the q components θ_Q, *i.e.* there are no supposed links between $(c_i)_{0\leq i\leq p}$ and $(d_j)_{1\leq j\leq q}$. Therefore, (15) and Assumption $\mathrm{Id}(\Theta)$ are satisfied for GLARCH(p,q) process.

Moreover, as it is well known for GARCH(p,q) processes, since $\sum_{j=1}^{q}|d_j| \leq \rho$ with $0 \leq \rho < 1$ from the expression of $\Theta_{p,q}(2)$, the roots $z_j(\theta)$ of the characteristic polynomial $\xi(z) = z^q - \sum_{j=1}^{q} b_j z^{q-j}$ satisfies $\sup_\theta \max_j |z_j(\theta)| < 1$: $(a_k(\theta))_k$ decreases exponentially fast towards 0. Then $\sum_{k=1}^{\infty}\sup_{\theta\in\Theta}|a_k(\theta)| < \infty$ and Assumption (S) is satisfied. Moreover, there exists $c_0 > 0$ such as $\sup_{\theta\in\Theta}|a_k(\theta)| \leq c_0\, k^{-3/4}$ for any $k \geq 1$. By considering the derivatives of equation (7), we also have $\sup_{\theta\in\Theta}\left\|\partial^j_{\theta^j} a_k(\theta)\right\| \leq c_j\, k^{-3/4}$ with $c_j > 0$ for any $k, j \geq 1$ and Assumption $C_{p+q+1}(\Theta)$ holds. □

Proof of Corollary 3.3. The assumptions of Corollary 3.3 are exactly the same as those of Proposition 3.1. It remains for us to prove that (13) could be satisfied under the condition (9), *i.e.* there exists $d(\theta) \in (0, 1/2)$ and $L_\theta(\cdot)$ a slowly varying function such that $a_j(\theta) = L_\theta(j)\, j^{d(\theta)-1}$ for $j \in \mathbb{N}^*$. Since $\theta \in \Theta$ a compact set, there exists $D \in (0, 1/2)$ such that $d(\theta) \leq D$ for any $\theta \in \Theta$. Moreover, since $L_\theta(\cdot)$ is a slow varying function and $\theta \in \Theta$ a compact subset of $\mathbb{R}^\ell$, there exists $C_L > 0$ such that $\sup_{\theta\in\Theta}\left|L_\theta(j)\right| \leq C_L\, j^{1/4-D/2}$ for any $j \in \mathbb{N}^*$, with $1/4 - D/2 > 0$. As a consequence,

$$\sup_{\theta\in\Theta}\left|a_j(\theta)\right| \leq C_L\, j^{1/4-D/2} j^{D-1} \leq C_L\, j^{D/2-3/4} \quad \text{for any } j \in \mathbb{N}^*.$$

With $\overline{d} = D/2 + 1/4 < 1/2$, Assumption $C_0(\Theta)$ is verified and Corollary 3.3 is established. □

Proof of Corollary 3.4. We have to prove that the assumptions of Corollary (3.3) are satisfied in this particular case of long memory LARCH(∞) process.
First, we have $\theta = {}^t(a_0, c, d) \in \Theta \subset \mathbb{R}^3$ and Assumption $C_3(\Theta)$ has to be verified. It is clear that $a_k : \theta \mapsto c\, k^{d-1}$ is a $\mathcal{C}^\infty$ function on Θ, with θ defined in (18). For $k \geq 3$, easy computations imply that:

$$\sup_{\theta\in\Theta}\left\|\partial^j_{\theta^j} a_k(\theta)\right\| \leq \sqrt{c_M^2 + j} \times \log k \times k^{d_M-1} \quad \text{for } 1 \leq j \leq 3.$$

Therefore, by considering for instance $\overline{d} = d_M/2 + 1/2 \in (d_M, 1/2)$, it is clear that there exists $C_a > 0$ such as (13) is satisfied and Assumption $C_3(\Theta)$ holds.
It remains to prove Assumption Id(Θ). This one will be verified by considering the equality $a_j(\theta) = a_j(\widetilde{\theta})$ for any $j \in \mathbb{N}$ where $\theta = {}^t(a_0, c, d)$ and $\widetilde{\theta} = {}^t(\widetilde{a}_0, \widetilde{c}, \widetilde{d})$. This implies $a_0 = \widetilde{a}_0$ and $c\, j^{d-1} = \widetilde{c}\, j^{\widetilde{d}-1}$ for any $j \in \mathbb{N}$, leading to $\theta = \widetilde{\theta}$: Assumption Id$(\Theta)$ is also satisfied. □

Proof of Theorem 3.1. Let $I_n(\theta)$ and $\widetilde{I}_n(\theta)$ be defined in (32). We follow a proof that is similar to the one of Theorem 2 in Davis and Dunsmuir (1997).
Let $v = \sqrt{n}(\theta - \theta^*) \in \mathbb{R}^\ell$ and define

$$\begin{aligned} W_n(v) &= \sum_{t=1}^{n} \Phi\big((X_{t-k})_{k\geq 0}, \theta^* + n^{-1/2}v\big) - \Phi\big((X_{t-k})_{k\geq 0}, \theta^*\big) = n\left(I_n(\theta) - I_n(\theta^*)\right) \\ \text{and} \quad \widetilde{W}_n(v) &= \sum_{t=1}^{n} \Phi\big((\widetilde{X}_{t-k})_{k\geq 0}, \theta^* + n^{-1/2}v\big) - \Phi\big((\widetilde{X}_{t-k})_{k\geq 0}, \theta^*\big) = n\left(I_n(\theta) - \widetilde{I}_n(\theta^*)\right). \end{aligned}$$

Then we are going to prove first that minimizing $\widetilde{I}_n(\theta)$ with respect to $\theta \in \Theta$ is equivalent to minimize $\widetilde{W}_n(v)$ with respect to $v \in \mathbb{R}^\ell$, which is also equivalent to minimize $W_n(v)$ with respect to $v \in \mathbb{R}^\ell$. Secondly, we will provide a limit theorem satisfied by $W_n(v)$ for any $v \in \mathbb{R}^\ell$. Then we are going to prove in 3. that $(W_n(\cdot))_n$ converges as a process of $\mathcal{C}(\mathbb{R}^\ell)$ (space of continuous functions on $\mathbb{R}^\ell$) to a limit process W. Hence the sequence of minimum of $\widetilde{W}_n$, *i.e* $(\widehat{v}_n)_n$ with $\widehat{v}_n = \sqrt{n}(\widehat{\theta}_n - \theta^*)$,

will converge in distribution to the distribution of the minimum of $W(\cdot)$.

1. For any $v \in \mathbb{R}^\ell$ and $n \geq 1$, we have:

$$\begin{aligned}
W_n(v) &= \sum_{t=1}^{n} \Big(\big|X_t\big| - \big|M_{\theta^*+n^{-1/2}v}(t)\big|\Big)^2 - \Big(\big|X_t\big| - \big|M_{\theta^*}(t)\big|\Big)^2 \\
&= \sum_{t=1}^{n} \Big(\big|X_t\big| - \Big|M_{\theta^*}(t) - \frac{1}{\sqrt{n}}\, {}^t v\, \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big|\Big)^2 - \Big(\big|X_t\big| - \big|M_{\theta^*}(t)\big|\Big)^2 \\
&= \sum_{t=1}^{n} \Big(\big|X_t\big| - \big|M_{\theta^*}(t)\big| - \frac{1}{\sqrt{n}}\, {}^t v\, \partial_\theta M_{\overline{\theta}_t^{(n)}}(t) \times \mathrm{sgn}(M_{\theta^*}(t))\Big)^2 - \Big(\big|X_t\big| - \big|M_{\theta^*}(t)\big|\Big)^2 \\
&= -\frac{2}{\sqrt{n}} \sum_{t=1}^{n} \big(\big|X_t\big| - \big|M_{\theta^*}(t)\big|\big) \times \mathrm{sgn}(M_{\theta^*}(t)) \times {}^t v\, \partial_\theta M_{\overline{\theta}_t^{(n)}}(t) + \frac{1}{n} \sum_{t=1}^{n} \big({}^t v\, \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\big)^2 \\
&= J_1^{(n)}(v) + J_2^{(n)}(v) \qquad (38)
\end{aligned}$$

with $\overline{\theta}_t^{(n)} = \alpha_t^{(n)}\, \theta^* + (1 - \alpha_t^{(n)})\big(\theta^* + n^{-1/2}v\big)$ where $\alpha_t^{(n)} \in [0,1]$ is given from the Taylor-Lagrange expansion.

Term $J_2^{(n)}(v)$: For any $v \in \mathbb{R}^\ell$, $\overline{\theta}_t^{(n)} \xrightarrow[n\to\infty]{a.s.} \theta^*$ and then

$$\Big|\partial_\theta M_{\overline{\theta}_t^{(n)}}(t) - \partial_\theta M^t_{\theta^*)}\Big| \xrightarrow[n\to\infty]{a.s.} 0 \quad \text{for any } t \in \mathbb{N}, \qquad (39)$$

since the functions $\theta \in \Theta \mapsto \partial_\theta a_i(\theta)$ are supposed to be continuous functions for any $i \in \mathbb{N}$. Then we obtain for any $v \in \mathbb{R}^\ell$:

$$\begin{aligned}
\Big|J_2^{(n)}(v) - \mathbb{E}\big[\big({}^t v\, \partial_\theta M_{\theta^*}(0)\big)^2\big]\Big| \leq \frac{1}{n} \sum_{t=1}^{n} \Big|\big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2 - \big({}^t v\, \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)^2\Big| \\
+ \Big|\frac{1}{n} \sum_{t=1}^{n} \big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2 - \mathbb{E}\big[\big({}^t v\, \partial_\theta M_{\theta^*}(0)\big)^2\big]\Big|.
\end{aligned}$$

Now using Cesaro Lemma we obtain from (39),

$$\Big|\frac{1}{n} \sum_{t=1}^{n} \big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2 - \mathbb{E}\big[\big({}^t v\, \partial_\theta M_{\theta^*}(0)\big)^2\big]\Big| \xrightarrow[n\to\infty]{a.s.} 0. \qquad (40)$$

Moreover, we have seen that there exist a function H such as $X_t = H((\xi_{t-j})_{j\geq 0})$ for any $t \in \mathbb{Z}$, and therefore $(X_t)_{t\in\mathbb{Z}}$ is a second order ergodic stationary sequence. Then, for any $v \in \mathbb{R}^\ell$, there exists a function $H_v : \mathbb{R}^{\mathbb{N}} \to [0,\infty)$ such as

$$\big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2 = H_v\big((\xi_{-j})_{j\geq 0}\big),$$

with also $\mathbb{E}\big[\big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2\big] < \infty$ (see 1. of Lemma 5.1). Then using Theorem 36.4 in Billingsley (1995), $\big(\big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2\big)_{t\in\mathbb{Z}}$ is an ergodic stationary sequence implying to:

$$\frac{1}{n} \sum_{t=1}^{n} \Big|\big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2 - \mathbb{E}\big[\big({}^t v\, \partial_\theta M_{\theta^*}(0)\big)^2\big]\Big| \xrightarrow[n\to\infty]{a.s.} 0. \qquad (41)$$

Finally, with (40) and (41), we obtain for any $v \in \mathbb{R}^\ell$,

$$J_2^{(n)}(v) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}\big[\big({}^t v\, \partial_\theta M_{\theta^*}(0)\big)^2\big] = {}^t v\, \Gamma_1^*\, v \quad \text{where} \quad \Gamma_1^* := \mathbb{E}\Big[\partial_\theta M_{\theta^*}(0) \times {}^t\partial_\theta M_{\theta^*}(0)\Big], \tag{42}$$

where Γ_1^* is as in (19).

Term $J_1^{(n)}(v)$: We also have

$$\begin{aligned} J_1^{(n)}(v) &= -\frac{2}{\sqrt{n}} \sum_{t=1}^n \big(\big|M_{\theta^*}(t)\,\xi_t\big| - \big|M_{\theta^*}(t)\big|\big) \times \operatorname{sgn}(M_{\theta^*}(t)) \times {}^t v\, \partial_\theta M_{\overline{\theta}_t^{(n)}}(t) \\ &= {}^t v \Big(-\frac{2}{\sqrt{n}} \sum_{t=1}^n \big(\big|\xi_t\big| - 1\big)\, M_{\theta^*}(t) \times \partial_\theta M_{\theta^*}(t) \\ &\qquad\qquad + \frac{2}{n} \sum_{t=1}^n \big(\big|\xi_t\big| - 1\big)\, M_{\theta^*}(t) \times \sqrt{n}\, \big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big) \\ &= {}^t v\, \big(K_1^{(n)}(v) + K_2^{(n)}(v)\big). \end{aligned}$$

We have $\Big(\big(\big|\xi_t\big| - 1\big)\, M_{\theta^*}(t) \times \partial_\theta M_{\theta^*}(t)\Big)_{t\in\mathbb{N}}$ that is a stationary ergodic martingale difference since with the σ-algebra $\mathcal{F}_t = \sigma\big\{(X_{t-k})_{k\geq 1}\big\}$,

$$\mathbb{E}\Big[\big(\big|\xi_t\big| - 1\big)\, M_{\theta^*}(t) \times \partial_\theta M_{\theta^*}(t) \;\Big|\; \mathcal{F}_t\Big] = \mathbb{E}\big[\big|\xi_t\big| - 1\big]\, \mathbb{E}\Big[M_{\theta^*}(t) \times \partial_\theta M_{\theta^*}(t)\Big] = 0,$$

because (X_t) is a causal process and ξ_t is independent of $\mathcal{F}_t$ and $\mathbb{E}\big[\big|\xi_0\big|\big] = 1$.
Now since $\Gamma_2^* := \mathbb{E}\Big[\big(M_{\theta^*}(0)\big)^2 \partial_\theta M_{\theta^*}(0) \times {}^t\partial_\theta M_{\theta^*}(0)\Big]$ is supposed to be a finite definite positive matrix (see also its expression in (20)),

$$\mathbb{E}\Big[\big(\big|\xi_0\big| - 1\big)^2 \big\|M_{\theta^*}(0) \times \partial_\theta M_{\theta^*}(0)\big\|^2\Big] = (\sigma_\xi^2 - 1)\, \mathbb{E}\Big[\big\|M_{\theta^*}(0) \times \partial_\theta M_{\theta^*}(0)\big\|^2\Big] < \infty.$$

Then the central limit for stationary ergodic martingale difference, Theorem 18.3 of Billingsley (1968) can be applied and we obtain for any $v \in \mathbb{R}^\ell$:

$$K_1^{(n)}(v) \xrightarrow[n\to\infty]{\mathcal{L}} K_1 \overset{\mathcal{L}}{\sim} \mathcal{N}\big(0\,,\, 4\,(\sigma_\xi^2 - 1)\,\Gamma_2^*\big). \tag{43}$$

For any $t \in \mathbb{Z}$, by the definition of $\overline{\theta}_t^{(n)}$, we have $\sqrt{n}\,(\overline{\theta}_t^{(n)} - \theta^*) = (1 - \alpha_t^{(n)})\, v$ with $\alpha_t^{(n)}$ a random variable of $\mathcal{F}_t$ bounded in $[0,1]$. Therefore, using a Taylor-Lagrange expansion, we also have

$$\Big|\sqrt{n}\, {}^t v\, \big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big| = (1 - \alpha_t^{(n)})^t v\, \partial_{\theta^2} M^t_{\overline{\overline{\theta}}_t^{(n)}}\, v \leq \sup_{\theta\in\Theta} \big\|\partial^2_{\theta^2} M_\theta(t)\big\|\, \big\|v\big\|^2 \quad \text{for any } t \in \mathbb{Z},$$

where $\overline{\overline{\theta}}_t^{(n)} = \beta_t^{(n)}\theta^* + (1 - \beta_t^{(n)})\overline{\theta}_t^{(n)}$ and $\beta_t^{(n)} \in [0,1]$.
We have $\mathbb{E}\Big[\Big(\sup_{\theta\in\Theta} \big\|\partial^2_{\theta^2} M_\theta(t)\big\|\Big)^2\Big] = \mathbb{E}\Big[\sup_{\theta\in\Theta} \big\|\partial^2_{\theta^2} M_\theta(0) - \partial^2_{\theta^2} \widetilde{M}_\theta(0)\big\|^2\Big]$. Using the same reasoning than in 3. of Lemma 5.1, $\mathbb{E}\Big[\sup_{\theta\in\Theta} \big\|\partial^2_{\theta^2} M_\theta(0) - \partial^2_{\theta^2} \widetilde{M}_\theta(0)\big\|^2\Big] < \infty$ when Assumption $C_{\ell+2}(\Theta)$ holds. As a consequence,

$$\mathbb{E}\Big[\Big(\sqrt{n}\, {}^t v\, \big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big)^2\Big] \leq \mathbb{E}\Big[\sup_{\theta\in\Theta} \big\|\partial^2_{\theta^2} M_\theta(t)\big\|^2\Big] \times \big\|v\big\|^4 < \infty.$$

Moreover, since $\overline{\theta}_t^{(n)} \in \mathcal{F}_t$, we also have $\big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big) \in \mathcal{F}_t$ and from the previous bound,

$$\begin{aligned}
&\mathbb{E}\Big[\Big|\big(|\xi_t|-1\big)\, M_{\theta^*}(t) \times \sqrt{n}\,{}^t v\big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big|\Big] \\
&\qquad = \mathbb{E}\big[\big||\xi_t|-1\big|\big] \times \mathbb{E}\Big[\Big|\big(\, M_{\theta^*}(t) \times \sqrt{n}\,{}^t v\big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big|\Big] \\
&\qquad \le \mathbb{E}\big[\big||\xi_t|-1\big|\big] \times \mathbb{E}\Big[\big(M_{\theta^*}(t)\big)^2\Big] \times \mathbb{E}\Big[\Big(\sqrt{n}\,{}^t v\big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big)^2\Big] < \infty. \quad (44)
\end{aligned}$$

Therefore $\big(|\xi_t|-1\big)\, M_{\theta^*}(t) \times \sqrt{n}\,{}^t v\big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)_t$ is a stationary causal sequence, and from (44), Theorem 36.4 in Billingsley (1995) implies that for any $v \in \mathbb{R}^\ell$,

$$ {}^t v\, K_2^{(n)}(v) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}\Big[\big(|\xi_t|-1\big)\, M_{\theta^*}(t) \times \sqrt{n}\,{}^t v\big(\partial_\theta M_{\theta^*}(t) - \partial_\theta M_{\overline{\theta}_t^{(n)}}(t)\big)\Big] = 0. \quad (45)$$

Finally, for any $v \in \mathbb{R}^\ell$, since $J_1^{(n)}(v) = {}^t v\,\big(K_1^{(n)}(v) + K_2^{(n)}(v)\big)$, then $J_1^{(n)}(v) \xrightarrow[n\to\infty]{\mathcal{L}} {}^t v\, K_1$ from (43) and (45), and with (42) this implies,

$$W_n(v) \xrightarrow[n\to\infty]{\mathcal{L}} {}^t v\,\Gamma_1^*\, v + {}^t v\, K_1 \quad \text{with} \quad K_1 \overset{\mathcal{L}}{\sim} \mathcal{N}\big(0\,,\, 4\,(\sigma_\xi^2 - 1)\,\Gamma_2^*\big). \quad (46)$$

2. Asymptotically, for any $v \in \mathbb{R}^\ell$, from 1., we know that the law of $W_n(v)$ is the same as the law of:

$$W_n'(v) = -\frac{2}{\sqrt{n}} \sum_{t=1}^n \big(|\xi_t| - 1\big)\, M_{\theta^*}(t) \times {}^t v\, \partial_\theta M_{\theta^*}(t) + \frac{1}{n} \sum_{t=1}^n \big({}^t v\, \partial_\theta M_{\theta^*}(t)\big)^2$$

and we deduce the same kind of result for the law of $\widetilde{W}_n(v)$, which is asymptotically equivalent to the one of:

$$\widetilde{W}_n'(v) = -\frac{2}{\sqrt{n}} \sum_{t=1}^n \big(|\xi_t| - 1\big)\, \widetilde{M}_{\theta^*}(t) \times {}^t v\, \partial_\theta \widetilde{M}_{\theta^*}(t) + \frac{1}{n} \sum_{t=1}^n \big({}^t v\, \partial_\theta \widetilde{M}_{\theta^*}(t)\big)^2.$$

Therefore we obtain:

$$\begin{aligned}
W_n'(v) - \widetilde{W}_n'(v) &= -\frac{2}{\sqrt{n}} \sum_{t=1}^n \big(|\xi_t| - 1\big)\, {}^t v \Big(M_{\theta^*}(t)\, \partial_\theta M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\, \partial_\theta \widetilde{M}_{\theta^*}(t)\Big) \\
&\quad + {}^t v \Big(\frac{1}{n} \sum_{s=1}^n \big(\partial_\theta M_{\theta^*}^s\, {}^t \partial_\theta M_{\theta^*}^s - \partial_\theta \widetilde{M}_{\theta^*}^s\, {}^t \partial_\theta \widetilde{M}_{\theta^*}^s\big)\Big)\, v \quad (47) \\
&= L_1^{(n)} + L_2^{(n)}. \quad (48)
\end{aligned}$$

Term $L_1^{(n)}$: Using the causality of (X_t), *i.e.* ξ_t independent to $\sigma\{X_{t-1}, X_{t-2}, \ldots\}$ for any $t \in \mathbb{Z}$, we deduce that:

$$\mathbb{E}\Big[\frac{1}{\sqrt{n}} \sum_{t=1}^n \big(|\xi_t| - 1\big)\, {}^t v \Big(M_{\theta^*}(t)\, \partial_\theta M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\, \partial_\theta \widetilde{M}_{\theta^*}(t)\Big)\Big] = 0, \quad (49)$$

since $\mathbb{E}[|\xi_t|] = 1$. Moreover, we have for any $t \ge 1$:

$$M_{\theta^*}(t)\, \partial_\theta M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\, \partial_\theta \widetilde{M}_{\theta^*}(t) = M_{\theta^*}(t)\, \big(\partial_\theta M_{\theta^*}(t) - \partial_\theta \widetilde{M}_{\theta^*}(t)\big) + \partial_\theta \widetilde{M}_{\theta^*}(t)\, \big(M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\big).$$

Using 1. and 2. of Lemma 5.1, we have for any $t \geq 1$:

$$\mathbb{E}\big[\big(M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\big)^2\big] \leq \widetilde{C}\, t^{2\overline{d}-1} \quad \text{and} \quad \mathbb{E}\big[\big\|\partial_\theta M_{\theta^*}(t) - \partial_\theta \widetilde{M}_{\theta^*}(t)\big\|^2\big] \leq \widetilde{C}\, t^{2\overline{d}-1}.$$

As a consequence, using Cauchy-Schwarz and Minkowski inequalities,

$$\big\|M_{\theta^*}(t)\,\partial_\theta M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\,\partial_\theta \widetilde{M}_{\theta^*}(t)\big\|_2 \leq \sqrt{\widetilde{C}}\,\big(\big\|M_{\theta^*}(0)\|_2 + \big\|\partial_\theta M_{\theta^*}(0)\|_2\big)\, t^{\overline{d}-1/2}. \tag{50}$$

Therefore, there exists $C > 0$ such as for any $n \geq 1$,

$$\begin{aligned}
&\mathbb{E}\Big[\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(|\xi_t| - 1\big)\,{}^t v\,\big(M_{\theta^*}(t)\,\partial_\theta M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\,\partial_\theta \widetilde{M}_{\theta^*}(t)\big)\Big)^2\Big] \\
&\qquad = \frac{1}{n}\,(\sigma_\xi^2 - 1)\,\|v\|^2 \sum_{t=1}^{n}\big\|M_{\theta^*}(t)\,\partial_\theta M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\,\partial_\theta \widetilde{M}_{\theta^*}(t)\big\|_2^2 \\
&\qquad \leq \frac{C}{n}\,(\sigma_\xi^2 - 1)\,\|v\|^2 \sum_{t=1}^{n} t^{2\overline{d}-1} \\
&\qquad \leq \frac{C}{n}\,(\sigma_\xi^2 - 1)\,\|v\|^2 \Big(1 + \int_1^n x^{2\overline{d}-1} dx\Big) \leq C'\,\|v\|^2\, n^{2\overline{d}-1}
\end{aligned}$$

with $C' > 0$ and for n large enough, using again the fact that for f a positive decreasing function $\sum_{k=1}^{n} f(k) \leq f(1) + \int_1^n f(t)dt$. Therefore, for any $v \in \mathbb{R}$,

$$\mathbb{E}\Big[\Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(|\xi_t| - 1\big)\,{}^t v\,\big(M_{\theta^*}(t)\,\partial_\theta M_{\theta^*}(t) - \widetilde{M}_{\theta^*}(t)\,\partial_\theta \widetilde{M}_{\theta^*}(t)\big)\Big)^2\Big] \underset{n\to\infty}{\longrightarrow} 0. \tag{51}$$

Therefore, from (49) and (51), we deduce $\mathbb{E}\big[L_1^{(n)}\big] = 0$ and $\mathrm{Var}\big(L_1^{(n)}\big) \underset{n\to\infty}{\longrightarrow} 0$. Using Bienaymé-Tchebytchev Inequality, this implies that for any $v \in \mathbb{R}^\ell$,

$$L_1^{(n)} \underset{n\to\infty}{\xrightarrow{\mathbb{P}}} 0. \tag{52}$$

Term $L_2^{(n)}$: Using the same method, we also obtain that there exist $C'' > 0$ such as for any $s \in \{1, \ldots, n\}$,

$$\big\|\partial_\theta M_{\theta^*}^s\,{}^t\partial_\theta M_{\theta^*}^s - \partial_\theta \widetilde{M}_{\theta^*}^s\,{}^t\partial_\theta \widetilde{M}_{\theta^*}^s\big\|_2 \leq C''\, s^{\overline{d}-1/2}. \tag{53}$$

Now with (53), we can use again the result established in part 2. of the proof of Proposition 3.1 based on the Corollary 1 of Kounias and Weng (1969) and obtain for any $n \in \mathbb{N}^*$,

$$\sum_{s=1}^{n}\frac{1}{s}\big\|\partial_\theta M_{\theta^*}^s\,{}^t\partial_\theta M_{\theta^*}^s - \partial_\theta \widetilde{M}_{\theta^*}^s\,{}^t\partial_\theta \widetilde{M}_{\theta^*}^s\big\|_1 \leq C'' \sum_{s=1}^{n} s^{\overline{d}-3/2} < \infty$$

since $\overline{d} < 1/2$. Therefore:

$$\frac{1}{n}\sum_{s=1}^{n}\big(\partial_\theta M_{\theta^*}^s\,{}^t\partial_\theta M_{\theta^*}^s - \partial_\theta \widetilde{M}_{\theta^*}^s\,{}^t\partial_\theta \widetilde{M}_{\theta^*}^s\big) \underset{n\to\infty}{\xrightarrow{a.s.}} 0, \tag{54}$$

implying

$$L_2^{(n)} \underset{n\to\infty}{\xrightarrow{a.s.}} 0 \qquad \text{for any } v \in \mathbb{R}^\ell. \tag{55}$$

Finally from (48), (52) and (55), we deduce that for any $v \in \mathbb{R}^\ell$,

$$\big|W_n(v) - \widetilde{W}_n(v)\big| \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Using (46), this implies

$$\widetilde{W}_n(v) \xrightarrow[n\to\infty]{\mathcal{L}} W(v) := {}^t v\,\Gamma_1^*\,v + {}^t v\,K_1 \quad \text{with} \quad K_1 \overset{\mathcal{L}}{\sim} \mathcal{N}\big(0\,,\,4\,(\sigma_\xi^2 - 1)\,\Gamma_2^*\big). \tag{56}$$

3. Now, using the same arguments than in the proof of Theorem 2 of Davis and Dunsmuir (1997), we deduce that finite distributions $(\widetilde{W}_n(v_1), \cdots, \widetilde{W}_n(v_k))$ converge to $(W(v_1), \cdots, W(v_k))$ for any $(v_1, \cdots, v_k) \in (\mathbb{R}^\ell)^k$. Moreover, always following the proof of Theorem 2 of Davis and Dunsmuir (1997), $(W_n(v))_v$ converges to $(W(v))_v$ as a process on the continuous function space $\mathcal{C}^0(\mathbb{R})$.
As a consequence, a maximum $\widehat{v} = \sqrt{n}\,\big(\widehat{\theta}_n - \theta^*\big)$ of $\widehat{W}_n(v)$ converges in distribution to the maximum of ${}^t v\,\Gamma_1^*\,v + {}^t v\,K_1$, which is $\overline{v} := -\frac{1}{2}\left(\Gamma_1^*\right)^{-1} K_1 \overset{\mathcal{L}}{\sim} \mathcal{N}\Big(0\,,\,(\sigma_\xi^2 - 1)\left(\Gamma_1^*\right)^{-1}\Gamma_2^*\left(\Gamma_1^*\right)^{-1}\Big)$ and this implies (21). □

Proof of Remark 3.4. Using the notations of the proof of Theorem 3.1, we have:

$$\widehat{\Gamma}_1 = \frac{1}{n}\sum_{s=1}^n \partial_\theta \widetilde{M}_{\widehat{\theta}_n}(s)\,{}^t\partial_\theta \widetilde{M}_{\widehat{\theta}_n}(s) \quad \text{and} \quad \widehat{\Gamma}_2 := \frac{1}{n}\sum_{s=1}^n \big(M_{\widehat{\theta}_n}(s)\big)^2 \partial_\theta \widetilde{M}_{\widehat{\theta}_n}(s)\,{}^t\partial_\theta \widetilde{M}_{\widehat{\theta}_n}(s).$$

Using the same computations than those required for establishing (53), we have for any $\theta \in \Theta$:

$$\big\|\partial_\theta M_\theta(s) \times {}^t\partial_\theta M_\theta(s) - \partial_\theta \widetilde{M}_\theta(s) \times {}^t\partial_\theta \widetilde{M}_\theta(s)\big\|_2 \le C''\, s^{\overline{d}-1/2}.$$

And following (54), we obtain for any $\theta \in \Theta$

$$\frac{1}{n}\sum_{s=1}^n \big(\partial_\theta M_\theta(s) \times {}^t\partial_\theta M_\theta(s) - \partial_\theta \widetilde{M}_\theta(s) \times {}^t\partial_\theta \widetilde{M}_\theta(s)\big) \xrightarrow[n\to\infty]{a.s.} 0$$

Using Theorem 2.2.1. in Straumann (2005), we deduce that:

$$\sup_{\theta\in\Theta}\Big|\frac{1}{n}\sum_{s=1}^n \big(\partial_\theta M_\theta(s) \times {}^t\partial_\theta M_\theta(s) - \partial_\theta \widetilde{M}_\theta(s) \times {}^t\partial_\theta \widetilde{M}_\theta(s)\big)\Big| \xrightarrow[n\to\infty]{a.s.} 0$$

and this implies

$$\Big|\frac{1}{n}\sum_{s=1}^n \big(\partial_\theta M_{\widehat{\theta}_n}(s) \times {}^t\partial_\theta M_{\widehat{\theta}_n}(s) - \partial_\theta \widetilde{M}_{\widehat{\theta}_n}(s) \times {}^t\partial_\theta \widetilde{M}_{\widehat{\theta}_n}(s)\big)\Big| \xrightarrow[n\to\infty]{a.s.} 0. \tag{57}$$

Moreover, for any $\theta \in \Theta$, since the process $\big(\partial_\theta M_\theta(s) \times {}^t\partial_\theta M_\theta(s)\big)_s$ is a stationary causal sequence such as $\mathbb{E}\big[\big\|\partial_\theta M_\theta(0) \times {}^t\partial_\theta M_\theta(0)\big\|\big] < \infty$, Theorem 36.4 in Billingsley (1995) implies :

$$\frac{1}{n}\sum_{s=1}^n \partial_\theta M_{\theta^*}(s) \times {}^t\partial_\theta M_{\theta^*}(s) \xrightarrow[n\to\infty]{a.s.} \Gamma_1^*.$$

But for any $s \in \{1, \ldots, n\}$, $\theta \in \Theta \mapsto \partial_\theta M_\theta(s)$ is a continuous function and since $\widehat{\theta}_n \xrightarrow[n\to\infty]{a.s.} \theta^*$,

$$\frac{1}{n}\sum_{s=1}^n \partial_\theta M_{\widehat{\theta}_n}(s) \times {}^t\partial_\theta M_{\widehat{\theta}_n}(s) \xrightarrow[n\to\infty]{a.s.} \Gamma_1^*. \tag{58}$$

Using (57) and (58), we deduce that $\widehat{\Gamma}_1 \xrightarrow[n\to\infty]{a.s.} \Gamma_1^*$.

Using the same reasoning as well as (50), we also deduce that $\widehat{\Gamma}_2 \xrightarrow[n\to\infty]{a.s.} \Gamma_2^*$. □

Proof of Corollary 3.5. Using exactly the same arguments as in the proof of Corollary 3.2 for establishing that Assumption $C_\ell(\Theta)$ holds with $\ell = p+q+1$, Assumption $C_{\ell+2}(\Theta)$ also holds. Moreover, for any GLARCH(p,q) process, the matrix Γ_1^* and Γ_2^* are positive definite matrix. Indeed, following the same reasoning as in the proof of Lemma 5 of Beran and Schützner (2009), we have for any $v \in \mathbb{R}^{p+q+1}$

$$ {}^t v\, \Gamma_1^*\, v = \mathbb{E}\Big[\big({}^t v\, \partial_\theta M_{\theta^*}(0)\big)^2\Big] \geq 0. $$

Assume that ${}^t v\, \partial_\theta M_{\theta^*}(0) = 0$. By stationarity, this implies ${}^t v\, \partial_\theta M_{\theta^*}(k) = 0$ for any $k \in \mathbb{Z}$. Using relation (7), we deduce that:

$$ \partial_\theta M_{\theta^*}(0) = \partial_\theta\big(c_0 + c_1\, X_{-1} + \cdots + c_p\, X_{t-p}\big) + \partial_\theta\big(d_1\, M_{\theta^*}(-1) + \cdots + d_q\, M_{\theta^*}(-q)\big). $$

Then:

$$ \left\{ \begin{array}{lcl} \partial_{c_0} M_{\theta^*}(0) & = & 1 + d_1\, \partial_{c_0} M_{\theta^*}(-1) + \cdots + d_q\, \partial_{c_0} M_{\theta^*}(-q) \\ \partial_{c_i} M_{\theta^*}(0) & = & X_{-i} + d_1\, \partial_{c_i} M_{\theta^*}(-1) + \cdots + d_q\, \partial_{c_i} M_{\theta^*}(-q) \quad 1 \leq i \leq p \\ \partial_{d_j} M_{\theta^*}(0) & = & M_{\theta^*}(-j) + d_1\, \partial_{d_j} M_{\theta^*}(-1) + \cdots + d_q\, \partial_{d_j} M_{\theta^*}(-q) \quad 1 \leq j \leq q \end{array} \right. $$

Therefore, if ${}^t v\, \partial_\theta M_{\theta^*}(0) = 0$, then $\big(1, X_{-1}, \ldots, X_{-p}, M_{\theta^*}(-1), \ldots, M_{\theta^*}(-q)\big)\, v = 0$. Such equation has no solution since a linear relationship (7) relies all these random variables to $M_{\theta^*}(0)$ (or these means that (X_t) would be a GLARCH$(p-1, q-1)$ process which is not possible since we have assumed that $P_{\theta_1^*}$ and $Q_{\theta_2^*}$ are coprime polynomials). Hence, ${}^t v\, \partial_\theta M_{\theta^*}(0) = 0$ implies $v = 0$: Γ_1^* is a positive definite matrix (and a similar reasoning leads to the same property satisfied by Γ_2^*). □

Proof of Corollary 3.6. As we had already written, in the case of a LARCH(p) process, we must have $\Theta \subset \Theta_p(4)$, which is defined in (6). Therefore, choosing Θ as defined in (23) guarantees that it is a compact subset of $\Theta_p(4)$. Moreover, since a LARCH(p) process is a particular case of a GLARCH(p,q) process, Corollary 3.5 is satisfied under the conditions of Corollary 3.6. □

Proof of Corollary 3.7. In the proof of Corollary 3.3, it was established that (13) could be satisfied under the condition (9), *i.e.* there exists $d(\theta) \in (0, 1/2)$ and $L_\theta(\cdot)$ a slowly varying function such that $a_j(\theta) = L_\theta(j)\, j^{d(\theta)-1}$ for $j \in \mathbb{N}^*$. The various assumptions required for the establishment of the central limit theorem (21) are also present in the assumptions of the Corollary 3.7. And condition (9) is not in contradiction when Assumption $C_{\ell+2}(\Theta)$ holds. □

Proof of Corollary 3.8. In the proof of Corollary 3.4, Assumption $C_3(\Theta)$ holds. But for $k \geq 3$, we also obtain $\sup_{\theta\in\Theta} \big\|\partial_{\theta^j}^j a_k(\theta)\big\| \leq \sqrt{c_M^2 + j} \times \log^4 k \times k^{d_M - 1}$ for $j = 4, 5$. Then always with $\overline{d} = d_M/2 + 1/2$, Assumption $C_5(\Theta)$ holds.
Moreover, *mutatis mutendis*, we can use again the proof of Lemma 5 of Beran and Schützner (2009) for proving that Γ_1^* and Γ_2^* are two positive definite matrix in this case of long memory LARCH(∞) process. □

Aknowledgments The author is very grateful to the referees and the editor for many relevant suggestions, corrections and comments that helped to notably improve the paper.

Data Availability Statement

Data available from request from the author.
Used softwares are available on `https://samm.univ-paris1.fr/Sofwares-Logiciels`

References

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

[2] Beran J. 2006. On location estimation for larch processes. *Journal of Multivariate Analysis* **97:** 1766–1782.

[3] Beran J., Schützner M. 2009. On approximate pseudo-maximum likelihood estimation for LARCH-processes. *Bernoulli* **15:** 1057–1081.

[4] Berkes I., Horváth, L. 2003. Asymptotic results for long memory larch sequences. *Annals of Applied Probability* **13:** 641–668.

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

[6] Billingsley P. 1995. *Probability and Measure.* New York: John Wiley & Sons Inc.

[7] Davis R., Dunsmuir W. 1997. Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability* **10:** 481–497.

[8] Doukhan P., Wintenberger O. 2008. Weakly dependent chains with infinite memory. *Stochastic Processes and Applications* **118:** 1997–2013.

[9] Francq C., Zakoïan J.-M. 2010. Inconsistency of the mle and inference based on weighted ls for larch models. *Journal of Econometrics* **159:** 151–165.

[10] Giraitis L., Leipus R., Robinson P., Surgailis D. 2004. Larch, leverage and long memory. *Journal of Financial Econometrics* **2:** 177–210.

[11] Giraitis L., Robinson P., Surgailis D. 2000. A model for long memory conditional heteroscedasticity. *The Annals of Applied Probability* **10:** 1002 – 1024.

[12] Giraitis, L., Leipus, R., Surgailis, D. 2009. ARCH(∞) Models and Long Memory Properties. In: Mikosch, T., KreiB, JP., Davis, R., Andersen, T. (eds) Handbook of Financial Time Series. Springer, Berlin

[13] Grublyte G, Surgailis D, Skarnulis A. 2017. Quasi-MLE for quadratic ARCH model with long memory. *J. Time Ser. Anal.* **38:** 535–551.

[14] Kounias E., Weng T.-S. 1969. An inequality and almost sure convergence. *Annals of Mathematical Statistics* **48:** 1091–1093.

[15] Ling S. 2007. Self-weighted and local quasi-maximum likelihood estimators for arma-garch/igarch models. *Journal of Econometrics* **140:** 849–873.

[16] Nielsen M.Ø., Noël A.L. 2021. To infinity and beyond: Efficient computation of ARCH(∞) models. *J. Time Ser. Anal.* **42:** 338–354.

[17] Parzen E. 1999. *Stochastic processes*. Classics in Applied Mathematics, SIAM.

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

[19] Robinson P., Zaffaroni P. 1997. Modelling nonlinearity and long memory in time series. *Fields Institute Communications* **11:** 161–170.

[20] Straumann D. 2005. *Estimation in conditionally heteroscedastic time series models*, volume 181 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.

[21] Truquet L. 2014. On a family of contrasts for parametric inference in degenerate arch models. *Econometric Theory* **30:** 1165–1206.