



**HAL**  
open science

# A new semi-parametric estimator for LARCH processes

Jean-Marc Bardet

► To cite this version:

| Jean-Marc Bardet. A new semi-parametric estimator for LARCH processes. 2021. hal-03400395v1

**HAL Id: hal-03400395**

**<https://hal.science/hal-03400395v1>**

Preprint submitted on 25 Oct 2021 (v1), last revised 23 Mar 2023 (v2)

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

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A new semi-parametric estimator for LARCH processes

JEAN-MARC BARDET <sup>1</sup>

<sup>1</sup>*University Paris 1 Panthéon-Sorbonne, SAMM, France, E-mail: [bardet@univ-paris1.fr](mailto:bardet@univ-paris1.fr)*

This paper aims at providing a new semi-parametric estimator for LARCH( $\infty$ ) processes, and therefore also for LARCH( $p$ ) or GLARCH( $p, q$ ) processes. This estimator is obtained from the minimization of a contrast leading to a least squares estimator of the absolute values of the process. The strong consistency and the asymptotic normality are showed, and the convergence happens with rate  $\sqrt{n}$  as well in cases of short or long memory. Numerical experiments confirm the theoretical results, and show that this new estimator clearly outperforms the smoothed quasi-maximum likelihood estimators or the weighted least square estimators often used for such processes.

*MSC 2010 subject classifications:* Primary 62F12; 62M10; secondary 91B84

## 1. Introduction

This paper is devoted to study the asymptotic properties of a new semi-parametric estimator for LARCH( $\infty$ ) processes. Such processes were first defined in [13] and also studied concerning their stationarity and dependence properties in [14], [9], [4] and [10]. LARCH( $\infty$ ) processes, which are conditionally heteroskedastic weak white noises, provide new perspectives for modelling financial data. Indeed, this model has the advantage over GARCH formulations to allow the volatility to be arbitrarily close to 0 and to may be long memory process.

Since this case will also be considered in this paper, we will use a semi-parametric class of LARCH( $\infty$ ) defined in (1) and assume that a trajectory  $(X_1, \dots, X_n)$  is observed, but the parameter  $\theta^*$  that defines the model is unknown (see more details in Section 2). Our goal will be to propose an estimator allowing to estimate  $\theta^*$  (and not a component of  $\theta^*$  such as the location parameter as it was done using M-estimator in [2]).

For numerous affine causal process, such as ARMA, GARCH, ARMA-GARCH, AR( $\infty$ ) or ARCH( $\infty$ ) processes, the Gaussian quasi-maximum likelihood (QML) provides a very accurate estimator (see more details in [1]). Even if a LARCH( $\infty$ ) process or its particular cases LARCH( $p$ ) or GLARCH( $p, q$ ) (see their definitions in (5) and (6)) are also causal affine time series, such a contrast cannot be used as is to estimate the parameter  $\theta^*$ . Indeed, the conditional variance of  $X_t$  can not be bounded close to 0 and this does not allow asymptotic results for such contrasts (see more details on this point in [3], [16] and especially in [8]). Beran and Schützner in [3] and Truquet in [16] proposed an interesting alternative of estimation based on a family of smooth approximations of the QML estimation and they establish the consistency and asymptotic normality of the estimator of  $\theta^*$  in cases of short or long memory. Franck and Zakoian in [8] 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 Truquet.

We propose a new estimator which is obtained by minimizing a least squares contrast of the absolute values of  $(X_t)$  (see its precise definition in (11), Section 3). A strong consistency and asymptotic normality are established for this estimator, under not too restrictive assumptions, which notably allow

to consider as well short and long memory. Moreover, only a fourth order moment of the white noise is required for the asymptotic normality (order 4 in [16] and 5 in [3] for the smoothed QML estimator, and 8 in [8], Assumption A12, for the weighted LS estimator). The convergence happens with rate  $\sqrt{n}$  as well for short and long memory LARCH( $\infty$ ) process, while it is  $n^\beta$  with  $0 < \beta < 1/2$  for the smoothed QMLE defined in [3] in this last case.

Monte-Carlo experiments confirm the asymptotic behavior of the estimator even for trajectories of not very large lengths. The performances of this new estimator are then compared to those obtained with the regularized QMLE (for which the choice of the regularization parameter is a real problem) and to those obtained with the weighted least squares estimator of [8]. The results of these comparisons show without any doubt the much faster convergence of this new estimator, especially compared to the smoothed QMLE.

The forthcoming Section 2 will be devoted to the definition and stationarity conditions of the considered LARCH( $\infty$ ) processes. The main results concerning the definition and the asymptotic behavior of the new estimator are stated in Section 3. Numerical experiments are proposed in Section 4 and proofs are established in Section 5.

## 2. Semi-parametric LARCH( $\infty$ ) processes

Denote  $\|\cdot\|$  the usual Euclidian norm for vectors or matrix and denote  $\|Z\|_p = \mathbb{E}[\|Z\|^p]^{1/p}$  for  $p \geq 1$  where  $Z$  is a random vector valued in  $\mathbb{R}^m$ ,  $m \in \mathbb{N}^*$ . Here we consider a LARCH( $\infty$ ) process introduced in [13] and also studied in [9], [10], [3], [8] or [16], which is defined by:

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

where:

- $(\xi_t)_{t \in \mathbb{Z}}$  is a sequence of symmetric centered independent random variables such as  $\|\xi_0\|_1 = 1$  and  $\|\xi_0\|_r < \infty$  with  $r \geq 2$ ;
- $\theta^* \in \mathbb{R}^d$ , is an unknown vector of parameters but  $d \in \mathbb{N}^*$  is known;
- For any  $j \in \mathbb{N}$ ,  $\theta \in \mathbb{R}^d \mapsto a_j(\theta) \in \mathbb{R}$  are known continuous functions and without lose of generality we will assume  $a_0(\theta) > 0$  for any  $\theta \in \mathbb{R}^d$  (the case  $a_0(\theta) = 0$  implies  $X_t = 0$  for any  $t \in \mathbb{Z}$ , see [9]).

For insuring the stationarity of  $(X_t)$  and the existence of  $\|X_t\|_2$  (see also [9]), assume that  $\theta^* \in \Theta(2)$ , with

$$\Theta(2) = \left\{ \theta \in \mathbb{R}^d, \|\xi_0\|_2^2 \sum_{j=1}^{\infty} a_j^2(\theta) < 1 \right\}. \quad (2)$$

**Remark 2.1.** Note that we assume  $\|\xi_0\|_1 = 1$  and not  $\|\xi_0\|_2 = 1$  as it is usually done. This will be explained by the expression of the estimator we will consider. However, the difference between those normalization choice only consists on a new parametrization, since, with  $\xi'_t = \xi_t / \|\xi_0\|_2$  for any  $t \in \mathbb{Z}$

and using the linearity of its expression, (1) could also be written

$$X_t = \xi_t' \left( a_0'(\theta^*) + \sum_{j=1}^{\infty} a_j'(\theta^*) X_{t-j} \right) \quad \text{where } \|\xi_0'\|_2 = 1 \text{ and } a_j'(\theta^*) = \|\xi_0\|_2 a_j(\theta^*) \text{ for any } j \in \mathbb{N}. \quad (3)$$

For instance, in case of a Gaussian white noise, we have  $\|\xi_0\|_2^2 = \sigma_\xi^2 = \pi/2$  when  $\|\xi_0\|_1 = 1$ .

In the sequel, we will also consider  $\|X_0\|_4$  and for this we define

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

Three interesting special cases of LARCH( $\infty$ ) processes can be mentioned:

1. The first one is composed by LARCH( $p$ ) processes, which are defined by

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

and therefore a LARCH( $p$ ) process is a LARCH( $\infty$ ) process defined in (1) with  $a_k(\theta) = a_k$  for  $0 \leq k \leq p$  and  $\theta = (a_0, a_1, \dots, a_p) \in (0, \infty) \times \mathbb{R}^p$ . For such LARCH( $p$ ), the sets  $\Theta(2)$  and  $\Theta(4)$  are directly deduced from their definitions.

2. A natural extension of LARCH( $p$ ) processes to be considered are GLARCH( $p, q$ ) processes, following the same procedure as the known transition from ARCH processes to GARCH processes. A GLARCH( $p, q$ ) process is defined by

$$X_t = \xi_t \sigma_t \quad \text{with} \quad \sigma_t = c_0 + \sum_{i=1}^p c_i X_{t-i} + \sum_{j=1}^q d_j \sigma_{t-j}, \quad \text{for any } t \in \mathbb{Z}. \quad (6)$$

For studying such a process, define 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 iterative equation (6) is equivalent to  $P(B)\sigma = Q(B)X$  where  $B$  is the usual backward operator. In the sequel we will assume that  $P$  and  $Q$  are coprime polynomials for  $\theta = \theta^*$ .

We define  $\theta = {}^t(c_0, c_1, \dots, d_1, \dots, d_q) \in (0, \infty) \times \mathbb{R}^{p+q}$  and the coefficients  $a_k(\theta)$  exponentially decrease to 0 when  $k \rightarrow \infty$  (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 (6) is  $\theta \in \Theta_{p,q}(2)$ , with

$$\Theta_{p,q}(2) = \left\{ \theta \in (0, \infty) \times \mathbb{R}^{p+q}, \sum_{i=1}^p d_i^2 + \|\xi_0\|_2^2 \sum_{j=1}^p c_j^2 < 1 \right\}. \quad (7)$$

The computation of  $\Theta(4)$  for such GLARCH( $p, q$ ) processes is not straightforward. In [10],  $\Theta(4)$  is simplified for GLARCH(1, 1) and it is established that:

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

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

$$a_j(\theta) = L_\theta(j) j^{d(\theta)-1} \quad \text{for } j \in \mathbb{N}^*. \quad (8)$$

This case has been especially considered in [14] and [3]. In this article, a semi-parametric estimation procedure has been studied when  $a_j(\theta) = c j^{1-d}$  for  $j \in \mathbb{N}^*$  and  $\theta = (a_0, c, d)$  (see more details hereafter). Note also that in such a case  $\sum_{j=1}^{\infty} |a_j(\theta)| = \infty$  but  $\sum_{j=1}^{\infty} a_j^2(\theta) < \infty$  or  $\sum_{j=1}^{\infty} a_j^4(\theta) < \infty$ .

### 3. A new estimator of LARCH parameters

#### 3.1. Definition and consistency of the estimator

We consider here a particular case of M-estimators for estimating  $\theta^*$  from an observed trajectory  $(X_1, \dots, X_n)$  of a stationary solution of (1). For this, let the following contrast function  $\Phi$  be defined for  $x \in \mathbb{R}^\infty$  and  $\theta \in \mathbb{R}^d$  by

$$\Phi(x, \theta) = \left( |x_1| - |a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta) x_{j+1}| \right)^2. \quad (9)$$

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

$$\tilde{X}_t = \begin{cases} X_t & \text{for } t \geq 1 \\ 0 & \text{for } t \leq 0. \end{cases} \quad (10)$$

Then define the following estimator:

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\text{Argmin}} \frac{1}{n} \sum_{t=1}^n \Phi((\tilde{X}_{t-k})_{k \geq 0}, \theta). \quad (11)$$

We add a classical identification condition:

**Assumption Id( $\Theta$ ):** If  $\theta, \theta' \in \Theta$ ,

$$(a_i(\theta) = a_i(\theta') \text{ for all } i \in \mathbb{N}) \implies (\theta = \theta'). \quad (12)$$

Then, we obtain the following conditions of the consistency of both the estimators:

**Proposition 3.1.** Assume that  $(X_t)$  is a stationary solution of (1) with  $\theta^* \in \Theta$  a compact set of  $\Theta(2)$ . Under Assumption Id( $\Theta$ ) and if

$$\sum_{k=1}^{\infty} \log k \times \sup_{\theta \in \Theta} \{a_k^2(\theta)\} < \infty, \quad (13)$$

then  $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta^*$ .

In case of LARCH( $p$ ) or GLARCH( $p, q$ ) process, Assumption  $\text{Id}(\Theta)$  as well as condition (13) are automatically satisfied and therefore:

**Corollary 3.1.** Assume that  $(X_t)$  is a stationary solution of a LARCH( $p$ ) or a GLARCH( $p, q$ ) process, respectively defined in (5) and (6) with  $\theta^* \in \Theta$  a compact set of  $\Theta(2)$ . Then  $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta^*$ .

The consistency of  $\hat{\theta}_n$  for long-memory LARCH( $\infty$ ) can also be deduced from Proposition 3.1, since condition (13) is satisfied in such a case:

**Corollary 3.2.** Assume that  $(X_t)$  is a stationary solution of a LARCH( $\infty$ ) defined in (1) where  $\theta^* \in \Theta$  a compact set of  $\Theta(2)$  and  $(a_j(\theta))$  satisfying (8) with  $0 \leq d(\theta) \leq \bar{d} < 1/2$  for any  $\theta \in \Theta$ . Then under Assumption  $\text{Id}(\Theta)$ ,  $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta^*$ .

**Corollary 3.3 (Example of long memory LARCH( $\infty$ ) studied in [3]).** Let  $a_0(\theta) = a_0$  and  $a_j(\theta) = c j^{d-1}$  for  $j \geq 1$ , set  $\theta = (a_0, c, d)$  with  $0 < \underline{a} \leq a_0 \leq \bar{a} < \infty$ ,  $0 \leq d \leq \bar{d} < 1/2$  and  $|c| \leq \bar{c}$  with  $\bar{c}^2 \|\xi_0\|_2^2 \sum_{i=1}^{\infty} j^{2\bar{d}-2} < 1$ . Then  $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta^*$ .

**Remark 3.1.** The conditions required for Proposition 3.1 and Corollaries 3.1 and 3.3 are weaker to those stated in Theorem 4.2. of [8] for a weighted LS estimator (see its definition in (22)) where  $\|\xi_0\|_4 < \infty$  is required, and in Theorem 4 of [3] for a smoothed QML estimator (see its definition in (21)) where  $\|\xi_0\|_3 < \infty$  is required for  $\mathbb{L}^1$  consistency. However, in [16], the strong consistency of smoothed QML estimator is obtained for LARCH( $p$ ) processes under weaker conditions (there exists  $s > 0$  such as  $\|\xi_0\|_s < \infty$ ).

### 3.2. Asymptotic normality of the estimator

In the sequel, for  $\psi : \theta \in \Theta \subset \mathbb{R}^d \mapsto \psi(\theta) \in \mathbb{R}$  such as  $\psi \in \mathcal{C}^1(\Theta)$ , denote:

$$\partial_{\theta} \psi(\theta) := \left( \frac{\partial}{\partial \theta_i} \psi(\theta) \right)_{1 \leq i \leq d} \quad \text{and} \quad \partial_{\theta^2}^2 \psi(\theta) := \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \psi(\theta) \right)_{1 \leq i, j \leq d}.$$

As it was already mentioned in [3], for establishing the asymptotic normality of  $\hat{\theta}_n$ , it is required to consider the derivatives in  $\theta$  of

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

But the existence of  $M_{\theta}^t$  derivatives could be problematic since the sequence  $(a_k(\theta))$  is not summable in case of long memory. Hence we will consider the following assumption:

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

Note that this assumption is not really restrictive since a stochastic process can always be replaced by a separable version (see Remark 1 in [3]). Then:

**Theorem 3.1.** Assume that  $(X_t)$  is a stationary solution of (1) with  $\theta^* \in \Theta$ , where  $\Theta$  is in the interior of a compact set included in  $\Theta(4)$ . Assume that Assumption S is satisfied and for any  $k \in \mathbb{N}$ , the functions  $a_k \in \mathcal{C}^2(\Theta)$  and such as there exist  $C_a > 0$  and  $\bar{d} < 1/2$  satisfying

$$\sup_{\theta \in \Theta} \left\{ |a_k(\theta)| + \|\partial_\theta a_k(\theta^*)\| \right\} \leq C_a k^{\bar{d}-1} \quad \text{for any } k \in \mathbb{N}^* \quad (15)$$

and such as the following matrix  $\Gamma_1^*$  and  $\Gamma_2^*$  are positive symmetric with

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

$$\text{with } \sigma_X^2 := \mathbb{E}[X_0^2] = \frac{a_0^2(\theta^*) \sigma_\xi^2}{1 - \sigma_\xi^2 \sum_{k=1}^{\infty} a_k^2(\theta^*)}; \quad (16)$$

$$\text{and } \Gamma_2^* := \mathbb{E} \left[ (M_{\theta^*}^0)^2 \partial_\theta M_{\theta^*}^0 \times {}^t \partial_\theta M_{\theta^*}^0 \right]. \quad (17)$$

Then, under Assumption Id( $\Theta$ ),

$$\sqrt{n} (\hat{\theta}_n - \theta^*) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N} \left( 0, (\sigma_\xi^2 - 1) (\Gamma_1^*)^{-1} \Gamma_2^* (\Gamma_1^*)^{-1} \right). \quad (18)$$

The expression of  $\Gamma_2^*$  is not easy to simplify, even in the simplest cases. This is not really a problem, since, as is usual, it is possible to use Slutsky's Lemma, to define the following estimators of  $\Gamma_1^*$  and  $\Gamma_2^*$ :

$$\left\{ \begin{array}{l} \hat{\Gamma}_1 := \frac{1}{n} \sum_{t=1}^n \left( \partial_\theta a_0(\hat{\theta}_n) + \sum_{k=1}^{t-1} \partial_\theta a_k(\hat{\theta}_n) X_{t-k} \right) \left( \partial_\theta a_0(\hat{\theta}_n) + \sum_{k=1}^{t-1} \partial_\theta a_k(\hat{\theta}_n) X_{t-k} \right) \\ \hat{\Gamma}_2 := \frac{1}{n} \sum_{t=1}^n \left( a_0(\hat{\theta}_n) + \sum_{k=1}^{t-1} a_k(\hat{\theta}_n) X_{t-k} \right)^2 \left( \partial_\theta a_0(\hat{\theta}_n) + \sum_{k=1}^{t-1} \partial_\theta a_k(\hat{\theta}_n) X_{t-k} \right) \times \\ \quad \times \left( \partial_\theta a_0(\hat{\theta}_n) + \sum_{k=1}^{t-1} \partial_\theta a_k(\hat{\theta}_n) X_{t-k} \right) \end{array} \right. ,$$

which are consistent estimators of  $\Gamma_1^*$  and  $\Gamma_2^*$  (ergodic theorem),

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

$$\text{with } \hat{\sigma}_\xi^2 := \frac{1}{n} \sum_{t=1}^n \frac{X_t^2}{\left( a_0(\hat{\theta}_n) + \sum_{k=1}^{t-1} a_k(\hat{\theta}_n) X_{t-k} \right)^2}. \quad (19)$$

Note that the consistency of  $\hat{\sigma}_\xi^2$  has been established in [8]. Such asymptotic normality of  $\hat{\theta}_n$  allows the computation of asymptotic confidence intervals or test's thresholds on  $\theta$ .

In case of LARCH( $p$ ) processes, Theorem 3.1 is satisfied under very simple assumptions:

**Corollary 3.4.** Assume that  $(X_t)$  is a stationary solution of a LARCH( $p$ ) defined in (5) with  $0 < a_0^* < \bar{a}$  where  $0 < \bar{a} < \infty$  and  $(a_1^*, \dots, a_p^*)$  such as

$$\|\xi_0\|_4^4 \sum_{j=1}^p a_j^{*4}(\theta) + 6 \|\xi_0\|_2^2 \sum_{j=1}^p a_j^{*2}(\theta) < 1.$$

Then, the central limit theorems (18) and (19) hold. Moreover, for LARCH( $p$ ) processes, considering the usual representation (3) described in Remark 2.1, we obtain:

$$\sqrt{n} (\hat{\sigma}_\xi^4 - \sigma_\xi^2)^{-1/2} (\hat{\Gamma}_1 \hat{\Gamma}_2^{-1} \hat{\Gamma}_1)^{1/2} \left( t(\hat{a}'_0, \dots, \hat{a}'_p) - t(a_0'^*, \dots, a_p'^*) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, I_{p+1}). \quad (20)$$

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

$$\begin{aligned} \Gamma_1^* &= \begin{pmatrix} 1 & 0 \\ 0 & \sigma_X^2 \end{pmatrix}, \quad \Gamma_2^* = \begin{pmatrix} a_0^{*2} + \sigma_X^2 & 2a_0 a_1 \sigma_X^2 \\ 2a_0 a_1 \sigma_X^2 & a_0^2 \sigma_X^2 + \mathbb{E}[X_0^4] \end{pmatrix} \\ &\implies (\Gamma_1^*)^{-1} \Gamma_2^* (\Gamma_1^*)^{-1} = \begin{pmatrix} a_0^{*2} + \sigma_X^2 & 2a_0^* a_1^* \\ 2a_0^* a_1^* & \frac{a_0^{*2}}{\sigma_X^2} + \frac{\mathbb{E}[X_0^4]}{\sigma_X^4} \end{pmatrix}, \end{aligned}$$

$$\text{where } \begin{cases} \sigma_X^2 = a_0^{*2} \sigma_\xi^2 (1 - \sigma_\xi^2 a_1^{*2})^{-1} \\ \mathbb{E}[X_0^4] = a_0^{*4} \mathbb{E}[\xi_0^4] (1 + 5 \sigma_\xi^2 a_1^{*2}) (1 - \sigma_\xi^2 a_1^{*2})^{-1} (1 - \mathbb{E}[\xi_0^4] a_1^{*4})^{-1}. \end{cases}$$

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

**Corollary 3.5.** Assume that  $(X_t)$  is a stationary solution of GLARCH( $p, q$ ) process defined in (6) with  $\theta^* = (c_0^*, c_1^*, \dots, c_p^*, d_1^*, \dots, d_q^*) \in \Theta$  the interior of a compact set of  $\Theta_{p,q}(4)$ . Then the central limit theorems (18) and (19) hold.

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

**Corollary 3.6.** Assume the conditions of Corollary 3.2 with  $\theta^* \in \Theta$ , where  $\Theta$  is in the interior of a compact set included in  $\Theta(4)$ . Assume also that for any  $j \in \mathbb{N}^*$ ,  $\theta \in \mathbb{R}^d \mapsto a_j(\theta) = L_\theta(j) j^{d(\theta)-1}$  as well as  $\theta \in \mathbb{R}^d \mapsto a_0(\theta)$  are  $\mathcal{C}^2(\Theta)$  functions such as (15) is satisfied. Then, under Assumption S and if  $\Gamma_1^*$  and  $\Gamma_2^*$  are positive definite matrix, the central limit theorems (18) and (19) hold.

**Corollary 3.7 (Example of long memory LARCH( $\infty$ ) studied in [3]).** Under the assumptions of Corollary 3.3 and if  $\bar{c}^4 \|\xi_0\|_4^4 \sum_{i=1}^\infty j^{4\bar{d}-4} + 6 \bar{c}^2 \|\xi_0\|_2^2 \sum_{i=1}^\infty j^{2\bar{d}-2} < 1$ , under Assumption S, then the central limit theorems (18) and (19) hold.

**Remark 3.2.** To our knowledge, the only result obtained for the estimation of the memory parameter  $d$  in the case of long memory LARCH( $\infty$ ) processes was obtained in [3] using the smoothed QML estimator. However, the expression of this estimator actually uses only a small part of the trajectory (whose size also depends on the parameter  $d$ !) to account for the strong memory. This leads to a rate of convergence in  $n^\beta$  with  $0 < \beta < 1/2$ , which is much less interesting than the rate in  $\sqrt{n}$  obtained with  $\hat{\theta}_n$ . The Monte-Carlo experiments will confirm these theoretical results and the much better performance of  $\hat{\theta}_n$ .



## 4. Numerical experiments

In this Section, we report the results of Monte-Carlo experiments realized on several LARCH processes. More precisely, 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( $\infty$ ) 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  $d = 0.1, 0.2$ , using the same example studied in [3] for its numerical illustrations.
- 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( $\infty$ ) process (as in [3]);
- Two distributions for  $\xi_0$  such as  $\mathbb{E}[|\xi_0|] = 1$ : a Gaussian  $\mathcal{N}(0, \pi/2)$  distribution denoted  $\mathcal{N}$  and a renormalized Student  $t(6)$  distribution with 6 freedom degrees.

For each choice of process, of length  $n$  and noise distribution, 1000 replications of independent trajectories of the LARCH process are generated.

Two other estimators will be compared to  $\hat{\theta}_n$ :

1. Following [3] and [16], the smooth approximation of the QMLE defined for  $h > 0$  by

$$\hat{\theta}_{QML}(h) := \underset{\theta \in \Theta}{\operatorname{Argmin}} \frac{1}{n} \sum_{t=1}^n \frac{h + X_t^2}{h + (M_\theta^t)^2} + \log(h + (M_\theta^t)^2). \quad (21)$$

A priori choice of  $h$  or data-driven  $\hat{h}$  is not an easy task, even if Truquet in [16] have given some indications. Hence, we will make appear the results obtained for 2 different values of  $h$  that provide the best performances. In case of the considered long memory LARCH( $\infty$ ) process, [3] proposed a modified version of  $\hat{\theta}_{QML}(h)$  and we will use its results.

2. Following [8], the weighted least square estimator defined by:

$$\hat{\theta}_{FZ} := \underset{\theta \in \Theta}{\operatorname{Argmin}} \frac{1}{n} \sum_{t=1}^n \tau_t (X_t^2 - (M_\theta^t)^2)^2, \quad (22)$$

where the weights ( $\tau_t$ ) are obtained for LARCH( $p$ ) or GLARCH( $p, q$ ) using an empirical rule proposed in [12]:

$$\tau_t = \left( \max \left( 1, \frac{1}{C} \sum_{i=1}^p |X_{t-i}| \mathbb{1}_{|X_{t-i}| > C} \right) \right)^{-4},$$

where  $C$  is computed as the 90% quantile of the absolute values ( $|X_1|, \dots, |X_n|$ ). In case of long memory LARCH( $\infty$ ), we replace  $p$  by  $t - 1$  in the definition of  $\tau_t$ .

**Remark 4.1.** *Following Remark 2.1, the comparisons between  $\hat{\theta}_n$ , obtained with parameters defined under the normalization condition  $\|\xi_0\|_1 = 1$ , and  $\hat{\theta}_{QML}(h)$  or  $\hat{\theta}_{FZ}$ , for which the normalization condition is  $\|\xi_0\|_2 = 1$ , require to modified certain estimators. Hence, for LARCH(2) process, we consider  $\|\xi_0\|_2 \hat{\theta}_{QML}(h)$  and  $\|\xi_0\|_2 \hat{\theta}_{FZ}$ , for GLARCH(1, 1) process, the same except for  $d_1$  where*

		$\hat{\theta}_n$			$\hat{\theta}_{FZ}$			$\hat{\theta}_{QML(2)}$			$\hat{\theta}_{QML(1)}$		
$\xi_0$ 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	1.303	0.163	0.181	1.968	0.263	0.249	2.505	0.346	0.267
	500	0.272	0.040	0.061	1.178	0.117	0.145	1.701	0.199	0.239	2.234	0.273	0.249
	1000	0.224	0.029	0.051	1.148	0.092	0.126	1.643	0.193	0.259	2.015	0.220	0.231
	2000	0.124	0.021	0.031	1.129	0.073	0.109	1.604	0.180	0.256	1.965	0.181	0.225
	5000	0.077	0.014	0.021	1.127	0.058	0.100	1.677	0.212	0.311	2.082	0.178	0.245

**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.

		$\hat{\theta}_n$			$\hat{\theta}_{FZ}$			$\hat{\theta}_{QML(1)}$			$\hat{\theta}_{QML(0.5)}$		
$\xi_0$ 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	0.553	0.165	0.207	0.739	0.101	0.184	0.734	0.107	0.185
	500	0.142	0.042	0.081	0.499	0.138	0.125	0.542	0.095	0.162	0.583	0.108	0.158
	1000	0.091	0.029	0.051	0.479	0.121	0.091	0.530	0.094	0.153	0.597	0.114	0.151
	2000	0.064	0.020	0.033	0.466	0.108	0.065	0.515	0.090	0.116	0.611	0.119	0.152
	5000	0.039	0.013	0.022	0.462	0.092	0.059	0.488	0.081	0.107	0.575	0.110	0.120

**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.

$\hat{d}_1^{QML}(h)$  and  $\hat{d}_1^{FZ}$  are considered as they are, and for LARCH( $\infty$ ) process, the same except for  $d$  where  $\hat{d}^{QML}(h)$  and  $\hat{d}^{FZ}$  are considered as they are. When the law of  $\xi_0$  is unknown, the comparison is still possible by using the estimator  $\hat{\sigma}_\xi$  defined in (19).

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

### Conclusions of the Monte-Carlo experiments:

- The convergence of  $\hat{\theta}_n$  happens with a rate  $\sqrt{n}$  for the 3 types of LARCH processes considered and the 2 noise distributions. This is also the case for the estimators  $\hat{\theta}_{QML}$  and  $\hat{\theta}_{FZ}$ , but only in the Gaussian framework, the case of a Student distribution  $t(6)$ , for which the moments of order 6 do not exist, making the convergence much slower.

		$\hat{\theta}_n$			$\hat{\theta}_{FZ}$			$\hat{d}_{QML}$
$d$	$n$	$a_0$	$c$	$d$	$a_0$	$c$	$d$	
$d = 0.1$	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

**Table 3.** Square roots of the MSE computed for estimators  $\hat{\theta}_n$  and  $\hat{\theta}_{FZ}$  of parameters  $a_0 = 1$ ,  $c = 1$  and  $d = 0.1$  or  $0.2$  of the LARCH( $\infty$ ) process computed from 300 independent replications, and for  $\hat{d}_{QML}(h)$  already computed in [3].

- Overall we can say that in the framework of a white Gaussian noise, the  $\hat{\theta}_n$  estimator converges twice as fast as  $\hat{\theta}_{FZ}$  and this for the 3 processes. We notice that even in the case of long memory, while this has not been shown in [8], the  $\hat{\theta}_{FZ}$  estimator provides convincing results, even if they are largely outperformed by those of  $\hat{\theta}_n$ . However, as already mentioned, the asymptotic normality of  $\hat{\theta}_{FZ}$  with rate of convergence  $\sqrt{n}$  is ensured only under the minimal assumption of a moment of order 8 for the white noise, which is illustrated by the poor numerical results performed with a Student's law  $t(6)$ .
- Finally, concerning the  $\hat{\theta}_{QML}$  estimator, it offers satisfactory performances comparable to those of  $\hat{\theta}_n$  only in one case, that of the GARCH(1, 1) with Gaussian noise, and this after the selection of an optimal regularization parameter (which is easily done in Monte-Carlo experiments, but which would otherwise require a data-driven procedure which does not exist at the moment). For the estimator of the long memory parameter  $d$  in the case of a LARCH( $\infty$ ) process, the estimator proposed in [3] has really catastrophic performances compared to those obtained with  $\hat{\theta}_{FZ}$  and especially those of  $\hat{\theta}_n$ .

## 5. Proofs

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

$$I_n(\theta) := \frac{1}{n} \sum_{t=1}^n \Phi((X_{t-k})_{k \geq 0}, \theta) \quad \text{and} \quad \tilde{I}_n(\theta) := \frac{1}{n} \sum_{t=1}^n \Phi((\tilde{X}_{t-k})_{k \geq 0}, \theta). \quad (23)$$

The proof will be stepped in 3 points:

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

$$I(\theta) := \mathbb{E}[\Phi((X_{-k})_{k \geq 0}, \theta)] \quad \text{for } \theta \in \Theta. \quad (24)$$

Indeed, from [7], there exists a function  $H : \mathbb{R}^\infty \rightarrow \mathbb{R}$  such as for any  $t \in \mathbb{Z}$ ,  $X_t = H((\xi_{t-k})_{k \geq 0})$  and therefore  $(X_t)_{t \in \mathbb{Z}}$  is a second order ergodic stationary sequence since  $r = 2$ . Following the same reasoning, we also have  $(\Phi((X_{-k})_{k \geq 0}, \theta))_{t \in \mathbb{Z}}$  that is an ergodic stationary sequence for any  $\theta \in \Theta$ ,

with  $\mathbb{E}[\Phi((X_{-k})_{k \geq 0}, \theta)] < \infty$  for any  $\theta \in \Theta$ . As a consequence, for any  $\theta \in \Theta$ ,

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

Now, using Theorem 2.2.1. in [15], we deduce that the previous ergodic theorem is also a uniform ergodic theorem and we obtain  $\sup_{\theta \in \Theta} |I_n(\theta) - I(\theta)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$ .

2. We also have  $\sup_{\theta \in \Theta} |I_n(\theta) - \tilde{I}_n(\theta)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$ . For establishing this result, first set:

$$M_\theta^t := a_0(\theta) + \sum_{j=1}^{\infty} a_j(\theta) X_{t-j} \quad \text{and} \quad \tilde{M}_\theta^t := a_0(\theta) + \sum_{j=1}^{t-1} a_j(\theta) X_{t-j}. \quad (25)$$

Then,

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

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

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

Therefore, using Cauchy-Schwarz and Minkowski inequalities, we obtain:

$$\begin{aligned} &\mathbb{E} \left[ \sup_{\theta \in \Theta} |\Phi((X_{t-k})_{k \geq 0}, \theta) - \Phi((\tilde{X}_{t-k})_{k \geq 0}, \theta)| \right] \\ &\leq \mathbb{E} \left[ \sup_{\theta \in \Theta} \left\{ |M_\theta^t - \tilde{M}_\theta^t| \right\} \left( 2\|X_t\| + \sup_{\theta \in \Theta} \left\{ |M_\theta^t| + |\tilde{M}_\theta^t| \right\} \right) \right] \\ &\leq \left( \mathbb{E} \left[ \sup_{\theta \in \Theta} \left\{ (M_\theta^t - \tilde{M}_\theta^t)^2 \right\} \right] \right)^{1/2} \left( 2\|X_t\|_2 + \left( \mathbb{E} \left[ \sup_{\theta \in \Theta} |M_\theta^t|^2 \right] \right)^{1/2} + \left( \mathbb{E} \left[ \sup_{\theta \in \Theta} |\tilde{M}_\theta^t|^2 \right] \right)^{1/2} \right). \end{aligned}$$

Now, from [9], we know that since  $\Theta$  is a compact set included in  $\Theta(2)$ , then there exists  $C > 0$  such as  $\left( 2\|X_t\|_2 + \left( \mathbb{E} \left[ \sup_{\theta \in \Theta} |M_\theta^t|^2 \right] \right)^{1/2} + \left( \mathbb{E} \left[ \sup_{\theta \in \Theta} |\tilde{M}_\theta^t|^2 \right] \right)^{1/2} \right) \leq C$ . Moreover, from the same reasoning than in Lemma 2, b/ of [3], there exists  $C' > 0$  such as

$$\mathbb{E} \left[ \sup_{\theta \in \Theta} \left\{ (M_\theta^t - \tilde{M}_\theta^t)^2 \right\} \right] \leq C' \sum_{j=t}^{\infty} \sup_{\theta \in \Theta} \{a_j^2(\theta)\}.$$

As a consequence, we deduce there exists  $C > 0$  such as:

$$\mathbb{E} \left[ \sup_{\theta \in \Theta} |\Phi((X_{t-k})_{k \geq 0}, \theta) - \Phi((\tilde{X}_{t-k})_{k \geq 0}, \theta)| \right] \leq C \sum_{j=t}^{\infty} \sup_{\theta \in \Theta} \{a_j^2(\theta)\}.$$

Then, using condition (13),

$$\begin{aligned} & \sum_{t=1}^n \frac{1}{t} \mathbb{E} \left[ \sup_{\theta \in \Theta} \left| \Phi((X_{t-k})_{k \geq 0}, \theta) - \Phi((\tilde{X}_{t-k})_{k \geq 0}, \theta) \right| \right] \\ & \leq C \sum_{t=1}^n \sup_{\theta \in \Theta} \{a_t^2(\theta)\} \left( \sum_{j=1}^t \frac{1}{j} \right) < \infty. \end{aligned}$$

This induces  $\sup_{\theta \in \Theta} |I_n(\theta) - \tilde{I}_n(\theta)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$  from Corollary 1 of [11], where it is established that

$$\sum_{t=1}^{\infty} \frac{\mathbb{E}[|Z_t|]}{b_t} < \infty \text{ implies } \frac{1}{b_n} \sum_{t=1}^n Z_t \xrightarrow[n \rightarrow \infty]{a.s.} 0 \text{ for a } \mathbb{L}^1 \text{ sequence of r.v. } (Z_t)_t.$$

3. The two previous points show us that  $\sup_{\theta \in \Theta} |\tilde{I}_n(\theta) - I(\theta)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$  with  $I$  defined in (24). The proof is achieved if we establish that  $\theta^*$  is the unique minimum of  $\theta \in \Theta \mapsto I(\theta)$ . This is induced by the following computations:

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

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

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

As we assumed that  $a_0(\cdot)$  is a positive function, using also Assumption Id( $\Theta$ ), we deduce that  $\theta = \theta^*$  is the only solution of the previous equality. As a consequence,  $\theta^*$  is the unique minimizer of  $I(\cdot)$  and since  $\sup_{\theta \in \Theta} |\tilde{I}_n(\theta) - I(\theta)| \xrightarrow[n \rightarrow \infty]{a.s.} 0$  and  $\hat{\theta}_n = \hat{\theta}_n = \underset{\theta \in \Theta}{\text{Argmin}} \tilde{I}_n(\theta)$ , we deduce that  $\hat{\theta}_n \xrightarrow[n \rightarrow \infty]{a.s.} \theta^*$ .  $\square$

**Proof of Corollary 3.1.** In case of LARCH( $p$ ) process,  $a_i(\theta) = a_i$  for  $0 \leq i \leq p$  and therefore (12) and Assumption Id( $\Theta$ ) are obviously satisfied.

For a GLARCH( $p, q$ ) process, set  $\theta_P = {}^t(c_0, c_1, \dots, c_p) \in (0, \infty) \times \mathbb{R}^p$ ,  $\theta_Q = {}^t(d_1, \dots, d_q) \in \mathbb{R}^q$  and therefore  $\theta = (\theta_P, \theta_Q)$ . Then  $\sigma = P_{\theta_P}^{-1}(B) Q_{\theta_Q}(B) X$ . It is clear that  $\theta_P \rightarrow P_{\theta_P}$  is an injective function, and it is the same for  $\theta_P \rightarrow P_{\theta_P}^{-1}$  and  $\theta_Q \rightarrow Q_{\theta_Q}$ . Finally it is also the same for  $\theta = {}^t(\theta_P, \theta_Q) \rightarrow P_{\theta_P}^{-1} \times Q_{\theta_Q}$ , because  $P_{\theta_P}$  and  $Q_{\theta_Q}$  are not zero polynomial and because  $\theta_P$  and

$\theta_Q$  have not common component. As a consequence, (12) and therefore Assumption Id( $\Theta$ ) are also satisfied for GLARCH( $p, q$ ) process.

Moreover, for any GLARCH( $p, q$ ) process,  $(a_k(\theta))_k$  satisfies  $\sup_{\theta \in \Theta} \sum_{k=0}^{\infty} |a_k(\theta)| < \infty$ . Therefore (13) is verified for such a process.  $\square$

**Proof of Corollary 3.2.** The only required proof concerns (8), which is obviously satisfied.  $\square$

**Proof of Corollary 3.3.** In this framework, Assumption Id( $\Theta$ ) and (8) are obviously satisfied.  $\square$

**Proof of Theorem 3.1.** Let  $I_n(\theta)$  and  $\tilde{I}_n(\theta)$  be defined in (23). We follow a proof that is similar to the one of Theorem 2 in [6].

Let  $v = \sqrt{n}(\theta - \theta^*) \in \mathbb{R}^d$  and define

$$W_n(v) = \sum_{t=1}^n \Phi((X_{t-k})_{k \geq 0}, \theta^* + n^{-1/2}v) - \Phi((X_{t-k})_{k \geq 0}, \theta^*) = n(I_n(\theta) - I_n(\theta^*))$$

$$\text{and } \tilde{W}_n(v) = \sum_{t=1}^n \Phi((\tilde{X}_{t-k})_{k \geq 0}, \theta^* + n^{-1/2}v) - \Phi((\tilde{X}_{t-k})_{k \geq 0}, \theta^*) = n(I_n(\theta) - \tilde{I}_n(\theta^*)).$$

Then we are going to prove first that minimizing  $\tilde{I}_n(\theta)$  is equivalent to minimizing  $\tilde{W}_n(v)$ , which is equivalent to minimizing  $W_n(v)$  with respect to  $v$ . As a consequence, there exists a sequence  $(\hat{v}_n)_n$  where  $\hat{v}_{LAV}$  is a minimizer of  $W_n(v)$  such as  $\hat{v}_n = \sqrt{n}(\hat{\theta}_n - \theta^*)$ . Secondly, we will provide a limit theorem satisfied by  $W_n(v)$ . Then we are going to prove in 3/ that  $(W_n(\cdot))_n$  converges as a process of  $\mathcal{C}(\mathbb{R}^d)$  (space of continuous functions on  $\mathbb{R}^d$ ) to a limit process  $W$ . Hence  $(\hat{v}_n)_n$  converges to the minimizer of  $\tilde{W}_n$ .

1. For  $v \in \mathbb{R}$ , we have for  $n$  large enough and using a Taylor-Lagrange expansion,

$$\begin{aligned} W_n(v) &= \sum_{t=1}^n \left( |X_t| - |M^t(\theta^* + n^{-1/2}v)| \right)^2 - \left( |X_t| - |M^t(\theta^*)| \right)^2 \\ &= \sum_{t=1}^n \left( |X_t| - \left| M^t(\theta^*) - \frac{1}{\sqrt{n}} {}^t v \partial_{\theta} M^t(\bar{\theta}_t^{(n)}) \right| \right)^2 - \left( |X_t| - |M^t(\theta^*)| \right)^2 \\ &= \sum_{t=1}^n \left( |X_t| - |M^t(\theta^*)| - \frac{1}{\sqrt{n}} {}^t v \partial_{\theta} M^t(\bar{\theta}_t^{(n)}) \times \text{sgn}(M^t(\theta^*)) \right)^2 - \left( |X_t| - |M^t(\theta^*)| \right)^2 \\ &= -\frac{2}{\sqrt{n}} \sum_{t=1}^n \left( |X_t| - |M^t(\theta^*)| \right) \times \text{sgn}(M^t(\theta^*)) \times {}^t v \partial_{\theta} M^t(\bar{\theta}_t^{(n)}) + \frac{1}{n} \sum_{t=1}^n \left( {}^t v \partial_{\theta} M^t(\bar{\theta}_t^{(n)}) \right)^2 \\ &= J_1^{(n)}(v) + J_2^{(n)}(v) \end{aligned} \tag{26}$$

with  $\bar{\theta}_t^{(n)} = \alpha_t^{(n)} \theta^* + (1 - \alpha_t^{(n)}) (\theta^* + n^{-1/2}v)$ . Therefore  $\bar{\theta}_t^{(n)} \xrightarrow[n \rightarrow \infty]{a.s.} \theta^*$  and then for any

$$\left| \partial_{\theta} M^t(\bar{\theta}_t^{(n)}) - \partial_{\theta} M^t(\bar{\theta}^*) \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0 \quad \text{for any } t \in \mathbb{N}, \tag{27}$$

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  $v \in \mathbb{R}$ :

$$\begin{aligned} \left| J_2^{(n)}(v) - \mathbb{E}[({}^t v \partial_\theta M^0(\theta^*))^2] \right| &\leq \frac{1}{n} \sum_{t=1}^n \left| ({}^t v \partial_\theta M^t(\theta^*))^2 - ({}^t v \partial_\theta M^t(\bar{\theta}_t^{(n)}))^2 \right| \\ &\quad + \left| \frac{1}{n} \sum_{t=1}^n ({}^t v \partial_\theta M^t(\theta^*))^2 - \mathbb{E}[({}^t v \partial_\theta M^0(\theta^*))^2] \right|. \end{aligned}$$

Now using Cesaro Lemma we obtain from (27),

$$\left| \frac{1}{n} \sum_{t=1}^n ({}^t v \partial_\theta M^t(\theta^*))^2 - \mathbb{E}[({}^t v \partial_\theta M^0(\theta^*))^2] \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (28)$$

Moreover, using the Ergodic Theorem applied to the stationary process  $\left( ({}^t v \partial_\theta M^t(\theta^*))^2 \right)_t$ , we also have:

$$\frac{1}{n} \sum_{t=1}^n \left| ({}^t v \partial_\theta M^t(\theta^*))^2 - \mathbb{E}[({}^t v \partial_\theta M^0(\theta^*))^2] \right| \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (29)$$

Finally, with (28) and (29), we obtain for any  $v \in \mathbb{R}$ ,

$$J_2^{(n)}(v) \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E}[({}^t v \partial_\theta M^0(\theta^*))^2] = {}^t v \Gamma_1^* v \quad \text{where} \quad \Gamma_1^* := \mathbb{E}[\partial_\theta M^0(\theta^*) \times {}^t \partial_\theta M^0(\theta^*)], \quad (30)$$

where the formula of  $\Gamma_1^*$  is made more explicit in (16). Now, we also have

$$\begin{aligned} J_1^{(n)}(v) &= -\frac{2}{\sqrt{n}} \sum_{t=1}^n (|M^t(\theta^*) \xi_t| - |M^t(\theta^*)|) \times \text{sgn}(M^t(\theta^*)) \times {}^t v \partial_\theta M^t(\bar{\theta}_t^{(n)}) \\ &= {}^t v \left( -\frac{2}{\sqrt{n}} \sum_{t=1}^n (|\xi_t| - 1) M^t(\theta^*) \times \partial_\theta M^t(\theta^*) \right. \\ &\quad \left. + \frac{2}{n} \sum_{t=1}^n (|\xi_t| - 1) M^t(\theta^*) \times \sqrt{n} (\partial_\theta M^t(\theta^*) - \partial_\theta M^t(\bar{\theta}_t^{(n)})) \right) \\ &= {}^t v (K_1^{(n)}(v) + K_2^{(n)}(v)). \end{aligned}$$

We have  $\left( (|\xi_t| - 1) M^t(\theta^*) \times \partial_\theta M^t(\theta^*) \right)_{t \in \mathbb{N}}$  that is martingale increments process since with the  $\sigma$ -algebra  $\mathcal{F}_t = \sigma\{(X_{t-k})_{k \geq 1}\}$ ,

$$\mathbb{E} \left[ (|\xi_t| - 1) M^t(\theta^*) \times \partial_\theta M^t(\theta^*) \mid \mathcal{F}_t \right] = \mathbb{E}[|\xi_t| - 1] \mathbb{E} \left[ M^t(\theta^*) \times \partial_\theta M^t(\theta^*) \right] = 0,$$

because  $(X_t)$  is a causal process and  $\xi_t$  is independent to  $\mathcal{F}_t$  and  $\mathbb{E}[|\xi_0|] = 1$ .

Now since  $\Gamma_2^* := \mathbb{E} \left[ (M^0(\theta^*))^2 \partial_\theta M^0(\theta^*) \times {}^t \partial_\theta M^0(\theta^*) \right]$  is supposed to be a finite definite positive

matrix (see also its expression in (17)),

$$\mathbb{E}\left[\left(|\xi_0| - 1\right)^2 \left\|M^0(\theta^*) \times \partial_\theta M^0(\theta^*)\right\|^2\right] = (\sigma_\xi^2 - 1) \mathbb{E}\left[\left\|M^0(\theta^*) \times \partial_\theta M^0(\theta^*)\right\|^2\right] < \infty.$$

Then a central limit for increment martingales can be applied (see for instance [5]) and we obtain for any  $v \in \mathbb{R}$ :

$$K_1^{(n)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} K_1 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, 4(\sigma_\xi^2 - 1)\Gamma_2^*). \quad (31)$$

Using a Taylor expansion, we also have

$$\mathbb{E}\left[\left(\sqrt{n}{}^t v (\partial_\theta M^t(\theta^*) - \partial_\theta M^t(\bar{\theta}_t^{(n)}))\right)^2\right] \leq \mathbb{E}\left[\left(\sup_{\theta \in \Theta} \|\partial_{\theta^2} M^t(\theta)\|\right)^2\right] \times \|v\|^2 < \infty.$$

Moreover,  $(\partial_\theta M^t(\theta^*) - \partial_\theta M^t(\bar{\theta}_t^{(n)})) \in \mathcal{F}_t$  and from the previous bound,

$$\begin{aligned} & \mathbb{E}\left[\left(|\xi_t| - 1\right) M^t(\theta^*) \times \sqrt{n}{}^t v (\partial_\theta M^t(\theta^*) - \partial_\theta M^t(\bar{\theta}_t^{(n)}))\right] \\ &= \mathbb{E}\left[|\xi_t| - 1\right] \times \mathbb{E}\left[\left(M^t(\theta^*) \times \sqrt{n}{}^t v (\partial_\theta M^t(\theta^*) - \partial_\theta M^t(\bar{\theta}_t^{(n)}))\right)\right] \\ &\leq \mathbb{E}\left[|\xi_t| - 1\right] \times \mathbb{E}\left[\left(M^t(\theta^*)\right)^2\right] \times \mathbb{E}\left[\left(\sqrt{n}{}^t v (\partial_\theta M^t(\theta^*) - \partial_\theta M^t(\bar{\theta}_t^{(n)}))\right)^2\right] < \infty. \end{aligned}$$

Therefore the ergodic theorem for causal stationary process can be applied and we obtain for any  $v \in \mathbb{R}$ ,

$${}^t v K_2^{(n)}(v) \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E}\left[\left(|\xi_t| - 1\right) M^t(\theta^*) \times \sqrt{n}{}^t v (\partial_\theta M^t(\theta^*) - \partial_\theta M^t(\bar{\theta}_t^{(n)}))\right] = 0. \quad (32)$$

Finally, for any  $v \in \mathbb{R}$ , since  $J_1^{(n)}(v) = {}^t v (K_1^{(n)}(v) + K_2^{(n)}(v))$ , then  $J_1^{(n)}(v) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} {}^t v K_1$  from (31) and (32), and with (30) this implies,

$$W_n(v) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} {}^t v \Gamma_1^* v + {}^t v K_1 \quad \text{with} \quad K_1 \overset{\mathcal{L}}{\sim} \mathcal{N}(0, 4(\sigma_\xi^2 - 1)\Gamma_2^*). \quad (33)$$

2/ Asymptotically, from part 1/, we know that the law of  $W_n(v)$  is the same as the law of:

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

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

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

Therefore we obtain:

$$W_n'(v) - \widetilde{W}_n'(v) = -\frac{2}{\sqrt{n}} \sum_{t=1}^n (|\xi_t| - 1) {}^t v \left( M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*) \right)$$



$$+{}^t v \left( \frac{1}{n} \sum_{t=1}^n (\partial_\theta M^t(\theta^*){}^t \partial_\theta M^t(\theta^*) - \partial_\theta \widetilde{M}^t(\theta^*){}^t \partial_\theta \widetilde{M}^t(\theta^*)) \right) v. \quad (34)$$

Now, from their definitions, and since (15) holds, we have:

$$\begin{aligned} \mathbb{E}[(M^t(\theta^*) - \widetilde{M}^t(\theta^*))^2] &= \sigma_X^2 \sum_{j=t}^{\infty} a_j^2(\theta^*) \\ \mathbb{E}[\|\partial_\theta M^t(\theta^*) - \partial_\theta \widetilde{M}^t(\theta^*)\|^2] &= \sigma_X^2 \sum_{j=t}^{\infty} \|\partial_\theta a_j(\theta^*)\|^2, \end{aligned}$$

with  $\sigma_X^2 = \mathbb{E}[X_0^2]$  defined in (16). This implies:

$$\begin{aligned} M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*) &= M^t(\theta^*) \left( \partial_\theta M^t(\theta^*) - \partial_\theta \widetilde{M}^t(\theta^*) \right) \\ &\quad + \partial_\theta \widetilde{M}^t(\theta^*) \left( M^t(\theta^*) - \widetilde{M}^t(\theta^*) \right) \\ \Rightarrow \quad \|M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*)\|_2 &\leq \sigma_X \left( \|M^0(\theta^*)\|_2 \left( \sum_{j=t}^{\infty} \|\partial_\theta a_j(\theta^*)\|^2 \right)^{1/2} \right. \\ &\quad \left. + \|\partial_\theta M^0(\theta^*)\|_2 \left( \sum_{j=t}^{\infty} a_j^2(\theta^*) \right)^{1/2} \right) \\ &\leq C \left( \left( \sum_{j=t}^{\infty} \|\partial_\theta a_j(\theta^*)\|^2 \right)^{1/2} + \left( \sum_{j=t}^{\infty} a_j^2(\theta^*) \right)^{1/2} \right), \quad (35) \end{aligned}$$

with a constant  $C > 0$  and using Cauchy-Schwarz and Minkowski inequalities. Then, using the causality of  $(X_t)$ , *i.e.*  $\xi_t$  independent to  $\sigma\{X_{t-1}, X_{t-2}, \dots\}$  for any  $t \in \mathbb{Z}$ , we deduce that:

$$\mathbb{E} \left[ \frac{1}{\sqrt{n}} \sum_{t=1}^n (|\xi_t| - 1) {}^t v \left( M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*) \right) \right] = 0,$$

since  $\mathbb{E}[|\xi_t|] = 1$ , and with (35),

$$\begin{aligned} &\mathbb{E} \left[ \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n (|\xi_t| - 1) {}^t v \left( M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*) \right) \right)^2 \right] \\ &= \frac{1}{n} (\sigma_\xi^2 - 1) \|v\|^2 \sum_{t=1}^n \|M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*)\|_2^2 \\ &\leq \frac{2C^2}{n} (\sigma_\xi^2 - 1) \|v\|^2 \sum_{t=1}^n \sum_{j=t}^{\infty} (\|\partial_\theta a_j(\theta^*)\|^2 + a_j^2(\theta^*)) \\ &\leq \frac{2C^2}{n} (\sigma_\xi^2 - 1) \|v\|^2 \sum_{t=1}^n \sum_{j=t}^{\infty} C_a j^{2\bar{d}-2} \end{aligned}$$

$$\leq \frac{2C^2}{n} (\sigma_\xi^2 - 1) \|v\|^2 \frac{1}{1-2\bar{d}} \sum_{t=1}^n C_a t^{2\bar{d}-1} \leq C' \|v\|^2 n^{2\bar{d}-1},$$

using classical comparisons between sums and integrals. Therefore, for any  $v \in \mathbb{R}$ ,

$$\mathbb{E} \left[ \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n (|\xi_t| - 1)^t v (M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*)) \right)^2 \right] \xrightarrow{n \rightarrow \infty} 0.$$

Using Bienaymé-Tchebychev Inequality, this implies

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n (|\xi_t| - 1)^t v (M^t(\theta^*) \partial_\theta M^t(\theta^*) - \widetilde{M}^t(\theta^*) \partial_\theta \widetilde{M}^t(\theta^*)) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0. \quad (36)$$

Using the same method, we also obtain that there exist  $C' > 0$  and  $C'' > 0$  such as:

$$\|\partial_\theta M^t(\theta^*)^t \partial_\theta M^t(\theta^*) - \partial_\theta \widetilde{M}^t(\theta^*)^t \partial_\theta \widetilde{M}^t(\theta^*)\|_2 \leq C' \left( \sum_{j=t}^{\infty} \|\partial_\theta a_j(\theta^*)\|^2 \right)^{1/2} \leq C'' t^{\bar{d}-1/2}. \quad (37)$$

Now with (37), we can use again the result established in part 2/ of the proof of Proposition 3.1 based on the Corollary 1 of [11]:

$$\sum_{i=1}^n \frac{1}{t} \|\partial_\theta M^t(\theta^*)^t \partial_\theta M^t(\theta^*) - \partial_\theta \widetilde{M}^t(\theta^*)^t \partial_\theta \widetilde{M}^t(\theta^*)\|_1 \leq C'' \sum_{i=1}^n t^{\bar{d}-3/2} < \infty$$

since  $\bar{d} < 1/2$  and therefore:

$$\frac{1}{n} \sum_{t=1}^n (\partial_\theta M^t(\theta^*)^t \partial_\theta M^t(\theta^*) - \partial_\theta \widetilde{M}^t(\theta^*)^t \partial_\theta \widetilde{M}^t(\theta^*)) \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (38)$$

Finally from (34), (36) and (38), we deduce that for any  $v \in \mathbb{R}$ ,

$$\begin{aligned} & |W_n(v) - \widetilde{W}_n(v)| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0 \\ \implies & \widetilde{W}_n(v) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} W(v) := {}^t v \Gamma_1^* v + {}^t v K_1 \quad \text{with} \quad K_1 \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, 4(\sigma_\xi^2 - 1) \Gamma_2^*). \end{aligned} \quad (39)$$

3/ Now, using the same arguments than in the proof of Theorem 2 of [6], we deduce that finite distributions  $(\widetilde{W}_n(v_1), \dots, \widetilde{W}_n(v_k))$  converge to  $(W(v_1), \dots, W(v_k))$  for any  $(v_1, \dots, v_k) \in (\mathbb{R}^d)^k$ . Moreover, always following the proof of Theorem 2 of [6],  $(W_n(v))_v$  converges to  $(W(v))_v$  as a process on the continuous function space  $\mathcal{C}^0(\mathbb{R})$ .

As a consequence, a maximum  $\widehat{v} = \sqrt{n}(\widehat{\theta}_n - \theta^*)$  of  $\widehat{W}_n(v)$  converges in distribution to the maximum of  ${}^t v \Gamma_1^* v + {}^t v K_1$ , which is  $\bar{v} := -\frac{1}{2} (\Gamma_1^*)^{-1} K_1 \stackrel{\mathcal{L}}{\sim} \mathcal{N}(0, (\sigma_\xi^2 - 1) (\Gamma_1^*)^{-1} \Gamma_2^* (\Gamma_1^*)^{-1})$  and this implies (18).  $\square$

**Proof of Corollary 3.4.** See the proof of Corollary 3.5 for the first part of the corollary (Assumption Id( $\Theta$ ) and Assumption S are satisfied, and  $\Gamma_1^*$  and  $\Gamma_2^*$  are positive definite matrix).

For establishing (20), we use  $a'_j = \sigma_\xi a_j$  for  $0 \leq j \leq p$  detailed in Remark 2.1.  $\square$

**Proof of Corollary 3.5.** For any GLARCH( $p, q$ ) process,  $(a_k(\theta))_k$  satisfies  $\sup_{\theta \in \Theta} \sum_{k=0}^{\infty} |a_k(\theta)| < \infty$ . Therefore Assumption S is automatically verified for such a process.

For any GLARCH( $p, q$ ) process, the matrix  $\Gamma_1^*$  and  $\Gamma_2^*$  are positive definite matrix. Indeed, following the same reasoning as in the proof of Lemma 5 of [3], we have for any  $v \in \mathbb{R}^{p+q+1}$

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

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 (6), we deduce that:

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

Then:

$$\begin{cases} \partial_{c_0} M_{\theta^*}^0 = 1 + d_1 \partial_{c_0} M_{\theta^*}^{-1} + \dots + d_q \partial_{c_0} M_{\theta^*}^{-q} \\ \partial_{c_i} M_{\theta^*}^0 = X_{-i} + d_1 \partial_{c_i} M_{\theta^*}^{-1} + \dots + d_q \partial_{c_i} M_{\theta^*}^{-q} & 1 \leq i \leq p \\ \partial_{d_j} M_{\theta^*}^0 = M_{\theta^*}^{-j} + d_1 \partial_{d_j} M_{\theta^*}^{-1} + \dots + d_q \partial_{d_j} M_{\theta^*}^{-q} & 1 \leq j \leq q \end{cases}$$

Therefore, if  ${}^t v \partial_{\theta} M_{\theta^*}^0 = 0$ , then  $(1, X_{-1}, \dots, X_{-p}, M_{\theta^*}^{-1}, \dots, M_{\theta^*}^{-q}) v = 0$ . Such equation has no solution since a linear relationship (6) 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  $v = 0$ :  $\Gamma_1^*$  is a positive definite matrix (and a similar reasoning leads to the same property satisfied by  $\Gamma_2^*$ ).  $\square$

**Proof of Corollary 3.6.** All the assumptions required for satisfying Theorem 3.1 are well stated.  $\square$

**Proof of Corollary 3.7.** For  $k \geq 1$ , we have  $\partial_{\theta} a_k(\theta) = {}^t (0, k^{d-1}, c \log(k) k^{d-1})$  and therefore we obtain  $\sum_{k=1}^{\infty} \|\partial_{\theta} a_k(\theta)\|^2 < \infty$  as it is required in (15).

Moreover, *mutatis mutandis*, we can use again the proof of Lemma 5 of [3] for proving that  $\Gamma_1^*$  and  $\Gamma_2^*$  are two positive definite matrix in this case of long memory LARCH( $\infty$ ) process.  $\square$

## References

- [1] J.-M. Bardet and O. Wintenberger. Asymptotic normality of the quasi-maximum likelihood estimator for multidimensional causal processes. *Ann. Statist.*, 37:2730–2759, 2009.
- [2] J. Beran. On location estimation for larch processes. *Journal of Multivariate Analysis*, 97:1766–1782, 2006.
- [3] J. Beran and M. Schützner. On approximate pseudo-maximum likelihood estimation for LARCH-processes. *Bernoulli*, 15(4):1057–1081, 2009.
- [4] I. Berkes and L. Horváth. Asymptotic results for long memory larch sequences. *Annals of Applied Probability*, 13:641–668, 2003.
- [5] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, 1968.
- [6] R. Davis and W. Dunsmuir. Least absolute deviation estimation for regression with arma errors. *Journal of Theoretical Probability*, 10:481–497, 1997.
- [7] P. Doukhan and O. Wintenberger. Weakly dependent chains with infinite memory. *Stochastic Processes and Applications*, 118:1997–2013, 2008.
- [8] C. Francq and J.-M. Zakoïan. Inconsistency of the mle and inference based on weighted ls for larch models. *Journal of Econometrics*, 159:151–165, 2010.

- [9] L. Giraitis, P.M. Robinson, and D. Surgailis. A model for long memory conditional heteroscedasticity. *The Annals of Applied Probability*, 10(3):1002 – 1024, 2000.
- [10] L. Giraitis, R. Leipus, P.M. Robinson, and D. Surgailis. Larch, leverage and long memory. *Journal of Financial Econometrics*, 2(2):177–210, 2004.
- [11] E.G. Kounias and T.-S. Weng. An inequality and almost sure convergence. *Annals of Mathematical Statistics*, 48:1091–1093, 1969.
- [12] S. Ling. Self-weighted and local quasi-maximum likelihood estimators for arma-garch/igarch models. *Journal of Econometrics*, 140:849–873, 2007.
- [13] P.M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47(1):67–84, 1991.
- [14] P.M. Robinson and P. Zaffaroni. Modelling nonlinearity and long memory in time series. *Fields Institute Communications*, 11:161–170, 1997.
- [15] D. Straumann. *Estimation in conditionally heteroscedastic time series models*, volume 181 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, 2005.
- [16] L. Truquet. On a family of contrasts for parametric inference in degenerate arch models. *Econometric Theory*, 30(6):1165–1206, 2014. .