Title	On the quasi-likelihood estimation for random coefficient autoregressions
Author(s)	Truquet, L; Yao, J
Citation	Statistics, 2012, v. 46 n. 4, p. 505-521
Issued Date	2012
URL	http://hdl.handle.net/10722/159911
Rights	This is an electronic version of an article published in Statistics, 2012, v. 46 n. 4, p. 505-521. The journal article is available online at: http://www.tandfonline.com/doi/abs/10.1080/02331888.2010.541557

ON THE QUASI-LIKELIHOOD ESTIMATION FOR RANDOM COEFFICIENT AUTOREGRESSIONS

L. TRUQUET AND J. YAO

ABSTRACT. We examine the Gaussian Quasi-Maximum Likelihood Estimator (QMLE) for random coefficient autoregressions. Consistency and asymptotic normality are established for general random coefficients and general correlation structure between these coefficients and the noise. In particular, the obtained results apply even if the stationary solution has infinite absolute mean or infinite variance. Next an application to integer-valued times series modeling is given which provides a novel alternative for traditional INAR-like models for these series.

1. INTRODUCTION

Random coefficient autoregressions (RCAR) appear in multiple applications such as econometrics, telecommunication or dynamic population models as well as in fundamental mathematics, see [14] for a review. A real or integer-valued p-th order RCAR takes the form

$$X_t = A_{t,1}X_{t-1} + \cdots + A_{t,p}X_{t-p} + \xi_t, \quad t \in \mathbb{Z}, \tag{1}$$

where the sequence

$$\Phi_t = (A_{t,1}, \ldots, A_{t,p}, \xi_t)\,, \quad t \in \mathbb{Z}, \tag{2}$$

of AR-coefficients and the noise is i.i.d. with possible correlations between the $(p+1)$ components.

The RCAR models, also called random difference equations, have been studied by several authors ([12, 21, 23] and [18]). Most of their theoretic properties are well known, including conditions for the existence and the uniqueness of a stationary solution, or for the existence of moments for the stationary distribution.

In this paper, we address the parameter estimation problem for the stationary RCAR model (1). A semi-parametric estimation procedure for a real-valued RCAR model (1) has

2000 *Mathematics Subject Classification.* Primary: 62M10 ; Secondary: 62M20 .

Key words and phrases. Quasi-likelihood estimation, random coefficients autoregressions, integer-valued time series.

been investigated by Basawa & Hwang [5]. More precisely, these authors investigated a Conditional Least Squares (CLS) as well as a Weighted Least Squares (WLS) estimator of the parameters $\theta = (\mathbb{E}\Phi_0, \mathbb{V}\Phi_0)$ using a three-step procedure.

It turns out that this three-step procedure as well as the conditional least squares can be avoided by using a Quasi-Maximum Likelihood Estimator (QMLE) as proposed in Aue & al. [4]. As it is well-known, when Φ_0 is Gaussian, the QMLE coincides with the (conditional) MLE and is then asymptotically more efficient than the least squares procedure studied in [5]. Aue & al. proved the consistency and the asymptotic normality of the QMLE for model (1) when $p = 1$ and the two components of $\Phi_0 = (A_0, \xi_0)$ are uncorrelated. In this paper we extend their result to the general stationary RCAR process, i.e with an arbitrary order p and possible correlations between the $p + 1$ components of Φ_0. This extension requires some technical care due to a more complex structure of the conditional variance in the general framework. As in [4], we show that the QMLE is asymptotically normal provided the fourth moment of the vector Φ_0 is finite. The latter condition does not impose any moment condition on the marginal X_0 (see the remark after Theorem 2). The same kind of results holds for the well-known GARCH processes (see Mikosch & Straumann [26] where the QMLE estimator is studied for several ARCH models).

Let us also mention two recent works about the convergence of the QMLE for the non-stationary RCAR process when $p = 1$ and the components of ξ_0 are uncorrelated. Berkes & al. [7] examines the asymptotic properties of the QMLE when the RCAR process is assumed to be non-stationary. In the same context, Aue & Horvátz [3] proves that the QMLE can be employed regardless whether the underlying process is stationary or not.

The paper is organized as follows. In Section 2, we recall some well known facts about the stationarity and moments of a general real-valued process (1). We also discuss some specific constraints for its integer-valued version. In Section 3, we state asymptotic results for the QMLE of parameter θ. In Section 4, we propose RCAR model as a novel alternative to the analysis of integer-valued times series. All proofs are postponed to the last section of the paper.

2. Preliminaries on RCAR models

We first recall several useful properties of the RCAR model (1). Let (B_t) be the sequence of companion matrices (of size $p \times p$)

$$B_t = \begin{pmatrix} A_{t,1} & A_{t,2} & \cdots & A_{t,p-1} & A_{t,p} \\ & & I_{p-1} & & 0 \end{pmatrix},$$

and $\|B\|$ be an operator norm on $\mathcal{M}_p$ (the set of real matrices of size $p \times p$) associated to an arbitrary vector norm $\|y\|$ on $\mathbb{R}^p$. Define the

$$\textbf{(H0):} \qquad \gamma < 0, \qquad \mathbb{E}\log^+ \|B_1\| < \infty, \quad \mathbb{E}\log^+ |\xi_1| < \infty \ , \tag{3}$$

where γ is the top-Lyapounov exponent of the sequence (B_t) defined by

$$\gamma = \inf_{t>0} \frac{1}{t} \mathbb{E}\log \|B_t \cdots B_1\| \ .$$

Here for a real number a, $a+ = \max(a, 0)$. It is well-known, see e.g. [13] that under **(H0)**, the RCAR equation (1) has a stationary solution given as the first coordinate of the following almost surely converging series

$$Z_t = (\xi_t, 0, \ldots, 0)' + \sum_{k=1}^{\infty} B_t \cdots B_{t-k+1} (\xi_{t-k}, 0, \ldots, 0)' \ . \tag{4}$$

Throughout the paper, A' denotes the transpose of a matrix A. The condition $\gamma < 0$ is also necessary for an non-anticipative stationary solution as established in [12].

Furthermore, we know conditions for the existence of the moments for the stationary solution. Define for any positive $s > 0$, the generalised moment function

$$k(s) = \inf_{t>0} (\mathbb{E}\|B_t \cdots B_1\|^s)^{\frac{1}{t}} \ . \tag{5}$$

Then it is known, see for instance [19], that

- there is an interval $[0, \sigma)$, possibly $[0, \infty)$, on which the function k is well defined (note that $k(0) = 1$ by definition);
- the function $\log k(s)$ is strictly convex on $[0, \sigma)$ and its derivative from right at zero equals γ.
- if for some $s > 0$, $k(s) < 1$ and $\mathbb{E}|\xi_1|^s < \infty$, then the stationary solution (Z_t) has a moment of order s, i.e. $\mathbb{E}\|Z_t\|^s < \infty$ (a fortiori $\mathbb{E}|X_t|^s < \infty$).

Note that for positive $s \sim 0$, $\log k(s) \sim \gamma s$. Therefore under **(H0)**, $\gamma < 0$, we have $\log k(s) < 0$ for some small $s > 0$. Hence $\mathbb{E}|X_t|^s < \infty$ once $\mathbb{E}|\xi_1|^s < \infty$ for small $s > 0$.

As for the auto-correlation function $\rho(k) = \mathrm{corr}(X_t, X_{t+k})$ (assuming the RCAR process X has a second moment), it is easy to see that the $\rho(k)$ satisfies the following Yule-Walker equation: for $k \in \mathbb{Z}$,

$$\rho(k+p+1) = a_1\rho(k+p) + \cdots + a_p\rho(k+1),$$

with $(a_1, \ldots, a_p) = \mathbb{E}(A_{0,1}, \ldots, A_{0,p})$. In other words, X is an AR(p) process in weak sense.

Let us develop the simplest special case where $p = 1$. The model (1) takes a simpler form

$$X_t = A_t X_{t-1} + \xi_t, \quad t \in \mathbb{Z}.$$

The top-Lyapounov exponent simply equals $\gamma = \mathbb{E}\log|A_1|$. Under **(H0)**, the stationary solution becomes

$$X_t = \xi_t + \sum_{k=1}^{\infty} A_t \cdots A_{t-k+1}\xi_{t-k} \ .$$

If $\mathbb{E}A_t^2 < 1$ and $\mathbb{E}\xi_t^2 < \infty$, X thus has a finite variance. Moreover, it is a weak AR(1) process; in particular its auto-correlation function equals $\rho(k) = a^k$ for $k \geq 0$ where $a = \mathbb{E}A_0$.

3. Parameter estimation

For the estimation problem of the RCAR model (1), let us first rewrite the model in the form

$$X_t = A_{t,1}X_{t-1} + \cdots + A_{t,p}X_{t-p} + \xi_t = \Phi_t' Y_t, \quad t \in \mathbb{Z}, \tag{6}$$

where $Y_t = (X_{t-1}, \ldots, X_{t-p}, 1)'$. The aim is then to estimate the distribution q of the vector Φ_0 of regression coefficients.

As stated in Introduction, we follow a semi-parametric approach as in Basawa & Hwang [5]. More precisely, no specific form will be assumed on the distribution q, and we are only interested in estimating the parameters

$$(m(q), v(q)), \quad m(q) = \mathbb{E}_q\Phi_0, \quad v(q) = \mathbb{V}_q\Phi_0, \tag{7}$$

i.e. the mean and variance parameters of the stochastic coefficients and the noise. Here and in the sequel, $\mathbb{P}_q$ (resp. $\mathbb{E}_q$ and $\mathbb{V}_q$) stand for the probability measure of the process (X_t) (resp. expectation and variance operator) when the probability measure of Φ_0 is q.

Let $\mathcal{F}_t = \sigma(X_{t-j} | j \in \mathbb{N})$ be the σ-algebra (or information set) generated by the variables up to time t. We have

$$m_t(q) = \mathbb{E}_q(X_t|\mathcal{F}_{t-1}) = Y_t' m(q), \qquad V_t(q) = \mathbb{V}_q(X_t|\mathcal{F}_{t-1}) = Y_t' v(q) Y_t. \tag{8}$$

To simplify the notations, we set $\ell = p+1$. Moreover we denote by $\mathcal{S}_\ell$ the set of real valued symmetric semi-definite positive matrices of size ℓ.

In order to state asymptotic results for the QMLE, we define more precisely the parameter space Θ for the mean vector and variance matrix $(m(q), v(q))$ of distribution q. To this end we consider the Cholesky decomposition of a positive semi-definite matrix. If q is the probability distribution of Φ_0, then we can write

$$v(q) = \begin{pmatrix} P & 0 \\ w' & 1 \end{pmatrix} \cdot \begin{pmatrix} D & 0 \\ \mathbf{0}' & c \end{pmatrix} \cdot \begin{pmatrix} P' & w \\ \mathbf{0} & 1 \end{pmatrix}, \tag{9}$$

where $w \in \mathbb{R}^p$, $c \in \mathbb{R}_+$, P is a $p \times p$ unit lower-triangular matrix and D is a $p \times p$ non-negative diagonal matrix. Note that PDP' is a Cholesky decomposition of $\mathbb{V}_q\left(A_{0,1}, \ldots, A_{0,p}\right)$. Of course if $\mathbb{V}_q\left(A_{0,1}, \ldots, A_{0,p}\right)$ is non-degenerate, decomposition (9) is unique. Otherwise, it is well-known (see [28] for quadratic ARCH models) that the decomposition (9) remains unique if we set to zero the elements of w' and the sub-diagonal elements of P in columns corresponding to the zero diagonal elements of D. Note that from (9), the conditional variance becomes

$$Y_0' v(q) Y_0 = c + \left(X_{-1} + v_1, \ldots, X_{-p} + v_p\right) PDP' \left(X_{-1} + v_1, \ldots, X_{-p} + v_p\right)', \tag{10}$$

with $v = (P')^{-1} w$. The condition $c > 0$ ensures that the conditional variance has a positive lower bound, which is a crucial regularity condition for the application of the QMLE. In this context the random coefficients $A_{0,1}, \ldots, A_{0,p}$ can be constant or more generally linearly dependent. Note also that we have from (10),

$$\inf_{x \in \mathbb{R}^p} (x', 1) v(q) \begin{pmatrix} x \\ 1 \end{pmatrix} = c = d_q^2,$$

where d_q denotes the $\mathbb{L}^2-$distance between ξ_0 and the vector subspace generated by the random coefficients $A_{0,1}, \ldots, A_{0,p}$ and the constant 1. Then if $c > 0$, ξ_0 does not write as a linear combination of the random coefficients and the constant 1, in particular ξ_0 is non-degenerate. The strong consistency of the QMLE will be proved in this general setting.

Let ω, δ, r, s be real numbers such that $\omega, r, s \geq 0$ and $\delta > 0$. Then we define the following compact subset K of $\mathbb{R}^d$ with $d = \frac{\ell(\ell+3)}{2}$:

$$K = [-r, r]^\ell \times [-s, s]^{\frac{p^2+p}{2}} \times [\omega, s]^p \times [\delta, s]. \tag{11}$$

Now we define an application $f : K \to \mathbb{R}^d$ as follows. For an element $k \in K$ with

$$k = \left(m_1, \ldots, m_\ell, P_{2,1}, \ldots, P_{p,1}, P_{3,2}, \ldots, P_{p,2}, \ldots, P_{p-1,p}, w_1, \ldots, w_p, d_1, \ldots, d_p, c\right),$$

let $f(k)$ the vector of $\mathbb{R}^d$ defined by

$$f(k) = (m_1, \dots, m_\ell, M_{1,1}, \dots, M_{\ell,1}, M_{2,2}, \dots, M_{\ell,p}, \dots, M_{\ell,\ell}), \tag{12}$$

with

$$M_{i,j} = \begin{cases} \sum_{h=1}^{\min(i,j)} d_h P_{i,h} P_{j,h}, \text{ if } 1 \le j \le i \le p, (\text{ setting } P_{h,h} = 1), \\ \sum_{h=1}^{j-1} P_{j,h} w_h + w_j, \text{ if } 1 \le j \le p \text{ and } i = \ell, \\ c + \sum_{h=1}^{p} w_h^2 d_h, \text{ if } i = j = \ell. \end{cases}$$

Observe that for a given $k \in K$, the coefficients $M_{i,j}$ of $f(k)$ are the lower diagonal elements of the positive semi-definite matrix $v(q)$ in (9).

Finally, we define the parametric space Θ as follows:

(H1): $\Theta = f(K)$ where $K = K(r, s, \omega, \delta)$ is a compact subset of $\mathbb{R}^d$ the form (11) and f is defined by (12).

Since f is a continuous function, Θ is a compact subset of $\mathbb{R}^d$. Moreover Θ can be trivially identified to a compact subset of $\mathbb{R}^\ell \times \mathcal{S}_\ell$, containing the mean vector and covariance matrix $(m(q), v(q))$ of the distribution q of the noise Φ. Then we will denote $m_t(\theta)$ (resp. $V_t(\theta)$) instead of $m_t(q)$ (resp. $V_t(q)$) in (8). If $\omega > 0$ in the definition of Θ, the uniqueness of the Cholesky decomposition (10) implies that f is one to one and we can identify Θ and K. If now $\omega = 0$, f is no more one to one and the parameter k is not identifiable. This is why we consider the parameter $\theta = f(k)$ instead of k. The rule of parameter k is to give a suitable parametrization of a compact subset of covariance matrices.

We are now ready to define the *Gaussian quasi-likelihood*. Suppose we observe $\{X_t\}$, $-p+1 \le t \le n$ where Φ_t has an unknown distribution q_0. The *Gaussian quasi-likelihood* of $(X_1, \dots, X_n)$ is defined, conditionally on $\{X_t\}$, $-p+1 \le t \le 0$, as

$$L_n(X_1, \dots, X_n; \theta) = \prod_{t=1}^{n} \left\{2\pi Y_t' v(q) Y_t\right\}^{-\frac{1}{2}} \exp\left\{-\frac{(X_t - Y_t' m(q))^2}{2 Y_t' v(q) Y_t}\right\}.$$

This is of course the exact (conditional) likelihood when the noise (Φ_t) is Gaussian.

The QMLE equals

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta} L_n(X_1, \dots, X_n;\ \theta) = \arg\min_{\theta \in \Theta} Q_n(\theta), \tag{13}$$

where

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} q_t(\theta);$$

$$q_t(\theta) = \frac{\left(X_t - Y_t' m(q)\right)^2}{Y_t' v(q) Y_t} + \ln Y_t' v(q) Y_t, \qquad t \geq 1.$$

Let

$$Q(\theta) = \mathbb{E}_{q_0}\left(\frac{\left(X_0 - Y_0' m(q)\right)^2}{Y_0' v(q) Y_0} + \ln\left(Y_0' v(q) Y_0\right)\right).$$

Assumption **(H1)** ensures the existence of a uniform positive lower bound for the conditional variance which will be useful to study the strong consistency of the QMLE. More precisely, we have from (10)

$$\inf_{\theta\in\Theta} V_t(\theta) \geq \delta, \text{ a.s.} \tag{14}$$

In order to study the asymptotic normality of the QMLE, it is necessary to assume that the true value θ_0 lies in the interior of Θ. In this case, $v(q_0)$ is positive definite and we can give a more precise lower bound for the conditional variance if we assume the coordinates of Φ_0 linearly independent for all $\theta \in \Theta$. The following lemma will be useful to prove the asymptotic normality under minimal moment conditions. From now on, we use the Euclidean norm for vectors or matrices.

Lemma 1. *If* **(H1)** *is satisfied with $\omega > 0$, then there exists a positive number h such that*

$$\inf_{\theta\in\Theta} V_t(\theta) \geq h\|Y_t\|^2, \text{ a.s.}$$

3.1. **Consistency of the QMLE $\hat{\theta}_n$.** The following hypotheses will also be required. In the sequel, any a.s. convergence refers to the actual distribution $\mathbb{P}_{q_0}$ of the process (X_t).

(H2): Under $\mathbb{P}_{q_0}$, the top-Lyapounov exponent γ of the sequence (B_t) is negative. Moreover, $\mathbb{E}_{q_0}\|\Phi_0\|^2 < \infty$.

(H3): For any matrix M and any vector m, we have

$$\left(m'Y_t, Y_t' M Y_t\right) = 0 \text{ a.s } \Rightarrow (m, M) = 0.$$

Note that **(H2)** does not imply that $\mathbb{E}_{q_0}|X_0|^2 < \infty$. Actually, we know only that for some small $s > 0$, $\mathbb{E}_{q_0}|X_0|^s < \infty$. On the other hand, Condition **(H3)** will ensure the identifiability of the parameter θ_0. We will give below a condition on the distribution of Φ_0 that implies **(H3)**.

Theorem 1. *Under Assumptions* **(H1)** *to* **(H3)**, *the QMLE* $\hat{\theta}_n$ *is strongly consistent:*

$$\lim_{n\to\infty} \hat{\theta}_n = \theta_0 \ a.s.$$

3.2. **Asymptotic normality of the QMLE** $\hat{\theta}_n$**.** In the following, if f is a real function defined on Θ, ∇f denotes its gradient and $\nabla^2 f$ its Hessian matrix.

For asymptotic normality, we need the following assumptions:

(H4): $\mathbb{E}_{q_0}\|\Phi_0\|^4 < \infty$;

(H5): The actual value θ_0 of θ is an interior point of Θ.

Theorem 2. *Under Assumptions* **(H1)- (H5)**, *with* $\omega > 0$ *in the definition of* Θ, *the QMLE* $\hat{\theta}_n$ *is asymptotically normal: as* $n \to \infty$,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \Rightarrow \mathcal{N}(0, F_0^{-1} G_0 F_0^{-1}) \quad \textit{in distribution},$$

where

$$\begin{aligned} F_0 &= \mathbb{E}_{q_0} \nabla^2 q_0(\theta_0) \\ &= \mathbb{E}_{q_0}\left[\{V_0^{-2}\nabla V_0(\nabla V_0)'\}(\theta_0)\right] + 2\mathbb{E}_{q_0}\left[\{V_0^{-1}\nabla m_0(\nabla m_0)'\}(\theta_0)\right], \end{aligned}$$

and with $Z_0 = \{X_0 - m_0(\theta_0)\}/V_0(\theta_0)$,

$$\begin{aligned} G_0 &= \mathbb{V}_{q_0}\{\nabla q_0(\theta_0)\} \\ &= \mathbb{E}_{q_0}\{Z_0^4(\nabla V_0 \nabla V_0')(\theta_0)\} \\ &\quad -\mathbb{E}_{q_0}\left[\{V_0^{-2}\nabla V_0 \nabla V_0'\}(\theta_0)\right] + 4\mathbb{E}_{q_0}\left[\{V_0^{-1}\nabla m_0 \nabla m_0'\}(\theta_0)\right] \\ &\quad +\mathbb{E}_{q_0}\left[Z_0^3\{\nabla V_0 \nabla m_0' + \nabla m_0 \nabla V_0'\}(\theta_0)\right]. \end{aligned}$$

Remark. It is worth noticing that the QMLE is asymptotically normal without assuming any particular moment condition for the stationary distribution of (X_t). As it was already mentioned the condition $\gamma < 0$ in Assumption **(H2)** ensures the existence of a number $s > 0$ such that $\mathbb{E}_{q_0}|X_0|^s < \infty$. But s can be arbitrary small even if Assumption **(H4)** holds. For example when $p = 1$, A_0 and ξ_0 are independent Gaussian random variables with mean 0 and respective variance σ^2 and 1, the condition for stationarity $\gamma < 0$ becomes

$$\sigma < \sigma_0, \qquad \sigma_0 \approx 2.742.$$

But it is easy to show that

$$\mathbb{E}_{q_0}|X_0|^s \geq \sigma^s \frac{2^{s/2}}{\sqrt{\pi}} \Gamma\left(\frac{s+1}{2}\right) \mathbb{E}_{q_0}|X_0|^s,$$

where Γ denotes the usual Gamma function. Then we deduce that $\mathbb{E}_{q_0}|X_0|^s = \infty$ provided

$$\sigma^s \frac{2^{s/2}}{\sqrt{\pi}} \Gamma\left(\frac{s+1}{2}\right) > 1. \tag{15}$$

For example if $\sigma > \sqrt{\pi/2} \sim 1.25$, $\mathbb{E}|X_0| = \infty$ and when $\sigma > 1.48$, $\mathbb{E}\sqrt{|X_0|} = \infty$. Nevertheless, Theorems 1 and 2 still apply to these cases.

In the case where $p = 1$ and Φ_0 has uncorrelated components, Theorem 1 and 2 have been obtained in [4] under similar conditions. In [5], the conditional least squares are first applied to X for the estimation of the mean of Φ_0 and next to X^2 for the estimation of its variance; a WLS estimation is also investigated to improve the efficiency. This three-step procedure requires the existence of the fourth moment of X and is therefore restrictive. Moreover when Φ_0 is Gaussian, the QMLE coincide with the conditional maximum likelihood estimator and is thus asymptotically efficient.

In [3], the authors study the RCAR process with $p = 1$ and $A_{0,1}$ independent of ξ_0. They show that the QMLE can be employed regardless whether the underlying process is stationary or not. In fact if the process is non-stationary, the variance of ξ_0 cannot be estimated consistently (see [7]), whereas the QMLE of the mean/variance of $A_{0,1}$ is shown to be always consistent and asymptotically normal. Our result for the general RCAR process with possible correlations between the components of Φ_0 do not apply in the non-stationary case (i.e when the top-Lyapunov exponent is non-negative). However the results given in [7] and [3] are stated only for the case $p = 1$ and when $A_{0,1}$ and ξ_0 are uncorrelated and further investigation is needed to study the general case considered in this paper. This difficult problem is beyond the scope of this paper.

3.3. **A sufficient condition for (H3).** In this section, we study the problem of parameter identification. We first give a sufficient condition on the distribution of Φ_0 in order to satisfy **(H3)**.

(ID) For all real numbers $a_1, \ldots, a_p, \alpha, \beta$ with $a_j \neq 0$ for $j = 1, \ldots, p$

$$\mathbb{P}_{q_0}\left(\sum_{j=1}^{p} a_j A_{0,j} + \xi_0 \in \{\alpha, \beta\}\right) < 1.$$

The implications **(ID)** $\Rightarrow$ **(H3)** is an easy consequence of the following lemma.

Lemma 2. *Suppose that* **(ID)** *holds. Let* $\{a_{i,j}, b_j, c/i, j \in \{1,\dots,p\}\}$ *be a family of real numbers satisfying* $a_{i,j} = a_{j,i}$ *and such that:*

$$\sum_{i,j=1}^{p} a_{i,j} X_{t-i} X_{t-j} + \sum_{j=1}^{p} b_j X_{t-j} + c = 0 \ \ a.s.$$

Then $\forall i, j \in \{1,\dots,p\}$, $a_{i,j} = b_j = c = 0$.

Remarks.

(1) Assumption **(ID1)** is satisfied when the coordinates of Φ_t are independent and when ξ_0 and at least one of the coefficients $A_{0,1},\dots,A_{0,p}$ are non degenerate. Indeed if **(ID)** is not satisfied, set $U = \sum_{j=1}^{p} a_j A_{0,j}$ with $a_j \neq 0$, and

$$\mathbb{P}_{q_0}\left(U + \xi_0 \in \{\alpha, \beta\}\right) = 1.$$

We have $\mathbb{V}_{q_0}(U) = \sum_{j=1}^{p} a_j^2 \mathbb{V}_{q_0}\left(A_{0,j}\right) > 0$ and U is non-degenerate. Now

$$1 = \mathbb{P}_{q_0}\left(U + \xi_0 \in \{\alpha, \beta\}\right) = \int \mathbb{P}_{q_0}\left(u + \xi_0 \in \{\alpha, \beta\}\right) \mathbb{P}_U(du),$$

then $\mathbb{P}\left(u + \xi_0 \in \{\alpha, \beta\}\right) = 1$ for $\mathbb{P}_U-$almost all u. Then ξ_0 has a two points distribution with support $\{\alpha - u, \beta - u\}$ for at least two different values of u (since U is non-degenerate). Since ξ_0 is non degenerate, this is not possible and we conclude that **(ID)** is always satisfied.

(2) When $p = 1$, Φ_0 uncorrelated and ξ_0 centered, assumption **(ID)** is also used by Aue et al. [4] in order to prove the asymptotic normality of the QMLE, as well as the condition

(16) $\mathbb{P}\left(A_{0,1}\xi_0 = 0\right) < 1.$

In order to prove the strong consistency of the QMLE, Aue et al. [4] used only the condition (16). We found that condition **(ID)** is sufficient for the proof of both consistency and asymptotic normality. Note that none of these assumptions imply the other. For example when $A_{0,1}$ has a Bernoulli distribution and $\xi_0 = (1 - A_{0,1})\widetilde{\xi}_0$ (see model (19)), where $\widetilde{\xi}$ is independent of $A_{0,1}$ and whose support is not concentrated in one or two points, **(ID)** is satisfied but (16) is not. On the other hand, if $A_{0,1}$ and δ are two independent Bernoulli random variables and $\xi_0 = (1 - A_0) + \delta$, (16) is satisfied but **(ID)** not.

One can mention that if the random variables $A_{0,1}$ and ξ_0 are correlated, condition (16) is no more sufficient for the identification of the parameter θ_0. Let us consider the

case $p = 1$ and give an example of model (1) such that Assumption (16) is satisfied and the parameters are not identifiable. Assume that the observed process X is defined by

$$X_t = A_t X_{t-1} + \xi_t$$

	A_t		
ξ_t	-1	0	1
-1	0	1/3	0
0	0	0	1/3
1	1/3	0	0

Note that the support of the stationary distribution is $\{-1, 2\}$. More, A_0, ξ_0 and 1 are linearly independent. We have $\mathbb{P}(A_{0,1}\xi_0 = 0) = \frac{2}{3} < 1$ and $\mathbb{E}_{q_0}(X_t/\mathcal{F}_{t-1}) = 0$. Moreover

$$\mathbb{V}_{q_0}(X_t|\mathcal{F}_{t-1}) = \frac{2}{3}X_{t-1}^2 - \frac{2}{3}X_{t-1} + \frac{2}{3} = 2.$$

Now for a real number β satisfying $-4/9 < \beta < 0$, we denote $\alpha = -2\beta$ and $\gamma = 2+4\beta$. In this case there exists a random vector Φ_0 with mean 0 and covariance matrix given by $\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$. More,

$$\alpha X_{t-1}^2 + 2\beta X_{t-1} + \gamma = 2 = \mathbb{V}_{q_0}(X_t/\mathcal{F}_{t-1}) \quad a.s,$$

and if $\beta \neq -1/3$, parameters are different.

4. Application to integer-valued time series

Finding accurate models for integer-valued time series has been an important goal for statisticians since a long time. Among firstly proposed models are the class of integer-valued auto-regressions (INAR) introduced in [1, 2, 25]. A general INAR(p) model ([15]) is defined as solution of

$$X_t = \sum_{i=1}^{p} a_i \circ X_{t-i} + \xi_t, \quad t \in \mathbb{Z}, \tag{17}$$

where $a_i \in (0, 1)$ satisfying $\sum_{i=1}^{p} a_i < 1$ and $a \circ X$ denotes the thinning operator of Steutel and Van Harn [29]. Recall that each of the thinning operations $a_i \circ X_{t-i}$ uses a series of Bernoulli-valued counting variables assumed independent from each other among different indexes i and

time epochs t, as well as from the noise $\xi = (\xi_t)$. Concerning the auto-correlation structure, INAR(p) models are examples of AR processes in the weak sense, i.e

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \eta_t, \quad t \in \mathbb{Z}, \tag{18}$$

where η is an uncorrelated process. Then INAR processes have the same auto-correlation structure as AR processes. INAR processes have possible extensions if we consider other probability distributions for the thinning operator $a_j\circ$ (for example a Poisson distribution instead of a Bernoulli distribution). Such extensions called generalized integer-valued autoregressive processes (GINAR) are considered by Latour [22].

Despite these nice properties, INAR or GINAR models have several limitations. For example they can be used to model positive time series only. However, negative-valued times series occur frequently, for example when one uses Box-Jenkins differences when the original series of counts display trend or seasonal effects. Moreover, due to the positivity of the regression coefficients (a_j) in (18), the auto-correlations of a GINAR process are always positive. Thus, time series with negative correlations cannot be treated by these models. Another drawback relates to the stationary distribution of the process. As shown in [29], the stationary distribution of an INAR(1) process is infinitely divisible; hence distributions such as binomial or categorical observations are excluded.

Recent attempts have been made to overcome these drawbacks. In Kachour and Yao [24], the authors propose to replace the thinning operation $a \circ X$ by $\langle aX \rangle$, that is the usual multiplication followed by a rounding step. Their model, named Rounded INAR model (RINAR), is applicable to negative-valued time series as well as to series having some negative auto-correlation coefficients. However, due to the discontinuity of the rounding operator, the CLS estimator proposed by these authors has no regular behaviour. For example, no CLT for the estimator is obtained yet. On the another hand and in order to weaken the constraining requirement for the stationary distribution to be infinitely divisible, Biswas and Song [11] proposed an alternative model using random coefficients. It will be seen below that their model can be viewed as a particular instance of what we present in this paper.

Here we propose to use the RCAR models as a novel alternative to the traditional INAR/GINAR models for integer-valued time series. To be specific, we consider the general RCAR model (1) but with integer-valued random coefficients and noise, i.e. the random vectors

$$\Phi_t = (A_{t,1}, \ldots, A_{t,p}, \xi_t), \quad t \in \mathbb{Z},$$

are assumed to take their values in $\mathbb{Z}$ for all t.
In this case, the condition $\gamma < 0$ in (3) holds if $\mathbb{P}(A_{0,1} = 0, \ldots, A_{0,p} = 0) > 0$. Indeed, it is easy to verify that $\mathbb{P}(B_p \cdots B_1 = 0) > 0$ and then $\mathbb{E}(\log \|B_p \ldots B_1\|) = -\infty$ which implies $\gamma = -\infty$.

All the properties presented in Sections 2 and 3 for both the model and the QMLE estimator apply equally here. To summarize, the RCAR model (1) with integer-valued random coefficients and noise has the following properties

- the model can take negative values as well as positive values ;
- it has the same auto-correlation structure than a standard AR(p) model ;
- the stationary solution is not necessarily infinitely divisible;
- under appropriate moment and identifiability conditions, the associated QMLE estimator is strongly consistent and asymptotically normal.

Therefore, this integer-valued RCAR model has overcome the indicated drawbacks of the INAR/GINAR model, as well as some difficulties of the RINAR model of [24].

4.1. **Comparison with a model proposed by Biswas and Song** [11]**.** The integer-valued AR process proposed in [11] can be presented as follows. Let (δ_t) be a sequence of independent and identically distributed random variables with $\mathbb{P}(\delta_0 = i) = a_i$ for $i = 1, \ldots, p$ and $\mathbb{P}(\delta_0 = \ell) = 1 - \sum_{i=1}^{p} a_i$, where $a_i \in (0,1)$, $\sum_{i=1}^{p} a_i < 1$ and $\ell = p + 1$. Then a weak AR specification discussed in [11] is

$$X_t = \sum_{i=1}^{p} \mathbf{1}_{\delta_t = i} X_{t-i} + \mathbf{1}_{\delta_t = \ell} \cdot \tilde{\xi}_t, \quad t \in \mathbb{Z}, \tag{19}$$

where the sequence $\tilde{\xi}$ is i.i.d and independent of the sequence δ. Note that this process is a particular instance of the general RCAR(p) model (1), with

$$\Phi_t = (\mathbf{1}_{\delta_t = 1}, \ldots, \mathbf{1}_{\delta_t = p}, \mathbf{1}_{\delta_t = \ell} \cdot \tilde{\xi}_t), \quad t \in \mathbb{Z}.$$

In particular, the components of Φ_t are strongly correlated. Furthermore by construction, the stationary distribution of (19) can have an arbitrary form, since this distribution is exactly that of the noise $\tilde{\xi}_t$. As for a general RCAR process, the model (19) has the same auto-correlation function than an AR(p) process with autoregression coefficients $a_i = \mathbb{E}_{q_0} A_{t,i} = \mathbb{P}(\delta_t = i)$, $i = 1, \ldots, p$. Since these coefficients are all positive, the underlying auto-correlation function is less general than for a general AR(p) process or the RCAR processes considered in this paper. In Section 3, we have studied the parametric estimation of RCAR processes in

TABLE 1. Parameters estimates of three fits of a RCAR(1) model for the whole series, its first and second halves respectively.

	a	a'	v	v'
whole series (1:120)	0.44	1.82	0.12	3.59
first half (1:60)	0.48	1.49	0.06	2.81
second half (61:120)	0.43	2.21	0.11	4.57

a very general setting. Note that the model (19) satisfies the assumptions **(H1)**, **(H4)** and **(ID)** when the support of $\widetilde{\xi}_0$ contains at least three distinct values.

4.2. **An application to real data.** The time series in Figure 1 represents the number of cases of E. Coli (O157:H7) infections observed over a four weeks period starting from January 1990 to December 2000. In total we have 143 observed values ranging from 0 to a maximum of 29 cases (December 1999)[1].

A visual inspection reveals that the maximum value of 29 corresponds to an exceptional event we might not be able to analyze correctly. Therefor we restrict our attention to the first 120 observations. A first attempt was made to fit a RCAR model for the whole time series. Despite the fact that the series has a much bigger variance by the end of the period than at the beginning, the sample ACF and PACF suggest that an AR(1) model is likely possible. We then fit a RCAR(1) model $X_t = A_t X_{t-1} + \xi_t$ where (A_t) and (ξ_t) are independent.

A closer look at the series also demonstrates that the second half of the series should have a bigger volatility than the first one. To confirm this, we made two more fits for both halves. The parameter estimates are displayed in Table 1.

Recall that the conditional variance of X_t equals to $vX_{t-1}^2 + v'$. For the first half of the series, the variance parameter estimate $\hat{v} = 0.06$ of the random coefficient is indeed non significant: its standard deviation is estimated to be 0.08. Therefore, a deterministic integer-valued AR(1) model is better suited to this half where the conditional variances could be a constant (homoscedasticity). For the second half, we found $\hat{v} = 0.11$ which is significantly non null indicating a possible heteroscedastic behaviour. This agrees with a higher volatility observed in this half which is also confirmed by a bigger value of the variance parameter $\hat{v}'$ of 4.57 against 2.81.

[1]The authors would like to thank Pierre Boivin from the Health and Social Services (Direction of public health, Department of infectious diseases), in Roberval (Québec) Canada, and Alain Latour from University of Grenoble (France) who kindly provided the data set.

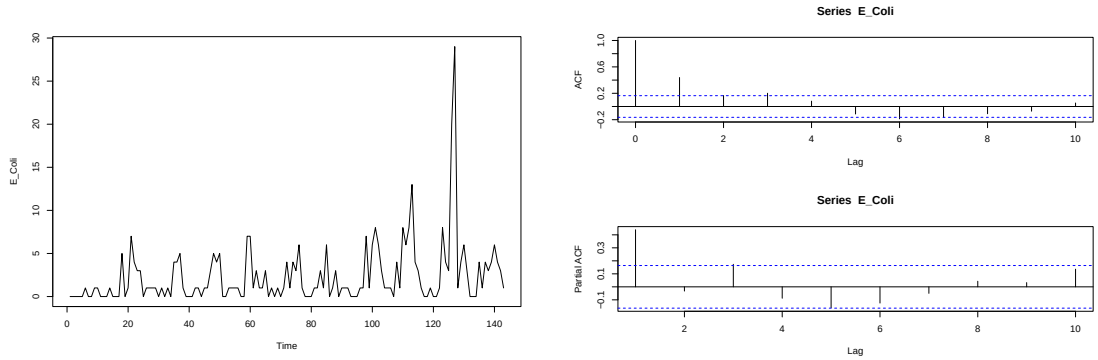

FIGURE 1. Series of E. Coli infections - left panel, from January 1990 to December 2000 with 143 data points, and its ACF and PACF - right panel.

5. PROOFS

Proof of Lemma 1. If q is the distribution of Φ_0, then using (8), we have:

$$\mathbb{V}_q\left(X_t|X_{t-1},\ldots,X_{t-p}\right) \geq \alpha(q)\|Y_t\|^2,$$

where $\alpha(q)$ denotes the smallest eigenvalue of the matrix $v(q)$. From the assumption **(H1)** with $\omega > 0$, $v(q)$ is positive definite and we have:

$$\theta = g(q) \in \Theta \Rightarrow \alpha(q) > 0.$$

Let $h = \inf_{g(q)\in\Theta} \alpha(q)$. Since Θ is supposed to be compact, the subset of real matrices $\{v(q)/g(q) \in \Theta\}$ is bounded. Moreover from the continuity of the determinant, $\inf_{g(q)\in\Theta} \det\left(v(q)\right) > 0$. Then h is positive number and Lemma 1 follows.□

Proof of Lemma 2.

- We first prove that for all $k \geq 1$, we have $\mathbb{P}\left(X_{t-1} \neq 0,\ldots,X_{t-k} \neq 0\right) > 0$. For $k = 1$, the result is obvious since if $X_0 = 0$ a.s then $\xi_0 = 0$ a.s, which is excluded by the assumption **(H1)**. We use an induction argument. If $\mathbb{P}_{q_0}\left(X_{t-1} \neq 0,\ldots,X_{t-k} \neq 0\right) > 0$ for an integer $k \geq 1$, then

$$\mathbb{P}_{q_0}\left(X_{t-1} \neq 0,\ldots,X_{t-k-1} \neq 0\right) = \int_{x_1,\ldots,x_k\neq 0} \mathbb{P}_{q_0}\left(\sum_{j=1}^{p} A_{0,j}x_j + \xi_0 \neq 0\right) \mathbb{P}_{X_{t-1}^-}(dx),$$

where $X_{t-1}^- = (X_{t-j})_{j\geq 2}$. Then if $\mathbb{P}_{q_0}\left(X_{t-1} \neq 0,\ldots,X_{t-k-1} \neq 0\right) = 0$, there exist $x_1,\ldots,x_p \in \mathbb{R}^p$ such that $\mathbb{P}_{q_0}\left(\sum_{j=1}^{p} A_{0,j}x_j + \xi_0 \neq 0\right) = 0$. But this contradicts the assumption **(H1)**, which ensures that $d_{q_0} > 0$, where d_{q_0} is the $\mathbb{L}^2-$distance between

ξ_0 and the vector subspace generated by the random coefficients $A_{0,1}, \ldots, A_{0,p}$ and the constant 1.

- We now prove Lemma 2. Suppose that the assumptions of Lemma 2 hold. For convenience, we set $a_{i,j} = A_{t,j} = b_j = 0$ for $t \in \mathbb{Z}$ and $i, j > p$. For $k \in \mathbb{N}^*$, we denote

$$q_{t,k} = \sum_{i,j \geq k} a_{i,j} X_{t-i} X_{t-j} + \sum_{j \geq k} b_j X_{t-j} + c.$$

We are going to show that:

$$q_{t,k} = 0 \text{ a.s} \Rightarrow \forall j \geq k \quad b_k = a_{k,j} = 0 \text{ and } q_{t,k+1} = 0 \text{ a.s.} \tag{20}$$

Then the conclusion of Lemma 2 will follow from a finite induction on $\{1, \ldots, p\}$ and the equality $q_{t,\ell} = c$.

Suppose that $q_{t,k} = 0$ a.s. Then we have

$$a_{k,k} X_{t-k}^2 + s\left(X_{t-k}^-\right) X_{t-k} + z\left(X_{t-k}^-\right) = 0 \text{ a.s,}$$

where $s\left(X_{t-k}^-\right) = 2\sum_{j \geq k+1} a_{k,j} X_{t-j} + b_k$ and $z\left(X_{t-k}^-\right) = q_{t,k+1}$ a.s. Then conditioning with respect to X_{t-k}^-, we have for $\mathbb{P}_{X_{t-k}^-}$ −almost all x:

$$a_{k,k} \left(\sum_{j=1}^{p} A_{0,j} x_j + \xi_0 \right)^2 + s(x) \left(\sum_{j=1}^{p} A_{0,j} x_j + \xi_0 \right) + z(x) = 0, \text{ a.s.,}$$

and from the first point of the proof, we can suppose $x_1, \ldots, x_p \neq 0$. Then if $a_{k,k} \neq 0$, there exists α, β such that

$$\mathbb{P}\left(\sum_{j=1}^{p} A_{0,j} x_j + \xi_0 \in \{\alpha, \beta\} \right) = 1,$$

which contradicts the assumption **(ID)**. We conclude that $a_{k,k} = 0$ and

$$s\left(X_{t-k}^-\right) \left(\sum_{j=1}^{p} A_{0,j} X_{t-j-k} + \xi_{t-k} \right) + z\left(X_{t-k}^-\right) = 0 \text{ a.s.}$$

In this case $s\left(X_{t-k}^-\right) = 0$ a.s, otherwise there exist x such that the random variable $\sum_{j=1}^{p} A_{0,j} x_j + \xi_{t-k}$ is degenerate which contradicts $d_{q_0} > 0$. If $s\left(X_{t-k}^-\right) = 0$ a.s, we easily deduce $a_{k,j} = b_k = 0$ for $j \geq k+1$. This easily follows by an induction argument, since the conditional variance is always positive (see 14). We conclude that $q_{t,k+1} = 0$ a.s and the proof is complete.□

Proof of Theorem 1. The proof follows the proof of theorem 2.1 in Jeantheau [20] who proved the consistency of the QMLE for general multivariate ARCH models (see theorem 2.1 of that paper). As in [20], we use the following Theorem which is a straightforward generalisation of Theorem 1.12 in Pfanzagl (1969) for i.i.d data.

Theorem 3. *Let $(Y_t)_{t\in\mathbb{Z}}$ be a strictly stationary and ergodic process, θ a parameter in Θ a compact of $\mathbb{R}^d$, and for $n \in \mathbb{N}^*$, let Q_n be a contrast function such that*

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} f\left(Y_t, \ldots, Y_{t-p}; \theta\right),$$

where f is a measurable function with real values and continuous in θ. Suppose that

(1) $\mathbb{E}\inf_{\theta\in\Theta} f\left(Y_0, \ldots, Y_{-p}; \theta\right) > -\infty$.

(2) $\theta \mapsto \mathbb{E}f\left(Y_0, \ldots, Y_{-p}; \theta\right)$ *has a unique finite minimum at θ_0.*

Then, the minimum contrast estimator $\hat{\theta}_n$ associated to Q_n is thus strongly consistent: $\lim_{n\to\infty}\hat{\theta}_n = \theta_0$ *a.s.*

We apply Theorem 3 setting $f\left(X_0, \ldots, X_{-p}; \theta\right) = q_0(\theta)$. Obviously f is continuous in θ.

- Using (14), we have $\inf_{\theta\in\Theta} f\left(Y_0, \ldots, Y_{-p}; \theta\right) \geq \ln\delta$, assumption 1) of Theorem 3 holds for the process X.
- We next prove that assumption 2) holds. Since Φ_t is independent of $\mathcal{F}_{t-1}$ for $t \in \mathbb{Z}$, we have using equation (1):

$$\mathbb{E}_{q_0}\left\{\frac{\left(X_0 - m_0(\theta_0)\right)^2}{V_0(\theta_0)}/\mathcal{F}_{t-1}\right\} = \frac{V_0(\theta_0) + \left(m_0(\theta_0) - m_0(\theta)\right)^2}{V_0(\theta_0)}.$$

  Since the last expression is positive and $\ln\left(V_0(\theta)\right) \geq \ln\left(\delta^2\right)$, the following expression can be deducted:

$$Q(\theta) = \mathbb{E}_{q_0}\left(\frac{V_0(\theta_0) + \left(m_0(\theta) - m_0(\theta_0)\right)^2}{V_0(\theta)} + \ln\left(V_0(\theta)\right)\right). \tag{21}$$

  Now we prove that $Q(\theta_0) = \mathbb{E}q_0(\theta_0) \in \mathbb{R}$ (from the first point we know that $Q(\theta_0)$ is well defined and $\in \mathbb{R}\cup\{\infty\}$. Let $s > 0$ such that $\mathbb{E}|X_0|^s < \infty$ (s exists by **(H2)**). Since $V_0(\theta_0)$ is a polynomial of degree 2 in $X_{-1}, \ldots, X_{-p}$, obviously $\mathbb{E}V_0(\theta_0)^{s/2} < \infty$. Then we obtain:

$$\mathbb{E}\left(\ln V_0(\theta_0)\right) = \frac{2}{s}\mathbb{E}\ln\left(V_0(\theta_0)^{s/2}\right) \leq \frac{2}{s}\ln\mathbb{E}V_0(\theta_0)^{s/2} < \infty. \tag{22}$$

  From (21) and (22) we conclude that $Q(\theta_0) < \infty$.

  Now we prove that for $\theta \in \Theta$, $Q(\theta) \geq Q(\theta_0)$ and the equality holds only when $\theta = \theta_0$.

Since $Q(\theta_0)$ is finite, the difference $Q(\theta) - Q(\theta_0)$ is well defined for $\theta \in \Theta$, and we get:

$$Q(\theta) - Q(\theta_0) = \mathbb{E}\left(\frac{V_0(\theta_0) + (m_0(\theta) - m_0(\theta_0))^2}{V_0(\theta)} - \ln\left(\frac{V_0(\theta_0)}{V_0(\theta)}\right) - 1\right)$$

Since $\big(x - \ln x \geq 1,\ \forall x > 0\big)$ and $\big(x - \ln x = 1 \Leftrightarrow x = 1\big)$ we derive $Q(\theta_0) \leq Q(\theta)$ and:

$$Q(\theta) = Q(\theta_0) \Rightarrow (m_0(\theta), V_0(\theta)) = (m_0(\theta_0), V_0(\theta_0)) \quad a.s$$

As $m_0(\theta) = m(q)'Y_0$ and $V_0(\theta) = Y_0' v(q) Y_0$, assumption **(H3)** implies that $\theta = \theta_0$ which proves that assumption 2) of Theorem 3 holds.

Then the consistency of the QMLE follows from Theorem 3. □

Proof of Theorem 2. Let $t \in \mathbb{Z}$. In Section 3, $q_t(\theta)$ has been defined as:

$$q_t(\theta) = \frac{(X_t - m_t(\theta))^2}{V_t(\theta)} + \ln V_t(\theta).$$

So the first and second derivatives are:

$$\nabla q_t(\theta) = \frac{\nabla V_t(\theta)}{V_t(\theta)}\left(1 - \frac{(X_t - m_t(\theta))^2}{V_t(\theta)}\right) - 2\frac{(X_t - m_t(\theta))\nabla m_t(\theta)}{V_t(\theta)} \tag{23}$$

$$\begin{aligned}\nabla^2 q_t(\theta) = \frac{1}{V_t(\theta)^2}\Bigg[&\nabla V_t(\theta)\nabla V_t(\theta)'\left(2\frac{(X_t - m_t(\theta))^2}{V_t(\theta)} - 1\right)\\ &+\nabla^2 V_t(\theta)\left(1 - \frac{(X_t - m_t(\theta))^2}{V_t(\theta)}\right) + 2(X_t - m_t(\theta))\nabla m_t(\theta)\nabla V_t(\theta)'\\ &+2V_t(\theta)\nabla m_t(\theta)\nabla m_t(\theta)' - 2V_t(\theta)(X_t - m_t(\theta))\nabla^2 m_t(\theta)\\ &+2\left(X_t - m_t(\theta)\right)\nabla V_t(\theta)\nabla m_t(\theta)'\Bigg]\end{aligned}$$

Lemmas 3 to 5 give important properties of $\nabla q_t(\theta)$ and $\nabla^2 q_t(\theta)$. They are required to prove Theorem 2

Lemma 3. *For all $\theta \in \Theta$, the sequences $\{\nabla q_t(\theta)\}_t$ and $\{\nabla^2 q_t(\theta)\}_t$ are ergodic and stationary.*

Proof of Lemma 3. From (23), one can see that for suitable measurable functions $f_\theta : \mathbb{R}^\ell \to \mathbb{R}^d$ and $g_\theta : \mathbb{R}^\ell \to \mathcal{M}_d$, where $\mathcal{M}_d$ denotes the space of square matrices of size d, we have $\nabla q_t(\theta) = f_\theta\left(X_t, \ldots, X_{t-p}\right)$ and $\nabla^2 q_t(\theta) = g_\theta\left(X_t, \ldots, X_{t-p}\right)$. Then the stationarity and the ergodicity of these two sequences follows from the ergodicity and stationarity of X.□

Recall that $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^d$ or the matrix norm associated with $\|\cdot\|$, as required.

Lemma 4. *We have:*

$$\mathbb{E}\left\|\nabla q_0(\theta_0)\right\|^2 < \infty \text{ and } \mathbb{E}\sup_{\theta\in\Theta}\left\|\nabla^2 q_0(\theta)\right\| < \infty.$$

Proof of Lemma 4. Recall that if P is a polynomial of degree q defined on $\mathbb{R}^p$, then there exist non-negative constants $d_0, \ldots, d_p$ such that:

$$\left|P\left(X_{-1}, \ldots, X_{-p}\right)\right| \leq d_0 + \sum_{j=1}^{p} d_j \left|X_{-j}\right|^q, \quad \text{a.s.}$$

Proof of the first assertion: $\mathbb{E}\left\|\nabla q_0(\theta_0)\right\|^2 < \infty$.

We have using (23):

$$\left\|\nabla q_0(\theta_0)\right\| \leq H(\theta_0) + G(\theta_0) \times \frac{\left(X_0 - m_0(\theta_0)\right)^2}{V_0(\theta_0)}, \tag{24}$$

where

$$H(\theta_0) = V_0(\theta_0)^{-2}\left(\|\nabla V_0(\theta_0)\| V_0(\theta_0) + \|\nabla m_0(\theta_0)\|^2 V_0(\theta_0)\right),$$
$$G(\theta_0) = \frac{\|\nabla V(\theta_0)\| + V_0(\theta_0)}{V_0(\theta_0)}.$$

Note that $H(\theta_0)$ and $G(\theta_0)$ are two bounded random variables. Indeed by Lemma 1, we have

$$V_t(\theta_0) \geq h\|Y_t\|^2,$$

and the numerator of $H(\theta_0)$ (resp $G(\theta_0)$) is a polynomial of degree 4 (resp. 2) in $X_{t-1}, \ldots, X_{t-p}$ and the boundedness of $H(\theta_0)$ and $G(\theta_0)$ follows easily from the remark stated at the beginning of the proof.

Moreover, as $X_0 - m_0(\theta_0) = \left(\Phi_0 - m(q_0)\right)' Y_0$, we have

$$\left(X_0 - m_0(\theta_0)\right)^2 \leq \|Y_0\|^2 \times \|\Phi_0 - m(q_0)\|^2,$$

which leads to the bound

$$\frac{\left(X_0 - m_0(\theta_0)\right)^2}{V_0(\theta_0)} \leq \frac{\|\Phi_0 - m(q_0)\|^2}{h}. \tag{25}$$

Then by assumption **(H4)**, the left hand side of the previous bound is a square integrable random variable. Thus, as $H(\theta_0)$ and $G(\theta_0)$ are bounded, the first assertion follows from the bound (24).

Proof of the second assertion: $\mathbb{E}_{q_0} \sup_{\theta\in\Theta} \left\|\nabla^2 q_0(\theta)\right\| < \infty$. The principle is the same than for the proof of the first assertion. Using the expression (**??**), assumption H8) and the lower bound given by Lemma 1, we can bound each term by a random variable not depending on θ and of the form

$$P\left(\|\Phi_t\|\right) \times \frac{Q(Y_0)}{\|Y_0\|^q},$$

where P is a polynomial of degree less than 4, Q is a polynomial of degree $q \leq 4$ in $X_{t-1}, \ldots, X_{t-p}$ In this case, each random variable of the form $\frac{Q(Y_0)}{\|Y_0\|^q}$ is bounded and the second assertion follows.□

Lemma 5. *The entries of the column vectors of the differential of the function* $\theta \mapsto (m_0(\theta), V_0(\theta))$ *evaluated at* θ_0 *are linearly independent random variables.*

Proof of Lemma 5. Suppose there exist constants $\lambda_1, \ldots, \lambda_d$ such that

$$\sum_{i=1}^{d} \lambda_i \partial_i m_0(\theta_0) = 0 \quad \text{a.s} \quad \text{or} \quad \sum_{i=1}^{d} \lambda_i \partial_i V_0(\theta_0) = 0 \quad \text{a.s},$$

where ∂_i denotes the partial derivative with respect to the $i-$th coordinate of $\theta \in \mathbb{R}^d$. Since we have:

$$m_0(\theta) = m(q)'Y_0, \text{ and } V_0(\theta) = Y_0' v(q) Y_0,$$

assumption **(H3)** ensures that

$$\sum_{i=1}^{d} \lambda_i \partial_i m(q) = 0 \text{ and } \sum_{i=1}^{d} \lambda_i \partial_i v(q) = 0,$$

and then $\lambda_1 = \ldots = \lambda_d = 0$ which completes the proof.□

Before giving the demonstration of Theorem 2, let us recall Theorem 4 from [26], which will be used to derive a uniform law of large numbers.

Theorem 4. *Let* Θ *be a compact set of* $\mathbb{R}^d$ *and* $(v_t)_{t\in\mathbb{Z}}$ *a stationary ergodic sequence of random elements with values in the space of continuous functions* $\mathcal{C}\left(\Theta, \mathbb{R}^k\right)$. *Then the uniform strong law of large numbers is implied by*

$$\mathbb{E} \sup_{\theta\in\Theta} \|v_0(\theta)\| < \infty.$$

We are now ready to prove Theorem 2.

Proof of Theorem 2. The technique for the proof of this theorem is very classical, we follow the proof given in [26, Theorem 2.2.1, p. 19]. Since $\theta \in \Theta^\circ$, using a Taylor expansion, we get:

$$0 = \nabla Q_n(\hat{\theta}_n) = \nabla Q_n(\theta_0) + \widetilde{M}_n \cdot (\hat{\theta}_n - \theta_0)$$

where $\widetilde{M}_n$ is the matrix of the second order derivatives, that is :

$$\widetilde{M}_n(i,j) = \partial^2_{ij} Q_n(\gamma_i), \quad 1 \leq i,\, j \leq d.$$

with $\|\hat{\theta}_n - \gamma_i\| \leq \|\hat{\theta}_n - \theta_0\|, \quad i = 1, \ldots, d$. Hence,

$$\sqrt{n} Q_n(\theta_0) = \sqrt{n}\widetilde{M}_n \cdot (\hat{\theta}_n - \theta_0)$$

By Lemma 3, $\left(\nabla^2 q_t(\theta_0)\right)_{t\in\mathbb{Z}}$ is an ergodic stationary sequence. According to Lemma 4, $\sup_{\theta\in\Theta} \left\|\nabla^2 q_0(\theta)\right\|$ is integrable. Then, we can apply Theorem 4 and since $\hat{\theta}_n \to_{n\to\infty} \theta_0$ a.s, we conclude that $\widetilde{M}_n \to_{n\to\infty} F_0 = \mathbb{E}\nabla^2 q_t(\theta_0)$ a.s. More, F_0 is non-singular. Indeed using (23), we have:

$$F_0 = \mathbb{E}\left(V_0(\theta_0)^{-2}\left\{\nabla V_0(\theta_0)\nabla V_0(\theta_0)' + 2V_0(\theta_0)\nabla m_0(\theta_0)\nabla m_0(\theta_0)'\right\}\right),$$

and using Lemma 5, it is easily shown that this matrix is positive-definite. More:

$$\sqrt{n} Q_n(\theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla q_t(\theta_0) \quad \text{and} \quad \mathbb{E}\left(\nabla q_t(\theta_0)/\mathcal{F}_{t-1}\right) = 0.$$

Since by Lemma 4, $\mathbb{E}\left\|\nabla q_0(\theta_0)\right\|^2 < \infty$, the sequence $(\nabla q_t(\theta_0))_t$ is an ergodic stationary $\mathcal{F}_t$-martingale difference sequence of finite variance. Then by [10, Theorem 23.1, p. 206], we have: $\sqrt{n} Q_n(\theta_0) \to_{n\to\infty} \mathcal{N}(0, G_0)$ in distribution, with $G_0 = \mathbb{E}\left(\nabla q_0(\theta_0)\nabla q_0(\theta_0)'\right)$. Consequently, we get:

$$\sqrt{n}(\hat{\theta_n} - \theta_0) \to_{n\to\infty} \mathcal{N}(F_0^{-1} G_0 F_0^{-1}).$$

The expression of G_0 follows from straightforward computations using the expression (**??**).

References

[1] Al-Osh, M. A., Alzaid., A. A. (1987) *First-order integer-valued autoregressive (INAR(1)) process.* J. Time Ser. Anal., **8** (3), 261-275.

[2] Al-Osh, M. A., Alzaid., A. A. (1990) *An integer-valued pth-order autoregressive structure (INAR(p)) process.* J. Appl. Probab., **27**(2), 314-324.

[3] Aue, A., Horváth, L. (2010) *Quasi-Likelihood estimation in stationary and nonstationary autoregressive models with random coefficients.* To appear in Statistica Sinica.

[4] Aue, A., Horváth, L., Steinbach, J. (2006) *Estimation in random coefficient autoregressive models.* J. Time Ser. Anal., **27**(1), 61-76.

[5] Basawa, I.V., Hwang, S.Y. (1998) *Parameter Estimation for generalized random coefficient autoregressive processes.* Journal of Statistical Planning and Inference, **68**, 323-337.

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

[7] Berkes, I., Horvath, L., Ling, S. (2009) *Estimation in nonstationary random coefficient autoregressive models.* J. Time Ser. Anal., **30** (4), 395-416.

[8] Black, F. (1976) *Studies in stock price volatility changes.* Proceedings of the Business and Economic Statistics Section, 177-181.

[9] Berkes, I., Horváth, L., Kokoszka, P. S. (2003) *GARCH processes: structure and estimation.* Bernoulli, **9**, 201-227.

[10] Billingsley, P. (1968) *Convergence of Probability Measures.* John Wiley, New York.

[11] Biswas, A., Songb, P. X. -K (2009) *Discrete-valued ARMA processes.* Statistics and Probability Letters, **79**, 1884-1889.

[12] Bougerol, P., Picard, N. (1992) *Strict stationarity of generalized autoregressive processes.* Ann. Probab., **20** (4), 1714-1730.

[13] Brandt, A., (1986) *The stochastic equation* $Y_{n+1} = A_n Y_n + B_n$ *with stationary coefficients.* Adv. in Appl. Probab., **18** (1), 211-220.

[14] Diaconis, P., Freedman, D. (1999) *Iterated random functions.* SIAM Rev., **41**, 45–76.

[15] Du, J.; Li. Y. (1991) *The integer-valued autoregressive* (INAR(p)) *model.* J. Time Ser. Anal., **12**, 129–142.

[16] Francq, C., Zakoïan, J-M. (2004) *Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes.* Bernoulli, **10**, 605-637.

[17] Francq, C., Makarova, S., Zakoïan, J-M. (2008) *A class of stochastic unit-root bilinear processes. Mixing properties and unit-root test.* Journal of Econometrics, **142** (1), 312-326.

[18] Goldie, C. M. (1991) *Implicit renewal theory and tails of solutions of random equations.* Ann. Appl. Probab., **1** (1), 126-166.

[19] Guivarc'h, Y., and Le Page, E. (2004) *Simplicité de spectres de Lyapunov et propriété d'isolation spectrale pour une famille d'opérateurs de transfert sur l'espace projectif.* In *Random Walks and Geometry (Workshop Vienna 2001)*, V. Kaimanovitch, Ed. De Gruyter, 181-259.

[20] Jeantheau, T. (1998) *Strong consistency of estimators for multivariate ARCH models.* Econometric Theory, **14** (1), 70-86.

[21] Kesten, H. (1973) *Random difference equations and renewal theory for products of random matrices.* Acta Math., **131**, 207-248.

[22] Latour, A. (1998) *Existence and stochastic structure of a non-negative integer-valued autoregressive process.* J. Time Ser. Anal., **19**, 439-455.

[23] Le Page, E. (1983) *Théorème de renouvellement pour les produits de matrices aléatoires. equations aux différences aléatoires.* In Séminaires de probabilités de Rennes.

[24] Kachour, M., Yao, J.F. (2009) *The first-order rounded integer-valued autoregressive (RINAR(1)) process.* J. Time Ser. Anal., **4**, 417-448.

[25] McKenzie, E. (1985) *Some simple models for discrete variate time series.* Water Resources Bulletin, **21**(4), 645-650.

[26] Mikosch, T., Straumann, D. (2006) *Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: A stochastic recurrence equations approach.* Ann. Statist., **34** (5), 2449-2495.

[27] Pfanzagl, J. (1969) *On the mesurability and consistency of minimum contrast estimates.* Metrika, **14**, 249-272.

[28] Sentana, E. (1995) *Quadratic ARCH models.* The Review of Economic Studies, **52** (4), 639-661.

[29] Steutel, F., Van Harn, K. (1979) *Discrete analogues of selfdecomposability and stability.* Ann. Prob., **2**, 893-899.

[30] Straumann, D. (2004) *Estimation in Conditionally Heteroscedastic Time Series Models.* Lecture Notes in Statistics. Springer Verlag.

[31] Weiss, A. A. (1986) *Asymptotic theory for ARCH models: estimation and testing.* Econometric theory **2**, 107-131.

Address:

Lionel TRUQUET, IRMAR, Université de Rennes 1, Campus de Beaulieu, 35042 Rennes, FRANCE
(lionel.truquet@univ-rennes1.fr)

Jian-feng YAO, Department of Statistics and Actuarial Science, University of Hong Kong, Pokfulam, HONG KONG
(jeffyao@hku.hk)