



HAL
open science

Generalized Gaussian quasi-maximum likelihood estimation for most common time series

Yakoub Boularouk, Jean-Marc Bardet

► **To cite this version:**

Yakoub Boularouk, Jean-Marc Bardet. Generalized Gaussian quasi-maximum likelihood estimation for most common time series. *Communications in Statistics - Theory and Methods*, 2022, 10.1080/03610926.2022.2103148 . hal-02902614

HAL Id: hal-02902614

<https://hal.science/hal-02902614v1>

Submitted on 20 Jul 2020

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.

Generalized Gaussian quasi-maximum likelihood estimation for most common time series

Yakoub Boularouk^a, Jean-Marc Bardet^b

^a*Institute of Science and Technology, Melilab laboratory, University of Mila, Algeria.*

^b*S.A.M.M., Université Panthéon-Sorbonne, 90, rue de Tolbiac, 75634, Paris, France.*

Abstract

We propose a consistent estimator for the parameter shape of the generalized gaussian noise in the class of causal time series including ARMA, AR(∞), GARCH, ARCH(∞), ARMA-GARCH, APARCH, ARMA-APARCH, ..., processes. As well we prove the consistency and the asymptotic normality of the Generalized Gaussian Quasi-Maximum Likelihood Estimator (GGQMLE) for this class of causal time series with any fixed parameter shape, which over-performs the efficiency of the classical Gaussian QMLE.

Keywords:

Primary MSC: 62M10 62M10, secondary 60G10

Quasi maximum likelihood, Efficiency of estimators, Strong consistency, Asymptotic normality, ARMA-ARCH processes.

1. Introduction

This paper is first devoted to estimate the parameter shape r_0 of the white noise density from an observed trajectory of an affine causal process. This class of time series was already defined and studied in [Duchesnes and Francq \(2008\)](#), [Bardet and Wintenberger \(2009\)](#) and [Bardet et al. \(2017\)](#). Hence, we will consider an observed sample (X_1, \dots, X_n) where $(X_t)_{t \in \mathbb{Z}}$ is a solution of the following equation:

$$X_t = M_{\theta^{0*}}((X_{t-k})_{k \geq 1}) \zeta_t + f_{\gamma^{0*}}((X_{t-k})_{k \geq 1}), \quad t \in \mathbb{Z}, \quad (1)$$

where

- $\theta^{0*} \in \Theta \subset \mathbb{R}^d$, $d \in \mathbb{N}^*$ and $\gamma^{0*} \in \Gamma \subset \mathbb{R}^e$, $e \in \mathbb{N}^*$ are two unknown vectors of parameters, also called the "true" parameters (but d and e are known);
- $(\zeta_t)_{t \in \mathbb{Z}}$ is a sequence of centred independent identically distributed random variables (i.i.d.r.v.) with symmetric probability distribution, i.e. $\zeta_0 \stackrel{\mathcal{L}}{=} -\zeta_0$, and such as there exists $r_0 \geq 1$ and $s \geq \min(2, r_0)$ satisfying

$$\mathbb{E}(\zeta_0) = 0, \quad \mathbb{E}(|\zeta_0|^{r_0}) = 1 \quad \text{and} \quad \mathbb{E}(|\zeta_0|^s) < \infty. \quad (2)$$

- For $x = (x_n)_{n \in \mathbb{N}} \in \mathbb{R}^\infty$ where \mathbb{R}^∞ is the space of real sequences with a finite number of non zero terms, $(\theta, (x_n)_{n \in \mathbb{N}}) \rightarrow M_\theta((x_n)_{n \in \mathbb{N}}) \in (0, \infty)$ and $(\gamma, (x_n)_{n \in \mathbb{N}}) \rightarrow f_\gamma((x_n)_{n \in \mathbb{N}}) \in \mathbb{R}$ are two known applications.

In [Bardet and Wintenberger \(2009\)](#) and [Bardet et al. \(2017\)](#), it was proved that all the most famous stationary time series used in econometrics, such as ARMA, AR(∞), GARCH, ARCH(∞), TARCH, ARMA-GARCH processes can be written as a causal stationary solution of (1).

Email addresses: y.boularouk@centre-univ-mila.dz (Yakoub Boularouk), bardet@univ-paris1.fr (Jean-Marc Bardet)

The maximum of the conditional quasi-likelihood method, Gaussian QMLE, is the most used to estimate the parameters of the models for stochastic processes. This method was mainly built using the Gaussian instrumental density, see for instance [Berkes *et al.* \(2003\)](#) or [Francq and Zakoian \(2004\)](#) for GARCH(p, q) processes, [Francq and Zakoian \(2013\)](#) for ARMA-GARCH processes, [Straumann and Mikosch \(2006\)](#) for general heteroskedastic models, and [Bardet and Wintenberger \(2009\)](#) for the general class of affine causal models. [Trindade *et al.* \(2006\)](#) studies the ARMA and GARCH models driven by asymmetric Laplace noise, [Bardet *et al.* \(2017\)](#) estimated the model parameters using the Laplacian QMLE, *i.e.* the estimator is constructed from a Laplacian conditional density. Even if the obtained estimators are consistent only under moment conditions on the conditional density, it could be. But theoretically, the divergence of true innovation density can greatly increase the variance of the estimates, increasing with the cost of ignoring true distribution innovation.

In order to avoid this arbitrary choice of QML conditional density, [Lii and Rosenblatt \(1992\)](#) proposed an approximate procedure of maximum average non-reversible moving average processes driven by a non-Gaussian noise, [Francq *et al.* \(2011\)](#) proposed a two stage non Gaussian QML estimation for GARCH processes based on generalized Gaussian errors, [Jianqing *et al.* \(2014\)](#) proposed a three step quasi-maximum likelihood procedure.

The generalized Gaussian density also known as the Generalized Error Distribution (denoted GED(r) with $r > 0$) or the power Gamma distribution is given by

$$g_r(x) = \frac{1}{2} \frac{r^{1-1/r}}{\Gamma(1/r)} e^{-\frac{|x|^r}{r}} \quad \text{for } x \in \mathbb{R}. \quad (3)$$

Note that g_1 is the Laplace density and g_2 is the Gaussian one. If Z_r follows a GED(r), then

$$\mathbb{E}(Z_r) = 0 \quad \text{and} \quad \mathbb{E}(|Z_r|^r) = 1. \quad (4)$$

Moreover, we have the following result and notation:

$$m_r(p) = \mathbb{E}(|Z_r|^p) = r^{\frac{p}{r}-1} \frac{\Gamma(\frac{p+1}{r})}{\Gamma(\frac{r+1}{r})} \quad \text{for any } p > 0. \quad (5)$$

In this paper, we propose a new two stage estimation procedure leading to a Pseudo Generalized Gaussian Quasi-Maximum Likelihood Estimator (PGGQMLE) in the general case of affine causal process.

1. Firstly, we assume that (ζ_t) in (1) is a white noise distributed following a GED(r_0) with $r_0 \geq 1$ an unknown parameter. Then r_0 is estimated by \widehat{r} using jointly the Gaussian and Laplacian QMLEs of θ_0 . The strong consistency of \widehat{r} is established.
2. Secondly, after showing that the Generalized Gaussian Quasi-Maximum Likelihood Estimator (GGQMLE) of θ_0 , *i.e.* a quasi-maximum likelihood estimator built using a GED(r) as an instrumental density (see below), is strongly consistent for any $r > 1$, by replacing r with \widehat{r} a PGGQMLE is obtained and its consistency is also established.

The following Section 2 provides the definitions and assumptions. Section 3 studies the estimator \widehat{r} of the parameter shape r and its consistency is established. Then, the asymptotic behavior of the PGGQML estimator is studied in Section 4, while the results of Monte-Carlo experiments are presented in Section 6.

2. Definition and assumptions

2.1. Definition of the estimator

Assume that (X_1, \dots, X_n) is an observed trajectory of X solution of (1) where $\theta^{0*} \in \Theta \subset \mathbb{R}^d$ and $\gamma^{0*} \in \Gamma \subset \mathbb{R}^e$ are unknown. For estimating θ^{0*} and γ^{0*} we consider the log-likelihood of (X_1, \dots, X_n) conditionally to (X_0, X_{-1}, \dots) . If g is the probability density (with respect to Lebesgue measure) of ζ_0 , then, from the affine causal definition of X , this conditional log-likelihood is equal to:

$$\sum_{t=1}^n \log \left(\frac{1}{M_\theta^t} g \left(\frac{X_t - f_\gamma^t}{M_\theta^t} \right) \right)$$

where $M_\theta^t := M_\theta(X_{t-1}, X_{t-2}, \dots)$ and $f_\gamma^t := f_\gamma(X_{t-1}, X_{t-2}, \dots)$, with the assumption that $M_\theta^t > 0$. However, M_θ^t and f_γ^t are generally not computable since X_0, X_{-1}, \dots are unknown. Thus, a quasi-log-likelihood is considered instead of the log-likelihood and it is defined by:

$$\log(QL_{(\theta, \gamma)}^{(g)}(X_1, \dots, X_n)) = \sum_{t=1}^n \log \left(\frac{1}{\widehat{M}_\theta^t} g \left(\frac{X_t - \widehat{f}_\gamma^t}{\widehat{M}_\theta^t} \right) \right),$$

with $\widehat{f}_\gamma^t := f_\gamma(X_{t-1}, \dots, X_1, u)$ and $\widehat{M}_\theta^t := M_\theta(X_{t-1}, \dots, X_1, u)$, where $u = (u_n)_{n \in \mathbb{N}}$ is a finitely non-zero sequence $(u_n)_{n \in \mathbb{N}}$. The choice of $(u_n)_{n \in \mathbb{N}}$ does not have any consequences on the asymptotic behaviour of L_n , and (u_n) could typically be chosen as a sequence of zeros. Finally, a Quasi-Maximum Likelihood Estimator (QMLE) of $(\theta^{0*}, \gamma^{0*})$ can be defined with the respect of the choice of g

$$\widehat{\phi}_n^{(g)} = (\widehat{\theta}_n^{(g)}, \widehat{\gamma}_n^{(g)}) := \text{Argmax}_{(\theta, \gamma) \in \Theta \times \Gamma} \log(QL_{(\theta, \gamma)}^{(g)}(X_1, \dots, X_n)).$$

The aim of this paper is to propose a new choice of the function g , while the QMLE is generally built with $g = g_2$ the standard Gaussian distribution (denoted Gaussian QMLE) or, less often, with $g = g_1$ the standard Laplacian distribution (denoted Laplacian QMLE). Now, we will consider moire generally $g = g_r$.

As a consequence, for any $1 \leq r \leq s$ with s defined in (2), then the equation (1) can be written again:

$$X_t = M_{\theta^{r*}}(X_{t-1}, X_{t-2}, \dots) \zeta_t^{(r)} + f_{\gamma^{0*}}(X_{t-1}, X_{t-2}, \dots), \quad t \in Z \quad (6)$$

with $\zeta_t^{(r)} = \zeta_t (\mathbb{E}(|\zeta_0|^r))^{-1/r}$, implying $\mathbb{E}(|\zeta_t^{(r)}|^r) = 1$ and $M_{\theta^{r*}} = (\mathbb{E}(|\zeta_0|^r))^{1/r} M_{\theta^{0*}}$. Then, we can define the Generalized Gaussian Quasi Maximum Likelihood $\widehat{\phi}_n^{(r)} = (\widehat{\theta}_n^{(r)}, \widehat{\gamma}_n^{(r)})$ of $\phi^{r*} = (\theta^{r*}, \gamma^{0*})$ that is defined by

$$\widehat{\phi}_n^{(r)} = (\widehat{\theta}_n^{(r)}, \widehat{\gamma}_n^{(r)}) := \text{Argmin}_{(\theta, \gamma) \in \Theta \times \Gamma} \sum_{t=1}^n \widehat{q}_t(\theta, \gamma) \quad \text{where} \quad \widehat{q}_t(\theta, \gamma) := \log(|\widehat{M}_\theta^t|) + \frac{1}{r} |\widehat{M}_\theta^t|^{-r} |X_t - \widehat{f}_\gamma^t|^r. \quad (7)$$

In other words, this estimator is equal to $\widehat{\phi}_n^{(g)}$ when $g = g_r$ the GED(r) density.

REMARK 2.1. We also see that for $r = 2$ the Generalized Gaussian contrast is the Gaussian contrast and for $r = 1$ it is the Laplacian one.

Bardet and Wintenberger (2009) and Bardet *et al.* (2017) have respectively proved the consistency and the asymptotic normality of $\widehat{\phi}_n^{(2)}$ and $\widehat{\phi}_n^{(1)}$ respectively. Using these both these consistent estimators, we begin with the estimation of the parameter shape r_0 when the distribution of ζ_0 is exactly a GED(r_0).

2.2. Existence and stationarity

First we will provide some sufficient conditions for insuring the existence and stationarity of a solution of (1) such as $\mathbb{E}[|X_0|^s] < \infty$ with $s \geq 1$. As it was already done in Doukhan and Wintenberger (2007), several Lipschitz-type inequalities on f_γ and M_θ can be used for obtaining this s -order stationarity of an ergodic causal solution of (1). First, denote $\|g_\phi\|_\Phi = \sup_{\phi \in \Phi} \|g_\phi\|$ where $\Phi \subset \mathbb{R}^{d+e}$ and $\|\cdot\|$ is the usual Euclidian norm. Now, let us introduce the generic symbol K_ϕ for any of function $K_\phi : \mathbb{R}^N \mapsto \mathbb{R}^m$ or $\mathcal{M}_m(\mathbb{R})$ (for instance $K_\phi = f_\phi$ or M_ϕ or their derivatives). For $k = 0, 1, 2$, define a Lipschitz assumption on function K_ϕ :

Assumption (A $_k(K, \Phi)$) $\forall x \in \mathbb{R}^\infty, \phi \in \Phi \mapsto K_\phi(x) \in C^k(\Phi)$ and $\partial_\phi^k K_\phi$ satisfies $\|\partial_\phi^k K_\phi(0)\|_\Phi < \infty$ and there exists a sequence $(\alpha_j^{(k)}(K, \Phi))_j$ of nonnegative numbers such that $\forall x, y \in \mathbb{R}^\infty$

$$\|\partial_\phi^k K_\phi(x) - \partial_\phi^k K_\phi(y)\|_\Phi \leq \sum_{j=1}^{\infty} \alpha_j^{(k)}(K, \Phi) |x_j - y_j|, \quad \text{with} \quad \sum_{j=1}^{\infty} \alpha_j^{(k)}(K, \Phi) < \infty.$$

For ensuring a stationary s -order solution of (1), where $s \geq 1$, define the set

$$\Phi(s) := \left\{ \phi = (\theta, \gamma) \in \mathbb{R}^{d+e}, (\mathbf{A}_0(f, \{\phi\})) \text{ and } (\mathbf{A}_0(M, \{\phi\})) \text{ hold,} \right. \\ \left. \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\phi\}) + (\mathbb{E}[|\zeta_0|^s])^{1/s} \sum_{j=1}^{\infty} \alpha_j^{(0)}(M, \{\phi\}) < 1 \right\}.$$

Then, from Doukhan and Wintenberger (2007), we obtain:

Proposition 2.1. *If $\phi^{0*} \in \Phi(s)$ for some $s \geq 1$, then there exists a unique causal (X_t) is independent of $(\zeta_i)_{i>t}$ for $t \in \mathbb{Z}$ solution X of (1), which is stationary, ergodic and satisfies $\mathbb{E}[|X_0|^s] < \infty$.*

Thus, the stationarity and r -order of a solution of (6) is ensured from this corollary:

Corollary 2.1. *If $\phi \in \Phi^r(r)$, with:*

$$\Phi^r(r) := \left\{ \phi = (\theta, \gamma) \in \mathbb{R}^{d+e}, (\mathbf{A}_0(f, \{\phi\})) \text{ and } (\mathbf{A}_0(M, \{\phi\})) \text{ hold,} \sum_{j=1}^{\infty} \alpha_j^{(0)}(f, \{\phi\}) + \sum_{j=1}^{\infty} \alpha_j^{(0)}(M, \{\phi\}) < 1 \right\}, \quad (8)$$

then there exists a unique causal solution X of (6), which is stationary, ergodic and satisfies $\mathbb{E}[|X_0|^r] < \infty$.

2.3. Additive assumptions required for the estimation

Fix some compact subset Φ of $\Phi(s) \subset \mathbb{R}^{d+e}$. We will consider the following assumptions:

(Ainf) There exists $\underline{M} > 0$ such that $\inf_{(\theta, \gamma) \in \Phi} M_{\theta}(x) \geq \underline{M}$ for all $x \in \mathbb{R}^{\infty}$.

(Id) For all $(\theta, \gamma) \in \Phi$, $(f_{\gamma}^t = f_{\gamma_0}^t \text{ and } M_{\theta}^t = M_{\theta_0}^t \text{ a.s.}) \Rightarrow \theta = \theta_0 \text{ and } \gamma = \gamma_0$.

(Var) One of the families $(\partial f_{\gamma_0}^t / \partial \gamma_i)_{1 \leq i \leq d}$ or $(\partial M_{\theta_0}^t / \partial \theta_i)_{1 \leq i \leq d}$ is a.e. linearly independent, where:

$$\frac{\partial f_{\gamma}^t}{\partial \gamma} := \frac{\partial f_{\gamma}^t}{\partial \gamma}(X_{t-1}, \dots) \text{ and } \frac{\partial M_{\theta}^t}{\partial \theta} := \frac{\partial M_{\theta}^t}{\partial \theta}(X_{t-1}, \dots).$$

The condition [(Id)] is a usual identifiability condition while the condition (Var) is needed for ensuring the finiteness of the asymptotic variance in the result on asymptotic normality.

3. Estimation of the parameter shape r_0

In this section we propose a method to estimate the parameter shape r_0 when (ζ_t) is supposed to exactly follow a $\text{GED}(r_0)$ with $r_0 \geq 1$.

3.1. Construction of the estimator in case of $\text{GED}(r_0)$ white noise

Assume now that ζ_0 follows a $\text{GED}(r_0)$ and (X_1, \dots, X_n) is an observed trajectory of (X_t) that satisfies (1). In this case, a straightforward relation can be established between $M_{\theta^{0*}}$ and $M_{\theta^{(r)}}$:

Lemma 3.1. *For any $r \geq 1$, when (X_1, \dots, X_n) is an observed trajectory of (X_t) that satisfies (1) and ζ_0 follows a $\text{GED}(r_0)$, then*

$$M_{\theta^{0*}} = M_{\theta^{(r)}} r^{\frac{1}{r} - \frac{1}{r_0}} \left(\frac{\Gamma(1 + \frac{1}{r_0})}{\Gamma(\frac{r+1}{r_0})} \right)^{1/r} \text{ and } f_{\gamma^{r*}}^t = f_{\gamma_0}^t. \quad (9)$$

Proof. Here we use the relation provided by the rewriting of (1), i.e.

$$M_{\theta^{0*}} = M_{\theta^{r*}} (\mathbb{E}[|\zeta_0|^r])^{-1/r} \quad (10)$$

and the moment equality (5) and that induce (9). \square

In the sequel we will consider two particular cases $r = 1$, corresponding to the Laplacian QMLE and $r = 2$, corresponding to the classical Gaussian QMLE. Hence using the results of Lemma 3.1 we obtain:

$$\left(\frac{M_{\theta^{1*}}^t}{M_{\theta^{2*}}^t}\right)^2 = \frac{\Gamma^2(\frac{2}{r_0})}{\Gamma(\frac{1}{r_0})\Gamma(\frac{3}{r_0})} := H(r_0). \quad (11)$$

This function H is a continuous and increasing function so it is invertible. The forthcoming Figure 1 exhibits the graph of H .

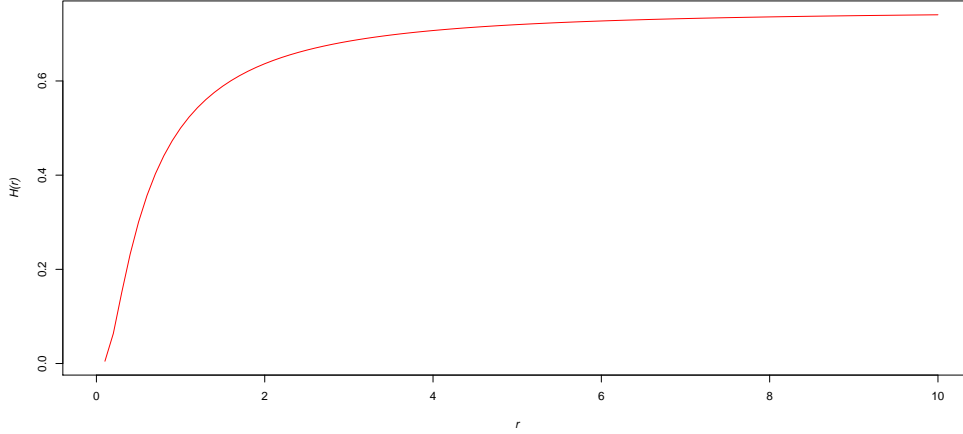


Figure 1: Graph of the function $H(r)$

The relation (11) leads to the definition of an estimator \widehat{r}_n of the parameter shape r_0 given by the expression

$$\widehat{r}_n = H^{-1}\left(\frac{1}{n} \sum_{i=1}^n \left(\frac{\widehat{M}_{\theta_n^{(1)}}^t}{\widehat{M}_{\theta_n^{(2)}}^t}\right)^2\right), \quad (12)$$

since θ^{1*} and θ^{2*} are unknown and can be estimated using QMLE estimators.

3.2. Consistency of \widehat{r}_n

Now we study the consistency of the proposed estimator \widehat{r}_n for the shape parameter r_0 .

Theorem 3.1. *Let X be a stationary solution of the equation (1) where ζ_0 follows a $GED(r_0)$ and $\phi^{0*} \in \Phi$, a compact subset of $\Phi(2)$. Assume also that Assumptions $(\mathbf{A}_0(f, \Phi))$, $(\mathbf{A}_0(M, \Phi))$, (\mathbf{Ainf}) and (\mathbf{Id}) hold with*

$$\alpha_j^{(0)}(f, \Phi) + \alpha_j^{(0)}(M, \Phi) = O(j^{-\ell}) \text{ for some } \ell > 3/2. \quad (13)$$

Then the estimator \widehat{r}_n is strongly consistent, i.e. $\widehat{r}_n \xrightarrow[n \rightarrow \infty]{a.s.} r_0$.

Proof of Theorem 3.1. It is clear that $\widehat{e}_t(\theta^{1*}, \theta^{2*}) := \left(\frac{\widehat{M}_t^{\theta^{1*}}}{\widehat{M}_t^{\theta^{2*}}}\right)^2 = H(r_0)$ for any $t \in \mathbb{Z}$ by definition of θ^{1*} and θ^{2*} . As a consequence, $h_n(\theta^{1*}, \theta^{2*}) := \frac{1}{n} \sum_{t=1}^n \widehat{e}_t(\theta^{1*}, \theta^{2*}) = H(r_0)$. But since the function $\theta \in \mathbb{R}^d \mapsto M_\theta(x)$ is supposed to be a continuous function for any $x \in \mathbb{R}^\infty$, the function $(\theta^1, \theta^2) \in \mathbb{R}^{2d} \mapsto h_n(\theta^1, \theta^2)$ is also almost surely a continuous function.

Bardet and Wintenberger (2009) proved that $\widehat{\theta}_n^2 \xrightarrow[n \rightarrow \infty]{a.s.} \theta^{2*}$, Bardet et al. (2017) proved that $\widehat{\theta}_n^1 \xrightarrow[n \rightarrow \infty]{a.s.} \theta^{1*}$ under the assumptions.

Therefore $h_n(\widehat{\theta}_n^1, \widehat{\theta}_n^2) - h_n(\theta^{1*}, \theta^{2*}) \xrightarrow[n \rightarrow \infty]{a.s.} 0$, i.e. $h_n(\widehat{\theta}_n^1, \widehat{\theta}_n^2) \xrightarrow[n \rightarrow \infty]{a.s.} H(r_0)$. Since the function H^{-1} is also a continuous function, the proof is established. \square

4. Consistency of the Pseudo-Generalized Gaussian Quasi-Maximum Likelihood Estimator

In this section we first study the consistency of the Generalized Gaussian Quasi-Maximum Likelihood Estimator (GGQMLE).

Theorem 4.1. Let X be a stationary solution of the equation (6) with $1 \leq r \leq s$ and $\phi^{r*} \in \Phi$ for a compact subset $\Phi \subset \Phi^r(r)$ defined in (8) under Assumptions $(\mathbf{A}_0(f, \Phi))$, $(\mathbf{A}_0(M, \Phi))$. Assume also that Assumptions (\mathbf{Ainf}) and (\mathbf{Id}) hold with

$$\alpha_j^{(0)}(f, \Phi) + \alpha_j^{(0)}(M, \Phi) = O(j^{-\ell}) \text{ for some } \ell > \max\left\{2 \frac{r+1}{s} - 1, 1\right\}. \quad (14)$$

Then, for any $r \geq 1$, $\widehat{\phi}_n^{(r)} \xrightarrow[n \rightarrow \infty]{a.s.} \phi^{r*}$ with ϕ^{r*} defined in (6).

REMARK 4.1. When $s \geq r + 1$, the bound in (14) is obtained with $\ell > 1$. As a consequence, long-range dependent processes such as FARIMA processes (that are particular cases of AR(∞) processes) can also be considered.

Proof of Theorem 4.1. Using the assumptions, (X_t) is also a solution of (6). The proof of the theorem is divided into two parts and follows the same procedure than in Bardet and Wintenberger (2009). Consider $(\widehat{q}_t(\theta, \gamma))_{0 \leq t \leq n}$ defined in (7) and $(q_t)_{t \in \mathbb{Z}}$ defined by

$$q_t(\theta, \gamma) := \log(|M_\theta^t|) + \frac{1}{r} |M_\theta^t|^{-r} |X_t - f_\gamma^t|^r. \quad (15)$$

1. We first prove that

$$\left\| \frac{1}{n} \sum_{t=1}^n q_t(\theta, \gamma) - \mathbb{E}(q_t(\theta, \gamma)) \right\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (16)$$

(i) In the same way and for the same reason in the proof of Theorem 1 of Bardet et al. (2017), a uniform (on Θ) strong law of large numbers is satisfied by $(q_t(\theta, \gamma))_{t \in \mathbb{Z}}$, which is a sequence of martingale increments, is implied by establishing $\mathbb{E}(\|q_t(\phi)\|_{\Phi}) < \infty$. But for all $t \in \mathbb{Z}$,

$$\begin{aligned} |q_t(\theta, \gamma)| &\leq \left| \log(|M_\theta^t|) \right| + \frac{1}{r} |M_\theta^t|^{-r} |X_t - f_\gamma^t|^r \\ &\leq \frac{|X_t - f_\gamma^t|^r}{r \underline{M}^r} + \left| \log(|\underline{M}|) \right| + \frac{|M_\theta^t|}{|\underline{M}|} \\ &\leq C (|X_t|^r + |f_\gamma^t|^r + |M_\theta^t| + 1), \end{aligned}$$

with $C > 0$ that does not depend on (θ, γ) . But, since $\Theta \subset \Phi(r)$ implying $\mathbb{E}(|X_t|^r) < \infty$, and as it was already proved in Bardet and Wintenberger (2009), $\mathbb{E}(\sup_{\theta \in \Theta} (|M_\theta^t|^r + |f_\gamma^t|^r)) < \infty$. Therefore $\mathbb{E}(\|q_0(\theta, \gamma)\|_{\Phi}) < \infty$ and (16) holds.

2. Moreover, we also have:

$$\left\| \frac{1}{n} \sum_{t=1}^n (q_t(\theta, \gamma) - \widehat{q}_t(\theta, \gamma)) \right\|_{\Theta} \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (17)$$

Indeed, for any $t \geq 1$ and $(\theta, \gamma) \in \Theta$, and using several times the mean value Theorem,

$$\begin{aligned}
r \left| \widehat{q}_t(\theta, \gamma) - q_t(\theta, \gamma) \right| &\leq r \left| \log(|\widehat{M}_\theta^t|) - \log(|M_\theta^t|) \right| + \left| |\widehat{M}_\theta^t|^{-r} |X_t - \widehat{f}_\gamma^t|^r - |M_\theta^t|^{-r} |X_t - f_\gamma^t|^r \right| \\
&\leq \frac{r}{\underline{M}} \left| \widehat{M}_\theta^t - M_\theta^t \right| + |X_t - f_\gamma^t|^r \times \left| |\widehat{M}_\theta^t|^{-r} - |M_\theta^t|^{-r} \right| \\
&\quad + |\widehat{M}_\theta^t|^{-r} \left| |X_t - \widehat{f}_\gamma^t|^r - |X_t - f_\gamma^t|^r \right| \\
&\leq \frac{r}{\underline{M}} \left| \widehat{M}_\theta^t - M_\theta^t \right| + 2^{r-1} (|X_t|^r + |f_\gamma^t|^r) \times \frac{r}{\underline{M}^{r+1}} \left| \widehat{M}_\theta^t - M_\theta^t \right| \\
&\quad + \frac{1}{\underline{M}^r} r |Z_\gamma|^{r-1} \left| \widehat{f}_\gamma^t - f_\gamma^t \right|
\end{aligned}$$

where $|Z_\gamma| \leq \max(|X_t - \widehat{f}_\gamma^t|, |X_t - f_\gamma^t|) \leq 2|X_t| + |\widehat{f}_\gamma^t| + |f_\gamma^t|$. Therefore, there exists $C > 0$ depending on r and \underline{M} such as

$$\left| \widehat{q}_t(\theta, \gamma) - q_t(\theta, \gamma) \right| \leq C \left((1 + |X_t|^r + |f_\gamma^t|^r) \left| \widehat{M}_\theta^t - M_\theta^t \right| + (|X_t|^{r-1} + |\widehat{f}_\gamma^t|^{r-1} + |f_\gamma^t|^{r-1}) \left| \widehat{f}_\gamma^t - f_\gamma^t \right| \right).$$

As a consequence, if $r \leq s \leq r+1$, using Hölder Inequality with $p = (r+1)/r$ and $q = r+1$, we obtain

$$\begin{aligned}
\mathbb{E} \left(\left\| \widehat{q}_t(\theta, \gamma) - q_t(\theta, \gamma) \right\|_\Theta^{s/(r+1)} \right) &\leq C \mathbb{E} \left((1 + |X_t|^r + \|f_\gamma^t\|_\Phi^r)^{s/(r+1)} \left\| \widehat{M}_\theta^t - M_\theta^t \right\|_\Phi^{s/(r+1)} \right. \\
&\quad \left. + (|X_t|^{r-1} + \|\widehat{f}_\gamma^t\|_\Phi^{r-1} + \|f_\gamma^t\|_\Phi^{r-1})^{s/(r+1)} \left\| \widehat{f}_\gamma^t - f_\gamma^t \right\|_\Phi^{s/(r+1)} \right) \\
&\leq C \left(1 + (\mathbb{E}(|X_t|^s))^{r/(r+1)} + (\mathbb{E}(\|f_\gamma^t\|_\Phi^s))^{r/(r+1)} \right) (\mathbb{E}(\left\| \widehat{M}_\theta^t - M_\theta^t \right\|_\Phi^s))^{1/(r+1)} \\
&\quad + C \left((\mathbb{E}(|X_t|^{s(r-1)/r})^{r/(r+1)} + (\mathbb{E}(\|f_\gamma^t\|_\Phi^{s(r-1)/r})^{r/(r+1)} + (\mathbb{E}(\|\widehat{f}_\gamma^t\|_\Phi^{s(r-1)/r})^{r/(r+1)}) \right) (\mathbb{E}(\left\| \widehat{f}_\gamma^t - f_\gamma^t \right\|_\Phi^s))^{1/(r+1)}.
\end{aligned}$$

Using again [Bardet and Wintenberger \(2009\)](#), we know that $\max(\mathbb{E}(|X_t|^s), \mathbb{E}(\|f_\gamma^t\|_\Phi^s), \mathbb{E}(\|\widehat{f}_\gamma^t\|_\Phi^s)) < \infty$. Moreover, we also have

$$\mathbb{E}(\left\| \widehat{M}_\theta^t - M_\theta^t \right\|_\Phi^s) \leq \mathbb{E}(|X_0|^s) \left(\sum_{k=t}^{\infty} \alpha_k^{(0)}(M, \Phi) \right)^s \quad \text{and} \quad \mathbb{E}(\left\| \widehat{f}_\gamma^t - f_\gamma^t \right\|_\Phi^s) \leq \mathbb{E}(|X_0|^s) \left(\sum_{k=t}^{\infty} \alpha_k^{(0)}(f, \Phi) \right)^s.$$

Finally, we obtain that there exists $C > 0$ not depending on t sur as for any $t \geq 1$,

$$\mathbb{E} \left(\left\| \widehat{q}_t(\theta, \gamma) - q_t(\theta, \gamma) \right\|_\Theta^{s/(r+1)} \right) \leq C \sum_{k=t}^{\infty} (\alpha_k^{(0)}(M, \Phi) + \alpha_k^{(0)}(f, \Phi))^{s/(r+1)} \leq C t^{1-(\ell s)/(r+1)}, \quad (18)$$

from (14). Using Corollary 1 of [Kounias and Weng \(1969\)](#), (17) is established if there exists $u \in (0, 1]$ such as

$$\sum_{t \geq 1} \frac{1}{t^u} \mathbb{E} \left(\left\| q_t(\theta, \gamma) - \widehat{q}_t(\theta, \gamma) \right\|_\Phi^u \right) < \infty. \quad (19)$$

It is such the case with $u = s/(r+1)$, since we have:

$$\sum_{t \geq 1} \frac{1}{t^{s/(r+1)}} \mathbb{E} \left(\left\| q_t(\theta, \gamma) - \widehat{q}_t(\theta, \gamma) \right\|_\Phi^{s/(r+1)} \right) \leq C \sum_{t \geq 1} t^{1-(\ell+1)s/(r+1)} < \infty.$$

In case of $s \geq r+1$, it is sufficient to consider the previous case with $s = r+1$.

3. From (16) and (17), we deduce

$$\left\| \frac{1}{n} \sum_{t=1}^n \widehat{q}_t(\theta, \gamma) - \mathbb{E}(q_t(\theta, \gamma)) \right\|_\Theta \xrightarrow[n \rightarrow \infty]{a.s.} 0. \quad (20)$$

Moreover, for $\phi = (\theta, \gamma) \in \Phi$, we study

$$L(\phi) = -\mathbb{E}(q_0(\phi)).$$

which can also be consider as a Kullback-Lieber discrepancy. Using $\mathbb{E}(|\zeta_t|^r) = 1$, we obtain:

$$\begin{aligned} L(\phi) &= -\frac{1}{r} \mathbb{E} \left(\log(|M_\theta^t|^r) + \left(\frac{M_{\theta^{r^*}}^t}{M_\theta^t} \right)^r \left| \zeta_t + \frac{f_{\gamma^{0^*}}^t - f_\gamma^t}{M_{\theta^{r^*}}^t} \right|^r \right) \\ \Rightarrow L(\phi^{r^*}) - L(\phi) &= \frac{1}{r} \mathbb{E} \left(\log \left(\left| \frac{M_\theta^t}{M_{\theta^{r^*}}^t} \right|^r \right) + \left| \frac{M_{\theta^{r^*}}^t}{M_\theta^t} \right|^r \left| \zeta_t + \frac{f_{\gamma^{0^*}}^t - f_\gamma^t}{M_{\theta^{r^*}}^t} \right|^r - 1 \right) \\ &= \frac{1}{r} \mathbb{E} \left(\log \left(\left| \frac{M_\theta^t}{M_{\theta^{r^*}}^t} \right|^r \right) - 1 + \left| \frac{M_{\theta^{r^*}}^t}{M_\theta^t} \right|^r \mathbb{E} \left(\left| \zeta_t + \frac{f_{\gamma^{0^*}}^t - f_\gamma^t}{M_{\theta^{r^*}}^t} \right|^r \mid (X_{t-k})_{k \geq 1} \right) \right). \end{aligned}$$

But for ζ_t following a symmetric probability distribution, for any $m \in \mathbb{R}^*$, $\mathbb{E}(|\zeta_t + m|^r) > \mathbb{E}(|\zeta_t|^r) = 1$. Therefore, for $\theta \neq \theta^{r^*}$, if $f_\gamma \neq f_{\gamma^{0^*}}$ (else $>$ is replaced by \geq),

$$\begin{aligned} L(\phi^{r^*}) - L(\phi) &> \frac{1}{r} \mathbb{E} \left(\log \left(\left| \frac{M_\theta^t}{M_{\theta^{r^*}}^t} \right|^r \right) - 1 + \left| \frac{M_{\theta^{r^*}}^t}{M_\theta^t} \right|^r \right) \\ &> \frac{1}{r} h \left(\left| \frac{M_\theta^t}{M_{\theta^{r^*}}^t} \right|^r \right), \end{aligned}$$

with $h(x) = \log(x) - 1 + x$. But for any $x \in (0, 1) \cup (1, \infty)$, $h(x) > 0$ and $h(1) = 0$. Therefore if $M_\theta \neq M_{\theta^{r^*}}$, $L(\phi^{r^*}) - L(\phi) > 0$ (> 0 is replaced by $= 0$ if $M_\theta = M_{\theta^{r^*}}$). This implies from Condition **(Id)** that $L(\phi^{r^*}) - L(\phi) > 0$ for all $\theta \in \Theta$, $\theta \neq \theta^{r^*}$. Hence a supremum of $L(\phi)$ is only reached for $\phi = \phi^{r^*}$, which is the unique maximum, and this the same behavior for $\sum_{t=1}^n \widehat{q}_t(\theta, \gamma)$ from (20), implying $\widehat{\phi}_n^{(r)} \xrightarrow[n \rightarrow \infty]{a.s.} \phi^{r^*}$. \square

Since $\phi^{r^*} = (\gamma^{r^*}, \theta^{0^*})$, it is clear that for $\widehat{\theta}^{(r)} \xrightarrow[n \rightarrow \infty]{a.s.} \theta^{0^*}$ for any $r \leq s$. Concerning the parameter θ , we will add the following assumption on the function M_θ with p a positive real number:

(HM)(p): For any $C > 0$ and $\theta \in \mathbb{R}^d$, $C \times M_\theta(\cdot) = M_{|C|^p \theta}(\cdot)$.

Examples: Assumption **(HM)(p)** is trivially satisfied in case of ARMA or AR(∞) processes with $p = 1$, but also for GARCH(p, q) processes with $p = 1$, or APARCH(p, δ, q) processes with $p = \delta$.

As a consequence, under Assumption **(HM)(p)**, we have:

$$\theta^{r^*} = (\mathbb{E}(|\zeta_0|^r))^{p/r} \theta^{0^*}.$$

Even if p, r_0 and r are known, the is generally unknown. But in the particular cas where ζ_0 follows a GED(r_0), using (5), we obtain:

$$\theta^{r^*} = \left(r_0^{\frac{r}{r_0} - 1} \frac{\Gamma(\frac{r+1}{r_0})}{\Gamma(\frac{r_0+1}{r_0})} \right)^{p/r} \theta^{0^*}.$$

Therefore, if ζ_0 follows a GED(r_0), $\widehat{\theta}^{(r)}$ can also be used for estimating θ^{0^*} for any $1 \leq r \leq s$. But which parameter r has to be chosen for estimated the parameters θ^{0^*} and γ^{0^*} ? Clearly, when r_0 is known, the choice $r = r_0$ is induced by the following property:

Property 4.1. Let X be a stationary solution of the equation (1) where ζ_0 follows a GED(r_0) and $\phi^{0^*} \in \Phi$, a compact subset of $\Phi(r_0)$. Assume also that Assumptions **(A₀)(f, Φ)**, **(A₀)(M, Φ)**, **(Ainf)** and **(Id)** hold with

$$\alpha_j^{(0)}(f, \Phi) + \alpha_j^{(0)}(M, \Phi) = O(j^{-\ell}) \text{ for some } \ell > 1. \quad (21)$$

Then the estimator $\widehat{\phi}^{(r_0)}$ is asymptotically efficient (its renormalized asymptotic covariance behaves as the Cramèr-Rao bound).

Proof of Property 4.1. Denote $\widetilde{\mathcal{L}}_n(\phi) = \frac{1}{n} \sum_{t=1}^n q_t(\theta, \gamma)$ and $\mathcal{L}_n(\phi) = \frac{1}{n} \log(f_{(X_1, \dots, X_n)}(X_1, \dots, X_n))$, where $f_{(X_1, \dots, X_n)}$ is the probability density of the vector (X_1, \dots, X_n) with respect to the Lebesgue measure when (ζ_t) is a white noise following a GED(r_0) distribution. Then, if $\widetilde{\phi}_n := \text{Argmax}_{\phi \in \Phi} \mathcal{L}_n(\phi)$ then $\widetilde{\phi}$ is the classical maximum likelihood estimator. Using [Daniels \(1961\)](#), we know that $\widetilde{\phi}$ is asymptotically efficient.

But $\mathcal{L}_n(\phi) = \widetilde{\mathcal{L}}_n(\phi) + \frac{1}{n} \log(f_{X_1}(X_1))$ and therefore $\mathcal{L}_n(\phi) = \widetilde{\mathcal{L}}_n(\phi) + O_{\mathbb{P}}(1/n)$.

Moreover $\widetilde{\mathcal{L}}_n(\phi) = \frac{1}{n} \sum_{t=1}^n \widehat{q}_t(\theta, \gamma) + O_{\mathbb{P}}(1/n)$ under condition (21). As a consequence, $\widetilde{\phi}_n = \widehat{\phi}^{(r_0)} + O_{\mathbb{P}}(1/n)$ and therefore $\widehat{\phi}^{(r_0)}$ is also asymptotically efficient. \square

What can be done if r_0 is unknown? We can use the previous estimator \widehat{r}_n of r_0 and plug-in it in the formula of the estimator of ϕ . Hence:

Property 4.2. Under the assumptions of Property 4.1 but with Φ , a compact subset of $\Phi(v)$ with $v = \max(r_0, 2)$, then $\widehat{\phi}_n^{(\widehat{r}_n)} \xrightarrow[n \rightarrow \infty]{a.s.} \phi^{0*}$ with ϕ^{0*} defined in (1).

Proof of Property 4.2. The function $r \in [1, \infty) \rightarrow \widehat{\phi}_n^{(r)}$ is a.s. a continuous function. As a consequence, using $\widehat{r}_n \xrightarrow[n \rightarrow \infty]{a.s.} r_0$ obtained in Theorem 3.1 and $\widehat{\phi}_n^{(r_0)} \xrightarrow[n \rightarrow \infty]{a.s.} \phi^{0*}$ obtained in Property 4.1, we obtain $\widehat{\phi}_n^{(\widehat{r}_n)} \xrightarrow[n \rightarrow \infty]{a.s.} \phi^{0*}$. \square

5. Examples

In this section, we explicitly apply the proposed two stage estimation procedure to several classical time series that are affine causal models: AR(p), GARCH(p, q) and APARCH(p, δ, q) processes.

1/ AR(p) processes: Consider the equation of an AR(p) process, *i.e.*

$$X_t = a^{0*} \zeta_t + \mu^{0*} + \sum_{i=1}^p \beta_i^{0*} X_{t-i}, \quad t \in \mathbb{Z}, \quad (22)$$

where $\mu^{0*} \in \mathbb{R}$, $a^{0*} > 0$, $\sum_{j=1}^{\infty} |\beta_j^{0*}| < 1$ and $(\zeta_t)_{t \in \mathbb{Z}}$ is a white noise satisfying (2).

Such AR(p) process is a special case of the affine causal process (1) with $M_\theta^t = a$ and $f_\gamma^t = \mu + \sum_{i=1}^p \beta_i X_{t-i}$ where $\theta = a$ and $\gamma = (\mu, \beta_1, \dots, \beta_p)$. For $0 < \alpha < \beta < \infty$, and $0 < \rho < 1$, define

$$\Theta = [\alpha, \beta] \quad \text{and} \quad \Gamma = \{(\mu, \beta_1, \dots, \beta_p) \in \mathbb{R}^{p+1}, |\mu| \leq \rho^{-1} \text{ and } \sum_{i=1}^p |\beta_i| \leq \rho\}.$$

Moreover, since M_θ^t is a real constant not depending of the time, the parameter shape estimator defined in 12 becomes,

$$\widehat{r}_n = H^{-1} \left(\left(\frac{\widehat{a}_n^{(1)}}{\widehat{a}_n^{(2)}} \right)^2 \right)$$

Then, the Generalized Gaussian Quasi-Maximum Likelihood Estimator $\widehat{\phi}_n^{(\widehat{r}_n)} = (\widehat{\theta}_n^{(\widehat{r}_n)}, \widehat{\gamma}_n^{(\widehat{r}_n)})$ is defined by

$$\widehat{\phi}_n^{(\widehat{r}_n)} := \text{Argmin}_{(\theta, \gamma) \in \Theta \times \Gamma} \sum_{t=1}^n \widehat{q}_t(\theta, \gamma) \quad \text{where} \quad \widehat{q}_t(\theta, \gamma) := \log(a) + \frac{1}{\widehat{r}_n} |a|^{-\widehat{r}_n} |X_t - (\mu + \sum_{i=1}^p \beta_i X_{t-i})|^{\widehat{r}_n}, \quad (23)$$

with $X_0 = X_{-1} = \dots = X_{1-p} = 0$ by convention.

Then the assumptions of Property 4.1 and 4.2 are satisfied since the Lipschitz coefficients of M_θ and f_γ and their derivatives decrease exponentially fast (see for instance [Bardet and Wintenberger \(2009\)](#)).

2/ GARCH process: The GARCH(p, q) process has been introduced by [Bollerslev \(1986\)](#) as the solution of equations

$$\begin{cases} X_t = \sigma_t \zeta_t, \\ \sigma_t^2 = \omega^{0*} + \sum_{i=1}^p \alpha_i^{0*} X_{t-i}^2 + \sum_{j=1}^q \beta_j^{0*} \sigma_{t-j}^2, \end{cases} \quad (24)$$

where $\omega^{0*} > 0$, $\alpha_i^{0*} \geq 0$ for $i = 1, \dots, p$ and $\beta_j^{0*} \geq 0$ for $j = 1, \dots, q$ with $\alpha_p^{0*}, \beta_q^{0*}$ positive and

$$\sum_{i=1}^p \alpha_i^{0*} + \sum_{j=1}^q \beta_j^{0*} < 1.$$

As a consequence $(X_t)_t$ is a stationary process such as $\mathbb{E}(|X_0|^s) < \infty$ (see for instance [Bardet and Wintenberger \(2009\)](#)). Such GARCH process is a special case of affine causal process (1) where $M_\theta^t = \sigma_t(\theta)$ and $f_\gamma^t = 0$, we denote here $\theta = (\omega, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q)$ and $(\sigma_t(\theta))_t$ satisfies the recurrence relationship

$$\sigma_t^2(\theta) = \omega + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2(\theta) \quad \text{for any } t \in \mathbb{Z}.$$

Now define Θ such as:

$$\Theta = \{\theta \in [0, \infty[^{p+q+1}, \rho \leq \omega \leq 1/\rho, \sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j \leq \rho'\},$$

with $0 < \rho, \rho' < 1$, and this ensuring the stationarity of (X_t) for any $\theta \in \Theta$.

Using Lemma 3.1 we get

$$\widehat{r}_n = H^{-1} \left(\frac{1}{n} \sum_{t=1}^n \frac{\widehat{\sigma}_t^2(\widehat{\theta}_n^{(1)})}{\widehat{\sigma}_t^2(\widehat{\theta}_n^{(2)})} \right) \quad (25)$$

where $\widehat{\sigma}_t^2$ is defined using $(X_i)_{i \in \mathbb{N}} = 0$.

We estimate θ by the Generalized Gaussian Quasi-Maximum Likelihood Estimator $\widehat{\theta}_n^{(\widehat{r}_n)}$ defined by

$$\widehat{\theta}_n^{(\widehat{r}_n)} := \text{Argmin}_{\theta \in \Theta} \sum_{t=1}^n \left(\log(|\widehat{M}_\theta^t|) + \frac{1}{\widehat{r}_n} \left| \frac{X_t}{\widehat{M}_\theta^t} \right|^{\widehat{r}_n} \right).$$

Then the assumptions of Property 4.1 and 4.2 are satisfied since the Lipschitz coefficients of M_θ and f_γ and their derivatives decrease exponentially fast (see for instance [Bardet and Wintenberger \(2009\)](#)).

3/ APARCH process: The APARCH(p, δ, q) processes have been introduced by [Ding et al. \(1993\)](#) as the solution of equations

$$\begin{cases} X_t = \sigma_t \zeta_t, \\ \sigma_t^\delta = \omega^{0*} + \sum_{i=1}^p \alpha_i^{0*} (|X_{t-i}| - \gamma_i^{0*} X_{t-i})^\delta + \sum_{j=1}^q \beta_j^{0*} \sigma_{t-j}^\delta, \end{cases} \quad (26)$$

where $\delta \geq 1$, $\omega^{0*} > 0$, $-1 < \gamma_i^{0*} < 1$ and $\alpha_i^{0*} \geq 0$ for $i = 1, \dots, p$, $\beta_j^{0*} \geq 0$ for $j = 1, \dots, q$ satisfying $\sum_{j=1}^q \beta_j^{0*} < 1$. In the sequel the parameter δ is supposed to be known.

More generally, we denote now

$$\theta = (\omega, \alpha_1, \dots, \alpha_p, \gamma_1, \dots, \gamma_p, \beta_1, \dots, \beta_q),$$

with $\omega > 0$, $-1 < \gamma_i < 1$ and $\alpha_i \geq 0$ for $i = 1, \dots, p$, $\beta_j \geq 0$ for $j = 1, \dots, q$ satisfying $\sum_{j=1}^q \beta_j^{0*} < 1$. Then conditions of stationarity of (X_t) can be deduced. Indeed, using L the usual backward operator such as $LX_t = X_{t-1}$, then $(1 - \sum_{j=1}^q \beta_j L^j)^{-1}$ exists and simple computations imply for $t \in \mathbb{Z}$:

$$\begin{aligned} \sigma_t^\delta(\theta) &= (1 - \sum_{j=1}^q \beta_j L^j)^{-1} \left[\omega + \sum_{i=1}^p \alpha_i (1 - \gamma_i)^\delta (\max(X_{t-i}, 0))^\delta + \alpha_i (1 + \gamma_i)^\delta (-\min(X_{t-i}, 0))^\delta \right] \\ &= b_0(\theta) + \sum_{i \geq 1} b_i^+(\theta) (\max(X_{t-i}, 0))^\delta + \sum_{i \geq 1} b_i^-(\theta) (\max(-X_{t-i}, 0))^\delta. \end{aligned}$$

where $b_0(\theta) = \omega(1 - \sum_{j=1}^q \beta_j)^{-1}$ and the coefficients $(b_i^+(\theta), b_i^-(\theta))_{i \geq 1}$ are defined by the recursion relations

$$\begin{cases} b_i^+(\theta) = \sum_{k=1}^q \beta_k b_{i-k}^+(\theta) + \alpha_i (1 - \gamma_i)^\delta & \text{with } \alpha_i (1 - \gamma_i) = 0 \text{ for } i > p \\ b_i^-(\theta) = \sum_{k=1}^q \beta_k b_{i-k}^-(\theta) + \alpha_i (1 + \gamma_i)^\delta & \text{with } \alpha_i (1 + \gamma_i) = 0 \text{ for } i > p \end{cases} \quad (27)$$

with $b_i^+(\theta) = b_i^-(\theta) = 0$ for $i \leq 0$.

As a consequence, an APARCH process is a special case of the affine causal process with $f_\theta^t \equiv 0$ and $M_\theta^t = \sigma_t(\theta)$. Therefore $\alpha_j^{(0)}(f, \Theta) = 0$ and simple computations imply $\alpha_j^{(0)}(M, \{\theta\}) = \sup_{\theta \in \Theta} \max(|b_j^+(\theta)|^{1/\delta}, |b_j^-(\theta)|^{1/\delta})$ and we established in [Bardet et al. \(2017\)](#) that the sequence $(\alpha_j^{(0)}(M, \{\theta\}))_i$ decreases exponentially fast. Thus, with $0 < \rho, \rho' < 1$ define the compact set Θ by

$$\Theta = \left\{ \theta \in \mathbb{R}^{2p+q+1} \mid \rho \leq \omega \leq 1/\rho \text{ and } \sum_{j=1}^{\infty} \max(|b_j^+|^{1/\delta}, |b_j^-|^{1/\delta}) \leq \rho' \right\}. \quad (28)$$

For $\theta^{0*} \in \Theta$, the estimator of the parameter shape r_0 in the case of an APARCH process is

$$\widehat{r}_n = H^{-1} \left(\frac{1}{n} \sum_{t=1}^n \frac{\widehat{\sigma}_t^2(\widehat{\theta}_n^{(1)})}{\widehat{\sigma}_t^2(\widehat{\theta}_n^{(2)})} \right). \quad (29)$$

Since we have an estimation \widehat{r}_n for the parameter shape r , we use it in the estimation of $(\widehat{\theta}_n^{\widehat{r}_n})$

$$\widehat{\theta}_n^{\widehat{r}_n} := \operatorname{Argmin}_{\theta \in \Theta} \sum_{t=1}^n \left(\log(|\widehat{M}_\theta^t|) + \frac{1}{\widehat{r}_n} \left| \frac{X_t}{\widehat{M}_\theta^t} \right|^{\widehat{r}_n} \right).$$

Then the assumptions of Property 4.1 and 4.2 are satisfied since the Lipschitz coefficients of M_θ and f_γ and their derivatives decrease exponentially fast (see for instance [Bardet et al. \(2017\)](#)).

6. Numerical Results

In order to illustrate the interest of the estimate of the parameter shape r , we realize Monte-Carlo experiments on the behavior of Gaussian QMLE, Laplace QMLE and the Pseudo Generalized Gaussian QMLE, for GARCH and APARCH processes and several sizes of sample ($n = 100$, $n = 1000$ and $n = 5000$). More precisely, we will consider:

- a GARCH(1, 1) process defined by $X_t = \sigma_t \zeta_t$ where $\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta \sigma_{t-1}^2$ with $\alpha_0 = 0.2$, $\alpha_1 = 0.4$ and $\beta = 0.2$;
- an APARCH(1, δ , 1) process defined by $X_t = \sigma_t \zeta_t$ where $\sigma_t^\delta = \alpha_0 + \alpha_1 (|X_{t-1}| - \gamma X_{t-1})^\delta + \beta \sigma_{t-1}^\delta$ and $\alpha_0 = 0.2$, $\alpha_1 = 0.4$, $\gamma = 0.8$, $\beta = 0.2$ and $\delta = 1.2$ (δ is supposed to be known).

6.1. Estimation of r

We first consider white noise $(\zeta_t)_{t \in \mathbb{Z}}$ such as the distribution of ζ_0 is a generalized Gaussian distributions for several values of r : $r = 1$ (Laplace distribution), $r = 1.3$, $r = 1.7$, $r = 2$ (Gaussian distribution) and $r = 2.6$. Using 1000 independent replications of both the processes, r is estimated by \widehat{r}_n defined in (12) and its root-mean-square error (RMSE) is computed and reported in forthcoming tables 1 and 2.

	$r = 1$	$r = 1.3$	$r = 1.7$	$r = 2$	$r = 2.6$
$n = 100$	0.362	0.366	0.650	1.007	1.381
$n = 1000$	0.064	0.084	0.122	0.158	0.196
$n = 5000$	0.030	0.039	0.051	0.071	0.086

Table 1: RMSE of \widehat{r}_n for the GARCH(1, 1) process particular case from 1000 independent replications.

Conclusion of the numerical results: The simulations exhibit that the larger the sample size n the smaller the RMSE of the parameter shape estimator \widehat{r}_n . They also show that the larger the shape parameter r the larger the RMSE of \widehat{r}_n . More precisely, it seems that the RMSE of \widehat{r}_n/r is only depending on n .

	$r = 1$	$r = 1.3$	$r = 1.7$	$r = 2$	$r = 2.6$
$n = 100$	0.522	0.672	1.560	2.946	2.386
$n = 1000$	0.073	0.090	0.756	0.218	0.271
$n = 5000$	0.029	0.036	0.067	0.089	0.117

Table 2: RMSE of \widehat{r}_n for the APARCH(1, 1) process particular case from 1000 independent replications.

6.2. Comparisons of Gaussian, Laplacian and Pseudo-Gaussian Generalized QMLE

In the sequel, we will consider white noises (ζ_t) where ζ_0 follows several different probability distributions with unit variance:

- the Gaussian distribution $\mathcal{N}(0, 1)$;
- the centered Laplacian distribution $\mathcal{L}(1/\sqrt{2})$;
- the Uniform distribution $\mathcal{U}([- \sqrt{3}, \sqrt{3}])$;
- the renormalized Student distribution $\sqrt{3/5} t_5$.

Hence, to allow comparison of the estimators, we have chosen the usual normalization of the model writing with unit variance white noise (ζ_t) . We first assume that ζ_0 follows a Generalized Gaussian distribution $\text{GED}(r_0)$ to within one multiplicative constant, with $r_0 \geq 1$, and therefore $\zeta_0 = (m_{r_0}(2))^{-1/2} Z_{r_0}$ for obtaining $\mathbb{E}(\zeta_0^2) = 1$. Then we can write with $r_0 \geq 1$: Hence, for $r \geq 1$, and for both the processes, we can write with $\mathbb{E}(\zeta_0^2) = 1$ and

$$X_t = M_{\theta^{2*}}^t \zeta_t = M_{\theta^{2*}}^t (m_{r_0}(2))^{-1/2} \zeta_t^{r_0}$$

with $(\zeta_t^{r_0})$ a white noise such as $\zeta_0^{r_0}$ follows the $\text{GED}(r_0)$ distribution defined in (3). Now consider $r \geq 1$. We can write:

$$X_t = M_{\theta^{2*}}^t \zeta_t = \left(\frac{(m_{r_0}(r))^{1/r}}{(m_{r_0}(2))^{1/2}} M_{\theta^{2*}}^t \right) \zeta_t^{(r)} = M_{\theta^{r*}}^t \zeta_t^{(r)},$$

with $\zeta_t^{(r)} = \zeta_t \frac{(m_{r_0}(2))^{1/2}}{(m_{r_0}(r))^{1/r}}$ and therefore $\mathbb{E}(|\zeta_t^{(r)}|^r) = 1$. As a consequence,

1. For GARCH(1, 1) process, we have $(M_{\theta^{2*}}^t)^2 = \alpha_0^{(2*)} + \alpha_1^{(2*)} X_{t-1}^2 + \beta^{(2*)} (M_{\theta^{2*}}^{t-1})^2$. Then we deduce

$$(M_{\theta^{r*}}^t)^2 = \frac{(m_{r_0}(r))^{2/r}}{m_{r_0}(2)} (\alpha_0^{(2*)} + \alpha_1^{(2*)} X_{t-1}^2) + \beta^{(2*)} (M_{\theta^{r*}}^{t-1})^2 = \alpha_0^{(r*)} + \alpha_1^{(2*)} X_{t-1}^2 + \beta^{(r*)} (M_{\theta^{r*}}^{t-1})^2$$

$$\implies \begin{cases} \alpha_0^{(r*)} &= \frac{(m_{r_0}(r))^{2/r}}{m_{r_0}(2)} \alpha_0^{(2*)} \\ \alpha_1^{(r*)} &= \frac{(m_{r_0}(r))^{2/r}}{m_{r_0}(2)} \alpha_1^{(2*)} \\ \beta^{(r*)} &= \beta^{(2*)} \end{cases} .$$

Thus, we will compare three estimators of $\theta^{(2*)} = (\alpha_0^{(2*)}, \alpha_1^{(2*)}, \beta_1^{(2*)})$:

- the classical Gaussian QMLE $\widehat{\theta}_n^{(2)} = (\widehat{\alpha}_0^{(2)}, \widehat{\alpha}_1^{(2)}, \widehat{\beta}^{(2)})$ defined in (7);

- the modified Laplacian QMLE $\widehat{\theta}_n^{(1)} = (\widehat{\alpha}_0^{(1)}, \widehat{\alpha}_1^{(1)}, \widehat{\beta}^{(1)})$ such as

$$\begin{cases} \widehat{\alpha}_0^{(1)} &= \frac{m_{\widehat{r}_n}(2)}{(m_{\widehat{r}_n}(1))^2} \widehat{\alpha}_0^{(1)} \\ \widehat{\alpha}_1^{(1)} &= \frac{m_{\widehat{r}_n}(2)}{(m_{\widehat{r}_n}(1))^2} \widehat{\alpha}_0^{(1)} \\ \widehat{\beta}^{(1)} &= \widehat{\beta}^{(1)} \end{cases}, \quad \text{where } \widehat{\theta}_n^{(1)} = (\widehat{\alpha}_0^{(1)}, \widehat{\alpha}_1^{(1)}, \widehat{\beta}^{(1)});$$

- the modified Pseudo-Gaussian Generalized QMLE $\widehat{\theta}_n^{(\widehat{r}_n)} = (\widehat{\alpha}_0^{(\widehat{r}_n)}, \widehat{\alpha}_1^{(\widehat{r}_n)}, \widehat{\beta}^{(\widehat{r}_n)})$ such as

$$\begin{cases} \widehat{\alpha}_0^{(\widehat{r}_n)} &= m_{\widehat{r}_n}(2) \widehat{\alpha}_0^{(\widehat{r}_n)} \\ \widehat{\alpha}_1^{(\widehat{r}_n)} &= m_{\widehat{r}_n}(2) \widehat{\alpha}_1^{(\widehat{r}_n)} \\ \widehat{\beta}^{(\widehat{r}_n)} &= \widehat{\beta}^{(\widehat{r}_n)} \end{cases}, \quad \text{where } \widehat{\theta}_n^{(\widehat{r}_n)} = (\widehat{\alpha}_0^{(\widehat{r}_n)}, \widehat{\alpha}_1^{(\widehat{r}_n)}, \widehat{\beta}^{(\widehat{r}_n)}).$$

- For the APARCH(1, δ , 1) process, we have $(M_{\theta^{2*}}^t)^\delta = \alpha_0^{(2*)} + \alpha_1^{(2*)}(|X_{t-1}| - \gamma^{(2*)} X_{t-1})^\delta + \beta^{(2*)} (M_{\theta^{2*}}^{t-1})^\delta$. Therefore we deduce

$$\begin{aligned} (M_{\theta^{r*}}^t)^\delta &= \frac{(m_{r_0}(r))^{\delta/r}}{(m_{r_0}(2))^{\delta/2}} (\alpha_0^{(2*)} + \alpha_1^{(2*)}(|X_{t-1}| - \gamma^{(2*)} X_{t-1})^\delta) + \beta^{(2*)} (M_{\theta^{r*}}^{t-1})^\delta \\ &= \alpha_0^{(r*)} + \alpha_1^{(r*)}(|X_{t-1}| - \gamma^{(r*)} X_{t-1})^\delta + \beta^{(r*)} (M_{\theta^{r*}}^{t-1})^\delta \\ &\implies \begin{cases} \alpha_0^{(r*)} &= \frac{(m_{r_0}(r))^{\delta/r}}{(m_{r_0}(2))^{\delta/2}} \alpha_0^{(2*)} \\ \alpha_1^{(r*)} &= \frac{(m_{r_0}(r))^{\delta/r}}{(m_{r_0}(2))^{\delta/2}} \alpha_1^{(2*)} \\ \gamma^{(r*)} &= \gamma^{(2*)} \\ \beta^{(r*)} &= \beta^{(2*)} \end{cases}. \end{aligned}$$

Thus, we will compare three estimators of $\theta^{(2*)} = (\alpha_0^{(2*)}, \alpha_1^{(2*)}, \gamma^{(2*)}, \beta^{(2*)})$:

- the classical Gaussian QMLE $\widehat{\theta}_n^{(2)} = (\widehat{\alpha}_0^{(2)}, \widehat{\alpha}_1^{(2)}, \widehat{\gamma}^{(2)}, \widehat{\beta}^{(2)})$ defined in (7);
- the modified Laplacian QMLE $\widehat{\theta}_n^{(1)} = (\widehat{\alpha}_0^{(1)}, \widehat{\alpha}_1^{(1)}, \widehat{\gamma}^{(1)}, \widehat{\beta}^{(1)})$ such as

$$\begin{cases} \widehat{\alpha}_0^{(1)} &= \frac{(m_{\widehat{r}_n}(2))^{\delta/2}}{(m_{\widehat{r}_n}(1))^\delta} \widehat{\alpha}_0^{(1)} \\ \widehat{\alpha}_1^{(1)} &= \frac{(m_{\widehat{r}_n}(2))^{\delta/2}}{(m_{\widehat{r}_n}(1))^\delta} \widehat{\alpha}_0^{(1)} \\ \widehat{\gamma}^{(1)} &= \widehat{\gamma}^{(1)} \\ \widehat{\beta}^{(1)} &= \widehat{\beta}^{(1)} \end{cases}, \quad \text{where } \widehat{\theta}_n^{(1)} = (\widehat{\alpha}_0^{(1)}, \widehat{\alpha}_1^{(1)}, \widehat{\gamma}^{(1)}, \widehat{\beta}^{(1)});$$

- the modified Pseudo-Gaussian Generalized QMLE $\widehat{\theta}_n^{(\widehat{r}_n)} = (\widehat{\alpha}_0^{(\widehat{r}_n)}, \widehat{\alpha}_1^{(\widehat{r}_n)}, \widehat{\beta}^{(\widehat{r}_n)})$ such as

$$\begin{cases} \widehat{\alpha}_0^{(\widehat{r}_n)} &= (m_{\widehat{r}_n}(2))^{\delta/2} \widehat{\alpha}_0^{(\widehat{r}_n)} \\ \widehat{\alpha}_1^{(\widehat{r}_n)} &= (m_{\widehat{r}_n}(2))^{\delta/2} \widehat{\alpha}_1^{(\widehat{r}_n)} \\ \widehat{\gamma}^{(\widehat{r}_n)} &= \widehat{\gamma}^{(\widehat{r}_n)} \\ \widehat{\beta}^{(\widehat{r}_n)} &= \widehat{\beta}^{(\widehat{r}_n)} \end{cases}, \quad \text{where } \widehat{\theta}_n^{(\widehat{r}_n)} = (\widehat{\alpha}_0^{(\widehat{r}_n)}, \widehat{\alpha}_1^{(\widehat{r}_n)}, \widehat{\beta}^{(\widehat{r}_n)}).$$

Remark: Since \widehat{r}_n is obtained from the reciprocal function H^{-1} , it could not be defined if the condition $\frac{1}{n} \sum_{t=1}^n \left(\frac{\widehat{M}_{\widehat{\theta}_n^{(1)}}^t}{\widehat{M}_{\widehat{\theta}_n^{(2)}}^t} \right)^2 <$

0.75 is not satisfied. As a consequence, in the numerical procedure we set $\widehat{r}_n = 10$ when $\frac{1}{n} \sum_{t=1}^n \left(\frac{\widehat{M}_{\widehat{\theta}_n^{(1)}}^t}{\widehat{M}_{\widehat{\theta}_n^{(2)}}^t} \right)^2 \geq 0.74$. This is a very common case when we take the uniform distribution as the one of the white noise.

The results are presented in Tables 3 and 4.

		\mathcal{L}			\mathcal{N}			t_5			\mathcal{U}		
		$\widehat{\theta}_n^{r_n}$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$	$\widehat{\theta}_n^{r_n}$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$	$\widehat{\theta}_n^{r_n}$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$	$\widehat{\theta}_n^{r_n}$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$
$n = 100$	α_0	0.622	0.096	0.102	0.207	0.089	0.088	0.968	0.101	0.105	1.094	0.094	0.079
	α_1	0.801	0.253	0.279	0.344	0.200	0.197	0.526	0.250	0.265	1.590	0.167	0.134
	β	0.392	0.230	0.273	0.294	0.226	0.235	0.351	0.229	0.269	0.129	0.221	0.190
	Sum	1.816	0.580	0.654	0.845	0.515	0.520	1.845	0.580	0.639	2.813	0.482	0.403
	\widehat{r}_n		1.323			2.540			1.647			8.872	
$n = 1000$	α_0	0.043	0.039	0.043	0.037	0.036	0.034	0.051	0.038	0.047	0.019	0.035	0.027
	α_1	0.102	0.092	0.102	0.068	0.063	0.060	0.107	0.084	0.110	0.036	0.049	0.036
	β	0.104	0.102	0.111	0.087	0.090	0.087	0.101	0.094	0.119	0.034	0.083	0.065
	Sum	0.250	0.234	0.256	0.192	0.189	0.180	0.259	0.216	0.277	0.088	0.167	0.130
	\widehat{r}_n		1.021			2.035			1.224			9.789	
$n = 5000$	α_0	0.018	0.017	0.019	0.016	0.016	0.015	0.022	0.017	0.023	0.015	0.016	0.012
	α_1	0.042	0.040	0.044	0.027	0.028	0.025	0.045	0.037	0.051	0.031	0.022	0.017
	β	0.046	0.046	0.051	0.037	0.039	0.037	0.045	0.044	0.063	0.014	0.036	0.028
	Sum	0.106	0.102	0.114	0.081	0.083	0.081	0.112	0.099	0.136	0.059	0.073	0.056
	\widehat{r}_n		1.001			2.008			1.198			9.995	

Table 3: Sample mean of \widehat{r}_n and Root-Mean Square Error of the components of $\widehat{\theta}_n^{r_n}$, $\widehat{\theta}_n^{(1)}$ and $\widehat{\theta}_n^{(2)}$ for the considered GARCH(1, 1) processes.

Conclusion of the numerical results: Firstly, the simulations exhibit that the larger the sample size n the smaller the RMSE of the estimators. Secondly, as we suspected, $\widehat{\theta}_n^{(1)}$ and $\widehat{\theta}_n^{(2)}$ give the best results when the white noise distribution is Laplace (respectively Gaussian). Thirdly, globally, it is $\widehat{\theta}_n^{r_n}$ which provides the best results when $n \geq 1000$ (otherwise for $n = 100$ a bad estimation of r_0^* can be damaging to it). For uniform and Student distributions that are not GED distributions, the procedure automatically searches for the nearest GED. The PGGQMLE estimator $\widehat{\theta}_n^{r_n}$ uses this to provide an estimator as close as possible to the one obtained by maximizing the "true" conditional quasi-likelihood.

		\mathcal{L}			\mathcal{N}			t_5			\mathcal{U}		
		$\widehat{\theta}_n^n$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$	$\widehat{\theta}_n^n$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$	$\widehat{\theta}_n^n$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$	$\widehat{\theta}_n^n$	$\widehat{\theta}_n^{(1)}$	$\widehat{\theta}_n^{(2)}$
$n = 100$	α_0	0.250	0.073	0.078	0.101	0.064	0.062	0.212	0.074	0.078	0.031	0.054	0.045
	α_1	0.411	0.200	0.212	0.221	0.149	0.141	0.363	0.189	0.209	0.069	0.115	0.094
	γ	0.386	0.295	0.307	0.245	0.225	0.213	0.347	0.276	0.298	0.128	0.182	0.162
	β	0.222	0.202	0.204	0.184	0.175	0.169	0.222	0.196	0.200	0.087	0.142	0.121
	Sum	1.269	0.770	0.801	0.751	0.613	0.584	1.145	0.733	0.785	0.314	0.493	0.422
	\widehat{r}_n		1.237			2.470			1.586			8.793	
$n = 1000$	α_0	0.029	0.027	0.029	0.020	0.019	0.018	0.033	0.026	0.033	0.011	0.018	0.013
	α_1	0.070	0.068	0.074	0.051	0.051	0.049	0.076	0.065	0.080	0.024	0.041	0.032
	γ	0.120	0.120	0.130	0.095	0.097	0.095	0.115	0.111	0.135	0.032	0.085	0.066
	β	0.081	0.081	0.086	0.054	0.056	0.053	0.076	0.073	0.094	0.019	0.048	0.037
	Sum	0.300	0.296	0.320	0.220	0.223	0.215	0.300	0.276	0.342	0.086	0.192	0.149
	\widehat{r}_n		1.021			2.029			1.227			9.739	
$n = 5000$	α_0	0.012	0.012	0.013	0.009	0.009	0.009	0.014	0.011	0.015	0.009	0.007	0.006
	α_1	0.032	0.031	0.035	0.022	0.022	0.021	0.034	0.030	0.040	0.020	0.018	0.014
	γ	0.055	0.055	0.061	0.039	0.042	0.039	0.054	0.053	0.071	0.013	0.035	0.027
	β	0.034	0.034	0.039	0.024	0.026	0.024	0.032	0.031	0.042	0.008	0.020	0.016
	Sum	0.134	0.132	0.147	0.094	0.099	0.092	0.135	0.125	0.168	0.050	0.081	0.062
	\widehat{r}_n		1.002			2.007			1.195			9.996	

Table 4: Sample mean of \widehat{r}_n and Root-Mean Square Error of the components of $\widehat{\theta}_n^n$, $\widehat{\theta}_n^{(1)}$ and $\widehat{\theta}_n^{(2)}$ for the considered APARCH(1, 1) processes.

References

- Bardet, J.-M. and Wintenberger, O. (2009) Asymptotic normality of the Quasi-Maximum likelihood estimator for multidimensional causal process. *Ann. Statist.*, **37**, 2730–2759.
- Bardet, J.-M. Boularouk, Y. and Djaballah, K. (2017) Asymptotic behaviour of the Laplacian quasi-maximum likelihood estimator of affine causal processes. *Electronic Journal of Statistics* **11**: 452–79.
- Berkes, I., Horváth, L. and Kokoszka, P. (2003) GARCH processes: structure and estimation. *Bernoulli*, **9**, 201–227.
- Daniels, H.E (1961). The asymptotic efficiency of a maximum likelihood estimator. In Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability (ed. J. Neyman & E.L. Scott), 151–163.
- Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity, *J. Econometrics*, **31**, 307–327.
- Ding, Z., Granger C.W.J. and Engle R.F. (1993) A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance*, **1**, 83–106.
- Doukhan, P. and Wintenberger, O. (2007) Weakly dependent chains with infinite memory. *Stochastic Process. Appl.*, **118**, 1997–2013.
- Duchenes, P. and Francq, C. (2008) On diagnostic checking time series models with portmanteau test statistics based on generalized inverses and 2-inverses. *COMPSTAT 2008, Proceedings in Computational Statistics*, 143–154.
- Francq, C. and Zakoian, J.-M. (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes, *Bernoulli*, **10**, 605–637.
- Francq, C., Lepage, G. and Zakoian, J.-M. (2011) Two-stage non Gaussian QML estimation of GARCH models and testing the efficiency of the Gaussian QMLE. *Journal of Econometrics*, **165**, 246–257.
- Francq, C. and Zakoian, J.-M. (2013) Optimal predictions of powers of conditionally heteroskedastic processes. *Journal of the Royal Statistical Society B*, **75**, 345–367.
- Jeantheau, T. (1998) Strong consistency of estimators for multivariate arch models. *Econometric Theory*, **14**, 70–86.
- Jianqing Fan, Lei Qi and Dacheng Xiu (2014) Quasi-Maximum Likelihood Estimation of GARCH Models With Heavy-Tailed Likelihoods. *Journal of Business and Economic Statistics*, **32**, 178–191
- Keh-Shin Lii, Murray Rosenblatt (1992) An Approximate Maximum Likelihood Estimation for Non-Gaussian Non-minimum Phase Moving Average Processes. *Journal of Multivariate Analysis*, **43**, 272–299
- Krengel, U. (1985) Ergodic Theorems. Gruyter Studies in Mathematics 6, Walter de Gruyter Co., Berlin.
- Kounias, E.G. and Weng, T.-S. (1969) An inequality and almost sure convergence. *Annals of Mathematical Statistics*, **40**, 1091–1093.
- Lumsdaine, R.L. (1996) Consistency and asymptotic normality of the quasi-maximum likelihood in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models. *Econometrica*, **64**, 575–596.
- Robinson, P.M. and Zaffaroni, P. (2006) Pseudo-maximum likelihood estimation of ARCH(∞) models. *Ann. Statist.* **34**, 1049–1074.

- Straumann, D. and Mikosch, T. (2006) Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach. *Ann. Statist.* **34**, 2449–2495.
- Trindade, A. A., Zhu, Y. and Andrews, B. (2010) Time series models with asymmetric Laplace innovations. *Journal of Statistical Computation and Simulation*, 80(12), 1317–1333.
- Weiss, A.A. (1986) Asymptotic theory for ARCH models estimation and testing. *Econometric Theory*, **2**, 107–131.